**Title**
Microchip Technology for High-Throughput Cancer Pathology



**Author**
Duberow, Jr, David Paul

**Publication Date**
2009

Peer reviewed|Thesis/dissertation

# Microchip Technology for High-Throughput Cancer Pathology

by

David Paul Duberow, Jr.

A dissertation submitted in partial satisfaction of the

requirements for the degree of

Doctor of Philosophy

in

Chemistry

in the

Graduate Division

of the

University of California, Berkeley

Committee in charge:

Professor Richard A. Mathies, Chair
Professor Evan R. Williams
Professor Martyn T. Smith

Fall 2009

Microchip Technology for High-Throughput Cancer Pathology

Abstract

Microchip Technology for High-Throughput Cancer Pathology

by

David Paul Duberow, Jr.

Doctor of Philosophy in Chemistry

University of California, Berkeley

Professor Richard A. Mathies, Chair

Novel methodologies for the molecular detection of human cancer have been advanced using microfabricated capillary electrophoresis devices as analytical platforms. These techniques enable direct and quantitative characterization of the unique biomolecular signature inherent to individual cancers and, through reduction of sample usage and analysis time, further the goal of routine genetic screening in the clinical setting.

The majority of the work detailed here applies a comparative sequencing technique known as Polymorphism Ratio Sequencing (PRS) to cancer detection with a focus on mutations in the mitochondrial DNA (mtDNA). First, a rigorous optimization of sample processing protocols was undertaken to improve PRS separations on a 96-lane microfabricated sequencing device. A modification of electrokinetic injection conditions has increased capillary success rates to nearly 100%, while the introduction of dynamic coatings and automated data processing has decreased analysis time by 75%. These optimized conditions were validated through a complete mtDNA sequence comparison of two unrelated individuals, uncovering 44 confirmed germline variations, eight of which were undetected in early PRS experiments. Further analysis of paired tumor and blood mtDNA from six individuals with lung and bladder cancer revealed three heteroplasmic somatic variants while uncovering 18 erroneous mutations identified in previous microarray analysis.

To establish potential clinical relevance, PRS was independently applied to a mitochondrial D-loop analysis of fourteen bladder cancer patients. A total of 21 somatic variations were identified, with seven patients harboring at least one mutation. Fifteen of these mutations were heteroplasmic, often occurring at low levels or problematic base locations inaccessible to conventional technologies. Where available, matched urine mtDNA was found to contain abundant populations of the mutant genotype, establishing the potential use of bodily fluids for noninvasive screening.

Finally, an integrated microdevice capable of PRS extension followed by inline purification and electrophoretic separation is presented. This device makes use of dual on-chip thermal cyclers and orthogonal Sanger extension primers to generate a complete set of PRS fragments prior to oligonucleotide-based capture and injection. Further integration with upstream sample processing steps, including single-cell capture and PCR amplification, is proposed, providing the framework for real-time mutant quantitation in microbiopsies, ultimately enabling full clinical integration.

To the Memory of
Margaret Ann Downing
and
Ruth Marie Duberow

# Contents

# List of Abbreviations

| | |
|---|---|
| 12S, 16S | 12S and 16S ribosomal RNA |
| ADP | adenosine diphosphate |
| ATP | adenosine triphosphate |
| ATPase6 (8) | ATP synthase subunits (6 and 8) |
| bp | base pairs |
| CE | capillary electrophoresis |
| CEPH | Centre d'Etude du Polymorphism Humain |
| COXI (II, III) | cytochrome *c* oxidase subunits (I, II, and III) |
| Cytb | cytochrome *b* |
| ddNTP | dideoxynucleotide triphosphate |
| D-loop | displacement loop |
| DNA | deoxyribonucleic acid |
| dNTP | deoxynucleotide triphosphate |
| DTW | dynamic time warp |
| EDTA | ethylenediaminetetraacetic acid |
| EOF | electroosmotic flow |
| ET | energy transfer |
| HEA | hydroxyethylacrylamide |
| HF | hydrofluoric acid |
| IVT | *in vitro* transcribed |
| LPA | linear polyacrylamide |
| mRNA | messenger RNA |
| mtDNA | mitochondrial DNA |
| NADH | nicotinamide adenine dinucleotide |
| NCI | National Cancer Institute |
| ND1 (2, 3, 4, 4L, 5, 6) | NADH dehydrogenase subunits (1-6) |
| NHGRI | National Human Genome Research Institute |
| NIH | National Institutes of Health |
| nt | nucleotide |
| OXPHOS | oxidative phosphorylation |
| PAH | polycyclic aromatic hydrocarbon |
| PCR | polymerase chain reaction |
| PDMA | polydimethylacrylamide |
| PDMS | polydimethylsiloxane |
| Ph | Philadelphia chromosome |
| PMMA | polymethylmethacrylate |
| PRS | polymorphism ratio sequencing |
| Rb | retinoblastoma protein |
| RNA | ribonucleic acid |

ROS . . . . . reactive oxygen species
rRNA . . . . . ribosomal RNA
RTD . . . . . resistive temperature detector
SDS . . . . . sodium dodecyl sulfate
S/N . . . . . signal-to-noise ratio
SNP . . . . . single nucleotide polymorphism
SSC . . . . . saline sodium citrate
STR . . . . . short tandem repeat
TAPS . . . . . 3-[tris(hydroxymethyl)methylamino]-1-propanesulfonic acid
TEMED . . . . . N,N,N',N'-tetramethylethylenediamine
TRAC . . . . . Transcript Analysis with Aid of Affinity Capture
tRNA . . . . . transfer RNA
TTE . . . . . tris-taps-EDTA
UV . . . . . ultraviolet
VTT . . . . . Valtion Teknillinen Tutkimuskeskus (Technical Research Centre of Finland)
µCAE . . . . . microfabricated capillary array electrophoresis
µDG . . . . . microfabricated droplet generator
µm . . . . . micrometer, or micron ($10^{-6}$ meters)
µM . . . . . micromolar ($10^{-6}$ moles per liter)

# List of Tables and Figures

# Acknowledgements

It has been alleged that behind every good man is a great woman. While being personally unable to verify this claim, I can attest that behind at least one struggling Ph.D. student is a small army of highly trained and well-appointed individuals—individuals who deserve the warmest, most sincere kind of gratitude. First and foremost, I must thank my parents, Joan and Dave Duberow, who played the roles of best friends, psychiatrists, and cheerleaders—in short, mom and dad—from 3,000 miles away these five years. Mom, thank you for being someone to vent my frustrations to, someone to laugh at life with, and someone willing to remind me of all the reasons not to quit grad school and become a clown. And Dad, thanks for taking more interest in my life out here in California than I often did.

To Grandpa, Grandma, Papa, and the rest of the family and friends back East, thanks for laying out the red carpet for me every time I came home—and for feigning an interest in molecular biology and microfluidics on those rare occasions when I actually wanted to talk about my research. Jim, thanks for being the best kind of brother a guy could ask for—and the most difficult kind of person to acknowledge without sounding mushy. Crazy Aunt Debbie, thank you for being my family link and daily companion here in the Bay Area for the first four years. I'll never forget the inventories, the road trips, and the late night conversations over Bailey's and Kahlua–although the exact content of these conversations escapes me at present. And of course, I would be damned in all eternity if I didn't acknowledge my partner Nick Osano, who encouraged me in times of failure, reveled with me in times of slightly less profound failure, and in the latter portion of my graduate career, tolerated my random midnight thesis-writing sessions with quiet understanding (or perhaps blissful indifference).

My experiences in the Mathies Lab have been rapturous at best and instructive at worst. In the past five years I have had the opportunity to work through some of the most frustrating challenges of my life in an atmosphere of scientific exploration and intellectual curiosity. For fostering this atmosphere I credit my thesis advisor, Richard Mathies. Thank you, Rich, for taking me in, putting up with my dramatic learning curve, and giving me a chance to succeed anyway.

However, the greatest breakthroughs of my graduate career were achieved with the day-to-day help of some of the most knowledgeable and creative thinkers I have ever met, beginning with the great Robert Blazej, the creator of PRS and probably the most intelligent, direct, and collected scientist I have ever had the privilege of working with. Robert still provides the example of mastery and professionalism to which I can only hope to aspire. Kanwar Singh, my partner in crime for the first three years, thank you for being there to help me turn the crank. Moreover though, thanks for being a good friend—and for bearing with me while I adapted to graduate life. The vast amount of debt I owe to Eric Chu is incalculable, not only for fabricating those beastly 96-lane chips every time I got a little too pressure-happy with the syringes, but also for introducing me to Romano's, the Cerrito Speakeasy, and along with Amanda, Yosemite. To my pho-po girls Sam, Numrin, and Stephanie—thanks for keeping the lab fun and tolerating the drama for the sake of science. To Nadia and Lotta, thanks for being there to bounce ideas, rants,

and stupid jokes off of on a nearly daily basis. Enjoy the notebooks, Nadia! Finally, I must gratefully acknowledge Jim Scherer, who after designing half of the devices I have used throughout my tenure at Berkeley, unwittingly became my inspiration for completing the mitochondrial sequencing project. Jim, I wish you good health, happiness, and all the rewards in life that, for all of the dedication and kind advice you have offered us in the lab over the years, you so richly deserve.

I owe my last remaining shred of sanity to the administrative proficiency of Mary Hammond, who in assisting me with paperwork, coordinating meetings, and counseling me in times of stress became a true friend and ally in the unending battle with bureaucracy. In this vein, the amazing staff in the College of Chemistry administrative offices—particularly Lynn Keithlin, Doty Valrey, and Aileen Harris—deserve tremendous thanks.

None of the comparative sequencing work I have completed would have been possible without the help of our collaborators at Johns Hopkins University, a motley cast of several very capable scientists, exactly zero of whom I have ever met. Dr. David Sidransky was instrumental in overseeing this collaboration, while Mohammad Hoque and Heike Bernert bravely faced the tissue banks of Johns Hopkins to provide samples for the PRS work. In later years, Samantha Maragh at NIST, as well as Santanu Dasgupta and Mariana Brait, were particularly gracious in their assistance with sample characterization and MitoChip-related inquiries. More locally, the pleasant professionalism and expertise of Jing Yi and Prof. Terry Speed in our own Statistics Department made for an excellent collaboration during the design of the automated PRS analysis software. Many thanks to you, Jing, for writing the core scripts that ultimately became "mtDNA_Analysis.m"—and for staring at PRS data almost as long as I did that year! Finally, despite our ill-fated collaboration with the department of otolaryngology at UCSF, I do wish to express my gratitude to Dr. Ivan El-Sayed, who helped me navigate the treacherous waters of human subjects protocol review, if only to be shot down by, of all things, our funding agency.

The microchip-based TRAC work spearheaded in this thesis was undertaken in collaboration with Hans Soderlund's group at VTT in Finland. In the six months I worked with Jari Rautio, the world's expert on all things TRAC, he proved in so many ways how deserving he was of his newly-awarded Ph.D. Many thanks to you, Jari, for introducing me to expression analysis and, perhaps more to the point, for the brief diversion from mitochondrial sequencing.

Finally, I must profess my undying feelings—which I *think* amount to gratitude—to my intellectual mentors at FHS and Behrend, the passionate and talented scientists who inspired my love of science at a young age: Candace Magraw, Gerry Drabina, Tom Spudich, Mary Chisholm, Chuck Yeung, and John Hall. I'll never forgive you guys.

# Chapter 1

## Chips and SNP's

-or-

## *I am Mito-Man!*

## 1.1 Historical Perspective on an Ancient Disease

Attacking the body from within by means of its own uncontrolled proliferation, cancer is a complex, elusive, and terrifying disease. Responsible for over 500,000 deaths per year, it has more than earned its place at the forefront of biomedical research, with more than $5.5 billion in cancer-related funding by the NIH in 2008 alone (Figure 1.1) [1-2]. While private and government agencies have mounted heroic efforts to understand and treat the over 200 documented forms of human cancer, the ultimate dream of a "cancer cure" remains unrealized, owing largely to our only recent ability to characterize the disease accurately.

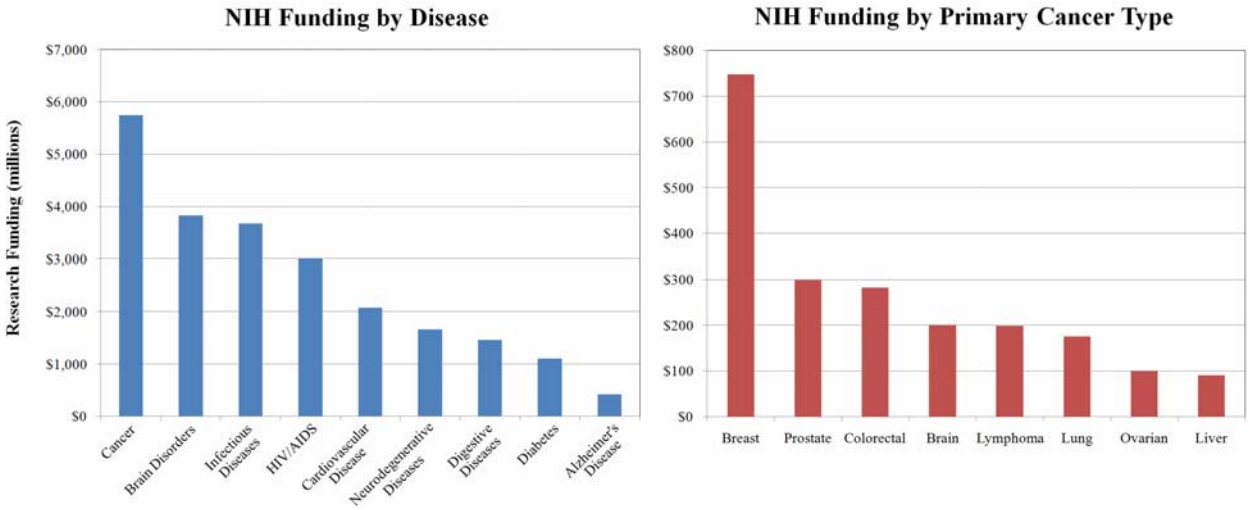



**Figure 1.1:** Relative NIH research funding for major diseases and cancer types in 2008. Garnering $5.75 billion, cancer research outspends all other disease classes by nearly $2 billion, with breast cancer receiving the greatest awards.

Although cancer has been a part of history since even the earliest records of human existence, our current understanding of its genetic causes and physiological changes has only taken shape within the past 50 years. For nearly 2000 years the origin of cancer was understood in the context of the Humoral Theory advanced by Hippocrates around 400 BC. Under this theory a local buildup of black bile, one of the four fundamental bodily fluids, was thought to result in tumor formation. Hippocrates' early theory was modified slightly, though not fundamentally altered, in the late middle ages with the speculation that a buildup and subsequent decomposition of lymphatic fluid was the base cause. This belief led noted surgeon John Hunter to begin prescribing the first surgeries to treat certain non-metastatic tumors in the 18th century. In 1838 Johannes Müller demonstrated that cancer was composed of diseased cells, although neither the origin nor the nature of the disease was clear. Over the next century further research revealed vague links between tumor development and environmental stresses such as chemicals, radiation exposure, and viral infection [3-4].

In 1953 groundbreaking work by James Watson and Francis Crick elucidated the molecular structure of DNA, which precipitated a deluge of discoveries regarding the nature of inheritance and the dependence of protein synthesis on the genetic code [5-6]. Genes responsible

for mediating such key biochemical processes as proliferation and specialization were identified, and alterations to these genes were found to be associated with abnormal growth. With greater understanding of the fundamental nature of cancer came novel treatments, generally in the form of biological or chemical agents directed toward tumor-specific genes or molecular targets [7-10]. Enhanced awareness of high-risk genotypes has also provided a platform for identifying genetically predisposed individuals and prescribing measures to preempt tumorigenesis [11-12].

While the scientific and medical communities have benefited greatly from an expanding knowledge base in the past 50 years, the ever-increasing rates of new cases—estimated at nearly 1.5 million nationally in 2009—serve as a sobering reminder that much remains to be learned about the onset, diagnosis, and treatment of this devastating disease. Cancer research remains the most-funded program for a single disease in the United States, with reports of novel therapies and discoveries of associated biological pathways consistently and prominently featured in scientific discourse. With the rapid emergence of new technologies for genetic profiling, an era of uniformly facile and effective cancer treatment may not be far on the horizon.

### 1.2 Genetic Basis for Tumorigenesis

Cancer-associated biomolecular variants wear a variety of costumes. The cells that comprise each organ, structure, or circulating fluid all contain unique biochemical signatures consisting of gene expression, protein modification, and metabolite production. Cancer is marked by a disruption in this fingerprint of expression; however the biochemical alterations in each cancer type are as unique as the cells themselves. Direct proteomic analysis has shown some success in uncovering consistently modified expression profiles in certain cancers [13-15]. Mostly, however, these effects share the underlying cause of genetic anomalies in two major classes of genes controlling cell growth, replication, and life span. Oncogenes, overactive or overexpressed forms of naturally occurring genes regulating cell growth and division (proto-oncogenes), promote tumor growth through an increase in their natural function. The first gene in this class was conclusively linked to human cancers by Robert Weinberg and Geoffrey Cooper in 1982 and was the first characterized member of the RAS family, a class of oncogenes responsible for cell proliferation, differentiation, adhesion, and migration [16]. One well-established mechanism invoking RAS-family proteins, the Ras-MAPK pathway, regulates replication by responding to external growth factors and triggering a cascade of protein kinases that ultimately facilitate the phosphorylation of the retinoblastoma protein (Rb). Deactivation of this protein allows the cell to pass through the $G_1$ restriction checkpoint and therefore, to replicate [17]. A hyperactive RAS protein may thus promote an otherwise disfavored replication event, resulting in excessive growth. Conversely, tumor suppressor genes, if mutated or underrepresented, promote tumor growth by loss of activity. Most infamous among this class of genes is TP53, the single most commonly altered gene observed in human cancers [18-19]. Originally believed to be an oncogene, TP53 plays a prominent role in confining the growth of damaged cells. In its active form, the p53 protein promotes the expression of p21, a protein that deactivates the CDK2 complex responsible for phosphorylating Rb. Thus, loss of p53 function produces unrestricted cell growth [20]. Changes in protein function may have origin in a number of biochemical factors and often arise from interaction with other oncogenetic species. Human papillomavirus, for example, expresses oncoproteins that inhibit both p53 and Rb, itself a tumor

suppressor [21-22]. Predominantly however, these functional changes are linked to structural alterations arising from damage to the encoding gene.

Localized genetic damage was suspected to play a role in tumorigenesis before the structure of DNA and the nature of replication were fully understood. In 1914 Theodor Boveri articulated a theory explaining tumor development in terms of somatic tissue deviating from the germline through abnormal chromosome arrangement [23]. It was not until the 1960's, however, that such chromosomal changes could be identified and studied precisely. The first cancer-associated genetic anomaly to be characterized rigorously was reported in 1973. The Ph, or "Philadelphia" chromosome, results from a reciprocal translocation of chromosomes 9 and 22 and has been observed in 90% of chronic myelogenous leukemia patients [24]. Translocations such as the Philadelphia abnormality involve the exchange of large sections of genetic code between chromosomes. A number of other long-range genetic alterations including rearrangements, insertions, and deletions can affect large DNA fragments and often fundamentally alter—or completely abolish—gene activity. Other more subtle changes in genetic structure are also possible: for example, a common regulatory mechanism for otherwise normal genes is methylation of noncoding promoter regions, which determine the extent to which a gene is expressed. Excessive methylation of promoter regions is capable of diminishing gene activity and in extreme cases may silence the gene completely [25].

The minutest form of genetic alteration is the point mutation, a change in a single nucleotide identity within the sequence. Despite their relatively modest structural effects, point mutations are highly prevalent and are capable of fundamentally altering biomolecular processes. A single-base substitution occurring within a codon for a key amino acid can substantially change the structure of the protein, dramatically altering its function. More severely, an insertion or deletion of one or two bases shifts the frame of translation, resulting in a completely different protein following the mutation point. Such single-base changes have been associated with a variety of disorders, both cancer-related and otherwise. For example, sickle-cell disease is caused by a single point mutation in the β-hemoglobin gene [26], while a deletion of three base pairs in the CFTR gene has been documented to cause 70% of all cystic fibrosis cases [27].

A certain proportion of genetic abnormalities are inherited in the germline, resulting in haplotypes that may predispose individuals to genetic diseases such as cancer. However it is not until this abnormal gene achieves dominance, typically through mutation or inactivation of the complementary "normal" chromosome, that cancer develops. In many cases, the causes of this acquired damage is environmental, the result of high-energy radiation or chemical carcinogens that stretch, break, or modify the DNA structure (Figure 1.2). Radiation damage from ultraviolet (UV) exposure most commonly takes the form of dimers between adjacent pyrimidines in the sequence, which can pair improperly during replication [28]. Chemical carcinogens come in a variety of forms but act predominantly by reacting directly with nucleobases, forming base adducts that pair improperly or contort the helix. One common class of carcinogens, the polycyclic aromatic hydrocarbons (PAH's), are themselves inert but form reactive metabolites *in vivo*. The metabolism of benzo[*a*]pyrene, a well-documented PAH found in cigarette smoke, involves the hydroxylation of the substrate via a double-epoxidation reaction catalyzed by cytochrome P450. The 7,8-diol 9,10-epoxy intermediate in this process has been shown to be carcinogenic, forming adducts with electron-donating nucleobase groups [29].

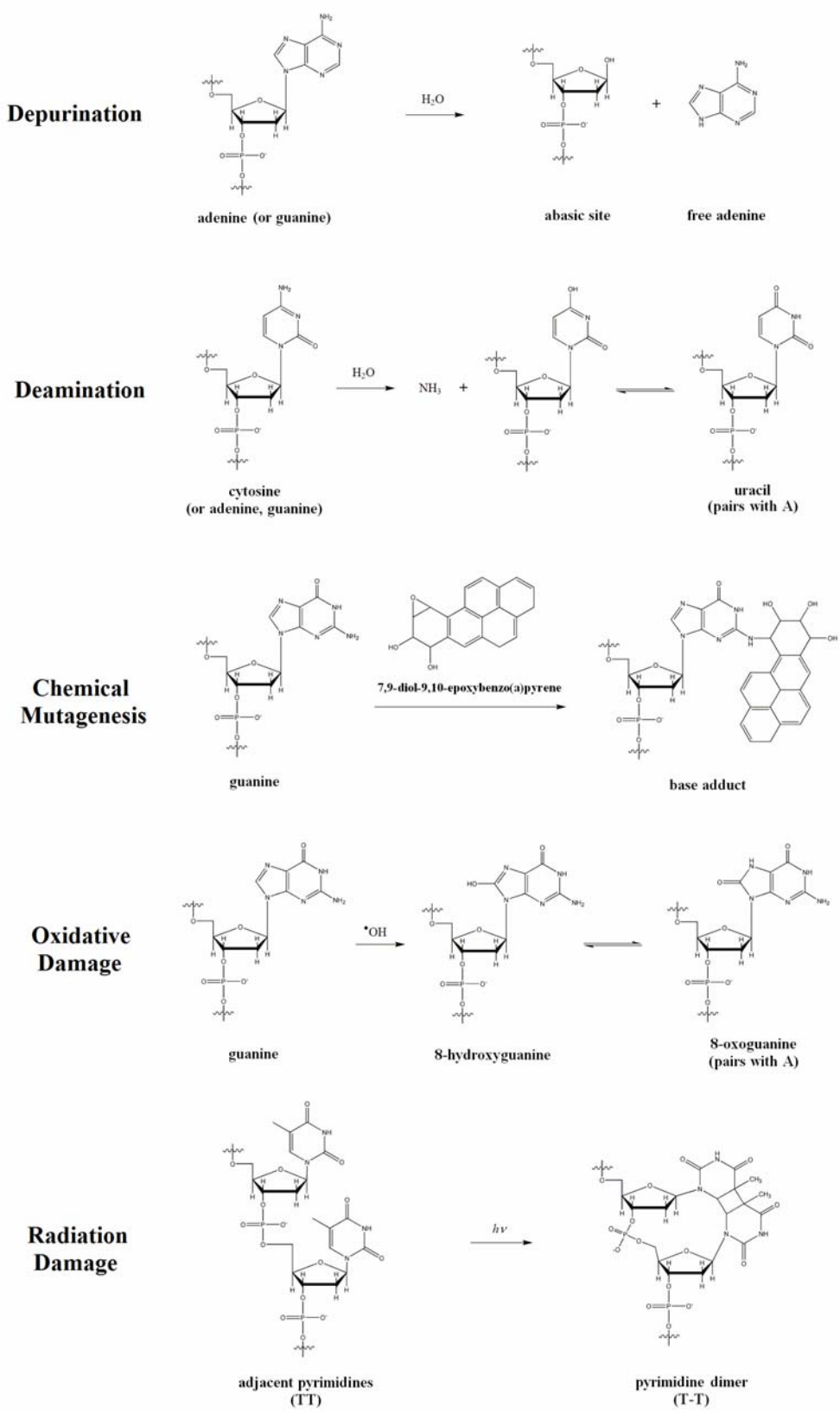



**Figure 1.2:** Examples of structural changes in some common DNA mutation classes.

In addition to these external mutagens, a number of species capable of damaging the structural integrity of DNA occur naturally as by-products of endogenous biochemical processes. Most notable among these are the reactive oxygen species (ROS) produced during respiration (Figure 1.3). Upon direct reaction with DNA, these ROS create base analogs that may be recognizable as nucleobases but pair differently with the complementary strand, producing mutations in daughter strands during replication [30].

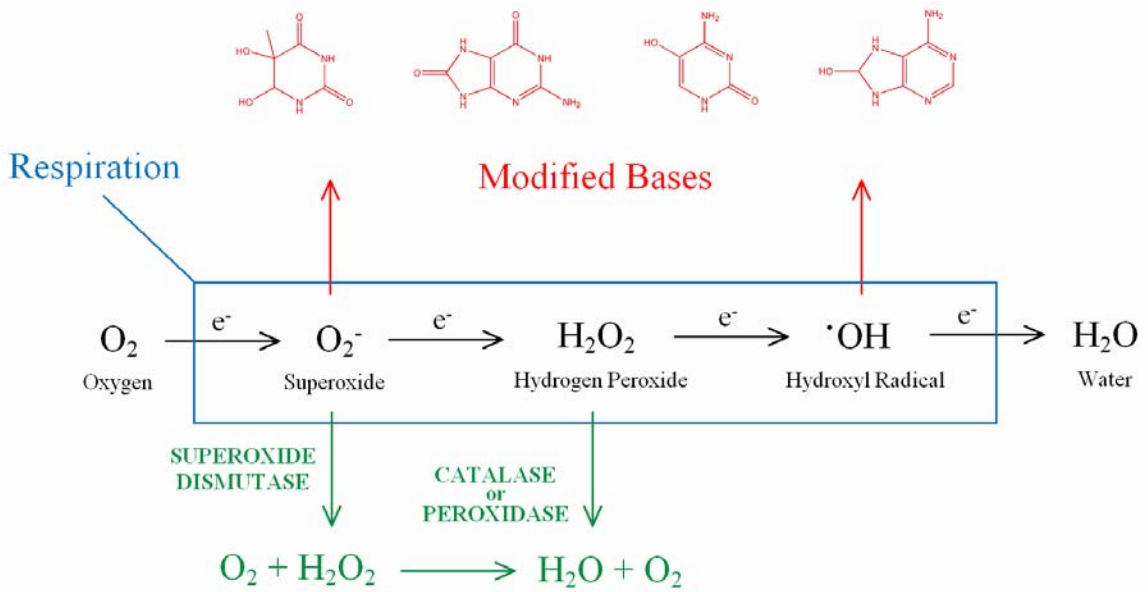



**Figure 1.3:** Generation of reactive oxygen species during respiration. ROS are produced by incomplete reduction of molecular oxygen at key electron transfer steps in the mitochondrial respiratory chain. These ROS react with DNA to form oxy- and hydroxyl- derivatives of natural nucleobases, often altering the pairing properties of these bases.

The cell's own transcription machinery provides a final layer of instability, occasionally producing misincorporations, slippage, and other polymerase errors which can become exacerbated by unchecked replication. In these cases, indeed in all cases of biomolecular mutation, the cell contains proofreading enzymes, coding redundancies, and repair mechanisms to minimize the effects of the damage; however it is the double-failure of the gene's structural integrity and the cell's defense systems that allows a malignant gene to achieve dominance.

Each of the mutation classes discussed above has been associated with at least one form of cancer, and these biochemical targets have been exploited for the purposes of molecular pathology. The advent of the polymerase chain reaction (PCR) for gene amplification has proven especially fortuitous in this vein, and highly sensitive and specific PCR assays have since been developed to probe for chromosomal changes, point mutations, and promoter alteration [31-33]. A particularly robust marker that lends itself well to PCR-based detection is microsatellite instability, or changes in number of short repeating motifs, or microsatellites, in a local sequence. Microsatellite alterations accumulate in genetically unstable environments such as cancer and can be detected readily by PCR amplification [34-35]. Although PCR-based assays are highly

sensitive, their widespread use in the discovery of novel cancer biomarkers is limited by their specificity. More general mutation information may be obtained by direct genomic sequencing; however conventional sequencing techniques are both intensive and expensive. As a more focused alternative, hybridization-based microarrays have shown promise as diagnostic monitors of the expression of multiple genetic targets [36-38].

The array of cancer-implicated genes is vast, and expression profiles are highly variable based on cancer type and stage. In response to an expanding wealth of information, a number of cancer variant databases have been developed [19, 39]. However these repositories, like our understanding of the fundamental causes of cancer, are incomplete, and further work is necessary to reveal the complete molecular fingerprint of tumor progression. To this end, the NIH has conceived an organized, statistical effort to determine the complete genetic signature of the full spectrum of human cancers. In its most ambitious form, the proposed "Cancer Genome Atlas" approaches this challenge through complete genomic sequencing of a statistically relevant pool of varying cancer types in an endeavor akin to thousands of Human Genome Projects. A pilot study of brain, lung, and ovarian cancers was launched by the NCI and NHGRI in 2005, and preliminary results from a multidimensional analysis of 206 glioblastomas were released in 2008 [40-41]. The results confirm the power of thorough genetic analysis in a wide sample base. Unfortunately, the prohibitive cost and time investment of DNA sequencing require that further improvements in comparative sequencing technology be developed and implemented to bring such full-scale mutation analysis within the realm of feasibility.

### 1.3 Mitochondrial Implication in Cancer Development

The most complex and diverse of the organelles, the mitochondrion plays both a vital and a deadly role in the life of the cell. Long known as the "powerhouse of the cell," the mitochondrion contains the machinery of electron transport, a necessary step in respiration, the oxidative process by which over 80% of the cell's energy is supplied. Beyond this most recognized role, however, the mitochondrion also participates in the synthesis of biomolecular building blocks, most notably amino acids, phospholipids, and heme, and possesses the power to commit suicide in cases of extreme cellular damage [42]. Mitochondrial dysfunction has been associated with a variety of degenerative diseases, and its role in cancer progression has been a topic of particular scrutiny in recent years.

The primary role of the mitochondrion is energy production through oxidative phosphorylation (OXPHOS) of adenosine diphosphate (ADP). The process of mitochondrial electron transport is reviewed in Figure 1.4(*A*) and consists of a series of oxidation/reduction reactions catalyzed by four enzymes, or complexes, bound to the inner mitochondrial membrane. Beginning with the oxidation of NADH, an electron carrier from upstream metabolic reactions, electrons are shuttled through the chain of complexes, as well as membrane-bound electron carriers coenzyme Q and cytochrome *c*. The chain terminates at complex IV with the reduction of molecular oxygen to water. During the process, complexes I, III, and IV actively transfer protons from the matrix to the intermembrane space. These "proton pumps" maintain an excess of protons in the intermembrane space, producing an electrochemical gradient across the inner membrane. The final complex in the respiratory chain uses this membrane potential to drive ATP synthesis through a cycle of conformational changes driven by a molecular motor.

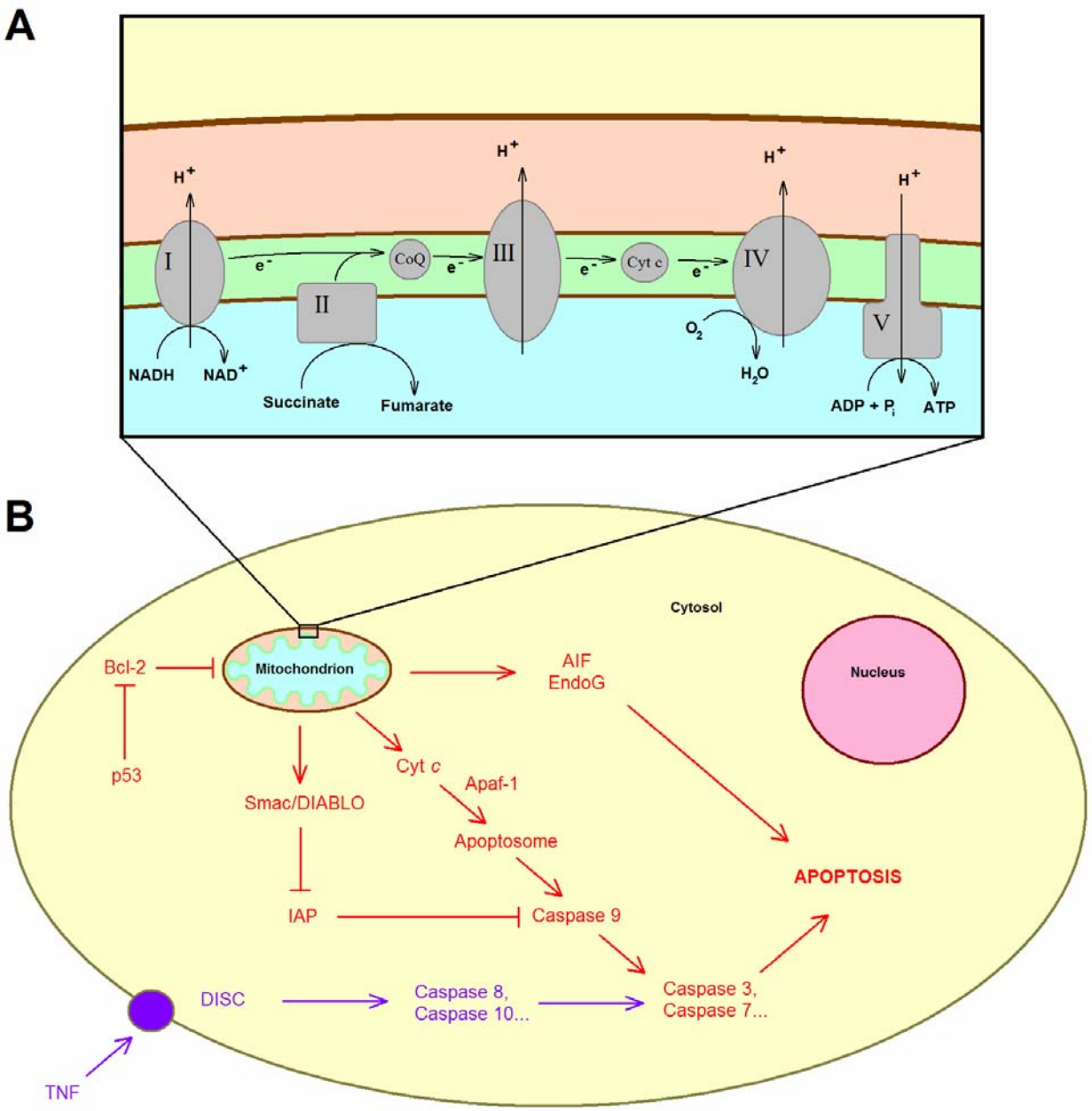


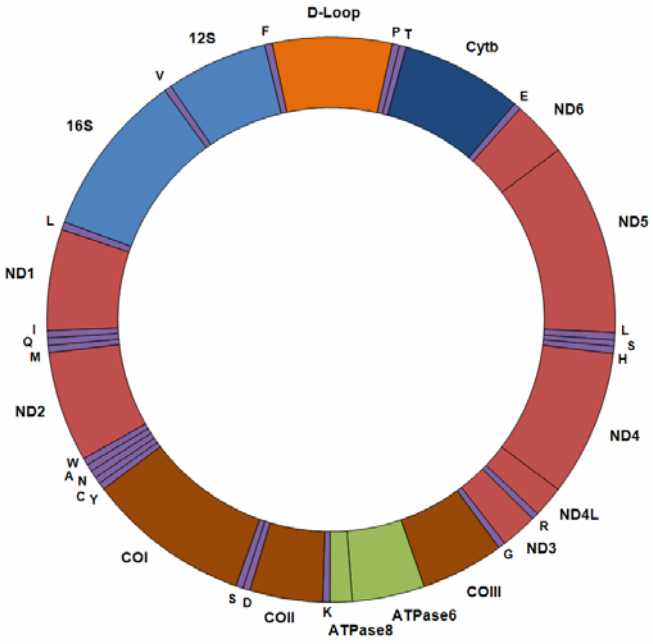

**Figure 1.4:** Mitochondrial roles in energy production and apoptosis. (*A*) A series of oxidation/reduction reactions facilitated by inner-membrane-bound complexes transfers electrons from NADH to oxygen as protons are pumped into the intermembrane space. The resulting electrochemical gradient drives generation of ATP at the final complex, providing the majority of the energy used in the cell. (*B*) Responding to triggers within the cell, pro-apoptotic factors are released from the intermembrane space. These enzymes activate a series of caspases which ultimately destroy the cell. The internal, mitochondria-dependent apoptosis pathway is indicated in red and is independent from the external, growth-factor-dependent pathway denoted in violet.

In performing its vital duties in the cell, the mitochondrion finds itself as the centerpiece of one of nature's more paradoxical quirks. In addition to facilitating oxidative phosphorylation and thus providing the majority of the cell's energy, the mitochondrion is also home to AIF, SMAC/DIABLO, and several other enzymes capable of initiating apoptosis, the natural programmed death pathway invoked under circumstances of environmental stress or irreparable cell damage (Figure 1.4(*B*)). Innocuous when confined in the mitochondrial intermembrane space, these enzymes are released into the cytosol upon stimulation of upstream apoptotic factors in a cascade known morbidly as the "death signal." A key step in this triggering pathway is an accumulation of the p53 tumor suppressor protein in response to excessive genetic damage. The p53 protein has been shown to interact with the Bcl-2 family of genes, the main inhibitors of mitochondrial membrane permeation. Once liberated, these mitochondria-bound proteins trigger the activation of a series of caspases, or apoptotic enzymes. Interestingly, the caspase activation pathway commences with the binding of Apaf-1 to cytochrome *c*, the electron carrier between complexes III and IV during normal OXPHOS. The cyt*c*-Apaf1 complex, the apoptosome, begins the caspase cascade by activating initiator caspase 9, which in turn activates downstream caspases such as 3 and 7, the enzymes that directly degrade cell contents. Independently, AIF and endonuclease G act to degrade DNA upon release from the mitochondria [43]. These apoptotic processes play a vital role in controlling tissue growth and eliminating damaged cells. Interruption of this pathway results in the accumulation of diseased, mutated, or otherwise damaged "zombie" cells and has been identified as a key factor in tumorigenesis.

Given its important and multi-faceted role in the cell's biochemical network, it is not surprising that mitochondrial dysfunction has been linked to aging and a variety of metabolic, cardiovascular, and degenerative diseases [44-45]. The first steps in linking such defects to tumorigenesis were made by Otto Warburg, who in 1956 proposed damage to the respiratory pathway as a key event in cancer development. Warburg further postulated that such stress would be marked by increased glucose consumption, as less efficient glycolysis pathways upstream of OXPHOS adapt to satisfy the energetic needs of the cell [46]. This increased glycolytic activity, even in the presence of oxygen, allows the neoplastic cell to circumvent normal OXPHOS, meet the high energetic needs of the cell, and thrive under stress. In later years this "Warburg Effect" has been observed experimentally, linked to specific genetic variants, and even exploited as a potential target for anticancer therapy in tumors resistant to traditional chemotherapies [47-48]. Subsequent efforts have linked observable changes such as altered respiratory chain content or increased ROS production to cancerous tissue [49-50]. However the bulk of studies linking mitochondrial damage to cancer have focused on genetic damage within the mitochondrion itself.

