Characterization of Mitochondrial DNA Heteroplasmy at Five Hotspots within the HVI Region of Post-Mortem, Formalin Fixed Paraffinized Human Liver Cells

by

JASON CHARLES KOLOWSKI

A dissertation submitted to the Graduate Faculty in Biochemistry in partial fulfillment
of the requirements for the degree of Doctor of Philosophy,
The City University of New York

2012

This manuscript has been read and accepted for the
Graduate Faculty in Biochemistry in satisfaction of the
dissertation requirement for the degree of Doctor of Philosophy.

Margaret Wallace, Ph.D.

Date — Chair of Examining Committee

Edward Kenelly, Ph.D.

Date — Executive Officer

Lawrence Kobilinsky, Ph.D.

Thomas Schmidt-Glenewinkel, Ph.D.

Mechthild Prinz, Ph.D.

Noelle Umback, Ph.D.

Supervisory Committee

THE CITY UNIVERSITY OF NEW YORK

Abstract

Characterization of Mitochondrial DNA Heteroplasmy at Five Hotspots within the HVI Region of Post-Mortem, Formalin Fixed Paraffinized Human Liver Cells

By

Jason Charles Kolowski

Advisor: Professor Margaret Wallace, Ph.D.

Within the field of mitochondrial DNA (mtDNA) analysis, heteroplasmy is a widely-recognized and yet poorly-understood event. Heteroplasmy is defined as the presence of more than one mitochondrial genome within a tissue sample from a single individual, such that the mtDNA sequence shows the presence of a mixed base or regions of homologous bases that vary in length. Due to the highly conserved nature of the mitochondrial genome, these heteroplasmic events occur at a variety of well-documented hotspots, a majority of which occur within the hypervariable control region that flanks the origin of replication. This control region is the same area that is tested in forensic mtDNA analysis, and is the most useful for establishing the link between evidentiary samples and maternally-related individuals due to the polymorphisms that accumulate in this region. However, when heteroplasmic events are uncovered in forensic mtDNA analysis, issues arise due to the lack of clear understanding of the origin of heteroplasmy. The occurrence of mtDNA heteroplasmy between different tissues within a single individual has been well-established, and heteroplasmic events have been shown to increase with age. However, what is presently unclear is whether or not mtDNA heteroplasmy exists within a single cell when heteroplasmy is present within a tissue. In this regard, three possibilities exist; 1) a single cell contains a solitary pool of mtDNA genomes (defined as homoplastic), and heteroplasmy exists as a mixture of different homoplastic cells within tissue, 2) a single cell

contains a mixture of multiple mtDNA genomes, and heteroplasmy is present within a single cell, or 3) heteroplasmic tissue contains a mixture of homoplastic and heteroplasmic cells due to random cellular distribution throughout the tissue. To investigate this question, laser dissection microscopy will be used to isolate individual cells from a tissue sample with known mtDNA heteroplasmy. Typing of single nucleotide polymorphisms at specific hotspots within the HVI region will then be done to detect possible mtDNA heteroplasmy within a single cell.

Acknowledgements

This research project was funded by grants from the Forensic Science Foundation, Inc., the CUNY Graduate Center Office of Research and Sponsored Programs, and by a private endowment. Special thanks to the following people who assisted throughout the course of the research: Dr. Yingying Tang, Stephanie Pack, Dr. Fred Pugliese, Nick Otis, Abe Zepeda, Tatiana Batson, Dr. Eli Shapiro, Dr. Zoran Budimlija, Dr. Grace Axler-Diperte, Dr. Adele Mitchell, David Markus, and the Dissertation Committee, consisting of Dr. Noelle Umback, Dr. Thomas Schmidt-Glenewinkel, Dr. Mechthild Prinz, Dr. Larry Kobilinsky, and Dr. Margaret Wallace.

Table of Contents

Copyright ii
Approvals iii
Abstract iv
Acknowledgements vi
Table of Contents vii
List of Figures x
List of Tables xvi
List of Terms & Abbreviations xviii
Chapter 1 - Introduction 1
Chapter 2 - Experimental Design 13
Sample Collection 13
Sample Selection 14
Tissue Block Processing 15
Cell Separation and Gross Tissue Generation 16
Gross Tissue Screening for Heteroplasmy 17
Cell-Staining 21
Pathological Examination of Separated Cells 23
Microscopy and Laser Dissection 24
Cell Lysis and HVI SNP Amplification 28
SNP Extension/Termination Probing 30
Capillary Electrophoresis and Fluorescent Detection 32
Data Analysis 33
Experimental Objectives 35
Chapter 3 - Experimental Results 39
Development of a Protocol for Cell Separation/Isolation and Generation of the Gross Tissue Samples 39
Cell Separation and Gross Tissue Creation 40
HVI SNP Assay Development and Validation 45
Validation of the Single-Cell Amplification Using the HVI SNP Primers 54
Primer Titration Studies and Multiplex Design 57
Extension Primer Design and Validation 74
Amplified DNA Purification and Protocol Validation 76
SNaPshot™ Assay and Protocol Validation 78

Cycle Number Titration and Melting Temperature Titration Studies on Individual Primer Sets 85
Initial Amplification and Agilent 2100 Bioanalyzer Data 85
ExoSAP-IT® Treatment and Secondary Agilent 2100 Bioanalyzer Data 98
SNaPshot™ Amplification and Analysis 109
Cycle Number Titration With Test Cells 112
Isolation and Amplification of the Test Cells 112
SNaPshot™ Analysis of the Test Cells 118
Pathology Screening Results 125
Extraction Negative Control Analysis 126
Gross Tissue Sample Analysis 127
Extension Primer Length Study 128
Re-analysis of the Gross Tissue and Length Study Data 136
Age vs. Heteroplasmy of the Gross Tissue 138
Pathology Follow-Up 139
SNaPshot™ Analysis of Single Cells 141
Isolation of Single Cells from Heteroplasmic Samples and Controls 141
Amplification and Agilent 2100 Bioanalyzer Data for Isolated Single Cells 148
SNaPshot™ Amplification of the Single Cells 158
SNaPshot™ Results of Sample 284 Cell Analysis 159
SNaPshot™ Results of Sample 315 Cell Analysis 163
SNaPshot™ Results of Sample 293 Cell Analysis 168
SNaPshot™ Results of Sample 314 Cell Analysis 173
SNaPshot™ Results of Sample 289 Cell Analysis 178
SNaPshot™ Results of Sample 349 Cell Analysis 181
SNaPshot™ Results of Sample 339 Cell Analysis 186
SNaPshot™ Results of Sample 355 (16390) Cell Analysis 191
SNaPshot™ Results of Sample 355 (16519) Cell Analysis 195
SNaPshot™ Results of Sample 369 Cell Analysis 198
Additional testing of non-heteroplasmic tissues- 261 and 342 203
Typing results of non-heteroplasmic tissue showing possible heteroplasmic cells (sample 261, with the 16390 primers) 203
Typing results of non-heteroplasmic tissue showing possible heteroplasmic cells (sample 342, with the 16390 primers) 213
Chapter 4 - Discussion 222

Project Design and Setup 222
Cell Separation and Gross Tissue Creation 223
Gross Tissue Screening 223
Age vs. Heteroplasmy of the Gross Tissue 227
Pathology Review of Cells 228
Microscopy & Laser Dissection Microscopy 229
Cell Lysis and One-Step Amplification 231
Extraction Negative Control Analysis Issues 236
SNaPshot™ Assay Issues 237
SNaPshot™ Results on Single Cells 246
Discussion of SNaPshot™ Results of Sample 284 (16069) Cell Analysis 248
Discussion of SNaPshot™ Results of Sample 315 (16069) Cell Analysis 256
Discussion of SNaPshot™ Results of Sample 293 (16223) Cell Analysis 264
Discussion of SNaPshot™ Results of Sample 314 (16223) Cell Analysis 274
Discussion of SNaPshot™ Results of Sample 289 (16324) Cell Analysis 280
Discussion of SNaPshot™ Results of Sample 349 (16324) Cell Analysis 285
Discussion of SNaPshot™ Results of Sample 339 (16390) Cell Analysis 292
Discussion of SNaPshot™ Results of Sample 355 (16390) Cell Analysis 298
Discussion of SNaPshot™ Results of Sample 355 (16519) Cell Analysis 304
Discussion of SNaPshot™ Results of Sample 369 (16519) Cell Analysis 309
Additional testing of non-heteroplasmic tissues- 261 and 342 316
Discussion of SNaPshot™ results of non-heteroplasmic tissue showing possible heteroplasmic cells (sample 261, with the 16390 primers) 316
Discussion of the SNaPshot™ results of non-heteroplasmic tissue showing possible heteroplasmic cells (sample 342, with the 16390 primers) 321
Implications of Results 331
Sources of Error & Countermeasures 331
Sensitivity of detection systems 332
Future Outcomes 333
Final Conclusions 337
Appendices 339
References 364

List of Figures

Figure 1- Point Heteroplasmy 4
Figure 2- Length Heteroplasmy 5
Figure 3- Age Distribution of Gross Tissue Samples 14
Figure 4- Tissue block, Sample 342 15
Figure 5- Tissue block, Sample 287 15
Figure 6- MitoTracker® Green FM Dye 21
Figure 7- Cells at 40 X, visible light with Nuclear Fast Red 22
Figure 8- Cells at 40 X, FITC filter w/ MitoTracker® Green FM dye 22
Figure 9- Sample 349, Tissue block, HE Stain, 100 X magnification 41
Figure 10- Sample 349, separated cells, NFR stain, 100 X magnification 41
Figure 11- Sample 342, Tissue block, HE Stain, 100 X magnification 42
Figure 12- Sample 342, separated cells, NFR stain, 100 X magnification 42
Figure 13- Sample 355, separated cells, NFR stain, 100 X magnification 43
Figure 14- Sample 261, Tissue block, HE Stain, 100 X magnification 44
Figure 15- Sample 261, separated cells, NFR stain, 100 X magnification 44
Figure 16- Agilent 2100 Bioanalyzer data for 16069 primers 49
Figure 17- Agilent 2100 Bioanalyzer data for 16223 primers 50
Figure 18- Agilent 2100 Bioanalyzer data for 16324 primers 51
Figure 19- Agilent 2100 Bioanalyzer data for 16390 primers 52
Figure 20- Agilent 2100 Bioanalyzer data for 16519 primers 53
Figure 21- Tissue 354, single cell with 16069 primers, 34-cycle amplification 55
Figure 22- Tissue 354, single cell with 16069 primers, 68-cycle amplification 56
Figure 23- Agilent 2100 Bioanalyzer data for 16069 primers, best yield at 5 µM 58
Figure 24- Agilent 2100 Bioanalyzer data for 16223 primers, best yield at 10 µM 59
Figure 25- Agilent 2100 Bioanalyzer data for 16324 primers, best yield at 10 µM 60
Figure 26- Agilent 2100 Bioanalyzer data for 16390 primers, best yield at 5 µM 61
Figure 27- Agilent 2100 Bioanalyzer data for 16519 primers, best yield at 5 µM 62
Figure 28- Agilent 2100 Bioanalyzer data for the Multiplex SNP Primers- 1st attempt .. 63
Figure 29- Agilent 2100 Bioanalyzer data for withholding study- no 16069 primers in Multiplex 66

Figure 30- Agilent 2100 Bioanalyzer data for withholding study- no 16223 primers in Multiplex 67
Figure 31- Agilent 2100 Bioanalyzer data for withholding study- no 16324 primers in Multiplex 68
Figure 32- Agilent 2100 Bioanalyzer data for withholding study- no 16390 primers in Multiplex 69
Figure 33- Agilent 2100 Bioanalyzer data for withholding study- no 16519 primers in Multiplex 70
Figure 34- Agilent 2100 Bioanalyzer data for optimized Multiplex with block amplification design 71
Figure 35- Agilent 2100 Bioanalyzer data for new Multiplex design; HL60 Positive Control DNA 73
Figure 36- Agilent 2100 Bioanalyzer data for new Multiplex design run against tissue sample 347 74
Figure 37- SNaPshot™ Positive Control (Applied Biosystems, 2000) 82
Figure 38- SNaPshot™ Positive Control actual SNP peaks 82
Figure 39- SNaPshot™ Multiplex data for HL60 - Positive Control DNA 84
Figure 40- Melting Temp. and Cycle # Titration Studies, Virtual Gel Image #1 88
Figure 41- Melting Temp. and Cycle # Titration Studies, Virtual Gel Image #2 89
Figure 42- Melting Temp. and Cycle # Titration Studies, Virtual Gel Image #3 90
Figure 43- Melting Temp. and Cycle # Titration Studies, Virtual Gel Image #4 91
Figure 44- Melting Temp. and Cycle # Titration Studies, Virtual Gel Image #5 92
Figure 45- Melting Temp. and Cycle # Titration Studies, Virtual Gel Image #6 93
Figure 46- Melting Temp. and Cycle # Titration Studies, Virtual Gel Image #7 94
Figure 47- Melting Temp. and Cycle # Titration Studies, Virtual Gel Image #8 95
Figure 48- Melting Temp. and Cycle # Titration Studies, Virtual Gel Image #9 96
Figure 49- Melting Temp. and Cycle # Titration Studies, Virtual Gel Image #10 97
Figure 50- Melting Temp. and Cycle # Post ExoSAP-IT®, Virtual Gel Image #1 99
Figure 51- Melting Temp. and Cycle # Post ExoSAP-IT®, Virtual Gel Image #2 100
Figure 52- Melting Temp. and Cycle # Post ExoSAP-IT®, Virtual Gel Image #3 101
Figure 53- Melting Temp. and Cycle # Post ExoSAP-IT®, Virtual Gel Image #4 102

Figure 54- Melting Temp. and Cycle # Post ExoSAP-IT®, Virtual Gel Image #5..... 103
Figure 55- Melting Temp. and Cycle # Post ExoSAP-IT®, Virtual Gel Image #6..... 104
Figure 56- Melting Temp. and Cycle # Post ExoSAP-IT®, Virtual Gel Image #7..... 105
Figure 57- Melting Temp. and Cycle # Post ExoSAP-IT®, Virtual Gel Image #8..... 106
Figure 58- Melting Temp. and Cycle # Post ExoSAP-IT®, Virtual Gel Image #9..... 107
Figure 59- Melting Temp. and Cycle # Post ExoSAP-IT®, Virtual Gel Image #10..... 108
Figure 60- Cycle # Titration Study on One-Step Single Cells, Virtual Gel Image #1 ... 114
Figure 61- Cycle # Titration Study on One-Step Single Cells, Virtual Gel Image #2 ... 115
Figure 62- Cycle # Titration Study on One-Step Single Cells, Virtual Gel Image #3 ... 116
Figure 63- Cycle # Titration Study on One-Step Single Cells, Virtual Gel Image #4 ... 117
Figure 64- Primer Focus Data from Applied Biosystems..... 131
Figure 65- Dual Primer Focus Data from Applied Biosystems..... 131
Figure 66- Agilent 2100 Bioanalyzer results for 34 16069 samples 133
Figure 67- Agilent 2100 Bioanalyzer results for 34 16223 samples 133
Figure 68- Agilent 2100 Bioanalyzer results for 34 16324 samples 134
Figure 69- Agilent 2100 Bioanalyzer results for 34 16390 samples 134
Figure 70- Agilent 2100 Bioanalyzer results for 34 16519 samples 135
Figure 71- Age vs. Instances of Heteroplasmy..... 139
Figure 72- Age vs. Instances of Heteroplasmy, pathology/mixture samples removed .. 141
Figure 73- 355 cells identified..... 144
Figure 74- 355 cells under FITC filter..... 145
Figure 75- 355 cells after LDM cut 146
Figure 76- 355 cells after 1st LDM catapult..... 147
Figure 77- 355 cells after 2nd LDM catapult..... 148
Figure 78- 284 sample cells and control cells, 16069 primers 151
Figure 79- 315 sample cells and control cells, 16069 primers 151
Figure 80- 293 sample cells and control cells, 16223 primers 152
Figure 81- 293 cell #46, 16223 primer 152
Figure 82- 314 sample cells and control cells, 16223 primers 153
Figure 83- 314 cell #1, 16223 primer 153
Figure 84- 289 sample cells and control cells, 16324 primers 154

Figure 85- 289 cell #4, 16324 primer 154
Figure 86- 349 sample cells and control cells, 16324 primers 155
Figure 87- 339 sample cells and control cells, 16390 primers 155
Figure 88- 339 cells with electrophoresis artifacts, 16390 primers 156
Figure 89- 355 sample cells and control cells, 16390 primers 156
Figure 90- 355 sample cells and control cells, 16519 primers 157
Figure 91- 369 sample cells and control cells, 16519 primers 157
Figure 92- Positive Control for 16390 primers, Sample 261 Cells 205
Figure 93- Amplification Negative Control for 16390 primers, Sample 261 Cells..... 206
Figure 94- Cycling Amplification Negative Control, 16390 primers, Sample 261 Cells206
Figure 95- Overlay of Agilent 2100 Bioanalyzer Data, 16390 primers, 261 cells 207
Figure 96- Overlay of Agilent 2100 Bioanalyzer Data, 16390 primers, 261 cells and controls 208
Figure 97- Overlay of Agilent 2100 Bioanalyzer Data, 16390 primers, 329 controls for 261 cells 208
Figure 98- Positive Control for 16390 primers, Sample 342 Cells 214
Figure 99- Amplification Negative Control for 16390 primers, Sample 342 Cells..... 215
Figure 100- Cycling Amplification Negative Control, 16390 primers, Sample 342 Cells215
Figure 101- Overlay of Agilent 2100 Bioanalyzer Data, 16390 primers, 342 cells 216
Figure 102- Overlay of Agilent 2100 Bioanalyzer Data, 16390 primers, 342 cells and control cells 216
Figure 103- Peak Heights for 284 cells, 16069 primers 250
Figure 104- T Peak Heights vs. C Peak Heights for 284 cells, 16069 primers 253
Figure 105- Peak Heights for 315 cells, 16069 primers 256
Figure 106- A Peak Heights vs. C Peak Heights for 315 cells, 16069 primers 260
Figure 107- Valid A Peak Heights vs. C Peak Heights for 315 cells, 16069 primers 260
Figure 108- Peak Heights for the 293 cells, 16223 primers 265
Figure 109- A Peak Heights vs. C Peak Heights for 293 cells, 16223 primers 268
Figure 110- Valid A Peak Heights vs. C Peak Heights for 293 cells, 16223 primers 269
Figure 111- T Peak Heights vs. C Peak Heights for the 293 cells, 16223 primers..... 271
Figure 112- Valid T Peak Heights vs. C Peak Heights for the 293 cells, 16223 primers272

Figure 113- Peak Heights for 314 cells, 16223 primers 274
Figure 114- T Peak Heights vs. C Peak Heights, 314 cells, 16223 primers 275
Figure 115- Valid T Peak Heights vs. C Peak Heights, 314 cells, 16223 primers 276
Figure 116- Peak Heights of 289 cells, 16324 primers..... 281
Figure 117- Peak Heights for 349 cells, 16324 primers 285
Figure 118- G Peak Heights vs. T Peak Heights for 349 cells, 16324 primers 288
Figure 119- Valid G Peak Heights vs. T Peak Heights for 349 cells, 16324 primers 289
Figure 120- Peak Heights for 339 cells, 16390 primers 292
Figure 121- T Peak Heights vs. C Peak Heights for 339 cells, 16390 primers 295
Figure 122- Valid T Peak Heights vs. C Peak Heights for 339 cells, 16390 primers..... 296
Figure 123- Peak Heights for 355 cells, 16390 primers 299
Figure 124- T Peak Heights vs. C Peak Heights for 355 cells, 16390 primers 301
Figure 125- Peak Heights for 355 cells, 16519 primers 304
Figure 126- Peak Heights for 369 cells, 16519 primers 309
Figure 127- C Peak Heights vs. G Peak Heights for 369 cells, 16519 primers 312
Figure 128- Valid C Peak Heights vs. G Peak Heights for 369 cells, 16519 primers 313
Figure 129- Peak Heights for 261 cells, 16390 primers 316
Figure 130- C Peak Heights at ~38 bp vs. C Peak Heights at ~36 bp, 261 cells, 16390 primers 318
Figure 131- T Peak Heights vs. C Peak Heights at ~36 bp, 261 cells, 16390 primers ... 318
Figure 132- T Peak Heights vs. C Peak Heights at ~38 bp, 261 cells, 16390 primers ... 319
Figure 133- Peak Heights of 342 cells, 16390 primers..... 322
Figure 134- G Peak Height vs. C Peak Height at ~37 bp, 342 cells, 16390 primers..... 323
Figure 135- C Peak Height at ~39 bp vs. C Peak Height at ~37 bp, 342 cells, 16390 primers 324
Figure 136- T Peak Height vs. C Peak Height at ~37 bp, 342 cells, 16390 primers 324
Figure 137- Valid G Peak Height vs. C Peak Height at ~37 bp, 342 cells, 16390 primers325
Figure 138- Valid C Peak Height at ~39 bp vs. C Peak Height at ~37 bp, 342 cells, 16390 primers 326
Figure 139- Valid T Peak Height vs. C Peak Height at ~37 bp, 342 cells, 16390 primers326
Figure 140- T and ~39bp C Peak Heights of 342 cells, 16390 primers..... 327

Figure 141- T Peak Height vs. C Peak Height at ~39 bp, 342 cells, 16390 primers 327
Figure 142- Valid T Peak Height vs. C Peak Height at ~39 bp, 342 cells, 16390 primers328

List of Tables

Table 1- SNP base-pair sizing 53
Table 2- Optimal SNP primer concentrations 63
Table 3- Setup scheme for primer withholding study 65
Table 4- Unmodified extension primers sequences 75
Table 5- Extension primer sequences modified with 5′ poly-thymine additions 75
Table 6- Data for ethanol-purified mtDNA samples, 0 days after purification 77
Table 7- Data for ethanol-purified mtDNA samples, 2 days after purification 77
Table 8- Setup scheme for SNaPshot™ assay validation 79
Table 9- SNaPshot™ Positive Control expected values (Applied Biosystems, 2000) 81
Table 10- Multiplex SNP peaks 83
Table 11- Cycle # and Tm Titration Samples Sent to SNaPshot™ 110
Table 12- Amplification Negative Controls from Cycle # and Tm Titration 110
Table 13- HL60 Replicate Samples From Cycle # and Tm Titration in SNaPshot™ 111
Table 14- Amplification Control Results for Test Cell in SNaPshot™ 118
Table 15- Tissues and SNP profiles for Cycle # Titration Study 119
Table 16- SNaPshot™ Results for Cycle # Titration Study, 16390 primers, test cells 120
Table 17- SNaPshot™ Results for Cycle # Titration Study, 16519 primers, test cells 120
Table 18- d-Rhodamine Dye Information 129
Table 19- Molecular Weights of dNTP bases 130
Table 20- Bin Sizes Based on Length Study 136
Table 21- Controls and Heteroplasmic Samples Selected for SNaPshot™ Analysis of Single Cells 143
Table 22- Control Sample Results for 284 Cells in SNaPshot™ 159
Table 23- SNaPshot™ Results for A/T heteroplasmic 284 Cells and Controls, 16069 Primers 161
Table 24- Control Sample Results for 315 Cells in SNaPshot™ 164
Table 25- SNaPshot™ Results for C/A heteroplasmic 315 Cells and Controls, 16069 Primers 165
Table 26- Control Sample Results for 293 Cells in SNaPshot™ 168
Table 27- SNaPshot™ Results for C/T heteroplasmic 293 Cells and Controls, 16223 Primers 170
Table 28- Control Sample Results for 314 Cells in SNaPshot™ 173
Table 29- SNaPshot™ Results for C/T heteroplasmic 314 Cells and Controls, 16223 Primers 175

Table 30- Control Sample Results for 289 Cells in SNaPshot™ 178
Table 31- SNaPshot™ Results for G/T heteroplasmic 289 Cells and Controls, 16324 Primers 179
Table 32- Control Sample Results for 349 Cells in SNaPshot™ 182
Table 33- SNaPshot™ Results for A/T heteroplasmic 349 Cells and Controls, 16324 Primers 183
Table 34- Control Sample Results for 339 Cells in SNaPshot™ 186
Table 35- SNaPshot™ Results for C/T heteroplasmic 339 Cells and Controls, 16390 Primers 188
Table 36- Control Sample Results for 355 (16390) Cells in SNaPshot™..... 191
Table 37- SNaPshot™ Results for C/T heteroplasmic 355 Cells and Controls, 16390 Primers 192
Table 38- Control Sample Results for 355 (16519) Cells in SNaPshot™..... 195
Table 39- SNaPshot™ Results for G/A heteroplasmic 355 Cells and Controls, 16519 Primers 196
Table 40- Control Sample Results for 369 Cells in SNaPshot™ 198
Table 41- SNaPshot™ Results for G/A heteroplasmic 369 Cells and Controls, 16519 Primers 200
Table 42- Control Sample Results for 261 Cells in SNaPshot™ 208
Table 43- SNaPshot™ Results for 261 Cells and Controls, 16390 Primers..... 210
Table 44- Control Sample Results for 342 Cells in SNaPshot™ 217
Table 45- SNaPshot™ Results for 342 Cells and Controls, 16390 Primers..... 218
Table 46- Comparison of AN/CAN control peak heights from SNaPshot™ data 241

List of Terms & Abbreviations

3′	3rd carbon in the DNA ribose molecule that has an attached hydroxyl group; serves at the connection point in the replicating DNA strand for incoming dNTPs
5′	5th carbon in the DNA ribose molecule that has an attached phosphate group; serves as the anchoring carbon for a newly-attached dNTP
A	adenine, one of the four bases of DNA
BLAST	basic local alignment search tool
bp	base-pair
C-stretch	cytosine-rich regions of the mitochondrial hypervariable regions
C	cytosine, one of the four bases of DNA
˚C	degrees Celsius
CCD	charge-coupled device
cm	centimeter
cycle	a unit of measure of a PCR reaction, consisting of denaturation, annealing and extension temperatures at defined time intervals
D-loop	displacement loop, denoted as the origin of replication in the circular mitochondrial genome
DNA	deoxyribonucleic acid
dNTP	deoxy-nucleoside triphosphate
ddNTP	di-deoxy nucleoside triphosphate
DOHMH	Department of Health and Mental Hygiene
ExoSAP-IT®	exonuclease / shrimp alkaline phosphatase reagent
FITC	common fluorescence filter used on CCD detectors, based on the wavelength of fluorescein isothiocyanate
G	guanine, one of the four bases of DNA

HPLC	high-performance liquid chromotography
HVI	hypervariable region I
HVII	hypervariable region II
IRB	Internal Review Board
l	liter
MALDI-TOF	matrix-assisted laser desorption/ionization – time-of-flight
ml	milliliter
mtDNA	mitochondrial DNA
ng	nanogram
nm	nanometer
nM	nanomolar
NYC	New York City
OCME	Office of Chief Medical Examiner
PALM	photo-activated localization microscopy
PCR	polymerase chain reaction
PEN	Polyethylene Naphthalate
rCRS	revised Cambridge Reference Sequence
RPM	rotations per minute
SAP	shrimp alkaline phosphatase
SNaPshot™	SNP-testing kit from Applied Biosystems Corporation
SNP	single nucleotide polymorphism
STR	short tandem repeat
T	thymine, one of the four bases of DNA

Taq	thermo-stable DNA polymerase harvested from *Thermus aquaticus*
Transversion	When used to reference a DNA substitution mutation (such as mtDNA heteroplasmy), it denotes the change of a purine to a pyrimidine, or vice versa
Transition	When used to reference a DNA substitution mutation (such as mtDNA heteroplasmy), it denotes the change of one purine base for another purine base, or a pyrimidine base for another pyrimidine base
µl	microliter
µm	micrometer
µM	micromolar

Chapter 1 - Introduction

Mitochondria are intracellular organelles found in eukaryotic cells, and are responsible for the energy production of such cells (Okamoto & Shaw, 2005). Mitochondria are unique organelles in that they contain their own DNA, referred to as mitochondrial DNA, or mtDNA. Mitochondrial DNA exists as circular DNA structures within the mitochondrial matrix, and each genome of human mtDNA contains approximately 16,569 base-pairs (bp) (Anderson et al., 1981; Gray, 1989). In the past decade, the application of decoding and understanding the mitochondrial genome has benefited the sciences by clarifying the cellular role of the mitochondrion in energy production, and also from a broader aspect, by tracing maternal lineages to show ancestry and human migratory patterns across the millennia of human history (Aquadro & Greenberg, 1983; Boles et al., 1995; Holland et al., 1995; Lutz et al., 1996; Paabo et al., 1988).

A remarkable amount of the mtDNA genome is devoted to encoding the genes for oxidative phosphorylation and cellular energy function (Bai et al., 2004; Bai et al., 2000). The human mtDNA genome is also unique in that it contains no introns, therefore, the genome is highly conserved (Bai et al., 2000; Gray, 1989). In order to maintain the functionality of the mtDNA genome, there are two highly polymorphic regions that flank the origin of replication within the displacement region (D-loop) of the mitochondrial DNA, and it is theorized that these polymorphic regions are designed to absorb mutations that would otherwise disrupt the highly conserved nature of the genome (Sigurðardóttir et al., 2000). These regions have been classified as hypervariable regions I and II, respectively (Anderson et al., 1981; Andrews et al., 1999). A third hypervariable region, region III, has also been identified within the mtDNA control region (Paneto et al., 2010). It is within these hypervariable regions that ancestry and forensics utilize

the maternal inheritance of the mtDNA genome to compare familial linkages (Allen et al., 1998; Benecke, 1997; Boles et al., 1995; Brown & Brown, 1994; Budowle et al., 2003; Butler & Levin, 1998; Corach et al., 1997; Giles et al., 1980; Gill et al., 1994; Hutchinson et al., 1974; Ivanov et al., 1996; Stone et al., 2001; Wilson & Allard, 2004; Wilson & Cann, 1992).

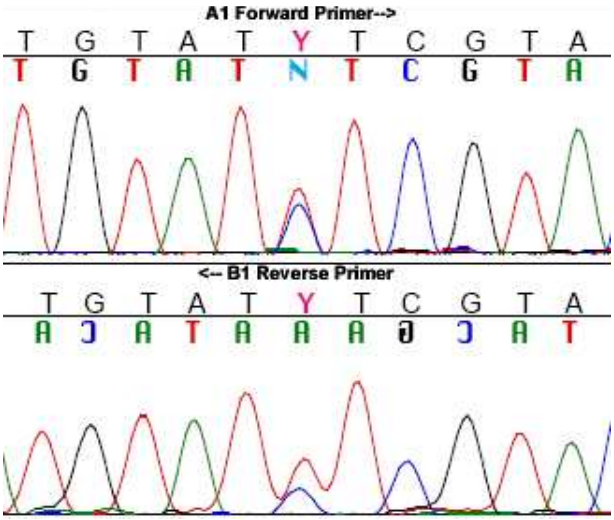
The ability of mtDNA to be utilized in forensic and anthropological settings, where samples that are typically encountered are degraded tissues, bones, and hair, is due to the high copy number of mitochondria in the cells (Bogenhagen & Clayton, 1974). The recovery of mtDNA from such samples is enhanced further due to the high number of mitochondrial genomes present within a single mitochondrion; depending on the cell type, it is estimated that each mitochondrion contains up to 100 genomes per mitochondrion, and each cell can contain 100 to 1000 or more mitochondria depending on cell type (Akane et al., 1993). The prevalence of the mtDNA genome within a single cell then provides a large number of template mtDNA that can be extracted and analyzed. In order to visualize mtDNA for forensic or ancestral analysis, the sequence of the mtDNA hypervariable regions must be generated. To accomplish this, the hypervariable regions are artificially replicated using polymerase chain reaction (PCR) technology (Mullis, 1990). During the replication of the hypervariable regions, fluorescently-tagged di-deoxynucleoside triphosphate terminator bases are randomly incorporated into the growing mtDNA strand (Sanger, 1981; Sanger et al., 1977). After approximately 30 rounds of PCR amplification, every possible position along the mtDNA sequence should be accounted for by a fluorescently-labeled terminator base. Since each base termination results in a unique size-specific fragment of mtDNA, these fluorescently-tagged fragments can be separated using electrophoresis (Butler et al., 1998; Holland & Parsons, 1999; Hopgood et al., 1992; Levin et al., 1999; Levin, Hancock et al., 2003; Stewart et al., 2003; Sullivan et al., 1992; Wilson, DiZinno et

al., 1995). At the end of the electrophoretic separation, the fragments sequentially pass through a laser beam and the fluorescent tag on the terminator base is activated. The tags specifically emit one of four possible wavelengths which are detected by a CCD camera within an automated sequencer. The wavelength detected by the instrument is then interpreted by the computer software as either being adenine, thymine, guanine, or cytosine. Each base is therefore represented by a discrete wavelength, and when analyzed in order, from shortest fragment to longest fragment, the sequence of the mtDNA is established. At times during this mtDNA sequencing, heteroplasmy is detected.

Heteroplasmy has been classically defined as the presence of more than one genome of mitochondrial DNA within an individual (Aquadro & Greenberg, 1983; Greenberg et al., 1983; Parsons et al., 1997). Several studies, both at the New York City Office of Chief Medical Examiner and at external laboratories, have shown that heteroplasmic events can vary between tissues from an individual (Alonso et al., 1996; Bendall et al., 1997; Budowle, Wilson et al., 1999b; Calloway et al., 2000; Holland et al., 1995; Isenberg & Moore, 1999; Jazin et al., 1996; Kolowski et al., 2004; Miller et al., 1996; Pfeiffer et al., 1999; Roberts & Calloway, 2011; Wilson, Polanskey et al., 1995). In such cases, one tissue type from a single individual is seen to contain no heteroplasmy (or a "clean" mtDNA sequence), and another tissue from the same individual displays a heteroplasmic event. To date, there are two classifications of heteroplasmic events that are commonly encountered during mtDNA analysis. The first is referred to as "point heteroplasmy," where a single location along the mtDNA sequence is found to contain two different bases of DNA. Since mtDNA sequencing analysis is a combination of all of the possible mtDNA genomes within the sample, this is not truly a mixed base in a single mtDNA genome, but the presence of two different genomes within the total sample. Therefore, the

sample consists of different mtDNA genomes containing one unique base at a specific position when compared to the other mtDNA genomes within the total sample. This point heteroplasmy is believed to be due to point mutations within the hypervariable regions that cause two different mtDNA sequences to manifest within one individual (Figure 1). When these point mutations occur in the mtDNA, they appear to be random, despite the molecular mechanisms driving twice as many possible transversions versus a transition of the original base (Carr, 2010). Although point heteroplasmy occurs in a moderate percentage of samples, when it does appear it manifests at specific hotspots within the mtDNA hypervariable regions with high regularity (Li et al., 1999; Santos et al., 2008).

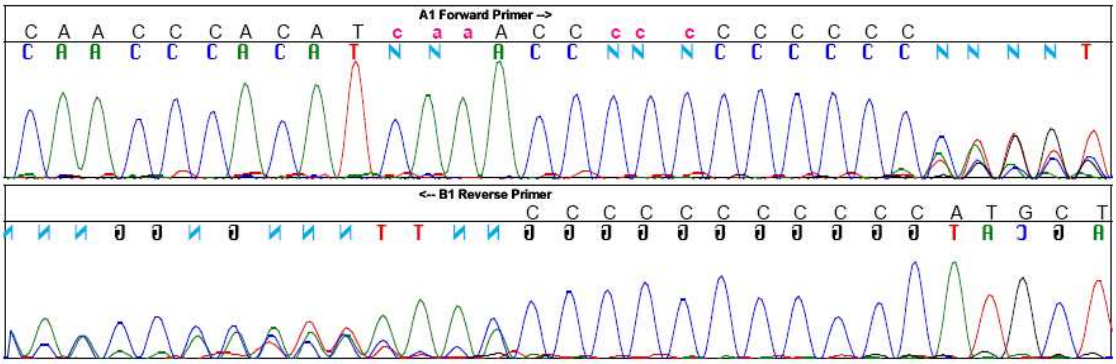
Figure 1- Point Heteroplasmy



The second type of commonly encountered mtDNA heteroplasmy is "length heteroplasmy." This type of heteroplasmy is observed at sequencing analysis as a mixture of different lengths of repeating DNA sequence, commonly within a cytosine-rich "C-stretch" within each hypervariable region. These "C-stretches" are sections of the mtDNA genome which contain up to twelve continuous cytosine residues, commonly interrupted by a thymine residue within the cytosine bases. *Taq* polymerase, which is commonly used in forensic and ancestral PCR analysis, has a moderate fidelity resulting in an inherent low error rate. This moderate fidelity causes the polymerase to "slip" when encountering more than eight continuous identical nucleotides during the PCR extension phase. Therefore, if the thymine that commonly

interrupts the cytosine region transitions to a cytosine, the "C-stretch" becomes longer than can be efficiently amplified with *Taq* polymerase. Due to this polymerase slippage, random relocation of the polymerase onto the template DNA creates various mis-reads of the template sequence, and therefore data past the "C-stretch" region is uninformative when this occurs (Figure 2). When length heteroplasmy is observed during sequencing analysis, various amplification strategies using additional primers must be employed to work around the affected "C-stretch" region to recover as much of the remaining mtDNA sequence as possible. Within the HVI region, the thymine that normally sits in the C-stretch at position 16189 is a highly publicized hotspot for cytosine transition occurrences.

Figure 2- Length Heteroplasmy



Within forensic settings, these two forms of heteroplasmy therefore constitute a paradoxical situation. When a point mutation event occurs and a point heteroplasmy is revealed, the presence of this seemingly rare event increases the odds that an unknown sample and a reference sample are the "same" if this point heteroplasmy is seen in both samples. However, due to the fact that different tissues from a single individual can yield different heteroplasmic results, combined with the high mutation rate within the HVI region of the mtDNA genome, the absence of such a point mutation from one sample when compared to another does not imply that the samples are unrelated; if all other common base changes are accounted for, the samples are

simply inconclusive (Salas et al., 2001). In fact, to remain as conservative as possible, forensic mtDNA analysis procedures use specific wordings such as “cannot be excluded” or “can be excluded” when reporting the results of mtDNA testing (SWGDAM, 2003). There is never any mention of “inclusions” or “matches” due to the issues surrounding the HVI mutation rate and the possible heteroplasmy that can result (Holland & Parsons, 1999).

To further compound the issue surrounding mtDNA analysis, the thymine-to-cytosine transition at position 16189 that ultimately leads to length heteroplasmy cannot be directly analyzed or interpreted. Within forensic settings standard protocol dictates that if length heteroplasmy is encountered, the event is simply reported as being present due to the issue of the fidelity of *Taq* polymerase and the inability to clearly resolve the actual sequence of the “C-stretch” region upon analysis. While attempts have been made to standardize the reporting of length heteroplasmy events, there has been disagreement in the past between laboratories on the interpretation of the results and the proper procedure for reporting the findings (Carracedo et al., 2000; Carrecedo et al., 1998; Miller & Budowle, 2001; Paneto et al., 2007; Tully et al., 2001; Wilson et al., 2002a; Wilson et al., 2002b). Recent attempts at standardizing the nomenclature and reported mitotype have been proposed to attempt to reconcile these issues within the forensic mtDNA field (Budowle et al., 2010; Polanskey et al., 2010).

To understand the basis of heteroplasmy, it is necessary to examine the origin of a cell. During oogenesis (female egg development), it is theorized that an isolation event selects for a single mitochondrion source, creating a “bottleneck” in which as many as 100,000 different mitochondria are randomly isolated down to a pool of approximately 200 mitochondria through cellular division (Marchington et al., 1998). As a result of this genetic bottleneck, any dominant mitochondrial type within the 200 mitochondria is preferentially selected for as the “dominant

homoplastic" mitochondrial DNA type, although minor types could be detected depending on the frequency of the minor contributing mtDNA (Melton, 2004). Note that this specific conclusion is conservative, implying that other mitotypes may be present in the single cell at low levels. However, because heteroplasmy is known to occur at mtDNA hotspots with relative frequency, this contradicts the bottleneck theory, and therefore the presence of two or more mtDNA genomes in a cell line could be the result of unequal partitioning of the cell during oogenesis, which is more akin to random genetic drift (Ashley et al., 1989; Chinnery et al., 2000).

A second theory on the nature of heteroplasmy focuses on the possible recombination of the mtDNA genomes before mitochondrion division, which leads to the mitochondrial heteroplasmy (Awadalla et al., 1999). Recombination events are known to occur later in development, commonly within specialized cells (Zsurka et al., 2007). Recombination events are important within the nuclear DNA, and are known to give rise to genetic variations which support the distribution of genes throughout a population (Kraytsberg et al., 2004). When applied to the mtDNA genomes, however, there is some disagreement as to whether or not mtDNA recombination occurs (Eyre-Walker & Awadalla, 2001; Eyre-Walker et al., 1999; Hoelzel et al., 1994; Macaulay et al., 1999; Zsurka et al., 2005). If recombination does occur, one of the major questions is how it leads to heteroplasmic events only in the hypervariable regions, and not within the conserved genes that make up a majority of the compact mtDNA genome (Ashley et al., 1989; Eyre-Walker & Awadalla, 2001; Eyre-Walker et al., 1999; Hoelzel et al., 1994; Macaulay et al., 1999; Zsurka et al., 2005). What is clear, however, is that if heteroplasmy rates are high due to recombination events, the rate only affects 1% or less of the mitochondrial genomes within a single individual, so the possibility of detecting the

heteroplasmy *in vivo* is close to zero (Kraytsberg et al., 2004; Slate & Gemmell, 2004; Stoneking et al., 1991; Zsurka et al., 2007).

A third theory of heteroplasmy is similar to the issues of mitochondrial recombination, but instead focuses on the fusion and fission of mitochondrial organelles, independent of recombination. Mitochondrial fusion is known to occur, and at the cellular level this fusion is responsible for the mixing of mitochondrial genomes (Bereiter-Hahn & Vöth, 1994; Chen & Chan, 2005; Chen et al., 2005; Nunnari et al., 1997; Okamoto & Shaw, 2005; Yaffe, 1999). The fusion and fission of the mitochondria play an important role in maintaining the morphology of the organelles within the cell, and assisting in cellular mechanisms of apoptosis (Lee et al., 2004; Legros et al., 2002; Perfettini et al., 2005). If an assumption is made that in a "primitive" or early-growth stage heteroplasmic cell there are unique mitochondria, each with a unique mtDNA genome set, then a scenario could exist in which random fusion and fission result in a random mixing of the two (or more) different mtDNA genomes within a single mitochondrion. In contrast, imagine that the original assumption still holds, but additionally there is a mechanism in which a unique mitochondrion recognizes and fuses only with another mitochondrion that contains the same mtDNA genomes. In this second scenario, the fusion and fission still occurs, but the mixing of the heteroplasmic mtDNA genomes at the organelle level does not occur. Out of the two scenarios, the first is more likely to occur due to simplicity, as the second scenario requires a high level of cellular mechanics. As a counter point, the second scenario cannot be completely discarded due to the cellular mechanism of lateral inhibition (Collier et al., 1996). Cellular differentiation is dictated by the lateral inhibition process, for example, how the development of a nerve cell directs the surrounding cells to become another cell type, to eliminate sympathetic crosstalk in the nervous system. It is possible that a similar mechanism

could be underlying the process of differentiating other specialized cells based on the energy requirement functioning of the cell, which would be tied directly to mitochondrial prevalence. Regardless of the possible scenario, the mitochondrial fusion and fission is known to occur, and either of the scenarios can explain the nature of cellular heteroplasmy but leaves the state of mtDNA heteroplasmy at the organelle level in question.

A fourth theory of heteroplasmic origins focuses on the possibility of a paternal contribution to the oocyte's mtDNA pool. It was typically believed that the sperm mitochondria do not physically enter the female egg, since the sperm mitochondria are in the tail region and the head of the sperm only interacts with the cell membrane of the oocyte (Sutovsky et al., 1996; Sutovsky et al., 1999). However, recent studies have shown that the tail region of the sperm is incorporated into the fertilized egg due to the paternal mitotype which can be detected during embryogenesis, but after embryo formation the paternal mtDNA genomes are somehow "lost" (Ankel-Simons & Cummins, 1996). The prior theory of the typical interaction explains the maternal-only contribution of the mtDNA genomes to the offspring. However, because of the possibility that the sperm cell enters the egg by some type of intracellular incorporation, the paternal mitochondria could contribute to the mtDNA pool for the developing offspring (Slate & Gemmell, 2004). Since the sperm mtDNA contribution would be low, most of the differences would not be detectible, but over time and billions of cellular divisions, the paternal contribution could show up as a heteroplasmic event upon analysis of the mtDNA (Schwartz & Vissing, 2002, 2004).

A fifth theory regarding mtDNA heteroplasmy ignores possible genetic origins and focuses instead on the influence of the amplification strategies that are commonly used to analyze and visualize the mtDNA sequence. *Taq* polymerase, which is commonly used in DNA

PCR amplification, is known to have a typical error rate of $1x10^4$. In terms of sequencing strategies *Taq* is considered a moderate-fidelity polymerase, and it would be expected to see a possible nucleotide mis-incorporation approximately every ten thousand bases using this polymerase (Chong et al., 2005; Gyllensten & Erlich, 1988; Hopgood et al., 1992; Levin, Hancock et al., 2003). It is also theorized that specific DNA sequences enhance the possibility of a base mis-incorporation, resulting in "hotspots" in the genome. These hotspot locations enable nucleotide-polymerase interactions that give rise to base incorporation mistakes, which ultimately manifest as heteroplasmy (Brandstatter & Parson, 2003; Zhang & Deisseroth, 1991). To overcome the amplification issues that might result in mtDNA heteroplasmy, it has been suggested to utilize a high-fidelity polymerase, or focus instead on probing single nucleotide polymorphisms (SNPs) to lower the odds of a base mis-incorporation compared to continuous DNA sequencing (Brandstatter & Parson, 2003; Budowle, 2001; Budowle et al., 2004; Kline et al., 2005). An additional factor affecting the detection and analysis of heteroplasmy may be the instrumentation used to detect the differences in the mitochondiral sequence. Recent studies have demonstrated that the detectable level of heteroplasmy can vary for a sample depending on the type of laboratory instrumentation and detection that is being used; in particular, the differences seen between heteroplasmy levels from capillary electrophoresis data and next-generation solid-phase sequencing data (Bandelt & Salas, 2011).

Another possible cause of mtDNA heteroplasmy is in relation to the age of the individual. It has been shown that mutations accumulate over time, and specific mutations in the mtDNA genome can correlate to the age of an individual. This phenomenon would be due to the production of reactive oxygen species that manifest within the organelle due to the energy-production role of the mitochondria. Over the lifespan of an individual, these reactive oxygen

species accumulate and cause oxidative stress on the mtDNA genomes, resulting in an increase of age-related mutations (Mullis, 1990; Richter et al., 1988). One such mutation is an age-dependent deletion event which occurs at the 4977 bp position in the coding region of the mtDNA genome, and has a devastating effect on several genes responsible for polypeptide and tRNA production at that location (Meissner et al., 1997; Porteous et al., 1998). As a result of this deletion, the mitochondrion loses the ability to transcribe the proteins necessary for cellular energy production. Additional age-related mutations within the mtDNA genome can give rise to specific mitochondrial-related illnesses (Goto et al., 1990). Due to the fact that several issues affect the mutation rate for the mtDNA genome, the common rate for mtDNA mutation is unknown (Lopez et al., 1997). However, it is known that as living organisms age, unchecked mutational events can compound, and in a small, conserved genome such as in the mitochondrion, these compound mutations in random mtDNA genomes could appear as heteroplasmic events upon sequence analysis (Chen et al., 2003; Wallace, 1995, 1997). While the other theories of the cellular origin of mtDNA heteroplasmy are more likely, the effects of age on the mtDNA genome have been shown to correlate with higher rates of heteroplasmy, and higher mutation rates in general (Calloway et al., 2000). However, the rate at which mtDNA heteroplasmy manifests in the hypervariable regions is still in question, and there is some evidence that the hypervariable control region's heteroplasmic rate might be static as the age of an individual increases (Lagerstrom-Fermer et al., 2001).

Regardless of the initial cause, it is known that mtDNA heteroplasmy occurs between tissues as a result of the cell differentiation, and therefore between the cells of different tissues. However, within tissues where heteroplasmy occurs, it is unclear whether the heteroplasmy occurs between the individual cells, between the mitochondria in the cell, or between the

genomes within an individual mitochondrion. In a 2008 study conducted by Sabine Lutz-Bonengel and colleagues, blood lymphocytes isolated using flow-cytometry displayed mainly homoplastic mtDNA types, indicating that blood heteroplasmy is a mixture of heteroplasmic cells (Lutz-Bonengel et al., 2008). A more recent 2010 study conducted by Joseph Reiner and his team used optical tweezers to isolate actual mitochondria from HL60 cells, and the testing of these isolates demonstrated heteroplasmy in the segregated mitochondrion at the same ratio that was seen in the heteroplasmic cells (Reiner et al., 2010). Clearly, the discrepancy between these two studies shows that are aspects to the overall question of cellular heteroplasmy that remain outstanding. What is clear is, however, is that each mitochondrion can contain approximately one-hundred genomes of mtDNA, and each cell can contain a range of hundreds to thousands of mitochondria. It is unclear when sampling even a single-cell source whether or not the heteroplasmic event at the gross tissue level is occurring within the mtDNA genomes of a single cell, or whether a single cell hosts a solitary, unique pool of mtDNA genomes.

The specific aim of this research is to gain a better understanding of how two or more mitochondrial genomes can occur within an individual, resulting in heteroplasmy. The question to be answered is; with isolation of a solitary cell from a tissue with known heteroplasmy and high-sensitivity detection of the mtDNA from a single cell, will heteroplasmic events still be seen? **Due to known events such as mitochondrial fusion/fission, mutational events, and oxidative stress, it is hypothesized that mtDNA heteroplasmy will be detectable within a single cell.** The expected outcome of the study will be a better understanding of the intracellular phenomenon of mitochondrial heteroplasmy, which could lead to a more structured protocol for identifying and categorizing heteroplasmy when encountered during mitochondrial DNA analysis.

Chapter 2 - Experimental Design

Sample Collection

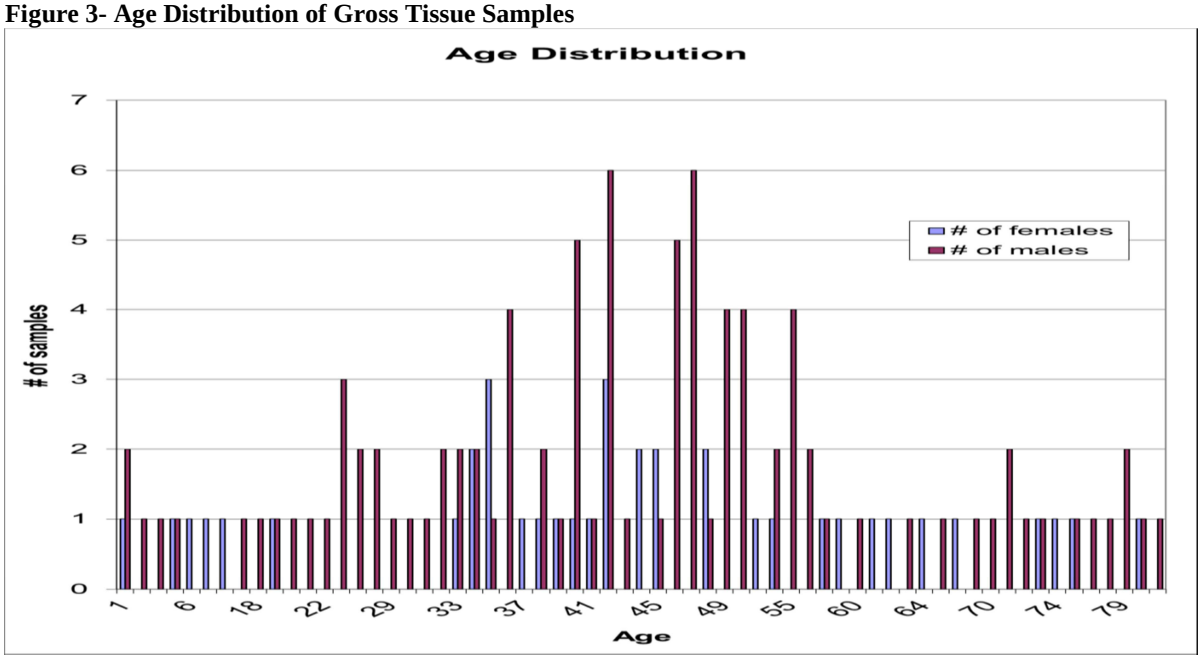
Post-mortem, formalin-fixed liver samples embedded in paraffin were collected from the Department of Histology at the Office of Chief Medical Examiner (OCME), New York City, in accordance with the OCME and New York City Institutional Review Board (IRB) guidelines. In order to properly screen for heteroplasmy based on known heteroplasmy rates, one-hundred and twenty-nine tissue samples were screened to find a pool of samples that contained heteroplasmy at one of five HVI hotspots.

The rationale for the use of liver samples was based on the stability of mtDNA in formalin fixed liver tissues. Previous studies have shown an excellent recovery of mtDNA from paraffin-embedded tissue samples, with the highest mtDNA recovery rate in liver samples (Miething et al., 2005). While there are other fixing techniques such as HumFix or methacarn that yield better mtDNA recovery results, these fixations are not typically performed on standard autopsy samples at the Office of Chief Medical Examiner in New York City (Uneyama et al., 2002). Therefore, the formalin-fixed samples were used. A secondary rationale for utilizing liver samples was that the tissue is constructed from different cell morphologies, and a large amount of interstitial space exists between liver cells. Therefore, cell separation can more readily be achieved. A tertiary rationale for the use of liver samples was the large number of mitochondria that are present in liver cells, which would aid in the analysis of a mitochondrial DNA profile from an individual cell.

Sample Selection

Samples were selected at random from a pool of paraffin tissue blocks designated as "waste tissue" by OCME-NYC and DOHMH-NYC guidelines. Samples were selected based on the noted presence of "liver" or "hepatic tissue" in the description on the package containing the tissue blocks. These tissue blocks were originally prepared from liver tissue cross-sections that had been submitted to the OCME histology laboratory following an internal autopsy examination. These liver tissue samples were further sectioned in the histology laboratory, placed into individual tissue cassettes, and finally fixed and embedded with paraffin. After the random selection of these tissue blocks was complete, the identifying information from each sample was used to cross-reference the autopsy records in order to note the age and sex of the tissue donor. Thereafter, the identifying OCME information was removed from the samples, and a random three-digit number was assigned to each sample in numerical order, selecting samples at random until all samples were accounted for. Based on the demographic information, the graph displayed in Figure 3 was prepared for all of the original tissue blocks used in the study.

Figure 3- Age Distribution of Gross Tissue Samples



The age ranges represented in the graph in Figure 3 are from 1 year of age to 83 years of age. Out of these 129 original samples, 35 are from females, the remaining 94 are from males. The average age of the samples is 43, with the median age of 42, demonstrating a Gaussian distribution of the samples across the age range.

Tissue Block Processing

Examples of these tissue blocks are seen in Figure 4 and Figure 5. Figure 4 is an example of how the tissue block was generally constructed, and what it looked like after it was sampled for the study. The larger fragment, which is missing the center section, is the liver sample. The smaller crescent-shaped fragment is a cutting of an artery from the same individual. The liver sample was identified by the marbling and mottled texture of the tissue. The section that was removed was then trimmed of excess paraffin and inserted into a water-filled Eppendorf™ tube and incubated in a water bath to remove the remaining paraffin in the tissue block. A similar block is seen in Figure 5, however the tissue is much darker and denser that was seen in Figure 4. This block was noted as containing liver and kidney. It was noted that the upper sample showed the two-tone undulations in the sample typical of renal tissue, while the bottom sample was more consistently dark and dense, which is typical for the tissue associated with the liver samples. The tissue that was identified as liver is the sample that was sectioned for testing in the study.

Figure 4- Tissue block, Sample 342

Figure 5- Tissue block, Sample 287

This process was repeated for each tissue block for each sample, and in the end, all of the isolated cells were verified by a pathologist to be uniform hepatic cells, which reinforced the aspect that the correct tissue source was identified and selected for each tissue block.

Cell Separation and Gross Tissue Generation

Each paraffin-embedded liver sample had a 1 cm cube of tissue removed from the sample, and this was heated in sterile water at 75 °C for 30 minutes to remove the paraffin from the sample. The tube was then cooled, and the paraffin ring from the top of the tube was removed and discarded. The sample was reheated in sterile water at 75 °C for another 1-2 sets of 30-minute incubations, until no more paraffin was released from the sample. The tissue was then sandwiched between the frosted ends of clean microscope slides, and mashed/ground to release the individual cells. The mashed tissue was placed into a cell-straining mesh basket (40 µm mesh) on top of a 50 ml conical tube, and washed with 1 ml of water. The filters containing the tissues underwent a centrifugal sieving technique at 1000 RPM for 5 minutes to collect the single cells in the liquid at the bottom of the 50 ml conical tube (Gebhardt, 2002). After passing through the 40 µm nylon mesh, the cell suspension was pelleted and transferred to 1.5 ml microcentrifuge tubes and stored at 3.7 °C. The tissue remnants trapped in the filter were also transferred to separate 1.5 ml microcentrifuge tubes and stored at 3.7 °C, as the gross tissue samples.

The rationale for separating the cells was to ensure that the same tissue source can be sampled for multiple laser-dissection isolations, without the need to re-cut the paraffin-embedded tissue block and risk missing a random heteroplasmic region in the tissue. Since the purpose of the research was to locate a heteroplasmic event in a single cell, the separation of the cells was a critical step.

Gross Tissue Screening for Heteroplasmy

To find tissue samples that contained heteroplasmy, the mtDNA in the gross tissue samples (from the tissue remnants in the filter baskets) were extracted by standard phenol/cholorform/isoamyl-alcohol organic extraction utilizing phase-lock gel from the Eppendorf Corporation (Westbury, NY). The extracted mtDNA was filtered and concentrated to 20 µl on Microcon YM100 filters from Millipore (Billerica, MA). The mtDNA in the gross tissue extracts was amplified using the HVI SNP multiplex primers (Appendix A) and *Taq* DNA polymerase, as per the standard guidelines (Anderson et al., 1981; Andrews et al., 1999; FBI, 2003; Isenberg & Moore, 1999). The amplification scheme for the SNP multiplex was as follows:

Soak at 95°C for 9 minutes
34 cycles: { 95°C for 10 seconds; 60°C for 30 seconds; 72°C for 30 seconds }
Storage soak at 4°C indefinitely

Post-amplification quantitation was performed on the Agilent 2100 Bioanalyzer Lab-on-a-Chip microfluidic fluorescence assay (Agilent Technologies, Santa Clara, CA). The primers and dNTP bases from the primary reaction were removed with ExoSAP-IT® (Affymetrix, Santa Clara, CA). The SNP assay was performed with the Applied Biosystems SNaPshot™ kits (Foster City, CA), using the SNP extension primers in a multiplex to screen the HVI region for heteroplasmy (Appendix A). The hypervariable region I was the focus of the study due to the well-characterized number of heteroplasmic hotspots in this region (Holland & Parsons, 1999).

The SNP testing methodology that was employed in this study is known as "extension-termination" typing (Li et al., 1999; Sobrino et al., 2005). This was accomplished by employing an extension primer to target the amplified region of HVI that contains the SNP base. The 3′ end

of the extension primer sits one base away from the SNP site in question, and the addition of a fluorescently tagged di-deoxynucleoside triphosphate terminator base to this extension primer identified the SNP base in question. Following post-amplification quantitation, the five HVI SNP locations for each sample were characterized using specific extension primers to target the HVI hotspot sites that were identified as having a high percentage of variability. The HVI SNPs were identified at the following positions on the mtDNA genome: 16069, 16223, 16324, 16390, and 16519 (Appendix B). Each of these positions were characterized in the literature as having a high percentage of variability based on sequencing analysis, which was derived from mtDNA population studies or the establishment of mtDNA databases (Baasner et al., 1998; Brandstatter, Parsons, Niederstatter et al., 2003; Brandstatter, Parsons, & Parson, 2003; Budowle & van Daal, 2008; Budowle, Wilson et al., 1999a; Finnila et al., 2001; Ingman & Gyllensten, 2003; Kline et al., 2005; Kong et al., 2003; Lutz et al., 1998; Maca-Meyer et al., 2003; Mishmar et al., 2003; Palanichamy et al., 2004; Stoneking, 2000; Tagliabracci et al., 2001; Thangaraj et al., 2005). The extension primers were also designed with differing lengths of poly-thymine tails at the 5′ end of the primer to ensure that all of the extension primers separated by at least four bp in length. The resulting amplicons of the extension primer amplification were then efficiently separated and detected using fluorescent detection capillary electrophoresis.

Electrophoresis and detection of the SNP bases was performed on an Applied Biosystems 3130*xl* multi-capillary electrophoresis array (Foster City, CA), in 96-well plates (Marino et al., 1996). The analysis of the SNP data was performed using SNP-typing presets and user-defined bins within the GeneMapper™ 3.0 software from Applied Biosystems (Foster City, CA).

Under electrophoresis, the extension primers separated based on charge and length, and the fluorescent tags of the ddNTP bases were read by the analyzer and translated into colored

peaks; green for adenine, yellow for cytosine (displayed as black), red for thymine, and blue for guanine. The SNP data for the gross tissue appeared as a set of five separate peaks in the electropherogram, with each peak corresponding to a specific extension primer, and therefore a specific HVI hotspot position. Mixtures of the colored peaks served as an indicator of heteroplasmy at the SNP hotspots. Since the standard (rCRS) base of each SNP was known, as well as the anticipated hotspot base change, the signal mixture that could be expected with instances of heteroplasmy would be known. For example, at position 16390, the standard base G is known to have a high frequency of transition with A. When targeted with the extension primer, the complementary SNP base to the standard G base would be a yellow signal peak for the C. The A base on the template would oppositely be identified with a red T signal. If such heteroplasmy existed in a sample, then there would be mixed signals resulting in a yellow-red peak mixture. This mixing, however, would be unequal and major/minor peak combinations would also occur, in which the colors of the major and minor peaks would be moderately indicative of the contribution of the mixture of the heteroplasmic bases in the sample. Fluorescent tags on each separate ddNTP would also modify the mobility of the SNP extension primer, resulting in a slight shift within the same bin for the peak of interest, allowing separation of two different colored peaks within the same bin to be resolved with the GeneMapper™ software when heteroplasmy was present for a SNP position. If the color was not mixed, and matched the color of either the standard rCRS base or the expected base-change for the HVI hotspot, it would be determined that heteroplasmy was not present at that particular SNP hotspot within the tissue. Using this process, the gross tissue samples were screened for heteroplasmy at the specific HVI SNP hotspots. The gross tissue samples that were determined to contain heteroplasmic hotspots at any SNP position were then selected for individual cell testing. A

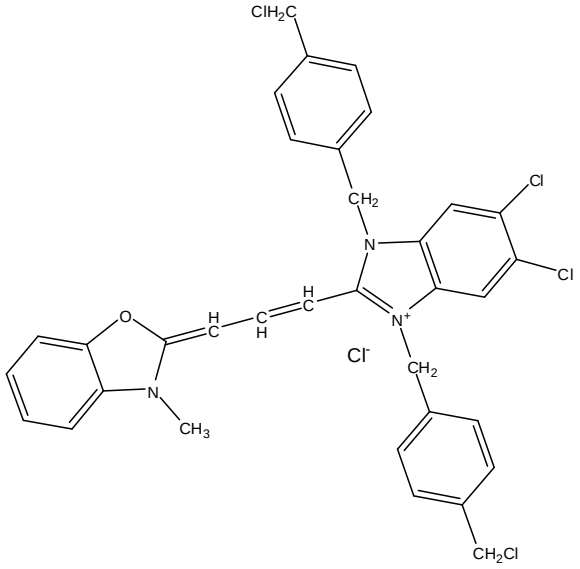
representative number of tissues that did not contain heteroplasmy at the SNP hotspots will be utilized as control samples for the individual cell testing assay. The number of control tissue samples was approximately 10% of the number of heteroplasmic tissues found in the gross tissue screening.

The rationale for screening the gross tissue samples for heteroplasmy using this methodology was that the detection of heteroplasmy would be done in a rapid and thorough manner for the five SNP hotspots in the HVI region. Organic extraction was necessary for the efficient and clean isolation of the mtDNA from the gross tissue samples. The rationale for probing the SNP sites using the extension/termination technique was that the SNP extension primers were designed to anneal to a target sequence that sits with the 3′ end of the extension primer one base away from the SNP site in question (Brandstatter, Parsons, & Parson, 2003; Li et al., 1999; Quintans et al., 2004; Vallone et al., 2004). The addition of the ddNTP terminator base to the extension primer resulted in an amplicon that was exactly one base longer than the original extension primer, with the new ddNTP base probing the nucleotide that was present in the template mtDNA, which was the SNP hotspot. Since the extension primers were all approximately 18-22 bp in length, a 5′ poly-thymine tail was added to four of the five extension primers to artificially create 4-bp differences in length among the extension primers. (Refer to the Appendix A for the extension primer sequences.) This aided in the electrophoretic separation of the extension primers when more than one heteroplasmic SNP was identified in the gross tissue sample. The length differences of the extension primers therefore allowed each SNP site to be uniquely resolved in the electropherogram for analysis. The addition of the poly-thymine bases was done as the extension primers were synthesized by the Integrated DNA Technologies, Inc. laboratory (Coralville, IA).

Cell-Staining

For all samples that were selected for further testing in the individual cell assay, the isolated cells were stained using MitoTracker® Green FM dye to selectively stain the mitochondria within the fixed cells. The freed-cell samples were stained with the MitoTracker® Green FM dye for 10-20 minutes in an aqueous solution that contained 10-200 nM of fluorescent probe (Invitrogen, 2006). MitoTracker® Green FM (Molecular Probes Inc., Eugene, OR) is a fluorescent dye that contains a thiol-reactive chloromethyl group that selectively stained the mitochondrial membrane (see Figure 6, below) (Buckman et al., 2001; Invitrogen, 2006; Poot et al., 1996; Salet & Moreno, 1990; Sutovsky et al., 1999).

Figure 6- MitoTracker® Green FM Dye



The cells were pelleted, washed with water, and fixed by evaporation and desiccation onto thermoplastic slides designed for the laser dissection microscope (Surgipath Medical Industries, Inc. Richmond, IL). The PALM microscope slide matrix is a synthetic polyethylene naphthalate (PEN) polymer that acts as a support for the cells and is capable of being cut with the PALM laser for micro-dissection of the samples (Carl Zeiss MicroImaging GmbH, Bernried,

Germany) (Zeiss, 2008). The fixed cells on the micro-dissection slides were then counterstained using Nuclear Fast Red to highlight the nuclei of the cells, which aids in the differentiation of a single intact cell from a cluster of cells (see Figure 7 & Figure 8) (Burton et al., 1998; Di Martino et al., 2004; Frank et al., 2007; Sams & Davies, 1967). The slides were then ready for fluorescent examination and laser dissection.

Figure 7- Cells at 40 X, visible light with Nuclear Fast Red

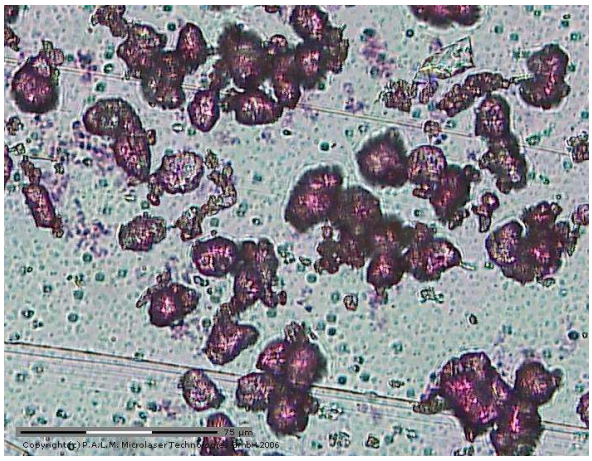
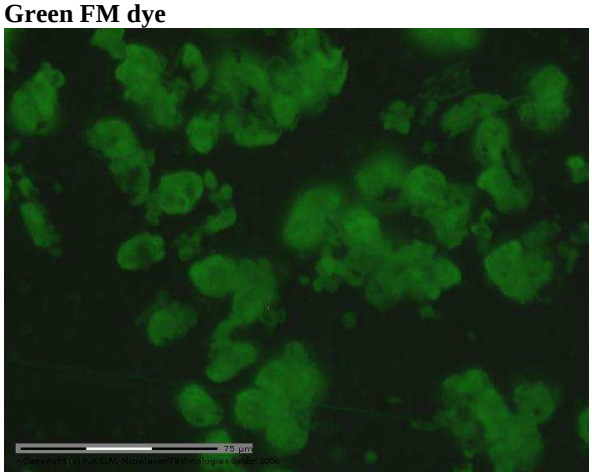


Figure 8- Cells at 40 X, FITC filter w/ MitoTracker® Green FM dye



The rationale for using the MitoTracker® Green FM dye was that it was the only fluorescent marker that highlighted mitochondria in formalin fixed tissues, and did so with low background and a high-emission signal that was readily detectable using the fluorescent imager on a microscope (Buckman et al., 2001; Invitrogen, 2006; Vassella et al., 1997). Since the fluorescent output of the dye was resonant to surrounding dye molecules, the clustering of the MitoTracker® Green FM molecules at the mitochondrial membrane led to a high level of green fluorescence, which indicated a large quantity of mitochondrial membranes within the cell, and therefore gave a visual indicator of a desired cell for isolation.

The MitoTracker® Green FM was therefore a specific dye for mitochondria, and because it was bound to the outside of the mitochondrial membrane, it would not affect the mtDNA after

extraction (Bestvater et al., 2002; Invitrogen, 2006; Kuznetsov et al., 1998). Other mitochondrial-specific dyes, such as Rhodamine 123, CalciumOrange-5N, and MitoTracker® Red, which all have a stronger binding affinity for the mitochondrial membrane, were not suitable for use in pre-fixed cells (Bolaños et al., 2008; Invitrogen, 2006; Minamikawa et al., 1999; Vassella et al., 1997). These other mitochondrial dyes also required active cellular respiration for the uptake and binding of the dye, and are therefore more appropriately used for studies in which tracking the movements of mitochondria or determining the fate of mitochondria in living cells is necessary (Bestvater et al., 2002; Bolaños et al., 2008; Buckman et al., 2001; Minamikawa et al., 1999; Vassella et al., 1997). The rationale for using the Nuclear Fast Red was that it helped differentiate intact cells from cellular debris, and allowed for the identification of a single cell apart from clusters of cells by identifying the nuclei.

Pathological Examination of Separated Cells

All of the slides were examined by a qualified pathologist to check for uniformity, and to confirm that the cells on the slide were hepatic cells, thereby confirming the use of liver tissue for the creation of the sample sets. The cells were also screened for any pathology that might eliminate the sample from further testing. Any samples that were identified as having questionable pathology had the original tissue blocks submitted for the preparation of cytological tissue cross-sections, stained with hematoxylin and eosin (HE) stain. The rationale for screening the cells for visual pathology was to eliminate any cells that might have been diseased (hepatitis, cirrhosis, etc.) or cancerous, as the former group would have contained non-functioning or dead mitochondria, and the latter group would have elevated mitochondrial levels to meet the demands of a rapidly metastasizing cell. The creation of histological slides from the tissue cross-

sections allowed confirmation of any diseased or cancerous tissues which flagged these problematic samples for removal from the study.

Microscopy and Laser Dissection

Once the mitochondria were stained using the dyes, the slides were examined and the cells were isolated using the PALM MicroBeam C laser dissection microscope. The PALM microscope is capable of high-level magnification of cells and cellular structures using the Carl Zeiss Axiovert 200 microscope, and with the proper fluorescent illuminator and filter, fluorescent image analysis of the labeled cells (Carl Zeiss MicroImaging GmbH, Bernried, Germany). Previous published studies into the effective isolation of cells and mitochondria focused on three different technologies: flow cytometry, laser-dissection microscopy, and optical tweezers. While it was seen that the optical tweezers were the most capable of isolating cells and individual mitochondrion, the laser-dissection microscopy was not ruled out as a viable instrument for the same purpose (Pflugradt et al., 2011). For the aspects of this project, optical tweezers were not available, and therefore not an option, and the laser-dissection microscope was deemed more than capable of isolating cells using the available magnification oculars. The cells were first visualized using the FITC fluorescent filter on the laser dissection microscope. The MitoTracker® Green FM dye was excited at 490 nm and the emission filter was set for detection at 516 nm. At 100 X to 200 X magnification, single cells were identified due to the green fluorescence within the cells that indicated the mitochondria had been selectively tagged with the MitoTracker® Green FM dye. This dye assisted in visualizing cells apart from vascular debris, glass chips, crystals, and ruptured cellular debris on the slide. Standard light microscopy (no fluorescence filter) was used to check whether the cell was intact and contained a nucleus, based on the staining of the Nuclear Fast Red dye. Also, the Nuclear Fast Red dye either showed

clusters of cells or overlapping cells that were not desirable for this project. Documentation of the selected cells was captured as a batch data file through the computer interface on the laser dissection microscope. The laser dissection of an individual cell was performed only when a single cell was identified on the slide.

The PALM incorporates a computer-assisted interface along with the microscope that allowed the user to select regions or sections of the sample that were cut out with a robotically-controlled chemical laser. In this process, the cell was outlined by the user via the computerized interface, the laser cut along this outline, and then "punched" the cut section into a micro-centrifuge tube cap above the stage, in a technique known as "catapulting" (Zeiss, 2008). Using the interface software on the laser dissection microscope, an oval region was established surrounding the selected individual cell, approximately 100-150 µm at the longest diameter, and the laser cut the microscope slide membrane along this indicated circular pathway. The cut membrane, now containing the single cell sample, was then launched off the slide into 15 µl of capture buffer (83.4% dH_2O, 16.6% bovine serum albumin (1.6 µg/µl)) contained in a sterile collection cap (Hunt & Finkelstein, 2004). This process was repeated a total of 50 times per slide, in order to collect up to 50 separate cells from a single tissue. The collection tubes were then sealed and labeled according to the random sample number and cell number, based on the order in which the samples were obtained. The slides were retained for future work.

In the process of performing validation studies on the laser dissection microscope, it was noted that there were a wide range of possible sources of error that could occur due to either the manual use of the instrument, or fluctuations in the robotic arm controlling the collection tube. Several protocols were developed to address these possible sources of error and minimize the effect that these sources of error would have on the collection of individual cells. The first

potential source of error that was noted was that the microscope slides had a wide degree of variability in the position of the slide when inserted into the spring-mounted holder on the robotic stage of the laser dissection microscope. Since the cells on the slide were being mapped for the purpose of returning to the slide over consecutive sessions to collect the appropriate number of cells, it was necessary for the digital map to overlay the same cells each time the slide was inserted. This exact overlay was impossible if the slide positioned differently each time. The digital map was created by tagging the X/Y/Z coordinates of each selected cell, and this position was locked relative to the objective of the microscope, not the stage. Therefore, movements in the slide and/or stage would shift the physical coordinates of the cells. However, it was possible to re-position the digital map as a whole if a common reference point could be established on each slide. To create this common reference point, a rectangle was cut into the slide matrix outside of the area containing the cells. This rectangle served as the reference point, and had a digital overlay of this rectangle present in the digital map. By inserting the slide is the same relative position (frosted/labeled end of slide toward the ocular of the microscope) and ensuring the slide was flush against the right side of the slide holder, the slide would be aligned in the same X/Y plane as before, and then only the digital map had to be moved to overlay the digital image of the reference rectangle with the physical rectangle cut. When this was done correctly, it was seen that the rest of the map would overlay the appropriate cells on the slide, and therefore allow rapid and accurate re-acquisition of the previously-mapped cells for the study.

A second potential source of error in the laser dissection microscope was the use of the collection cap and the fluctuations in the height of the cap above the stage. While the use of the reference rectangle re-aligned the cells, it only did so in the X/Y plane, and variability in the Z

plane meant that the collection cap height needed to be manually controlled for each collection. While this variable was capable of being manually controlled in the software, there was seen to be a wide range of variability in the cap height due to the flexibility of the metal arm that held the collection cap. This metal arm was designed to be flexible to ensure that if the cap came into hard contact with the slide, the arm would absorb the impact, and not break the slide as a result. However, the contact with the edge of the collection tube cap and the slide would result in the physical transfer of multiple cells, debris, or other tissue detritus to the cap, which would be unacceptable for the study. Therefore, a method was devised to visually ensure that the cap edge was not in physical contact with the slide by manually controlling the Z coordinate of the robotic stage. Confirmation of the appropriate cap height would also be seen if the digital image of the slide on the computer screen did not distort, meaning that the slide surface was not under pressure and being forced out of the plane of focus. If a cap did come into contact with the slide, it was discarded and a new cap was inserted to perform the next collection.

A third point of possible error in the laser dissection microscope was the use of the water/BSA buffer to ensconce the interior of the collection cap. The addition of the buffer allowed the liquid to better adhere to the hydrophobic interior of the cap rather than just using deionized water, which tended to bead up within the cap without the addition of the BSA. The use of the buffer caused a concave meniscus to form due to the capillary action of the liquid up the sides of the collection cap. When the liquid-filled cap was then inverted and inserted into the collection arm, this capillary action was further affected by gravity. When the cap was brought within a certain distance of the slide, even with visual confirmation that the cap was not physically touching the slide, the attraction between the hygroscopic slide membrane and the water would draw the liquid from the cap and cause it to pool on the slide surface. When this

occurred, it was impossible to know if additional, non-specified cells were taken up by osmosis or reverse capillary action back into the cap. At the same time, it was not recommended to attempt laser-dissection microscopy on a moist slide surface, as the refractive index of water droplets would disrupt the focus of the chemical laser and affect the ability to properly cut the slide membrane. When this error was seen to occur, the slide was removed and re-desiccated, and the cap was discarded. To ensure that this error did not occur, the Z-axis control was set to move slowly and under manual control to again provide visual conformation of the appropriate cap height above the slide. This proper height was again confirmed by the digital image of the slide surface, which would display any moisture on the slide surface as refractive spots, or droplets.

This methodology ensured that a solitary cell was isolated using the laser-dissection technique. The combination of both the fluorescence and light microscopy provided an accurate and reliable indication that a single cell was being identified and captured, and this allowed for the isolation of at least 50 individual cells per slide. If the cells were found to cluster or group on the slide, then such cells were ignored for the purpose of the study. The use of a capture buffer allowed the low volume of liquid to completely adhere to the collection cap and maximized the surface area available to catch the catapulted single cell.

Cell Lysis and HVI SNP Amplification

The individually captured cells were then amplified using a one-step PCR assay. The one-step amplification setup consisted of the HVI SNP region primers specific for the heteroplasmic SNP position seen in the gross tissue. AmpliTaq Gold® polymerase (Applied Biosystems, Foster City, CA) and additional PCR reagents were added in a final amplification volume of 25 µl for a set number of cycles based on the SNP primers used in the amplification

(Coble et al., 2004; Parsons & Coble, 2001; Vallone et al., 2004; Xiu-Cheng Fan et al., 2008). (Refer to Appendices A and B for primer sequences and the HVI SNP map). The temperature-cycling scheme for the one-step SNP amplification was as follows:

Soak at 95°C for 9 minutes
28-34 cycles: { 95°C for 10 seconds; 60°C for 30 seconds; 72°C for 30 seconds }
Storage soak at 4°C indefinitely

Using this setup, the individual cells were thermally lysed by performing standard Hot-Start PCR. This process relied on the 94°C, 9 minute cycle at the beginning of the amplification to lyse the cell and release the mtDNA, which was then followed by standard PCR amplification. Samples were run against both positive and negative controls from the one-step amplification step onward, to ensure quality and accuracy of the final data (Levin, Holland et al., 2003).

The amount of amplified mtDNA in each sample was quantitated using the Agilent 2100 Bioanalyzer Lab-on-a-Chip microfluidic fluorescence assay (Agilent Technologies, Santa Clara, CA). If the Agilent 2100 Bioanalyzer sizing data showed the amplified peaks at the expected bp sizes for the SNP amplicons but below the detection threshold (0.1 ng), the sample could be amplified for an additional 16 rounds of amplification for a total of 50 cycles of PCR amplification to establish an above-threshold signal (Barreto et al., 1996; Barros et al., 1997). This was established on a per-sample basis, since a second round of PCR amplification would be done on the samples, utilizing the amplified SNP region as the template DNA for the SNaPshot™ PCR amplification.

The rationale for performing a “whole-cell” lysis and one-step amplification was to eliminate the possible loss of the mtDNA that would possibly occur through a standard extraction/purification attempt on a single cell. Similar techniques are present in the literature

dealing with laser-dissected tissues and whole-cell amplification for both single-nucleotide polymorphism (SNP) and short-tandem repeat (STR) analysis following either an alkaline or Proteinase-K lysis (Rook et al., 2004). The potential pitfall in the one-step PCR amplification of the mtDNA from an individual cell was the possibility of a low-level of template mtDNA for the PCR process. If this was found to be the case following the post-amplification quantitation of the amplified samples, the samples could be re-amplified using 16 more cycles of amplification. To control for this possible additional amplification, two negative controls were established at the beginning of the one-step PCR amplification process; the amplification negative control (AN) and a cycling amplification negative control (CAN). The AN control, along with the HL60 positive control established the quality for the primary one-step amplification of the single cells, and the CAN control was only to be applied if the cells required additional amplification cycles, to control for the possible stochastic effects produced during additional rounds of amplification on the single cells. Precedent for this control set can be seen in published validations of other SNP-plex assays in which it was noted that contamination and primer quality issues are problematic in SNP typing at low-DNA input levels (Musgrave-Brown et al., 2007). However, the additional amplification cycles might not be necessary on low-yielding samples, since the purpose of the primary one-step amplification was to produce template DNA for the final SNaPshot™ assay in a concentration range of 0.2-0.4 pmol of amplified DNA.

SNP Extension/Termination Probing

The amplified single-cell mtDNA was first purified using an Exonuclease I / shrimp alkaline phosphatase (ExoSAP-IT®) cleanup, which removed the unincorporated dNTP bases from the first round of the one-step PCR process, as well as dephosphorylated the original SNP primers. The specific hypervariable SNP for the region previously identified with heteroplasmy

was then probed with the specific SNP extension primer and the SNP site was targeted with fluorescently tagged di-deoxynucleoside triphosphate (ddNTP) bases. (Refer to the Appendix A for the primer sequences for each extension primer.) The extension primers were added to the purified mtDNA amplicons along with the SNaPshot™ kit reagents that contained the labeled ddNTP bases, polymerase, and necessary buffers for the SNaPshot™ PCR assay (Applied Biosystems, Foster City, CA). The amplification scheme for the SNaPshot™ assay was as follows:

$$\text{25 cycles: } \begin{cases} 96°C \text{ for 10 seconds} \\ 50°C \text{ for 5 seconds} \\ 60°C \text{ for 30 seconds} \end{cases}$$

Storage soak at 4°C indefinitely

Following SNaPshot™ amplification, the samples were treated with 1.5 units of shrimp alkaline phosphatase (SAP) (Promega Corporation, Madison, WI) to dephosphorylate the unused extension primers prior to capillary electrophoresis.

The rationale for probing the SNP sites using the extension/termination technique was that the SNP extension primers are designed to anneal to a target sequence so that the 3′ end of the extension primer sits one base away from the SNP site in question (Brandstatter, Parsons, & Parson, 2003; Li et al., 1999; Podini & Vallone, 2009; Quintans et al., 2004; Vallone et al., 2004). The addition of the ddNTP terminator base to the extension primer resulted in an amplicon that was exactly one base longer than the original extension primer, with the new ddNTP base probing the SNP hotspot nucleotide that was present in the template mtDNA. Since the extension primers were all approximately 18-22 bp in length, a 5′ poly-thymine tail was added to four of the five extension primers to artificially create 4-bp differences in length among each extension primer. (Refer to the Appendix A for the extension primer sequences.) This aided in the electrophoretic separation of the extension primers when more than one

heteroplasmic SNP was simultaneously being detected in the individual cells. The length differences of the extension primers therefore allowed each SNP site to be uniquely resolved in the electropherogram for analysis.

Capillary Electrophoresis and Fluorescent Detection

The SAP-purified extension primers were mixed with a fluorescent size standard (Genescan-120 LIZ) and deionized formamide within a 96-well plate (Applied Biosystems, Foster City, CA). Detection of the amplified extension primers was performed on a 3130*xl* multi-capillary electrophoresis array (Applied Biosystems, Foster City, CA), in 96-well plates utilizing a 7% performance-optimized polymer (POP-7™), a 36 cm capillary array, and the E5 chemistry detection and analysis protocol. Under electrophoresis, the extension primers separated based on charge and length, and the fluorescent tags of the ddNTP bases were read by the analyzer and translated into colored peaks; green for adenine, yellow for cytosine (displayed as black), red for thymine, and blue for guanine. Mixtures of the colored peaks served as an indicator of heteroplasmy within the SNP hotspot regions. Since the standard (rCRS) base of each SNP was known, as well as the heteroplasmic base change based on the gross tissue analysis, the anticipated signal mixture for the heteroplasmic samples was known. For example, at position 16390, the standard base G is known to have a high frequency of transition with A. When targeted with the extension primer, the complementary SNP base to the standard G base would be a yellow signal peak for the C. The A base on the template would oppositely be identified with a red T signal. If such heteroplasmy existed in a sample, then there would be mixed signals resulting in a yellow-red peak mixture. This mixing, however, might be unequal and major/minor peak combinations might also occur, in which the colors of the major and minor peaks would be moderately indicative of the contribution of the mixture of the heteroplasmic

bases in the sample. Fluorescent tags on each separate ddNTP also had the propensity to modify the size of the SNP extension primer, resulting in a slight shift within the same bin for the peak of interest. This difference allowed for the separation of two differently colored peaks within the same bin to be resolved with the GeneMapper™ software (Applied Biosystems, Foster City, CA) when heteroplasmy was present for a SNP position. If multiple peaks were not present, and the resulting single peak matched that of either the standard rCRS base or the expected base-change for the HVI hotspot, it was determined that heteroplasmy was not present at that particular SNP hotspot within the cell.

Data Analysis

The data for each of the individual cell samples was presented as the standard mitotype based on the SNP hotspots that were probed with the extension primers, which were positions 16069, 16223, 16324, 16390, and 16519. The nucleotide base at each of these positions was displayed as the fluorescent DNA base that was detected at that position.

The gross tissue samples were also compared to measure the relative amount of heteroplasmy against the age of the individuals, to investigate if heteroplasmic events increased with age. This was determined not only for the samples that were found to contain heteroplasmy, but also for the samples that did not contain any heteroplasmic events and were not used past the SNP screening step for the study. It was theorized that fewer heteroplasmic events would occur at a younger age, which would support the theory of heteroplasmy manifestation due to oxidative damage or mutational events within the mitochondrial genomes. A similar correlation was also performed examining the relative amount of heteroplasmy against the gender of the individuals as well.

The rationale for presenting the data as the SNP profiles from each sample served to correlate the mtDNA profile derived by SNP typing the gross tissue with the data from the individual cells from the same sample. If heteroplasmy was detected in the gross tissue sample and was uniformly confirmed in the individual cells, it would be assumed that the cells reflected the overall mitotype of the gross tissue, and therefore there existed an equal partitioning of the cellular mtDNA genomes. If heteroplasmy was detected in the gross tissue but was not uniformly confirmed in the individual cells (some cells contained the heteroplasmy, some did not), then it would be assumed that the individual cells did not reflect the overall mitotype of the gross tissue and there was an unequal partitioning of the cellular mtDNA genomes. If the testing of the gross tissue showed that no heteroplasmy was present but individual cells from the tissue showed the presence of heteroplasmy, it would be assumed that the heteroplasmy seen in the cells was minor and did not reflect the overall mtDNA profile of the gross tissue, such that a dominant homoplastic mtDNA profile in the tissue was overwhelming the minor cellular heteroplasmy, and again there was an unequal partitioning of the cellular mtDNA genomes.

A potential pitfall at this stage would be if a different and/or additional heteroplasmic base, other than the one originally found in the gross tissue sample, appeared in the individual cells. This was possible since the individual cellular mtDNA SNP profile might be minor compared to the dominant tissue profile, but capable of being detected in the individual cell when not overwhelmed by the dominant gross tissue mtDNA profile. In another scenario, the testing of the 50 cells could uncover 50 different mitotypes from a tissue source, showing that a dominant mitotype was different for each separate cell. If the individual cells contained unique heteroplasmic positions when compared to other cells from the same tissue, verification of the

individual mitotypes would be done by re-examining the gross tissue data for an accumulation of all of the combined heteroplasmic sites.

Experimental Objectives

The specific objective of this study was to gain a better understanding of how mtDNA heteroplasmy manifests within a single cell, and if different mtDNA genomes were physically present within a single cell. Since it has been previously established that heteroplasmy exists between tissues from the same individual, and therefore between cells, the question was then formed: Is heteroplasmy present within a single cell?

To answer this question, 50 individual cells were isolated using laser dissection from post-mortem liver tissue containing known mtDNA heteroplasmy. If the isolation, acquisition, and detection of a heteroplasmic event from a single cell was achieved, it would then indicate that the cell hosted a mixture of mtDNA genomes. The distribution of these mtDNA genomes within the cell was then the next logical question, but outside the boundaries of the technology that was currently employed with this research.

A possibility that was considered in the approach to this project was that one cell might display a homoplastic mtDNA profile, but another cell from the same tissue might show a heteroplasmic mixture of mtDNA genomes based on the SNP testing. This scenario addressed the fundamental question of the research; the finding of even one cell that contained a heteroplasmic mix of mtDNA genomes would answer the question as to whether or not a single cell could host multiple mtDNA genomes. The possible difference of heteroplasmy between the cells would be a notable aspect of the research, and further illustrate the differentiation that might be occurring between cells within a single tissue sample. However, redundancy of the data

would be necessary to show significance in the samples, and a single instance of a solitary cell containing heteroplasmy would possibly require re-testing for verification.

A secondary consideration was that heteroplasmy from one cell might be different than the heteroplasmy from a second cell. If this occurred, it would be expected that the individual heteroplasmic instances would be confirmed within the set of fifty cells from the gross tissue sample from which the single cells originated. Also, a tertiary consideration was the "appearance" of a new heteroplasmic site in the single cells, uncovered during testing, which was not originally seen in the heteroplasmy of the gross tissue. Each of the above scenarios was a possibility, and was taken into consideration in the analysis of the data. An important counterpoint to address was the reverse consideration that heteroplasmy could be detected at the cellular level, but if the percentage of such cells possessing the heteroplasmy was small, the heteroplasmy may not have been primarily detected in the gross tissue. For this reason, a representative number of gross tissue samples that did not contain heteroplasmy were also screened in the individual cell SNP testing assay.

A fundamental paradox that also was addressed was that the multiple strategies that were used to examine heteroplasmy might, in turn, have been producing it as well. Heteroplasmic events, especially length heteroplasmy, are possibly linked to nucleotide mis-incorporation due to low- to moderate-fidelity polymerase, and if such a polymerase was used, there was no way to know if the heteroplasmic events seen in the initial screening of the gross tissue samples related to SNP analysis of the single cells in the same regard. A major aspect of this project was to relate the relevance of heteroplasmy within the parameters of forensic settings; therefore it was practical to utilize the same techniques for screening that would be used in forensics. This included the use of the moderate-fidelity thermostable *Taq* polymerase, and the standard mis-

incorporation rate that related to the *Taq* polymerase activity therefore was considered. The testing also focused solely on the HVI region of the mtDNA, and did not focus on any of the known coding regions within the remainder of the genome. HVI, along with HVII, are the common mtDNA regions tested in forensic applications. Finally, the SNP typing of the samples was designed to avoid the HVI C-stretch, which is a common problem addressed in forensic mtDNA analysis. As such, within this study, the issue of length heteroplasmy was not addressed in the screening of the cell's HVI hotspots, either at the screening of the gross tissue samples or the SNP testing of the individual cells.

Another issue considered was the age of the individual at the time of death, to investigate if there was a trend in the presence of heteroplasmic events. The age of the individual contributing the tissue was compared against the SNP testing results of the samples to investigate if there was an increase in heteroplasmic events with age. However, since each sampled cell was only compared against the other cells from the same tissue, the age-relationship of the heteroplasmic events from a single tissue source were null. Only the gross tissue samples were compared to correlate the age of the donors to investigate if the heteroplasmy rate was correlated to age. The gender of the samples was also investigated to see if there was any trend relating heteroplasmy to the sex of the individual, but again, the comparison was only done between the gross tissue samples and not between the cells of a single sample.

The data collected in this study was reported as SNP mitotypes: i.e. as the actual base determined by the fluorescent signal output at each of the SNP locations probed in the extension typing. These differences were reported for the SNP site tested within each cell, with the data collected from each tissue source grouped into the regular sets of fifty, based on the number of cells that were collected using the laser-dissection microscopy. Since the heteroplasmic events

in the gross tissue were verified with SNP-testing data from the single cells, both the original gross tissue data and the final individual cell data were presented in form of tables that listed the data for each sample (peak height, peak bp size) as well as the SNP profile for the sample.

Chapter 3 - Experimental Results

Development of a Protocol for Cell Separation/Isolation and Generation of the Gross Tissue Samples

In order to optimize the recovery of intact cells from the paraffin tissue blocks, several different methods were investigated to remove the paraffin, isolate intact cells, and generate usable tissue remnants for the gross tissue. Manual removal of the paraffin was attempted, but was unsuccessful due to interstitial invasion of the paraffin during histological fixation. Grinding and mashing was also attempted on paraffin-embedded test tissues, but no intact cells were observed upon microscopic analysis of the remnants. With the application of heat and repeated wash steps in water, however, the paraffin was capable of melting off the tissue and solidifying at the top of the liquid upon cooling. The paraffin remnants could be easily removed from the top of the tube, leaving pliable, intact tissue behind.

In order to effectively separate the cells from the tissue, several variations of mortars and pestles were used for various lengths of time and with varying pressure. None of these processes resulted in a large yield of intact cells, and would more often result in cell debris and cellular fragments upon microscopic examination. It was seen that the grinding action was effective in removing the cells from the tissue, but the conical environment of the mortar-pestle combination allowed no space for free cells to escape once separated from the tissue block. To remedy this, the pliable tissue blocks were placed between the frosted ends of sterile microscope slides and "mashed" together. This was seen to free individual cells from the tissue, and at the same time it provided the necessary room for the cells to laterally migrate between the planes of the microscope slides while remaining intact.

A variety of cell-strainers and filters were then investigated, and it was seen that a 40 µm mesh filter was the best at allowing intact cells to pass through the filter into a supernatant while retaining the remaining tissue on the filter for use as the gross tissue sample. Using this methodology, individual cells and a gross tissue sample could be generated from a paraffin-embedded tissue in under two hours.

Cell Separation and Gross Tissue Creation

The mashing process that was previously described in the experimental design section resulted in the successful separation of individual cells for the purpose of the study. However, it was seen that various levels of cell separation were present on the slides, based on the maceration of the tissue sample. These various levels were defined as follows:

- Less macerated: indicating large colonies of intact tissue
- More macerated: indicating less presence of intact tissue and more single cells
- Sparse: indicating virtually no intact tissue and a low density of cells across the slide surface.

In order to visualize these various levels of cell separation, a cross-section of the physical tissue block was compared against the separated cells from the same tissue. As seen in Figure 9, the tissue from sample 349 stained with HE stain, shows typical liver morphology. Figure 10 is an example of the “less macerated” cells, stained with Nuclear Fast Red (NFR) dye that resulted from the tissue mashing of sample 349. As it can be seen in the image of the cells, there are larger chunks of intact tissue, as well as individual cells throughout the slide.

Figure 9- Sample 349, Tissue block, HE Stain, 100 X magnification

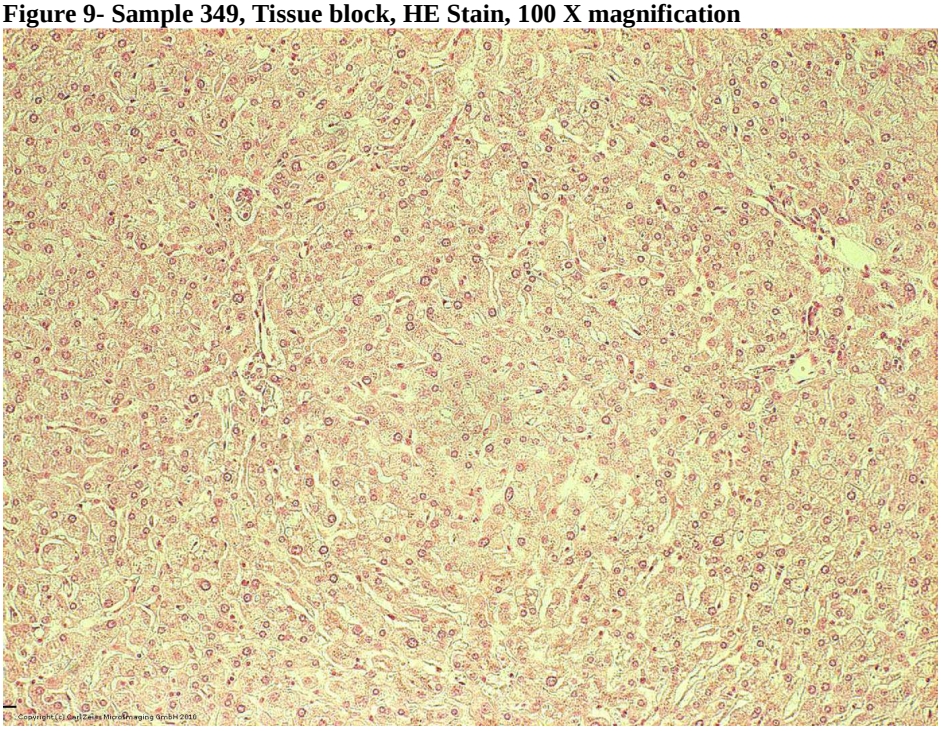
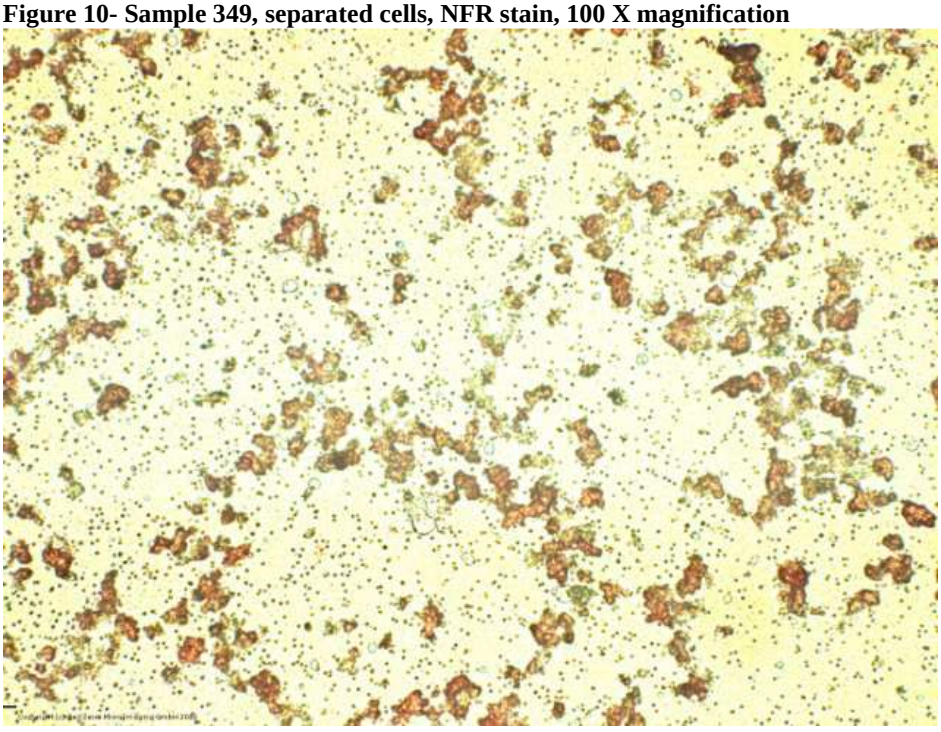


Figure 10- Sample 349, separated cells, NFR stain, 100 X magnification



In another example, the cells for sample 342 were classified as falling between “less macerated” and “more macerated”. The intact and HE-stained tissue for sample 342, seen in

Figure 11 shows the typical liver cell morphology, while Figure 12 is the image of the NFR stained and separated cells from the tissue block of sample 342. While there are large sections of intact tissue visible, there are also individual cells present, but in less density across the slide than seen in sample 349 (Figure 10, above).

Figure 11- Sample 342, Tissue block, HE Stain, 100 X magnification

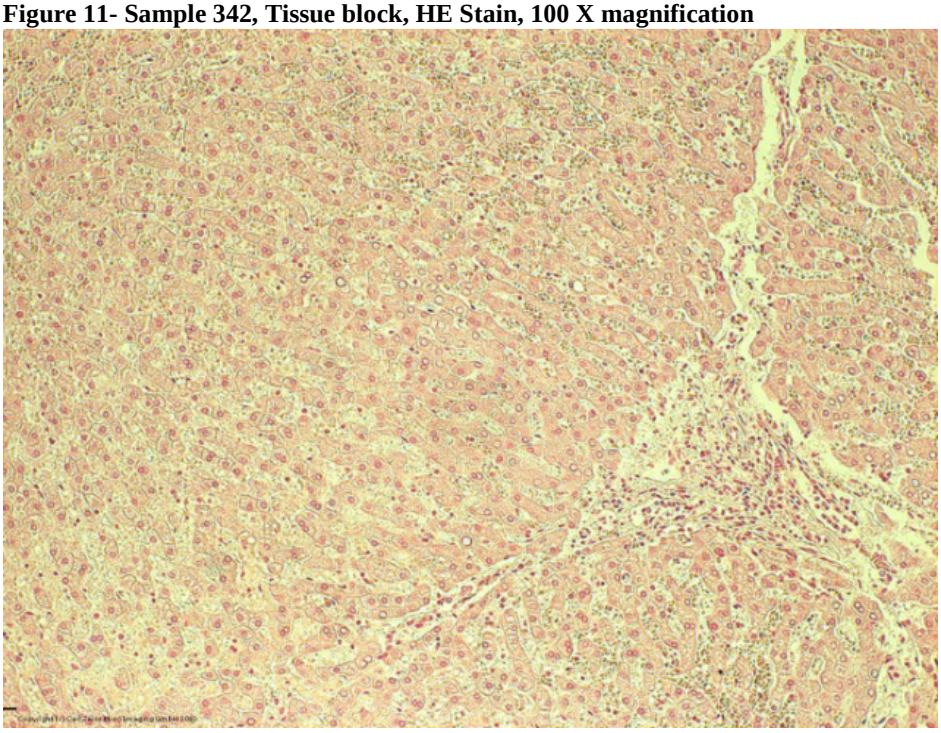
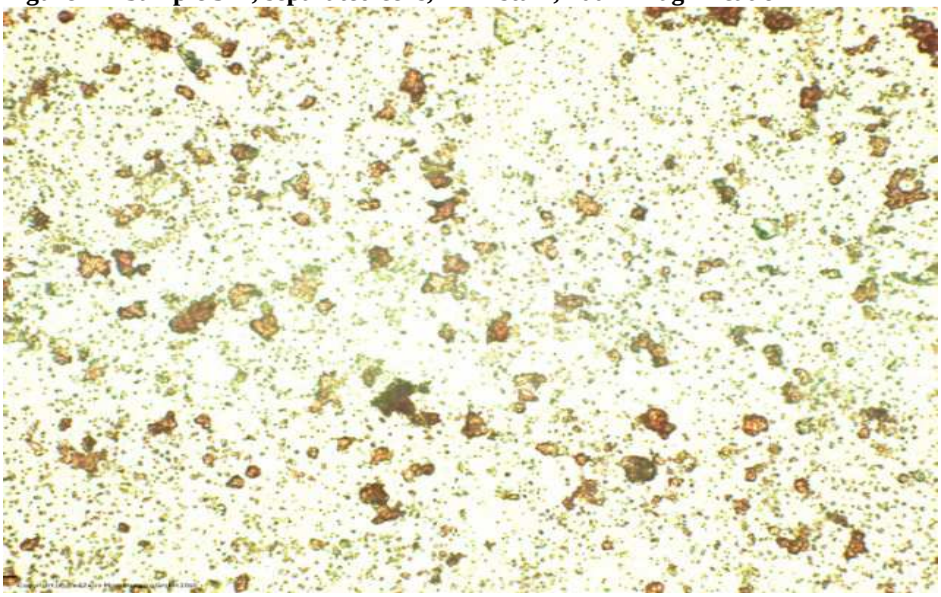
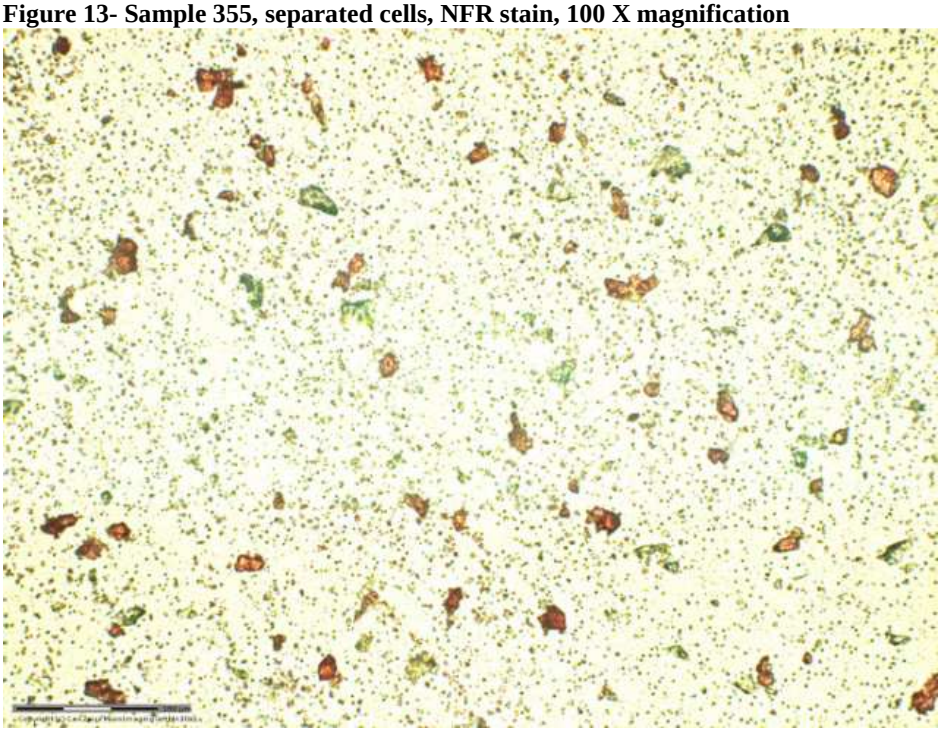


Figure 12- Sample 342, separated cells, NFR stain, 100 X magnification



In another example, the cells from sample 355 were seen to be “more macerated.” There was no intact tissue block remaining for sample 355 to show the intact tissue morphology, however Figure 13 is the image of the separated and NFR stained cells from the tissue block of sample 355. It can be seen in Figure 13 that there are again smaller clusters of tissue, as well as single cells, all of which are again dispersed less densely than was seen in the slide images for samples 349 and 342 (Figure 11 and Figure 12, respectively above).

Figure 13- Sample 355, separated cells, NFR stain, 100 X magnification



For a final example, the cells from sample 261 were classified as “sparse” when the slide was reviewed. The HE-stained intact tissue for sample 261, seen in Figure 14 shows typical liver cell morphology, while Figure 15 is the image of the NFR stained and separated cells from the tissue block of sample 261. It can be seen in Figure 15 that no clusters of cells are visible, and the cells that are present are single, intact cells populating a very low-density field across the slide surface.

Figure 14- Sample 261, Tissue block, HE Stain, 100 X magnification

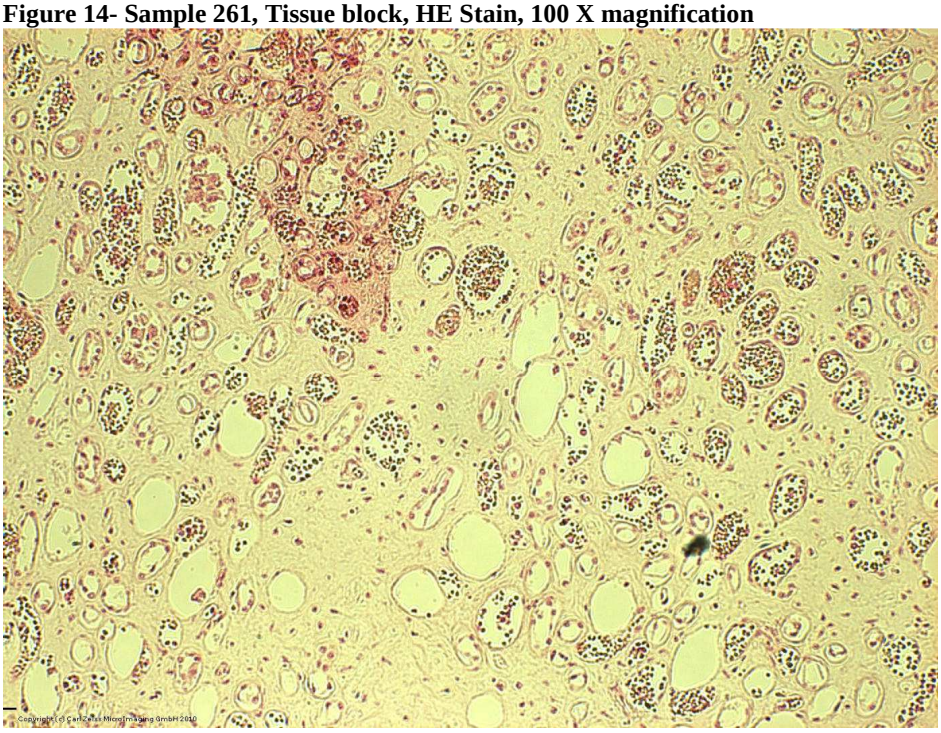
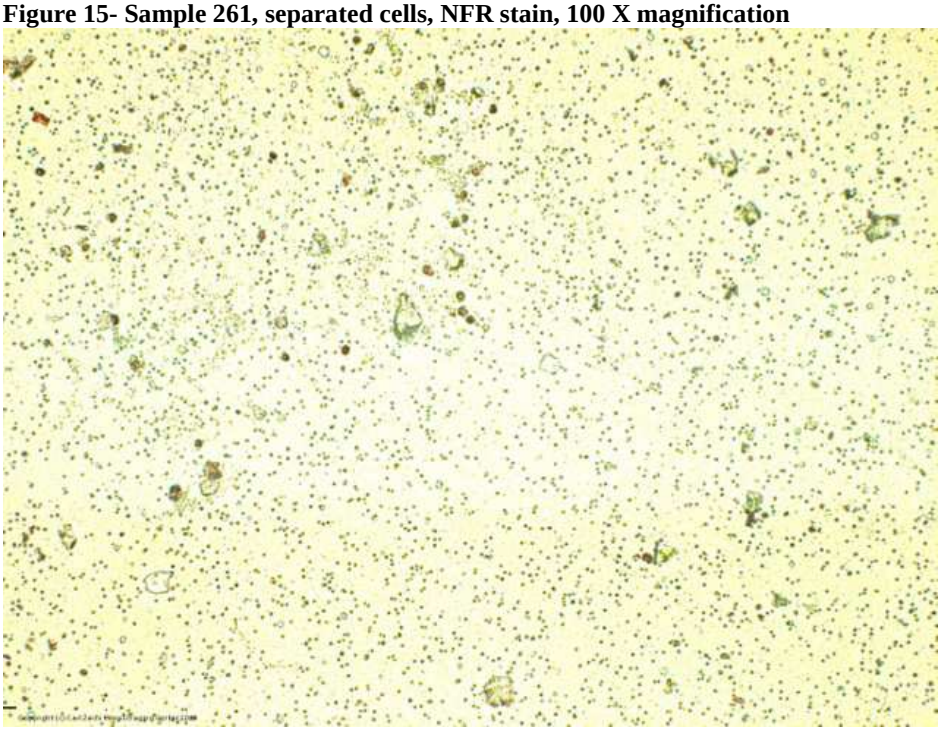


Figure 15- Sample 261, separated cells, NFR stain, 100 X magnification



HVI SNP Assay Development and Validation

Following the completion of the protocol design for capturing individual cells, work began on developing an HVI SNP assay that would be used to screen both the gross tissue and the individual cells for mtDNA heteroplasmy. The first step in this process was to identify the hotspot positions in the mtDNA HVI region. The rationale here was that point heteroplasmy was most often seen at these hotspot locations, so SNP screening would be the best way to uncover heteroplasmy within a cell. Several publications and online databases list known HVI hotspots, and while nearly every single location within HVI has some published instance of being a "hotspot," there were several locations that showed a much higher percentage of base-change prevalence. It was determined that the SNP assay would have to target these regions, and that while not all could be tested, specific locations could be identified that were evenly spaced out within the HVI region. Based on the publications and databases, the following locations were initially chosen:

16069, 16189, 16223, 16324, 16366, 16390, 16519

(Achilli et al., 2005; Achilli et al., 2004; Coble et al., 2004; Finnila et al., 2001; Ingman & Gyllensten, 2003; Ingman et al., 2000; Kong et al., 2003; Maca-Meyer et al., 2003; Mishmar et al., 2003; Palanichamy et al., 2004; Parsons & Coble, 2001; Thangaraj et al., 2005; Xiu-Cheng Fan et al., 2008)

The hotspot for 16069 was reported as a transition of C-T bases, occurring in a maximum frequency of 43 out of 241 samples, or 17.84% of the time, to a minimum frequency of 17 out of 192 samples, or 8.85% of the time (Coble et al., 2004; Finnila et al., 2001).

