Department of Biosciences and Nutrition
Karolinska Institutet, Stockholm, Sweden

SOMATIC MUTATIONS IN HEALTHY CELLS AND AGE-ASSOCIATED DISEASES

Hafdis Thorunn Helgadottir

Karolinska Institutet

Stockholm 2019

Published by Karolinska Institutet.
Printed by E-print AB 2019
Cover Image by Elvira Lundin

ISBN 978-91-7831-570-3

Somatic Mutations in Healthy Cells and Age-Associated Diseases

THESIS FOR DOCTORAL DEGREE (Ph.D.)

By

Hafdis Thorunn Helgadottir

Public Defence **Friday**, **September 27**, **2019**, **9:30**
Karolinska Institutet, **Gene** (5108), **Neo floor 5**, Campus Flemingsberg

Principal Supervisor:
Professor Maria Eriksson
Karolinska Institutet
Department of Biosciences and Nutrition

Co-supervisors:
Professor Magnus Nordenskjöld
Karolinska Institutet
Department of Molecular Medicine and Surgery

Professor Karin Dahlman-Wright
Karolinska Institutet
Department of Biosciences and Nutrition

Dr Irene Franco
Karolinska Institutet
Department of Biosciences and Nutrition

Opponent:
Docent Lars Forsberg
Uppsala University
Department of Immunology, Genetics and Pathology

Examination Board:
Docent Jaakko Patrakka
Karolinska Institutet
Department of Laboratory Medicine

Professor Thomas Nyström
University of Gothenburg
Department of Microbiology and Immunology

Professor Mats Nilsson
Stockholm University
Department of Biochemistry and Biophysics

To my family

ABSTRACT

Aging is a complex process that affects all living organisms. As we age, the biological functions are affected, resulting in a decline of the tissue and possibly age-related diseases. Several environmental and genetic factors have been suggested to contribute to aging. Among these factors, a progressive loss of genome integrity, caused by the occurrence of somatic mutations, is proposed as a cause of deterioration of cellular functions. The aim of this thesis was to analyze the effect of somatic mutations in healthy cells and analyze the contribution of somatic mutations to age-related diseases.

In **paper I**, we showed that satellite cells, stem cells of the skeletal muscle, accumulate 13 somatic mutations per genome per year during adult life. Although genes expressed in the skeletal muscle were protected from mutations by the DNA repair machinery, we observed that this protection was less efficient at increased age, resulting in higher mutation load in the exons of old compared to young satellite cells. A somatic mutation identified in a satellite cell was also detected in a small percentage of the cells of the muscle biopsy, suggesting that somatic mutations propagate from satellite cells to the differentiated muscle during adult age and might contribute to its age-related decline.

In **paper II**, we created a genetic atlas of somatic mutations in healthy cells from different tissues based on newly generated and publicly-available sequencing data. In contrast to the current view of a tissue-specific mutational profile, several cell types showed the same mutational profile despite coming from different tissues. Furthermore, two distinct cell types from the same tissue showed different mutational profiles and rates of mutation accumulation. Thanks to these data, we identified multiple factors influencing mutagen exposure and consequent mutational profiles. These factors include the cell´s localization within the tissue, the degree of differentiation and the presence of a protective stem cell niche. In addition, we identified an epithelial cell of the kidney that shows a unique distribution of mutations, characterized by mutation enrichment in highly transcribed genes. This pattern increases the chances of mutating a cancer-driver gene and is in agreement with an increased predisposition to cancer in this cell type. Finally, our analyses provide evidence of a decline of DNA-repair with aging.

In **paper III**, we identified somatic mutations in the brain of Alzheimer´s disease (AD) patients. Using ultra-deep sequencing and tailored bioinformatics analysis, we could detect low-frequency variants in bulk tissue. In total, 2.86 Mb of candidate genes and AD-linked genomic regions were included in the study, and 11 somatic single nucleotide variants (SNVs) were identified in AD brains, but none in non-AD brains. One variant was validated and predicted to affect transcription factor binding sites upstream of the *CD55* gene, possibly contributing to AD through the regulation of the complement system.

In **paper IV**, we showed that patients with end-stage chronic kidney disease (CKD) express progerin within their arterial media, the same mutated form of the protein lamin A found in premature aging patients. Importantly, we could identify the mutation that causes progeria, the *LMNA* c.1824C>T, in DNA extracted from the arteries. In total, we could identify the progerin protein or the mutation in 34 of the 40 CKD patients. DNA damage and increased proliferation were detected in the CKD patients, indicating extensive vascular regeneration. Our result suggests that progenitor cells carrying *LMNA* c.1824C>T contribute to the vascular pathology and thereby to the disease progression observed in CKD patients.

In conclusion, the work presented in this thesis provides a new understanding of the contribution of mutation accumulation in healthy cells with possible implications for aging and age-associated diseases.

LIST OF SCIENTIFIC PAPERS

I. **Somatic mutagenesis in satellite cells associates with human skeletal muscle aging.**
Irene Franco, Anna Johansson, Karl Olsson, Peter Vrtačnik, Pär Lundin, Hafdis T. Helgadottir, Malin Larsson, Gwladys Revêchon, Carla Bosia, Andrea Pagnani, Paolo Provero, Thomas Gustafsson, Helen Fischer, Maria Eriksson. *Nature Communications.* 2018 Feb 23;9(1):800. PMID: 29476074.

II. **Basal and mutagen-driven somatic mutagenesis shapes the genome of healthy human cells**.
Irene Franco[*], Hafdis T. Helgadottir[*], Aldo Moggio, Malin Larsson, Peter Vrtačnik, Anna Johansson, Nina Norgren, Pär Lundin, David Mas-Ponte, Johan Nordström, Torbjörn Lundgren, Peter Stenvinkel, Lars Wennberg, Fran Supek, Maria Eriksson
[*]Equal contribution
In revision

III. **Somatic mutation that affects transcription factor binding upstream of *CD55* in the temporal cortex of a late-onset Alzheimer disease patient.**
Hafdis T. Helgadottir, Pär Lundin, Emelie Wallén Arzt, Anna-Karin Lindström, Caroline Graff, Maria Eriksson. *Human Molecular Genetics.* 2019 Aug 15;28(16):2675-2685. PMID: 31216356.

IV. **The Hutchinson-Gilford progeria syndrome mutation, *LMNA* c.1824C>T, is a somatic mutation in patients with chronic kidney disease**.
Hafdis T. Helgadottir[*], Nikenza Viceconte[*], Anna Witasp, Agustin Sola Carvajal, Gwladys Revêchon, Daniel Whisenant, Ece Somuncular, Anne-Sofie Johansson, Emelie Wallén Arzt, Anders Thorell, Anne Babler, Susanne Ziegler, Dagmara McGuinness, Sidinh Luc, Rafael Kramann, Paul G. Shiels, Annika Wernerson, Peter Stenvinkel, Maria Eriksson.
[*]Equal contribution
Manuscript

ADDITIONAL ARTICLES

I. **Accumulation of progerin affects the symmetry of cell division and is associated with impaired Wnt signaling and the mislocalization of nuclear envelope proteins.**
Agustín Sola-Carvajal[*], Gwladys Revêchon[*], Hafdis T. Helgadottir, Daniel Whisenant, Robin Hagblom, Julia Döhla, Pekka Katajisto, David Brodin, Fredrik Fagerström-Billai, Nikenza Viceconte, Maria Eriksson. *The Journal of Investigative Dermatology*. 2019 May 23. pii: S0022-202X(19)31566-0. PMID: 31128203.
[*]Equal contribution

CONTENTS

LIST OF ABBREVIATIONS

AD	Alzheimer´s disease
bp	basepair
cDNA	Complementary DNA
CEP	Clonally expanded progenitors
CKD	Chronic kidney disease
CNV	Copy number variant
COSMIC	Catalogue Of Somatic Mutations In Cancer
ddPCR	Droplet digital PCR
EP	Epidermis
FACS	Fluorescence-activated cell sorter
HGPS	Hutchinson–Gilford progeria syndrome
HSCs	Hematopoietic stem cells
HSPCs	Hematopoietic stem and progenitor cells
Indel	Insert/deletion
KT	Kidney tubule
LCM	Laser capture microdissection
MB	Megabase
MDA	Multiple displacement amplification
MMR	Mismatch repair
PCR	Polymerase chain reaction
PFC	Prefrontal cortex
ROS	Reactive oxygen species
RT	Replication timing
SAT	Subcutaneous adipose tissue
SBS	Single-base-substitution
SC	Satellite cell
scDNA	Single-cell DNA
scRNA	Single-cell RNA
SCMDA	Single-cell multiple displacement amplification
SkinFB	Skin fibroblast
SkM	Skeletal muscle
SNV	Single-nucleotide variant
UDS	Ultra-deep targeted sequencing
UV	Ultraviolet
VAT	Visceral adipose tissue
VSMCs	Vascular smooth muscle cells
WGS	Whole-genome sequencing
WT	Wild-type

1 INTRODUCTION

1.1 AGING

During the last century the life expectancy has dramatically increased. At the beginning of 19th century life expectancy was around 50 years, but it is now over 80 years in western societies (1, 2). Improved life qualities, where people have access to better health care and food has led to increased life expectancy and a higher proportion of older individuals. Today, 12% of the world population is at least 60 years of age, and this percentage is expected to rise to 22% by the year 2050 (2).

Aging is a complex process that affects most biological functions and results in tissue decline. As the individual gets older, the likelihood of age-related diseases increases. The incidences of several diseases, such as cancer and cardiovascular diseases, dramatically increase after the age of 60 (3). Although twin studies have shown that heritability of human aging is 20-30% (4), few genetic factors contributing to aging are known. Genome-wide association studies on centenarians, longevity and healthy aging individuals have revealed that aging is a complex trait with several candidate genes but few well-replicated genes (5-7). The *SIRT1*, *LMNA* and *CDKN2A/B* are among the genes that have been linked with longevity and aging (8-10), but the strongest association has been observed in the *APOE* and the *FOXO3A* genes (6, 9, 11, 12). It has been suggested that healthy aging is due to reduced genetic risk or even protection against age-related diseases (7). Indeed, a variant in chromosome 5q33.3 linked to longevity, is associated with low blood pressure and reduced risk of death from a stroke (6).

Several factors have been suggested to contribute to aging, and summarized in nine hallmarks of aging: genomic instability, telomere attrition, epigenetic alterations, loss of proteostasis, deregulated nutrient sensing, mitochondrial dysfunction, cellular senescence, stem cell exhaustion and altered intercellular communication (13). These nine hallmarks have been divided into three categories: primary, antagonistic and integrative hallmarks (Figure 1). The primary hallmarks (genomic instability, telomere attrition, epigenetic alterations and loss of proteostasis) have adverse effects on the cells and are contributed, among others, by exogenous and endogenous DNA damage and inefficient DNA repair machinery. That leads to antagonistic hallmarks (deregulated nutrient sensing, mitochondrial dysfunction and cellular senescence) as responses to the damage and finally to integrative hallmarks (stem cell exhaustion and altered intercellular communication) as a result of the previous two (13).

One way to study the hallmarks of aging is to analyze premature aging syndromes (also known as progeroid syndromes) like Hutchinson-Gilford Progeria Syndrome (HGPS; OMIM 176670) (14-16), Werner syndrome (OMIM 277700) (17) and xeroderma pigmentosum (OMIM 278700) (18). These disorders show many features of physiological aging and share many of the hallmarks previously linked to aging (genome instability, telomere dysfunction, epigenetic changes, nuclear lamina alterations, cell cycle deregulation, senescence, metabolic defects, inflammation, stem cell exhaustion) (19). Several *in vitro* and *in vivo* models are

available for these disorders (19) that can be used to achieve better understandings of not only the progeroid syndromes but aging as well.

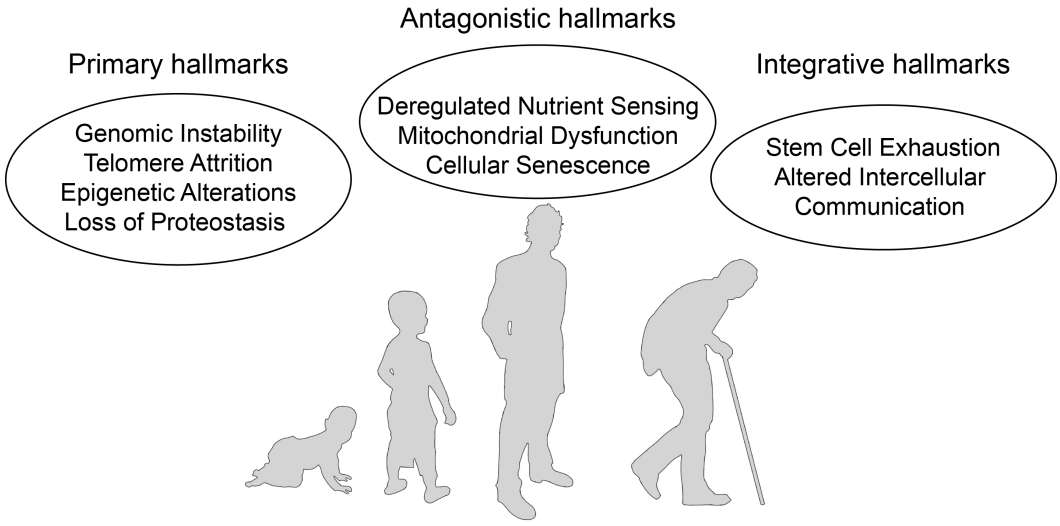

Figure 1: Hallmarks of aging (13). **Primary hallmarks:** Genomic instability, telomere attrition, epigenetic alterations and loss of proteostasis are believed to be the primary cause of damage. **Antagonistic hallmarks:** Deregulated nutrient sensing, mitochondrial dysfunction and cellular senescence are considered as a response to damage. **Integrative hallmarks:** Stem cell exhaustion and altered intercellular communication are the end result of the previous two, resulting in functional decline with aging.

1.2 MUTATIONS IN THE HUMAN GENOME

The human genome can harbor two types of mutations, germline and somatic. Germline mutations are found in the first cell of the embryo and propagate to every cell of an individual. Conversely, somatic mutations are acquired by each cell lineage during development and post-natal life. For this reason, somatic mutations are found only in a portion of the cells of an individual, or even in a single cell (Table 1).

Table 1: Comparison on germline and somatic mutations

Germline mutations	Somatic mutations
Found in every cell in the body	Most often occurring in single cells
Affecting all tissues	Most often only found in part of a tissue
Inherited from parents or acquired during gametogenesis	Acquired during development or anytime during life
Transmitted to offspring	Usually not transmitted to offspring
Effects the evolution	Do not have an effect on the evolution
Mutation rate 1.2×10^{-8} mutations per bp (20)	Mutation rate 2.8×10^{-7} mutations per bp (20)

Somatic mutations finally convert every tissue into a mosaic of cells, each carrying its unique genome (21-34). The frequency of a newly acquired mutation in the tissue depends on when the mutation arises. Mutations arising early during embryogenesis can be detected in substantial percentages of the cells and multiple tissues, while mutations acquired later in the lifespan usually affect fewer cells (Figure 2). It is unclear what effect somatic mutation accumulation has on the tissue and how it contributes to the decline of tissue homeostasis. However, it has been proposed to contribute to aging and diseases.

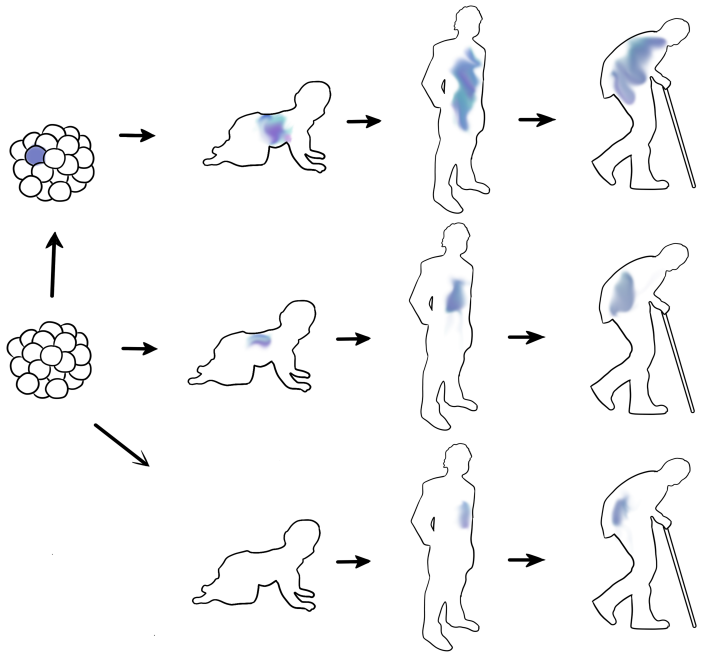

Figure 2: The frequency of the somatic mutations depends on when the mutation arises. Mutations arising early during embryogenesis are usually found in multiple tissues and in a high percentage of the cells composing each tissue. Mutations occurring after birth are usually found in individual cells. Only if mutations are in the germ cells they will be transferred to the offspring.

Different mutation types have a different impact on the cell depending on their size and genomic location. Mutations range from single nucleotide mutations to large genetic alterations, or even to a gain or a loss of a whole chromosome (22). Larger genetic alterations are rarer but affect more bases, while single base substitutions are common but only affect one nucleotide (34). Mutations in the coding regions can have severe consequences for the cell; however, non-coding regions, even if mutated or lost, are unlikely to cause a significant effect on the cell. Nevertheless, it is becoming more apparent that within these non-coding regions are regulatory regions that are important for cell function (35).

Mutations occur continuously and randomly hit the genome during our lifetime. DNA damage can be due to exogenous or endogenous mutagens. Examples of exogenous mutagens are ultraviolet (UV) radiation and various chemicals (36), while examples of endogenous mutagens are free radicals like reactive oxygen species (ROS), or spontaneous deamination of the DNA bases (37, 38). Reactive oxygen species are a result of our metabolism and can

lead to over 25 different oxidative DNA lesions where 8-Oxoguanine (8-oxoG) is the best-studied, leading to G:C>T:A mutations (39). Spontaneous deamination on cytosine and 5-methylcytosine in CpG sites are quite frequent and results in C>T mutations (38).

To counteract this continuous DNA lesions in the genome, we have an efficient repair mechanism (37). However, DNA repair does not have the same efficiency in all parts of the genome (39-41). During cell replication, the most vital genomic regions are replicated first and the least essential regions last (42). The mutation rate is increased in the later replicated regions (43, 44) indicating a selective repair mechanism. That is due to differential mismatch repair mechanisms (MMR) (40, 41) and transcription-coupled repair (39) that protect exons and regions necessary for the cell existence and function. During aging, the repair mechanism declines (45) contributing to increased mutational burden observed in older individuals, and to aging and age-related diseases as cancer (37, 39, 46).

Somatic mutations have been suggested to contribute to the regeneration of injured tissue (47). Studies on the liver and the esophagus have reported frequently mutated genes that promote proliferation without leading to tumor formation (24, 47). Furthermore, somatic mutations can rescue the cell from pathogenic mutations in a mechanism called revertant mosaicism (22) or somatic genetic rescue (48). This process involves mutations or recombination, resulting in a regain of the wild-type genotype or diminished effect of the pathogenic mutation. The acquired mutation can give the cell an increased fitness resulting in clonal expansion of that cell and generate somatic mosaicism within the tissue (47, 48).

Germline variants can contribute to somatic mutagenesis. Cancer provides the best-known examples where germline variants lead to increased mutational burden and specific somatic mutation patterns, as can be seen for variants in the *APOBEC* gene family or the *BRACA* genes (49, 50). Also, patients with xeroderma pigmentosum, caused by variants in DNA repair genes, have up to a 2.5 fold increased somatic mutagenesis compared to age-matched controls (27). Germline variants can indirectly lead to increased mutagenesis. Individuals carrying a nicotine-addiction variant have an increased risk of smoking (51), which creates a toxic environment for the cells, increased mutational burden, and tumorigenesis.

