A Non-Invasive Insight into Soft-Tissue Sarcomas

Seyed Hossein Moosavi

UNIVERSITY OF OSLO

June 2016

Department of Biosciences

Faculty of Mathematics and Natural Sciences

Department of Tumour Biology

Institute for Cancer Research

The Norwegian Radium Hospital

Oslo University Hospital

A Non-Invasive Insight into Soft-Tissue Sarcomas

Seyed Hossein Moosavi

Department of Biosciences

University of Oslo

May 2016

A Non-Invasive Insight into Soft-Tissue Sarcomas
Seyed Hossein Moosavi
http://www.duo.uio.no

Trykk: Reprosentralen, Universitetet i Oslo

Abstract

Introduction: Sarcomas are an infrequent and highly heterogeneous group of mesenchymal tumours, accounting for 1% of all human cancer worldwide, with more than 50 histological subtypes and incidence rate of about 20% among all paediatric solid malignant cancers. The aetiology of most of sarcomas is poorly understood, however, chromosomal aberrations are the major genetic event, underlying the pathogenesis of sarcomas. To date, molecular diagnosis of the disease has been confined to the aberrations in a few genes. However, manifestation of highly complex karyotype and diverse structural aberrations observed in many sarcomas has made the cytogenetic analysis inevitably laborious. The acquisition of tumour genetic profile for diagnostic and prognostic purposes by tumour biopsy has been long utilized in cancer care. Nevertheless, it has been unable to depict the complete tumour genetic landscape, neither spatially nor temporally. Owing to advances in the next-generation sequencing (NGS) methodologies, liquid biopsies of tumours by circulating cell-free DNA (cfDNA) have facilitated tumour genotyping and allows minimally invasive monitoring of tumour dynamics.

Material and Methods: We generated sequencing libraries from isolated tumour and matched normal DNA samples, as well as plasma cfDNA collected at the time of surgery. We sequenced the normal and tumour libraries by whole-exome sequencing (WES), and preformed targeted resequencing of cfDNA, utilizing the NCGC 900 cancer gene panel.

Result: We identified somatic mutations in the tumours from 14 high-grade soft-tissue sarcomas. We identified 288 somatic mutations in 6 tumour—plasma pairs, including damaging mutations in *TP53*, *RB1*, *TSC1*, *NRAS*, *MTOR*, *MAP3K4*, *ERBB2*, *SETD2*, and *ARID1B.*

Conclusion: Our results suggest that the detection of somatic mutations in cfDNA of STS patients is feasible. In addition, our initial results indicate that the detection of tumour heterogeneity is plausible. Our findings may be translated into the clinical setting for prognostic or predictive purposes in STS patients.

Acknowledgements

The work presented in here is the result of Master project carried out during Jan 2015-Jun 2016 at the Department of Tumour Biology, Institute for Cancer Research, Oslo University Hospital.

I would like to express my special thanks and acknowledgment to Dr. Heidi Maria Namløs, who walked me every steps along this path, not only for sharing the best of your knowledge and experiences with me, but your patience, kindness, assistance, positive and motivating personality. Thank you for having always the time to sit and discuss my issues, and untangling my difficulties with your smart ideas. As well, thank you for introducing me to beautiful Norwegian culture and trying to teach me the language, although I believe I still need to work on that for a little while.

I would like to thank my main supervisor Dr. Leonardo A. Meza-Zepeda for his significant knowledge, and great personality. Thank you for introducing me to the field of your interest and expertise.

Thank to both Dr. Meza-Zepeda and Dr. Namløs for challenging my capabilities and trusting in my abilities. I appreciate all excellent guides and resources that you provided me.

I would like to express my gratitude to Professor Ola Myklebost, whom his endeavours and significant lifetime achievements have laid the foundation of a unique group. I truly appreciate your trust and believe in me.

My colleagues and group members, especially Dr. Stine H. Kresse, Dr. Tale Barøy and Lars B. Aasheim, I want to thank them all, who were always helpful and created a friendly and interactive learning environment, and contributed to my educational success. Thank you for making delicious cakes all the time. My classmate and fellow master student at the department Wafa and all my friends, particularly Eta, Kjetil, Farzad and all who shared many best moments with me.

My parents, which I do believe without their motivation and support I was not at the place where I currently stand. I cannot thank you in a way that deserves you, but I hope my endeavours in life make you proud of your son.

Finally I would to appreciate the patients who have contributed to our understanding of cancer, and the great opportunity of study in Norway, offered by University of Oslo.

Hossein Moosavi
May 17th, 2016

Abbreviations

g	Gravity
µL	Microliter
µM	Micrometer
AA	Amino acid
AF	Allele frequency
bp	Base pair
cfDNA	Circulating Cell-Free DNA
ctDNA	Circulating Tumour DNA
Chr	Chromosome
CNA	Copy number aberration
CTC	Circulating tumour cell
ddPCR	Digital droplet polymerase chain reaction
DNA	Deoxyribonucleic acid
DP	Depth of read (coverage)
dsDNA	Double-stranded DNA
gDNA	Genomic DNA
IGV	Integrative Genomics Viewer
InDel	Insertion/Deletion
Kb	Kilobase
LMS	Leiomyosarcoma
miRNA	MicroRNA
NGS	Next-generation sequencing
nM	Nanomolar
ng	Nanogram
RNA	Ribonucleic acid
PCR	Polymerase chance reaction
PM	Personalized medicine
qRT-PCR	Quantitative real-time polymerase chain reaction
SBS	Sequencing-by-synthesis
SNV	Single nucleotide variation
SS	Synovial sarcoma
ssDNA	Single-stranded DNA
STDEV(SD)	Standard deviation
TKI	Tyrosine kinase inhibitor
UPS	Undifferentiated spindle/pleomorphic sarcoma
UV	Ultraviolet
WES	Whole-exome sequencing
WGS	Whole-genome sequencing

Table of contents

1.Introduction

1.1. An introduction to cancer

The term cancer refers to a collection of related diseases and is characterized by uncontrolled growth of abnormal cells (1). The fact that virtually all human cancers emerge from sequential aberrations in a common and defined set of critical genes and pathways and evolve over the a time timeframe of 20 to 30 years (2). Cancer is the leading cause of morbidity and mortality worldwide. In 2012, approximately 14 million new cases and 8.2 million cancer-related deaths were identified (3). According to the World Cancer Report, the annual global incidence of cancer is expected to increase to 19.3 million cases by 2025(3), with the global population growth as a critical determinants, and major changes in the lifestyle and diet as predicted players that can contribute the estimation (3).

The worldwide cancer incidence is highest for men, which most frequently develop lung, prostate, colorectal and stomach cancer. Breast, colorectal, lung and cervix are among cancers with the highest incidence in women. The highest mortality rate worldwide has been reported for lung, breast, liver, stomach, and colorectal cancers (3). In Norway, prostate, colorectal, breast, and lung are the four most common types of cancer. In Figure 1, a graphical overview of the age-standardised rate (ASR) of incidence and mortality of common cancers in Norway is shown.

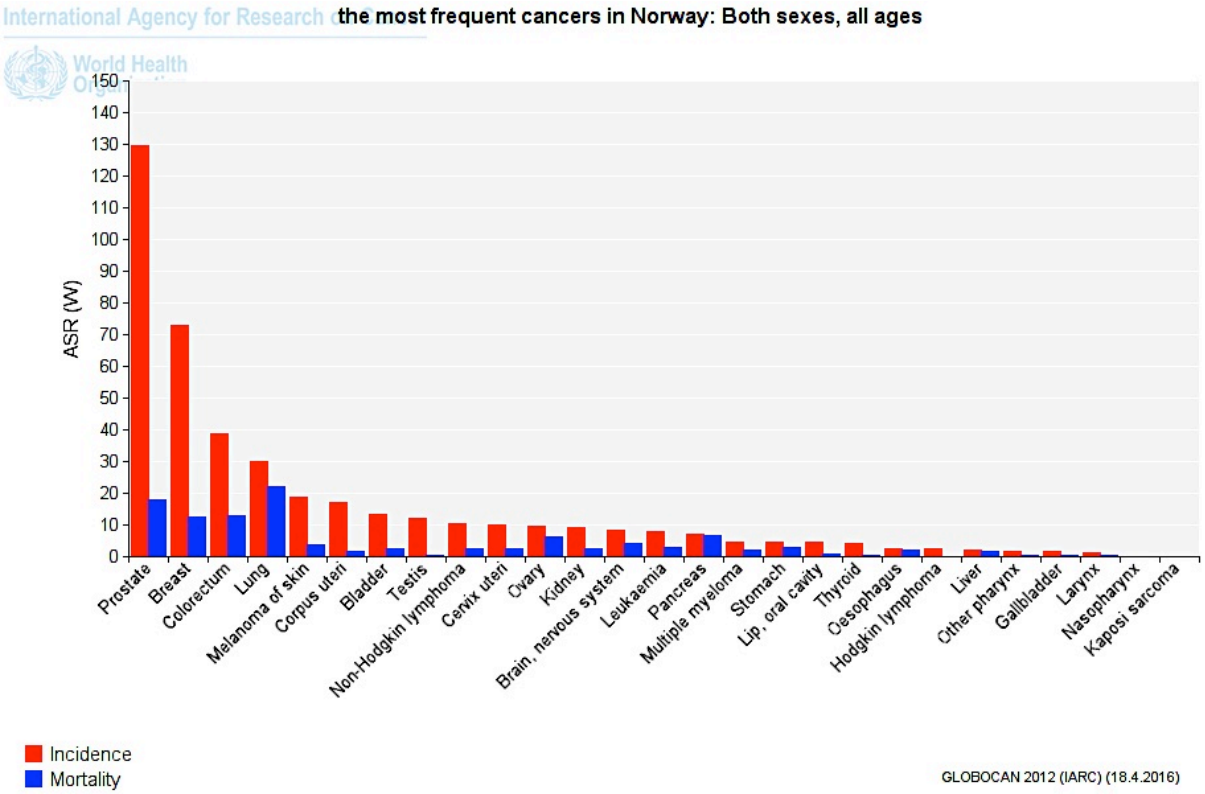
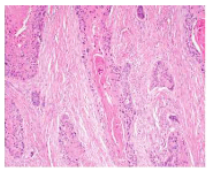
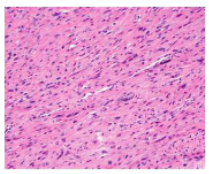
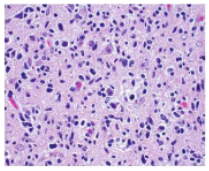
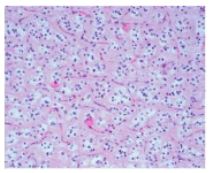
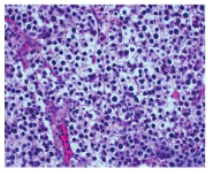


Figure 1. The number of new cases and deaths of common cancers, per 100,000 persons per year in Norway. The ASR is a weighted mean of the age-specific rate; the weights are taken from population distribution of the world standard population. GLOBOCAN 2012[1]

It has been demonstrated that cancer-related mortality can be decreased if two principles, screening and early diagnosis, in the healthcare management will be carefully fulfilled. It has been estimated that between 5—10% of all cancer cases have their root in hereditary genetic defects, whereas aetiology of the majority of cancers are ascribed to environmental risk factors. High-risk environmental factors include diet, smoking, alcohol consumption, environmental pollutants, stress, infections, obesity, and physical inactivity. Therefore, cessation and/or minimization of these factors can play a major role in cancer prevention [reviewed in (3, 4)]. A growing body of research have increased our understanding of certain environmental risk factors, such as smoking and UV exposure, that can jeopardize the function of cells and can significantly contribute to the risk of developing a lung and/or a melanoma cancer, for instances [reviewed in (2, 4)].

[1] Ferlay J, Soerjomataram I, Ervik M, Dikshit R, Eser S, Mathers C, Rebelo M, Parkin DM, Forman D, Bray, F. GLOBOCAN 2012 v1.0, Cancer Incidence and Mortality Worldwide: IARC CancerBase No. 11 [Internet]. Lyon, France: International Agency for Research on Cancer; 2013. Available from: http://globocan.iarc.fr, accessed on 10 Apr 2016.

Cancer, at its core, arises from an uncontrolled and rapid cells' proliferation that results in the abnormal growth of tissue, an especial characteristic of cancer known as "neoplasm". As a result of this consecutive excessive proliferation, tumours may form. Tumours can primarily be classified as "malignant" and "benign", with only malignant tumours being considered to have a cancerous nature. Dissemination of cancer cells into the blood stream may lead to invasion of cancer to adjacent and/or distal tissues, resulting in metastatic outgrowths [reviewed in (2, 5)]. Benign tumours, however, are without capability of invasion, and once they are removed they do not grow back. Depending on the tissue of origin, benign neoplasms can be classified in different categories, e.g. adenoma, fibroma or lipoma. Malignant neoplasms that form solid and hematopoietic tumours can be classified as carcinomas, sarcomas and lymphomas, non-solid tumours of the blood cells. The classification of solid tumours based on their tissue-of-origin has been shown in Figure 2. It is noteworthy to mention that the common cancers, which represented in Figure 1, commonly originate from epithelial tissues, thus are categorized within the carcinoma tumour type.

Solid Tumor Type	Image	Cell/Tissue of Origin	Benign Tumors Examples	Malignant Tumors Examples
Epithelial tumors		Squamous epithelium Transitional epithelium Glandular/ductal epithelium Neuroendocrine cells Internal organ specific Liver cell Kidney cell	Epithelioma (papilloma) Transitional cell papilloma Adenoma Carcinoid Hepatoma Renal cell adenoma	Squamous cell carcinoma Transitional cell carcinoma Adenocarcinoma Small cell carcinoma Hepatocellular carcinoma Renal cell carcinoma
Mesenchymal tumors		Fibroblast Fat cell Blood vessels Smooth muscle cell Striated muscle cell Cartilage Bone cell	Fibroma Lipoma Hemangioma Leiomyoma Rhabdomyoma Chondroma Osteoma	Fibrosarcoma Liposarcoma Angiosarcoma Leiomyosarcoma Rhabdomyosarcoma Chondrosarcoma Osteosarcoma
Tumors of neural cell precursors		Neuroblast	Ganglioneuroma	Neuroblastoma
Tumors of glial cells and neural supporting cells		Glial cells Meningeal cells Schwann cells	— Meningioma Schwannoma	Glioma Malignant meningioma Malignant schwannoma
Germ cell tumors		Embryonic cells	Teratoma	Embryonal carcinoma Teratocarcinoma Seminoma/dysgerminoma

Solid tumors are classified according to the tissue of origin. Usually, but not always, malignant tumors have names ending in "carcinoma" or "sarcoma."

Adapted from Damjanov, 2006.

Figure 2. Histological classifications of solid tumours are shown. Pathological investigation on tumour tissue has important diagnostic value in clinic. Adapted from Ventra Medical System[2]. Solid tumors: Principles and perspectives. Level 2.

1.1.1. Hallmarks of cancer

Scientists today have a profound understanding of the biology and underlying causes of a vast array of cancers. With the advent of new technologies and consequent wealth of information from the various fields in basic research on cancers, a broader and deeper insight of cancer biology has been achieved. In the light of these advances, important key molecular mechanisms underlying different types of cancer have been uncovered (2, 6).

[2] *Accessed online at http://www.ventana.com/_resources/swf/training/SolidTumors/files/assets/basic-html/toc.html*

Historically, across several altered functions within transformed cancer cells that were suggested as the cancer-initiating traits, six described distinctive and complementary characteristics were highlighted as the major cancer hallmarks that dictate malignant growth and proliferation pattern of the cancer cells and are manifested in almost every cancerous cells (1). Hanahan et al. were first to describe these hallmarks as acquired functional capabilities that allow cancer cells to survive, proliferate, and disseminate. These core hallmarks are evading programmed cell death (apoptosis), self-sufficiency in growth signals, evading growth-inhibition, limitless replicative potential, sustained tissue invasion and metastasis, and inducing angiogenesis. These functions are acquired in different tumour types via distinct mechanisms at various time points during tumorigenesis and well describe the diversity of neoplastic diseases (1, 7, 8). However, an increasing body of evidence by more recent studies have proposed additional, emerging hallmark features that may facilitate the development and progression of cancer(9). Emerging hallmarks, "evading immune destruction" and "reprograming of cellular energetics" underscore continuous support of cell growth and proliferation by alternation in cellular energy metabolism function, and evasion of cancer cells from immune system mediated elimination. Enabling characteristics, "genome instability and mutability" and "tumour-promoting inflammation", are considered as potential underlying molecular events contributing to acquisition of both core and emerging hallmarks. Genomic alternations, especially rare mutations among them, can govern hallmark capabilities. Moreover, inflammatory responses caused by innate immune cells can serve cancer cells to promote tumour formation and progression (9-12). A schematic representation of all discovered hallmarks of cancer is shown in Figure 3.

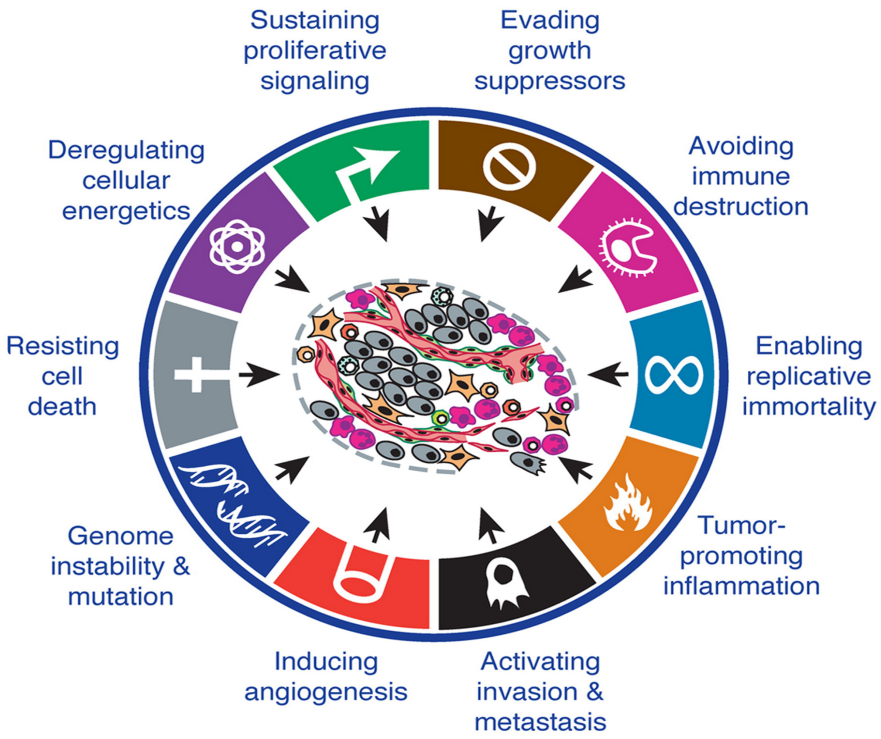


Figure 3. Hallmarks of cancer are shown. The hallmarks are nesseary for tumour growth and progression.Adapted from Hanahan D, Weinberg RA. Cell. 2011;144:646-674.

1.1.1.1. Genome instability

Despite the increased understanding of cancers at the genomic and epigenomic level, the evolutionary mechanisms recruited by tumours are not fully understood for most of cancer types (13, 14). Both genetic and epigenetic factors can affect cellular pathways (2, 9, 15, 16). The fact that cancer originates from a genetically normal cell, gaining alternations at the genomic and epigenomic level, resulting in hundreds of thousands of divergent cells by the mutational landscape, is widely accepted among the scientific community. Many of these mutations can have deleterious impact on the cells function but do not possess the ability to initiate cancer development. Such mutations are classified as “passenger” mutations. A sequential fraction of approximately 140 mutations are referred to as the cancer “driver”, (2) and are not only deleterious for the cellular function but also provide selective growth advantage that promotes tumorigenesis and subsequently cancer development. With the advent of large-scale, systematic genomic studies conducted on cancer genomes many cancer driver genes across various cancer subtypes has been revealed (17, 18). The most frequently

mutated cancer critical genes across 28 various cancer types are shown in Figure 4(19).

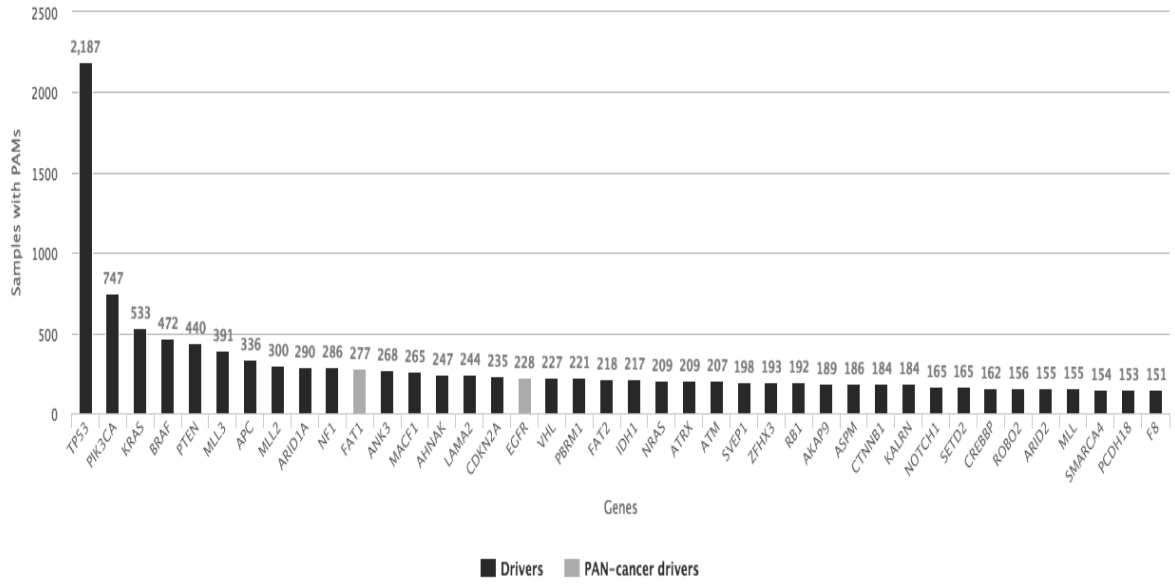


Figure 4. The analysis of 6,792 samples has revealed the most frequently mutated genes across 28 various cancers. PAMs on the vertical axis indicates protein-affecting mutations in the plotted genes. The dataset retrieved from the IntOGen database.

The cause of mutations has been ascribed to both endogenous and exogenous mutagens. Endogenous mechanisms such as infidelity of the DNA polymerase during replication mechanism, defective DNA repair machinery, and the enzymatic modifications of DNA, whereas UV light and inhaled tobacco smoke are instances of the most known carcinogens that well correlates with the increased risk of mutagenesis in normal cells [reviewed in (2, 4, 16)].

Across cancer of different types, 3 groups of genes are virtually always mutated—proto-oncogenes, tumour suppressor genes, and genes responsible for genome repair. Genes within these groups are particularly important in the regulation of cell essential pathways such as cell cycle and programmed cell death. Mutations in these genes can cause gain or loss of function, thereby driving the malignant growth and proliferation of the cells [reviewed in (2)]. Herein, to clarify the importance and consequences of mutations in these genes, first the definition of various mutations will be further explained.

Classically, mutations are broadly classified to germline and somatic mutations. The germline mutations are inherited and occurred during embryogenesis, and therefore are present in virtually every cell within the individual, whereas somatic mutations are acquired mutations present only in a subset of cells, a phenomenon commonly seen in the cancer cells [reviewed in (2)]. Approximately 90% of the mutations in cancers are due to acquired somatic alternations and therefor are not present in the normal tissues [reviewed in (2, 4, 20)]. Nearly 90% of the known somatic mutations in cancer genes have dominant effect at the cellular level, meaning that one mutation in only one allele is sufficient to initiate cancer development [reviewed in (5)].

Mutations may occur within regions of the genes that code essential proteins for normal cellular function, or may happen within regions that do not have directly impact on the function of the produced proteins. In this respect, region of the genes with essential protein-coding function, known as exons, encompass the majority of critical alternations observed in cancers, while mutations in the non-coding region sequence of the genes e.g. introns and intergenic regions, may indirectly affect the protein production of the genes by affecting regulatory elements (21).

The term "exome" refers to the complete set of protein-coding regions and approximately encompasses 1-2 % of the human genome. Mutations from another perspective are divided into point mutations, namely single nucleotide variation (SNV), insertions and deletions (Indels), and chromosome structural and numerical aberrations [reviewed in (5, 22)].

The point mutations are fundamentally of two types; transitions and transversions that occur at different rates, with rate of transition substitutions predominantly higher than transversions. Different impacts on the function of genes by these mutations have led to classification into A) *missense mutation*, which results in the incorporation of a different amino acid in the protein made by the same gene. B) *Nonsense mutation,* resulting in a shortened polypeptide due to the introduction of a premature termination codon. The resulting polypeptide may have no function or an improper activity.. C) *Silent mutation* (synonymous), where the mutation does not result in an amino acid change, due to the codon degeneracy. The effects of these changes varies in

the way they may influence e.g. protein folding, destabilize protein binding and/or active sites conformation [reviewed in (22)].

Indels refer to the insertions or deletions of one or few bases from the sequence. Depending on the number of bases inserted or removed, indels can be grouped into A) *Frameshift mutations,* if the number of nucleotides inserted or deleted is not multiple of three, which cause a change in the reading frame of transcript, and result in a completely different, or non-functional protein B) *in frame mutations,* in which the number of nucleotides inserted or deleted is a multiple of three, resulting in addition, deletion, or change of amino acid(s) in the protein. The *in frame* mutations may transform, enhance or reduce the activity of normal protein [reviewed in (22)]. Based on given information, both *nonsense* and *frameshift mutations* are considered to significantly impact protein function [reviewed in (2, 22)]. The chromosome aberrations refer to large-scale changes where a large area of gene or chromosome is affected. These changes include inversions, translocations, duplications, and numerical changes of the chromosomes, which can result in copy number aberration (CNA), which consequently can affect the level of gene expression. Moreover, chromosomal rearrangements play the key role in activation of human oncogenes and inactivation of tumour suppressors genes, a phenomenon observed in various cancers (23). Mutations that result in loss of function of tumour suppressor genes or gain of function in the proto-oncogenes conclude to oncogenic activity of the genes and significantly contribute to initiation and progression of the cancers. The protein product of these groups of genes are in control of cell proliferation, growth and apoptosis, and can be broadly divided into transcription factors, chromatin remodellers, growth factors, growth factor receptors, signal transducers, and apoptosis regulators (24).

Instances of well-known human proto-oncogenes where their oncogenic activity have been demonstrated in several cancers are *RAS*, *ERBB2*, *MYC* and *CCND1*. The oncogenic activity of known tumour suppressor genes such as *TP53, RB1, APC, NF1, PTEN and CDKN2A* due to loss of function mutations have been frequently observed across various cancer types (24) [reviewed in (5)].

The third group of genes with critical role in cancers that are commonly mutated in many cancers are DNA repair genes. The cellular DNA is continuously under attack by cellular and/or environmental mutagens as previously mentioned. Defects in the genome maintenance mechanisms, including base excision repair (BER), nucleotide excision repair (NER), transcription coupled repair (TCR), mismatch repair (MMR) and double strand break repair (DSBR) can potentially increase rate of spontaneous mutations capable of tumour development initiation during DNA replication event (20, 25). Enumerated mutations in currently 150 known human DNA repair genes may greatly affect genome integrity (25).

Deregulation of cancer's essential genes can also be triggered by epigenetic factors, which are defined as heritable changes in genome without altering the DNA sequence, thereby affecting the expression pattern of the genes (26). Disruption in the global methylation pattern, such as hypermethylation of the CpG islands in the promoter regions of tumour suppressor genes, can contribute to the development of cancer (26, 27). Furthermore, hypomethylation, *i.e.* a low level of methylation, has been observed in tumour tissue compared to corresponding matched normal tissue. The low level of methylation is believed to promote chromosomal rearrangements and activation of the transposable elements, which in consequence affects genomic instability.