## 1.4 The Mitochondrial Genome

By far most of the 87 proteins contained in the mitochondrial OXPHOS complexes are encoded by the genomic DNA. These genes are transcribed in the nucleus, and the resulting proteins are transported into the mitochondrion. The organelle does, however, contain its own genome, an artifact of its ancestral roots as a bacterial infection-turned-symbiosis. The mitochondrial DNA (mtDNA) is a compact, circular genome consisting of 37 genes distributed across 16,569 bp (Figure 1.5). It encodes 13 proteins, all of which form functional subunits in mitochondrial OXPHOS enzymes, 22 tRNA's, and two rRNA's. A noncoding "displacement loop" (or D-loop) containing promoters for transcription accounts for the majority of all variation in human mtDNA [51-52].




| Abbreviation | Gene | Complex | Base Position |
|---|---|---|---|
| D-Loop | Displacement Loop | --- | 16024 - 576 |
| F | tRNA Phenylalanine | --- | 577 - 647 |
| 12S | 12S Ribosomal RNA | --- | 648 - 1601 |
| V | tRNA Valine | --- | 1602 - 1670 |
| 16S | 16S Ribosomal RNA | --- | 1671 - 3229 |
| L | tRNA Leucine | --- | 3230 - 3304 |
| ND1 | NADH Dehydrogenase Subunit 1 | I | 3307 - 4262 |
| I | tRNA Isoleucine | --- | 4263 - 4331 |
| Q | tRNA Glutamine | --- | 4329 - 4400 |
| M | tRNA Methionine | --- | 4402 - 4469 |
| ND2 | NADH Dehydrogenase Subunit 2 | I | 4470 - 5511 |
| W | tRNA Tryptophan | --- | 5512 - 5579 |
| A | tRNA Alanine | --- | 5587 - 5655 |
| N | tRNA Asparagine | --- | 5657 - 5729 |
| C | tRNA Cysteine | --- | 5761 - 5826 |
| Y | tRNA Tyrosine | --- | 5826 - 5891 |
| COI | Cytochrome c Oxidase Subunit I | IV | 5904 - 7445 |
| S | tRNA Serine 1 | --- | 7446 - 7514 |
| D | tRNA Aspartic Acid | --- | 7518 - 7585 |
| COII | Cytochrome c Oxidase Subunit II | IV | 7586 - 8269 |
| K | tRNA Lysine | --- | 8295 - 8364 |
| ATPase8 | ATP Synthase F0 Subunit 8 | V | 8366 - 8572 |
| ATPase6 | ATP Synthase F0 Subunit 6 | V | 8527 - 9207 |
| COIII | Cytochrome c Oxidase Subunit III | IV | 9207 - 9990 |
| G | tRNA Glycine | --- | 9991 - 10058 |
| ND3 | NADH Dehydrogenase Subunit 3 | I | 10059 - 10404 |
| R | tRNA Arginine | --- | 10405 - 10469 |
| ND4L | NADH Dehydrogenase Subunit 4L | I | 10470 - 10766 |
| ND4 | NADH Dehydrogenase Subunit 4 | I | 10760 - 12137 |
| H | tRNA Histidine | --- | 12138 - 12206 |
| S | tRNA Serine 2 | --- | 12207 - 12265 |
| L | tRNA Leucine 2 | --- | 12266 - 12336 |
| ND5 | NADH Dehydrogenase Subunit 5 | I | 12337 - 14148 |
| ND6 | NADH Dehydrogenase Subunit 6 | I | 14149 - 14673 |
| E | tRNA Glutamic Acid | --- | 14674 - 14742 |
| Cytb | Cytochrome b | III | 14747 - 15887 |
| T | tRNA Threonine | --- | 15888 - 15953 |
| P | tRNA Proline | --- | 15956 - 16023 |

**Figure 1.5:** Organization of the mitochondrial genome. The highly compact, 16,569-bp mtDNA encodes 37 genes, including 13 proteins vital to oxidative phosphorylation.

The mitochondrial genome is attractive as a target for genetic analysis, both clinical and otherwise, for several reasons. Its relatively small size renders it an easy target for complete sequencing. Its high concentration—hundreds to thousands of copies per cell—allows for PCR amplification in the absence of abundant sample. This property also presents the phenomenon of heteroplasmy, genetic variation between mtDNA copies contained within the same cell. Traced spatially or temporally, drifts in heteroplasmy provide insight into the mutation process and the acquisition of dominance. Additionally, because of its independence from the nucleus, the

mtDNA is strictly maternally inherited, which makes the genome a particularly robust singular marker for ancestral determination and forensic identification [53-54].

A considerable body of literature exploring the frequency and consequences of mtDNA damage has been growing rapidly over the past fifteen years. Despite a several-hundred-fold redundancy within each cell, the mitochondrial genome is, on average, over ten times more susceptible to alteration than the nuclear genome, with an approximate germline mutation rate of 1% per million years [55]. This enhanced mutation rate is largely attributed to a relative lack of protection and repair within the mitochondria. Whereas nuclear-bound chromosomes are wound around protective histones which inhibit chemical modification of genetic material, mitochondria contain no such protective structures. Additionally, mitochondria possess only minimal repair mechanisms, predominantly in the form of base excision pathways directed toward specific base adducts. Nucleotide excision repair appears not to be present at all in mammalian mitochondria [56-57]. However most importantly, mitochondrial genes are especially sensitive to oxidative damage due to their proximity to electron transport. These processes expose the mtDNA to the reactive oxygen by-products of OXPHOS before they can be degraded enzymatically or diffuse through the cytosol, resulting in a higher mutation rate from endogenous damage than in the nucleus. Functional damage to mitochondrial genes is particularly harmful, as mutations of this type directly alter the function of OXPHOS subunits. Debilitating mutations of this type often impede electron transport, resulting in further ROS accumulation. Increased ROS production from a backfiring electron transport chain increases the likelihood of further genetic damage, thereby creating a loop of concurrent ROS and mutation accumulation. Indeed, this accumulation has been associated with aging and is believed to play a role in the progression of such degenerative diseases as Alzheimer's and Parkinson's [44, 58-60].

The relatively high mutation rate observed in the mitochondrial genome, combined with demonstrated mitochondrial implication in the metabolic and apoptotic changes observed in cancer, has served as the motivation for a large body of research centered on mtDNA variation. The exploitation of the mitochondrial genome as a potential cancer biomarker was spearheaded in 1998 by Polyak *et al*., who used complete mtDNA sequencing to monitor somatic changes in colorectal cancers. Since this initial work, considerable progress has been made in identifying cancer-associated mtDNA changes. Certain mitochondrial haplotypes have been shown to predispose individuals to breast, prostate, and renal cancers [61-62]. However, most attention has been devoted to single base changes or large-scale insertion/deletion mutations in cancer patients. While many of these variants appear to be randomly scattered throughout the genome, a handful of recurring mutation sites have become apparent over time. A common "hot spot" appears in the D310 region, a run of 12 cytosine residues in the D-loop interrupted by a single thymine [63-65]. This variant is particularly interesting, as the D310 sequence forms part of the complex that initiates heavy strand replication. Another common mutant is a 4977 bp deletion affecting ATPases 6 and 8, NADH dehydrogenase subunits 3-5, and COXIII, as well as 5 tRNA's. This alteration, the so-called "common deletion," has been observed in a variety of cancers and genetic diseases [66-67].

With the increased availability of high-throughput resequencing technologies, the field of tumor-associated mtDNA variation has grown rapidly in the past decade, beginning in earnest with the work of Fliss *et al*. in 2000 [68]. Using a direct sequencing approach, authors compared mtDNA (80% coverage) from 41 primary bladder, lung, and head and neck tumors to paired normal mtDNA from blood. When available, extracts of healthy tissue and associated bodily

fluids were also sequenced. In total, 292 sequence variants were identified, 39 of which were acquired mutations. At least one such somatic variant was present in 64% of bladder, 43% of lung, and 46% of head/neck tumors, with altered genomes containing between one and four mutations. Though not universally so, somatic mutations were largely detectable in associated bodily fluids (100% of urine, 80% of bronchoalveolar lavage, and 67% of saliva), thus suggesting the feasibility of mtDNA variation as a target for noninvasive cancer detection.

Since taking flight as a substantial field of study, a considerable degree of mtDNA variation has been observed in a wide range of cancers, including breast [66, 69-70], kidney [71], liver [72-73], colorectal [74-75], ovarian [76], gastric [77], esophageal [78], melanoma [79], pancreatic [80], thyroid [81-82], and brain [83]. Further work independent of Fliss *et al*. has uncovered additional variants in head and neck [84-85], bladder [86], and lung [67, 87-88] tumors as well. While a number of these studies have focused on tracking individual mutations or genes, several have approached the issue by sequencing the entire mitochondrial genome *ab initio* to identify any and all somatic variation. A survey of selected full-mtDNA studies is summarized in Table 1.1. In general, full mitochondrial sequencing of tumor tissue revealed variation in 40 to 80% of all individuals, with mutated mtDNA bearing typically one to three somatic single-base changes. The noncoding D-loop is almost universally the most common hot spot for mutations, with NADH dehydrogenase, cytochrome *c* oxidase, or ribosomal RNA genes bearing the majority of the coding mutations, depending on the study. Particularly noteworthy among these data are the findings of Mithani *et al*., suggesting an average of eight or more somatic mutations in unique cases of melanoma, with individual mutation loads as high as 42. While it may be tempting to draw conclusions based on these compiled data, the studies are too few, the sample sets are too small, and the available technology is too error-prone to assign specific mutation loci to individual cancers with any certainty. The wide range in the reported frequency of heteroplasmic mutations in these studies (17-83%) is especially troubling, indicating considerable variability between assays.

| Cancer Type | Sample Size | % of Patients with at Least One Mutation | Average # of Mutations | Total Mutations | % Heteroplasmic | Genes Affected | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | | | D-Loop | tRNA | 12S | 16S | ND1-6 | COI-III | ATP 6 | Cytb |
| Brain [83] | 15 | 40% (6 individuals) | 3 | 18 | 56 | 11 | 3 | --- | --- | 1 | 3 | --- | --- |
| Breast [70] | 19 | 74% (14 individuals) | 1.9 | 27 | 56 | 22 | --- | --- | 1 | 3 | --- | 1 | --- |
| Colorectal [74] | 10 | 70% (7 individuals) | 1.7 | 12 | 17 | --- | --- | 1 | 3 | 3 | 3 | --- | 2 |
| Kidney [71] | 8 | 63% (5 individuals) | 1.2 | 6 | 83 | 2 | --- | 1 | 1 | 2 | --- | --- | --- |
| Lung [87] | 11 | 45% (5 individuals) | 1.6 | 8 | 25 | 2 | 1 | 1 | 1 | 1 | 1 | --- | 1 |
| Melanoma [79] | 16 | 75% (12 individuals) | 8.3 | 100 | ----- | 25 | 5 | 3 | 7 | 39 | 6 | 4 | 11 |
| Ovarian [76] | 10 | 60% (6 individuals) | 1 | 6 | 33 | 2 | --- | 2 | 1 | --- | --- | --- | 1 |
| Pancreatic [80] | 5 | 80% (4 individuals) | 1 | 4 | 25 | --- | 1 | --- | 1 | 2 | --- | --- | --- |

**Table 1.1:** Results of selected full-mtDNA sequencing studies for various cancer types. Mutations in the mtDNA are observed in 40-80% of samples, with the majority of mutations occurring in the noncoding D-loop. Heteroplasmy data are not available for reference 79.

The recent explosion of mtDNA mutation data has led to a somewhat more concerted effort to coordinate new findings in the form of an online database known as MitoMap (http://www.mitomap.org/) [89-90]. A summary of cancer-associated mutation sites reported in this database, grouped by major gene location, is compiled in Figure 1.6. By far, the most commonly mutated region of the genome is the displacement loop. Spanning only 6.8% of the

genome, the D-loop accounts for over 24% of reported mutation sites. The high mutation frequency is commonly explained by a lack of selective pressure on noncoding regions; however, the presence of regulatory transcription factors in this highly variable region may be related to the observation of altered mtDNA content in cancers [91]. Mutations in the coding region are in general distributed evenly among the 15 major genes, with the exception of the complex I, or NADH dehydrogenase, genes. While comprising a combined 46% of the major gene span within the mtDNA, the seven mitochondrial NADH dehydrogenase subunits account for over 60% of the mutation sites within these genes. ND1, ND3, and ND4L and ND4 are particularly noteworthy mutation hot spots, each exhibiting a disproportionately high mutation rate relative to its size. Although NADH dehydrogenase contains a total of 45 subunits, damage to the seven mtDNA-encoded subunits may significantly impair the function of this vital complex, which combined with complex III, accounts for the highest production of ROS in the OXPHOS machinery [92].

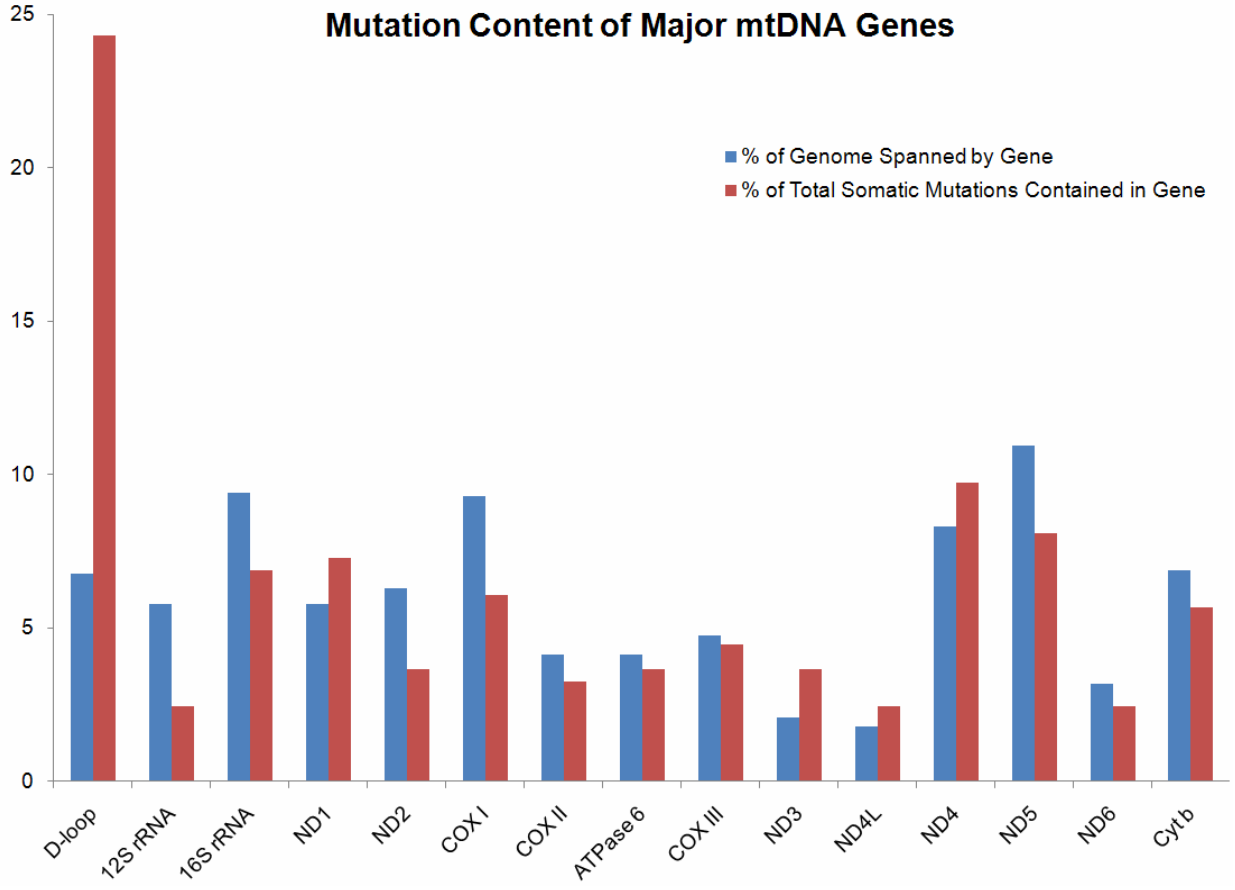



**Figure 1.6:** Relative sizes and cancer mutation frequencies for major noncoding and coding regions. The D-loop by far bears the highest and most vastly disproportionate frequency of reported cancer-associated mutations, although complex I genes also appear to be mutated at an unusually high frequency.

Despite the apparently striking frequency of mtDNA variation observed in tumor tissue, the functional significance of mtDNA mutations as a class has not been established unequivocally. In recent years focused studies have succeeded in identifying pathological links between specific mtDNA variation and tumorigenesis. Very recently an artificially overexpressed cytochrome *b* protein containing a 21 bp deletion was found to increase ROS production in a tumor cell line and normal splenocytes. This effect was accompanied by an increase in glycolysis and, in the case of the cancer cell line, increased tumor growth [86]. Months later, mtDNA containing mutations in the ND6 gene from a highly metastatic cell line were found to transfer metastatic potential to otherwise non-metastatic tumor cells in mice. Metastatic activity was decreased by prior treatment of the metastatic tumors with ROS scavengers, suggesting a direct link with ROS generation associated with a mtDNA mutation [93].

Given the variety of known tumor-associated mtDNA variations, it is not clear, nor will it be quickly determined, which of these mutations possess tumorigenic potential. Common in the mutation studies conducted thus far, however, is the observation that tumor-associated mutations are largely homoplasmic, with all mtDNA copies within the cell containing the identical basecall. This observation is in contrast to aging-associated mutations, which appear to be largely random and heteroplasmic. The observation of homoplasmy in such a vast proportion of tumor-associated mutations is highly suggestive of pathological relevance, either through direct tumor causation or through adaptation to thrive under stress [94].

## 1.5 Polymorphism Ratio Sequencing as a Tool for Quantitative Cancer Detection

Given the centrality of genetic variation in the onset and progression of cancer, technology for rapid, inexpensive DNA sequencing holds the key to unlocking the complete biomolecular signature of the disease. The method that has historically demonstrated greatest success in generating DNA sequencing data was introduced by Frederick Sanger in 1977 [95]. Sanger's chain termination-based method makes use of a polymerase reaction using chemically modified nucleotides to generate a series of "extension fragments," or partial copies of a template DNA sequence, each labeled and differentiable based on the identity of the terminating base. These fragments are then purified and separated based on size by electrophoresis through a viscous solution of polymer, typically 2-5% (w/v) linear polyacrylamide (LPA). The sequence of the original template strand is read by identifying the terminal base of each fragment from smallest to largest.

In the Sanger extension phase, template DNA is combined with a thermally stable DNA polymerase and a short oligonucleotide primer designed to hybridize selectively to a target sequence in the template marking the beginning of the region to be sequenced. The mixture is then diluted in a buffer containing the four deoxyribonucleotide triphosphates (dNTP's) and a limiting concentration of a particular dideoxyribonucleotide triphosphate (ddNTP). These ddNTP's lack the 3' hydroxyl group necessary for chain propagation, and thus incorporation of a ddNTP terminates the polymerization reaction. Random incorporation of ddNTP's results in a series of fragments, one for each occurrence of the target base in the template sequence. In general, analysis of a complete sequence requires the preparation of four reactions, each generating fragments with a different terminating base.

In its earliest form, Sanger extension fragments were radiolabeled by incorporation of a $^{32}$P-tagged dNTP and separated on a slab gel, yielding roughly 100 bases of data per sample. Protocols for extension using fluorescently labeled primers were developed in 1986, and the use of multiple fluorescent tags allowed for separation of all four extension reaction products on a single gel [96]. The next year, introduction of ddNTP's bearing unique fluorescent dyes facilitated the further multiplexing of the extension reactions [97]. Sensitivity of fluorescence-based sequencing was dramatically increased with the development of Förster energy-transfer dyes, which allow for efficient single-wavelength excitation of multiple fluorophores with a range of maximal emissions [98]. Combined with the advent of capillary electrophoresis (CE)-based separations, automated sample manipulation, and data processing technology, these improvements have revolutionized Sanger sequencing. State-of-the-art sequencing instruments are now fully automated and yield continuous, four-color read lengths of up to 1000 bases [99]. A typical contemporary Sanger sequencing scheme is detailed in Figure 1.7.

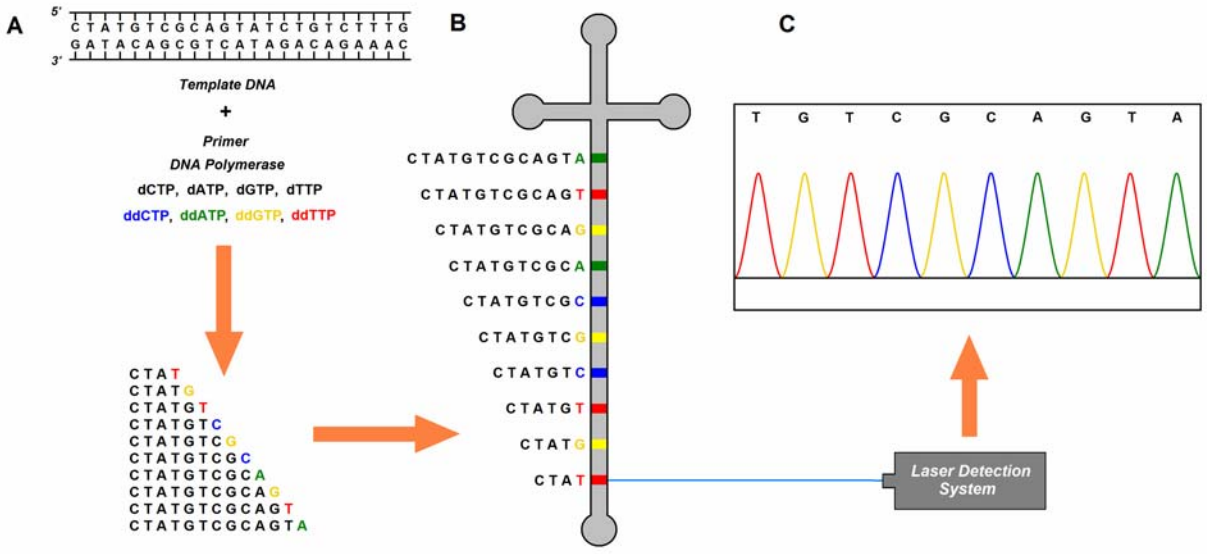



**Figure 1.7:** Modern Sanger sequencing with dye terminators. (*A*) A polymerase-catalyzed reaction generates a series of partial copies of the template DNA strand, with each copy bearing a fluorescent label indicating the terminating base. (*B*) These fragments are separated electrophoretically, and (*C*) the template sequence is read using the four-color electropherogram.

Despite being somewhat complicated both conceptually and practically, Sanger sequencing has generated gigabases of novel sequencing data and was vital in the completion of the Human Genome Project in the early 2000's [100]. Although novel technologies have been advanced to simplify and decrease the cost of DNA sequencing [101-104], Sanger extension remains, if perhaps not for long, the gold standard by virtue of its balance of read length, accuracy, and cost.

While detection of genetic variation is possible by direct comparison of conventional sequencing data, such an approach is problematic in that it relies on the assumption of lane-to-lane consistency as well as manual examination of largely overlapping traces. Direct sequencing also precludes the investigation of genetic heterogeneity within a sample, namely mitochondrial heteroplasmy. In 2003 Robert Blazej *et al*. addressed these concerns with the development of Polymorphism Ratio Sequencing (PRS), a novel Sanger extension-based technique for direct, quantitative comparative sequencing of two distinct DNA samples [105]. Recognizing that dye primer extension chemistry does not inextricably link any particular dye to a specific base or template genome, PRS employs a unique labeling scheme allowing for simultaneous electrophoretic separation of extension fragments from two samples (Figure 1.8). In separate reaction vessels, Sanger fragments corresponding to a particular terminal base are extended and labeled with distinct dyes identifying the template. These fragments are then pooled into a final sample mixture and separated, yielding an electropherogram that overlaps perfectly in all places, except in instances of genetic variation. Monitoring the squared difference between corresponding tracks provides a direct indication of variant position, producing a nonzero value only in regions of poor overlap. Using a standard four-color detection system, it is possible to monitor two bases simultaneously: therefore two samples (by convention A/C and G/T) are required to obtain a complete data set. Ultimately, PRS represents a shuffling of sequencing information for two genomes between two capillaries in a manner more appropriate for sequence comparison.

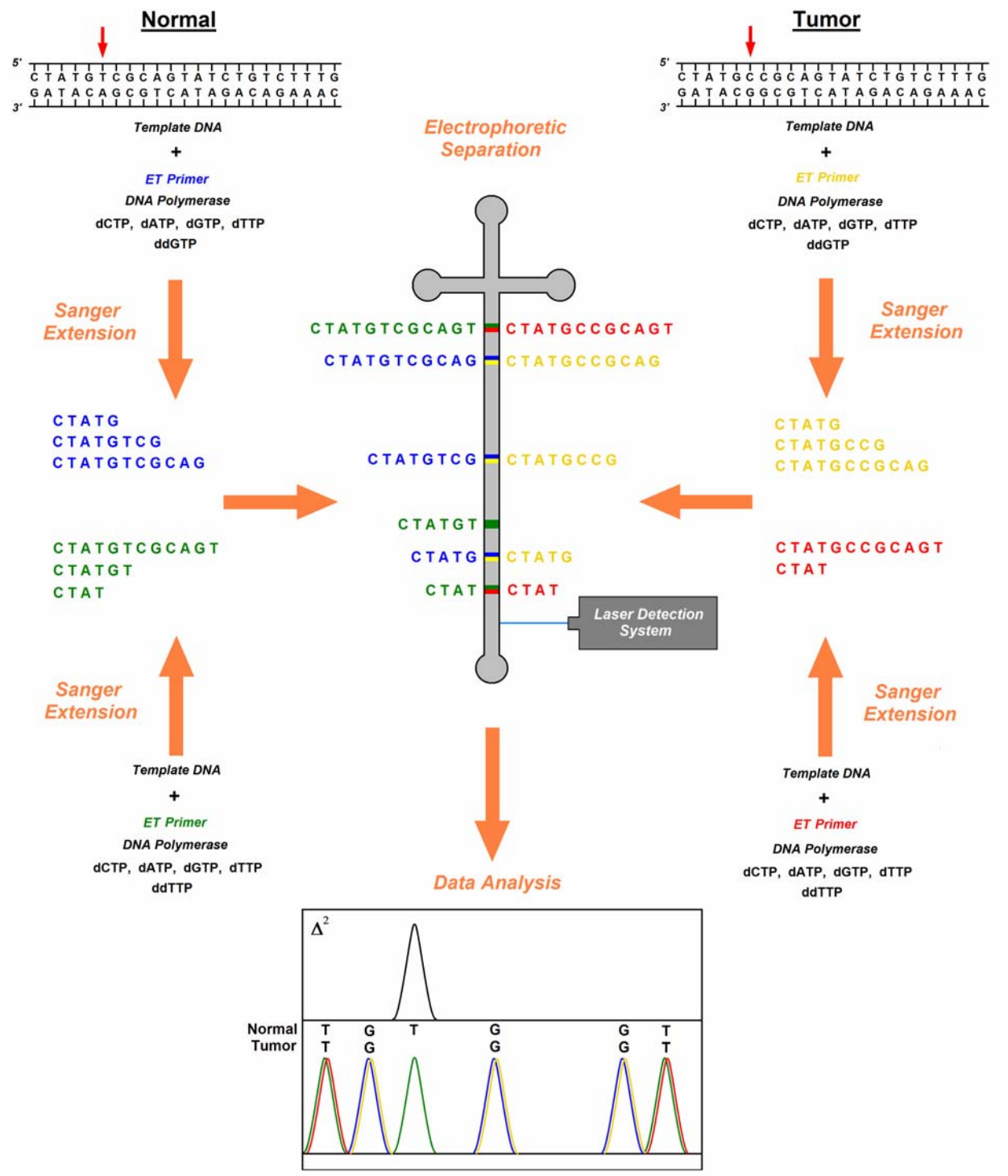



**Figure 1.8:** Polymorphism Ratio Sequencing. Corresponding Sanger fragments from two different templates are labeled using different dye primers. Pooled fragments are separated electrophoretically and analyzed for mismatches in the electropherogram. In this example, tumor DNA bearing a T→C mutation is compared to normal DNA from the same individual in a four-color G/T trace. The base change is indicated by a lone green peak with no corresponding red peak, as well as a single peak in the squared difference plot.

The inclusion of a control genome in each PRS sample serves two crucial purposes. Most fundamentally, the control sequence allows the user to tune the assay for direct detection of variation between specific samples. Somatic mutations are clearly differentiated from single nucleotide polymorphisms (SNP's) in the germline, as normal DNA from the individual is used as a reference. In the remaining regions of perfect overlap, the control genome also provides an internal standard that can be used for mutant genotype quantitation based on electropherogram peak ratios. In initial studies, this quantitation was found to be accurate to a detection limit of 5% minor allele population. In the context of mtDNA analysis, this quantitative accuracy allows for precise tracking of heteroplasmy shifts, potentially indicating the relative extent of tumor invasion in different samples.

As a diagnostic tool, PRS affords several advantages over alternate techniques for cancer genotyping. The collection of complete sequencing data in target genes allows for detection of novel markers on a patient-by-patient basis without prior knowledge of mutation sites, as is required in PCR, hybridization, and ligation-based SNP assays. In comparison to conventional sequencing, PRS circumvents disadvantages such as inability to quantitate and laborious sequence comparison by directly comparing the tumor genome to an internal control derived from the same individual. These advantages are particularly relevant in the context of recent phylogenetic studies suggesting high error rates in mtDNA sequencing comparisons, postulated to be the result of contamination or improper sample handling [106]. These reports have been corroborated by subsequent full-mtDNA sequencing studies performed with more rigorous quality controls in which authors conclude that previous reports of mutation frequencies may indeed have been overestimated [107].

In comparison to resequencing microarrays, PRS offers the advantage of accurate mutation detection with minimal dependence on local sequence variation and mutation type. Microarrays rely on uniform hybridization of labeled, fragmented sample DNA to short oligonucleotide probes fabricated onto the device and therefore require strict agreement with anticipated sequence. Multiple disturbances in local sequence give rise to inefficient hybridization and erroneous basecalls. The inability to discern single base insertions and deletions, as well as large-scale translocations, is also problematic [108]. Additionally, like many sequencing assays performed using other platforms, these microarrays have the fundamental flaw of using a single reference sequence for comparison, often the revised Cambridge sequence. The use of a standard reference is necessary for the identification of at-risk haplotypes, and given the present state of knowledge on functional mtDNA sequence variation, the selection of such a reference sequence is largely arbitrary. Nonetheless, no singular "healthy" sequence has been identified, and so any direct comparisons between individuals must be considered within a larger context.

## 1.6 Microfabrication as a Platform for Low-Cost, High-Throughput Genetic Testing

While genetic assays such as PCR, gene expression, and DNA sequencing undoubtedly hold the power to elucidate the molecular origin of disease and truly revolutionize the healthcare industry, the high cost and lengthy analysis times associated with these tests are major barriers to routine implementation. Using conventional extension kits in microliter-scale reactions, a single mtDNA sequence comparison consumes over $1,000 in reagents and requires 24 hours of laboratory time to complete. While parallelization and automation have been successful in addressing some of these concerns, miniaturization and integration wield the greatest potential in making personalized genetic analysis accessible to the general public.

Exploiting technologies developed for the computer chip industry, a number of research groups have progressed toward the goal of miniaturized bioanalysis systems through microfabrication. A typical procedure for the microfabrication of ultrafine features in a glass substrate is detailed in Figure 1.9. First, a glass wafer is cleaned and coated with a thin layer of photoresist. A mask bearing the features to be etched is then placed atop the wafer, and the exposed regions of photoresist are irradiated with UV light to transfer the pattern onto the wafer. After photoresist developing, glass in the exposed region is treated with concentrated hydrofluoric acid, isotropically etching microchannel features into the glass surface with a depth determined by HF exposure time. Channels are then sealed by bonding a blank glass wafer to the top of the etched surface [109-110].

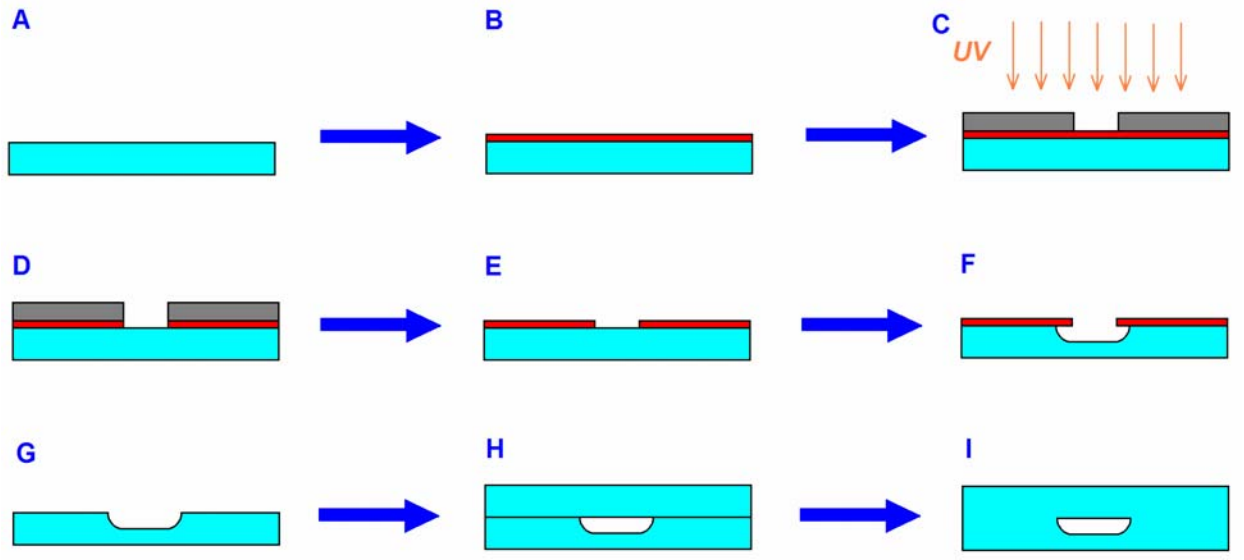



**Figure 1.9:** Microfabrication technique. A clean glass wafer (*A*) is spin-coated with photoresist (*B*). A photomask (*C*) is applied to the wafer, and UV radiation is used to expose the photoresist (*D*). After removing the photomask (*E*), glass in the exposed region is etched (*F*), and the photoresist is removed (*G*). A second blank wafer is applied to the top of the etched glass (*H*), and the chip features are sealed by thermal bonding (*I*).

In contrast to mainstream commercial devices for capillary electrophoresis, separations in a microfabricated device take place in channels tens to hundreds of micrometers in diameter. The resulting increase in surface-area-to-volume ratio allows for more efficient dissipation of Joule heating, which permits the application of higher separation potentials. As a result, electrophoresis in microfabricated capillaries is faster and better resolved than macroscale electrophoresis. In addition, the small sample volumes required for analysis allow for decreased reagent consumption and lower overall cost per assay.

Advancements in microfabrication technology have fostered impressive growth in the complexity and capability of bioanalytical microdevices in the past two decades. The first validation of a microfabricated capillary electrophoresis bioanalysis device was reported in 1993 and consisted of the separation of fluorescently labeled amino acids in 30-μm wide capillaries [111]. Separation of labeled DNA fragments was reported by Woolley *et al.* in 1994 [112], and subsequent design alterations yielded an improvement in resolution sufficient for DNA sequencing in 1995 [113]. With the success of these single-channel devices came a push for parallel analysis in the form of microfabricated capillary arrays. The first genotyping array chip consisted of twelve 6-cm long channels and was designed for fragment analysis [109]. More massively parallel 96-sample genotyping array chips were introduced in following years [110, 114], with a high-resolution sequencing counterpart following in 2002 [115]. Microfabrication reached its current peak of parallelism later that same year with the advent of a 384-lane genotyping chip [116].

With the development of oligonucleotide hybridization-based sample purification [117], as well as miniaturized devices for thermal control and fluidic manipulation, most recent emphasis has been on the integration of sample preparation with electrophoretic separation. The first tandem on-chip PCR-CE device was developed in 1996 [118], decreasing thermal cycling time from several hours to 15 minutes and total analysis time to 20 minutes. Integrated microvalves and pumps capable of transporting nanoliter volumes of solution microfluidically were introduced in 2003 [119] and have facilitated the development of portable instruments for forensic typing, pathogen detection, and space exploration [120-122].

The prospect of a microfabricated device capable of performing genetic analysis for cancer diagnosis represents a monumental step toward affordable, personalized treatment of a fundamentally individual disease. Such a device would not only have the potential to elucidate the key genetic variants responsible for tumorigenesis in the individual patient, but also provide the means for exploiting these variants as molecular markers of tumor invasion. The small sample sizes and rapid analysis times afforded by microfabrication speak to an era in which focused genetic analysis can be used as a real-time indicator of tumor excision, particularly in highly sensitive cancers such as those originating in the head and neck. Beyond the clinical setting, the demonstrated ease with which microfabricated bioanalysis devices scale up to arrays allows for multiplexed searches for genetic variants in several key genes. The elucidation of these variants will undoubtedly uncover the core pathways leading to tumorigenesis in the many varied forms of cancer and assist in bringing to fruition the ultimate goal of a complete understanding of cancer's biomolecular fingerprint.

## 1.7 Scope of the Dissertation

The focus of the work described herein is method development for the application of microfabricated capillary electrophoresis devices to the detection of human cancer. Primarily, this work has centered on the detection of tumor-associated somatic mutations in the mtDNA using PRS technology. First, a robust protocol for the successful, high-resolution separation of Sanger extension fragments in a high-throughput microfabricated capillary array electrophoresis (μCAE) bioprocessor was developed and validated through a complete mtDNA comparative sequence of two unrelated individuals from the CEPH institute (Centre d'Etude du Polymorphism Humain). The PRS methodologies were then directly tested against a comparable device for high-throughput mtDNA mutation detection, the Human MitoChip, correctly identifying 18 false mutations called by the MitoChip and detecting three additional heteroplasmic variations not previously identified.

Having validated the PRS protocols on non-clinical samples, a full-scale statistical investigation of bladder tumor-associated somatic mutations in the hypervariable mitochondrial D-Loop was undertaken. In parallel, a similar study of mtDNA variation in available urine extracts was also performed to explore the validity of correlating tumor-bound mutations with genetic markers found in associated body fluids. In total, 21 somatic variants were identified in 7 of 14 individuals. Of these, at least 15 variants exhibited some degree of heteroplasmy, with five representing population shifts from a heteroplasmic germline. In all cases, tumor-associated mutations were detected in matched urine, often more strongly than in the tumor itself.

Finally, the future of PRS as a clinical tool for real-time cancer detection is explored. A coding scheme allowing for semi-multiplexed primer-labeling of Sanger fragments is proposed, reducing the total number of extension reactions required for complete four-color PRS analysis by half. This scheme is employed in a proposed next-generation microdevice for integrated PRS sample preparation, purification, inline injection, and capillary electrophoresis. Taking full advantage of recent advancements in microfabrication technology in the fields of single-cell manipulation, on-chip thermal cycling, and high-resolution separation, the integrated PRS device reduces total sample analysis time from several days to mere hours, enabling the rapid and facile detection genetic variation.

Appendix A chronicles a brief foray into the enemy territory of gene expression analysis. The TRAC assay, or Transcript Analysis with Aid of Affinity Capture, was developed in the lab of Hans Söderlund at VTT (Technical Research Centre of Finland) for the purposes of monitoring the expression of cancer-associated nuclear genes. The facile accommodation of the TRAC assay on the 96-lane μCAE bioprocessor highlights the robust nature of μCAE technology, while initial steps toward on-chip purification and inline injection speak to the advantages of integration.

# Chapter 2

## Toward an Optimized, Streamlined Protocol for High-Throughput PRS

-or-

## *Mito-Man Faces his Demons…*

## 2.1 Introduction

The first challenge in advancing a microchip-based protocol for rapid, high-throughput PRS analysis of clinical samples was to develop a method for performing long-read PRS separations on a capillary electrophoresis microdevice capable of multiple parallel separations. In the first several months of experiments, it became apparent that having designed a device with the ability to generate massive quantities of high-quality data, the developers of the 96-lane sequencing chip selected for this endeavor had not learned how to operate it in a manner exploiting its full potential. An inconsistency in sample injection across the device was found to be the fundamental cause of pervasive multiple lane failures, a problem that was solved through the alteration of injection parameters.

Streamlining the data acquisition and analysis process was another issue to be addressed. Beginning with extracted DNA samples, a typical PRS comparative mtDNA sequence required 18 hours of sample preparation time, followed by three hours of electrophoresis and three months of manual data analysis. These processes were ameliorated through the implementation of a novel dynamic coating scheme and development of an automated, integrated data analysis program.

Finally, these alterations to the PRS scheme were validated through a variety of tests on both control and clinical samples. A full-mtDNA SNP detection assay was performed on two unrelated individuals using CEPH DNA, essentially repeating the experiments of the original PRS work in 2003. In addition to identifying all 36 previously identified SNP's, eight additional variants were uncovered. Partial-mtDNA somatic mutation analysis was also performed in three cases of lung and three cases of bladder cancer, uncovering three heteroplasmic variants in 16,288 bp of total sequence coverage. These results were compared to corresponding data from two versions of the Human MitoChip, a commercial mtDNA sequencing microarray, and correctly identified 14 false mutations erroneously called by MitoChip (v1.0) and 4 called by MitoChip (v2.0). The demonstrably impressive sensitivity of the optimized PRS protocols, particularly in comparison to the commercial device, establishes the applicability of PRS to accurate mutation detection in cancer research.

## 2.2 Improving Electrokinetic Injection Reproducibility

### 2.2.1: Electrokinetic Injection

The microfluidic platform for the majority of the PRS work advanced in this chapter is the 96-lane microfabricated capillary array electrophoresis bioprocessor developed by Paegel *et al*. in 2002 (Figure 2.1) [115]. This device features 96 electrophoretic separation channels, each 200 μm wide, 30 μm deep, and 17 cm long, arranged radially on a 150-mm glass wafer and converging on a common anode, through which separation matrix is loaded during chip preparation. The array is divided into 48 doublets, each consisting of two separation channels, each with independent sample reservoirs but sharing common cathode and waste wells. During electrophoresis, electrical potentials are applied by means of a 96-electrode ring which allows for independent addressing of the 96 sample wells. A polydimethylsiloxane (PDMS) moat ring containing 5x TTE (250 mM tris, 250 mM TAPS, 5 mM EDTA, pH 8.3) buffer is affixed to the

top of the chip and supplies voltage to the cathode and waste reservoirs. Separation at 150 V/cm followed by detection of laser induced fluorescence on the Berkeley Confocal Rotary Scanner [114] yields 96 four-color sequencing electropherograms in parallel. In initial tests, the sequencing array chip was capable of producing 1.7 kbp/min of high-quality sequencing data, with a total separation time under 30 minutes.