Although somatic mutations are not transmitted to the offspring, pathogenic mutations have been identified at a lower frequency in parents of affected offspring (52-56). For example, germline mutations in children with osteogenesis imperfecta and HGPS have been found as low-level somatic mutations in parents. That can result in a milder phenotype in the parent and recurrence of the disease among siblings (55-57).

1.3 METHODS TO IDENTIFY SOMATIC MUTATIONS

Identifying somatic mutations in the genome is more challenging than to identify germline mutations. Germline mutations are present in every cell and can be assessed in any bulk tissue. Most studies on germline mutations are conducted on blood DNA due to easy access. However, when identifying tissue-specific somatic mutations, tissue accessibility is a limiting factor, as well as the frequency of the mutations in the tissue.

The specificity and sensitivity of the mutation detection depend on the experimental method used and the bioinformatics analysis. The main experimental methods to identify somatic mutations are: i) *in vitro* clonal expansion of progenitor cells, ii) single-cell DNA (scDNA) and single-cell RNA (scRNA) sequencing and iii) bulk tissue analysis using ultra-deep targeted sequencing, exome sequencing, whole-genome sequencing or droplet digital PCR (ddPCR) (Figure 3). All these methods have pros and cons. The first two methods (clonal expansion and single-cell sequencing) give the possibility to identify mutations unique for every cell and to compare the mutagenesis between different cell types. Conversely, bulk tissue analyses only allow the detection of mutations that are found in a high percentage of cells in the tissue (>5% allele fraction in standard exome sequencing (58)). Additionally, bulk tissue often consists of several cell types. Therefore, it is not possible to identify what cell type is carrying the mutations detected, while specific cell types can be selected using the first two methods. The first method relies on progenitor or progenitor-like cells that need to be clonally expanded in culture, while post-mitotic cells can be analyzed with the other two methods.

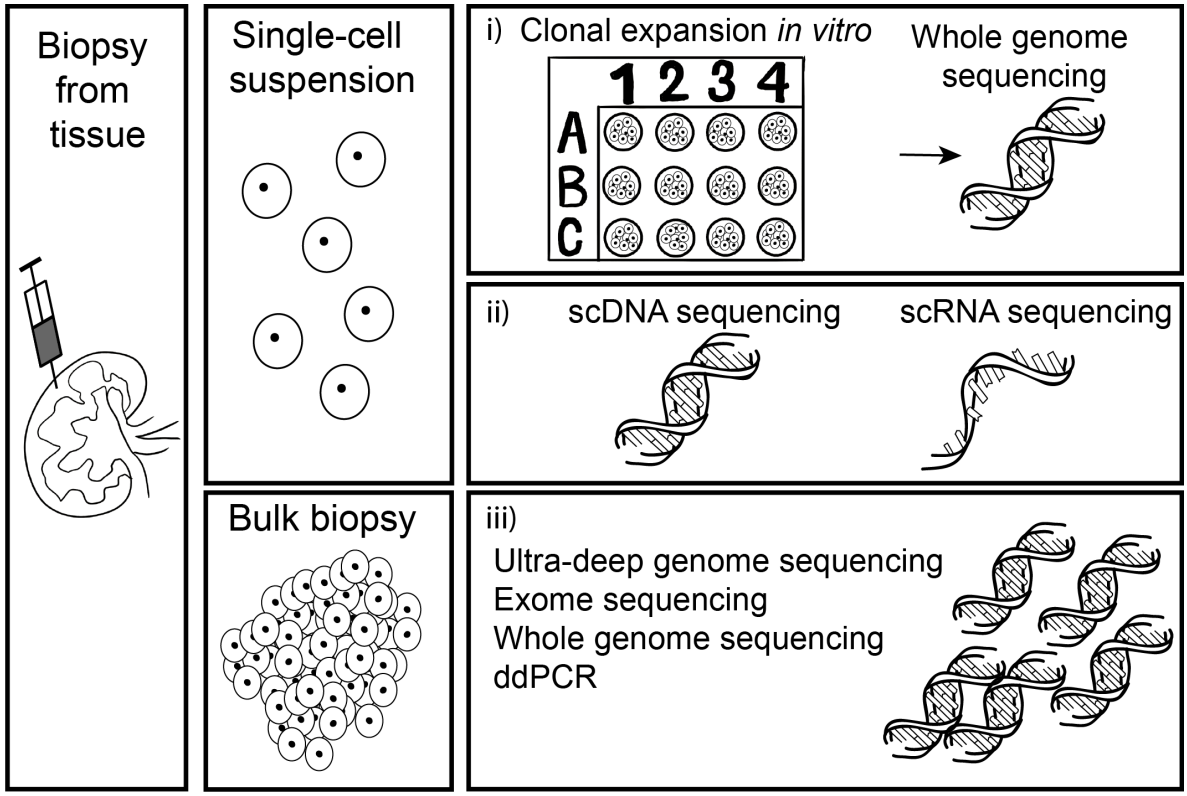

Figure 3: Three different methods to identify somatic mutations in tissues. Single-cell suspensions obtained from a tissue biopsy can either be i) clonally expanded in culture prior to whole-genome sequencing, or ii) single-cell DNA (scDNA) or RNA (scRNA) sequenced. iii) DNA can be isolated from a bulk biopsy for ultra-deep genome sequencing, exome sequencing, whole-genome sequencing or droplet digital PCR (ddPCR).

Identifying somatic mutations in single cells reveals the mutation load and pattern in different cell types and the relationship between cell lineages within a tissue. Several studies have been reported on cell-specific somatic mutations identified with clonal expansion of

progenitor cells (21, 30, 32, 59) or with scDNA (27, 28, 33, 60) or scRNA (61, 62) sequencing. These methods are based on the tissue being disrupted to obtain a single-cell suspension. That can be challenging depending on the tissue. Obtaining single-cell suspension from hard tissues like bones is more difficult than from soft tissues like the brain. The tissue sources are limited, and long-time storage can affect the quality of the DNA or RNA. Furthermore, cell types need to be known prior to sequencing. With RNA sequencing, cell types can be identified based on the expression values, but DNA sequencing does not provide such information about cell types. Therefore, cells need to be fluorescence-activated cell sorted (FACS) or tested in other ways before sequencing.

Since the clonally expanded progenitor cells are the clones of the same original cell, all variants that occurred prior to the *in* vitro cell culture are expected to have an allele frequency of 50%. The cells are expanded in culture to obtain enough DNA to get good genomic coverage and complexity. However, there is an increased risk of culture-induced variants that can affect the results. One approach to avoid bias from culture-induced variants is to analyze the distribution of allele frequency in the sample and set a threshold to avoid false variants. As well it needs to be kept in mind that growing cells in culture can lead to selection bias of cells carrying beneficial mutations. On the other hand, scDNA sequencing cells have no culture-induced mutations but need to be whole genome amplified. That can result in a high error rate (23, 34), among others, due to the limited DNA input leading to allelic imbalance, allelic dropout or false-positive calls (63).

Another approach is to simultaneously do gene expression analysis and mutation analysis in scRNA sequencing using SMART-Seq2 (61) or 10X Chromium (62). This approach can only detect variants in expressed genes, which is a small portion of the genome. In addition, these methods are less advanced than those based on DNA sequencing. ScRNA sequencing has high error rates due to reverse transcription of RNA to cDNA, and cDNA amplification prior to sequencing (61, 64). Furthermore, gene expression values can vary depending on the condition the cells are in, or between different experiments (64) and that affects the identification of somatic mutations. It is harder to identify somatic mutations in low-expressed genes since they have less sequencing depth than the high-expressed genes. Furthermore, bioinformatics analysis needs to be adjusted to identify somatic mutations in a large set of cells, and to link that to the transcriptomic data (62, 64).

When identifying rare somatic mutations in bulk DNA, specific questions or analyses of specific cell types are more beneficial. Such as to search for parental mosaicism of siblings with the same rare disease (52-56) or to search for mutations in candidate genes linked with the Alzheimer´s disease (AD; OMIM 104300) (65-68). Then, ddPCR (69) is valuable since it is a sensitive method to identify rare mutations (<1% allele frequency) in a bulk tissue (70). Another option is to use laser capture microdissection (LCM) to select cells of interest to enrich for particular cell types (29, 31, 71). The cells of interest are isolated from the tissue and genomic DNA extracted from the compiled cells to obtain enough for deep sequencing. With this method, colorectal crypts (29) and neurons in the brain (71) have been analyzed.

Bulk DNA sequencing generally has low error-rate and high genomic coverage, however, it has limits on how rare mutations can be identified and to obtain genome-wide sequence data of high depth can be costly. An average read depth of 30X coverage is sufficient to detect germline mutations in whole-genome sequencing. However, low-frequency somatic mutations are not easily detected and deeper sequencing is needed (<5% allele fraction in standard exome sequencing at 100X-150X coverage (58)).

Downstream bioinformatic analyses need to be tailored to different methods to be able to identify the somatic mutations. Several somatic variant callers such as MuTect1 (58), MuTect2 (58), Varscan2 (72) and Strelka2 (73) have been developed to identify somatic mutations in the cancer genome. They are more specific and sensitive than the germline mutation callers and use a reference sample to filter out the germline mutations. The mutation identification needs to be strict to avoid false-positive calls and contamination, but not too strict to induce false-negative calls. When identifying low-frequency somatic mutations in a non-cancer tissue, and when the sequencing depth is more than usually used for cancer projects, the default settings of the somatic callers can be too strict and need to be tailored.

1.4 SOMATIC MUTATIONAL PATTERN

The consequence of each mutation depends among others, on the genomic location. Large parts of the genome are non-coding regions where mutations are less likely to have an effect, and like previously mentioned, essential regions like exons, are protected from mutations (see 1.3). By using available databases like the Protein Atlas (74, 75), it is feasible to position the mutations in consideration to tissue-expressed genes or regulatory regions to explore mutation enrichment or depletion in these regions (76). Furthermore, the effect of each mutation can be analyzed using annotation tools, as VEP (77) and snpEff (78), that estimate its effect on the genome based on the genomic location and the mutation type.

Analyzing mutational signatures of somatic mutations is a new field. However, it is becoming a custom method to understand the biological mechanism behind somatic mutagenesis. Mutagenic compounds leave behind distinct mutational spectra comprised of different fractions of substitution types and sequencing context (26, 79-82). The most common method to analyze the mutation signatures is the 96-category classification where the tri-nucleotide context of the single-nucleotide variant (SNV) is analyzed. The six substitution types (C:G → A:T, C:G → G:C, C:G → T:A, T:A → A:T, T:A → C:G and T:A → G:C) and the first bases upstream and downstream of the mutation are identified (figure 4A). Each substitution type can appear in 16 different sequencing contexts, leading to 96 different tri-nucleotide contexts. For instance, the C>T mutation with adenine (A) at 5' and guanine (G) at 3' end is marked as A[C>T]G.

Mathematical analysis on these mutational spectra in the cancer genomes has revealed several mutational signatures (26, 79-82). An earlier version of the cancer signature catalog constituted of 30 signatures (80) and a recent version has identified 49 signatures for single-

base-substitution (SBS), 11 signatures for double-base-substitutions (DBS) and 17 signatures for small insertions and deletion (ID) (81).

The signatures have a distinct pattern that can be related to a specific mutagen or biological process. Accumulation of C>T mutations, especially CC>TT dinucleotide, are characteristics of UV exposure (36). It can be found in sun-exposed skin and are appointed to signature 7 (Figure 4B, SBS7a). Smoking gives another pattern with an excess of C>A mutations (79) and assigned to signature 4 (Figure 4B, SBS4). Signature 5 is characterized by a relative "flat" landscape (Figure 4B, SBS5) and has along with signature 1 been linked to aging (26).

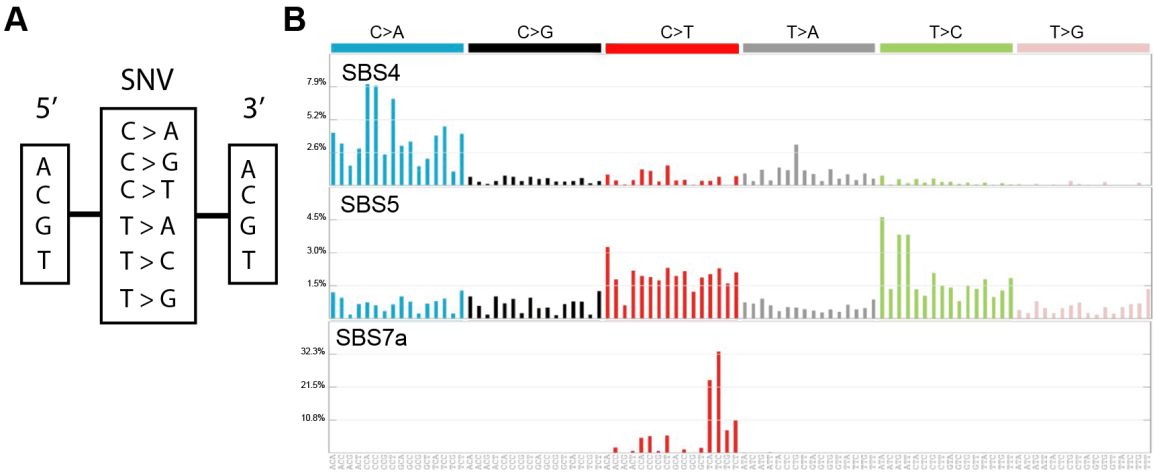

Figure 4: A) Tri-nucleotide sequencing context comprises of the substitution type (here SNV) and the nearest neighboring bases on 5' and 3' site. **B)** Examples of three signatures, SBS4, SBS5 and SBS7a (81). The X-axis mark the 96-categories and Y-axis represents the frequency of each category.

Studies have been conducted to validate and link signatures to specific mutagens, both endogenous and exogenous (39, 83-85). *In vitro* knockout experiments on DNA repair genes have validated the involvement of DNA repair pathways to specific signatures (83, 84) where MMR deficiency has been shown to be associated with signatures 6, 20 and 26 (84) and nucleotide excision repair (NER) to signature 8 (83). Furthermore, a study on environmental agents has confirmed solar radiation to signature 7, and mutagens in tobacco smoke to signature 4 (85). Nevertheless, many signatures have unknown etiology and further studies are needed.

Mutational signature analysis can be achieved and interpreted in different ways. Signatures can be extracted as *de novo* signatures (86), or compared directly for their similarity to the COSMIC signatures (87). The *de novo* approach allows identification of novel mutational processes, which is not possible to do with a direct comparison. However, both methods can be biased from overfitting of signatures and obscure differences between the signatures, especially those that have a "flat" mutational profile (88). Comparing the mutational signatures observed in healthy cells to the ones observed in cancer or knockout cells can provide valuable knowledge regarding somatic mutagenesis.

1.5 SOMATIC MUTATIONS IN HEALTHY TISSUES

Recently, several studies have been published on the accumulation of somatic SNVs in healthy tissues (21, 24, 27, 29-33, 59, 60). The main aims of the studies were to identify the mutation load of cells in different tissues at different ages, as well as to inspect the location and pattern of mutagenesis. These studies have either been based on whole-genome sequencing of clonally expanded progenitors (21, 30, 32, 59), selected tissue regions (24, 29, 31) or scDNA sequencing (27, 33, 60).

Even at birth, acquired somatic mutations can be detected in our cells, and throughout our lives, they continue to accumulate mutations (Figure 5A). Hematopoietic stem and progenitor cell (HSPC) in a newborn has around 40 somatic SNVs while in a 63-year old individual over 800 somatic SNVs are detected (59). Cell in the intestine of a 3-year old individual have accumulated around 250 somatic SNVs, while up to 3500 somatic SNVs are detected in an 87-year old individual (21).

Although most tissues show similar mutation accumulation, there are signs of more rapid accumulation in the sun-exposed skin (32) and colorectal crypts (29) (Figure 5A). The cells are exposed to different environmental and endogenous factors, and UV radiation could explain this mutation burden in the skin sample (32). It could be expected that cells in highly proliferating tissues show a higher accumulation rate than slowly proliferating or post-mitotic cells; indeed, cancer risk correlates with cell division rate (89, 90). However, brain neurons, which are post-mitotic cells, accumulate up to 40 SNVs/year (27), and epithelial cells in the esophagus, colon and intestine, which are highly proliferating cells, accumulate 36 SNVs/year (21, 24) (Figure 4B). Different experimental methods and analysis could explain this difference: the neurons are scDNA sequenced (27) while the colon and intestine cells are clonally expanded *in vitro* (21). Nevertheless, defective DNA damage repair mechanism and oxidative stress could explain this mutation accumulation in the aging brain cells (27).

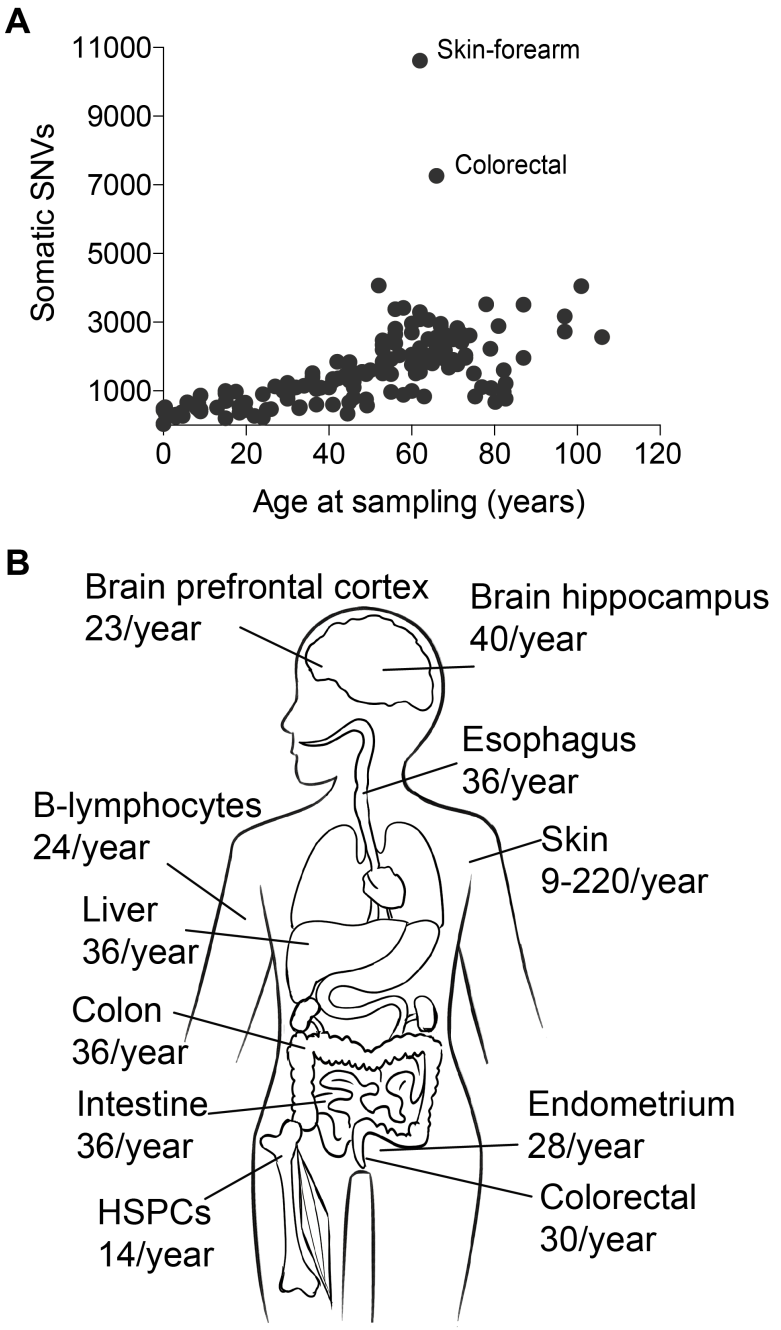

Figure 5: Somatic SNVs in different human tissues, as reported in representative studies and showed in table 2. **A)** The number of somatic SNVs (y-axis) in different tissues and at different ages (x-axis). Cells from the same individual were merged together; each dot in the plot represents the average somatic SNVs per cell per individual. **B)** Average yearly increase of somatic SNVs in different tissues of the human body.