1.1.2. Tumour heterogeneity and clonal expansion

Tumours, even of same histopathological subtype, exhibit genetic heterogeneity, a phenomenon known as intra- and inter-tumour heterogeneity. Differences in mutational spectrum between the primary tumour and metastatic lesions, between different metastatic lesions, and even within a single tumour have been described. The tumour heterogeneity was initially attributed to morphological and epigenetic plasticity, however, associations to genetic factors, particularly genetic instability, which explain existence of genetically divergent intra-tumour clones have been demonstrated [reviewed in (20, 28)]. More recently, a broader biological definition that encompasses all contributors to the development of cancer at genomic and epigenomic levels has been suggested as "cancer driver" in which a cell-autonomous or non–cell–autonomous alteration that can contribute to tumour evolution at any stage by promoting any of the cancer hallmarks (13). These alternations may endow

tumours the ability to pave various evolutionary directions, from early initiation to full-grown metastatic tumours, in the time line of cancers' development (13).

Tumour heterogeneity is also believed to play a role in acquired drug resistance in the tumours [reviewed in (20, 29)]. To further explain observed heterogeneity in the tumours, three paradigms have been suggested. In each of the hypothetical paradigms, potential factors that can contribute to heterogeneity of the tumour during progression of cancer have been demonstrated [reviewed in (20)].

The clonal evolution model originally described by Nowell in 1976, suggests expansion of one (monoclonal) or multiple (polyclonal) subpopulations during the event of a tumour's evolution, influenced by Darwinian selection forces. As a consequence, a fraction of cells within the tumour with different heritable traits are privileged (Figure 5A). In the second paradigm, the cancer stem cell paradigm, tumour heterogeneity is ascribed to either single or multiple mutations in the progenitor stem cells within the tumours, regardless of heritable traits (Figure 5B). This concept has been elucidated by different studies, the principle of cancer stem cell, specifically in the solid tumours, as the results of experiments vary considerably. In the third model, high diversity of tumour cells is attributed to random and incremental accumulation of mutations during tumour progression, known as the mutator phenotypes (Figure 5C) [reviewed in (30)].

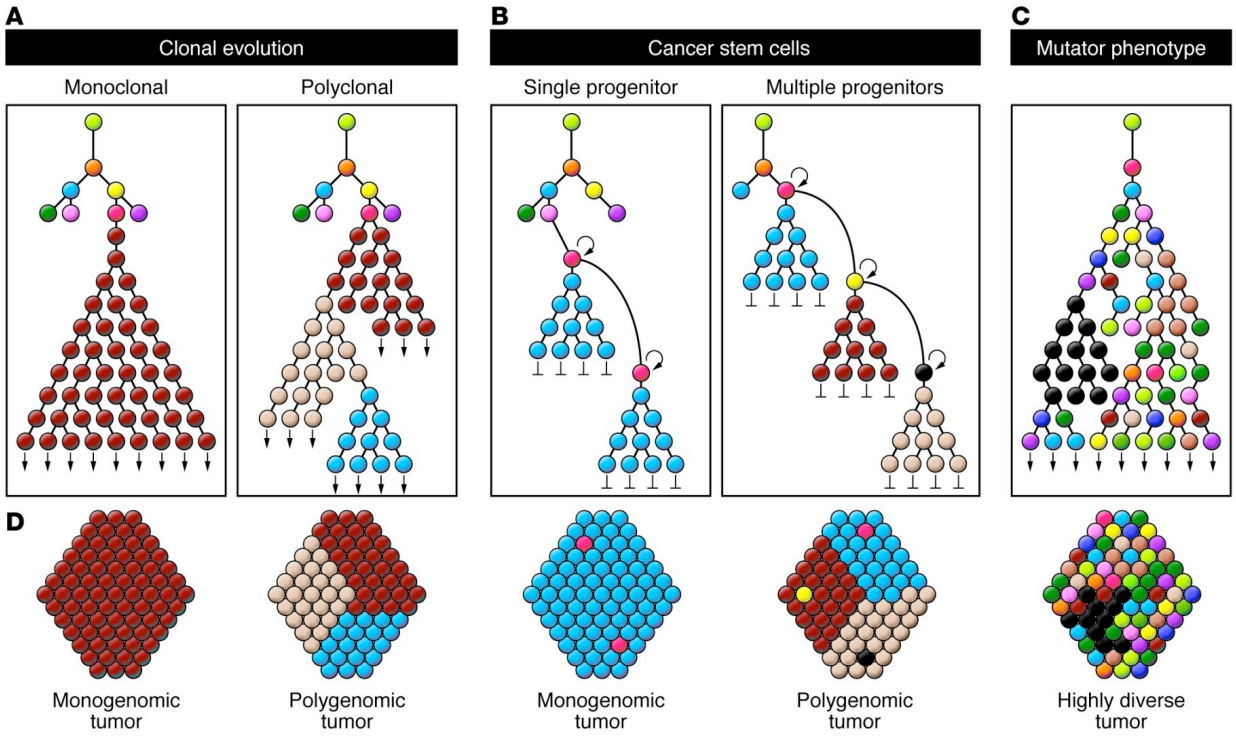


Figure 5. The hypothetical paradigm, underling tumour heterogeneity is shown (A–C). Different models for tumour evolution can give rise to distinct types of intra-tumour heterogeneity, typified here by: A) clonal evolution, B) cancer stem cell, and D) mutator phenotype models. (D) The different evolutionary paths result in several subpopulations with distinct spatial distributions. Adapted from Russnes et al., J Clin Invest. 2011;121(10):3810-3818. doi:10.1172/JCI57088.

As a consequence of all enumerated mechanisms above, subpopulations within virtually all tumours always have different genetic and epigenetic changes, resulting in distinct and heterogeneous subpopulations of cells within tumours of same cancer type in each individual [reviewed in (30)].

1.2. Sarcomas

1.2.1. Overview of sarcomas

Sarcomas are an infrequent heterogeneous group of malignant tumours of mainly mesenchymal origin, accounting for approximately 1 % of all adults and nearly 21% of all paediatric solid malignant cancers. Soft-tissue sarcomas (STS) compose the vast

majority of sarcomas at the time of diagnosis (31). In Norway, approximately 300 new cases of soft tissue sarcomas are diagnosed every year[3].

The term "sarcoma" is Greek and means fleshy ("sarcos") and tumour ("oma") [reviewed in (32)]. Sarcomas are primarily grouped into two main subtypes; soft-tissue and bone sarcomas. Sarcomas mainly develop as sporadic cases and their aetiology is mainly unknown. However, several hereditary genetic and environmental risk factors are demonstrated to associate with the development of sarcomas. Individuals with germline mutation in tumour suppressor genes such as *RB1*, *TP53, NF1,* and *APC* are shown to be considerably susceptible to develop sarcoma. The striking example of hereditary genetic risk factor that highly correlates with sarcoma development is Li-Fraumeni syndrome. Individuals with Li-Fraumeni syndrome are born with one defective copy of *TP53.* Overall, it has been demonstrated that *TP53* aberrations occurs in about 60% of all sarcomas (33).

From a cytogenetic standpoint, sarcomas can be broadly categorized into two groups. The first category includes subtypes that manifest relatively simple karyotype, encompassing up to 50% of all sarcomas, and commonly represent disease-specific chromosomal translocations, resulting in fusion genes that commonly encode for defective transcription factors, and a smaller group of fusion proteins with known deterrent consequences in the cell growth and differentiation. Examples of sarcomas within this category are Ewing's sarcoma, clear-cell sarcoma, synovial sarcoma (SS), and desmoplastic small round cell tumour. Commonly participated genes in translocations of these subtypes, for some prominent instances, are *EWSR1, FOXO1A, DDIT3*, *SYT, SSX, PAX,* and *ALK.* Overall, 50% of the affected genes by the translocations in this category belong to the *TET* family of transcriptional regulators. It has been demonstrated that about 20 % of the subtypes in this group have activating and/or inactivating oncogenic point mutations. Gastrointestinal stromal tumour (GIST) subtype is the prime example of this group in which aberrations in the *KIT*, *PDGFRA* and rarely *BRAF* proto-oncogenes play a major role in the pathology of the cancer. Other known instances of oncogenic mutations in sarcomas include gain of function

[3] http://www.sarkom.no/uploads/Arsrapport2013-2015.pdf

mutations in *PIK3CA* and loss of function mutations in *TSC1/TSC2* and *INI1* tumour suppressor genes (33, 34). Gene amplification events encompasses ~15% of this group, and the prime instances of genes involved in these events are murine double minute gene (*MDM2*) and cyclin-dependent kinase 4 (*CDK4*) with function in the regulation cell cycle [Reviewed in (32, 35)].

The second group is characterized by aneuploid karyotypes with complex structural defects with several rearrangements, deletions, and duplication. Sarcomas such as leiomyosarcoma (LMS), osteosarcoma and undifferentiated spindle/pleomorphic sarcoma are categorized in this group. The subtypes within this category do not manifest determinant alternations. However, presence of multiple amplifications in combination with several deletions in chromosomes that results in inactivation of tumour suppressor genes such as *RB1, TP53,* and *PTEN* has been demonstrated (34, 36).

Critical function of *TP53* and *RB1* in regulation of the G1/S transition during the cell cycle and initiation of the DNA damage-mediated responses that results to activation of repair mechanisms and/or apoptosis has been demonstrated, and is thought to be one of the major cause of various genetic alternation observed in this group. The non-homologous end-joining DNA repair pathway has also been demonstrated to play a role in development of sarcomas with complex karyotype(32).

Concordant with many other cancer types, constitutive and/or perturbed activation of essential growth factor signal transduction pathways such as PI3K/AKT and MAPK/ERK, and mammalian target of rapamycin (mTOR) have been also demonstrated in many sarcomas. In this respect, the structural changes of receptor tyrosine kinases (RTKs), including IGF-1R, C-KIT, C-MET, VEGFR-A, and FGFR, have been demonstrated to play a major role in the pathogenesis and progression of many sarcomas by contributing to cellular transformation (32, 37, 38).

1.2.2. Soft Tissue sarcoma

The World Health Organization has classified at least 50 distinct STS subtypes, based on their histological and morphological appearances [reviewed in (32, 39)]. STS tumours are most common in the muscles, cartilage, nerves, fat, and blood vessels. Lower and upper extremities together are the most common anatomic parts affected by STS. Retroperitoneal and visceral are another major sites influenced by STS (Figure 6) [reviewed in (39)]. The most common histologic subtypes of adolescent STS are as follow; undifferentiated spindle/pleomorphic sarcoma (UPS); liposarcoma; leiomyosarcoma (LMS); synovial sarcoma (SS); malignant peripheral nerve sheath tumour (MPNST); rhabdomyosarcoma (RMS), within which undifferentiated sarcomas and liposarcoma accounts for up to 45 % of all STS. In childhood STS, rhabdomyosarcoma, neuroblastoma, and the Ewing family are most common STS [reviewed in (39)].

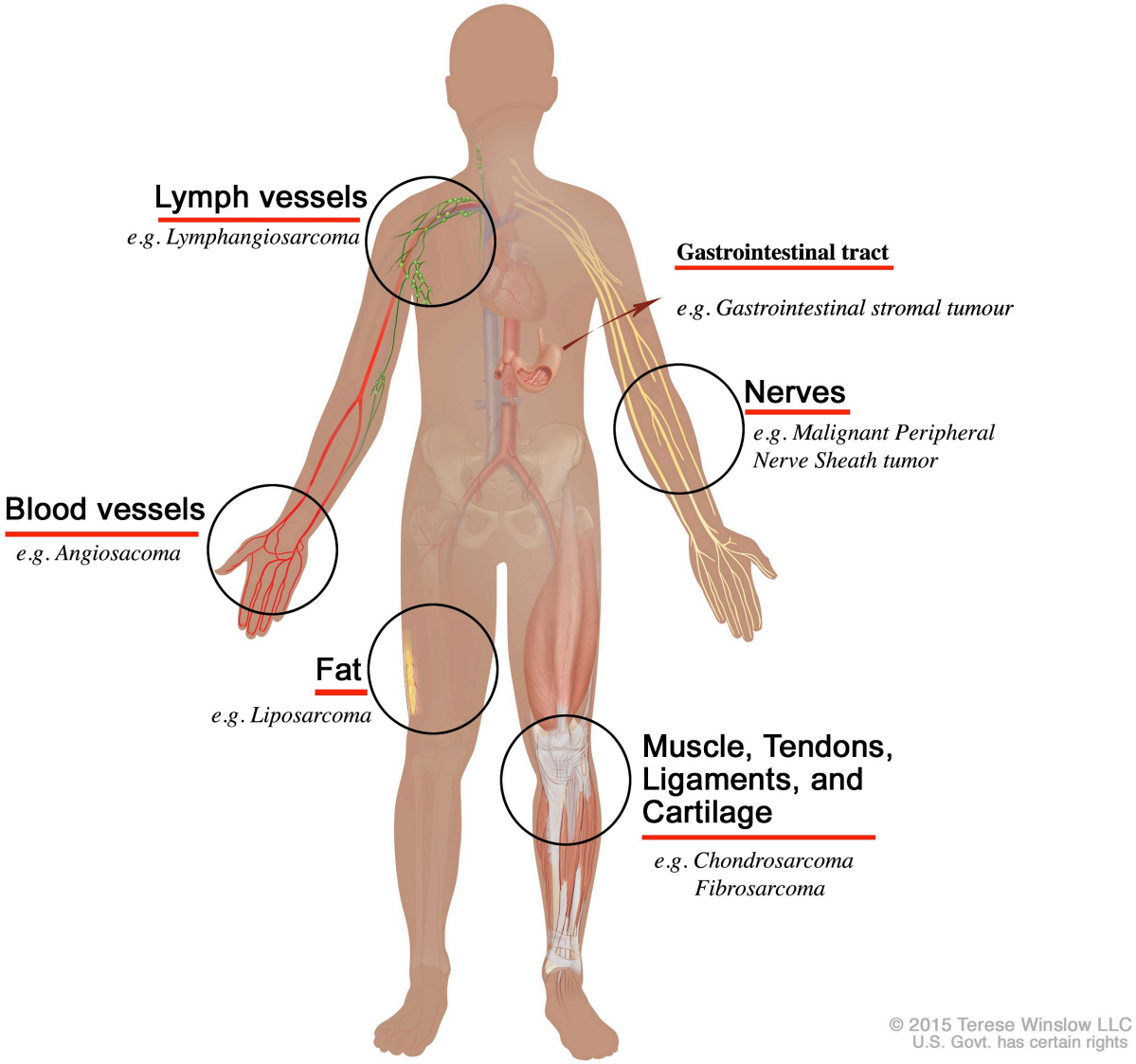


Figure6. Anatomic sites and tissues that soft-tissue sarcomas frequently occur are illustrated. Adapted from www.cancer.gov

Available guidelines in the management of STS vary in the criteria for assigning the tumours' grade and stage in STS. However, in assigning a grade score to a tumour the primacy of parameters such as necrosis, tumour differentiation, and mitotic activity are commonly emphasized. Moreover, metastasis at the time of diagnosis, tumour size, grade and location are poor prognostic determinants. Almost 75 % of soft tissue sarcomas are highly malignant [reviewed in (39)]. After treatment 11 % to 14 % of STS relapse locally, while 18% to 50% develop metastases [reviewed in (40, 41)]. The increased frequency of distant metastases and tumour size in high-grade STS has shown to strongly correlate. The most relapses are proposed to occur within five years.

The median time of 12 to 18 months for local recurrence and 12 months for distant recurrences has been demonstrated. The five-year survival rate in the range of 44 % to 91% for localized sarcomas and approximately 50% in patients with metastasis, with differences inferring from histology, grade of malignancy, and tumour size of the disease have been indicated [reviewed in (31, 40)]. Noteworthy, metastasis is the indicative of survival in patients with STS, and despite detection an elimination of metastasis, only a few patients may reveal prolonged remission. Given information on the lower survival rate in metastatic, high-grade sarcomas after diagnosis of metastases underpins the importance of early detection of recurrence as a pivotal criteria in the survival of the patients [reviewed in (40, 42)].

1.2.3. Current therapies and limitations in soft-tissue sarcomas

Accurate diagnosis of sarcomas is challenging, given their rarity and wide histological diversity [reviewed in (39)]. Although incremental developments in patient management guidelines and technological advances have resulted in better outcome of patients with sarcomas, early diagnosis is still an unmet issue.

To date, a multimodal protocol composed of complete surgical excision of tumour liaison in combination with adjuvant and neoadjuvant radiation and chemotherapy, based on tumour size, location, and aggressiveness is widely used therapeutic modality [reviewed in (39, 42)]. Nevertheless, it has been estimated that half of treated STS patients develop recurrence [reviewed in (40)]. This highlights the primacy of accurate diagnosis in determining the prognosis of the STS, particularly risk of metastasis [reviewed in (40, 41)].

Pre- and postoperative radiation and/or chemotherapy have been demonstrated useful to avoid local recurrence and reducing the risk of developing systemic metastasis after complete excision of tumour lesions. These goals have been achieved in specific tumour types such as Ewing sarcoma and rhabdomyosarcoma, however, the efficacy of these agents are still a matter of debate in many subtypes [reviewed in (41-43)]. Of note, it is widely accepted that among exogenous agents, ionizing radiation has been reported as the most common risk factor for development of often high-grade tumour with poor prognosis (44). Particularly, evident increased risk of both childhood bone and soft-tissue sarcomas after exposure to high-dose of radiation has been observed

(44, 45). Furthermore, resistance to radiation- and chemotherapy-induced cell death often occurs in cancer cells, including sarcomas (46, 47). Current surveillance strategies mostly relay on imaging methods, which lack optimal sensitivity to identify local recurrence after treatment [reviewed in (39)]. Delay in diagnosis greatly affects the outcome of patients with STS, therefore preventing and/or early identification of metastasis are cornerstone of sarcoma management [reviewed in (32, 38)]. Uncontrolled microscopic and/or systemic microscopic diseases are the major risk to life of sarcoma patients [reviewed in (40)].

If the cancer progress, due to latency in diagnosis, conventional treatment of choice for advanced STS, excluding GIST, is monotherapy by an anthracycline, mainly doxorubicin, or in combination with ifosfamide (48). However, administration of current chemotherapeutic treatments in unresectable metastatic STS has rather palliative intent (49, 50). Preoperative chemotherapy theoretically has several advantages on some specific histological subset of localized sarcoma tumours, however the impact of adjuvant chemotherapy on overall survival of patients with high-grade soft-tissue sarcomas is still unclear [reviewed in (42, 51)].

1.2.4. Targeted molecular therapy

There is a need of more cancer-specific therapies for non-responsive cancers. A precise classification of patients based on the molecular characteristics of cancer may result in optimal drug efficacy and usage, consequently making it possible to decrease drug cytotoxicity on non-responsive patients. To address the concern of individualized treatment strategies, the emerging field of personalized medicine (PM), a joint effort of medicine, biology, genetics, and computer sciences has been introduced. The current description of PM by the National Institutes of Health is “the science of individualized prevention and therapy” (52). With the fundamental concept of “one size does not fit all”, PM is attempting to translate tangible discoveries of cancer genomics to the “beside”.

The immediate goal of PM is to optimally customize preventive measures and therapeutic approaches, utilizing the individual’s clinical, genetic and genomic information while patients are still in the earlier stages of the disease (6, 52). Attempts

in the field of PM have considerably shortened the timeline between the discovery of disease-causing mutation and the development of effective drugs (53).

Considering sarcoma, efforts within investigation of novel systemic treatments against specific molecular targets have led to deeper insight of STS pathogenesis and brought new benefits for patients suffering from the cancer. With the growing amount of information about fundamental underlying molecular mechanisms with critical function in pathogenesis of sarcomas, targeted drugs have been developed to specifically target underlying deterrent mutations, whereas in the conventional treatment many subtypes were given unspecific treatment regardless of molecular aberrations in each patient. In STS much of the attempts in treatment of STS are particularly focused on tyrosine kinase inhibitors (TKIs) (Figure 7), due to few known causative oncogenic mutations described above (54).

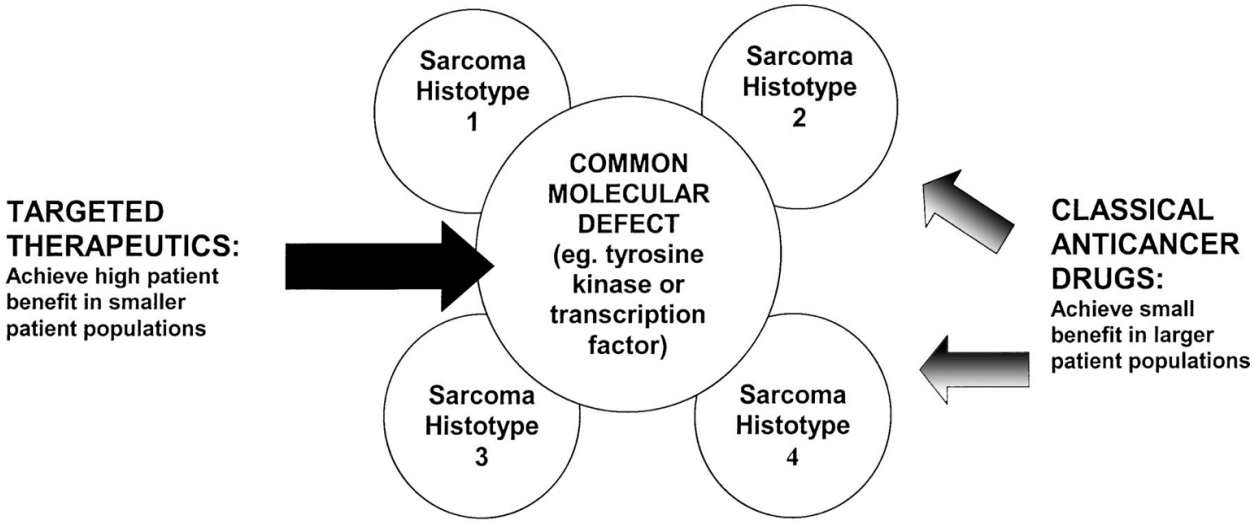


Figure7. The mechanism of action in the targeted therapeutics versus classical anticancer drugs is shown. Targeted drugs are designed specifically based on the underlying cancer driver mutation not tumours' subtypes, in contrast to non-specific drugs that underlying molecular mechanisms of individual tumours are not considered. Adapted from Borden et al., Clin Cancer Res. 2003 Jun;9(6):1941-56

The TKI imatinib mesylate is considered the first-line treatment for advanced or metastatic gastrointestinal stromal tumour (GIST) bearing mutation in exon 11 of *c-KIT*, which alone accounts for approximately 70 % of cases, with response rate of more than 60%. Another frequently mutated exon in GIST tumours has also been discovered in exon 9 and 13 of *c-KIT* and *PDGFRA,* which both are cell surface tyrosine kinase receptors. Imatinib has also been suggested to be effective in targeting fusion gene

caused by a translocation between *COL1A1 and PDGFB* in dermatofibrosarcoma protuberans (DFSP) and the related giant-cell fibrosarcoma. Classification of this kind, based on the underling driver mutation, has led to utilization of other targeted drugs in patients who were predicted to not respond to the imatinib. Prime example of this classification is development of another targeted drug, sunitinib malate, which now is a second-treatment option for a fraction of GIST patients with mutated *c-KIT* but not responsive to imatinib, and also another subset of advanced GIST patients, bearing mutation in *PDGFRA* gene (55-58). Another example of benefit of targeted drugs developed for STS is a multi-targeted TKI, pazopanib, which has been developed for a particular metastatic subtype of STS patients who failed to response to standard chemotherapy (59).

The information given above were just few examples of the benefits of targeted therapeutics that can be designed with more specificity and less cytotoxicity as the underlying molecular mechanisms that contributed to the development and progression of cancers become more and more uncovered (53, 60).

1.3. New strategies in the management of cancer patients

1.3.1. The introduction of sequencing technologies

The Human Genome Project (HGP) was a 13-year-long project that successfully completed in 2003 at the cost of approximately US$3 billion. With the advantage of utilizing the very first sequencing method introduced by Sanger in 1977, HGP was able to reveal the first sequence of human DNA, which laid the foundation for genomic research (22, 61, 62).

With further revolutionary advances of sequencing technology over the past decade and introduction of next-generation sequencing (NGS), faster, less expensive, more accurate sequencing of genome and production of enormous amounts of data (Figure 8) are feasible. Using sequencing technologies, analysing of whole-genomes (WGS),

whole-exome (WES), gene panels (targeted resequencing), complete RNA transcriptomes, and chromatin map (ChIP-seq) has been made possible. In the light of these technologies, today a whole human genome can be sequenced at the cost of US$1000 and in less than 24 hours.

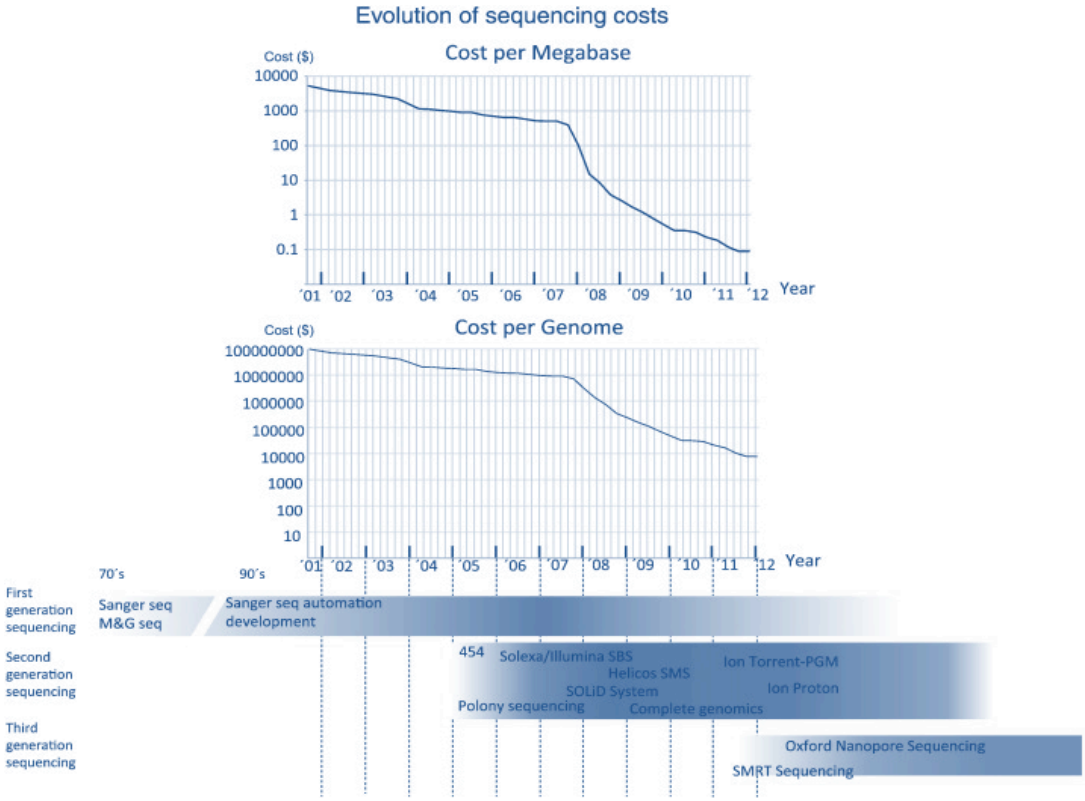


Figure8. Drop-down cost of sequencing per Mb as the sequencing technologies develop is represented. Price has been dramatically declined over past 5 years, concurrent with the introduction of new sequencing technologies. Adapted from Morey et al., Mol Genet Metab. 2013 Sep-Oct;110(1-2):3-24

With the benefit of sequencing technology, various databases of human genomes have been created. Wealth of data by projects such as The Cancer Genome Atlas (TCGA) [4]with the aim of identification of the key genomic changes in 50 major types and subtypes of cancer, and HapMap[5], which aims to determine the common patterns of DNA sequence variation in the human genome has been created. This provides powerful tools to enhance our understanding of genome-related health and disease issues. Recent advances in the field of genomic have raised the insight about the cancer genome and underlying aberrations in various cancers, and resulted in successful implementation of sequencing technologies for diagnosis and prognosis purposes (63-

[4] http://cancergenome.nih.gov
[5] https://hapmap.ncbi.nlm.nih.gov

70). In this respect, nowadays many studies on different aspects of cancers biology utilize sequencing-based strategies to interrogate cancer genomes.