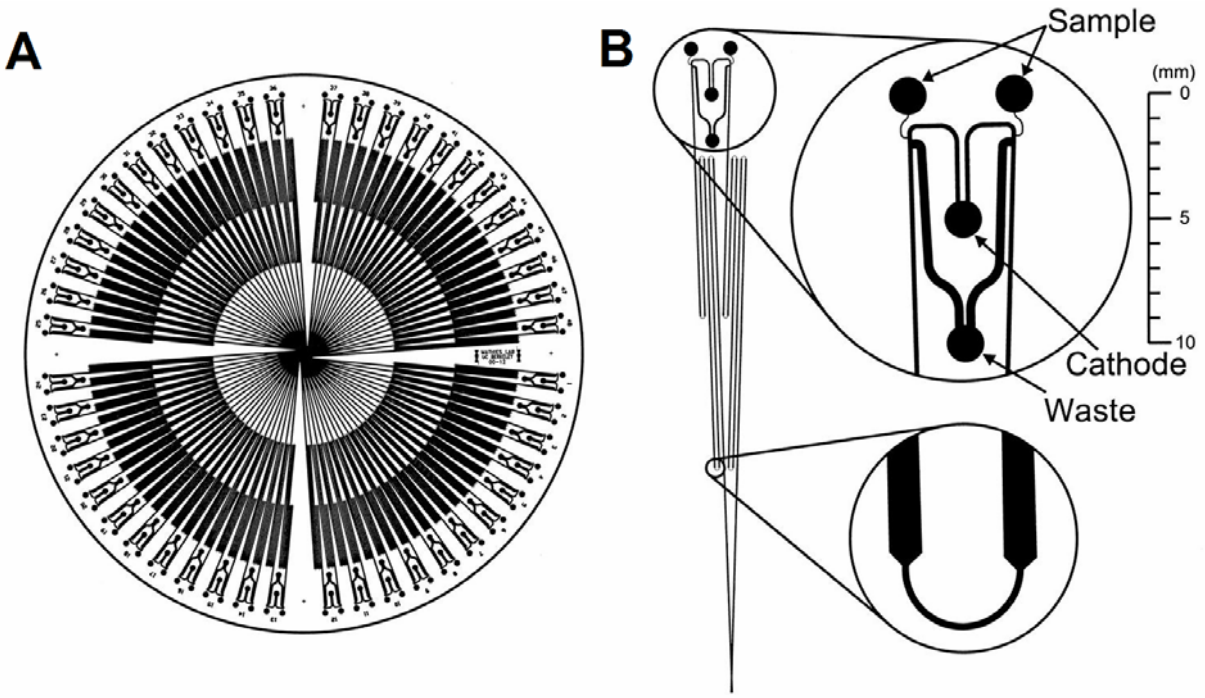



**Figure 2.1:** Layout of the μCAE bioprocessor. (*A*) The 96 17-cm separation channels are arrayed radially on a 150-mm wafer and are grouped into 48 doublets (*B*), with each doublet sharing common cathode and waste reservoirs. To accommodate the entire length of the capillaries on chip, each channel is folded back on itself four times, with each turn bearing tapered geometry to mitigate zone dispersion due to path length differences between inner and outer turn radii.

One of the most crucial steps in chip operation is the injection of sample into the separation capillary. In the array sequencing chip, as in most microfabricated capillary electrophoresis devices, the injection is performed electrokinetically by applying a positive electrical potential gradient from the sample reservoir across the capillary to a waste reservoir (Figure 2.2). Sample is driven out of the sample reservoir toward the waste through a channel crossing the separation capillary. Once a steady state concentration of sample is achieved in the channel intersection, the injection potentials are removed and the separation potentials are applied, with a slight positive "backbiasing" field remaining in the injector to prevent uninjected sample from streaming down the column during separation. The result of this approach is a well-defined initial sample band precisely tunable within tens of microns by channel geometry; however only a small fraction—less than one percent—of the sample is used for analysis. The inefficiencies and limitations of the electrokinetic injection are addressed in greater detail in Chapter 4; however this injection step presented a different obstacle in early experiments.

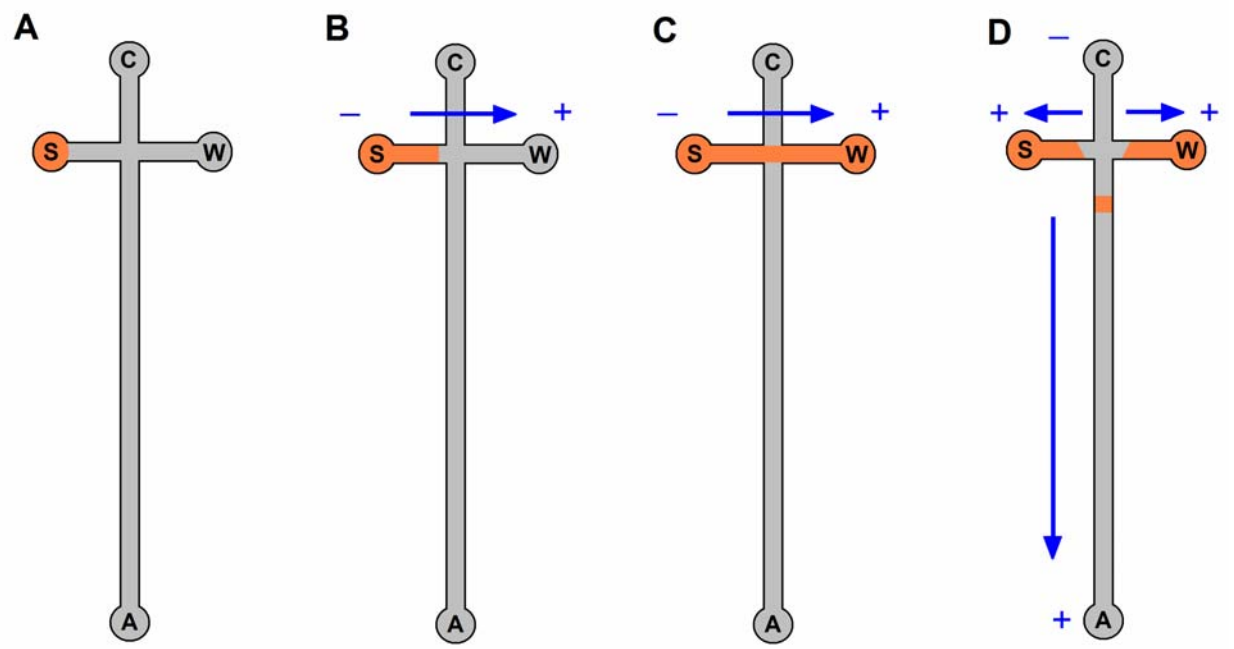



**Figure 2.2:** Electrokinetic injection in a simple cross. (*A*) Sample is pipetted into a sample reservoir. (*B*) A positive electric field is applied from this reservoir to a separate waste reservoir, driving the sample toward the waste and (*C*) filling the injection cross with sample. (*D*) Following injection, the separation potential is applied from cathode to anode, and a slightly positive field is maintained at the sample and waste to prevent sample leakage.

### 2.2.2: Irreproducible Data

In early exploratory experiments on the μCAE bioprocessor, results were plagued by unacceptably high lane failure rates, often approaching 100%. Aside from generally poor quality, electropherograms were highly inconsistent, with data from the same sample exhibiting considerable lane-to-lane and run-to-run variability. Lane failure was most commonly manifest as extremely low signal intensity, although in several cases complete detector saturation was observed. Often, both presentations were observed in adjacent capillaries, appearing in patterns similar to those depicted in Figure 2.4. Months of experiments altering sample preparation protocols and detection parameters were unsuccessful in producing acceptably consistent data. These anomalies were ultimately attributed to injection malfunctions by direct fluorescence imaging of the injection process.

### 2.2.3: Gel Overload and Clogging

During electrokinetic injection through a sieving matrix, extension fragments undergo separation according to their radii of gyration, the electric field applied, and the physical properties of the gel, just as they do in the separation capillary. Therefore, the temporal sample concentration "$C_{inj}(t)$" in the injection region is dependent not only on injection field "E" and sample solution concentration "C," but also on electrophoretic mobility "μ:"

$$C_{inj}(t) = \mu EACt$$

where "A" is the cross-sectional area of the capillary and "t" is time. In the absence of electroosmotic flow, the electrophoretic mobility of a particular ion in a medium with viscosity "η" is given by the relation:

$$\mu = q/6\pi\eta r$$

for ionic charge "q" radius of gyration "r." Thus, injection quantity is inversely proportional to ionic radius. For long injection pathways and low field strengths, this effect presents the potential for size biases in the injection band. An injection terminated before long fragments can migrate to the intersection results in an electropherogram with steadily decreasing signal intensity. Conversely, an excessively long injection runs the risk of depleting short fragment concentrations in the sample reservoir and ultimately in the injection cross, thus underrepresenting these fragments in the final electropherogram. The original injection protocols for the sequencing array chip were aimed at circumventing these effects by applying a field strength sufficient to induce biased reptation (500 V/cm). In the biased reptation regime, long DNA molecules depart from their native gyration in solution and align with electric field lines, breaking down the differential size bias that gives rise to electrophoretic separation. These injection parameters yield a consistent, but very high, injection concentration of all extension fragments almost instantaneously within the capillary.

For ideal (mid-dynamic-range) detection on the Berkeley Confocal Rotary Scanner, the concentration of each Sanger extension fragment must be on the order of 1 nM in the detection region and therefore, assuming negligible zone broadening, in the injection cross. However, the injected sample consists of a series of roughly 900 extension fragments. Assuming even termination across all base positions within the PCR template, the average length of these fragments is 450 nt. Given the approximate square-root dependence between single-stranded DNA fragment length and radius of gyration, the average effective length of these fragments is 400 nt, with a 13 nm radius of gyration [123]. Also present in the sample are nanomolar quantities of template DNA, which has been shown to occlude capillaries, although generally not for short PCR amplicons [124]. The resulting effective concentration of the PRS sample in the injection cross is 1 μM of a 400 nt extension fragment.

Exacerbating the high concentration of DNA fragments in the sample stream is the geometry of the injector system. Sample arms are approximately half as wide as the separation capillary, and a third as wide as the waste arm. In electrophoresis capillaries, electrical resistance is proportional to channel length and inversely proportional to each cross-sectional dimension. Consequently, resistance in the sample arm is over 15.5 times as large as it is in the

injection cross. Acting as a resistor series, this increased resistance manifests itself as a higher effective field strength in the sample arm, over three times the field strength in the injection cross and nearly four times the strength in the waste arm. Thus, a steady-state concentration of 1 μM in the injection cross corresponds to a 3 μM concentration in the sample arm. Early in the injection pathway, size-based separation is negligible. Therefore this full effective concentration enters the sample arms immediately after the injection is initiated. Such extreme DNA concentrations are capable of obstructing the separation matrix in the sample arm, an effect that can be observed visually by fluorescence imaging (Figure 2.3).

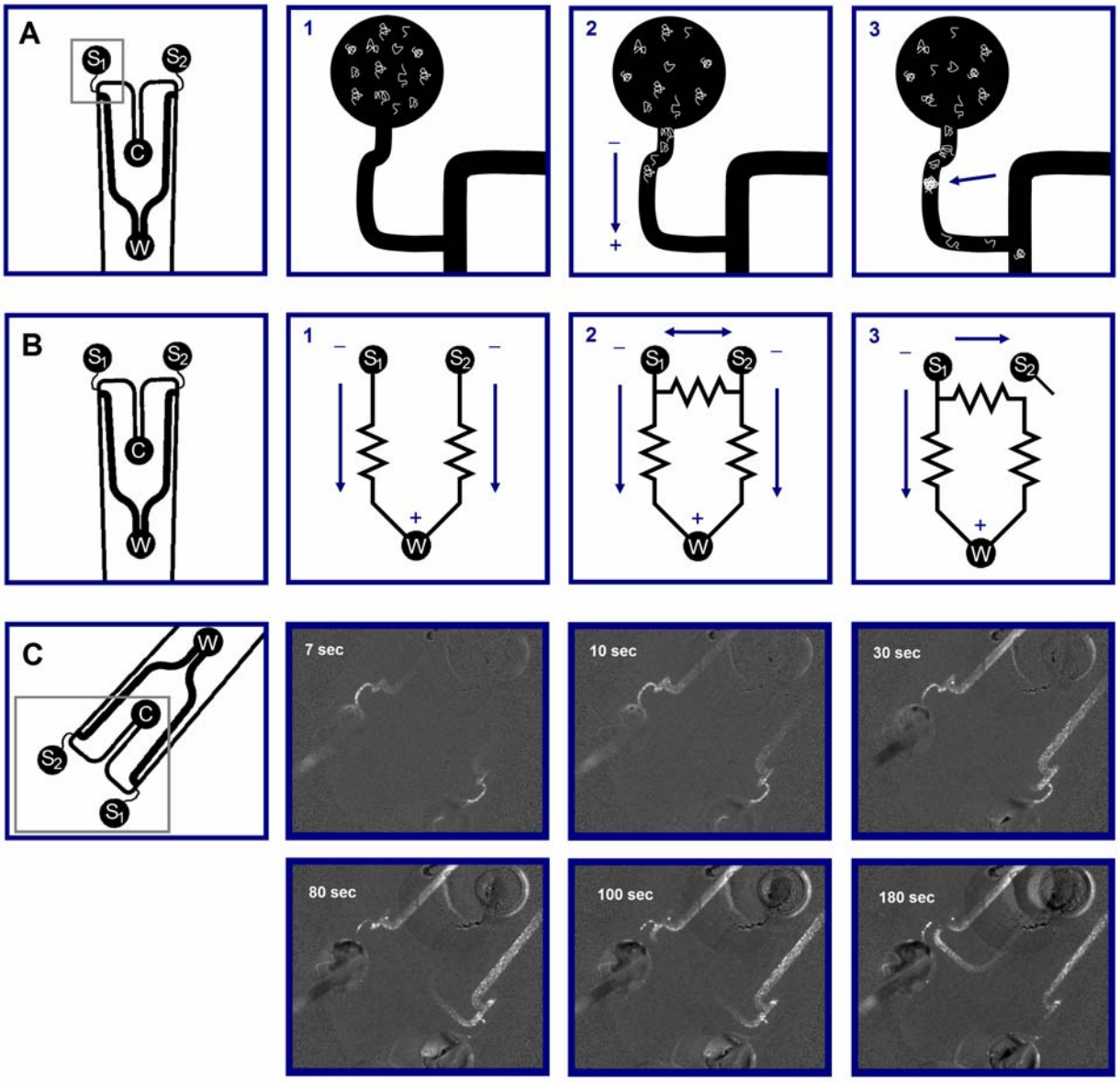



**Figure 2.3:** Gel clogging in the injection arm. (*A*) Sanger extension fragments from the sample reservoir (*1*) are highly concentrated by the high injection current in the sample arm (*2*), leading to gel obstruction and clot formation (*3*). (*B*) Although the injection structure was originally modeled as two independent circuits (*1*), the system behaves as a larger network due to the connectivity of the cathode (*2*). Disruption of the balance of resistance by matrix overload provides an alternate accessible sample pathway through the cathode arm (*3*). (*C*) Visible gel clogging and sample leakage at 500 V/cm. Clots form early in both sample arms but most severely in $S_2$, producing significant leakage from $S_1$ by the end of the injection (30 sec). At ~80 seconds, the arm adjacent to $S_1$ becomes more obstructed, reversing the leakage bias.

A final factor complicating the observed injection anomalies arises from the coupling of adjacent injection structures in a doublet design. It was originally assumed that during injection, each side of the doublet would act independently of the other. However, the common cathode reservoir, which is floated during injection, provides electrical connectivity between the twin injection crosses. As a result, each injector behaves in practice as a network of three resistors, with each side of the injection affecting the other. Such coupling is acceptable in the ideal circumstance of a perfectly balanced doublet, with each arm containing nearly equal-ionic-strength samples and no disruptions in current. The truly robust platform for high-throughput analysis, however, must be able to accommodate some degree of sample-to-sample variation, and as operated, the μCAE bioprocessor did not. The coupling at the cathode provides two possible sample injection pathways for each capillary (Figure 2.3(*B*)). In addition to the intended path from sample to waste, a drop in current in the adjacent capillary is capable of drawing sample toward the deficient side of the injector. This effect, along with the underlying obstruction of the injection pathways, can be observed directly by fluorescence imaging (Figure 2.3(*C*)). Depending on the magnitude of the drop, this effect can be severe, producing a common "saturation/flatline" pattern in which the unobstructed capillary yields a poorly resolved or completely saturated electropherogram while the clogged capillary gives little to no signal (Figure 2.4).

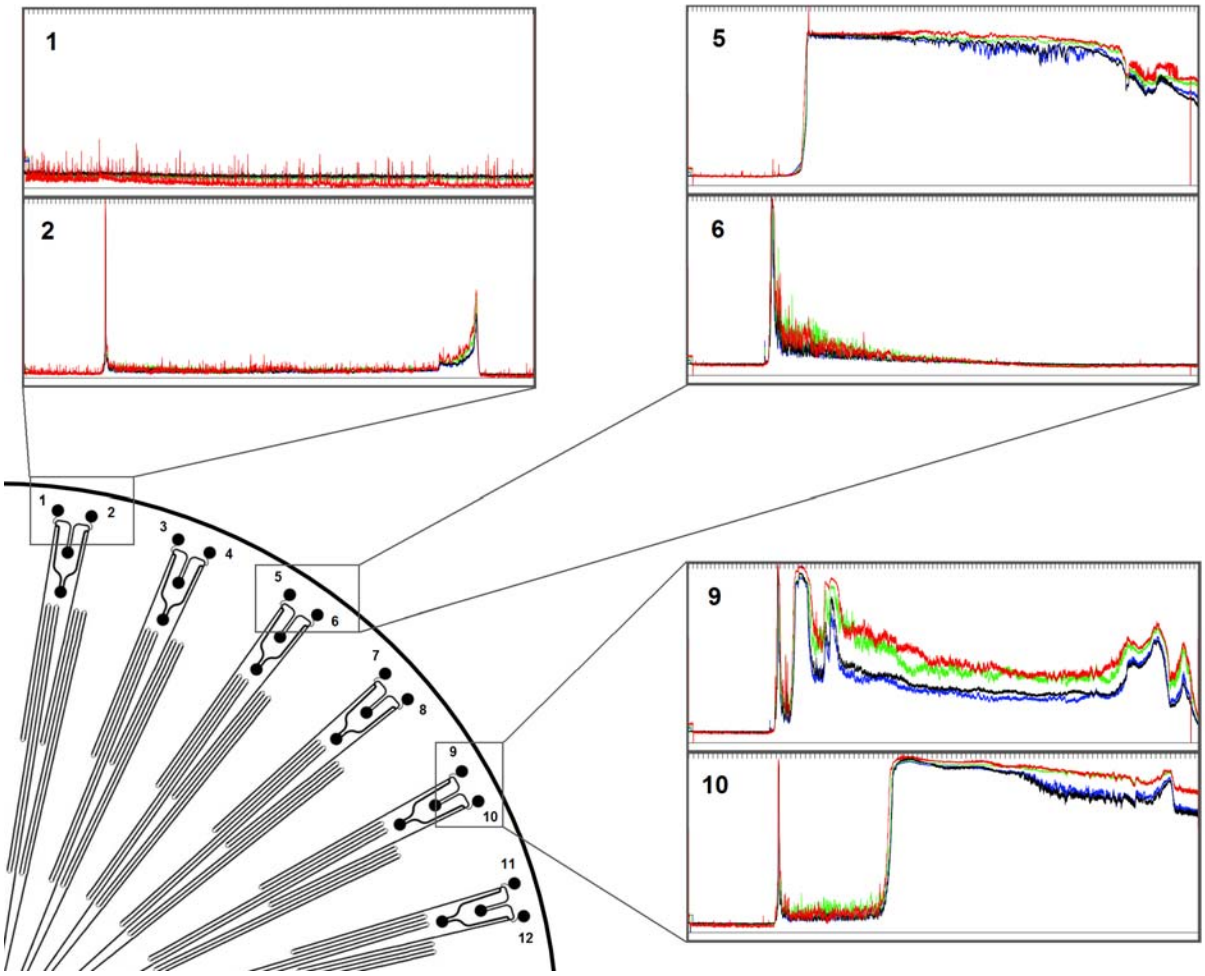



**Figure 2.4:** Manifestations of gel clogging in the context of the doublet design. Early formation of clots in both lanes of a doublet is typified by complete lane failure (*1*) or overall low signal (*2*) in both lanes. Clogging in one lane only may result in underinjection in the clogged channel (*6*) and total saturation in the other (*5*) due to sample leakage toward the cathode. Late-injection clogging may appear as loss of separation due to clot formation in the capillary (*7*) or delayed saturation due to loss of backbiasing (*8*).

Issues of inconsistent capillary performance and high lane failure rates had been observed in previous proof-of-concept studies with the sequencing array chip; however these difficulties were not fully understood and were often ignored due to the relative simplicity of the assays. The ability to circumvent these issues was at least in part due to the prior use of sequencing matrices synthesized in-house without rigorous control of polymer length and effective pore size. The resulting variability in gel properties likely produced an inconsistent obstruction threshold: therefore, a more permeable matrix, even if synthesized haphazardly, would yield more consistent injections across the array device and consequently, more successful separations. In the more carefully controlled context of a commercial sieving matrix, however, it was necessary to resolve the injection stability issues more robustly.

### 2.2.4: Eliminating Gel Overload

As stated above, the injector doublets—indeed all channels in the device—behave in principle like an electrical circuit, with a network of resistors separating termini at which electrical potential is held constant. Application of a potential gradient between termini results in a current within the channels. In a highly purified sample, this current is carried almost exclusively by the sample. Thus, current stability is a quantitative means of tracking the movement of sample in the capillary network and is highly effective in diagnosing injection and separation problems. For this reason, current was used in conjunction with fluorescence imaging as an indication of injection stability during optimization (Figure 2.5). Standard sequencing reactions prepared from mtDNA PCR products served as samples for these optimization experiments

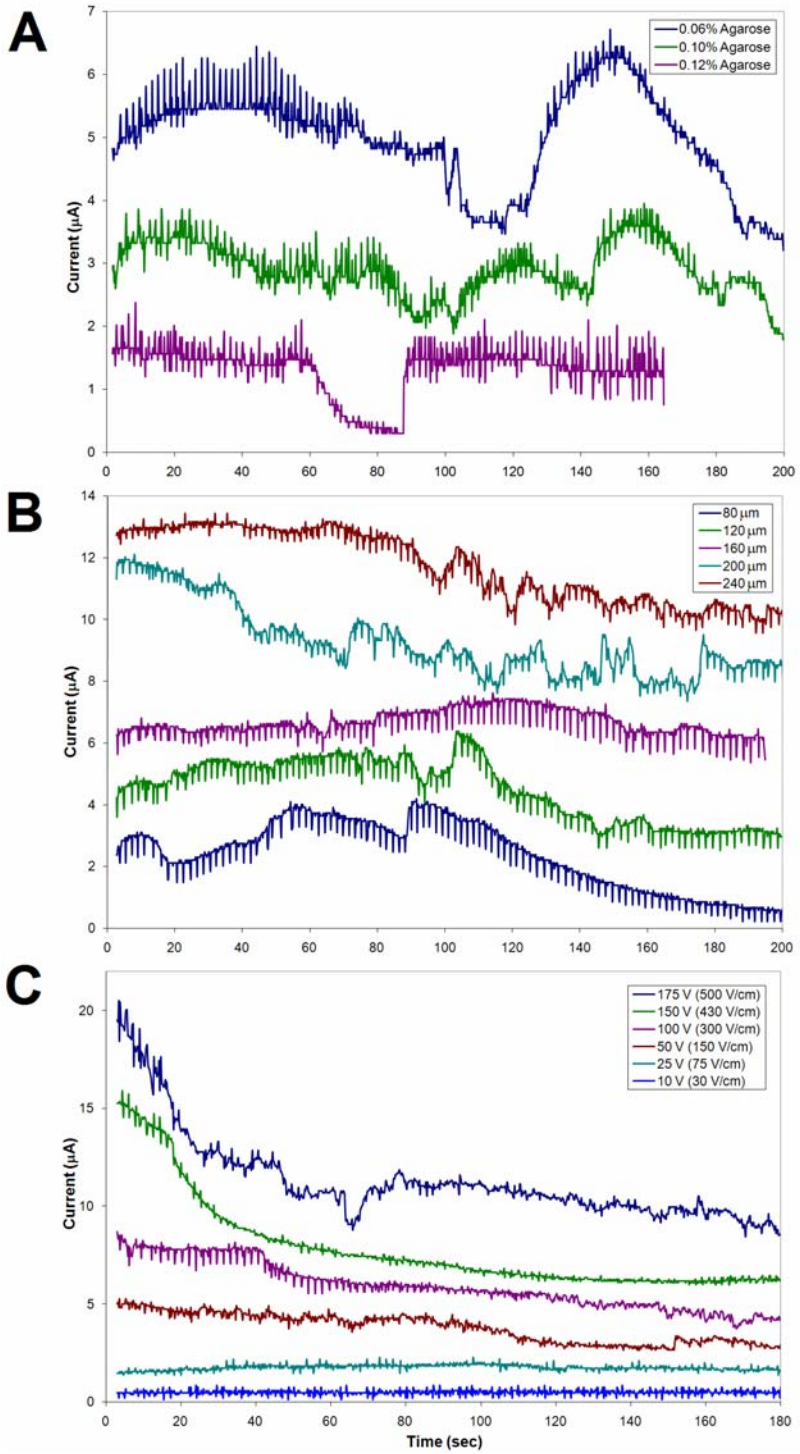



**Figure 2.5:** Optimization of injection stability. (*A*) Sample resuspension in low-concentration agarose suppresses low-mobility fragments and produces a transiently stable current at 150V/cm, with a current crash delay inversely correlated with agarose concentration. (*B*) Injection fields as low as 150V/cm produce unstable injection currents in experimental sample arms with widths ranging from 80-240 µm. (*C*) Decreasing injection potentials further to 25 and 10V (75 and 30 V/cm) yields stable, sustained currents, although visible clogging occurs even at 75V/cm.

A partial improvement in injection stability was achieved initially through the reduction of injection potentials. A reduction in injection field results in a decreased current and therefore, a lower concentration of extension fragments in the sample arm. This parameter was explored only cursorily at first, as low-field injections give rise to greater size bias and therefore require longer injection times to balance the intensities of short and long fragments. Longer injection times also introduce the risk of zone broadening due to diffusion as well as small ion depletion due to preferential injection. For these reasons an absolute injection potential of 50V (150 V/cm in the sample arm) was selected, although somewhat arbitrarily, as the minimum injection field studied. Unable to produce a reproducibly stable current alone, varying this parameter was a common thread in all subsequent optimization experiments.

Previously researchers have addressed the issue of gel clogging by resuspending their samples in low-concentration polymer solutions prior to injection. These "template suppression buffers" are designed to impede the injection of low-mobility strands, usually template from the extension reaction, and have even enjoyed success commercially. A variety of template suppression buffers have been employed, but ~0.1% agarose or polyacrylamide is a common standard [124-125]. The use of low-concentration agarose as a means of preventing sample overload was briefly explored. Solutions of 0.12%, 0.10%, and 0.06% agarose were prepared in filtered, deionized water and used as the sole resuspension buffer following sample purification. Current was monitored during injection at fields ranging from 500 to 150 V/cm. The 150V/cm injection is depicted in Fig 2.5(*A*). While injection from a template suppression buffer does yield a stable current transiently, the injection invariably crashes with visible sample clotting after several seconds. Interestingly, the timing of this breakdown is inversely correlated with agarose concentration, suggesting that increasing polymer concentration allows entangled clots to achieve mobility more readily and enter the sample channel.

As previously discussed, the narrow dimensions of the sample arm were a major factor in superconcentrating extension fragments during injection. In the original design of the μCAE device, this feature was intended to facilitate even loading of separation matrix throughout the injection structure by placing high fluidic resistance at the shortest port-to-port pathway. To explore the possibility of alleviating matrix overload by injector design, a test device was fabricated featuring identical doublets with varying sample arm widths (80, 120, 160, 200, and 240 μm). Separation matrix loading was indeed possible in all doublets, although complete loading in wide-arm doublets resulted in the loss of a large excess of gel through the sample arms. Each doublet was tested under varying injection potentials, with low-field (150V/cm) results summarized in Figure 2.5(*B*). Fluctuations in injection current were present in even the widest doublet at this field. It was thus concluded that increasing channel cross-section to any operationally feasible dimension would not alone support a stable current.

Ultimately, given the inability to resolve current instability and gel overload by any other means, the reduction of injection potentials was revisited as a singular approach to improving injection reproducibility. In addition to the higher-field injections previously attempted, injections at 25 and 10V (75 and 30 V/cm) were explored. Representative current profiles are compiled in Figure 2.5(*C*). A marked improvement in current stability is observed in successively lower injection fields, with the 75 and 30V/cm injections producing almost perfectly stable injections. Closer visual inspection of these lower-field injections by fluorescence imaging revealed considerable clot formation at 75V/cm, but not at 30V/cm, during

sustained injection. Therefore, the 30V/cm injection was explored further to determine appropriate timing parameters and performance under sequencing conditions.

Injection timing was determined by fluorescence imaging. Raw fluorescence intensity was monitored at several locations within the injection zone and normalized to a single scale to determine the point of maximum signal, presumed to correspond to a steady state of the largest proportion of extension fragments. As illustratd in Figure 2.6(*A*), this maximum occurs within 270-290 seconds of commencing the injection. To address concerns of zone broadening due to diffusion, injection band width was also monitored throughout the injection. Figure 2.6(*B*) confirms that the point of maximal fluorescence occurs in a regime of only minimal zone broadening, with over 90% of the final band width of 640 μm achieved within the first 60 seconds of the injection, when the raw fluorescence intensity rises most rapidly.

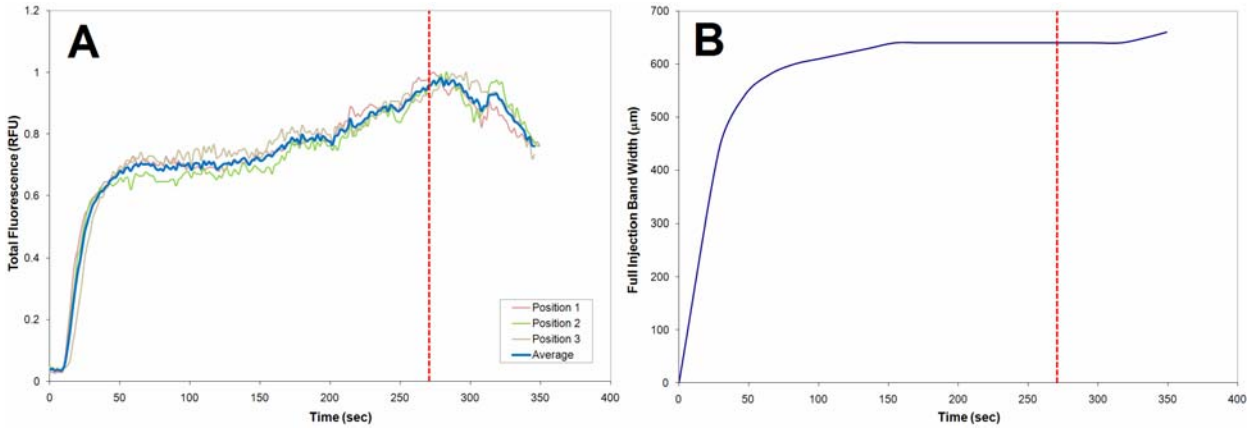



**Figure 2.6:** Electrokinetic injetion timing at 30 V/cm. (*A*) Total fluorescence as a function of injection time at several points within the injection band. The fluorescence maximum at approximately 270 seconds indicates a steady state concentration has been reached for a maximum percentage of extension fragments. (*B*) Minimal zone broadening is observed after approximately 60 seconds, confirming that maximum fluorescence occurs in a regime of only modest diffusion.

Despite the lengthy injection time requirements (ultimately defined at 270 seconds), initial tests on the μCAE chip using both standard sequencing and PRS samples were promising. Lane failures not directly attributable to chip fabrication, hardware failure, or channel obstruction prior to matrix introduction were virtually eliminated without significantly sacrificing the resolution necessary for sequencing-based analyses. Despite these improvements in reproducibility, however, one last procedural issue presented a fundamental obstacle to robust operation of the μCAE device.

## 2.3 Exploration of Dynamic Coating Protocols

At the pH of DNA electrophoresis buffers, the free silanol groups lining the walls of glass capillaries are deprotonated, resulting in a net negative charge along the walls of the channel. To preserve local electrostatic neutrality, an immobile layer of positive counterions from the solution forms along the capillary walls. This counterion layer attracts a second layer of relatively immobile anions, which is itself countered by a more diffuse, mobile layer of cations. During electrophoresis, this diffuse layer of solvated cations migrates toward the cathode, dragging the bulk solution in the capillary with it. This effect, known as electroosmotic flow (EOF), is advantageous in the absence of a sieving matrix and exploited in capillary zone electrophoresis to provide resistance to ionic electromigration. Gel electrophoresis, however, requires migration through a perfectly stationary matrix.

In sieving electrophoresis, EOF is circumvented through the use of coatings, chemical modifications which effectively neutralize capillary walls, thus abolishing the ionic multilayer and suppressing flow. The most historically prevalent approach to capillary coating was introduced by Stellan Hjertén in 1985. Hjertén's coating covalently attaches a layer of LPA directly to the capillary wall through a rigorous series of treatments that include cleaning with sodium hydroxide, alkylation of free silanol groups with a modified siloxane bearing a methacryl moiety capable of incorporating into LPA, and finally on-chip acrylamide polymerization [126]. Although this technique has proven successful and reliable in suppressing EOF, the coating procedure is laborious, requires 3-4 hours to complete, and involves high-pressure manipulation of hazardous chemicals. Additionally, the final on-chip polymerization step is highly sensitive to timing, solution homogeneity, and oxygen contamination: therefore, the quality of the Hjertén coating varies considerably between applications.

In recent years the cumbersome nature of the Hjertén coating has been addressed through the development of noncovalent dynamic coatings that are applied immediately before each separation. These coatings offer the advantages of minimal reagent processing, rapid application procedures with no *in situ* polymerization, and consistent quality. To aid in streamlining the PRS data collection process, polyDuramide, a dynamic coating consisting of an average ~2 MDa polymer of hydroxyethylacrylamide (HEA), was evaluated on the μCAE device for high-resolution applications. This work was conducted in cooperation with Annelise Barron's lab at Northwestern University. In contrast to the covalently attached Hjertén coating, polyDuramide adsorbs to capillary walls by hydrogen bonding. In initial studies performed in bare silica capillaries, the polyDuramide coating was able to suppress EOF for over 600 hours of continuous electrophoresis under sequencing conditions and, used in conjunction with a polyDuramide-based separation matrix, yielded read lengths exceeding 700 bp at 98.5% accuracy [127]. Procedurally, the polyDuramide coating is both simpler and less time-consuming than the Hjertén coating, requiring only a 15-minute wash with 1M hydrochloric acid, followed by a 15-minute treatment with a 0.25% (w/v) solution of the polymer. The entire process is sufficiently fast that a fresh coating can be applied immediately before each separation, thus increasing run-to-run consistency. The coating solution itself is polymerized off-chip in large quantities, thereby eliminating the further variability introduced by on-chip polymerization.

To evaluate the polyDuramide coating in the context of the μCAE device, a series of separations was performed on standard PRS samples after coating with either polyDuramide or

the Hjertén procedure. Resolution was monitored for three sets of neighboring peaks distributed throughout the trace and used as a direct indication of coating performance. In addition, two novel separation matrices from the Barron Lab were tested, both in conjunction with the polyDuramide coating. A 5% polydimethylacrylamide (PDMA) matrix reportedly increases resolution of larger, lower-mobility fragments, while a 4% copolymer of LPA and HEA was engineered to improve smaller-fragment resolution (both matrices prepared in 7 M urea). Figure 2.7 represents the compiled results of these experiments, with each result normalized to the corresponding resolution obtained using commercial GenomeLab Separation Gel LPA-1 (Beckman-Coulter) with a Hjertén coating. Each bar represents the average results for five samples, each run in duplicate during a single experiment on the μCAE chip. Within uncertainty, most of the explored matrix/coating permutations exhibited approximately similar performance; however a consistent and significant improvement in resolution (18% on average) was obtained throughout the trace using LPA-1 with a polyDuramide coating. The experimental sequencing matrices did not offer a significant improvement in resolution, although both did appear to resolve lower-mobility fragments slightly more effectively.

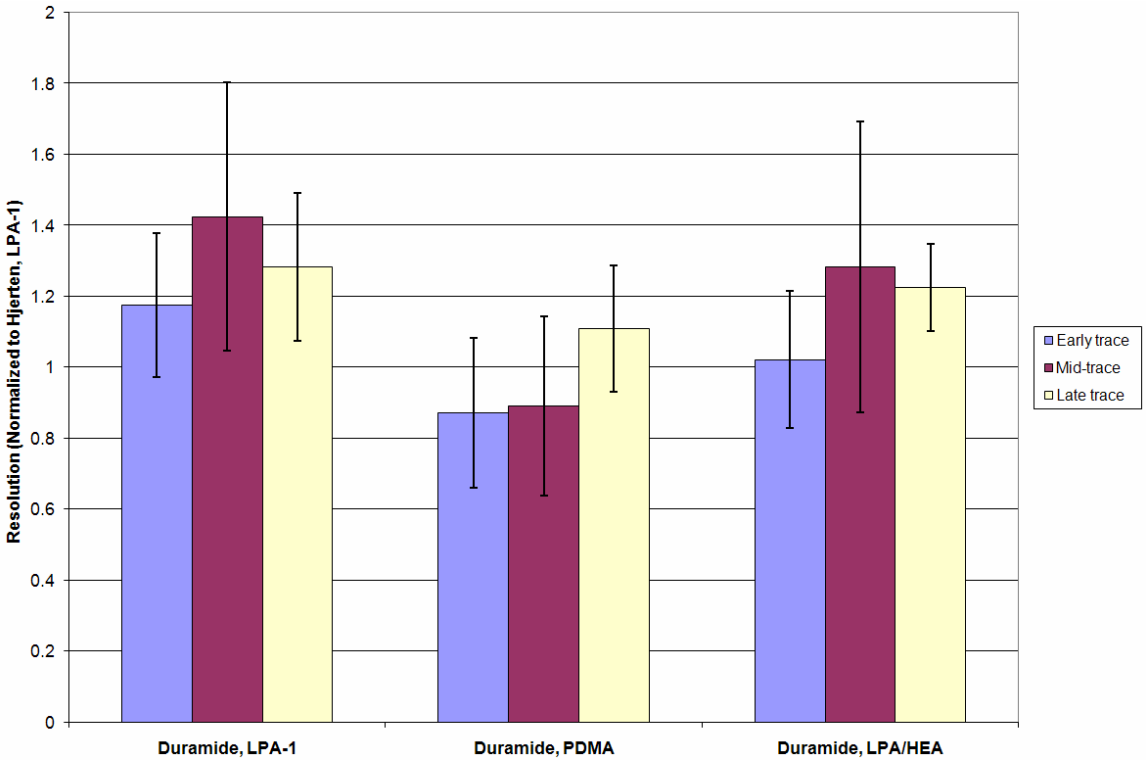



**Figure 2.7:** Resolution comparison of various coatings and separation matrices. Duplicate separations of five individual PRS samples are represented, and all data are normalized to the Hjertén coating and commercial LPA-1 matrix previously used. A combination of the polyDuramide dynamic coating and LPA-1 affords a consistent improvement in resolution throughout the sequencing trace, while other experimental separation matrices perform similarly to Hjertén/LPA-1 within uncertainty.

Subsequent to these results, the previously employed Hjertén coating has been replaced by polyDuramide. In addition to reducing total chip preparation time, this alteration has fostered a near 100% capillary success rate, increased run-to-run consistency, and higher-quality electropherograms. As exemplified in Figure 2.8, conventional sequencing data collected using the polyDuramide coating in conjunction with the new injection parameters have been impressive, yielding continuous read lengths surpassing 500 bases with 99% accuracy. Interestingly, many early-and mid-trace miscalls appear to be deletion errors, indicating poor calibration of conventional sequencing software with data obtained from the μCAE device.

