

Cell lineage tracing in human tissues using somatic mitochondrial mutations

THÈSE N° 7643 (2017)

PRÉSENTÉE LE 17 NOVEMBRE 2017
À LA FACULTÉ DES SCIENCES DE BASE
CHAIRE DE STATISTIQUE APPLIQUÉE
PROGRAMME DOCTORAL EN BIOTECHNOLOGIE ET GÉNIE BIOLOGIQUE

ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE

POUR L'OBTENTION DU GRADE DE DOCTEUR ÈS SCIENCES

PAR

Paul REFINETTI

acceptée sur proposition du jury:

Prof. P. De Los Rios, président du jury
Prof. S. Morgenthaler, Prof. P. O. Ekstrom, directeurs de thèse
Prof. W. Thilly, rapporteur
Prof. B. John, rapporteur
Prof. B. Deplanke, rapporteur



ÉCOLE POLYTECHNIQUE
FÉDÉRALE DE LAUSANNE

Suisse
2017

To see in the darkness...
Give your eyes the time to adjust

To my family, my parents and sisters...

Acknowledgements

A PhD is a journey. A journey that takes you from a young graduate full of knowledge to a young scientist that does not know anything.

It is a journey I have not, and never could have walked alone. There is no way to convey how grateful I am to everyone that has been around me during this time, I can only thank a few of you. Per Ekstrom, thesis co-supervisor, colleague, teacher, complete looney and friend. The success of this thesis has been the success of our two-man team. From conception, through realisation passing by the various steps of Limoncello Scoring, MeMeMeMeMe, CWOT...

William Thilly, you tried, and are still trying to teach me how to be a scientist. Most of the successful ways of thinking and ideas that make up this thesis came from you.

Christian Arstad, a colleague and a friend. You decided to believe in our project and join us. In doing so you made significant contribution to the study of mtDNA mutations.

William Pralong, a Mentor, a true mentor. Advised me not only in matters of science, but of people, society and personal life. You have been a mentor not only during my PhD, but ever since the day, 11 years ago, when I first walked into the newly started faculty of life sciences, and decided to join.

Masaru Kato, thank you for inviting me to the University of Tokyo, a life changing experience. Gian-Andrea Pasquinelli, Silvia Fittipaldi and Francesco Vasuri, friends and colleagues. Thank you for believing in my ideas and starting a successful collaboration.

My friends and PhD colleagues, Matthieu Delince, Georges Muller, Pierluigi Manti, Daria Rukina, Ulugbek Kamilov, Priscila Briquez, Natalie Brandenburg, Michiko Kanemitsu, and many others, some of whom have been my colleagues since first year, or even before that.

The young students, whom I had the privilege to supervise, Heinz Strasle, Mery Aschkan and Mercedes Ondik. You were a

My good friends around the world. Wherever I was, you were always close by, making sure I never forgot that the world is larger than a computer screen: Eric Merk, Hiromi Kimura, Christina Bjoraneset, Jeanne Guzman, Neelu Westvang, Tatiana Marino, Simen Reynolds, Yulia Fofanova, Anastasia Morienko, Gregory Gortz, Yulia Novikova, Michael King, Ricardo and Marcelo Camps, and many more.

Florence Driessens, my partner for most of my PhD career, and a great friend. I would not be who I am today without you.

Last, but not least Stephan Morgenthaler... You have supervised all the projects I did at EPFL, from Bachelor to Master through PhD. Thank you for continuously believing in me, and taking on projects that seemed impossible at first, in a field outside your main area of expertise. Even

Acknowledgements

though supervising me has not been easy you were always present, waiting when I needed time, and pushing when I needed pushing. I could not imagine having had a better supervisor. To those I forgot, thank you for forgiving me.

Lausanne, 18 January 2017

P. R.

Preface

This thesis emerged from Paulo Refinetti's interest in understanding tumor growth. He had done projects on this topic during his Master's studies, in particular for his Master's thesis, supervised and hosted by Prof. William G. Thilly from MIT. During his stay at MIT Paulo Refinetti learned key ideas and methods and also was introduced to Dr. Per Ekstrom who would become his co-supervisor and laboratory host during his Ph.D. studies. After some initial experiments and wide-ranging discussions, a Ph.D. project focussing on investigating the interaction between stromal cells and tumor cells was decided on. This interplay is generally believed to play a major role in cancer growth and progression. His Master's projects had prepared Paulo Refinetti for working with mutations both in the theoretical sense of mathematical biology and in the biological sense of detecting, characterising and quantifying them. Observing the spread of mutations in mitochondrial DNA is particularly attractive. Because there are many copies of the mitochondrial DNA present in each cell, the somatic mutational process is accelerated and it should thus be easier to find tumors with such mutations. Determining the multiplicity of mitochondrial DNA in cells was one of the questions Paulo Refinetti tackled and it led him to propose and develop a novel mathematical model for analysing the results of an rtPCR run. The first part of the thesis describes the adaptation of the CDCE method to human mitochondrial DNA and resulted in a low cost and highly effective tool. Paulo Refinetti also wrote software that can automatically analyse and score the output from CDCE runs. The effectiveness was demonstrated in an application to samples taken from many tumors, a comparison with next generation sequencing and in two clinical applications. The analysis of tumor samples showed that in some cases, somatic mutations in mitochondrial DNA that are present in high fractions are found. The second part of the thesis covers the application to samples taken from tumors. At the beginning, the traditional method relying on staining was believed to be able to distinguish stromal cells from tumor cells. This would have allowed to take samples from either source and score them separately. In the available samples, this turned out to be inconclusive. Taking small patches of cells that cover the whole of a tumor slice, and thus sample it systematically, turned out to be a better idea. In tumors, where a mutation in the mitochondrial DNA is present, the analysis of the patches gives a fairly precise picture of the spatial distribution of the mutation within the slice. The final step is taken by slicing a three dimensional tumor tissue into layers. Paulo Refinetti demonstrated the feasibility of this approach and applied it in two examples. The graphical representation of this data shows the spread of a mutation via the growth of a tumor.

Lausanne, 17 January 2017

Stephan Morgenthaler

Abstract

Lineage tracing is a basic concept in developmental biology, as well as diseases related to organ development. A reliable method to track cell lineages in human tissues has the potential to greatly contribute to the fields of oncogenesis, tissue renewal, ageing as well as wound healing.

Inherited mitochondrial mutations have been a successful tool in the understanding of human lineage and migrations throughout history. Similarly, somatic mitochondrial mutations have been used to study the patterns of cell division and renewal in some human tissues. Each cell has many copies of the mitochondrial genome, a mutation can therefore be present in any fraction of the total number of mitochondrial DNA copies. To study somatic mitochondrial mutations therefore requires mutations to be detected, and the mutant fraction to be quantified. This requirement makes the use of most sequencing techniques very difficult.

The first step in the development of lineage tracing methods using somatic mtDNA mutations as a marker, is a quantitative, fast and high throughput detection method. A two step PCR was designed to amplify first the whole mtDNA in large fragments without nuclear DNA amplification. Second, a set of small fragments that can be scanned for mutations using cycling temperature capillary electrophoresis. In the end 76% of the mitochondrial genome could be scanned for mutations. Tests showed that mutations present at a fraction as low as 1% of the total copy numbers could be detected and accurately quantified. Furthermore, the method could be applied to analyse micro-anatomical samples taken by laser capture micro dissection.

Lineage tracing in human tumors was achieved by scanning the mitochondrial DNA for mutations in samples extracted from macro-anatomical samples taken across surgically removed and freshly frozen tumor pieces. Tumors in which a mutation was detected at a sufficient fraction of the extracted DNA were finely sampled. The first attempt consisted in taking 286 samples across a slice of a Leydig cell tumor, in which two separate mutations were identified at a high enough fraction to justify further investigation. Both mutations strongly correlated suggesting a single lineage. Since no discernible pattern emerged, the results were validated by taking 96 samples on another slice of tissue, slightly deeper, and observing that in similar positions, similar mutant fractions could be reproduced.

The natural evolution was to take multiple successive slices of tissue. On each slice 96 samples, arranged in an 8×12 grid are then taken. Analysing the results from 19 layers gave the three dimensional reconstruction of a tumor lineage in a tissue volume. The results show that the lineage followed is quite diffused through the volume and does not occupy its majority.

A three dimensional reconstruction of a cell lineage was then established for two more tumors, and three metastasis. Under the assumption that a metastasis is seeded by a single tumor cell, the

Abstract

presence of the same mtDNA mutation in both primary tumor and metastasis guarantees that only cells carrying the mutation are tumor-derived. The three dimensional reconstruction of a primary tumor lineage through a metastasis is therefore the reconstruction of the patterns of metastasis growth in its host tissue. Observing that most of the metastasis volume is not composed of tumor derived cells suggests that metastatic cells grow by diffusing through their host tissue, or by recruiting host organ cells, as well as conditioning them to assume its abnormal morphology. Lineage tracing in humans has many further applications both inside the field of oncology and outside. Further study of primary tumor and metastasis will increase the data on the morphology of tumors growing in their host tissue. A large amount of data might allow the deduction of the underlying principles governing the tumor-environment relationship. Lineage tracing can also be applied to the study of other fields such as wound healing or arteriosclerosis. Most importantly, it has the potential to improve our understanding of human developmental biology.

Key words: Mitochondrial DNA, Lineage Tracing, Cycling Temperature Capillary Electrophoresis, Oncogenesis, Metastasis

Résumé

Le tracement de lignage est un concept de fondamental en biologie du développement ainsi que dans l'étude des maladies du développement. L'introduction d'une méthode capable d'identifier et de suivre les lignages cellulaires à l'intérieur de tissu humains bénéficierait l'étude du cancer, de la cicatrisation ainsi que du vieillissement.

Les mutations mitochondriales héréditaires, ont permis l'étude des migrations humaines au travers de l'histoire. Un parallèle existe entre les populations humaines dans le monde et les lignages cellulaires dans un tissu. En faisant ce parallèle, les mutations somatiques mitochondriales, ont été utilisées pour étudier les mécanismes de divisions cellulaires dans certains organes. Dans chaque cellule se trouvent plusieurs copies d'ADN mitochondrial. Une mutations, peut donc être présente dans une, plusieurs, ou toutes les copies d'ADN mitochondriales d'une cellule. Il en suit, que pour étudier les mutations somatiques mitochondriales, il n'est pas seulement nécessaire de les détecter, mais de quantifier les fractions auxquels elles sont présentes. Cette contrainte rend la plupart des techniques de séquençage génétique inutilisables.

Pour permettre l'étude de tracement de lignages en utilisant les mutations somatiques mitochondriales comme marqueur, a fallu développer une méthode quantitative, rapide et avec un haut débit. Une PCR en deux étapes a été mise au point pour, dans un premier temps, amplifier l'ADN mitochondriale dans son ensemble. En un deuxième temps, amplifier une série de petites séquences optimiser pour l'utilisation de électrophorèse capillaire avec température variable (CTCE). Cette méthode permet de détecter et quantifier des mutations présentes dans 76% de l'AND mitochondriale. Des tests de précision ont démontré qu'il est possible de détecter et quantifier les mutations jusqu'à une fraction de 1%. De plus, il est possible d'analyser des échantillons difficiles comme ceux prélever par « Laser capture microdissection »

La première étape dans le tracement de lignages dans un un tissu cancéreux est d'identifier des mutations qui peuvent être utilisées comme marqueur. Pour cela, l'AND extrait de larges coupes au travers de l'échantillon est étudié pour découvrir telles mutations dans le genome mitochondrial. L'analyse est poursuivie dans les tissu ou une mutations a été trouve. Dans un premier essai, 286 échantillons micro-anatomiques ont été prélevés sur une coupe histologique d'un cancer des cellules de Leydig, où 2 mutations de l'AND mitochondrial avaient été détectées. Une forte corrélation a été mesurée entre les deux mutations, ce qui indique qu'il existe un seul lignage. Aucune morphologie n'as pu être reconnue. Pour valider les résultats, 96 échantillons micro-anatomiques ont donc été prélevé sur une coupe histologique successive du même tissu. L'étape suivante a été de prendre plusieurs coupes histologiques successives, et de prélever 96 échantillons sur chacune. Les échantillons ont été place selon une grille 8 × 12, placée, autan que

Résumé

possible, au dessus de la précédente. En analysant de cette façon 19 coupes histologiques, il a été possible d'observer la disposition tridimensionnel d'un lignage à l'intérieur d'un tissu cancéreux. L'observation de ces résultats montre que le lignage suivit n'occupe qu'une minorité du volume tumoral total.

De cette façon, une reconstruction tridimensionnel de lignage cancéreux a été réaliser pour deux autres cancers primaux et 3 métastases. Sous l'hypothèse que les métastases ont pour origine une seule cellule cancéreuse, la présence de la même mutation mitochondrial dans le cancer primaire et la métastase garantie que seule les cellules ayant la mutation sont issue du cancer.

Les résultats montre que la majorité du volume des métastases n'est pas composée de cellules issues du cancer. Il est donc possible de conclure que les métastase grandissent en s'infiltrant dans le tissu environnante ou en recrutant les cellules de l'organe ou elles se trouvent. En même temps, les cellules issues du cancer conditionnent les cellules du tissu environnant pour assumer une morphologie anormal.

L'habilité à tracer des lignages cellulaires a plusieurs applications dans l'études des cancers et ailleurs. Une succession d'études sur des tumeurs primaires et des métastases a le potentiel d'élucider comment les tumeurs interagissent avec leur environnement. C'est une technique qui peut aussi être utilisé pour étudier l'artériosclérose e la cicatrisation. Surtout, c'est une technique avec laquelle il sera possible de mieux comprendre la biologie du développement.

Contents

Acknowledgements	i
Preface	iii
Abstract	v
Résumé	vii
List of figures	xiii
1 Introduction	1
1.1 Motivation	1
1.2 There are many copies of mitochondrial DNA in each cell	1
1.3 Lineage Tracing	2
1.4 Mitochondrial DNA	3
1.5 Tumor Lineages	5
2 Key Points	7
2.1 Homoplasmic mutations are less frequent than expected	7
2.2 Design a reliable primer set	7
2.3 Tumor cells cannot be distinguished by histology alone	8
2.4 Sampling needs to be unbiased	8
2.5 Lineage tracing must be in three dimensions	9
2.6 Following a marker is following a lineage, not all lineages	9
2.7 Tumors and metastatic growth are mostly made up of non-tumor cells	10
2.8 Automate what you can, keep improving your methods	10
3 Methods	11
3.1 Cycling Temperature Capillary Electrophoresis	11
3.2 Primer Design	13
3.2.1 Pre-amps	13
3.2.2 Second Round	13
3.3 Electropherogram Scoring	13
3.4 Laser Capture Microdissection	14

Contents

3.5	Real Time PCR	14
4	First Experiment	17
4.1	Motivation	17
4.2	Experiment	17
4.3	Results	18
4.4	Conclusions and Reflections	18
5	Sample Preparation and Staining	21
5.1	Introduction	21
5.2	Staining protocols	22
5.2.1	SYBR Green	22
5.2.2	Giemza	22
5.2.3	H&E	22
5.2.4	DAPI	22
5.2.5	Ethidium Bromide	22
5.2.6	Pan-cytokeratins	23
5.3	Results	23
5.3.1	Compatibility	23
5.3.2	Observing colon structure	25
5.4	Conclusion	25
6	Detecting Mutations in the Mitochondrial Genome	27
6.1	Resulting Publication: Cycling temperature capillary electrophoresis: A quantitative, fast and inexpensive method to detect mutations in mixed populations of human mitochondrial DNA	27
6.1.1	Short description	27
6.1.2	Significance	28
6.1.3	Personal Contribution	28
6.1.4	Status	28
7	Clinical Applications	53
7.1	Resulting Publication: Scanning the mitochondrial genome for mutations by cycling temperature capillary electrophoresis	53
7.1.1	Short description	53
7.1.2	Significance	54
7.1.3	Personal Contribution	54
7.1.4	Status	54
7.2	Resulting Publication: Intraperitoneal exfoliated tumor cells detected after surgical resection of rectal cancer. A prognostic factor on survival?	54
7.2.1	Short description	54
7.2.2	Significance	55
7.2.3	Personal Contribution	55

7.2.4	Status	55
8	Measuring Mitochondrial DNA Copy Number	57
8.1	Resulting Publication: Quantifying mitochondrial DNA copy number using robust regression to interpret real time PCR results	57
8.1.1	Short description	57
8.1.2	Significance	58
8.1.3	Personal Contribution	58
8.1.4	Status	58
9	Two-dimensional Lineage Tracing	77
9.1	Liver tumors	78
9.2	Resulting Publication:Tracing mitochondrial mutation in tumors	79
9.2.1	Short description	79
9.2.2	Significance	79
9.2.3	Personal Contribution	80
9.2.4	Status	80
10	Three-dimensional Lineage Tracing	109
10.1	Resulting Publication	110
10.1.1	Short description	110
10.1.2	Significance	111
10.1.3	Personal contribution	111
10.1.4	Status	111
11	Conclusion	131
11.1	Future	132
	Bibliography	141

List of Figures

3.1	Schematic representation of heteroduplex formation. In the a solution containing two variants of a double stranded DNA sequence, denaturing the DNA will cause the formation of homoduplexes and heteroduplexes. The heteroduplexes will have a lower melting temperature than the homoduplexes, and can be therefore separated using CTCE	12
3.2	Comparison of electropherograms from two samples analysed on the same fragment. The top one contains no mutation, and a single peak is visible. The bottom sample contains a mutation, and the typical four peak configuration is visible. The first two peaks represent the two possible homoduplexes, whereas the second two peaks represent the two heteroduplexes.	12
3.3	Custom built cap holder. Caps strips are placed on the holder and filled with lysis buffer. Samples fall into the caps by gravity. Only one sample is captured in each cap. The caps can then be placed on 96-wells plates and quickly centrifuged to drop the content into the well.	14
4.1	Example of a LCM sample from the first experiment. This specific sample was considered to be an epithelial sample. The sample is very large, and the thickness of the slice makes the image blurry	18
5.1	Varying thickness of colon tissue, stained with H&E. Thicknesses shown: 5, 10, 15, 20 and 25 μm	24
5.3	Rectal and prostate cancer, 12 μm slices, stained using pan-k antibody and counterstained with SYBR. Pictures taken during LCM experiment, the same area is imaged in both bright-field and in fluorescence, showing the IHC staining as well as the SYBR staining. The shapes visible were drawn using the LCM software for sample acquisition. They represent another attempt at selecting tumor areas.	25
9.1	Results from the two-dimensional analysis of a liver adenoma. The mutant fraction in each sample is represented by fraction of black in each circle. The size and position of each sample taken is accurately represented by the circles.	78

List of Figures

10.1 Current theory of metastasis formation. A single cell leaves a grown tumor and "seeds" a metastasis. If true, this hypothesis means that all tumor cells in a metastasis are descendent of a single tumor cells. Picture taken from the site johnhopkins.edu. 110

1 Introduction

1.1 Motivation

Every organism is, at its beginning a single cell. Development is how this cell divides, differentiates, and ultimately gives rise to a complete organism. Once the organism is born, development is not finished. The organism must grow, develop new function sometimes, and must self-repair, renew itself to continue functioning. Sometimes development fails, resulting in death, or dysfunction in the adult organism. To understand these processes is the goal of developmental biology.

All cells in the human body have a common ancestor. However, at one point during development, a groups of cells acquire characteristics that sets them apart from others, forming lineages. The characteristic that identifies a cell lineage and sets it apart from others is a marker. Understanding the behaviour of cell lineages as they grow to form tissues, is lineage tracing.

Lineage tracing requires the ability to detect the presence of a marker in a tissue, and then identifying its distribution through the tissue volume. Somatic mitochondrial mutations are potentially a very good marker for lineage tracing. Distinguishing the cells that carry a mutation from the others, a cell lineage can be defined. The objective of this thesis was to develop the methods to use somatic mtDNA mutations as a marker for lineage tracing. This requires first, the ability to detect the presence of mutations in a tissue volume. Second, to follow the mutation through the volume and identify areas that are part of the same lineage. Once developed, these methods will enable lineage tracing in human tissues. In doing so, improve scientific understanding of human developmental biology and carcinogenesis.

1.2 There are many copies of mitochondrial DNA in each cell

The core proteins of the oxidative phosphorylation chain are synthesised inside the mitochondria, and encoded in the mitochondrial DNA. Inside each cell, there is a large number of copies of

the mitochondrial DNA. The number seems to vary according to the specific needs of cells. For example, cardiac muscle cells have roughly 7000 mtDNA copies per cell, twice as many copies as skeletal muscle cells [1]. Since there are many copies of mtDNA in each cell, a mitochondrial mutation can be present in only one copy, all copies, or any fraction in between. A mutation present in all the copies of mtDNA in a cell is called homoplasmic. A mutation that is present in a fraction of the mtDNA copies in a cell is called heteroplasmic. There are few reliable estimates of mtDNA copy number in different cell types in humans and the turnover process of mtDNA in cells is poorly understood [2, 3]. Consequently the process through which mtDNA mutations arise and reach (or don't) homoplasmic levels is subject to speculation [4, 5, 2, 6].

1.3 Lineage Tracing

Lineage tracing is the study of the ancestry of the cells composing a multi-cellular organism. The current understanding of biology is that a single cell gives rise, via multiple divisions and differentiation steps, to a whole organ, organism or tumor. To investigate this process we require lineage markers, which are retained in the lineage of an original cell, and can be followed, or traced back. The earliest lineage marker used was a form of ink [7], which could be injected directly in cells of a developing embryo. If the embryo is transparent, the ink can be seen to propagate as the cells divide [7, 8, 9]. This technique has as a first drawback, that ink concentration decreases as cells divide, resulting in the signal being lost after some generations of cell division [8]. The second disadvantage is that it requires the embryo to be transparent to be followed. The organisms to which it can be applied are limited, and include the *Xenopus* [7] and the Zebra-fish. From the experiments with ink, the importance of developing perennial markers, that are not diluted as the cells divide, became apparent. The chimeric organisms were developed to answer that need. Chimeras were made using similar species of which the cells could be differentiated under the microscope. One of the best-known historical examples is the Chick-Quail chimera realized by Waddington [10]. Other examples include mouse-rat chimera [11, 12]. In these experiments, the different cells from different organism "randomly" distribute over the lineages of development. The observation of the born animal can then allow the determinations of common ancestry [8]. Among the major disadvantages of this technique are: the random distribution of cells in the lineages and the difficulties involved in making chimeras. With the development of fluorescent proteins (such as GFP), genetically inserted markers rapidly took the center stage. The expression of fluorescent proteins can be induced in a transgenic organism. Better understanding of molecular biology has given researchers the possibility of putting the expression of GFP under the control of specific promoters [13, 14]. Lineage tracing is then done by assuming that the activity of a transcription factor marks specifically a single cell lineage, which is often a strong assumption. Consequently ways to increase the selectivity of the marker's expression have been developed. A powerful tool is the Cre-Lox system. Cre recombinase has the ability of recombining together two Lox sites [9, 15, 16, 17]. The activity of Cre, can itself be put under the simultaneous control of a transcription factor and an external stimuli. External stimuli include tamoxifen or estrogen [18]. Many comprehensive reviews on

the subject exist (e.g. Kretzschmar et al. [8]). The important thing about the inducible Cre-Lox system for gene expression is that it will cause the marker to be constitutively expressed in all cells, that at the time of external activation, expressed the transcription factor under which the Cre recombinase expression was put under [9, 19]. This constitutive expression will then be passed on to all the progeny of that cell. Thus selectively marking the lineage. The descendants of the marked lineages can then be seen under fluorescence microscopy.

Markers mentioned up to now are an evolution of ink. They all rely on inserting in a group of cells of interest, something that can be seen later on. It is also possible to decide to follow a marker that arises spontaneously in cells. These "internal" markers present the great advantage of being applicable to human studies. Most internal markers are genetic markers. A lot of the development of internal markers has been motivated by the study of populations. Tracking populations throughout history is not so different from following stem cells throughout organogenesis. In both cases, the only available information is the final state, and in both cases there is a common ancestor a certain number of divisions back. An example of a successful marker is X chromosome inactivation in women heterozygous for the glucose-6-phosphate dehydrogenase (G6PD) gene. Two variants of the gene, both active, that produce G6PD type A and B respectively exist. The two protein variants of the enzyme can be easily distinguished by electrophoresis [20, 21, 22]. Tissues that are clonal, will have only one type of the protein, whereas if there is a polyclonal origin, the probability of having only one variant is small. This technique has been used successfully to establish the clonal origin of atherosclerotic plaque [22]. It does however have its limitations, the X chromosome carrying the B variant of G6PD is rare and found general only in women of african origin. Furthermore, it is not applicable in men. Thus limiting greatly its applicability to tissue from the general population.

1.4 Mitochondrial DNA

Mitochondrial DNA (mtDNA) is composed of a single circular chromosome of 16569 base pairs (hg38, GRCh38, Dec. 2013). It encodes for 22 tRNA, 13 protein subunits and two ribosomal RNA subunits. In each cell, many copies of the mtDNA are present. The number per cell is believed to vary according to cell type. Estimates of mtDNA copy number vary between a few hundreds to several thousands, also depending on the cell type in question [23, 24, 4](with highest recorded numbers of 8000 for cardiomyocytes) Mitochondrial DNA encodes the core proteins of the oxidative phosphorylation chain, as well as the tRNA and ribosome subunits required to synthesise them. Mitochondrion are supposed to be responsible for most of the intracellular ATP production [25, 26]. The study of mitochondrial DNA is key in understanding the principles of cellular metabolism, as it is responsible for a disproportionate part of the energy production for its length. Also, mtDNA is involved in mitochondrial diseases, and is thought to be involved in the process of ageing, as well as tumor development and possibly Alzheimer [27, 28, 29]. The importance of mtDNA can be seen in the cases where it is dysfunctional: Mitochondrial diseases. Mitochondrial diseases are diseases due to mitochondrial dysfunction, most often due to mutations in mtDNA. They have an estimated prevalence in the order of 1 person in 10 000. The

Chapter 1. Introduction

severity of the symptoms is dependent on the affected tissue and degree of heteroplasmy of the deleterious mutation [30, 31, 32, 33]. The diagnosis of mitochondrial mutation is often difficult. An individual affected by a mitochondrial mutation, can have the mutation present at a very low fraction throughout the body, and only reaching a mutant fraction high enough to cause harm in a specific tissue. This will make diagnosis through sequencing of mtDNA taken from blood particularly difficult. Furthermore, the detection of the mutation does not characterise the disease. Mitochondrial diseases often affect muscle and nerve tissue, examples are MERF and LHON [33, 31, 30, 34, 35].

mtDNA has been a very successful research tool to understand the patterns of human migration throughout the world [36, 37, 38]. Being maternal inherited only, and having a mutation rate slightly higher than nuclear DNA, it provides an ideal tool for the purpose. Understanding human migration involves determining the time in history when the last common ancestor happened between two human populations. Then reconstructing the migration patterns by filling in the gaps with genetic information from archeological remains. The main difficulty when using nuclear DNA to study human populations, is in the process of mixing that happens at each generation. Only one out of each pair of chromosomes is transmitted from parent to children, and each time a chromosome is not passed on, the line that it identifies becomes broken. Furthermore, there are the possibilities of crossover. mtDNA does not suffer from this drawback, and unbroken maternal lines of mtDNA can be traced back. Mutation rates per generation are also an important factor in the suitability to be a marker. Too low a mutation rate, will give too little temporal resolution, and separation in the populations will not be localisable in history. Whereas too high a mutate rate will give very good resolution, but will prevent the history to be extrapolated far back in time. High mutation rate means that the probability of identical independent mutations becomes too high for lineage to be uniquely identifiable. mtDNA is estimated to acquire between 0.3 and 1 mutation in the whole genome, each generation [39, 40]. Under the assumption that mutation rates are uniformly distributed, the probability of an identical, independent mutation is in the order of 10^{-4} . If generation times are in the order of 20 years, the hypothesis can be made that they are unlikely in the period of 50 000 to 100 000 years of human history that are of interest for the study of migrations. A further advantage is that with this high mutation rate, separations between populations can be localised in history with the precision of a few generations. Lineage tracing in developmental biology essentially works on the same principles as the study of human populations. Differences are found and last common ancestors identified. For very much the same reasons as for human lineages, mtDNA has proven an effective tool for lineage tracing inside human tissues. In most cases, humans are born with only one variant of mtDNA. Thus all the differences present can be hypothesised as having arisen due to somatic mutations. There are currently no reliable estimates of the mutation rate per cell division for the mtDNA. Measurements have been performed in cell lineages [41, 5, 42, 43]. They are either limited or difficult to translate to human tissue. However, it does seem to be low enough to justify the assumption of no independent identical mutation arising at a high fraction. The first problem faced when using mtDNA mutation as a marker for clonality is the mutant fraction. The presence of many copies of mtDNA in a cell, mean that a quantitative measure of the mutant fraction is

required to identify a cell clone. A successful example is the use of cytochrome-C-oxidase (COX) gene inactivation mutations. If more than 90% of the mtDNA in a cell carry a gene inactivating mutation on the COX gene, it will lead to an accumulation of cytochrome C that can be detected by IHC [44, 45, 6, 46, 4]. With this method, it has been possible to identify colon crypts as a single clonal unit [4], as well as to study the turnover patterns of human airway epithelium. The main advantage of this technique, is that it can be used directly on human tissue, and requires no "mutation discovery" a-priori. It does have a limitation in that COX gene inactivating mutations are quite rare, requiring the study of many tissues. Furthermore, no visible mutation can be defined as "neutral" and the question arises whether selective pressure is applied for or against it.

1.5 Tumor Lineages

Oncogenesis, the process through which cancers arise is still poorly understood. There seems to be a general consensus that cancers are a clonal disease, and therefore arise from a single cell [47, 48, 49, 50, 51, 52]. The idea could be traced originally to Boveri at the beginning of the 20th century. The idea then gained ground with the modelling and theories of Armitage and Doll [53, 54], according to which, the age specific incidence could be explained by a multi stage disease progression model. Using the X chromosome inactivation method, the clonal origin of Burkitt's cell lymphoma, myeloid leukaemia was established [55]. In most carcinomas, evidence to support a clonal origin of tumors seem scarce. Studies of cervical cancer, were inconclusive, and encountered the difficulty of being unable to select cancer cells from the tumor mass [20]. Clonal origin was claimed for head and neck tumor [56], although the study also suffers from difficulty in quantifying the fraction of tumor cells. Even though tumors are supposed to have a clonal origin, they form heterogeneous structures [50, 57, 58, 59]. Histologic analysis of tumors shows that they are composed of different tissue type with cells of different morphology [57]. Also, they form very heterogeneous populations [60], with tumors of the similar tissue, showing marked differences. The observation of heterogeneity is not a contradiction to clonal origin. It does however require either differentiation of tumor lineages during oncogenesis, or the recruitment of different cell types from the host tissue. If tumor lineages could be traced, then it would be possible to identify which portions are tumor derived, and thus what differentiation do tumor cells undergo, as well as which cells are recruited from the host tissue. Cancer lineage tracing has been investigated in the mouse using a wide variety of external markers. Example are fluorescent proteins inserted using recombinant viruses [58], or activated with the Cre-lox system [8, 15, 61, 62]. Since cancer is extremely rare in the mouse, most of the mouse cancer studies rely on the induction, or insertion of a tumor. It is possible to graft mice with a human tumor, which might have been labeled beforehand [63, 64, 65, 59]. Alternatively, many strategies have been proposed to induce a tumor in the mouse. Some recently used are the modification of gene expression, such as TP53 [66, 67, 15] or other tumor suppressor genes. Often activated using Cre-lox system [15, 8], or under the regulation of a particular transcription factor. A tumor often arises from these manipulations, but the similarity of these "induced mouse tumours" with a normally occurring human tumor is not clear.

Chapter 1. Introduction

Internal markers are most often genetic markers [68, 2, 69, 70]. They present the advantage of being applicable to the study of human tumors. The most successful is again the use of X chromosome inactivation [71, 55, 72, 73, 56, 74, 52]. Internal marker rely on the genetic heterogeneity intra-tumor or or between the tumor and its surrounding tissue. Genetic heterogeneity can be of many different types, such as allelic instability[47], chromosomal aberrations [51], single nucleotide polymorphism [75]. somatic mtDNA mutations have also, to a more limited extent been used to trace cell lineages in tumors [76, 77]. Although, to my knowledge, no extensive study using the labelling of COX inactivating mutation has been attempted in tumors.

Taken together, these results point to a general misunderstanding of the nature of solid tumors. There does not seem to be many methods that allow the direct study of human tumor tissue. The two most accurate ones, X chromosome inactivation and labelling of COX inactivation, very successful in the study of normal tissue, have had their difficulties being used on human tumors. Possible explanations might be that the X chromosome inactivation method can only be used on a small fraction of women from African descent. Whereas COX inactivation requires the presence of a very rare set of mtDNA mutations to occur at a high fraction. There does not seem to be a clear demonstration of the clonal origin of carcinomas in humans. In fact, there does not seem to be a clear consensus of what constitutes tumor tissue. Which was the main difficulty faced by the studies that used X chromosome inactivation to test the clonality of tumors. Any successful method, to analyse tumor lineages, therefore has to work under the assumption that it is impossible to select tumor areas *a-priori*. It requires the ability to analyse sufficiently small samples, in sufficient numbers, to allow systematic sampling of a tissue volume. This way, the patterns of tumor lineages could be reconstructed directly from the acquired data, without requiring a histological analysis of the sample. In doing so, it would be possible to reconstruct the patterns of clonal growth in tumor volume.

2 Key Points

2.1 Homoplasmic mutations are less frequent than expected

The project started with the idea that carcinomas, were composed of two distinct parts, a mesenchymal and an epithelial part [78, 66]. The question still existed whether the mesenchymal portion is related to the tumor cells, or a reaction of the surrounding "normal" tissue. Reviewing the literature, we made the hypothesis that roughly 80% of all tumors would carry a mitochondrial mutation at a homoplasmic, or near-homoplasmic level [79, 80, 81, 76]. If a tumor is a clonal group of cells, then a mtDNA mutation present at the tumor origin would be present in all tumor cells. Taking samples across a tumor sample would therefore detect mutations with a high frequency. The first experiment performed, attempted to determine whether the epithelial and mesenchymal portion of a tumor were clonal related. To do so, 8 large LCM samples were selected based on histological examination, for 10 tumors. Using a first round amplification from the literature and the first CTCE primer set of our design, 60% of the mtDNA was scanned across those samples. Results were inconclusive, as no mutations were detected in the tumor samples. Not only had the initial question not been answered, but the observation frequently reported in the literature had not been reproduced. If homoplasmic mutations were common in human tumors, how come we had found none? One possibility is that homoplasmic mtDNA mutations are rarer than initially supposed. One of the possible explanations is that most published results, are measured in mixed samples, in which the tumor fraction is estimated by histological observation.

2.2 Design a reliable primer set

The first primer set used [82], to amplify mtDNA was unreliable on LCM samples. DNA extracted from laser capture micro-dissection (LCM) samples is low in concentration, and sometimes difficult to amplify. We therefore needed to develop a new primer set that would prove more reliable. From our experience in primer design, a length of around 800bp seemed optimal. It was also important that the PCR result could be checked using capillary electrophoresis, instead of gel electrophoresis. This is when the idea came of placing a tail at the end of each primer. The

tail can then be amplified by a highly efficient "tail primer", which could also be labeled with a fluorophore for detection using capillary electrophoresis. The second round primer set was then modified and augmented to amplify on the new set of pre-amp and achieve a higher coverage. In the end we had a reliable set of primers, with less than 1% failure rate, that could be used on DNA extracted from LCM samples.

2.3 Tumor cells cannot be distinguished by histology alone

Trying to understand what could be defined as "tumor cells," epithelial and mesenchymal portions within a tumor, we were able to identify no consensus. A problem that had already been identified by Fialkow. Philip J. Fialkow in the 60's and 70's was trying to do something strikingly similar to us, and encountered similar problems. He was trying to demonstrate the clonal origin of tumors, as well as the origin of the different tissues present in a tumor mass [83, 55, 84, 72, 73, 20, 56, 85, 86, 87, 88]. He also came up against the difficulty of defining which parts of a tissue was tumor and which was not. His results from having multiple independent pathological examination of the same samples, seem to suggest that no clear consensus exists. In fact, there does not seem to be a way to define, based purely on morphology, which cells are tumor in origin and which are not, even when they all are part of the "anomalous mass" that constitutes the tumor. Imuno-histochemistry (IHC) does not offer an adequate solution. If it is impossible to define which areas in a tumor sample, are tumor in origin, and which are normal in origin, then any attempt at selecting structures based on morphology is bound to fail. This was one of the possible explanations for our earlier failures to identify mutations. Our observations also seem to point in the direction that there is no clear understanding of what tumor cells are. In the chapter 5, the difficulties of imaging tumor tissues are described. Trying different types of staining we were unable to obtain specific tumor staining or to recognise clear structures within the tumors. Furthermore the results from 2 dimensional lineage tracing, shown in chapter 9, demonstrate that the mtDNA mutations used as markers can be found across a slice of tumor tissue without apparent correlation to tissue morphology, further strengthening the hypothesis that in a tumor mass, cells of tumor origin cannot be distinguished by histology alone.

2.4 Sampling needs to be unbiased

There is no method that allows us to distinguish between tumor cells and non-tumor derived cells simply through observation. Any attempt to select structures when picking LCM samples is therefore bound to introduce biases. Tests with the LCM equipment were therefore performed to determine the size of sample that could be consistently taken: a circle of around $25000\mu m^2$. A standard 8×12 grid was then designed with 96 such samples. Sampling a tumor slice was then conducted by placing these grids and taking samples close by. An example of such a sampling pattern can be seen in figure 9.1, in which a large portion of the slice is covered in identical samples taken in a grid formation. Observation of these scans revealed that the mutations seem to be distributed across the surface without a discernible correlation with observable structures. The

2.5. Lineage tracing must be in three dimensions

test that ultimately verified that indeed we were tracing lineages, was the ability to reproduce the results in another slice. These results can be seen in the chapter 9 and form the main results presented in the paper described in section 9.2. These results allowed us to conclude that any attempt to select samples based on tissue morphology is going to be of little benefit and likely to introduce biases. It is therefore important to have a systematic sampling of the tissue analysed.

2.5 Lineage tracing must be in three dimensions

The paper described in section 9.2 shows that it is possible to find the same mutations in a separate slice of tissue, in similar position. The next logical step, is to sample consecutive slices. Sampling the same location in consecutive slices of the tissues, it would be possible to reconstruct the three dimensional patterns of cell lineages in human tumors. The first attempt was the Leydig cell tumor, which had already been subject to two dimensional scanning. At first all consecutive slices were taken, sampling and analysing them, we realised that the sampling in the z (depth) axis, was too fine in comparison to the sampling in the x and y directions. So after taking 12 consecutive slices of tissue, only one every 3 was placed on membrane. The result of the experiment was a success. Once the analysis was done and the first three dimensional plot of mutations made, we observed that there was a strong correlation of mutations in the z axis as well. Having proven it was possible, we decided to make more three dimensional analysis of human tumors, as described in chapter 10. They confirm that markers followed seem to form three dimensional structures that would not be recognised only by a two dimensional analysis. This confirms that to trace a cell lineage through a volume of tissue, a three dimensional study of the tissue is needed.

