Allelic imbalance among cancer associated genes in the pan-adenocarcinoma preferred mutations

Degree dissertation

Lo, Chunhsuan (羅　鈞軒)

Department of Genetics

School of Life Science

The Graduate University for Advanced Studies, SOKENDAI

Allelic imbalance among cancer associated genes in the pan-adenocarcinoma preferred mutations

Doctor of Philosophy

Author: Lo, Chunhsuan (羅　鈞軒)　　***Supervisor: Inoue, Ituro (井ノ上　逸朗)***

Department of Genetics

School of Life Science

The Graduate University for Advanced Studies, SOKENDAI

December 2022

The Graduate University for Advanced Studies, SOKENDAI

Human Genetics Laboratory, Department of Genomics and Evolutionary Biology, National Institute of Genetics (NIG), Japan.

DEFENSE COMMITTEE AND FINAL READING APPROVALS

of the dissertation submitted by

Lo, Chunhsuan（羅　鈞軒）

Dissertation Title: Allelic imbalance among cancer associated genes in the pan-adenocarcinoma preferred mutations

The Date of Final Oral Examination: January Sixth, 2023

The following individuals have read and discussed the dissertation submitted by student Chunhsuan LO (羅 鈞軒), and they evaluated the student's presentation and response to questions during the final oral examination. They finally decided that the student has passed the final oral examination.

Prof. Yasukazu NAKAMURA (中村 保一), Ph.D. Chair, Committee

Prof. Masanori ARITA (有田 正規), Ph.D. Member, Committee

Prof. Kazuhiro MAESHIMA (前島 一博), Ph.D. Member, Committee

Prof. Kazuho IKEO (池尾 一穂), Ph.D. Member, Committee

Prof. Hirofumi NAKAOKA (中岡 博史), Ph.D. Member, Committee

The final reading approval of the dissertation is granted by Yasukazu NAKAMURA, Ph.D., Chair of the Supervisory Committee. The dissertation is also approved by National Institute of Genetics (NIG) Japan and The Graduate University for Advanced Studies, SOKENDAI.

DEDICATION

I would like to dedicate my Mother (黄 玉霞) and my Father (羅 文瑞) for their invaluable guidance, heart-warming support and endless love. Also, I thank to my other Family members and close Friends for giving me backing and relieving my pressures during my doctoral studies in Japan.

ACKNOWLEDGMENTS

I would like to express deep gratitude to my advisor Prof. Ituro Inoue (井ノ上 逸朗) for the kind supervision. His immense contribution of inspiration, encouragement, and patience are priceless values and the reasons for the excitement and productivity of my research.

I would like to give special thanks to Prof. Hirofumi Nakaoka (中岡 博史), Prof. Naoko Fujito (藤戸 尚子) and Prof. Kazuyoshi Hosomichi (細道 一善) for assisting me to design the analysis and for providing the technical supports.

I am thankful to my advisory committees, Prof. Yasukazu Nakamura (中村 保一), Prof. Masanori Arita (有田 正規), Prof. Kazuhiro Maeshima (前島 一博) and Prof. Kazuho Ikeo (池尾 一穂) for their insightful comments and brilliant suggestions.

Especially, this work is kindly supported by National Institute of Genetics (NIG) Japan and The Graduate University for Advanced Studies, SOKENDAI. I would like to thank the NIG supercomputer team for providing me with parallel computing resources. The computations in this publication were partially performed on the NIG supercomputer at ROIS National Institute of Genetics. Also, I appreciated GTEx Consortium and TCGA group for sharing the bioinformatics data, and I appreciated the careful review of my manuscript by those online reviewers.

I appreciate my lab members for being willing to discuss and provide thoughtful comments on my research. Also, the sharing from all dear friends in NIG also delivered a warm encouragement. I am grateful to my family for always being by my side with a constant source of love and motivation. I would like to thank Japan and Taiwan exchange association for the financial support. Lastly, I would like to deeply thank all my friends for making my stay in Japan memorable.

ABSTRACT

To investigate the biological roles of allele-specific transcription, I analyzed large-scale pan-adenocarcinoma bulk RNA-Seq data from The Cancer Genome Atlas (TCGA) with paired genotype data. Among the genes being reported by the Catalogue of Somatic Mutations in Cancer (COSMIC) database, their heterozygous somatic mutation carriers were screened for RNA allelic imbalance (RAI) due to transcriptional efficiency. The biased aspects of allelic imbalance of those expressed mutations inside purity corrected tumor micro-environment (TME) can be further grouped into cancer driver ones and non-cancer driver ones. For cancer driver genes, the effects of oncogenesis or the impacts of insufficiency in tumor suppression can be intensified by increasing the expression of mutant alleles through the observed **allelic imbalance in transcriptional efficiency** (**AITE**) effect. This is in contrast to those mutated non-cancer driver genes, whose normal alleles are instead increased in the expression for buffering the mutant proteins inside TME.

Totally, my study analyzed large-scale pan-adenocarcinoma samples from TCGA program (n=3256). The genotypes and allele-specific copy numbers inside each gene were checked in advance by SNP-array and whole exome sequencing data (WESeq). Then, based on the paired RNA sequencing data (RNA-Seq) of the same homogenized tissue extract, the AITE effects were screened for those heterozygous positions without overlapping with allele-specific somatic copy number alteration (SCNA) regions.

