COMPARATIVE ANALYSIS OF MUTATIONAL LANDSCAPES IDENTIFIED FROM EXOME
SEQUENCING OF TUMOR TISSUES AND CIRCULATING TUMOR DNA

Miss Julanee Leenanitikul

จุฬาลงกรณ์มหาวิทยาลัย
CHULALONGKORN UNIVERSITY

A Thesis Submitted in Partial Fulfillment of the Requirements

for the Degree of Master of Science in Bioinformatics and Computational Biology

Inter-Department of Bioinformatics and Computational Biology

GRADUATE SCHOOL

Chulalongkorn University

Academic Year 2022

การวิเคราะห์เชิงเปรียบเทียบของภูมิทัศน์การกลายพันธุ์ที่ระบุจากลำดับเอกโซมของเนื้อเยื่อมะเร็ง
และดีเอ็นเอของมะเร็งในระบบหมุนเวียน

น.ส.จุฬาณี ลีนะนิธิกุล

Thesis Title	COMPARATIVE ANALYSIS OF MUTATIONAL LANDSCAPES IDENTIFIED FROM EXOME SEQUENCING OF TUMOR TISSUES AND CIRCULATING TUMOR DNA
By	Miss Julanee Leenanitikul
Field of Study	Bioinformatics and Computational Biology
Thesis Advisor	SIRA SRISWASDI, Ph.D.
Thesis Co Advisor	TRAIRAK PISITKUN, M.D.

Accepted by the GRADUATE SCHOOL, Chulalongkorn University in Partial Fulfillment of the Requirement for the Master of Science

.... Dean of the GRADUATE SCHOOL

(Associate Professor YOOTTHANA CHUPPUNNARAT, Ph.D.)

THESIS COMMITTEE

.... Chairman

(Professor SUPACHITRA CHADCHAWAN, Ph.D.)

.... Thesis Advisor

(SIRA SRISWASDI, Ph.D.)

.... Thesis Co-Advisor

(TRAIRAK PISITKUN, M.D.)

.... Examiner

(Assistant Professor MONNAT PONGPANICH, Ph.D.)

.... External Examiner

(Verayuth Praphanphoj, M.D.)

จุฬาณี ลีนะนิธิกุล : การวิเคราะห์เชิงเปรียบเทียบของภูมิทัศน์การกลายพันธุ์ที่ระบุจาก
ลำดับเอกโซมของเนื้อเยื่อมะเร็งและดีเอ็นเอของมะเร็งในระบบหมุนเวียน. (
COMPARATIVE ANALYSIS OF MUTATIONAL LANDSCAPES IDENTIFIED FROM
EXOME SEQUENCING OF TUMOR TISSUES AND CIRCULATING TUMOR DNA)
อ.ที่ปรึกษาหลัก : อ. ดร.สิระ ศรีสวัสดิ์, อ.ที่ปรึกษาร่วม : อ. นพ.ไตรรักษ์ พิสิษฐ์กุล

การหาลำดับของดีเอ็นเอมะเร็งจากเลือดในระบบหมุนเวียนถูกนำมาใช้เป็นวิธีทางเลือกที่
ไม่รุกรานสำหรับการวินิจฉัยโรคมะเร็งและการตรวจหาลักษณะการกลายพันธุ์ของมะเร็ง อย่างไรก็
ตาม สัดส่วนของดีเอ็นเอมะเร็งที่ได้จากการเก็บตัวอย่างเลือด มีปริมาณที่ต่ำประกอบกับปัจจัยอื่นๆ
ที่จำกัดความสามารถของการวิเคราะห์ดีเอ็นเอมะเร็งจากตัวอย่างเลือดที่จะทำร่วมกับการหาความ
สอดคล้องของแวเรียนซ์ที่จำเพาะต่อแวเรียนซ์จากตัวอย่างเนื้องอกและแวเรียนซ์ที่มีความสำคัญกับ
มะเร็งในทางการตรวจหาทางคลินิก ในการศึกษานี้ ได้ทำการหาลำดับเอกโซมทั้งหมดภายในดีเอ็น
เอมะเร็งจากเลือดในระบบหมุนเวียน ในการจับคู่ดีเอ็นเอมะเร็งจากเลือดในระบบหมุนเวียนและดี
เอ็นเอของมะเร็งที่ได้จากการเก็บชิ้นเนื้องอกในผู้ป่วยมะเร็ง 15 ราย เพื่อประเมินขอบเขตของความ
สอดคล้องกันระหว่างโปรไฟล์การกลายพันธุ์ที่ได้จากแหล่งข้อมูลทั้งสอง เราพบว่าในเลือดมีปริมาณ
สัดส่วนของดีเอ็นเอมะเร็งสูงถึง 16.4% ยังคงไม่เพียงพอสำหรับการตรวจหาแวเรียนซ์ที่จำเพาะต่อ
เนื้องอก แต่อย่างไรก็ตามเมื่อสัดส่วนดีเอ็นเอมะเร็งจากตัวอย่างเลือดเท่ากับ 30% หรือสูงกว่า
พบว่ามีความสอดคล้องกันสูงระหว่างแวเรียนซ์ที่ได้จากการตรวจชิ้นเนื้องอกกับแวเรียนซ์ที่พบในดี
เอ็นเอมะเร็งจากตัวอย่างเลือด จากการศึกษานี้ได้ผลสรุปการวิเคราะห์ความสามารถของดีเอ็นเอ
มะเร็งที่ได้จากการเก็บตัวอย่างเลือดว่าความสามารถตรวจจับความแตกต่างของเนื้องอกได้อย่าง
สม่ำเสมอและตรวจหายีนที่สำคัญที่เกี่ยวข้องกับมะเร็งได้ทั้งในผู้ป่วยที่มีทั้งการเก็บตัวอย่างจาก
หลายช่วงเวลาและผู้ป่วยมะเร็งที่มีทั้งเนื้องอกระยะแรกและระยะแพร่กระจาย

สาขาวิชา	ชีวสารสนเทศศาสตร์และ ชีววิทยาเชิงคอมพิวเตอร์ (สห สาขาวิชา)	ลายมือชื่อนิสิต ....
ปีการศึกษา	2565	ลายมือชื่อ อ.ที่ปรึกษาหลัก ....
		ลายมือชื่อ อ.ที่ปรึกษาร่วม ....

6380140520 : MAJOR BIOINFORMATICS AND COMPUTATIONAL BIOLOGY

KEYWORD: Circulating tumor DNA, Whole Exome Sequencing, Liquid biopsy, Next-generation sequencing

Julanee Leenanitikul : COMPARATIVE ANALYSIS OF MUTATIONAL LANDSCAPES IDENTIFIED FROM EXOME SEQUENCING OF TUMOR TISSUES AND CIRCULATING TUMOR DNA. Advisor: SIRA SRISWASDI, Ph.D. Co-advisor: TRAIRAK PISITKUN, M.D.

Next-generation sequencing of circulating tumor DNA (ctDNA) has been used as a noninvasive alternative for cancer diagnosis and characterization of tumor mutational landscape. However, low ctDNA fraction and other factors can limit the ability of ctDNA analysis to capture tumor-specific and actionable variants. In this study, whole-exome sequencings (WES) were performed on paired ctDNA and tumor biopsy in 15 cancer patients to assess the extent of concordance between mutational profiles derived from the two source materials. We found that up to 16.4% ctDNA fraction can still be insufficient for detecting tumor-specific variants and that high concordance with tumor biopsy is achieved when ctDNA fraction is 30% or higher. Most importantly, ctDNA analysis can consistently capture tumor heterogeneity and detect key cancer-related genes even in a patient with both primary and metastatic tumors.

Field of Study:	Bioinformatics and Computational Biology	Student's Signature ....
Academic Year:	2022	Advisor's Signature ....
		Co-advisor's Signature ....

ACKNOWLEDGEMENTS

TABLE OF CONTENTS

Page

จุฬาลงกรณ์มหาวิทยาลัย
CHULALONGKORN UNIVERSITY

LIST OF TABLES

LIST OF FIGURES

จุฬาลงกรณ์มหาวิทยาลัย
CHULALONGKORN UNIVERSITY

CHAPTER 1   INTRODUCTION

1.1 BACKGROUND AND MOTIVATION

Genetic variations in human genome are broad, ranging from single nucleotide variants (SNVs), indels, to copy number alteration (CNA). CNA includes large chromosomal aneuploidy events involving entire chromosomes. While most genetic variations have no apparent impact on human health, some manifest at the earliest stages of tumorigenesis and accumulate throughout subsequent tumor developments (2).

Genetic mutations are either germline or somatic. Germline mutations are inherited and not frequently unique to cancer population. On the other hand, somatic mutations occur during a person's lifetime and play a significant role in tumorigenesis (3). A typical cancer begins with somatic mutations in genes related to the regulation of cell division and cellular function which make the cells nonresponsive to signals that control cell growth or death. Uncontrolled cell growth subsequently results in tumor mass development. In the late stages, the tumor can become metastatic, where cancer cells spread into neighboring tissues or enter the bloodstream, traverse throughout the body, and form new tumors at different sites.

Tumor biopsy followed by histopathological examination or next generation sequencing is a standard approach for cancer diagnosis which can reveal molecular characteristics of the tumor tissue. However, tumor biopsy is invasive, may damage nearby tissues, is not applicable when the tumor locations are inaccessible, and cannot detect metastasis at distant sites (4, 5). Furthermore, because each tumor is heterogeneous, the mutation profiles identified from a biopsy would not accurately reflect the entire tumor mutational landscape. This introduces bias and error into the selection of personalized therapies (6). Consequently, a non-invasive technique that can capture the broad genetic heterogeneity of the tumor would be highly desirable.

To address the limitations of tumor biopsy, liquid biopsy was developed as an alternative method for cancer diagnosis by analyzing circulating tumor DNA (ctDNA) in the plasma from patients' blood samples, which are much easier to

collect than tissue samples. Hence, liquid biopsy is rapid, non-invasive, and has the potential to capture ctDNA shed from both primary tumor sites and metastatic sites in the patient. The technique can also be practically performed multiple times to monitor the course of treatment or the progression of the disease (7). Because ctDNA in the cancer patient predominantly derives from apoptosis and tumor cell necrosis throughout the body (8, 9) , analysis of ctDNA can yield insights into tumor-specific mutational landscape and intra-tumor heterogeneity that were not covered by traditional tissue biopsy (10). However, although ctDNA levels in cancer patients' bloodstream were reportedly higher than those in healthy individuals (11) , ctDNA analysis for cancer diagnosis still requires dedicated devices and techniques with high sensitivity due to low quantity of ctDNA and contamination of genomics DNA from normal cells (12, 13) .

Next-generation sequencing (NGS) has evolved dramatically in the past several years. NGS is a technology that parallelly sequences massive amounts of DNA sequences in a sample. It becomes more popular and widely used in clinical oncology due to its high accuracy and sensitivity for mutation detection (14, 15). Furthermore, it systematically provides genomic profiles from both tumor tissue and ctDNA samples (16, 17). Applications of NGS in cancer genomics mainly focuses on whole-exome sequencing (WES) and whole-genome sequencing (WGS) (18). However, whole genome sequencing requires high cost as it covers the full length of the genome. In situations where only some parts of the genome are of interest or where high-confidence variant detection is not needed, low-coverage sequencing can be more cost-effective (19, 20).

Ultra-low pass whole genome sequencing (ULP-WGS) or low-coverage sequencing (~0.1-0.3x coverage) is emerging as a cost-effective alternative that allows population-scale screening of mutations and copy number alterations for the entire genome. It produces a mere fraction of the data per sample compared to standard NGS techniques and relies on computational methods to fill in the missing information (21). ULP-WGS has been used to quantify tumor fractions (22-27) and estimate copy number alterations (28-31), etc.

In a previous study (32), ULP-WGS, coupled with ichorCNA software to estimate tumor fraction, was utilized to screen tumor with greater than 10% in ctDNA for performing subsequent whole-exome sequencing. This lets clinicians filter out samples with low tumor-specific DNA that would not benefit from further in-depth investigation and save costs. Furthermore, previous study also provides comprehensive ctDNA mutation profiles from multiple patients and shows high concordance between whole-exome sequencing results of ctDNA and tumor tissue samples. However, many issues regarding the applicability of ctDNA-based liquid biopsy have not been covered by previous studies, namely (i) whether the capability of this technique varies across cancer types, (ii) whether ctDNA can capture genetic variants from multiple tumor sites within the same patient, and (iii) whether the technique can effectively monitor cancer progression and treatment response across multiple timepoints.

This study obtained ctDNA, tumor, and normal DNA whole-exome sequencing data for bioinformatics analysis from the Queen Sirikit Centre for Breast Cancer (QSCBC) and from the neoantigen and cancer vaccine project of the Immunotherapy Excellence Center, King Chulalongkorn Memorial Hospital. QSCBC data were collected between 2019-2021 from 10 breast cancer and 1 sarcoma patients. Data from the Immunotherapy Excellence Center were collected between 2020-2021 from 4 gastrointestinal cancer patients and 1 melanoma patient with liver metastasis. The tumor fractions of all ctDNA samples were previously determined using ichorCNA prior to whole-exome sequencing. Solid tumor tissues were obtained from biopsy or surgery. Normal samples were obtained from peripheral blood mononuclear cells (PBMC). Circulating tumor DNA were obtained from blood samples. Sample collections, processing, and sequencing have been performed earlier. The scope of this thesis starts from the processing and bioinformatics analysis of raw DNA sequencing data onwards.

Our main purpose is to assess the concordance between whole-exome sequencing of ctDNA and tumor tissue samples in various cancer types and across multiple time points. In one case, the patient was diagnosed with multiple cancers, which would let us assess whether ctDNA analysis can capture mutation profiles

from multiple tumor sites. We will also determine the impact of tumor fraction in ctDNA sample on the quality of whole-exome sequencing. Our work indicates the potential of ctDNA-based liquid biopsy to supplement and/or replace tumor biopsy in clinical applications.

1.2 RESEARCH OBJECTIVE

1. To assess the concordance of mutational landscapes detected from liquid biopsy of circulating tumor DNA and tumor tissue in cancer patients
2. To evaluate the ability of liquid biopsy of circulating tumor DNA to capture genetic variants from multiple tumor sites within the same patient

1.3 SCOPE OF THE RESEACH

1.3.1.　To assess the impact of ctDNA fraction on the quality of WES by comparing the number of identified variants and the correlation of variant allele frequency estimated from ctDNA and tumor tissue sample.

1.3.2.　To evaluate the degree of concordance between mutational landscapes derived from ctDNA and tumor biopsy by calculating the fraction of common variants and the correlation of variant allele frequency

1.3.3.　To assess whether cancer-relevant variants can be consistently identified from exome sequencing of ctDNA

1.3.4.　To evaluate the ability of ctDNA to capture variants from multiple metastatic tumors within the same patient.

1.4 EXPECTED OUTCOMES

Circulating tumor DNA can capture entire genetic variation in tumor and mutational landscape of tumor and can be a representative of tumor biopsy in the further clinical applications of cancer.

CHAPTER 2   LITERATURE REVIEW

2.1 HALLMARK OF CANCER AND CARCINOGENESIS

Cancer is one of the leading causes of mortality. It is a disease in which the growth of specific cells become uncontrollable, and the affected cells eventually expand and spread to other parts of the body. Cells regularly divide and proliferate to form new cells to replace damaged and old cells. When this process goes awry, damaged or abnormal cells keep proliferating when they should not. Tumors, which are tissue masses, arise from the expansion of these cells. Tumors can be benign or cancerous (malignant). Cancerous tumors can eventually spread to other parts of the body and form new tumors via a process called metastasis. In contrast, non-cancerous (benign) tumors do not infiltrate nearby tissues and rarely reoccur after removal. However, benign tumors can still develop into excessively large tissue mass that can cause serious negative effects or even death, such as benign brain tumor (33, 34).