2.6 Following a marker is following a lineage, not all lineages

The first three dimensional analysis of tumor tissue for mtDNA mutations proved that it was possible to follow a tumor lineage throughout a volume of tissue. The presence of the marker guarantees tumor origin (if the basic assumptions of oncogenesis are true). The absence of a marker does not guarantee "non-tumor" origin. The mutation could have arisen at the beginning of the tumor development, and be present in the first tumor cell that gave origin to all the successive ones. In this case, cells without the mutation would be derived from non-tumor cells. In the other case, mutation identified and followed could have arisen at a later stage of oncogenesis, and only mark one of the many tumor lineages. The problem was solved by finding tumors and the corresponding metastasis carrying the same mutation. If the metastasis is started by only one, or a small number of tumor cells, then the presence of the same lineage as the primary tumor, guarantees that the original cell was carrying the mutation. Consequently, all tumor derived cells in the metastasis are identifiable by the presence of the mutation. The three dimensional analysis of tumor cell lineages in a metastasis is the reconstruction of the patterns of metastatic growth in the host tissue.

2.7 Tumors and metastatic growth are mostly made up of non-tumor cells

What is left to do is still very much work in progress, understanding the results. The observed patterns do not seem to validate any known model of tumor genesis. If tumors are clonal, then the observed results seem to indicate that the majority of the tumor volume, is not composed of tumor derived cells. The same is true for metastasis. This would contradict the beliefs that tumor grow as a cell colony, by expanding and pushing the remaining cells out. It would seem that very few tumor cells diffuse through the tissues in which they grow and re-condition it. The process and mechanisms involved in this process are not yet understood.

2.8 Automate what you can, keep improving your methods

Possibly the main reason the work presented here was possible, was that all the methods used, and analysis performed were improved and automated along the way. This shared belief between my supervisors and myself, as well as the dedication to optimise each step, proved crucial in the increase in throughput we achieved. For example, a software made to design the primers, which allowed a very large number of primer pairs to be designed, and then optimised, thus enabling us to reach the coverage we have. When the LCM capture system was becoming the limiting factor in the analysis, Per Ekstrom designed and built a collector to increase the throughput, from 100 samples a day to 800. Then came the problem of analysing the electropherogram, for which an automated solution was also designed. Towards the end of the work, the combination of all the improvements made, allowed us to make a complete three dimensional analysis of a tissue in one week, from mutation discovery, microtome, sampling, PCR, CTCE and data analysis.

3 Methods

3.1 Cycling Temperature Capillary Electrophoresis

Cycling temperature capillary electrophoresis (CTCE) is a method for detecting and quantifying DNA variations within a sample. DNA is a double stranded helix where each strand is complementary to the other. If a sample containing two variants is denatured (i.e. the double strands are separated into single-strands), and then allowed to anneal (i.e. reform the double-strand), homoduplexes and heteroduplexes will form, see figure 3.1. Homoduplexes and heteroduplexes have different melting temperature. CTCE exploits this difference in melting behaviour to separate them and quantify the relative fraction. The melting behaviour of double stranded DNA affects its migration speed through a gel filled capillary. Electrophoresis is used to force DNA migration. Since the charge of DNA is only dependent on its length, it is equal for variants of the same length. The resistance of the gel to the migration depends on the melting status. Coupling temperature variations with capillary electrophoresis, it is possible to force the different fragments of double stranded DNA to migrate at different speeds through the capillary. Since heteroduplexes have a lower melting temperature, they will spend more time in an open configuration than homoduplexes, and cause them to move slower. The separation allows for the relative quantity to be measured, which leads to the mutant fraction.

For CTCE to be effectively used, DNA fragments need to respect three conditions:

1. Length smaller than 180 base pairs
2. Monotonous (only increasing or decreasing) melting profile as a function of position.
3. A GC clamp at the warmer (presence ensured by condition 2) end of the fragment that will prevent complete separation of the strands of DNA analysed.

To separate mtDNA into fragments satisfying these conditions, a specific set of primer has been designed. The detailed description of the design and PCR conditions are described in section 6.1.

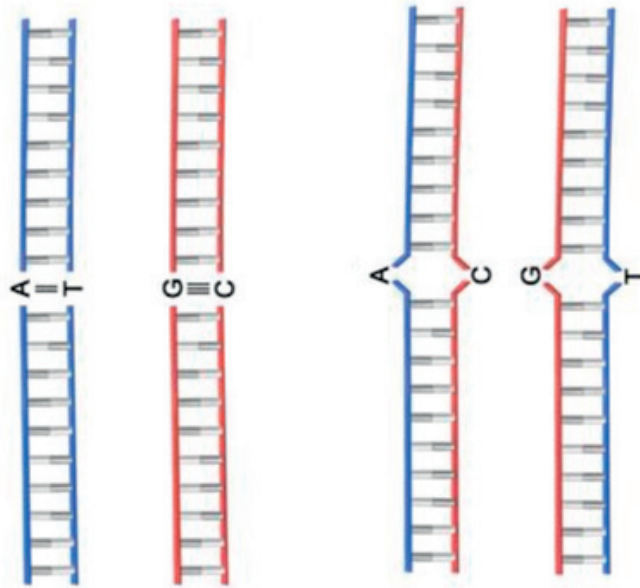


Figure 3.1: Schematic representation of heteroduplex formation. In a solution containing two variants of a double stranded DNA sequence, denaturing the DNA will cause the formation of homoduplexes and heteroduplexes. The heteroduplexes will have a lower melting temperature than the homoduplexes, and can be therefore separated using CTCE

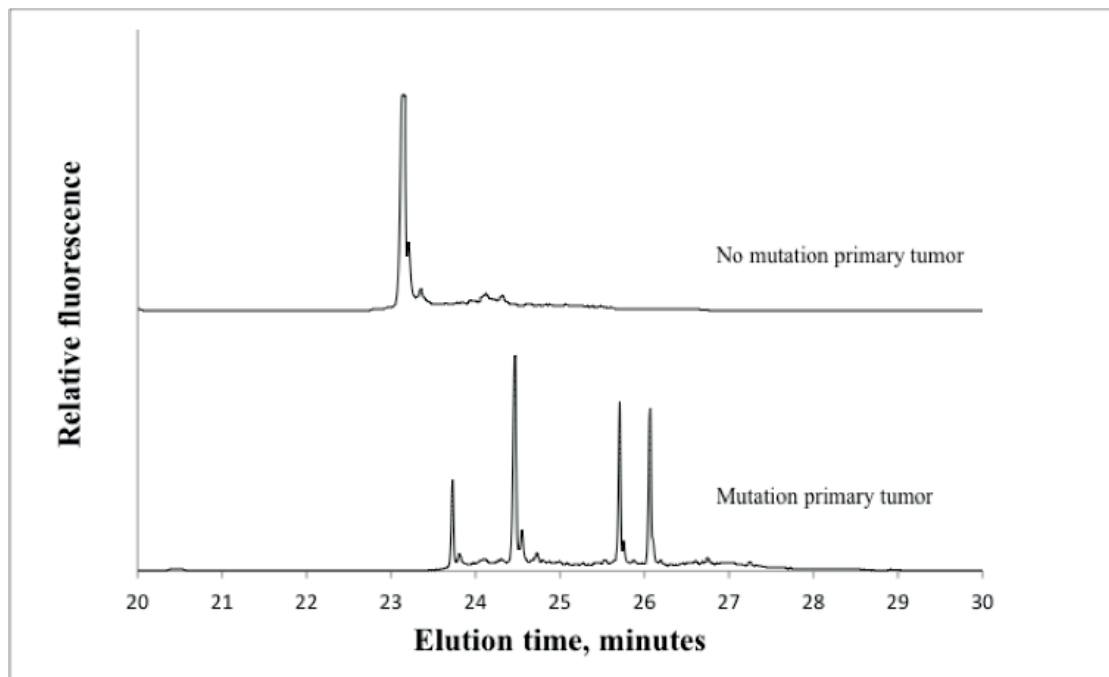


Figure 3.2: Comparison of electropherograms from two samples analysed on the same fragment. The top one contains no mutation, and a single peak is visible. The bottom sample contains a mutation, and the typical four peak configuration is visible. The first two peaks represent the two possible homoduplexes, whereas the second two peaks represent the two heteroduplexes.

3.2 Primer Design

3.2.1 Pre-amps

To selectively amplify the mtDNA without amplifying the nuclear genome, a set of 23 primer pairs was designed. They amplify overlapping fragments of lengths between 421 to 1030 base pairs. Together these fragments amplify with a slight overlap the whole mtDNA. Designing the fragments required the selection of areas with mismatches between the mtDNA and NUMT. Using BLAST (ncbi) the mtDNA was aligned to the human nuclear DNA. Manually inspecting the alignment results, areas with as many mismatches as possible were selected for as primer target. To check for PCR success, as well as to increase PCR efficiency, a 3-primer strategy was used. The idea is to add a short tail at the 3' end of each primer. The PCR is then performed with the forward and reverse primer, as well as with a third primer that amplifies on the tail. The tail primer is highly efficient, and can be labeled with a fluorophore. Labelling the tail primer, allows the PCR success to be checked using capillary electrophoresis instead of gel electrophoresis. Which highly increases the throughput.

3.2.2 Second Round

The second round fragments are amplified on the product of the first round PCR. There is therefore no risk of nuclear amplification even in the case of compatible primers. The second round fragments need to satisfy the requirements for efficient separation of mutants by CTCE. An automated algorithm was designed to select the fragments over the pre-amps. The output was a set of primers that could be amplified over each of the pre-amps. A clamp "GC-clamp" is attached to the primer that amplifies on the side of the fragment with the highest average melting temperature. PCR conditions were then optimised for each of the designed primer pairs. Primers that failed to amplify were shifted a couple of base pairs up or down on the pre-amp, and optimised again. The result was a set of 192 primers that covers 76% of the mtDNA.

3.3 Electropherogram Scoring

The output of CTCE is an electropherogram, a typical example of which can be seen in figure 3.2. As the figure shows, the presence of a mutation can be readily identified by the presence of multiple peaks. In the case of a single peak, an internal standard is needed to determine which allele is the one present. The mutant fraction can be estimated by measuring the relative peak heights, or areas (height and area are related). Manually identifying and measuring the peaks is a laborious process. A user interface has therefore been developed using Matlab. The interface also contains various functions for the automatic recognition and measurement of the peaks. It also allows manual selection of peaks after automated scoring. The development of automated tools, has dropped the analysis time by over 20-fold, thus making it possible to analyse the results from three dimensional tissue sampling.

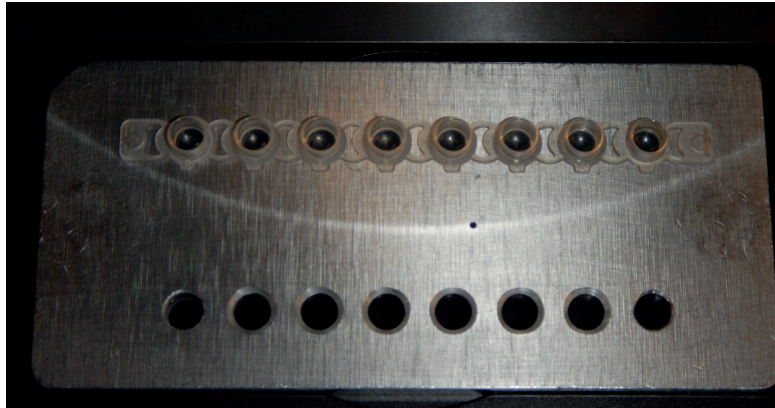


Figure 3.3: Custom built cap holder. Caps strips are placed on the holder and filled with lysis buffer. Samples fall into the caps by gravity. Only one sample is captured in each cap. The caps can then be placed on 96-wells plates and quickly centrifuged to drop the content into the well.

3.4 Laser Capture Microdissection

The objective of Laser Capture Microdissection (LCM) is to combine tissue capture with microscope imaging. The equipment used throughout for all the experiments mentioned here was a Leica LDM 6000. Samples are placed on thin PET membranes that can be cut by the laser. The Leica software allows for shapes to be drawn on the microscope camera output. The laser will then cut the shapes, cutting through both samples and membrane, which will drop by gravity into collector caps. To increase the throughput, we built our own cap holder, shown in figure 3.3. A holder that could be fitted with two 8-caps strips at a time. Which allowed for the collection of 16 samples with only one loading. The strips were filled with lysis buffer, and after collection could be directly transferred to a 96-well plate for incubation.

3.5 Real Time PCR

Real time PCR is considered a standard way to quantify DNA copy number. Coupled with reverse transcription it can also be used to quantify mRNA copy number, which relates to gene expression. The principle of the method is to perform a PCR in the presence of a fluorophore that is fluorescent when bound to double-stranded DNA. Fluorescent level is then measured after each PCR cycle. The output is a measure of DNA amount in the sample at each cycle, which can be used to estimate the relative copy number between two samples. The details can be found in the paper described in section 8.1.

There are various theories as to how the output of rtPCR should be analysed. The ΔC_T method is generally considered to be the standard. However, it still has major drawbacks. For example requires identical PCR efficiency between two samples being compared, which means that the PCR efficiency and fragments must be optimised. It has the considerable advantage of being implemented in a fully automated form in most software for rtPCR analysis, whereas most other

3.5. Real Time PCR

methods rely on the isolation of the exponential portion of the rtPCR signal. The alternative proposed is to use a robust regression to isolate the exponential portion of the signal. This allows the efficiency of the rtPCR to be estimated, and removes the requirement that two samples compared have the same efficiency. It also permits the measurement of PCR efficiency without cumbersome dilution curves, which could also introduce bias in the measurement.

4 First Experiment

4.1 Motivation

The project started with the idea that carcinomas, were composed of two distinct parts, a mesenchymal and an epithelial part [78, 66]. The question still existed whether the mesenchymal portion was related to the tumor cells, or was a reaction of the surrounding "normal" tissue. Reviewing the literature, we made the hypothesis that roughly 80% of all tumors would carry a mitochondrial mutation at a homoplasmic, or near-homoplasmic level [79, 82, 80, 81, 76]. If a tumor is a clonal group of cells, then a mtDNA mutation present at the tumor origin, would be present in all tumor cells. Taking tumors, and sampling the epithelial and mesenchymal portion of them, scanning the mtDNA for mutations, it would be possible to answer the initial question.

4.2 Experiment

We took 10 samples of colon tumors from a bio-bank, for which either normal tissue (taken from the colon a few centimetres away from the tumor) or blood from the same patient was available. The tumor samples were cut in $50\mu\text{m}$ slices and placed on membranes for LCM. The reason for a $50\mu\text{m}$ cut was to have at least one cell layer on the membrane, therefore guaranteeing DNA for the analysis. With the help of a pathologist from the Norwegian Radiumhospital, three samples that could be identified as mesenchymal and four samples that could be identified as epithelial were chosen for each tumor, an example of which can be seen in figure 4.1. In order to detect the presence of a mutation present in the whole tumor, an 8th sample was constructed for each tissue, namely an even mix of DNA extracted from each tumor sample and DNA extracted from normal tissue (or blood). We considered the possibility that all samples taken in the tumor area would carry a mutation at a homoplasmic level. To identify it as a mutation, comparison was needed against normal tissue or blood, thus motivating the construction of the 8th sample.

Scanning the mitochondrial DNA requires its selective amplification. I then had the idea of looking for a set of primer to amplify the mtDNA and then design the CTCE primers to amplify



Figure 4.1: Example of a LCM sample from the first experiment. This specific sample was considered to be an epithelial sample. The sample is very large, and the thickness of the slice makes the image blurry

on the product of the first round. The idea was not novel, and a primer set was found in the literature, that was guaranteed to amplify only the mtDNA [82]. The second round primers were designed automatically, using a software of my design, and the POLAND algorithm for the melting profile of a DNA fragment. After test and optimisation, roughly 60% of the mtDNA could be scanned for mutations using this initial set-up. At the time, we still held onto the belief that, frequently enough, a tumor is a bulk of clonal cells which regularly contains a homoplasmic mutation present in the whole of its mass. In which case, we should observe no mutation in the tumor samples, nor in the normal tissue. A mutation should however, be detected in the mixed sample of tumor and normal tissue.

4.3 Results

Results were inconclusive. No mutations were identified in the mixed samples, and only two mutations were detected in the individual LCM tumor samples. Not only had the initial question not been answered, but the observation frequently reported in the literature had not been reproduced. If homoplasmic mutations were common in human tumors, how come we had found none?

4.4 Conclusions and Reflections

The first observation was the relative high failure rate of PCR, one of the 9 first round reactions failed and only about 60% of the mtDNA was scanned. The PCR on extracted DNA was more consistent, but there was a reasonable frequent failure in the LCM samples which raised questions about the reliability of the primers to amplify in difficult situations. An explanation was that the fragments were too long, which made it too susceptible to variations we could not control. The results of the experiment were inconclusive. No mutations were found in the

4.4. Conclusions and Reflections

LCMD sample, or in the mixed samples. Having scanned 80 samples, and found no mutations, forced a re-evaluation of the method used to detect them, as well as our assumptions about mutation rates in tumors and normal tissue. Furthermore, it was clear that the distinction between epithelial and mesenchymal portion of a tumor was not obvious, as could be inferred from the literature. There does not even seem to be an easy way to define what constitutes tumor cells when histologically analysing a sample, it is at all possible. Figure 4.1 shows an example of an epithelial sample taken for this experiment. The preparation of the sample was delegated to the histology department of the Radiumhospital, which made the sectioning and staining. As is shown in figure 4.1, the sample is difficult to interpret, as the contrast is poor. It was then clear that better sample preparation was therefore needed. We still believed that somehow it was possible to distinguish tumor cells from non-tumor cells in a sample just by observation. It was then decided to attempt various experiments with histology. Then came the question of how frequent are mtDNA mutations in tumors. A better review of the literature available of mtDNA in tumors revealed a few inconsistencies in the results. Most of the claims of homoplasmy, relied on a pathological examination [79]. Furthermore, they were performed with methods that would not allow the quantification of the mutant fraction in a sample [89, 79]. In order to have a good estimate of the mutational spectra in human tumors, the experiment would have to be done "in-house." There was a high degree of failure in this experiment. One of the reasons seemed to be the high rate of PCR failure of the first round. A better primer set for first round PCR would need to be found.

5 Sample Preparation and Staining

5.1 Introduction

To be imaged with a microscope, tissues need to be stained. A concept that goes back to the beginning of microscopy and analysis of tissues with it. If the first cells were observed in 1665 by Robert Hooke, the first documented stain for tissues can be found less than a decade later, in 1673 by Antonie van Leeuwenhoek. Van Leeuwenhoek used saffron extract as a dye, later experimented with other known pigments such as carmine and madder. Even so, some attribute the first histological stain to Joseph von Gerlach, who in 1858 successfully used ammoniacal carmine to stain cerebellum. The objective of dyeing tissues is to create contrast. Some parts of cells will absorb some dyes, but not others. The unabsorbed dye can then be washed away leaving only some areas stained, thus forming patterns. Another consequence of differential dye absorption is the possibility of counter-staining, or staining with more than one dye. Since different stains attach to different chemical compounds, the staining choice will greatly impact what can be observed on the sample. Consequently, many applications have their own traditional stain.

One of the main shortcomings described in chapter 4 was the difficulty to interpret the images acquired of the samples. In order to improve the quality of our imaging, a variety of sample preparation protocols had to be tested. Also, if it was possible to reliably distinguish tumor cells and therefore select them, we needed to know. A variety of staining protocols were therefore experimented with, and applied to both normal and tumor tissue. Since different applications require different staining, it was important for us to develop a stain that suited our purpose.

5.2 Staining protocols

5.2.1 SYBR Green

SYBR green is a DNA marker, used mostly gel electrophoresis and for rtPCR. It binds to double stranded DNA and has a fluorescence peak around 520nm. It can therefore be visualised with a green fluorescent protein type microscope configuration. its main advantage is being easy to use, and supposedly specific for DNA. Fresh frozen samples were placed on LCM membranes or glass slides, then fixed by immersing in a 50% methanol and 50% acetone mix for 10min then air-dried. Staining with SYBR was obtained by immersing samples for 5 min in a 2X SYBR mix (diluting the original 10 000X SYBR 5000 times in water). SYBR staining is also possible in combination with IHC.

5.2.2 Giemza

Giemza is an traditional way of staining. Currently in use as standard in laboratories, and is also considered the reference technique for diagnosing the presence of the *Plasmodium* parasite[90]. Slides are fixed by 10min immersion in methanol, then air-dried. Staining is obtained by placing the slides 30min in a solution of Giemza (Meyer Hematoxylin) diluted 20 times in H₂O . It is then immersed in a solution of 1% acetic acid in H₂O for 30 seconds, a step known as differentiation. Then washed in H₂O and air-dried.

5.2.3 H&E

Hematotoxylin and eosin were purchased at Sigma Aldrich in their ready to use solution. Samples were fixed in methanol for 10 minutes and then air dried. Samples were placed in Hematoxylin for 30 min, then rinsed in 95% ethanol solution. Then dipped 8 times in Eosin, then rinsed again in 95% ethanol, then air-dried.

5.2.4 DAPI

DAPI is a blue fluorescent stain most often used to image live cells. DAPI stock solution with a concentration of 5mg/ml was diluted 1000 times in H₂O to create the staining solution. Samples were immersed for 5 min DAPI staining solution and then washed with H₂O and allowed to air-dry.

5.2.5 Ethidium Bromide

Ethidium Bromide (EthBr) is an intercalating agent that emits a strong fluorescence when bound to DNA. It is a standard tool for staining DNA when performing gel electrophoresis. It can also be used as a DNA stain in histological preparations. Our protocol consisted in fixing the samples

for 10 in methanol. Then the samples were incubated in a solution of $0.5\mu\text{g}/\text{ml}$ of EthBr in H_2O for 5 min and then washed with H_2O and allowed to air-dry

5.2.6 Pan-cytokeratins

Pan-cytokeratins (Pan-K) are a class of antibody that binds to a large amount of the known cytokeratins. It is reported as a marker of epithelial tissue. Cytokeratins antibody was purchased from Dako (Cytokeratin antibody, Clone AE1/AE3, Dako by Aligent, Santa Clara, CA), and protocol was optimised according to manufacturer's instructions. Samples were double fixed, by immersion 10 minutes in Ethanol at -20°C , then again 10 minutes in Acetone at -20°C , then allowed to air-dry. Samples were then washed with PBS and incubated with Block V (blocking solution supplied with the antibody by supplier) for 5 min. Washed with PBS and the again with PBS containing 0.5% bovine serum albumin (BSA) as well as 0.3% Triton-X100 (a detergent). Samples were then incubate with antibodies diluted 100 times, for 1h at room temperature in humid conditions. Washed with PBS and then incubated with HRP binding polymer for 30 minutes at room temperature in humid conditions then washed with PBS. Samples were then incubated for 12 min at room temperature with DAB chromogen (supplied with antibody by manufacturer) then washed with H_2O . Last, samples were counter-stained with SYBR for nucleus imaging.

5.3 Results

The first question is the thickness of the samples that was best for imaging. Figure 5.1 shows H&E staining of colon with varying thicknesses. If the thickness is too low, the amount of material deposited on the slide is too insubstantial. On the other hand, too thick slices cause the image to loose focus, most likely due to the small depth of focus of the microscopes used. $10\mu\text{m}$ was still seen as too thin and $15\mu\text{m}$ as too thick, the best compromise was chosen to be $12\mu\text{m}$.

5.3.1 Compatibility

Not all counter stains are compatible with an original stain. H&E is incompatible with any other stain. EthBr cannot be used for LCM, as the fluorescent emission required for imagined, burns through the PET membrane. Giemza has an autofluorescence similar to the one by DAPI, which makes them incompatible. SYBR is well compatible with Giemza, as well as with the antibodies protocols.

Chapter 5. Sample Preparation and Staining

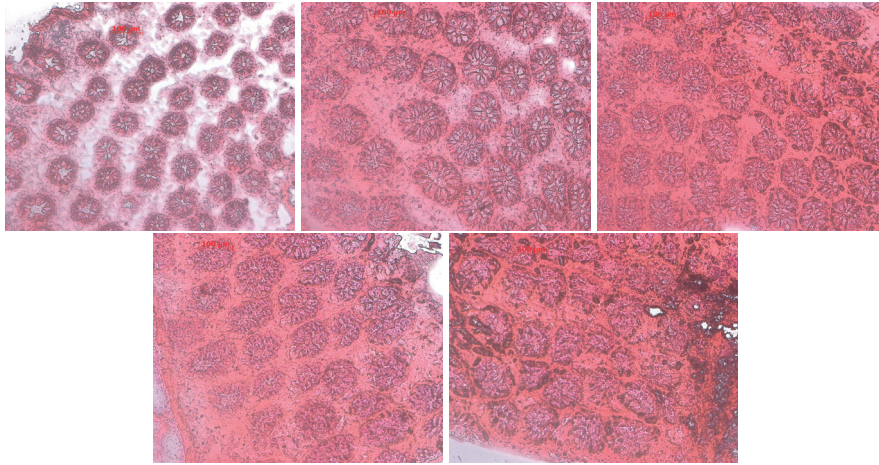
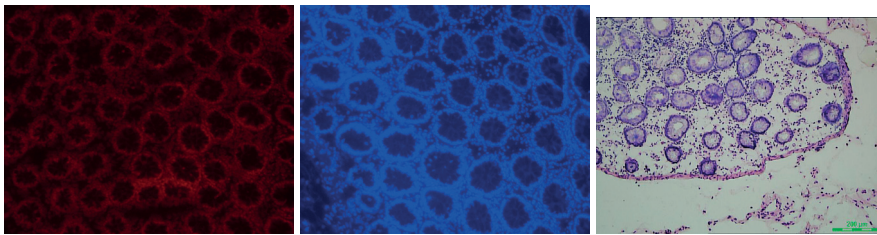
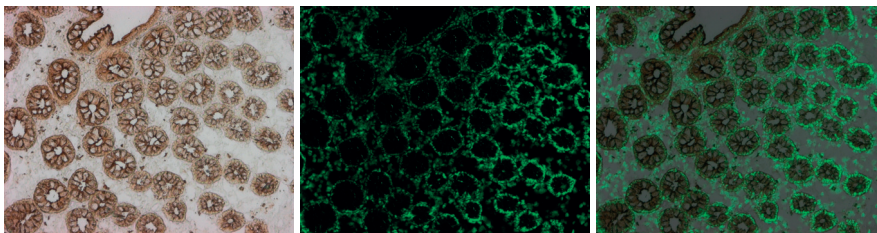


Figure 5.1: Varying thickness of colon tissue, stained with H&E. Thicknesses shown: 5, 10, 15, 20 and 25 μm



(a) Normal human colon, 12 μm cuts, stained using standard staining techniques. From left to right: Ethidium Bromide, DAPI, Giemza



(b) Normal human colon, stained using pan-K antibody and counterstained with SYBR. On the left, is the staining produced by the antibody. The crypts are clearly labeled as epithelial tissue, by the red-brown colouring given by the fluorophore associated with the antibodies. In the middle, is the fluorescent image showing SYBR staining. Crypt formations can be clearly distinguished by their ring-like structure of nucleus. On the right, the merged image of SYBR and antibody. Confirms that both method permit the clear distinction of epithelial and mesenchymal portions.

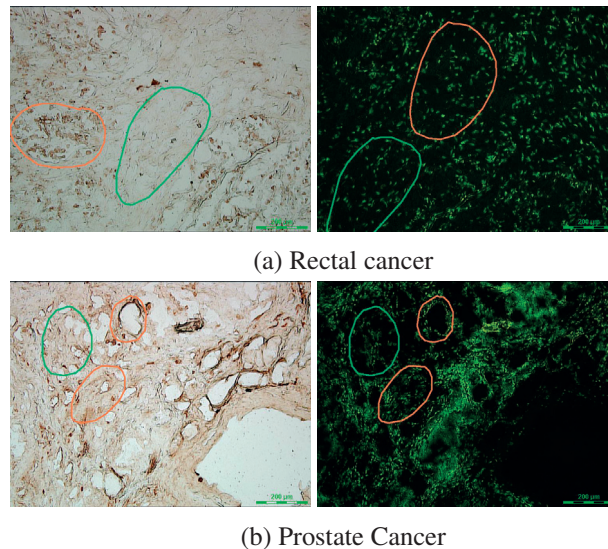


Figure 5.3: Rectal and prostate cancer, $12\mu\text{m}$ slices, stained using pan-k antibody and counter-stained with SYBR. Pictures taken during LCM experiment, the same area is imaged in both bright-field and in fluorescence, showing the IHC staining as well as the SYBR staining. The shapes visible were drawn using the LCM software for sample acquisition. They represent another attempt at selecting tumor areas.

5.3.2 Observing colon structure

The typical morphology is the crypt morphology. In all of the preparations shown, the crypts can easily be distinguished. The IHC clearly labels the epithelial cells of the crypts and does not mark the others. It is however very easy to distinguish the crypt formations in all of the staining. However, as can be seen in Figure 5.3, tumors showed no clear marking by IHC, neither are the clearly identifiable structures of the normal tissue visible.

5.4 Conclusion

In a normal tissue, such as the human colon, structures formed by the tissue are clearly identifiable, independently of the stains. As long as the slide preparation has been done carefully, all the stains will allow a careful observer, to recognise the tissue morphology and the defining structures, such as the colonic crypts. In tumors, these morphologies are lost. None of the stains experimented with here allowed a clear recognition of any structure bearing resemblance to normal tissue classification. Observing tumor tissues, it is observable that a sample is "not normal," however, no structures bearing clear resemblance to normal tissue features is observable. Also, if a tumor sample is composed of tumor cells and normal cells infiltrating it, no distinction seems to be clearly visible from histologic analysis. This observation seems counter-intuitive and to go against what is generally thought in the literature. However, it was already observed by Fialkow [56], that estimating the fraction of tumor cells in a sample is very difficult. The observations in

Chapter 5. Sample Preparation and Staining

his article suggests that there is no consistency between different pathological examinations of the same sample. Observations that validate the ones described here.

6 Detecting Mutations in the Mitochondrial Genome

The primer set used for the first round PCR was unreliable. To address the problem, we scanned the literature and looked for a series of primer sets and methods that could be used to amplify mtDNA. After trying a series of protocols, and not being able to have a reliable amplification of mtDNA on laser capture samples [91, 92], we decided to design our own pre-amp set. From our experience, a length of around 800bp seemed optimal. Furthermore, it was important that the PCR result could be checked using capillary electrophoresis, instead of gel electrophoresis. This is when the idea came of placing a tail at the end of each primer. A large portion of the second round primers that failed in the first attempt also needed to be re-designed. The results of this improvement are explained in [93]. The new set of primers consisted of 23 pre-amps (first round primers) and 192 second round primers, and allowed the scanning of 76% of the mtDNA.

6.1 Resulting Publication: Cycling temperature capillary electrophoresis: A quantitative, fast and inexpensive method to detect mutations in mixed populations of human mitochondrial DNA

6.1.1 Short description

This article describes CTCE as a method for the detection and quantification of mutations in mtDNA. It contains an in-depth description of the method with pre-amps, second round primers and cycling conditions. To validate the method, a sample containing multiple mutations is artificially created by mixing the DNA from 30 persons in equal amounts. The performance of the method is compared to next generation sequencing (NGS), using a standard protocol. The results clearly show that NGS is not quantitative, and is incapable of detecting mutations at a fraction below 30%. Also challenges involved with alignment of sequencing results are highlighted. In contrast, CTCE is shown to be able to accurately detect and quantify mutations present at fractions as low as 1%. The throughput and cost of CTCE for the detection of mtDNA mutation

Chapter 6. Detecting Mutations in the Mitochondrial Genome

is shown to be at least an order of magnitude better than NGS.

6.1.2 Significance

This is a methods paper. It describes and validates a method to detect and quantify mitochondrial mutations. CTCE is applicable in various situations, not limited to lineage tracing. Other researchers in the field of human mitochondrial DNA could benefit from a fast, quantitative and inexpensive way of detecting mutations. Possible applications include the diagnostic of mitochondrial diseases, forensic applications as well as tumor related questions. Furthermore, the performance of CTCE is often questioned, when compared to other methods such as NGS. Comparing the performance of the two methods, and showing the greater precision, throughput and lower cost of CTCE justifies our method choice. Successive papers can therefore make reference to this one.

6.1.3 Personal Contribution

In this article, I was responsible, together with Per Ekstrom, for the experimental design. I designed the primers, tested them and optimised them. Per Ekstrom came up with the idea of attaching a tail to the pre-amp primers. I designed the sequencing strategy and performed all the downstream data analysis. The interpretation of the results was a joint effort of all three authors, as was writing the paper.

6.1.4 Status

Published on 7th of May 2016, in Mitochondrion.

Cycling Temperature Capillary Electrophoresis: A quantitative, fast and inexpensive method to detect mutations in mixed populations of human mitochondrial DNA

Paulo Refinetti^a, Stephan Morgenthaler^a, Per O. Ekstrøm^b

^a*École Polytechnique Fédérale de Lausanne, EPFL FSB STAP, Station 8, Lausanne, Switzerland*

^b*Department of Tumor Biology, Norwegian Radiumhospital, Oslo, Norway*

Abstract

Cycling temperature capillary electrophoresis has been optimised for mutation detection in 76% of the mitochondrial genome. The method was tested on a mixed sample and compared to mutation detection by next generation sequencing. Out of 152 fragments 90 were concordant, 51 discordant and in 11 were semi-concordant. Dilution experiments show that cycling capillary electrophoresis has a detection limit of 1-3%. The detection limit of routine next generation sequencing was in the ranges of 15 to 30%. Cycling temperature capillary electrophoresis detect and accurate quantify mutations at a fraction of the cost and time required to perform a next generation sequencing analysis.

Keywords: mitochondrial DNA, heteroplasmy quantification, cycling temperature capillary electrophoresis, quantitative biology, High throughput

1. Introduction

1.1. Mitochondrial DNA

Mitochondrial DNA (mtDNA) is composed of a single circular chromosome of 16569 base pairs (hg38, GRCh38, Dec. 2013). It encodes for 22 tRNA, 13 protein subunits and two ribosomal RNA subunits. In each cell, many copies of the mtDNA are present. The number per cell is believed to vary according to cell type, however, few accurate estimates exist. Mitochondrial DNA is maternally inherited, which makes it well suited to study human migration. Quantitative mutational spectra of mtDNA are of great utility in medicine and biology. In

the following we briefly address a few of the applications, namely mitochondrial diseases, lineage tracing and tumour biology.

Mitochondrial diseases are prevalent in the order of 1 in 10 000 of the human population, with more or less severe symptoms depending on the affected tissue and degree of heteroplasmy of the deleterious mutation [1, 2, 3, 4]. This example shows the importance of having a quantitative mutation assay, because knowledge of the existence of a mutation will not characterise the disease.

There is currently much interest in using somatic mtDNA mutations as a marker for lineage tracing [5, 6, 7, 8]. Cytochrome C oxidase inactivation mutations have been successfully used as markers. The few attempts with sequencing, however, have been hampered by the slow speed and the excessive cost.

Since it is generally accepted that mtDNA mutations hold the answer to important biological questions, many methods have been developed to study them. Most require the isolation of mtDNA from nuclear DNA to prevent the amplification of pseudogenes (or NMUTs). This can be achieved by performing PCR selectively on mtDNA, and amplifying large portions that can then be analyzed [9, 10, 11, 12, 13, 14]. Methods range from whole mtDNA amplification using one primer pair [15, 14] to splitting it in many pieces [11, 12, 13]. Once the mtDNA has been amplified, it can be sequenced. Alternative methods to sequencing include temporal temperature gel electrophoresis [16], or restriction enzyme length polymorphism to detect insertions/deletions [17]. Sequencing, even after recent advances, remains expensive and slow if one has to analyse hundreds of samples.

To detect and quantify mutations double stranded DNA containing no mismatch (homoduplexes) from those containing one or more mismatches between the strands (heteroduplexes) can be separated with the use of denaturing capillary electrophoresis. Homoduplexes can be separated from heteroduplexes because of different melting temperatures [18, 19, 20]. Denaturing capillary electrophoresis was successfully used to quantify the error rates of DNA polymerase β , Taq DNA polymerase and Pfu DNA polymerase [21, 22, 23, 24]. Denaturing capillary electrophoresis was also used in its early forms to analyse mutations in the human mtDNA [25, 26], by scanning a 100bp portion for mutations. Recent advances on the denaturing capillary electrophoresis gave rise to cycling temperature capillary electrophoresis (CTCE). The use of CTCE omits stringent temperature control and fragments design [22]. Thus, allowing for analysis of many fragments simultaneously [27, 28, 29].

2. MATERIALS AND METHODS

2.1. General protocol

First, an initial round of PCR is performed using 23 primer pairs (Figure 1). An additional primer is used to increase efficiency and label the product. To check the presence of a product, capillary electrophoresis can be used. The second step is a second round of PCR using 193 primer pairs and cover 76% of the mitochondrial DNA (see supplementary material). If the optimised annealing temperature is used for each of the 193 fragments, the sensitivity is sufficient to detect mutations present at fractions of 1-2%. For the analysis of a single sample it is convenient to use the same annealing temperature on all 193 fragments. We used an average annealing temperature of 55°C which works on 152 fragments, which cover 62% of the total mtDNA. After the second round PCR has been performed, the resulting products are analysed using CTCE. We use a MegaBace1000 instrument, (GE Healthcare Life Sciences, CT, USA). The software was modified to that the built-in oven could cycle the temperature. The temperature is switched from high to low every 30 seconds during a period of 20 minutes. For optimal separation between the products, each fragment needs to be cycled at specific temperatures. If the fragments are set on the same support, it is possible either to make several runs at different cycling temperatures, or to use a single run with a larger temperature amplitude and slower switching frequency (see [22]).

For the purpose of comparing with modern sequencing methods, the whole mtDNA was amplified in nine fragments using the protocol described in Ramos et al. [11]. The resulting products were sent for library preparation and sequencing on an Illumina Hi-seq2000 at the University of Lausanne Integrated Genomics Facility. The results were then mapped using Bowtie2. All of the computations were performed on the EPFL basic science cluster.

