Designing Circulating-tumor cells (CTCs) detection method using co-measurement data of Single-Cell RNA and Targeted DNA Sequencing

By-

Omkar Chandra R

PHD17206

Under the supervision of

Dr. Vibhor Kumar and Dr. Say Li Kong

Indraprastha Institute of Information Technology Delhi

Dec 2022

Designing Circulating-tumor cells (CTCs) detection method using co-measurement data of Single-Cell RNA and Targeted DNA Sequencing

By-

Omkar Chandra R

PHD17206

in partial fulfillment of the requirements for the degree of Master of Technology

To

Indraprastha Institute of Information Technology Delhi

Dec 2022

Certificate

This is to certify that the thesis entitled "Designing Circulating-tumor cells (CTCs) detection method using co-measurement data of Single-Cell RNA and Targeted DNA Sequencing" being submitted by Omkar Chandra R to the Indraprastha Institute of Information Technology, Delhi, for the award of the Master of Technology, is an original research work carried out by him under my supervision. In my opinion, the thesis has reached the standards fulfilling the requirements of the regulations relating to the degree.

The results contained in this thesis have not been submitted in part or full to any other university or institute for the award of any degree/diploma.

Dr. Vibhor Kumar Dec 2022

Associate Professor

Department of Computational Biology

Indraprastha Institute of Information Technology Delhi, New Delhi 110 020

Acknowledgment

I am grateful to my advisor Dr. Vibhor Kumar, Associate Professor in the Department of Computational Biology IIITD, who has always allowed me to learn new scientific concepts spanning multiple domains and had interesting conversations over the years. Without whose invaluable advice and support, the present work would not have been possible. He taught me about perseverance and to become independent in my research work, and I thank him for making me part of his lab. I am also grateful to my co-advisor Dr. Say Li Kong, Genome Institute of Singapore, who has always helped to understand all the difficult concepts in experimental biology and for considering me to work on this project started by her, who also generated all the data used in this project.

For a student like me who had a wet lab background, who has learnt a lot of skills and knowledge in computational biology in this institute. I will always be in debt to Indraprastha Institute of Technology, Delhi (IIITD). IIITD has provided an excellent environment for me to keep learning. The infrastructure and technical capabilities, is unlike any other institute I have worked before, I will be always indebted to the institute and I hope I will be able to contribute back to the institute in some capacity so as to help future students like me who would join here.

I would like to extend my acknowledgement to all my friends and colleagues of my lab as well as the A-131 room of the R & D building, Computational Biology Department at IIITD. My thesis work results from collaboration and teamwork involving healthy discussions and regular feedback. We learned from each other and improved. I would also like to thank the administrative heads for MTech and PhD of IIITD for ensuring I would finish all the bureaucratic work on time and IT help desk IIITD for assisting in computing facilities.

I am grateful for my late parents. They constantly made efforts in nurturing me to be the human I am today. My mother and father, Mahalakshmi and Ramachandra have strongly ingrained the values of honesty, integrity, love and respect. I would not have been able to pursue and complete the degree at IIITD without their nurturing.

ABSTRACT

Abstract: Detection of cancer for diagnosis and tracking the prognosis of the patients using liquid biopsy procedure: the extraction of blood for the detection of cells is the least invasive procedure and relatively fast process compared to tissue biopsy, which would make the isolation and analysis of biological samples like exosomes, circulating tumor DNA (ctDNA), circulating tumor cells (CTCs), a preferred approach over solid biopsy via open surgery. Due to the sparsity of the overall count of CTCs in the blood and low genomic content recovery from them, it is hard to identify and characterize CTCs to make insights into the biology of the disease.

In this study, we analyzed the concurrent data of single-cell RNA-seq and targeted DNA-seq from 9 pancreatic cancer patients. We developed a pipeline that leverages the nature of concurrent data to identify CTCs from non-CTCs (blood cells).
We were able to identify CTCs using the mutation profile and transcriptomic profiles of CTCs. We made fundamental biological insights into pancreatic cancer disease by integrating mutational and transcriptomic profiles of the cells.

CONTENTS

- Acknowledgments ----------3
- Abstract ----------5
- List of Figures ----------7
- Result ----------13
    - Detection of CTCs using mutation profiles —----------14
    - Detection of CTCs using transcriptomic profiles —----------17
    - Integration of mutation and transcriptomic profiles----------21
    - Application of the pipeline on patient data (G53 and G57)-----24
- Conclusion ----------30
- Method ---------- 31
- Bibliography ----------32

LIST OF FIGURES

Figure 1	Overview	13
Figure 2	**Data quality information of DNA-seq of different libraries**	14
Figure 3	**Allele frequency**	15
Figure 4	**Median count comparison**	16
Figure 5	**Quality of RNA-seq profiles**	17
Figure 6	**tSNE embeddings**	19
Figure 7	**median allele count comparison between the newly predicted CTCs and non-CTCs in patient G50**	20
Figure 8	**Differential pathway enrichment between all the CTCs selected from both the approach and non-CTCs**	21
Figure 9	**Represents the correlation between minor allele per gene count and the pathway enrichment scores of all the selected CTCs in patient G50**	22
Figure 10	**Represents the correlation between minor allele per gene count and the pathway enrichment scores of all the non-CTCs in patient G50**	23
Figure 11	**Represents the correlation between minor allele per gene count and the pathway enrichment scores of all the non-CTCs in patient G50.**	24
Figure 12	**Differential pathway enrichment between all the CTCs selected from both the approach and non-CTCs in patient G53 (32 cells in total).**	25

Figure 13	**Represents the correlation between minor allele per gene count and the pathway enrichment scores of all the selected CTCs in patient G53 (32 cells in total)**	26
Figure 14	**First graph indicates the median minor allele between selected CTCs using mutational profile and non-CTCs in patient G57 and second graph shows median of minor allele between newly predicted CTCs using transcriptomic profile and non-CTCs in patient G57 (21 cells in total)**	27
Figure 15	**Differential pathway enrichment between all the CTCs selected from both the approach and non-CTCs in patient G57 (21 cells in total).**	28
Figure 16	**Represents the correlation between minor allele per gene count and the pathway enrichment scores of all the selected CTCs in patient G57 (21 cells in total).**	29
Figure 17	**Minor allele per gene across all patients' final selected CTCs**	30

Designing Circulating-tumor cells (CTCs) detection method using co-measurement data of Single-Cell RNA and Targeted DNA Sequencing

Introduction

The traditional approaches for chemotherapy regimen design is guided by analyzing the solid tumors which are obtained through open surgeries or invasive tumor biopsy sampling. Typically, histology and certain protein signatures are used to assess the prognosis and to develop multi-drug targeted therapy regimens [1,2]. However the cancer-free or survival outcomes of the patients improved only marginally because of the tumorigenicity, invasiveness and evasiveness of the tumor cells conferred by the changes at genomic and epigenomic level. Thus, keeping the cells resistant to the drugs. [3,4].
These approaches of analyzing just the proteomic and histological biomarkers are being slowly replaced by the modern approach of analyzing the solid tumors for signatures at single-cell resolution to dissect the heterogeneity of the tumor samples at multi-omics level (genomic, epigenomic, transcriptomic, proteomic etc) using the next-generation sequencing (NGS) data [5]. NGS of single-cell data of tumor cells offers the opportunity to understand the changes at transcriptomic and genomic level in detail because of the techniques' capabilities to detect low-frequency variants and variants at different biological context like alternatively spliced transcripts, gene fusion, single-nucleotide polymorphisms, small and long non-coding RNAs, post-transcriptional modifications and minor changes in gene expression can also be detected (given a significant number of samples) [6]. Moreover, single-cell sequencing data offers the ability to dissect the heterogeneity of the tumor tissue at single-cell resolution to identify molecular and cellular mechanisms to develop targeted therapies [7–9].

Accessibility of solid tumors from the patients is one of the major challenges because of the arising clinical complications associated with the invasive procedures necessary to obtain tissue at the time of initial diagnosis as well as over the course of disease treatment. Also, the amount of sample extraction is completely subjected to the pathophysiological conditions of the patients [10]. One of the biological challenges in the analysis of the solid tumors is the intra-tumor and inter-tumor heterogeneity and dynamic change in the genomic properties of the tumor tissues over time, practically making the assessments on the initially obtained tumor futile [11]. Also, the cost of patient care and the turnaround time for getting results using tissue biopsies can be

longer while the patient care might be of time-sensitive issue. Thus, considering these limitations newer ways of biopsy procedures have evolved like liquid biopsy which is generally rapid, precise and to a degree allows for the real-time treatment of the patient.

A liquid biopsy method, the extraction of blood for the detection of cells is the least invasive procedure and relatively fast process compared to tissue biopsy, which would make the isolation and analysis of biological samples like exosomes, circulating tumor DNA (ctDNA), circulating-tumor cells (CTCs), a preferred approach over solid biopsy via open surgery [12]. Also, analysis of CTCs in the blood sample allows for tracking the prognosis of the disease in real-time with least strain on the patient. Many studies in the past have shown that CTCs can be used as a marker to predict disease progression and survival in metastatic as well in the early-stage cancer patients [14]. Analysis of CTCs can provide clinically relevant as well as fundamental information that is not revealed with the analysis of ctDNA [15].

CTCs were first detected in the blood of a metastatic cancer patient by Ashworth in 1869 [16]. Circulating-tumor cells (CTCs) are the cells shed out from the solid tumor tissues during hematogenous spread of the cancer, and can be detected in the blood circulation. CTCs serve as the seeds for the formation of secondary tumors around the body. CTCs translocate to distant tissues and organs via blood vessels, adapt to the new microenvironment, and eventually seed, proliferate, and colonize to form metastases. The genome of the CTCs can potentially reveal the nature of tumor tissue and its origin [17,18]. A meta-analysis conducted by Qingtao et al. using data from twenty-one studies with 3997 subjects concluded that the CTCs had good diagnostic value for detecting lung cancer [19]. There have been many studies demonstrating the role of CTCs in designing the therapy regimens in breast cancer, pancreatic cancer, hepatic cancer, and oral cancer. [20–23].

Analysis of the CTCs not only provides insights into the clinical diagnosis and prognosis, it can also give fundamental insights into the basic biology of the cancer cells. Cancer metastasis occurs in an elaborate, sequential manner starting with detachment of cells from the primary tumor, followed by intravasation of the CTCs into bloodstream, further translocation to different organs through the circulatory system and extravasation into secondary sites, finally adoption to the new microenvironment and multiplication. [16,24]. Eslami-S et al. describe two types of CTCs in their review article based on their metastatic capacity: (1) those CTCs that can migrate and reach into the circulatory system but are not metastasis-competent [25] (2) CTCs that are metastasis-competent. This categorization is based on the seminal work laid down by Fidler et al., and they also showed that less than 0.01% of cancer cells injected in in vivo models could generate metastatic tumors. The highly heterogeneous nature of the cancer confers a different mechanism of metastasis among individual cells of the same tumor [25,26]. All these dynamic heterogeneity among the CTCs are because of the collective changes happening at the genotypic, phenotypic as well as at cellular levels. Therefore, it is imperative to understand the nature of CTCs at different levels using multi-omics data at single-cell resolution to develop new drugs and to design better diagnostic and treatment methods against cancer. CTC isolation from the blood sample, and obtaining multi-omics data at the single-cell resolution from CTCs using

assays and meaningful integration of such data is challenging and is a biotechnology and data science problem, respectively.

CTC isolation is the most difficult first step because of the extreme rarity of CTCs in the blood samples compared to hematologic cells (about 1 tumor cell per 1 billion blood cells).
There are different strategies used to isolate the CTCs from the peripheral blood samples.
One of the immuno-affinity strategies is CTC-chip, a microfluidic platform with antibody (EpCAM)-coated microposts under precisely controlled laminar flow conditions [27]. One of the earlier more efficient method in CTC isolation was described by Shannon et al., they described throughput microfluidic mixing device, the herringbone-chip, or “HB-Chip" which increases the mixing of blood cells through the generation of microvortices to significantly increase the number of interactions between target CTCs and the antibody-coated chip surface, thereby increasing the chances of capturing the CTCs [28]. One of the drawbacks of these methods is its inability to isolate CTCs without epithelial antigens.