2.1.1 HALLMARK OF CANCER

The broad genotypes of cancer cell are a consequence of six key abnormalities in cell physiology that collectively drive malignant cell development, namely 1) self-sufficiency in growth signals, 2) insensitivity to growth-inhibitory signals, 3) evasion of programmed cell death, 4) infinite replicative capacity, 5) persistent angiogenesis, and 6) tissue invasion and metastasis (Figure 1) (35).
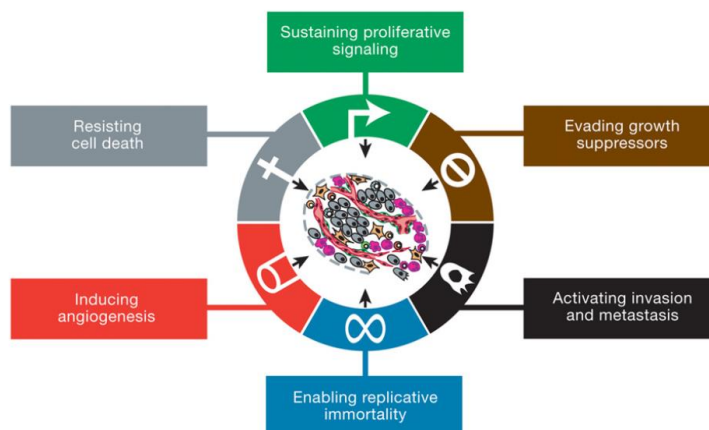
Figure 1 Hallmark of cancer consists of six steps
:1) self-sufficiency in growth signals 2) insensitivity to growth-inhibitory (antigrowth)
signals 3) evasion of programmed cell death (apoptosis) 4) infinite replicative
capacity 5) persistent angiogenesis and 6) tissue invasion and metastasis

2.1.2 CARCINOGENESIS

Carcinogenesis, or the process of cancer formation typically consists of three stages: initiation, promotion, and progression (Figure 2) (36-38).
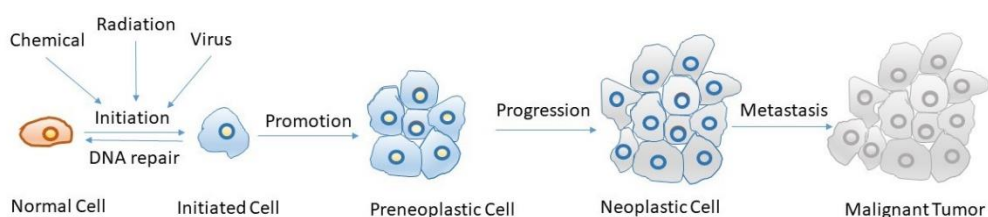
Figure 2 The three-phase process of carcinogenesis after carcinogen exposure and
the different scientific applications based on primary cancer animal models

1. Initiation

Cancer development begins with initiation. Initiators can be exogenous agents (e.g., cigarette smoke, radiation, virus) or endogenous substances (e.g., bile acids, reactive oxygen species) that cause DNA damages. When these damages are incorrectly repaired, the changes in DNA in damaged cell become permanent and irreversible (39). All daughter cells from the division of the mutant cell will also carry the same mutations. More exposure to the initiators naturally lead to greater risk of carcinogenesis. (37, 38)

2. Promotion

Promotion is the second stage that occurs on initiated cells and stimulate the cells to divide. The intracellular or extracellular environment of the cells or the promoters can influence cancer development. These factors enhance the clonal growth of mutated cells, which eventually leads to accumulation of tumor mass. However, promoters may not cause cancer on their own (37, 40).

3. Progression

The third stage progression is the transformation of a benign tumor into a neoplasm and finally malignancy. In this step, tumor cells compete with one another to survive, resulting in the emergence and selection of more mutations that make the tumor cells more aggressive. As the tumor grows in size, individual cells accumulate different mutations, leading to increased heterogeneity within the tumor **mass** (40).

2.2 GENETIC VARIATION IN CANCER

A genetic variation (or a mutation) is a permanent change in the DNA sequence, which provides instructions for cells to function. Most alterations to a person's DNA are harmless but certain variations can lead to genetic disorders. Hundreds of potential genetic changes can develop during each round of cell division (41).Genetic variations can be categorized into two major classes based on how they arise: 1) germline variations and 2) somatic variations (Figure 3)

2.2.1. GERMLINE VARIATIONS

Only around 5%–10% of all malignancies are caused by germline variations. Germline variations occur in a sperm cell or an egg cell and are directly transmitted from a parent to a child. Every cell in the newborn contains the same germline mutations. Hereditary cancers are those that result from pathogenic germline variations. To date, more than 50 different hereditary cancers have been identified (42).

2.2.2. SOMATIC VARIATIONS

The most frequent causes of cancer are somatic or acquired genomic variations, which result from the accumulation of DNA damages over the course of a person's lifetime. Different cells in a person's body contain different set of somatic variations. A few typical carcinogens are smoking, UV radiation, viruses, chemical exposures, and aging (42).
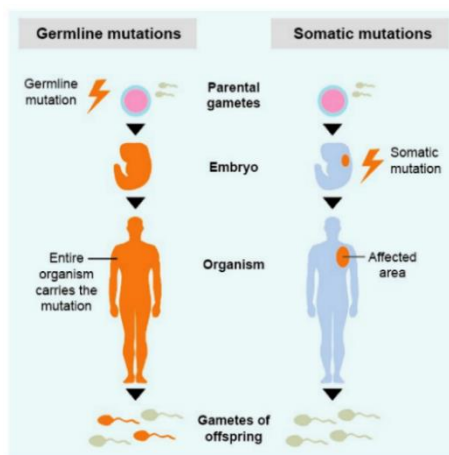
Figure 3 Genetic variations based on the tissue from which they arise

2.2.3 TYPES OF GENETIC VARIATIONS

Genetic variations can cause multiple types of changes on the DNA sequences, including small variations such as single nucleotide base substitutions, indels, and larger structural variations such as copy number alterations (CNA), which refers to major aneuploidy of individual genes, genomic regions, or whole chromosomes (43) (Figure 4).
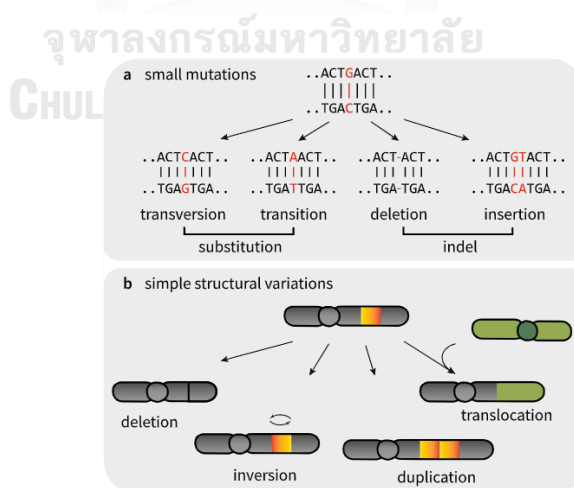
Figure 4 Two types of genetic variations consist of small and structural genetic variations

2.2.3.1 Small variation (1-50 bp)

1. Substitution

There are six possible substitutions for the Watson-Crick base pairings: C>A, C>G, C>T, T>A, T>C, and T>G (44). A substitution in a protein coding region could change a codon to one that encodes a different amino acid and result in a small change on the protein produced (also called missense mutation), or one that encodes the same amino acid and cause no change on the protein produced (also called silent mutation), or it could change an amino-acid-coding codon to a stop codon, resulting in an incomplete protein (also called nonsense mutation) (45).

Synonymous mutations, or silent mutations, are substitutions in the coding regions that do not affect the amino acid sequences of the produced proteins. Non-synonymous mutations are substitutions in the coding regions that affect the amino acid sequences of the produced proteins. Non-synonymous mutations consist of missense and nonsense mutations. Although synonymous mutations do not affect the amino acid sequences, they can impact gene functions in rare cases (46).

2. Indels

An insertion or a deletion affecting 2 or more nucleotide positions. Small and large indels are distinguished on the basis of length, with small indels spanning less than 1kb (47).

2.2.3.2 Structural variation (>50 bp)

Structural variations (SVs) are also known as chromosomal rearrangements. SVs involve rearrangements of large DNA segments that can affect DNA copy number (43).

1. Copy number variations (CNV)

Copy number variation is a form of structural variation that changes the copy count of genes or other genetic elements at certain genomic positions via a duplication or deletion event. Copy number variations account for 4.8-9.5% of the entire length of the human genome. Copy number variations are critical in mammals for generating diversity in the population but are also a cause of diseases (48).

2. Inversion

Inversions occur when a segment of DNA breaks off from the chromosome and is then reinserted into the original breakpoints in an inverted orientation. Although inversions do not result in DNA gains or losses, they can still have negative effects on gene function and regulation. For example, a gene can be disrupted by an inversion if its coding region is truncated or if its regulatory elements are relocated far away from the gene (49).

3. Translocation

Translocation occurs when a region of one chromosome is relocated to a nonhomologous chromosome or to a different location on the same chromosome.

Translocations place genes in novel linkage associations and create chromosomes with no homologous partner (50).

2.3 SOMATIC VARIANT IDENTIFICATION AND ANALYSIS

Variant calling is the process of analyzing next generation sequencing (NGS) data to identify the differences in genomic sequence of a person compared to references. This will reveal a large amount of genetic variation information, such as single nucleotide variation (SNV), insertion and deletion sites, structural variation sites, and copy number variations. Different analysis pipelines were developed specifically for identifying each type of genetic variants. This study focuses on the detection of somatic SNVs from a comparative analysis of tumor-normal paired samples from cancer patients (51).
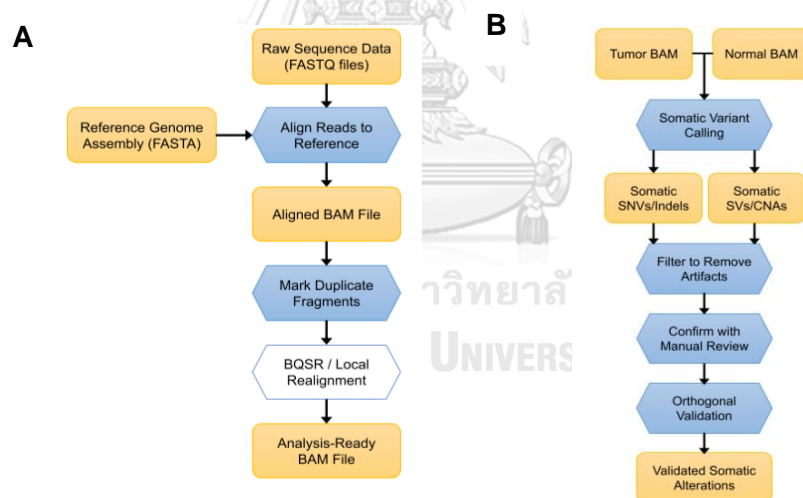
Figure 5 Sequencing pipeline of NGS analysis and somatic variant calling

2.3.1 ALIGNMENT AND PRE-PROCESSING OF NGS DATA

Before variant calling, NGS reads must be pre-processed and aligned to reference sequences. Alignment allows us to map each read to the corresponding genomic position and identify differences between DNA

sequences in the sample and the reference. Aligner software such as Burrows-Wheeler Aligner (BWA) can be used. The resulting alignments are stored in binary alignment (BAM) file format. After an initial alignment, the base quality scores may be recalibrated and a second round of realignment may be performed to reduce false-positive alignments around indels. The quality of the final results may be further checked to ensure that sufficient sequencing coverage was obtained and that there is no contamination (Figure 5A) (51).

2.3.2 SOMATIC VARIANT CALLING

To call somatic variants, a normal tissue or blood sample is required, in addition to tumor tissue sample, to rule out germline mutations that would also be present in normal cells (Figure 5B). Popular somatic mutation callers include MuTect2 (52), Strelka2 (53), and VarScan2 (54). Each caller has distinct advantages and disadvantages, and therefore combining results from two or more callers can provide higher sensitivity and specificity. Low tumor purity and low sequencing depth can make the detection of somatic mutations more challenging. Candidate somatic variants should be further screened to remove alignment artifacts and common variants that are unlikely to be pathogenic. Variant allele frequency, or the fraction of mutant allele, can be used to select variants of interest. Some studies suggested that utilizing a panel of normals (PoN), or a collection of DNA data from normal specimens, to identify and remove recurring sequencing artifacts is beneficial. Reads on both strands should support the same variant alleles without bias in read position, base quality, or mapping quality. High-quality SNVs or indels should also be present at extremely low frequency (MAF < 0.001). Finally, candidate variants should be visualized via tools such as Integrative Genomics Viewer (51).

2.4 PRECISION MEDICINE IN ONCOLOGY

Precision medicine is a new therapeutic approach that aims to treat and prevent the disease by analyzing each patient's unique characteristics and using that information to select the best treatments. Large-scale omics data, including genomics, proteomics, epigenomics, and metabolomics enable precision medicine by providing detailed information about the patients at molecular level. Models developed from these data to predict disease progression and treatment response help clinicians make the optimal therapeutic decisions. The Human Genome Project and advancements in NGS can be considered as the first step of precision medicine (55). In the context of cancer, NGS can discover clinically actionable variations (56). The effectiveness of precision medicine in the treatment of cancer has been shown in many researchs (57-59). The notion of somatic variations as the primary cause of cancer development is known as precision oncology (60).

2.5 TUMOR BIOPSY AND THE EMERGENCE OF LIQUID BIOPSY IN CLINICAL PRACTICE

2.5.1 TUMOR BIOPSY AND ITS ADVANTAGES AND DISADVANTAGES

Tissue biopsy is the extraction of tumor tissue using fine needle or open surgery. Subsequent genetic and histopathological assessment of these tissue samples can determine the type of tumor, which aids oncologists in developing a personalized treatment plan. Although tissue biopsy is important, it is invasive, expensive, time-consuming, and, most importantly, sometimes ineffective at capturing the broad genetic heterogeneity of a tumor. This is because only a small region of the tumor is sampled. Furthermore, repeating tissue biopsies multiple times to monitor tumor progression or treatment response is not feasible. Tissue biopsy is also not applicable when the tumor itself is undetectable, such as early-stage tumor and metastatic site.

2.5.2 LIQUID BIOPSY

Liquid biopsy is a recent non-invasive diagnostic and prognostic technique in precision oncology that has gained a lot of attention in recent years as a way to overcome the limits of tissue biopsies (61). The term liquid biopsy was first used to describe the study of circulating tumor cells (CTCs) in the bloodstream, but it is now mostly used to describe the analysis of circulating, or cell-free, tumor DNA (ctDNA). Liquid biopsy encompasses the analyses of any body fluids, including blood, plasma, serum, saliva, urine, and gastric juice, for clinical evaluation. Since 2004, when Guardant Health developed the first commercially accessible liquid biopsy test, clinical use of liquid biopsy has expanded considerably. In 2016, liquid biopsy (limited to ctDNA analysis) was formally adopted as a diagnostic method in major laboratories. Currently, multiple FDA-approved commercial liquid biopsy tests are available, such as the cobas® EGFR Mutation Test v2, which identifies whether specific patients with non-small cell lung cancer are eligible for EGFR tyrosine kinase inhibitors. Because liquid biopsy is non-invasive, it can be repeatedly performed to monitor tumor progression or treatment response. Furthermore, liquid biopsy can capture diverse tumor DNA that were shed from organs that are difficult to access via biopsy. Hence, genetic data obtained through liquid biopsy can be a better representative of the complete population of tumor cells. Importantly, liquid biopsy enables early tumor detection (61).

2.5.2.1 Cell-free DNA (cfDNA) and circulating tumor DNA (ctDNA)

Circulating cell-free DNA (cfDNA) and circulating tumor DNA are DNA that were shed from cells into the bloodstream. cfDNA were first identified in 1948 by Mandel and Metais in blood samples from healthy individuals. Later, ctDNA were identified in cancer patients'

blood. ctDNA was thought to originate from tumor tissue since it has several cancer-associated molecular features, including single nucleotide genetic variations, methylation, cancer-derived viral sequences, rearrangements, amplifications, microsatellite instability, and loss of heterozygosity. In addition to blood, cfDNA and ctDNA can also be obtained from cerebrospinal fluid, saliva, and in extremely low concentrations from urine. There are two proposed pathways for the release of DNA from cells into the bloodstream. The first is passive release of DNA through cell death (apoptosis or necrosis). The second is active secretion where living tumor cells release extracellular vesicles, including exosomes and prostasomes, that carry DNA fragments of lengths between 150 and 250 bp, (Figure 6). (8, 61, 62)
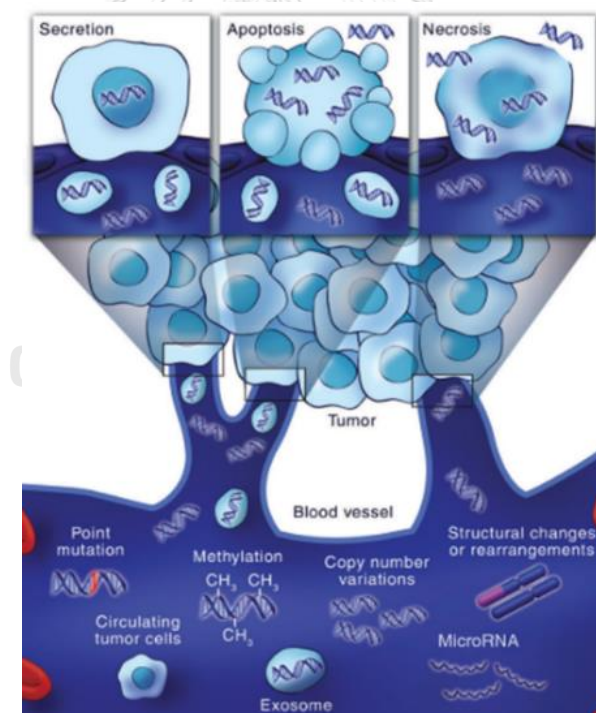
Figure 6 The process of release small fragments of cell-free DNA and into circulation by multiple mechanisms.