1.3.2. Cancer genome analysis with next-generation sequencing

Two broad categories of "targeted" and "untargeted" approaches have been defined. The untargeted approaches such as WGS and WES interrogate all or a relatively large proportion of genome, whereas a more defined region of the genome like a set of genes is the central focus in targeted approaches Prominent advantage of WGS is its capability to characterize all variant types, specifically structural rearrangements. WES strategy provide higher sensitivity in comparison to for analysis of somatic alternations, such as point mutations and small indels in exonic regions of the genes [reviewed in (22)].

In cancer research, capabilities of targeted NGS strategies have gained traction, particularly due to the advantages over untargeted approaches such as cost per run, specificity and relatively easier data analysis [reviewed in (22)]. In this regard, a panel of frequently mutated genes in cancer will be focus of investigation. Decision on selection of genes can be made upon wealth of cancer genomic datasets available such as TCGA International Cancer Genome Consortium [6], and Catalogue Of Somatic Mutations In Cancer[7]. Potential advantages of targeted resequencing approaches have been recapitulated in Table 1.

[6] ICGC; https://dcc.icgc.org

[7] COSMIC; http://cancer.sanger.ac.uk

Table 1. Advantages and disadvantages of targeted resequencing strategy are represented.

	Description	Advantages	Disadvantages
Targeted resequencing strategy	Determines the somatic aberrations of sequence in a specific set of genes	Usually cheaper than exome or genome sequencing, but this depends on the size of the gene panel	Does not provide information on regions outside of the gene panel
		Focussed on particular regions of interest and so data interpretation is easier	
		No concern regarding incidental findings, since only the regions of interest are sequenced	
		Can customize the panel to capture problematic regions that are difficult to sequence using exome or genome strategies	

The common main feature of all NGS platforms is their ability of massively parallel sequence several samples at the same time. To achieve this purpose, libraries of DNA material prior to sequencing must be generated. Several methods have been introduced for this purpose, though they differ in several parameters such as required initial input of DNA material, processing time, and other metrics, depending on downstream procedures (61, 71). Regardless of technology, there are several common steps in this workflow i.e. fragmentation, end-repair, adapters ligation, and in case of targeted approach enrichment of regions of interest, in chronological order (72).

The library preparation procedure for genomic DNA usually starts with fragmentation of DNA to a target size, which varies by the goals of experiments and depending on the platform used. Fragmentation performs by mechanical or chemical methods. Fragmentation yields randomly sheared fragments with a normal size distribution pattern around the target size. The next steps in the line, is to repair the ends of randomly fragmented DNA. During this step, fragments become blunt-ended and 5'-phosphorylated by dedicated polymerases and kinases. Consequently, a poly A overhang (3'- dA overhang) will be built upon 3'-ends, which allows for ligation of paired-end adapters (61).

If next-generation sequencing of only specific set of genes or whole-exom is demanded, enrichment of genomic region(s) of interest is required. Various strategies for this

purpose are commercially available, e.g. PCR–based approaches and capture–based approach that each comes with advantages and weak points (71) that has been discussed later.

The sequencing procedure for Illumina (Illumina Inc., CA, USA) HiSeq platform has been typified, since this platform has been used in this project. The Sequencing-By-Synthesis (SBS) technology has been utilized in the Hiseq platforms that uses four fluorescently labeled deoxynucleoside triphosphates i.e. ddATP, ddGTP, ddCTP and ddTTP, blocked at their 3′-OH. Upon incorporation into the newly synthesized DNA molecule, emitted fluorescent signal as result of liberation of florescent dyes from each dNTPs will be detected [reviewed in (73)]. It is also possible to "index" fragments prior to sequencing, which allows samples to be pooled and sequenced in parallel. During sequencing, emission wavelength along with the signal intensity determines base call quality. A quality score (Q-score), ranging from Q10 to Q40 as highest score, is the prediction of probability of incidental base incorporation during sequencing. A high quality score implies that a base call is more reliable and less likely to be incorrect [reviewed in (73)].

1.4. Exploiting biomarkers for dignostic and prognosis in cancer care

The term biomarker, a portmanteau of "biological marker" was first described by Hulka *et al*. in 1990 as "cellular, biochemical or molecular alterations that are measurable in biological media such as human tissues, cells, or fluids"(74). In 2001, the National Institutes of Health Biomarkers Definitions Working Group defined a biomarker as "a characteristic that is objectively measured and evaluated as an indicator of normal biological processes, pathogenic processes, or pharmacologic responses to a therapeutic intervention." (75) Clinical significance of biomarkers is attributed to their relevance and validity characteristics (76). Currently, circulating tumour cells (CTCs), circulating tumour DNA (ctDNA), exosome microvesicles (EMVs) and microRNA (miRNA) are all available alternative, alongside protein markers (Figure 9)[reviewed in (77-80)]. Herein, cfDNA as a promising and rapidly advancing biomarker will be of central focus, due to its relevance to the scope of present study.

However, applications of cfDNA in comparison to CTCs and miRNAs as potential biomarkers will be further discussed.

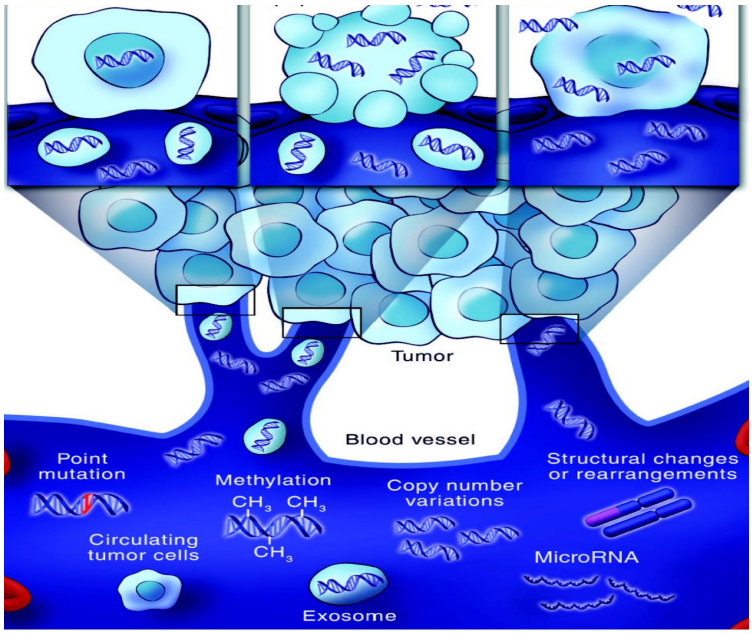


Figure 9 An overview of detectable biomarkers in the circulation and their potential applications are shown. Adapted from Diaz et al., J Clin Oncol. 2014 Feb 20;32*(6)*:579-86.

1.4.1. cfDNA as potential clinical biomarker

1.4.1.1. Biology of cfDNA

Circulating cell-free nucleic acid (cfDNA) in the blood of healthy individuals was first identified by Mandel and Métais in 1948 (81). The cfDNA is present in plasma, serum, urine and other bodily fluids from both healthy and individuals with disease [reviewed in (80, 82)]. Both genomic DNA (gDNA) and mitochondrial DNA (mtDNA), attached to nucleoprotein complexes, contribute to the overall concentration of cfDNA (83, 84).

The half-life in range from 15 minutes to several hours in circulation for cfDNA has been reported [reviewed in (80)]. Several mechanisms have been demonstrated to contribute to the levels of cfDNA in plasma, however apoptosis and necrosis are considered to be the two main mechanisms, contributing to the release of cfDNA into blood circulation (Figure 10). While apoptosis, of mostly hematopoietic lineage cells, contributes to cfDNA in the plasma of healthy individuals, occurrence of both mechanisms simultaneously in the tumour microenvironment has been considered to contribute to the total cfDNA concentration in cancer patients (83, 84).

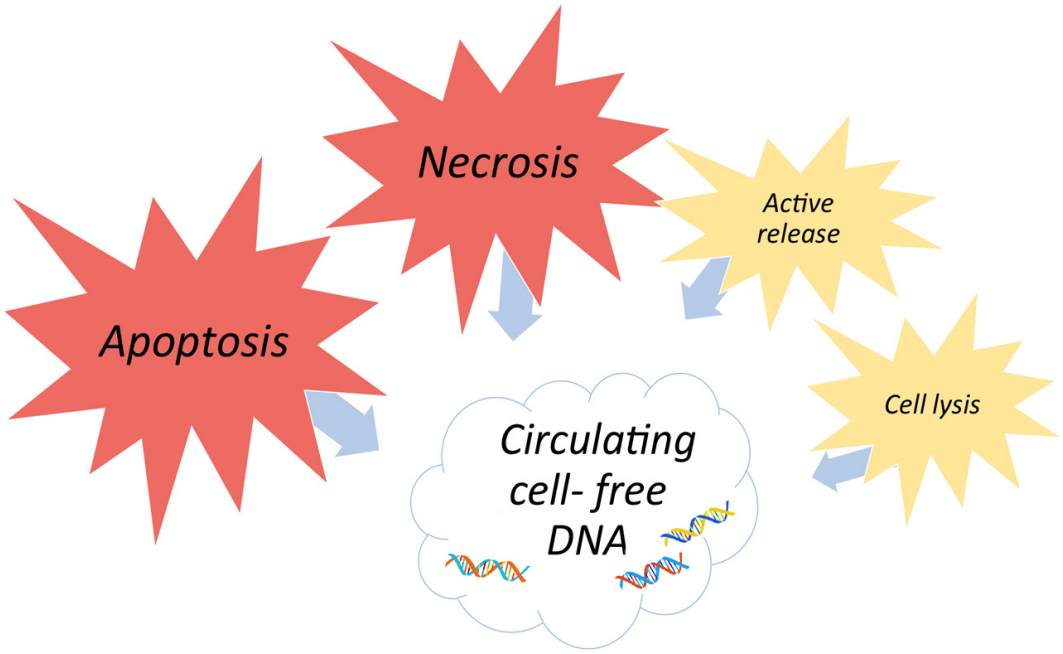


Figure 10. Cellular mechanisms that contribute to the total concentration of circulating cell-free DNA in the blood are shown. Every cell in the body undergoes these mechanisms, including cancer cells; therefore their DNA is released to the circulation.

Each of these mechanisms produces fragments of various sizes, ranging between small fragments of 70 to 200 bp, up to large fragments of approximately 21Kb (63, 83, 84). However, fragments at approximation of 180bp to 200bp are frequent, which corresponds to the size of nucleosome-bound DNA and mostly resemble the characteristic of apoptotic cell death (83).

The tumour-derived fraction of cfDNA (ctDNA) range from <0.1% up to 93% of total cfDNA molecules in plasma, which putatively depends on multi variables [reviewed in (80, 85)]. Several studies have demonstrated that the level of cfDNA in serum is higher than in compared matched plasma samples. However, this phenomenon has been ascribed to clotting of nucleated blood cells consequent to venepuncture, therefor it has been reasoned that the cfDNA is more contaminated with gDNA in comparison to purified cfDNA from plasma [reviewed in (86, 87)].

1.4.1.2. Clinical application of cfDNA

Increased concentration of cfDNA under abnormal pathological situations like cancer, stroke, myocardial infraction, autoimmune disease, intensive exercise and infections, in

comparison to normal pathological condition has been demonstrated. Presence of circulating foetal DNA in maternal plasma and serum has been also detected [reviewed in (88)]. Elevated total concentration of cfDNA in cancer status has been described to correlate with cancer type, stage and tumour burden, vascularity, and cellular turnover [reviewed in (80, 82)]. However, there has been no consensus on possible correlation between total cfDNA concentration (63) and tumour location and size [reviewed in (86)], and overall concentration of cfDNA between individuals in similar condition has been demonstrated to vary, considerably. For various cancer types, an average concentration of 180 ng/ml for circulating free DNA in plasma have been observed, whereas in healthy individual the mean concentration of 13 ng/ml has been reported [reviewed in (83)]. In this regards, initial efforts of studies on clinical utility of cfDNA, which was directed towards simple quantitative evaluation of cfDNA concentration in the circulation were inconsistence and sometimes contradictory [reviewed in (86-88)].

More recently, qualitative changes of cfDNA have been considered the most clinically important aspect of cfDNA analysis. The evidence of existence of tumour-specific oncogenic mutations of *N-RAS* and *K-RAS* in cfDNA has been reported from studies conducted as early as in 1994. Tumour genotyping with the purpose of identifying actionable mutations, such as *NRAS, EGFR, BRAF, PIK3CA, KRAS,* and *KIT* across various cancer types is the most immediate application of cfDNA analysis that can lead to the use of available therapeutic in cancer patients, harbouring similar oncogenic mutations [reviewed in (53, 85, 89)]. In addition, tumour-specific changes have been been demonstrated as a promising applications of cfDNA in early detection of tumour recurrence after surgery or treatment with curative intent (Figure 11). In this regard, investigation of detectable tumour-specific alternations in cfDNA has been suggested to inform about minimal residual disease and/or emergence of acquired resistance in refractory cancers [reviewed in (85, 89)]. Ever increasing evidence on the presence of tumour-specific alternations in cfDNA of cancer patients has encouraged the use of NGS strategies to interrogate variety of genetic and epigenetic aberrations in several malignancies in breast (90) and lung (91, 92), ovarian (93, 94) and few sporadic studies in other cancer types. These studies demonstrate that by analysing of cfDNA detection of tumour specific alternations with high sensitivity is feasible [reviewed in (85, 89, 95)]. Overall, analysis of cfDNA is a minimally invasive strategy that may guide

the course of therapy much faster and accurate than conventional diagnostic strategies, based on the molecular profile of cancer, which subsequently can also eliminate hazardous side effects of therapeutics on non-responsive patients [reviewed in (85, 89, 95, 96).

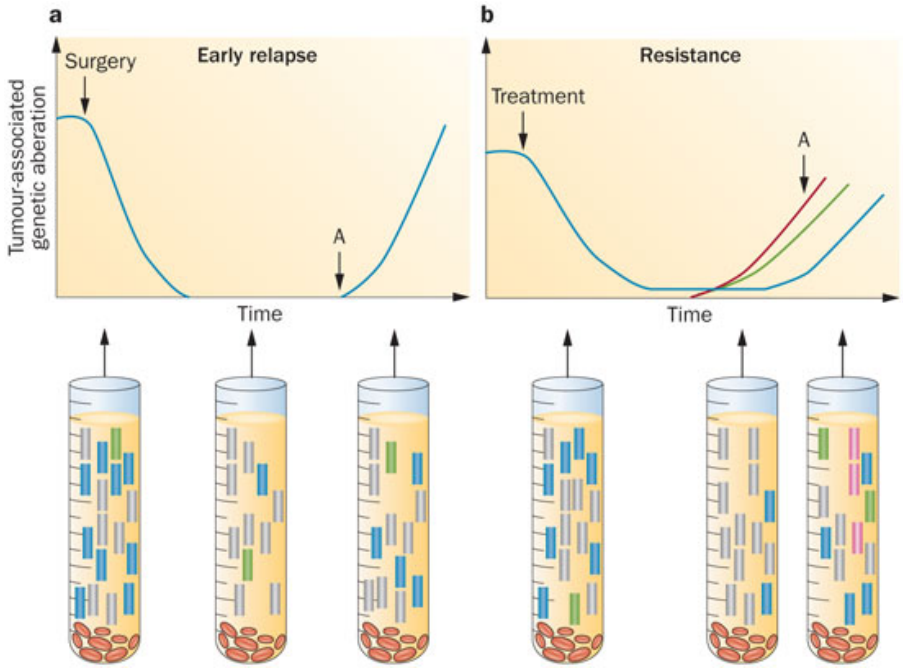


Figure 11. Potential clinical application of tumour-derived cfDNA in cancer is shown. In this illustration clinical scenarios following A) a surgical excision and B) a course of treatment are demonstrated. Part (a) illustrate that by monitoring of one or few tumour-specific mutations in the ctDNA the recurrence of the cancer can be rapidly detected. part (b), by genotyping tumour specific mutations emergence of new mutations that play a role in the resistance can be uncovered. Adapted from Crowley et al., Nat Rev Clin Oncol. 2013 Aug;10(8):472-84

Finally, It has been postulated that early detection of individuals at high risk of cancer, due to known inherited genetic risk factors, by serial monitoring of cfDNA is feasible [reviewed in (95)], however, extremely low level of cfDNA in healthy individual should be considered.

1.4.2. Introduction of liquid biopsy strategy

To overcome limitations posed by tissue biopsy in accurate and dynamic diagnosis and prognosis of malignancies concept of liquid biopsy has been developed. Advances in the field of genomics have paved the way for new methods to foster diagnostic, predictive and prognostic procedures. Emerging field of liquid biopsy offers advantages in early diagnosis, prognosis, and non-invasive monitoring of diseases at a

sensitive and dynamic fashion over conventional tissue biopsy approaches. For these purposes, liquid biopsy exploits information from enumerated biomarkers found in the circulation [reviewed in (85, 89, 92)]

Liquid biopsies have so far been able to tackle some of the current limitations that tissue sampling was unable to assists with. First, it allows serial biopsies of patients during the course of their disease, resulting in real-time monitoring of patients, at the cost of a blood draw. Second, in contrast to tissue biopsy that is limited both spatially and temporally, liquid biopsies have been able to delineate intra- and inter-tumour heterogeneity during the tumour development [reviewed in (85, 89)].

Both features underlined above are important benefits of liquid biopsies that can address concerns regarding tumour heterogeneity, where its particular subsequent implication is to guide clinical decision making and faster treatment initiate based on cancer genomic profile [reviewed in (85, 89)]

Another challenge in the management of cancer patient is posed by the lack of adequate sensitivity for detection of deep-tumour recurrence, common feature of sarcomas tumours, and micro-metastases in current imaging methods. Notably in sarcomas, radiations have been strongly associated with the elevated risk of cancer development (31, 45). Growing body of evidence has demonstrated that, for instance, small traces of cfDNA in the plasma harbouring tumour-specific mutations, can be detected by liquid biopsy approach, allowing for sensitive disease burden screening and recurrence detection (63, 97, 98) [reviewed in(95)].

Applications of liquid biopsies have been recapitulated in the Table 2 [reviewed in (95, 99)]. Despite the future promises, several issues that need to be addressed before the implementation of liquid biopsies in the clinical setting.

Table 2. Application of liquid biopsy in the clinical practice has been summarized.

Approach	Application
Diagnostic	Early detection
	Monitoring of minimal residual disease
Predictive	Evaluation of molecular heterogeneity
	Monitoring of tumor dynamics
	Identification of genetic determinants for targeted therapy
	Treatment response assessment
	Real-time investigation of tumour evolution
Prognostic	Evaluation of risk of relapse
	Evaluation of changes in tumor burden

1.5. Project background

The prospective study Circulating DNA in Sarcoma (CircSarc), has initiated the evaluation of the clinical impact of ctDNA as a biomarker for disease monitoring in STS patients. The prospective CircSarc study is a joint effort between surgeons, molecular biologists, pathologists, and bioinformaticians. The goal of the study is to recruit 30 patients. Patient material is collected at Oslo and Haukeland University Hospitals. For each patient, tumour material is being collected at the time of surgery, as well as EDTA blood samples before and after surgery, before and after each treatment cycle and at each routine control for five years or to death.

The immediate aim of this study is to evaluate the utility of liquid biopsies to monitor tumour burden by targeted resequencing. Use of liquid biopsies can be used for early detection of ctDNA and thus identify patients that have developed relapse or metastatic disease before this can be detected by standard clinical methods, which increases the possibility of an early intervention. Identification of new mutations in the ctDNA or changes in the frequency of the initial tumour mutations may give information on mutations possibly involved in drug resistance mechanisms. Moreover, analysis of sequencing dataset may reveal the new therapeutic targets that have been previously found in other cancer types. This provides the opportunity to implement FDA-approved drugs designed for those cancers in STS patients, a concept known as “drug repositioning”.

1.6. Aims of the study

The project described in this thesis is a part of the prospective study "Circulating DNA in Sarcoma" (CircSarc). CircSarc will evaluate the clinical impact of ctDNA as a biomarker for disease monitoring in STS patients. The immediate aim of this study is to evaluate the utility of plasma cfDNA to monitor tumour burden by targeted resequencing to identify recurrent tumours prior to clinical manifestation of the disease, which increases the possibility of early medical interventions in patients who are at risk of developing relapse or metastatic disease. The second goal of this study is to underpin patients who may benefit from neo-adjuvant and adjuvant chemotherapy based on the concentration of tumour-derived mutations in the plasma. In addition, as a result of blood-based real time monitoring of patients it might also be possible to separate patients with refractory cancer who may benefit from other therapeutic modules.

The work described in this master project is part of CircSarc study, and the overall aim is to evaluate the clinical impact of cfDNA in STS. More specifically, the detection of tumour-specific mutations in plasma cfDNA will be evaluated, and how well the cfDNA represent the mutational profile of the primary tumour. This will be done by identifying somatic mutations in primary STS tumours, establishing methods for sequencing of cfDNA, and identifying somatic mutations in cfDNA.

2.Material and methods

2.1. Patients

All clinical specimens including blood, tumour, and plasma samples were obtained from the patients recruited into the CircSarc or NoSarc studies. Both projects have been approved by the Regional Ethical Committee for Southern Norway (S-06133). Informed, written consent was obtained from patients before enrolment into the studies. The patients eligible to the study were above 18 years of age, had high-grade malignant soft tissue sarcoma in extremities and trunk wall, and had tumours that could be operated with wide or marginal surgical margins. Patients with MPNST, conditions associated with a risk of poor protocol compliance, patients that had received preoperative treatment (chemotherapy or radiation), and patients that had other cancers during the past three years before commencement of study were excluded from the study.

2.2. Tumour materials

Collected tumour tissues were collected and fresh frozen at -80˚C immediately after surgery. For each patient, the tumour piece was divided into 3 parts, and labelled A to C consecutively. Both parts A and C represented ends of the tumour piece, meanwhile B the middle. Part B was used for DNA extraction. From the junction of part B, part A and C were sectioned (4 μm) on a cryostat (Leica CM1950; Leica Microsystems Ltd, Wetzlar, Germany). Slides were stained with Haematoxylin and Eosin (H&E) and examined by a pathologist to confirm the presence of cancerous cells in the tumour piece.

Haematoxylin and Eosin (H&E staining was prepared as followed: Tumour slides were submerged in 4 % formalin for 4 min followed by rinsing in water. Then, slides were stained in Haematoxylin for 3 min, which colours nuclei into a deep blue-purple colour. Slides were rinsed in running water and immersed into ammonia solution for 10 s, and submerged in water for 4 min. Counterstaining of the slides, to achieve full cellular detail, were performed by submerging slides into Eosin for 1 min, which colours

eosinophilic structures in various shades of red or pink. Eosin-stained slides were dehydrated through consecutive steps in a serial dilution of ethanol, starting from 76 % up to 100 %. The slides were then mounted and examined by a sarcomas pathologist.

2.3. Normal blood and plasma materials

In the CircSarc study, blood samples were collected 1 day before surgery, 3 days after surgery and each time the patients came for routine control, approximately every 3 months. In addition, sample collection was carried out before and after administration of therapy, i.e. radiation or neo- adjacent chemotherapy (Figure 12). For this thesis project, only blood samples collected before surgery were used. The blood samples were collected in EDTA tubes (BD Vacutainer, USA) and processed within 2 hours from venepuncture to minimise contamination as a result of nucleated blood cells lysis. Of note, four of the plasma samples used in this study were collected under the NoSarc project protocol in which blood samples were collected in K3EDTA tubes (Streck Inc., Omaha, NE, USA) and processed within 24 hours after venepuncture. The collected samples were subjected to room temperature centrifugation at 820 x *g* for 10 min (fixed angle rotor), and the plasma layer was carefully transferred into new tubes without disturbing the buffy coat layer. The samples were then re-centrifuged at 10,000 x *g* for 10 min to pellet any possible cell debris and high molecular weight DNA attached to cell membranes. Then plasma was transferred into 2 mL collection tube (Qiagen), leaving behind 0.5 mL of supernatant to avoid possible carryover of residual cell debris. Processed plasma samples were stored at -80 °C until DNA was extracted. The buffy coat layer from the first centrifugation at 820 x *g*, representing normal cells, was frozen at -80 ˚C.

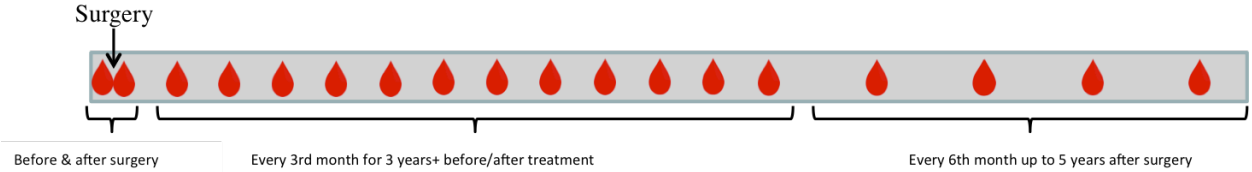


Figure 12. Representative timeline of sample collection for each patient in CircSarc study is shown.

2.4. Isolation of DNA

2.4.1. Isolation of tumour DNA

DNA from fresh frozen tumour material was isolated using the TissueLyser LT (Qiagen, Hilden, Germany) and the Wizard Genomic DNA Purification Kit (Promega, Wisconsin, US) according to the manufacturer's protocol.

Twenty-five mg of frozen tumour material was added to nuclei Lysis solution provided in the Wizard Genomic DNA Purification Kit, and pre-cold 5 mm stainless steel beads were added. The tube with DNA and beads was inserted into a TissueLyser LT shaker, and shacked for 40 s at 30 Hz. The homogenized tumour tissue solution was treated with Proteinase K (Promega, Madison, WI) and incubated at 55 °C for 3 hours with gentle shaking. Further steps included RNase treatment and isopropanol ethanol precipitation as described in the provided protocol.

2.4.2. Isolation of normal and circulating cell-free DNA

DNA purification from up to 1mL of buffy coat from EDTA blood was performed using QIAamp DNA Blood Midi Spin protocol (Qiagen), according to manufacturer's instruction. The isolated DNA was then stored at 4 °C.

Circulating cell-free nucleic acids in biological fluids are commonly attached to proteins or engulfed in vesicles; therefore an efficient lysis procedure to release and purify them is necessary. Purification of cfDNA from up to 2 ml of plasma was performed under highly denaturing conditions using QIAamp Circulating Nucleic Acid assay (Qiagen), according to manufacturer's instruction. In brief, the procedure comprises 4 steps (lyse, bind, wash, elute) and is carried out using QIAamp Mini columns on a vacuum manifold. After thawing, the plasma was re-centrifuged at 16,000 x *g* for 5 min, and the supernatant, leaving 50 µL in the tube, was transferred to a new tube and incubated with lysis buffer and proteinase K, ensuring inactivation of DNases/RNases and complete release of nucleic acids from other macro-molecules, i.e. proteins and lipids.

In addition, carrier RNA was added to enhance binding of circulating DNA to the QIAamp Mini membrane. The cfDNA was eluted with 25 µL of buffer AVE, and stored at -20 ˚C.