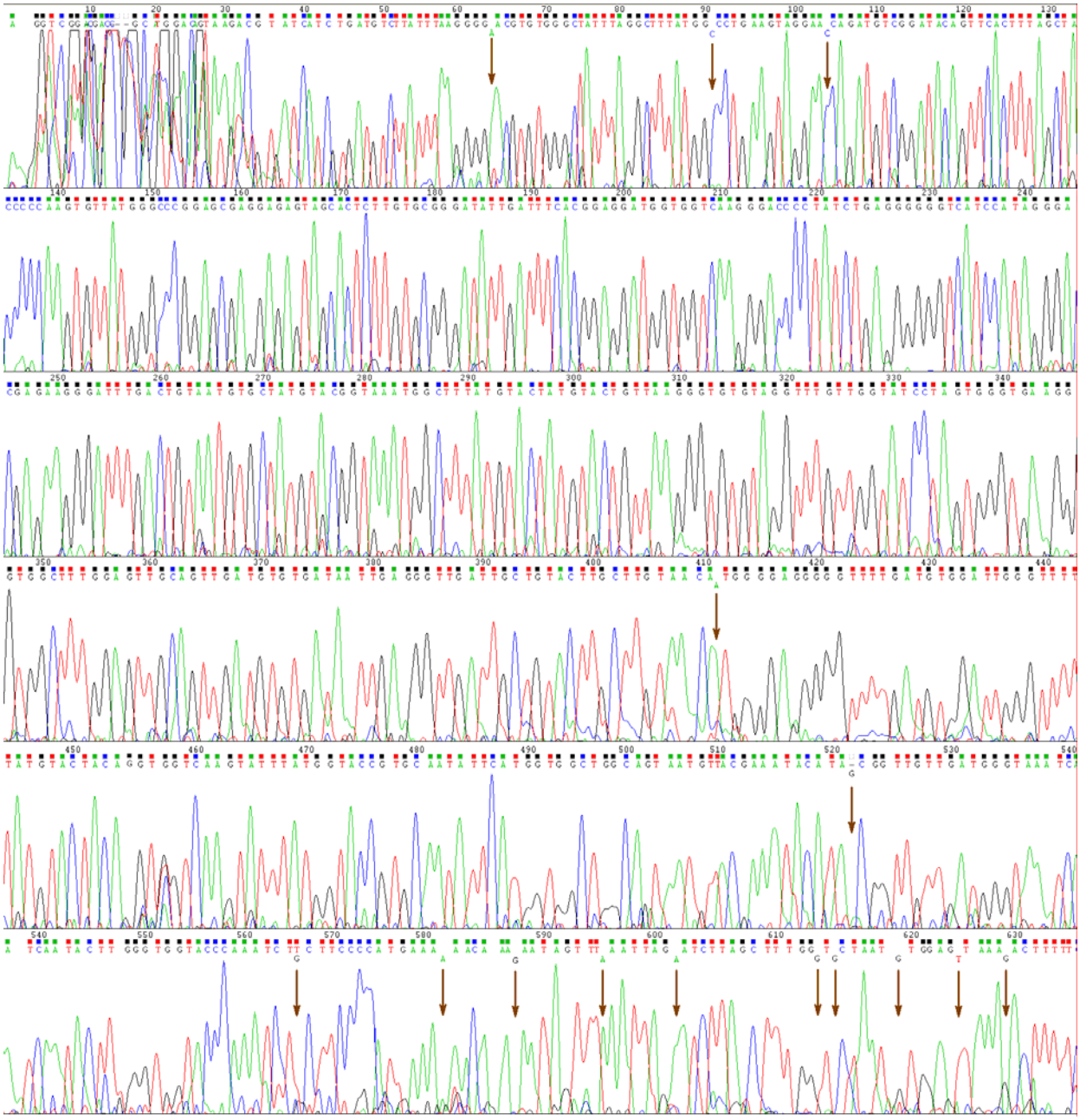


**Figure 2.8:** Standard sequencing using optimized coating and sample injection protocols. Improved operation conditions generate a high proportion of successful lanes with impressive accuracy, as typified by this result. Five errors from positions 65 to 581 yield a continuous 99% accurate read length of 516 bases. Most of these errors are deletion errors, suggesting a lack of synchrony between microchip results and conventional basecalling software.

## 2.4 Automating and Streamlining PRS Data Analysis

A final challenge in developing a high-throughput microchip-based PRS methodology was the design of specialized software for automated analysis of 96 comparative sequencing electroherograms. Despite its basis in conventional sequencing, PRS involves fundamentally different, although considerably less intensive, data processing, and is thus incompatible with standard basecalling software. Previously, these custom processing steps were performed manually on individual electropherograms using a combination of data analysis programs. Integration of these steps into a single, automated program was the final hurdle in implementing high-throughput PRS.

To determine the complete DNA sequence from a single trace, conventional basecalling software requires a well-resolved, four-color electropherogram with single-base resolution and regular peak intensities. Even under highly controlled and optimized labeling conditions, irregularities in base incorporation, formation of secondary structure, and variations in detector response render this degree of regularity difficult to achieve. Therefore, automated conventional sequencing software supplements raw data quality through a series of processing steps which assume minimal overlap between adjacent bases and a regular, often prescribed rate of fragment migration. Consequently, final sequencing electropherograms are highly manipulated and only qualitatively accurate. Such artificial processing is, however, acceptable for conventional sequencing data, as the objective is a series of discrete basecalls, each with only four possible identities.

In contrast, PRS emphasizes the quantitative detection of variation within a single electropherogram. Thus, the identification of specific basecalls is not necessary, while preservation of signal intensity is. The inclusion of a reference sequence provides an internal control by which both the intensity and temporal position of each base track can be calibrated, thus eliminating the need for short-range normalization. Additionally, and in further contrast to conventional sequencing data, PRS coding is designed to produce maximal overlap between base tracks in a single electropherogram. Therefore, the minimal-overlap algorithms employed in standard sequencing are completely incompatible with PRS data analysis.

Proper PRS data analysis occurs in seven main stages. Following data collection, the raw electropherograms are truncated at the beginning and end, eliminating data which, due to either polymerase error or the resolving power of the matrix, are of insufficient quality to analyze. This step also eliminates the high-intensity primer and PCR runoff peaks resulting from unincorporated dye primers and extension fragments spanning the entire PCR amplicon. The cropped electropherograms then undergo a matrix deconvolution in which spectral overlap between the four fluorescent probes is corrected. Standard noise filtration follows, along with baseline subtraction, and signal normalization to adjust the scale of all four tracks. The temporal alignment of overlapping tracks is then adjusted to account for slight differences in mobility between fluorescent probes. Finally, the squared difference plot is generated and variants are identified.

In previous PRS studies, the aforementioned data processing steps were performed predominantly using BaseFinder, a freeware genetic analysis program for single-lane processing [128]. Raw data files were examined visually, and regions of poor data were deleted manually. Matrix deconvolution was performed using a relation matrix containing the spectral properties of each dye. This matrix was defined by searching for and selecting minimally overlapping peaks

for each individual base from a standard sequencing trace. Similarly, mobility shift correction was performed manually by sliding individual base tracks until visual overlap was achieved. Finally, the BaseFinder-processed data file was transferred to Microsoft Excel, where local modifications to normalization and alignment were made and the squared difference plot was calculated. With few exceptions, each of these steps was performed manually for all 96 lanes. Scripting operations in BaseFinder allowed for some degree of automation; however, a great deal of personal attention was still required in the local normalization and alignment phase, and the entire data analysis process took nearly three months to complete.

The effort to automate and streamline this process in a single data analysis program was advanced in collaboration with Jing Yi from Professor Terry Speed's group in the Statistics Department at UC Berkeley. Working entirely in MATLAB, Jing developed scripts to perform each of the necessary operations in rapid, automated succession. Raw data files are first trimmed of primer and terminal peaks at fixed points in the electropherogram (7 and 27 minutes, respectively) and baseline-corrected using Tophat, a dynamic baseline subtraction algorithm. Spectral deconvolution is performed using a relation matrix defined independently in BaseFinder and entered manually by the user once at the beginning of analysis. Noise removal by Wavelet transform follows, along with global normalization relative to the mean signal in each track. The trace is then aligned for maximal overlap using a Dynamic Time Warping algorithm, and the squared difference plot is generated automatically, with possible variants (squared difference peaks more intense than three standard deviations above the mean) highlighted. Complete data analysis is performed automatically on all 96 data files in 6-7 hours, producing individual overlaid base tracks but also preserving partially processed data in separate folders for diagnostic purposes. Genetic variation is detected by manual inspection of the final traces, allowing for more focused inspection of variant regions to identify the nature of each mutation from sequence content.

Aside from automating and enhancing the rate of data processing, the integrated software differs fundamentally from previous BaseFinder-based analysis in two key respects—baseline subtraction and trace alignment. In the integrated analysis software, Tophat, a morphological method commonly used in graphic enhancement, is employed for baseline correction. Previously, a simple linear interpolation was performed between minima located in successive windows in the trace, and the value of the line connecting these minima was subtracted from each corresponding data point. Tophat improves this approach by performing baseline adjustment in a dynamic, nonlinear fashion. Instead of successive windows, a sliding window is employed. A local minimum is found within a window centered on a specific data point and saved into a vector at the midpoint. The window is then shifted by a single point and a new local minimum is found and recorded. The process continues for the entire trace, and the resulting minimum vector is subtracted from the raw data. The result is a more carefully baseline-corrected trace in which authentic data are clearly differentiated from background, as illustrated in Figure 2.9.

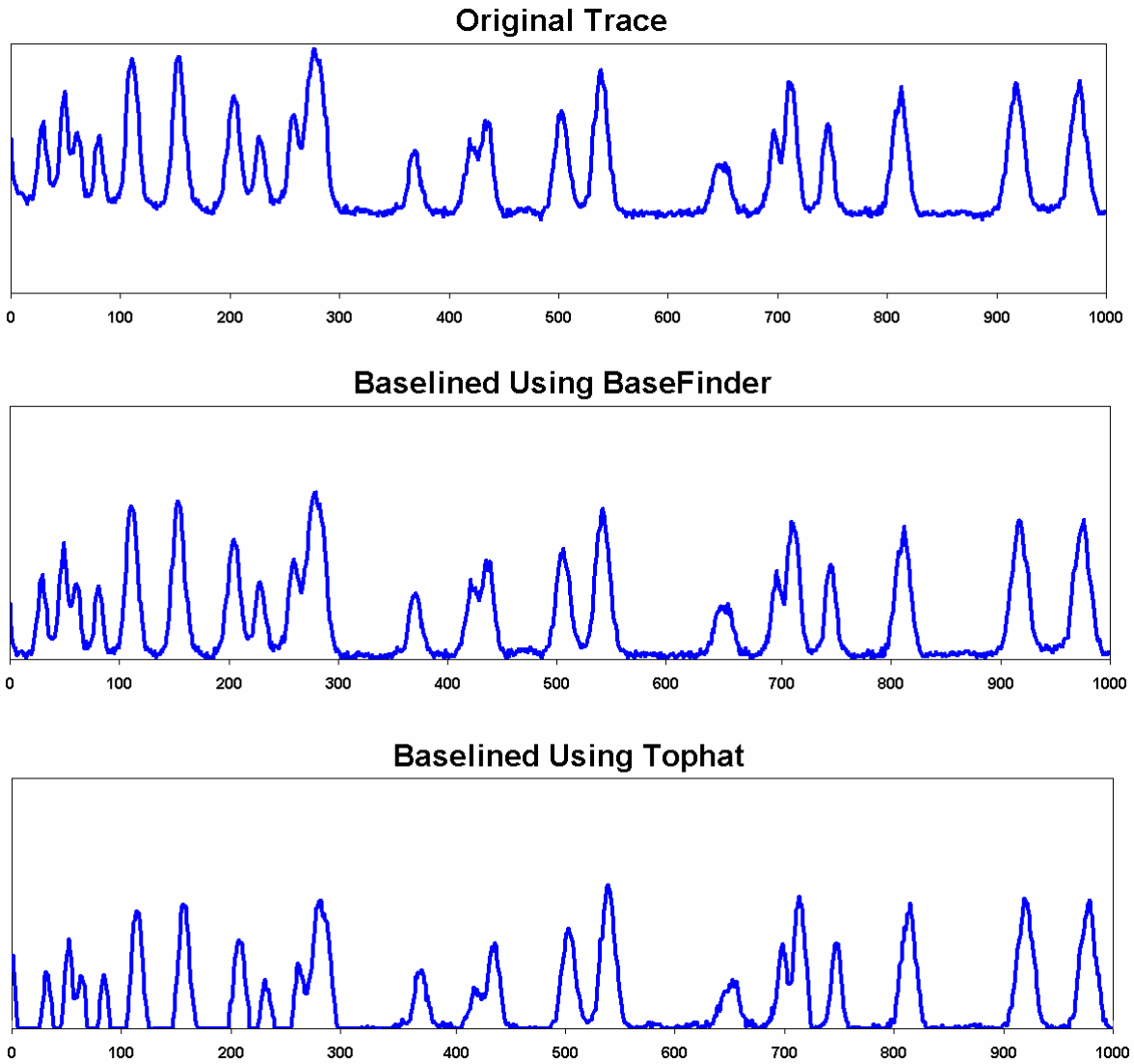



**Figure 2.9:** Tophat baseline correction in the integrated PRS analysis software. Tophat suppresses subtle, transient shifts in the baseline by defining a vector of local minima using a sliding window. The result is a more carefully baselined trace than that produced by the successive-window baselining algorithm in BaseFinder.

Trace alignment in the integrated program is performed using Dynamic Time Warping (DTW). Working in small windows with a size defined by the user, the DTW algorithm memorizes the optimal solutions to alignment sub-problems within the trace in an organized form, defining a similarity score to each potential solution at each point. The function then allows a range of trace-shifting "steps" in the space of (time frames in sample, time frames in reference) and determines the path through this space that maximizes the cumulative similarity score. This optimal path defines the mobility shift more dynamically than the global quadratic formula-fitting algorithm in BaseFinder, thereby correcting for subtle variations in migration times between bases and accommodating small-scale insertions and deletions.

Upon careful evaluation of the program, it was discovered that, although the DTW algorithm used to align sample and reference traces was successful in resolving isolated SNP's in regions of high similarity, it failed to recognize adjacent SNP's (genetic variants separated by five or fewer bases). This effect, typified by the double TACC→CACG polymorphisms at base position 16126 in Figure 2.10, was a consequence of a narrow DTW window. This parameter defines the amount of data considered when calculating local trace alignment solutions and must therefore be tuned to include a sufficient number of control peaks in regions containing variation. Increasing this parameter over a range of 31-601 scan positions yields an increased resolution of the two adjacent SNP's, with a maximum separation observed at a DTW parameter of 401 scan points (1/15 of the analyzed trace).

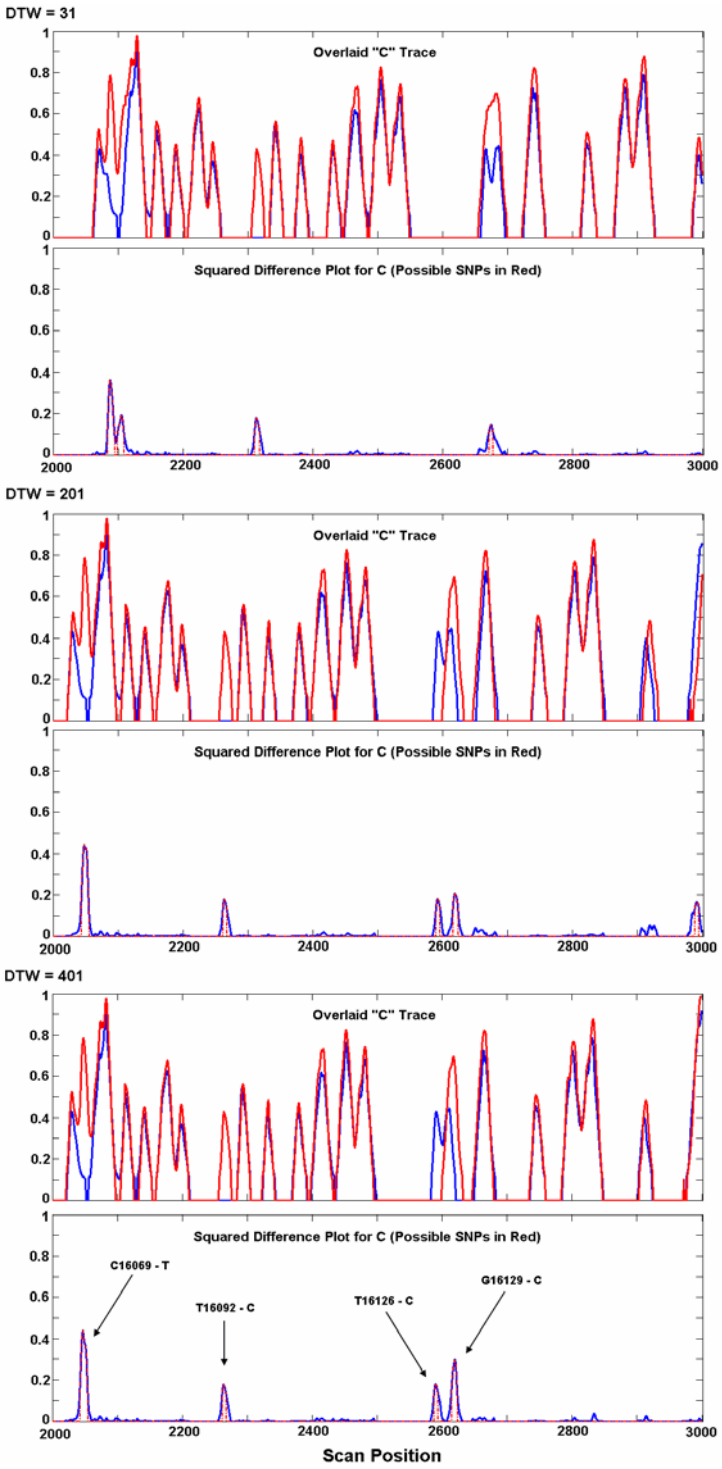



**Figure 2.10:** Optimization of DTW trace alignment in the integrated PRS analysis software. The trace context used in defining the quality score of each DTW solution is tunable by the user and must be sufficiently large to accommodate closely adjacent variants, typified by the TACC→CACG double-SNP at position 16126. The effects of varying DTW window sizes are shown, with 401 scan points (1/15 of the analyzed trace) yielding optimal detection of the two variants while preserving trace alignment in neighboring regions.

## 2.5 Validation of Optimized Conditions

### 2.5.1 SNP Detection Using Standard CEPH DNA

Following independent, focused optimization of coating, injection, and data analysis procedures, the combined protocol was tested in the context of a complete mtDNA sequence comparison between two unrelated individuals from the CEPH institute (CEPH ID's NA10859 and NA13116). This comparison provides a facile intermediate validation, as base-to-base agreement between different individuals is largely binary, with variants appearing as pure homoplasmic polymorphisms. Additionally, this analysis is an exact replication of the initial PRS proof-of-concept work reported in 2003, and thus presents a direct comparison between original and optimized techniques.

Generation of template for PRS extension was performed by PCR amplification of the entire mtDNA in 24 fragments with average length 893 bp. CEPH DNA (10 ng of total DNA extract) was combined with 7.5 pmol each of forward and reverse PCR primer (Integrated DNA Technologies), 2.5 μL of 10X PCR Buffer (-$MgCl_2$), 0.5 nmol of each dNTP, 37.5 nmol of $MgCl_2$, and 1.25U of Platinum TAQ DNA Polymerase (Invitrogen) and diluted to a total reaction volume of 25 μL with nuclease-free water (ISC BioExpress). PCR primers from Rieder *et al.* [129] were modified at the 5' end to include universal sequencing primer recognition (-40 M13 for forward primers and -28 M13 for reverse). PCR reactions were incubated in an MJ Research PTC-200 thermal cycler for 30 cycles of denaturation (94º C for 30 seconds), annealing (61º C for 45 seconds), and extension (72º C for 90 seconds). The thermal program was initiated with a 60-second denaturation at 94ºC and terminated with a final incubation at 72º C for 180 seconds. Primer pair 18 (base positions 11929-12793), required a lower annealing temperature for successful amplification and was thus annealed at 50º C. PCR products were purified enzymatically by incubation with 30 U of exonuclease I and 2 U of shrimp alkaline phosphatase (GE Healthcare) for 30 minutes at 37º C, followed by 15 minutes at 80º C. PCR products were quantified using the PicoGreen assay (Invitrogen) and diluted to a final concentration of 12 ng/μL.

Sanger extension was performed using energy transfer (ET) dye primer labeling chemistry. For each uniquely labeled set of extension fragments, 4 μL (48 ng) of purified PCR product were combined robotically with 800 fmol of the appropriate universal ET sequencing primer (FAM-R110, FAM-R6G, FAM-TAMRA, or FAM-ROX, emission maxima at 525, 555, 580, and 605 nm, respectively) [98] and 2 μL of DYEnamic Direct Cycle Sequencing Mix (GE Healthcare). The reaction was then diluted with water to a total volume of 10 μL. Thermal cycling consisted of 45 cycles of denaturation (95º C for 30 seconds), annealing (52º C for 15 seconds), and extension (72º C for 60 seconds). Extension fragments were pooled and purified by co-precipitation with 15 μg of glycogen (MP Biomedicals) in 3 μL of 7.5 M ammonium acetate (Fisher) and 110 μL of 100% ethanol (Rossville). After 60 minutes at -4º C, the reactions were centrifuged at 3700 rpm for 30 minutes and the pellets washed with 150 μL of 70% ethanol. The pellets were recollected by a second 5-minute centrifugation at 3700 and allowed to air dry before resuspension in 15 μL of 50% HiDi Formamide (Applied Biosystems).

Prior to analysis, microchannels were coated with polyDuramide. Channels were then primed with 1 mL of 1x TTE (50 mM tris, 50 mM TAPS, 1 mM EDTA, pH 8.3), and separation matrix (GenomeLab Separation Gel LPA-1, Beckman-Coulter) was loaded via the central anode

using a high-pressure gel loader [130]. Excess matrix was evacuated from the sample reservoirs, and 1.8 µL of sample were added to each well. The microdevice was then transferred to the Berkeley Confocal Rotary Scanner, where a ring bearing 96 electrodes was affixed to the sample reservoirs. Cathode, anode, and waste moats were filled with 5x TTE buffer, and electrical connections were made to each. PRS fragments were injected electrokinetically at 30 V/cm for 270 seconds, followed by separation at 150 V/cm for 30 minutes. Both the injection and separation were performed at 67º C. Data were analyzed using the integrated software.

The complete results of the full-mtDNA SNP analysis of CEPH samples NA10859 and NA13116 are compiled in Table 2.1. In total, 44 variants, all single-base substitutions or insertion/deletions, were found. All 36 of the variants uncovered in Blazej *et al*. were detected cleanly. In addition, eight novel polymorphisms not detected in this previous study were found, including one C→T transition not previously documented in mtDNA literature. All previously undetected variants were confirmed independently by conventional sequencing. The omission of these variants from the previous analysis is likely the result of low raw signal intensity and poor resolution arising from lane-to-lane variability under the previous high-field injection and Hjertén coating scheme. These variants also may have been overlooked during manual data analysis using segmented and sub-optimal protocols. The single C-insertion in the D310 poly-C tract is particularly problematic in this regard, as the sample and reference traces differ by only a single base in a long region of continuous overlap (12 cytosines). The ability to detect this and other challenging polymorphisms using the optimized PRS protocols confirms the increased accuracy and sensitivity achieved under these conditions and sets the scene for further validation using clinically relevant samples.

| Base Position | Gene | Base Change | AA Change |
|---|---|---|---|
| 185 | D-Loop | G → A | ---- |
| 188 | D-Loop | A → G | ---- |
| 217 | D-Loop | T → C | ---- |
| 228 | D-Loop | G → A | ---- |
| 303 | D-Loop | C → CC | ---- |
| 462 | D-Loop | C → T | ---- |
| 489 | D-Loop | T → C | ---- |
| 508 | D-Loop | A → G | ---- |
| 514 | D-Loop | CAC → C | ---- |
| 1811 | 16S rRNA | A → G | Syn |
| 3010 | 16S rRNA | G → A | Syn |
| 3720 | NADH Dehydrogenase 1 | A → G | Syn |
| 3849 | NADH Dehydrogenase 1 | G → A | Syn |
| 4553 | NADH Dehydrogenase 2 | T → C | Syn |
| 4736 | NADH Dehydrogenase 2 | T → C | Syn |
| 5390 | NADH Dehydrogenase 2 | A → G | Syn |
| 5426 | NADH Dehydrogenase 2 | T → C | Syn |
| 6045 | Cytochrome c Oxidase I | C → T | Syn |
| 6152 | Cytochrome c Oxidase I | T → C | Syn |
| 8473 | ATP Synthase 8 | T → C | Syn |
| 10398 | NADH Dehydrogenase 3 | A → G | Thr → Ala |
| 10876 | NADH Dehydrogenase 4 | A → G | Syn |

| Base Position | Gene | Base Change | AA Change |
|---|---|---|---|
| 11251 | NADH Dehydrogenase 4 | A → G | Syn |
| 11467 | NADH Dehydrogenase 4 | A → G | Syn |
| 12308 | tRNA Leucine 2 | A → G | ---- |
| 12372 | NADH Dehydrogenase 5 | G → A | Syn |
| 12558* | NADH Dehydrogenase 5 | C → T | Syn |
| 12612 | NADH Dehydrogenase 5 | A → G | Syn |
| 13020 | NADH Dehydrogenase 5 | T → C | Syn |
| 13708 | NADH Dehydrogenase 5 | G → A | Ala → Thr |
| 13734 | NADH Dehydrogenase 5 | T → C | Syn |
| 14798 | Cytochrome b | T → C | Phe → Leu |
| 15452 | Cytochrome b | C → A | Leu → Ile |
| 15907 | tRNA Threonine | A → G | ---- |
| 16051 | D-Loop | A → G | ---- |
| 16069 | D-Loop | C → T | ---- |
| 16092 | D-Loop | T → C | ---- |
| 16126 | D-Loop | T → C | ---- |
| 16129 | D-Loop | G → C | ---- |
| 16183 | D-Loop | A → C | ---- |
| 16189 | D-Loop | T → C | ---- |
| 16291 | D-Loop | C → A | ---- |
| 16362 | D-Loop | T → C | ---- |
| 16366 | D-Loop | C → T | ---- |

**Table 2.1:** Results of complete mtDNA SNP analysis of unrelated CEPH individuals NA10859 and NA13116. In total, 44 base changes were detected, including eight not detected in previous PRS analysis (indicated in blue). C12558→T is a novel polymorphism not previously reported in literature.

### 2.5.2 Mutation Detection in Clinically Relevant Tumor Samples

With the successful validation of the optimized PRS protocols on non-clinical samples, further characterization was undertaken on authentic mtDNA samples from cancer patients. To this end, six sets of paired total DNA (tumor and normal) were obtained from the laboratory of David Sidransky at Johns Hopkins University. Three cases of lung cancer and three bladder cancers were included in this sample set. In all cases, "tumor" DNA was extracted from sections of a solid primary tumor, while "normal" DNA was derived from whole blood. All samples were fully extracted at Johns Hopkins and included no identifiable information at the time of acquisition.

Prior to PRS analysis, all cancer samples were analyzed for somatic variation using the Human MitoChip (Affymetrix), a commercial microarray targeted for high-throughput mtDNA sequence analysis. In its original incarnation, the MitoChip consisted of roughly 300,000 "features," each 24 x 20 μm, and each synthetically modified with ~$10^6$ copies of a unique 25-nt probe designed to interrogate the identity of a particular base position in the sequence. The basecall at each position is determined by the relative hybridization of fluorescently-labeled mtDNA restriction fragments to four nearly-identical probes differing only by the middle (13th) base. Raw fluorescence intensities for each set of probes are measured and interpreted using an adaptive background genotyping algorithm in which a quality score is assigned to each possible basecall using a statistical model including raw fluorescence intensity, as well as the "expected" basecall based on a reference sequence and the likelihood of each particular mutation (for example, a transition between purine or pyrimidine bases is roughly twice as likely as a transversion, in which the ring structure of the base is altered). This original MitoChip (v1.0) was capable of sequencing both strands of all coding regions, with all genes except the 12S and 16S rRNA's tiled twice. The noncoding D-loop, however, was excluded due to size limitations. In early validations, the MitoChip (v1.0) yielded accurate sequencing data at 96.54% of all interrogated base positions [131]. In 2006, a second-generation MitoChip (v2.0) was introduced, increasing the redundancy of coding-region coverage to 100% and adding effectively 2x coverage of the D-loop. In addition, adjustments were made to improve the accuracy of the adaptive background genotyping software [132]. To compare the performance of PRS to that of both MitoChip generations, the three lung cancer samples were subjected to MitoChip (v1.0) analysis, while the three bladder cancer samples were analyzed using MitoChip (v2.0) [133].

Due to sample availability, it was necessary to undertake an abbreviated analysis, analyzing only mtDNA amplicons containing previously identified somatic mutations (PCR primer pairs 2, 5, 7, and 8 for lung sample 1; primers 5 and 7 for lung 2; primers 2, 4, and 5 for lung 3; primer pair 6 for bladder sample 1; primers 1 and 2 for bladder 2; and primer pair 1 for bladder 3). MitoChip (v2.0) samples were sufficiently abundant to allow complete PRS analysis of the D-loop (primer pairs 23 and 24) as well to characterize MitoChip performance in this newly added region. Samples were amplified, extended, purified, separated, and analyzed using previously detailed protocols. Results are tabulated in Table 2.2.

**MitoChip v1.0**

| Sample | Base Position | Gene | MitoChip | | PRS | | Direct CE | |
|---|---|---|---|---|---|---|---|---|
| | | | Blood | Tumor | Blood | Tumor | Blood | Tumor |
| Lung 1 | G 1719 | 16S rRNA | * | R | * | * | * | * |
| | A 3385 | NADH Dehydrogenase 1 | G | R | G | G | G | G |
| | C 3450 | NADH Dehydrogenase 1 | T | Y | T | T | T | T |
| | A 3480 | NADH Dehydrogenase 1 | * | R | * | * | * | * |
| | A 4901 | NADH Dehydrogenase 2 | G | R | G | G | G | G |
| | G 5773 | tRNA Cysteine | A | R | A | A | A | A |
| Lung 2 | C 3450 | NADH Dehydrogenase 1 | * | Y | * | * | * | * |
| | G 4580 | NADH Dehydrogenase 2 | A | R | A | A | A | A |
| | A 4901 | NADH Dehydrogenase 2 | * | R | * | * | * | * |
| Lung 3 | G1719 | 16S rRNA | G | A | A | A | A | A |
| | G 3022 | 16S rRNA | * | * | * | R | * | R |
| | A3480 | NADH Dehydrogenase 1 | A | G | G | G | G | G |
| | G 10427 | tRNA Arginine | * | R | * | * | * | * |
| | T 10885 | NADH Dehydrogenase 4 | * | Y | * | * | * | * |
| | A 11083 | NADH Dehydrogenase 4 | * | R | * | * | * | * |

**MitoChip v2.0**

| Sample | Base Position | Gene | MitoChip | | PRS | | Direct CE | |
|---|---|---|---|---|---|---|---|---|
| | | | Blood | Tumor | Blood | Tumor | Blood | Tumor |
| Bladder 1 | A4009 | NADH Dehydrogenase 1 | C | T | * | * | * | * |
| | C4011 | NADH Dehydrogenase 1 | * | * | T | T | T | T |
| Bladder 2 | A1188 | 12S rRNA | T | C | * | * | * | * |
| | T1189 | 12S rRNA | * | * | C | C | C | C |
| | A1810 | 16S rRNA | * | G | * | * | * | * |
| | A1811 | 16S rRNA | * | * | G | G | G | G |
| | C303 | D-loop | ---- | ---- | CCC | C | CCC | C |
| Bladder 3 | T720 | 12S rRNA | * | C | * | * | * | * |
| | T721 | 12S rRNA | * | * | C | C | C | C |
| | C303 | D-loop | ---- | ---- | CC | C | CC | C |

**Table 2.2:** Results of somatic mutation analysis of cancer samples by MitoChip (v1.0 and v2.0), PRS, and direct sequencing. Asterisks indicate agreement with the Revised Cambridge Reference Sequence; purine heteroplasmies are indicated by R; pyrimidine heteroplasmies by Y. In all cases of MitoChip-reported somatic variation, none was detected by PRS or conventional sequencing. In addition, three novel mutations (indicated in blue) not detected in MitoChip analysis were uncovered.

Surprisingly, in each of the 14 cases of somatic variation reported in MitoChip (v1.0) analysis, no variation was detected by PRS. Particularly noteworthy among these discrepancies are the two homoplasmic mutations in lung cancer sample 3. If indeed present, these variants would have been detected readily by PRS. Interestingly, while PRS detected no somatic variation at either base position, both basecalls are germline polymorphic relative to the revised Cambridge reference sequence [52]. In subsequent MitoChip (v1.0) analyses using an adjusted mutation-calling quality score threshold, both homoplasmic mutations were identified as false positives. The remaining 12 heteroplasmic variations, however, were upheld as accurate, although independent analysis by both PRS and conventional sequencing found no variation at these positions. In addition to these false positives, one authentic mutation (G3022→R) found during PRS analysis and subsequently verified by direct sequencing was not detected by the MitoChip. The omission of this variant is particularly troubling, as the heteroplasmic shift is over 60% mutant.

Similarly, all four somatic mutations detected in MitoChip (v2.0) analysis were found to be erroneous, exhibiting not even germline deviation from the standard Cambridge sequence. In each of these cases however, adjacent germline SNP's were uncovered within 1-2 bases of the reported mutation, with the polymorphic basecall matching the alleged mutant basecall in each instance. Two additional variants, both somatic deletions in the D310 tract, were also uncovered by PRS analysis. Subsequent verification by direct sequencing confirmed both of these mutations to be shifts from germline heteroplasmies, with both normal and tumor genomes exhibiting differing degrees of heteroplasmy in the D310 region. This class of mtDNA variation is particularly difficult to detect and had not been encountered in any previous PRS study.

The discrepancies between MitoChip and PRS results can be understood in the intrinsic approaches and limitations of each technique. A Sanger sequencing-based technique, PRS generates comparative sequencing data with little prior knowledge of the sample genome. Conversely the MitoChip, like all sequencing microarrays, involves hybridization of the sample to a series of synthetic oligonucleotide probes and thus requires strict adherence to a prescribed sequence. Consequently, accurate characterization at a particular base is highly dependent on local sequence variation in the sample, as well as G/C content in the 25-nt region surrounding the probed base and probe sequence homology with restriction fragments generated from other regions of the genome. This shortcoming may be related to the recurrent miscalls at base positions 1719, 3450, 3480, and 4901 during MitoChip (v1.0) analysis. In addition, the final data produced in MitoChip analysis are a series of distinct basecalls based on a statistical model applied at each base position. This approach tends to generate false mutation calls in regions involving any variation, germline or somatic, from the standard reference sequence, and is the most likely explanation for the four miscalls in the MitoChip (v2.0) analysis. To these limitations are added the nonlinear thermodynamics of hybridization, which preclude quantitation of mutant populations, and an inability to detect single-base insertions and deletions, which are not accounted for in probe design. The omission of the two D310 variants in bladder specimens 2 and 3 is explained by both, as both alterations represent heteroplasmic short-range deletions.

## 2.6 Conclusions

Substantial improvements in the collection and analysis of PRS data have been advanced. Implementation of these changes has fostered a four-fold reduction in total analysis time, and operation of the μCAE device has been simplified dramatically. A demonstrably higher sensitivity has been also achieved, as evidenced by the accurate detection of eight verifiable SNPs omitted from similar experiments using original PRS protocols. Moreover, the PRS assay has been successfully validated on clinically relevant samples from cancer patients, correctly identifying three heteroplasmic somatic variants in six specimens, including two deletions in the D-loop. These mutations were not detected in a separate analysis using a high-throughput commercial resequencing microarray, although 18 false mutations were erroneously called in microarray analysis (correctly characterized by PRS). These results showcase PRS as a promising technique for accurate and sensitive detection of genetic cancer markers, both established and novel, and lay the groundwork for full-scale screening of multiple biological samples, which will be addressed in the next chapter.

# Chapter 3

## High-Throughput Detection of Somatic Bladder Cancer Mutations in the Mitochondrial Genome by Polymorphism Ratio Sequencing

-or-

## *Mito-Man vs. Bladder Cancer*

## 3.1 Introduction

Historically known as the "powerhouse of the cell," the mitochondrion plays a powerful and paradoxical role in cell survival and proliferation. In addition to providing over 80% of the cell's energy through oxidative phosphorylation (OXPHOS), mitochondria also participate in the synthesis of such biomolecular building blocks as amino acids, phospholipids, and heme. More morbidly, permeation of the outer membrane in response to intracellular signaling cascades is a key step in apoptosis, the suicide pathway by which irreparably damaged cells are eliminated and tissue proliferation is restricted [43]. Mitochondrial dysfunction has been suspected as a major factor in tumor onset and progression since as early as Warburg [46], and recent years have seen experimental verification of altered mitochondrial content and function in human cancer [49-50].

In addition to harboring the machinery of life and death, the mitochondrion also contains its own compact, circular genome of 16,569 bp. The mitochondrial DNA (mtDNA) encodes 22 tRNA's, two rRNA's, and 13 proteins, all of which form functional subunits in mitochondrial OXPHOS structures. Additionally, a noncoding control region contains promoters and other factors regulating transcription. In the past decade, somatic mutations in the mtDNA have been uncovered in a wide range of cancers including head and neck [68, 84-85], breast [66, 69-70], kidney [71], liver [72-73], colorectal [74-75], pancreatic [80], brain [83], bladder [68, 86], and lung [67-68, 87-88]. More recently, the functional role of these mutations has begun to come into focus, with the demonstration of increased tumor growth and metastasis in the presence of coding mutations [86, 93].

The exploitation of mtDNA sequence variation as a viable tool in cancer detection and characterization presents a variety of unique benefits and challenges. The relative abundance of the mtDNA—hundreds to thousands of copies per cell—makes the genome an attractive target for genetic analysis, enabling the acquisition of high-quality data from minimal sample. In addition, its proximity to the reactive oxygen species (ROS) generated as by-products of electron transport renders the mtDNA more susceptible to damage and subsequent mutation, providing a direct window for observing the cycle of oxidative damage believed to be implicated in a variety of mitochondrial diseases [44, 58-60]. More problematic is the issue of heteroplasmy, the phenomenon by which multiple genetic populations may exist within a single cell as a result of mitochondrial multiplicity. While tracking heteroplasmy shifts over time may ultimately prove useful in monitoring tumor progression, assigning numerical significance to these changes is difficult with conventional sequencing, hybridization, and PCR-based technology due to the non-linearity and extensive data processing associated with these techniques.

Recently we introduced Polymorphism Ratio Sequencing (PRS), a novel Sanger sequencing-based technique that allows for direct and quantitative determination of sequence variation between two templates using the modified coding and pooling scheme depicted in Figure 3.1. In separate reaction vessels, Sanger extension fragments corresponding to the same terminating base are labeled with distinct dye primers identifying the template. These fragments are then pooled and separated electrophoretically, yielding an electropherogram that overlaps perfectly in all places except in instances of variation. Monitoring the squared difference between corresponding traces provides a direct indication of variant location, producing a nonzero value only in regions of poor overlap. Moreover, the inclusion of an internal control provides a standard for variant quantitation, with peak ratios linear in genotype population to a detection limit of 5% mutant frequency. With four-color detection capabilities, two bases (by

convention A/C and G/T) may be monitored simultaneously. The coupling of A/C and G/T requires multiple trace correlation to confirm single-base transitions, the most common mutation class, building in a layer of redundancy and reducing the effects of single-sample anomalies [105]. In initial validations, labeled and pooled fragments were analyzed on a 96-lane microfabricated capillary array electrophoresis (μCAE) bioprocessor. The modest sample injection volume afforded by the microdevice eliminates the need for abundant starting template and considerably reduces overall reagent cost. In addition, the microfabricated platform lends itself well to integration [118, 138, 145] and parallelization [110, 114-116], thus decreasing operational intensity while increasing throughput.