The comparison of advantages and disadvantages between liquid biopsy and tumor biopsy for clinical application are summarized in Figure 7 (63).
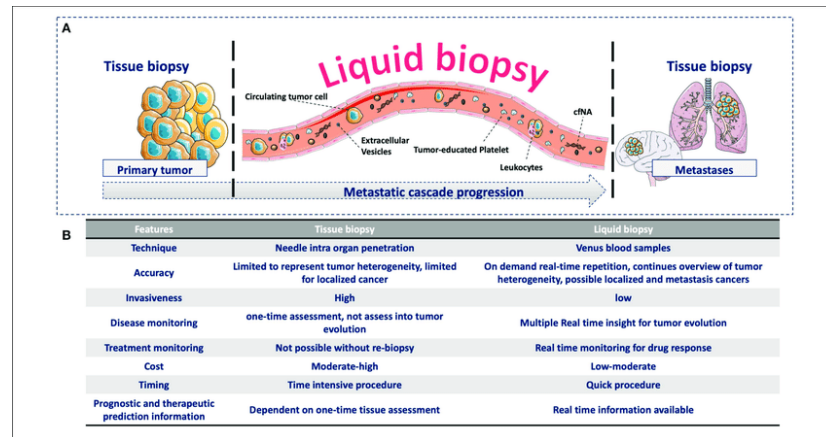
Figure 7 Advantages of liquid biopsy vs. tissue biopsy during the metastatic cascade.

2.6 NEXT GENERATION SEQUENCING (NGS) AND ITS APPLICATION IN CLINICAL ONCOLOGY

NGS is a massively parallel DNA/RNA sequencing technique that provides ultra-high throughput, scalability, and speed. NGS has revolutionized biological sciences, allowing labs to perform a wide range of applications and analyze biological systems at an unprecedented scale (56, 64). Typical applications of NGS are the analyses of whole genome or exome, also called whole genome sequencing (WGS) or whole exome sequencing (WES), respectively. NGS can also be applied to deeply sequence specific target regions, known as targeted sequencing, or to analyze RNA transcripts, known as RNA-Seq or called transcriptome sequencing, to discover novel RNA isoforms, splice sites, and to quantify gene expression levels, or to analyzes epigenetic factors such as genome-wide DNA methylation and DNA-protein bindings. In the context of cancer, NGS enables the detection of rare somatic variants, and tumor subclones (64).

In this study, we focused on the applications of NGS at genomics level (Table 1).

2.6.1. WHOLE-GENOME SEQUENCING (WGS)

WGS is the most thorough NGS approach what examines the whole 3.2 billion bases of the human genome. The tremendous decrease in sequencing cost has made WGS an affordable tool for genomics research. WGS can be applied to the genomes of diverse species. Sequencing coverage, which is the average number of reads that align to a specific location on the genome, is an important quality parameter for a WGS study. Variant discovery in human genome requires approximately 30× to 50× coverage to achieve an acceptable degree of confidence (65, 66).

2.6.2. WHOLE EXOME SEQUENCING (WES)

Whole exome sequencing is a targeted sequencing technique that focuses on exons which account for a little less than 2% of the human genome length. Because the majority of disease-causing mutations occur on exons, WES is a more cost-effective alternative to WGS for clinical applications. In WES, DNA fragments that correspond to exons are preferentially captured and sequenced. A typical WES analysis yields a sequencing coverage of 50x to 100× (65, 66).

Table 1 Comparison of whole genome sequencing (WGS), whole exome sequencing (WES)

Technique	Target Regions	Sequencing depth	Variants Detected	Advantages	Limitations
WGS	Entire genome	>30X	~4,000,000	• Identifies variants in all genomes • Detects genome rearrangements, CNV, novel SNP and structural variants • Uniform depth of sequencing	• Highest cost • Largest volume of data is produced • Require long and most complex analysis • Limited application in clinical diagnostic
WES	2% of genome	>50X~100X	~20,000	• Identifies variants in all protein-coding regions • Low cost compared to WGS	• Possibility to have incomplete exome coverage • Cannot detect non-coding and structural variants • Require exome capture or enrichment methods during library preparation

In addition to commonly used NGS methods, this study also utilized a low-coverage sequencing technique known as ultra-low pass genome sequencing (ULP-WGS) which can provide a rough overview of DNA copy number alteration at low cost.

2.6.3 ULTRA-LOW PASS GENOME SEQUENCING (ULP-WGS)

ULP-WGS is a low-cost, high-throughput DNA-sequencing method that can detect large-scale structural variations on the genomes with only 0.1x to 3x sequencing coverage. This method utilizes imputation algorithms to fill in the missing information and improve the accuracy of variant calling. In the field of cancer diagnosis, ULP-WGS can be used to identify somatic copy

number alterations (SCNAs) and estimate the fraction of tumor DNA among cfDNA in patients' blood sample (22-27)(28-31). This enables copy number variation profiling and tracking over time (32).

2.7 PREVIOUS STUDIES OF CIRCULATING TUMOR DNA

In a previous study (32), ULP-WGS coupled with ichorCNA software was used to estimate the tumor fraction in cfDNA and to screen patients with greater than 10% in ctDNA fraction for performing subsequent whole-exome sequencing (Figure 8). This lets clinicians save costs and time by filtering out samples with low tumor-specific DNA that would not benefit from further in-depth NGS investigation. Furthermore, several previous studies have shown high concordance between whole-exome sequencing results of ctDNA and tumor tissue samples.

However, many issues regarding the applicability of ctDNA-based liquid biopsy have not been covered by previous studies, namely (i) whether the capability of this technique varies across cancer types, (ii) whether ctDNA can capture genetic variants from multiple tumor sites within the same patient, and (iii) whether the technique can effectively monitor cancer progression and treatment response across multiple timepoints.
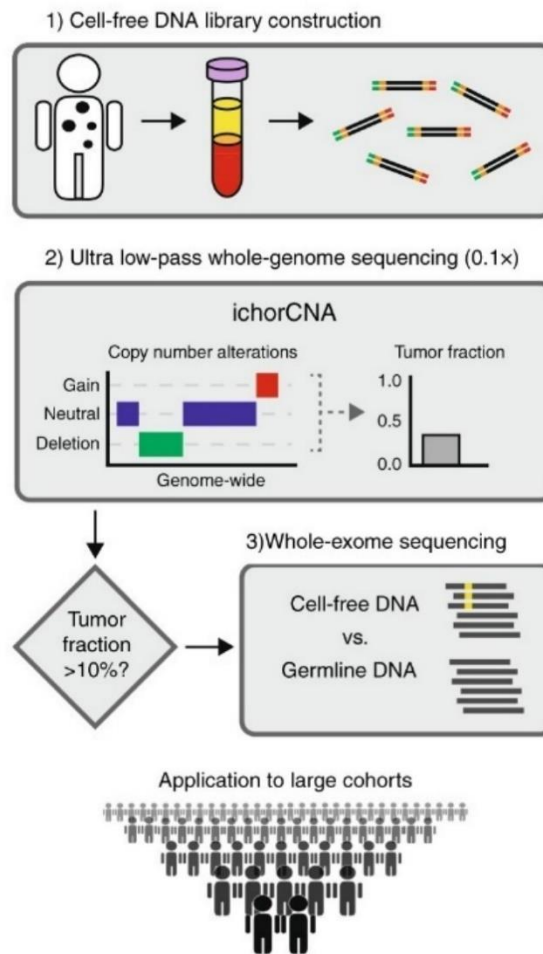
Figure 8 Previous study workflow in applying ULP-WGS with their developed software "ichorCNA" to simultaneously estimate copy number variation and quantify tumor fraction from cfDNA sample of cancer patients prior WES

CHAPTER 3   EXPERIMENTAL

3.1 BIOINFORMATICS ANALYSIS

Sequenced ULP-WGS of ctDNA data were mapped to the hg 19 (GRCh37) reference genome (0.1-0.3X genome-wide coverage). Analysis of the ULP-WGS data was quantified the tumor DNA fraction by ichorCNA software (32). The effect of tumor DNA fraction on variant discovery will be further explored. For WES, sequenced data were aligned to the hg38 (GRCh38) reference genome using the Burrows-Wheeler Aligner (BWA) (67) and processed using the standard Genome Analysis Toolkit (GATK) pipeline (68), comprising duplicate marking, indel realignment, and base quality score recalibration. Germline variants in PBMC samples were called with HaplotypeCaller (69), which also realigned reads onto the inferred haplotypes. HaplotypeCaller model uses ploidy in its genotype likelihood calculations which unsuitable for somatic variant calling. Somatic variants in ctDNA and tumor biopsy samples were called using 3 tools, Mutect2 (52) for calling low-frequency variants, Strelka**2** (53) for calling high-frequency variants, and VarScan (54) for calling high-frequency variants to supplement Strelka. Each tool identified somatic variants by comparing variants observed in ctDNA or tumor tissue against the reads found in normal DNA. Variant Call Format (VCF) files from 3 different variant callers were merged into single VCF file. The VCF file was converted to Mutation Annotation Format (MAF) files containing called variants of ctDNA and tumor biopsies and subjected to quality filtering. Candidate variants will be filtered out if the (1) variant allele frequency (VAF) in both ctDNA and tumor was lower than 1%. For each variant, the calculation of VAF were calculated from number of alternated sequencing reads found in total sequencing reads (2) VAF of normal tissue was higher than 5% (3) number of alternated reads in tumor was lower than 4 reads (4) number of total reads of normal or tumor DNA was less than 8 reads. Synonymous variants were included in this study in order to evaluate trend of concordance between patients' paired samples along with only

non-synonymous variants group. In addition, cancer relevant genes and cancer genetic mutation in this study were annotated by the Catalogue of Somatic Mutation in Cancer (COSMIC) database (44, 70).

3.2 CONCORDANCE ANALYSIS

The concordance between mutational profiles derived from WES of ctDNA and tumor was calculated at two levels. The first level is based on the number of (i) concordant variants and (ii) variants that were detected in only tumor tissue or only ctDNA sample (discordant variants) that passed the quality cutoff defined above. The numbers of concordant and discordant variants between ctDNA and paired tumor tissue would represent the ability of ctDNA to capture variants found in tumor tissue. Whereas the numbers of concordant and discordant variants between samples across time points would illustrate the drift of genetic variation over time. The Jaccard similarity index between the sets of variants will be calculated in order to evaluate the similarity between two sets.

The second level of concordance analysis is based on the Pearson's linear correlation of variant allele frequencies (VAF) of shared variants that were identified in both source materials. For the melanoma patient who has both skin and liver tumor tissue biopsies (MN01), the concordance between ctDNA and each tissue was evaluated separately.

3.3 CANCER-RELEVANT GENE CLASSIFICATION

To identify cancer relevant genes from ctDNA and matched tumor tissue samples, first we divided group of genes based on each cancer type and sorted genes by mutation frequency from highest to lowest mutation frequency of genes which were found among all patients. Top genes for each cancer type were compared with genes from Cancer Browser in COSMIC database (44) based on tissue

types related to cancer, and tier-1 genes, which have been documented with cancer-relevant activities, from Cancer Gene Census from COSMIC (71). For additional databases, Human Protein Atlas was also be used to identify the protein expression levels in cancer patient and in normal tissue related to the cancer type. ClinVar was used to assess clinical significance (e.g., pathogenic, likely pathogenic, uncertain significance, etc.). The mutation landscape was visualized using Maftools (72) (version 2.8.0) in R (version 4.1.3).

3.4 STATISTICAL ANALYSIS

Standard descriptive statistics were used to characterize the profile of mutation types (e.g., substitution, insertion, deletion), the level of overlap between ctDNA and tumor, the mutation landscape, and the profile of base substitutions (e.g., C>T/G>A, T>G/A>C) identified in each patient and cancer type. The level of overlap of identified variants between ctDNA and tumor tissue samples in each patient will be calculated as the number of shared variants detected in both ctDNA and tumor divided by the number of all tumor variants. Systematic differences in the profiles of mutation types or base substitutions between tumor and ctDNA variants will be further explored to identify possible underlying mechanisms, which may reflect the heterogeneity of the tumor. Moreover, the mutation types and base substitution profiles between ctDNA and tumor were compared to determine whether circulating tumor DNA can capture a similar profile as the tumor's.

Standard inferential statistics and Pearson's correlation coefficients were considered significant if the p-values were less than 0.01. Mann-Whitney U test to check whether VAFs of concordant variants are higher than those of discordant variants. Venn diagrams were visualized using ggvenn (version 0.1.9). This study used ggplot2 (version 3.3.6) for visualization in R.

3.5 PATIENT CHARACTERISTICS

ctDNA and tumor whole exome sequencing (WES) data from 18 samples from 15 cancer patients with 4 cancer types were analyzed (Table 2). There are nine patients with breast cancer, one patient with sarcoma, four patients with gastrointestinal cancer (one cecum cancer, two colon cancer, and one duodenum cancer), and one melanoma patient with tissues collected from both the primary skin site and a metastatic liver site. There were 2 patients who more than single paired ctDNA and tumor samples, BC01 with 2 different timepoints of paired samples and MN01 with 2 paired samples of ctDNA and tumor at skin site and liver site. Lower than 6 months of time interval between paired ctDNA and tumor tissue sample collection was defined as same timepoint (APPENDIX A). Due to external logistical factors in some timepoints, there is lack of paired ctDNA or paired tumor tissue sample in some cases. Thus, in concordance analysis, some of paired samples were compared from cross timepoint (APPENDIX B). In concordance analysis, the time interval between paired ctDNA and tumor tissue sample collections varied from 0 to 17 months, with a median of 3 months. In parallel, the data collection workflow starts with an ultra-low pass whole-genome sequencing (ULP-WGS) and whole exome sequencing of the blood samples to assess ctDNA fractions before continuing comparative analysis of paired ctDNA and tumor tissue samples (Figure 9).

Table 2. Patient characteristics

Cancer type	Number of samples
Breast	9 (60 %)
Sarcoma	1 (6.67%)
Gastrointestinal	4 (26.67%)
- Cecum	1
- Colon	2
- Duodenum	1
Melanoma	1 (6.67%)
Total	**15**
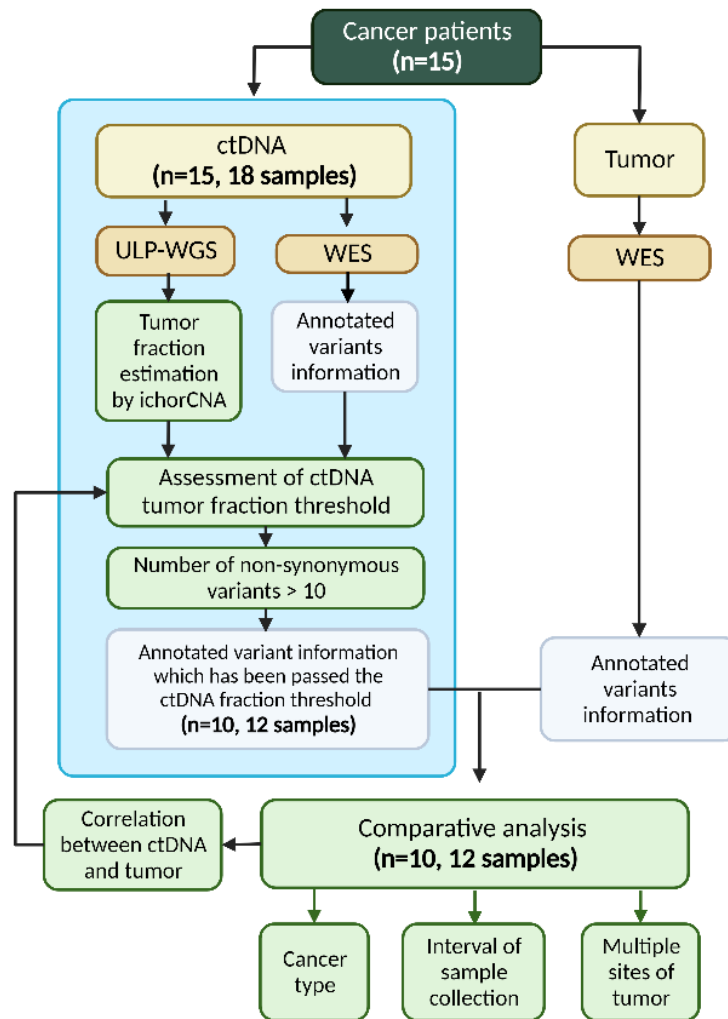
Figure 9 Study workflow of 15 cancer patients

Blue box indicates ctDNA protocol and analysis. After a preliminary round of WES

quality check, 6 samples with lower than 17% ctDNA fractions were also excluded

from comparative analysis with tumor WES.

CHAPTER 4   RESULTS AND DISCUSSIONS

In this study, we assessed the impact of the threshold for ctDNA fraction on the quality of whole exome sequencing (WES) of ctDNA samples, analyzed the degree of concordance between WES of ctDNA and tumor samples based on cancer types and tumor sites, analyzed the trend of variant allele frequency associated with concordance among paired samples and showed the potential drift in mutational profiles over time when paired ctDNA and tumor tissues and paired ctDNA samples were collected several months apart. Due to there are some reports about synonymous mutations relevant to human cancer (73, 74) also we assumed that including synonymous variants along with non-synonymous variants would make the degree of concordance between ctDNA and tumor clearer. Thus, we also evaluated the trend of results from only non-synonymous variants and all variants included synonymous variants whether including synonymous variants in analysis may improve the trend of overall results in order to suggest including synonymous variants in further analysis.