2.5. DNA quantification and quality control

The purity of tumour and normal genomic DNA were measured with NanoDrop 2000 spectrophotometer (Thermo Scientific, MA, USA). The purity of isolated DNA was determined by ratios of absorbance at A260/280 nm and A230/260 nm on the spectrophotometer. Ratio values of approximately 1.8, and 2.0 – 2.2 are accepted for pure nucleic acid, respectively. Considerable deviations from these values indicate the presence of protein or other contaminants that absorb strongly at or near these wavelengths.

Despite widely accepted utility of this method for quality measures, quantitative measure by spectrophotometry lacks specificity, since it measures both single stranded DNA (ssDNA), double stranded DNA (dsDNA), and RNA contaminations in the reaction. Accurate concentration measurement, which is crucial for the downstream library preparation procedure, was carried out by Qubit BR dsDNA assay kit (Life Technologies, CA, USA), according to manufacturer's instruction. Qubit fluorometer detects emitted fluorescent signals from binding of dye molecules to dsDNA in the reaction.

2.6. Next generation sequencing library preparation

DNA Fragmentation

Fragmentation was performed with Covaris S2 Ultrasonicator (Covaris, MA, USA) with the following setting: Duty Cycle 10%, Intensity 5, Cycles per burst 200, and Time 6 cycle of 60 seconds each, 4 °C temperature. The Covaris utilizes Adaptive Focused Acoustics technology to produce controllable cavitation by acoustic stream to mechanically shear the aqueous DNA sample. This method is highly predictable and reproducible in terms of output fragment size distribution, and effectively minimizes GC bias and avoids thermal damages. The fragment size has a profound impact on the

outcome of a target-enrichment experiment, with shorter fragments constantly being captured with higher specificity than longer ones, indicating that longer fragments probably contain a higher proportion of off-target sequence. Furthermore, longer fragments may have a higher potential for cross-hybridization.

A volume of 50 µL containing approximately 1,500 ng normal and/or tumour dsDNA was separately transferred to Snap-Cap microTUBE (Covaris) provided by manufacturer. Fragmentation setting was adjusted to recommend value provided with instrument for the target size of 150 bp to 200 bp DNA fragments. Of note, the fragmentation part was skipped in cfDNA library procedure.

Size selection and sample clean-up with paramagnetic beads

In the predefined polyethylene glycol (PEG) and salt concentration, ssDNA and dsDNA in the range of >100 bp up to 10kb can be reversibly captured with carboxyl-coated paramagnetic beads and separated from unwanted DNA fragments and substances in the sample reaction such as primers, dNTPs and adapter dimers that may influence steps in the downstream workflow (Figure 13). During the library preparation, clean-up steps following fragmentation of genomic/tumour DNA and PCR amplification were performed using Agencourt AMPure XP beads (Beckman Coulter Inc., CA, USA). To capture fragments size >100 bp, 50 µL of sample reaction, containing sheared DNA was added to 90 µL (1:1.8 ratio) of room temperatured homogeneous AMPure XP beads, followed by 2 wash steps with freshly prepared ethanol, according to manufacturer's protocol. PCR clean-up step using AMPure beads using the same protocol was carried out for each sample, following the provided protocol by the manufacturer.

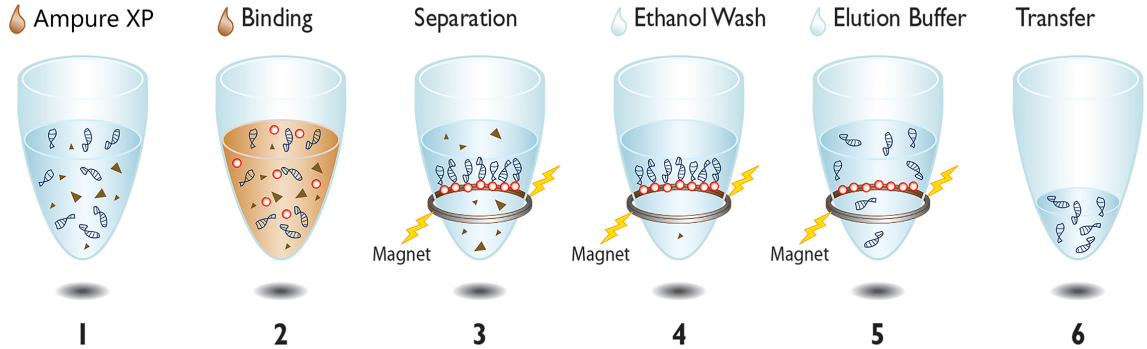


Figure 13. AMPure XP beads clean-up process is shown. Picture acquired from Beckman Coulter Inc., 1) Sample volume containing different fragment sizes, 2) Binding of DNA fragments to magnetic beads proportional to volume of beads, 3) Size selection and separation of fragments bound to magnetic beads, 4) Washing with Ethanol to discard other fragment size in reaction, 5)

Elution of bound DNA fragments from the magnetic particles, 6) Transfer elution away from the beads into a new tubes.

2.6.1. Normal and tumour DNA library construction

Sequencing libraries from tumour and normal DNA samples were generated by the SureSelect^{XT} Target Enrichment System for Illumina Paired-End Sequencing Library protocol (G7530-90000). For each sample, an individual indexed library was prepared. The main steps of the workflow were as follow; 1) library preparation, 2) hybridization and capture, 3) library indexing for multiplexed sequencing, each including different enzymatic steps. Detailed description is available in manufacturer's protocol. An overview of the steps in library preparation workflow is shown in the figurative depiction (Figure 14).

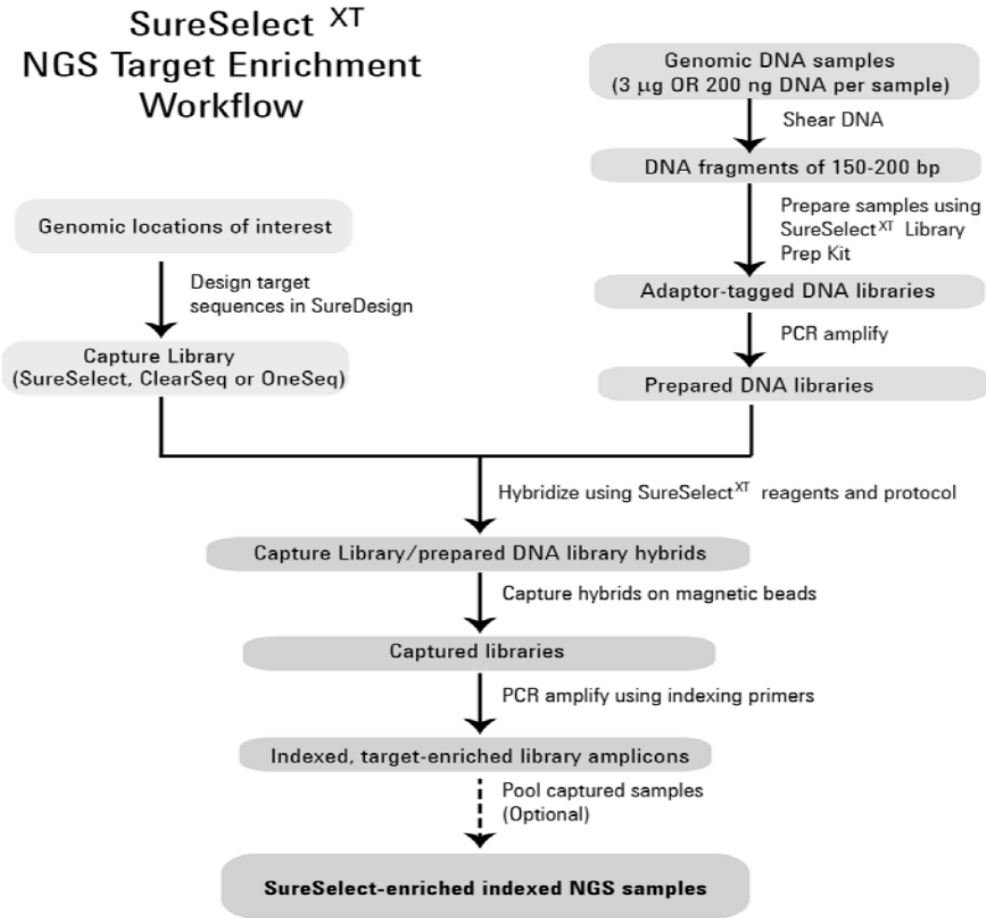


Figure 14. Target-enrichment library preparation workflow used for generation of sequencing libraries is shown. Graph acquired from Agilent Technologies SureSelect^{XT} Target Enrichment System for Illumina Paired-End Sequencing Library protocol.

Following DNA fragmentation to the desired length ends-repair of the DNA fragments was carried out. The introduced 5' overhangs, as result of fragmentation procedure of both normal and tumour DNA molecules, were enzymatically repaired by T4 DNA polymerase, Klenow DNA polymerase, and T4 polynucleotide kinase, provided by the Agilent SureSelect XT Library Prep kit ILM. The samples were purified with AMPure XP beads and subjected to 3' ends adenylation, following manufacturer's protocol. Consequently, Illumina compatible Paired-end adaptors were ligated to the A-tailed purified DNA by 6 cycles of PCR amplification, according to the manufacturer's protocol.

Quantity and fragment size of purified libraries were analysed with the Agilent 2200 TapeStation System on D1000 ScreenTape (Santa Clara, CA, USA) according to manufacturer's protocols. The Agilent 2200 TapeStation system is an automated electrophoresis-based platform, which separates DNA and RNA molecules based on size, and provides information on the sample size distribution and size–based quantification. D1000 ScreenTape is designed for analysing DNA molecules within the quantitative range of 0.1 – 50 ng/μL with size distribution between 35 and 1000 bp.

In the second step, the prepared library was subjected to hybridization and capturing. The hybridization reaction required initial concentration of 750 ng in a volume of 3.4 μL (221ng/μL). For libraries below the desired amount, the entire library volume was subjected to dehydration by vacuum concentrator at low heat and reconstituted with 3.4 μL of nuclease-free water.

DNA libraries were mixed with SureSelect Block Mix, RNase Block dilution and Capture Library Hybridization Mix. The Agilent SureSelect^{XT} Human All Exon v5 baits were used to capture whole 50 Mb of the human exome. RNase Block mix used to avoid RNase-induced degradation of RNA baits during sample preparation procedure. The hybridization reactions were subjected to overnight incubation at 65 °C, according to manufacturer's instruction. Dynabeads® MyOne™ Streptavidin T1 magnetic beads (Life Technologies, p/n 65602) were used to fish target regions captured on RNA baits out from the off-target molecules in the reaction, according to manufacturer's protocol

(Figure 15). The utilized capture libraries were composed of biotinylated 120mer RNA baits, with complementary sequence to the DNA libraries.

In the third step of the procedure the captured libraries were indexed with 8 bp indexes, following 10 cycles of post-capture PCR amplification, according to the manufacturer instruction. To quality control the indexed libraries, consequent to PCR clean up with the AMPure XP beads libraries were analysed for the size distribution and concentration using Agilent TapeStation on HS D1000 ScreenTape and reagent kit (Agilent technologies).

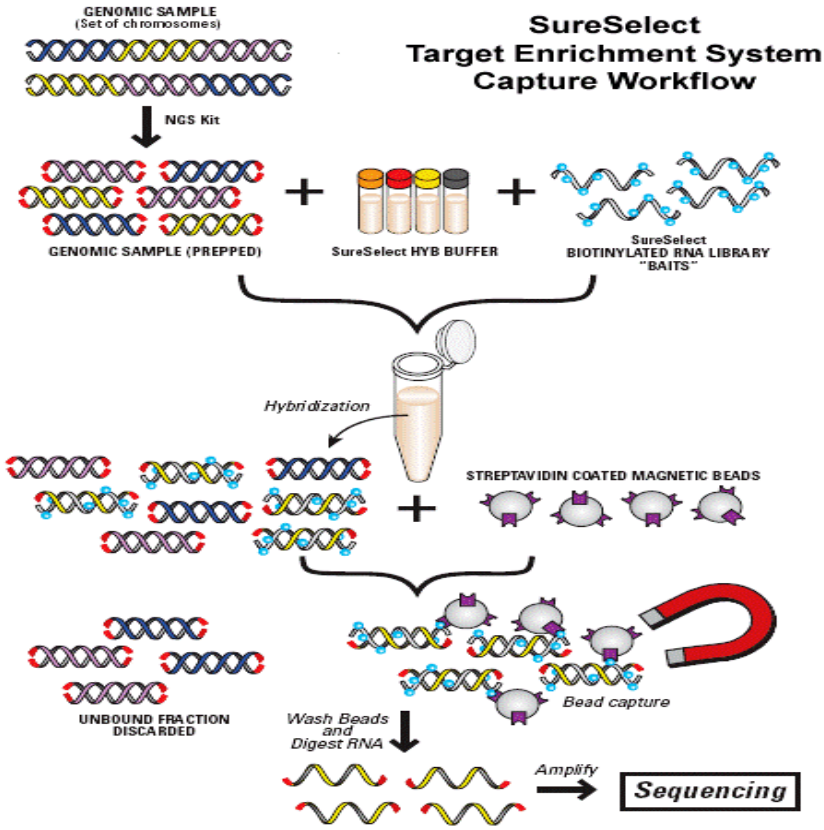


Figure 15. Schematic overview of the in-solution capture hybridization workflow is shown. Dynabeads® MyOne™ Streptavidin T1 were used for capturing regions of interests, specifically bound to RNA baits during library preparation. Image acquired from www.genomics.agilent.com

Based on the measured concentration on Agilent TapeStation on HS D1000 ScreenTape of target fragments peak within the recommended range, 250 bp to 350 bp, tumour and normal libraries were diluted to a centration of 4 nM. Prepared cfDNA libraries were similarly quantified and diluted to 5 nM. Prior to pooling samples, prepared dilutions were quantified with qPCR assay to verify the optimal molar concentration of

amplifiable libraries as an essential prerequisite for proper cluster generations during sequencing. The PCR assay only amplifies adapter-ligated molecules.

qPCR quantification was performed using the QPCR NGS Library Quantification protocol (Illumina, GA). Each library was treated by the Stratagene Brilliant III Ultra-Fast SYBR® Green QPCR Master Mix, containing a mutant *Taq* DNA polymerase, dNTPs, and double-stranded DNA-binding dye SYBR Green I according to manufacturer's protocol. The qPCR plate was set up in the following setting; Five 10-fold serial dilutions of the control template in range of 10 pM down to 0.001 pM and two 10-fold of generated libraries in duplicates were prepared. Moreover, 2 Non-Template Controls (NTC) were included to ensure no contamination.

2.6.2. Circulating cell free DNA library preparation

The plasma-purified cfDNA were processed using ThruPLEX Plasma-seq 12S kit (Rubicon Genomics, MI, USA). The protocol has been optimized to prepare libraries from as low as 1 ng cfDNA. The workflow consisted of 3 consecutive steps of repair, ligate, and amplification. To generate libraries from cfDNA, manufacturer's instruction was followed. Generated libraries were subjected to qPCR assay, which carried out on 7900HT Fast Real-Time PCR System (Applied Biosystems, CA, USA). The library amplification's master mix for RT-qPCR quantification was prepared according to the manufacturer protocol, and contained 8 nt Sanger indexes, provided with the kit, and 20X EvaGreen dye (Cat. no. 31000-T, Biotium, CA, USA) for monitoring the amplification of the libraries after each PCR cycles. To achieve optimal quantities and minimize PCR duplicate rates, the amplification reaction was stopped when the linear phase was reached, after 8 PCR cycles.

Generated libraries were quantified on Agilent TapeStation HS D5000 ScreenTape to better resolve small the small fragment-size, and quantify the desired fragment peaks. The schematic representation in Figure 16 shows the principles of ThruPLEX Plasma-seq library preparation at each step.

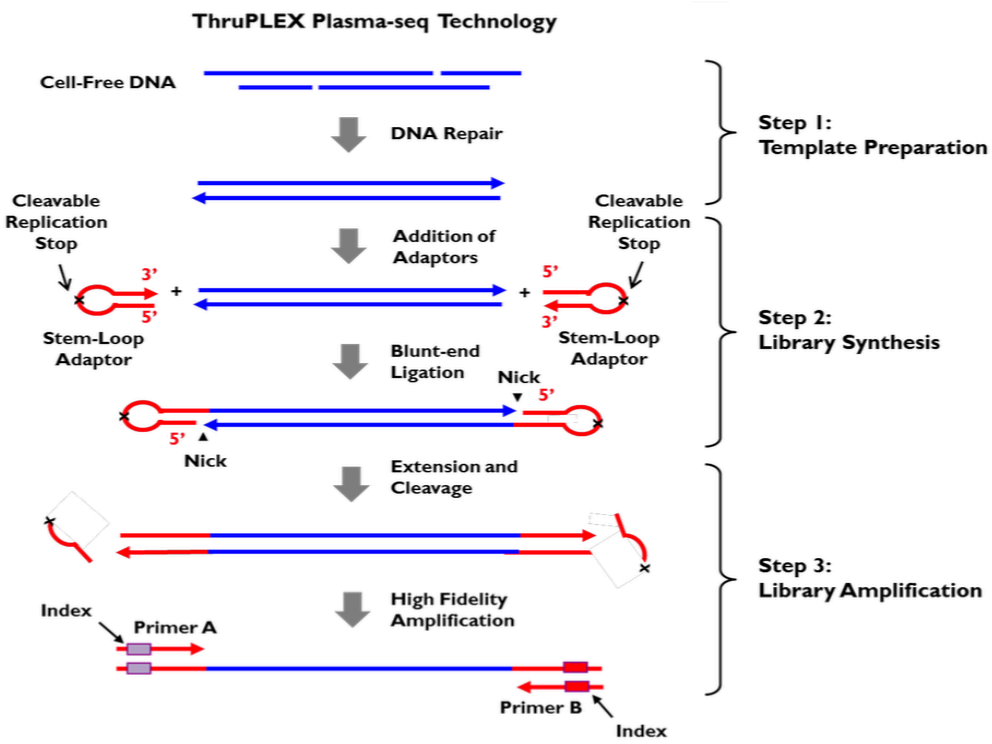


Figure 16. The three-step ThruPLEX Plasma-seq Technology is shown. The physical shape of thestem-loop adapters and the blunt-end ligation efficiently avoid formation of concatemers *and primer-dimer structures. Image acquired from ThruPLEX Plasma-seq Technology protocol, Rubicon genomics.*

Downstream enrichment of exonic regions of interest accomplished by Agilent SureSelectXT Target Enrichment system (G7530-90000), as previously described for generation of genomic normal/tumour DNA libraries. In this respect, amplified cfDNA fragmented using specifically designed (Agilent Inc.,) capture library for NCGC 900 genes panel with the approximate size of 3.5 Mb were captured. Guided protocol (RDM-153-002) provided by Rubicon Genomics followed entirely.

2.7. Next-generation sequencing and data processing

All prepared libraries were sequenced at the Genomic Core Facilities at Oslo University Hospital. The tumour sample libraries were sequenced at a mean coverage of > 200-fold, and the matched normal libraries were sequenced at >50-fold coverage on either HiSeq2500/4000 platforms. cfDNA libraries were sequenced on Illumina HiSeq 4000, with the average depth of >400-fold coverage. Paired-end reads approach of 100 bp was utilized for all sample types.

Real-time analysis and base calling were conducted by Illumina's software packages HSC2.0.2/RTA1.17.21.3. De-multiplexing was performed on the raw data and low-quality reads were filtered using Illumina CASAVA (v. 1.8.2). Consequently, sequenced reads were analysed through a custom pipeline (Figure 17) developed by the bioinformatics team of the Norwegian Cancer Genomics Consortium. Validated reads were mapped to human reference genome hg19 (build b37) with Burrows-Wheeler Alignment tool (BWA-mem)(100). Subsequently Picard tool[8] was used for sample sorting and duplicate marking. GATK[9] was used for two-step local realignment around indels, with tumour–normal pair samples. Paired-end information was checked for inconsistencies, and base-quality recalibration was performed by GATK. Somatic variant calling on the matching paired samples was performed by MuTect (101) and Strelka (102) callers. Using Integrative Genomics Viewer (IGV) (103), the overall alignment of reads and read depth around the target regions, quality of mapping and base calls, and base composition of reference sequence were visually interrogated.

8 http://broadinstittute.github.io/picard

9 https://www.broadinstitute.org/gatk

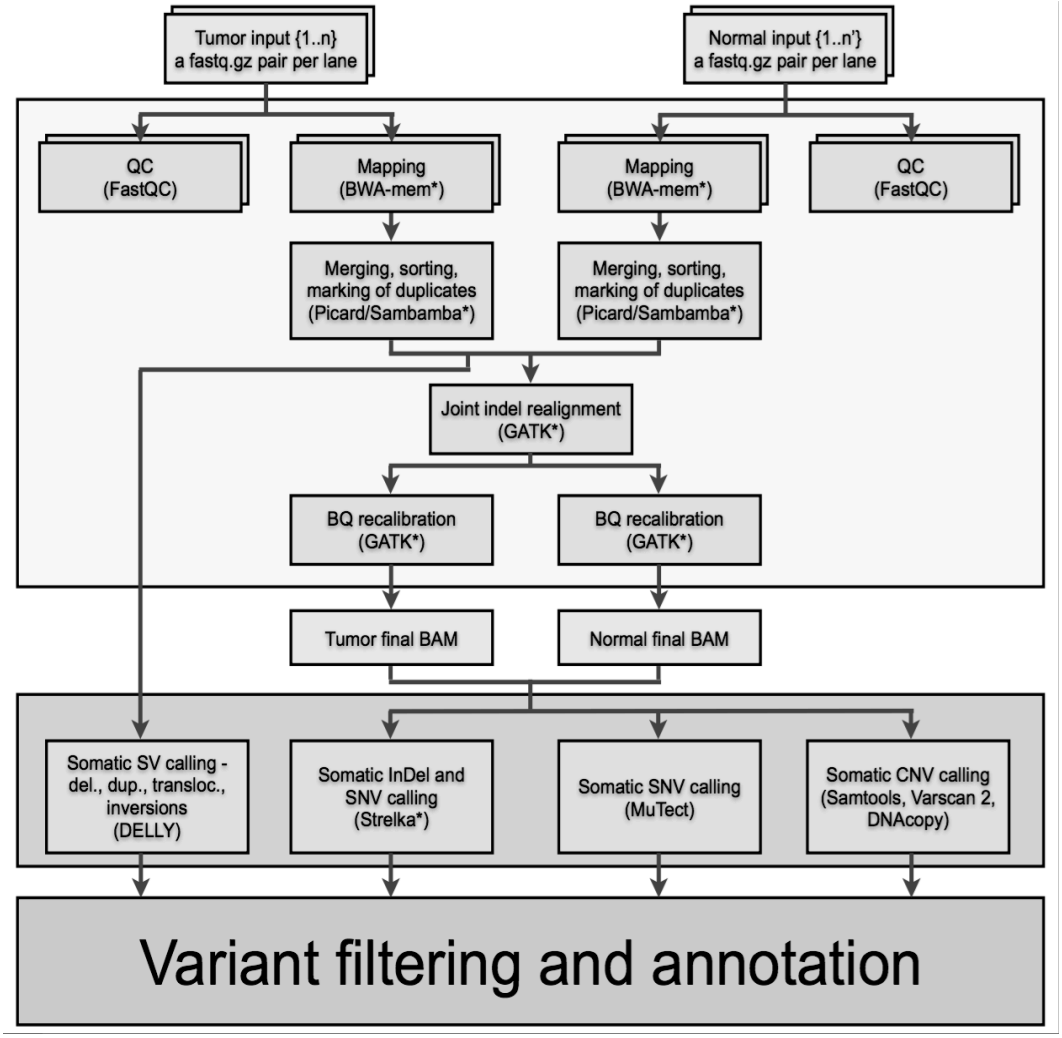


Figure 17. The workflow of custom bioinformatics pipeline utilized for mapping, sorting, filtering and realignment of the reads, in this project is shown.

3.Results

3.1. Clinical features of the patients

We examined 14 STS patients from 7 distinct subtypes (male n = 9, female n = 5) (Table 4). The major subtypes in this series were myxofibrosarcoma (n=5) and leiomyosarcoma (n=4). The mean age of the patients was 62.3 years. The presence of cancerous cells in each tumour specimen was confirmed by pathological examination of tumour slides, done at the Department of Pathology at Oslo University Hospital. A tumour cell content of >90% was reported in 11 of the 14 tumour samples, and for the remaining samples a tumour content between 50-80% was observed. The clinical characteristics of the patients are summarised in Table 3.

Table 3. Clinical features of the patients in the study are represented.

Sample	Subtype	Tumour	Tumour %	Status	Tumour Size (Cm)	Metastasis
CS01	MFS	Primary	>50%	DOD	17	N
CS02	DDLPS	Primary	>90%	NED	8	N
CS03	UPS	Primary	>90%	AWD	8	N
CS04	PLPS	Primary	>60%	NED	12	N
CS05	MFS	Primary	>90%	NED	10	N
CS06	PEComa	Primary	>90%	NED	11	N
CS07	MFS	Primary	>90%	NED	9	N
CS08	MFS	Primary	>90%	NED	36	N
CS09	LMS	Primary	>90%	AWD	11	YES
CS10	SS	Primary	>90%	NED	9	N
CS11	LMS	Primary	>90%	NED	13	YES
CS12	LMS	Primary	>80%	DOD	9	YES
CS13	LMS	Primary	>90%	DOD	26	N
CS14	MFS	Local recurrence	>90%	AWD	6	N

UPS: Undifferentiated pleomorphic sarcoma, DDLPS: Dedifferentiated liposarcoma, MFS: myxofibrosarcoma, LMS: leiomyosarcoma, SS: Synovial sarcoma, PLPS: Pleomorphic liposarcoma, PEComa: Malignant perivascular epitheliod cell tumour, AWD: Alive With Disease, NED: No Evidence of Disease, DOD: Dead Of Disease.

3.2. Assessment of next-generation sequencing libraries

The constructed tumour/normal libraries using the Agilent SureSelect^{XT} Human All Exon v5, and Plasma cfDNA libraries using TruePLEX Plasma-Seq kit were analysed on the Agilent TapeStation 2200 to examine size distribution and concentration of the final sequencing libraries.

An example of constructed library for genomic normal and/or tumour DNA is shown in Figure 18, comparing the size and concentration of fragmented DNA (input material) and the final libraries. As shown in the example electropherogram, the peaks in the middle, indicated with *b*, correspond to the post-capture indexed-libraries after amplification. Narrower distribution and escalated concentration of desired fragment size of the input material in comparison to the broader peaks indicated with *a*, which shows the size and concentration of the fragments, was observed. The constructed libraries had a peak of DNA fragments positioned between 250 and 350bp as reference range, as recommended in the protocol. Similarly, the size range of generated cfDNA libraries after measurement on the TapeStation 2200, were concordance with the manufacturer's recommended reference size.