2.2. Primer Design

2.2.1. Primers for the selective amplification of mtDNA without nuclear contamination

There are many regions of homology between mtDNA and nuclear DNA (nDNA). To prevent the amplification of these regions in the nuclear DNA (NMUTs or pseudogenes), the primers are designed not anneal to nDNA but only on mtDNA. Using the Cambridge reference sequence rCRS (NC.012920.1) for the mtDNA and BLAST (<http://blast.ncbi.nlm.nih.gov/>, 2015), the regions of homology were determined. Primers were selected in a way that would maximise the mismatches

between the nDNA and the mtDNA in the primer area. The length of the resulting amplicons was between 421 and 1156 bp. No attention was paid to possible secondary structures formed by the primer. A total of primer pairs were designed. The resulting fragments are shown in Figure 1. A short tail (CGC CCG CCG CGC CCC GCG) was added on the 5' end of all primers (both forwards and reverse). This addition allows the use of a third primer in the PCR reaction. The third primer anneals to the tail and carries a fluorescent ATTO 523 molecule. The consequence is an increase efficiency and the labelling of the product for detection using capillary electrophoresis. The primers are given in Table 1.

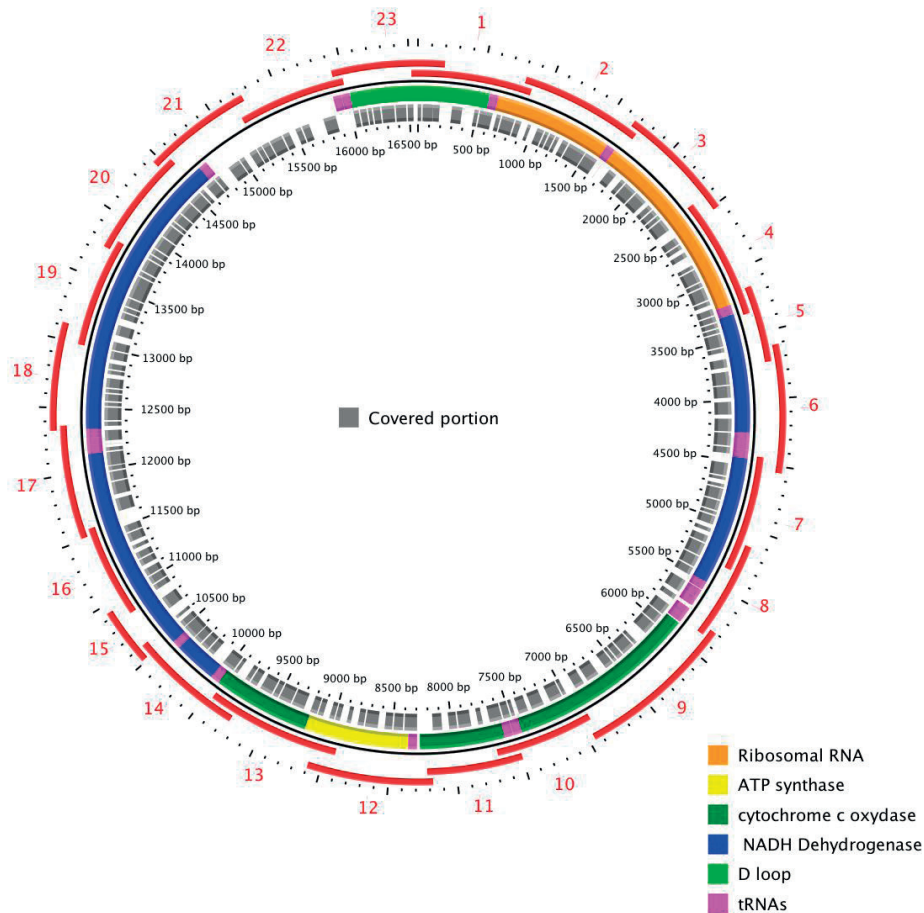


Figure 1: **PCR primer layout and coverage.** On the outside are the 23 fragments of first round PCR. Each fragment was selected to have a length of around 800 and as many mismatches between their annealing site in the mtDNA and homologous nDNA sites. The final result are fragments of between 421 and 1156 base pairs. On the inside is shown the effective coverage achievable by using the optimised annealing temperature for each second round PCR fragment (shown in grey).

2.2.2. Second Round PCR: amplification of fragments for analysis using CTCE

The design of the second round PCR fragments had to take into account the melting profile. As is described in [22], the fragments are required to have a monotone melting profile. Since the second round is performed on the PCR product of

Fragment	Start	End	Length	Forwards Primer	Reverse Primer	Annealing T
1	16521	880	928	CGCCGGCCGGCCGGCCGATAAAGCCATAAAT AGC CCA CA	TCGCCCCCGCCCGCCCGCCCAACCCCTGGGGTTA GTA TAG CT	54
2	816	1729	913	CGC CCG CCG CGC CCC GCG ACA GCA GTG ATT AAC CTT TAG CA	CGC CCG CGC CGC CCC GCGTGG TTT GGC TAA GGT TGT CTG GT	51.5
3	1665	2524	859	CGC CCG CCG CGC CCC GCGCGC TCT GAG CTA AAC CTA GCC CC	CGC CCG CGC CGC CCC GCG ATG CTA GAG GTG ATG TTT TTG GT	59
4	2402	3336	934	CGC CCG CCG CGC CCC GCG CAA GTC ATT ATT ACC CTC ACT GT	CGC CCG CGC CGC CCC GCG CAA TGA GGA GTA GGA GGT TGG CC	50.5
5	3149	3724	575	CGC CCG CCG CGC CCC GCGCCT ACT TCA CAA AGC GCC TTC CC	CGC CCG CGC CGC CCC GCG TGT TTG GGC TAC TGC TCG CAG TG	61
6	3610	4547	937	CGC CCG CCG CGC CCC GCGTCC TAT TTA TTC TAG CCA CCT CT	CGC CCG CGC CGC CCC GCG CAG TGC GAG CTT AGC GCT GTG AT	55.5
7	4447	5341	894	CGC CCG CCG CGC CCC GCGGTT GGT TAT ACC CTT CCC GTA CT	CGC CCG CGC CGC CCC GCG AGA GGT TAA GGA GGG TGA TGG TG	59
8	5126	5861	735	CGC CCG CCG CGC CCC GCGTCA ACT TAA ACT CCA GCA CCA CG	CGC CCG CGC CGC CCC GCG AAT CTA AAG ACA GGG GTT AGG CC	50.5
9	5800	6956	1156	CGC CCG CCG CGC CCC GCGAAT TCA ATA TGA AAA TCA CCT CG	CGC CCG CGC CGC CCC GCGCCT ACG GTG AAA AGA AAG ATG AA	62
10	6917	7671	754	CGC CCG CCG CGC CCC GCGTGC TCT GAG CCC TAG GAT TCA TC	CGC CCG CGC CGC CCC GCG TGA GGG CGT GAT CAT GAA AGG TG	55.5
11	7500	8216	716	CGC CCG CCG CGC CCC GCG CCT CCA TGA CTT TTT CAA AAA GG	CGC CCG CGC CGC CCC GCG GAC GAT GGG CAT GAA ACT GTG GT	54
12	8176	9096	920	CGC CCG CCG CGC CCC GCGCTG AAA TCT GTG GAG CAA ACC AC	CGC CCG CGC CGC CCC GCGAGT GTA GAG GGA AGG TTA ATG GT	54
13	8923	9953	1030	CGC CCG CCG CGC CCC GCG CAC ACC TAC ACC CCT TAT CCC CA	CGC CCG CGC CGC CCC GCG AAA CCA CAT CTA CAA AAT GCC AG	55.5
14	9752	10602	850	CGC CCG CCG CGC CCC GCGTGC AGT CTC CCT TCA CCA TTT CC	CGC CCG CGC CGC CCC GCGAGC TAT AAT GAA CAG CGA TAG TA	50.5
15	10519	10940	421	CGC CCG CCG CGC CCC GCGGTA GGA ATA CTA GTA TAT CGC TC	CGC CCG CGC CGC CCC GCG GGT GTC GGA GGA AAA GGT TGG GG	50
16	10852	11566	714	CGC CCG CCG CGC CCC GCGGCC TAA TTA TTA GCA TCA TCC CC	CGC CCG CGC CGC CCC GCG ATG CCT CAT AGG GAT AGT ACA AG	51
17	11515	12367	852	CGC CCG CCG CGC CCC GCGGCC TGA CAA AAC ACA TAG CCT AC	CGC CCG CGC CGC CCC GCG TAG GGT GGT TAT AGT AGT GTG CA	52.5
18	12333	13122	789	CGC CCG CCG CGC CCC GCGAAG TAA TAA CCA TGC ACA CTA CT	CGC CCG CGC CGC CCC GCGCGG ATG AGT AAG AAG AIT CCT GC	52.5
19	12990	13830	840	CGC CCG CCG CGC CCC GCG TAG CAG CAG CAG GCA AAT CAG CC	CGC CCG CGC CGC CCC GCG AGT CCT AGG AAA GTG ACA GCG AG	61
20	13748	14558	810	CGC CCG CCG CGC CCC GCGAAC ATT TCC CCC GCA TCC CCC TT	CGC CCG CGC CGC CCC GCGGTG TGT TAT TAT TCT GAA TTT TG	61
21	14455	15247	792	CGC CCG CCG CGC CCC GCGATC GCT GTA GTA TAT CCA AAG AC	CGC CCG CGC CGC CCC GCG CCT CCT CAG ATT CAT TGA ACT AG	51
22	15169	15993	824	CGC CCG CCG CGC CCC GCGGAG GGG CCA CAG TAA TTA CAA AC	CGC CCG CGC CGC CCC GCGTGG GTG CTA ATG GTG GAG TTA AA	51
23	15924	201	846	CGC CCG CCG CGC CCC GCG AAC CCG AGA CGA AAA CCT TTT TC	CGC CCG CGC CGC CCC GCG CTT TAG TAG GTA TGT TCG CCT GT	51

Table 1: First round PCR primers as used. The annealing temperature indicated has been found to be the optimal one using the PCR recipe described in the methodology. The primers are complete with the 5' short GC-tail.

the first round, the fragments can be designed without concern for homologies in the nDNA. An automated software was used to scan through the first round fragments to select sub-fragments with length between 50bp and 150bp and with monotone melting profile. Once the sub-fragment is selected, a complete GC-clamp is added to the 5' end of the "warm-end" primer (i.e. the primer located at the end of the fragment with the higher melting profile, whose existence is guaranteed by the monotonicity of the melting profile). The second round primers were all tested, and those found not to work were shifted two base pairs up or down in the sequence and optimised. This resulted in 193 fragments for whom PCR works. Together they cover 98% of the mtDNA. However, the portions of mtDNA in the primer cannot be scanned for mutations, thus the portion effectively analysed is reduced to 76% of the total human mitochondrial DNA. The table containing the list of second round primers and annealing temperature is in the supplementary materials.

2.3. PCR Conditions

As PCR buffer we use a 10X Thermopol buffer, which has the following components: 200mM Tris-HCl, 100mM (NH₄)₂SO₄, 100mM KCl, 20mM MgSO₄ and 1% Triton- X100. It is stabilised at pH 8.8 at 25°C .

2.3.1. First round PCR

First round samples were amplified using the following recipe. 1X Thermopol Buffer, 0.15μM forward primer, 0.15μM reverse primer, 0.15μM fluorescently labeled primer, dNTP 800μM , DNA template 0.3ng/μl and Taq DNA polymerase 0.07U/μl . The PCR temperature cycling conditions were as follows, 4 minutes initial denaturation at 94°C, then 38 cycles of 94°C, 15 seconds, 40 seconds annealing temperature (specific to each fragment) and extension time at 72°C for 150 seconds. All samples and amplified product were stored at -20°C until analysis.

Element	Concentration
10X Thermopol Buffer	0.1
Forward and reverse primer	0.3μM
"Short-GC" Primer	0.15μM
all dNTP	800μM
DNA template	0.3ng/μl
Taq DNA polymerase	0.07U/μl

Table 2: Protocol for first round of PCR in a final volume of 10μl

2.3.2. Second Round

Second round PCR was performed for all fragments using the recipe shown in Table 3. The following PCR cycling were used for all fragments; 94°C for 60 seconds followed by 30 cycles of 94°C for 15 seconds, 55°C for 30 seconds and 72°C for 60 seconds. All samples and amplified product were stored at -20°C.

Element	Concentration
10X Thermopol Buffer	0.1
Primer without "GC-clamp"	0.3 μ M
Primer with "GC-clamp"	0.15 μ M
Labeled "GC-clamp" Primer	0.15 μ M
all dNTP	500 μ M
Bovine Serum Albumine at 10mg/ml	1%
200 times diluted 1 st round of PCR	0.05 μ l/ μ l
Cloned Pfu DNA polymerase	0.1U/ μ l

Table 3: Protocol for second round of PCR in a final volume of 10 μ l

2.4. Capillary Electrophoresis

The PCR product after the first round was analysed using capillary electrophoresis (CE). Using the a MegaBace 1000, sample were injected at 5kV during 90 seconds. A DNA ladder for estimating the size of the PCR product is co-injected. We then run the electrophoresis for 200mins at 9kV with a constant temperature of 28°C.

2.4.1. Cycling Temperature Capillary Electrophoresis (CTCE)

CTCE was performed with a MegaBACE1000 DNA sequencer (GE Healthcare Life Sciences, CT, USA)). The instrument software was modified by replacing the "tmpr.nxe" file with an updated version obtained from Molecular Dynamics (now part of GE Healthcare). The update controls the cooling fan, and temperature limits in the registry. Cycling temperature changes and intervals (cycles) were specified in the "macro.ini" file under the section "Inject Samples and Run". Files and detailed descriptions are available upon request. The separation principle and general protocol of CTCE has previously been published [22]. The Polymer used to fill the capillary was obtained from The Gel Company (San Francisco, CA, USA) The electrophoresis buffer used in the anode and cathode side was a 1:10 dilution of a 10X running buffer. The 10X electrophoresis buffer was

made with the following components, 300mM tris, 1.0M TAPS, 10mM EDTA, pH adjusted to 8.00 and sterile filtered.

Before CTCE, samples are thawed at room temperature. Injection time is 25 seconds, at 10kV. The electrophoresis potential is 9kV and it takes between 30 and 50 mins. In the case where all 193 fragments are organised in 96 well plates with varying melting profiles, each plate was run six times with six different cycling temperatures for 20 cycles. The running temperatures are: Run 1 at 44-41°C ; Run 2 at 47-44°C; Run 3 at 50-47°C; Run 4 at 53-50°C; Run 5 at 56-53°C; Run 6 at 59-56°C. Using this strategy, it is possible to analyse simultaneously fragments with varying melting temperatures. It also has the added benefit of allowing a clearer identification of double stranded PCR products as they can be seen to shift in time as the temperature increases.

2.5. Next Generation Sequencing

The preparation of the sample for sequencing was performed using the method described by Ramos et.al.[11]. The DNA from the 30 subjects was amplified separately using the PCR condition as described in [11], for a total of $30 \times 9 = 270$ PCR. One μ l of product resulting from each of the PCR results was then taken and pooled together. The library preparation and sequencing was performed at the Integrated Genomics Facility of UNIL (Lausanne Genomic Technologies facility, <http://www.unil.ch/dafl/home.html>). A single library was prepared from the sample and sequencing was performed using a Hi-seq2000 instrument from Illumina (Illumina Inc, San Diego, CA). Output from sequencing was aligned to the rCRS (NC_012920.1) using Bowtie2. For a complete description of the options of Bowtie2, please refer to the manual at <http://bowtie-bio.sourceforge.net/bowtie2/manual.shtml>. In brief, Bowtie2 can run in two ways: local and end-to-end. The local mode aligns only the better matching portions or reads, while the rest are discarded. The end-to-end mode on the other hand forces the whole read (from end to end) to be aligned to the sequence. The local mode has higher computational cost. Bowtie2 contains pre-set options, for both the local and the end-to-end mode. These present options vary in their degree of selectivity and sensitivity. The more sensitive ones make use of longer seeds (ie: a sample from the read used to identify possible alignment areas). These are the following:

- very-fast (Same as: -D5-R1-N0-L22-iS,0,2.50)
- fast (Same as: -D 10 -R 2 -N 0 -L 22 -i S,0,2.50)

- sensitive (Same as: -D 15 -R 2 -L 22 -i S,1,1.15 the default in –end-to-end mode)
- very-sensitive (Same as: -D 20 -R 3 -N 0 -L 20 -i S,1,0.50)
- very-fast-local (Same as: -D 5 -R 1 -N 0 -L 25 -i S,1,2.00)
- fast-local (Same as: -D 10 -R 2 -N 0 -L 22 -i S,1,1.75)
- sensitive-local (Same as: -D 15 -R 2 -N 0 -L 20 -i S,1,0.75 the default in –local mode)
- very-sensitive-local (Same as: -D 20 -R 3 -N 0 -L 20 -i S,1,0.50)

The alignment was attempted with each of the eight preset options, four local and four end-to-end. The result of the mapping was then used to generate a pileup. The final result of the analysis was the number of times each base was read, as well as total coverage and quality score.

2.6. *Sample Description and Preparation*

The method described in this paper is optimised for the detection and quantification of mtDNA mutations in mixed samples. Consequently a sample containing a large number of variations in the sequence was created. DNA from 30 subjects was pooled together. For each sample, the starting point was frozen tissue from which a needle biopsy was taken. DNA was extracted using the Quiagen Mag-Attract system according to the manufacturer’s instructions. The samples were digested using Proteinase K and then inserted in the instrument for DNA extraction. One μ l of extracted DNA was taken from each sample and pooled together. The samples used were taken between 1997 and 2005 at the Norwegian Radium Hospital, as part of routine care with patient’s informed consent to store surgical discards for research purpose. Samples were anonymised with respect to patient information. The exact haplogroups of the individual samples is not known. However, all patients were Norwegians, and can therefore be expected come mainly from the H and U haplogroup[30]. Under Norwegian legislation, technical and methodological development work that uses anonymised biological material does not require approval by a research ethics committee.

3. RESULTS

3.1. First round amplification

To verify the effectiveness of the first round PCR, both capillary and gel electrophoresis were used. The resulting amplification are shown in Figure 2, which demonstrates PCR effectiveness.

3.2. Second round amplification

The 152 fragments cover 12496bp, or 75.4% of the total, this includes the primers. Since mutations in the primer area are not detectable, only 62% of the total mtDNA is effectively scanned for mutations. If all the fragments are amplified separately using the optimal annealing temperature, all of the 193 fragments will amplify, for a final coverage of 98% of the whole mtDNA thus resulting in an effective analysis of 76% of the total for the purpose of mutation discovery (see Figure 1). The primer set has so far been tested on individuals from North European populations.

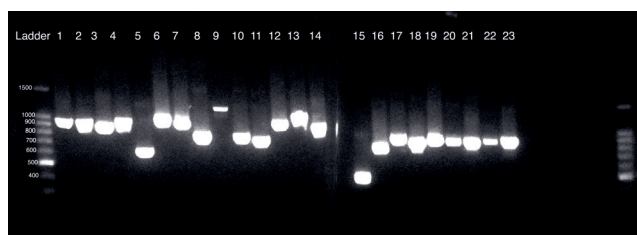


Figure 2: **Gel Electrophoresis of the first round PCR.** On the rightmost and leftmost lane is a GeneRuler 100 bp DNA Ladder from ThermoScientific. The sizes of the bands are indicated on the left. The intermediate lane contain the PCR products ordered from the first to the 23rd from left to right.

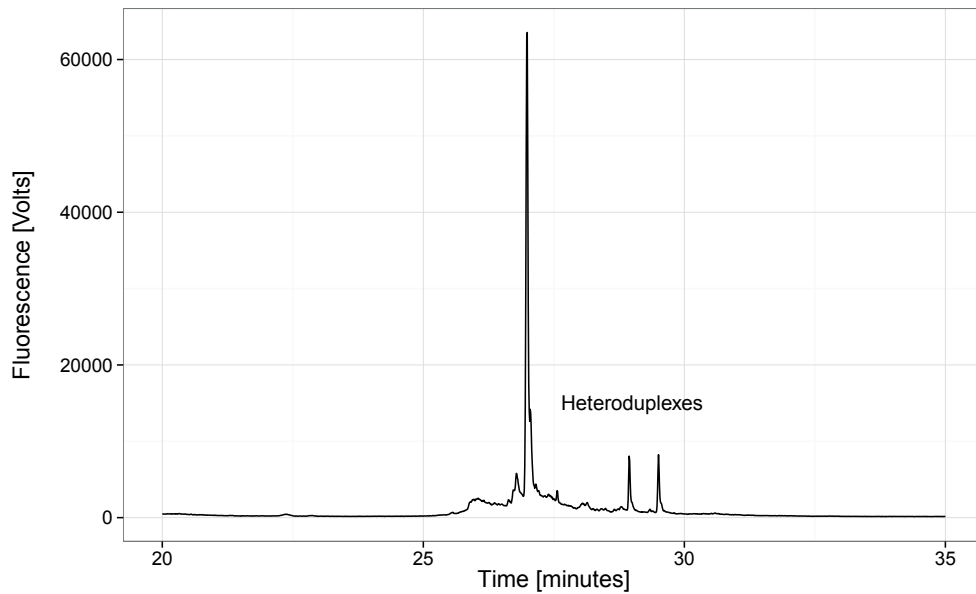


Figure 3: **CTCE output example.** The electropherogram display mutation detection of the fragment amplified between base pairs 13322 and 13401 of the human mtDNA. The mutant fraction is calculated by comparing the area under the heteroduplex peaks to the total area.

3.3. Precision of CTCE

To check the precision of CTCE, an experiment with serial dilutions was performed. The starting point is the DNA from two sources, one with a mutations at an appreciable level in one of the fragments and the other without any mutations in that fragment. A series of twelve wells is then prepared. The first well contains only the sample from the mutated DNA. The subsequent wells were constructed by mixing equal volumes of the previous well with the non mutated sample. This experiment was performed separately for two different samples with mutations in different fragments of the mtDNA. Figure 4 shows the logarithm of the mutant fraction as a function of the well number or the dilution. The mutant fractions obtained by CTCE cluster around a line, which demonstrates the precision of CTCE in determining mutant fractions down to a detection limit of around 1%. Note that the two regression lines do not have slope -1 . The slope would be -1 only if the DNA concentration in the initial two DNA extraction were identical. Differences in DNA extraction efficiency as well origin of the samples could accounts for the different concentration of mtDNA in the initial samples.

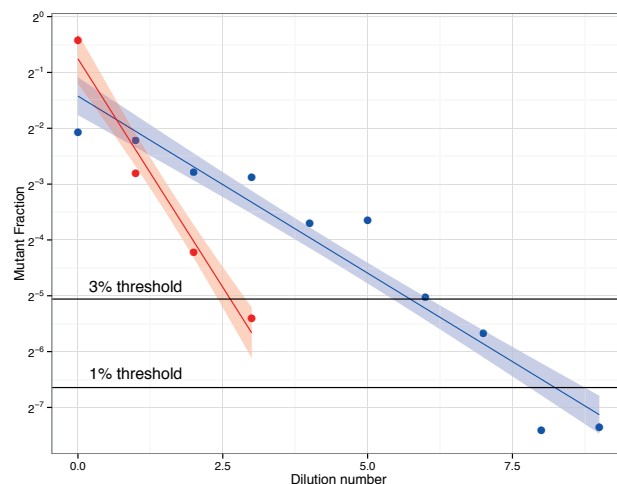


Figure 4: **Precision of CTCE.** The computed mutant fraction is shown as a function of dilution number. The two colours show the two different samples for which the dilution series of mixed samples into homogeneous sample was performed. How well the samples adhere to a straight line is a measure of the precision of CTCE when it comes to quantifying the mutant fraction in a mixed sample. The slopes are the result of the serial dilution series by a constant factor. We diluted the mixed sample by a factor 2 in volume, but since the initial mtDNA concentration in both samples is unknown, the slope is expected to be different from one. One might calculate the ratio of mtDNA concentration between both samples using the present figure.

3.4. Mutations

The use of CTCE allows the detection and quantification of mutations in any of the fragments. It does not, however, identify the nature of the mutation. The data from sequencing is thus arranged in a manner to make it comparable with CTCE by counting in each of the CTCE fragments the number of mutations found. Restricting the analysis to the 152 fragments in which mutations can be found using CTCE when processing the samples in parallel (Table 5).

Out of the 152 fragments analysed, 90 were perfectly concordant (same number of mutations identified), 11 were semi concordant, and 51 were discordant (see Figures 5 and 6).

3.5. Concordance of CTCE and sequencing

Figures 5 investigate the concordance and discordance between CTCE and sequencing. Figure 5 shows a good concordance in both presence and number of mutations in each fragment between the methods. Also clearly visible are fragments where CTCE finds mutations that are not validated by sequencing or vice versa. Table 5 shows an overall summary

Sequencing Parameter	
Reads	$\sim 339 \times 10^6$
Mapped reads	$\sim 60\%$ (alignment dependent)
Average number of times a base was read	596000
Minimum number of times a base was read	87000

Table 4: This table describes the output of sequencing with Illumina Hi-Seq 2000 and read alignment. The low percentage of mapped reads has been confirmed by another experiment where a 100bp homogeneous sample was sequenced (Per Ekstrom, currently unpublished results)

	CTCE	Sequencing
Total Mutations Identified	89	79
Fragments with mutations	59	51

Table 5: This table compares CTCE and sequencing in terms the mutations found and the number of fragments where a mutations was identified

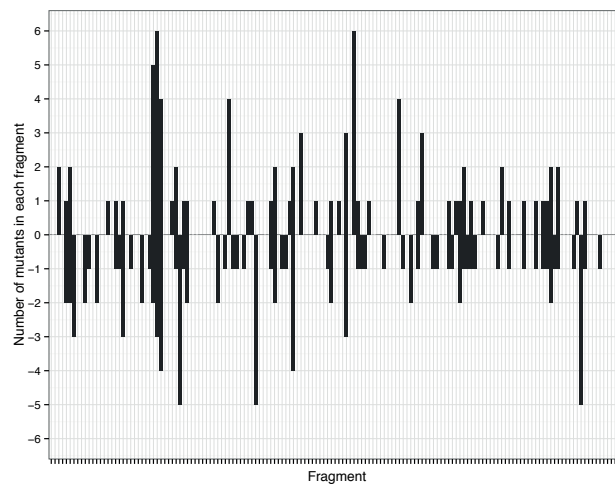


Figure 5: **Mutations in each Fragment.** For each fragment, the numbers of mutations found by CTCE and by Hi-seq. The positions are the fragment numbers. Each vertical line is a fragment, so a position without a bar is a fragment where no mutation was found. The top part of the plot shows the mutations detected using sequencing, and the lower panel (in negative) shows the mutations detected using CTCE in the same fragment.

4. DISCUSSION

4.1. Discordance of CTCE and sequencing

Figure 6 shows that there are clear discordance between sequencing and CTCE. Three problematic areas can be distinguished in the plot. There are two groups of fragments where sequencing detects five or more mutations and CTCE detects two or more. In this case the explanation could be an underestimation of the mutant number by CTCE. In fact, the complexity of the electropherogram increases exponentially with the number of mutations present in the fragment. The second problem concerns fragments where sequencing detects two or more mutations whereas CTCE finds none. These areas are most likely due to sequencing noise introduced in sequencing or post-processing of results. The third area of discordance are the cases where sequencing finds two or less mutations and CTCE finds more mutations. Since the detection limit of sequencing is in the order of 15% (Figure 8) this is to be expected.

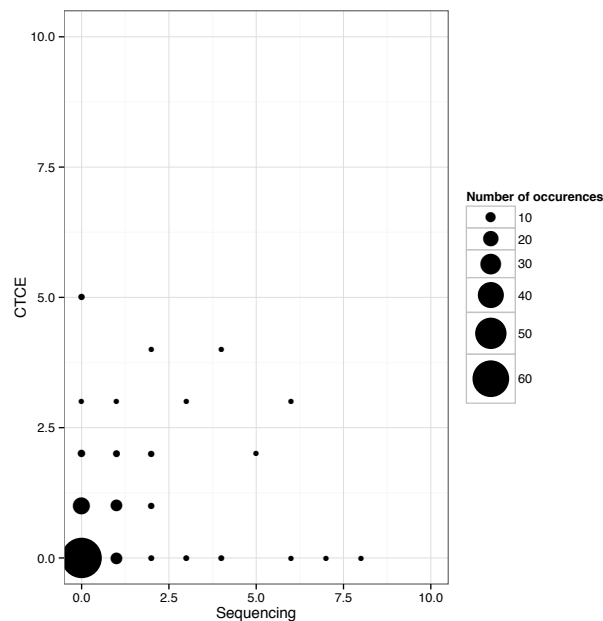


Figure 6: **Sequencing results as a function of CTCE results.** This plot correlates the number of mutations found by sequencing and by CTCE. The size of the dot represents the number of times that specific configuration has occurred. All the sequencing results shown here are those mapped using the "very-sensitive-local" method.

4.2. Alignment of sequencing results

The raw output from the Illumina Hi-Seq2000 are a series of short reads (roughly 100bp long). In the order of 10^8 reads are created in a sequencing run. Alignment is the process of determining which portion of the analysed sequence each read represent. The interpretation of sequencing results is heavily dependent on the alignment procedure. It is clear that low rate of alignment errors will create a background of apparent mutant fraction across the sequence. Figure 7 shows the spread of mutations and mutant fractions as detected using different alignment procedures on the same output. There are inconsistencies, with discordant mutant fractions for most mutations. There are also several cases of an alignment procedure detecting a mutation, even at a high fraction that is not confirmed by other alignment results. Inspection of the data reveals the alignments cluster into two groups with roughly similar behaviour. The end-to-end alignments, which are the Bowtie2 defaults, and the local alignments. Within the two groups, increasing stringency does not make a visible difference. It is not possible to decide which alignment method, local or end-to-end is more accurate using the observed output alone. Local alignment is a more conservative approach as it contain fewer mutations at lower fractions. Comparison of sequencing results with CTCE results shows a greater similarity between CTCE and local alignment than with end-to-end alignment. The different alignment procedures also succeed in aligning a different number of reads. End-to-end alignment typically aligns around 58% of the reads, whereas the local alignment reaches around 60%. We do not have enough data to test the significance of this finding. These results corroborate doubts about the usability of sequencing methods for the detection of low frequency mutations put forward in Taylor et al. [13] and Hancock et al. [31, 32], which suggests that the detection limit of mutant fractions by sequencing is somewhere between 20 and 30%. Though more accurate analyses are possible[33, 34, 35], they involve higher complexities and costs. The error rate estimate of 30% therefore reflects what is possible in routine analysis. Figure 8 shows the number of mutations found in the whole mtDNA when applying different thresholds per base pair. Assuming the number of mutations present in the sample is in the order of 100, the threshold for local alignment can be placed slightly below 5%, whereas the threshold for end-to-end alignment (used in [13]) would be slightly below 10%. Since the threshold is for each base pair, the total noise level is three times the threshold, thus roughly 30% for end-to-end alignments and 15% for end-to-end alignment. If the sample analysed is known to contain few mutations (in the order of one or two), the detection limit can be significantly reduced by applying more stringent alignment methods[33, 34, 35].

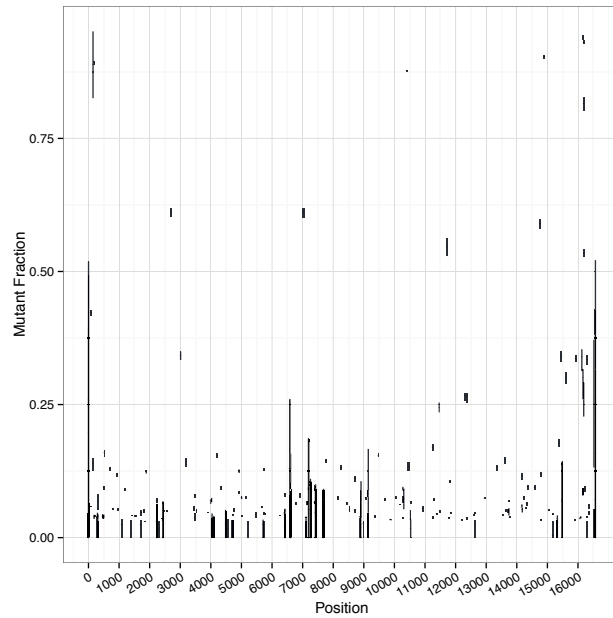


Figure 7: **Variations between alignment results.** The mutations found by next-generation sequencing as reported by different alignment methods are shown in this plot. The graphs shows for each position, the spread of mutant fraction found by the eight alignment method tested. The line covers the interval of the minimum fraction to the maximum fraction. A line that touches the x axis means that at least one method of alignment did not find a mutation at that position.

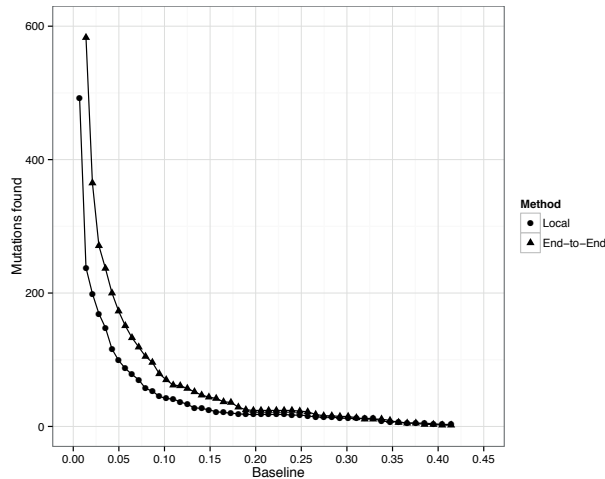


Figure 8: **Mutations found as a function of baseline.** For each position in the mtDNA we have a number of reads that result in an A, T, G and C. If any of these number is less than the baseline times the number of total reads at that position it is set to zero. To justify our choice of baseline, we show here the number of mutations found as a function of the baseline chosen. High baseline detects too few mutations, and inversely for low baseline. The chosen level of baseline was 5% for local alignment and 10% for end-to-end alignment.

4.3. Cost analysis

Once a method to cover a desired sequence of DNA is established, the cost of performing a CTCE analysis is many orders of magnitude smaller than sequencing. For the analysis of 62% of mtDNA of a single sample, as we performed here, we calculate that the total cost of the procedure is around 75 USD. A more detailed explanation is given in Table 6. The time it takes to perform the complete analysis is 20 hours, of which 4 hours are work and the rest is waiting time. The costs involved with next generation sequencing are much higher. For a detailed evaluation of the cost please refer to one of the many reviews [36, 37]. The cost of library preparation and sequencing by the facility at University de Lausanne was 1920 CHF (2200 USD at the time). All other costs are not taken into account as they are much smaller.

4.4. Conclusion

This paper proposes a fast, quantitative and inexpensive method of analysing mutations in a sample containing a mixed population of mtDNA. The method clearly performs better than sequencing in terms of cost, speed and precision, with

Experimental Step	Materials used	Approximate Cost
First round PCR, 23 reactions	General PCR reagents	5 USD
Second round PCR, 193 reactions	General PCR reagents	20 USD
CTCE	Capillary depreciation, Gell and Buffer	50 USD

Table 6: The table shows the cost of running a CTCE analysis on a single sample. It excludes the value or depreciation of the equipment and electricity. PCR thermocycler and capillary electrophoresis machinery should be available. We include all consumables.

the drawback of not identifying the mutation. It detects mutations at fractions as low as 1% and it furthermore determines their fraction accurately. Coupled with a cost below 100 USD per sample, high throughput and fast results, it can be applied in various areas of biology. For example in the quantification of mutant fractions in a large number of samples taken using Laser capture micro dissection.

Appendix A. Additional file — Second Round Primer list

An excel file containing the list of all second round primers complete with GC-clamp. The file show the set of primer sequences as well as which first round PCR template the primer amplifies on.

Appendix B. Competing interests

The authors declare that they have no competing interests.

Appendix C. Author's contributions

Primer design and Lab work has been performed by Ekstrom and Refinetti working together. The design of the experiment was a combined effort by all three authors. All authors contributed to writing and reviewing the manuscript. All authors have read and approved the manuscript for publication.

Appendix D. Acknowledgements

We would like to thank Prof. William G. Thilly for the contributions and guidance during the work that led to the current publication. We would also like

to thank Prof. Eivind Hovig for his constructive criticism. And last we would like to thank Prof. William Pralong, for his help in building the initial hypothesis and his support throughout.

- [1] S. DiMauro, E. A. Schon, Mitochondrial DNA mutations in human disease, *American journal of medical genetics* 106 (2001) 18–26.
- [2] L. C. Greaves, A. K. Reeve, R. W. Taylor, D. M. Turnbull, Mitochondrial DNA and disease., *The Journal of pathology* 226 (2012) 274–286.
- [3] C. B. Park, N.-G. Larsson, Mitochondrial DNA mutations in disease and aging., *The Journal of Cell Biology* 193 (2011) 809–818.
- [4] M. Zeviani, Mitochondrial disorders, *Brain* 127 (2004) 2153–2172.
- [5] J. K. Blackwood, S. C. Williamson, L. C. Greaves, L. Wilson, A. C. Rigas, R. Sandher, R. S. Pickard, C. N. Robson, D. M. Turnbull, R. W. Taylor, R. Heer, In situ lineage tracking of human prostatic epithelial stem cell fate reveals a common clonal origin for basal and luminal cells., *The Journal of pathology* 225 (2011) 181–188.
- [6] T. G. Fellous, S. A. C. McDonald, J. Burkert, A. Humphries, S. Islam, N. M. W. De-Alwis, L. Gutierrez-Gonzalez, P. J. Tadrous, G. Elia, H. M. Kocher, S. Bhattacharya, L. Mears, M. El-Bahrawy, D. M. Turnbull, R. W. Taylor, L. C. Greaves, P. F. Chinnery, C. P. Day, N. A. Wright, M. R. Alison, A methodological approach to tracing cell lineage in human epithelial tissues., *Stem cells (Dayton, Ohio)* 27 (2009) 1410–1420.
- [7] N. T. Gaisa, T. A. Graham, S. A. McDonald, R. Poulosom, A. Heidenreich, G. Jakse, R. Knuechel, N. A. Wright, Clonal architecture of human prostatic epithelium in benign and malignant conditions., *The Journal of pathology* 225 (2011) 172–180.
- [8] K. Kretzschmar, F. M. Watt, Lineage tracing., *Cell* 148 (2012) 33–45.
- [9] K. Khrapko, N. Bodyak, W. G. Thilly, N. J. van Orsouw, X. Zhang, H. A. Coller, T. T. Perls, M. Upton, J. Vijg, J. Y. Wei, Cell-by-cell scanning of whole mitochondrial genomes in aged human heart reveals a significant fraction of myocytes with clonally expanded deletions, *Nucleic acids research* 27 (1999) 2434–2441.