4.1 ASSESSMENT OF CTDNA TUMOR FRACTION THRESHOLD

To identify an appropriate cutoff for ctDNA fraction, the relationship between ctDNA fractions estimated from ULP-WGS and the number of variants identified from ctDNA WES was first analyzed. This shows that there is a positive correlation between ctDNA fraction and the number of identified non-synonymous variants with non-significant (Figure 10A, Pearson's correlation = 0.4657, p-value = 0.069) and the number of all variants with significant trend (Figure 10C, Pearson's correlation = 0.6202, p-value = 0.006). Furthermore, very low number of non-synonymous variants (≤10) and all variants included synonymous variants (≤90) were identified in all samples with ≤16.4% ctDNA fractions and that ctDNA fractions above 30% appeared to have no impact on the number of identified variants. The median number of identified non-synonymous variants and all variants for samples with ≤16.4% ctDNA

fractions are 3 and 10 variants while the median for samples with >30% ctDNA fractions are 95 and 296 variants. The copy number profiles for the five patients with low ctDNA fractions were shown in Figure 11.

When focusing the group of patients with sufficient numbers of identified variants (>30% ctDNA fractions), there is also a positive trend between ctDNA fractions and the concordance between ctDNA WES and tumor WES data on non-synonymous variants (Pearson's correlation = 0.4473, p-value = 0.1448, Figure 10B). In contrast, there is no correlation between ctDNA fractions and the concordance between ctDNA WES and tumor WES data in when all variants were included (Pearson's correlation = 0.0516, p-value = 0.8734, Figure 10D). The degree of concordance was calculated as the Pearson's correlation between the variant allele frequencies (VAF) of shared variants identified in both WES data. As the fluctuation in concordance remains high in all variants and only nonsynonymous regardless of ctDNA fraction, likely due to other factors from the patient and tumor, our results suggest that an appropriate cutoff for ctDNA fraction to ensure high quality WES data lies between 16% and 30%. This is in contrast with previous report (32) that recommended a rather lowest cutoff of 3% ctDNA fraction and 10% of optimal ctDNA fraction for performing WES. We found that a low number of variants were identified in all five samples with ctDNA fraction below 16.4% (7.20-16.42%). Although there may be other factors that could explain the low number of identified variants, our finding indicates that the cutoff for ctDNA fraction needs to be carefully established.
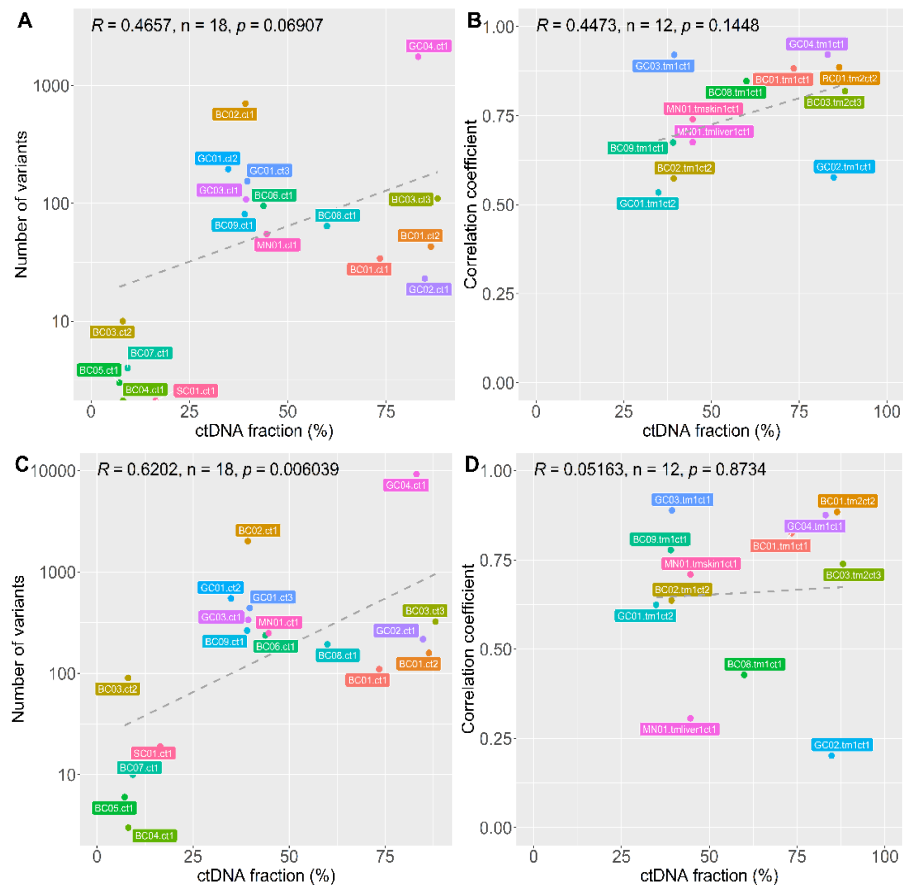
Figure 10 Assessment of ctDNA tumor fraction threshold

A) Comparison between estimated ctDNA fraction and the number of non-synonymous variants identified in ctDNA WES. Very low number of non-synonymous variants were identified when ctDNA fraction is 16.4% or lower. Samples with ctDNA fraction above 30% shows similar numbers of identified non-synonymous variants. The overall Pearson's correlation between ctDNA fractions and number of identified variants is 0.4657. B) Comparison between ctDNA fraction and the degree of concordance (Pearson's correlation) between variant allele frequency estimated from ctDNA and tumor tissue samples. Only samples with sufficient number of non-synonymous variants (ctDNA fraction above 30%) are shown. The is a positive trend between ctDNA fraction and concordance with Pearson's correlation of 0.4473. C), D) Similar plot as A), B) but used the number of all variants (Pearson's correlation = 0.6202) and the degree of concordance (Pearson's correlation) between variant allele frequency of all variants estimated from ctDNA and tumor tissue samples for

the comparison, respectively. Each dot represents a sample (BC: breast cancer, GC: gastrointestinal cancer, SC: sarcoma, MN: melanoma). The numbers following tm (tumor) and ct (ctDNA) indicate the sample collection time point (1 or 2 or 3).
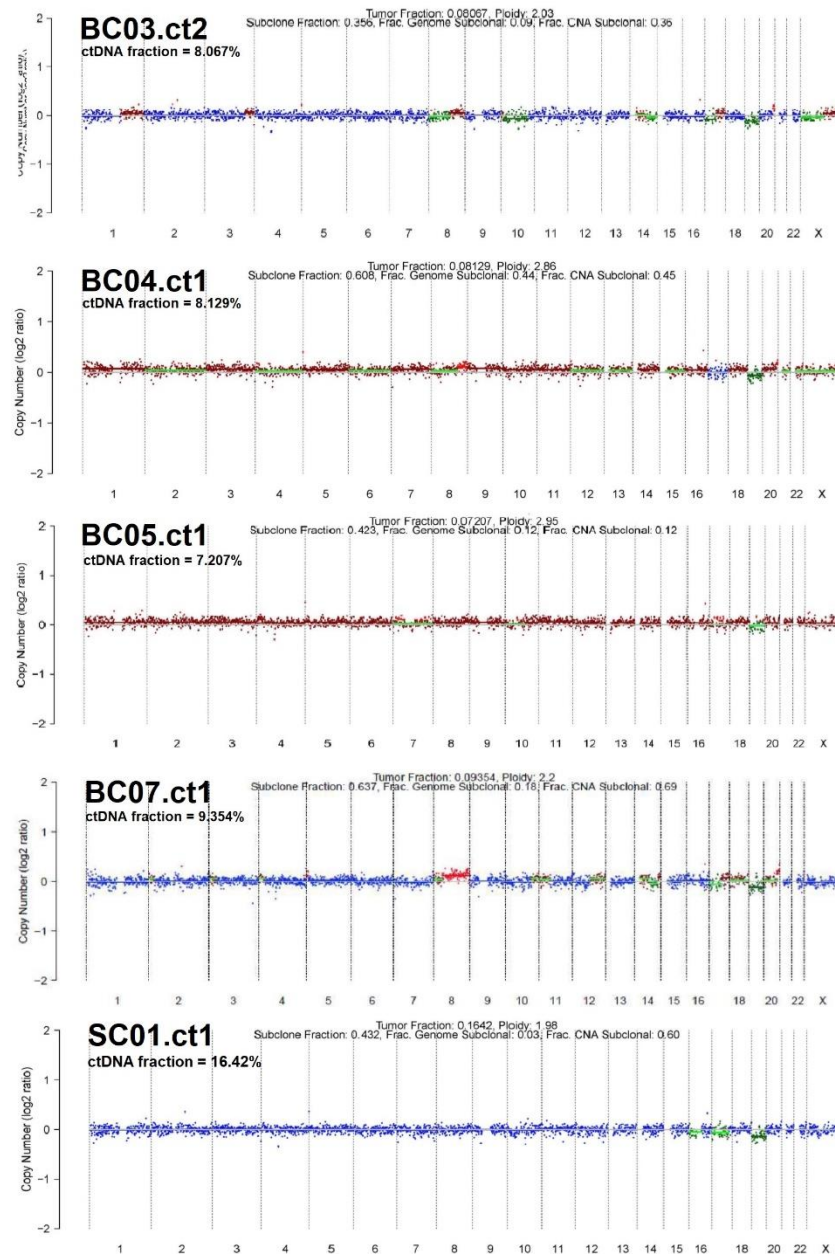
Figure 11 The copy number profiles for the five patients with low ctDNA fractions
From overall 18 samples of 15 patients, these five patients with ctDNA fraction lower than 16.42% were excluded from this study before comparative analysis. The

median number of identified non-synonymous variants and all variants from these samples are 10 (range from 3 to 90 variants) and 3 variants (range from 1-3 variants), respectively, which were insufficient for comparative analysis.

4.2 CONCORDANCE BETWEEN CTDNA WES AND TUMOR WES

As different cancer types may exhibit different degrees of heterogeneity and variations in mutational profiles that can affect the concordance between ctDNA WES and tumor WES, data from patients with different cancer types were analyzed separately. The overlap between variants identified from ctDNA and tumor tissues WES of breast cancer patients and gastrointestinal cancer patients were showed in Figure 12. For patients with multiple sample collection time points, concordances were calculated between WES of ctDNA and tumor samples that were collected closest to each other to minimize the impact from the drift in mutational profile over time (Figure 15D and 15H). This results in six ctDNA-tumor pairs from five breast cancer patients, four ctDNA-tumor pairs from four gastrointestinal cancer patients, and two ctDNA-tumor pairs from one melanoma patient were showed in APPENDIX B.

4.2.1 CONCORDANCE BETWEEN CTDNA WES AND TUMOR WES BASED ON CANCER TYPE

For breast cancer patients, the average Jaccard index for the overlap of identified all variants (Figure 13A) and non-synonymous variants (Figure 13B) from ctDNA WES and tumor WES are 0.41 (SD = 0.10) and 0.59 (SD = 0.16), with average time interval between sample collection of 6 months (range from 0 to 17 months). Variant allele frequencies (VAF) of shared variants estimated from ctDNA and tumor WES are moderately concordant with Pearson's correlation of 0.71 (SD = 0.16, Figure 13A) for all variants and 0.78 (SD = 0.13, Figure 13D) for only non-synonymous variants.

For gastrointestinal cancer patients, the average Jaccard index for the overlap of identified all variants (Figure 14C) and non-synonymous variants (Figure 14D) are 0.54 (SD = 0.25) and 0.64 (SD = 0.34), with average time interval between sample collection of 2 months. The high standard deviation is due to an outlier case with extremely low concordance (Jaccard Index of 0.15, 9 shared variants out of 61 identified in both ctDNA and tumor WES from only non-synonymous variants and Jaccard Index of 0.18, 161 shared variants out of 913 identified in both ctDNA and tumor WES from all variants). Nonetheless, the concordant in VAF of shared variants of overall gastrointestinal patient cases are still moderately high for both all variants (average Pearson's correlation = 0.64, SD = 0.32, Figure 14A) and only non-synonymous variants (average Pearson's correlation = 0.74, SD = 0.21, Figure 14D).

In all patients except GC02, the Jaccard indices of identified non-synonymous variants from ctDNA and tumor tissue were higher than those of all variants (including synonymous variants) (Figure 12). This indicates a higher variety of synonymous variants. Similarly, most of the correlations of variant allele frequency between ctDNA WES and tumor WES in patients were also higher when only non-synonymous variants were considered (except for BC02, BC09 and GC03 where inclusion of synonymous variants improve the correlation). Hence, while including synonymous variants can broaden the identified mutational landscape, it may also introduce noises into the analysis depending on the cases.

In addition, although WES of ctDNA and tumor may identify similar sets of variants with highly correlated variant allele frequencies (VAF) with Pearson's correlation reaching 0.89 for all variants and 0.92 for nonsynonymous variants in some patients (Figure 13 and 14), the actual VAFs can differ quite a lot (e.g., higher ctDNA VAF than tumor VAF in BC08 in Figure 13 and higher tumor VAF than ctDNA VAF in GC03 in Figure 14). This may not necessarily indicate poor quality of ctDNA WES because VAF in tumor WES can also be affected by tumor purity and intra-tumor heterogeneity (75-78).

The lack of strong correlation between ctDNA fraction and the concordance in VAFs between ctDNA and tumor (Figure 10B and 10D) suggested that the latter explanations are more likely. Anyway, this discrepancy would be difficult to resolve because WES of multiple biopsies of the same tumor would be required to establish the ground truth VAFs for comparing with results from ctDNA.

For concordance variants, VAFs are either consistently higher in tumor tissue or consistently higher in ctDNA (Figure 16). This consistency indicates the high quality of VAF estimated from ctDNA. At the same time, the fact that VAFs are sometimes higher in tumor and sometimes higher in ctDNA depending on the patients also illustrates the impact from several factors such as tumor heterogeneity, contamination of normal tissue, and the shedding rate of tumor DNA into the bloodstream that can heavily influence the VAF values. On the other hand, discordant variants are mostly rare variants with significantly lower VAFs than those of concordant variants (Wilcoxon rank sum test P-value < 0.01) as expected (Figure 17). Rare variants identified only ctDNA likely originated from cancer cells in the area not covered by the biopsy. Hence, ctDNA can complement biopsy to obtain a more complete tumor genomic information (79).
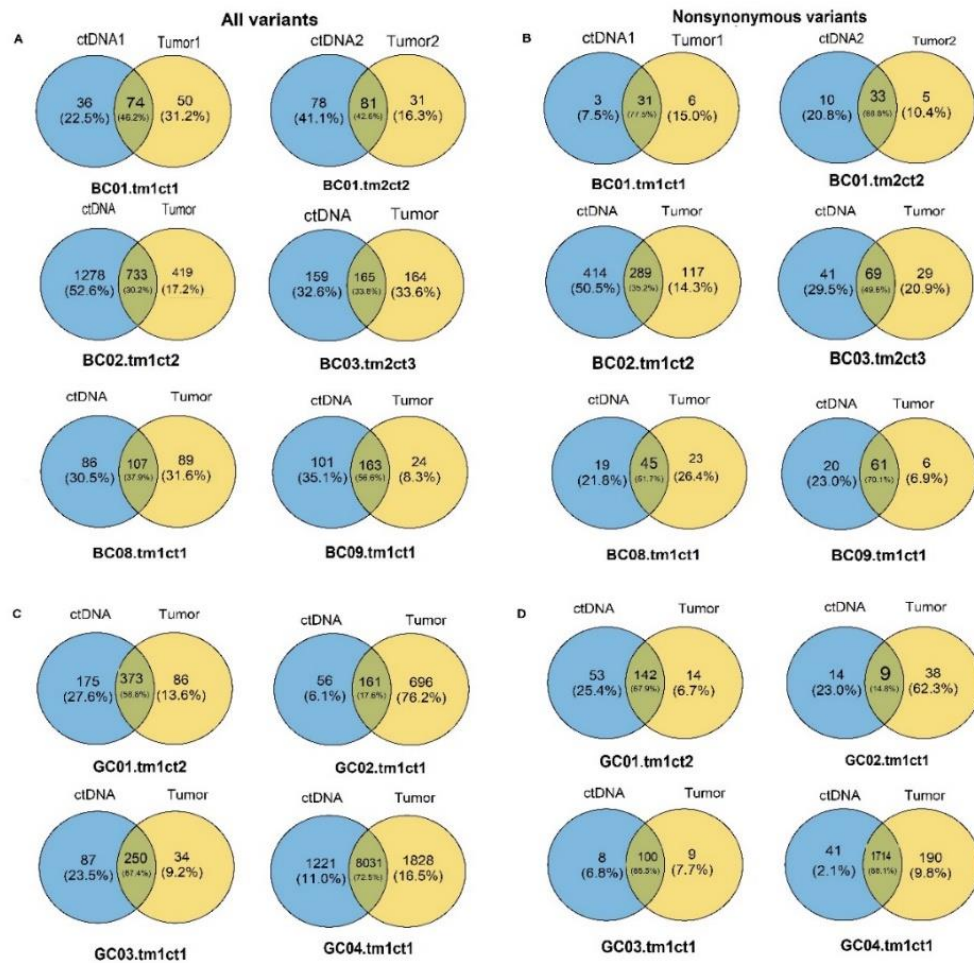
Figure 12 The overlap of shared variants between ctDNA and tumor as identified by WES.