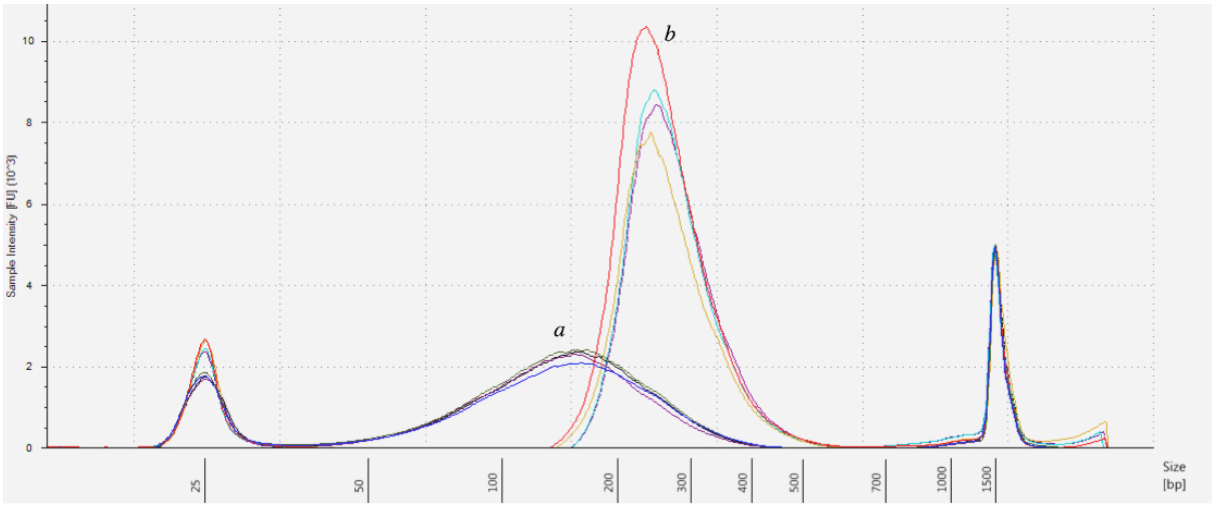


Figure 18. Agilent 2200 TapeStation D1000 ScreenTape assay for one cohort of prepared tumour libraries is shown. Peaks on either end indicate lower and upper markers. Verification of successful library preparation procedure by shift in size distribution and increase concentration of peaks (a) to optimal fragment size of approximately 225 to 275bp (b) was observed. The broader peaks in the middle indicate genomic DNA after fragmentation to target size of 150-200bp.

3.3. Sequencing statistics of high-throughput sequencing data

The whole exome of 28 tumour and normal libraries from high-grade STS patients were sequenced paired-end 2*100 bp on an Illumina HiSeq 2500 or HiSeq 4000 instruments using TruSeq SBS Kit v3. As part of a proof-of-concept experiment, 6 corresponding plasma were targeted resequenced using the same instruments and kits. A mean coverage of 112-fold with 80.8 % of the bases covered >50-fold for normal

libraries was achieved. A mean coverage of 236-fold, 405-fold for the tumour and cfDNA libraries was achieved, with 82.7 % and 97.5 % of the bases covered >100-fold, respectively. An average PCR duplicate rate of 10 %, 12% and 50% were observed for normal, tumour and cfDNA libraries, respectively (Table 4-5). In addition, per base sequence quality, per sequence quality scores, per sequence GC content parameters from FASTAQC reports were enquired.

Table 4. Table represents various sequencing metrics generated by Illumina HiSeq 2500/4000 sequencers. The samples CS01-CS07 are shown. cfDNA from these samples have not been sequenced

Sample	CS01	CS02	CS03	CS04	CS05	CS06	CS07
Read Pairs (Million)							
Normal	39.077	44.484	47.238	36.167	33.739	41.976	40.390
Tumour	80.128	98.100	85.952	81.870	76.527	93.894	88.892
Mean Coverage							
Normal	96.13	106.65	115.57	81.73	78.39	101.25	93.58
Tumour	187.69	233.7	203.79	195.42	175.36	217.82	205.68
% of bases read >100X							
Tumour	76.5	83.8	78.1	76.7	93.5	80.1	80.3
% of bases read >50X							
Normal	78.6	83.6	85	71.3	70	79.9	78.5
~ PCR Duplicate %							
Normal	4	6	5	11	9	8	12
Tumour	11	12	10	8	9	8	8

Table 5. The table represents various sequencing metrics generated by Illumina HiSeq 2500/4000 sequencers. The samples CS08-CS14 are shown. cfDNA from CS10 has not been sequenced.

Sample	CS08	CS09	CS10	CS11	CS12	CS13	CS14
Read Pairs (Million)							
Normal	34.727	37.619	36.955	60.536	131.090	76.315	27.576
Tumour	74.737	81.122	203.606	124.542	125.741	118.088	108.514
cfDNA	45.652	54.123		65.680	71.401	57.651	48.190
Mean Coverage							
Normal	82.72	86.77	89.77	132.97	270.47	166.42	67.86
Tumour	159.68	193.75	440.61	286.56	283.2	272.53	241.49
cfDNA	366.94	165.52		353.92	604.33	435.72	504.42
% of bases read >100X							
Tumour	68.3	73.5	97	90	89.4	86.9	84
cfDNA	97.9	92.9		98.2	99	98.4	98.5
% of bases read >50X							
Normal	71.5	74.9	99.2	88.3	98.3	91.7	60.8
~ PCR Duplicate %							
Normal	9	10	6	15	19	13	12
Tumour	6	14	11	12	17	19	21
cfDNA	48%	63%		60%	44%	50%	38%

Mean coverage (DP) only for samples with sequenced plasma is shown in Figure 19.

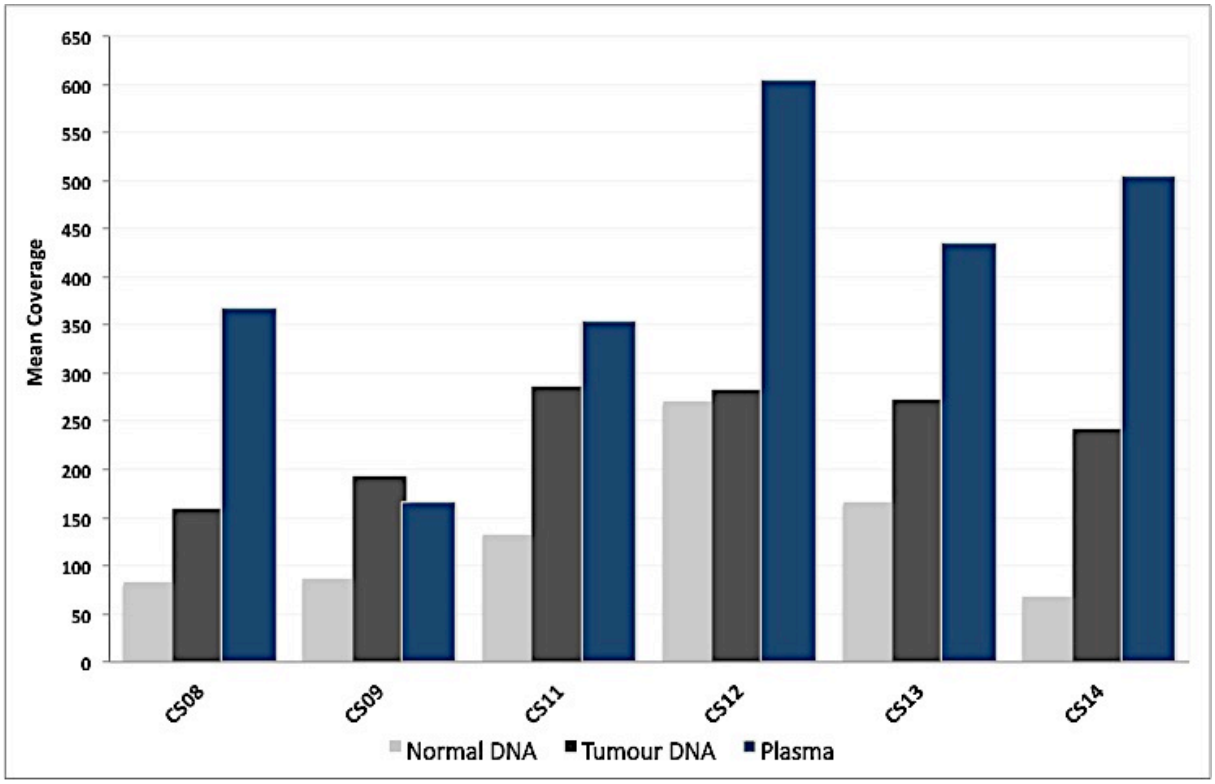


Figure 19. Shown is the mean coverage for normal, tumour and plasma samples for each patient. Plasma samples for all but one patient (CS09) have higher coverage in comparison to matching normal-tumour.

3.4. Somatic mutations

We exome sequenced 14 pairs of tumour–normal libraries from high-grade STS patients. Alterations that were not present in matched normal samples were annotated as somatic mutations. A total number of 14,854 somatic mutations were called by the pipeline. The number of somatic mutations, SNVs and Indels, in each of the tumour samples is represented in Table 6. The number of SNVs across the tumour samples was higher in comparison to the number of Indels. The number of non-coding mutations, within SNVs and/or Indels, was also higher than the number of coding mutations. Across the tumour samples, transversion substitutions occur at a slightly higher frequency than transitions, 9 out of 14 (Table 6).

Table 6. The table represents the samples from which libraries were generated.

Sample #	Blood	Tumour	SNVs	Transitions In SNV	Transversions In SNV	Indels	Coding Region Variants	Non-Coding Region variants
CS01	X	X	147	81	66	4	55	96
CS02	X	X	123	58	65	1	41	82
CS03	X	X	10,991	10,115	876	31	4112	6910
CS04	X	X	205	89	131	7	62	150
CS05	X	X	220	69	121	2	77	145
CS06	X	X	190	91	140	3	83	110
CS07	X	X	231	134	147	5	73	163
CS08*	X	X	281	111	103	1	97	185
CS09*	X	X	214	102	134	3	80	137
CS10	X	X	236	336	234	2	43	195
CS11*	X	X	570	111	155	39	112	221
CS12*	X	X	266	188	184	12	63	206
CS13*	X	X	373	275	340	26	75	323
CS14*	X	X	615	89	131	57	207	465

*Asterisk * indicates patients for which the plasma cfDNA was sequenced. Blood represents normal gDNA libraries, whereas tumour represent tumour gDNA libraries. The tumour's somatic variants in different categories are shown for each patient. The total number of coding and non-coding variants is equal to sum of SNVs and Indels in each sample. X indicates samples that have been used for library preparation and sequencing.*

For all sequenced tumours, somatic mutations in the context of A) transition/transversion, B) coding/non coding and C) variants class are shown in Figures 20—23. The number of somatic variants in sample CS03 was higher than the other 13 tumour samples; 11,022 somatic mutations in comparison to the average of 295 variant in other 13 tumours. This particular case has been further discussed in more detail. Similarities in the proportion of transversions or transitions substitutions across the tumours were observed (Figure 20).

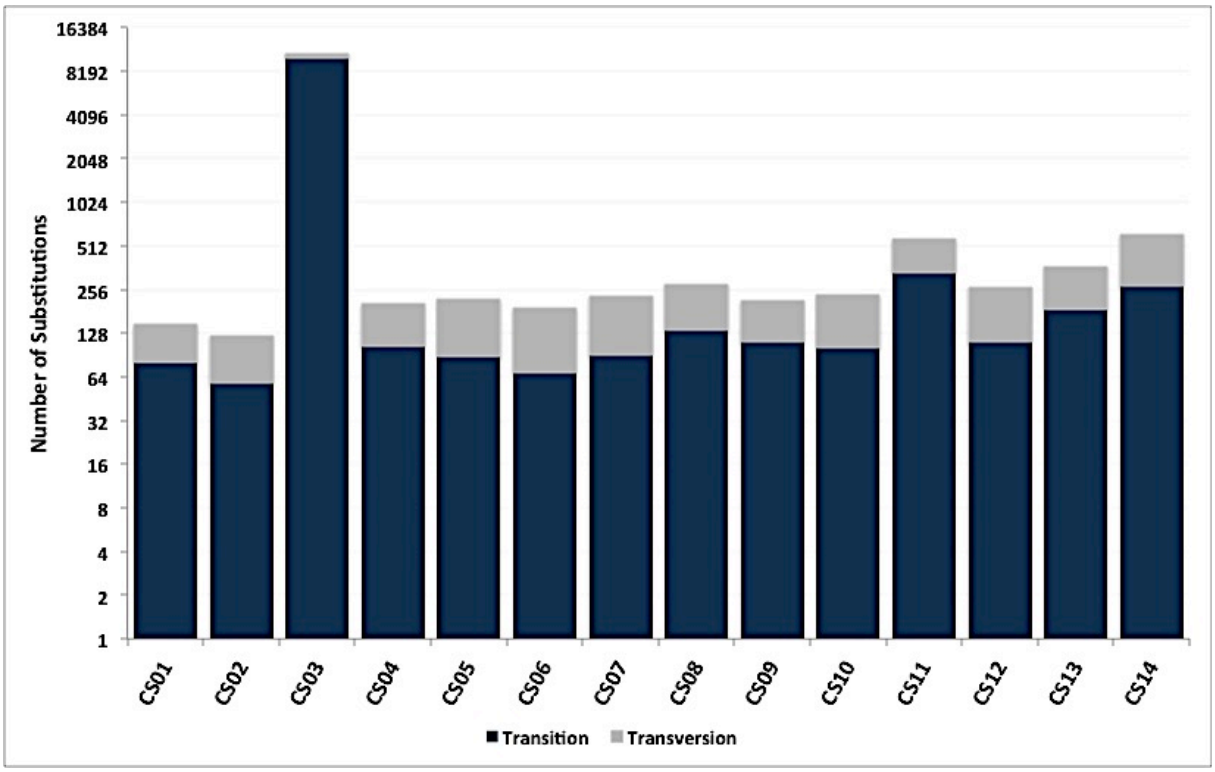


Figure 20. Proportion of SNVs across the tumour samples is shown. Total base changes in sample CS03 was strikingly higher in comparison with other samples. Y-axis is shown in logarithmic scale.

Among the substitutions, G>A:C>T substitutions were observed at ~57 % of the tumour samples, 8 out of 14 tumours (Figure 21). The observed pattern showed concordance with major mechanism for new mutations, which is deamination of 5'-methyl C, resulting in higher proportion of substitutions mentioned above. The average number of SNVs in comparison with the average number of indels, was higher by 76-fold across the tumours' mutations (Figure 22).

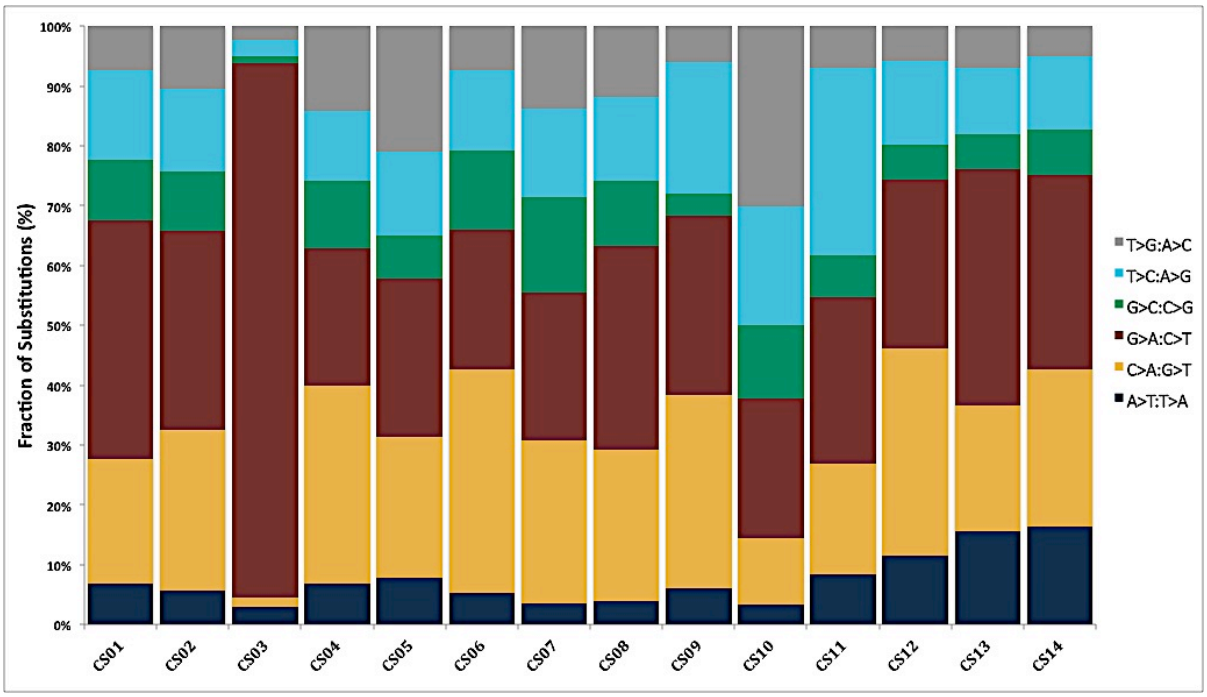


Figure 21. The proportion of base change patterns within and between the tumour samples is shown. The G>A: C>T transversions were dominant base change. The prevalence of C>A:G>T transversions in the 5 tumours, CS04, CS06, CS07, CS09, CS12 were predominant substitutions. T>G:A>C substitutions was the major base change in sample CS10.

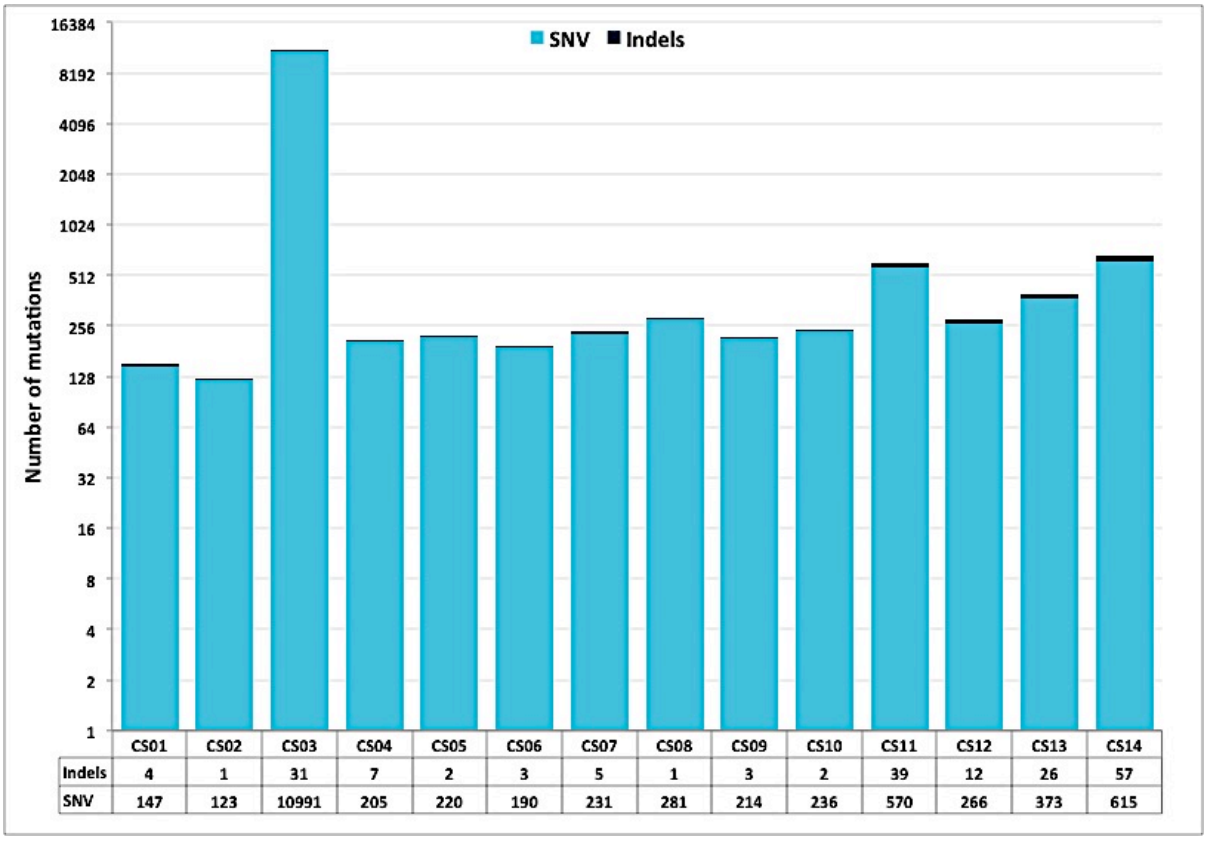


Figure 22. The number of SNVs and indels within and between tumour samples is shown. Y-axis is shown in logarithmic scale.

Detected somatic mutations were divided into coding and non-coding groups. Missense, nonsense, splice-site acceptors/donors, synonymous, and frame shift variants were defined as coding region mutations, whereas intronic, intergenic, UTRs and splice region mutations, excluding splice sites, were considered as the non-coding region mutations. On average, one-third of all detected somatic mutations in the tumours were grouped within the coding regions of the genes in each sample (Figure 23).

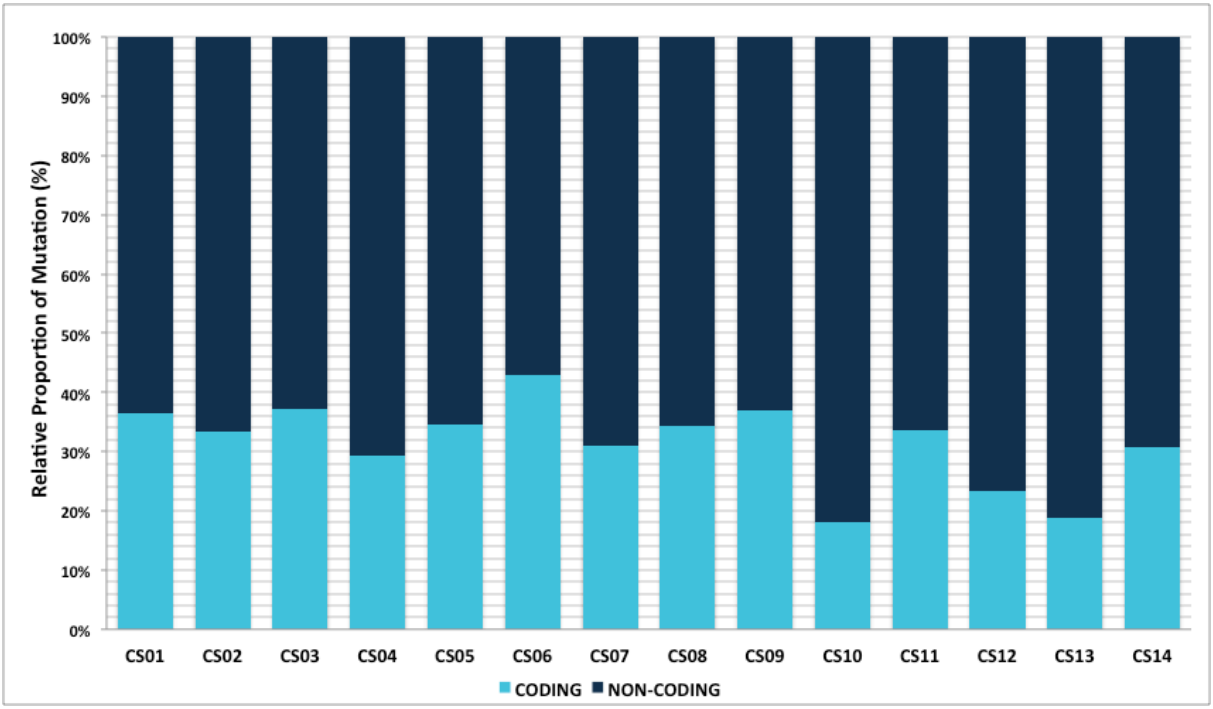


Figure 23. The proportion of the somatic mutations in the coding region in comparison to the non-coding region mutations is shown. More than two thirds of somatic mutations in the tumours occurred in non-coding regions. The data is presented at 100% scale.

Due to the considerable number of mutations in patient CS03, who had been previously diagnosed with UPS, the extended analysis of mutational signature was performed and signatures S7 and S1 were revealed, as defined in (104). In this analysis, substitution class, the sequence context at 3′ and 5′ of mutated base, and the transcribed strand on which mutation occurs were considered (Figure 24-25).

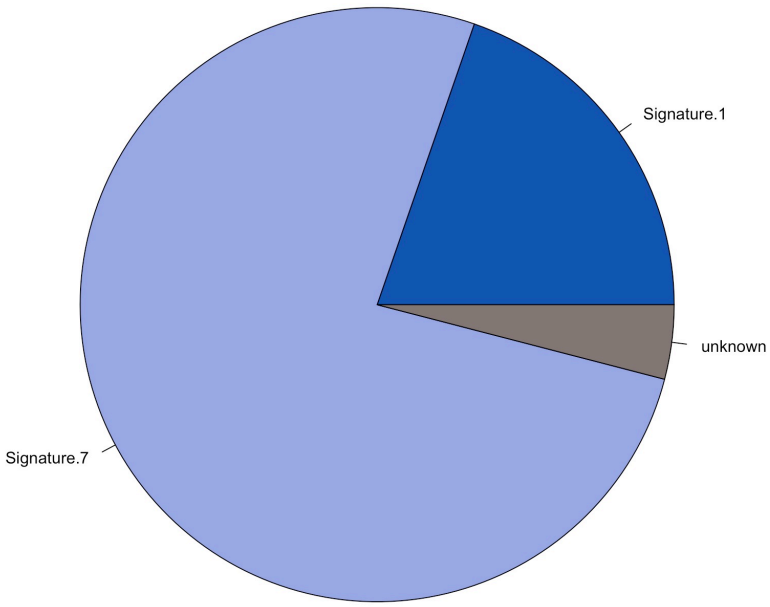


Figure 24. Figure represents concluded proportion of each correlated signatures from the analysis of mutational signatures, based on mutational catalogue of cancer. The signature 1 is characterized by considerable C>T substitutions and has been observed in almost all cancer types. The signature 7 with similar prevalence of C>T substitutions, has been demonstrated to predominantly correlate with malignant melanoma.

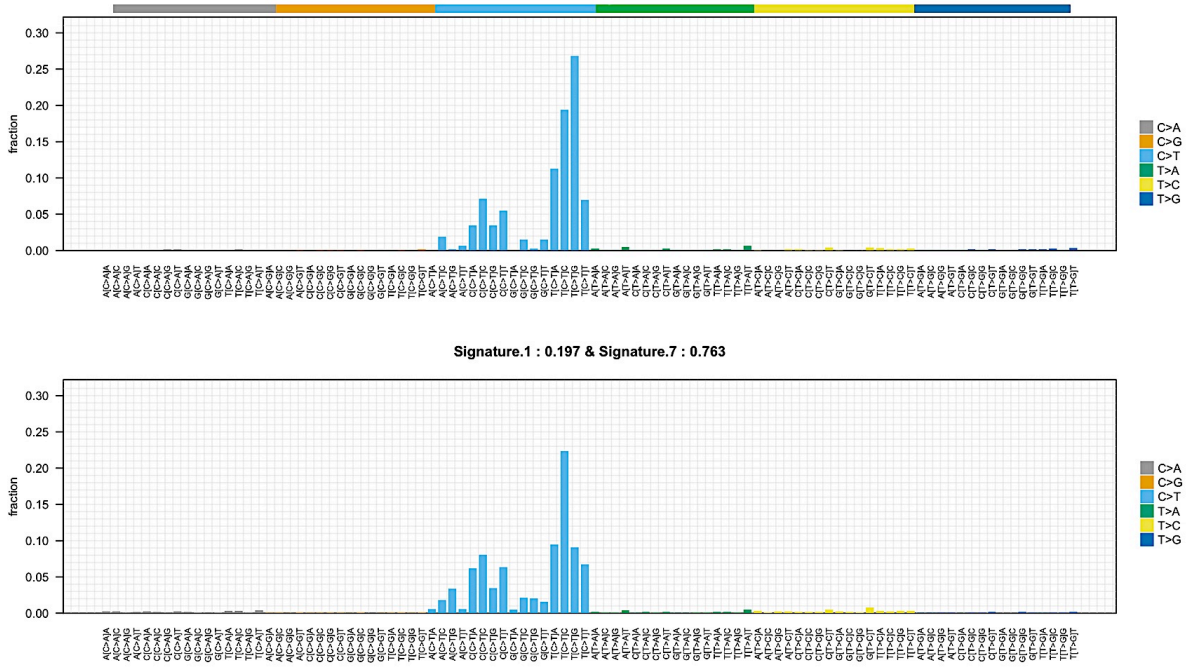


Figure 25. Mutational signature analysis in the patient CS03 concluded signatures S1 and S7.Proportion of C>T substitution in TCC, TCA, and TCG trinucleotides demonstrate higher ratio.