These results need to be taken with some consideration; different experimental methods and bioinformatical analyses were performed in these studies and could affect the result. Colon, intestine and liver (21), HSPCs (hematopoietic progenitor cells)(59), B-lymphocytes (33), skin (32), colorectal (29), brain PFC (prefrontal cortex) and hippocampus (27), esophagus (24), endometrium (31).

Table 2: Overview of studies on somatic mutation accumulation in healthy cells

Tissue	Study design	Age (years)	Samples/ Individuals	SNVs per sample (range)	SNV rate	Reference
Sequencing On Clonally Expanded Progenitors						
Colon	WGS CEP	9-67	21/6	400-3,383	36/year	(21)
Intestine	WGS CEP	3-87	14/9	245-3,516	36/year	(21)
Liver	WGS CEP	30-55	10/5	771-1,919	36/year	(21)
Blood-HSPCs	WGS CEP	0-63	22/7	0-1,014	14/year	(59)
Blood-HSPCs	WGS CEP	59	148/1	815–1,210	na	(30)
Skin fibroblasts	WGS CEP	58-62	10/2	581-12,743	9-220/year	(32)
Sequencing On Selected Tissue/Genome Regions						
Colorectal epithelial	WGS LCM	13-79	571/41	89-14,358	41/year	(29)
Endometrial epithelial	WGS LCM	19-81	182/18	225-2,890	28/year	(31)
Esophagus epithelium	WGS and UDS°	20-75	844/9	270-3,050*#	36/year*	(24)
Single-cell DNA Sequencing						
B lymphocytes	WGS SCMDA	0 - 106	56/14	237-11,765	24/year*	(33)
Fibroblasts	WGS SCMDA	6	6/1	927±371 (s.d)	na	(60)
Brain-prefrontal cortex	WGS MDA	0.4-82	93/15	300-7,000*	23/year	(27)
Brain-hippocampal dentate gyrus	WGS MDA	15-82	26/6	500-5,800*	40/year	(27)

Study design: WGS: Whole-genome sequencing; CEP: clonally expanded progenitors; LCM: laser capture microdissection; SCMDA: single-cell multiple displacement amplification; MDA: multiple displacement amplification – single-cell DNA sequencing; UDS°: Ultra-deep targeted sequencing on 74 cancer genes and SNPs. Age: age range of individuals used in the studies; Samples/Individuals: number of samples and individuals used in the studies. SNVs per sample: range of the SNVs identified in the samples, as reported in the studies, or estimated from figures in the citations (marked *). One study reported only the estimated numbers per individual (marked #). The SNV rate shows SNVs accumulated every year as reported in the studies, or estimated from figures (marked *). HSPCs: Hematopoietic stem and progenitor cells

Analyses on mutational signatures in healthy cells have revealed that signatures 1 and 5 increase with aging, and can be found in most tissues (26). However, they are detected in different proportions in different cell types (Figure 6). Signature 1 has been suggested to be more prominent in cells that have a high proliferation rate, while signature 5 is more noticeable in cells that divide slower (21, 59). Comparisons of cells in different tissues support this conclusion. The intestinal cells (colon and intestine) are fast-dividing cells and

have a higher proportion of signature 1 mutations, while the liver cells that are slow-dividing cells have a higher proportion of signature 5 mutations (21) (Figure 6). HSPCs are believed to be quiescent and have a higher proportion of signature 5 (59) (Figure 5). Furthermore, signature 5 is not observed in infants' brain neurons but appear with increased age in adolescence and adults (27).

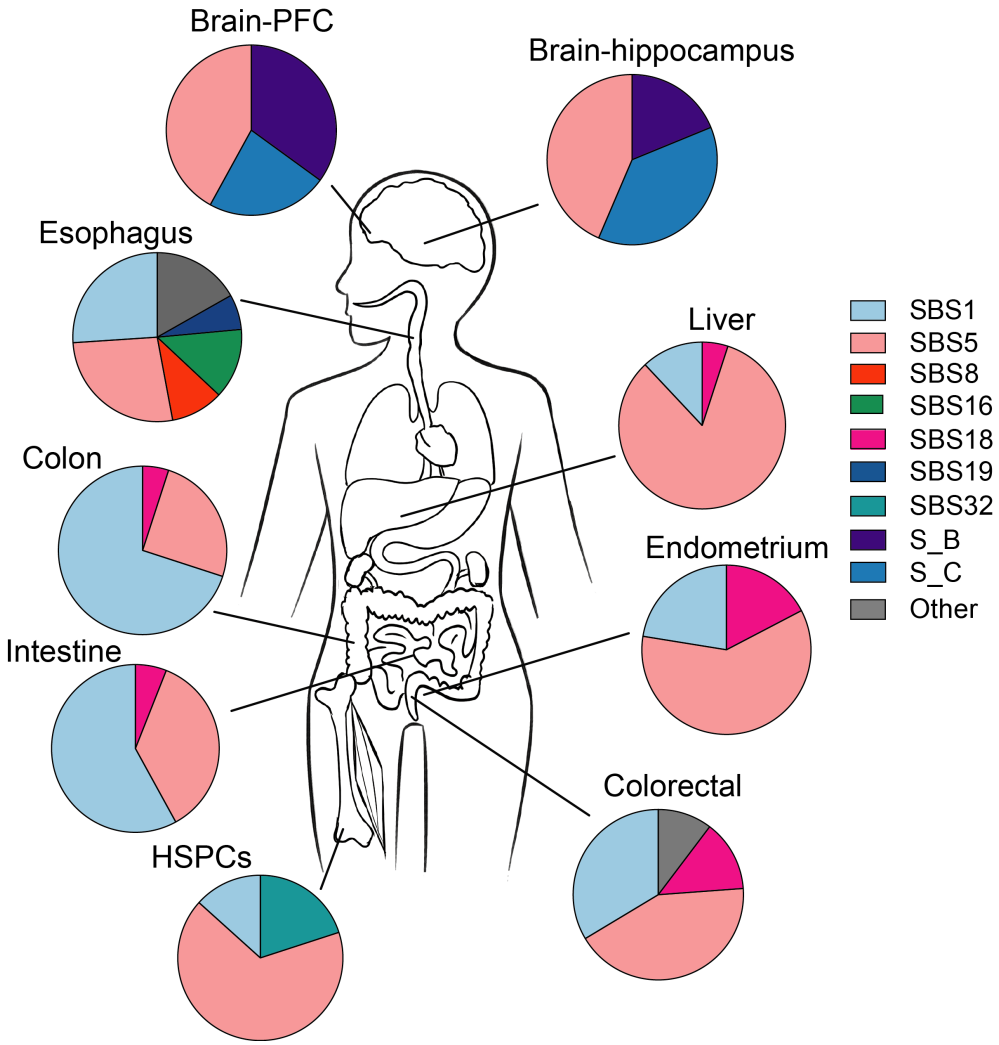

Figure 6: The mutational signatures as reported in representative papers: colon, intestine and liver (21), HSPCs (hematopoietic progenitor cells)(59), colorectal (29), or estimated from figures: Brain PFC (prefrontal cortex) and hippocampus (27), esophagus (24), endometrium (31). Reported signatures are from the COSMIC database or *de novo* signatures (S_B and S_C (27)). These results need to be taken with some consideration; different experimental methods and bioinformatical analyses were performed in these studies and could affect the outcome.

As discussed before, early replicating regions and exons are depleted from mutation accumulation (see 1.3). This is observed in the liver, colon, small intestine and skin (21, 32), although the liver shows less correlation to replication timing than the other tissues (21). Opposite to that are the neurons in the prefrontal cortex (PFC) that show mutation enrichment in exons, especially in genes involved in neural function, suggesting defective DNA damage repair mechanism (27). Furthermore, cancer-associated mutations, or so-called driver

mutations, are observed in healthy cells. The cells of the esophagus (24), colorectal (29) and endometrium (31) all have mutations linked to cancer in their respective tissues.

This different mutational signature and mutation load observed in different tissues show that both extrinsic factors like sun exposure, as well as intrinsic factors such as defective DNA repair mechanisms, affect the cells in a different manner.

1.6 SOMATIC MUTATIONS IN AGE-ASSOCIATED DISEASE

The contribution of somatic mutations to cancer is well known. Mutations that hit an oncogene or a tumor-suppressor gene can give that cell advantages and increased proliferation, leading to tumorigenesis. Many common diseases do not have known genetic causes, especially age-related diseases. That has brought up the hypothesis that somatic mutations might contribute to the diseases.

Suggestive pathogenic somatic mutations have been identified in several tissues leading to various diseases (54). Larger chromosomal aberrations such as a loss or a gain of a whole chromosome, or copy number variants (CNVs) have been linked with diseases (91). Loss of chromosome Y has been linked to a shorter lifetime, smoking, cancer and Alzheimer´s disease (AD) (92-94).

Somatic mutations have been linked to brain malformations (95, 96), and neurological diseases (96). Studies on the brain tissue have shown that it consists of mosaic cells resulting, among others, from transposons (97), CNVs (96) and SNVs (65-68, 98). Although some familial genetic risk factors are known to contribute to AD, most cases are of unknown causes. Therefore it has been suggested that somatic mutations could be the cause of the disease. Identifying somatic mutations in the AD brains is challenging, but recent studies have shed light on brain-specific SNVs in AD brains (Table 3).

While most studies have been focusing on targeted genes (65-68), two exome sequencing studies have reported several brain-specific SNVs (71, 98). In a recent study, both ultra-deep sequencing (584x) and LCM, to enrich neuronal cells, were used to identify in total 595 brain-specific SNVs in AD hippocampus. The mutation rate was calculated to be 22 somatic SNVs per year, where the mutations were contributed by signature 1, 5 and 18, previously linked with aging and oxidative stress (71). In addition, somatic gene recombination in the neurons of individuals with AD, as well as in healthy neurons, has been linked to AD (99). Cleaved protein from the gene *APP* forms plaques in the brain and causes AD. In this study, cDNA of the *APP* was recombined back into the genome, possibly contributing to the progression of AD (99).

Table 3: Overview of studies on somatic mutations in AD brains

Samples	Targets	Brain-specific SNVs in AD brains	Reference
52 AD and 11 non-AD – blood and brain	Deep exome sequencing (584x)	595 brain-specific SNVs	(71)
17 AD and 4 non-AD – blood and brain	Exome sequencing	~575 SNVs/patients	(98)
20 AD brains	Exons of 56 genes linked to common neurodegenerative disorders (genes include *APOE, APP, PSEN1, PSEN2, SORL1, UNC5C*)	15 brain SNVs in 9 AD brains – none in AD candidate genes	(65)
7 AD and 7 non-AD brains	*APP* gene	Somatic gene recombination	(99)
100 AD brains and 355 blood from 445 AD	11 AD candidate genes (*APP, PSEN1, PSEN2, VPS35, MARK4, SORL1, BACE1, NCSTN, PSENEN, APH1A, APH1B*)	Two brain-specific SNVs: one in the *APP* gene, and one in the *SORL1* gene	(66)
72 AD and 58 non-AD brains	4 AD candidate genes, 10 kb upstream and downstream (*APP, PSEN1, PSEN2, MAPT*)	Two SNVs in the *MAPT* gene, and one in the *PSEN2* gene	(67)
1 AD brain	*PSEN1*	Somatic and germline mosaicism of SNV in *PSEN1* exon 12	(68)

Somatic mutations have been suggested to contribute to kidney and cardiovascular diseases (54). They have been detected in patients with polycystic kidney disease (100), and milder cases of Alport syndrome have been linked to mosaicism (54). Mutations in vascular smooth muscle cells (VSMCs) in the arterial wall have been proposed to contribute to atherosclerosis where they are believed to lead to the formation of plaques in the arterial wall (101-103). Clonal expansion has been reported in several tissues leading to patches of cells from the same single cell (24, 104). Similar has been shown in the arterial wall where cells in the medial layer hyper-proliferate as a response to injury (102, 103). Therefore, cells carrying somatic mutations can be clonally expanded in the tissue and contribute to its decline.

2 AIMS

The aim of this thesis was to improve the understanding of genetic events in the development of aging and age-associated disease by studying somatic mutations in human biopsies. This aim was further divided into specific aims:

- To establish a genetic atlas of somatic mutations in various healthy human cells at young and old ages (papers I and II)

- To identify mutational patterns related to specific cell types and better understand mutagenesis in various tissues (paper I and II)

- To develop a bioinformatic strategy to identify somatic mutations in bulk tissue (paper III)

- To identify somatic mutations in the temporal cortex of early and late-onset AD patients and non-AD individuals using ultra-deep sequencing of candidate genes and non-coding regions and tailored bioinformatic analysis (paper III).

- To analyze the expression of progerin in the arteries of CKD patients and the occurrence of the *LMNA* c.1824C>T mutation as the cause of it. Consequently, to analyze the possible contribution of progerin accumulation to the disease pathogenesis in CKD patients (paper IV).

3 METHODOLOGY

3.1 MATERIAL

3.1.1 Human samples (Papers I-IV)

The studies in this thesis were all approved by the regional committee of ethics in Stockholm and adhered to the statutes of the Declaration of Helsinki.

In paper I, blood and biopsy from the leg muscle vastus lateralis were collected from three young (21-24 years) and four old (64-78 years) healthy individuals. From the biopsies, single satellite cells (SC/SkM) were expanded in culture and 29 clones sent to whole genome sequencing along with the blood samples. In addition, four satellite clones were sequenced as a positive control for culture-induced variants after being kept in culture for 50 days.

In paper II, blood and multi-tissue biopsies were collected from three living kidney donors of younger age (30-38 years) and three of older age (63-69 years). From the biopsies, in total 69 single-cell clones from the kidney tubules (KT), subcutaneous fat (SAT), visceral fat (VAT) and the keratinocytes (EP), as well as the blood, were subjected for whole genome sequencing.

In paper III, blood and brain temporal cortex were obtained from four early-onset (59-68 years), four late-onset (79-89 years) AD patients, and eight age-matched (53-88 years) non-AD individuals. In total 32 samples were subjected for ultra-deep targeted sequencing.

In paper IV, arteries were collected from 40 CKD patients (20-69 years) and eight controls (47-68 years) for immunofluorescence/immunohistochemistry (IF/IHC) staining and RNA extraction. Blood was obtained from 26 of the 40 CKD patients and 26 additional controls (21-81 years) for rare mutation detection.

3.1.2 Cells and animal models (Paper IV)

Lymphoblastoid and dermal fibroblast cell lines were obtained through the Coriell Cell Repository, Progeria Research Foundation and the International Werner Syndrome Registry. Tissue samples originated from individuals with progeria and each cell line carried a variant in the *LMNA* gene known to cause the progeria syndrome. The purpose of keeping the cell lines was to use them as a heterozygote standard for testing rare event detection using ddPCR.

Human primary aortic smooth muscle cells (ATCC, PCS-100-012) were used in an in vitro experiment where the effects of uremic serum were analyzed.

The HGPS knock-in mouse model (105) was generated to carry a mutation in the *Lmna* gene (c.1827C>T, p.G609G) leading to progerin, equivalent to the human *LMNA* mutation (c.1824C>T, p.G608G). The mice were used to analyze the effect progerin has on cells *in vivo*.

3.2 LABORATORY METHODS

3.2.1 Somatic mutations in single cells and bulk tissue (Papers I-IV)

Different methods can be performed to identify somatic mutations (see 1.4). Here, three methods were used; *in vitro* clonal expansion of single cells (papers I and II), ultra-deep sequencing of bulk tissue (paper III), and ddPCR of bulk tissue (paper IV).

In papers I and II, tissue biopsies were obtained and immediately digested to obtain single cells. The cells were plated and expanded in culture to obtain enough DNA for whole genome sequencing. Blood was collected from the same individuals and DNA extracted for whole genome sequencing. To validate the cell types, the cells were FACS sorted before culturing (paper I), and a selection of cultured clones was validated with q-PCR (paper I and II).

In papers I and II, the library preparation was performed using 25 ng of DNA starting material and a semiautomatic NeoPrep station using the Illumina TruSeq Nano Kit. The libraries of the blood DNA were prepared with Illumina TruSeq PCR-free library preparations. Sequencing was performed at NGI Sweden, Science for Life Laboratories, Stockholm, on Illumina HiSeq X, PE 2×150 bp.

In paper III, sample libraries were created using 4 μg of DNA (that represent >600,000 cells, assuming 6.6 pg DNA/cell) as a starting material. High amount of DNA was used to obtain the complexity needed to be able to detect rare mutations. The sample libraries were hybridized to an array containing regions linked to AD and consisted of 2.86 MB of the human genome.

The average output from a HiSeq 2500 lane is 143 million reads. The sequencing was paired-end with an insert size of 125 bp and the target region was 2.86 MB. The 32 libraries were run in two pools, each consisting of 16 equal sample libraries. To estimate the average base coverage following formula was used:

$$\frac{143\ M\ reads * 2\ paired\ end * 125\ bp\ insert\ size}{2.86\ Mb\ target\ size * 16\ samples\ in\ the\ pool}$$

In paper III, 32 sample libraries were individually captured on the sequencing array according to the provided protocol from the manufacturer (Roche, USA).

In paper IV, we used ddPCR to detect somatic mutations in arteries and blood of CKD patients and controls. DdPCR is a sensitive method to detect rare alleles in the presence of a common allele in a mosaic tissue sample (69). The solution is partitioned into 20,000 droplets containing the DNA sample and probe assay for the wild-type and mutant allele. The PCR amplification occurs within the droplets, which facilitates the detection of these rare events. The fluorescence light from the probes is then read, and the concentration and fractional abundance of the mutant allele calculated. DdPCR assays for SNVs in the *LMNA* gene and non-*LMNA* genes were designed and run on DNA from the arteries and blood of the CKD

and controls. Furthermore, assays were designed for non-*LMNA* SNVs that have similar effects on their respective protein as the LMNA SNVs have on the lamin A protein.

3.2.2 Validation of somatic mutations (Papers I, III, IV)

The ddPCR technology was used in papers I and III to validate the somatic mutations identified with sequencing, in paper IV to detect somatic mutations and in papers I and IV for absolute quantification.

The assays were designed and ordered from BioRad's web interface for absolute quantification (primer assays in paper I and IV) and rare event detection (primer-probe assays in paper I, III and IV). The procedure of droplet generation was performed according to the manufacturer's recommendations and the raw fluorescence data analyzed and exported from the manufacturer's software (QuantaSoft version 1.6, Bio-Rad). The samples were run in one or more wells, depending on DNA availability. When more than one well was tested for the same sample/assay, the replicated wells were merged.

3.3 BIOINFORMATIC ANALYSIS

3.3.1 Somatic variant identification (Paper I, II, III)

In total, whole-genome sequencing was performed on 40 samples for paper I, and 75 samples in paper II. In addition, ultra-deep sequencing on genomic regions covering in total 2.86 MB was done on 32 samples for paper III. This resulted in a large amount of data to be analyzed. Two pipelines were developed to identify somatic mutations in clonally expanded cells (papers I and II) and bulk tissue (paper III) (Figure 7).