An in vivo indwelling intravascular aphaeretic CTC isolation system called ^{HB}GO chip is designed to continuously collect CTCs directly from a peripheral vein, it uses a graphene oxide sheets functionalized with high-density anti-EpCAM antibodies and herringbone structures for better surface contact between the CTCs and the antibody coated sheets [29]. This method can capture CTCs in more numbers compared to *ex vivo* methods.
Liu et al. described an immuno-affinity based negative selection method which has the advantage of isolating CTCs with insufficient expression of the surface target markers, which may be missed by positive selection methods. They employed three enrichment methods: CD_{45} depletion (negative selection), EpCAM-positive selection, and the combination of both negative and positive selection. The highest recovery was observed in the sole CD_{45} depletion method [30].

There are biophysical feature strategies, one of the ways is using the size dependent separation. Payne et al., demonstrated that CTC enrichment in head and neck squamous cell carcinoma using Parsortix microfluidic CTC enrichment platform, the cell capture rate can be increased with fixative blood collection tube[31]. Wu et al. describe CTC detection and isolation based on the surface charger of cancer cells, a bioelectrical manifestation of the “Warburg effect”. Negative surface charge is a direct consequence of increased uptake of glucose by the cancer cells to meet the demands of the rapid nuclear division in them, which is exploited to capture the CTCs using the poly(ethyleneimine)-functionalized Fe3O4 nanoparticles, surface-decorated with protein corona. This method is able to isolate diverse CTC subpopulations of heteroploids and biomarker expression. Being a label-free and charge-based CTC method, it is demonstrated to be one of the efficient ways of liquid biopsy[32]. Mishra et al., describe ^{LP}CTC-iChip, an ultrahigh-throughput microfluidic chip that rapidly sorts through an entire leukapheresis product of over 6 billion nucleated cells, increasing CTC isolation capacity by two orders of magnitude, further application of magnetic field on the magnetically labeled leukocytes helps in negative depletion of leukocytes, increasing the potentially viable CTCs without bias for expression of specific tumor epitomes [33].

The next step after CTC isolation is the efficient generation of concurrent multi-omics data (like genomics, transcriptomics) from the single cells. There has been a number of studies by different groups to obtain the concurrent data from the isolated CTCs. Han et al. describe a workflow for concurrent evaluation of transcriptomic and genomic characteristics within the same single-cell on microfluidic platform, although it is a low throughput method, it has been one of the earliest work on concurrent data generation [34]. To address the issue of low throughput and slowness, Say Li et al. have developed concurrent sequencing of the transcriptome and targeted genomic regions (CORTAD-seq) within the same single cell on an automated microfluidic platform [35]. They successfully characterized lung cancer cell lines to dissect the cellular consequences of mutations that result in resistance to targeted therapy.

The final step is the biologically meaningful integrative analysis of the generated data from the CTCs. This project focuses on this step, we have developed a novel pipeline of identifying the CTCs using the concurrent genomic and transcriptomic profiles of isolated pancreatic CTCs generated in-house. The advantage of using the concurrent data is it allows for the characterization of the CTCs at two levels, genomic and transcriptomic. Therefore, even if the signature genes are not identified at the transcriptomic level, mutational signature at the genomic level would help in identifying the CTCs. Further, combination of both the profiles of CTCs can give novel insights into the biology of the pancreatic cancer itself. Pancreatic adenocarcinoma (PDAC) is one of the most malignant abdominal tumors. It is the seventh leading cause of tumor-related death worldwide. It correlates with age and is more prevalent in men than in women [36]. PDAC has poor prognosis, only 24% of people survive for 1 year and 9% live for 5 years [37]. Important aspect of pancreatic cancer is that the diagnostic tests are non-specific and may miss patients with early-stage disease, also PDAC and the other less common exocrine cancers are typically diagnosed at a late stage [38]. Because of these reasons it is imperative to develop proper diagnostic methods which can be employed in detection of the cancer to decrease the mortality rates.

In this study, we utilized the mutational signatures across 10 genes related to pancreatic cancer to estimate major and minor alleles based on the number of cells harboring a particular type of single-nucleotide polymorphism (SNP) across these 10 genes: \. Further, a cell is considered to be CTC if it harbors more number of minor alleles, in consideration of the fact that CTCs tend to accumulate few newer mutations over time in their genome, it tends to have more somatic mutations (minor allele) [39].

Result

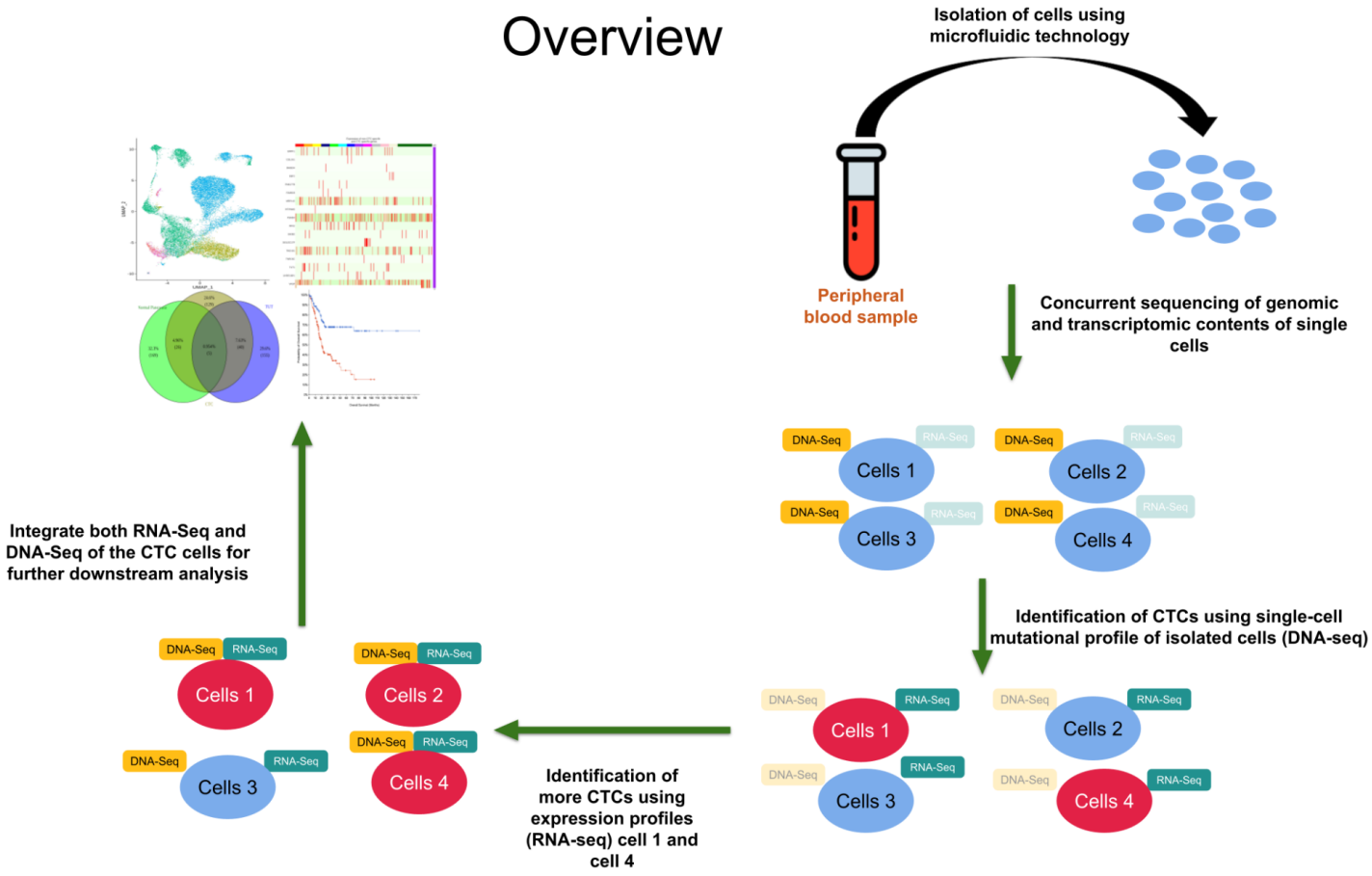

Figure 1. This represents the overall workflow of the project. The cells were isolated by size using microfluidic technology, subjected to concurrent sequencing of both DNA-seq and RNA-seq. In the first approach to identify CTCs, their respective mutation profiles are used. In the second approach, the identified CTCs' transcriptomic profiles are used to further detect CTCs which were previously not detected.

Detection of CTCs using mutational profiles

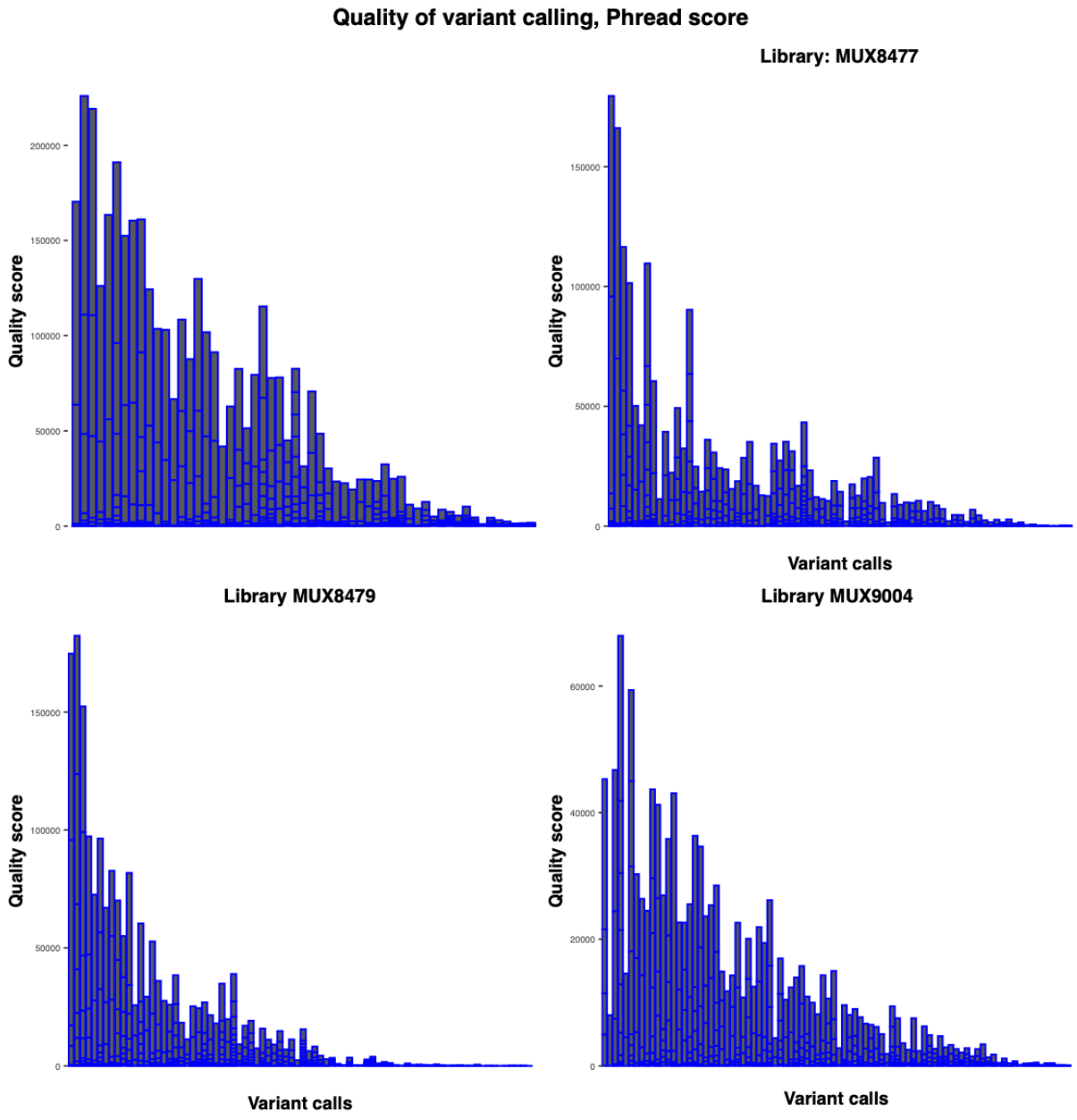

Figure 2. Data quality information of DNA-seq of different libraries.