- [10] L. Fendt, B. Zimmermann, M. Daniaux, W. Parson, Sequencing strategy for the whole mitochondrial genome resulting in high quality sequences, *BMC Genomics* 10 (2009) 139.
- [11] A. Ramos, C. Santos, L. Alvarez, R. Nogués, M. Aluja, Human mitochondrial DNA complete amplification and sequencing: A new validated primer set that prevents nuclear DNA sequences of mitochondrial origin co-amplification, *Electrophoresis* 30 (2009) 1587–1593.
- [12] A. Ramos, C. Santos, E. Barbena, L. Mateiu, L. Alvarez, R. Nogués, M. P. Aluja, Validated primer set that prevents nuclear DNA sequences of mitochondrial origin co-amplification: a revision based on the New Human Genome Reference Sequence (GRCh37)., *Electrophoresis* 32 (2011) 782–783.
- [13] R. W. Taylor, G. A. Taylor, S. E. Durham, D. M. Turnbull, The determination of complete human mitochondrial DNA sequences in single cells: implications for the study of somatic mitochondrial DNA point mutations., *Nucleic acids research* 29 (2001) E74–4.
- [14] C. H. Tengan, C. T. Moraes, Detection and analysis of mitochondrial DNA deletions by whole genome PCR, *Biochemical and molecular medicine* 58 (1996) 130–134.
- [15] N. D. Bodyak, Quantification and sequencing of somatic deleted mtDNA in single cells: evidence for partially duplicated mtDNA in aged human tissues, *Human Molecular Genetics* 10 (2001) 17–24.
- [16] D.-J. Tan, R.-K. Bai, L.-J. C. Wong, Comprehensive scanning of somatic mitochondrial DNA mutations in breast cancer., *Cancer research* 62 (2002) 972–976.
- [17] P. F. Chinnery, D. C. Samuels, J. Elson, D. M. Turnbull, Accumulation of mitochondrial DNA mutations in ageing, cancer, and mitochondrial disease: is there a common mechanism?, *The Lancet* 360 (2002) 1323–1325.
- [18] R. S. Cha, H. Zarbl, P. Keohavong, W. G. Thilly, Mismatch amplification mutation assay (MAMA): application to the c-H-ras gene., *PCR methods and applications* 2 (1992) 14–20.

- [19] K. Khrapko, H. Collier, P. André, X. Li, F. Foret, A. Belenky, B. Karger, W. Thilly, Mutational spectrometry without phenotypic selection: Human mitochondrial DNA, *Nucleic acids research* 25 (1997) 685–693.
- [20] X.-C. Li-Sucholeiki, A. Tomita-Mitchell, K. Arnold, B. J. Glassner, T. Thompson, J. V. Murthy, L. Berk, C. Lange, P.-M. Leong-Morgenthaler, D. MacDougall, J. Munro, D. Cannon, T. Mistry, A. Miller, C. Deka, B. Karger, K. M. Gillespie, P. O. Ekstrom, J. A. Todd, W. G. Thilly, Detection and frequency estimation of rare variants in pools of genomic DNA from large populations using mutational spectrometry, *Mutation Research/Fundamental and Molecular Mechanisms of Mutagenesis* 570 (2005) 267–280.
- [21] B. P. Muniappan, W. G. Thilly, The DNA polymerase beta replication error spectrum in the adenomatous polyposis coli gene contains human colon tumor mutational hotspots., *Cancer research* 62 (2002) 3271–3275.
- [22] P. O. Ekstrom, K. Khrapko, X.-C. Li-Sucholeiki, I. W. Hunter, W. G. Thilly, Analysis of mutational spectra by denaturing capillary electrophoresis, *Nature Protocols* 3 (2008) 1153–1166.
- [23] W. Zheng, K. Khrapko, H. A. Collier, W. G. Thilly, W. C. Copeland, Origins of human mitochondrial point mutations as DNA polymerase γ -mediated errors, *Mutation Research/Fundamental and Molecular Mechanisms of Mutagenesis* 599 (2006) 11–20.
- [24] P. André, A. Kim, K. Khrapko, W. Thilly, Fidelity and mutational spectrum of Pfu DNA polymerase on a human mitochondrial DNA sequence, *Genome research* 7 (1997) 843–852.
- [25] H. A. Collier, K. Khrapko, N. D. Bodyak, E. Nekhaeva, P. Herrero-Jimenez, W. G. Thilly, High frequency of homoplasmic mitochondrial DNA mutations in human tumors can be explained without selection - *Nature Genetics*, *Nature Genetics* 28 (2001) 147–150.
- [26] K. Khrapko, H. Collier, P. Andre, X. Li, J. Hanekamp, W. Thilly, Mitochondrial mutational spectra in human cells and tissues, *Proceedings of the National Academy of Sciences of the United States of America* 94 (1997) 13798–13803.

- [27] P. O. Ekstrøm, J. Bjørheim, G. Gaudernack, K.-E. Giercksky, Population Screening of Single-Nucleotide Polymorphisms Exemplified by Analysis of 8000 Alleles, *Journal of Biomolecular Screening* 7 (2002) 501–506.
- [28] P. O. Ekstrom, J. Bjørheim, W. G. Thilly, Technology to accelerate pangenomic scanning for unknown point mutations in exonic sequences: cycling temperature capillary electrophoresis (CTCE), *BMC Genetics* 8 (2007) 54.
- [29] P. O. Ekstrøm, D. J. Warren, W. G. Thilly, 22539319, *Electrophoresis* 33 (2012) 1162–1168.
- [30] G. Passarino, G. L. Cavalleri, A. A. Lin, L. L. Cavalli-Sforza, A.-L. Børresen-Dale, P. A. Underhill, Different genetic components in the Norwegian population revealed by the analysis of mtDNA and Y chromosome polymorphisms., *European journal of human genetics : EJHG* 10 (2002) 521–529.
- [31] D. K. Hancock, L. A. Tully, B. C. Levin, A Standard Reference Material to determine the sensitivity of techniques for detecting low-frequency mutations, SNPs, and heteroplasmies in mitochondrial DNA, *Genomics* 86 (2005) 446–461.
- [32] Hancock, Human Mitochondrial DNA—Amplification and Sequencing Standard Reference Materials (2003) 1–93.
- [33] M. Li, M. Stoneking, A new approach for detecting low-level mutations in next-generation sequence data, *Genome biology* 13 (2012) R34.
- [34] M. Li, A. Schönberg, M. Schaefer, R. Schroeder, I. Nasidze, M. Stoneking, Detecting heteroplasmy from high-throughput sequencing of complete human mitochondrial DNA genomes., *American Journal of Human Genetics* 87 (2010) 237–249.
- [35] A. Pyle, G. Hudson, I. J. Wilson, J. Coxhead, T. Smertenko, M. Herbert, M. Santibanez-Koref, P. F. Chinnery, Extreme-Depth Re-sequencing of Mitochondrial DNA Finds No Evidence of Paternal Transmission in Humans., *PLoS Genetics* 11 (2015) e1005040.
- [36] M. Gundry, J. Vijg, Direct mutation analysis by high-throughput sequencing: From germline to low-abundant, somatic variants. 729 (2012) 1–15.

- [37] E. R. Mardis, Next-generation DNA sequencing methods., *Annual Review of Genomics and Human Genetics* 9 (2008) 387–402.

7 Clinical Applications

After writing and submitting the paper [93] it was clear we had a solid method in hand. It had been designed, tried and tested, as well as proven superior to the standard NGS approach. The next logical step was to answer the question raised by the first experiment, and find out the frequency and distribution of mutations in tumors. We already had a large amount of DNA extracted from tumor samples, and many tumor samples in the freezer. Scanning the mtDNA in a large number of DNA samples, extracted from human tumors, could give us a much better understanding to the distribution and frequency and distribution of mtDNA mutations. The details of this work can be found in the paper [94].

7.1 Resulting Publication: Scanning the mitochondrial genome for mutations by cycling temperature capillary electrophoresis

7.1.1 Short description

The article describes the result of applying the method described in our first paper to 94 different human tumors. 12 tumors of 8 types are scanned for the presence of mitochondrial mutations, which would also alert us to significant differences between tumor types. The results show that a majority of tumors carry a mutations on the mitochondrial DNA, although often at low fractions. Also, the mutations are not uniformly distributed throughout the mtDNA. Around 20 out of 192 fragments contain almost 80% of all detected mutations. All mutations that were present at a fraction above 30% in the samples, were Sanger sequenced (SS), and the base pair substitution identified (mutations at a fraction below 30% cannot be identified in SS). No two identical mutations were found. The position, nature and amount of mutation does not depend on the cancer type. mtDNA mutations therefore happen in a process that is independent of cancer type.

7.1.2 Significance

The results that Christian Arstad obtained did not show any detectable difference in mtDNA mutations between various tumors. Also, it allowed us to better understand the mutational spectra of mtDNA in tumors. Finding that most mutations occur in only a small fraction of the genome, greatly streamlined the mutation discovery phase in lineage tracing. Scanning the hot-spots only gives a good chance of identifying markers, without scanning the whole mtDNA. A scanning of the hot-spots only represents 30 CTCE fragments, compared to 192 fragments for the whole mtDNA. This greatly increased the number of samples scanned, or reduced the necessary time. Furthermore, sequencing mutations identified with CTCE also offered another independent validation of the method. Last, it reinforced the validity of one of the hypothesis made in lineage tracing: "No independent occurrence of identical mutations." The fact that not a single one of the sequenced mutations was identical (even when many of them were on same fragment), greatly increased our certainty in the hypothesis.

7.1.3 Personal Contribution

Christian Arstad arrived in the lab shortly after we were certain we had a solid method to analyse and detect mutations. My supervisors and myself already were wondering about the distribution of mutations in the mitochondrial genome, so his arrival and proposition to take up the challenge came at a perfect time. Having designed the method, I assisted him to set up the experimental plan, and perform some of the original mtDNA scanning. I also contributed to analysing some of the data and writing the paper.

7.1.4 Status

Published on the 11th October 2016, in Mitochondrial DNA Part A. [94]

7.2 Resulting Publication: Intraperitoneal exfoliated tumor cells detected after surgical resection of rectal cancer. A prognostic factor on survival?

7.2.1 Short description

Intraperitonealcancer cells (IPCC) can be found in the peritoneal cavity following surgical removal of colon and rectal cancer. Part of the standard operating procedure, is to wash the peritoneal cavity with saline solution following surgical removal of the tissue. The fluid is then removed and can be used for analysis, a procedure known by its french name: "*lavage*". There is general debate in the community as to the effect of IPCC in cancer relapse, for example by being the source responsible for local metastasis or regrowth (a detailed literature review is

7.2. Resulting Publication: Intraperitoneal exfoliated tumor cells detected after surgical resection of rectal cancer. A prognostic factor on survival?

given in the paper Arstad et al, currently in press). In a previous paper by my co-supervisor Per Ekstrom, an attempt was made to detect IPCC using Kras mutations [95]. The results were difficult to interpret due to the low prevalence of Kras mutations. I was using mtDNA as a marker of tumor lineages, and so the idea came that they could also be used for the detection of IPCC. The material and extracted DNA was already available from the previous Kras study, and the method to perform the experiment existed. Furthermore, the hot-spots of mtDNA in cancers were known from the previous application, which meant that only the hot-spots could be scanned for markers. DNA extracted from primary tumors and from lavage fluid was available for 191 patients. Out of these 191 patients analyzed, 138 (72%) had mtDNA mutations in their primary tumors, out of which 45 (33%) had the same mutation detected in the lavage fluid, Thus guaranteeing the presence of IPCC. No relationship was found between the presence of IPCC and cancer survival, with an observation time of 5 years. The conclusion was that, we could well detect the presence of IPCC using a linear tracing approach, but IPCC has no effect on survival. Metastasis and local recurrence are therefore not a consequence of the presence of IPCC after surgery.

7.2.2 Significance

This paper demonstrated how CTCE and lineage tracing can be applied to clinical questions and applications. If the presence of IPCC was a prognostic factor for post-surgery survival, then the lineage tracing method we developed would have an immediate clinical application. In the other case, an important question of cancer metastasis would be answered: "Whether IPCC are responsible for the seeding of metastasis." In detecting the presence of the same mutation in both primary tumor and lavage fluid, the ability of our CTCE method to detecting IPCC was established. The results obtained did not support the theory that IPCC presence was a prognosis factor for tumor development. This invalidates theories that would see IPCC involved in the development of metastasis. An important finding that shows the ability of mtDNA mutations based lineage tracing methods to answer important questions of oncology.

7.2.3 Personal Contribution

Christian Arstad was at this point fully confident in the use of CTCE to detect and quantify mtDNA mutations. He therefore performed all of the experimental work himself. I contributed to the experimental design, statistical analysis of results and results interpretations. I also significantly contributed to writing the manuscript and responding to reviewers.

7.2.4 Status

Published on the 7th of June 2017, in BMC cancer [96].

8 Measuring Mitochondrial DNA Copy Number

The number of mtDNA copies in cells is currently unknown for most cell types. There is also speculation that the copy number could change in cancer tissues [97, 98]. Mathematical modelling of how mutations arise and become homoplasmic place a major emphasis on the importance of copy number [5, 4]. A significant variation in copy number from normal tissue to tumor tissue, could change the probability of mtDNA mutations becoming homoplasmic, or reaching a high fraction of the total number. In the interpretation of the mutational analysis, it was therefore important to have approximations of the copy number in our sampled tissue. We also needed to answer the concern that the areas that were found to carry a mutation at a homoplasmic level had a much lower copy number of mtDNA per cell than the other areas. A result that would compromise our hypothesis of lineage tracing that independent identical mutation at high fraction is a very rare event. The standard way in which relative copy number is quantified is real time PCR (rtPCR). When we first attempted to use rtPCR, we noticed high variations in the results. It was therefore decided that more investigation in the precision of rtPCR was needed if we wanted estimates of mtDNA copy numbers in our samples. The results of this research can found in section 8.1.

8.1 Resulting Publication: Quantifying mitochondrial DNA copy number using robust regression to interpret real time PCR results

8.1.1 Short description

This is a technical note, which describes an improved method to analyse rtPCR output data. The standard way to analyse rtPCR data is using the ΔC_T method which suffers from two major drawback: assuming that the efficiency is the same for all samples compared, and requiring cumbersome dilution series to estimate PCR efficiency. The first assumption is either wrong, or requires optimisation and cumbersome dilution series to validate. The robust regression method proposed in this note automatically isolates the exponential portion of the rtPCR output, and

Chapter 8. Measuring Mitochondrial DNA Copy Number

therefore measures PCR efficiency. This permits easy measurement of efficiency, without dilution, as well as the comparison between samples with different efficiencies.

8.1.2 Significance

The robust regression method was used to estimate mtDNA copy number in tumor and normal tissue. A possible explanation for the mtDNA mutations found in tumors could have been a much smaller number of mtDNA, and thus a much higher probability of a mutation reaching a high fraction. Verifying that there was no significant difference in mtDNA copy number between normal tissue and tumor tissue invalidated that explanation. In doing so, the hypothesis of no independently occurring high fraction mtDNA mutation was reinforced. The method described, allows the comparison of fewer samples than the ΔC_T method, which made possible the estimation of mtDNA copy number in LCM samples. The technique was therefore used to compare a sample with a homoplasmic mutation with wild type samples in the paper described in section 9.2. In a more general sense, the availability to the community of an R package that can make the analysis ad-hoc. Increased use of robust regression to analyse rtPCR results would result in a general improvement in the accuracy and simplicity of rtPCR as a method for research or diagnostic.

8.1.3 Personal Contribution

Per Ekstrom and myself designed the primers for rtPCR. I was then responsible for optimising them and performing the experimental work. Stephan Morgenthaler had the Idea of using a robust regression. I did all the implementation, testing and optimisation of the Data analysis method, as well as most of the writing.

8.1.4 Status

Accepted with minor revisions the 7th of July 2017, in BMC Research Notes.

Quantifying mitochondrial DNA copy number using robust regression to interpret real time PCR results

Paulo Refinetti^a, David Warren^c, Stephan Morgenthaler^a, Per O. Ekstrøm^b

^a*École Polytechnique Fédérale de Lausanne, EPFL FSB STAP, Station 8, Lausanne, Switzerland*

^b*Department of Tumor Biology, Norwegian Radiumhospital, Oslo, Norway*

^c*Department of Medical Biochemistry, Norwegian Radiumhospital, Oslo, Norway*

Abstract

Background Real time PCR is a widely used method to quantify relative DNA copy number. The data analysis of rtPCR output is still left with challenges. The most widely used method is the ΔC_t method, which requires all of the compared DNA fragments analysed to have the same PCR efficiency. An alternative method, that relies on fitting a robust regression to the rtPCR signal is proposed. The robust regression automatically fits the exponential portion of the curve. This methods reduces potential biases and allows the comparison of DNA fragments with different PCR efficiencies.

Findings The results from robust regression and ΔC_t methods are similar. The spread of results when the same PCR is performed 96 times is similar and can be approximated by a normal distribution in both cases. Estimating the efficiency with the ΔC_t method requires a dilution series, which is not required for the robust regression method. rtPCR is the used to quantify mtDNA copy number in three different tissues types: breast, colon and prostate, as well as associated tumors taken from the same patient, for a total of six tissues (from three patients). The mitochondrial copy number per cell is estimated between 200 and 300 copies per cell. When analysing the same tissue with the robust regression and the ΔC_t method, similar results are obtained, although the confidence ratio is slightly better for the robust regression.

Conclusion Using a robust regression to identify the exponential phase does not identical PCR efficiency when computing relative DNA concentration. This removes the requirement of optimising the reaction before as well as remove a source of potential bias. The method used has been implemented

as an R package for convenient use.

Keywords: , rtPCR, robust regression, mitochondrial DNA

1. Background

Mitochondrial DNA (mtDNA) is composed of a single circular chromosome of 16569 base pairs (hg38, GRCh38, Dec. 2013). It encodes for 22 tRNA, 13 protein subunits and two ribosomal RNA subunits. The proteins are key proteins of the oxidative phosphorylation chain. The tRNA and ribosomal subunits are required to process the genes and synthesize the proteins. Better measurements of mtDNA copy numbers would improve the understanding of mtDNA mutagenesis [1, 2, 3]. Furthermore, it is important to understand the growth and distributions of somatic mutations in mtDNA, which is crucial for gaining insights into the transition from heteroplasmy to homoplasmy in tumor cells, cell lineage in normal tissue and questions relating to metabolic changes in tumor development [4]. The symptoms and severity of mitochondrial diseases depends on the mutant fraction and on the tissue [5, 6, 7, 8]. Mitochondrial mutations also appear to be involved in cancer development [9, 10, 11, 12, 4], and aging [13, 14, 15, 16]. Somatic mitochondrial mutations are an important tool for lineage tracing [2, 17, 18, 19]. Most methods depend on the detection of mutation that have achieved homoplasmy or are very close. Accurate copy number estimates are important in the cellular mitochondrial dynamics, and therefore in the process through which cells become homoplasmic. The standard method for quantifying DNA copy number is real time PCR (rtPCR)[20, 21, 22]. Most methods rely on amplifying a mitochondrial and a nuclear fragment in separate reactions, with the template from the same sample[22, 12]. There has been much development in the analysis of rtPCR output, however, there are still challenges [23, 24].

Problem

The output of rtPCR reaction is the fluorescence measured at the end of each PCR cycle, which is proportional to the amount of double stranded DNA in the sample. It is a measure of the PCR reaction progression. Three phases can be observed:

Phase I: Lag phase: The signal is too low for the detector, only the noise is visible.

Phase II: Exponential phase: Signal grows exponentially with the number of cycles.

Phase III: Saturation phase: Signal increases sub-exponentially, or not at all as the PCR reaction saturates.

The dynamics of the PCR reaction can only be observed during Phase II. The signal can be modelled by an exponential function: $S = \alpha N E^c$. In this equation, N is the number of DNA copies at the start of the experiment, S is the signal, α is an unknown constant relating the copy number to the signal intensity and c is the cycle number. The constant α is related to parameters such as detection efficient or fluorescence per base pair, and it is assume that this value is constant and does not depend on the sample. The standard method to analyse rtPCR is the ΔC_T method. A signal threshold T is chosen, a little above the noise level. The C_T value is defined as the cycle number at which the signal crosses the threshold. It is calculated by taking a linear interpolation between the first signal value above the threshold and the one immediately below, then taking as C_T the value at which the line intersects the chosen threshold. If there are two samples, A and B , for which rtPCR signal have been obtained, threshold yields an equation relating the initial copy numbers of the two samples.

$$T = \alpha N_A E_A^{C_T^A} = \alpha N_B E_B^{C_T^B} \text{ or } \frac{N_A}{N_B} = \frac{E_B^{C_T^B}}{E_A^{C_T^A}}$$

Assuming equal efficiency for both reactions, $E_A = E_B = E$, the equation becomes

$$\frac{N_A}{N_B} = E^{\Delta C_T}$$

Where $\Delta C_T = C_T^A - C_T^B$. The ΔC_T method has a few clear flaws, which have already been pointed out and demonstrated by Karlen et al [23]. The first one is the assumption of equal efficiency which is essential to this method. If the fragments used are not the same, as is the case for the quantification of mtDNA, the reaction needs to be optimised to have equal efficiency. If PCR efficiency depends on initial DNA concentration, as some results suggests [22], this would introduce errors in the measurements.

Proposed Solution

The objective of rtPCR is to measure the relative initial copy number $\frac{N_A}{N_B}$ between two samples. The exponential phase is linear in the cycle number, taking the logarithm results in:

$$\log_2 S = c \log_2 E + \log_2(\alpha N), \log_2(\alpha N) = I \text{ the Intercept of the curve}$$

The logarithm is taken to be in base two unless indicated otherwise. If based on several consecutive values of (c,S), we determine the coefficients of this linear equation, the intercept term (the value at c=0) measured for two samples A and B yields the formula.

$$\frac{N_A}{N_B} = 2^{I_A - I_B}$$

The above formula relies only on the constancy of α and does not require the efficiency to be the same. It was shown that methods that rely on observed values of (c,S) during the exponential phase and a regression fit to obtain the constants perform much better than the ΔC_T method [23]. Furthermore, the slope of the regression determines the logarithm of the efficiency without relying on diluting the sample. The same exact mix could be analysed multiple time and the efficiency measured with greater precision. Because unequal efficiencies are permissible, the fitting of a regression line reduces the difficulties in optimising the PCR reactions and improves precision. Previous methods relied on the manual selection of the exponential phase [23, 24, 25, 26], or on cumbersome mathematical models. The solution proposed is to use a highly robust linear regression on the log of the signal. The robust regression minimises the median of the square residuals, instead of the mean squared residuals (Least squared method). Results show it reliably identifies the exponential phase of the rtPCR signal and gives a good measurement of the efficiency E and intercept I for each reaction. Using multiple replicas of the same experiment and taking averages, the precision of the estimates can be increased.

2. Materials and Methods

Tissue and DNA extraction

Anonymous surgical discharges were obtained after standardised informed consent. Tissue was stored at the surgical department at -70°C until DNA

extraction. According to Norwegian Law, technical and methodological development work that uses anonymised biological material does not require approval from research ethics committees (Web page, last access November 2016). Normal and tumor tissue was obtained from three different patients with 3 different tumor types (breast, prostate and colon). The normal tissue was taken at a distance of 10 to 15cm from the location of the tumor. A few milligrams were taken from each sample and subjected to DNA extraction.

DNA extraction

Samples were digested with proteinase K for 4 hours at 57°C in 300 μ l of digestion Buffer (Qiagen, Hilden, Germany) according to manufacturer's instructions. DNA was extracted from them using the Qiagen MagAttract DNA Mini-M48 Kit with a dedicated automatic solution also provided by Qiagen. The result is a DNA solution containing approximately 50ng of DNA per μ l .

Primers

Primers were designed using the rtPCR primer design tool of IDT (Integrated DNA technologies). The tool designs primers with similar efficiency, annealing temperature, and attempts to avoid the formation of tertiary structures. The nuclear and mitochondrial primer pairs were designed for simultaneous amplification. Table 1 shows the primer pairs.

Primer amplification conditions were then optimised to allow the same PCR conditions to be used for both pairs simultaneously. When designing the primers, attention was paid to homologies between nuclear and mitochondrial DNA sequences. The mitochondrial primer was chosen so that it could not amplify in the nuclear genome (and vice versa).

rtPCR condition

Real time PCR was performed using a BioRad CFX connect Real-time PCR detection System. PCR recipe was 2X Perfecta SYBR Green SuperMix for iQ (QuantaBio, Beverly, MA, USA), 0.2 μ M of each primer, for a final volume of 20 μ l . The PCR temperature cycling used: Initial denaturing at 94°C for 4min, followed by 45 cycles of denaturing at 94°C for 30 sec, annealing at 60°C for 30 sec and extension at 72°C for 1min.

Experiment

For both the mitochondrial and the nuclear primers, 96 replicas (a whole plate) of the same identical rtPCR are produced. A 2ml PCR mix was created (as described the section "rtPCR Condition"), to which $4\mu\text{l}$ of extracted DNA was added. The mix was spread on a PCR plate adding $20\mu\text{l}$ of it into each well. Serial dilutions for both primers were used to estimate PCR efficiency with the ΔC_T method, testing for changes in PCR efficiency and give an indication of precision. For both primers, the initial rtPCR mix is serial diluted into rtPCR mix without DNA, by a factor 5, for 6 steps. There are 16 replicas for each dilution for a total of $16 \times 6 = 96$ reactions. As an internal consistency check, the mtDNA copy numbers is measured for each tissue on extracted DNA and extracted DNA diluted 10. Each tissue therefore has 4 different rtPCR reactions: Nuclear DNA, mitochondrial DNA, nuclear DNA diluted by 10, and mitochondrial DNA diluted by 10. Each rtPCR reaction has 24 replicas, thus the total number is $4 \times 24 = 96$ (a complete 96 well plate).

Data analysis

Data analysis is performed using an R package developed specifically for the analysis of rtPCR results. The package, together with the codes used to generate the graphs and tables are included in the supplementary materials. The efficiency and intercept associated with each rtPCR reaction is estimated by fitting a robust linear regression on the base two logarithm of the signal. First the middle point is determined as the couple $(c_m, \log S_m)$, where $\log S_m$ is closest to $\frac{\max \log S + \min \log S}{2}$. Forcing passage through the middle point ensures the regression fits the exponential phase. This only requires the fitting of the slope.

The relative copy number between two experiments is defined as $\frac{N_A}{N_B} = 2^{I_A - I_B}$ which is estimated by taking the difference in the mean intercept,

$$\text{estimation of } \left(\frac{N_A}{N_B}\right) = 2^{(\bar{I}_A - \bar{I}_B)} = 2^{\hat{\Delta}I}$$

The intercept is assumed to follow a Normal distribution, which is justified by inspection of the results from 96 replicas. The 95% confidence interval for ΔI can therefore be estimated as:

$$C.I. = \hat{\Delta}I \pm W; W = q_{0.975}(t_{n_A+n_B-2}) \times \sqrt{\frac{\text{Var}[I_A]}{n_A} + \frac{\text{Var}[I_B]}{n_B}}$$

Where $q_{0.975}$ is the 97.5% quantile function, t_ν is the t distribution with ν degrees of freedom, and n_A and n_B are the number of replicas for A and B respectively and the variances are estimated from the replicated values. From this confidence interval, one derives an interval for the relative copy number of the form of $\frac{N_A}{N_B} \times 2^{\pm W}$, which is in fact a confidence ratio. 2^W is the maximal relative error in the copy number due to statistical uncertainty. One can interpret the $C.R.$ as follows:

$$P(C_{\text{relative}} \in [\frac{N_A}{N_B} - C.R., \frac{N_A}{N_B} \times C.R.]) = 95\%$$

In sum, the boundaries for confidence interval of the actual relative concentration, can be calculated by multiplying and dividing the estimated relative concentration by the $C.R.$. The baseline noise in a rtPCR reaction is estimated by taking the highest point for which the first derivative of the signal as a function of the cycle number is negative. The threshold to calculate the C_T value is chosen by taking the highest value for the baseline in an experiment. When relative concentrations are calculated between 2 samples, the same threshold was used to calculate the C_T value.

3. Findings and Discussion

Figure 1 shows the results of repeating 96 times the same reaction. The efficiency and intercept, calculated using the robust regression, as well as the C_T values are shown. In all three cases, the values group together with 5 outliers. These outliers are not PCR failures. They represent genuine variation in PCR performance on identical rtPCR mix. These results justifies the use of a normal approximation.

Figure 2 shows the result of the dilution series. For both nuclear and mitochondrial DNA, the measured concentration relative to the first point was calculated using the ΔC_T method and the robust regression method. The vertical axis is the logarithm base five of the relative concentration. The slope of the curve should be -5 , by construction.

It can be seen that the robust regression gives, in both cases, a slightly better slope than the C_T method. This difference does not appear to be significant. The dilution series can be used to estimate the efficiency of the PCR using the C_T method. If the efficiency is assume identical in all samples:

$$\frac{N_A}{N_B} = E^{\Delta C_T} \Rightarrow \log_5 \frac{N_A}{N_B} = \Delta C_T \log_5 E$$

The logarithm of the dilution factor, is linearly related to the ΔC_T , and the slope is the log of the Efficiency. Results can be seen in figure 3.

The estimates of mtDNA copy numbers can be seen on table 2. The estimates of mtDNA copy number are constant, ranging from 100 to 150 copies of mtDNA to nuclear DNA. The values need to be multiplied by 2 to obtain the number of mitochondria per cell. The confidence ratios are around 1.3. The ratio between the measured concentrations of mtDNA, and diluted mtDNA should be 10, by construction. The same is true for nuclear DNA. Taking into account the confidence ratios associated with the measurement, the diluted samples have indeed a copy number 10 times below their un-diluted counter-parts. Observing a CR of 1.3 is and a mitochondrial copy number of 200, corresponds to having a 95% certainty that the real value lies from 150 and 260. This precision was achieved with 24 replicas. The CR decays very slowly as a function of the number of samples. Using a robust regression to analyse rtPCR data presents major advantages over the ΔC_T method. First, it does not make the assumption of identical PCR efficiency between two samples. This reduces potential biases, and allows for the comparison of fragments/samples with clearly different efficiencies. It also allows the estimation of PCR efficiency without performing dilution series. There is reason to suspect that the efficiency and the initial concentration are not independent. To illustrate this point, the efficiency of the dilution series is shown in figure 3. If indeed the efficiency depends on the initial copy number, it would be a major flaw in the estimates made with the ΔC_T value. For comparison, the prostate tumor tissue is analysed using the ΔC_T method, and the results can be seen in table 3. The results are slightly different than those estimated using the robust regression. They are, however, coherent, if the larger confidence ratio is taken into account. It might also explain that the result measured, is much higher than the one observed using the robust regression. Karlen et al [23] also shows that the ΔC_T method performs well in the case of identical efficiencies, but perform poorly in other circumstances.

The robust regression method offers an alternative way to analyse rtPCR data. It's major advantage is not requiring identical efficiency between samples. This removes the necessity of PCR optimisation as well as remove a potential source of bias.

List of abbreviations used

- rtPCR: real time polymerase chain reaction

- DNA: deoxyribonucleic acid
- mtDNA: mitochondrial DNA
- tRNA: transcription ribonucleic acid

- [1] Coller, H.A., Khrapko, K., Bodyak, N.D., Nekhaeva, E., Herrero-Jimenez, P., Thilly, W.G.: High frequency of homoplasmic mitochondrial DNA mutations in human tumors can be explained without selection - Nature Genetics. *Nature Genetics* **28**(2), 147–150 (2001)
- [2] Taylor, R.W., Barron, M.J., Borthwick, G.M., Gospel, A., Chinnery, P.F., Samuels, D.C., Taylor, G.A., Plusa, S.M., Needham, S.J., Greaves, L.C., Kirkwood, T.B.L., Turnbull, D.M.: Mitochondrial DNA mutations in human colonic crypt stem cells. *Journal of Clinical Investigation* **112**(9), 1351–1360 (2003)
- [3] Marcelino, L.A., Thilly, W.G.: Mitochondrial mutagenesis in human cells and tissues. *Mutation Research/DNA Repair* **434**(3), 177–203 (1999)
- [4] Brandon, M., Baldi, P., Wallace, D.C.: Mitochondrial mutations in cancer. *Oncogene* **25**(34), 4647–4662 (2006)
- [5] Zeviani, M.: Mitochondrial disorders. *Brain* **127**(10), 2153–2172 (2004)
- [6] Schaefer, A.M., McFarland, R., Blakely, E.L., He, L., Whittaker, R.G., Taylor, R.W., Chinnery, P.F., Turnbull, D.M.: Prevalence of mitochondrial DNA disease in adults. *Annals of neurology* **63**(1), 35–39 (2008)
- [7] Shoffner, J.M., Wallace, D.C.: Oxidative phosphorylation diseases and mitochondrial DNA mutations: diagnosis and treatment. *Annual review of nutrition* (1994)
- [8] Brown, M., Starikovskaya, E., Derbeneva, O., Hosseini, S., Allen, J., Mikhailovskaya, I., Sukernik, R., Wallace, D.: The role of mtDNA background in disease expression: a new primary LHON mutation associated with Western Eurasian haplogroup J. *Human Genetics* **110**(2), 130–138 (2002)

- [9] Lam, E.T., Bracci, P.M., Holly, E.A., Chu, C., Poon, A., Wan, E., White, K., Kwok, P.-Y., Pawlikowska, L., Tranah, G.J.: Mitochondrial DNA sequence variation and risk of pancreatic cancer. *Cancer research* **72**(3), 686–695 (2012)
- [10] Carew, J.S., Huang, P.: Mitochondrial defects in cancer. *Molecular Cancer* **1**(1), 9 (2002)
- [11] Chatterjee, A., Mambo, E., Sidransky, D.: Mitochondrial DNA mutations in human cancer. *Oncogene* **25**(34), 4663–4674 (2006)
- [12] Grady, J.P., Murphy, J.L., Blakely, E.L., Haller, R.G., Taylor, R.W., Turnbull, D.M., Tuppen, H.A.L.: Accurate Measurement of Mitochondrial DNA Deletion Level and Copy Number Differences in Human Skeletal Muscle. *PloS one* **9**(12), 114462 (2014)
- [13] Kujoth, G.C.: Mitochondrial DNA Mutations and Apoptosis in Mammalian Aging. *Cancer research* **66**(15), 7386–7389 (2006)
- [14] Greaves, L.C., Turnbull, D.M.: Mitochondrial DNA mutations and ageing. *Biochimica et biophysica acta* **1790**(10), 1015–1020 (2009)
- [15] Lee, H.-C., Chang, C.-M., Chi, C.-W.: Somatic mutations of mitochondrial DNA in aging and cancer progression. *Ageing Research Reviews* **9**, 47–58 (2010)
- [16] Melova, S., Schneider, J.A., Coskun, P.E., Bennett, D.A., Wallace, D.C.: Mitochondrial DNA rearrangements in aging human brain and in situ PCR of mtDNA. *Neurobiology of aging* **20**(5), 565–571 (1999)
- [17] Fellous, T.G., McDonald, S.A.C., Burkert, J., Humphries, A., Islam, S., De-Alwis, N.M.W., Gutierrez-Gonzalez, L., Tadrous, P.J., Elia, G., Kocher, H.M., Bhattacharya, S., Mears, L., El-Bahrawy, M., Turnbull, D.M., Taylor, R.W., Greaves, L.C., Chinnery, P.F., Day, C.P., Wright, N.A., Alison, M.R.: A methodological approach to tracing cell lineage in human epithelial tissues. *Stem cells (Dayton, Ohio)* **27**(6), 1410–1420 (2009)
- [18] Blackwood, J.K., Williamson, S.C., Greaves, L.C., Wilson, L., Rigas, A.C., Sandher, R., Pickard, R.S., Robson, C.N., Turnbull, D.M., Taylor, R.W., Heer, R.: In situ lineage tracking of human prostatic epithelial

- stem cell fate reveals a common clonal origin for basal and luminal cells. *The Journal of pathology* **225**(2), 181–188 (2011)
- [19] Walther, V., Alison, M.R.: Cell lineage tracing in human epithelial tissues using mitochondrial DNA mutations as clonal markers. *Wiley interdisciplinary reviews. Developmental biology* **5**(1), 103–117 (2016)
- [20] Nicklas, J.A., Brooks, E.M., Hunter, T.C.: Development of a quantitative PCR (TaqMan) assay for relative mitochondrial DNA copy number and the common mitochondrial DNA deletion in the rat - Nicklas - 2004 - *Environmental and Molecular Mutagenesis* - Wiley Online Library. *Environmental and ...* (2004)
- [21] Andreu, A.L., Martinez, R., Marti, R., García-Arumí, E.: Quantification of mitochondrial DNA copy number: Pre-analytical factors. *Mitochondrion* **9**(4), 242–246 (2009)
- [22] Fernandez-Jimenez, N., Castellanos-Rubio, A., Plaza-Izurietta, L., Gutierrez, G., Irastorza, I., Castaño, L., Vitoria, J.C., Bilbao, J.R.: Accuracy in copy number calling by qPCR and PRT: a matter of DNA. *PloS one* **6**(12), 28910 (2011)
- [23] Karlen, Y., McNair, A., Perseguers, S., Mazza, C., Mermod, N.: Statistical significance of quantitative PCR. *BMC bioinformatics* **8**(1), 131 (2007)
- [24] Tichopad, A.: Standardized determination of real-time PCR efficiency from a single reaction set-up. *Nucleic acids research* **31**(20), 122–122 (2003)
- [25] Timken, M.D., Swango, K.L., Orrego, C., Buoncristiani, M.R.: A duplex real-time qPCR assay for the quantification of human nuclear and mitochondrial DNA in forensic samples: implications for quantifying DNA in degraded Samples. *Journal of Forensic science* (2005)
- [26] Ruijter, J.M., Ramakers, C., Hoogaars, W.M.H., Karlen, Y., Bakker, O., van den Hoff, M.J.B., Moorman, A.F.M.: Amplification efficiency: linking baseline and bias in the analysis of quantitative PCR data. *Nucleic acids research* **37**(6), 45–45 (2009)

Figures

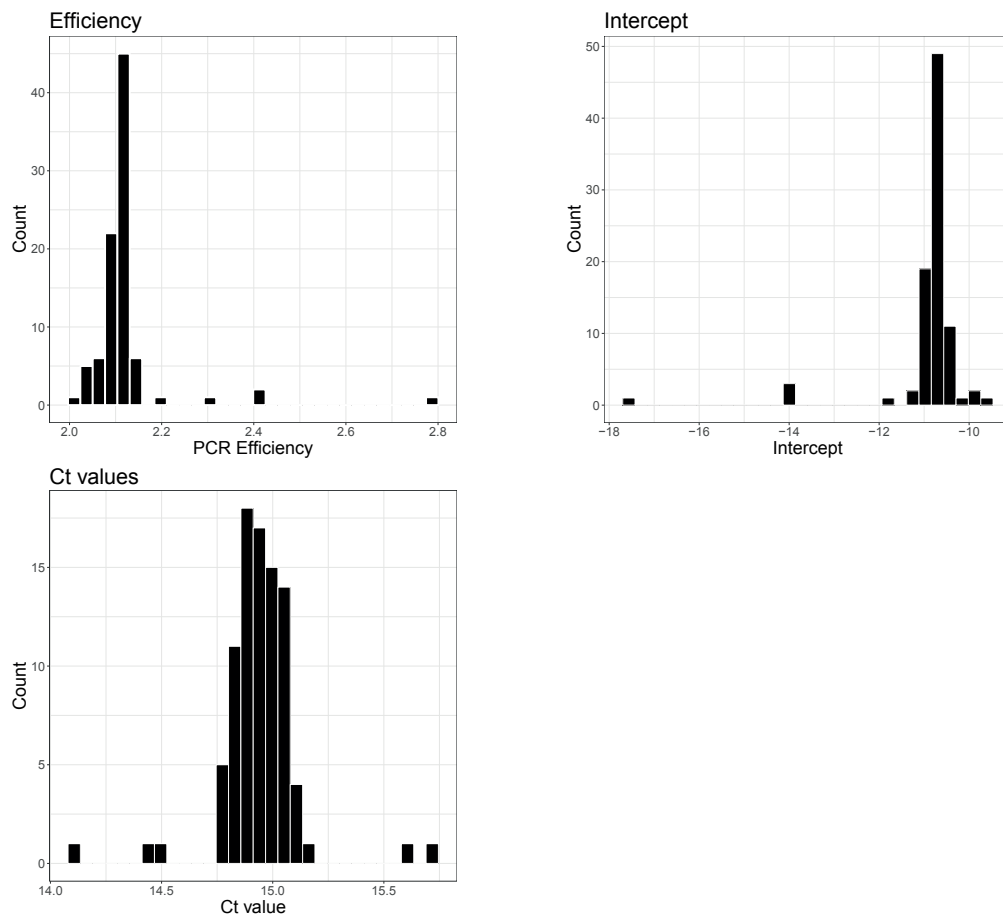


Figure 1: Histograms showing the distribution of Efficiency, intercept and C_T . Data shown from 94 replicas. For each reaction, the robust regression method is used to compute the efficiency and intercept.