For both pipelines, the raw files were processed in the same way (Figure 7, left panel). The raw files from the sequencer (BCL) were transformed to FASTQ files using the software bcl2fastq (www.illumina.com). The FASTQ files contain the raw unmapped reads as well as the sequence quality score and are in a format that could be further processed. The FASTQ files were aligned to the reference genome using Burrow-Wheeler aligner (BWA, (106)). The mapped BAM reads were checked and marked for PCR duplication, the base quality scores were recalibrated and indels were identified and the reads realigned. These steps are part of GATK Best Practices (107) and were performed to improve the variant calling.

To identify somatic mutations in clonally expanded cells (Figure 7, middle panel; papers I and II), three callers were used: HaplotypeCaller (107), MuTect2 (58) and Fermikit (108). Variants, SNVs and indels, that were identified by all three callers were subjected for further analysis as shown in figure 7. Each clone was compared to matching blood sample to identify somatic mutations in each cell.

To identify somatic variants in bulk sequencing (Figure 7 right panel; paper III), four different somatic callers: MuTect (58), MuTect2 (58), Strelka2 (73), and Varscan2 (72), were used in "somatic mode" with brain as tumor and blood as normal. Problematic regions containing segmental duplication, as well as variants with the comments "clustered events",

"poor mapping region", "nearby gap", "triallelic sites" and "strand artifact" from the MuTect callers were excluded. All variants identified, both germline and somatic, were further analyzed using the custom made somatic filtering showed in figure 7.

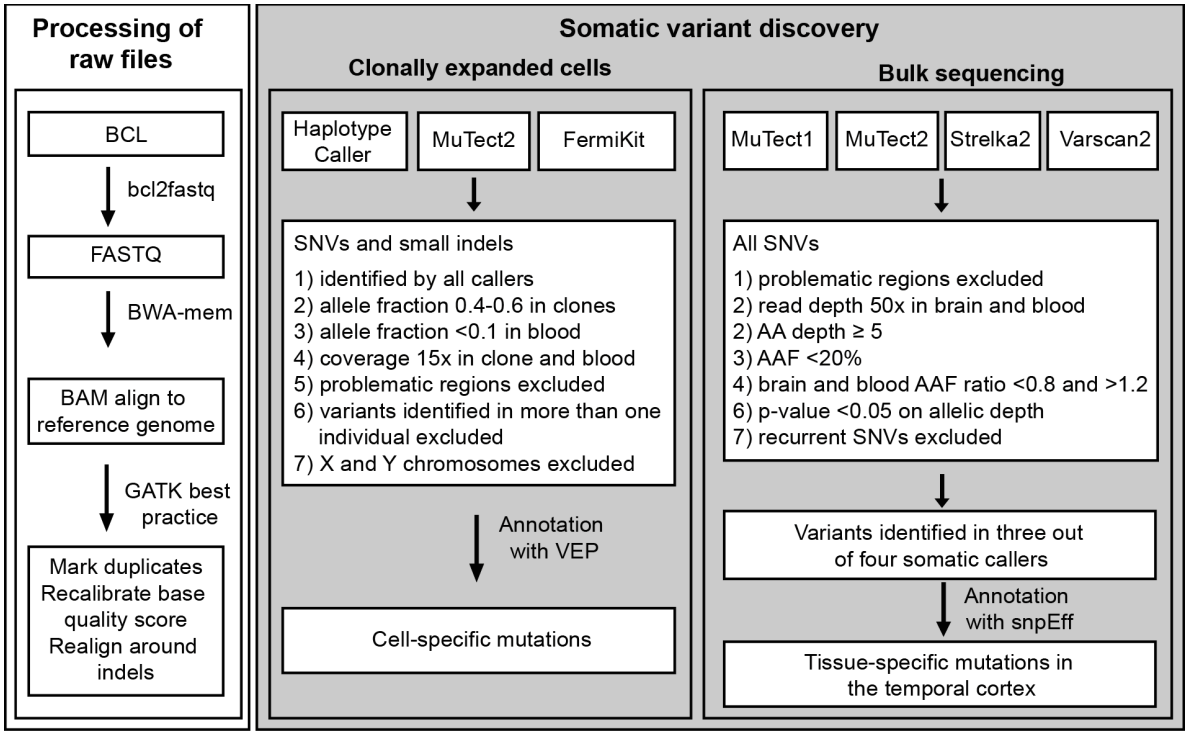

Figure 7: Schematic bioinformatic pipelines used in this thesis for identification of mutations in clonally expanded single-cells and in bulk tissue. "Processing of raw files" (Papers I-III), "Clonally expanded single-cells" (Papers I-II) and "Bulk sequencing" (Paper III). BCL: binary call files; BWA-mem: Burrow-Wheeler aligner; VEP: Variant Effect Predictor; AA: alternative allele; AAF: alternative allele frequency.

3.3.2 Mutational signatures and genomic distribution (paper I, II)

Analysis of mutational signatures was performed using the R packages *MutationalPatterns* (76) and *NMF*. Non-negative matrix factorization (NMF) was used where the cophenetic correlation coefficients determined the number of *de novo* mutational signatures needed. To understand the mechanism behind the *de novo* signatures, they are compared with the COSMIC signatures and the similarities between the mutational profiles calculated (76).

Analysis of the genomic distribution of mutations was performed in two ways:

1) To investigate the mutational enrichment or depletion in exons, introns, regulatory and conserved regions the R-package *MutationalPatterns* (76) was used. Information about tissue-specific genes was obtained from the Human Protein Atlas (http://proteinatlas.com) and conserved regions (PhastConsElements46way data) were downloaded from the UCSC genome browser. In short, coverage files were obtained for every sample, as well as the total number of somatic mutations identified in each sample. First, the probabilities (*Prob*) were calculated based on the total number of mutations (N_{mut}) and total bases covered (B_{cov}).

$$Prob = \frac{N_{mut}}{B_{cov}}$$

Then the expected number of mutations (*Exp*) within the region were calculated where the probabilities were multiplied with total numbers of bases within the region of interest (B_{ROI}):

$$Exp = Prob \times B_{ROI}$$

From the observed (*Obs*) and expected (*Exp*) numbers of mutations, log2(*Obs*/*Exp*) was calculated, where negative numbers represent mutation depletion and positive numbers represent mutation enrichment. Binominal test was used to obtain the *P*-value.

2) To investigate the relationship between the mutations and replication timing/H3K36me3/transcriptional levels/CTCF motif, regression analysis was performed (109, 110). The data on replication timing was obtained from the ENCODE project (RepliSeq) and divided into 6 bins ranging from the latest replicating (bin 0) to the earliest replicating (bin 5). The transcriptional (RNA-seq) levels and H3K36me3 histone marker were collected from Roadmap. Each dataset was divided into 4 bins, ranging from 0, containing non-expressed regions and absent from H3K36me3, up to bin 3 containing high expressed regions and highest abundance of H3K36me3.

3.3.3 *In silico* analysis of mutation effect (Paper I, II, III)

Majority of the mutations detected are located in introns or outside genes. Therefore it can be hard to estimate the effect of the mutations. Different sources are available online where it is possible to determine if the mutation is likely to have a damaging effect, or if it is located within an enhancer or transcription factor binding sites.

In paper I, data from the FANTOM5 project (fantom.gsc.riken.jp/data) was used to analyze regions that are expressed during myoblast to myotube differentiation. In paper I and II, CADD (Combined Annotation Dependent Depletion, (111)) was used to annotate the somatic variants, and Panther GO slim overrepresentation test (pantherdb.org) to identify overrepresented biological processes.

In paper III, candidate enhancers and transcription factors were obtained using GeneHancer (112) PROMO v3.0.2 (113, 114) and JASPAR v5.0_ALPHA (115).

4 ETHICAL CONSIDERATION

The use of human research subjects always requires consideration of ethical issues and is monitored by Ethical Review Boards. In this thesis, studies on human samples were conducted in all four papers. All the participants (or a next of kin acting as a proxy) gave their informed consent and participated voluntarily in the study.

In paper I, the biopsies were obtained exclusively for the purpose of the study. The sampling was performed to be as painless as possible for the individual. In paper II and IV, the biopsies were obtained during already planned surgeries and did not add to more discomfort for the individuals. In paper III, the brain was collected post-mortem.

In paper IV, mice were used as an animal model. The experiments were essential to verify the biological process, previously observed *in vitro*. The Ethical review board approved the animal studies, and all procedures were performed in accordance with the institutional guidelines and regulations. Minimum amounts of animals were used in the experiment, and animal health carefully monitored.

Furthermore, sequencing data from a living individual is considered sensitive personal data that follows strict legislation in Sweden, as with the new EU's General Data Protection Regulation (GDPR). The data has been kept in a protected server with a double authentication, designed to handle sensitive personal data (Bianca, Uppmax, www.uppmax.uu.se).

5 RESULT AND DISCUSSION

The thesis consists of four studies, where somatic mutations in human cells and their contribution to age-related diseases have been analyzed. In this section, the most relevant findings are summarized and discussed.

5.1 PAPER I

Rationale and study design

Somatic mutations accumulate with increased age and are believed to lead to functional decline. Although mutations in healthy cells have been studied (21-34), several questions are still unanswered. The occurrence of somatic mutations and its downstream effect of somatic mutation burden in the skeletal muscle (SkM) or its resident stem cell population (the satellite cells, SCs) were largely unknown. In **paper I**, the aim was to study somatic mutation accumulation in human SCs, analyzing the muscle of young and old healthy individuals.

Muscle biopsies from young (21-24 years, n=3) and old (64-78 years, n=4) healthy individuals were collected and single satellite cells clonally expanded in culture. In total, 29 SCs were subjected to WGS along with the blood sample from the same individuals. In addition, a subset of SCs was kept in culture for 50 days to analyze the effect of long-term culturing and culture-induced mutations. A bioinformatic pipeline was developed to identify somatic mutations in SCs. The distribution of somatic mutations, their age-related accumulation and the somatic mutational patterns were analyzed to understand the effect of somatic mutagenesis on the aging of the skeletal muscle.

With increased age, satellite cells accumulate mutations that propagate into the muscle and might contribute to functional decline

Somatic mutations in SCs increase with aging. An individual in his 20's has around 200 SNVs/SCs and an individual in his 80's has up to 1300 SNVs per SC (Table 4). Through our data, we calculated a mutation rate of 13 SNVs/genome/year in SCs.

The mutations showed a pattern that has been previously linked to aging. Around 25% of the mutations observed in SCs were contributed by signature 1, and around 45% of the mutations by signature 5. The remaining 30% were contributed by signature 8. In old SCs, the fraction of signature 5 was higher than in younger cells, while the fraction of signature 1 and 8 remained the same in young and old. Signature 18 has previously been linked to *in vitro* stress during cell culture and was observed in long-culture SCs.

In agreement with the protection of the functional regions of the genome from mutations (see 1.3), fewer mutations than expected were observed in exons and regulatory regions. Interestingly, this protection was somewhat lost in the older cells, potentially due to less efficient DNA repair machinery during aging. We observed that a higher mutation burden resulted in a slower proliferation of the cells from older muscles and that somatic mutations can propagate into the skeletal muscle. A variant found in one SC was found in 1.3% of the

alleles of genomic DNA and 1.9% of the RNA transcripts of the muscle biopsy from which the SC originated.

In summary, somatic mutations accumulate in the SCs with aging. We observed a loss of genome integrity with increased age where regions essential for SCs were more affected in the older individuals compared to the younger. A somatic mutation identified in one of the SCs was detected in the muscle biopsy (on both DNA and RNA level), indicating that the genomic defects observed can be propagated to the muscle. Hence, the accumulation of somatic mutations in the SCs might contribute to the age-related decline observed in muscle mass and function.

5.2 PAPER II

Rationale and study design

Although inter-individual mutagenesis is well studied, not much is known about the intra-individual genetic variation. For example, cells within the same tissue are believed to be exposed to the same mutagens and consequently show similar somatic mutation patterns. However, a comprehensive analysis of healthy tissues is still lacking. In **paper II**, the aim was to study somatic mutation accumulation in different tissues from the same individual, as well as to analyze mutagenesis among different tissues and ages. This provides information about the tissue-specific mutagenesis while excluding the variability due to different genetic background and environmental exposure.

Multi-tissue biopsies were collected from living kidney donors of younger (30-38 years, n=3) and older (63-69, n=3) age (Figure 8A). From the collected tissues (kidney tubules (KT, n=25), subcutaneous fat (SAT, n=22), visceral fat (VAT, n=20) and epidermis (EP, n=2)) progenitor cells were collected and clonally expanded in culture (Figure 8B). The bioinformatic pipeline created for paper I was used here to identify somatic mutations in different tissues and the mutational patterns analyzed. In addition, publicly available WGS data on healthy cells, generated with a similar strategy as in papers I and II, were used to explore the mutational profile in different tissues. Somatic mutations from tissue-matched cancers were also included in the analysis.

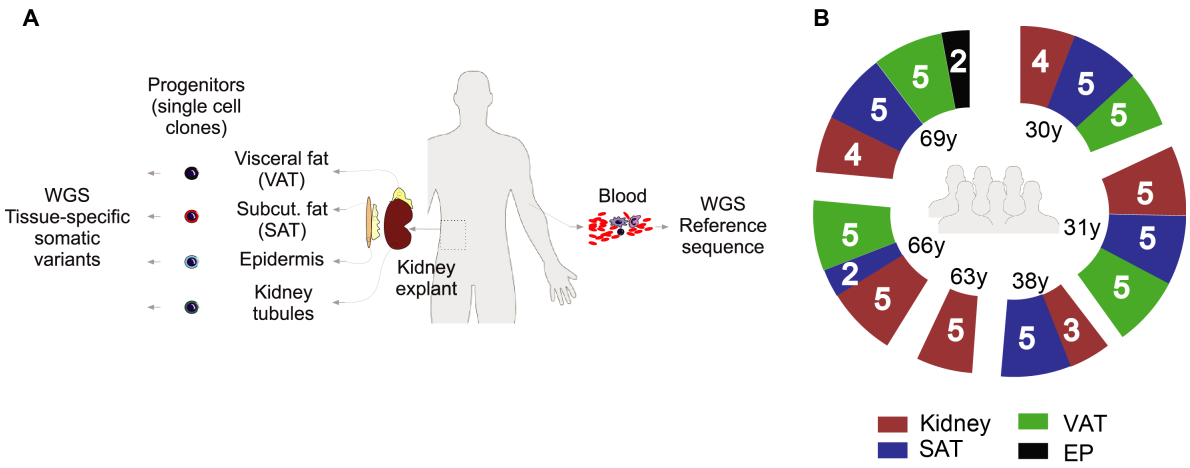

Figure 8: **A**) Experimental strategy used in paper II. Blood, kidney, subcutaneous fat (SAT), visceral fat (VAT) and skin (EP: epidermis) biopsies were obtained from living kidney donors. The blood was whole genome sequenced and used for germline variant detection. Single progenitor cells from kidney tubules, SAT, VAT and EP were clonally expanded in culture and whole-genome sequenced. **B**) Schematic summary of the sequenced samples. Two to five single genomes per biopsy were sequenced from six individuals.

Table 4: Combined results from paper I and paper II:

Tissue	Age (years)	Samples/ Individuals	SNVs per sample (range)	SNV rate	Indels per sample (range)	Reference
Skeletal muscle	21-78	29/7	266-1473	13/year	11-114	Paper I
Kidney tubules 1	30-69	11/6	342-1131	12/year	25-116	Paper II
Kidney tubules 2	31-69	14/5	807-4132	57/year	56-582	Paper II
Epidermis	69	2/1	1510;1639	na	28;32	Paper II
Subcutaneous fat	30-69	22/5	438-1558	17/year	36-109	Paper II
Visceral fat	30-69	20/4	684-2133	27/year	40-127	Paper II

Age: age range of individuals used in the studies; Samples/Individuals: number of samples and individuals used in the studies. SNVs per sample: range of the SNVs identified in the samples, normalized to coverage. The SNV rate shows SNVs accumulated every year. Indels per sample: range of the indels identified in the samples, normalized to coverage. na: not available.

Progenitor cells accumulate somatic mutations at different rate. A specific population of kidney tubule cells shows mutation enrichment in kidney-expressed genes and regulatory regions.

Somatic mutations were detected in all progenitor cells and, as observed in satellite cells (SkM: skeletal muscle) in paper I, increased with aging (Table 4). Although we had cells from four different tissues (KT, SAT, VAT, EP), analysis on the mutational context (i.e., tri-nucleotide context) showed that KT cells were divided into two groups (Figure 9A). The EP cells showed strong UV-induced mutagenesis with many C>T mutations while part of the KT cells, called KT2, showed fewer C>T mutations than SAT, VAT and the remaining KT, called KT1. Furthermore, a yearly increase of mutations was higher in KT2 cells (56.6 SNVs

and 8.0 indels per genome per year) compared to the other cell types (KT1 clones 11.7 SNVs and 1.4 indels; SAT 17.5 SNVs and 0.9 indels; VAT 27.2 SNVs and 1.4 indels) (Figure 9B).

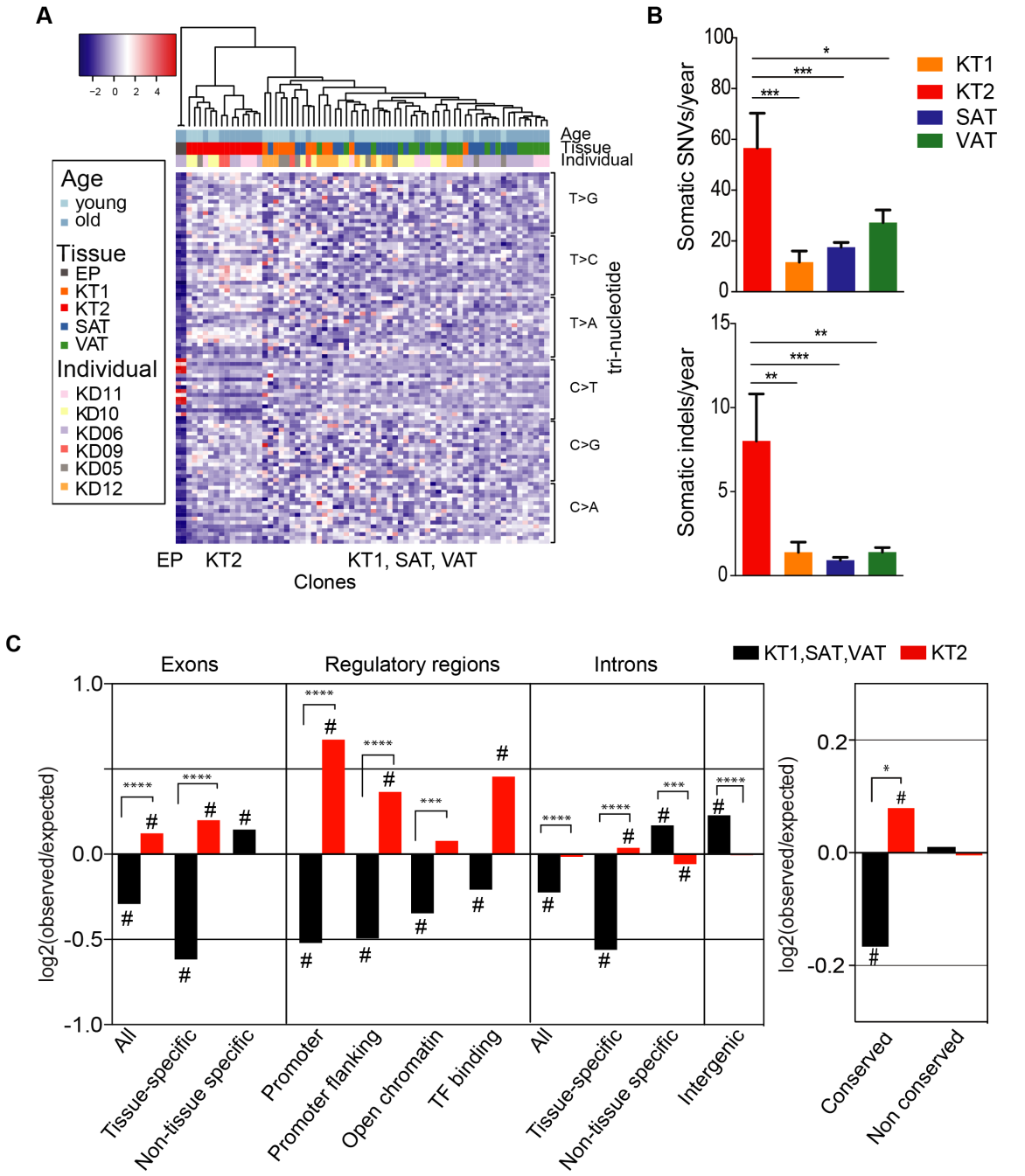

Figure 9: A) Heatmap showing the clustering of the 69 single genomes based on their tri-nucleotide context (vertical) and the relative amount of mutations in each category (horizontal). **B)** Average yearly increase of somatic SNVs and indels per tissue. **C)** Enrichment (upward bars) or depletion (downward bars) of somatic mutations in indicated genomic regions. # $P< 0.05$ one-sided binomial test; * $P<0.05$, ** $P<0.01$, *** $P < 0.001$ one-way ANOVA or two-sided t-test. EP=epidermis; KT1=kidney tubule 1; KT2=kidney tubule 1; SAT=subcutaneous fat; VAT=visceral fat

The mutations identified in KT2 cells were more likely to be pathogenic than those observed in other cell types obtained from the same donor. In the KT2 cells we observed mutational

enrichment in kidney-expressed genes (exons and introns) and regulatory regions, while KT1, SAT, and VAT cells showed mutation depletion in their tissue-expressed genes and regulatory regions. Similarly, we observed that while conserved regions were protected from mutations in KT1, SAT and VAT, they were enriched for mutations in KT2 cells (Figure 9C).