The hotspot for 16223 was reported as a transition of C-T bases, occurring in a maximum frequency of 50 out of 80 samples, or 62.5% of the time, to a minimum frequency of 6 out of 75 samples, or 8% of the time (Ingman et al., 2000; Palanichamy et al., 2004).

The hotspot for 16324 was reported as a T-C transition by Parsons, et al, but no frequency data was published in the various databases. However, this SNP site was evenly spaced in the HVI region among the other hotspots, so it was chosen for the SNP plex (Parsons & Coble, 2001).

The hotspot for 16390 was reported as a transition of G-A bases, occurring in a maximum frequency of 6 out of 48 samples, or 12.5% of the time, to a minimum frequency of 4 out of 44 samples, or 9.09% of the time (Maca-Meyer et al., 2003; Mishmar et al., 2003).

The hotspot for 16519 was reported as a transition of T-C bases, occurring in maximum frequency of 171 of 241 samples, or 70.95% of the time, to a minimum frequency of 15 out of 29 samples, or 39.46% of the time (Achilli et al., 2004; Coble et al., 2004). It should be noted that while 16519 technically falls outside of the defined HVI region from 16069 to 16324, it is still technically within the major non-coding hypervariable region.

The SNP sites 16189 and 16366 were eventually dropped from the SNP assay. The 16189 position is part of the HVI C-stretch; the hypervariable stretch of cystine bases that get interrupted at 16189 by a thymine base. Since the transition of this thymine to a cystine is the root cause for length heteroplasmy in the cell, and length heteroplasmy is a very common occurrence in mtDNA testing due to what is commonly seen as an “all-or-none” T-C change, this position was dropped since it was not informative. The 16366 position was abandoned after it was seen that the region between 16300 and 16400 was too crowded with SNP primers, and the

16366 primers were seen to be problematic based on the thermodynamic calculations for dimerization and secondary-structure formation.

For the remaining five SNP locations, it was determined that the hotspots would be probed using a SNP amplification/extension assay. This would be done by amplifying a target region around the SNP base of interest using forward and reverse primers, and then targeting the SNP position using an extension primer in which the 3′ end sits one base away from the SNP base in question. The addition of a labeled di-deoxynucleoside triphosphate terminator base at the 3′ end of the extension primer would then be analyzed to identify the SNP base. Since the SNP assay was designed to target five SNP regions, and the extension bases would then be used to analyze the SNP from each region, a total of fifteen primers needed to be designed, such that each SNP location was targeted with a forward, reverse, and extension primer. The rCRS was used as a template for primer design, and various forward and reverse primers and extension primers were developed based on published SNP assays or by using the rCRS sequence to target the five SNP locations (see Appendix B) (Vallone et al., 2004).

As noted in the Appendix A, the primers were unique for the targeted regions of the HVI SNP locations. The various primer sequences were then analyzed using the Primer3 software and it was seen that no self-dimerization or secondary structure was present in the proposed oligonucleotide sequences (Rozen & Skaletsky, 1998). The primer sequences were then analyzed using Auto-Dimer software to check on any potential dimerization between the different primers, and none was found (Vallone & Butler, 2004). The primer sequences were then run through BLAST using the following settings; BlastN algorithm, no complexity, expect 10 matches (Altschul et al., 1990). When this was done, it was seen that four of the five sets of primer sequences had the top alignment with the human mtDNA genome, with the next best

alignment score being several orders of magnitude less that the human mtDNA alignment. The exceptions were the 16519 forward and reverse primers. The 16519 primer sequences were seen to still have a secondary alignment with some nuclear genes, with the forward primer aligning to a location on Chromosome 11, with minor builds to other sites on Chromosomes 13, 5, and 4. The reverse primer was aligning to a site on Chromosome 17, with minor builds to sites on Chromosomes 1, 7, 8, 9, and the X-chromosome. It was established based on the BLAST data there should be no risk of amplification of the nuclear sites due to the distances between the forward and reverse primers, but minor secondary amplification might be expected for the 16519 primer set at these locations, and would be considered when analyzing the data.

The necessary forward and reverse primers were then ordered from Integrated DNA Technologies (Coralville, IA) in 100 µM concentrations. These were reconstituted to 100 µM primer stocks using sterile water, and working concentrations of each primer were made at 10 µM. Each of the individual primer pairs were then run against a standard extract of HL60 DNA (0.1 ng / 10 µl) (Collins, 1987) in a 25 µl reaction using the *Taq* polymerase, under the following conditions:

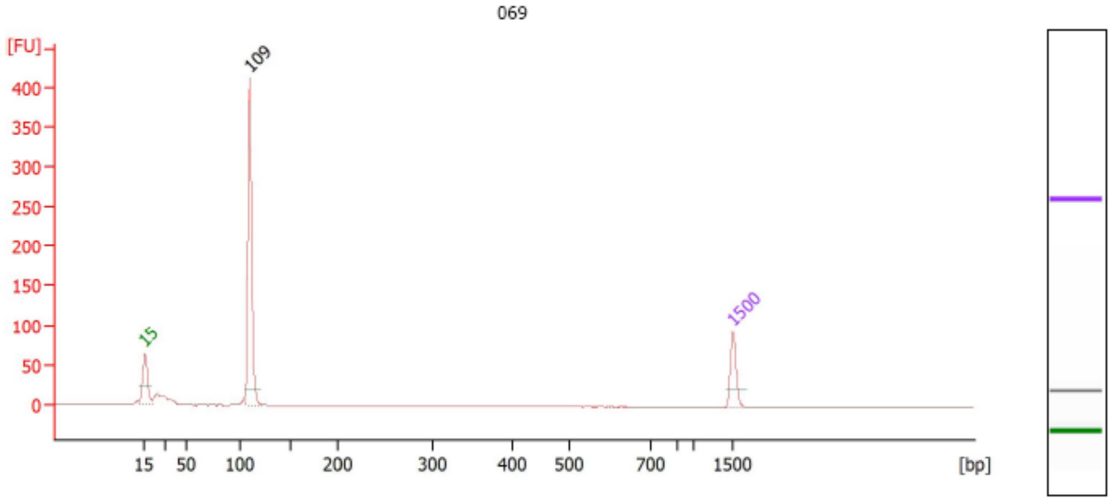
Soak at 95°C for 9 minutes
34 cycles:
- 95°C for 10 seconds
- 60°C for 30 seconds
- 72°C for 30 seconds

Storage soak at 4°C indefinitely

The amplified samples were then run on the Agilent 2100 Bioanalyzer (Agilent Technologies, Santa Clara, CA) to determine the quality and quantity of the amplified mtDNA. It was seen that each primer pair successfully amplified the separate HVI SNP regions based on the comparison between the measured amplicon size and the expected amplicon size from the rCRS map. The results from the first run are as follows:

(Note the Agilent 2100 Bioanalyzer data presented below contains the following: the electropherogram sizing graph, virtual gel lane, and the table of the peak data. The peaks in the graph are labeled with the "bp size" of each sample, and each graph will contain an upper and lower marker peak, at 15 and 1500 bp, respectively. The peak data table contains the peak size (in bp), concentration (in ng/µl), and molarity data (in nmol/L) from each peak in the sample.)

Figure 16- Agilent 2100 Bioanalyzer data for 16069 primers



Overall Results for sample 2 : 069

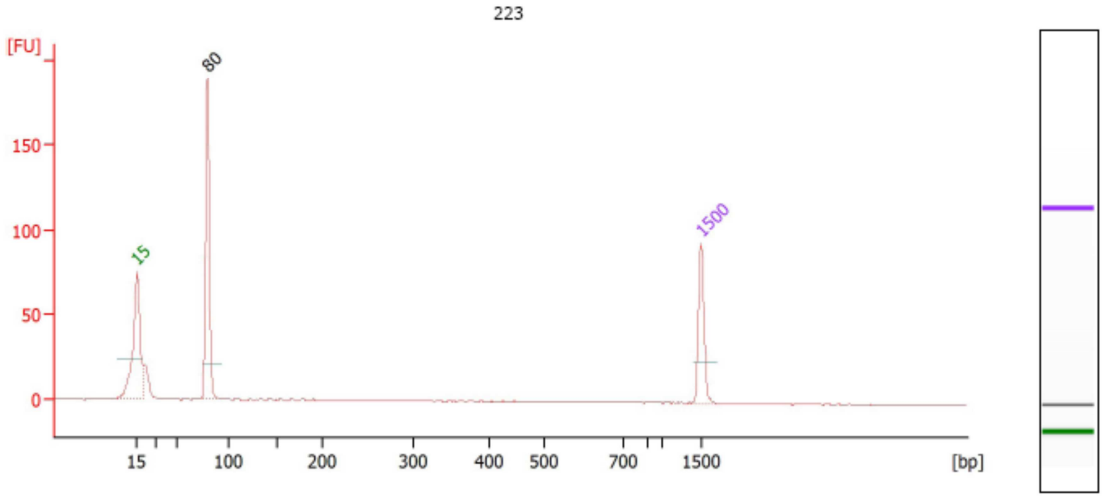
Number of peaks found: 1

Peak table for sample 2 : 069

Peak	Size [bp]	Conc. [ng/µl]	Molarity [nmol/l]	Observations	Aligned Migration Time [s]	Area	Time corrected area	Peak Height	Peak Width	% of Total
1	15	4.20	424.2	Lower Marker	43.00	40.2	101.1	63.8	1.5	0.0
2	109	13.98	193.6		55.50	192.7	377.0	413.2	1.9	100.0
3	1,500	2.10	2.1	Upper Marker	113.00	68.3	66.3	94.5	2.5	0.0

The associated amplification negative control showed no peaks for this sample.

Figure 17- Agilent 2100 Bioanalyzer data for 16223 primers



Overall Results for sample 4 : 223

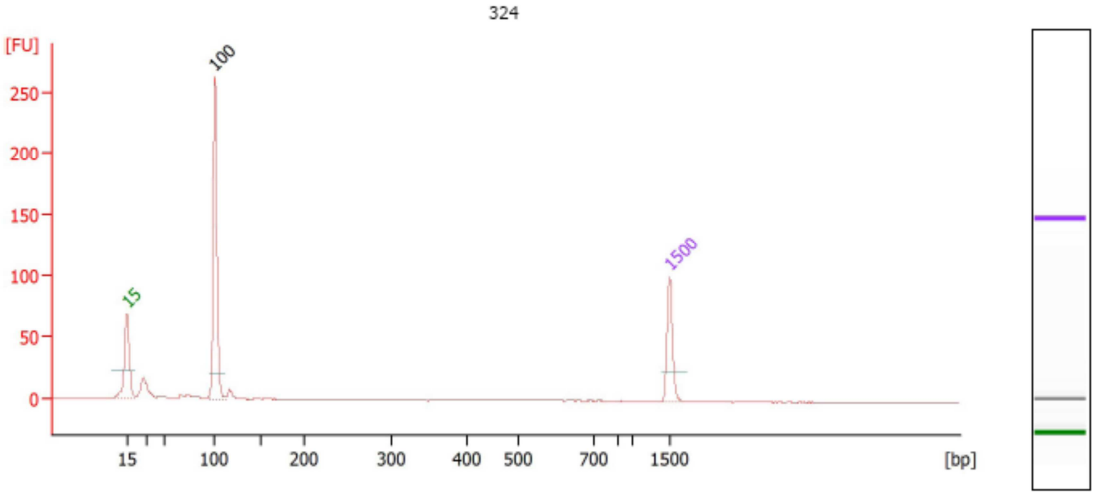
Number of peaks found: 1

Peak table for sample 4 : 223

Peak	Size [bp]	Conc. [ng/µl]	Molarity [nmol/l]	Observations	Aligned Migration Time [s]	Area	Time corrected area	Peak Height	Peak Width	% of Total
1	15	4.20	424.2	Lower Marker	43.00	66.8	171.0	75.2	3.3	0.0
2	80	7.21	137.2		51.76	89.0	190.5	190.4	2.5	100.0
3	1,500	2.10	2.1	Upper Marker	113.00	66.8	66.6	94.4	2.9	0.0

The associated amplification negative control showed no peaks for this sample.

Figure 18- Agilent 2100 Bioanalyzer data for 16324 primers



Overall Results for sample 6 : 324

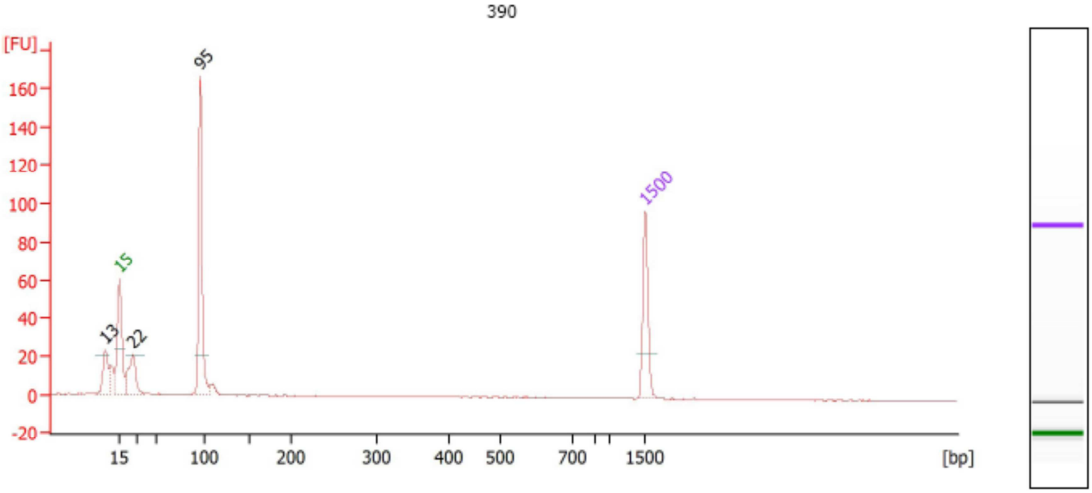
Number of peaks found: 1

Peak table for sample 6 : 324

Peak	Size [bp]	Conc. [ng/µl]	Molarity [nmol/l]	Observations	Aligned Migration Time [s]	Area	Time corrected area	Peak Height	Peak Width	% of Total
1	15	4.20	424.2	Lower Marker	43.00	46.0	120.2	69.7	3.2	0.0
2	100	8.85	134.2		54.36	122.4	255.8	263.5	2.0	100.0
3	1,500	2.10	2.1	Upper Marker	113.00	69.8	71.6	101.2	3.4	0.0

The associated amplification negative control showed no peaks for this sample.

Figure 19- Agilent 2100 Bioanalyzer data for 16390 primers



Overall Results for sample 8 : 390

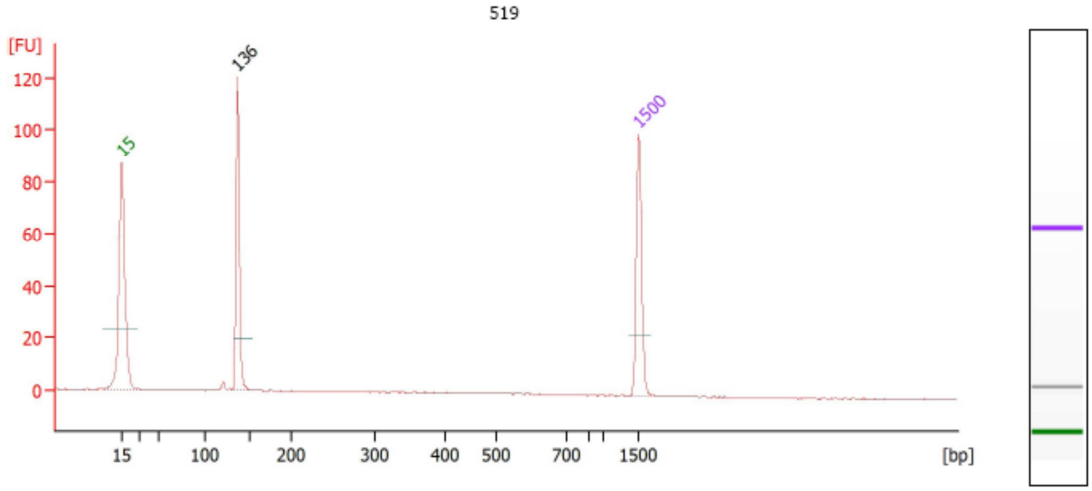
Number of peaks found: 2

Peak table for sample 8 : 390

Peak	Size [bp]	Conc. [ng/µl]	Molarity [nmol/l]	Observations	Aligned Migration Time [s]	Area	Time corrected area	Peak Height	Peak Width	% of Total
1	13	0.00	0.0		41.11	15.1	41.9	23.2	1.8	0.0
2	15	4.20	424.2	Lower Marker	43.00	36.1	95.9	60.4	1.4	0.0
3	22	2.54	173.9		44.77	20.0	51.1	20.7	2.7	21.7
4	95	5.76	91.4		53.79	72.2	155.4	167.6	1.8	78.3
5	1,500	2.10	2.1	Upper Marker	113.00	64.2	67.5	98.2	2.8	0.0

The associated amplification negative control showed no peaks for this sample.

Figure 20- Agilent 2100 Bioanalyzer data for 16519 primers



Overall Results for sample 10 : 519

Number of peaks found: 1

Peak table for sample 10 : 519

Peak	Size [bp]	Conc. [ng/µl]	Molarity [nmol/l]	Observations	Aligned Migration Time [s]	Area	Time corrected area	Peak Height	Peak Width	% of Total
1	15	4.20	424.2	Lower Marker	43.00	67.7	181.1	87.3	4.8	0.0
2	136	3.53	39.4		58.68	51.2	102.3	121.0	2.7	100.0
3	1,500	2.10	2.1	Upper Marker	113.00	68.7	73.0	100.5	3.1	0.0

The associated amplification negative control showed no peaks for this sample.

The results of the sizing for each individual primer set were established as presented in Table 1:

Table 1- SNP base-pair sizing

SNP Position	16069	16223	16324	16390	16519
Base-pair size	109	80	100	95	136

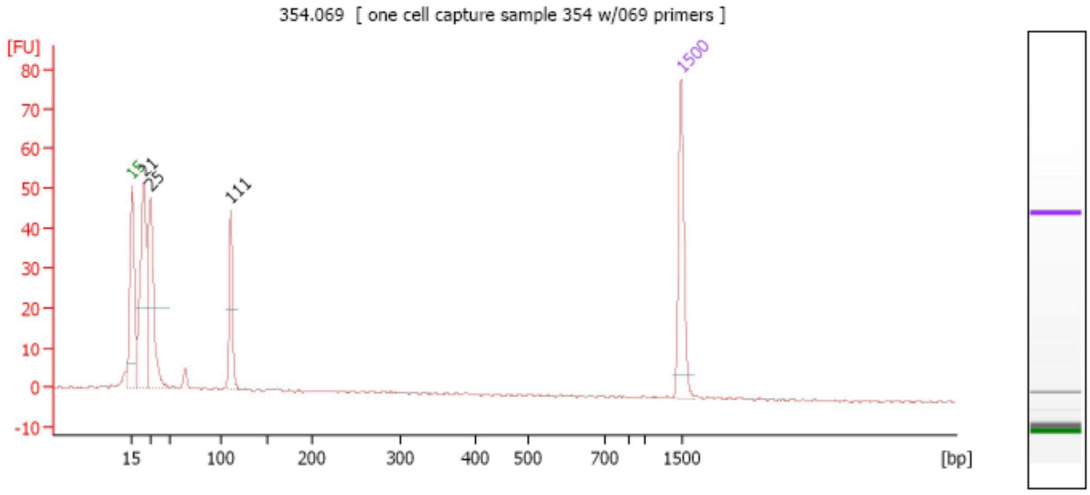
Validation of the Single-Cell Amplification Using the HVI SNP Primers

Once the primers were established as properly amplifying the HVI regions of interest for the SNP assay, it was decided that the critical factor in the project was whether or not the mtDNA from a single cell could be successfully amplified with the same primers. In order to validate the one-step amplification on a single cell, a total of 50 cells were collected by isolating 10 cells each from the following tissue samples; 336, 346, 340, 351, and 354. Each of these tissue samples had gone through the process of paraffin removal, mashing and cell filtration, and cell staining with the MitoTracker® Green FM and Nuclear Fast Red dyes. The cells from a single tissue were immobilized onto the PEN microscope slides and desiccated for 24 hours prior to laser dissection. The samples were screened for intact, single cells under 20 X - 40 X microscopy, and then using the PALM computer interface the single cells were cut and isolated, with each cell captured into a separate microcentrifuge tube that contained 18-20 µl of sterile water in the cap.

Based on the SNP primer validation data, it was seen that the highest yields from the 10 µM primers were from the 16069 primer pairs, so only these primers were run against one cell from each tissue source, using the standard 34-cycle amplification protocol. The cells were lysed and the mtDNA was amplified in this one-step reaction.

The rationale of the one-step HVI amplification is that at the beginning of the amplification, the 9-minute, 95 °C soak which activates the polymerase is also rupturing the cellular membranes, organelle membranes, and freeing the mtDNA into the surrounding amplification master mix. The water and PCR buffers would also be contributing osmotic pressure to the cells, further disrupting the membranes. The amplification of the 16069 primers against the cells from the five tissues produced the following results shown in Figure 21:

Figure 21- Tissue 354, single cell with 16069 primers, 34-cycle amplification



Overall Results for sample 8 : 354.069

Number of peaks found: 3

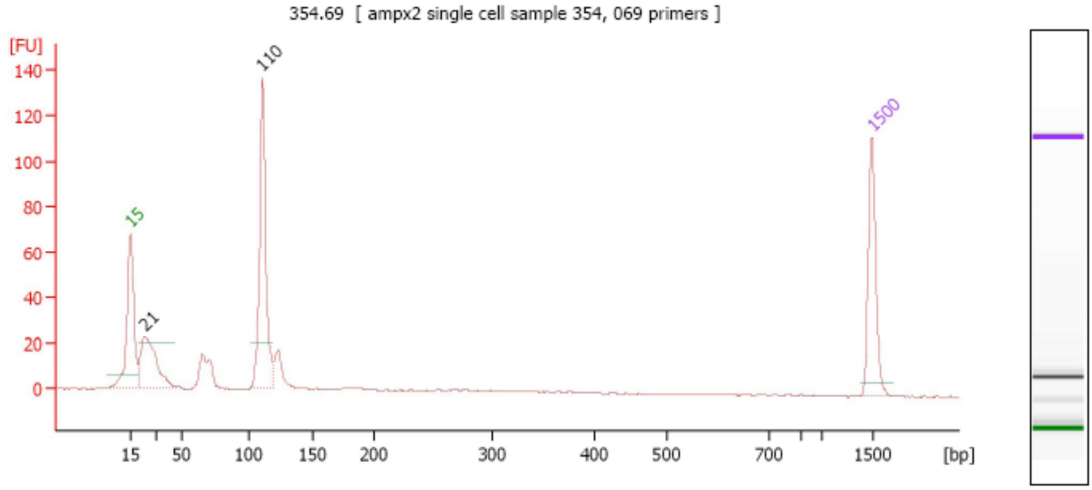
Peak table for sample 8 : 354.069

Peak	Size [bp]	Conc. [ng/µl]	Molarity [nmol/l]	Observations	Aligned Migration Time [s]	Area	Time corrected area	Peak Height	Peak Width	% of Total
1	15	4.20	424.2	Lower Marker	43.00	29.8	77.0	50.8	1.3	0.0
2	21	5.93	423.4		44.51	40.7	101.7	51.4	1.3	43.5
3	25	4.86	296.4		45.38	34.4	84.4	47.8	2.8	36.8
4	111	1.62	22.1		55.62	18.4	37.2	45.0	1.6	19.7
5	1,500	2.10	2.1	Upper Marker	113.00	56.4	57.0	80.5	2.8	0.0

Out of the five initial tissues, this was the only sample to have a measureable result, but it was clear that the expected amplicon size (111 bp) was present above the threshold of detection for the Agilent 2100 Bioanalyzer, and therefore amplification of the 16069 SNP region from a single cell was achieved.

At this point it became necessary to determine if the single-cell samples can be over-amplified to increase the yield and give measureable data. The samples were re-amplified for another 34 rounds of amplification to give a total of 68 cycles of amplification to the samples. When run on the Agilent 2100 Bioanalyzer, the samples were seen to have peaks in the 109-111 bp range, but the amplification negative control also contained peaks in this same range, meaning that low-level contaminants were also over-amplified. The results are presented in Figure 22:

Figure 22- Tissue 354, single cell with 16069 primers, 68-cycle amplification



Overall Results for sample 8 : 354.69

Number of peaks found: 2

Peak table for sample 8 : 354.69

Peak	Size [bp]	Conc. [ng/µl]	Molarity [nmol/l]	Observations	Aligned Migration Time [s]	Area	Time corrected area	Peak Height	Peak Width	% of Total
1	15	4.20	424.2	Lower Marker	43.00	44.1	113.7	67.7	3.0	0.0
2	21	3.51	254.3		44.42	33.1	82.8	22.9	3.5	29.5
3	110	5.03	69.2		55.51	78.9	160.1	136.7	2.1	70.5
4	1,500	2.10	2.1	Upper Marker	113.00	77.8	79.5	113.8	3.2	0.0

However, when looking at the results from the tissue 354 cell, it is clear that the second round of amplification increased the 16069 HVI region yield nearly 3 times. Since the over-amplification of 68 cycles resulted in a failed amplification negative control, it was decided that if additional rounds of amplification were needed, only 16 additional rounds would be done to give a total of 50 cycles of amplification.

The remainder of the cells were run against all of the primer pairs to determine if other individual cells could yield amplicons at the different SNP regions. The samples were run at 50 cycles of amplification using the one-step amplification strategy. In general, the amplification negative controls for the 16223 and 16324 amplifications showed minor amplification at the same size of the expected amplicon for those regions, so the data is inconclusive; however, every sample from the set resulted in a quantifiable peak on the Agilent 2100 Bioanalyzer. For the

cells that were amplified using the 16390 and 16519 primers, the amplification negative controls passed analysis, and for the 16390 samples, all six cells (one from each of the tissues) gave a quantifiable peak on the Agilent 2100 Bioanalyzer, with a range of 4.89 ng/µl to 6.65 ng/µl. It is important to note here that any of these samples would then be capable of being re-amplified in the SNaPshot™ system, based on the need to have an input of 0.2 pmol of template DNA for the SNaPshot™ reaction. Likewise, for the 16519 primers that were run across the different cells, the amplifications were successful for three of the six cells, resulting in an amplification yield range of 0.89 ng/µl to 4.22 ng/µl, which again is sufficient for SNaPshot™ analysis.

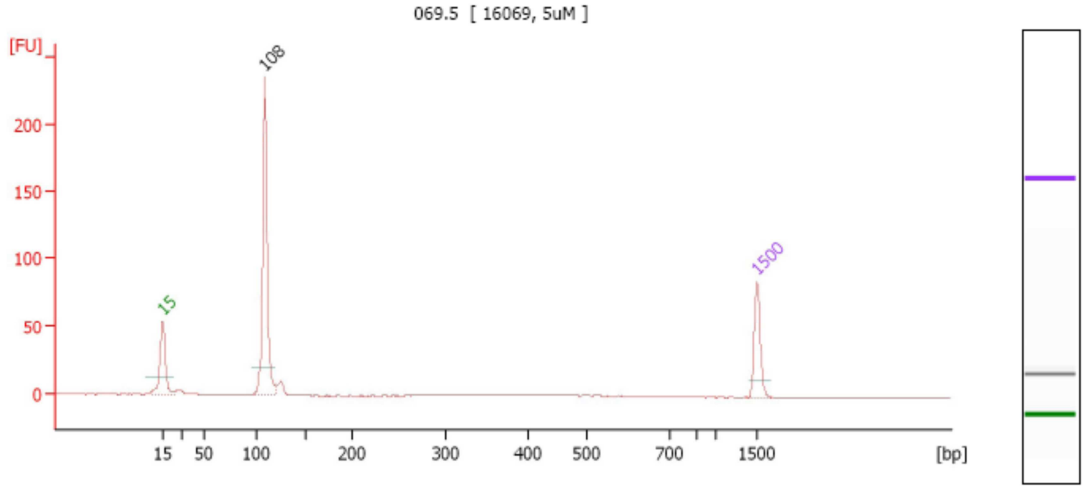
The following observations were made from these experiments: 1) mtDNA HVI amplification was accomplished on a single cell using a one-step amplification strategy, 2) the yield from these amplifications is sufficient for secondary amplification in the SNaPshot™ reaction, and 3) 50 cycles of amplification would be the maximum cycle threshold for the amplifications, and will only be used if the original 34 cycles of amplification results in an un-quantifiable peak on the Agilent 2100 Bioanalyzer.

Primer Titration Studies and Multiplex Design

The primers were seen to amplify the HVI SNP region as expected so the next step was to titrate the primers at various concentrations to find which primer concentration produced a standard yield of amplified DNA that was equal to the amplicons of the other primers, so that the final output for each primer set would result in evenly-matched peak heights after the multiplex amplification of the gross tissue samples. The primers were aliquotted from the 100 µM primer stock to the following concentrations: 1 µM, 5 µM, 10 µM, 15 µM, and 25 µM. The primers were run for each SNP region primer pair against HL60 template DNA (0.1 ng / 10 µl) using the standard amplification strategy, but increasing the cycles from 34 to 50 to verify that the

amplification negative controls from each primer titration would pass at the upper end of the previously established amplification range. The optimum yield for each of the five primer sets is shown in Figure 23 through Figure 27:

Figure 23- Agilent 2100 Bioanalyzer data for 16069 primers, best yield at 5 µM



Overall Results for sample 4 : 069.5

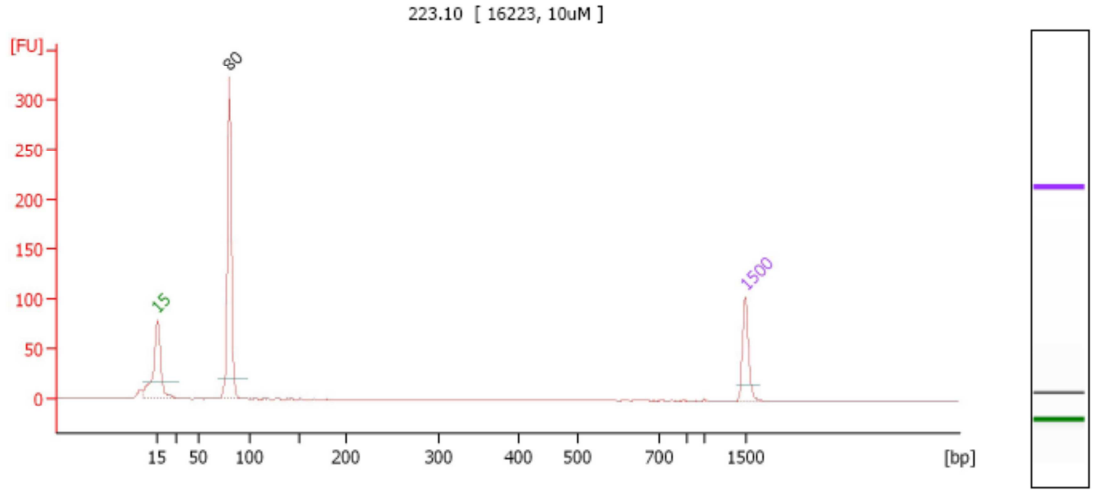
Number of peaks found: 1

Peak table for sample 4 : 069.5

Peak	Size [bp]	Conc. [ng/µl]	Molarity [nmol/l]	Observations	Aligned Migration Time [s]	Area	Time corrected area	Peak Height	Peak Width	% of Total
1	15	4.20	424.2	Lower Marker	43.00	38.2	91.1	53.6	3.3	0.0
2	108	9.96	139.2		55.05	133.6	250.9	236.0	2.9	100.0
3	1,500	2.10	2.1	Upper Marker	113.00	66.6	61.9	86.1	2.7	0.0

The associated amplification negative control showed no peaks for this sample.

Figure 24- Agilent 2100 Bioanalyzer data for 16223 primers, best yield at 10 µM



Overall Results for sample 6 : 223.10

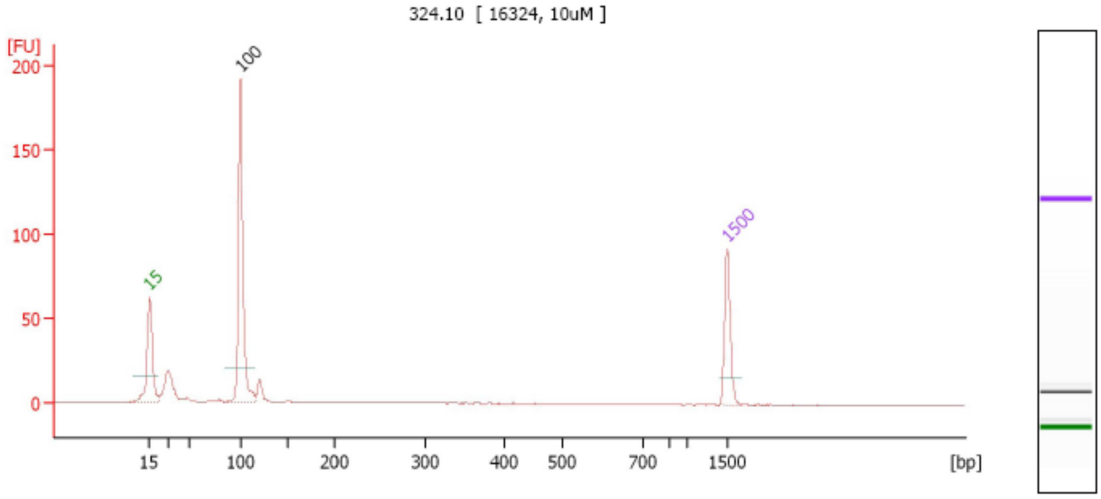
Number of peaks found: 1

Peak table for sample 6 : 223.10

Peak	Size [bp]	Conc. [ng/µl]	Molarity [nmol/l]	Observations	Aligned Migration Time [s]	Area	Time corrected area	Peak Height	Peak Width	% of Total
1	15	4.20	424.2	Lower Marker	43.00	74.6	181.2	79.1	4.3	0.0
2	80	11.05	208.9		51.58	162.8	332.0	322.8	3.6	100.0
3	1,500	2.10	2.1	Upper Marker	113.00	79.6	75.4	103.4	2.8	0.0

The associated amplification negative control showed no peaks for this sample.

Figure 25- Agilent 2100 Bioanalyzer data for 16324 primers, best yield at 10 µM



Overall Results for sample 6 : 324.10

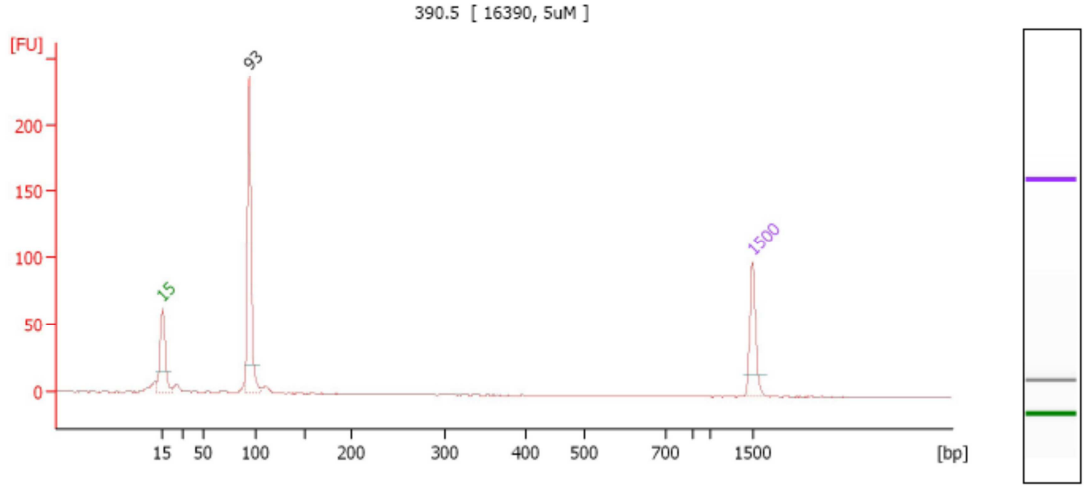
Number of peaks found: 1

Peak table for sample 6 : 324.10

Peak	Size [bp]	Conc. [ng/µl]	Molarity [nmol/l]	Observations	Aligned Migration Time [s]	Area	Time corrected area	Peak Height	Peak Width	% of Total
1	15	4.20	424.2	Lower Marker	43.00	43.6	107.8	62.0	3.2	0.0
2	100	8.12	123.1		54.04	107.7	214.0	192.5	3.7	100.0
3	1,500	2.10	2.1	Upper Marker	113.00	67.0	64.9	92.8	2.7	0.0

The associated amplification negative control showed no peaks for this sample.

Figure 26- Agilent 2100 Bioanalyzer data for 16390 primers, best yield at 5 µM



Overall Results for sample 4 : 390.5

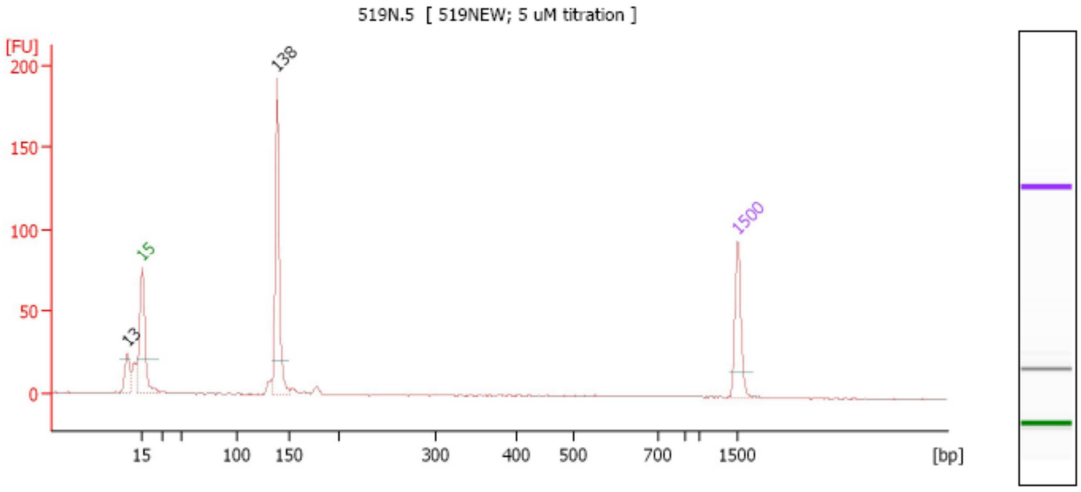
Number of peaks found: 1

Peak table for sample 4 : 390.5

Peak	Size [bp]	Conc. [ng/µl]	Molarity [nmol/l]	Observations	Aligned Migration Time [s]	Area	Time corrected area	Peak Height	Peak Width	% of Total
1	15	4.20	424.2	Lower Marker	43.00	43.2	103.3	63.5	1.8	0.0
2	93	8.05	130.7		53.26	118.6	230.7	238.0	2.0	100.0
3	1,500	2.10	2.1	Upper Marker	113.00	76.0	71.0	101.2	2.7	0.0

The associated amplification negative control showed no peaks for this sample.

Figure 27- Agilent 2100 Bioanalyzer data for 16519 primers, best yield at 5 µM



Overall Results for sample 4 : 519N.5

Number of peaks found: 1

Peak table for sample 4 : 519N.5

Peak	Size [bp]	Conc. [ng/µl]	Molarity [nmol/l]	Observations
1	13	0.00	0.0	
2	15	4.20	424.2	Lower Marker
3	138	6.11	67.1	
4	1,500	2.10	2.1	Upper Marker

The associated amplification negative control showed no peaks for this sample. The yellow "lock" symbol in the 2nd line of the above table is indicative of the manual labeling of the lower marker, since the shoulder peak at 13 bp was originally incorrectly labeled as such.

With the 16519 primers, there was the known issue of possible nuclear DNA co-amplification. Possibly as a result of this and the increased amplification cycle number, only the primers titrated at 1 µM and 5 µM passed. This was due to the amplification negative controls for the 15 µM and 25 µM primer titration samples containing peaks at the expected amplicon size, and the 10 µM primer titration sample did not display any amplicon peak. The best data overall was for the primers titrated at 25uM, giving 6.95 ng/µl at 138 bp, but the passing data was for the primer titration of 5µM, giving 6.11 ng/µl at 138 bp.

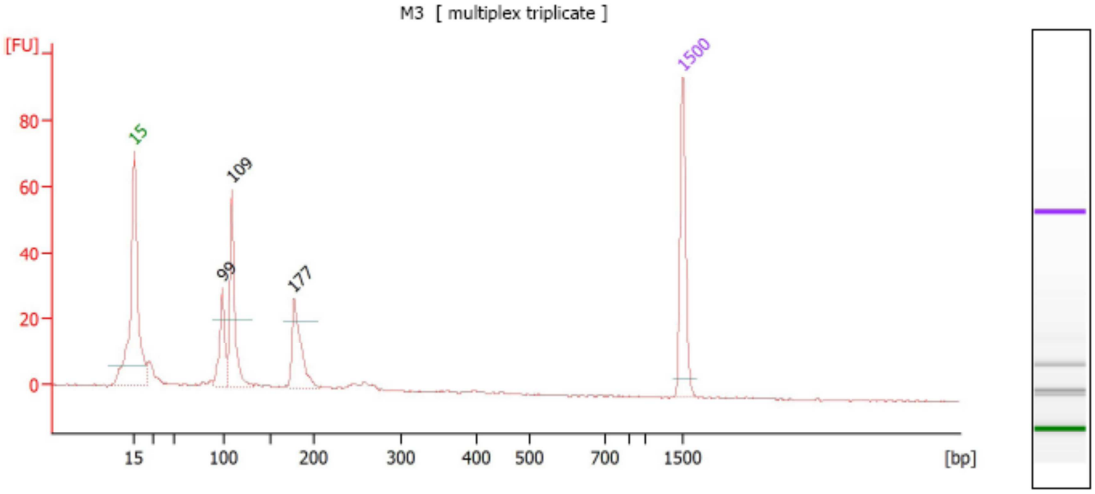
The optimal primer concentrations were established as follows in Table 2:

Table 2- Optimal SNP primer concentrations

SNP Position	16069	16223	16324	16390	16519
Primer conc.	5 µM	10 µM	10 µM	5 µM	5 µM

The primers were seen to optimally amplify the HVI SNP region as expected at specific concentrations of each primer, so the next step was to investigate if the primers worked cooperatively in a multiplex. Three replicate runs of the primer multiplex were then attempted, using the HL60 template DNA and standard 34-cycle amplification conditions as before, along with the optimal primer concentrations established for each primer set.

Figure 28- Agilent 2100 Bioanalyzer data for the Multiplex SNP Primers- 1st attempt



Overall Results for sample 4 : M3

Number of peaks found: 3

Peak table for sample 4 : M3

Peak	Size [bp]	Conc. [ng/µl]	Molarity [nmol/l]	Observations	Aligned Migration Time [s]	Area	Time corrected area	Peak Height	Peak Width	% of Total
1	15	4.20	424.2	Lower Marker	43.00	66.1	169.2	71.0	4.8	0.0
2	99	1.38	21.2		54.24	19.2	39.6	29.7	1.8	22.7
3	109	2.51	34.8		55.47	35.6	71.6	59.5	3.3	42.0
4	177	1.81	15.4		63.37	29.9	52.9	26.9	4.6	35.3
5	1,500	2.10	2.1	Upper Marker	113.00	70.5	71.3	96.8	3.1	0.0

All three of the runs revealed the same results, and the amplification negative for the sample passed. As it was seen in Figure 28, only three peaks were visible, with a double-sized peak around 109 bp. Based on the previous data for the single primer amplification, it was

expected that the amplicons for 16324 and 16390 would have resulted in a double-high peak being so close in size (95 bp and 100 bp, respectively). Also, the longer amplicon should have been the 16519 amplicon at 138 bp, which was missing in the multiplex set, and instead an even longer amplicon of 177 bp was appearing. The strongest peak at 109 corresponded perfectly with the 16069 amplicon, but it registered approximately twice the size of the surrounding peaks, which was inconsistent with the data from the individual primer titration study. It was hypothesized at this point that additional priming between various forward and reverse primers might be occurring to give the non-standard multiplex results. A sequence alignment analysis was then done to determine all of the possible amplicon sizes that could occur in the HVI region among all of the various forward and reverse SNP primers. The results of the amplifications are grouped into the SNP map of HVI in Appendix B.

In Appendix B, the numbers in the yellow boxes above the red brackets indicate the estimated amplicon size for each forward and reverse primer pair, followed by the actual amplicon size determined by the Agilent 2100 Bioanalyzer (in parentheses). Based on the HVI amplicon map, it was hypothesized that the 177 bp peak seen in the multiplex was a result of the amplification of the 16324 forward and 16390 reverse primers, and the 99 bp peak is the 16324 amplicon by itself. The 16223 and 16519 primers were not being activated in the multiplex, possibly due to an overwhelming amount of the other primers in competition for the template DNA. The thermodynamics of the amplification were also re-analyzed. Since the melting temperature of each primer pair was established during primer design, and individually each primer set was validated as properly working, then it was established that there might be competition within the multiplex to accomplish the proper melting temperature due to the physical presence of the other primer peaks.

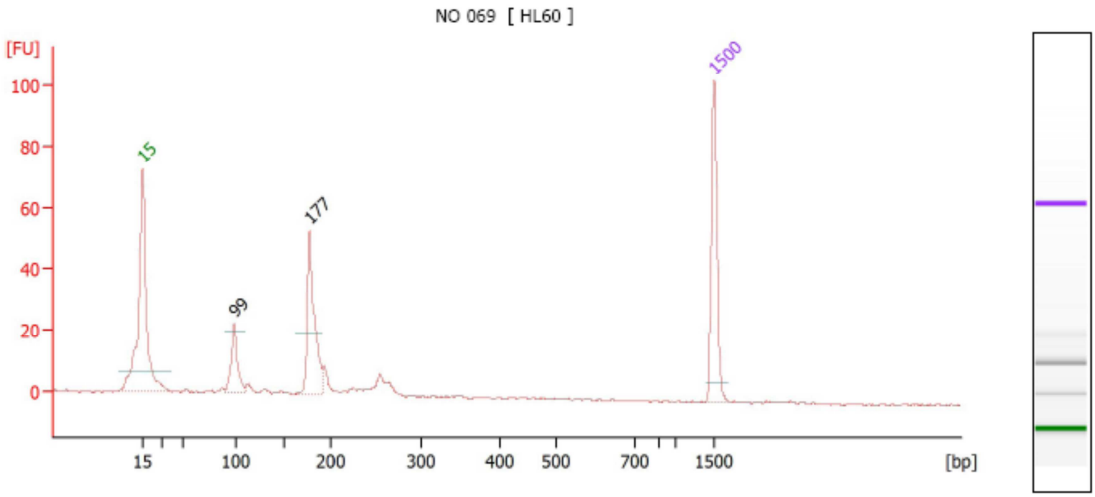
To screen for possible primer interference, an experiment was designed to re-run the multiplex by systematically withholding one of the primer sets to see if any of the primers are interfering with the overall multiplex efficiency. The scheme for the design of the amplification is presented in Table 3:

Table 3- Setup scheme for primer withholding study

Amplification # 1	WITHHELD	16223	16324	16390	16519
Amplification # 2	16069	WITHHELD	16324	16390	16519
Amplification # 3	16069	16223	WITHHELD	16390	16519
Amplification # 4	16069	16223	16324	WITHHELD	16519
Amplification # 5	16069	16223	16324	16390	WITHHELD

The samples in the withholding study were amplified using the HL60 template DNA using the same multiplex amplification conditions as before. Each of the primers were amplified at 5 µM concentrations to maintain primer concentration consistency. Note that this was a decrease in the previously established primer concentration for the 16223 and 16324 primers (see Table 2), which was done to equalize the size of these peaks to the 16069 peak for more even peak heights in the multiplex data. The amplification negative controls for each sample passed. The results for each of the five amplifications were as follows:

Figure 29- Agilent 2100 Bioanalyzer data for withholding study- no 16069 primers in Multiplex



Overall Results for sample 2 : NO 069

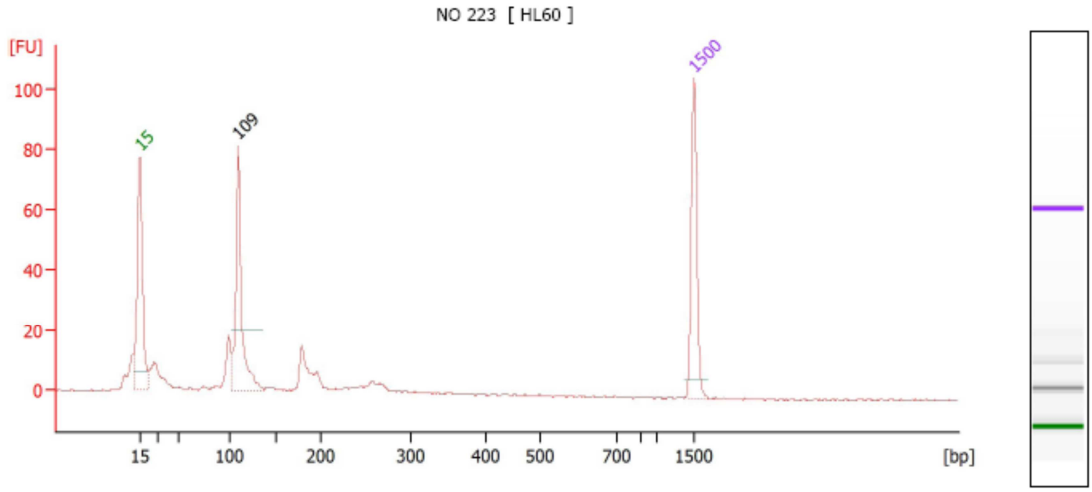
Number of peaks found: 2

Peak table for sample 2 : NO 069

Peak	Size [bp]	Conc. [ng/µl]	Molarity [nmol/l]	Observations	Aligned Migration Time [s]	Area	Time corrected area	Peak Height	Peak Width	% of Total
1	15	4.20	424.2	Lower Marker	43.00	71.2	178.7	72.7	6.5	0.0
2	99	1.19	18.3		54.23	18.0	36.1	22.4	2.5	26.9
3	177	2.72	23.2		63.43	48.9	84.2	53.8	3.3	73.1
4	1,500	2.10	2.1	Upper Marker	113.00	76.4	74.8	105.1	2.7	0.0

By withholding the 16069 primers, it was seen in Figure 29 that the 109 bp peak was not present, and the only resulting peaks are the 99 bp peak that was expected for 16324 and the 177 bp peak that was estimated to be the result of the 16324 forward and 16390 reverse primers. The amplicons for the 16223 and 16519 primer sets were still not present.

Figure 30- Agilent 2100 Bioanalyzer data for withholding study- no 16223 primers in Multiplex



Overall Results for sample 4 : NO 223

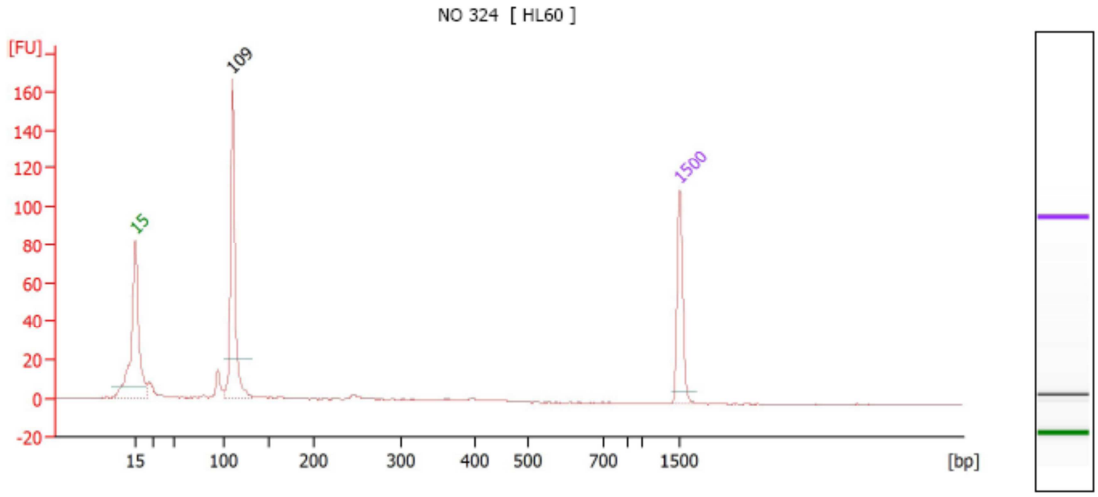
Number of peaks found: 1

Peak table for sample 4 : NO 223

Peak	Size [bp]	Conc. [ng/µl]	Molarity [nmol/l]	Observations	Aligned Migration Time [s]	Area	Time corrected area	Peak Height	Peak Width	% of Total
1	15	4.20	424.2	Lower Marker	43.00	54.1	137.6	77.7	1.9	0.0
2	109	3.75	52.2		55.44	58.5	116.8	81.6	4.0	100.0
3	1,500	2.10	2.1	Upper Marker	113.00	77.4	77.3	106.2	3.1	0.0

By withholding the 16223 primers, the 109 bp peak for 16069 was present in Figure 30, and unlabeled amplicons around 100 and 177 bp were still present, indicating the lack of the amplicon for the 16519 primers.

Figure 31- Agilent 2100 Bioanalyzer data for withholding study- no 16324 primers in Multiplex



Overall Results for sample 6 : NO 324

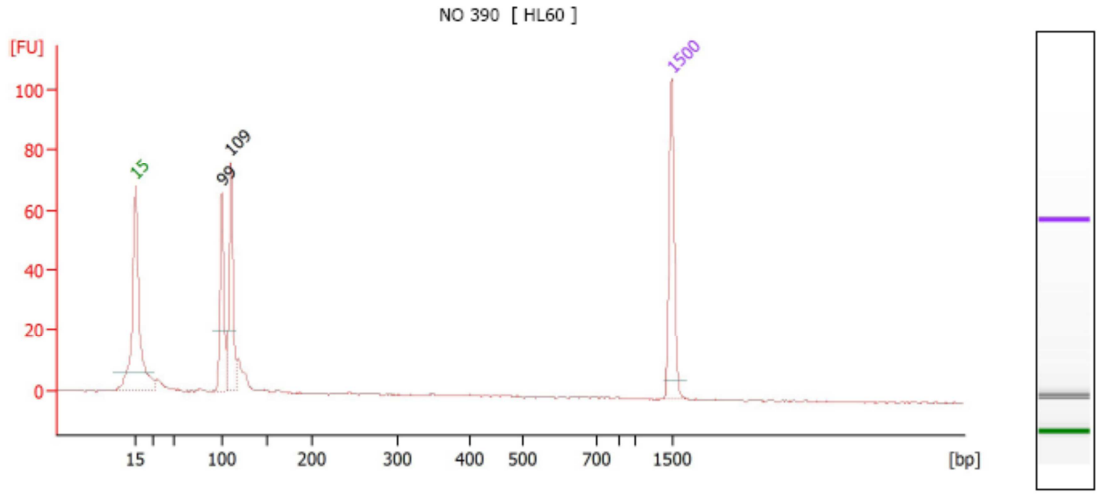
Number of peaks found: 1

Peak table for sample 6 : NO 324

Peak	Size [bp]	Conc. [ng/µl]	Molarity [nmol/l]	Observations	Aligned Migration Time [s]	Area	Time corrected area	Peak Height	Peak Width	% of Total
1	15	4.20	424.2	Lower Marker	43.00	77.6	200.1	82.7	4.5	0.0
2	109	5.76	79.8		55.50	93.8	189.8	166.9	3.8	100.0
3	1,500	2.10	2.1	Upper Marker	113.00	80.7	82.2	111.2	3.1	0.0

By withholding the 16324 primers, only the 109 bp peak of the 16069 amplicon was present in Figure 31, along with an unlabeled peak that was less than 100 bp. It is important to note here that neither the 100 nor 177 bp peaks were present, indicating that the 16324 primers played a role in the formation of those amplicons. The minor unlabeled peak that was less than 100 bp could be the 95 bp peak amplicon for the 16390 primers.

Figure 32- Agilent 2100 Bioanalyzer data for withholding study- no 16390 primers in Multiplex



Overall Results for sample 8 : NO 390

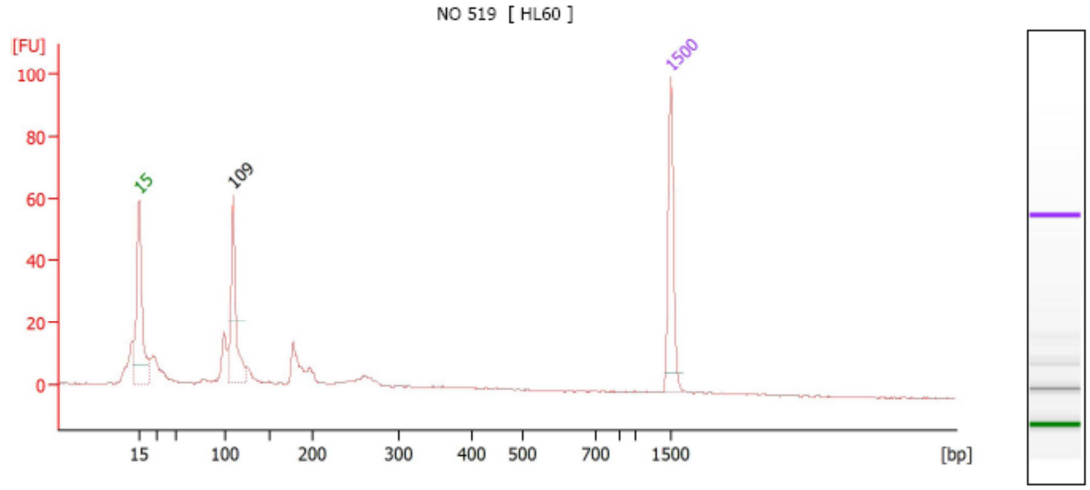
Number of peaks found: 2

Peak table for sample 8 : NO 390

Peak	Size [bp]	Conc. [ng/µl]	Molarity [nmol/l]	Observations	Aligned Migration Time [s]	Area	Time corrected area	Peak Height	Peak Width	% of Total
1	15	4.20	424.2	Lower Marker	43.00	60.8	157.4	68.1	5.5	0.0
2	99	2.33	35.5		54.31	32.8	68.2	66.3	1.8	46.7
3	109	2.61	36.1		55.50	37.5	76.3	76.0	1.3	53.3
4	1,500	2.10	2.1	Upper Marker	113.00	71.3	73.4	106.8	2.9	0.0

By withholding the 16390 primers, the 109 bp peak for the 16069 amplicon was present in Figure 32, as well as the 99 bp peak that was expected for the 16324 amplicon, but the 177 bp peak was not present due to the lack of the 16390 reverse primer, validating the hypothesis that the 177 bp peak was due to this primer interaction with the 16324 forward primer. It is again important to note that the amplicons for the 16223 and 16519 primers were not present.

Figure 33- Agilent 2100 Bioanalyzer data for withholding study- no 16519 primers in Multiplex



Overall Results for sample 10 : NO 519

Number of peaks found: 1

Peak table for sample 10 : NO 519

Peak	Size [bp]	Conc. [ng/µl]	Molarity [nmol/l]	Observations	Aligned Migration Time [s]	Area	Time corrected area	Peak Height	Peak Width	% of Total
1	15	4.20	424.2	Lower Marker	43.00	42.8	109.9	59.6	2.0	0.0
2	109	2.69	37.5		55.41	38.3	77.6	61.0	2.3	100.0
3	1,500	2.10	2.1	Upper Marker	113.00	70.7	72.4	101.9	2.6	0.0

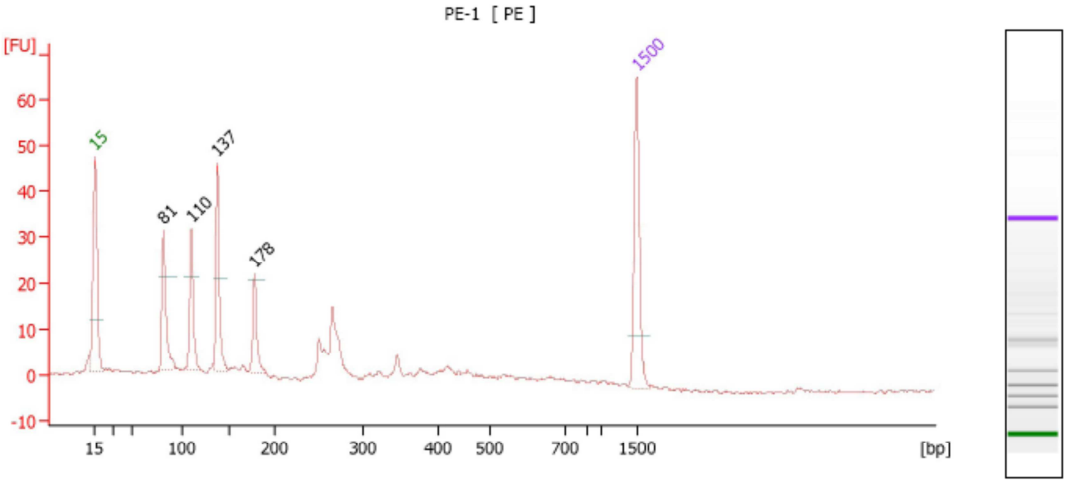
By withholding the 16519 primers, the 109 bp amplicon for the 16069 primers was again present in Figure 33, and unlabeled amplicons were also present at the 99 and 177 bp positions.

Two things were apparent based on this data: 1) internal priming was occurring between the 16324 forward and 16390 reverse primers, and 2) the 16223 and 16519 primers were not affected by the removal of any other primer pairs, indicating they were below an effective concentration for the multiplex. To this end, it was determined that the preferential internal priming between the 16324 forward primer and the 16390 reverse primer would be exploited to create a "block amplicon" that would then cover the entire region for the 16324 and 16390 hotspots, and the primers for the 16223 hotspot would be increased back to the original 10 µM concentration within the multiplex amplification setup. The rationale for the block amplicon was that the two corresponding SNP hotspots, 16324 and 16390 would be targeted with extension

primers on opposing strands, with the 16324 extension primers binding to the forward strand and the 16390 extension primers binding to the reverse strand, so there would be no competition of the two opposing primers during SNP analysis. This would also eliminate an amplicon peak from the Agilent 2100 Bioanalyzer runs, and allow the remaining four peaks to be more readily resolved.

The new multiplex amplification was established using the block amp design and an increased concentration of the 16223 primers, and was tested in triplicate against HL60 template DNA (0.1 ng / 10 µl) using the standard 34-cycle amplification conditions. The associated amplification negatives passed. The results of the multiplex amplification were as follows:

Figure 34- Agilent 2100 Bioanalyzer data for optimized Multiplex with block amplification design



Overall Results for sample 1 : PE-1

Number of peaks found: 4

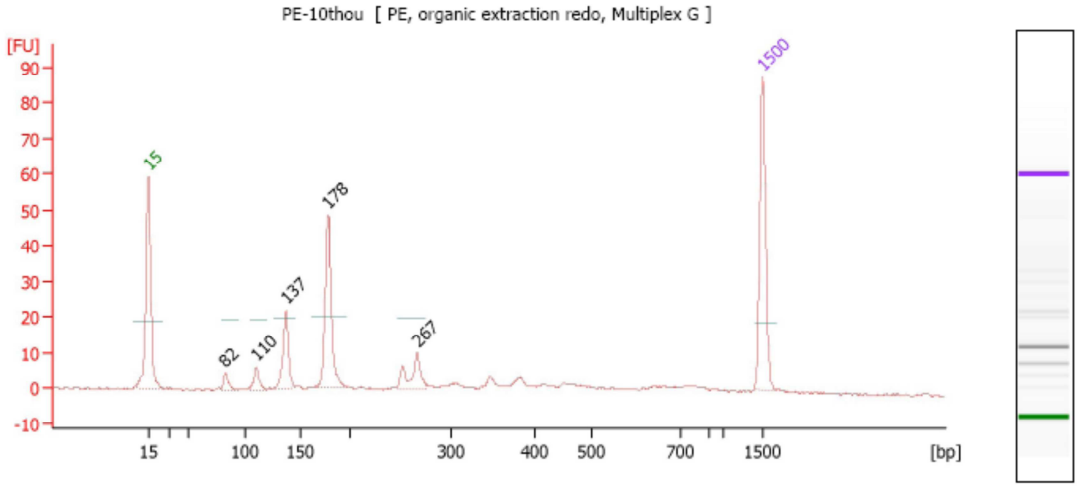
Peak table for sample 1 : PE-1

Peak	Size [bp]	Conc. [ng/µl]	Molarity [nmol/l]	Observations
1	15	4.20	424.2	Lower Marker
2	81	1.87	34.8	
3	110	1.51	20.8	
4	137	2.00	22.2	
5	178	1.00	8.5	
6	1,500	2.10	2.1	Upper Marker

The resulting multiplex amplification using the block amp and the increased 16223 primer concentration successfully produced the expected amplicons. As seen in the Agilent 2100

Bioanalyzer electropherogram in Figure 34, the 80 bp peak corresponded to the amplicon for the 16223 primers, the 108 bp peak corresponded to the amplicon for the 16069 primers, the 138 bp peak corresponded to the amplicon for the 16519 primers, and the 177 bp peak corresponded to the block amplicon of the 16324 forward and 16390 reverse primers. As further work progressed on the validation study, it was noted that when processing validation tissue samples that the 177 bp block amplicon was amplifying below the optimal concentration, and the 80 bp amplicon for the 16223 primers was over-amplifying. A slight modification in the primer concentrations was therefore carried out, decreasing the 16223 primer concentration to 5 µM and increasing the 16324 forward and 16390 reverse primer concentrations to 10 µM in the multiplex amplification. This modified multiplex setup was run against HL60 template DNA using the standard 34-cycle amplification setup. The amplification negative control for the associated samples passed. The results of the HL60 amplification were as follows:

Figure 35- Agilent 2100 Bioanalyzer data for new Multiplex design; HL60 Positive Control DNA



Overall Results for sample 1 : PE-10thou

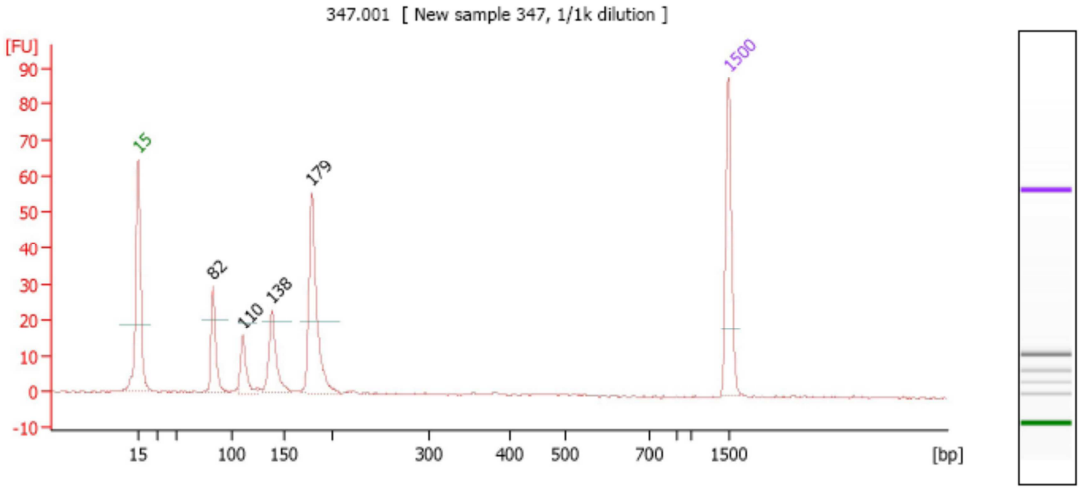
Number of peaks found: 5

Peak table for sample 1 : PE-10thou

Peak	Size [bp]	Conc. [ng/µl]	Molarity [nmol/l]	Observations
1	15	4.20	424.2	Lower Marker
2	82	0.29	5.3	
3	110	0.35	4.9	
4	137	1.11	12.3	
5	178	2.46	21.0	
6	267	0.74	4.2	
7	1,500	2.10	2.1	Upper Marker

While the new multiplex setup tended to skew the longer amplicons to higher post-amplification concentrations for HL60 as seen in Figure 35, it worked beautifully for the actual tissue samples in which the previous over- and under-amplification problems were noted. An example of a tissue extract, diluted at 1/1000 and amplified with the new multiplex is presented in Figure 36:

Figure 36- Agilent 2100 Bioanalyzer data for new Multiplex design run against tissue sample 347



Overall Results for sample 4 : 347.001

Number of peaks found: 4

Peak table for sample 4 : 347.001

Peak	Size [bp]	Conc. [ng/µl]	Molarity [nmol/l]	Observations
1	15	4.20	424.2	Lower Marker
2	82	1.43	26.3	
3	110	0.93	12.9	
4	138	1.55	17.1	
5	179	3.53	29.9	
6	1,500	2.10	2.1	Upper Marker

Based on these results, the multiplex SNP amplification was seen to be validated and was then used under the new multiplex conditions to process the gross tissue samples for this project.

Extension Primer Design and Validation

The implementation of the oligonucleotide extension primers for the SNaPshot™ assay had several additional primer design aspects that were required for the extension primers to be effective in targeting and locating the intended SNP hotspots in the HVI region. These extension primers needed to effectively target the SNP locations in question, but be smaller than 30-50 bp while still maintaining a unique length for each primer. In order to be properly separated and detected on the 3130*xl* capillary electrophoresis instrument, the extension primers needed to differ by at least four bp in length. The length differences between each extension primer were

accomplished with the use of poly-thymine tails at the 5′ end of the extension primers. As seen in the Appendix A, not all of the extension primers have the 5′ modification. The original primer sequences are presented in Table 4.

Table 4- Unmodified extension primers sequences

E16069 C-T	CCACCCAAGTATTGACT
E16223 C-T	AGCAAGTACAGCAATCAACC
E16324 T-C	TACATAGTACATAAAGCCAT
E16390 G-A	GATGGTGGTCAAGGGAC
E16519 T-C	TGTGGGCTATTTAGGCTTTATG

After the 5′ modification by adding the poly-thymine tails, the resulting primers are presented in Table 5, with the poly-thymine tail additions in subscript.

Table 5- Extension primer sequences modified with 5′ poly-thymine additions

E16069 C-T	CCACCCAAGTATTGACT
E16223 C-T	$_{T}$AGCAAGTACAGCAATCAACC
E16324 T-C	$_{TTTTT}$TACATAGTACATAAAGCCAT
E16390 G-A	$_{TTTTTTTTTTT}$GATGGTGGTCAAGGGAC
E16519 T-C	$_{TTTTTTTTTT}$TGTGGGCTATTTAGGCTTTATG

The additions of the thymine tails to four of the five extension primers allowed two conditions to be satisfied: 1) the primers were all separated by 4 bp, ideal for the separation on the capillary electrophoresis instrument, 2) the separation of the primers should result in the smallest primer being the 16069, followed in kind by the 16223, 16324, 16390, and finally the 16519. The thermodynamic measurements of these modified oligonucleotide primers all had a similar melting temperature, no auto-dimerization, no secondary structure, and no cross-dimerization with any of the other primers based on the analysis with the Primer3 software (Rozen & Skaletsky, 1998).

To check the amplification efficiency of the extension primers, each extension primer was run in a standard amplification against the opposing primer from the same HVI region of interest (see Appendix A and Appendix B for primer details and primer map). The extension primer and opposing primer set were each amplified at 10 µM concentration using HL60 as the

template DNA and the standard 34-cycle amplification setup, as used in the multiplex amplification assay. The resulting amplicons were run on the Agilent 2100 Bioanalyzer, and the expected amplicon sizes for each extension primer/opposing primer set were displayed. This data showed that the extension primers are targeting the expected regions of the HVI mtDNA, and since amplification occurred for each primer set, the extension primers are therefore capable of annealing and extending the corresponding bases. The extension primers were not titrated like the SNP multiplex primers, since the extension primers used in the SNaPshot™ assay noted a working range of input primer concentration between 0.05 µM and 1 µM. Another factor that was considered was the purification of the amplified multiplex mtDNA to remove the unincorporated primers and dNTP bases so that a second round of PCR could be performed with the extension primers.

Amplified DNA Purification and Protocol Validation

The first step in validating the extension primers for use in the SNaPshot™ assay was to evaluate the methods for purifying the initial amplification to remove the unincorporated primers and dNTPs. ExoSAP-IT® (USB Corporation, Cleveland, OH) was chosen due to the enzymatic combination of the exonuclease and shrimp alkaline phosphatase to purify the amplified mtDNA. Under the standard manufacturer's conditions, the purification was not efficient as secondary amplifications still produced the original multiplex peaks, indicating that the multiplex primers were not effectively removed. A second purification method using ethanol was attempted, in which the amplified mtDNA was precipitated and the supernatant removed, leaving purified mtDNA with none of the original PCR primers or dNTPs. This ethanol purification was seen to successfully remove the multiplex primers, but it was also noted that the ethanol-purified mtDNA quickly degraded after purification. This was seen in a time study in which various

volumes (1, 5, 8, and 10 µl) of ethanol-purified mtDNA were amplified using the extension primers and the associated opposing primers. The quantitation results of the amplifications were measured on the Agilent 2100 Bioanalyzer, and the results are presented in Table 6.

Table 6- Data for ethanol-purified mtDNA samples, 0 days after purification

	1 µL	**5 µL**	**8 µL**	**10 µL**
Peak #1	68 bp [1.78 ng/µl]	68 bp [1.40 ng/µl]	69 bp [1.23 ng/µl]	40 bp [0.88 ng/µl]
Peak #2	94 bp [0.42 ng/µl]	115 bp [1.53 ng/µl]	116 bp [1.34 ng/µl]	70 bp [1.13 ng/µl]
Peak #3	115 bp [1.53 ng/µl]	244 bp [1.38 ng/µl]	187 bp [0.37 ng/µl]	116 bp [1.00 ng/µl]
Peak #4	245 bp [1.10 ng/µl]	-	244 bp [1.60 ng/µl]	188 bp [0.48 ng/µl]
Peak #5	-	-	-	245 bp [1.55 ng/µl]
TOTAL [ ]	**4.83 ng/µl**	**4.31 ng/µl**	**4.54 ng/µl**	**4.04 ng/µl**

The purified mtDNA was stored at 4 ˚C for a total of 48 hours, and then re-sampled in the same volumes as before and again amplified with the same extension and opposing primers as before. The quantitation results of this second amplification were generated and are presented in Table 7.

Table 7- Data for ethanol-purified mtDNA samples, 2 days after purification

	1 µL	**5 µL**	**8 µL**	**10 µL**
Peak #1	68 bp [1.09 ng/µl]	68 bp [1.03 ng/µl]	69 bp [0.86 ng/µl]	69 bp [0.85 ng/µl]
Peak #2	117 bp [1.50 ng/µl]	84 bp [0.67 ng/µl]	84 bp [0.63 ng/µl]	84 bp [0.60 ng/µl]
Peak #3	244 bp [1.03 ng/µl]	117 bp [1.10 ng/µl]	118 bp [0.66 ng/µl]	117 bp [0.60 ng/µl]
Peak #4	-	245 bp [1.33 ng/µl]	244 bp [1.18 ng/µl]	243 bp [1.25 ng/µl]
TOTAL [ ]	**4.02 ng/µl**	**4.13 ng/µl**	**3.33 ng/µl**	**3.3 ng/µl**

The final data indicated that after 48 hours, there was a loss of mtDNA concentration and even loss of entire amplifiable peaks in all samples in the ethanol-purified mtDNA. This was

unacceptable since it was possible for purified samples to be unprocessed for up to a week or longer once the testing began on the single-cell samples.

A third option was then explored by revisiting the ExoSAP-IT® purification and applying the extended incubation times recommended in a SNaPshot™ protocol from Peter Vallone (Vallone et al., 2004). In this protocol, both the incubation and protein denaturation steps were significantly lengthened to provide more time for the ExoSAP-IT® enzymes to fully act on the amplified mtDNA. The recommended settings from Vallone were 90 minutes at 37 °C, 20 minutes at 80 °C, 5 minutes at 25 °C, and a final hold at 4 °C. Using these settings, purified multiplex mtDNA was generated and amplified using the extension primers and opposing primers. The quantitation results were obtained on the Agilent 2100 Bioanalyzer, and matched the expected amplicon sizes for the extension primer amplicons only, with no residual peaks from the multiplex amplification, indicating that the modified ExoSAP-IT® protocol was effectively removing the multiplex primers and dNTPs. To check the stability of the purified mtDNA, a time study was performed with the same setup used in the ethanol purification study, and repeated over 48 and 72 hours. For each replicate set of data, no significant loss of mtDNA was seen indicating that the mtDNA purified with the modified ExoSAP-IT® protocol was stable for use in the project.

SNaPshot™ Assay and Protocol Validation

SNaPshot™ kits were obtained from Applied Biosystems (Foster City, CA). To validate the kits, a series of HL60 amplicons were created for each SNP hotspot region using only the specific forward and reverse primers for that location, resulting in unique amplicons for 16069, 16223, 16324, 16390, and 16519 in separate tubes. A multiplex sample was also generated, along with the representative negative controls for each amplification set. This setup was

duplicated, in order for one set of amplicons to be purified with the ethanol precipitation process and the other with the modified ExoSAP-IT® protocol from Vallone, et al (2004). The ethanol precipitation was done on the same day as the SNaPshot™ assay to eliminate any possible effects of DNA degradation. For the SNaPshot™ assay, each amplified sample was aliquotted into four different tubes. Since the SNaPshot™ assay specified a range of amplified mtDNA input concentrations and a range of primer input concentrations, it was decided to modify these two variables across the four amplified aliquots, utilizing the scheme presented in Table 8.

Table 8- Setup scheme for SNaPshot™ assay validation

	0.2 µM primers	1.0 µM primers
0.2 pmol DNA	Sample 1	Sample 2
0.4 pmol DNA	Sample 3	Sample 4

Since the recommended input range of the DNA was from 0.01 pmol to 0.4 pmol in a 10 µl mixture, and the maximum DNA concentration was necessary to detect all primer binding sites, the maximum input (0.4 pmol) and half-maximum input (0.2 pmol) were attempted. The recommended input range of the primers was listed as 0.2 µM within the SNaPshot™ literature, with a note:

> "If a particular primer has a consistently low or high signal, increase or decrease the concentration of that primer. Successful results have been obtained using primers with concentrations that range between 0.05 µM and 1 µM in a six-primer mixture" (Applied-Biosystems, 2000).

Since there were five extension primers in the reaction, it was decided to test both the optimal primer concentration (0.2 µM) and the maximum recommended primer concentration (1.0 µM).

The HL60 samples were then amplified at the varying DNA input concentrations and primer concentrations, along with the necessary positive controls and negative controls for the SNaPshot™ reaction. The positive control and negative control for the SNaPshot™ kit was based on a specific DNA and primer set included with the SNaPshot™ kit, which was independent of the extension primers that were developed for this study. Therefore, a water blank sample was also run as a no-template control for the extension primers, which was not controlled by the negative control for the SNaPshot™ kit. The SNaPshot™ reaction was carried out under the following conditions-

25 cycles:
- 96 °C for 10 seconds
- 50 °C for 5 seconds
- 60 °C for 30 seconds

Storage soak at 4 °C indefinitely

Following the SNaPshot™ assay, the HL60 samples were treated with 1.5 units of SAP (shrimp alkaline phosphatase) (Promega Corporation, Madison, WI) and incubated for 60 minutes at 37 °C, followed by a 15 minute, 75 °C denaturation step, and then held at 4 °C indefinitely. Following SAP purification, 0.5 µl of each SNaPshot™ amplified sample was combined with 9 µl aliquots of Hi-Di formamide (Applied Biosystems, Foster City, CA) that also contained 0.5 µl of GS120-LIZ size standard (Applied Biosystems, Foster City, CA). These samples were heated and cooled to denature the target mtDNA, and injected onto a capillary array on the 3130*xl* genetic analyzer (Applied Biosystems, Foster City, CA), utilizing 7% performance-optimized polymer (POP-7™), a 36 cm capillary array, and the E5 chemistry detection and analysis protocol. Following electrophoresis and data collection, the data files were analyzed using GeneMapper™ software from Applied Biosystems (Foster City, CA).

It was seen in the analysis of the data that the individual extension primers for each specific region worked, but with a drastic difference in the peak heights of the ExoSAP-IT®

treated samples compared to the ethanol-precipitated samples in which the ExoSAP-IT® samples showed a 4-fold to 20-fold increase in the peak height values compared to the similar peaks in the ethanol-treated samples. For the multiplex samples, the ethanol precipitation failed to produce any viable SNP profile for any of the mtDNA/primer combinations, whereas the ExoSAP-IT® purified multiplex mtDNA gave a full and expected SNP profile for all five of the corresponding extension primers.

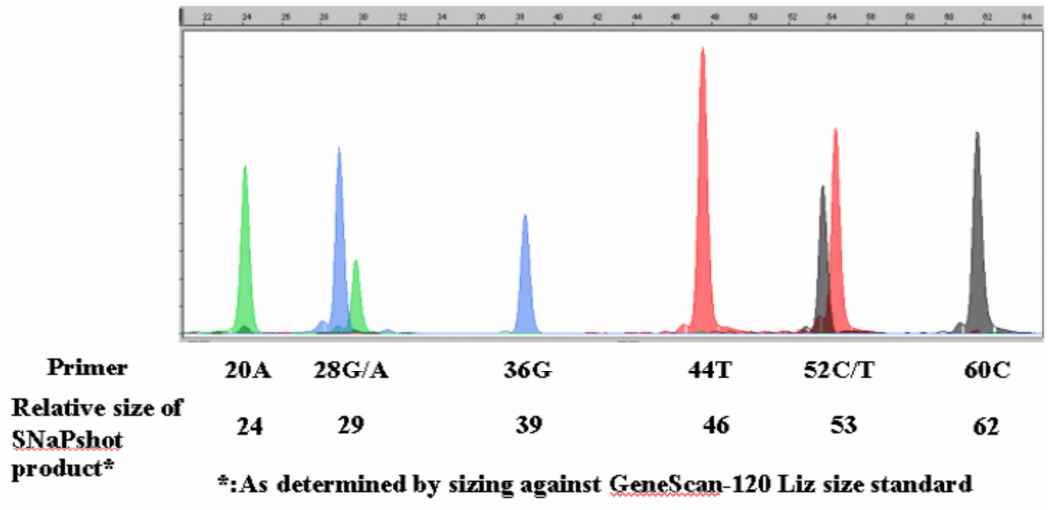
In regards to the mtDNA/primer input concentrations, it was noted that the combination of the 0.4 pmol mtDNA input and 1.0 µM primer input worked the best for the individual extension primers when applied to the separate amplicons, but caused the multiplex samples to become overblown. For the multiplex samples, the best peak heights and even signal output were seen in the 0.2 pmol mtDNA / 1.0 µM primer samples for the ExoSAP-IT® purified mtDNA.

Once the peak height and quality data was verified across the samples, it was necessary to determine the average migration time for each known peak and to check these times against the redundant datasets. For the SNaPshot™ positive control, the protocol listed the expected peaks according to color and migration pattern as presented in Table 9 and Figure 37 (Applied-Biosystems, 2000)-

Table 9- SNaPshot™ Positive Control expected values (Applied Biosystems, 2000)

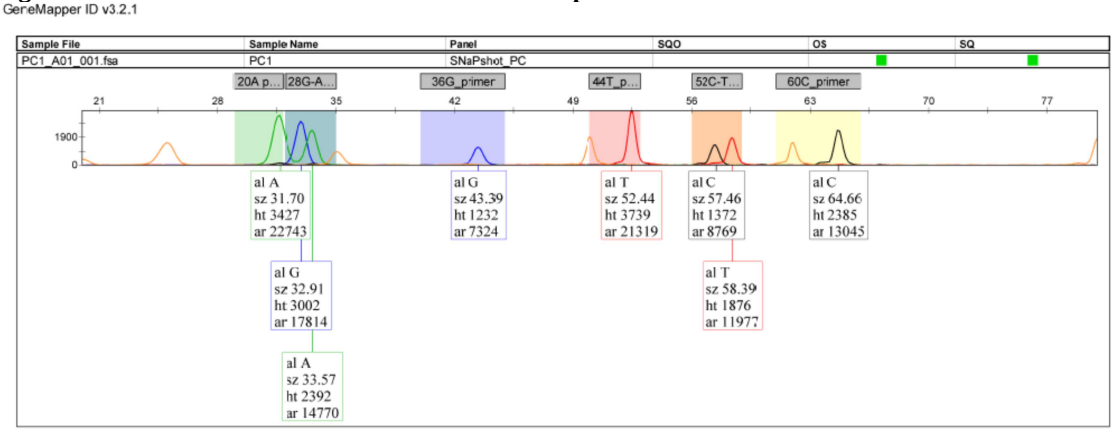
Multiplex Control Primer Mix	Length of Final Products (nt)	Signal Color	Heterozygosity
20A primer	21	Green	Homozygote
28G/A primer	29	Blue/green	Heterozygote
36G primer	37	Blue	Homozygote
44T primer	45	Red	Homozygote
52C/T primer	53	Black/red	Heterozygote
60C primer	61	Black	Homozygote

Figure 37- SNaPshot™ Positive Control (Applied Biosystems, 2000)



Based on the duplicate runs of the SNaPshot™ positive control, individual bins were defined within the GeneMapper™ software for each relative peak or peak set in the positive control, and a bin set was created to be used solely for the analysis of positive and negative SNaPshot™ kit controls. Applying the bin set to the actual data resulted in the data presented in Figure 38, in which the orange peaks represented the GS120-LIZ size standard, and the labeled peaks corresponded to the expected SNaPshot™ positive control SNP profile.

Figure 38- SNaPshot™ Positive Control actual SNP peaks



By running this same bin set against the SNaPshot™ negative control, no peaks were seen in any of the bins, or anywhere else in the data set. Therefore, the criteria for a valid

SNaPshot™ run was determined to correspond to the SNaPshot™ positive control displaying the peaks shown above, and the negative control showing no peak data in the defined bin set.

Once the bin set was defined for the SNaPshot™ controls, it was necessary to define a separate set of bin data for the expected peaks in the HVI mtDNA multiplex. By analyzing the SNP sizes of the individual extension primers on the HL60 mtDNA, each primer was able to be resolved at a specific bp size relative to the GS120-LIZ size standard. The average extension primer peak size for each of the five SNP hotspots is presented in Table 10.

Table 10- Multiplex SNP peaks

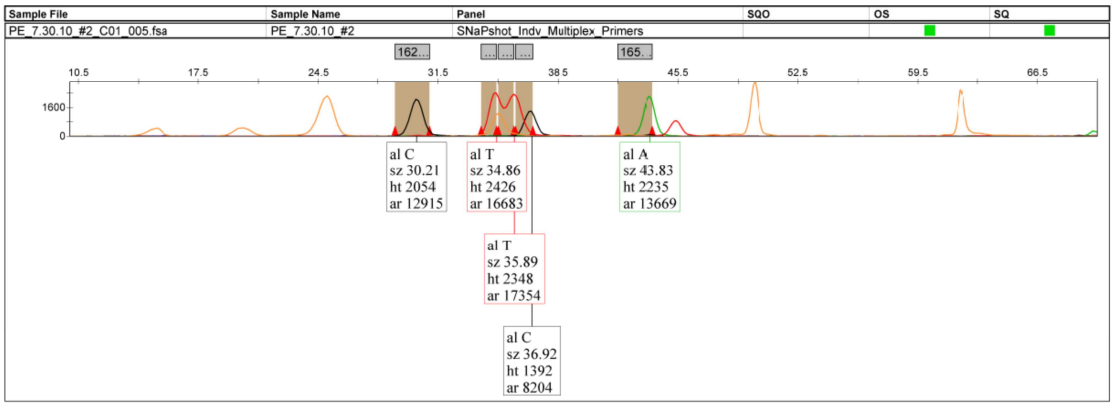
Hotspot	Average Extension Primer Peak Size (in bp)
16223	29.90
16324	34.47
16069	35.37
16390	36.52
16519	43.46

The data for the bp size of each extension primer was based on the redundant data from all eight of the corresponding samples for that primer (four samples run at the four varying DNA/primer input concentrations from the ExoSAP-IT® purification, and four samples run under the same conditions from the ethanol purification). Despite the ethanol data showing lower peak heights for the individual extension primers, the bp sizes were all seen to correspond to the ExoSAP-IT® data, and when analyzed, the standard deviation for size determination for each primer set across all eight samples was determined to be less than or equal to 0.2 bp. The individual HL60 extension primer peaks were then compared to the HL60 multiplex data, and the multiplex peaks were seen to be concordant to the individually run primer peaks for the HL60 mtDNA.

Using the HL60 primer peak and base-size data, two different bin sets were created for the purposes of analysis. One bin set was defined as the entire region from 20 bp to 50 bp, and was set to define any and all peaks (except the orange GS120-LIZ size standard peaks) that appeared within that region above the baseline threshold of 30 RFU. The second bin set defined each separate bin for the five extension primers based on the average sizes of the separate extension primers run against the HL60 mtDNA. The rationale for the two different bins is that heteroplasmy in the extension primers will appear as a different colored peak for the extension primer of interest, but this second peak may not be at the same position as the first peak due to a difference in the motility of the extension primer with a different base, due to the different sizes of the fluorescent dyes attached to each ddNTP terminator base.

Applying the individual bin set to the multiplex positive control (HL60 mtDNA), the profile shown in Figure 39 was obtained.

Figure 39- SNaPshot™ Multiplex data for HL60 - Positive Control DNA



The five expected peaks of the SNP assay were seen with the expected bases for the HL60 sequence at the specific hotspot positions; however, they were out of the expected order based on the original extension primer design. The smallest extension primer was designed to be 16069, however this peak was found to be around 35 bp, and the 16223 primer peak was seen to be closer to the 30 bp size. As previously mentioned, it is not unusual for the primers to be

affected by the addition of the dye-tagged ddNTP terminator base, and this affect might also explain the 1 bp separation between the 16324, 16069, and 16390 extension primers as well. It was interesting to note that the order of the SNP primer peaks are the same as was seen in the Agilent 2100 Bioanalyzer data when the multiplex peak data was examined, with 16223 being the smallest peak, followed by the peaks for 16069, 16519, and finally the block amp peak for 16324 and 16390.

Cycle Number Titration and Melting Temperature Titration Studies on Individual Primer Sets

Initial Amplification and Agilent 2100 Bioanalyzer Data

Based on the decision that only the primers for the previously identified heteroplasmic SNP positions would be run on the single cells, it was necessary to find the optimal cycle number and melting temperature for the single primer sets. It was apparent from the multiplex studies that the combination of all primers was influencing the amplification efficiency of the multiplex, and upon removal of key primer sets, the primer efficiency was seen to change. Two primer sets, 16390 and 16519, were optimized for the usage against the single cells. The 16390 primer represents the mid-range primers, which have been seen to generate peaks in the 83-109 bp range. The 16519 primers have been shown to create a longer amplicon, around 136 bp, so they were validated separately. To carry out this validation, HL60 control DNA (0.1 ng / 10 µl) (Collins, 1987) was amplified in a 25 µl reaction using the AmpliTaq Gold® polymerase to establish 45 samples that were amplified with the 16390 primers, and another 45 samples that were amplified using the 16519 primers. Fifteen amplification negative control samples were also established for each of the 16390 and 16519 primer sets. To test for the optimal melting

temperature simultaneously, three thermal cyclers were set up, each with different annealing temperatures, and each set to a total of 46 amplification cycles:

Thermal Cycler #1
Soak at 95 °C for 9 minutes
46 cycles: { 95 °C for 10 seconds; **55** °C for 30 seconds; 72 °C for 30 seconds }
Storage soak at 4 °C indefinitely

Thermal Cycler #2
Soak at 95 °C for 9 minutes
46 cycles: { 95 °C for 10 seconds; **57** °C for 30 seconds; 72 °C for 30 seconds }
Storage soak at 4 °C indefinitely

Thermal Cycler #3
Soak at 95 °C for 9 minutes
46 cycles: { 95 °C for 10 seconds; **59** °C for 30 seconds; 72 °C for 30 seconds }
Storage soak at 4 °C indefinitely

To each thermal cycler, the following samples were added:

- 15 of the HL60 samples w/ 16390 primers
- 5 of the negative control samples w/ 16390 primers
- 15 of the HL60 samples w/ 16519 primers
- 5 of the negative control samples w/ 16519 primers

In order to also test for the optimal cycle number, three of each of the HL60 samples and one of the negative control samples for each primer set was removed as the samples underwent amplification: at 22 cycles, at 28 cycles, at 34 cycles, at 40 cycles, and finally at 46 cycles. The amplified samples were then run on the Agilent 2100 Bioanalyzer (Agilent Technologies, Santa Clara, CA) to determine the quality and quantity of the amplified mtDNA. It was seen that each primer pair successfully amplified the separate 16390 and 16519 SNP regions based on the comparison between the measured amplicon size and the expected amplicon size from the rCRS map. In order to more efficiently view the results, the virtual gel images from the Agilent 2100

Bioanalyzer data were generated below (Figure 40 through Figure 49), with the sample names displayed above each lane of the virtual gel image. The sample name shows the cycle number, melting temperature, and primer set used for each sample. The negative control samples are denoted as "AN" for "Amplification Negative" control, and each AN is associated with the three samples that immediately follow the negative control. The samples amplified with the 16390 primers should have displayed bands around the 93 bp region. The samples amplified with the 16519 primers should have displayed bands around the 136 bp region.

Figure 40- Melting Temp. and Cycle # Titration Studies, Virtual Gel Image #1

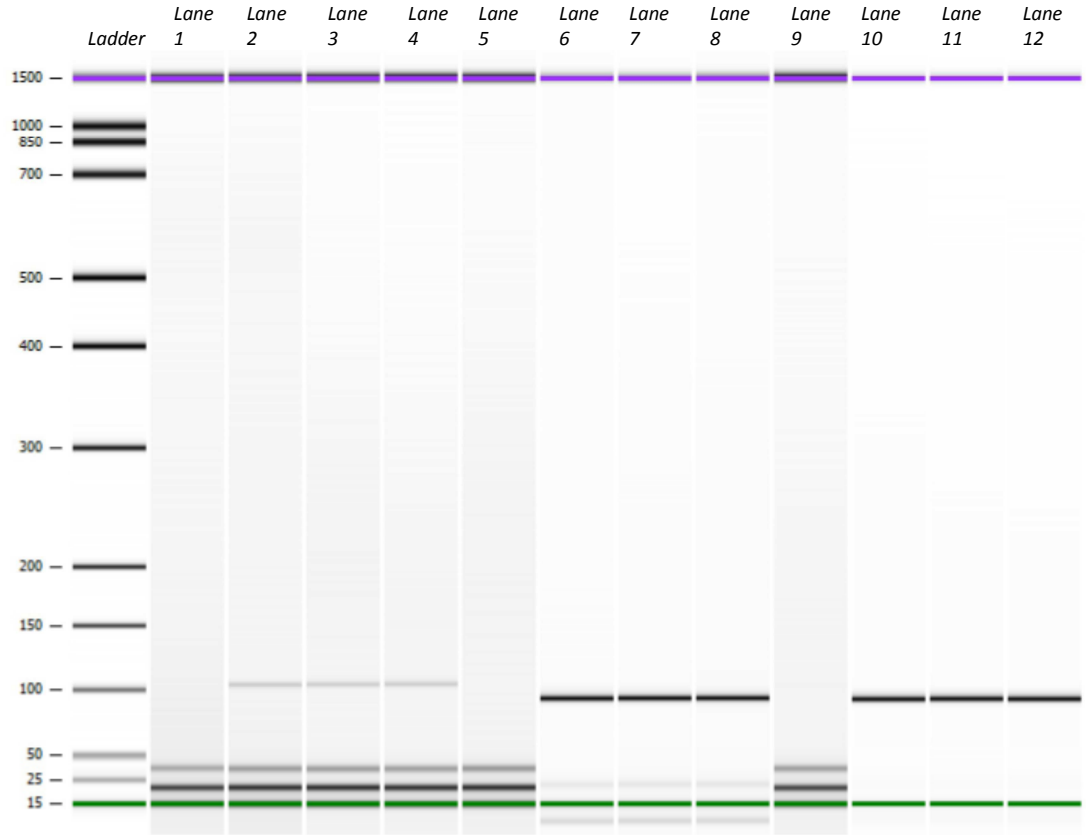


Figure Legend-

Lane 1	AN, 22 cycles, Melting Temp = 55	Lane 7	HL60, 28 cycles, Tm = 55, 16390
Lane 2	HL60, 22 cycles, Tm = 55, 16390	Lane 8	HL60, 28 cycles, Tm = 55, 16390
Lane 3	HL60, 22 cycles, Tm = 55, 16390	Lane 9	AN, 34 cycles, Melting Temp = 55
Lane 4	HL60, 22 cycles, Tm = 55, 16390	Lane 10	HL60, 34 cycles, Tm = 55, 16390
Lane 5	AN, 28 cycles, Melting Temp = 55	Lane 11	HL60, 34 cycles, Tm = 55, 16390
Lane 6	HL60, 28 cycles, Tm = 55, 16390	Lane 12	HL60, 34 cycles, Tm = 55, 16390

The amplification negative control samples (ANs) in Figure 40 displayed no bands at the 16390 amplicon position. At 55 ˚C, the 22-cycle samples were less concentrated than the 28- or 34-cycle samples.

Figure 41- Melting Temp. and Cycle # Titration Studies, Virtual Gel Image #2

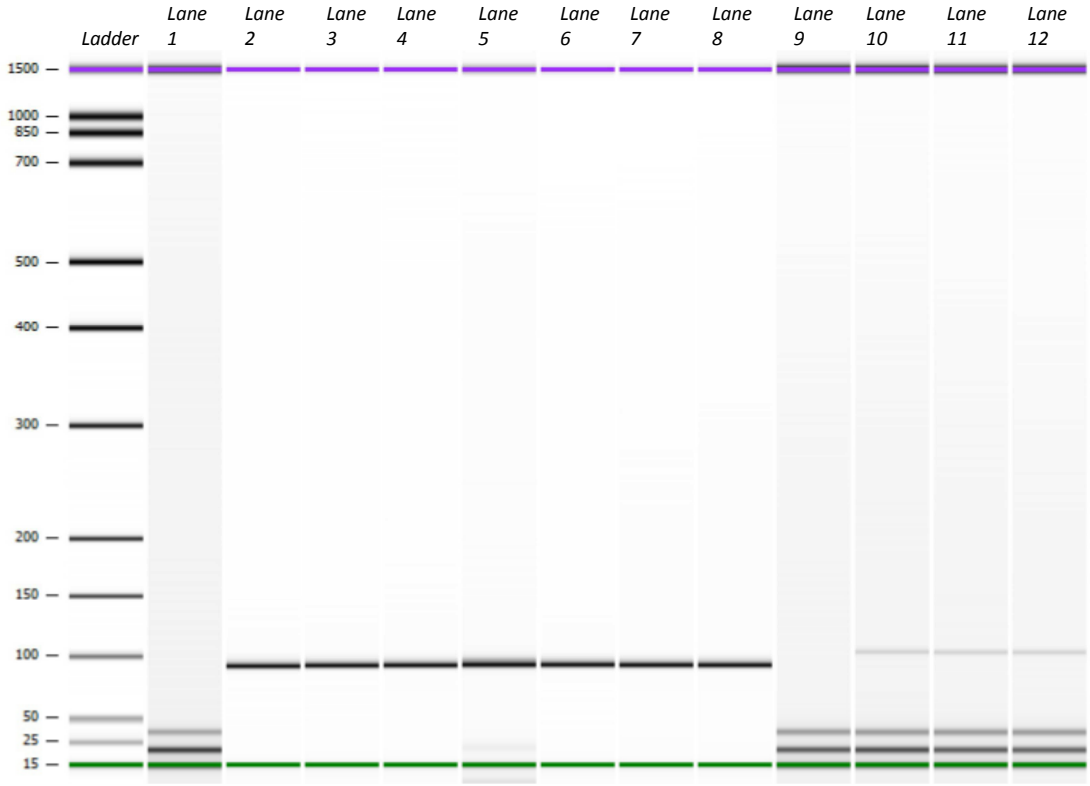


Figure Legend-

Lane 1	AN, 40 cycles, Melting Temp = 55	Lane 7	HL60, 46 cycles, Tm = 55, 16390
Lane 2	HL60, 40 cycles, Tm = 55, 16390	Lane 8	HL60, 46 cycles, Tm = 55, 16390
Lane 3	HL60, 40 cycles, Tm = 55, 16390	Lane 9	AN, 22 cycles, Melting Temp = 57
Lane 4	HL60, 40 cycles, Tm = 55, 16390	Lane 10	HL60, 22 cycles, Tm = 57, 16390
Lane 5	AN, 46 cycles, Melting Temp = 55	Lane 11	HL60, 22 cycles, Tm = 57, 16390
Lane 6	HL60, 46 cycles, Tm = 55, 16390	Lane 12	HL60, 22 cycles, Tm = 57, 16390

At 55 ˚C, the 40-cycle amplification passed, but the 46-cycle amplification failed due to bands at the 16390 position in the amplification negative control, as seen in Figure 41. For the 57 ˚C samples at 22 cycles, the amplification negative control passed but the samples were at a low concentration.

Figure 42- Melting Temp. and Cycle # Titration Studies, Virtual Gel Image #3

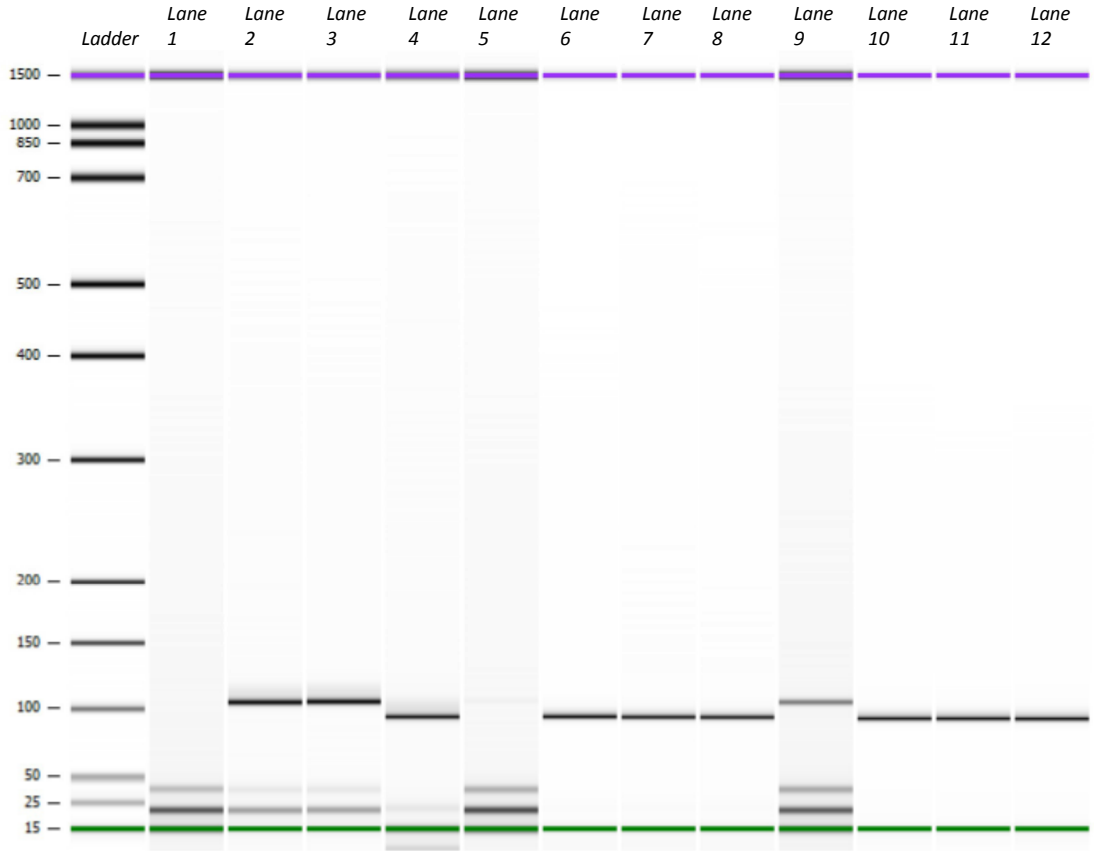


Figure Legend-

Lane 1	AN, 28 cycles, Melting Temp = 57	Lane 7	HL60, 34 cycles, Tm = 57, 16390
Lane 2	HL60, 28 cycles, Tm = 57, 16390	Lane 8	HL60, 34 cycles, Tm = 57, 16390
Lane 3	HL60, 28 cycles, Tm = 57, 16390	Lane 9	AN, 40 cycles, Melting Temp = 57
Lane 4	HL60, 28 cycles, Tm = 57, 16390	Lane 10	HL60, 40 cycles, Tm = 57, 16390
Lane 5	AN, 34 cycles, Melting Temp = 57	Lane 11	HL60, 40 cycles, Tm = 57, 16390
Lane 6	HL60, 34 cycles, Tm = 57, 16390	Lane 12	HL60, 40 cycles, Tm = 57, 16390

At 57 ˚C, the 28- and 34-cycle amplifications passed, but the 40-cycle amplification displayed a band in the 16390 position for the amplification negative control, as seen in Figure 42. Also, the bands in lanes 4 and 9 appeared to have shifted relative to the bands for the similar samples with the same cycle numbers and melting temperatures. Since these shifts resulted in bands that still were within the appropriate range of the amplicon, the bands were determined to be valid.

Figure 43- Melting Temp. and Cycle # Titration Studies, Virtual Gel Image #4

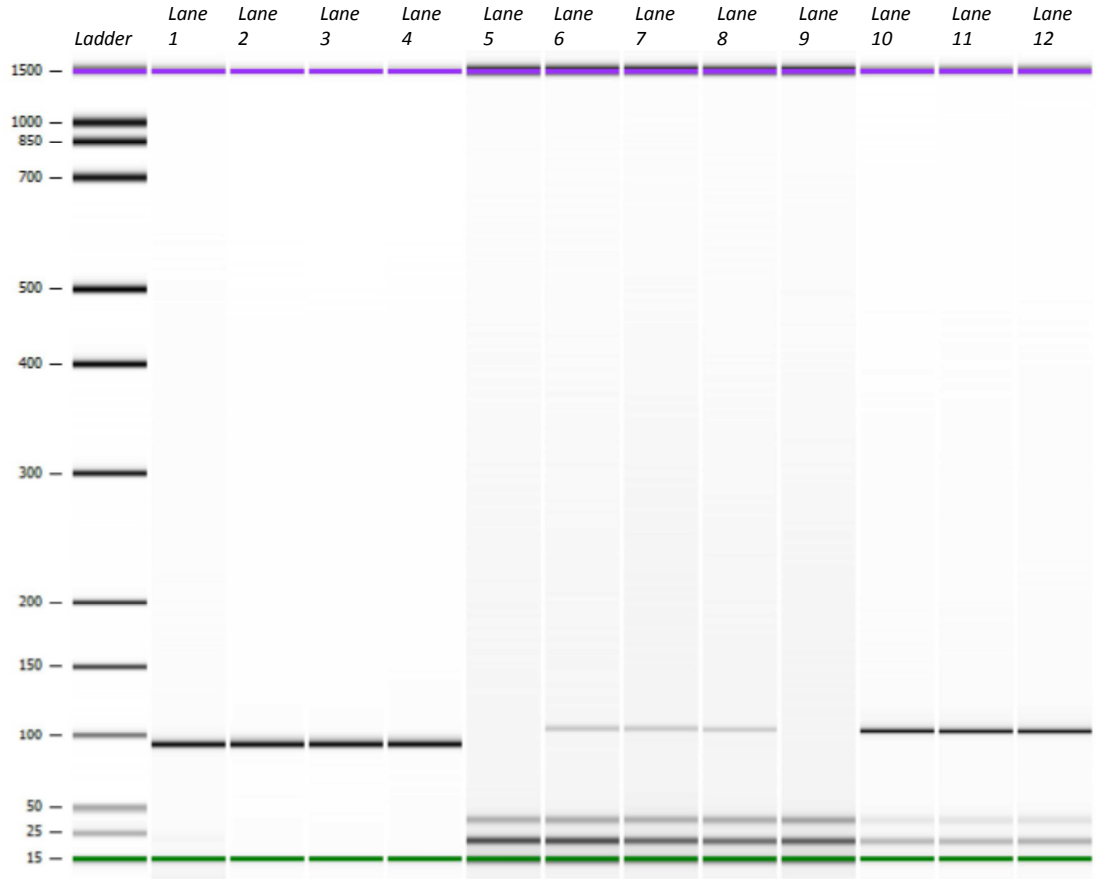


Figure Legend-

Lane 1	AN, 46 cycles, Melting Temp = 57	Lane 7	HL60, 22 cycles, Tm = 59, 16390
Lane 2	HL60, 46 cycles, Tm = 57, 16390	Lane 8	HL60, 22 cycles, Tm = 59, 16390
Lane 3	HL60, 46 cycles, Tm = 57, 16390	Lane 9	AN, 28 cycles, Melting Temp = 59
Lane 4	HL60, 46 cycles, Tm = 57, 16390	Lane 10	HL60, 28 cycles, Tm = 59, 16390
Lane 5	AN, 22 cycles, Melting Temp = 59	Lane 11	HL60, 28 cycles, Tm = 59, 16390
Lane 6	HL60, 22 cycles, Tm = 59, 16390	Lane 12	HL60, 28 cycles, Tm = 59, 16390

As seen in Figure 43, at 57 ˚C, the 46-cycle amplification failed due to bands at the 16390 position in the amplification negative control. For the 59 ˚C samples at 22- and 28-cycles, the amplification negative controls passed but the 22-cycle samples were at a low concentration.

Figure 44- Melting Temp. and Cycle # Titration Studies, Virtual Gel Image #5

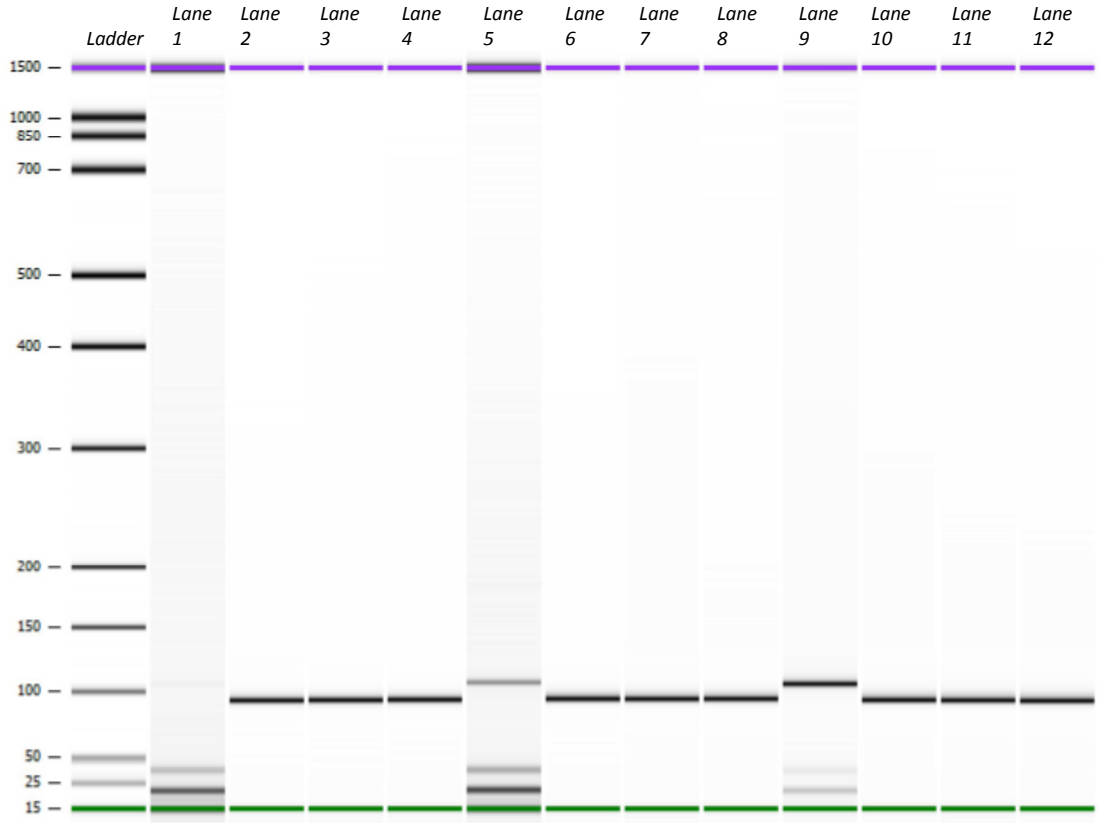


Figure Legend-

Lane 1	AN, 34 cycles, Melting Temp = 59	Lane 7	HL60, 40 cycles, Tm=59, 16390
Lane 2	HL60, 34 cycles, Tm=59, 16390	Lane 8	HL60, 40 cycles, Tm=59, 16390
Lane 3	HL60, 34 cycles, Tm=59, 16390	Lane 9	AN, 46 cycles, Melting Temp = 59
Lane 4	HL60, 34 cycles, Tm=59, 16390	Lane 10	HL60, 46 cycles, Tm=59, 16390
Lane 5	AN, 40 cycles, Melting Temp = 59	Lane 11	HL60, 46 cycles, Tm=59, 16390
Lane 6	HL60, 40 cycles, Tm=59, 16390	Lane 12	HL60, 46 cycles, Tm=59, 16390

At 59 ˚C, the 34-cycle amplifications passed, but the 40- and 46-cycle amplifications displayed bands in the 16390 position for the amplification negative controls, as seen in Figure 44. Also, the amplification negative controls bands appeared to have shifted relative to the other bands within the data set. However, the bands for the amplification negative controls were still within the range of appropriate size for the 16390 amplicon, and were deemed valid.