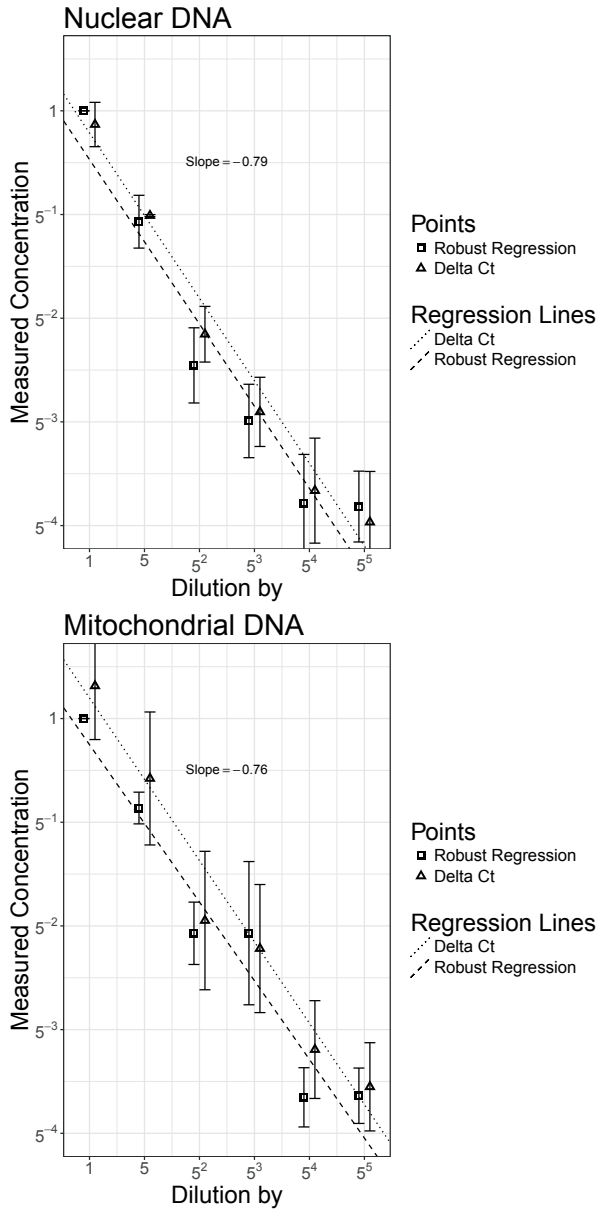


Figure 2: Dilution series: DNA concentration in the PCR mix is serially diluted by a factor 5, six times. The DNA concentration is measured using the robust regression method and the ΔC_T method. Each concentration is prepared in 16 replicas. The same procedure is reproduced for both nuclear and mitochondrial DNA. Regression lines are fitted using the least squared method. The data has been shifted slightly left and right of the true value for illustration purposes.

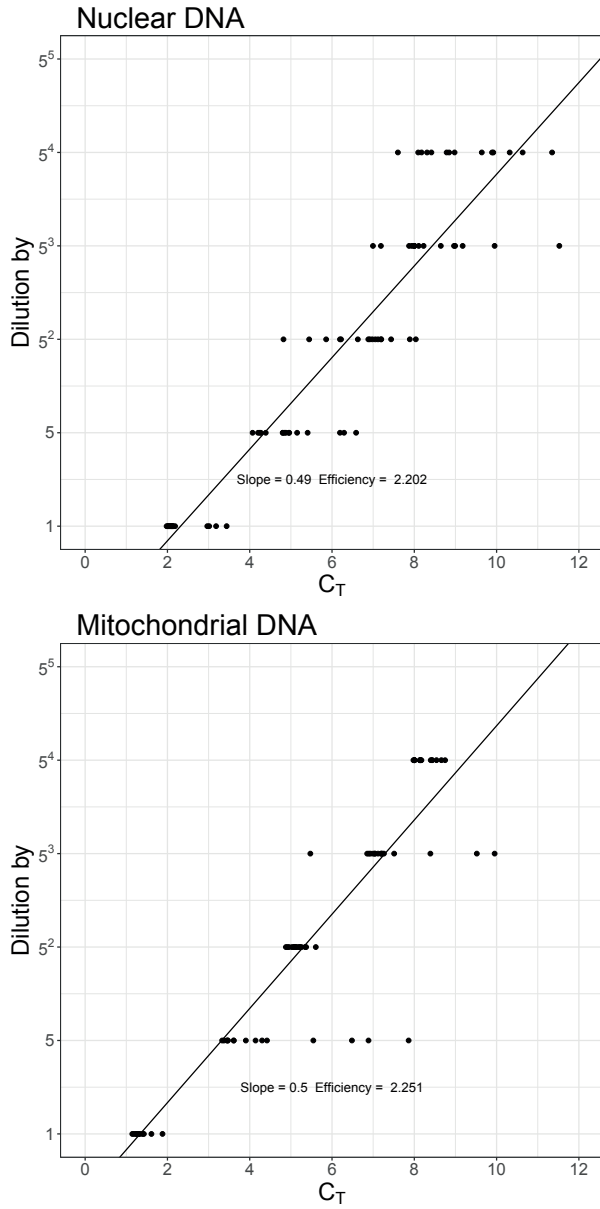


Figure 3: Calculating the efficiency using the ΔC_T method. The measured C_T values for each dilution are plotted. The slope of the fitted line can be used to estimate the efficiency.

Tables

Primer	Sequence (5' to 3')	Genome region
Mitochondrial Forward	ACA CCC TCC TAG CCT TAC TAC	chrM: 10087 - 10192
Mitochondrial Reverse	GAT ATA GGG TCG AAG CCG C	
Nuclear Forward	AGG GTA TCT GGG CTC TGG	chr11: 2170993-2171170
Nuclear Reverse	GGC TGA AAA GCT CCC GAT TAT	

Table 1: Primer sequences used to amplify nuclear and mitochondrial DNA sequences.

Tissue	Sample Type	Efficiency	Concentration	Relative Error
Breast Tumor	mtDNA	2.08	129.3	1.29
	mtDNA/10	2.13	6.87	1.47
	Nuclear	2.02	1	1
	Nuclear/10	1.99	0.16	1.4
Breast Normal	mtDNA	2.19	132.13	1.4
	mtDNA/10	2.02	38.54	1.36
	Nuclear	2.03	1	1
	Nuclear/10	2.02	0.09	1.38
Colon Tumor	mtDNA	2.13	132.72	1.35
	mtDNA/10	2.1	12.95	1.32
	Nuclear	2.01	1	1
	Nuclear/10	2.01	0.09	1.45
Colon Normal	mtDNA	2.11	219.19	1.5
	mtDNA/10	2.11	20.82	1.56
	Nuclear	2.05	1	1
	Nuclear/10	1.99	0.2	1.51
Prostate Tumor	mtDNA	2.17	134.14	1.41
	mtDNA/10	2.11	15.79	1.39
	Nuclear	2.04	1	1
	Nuclear/10	2	0.16	1.47
Prostate Normal	mtDNA	2.12	172.65	1.49
	mtDNA/10	2.08	22.21	1.47
	Nuclear	2.02	1	1
	Nuclear/10	2.01	0.13	1.55

Table 2: Relative concentration of samples for each tissue type. The nuclear DNA concentration is taken as reference, and is therefore 1. The relative error (confidence ratio) is also 1 as there is no uncertainty associated with it. The other concentrations are relative to nuclear DNA and the relative error associated with is. The values for the efficiency are the average efficiency over the replicas.

	Concentration	Relative Error
mtDNA	288.33	1.52
mtDNA/10	23.45	1.44
Nuclear	1.00	1.00
Nuclear/10	0.11	1.52

Table 3: Concentrations calculated using the ΔC_T method in Prostate Tumor sample. The confidence ratio is estimated using the t-interval.

9 Two-dimensional Lineage Tracing

At this point, we had a solid protocol for the analysis of mtDNA mutations in micro-anatomical samples. We understood the need to sample macroscopically beforehand, to identify the fractions that contained mutations that could be used as marker in the tumors. Our PCR reactions and CTCE conditions were optimised, as was our sample processing technique. However, we still held on to the idea that somehow, tumor areas could be identified. To identify tumor areas, we believed that a better sample preparation was needed. We still believed that one of the reasons that we failed in our first experiment was poor sample selection. Various tests with different staining were then made, and are described in the chapter 5. After many experiments and testings, no conclusive way of separating tumor cells from normal cells in the neighbourhood of the tumor was found. If it is impossible to define which areas in a tumor sample, are tumor in origin, and which are normal in origin, then any attempt at selecting structures based on morphology is bound to fail. The samples taken in the first experiment (chapter 4), were selected areas of the tumor mass, which together represented only a small fraction of the total area of a slice. Even if a mutation was present, it is possible we did not detect it. Also, the size of the sample was too large, as later results show. After coming up with an acceptable slice preparation protocol, experiments were made to characterise the behaviour of our LCM equipment. In previous experiments, samples were often lost due to miss falls. A 16 caps holder, sample collection was built and the protocol for collecting the samples and extracting the DNA was streamlined. Allowing 96 samples to be collected in 30min. Small samples often do not fall after being cut by the laser, possibly due to static electricity, or simply insufficient weight. Also, smaller samples contain less DNA, which results in a higher rate of PCR failure in the first round. A compromise therefore had to be found between sample size and resolution. Circles of $25000\mu\text{m}^2$ were selected as a good compromise. The same sample size was used from there onwards in all the LCM sampling.

Without ways of identifying tumor areas in samples, we decided to take slices across larger pieces of tissue, in which some portion of the tissue could be identified as normal. Scanning the whole mtDNA for each small LCM sample was deemed too time consuming. To identify pieces of tissues that were potential candidates for LCM analysis, slices of tissue were taken using the microtome. DNA extracted from a slice of tissue, contains a mix of mtDNA from all the cells in

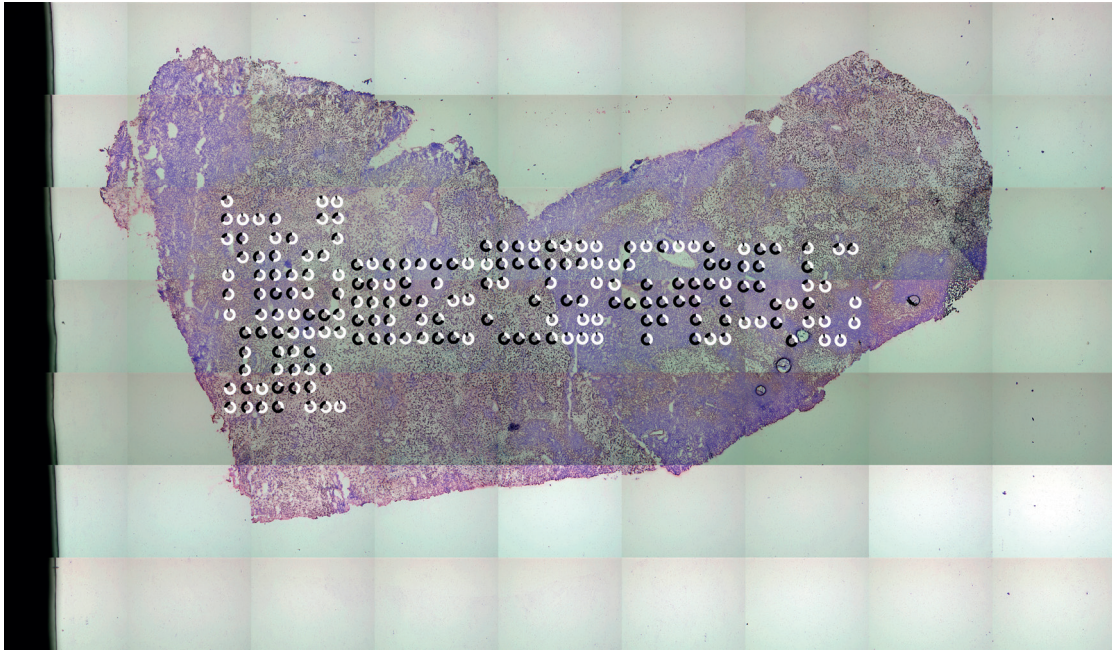


Figure 9.1: Results from the two-dimensional analysis of a liver adenoma. The mutant fraction in each sample is represented by fraction of black in each circle. The size and position of each sample taken is accurately represented by the circles.

that slice. Scanning the mtDNA mutation hotspots (identified in the paper 7.1) in the slice DNA identified the tissues that were potential candidates for LCM analysis. One example is the Leydig cell tumor, described in section 9.2. Two mutations were identified while scanning the mtDNA in the macroscopic slice, then almost 300 small samples were taken with LCM across the whole slice of tissue.

9.1 Liver tumors

Normal tissue can be identified as such from histologic analysis, when it is far removed from the tumor area. The non-normal area that includes the tumor can also be identified by histologic observation. The natural question is then what happens at the border. From a collaboration with the University hospital in Bologna, Italy, samples of liver tissues were obtained. These samples were large 2cm long samples, taken to cover the border between the tumor mass and the surrounding normal tissue. LCM sampling was then done across the section in order to see if any "border-phenomenon" could be observed. Using the same display syntax as in the paper shown in 9.2, an example of a currently unpublished result can be seen in figure 9.1. A total of 10 tumors have been analysed, and the analysis of the results is still in progress. Two other examples of liver analysis are shown in the paper described in section 9.2. They serve as a proof of concept to show that the method described in the article is applicable to other tissues.

The two dimensional lineage tracing method is being coupled with IHC analysis, and applied to

9.2. Resulting Publication: Tracing mitochondrial mutation in tumors

cholangiocarcinoma (bile duct cancer), hepatocellular carcinomas, adenomas of the liver, and focal nodular hyperplasia. Lineage tracing results might also reveal differences in the patterns of all three tumors. So far the results have not shown any clear border phenomenon, nor any clear differences between the tumor types.

9.2 Resulting Publication: Tracing mitochondrial mutation in tumors

9.2.1 Short description

This is the first major paper where lineage tracing results are published. It describes the two-dimensional analysis of a Leydig cell tumor. Two independent mutations are identified in a Leydig cell tumor by analysing the mtDNA hotspots on a macroscopic sample. 281 micro-anatomical samples are then taken scattered across a slice of the tissue. The same two fragments, in which a marker was identified, are then analysed in each of the fragments. The mutant fraction information is then displayed on top of the histologic image of the tissue. All mutant fractions were observed for both markers, ranging from wild type to homoplasmic mutant. There was a strong correlation between the mutant fraction of the two markers, where samples homoplasmic for one marker, were found to be homoplasmic for the other. Observing a high correlation between two markers suggests the presence of one cell lineage that carries both mutations. No correlation was observed between any identifiable structures and the presence, or absence, of mutations. In fact, the cell lineage traced, seems scattered across the tissue with no recognisable pattern. Finely sampling an area on another slice of the same tissue, showed a strong similarity between the position of the mutation, establishing that the cell lineage can be found on the same position of the tissue, slightly deeper.

9.2.2 Significance

In some ways, this is the first real lineage tracing paper. Detecting a mutation in a macroscopic sample, and then analysing it in micro-anatomical samples, showed no recognisable patterns. It was the observation of the same mutation in similar positions on another slice, sampled independently, that clearly established the fact that we had indeed a method for tracing cell lineages. Also, from reading the articles of Fialkow et al. the suspicion arose, that there was no clear way to identify tumor cells under histologic analysis. Observing that no recognisable patterns could be observed in the results validated this suspicion. It also opened many opportunities and asked many questions. Questions such as: "Is this pattern true for all tumors, or is it specific to this one?" "What fraction of a tumor volume is made up of tumor cells?" With the validated method in hand, many opportunities also arose to use it. If cell lineages can be identified on various slices of tissue, then maybe it is possible to take successive slices, and reconstruct in three dimensions the patterns formed by cell lineages in tumors. Maybe more data would allow a pattern of cell growth to emerge. Last, it was these observations that established that no selective

Chapter 9. Two-dimensional Lineage Tracing

sampling could be performed. All successive micro-anatomical sampling was performed by placing a standard grid of samples as small as they could be reliably taken using LCM.

9.2.3 Personal Contribution

The tissue was collected by Christian Arstad, who surgically removed the testicle from the patient, and snap froze the portion that was not required for pathological examination as procedure required. Sectioning staining and LCM was performed by myself, with the constant supervision and collaboration of Per Ekstrom. Data analysis and writing of the article was performed by Per Ekstrom and myself.

9.2.4 Status

Published on the 8th of April 2017 in BMC Clinical Pathology [99]

Mapping mitochondrial heteroplasmy in a Leydig tumor by laser capture micro-dissection and cycling temperature capillary electrophoresis

Paulo Refinetti¹, Christian Arstad², William G. Thilly³, Stephan Morgenthaler¹ and Per Olaf Ekstrøm²

Chair of Applied Statistics ⁽¹⁾, EPFL – FSB – STAP, 1015 Lausanne, Switzerland. Department of Tumor Biology ⁽²⁾, Institute for Cancer Research, The Norwegian Radium Hospital, Oslo, Norway. Laboratory in Metakaryotic Biology, Department of Biological Engineering ⁽³⁾, Massachusetts Institute of Technology, MA, USA.

Email Addresses:

- Paulo Refinetti (corresponding Author): paulo.refinetti@epfl.ch
- Christian Arstad: christian.arstad@studmed.uio.no
- William G. Thilly: thilly@mit.edu
- Stephan Morgenthaler: stephan.morgenthaler@epfl.ch
- Per O. Ekstrom: pok@rr-research.no

Keywords: capillary electrophoresis, mitochondrial mutations, heteroplasmy, homoplasmy, LCM, tumor heterogeneity

Abstract

Background: The growth of tumor cells is accompanied by mutations in nuclear and mitochondrial genomes creating marked genetic heterogeneity. Tumors also contain non-tumor cells of various origins. An observed somatic mitochondrial mutation would have occurred in a founding cell and spread through cell division. Micro-anatomical dissection of a tumor coupled with assays for mitochondrial point mutations permits new insights into this growth process. More generally, the ability to detect and trace, at a histological level, somatic mitochondrial mutations in human tissues and tumors, makes these mutations into markers for lineage tracing.

Method: A tumor was first sampled by a large punch biopsy and scanned for any significant degree of heteroplasmy in a set of sequences containing known mutational hotspots of the mitochondrial genome. A heteroplasmic tumor was sliced at a 12 μm thickness and placed on membranes. Laser capture micro-dissection was used to take 25000 μm^2 subsamples or spots. After DNA amplification, cycling temperature capillary electrophoresis (CTCE) was used on the laser captured samples to quantify mitochondrial mutant fractions.

Results: Of six testicular tumors studied, one, a Leydig tumor, was discovered to carry a detectable degree of heteroplasmy for two separate point mutations: a C \rightarrow T mutation at bp 64 and a T \rightarrow C mutation found at bp 152. From this tumor, 381 spots were sampled with laser capture micro-dissection. The ordered distribution of spots exhibited a wide range of fractions of the mutant sequences from 0 to 100% mutant copies. The two mutations co-distributed in the growing tumor indicating they were present on the same genome copies in the founding cell.

Conclusion: Laser capture microdissection of sliced tumor samples coupled with CTCE-based point mutation assays provides an effective and practical

means to obtain maps of mitochondrial mutational heteroplasmy within human tumors.

Introduction

The mitochondrial genome (mtDNA) is a circular DNA molecule [1,2] of around 16.5 kb. Multiple copies of mtDNA are found in each mitochondrion and a human cell may have between 100 and 10000 copies [3], depending on cell type. In early studies of somatic mtDNA mutations in tumors, it was suggested that somatic mitochondrial mutations can grow from a low frequency of occurrence (heteroplasmy) to become completely dominant (homoplasmy). Homoplasmy was reported to be found in more than half of human tumors [4-8]. Later studies, however, have found this number to be significantly lower [9,10]. Hypothetical paths to homoplasmy, however, can be offered as mathematical models based on random drift linked to mtDNA replication during cell division and in the absence of selection [5,11,12]. The basic hypothesis is that once a mutation arises in an organogenic stem cell as a single copy of mtDNA, this mutant copy may be retained in a stem cell lineage and by random distribution result in some stem cells to contain a significant fraction of a particular mutant mtDNA copy. If said stem cell becomes the first tumor stem cell then subsequent tumor stem cell divisions would be expected to continue random distribution of mutant copies so that after multiple divisions some descendant copies would contain no mutant copies (wild type homoplasmy) while others would have all copies carrying the mutant sequence (mutant homoplasmy) with most stem cells distributed with intermediate forms of mutant heteroplasmy. It is generally believed that most tumors are heterogeneous with infiltrations of non-tumor cells [13]. Sampling of only tumor-derived cells is therefore difficult. As non-tumor cells would not carry the mutation present in tumor cells it would

not be possible to distinguish between heteroplasmy and homoplasmy in a random tumor sample [14,15]. Pathological evaluation is often used to locate the tumor tissue. The success of this method depends on the experience of the pathologist and the quality of staining [8,16-18].

Somatic mtDNA mutations have been used as a marker of clonality in human tissues or tumors [11]. A successful method has been the immuno-histochemical marking of cytochrome oxidase (COX) deficient cells [19-21].

In Taylor et al. [3], selected individual colonic crypts were sequenced and sixty mitochondrial mutations were detected, of which half were homoplasmic. None of the mutations were identical. The probability of independent identical homoplasmic mutations can be assumed to be very low, as is also predicted by the mathematical model [11].

Laser capture microdissection (LCM) was introduced in 1996 [22]. The method uses a laser to cut selected areas of tissue resting on a membrane. The area of interest can be collected by gravity or by adherence and is transferred to post processing vials. LCM has been used to capture spots of tissue for the analysis of proteins, mRNA and DNA [23-26].

This study describes a method to trace somatic mitochondrial mutations through a tumor slice. In a first step, DNA from a large representative piece of tissue is analyzed to detect the presence of mtDNA mutations. If mutations are found, LCM is used to obtain several small spots of the tissue. Cycling capillary temperature electrophoresis (CTCE) [27-30] is used to detect and quantify mutant fractions in each sample. In the course of this study, more than 900 LCM samples were analyzed for the presence of mtDNA mutations.

Materials and Methods

Tissue samples

Surgical discharges were collected under informed consent at the surgical departments of Bærum sykehus (Vester Viken, Helse sør-øst, Norway) in 2015. Similarly, liver samples were collected as surgical discharges under informed consent at the University Hospital in Bologna (Policlinico Sant'Orsola-Malpighi, Bologna, Italy) as part of a different study.

According to the regional ethics comity (REC), “technical and methodological development work that uses anonymized biological material” does not require approval from REC

(https://helseforskning.etikkom.no/ikbViewer/page/reglerogrutiner/soknadsplikt/sokerikkerek?p_dim=34999&_ikbLanguageCode=us).

Six samples of testicular cancer, five germ cell tumors, and one Leydig cell tumor were snap-frozen in liquid nitrogen and stored at -70°C until DNA extraction was performed. Fresh frozen samples have been shown to have better DNA integrity than formalin fixed, paraffin embedded samples [31]. All samples were anonymized with an arbitrary number.

DNA extraction

A representative sample of each tissue was obtained by inserting a blunt end 19G 11/2 hypodermic needle (Microlance3, Becton Dickinson, Ireland) through the frozen tissue. DNA was extracted from the core sample by Blood & Cell Culture DNA Mini Kit (Qiagen, Valencia, California USA). DNA was eluted with elution buffer (as recommended by manufacturer) and stored at -20°C . The DNA from these samples was analyzed for somatic mitochondrial mutations.

First round PCR

Segments of mtDNA that have previously shown to contain many somatic mutations [32] were amplified with mitochondrial-specific primers to avoid amplification of homologous regions in the nuclear DNA. Five sets of

mitochondrial specific primer pairs were used, resulting in amplification product between 714 and 928 base pair in length (see Table 1).

The PCR reaction mixture contained 0.1 μ l of extracted DNA (~5ng), 0.8mM dNTPs (0.2mM of each dNTP) (VWR, Oslo, Norway), 1X Thermopol Buffer, 2mM MgSO₄, 0.075unit Taq/ μ l, 0.15 μ M of each forward, reverse and fluorescently labeled primer (Integrated DNA Technologies, Leuven, Belgium) and total reaction volume of 10 μ l. The temperature cycling was performed in an Eppendorf *Mastercycler ep gradient S* (Eppendorf, Hamburg, Germany) with an initial denaturation 94 °C for 240 seconds followed by cycling 38 times under the following conditions, denaturation at 94°C for 15 seconds, annealing for 40 seconds with temperature given in Table 1 and elongation at 72°C for 150 seconds.

Capillary electrophoresis

All first round amplification products were checked and visualized by capillary electrophoresis in MegaBACE 1000 DNA Analysis System (GE Healthcare Life Sciences, Pittsburgh, PA, USA). Samples were loaded into the capillaries from 96-well plates by electrokinetic injection at 161 V/cm for 20 seconds. The temperature of the capillary chamber was set to 27°C and electrophoresis was carried out at a constant field of 145 V/cm.

Second round PCR

Template for second round PCR was 0.8 μ l of a 1:200 dilution (first round PCR in H₂O). The templates were dispensed into 96-wells plates with a syringe dispenser (Hydra 96, Robbins Scientific, USA). To each well 10 μ l reaction mixture was added, consisting of 1xThermopol Reaction Buffer with 2 mM MgS04, 0.3 μ M forward primer, 0.15 μ M 1/2GC-tailed “reverse” primer,

0.15 μ M, 6-Carboxyfluorescein-GC-clamp primer, 500 μ M dNTP, 100 μ g Bovine Serum Albumine (Sigma-Aldrich, Oslo, Norway) and 0.75U Cloned Pfu DNA polymerase. Plates were sealed with two strips of electrical tape (Clas Ohlson, Oslo, Norway). The temperature cycling was repeated 25 times; 94°C for 15 seconds, annealing temperature (given in table 2) held for 30 seconds and extension at 72°C for 60 seconds.

Tumor analysis

Testicular tumor samples were obtained from six patients. Apparent normal tissues adjacent to the tumor were marked with a suture (by the surgeon responsible for removing the tumor) and snap frozen in liquid nitrogen. The suture was used as a reference when the sample was mounted in the cryotome. DNA extracted from the testicular tumor samples were analyzed for somatic mutations in the hotspot regions of the mtDNA. The fragments selected were based on the results of scanning 76% of the mtDNA in 94 tumor samples from 13 different tissues origin [32].

Tissue sectioning

Sample with somatic mitochondrial mutations was mounted to a cryostat sample holder with a water-soluble glycols and resins matrix (Tissue-Tek® O.C.T. Compound, Sakura, Finetek, USA). The sample holder temperature was set to -20°C and the knife temperature to -23°C. The cryostat (Leica CM1950) cut 12 μ m sections. Each cut was gently transferred to either a steel framed polyethylene naphthalate membrane (Leica, Leica Microsystems, Wetzlar, Germany) or glass microscope slide (Thermo Scientific, Gerhard Menzel, Braunschweig, Germany).

Cycling temperature capillary electrophoresis

CTCE analysis was performed for the selected fragments as previously described [28]. In brief; a 96-capillary DNA analyzer (MegaBACE 1000) was used to analyze 6-carboxyfluorescein labeled PCR products. Mutant PCR amplicons were separated from the wild type by cycling the temperature around the capillaries. The cycling temperature was based on the theoretical melting temperature, for each fragment, calculated by Poland's algorithm [29,33] in the implementation described by Steger [34,35]. The separation temperature proposed by the algorithms was adjusted based on the urea concentration in the matrix. The instrument was modified to allow for elevated temperature cycling [27,36]. Temperature cycling was programmed in the macro.ini file used by the Instrument Control Manager (ICM) software package (GE Healthcare Life Sciences, Pittsburgh, PA, USA). The injection and running electric fields were as given for the first round amplicons.

Internal standard

The two heteroplasmic mutations found in the initial sample analysis were re-amplified with a 5' ROX-labeled primer. These were used as internal standards during electrophoresis and were injected into all capillaries in all runs prior to sample injection. The internal standard serves as a control of the separating temperature and as a marker for the DNA mutations.

Tissue sectioning

Fixing and Staining

The membrane and the glass slide with the 12 μm tissue section were dried at room temperature and fixed in pure methanol (Sigma-Aldrich) for 10 minutes, followed by subsequently air drying of the methanol residue. The tissue was

stained with Giemsa azur eosin methylene blue solution (Merck, Damstadt, Germany) (diluted 1/20 with H₂O) for 30 minutes. Slides are then submerged into 1% acetic acid in H₂O solution for 30 seconds (differentiation) and immediately rinsed in water. The stained tissue was air dried prior to imaging by microscope.

Laser Capture Microdissection

A Leica LDM 6000 was used to take images of tissue sections mounted on membrane or glass slides. The software, Leica laser microdissection V6.7.1.3952, was used to control the microscope when taking pictures or selecting areas for laser capture microdissection and cutting. A hardware modification was made to the collection unit allowing for samples to be collected into two strips of 8 PCR caps (VWR, Oslo, Norway).

20 µl of a collection solution (1xThermopol buffer with Proteinase K, 0.27µg/µl) was added to each cap in the inverted strips. After cutting and collecting the selected areas by laser capture microdissection, the strips (with collection liquid and tissue) were mounted onto a 96-well PCR plate (Axygen, VWR, Oslo, Norway). The plate was briefly centrifuged and incubated at 56°C for 30 minutes. Deactivation of proteinase K was achieved by raising the temperature to 95°C for 1 minute. One microliter of incubated solution was used as template for the first round PCR (see above).

Results

Of the six testicular tumors analyzed, the sample from the Leydig cell tumor was found to have two mutations, one in fragment 2 and one in fragment 4 (Table 2). The two mutations were identified as base pair substitutions, C to T at base pair 64 and a T to C at base pair 152 (NC_012920.1). None of the other five tumors had any mutations in the fragments analyzed. Figure 1 shows the electropherograms for the fragments with a mutation. Finding only one sample with a mitochondrial mutation was less than expected, but is not significant when considering the small number of samples.

In the first step the DNA was extracted by needle biopsy taken through the tumor sample. It, therefore, contains a mixture of DNA from cells across the tumor. The presence of a detectable mutation fraction found in this way suggests that sampling by LCM might identify the same mutation. In the other tumors, no mutations were found, which suggests that none is present or the fraction is too low.

The sample with two mutations was taken to the next step in the procedure. Figure 2 shows an image of a section of the tissue with two mutations. The red circle indicates the area determined as normal by the surgeon. Ninety-three circular spots ($\sim 25000\mu\text{m}^2$) were dissected by LCM and dropped into separate caps. For each sample, fragments 2 and 4, in which a mutation was found, were analyzed.

To estimate the number of cells in a LCM sample, Giemsa stained dark blue nuclei were counted in 48 circles, by two researchers. The average number of nuclei observed was 31 with a standard deviation of 13 and a range of 7 to 76 nuclei/circle. The theoretical upper bound can be calculated assuming an average cell diameter of 20 μm . This gives a total of ~ 80 cells.

The mitochondrial DNA copy number in the LCM samples was determined by real-time PCR. LCM samples homoplasmic for both mutations (black circles) and wild type (white circles) were analyzed, and no difference in mtDNA copy number could be detected. The copy number was 300 per cell.

Figure 3 displays the LCM spots for which the mutant fraction could be determined for both mutations (color-coded by white and black circles). The combined LCM and PCR success rate was 86% and 76% for the two markers, respectively. In some samples, other mutants that do not co-elute with the internal standard have been marked with a different color. The inserted electropherograms demonstrate the allele separation and quantification of mutant fraction found in three different spots for each marker. The mutant fraction was calculated based on the peak heights and have previously been shown to have a linear relationship between signal and mutant fraction [29]. From these data, close to homoplasmic mutated regions could be identified for each marker.

Following this first analysis, a further 186 LCM samples were taken from the remaining part of the tissue section, as shown in Figure 4. The mutant fraction could be determined in a large fraction of samples (89% and 91%) as shown in Figure 4. The location of the spots for the second part of the analysis were selected by a subjective assessment of visible structures in Figure 4. Subsequent review of the data in Figures 3 and 4 did, however, not reveal any clear relation between apparent tissue structure and observed mutant fraction.

To validate the procedure, a new section approximately 50 μm below the first one was analyzed. To eliminate possible biases, the samples were collected as two fixed grids (8x6) in Figure 5. The success rate in determining the mutant fraction was 98% and 100%. The positions at which homoplasmic mutations can be observed in Figure 5, correlate with those in Figure 3. Six areas with near homoplasmic mutations in Figure 3, are matched to six nearby areas in Figure 5.

In addition, spots mostly show similar mutant fractions for both markers. The correlation was 0.75. It is reasonable to conclude that these mutations occur together, i.e., an mtDNA either contains both mutations or none.

The method was subsequently used to analyze two independent liver samples (a cholangio- and a hepato- cellular carcinoma). Both tumors had known mitochondrial mutations (study in progress). Each sample was subjected to LCM and mutation analysis in the grid-like sampling scheme as shown for the testis sample. Figure 6 displays the results of the 288 spots cut from each tissue. The upper section (cholangio carcinoma) is dominated by one mtDNA type and has heteroplasmic spots with a low mutant fraction in some areas. Only one spot, indicated by the arrow, is almost homoplasmic for the mutation, but in fact, the electropherogram showed the presence of a third mutation at a low fraction. The lower section of Figure 6 (hepatocellular carcinoma) has multiple spots of near homoplasmic mutations as well as several heteroplasmic spots.

Discussion

This paper describes a method to sample and analyze mutations from spots of a tumor tissue. The novelty lies in the small size of the samples or spots (25000 μm^2 or an average of 31 cell) and the systematic distribution of the samples. Since each cell carries multiple mitochondria and each mitochondrion has multiple copies of mtDNA, dissecting such a small number of cells gives enough templates for PCR amplification. This would not be sufficient for detecting mutations in nuclear DNA. In our studies we observed that it is better to distribute the spots on a regular grid, rather than placing them according to visible structures. The results give a demonstration of the efficacy of tracing the distribution of mutant mitochondrial genomes across a slice from a tissue or a tumor. This method could be used for studies of humans or experimental animals.

Because the LCM instrument depended on gravitation, spots smaller than 25000 μm^2 resulted in a higher laser capture failure rate (data not shown). Increasing the size of the spots risks dilution of the mutant fraction, leading to a potential sampling bias.

Because the probability of independent occurrence of an identical mutation in the mtDNA of a cell is very small, the observed results are consistent with a clonal expansion of cells. To place an upper bound on this probability, a fragment not having any mutation but situated in a hotspot region of the mtDNA (d-loop) was analyzed in 700 LCM samples. No mutations were found. The probability of independently acquiring a given mutation can thus safely be placed below 1/700. Some estimates put it much lower at around 10^{-30} [19]. To further strengthen the argument, macroscopic samples of 94 tumors were analyzed [32]. One hundred and forty-two high fraction mtDNA mutations were sequenced and none of them were identical.

In the specific case of the testicular tumor analyzed here, two homoplasmic mutations 88 base pairs apart are detected and traced. Since the presence of these mutations is highly correlated, it suggests that one cell lineage is being observed, and not a series of independent events.

The method described is sensitive and provides a quantitative way to explore any tissue for heteroplasmy. Under our optimized conditions, the cost of analyzing 300 spots is about \$30 (not including the cost of a trained technician working for three days). Once heteroplasmy with a sufficiently large mutant fraction has been established, it can, for example, be used to explore the resection margins of a tumor. Furthermore, if a continuous stack of sections cut by microtome were available, investigating mitochondrial mutations in three dimensions becomes possible. This would advance the field of three-dimensional lineage tracing in humans.

Conclusion

The method presented in this study can be used for lineage tracing in humans. Its capacity was demonstrated in human tumors and can be extended to the study of metastasis and normal tissue. Using a regular sampling grid of LCM spots is shown to be essential to study tumor heterogeneity.

References

1. Andrews RM, Kubacka I, Chinnery PF, Lightowlers RN, Turnbull DM, Howell N. Reanalysis and revision of the Cambridge reference sequence for human mitochondrial DNA. *Nat Genet.* 1999;23:147.
2. Anderson S, Bankier AT, Barrell BG, de Bruijn MH, Coulson AR, Drouin J, et al. Sequence and organization of the human mitochondrial genome. *Nature.* 1981;290:457–65.
3. Taylor RW, Barron MJ, Borthwick GM, Gospel A, Chinnery PF, Samuels DC, et al. Mitochondrial DNA mutations in human colonic crypt stem cells. *J. Clin. Invest.* 2003;112:1351–60.
4. Fliss MS. Facile Detection of Mitochondrial DNA Mutations in Tumors and Bodily Fluids. *Science.* 2000;287:2017–9.
5. Jones JB, Song JJ, Hempen PM, Parmigiani G, Hruban RH, Kern SE. Detection of mitochondrial DNA mutations in pancreatic cancer offers a “mass-”ive advantage over detection of nuclear DNA mutations. *Cancer Res.* 2001.
6. Liu VW, Shi HH, Cheung AN, Chiu PM, Leung TW, Nagley P, et al. High incidence of somatic mitochondrial DNA mutations in human ovarian carcinomas. *Cancer Res.* 2001;61:5998–6001.
7. Polyak K, Li Y, Zhu H, Lengauer C, Willson JK, Markowitz SD, et al. Somatic mutations of the mitochondrial genome in human colorectal tumours. *Nat Genet.* 1998;20:291–3.
8. Pejovic T, Ladner D, Intengan M, Zheng K, Fairchild T, Dillon D, et al. Somatic D-loop mitochondrial DNA mutations are frequent in uterine serous carcinoma. *Eur. J. Cancer.* 2004;40:2519–24.
9. He Y, Wu J, Dressman DC, Iacobuzio-Donahue C, Markowitz SD, Velculescu VE, et al. Heteroplasmic mitochondrial DNA mutations in normal

and tumour cells. *Nature*. 2010;464:610–4.