Meta-analysis of somatic mutations in healthy tissue identifies basal and mutagen-induced mutational patterns

To build a comprehensive atlas of somatic mutagenesis of healthy cells in different tissues, the analysis was extended to available datasets obtained from clonally expanded cells. Previous mutational analyses had been performed on stem cells from liver, intestine and colon (21), skin fibroblasts (25), blood progenitors (30) and the SkM satellite cells and long-cultured SkM satellite cells from paper I. These genomes, in total 92, were added to the analysis, as well as WGS data from tissues-matched cancer cells (n=192). Unsupervised clustering on the mutational context showed that KT1, SAT, VAT, SkM and blood clustered together, indicating similar mutational signatures. They were defined as "common progenitors" and grouped for further analyses (Figure 10). In addition, we could observe that cancer genomes often clustered in near proximity of their tissue-matched healthy genome.

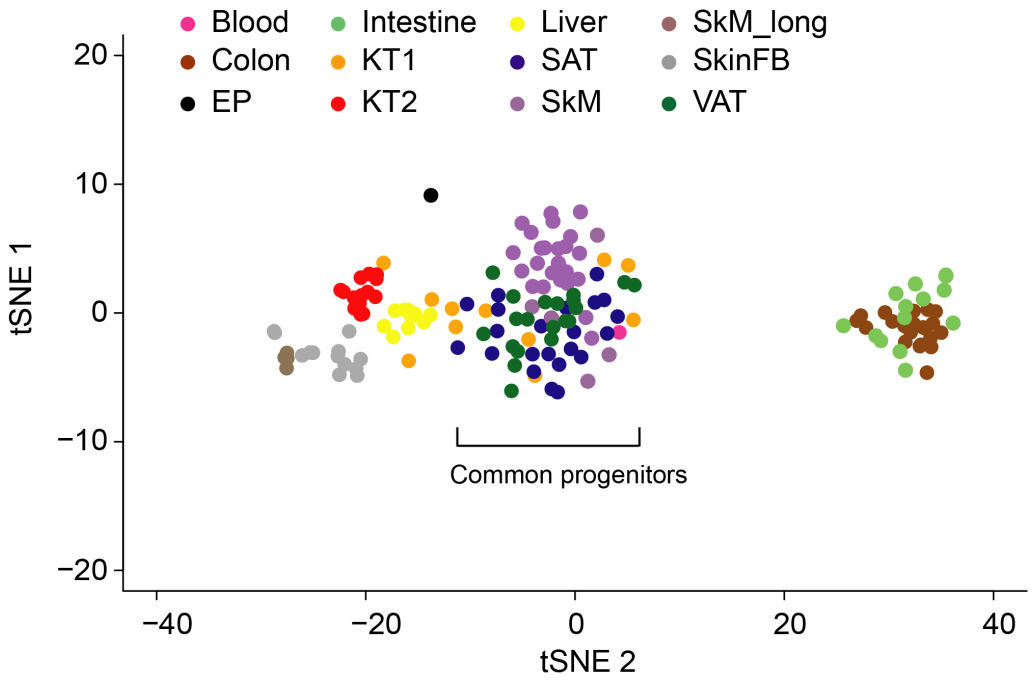

Figure 10: tSNE plot of the tri-nucleotide context of the somatic SNVs observed in the 69 genomes from EP (epidermis), KT1 (kidney tubule 1), KT2 (kidney tubule 2), SAT (subcutaneous fat), VAT (visceral fat), as well as the 92 genomes from skeletal muscle (SkM, paper I), and from public datasets: liver, intestine and colon (21), skin fibroblasts (25), blood progenitors (30).

Mutational signatures were analyzed for 353 WGS genomes included in the meta-analysis. By applying the most recent classification of mutational signatures (81), we could identify eight mutational signatures. Overall, three main signatures (signature 1, 3/8 and 5) were

identified in all cell types (Figure 11) and increased with aging. We defined this signature combination as a "basal mutagenesis" representing the inevitable result of the core cellular process. However, the contribution of these three signatures to the mutation burden of each genome was different depending on the cell type. Signature 1, 3/8 and 5 contributed to the majority of the mutations detected in the common progenitors (blood, SAT, VAT, SkM and KT2). The mutations detected in the intestine and colon samples were mostly signature 1, while the KT2 and liver samples showed a low signature 1 contribution but higher signature 5 (Figure 11). That is in agreement with signature 1 detected in cells with a high proliferation rate, while signature 5 is more prominent in slower dividing cells (21, 59).

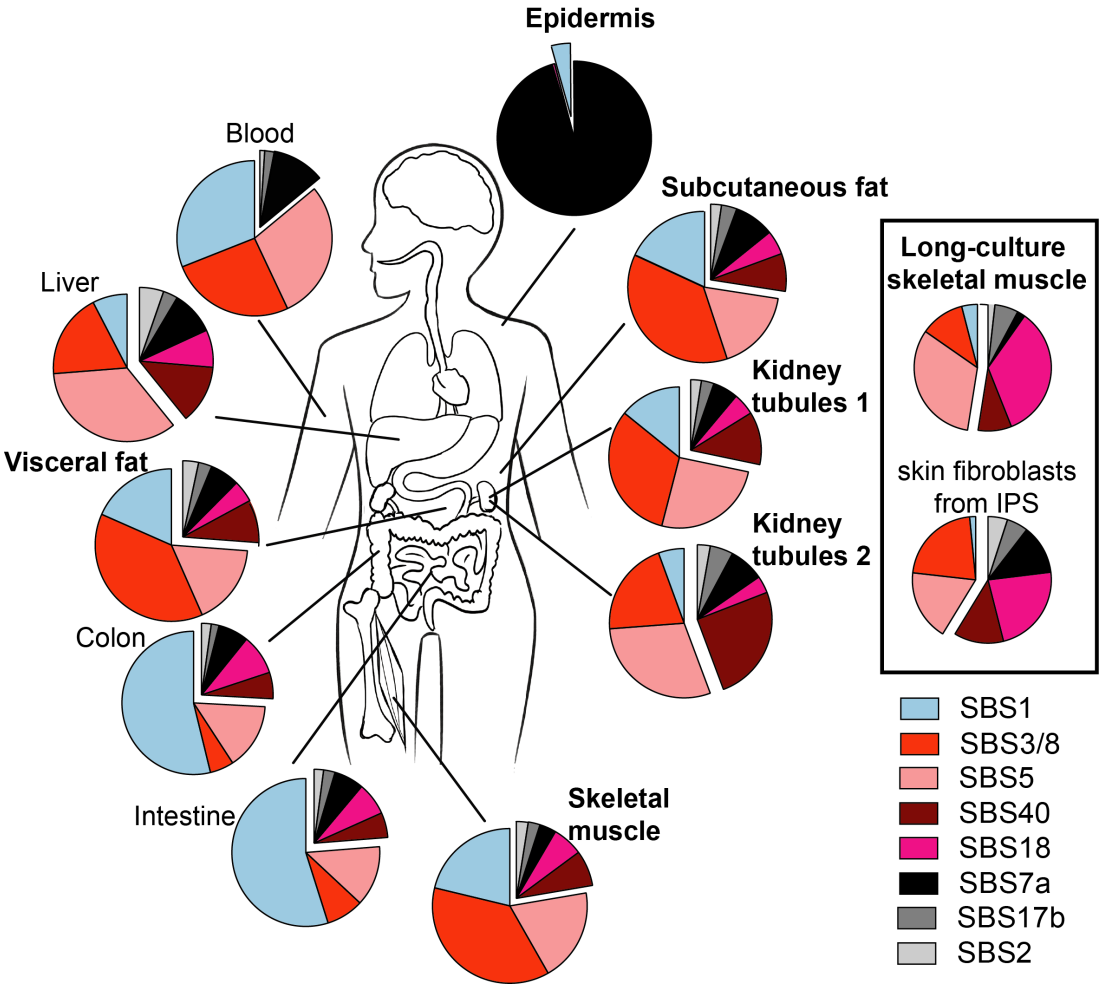

Figure 11: The relative contribution of the eight mutational signatures in healthy cells. The basal mutagenesis, defined as signatures 1, 3/8 and 5, are showed together while specific mutagenic signatures are pulled out from the pie chart. In the box to the right are the long culture skeletal muscle cells as well as the skin fibroblasts, both that were kept long in culture and showed higher proportion of signature 18, linked with *in vitro* culture stress. These results need to be taken with some consideration; the data is taken from different studies where different analyses were conducted to identify somatic mutations. Small shifts in the relative distribution of signatures might occur. Bold marked tissues represent data presented in paper I and II. Other tissues are colon, intestine and liver (21), blood (30) and skin fibroblasts from IPS (25).

Other cells showed specific signature induced by tissue-specific mutagens. In the skin, most mutations were contributed by signature 7a; a strong UV-induced signature and observed in melanoma cancer. Consistently with that melanoma cancer cells clustered in close proximity of the healthy epidermal cells. Skin fibroblast (25) and skeletal muscle cells from paper I that were kept long in culture (SkM_long) showed high contribution of signature 18, previously linked to *in vitro* mutagenesis (21). Signature 40 was mainly observed in KT2 as well as in the two kidney cancers (clear cell and papillary renal cell carcinoma (KIRC, KIRP)).

KT2 cells originate from the proximal tubules

To further define the subsets of KT2 and KT1, markers for different kidney cells were tested using FACS and qPCR. Both KT1 and KT2 expressed the markers of kidney progenitors (CD133, CD24 and PAX2). KT2 clones expressed the markers of proximal tubules, while the KT1 clones were generally negative for all markers of differentiated kidney tubule cells. No kidney clones expressed markers of distal tubules. A marker of tubule damage was observed at different degrees in all KT clones, especially in KT2.

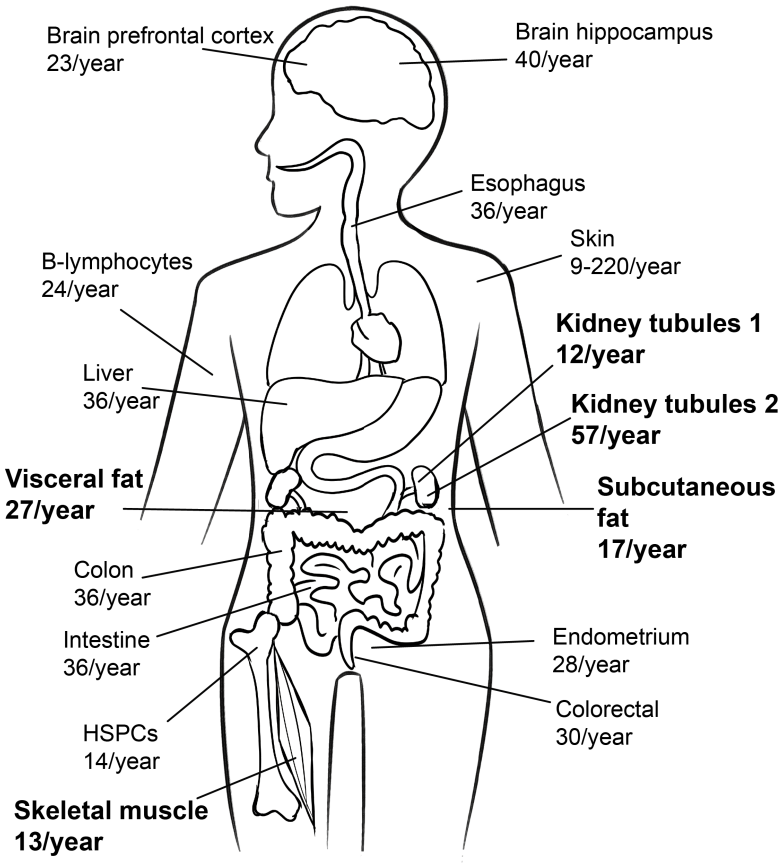

Figure 12: Average yearly increase of somatic SNVs in different tissues of the human body. Bold marked tissues represent data presented in paper I and II, other tissues as reported in representative studies (see figure 4 and table 2 in chapter 1.6). These results need to be taken with some consideration; different experimental methods and bioinformatical analyses were performed in these studies and could affect the outcome. Colon, intestine and liver (21),

HSPCs (hematopoietic progenitor cells)(59), B-lymphocytes (33), skin (32), colorectal (29), brain PFC (prefrontal cortex) and hippocampus (27), esophagus (24), endometrium (31).

Taken together, here we present whole genome sequencing data and high confident somatic variants. All cells were clonally expanded single human cells and provide a reliable comparison of tissue-specific differences. Compared to previously published data (see 1.6), the progenitors of SAT, VAT, SkM and KT1 show a yearly accumulation similar to that observed in most tissues, while the mutation accumulation in KT2 cells is among the highest reported in healthy cells (Figure 12).

Previously, signatures 1 and 5 have been described as the clock-like signatures representing the unavoidable mutational accumulation in cells with aging (26). Here, we expand that concept and define "basal mutagenesis" as a combination of signatures 1, 5 and 3/8 displaying the inevitable mutation accumulation as consequences of core cellular process during aging. In addition, we observed tissue-specific mutagens where signature 7 (high UV-induced mutations) was observed in EP and signature 18 (*in vitro* cell culture stress) in cells kept long in culture (SkM-long and SkinFB).

KT2 cells showed a mutational profile previously linked to clear cell and papillary cell renal cell carcinoma (KIRC and KIRP, respectively). These two kidney cancers originate from the proximal tubules, while the third kidney cancer tested, chromophobe renal cell carcinoma (KICH), originate from the distal tubules and cluster away from KT2 and KT1. Overall, based on the analysis of mutational patterns and expression markers, we conclude that KT2 are cells located in the proximal tubules, while KT1 are uncommitted progenitors of the kidney tubule.

Common progenitors apart from KT1 (e.g. SAT, VAT, SkM and blood) showed mutation depletion in early replicated regions, open chromatin (defined as H2K36me3-rich) and transcribed genes. However, in KT2 cells, this pattern was altered where early replicated regions, open chromatic and transcribed genes were not as protected from mutations as the other cells. Notably, highly transcribed genes were enriched for mutations in KT2 cells and showed an altered pattern of transcriptional-strand asymmetry. Together this indicates that the DNA repair mechanism is active in the common progenitors, while some form of transcription-coupled mutagenic processes (116) was observed in KT2 cells and MMR was not as effective as in common progenitors.

In summary, genome-wide analysis on somatic mutagenesis in human cells provides an understanding of the changes in genome integrity during aging and elucidates the mutagenic process occurring within the cells of our body.

5.3 PAPER III

Rationale and study design

Alzheimer's disease (AD) is a common neurodegenerative disease classified into early-onset and late-onset AD. Familial cases are known with suggestive genetic causes; however, most AD cases are sporadic with unknown causes. Most studies have been done on blood, searching for germline risk factors (117); however, recent studies have shown that somatic mutations accumulate in the brain of AD patients (66-68, 98) although pathogenic consequences of those variants are unknown. In **paper III,** the aim was to identify somatic mutations in the brain of AD and non-AD individuals.

Blood and brain (temporal cortex) samples were obtained from early-onset and late-onset AD patients (n=8) as well as age- and gender-matched non-AD individuals (n=8). Eleven genomic regions containing 28 genes and regulatory regions, in total 2.86 MB, were selected for targeted sequencing. These regions contain AD candidate genes or regions that have been linked to AD by association studies (117). Bioinformatic analysis was performed using four different somatic callers (MuTect1 (58), MuTect2 (58), Strelka2 (73) and Varscan2 (72)) and downstream filtering was applied to identify tissue-specific somatic SNVs in the brain (see 3.4.1).

Somatic mutations were identified in the brain of AD patients

The sequencing libraries were created using 4 μg of genomic DNA, which represent over 600,000 cells (estimating 6.6 pg/cell). This high DNA input resulted in an average coverage of 698X ± 23X (mean ± SEM) across the captured targets. The tailored bioinformatics pipeline created for this study consisted of four callers, where variants were further filtered based on, among others, read depth and alternative allele depth.

In total, 11 brain-specific SNVs were identified in the temporal cortex in the AD brains and none in the controls. One of the SNV was identified by all the four somatic callers, and the remaining 10 were identified by three of the callers. The SNV identified by all callers, chr1:207461994C>T, is located upstream of the *CD55* gene and was detected at 1% allele frequency in the sequencing. It was validated to be present at 0.4% allele frequency in the brain of a late-onset AD patient but not in the blood. It was not detected in the brain of the other individuals. *In silico* analysis predicted the variant to affect transcription factor binding sites upstream of the *CD55* gene. The *CD55* gene is an inhibitor of the complement system, a branch of the immune reaction whose function has been linked to AD (118-120).

Few brain-specific mutations have been found in AD brain samples (65, 66). Identifying somatic mutations in bulk DNA is challenging and limited to parts of the genome (65-68, 98). However, a recent study suggests that using LCM to enrich neurons is a promising method to identify brain-specific mutations (71).

For this study, the default settings of the somatic callers proved to be too strict and therefore a tailored downstream bioinformatics analysis was created. This study emphasizes the need to

create somatic mutation callers that are not designed for cancer analysis but to analyze rare somatic mutations in bulk DNA.

Taken together, somatic mutations occur in the brain and can be detected at low frequency using ultra-deep sequencing and comprehensive bioinformatic downstream analysis. Our results suggest that targeting regulatory regions in addition to the exons could increase the understanding of the molecular basis of both early- and late-onset AD. The confirmed variant might contribute to AD via the regulation of the complement system, however further studies are needed to confirm the significance of the findings for AD disease pathogenesis.