To detect the CTCs using the mutational profile, three steps were followed:

Step 1. Characterization of alleles as major and minor alleles

Based on the allele frequency it is possible to classify the alleles as major and minor alleles. Allele frequency for a patient is the ratio of the number of times an allele occurs in all the cells to the total number of alleles. An allele is considered to be a minor allele if its ratio is less

than 0.2 and an allele is considered to be a major allele if its ratio is considered to be more than 0.8. A major allele which is present in the majority of the cells represent germline mutations and a minor allele which is present in the minority of the cells represent somatic mutations. See Figure 3. to check the allele frequencies and

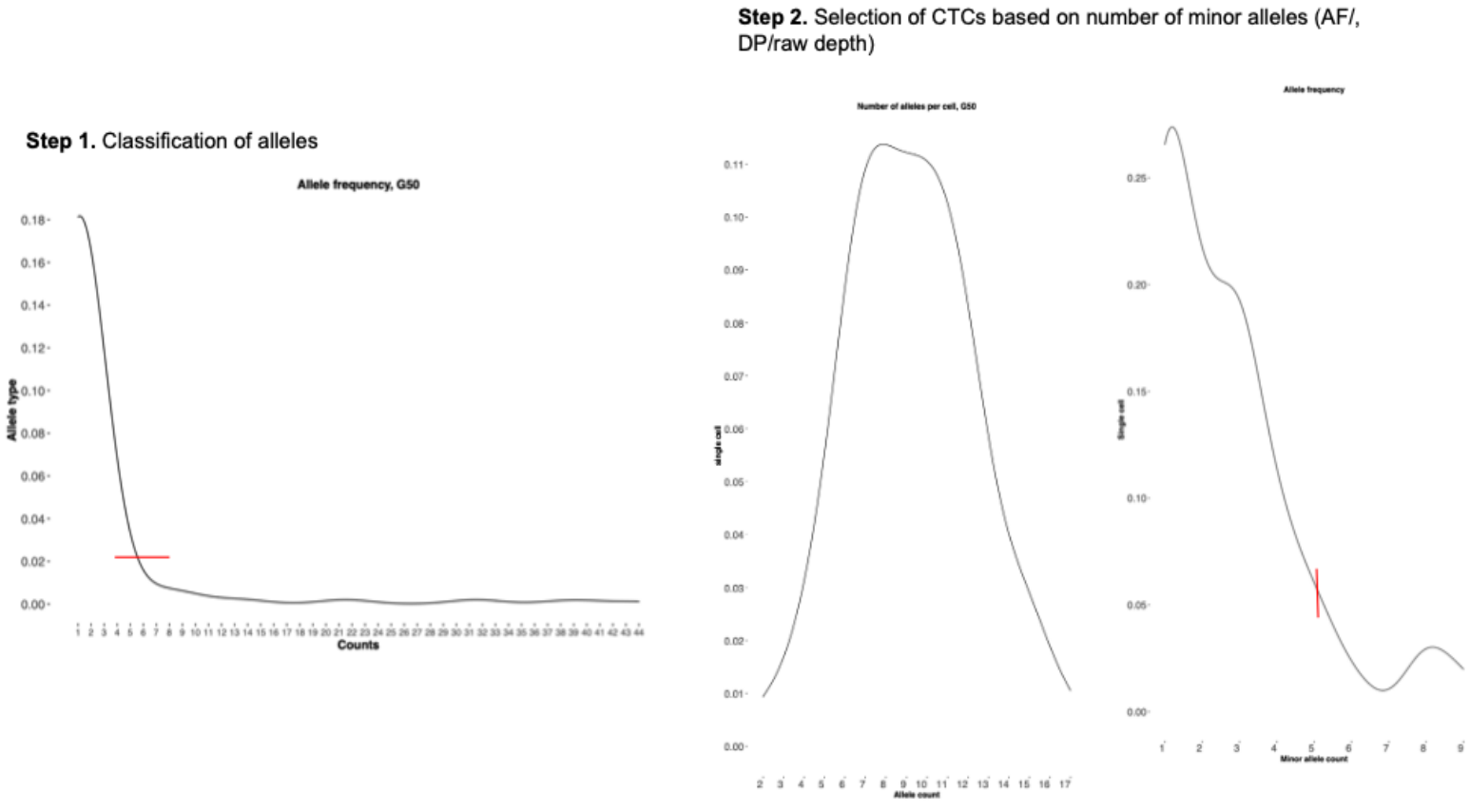

Figure 3. First graph depicts the allele frequency, the y-axis represents the allele type and x-axis represents the count. Second graph shows the distribution of all the allele frequency and distribution of minor alleles in patient G50.

Step 2. Characterization of alleles as major and minor alleles

Based on the fact that CTCs progressively keep acquiring new alleles, they are expected to have more minor alleles compared to the blood cells. There the cells which had a higher

number of minor alleles are considered as CTCs. Check Figure 2 for the minor allele distribution.

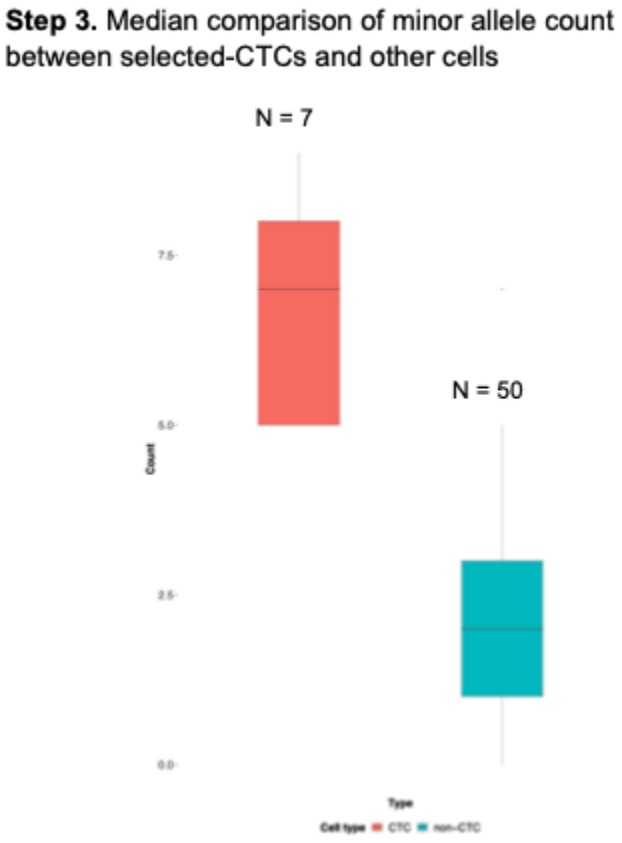

Figure 4. First graph depicts the median count comparison: the y-axis represents the count and x-axis represents the type of cell.

Step 3. Characterization of alleles as major and minor alleles

It is expected that the CTCs would harbor more number of minor alleles than the non-CTCs and this observation is seen in the patient G50 as presented in the Figure 4.

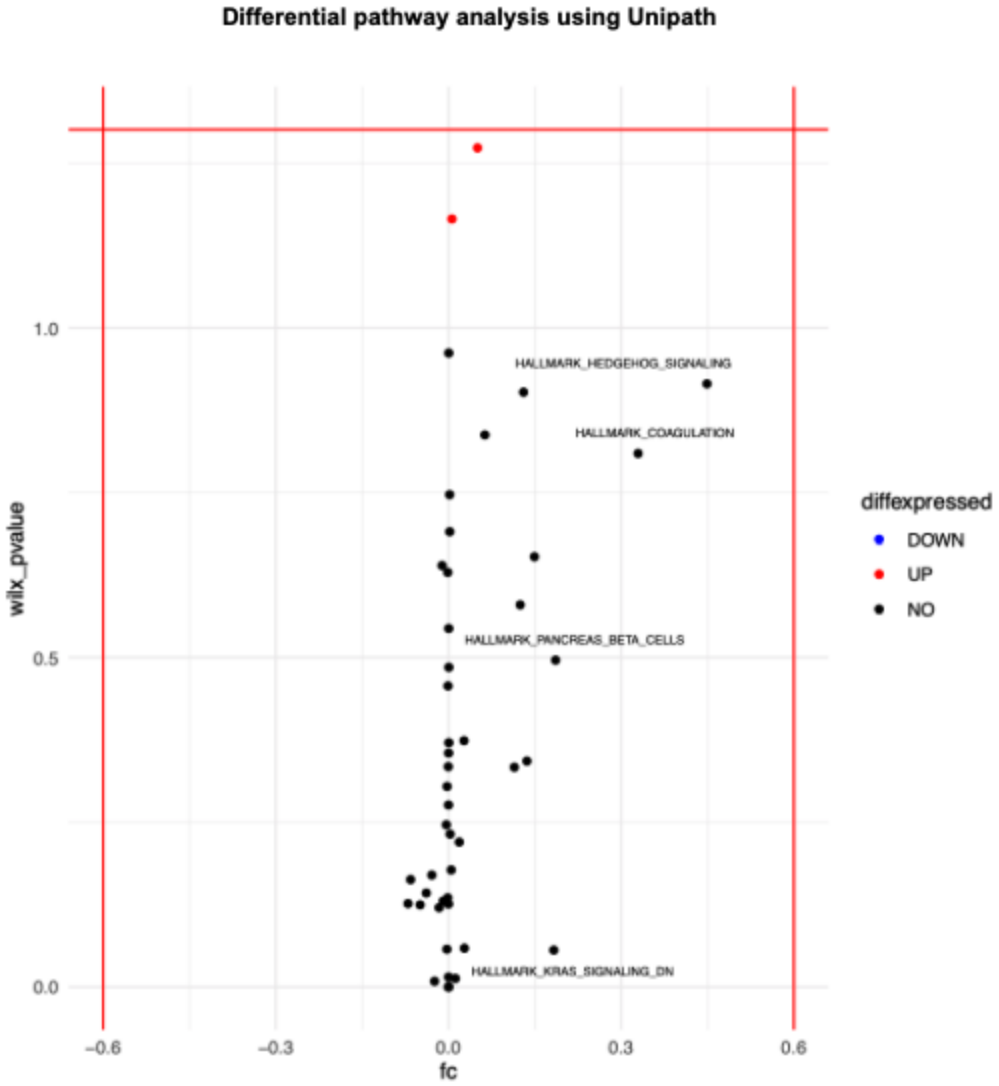

Figure 4.1. First graph depicts the allele frequency, the y-axis represents the count and x-axis represents the type of cell.

To verify if the identified cells are CTCs, differential pathway analysis is done using the scores generated by the Unipath tool. Pancreatic cancer specific pathways are observed in the CTCs with more fold change compared to non-CTCs, see Figure 4.1.

Detection of CTCs using transcription profiles

The identified CTCs' transcriptomic profiles are subjected to PU learning to further detect more CTCs which couldn't have been detected in the first approach using mutation profiles.
Check Figure 5 and Figure 6 for RNA-seq quality.