Figure 45- Melting Temp. and Cycle # Titration Studies, Virtual Gel Image #6

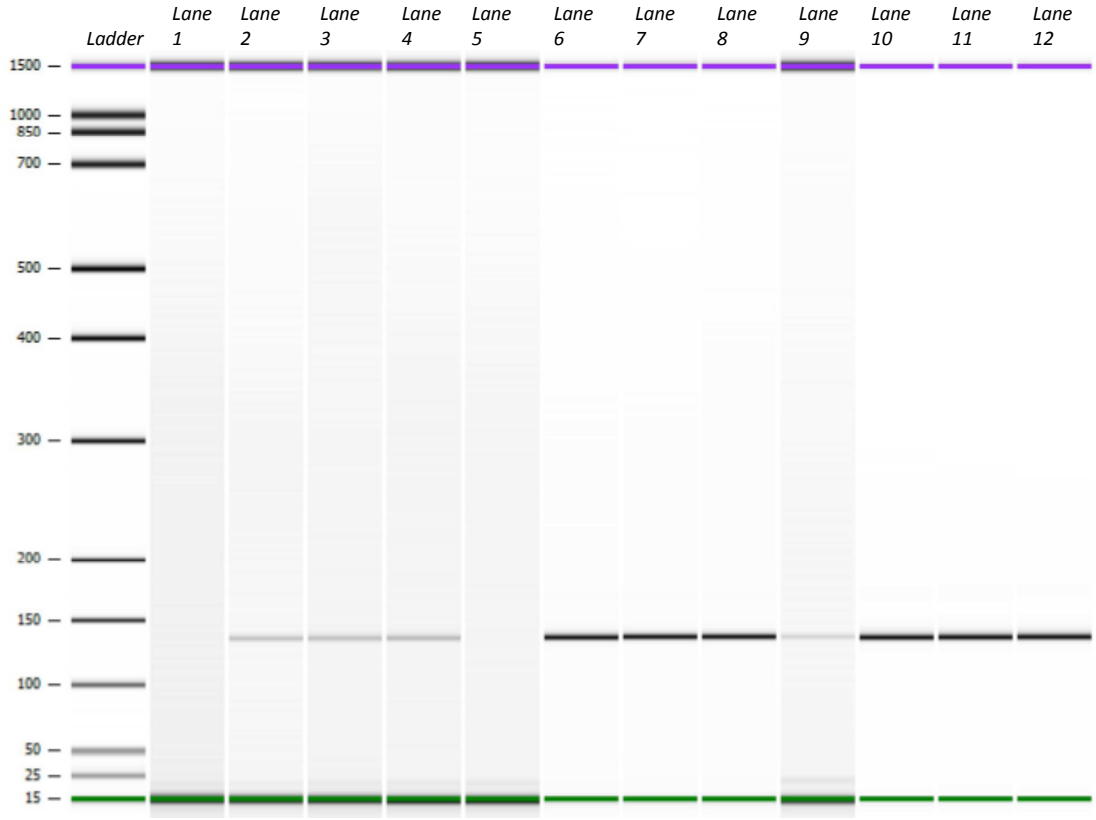


Figure Legend-

Lane 1	AN, 22 cycles, Melting Temp = 55	Lane 7	HL60, 28 cycles, Tm = 55, 16519
Lane 2	HL60, 22 cycles, Tm = 55, 16519	Lane 8	HL60, 28 cycles, Tm = 55, 16519
Lane 3	HL60, 22 cycles, Tm = 55, 16519	Lane 9	AN, 34 cycles, Melting Temp = 55
Lane 4	HL60, 22 cycles, Tm = 55, 16519	Lane 10	HL60, 34 cycles, Tm = 55, 16519
Lane 5	AN, 28 cycles, Melting Temp = 55	Lane 11	HL60, 34 cycles, Tm = 55, 16519
Lane 6	HL60, 28 cycles, Tm = 55, 16519	Lane 12	HL60, 34 cycles, Tm = 55, 16519

At 55 ˚C, the negative control samples for the 22- and 28-cycle amplifications displayed no bands at the 16519 amplicon position as seen in Figure 45. The 22-cycle samples were less concentrated than the 28- or 34-cycle samples. The 34-cycle samples failed, however, due to a band at the 16519 amplicon position in the amplification negative control.

Figure 46- Melting Temp. and Cycle # Titration Studies, Virtual Gel Image #7

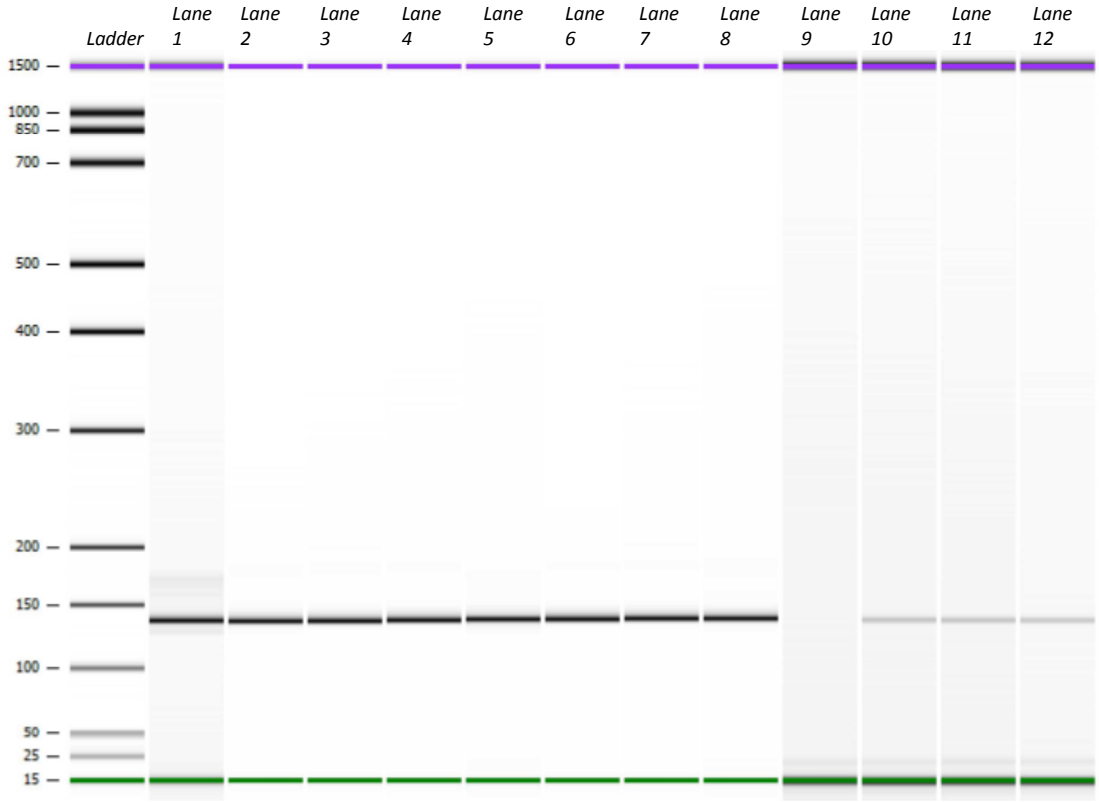


Figure Legend-

Lane 1	AN, 40 cycles, Melting Temp = 55	Lane 7	HL60, 46 cycles, Tm = 55, 16519
Lane 2	HL60, 40 cycles, Tm = 55, 16519	Lane 8	HL60, 46 cycles, Tm = 55, 16519
Lane 3	HL60, 40 cycles, Tm = 55, 16519	Lane 9	AN, 22 cycles, Melting Temp = 57
Lane 4	HL60, 40 cycles, Tm = 55, 16519	Lane 10	HL60, 22 cycles, Tm = 57, 16519
Lane 5	AN, 46 cycles, Melting Temp = 55	Lane 11	HL60, 22 cycles, Tm = 57, 16519
Lane 6	HL60, 46 cycles, Tm = 55, 16519	Lane 12	HL60, 22 cycles, Tm = 57, 16519

As seen in Figure 46, at 55 ˚C, both the 40- and the 46-cycle amplifications failed due to bands at the 16519 position in the amplification negative control. For the 57 ˚C samples at 22 cycles, the amplification negative control passed but the samples were at a low concentration.

Figure 47- Melting Temp. and Cycle # Titration Studies, Virtual Gel Image #8

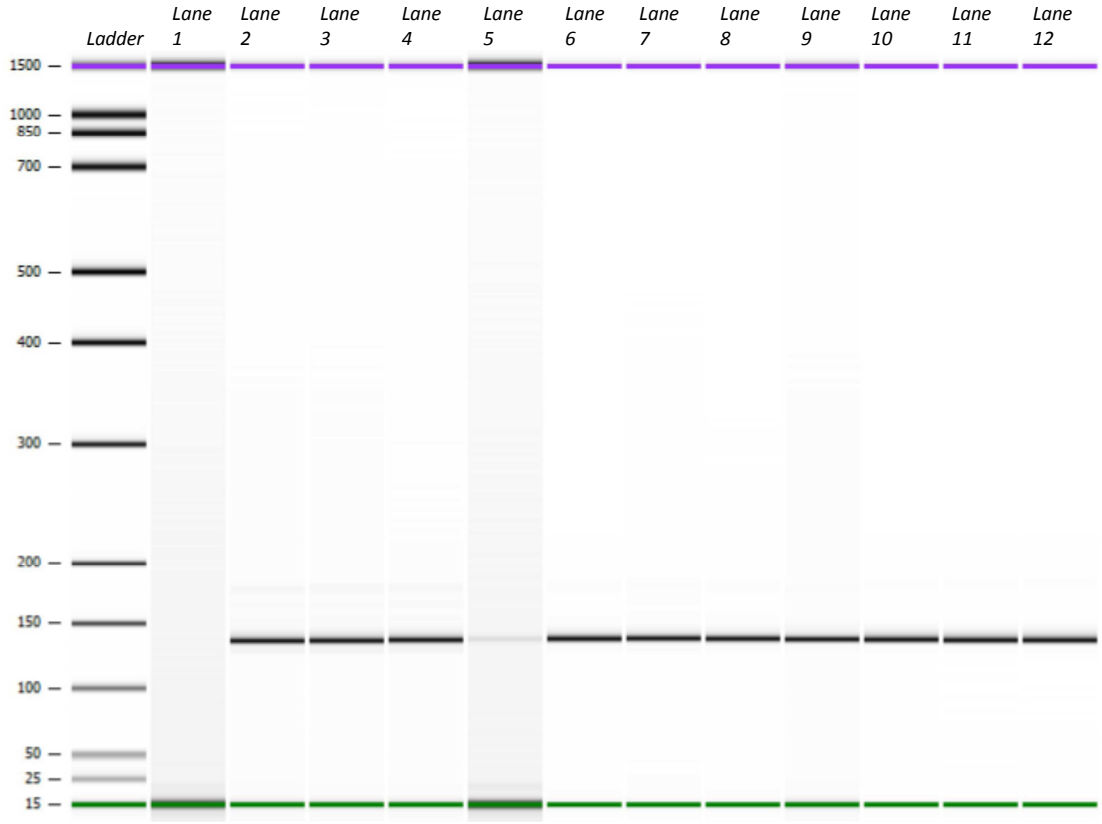


Figure Legend-

Lane 1	AN, 28 cycles, Melting Temp = 57	Lane 7	HL60, 34 cycles, Tm = 57, 16519
Lane 2	HL60, 28 cycles, Tm = 57, 16519	Lane 8	HL60, 34 cycles, Tm = 57, 16519
Lane 3	HL60, 28 cycles, Tm = 57, 16519	Lane 9	AN, 40 cycles, Melting Temp = 57
Lane 4	HL60, 28 cycles, Tm = 57, 16519	Lane 10	HL60, 40 cycles, Tm = 57, 16519
Lane 5	AN, 34 cycles, Melting Temp = 57	Lane 11	HL60, 40 cycles, Tm = 57, 16519
Lane 6	HL60, 34 cycles, Tm = 57, 16519	Lane 12	HL60, 40 cycles, Tm = 57, 16519

At 57 ˚C, the 28-cycle amplification passed, but the 34- and 40-cycle amplifications displayed a band in the 16519 position for the amplification negative control as seen in Figure 47. While the AN displayed a weak band in the 34-cycle amplification negative control, the 40-cycle amplification negative control was equal to the concentration of the associated 40-cycle samples.

Figure 48- Melting Temp. and Cycle # Titration Studies, Virtual Gel Image #9

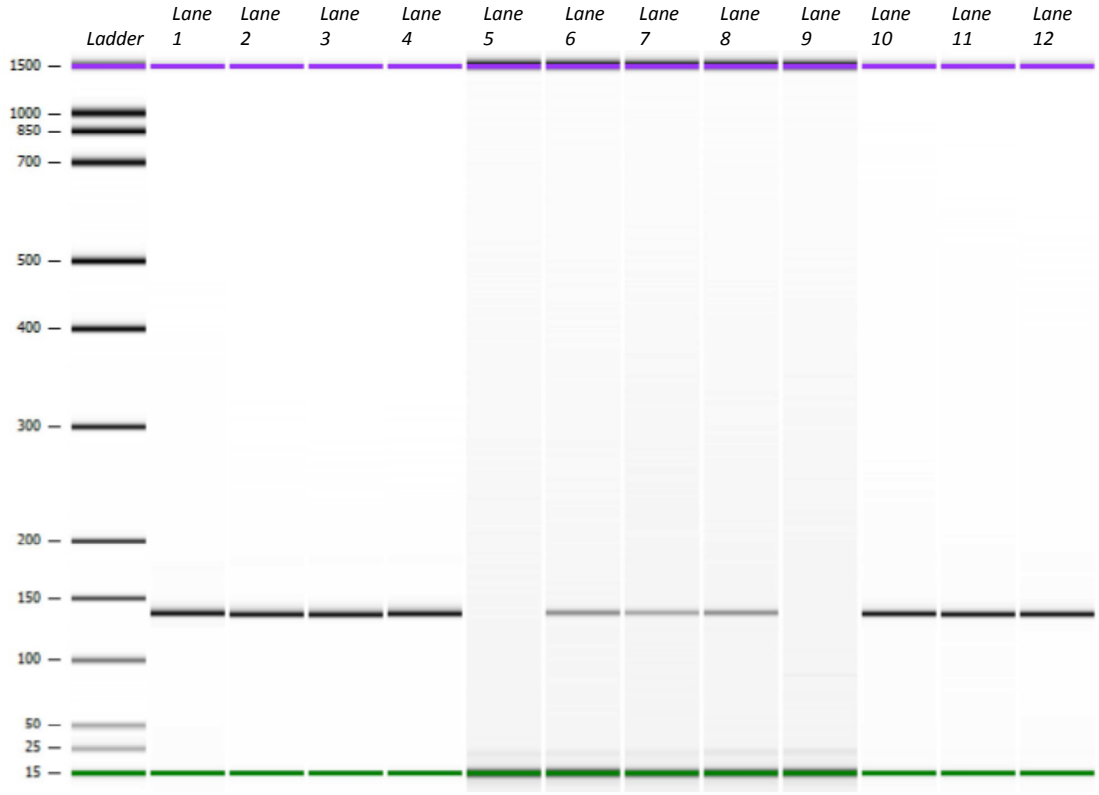


Figure Legend-

Lane 1	AN, 46 cycles, Melting Temp = 57	Lane 7	HL60, 22 cycles, Tm = 59, 16519
Lane 2	HL60, 46 cycles, Tm = 57, 16519	Lane 8	HL60, 22 cycles, Tm = 59, 16519
Lane 3	HL60, 46 cycles, Tm = 57, 16519	Lane 9	AN, 28 cycles, Melting Temp = 59
Lane 4	HL60, 46 cycles, Tm = 57, 16519	Lane 10	HL60, 28 cycles, Tm = 59, 16519
Lane 5	AN, 22 cycles, Melting Temp = 59	Lane 11	HL60, 28 cycles, Tm = 59, 16519
Lane 6	HL60, 22 cycles, Tm = 59, 16519	Lane 12	HL60, 28 cycles, Tm = 59, 16519

At 57 ˚C, the 46-cycle amplification failed due to bands at the 16519 position in the amplification negative control as seen in Figure 48. For the 59 ˚C samples at 22- and 28-cycles, the amplification negative controls passed but the 22-cycle samples were at a low concentration.

Figure 49- Melting Temp. and Cycle # Titration Studies, Virtual Gel Image #10

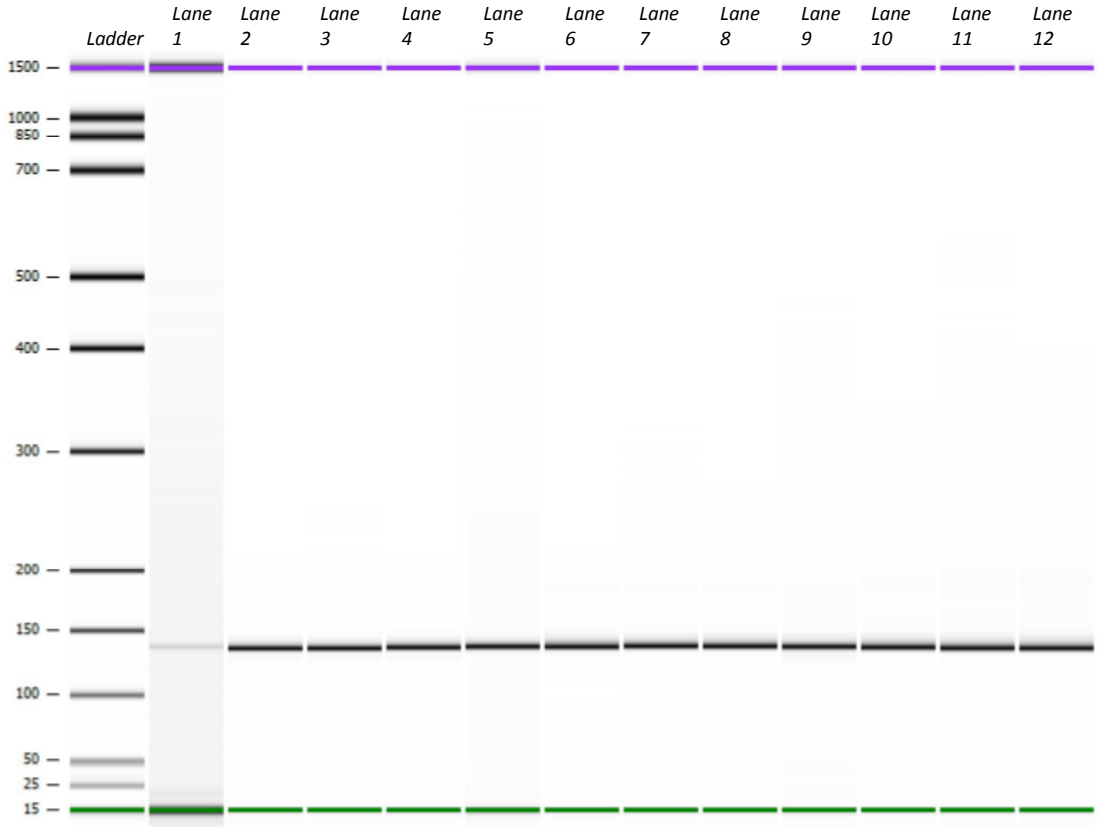


Figure Legend-

Lane 1	AN, 34 cycles, Melting Temp = 59	Lane 7	HL60, 40 cycles, Tm=59, 16519
Lane 2	HL60, 34 cycles, Tm=59, 16519	Lane 8	HL60, 40 cycles, Tm=59, 16519
Lane 3	HL60, 34 cycles, Tm=59, 16519	Lane 9	AN, 46 cycles, Melting Temp = 59
Lane 4	HL60, 34 cycles, Tm=59, 16519	Lane 10	HL60, 46 cycles, Tm=59, 16519
Lane 5	AN, 40 cycles, Melting Temp = 59	Lane 11	HL60, 46 cycles, Tm=59, 16519
Lane 6	HL60, 40 cycles, Tm=59, 16519	Lane 12	HL60, 46 cycles, Tm=59, 16519

At 59 ˚C, the 34-, 40- and 46-cycle amplifications displayed bands in the 16519 position for the amplification negative controls as seen in Figure 49. The AN band was weak at the 34-cycle amplification negative control, but the bands for the 40- and 46-cycle amplification negative controls were equal in concentration to the corresponding 40- and 46-cycle samples.

ExoSAP-IT® Treatment and Secondary Agilent 2100 Bioanalyzer Data

All of the amplified samples were then treated with ExoSAP-IT® using the modified Vallone protocol (Vallone et al., 2004). The samples were re-run on the Agilent 2100 Bioanalyzer to see if there was any loss or change in signal following the ExoSAP-IT® treatment. In order to more efficiently view the results of the ExoSAP-IT® treated samples, the virtual gel images from the Agilent 2100 Bioanalyzer data were again generated (Figure 50 through Figure 59), with the sample names displayed above each lane of the virtual gel image. The sample names show the cycle number, melting temperature, and primer set used for each sample. The negative control samples are denoted as "AN" for "Amplification Negative" control, and each AN is associated with the three samples that immediately follow the negative control. The samples purified with ExoSAP-IT® should have displayed bands around the 93 bp region for the 16390 primers, and around the 136 bp region for the samples amplified with the 16519 primers.

Figure 50- Melting Temp. and Cycle # Post ExoSAP-IT®, Virtual Gel Image #1

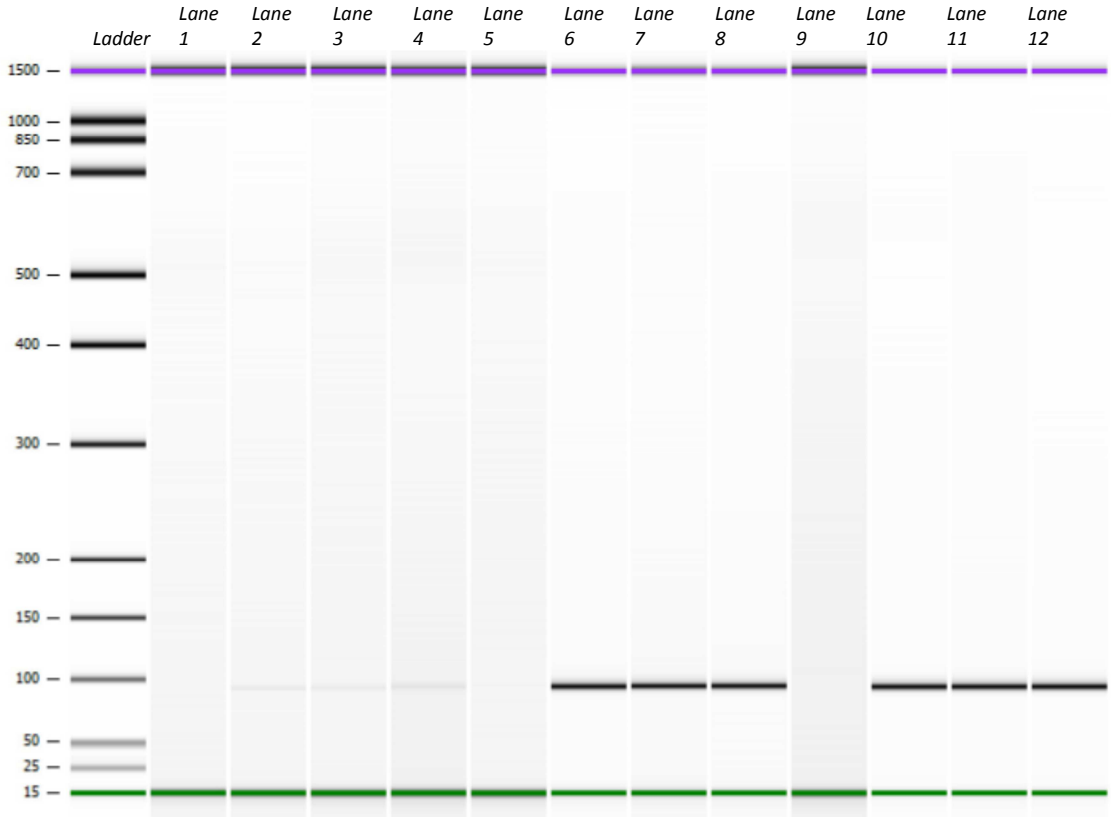


Figure Legend-

Lane 1	AN, 22 cycles, Melting Temp = 55	Lane 7	HL60, 28 cycles, Tm = 55, 16390
Lane 2	HL60, 22 cycles, Tm = 55, 16390	Lane 8	HL60, 28 cycles, Tm = 55, 16390
Lane 3	HL60, 22 cycles, Tm = 55, 16390	Lane 9	AN, 34 cycles, Melting Temp = 55
Lane 4	HL60, 22 cycles, Tm = 55, 16390	Lane 10	HL60, 34 cycles, Tm = 55, 16390
Lane 5	AN, 28 cycles, Melting Temp = 55	Lane 11	HL60, 34 cycles, Tm = 55, 16390
Lane 6	HL60, 28 cycles, Tm = 55, 16390	Lane 12	HL60, 34 cycles, Tm = 55, 16390

The negative control samples (AN) above displayed no bands at the 16390 amplicon position as seen in Figure 50. At 55 ˚C and after ExoSAP-IT® treatment, the 22-cycle samples were less concentrated than the 28- or 34-cycle samples.

Figure 51- Melting Temp. and Cycle # Post ExoSAP-IT®, Virtual Gel Image #2

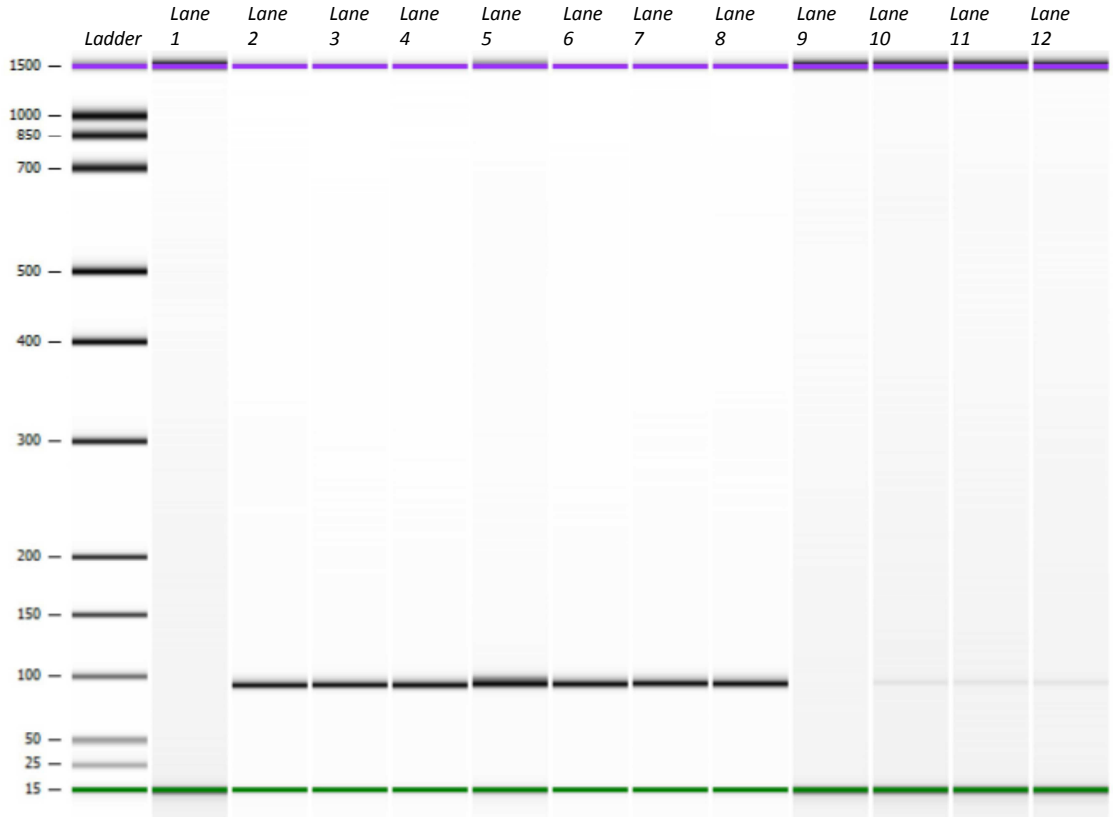


Figure Legend-

Lane 1	AN, 40 cycles, Melting Temp = 55	Lane 7	HL60, 46 cycles, Tm = 55, 16390
Lane 2	HL60, 40 cycles, Tm = 55, 16390	Lane 8	HL60, 46 cycles, Tm = 55, 16390
Lane 3	HL60, 40 cycles, Tm = 55, 16390	Lane 9	AN, 22 cycles, Melting Temp = 57
Lane 4	HL60, 40 cycles, Tm = 55, 16390	Lane 10	HL60, 22 cycles, Tm = 57, 16390
Lane 5	AN, 46 cycles, Melting Temp = 55	Lane 11	HL60, 22 cycles, Tm = 57, 16390
Lane 6	HL60, 46 cycles, Tm = 55, 16390	Lane 12	HL60, 22 cycles, Tm = 57, 16390

As seen in Figure 51, at 55 ˚C and after ExoSAP-IT® treatment, the 40-cycle amplification passed, but the 46-cycle amplification again failed due to bands at the 16390 position in the amplification negative control. For the 57 ˚C samples at 22 cycles after ExoSAP-IT® treatment, the amplification negative control passed but the samples were at a low concentration.

Figure 52- Melting Temp. and Cycle # Post ExoSAP-IT®, Virtual Gel Image #3

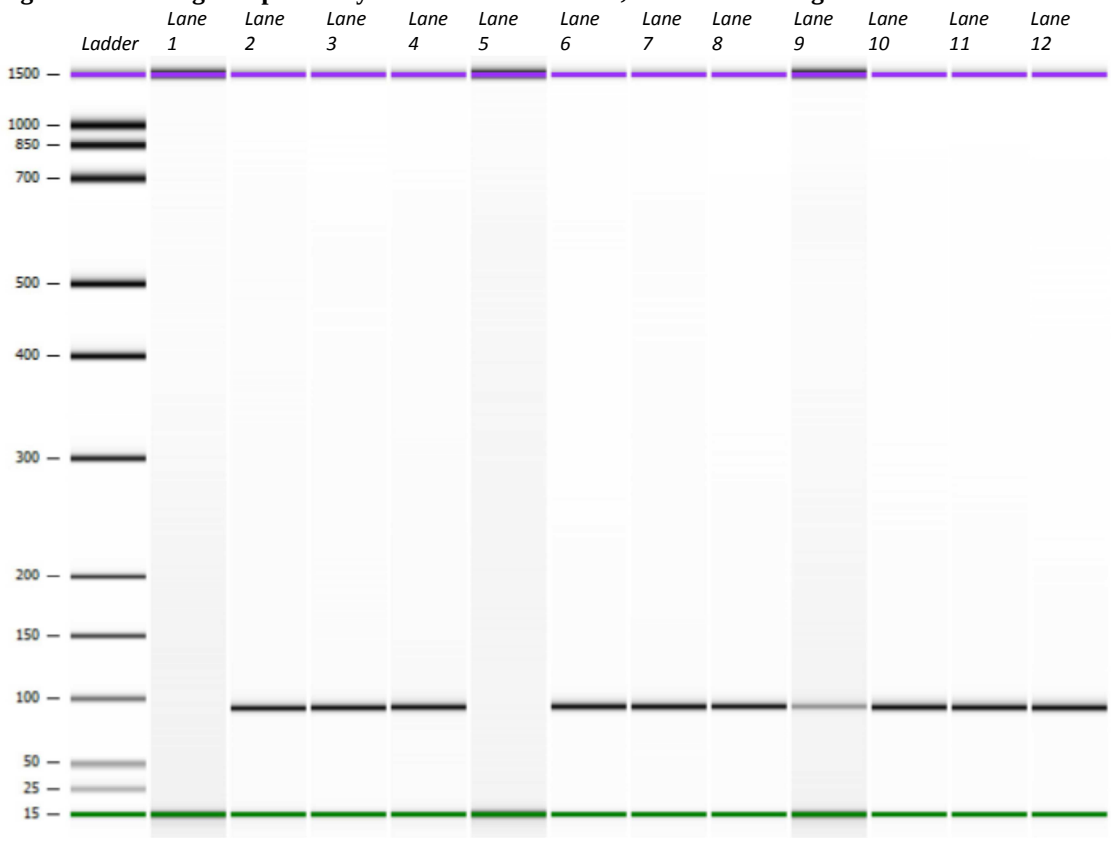


Figure Legend-

Lane 1	AN, 28 cycles, Melting Temp = 57	Lane 7	HL60, 34 cycles, Tm = 57, 16390
Lane 2	HL60, 28 cycles, Tm = 57, 16390	Lane 8	HL60, 34 cycles, Tm = 57, 16390
Lane 3	HL60, 28 cycles, Tm = 57, 16390	Lane 9	AN, 40 cycles, Melting Temp = 57
Lane 4	HL60, 28 cycles, Tm = 57, 16390	Lane 10	HL60, 40 cycles, Tm = 57, 16390
Lane 5	AN, 34 cycles, Melting Temp = 57	Lane 11	HL60, 40 cycles, Tm = 57, 16390
Lane 6	HL60, 34 cycles, Tm = 57, 16390	Lane 12	HL60, 40 cycles, Tm = 57, 16390

At 57 ˚C and after ExoSAP-IT® treatment, the 28- and 34-cycle amplifications passed, but the 40-cycle amplification displayed a band in the 16390 position for the amplification negative control as seen in Figure 52. There were no band shifts observed in the samples compared to the pre-ExoSAP-IT® data presented in Figure 42.

Figure 53- Melting Temp. and Cycle # Post ExoSAP-IT®, Virtual Gel Image #4

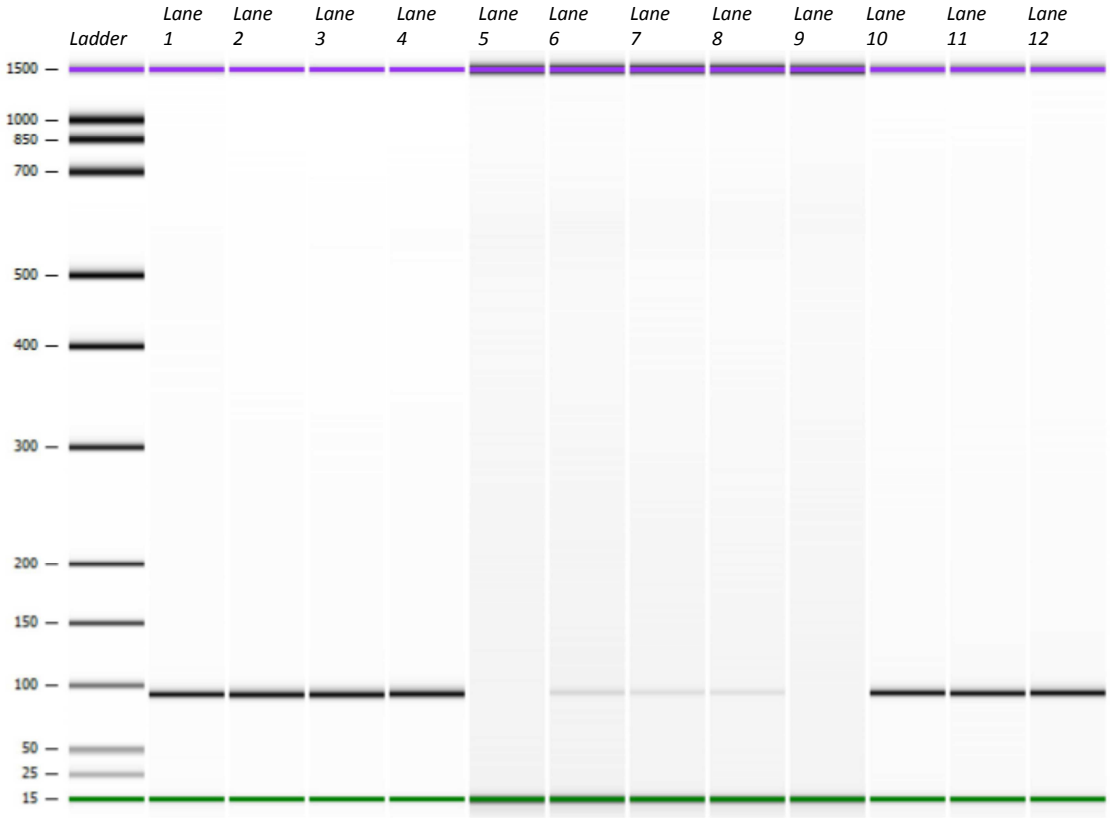


Figure Legend-

Lane 1	AN, 46 cycles, Melting Temp = 57	Lane 7	HL60, 22 cycles, Tm = 59, 16390
Lane 2	HL60, 46 cycles, Tm = 57, 16390	Lane 8	HL60, 22 cycles, Tm = 59, 16390
Lane 3	HL60, 46 cycles, Tm = 57, 16390	Lane 9	AN, 28 cycles, Melting Temp = 59
Lane 4	HL60, 46 cycles, Tm = 57, 16390	Lane 10	HL60, 28 cycles, Tm = 59, 16390
Lane 5	AN, 22 cycles, Melting Temp = 59	Lane 11	HL60, 28 cycles, Tm = 59, 16390
Lane 6	HL60, 22 cycles, Tm = 59, 16390	Lane 12	HL60, 28 cycles, Tm = 59, 16390

At 57 ˚C and after ExoSAP-IT® treatment, the 46-cycle amplification again failed due to bands at the 16390 position in the amplification negative control as seen in Figure 53. For the 59 ˚C samples at 22- and 28-cycles after ExoSAP-IT® treatment, the amplification negative controls passed but the 22-cycle samples were at a low concentration.

Figure 54- Melting Temp. and Cycle # Post ExoSAP-IT®, Virtual Gel Image #5

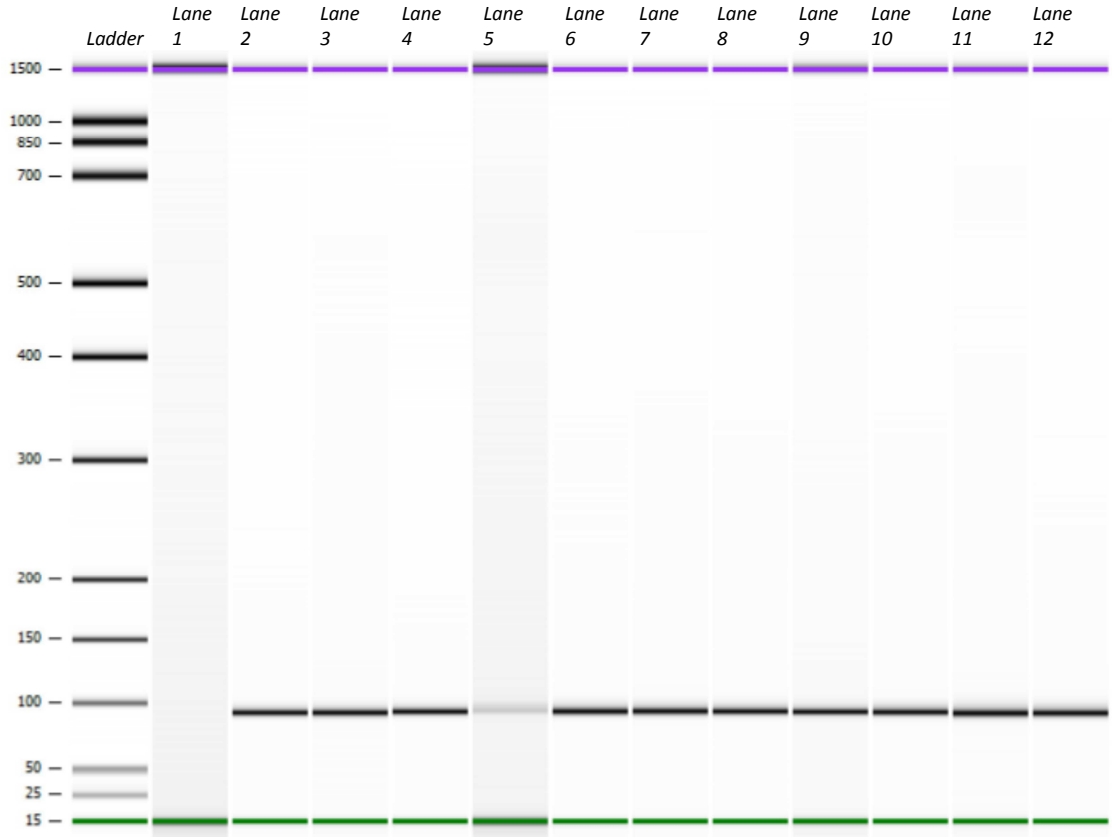


Figure Legend-

Lane 1	AN, 34 cycles, Melting Temp = 59	Lane 7	HL60, 40 cycles, Tm=59, 16390
Lane 2	HL60, 34 cycles, Tm=59, 16390	Lane 8	HL60, 40 cycles, Tm=59, 16390
Lane 3	HL60, 34 cycles, Tm=59, 16390	Lane 9	AN, 46 cycles, Melting Temp = 59
Lane 4	HL60, 34 cycles, Tm=59, 16390	Lane 10	HL60, 46 cycles, Tm=59, 16390
Lane 5	AN, 40 cycles, Melting Temp = 59	Lane 11	HL60, 46 cycles, Tm=59, 16390
Lane 6	HL60, 40 cycles, Tm=59, 16390	Lane 12	HL60, 46 cycles, Tm=59, 16390

As seen in Figure 54, at 59 ˚C and after ExoSAP-IT® treatment, the 34-cycle amplifications passed, but the 40- and 46-cycle amplifications displayed bands in the 16390 position for the amplification negative controls. There were no band shifts observed in the samples compared to the pre-ExoSAP-IT® data presented in Figure 44.

Figure 55- Melting Temp. and Cycle # Post ExoSAP-IT®, Virtual Gel Image #6

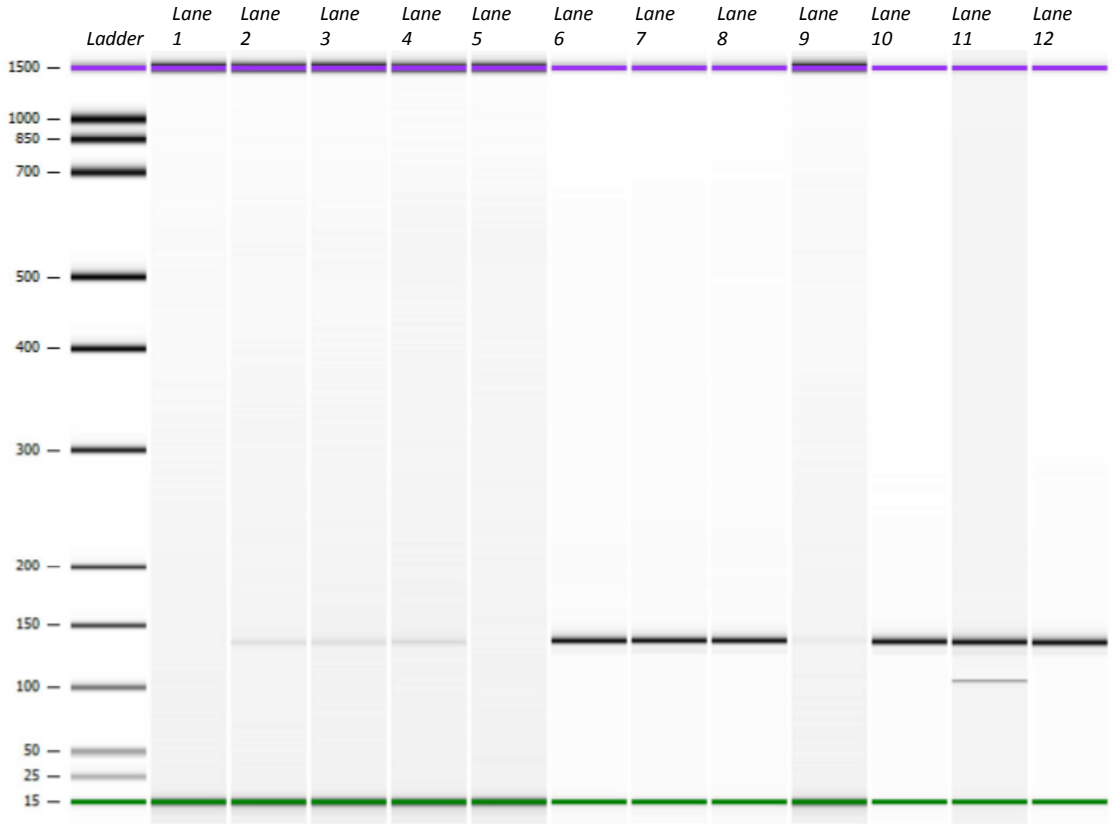


Figure Legend-

Lane 1	AN, 22 cycles, Melting Temp = 55	Lane 7	HL60, 28 cycles, Tm = 55, 16519
Lane 2	HL60, 22 cycles, Tm = 55, 16519	Lane 8	HL60, 28 cycles, Tm = 55, 16519
Lane 3	HL60, 22 cycles, Tm = 55, 16519	Lane 9	AN, 34 cycles, Melting Temp = 55
Lane 4	HL60, 22 cycles, Tm = 55, 16519	Lane 10	HL60, 34 cycles, Tm = 55, 16519
Lane 5	AN, 28 cycles, Melting Temp = 55	Lane 11	HL60, 34 cycles, Tm = 55, 16519
Lane 6	HL60, 28 cycles, Tm = 55, 16519	Lane 12	HL60, 34 cycles, Tm = 55, 16519

At 55 ˚C and after ExoSAP-IT® treatment, the negative control samples for the 22- and 28-cycle amplifications displayed no bands at the 16519 amplicon position as seen in Figure 55. The 22-cycle samples were less concentrated than the 28- or 34-cycle samples. The 34-cycle samples failed after ExoSAP-IT® treatment, however, due to a band at the 16519 amplicon position in the amplification negative control. An additional band was also observed in lane 11, which was not seen in the pre-ExoSAP-IT data presented in Figure 45.

Figure 56- Melting Temp. and Cycle # Post ExoSAP-IT®, Virtual Gel Image #7

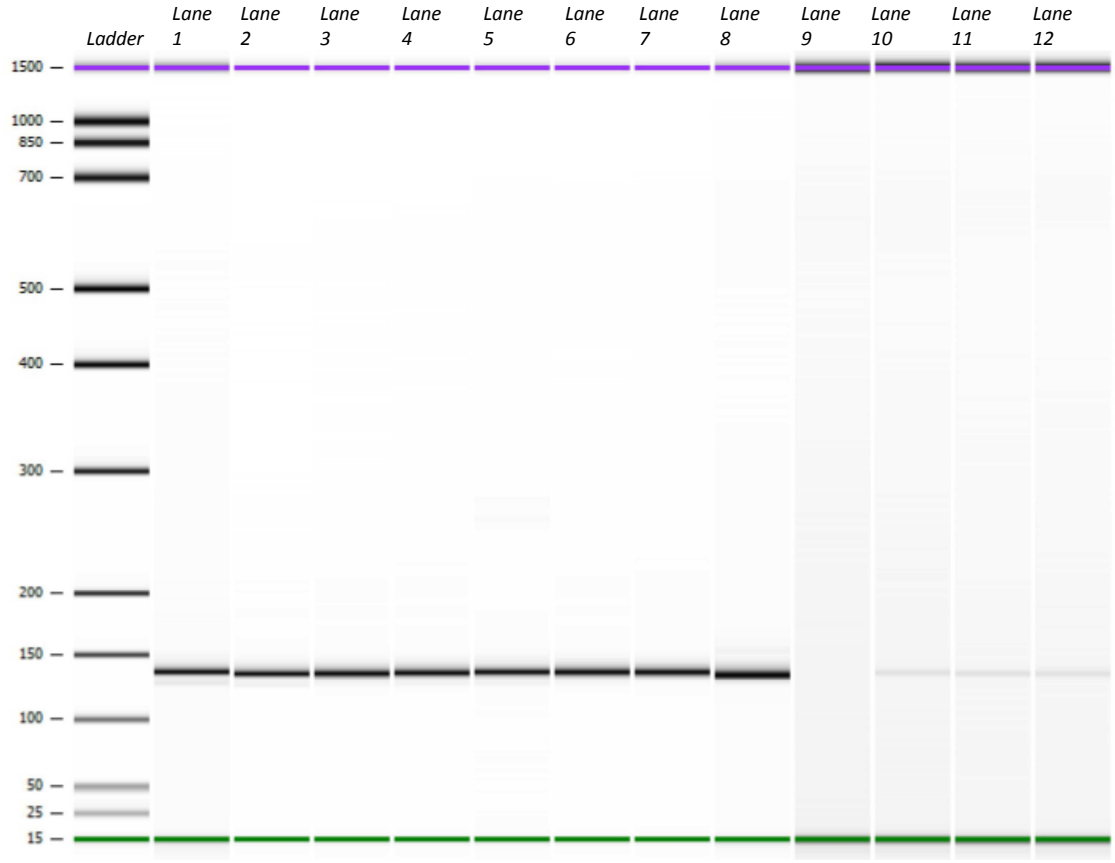


Figure Legend-

Lane 1	AN, 40 cycles, Melting Temp = 55	Lane 7	HL60, 46 cycles, Tm = 55, 16519
Lane 2	HL60, 40 cycles, Tm = 55, 16519	Lane 8	HL60, 46 cycles, Tm = 55, 16519
Lane 3	HL60, 40 cycles, Tm = 55, 16519	Lane 9	AN, 22 cycles, Melting Temp = 57
Lane 4	HL60, 40 cycles, Tm = 55, 16519	Lane 10	HL60, 22 cycles, Tm = 57, 16519
Lane 5	AN, 46 cycles, Melting Temp = 55	Lane 11	HL60, 22 cycles, Tm = 57, 16519
Lane 6	HL60, 46 cycles, Tm = 55, 16519	Lane 12	HL60, 22 cycles, Tm = 57, 16519

As seen in Figure 56, at 55 ˚C and after ExoSAP-IT® treatment, both the 40- and the 46-cycle amplifications again failed due to bands at the 16519 position in the amplification negative control. For the 57 ˚C samples at 22 cycles after ExoSAP-IT® treatment, the amplification negative control passed but the samples were at a low concentration.

Figure 57- Melting Temp. and Cycle # Post ExoSAP-IT®, Virtual Gel Image #8

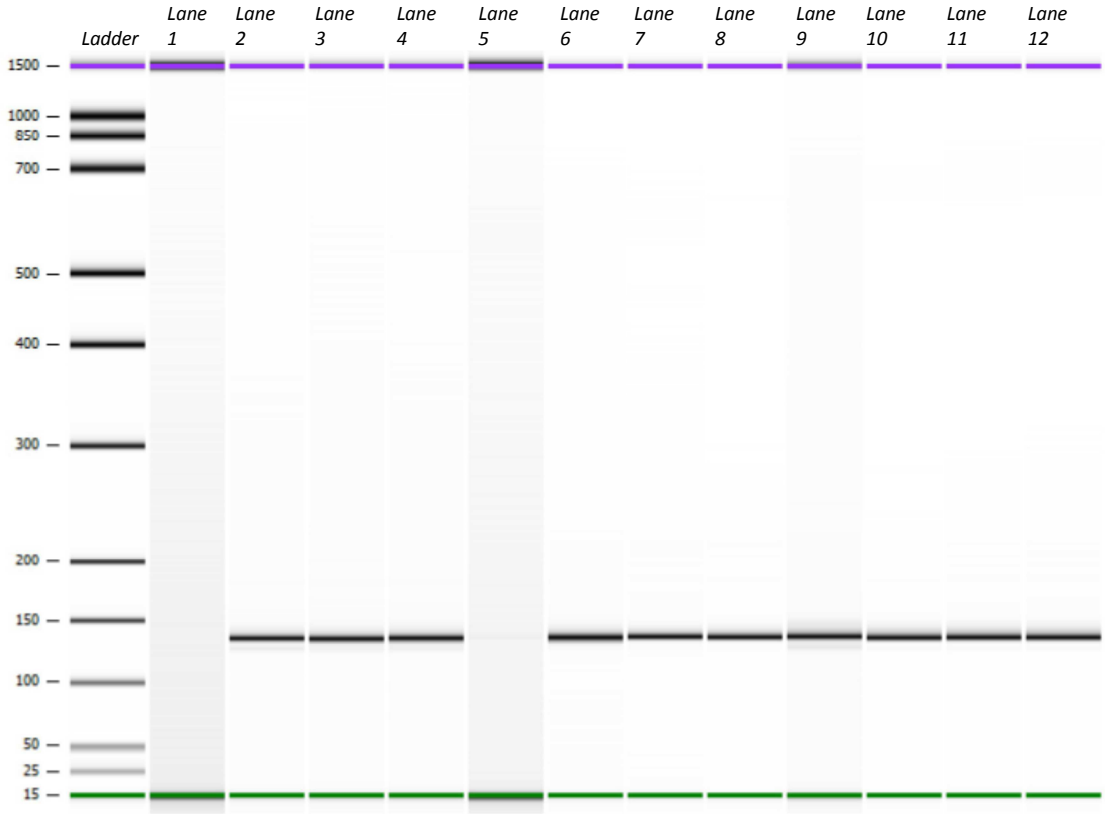


Figure Legend-

Lane 1	AN, 28 cycles, Melting Temp = 57	Lane 7	HL60, 34 cycles, Tm = 57, 16519
Lane 2	HL60, 28 cycles, Tm = 57, 16519	Lane 8	HL60, 34 cycles, Tm = 57, 16519
Lane 3	HL60, 28 cycles, Tm = 57, 16519	Lane 9	AN, 30 cycles, Melting Temp = 57
Lane 4	HL60, 28 cycles, Tm = 57, 16519	Lane 10	HL60, 40 cycles, Tm = 57, 16519
Lane 5	AN, 34 cycles, Melting Temp = 57	Lane 11	HL60, 40 cycles, Tm = 57, 16519
Lane 6	HL60, 34 cycles, Tm = 57, 16519	Lane 12	HL60, 40 cycles, Tm = 57, 16519

At 57 ˚C and after ExoSAP-IT® treatment, the 28- and 34-cycle amplifications passed, while the 40-cycle amplification displayed a band in the 16519 position for the amplification negative control as seen in Figure 57. Prior to the ExoSAP-IT® treatment, the 34-cycle amplification negative showed a weak band in the negative control (Figure 47), but after ExoSAP-IT® treatment the samples now passed.

Figure 58- Melting Temp. and Cycle # Post ExoSAP-IT®, Virtual Gel Image #9

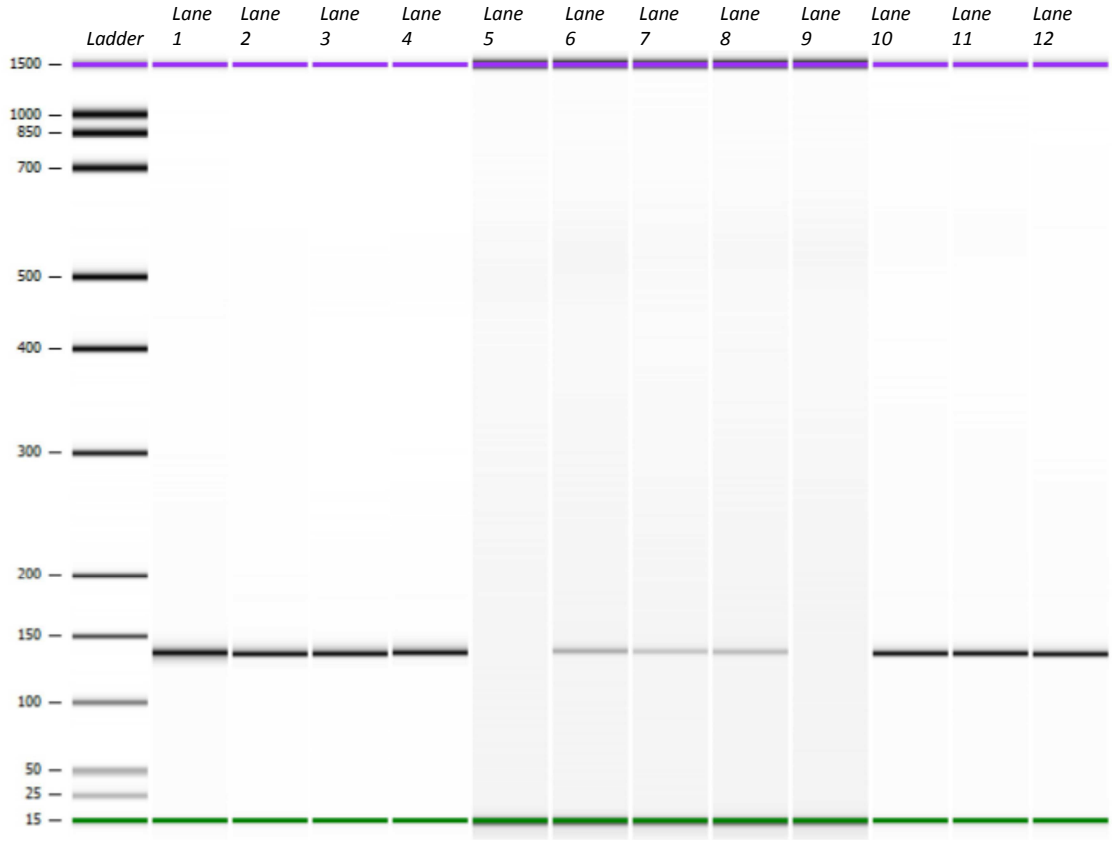


Figure Legend-

Lane 1	AN, 46 cycles, Melting Temp = 57	Lane 7	HL60, 22 cycles, Tm = 59, 16519
Lane 2	HL60, 46 cycles, Tm = 57, 16519	Lane 8	HL60, 22 cycles, Tm = 59, 16519
Lane 3	HL60, 46 cycles, Tm = 57, 16519	Lane 9	AN, 28 cycles, Melting Temp = 59
Lane 4	HL60, 46 cycles, Tm = 57, 16519	Lane 10	HL60, 28 cycles, Tm = 59, 16519
Lane 5	AN, 22 cycles, Melting Temp = 59	Lane 11	HL60, 28 cycles, Tm = 59, 16519
Lane 6	HL60, 22 cycles, Tm = 59, 16519	Lane 12	HL60, 28 cycles, Tm = 59, 16519

At 57 ˚C and after ExoSAP-IT® treatment, the 46-cycle amplification failed due to bands at the 16519 position in the amplification negative control, as seen in Figure 58. For the 59 ˚C samples at 22- and 28-cycles after ExoSAP-IT® treatment, the amplification negative controls passed but the 22-cycle samples were at a low concentration.

Figure 59- Melting Temp. and Cycle # Post ExoSAP-IT®, Virtual Gel Image #10

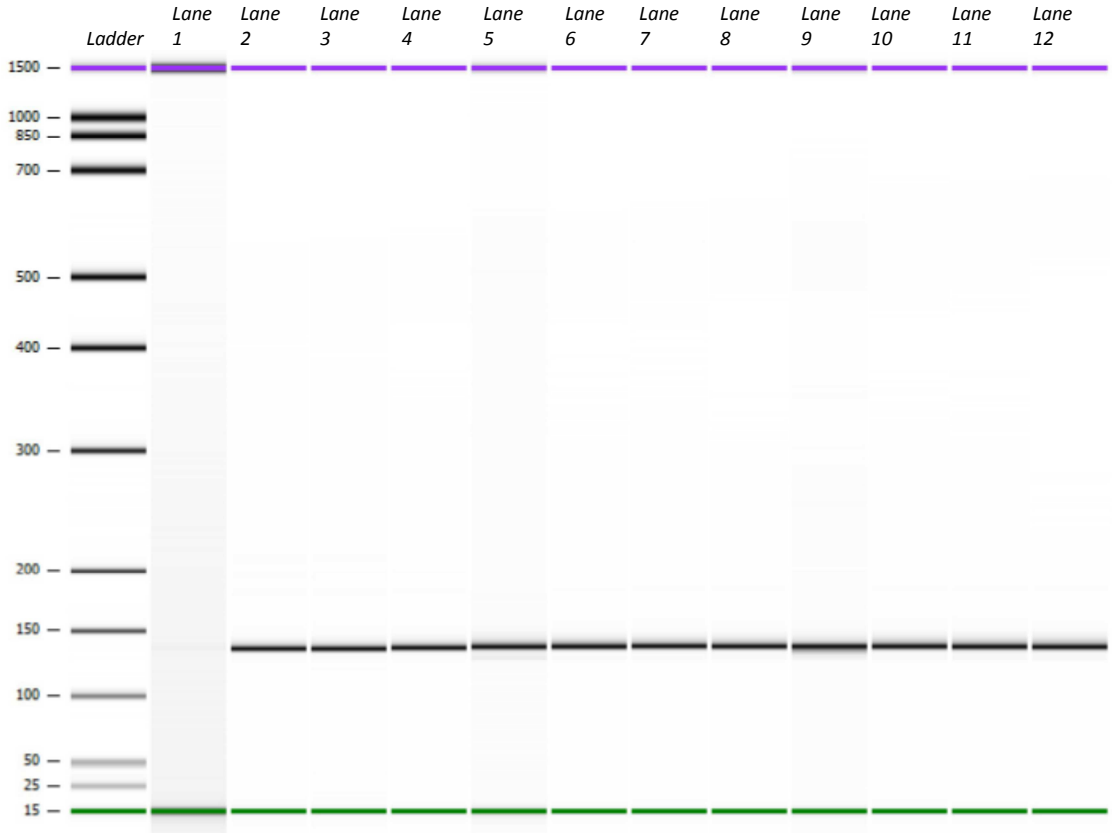


Figure Legend-

Lane 1	AN, 34 cycles, Melting Temp = 59	Lane 7	HL60, 40 cycles, Tm=59, 16519
Lane 2	HL60, 34 cycles, Tm=59, 16519	Lane 8	HL60, 40 cycles, Tm=59, 16519
Lane 3	HL60, 34 cycles, Tm=59, 16519	Lane 9	AN, 46 cycles, Melting Temp = 59
Lane 4	HL60, 34 cycles, Tm=59, 16519	Lane 10	HL60, 46 cycles, Tm=59, 16519
Lane 5	AN, 40 cycles, Melting Temp = 59	Lane 11	HL60, 46 cycles, Tm=59, 16519
Lane 6	HL60, 40 cycles, Tm=59, 16519	Lane 12	HL60, 46 cycles, Tm=59, 16519

As seen in Figure 59, at 59 ˚C and after ExoSAP-IT® treatment, the 34-cycle amplification negative control contained no bands, where prior to ExoSAP-IT® treatment, the sample displayed a band (Figure 49) and had failed. The 40- and 46-cycle amplifications displayed bands in the 16519 position for the amplification negative controls after ExoSAP-IT® treatment.

Based on the data from the amplified samples before and after ExoSAP-IT® treatment, the best results overall for the 16390 primers were seen to occur at 34 cycles and 57 ˚C, seen in Figure 42 and Figure 52. The best results overall for the 16519 primers were seen to occur at 28 cycles and 57 ˚C, also seen in Figure 42 and Figure 52. The data was determined to pass when the amplification negative control was indeed negative (showed no band on the Agilent 2100 Bioanalyzer data for the expected amplicon size) and all three replicates of the HL60 samples were similar in amplicon size and concentration. The next-best results for the 16390 primers, despite contaminated amplification negative controls, were seen in the 40-cycle data for both 57 ˚C and 59 ˚C. The next-best results for the 16519 primers were seen at 34 cycles and 57 ˚C. Based on this data, it was determined that 57 ˚C was the optimal annealing temperature for both of the primer sets, but the primers' efficiency resulted in different amplified DNA concentrations at different cycle numbers for the 16390 primers (and therefore the rest of the primers that generated similarly-sized amplicons) and 16519 primer set (which generated a larger amplicon).

SNaPshot™ Amplification and Analysis

The next step in the validation process was to send the validation samples noted as "best" and "next best" to SNaPshot™ amplification and analysis. Since the samples were HL60 positive control mtDNA, the expected base in HL60 at position 16390 was a "C", and at position 16519 the base was an "A." Based on the prior establishment of the SNaPshot™ sizing map, the 16390 extension primer should resolve around 37 bp, and the 16519 extension primer should resolve around 43 bp in the SNaPshot™ electropherogram. Using the SNaPshot™ assay, the samples outlined in Table 11 and their associated negative controls were amplified with the region-specific extension primers, SAP purified, and run on the 3130*xl* capillary electrophoresis instrument:

Table 11- Cycle # and Tm Titration Samples Sent to SNaPshot™

	16390 primer sets	16519 primer sets
Best Results	34 cycles @ 57 ˚C	28 cycles @ 57 ˚C
Next Best Results	40 cycles @ 57 ˚C	34 cycles @ 57 ˚C
	40 cycles @ 59 ˚C	

Upon analysis, the SNaPshot™ internal positive and negative controls were seen to pass. The amplification negative controls from the titration samples were analyzed first. The results are presented in Table 12:

Table 12- Amplification Negative Controls from Cycle # and Tm Titration

	Titration Sample	SNaPshot™ Results	Peak Height (in RFU)
16390 AN Control	34 cycles @ 57 ˚C	C peak @ ~37 bp	383
	40 cycles @ 57 ˚C	C peak @ ~37 bp	3497
	40 cycles @ 59 ˚C	C peak @ ~37 bp	1155
16519 AN Control	28 cycles @ 57 ˚C	G peak @ ~43 bp	98
	34 cycles @ 57 ˚C	G peak @ ~43 bp	906

The results of the amplification negative controls for the 16519 extension primers in SNaPshot™ generated a minor G peak for both sets, when the assay was screening for an A peak in the HL60 samples. Also, the data for the 16519 samples confirmed that the 28-cycle amplification was ideal, only showing the minor peak in the amplification negative control at 98 RFU, compared to 906 RFU for the 34-cycle negative control sample. However, since neither of the peaks was the expected A peak, the G peaks were believed to be background amplification resulting from the several rounds of nested PCR processing. The data for the 16390 amplification negative controls were interpreted in a slightly different way. Because C peaks were present, which was the same peak at 16390 in HL60, these controls established the negative control threshold that the samples must surpass to be deemed valid. Therefore, the 34-cycle, 57 ˚C samples must show a C peak greater than 383 RFU for that sample to be deemed valid. If any of the samples were at or below the amplification negative threshold, they were deemed

inconclusive, since the data could not be interpreted either as valid or failed due to background contamination. Despite the presence of the C peak in the amplification negative control samples, however, it was clear that the 34-cycle amplification at 57 ˚C was the most successful in terms of the peak height of the C peak. Using the 16390 amplification threshold and the results of the passing 16519 amplification negative controls, the remaining HL60 samples were analyzed, and the results are presented in Table 13:

Table 13- HL60 Replicate Samples From Cycle # and Tm Titration in SNaPshot™

	Titration Sample	SNaPshot™ result	Peak Height (in RFU)	Average Peak Height (in RFU)	Standard Deviation
16390 samples	34 cycles @ 57 ˚C	C peak @ ~37 bp	7016		
	"	C peak @ ~37 bp	6014	6460	509
	"	C peak @ ~37 bp	6351		
	40 cycles @ 57 ˚C	C peak @ ~37 bp	7201		
	"	C peak @ ~37 bp	5816	6650	735
	"	C peak @ ~37 bp	6933		
	40 cycles @ 59 ˚C	C peak @ ~37 bp	5813		
	"	C peak @ ~37 bp	7896	6390	1315
	"	C peak @ ~37 bp	5462		
16519 samples	28 cycles @ 57 ˚C	A peak @ ~43 bp	8190		
	"	A peak @ ~43 bp	8148	8183	32
	"	A peak @ ~43 bp	8212		
	34 cycles @ 57 ˚C	A peak @ ~43 bp	8101		
	"	A peak @ ~43 bp	8107	8087	28
	"	A peak @ ~43 bp	8055		

Based on this data, all of the samples typed correctly in the SNaPshot™ assay using the specific SNP extension primers. Also, all of the 16390 samples were well above the amplification negative thresholds established for the respective negative controls, and were therefore valid. It was established that the 57 ˚C annealing temperature produced the best results in the SNaPshot™ assay, due to the 59 ˚C samples displaying a high standard deviation and lower average RFU when compared to the 57 ˚C samples. Based on the data, it was therefore determined that the optimal annealing temperature for the 16390 and 16519 primers was 57 ˚C,

and the optimization of the amplification cycle number would be further investigated by applying this annealing temperature to single cells using the one-step amplification.

Cycle Number Titration With Test Cells

Isolation and Amplification of the Test Cells

In order to validate the proper cycle number for PCR amplification of the single primers, 30 test cells were collected by randomly selecting five gross tissue samples and capturing six individual cells from each slide, using the established laser dissection methodology. The cells were isolated into amplification tubes containing 15 µl of capture buffer (83.4% dH_2O, 16.6% bovine serum albumin (1.6 µg/µl)), and then the remaining PCR reagents were added along with specific SNP primers. The 16390 primers were added to fifteen cells: three cells from five samples each. The 16519 primers were added to the remaining fifteen cells: three cells from five samples each. The single cells were amplified in the one-step amplification process using the 57 ˚C annealing temperature established in the previous study, but to test for the optimal cycle number, one set of samples were removed at 28 cycles, the second set was removed at 34 cycles, and the final set was run for a total of 40 cycles of PCR amplification. For each set of five samples run with a particular primer at a set cycle number, a positive control HL60 sample and amplification negative control was established.

All of the amplified samples were then treated with ExoSAP-IT® using the modified Vallone protocol (Vallone et al., 2004). The samples were run on the Agilent 2100 Bioanalyzer. In order to more efficiently view the results of the one-step amplification on the single cell samples, the virtual gel images from the Agilent 2100 Bioanalyzer data were generated, with the sample names displayed above each lane of the virtual gel image. The sample name shows the cycle number, melting temperature, and primer set used for each sample. A positive control,

denoted as “PE” was associated with each sample set. The data is presented below (Figure 60 through Figure 63). The negative control samples are denoted as “AN” for “Amplification Negative” control, and are associated with the respective positive control sample and the five samples that immediately follow each negative control. The positive control HL60 samples and the one-step cell samples amplified and purified with ExoSAP-IT® were expected to show bands around the 93 bp region for the 16390 primers, and around the 136 bp region for the samples amplified with the 16519 primers. The Agilent 2100 Bioanalyzer results are as follows:

Figure 60- Cycle # Titration Study on One-Step Single Cells, Virtual Gel Image #1

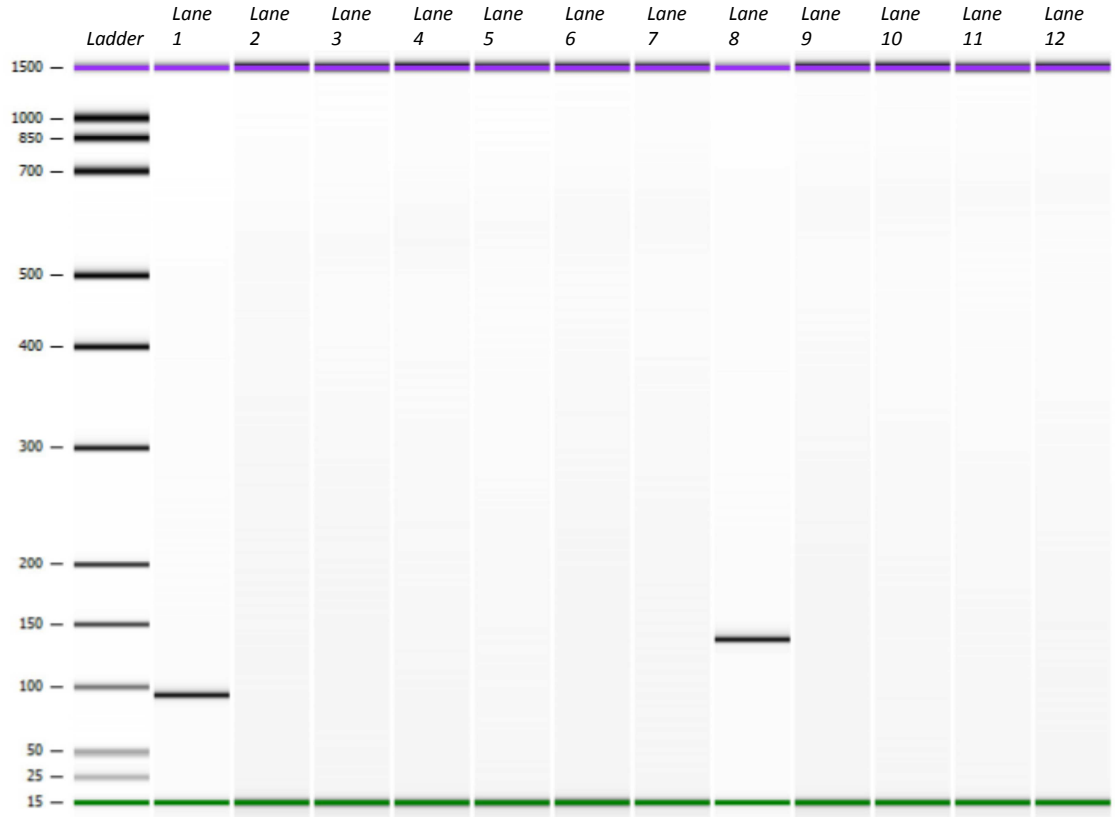


Figure Legend-

Lane 1	PE 16390 - 28 cycles	Lane 7	Tissue 357 16390 - 28 cycles
Lane 2	AN 16390 - 28 cycles	Lane 8	PE 16519 - 28 cycles
Lane 3	Tissue 251 16390 - 28 cycles	Lane 9	AN 16519 - 28 cycles
Lane 4	Tissue 287 16390 - 28 cycles	Lane 10	Tissue 251 16519 - 28 cycles
Lane 5	Tissue 311 16390 - 28 cycles	Lane 11	Tissue 287 16519 - 28 cycles
Lane 6	Tissue 345 16390 - 28 cycles	Lane 12	Tissue 311 16519 - 28 cycles

As seen in Figure 60, the positive control (PE) samples for both the 16390 and 16519 primer sets amplified at 28 cycles produced the proper sized amplicons for the respective primer sets, and the amplification negative (AN) samples did not display any bands, but neither did any of the cells for any of the 16390 or 16519 primers amplified at 28 cycles.

Figure 61- Cycle # Titration Study on One-Step Single Cells, Virtual Gel Image #2

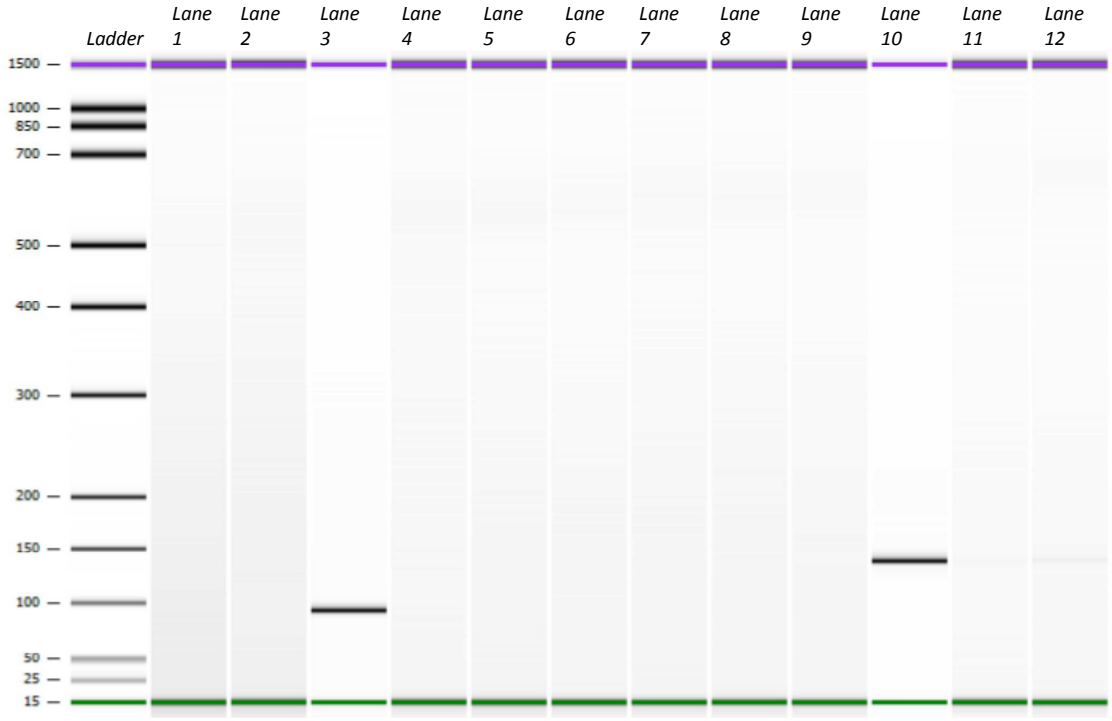


Figure Legend-

Lane 1	Tissue 345 16519 -28 cycles	Lane 7	Tissue 311 16390 -34 cycles
Lane 2	Tissue 357 16519 -28 cycles	Lane 8	Tissue 345 16390 -34 cycles
Lane 3	PE 16390 -34 cycles	Lane 9	Tissue 357 16390 -34 cycles
Lane 4	AN 16390 -34 cycles	Lane 10	PE 16519 -34 cycles
Lane 5	Tissue 251 16390 -34 cycles	Lane 11	AN 16519 -34 cycles
Lane 6	Tissue 287 16390 -34 cycles	Lane 12	Tissue 251 16519 -34 cycles

The remainder of the 28-cycle, 16519 tissue samples are shown in the first two lanes, and did not display any bands, as seen in Figure 61. The positive control (PE) samples for both the 16390 and 16519 primer sets amplified at 34 cycles produced the proper sized amplicons for the respective primer sets, and the amplification negative (AN) samples did not display any bands, but neither did any of the cells for any of the 16390 primers amplified at 34 cycles. A faint band (Lane 12) was visible in the first cell sample amplified with the 16519 primers at 34 cycles.

Figure 62- Cycle # Titration Study on One-Step Single Cells, Virtual Gel Image #3

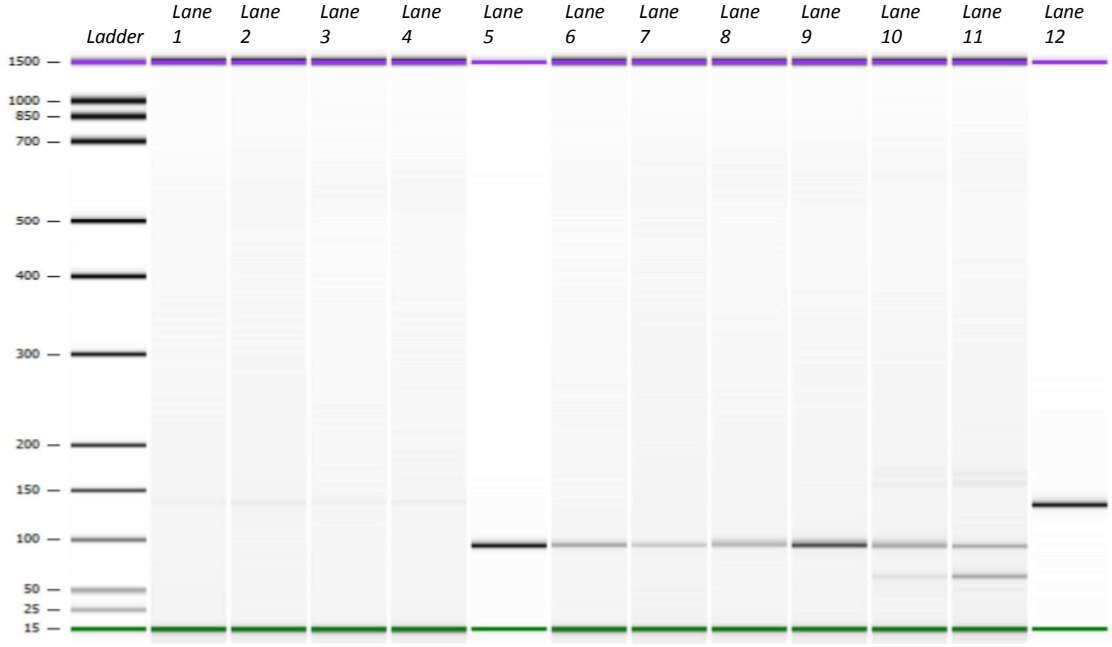


Figure Legend-

Lane 1	Tissue 287 16519 -34 cycles	Lane 7	Tissue 251 16390 -40 cycles
Lane 2	Tissue 311 16519 -34 cycles	Lane 8	Tissue 287 16390 -40 cycles
Lane 3	Tissue 345 16519 -34 cycles	Lane 9	Tissue 311 16390 -40 cycles
Lane 4	Tissue 357 16519 -34 cycles	Lane 10	Tissue 345 16390 -40 cycles
Lane 5	PE 16390 -40 cycles	Lane 11	Tissue 357 16390 -40 cycles
Lane 6	AN 16390 -40 cycles	Lane 12	PE 16519 -40 cycles

As seen in Figure 62, the remainder of the 34-cycle, 16519 tissue samples are shown in the first four lanes, and displayed faint bands consistent with the 16519 primer amplicon. The positive control (PE) samples for both the 16390 and 16519 primer sets amplified at 40 cycles produced the proper sized amplicons for the respective primer sets, but the amplification negative (AN) sample for the 16390 primers did display a band (Lane 6), which indicated contamination at 40 cycles. The remainder of the 40 cycle, 16390 samples were at varying concentrations, and samples 345 and 357 (Lanes 10 & 11) showed additional bands which indicated that non-specific amplification possibly occurred within these cell samples.

Figure 63- Cycle # Titration Study on One-Step Single Cells, Virtual Gel Image #4

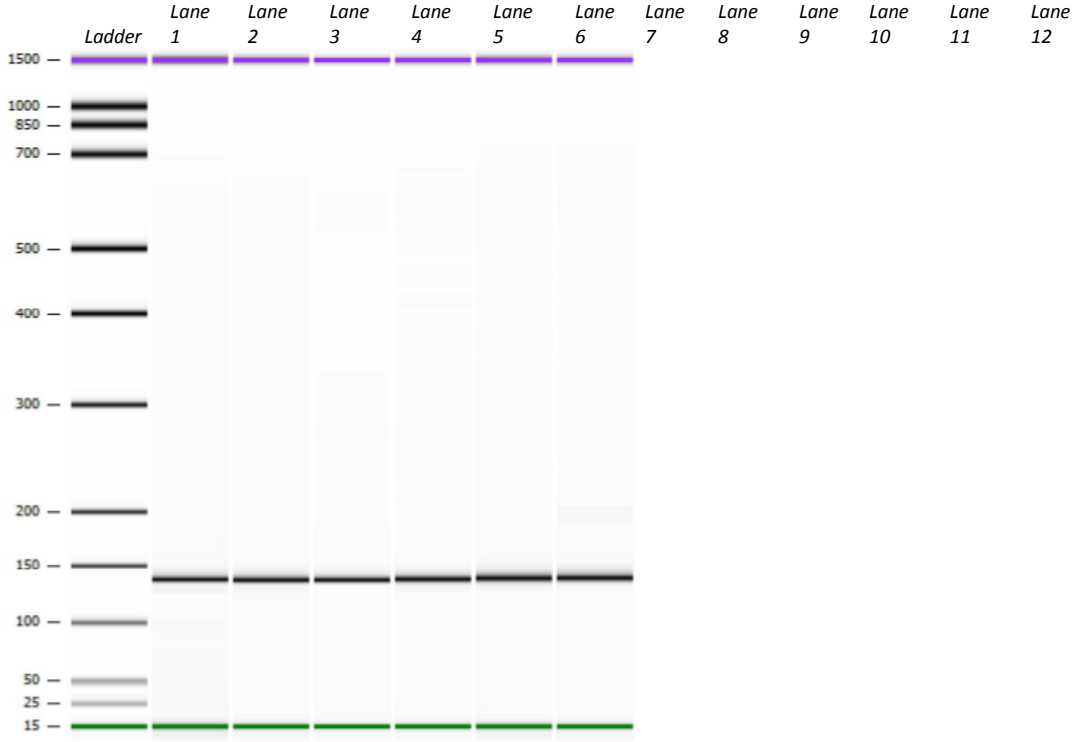


Figure Legend-

Lane 1	AN 16519 -40 cycles	Lane 7	(blank)
Lane 2	Tissue 251 16519 -40 cycles	Lane 8	(blank)
Lane 3	Tissue 287 16519 -40 cycles	Lane 9	(blank)
Lane 4	Tissue 311 16519 -40 cycles	Lane 10	(blank)
Lane 5	Tissue 345 16519 -40 cycles	Lane 11	(blank)
Lane 6	Tissue 357 16519 -40 cycles	Lane 12	(blank)

The remainder of the 40-cycle, 16519 tissue samples are shown in Figure 63, and were all displaying strong bands consistent with the 16519 primer amplicon. The amplification negative (AN) sample for the 40-cycle, 16519 primers displayed a band (Lane 1), which indicated contamination at 40 cycles. Only one sample, tissue 357, displayed additional bands (Lane 6) which indicated that additional non-specific amplification possibly occurred within these cell samples.

SNaPshot™ Analysis of the Test Cells

All of the test cell samples were then sent to SNaPshot™ amplification, regardless of the status of the amplification negative controls based on the Agilent 2100 Bioanalyzer data. Each sample was amplified using the SNaPshot™ kit and the region-specific extension primer as per the kit guidelines. The samples were then SAP purified and run on the 3130*xl* capillary electrophoresis instrument. Upon analysis, the SNaPshot™ internal positive and negative controls passed. The amplification positive and negative controls from the test cell samples were first analyzed. For reference, the HL60 DNA type at 16390 was a C base, and at 16519 it was an A base. The results are presented in Table 14:

Table 14- Amplification Control Results for Test Cell in SNaPshot™

Amplification Cycle #	Control Sample	Primer Set	SNP Base Detected	Peak Height (in RFU)
28	AN	16390	-N/A-	-N/A-
34	AN	16390	C	208
40	AN	16390	C	7889
28	PE	16390	C	5548
34	PE	16390	C	7327
40	PE	16390	C	6290
28	AN	16519	G	159
34	AN	16519	G	5143
40	AN	16519	G	8670
28	PE	16519	A	8134
34	PE	16519	A	8093
40	PE	16519	A	7953

The positive control samples typed correctly at all of the amplification cycles for both primers. The amplification negative controls were negative for the 16390 samples at 28 cycles, and gave a very low peak height at 34 cycles as well. The 40-cycle amplification negative control for both primer sets, as well as the 34-cycle amplification negative control sample for the 16519 primers were equal and/or greater in peak height to the relative positive control samples. Establishing an amplification negative threshold for these samples would have been useless as

some of the positive controls would have failed. Therefore, the 34-cycle amplification negative control was used to set the negative control threshold for the 16390 samples at 208 RFU. The 28-cycle amplification negative control was used to set the negative control threshold for the 16519 samples at 159 RFU. The samples and positive controls would have needed to be higher than these thresholds for the relative primer samples in order for the samples to pass. In the samples above, both of the relative positive controls passed based on these negative control thresholds. It was also noted that the amplification negative controls for the 16519 primers displayed a G base in the data. While this did not correspond to the A base in the HL60 positive control sample, it could have been due to contamination from any of the other test cells in the sample set, or it could have been background amplification, and could not be ignored in either scenario.

The SNaPshot™ results for the test-cell samples were then analyzed. For reference, the 16390 SNP amplicon resolved around 36 bp, and the 16519 SNP amplicon resolved around 42 bp. The tissues that were selected for the cycle-number titration are outlined in Table 15, along with the previously-established SNP profile for the 16390 and 16519 SNP positions:

Table 15- Tissues and SNP profiles for Cycle # Titration Study

Tissue #	16390 SNP	16519 SNP
251	C	A
287	C	G
311	C	G
345	C	A/C
357	C	G

The results of the SNaPshot™ analysis are presented in Table 16 and Table 17:

Table 16- SNaPshot™ Results for Cycle # Titration Study, 16390 primers, test cells

Amplification Cycle #	Test Cell Sample	Primer Set	SNP Base Detected	Peak Height (in RFU)
28	251	16390	C	39
34	251	16390	C	208
40	251	16390	C^{*+}	4668
28	287	16390	C	32
34	287	16390	C^{Δ}	528
40	287	16390	$C^{\Delta+}$	3148
28	311	16390	C^{nc}	below threshold
34	311	16390	C^{Δ}	289
40	311	16390	$C^{\Delta+}$	6603
28	345	16390	C	43
34	345	16390	C^{Δ}	532
40	345	16390	$C^{\Delta+}$	4469
28	357	16390	C	33
34	357	16390	C	670
40	357	16390	C	5640

Table Legend-
* = incorrect position
$^{+}$ = additional minor T peak at ~40 bp
$^{\Delta}$ = additional minor C peak at ~39 bp
^{nc} = peak not automatically called, could not be sized

Table 17- SNaPshot™ Results for Cycle # Titration Study, 16519 primers, test cells

Amplification Cycle #	Test Cell Sample	Primer Set	SNP Base Detected	Peak Height (in RFU)
28	251	16519	G	179
34	251	16519	G^{ob}	8199
40	251	16519	G^{ob}	8485
28	287	16519	G	78
34	287	16519	G	6136
40	287	16519	G^{ob}	8356
28	311	16519	G	135
34	311	16519	G	6908
40	311	16519	G^{ob}	8553
28	345	16519	G	230
34	345	16519	G	3780
40	345	16519	G^{ob}	8524
28	357	16519	G	197
34	357	16519	G	6285
40	357	16519	G^{ob}	8544

Table Legend-
^{ob} = peak overblown past matrix threshold, shows pull-up of other spectral colors

For the 16390 samples, it was clear that the 28-cycle samples were below the negative control threshold, and could not be successfully interpreted. For the 34-cycle samples, the 16390 primers successfully amplified four out of the five samples to give peak heights greater than the negative control threshold, and the fifth sample was exactly at 208 RFU, which was exactly the same as the negative control threshold for the 16390 samples at 34 cycles. As the 40-cycle, 16390 primer amplified samples were analyzed, it was clear that the increased cycle number was showing artifacts in the samples, such as additional minor peaks occurring at 39 and 40 bp, compared to the 37 bp 16390 peak in the electropherogram. Overall, for the 16390 primers, 34-cycles of amplification appeared to generate optimal results above the amplification threshold, with the correct base called for each sample. This same amplification strategy would then apply to the 16069, 16223, and 16324 primers as well, since these all generated similarly-sized amplicons.

For the 16519 samples, the 28-cycle samples appeared to give optimal peak heights for the samples, with three of the five samples showing peak heights above the 159 RFU negative control threshold for 28 cycles. However, the peak-heights were low, and the peak heights that were seen in the samples at 34 cycles were more even. If the amplification negative control for the 34-cycle, 16519 primer samples was established instead as the negative control threshold, the new threshold became 5143 RFU. With this new negative control threshold, four out of the five samples passed analysis; however, one of the samples was overblown and difficult to interpret. This difficulty was due to the fact that when a sample on the 3130*xl* capillary electrophoresis was above the maximum concentration for detection on the instrument, the fluorescent signal overloaded the CCD camera and caused artifacts to appear within the peak, known as pull-ups. This pull-up artifact was due to additional colors that were normally removed by an established

subtraction matrix now being detected in a greater RFU quantity than the matrix standards, and causing the other colored peaks to appear under the overblown peak in the electropherogram. This is a critical problem in SNP analysis, where polymorphisms (or in the specific case of this study, heteroplasmy) are detected as differences of bases at the end of an extension primer, displayed as the mixing of two or more colored bases within a range of the same position.

Overblown samples were simple to identify due a specific shape of the sample peak and also by the peak height, but the presence of the overblown peaks did make SNP analysis impossible for the affected samples. For this reason, it was seen that the 40-cycle samples were un-interpretable for the 16519 primers, and the most optimal and conservative cycle number for the 16519 primers was 28 cycles, leaving the possibility for additional amplification cycles to be performed on the samples if necessary in the future.

An interesting observation from the 28-cycle samples was that despite the presence of no bands in the Agilent 2100 Bioanalyzer data, and no additional cycling of the samples, SNaPshot™ results were obtained for nine of the ten samples, ignoring the negative control thresholds for each primer set. This same phenomenon was also noted for the 34-cycle samples as well, where in the 16390 Agilent 2100 Bioanalyzer data did not contain any bands, and the 16519 Agilent 2100 Bioanalyzer data contained faint bands, the results from the SNaPshot™ analysis were optimal for nine of the ten samples, with the tenth sample actually being overblown. The implication of this observation was that additional cycles of amplification may not be necessary when the single-cell samples appear to be negative on the Agilent 2100 Bioanalyzer. Specifically, the occurrence of the SNaPshot™ data at those specific cycle numbers shows that the single-cell samples were at an amplified concentration appropriate for the SNaPshot™ assay to produce results. Therefore, the lower end of the working detection

range of the Agilent 2100 Bioanalyzer assay was higher than the lowest cutoff concentration for the SNaPshot™ assay. As such, negative Agilent 2100 Bioanalyzer results can and will be sent to SNaPshot™ amplification as long as the relative positive and negative controls pass based on the analysis of the Agilent 2100 Bioanalyzer data.

The final issue with the 16519 primers was that two of the samples, 251 and 345, were previously identified in the gross tissue analysis as possessing an A base at the 16519 position (see Table 15). The SNP analysis of the single cells did not show this A peak, but instead showed a G peak. The G peak was present at each of the cycle number replicates, and increased in peak height as the cycle number increased as well. Based on the analysis of the HL60 positive control DNA, it was verified that the 16519 primers are capable of detecting and identifying an A peak when present, so it was interesting that the cells were uniformly showing a G peak upon analysis. Referring back to the data generated from the gross tissue SNP typing using the multiplex (data not shown), it was seen that the 16519 A in sample 251 had a peak height of 114 RFU, and for the 345 tissue the A peak was at 1351 RFU. While the 251 sample 16519 peak was below the average peak height for the rest of the data, the 345 tissue 16519 A peak was above average, and both peaks are therefore believed to be true values for the gross tissue.

The fact that the single cell samples from the same tissues were giving different values at the 16519 position is a curiosity, but not unexpected or unprecedented. The five positions identified in the HVI mitochondrial region are hotspots for base-change mutational events. The point of this research was screening these hotspots for differences in the bases at these positions. It was previously postulated that a tissue might be displaying a homoplastic base where underlying cells might have a different base, but the culmination of a majority of one type either cancels or overwhelms the minor heteroplasmic base change. This might be what is occurring in

the 16519 position of the test cells. It is possible that within the 251 and 345 tissues, there did exist an A/G heteroplasmy, but the minor G was so minor that when several thousands of cells in the gross tissue were extracted, amplified, and tested, the major A profile overwhelmed the minor G base to the point where it was not even detected on the SNaPshot™ assay. However, when six random cells from 251 and 345 were ablated and tested individually, these six cells displayed the G base as the major, and the A was either minor or not there, due to the fact that it was not detected in the SNaPshot™ assay of the single cells.

While the purpose of this series of validation experiments was designed to establish the optimal cycle number that should be applied to the single primer amplifications for 16390 and 16519, the issues surrounding the profiles uncovered at 16519 in the test cells relate directly to a fundamental issue with the overall project: the detection of heteroplasmy was limited by the sensitivity of the SNaPshot™ assay and the detection threshold of the 3130*xl* capillary electrophoresis machine, as well as the limiting factors on the amplification process such as the reagents, primers, polymerase, and dNTPs. What final determinations were made as to the presence or absence of heteroplasmy in the single cells had to account for these limiting factors and sensitivity thresholds. Also, since the data from the test cells in 16519 showed that one postulated scenario was plausible, then others might have been as well. Since possible low-level heteroplasmy was found in a sample where none was previously believed to exist, it was equally likely that single cells from the known heteroplasmic samples might end up displaying a single homoplastic mitochondrial SNP profile as well.

The end result of the cycle number and melting temperature optimization study, notwithstanding the results of the 16519 primers in the test cells, showed that 57 ˚C was the optimal annealing temperature for both the 16390 and 16519 primer sets, and by association, the

rest of the primer sets as well. For the single-primer amplification set, the 16390 primers were optimized at 34 cycles, which is also extended to the 16069, 16223, and 16324 primer sets as well due to approximately similar amplicon size. The 16519 primers were separately optimized at 28 cycles. These parameters allowed for optimal amplification of the individual primers on the single cells from known heteroplasmic tissues for the purposes of screening the individual cells for mitochondrial heteroplasmy.

Pathology Screening Results

The slides were visually analyzed by Dr. Zoran Budimlija from the Office of Chief Medical Examiner in New York City. The slides were examined under 100 X to 500 X magnification using standard stereomicroscopes under normal Köhler illumination. The cells were noted as either "uniform hepatic cells" or problematic, with varying reasons noted as to the problems seen in the cells. Out of the 109 slides prepared from the tissue blocks, seven samples were identified as not being liver cells and these were removed from the study. An additional twenty-three samples contained cells of questionable pathology. Out of these twenty-three samples, the visual results indicated either "morphologically changed hepatic cells," "non-uniform cells," "questionable cells," or "questionable pathology." These samples were identified, and flagged for additional processing. The original tissue blocks for each of the twenty-three samples were sent for histological cross section preparation, and stained with hematoxylin and eosin (HE) stain. These stained cross sections were re-examined by Dr. Budimlija, and out of the twenty-three slides, three were identified as containing pre-cirrhotic morphology, six were flagged for in-depth microscopic examination, and the remainder of the thirteen slides were classified as normal liver tissue with the exception of sample 275, which was noted as only containing kidney tissue on the cross-sectioned slide. It is possible that the hepatic

tissue was consumed from tissue block 275 prior to the creation of the histological cross-section, but as a precaution, sample 275 was removed from the study altogether. Based on the pathological examinations, a total of seventeen slides were found to have problematic pathology or not to contain liver cells, and these were removed from the study. Therefore, a total of 92 slides were determined to contain uniform hepatic cells that were viable for cell isolation and one-step amplification, contingent upon the results of the gross tissue analysis of the same tissues.

Extraction Negative Control Analysis

For each organic extraction set of the gross tissue remnants, an extraction negative control was generated that accounted for 5 to 10 individual samples. Following the validation of the SNaPshot™ assay and the creation of the analysis protocols, each extraction negative control was amplified at a neat concentration in the multiplex amplification and run at full input volume in the SNaPshot™ assay. The extraction negative controls would only be considered valid if no peaks were present anywhere in the HVI SNP region, using either the individual bins or the multiplex region bin for analysis. The extraction negative controls that were tested accounted for a total of 92 tissue samples, based on the viable tissues identified in the pathology review of the cells. After the initial round of analysis on the extraction negative controls, two separate extraction negative controls failed, each representing five samples. These ten samples were re-extracted, and upon reanalysis, the new extraction negative controls were seen to successfully pass analysis, resulting in a total of 92 viable samples from the gross tissue processing. The corresponding amplification positive and negative controls for the amplification of the extraction negative controls were also analyzed, and were deemed valid. Under the established guidelines for the analysis of the controls, the extraction negative controls accounting for 92 samples were

deemed valid. However, it was noted that one sample, sample #347, was exhausted after the initial extraction and re-extraction process. Since no additional tissue was available for this sample that could serve to confirm future testing, the sample #347 was removed from the study, resulting in 91 viable gross tissue samples for the purposes of the study.

Gross Tissue Sample Analysis

The 91 valid gross tissue extracts were amplified in the multiplex SNP assay, and quantitated on the Agilent 2100 Bioanalyzer. One sample gave no quantitation data following amplification, and four other samples were linked to a contaminated amplification negative control, and were thus deemed invalid and abandoned, leaving 86 viable samples for SNaPshot™ analysis of the gross tissue samples. The 86 samples were then purified with the modified ExoSAP-IT® protocol, and amplified using the SNaPshot™ amplification kit and the SNP extension primers. Following sample reinjections, reruns, and confirmatory amplification runs, a total of 122 run files were generated with SNaPshot™ data. The data from the GeneMapper™ analysis software was exported as comma separated values to an Excel spreadsheet, and a VBA macro was executed to find any samples in which two different SNP peaks fell within a range of amplicon sizes (+/- 4 bp), which would indicate heteroplasmy at that particular SNP position. Several samples were noted as containing mixed peaks within the pre-defined SNP bins, but several samples also had significant peaks outside of the defined bins. When the data was re-analyzed using the multiplex macro instead of the individual bin macro, several samples were seen to contain more than the expected five SNP peaks, with some samples containing up to seven valid peaks. Two specific samples were selected from the gross tissue samples to act as controls for the remainder of the study; samples 288 and 329. Neither of these samples showed

any heteroplasmy in the multiplex SNP assay for the gross tissue, and each sample generated SNP data for all five SNP positions.

Extension Primer Length Study

It was noted in the processing of the gross tissue that several peaks were found outside the previously established bins when the multiplex data was analyzed. Since the analysis bins were based on the peaks of the HL60 mtDNA which accounted for only one peak at each SNP position, the bin sizes did not reflect all of the possible extension primer peaks for each extension primer. It was for this reason that the analysis of the GeneMapper™ data was done with a +/- 4 bp window in the Excel macro, however, this setting was arbitrary based on what was published about the fluorescent dyes attached to the ddNTP bases (Applied-Biosystems, 2000). Since each fluorescent dye had a unique molecular structure that allowed for the difference in fluorescent wavelength, each molecular weight of the dye was also different, and this difference affected the overall mobility of the extension primer when different dyes were attached.

Due to the multiplex setup for this study, and based on the HL60 mtDNA analysis, several of the previously established bins were only 1-3 bp wide, and three of the bins were within 6-8 bp of each other. However, due to the possible range of the extension primers, the data from one primer bin was most likely spilling into other data bins when additional peaks were being detected that were not originally in the HL60 mtDNA SNaPshot™ profile. This led to the interpretation that the "heteroplasmy" at 16390 might not have been heteroplasmy at 16390, but overlapping data from 16069, 16324, and 16390 instead.

When the raw SNP peak data was overlaid from all of the valid gross tissue (see Appendix C for the overlay detail), the pattern of extraneous peaks outside of the previously defined bins was clear, and additional work was therefore required to correctly define the true

sizes of the individual SNP bins for the study. In order to begin investigating the length effects occurring in the SNP data, initial estimations were formed based on established data for the extension primers and the fluorescently dye-labeled ddNTP bases. Using the known molecular weight (MW) of each of these components, calculations were done to estimate the outcomes of each possible base addition to the extension primers. First, the fluorescent 4,7-dichloro-substituted rhodamine (d-Rhodamine) dye groups on the ddNTP bases were outlined as presented in

Table 18:

Table 18- d-Rhodamine Dye Information

ddNTP	Dye Label	Color	MW
A	dR6G	Green	479
C	dTAMRA	Black	916
G	dR110	Blue	380
T(U)	dROX	Red	1035

The molecular weights above are the product of the ddNTP base plus the succinyl ester dye. Using this data, the estimated differences between the 16390E C and the 16390E T primers could be calculated:

16390 extension primer = 14101.28 MW units (g/mol), so

16390E + ddCTP = 15017.28
16390E + ddTTP = 15136.28

This data was empirically confirmed by the SNP data, where in the 16390 extension primer the C normally resolved around 37 bp, and the T peak normally resolved around 40 bp. Examination of the various ddNTP bases without the fluorescent dyes displayed the commonality of these unmodified bases, with an average molecular weight of ~500 g/mol units, as presented in Table 19:

Table 19- Molecular Weights of dNTP bases

dNTP	MW
dATP	488.16
dCTP	463.13
dGTP	504.16
dTTP	479.14
dUTP	465.12

In essence, the commonality of the dNTP bases left only the molecular weight of the fluorescent dye as the true modifier of the extension primers. However, when examining the molecular weight differences between the 16390 extension primers labeled with a C versus a T base, the difference between these constructs was only 119 molecular weight units. At this point the question was posed; is this difference enough to validate the C/T shift seen in the data? The data was reviewed, and it was verified that the T did migrate farther than the C in the electropherogram data, with the T base appearing around 40 bp and the C base appearing around 37 bp. It was also considered that an additional shift might have occurred due to the mobility of the primer through the POP7 polymer, which is viscous and known to affect the mobility of the dye groups. A third consideration for the sizing differences was the secondary structure of the extension primer to the dye group (ionic, or attractive forces in the dye groups to the polar DNA molecule), and vice-versa, the possibility of a 3-dimensional structure of the dye group that would affect the overall mobility of the extension primer.

To address these issues before carrying out the length study, it was deemed necessary to answer the questions surrounding the resolution of the 3130*xl* capillary electrophoresis instrument. Utilizing POP-7 within a 36 cm capillary array should have been able to resolve less than 0.1 bp, since normal DNA sequencing was capable of this same resolution under POP-6 / 50 cm capillary conditions. What was discovered within supplemental documentation from Applied

Biosystems was as follows (ABI SNaPshot™ supplemental documentation for POP-7 on the 3130*xl*):

Figure 64- Primer Focus Data from Applied Biosystems

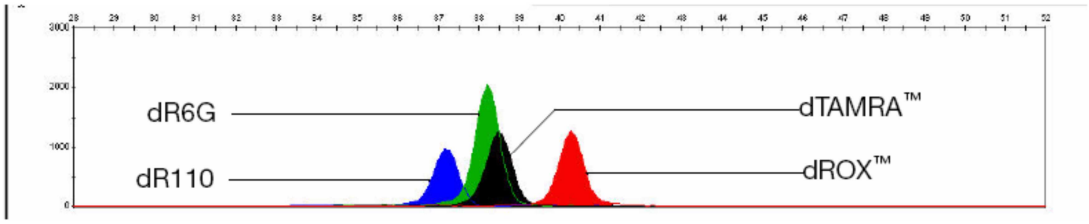


Figure 11 Primer Focus® Kit samples run with POP-7 polymer on a 3130*xl* Genetic Analyzer with a 36-cm capillary array. Mobility of a 32-mer oligonucleotide primer focus product is shown for each of the 4 dyes: dR110 (blue), dR6G (green), dTAMRA™ (yellow) and dROX™ (red).

Figure 64 above shows an example of the separation of the SNaPshot™ dyes, and therefore all of the possible SNP base additions, on a single 32-mer extension primer. What is important to note here is the separation between the C (black) and T (red) peaks for the single 32-mer oligonucleotide product. The same approximate separation of the C and T was seen in the 16390 peak data, with the C peak eluting around 37 bp and the red peak around 40 bp. It was also noted that the order of the primers in Figure 64 was the expected separation pattern that should have occurred based solely on the molecular weight differences of the d-Rhodamine dye groups (see Table 19, above).

Figure 65- Dual Primer Focus Data from Applied Biosystems

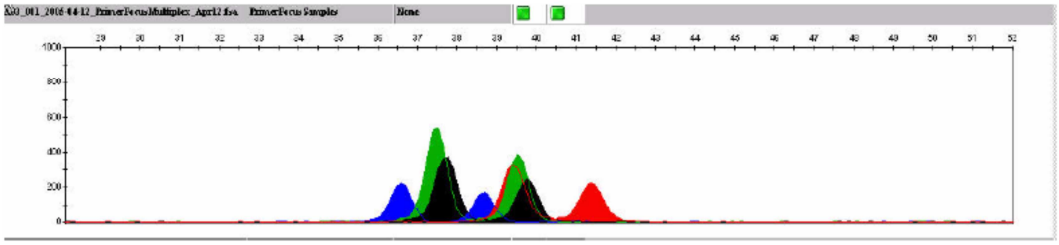


Figure 12 Multiplexed Primer Focus Kit samples run with POP-7 polymer on a 3130*xl* Genetic Analyzer with a 36-cm capillary array. Mobility of two primer focus oligonucleotide products, the 32-mer shown in Figure 11 and a 36-mer when mixed in equal amounts and run in the same capillary.

Figure 65, above, shows an example of the separation of 32-mer and 36-mer oligonucleotide products, showing that the four bp separation is enough to separate all possible SNP dyes for each oligonucleotide product, but at the same time each dye is separated by one or more bp in length. Note that the separation was still evident for each of the primer peaks between each oligonucleotide product. This indicates that the T peaks seen in the SNP data at 40 bp are not heteroplasmic peaks to the C peaks at 37 bp within the 16390 extension primer data, and the seven instances of gross-tissue heteroplasmy were therefore incorrectly identified as such.

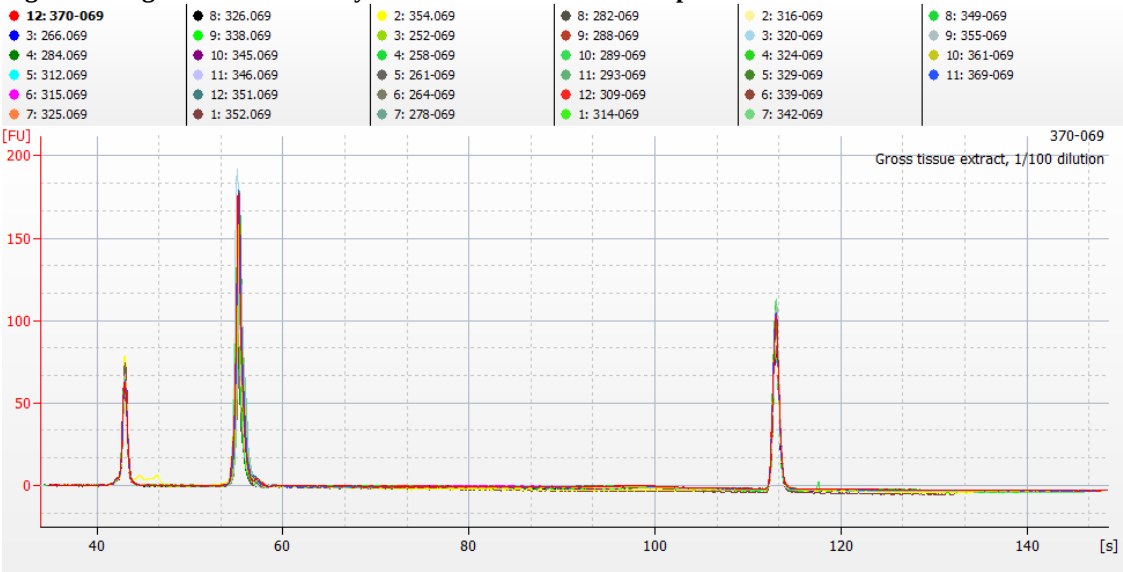
To investigate the issue with the primer length, the individual primer sets were amplified within the gross tissue samples that were previously identified as containing multiple peaks, as well as the control samples 288 and 329. This resulted in 34 samples amplified against the single primer sets. The gross tissue samples were amplified in a 25 µl reaction using the AmpliTaq Gold® polymerase, under the following conditions:

Soak at 95°C for 9 minutes
34 cycles:
- 95°C for 10 seconds
- 57°C for 30 seconds
- 72°C for 30 seconds
Storage soak at 4°C indefinitely

An HL60 positive control and amplification negative control were established for each primer amplification set. Following amplification, the amplified samples were treated with ExoSAP-IT® using the Vallone protocol (Vallone et al., 2004). The amplified samples and controls were run on the Agilent 2100 Bioanalyzer. The positive control for each of the five SNP locations displayed the appropriately-sized bp peak, and the negative control at each of the five SNP locations showed no bands. (Refer to Appendix D for the corresponding Agilent 2100 Bioanalyzer data figures for the positive and negative controls.)

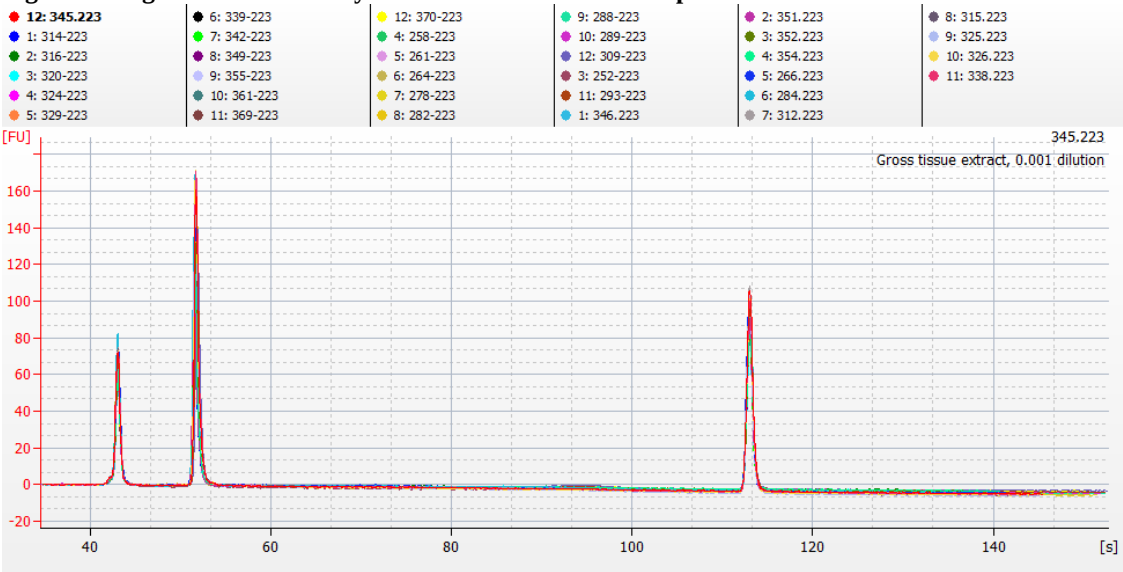
The 34 gross tissue samples were analyzed, and all samples were seen to generate the appropriate 16069 primer peak (see Figure 66 below).