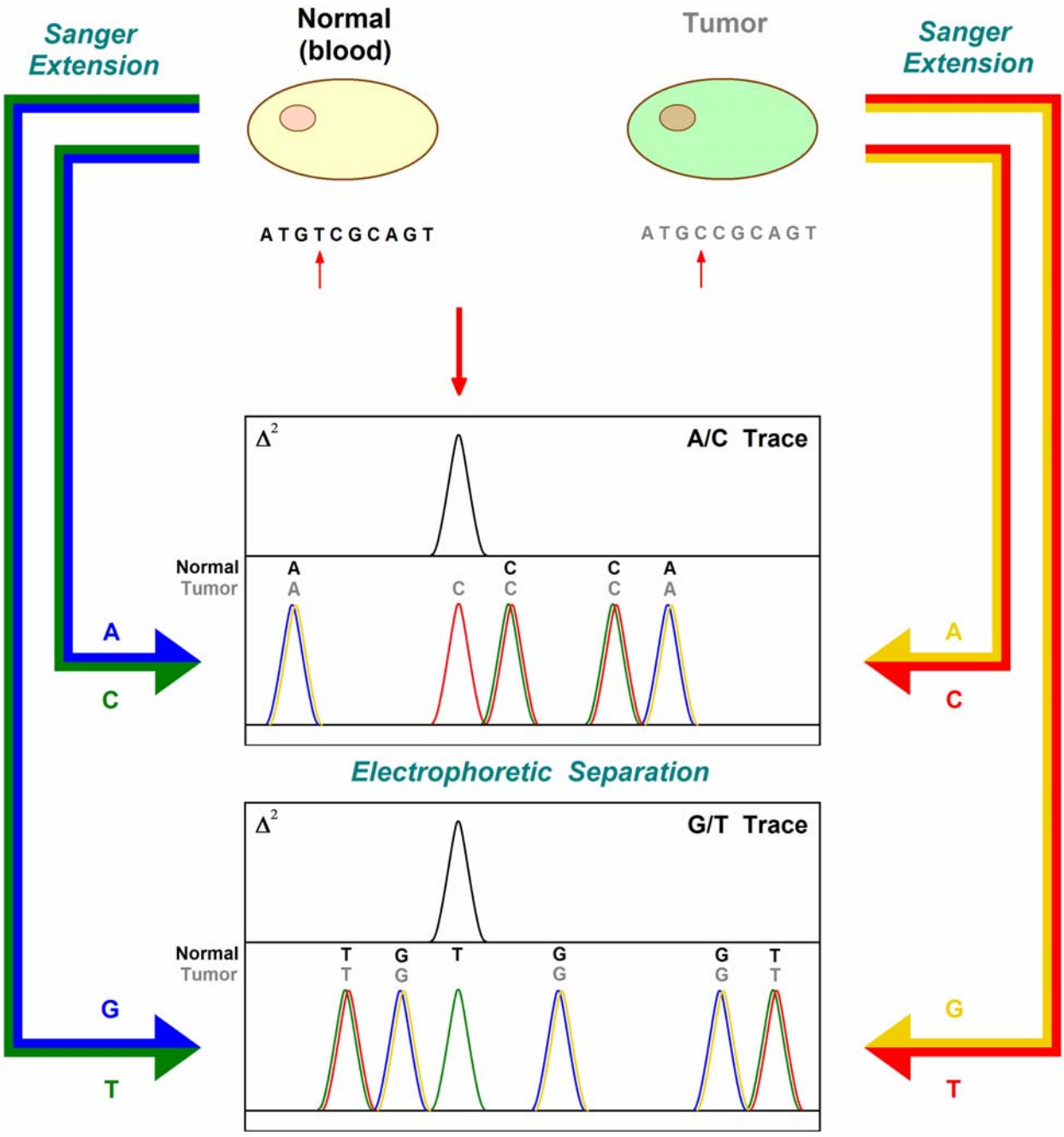



**Figure 3.1:** PRS labeling and pooling scheme for detection of tumor mutations. Like-terminating bases from normal and tumor samples are independently Sanger-extended and labeled with differing dye primers to identify the template. Extension fragments are then pooled in two four-color-compatible mixtures (by convention, A/C and G/T) and separated, yielding electropherograms with perfect two-color overlap except in regions of variation. Mutations are readily detected as single peaks in a squared difference plot and characterized based on the four-color peak signature. In this case, a T→C transition is identified as a lone "T normal" (green) peak in the G/T trace and a corresponding lone "C tumor" (red) peak in the A/C trace.

In this report we detail the application of microchip-based PRS analysis to the high-throughput detection of somatic mtDNA variants in a pool of 14 bladder cancer cases. In all cases, mtDNA extracted from blood serves as a normal internal control by which tumor-associated mutations are differentiated from germline polymorphisms. The analysis of available matched urine mtDNA is also discussed to explore the use of bodily fluids as potential noninvasive indicators of tumor infiltration or recurrence under a PRS analysis scheme. The results obtained here highlight the advantages of PRS as a technique for sensitive, high-throughput mutation detection in clinical samples.

## 3.2 Materials and Methods

### 3.2.1 Microfabrication

The design and fabrication of the μCAE device have been described previously [109-110, 115]. Briefly, the device consists of 96 separation capillaries arrayed radially on a 150-mm Borofloat glass wafer and converging on a common anode. Each capillary is 17 cm long, 200 μm wide, and 30 μm deep. Adjacent capillaries are coupled by common cathode and waste reservoirs, forming 48 doublets, each with a common injection structure. Microchannel features are formed by isotropic wet etching with concentrated HF and sealed by thermally bonding a second blank wafer to the channel layer. Fluidic access ports are diamond-drilled into the etched layer. A final layer of 1/8"-thick Borofloat rings replacing the PMMA buffer moats previously employed [115] is attached to the top of the device in a second thermal bonding step.

### 3.2.2 Sample Acquisition and Quantitation

Extracted total DNA samples from blood and tumor were obtained for 14 individuals with bladder cancer from the lab of Dr. David Sidransky at Johns Hopkins University. Where available (5 of 14 individuals), matched urine DNA was also acquired. All samples were quantitated using the PicoGreen assay (Invitrogen) and the grouping by individual verified by PowerPlex 16 STR genotyping analysis (Promega).

### 3.2.3 Template Amplification

Generation of template for PRS extension was performed by PCR amplification of the entire mitochondrial D-loop in two overlapping fragments of length 765 and 954 bp. Total DNA template (10 ng) was combined with 7.5 pmol each of forward and reverse PCR primer (Integrated DNA Technologies), 2.5 μL of 10X PCR Buffer (-$MgCl_2$), 0.5 nmol of each dNTP, 37.5 nmol of $MgCl_2$, and 1.25U of Platinum TAQ DNA Polymerase (Invitrogen) and diluted to a total reaction volume of 25 μL with nuclease-free water (ISC BioExpress). PCR primer sets 23 and 24 from Rieder *et al*. [129] were modified at the 5' end to include universal sequencing primer recognition (-40 M13 for forward primers and -28 M13 for reverse). PCR reactions were incubated in an MJ Research PTC-200 thermal cycler for 30 cycles of denaturation (94º C for 30

seconds), annealing (61º C for 45 seconds), and extension (72º C for 90 seconds). The thermal program was initiated with a 60-second denaturation at 94ºC and terminated with a final incubation at 72º C for 180 seconds.

PCR products were purified enzymatically by incubation with 30 U of exonuclease I and 2 U of shrimp alkaline phosphatase (GE Healthcare) for 30 minutes at 37º C, followed by 15 minutes at 80º C. PCR products were quantitated by PicoGreen and diluted to a final concentration of 12 ng/μL.

### 3.2.4 PRS Extension and Pooling

Sanger extension is performed using energy transfer (ET) dye primer labeling chemistry and has been reported previously in detail. For each uniquely labeled set of extension fragments, 4 μL (48 ng) of purified PCR product were combined robotically with 800 fmol (2 μL of 400 nM) of the appropriate universal ET sequencing primer (FAM-R110, FAM R6G, FAM-TAMRA, and FAM-ROX, with emission maxima 525, 555, 580, and 605 nm, respectively) [98] and 2 μL of DYEnamic Direct Cycle Sequencing Mix (GE Healthcare). The reaction was then diluted with water to a total volume of 10 μL. Thermal cycling consisted of 45 cycles of denaturation (95º C for 30 seconds), annealing (52º C for 15 seconds), and extension (72º C for 60 seconds). Extension fragments were pooled and purified by co-precipitation with 15 μg of glycogen (MP Biomedicals) in 3 μL of 7.5 M ammonium acetate (Fisher) and 110 μL of 100% ethanol (Rossville). After 60 minutes at -4º C, the reactions were centrifuged at 3700 rpm for 30 minutes and the pellets washed with 150 μL of 70% ethanol. The pellets were recollected by a second 5-minute centrifugation at 3700 and allowed to air dry before resuspension in 15 μL of 50% HiDi Formamide (Applied Biosystems).

### 3.2.5 Microchip Preparation and Operation

Prior to analysis, microchannels were coated using a modified Hjertén procedure [126]. Channels were primed with 1 mL of 1x TTE (50 mM tris, 50 mM TAPS, 1 mM EDTA, pH 8.3), and separation matrix (GenomeLab Separation Gel LPA-1, Beckman-Coulter) was loaded via the central anode using a high-pressure gel loader [130]. Excess matrix was evacuated from the sample reservoirs, and 1.8 μL of sample were added to each well. The microdevice was then transferred to the Berkeley Confocal Rotary Scanner [114], where a ring bearing 96 electrodes was affixed to the sample reservoirs. Cathode, anode, and waste moats were filled with 5x TTE buffer, and electrical connections were made to each. PRS fragments were injected electrokinetically at 30 V/cm for 270 seconds, followed by separation at 150 V/cm for 30 minutes. Both the injection and separation were performed at 67º C.

### 3.2.6 Data Analysis

Following data collection, electropherograms were processed using an automated MATLAB-based data analysis program. Raw electropherograms were first truncated, leaving only data between 7-27 minutes in the trace and removing primer and PCR runoff peaks. Data were baseline-corrected with a dynamic top-hat algorithm, and spectral deconvolution was performed using four-color peak intensities from a PCR-generated ET primer standard. Following global normalization of signal intensities, a dynamic time warp algorithm was employed to align normal and tumor traces temporally, using maximal overlap in nonvariant regions to define a unique quality score at each data point. Squared difference maxima exceeding three standard deviations were identified automatically as possible mutations. Secondary inspection of these regions using sequence context and four-color peak signature confirmed the presence of authentic variants.

## 3.3 Results

Comprehensive variant data are compiled in Table 3.1. Of 14 individuals with verifiably complete sample sets, seven (50%) harbored at least one somatic D-loop mutation. As confirmed by MITOMAP, the online mtDNA database [90], 10 of these variants are novel somatic mutations previously unreported in literature. One, the A15927→G transition in Sample 6, is an overall novel variant from the Cambridge Reference Sequence [52]. Consistent with previous D-loop mutation analyses, the most pervasive mutations between individuals were insertion/deletion changes in unstable repeat segments, notably the D310 tract (base positions 303-315) and the poly-C region from 16184 to 16193. Heteroplasmic mutations account for at least 15 of 21 (71%) total variants, appearing in both single base substitutions and insertion/deletion variants. Among these, five mutations occurring in Samples 6 and 13 appear to be population shifts from an already heteroplasmic germline. Where detected in the associated tumor, somatic mutations were also detected in urine, although to varying degrees. Typically these changes were more pronounced in urine than in the tumor itself, appearing as heteroplasmies biased more strongly toward the mutant genotype.

| Sample | Position | Normal | Tumor | Urine |
|---|---|---|---|---|
| 2 | 200 | G | R | R |
| | 303 | CCC | C* | C |
| 4 | 16183 | A | C | C |
| | 16184 | C | CC* | CC |
| 5 | 16184 | C | CC | CC |
| 6 | 152 | T | Y | --- |
| | 153 | G | R | --- |
| | 225 | A | R | --- |
| | 303 | CC | C | --- |
| | 15927 | R (A>G) | R (G>A) | --- |
| | 16126 | T | Y | --- |
| | 16163 | A | R | --- |
| | 16184 | CC | C | --- |
| | 16186 | C | Y | --- |
| | 16223 | Y (T>C) | Y (C>T) | --- |
| | 16278 | Y (T>C) | Y (C>T) | --- |
| | 16296 | C | Y | --- |
| 8 | 217 | T | C | --- |
| 10 | 16184 | C | CC | --- |
| 13 | 303 | CCC* | C | --- |
| | 514 | C* | CAC | --- |

R = Purine Heteroplasmy

Y = Pyrimidine Heteroplasmy

* = Insertion/Deletion Heteroplasmy

**Table 3.1:** Compiled D-loop mutation results. A total of 21 somatic variants were detected in 7 of 14 samples. At least 15 of these variants are heteroplasmic, and five represent population shifts from a heteroplasmic germline. All mutations detected in individuals with available urine DNA were readily detected in the bodily fluid.

With a total of twelve somatic variants, Sample 6 was by far the most frequently mutated individual. Selected representative data from this set are illustrated in Figure 3.2. In addition to single-base deletions in both major poly-C tracts, nine transition substitutions, all heteroplasmic, are present. Of these, three represent shifts from a heteroplasmic germline. These heteroplasmy drifts are typified by the A→G shift at base position 15927 and the T→C shift at position 16278. Notably, many of these mutation sites (5 of 12) were germline polymorphic relative to the Revised Cambridge Reference Sequence and may be indicative of a high-risk mitochondrial haplotype.

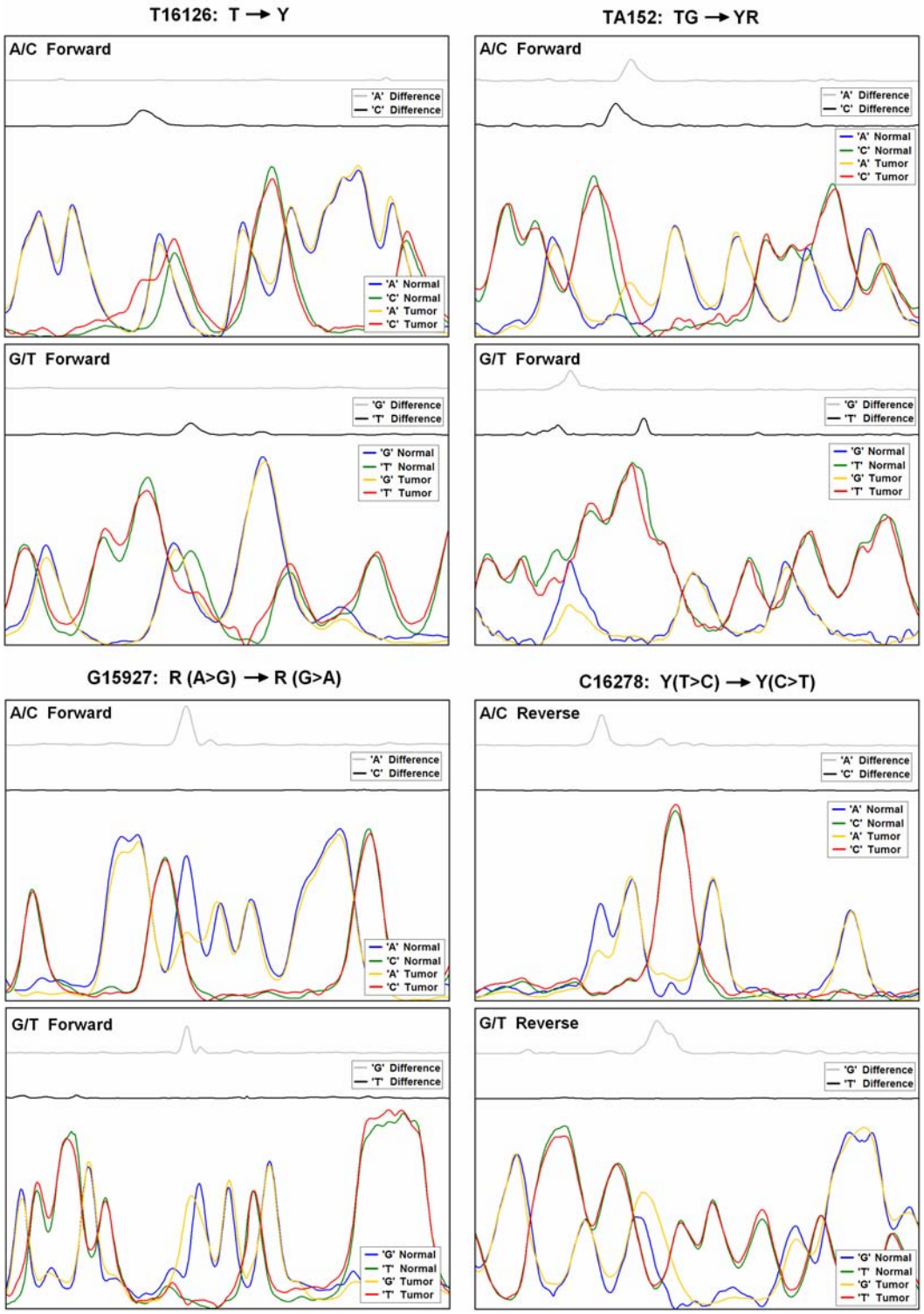



**Figure 3.2:** Selected PRS data from highly mutant Sample 6 (normal/tumor). Mutations detected in this sample were predominantly heteroplasmic, as typified by the T→Y transition at T16126 and the double TG→YR transitions at TA152. Also detected were shifts in germline heteroplasmies such as the R(A>G)→R(G>A) transition at G15927 and the Y(T>C)→Y(C>T) transition at C16278 (read in reverse complement as R(A>G)→R(G>A)).

Selected data for mutations detected in matched tumor and urine are compared in Figure 3.3. In all instances of tumor-associated mutation, the corresponding variant was detectable in urine. Curiously, the majority of these variants (3 of 5) are more strongly pronounced in the urine than in tumor, as indicated by both a striking mismatch in the overlaid traces and the intensity of the squared difference in these regions. This observation is true universally for Sample 2, the origin of both the G200→R transition and the D310 deletion in Figure 3.3. In Sample 5, the single-C insertion is equally apparent in tumor and urine, as is the A16183→C transversion in Sample 4. Conversely, the CC insertion in Sample 4 is less apparent in tumor than urine, more closely resembling a homoplasmic C insertion in tumor. In all cases, however, somatic variation is readily detected in the PRS data from both tumor and urine.

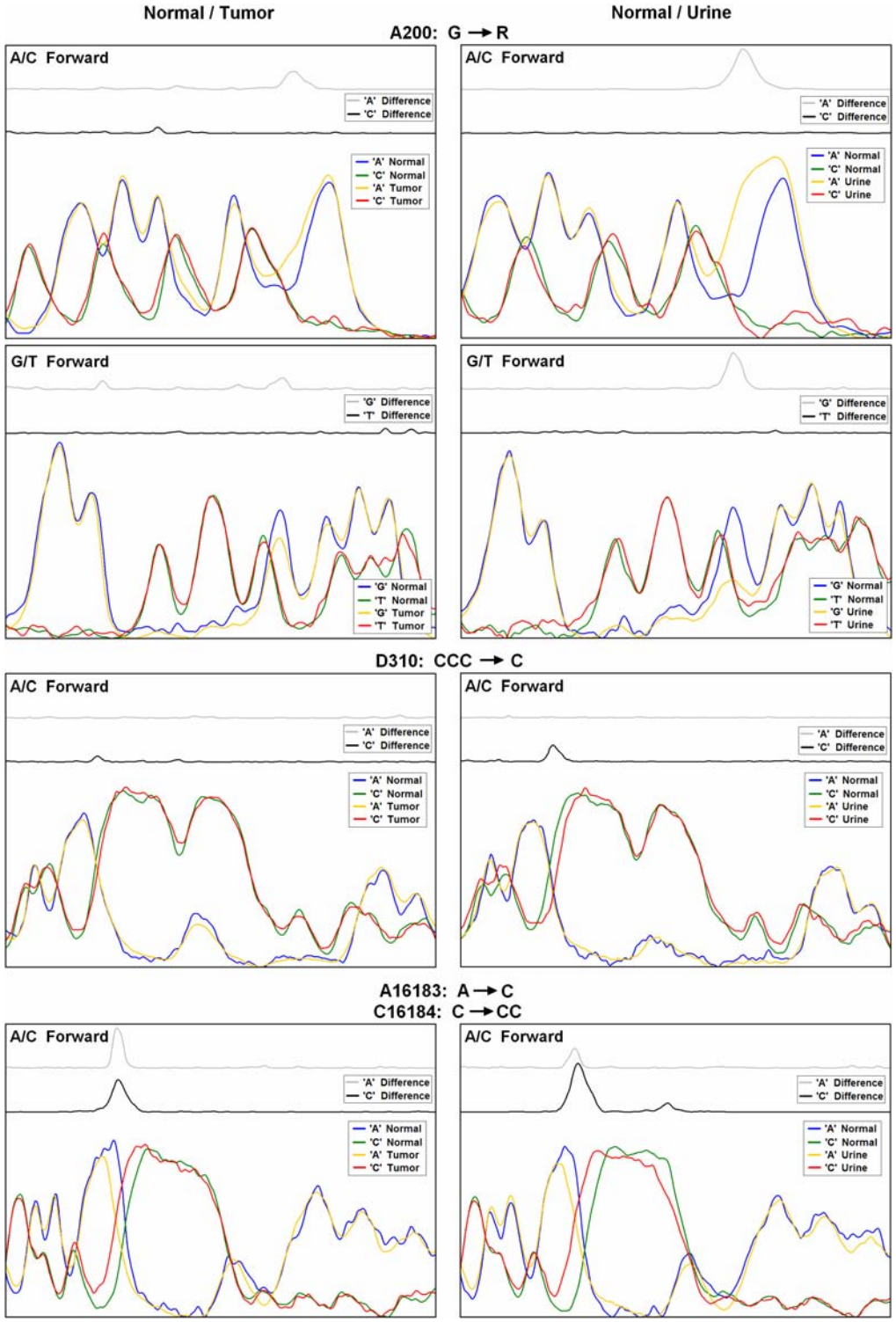



**Figure 3.3:** Sample PRS data for matched normal/tumor and normal/urine. In 3 of 5 cases, the base in question was more strongly variant in urine than tumor, as evidenced by both the G200→R transition and the D310 deletion (both Sample 2). The A16183→C transversion in Sample 4 is more consistent in intensity between tumor and urine, although a single-C insertion in the same poly-C region was more readily observed in urine.

## 3.4 Discussion

Owing to its uncommonly high mutation frequency and the variety in detected alterations, Sample 6 represents an interesting singular example of the nature of mitochondrial mutations observed in bladder cancer. In addition to the sheer number of variants, the mutations detected in this sample are almost universally heteroplasmic, with three representing shifts in heteroplasmy from germline to tumor. The high proportion of heteroplasmic drift observed in this sample is consistent with a model of mtDNA mutation accumulation originating from enhanced ROS generation by a damaged electron transport chain. While insufficient to establish functional relevance, the detection of four alterations within the heavy strand origin—two in highly conserved regions—suggests a potential effect on replication efficiency. More interestingly, the A→G heteroplasmy shift at base position 15927 occurs in the tRNA threonine gene, 97 bp outside of the control region. The detection of this variant in the modest fraction of coding sequence (450 bp) spanned by the PCR amplicons employed here corroborates previous findings of somatic mtDNA coding mutations in cancer [70-71, 74, 76, 79-80, 83, 87] and suggests their presence at reasonably high frequency. Complete mtDNA sequencing in coding regions may thus uncover an even greater number of mutations while providing insight into the biomolecular mechanism of respiratory chain failure. It should be noted that full-mtDNA PRS is possible using the 24-amplicon scheme of Rieder *et al*. and can be performed handily for a single individual on a single 96-lane μCAE device.

The observation of higher mutant populations in urine than in paired tumor, though not universal, is a common theme in these results. Intuition suggests that any cancer-associated mutation should be more readily detected in a carefully microdissected tumor sample than in the random collection of cells shed by the total urothelium. Thus, even equal mutant representation in tumor and urine, as is the case in Sample 5, is surprising. The inconsistency in the relative representation of the two mutations detected in Sample 4 between tumor and urine rules out imprecise microdissection as a sole explanation and suggests authentic biological relevance. Indeed, equal or greater mutant representation in urine has been observed in previous sequencing studies of matched tumor and bodily fluid [68]. Most probably, the excessive mutant populations in urine are indicative of an enhanced rate of cell loss from the tumor relative to the healthy epithelium, coupled with the advanced effects of mutation accumulation at the time of loss, a mechanism most likely linked with the acquisition of metastatic potential in advanced tumors.

Also noteworthy among mutated individuals is Sample 13, in which a germline heteroplasmic two-base insertion/deletion between the CA514 dinucleotide repeat and the D310 tract is perfectly corrected in the tumor. The PRS traces containing the double indel are illustrated in Figure 3.4, the mutation appearing as a series of minor normal pre-peaks appearing before the more intense dominant peaks, traceable by the tumor tracks between the CA insertion at 514 and the CC deletion at 303. This variant is particularly difficult to detect and characterize even by PRS, as complete data analysis requires the recognition and extraction of a systematic trend relative to the expected track, which in this example is represented by a homoplasmic tumor.

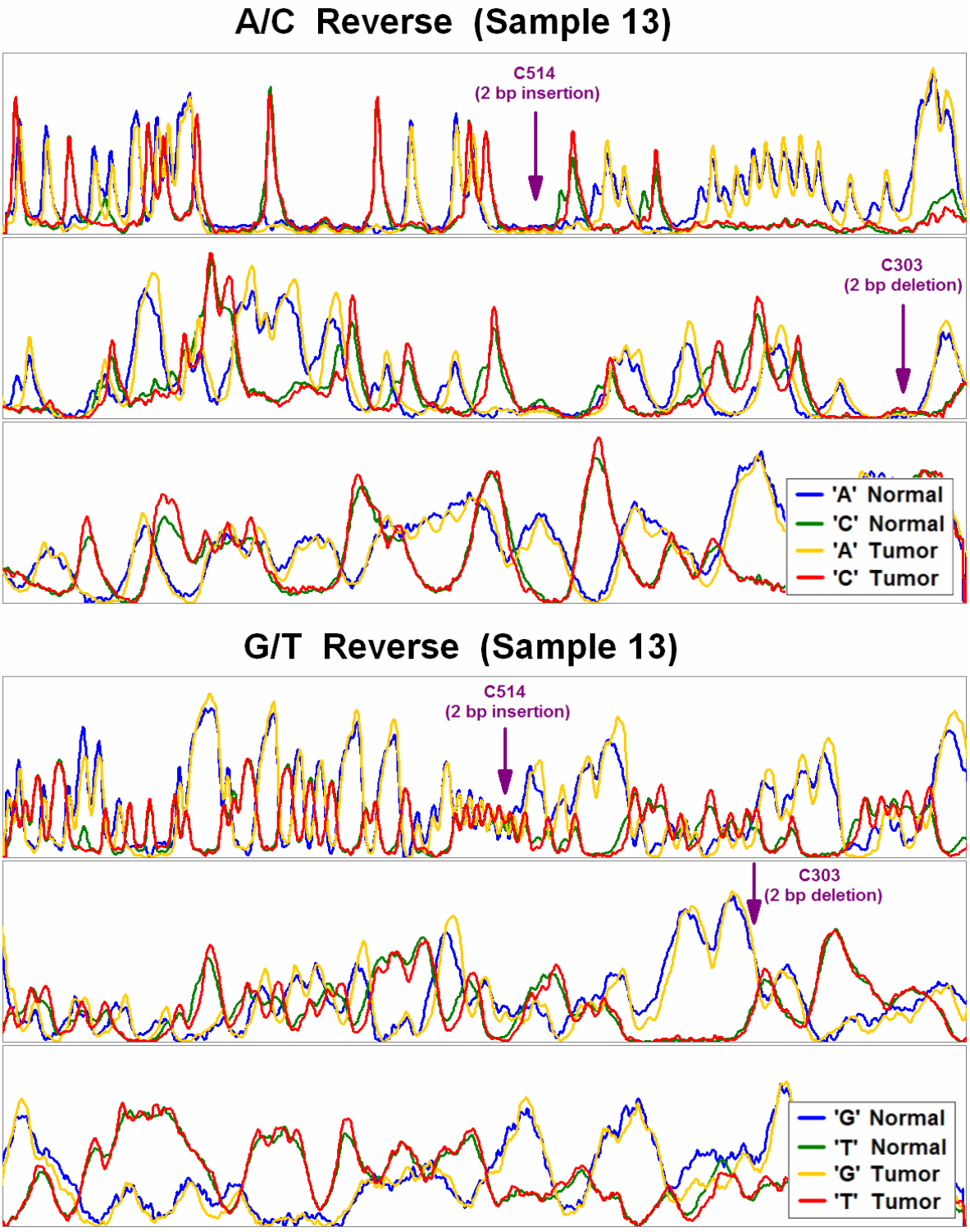



**Figure 3.4:** Double heteroplasmic insertion/deletion in Sample 13. A two-base splitting of all normal traces between C514 and C303 indicates a germline heteroplasmic in a double insertion/deletion between these positions (C514→CAC and CCC303→C). The absence of splitting in the tumor traces indicates a somatic shift toward the dominant germline population, producing homoplasmy in tumor.

While mainstream sequencing techniques, both Sanger-based and next-generation, possess the ability to generate sequencing data in high volume, the sheer expanse of data and lack of quantitative controls limit their performance in mutation detection, particularly in the context of genetically heterogeneous samples. In contrast, PRS employs a normal internal control to highlight differences in sequence character, thereby allowing for direct variant detection without the need for multiple trace analysis and manual sequence comparison. Moreover, the internal control affords heightened sensitivity to subtle differences in sequence, as evidenced by the heteroplasmic transition at A200 in Sample 2. While striking in urine, this particular variation is weak in the tumor and would likely have been missed by conventional sequencing, in which a 25% signal intensity cutoff is commonly employed for basecalling purposes.

As a highly controlled Sanger-based technique, PRS also offers distinct advantages over microarray technology, currently among the leading techniques for high-throughput mutation detection. The quantitative relationship between peak ratio and genotype population allows for accurate mutation determination, in contrast to the fundamentally nonlinear and probe-dependent hybridization approach employed by microarrays [108]. Additionally, PRS requires minimal data processing, thereby preserving a maximal amount of raw data while eliminating the need for statistical modeling. The use of an internal standard derived from the individual's own germline also circumvents comparison to an arbitrary sequence standard such as the Revised Cambridge Reference mtDNA Sequence, an algorithm commonly built into microarray analysis scripts. Consequently, authentic somatic mutants are identified cleanly without the confounding influence of deviation from an "expected" germline.

As cancer therapy enters an age of personalized healthcare, the requirement of accuracy in comparative sequencing technology will be paramount in ensuring proper diagnosis and treatment. The common PHRED-20 standard in conventional sequencing corresponds to 99% confidence in a particular basecall. Translated to sequence accuracy, this standard permits an unacceptable 17 errors per individual in the span of sequence analyzed here. Indeed, these concerns rise in concurrence with recent phylogenetic studies indicating high error rates in published sequencing-based mutation studies [106]. In addition to addressing these concerns through rigorous controls and quantitation, PRS presents the opportunity to interpret genetic heterogeneity, either in the context of mitochondrial heteroplasmy as observed here, or in the analysis of potentially heterogeneous tissue, as is observed in tumor margins. A somatic mutation reliably linked with a particular cancer may therefore be exploited by PRS as an indicator of tumor invasion and used to define margins either prior to or during excision. Such an approach is applicable to any variant for which flanking primers are known, enlisting a library of thousands of potential markers. Coupled with the small sample sizes and rapid analysis times enabled by microchip analysis, a variety of key genes may be analyzed in parallel, elucidating the complete and unique genetic signatures of individual cancers at minimal cost. Integration of massively parallel μCAE separation technology with upstream PRS labeling and purification processes [134, 135] will ultimately foster more rapid and facile analyses, while further integration with emerging technology for single-copy genetic analysis [139] will enable real-time cancer pathology on the individual cellular level.

## 3.5 Conclusion

Polymorphism Ratio Sequencing, a novel technique for quantitative detection of genetic variation, has been applied to a sequence comparison of normal and bladder tumor mtDNA in the highly variable control region, identifying a high frequency of somatic mutations in half of the individuals tested. These mutations are predominantly heteroplasmic and often appear at low levels, emphasizing the need for sensitivity in the sequencing-based characterization of cancer. Mutants detected in tumor were readily detected in matched urine DNA where available, suggesting testing of bodily fluid as a feasible alternative to tissue analysis for tumor detection and characterization. The results obtained here emphasize the sensitivity and accuracy of PRS methodology and establish the technique as a viable approach to high-throughput detection of cancer-associated genetic markers.

## Acknowledgements

The authors gratefully acknowledge Nadia Del Bueno, Samantha Cronier, Numrin Thaitrong, Jim Scherer, and Peng Liu for valuable consultation, as well as Santanu Dasgupta, Mariana Brait Rodrigues DeOliveira, Mohammad Hoque, and David Sidransky at Johns Hopkins University for assistance in sample procurement. The involvement of Jing Yi and Terry Speed from the UC Berkeley Statistics Department was central to the design of the automated PRS analysis software. Microfabrication was performed by Eric Chu at the UC Berkeley Microlab. This work was supported by NIH grants #HG03329 and #CA77664.

# Chapter 4

## Prospects and Conclusions

-or-

## *Mito-Man vs. the World!!!*

## 4.1 Introduction

The results of the previous chapters have established the applicability of PRS in the quantitative detection of genetic cancer markers, as well as its superiority over leading commercial techniques. In this final chapter, I discuss the prospects for future application of PRS methodology using microfabrication as a vehicle to greater accessibility. As the goal of personalized healthcare through individual genetic analysis draws nearer, the final strides will be made through decreased reagent cost and sample analysis time. Integration of on-chip sample processing on the nanoliter scale with quantitative sample recovery, rapid separation by μCAE, and sensitive fluorescence detection will eliminate the inefficient sample manipulation currently required for analysis. Novel technologies are emerging on all fronts and hold the potential for realizing this goal.

## 4.2: Toward Fully Integrated Comparative Sequencing

For the past decade, the dogma of microanalysis has held that once an assay has been developed and validated on a microfabricated platform, its throughput should be increased through integration and parallelization. In its present state, PRS is readily applicable to high-throughput analysis systems: indeed, the validations presented here were all performed on a massively parallel device. Thus, the next challenge in bringing PRS to microchip maturity is integration of upstream sample processing steps with on-chip separation and analysis.

An overall PRS sample processing scheme is presented in Figure 4.1, along with an inventory of currently developed microfabricated devices capable of performing these tasks. Following collection of raw biological sample, cells must first be lysed and the DNA extracted. Purified DNA template is then amplified by PCR, and the amplification products are themselves purified to remove unextended primers and unincorporated dNTP's. Sanger extension is then performed using PCR product as template, followed by PRS pooling and a final purification step to remove unincorporated dNTP's, ddNTP's, enzymes, and buffer from the extension reaction. Sample is then injected into the capillary, separated, and detected, with subsequent data processing and mutation identification steps completing the analysis.

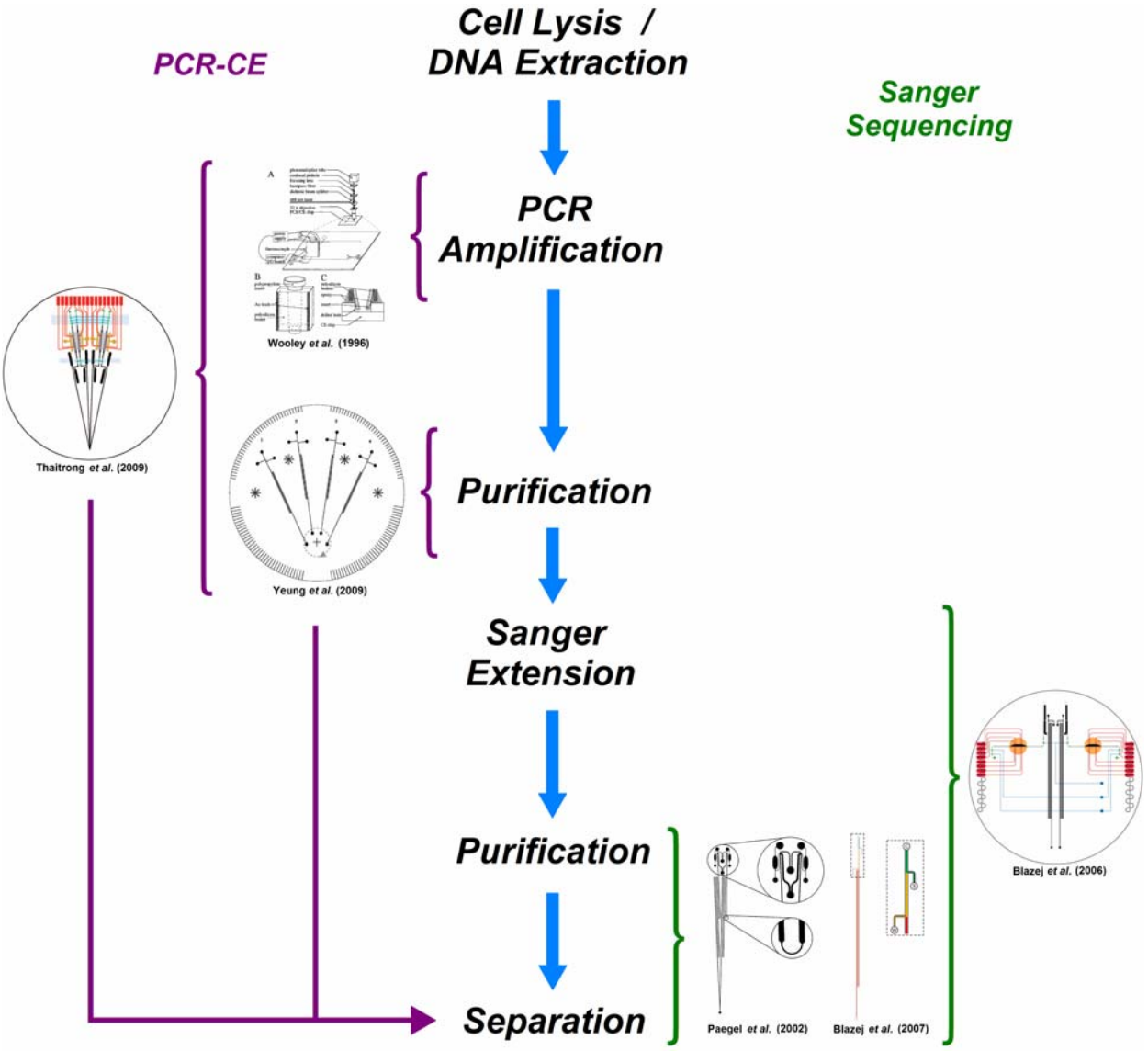



**Figure 4.1:** Forerunners of integrated PRS. Since as early as 1996 (Woolley *et al.*), thermal cycling has been performed in tandem with CE separations on microchips. Development of affinity matrix-based sample purification (Paegel *et al.*, 2002) and subsequent inline modifications (Blazej *et al.*, 2007, Yeung *et al.*, 2009) provided a method for efficient cleanup and quantitative injection of labeled sample. Total integration of thermal cycling, purification, and CE has been achieved for both conventional sequencing (Blazej *et al.*, 2006) and PCR (Thaitrong *et al.*, 2009).

The first step toward integrated sample preparation and electrophoretic analysis of DNA fragments on a single microchip came in 1996, with the development of a tandem PCR-CE device. This device was capable of performing PCR amplification in a microfabricated 20-μL thermal cycling reactor using miniature off-chip heaters and temperature sensors. Following extension, PCR product was directly injected through a microfabricated injection arm filled with sieving matrix into a 4 cm-long separation capillary. The reduced thermal mass of the PCR chamber allowed for more efficient heating and cooling, reducing the total reaction time from two hours to 15 minutes [118]. While this tandem PCR-CE device represented the first functional coupling of microfabricated thermal cycling and separation structures, the device was in fact a hybrid of two separate units, and the lack of an intervening purification step severely limited injection efficiency.

The issue of integrated sample purification was addressed in 2002 with the advent of integrated affinity-based oligonucleotide capture. This technique makes use of functionalized acrylamide polymers, selectively immobilizing extended DNA fragments and isolating pure sample without the need for off-chip precipitation or pre-electrophoresis in an orthogonal dimension. The capture matrix is prepared by introducing a specific 20-base oligonucleotide complementary to a common sequence on the desired products into the reaction cocktail during acrylamide polymerization. This capture oligo is modified at the 5' end with an acrydite moiety capable of incorporating into LPA chains. The resulting polymer bears a covalently attached affinity probe designed to hybridize selectively to the desired products. Sample is purified by driving raw thermal cycling reaction products electrophoretically through the capture matrix. Target molecules hybridize quantitatively to the matrix and become immobilized, while buffers, unincorporated nucleotides, and enzymes electrophorese through the gel to a waste reservoir. The bound products are then released thermally prior to injection and separation. In its first on-chip application, this technology was used to capture sequencing reaction products generated off-chip, yielding 99%-accurate read lengths over 500 bases on a 16-cm long separation capillary, and reducing an hour-long precipitation-based procedure to a mere two minutes [117].

Building on these technologies, complete integration of thermal cycling, product purification, injection, and separation was achieved in 2006 with the first integrated sequencing device. Starting from 1 fmol of DNA template, multiplexed four-color Sanger extension is carried out using dye-terminator labeling in a single 250-nL reactor. Following rapid thermal cycling, the extension products are pumped pneumatically into a hold chamber, from which a positive capture potential is applied toward a separate capture chamber containing an oligonucleotide affinity capture matrix. Sanger fragments extended beyond the 12th base in the sequence are captured and purified from buffer and unincorporated reagents from the sequencing reaction. Extension fragments are then released at separation temperature and injected electrokinetically into a 30-cm long capillary, where electrophoresis at 167 V/cm produces a high-quality electropherogram (read lengths up to 556 at 99% accuracy) within 35 minutes [134].

Beyond establishing the first fully-integrated device for thermal cycling, purification, and separation of sequencing samples, this device also represents the synergistic potential of a number of novel microfabricated structures developed and characterized independently since the first generation device. A notable improvement over the first PCR-CE device, the thermal cycling reactor is not only 100 times smaller, but also etched directly onto the same glass substrate as the separation channels. The total integration of thermal cycling and separation on a single microchip eliminates the need for a fluidic junction, a feature that proved problematic in

early integrated devices. The new device also makes use of pneumatic valves and pumps capable of manipulating nanoliter-quantities of solution within chambers and channels. The design and operation of these valves are detailed in Figure 4.2. Each valve consists of a discontinuity in the channel, bridged by a displacement chamber patterned into a separate glass layer in the device. When the valve is at rest, this chamber, and therefore the two sides of the channel, are isolated fluidically by a thin PDMS membrane between the glass layers. The valve is opened pneumatically by applying a vacuum to the displacement chamber, causing the PDMS to deflect and allowing solution to flow freely through the junction. A single independent valve restricts sample flow, while three valves actuated in series act as a pump capable of manipulating volumes of solution tunable by valve dimensions [119]. On the integrated device, these microvalves serve not only to sequester extension reagents from the sequencing capillary prior to capture, but also to pump sample into the hold chamber during capture. Finally, the development of resistive temperature detectors (RTD's) patterned directly onto the device allow for precise monitoring of the temperature inside the reactor during thermal cycling.