5.4 PAPER IV

Rationale and study design

Children born with the Hutchinsons-Gilford progeria syndrome (HGPS; progeria) show clinical features of premature aging like growth retardation, loss of subcutaneous fat and fragile bones (16). The main cause of progeria is a mutation in the *LMNA* gene, c.1824C>T (14, 15). The mutation creates a cryptic splice site leading to a 150 bp deletion that results in a truncated *LMNA* protein called progerin. The children have severe cardiovascular diseases and die in their teenage years from cardiovascular complications (16).

The vascular phenotype observed in the HGPS children is very similar to that observed in individuals with end-stage CKD. Pathology changes consist of tissue degeneration of arteries (Figure 13A: cross-section of an arterial wall) and include intima thickness, media calcification, loss of VSMCs and adventitial fibrosis (121, 122). These phenotypic similarities could indicate a shared component between these two different conditions. Indeed, progerin expression has been found at low levels in healthy tissues, including arteries (122-124). In **paper IV**, the aim was to analyze the presence of progerin in the arteries of CKD patients and its causes and contribute to the disease.

Arteries were collected from CKD patients (n=40) and controls (n=8), to assess the frequency of progerin positive cells in the tissue. Calcification of the arteries was classified, and cell density measured. The arterial sections were tested for markers of inflammation, cell proliferation, senescence and DNA damage. To assess the somatic mutations in the *LMNA* gene that lead to progerin, DNA from artery sections (CKD n=40, controls n=6) and blood (CKD n=26, controls n=26) were collected.

Progerin and the *LMNA* c.1824C>T variant are observed in the CKD arteries

Increased calcification was observed to correlate with increased age, and reduced cell density, especially in the arterial media of the CKD patients. Progerin positive cells were identified in 28 out of 40 patients with a frequency of 0.2-7.4%. The majority of the cells were scattered in the media, while clusters of positive cells were observed in two of the patients (Figure 13B). The frequency of progerin positive cells correlated with years of disease (Figure 13C), but not with age at sampling. Out of six control arteries, only one showed weak progerin staining.

The presence of progerin was further confirmed with the detection of the progerin transcripts in all CKD tested (n=19) but none of the controls (n=8).

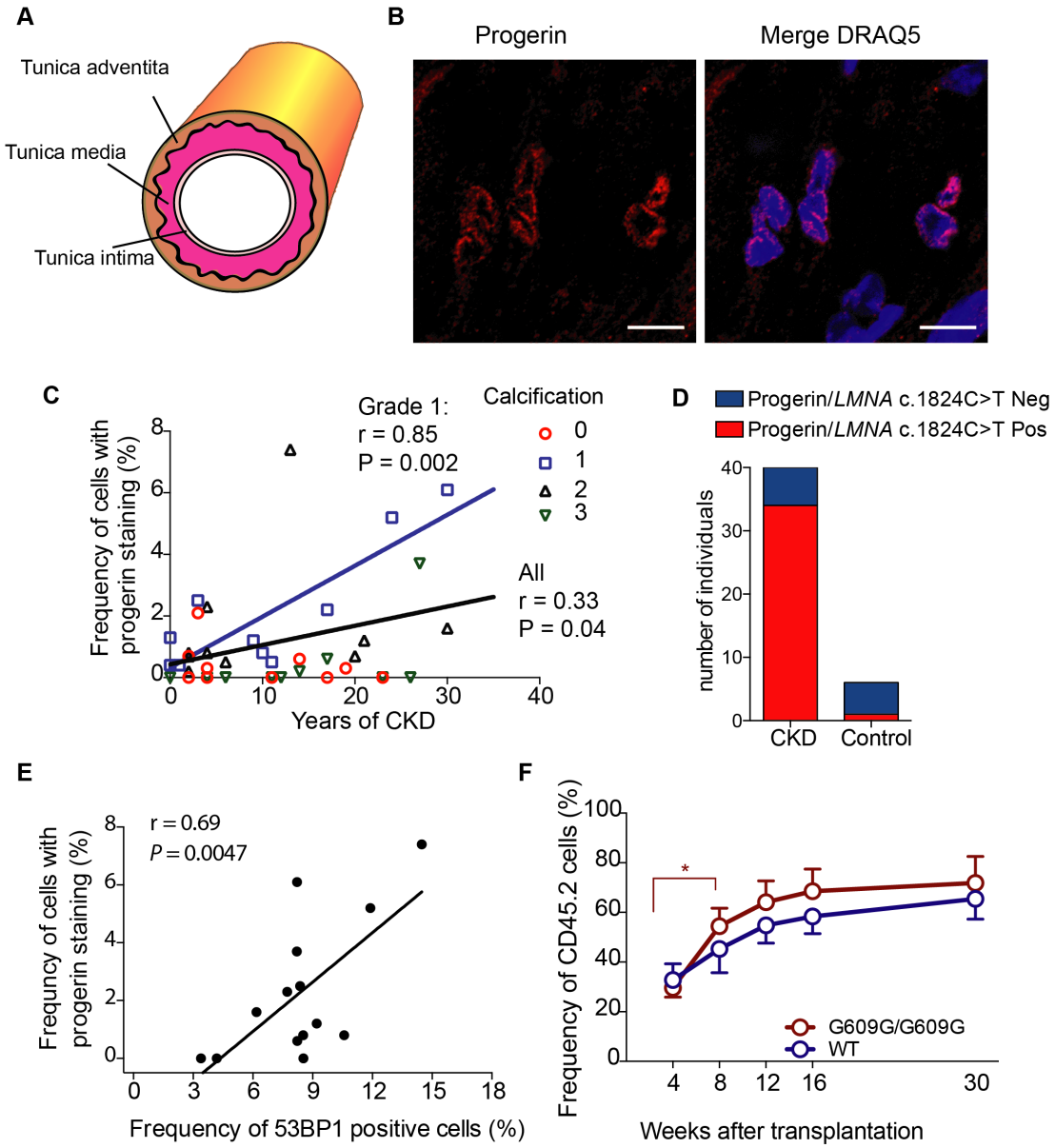

Figure 13: Progerin in CKD arteries and its effect *in vivo*. **A)** Cross-section of the arterial wall. The innermost layer, tunica intima, is a simple epithelial layer, the middle layer, tunica media, contains VSMCs, and the outer layer, tunica adventitia, contains elastic fibers. **B**) Clusters of progerin positive cells in the tunica medium of CKD arteries. Scale bars: 10 μm. **C)** Frequency of progerin positive cells (y-axis) in all CKD patients showed a correlation with the number of years of disease (x-axis), especially in CKD patients with calcification grade 1. **D)** The progerin protein, progerin transcript or the *LMNA* c.1824C>T mutation was detected in 34 out of 40 CKD patients and one out of eight control. **E)** Frequency of DNA damage measured with 53BP1 (x-axis) correlated with progerin positive cells (y-axis) (n=14). **F)** Frequency of CD45.2 population from *Lmna*G609G/G609G (red, n=5) and wild-type (blue, n=4) mice in recipient wild-type mice at 4, 8, 12, 16 and 30 weeks after transplantation. Data are shown as mean and SEM.

To identify the underlying molecular mechanism for the progerin expression in the CKD arteries, we designed a ddPCR assay to analyze the presence of the *LMNA* c.1824C>T variant. DNA was isolated from the paraffin-embedded slides containing the epigastric arteries from the 40 CKD patients and the six controls previously stained for progerin. The DNA isolation was successful for 27 CKD and one control. The variant was detected in 16 CKD patients and one control; the same control as previously had shown weak progerin staining. The fractional abundance of the variant ranged from 0.73%-47.8%. In total, we detected the progerin expression or the progerin variant (*LMNA* c.1824C>T) in 34 out of 40 CKD and 1 out of 8 controls (Figure 13D).

Increased proliferation and DNA damage in the CKD arteries

The arterial tissue is characterized by low regeneration in basal conditions; however, increased proliferation is observed as a response to tissue damage (102, 103). Staining for PHH3, a marker of increased proliferation, identified proliferating cells in nine out of 11 CKD arteries. Most of the proliferating cells were detected in the adventitial, but as well in the medial layer.

A downstream effect of progerin expression is increased DNA damage and senescence (125). The analysis showed a significant positive correlation between the frequency of cells with activated DNA damage response and the frequency of progerin positive cells in artery sections (n=18; r=0.63, *P*=0.005, and n=14; r=0.66, *P*=0.0097, for ATR and 53BP1 respectively) (Figure 13E). Markers of inflammation and senescence indicated a modest vascular and systemic inflammation and activation of a senescence-associated secretory phenotype (SASP).

The effect of the progerin mutation *in vivo*

Progerin is associated with premature senescence and its accumulation has detrimental effects on the cell. Therefore we wanted to assess its effect on stem cells *in vivo*. To achieve that, we performed a competitive hematopoietic stem cell (HSC) transplantation in mice. This is an established method to test the functionality of stem cells *in vivo.* We used HGPS knock-in mice that carry the progerin mutation *Lmna* c.1827C>T, equivalent to human *LMNA* c.1824C>T (105).

HSCs from 22-week-old HGPS and wild-type mice were collected. At that time, HGPS mice are near their endpoint and have fewer HSCs in their bone marrow than age-matched wild-type mice. The HSCs were injected into recipient wild-type donor mice and peripheral blood samples collected at different time points. Flow cytometry was used to analyze the proportion of transplanted donor cells (CD45.2) to recipient cells (CD45.1). We observed a significant increased initial proliferation in the progeroid donor cells, compared to wild type donor cells. Moreover, the progeroid donor cells were still present 30 weeks after transplantation (Figure 13F). These results show that HSCs with the progerin mutation are viable and do not undergo senescence *in vivo*.

Here we report progerin positive cells as well as the progerin variant, *LMNA* c.1824C>T, in the arteries of CKD patients. In two of the CKD arteries, clusters of progerin positive cells were detected, moreover, a high fractional abundance of *LMNA* c.1824C>T variant was observed in the CKD arteries (up to 47.8%). This suggests a clonal expansion of a progenitor cell carrying the variant leading to progerin expression. Studies have shown that somatic mutations clonally expand in proliferating tissue (24, 104). Although the arterial wall does not generally contain proliferating cells, increased proliferation was detected in the CKD patients. The origin of proliferating cells in the arterial wall is unknown; however, it has been shown that some VSMCs are capable of proliferating. They are suggested to be specific hyper-proliferating cells that create patches of cells originating from the same single cell (103), potentially carrying pathogenic mutations (101, 102).

Expression of progerin or the *LMNA* c.1824C>T variant has been shown to lead to increased cell proliferation (126, 127); yet, progerin is toxic for the cell and high levels of progerin lead to senescence (128, 129). We observed that hematopoietic progenitor cells with the progerin mutation have initial growth advantages and do not undergo senescence.

Taken together, progerin is expressed in the CKD arteries and might contribute to the vascular phenotype observed in the CKD patients. Our results suggest that the *LMNA* c.1824C>T mutation is the cause of the expression of progerin and that the mutation might have occurred in a VSMC or a progenitor cell activated during the vascular regeneration. The proliferation during vascular regeneration might have promoted the clonal expansion of mutated clones within the tissue. The final effect of this expansion is an increased tissue-expression of progerin, which finally contributes to the worsening of the pathology.

6 CONCLUSIONS AND FUTURE PERSPECTIVE

Germline mutations and their contribution to diseases have been extensively studied, leading to the identification of genetic risk factors and improved knowledge about our genome. Conversely, somatic mutagenesis has only started to be explored. In addition, the study of somatic mutagenesis has often been limited to cancer. Very little information is available about somatic mutation accumulation during physiological aging, as well as in pathologies other than cancer. In this thesis, I analyzed somatic mutations in healthy tissues and age-related diseases, such as Alzheimer´s disease and chronic kidney disease. This work provides advancement in our understanding of the mutagenesis in different tissues and its possible contribution to age-associated diseases.

In **paper I** and **paper II**, we showed that somatic mutations occur in healthy tissues and the somatic mutation burden increases with increased age. We performed a genome-wide meta-analysis on healthy cells from various tissues and at different ages. We characterized a process that we called "basal mutagenesis" and that is the consequence of mutation accumulation during cell core functions. Furthermore, we observed that the pattern of mutations in a cell is not strictly defined by the tissue of origin. In fact, even in the same tissue, there might be cells exposed to different mutagens that demonstrate different mutation patterns, as exemplified by the kidney in our study. Our results provide a better understanding of the processes leading to the loss of genome integrity in different tissues and during aging. For example, a specific cell type in the kidney (proximal tubule epithelial cell) showed a mutation landscape characterized by high mutation accumulation and enrichment in regulatory regions and expressed genes. This pattern favors the accumulation of mutations in functional genes, thus explaining the high propensity to cancer initiation observed in this particular cell type. Further analysis of mutational patterns in healthy cells and cancer will enhance our understandings of mutagenesis and provide new perspectives in preventing cancer and other age-associated diseases. Future studies will also be of importance in the evaluation of the functional significance of an increased mutation burden in progenitor cells and how that relates to reduced tissue function as seen with increased age.

In **paper III**, we showed that somatic mutations occur in the brain of AD patients. It is challenging to detect variants at low frequency in a bulk tissue; however, we demonstrated that it is possible to obtain meaningful results by using ultra-deep sequencing and tailored bioinformatics analysis. For example, the variant upstream of the gene *CD55* that we identified in an AD patient was predicted to affect transcription factor binding sites and might contribute to AD through the complement system. The importance of somatic mutations in age-related diseases such as AD is gaining interest, but their role in the disease process is unknown. Further studies are needed to analyze the effect and its contribution to AD. More importantly, while most studies have focused on the coding regions for the occurrence of somatic mutations (65, 66, 68, 71, 98), we analyzed non-coding regions including regulatory regions. Our results emphasize the importance of including non-coding genomic regions in future studies on somatic mutagenesis to increase the understanding of AD.

In **paper IV**, we identified progerin and the *LMNA* c.1824C>T variant in the arteries of CKD patients. Furthermore, we detected increased proliferation and DNA damage in these arteries. Our data suggest extensive vascular regeneration in the arterial wall from progenitor cells carrying the *LMNA* c.1824C>T variant. That leads to increased progerin within the tissue, further contributing to the disease progression. Progerin expression has been found at low levels in healthy tissues (122-124), but to our knowledge, this is the first time that the *LMNA* c.1824C>T variant is detected in human samples from non-progeroid patients. Further analyses are needed to define how low levels of progerin affect a cell in healthy tissue. Our results show that somatic mutations can become a severe risk in regenerated tissues. Additional functional studies on the pathogenic mutation found in the CKD arteries are needed to understand the effect on the tissue and its contribution to the disease progression. Moreover, studying rare diseases like HGPS can help us to identify new treatments for more common diseases, as CKD. Despite being rare, several treatment strategies and clinical trials are currently ongoing in HGPS patients and models. Our findings suggest that those strategies should be tested in CKD models.

At the beginning of my Ph.D. research studies, the mutation burden of somatic SNVs in healthy tissues was relatively unknown. Moreover, although it has been suggested that somatic mutations contribute to age-associated diseases, the identification of pathogenic somatic mutations has been challenging. However, during these past years, several studies have been published on mutagenesis in healthy cells during aging and how somatic mutations contribute to diseases. The studies presented in this thesis shed further light on the somatic mutagenesis in different tissues and cell types, and its possible consequences on the cell and the tissue.

7 ACKNOWLEDGEMENTS

To my main supervisor, ***Maria Eriksson***. No matter if you have a deadline the next day, you always have time for us. Thanks for everything these past years! It has been a pleasure working with you. I have enjoyed dropping by your office and accidently spending hours discussing different projects, speculating on new theories or planning the next experiment. I lacked knowledge in wet lab experiments that you could provide me and you know my weak spots to push me forward to where we are today. I am sure our science paths will cross somewhere in the future.

Magnus Nordenskjöld, ***Karin Dahlman-Wright*** and ***Irene Franco***. Thanks for being my co-supervisors and for your support. Thanks to my mentor, ***Ingileif Jónsdóttir***.

Thanks to ***Karl Ekwall***, chairman of the Department of Biosciences and Nutrition and former chairman ***Karin Dahlman-Wright***, for providing excellent research place.

The past and present of ME group: What a great group it has been! Collection of people from France, Italy, Germany, Spain, Ireland, Slovenia, Sweden, Canada, the Netherlands, India and Iceland, that are so different but still fit so well together! ***Gwladys Revêchon***, it has been quite a journey! We had our registration seminars the same day, my halftime was week before yours, and now you defended your thesis a week before me. Can we say we are PhD-twins? It has been great sitting next to you all this time, both at the office as in the lab. It has been real pleasure to get to know you, discussing castles, cats, and of course science. You are a hard-working scientist, always ready to start a new experiment early in the morning or late afternoons. Most of all, you are always positive and ready for my "great" ideas… I still think our board games will be success. Best of luck with whatever you plan in the future. ***Daniel Whisenant***, we never let you forget that you are the youngest PhD student, although you are definitely not the youngest in the group! You are such an optimistic person and fit so perfect into the group. You have so many great ideas and work so hard towards your goals, an inspiration for us all! Enjoy the rest of your PhD study! ***Peter Vrtacnik***, you have always been willing to assist me in my projects and asked the questions I had not thought about. I have enjoyed your company at work and outside. Thanks for all the runs in Visättra and around Flemingsberg, and for the bbq parties. Enjoy your paternal leave and see you around. ***Irene Franco***, you are very determined hard-working scientists. You know so much about your research field and have clear view on your studies. It has been nice working with you discussing our projects. Thanks for the collaboration this past year and the assist with this thesis.

Pär Lundin**,** thanks for having the patience to introduce me to the beauty of bioinformatics, and teaching me to write scripts so I could analyze the sequencing data. I know it has not been an easy task, especially in the very beginning when I didn´t want to learn it. You have always been there when I have needed help with the analyses, or when I accidently drowned the Uppmax clusters by starting a for-loop that started several other for-loops! ***Agustin Sola Carvajal***, I could write a whole page for you! You know everything in the lab and you are

always ready to assist and answer my questions. You are a great friend at work and outside work, always ready for some crazy race even though I don´t think you will ever run Tjurruset with me again. Thanks for everything! ***Charlotte Strandgren***, it was nice sitting next to you in Novum and discussing our projects. I could always ask you if I had some questions regarding the paperwork for the PhD study, you are such an organized person. Thanks for the help! ***Nikenza Viceconte***, hard-working scientist that is ready to fly to Stockholm over a weekend to do IF stainings! Thanks for your collaboration. ***Tomás McKenna***, you were finishing your PhD study when I started and you made sure I felt welcome to the group. You kept on trying to get us all to climbing, camping or on skiing trips, and invited us to pancakes to your house. You are so full of ideas and started the Dead Cell Society with us. ***Emelie Wallén Arzt***, you started just two months after me. You tried to keep order on us, keeping the lab in place and making sure we all went together for lunch. Something we still do. You are such a nice person, and it is a shame we don´t meet more often. We really should take this beer we have planned for a very very long time! ***Robin Hagblom***, it is always so fun to be around you and talk to you. You are always ready for some funny business… specially when Tomás is around! Thanks for all the help in the lab and with the proofreading in the past. Thanks to past and present students in the group: ***Sofia Forsblom***, ***Oliver Pajong***, ***Calle Blomqvist***, ***Saskia Molina Pineau***, ***Sophie Schröder, Sharbari Das.***

Serena Barilla, you are such a nice person, always checking on us and making us feel better. I will miss discussing important and less important stuff with you, especially about pasta and what is allowed and not allowed on pizza! Thanks for everything at work and outside work, specially the trip to Iceland with ***Gwladys*** and ***Agus***. One of these days, I will go to your beautiful country.