A) B) The overlap of all variants included synonymous variants and only non-synonymous variants between paired WES samples of breast cancer patients, respectively C) D) Similar to A) B) but from gastrointestinal cancer patients.
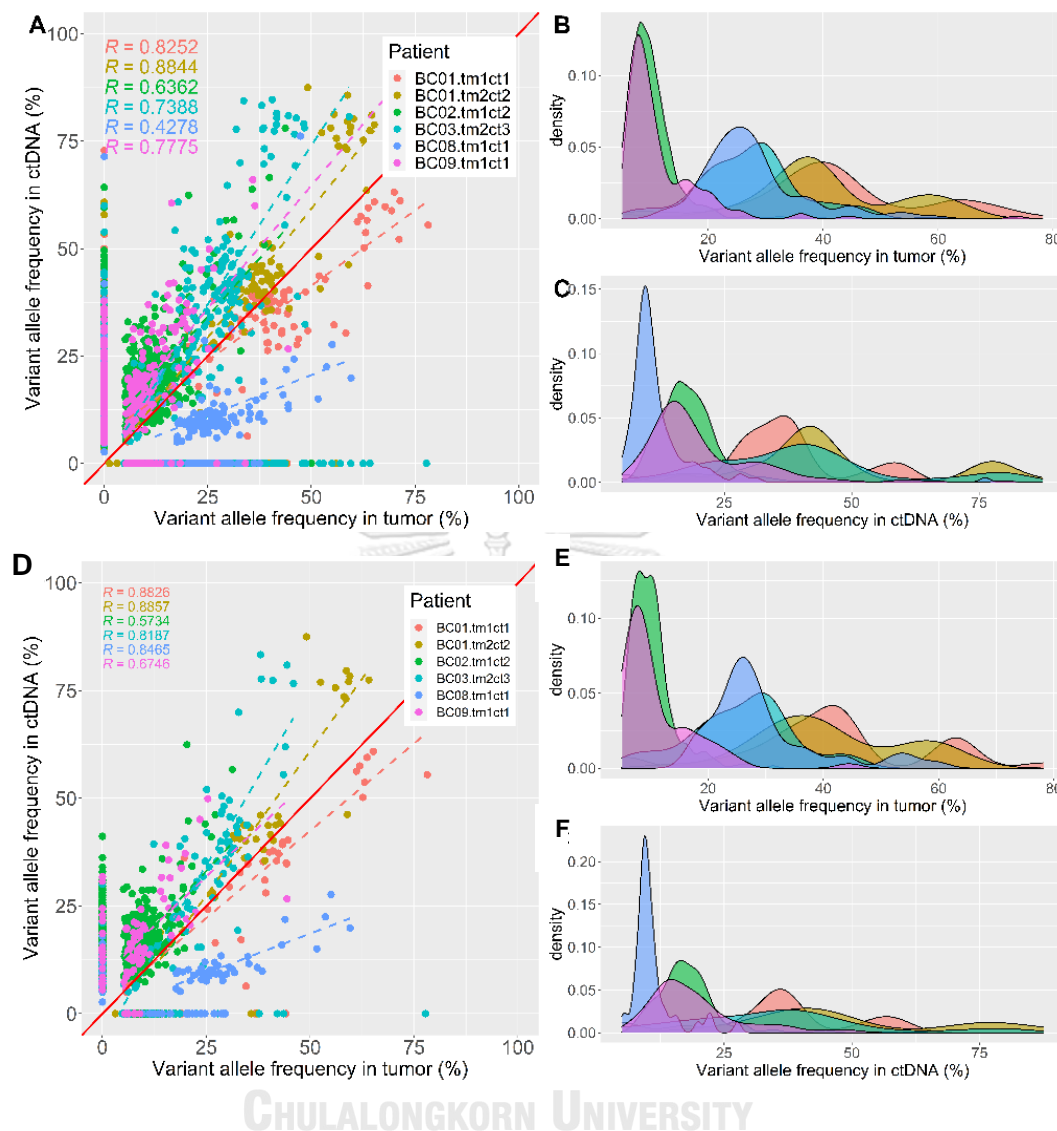
Figure 13 Concordance between ctDNA WES and tumor WES in breast cancer patients

A) Correlation between variant allele frequencies (VAF) of shared all variants identified in both ctDNA and tumor WES for each breast cancer patient. B) Distribution VAF in tumor. C) Distribution of VAF in ctDNA. D) E) F) Similar plots as A) B) C) but from shared only non-synonymous variants. Dashed lines indicate the best linear fits. Each color represents a patient.
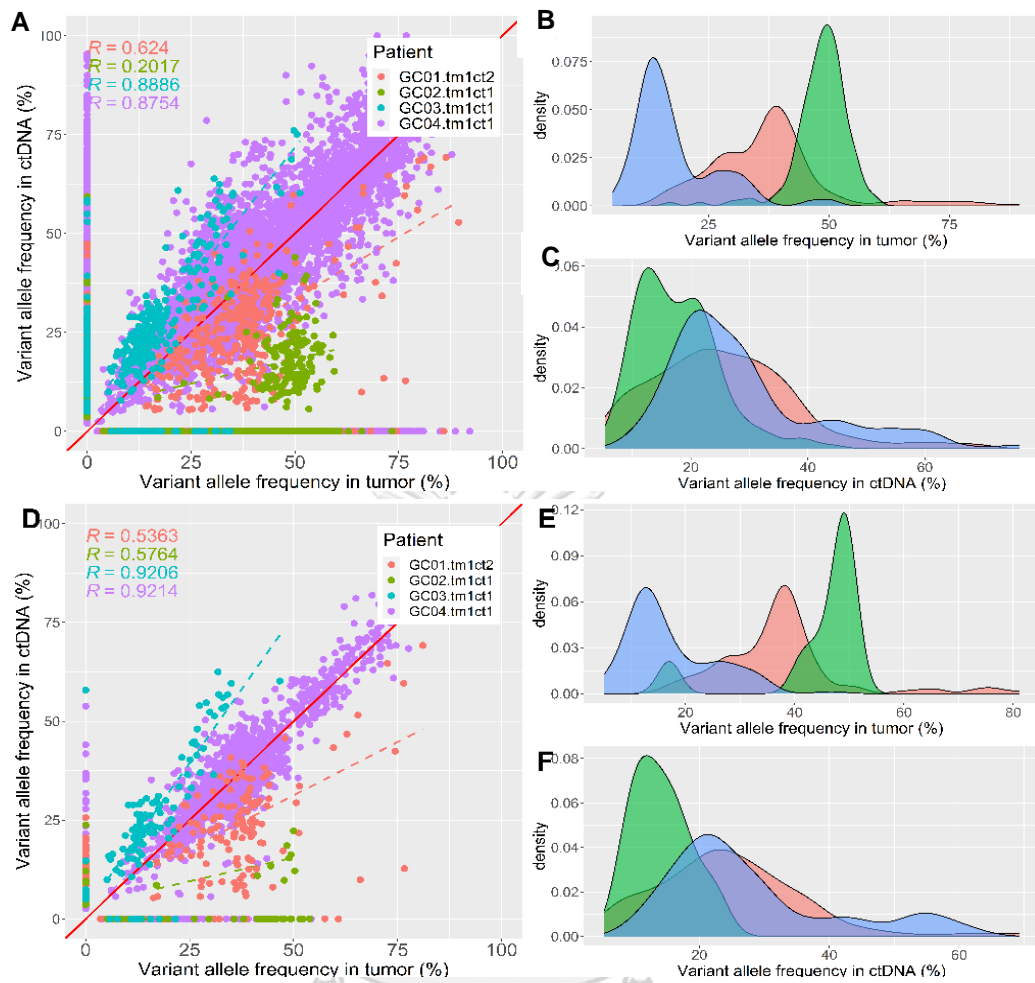
Figure 14 Concordance between ctDNA WES and tumor WES in gastrointestinal cancer patients

4.2.2 CTDNA CAPTURES MUTATIONAL PROFILES FROM MULTIPLE TUMOR SITES

For the melanoma patient with two tumor sites, a primary skin site and a metastatic liver site, tissues from both sites were collected within the same week, and ctDNA sample was collected in the following month. ctDNA WES is slightly more concordant with WES of the skin tissue than with WES of the liver tissue both when considering all variants (Figure 15A) and when considering only non-synonymous variants (Figure 15E). The Jaccard indices for the overlap of identified all variants are 0.11 between ctDNA and skin site and 0.22 between ctDNA and liver site, respectively. The Jaccard indices for the overlap of identified non-synonymous variants are 0.23 and 0.47, respectively. The Pearson's correlations of VAF estimated from ctDNA and tumor tissue are 0.71 and 0.31, respectively, for all variants, and 0.74 and 0.68, respectively, for non-synonymous variants. Furthermore, ctDNA can capture 56 total variants (Figure 15D) and 22 non-synonymous variants (Figure 15H) that overlap between the two tumor sites as well as 38 and 12 additional total and non-synonymous variants, respectively, that are unique to one tumor site. It is interesting that the mutational profile identified from WES of ctDNA is more concordant with the metastatic liver site than the primary skin site (Figure 15D and Figure 15H), while the correlation of VAF between ctDNA and tumor from primary site is still higher than between ctDNA and the metastases site (Figure 15A and 15E). This may be due to two reasons. First, more DNA molecules may be shed from metastatic cells into the bloodstream compared to the primary site. Second, the lower number of unique variants from the metastatic site suggests that the metastatic tumor is less heterogeneous than the primary tumor in terms of variation but the higher VAFs in the primary site still indicate a better relative concentration of primary site variants in ctDNA sample.
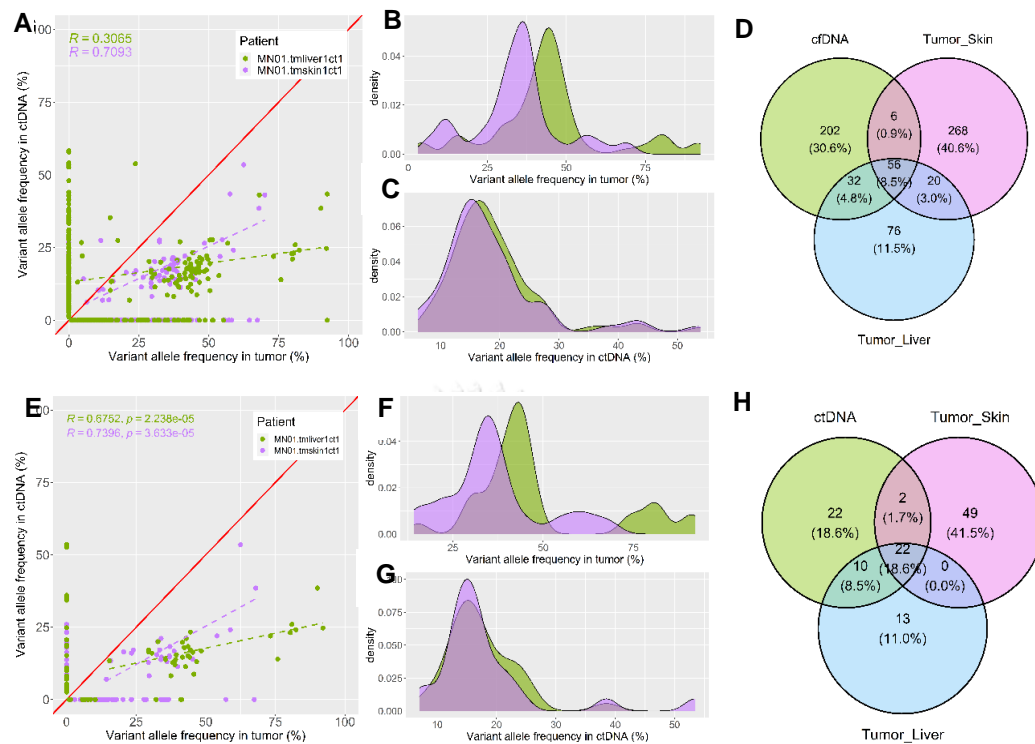
Figure 15 Concordance between ctDNA WES and tumor WES in melanoma patient with multiple tumor sites.
A) Correlation between variant allele frequencies (VAF) of shared all variants identified in both ctDNA and tumor WES for a melanoma patient with two tumor sites. B) Distribution VAF in tumor. C) Distribution of VAF in ctDNA. D) Overlap between all variants identified from ctDNA and tumor tissues from different sites of the melanoma patient. E) F) G) H) Similar plots as A) B) C) D) but from shared only non-synonymous variants.
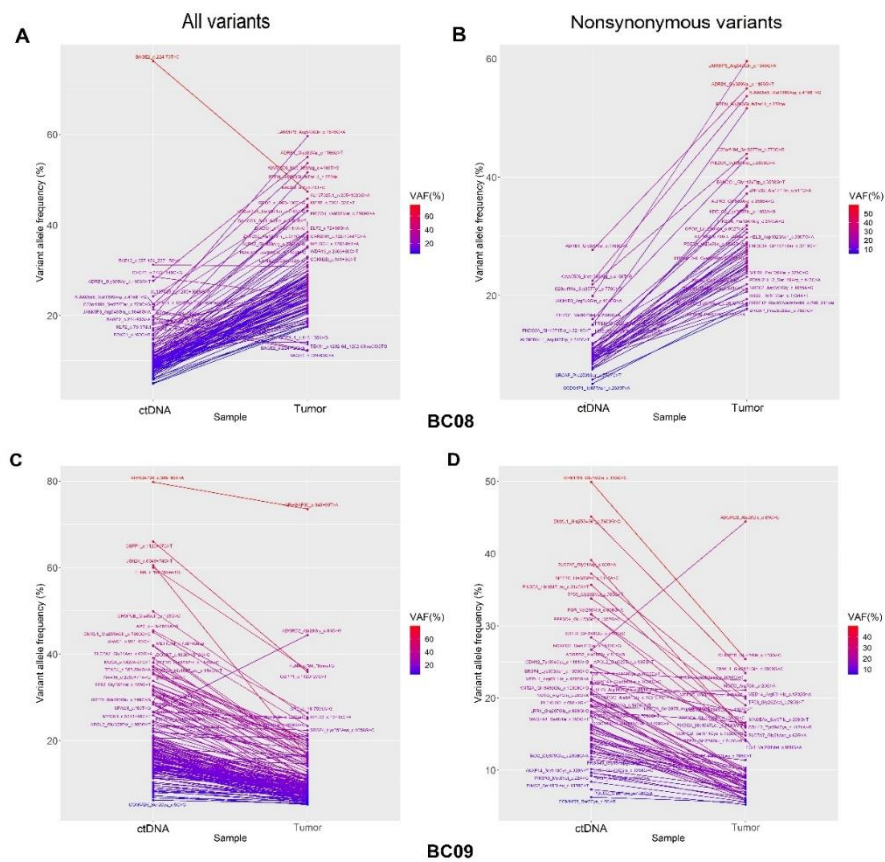
Figure 16 Line plot of paired allele frequencies of concordant variants from tumor and ctDNA which were collected in the same or closest timepoint.
A) B) showed trends of VAF tend to be higher in tumor among all variants and only non-synonymous variants, respectively while C) D) showed VAF trends tend to be higher in ctDNA.
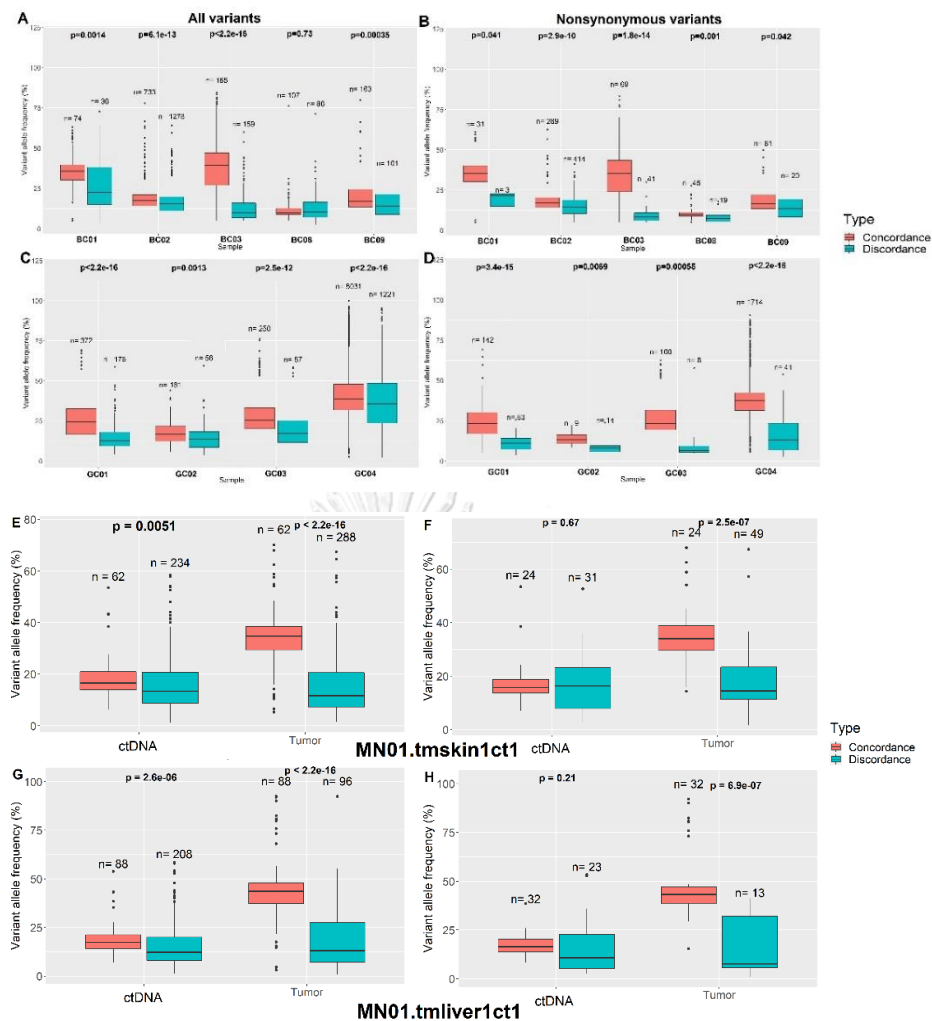
Figure 17 Boxplot of variant allele frequency of ctDNA concordant variant and discordant variants.