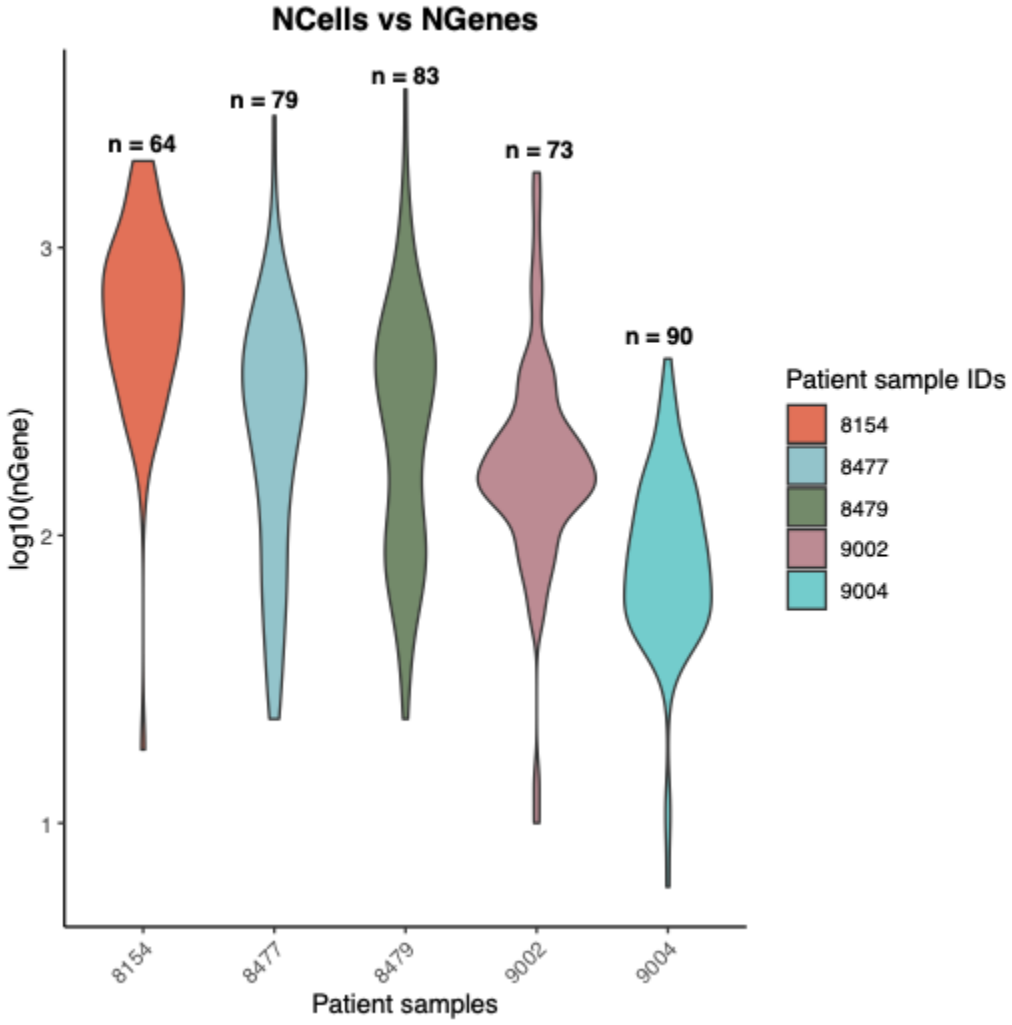

Figure 5. Represents the library quality of RNA-seq profiles of multiple libraries.

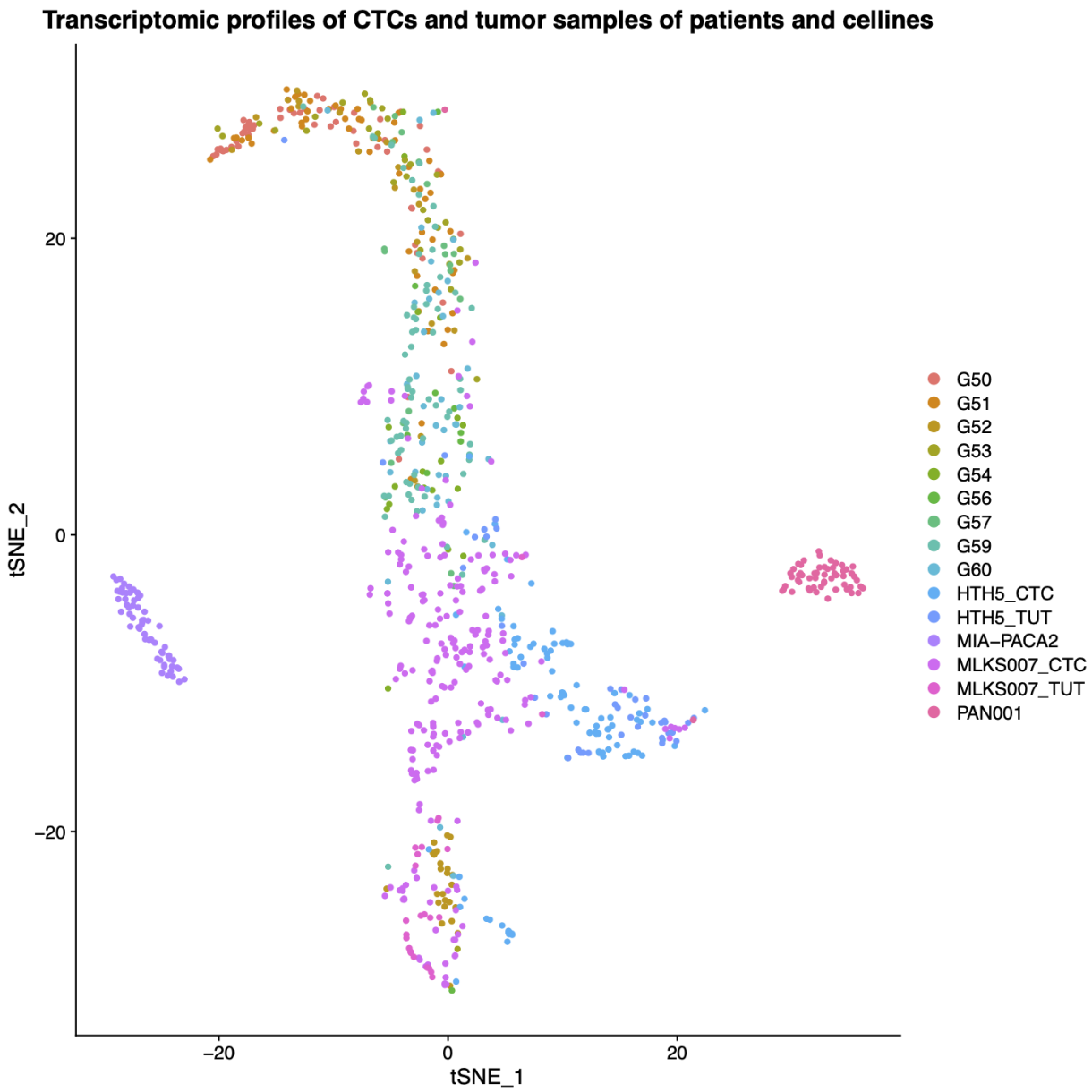

Figure 6. Represents the tSNE embeddings of all the cells of the different patients using RNA-seq profiles.

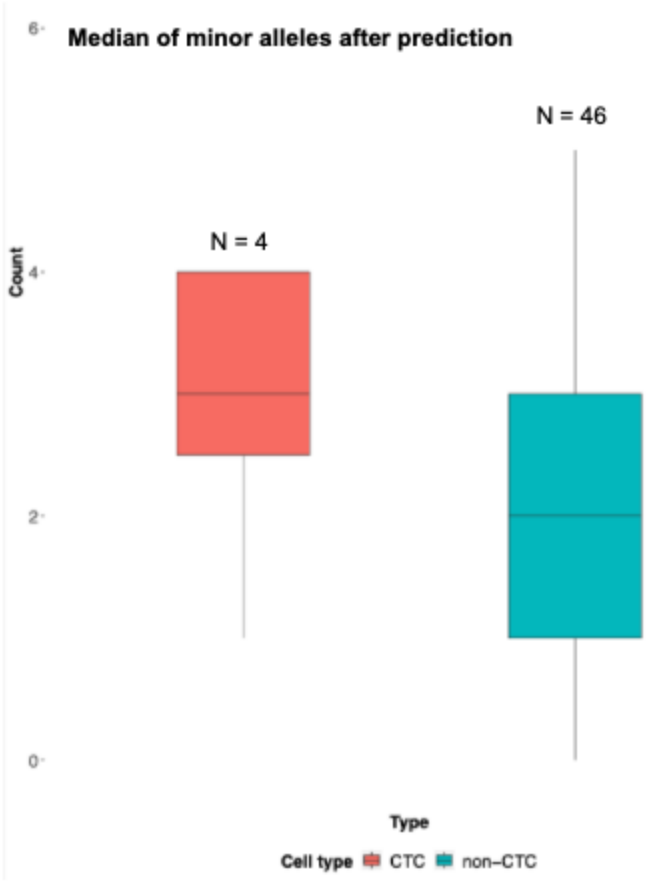

Figure 7. This graph represents the median allele count comparison between the newly predicted CTCs and non-CTCs in patient G50.

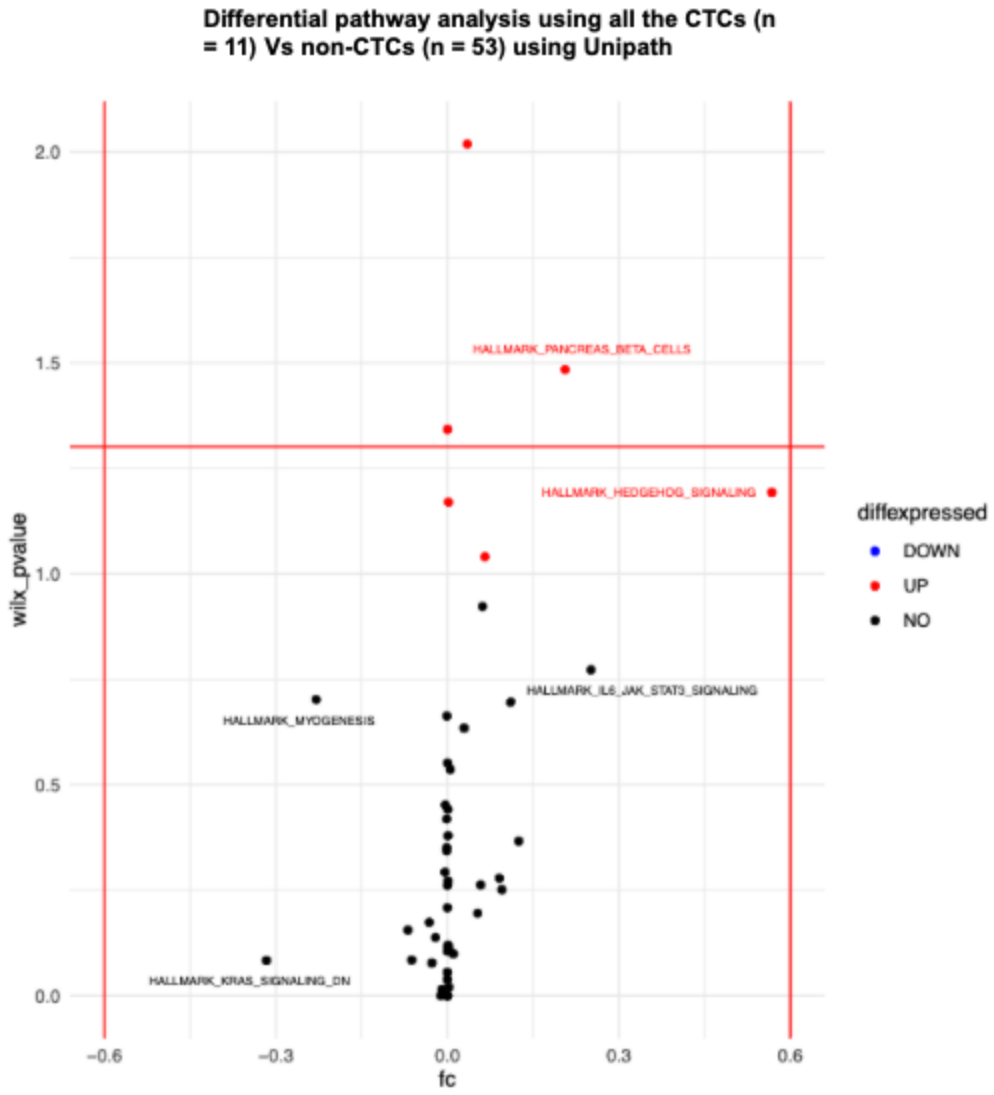

Figure 8. Differential pathway enrichment between all the CTCs selected from both the approach and non-CTCs.