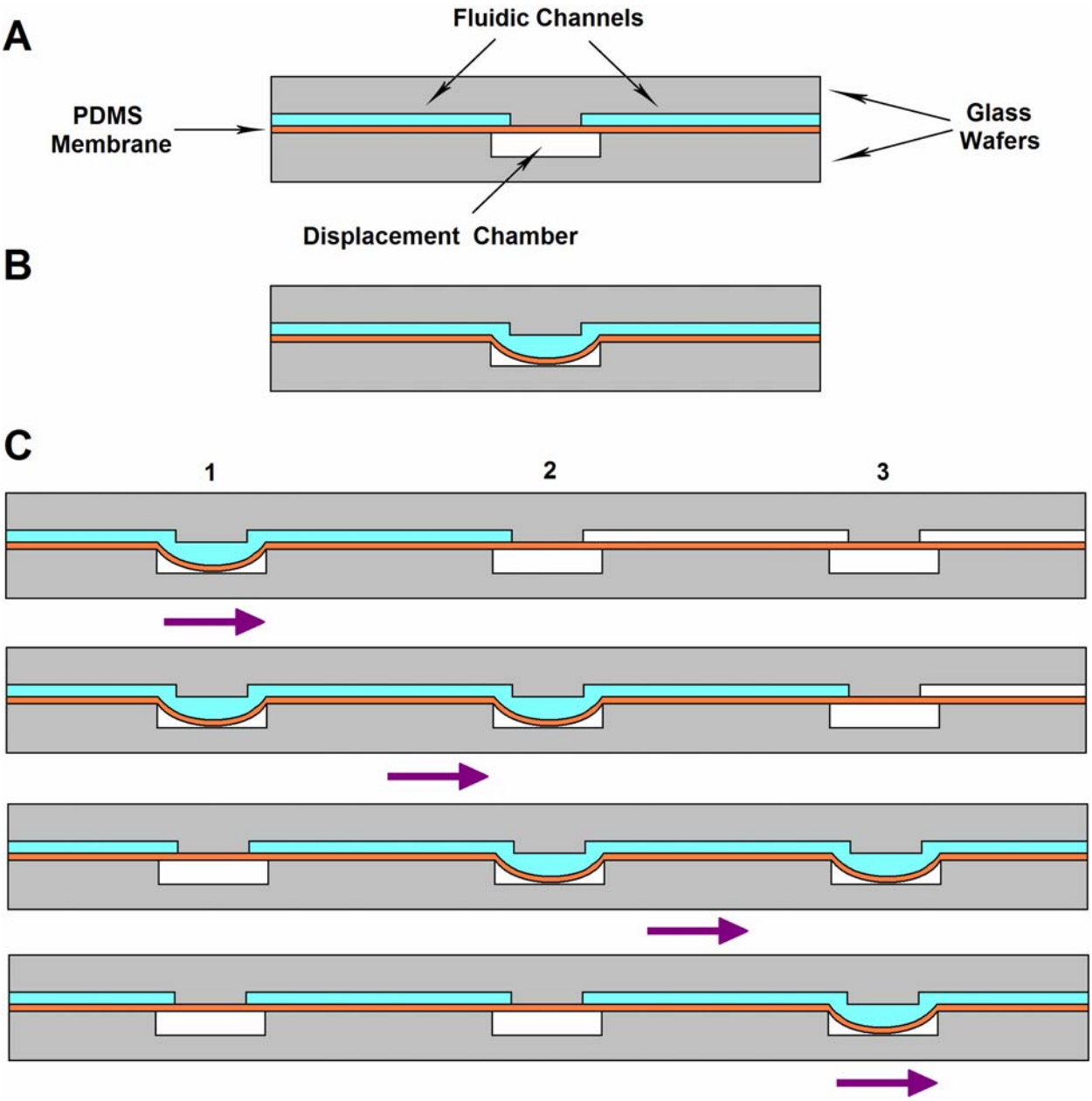



**Figure 4.2:** Design and operation of integrated microvalves for fluidic isolation and pumping. (*A*) The two sides of the channel are isolated by a PDMS membrane held in place by positive pressure in a displacement chamber below. (*B*) The valve is opened by applying a vacuum to the displacement chamber, forcing the membrane to deflect into the chamber and creating a connection between channels. (*C*) Serial activation of three valves in parallel allows for active pumping of fluid volumes directly tunable by valve dimensions.

While the integrated sequencing device vastly reduced both the volume of reagent required for Sanger extension (from 10 μL to 250 nL) and total analysis time (from four hours to under two), one final design facet rendered the device orders of magnitude less efficient than optimal. The electrokinetic injection cross has been a ubiquitous approach to sample introduction since nearly the dawn of microfluidics. While the cross injection allows for the precise tuning of resolution and signal intensity by providing a direct control for injection band width, the electrophoretic size bias renders the cross-injector highly dependent on timing. In the case of the integrated sequencing device, this injection timing was sensitive to less than one second within an overall 4-second injection. Moreover, the injection cross is immensely inefficient, with more than 99% of labeled products retained in the sample and waste reservoirs during analysis. The prospect of improving the sample introduction scheme to yield quantitative injection of purified sample represents one of the final steps in realizing the ultimate efficiency of microchip-based DNA analysis.

Realizing that properly optimized affinity capture produces a tightly defined band of sample within the capture matrix, Blazej *et al*. exploited this technology with the introduction of a prototype device for quantitative inline injection and Sanger sequencing in 2007. In contrast to previous oligonucleotide-based capture structures, the inline device employs an intervening plug of affinity matrix directly between the cathode and anode in the separation capillary. Unpurified sequencing reaction products generated off-chip are captured and washed electrophoretically through perpendicular sample and waste arms, leaving a ~150-μm band of purified extension fragments. After thermal release, 100% of these fragments are directly injected into the capillary, followed by separation and quantitative detection of all fluorescently labeled fragments. In initial tests, the inline injection enabled accurate sequencing from 100 amol of template, although continuous read length suffered due to the unique properties of the capture matrix employed. Most problematic among these properties is the net negative charge imparted to the matrix as a result of the covalently bound capture oligonucleotide. This negative charge causes the capture matrix, and therefore the separation matrix, to migrate toward the anode at separation potentials, thereby altering the resolution of the separation over time. Additional matrix-based concerns include differential thermal expansion and fragment mobilities within adjacent matrices [135].

## 4.3: A Novel Microdevice for High-Throughput, Integrated PRS

In the effort to integrate separation with on-chip sample labeling steps, PRS presents unique challenges owing to its dual-template pooling scheme. In contrast to contemporary automated Sanger sequencing, in which extension multiplexity is limited only by template identity, the primer-based labeling chemistry of PRS requires the primer to identify both the template and terminating base. Consequently, the Sanger extension step of PRS cannot be multiplexed in the same manner as conventional dye-terminator sequencing, in which fragments for all four terminating bases can be generated in a single reaction.

A proposed scheme for partially multiplexing the PRS extension reaction is presented in Figure 4.3. This process makes use of specialized PCR primers which contain tails for both sequencing primer recognition and oligonucleotide capture. One of two orthogonal universal sequencing primers (for example, the -40 and -28 M13 sequences employed in early PRS

studies) is appended to the far 5' end of each PCR primer. These sequences are unique to the template source, in this case normal or tumor, and allow for simultaneous labeling of all like extension fragments for both templates in a single reactor. Immediately 3' to this sequencing primer region, a common capture sequence roughly 20 nt in length is inserted, allowing for purification of all fragments extended beyond ~10 nt but excluding all unextended dye primers. The PCR priming sequence is attached at the far 3' end, producing a PCR primer ~60 nt in length. If both forward and reverse sequences are desired, reverse PCR primers may be designed similarly using two additional orthogonal sequencing primers, although a common capture sequence may be used.

Following PCR and purification, products from different templates are pooled and combined with sequencing primers matching the two universal sequences, each bearing a different fluorescent label. Buffer, dNTP's, and a single ddNTP are also added prior to thermal cycling. The resulting Sanger extension produces a series of fragments terminating with the same base but bearing a unique label identifying its source template. For four-color detection, termination products for another base may be pooled at this point. These fragments are generated using the same two universal sequences but must bear distinct labels. Extension products are then captured and purified on-chip by oligonucleotide hybridization, followed by thermal release, injection, and separation.

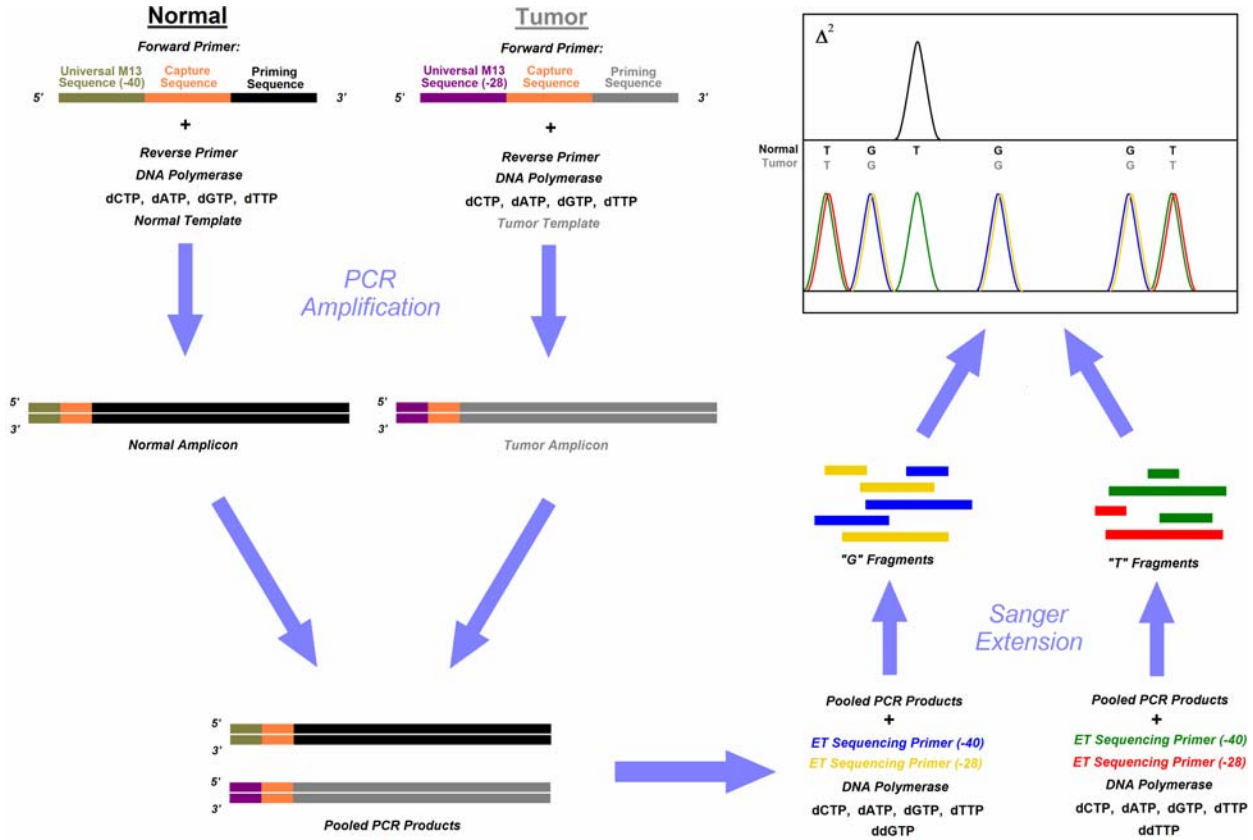



**Figure 4.3:** Labeling scheme for multiplexed PRS on an integrated platform. Tumor and normal templates are uniquely labeled with tailed forward PCR primers bearing recognition sequences for distinct sequencing primers, as well as a common capture probe. After PCR, amplicons are pooled and split into two reactors, where sequencing primers with distinct sequences and fluorescent labels generate like-terminating fragments for both templates in the same reaction. Extension fragments are then pooled, purified, and separated on-chip.

The integrated sequencing device of Blazej *et al*. provides a useful model system for the integration of PRS sample preparation, injection, and separation. Unfortunately, even under the maximally multiplexed scheme detailed above, PRS extension of a complete four-color sample requires two thermal cycling reactions. In addition, the wasteful and temperamental electrokinetic injection scheme employed by the integrated device fails to address the long-term concerns of robustness and efficiency. Both deficiencies are directly addressable by modest structural changes to the device, the former most straightforwardly by addition of a second reactor and the latter by introducing an inline capture and injection mechanism. The relatively poor performance of the original inline device, however, necessitates further optimization to the injection scheme before it can reliably be applied to sequencing-based assays requiring high accuracy and throughput.

It was originally hypothesized that the poor performance of the inline sequencing device was due primarily to loss of resolution resulting from affinity matrix migration during separation. This concern has been addressed in recent months through the development of techniques for on-chip photopolymerization of acrylamide matrices. In conventional polymerization, chain extension progresses by means of a radical reaction initiated by an agent such as ammonium

persulfate, which undergoes spontaneous radical decomposition in solution [126]. In a photopolymerization scheme, this agent is replaced by an otherwise stable compound that generates radicals upon irradiation, typically in the ultraviolet. The matrix is formed directly in the channel by sectioning off a selected region with an opaque photomask. A pre-polymer solution containing acrylamide monomer, buffer, and photoinitiator is flowed into the entire channel, and the unmasked region is irradiated for several minutes while the reaction progresses. The mask is then removed and unpolymerized solution is evacuated, leaving a clearly defined plug of polymer strictly within the unmasked region of the channel. Photopolymerization has been used for years as an alternative to off-chip polymerization owing to its shorter polymerization time and simpler handling compared to acrylamide gels polymerized off-chip [136].

As an approach to preparing the affinity capture matrix used in sample purification, photopolymerization provides a number of distinct advantages over off-chip polymer synthesis. The use of a photomask to restrict the illuminated portion of the channel during polymerization allows for precise control over the location, size, and shape of the gel. Previous multi-gel systems have relied on channel geometry to define regions containing different matrices; however these devices have proven problematic in terms of matrix loading. Additionally, photopolymerization in coated capillaries allows the coating to incorporate into the matrix polymer, anchoring the matrix to the capillary walls. Cross-linking the affinity matrix with modest concentrations of bis-acrylamide applies this restrictive force throughout the cross-section of the matrix, eliminating the electromigration and expansion issues that plagued early integrated sequencing devices. In first-generation applications in the Mathies lab, photopolymerized capture matrices have achieved impressive success, first in the capture and injection of PCR fragments generated off-chip for forensic typing [137], followed by integration with upstream PCR amplification for pathogen detection [138].

Figure 4.4 presents a proposed design for an integrated microdevice capable of multiplexed Sanger extension for four-color PRS samples, inline purification of extension fragments, and high-throughput electrophoretic separation. The device features an array of 48 separation doublets, each equipped with twin 17-cm long capillaries bearing inline injection heads, isolated sample inlets, and structures for Sanger extension. Sample processing for each lane occurs in two 250-nL thermal cycling reactors, operated in parallel and coupled downstream by a common sample injection channel. Following thermal cycling, extension products are pooled in this common channel and directed into a hold chamber by twin microvalve pumping systems. Common pneumatic lines ensure precise activation of all pumps in parallel. Sample from the hold chamber is electrophoresed toward the waste reservoir through an inline capture matrix photopolymerized between sample and waste junctions in the separation capillary. The use of a denaturing separation matrix necessitates a long (5.2-mm) capture plug to prevent diffusion of urea into the capture region. After affinity capture and electrophoretic wash, the resulting capture band is thermally released and directly injected into the capillary for separation. With an overall operation scheme akin to that of the integrated sequencing chip, the entire process from sample introduction to data acquisition can be performed in under two hours. Unlike the existing integrated device, however, the compact capillary design readily lends itself to radial parallelization, with sufficient area for 96 lanes on a single 200-mm wafer. With each doublet capable of complete PRS analysis of a single PCR amplicon, this integrated array device can be used to screen 48 unique genetic loci for somatic variation in parallel.

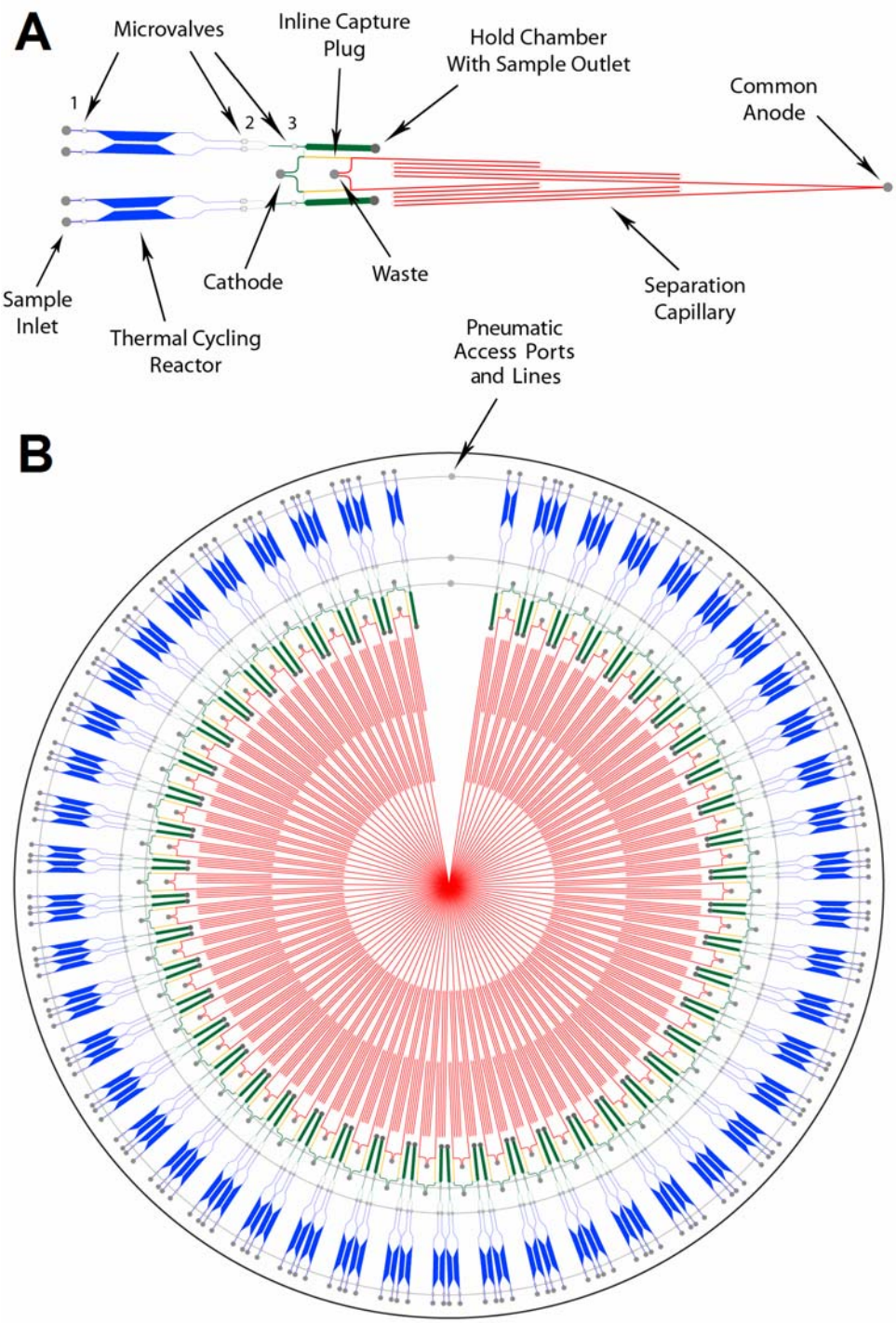



**Figure 4.4:** Array device for integrated PRS. (*A*) Each doublet consists of twin separation capillaries, each with a devoted pair of thermal cycling reactors for four-color PRS labeling of a single template. Microvalves 1 and 2 restrict the flow of extension cocktail during cycling, while a third allows for pumping products into the hold chamber prior to inline capture and injection. Common cathode and waste reservoirs reduce the complexity of access ports. (*B*) Forty-eight doublets are arrayed radially on a 200-mm diameter wafer, with pneumatic access for the three valve circuits located along a common radius.

In developing an optimized protocol for operating the integrated PRS device, a key concern will be introduction of extension cocktail into the thermal cycling reactors without cross-contamination via the common injection channel. While this coupling occurs considerably downstream of the reaction center, some degree of reagent mixing is to be expected. Closing all three microvalves during thermal cycling will limit diffusion from this intersection; however all valves must be open during introduction of the pre-extension solutions. Any degree of cross-contamination between reactors would produce false terminations throughout the trace: therefore, both reactors and channels must be filled at exactly the same rate to prevent leakage between reaction arms. In principle, this may be achieved by self-priming the pumping system from pools of sample over the inlet ports. Uneven priming may confound sample introduction, although the symmetric positive pressure provided by the two pumping systems should prevent leakage. Alternatively, the two reactor arms may be decoupled during sample introduction by adding a sample outlet arm immediately downstream of each reactor, with each sample outlet bearing a flow restriction valve (4). By closing valve 2 and opening valves 1 and 4, sample may be introduced into each reactor independently by simple capillary action without the risk of cross-contamination.

A second concern in operating the integrated PRS device arises from the differential mobilities of extension fragments between capture and separation matrices. The use of commercial separation matrices designed for long-read sequencing provides a rigorous control over both the quality and consistency of matrix properties; however these properties differ greatly from the photopolymerized capture matrix formed on-chip. As a result, the mobilities of the extension fragments released from the capture matrix will likely differ between the two gels. Optimally, this mobility differential should be biased toward the capture matrix, with relatively low mobility in the separation matrix giving rise to sample stacking during injection. As relative mobility depends on the effective pore size of each matrix, the properties of the capture matrix must be tuned to permit greater mobility during injection without inhibiting capture efficiency. Acrylamide concentration and degree of cross-linking will be key variables in this endeavor. Ionic strength stacking effects may also be exploited at this stage to decrease the effective band size injected into the separation matrix.

The device detailed above represents complete integration of the latter half of the PRS sample processing procedure. Additional improvements in the efficiency of sample preparation may be afforded by further integration of PCR amplification prior to Sanger extension. Quantitative recovery and purification of PCR products also affords the advantage of tuning the absolute quantities of secondary template subsequently delivered to the extension reaction, eliminating in all six hours of purification, quantitation, and dilution. While possible in principle, tandem PCR-Capture-PRS-Capture-CE presents fundamental design challenges. Aside from the increased structural complexity arising from additional reaction chambers and channels, the combination of reagents for both PCR and Sanger extension within the same fluidic circuit presents the potential for reagent interaction and incompatibility. These effects may be lessened through careful design of microvalve and channel geometries such that reagents from upstream processing steps do not contaminate downstream events. The use of photopolymerization to tune precisely the size and shape of capture matrices for both PCR and Sanger products will be of particular importance in this endeavor.

## 4.4: Single-Cell PRS

Among the greatest strengths of PRS is its ability to detect and quantitate multiple genetic populations within a single sample. This feature is especially advantageous in the context of mtDNA analysis, in which the potential for heteroplasmy is encountered. Indeed, the results of decades of cancer research on mitochondrial genetics, combined with those detailed here, indicate that heteroplasmic drift comprises a significant, if not majority, proportion of cancer-associated variation. In the context of these experiments, however, the use of the term heteroplasmy is highly assumptive. Conventional assays for mutation detection require template quantities on the order of 5 ng total DNA, or approximately 1000 cells. Therefore, sequence information derived from a particular sample represents the average genotype of all cells constituting the original biopsy. If indeed these cells are genetically identical, then multiple basecalls at a particular position indicate authentic heteroplasmy. On this macroscale, however, even PRS cannot distinguish between intracellular heteroplasmy and intercellular heterogeneity. Each provides a potentially useful approach for tumor analysis: sample heterogeneity may indicate the extent of tumor invasion in surrounding tissue, while a heteroplasmy drift may provide insight into the molecular causes of tumor progression. It is the ability to distinguish between these unique pathological questions that presents the current challenge.

A major hurdle to single-cell sequencing on-chip was overcome recently with the development of a microfabricated platform for emulsion bead PCR. In 2008, Kumaresan *et al*. introduced the microfabricated droplet generator (μDG), a device designed to produce monodisperse, nanoliter-volume droplets of PCR cocktail suspended in carrier oil for discrete amplification [139]. The μDG consists of a microfabricated nozzle for controlled injection of a PCR reagent stream into a sheath oil flow. The template molecules (or cells) are introduced into the PCR mix in a statistically dilute concentration so that a single template is present in an emulsion reaction droplet. Microbeads bearing the appropriate reverse primer are introduced into the mixture in a similar fashion. PCR mix is directed through the nozzle by an integrated pump consisting of three microvalves in series. This pump produces a pulsatile flow that facilitates bead movement in the microchannel and aids in droplet formation at the nozzle, providing a perfect 1:1 correspondence between pumping frequency and droplet formation frequency. Thus, the droplet volume is controlled precisely by the volume of the pump valves. After droplet generation, the emulsion is collected and thermally cycled. Successful amplification occurs in individual droplets containing both a bead and a template molecule, producing a fluorescent bead through the incorporation of a fluorescently labeled forward primer. These beads can then be analyzed or sorted in a fluorescence activated cell sorter. The overall procedure is illustrated schematically in Figure 4.5.

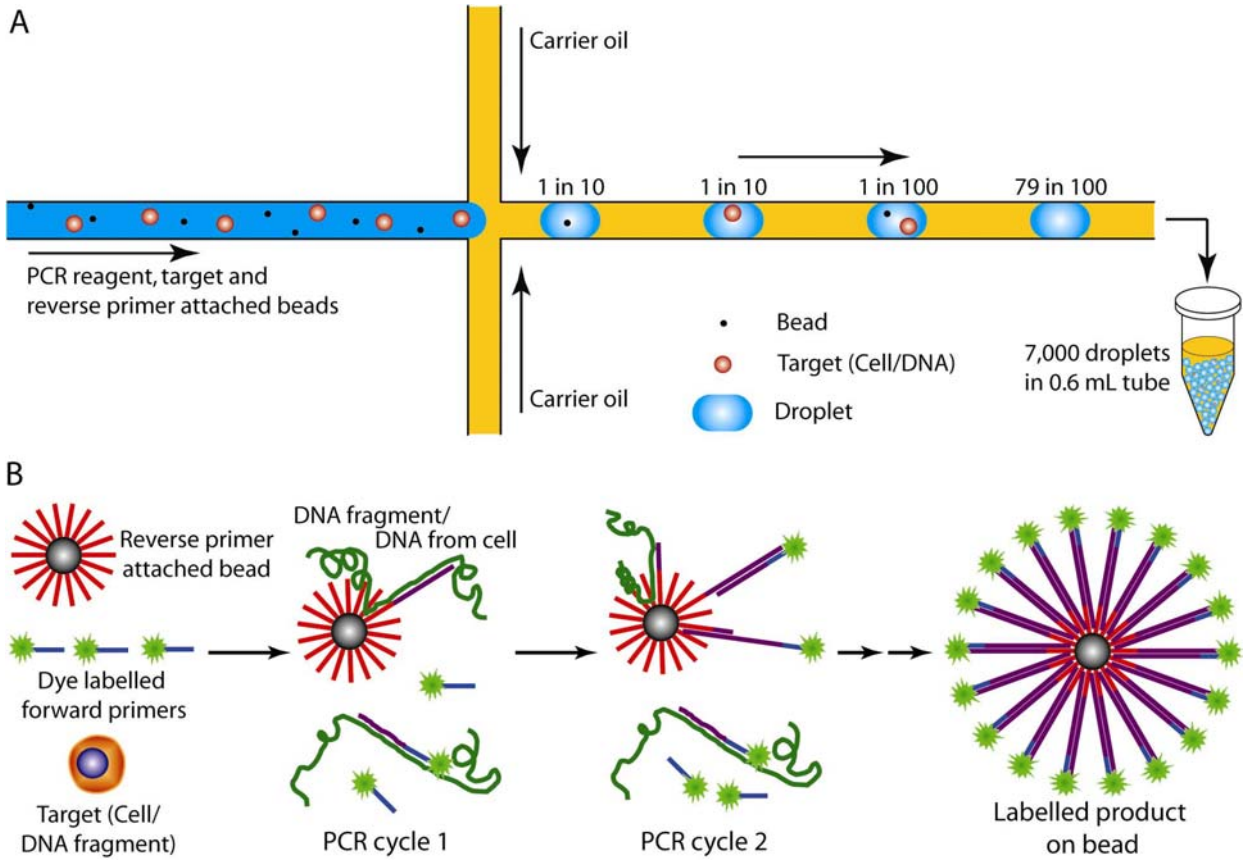



**Figure 4.5:** Emulsion bead PCR analysis method. *(A)* Statistically dilute target DNA/cells and beads are mixed with PCR reagent (blue) and pumped through the µDG to produce monodisperse droplets in carrier oil (yellow) at the cross-injector. These droplets are routed into a tube for thermal cycling. *(B)* Each functional droplet consists of a primer-labeled bead, a single template, and PCR buffer. Subsequent steps of PCR generate dye-labeled product on the bead surface.

Under optimized conditions, the µDG is capable of generating 150 amol of PCR product on a single bead, all of which is derived from a single DNA template. Using whole cells as PCR template for mtDNA analysis provides a ~100-fold increase in template concentration, although bead saturation is also a limiting factor in determining PCR yield. Nonetheless, the demonstrated yield of 150 amol is well beyond the 100 amol of template required for successful sequencing on the inline device, thus establishing the feasibility of single-cell comparative sequencing using a single bead from the µDG as extension template. Ultimately, direct coupling of the µDG with integrated PCR cycling capabilities, as well as the downstream Sanger extension steps proposed above, presents the potential to eliminate manual sample handling and decrease overall analysis time to mere hours.

## 4.5 Conclusions: The Future of Comparative Sequencing in Cancer Research

The use of short-range mutations, both germline and somatic, in cancer detection continues, inspired by an ever-growing body of scientific knowledge regarding genetic integrity in cancer. In 2008 over 1,600 scholarly publications were released on the highly variable p53 gene alone, with over 1,800 in 2006. Somatic variants in the mtDNA, while not universally established as causative, predictive, or reliably indicative, continue to garner strong interest within the field of cancer research owing to the genome's small size, abundance, and direct connection to pathways fundamentally linked to cancer progression. Despite the relatively uncertain relevance of many of these individual mutations, both nuclear and mitochondrial, the power of single-base variants to effect a drastic change in phenotype has been well established. Given the fundamentally genetic nature of cancer, as well as the growing understanding in the scientific community regarding the genetic indicators of predisposition, prognosis, and treatment response, DNA mutation analysis will remain at the forefront of cancer research, and ultimately treatment, as the specific biomolecular influences on cancer progression come into greater focus.

As a cancer detection tool, PRS affords the benefits of accuracy, sensitivity, and quantitation, all achieved through the use of an internal sequencing control. In the clinical setting, these features hold the potential to exploit cancer's genetic origins to address issues of characterization and treatment without the use of costly and error-prone conventional sequencing. Additionally, the ability to detect somatic variation without prior knowledge of its specific location allows for the discovery of novel mutations in any gene for which appropriate priming sequences are known. With the completion of the human genome, this presents a nearly boundless library of potential biomarkers within both nuclear and mitochondrial genomes. In the clinical setting on a high-throughput array, a specific tumor may thus be screened in up to 48 unique genetic regions, providing an individualized molecular fingerprint that can be used to track tumor progression, gauge prognosis, and monitor response to therapy.

The novel microdevices advanced recently, combined with those proposed here, set the stage for more rapid, efficient, and carefully controlled mutation detection using comparative sequencing-based technologies. The reduction in reagent consumption, operational complexity, and manual sample manipulation afforded by these advancements will also result in a substantially less expensive assay, allowing for more widespread application in the clinical setting and the realization of a truly personalized approach to genetic diseases, both cancer and otherwise. The results of these chapters suggest that PRS holds the potential to play a major, and in fact revolutionary, role in this new age of medicine.

# Appendix A

## High-Throughput Genetic Profiling by On-Chip TRAC Analysis

-or-

## *Mito-Man's Ally in the Gene Expression World*

## A.1 Introduction

While comparative DNA sequencing directly probes the fundamental genetic causes of cancer, gene expression assays provide the link between the genetic code and the proteins that influence intracellular chemical dynamics. In this section I present preliminary experiments undertaken to explore the feasibility of performing sensitive, high-throughput expression analysis of cancer samples on microfabricated platforms. The TRAC assay, or Transcript Analysis with Aid of Affinity Capture, was developed in 2003 by Hans Söderlund at Valtion Teknillinen Tutkimuskeskus (VTT, the Technical Research Centre of Finland) and provides a simple, sensitive scheme for the quantitation of key genes implicated in cancer. The experiments detailed in the following paragraphs represent the beginning of collaborative work between VTT and the Mathies Lab on the development of technology for sensitive, high-throughput bioanalysis on microfabricated devices.

## A.2 The TRAC Assay

The overall scheme of the TRAC method is illustrated in Figure A.1. Target mRNA's are first hybridized with a pool of gene-specific detection probes, as well as a biotinylated oligo-dT capture probe directed toward the common poly-A tail on the mRNA (*A*). Alternatively RNA can be biotinylated or capture probes specific for each target mRNA can be used. The RNA-probe complexes are then captured on streptavidin-coated magnetic particles (*B*). Unbound sample material and nonspecifically bound detection probes are removed, and the target-specific probes are eluted and analyzed by capillary electrophoresis (*C*). Target mRNA's are identified by migration time and fluorophore identity, while peak intensities are quantitatively correlated with the starting amount of the target mRNA. TRAC-based analysis is advantageous in that crude cell lysates may be used directly as samples with minimal upstream sample processing. Additionally, the technique lends itself well to multiplexed, high-throughput analysis, with both hybridization and separation steps easily automated using magnetic bead particle processors and highly parallel genetic analysis instrumentation. Starting from sample collection, the entire TRAC assay can be carried out in few hours.

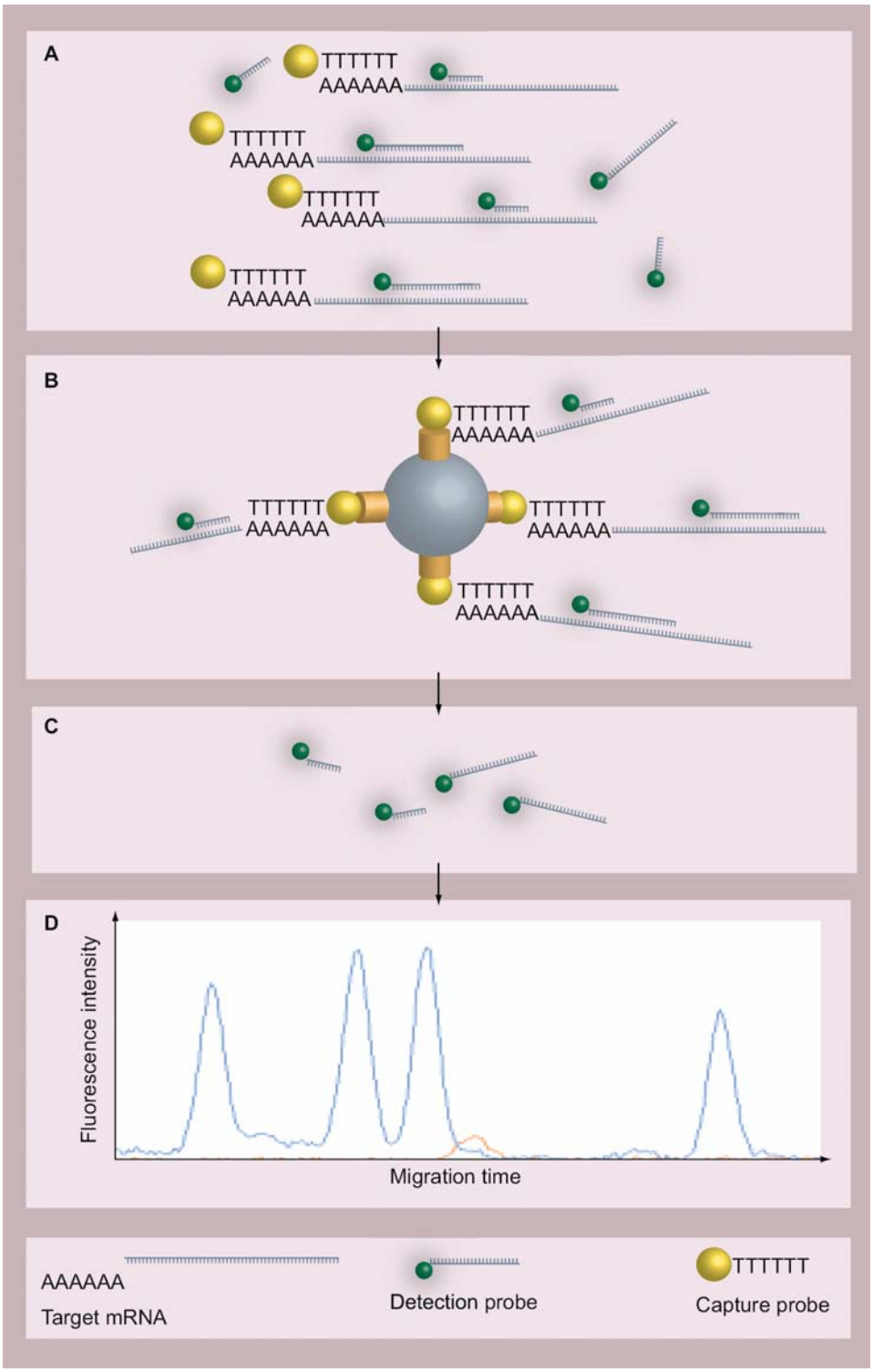



**Figure A.1:** Overall TRAC scheme (reproduced with permission from Rautio *et al*. [140]). (*A*) Total RNA from lysed cells are incubated with a biotinylated oligo-dT (25T) and a pool of fluorescent probes designed to bind quantitatively with specific mRNA targets within the lysate. (*B*) mRNA-probe hybrids are captured on streptavidin-coated magnetic beads and washed. (*C*) Probes are then dehybridized from their target mRNA's and (*D*) separated electrophoretically.

TRAC enables rapid and multiplex transcript analysis from large numbers of samples and has been used in a variety of applications, including quantitative detection of bacterial populations [141], rapid transcript analysis of microbial cultures [142], and monitoring of biotechnical processes in fungal organisms [143]. Using amplifiable DNA probes and PCR, TRAC can be used for highly sensitive transcript detection [144]. Recently TRAC has been used to analyze the expression profiles of selected gene markers from lysed colon cancer cells in a 96-well format [140].

Like TRAC, a wide variety of biological assays hold exciting promise in applications of genetic research, typing, and medical treatment. Unfortunately, these techniques are often prohibitively costly in terms of both money and time. The development of microfabricated devices for DNA extension, capture, and separation provides the foundation for rendering such large-scale genetic analysis feasible. With the recent emphasis on personalized healthcare, the field of microchip-based genetic analysis has grown rapidly in the past decade.

Of particular relevance to TRAC analysis is the recent advancement in affinity matrix-based oligonucleotide capture technology, whereby DNA hybridization is used to isolate and concentrate the desired analyte on-chip prior to electrophoretic analysis. The affinity matrix is prepared by introducing an acrydite-modified oligonucleotide complementary to a common sequence within the desired products into the reaction cocktail during acrylamide polymerization. The resulting polymer bears a covalently attached affinity probe capable of hybridizing to the analyte of interest. By electrophoretically driving unpurified PCR or Sanger sequencing products through the capture matrix, the target molecules hybridize quantitatively to the matrix and become immobilized. In addition to providing a method for efficient on-chip preconcentration, this affinity matrix capture technique also purifies the product, separating the probes of interest from PCR buffer and unincorporated nucleotides. The bound products are then released thermally prior to electrophoretic separation. This technology has been applied to a variety of applications, including capture of unpurified reaction products prepared off-chip [117] and integrated Sanger sequencing-capture-CE [134]. In particular, the integrated on-chip PCR-capture-CE device recently developed by Toriello *et al*. [145] has been of interest for this collaboration.

Integrated oligonucleotide-based affinity capture provides the potential for simplifying conventional TRAC protocols, essentially replacing the bead-based capture scheme currently in use. The elimination of this manual step will also improve the quantitative recovery of sample, increasing both sensitivity and accuracy. A microdevice capable of capturing probe-labeled target mRNA directly from the hybridization cocktail, followed by quantitative release and high-resolution separation of probes, is the ultimate goal of this endeavor. The collaborative work detailed here, however, was aimed primarily at exploring microchip-based TRAC analysis in a variety of microchip formats and characterizing poly-A capture properties. Subsequent work by Nadia Del Bueno has continued in this direction, improving the robustness of the technique and advancing the goal of more complete integration.

## A.3 Sensitivity Tests

Initial microchip-based TRAC experiments were directed toward characterizing the resolving capacity and sensitivity of the μCAE device and maximizing both for optimal analysis of TRAC probes. The separation of such small, closely-spaced fragments presented a number of unique challenges in these preliminary tests. Although the μCAE device had previously demonstrated impressive performance under sequencing conditions, achieving the single-base resolution necessary for consistent 400-500 base read lengths, these conditions performed optimally in the 100-500 nt range. Therefore, these parameters were ill-suited for separating fragments in the 20-50 nt range, which require low-field separations through high concentrations of polyacrylamide to facilitate more efficient sieving [146].