10. Salas A, Yao Y-G, Macaulay V, Vega A, Carracedo Á, Bandelt H-J. A Critical Reassessment of the Role of Mitochondria in Tumorigenesis. Turnbull D, editor. *PLoS Med*. 2005;2:e296.

11. Coller HA, Khrapko K, Bodyak ND, Nekhaeva E, Herrero-Jimenez P, Thilly WG. High frequency of homoplasmic mitochondrial DNA mutations in human tumors can be explained without selection. *Nat Genet*. 2001;28:147–50.

12. Sudo H, Li-Sucholeiki X-C, Marcelino LA, Gruhl AN, Herrero-Jimenez P, Zarbl H, et al. Fetal–juvenile origins of point mutations in the adult human tracheal–bronchial epithelium: Absence of detectable effects of age, gender or smoking status. *Mutation Research/Fundamental and Molecular Mechanisms of Mutagenesis*. 2008;646:25–40.

13. Fisher R, Pusztai L, Swanton C. Cancer heterogeneity: implications for targeted therapeutics. *British Journal of Cancer*. 2013;108:479–85.

14. McMahon S, LaFramboise T. Mutational patterns in the breast cancer mitochondrial genome, with clinical correlates. *Carcinogenesis*. 2014;35:1046–54.

15. Ju YS, Alexandrov LB, Gerstung M, Martincorena I. Origins and functional consequences of somatic mitochondrial DNA mutations in human cancer | *eLife*. *eLife Sciences*. 2014.

16. Chen JZ, Gokden N, Greene GF, Mukunyadzi P, Kadlubar FF. Extensive somatic mitochondrial mutations in primary prostate cancer using laser capture microdissection. *Cancer Res*. 2002;62:6470–4.

17. Fialkow PJ, Martin GM, Klein G, Clifford P, Singh S. Evidence for a clonal origin of head and neck tumors. *International Journal of Cancer*. 1972;9:133–42.

18. Smits AJJ, Kummer JA, de Bruin PC, Bol M, van den Tweel JG, Seldenrijk KA, et al. The estimation of tumor cell percentage for molecular testing by pathologists is not accurate. *Mod. Pathol*. 2014;27:168–74.

19. Walther V, Alison MR. Cell lineage tracing in human epithelial tissues using mitochondrial DNA mutations as clonal markers. *WIREs Dev Biol*. 2015;5:103–17.

20. Blackwood JK, Williamson SC, Greaves LC, Wilson L, Rigas AC, Sandher R, et al. In situ lineage tracking of human prostatic epithelial stem cell fate

reveals a common clonal origin for basal and luminal cells. *J. Pathol.* 2011;225:181–8.

21. Cottrell DA, Blakely EL, Johnson MA, Ince PG, Borthwick GM, Turnbull DM. Cytochrome c oxidase deficient cells accumulate in the hippocampus and choroid plexus with age. *Neurobiol. Aging.* 2001;22:265–72.

22. Emmert-Buck MR, Bonner RF, Smith PD, Chuaqui RF, Zhuang Z, Goldstein SR, et al. Laser capture microdissection. *Science.* 1996;274:998–1001.

23. Sluka P, O'Donnell L, McLachlan RI, Stanton PG. Application of laser-capture microdissection to analysis of gene expression in the testis. *Prog Histochem Cytochem.* 2008;42:173–201.

24. Craven RA, Banks RE. Laser capture microdissection and proteomics: possibilities and limitation. *Proteomics.* WILEY-VCH Verlag GmbH; 2001;1:1200–4.

25. Elkahloun AG, Gaudet J, Robinson GS. In Situ Gene Expression Analysis of Cancer Using Laser Capture Microdissection, Microarrays and Real Time Quantitative PCR. *Cancer Biology & Therapy.* Taylor & Francis; 2002;1:353–7.

26. He M, Rosen J, Mangiameli D, Libutti SK. Cancer development and progression. *Adv. Exp. Med. Biol.* New York, NY: Springer New York; 2007;593:117–33.

27. Ekstrom PO, Khrapko K, Li-Sucholeiki X-C, Hunter IW, Thilly WG. Analysis of mutational spectra by denaturing capillary electrophoresis. *Nat Protoc.* 2008;3:1153–66.

28. Ekstrøm PO, Warren DJ, Thilly WG. Separation principles of cycling temperature capillary electrophoresis. *Electrophoresis.* 2012;33:1162–8.

29. Refinetti P, Morgenthaler S, Ekstrom PO. Cycling temperature capillary electrophoresis: A quantitative, fast and inexpensive method to detect mutations in mixed populations of human mitochondrial DNA. *Mitochondrion.* 2016;29:65–74.

30. Ekstrøm PO. Population Screening of Single-Nucleotide Polymorphisms Exemplified by Analysis of 8000 Alleles. *Journal of Biomolecular Screening.* 2002;7:501–6.

31. Srinivasan M, Sedmak D, Jewell S. Effect of Fixatives and Tissue

Processing on the Content and Integrity of Nucleic Acids. The American Journal of Pathology. Elsevier; 2002;161:1961–71.

32. Arstad C, Refinetti P, Warren D, Giercksky K-E, Ekstrøm PO. Scanning the mitochondrial genome for mutations by cycling temperature capillary electrophoresis. Mitochondrial DNA Part A. 2016;:1–12.

33. Poland D. Recursion relation generation of probability profiles for specific-sequence macromolecules with long-range correlations. Biopolymers. 1974;13:1859–71.

34. Steger G. Thermal denaturation of double-stranded nucleic acids: prediction of temperatures critical for gradient gel electrophoresis and polymerase chain reaction. Nucleic acids research. 1994;22:2760–8.

35. Fixman M, Freire JJ. Theory of DNA melting curves. Biopolymers. 1977;16:2693–704.

36. Ekstrom PO, Bjørheim J, Thilly WG. Technology to accelerate pangenomic scanning for unknown point mutations in exonic sequences: cycling temperature capillary electrophoresis (CTCE). BMC Genet. 2007;8:54.

Declarations

Ethics

According to the regional ethics comity (REC), “Technical and methodological development work that uses anonymised biological material” does not require approval from REC

(https://helseforskning.etikkom.no/ikbViewer/page/reglerogrutiner/soknadsplikt/sokerikkerek?p_dim=34999&_ikbLanguageCode=us).

Competing Interests

The authors have declared no conflict of interest.

Author's contribution

PR designed the primer and tested the PCR conditions. PR and POE performed the LCM and the mutation detection. CA collected the samples and scanned the samples. WGT and SM supervised the study. All authors contributed to the writing of the manuscript and have read and approved the final version.

Availability of data

All data in their various phases of processing are available upon request. Are available upon request also all the informatics that have been adapted for the processing.

Acknowledgements

We would like to thank Prof. Eivind Hovig for interesting discussions and for his support. This work received financial support from the Torsteds legacy foundation (Oslo, Norway).

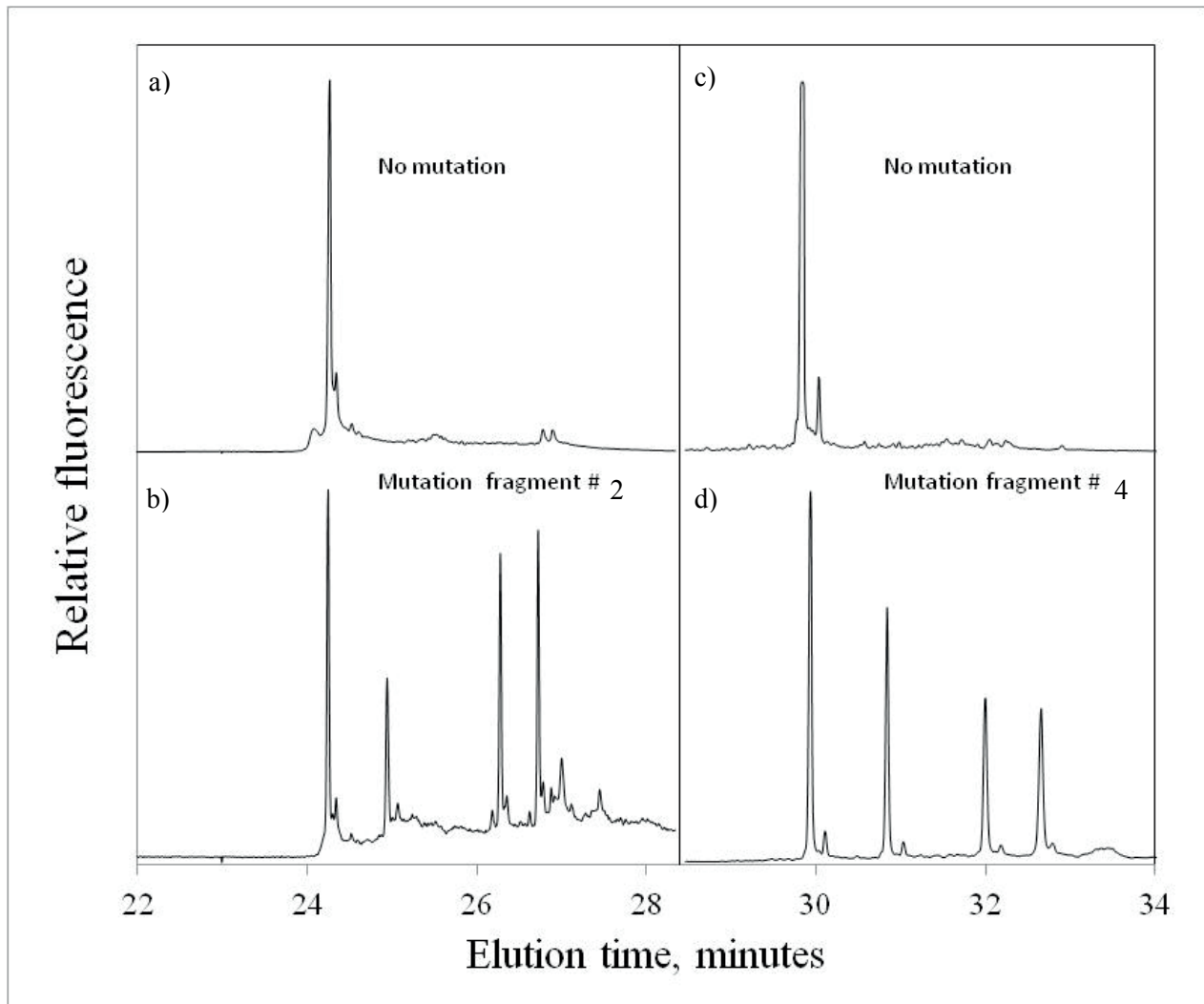
Abbreviations

- CE - Capillary Electrophoresis
- CTCE - Cycling temperature Capillary Electrophoresis
- LCM - Laser capture microdissection.

Figure legends

Figure 1.

Electropherogram of mutated fragments detected by cycling temperature capillary electrophoresis. Electropherograms a) and c) show non-mutated



samples, while electropherograms b) and d) show the two mutations found in one sample.

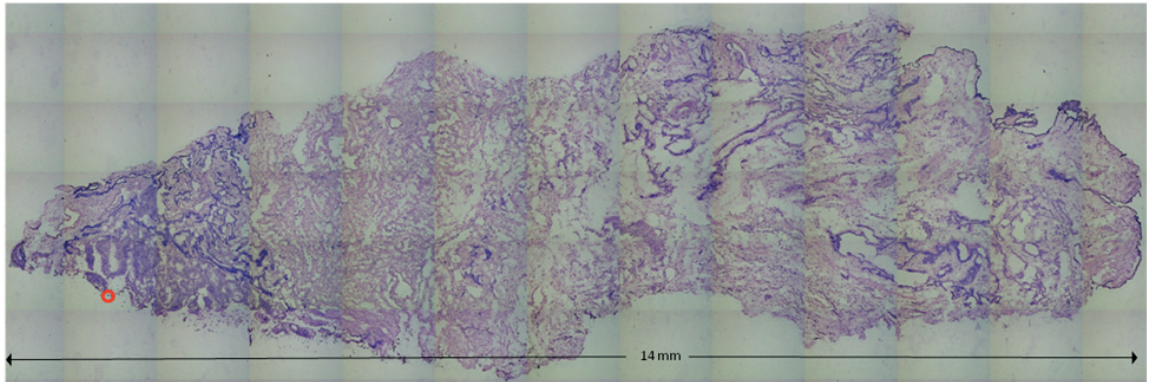


Figure 2.

Overview of sample harboring two mutations. The red circle indicates suture position attached by the surgeon to indicate normal tissue. Image was taken with the LCM microscope by scanning the sample

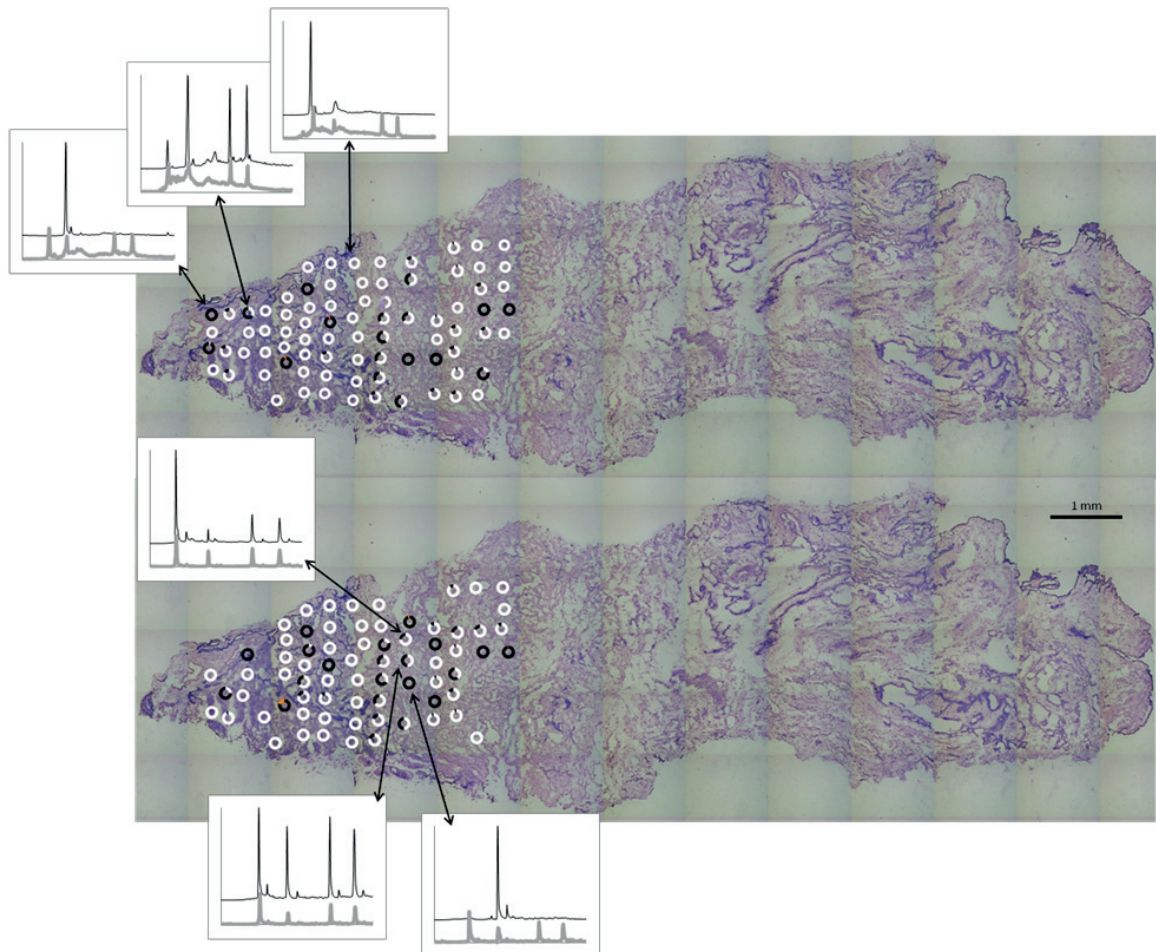


Figure 3.

Laser capture microdissected spots with mutants fractions. Marker detected by primer set # 2 is on the top and marker detected by primer set #4 is on the bottom. White is associated with the first peak (homoduplex) in electropherogram, and black is associated with the second peak (homoduplex). The fraction of white and black in each circle, represent the fraction of the first and second variant respectively in each sample. Other color presents mutants not aligning with the internal standard in the electropherogram. Inserted electropherograms represent homoplasmic regions and heteroplasmic areas (e.g mix of two cell population).

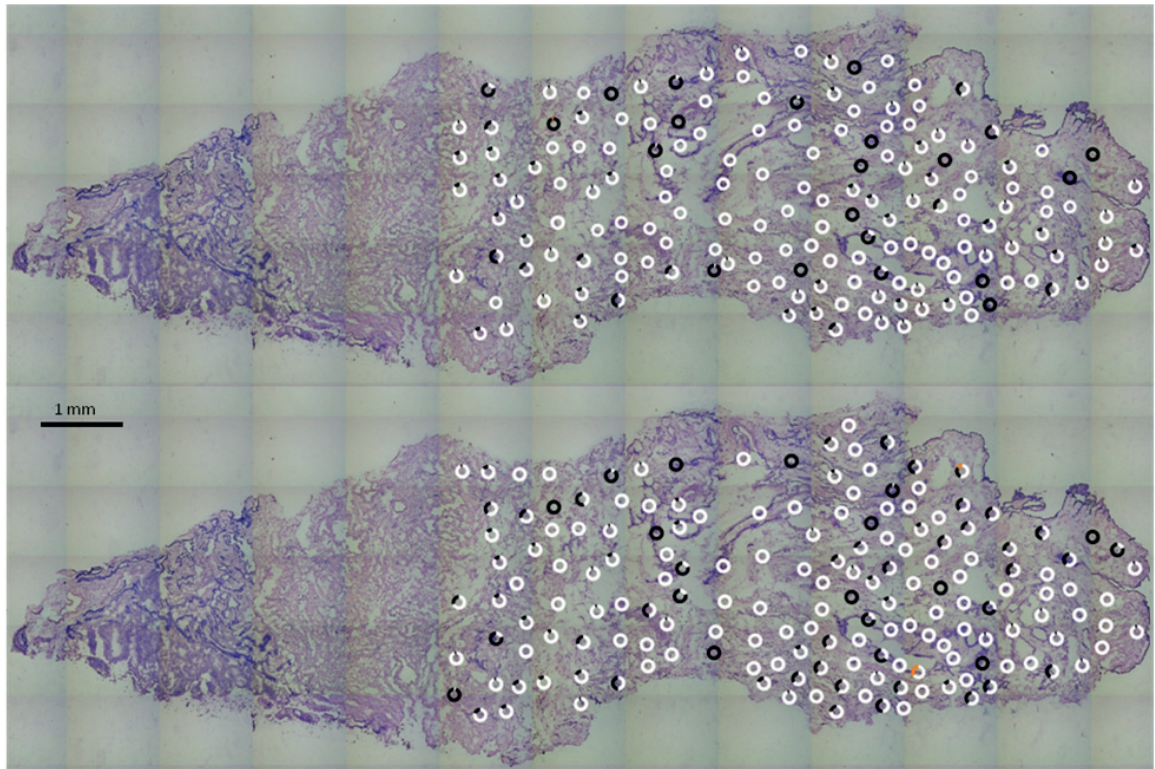


Figure 4.

Laser capture microdissected spots with mutants fractions. Marker detected by primer set # 2 is on the top and marker detected by primer set #4 is on the bottom. White is associated with the first peak (homoduplex) in electropherogram, and black is associated with the second peak (homoduplex). The fraction of white and black in each circle, represent the fraction of the first and second variant respectively in each sample. Other color presents mutants not aligning with the internal standard in the electropherogram. Inserted electropherograms represent homoplasmic regions and heteroplasmic areas (e.g mix of two cell population).

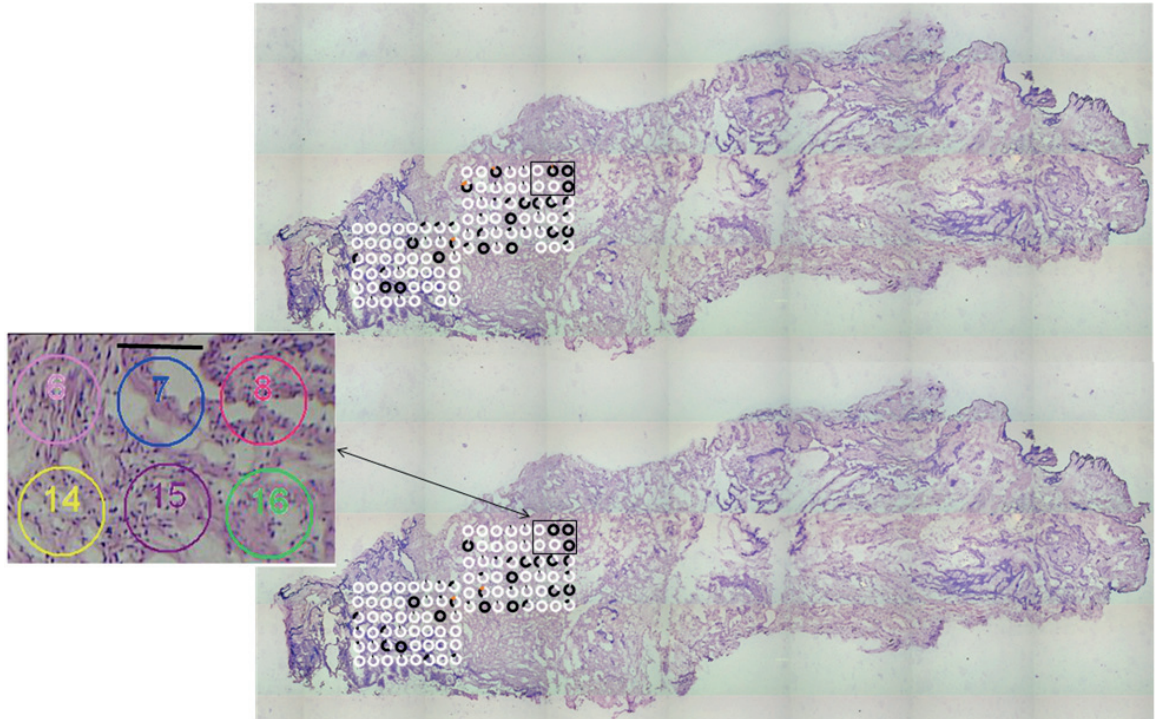


Figure 5.

Focused laser capture microdissection in a grid of 8x6 spots. Marker detected by primer set # 2 is on the top and marker detected by primer set #4 is on the bottom. Enlargement show adjacent different homoplasmic areas. Scale bar is 178 μ m long.

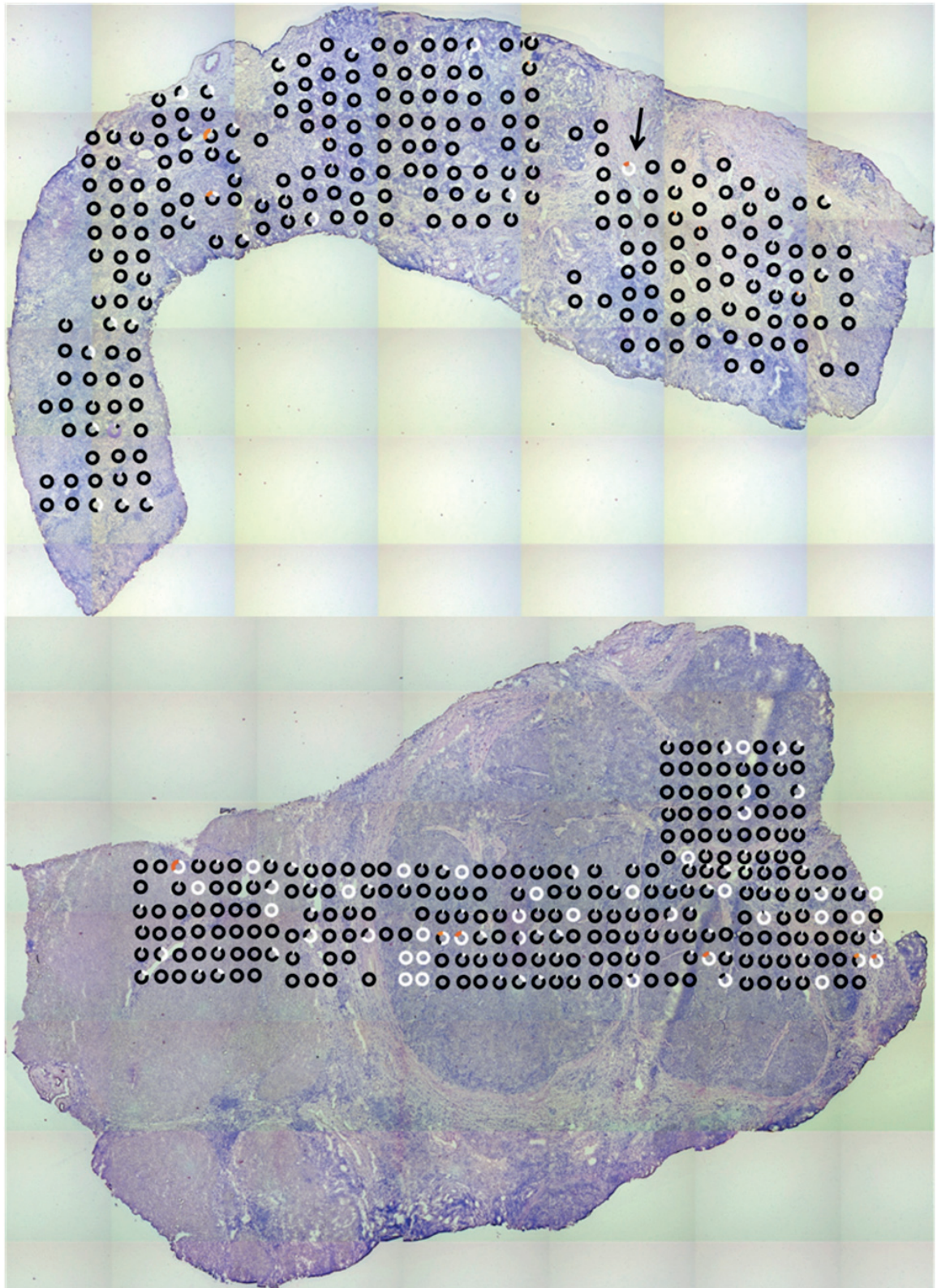


Figure 6.

Laser capture microdissection of two liver samples. Top image is a cholangiocellular carcinoma, the lower is a hepatocellular carcinoma. The first peak in the electropherogram (homoduplex) is presented as white while second peak (homoduplex) is black. Other colors present represent mutants not aligning with the internal standard in the electropherogram (i.e. other variants).

Table 1**Primers used to selectively amplify mtDNA.**

#	Start (bp)	End (bp)	Length (bp)	"Forward" primer (5' - 3')	"Reverse" primer (5' - 3')	Annealing temperature (°C)
23	15924	201	846	*AACCGGAGACGAAAACCTTTTC	*CTTTAGTAGGTATGTTGCGCTGT	51
1	16521	880	928	*CCATAAAGCCTAAATAGCCACA	*CCAACCCTGGGGTTAGTATAGCT	54
10	6917	7671	754	*TGCTCTGAGCCCTAGGATTCATC	*TGAGGGCGTGATCATGAAAGGTG	55.5
16	10852	11566	714	*GCCTAATTATTAGCATCATCCCC	*ATGCCTCATAGGGATAGTACAAG	51
22	15169	15993	824	*GAGGGGCCACAGTAATTACAAAC	*TGGGTGCTAATGGTGGAGTTAAA	51
				*=tail sequence (CGCCCGCCGCCCCGCG)		

Table 2

Primers used to amplify fragments suitable for detecting mutation by CTCE.

#	Start(bp)	End(bp)	Template,fragment# fromfirstPCR	"Forward"primer(5'-3')	"Reverse"primer(5'-3')	Annealing temperature(°C)
1	16569	25	1	*TGCATGGAGAGCTCCCGTGAGTGG	CCCCTTAAATAAGACATCACGAT	52
2	42	126	1	*ATTAACCACTCACGGGAGCTCTC	AGGATGAGGCAGGAATCAAAGAC	55
4	131	181	1	*CACCTATGTCGCAGTATCTGTC	CACACTTAGTAAGTATGTTCCG	55
6	483	513	1	*GGGGTTAGCAGCGGTGTGTGTGTG	TCCCCTCCATACTACTAATCT	55
7	530	633	1	*TACCAGCACACACACCCGCTG	CAAACCTATTTGTTTATGGGGTGA	55
8	673	705	1	*TTAGAGGGTGAACCTACTGGAAC	GGTTTGGTCTAGCCTTCTATT	58
82	7031	7134	10	*ACGACACGTAACGTTGTAGCC	AATATGATAGTAAAATGGATTT	52
84	7340	7416	10	*CTTCTTCCCACAACACTTTCTC	TCTCAAATCATGAAAATTATTAAT	55
125	11029	11086	16	*TTAGAGGGGGTGTAGGGGGT	CATCCCTCTACTATTTTAAACC	58
127	11193	11243	16	*ACCAGCCAGAACGCCTGAACGCA	GGTGTGTGAGTGTAAATTAGTG	55
128	11283	11311	16	*TGTGCCTGCGTTCAGGCGTCTGG	TAATCATATTTATATCTTCTTC	60
130	11437	11492	16	*TTGACCCAGCGATGGGGGCTTCGA	GAGCCAACAACCTAATATGACTA	55
176	15201	15257	22	*AGAATCGTGTGAGGGTGGGACTGT	AGTAATTACAACTTACTATCCG	60
177	15274	15377	22	*AGTAGACAGTCCCACCCTCACAC	GGTGATTTATCGGAATGGGAGG	60
178	15394	15448	22	*CTAGGAATCACCTCCATTCGGA	TAAATGTCATTAAAGGAGAGAAGGAA	55
181	15761	15864	22	*ACCTCCTATTCTAACCTGAATC	CAGGCCATTTGAGTATTTTGT	55
184	16080	16130	23	*CAAGTATTGACTCACCCATCAAC	ACAGGTGGTCAAGTATTTATGGTA	57
185	16170	16272	23	*GTGGGTGAGGGGTGGCTTTGGAGT	CCAATCCACATCAAACCCCTC	56
187	16263	16366	23	*AACTGCAACTCAAAGCCACCCC	CCCTATCTGAGGGGGTTCATCCAT	58
				*=tailsequence(CCCGCCGCCCGCCCGGG)		
				GC-Clamp=(6FAM-CGCCCGCCGCGCCCGCGCCCGTCCCGCCGCCCGCCCGGG)		

10 Three-dimensional Lineage Tracing

From the results described in chapter 5, the optimal thickness of a slice was established at $12\mu\text{m}$. So the question was, how many consecutive samples of $12\mu\text{m}$ could be taken using a microtome. The first attempt, performed on the cell tumor used for the paper 9.2 resulted in 24 cuts, the first 12 are successive, and the last 12 were taken with $2 \times 12\mu\text{m}$ interval between them. The same 96 sample grid was taken on 18 of these slices. The results can be seen in the paper described in 10.1. It confirms a certain degree of conservation of the mutant fractions across the z axis. Furthermore, it also established that sampling every slice was oversampling, and depth was being sacrificed for too much resolution.

Viewing the three dimensional plot of two independent markers in the cell tumor, we became convinced that we could identify cell lineages in tumors, and trace them in 3D. The question remained: "What cell lineage is being tracked?" In fact, no guarantee exists that the cell lineage that carries the mtDNA mutation was the only tumor lineage. A tumor lineage is being tracked, so the presence of the mutation marks the presence of tumor cells, but the inverse needs not be the case. Furthermore, the majority of the scanned volume does not contain the mutation. Thus, if the cell line followed was indeed the major tumor lineage, then it would mean that the majority of a tumor volume would not be composed of tumor-derived cells. A condition was needed, where the presence of the marker, would guarantee tumor origin, and its absence would guarantee it is not tumor derived. Figure 10.1 shows a textbook illustration of how metastasis are thought to form. If the model is true, then the metastasis has at its origin one, or very few tumor cells from the primary tumor. We hypothesised that, a metastasis containing the same mtDNA mutation as found in the primary tumor, would be founded by a tumor cell marked with that mutation. Consequently, all the tumor derived cells in the metastasis would contain the mutation. The key is to be able to identify the same mutation in both the primary tumor and the metastasis.

The sampling protocol chosen was to sample one out of each five. Ideally, 70 successive $12\mu\text{m}$ slices of tissue would be taken. Every 5th slice is placed on membrane for LCM sampling, and the other four are placed on glass, stained and imaged. This results in deeper the analysis of a larger volume by increasing the depth of tissue sampled, without sacrificing the ability to

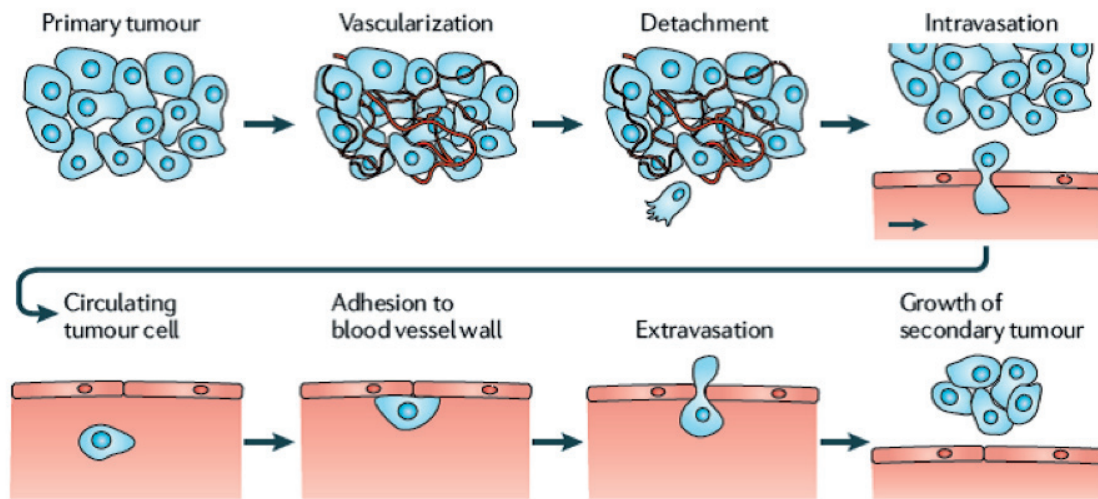


Figure 10.1: Current theory of metastasis formation. A single cell leaves a grown tumor and "seeds" a metastasis. If true, this hypothesis means that all tumor cells in a metastasis are descendent of a single tumor cells. Picture taken from the site johnhopkins.edu.

image the tissue continuously. Protocol and results can be seen in the paper described in 10.1.

10.1 Resulting Publication

10.1.1 Short description

The paper describes the result of our three dimensional lineage tracing. Using the method described in the chapter 9, mtDNA mutations that can serve as marker for lineage tracing are identified in five primary tumors, breast, bladder, colon, liver and cell (the same tissue described in chapter 9). In the bladder tumor case, the two markers found in the primary tumor are then also found in the lymph node metastasis. The primary breast tumor has only one marker, and it was identified in both a lymph node metastasis and in a liver metastasis. Three-dimensional reconstructions were not possible for the primary bladder tumor and breast tumor, due to poor sample conditions, but were possible for their metastasis. In the end, a total of 8 three dimensional reconstruction, for 6 tissues (2 tissues had two markers, the bladder tumor metastasis and the cell tumor) were obtained. A strong correlation was observed between both markers followed in the cell tumor and in the bladder tumor metastasis. Also, the majority of the volume analysed is not made of cells that contain the mutation. An observation that is true in primary tumors and in the metastasis and is difficult to explain using standard oncogenesis theory. A possible interpretation is that tumor lineages grow by "percolating" through the surrounding tissue and conditioning it to assume the abnormal morphology observed.

10.1.2 Significance

It is probably the most significant paper of the thesis. In it is the definitive proof that it is possible to perform lineage tracing in three dimensions. A tissue is a volume, thus to properly follow a cell lineage it must be followed in three dimensions. Lineage tracing in two dimensions only gives a cut across the structures formed by lineages as they grow. The result resembles a random speckling of mutations. However, three dimensionally following mtDNA mutations revealed some form of connection between the various areas carrying the mutation.

10.1.3 Personal contribution

Samples came from different sources. I was responsible for making the slices and preparing them, as well as taking the LCM samples and processing them, with the almost constant supervision and advice of Per Ekstrom. I also designed and implemented the software that was used to analyse the electropherograms and performed the analysis. The plotting of the results was a joint effort of Per Ekstrom and Myself, as was writing the paper.