Most of concordant groups had high VAF than discordant groups of ctDNA. A) B) from all variants and only non-synonymous variants of breast cancer patient group, respectively C) D) Similar plots as A) B) but from gastrointestinal cancer E) F) Similar plots as A) B) but from a melanoma patient which the concordant variants were from ctDNA and tumor at skin site G) H) Similar plots as E) F) but from tumor at liver site.

4.3 THE DRIFT IN MUTATIONAL PROFILES OVER TIME

Because ctDNA samples were collected at different time interval after tumor tissue biopsies in each case, the extent of the drift of mutational profiles captured by ctDNA over time can also be indirectly evaluated. There is almost no correlation between the amount of time interval between sample collections and the degree of concordance in estimated VAF (Figure 18A, Pearson's correlation = -0.15, p-value = 0.6206) when all variants were considered and a moderate negative correlation when only non-synonymous variants were considered (Figure 18B, Pearson's correlation = -0.49, p-value = 0.0753). The number of discordant variants (unique to ctDNA or to tumor WES) also increases, from a median of 147.5-192 total variants and 18.5-25.5 non-synonymous variants to a median of 130-87.5 and 30.5-18.5, respective, as time passes. A similar finding has been reported (49). However, the drop in concordance, which likely resulted from the drift in mutational profiles as the tumor progresses, reiterate the need for inexpensive, non-invasive technique that can be repeatedly performed to monitor the molecular status of the tumor over time
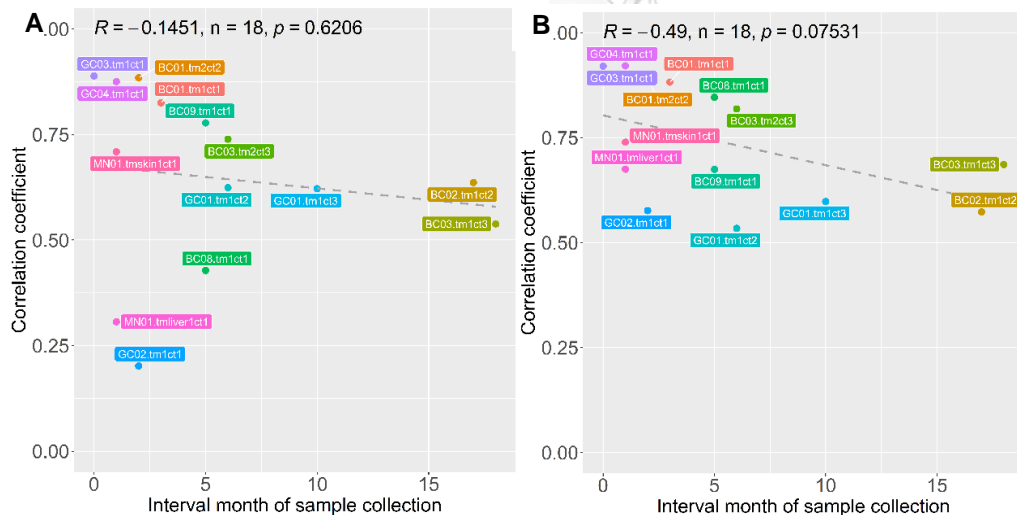
Figure 18 Negative relationship between the time interval between ctDNA and tumor tissue sample collections and the degree of concordance of estimated VAFs A) from all variants (Pearson's correlation = -0.15) B) from only non-synonymous variants (Pearson's correlation = -0.49). Each dot labeled with patient code followed by tumor and ctDNA sample with number of timepoint, respectively.

As there are two cases where ctDNA samples were collected at two timepoints, we can evaluate the drift in mutational profiles overtime in these patients. BC01 patient has two ctDNA and tumor sample timepoints while GC01 has two ctDNA timepoints and a single tumor sample timepoint (Figure 19). The results of BC01 and GC01 showed most variations of ctDNA at 2nd timepoint has concordant with ctDNA and tumor from 1st timepoint. It indicated capability in capturing major variants which highly presented in this patient overtime. In addition, the drift in mutational profile of ctDNA 2nd timepoint of BC01 were only detected by tumor from 2nd timepoint which were collected 2 months previously. There are 11 and 2 variants of all variants and only non-synonymous of ctDNA 2nd timepoint were detected in only tumor at 2nd timepoint. These 2 non-synonymous variants are TEX13D_Ala135Val_c.404C>T and ZNF367_Arg109LeufsTer10_c.326_348del which TEX13D gene has been reported highly point mutation in breast cancer patient from COSMIC database but there is no report in variant level. This indicates that ctDNA capable to capture cancer-relevant gene and show the relationship of reiterated ctDNA along with tumor tissue progression when time passed.

All variants   **Nonsynonymous variants**
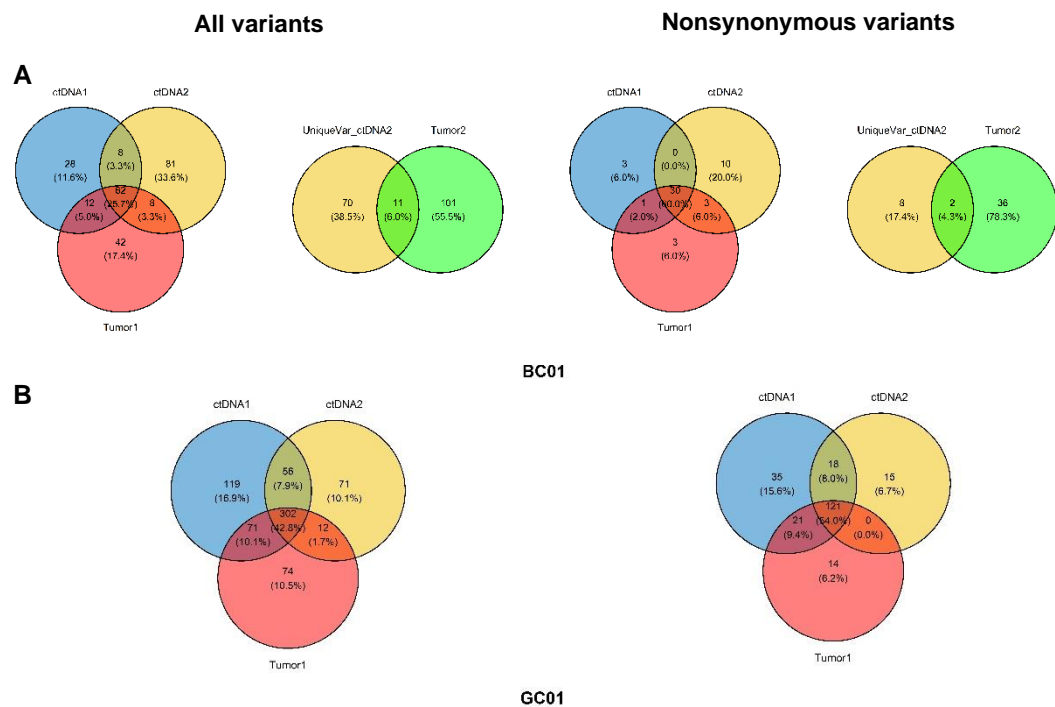
Figure 19 Concordance between ctDNA WES from different timepoints and tumor WES show drift of mutation

A) From BC01 patient who has 2 different timepoint of ctDNA and tumor samples. Unique variants of ctDNA sample (yellow) in 3 Venn diagram, ctDNA two timepoints and the first timepoint of tumor, was compared with second timepoint of tumor sample (green) B) GC01 patient who has 2 timepoints of ctDNA samples and single timepoint of tumor sample.

4.4 WES OF CTDNA AND TUMOR TISSUE IDENTIFY SIMILAR MUTATIONAL LANDSCAPES

The ability of ctDNA WES to capture clinically relevant characteristics of the tumor mutational landscape was evaluated based on whether ctDNA WES can identify similar mutation profile (Figure 20) and base substitution frequencies (Figure 21), which could be matched against known mutational signatures to identify the underlying molecular mechanism of cancer, and variants of the top cancer-related genes as tumor WES.

From Figure 20 found the consistent mutation type profile between ctDNA and tumor samples in individual patients from both non-synonymous and all variants. The profile of mutation type in all variants included the variants from non-coding regions (i.e., 5' UTR, 3'UTR, splice region, intron region) which were reported direct and indirect effects on cancer were also considered (73, 80, 81). Interestingly, even though we performed whole-exome sequencing but the results were able to capture other mutation types which from non-exonic regions, especially intron. Intron mutations were captured at higher rates than most frequent mutations in exon such as missense mutation and were found at the highest number of mutations than all mutation types from both ctDNA and tumor samples. This can be explained by evidence from the previous study that significantly high number of mutations from non-exonic regions can be obtained from exome sequencing data of breast cancer patients. When comparing the number of mutations from exonic regions with non-exonic regions, the number of mutations in non-exonic regions was higher than in exonic regions 1.3 times up to 4.6 times based on different capture reagents in library preparation step. Also, there were reports of 50-64.5% of sequenced bases outside target regions were captured in library preparation step of exome sequencing (82, 83). Despite the fact that a captured library is highly enriched for target regions, a significant fraction of DNA fragments still fall outside of target regions, which is dependent on capture efficiency (84).

For the comparison between profile of base substitution between ctDNA and tumor samples in overall patients showed in pair from left to right, respectively (Figure 21). There were also consistent in all pairs of ctDNA and tumor samples. Moreover, individual patients' signatures from base substitutions can be different or similar to other patients in the same cancer type. For instance, the predominant number of C>T base substitutions which associated with signature 1A/B, 6, 7, 11, 15, or 19 were predominantly found in breast cancer patients, BC01, BC08, BC09, also in gastrointestinal cancer patients, GC03 and GC04. The signatures were found in

patients can indicate underlying mechanism such as signature 1/B associated with elevated rate of spontaneous deamination of 5-methyl-cytosine which mostly found in many cancer types (85).

In Figure 22, Top genes for each cancer type were downloaded from the Cancer Gene Census via COSMIC database (also called tier-1 genes). In almost all cases, the same variants of tier-1 genes and the same base substitution profiles identified in tumor WES could also be identified in ctDNA WES. For breast cancer patients, known oncogenes and tumor suppressor genes such as ESR1, KRAS, PIK3CA, PIK3R1, MUC16 were consistently identified in both ctDNA and tumor WES across patients (Figure 21A). A discordant case, such as the gain of PICK3CA variant in ctDNA WES of patient BC02, do not appear to be correlated with low ctDNA fractions nor low VAFs. For gastrointestinal cancer patients, 8 tier-1 genes were identified among top frequently mutated genes, namely APC, CASP8, GRIN2A, MYH9, TP53, ASXL1, CDH11, and KRAS. All variants of tier1 genes were detected in both ctDNA and tumor samples (Figure 21B). For the melanoma patient with two tumor sites, three tier-1 genes, namely PSIP1, RSPO2, and SF3B1, were identified in both tumor tissues and ctDNA (Figure 15C). Three additional variants of tier-1 genes (BRCA2, MAF, and HOXD11) were identified exclusively in one sample.

From mutational landscape of all samples, WES of ctDNA can consistently capture top cancer-related genes (Cancer Gene Census (71)) identified in tumor tissues from all cancer types (Figure 22), namely ESR1, KRAS, PIK3CA, PIK3R1and MUC16 genes for breast cancer, APC, CASP8, TP53, KRAS, CDH11, GRIN2A, ASXL1 and MYH9 genes for gastrointestinal cancer, and PSIP1, RSPO2, and SF3B1 genes for melanoma/liver cancer. CtDNA WES can also identify additional top cancer genes not found in tumor, but further validations would be required. Overall, these findings indicate that WES of ctDNA is reliable for probing the genetic profiles of tumor.

Top known cancer-related genes in breast cancer patient group, including ESR1, KRAS, PIK3CA, PIK3R1, FAT1 (86-88), MED12 (89, 90) and MUC16 (91-93) genes had been associated with breast cancer. Mutation of ESR1 gene is a common cause of acquired resistance to therapy with an aromatase inhibitor in metastatic cancer. ESR1 may play a role in metastatic progression in breast cancer (94, 95). PIK3CA and PIK3R1 mutations can lead to oncogenesis and hyperactivity of PI3K signaling pathway which is involved with cell cycle, growth and proliferation (96).

Furthermore, KRAS mutations, which is an important factor in tumor initiation, progression, metastatic formation in early stage of colorectal cancer (97) as well as in gastrointestinal carcinomas (98) were detected in both breast cancer and gastrointestinal cancer groups. The top 20 frequently detected genes in gastrointestinal cancer group include known cancer genes, including APC, CASP8, TP53, KRAS, CDH11, GRIN2A, ASXL1 and MYH9. APC have been detected in the majority of colorectal tumors (99). In addition, APC is a potential prognostic biomarker in colorectal cancer (100), as well as GRIN2A (101, 102), ASXL1 (103), MYH9 (104), TP53 (105), CDH11 (106), and CASP8 (107). CASP8 is associated with apoptosis pathway which is one of the mechanisms that generate cell-free tumor DNA in the bloodstream. Presumably, CASP8 mutations may be linked to the level of ctDNA fractions (108).

For the melanoma patient, PSIP1, RSPO2, and SF3B1 gene, which were reported in Cancer Gene Census, are among 22 genes that were found to be mutated in both ctDNA and the two tumor sites. Only SF3B1 gene had been directly reported as a likely pathogenic and common mutation in melanoma (109-111). Although ctDNA can capture additional mutations, these mutations were not strongly associated with melanoma.
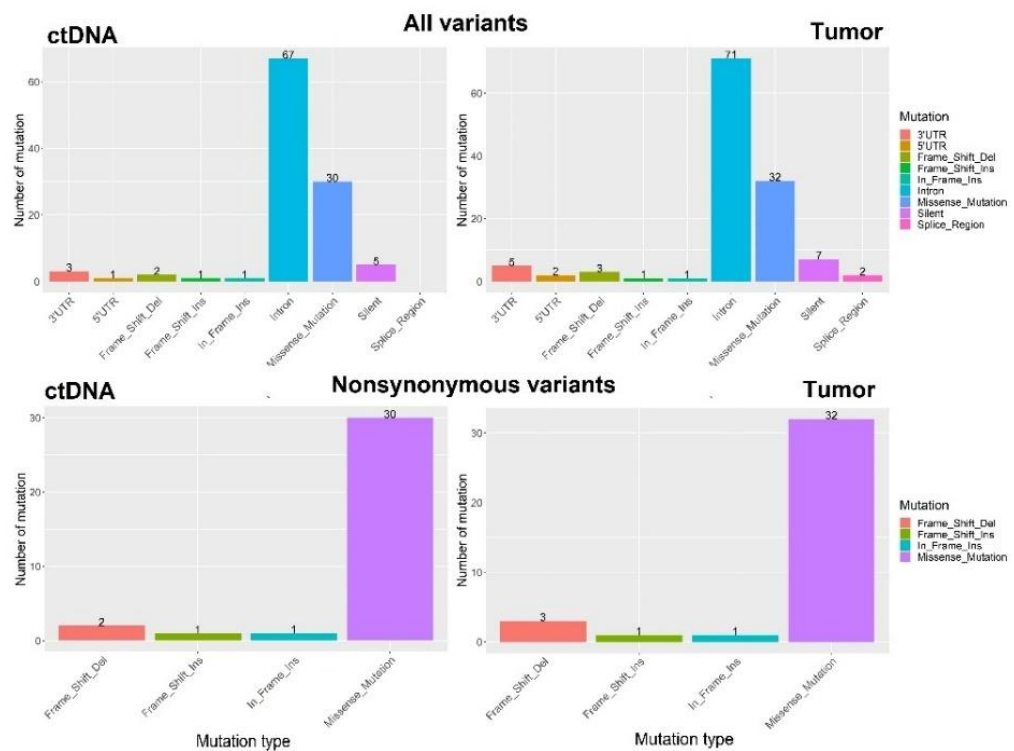
Figure 20 The profile of mutation type