In comparison between the 14,854 total numbers of somatic variants across all the tumours and a total of 459 cancer-driver genes retrieved from IntOGen databse, 100 known cancer driver genes were found mutated within the coding regions of genes across all the tumours. Noteworthy, 61 of these 100 cancer genes were detected in the tumour sample CS03. Thirteen cancer driver genes, *DHX9, TRIP10, SETDB1, TSC1, FAT1, ASH1L, PTPRF, MED12, TCF4, RB1, TAOK1*, *TP53* and *PCSK5* were observed within more than one tumour. *TP53* mutations were detected in 7 out of 14 tumour samples, where 6 of these mutations were located in the DNA-binding domain of the gene, and the remaining was located in the tetramer domain of the protein (Table 7). In addition, mutations in *NRAS* and *NOTCH1*, which are among frequently mutated cancer driver genes across various cancer types, were also identified among the somatic mutations.

Table 7. Mutations identified in TP53 gene are represented.

Mutated *TP53*	AA change	Consequence
g.chr17:7577597 GT>G	*p.D228fs*	*Frame shift*
g.chr17:7578503C>T	*p.V143M*	*Missense*
g.chr17:7578413C>T	*p.V173A*	*Missense*
g.chr17:7578406C>T	*p.R175H*	*Missense*
g.chr17:7578413C>A	*p.V173L*	*Missense*
g.chr17:7574013_7574013G>GA	*p.F338fs*	*Frame shift*
g.chr17:7578212G>A	*p.R213**	*Nonsense*

*AA: Amino Acid. § Indicates the samples for which the somatic mutation has been identified, but not been confirmed in previous studies. Asterisk * indicates a stop codon.*

3.5. Assessment of plasma cfDNA mutations

Somatic mutations in tumours and plasma ctDNA were identified following the strategy described below for 2 datasets; one dataset of tumour samples and one dataset for plasma samples (Figure 26). Sequencing libraries for 6 matched tumour-plasma pairs, out of 14 sequenced tumour samples, were generated. Through the pipeline used for tumour analysis, germline variants in tumour and plasma were excluded using the normal sample from the same individual as a control, generating datasets of somatic mutations. To each datasets, the two following filters were applied; strong call confidence, or non-synonymous variant. The remaining mutations were then manually investigated by IGV.

From the 6 tumour pairs, sequenced by exom libraries, 2,457 somatic variants were identified. Of these, 164 variants coincided with the genes in the NCGC 900 gene panel, used to sequence the corresponding plasma samples. By applying the filters mentioned above, a total of 80 somatic calls remained. From the 6 plasma pairs, sequenced using the 900 genes panel, 828 somatic variants were identified. Of the 80 variants identified in tumour, 26 mutations were also overlapping with the somatic mutations identified in the plasma samples. By manual inspection of the plasma BAM files (criteria explained in methods section), another 5 variants that had not been called through the pipeline were identified in the plasma. In total, 31 variants were detected to be present both in tumour and plasma from this investigation, and the remaining 49 variants were only present in the tumours, but not in the plasma samples.

In a similar strategy described for the tumour samples, the 828 identified variants in plasma were filtered, leaving only strong calls and/or non-synonymous variants. Subsequent to excluding the variants that had already been detected in both tumours and plasma, a total of 405 somatic mutations remained. Manual investigation of the BAM files enquired 288 somatic variants that have been called correctly, being present only in plasma ctDNA, while the remaining 117 confirmed in neither plasma nor tumour samples. This sum has been broken down for each patient in Table 8.

Table 8. The number of variants that were detected in plasma after applying the filters is shown.

Plasma sample	**Initial called variants (total 405)**	**Detected variants (total 288)**	**Mean DP**
CS08	56	29	370
CS09	24	24	167
CS11	24	9	353
CS12	242	195	573
CS13	23	19	342
CS14	36	32	444

The number of detected variants indicates those variants that in the manual investigation implied to be real variants, due to the fact that no apparent artefacts was observed. The mean coverage (DP) of only detected variants in each patient is represented.

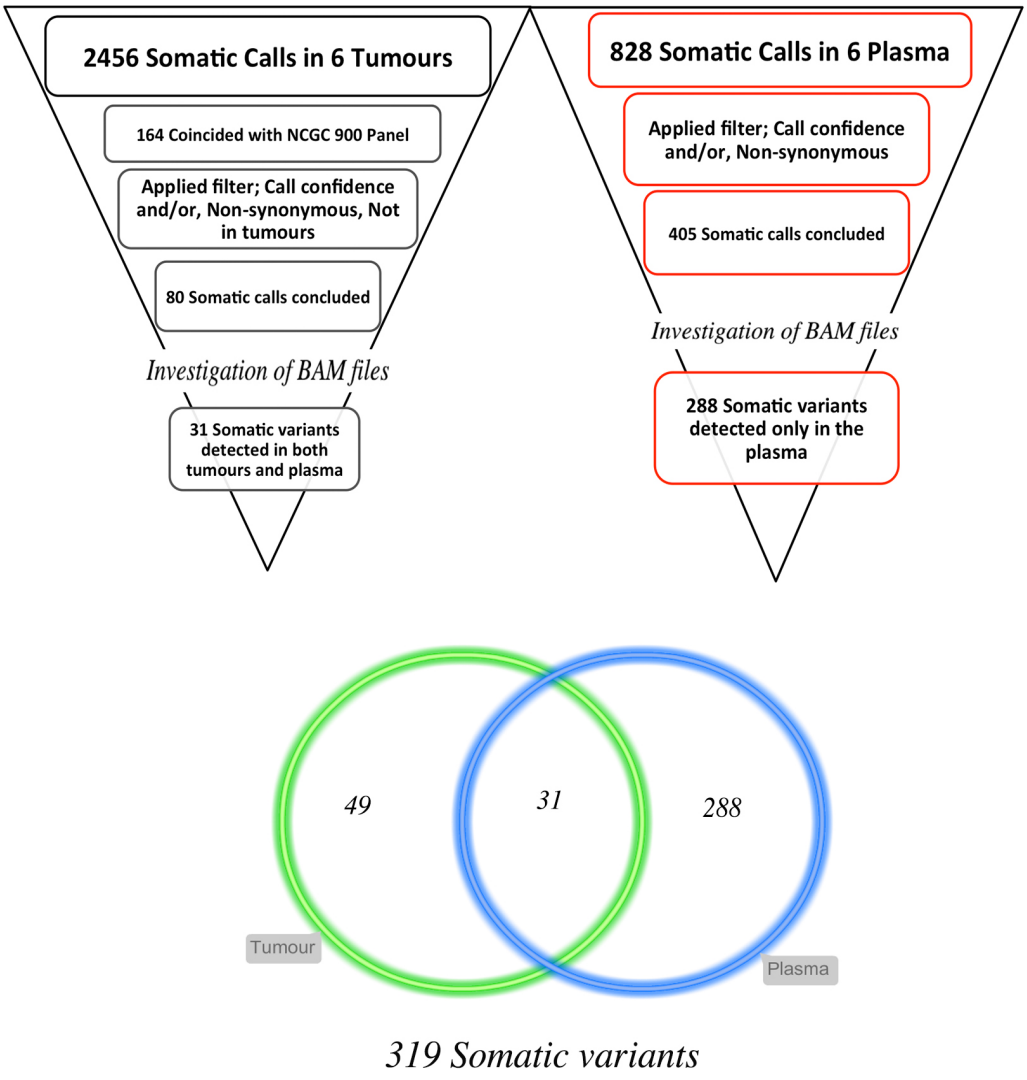


Figure 26. The strategy for detection of represented number of somatic mutations in the plasma ctDNA is shown. For details refer to the text. T+P indicates proportion of variants that has been detected in both tumour and plasma samples, whereas P indicates the somatic variants present only in the plasma.

In total, 319 somatic mutations were detected, 31 in tumour and plasma and 288 only in plasma (Table 9). It is noteworthy to mention that among the mutations that were present only in plasma, two-thirds (195/288) was found in the patient CS12, with an aggressive cancer type and metastases described at the time of diagnosis. Only some of the detected somatic mutations for patient CS12 are shown in Table 9. Interestingly, mutations in several known cancer driver genes, including *EGFR, FGFR1, FGFR3, ERBB2, MTOR, MAP3K4, SETD2* and *ARID1B* were detected only in the plasma ctDNA somatic mutations. *FGFR3* had been previously reported (COSM5504814) in COSMIC database and *MTOR* (p.E2264X) and *ERBB2* (p.P944H) was detected in plasma from 2

patients. Other instances of detected mutations in plasma that were detected in 3 or more patients are; *MAP3K4* (p.P1205H), *ARID1B* (p.P426H), *DYRK1B* (p.G570C), *CDH20* (p.P195), *KM2TB* (p.P1045T), and *FCGR2A* (p.P47T).

Table 9. Detected somatic mutations in the plasma ctDNA for each patient are represented.

Sample	Detected somatic variants in both tumour and plasma ctDNA	Somatic variants which were only present in the plasma
CS08	*PTPRF,* ***EPHB6, TP53***	***ARID1B, MXRA5, FCGR2A, KMT2B, CRTC1, DYRK1B, POLD1, FLT4, NLRP7, ITPKB, MAP3K4, LCTL, ASXL2, PLCG2, PTCH1, FLNC, SMURF1, PIK3C3, COL2A1, MAP2K5, MLH3, MUC17****, ALK, ALOX12B, TBX3, RET, MUC16* ***KMT2D*** (chr 12:49422844 C>A) ***KMT2D*** (chr 12:49420390 C>A)
CS09	***NRAS***	***ADAMTS20, PRG4, GNAS, KMT2C***
CS11	***MAP3K9****,* ***TP53****,* ***PIK3R5****, CDH2, CNBD1, LRP2,* ***RUNX1****, TP63, IGF2, MST1 (chr 3:49723274G>A), BMPR1B,* ***MXRA5****, SFTPA1,* ***MED12****,* ***MST1*** *(chr 3:49723881G>C),* ***RYR2****,* ***ACVR1***	***ARIBD1****,* ***TSC1****,* ***KMT2B****,* ***CDH20, FCGR2A, MAP3K4,*** *CACNA1E, ALK, AZGP1*
CS12*	***TP53****,* ***CIC****,* ***RB1****, GAB2*	***DNM2, CDH20****,* ***FANCA****,* ***MAPK7****,* **PRKCA**, ***LRP2****,* ***FGFR1****,* ***FGFR3****,* ***EGFR****,* ***CHEK1, MTOR, KMT2B*** *(*chr *19: 36214707C>A),* ***ERBB2*** *(*chr *17:37879825C>A),* ***ERBB2*** *(*chr *17:37882065C>A),* ***AXIN1*** *(*chr *16:339531C>A),* ***AXIN1*** *(*chr *16:339498C>A),* ***CIC*** *(*chr *19:42797844C>A),* ***SETD2*** *(*chr *3:47098557C>A),* ***SETD2*** *(*chr *3:47139564C>A)*
CS13	***RYR2****,* ***PKHD1****,* ***TP53****,* ***TAOK1, RB1***	***POU2AF1, CHD4, LCTL, ZNF276, CRTC1, DYRK1B, CACNA2D3, ROBO2, SLIT2, ESR1, MAP3K4, PTCH1****,* ***ARID1B****,* ***KMT2B,*** *TBX3, AKT2, FLT4, CARD11, SMARCA2*
CS14	*PARP4*	*ESR2, PML, GREB1, FLT4, IRF4, XRCC2, CARD11****, MTOR, ARID1B, CACNA1C, CDH20, DYRK1B, ELK3, EPHB6, NUP93, ERBB2, ZNF276, FUBP1, DNM2, NOTCH3, HLA-B, KMT2B, KMT2D, LCTL, MAP3K4, NEIL2, NLRP8, PARP10, RAD51, ROBO2, ZNF180, PTCH1***

Exonic mutations are presented in bold. Not all the detected plasma somatic calls for patient CS12 are shown. A considerable number of variants in CS12 were detected more than one time in a gene.

The average AF and mean coverage (DP) for the detected somatic mutations is shown in Table 10. The somatic calls were divided in 3 series; series 1 included somatic mutations that were detected both in tumour and plasma (a total of 31), series 2

includes 49 tumour mutations, out of 80 somatic mutations, that were identified in the tumour but not in the plasma, and series 3 includes mutations that were only detected in the plasma ctDNA (a total of 288).

Table 10. The table represents the mean and standard deviations (STDEV) values of AF and DP for 3 data series

Mutation series	Mean AF (P)	Mean AF (T)	Mean DP (P)	Mean DP (T)
Series 1	18% ± 14%	43% ± 25%	325 ± 159	175 ± 105
Series 2		18% ±13 %	377 ± 227	243 ± 220
Series 3	4%		511 +/-215	279 ±164

Series 1 includes the somatic mutations that were detected in both plasma and tumour (the total of 31). Series 2 represents mutations among the 80 tumour mutations that were detected only in the tumours. Series 3 includes mutations that were detected only in the plasma ctDNA. (P) and (T) indicate plasma and tumour, respectively.

Among the somatic mutations that were detected in both tumour and plasma ctDNA, 6 deletions were identified; two deletions at the exon 3 of *RB1* gene in the patients CS12 and CS13 were observed. In the patient CS12 a deletion in *GAB2* gene was also detected. In the patient CS11, a deletion in *MED12*, and a deletion in *RYR2* were found. An insertion in *TSC1,* resulting in a frame shift elongation, which putatively increases the risk of developing a rare and sporadic form of sarcoma known as PEComa at patients CS06 was identified. Interestingly, a frame shift insertion with 80 G>GTC base insertion in tumour suppressor gene *TSC1* (p.D510fs) were detected only in the plasma ctDNA from patient CS11.

The allele frequency (AF) of detected mutations in plasma ctDNA in comparison to matched tumour were commonly lower and absent in the matched normal DNA. However, for a few mutations that had been called through the pipeline, a relatively high AF in the normal sample was observed. In Figures 27—33, tumour somatic mutations are shown. In general, AF in tumour was observed to be higher in comparison to plasma.

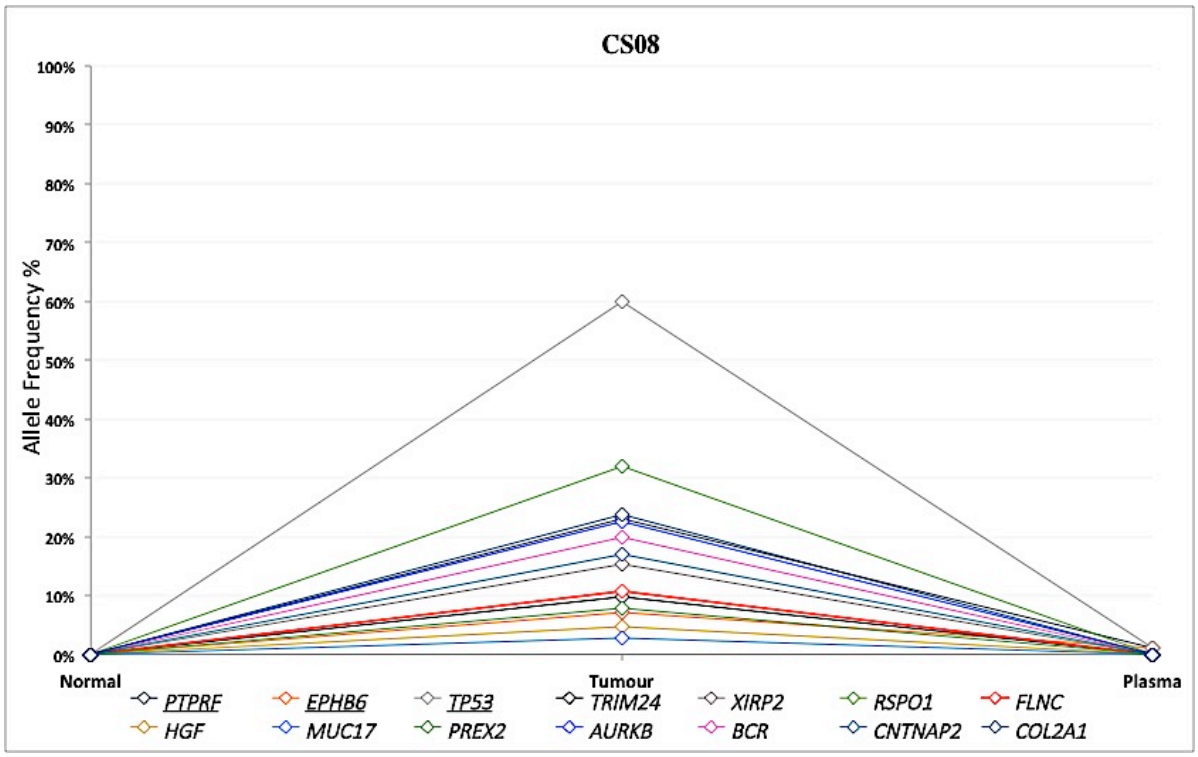


Figure 27. Somatic mutations in the patient CS08 is shown. Underlined genes refer to the mutations that were found in both tumour and plasma.

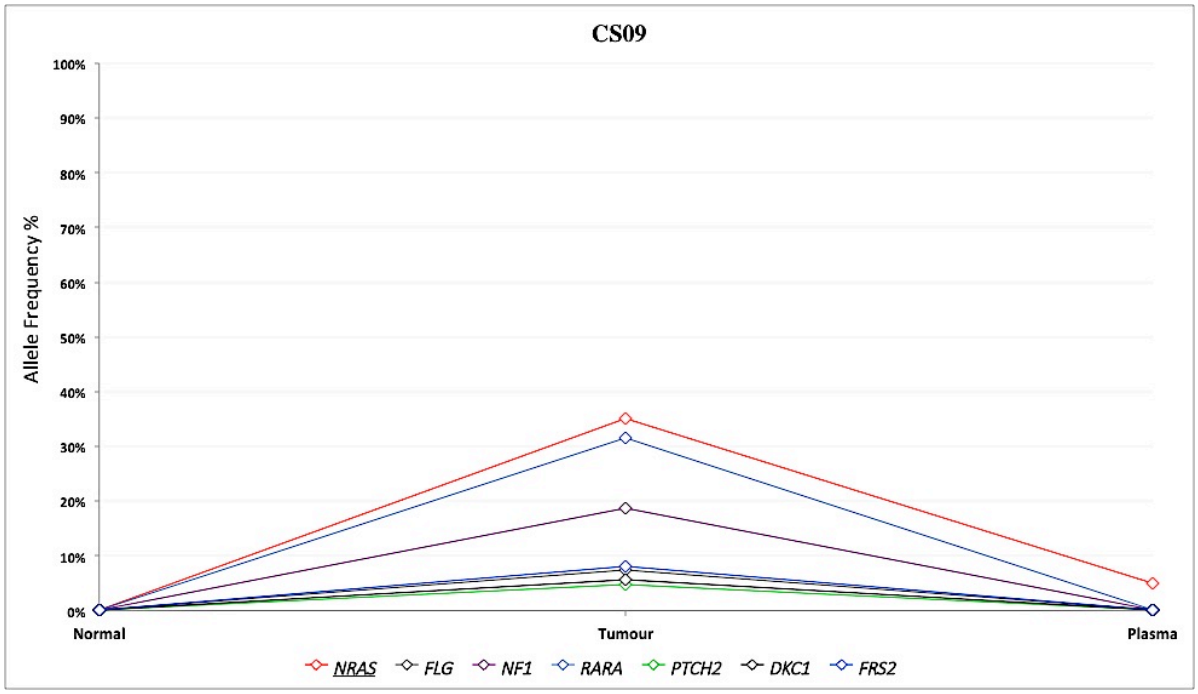


Figure 28. Somatic mutations in the patient CS09 is shown. Underlined genes refer to the mutations that were found in both tumour and plasma.

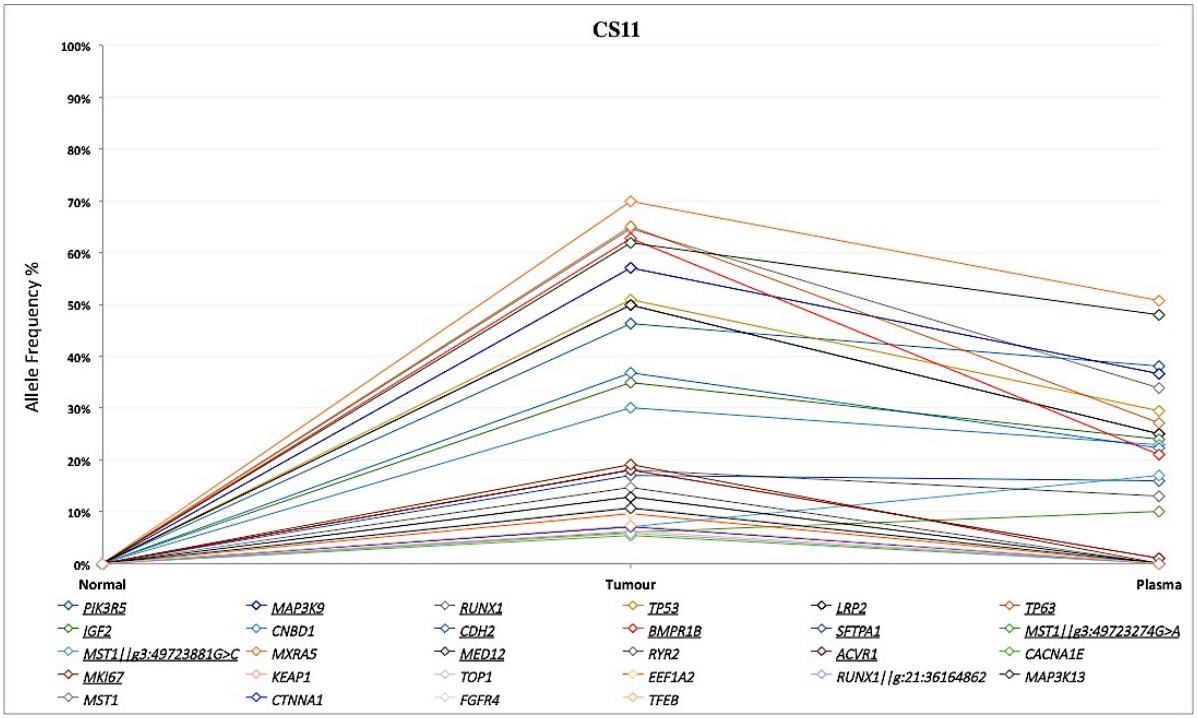


Figure 29. Somatic mutations in the patient CS11 is shown. The genes with more than one mutation are presented with genomic positions. Underlined genes refer to the mutations that were found in both tumour and plasma.

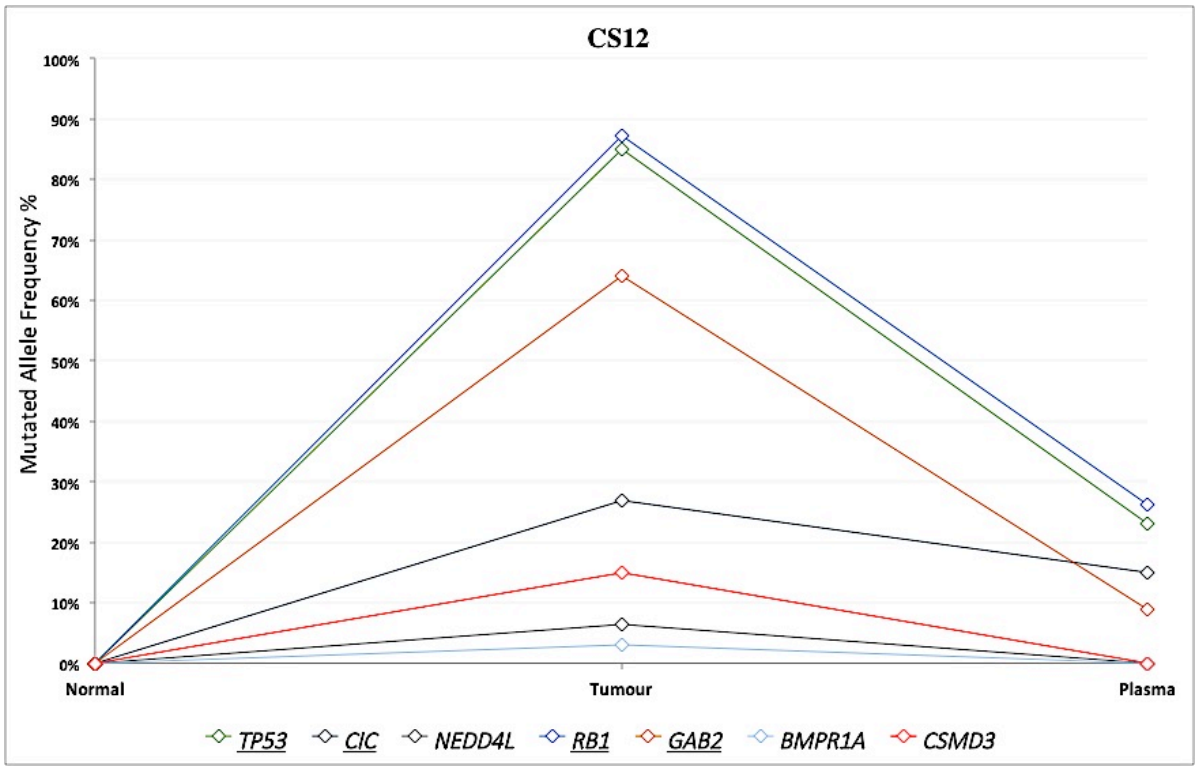


Figure 30. Somatic mutations in the patient CS12 is shown. Underlined genes refer to the mutations that were found in both tumour and plasma.

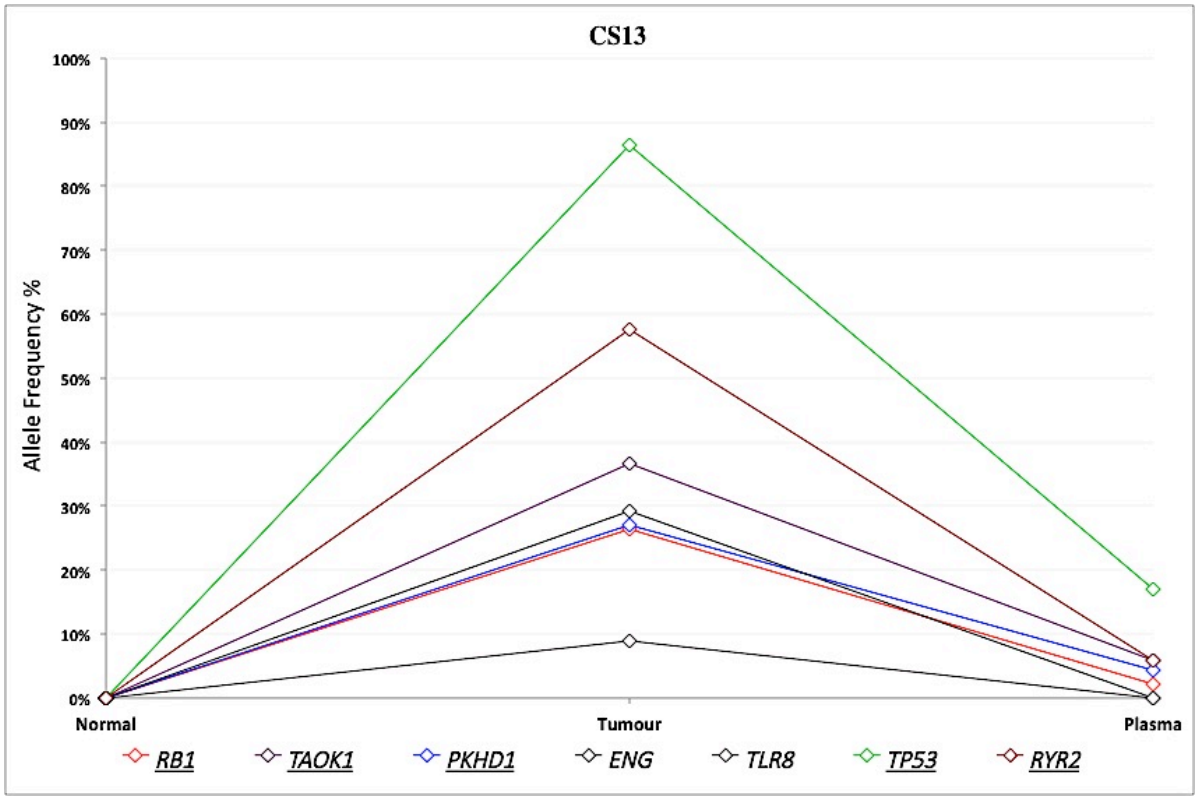


Figure 31. Somatic mutations in patient CS13 is shown. Underlined genes refer to the mutations that were found in both tumour and plasma.