10.1.4 Status

In preparation

3D tracing of human tumor cell lineages in primary tumor and metastasis

Paulo Refinetti^a, Stephan Morgenthaler^a, William G. Thilly^c, Per O. Ekstrøm^b

^a*École Polytechnique Fédérale de Lausanne, Chaire de Statistiques Appliquées, Station 8, Lausanne, Switzerland*

^b*Department of Tumor Biology, Norwegian Radiumhospital, Oslo, Norway*

^c*Laboratory of Epidemiology, Massachusetts Institute of Technology, Cambridge, MA*

Abstract

Background

Results

Conclusion

Keywords:

1. Background

Tumors are thought to be clonal lesions, originating from a single cell, that grows and goes on to create a tumor mass [1, 2, 3]. As with developmental biology, the question arises as to how a single cell gives rise to the observed result. In other words, what are the patterns of growth of cell lineage. Since the question is similar in developmental biology and tumors, it follows that similar techniques have been employed to study the patterns of cell lineages. Tracing cell lineages in human tissue, both normal and tumor requires the identification of markers, that identify individual lineages and allow their patterns of growth to be mapped. Somatic mitochondrial DNA (mtDNA) mutations are a strong candidate to serve as markers of clonality to study lineage tracing, and have been successfully used in normal human tissues [4, 5, 6]. Laser capture micro dissection (LCM), is a technique that allows the selection of small groups of cells from a histologic preparation [7]. In normal tissue, it is therefore possible to select micro-anatomical samples that contain cells from one tissue type from an organ. The reason being that the composition of an organ can be readily identified by histology and staining

procedures. Application of lineage tracing methods from normal tissue to tumor tissue has been less conclusive. The first reason is the difficulty in identifying different tissue origins within a tumor tissue. While it is possible to recognise a tissue section as "not normal", it is often impossible to determine which portions are tumor [8]. Attempts to selectively sample tumor tissue can therefore only lead to a biased sampling of the selected tissue [1]. Another often encountered issue is the presence of a signal. In normal tissue, clones tend to be reasonably small, for example one crypt in the human colon [6]. Consequently, even rare mutational events can be observed by sampling a reasonably small amount of tissue, due to the number of clonal units (crypts for the colon) that can be studied. In tumors, due to their origin as one cell that grows, lineages tend to be fewer and larger. Attempts to detect gene inactivating mutations in cytochrome C oxydase, in human adenomas have been hampered by the small number of adenomas that actually develop such mutations [9]. When a mutation was detected, it usually covered a large portion of the total adenoma. Another difficulty when performing lineage tracing is that tissue are a three dimensional. Studying lineage tracing on a slice of tissues only gives a cross section of the real pattern of cell growth. To understand the actual patters, a three dimensional analysis is required. Sampling a volume requires a dramatically larger number of sample than sampling a surface. Most methods that are capable of detecting and quantifying mtDNA mutations do not have the throughput and precision to satisfy this constrain [10].

In this paper, the authors propose to use a method that has been shown to effectively detect and quantify mtDNA mutations in two dimensions, and extend it to three dimensions. In a previous work cycling temperature capillary electrophoresis (CTCE) is used to detect and quantify mtDNA mutations in lied cell tumor. The mutations found are then analysed over a large number of micro-anatomical samples taken across a slice of tumor tissue (Refinetti et al, under review). The results confirmed the initial hypothesis that it is almost impossible to select for a cell lineage by histologic analysis. Using CTCE as a method to detect mtDNA mutations as markers of clonality will give the precision and throughput for a three dimensional analysis. Also, sample will be taken on a standard grid formation at a similar position on each successive slice of tissue. In doing so, no bias will be introduced in the sample.

2. Materials and Methods

2.1. Sample description and origin

Samples were collected from 5 different patients, and are as follow:

- Patient 1: Primary breast tumor, Liver metastasis, lymph node metastasis
- Patient 2: Primary bladder tumor, Lymph node metastasis, primary tumor.
- Patient 3: Primary hepatocellular carcinoma
- Patient 4: Primary colon cancer
- Patient 5: Primary Leighdig cell tumor (also described in Refinetti et al. under review)

Samples from Patients 1, 2 and 4 were collected at the Norwegian Radiumhospital with informed consent. Sample from patient number 5 was collected at Bærum sykehus (Vester Viken, Helse sør-øst, Norway). Samples from patient 3 was collected at the surgical department of the university hospital of Bologna (Italy). All samples were collected with informed consent according to local law and were anonymised prior to us using. According to the Norwegian law, "Technical and methodological development work that uses anonymised biological material" does not require approval from research ethics committee (REC) (Web page, last access November 2016). Following surgical resections, all samples were snap frozen by immersion in liquid nitrogen within 10 min. They were then kept below -30°C until sectioning. No fixing or embedding was done before sectioning.

2.2. DNA extraction

Two consecutive $50\mu\text{m}$ slices were placed in an eppendorf tube an DNA was extracted using the Mag-attract system by Macherey-Nagel. In brief, samples were incubated with Proteinase K in lysis buffer for 3h at 50°C . Samples were then placed in a dedicated DNA extraction robot by Macherey-Nagel, specifically for the use of the Mag-attract system. Result is DNA at a rough concentration go $50\text{ng}/\mu\text{l}$.

2.3. Tissue sectioning

Sampling of tumor tissue, is obtained by taking consecutive slices with the microtome and using LCM to sample an 8×12 grid on the slices. Samples were mounted to a cryostat sample holder with a water-soluble glycols and resins matrix (Tissue-Tek O.C.T. Compound, Sakura, Finetek, USA) and placed in a cryotome (Leica CM1950) with room temperature of -20°C and knife temperature of -23°C . Small adjustments were made to those temperatures if it was necessary to improve cutting performance. 2 consecutive $50\mu\text{m}$ slices were taken and placed in a Eppendorf (Eppendorf, Germany) tube. Next, 70 consecutive $12\mu\text{m}$ slices were then taken. Every 5th slice (i.e. numbers 1, 6 ... 66) was placed on PET (Leica, Leica Microsystems, Wetzlar, Germany) membrane for LCM. All other cuts were placed on glass slide (SuperFrost by Thermo Scientific, Gerhard Menzel, Braunschweig, Germany) for imaging.

2.4. Fixing and Staining

The membranes and glass slides with the $12\mu\text{m}$ tissue section were dried at room temperature and fixed in pure methanol (Sigma-Aldrich) for 10 minutes, followed by subsequently air drying of the methanol residue. The tissue was stained with Giemsa azur eosin methylene blue solution (Merck, Darmstadt, Germany) (diluted 1/20 with H_2O) for 30 minutes. Slides are then submerged into 1% acetic acid in H_2O solution for 30 seconds (differentiation) and immediately rinsed in water. The stained tissue was air dried prior to imaging by microscope.

2.5. Laser Capture Microdissection

A Leica LDM 6000 was used to take images of tissue sections mounted on membrane or glass slides. The software, Leica laser microdissection V6.7.1.3952, was used to control the microscope when taking pictures or selecting areas for laser capture microdissection and cutting. A hardware modification was made to the collection unit allowing for samples to be collected into two strips of 8 PCR caps (VWR, Oslo, Norway). The Leica software was used to design a standard sampling pattern of 96 circles with an area of about $25000\mu\text{m}^2$, placed regularly on an 8×12 grid. The same grid was sampled on each membrane section. Observation of the in-between samples placed on glass slides, allowed the grids to be placed on the position corresponding to the previous sampling. $20\mu\text{l}$ of a collection solution (1xThermopol buffer with Proteinase K ($0.27\mu\text{g}/\mu\text{l}$)) was added to each

cap in the inverted strips. After cutting and collecting the selected areas by laser capture microdissection, the strips (with collection liquid and tissue) were mounted onto a 96-well PCR plate (Axygen, VWR, Oslo, Norway). The plate was briefly centrifuged and incubated at 56°C for 30 minutes. Deactivation of proteinase K was achieved by raising the temperature to 95°C for 1 minute. One microliter of incubated solution was used as template for the first round PCR.

2.6. PCR conditions

2.6.1. First Round

Areas of mitochondrial DNA (mtDNA) that previously has shown to contain many somatic mutations [11] were amplified with mitochondrial specific primers to avoid amplification of homology regions in the nuclear DNA (NUMTs). Five sets of mitochondrial specific primer pairs were used, resulting in amplification product between 714 and 928 base pair in length (see table 1). The PCR reaction mixture contained 0.1 μ l of extracted DNA, 0.8mM dNTPs (0.2mM of each dNTP) (VWR, Oslo, Norway), 1X Thermopol Buffer, 2mM MgSO₄, 0.075unit Taq/ μ l, 0.15 μ M of each forward, reverse and fluorescently labeled primer (Integrated DNA Technologies, Leuven, Belgium) in a total reaction volume of 10 μ l. The temperature cycling was performed in an Eppendorf Mastercycler ep gradient S (Eppendorf, Hamburg, Germany) with an initial denaturation 94 °C for 240 seconds followed by cycling 38 time under the following conditions, denaturation at 94 °C for 15 seconds, annealing for 40 seconds with temperature given in Table 1 and elongation at 72°C for 150 seconds.

2.6.2. Second Round

Template for second round PCR was 0.8 μ l of a 1:200 dilution (first round PCR in H₂O). The templates were dispensed into 96-wells plates with a syringe dispenser (Hydra 96, Robbins Scientific, USA). To each well 10 μ l reaction mixture was added, consisting of 1xThermopol Reaction Buffer with 2 mM MgSO₄, 0.3 μ M forward primer, 0.15 μ M 1/2GC-tailed "reverse" primer, 0.15 μ M, 6-Carboxyfluorescein-GC-clamp primer, 500 μ M dNTP, 100 μ g Bovine Serum Albumine (Sigma-Aldrich, Oslo, Norway) and 0.75U Cloned Pfu DNA polymerase. Plates were sealed with two strips of electrical tape (Clas Ohlson, Oslo, Norway). The temperature cycling was repeated 25 times; 94°C for 15 seconds, annealing temperature (given in table 2) held for 30 seconds and extension at 72°C for 60 seconds.

#	Start (bp)	End (bp)	Length (bp)	"Forward" primer (5' - 3')	"Reverse" primer (5' - 3')	Annealing temperature (°C)
23	15924	201	846	*AACCGGAGACGAAAACCTTTTC	*CTTTAGTAGGTATGTTGCCTGT	51
1	16521	880	928	*CCATAAAGCCTAAATAGCCCACA	*CCAACCCTGGGGTTAGTATAGCT	54
10	6917	7671	754	*TGCTCTGAGCCCTAGGATTCATC	*TGAGGGCGTGATCATGAAAGGTG	55.5
16	10852	11566	714	*GCCTAATTATTAGCATCATCCCC	*ATGCCTCATAGGGATAGTACAAG	51
22	15169	15993	824	*GAGGGGCCACAGTAATTACAAAC	*TGGGTGCTAATGGTGGAGTTAAA	51
				*tail sequence (CGCCCGCCGCCCGCG)		

Table 1: First round primers used.

#	Start(bp)	End(bp)	Template,fragment# fromfirstPCR	"Forward" primer(5'-3')	"Reverse" primer(5'-3')	Annealing temperature(°C)
1	16569	25	1	*TGCATGGAGAGTCCCCTGAGTGG	CCCCTTAAATAAGACATCACGAT	52
2	42	126	1	*ATTAACCCTACACGGAGCTCTC	AGGATGAGGCAGGAATCAAAGAC	55
4	131	181	1	*CACCCATGTCGCAGTATCTGTC	CACACTTTAGTAAGTATGTTCCG	55
6	483	513	1	*GGGGTTAGCAGCGGTGTGTGTG	TCCCACTCCCATACTACTAATCT	55
7	530	633	1	*TACCAGCACACACACCCGCTG	CAAACCTATTTGTTATGGGGTGA	55
8	673	705	1	*TTAGAGGGTGAACCTACTGGAAC	GGTTTGGTCTAGCCTTTCTATT	58
82	7031	7134	10	*ACGACACGTAACGTTGTAGCC	AATATGATAGTAAAATGGATTTT	52
84	7340	7416	10	*CTTCTTCCCACACACTTCTC	TCTCAAATCATGAAAATTAATAAT	55
125	11029	11086	16	*TTAGGAGGGGGTGTAGGGGGT	CATCCCTCTACTATTTTTAACC	58
127	11193	11243	16	*ACCAGCCAGAACCCTGAACGCA	GGTGTGTGAGTGTAAATTAGTG	55
128	11283	11311	16	*TGTGCTGCGTTCAGGCGTTCTGG	TAATCATATTTATATCTTCTC	60
130	11437	11492	16	*TTGACCCAGCGATGGGGCTTCGA	GAGCCAACAACCTAATATGACTA	55
176	15201	15257	22	*AGAATCGTGTGAGGGTGGGACTGT	AGTAATTACAACTTACTATCCG	60
177	15274	15377	22	*AGTAGACAGTCCCACCCTCACAC	GGTGATTTTATCGGAATGGGAGG	60
178	15394	15448	22	*CTAGGAATCACCTCCCATTCGGA	TAATGTCATTAAAGGAGAGAAGGAA	55
181	15761	15864	22	*ACCTCCTCATCTAACCTGAATC	CAGGCCATTGAGTATTTGTTT	55
184	16080	16130	23	*CAAGTATTGACTCACCATCAAC	ACAGGTGGTCAAGTATTTATGGTA	57
185	16170	16272	23	*GTGGGTGAGGGGTGGCTTTGGAGT	CCAAATCCACATCAAACCCCCTC	56
187	16263	16366	23	*AACTGCAACTCAAAGCCACCC	CCCTATCTGAGGGGGTCAATCCAT	58
				*tail sequence(CCCGCCGCCCGCCCGG)		
				GC-Clamp=(6FAM-CGCCCGCCCGCCCGCCCGTCCCGCCGCCCGCCCGCCCGG)		

Table 2: Second round primer pairs used

2.7. Cycling Temperature Capillary Electrophoresis

CTCE analysis was performed for the selected fragments as previously described. In brief; a 96-capillary DNA analyzer (MegaBACE 1000) was used to analyze 6-carboxyfluorescein labeled PCR products. Mutant PCR amplicons were separated from the wild type by cycling the temperature around the capillaries. The cycling temperature was based on the theoretical melting temperature, for each fragment, calculated by Poland's algorithm [12] in the implementation described by Steger [12, 13]. The separation temperature proposed by the algorithms was adjusted based on the urea concentration in the matrix. The instrument was modified to allow for elevated temperature cycling [14, 15]. Temperature cycling was programmed in the macro.ini

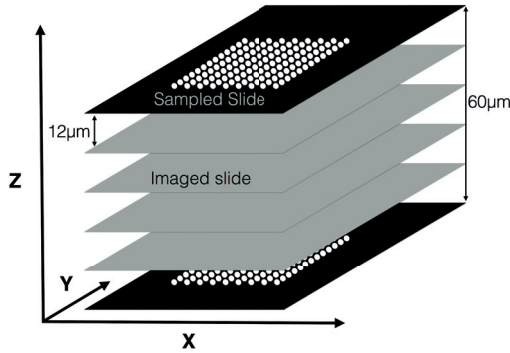


Figure 1: Schematic representation of sampling procedure. The sampled slides are placed on LCM membranes, and 96 samples are taken in the form of an 8×12 grid. One slide every 5 is sampled, and the ones in-between are placed on glass slides for imaging. The above schematic has been repeated for a total of 70 slides of which 14 are sample and 56 are only imaged.

file used by the Instrument Control Manager (ICM) software package (GE Healthcare Life Sciences, Pittsburgh, PA, USA). The injection and running electric fields were as given for the first round amplicons.

2.8. Data visualisation

The output of the analysis is the mutant fraction in the samples taken. Understanding these results requires the information to be plotted in three dimensions. The hypothesis is made that the mutation analysed, marks a cell lineage within the tissue. For each sample, the depth, within the tissue, is known by the slice number (assuming the microtome has infinite precision). Assuming also that the grid is positioned always in the same way on each sampled slide, it is easy to assign a cartesian coordinate to each sample taken. Also, the shape of the sample is constant, a disc of $25000\mu m^2$ with a height of $12\mu m$. A total of 21 points are then randomly distributed within each disk of a sample, and assigned a value 0 or 1. The number of 1 assigned is proportional to the mutant fraction measured in the sample. A mutant fraction of 100% will be represented by 21 points with value 1, randomly distributed in space within the volume occupied by the sample in space. Linear interpolation is then used to obtain a surface plot of the resulting "cloud" of points in the volume.

Patient	Tissue	3D	Markers	Sample Taken	Measured Samples
1	Primary Breast tumor	no	1	672	524
	Lymph node metastasis	yes	1	1344	571
	Liver Metastasis	yes	1	1344	708
2	Primary Bladder cancer	no	2	288	287, 229
	Lymph node metastasis	yes	2	1344	986, 611
3	hepatocellular carcinoma	yes	1	1344	956
4	Colon Cancer	yes	1	1344	598
5	leydig cell tumor	yes	2	1824	1728,1824

Table 3: Summary of samples analysed. For each patient, the markers analysed are the same across tissues. In some tissues, due to their states or size, it was impossible to establish a 3D reconstruction. 3D reconstruction was only established when multiple samples could be taken consecutively. The number of samples taken is the number of micro anatomical pieces taken using LCM. The measured samples is the number of samples for which a mutant fraction could be successfully calculated. In the cases with two marker, the number of samples for each marker are separated by a coma.

3. Results

As can be seen in table 3, a three dimensional reconstruction of mutant fraction has been possible for six tissues. Of these six tissues, two have two markers, for a total of 8 three dimensional reconstructions. Not every sample successfully gives information. Lack of cells in the sample (due to the morphology of the tissue) or PCR failure, or difficult to interpret signal can contribute to the absence of information from a sample. The total number of samples taken, as well as the number of successful ones are also shown in table 3. The visualisation of the mitochondrial markers in three dimensions, gives a representation of the patters of cell lineage.

Figure 2 shows the histogram of mutant fractions for all analysed markers and tissues. It can be observed that all distributions are heavily skewed towards the wild type. Thus, a majority of samples taken are are mostly wild-type. In the tissues for which 2 markers are present, both markers seem to have a very similar distribution. Suggesting that the mutation are found together in the same samples, thus independently marking the same lineage.

The three dimensional reconstruction of the cell lineages traced in each tissue are online and publicly accessed in a form that allows for exploration. A snap picture of each can be seen in figures 4 to 6. An interac-

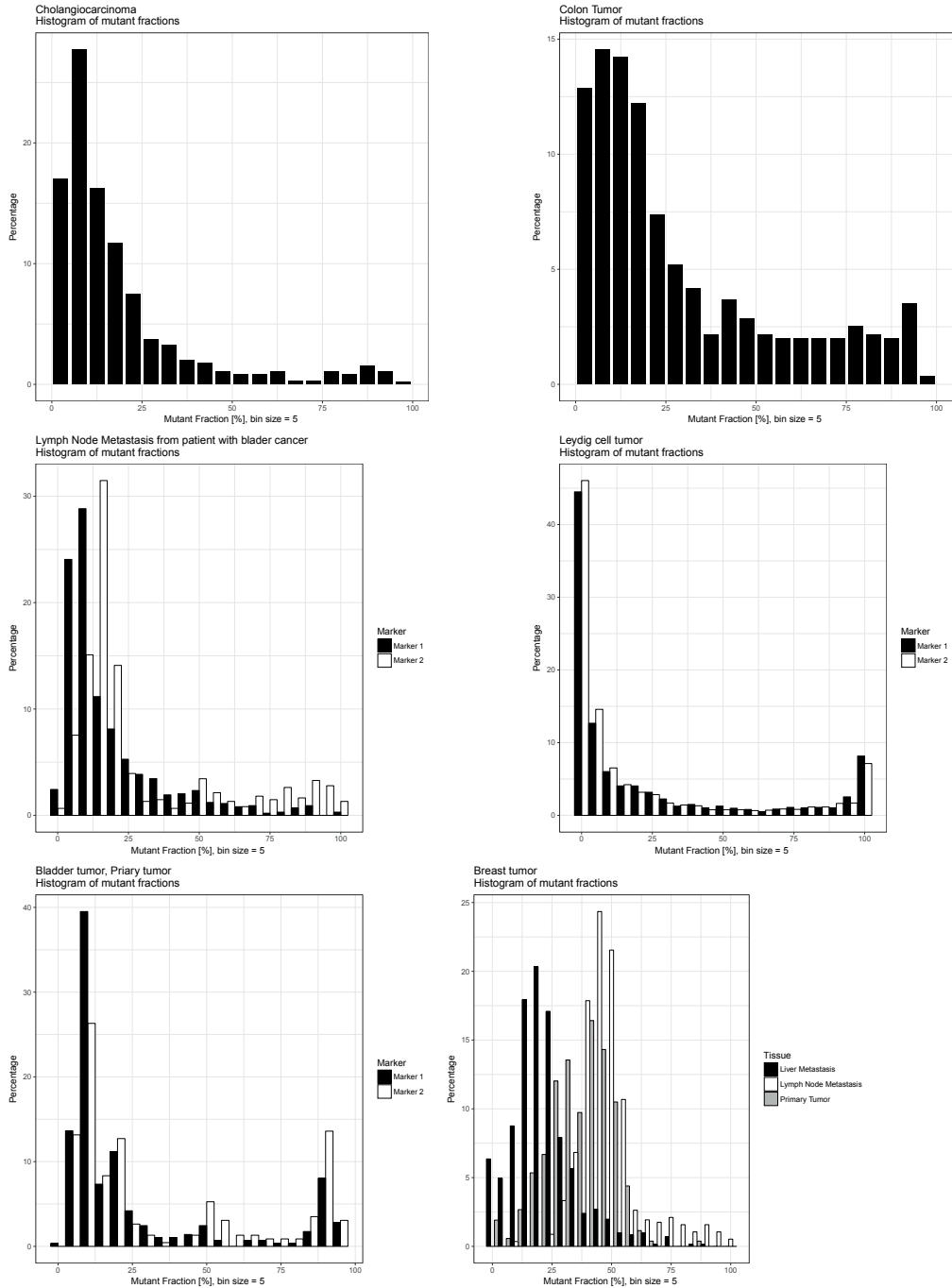


Figure 2: Histogram of mutant fraction. The mutant fractions for each tissue are summarised as histograms. The colon cancer and the hepatocellular carcinoma, have only one tissue and one marker, for which the distribution of the mutant fraction i is shown. The breast tumor, has only one marker followed, and three tissues, for an easier comparison, the mutant fraction of the three tissues is shown in the same histogram. The tissues with more than one marker are plotted individual, with both marker shown, this is the case for the leydig cell tumor, as well as both tissues of the bladder tumor.

tive plot of all the samples and all markers available can be seen online. A page has been dedicated in the Radiumhospital website for this purpose: <http://www.ous-research.no/hovig/?k=hovig%2FEkstrom&aid=16790>. The graphs show both the reconstruction of the lineage carrying the mutation, and of those not carrying it. By observing the reconstruction of the lineages carrying the mutations, the first thing that can be seen is the sparsity of the mutations. Taking the example of the leydig cell tumor, the majority of the volume does not carry the mutation. There is a layer in the middle of the tissue with a higher fraction than the rest, and there seems to be some sort of connection between the various areas carrying the mutation. Also, a strong similarity between marker 1 and marker 2 of the leydig cell tumor can be observed. This observation supports the theory that there is one cell lineage that carries both markers. By observing the interactive reconstruction of the leydig cell tumor, it can also be noted that the majority of the volume studied is not composed by the lineage that carries both markers. Similarly, in none of the studied tissues, was a lineage that fills the majority of the volume found.

4. Discussion

Some of the tissues have a high percentage of failed samples. Sample failure can be quite common since many procedures are performed in series, and failure in one of them will compromise the final result. The most probable source of sample failure is failure to collect it. The LCM system used relies on gravity for the sample cut by the laser to fall in the well underneath. Due to the small size and weight of the sample, as well as the static electricity present on the membrane, samples sometime don't fall. Also, high standard of quality were kept, and only samples in which a mutant fraction could be cleanly measured, with a precision below 3-5%, have been kept. The automated analysis of the electropherogram require a very high quality signal as well as internal standard, which also contributed partly to failure rate. The ability to make a three dimensional analysis of cell cell lineages in tumor depends on the quality of the tissue. Both primary tumors of the breast and prostate were low quality tissues. They were small and poorly conserved. This made the taking of 70 consecutive high-quality slices impossible. A smaller number of samples was therefore taken for the purpose of understanding the distribution of the mutations. Having the data available, the 3D plotting was still made available for viewing.

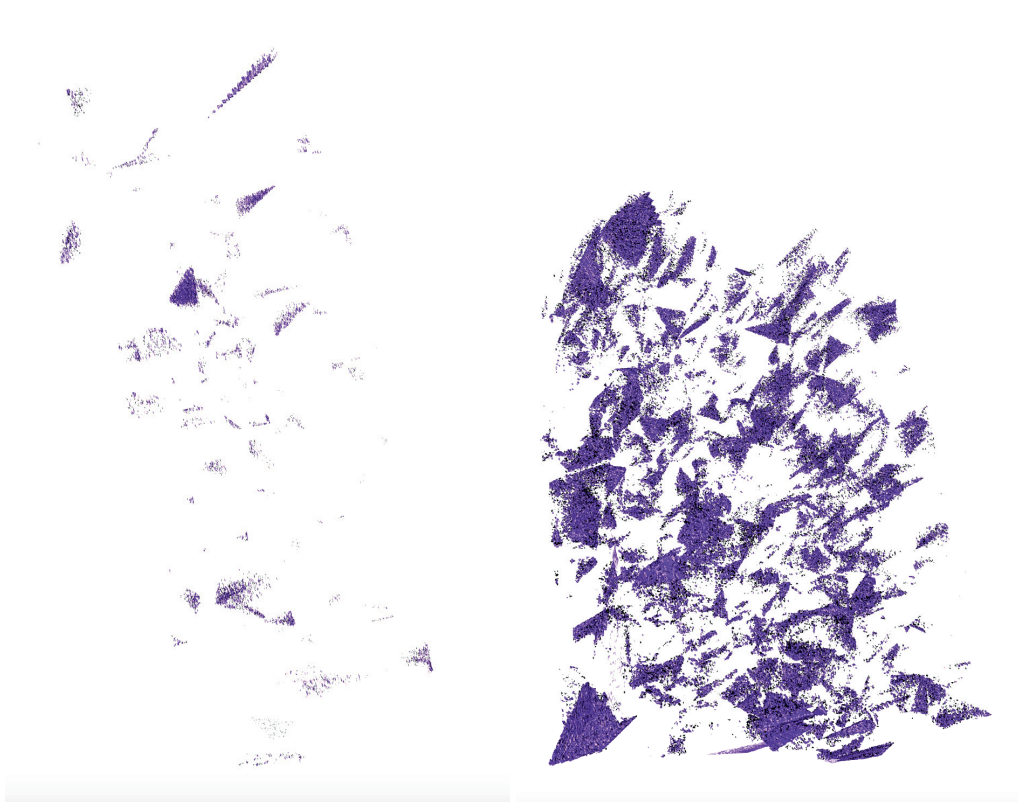


Figure 3: Snapshot of 3D reconstruction of the metastasis of the breast tumor metastasis. Liver metastasis is on the left and lymph node metastasis on the right. The same marker is tracked in both images

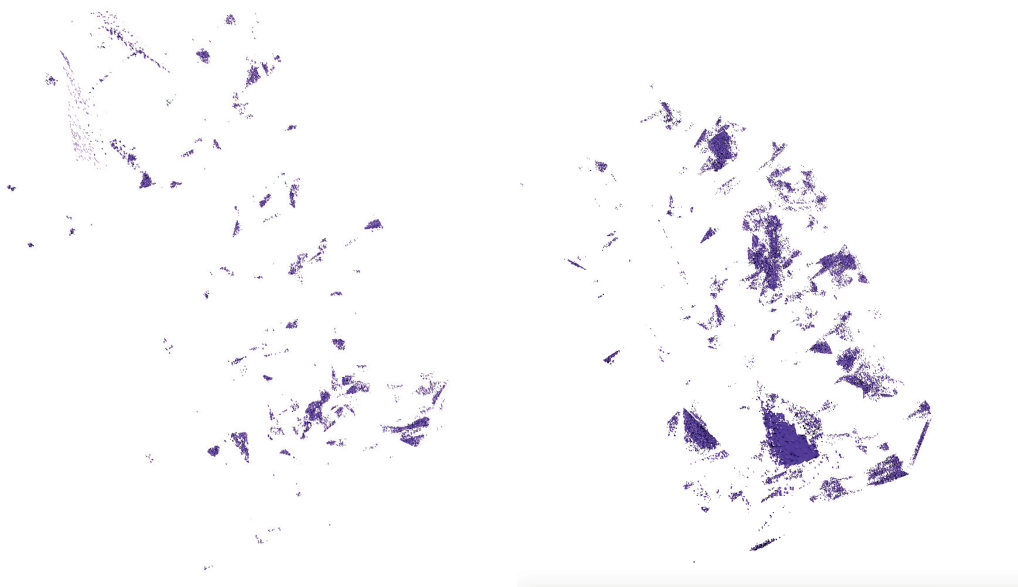


Figure 4: Snapshot of 3D reconstruction from lymph node metastasis of bladder tumor. On the left is the reconstruction of marker 1, and on the right marker 2.

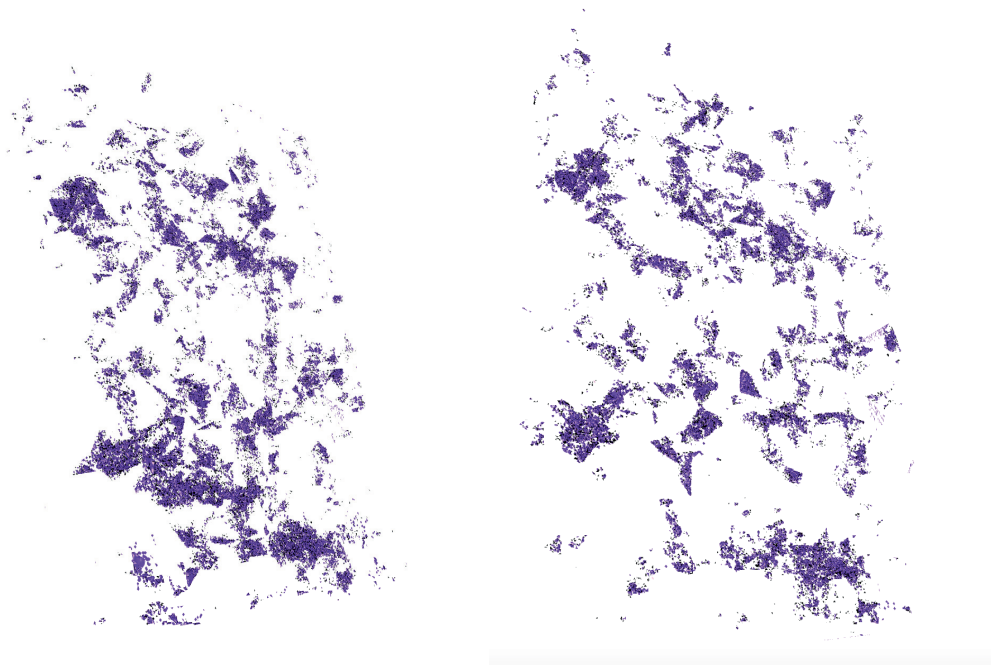


Figure 5: Snapshot of 3D reconstruction of leydig cell tumor. On the left is the reconstruction of marker 1, and on the right marker 2.

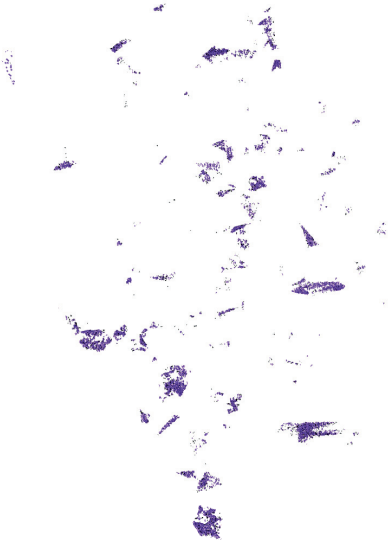


Figure 6: Snapshot of 3D reconstruction of hepatocellular carcinoma.

One of the main hypothesis made here is that following a mtDNA mutation is following a cell lineage. The presence of internal standard guarantees that the mutation being followed is the same in each sample in which it is found. The probability of two lineages carrying the same mutation is estimated to be bellow $1/700$ (Refinetti et al, under review). Other research place the estimate much lower than that [4, 5]. Also, in the cases where two markers are present, it is clear that they are strongly correlated, as figure 7 clearly shows. For example the leydig cell tumor, the histogram clearly shows that the two markers have similar distributions. Also, looking at the three dimensional reconstruction of them seem to indicate a strong similarity. With areas carrying one mutation having the other one as well. The probability of having 2 identical mutations occurring independently in two separate cell lines would be even smaller than having just one.

Standard cancer development theory suggests that metastasis are seeded by one, or few cells [16, 17, 18, 19, 20]. In the case of the breast tumor, two metastasis carry the same mutation that is found in the primary tumor. If the metastasis is necessarily seeded by a tumor derived-cell, it further reinforces the hypothesis that the lineage marked by the mutation followed is indeed

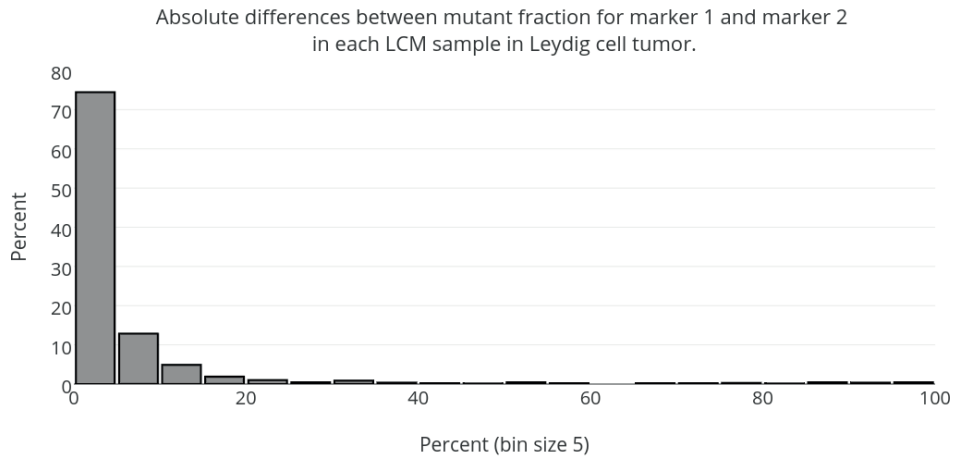


Figure 7: The absolute value of the difference between the mutant fraction of marker 1 and marker 2 is taken for each LCM sample of the leydig cell tumor. The histogram is then made with the resulting values. As can be seen, the majority of the values cluster around 0, which demonstrates a strong correlation between the two.

a tumor lineage. Furthermore, if it is true that, one, or very few cells were at the origin of the metastasis, then the only tumor lineage present in the metastasis, is the one marked by the mutation followed. This is true for both metastasis. In the primary tumor, there is no guarantee that no other tumor lineage exist. In the metastasis, however, there is reason to believe that the only tumor lineage present is the one followed by the marker. What is being observed in the three-dimensional reconstruction of the lineage, is therefore the distribution of tumor cells in a metastasis. The areas without the marker are not of tumor origin and have been recruited locally. Observing the pattern shown in figure shows that the majority of the volume is not composed of tumor derived cells. The cells derived from the tumor are scattered in small clusters that are sometimes connected. It almost looks like a percolation pattern i which the tumor cells would defuse through the tissue and then make small local growth. The presence of these cells however, disrupt the local tissue to assume a "disorganised" morphology.

5. Conclusion

Observation of three-dimensional reconstructions of tumor cell lineages seem to suggest that the majority of the tumor volume is not composed of tumor-derived cells. An observation that is further supported by the study of metastasis. If a metastasis is started by a single cell, then the descendent of this original cell must spread and diffuse (or percolate) through the host tissue, and in doing so make condition it to adopt the observed abnormal morphology. The spacial correlation between multiple markers strongly supports the credibility of the observations made as well as their consequence. As does the coherence of the observed structures in three dimensions. Hallmarks of cancer derived models of oncogenesis do not seem capable of explaining the observations made here.

Refference

- [1] FIALKOW, P.: Clonal origin of human tumors. *Biochimica et Biophysica Acta (BBA) - Reviews on Cancer* **458**(3), 283–321 (1976)
- [2] Fialkow, P.J.: Clonal origin of human tumors. *Annual Review of Medicine* (1979)
- [3] Sidransky, D., Frost, P., Von Eschenbach, A.: Clonal Origin of Bladder Cancer — *NEJM. ... England Journal of ...* (1992)
- [4] Walther, V., Alison, M.R.: Cell lineage tracing in human epithelial tissues using mitochondrial DNA mutations as clonal markers. *Wiley interdisciplinary reviews. Developmental biology* **5**(1), 103–117 (2016)
- [5] Fellous, T.G., McDonald, S.A.C., Burkert, J., Humphries, A., Islam, S., De-Alwis, N.M.W., Gutierrez-Gonzalez, L., Tadrous, P.J., Elia, G., Kocher, H.M., Bhattacharya, S., Mears, L., El-Bahrawy, M., Turnbull, D.M., Taylor, R.W., Greaves, L.C., Chinnery, P.F., Day, C.P., Wright, N.A., Alison, M.R.: A methodological approach to tracing cell lineage in human epithelial tissues. *Stem cells (Dayton, Ohio)* **27**(6), 1410–1420 (2009)
- [6] Taylor, R.W., Barron, M.J., Borthwick, G.M., Gospel, A., Chinnery, P.F., Samuels, D.C., Taylor, G.A., Plusa, S.M., Needham, S.J., Greaves,

- L.C., Kirkwood, T.B.L., Turnbull, D.M.: Mitochondrial DNA mutations in human colonic crypt stem cells. *Journal of Clinical Investigation* **112**(9), 1351–1360 (2003)
- [7] Simone, N.L., Bonner, R.F., Gillespie, J.W., Emmert-Buck, M.R., Liotta, L.A.: Laser-capture microdissection: opening the microscopic frontier to molecular analysis. *Trends in Genetics* **14**(7), 272–276 (1998)
- [8] Fialkow, P.J., Martin, G.M., Klein, G., Clifford, P., Singh, S.: Evidence for a clonal origin of head and neck tumors. *International Journal of Cancer* **9**(1), 133–142 (1972)
- [9] Humphries, A., Cereser, B., Gay, L.J., Miller, D.S.J., Das, B., Gutteridge, A., Elia, G., Nye, E., Jeffery, R., Poulsom, R., Novelli, M.R., Rodriguez-Justo, M., McDonald, S.A.C., Wright, N.A., Graham, T.A.: Lineage tracing reveals multipotent stem cells maintain human adenomas and the pattern of clonal expansion in tumor evolution. *Proceedings of the National Academy of Sciences* **110**(27), 2490–2499 (2013)
- [10] Refinetti, P., Morgenthaler, S., Ekstrom, P.O.: Cycling temperature capillary electrophoresis: A quantitative, fast and inexpensive method to detect mutations in mixed populations of human mitochondrial DNA. *Mitochondrion* **29**, 65–74 (2016)
- [11] Arstad, C., Refinetti, P., Warren, D., Giercksky, K.-E., Ekstrøm, P.O.: Scanning the mitochondrial genome for mutations by cycling temperature capillary electrophoresis. *Mitochondrial DNA. Part A, DNA mapping, sequencing, and analysis*, 1–12 (2016)
- [12] Steger, G.: Thermal denaturation of double-stranded nucleic acids: prediction of temperatures critical for gradient gel electrophoresis and polymerase chain reaction. *Nucleic acids research* **22**(14), 2760–2768 (1994)
- [13] Fixman, M., Freire, J.J.: Theory of DNA melting curves. *Biopolymers* **16**(12), 2693–2704 (1977)
- [14] Ekstrom, P.O., Khrapko, K., Li-Sucholeiki, X.-C., Hunter, I.W., Thilly, W.G.: Analysis of mutational spectra by denaturing capillary electrophoresis. *Nature Protocols* **3**(7), 1153–1166 (2008)

- [15] Ekstrom, P.O., Bjørheim, J., Thilly, W.G.: Technology to accelerate pangenomic scanning for unknown point mutations in exonic sequences: cycling temperature capillary electrophoresis (CTCE). *BMC Genetics* **8**(1), 54 (2007)
- [16] Hanahan, D., Weinberg, R.A.: Hallmarks of Cancer: The Next Generation. *Cell* **144**(5), 646–674 (2011)
- [17] Hanahan, D., Weinberg, R.A.: The Hallmarks of Cancer. *Cell* **100**(1), 57–70 (2000)
- [18] Liu, W., Laitinen, S., Khan, S., Vihinen, M., Kowalski, J., Yu, G., Chen, L., Ewing, C.M., Eisenberger, M.A., Carducci, M.A., Nelson, W.G., Yegnasubramanian, S., Luo, J., Wang, Y., Xu, J., Isaacs, W.B., Visakorpi, T., Bova, G.S.: Copy number analysis indicates monoclonal origin of lethal metastatic prostate cancer. *Nature medicine* **15**(5), 559–565 (2009)
- [19] Fidler, I., Hart, I.: Biological diversity in metastatic neoplasms: origins and implications. *Science (New York, N.Y.)* **217**(4564), 998–1003 (1982)
- [20] Polyak, K.: Breast cancer: origins and evolution. *Journal of Clinical Investigation* **117**(11), 3155–3163 (2007)

Chapter 10. Three-dimensional Lineage Tracing

its

11 Conclusion

It is possible to trace cell lineages in human tissues. A bold but accurate statement.