An example profile of mutation type from single patient which compared between

ctDNA (left) and tumor (right) sample from all variants (top) which included

synonymous and non-coding regions.
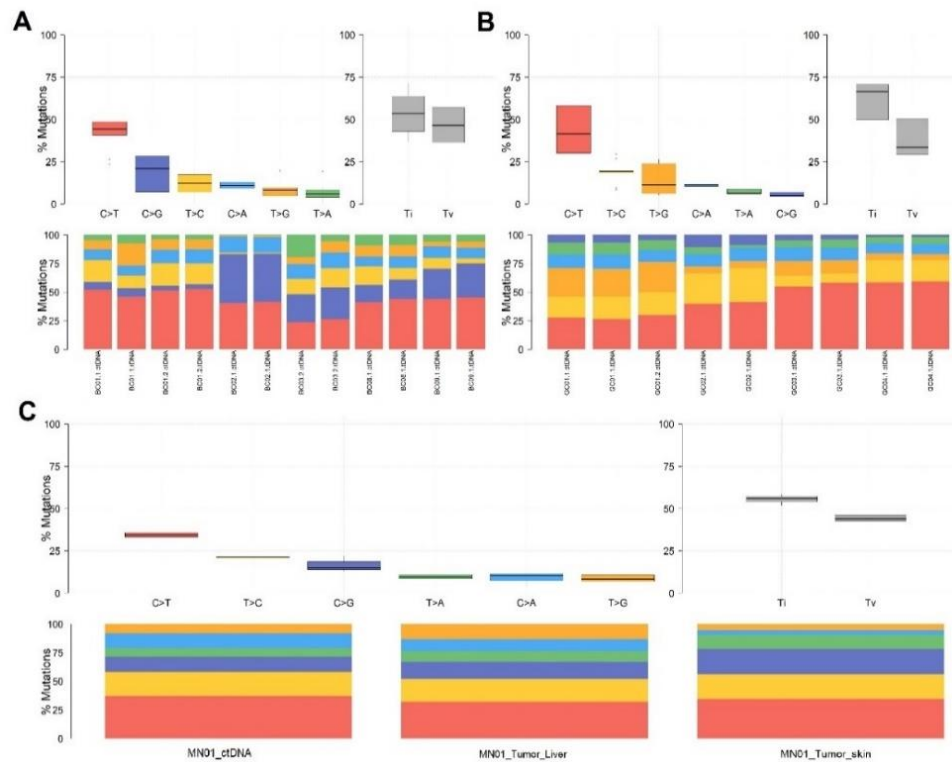
Figure 21 The profile of base substitution

A) The profile of base substitutions in 5 breast cancer patients consists of 6 pairs of ctDNA and tumor samples which there is a BC01 patient with two timepoints. Top left box plot showed overall base substitutions from all patients and percentage of mutations were calculated from number of each base substitution type with total number of base substitutions in each sample. C>T, C>G, T>C, C>A, T>G, and T>A base substitutions were indicated in red, purple, yellow, blue, orange, and green color, respectively. Top right box plot showed type of base substitution, transition (Ti) and transversion (Tv), from overall patients with percentage of mutations. The overall profile of base substitutions in the group showed on the bottom. B) as similar to A) but in 4 gastrointestinal cancer patients with GC01 patient, there is two ctDNA samples from different timepoint with one tumor sample. C) as similar to A) and B) but in a melanoma patient with two multiple tumor sites and a ctDNA sample.

Figure 22 ctDNA WES identifies similar mutation landscape of top cancer-related genes as tumor tissue's.

Top cancer-related genes for each cancer type were downloaded from Cancer Gene Census via COSMIC database. Genes were ordered based on mutation frequencies across patients of the same cancer type (Frequencies shown on the right y-axis). Tumor mutational burden (TMB) for each tumor is indicated at the top of the plot. Sample types and patient IDs are indicated with colors just below the plot. Base substitution profiles are shown at the bottom of the plot. A) The mutational profiles identified from ctDNA and tumor WES of breast cancer patients. B) Similar plot for gastrointestinal cancer patients. C) Similar plot for the melanoma patient with two tumor sites.

CHAPTER 5   CONCLUSIONS AND RECOMMENDATIONS

Our study has assessed the capability of ctDNA to capture tumor mutational profile and evaluated concordance between tumor and ctDNA variants from different sampling time points. We have shown that the level of ctDNA fraction is related to the degree of correlation between somatic variants identified in ctDNA and tumor tissue samples. Our result indicates that increasing the ctDNA fraction threshold from previous report can improve the quality of WES analysis and variant calling of ctDNA samples.   We found that including synonymous variants into the concordance analysis improves the correlation between ctDNA and tumor in some cases. Hence, we recommend doing this on a case-by-case basis. In addition, we identify small changes in mutation profiles over time that do not significantly affects the degree of concordance between ctDNA and tumor samples that were collected several months apart. However, the shorter gap time between tumor tissue and ctDNA sample collection led to higher concordance level. Lastly, WES analysis of ctDNA has potential to capture genetic alterations from multiple tumor sites within the same patients and complement WES analysis of tumor tissue. From these advantages, ctDNA analysis has a strong potential for clinical diagnostic, prognostic, and treatment follow-up.

Limitation of this study

The number of patients is not high enough to establish statistical significance regarding the concordance between whole exome sequencing of ctDNA and tumor tissue, especially as there are multiple factors, such as cancer type, ctDNA fraction, and time interval between sample collections, that contribute to the variability of mutational profiles identified from ctDNA. Due to logistical and patient health issues, we could not obtain samples from more time points to analyze the drift in mutational profiles over time quantitatively. These issues also lengthened the time interval between tumor tissue biopsy and ctDNA sample collections, which consequently affected the level of concordance. The analysis of the drift in mutational profiles over time was not ideal because the trend was based on data

from multiple patients with different molecular characteristics and it lacks validation. Nonetheless, this study's cohort is still consisted of samples with diverse molecular signatures and ctDNA characteristics and should be able to represent real applications of whole exome sequencing on ctDNA samples.

APPENDICES

APPENDIX A

Patient	Type	Project	Tumor1	ctDNA1	ULP-WGS1	TF	*	**	Tumor2	ctDNA2	ULP-WGS2	TF	*	**	Tumor3	ctDNA3	ULP-WGS3	TF	*
BC01	Breast	QSCBC	12/2019	03/2020	12/2019	73.44	3	3	05/2021	07/2021	07/2021	86.38	2	0	-	-	-	-	-
BC02	Breast	QSCBC	07/2019	-	-	-	-	-	-	12/2020	12/2020	39.24	-	0	-	-	-	-	-
BC03	Breast	QSCBC	09/2019	-	-	-	-	-	09/2020	07/2020	06/2020	8.07	2	1	-	03/2021	03/2021	88.08	0
BC04	Breast	QSCBC	-	03/2020	01/2020	8.13	-	2	-	-	-	-	-	-	-	-	-	-	-
BC05	Breast	QSCBC	05/2020	04/2020	01/2020	7.2	1	3	-	-	-	-	-	-	-	-	-	-	-
BC06	Breast	QSCBC	-	03/2020	02/2020	43.84	-	1	-	-	-	-	-	-	-	-	-	-	-
BC07	Breast	QSCBC	-	03/2020	12/2019	9.35	-	3	03/2021	-	-	-	-	-	-	-	-	-	-
BC08	Breast	QSCBC	06/2021	11/2021	10/2021	59.95	5	1	-	-	-	-	-	-	-	-	-	-	-
BC09	Breast	QSCBC	06/2021	11/2021	11/2021	39.06	5	0	-	-	-	-	-	-	-	-	-	-	-
GC01	Colon	MedOnco	09/2020	-	09/2020	56.03	-	-	-	03/2021	01/2021	34.86	-	2	-	07/2021	07/2021	39.72	0
GC02	Colon	MedOnco	12/2020	02/2021	10/2020	84.83	2	4	-	-	-	-	-	-	-	-	-	-	-
GC03	Cecum	MedOnco	12/2020	12/2020	10/2020	39.38	0	2	-	-	-	-	-	-	-	-	-	-	-
GC04	Duodenum	MedOnco	01/2021	02/2021	01/2021	83.13	1	1	-	-	-	-	-	-	-	-	-	-	-
MN01	Melanoma	MedOnco	02/2021	03/2021	01/2021	44.63	1	2	-	-	-	-	-	-	-	-	-	-	-
SC01	Sarcoma	QSCBC	08/2020	07/2020	06/2020	16.42	1	1	06/2021	-	04/2021	11.16	-	-	-	-	-	-	-

* Interval time between tumor sample collection and blood sample collection of ctDNA WES (months), ** Interval time between blood sample collection of ctDNA WES and ctDNA ULP-WGS (months), The number following by Tumor, ctDNA and ULP-WGS defined the timepoint of sample collection which indicate sample collection date, **TF** indicates ctDNA fraction

APPENDIX B

Patient	Tumor	ctDNA	TF	Interval time	Pearson's correlation (All variants)	Pearson's correlation (Only non-synonymous)	level of overlap (% of ctDNA variant found in tumor) (All variants)	level of overlap (% of ctDNA variant found in tumor) (Only non-synonymous)	level of overlap (% of tumor variant found in ctDNA) (All variants)	level of overlap (% of tumor variant found in ctDNA) (Only non-synonymous)	Jaccard index (All variants)	Jaccard index (only non-synonymous)
BC01.tm1ct1	12/2019	03/2020	73.44	3	0.8252	0.8826	59.68	83.78	67.27	91.18	0.463	0.7750
BC01.tm2ct2	05/2021	07/2021	86.38	2	0.8844	0.8857	72.32	86.84	50.94	76.74	0.426	0.6875
BC02.tm1ct2	07/2019	12/2020	39.24	17	0.6362	0.5734	63.63	71.18	36.45	41.11	0.302	0.3524
BC03.tm2ct3	09/2020	03/2021	88.08	6	0.7388	0.8187	50.15	70.41	50.93	62.73	0.338	0.4964
BC08.tm1ct1	06/2021	11/2021	59.95	5	0.4278	0.8465	54.59	66.18	55.44	70.31	0.379	0.5172
BC09.tm1ct1	06/2021	11/2021	39.06	5	0.7775	0.6746	87.17	91.04	61.74	75.31	0.566	0.7011
GC01.tm1ct2	09/2020	03/2021	34.86	6	0.6238	0.5363	81.26	91.03	68.07	72.82	0.588	0.6794
GC02.tm1ct1	12/2020	02/2021	84.83	2	0.2017	0.5764	18.79	19.15	74.19	39.13	0.176	0.1475
GC03.tm1ct1	12/2020	12/2020	39.38	0	0.8886	0.9206	88.03	91.74	74.18	92.59	0.674	0.8547
GC04.tm1ct1	01/2021	02/2021	83.13	1	0.8754	0.9214	81.46	90.02	86.80	97.66	0.725	0.8812
MN01.tmskin1ct1	02/2021	03/2021	44.63	1	0.7093	0.7396	17.71	32.88	20.95	43.64	0.106	0.2308
MN01.tmliver1ct1	02/2021	03/2021	44.63	1	0.3065	0.6752	47.83	71.11	29.73	58.18	0.224	0.4706

REFERENCES

1.   Eichler EE. Genetic Variation, Comparative Genomics, and the Diagnosis of Disease. New England Journal of Medicine. 2019;381(1):64-74.

2.   Leary RJ, Sausen M, Kinde I, Papadopoulos N, Carpten JD, Craig D, et al. Detection of chromosomal alterations in the circulation of cancer patients with whole-genome sequencing. Sci Transl Med. 2012;4(162):162ra54.

3.   Goya R, Sun MG, Morin RD, Leung G, Ha G, Wiegand KC, et al. SNVMix: predicting single nucleotide variants from next-generation sequencing of tumors. Bioinformatics. 2010;26(6):730-6.

4.   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.

5.   Shohdy KS, West H. Circulating Tumor DNA Testing—Liquid Biopsy of a Cancer. JAMA Oncology. 2020;6(5):792-.

6.   Winn JS, Hasse Z, Slifker M, Pei J, Arisi-Fernandez SM, Talarchek JN, et al. Genetic Variants Detected Using Cell-Free DNA from Blood and Tumor Samples in Patients with Inflammatory Breast Cancer. Int J Mol Sci. 2020;21(4).

7.   Ignatiadis M, Sledge GW, Jeffrey SS. Liquid biopsy enters the clinic - implementation issues and future challenges. Nat Rev Clin Oncol. 2021;18(5):297-312.

8.   Diaz LA, Jr., Bardelli A. Liquid biopsies: genotyping circulating tumor DNA. J Clin Oncol. 2014;32(6):579-86.

9.   Morris VK, Strickler JH. Use of Circulating Cell-Free DNA to Guide Precision Medicine in Patients with Colorectal Cancer. Annu Rev Med. 2021;72:399-413.

10.   Pereira B, Chen CT, Goyal L, Walmsley C, Pinto CJ, Baiev I, et al. Cell-free DNA captures tumor heterogeneity and driver alterations in rapid autopsies with pre-treated metastatic cancer. Nat Commun. 2021;12(1):3199.

11.   Yan YY, Guo QR, Wang FH, Adhikari R, Zhu ZY, Zhang HY, et al. Cell-Free DNA: Hope and Potential Application in Cancer. Front Cell Dev Biol. 2021;9:639233.

12.   Volik S, Alcaide M, Morin RD, Collins C. Cell-free DNA (cfDNA): Clinical Significance and Utility in Cancer Shaped By Emerging Technologies. Mol Cancer Res.

2016;14(10):898-908.

13.   Zhang L, Liang Y, Li S, Zeng F, Meng Y, Chen Z, et al. The interplay of circulating tumor DNA and chromatin modification, therapeutic resistance, and metastasis. Molecular Cancer. 2019;18(1):36.

14.   Sabour L, Sabour M, Ghorbian S. Clinical Applications of Next-Generation Sequencing in Cancer Diagnosis. Pathol Oncol Res. 2017;23(2):225-34.

15.   Singh RR. Next-Generation Sequencing in High-Sensitive Detection of Mutations in Tumors: Challenges, Advances, and Applications. J Mol Diagn. 2020;22(8):994-1007.

16.   Liu B, Morrison CD, Johnson CS, Trump DL, Qin M, Conroy JC, et al. Computational methods for detecting copy number variations in cancer genome using next generation sequencing: principles and challenges. Oncotarget. 2013;4(11):1868-81.

17.   Jiang X, Li H, Liu J, Sun H, Zhang L, Li W, et al. Feasibility Analysis of Cell-Free DNA Derived from Plasma of Lung Cancer Patients for Next-Generation Sequencing. Biopreserv Biobank. 2020;18(2):117-21.

18.   Shyr D, Liu Q. Next generation sequencing in cancer research and clinical application. Biol Proced Online. 2013;15(1):4.

19.   Nguyen B, Wong NC, Semple T, Clark M, Wong SQ, Leslie C, et al. Low-coverage whole-genome sequencing of extracellular vesicle-associated DNA in patients with metastatic cancer. Scientific Reports. 2021;11(1):4016.

20.   Homburger JR, Neben CL, Mishne G, Zhou AY, Kathiresan S, Khera AV. Low coverage whole genome sequencing enables accurate assessment of common variants and calculation of genome-wide polygenic scores. Genome Medicine. 2019;11(1):74.

21.   Wang IP. Low-Coverage Whole Genome Sequencing: Learning From Less Data 2019 [Available from: https://www.cancer.gov/about-nci/organization/ccg/blog/2019/low-coverage-seq.

22.   Weber ZT, Collier KA, Tallman D, Forman J, Shukla S, Asad S, et al. Modeling clonal structure over narrow time frames via circulating tumor DNA in metastatic breast cancer. Genome Medicine. 2021;13(1):89.

23.   Choudhury AD, Werner L, Francini E, Wei XX, Ha G, Freeman SS, et al. Tumor fraction in cell-free DNA as a biomarker in prostate cancer. JCI Insight. 2018;3(21).

24.   Suppan C, Brcic I, Tiran V, Mueller HD, Posch F, Auer M, et al. Untargeted

Assessment of Tumor Fractions in Plasma for Monitoring and Prognostication from Metastatic Breast Cancer Patients Undergoing Systemic Treatment. Cancers (Basel). 2019;11(8).

25.   Shulman DS, Klega K, Imamovic-Tuco A, Clapp A, Nag A, Thorner AR, et al. Detection of circulating tumour DNA is associated with inferior outcomes in Ewing sarcoma and osteosarcoma: a report from the Children's Oncology Group. Br J Cancer. 2018;119(5):615-21.

26.   Chen X, Chang CW, Spoerke JM, Yoh KE, Kapoor V, Baudo C, et al. Low-pass Whole-genome Sequencing of Circulating Cell-free DNA Demonstrates Dynamic Changes in Genomic Copy Number in a Squamous Lung Cancer Clinical Cohort. Clin Cancer Res. 2019;25(7):2254-63.

27.   Wan N, Weinberg D, Liu TY, Niehaus K, Ariazi EA, Delubac D, et al. Machine learning enables detection of early-stage colorectal cancer by whole-genome sequencing of plasma cell-free DNA. BMC Cancer. 2019;19(1):832.

28.   Manzoni M, Pompa A, Fabris S, Pelizzoni F, Ciceri G, Seia M, et al. Limits and Applications of Genomic Analysis of Circulating Tumor DNA as a Liquid Biopsy in Asymptomatic Forms of Multiple Myeloma. Hemasphere. 2020;4(4):e402.

29.   Fu J, Guo W, Yan C, Lv Z, Wang Y, Wang Z, et al. Combining targeted sequencing and ultra-low-pass whole-genome sequencing for accurate somatic copy number alteration detection. Funct Integr Genomics. 2021;21(2):161-9.