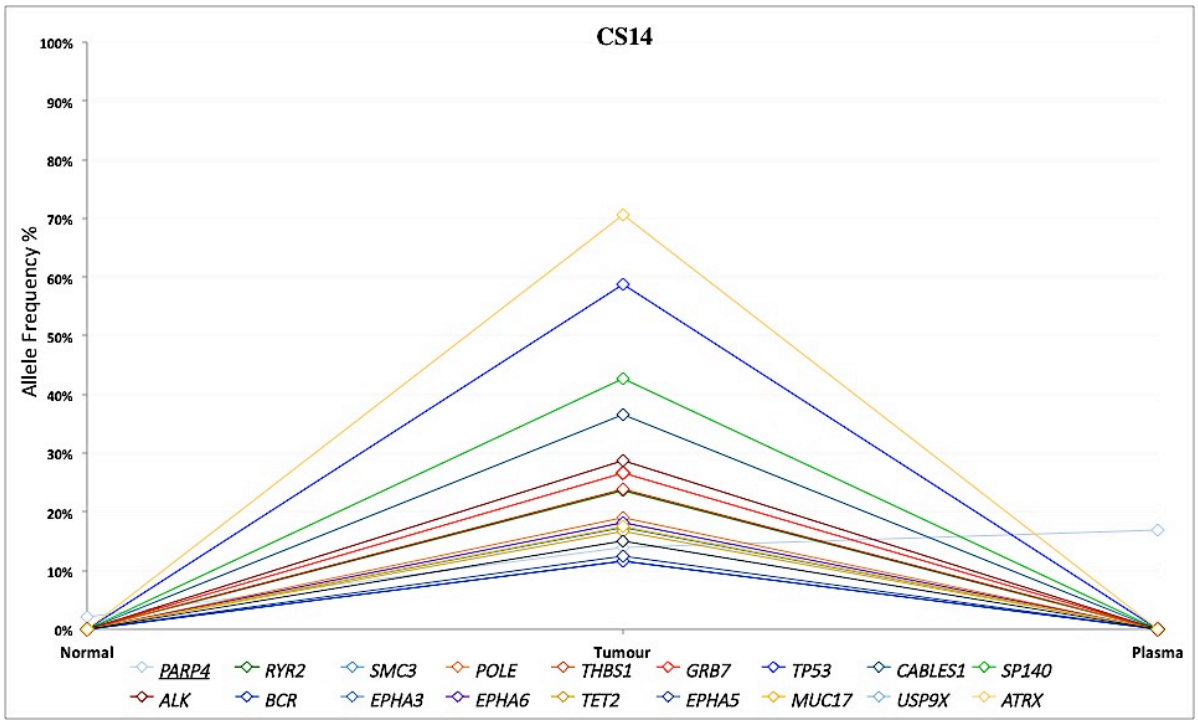


Figure 32. Somatic mutations in patient CS14 is shown. Underlined genes refer to the mutations that were found in both tumour and plasma.

4.Discussion

Soft-tissue sarcomas are a rare and heterogeneous group of mesenchymal tumours with more than 50 histologically distinct subtypes with high variation in their clinical features. Many soft tissue sarcomas will recur locally and/or metastasize, and the median time to local recurrence is 12 to 18 months and to metastasis around 1 year (40). The overall five-year survival of all soft tissue sarcoma patients is approximately 50% (40) and the long-term survival of patients with metastatic disease is poor. Current monitoring strategies, i.e. imaging approaches, do not possess optimal sensitivity to detect micrometastasis and/or deep-located small tumours, which as a consequence may result in inaccurate diagnosis and prognosis of the cancer during a course of therapy (41, 51). Thus, there is a clinical need to develop surrogate approaches in order to improve accuracy of assays for early detection, prognosis and monitoring of STS patients. Owing to the advances in the genomic tools during past decades, monitoring of cancer patients with high sensitivity and specificity in the context of liquid biopsy has been gained traction. It has been demonstrated that spatial and temporal heterogeneity of the tumours can be depicted by serial liquid biopsies [reviewed in (80, 85, 95, 99, 105)]. In this project, we generated 34 libraries from 14 genomic normal—tumour pairs and 6 corresponding cfDNA samples. By WES and targeted resequencing we aimed to detect tumour-specific mutations in the tumour and identify these in circulating tumour DNA from high-grade STS patients, which be further utilized as a personalized biomarker for tumour burden monitoring in STS patients.

4.1. Detected somatic variants in the tumour samples

We sequenced 14 soft tissue sarcoma tumours using WES, and identified somatic mutations. By comparing the somatic mutations in our tumour dataset with a cancer driver gene dataset retrieved from IntOGen (19), 100 cancer driver genes were found mutated in their coding region, across the 14 tumour samples. Although, biological aspects and pathogenesis of discovered genes in our samples are not a main objective

of this study, genes with diagnostic and/or prognostic values related to STS present in our samples will be briefly discussed.
TP53 tumour suppressor gene, frequently mutated in osteosarcoma and STS (31, 36), was found to be mutated in 7 out of 14 tumour samples. Mutations in *RB1* were also detected in 3 of the patients, which this will be briefly discussed further in the text.

It is noteworthy to mention that none of the mentioned cancer driver genes were detected in the patients CS02 and CS04. Both patients were diagnosed with liposarcoma. The patient CS02 was diagnosed with dedifferentiated liposarcoma, and the patient CS04 had a pleomorphic liposarcoma that is the rarest subtype of liposarcomas with complex structural rearrangements and high risk of local recurrence and metastasis. Oncogenic activity of *TP53, NF1, RB1, MDM2, CDK4, HMGA2 and TSPAN31* in these subtypes have previously been demonstrated [reviewed in (106)], however they were not identified in our tumour samples. Nevertheless, a subset of these tumours carry frequent amplification of *CDK4* and *MDM2*, and these tumours have normal *TP53* and *RB1* [reviewed in (106)], which may explain the absence of mutated cancer driver genes in these patients

A total of 3 non-synonymous mutations were identified in *TSC1,* missense (p.D964N), insertion (p.L557fs), and insertion (p.D510fs) mutations, identified in patients CS03(UPS), CS06(malignant PEComas) and CS11(LMS), respectively. The somatic insertion (p.D510fs) was only observed in the plasma of patient CS11. Patients with oncogenic mutations in this gene are thought to develop a family of perivascular epithelioid cell tumors, including angiomyolipoma, lymphangioleiomyomatosis (LAM) and PEComa (107). Alternations in *TSC1* and *TSC2* have been found to impact the mTOR-signalling pathway (mTORC1) in patients diagnosed with PEComa, with no currently effective therapy for malignant PEComas (38).

Revealing mutated *TSC1* gene as an underlying causative of the cancer is clinically significant since it may guide the treatment in patients harbouring this mutation. In a study by Wagner et al., they observed significant clinical responses in three patients treated with an off-label inhibitor of mTORC1 protein, sirolimus, which has alerted expression as a result of loss of function in *TSC1/TSC2* tumour suppressor genes (108).

Interestingly, a frameshift elongation mutation in *TSC1* (p.L557fs) with allele frequency of 43% was identified in the tumour somatic mutation of patients CS06, which had histologically been diagnosed with PEComa subtype. A deletion and missense mutations at the denoted codon had previously been reported for bladder cancer (COSM28285) and adenocarcinoma (COSM3905173) in the COSMIC database.

Based on the information give above, distinguishing the subtype of STS in patients harbouring mutation in *TSC1* seems to be feasible where the clinical descion making based on histopathological results was difficult, as was the case for the patient diagnosed with malignant PEComa. Mutated *TSC1* in the patients CS03 and CS11 was also identified, but no damaging or deleterious impact of these mutations had been previously reported.

The mutational profile of patient CS03 deviated from the other samples, having 40-fold more somatic mutations and far more transversions (G>A:C>T substitutions) than transitions. In an extended analysis of tumour's mutational signature based on mutation catalogue of Alexandrove et al. (104), signatures S1 and S7 were revealed. The mutational process computes mutational signatures on the basis of the trinucleotide frequency of the human genome, and the overly of calculated probability against the human genome trinucleotide frequency precludes various signatures, operative across cancer types. The signature 1 is characterized by considerable C>T substitutions at NpCpG trinucleotides, in virtually all cancer types analysed. It has been postulated that the relatively elevated rate of spontaneous deamination of 5-methyl-cytosine associates with aging, as an underlying mechanism of this substitution. This finding correlated with the patient's age. The signature S7, which exhibits large numbers of C>T substitutions, has previously been observed in squamous carcinoma of the head and neck and strongly correlates with malignant melanoma. It has been demonstrated that the manifested pattern in this signature correlates with the mutation pattern of UV-induced DNA damages, which mainly causes formation of pyrimidine-pyrimidine photodimers. Moreover, signature S7 exhibits a strong transcribed strand bias, indicating that mutation prevalence between transcribed and untranscribed strands varies, with higher prevalence of C>T mutations on the

untranscribed strand (104). This observation and concluded mutational signatures led us to postulate that the cancer may be a malignant melanoma rather than UPS.

However, due to the fact that UV-induced lesions thought to mainly be repaired by nucleotide excision repair pathway during transcription (25), we sought mutated DNA repair genes in the identified somatic mutations of the patient. The phenomenon that cancers with defective DNA repair manifest considerable number of mutations has been previously described (2). The list of DNA repairs genes from (http://www.dnarepairgenes.com/), and the most frequently mutated genes in malignant melanoma were extracted from tumour portal (109), and COSMIC databases (110), respectively. Interestingly, 21 DNA repair genes, among a list of 376 genes, were found mutated in the sample. In addition, in a search for frequently mutated genes in melanoma, a total of 14 genes, *FAT4, GRIN2A, ROS1, KMT2C, KDR, KMT2D, CARD11, RAC1, XIRP2, PPP6C, LCTL, ACO1, ANK3,* and *MXRA5*, found at the tumour portal and COSMIC databases, were mutated in tumour CS03. Based on this observation, the presence of considerable number of mutations in the DNA-repair genes and genes that are reported to be frequently mutated in malignant melanoma may further explain the cancer type. As well, observed striking number of detected mutations in this tumour may be explained with the fact that several DNA-repair genes had mutated in this tumour. However, this hypothesis needs to be further investigated and validated.

Patient CS03 was diagnosed with a UPS. UPS is a very heterogeneous group of STS tumour, and is the given diagnosis for any STS that manifest no identifiable line of differentiation. Further investigation of the tumour clinical characteristics, such as tumour localization (subcutaneous) and histology could indicate the chance of undifferentiated melanoma. The patient was recently presented with a local recurrence at the site of the primary tumour. As a consequence of our findings, histopathological examination was performed once more for both the primary tumour and the recurrence. The immunohistochemical panel that included markers for melanoma, S100, Melan-A and SOX-10 and HMB45, which were negative and consequently the tumour tissue could not be reclassified as a melanoma, but may be given a diagnosis as "possible UPS or melanoma". As a follow-up, we will do RNA-Seq of the tumour material as an appropriate alternative way to investigate whether melanoma gene

expression signatures are present in the tumour. A change in diagnosis from UPS to malignant melanoma would change the treatment possibilities for this patient in a metastatic setting.

4.2. Detected somatic variants in the plasma ctDNA

One of the goals of this study was to evaluate detection of tumour-specific mutations in plasma cfDNA. In an analysis of six tumour—plasma pairs, a total number of 319 somatic variants were detected in plasma.

Initially, 80 somatic mutations were detected in tumour (within the NCGC 900 gene panel), and among these 31 somatic mutations were detected to be present in both tumour and plasma. Through a manual investigation of mapped reads in IGV, in order to avoid erroneous calls, absence of variants in the matched normal, quality of reads in the targeted region, and presence of variants on both forward and reverse strands was taken into the consideration. Some mutations were detected in the tumour, but were not automatically identified to be present in the corresponding plasma through the pipeline, which we reasoned that the low AF of these mutations in the plasma ctDNA is the probable explanation.

Among the 31 somatic mutations, frameshift deletions in exon 3 (p.Q93fs) and 7 (p.F226fs) of *RB1* were found present in both tumour and ctDNA. A point mutation in the same position of exon 3 has previously been reported and confirmed in one breast cancer sample (COSM1477250). A mutation at the same position in exon 7 has been reported in a prostate cancer sample (COSM1470706). Aberrations in the *RB1* gene are a well-known genetic risk factor for development of osteosarcoma and have also frequently been observed in LMS (34, 36). Both patients harbouring deletions in *RB1* had interestingly LMS tumour type. In addition to mentioned deletions, a stop gained mutation in *RB1* (p.C61*) in tumour CS06 was detected for which cfDNA had not been sequenced were manifested both in tumour and ctDNA. The fact that mutations in *RB1* play a role in driving cancer and can be detected in cfDNA, make this gene a good biomarker for investigation of disease burden after surgical excision of tumour.

Mutated *TP53* in 4 of plasma samples, p.V173M, p.R175H, p.V143M, p.V173L were detected.

The plasma cfDNA was sequenced to a mean coverage of 405-fold, where the 6 corresponding tumour were had a mean coverage of 135-fold. Nevertheless, several mutations were only detected in tumour and not in plasma. The mutations that were detected in both tumour and plasma manifested a trend of higher average AF in tumours (43%) than those that were only detected in tumours (18%). We were able to detect mutations with an average AF of around 18 % in plasma, with probably less ability to pick up those at lower frequencies although mutations with AF down to as low as 1% were also detected.

Of the 288 somatic calls detected only in plasma, as many as 195 belonged to patient CS12, which was obviously higher than the average number of somatic variants in the plasma from the other patients. The patient had been diagnosed with high-grade aggressive LMS characterized with metastasis at the time of diagnosis and new metastases appeared few months later, and the patient finally died of the disease. However, one may argue that the observed mutation may be due to false-positive sequencing error, since the variants were only detected in the plasma with low AF and majority of ~96% observed substitutions were observed to be C>A transversion, manly at CCC>CAC and/or CCCG>CCAG context. C>A:G>T substitution was also predominant substitution in the tumour sample of this patient, with C>A alone accounting for 16% of the substitutions in the tumour. It is also important to mention that in the total number of reported somatic variants that were only detected in the ctDNA across all 6 samples, C>A substitution composed ~96% of all substitution. The only remained scepticism may be pointed towards library preparation procedure, which might have resulted in the introduction of this particular pattern of bases changes in the plasma. Previously we have experienced similar pattern of base substitutions, with C>A the predominant change, in the sequenced libraries of a serially collected plasma sample, not included in this study, with a similar kit (ThruPLEX DNA-seq Kit) from the same manufacturer. Extended analysis of that sample did not manifest oxidation pattern, which we first assumed was the rational for observed pattern of substitution. We enquired the manufacturer of the library preparation kit,

however, no such observation had been seen or reported before. At this time, we are not able to justify the validity or erroneous of this observation, however, the discussed facts may imply the validity of the results.

Overall, for the somatic variants that were only present in the plasma ctDNA, an average AF of 4% was observed, commonly below 8% and majority at around 3%. The mean DP was clearly higher than for the mutations that were detected in both tumour and plasma. Aside from the sensitivity of detection and incidental chance of technical biases, the result of our observation in the plasma, concord with an important aspect of our analysis, which was the detection of somatic mutations that were present in the plasma ctDNA but not in the corresponding normal—tumour pairs. Several intriguing interpretations of these results might be hypothesised. A possible explanation can be root from the inherent limitation of tumour sampling in regards to portrait the tumour heterogeneity. The tumour tissue may be heterogeneous, and is it unlikely to capture this heterogeneity with a small piece of single tissue biopsy (13, 28), thus the subclones that have not present in the isolated tumour DNA, may have contributed to the presence of mutations that were observed only in the plasma ctDNA. Two of the patients with sequenced tumour—plasma had metastasis at the time of diagnosis and other 2 had later developed a metastasis. Thus, it might be argued that the low AF observed for these mutations may inform about presence of undiagnosed contributing micro-metastasis [reviewed in (85, 95)], due to the limitations of screening strategies.

It is also important to mention that for some of the variants that were initially detected both in the tumour—plasma pairs, a relatively high level of AF in the normal DNA was observed. Although variants with this feature were not considered as real variants, due to the fact that it was not feasible to justify if the observed allele frequency in the plasma is only due to the technical sequencing biases that has been introduced into the normal DNA sequence such as base miss incorporation by polymerase. However, upon further investigation of the sequencing data the base composition of the reference sequence around the region of these variants was observed to be composed of repeated bases, which reinforced the sequencing bias. The assumption that the AF of normal sequence is mainly due to technical error was reinforced by the fact that all the detected variants were only sequenced from either forward/reverse strands. This

observation may imply that the callers are able to distinguish sequencing errors in the samples, which paly a role when high sensitivity is needed. Although these variants may finally be removed from the results, but they may be manually investigated or compared with other available sequencing dataset for the same sample.

4.3. Future perspectives for monitoring of cfDNA

As part of this thesis we have established methodologies to identify somatic mutations in cfDNA in plasma. In the continuation of this project, longitudinal monitoring of the samples using plasma will be performed. Thus, our results may provoke curiosity, which mutations from the detected mutations are then more suitable to be utilized as a personal biomarker for surveillance of the disease burden? As well, what strategies would best fit into this perspective?

For the purpose of monitoring, there are several approaches that can be followed for this purpose, and a few possibilities will be discussed. Targeted resequencing precludes escalated costs of sequencing while yielding more sensitivity, and also considerably simplifies the data processing workflow by being more time- and cost effective in terms of data analysis, need for informatics infrastructure, and patient sensitive data management [reviewed in (61, 69, 111)]. With this in mind, it is possible to continue targeted sequencing strategy using the NCGC 900 gene panel, since it provides the possibility to follow all mutations that were initially detected in the first plasma sample, as well as new mutations that may arise over time due to cancer progression. However, although it provides high with higher sensitivity in comparison to WES strategy, still the provided sensitivity was not maximized, as many of the mutations in tumour remained unidentified, and the costs are still high. As previously mentioned, sometimes extremely low concentration of ctDNA is present in a background of cfDNA, which considerably impacts the identification of queried mutations, thus requiring high depth of coverage in order to identify mutations [reviewed in (73, 112)]. This issue becomes even more highlighted when the total concentration of ctDNA is low, due to stage and aggressiveness of tumour (96) [reviewed in (86)].

As a second alternative, to reinforce the sensitivity of the assay, it is possible to further reduce the scale of analysis and interrogate a much smaller set of genes, while keeping the high throughput ability and increased sensitivity. There are many commercially available gene sets that have been developed to target the most clinically relevant genes, which are found to be frequently mutated in cancer. Several methodologies have been developed [reviewed in (85)], claiming extreme sensitivity that makes detection of 1 mutated ctDNA fragment feasible. Noteworthy, the majority of these techniques are only capable of investigating a relatively small set of genes, with prior knowledge of mutations in interrogated exons. For instance, we have established a collaboration to utilize the newly established ultrasensitive SiMSen-Seq method (113). It provides multiplexed, PCR-based barcoding of cfDNA that can be detected with next-generation sequencing with a sensitivity of 0.1%. As a result of the unique barcodes, a high level of confidence for detected mutations following massively parallel sequencing will be achieved. In simple words, discrimination of real mutations from false positive mutations becomes feasible.

Digital droplet PCR (ddPCR) on selected mutations is another option. This method provides very high sensitivity and specificity, but the drawback is, as for standard PCR, that a custom design of specific complementary primers with high specificity has to be done for each mutation, which is both laborious and costly. Thus, the use of ddPCR is more sensible when the same mutation in many samples will be investigated.

In regards to the choice of mutation, maybe the safest alternative for this purpose would be to select mutations present in both tumour and plasma, these mutations are more likely to be kept by natural selection forces on tumours since they are more important for tumour development and it has been demonstrated that majority, if not all, of the mutations in metastatic lesions are already present in the considerable number of cells in the primary tumours [reviewed in (2)]. As well, relatively higher AF that does not require to be deeply sequenced can be taken into account for this selection. Overall, based on the specific goal of monitoring, any mutation may be used as a biomarker, including mutations present in non-coding regions such as introns, as well as synonymous mutations. With an optimal choice of methods, longitudinal monitoring using liquid biopsies will allow new non-invasive means of studying cancer patients in terms of

monitoring minimal residual disease, response and resistance to therapy and tumour evolution.

4.4. Pre-analytical considerations of cfDNA analysis

Analysing cfDNA as a biomarker of choice is not without complications. One of the main difficulties of working with cfDNA is the lack of widely accepted standard operation procedures. Many studies have reported various factors that may impact the quantity and quality of the cfDNA during sampling and processing of plasma or serum. These factors include the type of tube used, containing or without preservative, the time interval between venepuncture and blood processing, time- and temperature-dependent storage conditions, centrifugation forces, and effect of freeze-thaw cycles [reviewed in (87, 88)] There are also considerable opposing findings in terms of cancer-dependent variables that may affect total cfDNA concentration that has been described previously in the text [reviewed in (86, 88)]. However, the amount of cfDNA do not reflect the quality and quantity of the tumour-derived proportion of cfDNA, and these factors can significantly impact the quality of downstream libraries and sequencing results.

Pre-analytical factors, e.g. purification and quantification strategies, are also a major obstacle in the cfDNA analysis. Lack of consensus on a robust and streamlined assay in order to be able to yield adequate and reproducible amounts of cfDNA, in a timely manner independent of cancer variable factors, is another prominent pitfall in the workflow of cfDNA analysis. There are various assays for purification of cfDNA from blood serum and/or plasma with different performance and workflow (114). Such assays are for instances conventional extraction with organic solvents, stepwise elution of cfDNA based on selective binding characteristic of silica-membrane technology (115), size-selective magnetic beads (116). The variability between utilized methods has resulted in incomparable results. Of note, many of these assays differ on the amount of initial plasma or serum needed for optimal yields. In an attempt of standardization of cfDNA measurement, Devonshire et al, compared the efficacy of the 3 most used commercially available cfDNA extraction kits and concluded that the kit,

QIAamp circulating nucleic acid, had better performance (114), which was the same kit we used in our study.

4.5. Circulating tumour DNA as an eligible biomarker for liquid biopsy

Although promising, there are several studies scrutinizing the investigation of ctDNA as a promising biomarker in the context of liquid biopsy. The lack of consensus on the primacy of biomarkers for interrogating tumour aberrations is a matter of debate between various groups. The rationale behind our choice of biomarker in this study was the significant observation among many studies as previously described, where cfDNA had been able to prove the presence of interrogated mutations. Moreover, available literature on the significance of each blood-based biomarkers confirms cfDNA potentials for the subjective of our study (91, 92, 94, 96, 117).

The most debatable biomarker across research groups that has been also central focus of many studies in comparison with cfDNA, is circulating tumour cells (CTCs). CTCs are also shed into the bloodstream or other biofluids, and can be captured and concentrated for mutation analysis in cancer patients [reviewed in (78, 95, 118)]. CTCs may seem more promising, especially when biological aspects of tumour, or temporal changes at the transcription level are investigated [reviewed in (86)]. Nevertheless, extremely low numbers of CTCs, with one CTC per $\sim 10^7$ white blood cells (WBCs) per millilitre of blood is one the major challenges in their detection [reviewed in (78, 86, 118)]. For this reason, extensive effort has been invested in developing various isolation strategies to facilitate the capture and concentration of CTCs. For some instances, antibody-based capture assays, functional-characteristics assays, imaging-based assays, and physical property-based technologies are suggested [reviewed in (78, 118)]. Aside from technological limitations in robust discrimination of CTC from crowd of blood cells, lack of well-characterized CTCs markers for validation purpose poses a challenge in their utility. Capturing adequate number of CTCs requires relatively high amounts of blood sample in comparison with required amount for optimum cfDNA isolation workflow. As well, detection of heterogeneity and tumour

dynamics by analysing only a single disseminated CTC is relatively unlikely [reviewed in (78, 95, 118)].

Given information has been extensively interrogated in several studies. In a research by Dawson et al., the ability to detect CTCs over cfDNA with somatic mutations in *PIK3CA* and *TP53* genes in 30 breast cancer patients was 10% lower for CTCs. Furthermore, the number of patients with increased level of detectable cfDNA before manifestation of the disease was more than double of those investigated for CTCs (92). In a clinical trial on a cohort of 41 advanced NSCLC treated with pertuzumab and erlotinib, the association of ctDNA and CTCs with detectable level of oncogenic mutations showed a greater sensitivity of ctDNA over CTCs (119). Another comparative study by Bettegowda et al., targeted tumour-specific rearrangements in 16 patients and demonstrated higher rate of detectable mutations in ctDNA over CTCs, 81% against 0%, respectively (117). However, in contrast to other studies, investigated *EGFR* mutation in patients with metastatic non–small-cell lung cancer was detected at 92% of examined CTCs and only in 33% of matched plasma ctDNA (120). Experiments described above are a few examples in support of both biomarkers, which make drawing a concrete conclusion formidable. Overall, relatively higher level of ctDNA, ease of collection and analysis in comparison to CTCs are advantages ctDNA represents.

Taken together, utility of ctDNA in management of cancer patient has been emphasised. However, on the other side, analysis of whole cell, including protein, RNA, and DNA analysis by CTCs are feasible, and it may better represent the tumour heterogeneity only if discrepancy on the number of CTCs needed to optimally capture heterogeneity could be addressed. In this regard, complementary utilization of CTCs and cfDNA based on the technical and sample availability and the goals of the study has been recommended [reviewed in (78, 95, 118)].

Recently identified short tumour-originated microRNAs (miRNAs) in the plasma, have been proposed to have significant potential in cancer diagnostic, particularly due to strong correlations between their deregulated expression and cancer progression (121, 122). miRNAs negatively regulate gene expression in a diverse range of cell

pathways, including apoptosis, proliferation, metastasis and epithelial to mesenchymal transition . Moreover, miRNAs are shown to be highly stable under storage and handling conditions, which is an essential feature in blood-borne biomarker with extremely low abundance [reviewed in (79)]. However, studies on the blood-based miRNAs have just begun and challenges ahead are purification of inherent extremely low abundance, lack of unifying structural features that allow for selective isolation and/or manipulation, small size (~20–30bp), and relatively lengthy and complex library preparation [reviewed in (77, 79)].

5.Conclusion

In this study, we sequenced 14 matched-normal pairs, and 6 corresponding plasma samples by WES and targeted resequencing strategies to evaluate whether tumour mutations can be detected in the plasma-borne ctDNA from STS patients. The finding of this project may be interpreted in several categories. First, as it was typified in the patient CS14 with mutated *TSC1*, it may be comprehended that underlying mutation may help guide the diagnosis of STS, which are difficult to diagnose in the clinical setting, due to lack of specific IHC tests to distinguish the specific subtype. Secondly, analysis of tumour mutational profiles of particularly hard to diagnose STS subtypes, such as UPS, which explained in the patient CS03, by revealing the mutational signatures may provides clues on the type of cancer that may also guide the clinical decision-making procedures. Although our data suggest that *somatic* variants may hold significance in prognostic and diagnostic and therapeutic relevance of STS, these will need to be further assessed in new studies to more accurately assess the biological consequence of the identified somatic mutations.