To past and present lunch and coffee friends. ***Marko Giudici*** and ***Saioa Goni***, thanks for all the nice times in Novum and discussions while waiting for the coffee to cook. ***Ohiane García***, ***Sandra Scharaw***, ***Rodrigo Lozano***, ***Anna Webb***, ***Ana Amaral***, ***Sridharan Ganesan*****,** thanks for nice times in Novum and Neo.

Thanks to everyone that have assisted and showed up for our ***Dead Cell Society***. It started small in Novum and has grown quite a lot in Neo. It has been pleasure participating in this work and create opportunities to meet other members in Novum and Neo. I hope that DSC events will continue in the future. Best of luck with that.

Past and present members of BioNut, thanks for creating a nice working environment in Novum and Neo. Member of the neighboring groups in Neo, ***ERA*** and ***MBE***. It has been nice sharing the corner with you. Thanks to the staff at the administrative, the IT-unit and the service unit at the department. Special thanks to ***Monica Ahlberg*** for taking such a great care of me and for all the help regarding the PhD study. ***Inger Moge***, ***Johan Dethlefsen***, and others at the Purchasing and lab. service.

Thanks to all my co-authors for great collaboration. Special thanks to ***Peter Stenvinkel***, ***Anna Witasp***, ***Caroline Graff***, ***Fran Supek*** and to the members of NBIS Bioinformatics

Long-term Support (WABI): ***Björn Nystedt***, ***Anna Johansson***, ***Malin Larsson*** and ***Nina Norgren***.

The ***Lundins***: ***Pär***, ***Maja***, ***Elvira, Daniel***. Thanks for all the support through the years, you have taken me as one of the family and helped me so much. I wouldn´t be here without you. Thanks ***Elvira*** for the drawing on the cover and in this thesis. Thanks to ***Sonny Lundin*** for assist with the proofreading.

My friends at home and abroad, thanks for your support throughout the years.

And most of all, to my family. My mom ***Sjöfn***, my sister ***Dagrún***, my brothers ***Jón Þór***, ***Ómar*** and ***Björgvin***, my brother-in-law and my sisters-in-law, my seven nieces and seven nephews. Thanks for all the support and your love. We are so close family so being away has not always been easy, especially when new nieces or nephews have arrived. You have kept me updated on what is going so when I come for a visit it feels like I never left. Takk fyrir allt í gegnum árin!

8 REFERENCES

1. Dong X, Milholland B, Vijg J. Evidence for a limit to human lifespan. Nature. 2016 Oct 13;538(7624):257-9. PubMed PMID: 27706136.

2. WHO. World Health Statistics 2016: Monitoring health for the SDGs: World Health Organization; 2016. Available from: https://http://www.who.int/gho/publications/world_health_statistics/2016/en/.

3. Calderon-Larranaga A, Vetrano DL, Onder G, Gimeno-Feliu LA, Coscollar-Santaliestra C, Carfi A, et al. Assessing and Measuring Chronic Multimorbidity in the Older Population: A Proposal for Its Operationalization. The journals of gerontology Series A, Biological sciences and medical sciences. 2017 Oct 1;72(10):1417-23. PubMed PMID: 28003375. Pubmed Central PMCID: 5861938.

4. Herskind AM, McGue M, Holm NV, Sorensen TI, Harvald B, Vaupel JW. The heritability of human longevity: a population-based study of 2872 Danish twin pairs born 1870-1900. Human genetics. 1996 Mar;97(3):319-23. PubMed PMID: 8786073.

5. Gierman HJ, Fortney K, Roach JC, Coles NS, Li H, Glusman G, et al. Whole-genome sequencing of the world's oldest people. PloS one. 2014 Nov 12;9(11):e112430. PubMed PMID: 25390934. Pubmed Central PMCID: 4229186. English.

6. Deelen J, Beekman M, Uh HW, Broer L, Ayers KL, Tan Q, et al. Genome-wide association meta-analysis of human longevity identifies a novel locus conferring survival beyond 90 years of age. Human molecular genetics. 2014 Aug 15;23(16):4420-32. PubMed PMID: 24688116. Pubmed Central PMCID: 4103672.

7. Erikson GA, Bodian DL, Rueda M, Molparia B, Scott ER, Scott-Van Zeeland AA, et al. Whole-Genome Sequencing of a Healthy Aging Cohort. Cell. 2016 May 05;165(4):1002-11. PubMed PMID: 27114037. Pubmed Central PMCID: 4860090.

8. Conneely KN, Capell BC, Erdos MR, Sebastiani P, Solovieff N, Swift AJ, et al. Human longevity and common variations in the LMNA gene: a meta-analysis. Aging cell. 2012 Jun;11(3):475-81. PubMed PMID: 22340368. Pubmed Central PMCID: 3350595.

9. Deelen J, Evans DS, Arking DE, Tesi N, Nygaard M, Liu X, et al. A meta-analysis of genome-wide association studies identifies multiple longevity genes. Nature communications. 2019 Aug 14;10(1):3669. PubMed PMID: 31413261.

10. Tacutu R, Thornton D, Johnson E, Budovsky A, Barardo D, Craig T, et al. Human Ageing Genomic Resources: new and updated databases. Nucleic acids research. 2018 Jan 4;46(D1):D1083-D90. PubMed PMID: 29121237. Pubmed Central PMCID: 5753192.

11. Murabito JM, Yuan R, Lunetta KL. The search for longevity and healthy aging genes: insights from epidemiological studies and samples of long-lived individuals. The journals of gerontology Series A, Biological sciences and medical sciences. 2012 May;67(5):470-9. PubMed PMID: 22499766. Pubmed Central PMCID: 3326242.

12. Beekman M, Blanche H, Perola M, Hervonen A, Bezrukov V, Sikora E, et al. Genome-wide linkage analysis for human longevity: Genetics of Healthy Aging Study. Aging cell. 2013 Apr;12(2):184-93. PubMed PMID: 23286790. Pubmed Central PMCID: 3725963.

13. Lopez-Otin C, Blasco MA, Partridge L, Serrano M, Kroemer G. The hallmarks of aging. Cell. 2013 Jun 06;153(6):1194-217. PubMed PMID: 23746838. Pubmed Central PMCID: 3836174.

14. Eriksson M, Brown WT, Gordon LB, Glynn MW, Singer J, Scott L, et al. Recurrent de novo point mutations in lamin A cause Hutchinson-Gilford progeria syndrome. Nature. 2003 May 15;423(6937):293-8. PubMed PMID: 12714972.

15. De Sandre-Giovannoli A, Bernard R, Cau P, Navarro C, Amiel J, Boccaccio I, et al. Lamin a truncation in Hutchinson-Gilford progeria. Science. 2003 Jun 27;300(5628):2055. PubMed PMID: 12702809.

16. Merideth MA, Gordon LB, Clauss S, Sachdev V, Smith AC, Perry MB, et al. Phenotype and course of Hutchinson-Gilford progeria syndrome. The New England journal of medicine. 2008 Feb 07;358(6):592-604. PubMed PMID: 18256394. Pubmed Central PMCID: 2940940.

17. Goto M, Rubenstein M, Weber J, Woods K, Drayna D. Genetic linkage of Werner's syndrome to five markers on chromosome 8. Nature. 1992 Feb 20;355(6362):735-8. PubMed PMID: 1741060.

18. Brooks PJ. The 8,5'-cyclopurine-2'-deoxynucleosides: candidate neurodegenerative DNA lesions in xeroderma pigmentosum, and unique probes of transcription and nucleotide excision repair. DNA repair. 2008 Jul 1;7(7):1168-79. PubMed PMID: 18495558. Pubmed Central PMCID: 2797313.

19. Carrero D, Soria-Valles C, Lopez-Otin C. Hallmarks of progeroid syndromes: lessons from mice and reprogrammed cells. Disease models & mechanisms. 2016 Jul 1;9(7):719-35. PubMed PMID: 27482812. Pubmed Central PMCID: 4958309.

20. Milholland B, Dong X, Zhang L, Hao X, Suh Y, Vijg J. Differences between germline and somatic mutation rates in humans and mice. Nature communications. 2017 May 9;8:15183. PubMed PMID: 28485371. Pubmed Central PMCID: 5436103.

21. Blokzijl F, de Ligt J, Jager M, Sasselli V, Roerink S, Sasaki N, et al. Tissue-specific mutation accumulation in human adult stem cells during life. Nature. 2016 Oct 13;538(7624):260-4. PubMed PMID: 27698416. Pubmed Central PMCID: 5536223.

22. Forsberg LA, Gisselsson D, Dumanski JP. Mosaicism in health and disease - clones picking up speed. Nature reviews Genetics. 2017 Feb;18(2):128-42. PubMed PMID: 27941868.

23. Martincorena I, Campbell PJ. Somatic mutation in cancer and normal cells. Science. 2015 Sep 25;349(6255):1483-9. PubMed PMID: 26404825.

24. Martincorena I, Fowler JC, Wabik A, Lawson ARJ, Abascal F, Hall MWJ, et al. Somatic mutant clones colonize the human esophagus with age. Science. 2018 Oct 18. PubMed PMID: 30337457.

25. Abyzov A, Tomasini L, Zhou B, Vasmatzis N, Coppola G, Amenduni M, et al. One thousand somatic SNVs per skin fibroblast cell set baseline of mosaic mutational load with patterns that suggest proliferative origin. Genome research. 2017 Apr;27(4):512-23. PubMed PMID: 28235832. Pubmed Central PMCID: 5378170.

26. Alexandrov LB, Jones PH, Wedge DC, Sale JE, Campbell PJ, Nik-Zainal S, et al. Clock-like mutational processes in human somatic cells. Nature genetics. 2015 Dec;47(12):1402-7. PubMed PMID: 26551669. Pubmed Central PMCID: 4783858.

27. Lodato MA, Rodin RE, Bohrson CL, Coulter ME, Barton AR, Kwon M, et al. Aging and neurodegeneration are associated with increased mutations in single human neurons. Science. 2018 Feb 2;359(6375):555-9. PubMed PMID: 29217584. Pubmed Central PMCID: 5831169.

28. Lodato MA, Woodworth MB, Lee S, Evrony GD, Mehta BK, Karger A, et al. Somatic mutation in single human neurons tracks developmental and transcriptional history. Science. 2015 Oct 02;350(6256):94-8. PubMed PMID: 26430121. Pubmed Central PMCID: 4664477.

29. Lee-Six H, Ellis P, Osborne RJ, Sanders MA, Moore L, Georgakopoulos N, et al. The landscape of somatic mutation in normal colorectal epithelial cells. bioRxiv. 2018:416800.

30. Lee-Six H, Obro NF, Shepherd MS, Grossmann S, Dawson K, Belmonte M, et al. Population dynamics of normal human blood inferred from somatic mutations. Nature. 2018 Sep;561(7724):473-8. PubMed PMID: 30185910. Pubmed Central PMCID: 6163040.

31. Moore L, Leongamornlert D, Coorens THH, Sanders MA, Ellis P, Dawson K, et al. The mutational landscape of normal human endometrial epithelium. bioRxiv. 2018:505685.

32. Saini N, Roberts SA, Klimczak LJ, Chan K, Grimm SA, Dai S, et al. The Impact of Environmental and Endogenous Damage on Somatic Mutation Load in Human Skin Fibroblasts. PLoS genetics. 2016 Oct;12(10):e1006385. PubMed PMID: 27788131. Pubmed Central PMCID: 5082821.

33. Zhang L, Dong X, Lee M, Maslov AY, Wang T, Vijg J. Single-cell whole-genome sequencing reveals the functional landscape of somatic mutations in B lymphocytes across the human lifespan. Proceedings of the National Academy of Sciences of the United States of America. 2019 Apr 16. PubMed PMID: 30992375.

34. Zhang L, Vijg J. Somatic Mutagenesis in Mammals and Its Implications for Human Disease and Aging. Annual review of genetics. 2018 Nov 23;52:397-419. PubMed PMID: 30212236.

35. Encode Project Consortium. An integrated encyclopedia of DNA elements in the human genome. Nature. 2012 Sep 06;489(7414):57-74. PubMed PMID: 22955616. Pubmed Central PMCID: 3439153.

36. Brash DE. UV signature mutations. Photochemistry and photobiology. 2015 Jan-Feb;91(1):15-26. PubMed PMID: 25354245. Pubmed Central PMCID: 4294947.

37. Hoeijmakers JH. DNA damage, aging, and cancer. The New England journal of medicine. 2009 Oct 08;361(15):1475-85. PubMed PMID: 19812404.

38. Akbari M, Krokan HE. Cytotoxicity and mutagenicity of endogenous DNA base lesions as potential cause of human aging. Mechanisms of ageing and development. 2008 Jul-Aug;129(7-8):353-65. PubMed PMID: 18355895.

39. Helleday T, Eshtad S, Nik-Zainal S. Mechanisms underlying mutational signatures in human cancers. Nature reviews Genetics. 2014 Sep;15(9):585-98. PubMed PMID: 24981601. Pubmed Central PMCID: 6044419.

40. Frigola J, Sabarinathan R, Mularoni L, Muinos F, Gonzalez-Perez A, Lopez-Bigas N. Reduced mutation rate in exons due to differential mismatch repair. Nature genetics. 2017 Dec;49(12):1684-92. PubMed PMID: 29106418. Pubmed Central PMCID: 5712219.

41. Supek F, Lehner B. Differential DNA mismatch repair underlies mutation rate variation across the human genome. Nature. 2015 May 07;521(7550):81-4. PubMed PMID: 25707793. Pubmed Central PMCID: 4425546.

42. Encode Project Consortium, Birney E, Stamatoyannopoulos JA, Dutta A, Guigo R, Gingeras TR, et al. Identification and analysis of functional elements in 1% of the human genome by the ENCODE pilot project. Nature. 2007 Jun 14;447(7146):799-816. PubMed PMID: 17571346. Pubmed Central PMCID: 2212820.

43. Liu L, De S, Michor F. DNA replication timing and higher-order nuclear organization determine single-nucleotide substitution patterns in cancer genomes. Nature communications. 2013;4:1502. PubMed PMID: 23422670. Pubmed Central PMCID: 3633418.

44. Stamatoyannopoulos JA, Adzhubei I, Thurman RE, Kryukov GV, Mirkin SM, Sunyaev SR. Human mutation rate associated with DNA replication timing. Nature genetics. 2009 Apr;41(4):393-5. PubMed PMID: 19287383. Pubmed Central PMCID: 2914101.

45. Gorbunova V, Seluanov A, Mao Z, Hine C. Changes in DNA repair during aging. Nucleic acids research. 2007;35(22):7466-74. PubMed PMID: 17913742. Pubmed Central PMCID: 2190694.

46. Vogelstein B, Papadopoulos N, Velculescu VE, Zhou S, Diaz LA, Jr., Kinzler KW. Cancer genome landscapes. Science. 2013 Mar 29;339(6127):1546-58. PubMed PMID: 23539594. Pubmed Central PMCID: 3749880.

47. Zhu M, Lu T, Jia Y, Luo X, Gopal P, Li L, et al. Somatic Mutations Increase Hepatic Clonal Fitness and Regeneration in Chronic Liver Disease. Cell. 2019 Apr 18;177(3):608-21 e12. PubMed PMID: 30955891.

48. Revy P, Kannengiesser C, Fischer A. Somatic genetic rescue in Mendelian haematopoietic diseases. Nature reviews Genetics. 2019 Jun 11. PubMed PMID: 31186537.

49. Middlebrooks CD, Banday AR, Matsuda K, Udquim KI, Onabajo OO, Paquin A, et al. Association of germline variants in the APOBEC3 region with cancer risk and enrichment with APOBEC-signature mutations in tumors. Nature genetics. 2016 Nov;48(11):1330-8. PubMed PMID: 27643540. Pubmed Central PMCID: 6583788.

50. Waszak SM, Tiao G, Zhu B, Rausch T, Muyas F, Rodríguez-Martín B, et al. Germline determinants of the somatic mutation landscape in 2,642 cancer genomes. bioRxiv. 2017 2017-01-01 00:00:00.

51. Thorgeirsson TE, Steinberg S, Reginsson GW, Bjornsdottir G, Rafnar T, Jonsdottir I, et al. A rare missense mutation in CHRNA4 associates with smoking behavior and its consequences. Molecular psychiatry. 2016 May;21(5):594-600. PubMed PMID: 26952864. Pubmed Central PMCID: 5414061.

52. Jonsson H, Sulem P, Arnadottir GA, Palsson G, Eggertsson HP, Kristmundsdottir S, et al. Multiple transmissions of de novo mutations in families. Nature genetics. 2018 Nov 5. PubMed PMID: 30397338.

53. Campbell IM, Yuan B, Robberecht C, Pfundt R, Szafranski P, McEntagart ME, et al. Parental somatic mosaicism is underrecognized and influences recurrence risk of genomic disorders. American journal of human genetics. 2014 Aug 07;95(2):173-82. PubMed PMID: 25087610. Pubmed Central PMCID: 4129404.

54. Erickson RP. Somatic gene mutation and human disease other than cancer: an update. Mutation research. 2010 Oct;705(2):96-106. PubMed PMID: 20399892.

55. Frederiksen AL, Duno M, Johnsen IB, Nielsen MF, Kroigard AB. Asymptomatic parental mosaicism for osteogenesis imperfecta associated with a new splice site mutation in COL1A2. Clinical case reports. 2016 Oct;4(10):972-8. PubMed PMID: 27761249. Pubmed Central PMCID: 5054473.

56. Pyott SM, Pepin MG, Schwarze U, Yang K, Smith G, Byers PH. Recurrence of perinatal lethal osteogenesis imperfecta in sibships: parsing the risk between parental mosaicism for dominant mutations and autosomal recessive inheritance. Genetics in medicine : official journal of the American College of Medical Genetics. 2011 Feb;13(2):125-30. PubMed PMID: 21239989.

57. Wuyts W, Biervliet M, Reyniers E, D'Apice MR, Novelli G, Storm K. Somatic and gonadal mosaicism in Hutchinson-Gilford progeria. American journal of medical genetics Part A. 2005 May 15;135(1):66-8. PubMed PMID: 15793835.

58. Cibulskis K, Lawrence MS, Carter SL, Sivachenko A, Jaffe D, Sougnez C, et al. Sensitive detection of somatic point mutations in impure and heterogeneous cancer samples. Nature biotechnology. 2013 Mar;31(3):213-9. PubMed PMID: 23396013. Pubmed Central PMCID: 3833702.

59. Osorio FG, Rosendahl Huber A, Oka R, Verheul M, Patel SH, Hasaart K, et al. Somatic Mutations Reveal Lineage Relationships and Age-Related Mutagenesis in Human Hematopoiesis. Cell reports. 2018 Nov 27;25(9):2308-16 e4. PubMed PMID: 30485801.

60. Dong X, Zhang L, Milholland B, Lee M, Maslov AY, Wang T, et al. Accurate identification of single-nucleotide variants in whole-genome-amplified single cells. Nature methods. 2017 May;14(5):491-3. PubMed PMID: 28319112. Pubmed Central PMCID: 5408311.

61. Enge M, Arda HE, Mignardi M, Beausang J, Bottino R, Kim SK, et al. Single-Cell Analysis of Human Pancreas Reveals Transcriptional Signatures of Aging and Somatic Mutation Patterns. Cell. 2017 Oct 5;171(2):321-30 e14. PubMed PMID: 28965763. Pubmed Central PMCID: 6047899.

62. Nam AS, Kim KT, Chaligne R, Izzo F, Ang C, Taylor J, et al. Somatic mutations and cell identity linked by Genotyping of Transcriptomes. Nature. 2019 Jul 3;571(7765):355-60. PubMed PMID: 31270458.