Figure 66- Agilent 2100 Bioanalyzer results for 34 16069 samples



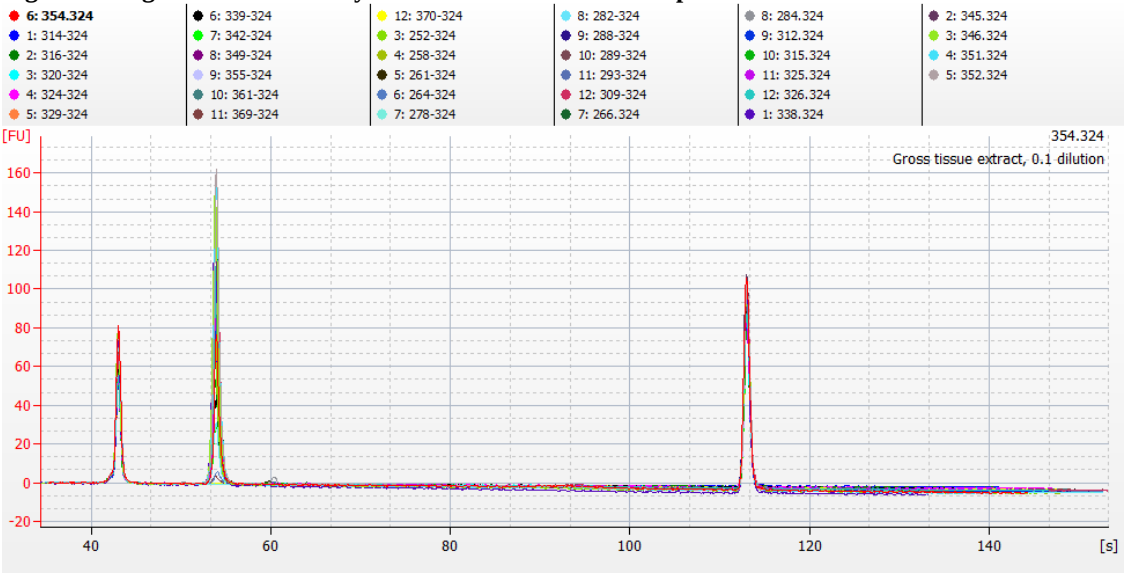
The 34 gross tissue samples were analyzed, and all samples were seen to generate the appropriate 16223 primer peak (see Figure 67 below).

Figure 67- Agilent 2100 Bioanalyzer results for 34 16223 samples



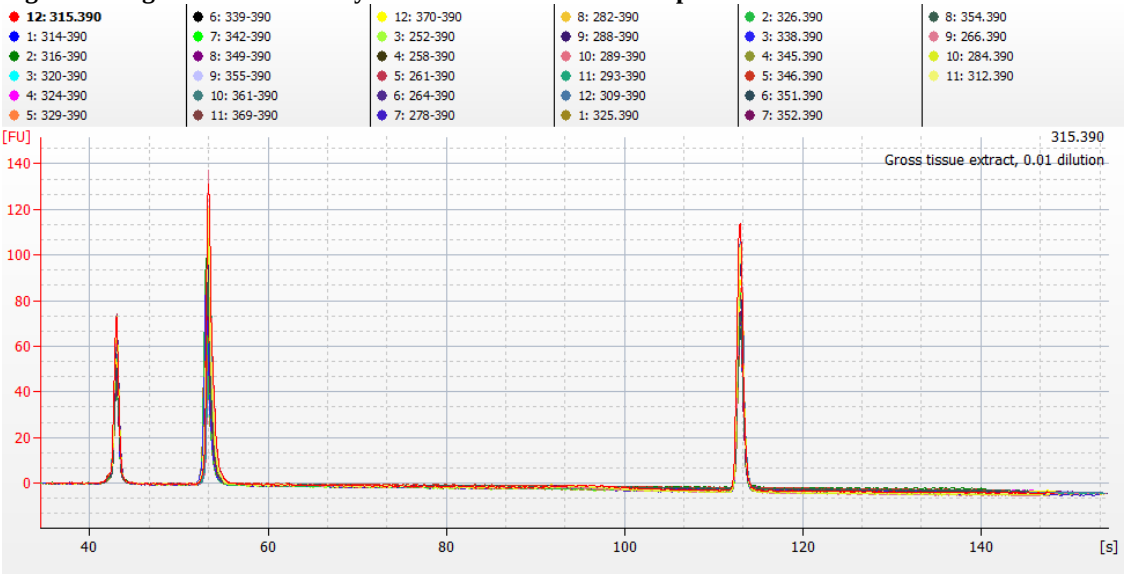
The 34 gross tissue samples were analyzed, and all samples were seen to generate the appropriate 16324 primer peak (see Figure 68 below).

Figure 68- Agilent 2100 Bioanalyzer results for 34 16324 samples



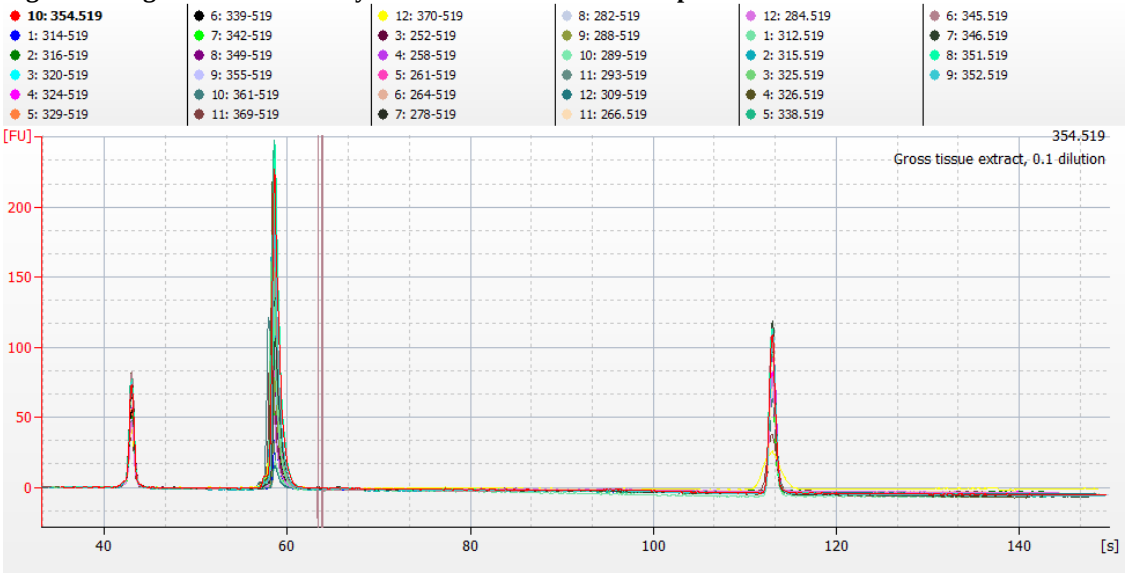
The 34 gross tissue samples were analyzed, and all samples were seen to generate the appropriate 16390 primer peak (see Figure 69 below).

Figure 69- Agilent 2100 Bioanalyzer results for 34 16390 samples



The 34 gross tissue samples were analyzed, and all samples were seen to generate the appropriate 16519 primer peak (see Figure 70 below).

Figure 70- Agilent 2100 Bioanalyzer results for 34 16519 samples



All of the samples were then amplified using the SNaPshot™ kit, using the standard amplification conditions. The samples were purified with SAP, mixed with GS120-LIZ size standard and deionized formamide and loaded onto 96-well plates for 3130*xl* capillary electrophoresis. Aliquots of each SNaPshot®-amplified primer set within a specific tissue sample were pooled together, and the 5-primer pooled mixture was aliquotted with size standard and formamide and loaded onto the electrophoresis plate as well. Similar pooling was done with the positive and negative control samples. The individual and pooled samples were run under normal SNaPshot™ electrophoresis conditions, utilizing 7% performance-optimized polymer (POP-7™), a 36 cm capillary array, and the E5 chemistry detection and analysis protocol.

Using the GeneMapper™ software, the samples were analyzed using only the multiplex analysis setting, which did not define bins and instead called all above-threshold samples (30 RFU and above) within the multiplex range from 20-50 bp. Using this analysis, the possible peaks for each primer within each gross tissue sample were determined, and this was compared

to the pooled data sample for each tissue sample as well. The positive and negative controls were also analyzed in the same way, and using this analysis, the positive control samples were seen to pass, and the amplification negative control samples established the negative control threshold for remainder of the sample analysis. When the negative control threshold was applied to the samples, only three samples were seen to contain below-threshold C peaks, and only at the 16390 positions. In each of these samples, a major T peak was present at the 16390 position instead, indicating that a possible heteroplasmic C peak was present, but since they were below the negative control threshold, these samples were inconclusive for analysis at the 16390 SNP position. All other peaks in every sample were above the negative control threshold, and were therefore deemed valid peaks.

Based on the analysis of the single primers against the gross tissue samples, the new individual bin sets were determined, as displayed in Table 20:

Table 20- Bin Sizes Based on Length Study

SNP Primers	Bin Range
16069	31 to 37 bp
16223	28 to 35 bp
16324	30 to 37 bp
16390	34 to 42 bp
16519	39.5 to 45.5 bp

Only the specific bin set above was used to analyze the one-step cells and/or individual primer amplicons from the gross tissue, based on the SNP primers utilized in the testing for mtDNA heteroplasmy.

Re-analysis of the Gross Tissue and Length Study Data

Based on the early interpretation of the gross tissue data, it was incorrectly deduced that only seven samples contained heteroplasmy, and only at the 16390 and 16519 locations. Prior to the length study, three samples were selected to screen individual cells as a proof of concept

study, samples 261 and 342 for 16390 and 355 for 16519. When the cells from these samples were analyzed using the specific extension primers, it was seen that the 16390 samples contained an unusual level of heteroplasmy within an odd SNP peak pattern, and the 16519 sample was also displaying a divergent result, in which the expected peak was not present in the sample. The issues noted with the individual cells from these samples led to the design of the length study. (Note: The raw data from 261, 342, and 355 were re-analyzed following the length study, and the results of these analyses are presented later in the discussion.)

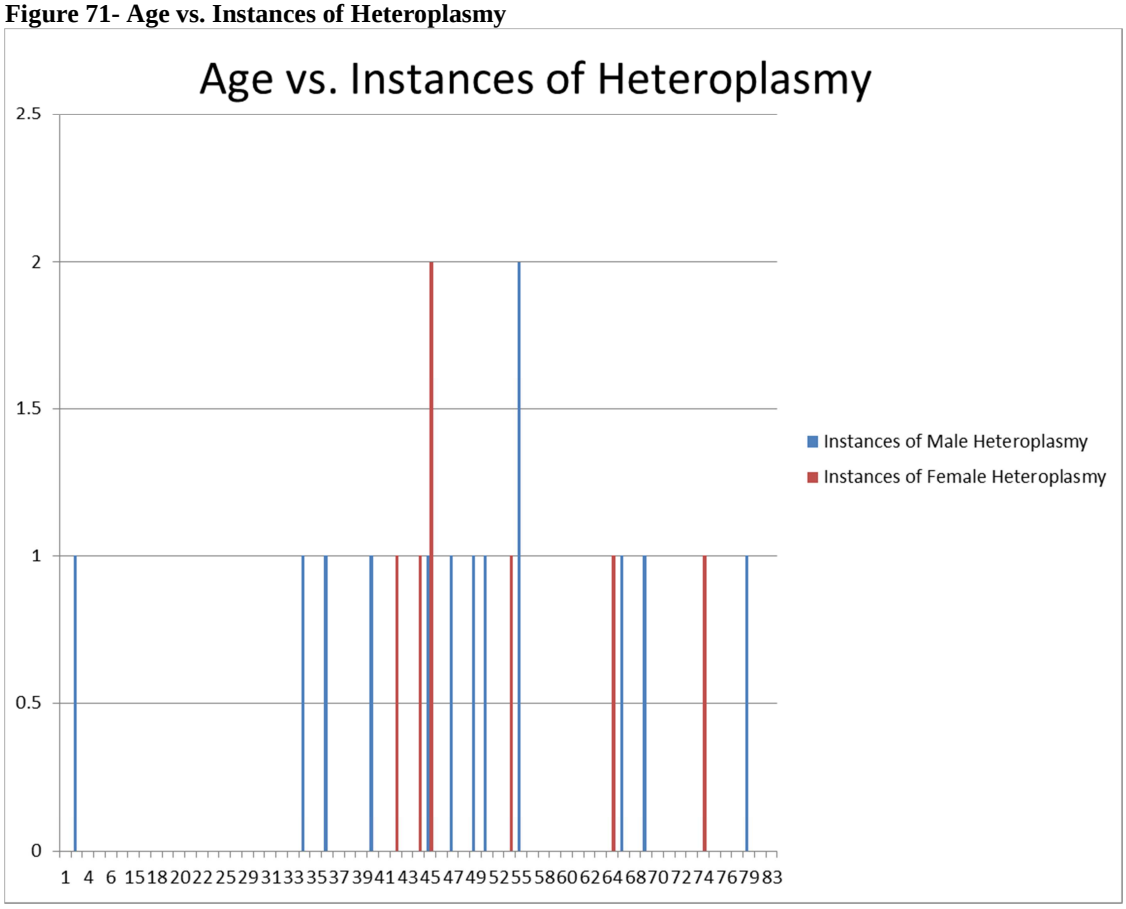
Using the new SNP bin sizes, the gross tissue data was re-analyzed, taking into account the possible crossover within the 16069, 16223, 16324, and 16390 data. Using this new analysis within an Excel macro, sixteen gross tissue samples were found to contain twenty possible HVI hotspot heteroplasmic positions, with representative coverage of all of the SNP positions throughout the sixteen samples. A final confirmation was done by visually confirming the mixed peaks in the electropherograms from each of the sixteen samples. The detailed results of the heteroplasmic gross tissues are presented in Appendix E: Detail Table of Heteroplasmic Tissue Samples.

Based on the post-length study reanalysis, samples 288 and 329 were again seen to contain a homoplastic SNP profile at each SNP position, and therefore these two samples were confirmed as the control samples for the remainder of the study. When the heteroplasmic single cells were tested, five control cells from sample 288 and five control cells from sample 329 were added to each set of fifty cells during the one-step amplification. This was done to control for the amplification of additional low-level heteroplasmy in the heteroplasmic single cells using control cells with expected homoplastic SNP profiles.

Age vs. Heteroplasmy of the Gross Tissue

Once the heteroplasmic samples were identified within the gross tissues, the known age and sex of the tissue donors was evaluated against the heteroplasmic findings. The results of this comparison are seen in Figure 71. Although the original age range of the study was seen to cover from 1 year of age up to 83 years of age across the 129 original samples (see Figure 3), the heteroplasmic samples were concentrated in a much tighter age range, with fourteen of the twenty heteroplasmic samples occurring between the ages of 34 and 55, with one sample occurring at age 2, and the remaining five samples loosely clustered between the ages of 64 and 79. Out of the twenty heteroplasmic samples, it was seen that seven of the samples were female, and thirteen of the samples were male, which would reflect a heteroplasmy rate of 20% in the female samples, and 13.8% in the male samples based on the screening of the gross tissue. Within the scope of the data set, it was noted that no heteroplasmy was seen in the samples between the ages of 3 and 33, with all but one instance of heteroplasmy occurring at and above the age of 34.

Figure 71- Age vs. Instances of Heteroplasmy



Pathology Follow-Up

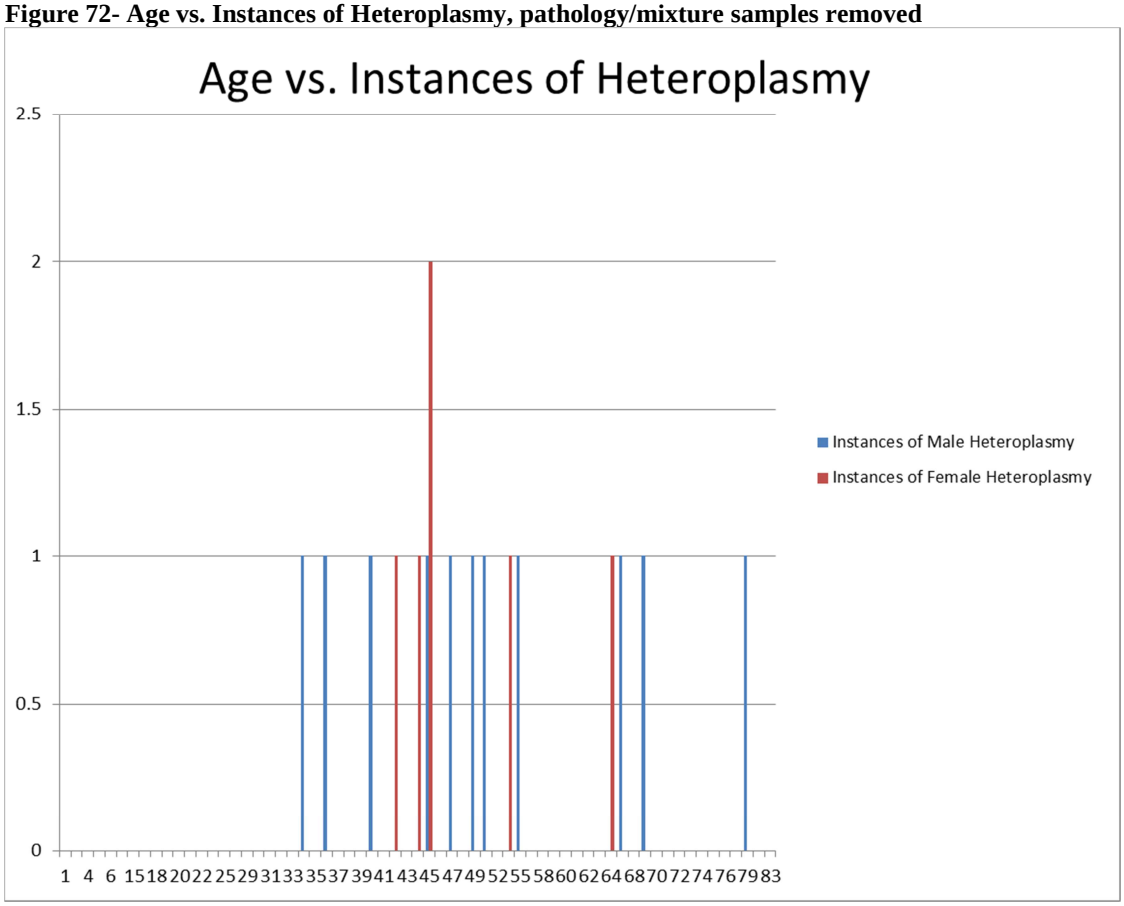
When the length study was performed, several tissues were extracted and tested for the generation of the new bin size data. Any of the samples found to contain possible heteroplasmy were included in the reanalysis data along with the original samples. Four such samples from the length study that were found to contain heteroplasmy were previously eliminated due to questionable pathology: sample 312, sample 338, sample 345, and samples 354. The use of these samples in the validation was not problematic since the length study data from these samples corresponded with the other samples and were therefore deemed concordant. However, it was noted that sample 345 was one of two identified heteroplasmic samples that contained three possible heteroplasmic hotspots based on the gross tissue analysis; the other sample that

contained three heteroplasmic hotspots was sample 261, which had no pathological problems. It was decided at this point, however, that no further testing would be performed on the sample 261. Therefore, with the elimination of sample 261 along with the four questionable pathology samples, the list of possible heteroplasmic gross tissue samples contained fifteen samples representing nineteen different heteroplasmic locations, which again covered all five of the representative SNP locations throughout all of the fifteen samples (see Appendix E for the detail on peak size, height, area, and ratios).

After the questionable samples were removed from the pool of eligible heteroplasmic samples, the age range of the remaining fifteen eligible heteroplasmic samples was again plotted, and the data was again seen to be concentrated in the age range of 34 and 55, but the majority of the previous outlier data was now gone (see Figure 72, below). With the removal of the questionable samples, the remaining seventeen samples were clustered in the main group of 34 to 55 years of age, and a second group of three samples were clustered around 64 to 69 years of age, with a single outlier of a 79 year-old male. Out of these seventeen remaining heteroplasmic samples, it was seen that six of the samples were female, and eleven of the samples were male, which would then reflect a heteroplasmy rate (corrected for the removal of the questionable/pathological samples) of 20% in the female samples, and 15% in the male samples, within the scope of the data set based on the screening of the gross tissue. Although based on a relatively small sample set (gross tissues, n=129, tested samples, n=109, after questionable/pathological samples removed, n=92), these heteroplasmy rates are consistent and in range with the heteroplasmy rates seen in the population, as per the literature (Bendall et al., 1997; Budowle et al., 2002a; Budowle et al., 2002b; Calloway et al., 2000; D'Eustachio, 2002; Grzybowski, 2000, 2001; Grzybowski et al., 2003; Irwin et al., 2009; Lagerstrom-Fermer et al.,

2001; Li et al., 1999; Roberts & Calloway, 2011; Salas et al., 2001; Tully et al., 2000; Wilson et al., 1997; Zsurka et al., 2007).

Figure 72- Age vs. Instances of Heteroplasmy, pathology/mixture samples removed



SNaPshot™ Analysis of Single Cells

Isolation of Single Cells from Heteroplasmic Samples and Controls

Based on funding and resource constraints, it was determined that only ten samples would be selected for single cell isolation and SNaPshot™ testing. Due to representative coverage of every SNP position within the remaining fifteen samples, it was decided that two samples would be chosen for each SNP site and tested. The criteria for selection was as follows:

1) Samples with the lowest ratio of heteroplasmic peak heights are optimal, which would indicate strong signals from both of the heteroplasmic peaks within the gross tissue sample.
2) If multiple samples exist for a SNP location with roughly similar peak height ratios, two different heteroplasmic mixtures will be selected to test for varying coverage of the SNP heteroplasmy at the position.

Based on these criteria, the following samples were selected for single-cell isolation and one-step amplification, as presented in Table 21 (see Appendix E for detail on peak size, height, area, and ratios).

Table 21- Controls and Heteroplasmic Samples Selected for SNaPshot™ Analysis of Single Cells

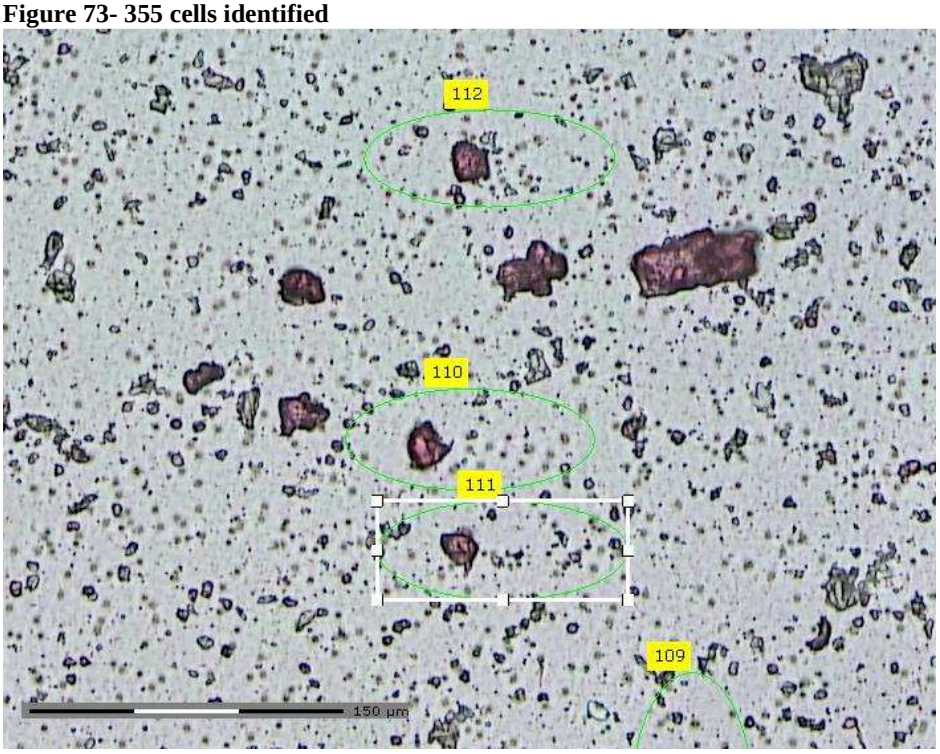
Sample	16069 SNP Location	Allele	16223 SNP Location	Allele	16324 SNP Location	Allele	16390 SNP Location	Allele	16519 SNP Location	Allele
288	16069	C	16223	C	16324	T	16390	C	16519	A
329	16069	C	16223	C	16324	T	16390	C	16519	A
284	16069	T A								
289					16324	T G				
293			16223	T C						
314			16223	T C						
315	16069	C A								
339							16390	T C		
349					16324	T A				
355							16390	C T	16519	G A
369									16519	G A

For each of the selected heteroplasmic tissue samples and the two control tissue samples, fifty cells were then isolated for analysis for the representative heteroplasmic SNP position. The cells that were utilized were from the previously prepared microscope slides. The cells were identified, isolated, and collected on the PALM laser-dissection microscope. Each cell was individually collected in 15 µl of capture buffer (83.4% dH_2O, 16.6% bovine serum albumin (1.6

µg/µl)) within a sterile collection tube. The addition of the BSA to the capture buffer allowed the liquid to completely ensconce the interior of the collection tube lid, providing a maximum target area for the ablated cell within the normally hydrophobic interior of the plastic cap. The total volume of the capture buffer contained the same volumes of water and BSA as the targeted amplification, so the entire capture buffer volume was applied to the one-step amplification of the individual cells.

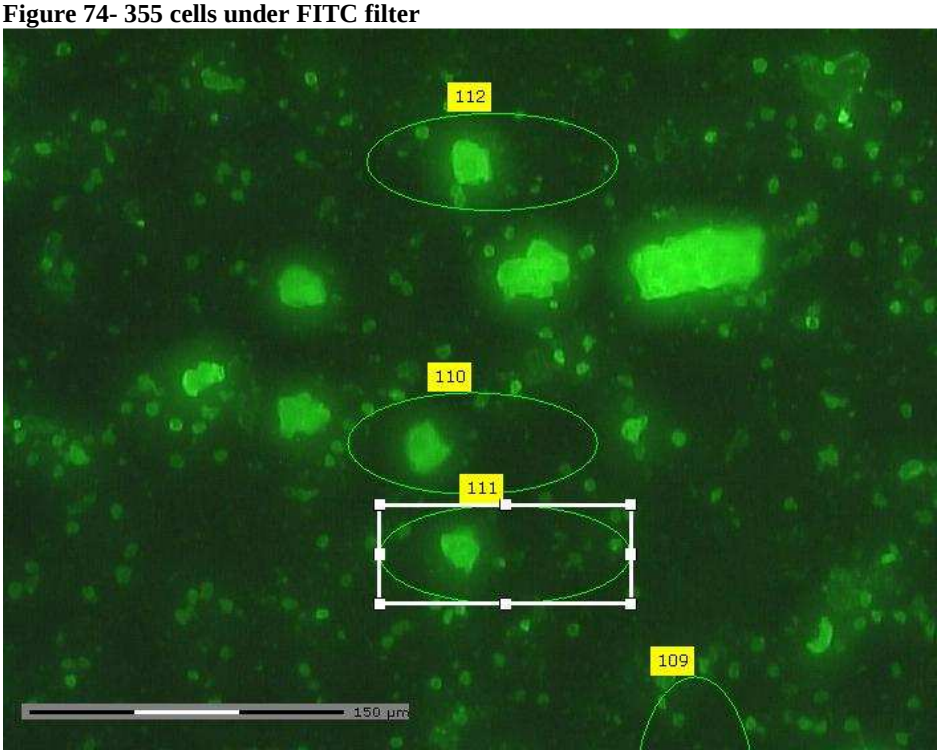
An example of this process is shown in Figure 73 through Figure 77, below. In the example, the field of view shows separated cells from the 355 tissue sample. The computer-outlined regions, which make up the digital map of the cells, can be seen in Figure 73 as the numbered green circles. For this example, note the proximity of the 110 and 111 numbered cells.

Figure 73- 355 cells identified



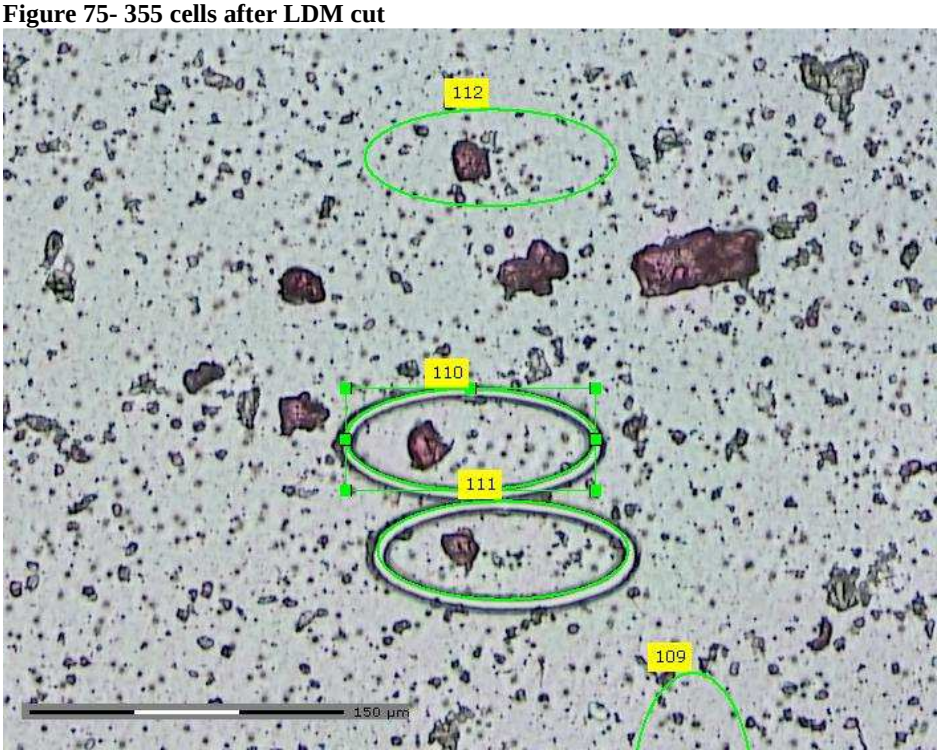
Under the FITC filter, it is shown in Figure 74 that the 110 and 111 cells each contain a single nuclear void, and along with the light-field image from Figure 73, indicate that the cells are single cells and therefore eligible for individual collection.

Figure 74- 355 cells under FITC filter



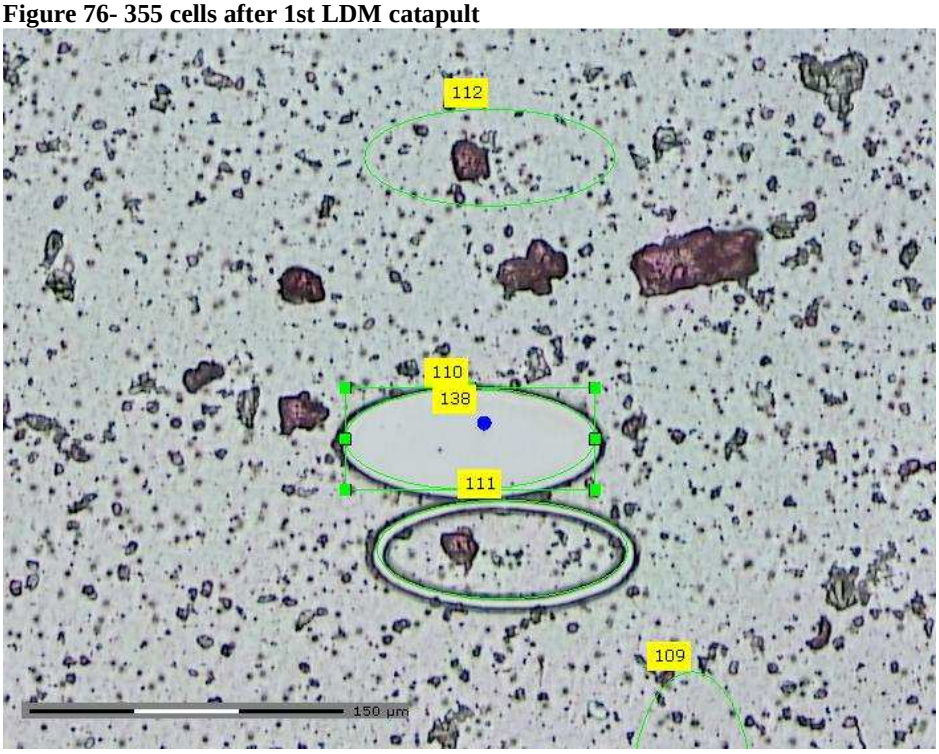
Once identified, the cells were each cut using the laser microscope. This involved selecting each sample individually and setting the laser dissection microscope to activate the laser and cut along the indicated oval pathway. As the laser is activated, the robotic stage moved to allow the fixed laser to cut along the green circular pathways. As seen in Figure 75, this can be done with such precision as to not disturb the surrounding cells. In this example, cell 110 was first cut, followed by cell 111, and they are both still present on the slide and ready for catapult capture into individual collection tube caps.

Figure 75- 355 cells after LDM cut



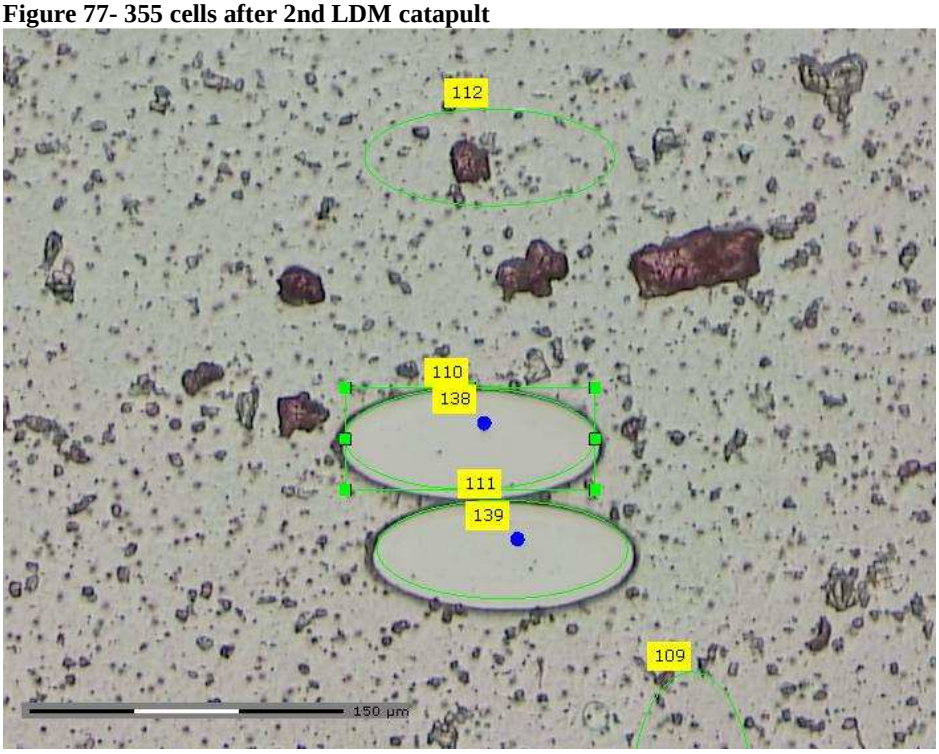
Once the cut was performed on the slide surface, the liquid-filled collection tube would be attached to the laser dissection microscope collection tube holder, and moved into place above the slide. The height was manually controlled using the computer interface to verify that neither the cap nor any liquid came into contact with the slide. Once confirmed, the 110 cell was then selected with a single catapult launch point, and was launched off the slide surface into the collection cap. As is seen in Figure 76, the application of the focused, pulsed laser on the slide surface only affected the targeted cell, and did not affect the 111 cell, or any of the other cells or debris surrounding the 110 cell.

Figure 76- 355 cells after 1st LDM catapult



After the 110 cell collection tube was removed and closed, a second tube was inserted onto the tube collection arm of the laser dissection microscope, and was again moved into place above the slide. The visual check was done to ensure that the proper collection tube cap height was valid, and the 111 cell was then selected and catapulted off the surface of the slide. As seen in Figure 77, this cell was also successfully removed from the slide surface without disrupting any of the surrounding cells or debris.

Figure 77- 355 cells after 2nd LDM catapult



Amplification and Agilent 2100 Bioanalyzer Data for Isolated Single Cells

The cells from the heteroplasmic samples and controls were one-step amplified using the specific SNP primers per sample based on the amplification settings that were previously validated. The cells from samples containing possible heteroplasmy at 16069, 16223, 16324, and 16390 were run using 34 cycles of amplification during the one-step amplification process. The cells from the samples containing possible heteroplasmy at 16519 were run using 28 cycles of amplification for the one-step amplification process. To each set of fifty heteroplasmic sample cells, five control cells were added to the amplification set from sample 288, and five control cells were also added from sample 329, giving a total of sixty cells that were amplified within each test set. To control for the amplification, a positive control (PE), amplification negative control (AN), and cycling amplification negative control (CAN) were established for each set of 60 cells. The positive control and the amplification negative control were run at the

standard number of amplification cycles specific for each primer set, and were used to establish the negative control threshold after SNaPshot™ analysis was complete. The cycling amplification negative control (CAN) was only to be used if it was determined that the samples required further rounds of primary amplification following Agilent 2100 Bioanalyzer sizing analysis, before the SNaPshot™ amplification occurred.

Based on funding and resource constraints, only the controls and a representative number of samples for each set of sixty cells were run on the Agilent 2100 Bioanalyzer. A total of twenty-four samples were run from each amplification set, which included the positive control, the amplification negative control, the cycling negative control, seventeen of the heteroplasmic sample cell samples, and four of the control cell samples. This was done, based mainly on the resource constraints of the study, but also on the empirical data from the validations, in which the one-step amplified single cells would not be expected to display a visible peak in the Agilent 2100 Bioanalyzer data. As noted in the cycling number validation study using the test cells, negative results on the Agilent 2100 Bioanalyzer data did not indicate that the single-cell samples could not transition to SNaPshot™ analysis, based on the low-end level of the detection of the Agilent 2100 Bioanalyzer being higher than the low-end input concentration for the SNaPshot™ assay. In order for the samples to effectively pass Agilent 2100 Bioanalyzer sizing analysis, however, the positive control must have displayed the appropriate band and the negative controls must have been negative.

Following the one-step amplification, all of the amplified single-cell samples were treated with ExoSAP-IT® using the Vallone protocol (Vallone et al., 2004). The representative amplified single cells and controls were then run on the Agilent 2100 Bioanalyzer. Refer to

Appendix F for the Agilent 2100 Bioanalyzer electropherograms for the one-step amplification positive and negative control results.

The positive controls for the 16069 primers, controlling the 284 and 315 samples and control cells, were seen at the appropriate bp size, and the negative controls for the sample sets displayed no amplified peaks.

The positive controls for the 16223 primers, controlling the 293 and 314 samples and control cells, were seen at the appropriate bp size, and the negative controls for the sample sets displayed no amplified peaks.

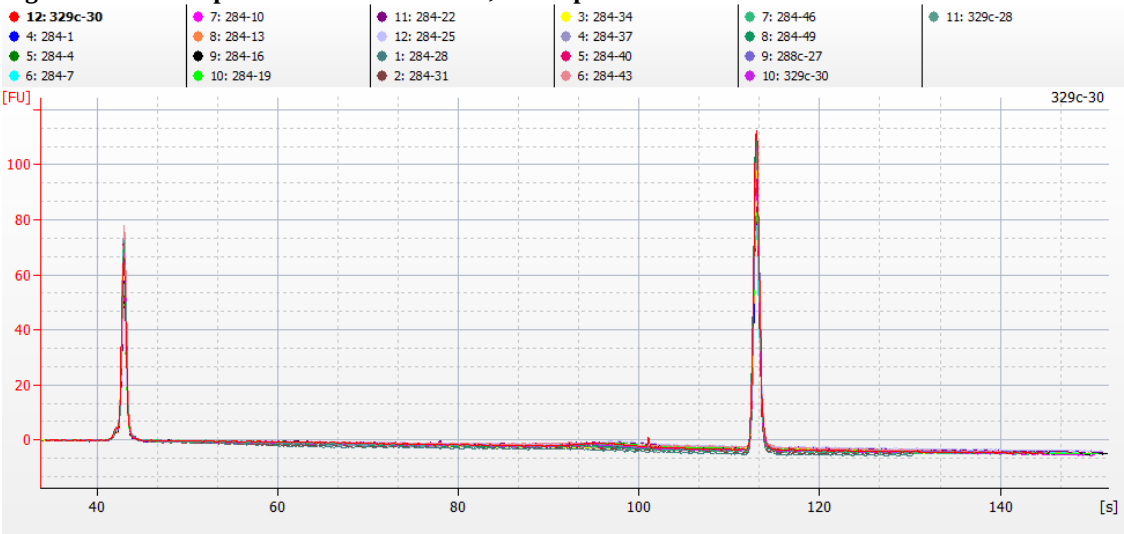
The positive controls for the 16324 primers, controlling the 289 and 349 samples and control cells, were seen at the appropriate bp size, and the negative controls for the sample sets displayed no amplified peaks.

The positive controls for the 16390 primers, controlling the 339 and 355 samples and control cells, were seen at the appropriate bp size, and the negative controls for the sample sets displayed no amplified peaks.

The positive controls for the 16519 primers, controlling the 355 and 369 samples and control cells, were seen at the appropriate bp size, and the negative controls for the sample sets displayed no amplified peaks.

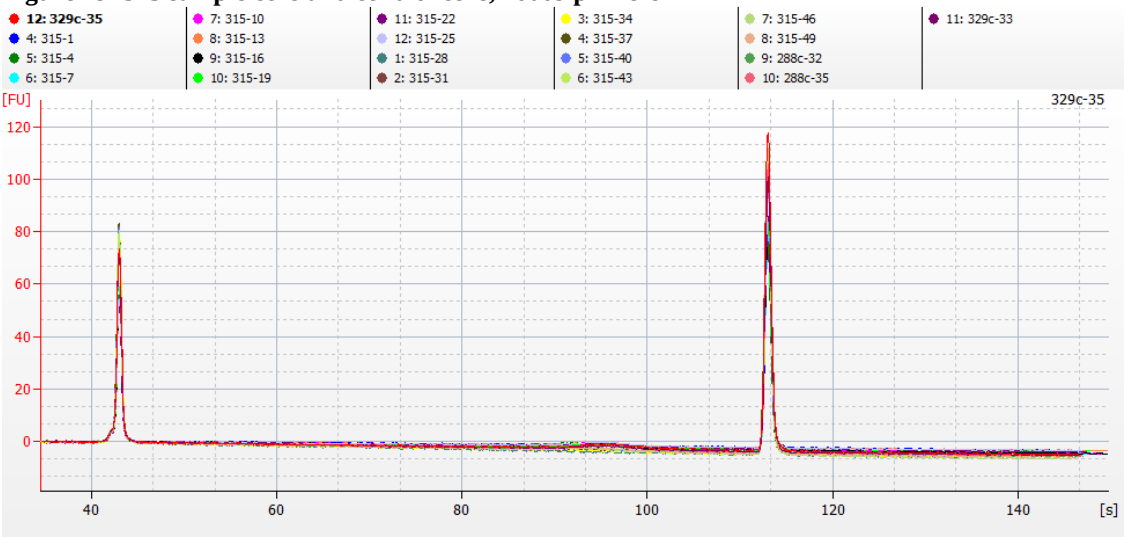
When the Agilent 2100 Bioanalyzer data for the single-cell samples was analyzed, it was seen that no detectable peaks were present for any of the single cell samples or control cell samples. In Figure 78 below, the twenty-one sample electropherograms for the representative cells and controls are overlaid, and no peaks are present in the 109 bp region for the 284 sample cells and control cells for primer 16069.

Figure 78- 284 sample cells and control cells, 16069 primers



In Figure 79 below, the twenty-one sample electropherograms for the representative cells and controls are overlaid, and no peaks are present in the 109 bp region for the 315 sample cells and control cells for primer 16069.

Figure 79- 315 sample cells and control cells, 16069 primers



In Figure 80 below, the twenty-one sample electropherograms for the representative cells and controls are overlaid, and for twenty of the 293 sample cells and control cells, no peaks are present in the 81 bp region for primer 16223. The cell #46 sample displayed a peak in the Agilent 2100 Bioanalyzer electropherogram, and this peak was above the detection threshold for

the Agilent 2100 Bioanalyzer software and was correctly identified at the appropriate base-size pair for the 16223 primers, as seen in Figure 81, below.

Figure 80- 293 sample cells and control cells, 16223 primers

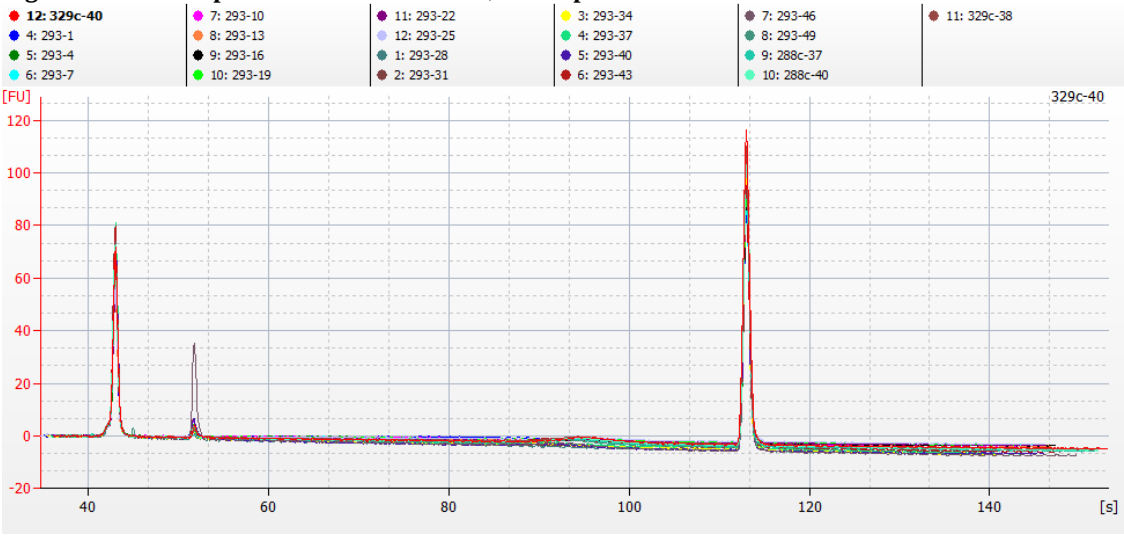
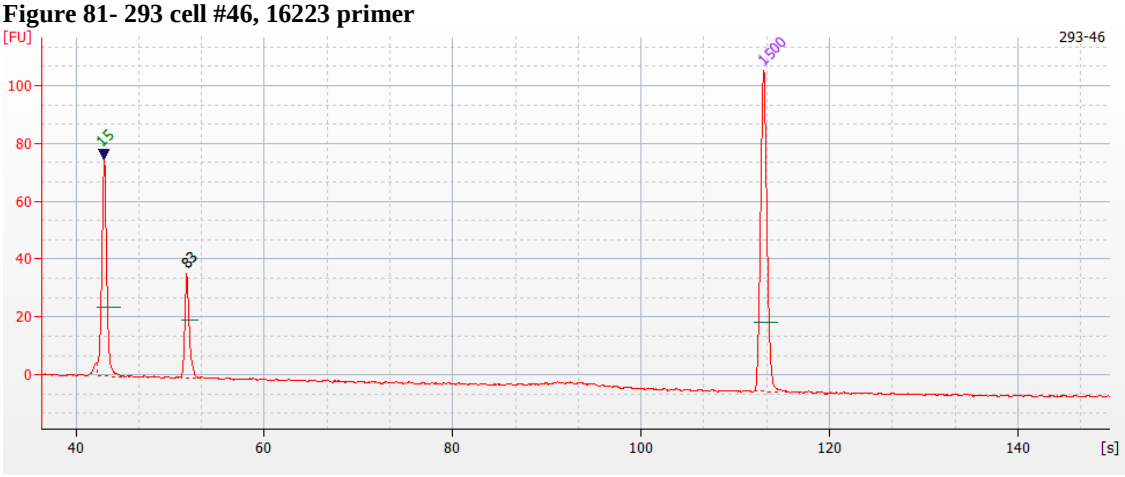


Figure 81- 293 cell #46, 16223 primer



In Figure 82 below, the twenty-one sample electropherograms for the representative cells and controls are overlaid, and for twenty of the 314 sample cells and control cells, no peaks are present in the 81 bp region for primer 16223. The cell #1 sample displayed a peak in the Agilent 2100 Bioanalyzer electropherogram, but the peak was below the detection threshold for the Agilent 2100 Bioanalyzer software and was therefore not identified, as seen in Figure 83, below.

Figure 82- 314 sample cells and control cells, 16223 primers

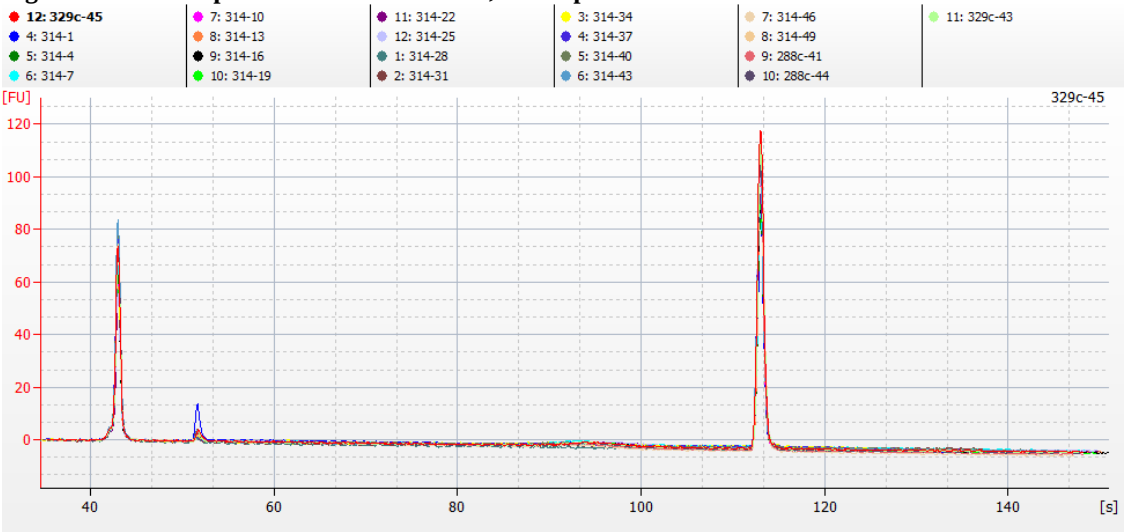
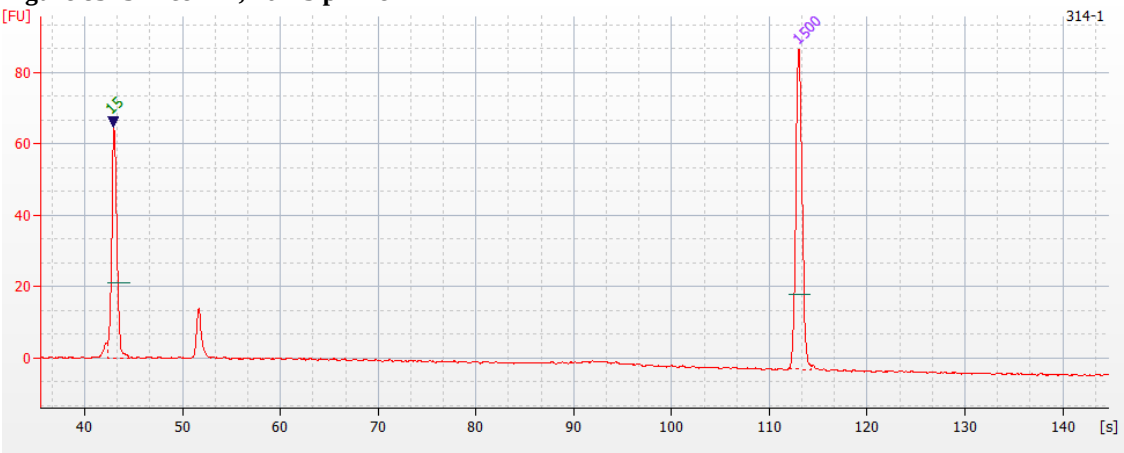


Figure 83- 314 cell #1, 16223 primer



In Figure 84 below, the twenty-one sample electropherograms for the representative cells and controls are overlaid, and for twenty of the 289 sample cells and control cells, no peaks are present in the 99 bp region for primer 16324. The cell #4 sample displayed a peak in the Agilent 2100 Bioanalyzer electropherogram, but the peak was below the detection threshold for the Agilent 2100 Bioanalyzer software and was therefore not identified, as seen in Figure 85, below.

Figure 84- 289 sample cells and control cells, 16324 primers

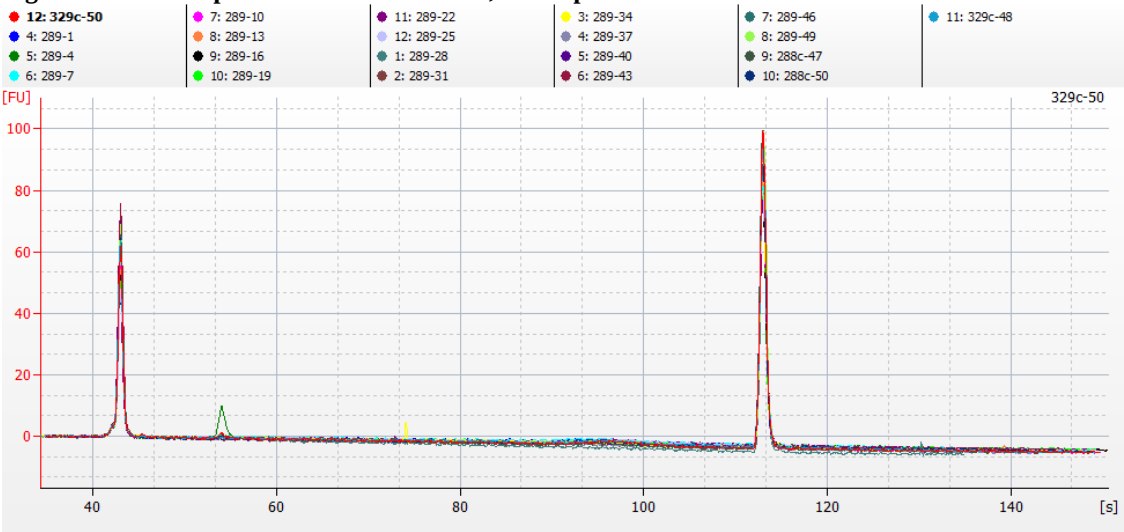
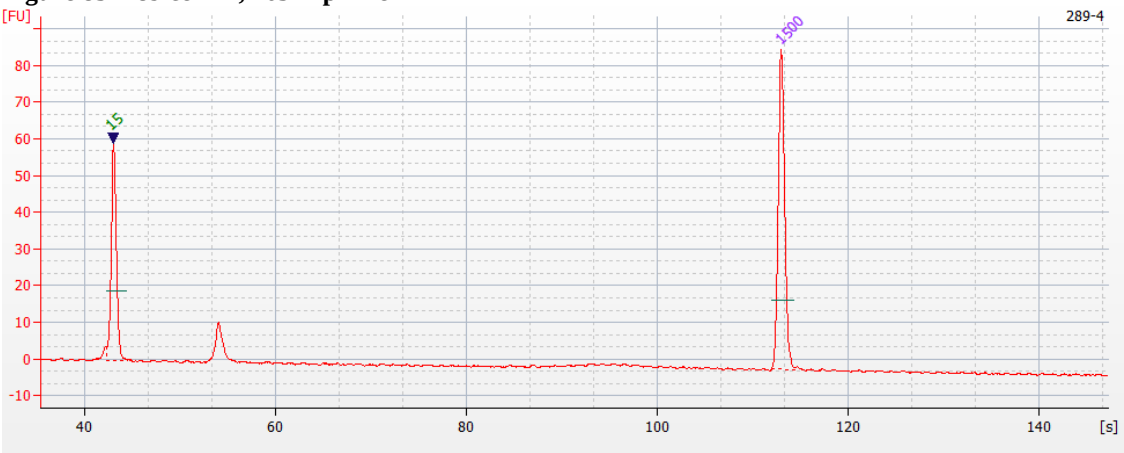
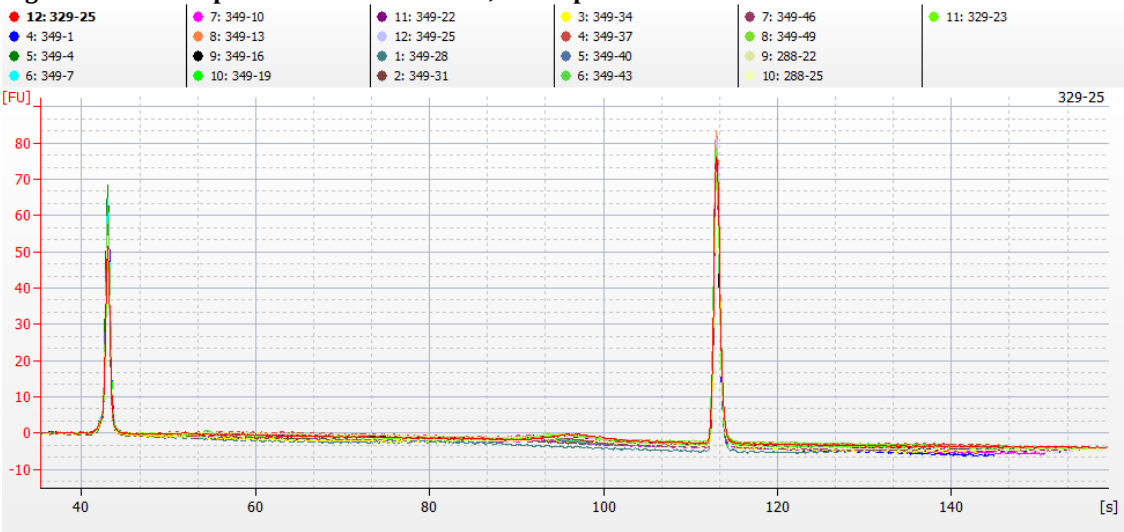


Figure 85- 289 cell #4, 16324 primer



In Figure 86 below, the twenty-one sample electropherograms for the representative cells and controls are overlaid, and no peaks are present in the 99 bp region for the 349 sample cells and control cells for primer 16324.

Figure 86- 349 sample cells and control cells, 16324 primers



In Figure 87 below, seventeen of the sample electropherograms for the representative cells and controls are overlaid, and for the 339 sample cells and control cells, no peaks are present in the 93 bp region for primer 16390. The four remaining samples for the 339 cells displayed electrophoresis artifacts in the data, but none of the samples displayed an above-threshold peak in the 93 bp region for the 339 cells, as seen in Figure 88 below.

Figure 87- 339 sample cells and control cells, 16390 primers

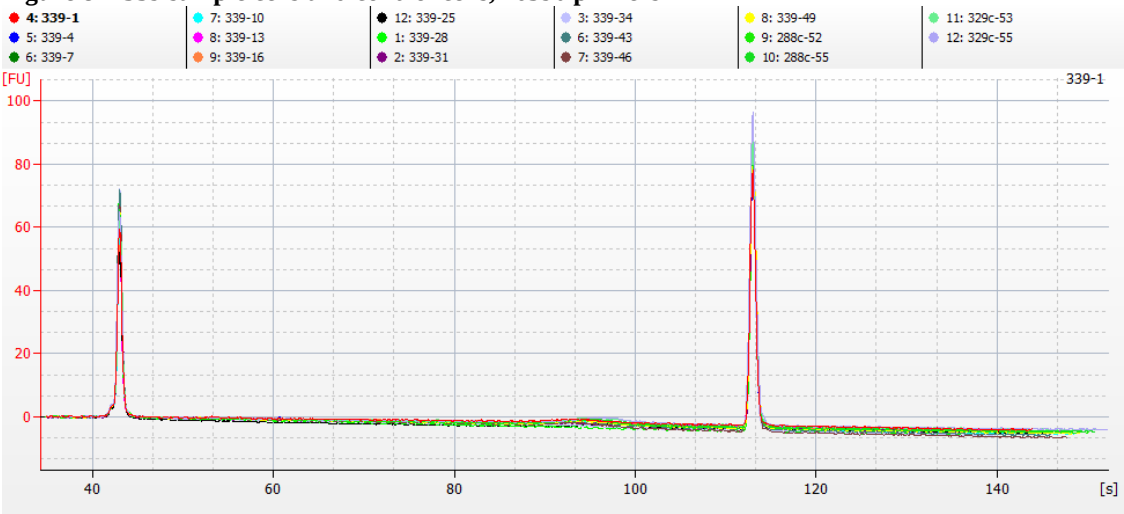
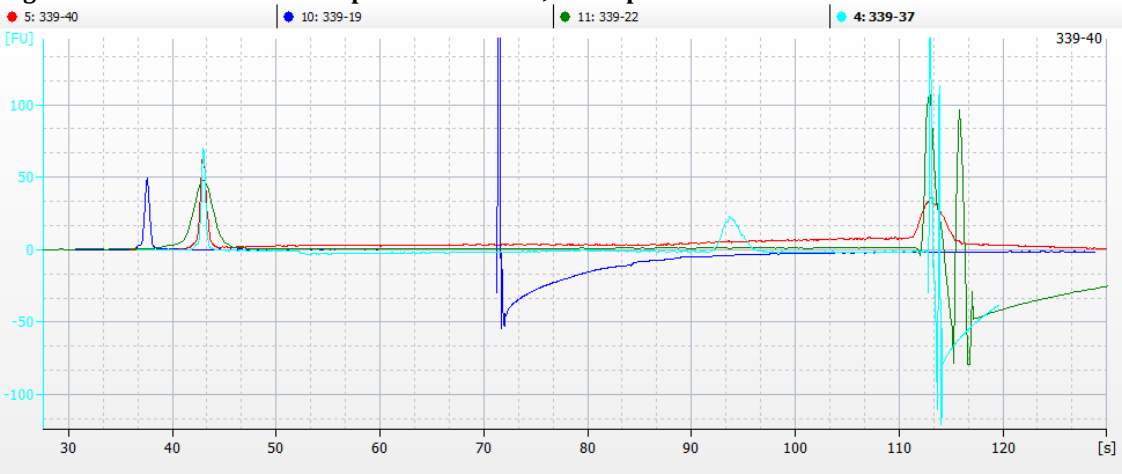
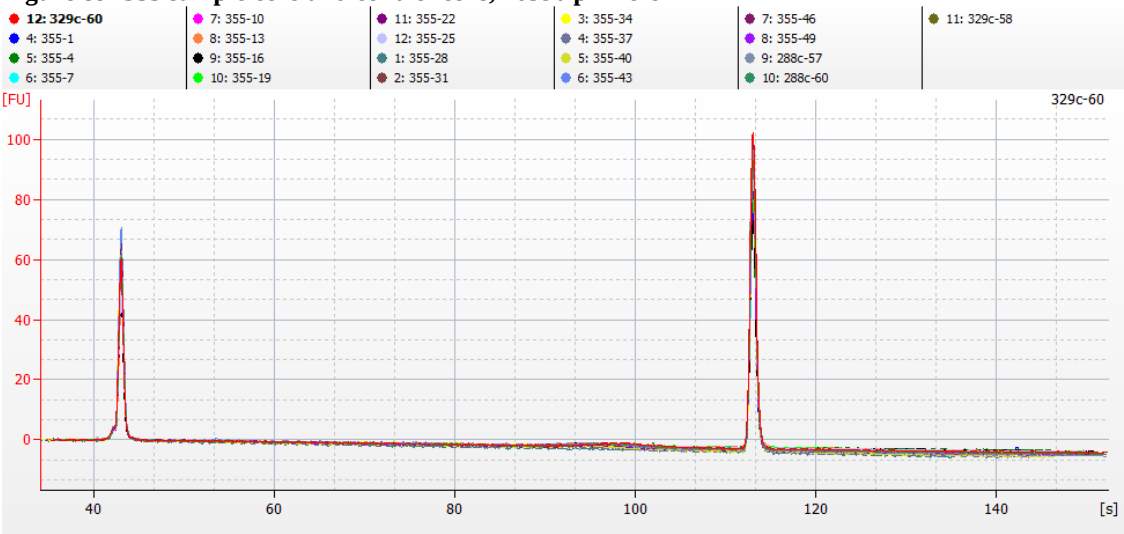


Figure 88- 339 cells with electrophoresis artifacts, 16390 primers



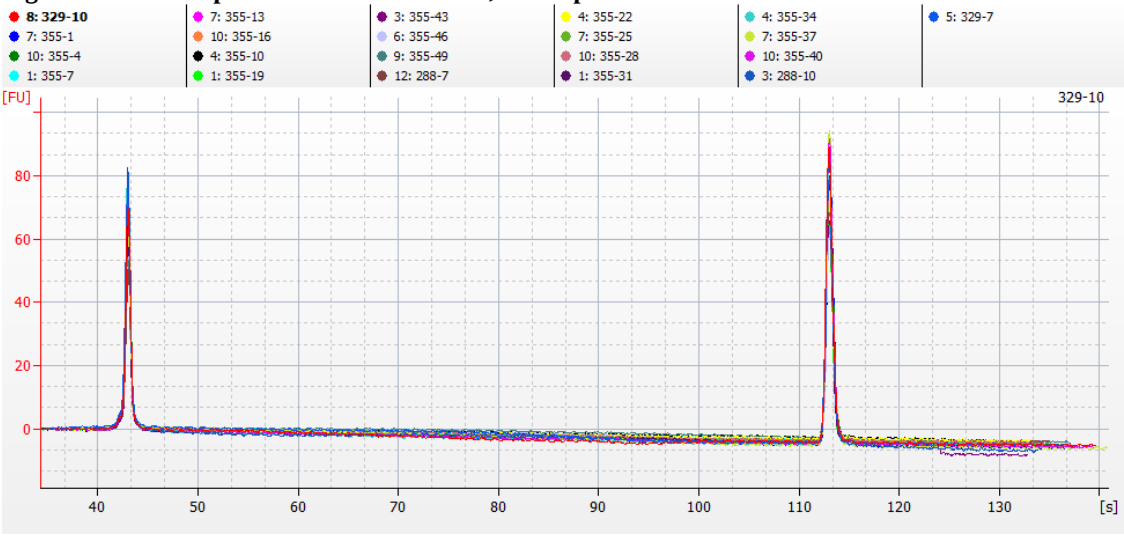
In Figure 89 below, the twenty-one sample electropherograms for the representative cells and controls are overlaid, and no peaks are present in the 93 bp region for the 355 sample cells and control cells for primer 16390.

Figure 89- 355 sample cells and control cells, 16390 primers



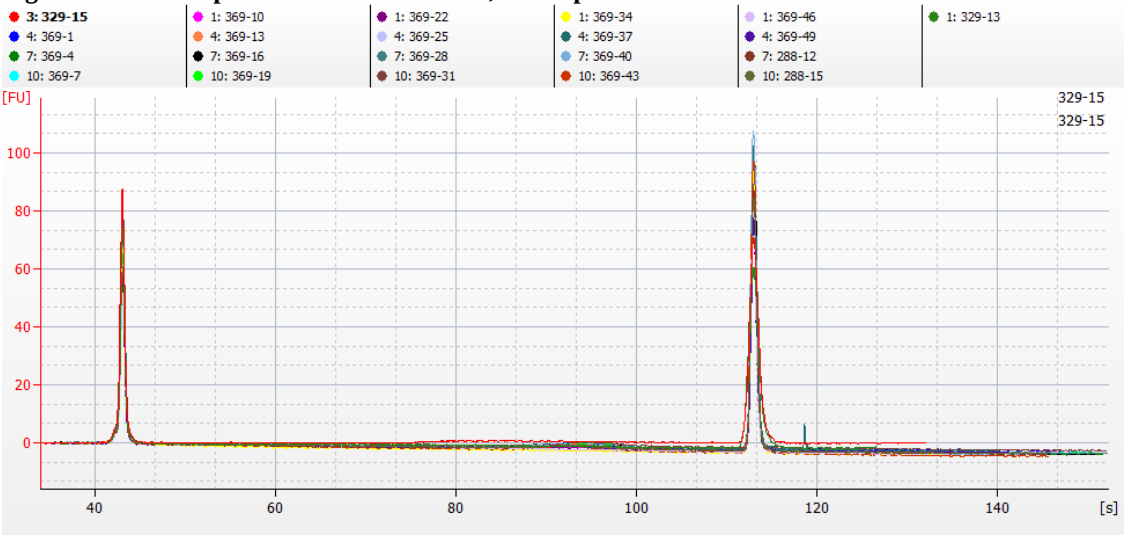
In Figure 90 below, the twenty-one sample electropherograms for the representative cells and controls are overlaid, and no peaks are present in the 137 bp region for the 355 sample cells and control cells for primer 16519.

Figure 90- 355 sample cells and control cells, 16519 primers



In Figure 91 below, the twenty-one sample electropherograms for the representative cells and controls are overlaid, and no peaks are present in the 137 bp region for the 369 sample cells and control cells for primer 16519.

Figure 91- 369 sample cells and control cells, 16519 primers



SNaPshot™ Amplification of the Single Cells

The single-cell samples and controls were amplified in the SNaPshot™ system based on the observations made in the validation of the amplification cycle number regarding negative Agilent 2100 Bioanalyzer results for the single-cell samples. The SNaPshot™ amplification was performed based on the data that the positive controls for each specific primer set displayed the appropriate-sized amplicon, and the negative controls displayed negative results. The samples were amplified in the SNaPshot™ system as per the standard kit guidelines using the specific SNP extension primer for the related samples: the 16069 extension primers for the 284 and 315 cells and controls, the 16223 extension primers for the 293 and 314 cells and controls, the 16324 extension primers for the 289 and 349 cells and controls, the 16390 extension primers for the 339 and 355 cells and controls, and the 16519 extension primer for the 355 and 369 cells and controls. In order to maintain a conservative approach to the controls, the PE was run at the optimal input of 0.2 pg of amplified DNA, which required dilution. The AN and CAN samples were run at a neat input concentration based on the similar level of sample input for all of the single-cell samples in the sample set. For example, in a scenario where three samples are amplified in the SNaPshot™ assay, one sample requires a 1/100 dilution to achieve 0.2 pg of amplified DNA, one sample requires a 1/10 dilution to achieve 0.2 pg of amplified DNA, and the third sample is not diluted (input neat) into the SNaPshot™ assay. In this scenario, the negative controls would be run neat (not diluted) to take into account the highest possible level of sample input into the system. The rationale here was that any SNP peaks seen in these negative controls will be at the highest possible level for all samples, with diluted samples theoretically containing less contaminant (if present) due to the dilution, and therefore will be the most stringent and conservative interpretation of the negative controls. For the single cells, the AN and the CAN

were amplified neat, as were all of the single-cell samples since the Agilent 2100 Bioanalyzer data did not detect a quantity of amplified DNA in these samples. Following SNaPshot™ amplification, the samples were purified with SAP, run on the 3130*xl* capillary electrophoresis instrument utilizing 7% performance optimized polymer (POP-7™), a 36 cm capillary array, and the E5 chemistry detection and analysis protocol.

SNaPshot™ Results of Sample 284 Cell Analysis

Upon analysis, the SNaPshot™ internal positive and negative controls were seen to pass. The amplification positive and negative controls from the 284 cell samples were analyzed first. For reference, the HL60 DNA type at 16069 should be a T base around 34 bp. The results are presented in Table 22:

Table 22- Control Sample Results for 284 Cells in SNaPshot™

Control Sample	Primer Set	SNP Base Detected	Base-pair size	Peak Height (in RFU)
PE	16069	T	34.62	8139
AN	16069	C	31.93	520
CAN	16069	C	32.1	41

The positive control passed, and the negative control threshold was established as 520 RFU for ~32 bp sized peaks based on the peak seen in the amplification negative control. The cycling amplification negative control (CAN) for these samples did show peaks as well, but since none of the samples were amplified with additional cycles, this control would not be analyzed any further.

The fifty single cell samples for 284 as well as the ten control cell samples were then analyzed. Based on the gross tissue analysis of tissue 284, the 16069 extension primer showed an A/T heteroplasmy. While this A/T heteroplasmy was expected based on the gross tissue analysis, C/T peaks were instead found (see the discussion for analysis of this issue).

The results of the SNaPshot™ SNP typing for each cell set are displayed in tables, such as Table 23 below. For each table that follows the individual cell data sets, the tables are divided into sets of columns, with one column for each of the bases of DNA detected in the various cell samples. The first set of columns reflects any of the first peaks detected in the heteroplasmic samples. The second set of columns reflects any secondary peaks detected in the samples, and so on for each column. Each set of the main peak columns contains individual data columns for the allele detected, as well as the size of the peak in bp, the height of the peak in RFU, and the area of the peak. Within each table, any sample name followed by a lowercase "c" is a control cell sample.

Table 23- SNaPshot™ Results for A/T heteroplasmic 284 Cells and Controls, 16069 Primers

Sample Name	Allele 1	Size 1	Height 1	Peak Area 1	Allele 2	Size 2	Height 2	Peak Area 2
PE-284					T	34.62	8139	57833
AN-284	C	31.93	520	3580				
CAN-284	C	32.1	41	306				
1-284	C	31.94	725	5073				
2-284	C	32.03	78	544				
3-284	C	32.09	573	3970				
4-284	C	31.94	1309	9020				
5-284	C	31.89	673	4575				
6-284	C	31.83	492	3372				
7-284								
8-284	C	31.89	363	2497				
9-284	C	32.05	298	2044				
10-284	C	31.94	440	3058				
11-284	C	32.01	1804	12360				
12-284	C	32.04	1572	10625				
13-284	C	31.91	779	5275				
14-284	C	31.91	761	5114				
15-284	C	31.89	153	1033				
16-284	C	31.91	3463	23573				
17-284	C	31.89	591	3982				
18-284	C	32.02	297	2038				
19-284	C	31.96	1155	7832				
20-284	C	31.91	803	5462				
21-284	C	31.91	468	3185				
22-284	C	31.91	1470	9941				
23-284								
24-284	C	31.84	1245	8467				
25-284	C	31.91	816	5510				
26-284					T	34.49	994	6925
27-284								
28-284	C	31.97	875	5959				
29-284	C	31.91	1867	12545				
30-284	C	31.91	466	3136				

Table 23- SNaPshot™ Results for A/T heteroplasmic 284 Cells and Controls, 16069 Primers (continued)

Sample Name	Allele 1	Size 1	Height 1	Peak Area 1	Allele 2	Size 2	Height 2	Peak Area 2
31-284	C	32	610	4117	T	34.62	1177	8141
32-284	C	31.91	1189	8133				
33-284	C	31.89	893	6136				
34-284	C	31.89	129	917				
35-284	C	31.91	812	5574				
36-284	C	31.83	240	1659				
37-284	C	31.97	225	1560	T	34.61	3781	26101
38-284	C	31.89	243	1659				
39-284								
40-284	C	32	416	2833				
41-284	C	31.91	794	5325				
42-284								
43-284					T	34.5	560	3828
44-284								
45-284	C	31.95	311	2092				
46-284	C	31.91	127	855				
47-284	C	31.91	140	936				
48-284	C	32.04	274	1876				
49-284	C	31.89	1840	12346				
50-284								
26-288c	C	32	479	3218				
27-288c	C	31.98	1114	7674				
28-288c								
29-288c	C	31.91	1020	6811				
30-288c	C	31.91	366	2473				
26-329c	C	31.91	240	1611				
27-329c	C	31.98	507	3412				
28-329c	C	31.89	1664	11366				
29-329c	C	31.91	217	1448				
30-329c	C	31.91	1353	9084	T	34.5	1864	12700

Using conditional formatting within the Microsoft® Excel 2010 software, the negative control threshold of 520 RFU was applied to the data and any ~32 bp C peak heights less than or equal to this threshold were labeled as red, and any ~32 bp C peak heights above this threshold were labeled as green. As the data was analyzed, it was seen that seven of the fifty samples and one of the ten controls gave no data, and these sample names were changed from black to red text (samples 7, 23, 27, 39, 42, 44, 50, and control sample 28-288c from the 284 cells). For the remaining samples, it was seen that eighteen of the samples were displaying peaks below the negative control threshold of 520 RFU for the ~32 bp sized peaks. For the control cells, five of the cells were also below this negative control threshold.

However, a second issue was occurring in the data as well. An additional T peak occurred in four random samples and one of the control samples. When this T peak occurred, it was seen to be approximately 34 bp in size. Because the amplification negative sample did not display a T peak, no negative control threshold for the T peak could be set, and therefore any visible T peak around 34 bp would be deemed a valid peak. When this was applied to the data, one sample that was seen to contain a below-threshold C peak around 31 bp also contained a valid T peak around 34 bp: sample 37. The analysis parameters for this sample will be addressed in the discussion. Therefore, by eliminating the samples that only contained a single below-threshold C peak at 31 bp, the remaining twenty-six viable cell samples (samples 1, 3, 4, 5, 11, 12, 13, 14, 16, 17, 19, 20, 22, 24, 25, 26, 28, 29, 31, 32, 33, 35, 37, 41, 43, and 49) and four control samples (27-288c, 29-288c, 28-329c, and 30-329c) were deemed valid for interpretation.

SNaPshot™ Results of Sample 315 Cell Analysis

Upon analysis, the SNaPshot™ internal positive and negative controls were seen to pass. The amplification positive and negative controls from the 315 cell samples were first analyzed.

For reference, the HL60 DNA type at 16069 should be a T base around 37 bp. The results are presented in Table 24:

Table 24- Control Sample Results for 315 Cells in SNaPshot™

Control Sample	Primer Set	SNP Base Detected	Base-pair size	Peak Height (in RFU)
PE	16069	T	34.49	8341
AN	16069	C	31.91	295
CAN	16069	N/A	N/A	N/A

The positive control passed, and the negative control threshold was established as 295 RFU for ~32 bp sized peaks based on the peak seen in the amplification negative control. The cycling amplification negative control (CAN) for these samples did not show peaks, but since none of the samples were amplified with additional cycles, this control would not be analyzed any further.

The fifty single cell samples for 315 as well as the ten control cell samples were then analyzed. Based on the gross tissue analysis, the 16069 extension primer showed a C/A heteroplasmy. While this C/A heteroplasmy was expected based on the gross tissue analysis, C/A/T peaks were instead found (see the discussion for analysis of this issue). The results of the SNaPshot™ SNP typing are displayed in Table 25 below.

Table 25- SNaPshot™ Results for C/A heteroplasmic 315 Cells and Controls, 16069 Primers

Sample Name	Allele 1	Size 1	Height 1	Peak Area 1	Allele 2	Size 2	Height 2	Peak Area 2	Allele 3	Size 3	Height 3	Peak Area 3
PE-315									T	34.49	8341	62476
AN-315					C	31.91	295	1999				
CAN-315												
1-315					C	31.91	531	3662				
2-315												
3-315					C	31.96	47	374	T	34.5	425	2879
4-315					C	31.89	646	4400				
5-315					C	31.88	862	5844				
6-315					C	31.95	263	1787				
7-315					C	31.89	2192	14768				
8-315												
9-315					C	31.91	260	1769				
10-315												
11-315					C	31.89	7769	53231				
12-315					C	31.91	1338	9002				
13-315					C	31.91	787	5317				
14-315					C	31.91	2562	17195				
15-315					C	31.99	622	4213				
16-315												
17-315					C	31.91	676	4589				
18-315												
19-315												
20-315					C	31.88	857	5831				
21-315					C	31.91	449	3035				
22-315					C	31.97	1452	9805				
23-315					C	31.83	260	1769				
24-315					C	31.91	109	747				
25-315												
26-315					C	32	193	1337				
27-315	A	30.95	429	2853	C	31.89	700	5427				
28-315												
29-315												
30-315												

Table 25- SNaPshot™ Results for C/A heteroplasmic 315 Cells and Controls, 16069 Primers (continued)

Sample Name	Allele 1	Size 1	Height 1	Peak Area 1	Allele 2	Size 2	Height 2	Peak Area 2	Allele 3	Size 3	Height 3	Peak Area 3
31-315	A	30.98	85	580	C	31.91	133	1041				
32-315	A	30.81	57	388	C	32.02	101	778				
33-315	A	30.82	107	713	C	31.89	142	1125				
34-315												
35-315	A	30.98	171	1130	C	31.91	243	2117				
36-315	A	30.81	51	349	C	31.89	75	617				
37-315												
38-315	A	30.98	227	1462	C	31.91	305	2348				
39-315	A	30.81	158	1035	C	31.75	193	1488				
40-315												
41-315	A	30.84	75	498	C	31.91	88	716				
42-315	A	31.04	55	354	C	31.96	67	561				
43-315	A	30.88	264	1766	C	31.81	335	2605				
44-315												
45-315	A	30.8	107	716	C	31.85	108	855				
46-315	A	30.88	278	1866	C	31.8	340	2713				
47-315	A	30.98	488	3247	C	31.91	645	5109				
48-315												
49-315					C	31.83	36	281				
50-315	A	30.91	42	283	C	31.83	50	386				
31-288c	A	30.9	48	310	C	31.82	59	461				
32-288c	A	30.88	123	821	C	31.81	182	1447				
33-288c	A	30.91	376	2553	C	31.83	663	5052				
34-288c	A	30.81	131	851	C	31.75	168	1335				
35-288c												
31-329c	A	30.98	58	373	C	31.91	74	582				
32-329c	A	30.88	155	981	C	31.84	184	1421				
33-329c												
34-329c	A	30.81	113	729	C	31.85	171	1362				
35-329c												

Using conditional formatting within the Microsoft® Excel 2010 software, the negative control threshold of 295 RFU was applied to the data and any ~32 bp C peak heights less than or equal to this threshold were labeled as red, and any ~32 bp C peak heights above this threshold were labeled as green. As the data was analyzed, it was seen that fifteen of the fifty samples and three of the ten controls gave no data, and these sample names were changed from black to red text (samples 2, 8, 10, 16, 18, 19, 25, 28, 29, 30, 34, 37, 40, 44, 48, and control samples 35-288c, 33-329c, and 35-329c from the 315 cells). For the remaining samples, it was seen that seventeen of the samples were displaying C peaks below the negative control threshold of 295 RFU for the 32 bp sized peaks. For the control cells, six of the cells were also below this negative control threshold.

However, a second issue was identified in the data as well. An additional T peak occurred in one random sample. When this T peak occurred, it was seen to be around 34 bp in size. Because the amplification negative sample did not display a T peak, no negative control threshold for the T peak could be set, and therefore any visible T peak around 34 bp would be deemed a valid peak. When this was applied to the data, one sample that was seen to contain a below-threshold C peak around 32 bp also contained a valid T peak around 34 bp: sample 3. The analysis parameters for this sample will be addressed in the discussion.

A third issue was occurring in the data as well. An additional A peak occurred in fifteen random samples and seven control samples. When this A peak occurred, it was approximately 31 bp in size. Because the amplification negative sample did not display an A peak, no negative control threshold for the A peak could be set, and therefore any visible A peak around 31 bp would be deemed a valid peak. When this was applied to the data, ten samples and six controls that were seen to contain a below-threshold C peak around 32 bp also contained a valid A peak

around 31 bp: samples 31, 32, 33, 35, 36, 39, 41, 42, 45, 50, and control samples 31-288c, 32-288c, 34-288c, 31-329c, 32-329c, and 34-329c from the 315 cells. The analysis parameters for these seventeen samples will be addressed in the discussion. Samples 2, 6, 8, 9, 10, 16, 18, 19, 23, 24, 25, 26, 28, 29, 30, 34, 37, 40, 44, 48, and 49 were deemed fully invalid, along with the control samples 35-288c, 33-329c and 35-329c. The remaining twenty-nine viable cell samples and seven controls samples were deemed valid for interpretation.

SNaPshot™ Results of Sample 293 Cell Analysis

Upon analysis, the SNaPshot™ internal positive and negative controls were seen to pass. The amplification positive and negative controls from the 293 cell samples were first analyzed. For reference, the HL60 DNA type at 16223 should be a C base around 29 bp. The results are presented in Table 26:

Table 26- Control Sample Results for 293 Cells in SNaPshot™

Control Sample	Primer Set	SNP Base Detected	Base-pair size	Peak Height (in RFU)
PE	16223	C	29.5	8073
AN	16223	C	29.62	593
		A	31.68	363
		T	32.63	164
CAN	16223	C	29.59	705
		A	31.62	460
		T	32.68	246

The positive control passed, and the negative control threshold was established as 593 RFU for ~29 bp sized C peaks based on the peak seen in the amplification negative control. A second peak was also seen in the AN control around 31 bp, and this A peak had a peak height of 363 RFU. A third peak, a T, was also seen in the AN control around 32 bp with a peak height of 164 RFU. Since these additional peaks were unexpected, the size and peak heights of the second and third peaks were noted. The cycling amplification negative control (CAN) for these samples

did show peaks as well, but since none of the samples were amplified with additional cycles, this control would not be analyzed any further.

The fifty single cell samples for 293 as well as the ten control cell samples were then analyzed. Based on the gross tissue analysis, the 16223 extension primer showed a C/T heteroplasmy. While this C/T heteroplasmy was expected based on the gross tissue analysis, C/A/T peaks were instead found (see the discussion for analysis of this issue). The results of the SNaPshot™ SNP typing are displayed in Table 27 below.

Table 27- SNaPshot™ Results for C/T heteroplasmic 293 Cells and Controls, 16223 Primers

Sample Name	Allele 1	Size 1	Height 1	Peak Area 1	Allele 2	Size 2	Height 2	Peak Area 2	Allele 3	Size 3	Height 3	Peak Area 3
PE-293	C	29.5	8073	52954								
AN-293	C	29.62	593	3901	A	31.68	363	2625	T	32.63	164	1077
CAN-293	C	29.59	705	4487	A	31.62	460	3347	T	32.68	246	1879
1-293	C	29.64	497	3146	A	31.71	292	2110	T	32.78	138	886
2-293	C	29.47	300	1929	A	31.68	214	1518	T	32.76	87	581
3-293	C	29.64	2941	18709	A	31.71	2043	13866	T	32.78	866	5620
4-293	C	29.64	781	4981	A	31.71	627	4472	T	32.78	280	2025
5-293	C	29.64	925	5845	A	31.7	745	5312	T	32.78	400	3314
6-293	C	29.58	1293	8137	A	31.62	638	4537	T	32.68	336	2177
7-293	C	29.59	773	4931	A	31.76	496	3598	T	32.68	264	2001
8-293	C	29.64	746	4728	A	31.71	432	3093	T	32.78	207	1342
9-293	C	29.64	491	3031	A	31.7	571	4021	T	32.64	316	2771
10-293	C	29.64	850	5436	A	31.71	532	3761	T	32.78	241	1584
11-293	C	29.59	795	5098	A	31.76	634	4561	T	32.82	350	2939
12-293	C	29.59	1349	8686	A	31.62	744	5502	T	32.68	442	3463
13-293	C	29.5	1330	8337	A	31.71	819	5942	T	32.65	399	2704
14-293	C	29.47	847	5734	A	31.54	479	3742	T	32.62	229	1659
15-293	C	29.57	614	3961	A	31.61	323	2320	T	32.67	163	1062
16-293	C	29.59	918	5838	A	31.62	540	3848	T	32.68	270	1842
17-293	C	29.59	455	2905	A	31.62	324	2313	T	32.68	141	902
18-293	C	31.89	1005	7467	A	31.89	893	5936	T	32.68	1005	10797
19-293	C	29.59	2826	18072	A	31.76	219	1551	T	32.81	484	4041
20-293	C	29.59	1057	6692	A	31.76	700	5004	T	32.68	344	2520
21-293	C	29.5	989	6228	A	31.7	530	3754	T	32.64	348	2786
22-293	C	29.5	1322	8470	A	31.71	696	5029	T	32.65	345	2254
23-293	C	29.64	499	3184	A	31.71	405	2856	T	32.65	190	1485
24-293	C	29.5	630	3938	A	31.57	403	2872	T	32.64	210	1525
25-293	C	29.5	1590	10121	A	31.7	283	1927	T	32.64	351	2848
26-293	C	29.68	717	4828	A	31.68	342	2565	T	32.72	185	1246
27-293	C	29.59	980	6366	A	31.63	606	4357	T	32.69	295	2103
28-293	C	29.59	1208	7920	A	31.63	592	4496	T	32.69	307	2039
29-293	C	29.5	1817	11352	A	31.57	180	1287	T	32.64	282	2093
30-293	C	29.64	1915	12231								

Table 27- SNaPshot™ Results for C/T heteroplasmic 293 Cells and Controls, 16223 Primers (continued)

Sample Name	Allele 1	Size 1	Height 1	Peak Area 1	Allele 2	Size 2	Height 2	Peak Area 2	Allele 3	Size 3	Height 3	Peak Area 3
31-293	C	29.67	1963	12675					T	32.59	126	1104
32-293	C	29.59	2830	17792					T	32.55	652	4612
33-293	C	29.5	1644	10403								
34-293	C	29.53	1242	7943								
35-293	C	29.64	2379	15048								
36-293	C	29.5	1634	10350					T	32.65	1299	8570
37-293	C	29.64	2629	16568					T	32.65	378	2817
38-293	C	29.5	1983	12345								
39-293	C	29.62	2866	18006					T	32.7	229	1818
40-293	C	29.58	3281	20385					T	32.55	699	4891
41-293	C	29.5	3347	21216								
42-293	C	29.59	2180	13652								
43-293	C	29.59	3276	20860								
44-293	C	29.5	2625	16594					T	32.65	247	1957
45-293	C	29.56	1836	11821					T	32.53	302	2225
46-293	C	29.59	7100	45399								
47-293	C	29.5	1646	10534								
48-293									T	32.55	4879	31839
49-293	C	29.59	3676	23208					T	32.55	408	3075
50-293	C	29.47	3083	19659					T	32.63	488	3556
36-288c	C	29.47	1819	11524								
37-288c	C	29.59	1426	9108					T	32.69	426	2966
38-288c	C	29.59	2924	18269					T	32.55	346	2580
39-288c	C	29.64	5899	37078					T	32.65	395	3153
40-288c	C	29.56	2567	16000								
36-329c	C	29.64	1718	10915					T	32.65	595	3986
37-329c	C	29.56	3527	22009					T	32.61	902	6144
38-329c	C	29.5	2082	13059					T	32.51	691	4689
39-329c	C	29.5	2490	15666					T	32.65	593	4062
40-329c	C	29.61	2829	17540					T	32.57	368	2740

Using conditional formatting within the Microsoft® Excel 2010 software, the negative control threshold of 593 RFU was applied to the data and any ~29 bp C peak heights less than or equal to this threshold were labeled as red, and any ~29 bp C peak heights above this threshold were labeled as green. As the data was analyzed, it was seen that all of the samples and controls contained data. Five of the samples displayed peaks below the negative control threshold of 593 RFU for the 29 bp sized C peaks. For the control cells, none of the cells were below this negative control threshold.

However, a second issue was occurring in the data as well. As was seen in the amplification negative control, an additional A peak was occurring in twenty-nine samples. When this A peak occurred, it was approximately 32 bp in size. Therefore, for the samples affected by this A peak, additional conditional formatting was applied to the table to establish the negative control threshold of 363 RFU for the A peak seen in the amplification negative control. When this was applied to the data, only eight samples were below this 363 RFU negative control threshold, and three of these same samples were also already below the 29 bp negative control threshold as well for the C peaks, and therefore were already ineligible for analysis (samples 1, 2, and 17 from the 293 cells). The remaining twenty-one samples that contained an A peak were compared against the samples that also contained a viable C peak at ~29 bp above the negative control threshold, and nineteen samples were seen to pass both negative control thresholds (samples 9 and 23 contained a below-threshold C peak around 29 bp, but an above threshold A peak around 32 bp).

A third issue was occurring in the data as well. As it was noted in the amplification negative control, an additional T peak was occurring in forty samples and eight of the control cell samples. When this T peak occurred, it was approximately 33 bp in size. Therefore, for the

samples affected by this T peak, additional conditional formatting was applied to the table to establish the negative control threshold of 164 RFU for the T peak seen in the amplification negative control. When this was applied to the data, only five samples were seen to be below this 164 RFU negative control threshold, and four of these same samples were also already below the 29 bp negative control threshold as well for the C peaks or below the 32 bp negative control threshold for the A peaks, and therefore were already ineligible for analysis (samples 1, 2, 15, and 17 for the 293 cells). The remaining thirty-five samples that contained a T peak were compared against the samples that also contained a viable C peak around 29 bp above the negative control threshold, and thirty-four samples and eight control cell samples were seen to pass both negative control thresholds. Samples 1, 2, and 17 were deemed fully invalid, leaving the remaining forty-seven viable cell samples and ten controls samples valid for interpretation.

SNaPshot™ Results of Sample 314 Cell Analysis

Upon analysis, the SNaPshot™ internal positive and negative controls were seen to pass. The amplification positive and negative controls from the 314 cell samples were first analyzed. For reference, the HL60 DNA type at 16223 should be a C base around 29 bp. The results are presented in Table 28:

Table 28- Control Sample Results for 314 Cells in SNaPshot™

Control Sample	Primer Set	SNP Base Detected	Base-pair size	Peak Height (in RFU)
PE	16223	C	29.66	5806
AN	16223	C	29.64	1584
CAN	16223	C	29.67	1698

The positive control passed, and the negative control threshold was established as 1584 RFU for ~29 bp sized peaks based on the peak seen in the amplification negative control. The cycling amplification negative control (CAN) for these samples did show peaks as well, but

since none of the samples were amplified with additional cycles, this control would not be analyzed any further.

The fifty single cell samples for 314 as well as the ten control cell samples were then analyzed. Based on the gross tissue analysis, the 16223 extension primer showed a C/T heteroplasmy. While this C/T heteroplasmy was expected based on the gross tissue analysis, G/C/T peaks were instead found (see the discussion for analysis of this issue). The results of the SNaPshot™ SNP typing are displayed in Table 29, below.

Table 29- SNaPshot™ Results for C/T heteroplasmic 314 Cells and Controls, 16223 Primers

Sample Name	Allele 1	Size 1	Height 1	Peak Area 1	Allele 2	Size 2	Height 2	Peak Area 2	Allele 3	Size 3	Height 3	Peak Area 3
PE-314					C	29.66	5806	36633				
AN-314					C	29.64	1584	10039				
CAN-314					C	29.67	1698	10677				
1-314					C	29.59	2750	17414	T	32.68	128	1154
2-314					C	29.58	1212	7724				
3-314					C	29.77	1224	7901	T	32.77	6318	42284
4-314					C	29.64	2699	17026				
5-314					C	29.63	1924	12129	T	32.64	558	3770
6-314					C	29.49	2254	14343	T	32.64	557	3841
7-314					C	29.58	1876	11795				
8-314					C	29.61	2723	17253	T	32.7	347	2591
9-314					C	29.64	3020	19048				
10-314					C	29.67	1123	7194	T	32.85	49	480
11-314					C	29.7	2425	15516	T	32.74	564	4056
12-314					C	29.72	3655	23073	T	32.68	467	3452
13-314					C	29.66	2222	13872	T	32.66	347	2576
14-314					C	29.58	3175	19698	T	32.62	671	4693
15-314					C	29.58	2233	14011	T	32.76	69	758
16-314					C	29.67	1801	11261				
17-314					C	29.58	1200	7481	T	32.63	811	5351
18-314					C	29.64	408	2584				
19-314					C	29.61	2269	14283				
20-314					C	29.67	2125	13177	T	32.66	622	4241
21-314					C	29.55	3338	20888				
22-314					C	29.58	3100	19417	T	32.63	734	4988
23-314					C	29.64	1738	10876	T	32.78	640	4407
24-314					C	29.67	2117	13197	T	32.66	76	747
25-314					C	29.67	3520	21823	T	32.66	673	4846
26-314					C	29.61	1654	10441	T	32.7	273	2034
27-314					C	29.75	1908	12154				
28-314					C	29.67	2561	15871	T	32.66	401	2804
29-314					C	29.55	1752	10937	T	32.74	1044	6948
30-314					C	29.58	1915	11959				

Table 29- SNaPshot™ Results for C/T heteroplasmic 314 Cells and Controls, 16223 Primers (continued)

Sample Name	Allele 1	Size 1	Height 1	Peak Area 1	Allele 2	Size 2	Height 2	Peak Area 2	Allele 3	Size 3	Height 3	Peak Area 3
31-314					C	29.5	2106	13219				
32-314	G	28.21	348	2163	C	29.61	2962	18511				
33-314					C	29.67	605	3841	T	32.66	3309	21450
34-314					C	29.64	964	6185	T	32.64	443	2987
35-314					C	29.56	3594	22434				
36-314					C	29.67	1888	11675	T	32.66	333	2334
37-314					C	29.7	1720	10641				
38-314					C	29.66	4621	28721				
39-314					C	29.5	1233	7757				
40-314					C	29.55	2590	15995				
41-314					C	29.66	2163	13578	T	32.66	202	1584
42-314					C	29.56	663	4270	T	32.61	4157	27219
43-314					C	29.53	3195	20340				
44-314					C	29.63	1683	10640	T	32.64	276	2017
45-314					C	29.64	3045	18936				
46-314					C	29.5	3254	20545	T	32.64	323	2531
47-314					C	29.5	943	6006	T	32.64	586	3939
48-314					C	29.67	1483	9390				
49-314					C	29.61	2404	15107	T	32.7	1025	6973
50-314					C	29.58	2059	12981				
41-288c					C	29.5	2043	12998				
42-288c	G	28.22	250	1567	C	29.62	3051	19288	T	32.58	463	3473
43-288c					C	29.61	2879	17976	T	32.57	705	4898
44-288c					C	29.67	2669	16542				
45-288c					C	29.52	2314	14544				
41-329c	G	28.12	1217	7535	C	29.55	1688	10431				
42-329c					C	29.66	2340	14512	T	32.66	328	2462
43-329c					C	29.64	2349	14904				
44-329c					C	29.64	1701	10626	T	32.72	607	4186
45-329c					C	29.66	3171	19744	T	32.66	821	5663

Using conditional formatting within the Microsoft® Excel 2010 software, the negative control threshold of 1584 RFU was applied to the data and any ~29 bp C peak heights less than or equal to this threshold were labeled as red, and any ~29 bp C peak heights above this threshold were labeled as green. As the data was analyzed, all of the samples displayed data. It was noted that eleven of the samples were displaying peaks below the negative control threshold of 1584 RFU for the 29 bp sized C peaks: samples 2, 3, 10, 17, 18, 33, 34, 39, 42, 47, and 48. For the control cells, none of the cells were below this negative control threshold. This left 39 viable cell samples and all ten control samples available for interpretation.

A second issue was present in the data as well. An additional T peak displayed in twenty-nine random cell samples and five control samples. When this T peak occurred, it was approximately 33 bp in size. Because the amplification negative sample did not display a T peak, no negative control threshold for the T peak could be set, and therefore any visible T peak around 32 bp would be deemed a valid peak. When this was applied to the data, seven samples that were seen to contain a below-threshold C peak around 29 bp also contained a valid T peak around 32 bp: samples 3, 10, 17, 33, 34, 42, and 47. The analysis parameters for these seven samples will be addressed in the discussion. Therefore, by eliminating the samples that only contained a single below-threshold C peak at 29 bp, the remaining forty-six viable cell samples and ten controls samples were deemed valid for interpretation.

A third issue was also occurring in the data. An additional G peak displayed in one sample and two control samples. When this G peak occurred, it was seen to be around 28 bp in size. Because the amplification negative sample did not display a G peak, no negative control threshold for the G peak could be set, and therefore any visible G peak around 28 bp would be deemed a valid peak. When this was applied to the data, all of the samples that were seen to

contain a G peak around 28 bp also contained other valid peaks as well. Therefore, no additional samples were eliminated from the set, samples 2, 18, 39, and 48 were deemed fully invalid, and the remaining forty-six viable cell samples and ten controls samples were deemed valid for interpretation.

SNaPshot™ Results of Sample 289 Cell Analysis

Upon analysis, the SNaPshot™ internal positive and negative controls passed. The amplification positive and negative controls from the 289 cell samples were first analyzed. For reference, the HL60 DNA type at 16324 should be a T base around 34 bp. The results are presented in Table 30:

Table 30- Control Sample Results for 289 Cells in SNaPshot™

Control Sample	Primer Set	SNP Base Detected	Base-pair size	Peak Height (in RFU)
PE	16324	T	34.36	8433
AN	16324	T	34.26	494
CAN	16324	T	34.25	1998

The positive control passed, and the negative control threshold was established as 494 RFU for ~34 bp sized peaks based on the peak seen in the amplification negative control. The cycling amplification negative control (CAN) for these samples did show peaks as well, but since none of the samples were amplified with additional cycles, this control would not be analyzed any further.

The fifty single cell samples for 289 as well as the ten control cell samples were then analyzed. Based on the gross tissue analysis, the 16324 extension primer showed a G/T heteroplasmy. The results of the SNaPshot™ SNP typing are displayed in Table 31, below.

Table 31- SNaPshot™ Results for G/T heteroplasmic 289 Cells and Controls, 16324 Primers

Sample Name	Allele 1	Size 1	Height 1	Peak Area 1	Allele 2	Size 2	Height 2	Peak Area 2
PE-289					T	34.36	8433	59053
AN-289					T	34.26	494	2984
CAN-289					T	34.25	1998	11971
1-289					T	34.24	1370	8237
2-289					T	34.35	1006	6137
3-289					T	34.25	2428	14711
4-289					T	34.26	1592	9499
5-289					T	34.22	7656	45854
6-289					T	34.37	2787	17167
7-289					T	34.24	1823	10958
8-289					T	34.24	1802	10812
9-289					T	34.24	1512	9026
10-289					T	34.24	1048	6330
11-289					T	34.25	2093	12609
12-289					T	34.36	2766	16619
13-289					T	34.22	1440	8592
14-289					T	34.24	1461	8687
15-289					T	34.24	1134	6782
16-289					T	34.36	2079	12537
17-289					T	34.22	4256	25476
18-289					T	34.23	1022	6210
19-289	G	31.49	30	292	T	34.24	4333	25866
20-289					T	34.22	1477	8852
21-289					T	34.22	1743	10529
22-289					T	34.24	476	2841
23-289					T	34.24	1826	10941
24-289					T	34.35	1874	11248
25-289					T	34.23	1126	6737
26-289					T	34.24	2018	12151
27-289					T	34.24	2944	18009
28-289					T	34.24	2645	15932
29-289					T	34.22	1391	8320
30-289					T	34.22	2048	12187

Table 31- SNaPshot™ Results for G/T heteroplasmic 289 Cells and Controls, 16324 Primers (continued)

Sample Name	Allele 1	Size 1	Height 1	Peak Area 1	Allele 2	Size 2	Height 2	Peak Area 2
31-289					T	34.24	975	5814
32-289					T	34.24	2078	12440
33-289					T	34.22	710	4247
34-289					T	34.23	466	2782
35-289					T	34.23	2643	15917
36-289					T	34.36	2701	16183
37-289					T	34.23	2525	15082
38-289					T	34.22	7587	44751
39-289					T	34.22	2168	12925
40-289					T	34.22	1372	8249
41-289					T	34.23	1249	7390
42-289					T	34.24	1336	8167
43-289					T	34.24	1959	11823
44-289					T	34.24	1278	7585
45-289					T	34.23	1683	10019
46-289					T	34.24	2062	12339
47-289					T	34.23	1891	11299
48-289					T	34.23	834	4988
49-289					T	34.23	1520	9045
50-289					T	34.22	1511	9062
46-288c					T	34.23	1861	11240
47-288c					T	34.24	3446	20830
48-288c					T	34.24	2141	13036
49-288c					T	34.24	6964	41887
50-288c					T	34.22	2695	16247
46-329c					T	34.22	2110	12568
47-329c					T	34.22	2640	15713
48-329c					T	34.22	2097	12497
49-329c					T	34.22	1346	8021
50-329c					T	34.24	3035	18538

Using conditional formatting within the Microsoft® Excel 2010 software, the negative control threshold of 494 RFU was applied to the data and any ~34 bp peak heights less than or equal to this threshold were labeled as red, and any ~34 bp peak heights above this threshold were labeled as green. As the data was analyzed, it was seen that all of the fifty samples and all ten of controls displayed data. It was seen that two of the samples were displaying peaks below the negative control threshold of 494 RFU for the 34 bp sized T peaks (samples 22 and 34 for the 289 cells). For the control cells, none of the cells were below this negative control threshold. This left forty-eight viable cell samples and ten controls samples available for interpretation.

A second issue was noted in the data as well. An additional G peak occurred in one samples. When this G peak occurred, it was approximately 31 bp in size. Because the amplification negative sample did not display a G peak, no negative control threshold for the G peak could be set, and therefore any visible G peak around 31 bp would be deemed a valid peak. When this was applied to the data, the sample that contained the G peak at 31 bp contained a valid T peak as well at 34 bp, which was sample 19. Therefore, by eliminating the samples that contained the below-threshold T peak at 34 bp, the remaining forty-eight viable cell samples and ten controls samples were deemed valid for interpretation.

SNaPshot™ Results of Sample 349 Cell Analysis

Upon analysis, the SNaPshot™ internal positive and negative controls passed. The amplification positive and negative controls from the 349 cell samples were first analyzed. For reference, the HL60 DNA type at 16324 should be a T base around 34 bp. The results are presented in Table 32:

Table 32- Control Sample Results for 349 Cells in SNaPshot™

Control Sample	Primer Set	SNP Base Detected	Base-pair size	Peak Height (in RFU)
PE	16324	T	34.88	8346
AN	16324	G	31.42	34
		T	34.64	1156
CAN	16324	G	31.7	41
		T	34.76	498

The positive control passed, and the negative control threshold was established as 1156 RFU for ~34 bp sized peaks based on the T peak seen in the amplification negative control. A second peak was also seen around 31 bp, and this G peak had a peak height of 34 RFU. Since this additional peak was unexpected, the size and peak height of the second peak was noted. The cycling amplification negative control (CAN) for these samples did show peaks as well, but since none of the samples were amplified with additional cycles, this control would not be analyzed any further.

The fifty single cell samples for 349 as well as the ten control cell samples were then analyzed. Based on the gross tissue analysis, the 16324 extension primer showed an A/T heteroplasmy. While this A/T heteroplasmy was expected based on the gross tissue analysis, G/T peaks were found instead (see the discussion for analysis of this issue). The results of the SNaPshot™ SNP typing are displayed in Table 33, below.

Table 33- SNaPshot™ Results for A/T heteroplasmic 349 Cells and Controls, 16324 Primers

Sample Name	Allele 1	Size 1	Height 1	Peak Area 1	Allele 2	Size 2	Height 2	Peak Area 2
PE_349					T	34.88	8346	63484
AN_349	G	31.42	34	429	T	34.64	1156	7615
CAN_349	G	31.7	41	541	T	34.76	498	3255
1_349					T	34.76	706	4652
2_349					T	34.64	794	5251
3_349	G	31.55	43	532	T	34.76	1444	9580
4_349	G	31.5	31	438	T	34.63	1040	6884
5_349	G	31.65	35	341	T	34.64	381	2500
6_349	G	31.58	36	329	T	34.75	1118	7344
7_349	G	31.71	31	290	T	34.75	1701	11135
8_349					T	34.76	511	3354
9_349	G	31.5	34	464	T	34.76	1183	7707
10_349					T	34.76	1024	6836
11_349	G	31.6	40	411	T	34.64	1015	6743
12_349	G	31.43	35	355	T	34.52	1864	12225
13_349	G	31.28	34	302	T	34.52	1019	6687
14_349					T	34.51	855	5566
15_349	G	31.28	30	294	T	34.52	716	4730
16_349	G	31.58	30	303	T	34.63	1317	8579
17_349					T	34.39	1009	6699
18_349					T	34.52	200	1285
19_349	G	31.37	42	511	T	34.63	786	5060
20_349					T	34.52	1406	9167
21_349	G	31.32	33	317	T	34.63	954	6192
22_349	G	31.3	37	332	T	34.52	418	2728
23_349	G	31.45	32	302	T	34.63	1055	6811
24_349								
25_349					T	34.52	1013	6523
26_349					T	34.51	490	3180
27_349	G	31.6	39	497	T	34.65	1537	10216
28_349	G	31.25	35	356	T	34.51	867	5666
29_349	G	31.53	30	290	T	34.64	1349	8842
30_349					T	34.64	637	4194

Table 33- SNaPshot™ Results for A/T heteroplasmic 349 Cells and Controls, 16324 Primers (continued)

Sample Name	Allele 1	Size 1	Height 1	Peak Area 1	Allele 2	Size 2	Height 2	Peak Area 2
31_349					T	34.51	579	3778
32_349	G	31.27	31	283	T	34.51	1402	9137
33_349					T	34.63	950	6168
34_349					T	34.51	477	3116
35_349	G	31.43	38	421	T	34.64	1474	9749
36_349	G	31.58	31	370	T	34.53	1070	6903
37_349	G	31.37	32	297	T	34.51	1434	9307
38_349	G	31.46	37	359	T	34.51	1049	6738
39_349	G	31.37	33	421	T	34.51	1152	7483
40_349	G	31.51	31	290	T	34.64	545	3572
41_349	G	31.24	34	432	T	34.53	673	4633
42_349					T	34.51	662	4307
43_349	G	31.46	36	442	T	34.64	4201	28113
44_349	G	31.33	31	280	T	34.63	1330	8701
45_349					T	34.51	1658	10795
46_349	G	31.43	34	359	T	34.52	169	1055
47_349					T	34.51	407	2591
48_349					T	34.51	298	1913
49_349	G	31.3	36	528	T	34.52	1541	10002
50_349					T	34.51	697	4582
21_288c	G	31.38	30	286	T	34.51	1291	8375
22_288c	G	31.43	33	431	T	34.52	1118	7314
23_288c					T	34.51	1264	8198
24_288c	G	31.5	30	355	T	34.51	1759	11385
25_288c	G	31.38	30	294	T	34.51	1079	6949
21_329c								
22_329c	G	31.25	35	354	T	34.51	2183	14018
23_329c					T	34.51	2637	17159
24_329c					T	34.52	790	5146
25_329c	G	31.25	37	467	T	34.51	1160	7397

Using conditional formatting within the Microsoft® Excel 2010 software, the negative control threshold of 1156 RFU was applied to the data and any ~34 bp T peak heights less than or equal to this threshold were labeled as red, and any ~34 bp peak heights above this threshold were labeled as green. As the data was analyzed, it was seen that only one of the fifty samples and one of the ten controls gave no data, and these sample names were changed from black to red text (sample 24, and control sample 21-329c from the 329 cells). For the remaining samples, it was seen that thirty-four of the samples were displaying peaks below the negative control threshold of 1156 RFU for the 34 bp sized T peaks. For the control cells, three of the cells were also below this negative control threshold. This left sixteen viable cell samples and six control samples available for interpretation.

A second issue was occurring in the data as well. As it was seen in the amplification negative control, an additional G peak was detected in thirty random samples and six control samples. When this G peak occurred, it was approximately 31 bp in size. Therefore, for the samples affected by this G peak, additional conditional formatting was applied to the data to establish the negative control threshold of 34 RFU for the G peak seen in the amplification negative control. When this was applied to the data, seventeen of the thirty samples and four of the six controls were seen to be below this 34 RFU negative control threshold. Ten of these samples and two of the controls already contained a below-threshold T peak around 34 bp, and were marked as ineligible for analysis. It was seen that six of the samples and two of the controls contained above-threshold signals for both the G peaks around 31 bp and the T peaks around 34 bp: samples 3, 12, 27, 35, 43, 49, and control samples 22- 329c and 25-329c from the 349 cells. It was seen in the data that seven samples and two controls contained a valid T peak around 34 bp while also displaying a below-threshold G peak around 31 bp: samples 7, 9, 16, 29,

32, 37, 44, and controls 21-288c and 24-288c from the 349 cells. It was also noted that seven samples contained a valid G peak around 31 bp with an invalid T peak: samples 5, 6, 11, 19, 22, 28, and 38. Finally, it was seen in the data that only two samples and two controls contained a valid T peak around 34 bp with no G peak displayed in the data: samples 20, 45, and controls 23-288c and 23-329c from the 349 cells. The twenty-two viable cell samples (3, 5, 6, 7, 9, 11, 12, 16, 19, 20, 22, 27, 28, 29, 32, 35, 37, 38, 43, 44, 45, and 49) and six controls samples (22-329c, 23-329c, 25-329c, 21-288c, 23-288c and 24-288c) were deemed valid for interpretation. The remainder of the samples will be addressed in the discussion.