30.   Jahangiri L, Hurst T. Assessing the Concordance of Genomic Alterations between Circulating-Free DNA and Tumour Tissue in Cancer Patients. Cancers (Basel). 2019;11(12).

31.   Bos MK, Angus L, Nasserinejad K, Jager A, Jansen M, Martens JWM, et al. Whole exome sequencing of cell-free DNA - A systematic review and Bayesian individual patient data meta-analysis. Cancer Treat Rev. 2020;83:101951.

32.   Adalsteinsson VA, Ha G, Freeman SS, Choudhury AD, Stover DG, Parsons HA, et al. Scalable whole-exome sequencing of cell-free DNA reveals high concordance with metastatic tumors. Nature Communications. 2017;8(1):1324.

33.   Institute NC. What Is Cancer? 2021 [Available from: https://www.cancer.gov/about-cancer/understanding/what-is-cancer.

34.   Cooper G. The Development and Causes of Cancer. The Cell: A Molecular Approach: Sunderland (MA): Sinauer Associates; 2000.

35.   Hanahan D, Weinberg RA. Hallmarks of cancer: the next generation. Cell. 2011;144(5):646-74.

36.   Taningher M, Saccomanno G, Santi L, Grilli S, Parodi S. Quantitative predictability of carcinogenicity of the covalent binding index of chemicals to DNA: comparison of the in vivo and in vitro assays. Environ Health Perspect. 1990;84:183-92.

37.   Liu Y, Yin T, Feng Y, Cona MM, Huang G, Liu J, et al. Mammalian models of chemically induced primary malignancies exploitable for imaging-based preclinical theragnostic research. Quant Imaging Med Surg. 2015;5(5):708-29.

38.   Carcinogenesis 2020 [updated 2020/8/14/. Available from: https://med.libretexts.org/@go/page/2161.

39.   Oliveira P, Colaço A, Chaves R, Guedes-Pinto H, P L, Lopes C. Chemical carcinogenesis. Anais da Academia Brasileira de Ciências. 2008;79:593-616.

40.   Saria MG. Overview of cancer. 2018.

41.   Bailey R. Genetic Variation Definition, Causes, and Examples 2021 [Available from: https://www.thoughtco.com/genetic-variation-373457.

42.   Mahon MS. Germline and Somatic Variants: What Is the Difference? 2020 [updated 2020 July 14. Available from: https://voice.ons.org/news-and-views/germline-and-somatic-variants-what-is-the-difference.

43.   Yi K, Ju YS. Patterns and mechanisms of structural variations in human cancer. Experimental & Molecular Medicine. 2018;50(8):1-11.

44.   Tate JG, Bamford S, Jubb HC, Sondka Z, Beare DM, Bindal N, et al. COSMIC: the Catalogue Of Somatic Mutations In Cancer. Nucleic Acids Research. 2018;47(D1):D941-D7.

45.   team tUE. Types of mutations: UC Museum of Paleontology 2022 [Available from: https://evolution.berkeley.edu/dna-and-mutations/types-of-mutations/.

46.   Scoville H. Synonymous vs. Nonsynonymous Mutations 2021 [Available from: https://www.thoughtco.com/synonymous-vs-nonsynonymous-mutations-1224600.

47.   Sehn JK. Chapter 9 - Insertions and Deletions (Indels). In: Kulkarni S, Pfeifer J, editors. Clinical Genomics. Boston: Academic Press; 2015. p. 129-50.

48.   Hastings PJ, Lupski JR, Rosenberg SM, Ira G. Mechanisms of change in gene copy number. Nature Reviews Genetics. 2009;10(8):551-64.

49.   Gallegos H. Inversion Mutation in Chromosome 2021 [Available from: https://study.com/learn/lesson/inversion-genetic-mutation-chromosomes.html#:~:text=An%20inversion%20mutation%20is%20one,it%20more%20difficult%20to%20read.

50.   O'Connor C. Human Chromosome Translocations and Cancer. Nature Education. 2008;1(1):56.

51.   Koboldt DC. Best practices for variant calling in clinical sequencing. Genome Medicine. 2020;12(1):91.

52.   Cibulskis K, Lawrence MS, Carter SL, Sivachenko A, Jaffe D, Sougnez C, et al. Sensitive detection of somatic point mutations in impure and heterogeneous cancer samples. Nature Biotechnology. 2013;31(3):213-9.

53.   Kim S, Scheffler K, Halpern AL, Bekritsky MA, Noh E, Källberg M, et al. Strelka2: fast and accurate calling of germline and somatic variants. Nature Methods. 2018;15(8):591-4.

54.   Koboldt DC, Zhang Q, Larson DE, Shen D, McLellan MD, Lin L, et al. VarScan 2: somatic mutation and copy number alteration discovery in cancer by exome sequencing. Genome Res. 2012;22(3):568-76.

55.   Wang G, Yandi X, Wei Z, Yuan Y, Qiu M, Sun C, et al. Toward the use of precision medicine for the treatment of head and neck squamous cell carcinoma. Oncotarget. 2015;8.

56.   Morash M, Mitchell H, Beltran H, Elemento O, Pathak J. The Role of Next-Generation Sequencing in Precision Medicine: A Review of Outcomes in Oncology. J Pers Med. 2018;8(3).

57.   Tsimberidou AM, Iskander NG, Hong DS, Wheler JJ, Falchook GS, Fu S, et al. Personalized medicine in a phase I clinical trials program: the MD Anderson Cancer Center initiative. Clin Cancer Res. 2012;18(22):6373-83.

58.   Radovich M, Kiel PJ, Nance SM, Niland EE, Parsley ME, Ferguson ME, et al. Clinical benefit of a precision medicine based approach for guiding treatment of refractory cancers. Oncotarget. 2016;7(35):56491-500.

59.   Schwaederle M, Parker BA, Schwab RB, Daniels GA, Piccioni DE, Kesari S, et al. Precision Oncology: The UC San Diego Moores Cancer Center PREDICT Experience. Molecular Cancer Therapeutics. 2016;15(4):743-52.

60.   Schwartzberg L, Kim ES, Liu D, Schrag D. Precision Oncology: Who, How, What, When, and When Not? American Society of Clinical Oncology Educational Book. 2017(37):160-9.

61.   Hirahata T, Ul Quraish R, Quraish AU, Ul Quraish S, Naz M, Razzaq MA. Liquid Biopsy: A Distinctive Approach to the Diagnosis and Prognosis of Cancer. Cancer Inform. 2022;21:11769351221076062.

62.   Hu Z, Chen H, Long Y, Li P, Gu Y. The main sources of circulating cell-free DNA: Apoptosis, necrosis and active secretion. Crit Rev Oncol Hematol. 2021;157:103166.

63.   Eslami-S Z, Cortés-Hernández L, Alix-Panabières C. The Metastatic Cascade as the Basis for Liquid Biopsy Development. Frontiers in Oncology. 2020;10:1055.

64.   Zhong Y, Xu F, Wu J, Schubert J, Li MM. Application of Next Generation Sequencing in Laboratory Medicine. Ann Lab Med. 2021;41(1):25-43.

65.   Illumina. An introduction to Next-Generation Sequencing Technology 2022 [Available from: www.illumina.com/technology/next-generation-sequencing.html.

66.   Strianese O, Rizzo F, Ciccarelli M, Galasso G, D'Agostino Y, Salvati A, et al. Precision and Personalized Medicine: How Genomic Approach Improves the Management of Cardiovascular and Neurodegenerative Disease. Genes (Basel). 2020;11(7).

67.   Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics. 2009;25(14):1754-60.

68.   McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernytsky A, et al. The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. Genome Res. 2010;20(9):1297-303.

69.   Poplin R, Ruano-Rubio V, DePristo MA, Fennell TJ, Carneiro MO, Van der Auwera GA, et al. Scaling accurate genetic variant discovery to tens of thousands of samples. bioRxiv. 2018:201178.

70.   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.

71.   Sondka Z, Bamford S, Cole CG, Ward SA, Dunham I, Forbes SA. The COSMIC Cancer Gene Census: describing genetic dysfunction across all human cancers. Nature Reviews Cancer. 2018;18(11):696-705.

72.   Mayakonda A, Lin DC, Assenov Y, Plass C, Koeffler HP. Maftools: efficient and comprehensive analysis of somatic variants in cancer. Genome Res. 2018;28(11):1747-56.

73.   Supek F, Miñana B, Valcárcel J, Gabaldón T, Lehner B. Synonymous Mutations Frequently Act as Driver Mutations in Human Cancers. Cell. 2014;156(6):1324-35.

74.   Gotea V, Gartner JJ, Qutob N, Elnitski L, Samuels Y. The functional relevance of somatic synonymous mutations in melanoma and other cancers. Pigment Cell Melanoma Res. 2015;28(6):673-84.

75.   Haider S, Tyekucheva S, Prandi D, Fox NS, Ahn J, Xu AW, et al. Systematic Assessment of Tumor Purity and Its Clinical Implications. JCO Precision Oncology. 2020(4):995-1005.

76.   Shin H-T, Choi Y-L, Yun JW, Kim NKD, Kim S-Y, Jeon HJ, et al. Prevalence and detection of low-allele-fraction variants in clinical cancer samples. Nature Communications. 2017;8(1):1377.

77.   Patel NM, Jo H, Eberhard DA, Yin X, Hayward MC, Stein MK, et al. Improved Tumor Purity Metrics in Next-generation Sequencing for Clinical Practice: The Integrated Interpretation of Neoplastic Cellularity and Sequencing Results (IINCaSe) Approach. Appl Immunohistochem Mol Morphol. 2019;27(10):764-72.

78.   Mao Y, Feng Q, Zheng P, Yang L, Liu T, Xu Y, et al. Low tumor purity is associated with poor prognosis, heavy mutation burden, and intense immune phenotype in colon cancer. Cancer Manag Res. 2018;10:3569-77.

79.   Xu B, Shan G, Wu Q, Li W, Wang H, Li H, et al. Concordance of Genomic Alterations between Circulating Tumor DNA and Matched Tumor Tissue in Chinese Patients with Breast Cancer. J Oncol. 2020;2020:4259293.

80.   Schuster SL, Hsieh AC. The Untranslated Regions of mRNAs in Cancer. Trends Cancer. 2019;5(4):245-62.

81.   Cuykendall TN, Rubin MA, Khurana E. Non-coding genetic variation in cancer.

Curr Opin Syst Biol. 2017;1:9-15.

82.   Ng SB, Buckingham KJ, Lee C, Bigham AW, Tabor HK, Dent KM, et al. Exome sequencing identifies the cause of a mendelian disorder. Nature Genetics. 2010;42(1):30-5.

83.   Yi X, Liang Y, Huerta-Sanchez E, Jin X, Cuo ZX, Pool JE, et al. Sequencing of 50 human exomes reveals adaptation to high altitude. Science. 2010;329(5987):75-8.

84.   Guo Y, Long J, He J, Li CI, Cai Q, Shu XO, et al. Exome sequencing generates high quality data in non-target regions. BMC Genomics. 2012;13:194.

85.   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.

86.   Lee S, Stewart S, Nagtegaal I, Luo J, Wu Y, Colditz G, et al. Differentially expressed genes regulating the progression of ductal carcinoma in situ to invasive breast cancer. Cancer Res. 2012;72(17):4574-86.

87.   Peng Z, Gong Y, Liang X. Role of FAT1 in health and disease. Oncol Lett. 2021;21(5):398.

88.   Katoh M. Function and cancer genomics of FAT family genes (review). Int J Oncol. 2012;41(6):1913-8.

89.   Laé M, Gardrat S, Rondeau S, Richardot C, Caly M, Chemlali W, et al. MED12 mutations in breast phyllodes tumors: evidence of temporal tumoral heterogeneity and identification of associated critical signaling pathways. Oncotarget. 2016;7(51):84428-38.

90.   Lim WK, Ong CK, Tan J, Thike AA, Ng CCY, Rajasegaran V, et al. Exome sequencing identifies highly recurrent MED12 somatic mutations in breast fibroadenoma. Nature Genetics. 2014;46(8):877-80.

91.   Aithal A, Rauth S, Kshirsagar P, Shah A, Lakshmanan I, Junker WM, et al. MUC16 as a novel target for cancer therapy. Expert Opin Ther Targets. 2018;22(8):675-86.

92.   Lakshmanan I, Ponnusamy MP, Das S, Chakraborty S, Haridas D, Mukhopadhyay P, et al. MUC16 induced rapid G2/M transition via interactions with JAK2 for increased proliferation and anti-apoptosis in breast cancer cells. Oncogene. 2012;31(7):805-17.

93.   Keup C, Benyaa K, Hauch S, Sprenger-Haussels M, Tewes M, Mach P, et al. Targeted deep sequencing revealed variants in cell-free DNA of hormone receptor-positive metastatic breast cancer patients. Cell Mol Life Sci. 2020;77(3):497-509.

94. Brett JO, Spring LM, Bardia A, Wander SA. ESR1 mutation as an emerging clinical biomarker in metastatic hormone receptor-positive breast cancer. Breast Cancer Research. 2021;23(1):85.

95. Dustin D, Gu G, Fuqua SAW. ESR1 mutations in breast cancer. Cancer. 2019;125(21):3714-28.

96. Chen L, Yang L, Yao L, Kuang XY, Zuo WJ, Li S, et al. Characterization of PIK3CA and PIK3R1 somatic mutations in Chinese breast cancer patients. Nat Commun. 2018;9(1):1357.

97. Smakman N, Borel Rinkes IH, Voest EE, Kranenburg O. Control of colorectal metastasis formation by K-Ras. Biochim Biophys Acta. 2005;1756(2):103-14.

98. Velho S, Corso G, Oliveira C, Seruca R. KRAS signaling pathway alterations in microsatellite unstable gastrointestinal cancers. Adv Cancer Res. 2010;109:123-43.

99. Fearnhead NS, Britton MP, Bodmer WF. The ABC of APC. Hum Mol Genet. 2001;10(7):721-33.

100. Tanaka N, Mashima T, Mizutani A, Sato A, Aoyama A, Gong B, et al. APC Mutations as a Potential Biomarker for Sensitivity to Tankyrase Inhibitors in Colorectal Cancer. Molecular Cancer Therapeutics. 2017;16(4):752-62.

101. Palaniappan A, Ramar K, Ramalingam S. Computational Identification of Novel Stage-Specific Biomarkers in Colorectal Cancer Progression. PLoS One. 2016;11(5):e0156665.

102. Liu D, Ma X, Yang F, Xiao D, Jia Y, Wang Y. Discovery and validation of methylated-differentially expressed genes in Helicobacter pylori-induced gastric cancer. Cancer Gene Therapy. 2020;27(6):473-85.

103. Lee JH, Lee JH, Ahn BK, Paik SS, Kim H, Lee KH. Loss of ASXL1 expression is associated with lymph node metastasis in colorectal cancer. Indian J Pathol Microbiol. 2020;63(2):221-5.

104. Wang B, Qi X, Liu J, Zhou R, Lin C, Shangguan J, et al. MYH9 Promotes Growth and Metastasis via Activation of MAPK/AKT Signaling in Colorectal Cancer. J Cancer. 2019;10(4):874-84.

105. Rokavec M, Li H, Jiang L, Hermeking H. The p53/microRNA connection in gastrointestinal cancer. Clin Exp Gastroenterol. 2014;7:395-413.

106.   Chen PF, Wang F, Nie JY, Feng JR, Liu J, Zhou R, et al. Co-expression network analysis identified CDH11 in association with progression and prognosis in gastric cancer. Onco Targets Ther. 2018;11:6425-36.

107.   Gu D, Du M, Tang C, Chu H, Xu Z, Huo X, et al. Functional polymorphisms in apoptosis pathway genes and survival in patients with gastric cancer. Environmental and Molecular Mutagenesis. 2014;55(5):421-7.

108.   Heitzer E, Auinger L, Speicher MR. Cell-Free DNA and Apoptosis: How Dead Cells Inform About the Living. Trends in Molecular Medicine. 2020;26(5):519-28.

109.   Harbour JW, Roberson EDO, Anbunathan H, Onken MD, Worley LA, Bowcock AM. Recurrent mutations at codon 625 of the splicing factor SF3B1 in uveal melanoma. Nature Genetics. 2013;45(2):133-5.

110.   Kong Y, Krauthammer M, Halaban R. Rare SF3B1 R625 mutations in cutaneous melanoma. Melanoma Res. 2014;24(4):332-4.

111.   Hintzsche JD, Gorden NT, Amato CM, Kim J, Wuensch KE, Robinson SE, et al. Whole-exome sequencing identifies recurrent SF3B1 R625 mutation and comutation of NF1 and KIT in mucosal melanoma. Melanoma Res. 2017;27(3):189-99.

จุฬาลงกรณ์มหาวิทยาลัย
CHULALONGKORN UNIVERSITY

VITA

NAME   Julanee Leenanitikul

DATE OF BIRTH   05 Dec 1997

PLACE OF BIRTH   Phang-nga,Thailand

INSTITUTIONS ATTENDED   6th July 2020

HOME ADDRESS   1/4 Village No.4 Tesaban9 Lane Petchrkasem Road Thai-
                          mueang Sub-district, Thai-mueang District, Phang-nga,
                          82120