We were not the first to attempt studying lineage tracing in humans, and their tumors. As mentioned in the Introduction (chapter 1), many researchers have attempted, with varying degrees of success, to solve the problem. We are not even the first to use somatic mitochondrial mutations as a marker. However, none of the previous approaches were completely successful in providing a general applicable method.

The contribution of this thesis is a solution to the problem. Optimising CTCE to detect and quantify mtDNA mutations in samples, has dropped the cost of the analysis, and improved the throughput and the precision. We could then thus process a much higher number of samples than anyone before, opening up the possibility of sampling a volume. Since cell lineages grow in a volume, a three dimensional analysis is the only way to get an accurate observation of the processes involved. Also, by taking a critical view on the histo-pathological analysis of cancers, we deduced that selective sampling was the wrong way to proceed. Systematic sampling (grid of identical small samples) of tumor slices showed no predictable pattern of mutations. This irregular dispersion of mutations required an independent validation by extending the observations in to the third dimension, the depth of the tissue. This in turn led to the visualisation of three dimensional reconstruction of cell lineages. Possibly for the first time with good precision and resolution in tumors. The final result is a tried and tested protocol for the analysis of cell lineages in three dimension in human tissues. The observations clearly show that tumor lineages do not grow in an organised fashion. It would seem from, the observations made that tumors grow by infiltrating the local tissue. This infiltration causes the tissue to acquire its abnormal morphology. The mechanisms and processes through which this happens are not fully understood. Another possible explanation for the observed results, could be the seeding of metastasis by multiple cells. It has been suggested, in a recent paper by Cheung et al. [100], that in the mouse model, metastasis can be seeded by a cluster of cells from the primary tumor. If such was also the case in humans, then the metastasis would contain a similar mix of lineages as the primary tumor, and could therefore have only a small fraction of its volume filled by the lineage carrying the markers

we followed. This explanation challenges the current understanding of how tumor metastasis originate, while upholding the hypothesis that most of the tumor mass is composed of tumor derived cells. It is clear that an unbiased sampling of tumor tissue in three dimensions, coupled with a reliable analysis method, leads to surprising observations that cannot be explained by the current oncogenesis paradigm.

11.1 Future

Among the first direct applications of the methods described in this thesis is the study of mitochondrial mutations. Even if it is not lineage tracing *per se*, our CTCE based method for the detection of mtDNA mutations could contribute significantly to the field. As has been mentioned already, mitochondrial diseases are caused by deleterious mtDNA mutations reaching high fraction in some tissues. Sequencing the mtDNA in the blood has no or little chance of picking up the signal, because it is insufficiently sensitive (anything below 30% can be missed). Our method, with 1% sensitivity, could pick it up and therefore place a definitive diagnosis on many patients for which current methods cannot do it.

The ability to study cell lineages directly in human tissues offers many further possibilities. There are a large number of unanswered question of human biology and medicine that could be answered or elucidated by applying the techniques developed in this thesis. The various observations relating to cancers we described in this thesis should already raise many questions in the reader's mind. What is a tumor really? Is it the bulk of the anomalous mass or the small group of cells causing the surrounding tissue to behave as it does? Does this small group of cells even exist? Maybe, more three dimensional reconstructions of tumor lineages can provide answers. Our observations seem to indicate that the majority of the tumor mass, or its metastasis, is not composed of tumor derived cells. And yet, visual observation of the sample, shows an anomalous mass of cells. This points to a poorly understood interaction between tumor derived cells and surrounding tissue.

A similar example could be wound healing. While cancer is an uncontrolled proliferation of cells, wound healing is a controlled proliferation of cells, that still requires further investigation. Mapping the lineages that compose a human scar would shed light onto the process through which they are formed. The origin of the cells that form the scar could also lead to the better understanding of why some wounds do not heal, a major problem in traumatology and surgery, or inversely, why some wounds "over-heal" forming keloids. Understanding the relationship between cell lineages in the human body would lead to understanding the process of growth and differentiation that brings us, humans, from a single cell to a fully functional organism. Repeating experiments similar to the ones we described, and increasing the number of samples taken, would reconstruct the various lineages that compose human organs. In doing so, we would better understand the differentiation paths taken by human stem cells during development.

Bibliography

- [1] F J Miller. Precise determination of mitochondrial DNA copy number in human skeletal and cardiac muscle by a PCR-based assay: lack of change of copy number with age. *Nucleic acids research*, 31(11):61e–61, June 2003.
- [2] Tariq G Fellous, Stuart A C McDonald, Julia Burkert, Adam Humphries, Shahriar Islam, Nemantha M W De-Alwis, Lydia Gutierrez-Gonzalez, Paul J Tadrous, George Elia, Hemant M Kocher, Satyajit Bhattacharya, Lisa Mears, Mona El-Bahrawy, Douglas M Turnbull, Robert W Taylor, Laura C Greaves, Patrick F Chinnery, Christopher P Day, Nicholas A Wright, and Malcolm R Alison. A Methodological Approach to Tracing Cell Lineage in Human Epithelial Tissues. *Stem cells (Dayton, Ohio)*, 27(6):1410–1420, June 2009.
- [3] J L Elson, D C Samuels, D M Turnbull, and P F Chinnery. Random Intracellular Drift Explains the Clonal Expansion of Mitochondrial DNA Mutations with Age. *The American Journal of Human Genetics*, 68(3):802–806, March 2001.
- [4] Robert W Taylor, Martin J Barron, Gillian M Borthwick, Amy Gospel, Patrick F Chinnery, David C Samuels, Geoffrey A Taylor, Stefan M Plusa, Stephanie J Needham, Laura C Greaves, Thomas B L Kirkwood, and Douglass M Turnbull. Mitochondrial DNA mutations in human colonic crypt stem cells. *Journal of Clinical Investigation*, 112(9):1351–1360, November 2003.
- [5] H A Coller, K Khrapko, N D Bodyak, E Nekhaeva, P Herrero-Jimenez, and W G Thilly. High frequency of homoplasmic mitochondrial DNA mutations in human tumors can be explained without selection. *Nature Genetics*, 28(2):147–150, June 2001.
- [6] Viola Walther and Malcolm R Alison. Cell lineage tracing in human epithelial tissues using mitochondrial DNA mutations as clonal markers. *Wiley Interdisciplinary Reviews: Developmental Biology*, 5(1):103–117, August 2015.
- [7] Walther Vogt. Gestaltungsanalyse am Amphibienkeim mit örtlicher Vitalfärbung. Vorwort über Wege und Ziele. *Wilhelm Roux' Archiv für Entwicklungsmechanik der Organismen*, 106(1-4):542–610, December 1925.

Bibliography

- [8] Kai Kretzschmar and Fiona M Watt. Lineage Tracing. *Cell*, 148(1-2):33–45, January 2012.
- [9] Margaret E Buckingham and Sigolène M Meilhac. Tracing Cells for Tracking Cell Lineage and Clonal Behavior. *Developmental Cell*, 21(3):394–409, September 2011.
- [10] E Zwilling. Induction of the Olfactory Placode by the Forebrain in *Rana pipiens*. *Experimental Biology and Medicine*, 31(8):933–935, May 1934.
- [11] ANDRZEJ K TARKOWSKI. Mouse Chimæras Developed from Fused Eggs. *Nature*, 190(4779):857–860, June 1961.
- [12] B Mintz. Genetic Mosaicism in Adult Mice of Quadriparental Lineage. *Science (New York, N.Y.)*, 148(3674):1232–1233, May 1965.
- [13] Cédric Blanpain and Elaine Fuchs. Epidermal Stem Cells of the Skin. *Annual Review of Cell and Developmental Biology*, 22(1):339–373, November 2006.
- [14] K Hatta, T Schilling, R BreMiller, and C Kimmel. Specification of jaw muscle identity in zebrafish: correlation with engrailed-homeoprotein expression. *Science (New York, N.Y.)*, 250(4982):802–805, November 1990.
- [15] Andrew D Rhim, Emily T Mirek, Nicole M Aiello, Anirban Maitra, Jennifer M Bailey, Florencia McAllister, Maximilian Reichert, Gregory L Beatty, Anil K Rustgi, Robert H Vonderheide, Steven D Leach, and Ben Z Stanger. EMT and Dissemination Precede Pancreatic Tumor Formation. *Cell*, 148(1-2):349–361, January 2012.
- [16] Kristin Hope and Mickie Bhatia. Clonal interrogation of stem cells. *Nature Methods*, 8(4):S36–S40, April 2011.
- [17] P C Orban, D Chui, and J D Marth. Tissue- and site-specific DNA recombination in transgenic mice. *Proceedings of the National Academy of Sciences*, 89(15):6861–6865, August 1992.
- [18] Daniel Metzger and Pierre Chambon. Site- and Time-Specific Gene Targeting in the Mouse. *Methods (San Diego, Calif.)*, 24(1):71–80, May 2001.
- [19] Alexandra Van Keymeulen and Cédric Blanpain. Tracing epithelial stem cells during development, homeostasis, and repair. *The Journal of Cell Biology*, 197(5):575–584, May 2012.
- [20] P J Fialkow. Clonal Origin of Human Tumors. *Annual Review of Medicine*, 30(1):135–143, February 1979.
- [21] R G Wiggans, P V Woolley, Tarilyn Smythe, D Hoth, J S Macdonald, Linda Green, and P S Schein. Phase-II trial of tamoxifen in advanced breast cancer. *Cancer Chemotherapy and Pharmacology*, 3(1), August 1979.

- [22] E P Benditt and J M Benditt. Evidence for a Monoclonal Origin of Human Atherosclerotic Plaques. *Proceedings of the National Academy of Sciences*, 70(6):1753–1756, June 1973.
- [23] A Alonso. Real-time PCR designs to estimate nuclear and mitochondrial DNA copy number in forensic and ancient DNA studies. *Forensic science international*, 139(2-3):141–149, January 2004.
- [24] Helen A L Tuppen, Emma L Blakely, Douglass M Turnbull, and Robert W Taylor. Mitochondrial DNA mutations and human disease. *Biochimica et Biophysica Acta (BBA) - Bioenergetics*, 1797(2):113–128, February 2010.
- [25] Robert N Lightowlers, Patrick F Chinnery, Douglass M Turnbull, and Neil Howell. Mammalian mitochondrial genetics: heredity, heteroplasmy and disease. *Trends in Genetics*, 13(11):450–455, November 1997.
- [26] Brigitte Pakendorf and Mark Stoneking. MITOCHONDRIAL DNA AND HUMAN EVOLUTION. *Annual Review of Genomics and Human Genetics*, 6(1):165–183, September 2005.
- [27] Laura C Greaves and Doug M Turnbull. Mitochondrial DNA mutations and ageing. *Biochimica et Biophysica Acta (BBA) - General Subjects*, 1790(10):1015–1020, October 2009.
- [28] J Shoffner. Oxidative Phosphorylation Diseases and Mitochondrial DNA Mutations: Diagnosis and Treatment. *Annual Review of Nutrition*, 14(1):535–568, January 1994.
- [29] Manuel B Graeber and Ulrich Müller. Recent developments in the molecular genetics of mitochondrial disorders. *Journal of the Neurological Sciences*, 153(2):251–263, January 1998.
- [30] Salvatore DiMauro and Eric A Schon. Mitochondrial DNA mutations in human disease. *American journal of medical genetics*, 106(1):18–26, 2001.
- [31] Laura C Greaves, Amy K Reeve, Robert W Taylor, and Doug M Turnbull. Mitochondrial DNA and disease. *The Journal of pathology*, 226(2):274–286, November 2011.
- [32] Chan Bae Park and Nils-Göran Larsson. Mitochondrial DNA mutations in disease and aging. *The Journal of Cell Biology*, 193(5):809–818, May 2011.
- [33] M Zeviani. Mitochondrial disorders. *Brain*, 127(10):2153–2172, September 2004.
- [34] R G Boles. Quantification of Mitochondrial DNA Heteroplasmy by Temporal Temperature Gradient Gel Electrophoresis. *Clinical chemistry*, 49(1):198–200, January 2003.
- [35] Michael D Brown and Douglas C Wallace. Molecular basis of mitochondrial DNA disease. *Journal of Bioenergetics and Biomembranes*, 26(3):273–289, June 1994.

Bibliography

- [36] L L Cavalli-Sforza and E Minch. Paleolithic and Neolithic Lineages in the European Mitochondrial Gene Pool. *The American Journal of Human Genetics*, 61(1):247–251, July 1997.
- [37] I B Colson, M B Richards, J F Bailey, B C Sykes, and R E M Hedges. DNA Analysis of Seven Human Skeletons Excavated from the Terp of Wijnaldum. *Journal of Archaeological Science*, 24(10):911–917, October 1997.
- [38] Martin Richards, Vincent Macaulay, Eileen Hickey, Emilce Vega, Bryan Sykes, Valentina Guida, Chiara Rengo, Daniele Sellitto, Fulvio Cruciani, Toomas Kivisild, Richard Villems, Mark Thomas, Serge Rychkov, Oksana Rychkov, Yuri Rychkov, Mukaddes Gölge, Dimitar Dimitrov, Emmeline Hill, Dan Bradley, Valentino Romano, Francesco Calì, Giuseppe Vona, Andrew Demaine, Surinder Papiha, Costas Triantaphyllidis, Gheorghe Stefanescu, Jiří Hatina, Michele Belledi, Anna Di Rienzo, Andrea Novelletto, Ariella Oppenheim, Søren Nørby, Nadia Al-Zaheri, Silvana Santachiara-Benerecetti, Rosaria Scozzari, Antonio Torroni, and Hans-Jürgen Bandelt. Tracing European Founder Lineages in the Near Eastern mtDNA Pool. *The American Journal of Human Genetics*, 67(5):1251–1276, November 2000.
- [39] B M Henn, C R Gignoux, M W Feldman, and J L Mountain. Characterizing the Time Dependency of Human Mitochondrial DNA Mutation Rate Estimates. *Molecular Biology and Evolution*, 26(1):217–230, October 2008.
- [40] Ron Lundstrom, Simon Tavaré, and R H Ward. Modeling the evolution of the human mitochondrial genome. *Mathematical biosciences*, 112(2):319–335, December 1992.
- [41] Christine Reder. *Acta Biotheoretica*, 49(4):235–245, 2001.
- [42] Corinna Herrnstadt, Gwen Preston, Richard Andrews, Patrick Chinnery, Robert N Lightowlers, Douglass M Turnbull, Iwona Kubacka, and Neil Howell. A high frequency of mtDNA polymorphisms in HeLa cell sublines. *Mutation Research/Fundamental and Molecular Mechanisms of Mutagenesis*, 501(1-2):19–28, April 2002.
- [43] K Khrapko, H A Coller, P C André, X C Li, J S Hanekamp, and W G Thilly. Mitochondrial mutational spectra in human cells and tissues. *Proceedings of the National Academy of Sciences*, 94(25):13798–13803, December 1997.
- [44] John K Blackwood, Stuart C Williamson, Laura C Greaves, Laura Wilson, Anastasia C Rigas, Raveen Sandher, Robert S Pickard, Craig N Robson, Douglass M Turnbull, Robert W Taylor, and Rakesh Heer. In situ lineage tracking of human prostatic epithelial stem cell fate reveals a common clonal origin for basal and luminal cells. *The Journal of pathology*, 225(2):181–188, September 2011.
- [45] Patrick F Chinnery, David C Samuels, Joanna Elson, and Douglass M Turnbull. Accumulation of mitochondrial DNA mutations in ageing, cancer, and mitochondrial disease: is there a common mechanism? *The Lancet*, 360(9342):1323–1325, October 2002.

- [46] D A Cottrell, E L Blakely, M A Johnson, P G Ince, G M Borthwick, and D M Turnbull. Cytochrome c oxidase deficient cells accumulate in the hippocampus and choroid plexus with age. *Neurobiology of aging*, 22(2):265–272, March 2001.
- [47] Kenji Takahashi, Takashi Kohno, Shingo Matsumoto, Yukihiro Nakanishi, Yasuhito Arai, Toshiyoshi Fujiwara, Noriaki Tanaka, and Jun Yokota. Clonality and heterogeneity of pulmonary blastoma from the viewpoint of genetic alterations: A case report. *Lung Cancer*, 57(1):103–108, July 2007.
- [48] Kornelia Polyak. Breast cancer: origins and evolution. *Journal of Clinical Investigation*, 117(11):3155–3163, November 2007.
- [49] Wennuan Liu, Sari Laitinen, Sofia Khan, Mauno Vihinen, Jeanne Kowalski, Guoqiang Yu, Li Chen, Charles M Ewing, Mario A Eisenberger, Michael A Carducci, William G Nelson, Srinivasan Yegnasubramanian, Jun Luo, Yue Wang, Jianfeng Xu, William B Isaacs, Tapio Visakorpi, and G Steven Bova. Copy number analysis indicates monoclonal origin of lethal metastatic prostate cancer. *Nature medicine*, 15(5):559–565, April 2009.
- [50] Douglas Hanahan and Robert A Weinberg. Hallmarks of Cancer: The Next Generation. *Cell*, 144(5):646–674, March 2011.
- [51] Nicholas E Navin and James Hicks. Tracing the tumor lineage. *Molecular oncology*, 4(3):267–283, May 2010.
- [52] P Nowell. The clonal evolution of tumor cell populations. *Science (New York, N.Y.)*, 194(4260):23–28, October 1976.
- [53] P ARMITAGE and R DOLL. The Age Distribution of Cancer and a Multi-stage Theory of Carcinogenesis. *British Journal of Cancer*, 8(1):1–12, March 1954.
- [54] P ARMITAGE and R DOLL. A Two-stage Theory of Carcinogenesis in Relation to the Age Distribution of Human Cancer. *British Journal of Cancer*, 11(2):161–169, June 1957.
- [55] Philip J Fialkow. The Origin and Development of Human Tumors Studied with Cell Markers. *New England Journal of Medicine*, 291(1):26–35, July 1974.
- [56] Philip J Fialkow, George M Martin, George Klein, Peter Clifford, and Surjit Singh. Evidence for a clonal origin of head and neck tumors. *International Journal of Cancer*, 9(1):133–142, January 1972.
- [57] Douglas Hanahan and Robert A Weinberg. The Hallmarks of Cancer. *Cell*, 100(1):57–70, January 2000.
- [58] Brett Hall, Jennifer Dembinski, A Sasser, Matus Studeny, Michael Andreeff, and Frank Marini. Mesenchymal Stem Cells in Cancer: Tumor-Associated Fibroblasts and Cell-Based Delivery Vehicles. *International Journal of Hematology*, 86(1):8–16, July 2007.

Bibliography

- [59] Raghu Kalluri and Michael Zeisberg. Fibroblasts in cancer. *Nat Rev Cancer*, 6(5):392–401, March 2006.
- [60] So Yeon Park, Mithat Gönen, Hee Jung Kim, Franziska Michor, and Kornelia Polyak. Cellular and genetic diversity in the progression of in situ human breast carcinomas to an invasive phenotype. *Journal of Clinical Investigation*, 120(2):636–644, February 2010.
- [61] Gregory Driessens, Benjamin Beck, Amélie Caauwe, Benjamin D Simons, and Cédric Blanpain. Defining the mode of tumour growth by clonal analysis. *Nature*, 488(7412):527–530, August 2012.
- [62] Stewart Sell. Stem cell origin of cancer and differentiation therapy. *Critical Reviews in Oncology/Hematology*, 51(1):1–28, July 2004.
- [63] Thea D Tlsty. Stromal cells can contribute oncogenic signals. *Seminars in cancer biology*, 11(2):97–104, April 2001.
- [64] Gerald R Cunha, Simon W Hayward, Y Z Wang, and William A Ricke. Role of the stromal microenvironment in carcinogenesis of the prostate. *International Journal of Cancer*, 107(1):1–10, August 2003.
- [65] Thea D Tlsty and Patrick W Hein. Know thy neighbor: stromal cells can contribute oncogenic signals. *Current Opinion in Genetics & Development*, 11(1):54–59, February 2001.
- [66] Kornelia Polyak, Izhak Haviv, and Ian G Campbell. Co-evolution of tumor cells and their microenvironment. *Trends in Genetics*, 25(1):30–38, January 2009.
- [67] Alexander B Mohseny and Pancras C W Hogendoorn. Concise Review: Mesenchymal Tumors: When Stem Cells Go Mad. *Stem cells (Dayton, Ohio)*, 29(3):397–403, March 2011.
- [68] Nadine T Gaisa, Trevor A Graham, Stuart Ac McDonald, Richard Poulson, Axel Heidenreich, Gerhard Jakse, Ruth Knuechel, and Nicholas A Wright. Clonal architecture of human prostatic epithelium in benign and malignant conditions. *The Journal of pathology*, 225(2):172–180, August 2011.
- [69] Jason C Mills and Ramesh A Shivdasani. Gastric Epithelial Stem Cells. *Gastroenterology*, 140(2):412–424, February 2011.
- [70] Yong-Gang Yao, Richard W Childs, Sachiko Kajigaya, J Philip McCoy Jr., and Neal S Young. Mitochondrial DNA Sequence Heterogeneity of Single CD34 +Cells After Nonmyeloablative Allogeneic Stem Cell Transplantation. *Stem cells (Dayton, Ohio)*, 25(10):2670–2676, October 2007.
- [71] Ronald D Barr and Philip J Fialkow. Clonal Origin of Chronic Myelocytic Leukemia. *New England Journal of Medicine*, 289(6):307–309, August 1973.

- [72] Philip J Fialkow, Robert J Jacobson, and Thalia Papayannopoulou. Chronic myelocytic leukemia: Clonal origin in a stem cell common to the granulocyte, erythrocyte, platelet and monocyte/macrophage. *The American Journal of Medicine*, 63(1):125–130, July 1977.
- [73] P J Fialkow, S M Gartler, and A Yoshida. Clonal origin of chronic myelocytic leukemia in man. *Proceedings of the National Academy of Sciences*, 58(4):1468–1471, October 1967.
- [74] G L Chen and J T Prchal. X-linked clonality testing: interpretation and limitations. *Blood*, 110(5):1411–1419, September 2007.
- [75] Maho Shibata and Michael M Shen. The roots of cancer: Stem cells and the basis for tumor heterogeneity. *BioEssays*, 35(3):253–260, October 2012.
- [76] B Linnartz. Comprehensive Scanning of Somatic Mitochondrial DNA Alterations in Acute Leukemia Developing from Myelodysplastic Syndromes. *Cancer research*, 64(6):1966–1971, March 2004.
- [77] Douglas C Wallace. Mitochondria and cancer. *Nat Rev Cancer*, 12(10):685–698, September 2012.
- [78] Michael W Klymkowsky and Pierre Savagner. Epithelial-Mesenchymal Transition. *The American Journal of Pathology*, 174(5):1588–1593, May 2009.
- [79] Bert Vogelstein, Kornelia Polyak, Yunbo Li, Hong Zhu, Christoph Lengauer, James K V Willson, Sanford D Markowitz, Michael A Trush, and Kenneth W Kinzler. *Nature Genetics*, 20(3):291–293, November 1998.
- [80] Pen-Hui Yin, Cheng-Chung Wu, Jin-Ching Lin, Chin-Wen Chi, Yau-Huei Wei, and Hsin-Chen Lee. Somatic mutations of mitochondrial genome in hepatocellular carcinoma. *Mitochondrion*, 10(2):174–182, March 2010.
- [81] Aekyong Kim. Mitochondrial DNA Somatic Mutation in Cancer. *Toxicological Research*, 30(4):235–242, December 2014.
- [82] Amanda Ramos, Cristina Santos, Luis Alvarez, Ramon Nogués, and Maria Pilar Aluja. Human mitochondrial DNA complete amplification and sequencing: A new validated primer set that prevents nuclear DNA sequences of mitochondrial origin co-amplification. *Electrophoresis*, 30(9):1587–1593, May 2009.
- [83] P FIALKOW. Clonal origin of human tumors. *Biochimica et Biophysica Acta (BBA) - Reviews on Cancer*, 458(3):283–321, October 1976.
- [84] Philip J Fialkow, A Michael Denman, Robert J Jacobson, and Mark N Lowenthal. Chronic Myelocytic Leukemia. *Journal of Clinical Investigation*, 62(4):815–823, October 1978.
- [85] J M Friedman and Philip I Fialkow. Cell Marker Studies of Human Tumorigenesis. *Immunological Reviews*, 28(1):17–33, January 1976.

Bibliography

- [86] P FIALKOW. CLONAL ORIGIN FOR INDIVIDUAL BURKTTT TUMOURS. *The Lancet*, 295(7643):384–386, February 1970.
- [87] J M Friedman and Philip I Fialkow. Cell Marker Studies of Human Tumorigenesis. *Immunological Reviews*, 28(1):17–33, January 1976.
- [88] Masaharu KOBAYASHI. Seals for Vacuums. *SHINKU*, 10(11):388–394, 1967.
- [89] Khaled K Abu-Amero, Ali S Alzahrani, Minjing Zou, and Yufei Shi. High frequency of somatic mitochondrial DNA mutations in human thyroid carcinomas and complex I respiratory defect in thyroid cancer cell lines. *Oncogene*, 24(8):1455–1460, December 2004.
- [90] J J Barcia. The Giemsa Stain: Its History and Applications. *International Journal of Surgical Pathology*, 15(3):292–296, July 2007.
- [91] Célia H Tengan and Carlos T Moraes. Detection and Analysis of Mitochondrial DNA Deletions by Whole Genome PCR. *Biochemical and Molecular Medicine*, 58(1):130–134, June 1996.
- [92] K Khrapko, N Bodyak, W G Thilly, N J van Orsouw, X Zhang, H A Coller, T T Perls, M Upton, J Vijg, and J Y Wei. Cell-by-cell scanning of whole mitochondrial genomes in aged human heart reveals a significant fraction of myocytes with clonally expanded deletions. *Nucleic acids research*, 27(11):2434–2441, January 1999.
- [93] Paulo Refinetti, Stephan Morgenthaler, and Per O Ekstrom. Cycling temperature capillary electrophoresis: A quantitative, fast and inexpensive method to detect mutations in mixed populations of human mitochondrial DNA. *Mitochondrion*, 29:65–74, July 2016.
- [94] Christian Arstad, Paulo Refinetti, David Warren, Karl-Erik Giercksky, and Per Olaf Ekstrøm. Scanning the mitochondrial genome for mutations by cycling temperature capillary electrophoresis. *Mitochondrial DNA Part A*, pages 1–12, October 2016.
- [95] Annette Torgunrud Kristensen, Johan N Wiig, Stein G Larsen, Karl-Erik Giercksky, and Per O Ekstrom. Molecular detection (k-ras) of exfoliated tumour cells in the pelvis is a prognostic factor after resection of rectal cancer? *BMC Cancer*, 8(1):224, July 2008.
- [96] Christian Arstad, Paulo Refinetti, Annette Torgunrud Kristensen, Karl-Erik Giercksky, and Per Olaf Ekstrøm. Is detection of intraperitoneal exfoliated tumor cells after surgical resection of rectal cancer a prognostic factor of survival? *BMC Cancer*, 17(1):406, June 2017.
- [97] P H Yin, H C Lee, G Y Chau, Y T Wu, S H Li, W Y Lui, Y H Wei, T Y Liu, and C W Chi. Alteration of the copy number and deletion of mitochondrial DNA in human hepatocellular carcinoma. *British Journal of Cancer*, May 2004.

- [98] Antonio Salas, Yong-Gang Yao, Vincent Macaulay, Ana Vega, Ángel Carracedo, and Hans-Jürgen Bandelt. A Critical Reassessment of the Role of Mitochondria in Tumorigenesis. *PLoS Medicine*, 2(11):e296, October 2005.
- [99] Paulo Refinetti, Christian Arstad, William G Thilly, Stephan Morgenthaler, and Per Olaf Ekstrøm. Mapping mitochondrial heteroplasmy in a Leydig tumor by laser capture microdissection and cycling temperature capillary electrophoresis. *BMC clinical pathology*, 17(1):6, 2017.
- [100] Kevin J Cheung and Andrew J Ewald. A collective route to metastasis: Seeding by tumor cell clusters. *Science (New York, N.Y.)*, 352(6282):167–169, April 2016.



Paulo Refinetti

BIRTH

Born 26/10/1988
in Paris, France

NATIONALITY

Italian, Brazilian

CIVIL STATUS

Single, no children

PERMANENT ADDRESS

Via Rossini 2
Carnago (VA), 21040
Italy

PHONE

+41 76 222 07 95

FAX

+39 0331 98 53 77

EMAIL

pauloref@mit.edu
paulo.refinetti@gmail.com
paulo.refinetti@epfl.ch

Profile

Think, ask questions and question the answers, is my practice. Validate the Way, choose the goal, perfect the method, is my philosophy. Develop reasoning, add value and solve problems, is my Way forward.

Education

SWISS FEDERAL INSTITUTE OF TECHNOLOGY (EPFL), LAUSANNE SWITZERLAND— 2011-NOW

PhD in Applied Statistics, in collaboration with MIT and Norwegian Radium Hospital.

Most knowledge of developmental biology today, comes from animal models. Although instructive, they are insufficient to understand pathological processes such as cancer, failed wound healing, atherosclerosis, etc.

My goal is to allow the study of developmental biology in humans. Combining Mathematical, Engineering and Biological techniques, we develop the methods to trace cell lineages directly in human tissues. With my supervisors, we have developed the biological, mechanical and software tools to identify cell lineages in human tumors, and follow them in 3D both in the primary tumors and in the metastasis. The method can now be expanded for both diagnostic purposes and the study of organogenesis in humans.

Supervision of Prof. S. Morgenthaler, Chair of applied statistics, EPFL. Co-supervision of Prof W.G. Thilly, Laboratory of epidemiology and stem cell biology, MIT. Co-supervision of Dr. Per Ekstrom, Department of Cancer Research, Radiumhospital, Oslo.

Courses followed in Algorithm Design, Statistical Mechanics and Mathematical Statistics. Courses taught in Probability, Statistics and Analytical Mechanics.

SWISS FEDERAL INSTITUTE OF TECHNOLOGY (EPFL), LAUSANNE, SWITZERLAND — 2009-2011

Masters in Bio-Engineering from the School of Life Sciences with chosen additional options in Biomechanics, Statistical Physics, Biophysics, Image Processing, Signal Processing

Master thesis at MIT in Epidemiology, supervised by Prof. William G. Thilly and Prof. Stephan Morgenthaler

SWISS FEDERAL INSTITUTE OF TECHNOLOGY (EPFL), LAUSANNE, SWITZERLAND — 2006-2009

Bachelor in Engineering from the School of Life Sciences
Bachelor project on Fractal geometry and Cancer supervised by Prof. Stephan Morgenthaler

SCUOLA EUROPEA VARESE, VARESE ITALY — 2002-2006

European Baccalaureat, 1st language French, 2nd language English with chosen additional options in Mathematics, Physics, Chemistry, Economics and Philosophy
Finished with best Baccalaureat of the class of 2006 with a GPA of 88/100



BIRTH

Born 26/10/1988
in Paris, France

NATIONALITY

Italian, Brazilian

CIVIL STATUS

Single, no children

PERMANENT ADDRESS

Via Rossini 2
Carnago (VA), 21040
Italy

PHONE

+41 76 222 07 95

FAX

+39 0331 98 53 77

EMAIL

pauloref@mit.edu
paulo.refinetti@gmail.com
paulo.refinetti@epfl.ch

Experience

OSLO RADIUMHOSPITAL, OSLO, NORWAY — 2011-ONGOING

Designing methods for lineage tracing in humans. Improving the methods of mutational spectrometry to follow mitochondrial DNA in human tissues.

MASSACHUSETTS INSTITUTE OF TECHNOLOGY, BOSTON, USA — 2010-2011

Visiting student in the Laboratory of Epidemiology supervised by Prof. William G. Thilly
Master thesis in Epidemiology.

GERMAN CANCER RESEARCH CENTER (DKFZ), HEIDELBERG, GERMANY — 2011

Working on the Swedish Family cancer database. The goal of the stay was to test the hypothesis of a familial risk for cancer. This included designing and implementing specific statistical tests.

GLOBALFOOD, SAO PAULO, BRAZIL — 2004

Internship in Laboratory practices
Assistance in product development in a leading food additives company

Languages

	SPOKEN	WRITTEN
French	Native Language	Native Language
Portuguese	Native Language	Native Language
Italian	Native Language	Native Language
English	Language of Studies	Language of Studies
German	Good	Basic Working Knowledge
Japanese	Conversational	Basic Knowledge
Spanish	Basic Working Knowledge	Basic Working Knowledge

Scientific Skills

MATHEMATICS

Mathematical Statistics, Applied Statistics, Analysis, Numerical Analysis

PROGRAMMING

Algorithm Design, Software Architecture, Database Management Systems

BIOLOGY

Quantitative Biology, Epidemiology, Molecular Biology, Genetics, Histology, Cell Culture

ENGINEERING

Image Processing, Mechanics, Material Sciences, Thermodynamics

PROGRAMMING LANGUAGES

C++, Objective C, perl, MATLAB, R, Java, HTML, CSS, Java-script, bash, SAS, python

OFFICE SOFTWARE

LaTex, Microsoft Office, iWork, NeoOffice, OpenOffice



BIRTH

Born 26/10/1988
in Paris, France

NATIONALITY

Italian, Brazilian

CIVIL STATUS

Single, no children

PERMANENT ADDRESS

Via Rossini 2
Carnago (VA), 21040
Italy

PHONE

+41 76 222 07 95

FAX

+39 0331 98 53 77

EMAIL

pauloref@mit.edu
paulo.refinetti@gmail.com
paulo.refinetti@epfl.ch

Special Responsibilities

STUDENT COUNCIL MEMBER, HIGH SCHOOL

Member of the student council at the Scuola Europea di Varese
Responsible for communicating and representing the interests of the French speaking students in front of the principles

HOSTING CONFERENCES FEATURING HIGH PROFILE INDUSTRIALS, EPFL

Member of SV-Industry, an association responsible for inviting leading professionals in the field of Life Sciences to share experiences with students

Special Interests

SPORTS

Ski, Golf (hcp 6.5), Water-ski, Kendo, Judo, Diving (Rescue Diver), Dancing

HOBBIES

Cooking, Go (5 kyu), History, Traveling, Wine tasting, Ancient arts, Playing piano

Awards

PRIX SSV DU MEILLEUR PROJET DE MASTERS EN BIOSCIENCES COMPUTATIONELLES

Prize for best master project in computational biosciences. Discerned by the School of Life Sciences, EPFL.

BEST POSTER PRESENTATION AT KYOTO BREAST CANCER CONSENSUS CONFERENCE 2016

Prize for best poster presentation at the Kyoto breast cancer consensus conference, in March 2016. Hosted by The Lancet.



Publications

PEER-REVIEW JOURNALS

Refinetti, P., Morgenthaler, S., & Ekstrom, P. O. (2016). Cycling temperature capillary electrophoresis: A quantitative, fast and inexpensive method to detect mutations in mixed populations of human mitochondrial DNA. *Mitochondrion*, 29, 65–74. <http://doi.org/10.1016/j.mito.2016.04.006>

Arstad, C., Refinetti, P., Warren, D., Giercksky, K.-E., & Ekstrøm, P. O. (2016). Scanning the mitochondrial genome for mutations by cycling temperature capillary electrophoresis. *Mitochondrial DNA. Part a, DNA Mapping, Sequencing, and Analysis*, 1–12. <http://doi.org/10.1080/24701394.2016.1233532>

Kini, L. G., Herrero-Jimenez, P., Kamath, T., Sanghvi, J., Gutierrez, E., Hensle, D., et al. (2013). Mutator/Hypermutable fetal/juvenile metakaryotic stem cells and human colorectal carcinogenesis. *Frontiers in Oncology*, 3, 267. <http://doi.org/10.3389/fonc.2013.00267>

Refinetti, P., Morgenthaler, S., Thilly, W.G. and Ekstrom, P.O., 2014. Mapping cell lineage growth in human tumors using somatic mitochondrial mutations. *International Journal Of Molecular Medicine*, 34, pp.S17-S17.

CONFERENCE PAPERS

Paulo Refinetti, Stephan Morgenthaler, Per O. Ekstrom. 3D mapping of somatic mitochondrial mutations in human tumor. A matter of Life and Death, Mechanisms and Relevance of Cell Death for Cancer Biology and Treatment. EACR conference series 2016

Paulo Refinetti, Stephan Morgenthaler, Per O. Ekstrom . 3D cell lineage tracing from tumor to metastasis . 8th Asian Oncology Summit held alongside the 12th Annual Conference of the OOTR and Kyoto Breast Cancer Consensus Conference 2016.

Paulo Refinetti, Stephan Morgenthaler, Per O. Ekstrom. Cycling temperature capillary electrophoresis: Adopting an orphan method for 3 dimensional lineage tracing in human tumors. Global Summit on Rare Diseases & Orphan Drugs, 2016.

BIRTH

Born 26/10/1988
in Paris, France

NATIONALITY

Italian, Brazilian

CIVIL STATUS

Single, no children

PERMANENT ADDRESS

Via Rossini 2
Carnago (VA), 21040
Italy

PHONE

+41 76 222 07 95

FAX

+39 0331 98 53 77

EMAIL

pauloref@mit.edu
paulo.refinetti@gmail.com
paulo.refinetti@epfl.ch