Integration of mutation and transcriptomic profiles

After identifying the CTCs by these two methods, the mutation profiles and the transcriptomic profiles of the CTCs are integrated together to check for the correlation between the number of minor allele counts per gene and the pathway enriched.

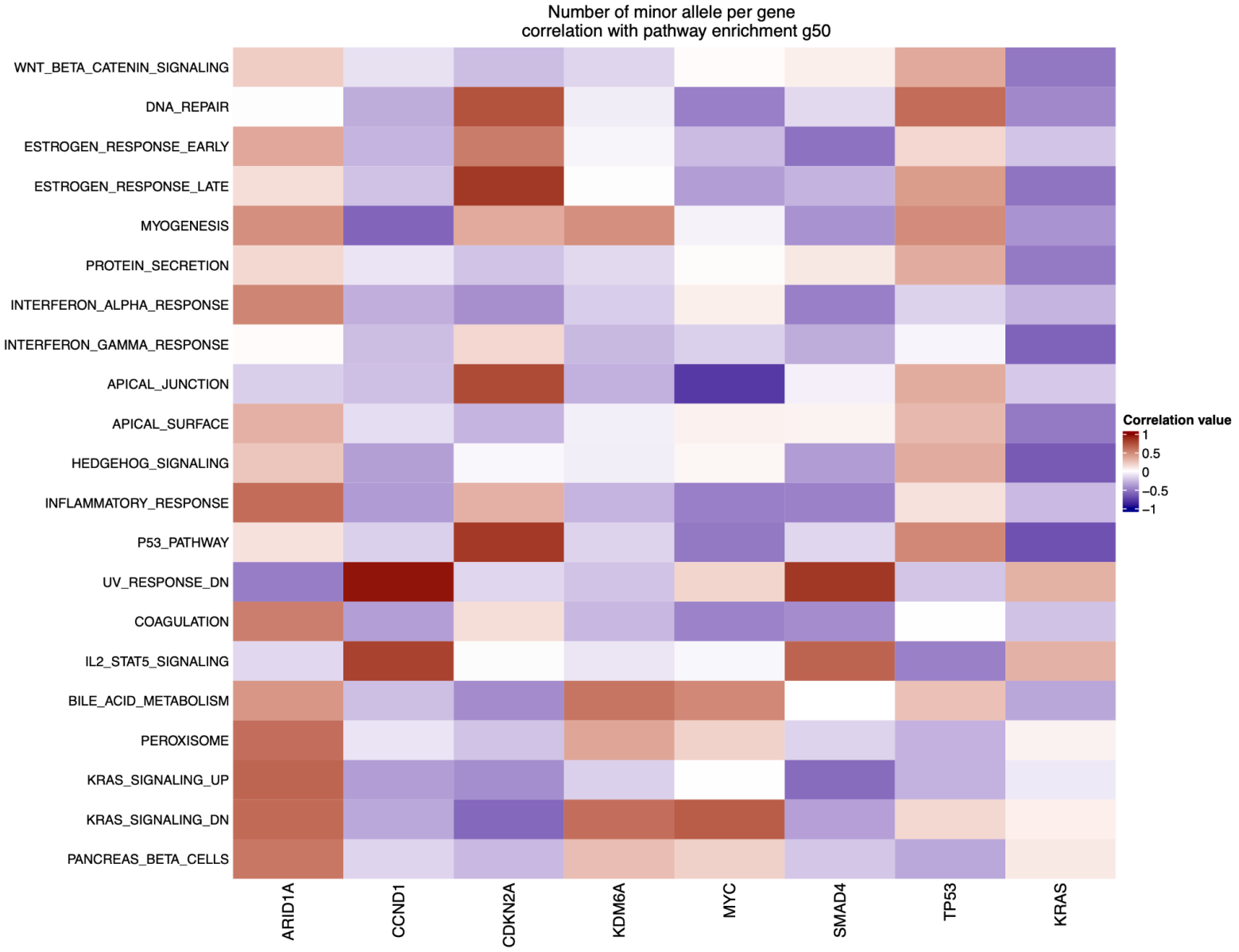

Figure 9. Represents the correlation between minor allele per gene count and the pathway enrichment scores of all the selected CTCs in patient G50.

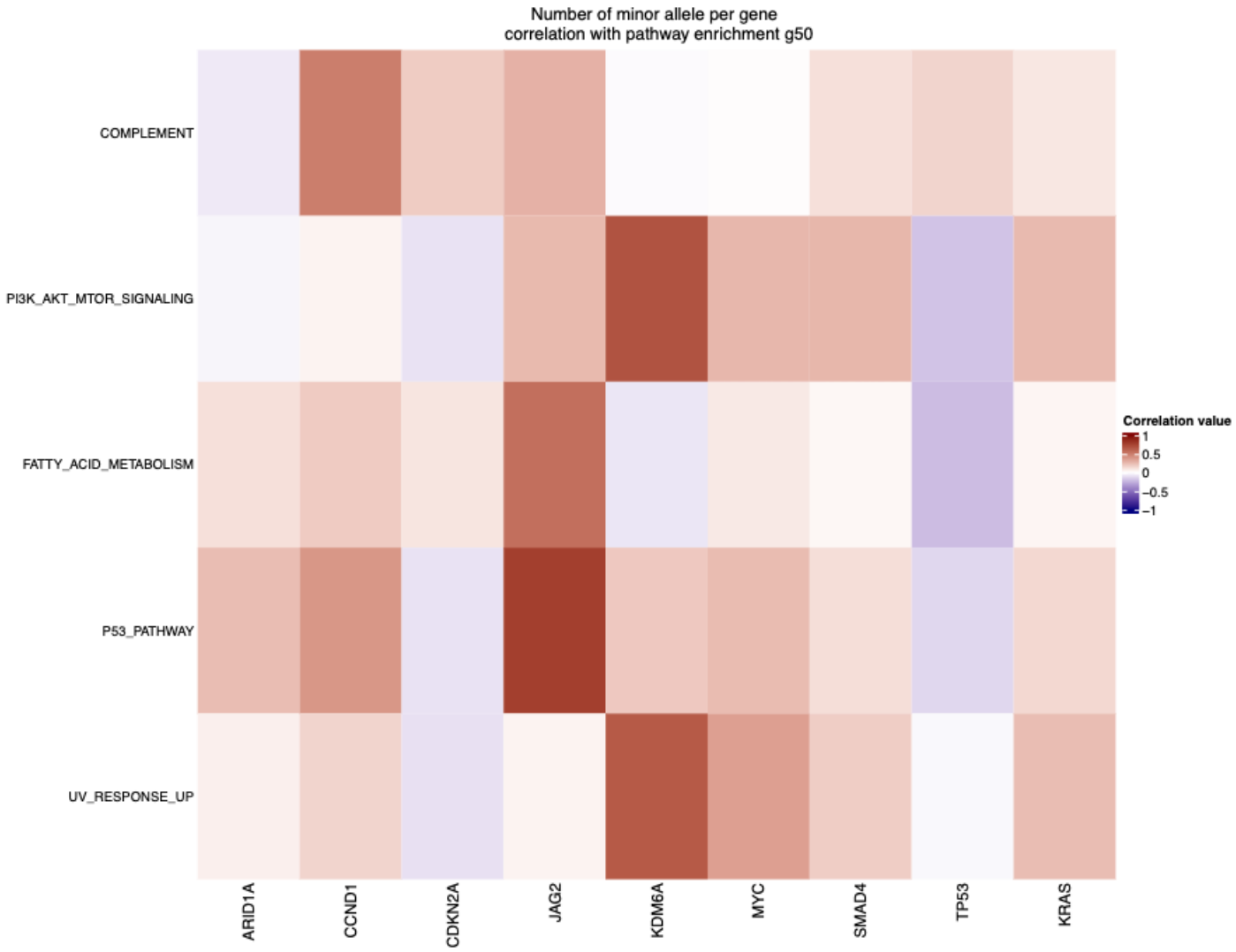

Figure 10. Represents the correlation between minor allele per gene count and the pathway enrichment scores of all the non-CTCs in patient G50.

The above methods of identification of CTCs using mutational profiles and transcriptional profiles were carried out on all the patient data (see Figure 11) and the next section shows results of two of the patients.

Patient ID	No. of cells	Type	Library ID (RNA-seq/DNA-seq)	Platform
MIA-PACA2 cell line*	60	Cell line	MUX7643/MUX7651	C1-96
G50/KCL 09112018	57	CTC	MUX8153/MUX8154	C1-96
G51/PIA 13122018	55	CTC	MUX8476/MUX8477	C1-96
G52/ASY 13122018	51	CTC	MUX8478/MUX8479	C1-96
PAN001 17122018	72	Tumour-derived line	MUX8406/MUX8407	C1-96
G53/Y-C 18122018	32	CTC	MUX8478/MUX8479	C1-96
G54/CJG 20122018	24	CTC	MUX8476/MUX8477	C1-96
G56/S-S 25012019	10	CTC	MUX9001/ MUX9002	C1-96
G57/T-K 04022019	21	CTC	MUX9001/ MUX9002	C1-96
G60/BKY 18032019	42	CTC	MUX9001/ MUX9002	C1-96
G59/LCB 13032019	90	CTC	MUX9003/ MUX9004	C1-96

Figure 11. Represents the correlation between minor allele per gene count and the pathway enrichment scores of all the non-CTCs in patient G50.