63. Wang Y, Navin NE. Advances and applications of single-cell sequencing technologies. Molecular cell. 2015 May 21;58(4):598-609. PubMed PMID: 26000845. Pubmed Central PMCID: 4441954.

64. Poirion O, Zhu X, Ching T, Garmire LX. Using single nucleotide variations in single-cell RNA-seq to identify subpopulations and genotype-phenotype linkage. Nature communications. 2018 Nov 20;9(1):4892. PubMed PMID: 30459309. Pubmed Central PMCID: 6244222.

65. Keogh MJ, Wei W, Aryaman J, Walker L, van den Ameele J, Coxhead J, et al. High prevalence of focal and multi-focal somatic genetic variants in the human brain. Nature communications. 2018 Oct 15;9(1):4257. PubMed PMID: 30323172. Pubmed Central PMCID: 6189186.

66. Nicolas G, Acuna-Hidalgo R, Keogh MJ, Quenez O, Steehouwer M, Lelieveld S, et al. Somatic variants in autosomal dominant genes are a rare cause of sporadic Alzheimer's disease. Alzheimer's & dementia : the journal of the Alzheimer's Association. 2018 Dec;14(12):1632-9. PubMed PMID: 30114415.

67. Sala Frigerio C, Lau P, Troakes C, Deramecourt V, Gele P, Van Loo P, et al. On the identification of low allele frequency mosaic mutations in the brains of Alzheimer's disease patients. Alzheimer's & dementia : the journal of the Alzheimer's Association. 2015 Nov;11(11):1265-76. PubMed PMID: 25937274.

68. Beck JA, Poulter M, Campbell TA, Uphill JB, Adamson G, Geddes JF, et al. Somatic and germline mosaicism in sporadic early-onset Alzheimer's disease. Human molecular genetics. 2004 Jun 15;13(12):1219-24. PubMed PMID: 15115757.

69. Hindson BJ, Ness KD, Masquelier DA, Belgrader P, Heredia NJ, Makarewicz AJ, et al. High-throughput droplet digital PCR system for absolute quantitation of DNA copy number. Analytical chemistry. 2011 Nov 15;83(22):8604-10. PubMed PMID: 22035192. Pubmed Central PMCID: 3216358.

70. Miotke L, Lau BT, Rumma RT, Ji HP. High sensitivity detection and quantitation of DNA copy number and single nucleotide variants with single color droplet digital PCR. Analytical chemistry. 2014 Mar 04;86(5):2618-24. PubMed PMID: 24483992. Pubmed Central PMCID: 3982983.

71. Park JS, Lee J, Jung ES, Kim MH, Kim IB, Son H, et al. Brain somatic mutations observed in Alzheimer's disease associated with aging and dysregulation of tau phosphorylation. Nature communications. 2019 Jul 12;10(1):3090. PubMed PMID: 31300647. Pubmed Central PMCID: 6626023.

72. Koboldt DC, Zhang Q, Larson DE, Shen D, McLellan MD, Lin L, et al. VarScan 2: somatic mutation and copy number alteration discovery in cancer by exome sequencing. Genome research. 2012 Mar;22(3):568-76. PubMed PMID: 22300766. Pubmed Central PMCID: 3290792.

73. Kim S, Scheffler K, Halpern AL, Bekritsky MA, Noh E, Kallberg M, et al. Strelka2: fast and accurate calling of germline and somatic variants. Nature methods. 2018 Aug;15(8):591-4. PubMed PMID: 30013048.

74. Uhlen M, Fagerberg L, Hallstrom BM, Lindskog C, Oksvold P, Mardinoglu A, et al. Proteomics. Tissue-based map of the human proteome. Science. 2015 Jan 23;347(6220):1260419. PubMed PMID: 25613900.

75. Uhlen M, Oksvold P, Fagerberg L, Lundberg E, Jonasson K, Forsberg M, et al. Towards a knowledge-based Human Protein Atlas. Nature biotechnology. 2010 Dec;28(12):1248-50. PubMed PMID: 21139605.

76. Blokzijl F, Janssen R, van Boxtel R, Cuppen E. MutationalPatterns: comprehensive genome-wide analysis of mutational processes. Genome medicine. 2018 Apr 25;10(1):33. PubMed PMID: 29695279. Pubmed Central PMCID: 5922316.

77. McLaren W, Gil L, Hunt SE, Riat HS, Ritchie GR, Thormann A, et al. The Ensembl Variant Effect Predictor. Genome biology. 2016 Jun 6;17(1):122. PubMed PMID: 27268795. Pubmed Central PMCID: 4893825.

78. Cingolani P, Platts A, Wang le L, Coon M, Nguyen T, Wang L, et al. A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of Drosophila melanogaster strain w1118; iso-2; iso-3. Fly. 2012 Apr-Jun;6(2):80-92. PubMed PMID: 22728672. Pubmed Central PMCID: 3679285.

79. Alexandrov LB, Ju YS, Haase K, Van Loo P, Martincorena I, Nik-Zainal S, et al. Mutational signatures associated with tobacco smoking in human cancer. Science. 2016 Nov 4;354(6312):618-22. PubMed PMID: 27811275. Pubmed Central PMCID: 6141049.

80. Alexandrov LB, Nik-Zainal S, Wedge DC, Aparicio SA, Behjati S, Biankin AV, et al. Signatures of mutational processes in human cancer. Nature. 2013 Aug 22;500(7463):415-21. PubMed PMID: 23945592. Pubmed Central PMCID: 3776390.

81. Alexandrov LB, Kim J, Haradhvala NJ, Huang MN, Ng AWT, Wu Y, et al. The Repertoire of Mutational Signatures in Human Cancer. bioRxiv. 2019:322859.

82. Nik-Zainal S, Alexandrov LB, Wedge DC, Van Loo P, Greenman CD, Raine K, et al. Mutational processes molding the genomes of 21 breast cancers. Cell. 2012 May 25;149(5):979-93. PubMed PMID: 22608084. Pubmed Central PMCID: 3414841.

83. Jager M, Blokzijl F, Kuijk E, Bertl J, Vougioukalaki M, Janssen R, et al. Deficiency of nucleotide excision repair is associated with mutational signature observed in cancer. Genome research. 2019 Jul;29(7):1067-77. PubMed PMID: 31221724.

84. Zou X, Owusu M, Harris R, Jackson SP, Loizou JI, Nik-Zainal S. Validating the concept of mutational signatures with isogenic cell models. Nature communications. 2018 May 1;9(1):1744. PubMed PMID: 29717121. Pubmed Central PMCID: 5931590.

85. Kucab JE, Zou X, Morganella S, Joel M, Nanda AS, Nagy E, et al. A Compendium of Mutational Signatures of Environmental Agents. Cell. 2019 May 2;177(4):821-36 e16. PubMed PMID: 30982602. Pubmed Central PMCID: 6506336.

86. Alexandrov LB, Nik-Zainal S, Wedge DC, Campbell PJ, Stratton MR. Deciphering signatures of mutational processes operative in human cancer. Cell reports. 2013 Jan 31;3(1):246-59. PubMed PMID: 23318258. Pubmed Central PMCID: 3588146.

87. Rosenthal R, McGranahan N, Herrero J, Taylor BS, Swanton C. DeconstructSigs: delineating mutational processes in single tumors distinguishes DNA repair deficiencies and patterns of carcinoma evolution. Genome biology. 2016 Feb 22;17:31. PubMed PMID: 26899170. Pubmed Central PMCID: 4762164.

88. Maura F, Degasperi A, Nadeu F, Leongamornlert D, Davies H, Moore L, et al. A practical guide for mutational signature analysis in hematological malignancies. Nature communications. 2019 Jul 5;10(1):2969. PubMed PMID: 31278357. Pubmed Central PMCID: 6611883.

89. Milholland B, Auton A, Suh Y, Vijg J. Age-related somatic mutations in the cancer genome. Oncotarget. 2015 Sep 22;6(28):24627-35. PubMed PMID: 26384365. Pubmed Central PMCID: 4694783.

90. Tomasetti C, Vogelstein B. Cancer etiology. Variation in cancer risk among tissues can be explained by the number of stem cell divisions. Science. 2015 Jan 2;347(6217):78-81. PubMed PMID: 25554788. Pubmed Central PMCID: 4446723.

91. Zhang F, Gu W, Hurles ME, Lupski JR. Copy number variation in human health, disease, and evolution. Annual review of genomics and human genetics. 2009;10:451-81. PubMed PMID: 19715442. Pubmed Central PMCID: 4472309.

92. Forsberg LA, Rasi C, Malmqvist N, Davies H, Pasupulati S, Pakalapati G, et al. Mosaic loss of chromosome Y in peripheral blood is associated with shorter survival and higher risk of cancer. Nature genetics. 2014 Jun;46(6):624-8. PubMed PMID: 24777449. Pubmed Central PMCID: 5536222.

93. Dumanski JP, Rasi C, Lonn M, Davies H, Ingelsson M, Giedraitis V, et al. Mutagenesis. Smoking is associated with mosaic loss of chromosome Y. Science. 2015 Jan 2;347(6217):81-3. PubMed PMID: 25477213. Pubmed Central PMCID: 4356728.

94. Dumanski JP, Lambert JC, Rasi C, Giedraitis V, Davies H, Grenier-Boley B, et al. Mosaic Loss of Chromosome Y in Blood Is Associated with Alzheimer Disease. American journal of human genetics. 2016 Jun 02;98(6):1208-19. PubMed PMID: 27231129. Pubmed Central PMCID: 4908225.

95. Jamuar SS, Lam AT, Kircher M, D'Gama AM, Wang J, Barry BJ, et al. Somatic mutations in cerebral cortical malformations. The New England journal of medicine. 2014 Aug 21;371(8):733-43. PubMed PMID: 25140959. Pubmed Central PMCID: 4274952.

96. Poduri A, Evrony GD, Cai X, Walsh CA. Somatic mutation, genomic variation, and neurological disease. Science. 2013 Jul 05;341(6141):1237758. PubMed PMID: 23828942. Pubmed Central PMCID: 3909954.

97. Baillie JK, Barnett MW, Upton KR, Gerhardt DJ, Richmond TA, De Sapio F, et al. Somatic retrotransposition alters the genetic landscape of the human brain. Nature. 2011 Oct 30;479(7374):534-7. PubMed PMID: 22037309. Pubmed Central PMCID: 3224101.

98. Parcerisas A, Rubio SE, Muhaisen A, Gomez-Ramos A, Pujadas L, Puiggros M, et al. Somatic signature of brain-specific single nucleotide variations in sporadic Alzheimer's disease. Journal of Alzheimer's disease : JAD. 2014;42(4):1357-82. PubMed PMID: 25024348.

99. Lee MH, Siddoway B, Kaeser GE, Segota I, Rivera R, Romanow WJ, et al. Somatic APP gene recombination in Alzheimer's disease and normal neurons. Nature. 2018 Nov;563(7733):639-45. PubMed PMID: 30464338. Pubmed Central PMCID: 6391999.

100. Tan AY, Zhang T, Michaeel A, Blumenfeld J, Liu G, Zhang W, et al. Somatic Mutations in Renal Cyst Epithelium in Autosomal Dominant Polycystic Kidney Disease. Journal of the American Society of Nephrology : JASN. 2018 Aug;29(8):2139-56. PubMed PMID: 30042192. Pubmed Central PMCID: 6065086.

101. De Flora S, Izzotti A. Mutagenesis and cardiovascular diseases Molecular mechanisms, risk factors, and protective factors. Mutation research. 2007 Aug 1;621(1-2):5-17. PubMed PMID: 17383689.

102. Weakley SM, Jiang J, Kougias P, Lin PH, Yao Q, Brunicardi FC, et al. Role of somatic mutations in vascular disease formation. Expert review of molecular diagnostics. 2010 Mar;10(2):173-85. PubMed PMID: 20214536. Pubmed Central PMCID: 2920047.

103. Chappell J, Harman JL, Narasimhan VM, Yu H, Foote K, Simons BD, et al. Extensive Proliferation of a Subset of Differentiated, yet Plastic, Medial Vascular Smooth Muscle Cells

Contributes to Neointimal Formation in Mouse Injury and Atherosclerosis Models. Circulation research. 2016 Dec 9;119(12):1313-23. PubMed PMID: 27682618. Pubmed Central PMCID: 5149073.

104. Yizhak K, Aguet F, Kim J, Hess JM, Kubler K, Grimsby J, et al. RNA sequence analysis reveals macroscopic somatic clonal expansion across normal tissues. Science. 2019 Jun 7;364(6444). PubMed PMID: 31171663.

105. Osorio FG, Navarro CL, Cadinanos J, Lopez-Mejia IC, Quiros PM, Bartoli C, et al. Splicing-directed therapy in a new mouse model of human accelerated aging. Science translational medicine. 2011 Oct 26;3(106):106ra7. PubMed PMID: 22030750.

106. Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics. 2009 Jul 15;25(14):1754-60. PubMed PMID: 19451168. Pubmed Central PMCID: 2705234.

107. McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernytsky A, et al. The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. Genome research. 2010 Sep;20(9):1297-303. PubMed PMID: 20644199. Pubmed Central PMCID: 2928508.

108. Li H. FermiKit: assembly-based variant calling for Illumina resequencing data. Bioinformatics. 2015 Nov 15;31(22):3694-6. PubMed PMID: 26220959. Pubmed Central PMCID: 4757955.

109. Avgustinova A, Symeonidi A, Castellanos A, Urdiroz-Urricelqui U, Sole-Boldo L, Martin M, et al. Loss of G9a preserves mutation patterns but increases chromatin accessibility, genomic instability and aggressiveness in skin tumours. Nature cell biology. 2018 Dec;20(12):1400-9. PubMed PMID: 30455462.

110. Supek F, Lehner B. Clustered Mutation Signatures Reveal that Error-Prone DNA Repair Targets Mutations to Active Genes. Cell. 2017 Jul 27;170(3):534-47 e23. PubMed PMID: 28753428.

111. Kircher M, Witten DM, Jain P, O'Roak BJ, Cooper GM, Shendure J. A general framework for estimating the relative pathogenicity of human genetic variants. Nature genetics. 2014 Mar;46(3):310-5. PubMed PMID: 24487276. Pubmed Central PMCID: 3992975.

112. Fishilevich S, Nudel R, Rappaport N, Hadar R, Plaschkes I, Iny Stein T, et al. GeneHancer: genome-wide integration of enhancers and target genes in GeneCards. Database : the journal of biological databases and curation. 2017 Jan 1;2017. PubMed PMID: 28605766. Pubmed Central PMCID: 5467550.

113. Farre D, Roset R, Huerta M, Adsuara JE, Rosello L, Alba MM, et al. Identification of patterns in biological sequences at the ALGGEN server: PROMO and MALGEN. Nucleic acids research. 2003 Jul 1;31(13):3651-3. PubMed PMID: 12824386. Pubmed Central PMCID: 169011.

114. Messeguer X, Escudero R, Farre D, Nunez O, Martinez J, Alba MM. PROMO: detection of known transcription regulatory elements using species-tailored searches. Bioinformatics. 2002 Feb;18(2):333-4. PubMed PMID: 11847087.

115. Mathelier A, Zhao X, Zhang AW, Parcy F, Worsley-Hunt R, Arenillas DJ, et al. JASPAR 2014: an extensively expanded and updated open-access database of transcription factor binding profiles. Nucleic acids research. 2014 Jan;42(Database issue):D142-7. PubMed PMID: 24194598. Pubmed Central PMCID: 3965086.

116. Haradhvala NJ, Polak P, Stojanov P, Covington KR, Shinbrot E, Hess JM, et al. Mutational Strand Asymmetries in Cancer Genomes Reveal Mechanisms of DNA Damage and Repair. Cell. 2016 Jan 28;164(3):538-49. PubMed PMID: 26806129. Pubmed Central PMCID: 4753048.

117. Lambert JC, Ibrahim-Verbaas CA, Harold D, Naj AC, Sims R, Bellenguez C, et al. Meta-analysis of 74,046 individuals identifies 11 new susceptibility loci for Alzheimer's disease. Nature genetics. 2013 Dec;45(12):1452-8. PubMed PMID: 24162737. Pubmed Central PMCID: 3896259.

118. Yasojima K, Schwab C, McGeer EG, McGeer PL. Up-regulated production and activation of the complement system in Alzheimer's disease brain. The American journal of pathology. 1999 Mar;154(3):927-36. PubMed PMID: 10079271. Pubmed Central PMCID: 1866427.

119. Morgan BP. Complement in the pathogenesis of Alzheimer's disease. Seminars in immunopathology. 2018 Jan;40(1):113-24. PubMed PMID: 29134267. Pubmed Central PMCID: 5794825.

120. Kolev MV, Ruseva MM, Harris CL, Morgan BP, Donev RM. Implication of complement system and its regulators in Alzheimer's disease. Current neuropharmacology. 2009 Mar;7(1):1-8. PubMed PMID: 19721814. Pubmed Central PMCID: 2724661.

121. Simone S, Loverre A, Cariello M, Divella C, Castellano G, Gesualdo L, et al. Arteriovenous fistula stenosis in hemodialysis patients is characterized by an increased adventitial fibrosis. Journal of nephrology. 2014 Oct;27(5):555-62. PubMed PMID: 24574138.

122. Olive M, Harten I, Mitchell R, Beers JK, Djabali K, Cao K, et al. Cardiovascular pathology in Hutchinson-Gilford progeria: correlation with the vascular pathology of aging. Arteriosclerosis, thrombosis, and vascular biology. 2010 Nov;30(11):2301-9. PubMed PMID: 20798379. Pubmed Central PMCID: 2965471.

123. McClintock D, Ratner D, Lokuge M, Owens DM, Gordon LB, Collins FS, et al. The mutant form of lamin A that causes Hutchinson-Gilford progeria is a biomarker of cellular aging in human skin. PloS one. 2007 Dec 05;2(12):e1269. PubMed PMID: 18060063. Pubmed Central PMCID: 2092390.

124. Scaffidi P, Misteli T. Lamin A-dependent nuclear defects in human aging. Science. 2006 May 19;312(5776):1059-63. PubMed PMID: 16645051. Pubmed Central PMCID: 1855250.

125. Gordon LB, Rothman FG, Lopez-Otin C, Misteli T. Progeria: a paradigm for translational medicine. Cell. 2014 Jan 30;156(3):400-7. PubMed PMID: 24485450.

126. Revechon G, Viceconte N, McKenna T, Sola Carvajal A, Vrtacnik P, Stenvinkel P, et al. Rare progerin-expressing preadipocytes and adipocytes contribute to tissue depletion over time. Scientific reports. 2017 Jun 30;7(1):4405. PubMed PMID: 28667315. Pubmed Central PMCID: 5493617.

127. Sagelius H, Rosengardten Y, Hanif M, Erdos MR, Rozell B, Collins FS, et al. Targeted transgenic expression of the mutation causing Hutchinson-Gilford progeria syndrome leads to proliferative and degenerative epidermal disease. Journal of cell science. 2008 Apr 01;121(Pt 7):969-78. PubMed PMID: 18334552.

128. Goldman RD, Shumaker DK, Erdos MR, Eriksson M, Goldman AE, Gordon LB, et al. Accumulation of mutant lamin A causes progressive changes in nuclear architecture in Hutchinson-Gilford progeria syndrome. Proceedings of the National Academy of Sciences of the United States of America. 2004 Jun 15;101(24):8963-8. PubMed PMID: 15184648. Pubmed Central PMCID: 428455.

129. McKenna T, Rosengardten Y, Viceconte N, Baek JH, Grochova D, Eriksson M. Embryonic expression of the common progeroid lamin A splice mutation arrests postnatal skin development. Aging cell. 2014 Apr;13(2):292-302. PubMed PMID: 24305605. Pubmed Central PMCID: 4331787.