To address the issue of resolution, a custom separation matrix was synthesized in-house by Dr. Thomas Chiesl. The high LPA concentrations required for effective TRAC separation necessitated the preferential synthesis of low-MW polymers to preserve sufficiently low viscosity for gel manipulation. Low-concentration (3%) acrylamide was polymerized under chain transfer conditions (3% isopropanol), facilitating early termination and generation of LPA strands in the 300-600 kDa range. Following dialysis, the polymer was freeze-dried for storage and, prior to use, resuspended to a final maximum concentration of 7.5% in 7 M urea.

| Probe ID | Gene Function | Probe Length | Sequence |
|---|---|---|---|
| PLAUR123 | Urokinase Receptor | 30 | CTAACCCACACACAACCTCGGTAAGGCTGG |
| MMP9 | Matrix Metallopeptidase | 32 | CCGCGATGGCGTCGAAGATGTTCACGTTGCAG |
| MMP7 | Matrix Metallopeptidase | 34 | ATCTCCATTTCCATAGGTTGGATACATCACTGCA |
| p53 | Tumor Suppressor | 38 | GGGTGTGGGATGGGGTGAGATTTCCTTTTAGGTACTAA |
| MMP2 | Matrix Metallopeptidase | 40 | TCTGGAATCTTAAATTACCAGGTAGGAGTGAGAATGCTGA |
| MSH2 | Mismatch Repair | 42 | CTTTCTTCACCTGATAAAGCATAGTTAAGGTCTCTTCAGTGG |
| PRSS123 | Serine Protease | 44 | CCTCAAGGAAGCCCACACAGAACATGTTGTTGGTAATCTTTCCA |
| SPINK1 | Serine Peptidase Inhibitor | 48 | CCTGATGGGATTTCAAAACCTTGGTTCTCAGCAAGGCCCAGATTTTTG |
| uPA | Urokinase | 50 | ACTAGGCTAAAAGGAAGGGATAACTGGCCAAGAAAGGGACATCTATGTGA |

**Table A.1:** Selection and design of TRAC probes for colon cancer screening

For the preliminary evaluation of on-chip probe separation using the specialized TRAC polymer, an equimolar pool of nine TRAC probes with a minimum spacing of 2 nt was employed (Table A.1). This pool was run in dilution series (500 nM - 1 pM) on the μCAE device and analyzed for sensitivity and resolution. Samples were analyzed by injection at 30 V/cm for 270 seconds, followed by a 15-minute separation at 110 V/cm. Both injection and separation were carried out at 67º C. As indicated in Figure A.2 (*A*), satisfactory resolution was achieved under these conditions for a probe spacing of 2 nt. However, using this simple cross-injection approach, the minimum detectable probe concentration was 10 pM. Given that the TRAC assay was developed to perform optimally with a detection system capable of 0.1-1 pM sensitivity, it was necessary to employ a number of modifications to increase signal-to-noise, including enhancements to the optical detection system, careful data integration window selection, and post-data-collection noise filtration. A marked increase in raw signal was achieved operationally through the use of a stacked injection, consisting of a standard cross-injection for 270 seconds (at 30 V/cm), followed by direct injection of analyte from the sample

reservoir toward the anode under separation potentials for five seconds. This injection scheme increases signal by nearly ten-fold; however the resulting increase in injection band width necessitated a 4-nt probe spacing for satisfactory resolution and peak quantitation (Figure A.2(*B*)). For the purposes of these initial experiments, it was concluded that the use of this five-probe pool was acceptable. Some degree of initial success was also achieved in releasing a simple oligo-dA-FAM label from oligo-dT-labeled beads directly on-chip (release at separation temperature of 67º C, Figure A.2 (*C*)); however probe release from bead-bound mRNA could not be achieved under these conditions.

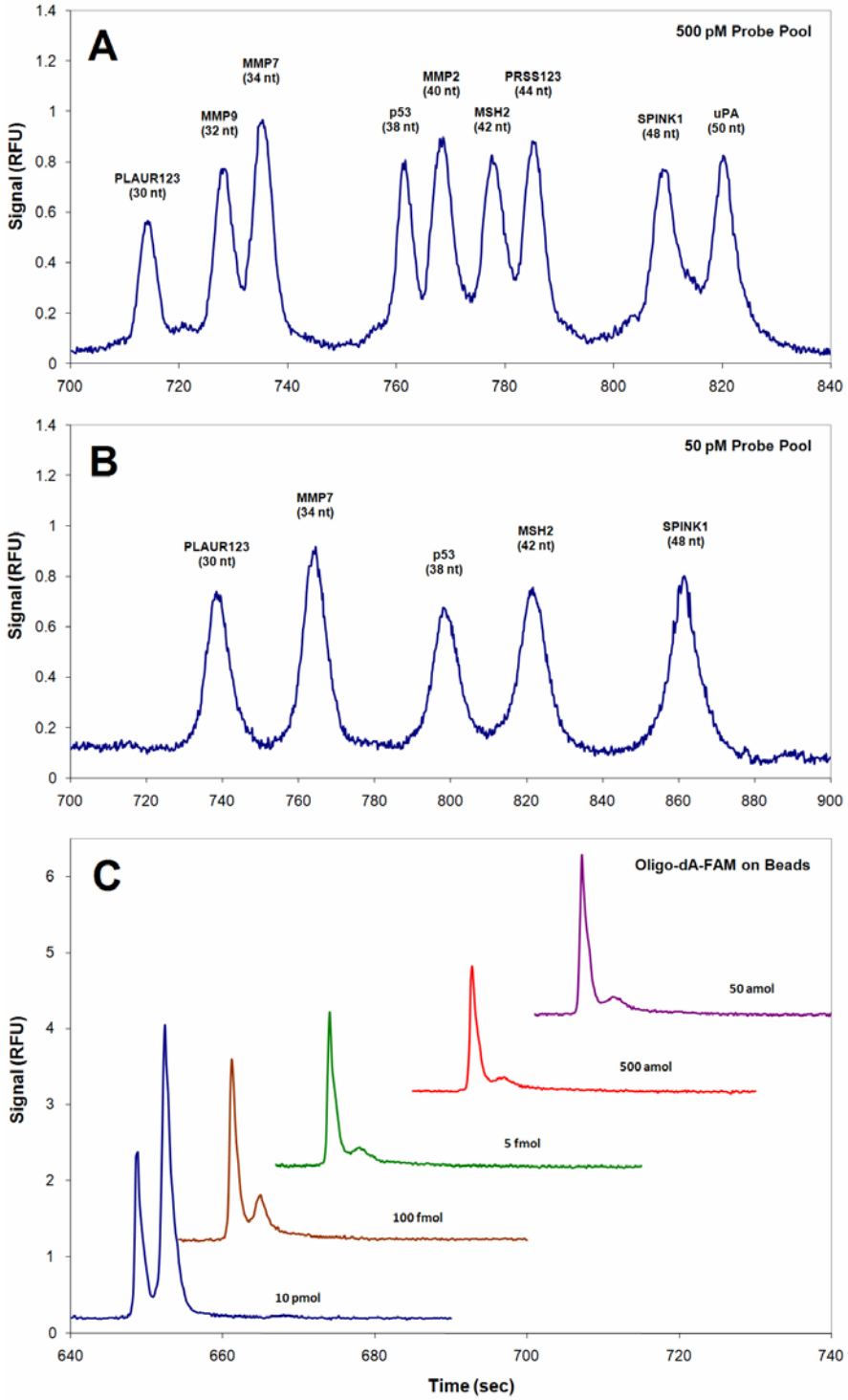



**Figure A.2:** Initial TRAC experiments on the µCAE bioprocessor. (*A*) Specialized polymers and separation conditions were successful in achieving 2-nt resolution using a standard cross injection. (*B*) Injection stacking yields increased sensitivity at the cost of resolution. (*C*) On-chip release of oligo-dA-FAM directly from bead-bound oligo-dT. Quantities listed correspond to starting amounts of oligo-dA-FAM hybridized in constant concentrations of biotinylated oligo-dT and streptavidin-coated beads.

Using the sparser pool of five TRAC probes in conjunction with the operational modifications listed above, a detection limit approaching 1 pM was achieved on the array device. Signal for the 1.25 pM probe pool was distinguishable from background in 50% of individual traces (Figure A.3(*A*)), but not consistently or within uncertainty (*B*). In previous TRAC experiments on the ABI 3100, limits of detection solidly between 0.1 pM and 1 pM were observed, although parallel analysis of TRAC samples prepared in Berkeley using the ABI 3730 achieved a detection limit similar to that of the μCAE device. Nonetheless, it was decided to test the μCAE system using TRAC performed on a series of *in vitro* transcribed mRNA (IVT RNA) pooled in varying quantities (0.1-446 fmol per reaction). The results of these analyses are shown in Figure A.3 (*C*-*D*) and indicate a minimum starting mRNA concentration of 300-400 amol is necessary for effective hybridization and detection. Reproducibility over multiple replicates was acceptable for quantitative TRAC analysis, with an average 15% coefficient of variation; however, the dynamic range of the system was problematic, with starting mRNA quantities of 5 fmol or higher producing saturating peaks.

TRAC samples with IVT RNA templates were prepared by pooling five different full-length IVT RNA's in different quantities for multiplex hybridization. These IVT RNA's were analogues for transcripts of the following genes implicated in colon cancer: PLAUR (1743 nt), MMP7 (1127 nt), p53 (2628 nt), MSH2 (3145 nt), and SPINK (438 nt). Hybridization was performed for 2 hours at 60º C in a 50 μL solution of 4x saline sodium citrate buffer (SSC), 1x Denhardt solution, 0.1% sodium dodecyl sulfate (SDS), 66 nM oligo dT-biotin, and 20 nM detection probes. To each hybridized sample, 50 μg of MyOne DynaBeads were added, and the mixture was incubated at room temperature for 30 minutes. Beads were washed three times with 0.1x SSC/0.1% SDS solution and collected magnetically. Following resuspension in 5 μL of 100% Hi-Di formamide, probes were eluted with a 30-minute incubation at 37º C, after which the beads were collected and discarded.

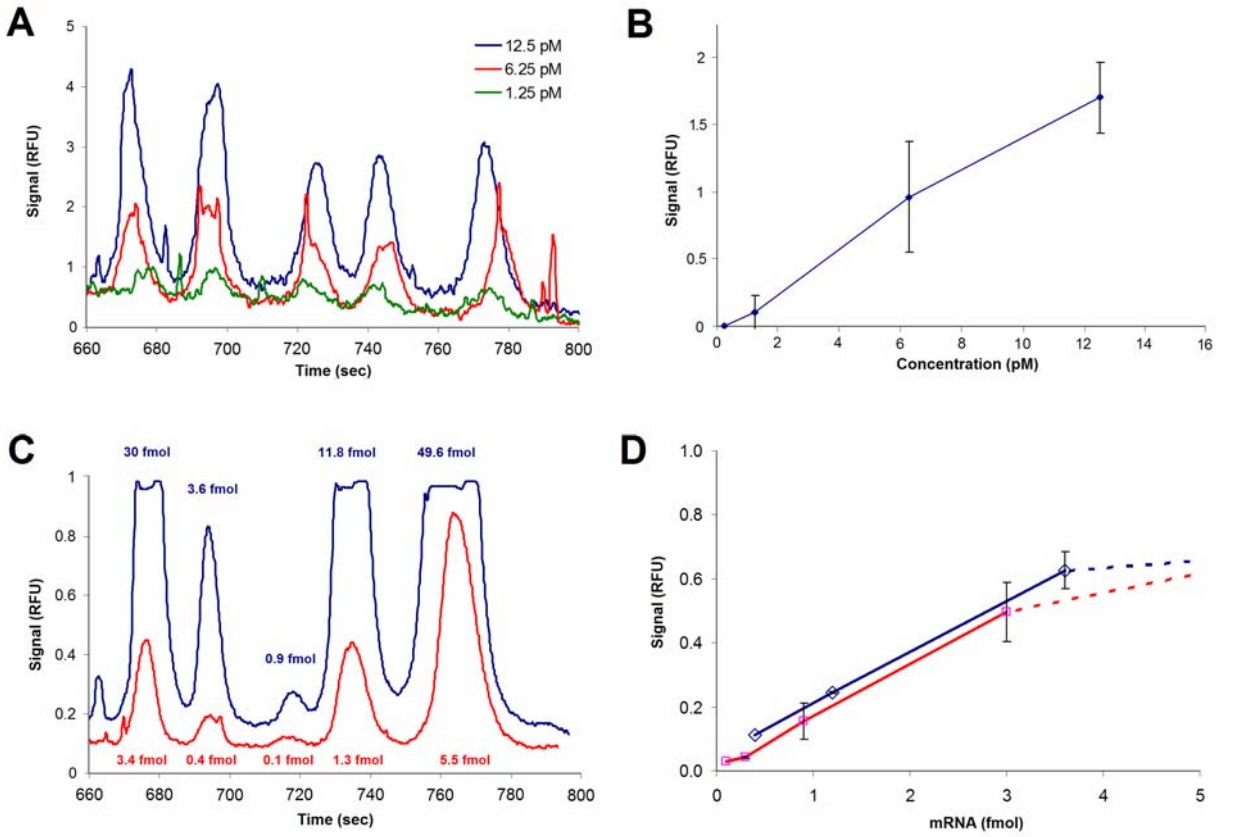



**Figure A.3:** Sensitivity experiments on the μCAE bioprocessor. (*A*) Using a 5-second stacked injection, a differentiable signal for the 1.25 pM probe pool is observed in ~50% of replicates, although this limit of detection is not achieved within uncertainty (*B*). (*C*) Full TRAC analysis of IVT RNA's spiked in varying quantities in the hybridization cocktail. Saturating signals are observed for RNA quantities above 5.5 fmol, although (*D*) the linear range extends only to 3.6 fmol.

## A.4 Inline Capture

Given that only marginally acceptable sensitivity was achieved on the array device and that this sensitivity was achieved at the cost of resolution (and consequently, the elimination of half of the original probe pool), it was decided that inline capture and preconcentration of mRNA probe hybrids offered the greatest promise in increasing both sensitivity and resolution to the levels of commercial-grade analytical platforms. For the preliminary purposes of characterizing the thermodynamics of oligo-dT capture and release, a small integrated microdevice designed for inline capture and separation of labeled, off-chip PCR products was employed. The design of this device is depicted in Figure A.4 [145]. The device features four parallel channels, each with devoted chambers for waste and capture. Upstream of the PCR reactors are PDMS pumps capable of manipulating 30 nL of sample solution per stroke. These pumps allow for the precise delivery of minute volumes of sample into the microchannel region, followed by quantitative capture and injection.

The operation of the capture-CE device is depicted in Figure A.4 (*a-f*). A capture matrix (denoted in yellow) is loaded through the cathode reservoir in each lane until the matrix front reaches the intersection with the separation capillary. The separation capillaries are then loaded with a separation matrix (denoted in red) via the common anode, and excess matrix is evacuated through the waste reservoir (*a*). Sample (green) is pipetted onto the sample inlet and pumped into the capillary system (*b*). As sample is pumped into the hold chamber (*c*), an electrical potential is applied to drive sample toward the capture chamber electrokinetically (*d*), forming a band of captured products at the interface of the capture matrix. This band is then washed with 1x TTE, leaving only labeled fragments immobilized by hybridization in a purified inline capture plug (*e*). Finally, the temperature of the device is raised and separation potentials are applied, releasing the sample into the separation capillary.

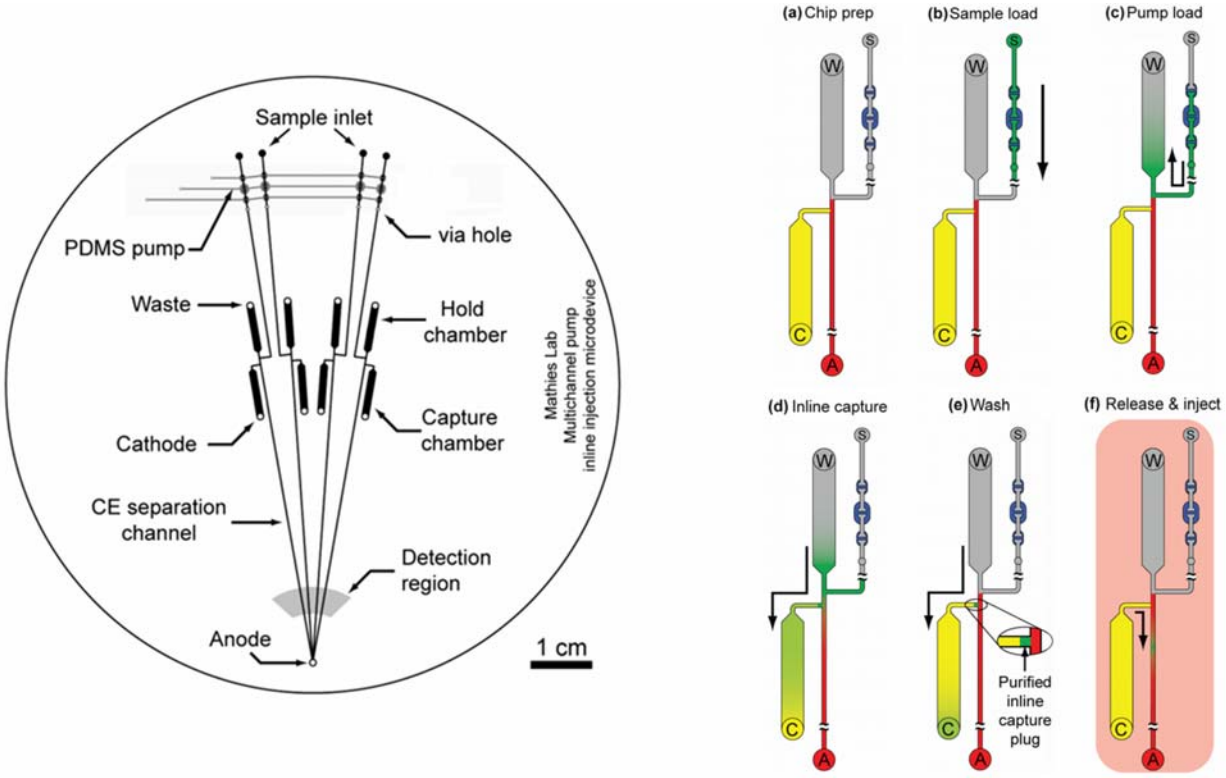



**Figure A.4:** Design and operation of the capture-CE chip. The microdevice features four 5-cm long separation capillaries, each equipped with PDMS micropumps for fluidic manipulation and chambers to facilitate sample injection. The chip is first loaded (*a*) with capture matrix (yellow) and separation matrix (red). Sample is pipetted onto the sample inlet (*b*) and pumped into the hold chamber (*c*). While pumping, an injection potential is applied from waste to cathode (*d*), followed by a wash from buffer at injection potentials, producing a narrow plug of sample at the capture matrix interface (*e*). The captured injection band is then released thermally prior to the application of the separation potential (*f*).

While the integrated capture-CE device provides a useful tool for optimizing the oligo-dT capture properties, which should be universal for all target mRNA-probe hybrids, its poor resolving capacity renders it only minimally useful for full TRAC analysis. In contrast to the 17-cm long channels on the μCAE device, the separation capillary is only 5 cm in length, limiting the opportunity for analyte to interact with the sieving matrix. Also problematic is the geometry of the capture region, as the capture plug must first complete a 90-degree turn upon entering the separation capillary. The distorted field lines in this turn result in zone elongation and peak tailing. Additionally, the proximity of the capture region to the separation column necessitates the use of a non-denaturing separation matrix, further limiting resolution by suppressing fragment size differentials. For these reasons a number of design modifications will be necessary if high-resolution separations of small fragments, and thus effective TRAC, are to be achieved. Nonetheless, the integrated capture-CE chip represents a useful model system for exploring capture characteristics and was therefore employed in these preliminary tests.

Representative TRAC samples, either oligo-dA-FAM or IVT RNA-probe hybrids, were captured using a 6.5% LPA capture matrix copolymerized with 20 μM of an acrydite-modified 25-nt oligo-dT capture probe ($T_M$ = 37.3º C, Integrated DNA Technologies). Polymerization was conducted in a 2-mL solution containing 6.5% (w/v) acrylamide monomer (Bio-Rad), 1x TTE, and 20 μM capture oligo. After sparging the solution with nitrogen at 4º C for one hour, 1.5 μL of 10% N,N,N',N'-tetramethylethylenediamine (TEMED, Bio-Rad) and 1.5 μL of 10% ammonium persulfate (Sigma) were added to initiate radical generation. Polymerization was allowed to continue at 4º C overnight. A non-functionalized sieving matrix of 5% LPA was synthesized in similar fashion.

Samples for on-chip TRAC were prepared by incubation of target IVT RNA's (either 890 pM SPINK or 530 pM MSH2) in a 5μL hybridization cocktail containing 4x SSC, 1x Denhardt solution, 0.1% SDS, and 10 nM of the appropriate TRAC probe. Hybridization was carried out at 60º C for one hour. Following capillary coating (1:1 solution of methanol:DEH-100 coating solution, The Gel Company), as well as loading of capture and separation matrices, sample was introduced into the capillaries either using the PDMS pumps (three strokes at 30 nL per stroke) or by directly pipetting 4 μL of sample into the waste reservoir. Inline capture was performed at 30º C by application of a 100 V/cm capture field from waste to cathode either during pumping or for 10 minutes after pipetting sample into the hold chamber. Following capture, the hold chamber was evacuated completely and refilled with 1x TTE buffer. The injection band was then washed electrophoretically for 4 minutes at capture temperature and potential. A second, 4.5-minute wash (170 V/cm) was performed from anode to cathode at 30º C. After washing, the temperature was raised to 80º C, and the separation potential (150 V/cm) was applied for 5 minutes.

The results of both oligo-dA-FAM and TRAC capture are depicted in Figure A.5. On-chip capture of oligo-dA-FAM was observed readily under both sample introduction schemes. In Figure A.5(*A*), 90 nL of a 500 pM oligo-dA-FAM solution (27 x $10^6$ molecules) was introduced via the PDMS pumps, yielding a signal-to-noise ratio of 150. A 5 pM solution was then tested, although direct injection of 4 μL (12 x $10^6$ molecules) from the waste reservoir was necessary to achieve a signal-to-noise ratio of 17 (Figure A.5(*B*)). The dramatic loss of signal after a mere two-fold reduction in total sample is likely the result of inefficient capture and excessive washing, as neither had been optimized for oligo-dT capture. More problematic however, are the results of the on-chip capture of TRAC samples from IVT RNA's. Figure

A.5(*C*) represents two replicates of intended capture for 4 μL of a monoplex target (890 pM SPINK-probe hybrid, 2.1 x $10^3$ total molecules). Aside from drastic inconsistency between lanes, the individual results are troubling: complete lane failure in one capillary and multiple peaks in the other. Interestingly, the peak profile in the blue trace is vaguely similar to the complete 9-probe pool in the original resolution tests. For a monoplex TRAC sample, however, a single, clean probe peak is expected. Trace quality, though somewhat more consistent between lanes, is equally poor with pump-loading of the same sample (90 nL, 48 x $10^6$ total molecules, Figure A.5(*D*)).

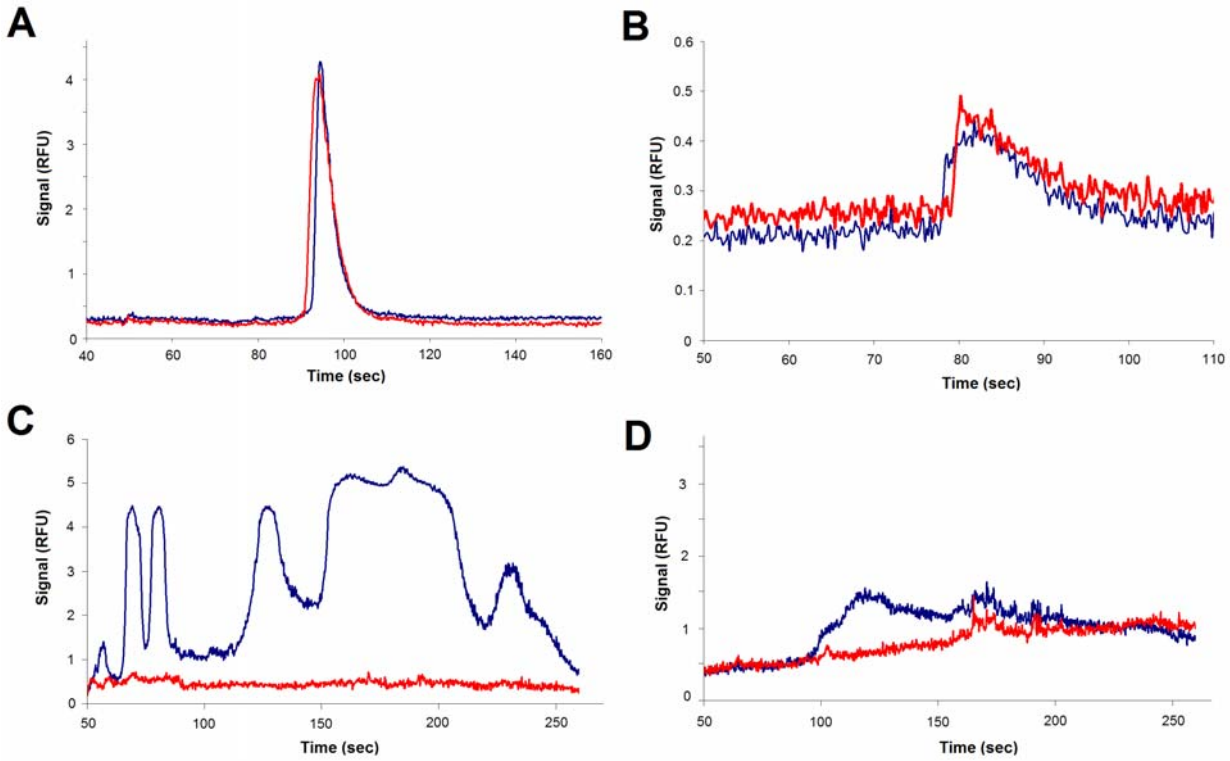



**Figure A.5:** On-chip capture of TRAC samples. Oligo-dA-FAM is readily captured and detected after injection of (*A*) 27 x $10^6$ molecules (S/N = 150) and (*B*) 12 x $10^6$ molecules (S/N = 17). Injection of RNA-probe hybrids for SPINK (438 nt) yield inconsistent, poor-quality traces for all injection quantities. Injection of 2.1 x $10^3$ molecules (*C*) and 48 x $10^6$ molecules (*D*) are shown.

Given the high concentrations and large size of the RNA-probe hybrid samples, it was hypothesized that gel overloading was inhibiting sample injection into the intended capture region. Because of the design of the capture-CE device, sample injected from the hold chamber must first pass through a small region of separation matrix before encountering the capture matrix. Gel clogging, if any, would tend to occur in this region of the main separation capillary. This suspicion was confirmed by altering the injection of a single IVT RNA TRAC sample (90 nL of 530 pM MSH2, 29 x $10^6$ molecules) in four parallel lanes of the capture-CE device. After standard capture and washing steps, two lanes were injected and separated under normal

conditions (150V/cm from cathode to anode), while the other two were electrophoresed directly from waste to anode (150 V/cm). As illustrated in Figure A.6(*A*), the standard injection and separation scheme yields a consistent, though very weak and broad, peak in both lanes. Conversely, a single, sharp peak corresponding to probes eluted from RNA hybrids tangled in the separation matrix interface is observed following direct injection from waste to anode (Figure A.6(*B*)). The second, lower-intensity peak in each replicate most likely represents the smaller fraction of RNA captured in the affinity matrix as the probes slowly leak into the capillary.

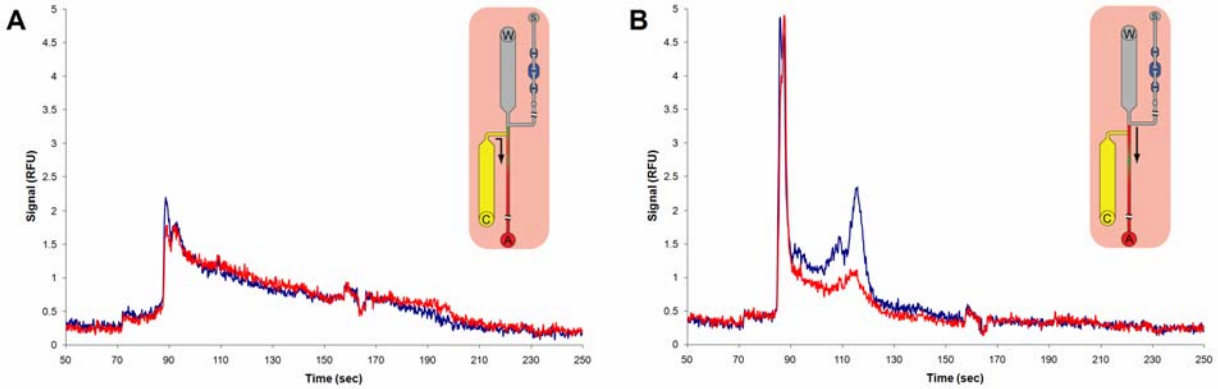



**Figure A.6:** Altered injection of IVT RNA samples. Parallel capture and wash of 29 x $10^6$ molecules of a single MSH2 (3145 nt) RNA-probe hybrid, followed by injection from (*A*) the intended capture region or (*B*) the hold reservoir, reveal gel clogging at the separation matrix interface.

The results of these preliminary capture tests, though not successful in the intended sense, were nonetheless enlightening. More optimization was necessary to ensure the most efficient capture, wash, and release of TRAC samples on-chip. Moreover, the positioning of the capture region at a right angle to the separation capillary was not acceptable for mRNA analysis. A truly in-line injection system with the capture matrix directly at the sample interface would ensure that all labeled RNA's would, by affinity or by tangling, capture in the intended region. A method for loading the capture-CE chip with affinity and separation matrices in this manner was developed on paper; however, isolating the capture matrix in the 1-mm offset between the hold chamber and the capture chamber proved to be difficult logistically. To remedy this problem, our attentions turned to an emerging technology in the Mathies Lab for precise on-chip tuning of matrix properties.

## A.5 Photopolymerization

The advantages of photopolymerization in the design and synthesis of affinity capture matrices are discussed in greater detail in Section 4.3. The exploitation of photopolymerization technology allows for in-chip polymerization of polyacrylamide capture matrices using photoinitiated radical reactions. Briefly, a short portion of the chip is sectioned off using an opaque mask. A solution containing acrylamide monomers, buffer, the acrydite-modified

capture oligonucleotide, and a light-sensitive molecule designed to initiate polymerization upon UV irradiation is flowed into the channels and allowed to equilibrate fluidically. The unmasked region is then irradiated for several minutes and the channels evacuated of unpolymerized solution. The result is a clearly defined plug of capture matrix covalently linked to the channel walls. This photopolymerization approach allows for precise control over the location, size, and shape of the capture plug, thus eliminating the difficulties previously encountered in loading capture matrices polymerized off-chip.

The endeavor to design a truly inline microdevice for the separation of TRAC samples was hastened by the concurrent development of a prototype chip for early development and characterization of photopolymerization protocols. The chip features a series of independent 6-cm separation capillaries, each bearing a generic twin-T injection head with a 0.1-cm offset between 0.5-cm sample and waste arms. As in the previous capture-CE device, the short capillary length limits the maximum resolution achievable by the device. Although the injection band does not directly contact the separation matrix, the proximity of separation region and the small size of the capture plug present non-trivial issues in the use of denaturing separation gels. Nonetheless, the prototype chip served as a useful platform for defining photopolymerization protocols and optimizing capture parameters for oligo-dT affinity matrices.

Following several months of effort by Nadia Del Bueno in the Mathies Lab, a partially optimized procedure for on-chip acrylamide polymerization was designed using riboflavin as the photoinitiator. All solutions were prepared at 5% polyacrylamide with 5% cross-linking using a stock 30% acrylamide solution (19:1 acrylamide:bis, Bio-Rad). Monomer solutions also contained 1x TTE and 20 μM acrydite-modified oligo-dT (25 nt), diluted to a final volume of 993 μL. The monomer solution was sparged under nitrogen for 11 minutes and combined with 6 μL of 0.1% (w/v) riboflavin (Sigma) and 1.25 μL 100% TEMED. A chrome photomask was placed on a microscope stand equipped with a model LH-M100CB-1 mercury arc lamp (Nikon), and the photopolymerization chip was carefully positioned above the mask, exposing all regions to contain photopolymer. After coating with DEH-100 (50% in methanol for 1 minute), monomer solution was pipetted into the anode and allowed to fill the entire channel system by capillary action. A thin layer of PDMS was applied to each fluidic port, and the solution was allowed to equilibrate for a total of ten minutes to minimize hydrodynamic flow. Polymerization was initiated by a 10-minute UV (330-380 nm) irradiation from below the photomask, after which excess monomer solution was evacuated and the plug inspected visually for structural integrity.

Following a thorough rinse of non-polymerized regions with filtered, deionized water, 1x TTE was loaded into the separation capillary, and 4 μL of sample were added to the pre-injection region (cathode and sample arms). A capture potential of 100 V/cm was applied from both cathode and sample reservoirs to the waste reservoir and was allowed to proceed for 10 minutes at 30º C. After capture, all buffers and solutions were replaced with fresh 1x TTE and a 5-minute electrophoretic wash was performed at capture potential and temperature. A non-denaturing 5% LPA separation matrix was then loaded manually into the separation region, and the pre-injection region was filled with 1x TTE buffer. The microdevice was raised to a separation temperature of 80º C, and a separation potential of 150 V/cm was applied for 5 minutes. The overall chip operation scheme is detailed in Figure A.7.

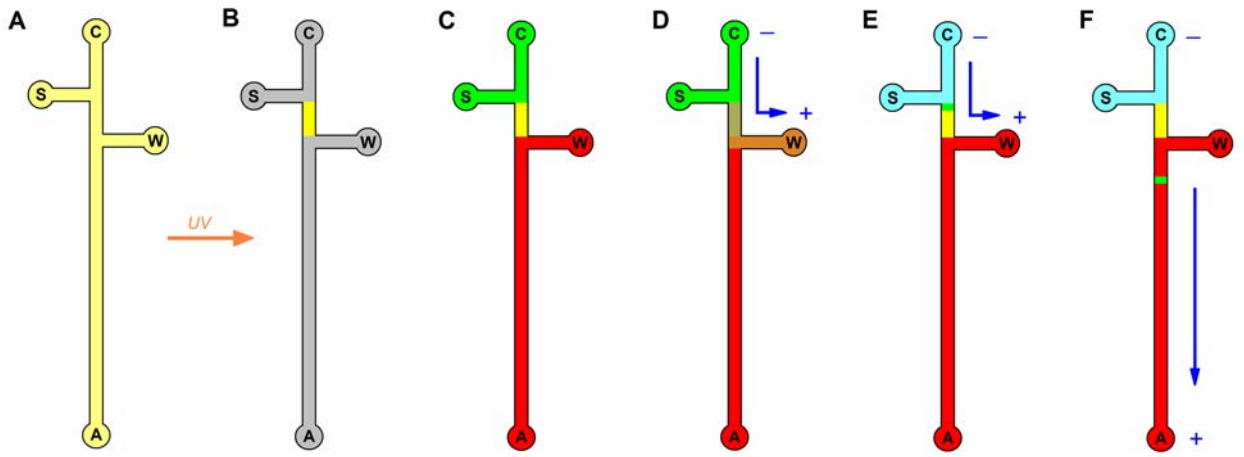



**Figure A.7:** TRAC in a photopolymerized capture gel. (*A*) The capillary is first filled with monomer solution. A specific region of the capillary is masked off and irradiated with UV, initiating acrylamide polymerization and (*B*) producing a clearly defined plug of polymer covalently attached to the capillary wall. (*C*) Separation matrix (red) and sample (green) are then loaded into the capillary. (*D*) Sample is electrophoresed through the capture matrix, leaving a purified inline sample band. (*E*) The band is then washed electrophoretically and (*F*) injected into the separation capillary under separation conditions.

A number of capture experiments were conducted in this manner to establish the ability to conduct oligonucleotide affinity capture in a photopolymerized gel and explore potential optimization parameters. The primary challenge in these early experiments was achieving consistent polymerization of a rigid, well-defined capture gel in multiple capillaries. This goal was hindered by a number of operational difficulties. Primarily, the dependence of the polymerization reaction on oxygen as a radical acceptor renders successful polymer formation highly dependent on sparging time, nitrogen flow rate, and initial amount of oxygen in solution. Careful control of these parameters is therefore vital. Additionally, hydrodynamic flow during irradiation can, depending on its severity, weaken an otherwise well-defined gel, initiate polymerization outside the intended region, or prevent polymerization altogether. This effect, due at least in part to the lopsided evaporation of monomer solution at the three fluidic ports near the cathode (relative to one at the anode), was not fully banished by the use of the PDMS barriers. Each of these factors affects not only lane-to-lane variability in photogel consistency, but also the homogeneity of individual plugs. The latter effect is exacerbated by the non-uniform radiation intensity across the illuminated region of the channel. A uniform gel is necessary for the formation of a well-defined sample plug and is consequently a factor in minimizing injection zone broadening.

Despite these difficulties, several successful capture-CE results were obtained in this exploratory phase, with two particularly promising results obtained from picomolar concentrations of oligo-dA-FAM shown in Figure A.8. The consistent generation of clean, single peaks, despite inconsistencies in signal intensity, establishes that oligonucleotide-based affinity capture is possible in polyacrylamide matrices photopolymerized on-chip and suggests that further optimization will yield quantitatively useful data, ultimately with TRAC samples.

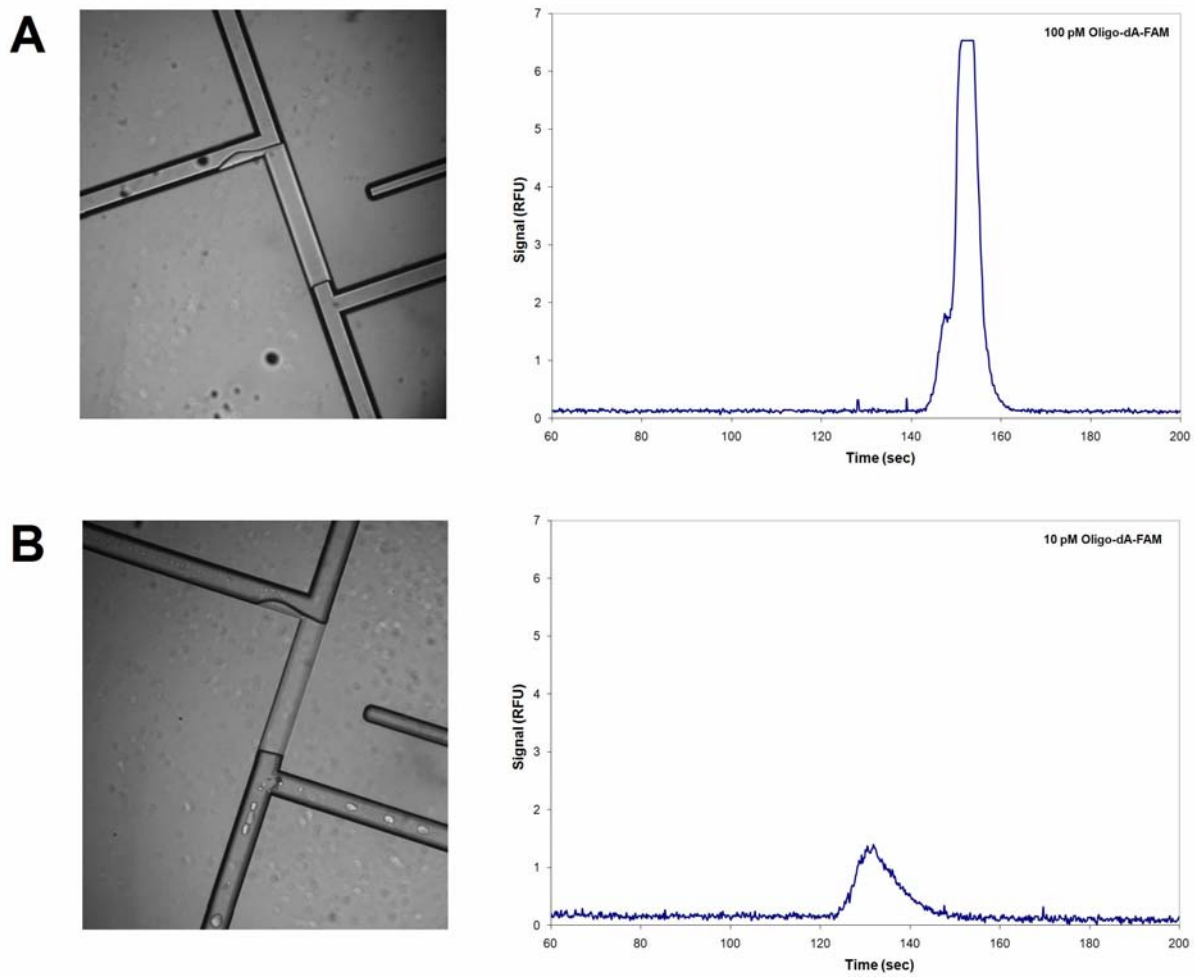



**Figure A.8:** On-chip capture of oligo-dA-FAM in photopolymerized capture matrices. (*A*) Saturated signal is observed after injection of 240 x $10^6$ molecules, while (*B*) 24 x $10^6$ molecules produce a clear peak with S/N = 47. Consistent formation of clearly defined capture plugs is observed in the leftmost panels.