SNaPshot™ Results of Sample 339 Cell Analysis

Upon analysis, the SNaPshot™ internal positive and negative controls were seen to pass. The amplification positive and negative controls from the 339 cell samples were first analyzed. For reference, the HL60 DNA type at 16390 should be a C base around 36 bp. The results are presented in Table 34:

Table 34- Control Sample Results for 339 Cells in SNaPshot™

Control Sample	Primer Set	SNP Base Detected	Base-pair size	Peak Height (in RFU)
PE	16390	C	36.38	4779
AN	16390	C	36.23	272
		T	39.19	128
CAN	16390	C	36.21	128

The positive control passed, and the negative control threshold was established as 272 RFU for ~36 bp sized peaks based on the peak seen in the amplification negative control. A second peak was also seen around 39 bp, and this T peak had a peak height of 128 RFU. Since this additional peak was unexpected, the size and peak height of the second peak was noted. The cycling amplification negative control (CAN) for these samples did show peaks as well, but

since none of the samples were amplified with additional cycles, this control would not be analyzed any further.

The fifty single cell samples for 339 as well as the ten control cell samples were then analyzed. Based on the gross tissue analysis, the 16390 extension primer showed a C/T heteroplasmy. The results of the SNaPshot™ SNP typing are displayed in Table 35, below.

Table 35- SNaPshot™ Results for C/T heteroplasmic 339 Cells and Controls, 16390 Primers

Sample Name	Allele 1	Size 1	Height 1	Peak Area 1	Allele 2	Size 2	Height 2	Peak Area 2
PE-339	C	36.38	4779	27931				
AN-339	C	36.23	272	1641	T	39.19	128	770
CAN-339	C	36.21	128	759				
1-339	C	36.23	724	4297				
2-339	C	36.36	196	1202				
3-339	C	36.1	1241	7290				
4-339	C	36.23	958	5587	T	39.19	234	1418
5-339	C	36.24	504	2921				
6-339	C	36.24	1100	6405				
7-339	C	36.23	842	5068				
8-339	C	36.22	361	2100				
9-339	C	36.36	893	5274				
10-339	C	36.23	660	3926	T	39.16	237	1415
11-339	C	36.22	348	2129				
12-339	C	36.37	795	4729				
13-339	C	36.22	159	935	T	39.16	231	1392
14-339	C	36.23	684	4013				
15-339	C	36.23	3122	18043				
16-339	C	36.23	1250	7270				
17-339	C	36.22	561	3309				
18-339	C	36.36	210	1248	T	39.16	115	694
19-339	C	36.23	171	1051				
20-339	C	36.23	501	2982				
21-339	C	36.23	613	3572				
22-339	C	36.22	238	1440	T	39.16	257	1640
23-339	C	36.24	1553	9217				
24-339	C	36.23	562	3266				
25-339	C	36.37	532	3151				
26-339	C	36.23	1625	9408	T	39.05	149	893
27-339	C	36.22	202	1202				
28-339	C	36.24	1085	6272	T	39.2	111	679
29-339	C	36.24	492	2881				
30-339	C	36.37	689	3995				

Table 35- SNaPshot™ Results for C/T heteroplasmic 339 Cells and Controls, 16390 Primers (continued)

Sample Name	Allele 1	Size 1	Height 1	Peak Area 1	Allele 2	Size 2	Height 2	Peak Area 2
31-339	C	36.37	413	2433	T	39.18	569	3421
32-339	C	36.23	356	2085				
33-339	C	36.37	254	1491	T	39.18	92	547
34-339	C	36.38	1018	5962				
35-339	C	36.37	866	5076				
36-339	C	36.35	661	4017				
37-339	C	36.36	550	3187				
38-339	C	36.36	571	3357				
39-339	C	36.37	1038	5968				
40-339	C	36.36	783	4604				
41-339	C	36.11	334	1972				
42-339	C	36.23	596	3513	T	39.05	231	1385
43-339	C	36.09	954	5677				
44-339	C	36.22	618	3636				
45-339	C	36.23	126	749				
46-339	C	36.22	371	2228				
47-339								
48-339	C	36.23	200	1194	T	39.05	137	813
49-339	C	36.1	826	4802				
50-339	C	36.23	528	3072				
51-288c	C	36.23	393	2286	T	39.03	75	474
52-288c	C	36.22	480	2898				
53-288c	C	36.36	201	1194	T	39.16	354	2103
54-288c	C	36.23	725	4205				
55-288c	C	36.22	351	2021				
51-329c	C	36.24	195	1147				
52-329c	C	36.36	254	1512				
53-329c	C	36.24	1028	5941				
54-329c	C	36.24	283	1661	T	39.2	38	246
55-329c	C	36.23	607	3488				

Using conditional formatting within the Microsoft® Excel 2010 software, the negative control threshold of 272 RFU was applied to the data and any ~36 bp C peak heights less than or equal to this threshold were labeled as red, and any ~36 bp C peak heights above this threshold were labeled as green. As the data was analyzed, it was seen that only one of the fifty samples gave no data, and these sample names were changed from black to red text (sample 47). For the remaining samples, it was seen that nine of the samples were displaying peaks below the negative control threshold of 272 RFU for the 36 bp sized C peaks. For the control cells, three of the cells were also below this negative control threshold. This left forty viable cell samples and seven controls samples available for interpretation.

However, a second issue was occurring in the data as well. As it was seen in the amplification negative control, an additional T peak displayed in eleven random samples and three control samples. When this T peak occurred, it was approximately 39 bp in size. Therefore, for the samples affected by this T peak, additional conditional formatting was applied to the data to establish the negative control threshold of 128 RFU for the T peak seen in the amplification negative control. When this was applied to the data, only three samples were seen to be below this 128 RFU negative control threshold, however two these same samples were also already below the 36 bp negative control threshold as well for the C peaks, and therefore were already ineligible for analysis. Three other samples that contained valid C peaks were seen to contain invalid T peaks, which now made these samples valid for interpretation. The remaining nine samples and three controls that contained a T peak were compared against the samples that also contained a viable C peak at ~36 bp above the negative control threshold, and five samples and one control were seen to pass both negative control thresholds: samples 4, 10, 26, 31, 42, and control 53-288c from the 339 cells. Samples 2, 18, 19, 27, 33, 45, and 47 , and controls 51-329c

and 52-329c were deemed fully invalid, which left a total of forty-three viable cell samples and eight controls samples that were deemed valid for interpretation.

SNaPshot™ Results of Sample 355 (16390) Cell Analysis

Upon analysis, the SNaPshot™ internal positive and negative controls were seen to pass. The amplification positive and negative controls from the 355 cell samples were first analyzed. For reference, the HL60 DNA type at 16390 should be a C base around 36 bp. The results are presented in Table 36:

Table 36- Control Sample Results for 355 (16390) Cells in SNaPshot™

Control Sample	Primer Set	SNP Base Detected	Base-pair size	Peak Height (in RFU)
PE	16390	C	36.24	5870
AN	16390	C	36.22	2677
		T	39.13	138
CAN	16390	C	36.22	177

The positive control passed, and the negative control threshold was established as 2677 RFU for ~36 bp sized C peaks based on the C peak seen in the amplification negative control. A second peak was also seen around 39 bp, and this T peak had a peak height of 138 RFU. Since this additional peak was unexpected, the size and peak height of the second peak was noted. The cycling amplification negative control (CAN) for these samples did show peaks as well, but since none of the samples were amplified with additional cycles, this control would not be analyzed any further.

The fifty single cell samples for 355 as well as the ten control cell samples were then analyzed. Based on the gross tissue analysis, the 16390 extension primer showed a C/T heteroplasmy. The results of the SNaPshot™ SNP typing are displayed in Table 37, below.

Table 37- SNaPshot™ Results for C/T heteroplasmic 355 Cells and Controls, 16390 Primers

Sample Name	Allele 1	Size 1	Height 1	Peak Area 1	Allele 2	Size 2	Height 2	Peak Area 2
PE-355	C	36.24	5870	34300				
AN-355	C	36.22	2677	15915	T	39.13	138	851
CAN-355	C	36.22	177	1029				
1-355	C	36.36	71	456				
2-355	C	36.35	160	965				
3-355	C	36.22	625	3695				
4-355	C	36.22	902	5320				
5-355	C	36.23	347	2061				
6-355	C	36.36	1066	6208				
7-355	C	36.23	330	1934				
8-355	C	36.22	156	930				
9-355	C	36.24	3096	17907				
10-355	C	36.23	513	3022	T	39.03	205	1251
11-355	C	36.21	572	3363	T	39.12	281	1714
12-355	C	36.23	351	2036	T	39.03	252	1526
13-355	C	36.24	767	4528	T	39.06	475	2842
14-355	C	36.22	142	852				
15-355	C	36.22	622	3643				
16-355	C	36.22	447	2615				
17-355	C	36.24	432	2519				
18-355	C	36.22	354	2100				
19-355	C	36.22	867	5059				
20-355	C	36.36	700	4150				
21-355	C	36.23	588	3386				
22-355	C	36.36	562	3307				
23-355	C	36.23	319	1877	T	39.16	284	1741
24-355	C	36.23	491	2894	T	39.05	132	782
25-355	C	36.09	369	2161	T	39.18	285	1705
26-355	C	36.35	141	836				
27-355	C	36.21	219	1314	T	38.99	309	1878
28-355	C	36.23	280	1644				
29-355	C	36.37	206	1194				
30-355	C	36.22	124	743	T	39.02	5952	35628

Table 37- SNaPshot™ Results for C/T heteroplasmic 355 Cells and Controls, 16390 Primers (continued)

Sample Name	Allele 1	Size 1	Height 1	Peak Area 1	Allele 2	Size 2	Height 2	Peak Area 2
31-355	C	36.35	453	2662				
32-355					T	39.12	129	777
33-355	C	36.36	402	2382				
34-355	C	36.22	274	1630				
35-355	C	36.22	792	4701				
36-355	C	36.24	745	4350				
37-355	C	36.36	280	1623				
38-355	C	36.36	178	1047	T	39.16	293	1759
39-355	C	36.1	425	2531				
40-355	C	36.24	289	1692				
41-355	C	36.36	1049	6034	T	39.17	310	1882
42-355	C	36.23	535	3143				
43-355	C	36.22	1144	6766	T	39.13	454	2747
44-355	C	36.22	1563	9062				
45-355	C	36.22	133	783				
46-355	C	36.09	338	2002				
47-355	C	36.22	209	1238				
48-355	C	36.1	187	1088				
49-355	C	36.09	455	2682				
50-355	C	36.22	560	3247				
56-288c	C	36.22	96	557	T	39.15	233	1425
57-288c	C	36.21	686	4058				
58-288c	C	36.23	287	1682				
59-288c	C	36.37	469	2729				
60-288c	C	36.22	721	4212				
56-329c	C	36.24	305	1760				
57-329c	C	36.23	602	3581				
58-329c								
59-329c	C	36.36	533	3148				
60-329c	C	36.1	704	4055				

Using conditional formatting within the Microsoft® Excel 2010 software, the negative control threshold of 2677 RFU was applied to the data and any ~36 bp C peak heights less than or equal to this threshold were labeled as red, and any ~36 bp C peak heights above this threshold were labeled as green. As the data was analyzed, it was seen that only one of the ten controls gave no data, and this sample name was changed from black to red text (control sample 58-329c from the 355 cells). For the remaining samples, it was seen that forty-nine of the samples were displaying C peaks below the negative control threshold of 2677 RFU for the 36 bp sized C peaks. For the control cells, all nine of the remaining cells were also below this negative control threshold. This left only one viable cell sample and no controls samples available for interpretation.

However, a second issue was occurring in the data as well. As it was seen in the amplification negative control, an additional T peak was occurring in thirteen random samples and one control samples. When this T peak occurred, it was approximately 39 bp in size. Therefore, for the samples affected by this T peak, additional conditional formatting was applied to the data to establish the negative control threshold of 138 RFU for the T peak seen in the amplification negative control. When this was applied to the data, only two samples (samples 24 and 32) were seen to be below this 138 RFU negative control threshold. Sample 24 also contained a below-threshold C peak, and was therefore already ineligible for analysis. Sample 32 did not display any C peak at 36 bp, so the resulting below-threshold T peak alone resulted in the sample being invalid. The twelve samples (samples 9, 10, 11, 12, 13, 23, 25, 27, 30, 38, 41, and 43) and one control (56-288c) that contained any above-threshold peaks were deemed valid for interpretation. The remainder of the samples will be addressed in the discussion.

SNaPshot™ Results of Sample 355 (16519) Cell Analysis

Upon analysis, the SNaPshot™ internal positive and negative controls were seen to pass. The amplification positive and negative controls from the 355 cell samples were first analyzed. For reference, the HL60 DNA type at 16519 should have been an A base around 43 bp. The results are presented in Table 38:

Table 38- Control Sample Results for 355 (16519) Cells in SNaPshot™

Control Sample	Primer Set	SNP Base Detected	Base-pair size	Peak Height (in RFU)
PE	16519	A	44.02	6082
AN	16519	G	43.36	70
CAN	16519	N/A	N/A	N/A

The positive control passed, and the negative control threshold was established as 70 RFU for ~43 bp sized peaks based on the peak seen in the amplification negative control. The cycling amplification negative control (CAN) for these samples did not show any peak and was also negative, but since none of the samples were amplified with additional cycles, this control would not be analyzed any further.

The fifty single cell samples for 355 as well as the ten control cell samples were then analyzed. Based on the gross tissue analysis, the 16519 extension primer showed a G/A heteroplasmy. The 355 cells, however, only displayed homoplastic G peaks (see the discussion for analysis of this issue). The results of the SNaPshot™ SNP typing are displayed in Table 39, below.

Table 39- SNaPshot™ Results for G/A heteroplasmic 355 Cells and Controls, 16519 Primers

Sample Name	Allele 1	Size 1	Height 1	Peak Area 1	Allele 2	Size 2	Height 2	Peak Area 2
PE-355					A	44.02	6082	38961
AN-355	G	43.36	70	413				
CAN-355								
1-355	G	43.16	135	825				
2-355	G	43.24	72	436				
3-355	G	43.3	371	2257				
4-355	G	43.23	173	1075				
5-355	G	43.33	75	447				
6-355	G	43.31	97	603				
7-355	G	43.24	115	701				
8-355	G	43.23	157	981				
9-355	G	43.24	149	881				
10-355	G	43.3	103	647				
11-355	G	43.38	95	584				
12-355	G	43.24	163	1009				
13-355	G	43.3	79	482				
14-355	G	43.07	133	834				
15-355	G	43.23	128	799				
16-355	G	43.13	188	1160				
17-355	G	43.18	152	959				
18-355	G	43.16	104	657				
19-355	G	43.35	191	1187				
20-355	G	43.17	288	1777				
21-355	G	43.24	185	1171				
22-355	G	43.3	60	381				
23-355	G	43.23	104	628				
24-355	G	43.29	123	751				
25-355	G	43.23	232	1455				
26-355	G	43.3	159	1016				
27-355	G	43.29	172	1059				
28-355	G	43.24	153	941				
29-355								
30-355	G	43.24	104	662				

Table 39- SNaPshot™ Results for G/A heteroplasmic 355 Cells and Controls, 16519 Primers (continued)

Sample Name	Allele 1	Size 1	Height 1	Peak Area 1	Allele 2	Size 2	Height 2	Peak Area 2
31-355	G	43.28	112	709				
32-355	G	43.31	224	1377				
33-355	G	43.24	91	567				
34-355	G	43.29	105	670				
35-355	G	43.29	301	1871				
36-355	G	43.24	107	689				
37-355	G	43.2	214	1350				
38-355	G	43.07	147	926				
39-355	G	43.24	158	977				
40-355	G	43.23	65	413				
41-355	G	43.24	172	1082				
42-355	G	43.24	97	611				
43-355	G	43.28	350	2207				
44-355	G	43.24	196	1252				
45-355	G	43.23	169	1060				
46-355	G	43.28	117	725				
47-355	G	43.24	92	565				
48-355	G	43.16	105	665				
49-355	G	43.31	130	820				
50-355	G	43.2	134	833				
6-288c	G	43.32	175	1083				
7-288c	G	43.29	251	1628				
8-288c	G	43.24	183	1101				
9-288c	G	43.15	173	1075				
10-288c	G	43.15	99	595				
6-329c	G	43.24	91	544				
7-329c	G	43.23	248	1514				
8-329c	G	43.24	158	951				
9-329c	G	43.24	90	564				
10-329c	G	43.3	218	1330				

Using conditional formatting within the Microsoft® Excel 2010 software, the negative control threshold of 70 RFU was applied to the data and any ~43 bp peak heights less than or equal to this threshold were labeled as red, and any ~43 bp peak heights above this threshold were labeled as green. As the data was analyzed, it was seen that only one of the fifty 355 cells samples gave no data, and this sample name was changed from black to red text (sample 29 from the 355 cells). For the remaining samples, it was seen that two of the samples were displaying peaks below the negative control threshold of 70 RFU for the 43 bp sized peaks. For the control cells, none of the cells were below this negative control threshold. Samples 22, 29, and 40 were therefore deemed fully invalid, and the remaining forty-seven viable cell samples and ten controls samples were deemed valid for interpretation.

SNaPshot™ Results of Sample 369 Cell Analysis

Upon analysis, the SNaPshot™ internal positive and negative controls were seen to pass. The amplification positive and negative controls from the 369 cell samples were first analyzed. For reference, the HL60 DNA type at 16519 should be an A base around 43 bp. The results are presented in Table 40:

Table 40- Control Sample Results for 369 Cells in SNaPshot™

Control Sample	Primer Set	SNP Base Detected	Base-pair size	Peak Height (in RFU)
PE	16519	A	43.35	8223
AN	16519	G	42.65	197
CAN	16519	G	42.79	669

The positive control passed, and the negative control threshold was established as 197 RFU for ~42 bp sized G peaks based on the G peak seen in the amplification negative control. The cycling amplification negative control (CAN) for these samples did show peaks as well, but

since none of the samples were amplified with additional cycles, this control would not be analyzed any further.

The fifty single cell samples for 369 as well as the ten control cell samples were then analyzed. Based on the gross tissue analysis, the 16519 extension primer showed a G/A heteroplasmy. While this G/A heteroplasmy was expected based on the gross tissue analysis, G/A/C peaks were found instead (see the discussion for analysis of this issue). The results of the SNaPshot™ SNP typing are displayed in Table 41, below.

Table 41- SNaPshot™ Results for G/A heteroplasmic 369 Cells and Controls, 16519 Primers

Sample Name	Allele 1	Size 1	Height 1	Peak Area 1	Allele 2	Size 2	Height 2	Peak Area 2	Allele 3	Size 3	Height 3	Peak Area 3
PE-369					A	43.35	8223	56053				
AN-369	G	42.65	197	1310								
CAN-369	G	42.79	669	4390								
1-369	G	42.79	691	4524					C	43.91	48	325
2-369	G	42.59	330	2178								
3-369	G	42.64	719	4668					C	43.76	68	449
4-369	G	42.72	624	3999					C	43.85	57	383
5-369	G	42.72	169	1082					C	43.85	33	218
6-369	G	42.72	415	2702					C	43.85	88	616
7-369	G	42.72	653	4179					C	43.85	87	572
8-369	G	42.69	272	1818					C	43.76	45	328
9-369	G	42.79	673	4353					C	43.91	67	443
10-369	G	42.58	601	3967					C	43.69	71	509
11-369	G	42.59	473	3225					C	43.69	46	305
12-369	G	42.66	591	3768					C	43.78	63	458
13-369	G	42.74	688	4528					C	43.99	35	258
14-369	G	42.58	464	3029					C	43.83	48	344
15-369	G	42.65	496	3188					C	43.76	38	272
16-369	G	42.79	460	3031					C	43.91	78	519
17-369	G	42.66	148	982					C	43.9	31	222
18-369	G	42.72	184	1193					C	43.83	43	311
19-369	G	42.65	537	3504					C	43.77	70	467
20-369	G	42.79	471	3109					C	43.9	69	503
21-369	G	42.73	941	6063					C	43.86	83	582
22-369	G	42.74	722	4697					C	43.86	52	373
23-369	G	42.65	768	4954					C	43.89	81	541
24-369	G	42.71	368	2358					C	43.84	37	250
25-369	G	42.79	817	5328					C	43.78	53	383
26-369	G	42.64	544	3596					C	43.88	42	287
27-369	G	42.65	664	4426					C	43.75	65	449
28-369	G	42.72	397	2600					C	43.85	71	499
29-369	G	42.72	420	2704					C	43.97	51	330
30-369	G	42.72	514	3332					C	43.85	48	313

Table 41- SNaPshot™ Results for G/A heteroplasmic 369 Cells and Controls, 16519 Primers (continued)

Sample Name	Allele 1	Size 1	Height 1	Peak Area 1	Allele 2	Size 2	Height 2	Peak Area 2	Allele 3	Size 3	Height 3	Peak Area 3
31-369	G	42.65	512	3303								
32-369	G	42.73	848	5566					C	43.84	62	418
33-369	G	42.66	350	2264								
34-369	G	42.59	559	3653								
35-369	G	42.72	844	5432								
36-369	G	42.65	838	5465								
37-369	G	42.74	1819	11611					C	43.86	64	460
38-369	G	42.59	1671	10808					C	43.84	108	737
39-369	G	42.66	535	3480								
40-369	G	42.66	732	4854								
41-369	G	42.65	1233	7977					C	43.89	44	288
42-369	G	42.59	548	3606								
43-369	G	42.63	1413	9548								
44-369	G	42.65	688	4372								
45-369	G	42.64	1163	7503								
46-369	G	42.58	1112	7351								
47-369	G	42.66	598	3872								
48-369	G	42.65	462	2958								
49-369	G	42.7	926	6044								
50-369	G	42.6	525	3429								
11-288c	G	42.73	634	4086								
12-288c	G	42.51	1251	8154								
13-288c	G	42.58	673	4374								
14-288c	G	42.8	839	5444								
15-288c	G	42.64	753	4759								
11-329c	G	42.6	794	5069								
12-329c	G	42.6	1219	7827								
13-329c	G	42.67	802	5171								
14-329c	G	42.66	1142	7313								
15-329c	G	42.6	983	6270								

Using conditional formatting within the Microsoft® Excel 2010 software, the negative control threshold of 197 RFU was applied to the data and any ~42 bp G peak heights less than or equal to this threshold were labeled as red, and any ~42 bp G peak heights above this threshold were labeled as green. As the data was analyzed, it was seen that all samples and controls displayed data. It was seen that three of the samples were displaying peaks below the negative control threshold of 197 RFU for the 42 bp sized G peaks: samples 5, 17, 18 from the 369 cells. For the rest of the samples and control cells, all of the cells were above this negative control threshold, leaving forty-seven viable cell samples and ten controls samples available for interpretation.

A second issue was noted in the data. An additional C peak was occurring in thirty-three samples. When this C peak occurred, it was seen to be around 43 bp in size. Because the amplification negative sample did not display a C peak, no negative control threshold for the C peak could be set, and therefore any visible C peak around 43 bp would be deemed a valid peak. When this was applied to the data, the three samples that were seen to contain a below-threshold G peak around 42 bp also contained a valid C peak around 43 bp. The analysis parameters for these three samples will be addressed in the discussion. None of the control samples displayed this additional C peak around 43 bp. Therefore, no samples were eliminated from the sample set and all fifty samples and the ten controls were deemed viable and valid for interpretation.

Additional testing of non-heteroplasmic tissues- 261 and 342

Prior to the length study, the tissues 261 and 342 were initially analyzed and mistakenly interpreted as containing heteroplasmy at 16390. This heteroplasmy was later determined to be a combination of the signals from the 16069, 16390, and 16324 extension primers within the original 16390 bin, due to the shift in the signals of each extension primer caused by the fluorescently tagged ddNTP base. Prior to the completion and analysis of the length study, fifty cells were isolated from each of these tissues and tested for what was mistakenly believed to be 16390 heteroplasmy, in an attempt to gather some preliminary data for the study. The initial analysis of these cells showed heteroplasmy in the individual cells. However, upon the gross tissue reanalysis following the completion of length study, neither the 261 nor 342 gross tissues displayed a heteroplasmic peak for the 16390 position. The results of the SNaPshot™ data from these cells are presented below.

Typing results of non-heteroplasmic tissue showing possible heteroplasmic cells (sample 261, with the 16390 primers)

For the 261 tissue samples, 50 cells were isolated for analysis for the representative 16390 SNP position. The cells that were utilized were from the previously prepared microscope slides. The cells were identified, isolated, and collected on the PALM laser-dissection microscope utilizing the same methodology as previously described.

The cells from the 261 sample and controls were one-step amplified using the 16390 SNP primers based on the amplification settings that were previously validated. The cells from the 261 sample were run using 34 cycles of amplification during the one-step amplification process. To each set of fifty sample cells, five control cells were added to the amplification set from sample 288, and five control cells were also added from sample 329, giving a total of sixty cells

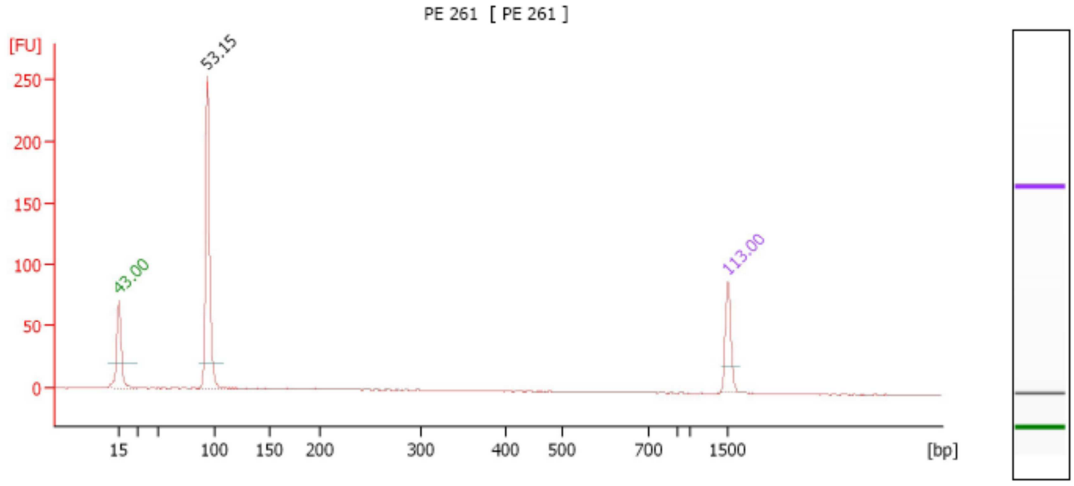
that were amplified within each test set. To control for the amplification, a positive control (PE), amplification negative control (AN), and cycling amplification negative control (CAN) were established for each set of 60 cells. The positive control and the amplification negative control were run at the standard number of amplification cycles specific for each primer set, and were used to establish the negative control threshold after SNaPshot™ analysis was complete. The cycling amplification negative control (CAN) was only to be used if it was determined that the samples required further rounds of primary amplification following Agilent 2100 Bioanalyzer sizing analysis, before SNaPshot™ amplification occurred.

All of the controls and samples for each set of sixty cells were run on the Agilent 2100 Bioanalyzer. Based on the empirical data from the validations, the one-step amplified single cells would not be expected to display a visible peak in the Agilent 2100 Bioanalyzer data. As noted in the cycling number validation study using the test cells, negative results on the Agilent 2100 Bioanalyzer data did not indicate that the single-cell samples could not transition to SNaPshot™ analysis, based on the low-end level of the detection of the Agilent 2100 Bioanalyzer being higher than the low-end input concentration for the SNaPshot™ assay. In order for the samples to effectively pass Agilent 2100 Bioanalyzer sizing analysis, however, the positive control must have displayed the appropriate peak and the negative controls must have been negative.

Following the one-step amplification, all of the one-step amplified samples were treated with ExoSAP-IT® using the Vallone protocol (Vallone et al., 2004). The representative amplified single 261 cells and controls were then run on the Agilent 2100 Bioanalyzer. The positive control for the 16390 primer was seen at the appropriate bp size (see Figure 92, below),

and the negative controls for the sample set showed no bands (see Figure 93, and Figure 94 below).

Figure 92- Positive Control for 16390 primers, Sample 261 Cells



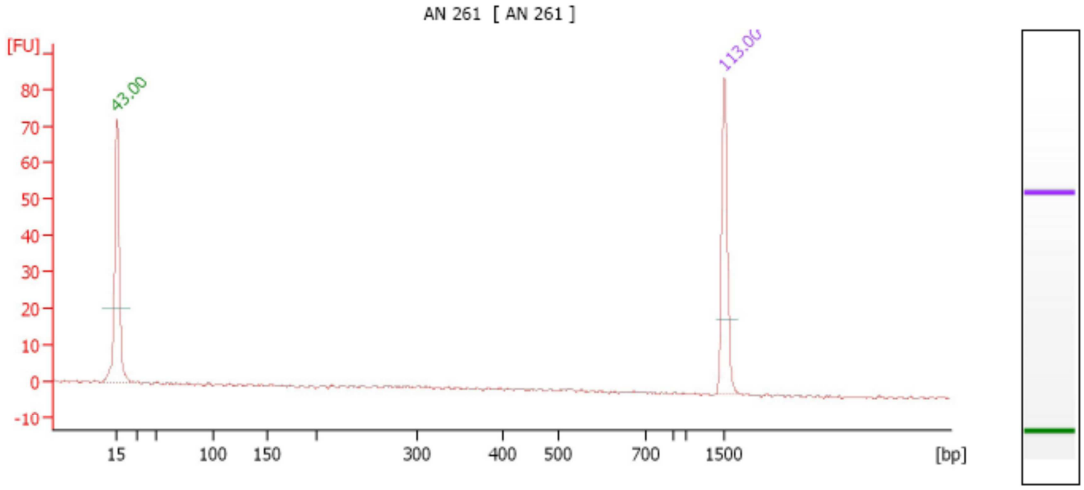
Overall Results for sample 1 : PE 261

Number of peaks found: 1

Peak table for sample 1 : PE 261

Peak	Size [bp]	Conc. [ng/µl]	Molarity [nmol/l]	Observations
1	15	4.20	424.2	Lower Marker
2	93	10.51	171.6	
3	1,500	2.10	2.1	Upper Marker

Figure 93- Amplification Negative Control for 16390 primers, Sample 261 Cells



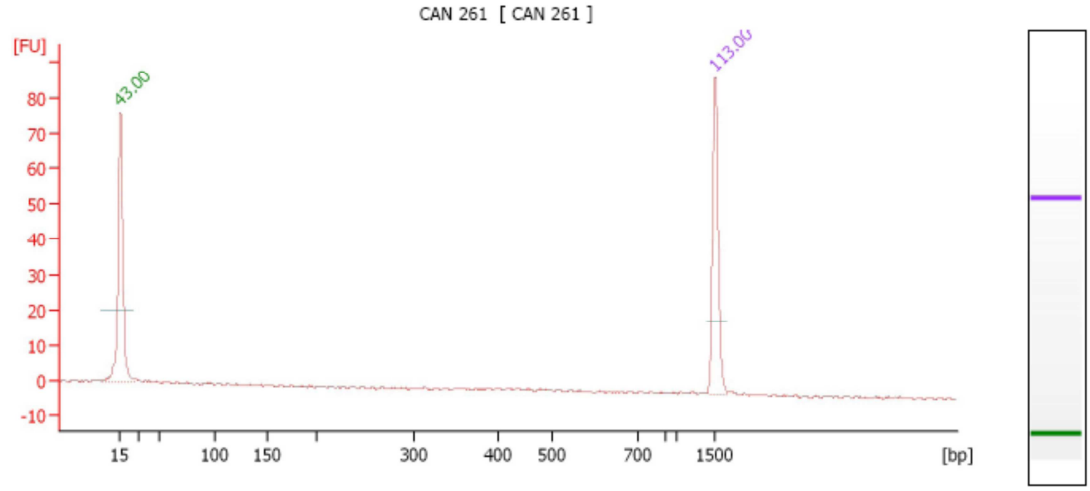
Overall Results for sample 2 : AN 261

Number of peaks found: 0

Peak table for sample 2 : AN 261

Peak	Size [bp]	Conc. [ng/µl]	Molarity [nmol/l]	Observations
1	15	4.20	424.2	Lower Marker
2	1,500	2.10	2.1	Upper Marker

Figure 94- Cycling Amplification Negative Control, 16390 primers, Sample 261 Cells



Overall Results for sample 3 : CAN 261

Number of peaks found: 0

Peak table for sample 3 : CAN 261

Peak	Size [bp]	Conc. [ng/µl]	Molarity [nmol/l]	Observations
1	15	4.20	424.2	Lower Marker
2	1,500	2.10	2.1	Upper Marker

In Figure 95 below, the first 48 sample electropherograms for the 261 cells are overlaid, and it is clear that no peaks are present in the 93 bp region. The remaining data for the last 2 samples of the 261 cells, as well as seven of the ten control samples for the 16390 primers can be seen in Figure 96. The last three control samples, from 329 (329-18, 329-19, and 329-20) showed a peak in the 93 bp region (Figure 97), which was consistent with the 16390 primers, and was determined to be valid amplification of the single cells giving a detectable signal on the Agilent 2100 Bioanalyzer.

Figure 95- Overlay of Agilent 2100 Bioanalyzer Data, 16390 primers, 261 cells

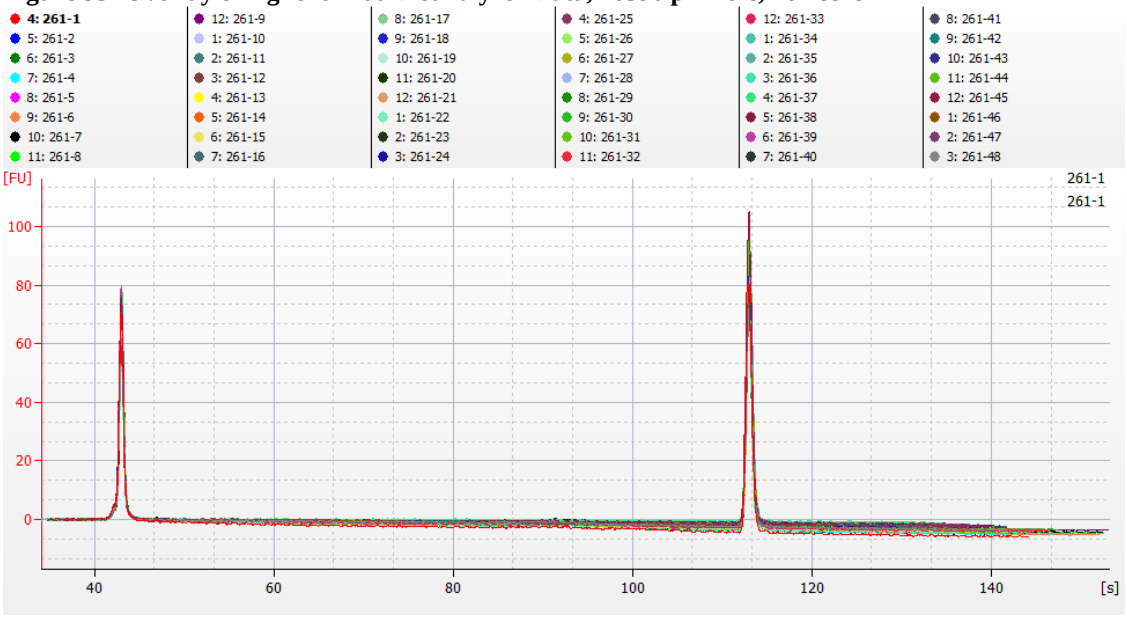


Figure 96- Overlay of Agilent 2100 Bioanalyzer Data, 16390 primers, 261 cells and controls

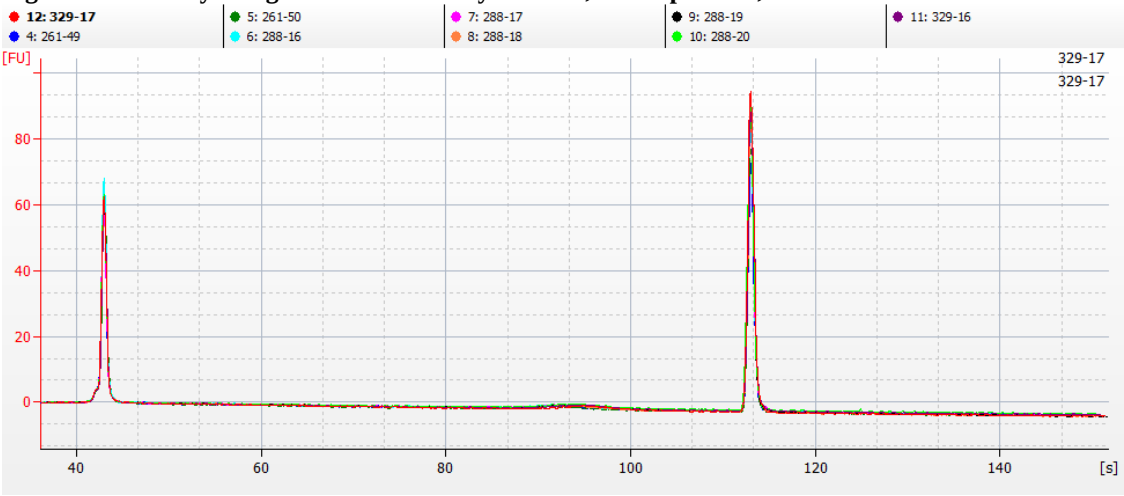
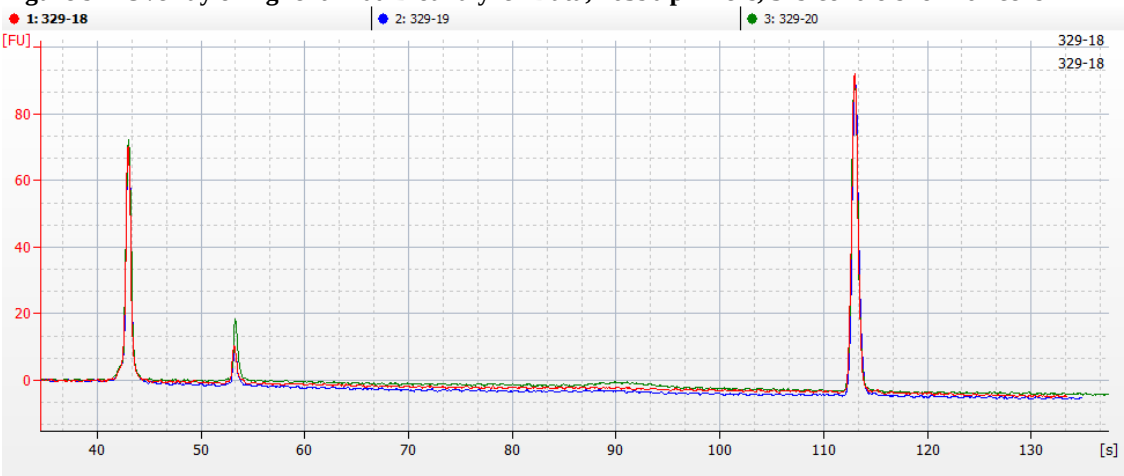


Figure 97- Overlay of Agilent 2100 Bioanalyzer Data, 16390 primers, 329 controls for 261 cells



Upon analysis, the SNaPshot™ internal positive and negative controls passed. The amplification positive and negative controls from the 261 cell samples were first analyzed. For reference, the HL60 DNA type at 16390 should be a C base around 36 bp. The results were as presented in Table 42:

Table 42- Control Sample Results for 261 Cells in SNaPshot™

Control Sample	Primer Set	SNP Base Detected	Base-pair size	Peak Height (in RFU)
PE	16390	C	36.09	5524
AN	16390	C	36.08	53
CAN	16390	C	36.21	431

The positive control passed, and the negative control threshold was established as 53 RFU for ~36 bp sized peaks based on the peak seen in the amplification negative control. The cycling amplification negative control (CAN) for these samples did show peaks as well, but since none of the samples were amplified with additional cycles, this control would not be analyzed any further.

The fifty single cell samples for 261 as well as the ten control cell samples were then analyzed. As seen below in Table 43, the results of the SNaPshot™ SNP typing is shown.

Table 43- SNaPshot™ Results for 261 Cells and Controls, 16390 Primers

Sample Name	Allele 1	Size 1	Height 1	Peak Area 1	Allele 2	Size 2	Height 2	Peak Area 2	Allele 3	Size 3	Height 3	Peak Area 3
PE-261	C	36.09	5524	33588								
AN-261	C	36.08	53	425								
CAN-261	C	36.21	431	3204								
1-261	C	36.34	92	742								
2-261	C	36.34	69	584								
3-261												
4-261	C	36.34	86	679								
5-261	C	36.48	61	442								
6-261	C	36.48	178	1367								
7-261	C	36.34	217	1621	C	38.44	156	959	T	39.22	42	457
8-261	C	36.2	68	550								
9-261	C	36.34	225	1725								
10-261	C	36.19	123	928	C	38.41	178	1083	T	39.18	50	392
11-261	C	36.19	260	2032								
12-261	C	36.35	101	716	C	38.48	249	1500	T	39.27	64	471
13-261	C	36.36	193	1423								
14-261	C	36.35	69	504								
15-261	C	38.59	93	576								
16-261	C	36.34	539	3989	C	38.45	189	1212				
17-261	C	36.22	114	888								
18-261	C	36.09	121	949								
19-261	C	36.21	206	1547	C	38.47	166	1019	T	39.12	48	461
20-261	C	36.2	87	648								
21-261	C	36.36	108	763	C	38.49	129	779	T	39.27	36	294
22-261	C	36.49	135	997								
23-261												
24-261	C	36.34	45	341								
25-261	C	36.34	211	1569								
26-261	C	36.33	151	1110	C	38.42	132	838				
27-261	C	36.06	176	1359								
28-261	C	36.48	159	1178	C	38.46	219	1380	T	39.24	62	487
29-261	C	36.36	109	779	C	38.49	386	2338	T	39.27	109	770
30-261	C	36.47	314	2330								

Table 43- SNaPshot™ Results for 261 Cells and Controls, 16390 Primers (continued)

Sample Name	Allele 1	Size 1	Height 1	Peak Area 1	Allele 2	Size 2	Height 2	Peak Area 2	Allele 3	Size 3	Height 3	Peak Area 3
31-261	C	36.34	107	774								
32-261					C	38.47	205	1194	T	39.26	56	359
33-261	C	36.48	273	2046								
34-261	C	36.33	54	434								
35-261	C	36.21	138	1043								
36-261	C	36.48	116	831	C	38.46	75	475				
37-261	C	36.48	351	2583								
38-261					C	38.49	374	2236	T	39.27	99	655
39-261	C	36.21	212	1565								
40-261	C	36.33	90	685								
41-261	C	36.34	202	1483	C	38.45	393	2370	T	39.23	108	817
42-261	C	36.21	163	1254								
43-261	C	36.06	104	821	C	38.39	140	907	T	39.03	41	333
44-261	C	36.48	177	1312								
45-261	C	36.34	220	1650	C	38.31	46	302				
46-261	C	36.48	589	4273	C	38.6	145	933	T	39.25	41	264
47-261												
48-261	C	36.34	113	837	C	38.45	85	520				
49-261	C	36.33	273	2016								
50-261	C	36.36	45	351								
16-288c	C	36.34	255	1833								
17-288c	C	36.18	310	2188								
18-288c	C	36.35	111	797	C	38.48	139	851	T	39.13	41	311
19-288c	C	36.35	133	926	C	38.48	194	1175	T	39.14	60	449
20-288c	C	36.21	259	1624								
16-329c	C	36.23	289	1875	C	38.5	194	1189	T	39.3	104	669
17-329c	C	36.36	43	270	C	38.49	166	999	T	39.27	71	466
18-329c	C	36.35	2702	18128	C	38.48	161	1219				
19-329c	C	36.33	2982	18734								
20-329c	C	36.35	4157	26230	C	38.46	121	1062	T	39.24	78	587

Using conditional formatting within the Microsoft® Excel 2010 software, the negative control threshold of 53 RFU was applied to the data and any ~36 bp peak heights less than or equal to this threshold were labeled as red, and any ~36 bp peak heights above this threshold were labeled as green. As the data was analyzed, it was seen that only three of the fifty samples gave no data, and these sample names were changed from black to red text (samples 3-261, 23-261, and 47-261). For the remaining samples, it was seen that two of the samples were displaying peaks below the negative control threshold of 53 RFU for the ~36 bp sized peaks. For the control cells, one of the cells contained a ~36 bp peak below this negative control threshold, but contained other valid peak data for that control sample. This left forty-five viable cell samples and four control samples available for interpretation. However, secondary and tertiary issues were occurring in the data as well. An additional C peak at ~38 bp was occurring in seventeen random samples and six control samples. An additional T peak at ~39 bp was occurring in twelve samples and five control samples. Neither of these additional peaks had a corresponding peak in the amplification negative control, so all peaks are valid. Out of the seventeen instances of the C peak at ~38 bp, only five were only this additional C peak, acting as a second peak to the original C peak at ~36 bp. For every instance of the T peak at ~39 bp, it was associated with a C peak at ~38 bp. However, not every instance of the C-T peaks at ~38 and ~39 bp was associated with a C peak at ~36 bp. Two samples, 32-261 and 38-261 contained only these C and T peaks at ~38 and ~39 bp, and did not display the C peak at ~36 bp. For the 261 cells, there was never an instance of the C peak at ~38 bp or the T peak at ~39 bp displayed on its own. The T peaks were always associated with a C peak at ~38 bp, and the C peak at ~38 bp was either associated with a T peak at ~39 bp, the C peak at ~36 bp, or both of these peaks.

Typing results of non-heteroplasmic tissue showing possible heteroplasmic cells (sample 342, with the 16390 primers)

For the 342 tissue sample, 50 cells were isolated for analysis for the 16390 SNP position. The cells that were utilized were from the previously prepared microscope slides. The cells were identified, isolated, and collected on the PALM laser-dissection microscope utilizing the same methodology as previously described.

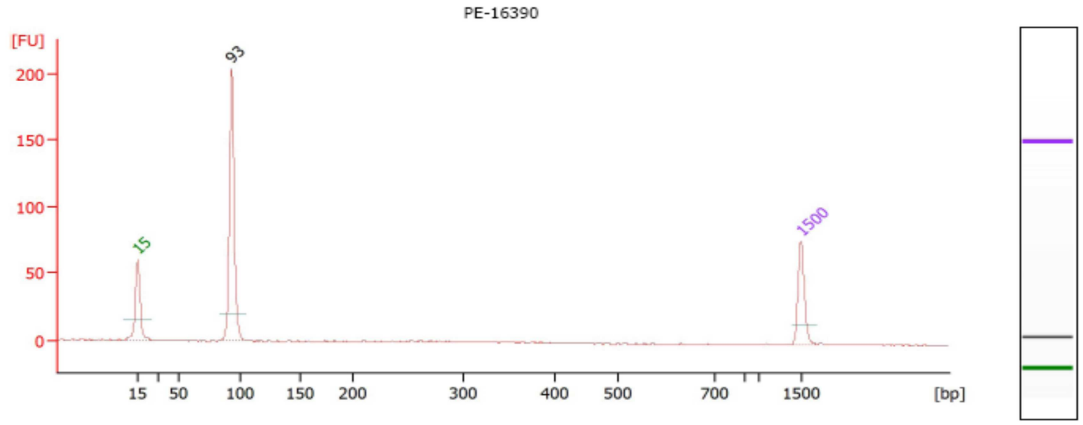
The cells from the 342 samples and controls were one-step amplified using the 16390 SNP primers per sample based on the amplification settings that were previously validated. The cells from the 342 samples were run using 34 cycles of amplification during the one-step amplification process. To the set of fifty 342 sample cells, five control cells were added to the amplification set from sample 288, and five control cells were also added from sample 329, giving a total of sixty cells that were amplified within the test set. To control for the amplification, a positive control (PE), amplification negative control (AN), and cycling amplification negative control (CAN) were established for the set of sixty cells. The positive control and the amplification negative control were run at the standard number of amplification cycles specific for the 16390 primer set, and were used to establish the negative control threshold after SNaPshot™ analysis was complete. The cycling amplification negative control (CAN) was only to be used if it was determined that the samples required further rounds of primary amplification following Agilent 2100 Bioanalyzer sizing analysis, before SNaPshot™ amplification occurred.

Following the one-step amplification, all of the one-step amplified samples were treated with ExoSAP-IT® using the Vallone protocol (Vallone et al., 2004). The representative amplified single 342 cells and controls were then run on the Agilent 2100 Bioanalyzer. Based on

the empirical data from the validations, the one-step amplified single cells would not be expected to display a visible peak in the Agilent 2100 Bioanalyzer data. As noted in the cycling number validation study using the test cells, negative results on the Agilent 2100 Bioanalyzer data did not indicate that the single-cell samples could not transition to SNaPshot™ analysis, based on the low-end level of the detection of the Agilent 2100 Bioanalyzer being higher than the low-end input concentration for the SNaPshot™ assay. In order for the samples to effectively pass Agilent 2100 Bioanalyzer sizing analysis, however, the positive control must have displayed the appropriate band and the negative controls must have been negative.

The positive control for the 16390 primer was seen at the appropriate bp size (see Figure 98 below), and the negative controls for the sample sets showed no bands (see Figure 99, and Figure 100 below).

Figure 98- Positive Control for 16390 primers, Sample 342 Cells



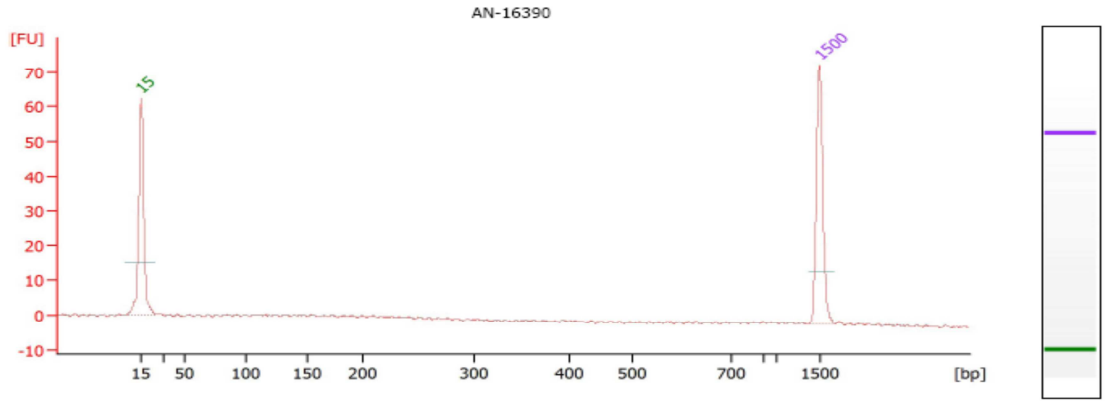
Overall Results for sample 1 : PE-16390

Number of peaks found: 1

Peak table for sample 1 : PE-16390

Peak	Size [bp]	Conc. [ng/µl]	Molarity [nmol/l]	Observations	Aligned Migration Time [s]	Area	Time corrected area	Peak Height	Peak Width	% of Total
1	15	4.20	424.2	Lower Marker	43.00	38.3	98.2	60.5	3.0	0.0
2	93	9.95	162.0		52.94	124.6	252.0	204.5	2.7	100.0
3	1,500	2.10	2.1	Upper Marker	113.00	64.7	57.4	77.7	2.8	0.0

Figure 99- Amplification Negative Control for 16390 primers, Sample 342 Cells



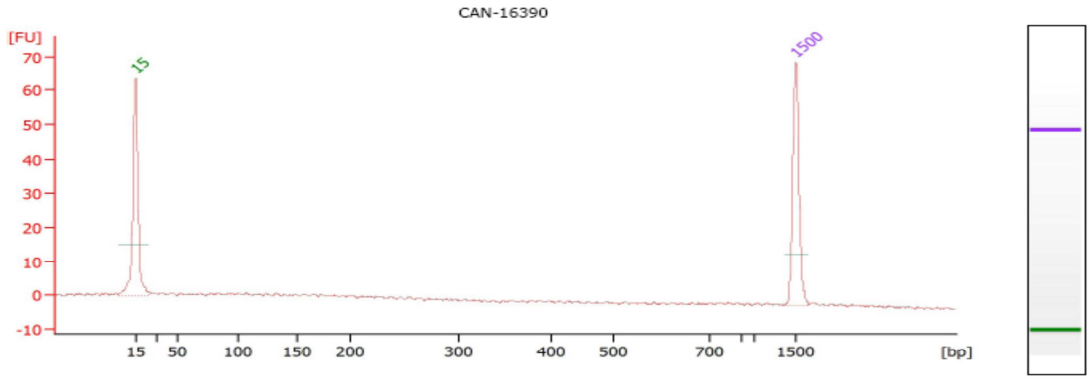
Overall Results for sample 2 : AN-16390

Number of peaks found: 0

Peak table for sample 2 : AN-16390

Peak	Size [bp]	Conc. [ng/µl]	Molarity [nmol/l]	Observations	Aligned Migration Time [s]	Area	Time corrected area	Peak Height	Peak Width	% of Total
1	15	4.20	424.2	Lower Marker	43.00	40.4	103.1	62.4	3.2	0.0
2	1,500	2.10	2.1	Upper Marker	113.00	61.2	54.2	74.8	2.7	0.0

Figure 100- Cycling Amplification Negative Control, 16390 primers, Sample 342 Cells



Overall Results for sample 3 : CAN-16390

Number of peaks found: 0

Peak table for sample 3 : CAN-16390

Peak	Size [bp]	Conc. [ng/µl]	Molarity [nmol/l]	Observations	Aligned Migration Time [s]	Area	Time corrected area	Peak Height	Peak Width	% of Total
1	15	4.20	424.2	Lower Marker	43.00	43.0	110.4	63.9	3.2	0.0
2	1,500	2.10	2.1	Upper Marker	113.00	58.3	52.3	71.6	2.6	0.0

In Figure 101 below, the first 48 sample electropherograms for the 342 cells are overlaid, and it is clear that no peaks are present in the 93 bp region. The remaining data for the last 2

samples of the 342 cells, as well as the ten control samples for the 16390 primers can be seen in Figure 102.

Figure 101- Overlay of Agilent 2100 Bioanalyzer Data, 16390 primers, 342 cells

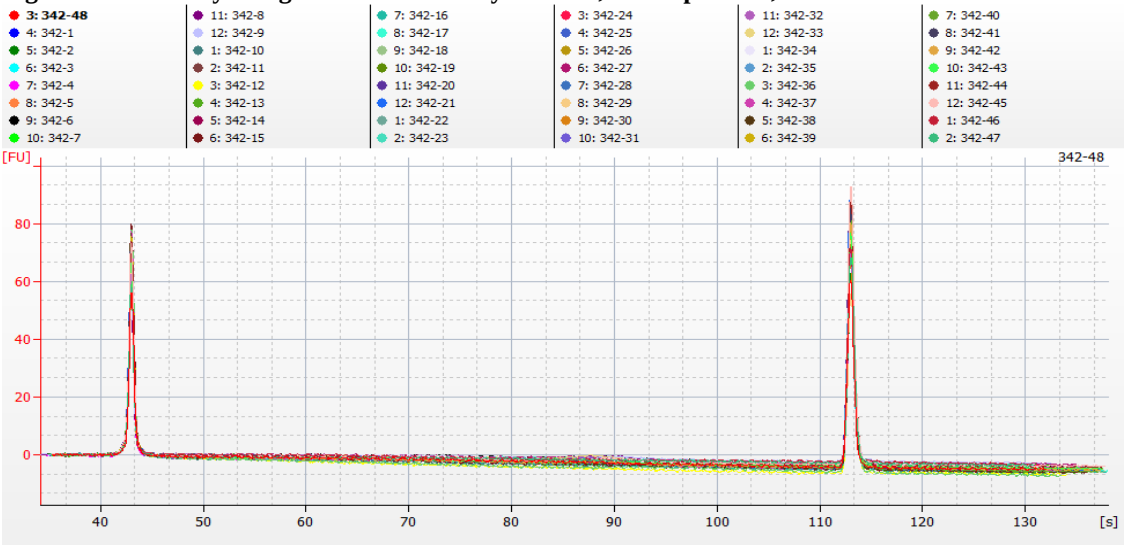
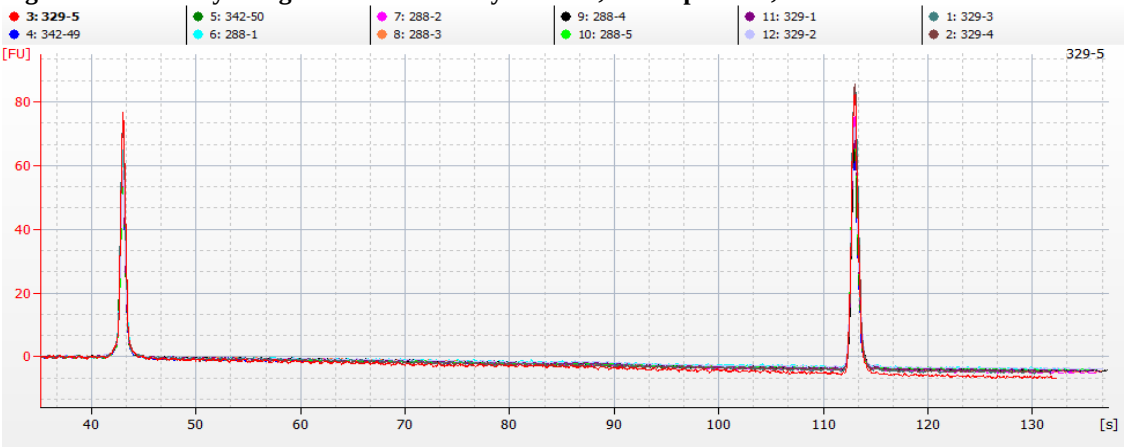


Figure 102- Overlay of Agilent 2100 Bioanalyzer Data, 16390 primers, 342 cells and control cells



The samples were then amplified in the SNaPshot™ assay, purified with SAP, and run on the 3130*xl* capillary electrophoresis instrument. Upon analysis, the SNaPshot™ internal positive and negative controls passed. The amplification positive and negative controls from the 342 cell samples were first analyzed. For reference, the HL60 DNA type at 16390 should be a C base around 37 bp. The results are presented in Table 44:

Table 44- Control Sample Results for 342 Cells in SNaPshot™

Control Sample	Primer Set	SNP Base Detected	Base-pair size	Peak Height (in RFU)
PE	16390	C	37.12	4551
AN	16390	C	37.06	236
		G	38.93	71
		C	39.36	141
		T	40.21	98
CAN	16390	C	37.10	369
		G	39.0	97
		C	39.43	180
		T	40.29	164

The positive control passed, and the negative control threshold was established as 236 RFU for ~37 bp sized peaks based on the C peak seen in the amplification negative control. A second peak was also seen around 39 bp, and this G peak had a peak height of 71 RFU. A third peak, another C, was also seen around 39.5 bp with a peak height of 141 RFU. A fourth peak, a T, was seen around 40 bp with a peak height of 98 RFU. Since these additional peaks were unexpected, the size and peak heights of the second, third, and fourth peaks were noted. The cycling amplification negative control (CAN) for these samples did show peaks as well, but since none of the samples were amplified with additional cycles, this control would not be analyzed any further.

The fifty single cell samples for 342 as well as the ten control cell samples were then analyzed. Based on the gross tissue analysis of sample 342, the 16390 extension primer displayed a homoplastic C base. As seen below in Table 45, the results of the SNaPshot™ SNP typing are shown.

Table 45- SNaPshot™ Results for 342 Cells and Controls, 16390 Primers

Sample Name	Allele 1	Size 1	Height 1	Peak Area 1	Allele 2	Size 2	Height 2	Peak Area 2	Allele 3	Size 3	Height 3	Peak Area 3	Allele 4	Size 4	Height 4	Peak Area 4
PE-342	C	37.12	4551	26709												
AN-342	C	37.06	236	1405	G	38.93	71	487	C	39.36	141	806	T	40.21	98	664
CAN-342	C	37.1	369	2137	G	39	97	618	C	39.43	180	1032	T	40.29	164	1119
1-342	C	37.22	30	187					C	39.38	86	469	T	40.23	161	991
2-342	C	37.18	206	1296					C	39.29	69	429	T	40.13	32	265
3-342	C	36.92	875	5152					C	39.36	378	2249	T	40.07	486	3172
4-342																
5-342	C	37.07	459	2767					C	39.23	421	2465	T	40.07	522	3249
6-342	C	37.2	1169	6883												
7-342	C	37.07	224	1360												
8-342	C	37.07	358	2178												
9-342																
10-342	C	37.36	1053	6343	G	39.08	54	352								
11-342	C	37.26	666	4041					C	39.37	290	1758	T	40.21	176	1256
12-342	C	37.22	777	4611					C	39.38	208	1223	T	40.23	144	1018
13-342																
14-342	C	37.08	193	1191												
15-342	C	37.06	91	561					C	39.22	226	1291	T	40.07	297	1858
16-342	C	37.22	332	2005					C	39.38	188	1199	T	40.23	143	974
17-342	C	37.23	125	762												
18-342	C	37.36	310	1980					C	39.36	109	639	T	40.2	43	415
19-342									C	39.33	303	1713	T	40.17	659	4019
20-342	C	37.23	182	1093					C	39.4	122	715	T	40.25	156	1028
21-342	C	37.23	207	1192												
22-342	C	37.22	192	1130					C	39.24	56	314	T	40.09	132	801
23-342									C	39.43	113	637	T	39.43	113	1826
24-342	C	37.07	217	1296					C	39.38	55	321	T	40.22	241	1441
25-342					G	39.1	68	461	C	39.38	171	952	T	40.23	196	1215
26-342	C	37.19	401	2385					C	39.33	185	1077	T	40.17	221	1413
27-342	C	37.21	188	1139												
28-342	C	37.22	711	4138												
29-342	C	37.21	766	4684												
30-342	C	37.22	170	1001												

Table 45- SNaPshot™ Results for 342 Cells and Controls, 16390 Primers (continued)

Sample Name	Allele 1	Size 1	Height 1	Peak Area 1	Allele 2	Size 2	Height 2	Peak Area 2	Allele 3	Size 3	Height 3	Peak Area 3	Allele 4	Size 4	Height 4	Peak Area 4
31-342	C	37.06	151	869					C	39.36	120	721	T	40.21	148	916
32-342	C	37.07	405	2419												
33-342	C	37.23	373	2156					C	39.4	67	384	T	40.25	506	3033
34-342	C	37.2	132	800					C	39.34	242	1392	T	40.18	179	1145
35-342									C	39.33	261	1512	T	39.33	261	735
36-342	C	37.2	111	676	G	39.06	89	609								
37-342	C	37.22	221	1410	G	39.1	63	469	C	39.24	162	878	T	40.09	50	590
38-342	C	37.22	107	611	G	39.1	148	1084	C	39.24	177	928	T	40.23	56	447
39-342	C	37.22	299	1880	G	38.95	61	428								
40-342	C	37.06	268	1822	G	39.07	59	413								
41-342	C	37.06	194	1171	G	38.93	95	637								
42-342									C	39.19	245	1468	T	40.03	70	493
43-342	C	37.05	172	1143												
44-342	C	37.07	678	3957												
45-342	C	36.99	267	1789												
46-342																
47-342	C	37.07	88	527					C	39.38	119	689	T	40.23	42	314
48-342	C	37.19	292	1817	G	39.04	97	668	C	39.18	114	642	T	40.16	58	568
49-342	C	37.05	1082	6566	G	38.9	295	2028								
50-342	C	37.22	164	949					C	39.24	264	1509	T	40.09	139	927
1-288c	C	37.16	161	997	G	38.99	68	454								
2-288c	C	37.04	204	1294	G	39.02	161	1124								
3-288c	C	37.2	114	766												
4-288c	C	37.2	181	1055					C	39.34	199	1149	T	40.18	188	1197
5-288c	C	37.19	72	458					C	39.33	250	1410	T	40.17	72	522
1-329c	C	37.21	621	3877												
2-329c	C	37.36	136	910												
3-329c	C	37.22	359	2107					C	39.38	41	252	T	40.22	227	1387
4-329c																
5-329c	C	37.06	379	2297					C	39.22	236	1395	T	40.07	136	929

As the data was analyzed, it was seen that only four of the fifty samples and one of the ten controls gave no data, and these sample names were changed from black to red text (samples 4, 9, 13, 46, and control sample 4 from the 342 cells). For the remaining samples, it was seen that twenty-one of the samples were displaying peaks below the negative control threshold of 236 RFU for the ~37 bp sized C peaks, leaving twenty samples that contained valid C peaks at ~37 bp. Using conditional formatting, the negative control threshold of 236 RFU was applied to the data and any ~37 bp peak heights less than or equal to this threshold were labeled as red, and any ~37 bp C peak heights above this threshold were labeled as green. For the control cells, six of the cells were also below this negative control threshold, leaving three control cells with valid C peaks at the ~37 bp size. However, other issues were occurring in the data. An additional G peak was occurring in ten random samples and two control samples. As it was seen in the amplification negative control, when this G peak occurred, it was seen to be around ~38 bp in size. Therefore, for the samples affected by this G peak, additional conditional formatting was applied to the data to establish the negative control threshold of 71 RFU for the G peak seen in the amplification negative control. When this was applied, only five samples and one of the control samples were seen to be above this 71 RFU negative control threshold, thereby containing valid G peak data. A third issue was the presence of additional C and T peaks seen in the amplification negative control, with these same peaks appearing in a total of twenty-six out of the fifty cells, and in four of the control cells as well. When this additional C peak occurred, it was seen to be around ~39 bp in size, and the T peak was around ~40 bp in size. It was noted that in the 342 cells, for every instance of the C peak at ~39 bp, there was a corresponding T peak at ~40 bp, however, due to the amplification negative control threshold, either none, one or both of the peaks were invalid at random throughout the 342 samples. Therefore, for the samples

affected by the additional C peak at ~39 bp, conditional formatting was applied to the data to establish the negative control threshold of 141 RFU for the C peak at ~39 bp seen in the amplification negative control. For the samples affected by the additional T peak at ~40 bp, conditional formatting was applied to the table to establish the negative control threshold of 98 RFU for the T peak seen in the amplification negative control. When this was applied to the data, out of the twenty-six 342 cells that contained the additional C and T peaks, four cells contained totally invalid C and T peak data, three cells contained only valid C peak data with invalid T peak data, seven cells contained only valid T peak data with invalid C peak data, and twelve cells contained both valid C and T peak data. For the control cells, out of the four cells that contained the additional C and T peaks, one cell contained only valid C peak data with invalid T peak data, one cell contained valid T peak data with invalid C peak data, and two cells contained both valid C and T peak data.

Chapter 4 - Discussion

Project Design and Setup

In a review of the final outcomes of the project and the various processes that occurred along the way, it was determined that the original scope of the project was appropriate. By focusing on the possible heteroplasmy in HVI hotspots, the scope of the project was limited to a set of targets within a fixed range of the mitochondrial genome, which in turn limited the overall primer design necessary for the development of the SNaPshot™ assay. The testing of the HVI hotspots was also ideal due to the availability of online databases containing the previously mapped and verified SNP data. If additional hotspots had been chosen within the mitochondrial HVI or HVII regions, this would have increased the complexity of the initial amplification processes and resulted in an increased consumption of the funding and resources that were allocated for primer and amplification reagents. Establishing this narrow scope at the beginning of the study was important because as the project proceeded, problems surrounding the five HVI hotspots compounded into further issues with the validation of the multiplex amplification, and further still with the issues that were uncovered during the length study of the extension primers with in the SNaPshot™ assay.

The narrow scope was also seen to be appropriate in the decision to use only liver tissue, which allowed for a constant average level of mitochondria per cell, uniform cell sizes, and availability of the tissue for use in the project. Initially, the project focused on the use of liver samples collected from formalin stock jars housed at the histology department of the New York City Office of Chief Medical Examiner. While these tissues were readily available and easily malleable for the cell separation technique, it was seen that the long-term formalin exposure had effectively destroyed all of the nucleic acid within the cells, and these tissues were therefore not

beneficial for use in the project. The study then turned to paraffin tissue blocks, also housed at the histology department of the New York City Office of Chief Medical Examiner. While the tissues in these paraffin blocks were also formalin-fixed, the total formalin exposure was in the timeframe of 12-36 hours. The tissues were then rinsed, sectioned, and encased in paraffin to preserve the tissue and provide support for histological microscopic slide preparation. These paraffin-embedded tissues were therefore ideal for the project, the only challenge being the effective removal of the paraffin from the tissue to allow the isolation of the single cells.

Cell Separation and Gross Tissue Creation

While there were clear visual and physical differences in the cell separation slides, the level of maceration (ranging from less macerated to sparse) had no impact on the ability of the individual cells to be identified and isolated for each tissue. Due to the fact that each level of maceration produced intact, single cells that were readily identifiable and capable of being isolated using the laser dissection microscope, the classification of the "level of maceration" did not affect the collection of the single cells from the heteroplasmic tissues.

Gross Tissue Screening

The tissue samples that remained on the mesh filters following cell separation served as the source for the gross tissue that was organically extracted and tested with the SNP multiplex. The use of this tissue remnant fraction served dual purposes; it allowed for the gross tissue to represent the same cell set from which the separated cells originated, while also eliminating the need for a second cut of the paraffin tissue block, paraffin removal, and additional chemical steps to break down the intact tissue block for organic extraction of the mtDNA. While this tissue was used to generate the extracted mtDNA from which the samples were screened for possible

heteroplasmy at the five SNP hotspots, it was originally run using the SNP multiplex and analyzed *before* the length study was performed. As a result, the initial interpretation of the gross tissue SNP data was incorrect due to the electrophoretic shift caused by the dye-labeled bases as they were added to each of the extension primers.

The length study was determined to be necessary after it was noted that several peaks were occurring outside of the previously defined bins for the SNaPshot™ analysis in the proof-of-concept single cell runs, as well as in the reanalysis of the preliminary SNP data from the gross tissues. In the process of validating the original SNaPshot™ multiplex, the analysis bin sizes were defined based on the common peaks that were present in the HL60 control mtDNA SNaPshot™ profile. If the mobility of any heteroplasmic peak fell outside the defined individual bin set, the peak itself would not be labeled and identified by the GeneMapper™ software. The original idea to counter this was the establishment of a single bin set which was defined as the extended length of the entire primer migratory region. This multiplex region bin would then call all peaks, regardless of size, that occurred within the entire SNaPshot™ primer region. The use of these two different bin sets therefore required two replicate analyses of the SNP data; first with the individual bins to ensure that the HL60 positive and negative controls were in concordance and that each sample contained valid data at each of the expected primer locations, and then a second analysis using the multiplex region bin to locate any peaks that happened to fall outside of the individual bin regions that could be indicative of heteroplasmy at any of the HVI hotspots. The initial analysis of the gross tissue data displayed several populating peaks within the defined individual bins, and uncovered seven samples that were incorrectly identified as heteroplasmic. The use of the multiplex region bin did uncover peaks outside the individual bins, but these peaks appeared to occur randomly upon analysis of the individual samples, and it

was incorrectly assumed that these were electrophoretic artifacts, and not real peaks. The ABI manual for SNaPshot™ described artifact peaks in vague terms, and none of the extraneous peaks from the gross tissue were detected during the validation of the SNP multiplex, reinforcing the hypothesis that these peaks were artifact data (Applied-Biosystems, 2000). However, subsequent rounds of reanalysis of the original SNP data from the gross tissue began to demonstrate a pattern in these extraneous peaks, and it was noted that occasionally, for example, a C peak and an A peak in the same bin were separated by three or more base-pairs (bp), which was unexpected. The initial research into this issue uncovered the possibility that the extension primers themselves might be responsible for the variation in the SNP data, leading to these extraneous peaks. As it was noted in the ABI SNaPshot™ manual (Applied-Biosystems, 2000):

> "In fragment analysis, a fragment is assigned a size based on its relative mobility to size standards as it migrates through the polymer. Any change in run conditions, such as capillary array length or polymer type, affects fragment mobility. Sizing differences resulting from mobility changes between various types of polymer are more apparent for sequences < 50 base-pairs (bp). Due to the nature of the polymer, smaller fragments (< 50 bp) run on POP-7™ polymer on the 3130/*xl* Genetic Analyzers and on the 3730/*xl* DNA Analyzers may have slightly different mobilities. To avoid sizing inconsistency with smaller fragments, Applied Biosystems recommends that genotyping projects be started and completed on the same instrument, using consistent run conditions. Avoid inconsistencies by creating single-base extension primers no shorter than 25 nucleotides (nts). In addition, the spacing should be at least 6-8 nts for primers that are less than 30 nts and 4-6 nts for primers that are greater than 30 nts. Alternatively, genotypes can be standardized to one platform by running representative alleles across all platforms and creating instrument-specific bin sets with standardized bin names in GeneMapper Software v3.7 or higher."

Based on this information, it was estimated that the design of the extension primers for 16069 and 16223 might have been too short for optimal detection (see Appendix A for the extension primer modifications and overall primer lengths). Even though these primers were shown to be effective in the HL60 multiplex, and somewhat effective in the initial gross tissue analysis, the length of the primers was perceived to be one reason that the primers failed to detect

any SNP positions in the single-cell samples. Also, ABI recommended that for primers less than thirty nucleotides, the separation should be around six to eight bp, but based on the validation of the HVI multiplex, the four bp differences were effective, most likely due to the use of only a small number of SNP's in the multiplex.

Based on the above information, it was realized that the ABI recommendations for the primer length were based on the probability of the shift in the detected primer size with the addition of the fluorescently-labeled ddNTP base. The overall primer length was minor compared to the effect of the dye-labeled primers on the final detected size of the extension primer in the SNP data. This explained why the peaks in the SNP multiplex were not exactly four bp apart; the dyes either "pushed" or "dragged" the extension primer to a slightly different size than what was expected. For example, the 16324 and 16390 extension primer peaks were seen to be close in the Agilent 2100 Bioanalyzer sizing (71 and 75 bp) but in the SNaPshot™ data, the peaks were nearly overlaid. This was due to the 16324T peak being "pushed forward" in the sizing of the peak, and the 16390C peak being "dragged back" from the original expected base sizes, causing these two peaks from different extension primers to appear in a single individual bin (for 16390) and ***appear as heteroplasmy***. This data indicated that the individual SNP bins were not correctly defined to appropriately account for the probable shifts in the SNP peak data when the different dye-labeled terminator bases were added to the extension primers.

Based on the information regarding the size-shift of the dye-labeled primers, several aspects of the study became clear. There was no way that a true SNP heteroplasmic peak will ever result in a true "overlaid" peak. The detection of heteroplasmy was thus understood to be the detection of more than one peak within an established bin of a defined size. The original individual bin sets were initially adjusted to account for a +/- 4 bp window, which is how the

original (incorrect) heteroplasmic samples were discovered, but it was determined that this window needed to be expanded for each primer set to account for the shift in the extension primers with the addition of the dye-labeled ddNTP bases. What was originally believed to be amplification artifacts, then, were most likely valid peaks that were shifted upstream/downstream from the 16223 and 16324 primer sets, and appearing within the 16390 and 16069 bins.

The extension primer length study was then carried out and the new bin sizes were determined as described in the Experimental Results section. The results of this study demonstrated that the different dye-labeled terminator bases, when added to the extension primers, shifted the range of the extension primers. Establishing the range of this shift then allowed for the expansion of the bins for each extension primer, and defined the possible peaks for each extension primer as well. The result of the primer length study was a clear primer map for all five of the extension primers, with the possible bases and their respective mobilities (positions) for each primer represented throughout the electropherogram. The new bins were then applied in the GeneMapper™ analysis software, and the raw SNaPshot™ data from the valid gross tissue samples was re-analyzed. The results of applying the corrected bins and bin sizes then accounted for every peak within the gross tissue SNaPshot™ data, and resulted in the identification of sixteen samples that contained the twenty possible heteroplasmic SNP locations. (See the Experimental Results section, as well as Appendix E for the full chart data).

Age vs. Heteroplasmy of the Gross Tissue

Once the length study was complete and the heteroplasmic gross tissues were identified, it was desirable to plot the known age and sex of the tissue donors against the instances of heteroplasmy. Within the scope of the data set, it can be inferred that there is an increased correlation with heteroplasmic occurrences in HVI hotspots and increased age.

Pathology Review of Cells

Following the length study and the resulting generation of the twenty heteroplasmic samples, it was determined that the laser-dissection microscopy slides needed to be examined by a qualified pathologist to determine if any of the samples contained either cancerous or otherwise abnormal cells. This was important for the study because cancerous hepatic cells would require an overabundance of mitochondrial activity to support the rapid cell growth. The highly productive mitochondrial genomes would be under further oxidative stress or facing increased mutational events leading to an increase in heteroplasmic events (He et al., 2010). Additionally, liver-specific diseases such as cirrhosis or hepatitis would be detrimental to the study as well, due to the fibrous scar tissue that occurs in the hepatic tissue as a result of cell death from exposure to toxins, alcohol, viruses, and fat. This scar tissue would cause the tissue to shrink and die, which would cut off the flow of nutrients to the surrounding cells. This in turn would lead to further cell death, with degradation of the nuclear and mitochondrial DNA occurring simultaneously. Therefore, the identification and elimination of any tissues possessing cancerous or diseased cells would remove any tissues that would be demonstrating increased or degraded mitochondrial DNA activity, which could risk being incorrectly interpreted as heteroplasmy within the mtDNA SNP profile.

Based on the final pathological findings, two samples that contained heteroplasmy (based on the length study data) were removed from the study. One of these samples, 345, had shown three separate SNP positions containing heteroplasmy: 16069 A/C, 16223 C/T, and 16519 A/G. These SNaPshot™ results, combined with the results of pathological analysis of the gross tissue justified the removal of this sample from the project, but even without the pathological findings, the results were suspect. The instance of more than one heteroplasmic point in the genome, even

at hotspots such as the ones targeted by the SNP analysis, would commonly be seen as questionable, raising concerns ranging from possible contamination to extreme mutational events in the cells. Typically in forensic mtDNA sequencing, the finding of two heteroplasmic bases in a sequence of over 300 nucleotides would be cause for concern, and three or more such heteroplasmic positions would signal a mixture of two or more separate mtDNA profiles, either from inadvertent or direct contamination of an additional mtDNA source (Budowle, 1999; Budowle et al., 2010; Wilson, DiZinno et al., 1995; Wilson et al., 1993). While the removal of sample 345 was justified through pathological analysis, a second sample in the heteroplasmic tissues was also seen to contain three heteroplasmic SNP positions: sample 261. SNaPshot™ analysis of sample 261 produced the following: 16223 C/T, 16324 G/T, and 16519 G/A. Microscopic examination of the 261 cells revealed normal hepatic cells with no visible pathology. Due the rationale of facing three possible heteroplasmic points without pathological proof of questionable cells, this sample was ultimately removed from the pool of eligible heteroplasmic tissues due to concerns over possible contamination and/or mixtures of non-specific cells.

Microscopy & Laser Dissection Microscopy

In the process of developing the protocol for the use of the laser dissection microscope, it was found that two out of the three potential sources of error were identified as issues with the robotic movement and control of the stage. It was seen that these general errors could be controlled by establishing an order of the processes used to collect the cells from the slides. Laser dissection microscopy operates based on the energy of the laser and the speed of the robotically controlled microscope stage; the combination of the two relates to how efficiently the cells are collected. It was seen that the most effective collection of the cells was obtained when

the identified cell was cut from the slide first. This allowed the cell and cut slide membrane to sit on the slide without the influence (ionic, hydrophilic) of the collection cap. The collection cap was then moved into the appropriate position above the cut cell, and then assuming the height of the collection cap above the slide surface was confirmed and valid, the cell would be catapulted off of the slide and into the collection cap. The collection cap would then move up and away from the slide to allow for removal and closure of the collection tube. While time-consuming, this manual control of the robotic stage allowed the greatest control over the collection cap height above the slide, and therefore minimized the error in cell collection. By adhering to this methodology, the unique separation and collection of the single cells was seen repeatedly during cell collection, re-affirming the rationale that a single cell was successfully collected when the proper collection procedures were followed using the laser dissection microscope.

The final point of error identified in the process of laser dissection microscopy was found to be the inability to visually confirm the collection of the captured cell within the microcentrifuge tube cap. While the laser dissection microscope contained the ability to microscopically inspect the interior of the cap, there was no way to realistically find the single cut cell that was on the order of 20-30 microns wide within a possible field ¼ inch in diameter, and of varying focal depths due to the concave meniscus of the liquid inside the cap. Therefore, it was assumed that if the cell was seen to successfully leave the slide during to the laser catapulting, and the cap had been at the appropriate height, the cell was collected. Due to the inability to visually confirm the physical cell collection, the study relied on the indirect verification of the presence of amplified mtDNA from the single-cell samples to show that the collection of the cells was successful.

Cell Lysis and One-Step Amplification

As the samples moved on to cell lysis and one-step amplification, it was again noted that there was no confirmation of the presence of the single cells in the collection tubes. An additional potential source of error was identified in the manual transfer of the collected cells, from the 1 ml collection tubes to the 0.5 ml amplification tubes. In order to carry out this procedure, the 1 ml collection tubes were centrifuged for three minutes at high-speed to bring the collection buffer and the single cell out of the collection tube cap and into the bottom of the tube. The 15 µl of collection buffer and the single cell were then transferred using a large bore 1 ml pipette tip, to ensure that the separated cell and the surrounding support membrane would be taken up along with the collection buffer. Since the collection buffer already contained the water and BSA components used in the amplification setup, the entirety of the collection buffer needed to be transferred to ensure that the correct volumes of water and BSA were present in the one-step amplification. As before, there was no way to confirm the presence of the single cell in the original 1 ml tube cap, or the 0.5 ml tube after the centrifuging and transfer of the capture buffer. As the individual PCR reagents were added to the 0.5 ml amplification tube, the tube was mixed and re-centrifuged prior to the one-step amplification. Following the one-step amplification, the samples were run on the Agilent 2100 Bioanalyzer for quantitation and sizing. At this point, it was unknown if the tubes contained successfully lysed and amplified single cells. However, the positive control contained 100 ng of HL60 DNA prior to amplification, and the negative control tubes, the AN and CAN, should have contained no detectable DNA. The review of the Agilent 2100 Bioanalyzer data for the single cells therefore served to demonstrate that the positive control contained detectable DNA at the correct size, and the negative controls were indeed negative. For the single cell samples, it was assumed that if a peak equal to or greater than the

size of the positive control would have been seen, it could have been indicative of contamination or the presence of more than one cell. The fact that only a small percentage of the single-cell samples showed any peak data whatsoever was seen as a positive indication that the tubes did not have widespread contamination, or additional cells present in the sample. Because the Agilent 2100 Bioanalyzer was then used to screen the one-step amplified samples for contamination rather than actual peak sizing, twenty-four representative samples could be screened (which included the positive and negative controls) to account for the sixty total samples throughout the amplification set. The benefit of using a representative set of samples was that it reduced the amount of reagents and consumables necessary for the Agilent 2100 Bioanalyzer assay. At the same time, it left the option available to run additional samples if any of the representative samples displayed peak data equal to or greater than the size of the positive control, which would have indicated a problem with the cell collection or amplification setup. The question that was then posed by the review of the Agilent 2100 Bioanalyzer data was whether or not the lack of a sizing peak in the data was due to a true below-threshold level of amplified mtDNA from the single cell, or if it was instead a negative sample. Due to the limits of the sensitivity of the Agilent 2100 Bioanalyzer, the data gave no indication as to why the samples did not show a peak. While there were other sizing instruments available, such as the UV-Spec NanoDrop™, this option was decided against due to the fact that the spectrophotometer used virtual calibrators for the DNA quantitation assay, and there was no way to account for the low-end sensitivity of the detection based on the overall working range of the instrument using these virtual calibrators. Also, the consumption of a larger volume of amplified mtDNA and the low throughput of the instrument added to the disadvantages of use of the spectrophotometer. Other options, such as gel electrophoresis using agarose or polyacrylamide also failed to account for the low end of

detection of the amplified mtDNA from a single cell, and were likewise not attempted. Additional considerations regarding the budget and available resources were taken into account when considering these alternative measurements and these financial factors also led to the decision to use the Agilent 2100 Bioanalyzer following the one-step amplification of the single cells.

A question was raised at this point as to if the concentration of the amplified mtDNA could be estimated or calculated based on anything other than the Agilent 2100 Bioanalyzer data. Previous studies investigating ancient DNA and the sequencing of Neanderthal DNA proved the applicability of the input and analysis of low-level DNA testing (Hofreiter et al., 2001; Noonan et al., 2006). An in-depth study of SNP typing of ancient DNA samples on a MALDI-TOF instrument platform demonstrated the successful analysis of picogram quantities of both mtDNA and nDNA samples (Mendisco et al., 2011). The separation chemistry used on the Agilent 2100 Bioanalyzer for this study had a working range of 0.5-50 ng/µl, so can be estimated that the amplified single cell samples that generated no detectable peaks were at a concentration below 500 pg/µl. Without further data points for comparison any attempted calculation of the actual mtDNA concentration would be arbitrary. If applied to the input volume calculation for the SNaPshot™ assay, this arbitrary value could negatively affect the success of the SNaPshot™ amplification and detection of the HVI SNPs.

Based on the analysis of the Agilent 2100 Bioanalyzer data on the single cells, and also based on the knowledge of the length study, it was determined that the SNaPshot™ amplification would ultimately address the question as to whether or not a single cell was successfully captured, transferred, and amplified. The initial single-cell amplification trials were successful in demonstrating that a captured cell could be successfully amplified and detected after

SNaPshot®, but the multiplex amplification of these single cells was seen to be problematic. When the decision was made to use only the specific extension primer for the noted heteroplasmic SNP in the single cells, this reduced the overall complexity of amplifying and analyzing the final samples, but not before the validation of the multiplex was established in SNaPshot®. In retrospect, the use of the multiplex was beneficial to the analysis of the gross tissue, but was unnecessary for the single cells. When designing the multiplex assay of the five SNP locations, it was determined to pool the amplification primers and amplify the extracted mtDNA, rather than reverse the approach and amplify the individual primers, run the individual SNaPshot™ extension primers, and then pool the SNaPshot™ amplicons prior to CE analysis. While the latter would have reduced the complexity of the validation process by not having to configure balanced melting temperatures and cycle-numbers for the multiplex assay, it would not have been cost-effective in that it would have required a four-fold increase in the SNaPshot™ reagents, which were already established as a limiting financial factor in the study.

Despite the complexity of validating the multiplex amplification for use in the project, there were inherent benefits from the use of the multiplex. The ability to screen the five SNP hotspots across a wide number of samples was critical in the design of the early project, and this was achieved using the multiplex setup. A second benefit of the multiplex was the decision in the design of the project to limit the multiplex to only five of the HVI hotspot locations. While more locations were reported in the literature, limiting the project to only five of these hotspots allowed for a spread of the data across the HVI fragment and permitted for a reasonable time-frame to validate the multiplex design. If the multiplex had been increased to include additional SNP hotspot locations, this would have increased the complexity of the multiplex, the complexity of extension primer design, and required additional time to balance the multiplex

reaction for all of the additional primers. The validation and testing of only five hotspot locations required just under one year of testing, analysis, and retesting to ensure that the multiplex was operating properly, and this timeframe would have been increased considerably with the addition of other SNP locations.

Despite the complexity and the time investment in the validation of the multiplex, the amplification was ultimately seen to perform to the original expectations, and was critical in the identification of the heteroplasmic gross tissue samples. However, the results of the SNaPshot™ data were not fully understood until after the analysis and interpretation of the length study were determined and the results of such were applied back to the original gross tissue multiplex data. While this detour in the project plan did delay the timeframe somewhat, it was critical for this length study to be done so that the proper interpretation of the gross tissue multiplex data could be carried out. In hindsight, this error could have been avoided by the use of a commercially available primer focus kit from Applied Biosystems (Applied-Biosystems, 2000). This commercially available kit would have artificially applied the four different dye-labeled ddNTP bases to the extension primers and allowed a rapid assessment of the separation size and expected pattern of each extension primer. However, at the beginning of the study, it was unclear why this kit would be of value to the study, and it was decided to allocate the limited funds towards the maximum number of SNaPshot™ amplification kits that could be purchased. This error, while costly in the sense of time spent on the project, was abated by the creation of the internal primer length study. The internal length study produced the same results that would have been seen in the use of the commercial kit, for much less of the cost of the commercial kit when averaging the cost of the individual SNaPshot™ reactions that were used to carry out the length study. The ultimate effect on the SNP multiplex validation was simply that the value of

the length study was not identified in the initial development plans, and additional time was then required at the latter part of the project to rectify this error. Ultimately, the combination of the validation of the multiplex, the addition of the length study, and the final interpretation of the gross tissue data and the single-cell data was seen to be the majority of the time invested in the project as a whole. The successful identification and interpretation of the data hinged on the validity of the multiplex and SNaPshot™ amplifications, so in the end, the time investment in these critical processes was seen to be beneficial to the study.

Extraction Negative Control Analysis Issues

In the review of the project design, the question was raised as to the elimination of samples based on the success of the extraction negative control data. The issue of identifying contamination in the extraction negative control and then eliminating the corresponding samples effectively reduced the overall number of eligible samples for the study, but might have also eliminated a true heteroplasmic sample before it could be screened with the multiplex assay. To counter this impression of sample loss, the extraction negative controls were established to control for small extraction sets containing between five and twelve samples, such that if a negative control failed, a large bulk of samples would not be compromised. Also, the establishment of the extraction negative control as a check against contamination was scientifically sound, and extremely conservative. While it is true that the only contamination in the extraction set might have been in the extraction negative control, and the corresponding samples may not have been affected, it was not seen as sound scientific practice to ignore the results of the negative controls just to bolster the number of available samples for the study. A second claim, of establishing an extraction negative control threshold (as was done for the amplification negative controls in the multiplex and SNaPshot™ analysis) was also seen as

imprudent at the extraction level, in such that it was possible to establish a clean extraction negative control with proper adherence to the laboratory decontamination protocols along with proper reagent and sample handling practices. It was seen in all passing extraction negative control samples that no detectable level of DNA was present at the amplification level, therefore ensuring that the corresponding gross tissue sample extracts would also be free of contamination and eligible for amplification in the SNP multiplex.

SNaPshot™ Assay Issues

Throughout the processing and analysis of the gross tissues and single cells in the SNaPshot™ amplification system, several issues were noted. The prominent issue throughout the process was the question of how to appropriately analyze the negative controls from the nested PCR reaction. Specifically, the analysis of the amplification negative (AN) controls and the cycling amplification negative (CAN) controls. Due to the nature of the nested PCR reaction, it was seen throughout the validations as well as the literature review that the nested amplification of the negative controls could be generating interpretable data that would need to be addressed (Budowle et al., 2002a; Budowle et al., 2002b; Budowle, Wilson, & DiZinno, 1999; Carrecedo et al., 1998; D'Eustachio, 2002; Grzybowski, 2000, 2001; Grzybowski et al., 2003; Musgrave-Brown et al., 2007). It was noted during the validation of the multiplex that the AN and CAN controls did produce data, but the peaks heights for these controls were several-fold lower than the data generated for the positive control and tissue samples. Throughout the validation and testing phase, it was seen that if the AN and CAN data values were used to establish a baseline for a minimum interpretable peak height, a majority of the data would then be eligible for interpretation, and lower-quality or questionable data would be excluded. This was shown to be reproducible in both the gross tissue test samples as well as the single-cell test

samples, and was viewed as a sound and conservative approach to the interpretation of the electropherogram data.

The peak height values were chosen to establish this negative control threshold rather than the peak area. The peak area took into account the area under the peak and did provide information as to the quantity of the data signal. However, issues such as peak broadening and shoulder peaks, due to electrophoretic effects and proximity to other peak data, skewed this peak area value. These effects led to an over-estimation of the true peak area value. Therefore, the peak height, which was the Y-axis value of the peak data from the electropherogram, was seen as an appropriate value from which to evaluate the results. When aligned with the peak size (which in turn related to the SNP analysis bin and therefore the SNP hotspot in question), all peaks from the same peak size were analyzed together and compared against the negative control threshold; the peaks with heights below the negative control threshold were excluded from the study, and the peaks above this value were deemed valid for interpretation.

The question was then raised as to the validity of applying this negative control threshold to the data set. Traditionally, any signal in a negative control would indicate that the negative control failed, which would commonly invalidate the associated samples linked to the control. In this study, the negative controls were seen to pass at the first level of amplification and analysis, demonstrated in the Agilent 2100 Bioanalyzer data for each amplification negative control and the cycling amplification negative control. It was only following the nested PCR of the SNaPshot assay in which detectable signals appeared in the negative controls. While it's true that the overall amplification cycling numbers were standard for the initial amplification (34 cycles) and for the SNaPshot amplification (25 cycles), the combined total of the amplification cycles approaches the original cycling numbers controversially reported by Grzybowski et al.,

which was later heavily attacked upon peer review (Budowle et al., 2002a; Budowle et al., 2002b; D'Eustachio, 2002; Grzybowski, 2000, 2001; Grzybowski et al., 2003). To guard against the erroneous report of artificial heteroplasmy, the best practice for the SNP data was determined to be the establishment of the amplification threshold based on the negative control. While other samples in the data sets might have actually had a true negative result, and would better serve as placeholder negative control samples, the consistent and conservative application of the negative control threshold ensured that no data under this threshold was utilized for the interpretation of the results. While this might have actually removed valid heteroplasmic samples (or a single value from a heteroplasmic sample, creating a homoplastic sample instead), this was seen to be acceptable as long as the use of the negative control threshold was consistently applied to every data set.

At another point during the testing and validation, the question was posed as to which negative control, the AN or CAN, was more appropriate for the use of establishing the negative control threshold. As both samples were created simultaneously, and carried through the testing and analysis simultaneously, it was determined that either could serve as the source of the negative control threshold. The CAN control was originally established during the validation to control for any additional rounds of multiplex or single-primer amplification that would have been required following the initial Agilent 2100 Bioanalyzer analysis of the data. In practice, neither the gross tissue samples nor the single cell samples required this additional amplification, so the CAN was never utilized for the intended purpose as a secondary amplification control. However, because the AN and CAN controls were therefore parallel identical controls, it was decided that arbitrary selection of one control over the other to establish the negative control threshold would not be appropriate. It was decided that the AN would be the sole source of the

negative control threshold, and the CAN values would be noted but not analyzed further in the process of interpreting the SNaPshot™ peak data for both the gross tissues and the single cells. In the review of the final single cell data, it was noted that if the CAN would have been used as the negative control threshold, then eight of the CAN values for the data would have been less than the AN value, while seven of the CAN values would have been greater than the AN value, so the results would have been negligible on the final data analysis (see Table 46, below). What is important to note in the review of the control peak height data is that the AN peak heights were never greater than the PE peak heights, which was critical for the successful analysis of the corresponding data.

Table 46- Comparison of AN/CAN control peak heights from SNaPshot™ data

Sample	Positive Control	Amplification Negative Control	Cycling Amplification Negative Control	CAN +/- compared to AN
284-16069	8139 T	520 C	41 C	-
315-16069	8341 T	295 C	Neg	-
293-16223	8073 C	593 C 363 A 164 T	705 C 460 A 246 T	+ + +
314-16223	5806 C	1584 C	1698 C	+
289-16324	8433 T	494 T	1998 T	+
349-16324	8346 T	34 G 1156 T	41 G 498 T	+ -
339-16390	4779 C	272 C 128 T	128 C Neg	- -
355-16390	5870 C	2677 C 138 T	177 C Neg	- -
355-16519	6082 A	70 G	Neg	-
369-16519	8223 A	197 G	669 G	+

Aside from the question of the use of the AN over the CAN, another issue brought to light in the analysis of the controls was the necessity of establishing a novel negative control for the SNaPshot™ amplification step. While the SNaPshot™ reaction had internal controls which were denoted as the SNaPshot™ Positive Control (PC) and SNaPshot™ Negative Control (NC), these controls were only controlling for the actual SNaPshot™ reagents and reaction, and did not include the specific extension primers used to probe the amplified mtDNA. The primers added to these SNaPshot™ controls were included in the SNaPshot™ kit from Applied Biosystems, and as such, controlled only for the validity and effectiveness of the SNaPshot™ extension/termination PCR reaction. It was debated as to whether or not a novel negative control

should have been established at this step to control for the addition of the extension primer to the SNaPshot™ reaction, but it was ultimately determined to be unnecessary. The AN and CAN controls effectively served this purpose, by controlling for the addition of the extension primers but also ultimately controlling for the nested PCR reaction itself. If a novel negative control was established at the SNaPshot™ step, it would have only served to contain the SNaPshot™ reagents and specific extension primer(s) for the multiplex or for the SNP of interest. However, the AN/CAN controls already served to control for the addition of the extension primers to known negative samples (based on the Agilent 2100 Bioanalyzer data) and the SNaPshot™ Negative Control controlled for the actual SNaPshot™ reagents. The combination of the AN and NC controls therefore served to effectively control for the SNaPshot™ reaction with the extension primers. A benefit to not establishing an additional negative control was that the SNaPshot™ reagents were not consumed further in the establishment of an unnecessary sample, reserving these expensive reagents instead for further analysis of the project samples.

A final issue encountered with the SNaPshot™ amplification was the design of the extension primers. Based on the available literature and discussions with colleagues with experience in the use of the SNaPshot™ amplification, it was decided that the length modifications for the extension primers would be done using thymine bases (Axler-Diperte, 2010; Brandstatter, Parsons, Niederstatter et al., 2003; Brandstatter, Parsons, & Parson, 2003; Coble et al., 2006; Kline et al., 2005; Li et al., 1999; Parsons, 2006; Sobrino et al., 2005; Vallone et al., 2004). It was also decided that a four base-pair separation between the extension primer sizes would be sufficient to separate the extension primers within a multiplex analysis. It was not until the length study was performed and analyzed that it was realized that the four base-pair separation was the minimum requirement, and while it allowed for the accurate interpretation of

the SNaPshot™ peaks, the additional separation of up to six base-pairs between the extension primers would have been more beneficial to the separation and analysis of the dye-labeled extension primers. Ultimately, however, the major design flaw in the extension primers was seen to be the use of the thymine bases as the primer length modifiers. While investigating the issues surrounding the 16519 data (in which G peaks were detected in the single cell data where A peaks were expected based on the gross tissue analysis), it was noted that one factor affecting this questionable data was the thymine bases on the extension primers. Based on the Applied Biosystems SNaPshot™ Multiplex Kit Protocol, the following was noted (Applied-Biosystems, 2000):

> "Poly (dT), poly (dA), poly (dC), and poly (dGACT) are 5′ non-homologous tails which are predicted to have minimal secondary structures. They have all been used successfully. Generally the signal patterns are not affected by the kinds of tails that are used. The 5′ poly (dT) tails however may interfere with the addition of 3′ ddA." (Emphasis added.)

The 16519 extension primers were the most heavily modified extension primers in the analysis set, and as a result, required eleven thymine bases to be added to the 5′ end of the extension primer. However, the original 5′ base in the unmodified 16519 extension primer was already a thymine base, which resulted in a total of twelve thymine residues on the 5′ end of the 16519 extension primer. When applied to the 16519 single cell samples, the resulting G peaks and the lack of the expected A peak might have been as a result of the interference of the poly-thymine tail, as noted in the SNaPshot™ Kit Protocol. Similar results were seen in the 349 cells tested with the 16324 extension primers and the 284 cells, tested with the 16069 extension primers. In the 349 cells, A/T heteroplasmy was expected based on the gross tissue analysis, but G and T peaks were found instead in the single cells. In the 284 cells, A/T heteroplasmy was also expected based on the gross tissue analysis, but C and T peaks were found instead in the

single cells. For these primers, the 16324 extension primer was modified with five thymine bases. However, the original 5′ base in the unmodified 16324 extension primer was already a thymine base, which resulted in a total of six thymine residues on the 5′ end of the 16324 extension primer. The 16069 extension primer did not contain a 5′ thymine base in the original sequence, and the 16069 extension primer was not modified with any thymine residues. Since the thymine modifications seemed to affect the 16519 extension primers but issues of non-addition of an expected A base were also noted in extension primers that were less modified and completely unmodified as well, there must have been an additional factor affecting the non-addition of the A base.

Ultimately, this non-addition of the A base may have been linked to the available concentration of the template mtDNA. The original gross-tissue screening of the 355 and 369 samples resulted in heteroplasmic G/A peaks at 16519, but when the single cells were analyzed, only G peaks were detected. More importantly, the gross tissue testing of the control sets 288 and 329 also revealed A peaks at 16519, but when the single cells were tested as controls with the 16519 primers, the cells generated clean data showing only G peaks across all twenty control cells between the two 16519 cellular amplification data sets. The concentration of the amplified mtDNA in these gross tissue samples was high enough to be detected in the Agilent 2100 Bioanalyzer assay and therefore allowed the optimal concentration of mtDNA template to be applied to the SNaPshot™ reaction, either in neat or diluted sample input volumes. In the case of the single cells, no detectable Agilent 2100 Bioanalyzer values resulted in a maximum neat input volume of the amplified single cell mtDNA, which may or may not have been within the optimal SNaPshot™ input concentration (Edenberg & Liu, 2009). As a result, the extension primers may have been simultaneously affected by the non-optimal amplified mtDNA template

concentration and the poly-thymine tail, resulting in the generation of only the G peak data and lacking the A peak. Similar results to the less modified 16324 extension primer and the unmodified 16069 extension primer might have also been due to this issue of the input mtDNA concentration.

The notion of mtDNA concentration dependency on the accuracy of the extension primer base addition was reinforced by the HL60 mtDNA positive control for the 16519 single cell samples, which successfully displayed an A peak after having a detectable mtDNA concentration on the Agilent 2100 Bioanalyzer and a calculated optimal mtDNA input amount into the SNaPshot™ assay. Since the HL60 positive control input mtDNA remained constant throughout the gross tissue testing and the single cell testing, the expected A peak seen in the gross tissue positive control remained consistent throughout the testing of the single cells. The end result of the 16519 analysis was that the data remained inconclusive for the G/A heteroplasmy in light of the possible extension primer issues. Unfortunately, the data concerning the 16069, 16324, and 16519 single cells in which the extension primers were screening for an A base was not generated and interpreted until the very end of the study, at which time the available funding was exhausted. Additional work to correct the extension primers would have resulted in the purchase and validation of new 16324 and 16519 extension primers with an alternate poly-nucleotide modifier, as well as purchasing additional SNaPshot™ reagents to re-run the 16324 and 16519 single cell amplicons. Since no modifications were done to the 16069 extension primers, it is unclear what could have been done to enhance the accuracy of the SNaPshot™ reaction. Additional funding could have been applied to attempt to detect lower levels of amplified mtDNA prior to the SNaPshot™ reaction to ensure that the proper input concentration was met.

SNaPshot™ Results on Single Cells

At the end of the study, a total of ten heteroplasmic samples were tested, each generating fifty cells for SNaPshot™ testing. Along with these ten heteroplasmic samples, two non-heteroplasmic control samples, tissues 288 and 329, each also generated fifty individual cells that were used as internal controls along with the heteroplasmic cell sets. In addition to these twelve samples, yielding a total of 600 individual cells, there were also two additional tissue sets that had cells isolated and tested in SNaPshot®. These two samples, 261 and 342, were both mistakenly identified as heteroplasmic samples prior to the interpretation of the length study data. As such, cells were collected and tested with SNaPshot™ from each of these samples, based on the interpretation of the data (at the time) showing possible heteroplasmy at 16390 in each sample. Following the length study, sample 261 was seen to retain the status of a possible heteroplasmic sample, but did not contain a heteroplasmic SNP at 16390. Rather, this 261 sample was one of the two samples that displayed three heteroplasmic points and was removed from the study as a questionable mixture sample, along with the other samples that contained problematic pathology. Sample 342, on the other hand, was shown not to contain heteroplasmy following the length study and reanalysis of the gross tissue data, so in essence the testing of the single cells from the 342 sample should act as an additional non-heteroplasmic control specifically for the 16390 primers. As with the heteroplasmic cell sets, the 261 and 342 cells also contained control cells from the 288 and 329 control cells as well, and these were also tested simultaneously in the SNaPshot™ amplification. The results from both the 261 and 342 16390 primer analysis will be discussed in greater detail at the end of this section, following the discussion on the ten heteroplasmic and two control cell sets.

In order to examine whether or not the heteroplasmic cells contained unique signals compared to artifacts (such as pull-up or non-specific dye artifacts), the peak heights from only the heteroplasmic cells were charted on X-Y scatter graphs, along with the negative control threshold from either one or both of the peak signals, if present. If the data points above one or both thresholds fell onto any normal slope of a line containing a (0,0) intercept, resulting in a best-fit R^2 value greater than 0.9, then it would be interpreted that the peak signals were dependent on each other and may be artifacts of the amplification/electrophoresis rather than true heteroplasmic signals. If the points do not have a best-fit value greater than 0.9, then it would be interpreted as the heteroplasmic base signals are independently occurring within the data. If a best-fit R^2 value was seen to be negative, it would be interpreted that the data set was following a non-linear trend, as the application of the best-fit value to non-linear data points typically result in negative R^2 values. These negative best-fit R^2 values would therefore be interpreted as occurring due to unique and individual values from each plotted peak height, indicating valid heteroplasmy.

In total, seven hundred and twenty cells were collected, amplified, purified, re-amplified with the SNaPshot™ system, and analyzed to generate the following results. As it has been previously noted, the SNaPshot™ data was the only verification that the single cells were successfully isolated and collected from the laser dissection microscopy process. Across the ten heteroplasmic cell sets, there were a total of twenty-five single-cell samples that gave no data. For the two non-heteroplasmic control cell sets, there were a total of seven single-cell samples that gave no data. In the additional 261 and 342 cell sets, an additional seven single-cell samples gave no data. If it is assumed that the lack of peak data is an indicator of a failed cell collection, the cell collection efficiency was 94.5%, with 681 of 720 cells successfully collected. If,

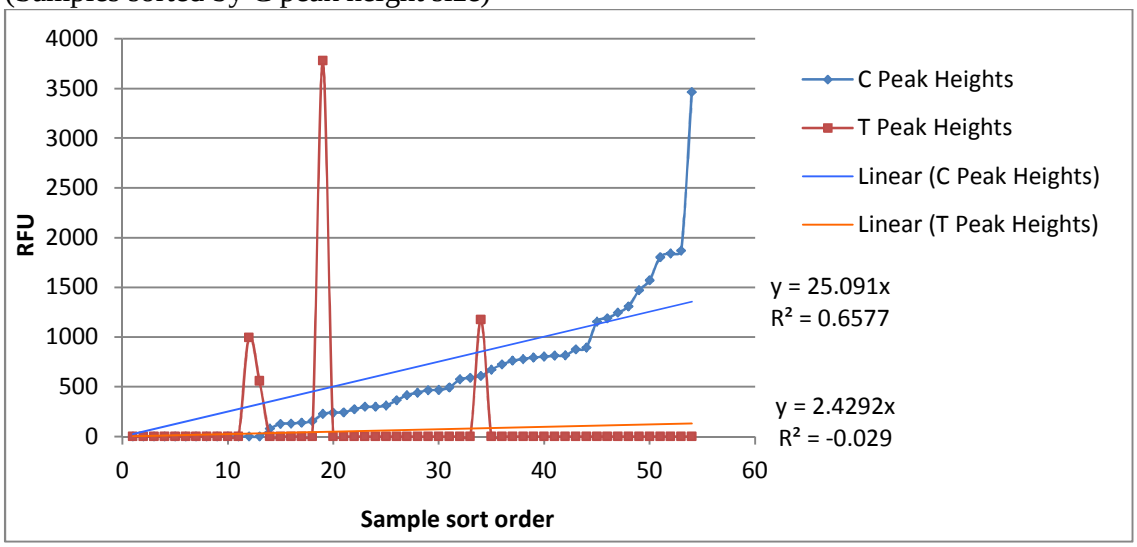
however, the negative results of the SNaPshot™ analysis are simply due to a below-threshold signal, and the cell was actually collected, the success rate would only increase, so the original estimate of 94.5% is more conservative, and shows that the cell collection methodology using laser dissection microscopy was successful in the identification and isolation of the single cells for the study. The consistency of the data across each fifty-cell sample set was another indicator that the collection was successful, such that stray or random SNP bases were not appearing in any significant frequency throughout the SNaPshot™ data. This was a more difficult aspect to measure, however, due to the inability of the A peaks to consistently generate within the single cells due to either the possible interference from the poly-thymine tails on the extension primers or issues related to the mtDNA concentration of the single cells in the SNaPshot™ assay. The issues of the non-addition of the ddATP bases notwithstanding, there were only three instances of unexpected peaks appearing in the single cell data, which is a second indicator as to the efficiency of the laser dissection microscopy in the isolation and collection of the single cells. It was ultimately seen that the laser dissection microscopy was effective at identifying and isolating single cells, but required strict adherence to the collection guidelines regarding the height of the collection tube cap above the slide, the use of the capture buffer, and the use of the reference cuts to repeatedly orient the microscope slide throughout repeated cutting events.

Discussion of SNaPshot™ Results of Sample 284 (16069) Cell Analysis

When the data from the twenty-six viable 284 cells and four valid control cells were analyzed, no instances of A/T heteroplasmy were present in any of the 284 cells (refer to Table 23 in the results section). The 284 cells, along with the 288 and 329 control cells only displayed either solitary C peaks or C/T peaks in the data. Out of the 284 cells, one sample (37-284) displayed a below-threshold C peak but did contain valid T peak data, so the C peak for this

sample was inconclusive and the T peak was valid. Two other samples (26-284 and 43-284) contained only valid T peaks, with no C peak data present. Out of the remaining twenty-three samples, all displayed C peaks at ~32 bp, and one of these samples also contained a T peak at ~34 bp (31-284). Typically the C peak height ranged from 500 to 3400 RFU, while the T peaks, when present, ranged in peak heights from 500 to 3700 RFU, so there was no clear indication of major or minor signals (see Figure 103). No A peaks were seen in any of the 284 cells or control cells whatsoever.

Figure 103- Peak Heights for 284 cells, 16069 primers
(Samples sorted by C peak height size)



Since none of the cells were seen to contain any A peaks in the ~32 bp region, it could be determined that one of three scenarios were occurring within these samples. The first scenario was that the tested cells were reflecting a true heteroplasmic C/T bases within the 16069 position of the mtDNA genomes. The second scenario was that the presence of the A base at the 16069 position was so minor within the mitochondrial DNA of the individual cells that it was being overwhelmed by the major C and T base signals. The third possibility was the issue of the non-addition of the ddATP bases in the extension primers due to the concentration of the mtDNA within the single cell samples.

When the original SNaPshot™ result for the gross tissue sample 284 was re-examined, it was seen that the T/A heteroplasmy ratio for the 16069 base position was not 1:1. Instead, within the gross tissue sample, the T peak was seen at 8417 RFU, where the A peak appeared at 983 RFU, and no C peak was detected at all. If it was assumed that the population of heteroplasmic mitochondrial DNA is not heterogeneous, and was split among unique cells (scenario #1 above), then within the viable 284 sample cells, a single cell containing the A base was not found. If it was assumed instead that a heterogeneous population of the heteroplasmic

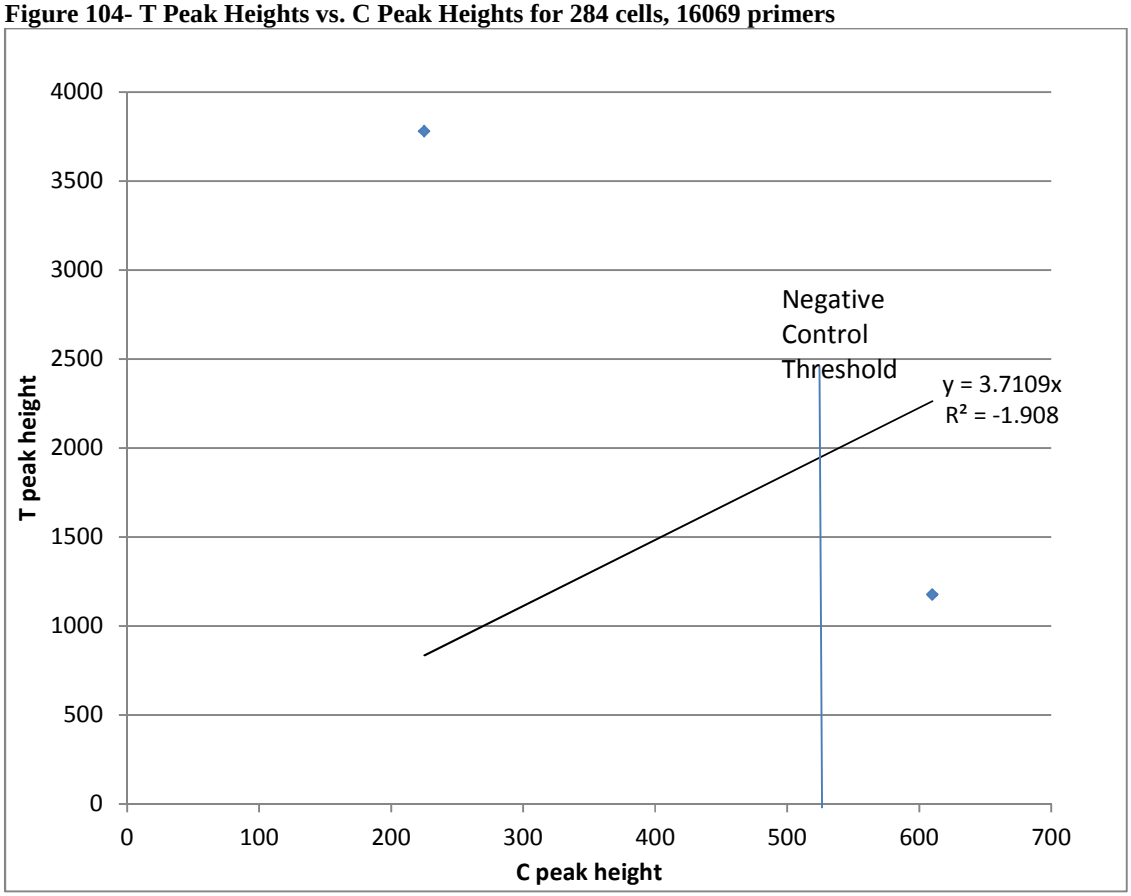
mitochondrial DNA exists within the cells (scenario #2 above), then the T bases are occurring over eight times more frequently than the minor A bases. The third scenario concerning the non-addition of the ddATP in the presence of the poly-thymine modified extension primer has already been identified as a source of error in the study, and was seen to be the most likely reason for the lack of A peaks in the data, especially for the 16519 extension primer with the longest poly-thymine tail. However, in the case of the 16069 primers, they were unmodified, and contained no thymine residue at the 5′ position.