Application of the pipeline on patient data (G53 and G57)

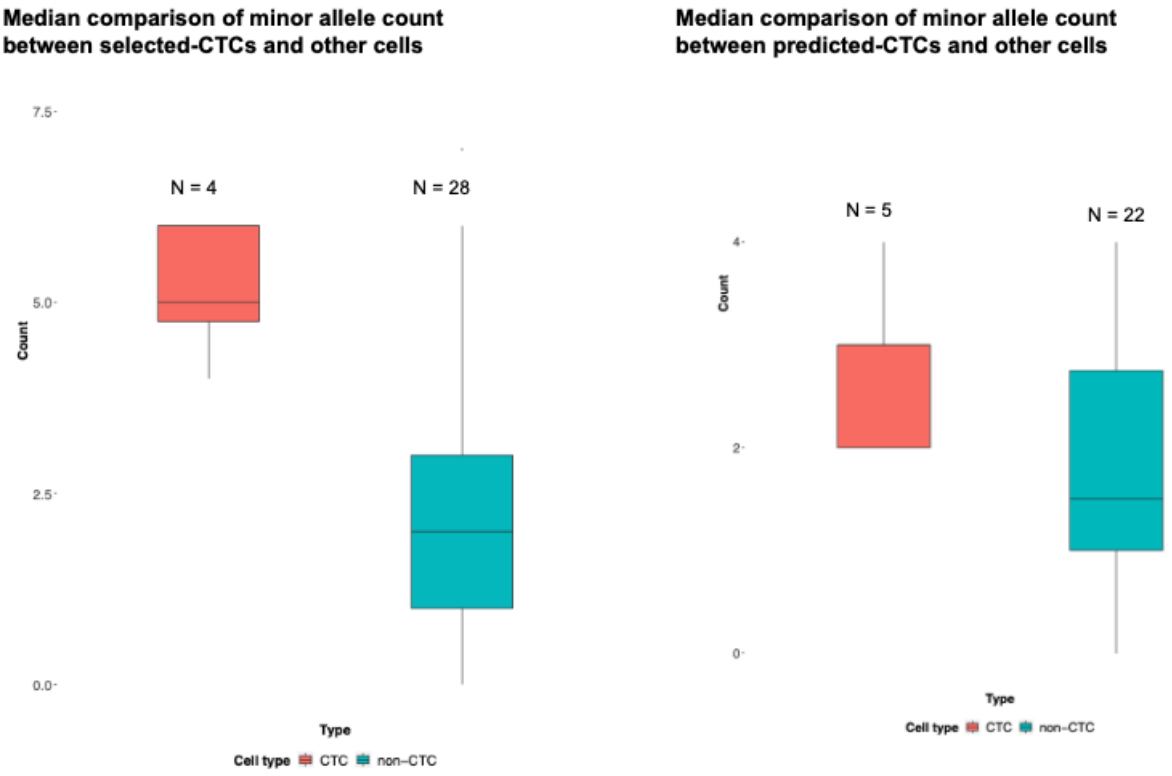

Figure 11. First graph indicates the median minor allele between selected CTCs using mutational profile and non-CTCs in patient G53 and second graph shows median of minor allele between newly predicted CTCs using transcriptomic profile and non-CTCs in patient G53 (32 cells in total).

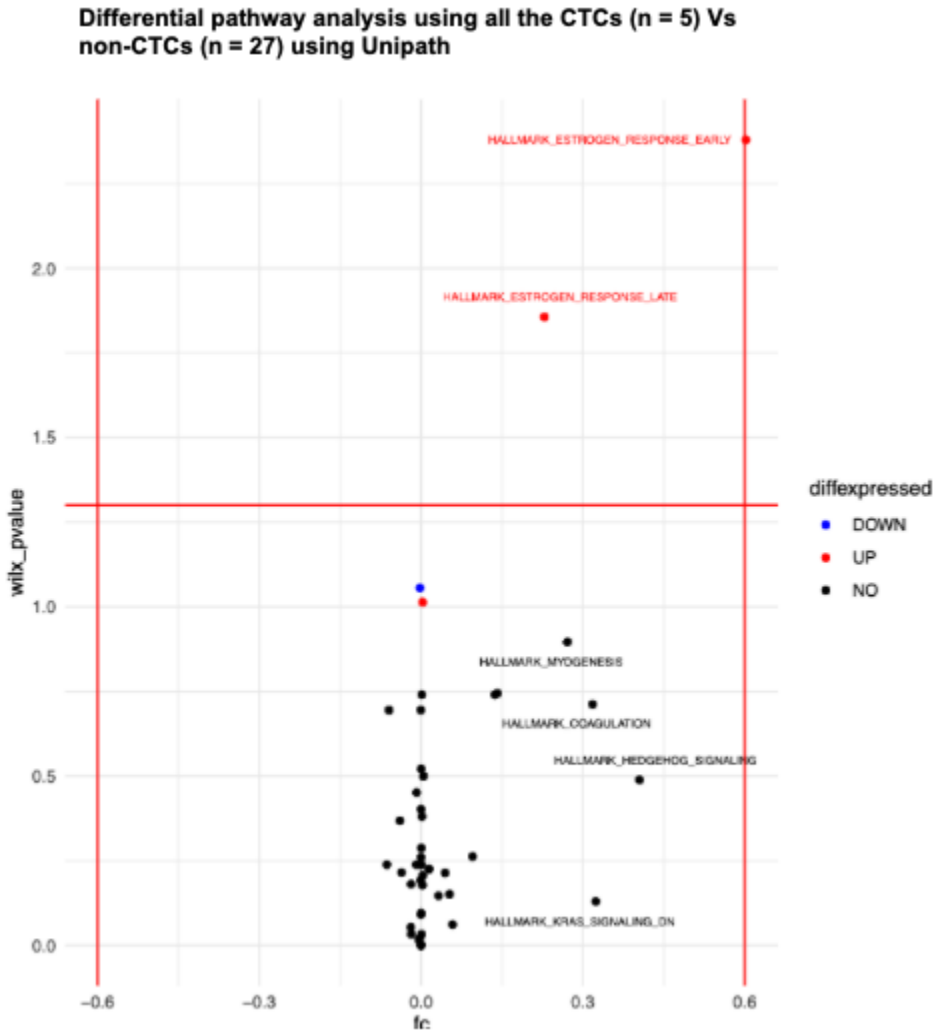

Figure 12. Differential pathway enrichment between all the CTCs selected from both the approach and non-CTCs in patient G53 (32 cells in total).

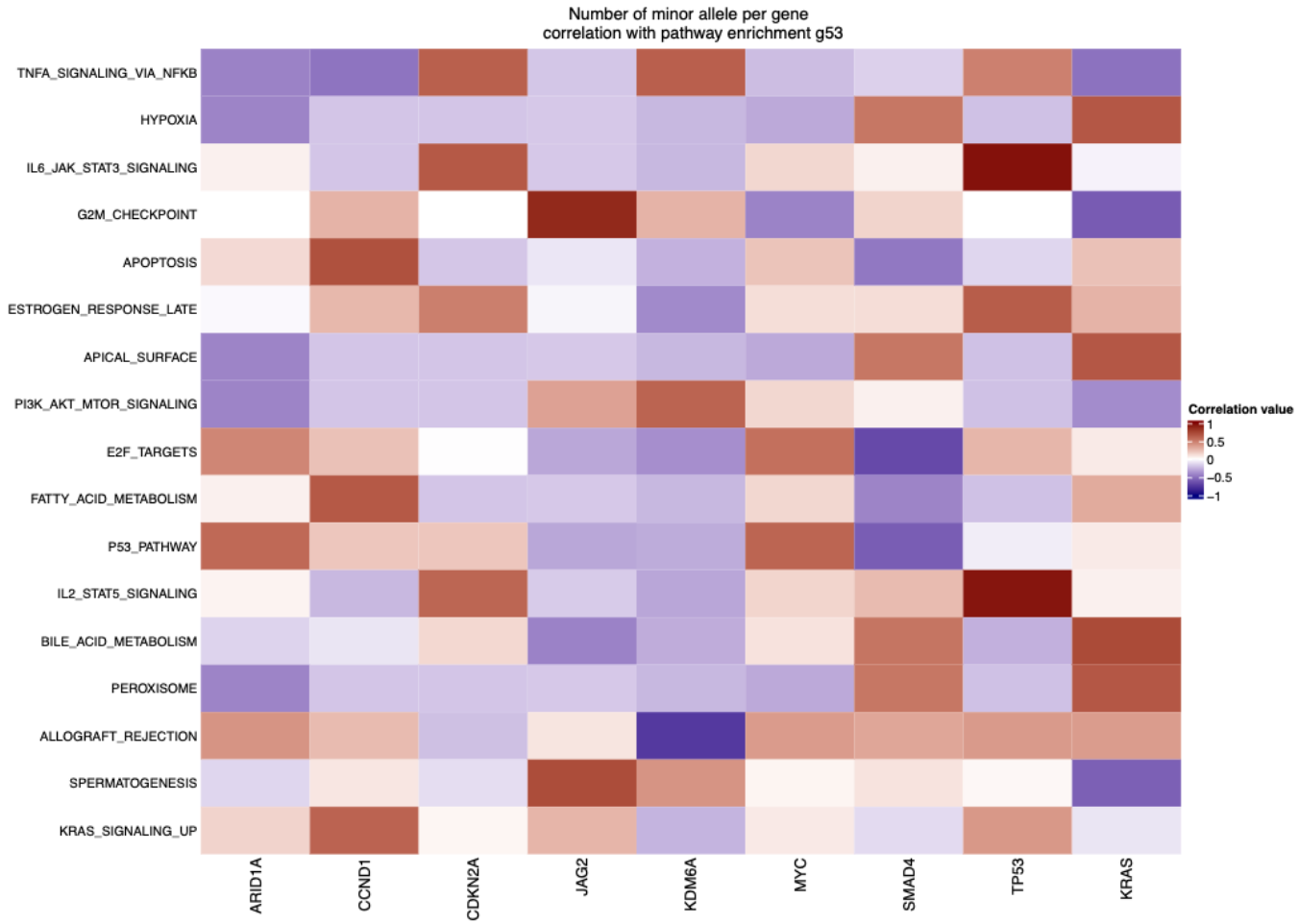

Figure 13. Represents the correlation between minor allele per gene count and the pathway enrichment scores of all the selected CTCs in patient G53 (32 cells in total).

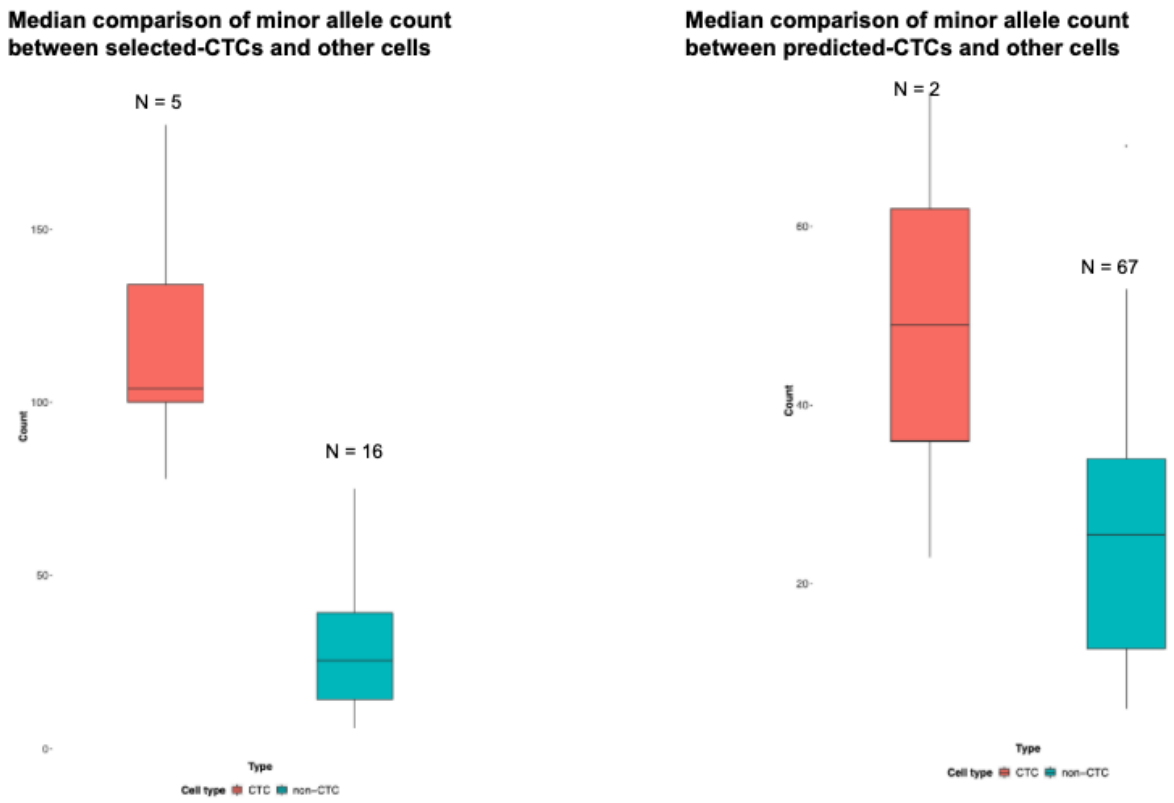

Figure 14. First graph indicates the median minor allele between selected CTCs using mutational profile and non-CTCs in patient G57 and second graph shows median of minor allele between newly predicted CTCs using transcriptomic profile and non-CTCs in patient G57 (21 cells in total).