In regards to the liquid biopsy strategy that in this project was used for detection of *somatic* mutations in the ctDNA from STS, our results suggest the feasibility of this implication. The proof-of-concept of this conclusion are 31 mutations that were found mutated both in tumour and plasma. As well, the total number of 288 mutations was found only in the plasma that may implies several facts. The most interesting speculation would be to think of these mutations, which might have been detected due to the higher sequencing coverage in comparison the sequenced tumour samples, as a sign of tumour heterogeneity or possible recurrence of the disease. This is also important to remind that 4 patients for whom the plasma was sequenced had metastatic tumours. However, from the other hand the reliability of detected mutations may be enquired. Based on the criteria that we followed the chance of stochastic error is unlikely but not far-fetched, and as it mentioned a pattern of substitutions in mutations where observe, which investigation of recurrence of this specific substitution pattern in future studies of plasma samples may provide more insights into the possibility of technical issues. Nevertheless, detected mutations in the plasma need to be further verified, preferentially by more sensitive strategies such as small

genes panels of frequently mutated genes in cancer and methods with extreme sensitivity such as ddPCR.

In this regard, the liquid biopsies as a non-invasive strategy in cancer care, regardless of biomarker used for analysis, may provide a safe platform to surrogate sub-optimal clinical methodologies in the management of cancers by diagnosis and tracking the cancer evolutionary blueprints, and may soon be able to encourage their utility in routine clinical settings.

6.Future perspectives

In the present study, we only sequenced 6 plasma samples from the available patients material for evaluation and detection of tumour-specific mutations in the plasma. By agreement, 30 high-grade STS patients we be followed for maximum of 5 years, and at the schedule time points the blood and plasma material will be collected. This will provide us the possibility to monitor the tumour-specific mutations in the plasma longitudinally. Accordingly, it will be feasible to draw a more accurate conclusion than this time, for several aspects. First, the level of confidence for currently detected mutation will be reinforced, if specific mutations would be observed in multiple time points. Second, whether results of observation from longitudinal monitoring correlates with the tumour progression and early detection of metastasis. Third, by continuation of monitoring for tumour –specific mutation in the plasma we might be able to detect actionable mutations for drug repurposing. Also informative would be to sequence tumour samples deeper to determine whether the resulting discrepancy in the number of detected mutations in the plasma in comparison to tumours is matter of higher sensitivity, tumour heterogeneity or unexplained technical biases.

Overall, presence of various biomarkers in the blood is indisputable, and the research on the developing technologies and strategies to exploits information is growing at an unprecedented speed. Technological instruments with adequate analytical sensitivity in order to produce tangible findings for clinical decision-making within a rational timespan can significantly speed up the implementation of biomarkers in the clinical setting. Synergic collaborations between the researchers in the fields of genomics and bioinformatics for generating a robust pipeline to analyse enormous amount of NGS data into the biological findings is strongly demanded. Lastly, aggregation of sequencing data and creation of validated databases of information on various cancer types, as results of international consortiums such as TCGA and ICGC can significantly contributes to our understanding of cancer genome, and provide dedicated path for the research in scientific communities.

The choice of proper biomarkers for liquid biopsy is still a matter of debate, however this can be considered that each biomarkers may provide several advantages that vary

from each others, thus on the basis of immediate aim and available resources for each study their utility can be prioritized. In addition, technological instruments with adequate analytical sensitivity in order to produce tangible findings for clinical decision-making within a rational timespan can significantly speed up the implementation of biomarkers in the clinical setting.

On the other hand, synergic collaborations between the researchers in the fields of genomics and bioinformatics for generating a robust pipeline to analyse enormous amount of HTS data into the biological findings is strongly demanded. Lastly, aggregation of sequencing data and creation of validated databases of information on various cancer types, as results of international consortiums such as TCGA and ICGC can significantly contributes to our understanding of cancer genome, and provide dedicated path for the research in scientific communities.

7.Reference

1. Hanahan D, Weinberg RA. The hallmarks of cancer. Cell. 2000;100(1):57-70.
2. Vogelstein B, Papadopoulos N, Velculescu VE, Zhou S, Diaz LA, Jr., Kinzler KW. Cancer genome landscapes. Science. 2013;339(6127):1546-58.
3. World Cancer Report 2014. World Health Organization (WHO), 2014 9789283204299.
4. Anand P, Kunnumakkara AB, Sundaram C, Harikumar KB, Tharakan ST, Lai OS, et al. Cancer is a preventable disease that requires major lifestyle changes. Pharm Res. 2008;25(9):2097-116.
5. Stratton MR, Campbell PJ, Futreal PA. The cancer genome. Nature. 2009;458(7239):719-24.
6. Chin L, Andersen JN, Futreal PA. Cancer genomics: from discovery science to personalized medicine. Nat Med. 2011;17(3):297-303.
7. Lemmon MA, Schlessinger J. Cell signaling by receptor tyrosine kinases. Cell. 2010;141(7):1117-34.
8. Witsch E, Sela M, Yarden Y. Roles for growth factors in cancer progression. Physiology (Bethesda). 2010;25(2):85-101.
9. Hanahan D, Weinberg RA. Hallmarks of cancer: the next generation. Cell. 2011;144(5):646-74.
10. Colotta F, Allavena P, Sica A, Garlanda C, Mantovani A. Cancer-related inflammation, the seventh hallmark of cancer: links to genetic instability. Carcinogenesis. 2009;30(7):1073-81.
11. Luo J, Solimini NL, Elledge SJ. Principles of cancer therapy: oncogene and non-oncogene addiction. Cell. 2009;136(5):823-37.
12. Negrini S, Gorgoulis VG, Halazonetis TD. Genomic instability--an evolving hallmark of cancer. Nat Rev Mol Cell Biol. 2010;11(3):220-8.
13. Alizadeh AA, Aranda V, Bardelli A, Blanpain C, Bock C, Borowski C, et al. Toward understanding and exploiting tumor heterogeneity. Nat Med. 2015;21(8):846-53.
14. Navin NE, Hicks J. Tracing the tumor lineage. Mol Oncol. 2010;4(3):267-83.
15. Esteller M. Cancer epigenomics: DNA methylomes and histone-modification maps. Nat Rev Genet. 2007;8(4):286-98.
16. Chin L, Gray JW. Translating insights from the cancer genome into clinical practice. Nature. 2008;452(7187):553-63.
17. International Cancer Genome C, Hudson TJ, Anderson W, Artez A, Barker AD, Bell C, et al. International network of cancer genome projects. Nature. 2010;464(7291):993-8.
18. Futreal PA, Coin L, Marshall M, Down T, Hubbard T, Wooster R, et al. A census of human cancer genes. Nat Rev Cancer. 2004;4(3):177-83.
19. Gonzalez-Perez A, Perez-Llamas C, Deu-Pons J, Tamborero D, Schroeder MP, Jene-Sanz A, et al. IntOGen-mutations identifies cancer drivers across tumor types. Nat Methods. 2013;10(11):1081-2.
20. Salk JJ, Fox EJ, Loeb LA. Mutational heterogeneity in human cancers: origin and consequences. Annu Rev Pathol. 2010;5:51-75.
21. Supek F, Minana B, Valcarcel J, Gabaldon T, Lehner B. Synonymous mutations frequently act as driver mutations in human cancers. Cell. 2014;156(6):1324-35.

22. Moorcraft SY, Gonzalez D, Walker BA. Understanding next generation sequencing in oncology: A guide for oncologists. Crit Rev Oncol Hematol. 2015;96(3):463-74.
23. Albertson DG, Collins C, McCormick F, Gray JW. Chromosome aberrations in solid tumors. Nat Genet. 2003;34(4):369-76.
24. Lodish H BA, Zipursky SL, Matsudaira P, Baltimore D, Darnell J. Proto-Oncogenes and Tumor-Suppressor Genes. Molecular Cell Biology. 4th ed. New York: W. H. Freeman; 2000.
25. Wood RD, Mitchell M, Lindahl T. Human DNA repair genes, 2005. Mutat Res. 2005;577(1-2):275-83.
26. Esteller M. Epigenetics in cancer. N Engl J Med. 2008;358(11):1148-59.
27. Berdasco M, Esteller M. Aberrant epigenetic landscape in cancer: how cellular identity goes awry. Dev Cell. 2010;19(5):698-711.
28. Marusyk A, Polyak K. Tumor heterogeneity: causes and consequences. Biochim Biophys Acta. 2010;1805(1):105-17.
29. Zahreddine H, Borden KL. Mechanisms and insights into drug resistance in cancer. Front Pharmacol. 2013;4:28.
30. Greaves M, Maley CC. Clonal evolution in cancer. Nature. 2012;481(7381):306-13.
31. Burningham Z, Hashibe M, Spector L, Schiffman JD. The epidemiology of sarcoma. Clin Sarcoma Res. 2012;2(1):14.
32. Helman LJ, Meltzer P. Mechanisms of sarcoma development. Nat Rev Cancer. 2003;3(9):685-94.
33. Matushansky I, Maki RG. Mechanisms of sarcomagenesis. Hematol Oncol Clin North Am. 2005;19(3):427-49, v.
34. Guillou L, Aurias A. Soft tissue sarcomas with complex genomic profiles. Virchows Arch. 2010;456(2):201-17.
35. Jain S, Xu R, Prieto VG, Lee P. Molecular classification of soft tissue sarcomas and its clinical applications. Int J Clin Exp Pathol. 2010;3(4):416-28.
36. Quesada J, Amato R. The molecular biology of soft-tissue sarcomas and current trends in therapy. Sarcoma. 2012;2012:849456.
37. Xiao W, Mohseny AB, Hogendoorn PC, Cleton-Jansen AM. Mesenchymal stem cell transformation and sarcoma genesis. Clin Sarcoma Res. 2013;3(1):10.
38. Wan X, Helman LJ. The biology behind mTOR inhibition in sarcoma. Oncologist. 2007;12(8):1007-18.
39. Fletcher CDM. World Health Organisation classification of tumours. Pathology and genetics of tumours of soft tissue and bone. IARC Press. 2013.
40. Sawamura C, Matsumoto S, Shimoji T, Okawa A, Ae K. How long should we follow patients with soft tissue sarcomas? Clin Orthop Relat Res. 2014;472(3):842-8.
41. Grimer R, Judson I, Peake D, Seddon B. Guidelines for the management of soft tissue sarcomas. Sarcoma. 2010;2010:506182.
42. Clark MA, Fisher C, Judson I, Thomas JM. Soft-tissue sarcomas in adults. N Engl J Med. 2005;353(7):701-11.
43. Corey RM, Swett K, Ward WG. Epidemiology and survivorship of soft tissue sarcomas in adults: a national cancer database report. Cancer Med. 2014;3(5):1404-15.
44. Mertens F, Larramendy M, Gustavsson A, Gisselsson D, Rydholm A, Brosjo O, et al. Radiation-associated sarcomas are characterized by complex karyotypes with frequent rearrangements of chromosome arm 3p. Cancer Genet Cytogenet. 2000;116(2):89-96.

45. Berrington de Gonzalez A, Kutsenko A, Rajaraman P. Sarcoma risk after radiation exposure. Clin Sarcoma Res. 2012;2(1):18.
46. Skvortsova I, Debbage P, Kumar V, Skvortsov S. Radiation resistance: Cancer stem cells (CSCs) and their enigmatic pro-survival signaling. Semin Cancer Biol. 2015;35:39-44.
47. Kim Y, Joo KM, Jin J, Nam DH. Cancer stem cells and their mechanism of chemo-radiation resistance. Int J Stem Cells. 2009;2(2):109-14.
48. Sleijfer S, Ouali M, van Glabbeke M, Krarup-Hansen A, Rodenhuis S, Le Cesne A, et al. Prognostic and predictive factors for outcome to first-line ifosfamide-containing chemotherapy for adult patients with advanced soft tissue sarcomas: an exploratory, retrospective analysis on large series from the European Organization for Research and Treatment of Cancer-Soft Tissue and Bone Sarcoma Group (EORTC-STBSG). Eur J Cancer. 2010;46(1):72-83.
49. Blay JY, van Glabbeke M, Verweij J, van Oosterom AT, Le Cesne A, Oosterhuis JW, et al. Advanced soft-tissue sarcoma: a disease that is potentially curable for a subset of patients treated with chemotherapy. Eur J Cancer. 2003;39(1):64-9.
50. Karavasilis V, Seddon BM, Ashley S, Al-Muderis O, Fisher C, Judson I. Significant clinical benefit of first-line palliative chemotherapy in advanced soft-tissue sarcoma: retrospective analysis and identification of prognostic factors in 488 patients. Cancer. 2008;112(7):1585-91.
51. Cormier JN, Pollock RE. Soft tissue sarcomas. CA Cancer J Clin. 2004;54(2):94-109.
52. Cornetta K, Brown CG. Balancing personalized medicine and personalized care. Acad Med. 2013;88(3):309-13.
53. Benson JD, Chen YN, Cornell-Kennon SA, Dorsch M, Kim S, Leszczyniecka M, et al. Validating cancer drug targets. Nature. 2006;441(7092):451-6.
54. Borden EC, Baker LH, Bell RS, Bramwell V, Demetri GD, Eisenberg BL, et al. Soft tissue sarcomas of adults: state of the translational science. Clin Cancer Res. 2003;9(6):1941-56.
55. Joensuu H, Roberts PJ, Sarlomo-Rikala M, Andersson LC, Tervahartiala P, Tuveson D, et al. Effect of the tyrosine kinase inhibitor STI571 in a patient with a metastatic gastrointestinal stromal tumor. N Engl J Med. 2001;344(14):1052-6.
56. Demetri GD, van Oosterom AT, Garrett CR, Blackstein ME, Shah MH, Verweij J, et al. Efficacy and safety of sunitinib in patients with advanced gastrointestinal stromal tumour after failure of imatinib: a randomised controlled trial. Lancet. 2006;368(9544):1329-38.
57. Seidel C, Fenner M, Langer F, Bantel H, Ganser A, Grunwald V. Imatinib-induced liver cirrhosis in a patient with advanced gastrointestinal stroma tumor (GIST). BMC Cancer. 2012;12:186.
58. Rutkowski P, Van Glabbeke M, Rankin CJ, Ruka W, Rubin BP, Debiec-Rychter M, et al. Imatinib mesylate in advanced dermatofibrosarcoma protuberans: pooled analysis of two phase II clinical trials. J Clin Oncol. 2010;28(10):1772-9.
59. van der Graaf WT, Blay JY, Chawla SP, Kim DW, Bui-Nguyen B, Casali PG, et al. Pazopanib for metastatic soft-tissue sarcoma (PALETTE): a randomised, double-blind, placebo-controlled phase 3 trial. Lancet. 2012;379(9829):1879-86.
60. Martin GS. Cell signaling and cancer. Cancer Cell. 2003;4(3):167-74.
61. Morey M, Fernandez-Marmiesse A, Castineiras D, Fraga JM, Couce ML, Cocho JA. A glimpse into past, present, and future DNA sequencing. Mol Genet Metab. 2013;110(1-2):3-24.

62. Sanger F, Nicklen S, Coulson AR. DNA sequencing with chain-terminating inhibitors. Proc Natl Acad Sci U S A. 1977;74(12):5463-7.
63. Diehl F, Li M, Dressman D, He Y, Shen D, Szabo S, et al. Detection and quantification of mutations in the plasma of patients with colorectal tumors. Proc Natl Acad Sci U S A. 2005;102(45):16368-73.
64. Wagle N, Berger MF, Davis MJ, Blumenstiel B, Defelice M, Pochanard P, et al. High-throughput detection of actionable genomic alterations in clinical tumor samples by targeted, massively parallel sequencing. Cancer Discov. 2012;2(1):82-93.
65. Kinde I, Wu J, Papadopoulos N, Kinzler KW, Vogelstein B. Detection and quantification of rare mutations with massively parallel sequencing. Proc Natl Acad Sci U S A. 2011;108(23):9530-5.
66. Leary RJ, Kinde I, Diehl F, Schmidt K, Clouser C, Duncan C, et al. Development of personalized tumor biomarkers using massively parallel sequencing. Sci Transl Med. 2010;2(20):20ra14.
67. Pritchard CC, Salipante SJ, Koehler K, Smith C, Scroggins S, Wood B, et al. Validation and implementation of targeted capture and sequencing for the detection of actionable mutation, copy number variation, and gene rearrangement in clinical cancer specimens. J Mol Diagn. 2014;16(1):56-67.
68. Bianchi DW, Platt LD, Goldberg JD, Abuhamad AZ, Sehnert AJ, Rava RP, et al. Genome-wide fetal aneuploidy detection by maternal plasma DNA sequencing. Obstet Gynecol. 2012;119(5):890-901.
69. Meldrum C, Doyle MA, Tothill RW. Next-generation sequencing for cancer diagnostics: a practical perspective. Clin Biochem Rev. 2011;32(4):177-95.
70. Lo YM, Chan KC, Sun H, Chen EZ, Jiang P, Lun FM, et al. Maternal plasma DNA sequencing reveals the genome-wide genetic and mutational profile of the fetus. Sci Transl Med. 2010;2(61):61ra91.
71. Mamanova L, Coffey AJ, Scott CE, Kozarewa I, Turner EH, Kumar A, et al. Target-enrichment strategies for next-generation sequencing. Nat Methods. 2010;7(2):111-8.
72. Ohashi H, Hasegawa M, Wakimoto K, Miyamoto-Sato E. Next-generation technologies for multiomics approaches including interactome sequencing. Biomed Res Int. 2015;2015:104209.
73. Metzker ML. Sequencing technologies - the next generation. Nat Rev Genet. 2010;11(1):31-46.
74. Mayeux R. Biomarkers: potential uses and limitations. NeuroRx. 2004;1(2):182-8.
75. Biomarkers Definitions Working G. Biomarkers and surrogate endpoints: preferred definitions and conceptual framework. Clin Pharmacol Ther. 2001;69(3):89-95.
76. Strimbu K, Tavel JA. What are biomarkers? Curr Opin HIV AIDS. 2010;5(6):463-6.
77. Schwarzenbach H, Nishida N, Calin GA, Pantel K. Clinical relevance of circulating cell-free microRNAs in cancer. Nat Rev Clin Oncol. 2014;11(3):145-56.
78. Krebs MG, Metcalf RL, Carter L, Brady G, Blackhall FH, Dive C. Molecular analysis of circulating tumour cells-biology and biomarkers. Nat Rev Clin Oncol. 2014;11(3):129-44.
79. Kosaka N, Iguchi H, Ochiya T. Circulating microRNA in body fluid: a new potential biomarker for cancer diagnosis and prognosis. Cancer Sci. 2010;101(10):2087-92.

80. Fleischhacker M, Schmidt B. Circulating nucleic acids (CNAs) and cancer--a survey. Biochim Biophys Acta. 2007;1775(1):181-232.
81. Mandel P, Métais P. Les acides nucléiques du plasma sanguin chez l'homme. C R Acad Sci Paris. 1948;142:241-3.
82. van der Vaart M, Pretorius PJ. Circulating DNA. Its origin and fluctuation. Ann N Y Acad Sci. 2008;1137:18-26.
83. Jahr S, Hentze H, Englisch S, Hardt D, Fackelmayer FO, Hesch RD, et al. DNA fragments in the blood plasma of cancer patients: quantitations and evidence for their origin from apoptotic and necrotic cells. Cancer Res. 2001;61(4):1659-65.
84. Stroun M, Maurice P, Vasioukhin V, Lyautey J, Lederrey C, Lefort F, et al. The origin and mechanism of circulating DNA. Ann N Y Acad Sci. 2000;906:161-8.
85. Diaz LA, Jr., Bardelli A. Liquid biopsies: genotyping circulating tumor DNA. J Clin Oncol. 2014;32(6):579-86.
86. Schwarzenbach H, Hoon DS, Pantel K. Cell-free nucleic acids as biomarkers in cancer patients. Nat Rev Cancer. 2011;11(6):426-37.
87. El Messaoudi S, Rolet F, Mouliere F, Thierry AR. Circulating cell free DNA: Preanalytical considerations. Clin Chim Acta. 2013;424:222-30.
88. Jung K, Fleischhacker M, Rabien A. Cell-free DNA in the blood as a solid tumor biomarker--a critical appraisal of the literature. Clin Chim Acta. 2010;411(21-22):1611-24.
89. Crowley E, Di Nicolantonio F, Loupakis F, Bardelli A. Liquid biopsy: monitoring cancer-genetics in the blood. Nat Rev Clin Oncol. 2013;10(8):472-84.
90. Newman AM, Bratman SV, To J, Wynne JF, Eclov NC, Modlin LA, et al. An ultrasensitive method for quantitating circulating tumor DNA with broad patient coverage. Nat Med. 2014;20(5):548-54.
91. Murtaza M, Dawson SJ, Tsui DW, Gale D, Forshew T, Piskorz AM, et al. Non-invasive analysis of acquired resistance to cancer therapy by sequencing of plasma DNA. Nature. 2013;497(7447):108-12.
92. Dawson SJ, Tsui DW, Murtaza M, Biggs H, Rueda OM, Chin SF, et al. Analysis of circulating tumor DNA to monitor metastatic breast cancer. N Engl J Med. 2013;368(13):1199-209.
93. Kinde I, Bettegowda C, Wang Y, Wu J, Agrawal N, Shih Ie M, et al. Evaluation of DNA from the Papanicolaou test to detect ovarian and endometrial cancers. Sci Transl Med. 2013;5(167):167ra4.
94. Forshew T, Murtaza M, Parkinson C, Gale D, Tsui DW, Kaper F, et al. Noninvasive identification and monitoring of cancer mutations by targeted deep sequencing of plasma DNA. Sci Transl Med. 2012;4(136):136ra68.
95. Haber DA, Velculescu VE. Blood-based analyses of cancer: circulating tumor cells and circulating tumor DNA. Cancer Discov. 2014;4(6):650-61.
96. Diehl F, Schmidt K, Choti MA, Romans K, Goodman S, Li M, et al. Circulating mutant DNA to assess tumor dynamics. Nat Med. 2008;14(9):985-90.
97. Vasioukhin V, Anker P, Maurice P, Lyautey J, Lederrey C, Stroun M. Point mutations of the N-ras gene in the blood plasma DNA of patients with myelodysplastic syndrome or acute myelogenous leukaemia. Br J Haematol. 1994;86(4):774-9.
98. Anker P, Mulcahy H, Chen XQ, Stroun M. Detection of circulating tumour DNA in the blood (plasma/serum) of cancer patients. Cancer Metastasis Rev. 1999;18(1):65-73.
99. Heitzer E, Ulz P, Geigl JB. Circulating tumor DNA as a liquid biopsy for cancer. Clin Chem. 2015;61(1):112-23.

100. Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics. 2009;25(14):1754-60.
101. 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. Nat Biotechnol. 2013;31(3):213-9.
102. Saunders CT, Wong WS, Swamy S, Becq J, Murray LJ, Cheetham RK. Strelka: accurate somatic small-variant calling from sequenced tumor-normal sample pairs. Bioinformatics. 2012;28(14):1811-7.
103. Robinson JT, Thorvaldsdottir H, Winckler W, Guttman M, Lander ES, Getz G, et al. Integrative genomics viewer. Nat Biotechnol. 2011;29(1):24-6.
104. Alexandrov LB, Nik-Zainal S, Wedge DC, Aparicio SA, Behjati S, Biankin AV, et al. Signatures of mutational processes in human cancer. Nature. 2013;500(7463):415-21.
105. Alix-Panabieres C, Schwarzenbach H, Pantel K. Circulating tumor cells and circulating tumor DNA. Annu Rev Med. 2012;63:199-215.
106. Conyers R, Young S, Thomas DM. Liposarcoma: molecular genetics and therapeutics. Sarcoma. 2011;2011:483154.
107. Martignoni G, Pea M, Reghellin D, Zamboni G, Bonetti F. PEComas: the past, the present and the future. Virchows Arch. 2008;452(2):119-32.
108. Wagner AJ, Malinowska-Kolodziej I, Morgan JA, Qin W, Fletcher CD, Vena N, et al. Clinical activity of mTOR inhibition with sirolimus in malignant perivascular epithelioid cell tumors: targeting the pathogenic activation of mTORC1 in tumors. J Clin Oncol. 2010;28(5):835-40.
109. Lawrence MS, Stojanov P, Mermel CH, Robinson JT, Garraway LA, Golub TR, et al. Discovery and saturation analysis of cancer genes across 21 tumour types. Nature. 2014;505(7484):495-501.
110. Forbes SA, Beare D, Gunasekaran P, Leung K, Bindal N, Boutselakis H, et al. COSMIC: exploring the world's knowledge of somatic mutations in human cancer. Nucleic Acids Res. 2015;43(Database issue):D805-11.
111. Mardis ER. The impact of next-generation sequencing technology on genetics. Trends Genet. 2008;24(3):133-41.
112. Sims D, Sudbery I, Ilott NE, Heger A, Ponting CP. Sequencing depth and coverage: key considerations in genomic analyses. Nat Rev Genet. 2014;15(2):121-32.
113. Stahlberg A, Krzyzanowski PM, Jackson JB, Egyud M, Stein L, Godfrey TE. Simple, multiplexed, PCR-based barcoding of DNA enables sensitive mutation detection in liquid biopsies using sequencing. Nucleic Acids Res. 2016.
114. Devonshire AS, Whale AS, Gutteridge A, Jones G, Cowen S, Foy CA, et al. Towards standardisation of cell-free DNA measurement in plasma: controls for extraction efficiency, fragment size bias and quantification. Anal Bioanal Chem. 2014;406(26):6499-512.
115. Xue X, Teare MD, Holen I, Zhu YM, Woll PJ. Optimizing the yield and utility of circulating cell-free DNA from plasma and serum. Clin Chim Acta. 2009;404(2):100-4.
116. Stemmer C, Beau-Faller M, Pencreac'h E, Guerin E, Schneider A, Jaqmin D, et al. Use of magnetic beads for plasma cell-free DNA extraction: toward automation of plasma DNA analysis for molecular diagnostics. Clin Chem. 2003;49(11):1953-5.
117. Bettegowda C, Sausen M, Leary RJ, Kinde I, Wang Y, Agrawal N, et al. Detection of circulating tumor DNA in early- and late-stage human malignancies. Sci Transl Med. 2014;6(224):224ra24.
118. Alix-Panabieres C, Pantel K. Challenges in circulating tumour cell research. Nat Rev Cancer. 2014;14(9):623-31.

119. Punnoose EA, Atwal S, Liu W, Raja R, Fine BM, Hughes BG, et al. Evaluation of circulating tumor cells and circulating tumor DNA in non-small cell lung cancer: association with clinical endpoints in a phase II clinical trial of pertuzumab and erlotinib. Clin Cancer Res. 2012;18(8):2391-401.
120. Maheswaran S, Sequist LV, Nagrath S, Ulkus L, Brannigan B, Collura CV, et al. Detection of mutations in EGFR in circulating lung-cancer cells. N Engl J Med. 2008;359(4):366-77.
121. Mitchell PS, Parkin RK, Kroh EM, Fritz BR, Wyman SK, Pogosova-Agadjanyan EL, et al. Circulating microRNAs as stable blood-based markers for cancer detection. Proc Natl Acad Sci U S A. 2008;105(30):10513-8.
122. Roth C, Rack B, Muller V, Janni W, Pantel K, Schwarzenbach H. Circulating microRNAs as blood-based markers for patients with primary and metastatic breast cancer. Breast Cancer Res. 2010;12(6):R90.