Apart from the known issue with the extension primer and the ddATP non-addition in combination with the mtDNA input concentration, the first scenario was less likely, due to the previous assumption that by sampling fifty cells from a known heteroplasmic tissue at least one cell should show the base change, based on known rates of heteroplasmy in such tissues. Assuming the heterogeneous spread of the mitochondrial DNA was equal within the cells and tissue sample, this could have contributed to the results of the gross tissue heteroplasmy and the divergent C/T heteroplasmic individual cells within the 284 sample set. This issue ultimately focuses on the mtDNA concentration in the samples. Since thousands of cells contributed mtDNA in the gross tissue sample, the concentration of the mtDNA could be measured on the Agilent 2100 Bioanalyzer, and the culmination of the signals of both bases was above the baseline but was not reflective of the unequal nature of the C/T signal. At the same time, the A/T concentrations in the gross tissue were possibly in such high concentrations as to overwhelm a minor C signal also present in the data. In a single individual cell, the mtDNA concentration was not detectable on the Agilent 2100 Bioanalyzer, and the prevalence of any heteroplasmic A base was never at a high enough concentration to be detected above the major signal of the C base within the single cells. Since the A base was unable to be detected, this allowed the T base

signal to come through, but only when the T base was not overwhelmed by the C peak data. This is somewhat verified based on the graph of the C peak heights in relation to the T peak heights (Figure 103). As the C peak heights increased, no T peaks were present in the data. It was seen throughout the data that the addition of the ddATP bases was problematic when the mtDNA concentration was not optimal. For the 284 cells, it appears that the decrease in the overwhelming C peak signal allowed the T peak to be detected. The lack of the A base addition could have caused the previously undetected C base to be successfully detected in the single cells. Therefore, the mtDNA concentration of the single cells most likely led to results that were seen for the 284 cells.

In order to determine if the heteroplasmic data for the 284 cells was due to amplification and/or electrophoresis artifacts or not, the T and C peak heights from the heteroplasmic cells were plotted, as shown in Figure 104. Due to the fact that only two cells from the 284 set had detectable heteroplasmy, and only one of these was above the negative control threshold and valid for interpretation, it was impossible to determine if the heteroplasmic data was real or due to artifacts.

Figure 104- T Peak Heights vs. C Peak Heights for 284 cells, 16069 primers



When examining the 288 and 329 control cells that were run along with the 284 cells, one instance of heteroplasmy was seen; sample 30-329c displayed as a C/T heteroplasmy. The remaining three valid control cells all displayed a solitary valid C peak. As seen in Table 23, (pgs. 161-162), these control samples all showed the presence of a C base at the ~32 bp size. In the gross tissue analysis of these samples, the 288 and 329 tissues displayed C bases at the 16069 position for each of the samples. This calls into question the validity of the controls, and the specificity of the 16069 extension primers. Once again, this was most likely reflective of the issue of the non-addition of the ddATP base primarily due to the mtDNA concentration in the sample, since the 16069 extension primer was unmodified and contains no poly-thymine bases at the 5′ position. However, the presence of the T peak in the control sample leads to question if

the T peak was truly present in the control cells, or if it was a random addition to the extension primer. If the T peak in the single control cell was a true base, it means that it was at a high enough concentration in the sample cell to be detected and signaled by the 16069 extension primers. If that was the case, then why was the T peak not seen in the gross tissue analysis of the control tissues for 288 and 329? If the T peak was the result of random, non-specific base addition to the 16069 extension primer in the presence of low mtDNA concentration, then a bigger problem exists. If this T peak was random, non-specific addition, then the 284 single cell samples would be affected as well, resulting in inconclusive data for the 284 cell set due to the untrustworthiness of the results. However, the T peak was expected in the 284 cell set as part of the possible A/T heteroplasmy based on the gross tissue, so the presence of the T peak in the 284 cells is not believed to be non-specific addition, but instead true data. Also, it would be unlikely for the T base displaying in the control sample to be any type of artifact, since a fluorescently-labeled ddNTP base must be adding to the extension primer for the samples to be detected, and this can only be occurring when the template strand contains the corresponding A base.

To address the specificity of the primers, it should be noted that while no A bases were present in the 284 cells, the control cells, or any of the positive or negative controls, the 16069 primers did in fact show A base addition in the second 16069 cells set (cells 315, see next discussion section for details). Ergo, an A base can be properly added and detected in the 16069 extension primers. This reinforces the notion that the ddATP addition is tied to both the mtDNA concentration of the sample as well as the presence of the modified poly-thymine tails on the extension primers, since the 315 cells showed a viable A base signal and the 16069 extension primers contained no 5′ modification to interfere with the A base addition. In the case of the 284

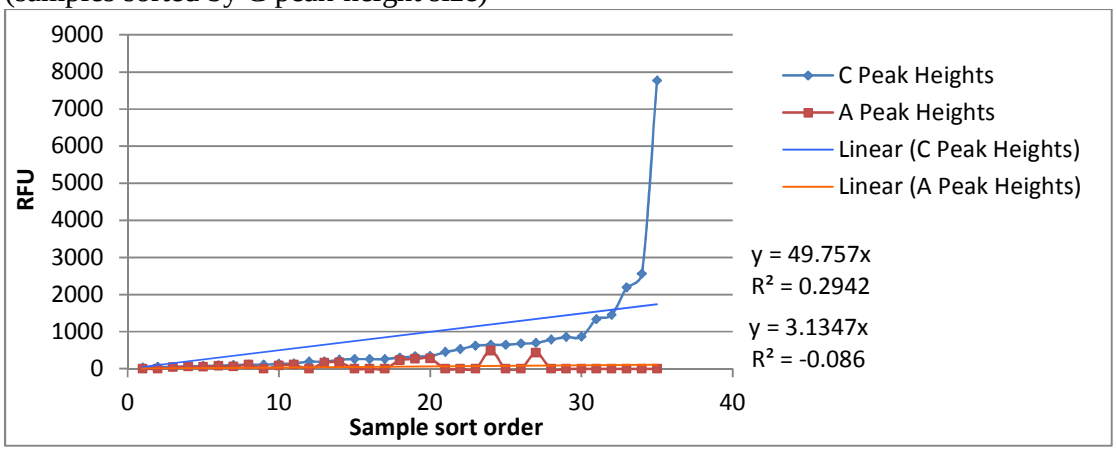
cells, the limiting factor for the A base addition might have been more affected by the available mtDNA concentration in the sample, causing the A bases not be detected.

As to the issue of the validity of the control samples, the C bases that were detected above the negative control threshold were of the correct bp size for the 16069 extension primer. Only three valid control cells are uniformly displaying the C base at 16069 along with the C/T mixture for the fourth control sample 30-329c. The gross tissues for the controls did not show an indication of any base other than the C at 16069 and as such, the implications of the controls samples are open to interpretation. However, the original rationale of targeting the hotspots in the HVI section of the mtDNA genome was to screen for higher instances of heteroplasmy, so it is not unlikely that the 30-329c sample might contain a valid T signal at the 16069 position within the single cell. If this is case, that solitary sample might not serve as a valid control, but would be an important data point to show heteroplasmy at the cellular level where no heteroplasmy was seen in the gross tissue. The original scope of the research allowed for this possibility to occur within the cells, and now if it is interpreted as a valid type, it is important to note.

Discussion of SNaPshot™ Results of Sample 315 (16069) Cell Analysis

When the data from the twenty-nine viable 315 cells and seven valid control cells was analyzed, it was seen that five instances of C/A heteroplasmy were present in the 315 cells (refer to Table 25 in the results section). The 315 cells, along with the 288 and 329 control cells displayed either solitary C peaks or C/A peaks in the data. Out of the 315 cells, one sample (3-315) displayed a below-threshold C peak but did contain valid T peak data, so the C peak for this sample was inconclusive and the T peak was valid. This was the only sample in the 315 cells and controls to display a T peak. Out of the remaining twenty-eight samples, ten samples displayed a valid A peak with a below-threshold C peak, so only the A peak was valid for those samples. The remaining eighteen samples consisted of the five C/A heteroplasmic samples and thirteen samples that only displayed a valid C peak. Typically the C peak height ranged from 300 to 7700 RFU, while the A peaks, when present, ranged in peak heights from 40 to 480 RFU, so the C peaks were seen as the major signals (Figure 105). Seven of the ten controls displayed a valid A peak, with only one of those seven control cells also showing a valid C peak (33-288c).

Figure 105- Peak Heights for 315 cells, 16069 primers
(samples sorted by C peak-height size)



Since only five of the cells were seen to contain any C/A heteroplasmy, it could be determined that one of three scenarios were occurring within the 315 cell samples. The first

scenario was that the viable cells were reflecting true heteroplasmic C/A bases or a homoplastic C base within the 16069 position of the mtDNA genomes. The second scenario was that the presence of the A base at the 16069 position was so minor within the mitochondrial DNA in all of the individual cells that it was being overwhelmed by the major C base signals, particularly in the samples which only displayed a solitary C base. The third possibility was the issue of the non-addition of the ddATP bases in the extension primers due to the concentration of the mtDNA within the single cell samples.

When the original SNaPshot™ result for the gross tissue sample 315 was re-examined, it was seen that the C/A heteroplasmy ratio for the 16069 base position was not 1:1. Instead, within the gross tissue sample, the C peak was seen at 8553 RFU, where the A peak appeared at 669 RFU, and no T peak was detected at all. If it was assumed that the population of heteroplasmic mitochondrial DNA is not heterogeneous, and was split among unique cells (scenario #1 above), then within the viable 315 sample cells, the five cells that contained this unequal partitioning of the heteroplasmic mtDNA were found. If it is assumed instead that a heterogeneous population of the heteroplasmic mitochondrial DNA exists within the cells (scenario #2 above), then the C bases are occurring over twelve times more frequently than the minor A bases, which is also somewhat verified by the 315 cell data. The third scenario concerning the non-addition of the ddATP due to the poly-thymine modified extension primer in combination with the mtDNA input concentration has already been identified as a source of error in the study, and was seen to be the most likely reason for the lack of A peaks in the other data sets, especially for the 16519 extension primer with the longest poly-thymine tail. However, in the case of the 16069 primers, they were unmodified, and contain no thymine residue at the 5′ position. It is critical to note that the 315 cells did show valid A peaks, which reinforces the

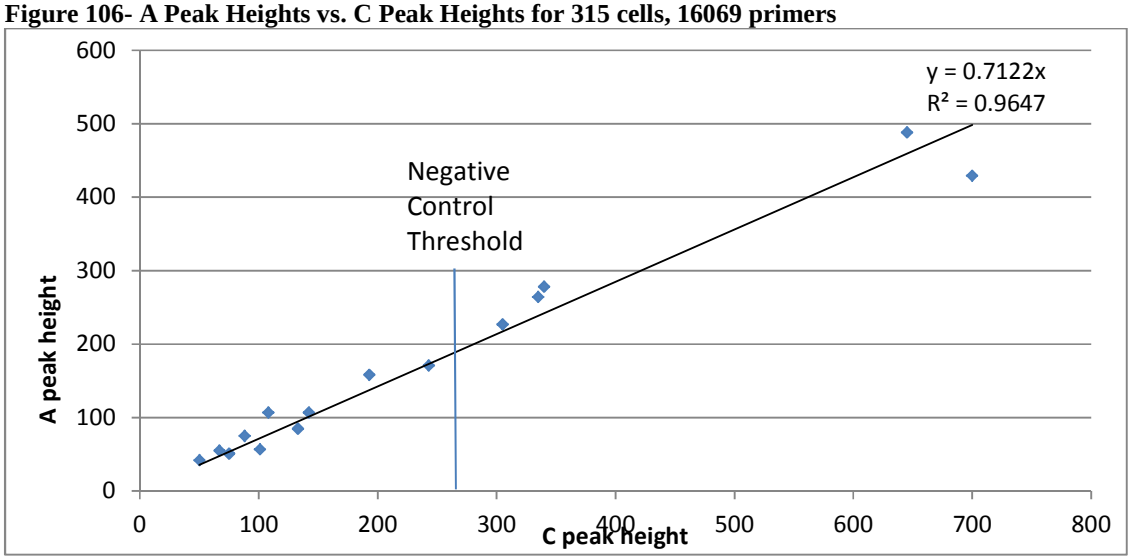
theory that part of the ddATP addition is tied to the poly-thymine modified extension primers. The 16069 extension primer successfully displayed the A peak in the 315 cells as a result of the unmodified extension primers.

Apart from the known issue of the ddATP non-addition, the second scenario was more likely, due to the previous assumption that by sampling fifty cells from a known heteroplasmic tissue at least one cell should show the base change, based on known rates of heteroplasmy in such tissues. Assuming the heterogeneous spread of the mitochondrial DNA was equal within the cells and tissue sample, this could have contributed to the results of the gross tissue heteroplasmy within the 315 sample set. In fact, for the ten 315 cell samples that displayed only valid A peaks, every sample in that set of ten also contained an invalid C peak that was below the amplification negative control threshold. Had these C peaks been above the threshold, an even higher number of heteroplasmic cells would have been uncovered in the 315 cells. This issue ultimately focuses on the mtDNA concentration in the samples. Since thousands of cells contributed mtDNA in the gross tissue sample, the concentration of the mtDNA could be measured on the Agilent 2100 Bioanalyzer, and the culmination of the signals of both bases was therefore above the baseline but was not reflective of the unequal nature of the C/A signal. Similarly, the C/A concentrations in the gross tissue were possibly in such high concentrations as to be detected in the SNaPshot™ assay and reflect this difference in the quantity of the C peaks compared to the A peaks. In the individual cells, the mtDNA concentration was not detectable on the Agilent 2100 Bioanalyzer, and the prevalence of any heteroplasmic A base was only at a high enough concentration in five of the fifty samples to be detected above the major signal of the C base within the single cells. This is somewhat verified based on the graph of the C peak heights in relation to the A peak heights (Figure 105). As the C peak heights increased, the A

peaks were not present. It was seen throughout the data that the addition of the ddATP bases was problematic when the mtDNA concentration was not optimal. For the 315 cells, it appears that the decrease in the overwhelming C peak signal allowed the A peak to be detected. The lack of the A bases in the cells that displayed only C peaks might be due to this non-addition of a valid A peak (as in scenario #2) or it is instead reflective of the actual homoplastic C profile in those cells (as in scenario #1). Therefore, the mtDNA concentration of the single cells most likely led to the results that were seen for the 315 cells.

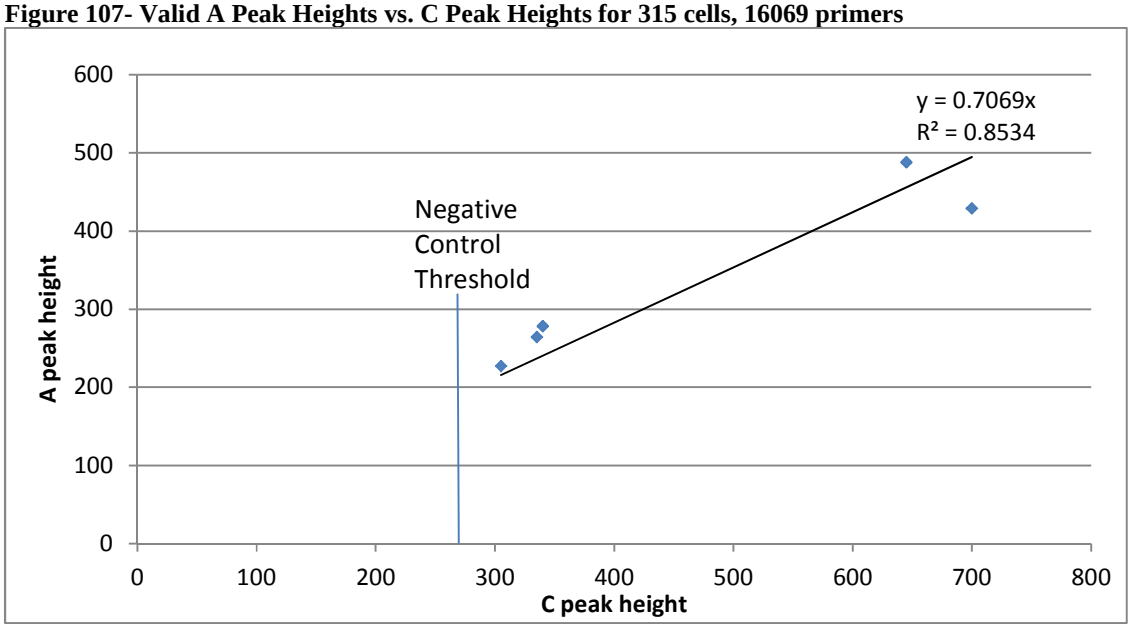
In order to determine if the heteroplasmic data for the 315 cells was due to amplification and/or electrophoresis artifacts or not, the A and C peak heights from the heteroplasmic cells were plotted, as shown in Figure 106. Since the C peaks contained a negative control threshold, the majority of the heteroplasmic points fall below this threshold, leaving only five valid heteroplasmic 315 cells. However, as seen in Figure 106, the totality of the heteroplasmic points fall on a common slope with a best-fit R^2 value of 0.96, thereby demonstrating a strong correlation of the A and C peaks. This data would indicate that the heteroplasmic A and C peaks in the 315 cells are not independent events and are possibly linked, therefore invalidating all of the heteroplasmic 315 cells due to potentially displaying amplification artifacts and not true heteroplasmy.

Figure 106- A Peak Heights vs. C Peak Heights for 315 cells, 16069 primers



However, when the invalid data points are removed from the plot (Figure 107), the best-fit R^2 value for the remaining points drops to 0.85, thereby indicating that while still close, the points are less correlated that before. Whether or not a true best fit can be established with only five data points a valid question, and it is possible that with additional valid heteroplasmic data points this best-fit value could go above the 0.90 threshold yet again.

Figure 107- Valid A Peak Heights vs. C Peak Heights for 315 cells, 16069 primers



When examining the 288 and 329 control cells that were run along with the 315 cells, one instance of heteroplasmy was seen; sample 33-288c displayed as a C/A heteroplasmy. The remaining six valid control cells all displayed a valid A peak but also contained C peaks that were below the negative control threshold, and were therefore invalid. As seen in Table 25, (pgs. 165-166), these control samples all showed the presence of an A base at the ~31 bp size, with the one valid C base appearing at ~ 32 bp. In the gross tissue analysis of these samples, the 288 and 329 samples were seen to be C bases at the 16069 position for each of the samples. This calls into question the validity of the controls, and the specificity of the 16069 extension primers. Once again, this is seen to be most likely reflective of the issue of the mtDNA concentration in the sample. However, in this case, the A peak was successfully added to the extension primers when the cells were at low concentrations, rather than the reverse. If that was the case, then why was the A peak not seen in the gross tissue analysis of the control tissues for 288 and 329? If the A peak was the result of random, non-specific base addition to the 16069 extension primer in the presence of low mtDNA concentration, then a bigger problem exists. If this A peak is random, non-specific addition, then the 315 single cell samples would be affected as well, resulting in totally inconclusive data for the 315 cell set due to the untrustworthiness of the results. However, the original rationale of the study design focused on the hotspot locations in the HVI mitochondrial genome. It is not unlikely that heteroplasmy would be found in these cells at these positions, even after not having detected heteroplasmy in the gross tissue. The design of the project proposed this exact occurrence. It might be possible that in the gross tissue of the control samples, there was an overwhelming C signal, which blocked out the ability of the A base to be reasonably detected. When isolating a single cell from these samples that contained this C/A

heteroplasmy, the overall concentration of the C base signal was decreased, allowing the A base to now be detected and displayed.

This might have only been possible for the 16069 extension primer as well, since it was the only extension primer to not contain any 5′ poly-thymine modification that was believed to interfere with the ddATP base addition in low mtDNA concentration. Also, it would be unlikely for the A base displaying in the control sample to be any type of artifact, since a fluorescently-labeled ddNTP base must be adding to the extension primer for the samples to be detected, and this can only be occurring when the template strand contains the corresponding T base.

To address the specificity of the primers, it should be noted that viable A bases were present in the 315 cells and the control cells, therefore an A base can be properly added and detected in the 16069 extension primers. This reinforces the hypothesis that the ddATP addition is tied to both the mtDNA concentration of the sample as well as the presence of the modified poly-thymine tails on the extension primers, since these 315 cells showed a viable A base signal and the 16069 extension primers contained no 5′ modification to interfere with the A base addition.

As to the issue of the validity of the control samples, the single C base that was detected above the negative control threshold was of the correct bp size for the 16069 extension primer. Since the valid control cells are uniformly displaying the A base at 16069 along with the C/A for 33-288c, the implications of the control samples are open to interpretation. None of the control cells displayed a T peak, as was seen in the 3-315 cell sample. This solitary T peak was a valid peak with a peak height of 425, and could have only occurred if the fluorescent ddNTP was added to the extension primer, with a corresponding A base in the template strand. Since no other instance of T peaks were seen in the other cells or controls for 315, it is assumed that this is

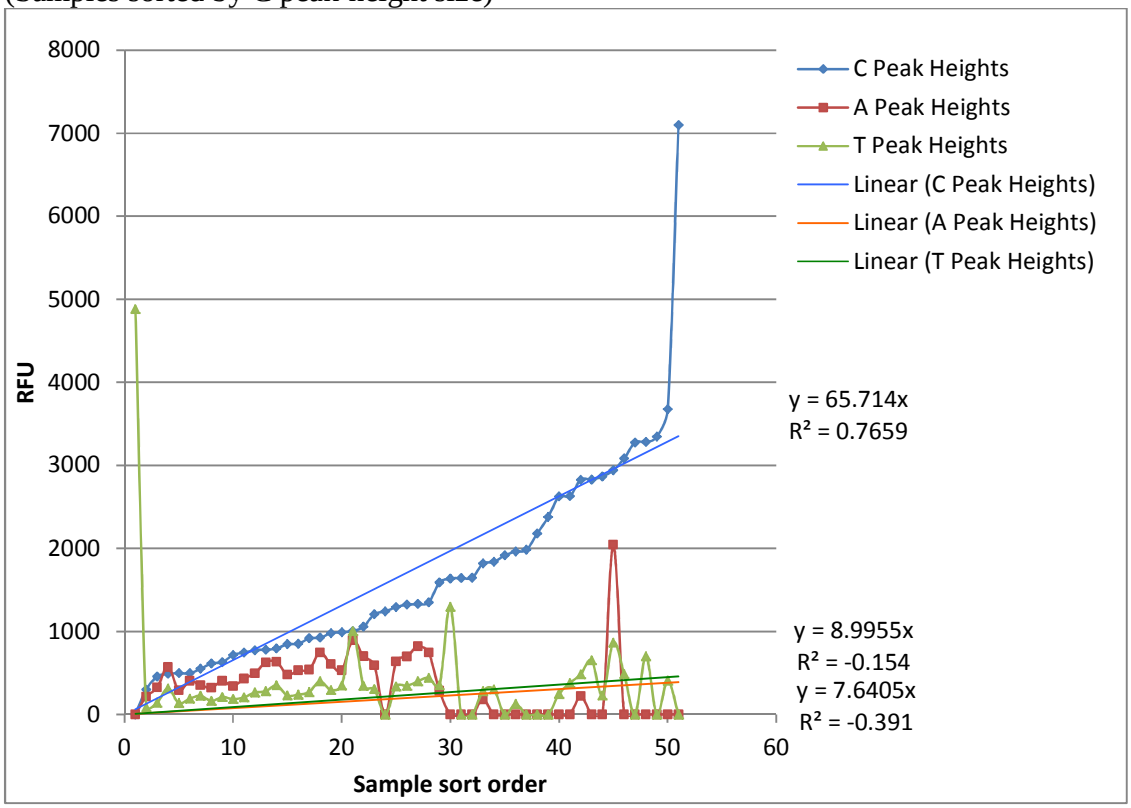
a valid T signal, but is not indicative of additional heteroplasmy since it was not verified in any of the other 315 cells or controls. This sample did contain a below threshold C peak, which was invalid, but compared to the T peak, the T was the major signal for that particular cell sample. On the whole, the negative control threshold for the 315 cells eliminated several samples containing a C signal that could have lent weight to the veracity of the heteroplasmy in the control cells and additional heteroplasmic 315 cells as well. However, the elimination of the invalid samples by use of the negative control threshold was the most conservative way to guard against false signals from artifacts, non-specific addition, and contamination. If the control cells are seen to be viable with the presence of the A signals and the incorporation of a single C/A heteroplasmic sample, then the preponderance of the 315 cells are viable and seen as valid instances of heteroplasmy in the single cells, within the limited scope of the five cells that contained this heteroplasmy.

Discussion of SNaPshot™ Results of Sample 293 (16223) Cell Analysis

When the data from the forty-seven viable 293 cells and ten valid control cells was analyzed, it was seen that a majority of the cells displayed some form of C/T heteroplasmy (refer to Table 27 in the results section). The 293 cells, along with the 288 and 329 control cells displayed either solitary C peaks or C/T peaks in the data. Only one sample (48-293) displayed a solitary T peak. The first twenty-nine consecutive single cells samples of the 293 cells also displayed an A peak as well. Out of the 293 cells, two samples (9-293 and 23-293) displayed a below-threshold C peak but did contain valid A and T peak data, so the C peak for this sample was inconclusive and the A and T peaks were valid. Two other samples (15-293 and 31-293) contained only valid C peaks due to invalid T or A/T peaks. Four other samples (19-293, 25-293, 26-293, and 29-293) contained valid C/T peaks and invalid A peaks. Out of the remaining thirty-seven samples, nine samples displayed only solitary C peaks at ~30 bp, nine others displayed heteroplasmic C/T bases with the C at ~30 bp and the T at ~33 bp. The remaining 19 single-cell samples displayed the C/A/T heteroplasmy, with the C base at ~30 bp, the A base at ~31 bp, and the T base at ~33 bp. Typically the C peak height ranged from 600 to 3600 RFU, while the A peaks, when present, ranged in peak heights from 400 to 2000 RFU. The T peaks, when present, ranged in peak heights from 200 to 4800, so there was no clear indication of major or minor signals. However, it was noted that as the C peak heights increased the A peaks were no longer present in the samples. It was noted that for the first 29 consecutive samples, where the A peak was seen (before the application of the amplification negative threshold) the C peak heights ranges from 600 to 1500 RFU, with two of the samples actually at 2800-2900 RFU. The 21 consecutive end samples, where no A peaks were detected, the C peak heights ranged from 1200 to over 3000 RFU, with one sample at 7100 RFU (Figure 108). No A peaks were seen in

any of the control cells whatsoever, with all ten control samples displaying C peaks ranging from 1400 to over 5800 RFU.

Figure 108- Peak Heights for the 293 cells, 16223 primers
(Samples sorted by C peak-height size)



Since thirty-four of the cells were seen to contain some form of C/T or C/A/T heteroplasmy, it could be determined that one of three scenarios were occurring within the 293 cell samples. The first scenario was that the viable cells were reflecting the true heteroplasmic C/T or C/A/T bases or homoplastic T base or C base within the 16223 position of the mtDNA genomes. The second scenario was that the presence of the A and T bases at the 16223 position was so minor within the mitochondrial DNA of the individual cells that it was being overwhelmed by the major C base signals, particularly in the samples which only displayed a solitary C base, or in the case of the missing A bases, only the C/T bases. The third possibility

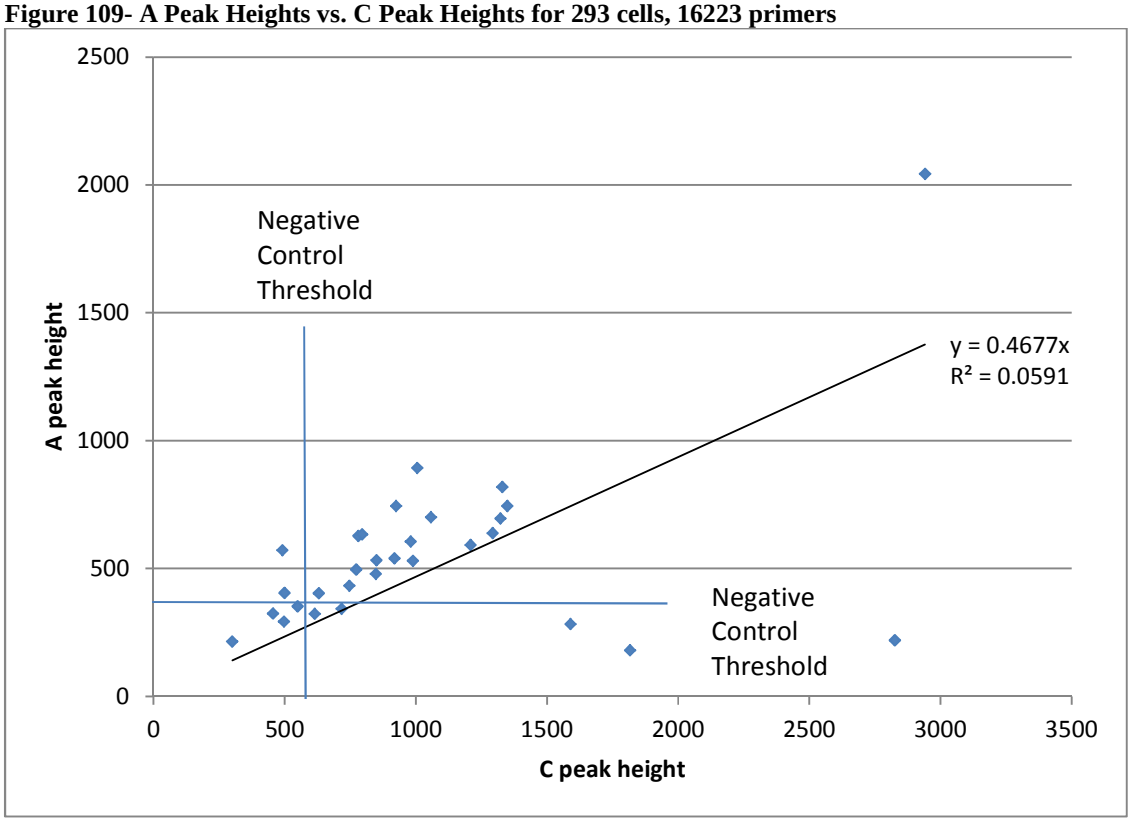
was the issue of the non-addition of the ddATP bases due to the extension primers in combination with the input concentration of the mtDNA within the cell sample.

When the original SNaPshot™ result for the gross tissue sample 293 was re-examined, it was seen that the T/C heteroplasmy ratio for the 16223 base positions was not 1:1. Instead, within the gross tissue sample, the T peak was seen at 1906 RFU, where the C peak appeared at 67 RFU, and no A peak was detected at all. If it was assumed that the population of heteroplasmic mitochondrial DNA was not heterogeneous, and was split among unique cells (scenario #1 above), then thirty-four cells that contained this unequal partitioning of the heteroplasmic mtDNA were found. If it was assumed instead that a heterogeneous population of the heteroplasmic mitochondrial DNA exists within the cells (scenario #2 above), then the T bases are occurring over twenty-eight times more frequently than the minor C bases. The third scenario concerning the non-addition of the ddATP due to the poly-thymine modified extension primer and the mtDNA input concentration has already been identified as a source of error in the study, and was seen to be the most likely reason for the lack of A peaks in the data, especially for the 16519 extension primer with the longest poly-thymine tail. However, in the case of the 16223 primers, they were modified with only one thymine residue at the 5′ position. Apart from the known issue with the ddATP addition, the second scenario was more likely, due to the previous assumption that by sampling fifty cells from a known heteroplasmic tissue at least one cell should show the base change, based on known rates of heteroplasmy in such tissues. Assuming the heterogeneous spread of the mitochondrial DNA was equal within the cells and tissue sample (scenario #2, above), this could have contributed to the results of the gross tissue heteroplasmy and the divergent C/T and C/A/T heteroplasmic individual cells within the 293 sample set. This issue ultimately focuses on the mtDNA concentration in the samples. Since

thousands of cells contributed mtDNA in the gross tissue sample, the concentration of the mtDNA could be measured on the Agilent 2100 Bioanalyzer, and the culmination of the signals was above the baseline but was not reflective of the unequal nature of the C/A/T signal. At the same time, the C/T concentrations in the gross tissue were possibly in such high concentrations as to overwhelm a minor A signal also present in the data. In an individual cell, the mtDNA concentration was not detectable on the Agilent 2100 Bioanalyzer, but the overall concentration of the C/T signal was lower, allowing the A base to be detected and not overwhelmed by the amplified C/T signal. This is somewhat verified based on the graph of the C peak heights in relation to if the A peak is present or not (Figure 108). As the C peak heights increase, the presence of A peak becomes less and less. It was seen throughout the data that the addition of the ddATP bases was problematic when the mtDNA concentration was not optimal. For the 293 cells, it appears that the decrease in the overwhelming C peak signal allowed the A peak to be detected. Therefore, the mtDNA concentration of the single cells most likely led to results that were seen for the 293 cells.

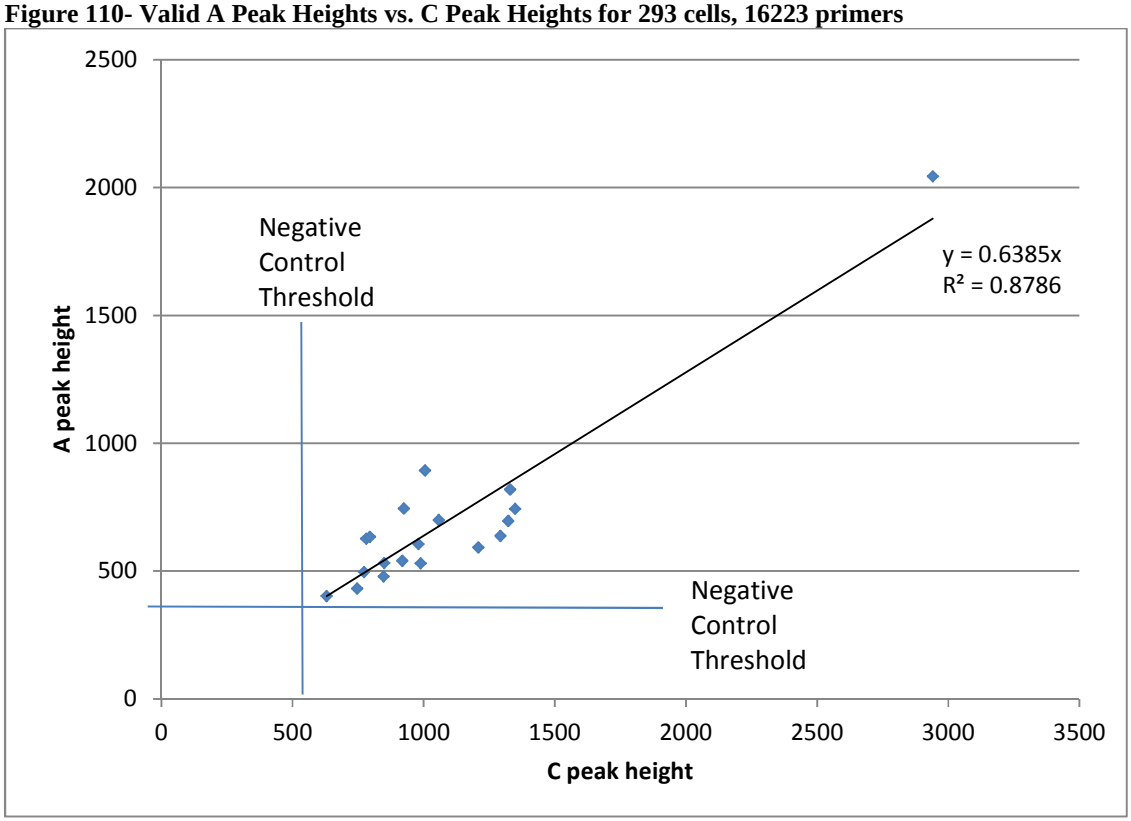
In order to determine if the heteroplasmic data for the 293 cells was due to amplification and/or electrophoresis artifacts or not, the A and C peak heights from the heteroplasmic cells were plotted, as shown in Figure 109. The negative control thresholds from both the A and C peaks boundary the valid cells, and invalidate eleven cells from the data set. The resulting best-fit R^2 value is 0.05, reflecting little correlation between the A and C peaks, indicating that they are independently appearing in the heteroplasmic cells.

Figure 109- A Peak Heights vs. C Peak Heights for 293 cells, 16223 primers



However, when the invalid cells were removed from the plot, the best-fit R^2 value improved dramatically, to 0.87 (see Figure 110). While this is below the 0.9 cutoff for correlation, it is very close, indicating that there might be more of a moderate correlation between the data points than true independent heteroplasmy. Since this new best-fit value is calculated by a representative number of valid cells, it would be unlikely to become less with the addition of additional valid data points, therefore, the 293 cells may not be exhibiting truly independent heteroplasmic peaks.

Figure 110- Valid A Peak Heights vs. C Peak Heights for 293 cells, 16223 primers

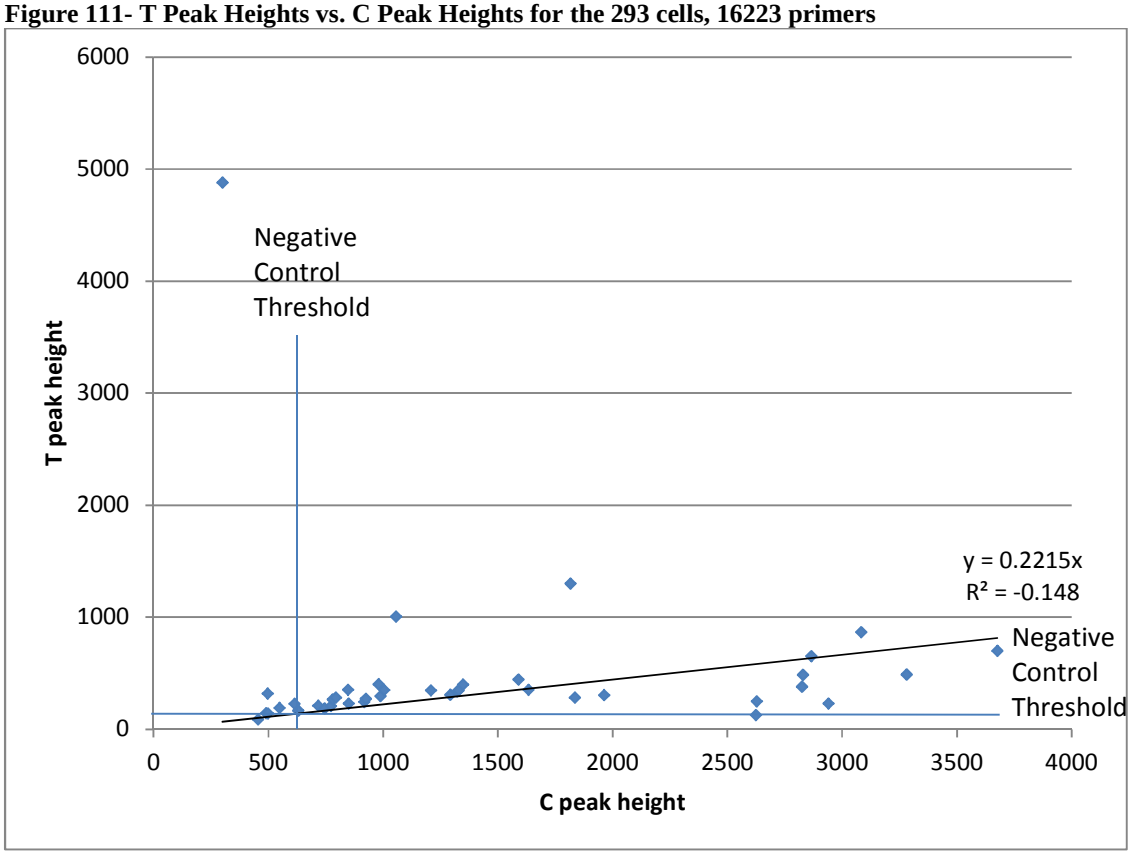


When examining the 288 and 329 control cells that were run along with the 293 cells, eight instances of C/T heteroplasmy were seen; only samples 36-288c and 40-288c contained solitary C peaks. As seen in Table 27, (pgs. 170-171), these control samples all showed the presence of a C base at the ~30 bp size, and when the T base was present, it was at the ~33 bp size. In the gross tissue analysis of these samples, the 288 and 329 samples were seen to be C bases at the 16223 position for each of the samples. This calls into question the validity of the controls, and the specificity of the 16223 extension primers. The presence of the T peak in the control samples opens the question if the T peak was truly present in the control cells, or if it was a random addition to the extension primer. If the T peak in the control cells was a true base, it means that it was at a high enough concentration in the sample cells to be detected and signaled by the 16223 extension primers. If that is the case, then why was the T peak not seen in the gross tissue analysis of the control tissues for 288 and 329? If the T peak was the result of

random, non-specific base addition to the 16223 extension primer in the presence of low mtDNA concentration, then a bigger problem exists. If this T peak was random, non-specific addition, then the 293 single cell samples would be affected as well, resulting in totally inconclusive data for the 293 cell set due to the untrustworthiness of the results. However, the T peak was expected in the 293 cell set as part of the possible C/T heteroplasmy based on the gross tissue, so the presence of the T peak in the 293 cells was not believed to be non-specific addition, but instead true data. Also, it would be unlikely for the T base displaying in the control samples to be any type of artifact, since a fluorescently-labeled ddNTP base must be adding to the extension primer for the samples to be detected, and this can only be occurring when the template strand contains the corresponding A base. It is also seen that the presence of the T base was not affected by the addition of the C base to the extension primer in the way that affected the A base addition (Figure 111). There was no decrease in the T peak signal as the C peak signal increased.

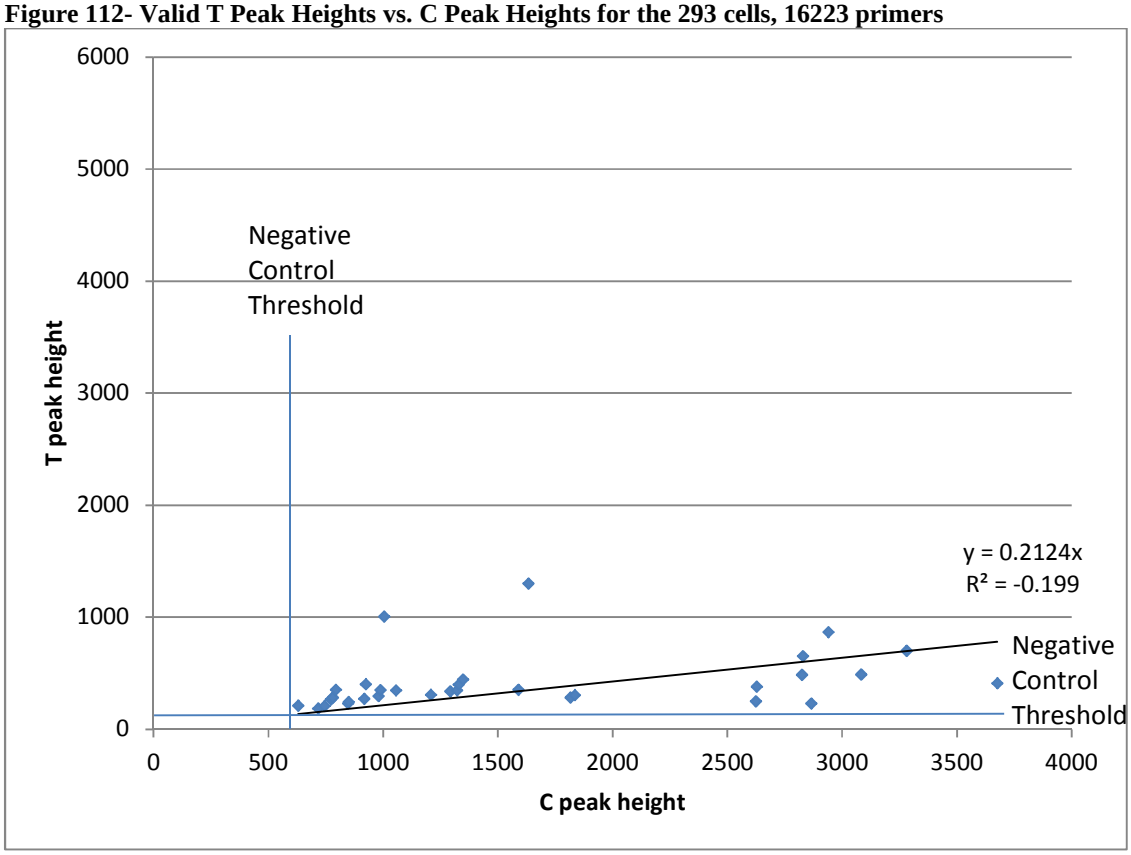
In order to determine if the heteroplasmic data for the 293 cells was due to amplification and/or electrophoresis artifacts or not, the T and C peak heights from the heteroplasmic cells were plotted, as shown in Figure 111. A majority of the data points for the 293 T/C cells are seen to be above both negative control thresholds, and therefore valid. The best-fit R^2 value for the data, however, was noted as -0.14, indicating that the data points may be non-linear, and therefore the best-fit correlation would not be useful is establishing the independence of the heteroplasmic data.

Figure 111- T Peak Heights vs. C Peak Heights for the 293 cells, 16223 primers



The removal of the invalid data points from the 293 T/C cells does little to change the overall best-fit calculation, as seen in Figure 112. The valid cells still show a negative R^2 value of -0.19, confirming that the cells are following a non-linear regression on the scatter plot. However, the variability of the data points does demonstrate the common trend noted before, where the increase in the C peak height has little effect of the overall T peak height, providing a strong indication that the T and C heteroplasmic peaks in the 293 cells are independent events, and therefore demonstrating valid heteroplasmy.

Figure 112- Valid T Peak Heights vs. C Peak Heights for the 293 cells, 16223 primers



To address the specificity of the primers, it should be noted that the presence of the A bases in the 293 cells and negative controls reinforces the notion that the ddATP addition is tied to both the mtDNA concentration of the sample as well as the presence of the modified poly-thymine tails on the extension primers, since the 293 cells showed a viable A base signal and the 16223 extension primers contained only a single 5′ thymine modification, which was probably not enough to interfere with the A base addition. In the case of the 293 cells, the limiting factor for the A base addition might have been more affected by the available mtDNA concentration in the sample, causing the A bases not to be detected when in the presence of larger C/T peak heights.

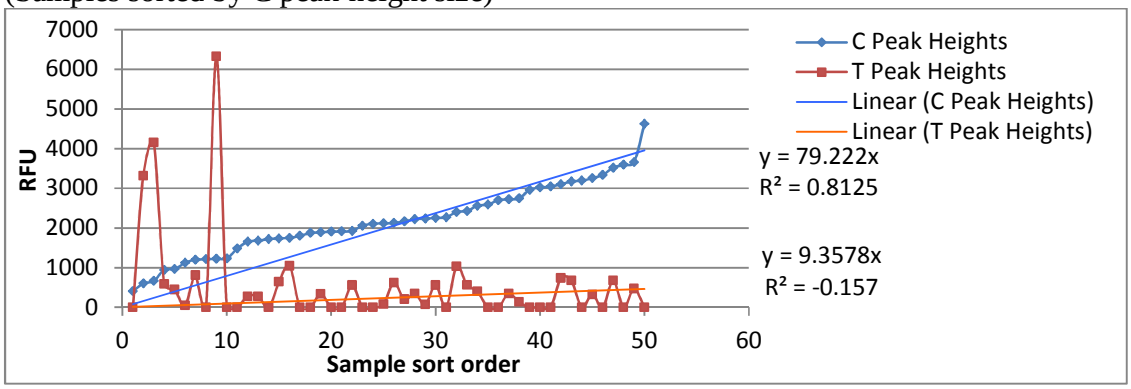
As to the issue of the validity of the control samples, the T bases that were detected above the negative control threshold were of the correct bp size for the 16223 extension primer. Since

the valid control cells are uniformly displaying the C base at 16223 along with the T base in eight of the samples, while the gross tissues for the controls did not show any indication of any base other than the C at 16223, the implications of the controls samples are open to interpretation. However, the original project design of targeting the hotspots in the HVI section of the mtDNA genome was to screen for higher instances of heteroplasmy, so it is not unlikely that the 288 and 329 samples might contain a valid T signal at the 16223 position within the single cell. If this is case, the samples might not serve as valid controls, but would be an important data point to show heteroplasmy at the cellular level where no heteroplasmy was seen in the gross tissue. The original scope of the research allowed for this possibility to occur within the cells, and if it is interpreted as a valid type, it is important to note.

Discussion of SNaPshot™ Results of Sample 314 (16223) Cell Analysis

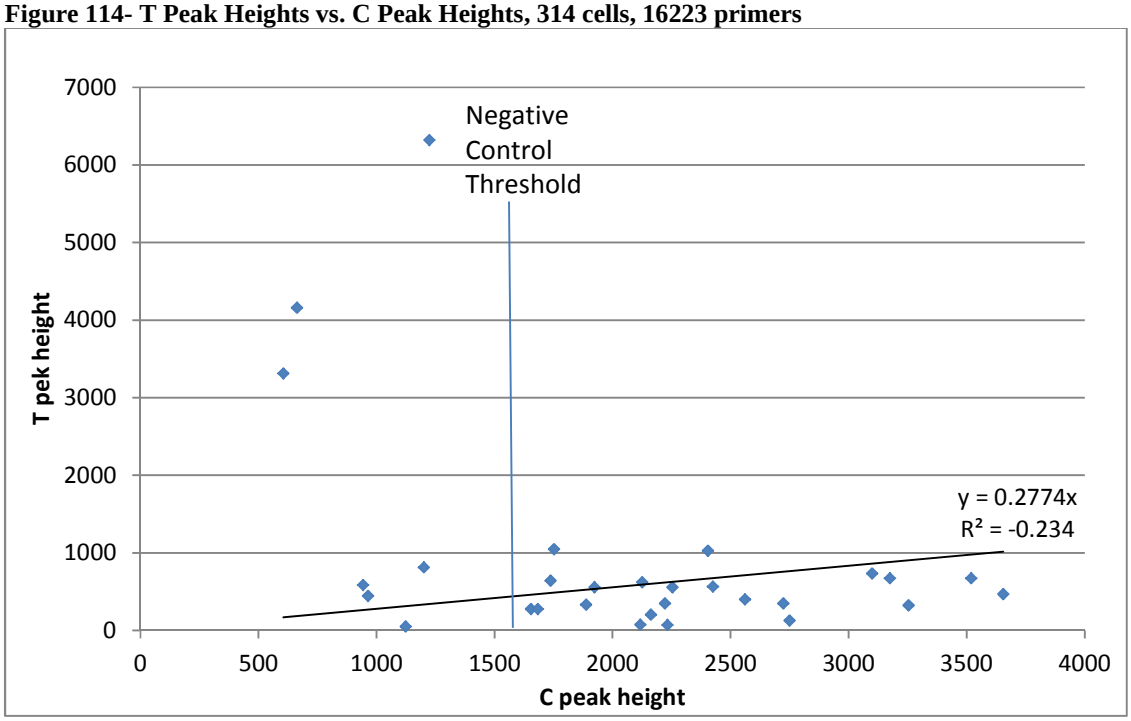
When the data from the forty-six viable 314 cells and ten valid control cells was analyzed, it was seen that a majority of the cells displayed some form of C/T heteroplasmy (refer to Table 29 in the results section). The 314 cells, along with the 288 and 329 control cells displayed either solitary C peaks or C/T peaks in the data. One sample cell (32-314) and two control cells (42-288c and 41-329c) also displayed a G peak as well. Out of the 314 cells, seven samples displayed a below-threshold C peak but did contain valid T peak data, so the C peak for this sample was inconclusive and the T peaks were valid. Out of the remaining thirty-nine samples, sixteen samples displayed only solitary C peaks at ~30 bp, while twenty-two others displayed C/T heteroplasmic bases with the C at ~30 bp and the T at ~33 bp. The remaining single cell sample displayed the G/C heteroplasmy, with the C base at ~30 bp, and the G base at ~28 bp. Typically the C peak heights ranged from 400 to 4600 RFU, while the T peaks, when present, ranged in peak heights from 40 to over 6300 RFU. The C peaks were the major peaks in the cells, when either the C or T was present (Figure 113). It was observed that as the C peak heights increased there appeared to be no effect on the addition or presence of the T peaks in the heteroplasmic cells.

Figure 113- Peak Heights for 314 cells, 16223 primers
(Samples sorted by C peak-height size)



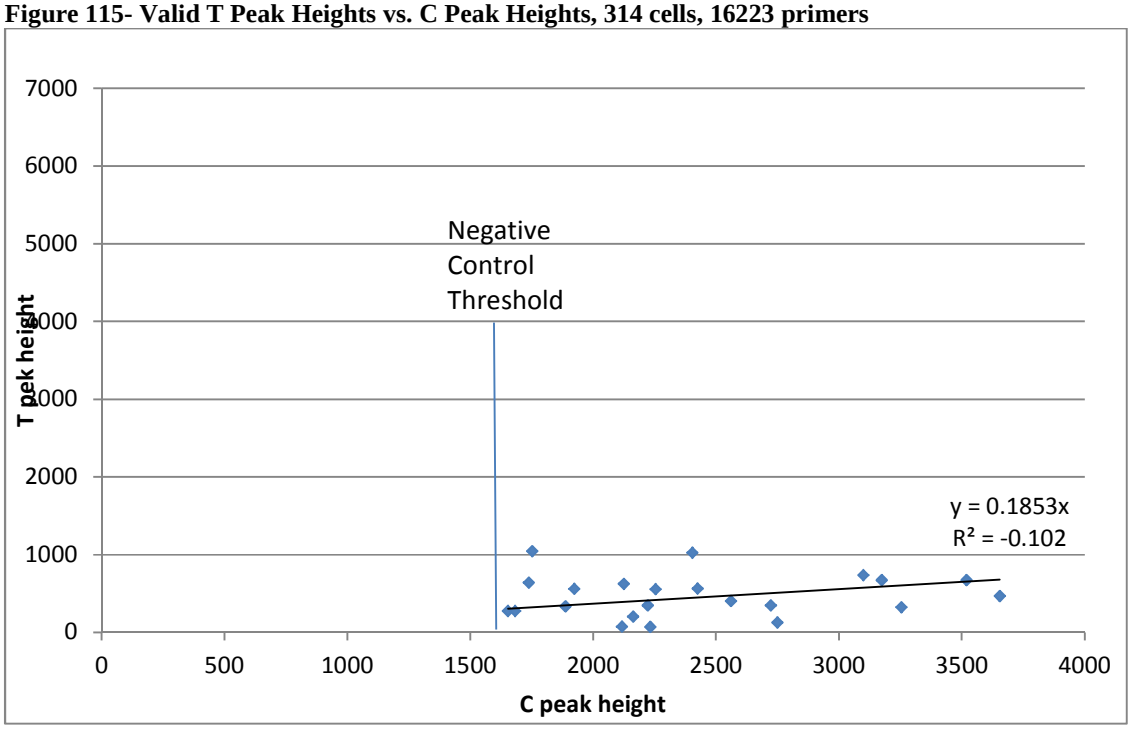
In order to determine if the heteroplasmic data for the 314 cells was due to amplification and/or electrophoresis artifacts or not, the T and C peak heights from the heteroplasmic cells were plotted, as shown in Figure 114. With only the negative control threshold to apply to the C peak heights, it is seen that when the heteroplasmic cells were plotted, a majority of the data points were valid. However, the best-fit R^2 value was negative (-0.23) which would indicate that the data points are non-linear. This reinforces the notion that the heteroplasmic cells are displaying unique and independent base signals.

Figure 114- T Peak Heights vs. C Peak Heights, 314 cells, 16223 primers



When the invalid cells were removed from the plot, the best-fit R^2 value did not improve, and remained negative, reinforcing the notion that the T and C signals from the heteroplasmic cells were unique and independent signals (Figure 115).

Figure 115- Valid T Peak Heights vs. C Peak Heights, 314 cells, 16223 primers



For the control cells, all ten control samples displayed C peaks at ~30 bp, ranging from 1700 to over 3100 RFU. Five of the control cells also displayed T peaks at ~33 bp, ranging from 300 to over 800 RFU. Two control samples also displayed G bases at ~28 bp, at 250 and 1217 RFU, respectively. For the 42-288c sample, all three heteroplasmic bases were present, reflecting a G/C/T heteroplasmy. For the 41-329c sample, only the G/C bases were present. The remaining four control samples displayed a solitary C peak.

Since twenty-two of the cells were seen to contain some form of C/T heteroplasmy (plus the one containing G/C heteroplasmy), it could be determined that one of three scenarios was occurring within the 314 cell samples. The first scenario was that the viable cells were reflecting a true heteroplasmic C/T or G/C bases and homoplastic C bases within the 16223 position of the mtDNA genomes. The second scenario was that the presence of the T and G bases at the 16223 position was so minor within the mitochondrial DNA of the individual cells that it was being overwhelmed by the major C base signals, particularly in the samples which only displayed a

solitary C base. The third possibility was the issue of the non-addition of the ddNTP bases in the extension primers due to the modified extension primers and/or concentration of the mtDNA within the single cell sample.

When the original SNaPshot™ result for the gross tissue sample 314 was re-examined, it was seen that the T/C heteroplasmy ratio for the 16223 base positions was not 1:1. Instead, within the gross tissue sample, the T peak was seen at 1924 RFU, where the C peak appeared at 119 RFU, and no G peak was detected at all. If it was assumed that the population of heteroplasmic mitochondrial DNA was not heterogeneous, and was split among unique cells (scenario #1 above), then twenty-two cells that contained this unequal partitioning of the heteroplasmic mtDNA were found. If it was assumed instead that a heterogeneous population of the heteroplasmic mitochondrial DNA exists within the cells (scenario #2 above), then the T bases occurred over sixteen times more frequently than the minor C bases. The third scenario concerning the non-addition of the ddNTP due to the poly-thymine modified extension primer and mtDNA concentration has already been identified as a source of error in the study, and was seen to be the most likely reason for the lack of A peaks in the data, especially for the 16519 extension primer with the longest poly-thymine tail. However, in the case of the 16223 primers, they were modified with only one thymine residue at the 5′ position, and no A peaks were detected in the gross tissue or single cells for 314. However, the effect of the non-optimal mtDNA concentration in the single cells cannot be ignored. Apart from the known issue with the ddNTP non-addition, the second scenario was more likely, due to the previous assumption that by sampling fifty cells from a known heteroplasmic tissue at least one cell should show the base change, based on known rates of heteroplasmy in such tissues. Assuming the heterogeneous spread of the mitochondrial DNA was equal within the cells and tissue sample, this could have

contributed to the results of the gross tissue heteroplasmy and the C/T heteroplasmic individual cells within the 314 sample set. This issue ultimately focuses on the mtDNA concentration in the samples. Since thousands of cells contributed mtDNA in the gross tissue sample, the concentration of the mtDNA could be measured on the Agilent 2100 Bioanalyzer, and the culmination of the signals of both bases was above the baseline but was not reflective of the unequal nature of the C/T signal. At the same time, the C/T concentrations in the gross tissue were possibly in such high concentrations as to overwhelm a minor G signal also present in the data. In an individual cell, the mtDNA concentration was not detectable on the Agilent 2100 Bioanalyzer, but the overall concentration of the C/T signal was lower, allowing the T base to be detected and not overwhelmed by the C signal. However, this was seen as a reverse of the gross tissue results, in which the T peaks were major in the gross tissue, and the C peak was minor. This might be indicative of the concentration of the mtDNA in the single cells compared to the gross tissue. It is possible that the ratio of the C to T peaks is closer to 1:1 on average for the tissue, and that fluctuation in this ratio is seen at the cellular level, indicating that the spread of the heteroplasmic mtDNA may not be evenly distributed throughout the cells. This observation is reinforced in the 314 cells by the presence of the random G peaks in the cells and control samples. If these G peaks are real, and reflective of the heteroplasmic peaks in the cells, then it reinforces the hypothesis that the spread of the heteroplasmic samples was heterogeneous, and these minor peaks become detectable at the cellular level when not overwhelmed by the major peaks of the sample. In the case of the 314 cells, these results would indicate that heteroplasmy in the single cells was most likely present, and was spread among the cells in the heteroplasmic tissue sample.

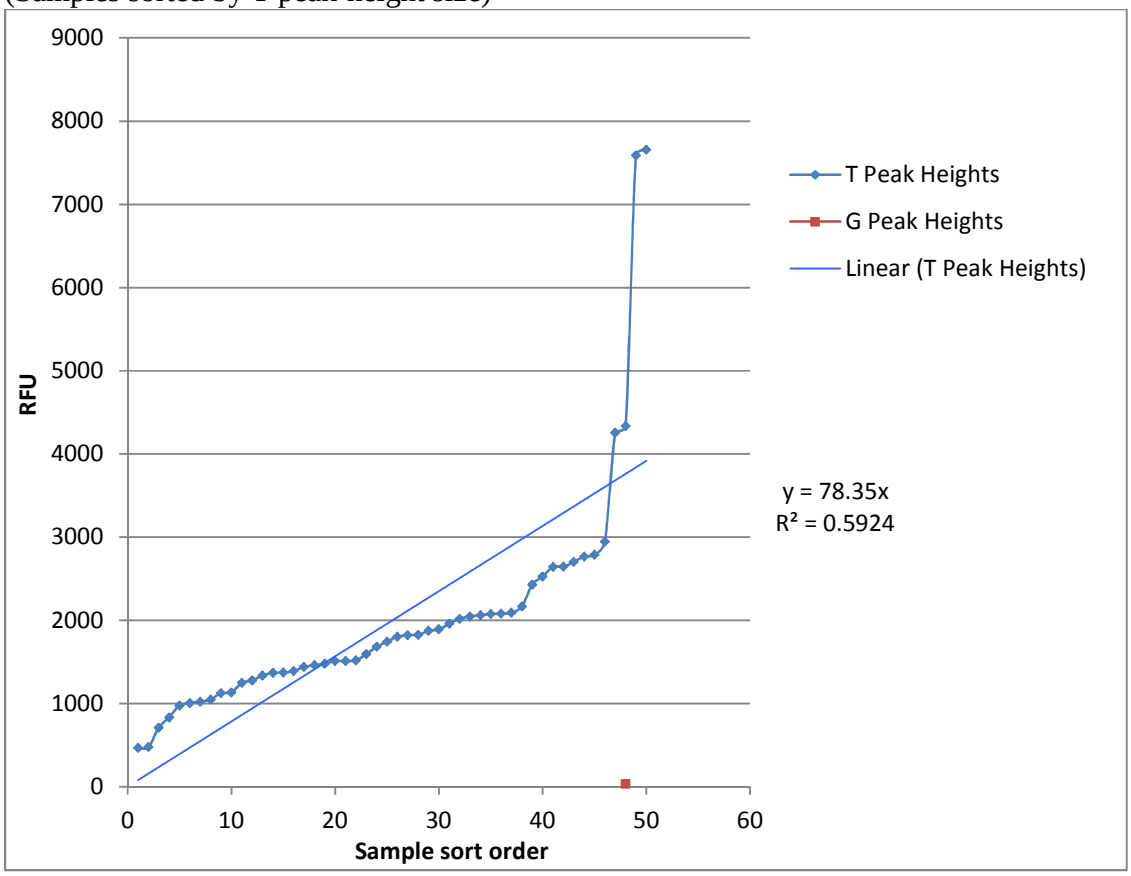
When examining the 288 and 329 control cells that were run with the 314 cells, five instances of C/T heteroplasmy were seen. As seen in Table 29, (pgs. 175-176), these control samples all displayed the presence of a C base at the ~30 bp size. When the T base was present, it was at the ~33 bp size, and the G base, when present in two samples, was at the ~28 bp size. In the gross tissue analysis of these samples, the 288 and 329 samples were seen to be C bases at the 16223 position for each of the samples. This calls into question the validity of the controls, and the specificity of the 16223 extension primers. The presence of the T and G peaks in the control samples opens the question if these peaks are truly present in the control cells, or if it is a random addition to the extension primer. If the T or G peak in the control cells is a true base, it means that it was at a high enough concentration in the sample cells to be detected and signaled by the 16223 extension primers. If that is the case, then why was the T or G peak not seen in the gross tissue analysis of the control tissues for 288 and 329? If the T or G peak was the result of random, non-specific base addition to the 16223 extension primer in the presence of low mtDNA concentration, then a bigger problem exists. If this T or G peak was random, non-specific addition, then the 314 single cell samples would be affected as well, resulting in totally inconclusive data for the 314 cell set due to the untrustworthiness of the results. However, the T peak was expected in the 314 cell set as part of the possible C/T heteroplasmy based on the gross tissue, so the presence of the T peak in the 293 cells was not believed to be non-specific addition, but instead true data. The G peak, however, was not seen in the original gross tissue heteroplasmy, but was present in one of the single cell samples. It would also be unlikely for the T or G base displaying in the control samples to be any type of artifact, since a fluorescently-labeled ddNTP base must be adding to the extension primer for the samples to be detected, and this can only occur when the template strand contains the corresponding A or C base.

As to the issue of the validity of the control samples, the T bases that were detected above the negative control threshold were of the correct bp size for the 16223 extension primer. Since the valid control cells are uniformly displaying the C base at 16223 along with the T base in five of the samples, while the gross tissues for the controls did not show any indication of any base other than the C at 16223, the implications of the controls samples are open to interpretation. However, the original project design of targeting the hotspots in the HVI section of the mtDNA genome was to screen for higher instances of heteroplasmy, so it is not unlikely that the 288 and 329 samples might contain a valid T or G signal at the 16223 position within the single cell. If this is case, the samples might not serve as valid controls, but would be an important data point to show heteroplasmy at the cellular level where no heteroplasmy was seen in the gross tissue. The original scope of the research allowed for this possibility to occur within the cells, and if it is interpreted as a valid type, it is important to note.

Discussion of SNaPshot™ Results of Sample 289 (16324) Cell Analysis

When the data from the forty-eight viable 289 cells and ten control cells was analyzed, one instance of G/T heteroplasmy was present in the 289 cells; sample 19-289 (refer to Table 31 in the results section). The remaining 289 cells, along with the 288 and 329 control cells only displayed solitary T peaks in the data. Out of the valid forty-eight samples, the T peak displayed at ~34 bp, and the single G peak appeared at ~31 bp. The T peak heights ranged from 400 to over 4300 RFU. The solitary G peak height was 30 RFU. The T peak height was greater than the G peak, so the G peak was minor (Figure 116).

Figure 116- Peak Heights of 289 cells, 16324 primers
(Samples sorted by T peak-height size)



Since only one of the cells was seen to contain a G peaks in the ~31 bp region, it could be determined that one of three scenarios was occurring within these samples. The first scenario was that the cells were reflecting true homoplastic T bases and a single instance of heteroplasmic G/T bases within the 16324 position of the mtDNA genomes. The second scenario was that the presence of the G base at the 16324 position was so minor within the mitochondrial DNA of the individual cells that it was being overwhelmed by the major T base signals. The third possibility was the issue of the non-addition of the ddNTP bases in the extension primers due to the modified 5′ poly-thymine tails and the mtDNA input concentration.

When the original SNaPshot™ result for the gross tissue sample 289 was re-examined, it was seen that the G/T heteroplasmy ratio for the 16324 base positions was not 1:1. Instead,

within the gross tissue sample, the T peak was seen around 2408 RFU, where the G peak appeared at 103 RFU. If it was assumed that the population of heteroplasmic mitochondrial DNA was not heterogeneous, and was split among unique cells (scenario #1 above), then within the viable 289 sample cells, a single cell containing the G base was found, but only once out of the fifty supposedly heteroplasmic cells. If it was assumed instead that a heterogeneous population of the heteroplasmic mitochondrial DNA exists within the cells (scenario #2 above), then the T bases are occurring over twenty-three times more frequently than the minor G bases. The third scenario concerning the non-addition of the ddNTP due to the poly-thymine modified extension primer in combination with the mtDNA input concentration has already been identified as a source of error in the study, and was seen to be the most likely reason for the lack of A peaks in the data, especially for the 16519 extension primer with the longest poly-thymine tail. Apart from the known issue with the extension primer and the ddNTP addition, the second scenario was more likely, due to the previous assumption that by sampling fifty cells from a known heteroplasmic tissue at least one cell should show the above-threshold heteroplasmic bases, based on known rates of heteroplasmy in such tissues, which is reflected in the results for the 289 cells. Assuming the heterogeneous spread of the mitochondrial DNA was equal within the cells and tissue sample, this could have contributed to the results of the gross tissue heteroplasmy and the apparent individual homoplastic T cells within the 289 sample set. This issue ultimately focuses on the mtDNA concentration in the samples. Since thousands of cells contributed mtDNA in the gross tissue sample, the concentration of the mtDNA could be measured on the Agilent 2100 Bioanalyzer, and the culmination of the signals of both bases was above the baseline but was not reflective of the unequal nature of the G/T signal. At the same time, the G concentrations in the gross tissue were in high enough concentrations as to be

detected within the data. In a single individual cell, the mtDNA concentration was not detectable on the Agilent 2100 Bioanalyzer, and the prevalence of the heteroplasmic G base was only at a high enough concentration in a single sample to be detected above the major signal of the T base within the single cells. Combined with the issue of the non-addition of the ddNTP due to the poly-thymine modified extension primers and mtDNA concentration, the previously minor G base was only at a high enough concentration in one cell to be probed and successfully detected, leading to the single heteroplasmic cell result. Therefore, the mtDNA concentration along with the non-addition of the ddNTP bases due to the poly-thymine modified extension primers most likely led to results that were seen for the 289 cells.

Since only one of the 289 cells exhibited heteroplasmy, the data from the peak heights of each signal was not plotted since the scatter plot would not be informative with only one data point.

When examining the 288 and 329 control cells that were run along with the 289 cells, no instances of heteroplasmy were seen. As seen in Table 31, (pgs. 179-180), these samples only displayed the presence of the T base at the ~34 bp size. This correctly reflected the T base seen in the 16324 position for the gross tissues of the control samples, and thereby demonstrates the validity of the control samples.

To address the specificity of the primers, it should be noted that the HL60 positive control sample displayed the appropriate T peak at ~34 bp with a peak height of 8433 RFU, so the fact that an T base can be added and detected in the 16324 extension primers is known to properly occur. This reinforces the notion that the ddNTP addition is tied to the mtDNA concentration of the sample, since the HL60 positive control consistently showed a detectable Agilent 2100 Bioanalyzer value for the mtDNA concentration. Also, it would be unlikely for the

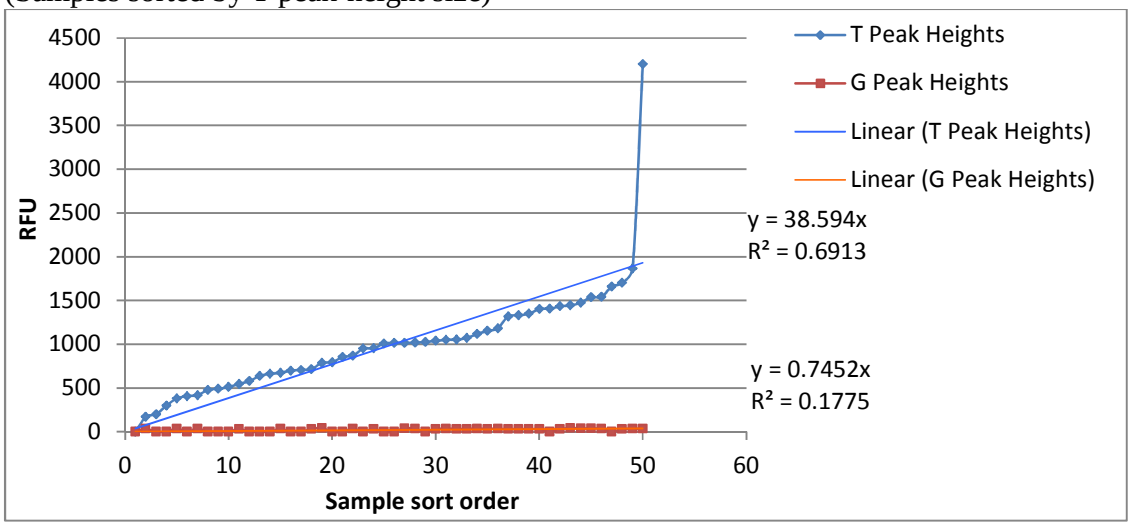
T bases displaying in the control samples to be any type of artifact, since a fluorescently-labeled ddNTP base must be adding to the extension primer for the samples to be detected, and this can only be occurring when the template strand contains the corresponding A base.

As to the issue of the validity of the control samples, the T bases that were detected were above the negative control threshold, and are of the correct bp size for the 16324 extension primer. Since the control cells were uniformly displaying the T base at 16324, the controls passed.

Discussion of SNaPshot™ Results of Sample 349 (16324) Cell Analysis

When the data from the twenty-two viable 349 cells and six valid control cells was analyzed, no instances of A/T heteroplasmy were present in any of the 349 cells (refer to Table 33 in the results section). The 349 cells, along with the 288 and 329 control cells only displayed either solitary T peaks or G/T peaks in the data. Out of the 349 cells, seven samples displayed a below-threshold T peak but did contain valid G peak data, so the T peak for these samples was inconclusive and the G peak was valid. Seven other samples displayed a below-threshold G peak but did contain valid T peak data, so the G peak for these samples was inconclusive and the T peak was valid. Two other samples (20-349 and 45-349) contained only valid T peaks, with no G peak data present. Out of the remaining six valid samples, all displayed a valid mixture of T peaks at ~34 bp and G peaks at ~31 bp. Typically the T peak height ranged from 200 to 4200 RFU, while the G peaks, when present, ranged in peak heights from 30 to 43 RFU, so the T peaks were seen as the major signal (Figure 117). No A peaks were seen in any of the 349 cells or control cells whatsoever.

Figure 117- Peak Heights for 349 cells, 16324 primers
(Samples sorted by T peak-height size)



Since none of the cells contained any A peaks in the ~33 bp region, it could be determined that one of three scenarios was occurring within these samples. The first scenario was that the cells were reflecting a true heteroplasmic G/T bases within the 16324 position of the mtDNA genomes. The second scenario was that the presence of the A base at the 16324 position was so minor within the mitochondrial DNA of the individual cells that it was being overwhelmed by the major G and T base signals. The third possibility was the issue of the non-addition of the ddATP bases in the extension primers due to the modified 5′ poly-thymine tails in combination with the mtDNA input concentration.

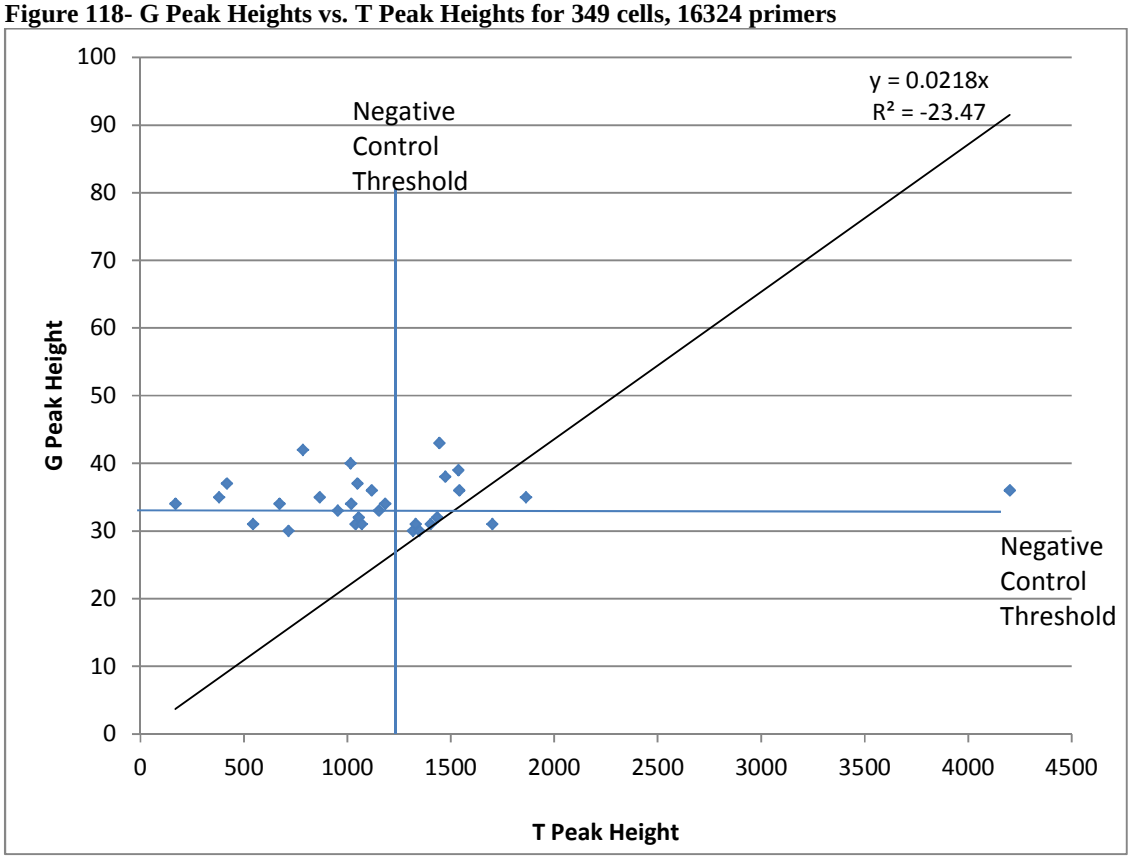
When the original SNaPshot™ result for the gross tissue sample 349 was re-examined, it was seen that the A/T heteroplasmy ratio for the 16324 base positions was not 1:1. Instead, within the gross tissue sample, the T peak was seen at 2205 RFU, where the A peak appeared at 101 RFU, and no G peak was detected at all. If it was assumed that the population of heteroplasmic mitochondrial DNA is not heterogeneous, and was split among unique cells (scenario #1 above), then within the viable 349 sample cells, a single cell containing the A base was not found. If it is assumed instead that a heterogeneous population of the heteroplasmic mitochondrial DNA exists within the cells (scenario #2 above), then the T bases are occurring over twenty-one times more frequently than the minor A bases. The third scenario concerning the non-addition of the ddATP due to the poly-thymine modified extension primer and mtDNA concentration has already been identified as a source of error in the study, and is seen to be the most likely reason for the lack of A peaks in the data, especially for the 16519 extension primer with the longest poly-thymine tail. In the case of the 16324 primers, they were also modified, and contain six thymine residues at the 5′ position. Apart from the known issue with the extension primer and the ddATP addition, the second scenario was more likely, due to the

previous assumption that by sampling fifty cells from a known heteroplasmic tissue at least one cell should show the base change, based on known rates of heteroplasmy in such tissues. Assuming the heterogeneity of the mitochondrial DNA was equal within the cells and tissue sample, this could have contributed to the results of the gross tissue heteroplasmy and the divergent G/T heteroplasmic individual cells within the 349 sample set. This issue ultimately focuses on the mtDNA concentration in the samples. Since thousands of cells contributed mtDNA in the gross tissue sample, the concentration of the mtDNA could be measured on the Agilent 2100 Bioanalyzer, and the culmination of the signals was above the baseline but was not reflective of the unequal nature of the A/T signal. At the same time, the A/T concentrations in the gross tissue were in such high concentrations as to overwhelm a minor G signal also present in the data. In a single individual cell, the mtDNA concentration was not detectable on the Agilent 2100 Bioanalyzer, and the prevalence of any heteroplasmic A base was never at a high enough concentration to be detected above the major signal of the T base within the single cells. This is somewhat verified based on the graph of the G peak heights in relation to the T peak heights (Figure 117). As the T peak heights increase, there appears to be no effect on the presence of the G peaks, so there is no correlation between these two peaks. It was seen throughout the data that the addition of the ddATP bases was problematic when the mtDNA concentration was not optimal. The possibility exists that the lack of the A bases caused the previously undetected G base to be successfully detected in the single cells. Therefore, the mtDNA concentration of the single cells most likely led to results that were seen for the 349 cells.

In order to determine if the heteroplasmic data for the 349 cells was due to amplification and/or electrophoresis artifacts or not, the G and T peak heights from the heteroplasmic cells

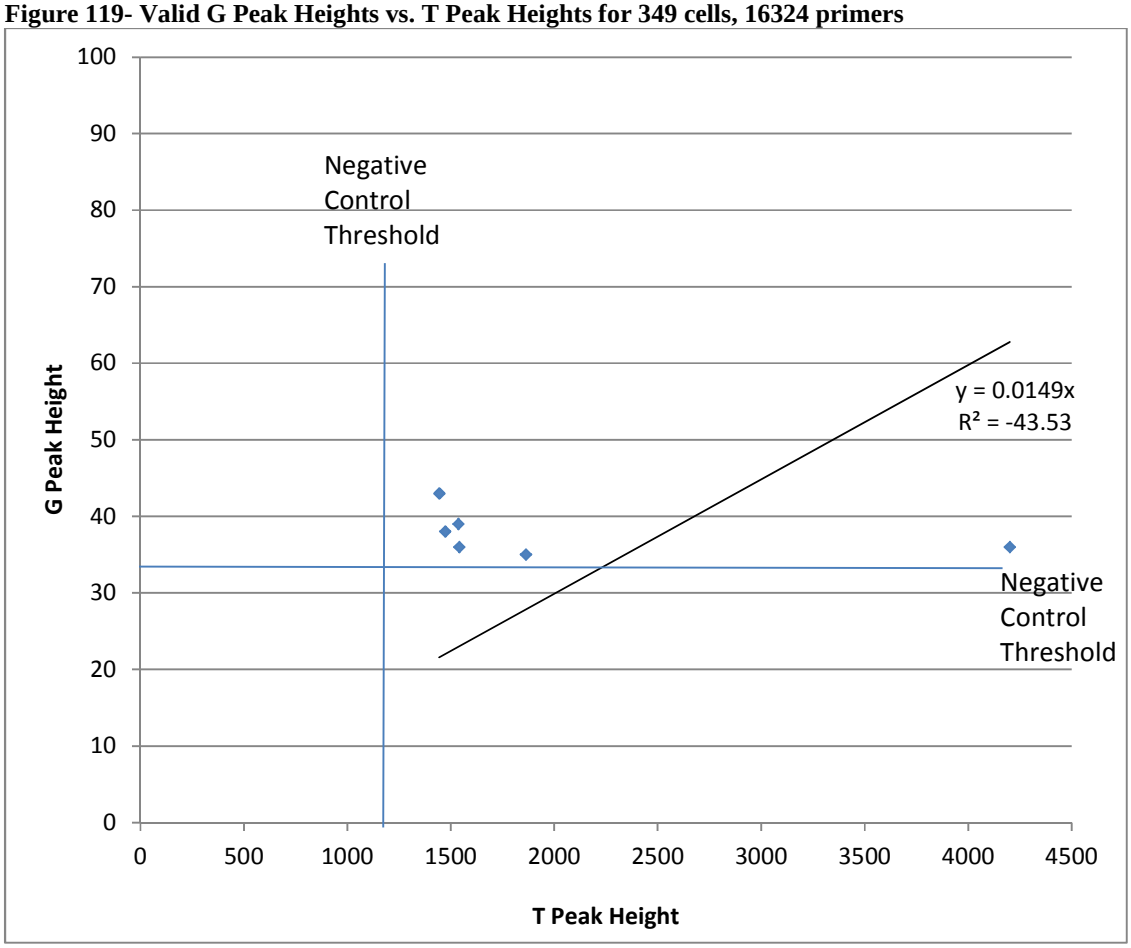
were plotted, as shown in Figure 118. Based on the negative control thresholds for both the G and T peaks, only six heteroplasmic cells were seen to be valid, but exhibited a wide range of variability. The resulting best-fit R^2 for all of the heteroplasmic cells was shown to be -23.4, indicating that the data points are highly non-linear and therefore the heteroplasmic G and T peak signals are probably unique and independent peak signals.

Figure 118- G Peak Heights vs. T Peak Heights for 349 cells, 16324 primers



When the invalid cells were removed from the plot, the best-fit R^2 value did not improve, and remained negative, reinforcing the notion that the G and T signals from the heteroplasmic cells were unique and independent signals (Figure 119).

Figure 119- Valid G Peak Heights vs. T Peak Heights for 349 cells, 16324 primers



When examining the 288 and 329 control cells that were run along with the 349 cells, two instances of heteroplasmy were seen; sample 22-329c and 25-329c displayed a G/T heteroplasmy. The remaining four valid control cells all displayed a valid T peak, with two of these controls displaying a solitary T peak, and the remaining two displaying a valid T peak as well as an invalid G peak. As seen in Table 33, (pgs. 183-184), these control samples all showed the presence of a G base at the ~31 bp size and the T base at ~34 bp size. In the gross tissue analysis of these samples, the 288 and 329 samples were seen to be T bases at the 16324 position

for each of the samples. This calls into question the validity of the controls, and the specificity of the 16324 extension primers. The presence of the G peak in the control samples opens the question whether the G peak was truly present in the control cells, or if it was a random addition to the extension primer. If the G peak in the control cells was a true base, it means that it was at a high enough concentration in the sample cells to be detected and signaled by the 16324 extension primers. If that was the case, then why was the G peak not seen in the gross tissue analysis of the control tissues for 288 and 329? If the G peak was the result of random, non-specific base addition to the 16324 extension primer in the presence of low mtDNA concentration, then a bigger problem exists. If this G peak was random, non-specific addition, then the 349 single cell samples would be affected as well, resulting in totally inconclusive data for the 349 cell set due to the untrustworthiness of the results. However, the T peak was expected in the 349 cell set as part of the possible A/T heteroplasmy based on the gross tissue, so the presence of the T peak in the 349 cells is not believed to be non-specific addition, but instead true data. Also, it would be unlikely for the G base displaying in the control sample to be any type of artifact, since a fluorescently-labeled ddNTP base must be adding to the extension primer for the samples to be detected, and this can only be occurring when the template strand contains the corresponding C base.

To address the specificity of the primers, it should be noted that while no A bases were present in the 349 cells, the control cells, or any of the positive or negative controls, the 16324 primers did display the A base addition in the gross tissue sample for 349, so the fact that an A base can be added and detected in the 16324 extension primers is known to properly occur. This reinforces the hypothesis that the ddATP addition is tied to both the mtDNA concentration of the sample as well as the presence of the modified poly-thymine tails on the extension primers, since

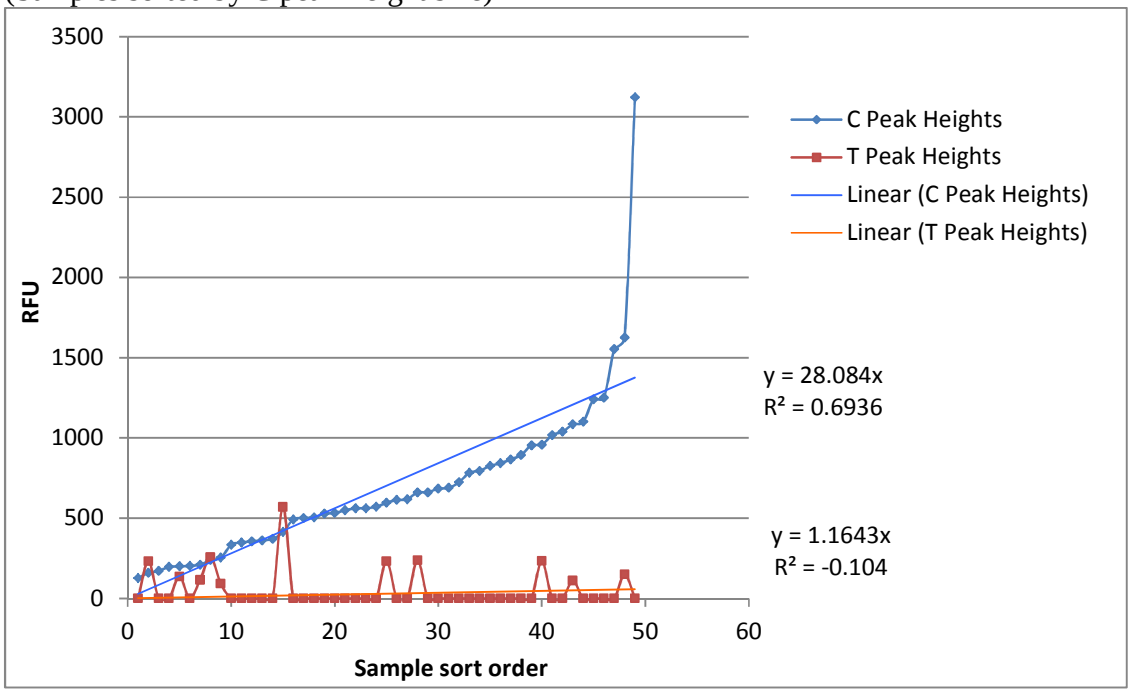
the 349 gross tissue showed a viable A base signal and the 16324 extension primers contained a 5′ modification to interfere with the A base addition. In the case of the 349 cells, the limiting factor for the A base addition might have been more affected by the available mtDNA concentration in the sample, causing the A bases not to be detected.

As to the issue of the validity of the control samples, the G bases that were detected above the negative control threshold were of the correct bp size for the 16324 extension primer. Since the valid control cells uniformly displayed the G base at 16324 along with the G/T based on the 349 cells, while the gross tissues for the controls did not show any indication of any base other than the T at 16324, the implications of the controls samples are open to interpretation. However, the original design of targeting the hotspots in the HVI section of the mtDNA genome was to screen for higher instances of heteroplasmy, it is not unlikely that the control samples might contain a valid G signal at the 16324 position within the single cell. If this is case, that solitary sample might not serve as a valid control, but would be an important data point to show heteroplasmy at the cellular level where no heteroplasmy was seen in the gross tissue. The original scope of the research allowed for this possibility to occur within the cells, and if it is interpreted as a valid type, it is important to note.

Discussion of SNaPshot™ Results of Sample 339 (16390) Cell Analysis

When the data from the forty-three viable 339 cells and eight valid control cells was analyzed, five instances of C/T heteroplasmy were present in the 339 cells (4-339, 10-339, 26-339, 31-339, and 42-339) (refer to Table 35 in the results section). The remaining 339 cells, along with the 288 and 329 control cells either displayed solitary C peaks in the data, or valid T peaks alongside invalid C peaks. Out of the valid forty-three samples, the C peak displayed at ~36 bp, and the T peak appeared at ~39 bp. The C peak heights ranged from 120 to over 3100 RFU. The T peak height ranged from 90 to over 560 RFU. The C peak height was greater than the T peak, so the T peaks were minor (Figure 120).

Figure 120- Peak Heights for 339 cells, 16390 primers
(Samples sorted by C peak-height size)



Since only five of the cells were seen to contain heteroplasmic C/T peaks, it could be determined that one of three scenarios was occurring within these samples. The first scenario was that the cells were reflecting true heteroplasmic C/T bases and homoplastic C bases within

the 16390 position of the mtDNA genomes. The second scenario was that the presence of the T base at the 16390 position was so minor within the mitochondrial DNA of the individual cells that it was being overwhelmed by the major C base signals. The third possibility was the issue of the non-addition of the ddNTP bases to the extension primers due to the modified 5′ poly-thymine tails and mtDNA input concentration.

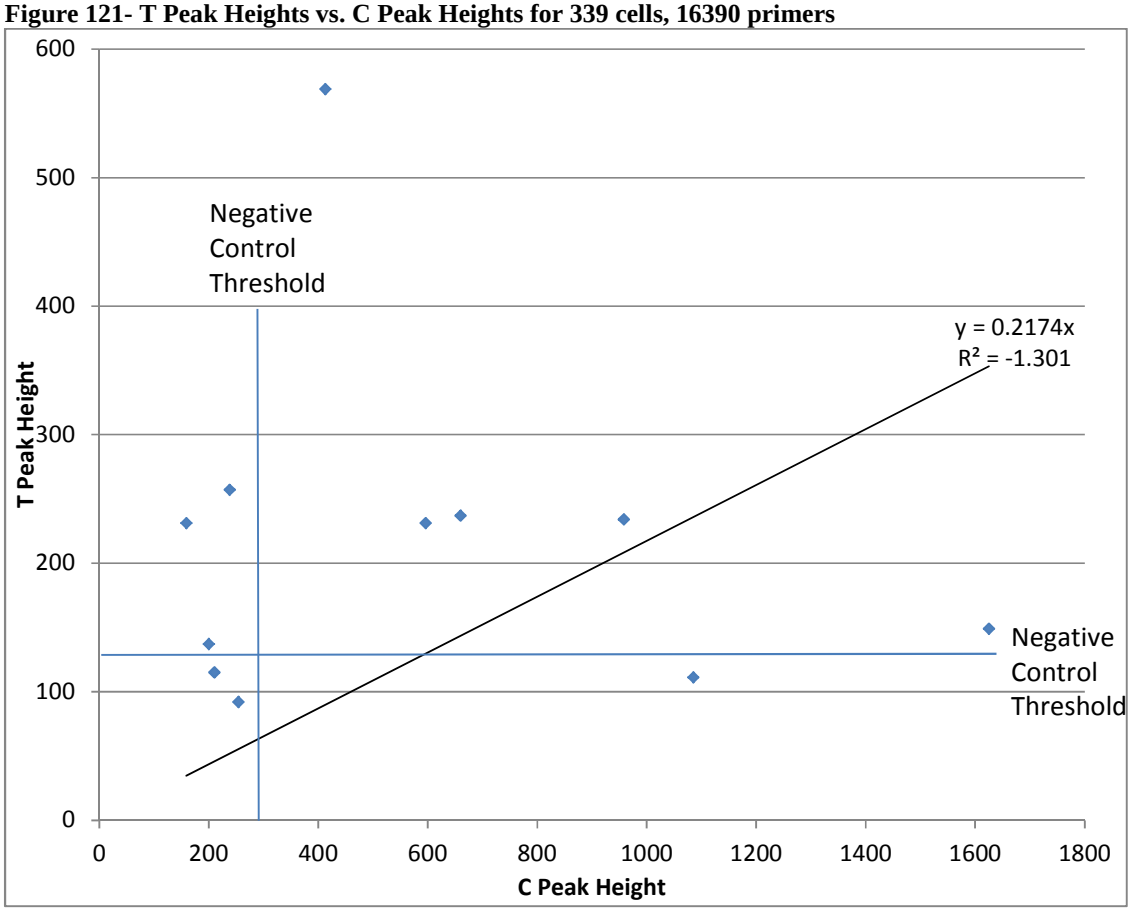
When the original SNaPshot™ result for the gross tissue sample 339 was re-examined, it was seen that the C/T heteroplasmy ratio for the 16390 base positions was not 1:1. Instead, within the gross tissue sample, the T peak was seen around 1618 RFU, where the C peak appeared at 96 RFU. If it was assumed that the population of heteroplasmic mitochondrial DNA is not heterogeneous, and was split among unique cells (scenario #1 above), then within the viable 339 sample cells, cells containing the C/T heteroplasmic bases were found, but only for five out of the fifty supposedly heteroplasmic cells. If it was assumed instead that a heterogeneous population of the heteroplasmic mitochondrial DNA exists within the cells (scenario #2 above), then the T bases are occurring over sixteen times more frequently than the minor C bases. The third scenario concerning the non-addition of the ddNTP due to the poly-thymine modified extension primers and mtDNA concentration has already been identified as a source of error in the study, and was seen to be the most likely reason for the lack of A peaks in the data, especially for the 16519 extension primer with the longest poly-thymine tail. Apart from the known issue with the extension primer and the ddNTP addition, the second scenario was more likely, due to the previous assumption that by sampling fifty cells from a known heteroplasmic tissue at least one cell should show the above-threshold base change, based on known rates of heteroplasmy in such tissues, which is reflected in the results for the 339 cells. Assuming the heterogeneous spread of the mitochondrial DNA was equal within the cells and

tissue sample, this could have contributed to the results of the gross tissue heteroplasmy and the apparent individual homoplastic C cells within the 339 sample set. This issue ultimately focuses on the mtDNA concentration in the samples. Since thousands of cells contributed mtDNA in the gross tissue sample, the concentration of the mtDNA could be measured on the Agilent 2100 Bioanalyzer, and the culmination of the signals of both bases was above the baseline but was not reflective of the unequal nature of the C/T signal. At the same time, the C base signal in the gross tissue was possibly in high enough concentration as to be detected and present in the data. In a single individual cell, the mtDNA concentration was not detectable on the Agilent 2100 Bioanalyzer, and the prevalence of any heteroplasmic T base was never at a high enough concentration to be detected above the major signal of the C base within the single cells. Combined with the issue of the non-addition of the ddNTP in the presence of the poly-thymine modified extension primers, the minor T base was only at a high enough concentration in five cells to be probed and successfully detected, leading to the heteroplasmic cell results. There did not appear to be a correlation between the presence of the C and T peaks, indicating that the detection of the T peaks was not limited by the presence of a higher signal from the C peaks (Figure 120). Therefore, the mtDNA concentration along with the non-addition of the ddNTP bases due to the poly-thymine modified extension primers most likely led to results that were seen for the 339 cells.

In order to determine if the heteroplasmic data for the 339 cells was due to amplification and/or electrophoresis artifacts or not, the T and C peak heights from the heteroplasmic cells were plotted, as shown in Figure 121. Due to the negative control thresholds from both the T and C peak heights, only five cells were seen to be valid out of the set of heteroplasmic cells.

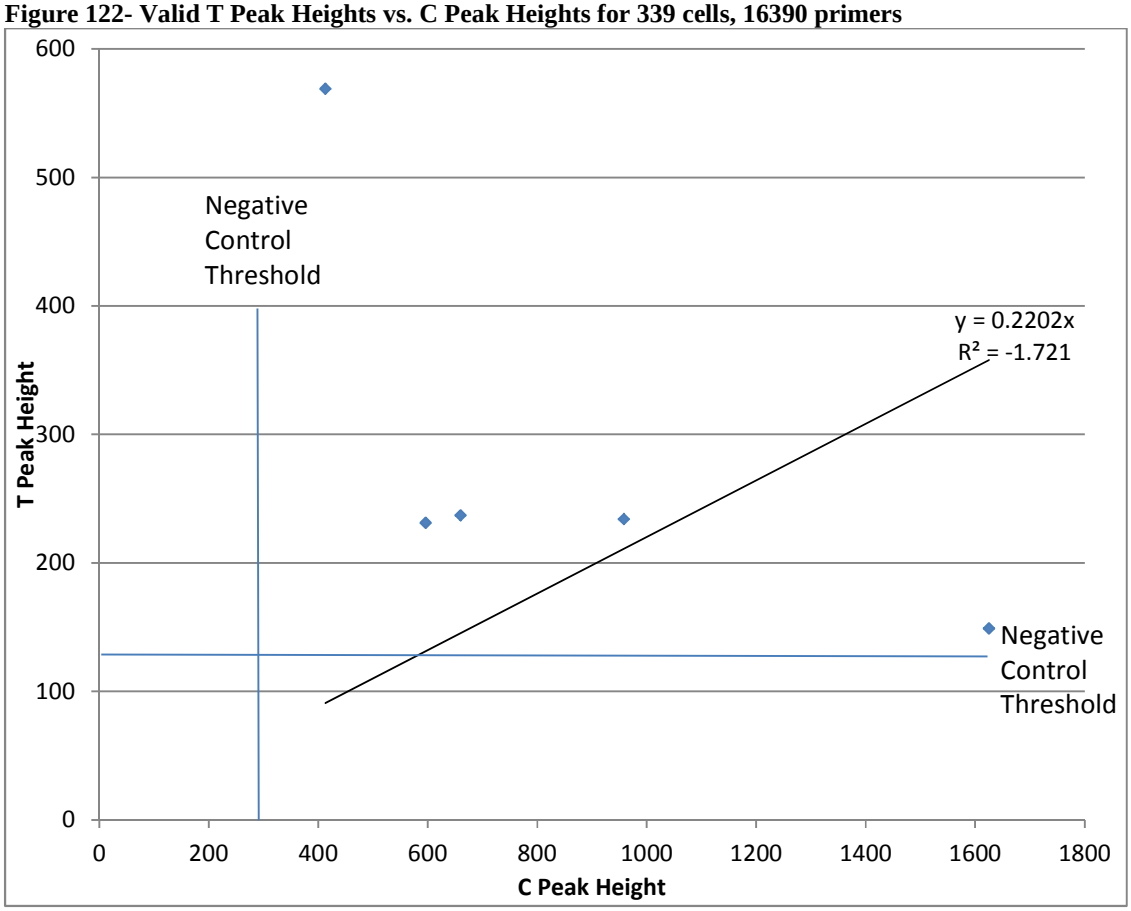
However, the overall best-fit R^2 value was negative (-1.30) indicating that the data is non-linear, and therefore may be unique and independent heteroplasmic base signals within the 339 cells.

Figure 121- T Peak Heights vs. C Peak Heights for 339 cells, 16390 primers



When the invalid cells were removed from the plot, the best-fit R^2 value did not improve, and remained negative, reinforcing the notion that the T and C signals from the heteroplasmic cells were unique and independent signals (Figure 122).

Figure 122- Valid T Peak Heights vs. C Peak Heights for 339 cells, 16390 primers



When examining the 288 and 329 control cells that were run along with the 339 cells, no instances of heteroplasmy were seen. As seen in Table 35, (pgs. 188-189), these samples all showed the presence of either a valid C base at the ~36 bp size, or a T base at the ~39 bp size. For two samples displaying a valid C base, invalid T bases were also present (51-288c and 54-329c). The single valid T base was in the presence of an invalid C base for control cell 53-288c. For the seven valid control cells displaying the C base, this correctly reflected the C base seen in the 16390 position for the gross tissues of the control samples, and thereby demonstrates the validity of the control samples. For the one control sample displaying the valid T, the results are open to interpretation, along with the additional control samples that displayed valid C bases with invalid T bases.

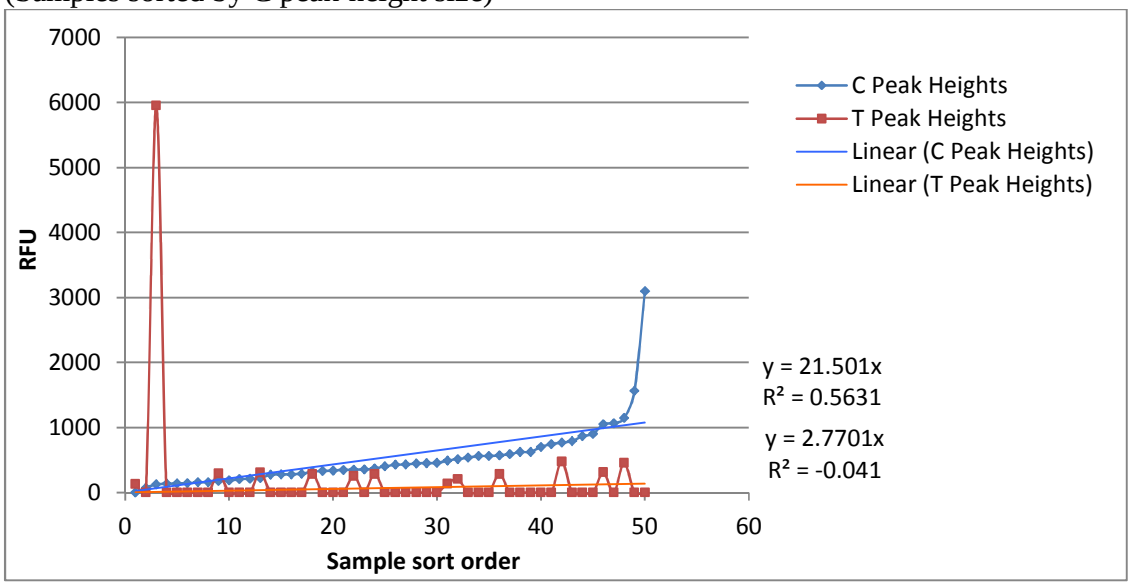
To address the specificity of the primers, it should be noted that the HL60 positive control sample displayed the appropriate C peak at ~36 bp with a peak height of 4779 RFU, so the fact that an C base can be added and detected in the 16390 extension primers is known to properly occur. This reinforces the notion that the ddNTP addition is tied to the mtDNA concentration of the sample, since the HL60 positive control consistently showed a detectable Agilent 2100 Bioanalyzer value for the mtDNA concentration. Also, it would be unlikely for the T bases displayed in the control samples to be any type of artifact, since a fluorescently-labeled ddNTP base must be adding to the extension primer for the samples to be detected, and this can only be occurring when the template strand contains the corresponding A base.

As to the issue of the validity of the control samples, the T bases that were detected were above the negative control threshold, and were of the correct bp size for the 16390 extension primer. Since the control cells were displaying the C base at 16390, as well as the solitary T base, the implications of the control samples are open to interpretation. However, the original design of targeting the hotspots in the HVI section of the mtDNA genome was to screen for higher instances of heteroplasmy, so it is not unlikely that the control samples might contain a valid T signal at the 16390 position within the single cell. If this is case, that solitary sample containing the valid T signal might not serve as a valid control, but would be an important data point to show heteroplasmy at the cellular level where no heteroplasmy was seen in the gross tissue. The original scope of the research allowed for this possibility to occur within the cells, and if it is interpreted as a valid type, it is important to note.

Discussion of SNaPshot™ Results of Sample 355 (16390) Cell Analysis

When the data from the twelve viable 355 cells and one viable control cell was analyzed, it was seen that no instances of C/T heteroplasmy were present in any of the 355 cells due to an extremely high amplification negative control threshold that invalidated all but one C peak (refer to Table 37 in the results section). The twelve valid samples either showed a solitary C peak (9-355) or valid T peak data against invalid C peak data. Only one sample displayed both invalid C and T peak data (24-355). Another solitary sample displayed no C peak data, and the T peak data was below the negative control threshold, and was therefore invalid (32-355). Out of the fifty cells, none were negative, indicating that had it not been for the high negative control threshold for the C peak data, the amplification would have been successful. For the C peaks, the peak heights were seen to range from 70 to over 3000 RFU, but due to the negative control threshold at 2677 RFU, all but the one sample over 3000 RFU were invalid for analysis. The T peaks ranged in height from 130 to over 5900 RFU, when the T peak was present. While the C peak heights are larger overall, it is unclear if there was a major or minor contributing peak within the 355 cells for the 16390 primers (Figure 123).

Figure 123- Peak Heights for 355 cells, 16390 primers
(Samples sorted by C peak-height size)



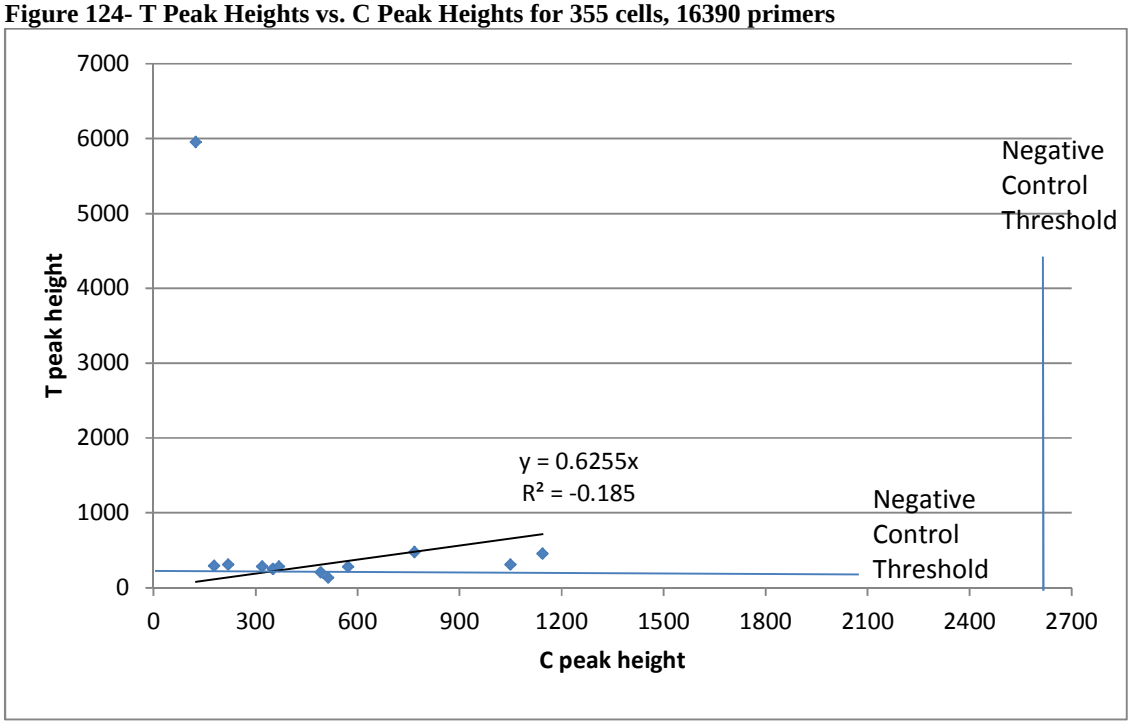
Since none of the cells were seen to contain any heteroplasmic peaks in the ~36 to ~39 bp region, it could be determined that one of three scenarios was occurring within these samples. The first scenario was that the cells were reflecting a true homoplastic C base within the 16390 position of the mtDNA genomes. The second scenario was that the presence of the T base at the 16390 position was so minor within the mitochondrial DNA of the individual cells that it was being overwhelmed by the larger C base signal. The third possibility was the issue of the non-addition of the ddNTP bases to the extension primers due to the modified 5′ poly-thymine tails and mtDNA input concentration.

When the original SNaPshot™ result for the gross tissue sample 355 was re-examined, it was seen that the C/T heteroplasmy ratio for the 16390 base positions was close to 1:1. Within the gross tissue sample, the C peak was seen around 60 RFU, where the T peak appeared at 53 RFU. If it was assumed that the population of heteroplasmic mitochondrial DNA was not heterogeneous, and was split among unique cells (scenario #1 above), then within the twelve viable sample cells, a single cell containing both the C and T base was not found. If it is

assumed instead that a heterogeneous population of the heteroplasmic mitochondrial DNA exists within the cells (scenario #2 above), then the C bases are occurring at the same frequency as the T bases, which was not seen in the data. The third scenario concerning the non-addition of the ddNTP due to the poly-thymine modified extension primer and mtDNA concentration has already been identified as a source of error in the study, and was seen to be the most likely reason for the lack of A peaks in the data, especially for the 16519 extension primer with the longest poly-thymine tail. Apart from the known issue with the extension primer and the ddNTP addition, the second scenario was more likely, due to the previous assumption that by sampling fifty cells from a known heteroplasmic tissue at least one cell should show the base change, based on known rates of heteroplasmy in such tissues. Assuming the heterogeneous spread of the mitochondrial DNA was equal within the cells and tissue sample, this could have contributed to the results of the gross tissue heteroplasmy and the apparent homoplastic individual cells within the 355 sample set. This issue ultimately focuses on the mtDNA concentration in the samples. Since thousands of cells contributed mtDNA in the gross tissue sample, the concentration of the mtDNA could be measured on the Agilent 2100 Bioanalyzer, and the culmination of the signals of both bases was above the baseline but was not reflective of the mixed nature of the C/T signal. In an individual cell, the mtDNA concentration was not detectable on the Agilent 2100 Bioanalyzer, and the prevalence of any heteroplasmic T base was never at a high enough concentration to be detected above the signal of the C base within the single cells. Therefore, the mtDNA concentration most likely led to results that were seen for the 355 cells. There appeared to be no correlation between the addition of the C bases and the presence of the T bases, so the possibility of the interference of the ddNTP addition from the poly-thymine modified extension primer is ruled out (Figure 123).

In order to determine if the heteroplasmic data for the 335 cells was due to amplification and/or electrophoresis artifacts or not, the T and C peak heights from the heteroplasmic cells were plotted, as shown in Figure 124. Due to the negative control thresholds for both the T and C peak heights, none of the heteroplasmic cells were valid due to the high negative control threshold on the C peak heights (2677 RFU). Although the best-fit R^2 value is negative, indicating unique and independent base signals, the fact that none of the cells are valid for analysis results in inconclusive data for the 355 cells with the 16390 primers. Since all of the heteroplasmic cells were invalid, there was no attempt to re-plot the data for only valid cells.