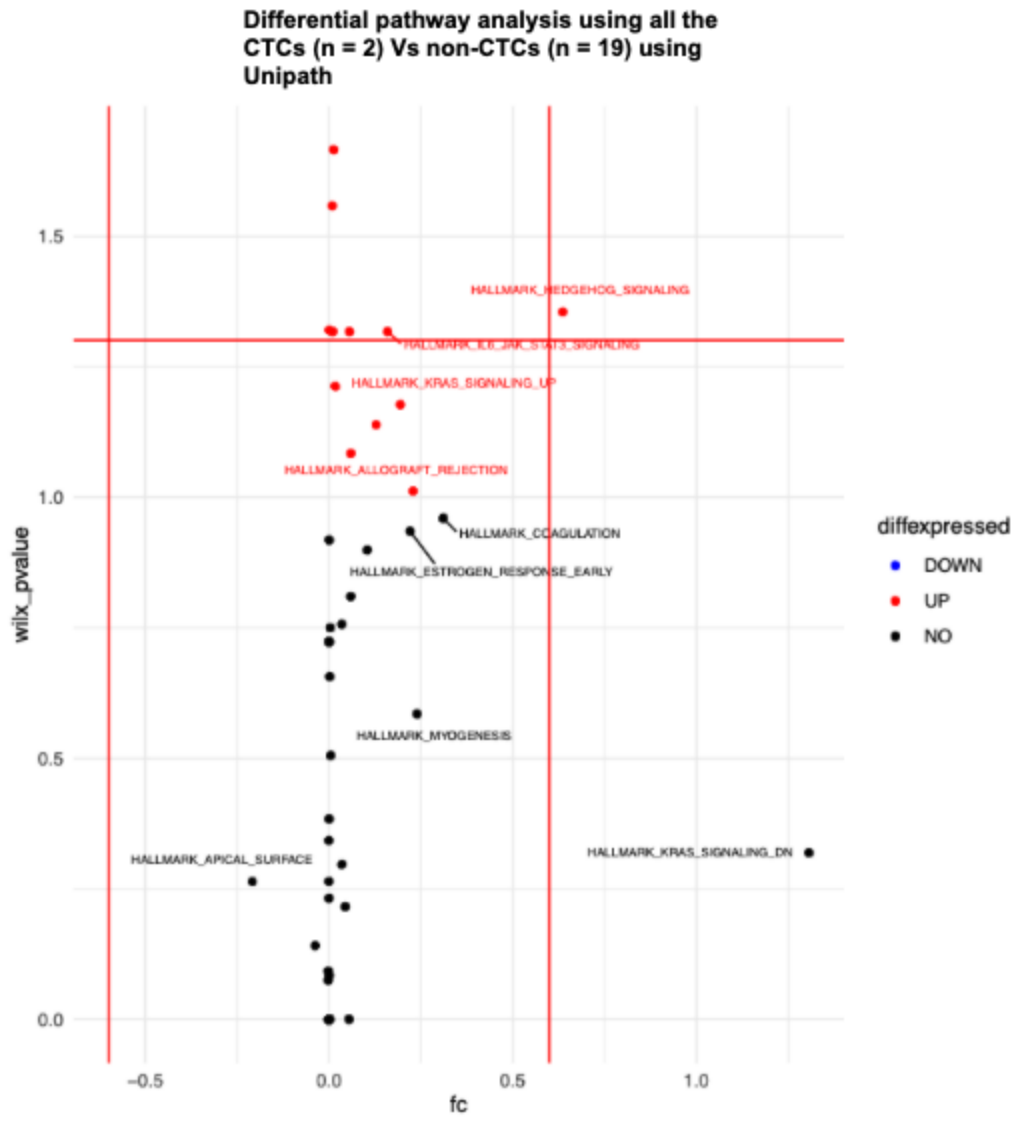

Figure 15. Differential pathway enrichment between all the CTCs selected from both the approach and non-CTCs in patient G57 (21 cells in total).

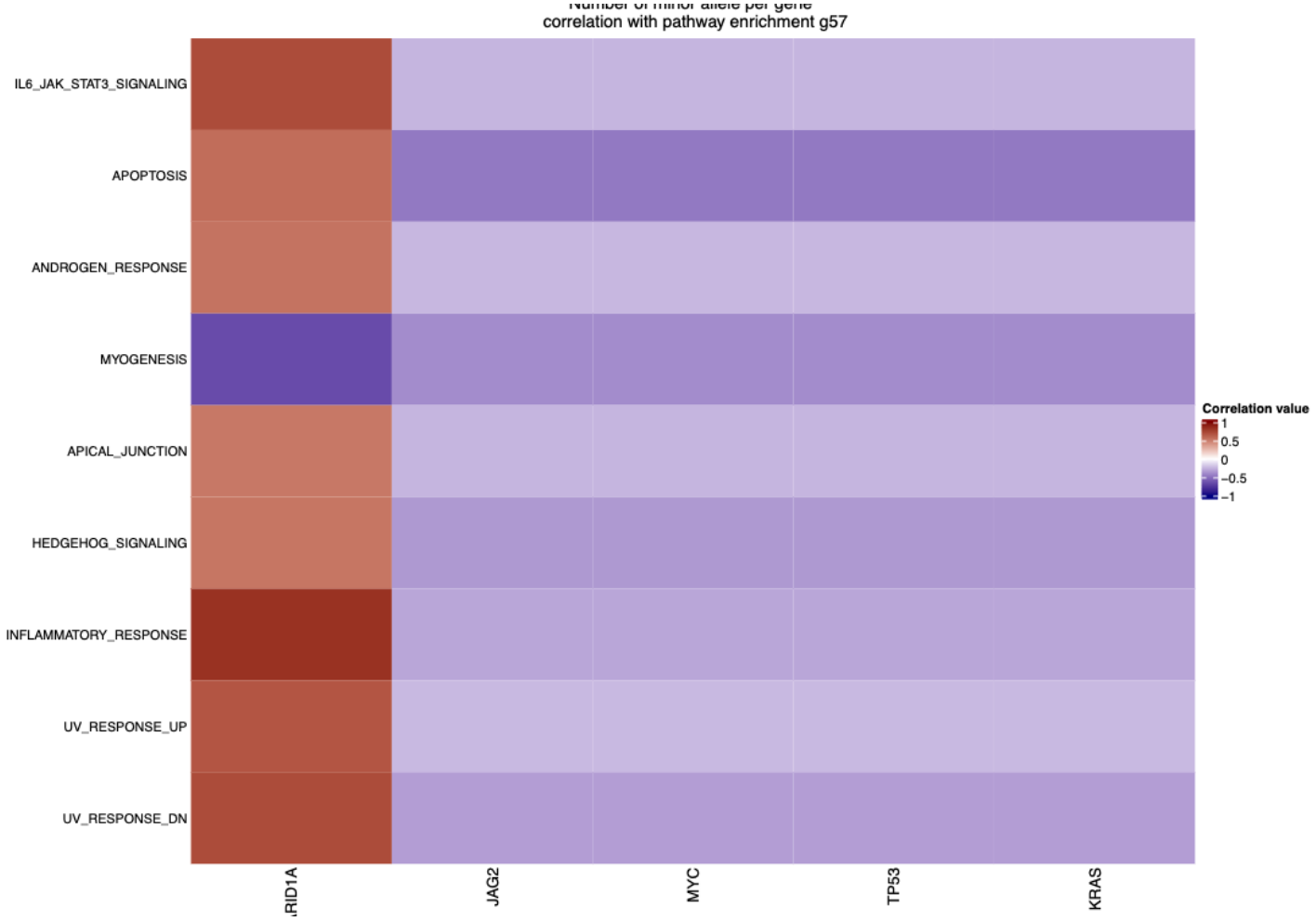

Figure 16. Represents the correlation between minor allele per gene count and the pathway enrichment scores of all the selected CTCs in patient G57 (21 cells in total).

Conclusion

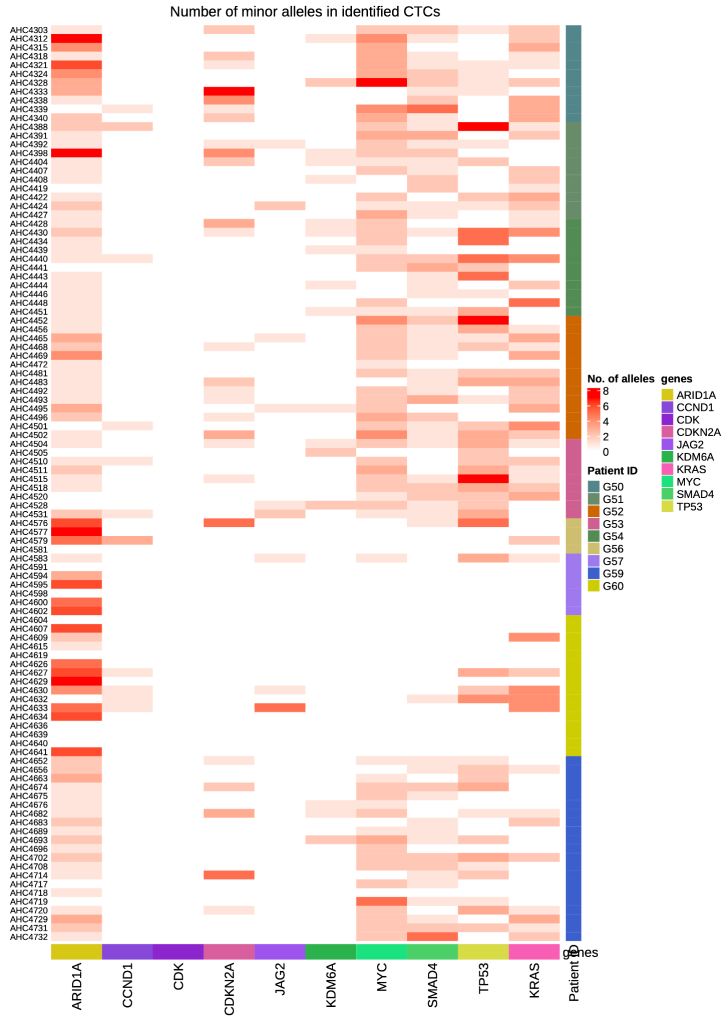

Figure 17: Minor allele per gene across all patients' final selected CTCs

CTCs cells exhibit heterogeneity as well as homology in mutation profiles across patient samples check Figure 13 and Figure 16. This is observed even in the transcriptomic profiles of the patients, see Figure 12 and Figure 15. This observation indicates the nature of pancreatic cancer itself, the reason it is hard to treat using current therapy regimens.

By showing the correlation between the observed mutations with the pathway, it can be concluded that the minor allele count can be used to identify the CTCs. The KRAS pathway is highly dysregulated in the CTCs across all the patient samples. ARID1A has been reported very less to harbor mutations, less than 10% among all the TCGA

pancreatic cancer samples shows mutation in ARID1A gene but we found a high number of mutations on ARID1A gene across all the patients in all the identified CTCs, see Figure 17.

The results from this study indicates that targeted sequencing has the ability for rapid identification of cancer for diagnosis and prognosis and combining concurrent data of multiple profiles can help explain pathobiology of the pancreatic cancer to develop effective therapy regimens.