At the conclusion of this collaboration, efforts were underway to optimize photopolymerization protocols, thereby improving the stability and consistency of capture plugs further. Continued efforts in subsequent collaborations between the Mathies Lab and VTT have yielded additional improvements in injection efficiencies, as well as novel microdevices capable of performing multiplexed TRAC analysis with on-chip probe capture and high-resolution separation. Published results are forthcoming.

## A.6 Conclusions

The marriage of TRAC-based gene expression analysis with microfabricated separation and detection devices holds great promise for enhancing the simplicity, sensitivity, and widespread applicability of gene expression analysis; however a number of adjustments must be made to existing technologies, both structurally and operationally, before these improvements can be realized. The primary challenge is developing an inline capture scheme that allows for consistent, quantitative recovery of TRAC probes as well as effective release and injection. The continued exploration of photopolymerized capture matrices will be crucial in this regard, as the results detailed here suggest further optimization is necessary to achieve acceptable reproducibility. Key variables in capture gel synthesis include monomer solution sparging and UV exposure times, as well as monomer and cross-linker concentrations. Capture temperature and field will also require attention if quantitative inline injection is to be achieved. In approaching these capture-based issues, a careful consideration of the unique mobility problem presented by the large mRNA-probe complexes will be vital.

Increasing resolution to the standard originally achieved on the 96-lane μCAE device will necessitate the development of longer-channel devices designed for small-fragment analysis. Field-amplified sample stacking may also be exploited to decrease injection band width prior to separation. The further development and optimization of specialized denaturing matrices for high-resolution separation of TRAC probes will provide an additional approach to improving separation quality, although the microchip design will have to be tailored to prevent urea from the separation matrix from inhibiting inline capture. Collaborative efforts toward realizing these goals are ongoing in the Mathies Lab and at VTT.

# References


1. *Cancer Facts & Figures: 2009*. Atlanta, GA: American Cancer Society, 2009.

2. *Estimates of Funding for Various Research, Condition, and Disease Categories (RCDC)*. 2009, National Institutes of Health.

3. Weinstein, I.B., and Case, K. The History of Cancer Research: Introducing an AACR Centennial Series. *Cancer Research*, 2008. **68**(17): p. 6861-6862.

4. *The History of Cancer*. Atlanta, GA: American Cancer Society, 2009.

5. Watson, J.D., and Crick, F.H.C. Molecular Structure of Nucleic Acids – A Structure for Deoxyribose Nucleic Acid. *Nature*, 1953. **171**(4356): p. 737-738.

6. Nirenberg, M. and Matthaei, J.H. Dependence of Cell-Free Protein Synthesis in E. coli Upon Naturally Occurring or Synthetic Polyribonucleotides. *Proceedings of the National Academy of Sciences of the United States of America*, 1961. **47**(10): p. 1588-1602.

7. Rosenberg, S.A. Immunotherapy and Gene Therapy of Cancer. *Cancer Research*, 1991. **51**(SUPPL): p. 5074s-5079s.

8. Blattman, J.N. and Greenberg, P.D. Cancer Immunotherapy: A Treatment for the Masses. *Science*, 2004. **305**(5681): p. 200-205.

9. Izquierdo, M. Short Interfering RNAs as a Tool for Cancer Gene Therapy. *Cancer Gene Therapy*, 2005. **12**(3): p. 217-227.

10. Curley, S.A., Cherukuri, P., Briggs, K., Patra, C.R., Upton, M., Dolson, E., and Mukherjee, P. Noninvasive Radiofrequency Field-Induced Hyperthermic Cytotoxicity in Human Cancer Cells Using Cetuximab-Targeted Gold Nanoparticles. *Journal of Experimental Therapeutics and Oncology*, 2008. **7**(4): p. 313-326.

11. King, M.C., Marks, J.H., and Mandell, J.B. Breast and Ovarian Cancer Risks Due to Inherited Mutations in *BRCA1* and *BRCA2*. *Science*, 2003. **302**(5645): p. 643-646.

12. Ponder, B. Genetic Testing for Cancer Risk. *Science*, 1997. **278**(5340): p. 1050-1054.

13. Srinivas, P.R., Verma, M., Zhao, Y., and Srivastava, S. Proteomics for Cancer Biomarker Discovery. *Clinical Chemistry*, 2002. **48**(8): p. 1160-1169.

14. Wulfkuhle, J.D., Liotta, L.A., and Petricoin, E.F. Proteomic Applications for the Early Detection of Cancer. *Nature Reviews Cancer*, 2003. **3**(4): p. 267-275.

15. Alaiya, A. Al-Mohanna, M., and Linder, S. Clinical Cancer Proteomics: Promises and Pitfalls. *Journal of Proteome Research*, 2005. **4**(4): p. 1213-1222.

16. Cooper, G.M. Cellular Transforming Genes. *Science*, 1982. **217**(4562): p. 801-806.

17. Wennerberg, K., Rossman, K.L., and Der, C.J. The Ras Superfamily at a Glance. *Journal of Cellular Science*, 2005. **118**(5): p. 843-846.

18. Oliver, M., Petitjean, A., Marcel, V., Pétré, A., Mounawar, M., Plymoth, A., de Fromentel, C.C., and Hainaut, P. Recent Advances in p53 Research: An Interdisciplinary Perspective. *Cancer Gene Therapy*, 2009. **16**(1): p. 1-12.

19. Petitjean, A., Mathe, E., Kato, S., Ishioka, C., Tavtigian, S.V., Hainaut, P., and Oliver, M. Impact of Mutant p53 Functional Properties on TP53 Mutation Patters and Tumor Phenotype: Lessons from Recent Developments in the IARC TP53 Database. *Human Mutation*, 2007. **28**(6): p. 622-629.

20. Sherr, C.J. and McCormick, F. The RB and p53 Pathways in Cancer. *Cancer Cell*, 2002. **2**(2): p. 103-112.

21. Scheffner, M., Werness, B.A., Huibreqtse, J.M., Levine, A.J., and Howley, P.M. The E6 Oncoprotein Encoded by Human Papillomavirus Types 16 and 18 Promotes the Degradation of p53. *Cell*, 1990. **63**(6): p. 1129-1136.

22. Dyson, N., Howley, P.M., Kunger, K., and Harlow, E. The Human Papilloma Virus-16 E7 Oncoprotein is Able to Bind to the Retinoblastoma Gene Product. *Science*, 1989. **243**(4893): p. 934-937.

23. Boveri, T. and Harris, H. Concerning the Origin of Malignat Tumours by Theodor Boveri, Translated and Annotated by Henry Harris. *Journal of Cell Science*, 2008. **121**(S1): p. 1-84.

24. Rowley, J.D. A New Consistent Chromosomal Abnormality in Chronic Myelogenous Leukemia Identified by Quinacrine Fluorescence and Giesma Staining. *Nature*, 1973. **243**(5405): p. 290-293.

25. Merlo, A., Herman, J.G., Mao, L., Lee, D.J., Gabrielson, E., Burger, P.C., Baylin, S.B., and Sidransky, D. 5' CpG Island Methylation is Associated with Transcriptional Silencing of the Tumor Suppressor p16/CDKN2/MTS1 in Human Cancers. *Nature Medicine*, 1995. **1**(7): p. 686-692.

26. Marotta, C.A., Wilson, J.T., Forget, B.G., and Weissman, S.M. Human β-Globin Messenger RNA. *Journal of Biological Chemistry*, 1977. **252**(14): p. 5040-5053.

27. Kerem, B.S., Rommens, J.M., Buchanan, J.A., Markiewicz, D., Cox, T.K., Chakravarti, A., Buchwald, M., and Tsui, L.C. Identification of the Cystic Fibrosis Gene: Genetic Analysis. *Science*, 1989. **245**(4922): p. 1073-1080.

28. Cadet, J. Anselmino, C., Douki, T., and Voituriez, L. Photochemistry of Nucleic Acids in Cells. *Journal of Photochemistry and Photobiology B: Biology*, 1992. **15**(4): p. 277-298.

29. Wood, A.W., Wislocki, P.G., Chang, R.L., Levin, W., Lu, A.Y.H., Yagi, H., Hernandez, O., Jerina, D.M., and Conney, A.H. Mutagenicity and Cytotoxicity of Benzo(*a*)pyrene Benzo-Ring Epoxides. *Cancer Research*, 1976. **36**(9 Pt 1): p. 3358-3366.

30. Valko, M., Izakovic, M., Mazur, M., Rhodes, C.J., and Telser, J. Role of Oxygen Radicals in DNA Damage and Cancer Incidence. *Molecular and Cellular Biochemistry*, 2004. **266**(1-2): p. 37-56.

31. Liu, Y., Hernandez, A.M., Shibata, D., and Cortopassi, G.A. *BCL2* Translocation Frequency Rises with Age in Humans. *Proceedings of the National Academy of Sciences of the United States of America*, 1994. **91**(19): p. 8910-8914.

32. Newton, C.R., Graham, A., Heptinstall, L.E., Powell, S.J., Summers, C., Kalsheker, N., Smith, J.C., and Markham, A.F. Analysis of Any Point Mutation in DNA. The Amplification Refractory Mutation System (ARMS). *Nucleic Acids Research*, 1989. **17**(7): p. 2503-2516.

33. Herman, J.G., Graff, J.R., Myöhänen, S., Nelkin, B.D., and Baylin, S.B. Methylation-Specific PCR: A Novel PCR Assay for Methylation Status of CpG Islands. *Proceedings of the National Academy of Sciences of the United States of America*, 1996. **93**(18): p. 9821-9826.

34. Mao, L., Lee, D.J., Tockman, M.S., Erozan, Y.S., Askin, F., and Sidransky, D. Microsatellite Alterations as Clonal Markers for the Detection of Human Cancer. *Proceedings of the National Academy of Sciences of the United States of America*, 1994. **91**(21): p. 9871-9875.

35. Nawroz, H., Koch, W., Anker, P., Stroun, M., and Sidransky, D. Microsatellite Alterations in Serum DNA of Head and Neck Cancer Patients. *Nature Medicine*, 1996. **2**(9): p. 1035-1037.

36. Dhanasekaran, S.M., Barrette, T.R., Ghosh, D., Shah, R., Varambally, S., Kurachi, K., Pienta, K.J., Rubin, M.A., and Chinnaiyan, A.M. Delineation of Prognostic Biomarkers in Prostate Cancer. *Nature*, 2001. **412**(6849): p. 822-826.

37. van't Veer, L.J., Dai, H., van de Vijver, M.J., He, Y.D., Hart, A.A.M., Mao, M., Peterse, H.L., van der Kooy, K., Marton, M.J., Witteveen, A.T., Schreiber, G.J., Kerkhoven, R.M., Roberts, C., Linsley, P.S., Bernards, R., and Friend, S.H. Gene Expression Profiling Predicts Clinical Outcome of Breast Cancer. *Nature*, 2002. **415**(6871): p. 530-536.

38. Dumur, C.I., Dechsukhum, C., Ware, J.L., Cofield, S.S., Best, A.M., Wilkinson, D.S., Garrett, C.T., and Ferreira-Gonzalez, A. Genome-Wide Detection of LOH in Prostate Cancer Using Human SNP Microarray Technology. *Genomics*, 2003. **81**(3): p. 260-269.

39. Futreal, P.A., Coin, L., Marshall, M., Down, T., Hubbard, T., Wooster, R., Rahman, N., and Stratton, M.R. A Census of Human Cancer Genes. *Nature Reviews Cancer*, 2004. **4**(3): p. 177-183.

40. *The Cancer Genome Atlas (TCGA) Pilot Project: Charting a New Course for Prevention, Diagnosis, and Treatment of Cancer*. 2007, National Institutes of Health.

41. The Cancer Genome Atlas Research Network. Comprehensive Genomic Characterization Defines Human Glioblastoma Genes and Core Pathways. *Nature*, 2008. **455**(7216): p. 1061-1068.

42. Attardi, G. Biogenesis of Mitochondria. *Annual Review of Cell Biology*, 1988. **4**: p. 280-333.

43. Jiang, X. and Wang, X. Cytochrome *C*-Mediated Apoptosis. *Annual Review of Biochemistry*, 2004. **73**: p. 87-106.

44. Wallace, D.C. A Mitochondrial Paradigm of Metabolic and Degenerative Diseases, Aging, and Cancer: A Dawn for Evolutionary Medicine. *Annual Review of Genetics*, 2005. **39**: p. 357-407.

45. Ballinger, S.W. Mitochondrial Dysfunction in Cardiovascular Disease. *Free Radical Biology & Medicine*, 2005. **38**(10): p. 1278-1295.

46. Warburg, O. On the Origin of Cancer Cells. *Science*, 1956. **123**(3191): p. 309-314.

47. Kim, J.W. and Dang, C.V. Cancer's Molecular Sweet Tooth and the Warburg Effect. *Cancer Research*, 2006. ***66***(18): p. 8927-8930.

48. Pelicano, H., Martin, D.S., Xu, R.H., and Huang, P. Glycolysis Inhibition for Anticancer Treatment. *Oncogene*, 2006. **25**(34): p. 4633-4646.

49. Simonnet, H., Alazard, N., Pfeiffer, K., Gallou, C., Béroud, C., Demont, J., Bouvier, R., Schägger, H., and Godinot, C. Low Mitochondrial Respiratory Chain Content Correlates with Tumor Aggressiveness in Renal Cell Carcinoma. *Carcinogenesis*, 2002. **23**(5): p. 759-768.

50. Pelicano, H., Carney, D., and Huang, P. ROS Stress in Cancer Cells and Therapeutic Implications. *Drug Resistance Updates*, 2004. **7**(2): p. 97-110.

51. Anderson, S., Bankier, A.T., Barrell, B.G., de Bruijn, M.H.L., Coulson, A.R., Drouin, J., Eperson, I.C., Nierlich, D.P., Roe, B.A., Sanger, F., Schreier, P.H., Smith, A.J.H., Staden, R., and Young, I.G. Sequence and Organization of the Human Mitochondrial Genome. *Nature*, 1981. **290**(5806): p. 457-465.

52. Andrews, R.M., Kubacka, I., Chinnery, P.F., Lightowlers, R.N., Turnbull, D.M., and Howell, N. Reanalysis and Revision of the Cambridge Reference Sequence for Human Mitochondrial DNA. *Nature Genetics* 1999. **23**(2): p. 147.

53. Cann, R.L., Stoneking, M., and Wilson, A.C. Mitochondrial DNA and Human Evolution. *Nature*, 1987. **325**(6099): p. 31-36.

54. Holland, M.M. and Parsons, T.J. Mitochondrial DNA Sequence Analysis—Validation and Use for Forensic Casework. *Forensic Science Review*, 1999. **11**(1): p. 21-50.

55. Pesole, G., Gissi, C., DeChirico, A., and Saccone, C. Nucleotide Substitution Rate of Mammalian Mitochondrial Genomes. *Journal of Molecular Evolution*, 1999. **48**(4): p. 427-434.

56. Croteau, D.L., and Bohr, V.A. Repair of Oxidative Damage to Nuclear and Mitochondrial DNA in Mammalian Cells. *Journal of Biological Chemistry*, 1997. **272**(41): p. 25409-25412.

57. Larsen, N.B., Rasmussen, M., and Rasmussen, L.J. Nuclear and Mitochondrial DNA Repair: Similar Pathways? *Mitochondrion*, 2005. **5**(2): p. 89-108.

58. Attardi, G. Role of Mitochondrial DNA in Human Aging. *Mitochondrion*, 2002. **2**(1-2): p. 27-37.

59. Kujoth, G.C., Leeuwenburgh, C., and Prolla, T.A. Mitochondrial DNA Mutations and Apoptosis in Mammalian Aging. *Cancer Research*, 2006. **66**(15): p. 7386-7389.

60. Mecocci, P., MacGarvey, U., and Beal, M.F. Oxidative Damage to Mitochondrial DNA is Increased in Alzheimer's Disease. *Annals of Neurology*, 1994. **36**(5): p. 747-751.

61. Bai, R.K., Leal, S.M., Covarrubias, D., Liu, A., and Wong, L.J.C. Mitochondrial Genetic Background Modifies Breast Cancer Risk. *Cancer Research*, 2007. **67**(10): p. 4687-4695.

62. Booker, L.M., Habermacher, G.M., Jessie, B.C., Sun, Q.C., Baumann, A.K., Amin, M., Lim, S.D., Fernandez-Golarz, C., Lyles, R.H., Brown, M.D., Marshall, F.F., and Petros, J.A. North American White Haplogroups in Prostate and Renal Cancer. *Journal of Urology*, 2006. **175**(2): p. 468-473.

63. Sanchez-Cespedes, M., Parrella, P., Nomoto, S., Cohen, D., Xiao, Y., Esteller, M., Jeronimo, C., Jordan, R.C.K., Nicol, T., Koch, W.M., Schoenberg, M., Mazzarelli, P., Fazio, V.M., and Sidransky, D. Identification of a Mononucleotide Repeat as a Major Target for Mitochondrial DNA Alterations in Human Tumors. *Cancer Research*, 2001. **61**(19): p. 7015-7019.

64. Parrella, P., Seripa, D., Matera, M.G., Rabitti, C., Rinaldi, M., Mazzarelli, P., Gravina, C., Gallucci, M., Altomare, V., Flammia, G., Casalino, B., Benedetti-Panici, P.L., and Fazio, V.M. Mutations of the D310 Mitochondrial Mononucleotide Repeat in Primary Tumors and Cytological Specimens. *Cancer Letters*, 2003. **190**(1): p. 73-77.

65. Tang, M., Baez, S., Pruyas, M., Diaz, A., Calvo, A., Riquelme, E., and Wistuba, I.I. Mitochondrial DNA Mutation at the D310 (Displacement Loop) Mononucleotide Sequence in the Pathogenesis of Gallbladder Carcinoma. *Clinical Cancer Research*, 2004. **10**(3): p. 1041-1046.

66. Ye, C. Shu, X.O., Wen, W., Pierce, L., Courtney, R., Gao, Y.T., Zheng, W., and Cai, Q. Quantitative Analysis of Mitochondrial DNA 4977-bp Deletion in Sporadic Breast Cancer and Benign Breast Diseases. *Breast Cancer Research and Treatment*, 2008. **108**(3): p. 427-434.

67. Dai, J.G., Xiao, Y.B., Min, J.X., Zhang, G.Q., Yao, K., and Zhou, R.J. Mitochondrial DNA 4977 BP Deletion Mutations in Lung Carcinoma. *Indian Journal of Cancer*, 2006. **43**(1): p. 20-25.

68. Fliss, M.S., Usadel, H., Caballero, O.L., Wu, L., Buta, M.R., Eleff, S.M., Jen, J., and Sidransky, D. Facile Detection of Mitochondrial DNA Mutations in Tumors and Bodily Fluids. *Science*, 2000. **287**(5460): p. 2017-2019.

69. Parrella, P., Xiao, Y., Fliss, M., Sanchez-Cespedes, M., Mazzarelli, P., Rinaldi, M., Nicol, T., Gabrielson, E., Cuomo, C., Cohen, D., Pandit, S., Spencer, M., Rabitti, C., Fazio, V.M., and Sidransky, D. Detection of Mitochondrial DNA Mutations in Primary Breast Cancer and Fine-Needle Aspirates. *Cancer Research*, 2001. **61**(20): p. 7623-7626.

70. Tan, D.J., Bai, R.K., and Wong, L.J.C. Comprehensive Scanning of Somatic Mitochondrial DNA Mutations in Breast Cancer. *Cancer Research*, 2002. **62**(4): p. 972-976.

71. Nagy, A., Wilhelm, M., Sükösd, F., Ljungberg, B., and Kovacs, G. Somatic Mitochondrial DNA Mutations in Human Chromophobe Renal Cell Carcinomas. *Genes, Chromosomes, and Cancer*, 2002. **35**(3): p. 256-260.

72. Nomoto, S., Yamashita, K., Koshikawa, K., Nakao, A., and Sidransky, D. Mitochondrial D-loop Mutations as Clonal Markers in Multicentric Hepatocellular Carcinoma and Plasma. *Clinical Cancer Research*, 2002. **8**(2): p. 481-487.

73. Okochi, O., Hibi, K., Uemura, T., Inoue, S., Takeda, S., Kaneko, T., and Nakao, A. Detection of Mitochondrial DNA Alterations in the Serum of Hepatocellular Carcinoma Patients. *Clinical Cancer Research*, 2002. **8**(9): p. 2875-2878.

74. Polyak, K., Li, Y., Zhu, H., Lengauer, C., Willson, J.K.V., Markowitz, S.D., Trush, M.A., Kinzler, K.W., and Vogelstein, B. Somatic Mutations of the Mitochondrial Genome in Human Colorectal Tumours. *Nature Genetics*, 1998. **20**(3): p. 291-293.

75. Hibi, K., Nakayama, H., Yamazaki, T., Takase, T., Taguchi, M., Kasai, Y., Ito, K., Akiyama, S., and Nakao, A. Detection of Mitochondrial DNA Alterations in Primary Tumors and Corresponding Serum of Colorectal Cancer Patients. *International Journal of Cancer*, 2001. **94**(3): p. 429-431.

76. Liu, V.W.S., Shi, H.H., Cheung, A.N.Y., Chiu, P.M., Leung, T.W., Nagley, P., Wong, L.C., and Ngan, H.Y.S. High Incidence of Somatic Mitochondrial DNA Mutations in Human Ovarian Carcinomas. *Cancer Research*, 2001. **61**(16): p. 5998-6001.

77. Burgart, L.J., Zheng, J., Shu, Q., Strickler, J.G., and Shibata, D. Somatic Mitochondrial Mutation in Gastric Cancer. *American Journal of Pathology*, 1995. **147**(4): p. 1105-1111.

78. Kumimoto, H., Yamane, Y., Nishimoto, Y., Fukami, H., Shinoda, M., Hatooka, S., and Ishizaki, K. Frequent Somatic Mutations of Mitchondrial DNA in Esophageal Squamous Cell Carcinoma. *International Journal of Cancer*, 2004. **108**(2): p. 228-231.

79. Mithani, S.K., Smith, I.M., Topalian, S.L., and Califano, J.A. Nonsynonymous Somatic Mitochondrial Mutations Occur in the Majority of Cutaneous Melanomas. *Melanoma Research*, 2008. **18**(3): p. 214-219.

80. Jones, J.B., Song, J.J., Hempen, P.M., Parmigiani, G., Hruban, R.H., and Kern, S.E. Detection of Mitochondrial DNA Mutations in Pancreatic Cancer Offers a "Mass"-ive Advantage over Detection of Nuclear DNA Mutations. *Cancer Research*, 2001. **61**(4): p. 1299-1304.

81. Tong, B.C., Ha, P.K., Dhir, K., Xing, M., Westra, W.H., Sidransky, D., and Califano, J.A. Mitochondrial DNA Alterations in Thyroid Cancer. *Journal of Surgical Oncology*, 2003. **82**(3): p. 170-173.

82. Lohrer, H.D., Hieber, L., and Zitzelsberger, H. Differential Mutation Frequency in Mitochondrial DNA from Thyroid Tumours. *Carcinogenesis*, 2002. **23**(10): p. 1577-1582.

83. Wong, L.J.C., Lueth, M., Li, X.N., Lau, C.C., and Vogel, H. Detection of Mitochondrial DNA Mutations in the Tumor and Cerebrospinal Fluid of Medulloblastoma Patients. *Cancer Research*, 2003. **63**(14): p. 3866-3871.

84. Mithani, S.K., Taube, J.M., Zhou, S., Smith, I.M., Koch, W.M., Westra, W.H., and Califano, J.A. Mitochondrial Mutations Are a Late Event in the Progression of Head and Neck Squamous Cell Cancer. *Clinical Cancer Research*, 2007. **13**(15): p. 4331-4335.

85. Mithani, S.K., Smith, I.M., Zhou, S., Gray, A., Koch, W.M., Maitra, A., and Califano, J.A. Mutations in Salivary Rinses of Patients with Head and Neck Cancer. *Clinical Cancer Research*, 2007. **13**(24): p. 7335-7340.

86. Dasgupta, S., Hoque, M.O., Upadhyay, S., and Sidransky, D., Mitochondrial Cytochrome B Gene Mutation Promotes Tumor Growth in Bladder Cancer. *Cancer Research*, 2008. **68**(3): p. 700-706.

87. Jakupciak, J.P., Wang, W., Markowitz, M.E., Ally, D., Coble, M., Srivastava, S., Maitra, A., Barker, P.E., Sidransky, D., and O'Connell, C.D. Mitochondrial DNA as a Cancer Biomarker. *Journal of Molecular Diagnostics*, 2005. **7**(2): p. 258-267.

88. Suzuki, M., Toyooka, S., Miyajima, K., Iizasa, T., Fujisawa, T., Bekele, N.B., and Gazdar, A.F. Alterations in the Mitochondrial Displacement Loop in Lung Cancers. *Clinical Cancer Research*, 2003. **9**(15): p. 5636-5641.

89. Kogelnik, A.M., Lott, M.T., Brown, M.D., Navathe, S.B., and Wallace, D.C. MITOMAP: A Human Mitochondrial Genome Database. *Nucleic Acids Research*, 1996. **24**(1): p. 177-179. URL: http://www.mitomap.org/.

90. Ruiz-Pesini, E., Lott, M.T., Procaccio, V., Poole, J., Brandon, M.C., Mishmar, D., Yi, C., Kreuziger, J., Baldi, P., and Wallace, D.C. An Enhanced MITOMAP with a global mtDNA Mutational Phylogeny. *Nucleic Acids Research*, 2007. **35**(Database Issue): p. D823-D828. URL: http://www.mitomap.org/.

91. Higuchi, M. Regulation of Mitochondrial DNA Content and Cancer. *Mitochondrion*, 2007. **7**(1-2): p. 53-57.

92. Hirst, J., King, M.S., and Pryde, K.R. The Production of Reactive Oxygen Species by Complex I. *Biochemical Society Transactions*, 2008. **36**(5): p. 976-980.

93. Ishikawa, K., Takenaga, K., Akimoto, M., Koshikawa, N., Yamaguchi, A., Imanishi, H., Nakada, K., Honma, Y., and Hayashi, J.I. ROS-Generating Mitochondrial DNA Mutations Can Regulate Tumor Cell Metastasis. *Science*, 2008. **320**(5876): p. 661-664.

94. Brandon, M., Baldi, P., and Wallace, D.C. Mitochondrial Mutations in Cancer. *Oncogene*, 2006. **25**(34): p. 4647-4662.

95. Sanger, F., Nicklen, S., and Coulson, A.R. DNA Sequencing with Chain-Terminating Inhibitors. *Proceedings of the National Academy of Sciences of the United States of America*, 1977. **74**(12): p. 5463-5467.

96. Smith, L.M., Sanders, J.Z., Kaiser, R.J., Hughes, P., Dodd, C., Connell, C.R., Heiner, C., Kent, S.B.H., and Hood, L.E. Fluorescence Detection in Automated DNA-Sequence Analysis. *Nature*, 1986. **321**(6071): p. 674-679.

97. Prober, J.M., Trainor, G.L., Dam, R.J., Hobbs, F.W., Robertson, C.W., Zagursky, R.J., Cocuzza, A.J., Jensen, M.A., and Baumeister, K. A System for Rapid DNA Sequencing with Fluorescent Chain-Terminating Dideoxynucleotides. *Science*, 1987. **238**(4825): p. 336-341.

98. Ju, J., Ruan, C., Fuller, C.W., Glazer, A.N., and Mathies, R.A. Fluorescence Energy Transfer Dye-Labeled Primers for DNA Sequencing and Analysis. *Proceedings of the National Academy of Sciences of the United States of America*, 1995. **92**(10): p. 4347-4351.

99. Shendure, J. and Ji, H. Next-Generation DNA Sequencing. *Nature Biotechnology*, 2008. **26**(10): p. 1135-1145.

100. Venter, J.C., *et al*. The Sequence of the Human Genome. *Science*, 2001. **291**(5507): p. 1304-1351.

101. Ronaghi, M., Karamohamed, S., Pettersson, B., Ulhén, M., and Nyrén, P. Real-Time DNA Sequencing Using Detection of Pyrophosphate Release. *Analytical Biochemistry*, 1996. **242**(1): p. 84-89.

102. Metzker, M.L., Raghavachari, R., Richards, S., Jacutin, S.E., Civitello, A., Burgess, K., and Gibbs, R.A. Termination of DNA Synthesis by Novel 3'-Modified-Deoxyribonucleoside 5'-Triphosphates. *Nucleic Acids Research*, 1994. **22**(20): p. 4259-4267.

103. Kasianowicz, J.J., Brandin, E., Branton, D., and Deamer, D.W. Characterization of Individual Polynucleotide Molecules Using a Membrane Channel. *Proceedings of the National Academy of Sciences of the United States of America*, 1996. **93**(24): p. 13770-13773.

104. Shendure, J., Porreca, G.J., Beppas, N.B., Lin, X., McCutcheon, J.P., Rosenbaum, A.M., Wang, M.D., Zhang, K., Mitra, R.D., and Church, G.M. Accurate Multiplex Polony Sequencing of an Evolved Bacterial Genome. *Science*, 2005. **309**(5741): p. 1728-1732.

105. Blazej, R.G., Paegel, B.M., and Mathies, R.A. Polymorphism Ratio Sequencing: A New Approach for Single Nucleotide Polymorphism Discovery and Genotyping. *Genome Research*, 2003. **13**(2): p. 287-293.

106. Salas, A., Yao, Y.G., Macaulay, V., Vega, A., Carracedo, A., and Bandelt, H.J. A Critical Reassessment of the Role of Mitochondria in Tumorigenesis. *Public Library of Science Medicine,* 2005. **2**(11): p. e296.

107. Wang, C.Y., Wang, H.W., Yao, Y.G., Kong, Q.P., and Zhang, Y.P. Somatic Mutations of Mitochondrial Genome in Early Stage Breast Cancer. *International Journal of Cancer*, 2007. **121**(6): p. 1253-1256.

108. Kothapalli, R. Yoder, S.J., Mane, S., and Loughran, T.P. Microarray Results: How Accurate Are They? *BioMed Central Bioinformatics*, 2002. **3**(22).

109. Woolley, A.T., Sensabaugh, G.F., and Mathies, R.A. High-Speed DNA Genotyping Using Microfabricated Capillary Array Electrophoresis Chips. *Analytical Chemistry*, 1997. **69**(11): p. 2182-2186.

110. Simpson, P.C., Roach, D., Woolley, A.T., Thorsen, T., Johnston, R., Sensabaugh, G.F., and Mathies, R.A. High-Throughput Genetic Analysis Using Microfabricated 96-Sample Capillary Array Electrophoresis Microplates. *Proceedings of the National Academy of Sciences of the United States of America*, 1998. **95**(5): p. 2256-2261.

111. Harrison, D.J., Fluri, K., Seiler, K., Fan, Z., Effenhauser, C.S., and Manz, A. Micromachining a Miniaturized Capillary Electrophoresis-Based Chemical-Analysis System on a Chip. *Science*, 1993. **261**(5123): p. 895-897.

112. Woolley, A.T. and Mathies, R.A. Ultra-High-Speed DNA Fragment Separations Using Microfabricated Capillary Array Electrophoresis Chips. *Proceedings of the National Academy of Sciences of the United States of America*, 1994. **91**(24): p. 11348-11352.

113. Woolley, A.T. and Mathies, R.A. Ultra-High-Speed DNA Sequencing Using Capillary Electrophoresis Chips. *Analytical Chemistry*, 1995. **67**(20): p. 3676-3680.

114. Shi, Y., Simpson, P.C., Scherer, J.R., Wexler, D., Skibola, C., Smith, M.T., and Mathies, R.A. Radial Capillary Array Electrophoresis Microplate and Scanner for High-Performance Nucleic Acid Analysis. *Analytical Chemistry*, 1999. **71**(23): p. 5354-5361.

115. Paegel, B.M., Emrich, C.A., Wedemayer, G.J., Scherer, J.R., and Mathies, R.A. High Throughput DNA Sequencing with a Microfabricated 96-lane Capillary Array Electrophoresis Bioprocessor. *Proceedings of the National Academy of Sciences of the United States of America*, 2002. **99**(2): p. 574-579.

116. Emrich, C.A., Tian, H., Medintz, I.L., and Mathies, R.A. Microfabricated 384-Lane Capillary Array Electrophoresis Bioanalyzer for Ultrahigh-Throughput Genetic Analysis. *Analytical Chemistry*, 2002. **74**(19): p. 5076-5083.

117. Paegel, B.M., Yeung, S.H.I., and Mathies, R.A. Microchip Bioprocessor for Integrated Nanovolume Sample Purification and DNA Sequencing. *Analytical Chemistry*, 2002. **74**(19): p. 5092-5098.

118. Woolley, A.T., Hadley, D., Landre, P., deMello, A.J., Mathies, R.A., and Northrup, M.A. Functional Integration of PCR Amplification and Capillary Electrophoresis in a Microfabricated DNA Analysis Device. *Analytical Chemistry* 1996. **68**(23): p. 4081-4086.

119. Grover, W.H., Skelley, A.M., Liu, C.N., Lagally, E.T., and Mathies, R.A. Monolithic Membrane Valves and Diaphragm Pumps for Practical Large-Scale Integration into Glass Microfluidic Devices. Sensors *and Actuators B-Chemical*, 2003. **89**(3): p. 315-323.

120. Liu, P., Yeung, S.H.I., Crenshaw, K.A., Crouse, C.A., Scherer, J.R., and Mathies, R.A. Real-Time Forensic DNA Analysis at a Crime Scene Using a Portable Microchip Analyzer. *Forensic Science International: Genetics*, 2008. **2**(4): p. 301-309.

121. Lagally, E.T., Scherer, J.R., Blazej, R.G., Toriello, N.M., Diep, B.A., Ramchandani, M., Sensabaugh, G.F., Riley, L.W., and Mathies, R.A. Integrated Portable Genetic Analysis Microsystem for Pathogen/Infectious Disease Detection. *Analytical Chemistry*, 2004. **76**(11): p. 3162-3170.

122. Skelley, A.M ., Scherer, J.R., Aubrey, A.D., Grover, W.H., Ivester, R.H.C., Ehrenfreund, P., Grunthaner, F.J., Bada, J.L., and Mathies, R.A. Development and Evaluation of a Microdevice for Amino Acid Biomarker Detection and Analysis on Mars. *Proceedings of the National Academy of Sciences of the United States of America*, 2005. **102**(4): p. 1041-1046.

123. Kopecka, K., Drouin, G., and Slater, G.W. Capillary Electrophoresis Sequencing of Small ssDNA Moleules Versus the Ogston Regime: Fitting Data and Interpreting Parameters. *Electrophoresis*, 2004. **25**(14): p. 2177-2185.

124. Vatcher, G., Smailus, D., Krzywinski, M., Guin, R., Stott, J., Tsai, M., Chan, S., Pandoh, P., Yang, G., Asano, J., Olson, T., Prabhu, A.L., Coope, R., Marziali, A., Schein, J., Jones, S., and Marra, M. Resuspension of DNA Sequencing Reaction Products in Agarose Increases Sequence Quality on an Automated Sequencer. *Biotechniques*, 2002. **33**(3): p. 532-534.

125. Almira, E.C., Panayotova, N., and Farmerie, W.G. Capillary DNA Sequencing: Maximizing the Sequence Output. *Journal of Biomolecular Techniques*, 2003. **14**(4): p. 270-277.

126. Hjertén, S. Elimination of Electroendosmosis and Solute Adsorption. *Journal of Chromatography*, 1985. **347**: p. 191-198.

127. Albarghouthi, M.N., Buchholz, B.A., Huiberts, P.J., Stein, T.M., and Barron, A.E. Poly-N-hydroxyethylacrylamide (polyDuramide): A Novel, Hydrophilic, Self-Coating Polymer for DNA Sequencing by Capillary Electrophoresis. *Electrophoresis*, 2002. **23**(10): p.1429-1440.

128. Giddings, M.C., Severin, J., Westphall, M., Wu, J., Smith, L.M. A Software System for Data Analysis in Automated DNA Sequencing. *Genome Research*, 1998. **8**(6): p. 644-665.

129. Rieder, M.J., Taylor, S.L., Tobe, V.O., and Nickerson, D.A. Automating the Identification of DNA Variations Using Quality-Based Fluorescence Re-sequencing: Analysis of the Human Mitochondrial Genome. *Nucleic Acids Research*, 1998. **26**(4): p. 967-973.

130. Scherer, J.R., Paegel, B.M., Wedemayer, G.J., Emrich, C.A., Lo, J., Medintz, I.L., and Mathies, R.A. High-Pressure Gel Loader for Capillary Array Electrophoresis Microchannel Plates. *Biotechniques*, 2001. **31**(5): p. 1150-1152.

131. Maitra, A., Cohen, Y., Gillespie, S.E.D., Mambo, E., Fukushima, N., Hoque, M.O., Shah, N., Goggins, M., Califano, J., Sidransky, D., and Chakravarti, A. The Human MitoChip: A High-Throughput Sequencing Microarray for Mitochondrial Mutation Detection. *Genome Research*, 2004. **14**(5): p. 812-819.

132. Zhou, S., Kassauei, K., Cutler, D.J., Kennedy, G.C., Sidransky, D., Maitra, A., and Califano, J. An Oligonucleotide Microarray for High-Throughput Sequencing of the Mitochondrial Genome. *Journal of Molecular Diagnostics*, 2006. **8**(4): p. 476-482.

133. Jakupciak, J.P., Maragh, S., Markowitz, M.E., Greenberg, A.K., Hoque, M.O., Maitra, A., Barker, P.E., Wagner, P.D., Rom, W.N., Srivastava, S., Sidransky, D., and O'Connell, C.D. Performance of Mitochondrial DNA Mutations Detecting Early Stage Cancer. *BioMed Central Cancer*, 2008. **8**(285).

134. Blazej, R.G., Kumaresan, P., and Mathies, R.A. Microfabricated Bioprocessor for Integrated Nanoliter-Scale Sanger DNA Sequencing. *Proceedings of the National Academy of Sciences of the United States of America*, 2006. **103**(19): p. 7240-7245.

135. Blazej, R.G., Kumaresan, P., Cronier, S.A., and Mathies, R.A. Inline Injection Microdevice for Attomole-Scale Sanger DNA Sequencing. *Analytical Chemistry*, 2007. **79**(12): p. 4499-4506.

136. Brahmasandra, S.N., Ugaz, V.M., Burke, D.T., Mastrangelo, C.H., and Burns. M.A. Electrophoresis in Microfabricated Devices Using Photopolymerized Polyacrylamide Gels and Electrode-Defined Sample Injection. *Electrophoresis*, 2001. **22**(2): p. 300-311.

137. Yeung, S.H.I., Liu, P., Del Bueno, N., Greenspoon, S.A., and Mathies, R.A. Integrated Sample Cleanup-Capillary Electrophoresis Microchip for High-Performance Short Tandem Repeat Genetic Analysis. *Analytical Chemistry*, 2009. **81**(1): p. 210-217.

138. Thaitrong, N., Toriello, N.M., Del Bueno, N., and Mathies, R.A. Polymerase Chain Reaction-Capillary Electrophoresis Genetic Analysis Microdevice with In-Line Affinity Capture Sample Injection. *Analytical Chemistry*, 2009. **81**(4): p. 1371-1377.

139. Kumaresan, P., Yang, C.J., Cronier, S.A., Blazej, R.G., and Mathies, R.A. High-throughput Single Copy DNA Amplification and Cell Analysis in Engineered Nanoliter Droplets. *Analytical Chemistry*, 2008. **80**(10): p. 3522-3529.

140. Rautio, J.J., Satokari, R., Vehmaan-Kreula, P. Serkkola, E., and Söderlund, H. TRAC in High-content Gene Expression Analysis: Applications in Microbial Population Studies, Process Biotechnology and Biomedical Research. *Expert Review of Molecular Diagnostics*, 2008. **8**(4): p. 379-385.

141. Satokari, R.M., Kataja, K., and Söderlund. H. Multiplexed Quantification of Bacterial 16S rRNA by Solution Hybridization with Oligonucleotide Probes and Affinity Capture. *Microbial Ecology*, 2005. **50**(1): p. 120-127.

142. Rautio, J.J., Kataja, K., Satokari, R., Penttilä, M., Söderlund, H., and Saloheimo, M. Rapid and Multiplexed Transript Analysis of Microbial Cultures Using Capillary Electrophoresis-Detectable Oligonucleotide Probe Pools. *Journal of Microbiological Methods*, 2006. **65**(3): p. 404-416.

143. Rautio, J.J., Bailey, M., Kivioja, T., Söderlund, H., Penttilä, M., and Saloheimo, M. Physiological Evaluation of the Filamentous Fungus Trichoderma Reesei in Production Processes by Marker Gene Expression Analysis. *BioMed Central Biotechnology*, 2007. **7**(28).

144. Kataja, K., Satokari, R.M., Arvas, M., Takkinen, K., and Söderlund, H. A Highly Sensitive and Multiplexed Method for Focused Transcript Analysis. *Journal of Microbiological Methods*, 2006. **67**(1): p. 102-113.

145. Toriello, N.M., Liu, C.N., Blazej, R.G., Thaitrong, N., and Mathies, R.A. Integrated Affinity Capture, Purification, and Capillary Electrophoresis Microdevice for Quantitative Double-Stranded DNA Analysis. *Analytical Chemistry*, 2007. **79**(22): p. 8549-8556.

146. Heller, C. Separation of Double-Stranded and Single-Stranded DNA in Polymer Solutions: II. Separation, Peak Width and Resolution. *Electrophoresis*, 1999. **20**(10): p. 1978-1986.