Figure 124- T Peak Heights vs. C Peak Heights for 355 cells, 16390 primers



When examining the 288 and 329 control cells that were run along with the 355 cell, no instances of heteroplasmy were seen. However, as seen in Table 37, (pgs. 192-193), nine of these ten samples displayed an invalid C base at the ~36 bp size. One of the control samples was negative (58-329c). The only sample to contain a valid peak was 56-288c, which displayed a valid T peak at ~39 bp, alongside an invalid C peak. This was a problem, because in the gross

tissue analysis of these samples, the 288 and 329 samples were seen to be C bases at the 16390 position for each of the samples. This calls into question the validity of the controls, and the specificity of the 16390 extension primers. Once again, this was seen to be most likely reflective of the issue of the mtDNA concentration in the sample alongside the high amplification negative control threshold. None of the C peaks in the control cells were above this threshold; thereby causing the C peaks to be invalid. However, the presence of the T peak in the control sample opens the question of whether the T peak was truly present in the control cells, or if it was a random addition to the extension primer. If the T peak in the control cells was a true base, it means that it was at a high enough concentration in the sample cells to be detected and signaled by the 16390 extension primers. If that was the case, then why was the T peak not seen in the gross tissue analysis of the control tissues for 288? If it was assumed in this case that the concentration of the mtDNA genomes containing an T base were at a higher concentration, but the inefficiency of the 16390 extension primers kept this from being detected, then the T peak might have been real. However, if the T peak is the result of random, non-specific base addition to the 16390 extension primer in the presence of low mtDNA concentration, then a bigger problem exists. If this T peak is random, non-specific addition, then the 355 single cell samples would be affected as well, resulting in totally inconclusive data for the 355 cell set due to the untrustworthiness of the results. However, the T peak was expected as a valid heteroplasmic peak in the 355 cells for the 16390 primers, so it is believed to be valid in the 355 cells, which leaves the validity of the T peak in the control cell in question.

To address the specificity of the primers, it should be noted that the HL60 positive control sample displayed the appropriate C peak at ~36 bp with a peak height of 5870 RFU, so the fact that an C base can be added and detected in the 16390 extension primers is known to

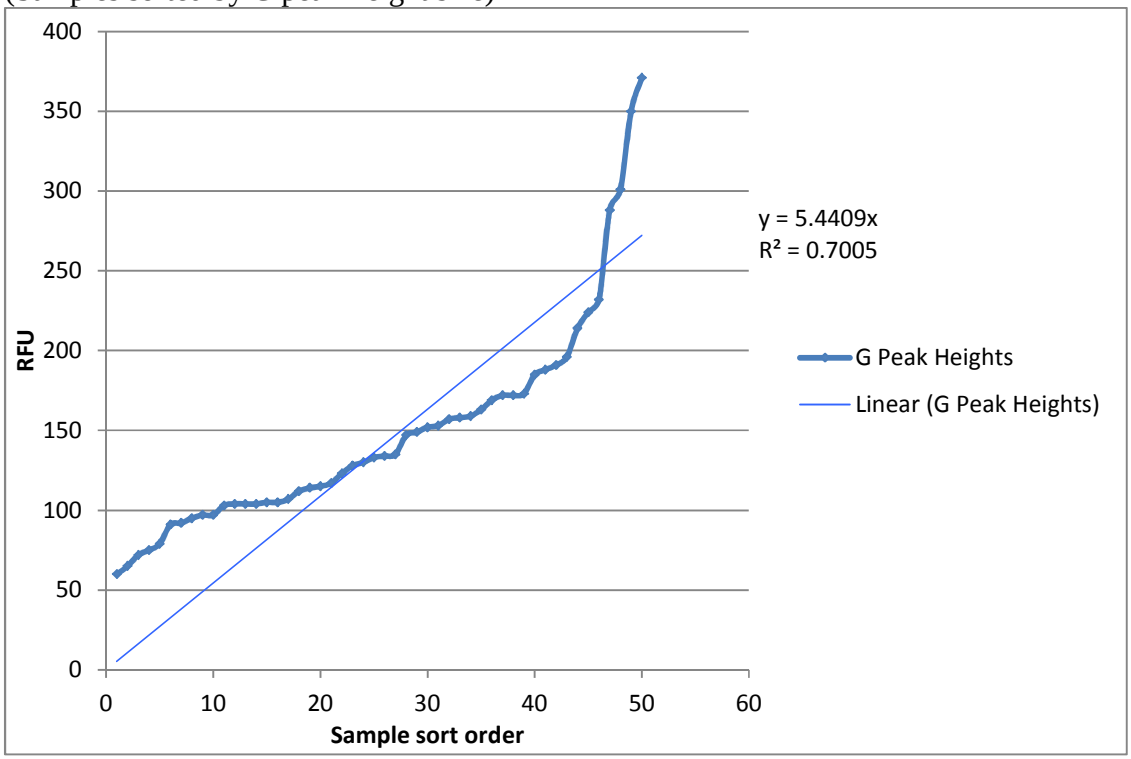
properly occur. Also, it would be unlikely for the T base displaying in the control sample to be any type of artifact, since a fluorescently-labeled ddNTP base must be adding to the extension primer for the samples to be detected, and this can only be occurring when the template strand contains the corresponding A base.

As to the issue of the validity of the control samples, the T base that was detected was above the negative control threshold, and was of the correct bp size for the 16390 extension primer. Since only one control cell was displaying the T base at 16390 with no valid C peaks, the implications of the controls samples are open to interpretation. If the control cell results for the 16390 extension primers are inconclusive, then the results of the 355 cells are as well.

Discussion of SNaPshot™ Results of Sample 355 (16519) Cell Analysis

When the data from the forty-seven viable 355 cells and ten control cells was analyzed, it was seen that no instances of G/A heteroplasmy were present in any of the 355 cells (refer to Table 39 in the results section). The 355 cells, along with the 288 and 329 control cells only displayed solitary G peaks in the data. Out of the fifty 355 cells, one sample (29-355) displayed no data and was negative. Two other samples (22-355 and 40-355) were below the negative control threshold and were therefore inconclusive and will not be included in the analysis. The G peak heights ranged from 60 to 371 RFU (Figure 125). No A peaks were seen in any of the 355 cells or control cells whatsoever.

Figure 125- Peak Heights for 355 cells, 16519 primers
(Samples sorted by G peak-height size)



Since none of the cells were seen to contain any heteroplasmic peaks in the ~43 bp region, it could be determined that one of three scenarios was occurring within these samples.

The first scenario was that the cells were reflecting a true homoplastic G base within the 16519 position of the mtDNA genomes. The second scenario was that the presence of the A base at the 16519 position was so minor within the mitochondrial DNA of the individual cells that it was being overwhelmed by the major G base signal. The third possibility was the issue of the non-addition of the ddATP bases due to the extension primers with the modified 5′ poly-thymine tails combined with the mtDNA input concentration.

When the original SNaPshot™ result for the gross tissue sample 369 was re-examined, it was seen that the G/A heteroplasmy ratio for the 16519 base positions was not 1:1. Instead, within the gross tissue sample, the G peak was 564 RFU, where the A peak was at 42 RFU. If it was assumed that the population of heteroplasmic mitochondrial DNA was not heterogeneous, and was split among unique cells (scenario #1 above), then within the forty-seven viable sample cells, a single cell containing the A base was not found. If it was assumed instead that a heterogeneous population of the heteroplasmic mitochondrial DNA exists within the cells (scenario #2 above), then the G bases are occurring over thirteen times more frequently than the minor A bases. The third scenario concerning the non-addition of the ddATP due to the poly-thymine modified extension primer and mtDNA concentration has already been identified as a source of error in the study, and was seen to be the most likely reason for the lack of A peaks in the data, especially for the 16519 extension primer with the longest poly-thymine tail. Apart from the known issue with the extension primer and the ddATP addition, the second scenario was more likely, due to the previous assumption that by sampling fifty cells from a known heteroplasmic tissue at least one cell should show the base change, based on known rates of heteroplasmy in such tissues. Assuming the heterogeneous spread of the mitochondrial DNA was equal within the cells and tissue sample, this could have contributed to the results of the

gross tissue heteroplasmy and the apparent homoplastic individual cells within the 355 sample set. This issue ultimately focuses on the mtDNA concentration in the samples. Since thousands of cells contributed mtDNA in the gross tissue sample, the concentration of the mtDNA could be measured on the Agilent 2100 Bioanalyzer, and the culmination of the signals of both bases was above the baseline but was not reflective of the unequal nature of the G/A signal. In a single individual cell, the mtDNA concentration was not detectable on the Agilent 2100 Bioanalyzer, and the prevalence of any heteroplasmic A base was never at a high enough concentration to be detected above the major signal of the G base within the single cells. Therefore, the mtDNA concentration along with the non-addition of the ddATP bases due to the poly-thymine modified extension primers most likely led to results that were seen for the 355 cells.

When examining the 288 and 329 control cells that were run along with the 355 cell, no instances of heteroplasmy were seen. However, as seen in Table 39, (pgs. 196-197), these samples all showed the presence of a G base at the ~43 bp size. This was a problem, because in the gross tissue analysis of these samples, the 288 and 329 samples were seen to be A bases at the 16519 position for each of the samples. This calls into question the validity of the controls, and the specificity of the 16519 extension primers. Once again, this was seen to be most likely reflective of the issue of the non-addition of the ddATP base with the poly-thymine modified extension primers in conjunction with the mtDNA concentration in the sample. However, the presence of the G peak in the control samples opens the question of whether the G peak was truly present in the control cells, or if it was a random addition to the extension primer. If the G peak in the control cells was a true base, it means that it was at a high enough concentration in the sample cells to be detected and signaled by the 16519 extension primers. If that was the case, then why was the G peak not seen in the gross tissue analysis of the control tissues for 288 and

329? If it was assumed in this case that the concentration of the mtDNA genomes containing an A base were at a higher concentration, but the ultimate inefficiency of the 16519 extension primers kept this from being detected, then the G peak might be real. However, if the G peak was the result of random, non-specific base addition to the 16519 extension primer in the presence of low mtDNA concentration, then a bigger problem exists. If this G peak was random, non-specific addition, then the 355 single cell samples would be affected as well, resulting in totally inconclusive data for the 355 cell set due to the untrustworthiness of the results.

To address the specificity of the primers, it should be noted that the HL60 positive control sample displayed the appropriate A peak at 44 bp with a peak height of 6082 RFU, so the fact that an A base can be added and detected in the 16519 extension primers is known to properly occur. This reinforces the notion that the ddATP addition is tied to the mtDNA concentration of the sample, since the HL60 positive control consistently showed a detectable Agilent 2100 Bioanalyzer value for the mtDNA concentration. Also, it would be unlikely for the G bases displaying in the control samples to be any type of artifact, since a fluorescently-labeled ddNTP base must be adding to the extension primer for the samples to be detected, and this can only be occurring when the template strand contains the corresponding C base.

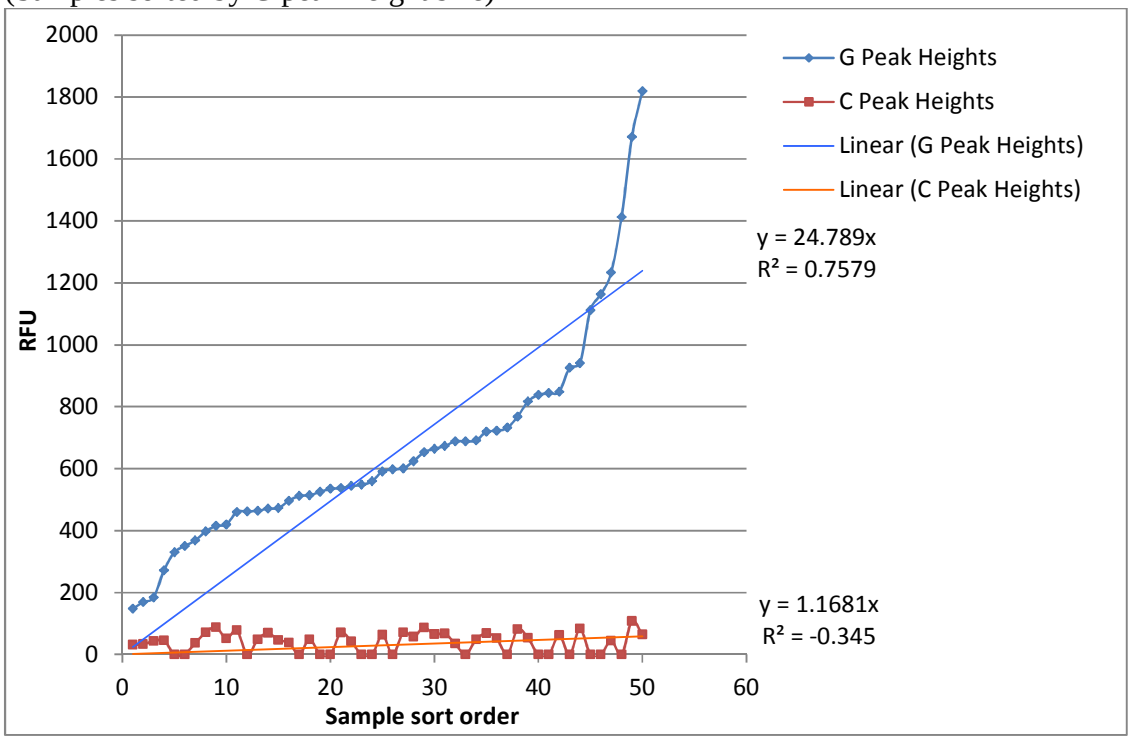
As to the issue of the validity of the control samples, the G bases that were detected were above the negative control threshold, and are of the correct bp size for the 16519 extension primer. Since the control cells are uniformly displaying the G base at 16519, while the gross tissue did not show any indication of any base other than the A at 16519, the implications of the controls samples are open to interpretation. However, since a different base is present at 16519 when compared to the gross tissue results, it is unclear if the control samples are valid or not. If

the control cell results for the 16519 extension primers are inconclusive, then the results of the 355 cells are as well.

Discussion of SNaPshot™ Results of Sample 369 (16519) Cell Analysis

When the data from the fifty viable 369 cells and ten control cells was analyzed, it was seen that no instances of G/A heteroplasmy were present in any of the 369 cells (refer to Table 41 in the results section). The 369 cells, along with the 288 and 329 control cells displayed either solitary G peaks or G/C peaks in the data. Out of the 369 cells, three samples (5-369, 17-369, and 18-369) displayed below-threshold G peak data but did contain valid C peak data, so the G peaks for those samples were inconclusive and the C peaks were valid. Out of the remaining forty-seven samples, all displayed G peaks at ~43 bp, and thirty-three of the samples also contained a C peak at ~44 bp. Typically the G peak height was 10 X that of C peaks, so the C peaks were minor (Figure 126). No A peaks were seen in any of the 369 cells or control cells whatsoever.

Figure 126- Peak Heights for 369 cells, 16519 primers
(Samples sorted by G peak-height size)



Since none of the cells were seen to contain any A peaks in the ~43 bp region, it could be determined that one of three scenarios were occurring within these samples. The first scenario was that the cells were reflecting true heteroplasmic G/C bases or homoplastic G bases within the 16519 position of the mtDNA genomes. The second scenario was that the presence of the A base at the 16519 position was so minor within the mitochondrial DNA of the individual cells that it was being overwhelmed by the major G and C base signals. The third possibility was the issue of the non-addition of the ddATP bases in the extension primers due to the modified 5′ poly-thymine tails in combination with the mtDNA input concentration.

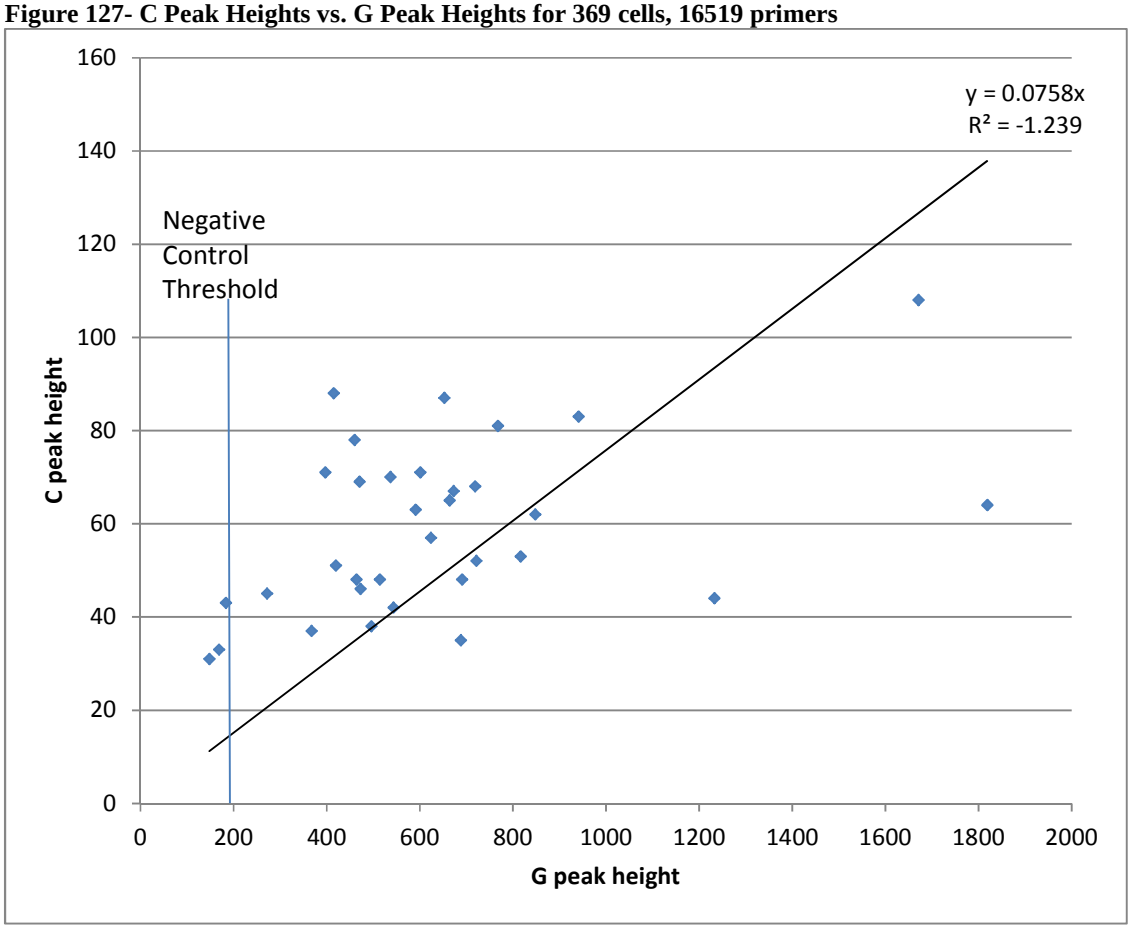
When the original SNaPshot™ result for the gross tissue sample 369 was re-examined, it was seen that the G/A heteroplasmy ratio for the 16519 base positions was not 1:1. Instead, within the gross tissue sample, the G peak was seen around 146 RFU, where the A peak appeared at 39 RFU, and no C peak was detected at all. If it was assumed that the population of heteroplasmic mitochondrial DNA was not heterogeneous, and was split among unique cells (scenario #1 above), then within the viable 369 sample cells, a single cell containing the A base was not found. If it was assumed instead that a heterogeneous population of the heteroplasmic mitochondrial DNA existed within the cells (scenario #2 above), then the G bases are occurring over four times more frequently than the minor A bases. The third scenario concerning the non-addition of the ddATP due to the poly-thymine modified extension primer and mtDNA concentration has already been identified as a source of error in the study, and was seen to be the most likely reason for the lack of A peaks in the data, especially for the 16519 extension primer with the longest poly-thymine tail. Apart from the known issue with the extension primer and the ddATP addition, the second scenario was more likely, due to the previous assumption that by sampling fifty cells from a known heteroplasmic tissue at least one cell should show the base

change, based on known rates of heteroplasmy in such tissues. Assuming the heterogeneity of the mitochondrial DNA was equal within the cells and tissue sample, this could have contributed to the results of the gross tissue heteroplasmy and the divergent G/C heteroplasmic individual cells within the 369 sample set. This issue ultimately focuses on the mtDNA concentration in the samples. Since thousands of cells contributed mtDNA in the gross tissue sample, the concentration of the mtDNA could be measured on the Agilent 2100 Bioanalyzer, and the culmination of the signals of both bases was above the baseline but was not reflective of the unequal nature of the G/A signal. At the same time, the G/A concentrations in the gross tissue were possibly in such high concentrations as to overwhelm a minor C signal also present in the data. In a single individual cell, the mtDNA concentration was not detectable on the Agilent 2100 Bioanalyzer, and the prevalence of any heteroplasmic A base was never at a high enough concentration to be detected above the major signal of the G base within the single cells. Combined with the issue of the non-addition of the ddATP in the presence of the poly-thymine modified extension primers, the previously minor C base was now capable of being probed and successfully detected in the single cells. This is somewhat verified based on the graph of the C peak heights in relation to the G peak heights (Figure 126). As the C peaks increase in height, the G peak heights do not seem to be affected. For the 369 cells, it appears that there is no correlation between the presence of the C peaks and G peaks. Therefore, the mtDNA concentration along with the non-addition of the ddATP bases due to the poly-thymine modified extension primers most likely led to results that were seen for the 369 cells.

In order to determine if the heteroplasmic data for the 369 cells was due to amplification and/or electrophoresis artifacts or not, the C and G peak heights from the heteroplasmic cells were plotted, as shown in Figure 127. Due to the low negative control threshold on the G peaks,

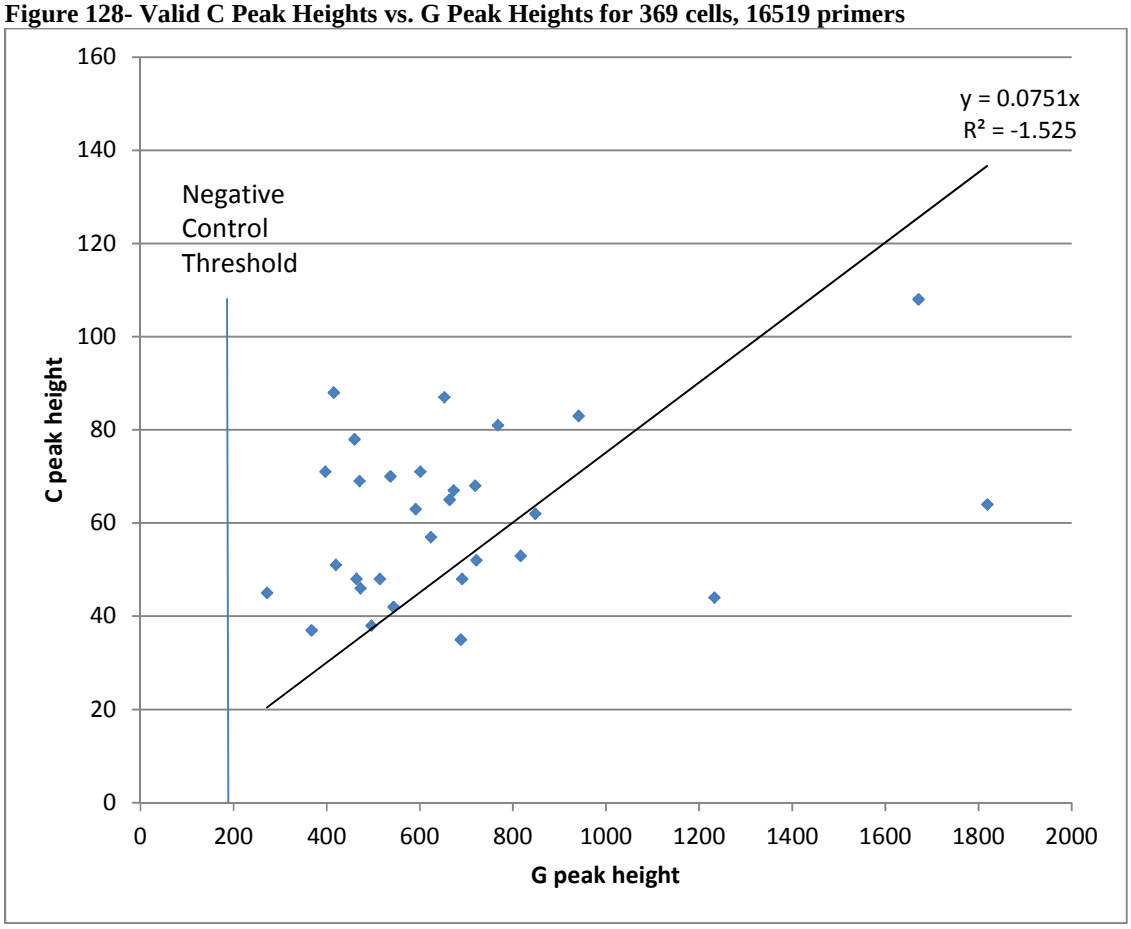
only three cells were seen to be invalid, leaving a majority of the 369 cells valid for interpretation. However, the best-fit R^2 value for the heteroplasmic cells was seen to be negative (-1.23) which would indicate that the data was non-linear, resulting in the interpretation that the C and G peaks are unique and independent signals in the heteroplasmic 369 cells.

Figure 127- C Peak Heights vs. G Peak Heights for 369 cells, 16519 primers



When the invalid cells were removed from the plot, the best-fit R^2 value did not improve, and remained negative, reinforcing the notion that the C and G signals from the heteroplasmic cells were unique and independent signals (Figure 128).

Figure 128- Valid C Peak Heights vs. G Peak Heights for 369 cells, 16519 primers



When examining the 288 and 329 control cells that were run along with the 369 cell, no instances of heteroplasmy were seen. However, as seen in Table 41, (pgs. 200-201), these samples all showed the presence of a G base at the ~43 bp size. This was a problem, because in the gross tissue analysis of these samples, the 288 and 329 samples were seen to display A bases at the 16519 position for each of the samples. This calls into question the validity of the controls, and the specificity of the 16519 extension primers. Once again, this is seen to be most likely reflective of the issue of the non-addition of the ddATP base with the poly-thymine modified extension primers in conjunction with the mtDNA concentration in the sample. However, the presence of the G peak in the control samples opens the question of whether the G peak is truly present in the control cells, or if it was a random addition to the extension primer. If

the G peak in the control cells was a true base, it means that it was at a high enough concentration in the sample cells to be detected and signaled by the 16519 extension primers. If that was the case, then why was the G peak not seen in the gross tissue analysis of the control tissues for 288 and 329? If it was assumed in this case that the concentration of the mtDNA genomes containing an A base were at a higher concentration, but the ultimate inefficiency of the 16519 extension primers kept this from being detected, then the G peak might be real. However, if the G peak was the result of random, non-specific base addition to the 16519 extension primer in the presence of low mtDNA concentration, then a bigger problem exists. If this G peak was random, non-specific addition, then the 355 single cell samples would be affected as well, resulting in totally inconclusive data for the 355 cell set due to the untrustworthiness of the results.

To address the specificity of the primers, it should be noted that the HL60 positive control sample displayed the appropriate A peak at 43 bp with a peak height of 8223 RFU, so the fact that an A base can be added and detected in the 16519 extension primers is known to properly occur. This reinforces the hypothesis that the ddATP addition is tied to the mtDNA concentration of the sample, since the HL60 positive control consistently showed a detectable Agilent 2100 Bioanalyzer value for the mtDNA concentration. Also, it would be unlikely for the G bases displaying in the control samples to be any type of artifact, since a fluorescently-labeled ddNTP base must be adding to the extension primer for the samples to be detected, and this can only occur when the template strand contains the corresponding C base.

As to the issue of the validity of the control samples, the G bases that were detected were above the negative control threshold, and are of the correct bp size for the 16519 extension primer. Since the control cells are uniformly displaying the G base at 16519, while the gross

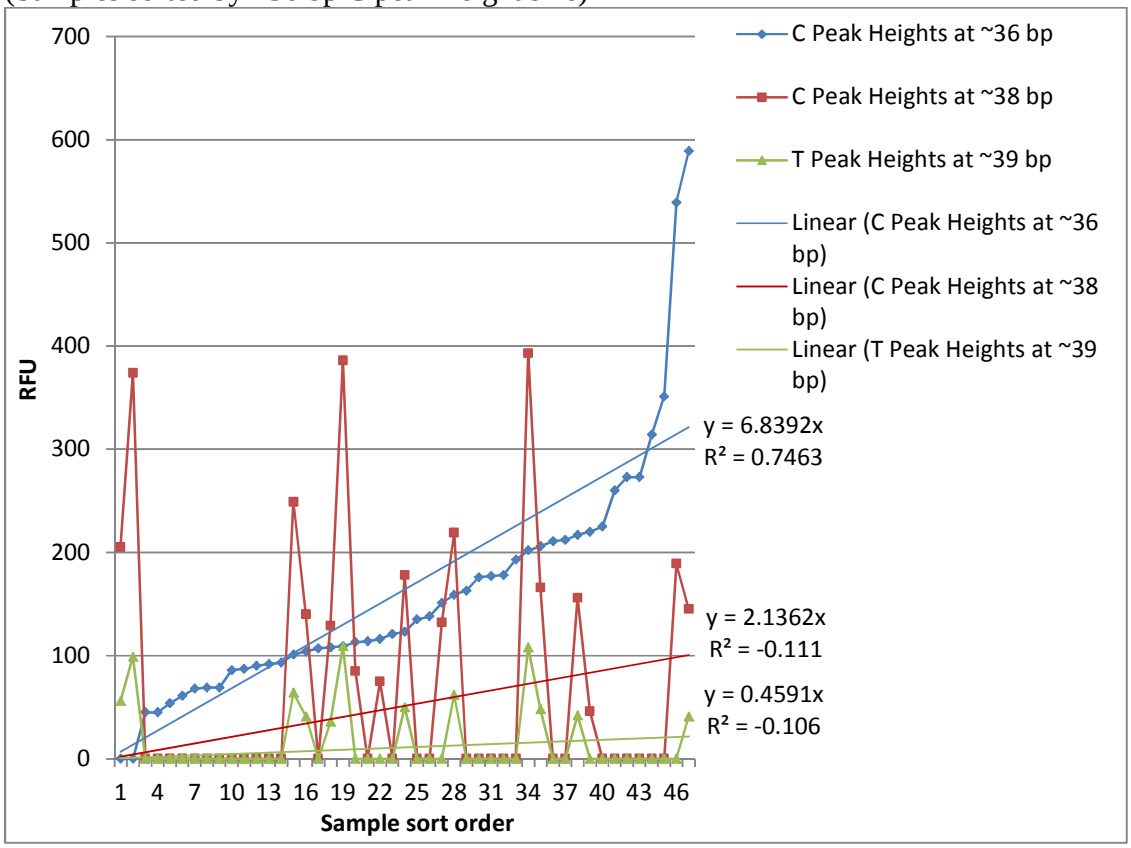
tissue did not show any indication of any base other than the A at 16519, the implications of the controls samples are open to interpretation. However, since a different base is present at 16519 when compared to the gross tissue results, it is unclear if the control samples are valid or not. If the control cell results for the 16519 extension primers are therefore inconclusive, then the results of the 369 cells are as well.

Additional testing of non-heteroplasmic tissues- 261 and 342

Discussion of SNaPshot™ results of non-heteroplasmic tissue showing possible heteroplasmic cells (sample 261, with the 16390 primers)

Based on the gross tissue analysis following the length study, sample 261 was seen to contain a solitary C peak around ~36 bp for the 16390 primer. For the 261 cells amplified with the 16390 primers, the peak heights for the C peaks at ~36 bp ranged from 45 to over 580 RFU. The peak heights for the C peaks at ~38 bp ranged from 46 to over 390 RFU, and the T peaks ranged in height from 36 to over 100 RFU (Figure 129). The C peak at ~36 bp is the major peak for the sample, with the additional C and T peaks being minor signals to the cells when they occur.

Figure 129- Peak Heights for 261 cells, 16390 primers
(Samples sorted by ~36 bp C peak-height size)



For the control samples for the 261 cells, nine of the ten samples displayed a valid C peak at ~36 bp. Six of the cells also displayed a C peak at ~38 bp, and five of these samples also contained a T peak at ~39 bp. For the one control sample that contained an invalid C peak at ~36 bp, the C peak at ~38 bp and the T peak at ~39 bp were present as well.

There does not appear to be any correlation with the appearance of the C peak at ~36 bp and the presence of either the additional C peak at ~38 bp or the T peak at ~39 bp (Figure 129). In order to determine if the heteroplasmic data for the 261 cells was due to amplification and/or electrophoresis artifacts or not, the C and T peak heights from the heteroplasmic cells were plotted, as shown in Figure 130, Figure 131, and Figure 132. Due to the negative control threshold only applying to the C peaks at ~36 bp, and no heteroplasmic cells were present below this threshold, all of the resulting heteroplasmic cells were determined to be valid for interpretation. When comparing the ~38 bp C peaks (Figure 130) and the T peaks (Figure 131) to the C peaks at ~36 bp, it is seen that the data points are highly variable, and both instances display a best-fit R^2 value that is negative, indicating that the data points in each instance are non-linear, and possibly indicative of unique and independent signals from the peaks in the heteroplasmic 261 cells.

Figure 130- C Peak Heights at ~38 bp vs. C Peak Heights at ~36 bp, 261 cells, 16390 primers

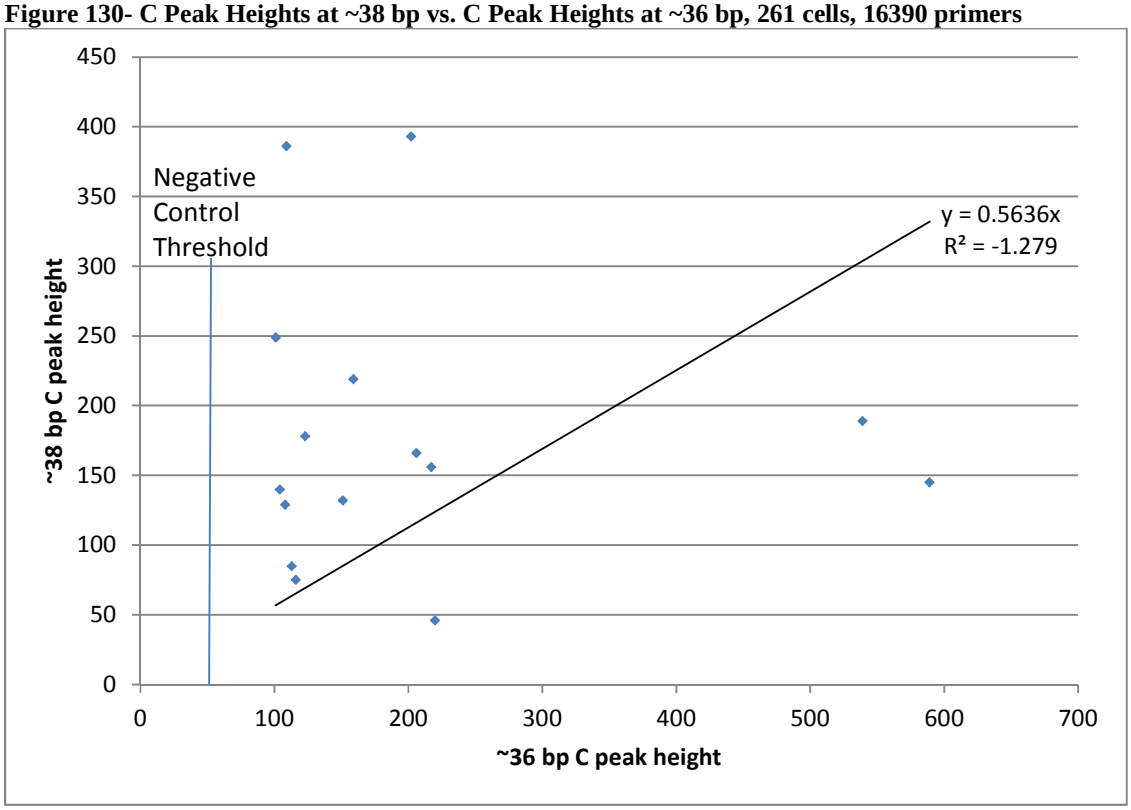
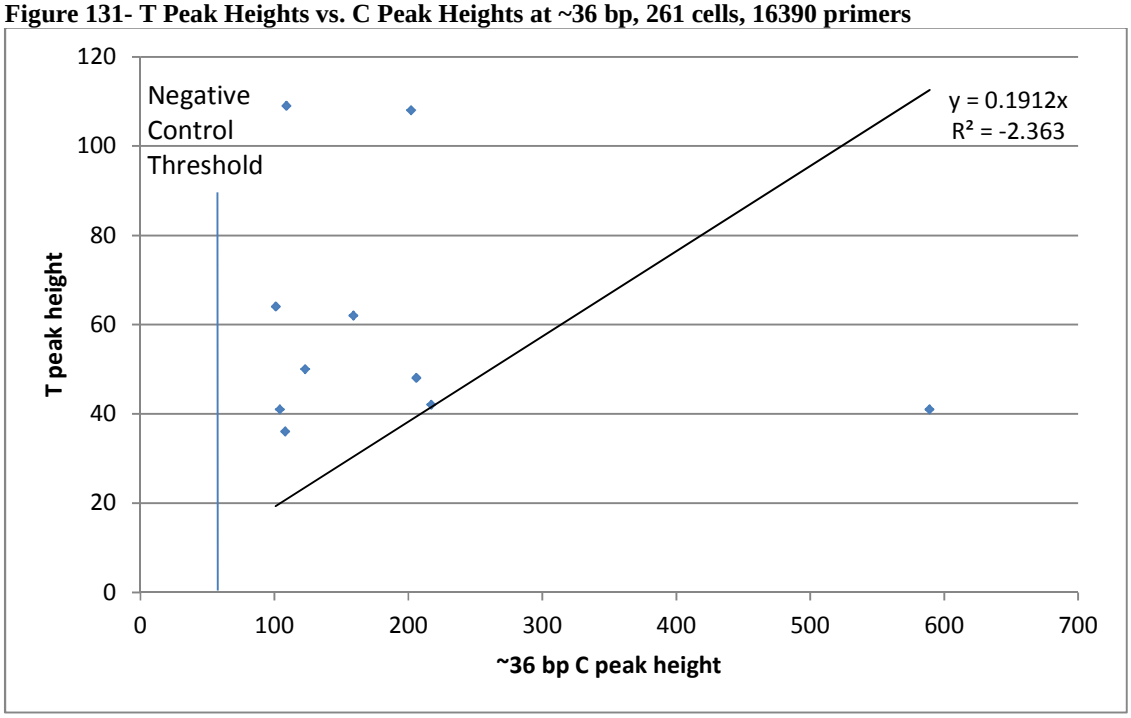
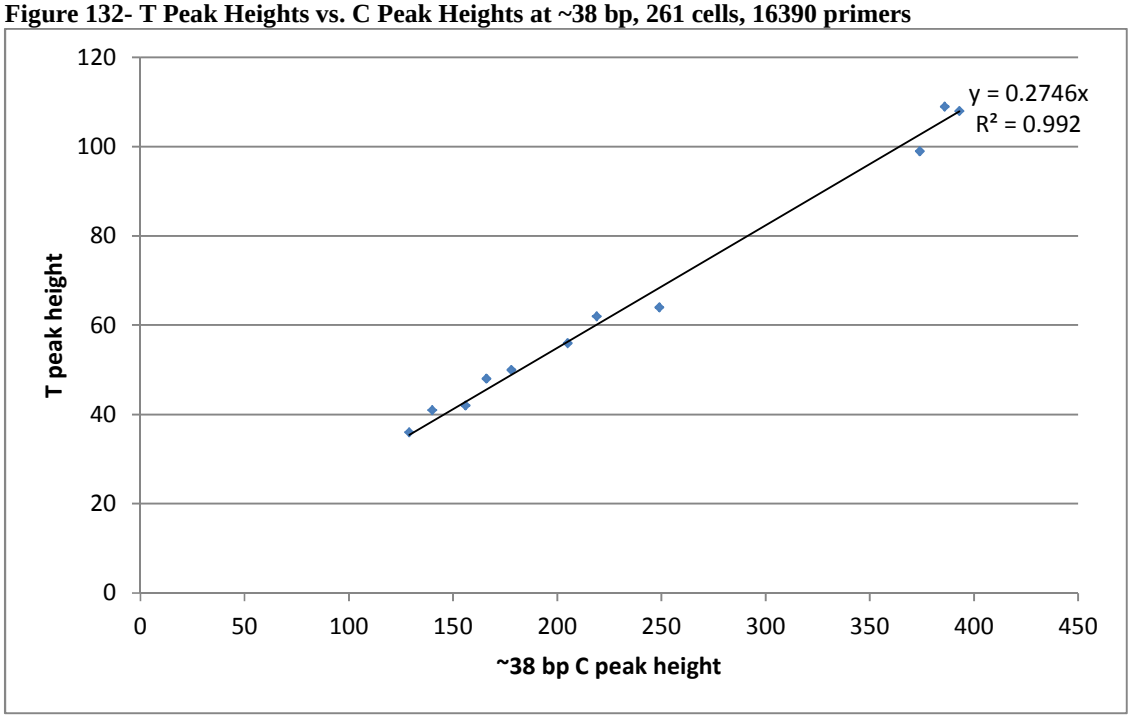


Figure 131- T Peak Heights vs. C Peak Heights at ~36 bp, 261 cells, 16390 primers



However, as noted when the T peaks were compared directly to the C peaks at ~38 bp (Figure 132), there was nearly a 1:1 correlation of these two peaks, which is also graphically demonstrated in Figure 129 as well. With a best-fit R^2 value of 0.99 for the T and ~38 bp C peak-height plots, it is clear that these peaks were the result of an amplification or electrophoresis artifact of some kind, but as to which peak is real and which peak is the artifact remains an outstanding question. If it is assumed that the true peak demonstrated the highest peak height between the two signals, then the C peaks at ~38 bp would be the true peak with the T peaks being the artifact. However, the presence of a heteroplasmic C peak at ~36 bp versus ~38 bp is also questionable, so it is unclear what the true answer for this data set is simply based on the data plots of the heteroplasmic peak heights.

Figure 132- T Peak Heights vs. C Peak Heights at ~38 bp, 261 cells, 16390 primers



Based on the gross tissue analysis, the 16390 extension primer showed a homoplastic C base, but did display heteroplasmy at the 16223, 16324, and 16519 hotspots. Sample 261 was only one of two samples to display more than two instances of heteroplasmy, with the second

sample (sample 345) having been removed from the study over concerns of the tissue sample pathology. The fact that the 261 sample displayed heteroplasmy at three separate locations was troublesome, and led to the assumption that the sample was a mixture of two separate tissues, possibly introduced in the histology fixation stage (Popiolek et al., 2003). Additional support for this assumption is seen in common rules regarding heteroplasmy in mitochondrial DNA analysis. In a sample, if heteroplasmy is seen twice (as in two separate locations of point heteroplasmy within one sample) it is recommended that the sample be retested to confirm the heteroplasmic points, because the recorded instances of valid two-point heteroplasmy is rare, but not impossible. Therefore, the presence of three or more heteroplasmic points in a sample is regarded as indicative of a mixture of two mtDNA sources, and the sample should not be processed further. This was the rationale for not testing the heteroplasmic positions at 261, but since the 16390 primers were already attempted against the 261 cells, it was an opportunity to test the effectiveness of the SNaPshot™ assay against a supposedly homoplastic hotspot at 16390. The results, seen above, do indicate that a mixture of mtDNA might be present in the sample due to the mixture of results seen at the ~36, ~38, and ~39 bp positions, but is difficult to determine based on the issues of the mtDNA concentration applied to the SNaPshot™ reaction, as well as the issue of the proper ddNTP base addition in conjunction with the mtDNA concentration.

There is also evidence to show that the 261 cells may not contain a mixture of mtDNA, but do contain valid heteroplasmy. If the control cells for the 261 samples are any indicator, it would be assumed that the results are somewhat valid since the control tissues of 288 and 329 did not contain any heteroplasmic peaks in the gross tissue analysis, but the cells in the 16390 primers are showing what appears to be heteroplasmic peaks. The results of the 261 cells are

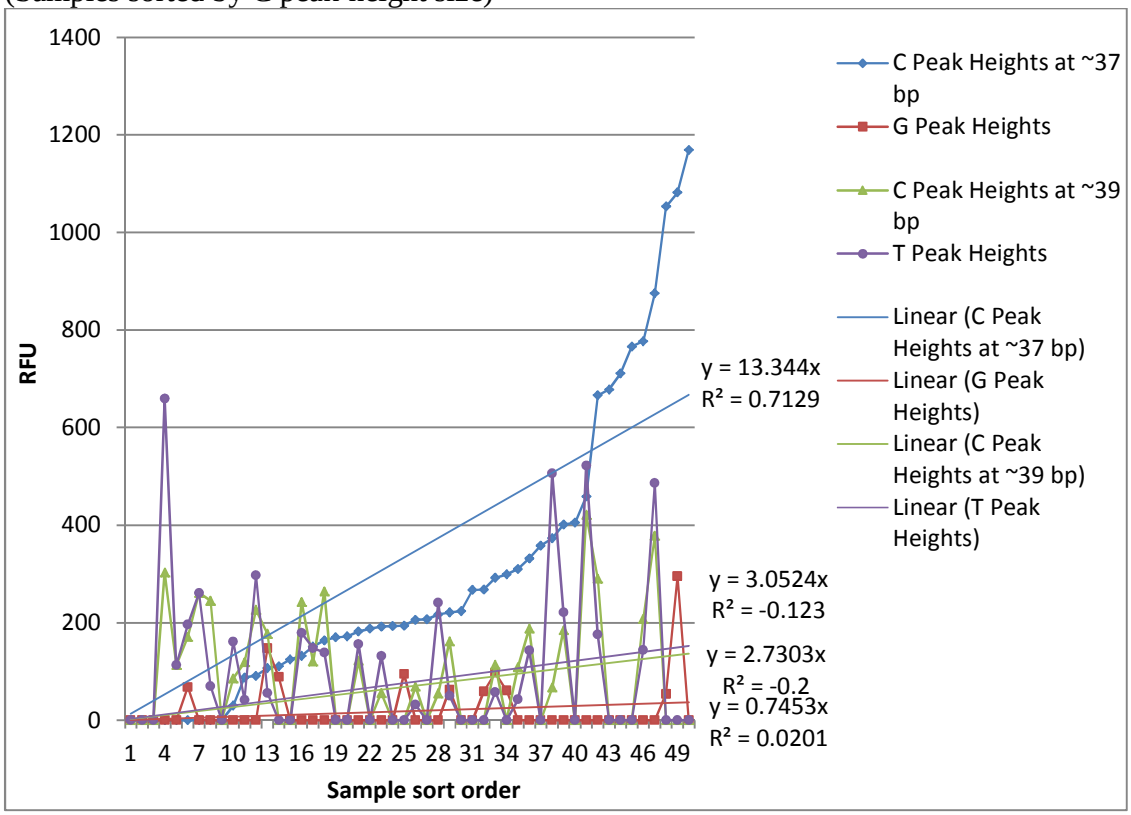
fascinating from the standpoint that the gross tissue did not display heteroplasmy at the 16390 position but it is being seen in the cells, which is a repeat of what was seen in the control cells from the ten valid test sets. Again, this might be tied to the overall concentration of the heteroplasmic bases in the gross tissue compared to the cells. In the case of the 261 cells (assuming no mixture of mtDNA), what might have been occurring was that the overwhelming C signal at ~36 bp caused the gross tissue to display a solitary C base, but when the single cells were tested, the overall amount of this ~36 bp C signal was diminished, allowing the additional ~38 bp C peak and ~39 bp T peak to be detected. This again highlights the idea that the mitochondria are heterogeneous within the cells, such that some of the 261 cells displayed the solitary C peak at ~36 bp, while others displayed the combination of two or three peaks. This reinforces the hypothesis that at the gross tissue level the dominant signal of the ~36 bp C peak overwhelmed the additional signals, and was only able to be overcome by the additional peak signals when the overall concentration of the ~36 bp C peak signal was limited to a single cell. However, if the 261 cells are actually a mixture of two or more mtDNA sources, then the results are inconclusive and should not be interpreted.

Discussion of the SNaPshot™ results of non-heteroplasmic tissue showing possible heteroplasmic cells (sample 342, with the 16390 primers)

Based on the gross tissue analysis following the length study, sample 342 was seen to contain a solitary C peak around ~36 bp for the 16390 primer. When examining the 342 cell data, the peak-height range for the C peak at ~36 bp suggests that this C peak is the dominant signal in the cell, with a range of peak heights seen from 30 to over 1100 RFU (Figure 133). There are only six out of the fifty cells that contained a valid C peak signal at ~36 bp along with valid C/T peaks at ~39-40 bp (cells 3-342, 5-342, 11-342, 12-342, 16-342, and 26-342). There

are only two cells that contained a valid C peak at ~36 bp and a valid G peak at ~38 bp (48-342 and 49-342).

Figure 133- Peak Heights of 342 cells, 16390 primers
(Samples sorted by C peak-height size)



However, out of the fifty 342 cells, only forty-one contain a C peak at ~36 bp, and out of these forty-one cells, only twenty are valid C peaks. Out of the ten control cells, only nine contained the C peak at ~36 bp, and out of these, only three of the nine were valid. This indicates that while the C peak at ~36 bp might be dominant, it was not present in every cell, and even when present, only half of the samples displayed a valid peak over what was a low-to-moderate amplification negative control threshold of 236 RFU. The fact that additional peaks are appearing in the cells from this sample again leads to the question; is this actual cellular-level heteroplasmy that is being seen due to a drop in the dominant signal, or is this some form of contamination? Based on the analysis of the 342 cell data, there appeared to be no correlation

with the presence of the C peak at ~37 bp with: the G peak at ~38 bp (Figure 134) the C peak at 39 bp (Figure 135), or the additional T peak at ~40 bp (Figure 136). Each of the respective best-fit R^2 values below are negative, indicating non-linear data for each comparison, demonstrating the possibility that the heteroplasmic instances are unique and independent base signals in the heteroplasmic cells.

Figure 134- G Peak Height vs. C Peak Height at ~37 bp, 342 cells, 16390 primers

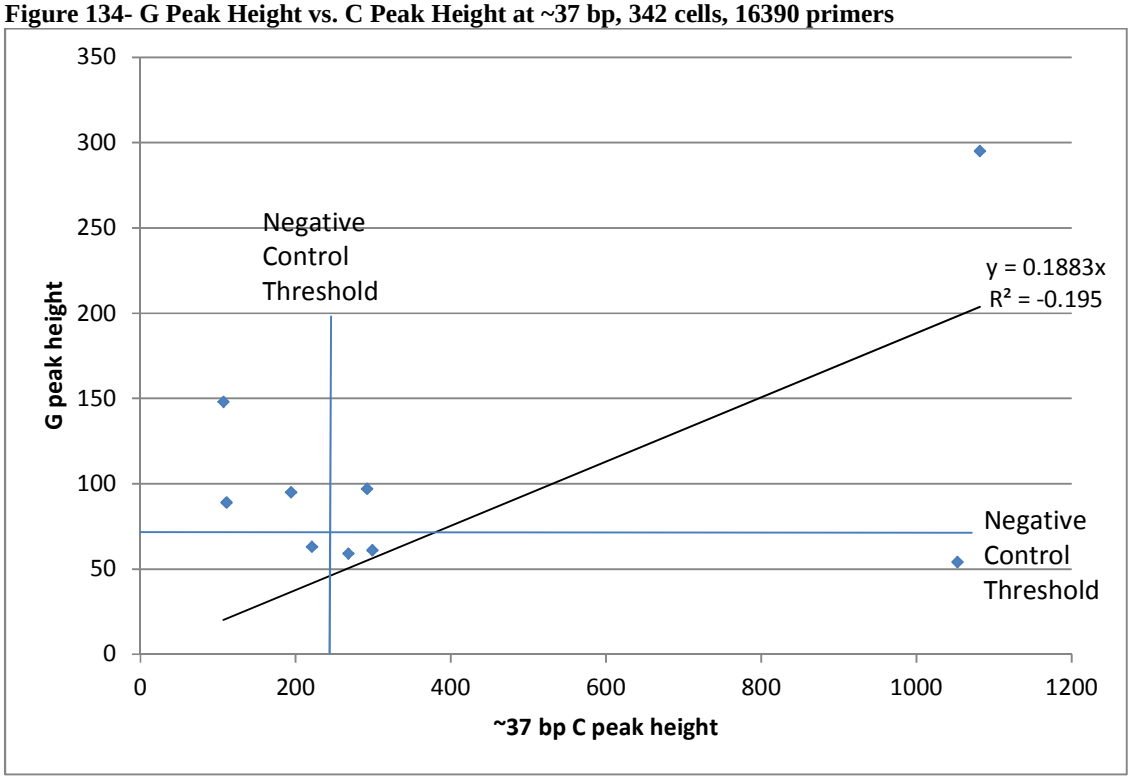


Figure 135- C Peak Height at ~39 bp vs. C Peak Height at ~37 bp, 342 cells, 16390 primers

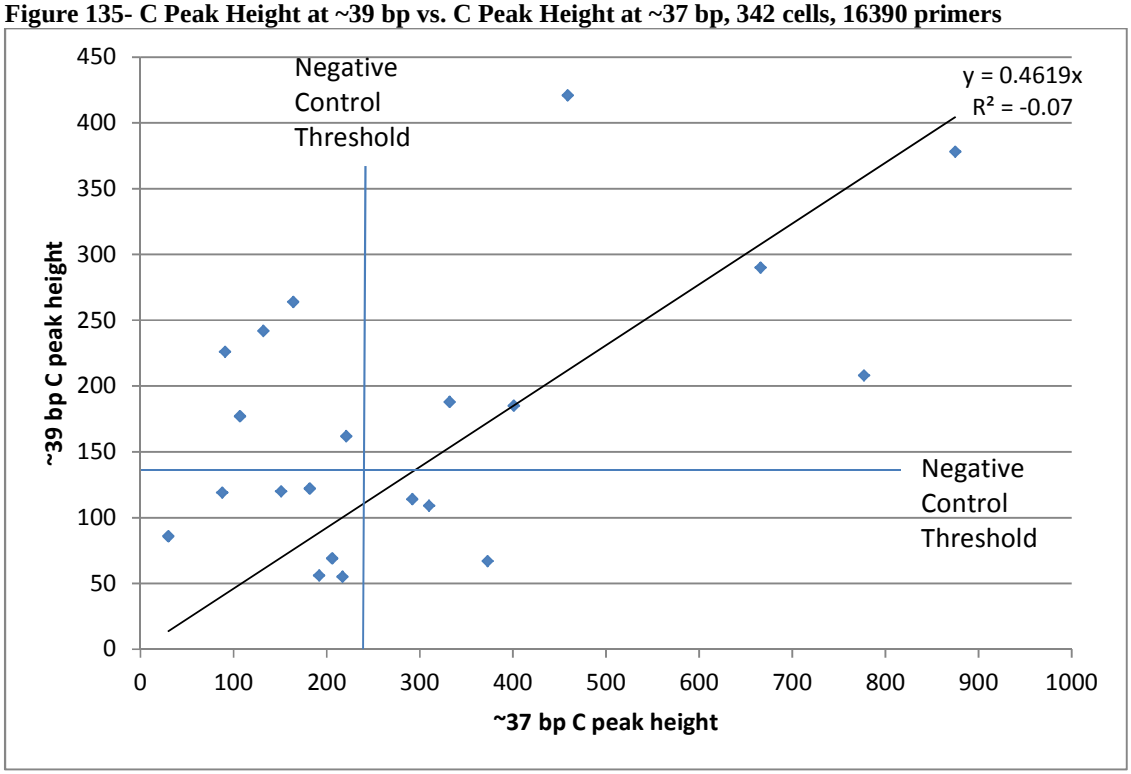
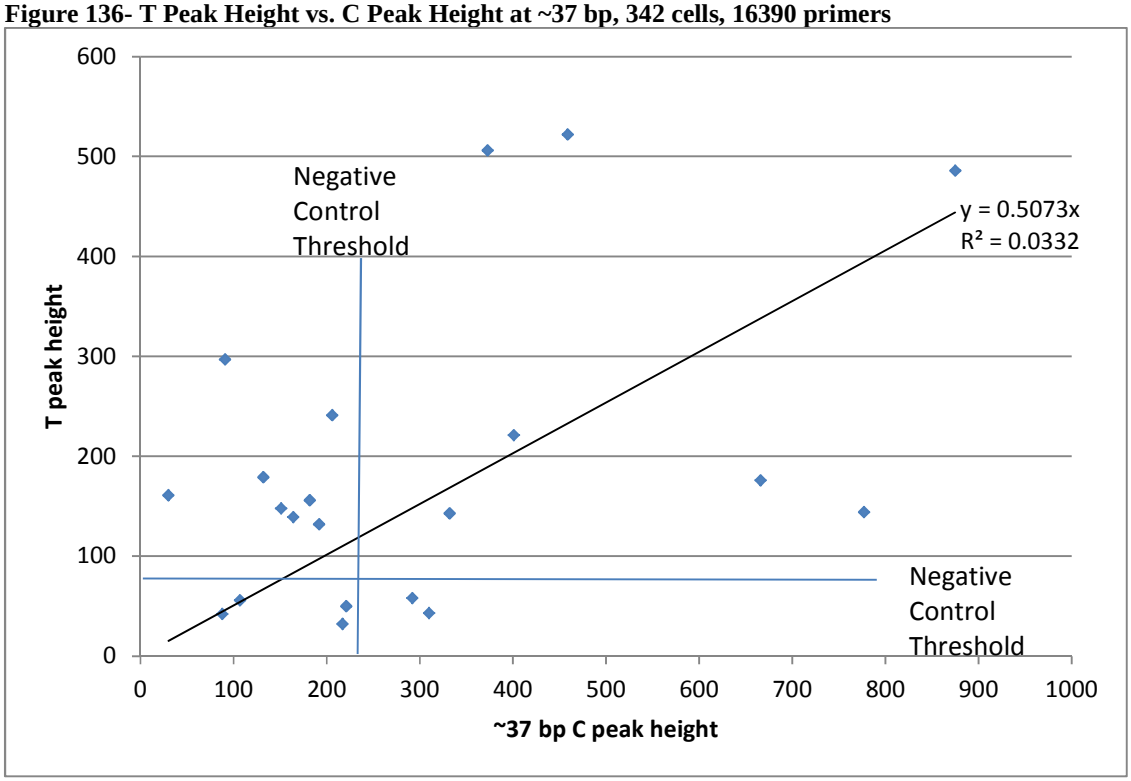


Figure 136- T Peak Height vs. C Peak Height at ~37 bp, 342 cells, 16390 primers



When the invalid cells were removed from the plots, the best-fit R^2 value did not improve, for the comparison of the ~37 bp C peak to the ~39 bp C peak (Figure 138) and the T peak to the ~37 bp C peak (Figure 139) and remained negative, reinforcing the notion that the T and C signals from the heteroplasmic cells were unique and independent signals. The resulting valid cells from the comparison of the G peaks to the ~37 bp C peaks resulted in only two valid data points, which artificially skewed the best-fit to 0.98, which is not a valid comparison with only two data points in the plot (Figure 137).

Figure 137- Valid G Peak Height vs. C Peak Height at ~37 bp, 342 cells, 16390 primers

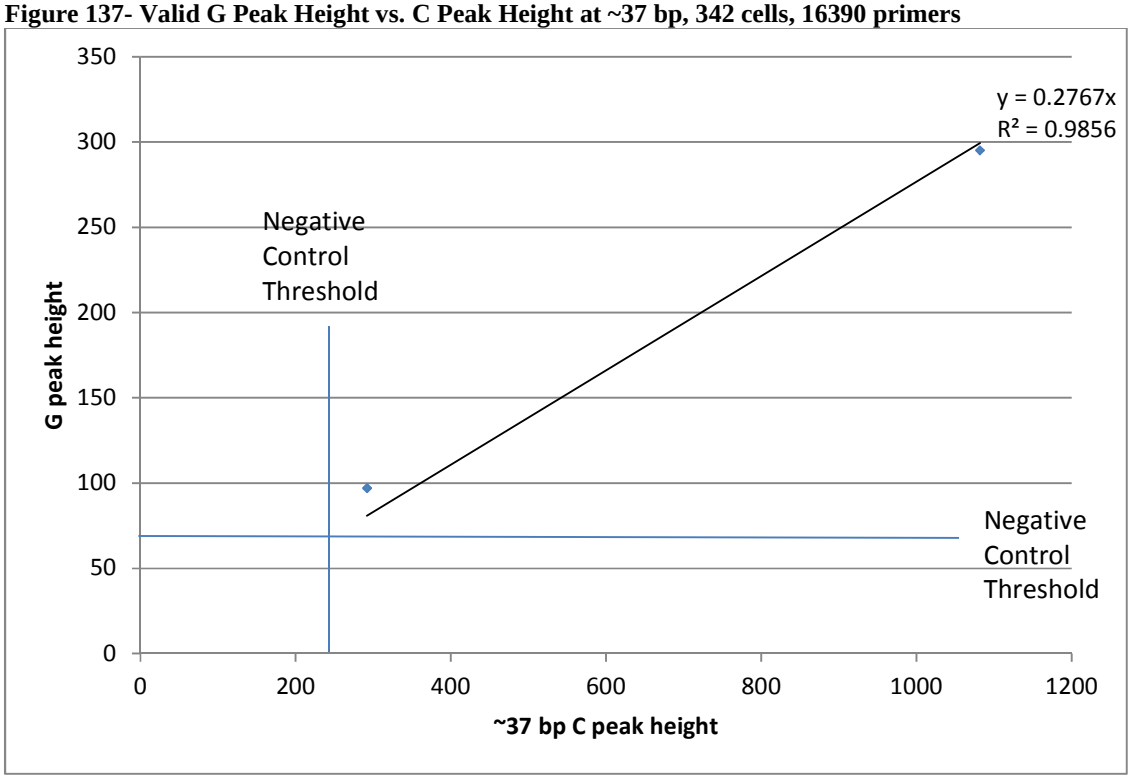


Figure 138- Valid C Peak Height at ~39 bp vs. C Peak Height at ~37 bp, 342 cells, 16390 primers

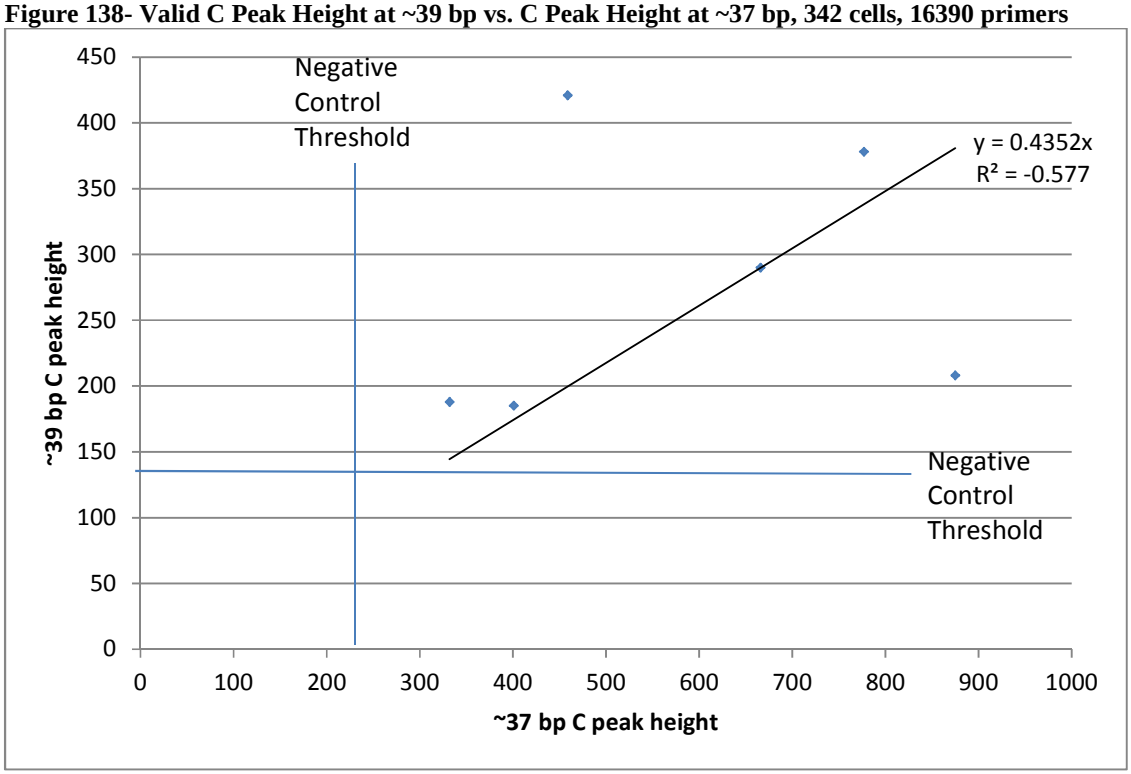
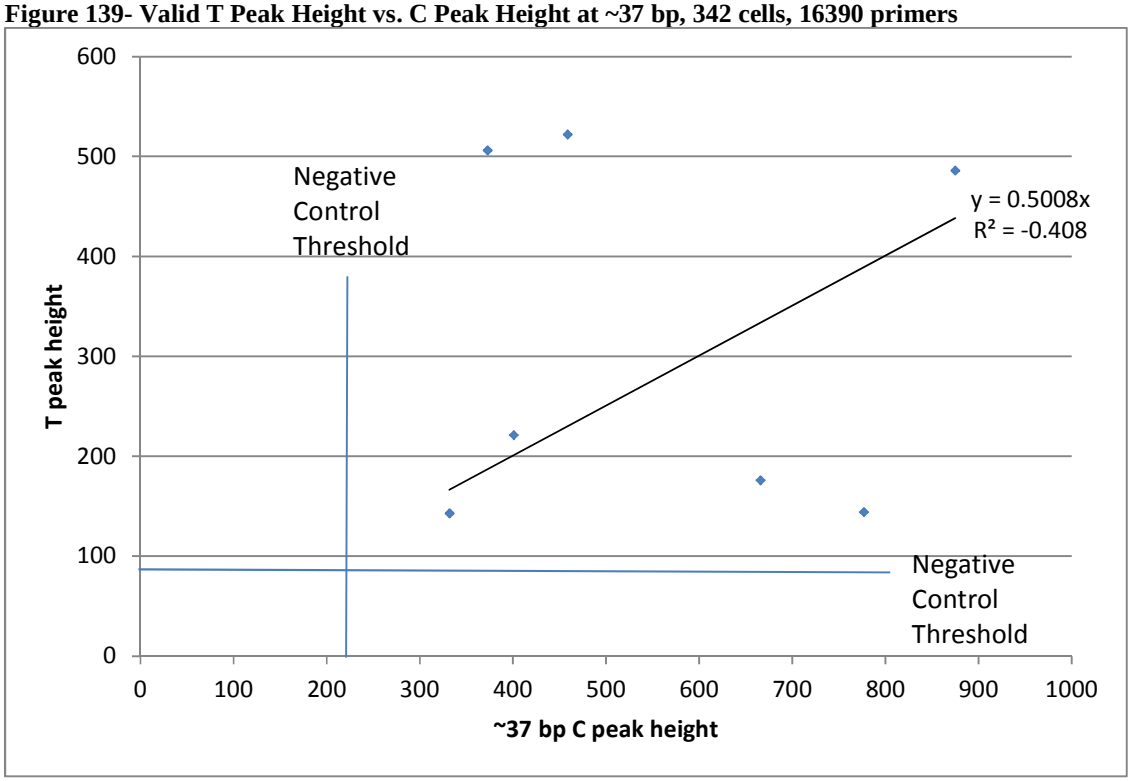


Figure 139- Valid T Peak Height vs. C Peak Height at ~37 bp, 342 cells, 16390 primers



The C and T peak at ~39 and ~40 bp do appear to be linked, in such that the peaks always appear as a pair (Figure 140), but when the peak heights are graphed on the scatter plot, the best-fit R^2 value is only 0.27, indicating very little possibility that the peaks are correlated (Figure 141).

Figure 140- T and ~39bp C Peak Heights of 342 cells, 16390 primers

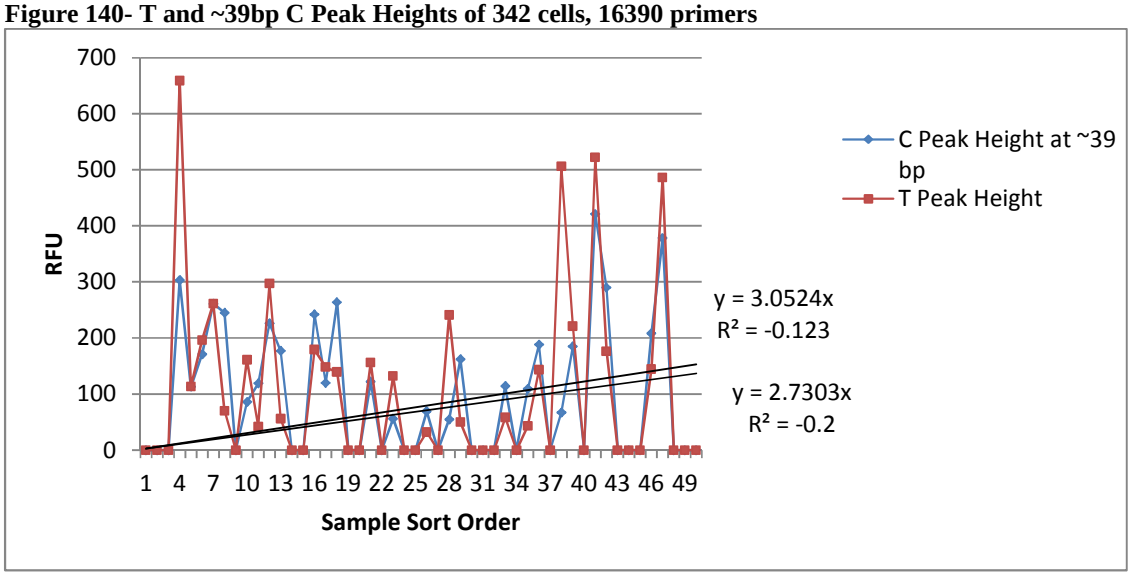
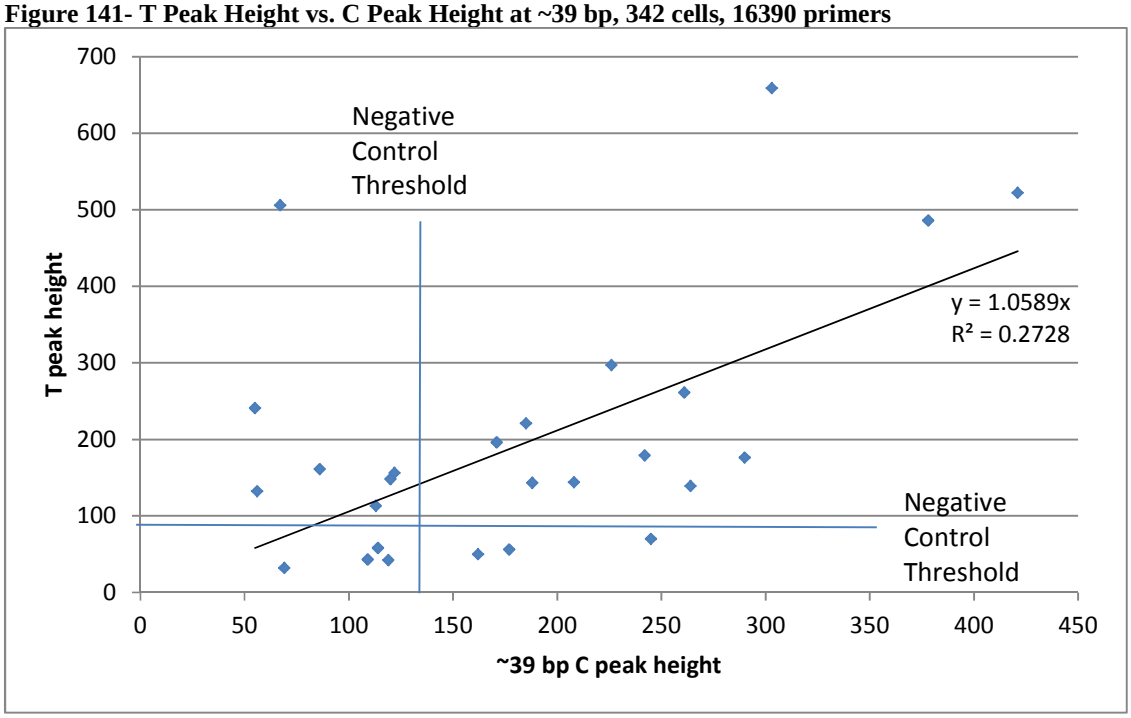
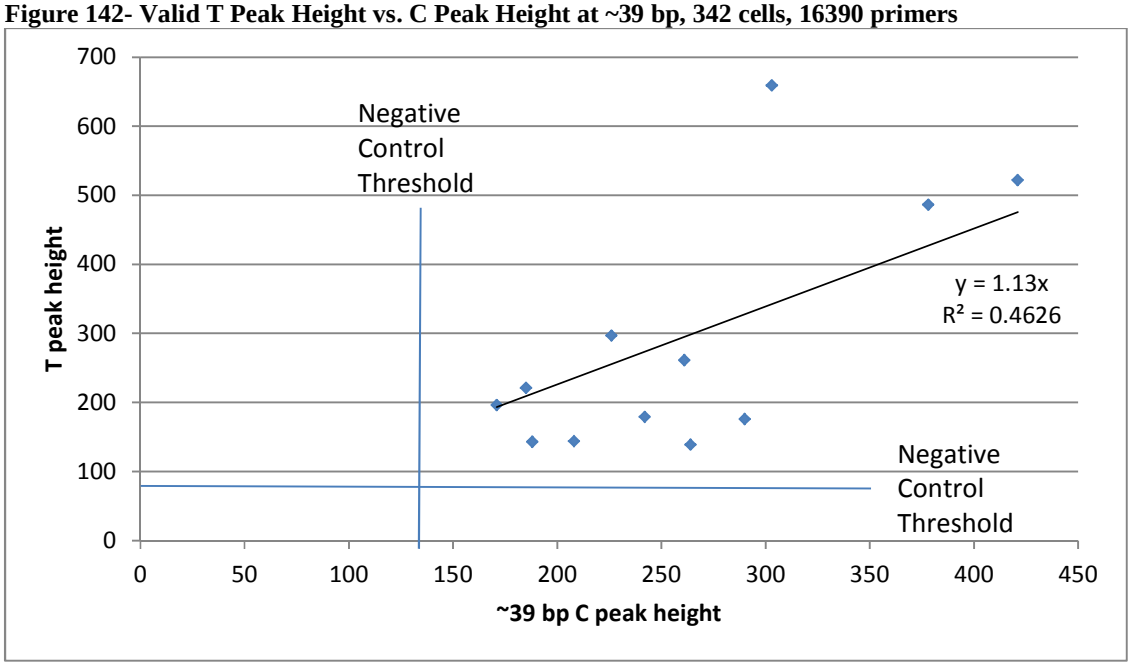


Figure 141- T Peak Height vs. C Peak Height at ~39 bp, 342 cells, 16390 primers



When the invalid cells were removed from the plot, the best-fit R^2 value did not drastically improve, reinforcing the notion that the T and C signals from the heteroplasmic cells were unique and independent signals (Figure 142).

Figure 142- Valid T Peak Height vs. C Peak Height at ~39 bp, 342 cells, 16390 primers



By examining the control cells, it is noted that both the 288 and 329 controls displayed at solitary C base at the 16390 primer position in the gross tissue analysis, however, the prior analysis of the 339 and 355 cells at the 16390 primer region both revealed additional T peaks ~39 bp in the control cells as well (see Table 35 on page 188 and Table 37 on page 196 for reference). A contaminant cannot be ruled out when looking at the 342 cellular data, especially in light of the prior assessment of the 261 cells, in which the standard mitochondrial analysis rules were applied and more than two instances of heteroplasmy were seen to be indicative of a contamination event (Wilson et al., 1993). However, that analysis was based on the presence of more than one heteroplasmic hotspot for the 261 cells, which is not the case for the 342 cells. In the 342 cells, what occurred was the presence of more than one heteroplasmic base in the single 16390 primer position. The complication with the 342 cells was the presence of two seemingly

independent C peak signals. It was originally believed that the presence of the T peak at ~40 bp was shifting the original ~36 bp C peak to ~39 bp, but in six cells, both the ~36 and ~39 bp C peaks are present, invalidating the idea of a single-base shift. If both C peaks are valid, then it was possible that other events were occurring in the SNaPshot™ assay independent of the 16390 extension primer. For both C peaks to occur within the data roughly 3-4 RFU apart, it would be reasonable to assume that an additional base was somehow being inserted into the extension primer, and perhaps the C peak at ~39 bp is actually probing the 16389 position instead (the 16390 extension primer detects on the reverse strand in the opposite direction). If this was the case, then inefficient ExoSAP-IT® might be the cause of leftover dNTP bases that would allow additional priming and detection of this peak. However, for the C and T peaks at ~39 and ~40 bp to appear, it would mean that this 16389 position would also be highly heteroplasmic, and there is no indication of this in the current literature. This would also mean that the ExoSAP-IT® process was somewhat successful at the same time, allowing the ~36 bp C peak and the ~38 bp G peak to be properly detected.

Another possible reason for the 342 results might have nothing to do with the 16390 extension primer, and instead be due to the primer-binding region in the mtDNA genome. Because the 16390 extension primer is twenty-nine nucleotides in length (after the poly-thymine modification) and it binds to the reverse strand on the mtDNA genome, it would effectively cover the 16391 to 16420 base positions. In the literature, there are several other known SNPs that occur within this region, specifically at 16391, 16398, 16399, and 16400. While the 16390 heteroplasmy is seen in the data to be the most prevalent SNP in that region (seen in 28 out of 1404 individuals, or 1.99%) the 16391 SNP is seen in 17 out of the 1404 individuals, and the 16399 is seen in 13 out of the 1404 individuals. The 16398 and 16400 SNP are each only seen

once in the data set of 1404 individuals (Achilli et al., 2005; Achilli et al., 2004; Coble et al., 2004; Finnila et al., 2001; Herrnstadt et al., 2002; Ingman et al., 2000; Kong et al., 2003; Maca-Meyer et al., 2003; Mishmar et al., 2003; Palanichamy et al., 2004; Thangaraj et al., 2005). If these additional SNP locations are displaying heteroplasmic bases, it might be interfering with the binding of the 16390 extension primer. If the primer binds upstream or downstream of the 16390 target, then what might be occurring is SNP detection of another location apart from the 16390 SNP, in which the C/T mixture is present in the mtDNA genome. However, the binding of the ddNTP base at the new position would not alter the overall length of the extension primer to allow the longer C peak at ~39 bp to appear, so additional bp addition still might have occurred in the sample. It is reasonable to assume that either of these scenarios is occurring at the 16390 extension primer and to continue to produce a list of additional explanations would only be speculative at best.

The true question behind the instance of heteroplasmy in the 342 cells is not how it could be occurring, but if it was true heteroplasmy in light of the gross tissue results. Again, this was seen to be a function of the relative concentration of the dominant peak signal in the gross tissue, which possibly overwhelmed the minor signals when compounded with the signals from all of the other cells tested in the gross tissue. This dominant signal was diminished when tested in the single cell, allowing the minor signals to be detected above the negative control threshold. Whether or not these additional signals are true peaks remains a question, which leads to an inconclusive finding for the analysis of the 342 cells. Without a solid explanation of the dual C peaks in the 342 cells, the data is inconclusive and cannot be reliably interpreted. The fact that multiple peaks were detected at random throughout the 342 cells is important to note, and does indicate that more than just a static homoplastic C peak is present in the cells.

Implications of Results

Sources of Error & Countermeasures

Several sources of error were identified throughout the course of the study, and steps were taken to reduce the overall impact of these possible errors. The stringent analytical guidelines applied to the extraction negative controls and the amplification negative controls, which respectively established the viable test samples for the study, were critical for the conservative analysis of the data. The addition of the cycling amplification negative (CAN) was important to establish a separate control for the increased amplification cycles of the samples, although it was not used for any of the cells tested in the study.

It was important from the beginning to establish the proper baseline for heteroplasmy in the gross tissue samples. The multiple amplifications and analysis of the gross tissue samples throughout the validation and length study was necessary to establish a redundant set of data that controlled for the instances of heteroplasmy. When heteroplasmy was seen, it confirmed and identified the presence of heteroplasmic tissues before the cells were isolated and tested. While the 261 and 342 tissues and the 288 and 329 controls were an exception to this, it was seen that the data produced from single cells throughout these four tissues generated interesting results. This data showed that the prevalence of multiple bases at the cellular level might be more widespread than what was observed in the apparent homoplastic gross tissue results.

When working with the gross tissue, it was important to have the visual confirmation of both the pathology and cell origin in order to confidently demonstrate a uniform sample set, which then served to standardize the sample set for the rest of the study. The removal of the samples that contained questionable pathology was critical to the study, in that cancerous or diseased tissues could have contained artificially high levels of heteroplasmy due to mtDNA

mutations and degradation, which would have skewed the results of the study toward a higher percentage of heteroplasmy in the cells.

The ultimate source of error was determined to be human error when handling the tissues and cells. This was especially relevant in the process of laser-dissection of the single cells from the microscope slides, which required strict adherence to the collection guidelines in order to guard against the possibility of multiple cells entering the capture tube. While every step was taken to ensure that these guidelines were followed during the course of collecting over seven-hundred cells, it is reasonable to expect that at even a 1% error rate as many as seven capture tubes might have contained multiple cells. At a 10% error rate, as many as seventy capture tubes might have contained multiple cells. This rate is truly unknown, due to the limitation that no visual confirmation could be performed to check the efficiency of the cell capture. The indirect inference as to the success of the cell capture came only during the analysis of the SNaPshot™ data, in which successful data indicated a successful cell capture. This not only related to the ability to discern between a single cell or multiple cell captures, but the ability to detect if a cell was even captured at all. It is assumed that the negative cell samples throughout the SNaPshot™ data were indicative of the lack of a successfully captured cell. However, it is possible that the amplification of the mtDNA from a single cell was below both the detectable threshold of the Agilent 2100 Bioanalyzer as well as the optimal input concentration for successful SNaPshot™ analysis, which led to the negative results for these samples.

Sensitivity of detection systems

A limitation of the study was the inability to reliably detect the concentration of the amplified mtDNA from the single cells using the Agilent 2100 Bioanalyzer, which was found to be critical for the optimal downstream use of the SNaPshot™ SNP testing kit. The lack of

detection of A peaks in the single cell data is most likely tied to the design of the extension primers (containing 5′ poly-thymine modifications) and to an equal degree, the mtDNA concentration present in the single cell samples. The primary one-step amplification was successful in amplifying the mtDNA from a single cell, but not in high enough quantities to be readily detected on the Agilent 2100 Bioanalyzer. This led to a non-optimal input volume of the amplified mtDNA from the single cells into the SNaPshot™ assay, in which the non-ideal mtDNA concentration was compounded with the problems of the poly-thymine modified extension primers. This non-optimal mtDNA concentration may have resulted in the non-addition of the ddATP base, which affected the final outcome of the SNP data by not properly displaying the true heteroplasmy within the cell. Alternately, this non-optimal mtDNA concentration may have caused a random addition of another ddNTP base, which would then have led to non-specific results displaying the possible false presence of heteroplasmy in the affected cells.

Future Outcomes

Based on what was learned from the study, several factors would need further consideration and development if a similar project was to proceed in the future. The appropriate design of the extension primers would be a critical step in the design of a future study. The elimination of the poly-thymine tails would ideally allow the addition of the A bases to the SNP locations of interest, and larger length differences would ideally be employed to allow better separation of the SNaPshot™ peak data. To better establish the extension primer expected results, the commercially available Primer Focus kit would be ideally employed in a future study, or something similar to the length study would need to be executed using the individual

extension primers to determine the possible outcomes of each extension primer prior to critical sample analysis.

In terms of the amplification strategy in the future, the similar validation for cycle number and melting temperature efficiency would need to be carried out with the new primer sets. However, creative primer design based on the SNP locations of interest might enable additional "block amplifications" to be performed (similar to the 16324-16390 amplification that was used in this study), which lowered the overall number of necessary amplification primers in the multiplex. In this same vein, it also would be critical to determine if a multiplex is even necessary in the future, in such that the individual primers could be optimized, individually amplified and run in the SNaPshot™ system, and then combined for electrophoresis and analysis. In this scenario, it would require a larger number of critical reagents, but would not require the intensive optimization of a multiplex amplification, which would instead result in a time savings.

In terms of the detection systems, in the future there will remain the challenge of detecting the amplified mtDNA from a single cell. A system with a higher level of sensitivity would need to be found and validated for use in detecting the amplified mtDNA concentration from the single cells, so that an ideal mtDNA concentration could be measured and applied to the SNaPshot™ assay. This would hopefully result in a more optimized SNaPshot™ amplification, leading to higher quality results within the single cells. However, it is important to note that recent studies comparing current high-sensitivity/high-throughput platforms have found that the detectable heteroplasmy levels change based on the amplification system or instrument used in the analysis (Bandelt & Salas, 2011; Lutz-Bonengel et al., 2008; Naue et al., 2011). The incorporation of any new detection system in the future would need to be compared to current or

past detection systems using known heteroplasmic samples to find if this same dependent phenomenon is also occurring.

A final change to consider in the future is the use of the capillary electrophoresis versus other instrumentation such as HPLC, MALDI-TOF, or a next-generation sequencing platform. The MALDI-TOF in particular has an inherently higher degree of sensitivity for the detection of the mass differences in the extension primers, but the increase in the sensitivity is also correlated to an increase in the cost of running a sample, such that the terminator ddNTPs required for MALDI-TOF analysis are extremely expensive and would be a limiting factor in the budget of any future study considerations. The recent introduction of the next-generation, solid-phase sequencing platforms are also expensive, but the apparent ability of these instruments to resolve mixtures in mtDNA samples, as well as allowing for the simultaneous detection of the amplified sample on a traditional capillary electrophoresis instrument are seen as positive potentials of this new technology (Calloway et al., 2012; Dixon, 2011). The major difference between the use of the next-generation sequencing and traditional Sanger sequencing, however, is the application of "similarity scores" in the analysis of the next-generation sequencing data rather than actual sequence analysis to identify variants in the nDNA/mtDNA samples. While this apparently works with a highly reproducible success rate, it would require an entirely new approach to the detection and interpretation of heteroplasmic positions in the mtDNA genome (Homer et al., 2009).

If traditional capillary electrophoresis is to continue to be employed in the future, the proposed optimization and changes to the extension primers would be essential, and should result in a wider separation of the electrophoretic peak data. The outcome of this simple change,

utilizing the current technology, would ideally be a clearer and more concise SNaPshot™ data set for analysis and interpretation of the sequencing results.

Final Conclusions

It is believed based on the analysis of the data that heteroplasmy was found in some of the single cell samples. However, the data also demonstrated that the detectable presence of the heteroplasmy was not uniform since it was not seen in every individual cell. Many of the single-cells sets contained inconclusive data, but overall the data displayed possible conditions in the cells in which some form of heteroplasmy is occurring within the individual cells for all tissues.

To control for this analysis of heteroplasmy, the "non-heteroplasmic" 288 and 329 cells were used. Whether or not the presence of heteroplasmy in the control cell SNaPshot™ results invalidates the control samples, and therefore calls into question the validity of the related sample results, is a matter of debate. The presence of multiple peaks in both the control cells and the in the sample cells is randomly distributed, and not uniform, so contamination should be ruled out for the original ten samples that contained heteroplasmy in the gross tissue. It is unclear if the results from the additional cells from 261 or 342 were as of the result of contamination or not. If it is therefore assumed that the presence of two or more valid peaks within the test cells are real, then an explanation of this data would be the presence of heteroplasmy in the single cells. The overall concentration of the mtDNA in the sample is correlated to the successful detection of heteroplasmy within the final results of the SNaPshot™ data. It appears that high levels of dominant mtDNA types overwhelmed lower levels of minor types at high mtDNA concentrations, but at lower concentrations the dominant mtDNA was decreased allowing the minor types to be detected if the design of the SNP extension primer allowed for the base addition. This was found for the control cells as well, assuming that the previously established homoplastic samples related to the culmination of the dominant SNP types found for the control tissues. The true heteroplasmic instances uncovered via SNaPshot™

analysis of the control cells were therefore due to the decrease of the dominant mtDNA type and detection of the minor heteroplasmic mtDNA profile within the single cell samples.

Appendices

Appendix A: Oligonucleotide PCR Primer Sequences
(In-house design unless noted otherwise)

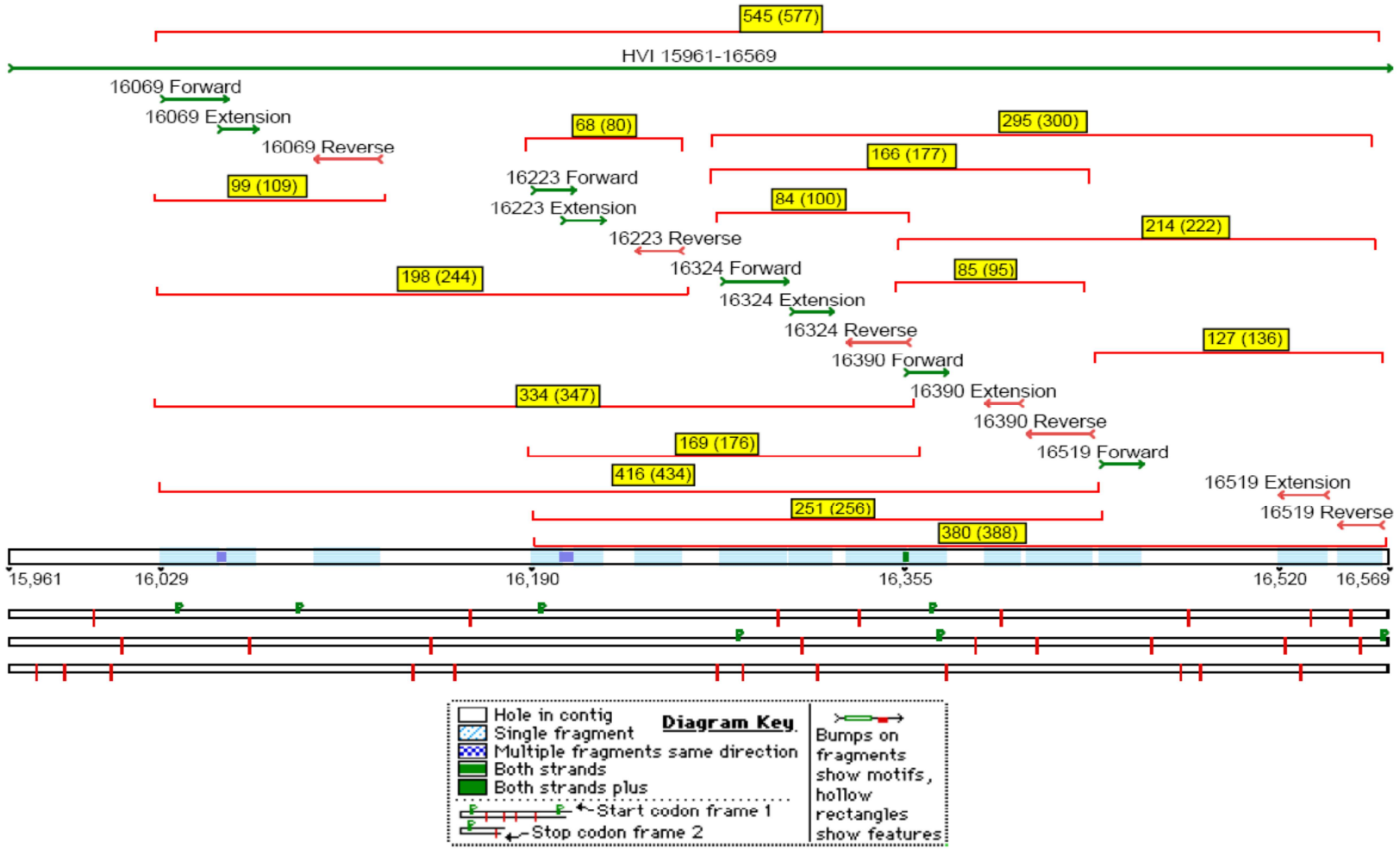
Region	Primer	Nucleotide (base) Sequence	Size (no. of bases)
16069[1]	F16069	5′- ACGTTGGATGGAAGCAGATTTGGGTACCAC-3′	30
	R16069	5′- ACGTTGGATGGTGGCTGGCAGTAATGTACG-3′	30
	E16069	5′- CCACCCAAGTATTGACT-3′	17
16223	F16223	5′-CCCCATGCTTACAAGCAAGTACAGCAATCAA-3′	31
	R16223	5′-GGTGGCTTTGGAGTTGCAGTTGATGTGTGA-3′	30
	E16223	5′- TAGCAAGTACAGCAATCAACC-3′	21
16324	F16324	5′- ACGTTGGATGAACCTACCCACCCTTAACAG-3′	30
	R16324	5′- AAGGGATTTGACTGTAATGTGCTATGTA-3′	28
	E16324	5′- TTTTTTACATAGTACATAAAGCCAT-3′	25
16390	F16390	5′- CTTCTCGTCCCCATGGATGA-3′	20
	R16390	5′- ACGTTGGATGGCGGGATATTGATTTCACGG-3′	30
	E16390	5′- TTTTTTTTTTTTGATGGTGGTCAAGGGAC-3′	29
16519[2]	F16519	5′- TACTCTCCTCGCTCCGGGCCCATAACA-3′	27
	R16519	5′- CATCGTGATGTCTTATTTAAGGGGAA-3′	26
	E16519	5′- TTTTTTTTTTTTGTGGGCTATTTAGGCTTTATG-3′	33

[1] Xiu-Cheng Fan et al. 2008
[2] Vallone et al. 2004

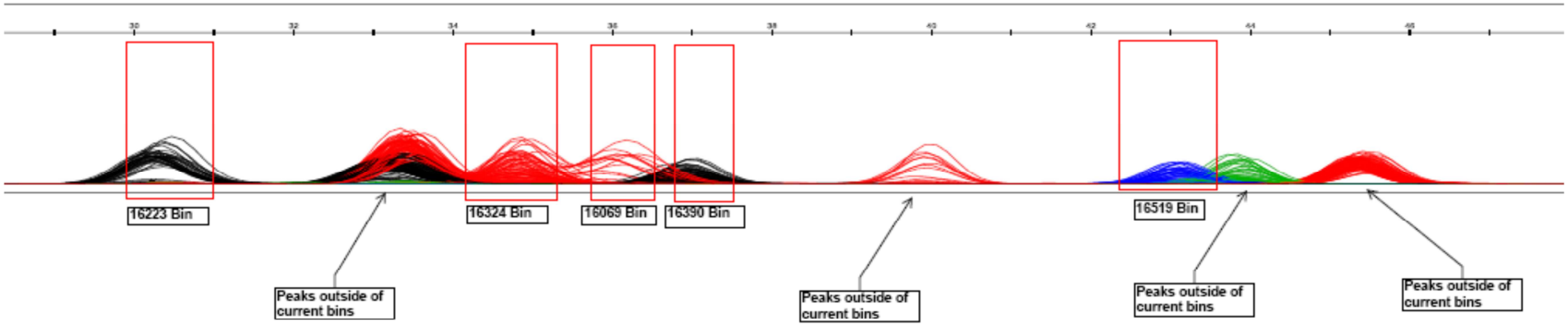
Table Legend-
F##### = Forward Primer
R##### = Reverse Primer
E##### = Extension Primer

Appendix B: SNP Map of HVI Hotspots

Estimated amplicon sizes are the red brackets with yellow numbers.

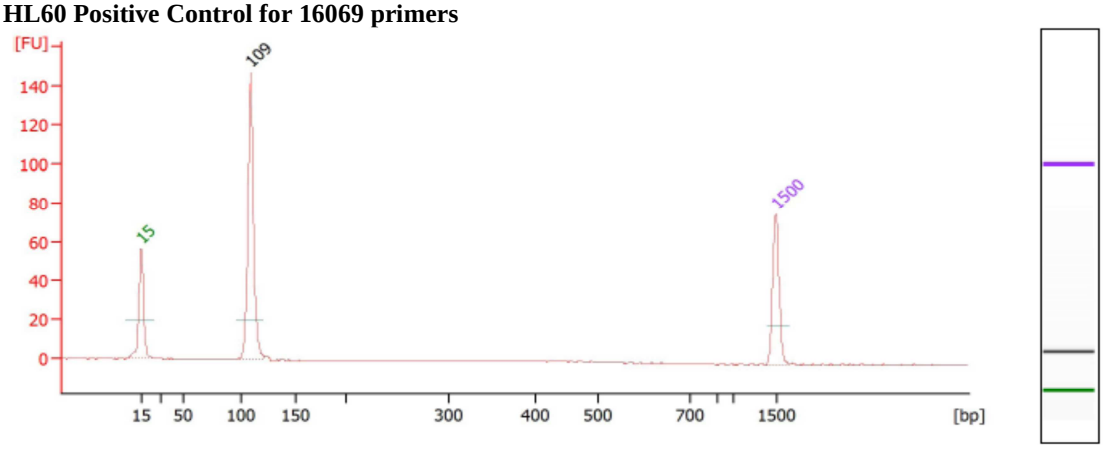


Appendix C: Overlay of SNP peak data from gross tissues



Appendix D: Agilent 2100 Bioanalyzer results for the length study positive and negative controls

HL60 Positive Control for 16069 primers



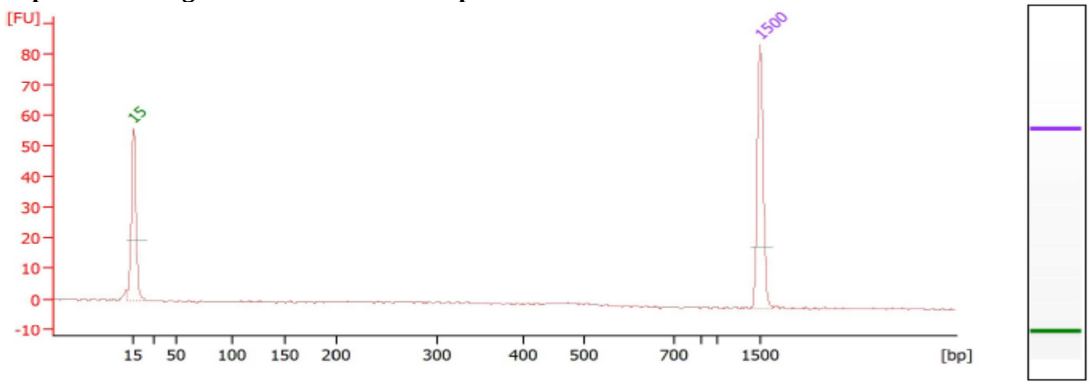
Overall Results for sample 1 : PE-069

Number of peaks found: 1

Peak table for sample 1 : PE-069

Peak	Size [bp]	Conc. [ng/µl]	Molarity [nmol/l]	Observations	Aligned Migration Time [s]	Area	Time corrected area	Peak Height	Peak Width	% of Total
1	15	4.20	424.2	Lower Marker	43.00	35.0	90.1	56.3	3.2	0.0
2	109	8.01	111.3		55.08	104.0	205.0	147.8	3.1	100.0
3	1,500	2.10	2.1	Upper Marker	113.00	64.5	59.7	77.0	2.5	0.0

Amplification Negative Control for 16069 primers



Overall Results for sample 2 : AN-069

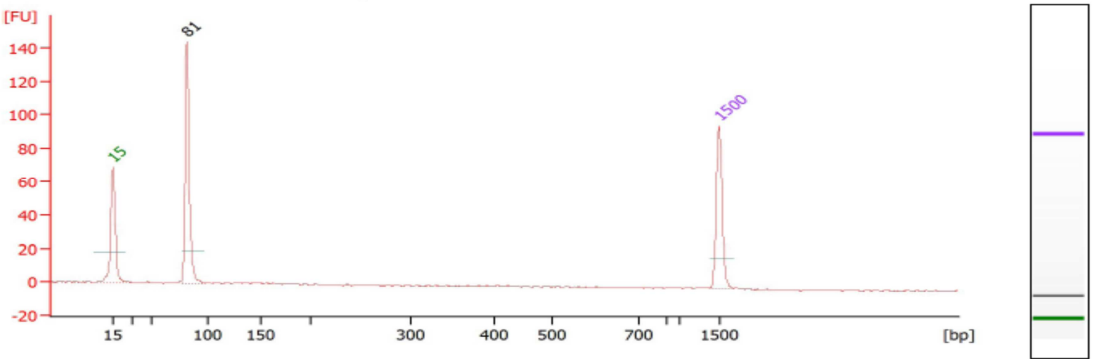
Number of peaks found: 0

Peak table for sample 2 : AN-069

Peak	Size [bp]	Conc. [ng/µl]	Molarity [nmol/l]	Observations	Aligned Migration Time [s]	Area	Time corrected area	Peak Height	Peak Width	% of Total
1	15	4.20	424.2	Lower Marker	43.00	33.5	86.2	55.9	2.2	0.0
2	1,500	2.10	2.1	Upper Marker	113.00	65.2	60.6	86.2	2.5	0.0

Appendix D: Agilent 2100 Bioanalyzer results for the length study positive and negative controls (Continued)

HL60 Positive Control for 16223 primers



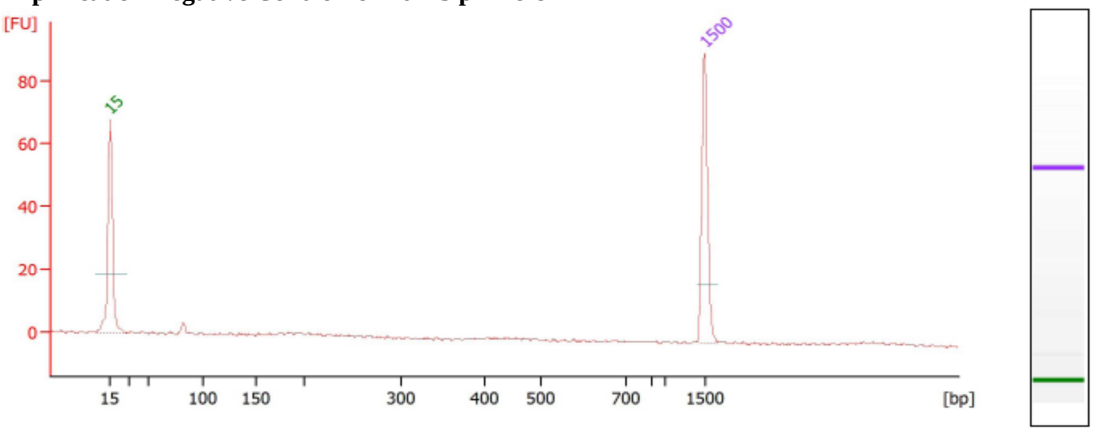
Overall Results for sample 1 : PE-223

Number of peaks found: 1

Peak table for sample 1 : PE-223

Peak	Size [bp]	Conc. [ng/µl]	Molarity [nmol/l]	Observations	Aligned Migration Time [s]	Area	Time corrected area	Peak Height	Peak Width	% of Total
1	15	4.20	424.2	Lower Marker	43.00	42.6	115.4	69.5	3.7	0.0
2	81	5.48	102.6		51.52	75.7	168.4	144.6	2.5	100.0
3	1,500	2.10	2.1	Upper Marker	113.00	74.3	72.2	97.8	2.7	0.0

Amplification Negative Control for 16223 primers



Overall Results for sample 2 : AN-223

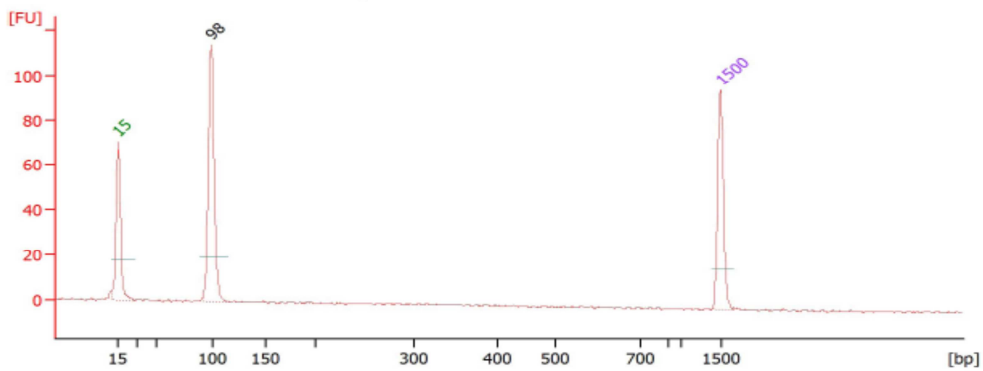
Number of peaks found: 0

Peak table for sample 2 : AN-223

Peak	Size [bp]	Conc. [ng/µl]	Molarity [nmol/l]	Observations	Aligned Migration Time [s]	Area	Time corrected area	Peak Height	Peak Width	% of Total
1	15	4.20	424.2	Lower Marker	43.00	41.3	112.8	67.8	3.8	0.0
2	1,500	2.10	2.1	Upper Marker	113.00	68.7	67.8	92.5	2.5	0.0

Appendix D: Agilent 2100 Bioanalyzer results for the length study positive and negative controls (Continued)

HL60 Positive Control for 16324 primers



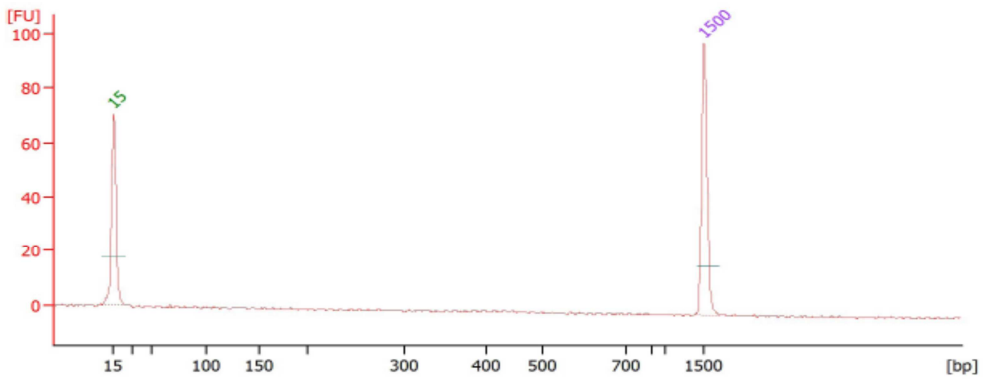
Overall Results for sample 1 : PE-324

Number of peaks found: 1

Peak table for sample 1 : PE-324

Peak	Size [bp]	Conc. [ng/µl]	Molarity [nmol/l]	Observations	Aligned Migration Time [s]	Area	Time corrected area	Peak Height	Peak Width	% of Total
1	15	4.20	424.2	Lower Marker	43.00	41.1	111.1	70.7	2.7	0.0
2	98	5.99	92.5		53.82	85.1	180.7	114.7	3.5	100.0
3	1,500	2.10	2.1	Upper Marker	113.00	72.1	70.2	97.8	2.5	0.0

Amplification Negative Control for 16324 primers



Overall Results for sample 2 : AN-324

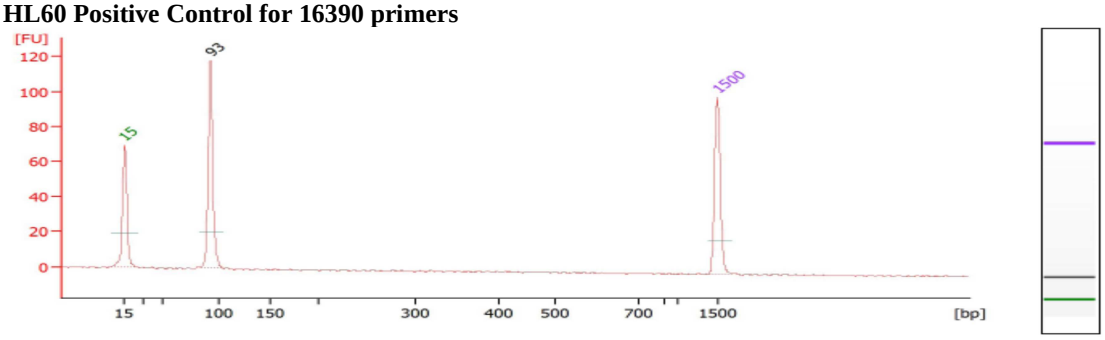
Number of peaks found: 0

Peak table for sample 2 : AN-324

Peak	Size [bp]	Conc. [ng/µl]	Molarity [nmol/l]	Observations	Aligned Migration Time [s]	Area	Time corrected area	Peak Height	Peak Width	% of Total
1	15	4.20	424.2	Lower Marker	43.00	41.4	113.1	70.7	2.7	0.0
2	1,500	2.10	2.1	Upper Marker	113.00	72.7	72.0	100.6	2.8	0.0

Appendix D: Agilent 2100 Bioanalyzer results for the length study positive and negative controls (Continued)

HL60 Positive Control for 16390 primers



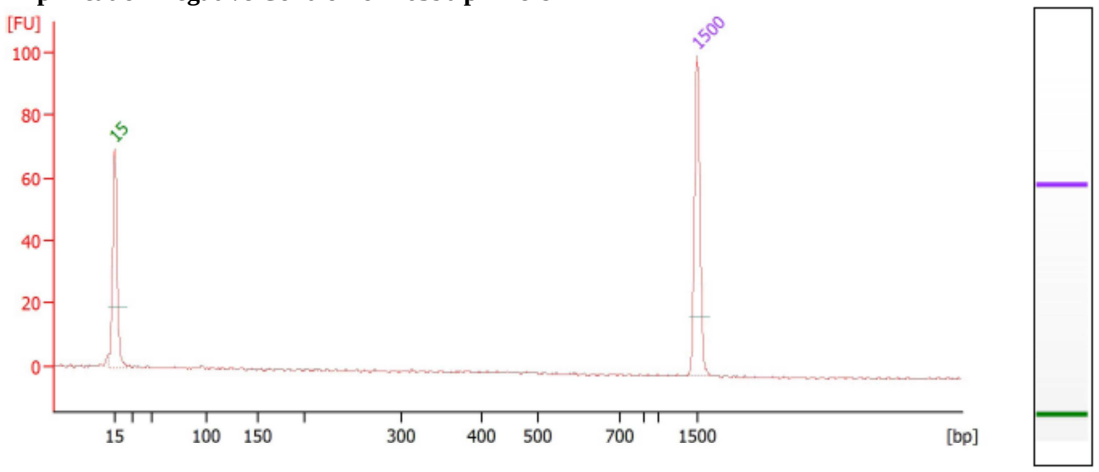
Overall Results for sample 1 : PE-390

Number of peaks found: 1

Peak table for sample 1 : PE-390

Peak	Size [bp]	Conc. [ng/µl]	Molarity [nmol/l]	Observations	Aligned Migration Time [s]	Area	Time corrected area	Peak Height	Peak Width	% of Total
1	15	4.20	424.2	Lower Marker	43.00	40.5	109.6	69.8	3.2	0.0
2	93	4.67	76.2		53.18	67.9	146.4	118.5	2.7	100.0
3	1,500	2.10	2.1	Upper Marker	113.00	75.0	73.8	101.3	2.9	0.0

Amplification Negative Control for 16390 primers



Overall Results for sample 2 : AN-390

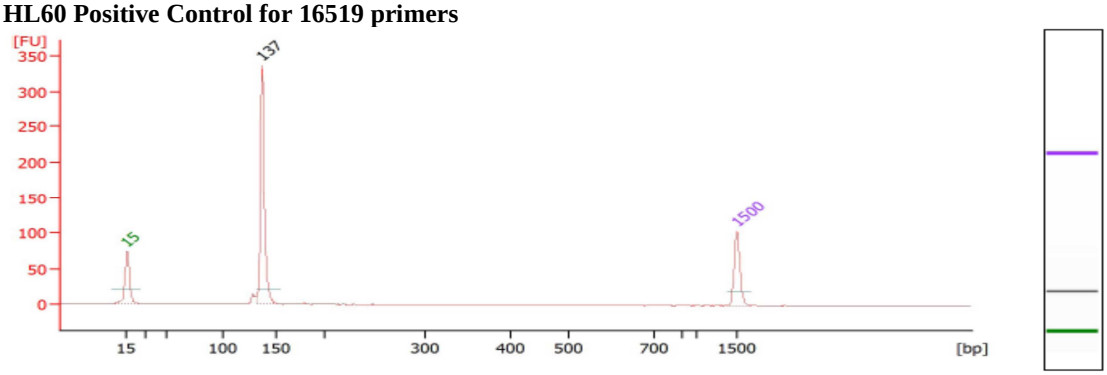
Number of peaks found: 0

Peak table for sample 2 : AN-390

Peak	Size [bp]	Conc. [ng/µl]	Molarity [nmol/l]	Observations	Aligned Migration Time [s]	Area	Time corrected area	Peak Height	Peak Width	% of Total
1	15	4.20	424.2	Lower Marker	43.00	40.2	109.6	69.6	2.2	0.0
2	1,500	2.10	2.1	Upper Marker	113.00	72.7	72.6	102.3	2.7	0.0

Appendix D: Agilent 2100 Bioanalyzer results for the length study positive and negative controls (Continued)