Method

CTC Isolation
A total of 7.5ml EDTA blood was collected from the patients. The CTC isolation was performed using the ClearCell FX instrument following the procedures recommended by the manufacturer. Briefly, blood lysis was conducted by addition of RBC lysis buffer in the ratio of 1:3 (blood: RBC lysis buffer). The mixture was inverted 10 times and incubated at room temperature for 10 minutes, followed by centrifugation at 500g for 10 minutes. The cell pellet was resuspended in 4 ml of resuspension buffer. The sample was processed using Protocol 4 for higher purity CTC isolation.

Tumor Tissues Dissociation
The tumor was washed thrice with cold washing buffer (DMEM/F12 (Gibco), 5% FBS (Standard, Gibco) supplemented with 1% 10,000U/ml penicillin–streptomycin (Gibco)). and chopped into smaller pieces with a sterile scalpel blade (Aesculap) followed by an incubation at 37 °C in 10 ml collagenase + dispase concoction (1 mg/ml) with shaking for 2 to 3 h. The suspensions were repeatedly washed with a washing buffer for 5 times, and passed through the pre-wet 40 µm cell strainers (BD Falcon, San Jose, CA, USA) for single cells. The cells were counted using the Moxi™ Z Mini Automated Cell Counter (ORFLO Technologies).

PCR Primer Design
We designed thirteen pairs of PCR primers at the intronic regions of 10 genes that are most frequently altered in pancreatic cancers (ARID1A, CCND1, CDK6, CDKN2A, JAG2, KDM6A, MYC, SMAD4, TP53 and KRAS) that generate PCR products of approximately 500-800 bp with annealing temperature around 60∞C.

Library Preparation for Next-Generation Sequencing
The library preparation was carried out using the Nextera XT Sample Preparation kit (Illumina) with minor modifications and one quarter of the recommended volume. Briefly, the cDNA and gDNA amplicons were diluted to the range of 0.1 – 0.3ng/µl with C1 Harvest Reagent (Fluidigm) followed by tagmentation at 55∞C for 10min. The tagmentation was inactivated by addition of NT Buffer. The Nextera PCR Master Mix and XT Indices were added into the sample. PCR

amplification was carried out with the following conditions: 72∞C for 3min; 95∞C for 30sec, 12 cycles of 95∞C for 10 sec, 55∞C for 30sec, 72∞C for 1 min; 72∞C for 5 min. The amplified products were pooled together and purified twice with 0.9X AMPure XP SPRI beads (Beckman Coulter). The cDNA libraries were sequenced on the NextSeq (Illumina) platform for 76bp single read sequencing run while the gDNA libraries were sequenced on the MiSeq (Illumina) platform for 250 bp paired-end sequencing run.

Data Analysis

Mutational profile analysis: BWA and GATK tools were used to generate VCF files.
Transcriptomic profile analysis: STAR and RSEM tools were used to generate raw counts.
Pathway enrichment: Unipath tool was used for pathway enrichment analysis.

Bibliography

1. Volpe A, Finelli A, Gill IS, Jewett MAS, Martignoni G, Polascik TJ, et al. Rationale for percutaneous biopsy and histologic characterisation of renal tumours. Eur Urol. 2012;62: 491–504.

2. Protein signatures for survival and recurrence in metastatic melanoma. J Proteomics. 2011;74: 1002–1014.

3. The rise of DNA methylation and the importance of chromatin on multidrug resistance in cancer. Exp Cell Res. 2003;290: 177–194.

4. Szakács G, Paterson JK, Ludwig JA, Booth-Genthe C, Gottesman MM. Targeting multidrug resistance in cancer. Nat Rev Drug Discov. 2006;5: 219–234.

5. Dimitrakopoulos C, Hindupur SK, Colombi M, Liko D, Ng CKY, Piscuoglio S, et al. Multi-omics data integration reveals novel drug targets in hepatocellular carcinoma. BMC Genomics. 2021;22: 1–26.

6. Serratì S, De Summa S, Pilato B, Petriella D, Lacalamita R, Tommasi S, et al. Next-generation sequencing: advances and applications in cancer diagnosis. Onco Targets Ther. 2016;9: 7355.

7. Single-Cell Analyses Inform Mechanisms of Myeloid-Targeted Therapies in Colon Cancer. Cell. 2020;181: 442–459.e29.

8. Kim D, Kobayashi T, Voisin B, Jo J-H, Sakamoto K, Jin S-P, et al. Targeted therapy guided by single-cell transcriptomic analysis in drug-induced hypersensitivity syndrome: a case report. Nat Med. 2020;26: 236–243.

9. Kong SL, Liu X, Tan SJ, Tai JA, Phua LY, Poh HM, et al. Complementary Sequential Circulating Tumor Cell (CTC) and Cell-Free Tumor DNA (ctDNA) Profiling Reveals Metastatic Heterogeneity and Genomic Changes in Lung Cancer and Breast Cancer. Front Oncol. 2021;11. doi:10.3389/fonc.2021.698551

10. Verma M, Barh D. Progress and Challenges in Precision Medicine. Academic Press; 2016.

11. Bedard PL, Hansen AR, Ratain MJ, Siu LL. Tumour heterogeneity in the clinic. Nature. 2013;501: 355–364.

12. Bai Y, Zhao H. Liquid biopsy in tumors: opportunities and challenges. Annals of Translational Medicine. 2018;6. doi:10.21037/atm.2018.11.31

13. Merker JD, Oxnard GR, Compton C, Diehn M, Hurley P, Lazar AJ, et al. Circulating Tumor DNA Analysis in Patients With Cancer: American Society of Clinical Oncology and College of American Pathologists Joint Review. J Clin Oncol. 2018;36: 1631–1641.

14. Piñeiro R. Circulating Tumor Cells in Breast Cancer Metastatic Disease. Springer Nature; 2020.

15. Keller L, Pantel K. Unravelling tumour heterogeneity by single-cell profiling of circulating tumour cells. Nat Rev Cancer. 2019;19: 553–567.

16. Tr A. A case of cancer in which cells similar to those in the tumours were seen in the blood after death. Australas Med J. 1869;14: 146.

17. Hofman P, Calin GA, Mani SA. Towards New Promising Discoveries for Lung Cancer Patients: A Selection of Papers from the First Joint Meeting on Lung Cancer of the FHU OncoAge (Nice, France) and the MD Anderson Cancer Center (Houston, TX, USA). MDPI; 2019.

18. Plaks V, Koopman CD, Werb Z. Circulating Tumor Cells. Science. 2013;341. doi:10.1126/science.1235226

19. Zhao Q, Yuan Z, Wang H, Zhang H, Duan G, Zhang X. Role of circulating tumor cells in diagnosis of lung cancer: a systematic review and meta-analysis. J Int Med Res. 2021;49. doi:10.1177/0300060521994926

20. Kujur PK, Flores BCT, Ramalingam N, Chinen LTD, Jeffrey SS. Advances in the Characterization of Circulating Tumor Cells in Metastatic Breast Cancer: Single Cell Analyses and Interactions, and Patient-Derived Models for Drug Testing. Adv Exp Med Biol. 2020;1220: 61–80.

21. Yin L, Pu N, Thompson E, Miao Y, Wolfgang C, Yu J. Improved Assessment of Response Status in Patients with Pancreatic Cancer Treated with Neoadjuvant Therapy using Somatic Mutations and Liquid Biopsy Analysis. Clin Cancer Res. 2021;27: 740–748.

22. Inhestern J, Oertel K, Stemmann V, Schmalenberg H, Dietz A, Rotter N, et al. Prognostic Role of Circulating Tumor Cells during Induction Chemotherapy Followed by Curative Surgery Combined with Postoperative Radiotherapy in Patients with Locally Advanced Oral and Oropharyngeal Squamous Cell Cancer. PLoS One. 2015;10: e0132901.

23. Ou H, Huang Y, Xiang L, Chen Z, Fang Y, Lin Y, et al. Circulating Tumor Cell Phenotype Indicates Poor Survival and Recurrence After Surgery for Hepatocellular Carcinoma. Dig Dis Sci. 2018;63: 2373–2380.

24. Pantel K, Speicher MR. The biology of circulating tumor cells. Oncogene. 2016;35: 1216–1224.

25. Eslami-S Z, Cortés-Hernández LE, Thomas F, Pantel K, Alix-Panabières C. Functional

analysis of circulating tumour cells: the KEY to understand the biology of the metastatic cascade. Br J Cancer. 2022;127: 800–810.

26. Fidler IJ. Metastasis: Quantitative Analysis of Distribution and Fate of Tumor Emboli Labeled With 125I-5-Iodo-2′ -deoxyuridine23. J Natl Cancer Inst. 1970;45: 773–782.

27. Nagrath S, Sequist LV, Maheswaran S, Bell DW, Irimia D, Ulkus L, et al. Isolation of rare circulating tumour cells in cancer patients by microchip technology. Nature. 2007;450: 1235–1239.

28. Stott SL, Hsu C-H, Tsukrov DI, Yu M, Miyamoto DT, Waltman BA, et al. Isolation of circulating tumor cells using a microvortex-generating herringbone-chip. Proc Natl Acad Sci U S A. 2010;107: 18392–18397.

29. Kim TH, Wang Y, Oliver CR, Thamm DH, Cooling L, Paoletti C, et al. A temporary indwelling intravascular aphaeretic system for in vivo enrichment of circulating tumor cells. Nat Commun. 2019;10: 1478.

30. Liu Z, Fusi A, Klopocki E, Schmittel A, Tinhofer I, Nonnenmacher A, et al. Negative enrichment by immunomagnetic nanobeads for unbiased characterization of circulating tumor cells from peripheral blood of cancer patients. J Transl Med. 2011;9: 1–8.

31. Payne K, Brooks JM, Taylor GS, Batis N, Noyvert B, Pan Y, et al. Immediate Sample Fixation Increases Circulating Tumour Cell (CTC) Capture and Preserves Phenotype in Head and Neck Squamous Cell Carcinoma: Towards a Standardised Approach to Microfluidic CTC Biomarker Discovery. Cancers . 2021;13: 5519.

32. Wu S, Gu L, Qin J, Zhang L, Sun F, Liu Z, et al. Rapid Label-Free Isolation of Circulating Tumor Cells from Patients' Peripheral Blood Using Electrically Charged FeO Nanoparticles. ACS Appl Mater Interfaces. 2020;12: 4193–4203.

33. Mishra A, Dubash TD, Edd JF, Jewett MK, Garre SG, Karabacak NM, et al. Ultrahigh-throughput magnetic sorting of large blood volumes for epitope-agnostic isolation of circulating tumor cells. Proc Natl Acad Sci U S A. 2020;117: 16839–16847.

34. Han L, Zi X, Garmire LX, Wu Y, Weissman SM, Pan X, et al. Co-detection and sequencing of genes and transcripts from the same single cells facilitated by a microfluidics platform. Sci Rep. 2014;4: 6485.

35. Kong SL, Li H, Tai JA, Courtois ET, Poh HM, Lau DP, et al. Concurrent Single-Cell RNA and Targeted DNA Sequencing on an Automated Platform for Comeasurement of Genomic and Transcriptomic Signatures. Clin Chem. 2019;65: 272–281.

36. Bray F, Ferlay J, Soerjomataram I, Siegel RL, Torre LA, Jemal A. Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. CA Cancer J Clin. 2018;68: 394–424.

37. Stewart BW, Wild CP. World Cancer Report 2014. 2014.

38. De La Cruz MSD, Young AP, Ruffin MT. Diagnosis and management of pancreatic cancer. Am Fam Physician. 2014;89: 626–632.

39. Martincorena I, Campbell PJ. Somatic mutation in cancer and normal cells. Science. 2015.

pp. 1483–1489. doi:10.1126/science.aab4082

40. Navin NE, Hicks J. Tracing the tumor lineage. Mol Oncol. 2010;4: 267.