HL60 Positive Control for 16519 primers



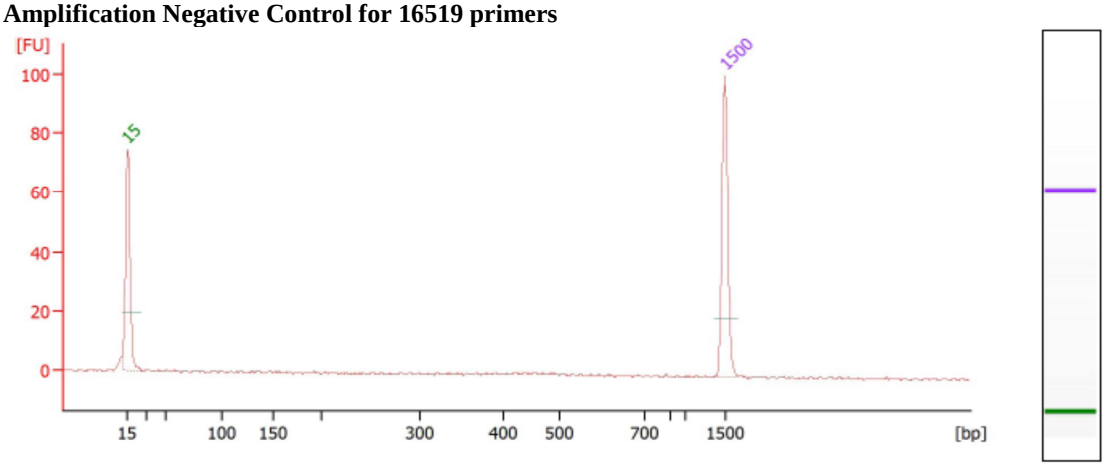
Overall Results for sample 1 : PE-519

Number of peaks found: 1

Peak table for sample 1 : PE-519

Peak	Size [bp]	Conc. [ng/µl]	Molarity [nmol/l]	Observations	Aligned Migration Time [s]	Area	Time corrected area	Peak Height	Peak Width	% of Total
1	15	4.20	424.2	Lower Marker	43.00	45.3	121.2	74.3	3.3	0.0
2	137	10.29	113.5		58.54	168.4	323.6	337.4	2.6	100.0
3	1,500	2.10	2.1	Upper Marker	113.00	77.4	74.7	104.6	2.7	0.0

Amplification Negative Control for 16519 primers



Overall Results for sample 2 : AN-519

Number of peaks found: 0

Peak table for sample 2 : AN-519

Peak	Size [bp]	Conc. [ng/µl]	Molarity [nmol/l]	Observations	Aligned Migration Time [s]	Area	Time corrected area	Peak Height	Peak Width	% of Total
1	15	4.20	424.2	Lower Marker	43.00	43.5	117.1	74.7	2.4	0.0
2	1,500	2.10	2.1	Upper Marker	113.00	75.5	73.8	101.9	2.8	0.0

Appendix E: Detail Table of Heteroplasmic Tissue Samples

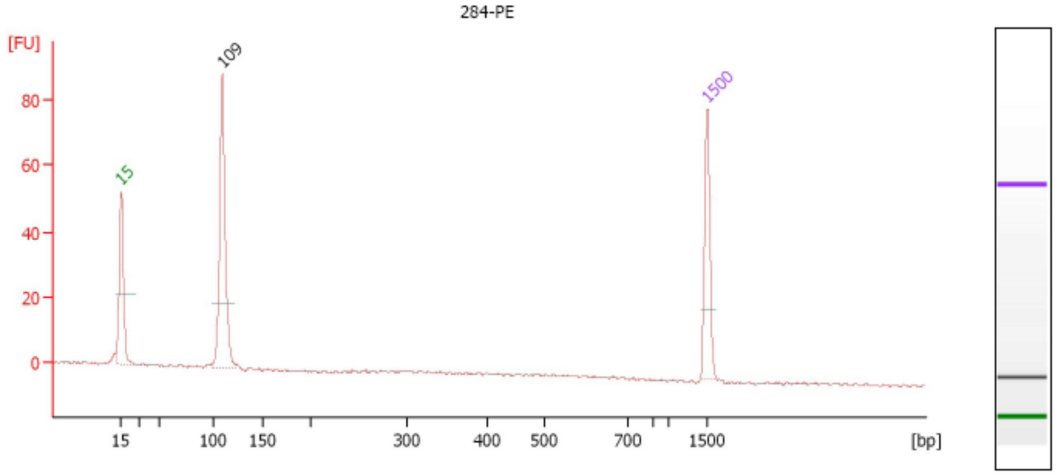
	16069					16223					16324					16390					16519					
Sample	SNP Location	Allele	Size	Height	Peak Area	SNP Location	Allele	Size	Height	Peak Area	SNP Location	Allele	Size	Height	Peak Area	SNP Location	Allele	Size	Height	Peak Area	SNP Location	Allele	Size	Height	Peak Area	Ratio
288	16069	C	33.18	1471	10206	16223	C	30.13	1724	10892	16324	T	34.72	1054	6554	16390	C	37.06	588	3684	16519	A	43.84	1659	10078	n/a
329	16069	C	33.09	1103	7510	16223	C	30.24	1305	8081	16324	T	34.6	386	2392	16390	C	36.93	249	1513	16519	A	43.76	130	799	n/a
261						16223	T	33.37	2570	17023																9.66
							C	30.25	266	1680																
											16324	T	34.47	1003	5752											9.93
												G	31.48	101	775											
																					16519	A	43.78	263	1632	1.34
																						G	42.99	197	1147	
266	16069	C	33.03	8577	83594																					16.69
		T	35.13	514	4809																					
284	16069	T	34.88	8417	78057																					8.56
		A	32.47	983	7159																					
																					16519	G	43.08	8815	79112	19.46
																						C	44.08	453	2915	
289	16069	C	32.7	8370	59965																					34.59
		T	35.28	242	2173																					
											16324	T	34.33	2408	14085											23.38
												G	31.45	103	877											
293						16223	T	33.33	1906	12387																28.45
							C	30.28	67	427																
312						16223	C	29.84	8253	57257																42.54
							T	32.93	194	2197																
314						16223	T	33.34	1924	12253																16.17
							C	30.31	119	751																
315	16069	C	32.27	8553	76400																					12.78
		A	31.51	669	6104																					
320						16223	T	33.23	2285	15257																27.53
							C	30.24	83	483																
326	16069	C	32.71	8493	65060																					16.33
		A	31.66	520	3598																					

Appendix E: Detail Table of Heteroplasmic Tissue Samples (continued)

	16069					16223					16324					16390					16519					
Sample	SNP Location	Allele	Size	Height	Peak Area	SNP Location	Allele	Size	Height	Peak Area	SNP Location	Allele	Size	Height	Peak Area	SNP Location	Allele	Size	Height	Peak Area	SNP Location	Allele	Size	Height	Peak Area	Ratio
338						16223	C	30	6237	40789																22.20
							T	33.08	281	2379																
339																16390	T	39.3	1618	9349						16.85
																	C	38.51	96	656						
345	16069	C	32.57	6982	50076																					7.67
		A	31.52	910	5958																					
						16223	T	34.73	8417	74240																1.25
							C	29.9	6718	44217																
																					16519	A	43.66	4517	30851	19.90
																						G	42.91	227	1441	
346	16069	C	32.48	7615	54509																					30.96
		A	31.45	246	1536																					
349											16324	T	34.48	2205	21568											21.83
												A	33.2	101	862											
351	16069	C	32.22	8573	76401																					33.23
		T	34.88	258	2659																					
						16223	C	29.9	6845	44491																32.75
							T	33.03	209	2127																
352						16223	C	29.92	5410	35312																51.52
							T	32.96	105	1280																
354	16069	C	32.14	8010	60677																					19.93
		A	31.28	402	2478																					
355																16390	C	36.94	60	506						1.13
																	T	39.82	53	262						
																					16519	G	42.99	564	3335	13.43
																						A	43.79	42	275	
369																					16519	G	43.14	146	892	3.74
																						A	43.94	39	265	

Appendix F: Agilent 2100 Bioanalyzer results for the one-step amplification controls

Positive Control for 16069 primers, Sample 284 Cells



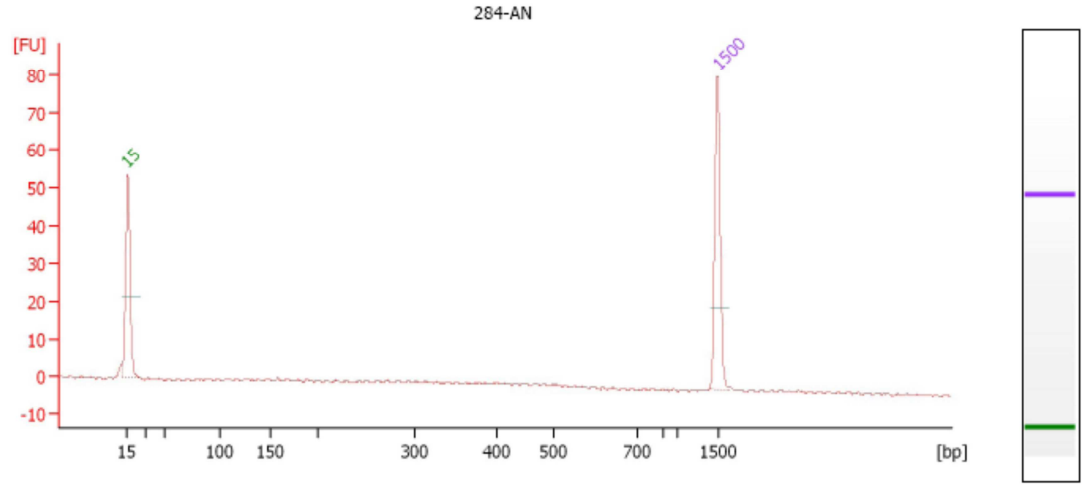
Overall Results for sample 1 : 284-PE

Number of peaks found: 1

Peak table for sample 1 : 284-PE

Peak		Size [bp]	Conc. [ng/µl]	Molarity [nmol/l]	Observations
1	◀	15	4.20	424.2	Lower Marker
2		109	5.52	76.9	
3	▶	1,500	2.10	2.1	Upper Marker

Amplification Negative Control for 16069 primers, Sample 284 Cells



Overall Results for sample 2 : 284-AN

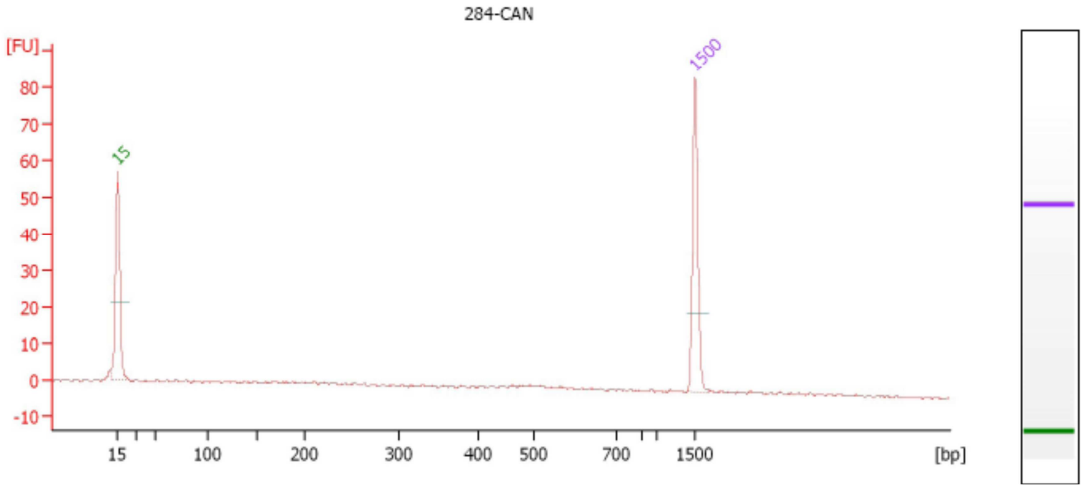
Number of peaks found: 0

Peak table for sample 2 : 284-AN

Peak		Size [bp]	Conc. [ng/µl]	Molarity [nmol/l]	Observations
1	◀	15	4.20	424.2	Lower Marker
2	▶	1,500	2.10	2.1	Upper Marker

Appendix F: Agilent 2100 Bioanalyzer results for the one-step amplification controls (continued)

Cycling Amplification Negative Control, 16069 primers, Sample 284 Cells



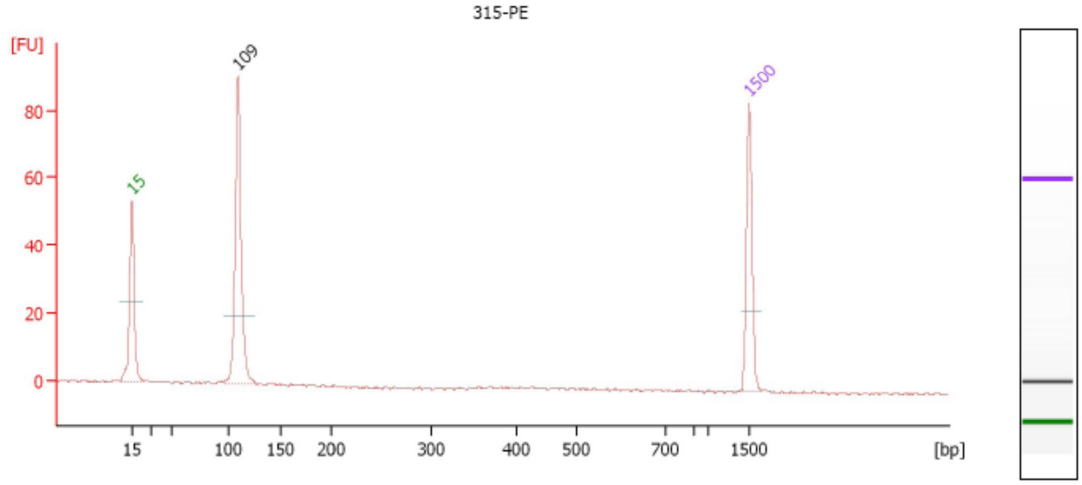
Overall Results for sample 3 : 284-CAN

Number of peaks found: 0

Peak table for sample 3 : 284-CAN

Peak	Size [bp]	Conc. [ng/µl]	Molarity [nmol/l]	Observations
1	15	4.20	424.2	Lower Marker
2	1,500	2.10	2.1	Upper Marker

Positive Control for 16069 primers, Sample 315 Cells



Overall Results for sample 1 : 315-PE

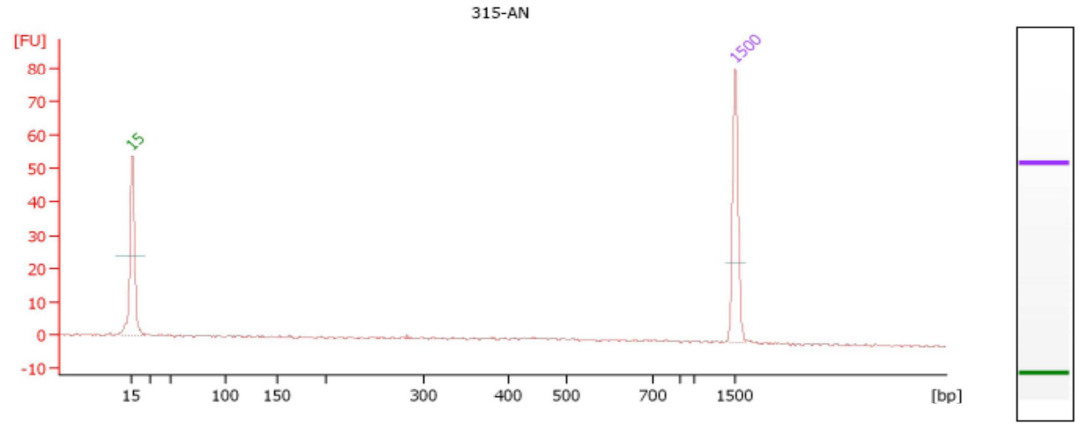
Number of peaks found: 1

Peak table for sample 1 : 315-PE

Peak	Size [bp]	Conc. [ng/µl]	Molarity [nmol/l]	Observations
1	15	4.20	424.2	Lower Marker
2	109	5.46	75.8	
3	1,500	2.10	2.1	Upper Marker

Appendix F: Agilent 2100 Bioanalyzer results for the one-step amplification controls (continued)

Amplification Negative Control for 16069 primers, Sample 315 Cells



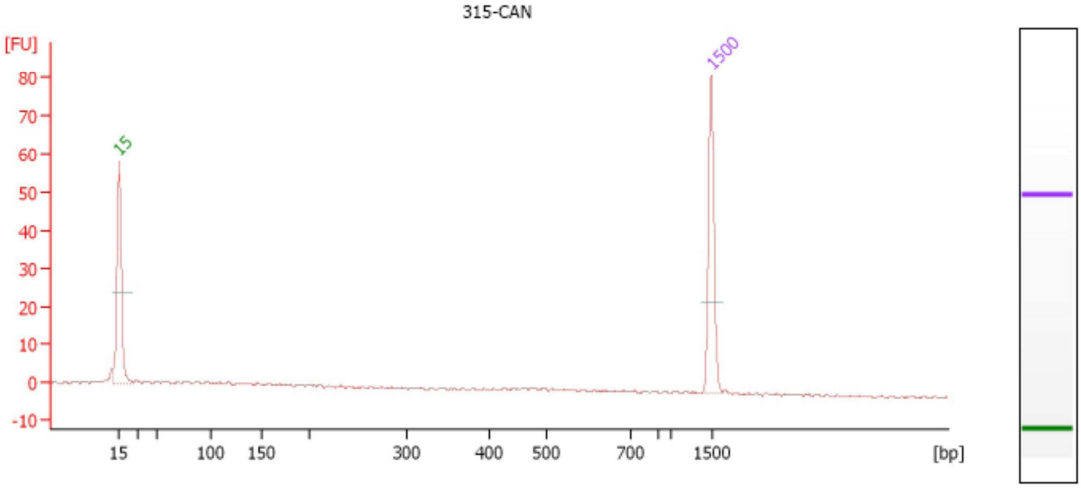
Overall Results for sample 2 : 315-AN

Number of peaks found: 0

Peak table for sample 2 : 315-AN

Peak	Size [bp]	Conc. [ng/µl]	Molarity [nmol/l]	Observations
1	15	4.20	424.2	Lower Marker
2	1,500	2.10	2.1	Upper Marker

Cycling Amplification Negative Control, 16069 primers, Sample 315 Cells



Overall Results for sample 3 : 315-CAN

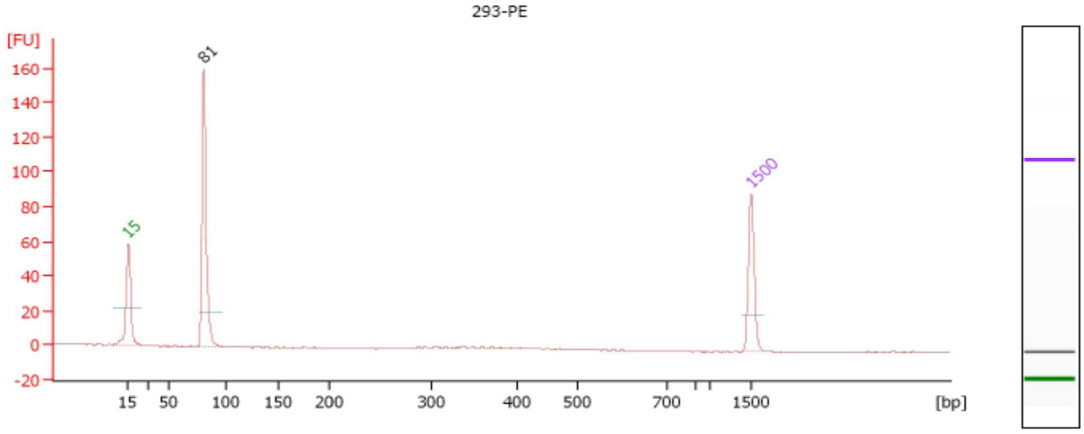
Number of peaks found: 0

Peak table for sample 3 : 315-CAN

Peak	Size [bp]	Conc. [ng/µl]	Molarity [nmol/l]	Observations
1	15	4.20	424.2	Lower Marker
2	1,500	2.10	2.1	Upper Marker

Appendix F: Agilent 2100 Bioanalyzer results for the one-step amplification controls (continued)

Positive Control for 16223 primers, Sample 293 Cells



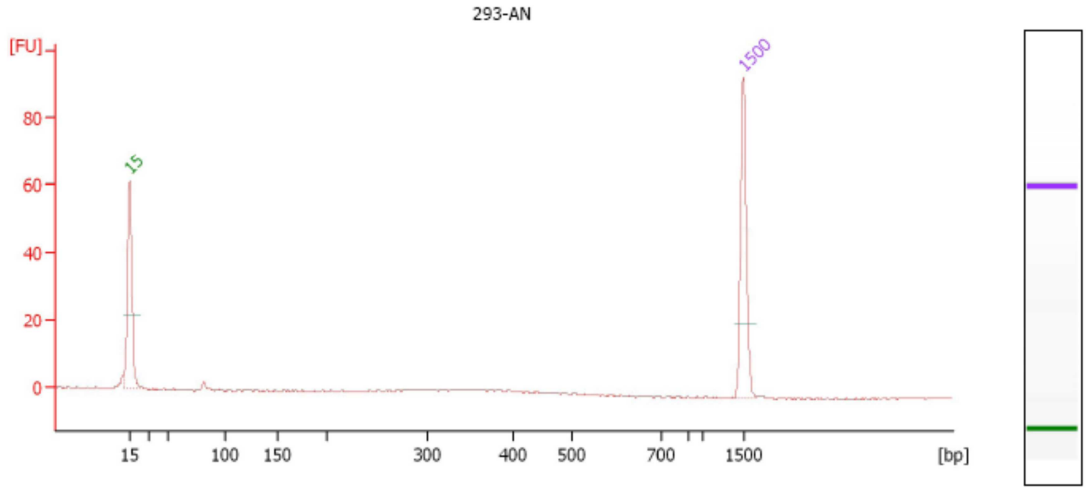
Overall Results for sample 1 : 293-PE

Number of peaks found: 1

Peak table for sample 1 : 293-PE

Peak		Size [bp]	Conc. [ng/µl]	Molarity [nmol/l]	Observations
1	◀	15	4.20	424.2	Lower Marker
2		81	6.56	123.1	
3	▶	1,500	2.10	2.1	Upper Marker

Amplification Negative Control for 16223 primers, Sample 293 Cells



Overall Results for sample 2 : 293-AN

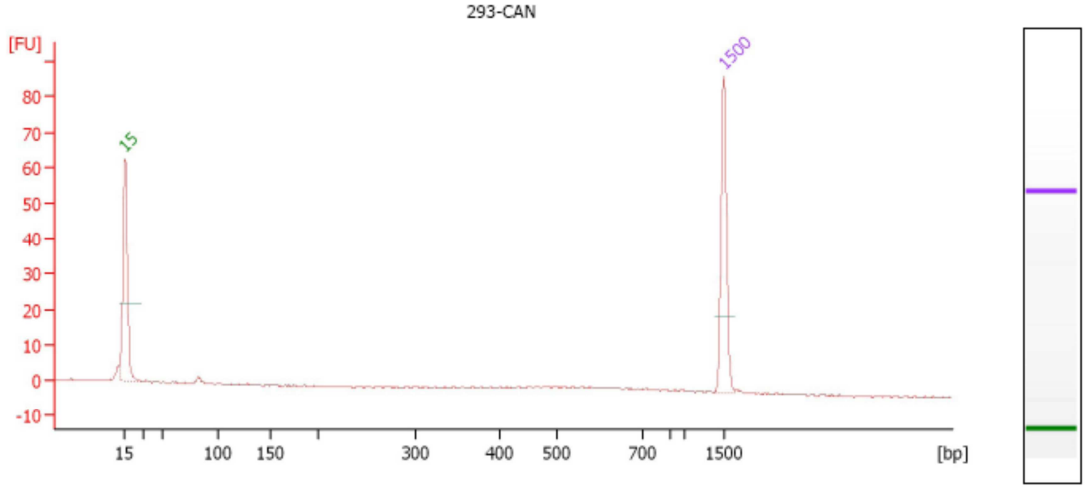
Number of peaks found: 0

Peak table for sample 2 : 293-AN

Peak		Size [bp]	Conc. [ng/µl]	Molarity [nmol/l]	Observations
1	◀	15	4.20	424.2	Lower Marker
2	▶	1,500	2.10	2.1	Upper Marker

Appendix F: Agilent 2100 Bioanalyzer results for the one-step amplification controls (continued)

Cycling Amplification Negative Control, 16223 primers, Sample 293 Cells



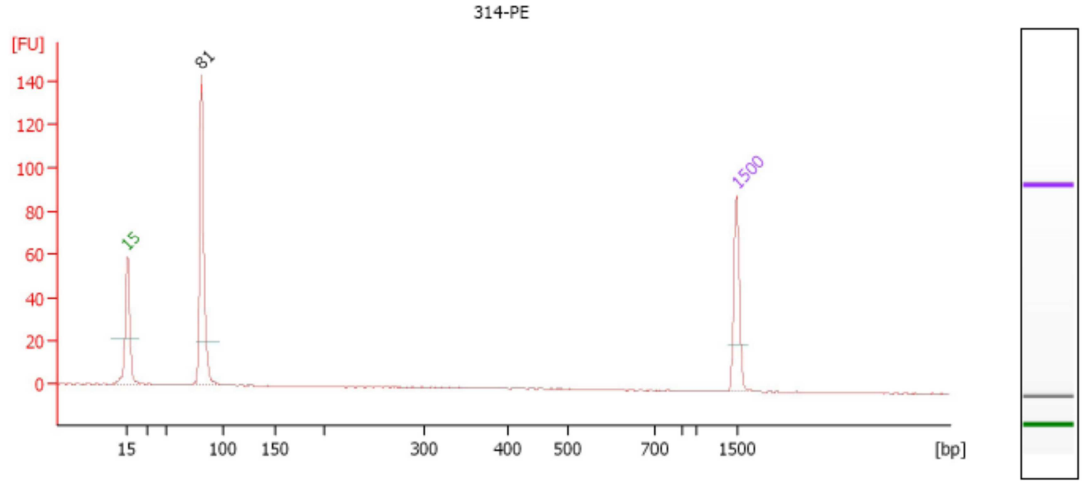
Overall Results for sample 3 : 293-CAN

Number of peaks found: 0

Peak table for sample 3 : 293-CAN

Peak	Size [bp]	Conc. [ng/µl]	Molarity [nmol/l]	Observations
1	15	4.20	424.2	Lower Marker
2	1,500	2.10	2.1	Upper Marker

Positive Control for 16223 primers, Sample 314 Cells



Overall Results for sample 1 : 314-PE

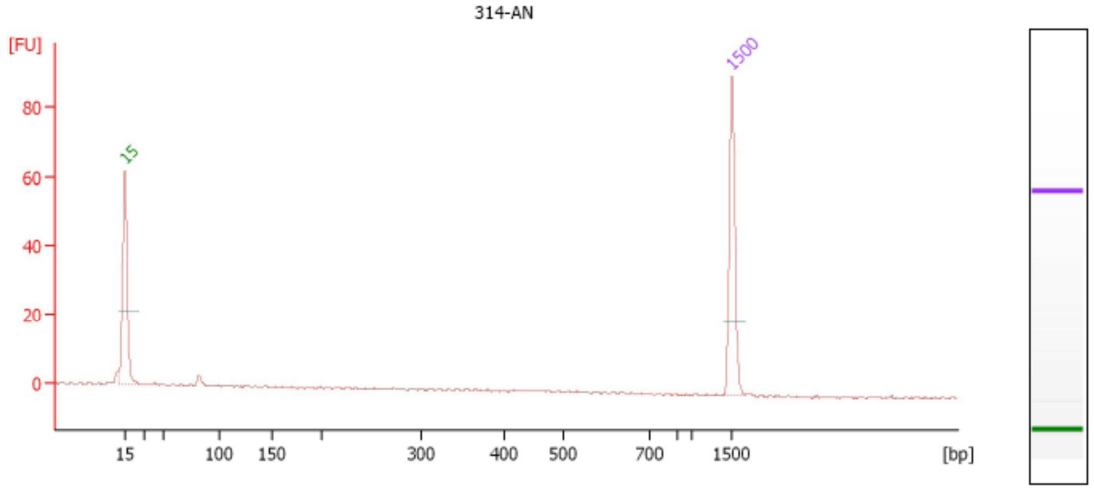
Number of peaks found: 1

Peak table for sample 1 : 314-PE

Peak	Size [bp]	Conc. [ng/µl]	Molarity [nmol/l]	Observations
1	15	4.20	424.2	Lower Marker
2	81	6.41	120.2	
3	1,500	2.10	2.1	Upper Marker

Appendix F: Agilent 2100 Bioanalyzer results for the one-step amplification controls (continued)

Amplification Negative Control for 16223 primers, Sample 314 Cells



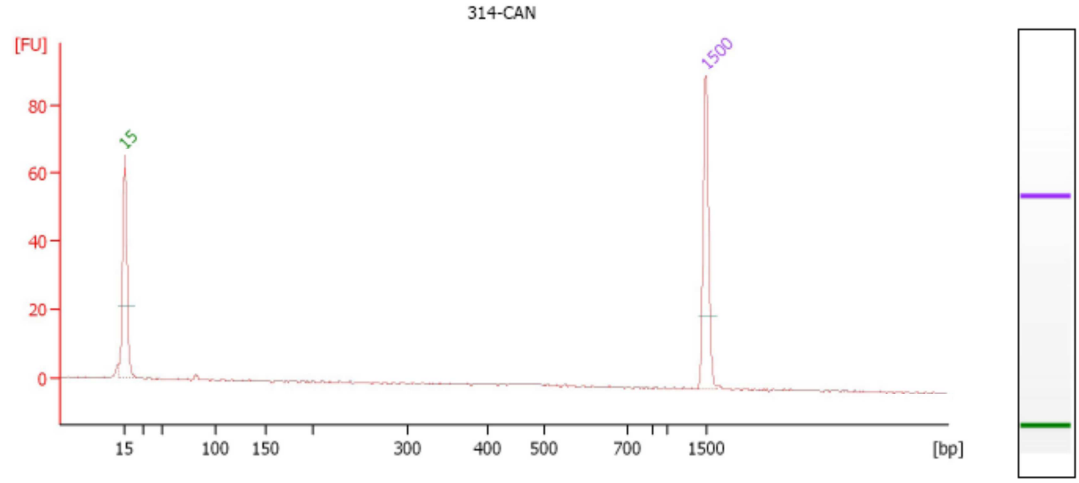
Overall Results for sample 2 : 314-AN

Number of peaks found: 0

Peak table for sample 2 : 314-AN

Peak		Size [bp]	Conc. [ng/µl]	Molarity [nmol/l]	Observations
1	◀	15	4.20	424.2	Lower Marker
2	▶	1,500	2.10	2.1	Upper Marker

Cycling Amplification Negative Control, 16223 primers, Sample 314 Cells



Overall Results for sample 3 : 314-CAN

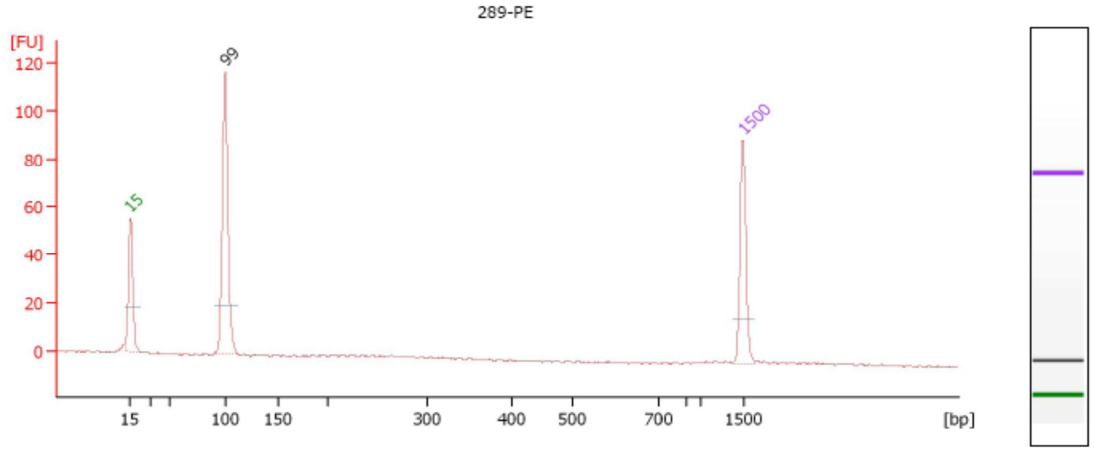
Number of peaks found: 0

Peak table for sample 3 : 314-CAN

Peak		Size [bp]	Conc. [ng/µl]	Molarity [nmol/l]	Observations
1	◀	15	4.20	424.2	Lower Marker
2	▶	1,500	2.10	2.1	Upper Marker

Appendix F: Agilent 2100 Bioanalyzer results for the one-step amplification controls (continued)

Positive Control for 16324 primers, Sample 289 Cells



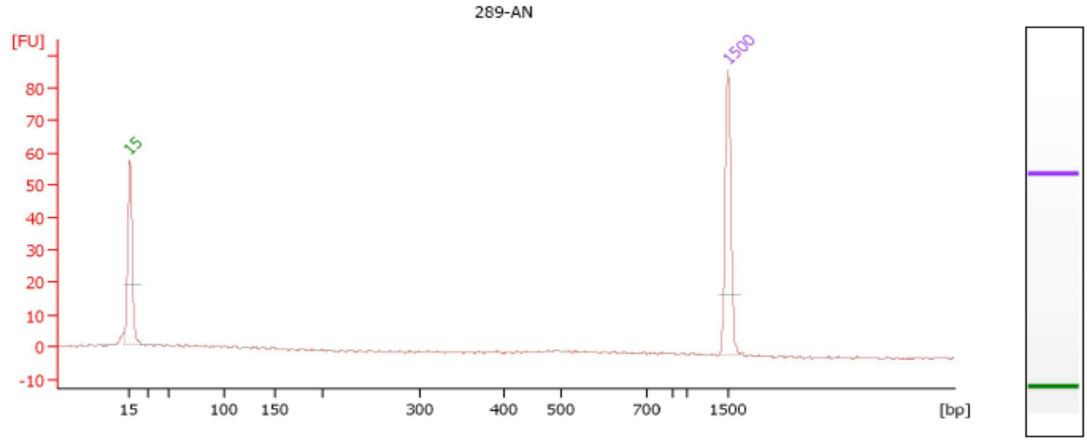
Overall Results for sample 1 : 289-PE

Number of peaks found: 1

Peak table for sample 1 : 289-PE

Peak		Size [bp]	Conc. [ng/µl]	Molarity [nmol/l]	Observations
1	◀	15	4.20	424.2	Lower Marker
2		99	5.86	90.0	
3	▶	1,500	2.10	2.1	Upper Marker

Amplification Negative Control for 16324 primers, Sample 289 Cells



Overall Results for sample 2 : 289-AN

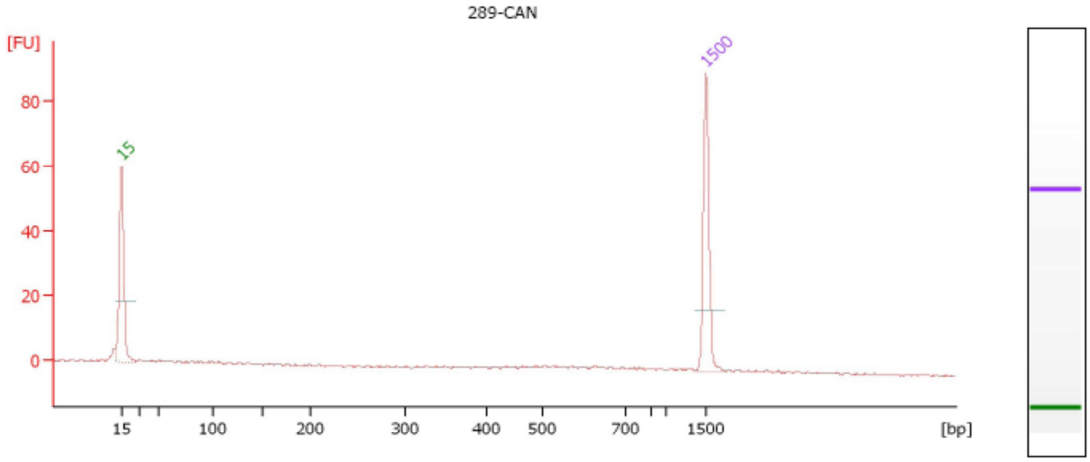
Number of peaks found: 0

Peak table for sample 2 : 289-AN

Peak		Size [bp]	Conc. [ng/µl]	Molarity [nmol/l]	Observations
1	◀	15	4.20	424.2	Lower Marker
2	▶	1,500	2.10	2.1	Upper Marker

Appendix F: Agilent 2100 Bioanalyzer results for the one-step amplification controls (continued)

Cycling Amplification Negative Control, 16324 primers, Sample 289 Cells



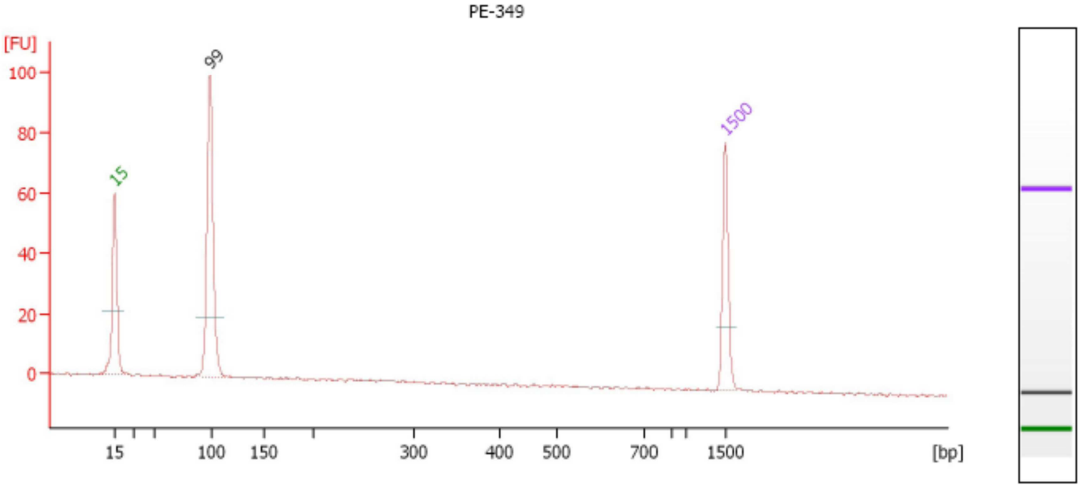
Overall Results for sample 3 : 289-CAN

Number of peaks found: 0

Peak table for sample 3 : 289-CAN

Peak		Size [bp]	Conc. [ng/µl]	Molarity [nmol/l]	Observations
1	◀	15	4.20	424.2	Lower Marker
2	▶	1,500	2.10	2.1	Upper Marker

Positive Control for 16324 primers, Sample 349 Cells



Overall Results for sample 1 : PE-349

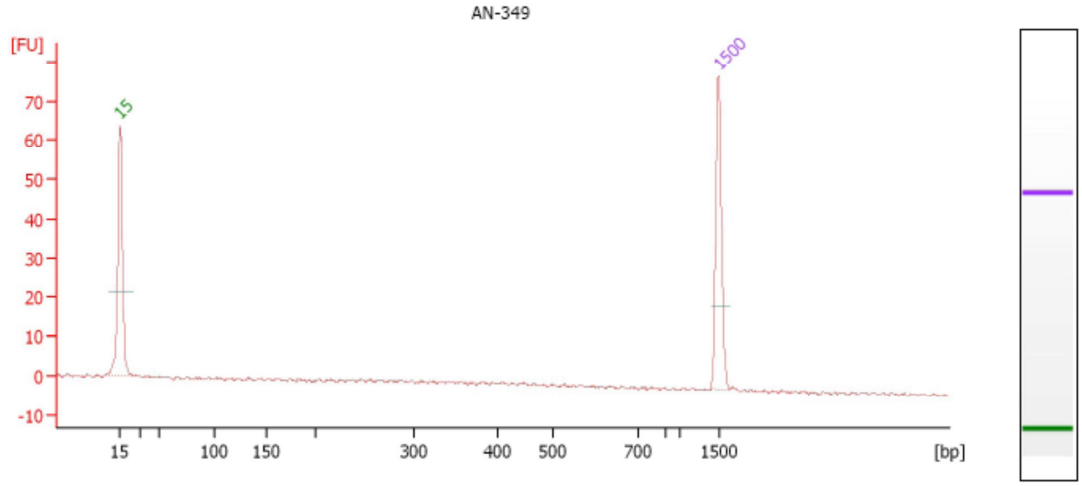
Number of peaks found: 1

Peak table for sample 1 : PE-349

Peak		Size [bp]	Conc. [ng/µl]	Molarity [nmol/l]	Observations
1	◀	15	4.20	424.2	Lower Marker
2		99	6.64	101.8	
3	▶	1,500	2.10	2.1	Upper Marker

Appendix F: Agilent 2100 Bioanalyzer results for the one-step amplification controls (continued)

Amplification Negative Control for 16324 primers, Sample 349 Cells



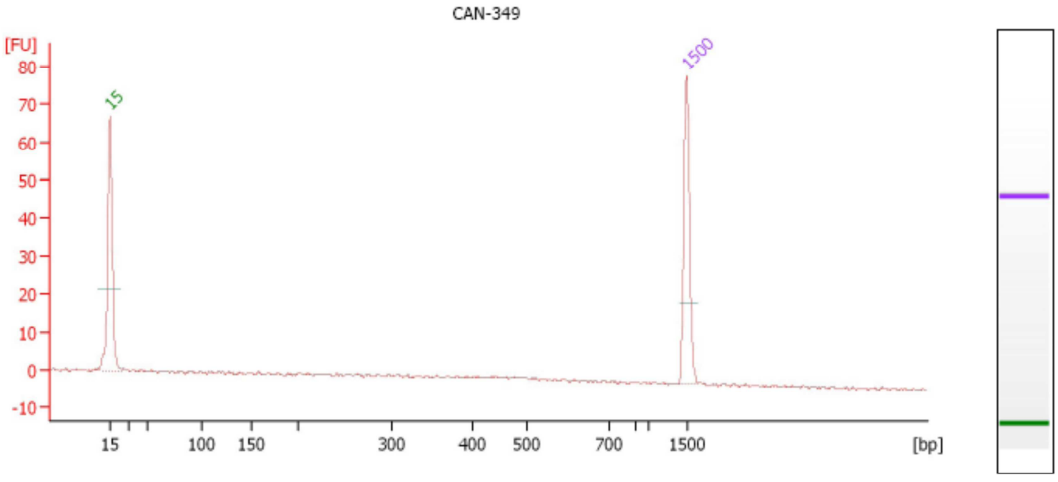
Overall Results for sample 2 : AN-349

Number of peaks found: 0

Peak table for sample 2 : AN-349

Peak	Size [bp]	Conc. [ng/µl]	Molarity [nmol/l]	Observations
1	15	4.20	424.2	Lower Marker
2	1,500	2.10	2.1	Upper Marker

Cycling Amplification Negative Control, 16324 primers, Sample 349 Cells



Overall Results for sample 3 : CAN-349

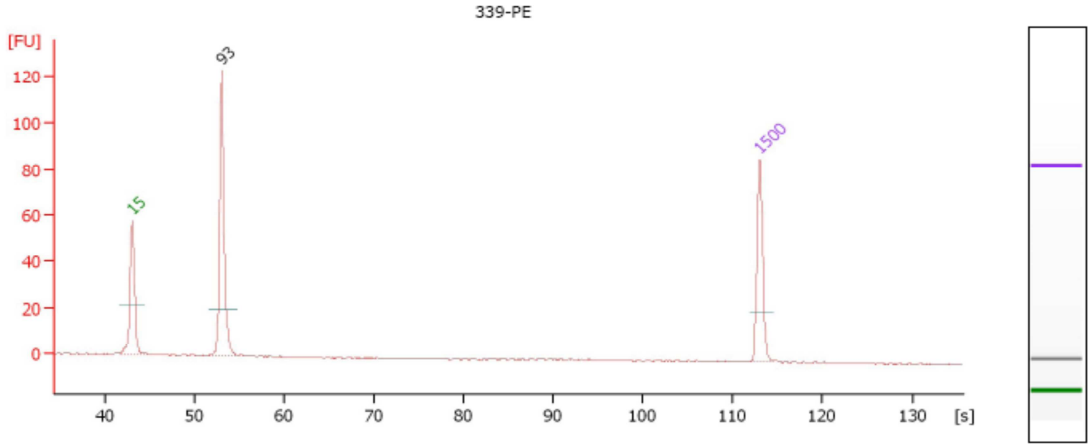
Number of peaks found: 0

Peak table for sample 3 : CAN-349

Peak	Size [bp]	Conc. [ng/µl]	Molarity [nmol/l]	Observations
1	15	4.20	424.2	Lower Marker
2	1,500	2.10	2.1	Upper Marker

Appendix F: Agilent 2100 Bioanalyzer results for the one-step amplification controls (continued)

Positive Control for 16390 primers, Sample 339 Cells



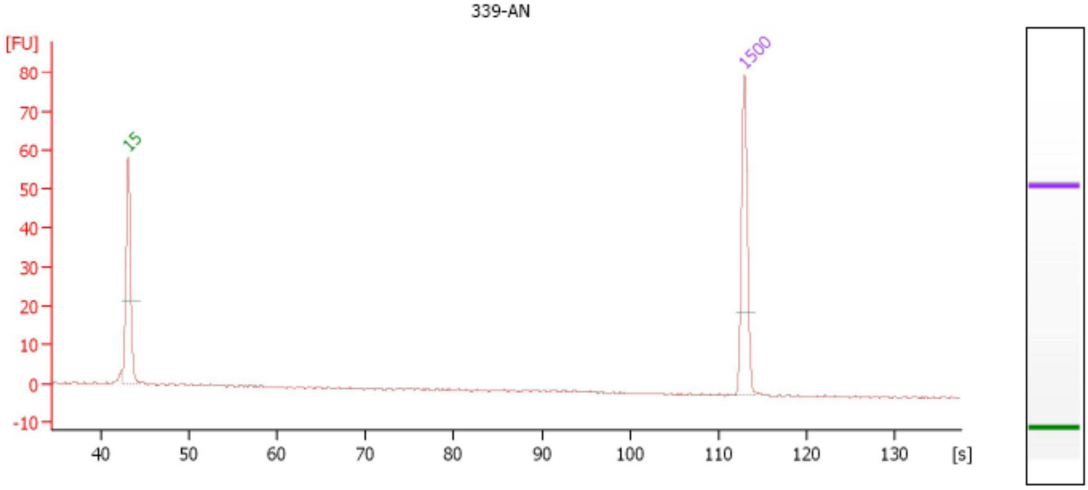
Overall Results for sample 1 : 339-PE

Number of peaks found: 1

Peak table for sample 1 : 339-PE

Peak		Size [bp]	Conc. [ng/µl]	Molarity [nmol/l]	Observations
1	◀	15	4.20	424.2	Lower Marker
2		93	5.51	90.1	
3	▶	1,500	2.10	2.1	Upper Marker

Amplification Negative Control for 16390 primers, Sample 339 Cells



Overall Results for sample 2 : 339-AN

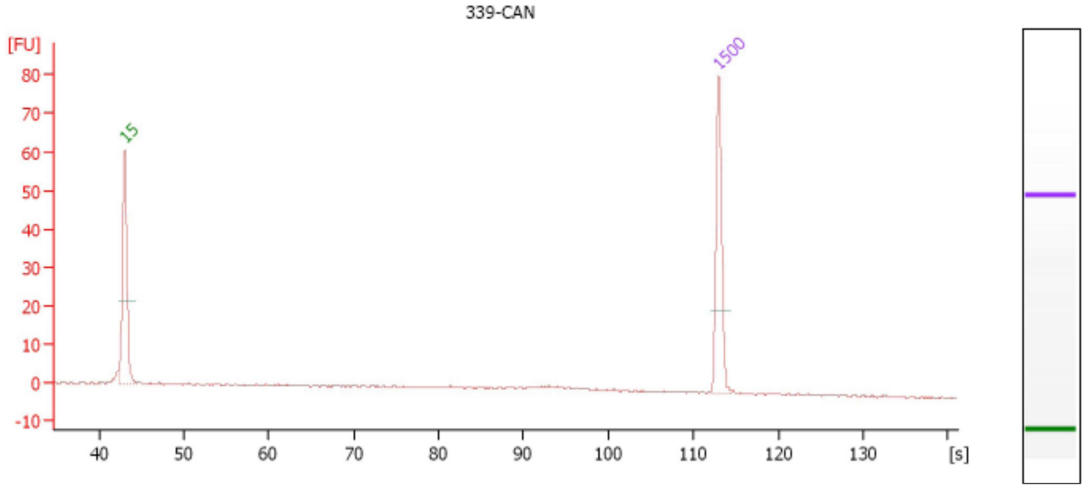
Number of peaks found: 0

Peak table for sample 2 : 339-AN

Peak		Size [bp]	Conc. [ng/µl]	Molarity [nmol/l]	Observations
1	◀	15	4.20	424.2	Lower Marker
2	▶	1,500	2.10	2.1	Upper Marker

Appendix F: Agilent 2100 Bioanalyzer results for the one-step amplification controls (continued)

Cycling Amplification Negative Control, 16390 primers, Sample 339 Cells



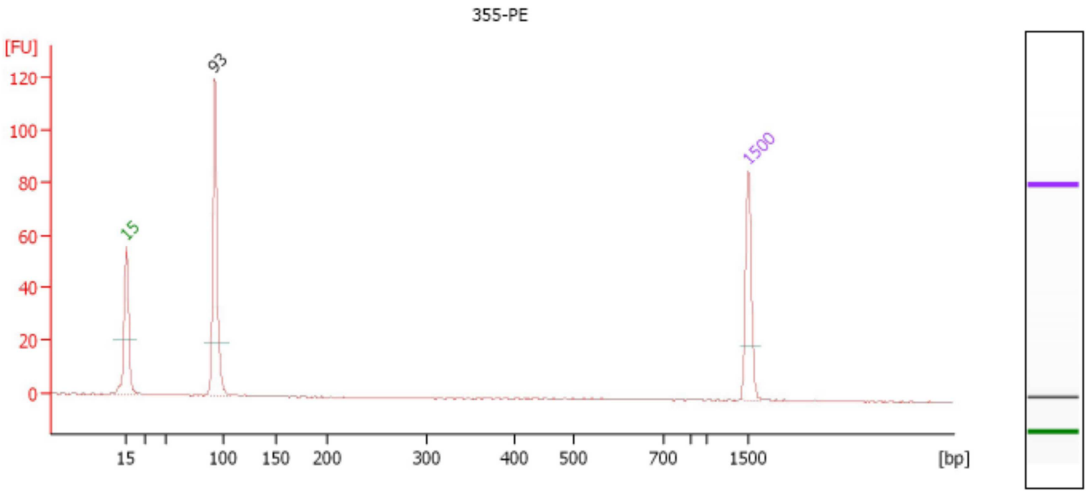
Overall Results for sample 3 : 339-CAN

Number of peaks found: 0

Peak table for sample 3 : 339-CAN

Peak	Size [bp]	Conc. [ng/µl]	Molarity [nmol/l]	Observations
1	15	4.20	424.2	Lower Marker
2	1,500	2.10	2.1	Upper Marker

Positive Control for 16390 primers, Sample 355 Cells



Overall Results for sample 1 : 355-PE

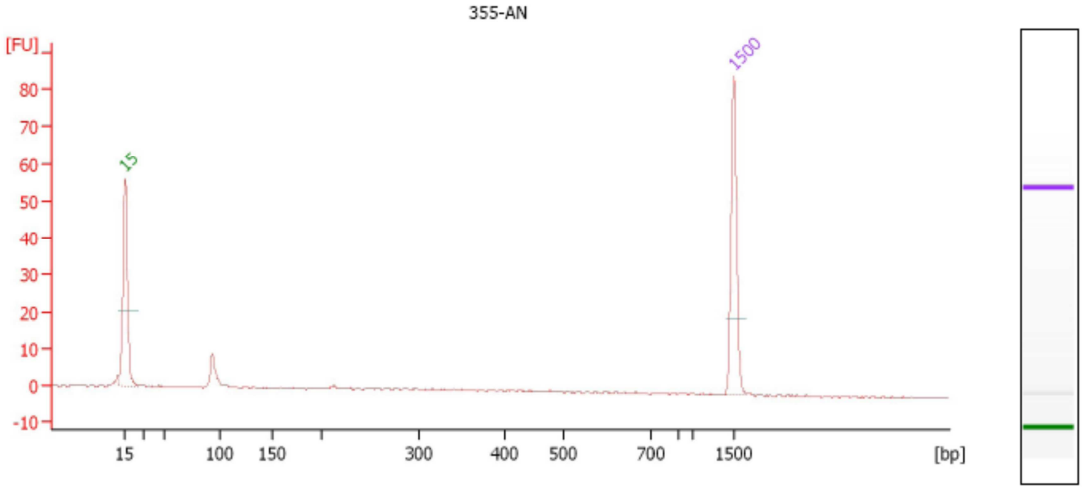
Number of peaks found: 1

Peak table for sample 1 : 355-PE

Peak	Size [bp]	Conc. [ng/µl]	Molarity [nmol/l]	Observations
1	15	4.20	424.2	Lower Marker
2	93	5.11	83.7	
3	1,500	2.10	2.1	Upper Marker

Appendix F: Agilent 2100 Bioanalyzer results for the one-step amplification controls (continued)

Amplification Negative Control for 16390 primers, Sample 355 Cells



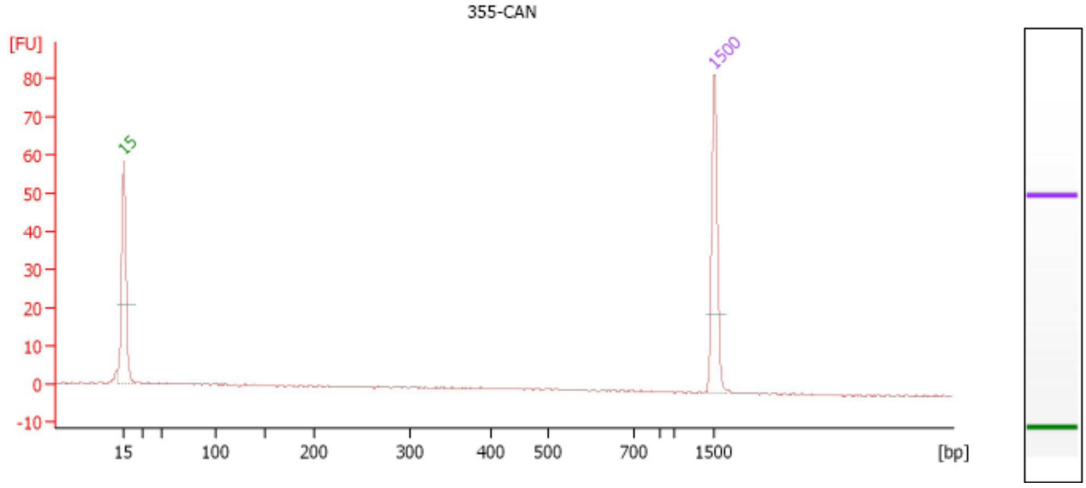
Overall Results for sample 2 : 355-AN

Number of peaks found: 0

Peak table for sample 2 : 355-AN

Peak	Size [bp]	Conc. [ng/µl]	Molarity [nmol/l]	Observations
1	15	4.20	424.2	Lower Marker
2	1,500	2.10	2.1	Upper Marker

Cycling Amplification Negative Control, 16390 primers, Sample 355 Cells



Overall Results for sample 3 : 355-CAN

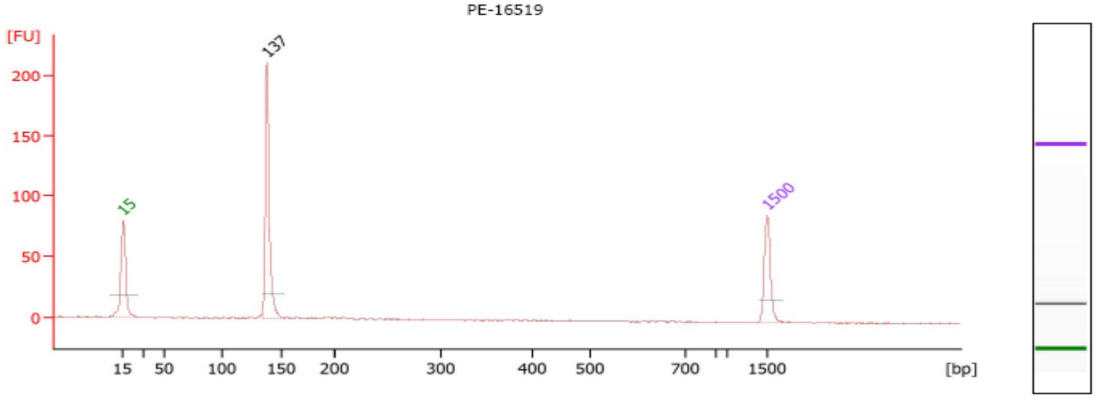
Number of peaks found: 0

Peak table for sample 3 : 355-CAN

Peak	Size [bp]	Conc. [ng/µl]	Molarity [nmol/l]	Observations
1	15	4.20	424.2	Lower Marker
2	1,500	2.10	2.1	Upper Marker

Appendix F: Agilent 2100 Bioanalyzer results for the one-step amplification controls (continued)

Positive Control for 16519 primers, Sample 355 Cells



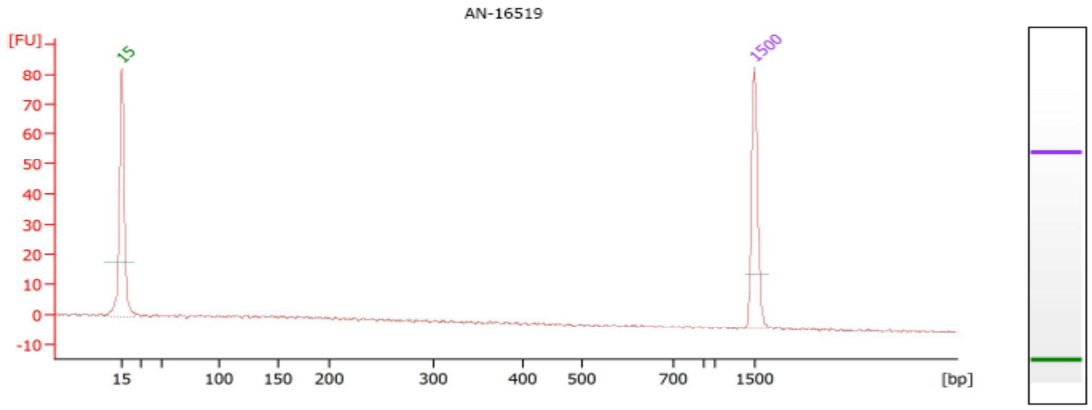
Overall Results for sample 4 : PE-16519

Number of peaks found: 1

Peak table for sample 4 : PE-16519

Peak	Size [bp]	Conc. [ng/µl]	Molarity [nmol/l]	Observations	Aligned Migration Time [s]	Area	Time corrected area	Peak Height	Peak Width	% of Total
1	15	4.20	424.2	Lower Marker	43.00	49.9	132.6	80.3	3.1	0.0
2	137	7.44	82.2		58.63	106.8	200.2	212.3	2.3	100.0
3	1,500	2.10	2.1	Upper Marker	113.00	67.9	62.9	88.1	2.6	0.0

Amplification Negative Control for 16519 primers, Sample 355 Cells



Overall Results for sample 5 : AN-16519

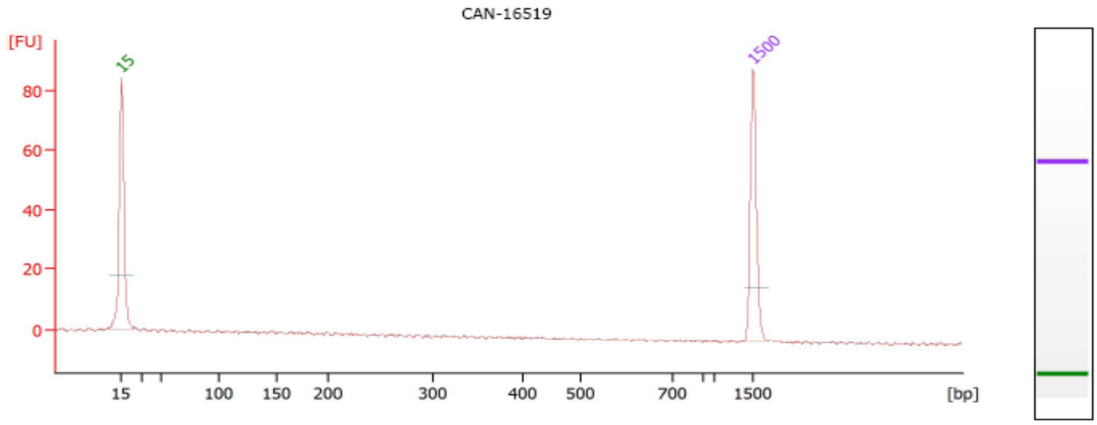
Number of peaks found: 0

Peak table for sample 5 : AN-16519

Peak	Size [bp]	Conc. [ng/µl]	Molarity [nmol/l]	Observations	Aligned Migration Time [s]	Area	Time corrected area	Peak Height	Peak Width	% of Total
1	15	4.20	424.2	Lower Marker	43.00	52.9	141.4	82.9	3.3	0.0
2	1,500	2.10	2.1	Upper Marker	113.00	68.6	64.2	87.3	2.6	0.0

Appendix F: Agilent 2100 Bioanalyzer results for the one-step amplification controls (continued)

Cycling Amplification Negative Control, 16519 primers, Sample 355 Cells



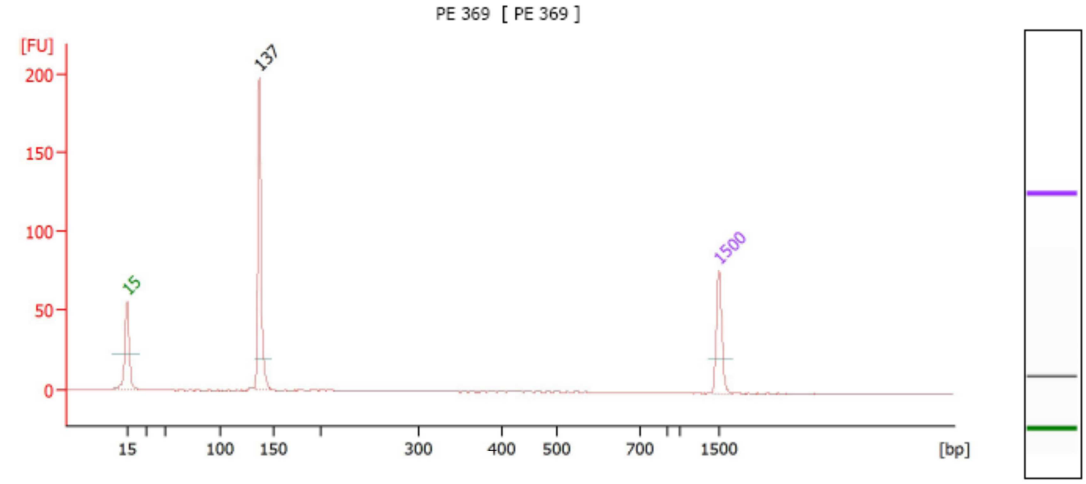
Overall Results for sample 6 : CAN-16519

Number of peaks found: 0

Peak table for sample 6 : CAN-16519

Peak	Size [bp]	Conc. [ng/µl]	Molarity [nmol/l]	Observations	Aligned Migration Time [s]	Area	Time corrected area	Peak Height	Peak Width	% of Total
1	15	4.20	424.2	Lower Marker	43.00	52.4	140.5	84.7	2.6	0.0
2	1,500	2.10	2.1	Upper Marker	113.00	70.1	66.0	91.6	2.6	0.0

Positive Control for 16519 primers, Sample 369 Cells



Overall Results for sample 1 : PE 369

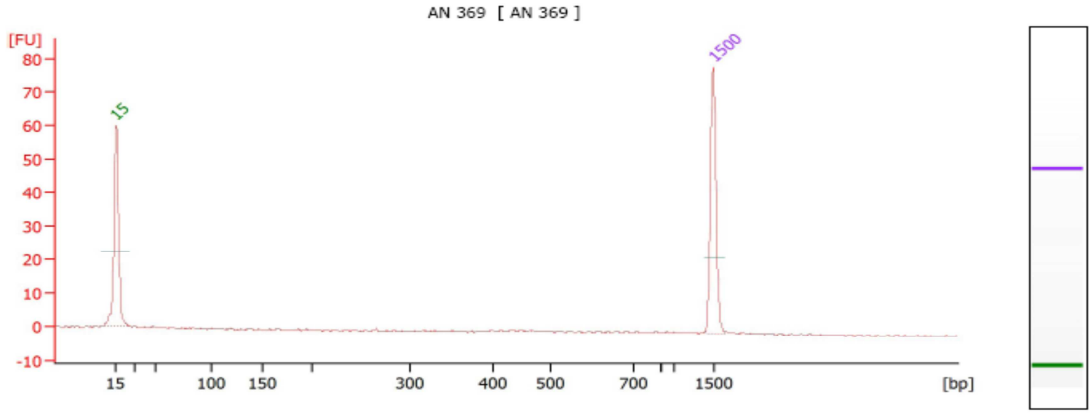
Number of peaks found: 1

Peak table for sample 1 : PE 369

Peak	Size [bp]	Conc. [ng/µl]	Molarity [nmol/l]	Observations	Aligned Migration Time [s]	Area	Time corrected area	Peak Height	Peak Width	% of Total
1	15	4.20	424.2	Lower Marker	43.00	33.1	89.8	55.2	3.3	0.0
2	137	7.29	80.8		58.64	85.9	166.4	198.5	2.1	100.0
3	1,500	2.10	2.1	Upper Marker	113.00	55.8	54.2	77.8	3.0	0.0

Appendix F: Agilent 2100 Bioanalyzer results for the one-step amplification controls (continued)

Amplification Negative Control for 16519 primers, Sample 369 Cells



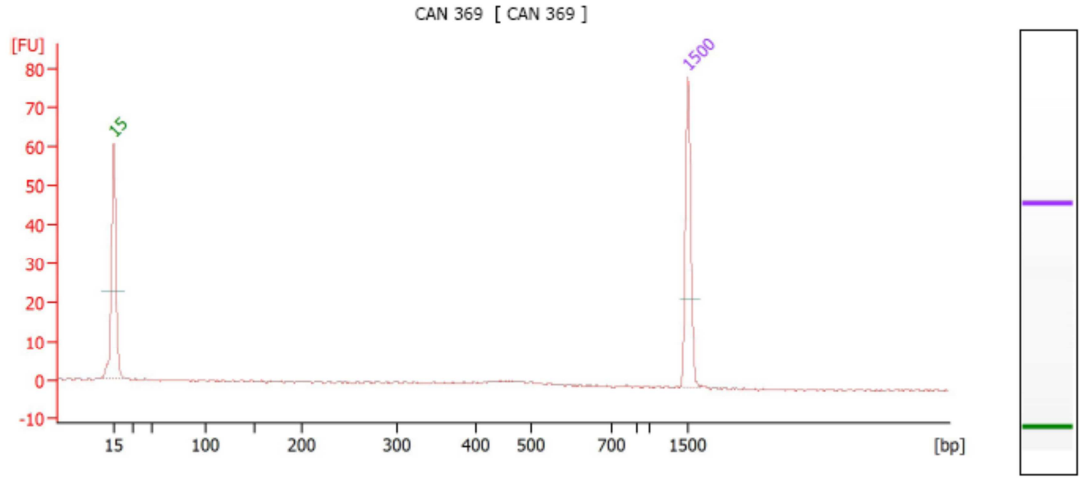
Overall Results for sample 2 : **AN 369**

Number of peaks found: 0

Peak table for sample 2 : **AN 369**

Peak	Size [bp]	Conc. [ng/µl]	Molarity [nmol/l]	Observations	Aligned Migration Time [s]	Area	Time corrected area	Peak Height	Peak Width	% of Total
1	15	4.20	424.2	Lower Marker	43.00	36.1	98.0	59.9	3.4	0.0
2	1,500	2.10	2.1	Upper Marker	113.00	57.7	56.4	79.6	2.5	0.0

Cycling Amplification Negative Control, 16519 primers, Sample 369 Cells



Overall Results for sample 3 : **CAN 369**

Number of peaks found: 0

Peak table for sample 3 : **CAN 369**

Peak	Size [bp]	Conc. [ng/µl]	Molarity [nmol/l]	Observations	Aligned Migration Time [s]	Area	Time corrected area	Peak Height	Peak Width	% of Total
1	15	4.20	424.2	Lower Marker	43.00	35.9	98.7	60.8	2.9	0.0
2	1,500	2.10	2.1	Upper Marker	113.00	56.2	56.1	80.3	2.4	0.0

References

Achilli, A., Rengo, C., Battaglia, V., Pala, M., Olivieri, A., Fornarino, S., et al. (2005). Saami and Berbers--an unexpected mitochondrial DNA link. *Am J Hum Genet, 76*(5), 883-886.

Achilli, A., Rengo, C., Magri, C., Battaglia, V., Olivieri, A., Scozzari, R., et al. (2004). The molecular dissection of mtDNA haplogroup H confirms that the Franco-Cantabrian glacial refuge was a major source for the European gene pool. *Am J Hum Genet, 75*(5), 910-918.

Akane, A., Shiono, H., Matsubara, K., Nakamura, H., Hasegawa, M., & Kagawa, M. (1993). Purification of forensic specimens for the polymerase chain reaction (PCR) analysis. *J Forensic Sci, 38*(3), 691-701.

Allen, M., Engstrom, A.-S., Meyers, S., Handt, O., Saldeen, T., von Haeseler, A., et al. (1998). Mitochondrial DNA sequencing of shed hairs and saliva on robbery caps: sensitivity and matching probabilities. *J Forensic Sci, 43*, 453-464.

Alonso, A., Martin, P., Albarran, C., Garcia, O., & Sancho, M. (1996). Rapid detection of sequence polymorphisms in the human mitochondrial DNA control region by polymerase chain reaction and single-strand conformation analysis in mutation detection enhancement gels. *Electrophoresis, 17*, 1299-1301.

Altschul, S. F., Gish, W., Miller, W., Myers, E. W., & Lipman, D. J. (1990). Basic local alignment serach tool. *J Mol. Biol., 315*(3), 403-410.

Anderson, S., Bankier, A. T., Barrell, B. G., de Bruijn, M. H., Coulson, A. R., Drouin, J., et al. (1981). Sequence and organization of the human mitochondrial genome. *Nature, 290*(5806), 457-465.

Andrews, R. M., Kubacka, I., Chinnery, P. F., Lightowlers, R. N., Turnbull, D. M., & Howell, N. (1999). Reanalysis and revision of the Cambridge reference sequence for human mitochondrial DNA. *Nature Genetics, 23*, 147.

Ankel-Simons, F., & Cummins, Jim M. (1996). Misconceptions about mitochondria and mammalian fertilization: Implications for theories on human evolution. *Proc. Natl. Acad. Sci., 93*(24), 13859-13863.

Applied-Biosystems. (2000). ABI PRISM® SNaPshot™ Multiplex Kit - Protocol. Foster City, CA: Applied Biosystems.

Aquadro, C. F., & Greenberg, B. D. (1983). Human mitochondrial DNA variation and evolution: analysis of nucleotide sequences from seven individuals. *Genetics, 103*(2), 287-312.

Ashley, M. V., Laipis, P. J., & Hauswirth, W. W. (1989). Rapid segregation of heteroplasmic bovine mitochondria. *Nucl. Acids Res., 17*(18), 7325-7331.

Awadalla, P., Eyre-Walker, A., & Smith, J. M. (1999). Linkage disequilibrium and recombination in hominid mitochondrial DNA. *Science, 286*, 2524-2525.

Axler-Diperte, G. (2010). Personal communication with Dr. Grace Axler-Diperte (pp. SNaPshot setup and design).

Baasner, A., Schafer, C., Junge, A., & Madea, B. (1998). Polymorphic sites in human mitochondrial DNA control region sequences: population data and maternal inheritance. *Forensic Sci Int, 98*(3), 169-178.

Bai, Y., Hu, P., Park, J. S., Deng, J. H., Song, X., Chomyn, A., et al. (2004). Genetic and functional analysis of mitochondrial DNA-encoded complex I genes. *Ann. N.Y. Acad. Sci., 1011*, 272-283.

Bai, Y., Shakeley, R. M., & Attardi, G. (2000). Tight control of respiration by NADH dehydrogenase ND5 subunit gene expression in mouse mitochondria. *Molecular and Cellular Biology, 20*(3), 805-815.

Bandelt, H., & Salas, A. (2011). Current Next Generation Sequencing technology may not meet forensic standards. *Forensic Sci. Int. Genet., 6*(1), 143-145.

Barreto, G., Vago, A. R., Ginther, C., Simpson, A. J. G., & Pena, S. D. J. (1996). Mitochondrial D-loop "signatures" produced by low-stringency single specific primer PCR constitute a simple comparative human identity test. *Am J Hum Genet, 58*, 609-616.

Barros, F., Lareu, M. V., Salas, A., & Carracedo, A. (1997). Rapid and enhanced detection of mitochondrial DNA variation using single-strand conformation analysis of superposed restriction enzyme fragments from polymerase chain reaction-amplified products. *Electrophoresis, 18*, 52-54.

Bendall, K. E., Macaulay, V. A., & Sykes, B. C. (1997). Variable levels of a heteroplasmic point mutation in individual hair roots. *Am J Hum Genet, 61*(6), 1303-1308.

Benecke, M. (1997). DNA typing in forensic medicine and in criminal investigations: a current survey. *Naturwissenschaften, 84*, 181-188.

Bereiter-Hahn, J., & Vöth, M. (1994). Dynamics of mitochondria in living cells: shape changes, dislocations, fusion, and fission of mitochondria. *Microsc Res Tech., 27*(3), 198-219.

Bestvater, F., Spiess, E., Stobrawa, G., Hacker, M., Feurer, T., Porwol, T., et al. (2002). Two-photon fluorescence absorption and emission spectra of dyes relevant for cell imaging. *Journal of Microscopy, 208*(2), 108-115.

Bogenhagen, D., & Clayton, D. A. (1974). The number of mitochondrial dexoyribonucleic acid genomes in mouse L and human HeLa cells. Quantitative isolation of mitochondrial deoxyribonucleic acid. *J. Biol. Chem., 249*(24), 7991-7995.

Bolaños, P., Guillen, A., Rojas, H., Boncompagni, S., & Caputo, C. (2008). The use of CalciumOrange-5N as a specific marker of mitochondrial Ca2+ in mouse skeletal muscle fibers. *Pflügers Archiv - European Journal of Physiology, 455*(4), 721-731.

Boles, T. C., Snow, C. C., & Stover, E. S. (1995). Forensic DNA testing on skeletal remains from mass graves: A pilot project in Guatemala. *J Forensic Sci, 40*, 349-355.

Brandstatter, A., & Parson, W. (2003). Mitochondrial DNA heteroplasmy or artifacts- a matter of the amplification strategy? *Int J Legal Med, 117*, 180-184.

Brandstatter, A., Parsons, T. J., Niederstatter, H., & Parson, W. (2003). Rapid screening of mtDNA coding region SNPs for the identification of Caucasian haplogroups. *Int J Legal Med, 117*(5), 291-298.

Brandstatter, A., Parsons, T. J., & Parson, W. (2003). Rapid screening of mtDNA coding region SNPs for the identification of west European Caucasian haplogroups. *Int J Legal Med, 117*(5), 291-298.

Brown, T. A., & Brown, K. A. (1994). Ancient DNA: using molecular biology to explore the past. *Bioessays, 16*(10), 719-726.

Buckman, J. F., Hernández, H., Kress, G. J., Votyakova, T. V., Pal, S., & Reynolds, I. J. (2001). MitoTracker labeling in primary neuronal and astrocytic cultures: influence of mitochondrial membrane potential and oxidants. *Journal of Neuroscience Methods, 104*(2), 165-176.

Budowle, B. (1999). Interpretation guidelines for mitochondrial DNA sequencing, tenth international symposium on human identification., from http://www.promega.com/ussymp10proc/default.htm

Budowle, B. (2001). A robust SNP typing methodology: overcoming limitations of mitochondrial analysis.: FBI.

Budowle, B., Allard, M. W., & Wilson, M. (2002a). Letter to the Editor: Characterization of heteroplasmy and hypervariable sites in HVI: critique of D'Eustachio's interpretations. *Forensic Sci Int, 130*, 68-70.

Budowle, B., Allard, M. W., & Wilson, M. R. (2002b). Critique of interpretation of high levels of heteroplasmy in the human mitochondrial region I from hair. *Forensic Sci Int, 126*, 30-33.

Budowle, B., Allard, M. W., Wilson, M. R., & Chakraborty, R. (2003). Forensics and mitochondrial DNA: applications, debates, and foundations. *Annu Rev Genomics Hum Genet, 4*, 119-141.

Budowle, B., Planz, J. V., Campbell, R. S., & Eisenberg, A. J. (2004). Single nucleotide polymorphisms and microarray technology in forensic genetics - development and application to mitochondrial DNA. *Forensic Science Review 16*(1), 21-36.

Budowle, B., Polanskey, D., Fisher, C. L., Den Hartog, B. K., Kepler, R. B., & Elling, J. W. (2010). Automated Alignment and Nomenclature for Consistent Treatment of Polymorphisms in the Human Mitochondrial DNA Control Region. *J Forensic Sci, 55*(5), 1190-1195.

Budowle, B., & van Daal, A. (2008). Forensically relevant SNP classes. *Biotechniques, 44*(5), 603-608, 610.

Budowle, B., Wilson, M., & DiZinno. (1999). Comments on Holland and Parsons's "Mitochondrial DNA sequence analysis - validation and use for forensic casework". *Forensic Science Review 11*(2), 175.

Budowle, B., Wilson, M. R., DiZinno, J. A., Stauffer, C., Fasano, M. A., Holland, M. M., et al. (1999a). Mitochondrial DNA regions HVI and HVII population data. *Forensic Sci Int, 103*(1), 23-35.

Budowle, B., Wilson, M. R., DiZinno, J. A., Stauffer, C., Fasano, M. A., Holland, M. M., et al. (1999b). Mitochondrial DNA regions HVI and HVII population data. *Forensic Sci Int, 103*, 23-35.

Burton, M. P., Schneider, B. G., Brown, R., Escamilla-Ponce, N., & Gulley, M. L. (1998). Comparison of histologic stains for use in PCR analysis of microdissected, paraffin-embedded tissues. *Biotechniques, 24*(1), 86-92.

Butler, J., & Levin, B. C. (1998). Forensic applications of mitochondrial DNA. *Trends Biotechnol, 16*(4), 158-162.

Butler, J., Wilson, M., & Reeder, D. (1998). Rapid mitochondrial DNA typing using restriction enzyme digestion of polymerase chain reaction amplicons followed by capillary electrophoresis separation with laser-induced fluorescence detection. *Electrophoresis, 19*(1), 119-124.

Calloway, C., Pasta, S., Parodi, C., & Erilch, H. (2012). A Sensitive Multiplex PCR Based Next-Generation Sequencing Assay for Resolution of Mixtures and Analysis of Forensic Samples. *Proceedings of the 2012 AAFS Meeting/ Atlanta, GA, 18*, 81-82.

Calloway, C. D., Reynolds, R. L., Herrin, G. L., & Anderson, W. W. (2000). The frequency of heteroplasmy in the HVII region of mtDNA differs across tissue types and increases with age. *Am J Hum Genet 66*(1384-1397).

Carr, S. M. (2010). Transition versus Transversion mutations. 2010, from http://www.mun.ca/biology/scarr/Transitions_vs_Transversions.html

Carracedo, A., Bar, W., Lincoln, P., Mayr, W., Morling, N., Olaisen, B., et al. (2000). DNA commission of the international society for forensic genetics: guidelines for mitochondrial DNA typing. *Forensic Sci Int, 110*, 79-85.

Carrecedo, A., D'Aloja, E., Dupuy, B., Jangbland, A., Karjalainen, M., Lambert, C., et al. (1998). Reproducibility of mtDNA analysis between laboratories: a report of the European DNA profiling group (EDNAP). *Forensic Sci Int, 97*, 167-170.

Chen, H., & Chan, D. C. (2005). Emerging functions of mammalian mitochondrial fusion and fission. *Hum. Mol. Genet., 14*(suppl_2), R283-289.

Chen, H., Chomyn, A., & Chan, D. C. (2005). Disruption of Fusion Results in Mitochondrial Heterogeneity and Dysfunction. *J. Biol. Chem., 280*(28), 26185-26192.

Chen, J. Z., Gokden, N., Greene, G. F., Green, B., & Kadlubar, F. F. (2003). Simultaneous generation of multiple mitochondrial DNA mutations in human prostate tumors suggests mitochondrial hyper-mutagenesis. *Carcinogenesis, 24*(9), 1481-1487.

Chinnery, P. F., Thorburn, D. R., Samuels, D. C., White, S. L., Hans-Heinrik, Dahl, M., et al. (2000). The inheritance of mitochondrial DNA heteroplasmy: random drift, selection, or both? *Trends in Genetics, 16*(11), 500-505.

Chong, M. D., Calloway, C. D., Klein, S. B., Orrego, C., & Buoncristiani, M. R. (2005). Optimization of a duplex amplification and sequencing strategy for the HVI/HVII regions of human mitochondrial DNA for forensic casework. *Forensic Sci Int, 154*(2-3), 137-148.

Coble, M., Just, R., O'Callaghan, J., Letmanyi, I., Peterson, C., Irwin, J., et al. (2004). Single nucleotide polymorphisms over the entire mtDNA genome that increase the power of forensic testing in Caucasians. *Int J Legal Med, 118*(3), 137-146.

Coble, M. D., Vallone, P. M., Just, R. S., Diegoli, T. M., Smith, B. C., & Parsons, T. J. (2006). Effective strategies for forensic analysis in the mitochondrial DNA coding region. *Int J Legal Med, 120*(1), 27-32.

Collier, J. R., Monk, N. A. M., Maini, P. K., & Lewis, J. H. (1996). Pattern Formation by Lateral Inhibition with Feedback: a Mathematical Model of Delta-Notch Intercellular Signalling. *Journal of Theoretical Biology, 183*(4), 429-446.

Collins, S. J. (1987). The HL-60 promyelocytic leukemia cell line: proliferation, differentiation, and cellular oncogene expression. *Blood, 70*, 1233-1244.

Corach, D., Sala, A., Penacino, G., Iannucci, N., Bernardi, P., Doretti, M., et al. (1997). Additional approaches to DNA typing of skeletal remains: The search for "missing" persons killed during the last dictatorship in Argentina. *Electrophoresis, 18*, 1608-1612.

D'Eustachio, P. D. (2002). Letter to the Editor: High levels of mitochondrial DNA heteroplasmy in human hairs by Budowle et al. *Forensic Sci Int, 129*, 35-42.

Di Martino, D., Giuffre, G., Staiti, N., Simone, A., Le Donne, M., & Saravo, L. (2004). Single sperm cell isolation by laser microdissection. *Forensic Sci Int, 146 Suppl*, S151-153.

Dixon, G. B. (2011). Development of a Single Mitochondrial DNA Amplification Strategy for Two Platforms: Next Generation and Sanger Sequencing from the Same Amplicon Library. *Proceedings of the 2012 AAFS Meeting/ Atlanta, GA, 18*, 80-81.

Edenberg, H. J., & Liu, Y. (2009). Laboratory Methods for High-Throughput Genotyping. *Cold Spring Harbor Protocols, 2009*(11).

Eyre-Walker, A., & Awadalla, P. (2001). Does human mtDNA recombine? *Journal of Molecular Evolution, 53*, 430-435.

Eyre-Walker, A., Smith, N. H., & Smith, J. M. (1999). Mitochondrial DNA recombination-reasons to panic. Reply to Macaulay et al. *Proceedings of the Royal Society of London., 266*, 2041-2042.

FBI. (2003). *Revised SWGDAM validation guidelines.* Retrieved 2010. from www.cstl.nist.gov/strbase/validation/SWGDAM_Validation.doc.

Finnila, S., Lehtonen, M. S., & Majamaa, K. (2001). Phylogenetic network for European mtDNA. *Am J Hum Genet, 68*(6), 1475-1484.

Frank, M., Dapson, R., Wickersham, T., & Kiernan, J. (2007). Certification procedures for nuclear fast red (Kernechtrot), CI 60760. *Biotech Histochem, 82*(1), 35-39.

Gebhardt, R. (2002). Liver Epithelial Cell Culture. In C. Wise (Ed.), *Methods in Molecular Biology, Epithelial Cell Culture Protocols* (Vol. 188, pp. 53-63). Totowa, NJ: Humana Press.

Giles, R. E., Blanc, H., Cann, H. M., & Wallace, D. C. (1980). Maternal inheritance of human mitochondrial DNA. . *Proc. Natl. Acad. Sci., 77*(11), 6715-6719.

Gill, P., Ivanov, P. L., Kimpton, C., Piercy, R., Benson, N., Tully, G., et al. (1994). Identification of the remains of the Romanov family by DNA analysis. *Nature Genetics, 6*, 130-135.

Goto, Y., Nonaka, I., & Horai, S. (1990). A mutation in the tRNA Leu(UUR) gene associated with the MELAS subgroup of mitochondrial encephalomyopathies. *Nature, 348*, 651-653.

Gray, M. W. (1989). Origin and evolution of mitochondrial DNA. *Annu Rev Cell Biol, 5*, 25-50.

Greenberg, B. D., Newbold, J. E., & Sugino, A. (1983). Intraspecific nucleotide sequence variability surrounding the origin of replication in human mitochondrial DNA. *Gene, 21*(1-2), 33-49.

Grzybowski, T. (2000). Extremely high levels of human mitochondrial DNA heteroplasmy in single hair roots. *Electrophoresis, 21*, 548-553.

Grzybowski, T. (2001). Response to "Extremely high levels of human mitochondrial DNA heteroplasmy in single hair roots". *Electrophoresis, 22*, 180-182.

Grzybowski, T., Malyarchuk, B., Czarny, J., Miścicka-Sliwka, D., & Kotzbach, R. (2003). High levels of mitochondrial DNA heteroplasmy in single hair roots: Reanalysis and revision. *Electrophoresis, 24*, 1159-1165.

Gyllensten, U. B., & Erlich, H. A. (1988). Generation of single-stranded DNA by the polymerase chain reaction and its application to direct sequencing of the HLA-DQA locus. *Proc. Natl. Acad. Sci., 85*, 7652-7656.

He, Y., Wu, J., Dressman, D. C., Iacobuzio-Donahue, C., Markowitz, S. D., Velculescu, V. E., et al. (2010). Heteroplasmic mitochondrial DNA mutations in normal and tumour cells. *Nature, 464*(7288), 610-614.

Herrnstadt, C., Elson, J., Fahy, E., Preston, G., Turnbull, D., Anderson, C., et al. (2002). Reduced-median-network analysis of complete mitochondrial DNA coding-region sequences for the major African, Asian, and European haplogroups. *Am J Hum Genet, 70*(5), 1152-1171.

Hoelzel, A. R., Lopez, J. V., Dover, G. A., & O'Brien, S. J. (1994). Rapid evolution of a heteroplasmic repetitive sequence in the mitochondrial DNA control region of carnivores. *Journal of Molecular Evolution, 39*, 191-199.

Hofreiter, M., Serre, D., Poinar, H. N., Kuch, M., & Paabo, S. (2001). Ancient DNA. *Nat Rev Genet, 2*(5), 353-359.

Holland, M., Fisher, D., Roby, R., Ruderman, J., Bryson, C., & Weedn, V. (1995). Mitochondrial DNA sequence analysis of human remains. *Crime Laboratory Digest, 22*(4), 109-122.

Holland, M., & Parsons, T. J. (1999). Mitochondrial DNA sequence analysis - validation and use for forensic casework. *Forensic Science Review, 11*(1), 21-50.

Homer, N., Merriman, B., & Nelson, S. F. (2009). Local alignment of two-base encoded DNA sequence. *BMC Bioinformatics, 10*(175).

Hopgood, R., Sullivan, K. M., & Gill, P. (1992). Strategies for automated sequencing of human mitochondrial DNA directly from PCR products. *BioTechniques, 13*(1), 82-92.

Hunt, J., & Finkelstein, S. (2004). Microdissection techniques for molecular testing in surgical pathology. *Arch Pathol Lab Med, 128*(12), 1372-1378.

Hutchinson, C., Newbold, J., Potter, S., & Edgell, M. (1974). Maternal inheritance of mammalian mitochondrial DNA. *Nature 251*, 536-538.

Ingman, M., & Gyllensten, U. (2003). Mitochondrial Genome Variation and Evolutionary History of Australian and New Guinean Aborigines. *Genome Research, 13*(7), 1600-1606.

Ingman, M., Kaessmann, H., Paabo, S., & Gyllensten, U. (2000). Mitochondrial genome variation and the origin of modern humans. *Nature, 408*(6813), 708-713.

Invitrogen. (2006). MitoTracker® and MitoFluor™ mitochondrion-selective probes. (pp. 1-6). Eugene, OR.

Irwin, J. A., Saunier, J. L., Niederstatter, H., Strouss, K. M., Sturk, K. A., Diegoli, T. M., et al. (2009). Investigation of heteroplasmy in the human mitochondrial DNA control region: a synthesis of observations from more than 5000 global population samples. *J Mol Evol, 68*(5), 516-527.

Isenberg, A. R., & Moore, J. M. (1999). Mitochondrial DNA analysis at the FBI laboratory. *Forensic Science Communications, 1*(2).

Ivanov, P. L., Wadhams, M. J., Roby, R. K., Holland, M. M., Weedn, V. W., & Parsons, T. J. (1996). Mitochondrial DNA sequence heteroplasmy in the Grand Duke of Russia Georij Romanov establishes the authenticity of the remains of Tsar Nicholas II. *Nature Genetics, 12*, 417-420.

Jazin, E. E., Cavelier, L., Eriksson, I., Oreland, L., & Gyllensten, U. (1996). Human brain contains high levels of heteroplasmy in the noncoding regions of mitochondrial DNA. *Proc. Natl. Acad. Sci., 93*, 12382-12387.

Kline, M. C., Vallone, P. M., Redman, J. W., Duewer, D. L., Calloway, C. D., & Butler, J. M. (2005). Mitochondrial DNA typing screens with control region and coding region SNPs. *J Forensic Sci, 50*(2), 377-385.

Kolowski, J. C., Goncharoff, P., & Fisher, D. R. (2004). Validation Studies for Mitochondrial Laboratory for the Department of Forensic Biology: Office of Chief Medical Examiner, NYC.

Kong, Q. P., Yao, Y. G., Sun, C., Bandelt, H. J., Zhu, C. L., & Zhang, Y. P. (2003). Phylogeny of east Asian mitochondrial DNA lineages inferred from complete sequences. *Am J Hum Genet, 73*(3), 671-676.

Kraytsberg, Y., Schwartz, M., Brown, T. A., Ebralidse, K., Kunz, W. S., Clayton, D. A., et al. (2004). Recombination of human mitochondrial DNA. *Science, 304*(5673), 981-.

Kuznetsov, A. V., Mayboroda, O., Kunz, D., Winkler, K., Schubert, W., & Kunz, W. S. (1998). Functional imaging of mitochondria in saponin-permeabilized mice muscle fibers. *J. Cell Biol., 140*(5), 1091-1099.

Lagerstrom-Fermer, M., Olsson, C., Forsgren, L., & Syvanen, A. (2001). Heteroplasmy of the human mtDNA control region remains constant during life. *Am J Hum Genet, 68*, 1299-1301.

Lee, Y.-j., Jeong, S.-Y., Karbowski, M., Smith, C. L., & Youle, R. J. (2004). Roles of the Mammalian Mitochondrial Fission and Fusion Mediators Fis1, Drp1, and Opa1 in Apoptosis. *Mol. Biol. Cell, 15*(11), 5001-5011.

Legros, F., Lombes, A., Frachon, P., & Rojo, M. (2002). Mitochondrial Fusion in Human Cells Is Efficient, Requires the Inner Membrane Potential, and Is Mediated by Mitofusins. *Mol. Biol. Cell, 13*(12), 4343-4354.

Levin, B. C., Cheng, H., & Reeder, D. J. (1999). A Human Mitochondrial DNA Standard Rererence Material for Quality Control in Forensic Identification, Medical Diagnosis, and Mutation Detection. *Genomics, 55*, 135-146.

Levin, B. C., Hancock, D. K., Holland, K. A., H.Cheng, & Richie, K. L. (2003). *Human Mitochondrial DNA- Amplification and Sequencing- Standard Reference Materials- SRM 2392 and SRM 2392-I*. Retrieved 2010. from www.nist.gov/srm/upload/sp260-155.pdf.

Levin, B. C., Holland, K. A., Hancock, D. K., Coble, M., Parsons, T. J., Kienker, L. J., et al. (2003). Comparison of the Complete mtDNA Genome Sequences of Human Cell Lines - HL-60 and GM10742A- From Individuals with Pro-Myelocytic Leukemia and Leber Hereditary Optic Neuropathy, Respectively, and the Inclusion of HL-60 in the NIST Human Mitochondiral DNA Standard Reference Material - SRM 2392-I. *Mitochondrion*(2), 387-400.

Li, J., Butler, J. M., Tan, Y., Lin, H., Royer, S., Ohler, L., et al. (1999). Single nucleotide polymorphism determination using primer extension and time-of-flight mass spectrometry. *Electrophoresis, 20*(6), 1258-1265.

Lopez, J. V., Culver, M., Stephens, J. C., Johnson, W. E., & O'Brien, S. J. (1997). Rates of nuclear and cytoplasmic mitochondrial DNA sequence divergence in mammals. *Mol. Biol. Evol., 14*(3), 277-286.

Lutz-Bonengel, S., Sänger, T., Parson, W., Müller, H., Ellwart, J., Follo, M., et al. (2008). Single lymphocytes from two healthy individuals with mitochondrial point heteroplasmy are mainly homoplasmic. *Int J Legal Med, 122*(3), 189-197.

Lutz, S., Weisser, H. J., Heizmann, J., & Pollak, S. (1996). mtDNA as a tool for identification of human remains. *Int J Legal Med, 109*, 205-209.

Lutz, S., Weisser, H. J., Heizmann, J., & Pollak, S. (1998). Location and frequency of polymorphic positions in the mtDNA control region of individuals from Germany. *Int J Legal Med, 111*(2), 67-77.

Maca-Meyer, N., Gonzalez, A. M., Pestano, J., Flores, C., Larruga, J. M., & Cabrera, V. M. (2003). Mitochondrial DNA transit between West Asia and North Africa inferred from U6 phylogeography. *BMC Genet, 4*, 15.

Macaulay, V., Richards, M., & Sykes, B. (1999). Mitochondrial DNA recombination-no need to panic. *Proceedings of the Royal Society of London. , 266*, 2037-2039.

Marchington, D. R., Macaulay, V., Hartshorne, G. M., Barlow, D., & Poulton, J. (1998). Evidence from human oocytes for a genetic bottleneck in an mtDNA disease. *Am J Hum Genet, 63*(3), 769-775.

Marino, M. A., Weaver, K. R., Tully, L. A., Girard, J. E., & Belgrader, P. (1996). Characterization of mitochondrial DNA using low-stringency single specific primer amplification analyzed by laser induced fluorescence- capillary electrophoresis. *Electrophoresis, 17,* 1499-1504.

Meissner, C., vonWurmb, N., & Oehmichen, M. (1997). Detection of the age-dependent 4977 bp deletion of mitochondrial DNA. *Int J Legal Med, 110,* 288-291.

Melton, T. (2004). Mitochondrial DNA heteroplasmy. *Forensic Science Review, 16*(1), 2-20.

Mendisco, F., Keyser, C., Hollard, C., Seldes, V., Nielsen, A. E., Crubézy, E., et al. (2011). Application of the iPLEXTM Gold SNP genotyping method for the analysis of Amerindian ancient DNA samples: Benefits for ancient population studies. *Electrophoresis, 32*(3-4), 386-393.

Miething, F., Hering, S., Hanschke, B., & Dressler, J. (2005). The Effect of Fixation to the Degradation of Nuclear and Mitochondrial DNA in Different Tissues. *J. Histochem. Cytochem.*, jhc.5B6726.2005.

Miller, K., & Budowle, B. (2001). A compendium of human mitochondrial DNA control region: development of an international standard forensic database. *Croat Med J, 42*(3), 315-327.

Miller, K., Dawson, J. L., & Hagelberg, E. (1996). A concordance of nucleotide substitutions in the first and second hypervariable segments of the human mtDNA control region. *Int J Legal Med, 109*(3), 107-113.

Minamikawa, T., Sriratana, A., Williams, D. A., Bowser, D. N., Hill, J. S., & Nagley, P. (1999). Chloromethyl-X-rosamine (Mito Tracker Red) photosensitizes mitochondria and induces apoptosis in intact human cells. *Journal of Cell Science, 112*(14), 2419-2430.

Mishmar, D., Ruiz-Pesini, E., Golik, P., Macaulay, V., Clark, A. G., Hosseini, S., et al. (2003). Natural selection shaped regional mtDNA variation in humans. *Proc Natl Acad Sci U S A, 100*(1), 171-176.

Mullis, K. B. (1990). The unusual origin of the polymerase chain reaction. *Scientific American, April 1990,* 56-63.

Musgrave-Brown, E., Ballard, D., Balogh, K., Bender, K., Berger, B., Bogus, M., et al. (2007). Forensic validation of the SNPforID 52-plex assay. *Forensic Sci Int. Genet, 1*(2), 186-190.

Naue, J., Sänger, T., Schmidt, U., Klein, R., & Lutz-Bonengel, S. (2011). Factors affecting the detection and quantification of mitochondrial point heteroplasmy using Sanger sequencing and SNaPshot minisequencing. *Int J Legal Med, 125*(3), 427-436.

Noonan, J. P., Coop, G., Kudaravalli, S., Smith, D., Krause, J., Alessi, J., et al. (2006). Sequencing and Analysis of Neanderthal Genomic DNA. *Science, 314*(5802), 1113-1118.

Nunnari, J., Marshall, W. F., Straight, A., Murray, A., Sedat, J. W., & Walter, P. (1997). Mitochondrial transmission during mating in Saccharomyces cerevisiae is determined by mitochondrial fusion and fission and the intramitochondrial segregation of mitochondrial DNA. *Mol. Biol. Cell, 8*(7), 1233-1242.

Okamoto, K., & Shaw, J. M. (2005). Mitochondrial morphology and dynamics in yeast and multicellular eukaryotes. *Annual Review of Genetics, 39*(1), 503-536.

Paabo, S., Gifford, J. A., & Wilson, A. C. (1988). Mitochondrial DNA sequences from a 7000-year old brain. *Nucleic Acid Research, 16*, 9775-9778.

Palanichamy, M. G., Sun, C., Agrawal, S., Bandelt, H. J., Kong, Q. P., Khan, F., et al. (2004). Phylogeny of mitochondrial DNA macrohaplogroup N in India, based on complete sequencing: implications for the peopling of South Asia. *Am J Hum Genet, 75*(6), 966-978.

Paneto, G. G., Longo, L. V. G., Martins, J. A., de Camargo, M. A., Costa, J. C., de Mello, A. C. O., et al. (2010). Heteroplasmy in Hair: Study of Mitochondrial DNA Third Hypervariable Region in Hair and Blood Samples. *J Forensic Sci, 55*(3), 715-718.

Paneto, G. G., Martins, J. A., Longo, L. V. G., Pereira, G. A., Freschi, A., Alvarenga, V. L. S., et al. (2007). Heteroplasmy in hair: Differences among hair and blood from the same individuals are still a matter of debate. *Forensic Sci Int, 173*(2-3), 117-121.

Parsons, T. J. (2006). *Mitochondrial DNA Genome Sequencing and SNP Assay Development for Increased Power of Discrimination*. Retrieved. from http://www.ncjrs.gov/App/Publications/abstract.aspx?ID=235000.

Parsons, T. J., & Coble, M. D. (2001). Increasing the forensic discrimination of mitochondrial DNA testing through analysis of the entire mitochondrial DNA genome. *Croat Med J, 42*(3), 304-309.

Parsons, T. J., Muniec, D. S., Sullivan, K., Woodyatt, N., Allston-Greiner, R., Wilson, M. R., et al. (1997). A high observed substitution rate in the human mitochondrial DNA control region. *Nature Genetics, 15*, 363-368.

Perfettini, J.-L., Roumier, T., & Kroemer, G. (2005). Mitochondrial fusion and fission in the control of apoptosis. *Trends Cell Bio., 15*(4), 179-183.

Pfeiffer, H., Lutz-Bonengel, S., Pollak, S., Fimmers, R., Baur, M. P., & Brinkmann, B. (1999). Mitochondrial DNA typing from human auxillary, pubic and head hair shafts - success rates and sequence comparisons. *Int J Legal Med, 112*, 287-290.

Pflugradt, R., Schmidt, U., Landenberger, B., Sanger, T., & Lutz-Bonengel, S. (2011). A novel and effective separation method for single mitochondria analysis. *Mitochondrion, 11*(2), 308-314.

Podini, D., & Vallone, P. M. (2009). Chapter 23- SNP Genotyping Using Multiplex Single Base Primer Extension Assays. In *Single Nucleotide Polymorphisms* (Vol. 578, pp. 379-391): Springer.

Polanskey, D., Den Hartog, B. K., Elling, J. W., Fisher, C. L., Kepler, R. B., & Budowle, B. (2010). Comparison of Mitotyper Rules and Phylogenetic-based mtDNA Nomenclature Systems. *J Forensic Sci, 55*(5), 1184-1189.

Poot, M., Zhang, Y.-Z., Kraemer, J. A., Wells, K. S., Jones, L. J., Hanzel, D. K., et al. (1996). Analysis of mitochondrial morphology and function with novel fixable fluorescent stains. *Journal of Histochemistry and Cytochemistry, 44*(12), 1363-1372.

Popiolek, D. A., Prinz, M. K., West, A. B., Nazzaruolo, B. L., Estacio, S. M., & Budimlija, Z. M. (2003). Multiplex DNA Short Tandem Repeat Analysis. *American Journal of Clinical Pathology, 120*(5), 746-751.

Porteous, W. K., James, A. M., Sheard, P. W., Porteous, C. M., Packer, M. A., Hyslop, S. J., et al. (1998). Bioenergetic consequences of accumulating the common 4977-bp mitochondrial DNA deletion. *European Journal of Biochemistry, 257*(1), 192-201.

Quintans, B., Alvarez-Iglesias, V., Salas, A., Phillips, C., Lareu, M. V., & Carracedo, A. (2004). Typing of mitochondrial DNA coding region SNPs of forensic and anthropological interest using SNaPshot™ minisequencing. *Forensic Sci Int, 140*(2-3), 251-257.

Reiner, J. E., Kishore, R. B., Levin, B. C., Albanetti, T., Boire, N., Knipe, A., et al. (2010). Detection of Heteroplasmic Mitochondrial DNA in Single Mitochondria. *PLoS ONE, 5*(12).

Richter, C., Park, J.-W., & Ames, B. N. (1988). Normal oxidative damage to mitochondrial and nuclear DNA is extensive. *Proc. Natl. Acad. Sci., 85*(17), 6465-6467.

Roberts, K. A., & Calloway, C. (2011). Characterization of Mitochondrial DNA Sequence Heteroplasmy in Blood Tissue and Hair as a Function of Hair Morphology. *J Forensic Sci, 56*(1), 46-60.

Rook, M. S., Delach, S. M., Deyneko, G., Worlock, A., & Wolfe, J. L. (2004). Whole Genome Amplification of DNA from Laser Capture-Microdissected Tissue for High-Throughput Single Nucleotide Polymorphism and Short Tandem Repeat Genotyping. *Am J Pathol, 164*(1), 23-33.

Rozen, S., & Skaletsky, H. J. (1998). Primer3, from http://www-genome.wi.mit.edu/genome_software/other/primer3.html.

Salas, A., Lareu, M. V., & Carracedo, A. (2001). Heteroplasmy in mtDNA and the weight of evidence in forensic mtDNA analysis: a case report. *Int J Legal Med, 114*(3), 186-190.

Salet, C., & Moreno, G. (1990). New trends in photobiology photosensitization of mitochondria. Molecular and cellular aspects. *Journal of Photochemistry and Photobiology B: Biology, 5*(2), 133-150.

Sams, A., & Davies, F. M. R. (1967). Commercial Varieties of Nuclear Fast Red; Their Behaviour in Staining After Autoradiography. *Biotechnic & Histochemistry, 42*(6), 269-276.

Sanger, F. (1981). Determination of nucleotide sequences in DNA. *Science, 214*(4526), 1205-1210.

Sanger, F., Nicklen, S., & Coulson, A. R. (1977). DNA sequencing with chain-terminating inhibitors. *Proc. Natl. Acad. Sci. USA, 74*(12), 5463-5467.

Santos, C., Sierra, B., Álvarez, L., Ramos, A., Fernández, E., Nogués, R., et al. (2008). Frequency and Pattern of Heteroplasmy in the Control Region of Human Mitochondrial DNA. *Journal of Molecular Evolution, 67*(2), 191-200.

Schwartz, M., & Vissing, J. (2002). Paternal inheritance of mitochondrial DNA. *N Engl J Med, 347*(8), 576-580.

Schwartz, M., & Vissing, J. (2004). No evidence for paternal inheritance of mtDNA in patients with sporadic mtDNA mutations. *Journal of the neurological sciences, 218*(1), 99-101.

Sigurðardóttir, S., Helgason, A., Gulcher, J., Stefansson, K., & Donnelly, P. (2000). The mutation rate in the human mtDNA control region. *Am J Hum Genet, 66*, 1599-1609.

Slate, J., & Gemmell, N. J. (2004). Eve 'n' Steve: recombination of human mitochondrial DNA *Trends in Ecology & Evolution, 19*(11), 561-563.

Sobrino, B., Brion, M., & Carracedo, A. (2005). SNPs in forensic genetics: a review on SNP typing methodologies. *Forensic Sci Int, 154*(2-3), 181-194.

Stewart, J., Aagaard, P., Pokorak, E., Polanskey, D., & Budowle, B. (2003). Evaluation of a multicapillary electrophoresis instrument for mitochondrial DNA typing. *J Forensic Sci, 48*(3), 571-580.

Stone, A. C., Starrs, J. E., & Stoneking, M. (2001). Mitochondrial DNA analysis of the presumptive remains of Jesse James. *J Forensic Sci, 46*(1), 173-176.

Stoneking, M. (2000). Hypervariable sites in the mtDNA control region are mutational hotspots. *Am J Hum Genet, 67*(4), 1029-1032.

Stoneking, M., Hedgecock, D., Higuchi, R. G., Vigilant, L., & Erlich, H. A. (1991). Population variation of human mtDNA control region sequences detected by enzymatic amplification and sequence-specific oligonucleotide probes. . *Am J Hum Genet 48*, 370-382.

Sullivan, K., Hopgood, R., & Gill, P. (1992). Identification of human remains by amplification and automated sequencing of mitochondrial DNA. *Int J Legal Med, 105*, 83-86.

Sutovsky, P., Navara, C. S., & Schatten, G. (1996). Fate of the sperm mitochondria, and the incorporation, conversion, and disassembly of the sperm tail structures during bovine fertilization. *Biol Reprod, 55*(6), 1195-1205.

Sutovsky, P., Ramalho-Santos, J., Moreno, R. D., Oko, R., Hewitson, L., & Schatten, G. (1999). On-stage selection of single round spermatids using a vital, mitochondrion-specific fluorescent probe MitoTrackerTM and high resolution differential interference contrast microscopy. *Hum. Reprod., 14*(9), 2301-2312.

SWGDAM. (2003). *Guidelines for mitochondrial DNA (mtDNA) nucleotide sequence interpretation.* Retrieved 2010. from http://www.fbi.gov/hq/lab/fsc/backissu/april2003/swgdammitodna.htm

Tagliabracci, A., Turchi, C., Buscemi, L., & Sassaroli, C. (2001). Polymorphism of the mitochondrial DNA control region in Italians. *Int J Legal Med, 114*(4-5), 224-228.

Thangaraj, K., Chaubey, G., Kivisild, T., Reddy, A. G., Singh, V. K., Rasalkar, A. A., et al. (2005). Reconstructing the origin of Andaman Islanders. *Science, 308*(5724), 996.

Tully, G., Bar, W., Brinkmann, B., Carracedo, A., Gill, P., Morling, N., et al. (2001). Considerations by the European DNA profiling (EDNAP) group on the working practices, nomenclature and interpretation of mitochondrial DNA profiles. *Forensic Sci Int, 124*, 83-91.

Tully, L. A., Parsons, T. J., Steighner, R. J., Holland, M. M., Marino, M. A., & Prenger, V. L. (2000). A sensitive denaturing gradient-gel electrophoresis assay reveals a high frequency of heteroplasmy in hypervariable region I of the human mtDNA control region. *Am J Hum Genet, 67*, 432-443.

Uneyama, C., Shibutani, M., Masutomi, N., Takagi, H., & Hirose, M. (2002). Methacarn Fixation for Genomic DNA Analysis in Microdissected, Paraffin-embedded Tissue Specimens. *J. Histochem. Cytochem., 50*(9), 1237-1245.

Vallone, P., & Butler, J. (2004). AutoDimer: a screening tool for primer-dimer and hairpin structures. . *Biotechniques. , 37*(2), 226-231.

Vallone, P. M., Just, R. S., Coble, M. D., Butler, J. M., & Parsons, T. J. (2004). A multiplex allele-specific primer extension assay for forensically informative SNPs distributed throughout the mitochondrial genome. *Int J Legal Med, 118*(3), 147-157.

Vassella, E., Straesser, K., & Boshart, M. (1997). A mitochondrion-specific dye for multicolour fluorescent imaging of Trypanosoma brucei. *Molecular and Biochemical Parasitology, 90*(1), 381-385.

Wallace, D. (1995). Mitochondrial DNA variation in human evolution, degerative disease, and aging. *Am J Hum Genet, 57*, 201-223.

Wallace, D. (1997). Mitochondrial DNA in aging and disease. *Scientific American, August*, 40-47.

Wilson, M., Allard, M., Monson, K., Miller, K., & Budowle, B. (2002a). Further Discussion of the Consistent Treatment of Length Variants in the Human Mitochondrial DNA Control Region. *Forensic Science Communications, 4*(4), 1-20.

Wilson, M., & Allard, M. W. (2004). Phylogenetics and mitochondrial DNA. *Forensic Science Review, 16*(1), 37-62.

Wilson, M., Allard, M. W., Monson, K., Miller, K., & Budowle, B. (2002b). Recommendations for consistent treatment of length variants in the human mitochondrial DNA control region. *Forensic Sci Int, 129*(1), 35-42.

Wilson, M., & Cann, R. L. (1992). The recent African genesis of humans. *Scientific American, April 1992*, 68-73.

Wilson, M., DiZinno, J. A., Polanskey, D., Replogle, J., & Budowle, B. (1995). Validation of mitochondrial DNA sequencing for forensic casework analysis. *Int J Legal Med 108*, 68-74.

Wilson, M., Holland, M., Stoneking, M., DiZinno, J., & Budowle, B. (1993). Guidelines for the use of mitochondrial DNA sequencing in forensic science. *Crime Laboratory Digest, 20*, 68-77.

Wilson, M., Polanskey, D., Butler, J., DiZinno, J., Replogle, J., & Budowle, B. (1995). Extraction, PCR amplification, and sequencing of mitochondrial DNA from human hair shafts. *BioTechniques 18*(4), 662-669.

Wilson, M., Polanskey, D., Replogle, J., DiZinno, J. A., & Budowle, B. (1997). A family exhibiting heteroplasmy in the human mitochondiral DNA control region reveals both somatic mosaicism and pronounced segregation of mitotypes. *Human Genetics, 100*, 167-171.

Xiu-Cheng Fan, A., Garritsen, H. S., Tarhouny, S. E., Morris, M., Hahn, S., Holzgreve, W., et al. (2008). A rapid and accurate approach to identify single nucleotide polymorphisms of mitochondrial DNA using MALDI-TOF mass spectrometry. *Clin Chem Lab Med, 46*(3), 299-305.

Yaffe, M. P. (1999). The Machinery of Mitochondrial Inheritance and Behavior. *Science, 283*(5407), 1493-1497.

Zeiss. (2008). PALM microLaser systems: Carl Zeiss Corp.

Zhang, W., & Deisseroth, A. B. (1991). Effect of nucleotide concentration on specificty of sequence amplification. *BioTechniques, 11*(1), 60-62.

Zsurka, G., Hampel, K. G., Kudina, T., Kornblum, C., Kraytsberg, Y., Elger, C. E., et al. (2007). Inheritance of mitochondrial DNA recombinants in double-heteroplasmic families: potential implications for phylogenetic analysis. . *Am J Hum Genet, 80*(2), 298-305.

Zsurka, G., Kraytsberg, Y., Kudina, T., Kornblum, C., Elger, C. E., Khrapko, K., et al. (2005). Recombination of mitochondrial DNA in skeletal muscle of individuals with multiple mitochondrial DNA heteroplasmy. *Nature Genetics, 37*(8), 873-877.