Among those expressed somatic mutations, most nonsynonymous SNV mutations of well-known **cancer driver genes biased their transcriptional amount toward the mutant allele** inside TME in statistically significant ways, where the average mutant allele frequency (MAF) in RNA-Seq was significantly **higher** than that in WESeq (Welch's t-test, $p = 4.86 \times 10^{-3}$ for *TP53*; $p = 3.20 \times 10^{-12}$ for *KRAS*; $p = 9.75 \times 10^{-4}$ for *CTNNB1*; $p = 3.80 \times 10^{-3}$ for *PTEN*). In contrast, the nonsynonymous SNV mutations of **non-cancer driver (passenger) genes biased their transcriptional amount toward the normal (reference) allele** inside TME, where the average MAF in RNA-Seq was significantly **lower** than that in WESeq (Welch's t-test, $p = 2.33 \times 10^{-20}$ for *COL3A1*; $p = 5.82 \times 10^{-23}$ for *MYH11*; $p = 2.59 \times 10^{-5}$ for *PDGFRB*; $p = 4.41 \times 10^{-3}$ for *TNC*; $p = 1.01 \times 10^{-12}$ for *PTPRC*). Especially, the bias extent of those AITE genes with increased mutant allele expression could be found in positive correlation with positive selection signals by dN/dS ratio (R-pearson = 0.17 and P = 0.17 for corresponding to the average of alternate allele fraction in RNA among cancer patients; R-pearson = 0.13 and P = 0.24 for corresponding to the proportion of cases showing alternate allele bias).

Overall, the tumors showed the significantly higher AITE occurring rates in their genomes when being compared to the paired-solid-normal tissues (Welch's t-test, $p < 2.2 \times 10^{-16}$, n=152). By further comparing their whole genome sequencing data (WGSeq), more *de novo* transposable elements (TE) insertion sites were also detected inside tumors than that inside normal tissues with significance (Welch's t-test, p=0.0456, n=27). Within 1 Mbp around of AITE genes, tumor-specific *de novo* TE insertions were discovered to be connected to AITE occurrence.

Finally, the single-cell RNA-Seq analysis has further revealed the intra-tissue heterogeneity inside TME for the cancer cells specific transcriptomes, which supports my current conclusions about allelic imbalance.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF ABBREVIATIONS

AFD	Allele Frequency Difference
AITE	Allelic Imbalance in Transcriptional Efficiency
ATAC-Seq	Assay for Transposase-Accessible Chromatin using sequencing
CGC genes	Cancer Gene Census reported genes (Cancer-associated genes)
CGH	Comparative Genomic Hybridization
CNA	Copy Number Alteration
CNV	Copy Number Variation
COSMIC	Catalogue of Somatic Mutations In Cancer
CRE	Cis-Regulatory Element
CTCF	CCCTC-binding zinc finger protein
GO	Gene Ontology
GTEx	Genotype-Tissue Expression project
GWAS	Genome-wide association studies
HLA	Human Leukocyte Antigen
ICGC	International Cancer Genome Consortium
InDel	Insertion and Deletion
LOH	Loss of Heterozygosity
MAF	Mutant Allele Frequency
NGS	Next-generation Sequencing technology
RAI	RNA Allelic Imbalance
RNA-Seq	RNA Sequencing data (Whole transcriptome sequencing data)
SCNA	Somatic Copy Number Alteration
scRNA-Seq	single-cells RNA-Seq
SNP	Single Nucleotide Polymorphism
SNV	Single Nucleotide Variant
TAD	Topologically Associated Domain
TASC	Tumor-Associated Stromal Cell
TCGA	The Cancer Genome Atlas program
TE	Transposable Element
TME	Tumor Micro-Environment
TRE	Trans-Regulatory Element
TSG	Tumor Suppressor Genes
VAF	Variant Allele Fraction
WESeq	Whole Exome Sequencing data
WGSeq	Whole Genome Sequencing data

CHAPTER ONE: BACKGROUNDS

This chapter begins with some basic background introductions. I will introduce sequencing technology, big data analysis of genome sequences, and cancer studies.

1.1 Revolutions in genotyping technologies

Since scientists solved the chemical structures of DNA molecules (1), various innovative instrumentations and reagents for identifying the segments of genetic sequences have been swiftly developed. It also brought those array-based technologies for high-throughput genotyping.

By Sanger sequencing methods (first generation), the genotyping of an entire human genome had taken 13 years until scientists declared the completion of Human Genome Project (2) in 2003. After that, the first release of high-throughput sequencing platform emerged in the mid-2000s, which made a 50,000-fold cost drop in genome sequencing industries (3). This advancement enabled individual multi-genomics investigation and realized medicine with personalized precision. In order to distinguish it from the capillary electrophoresis-based traditional Sanger method, scientists began calling this short-reads-shot-gun sequencing method with camera-based detector as Next Generation Sequencing (NGS) technology.

When it came to around the 2020s, scientists further developed single-cell sequencing technologies, which can be performed either in NGS shotgun sequencing platform (second generation) or in single molecular platform (third generation). Such single-cell sequencing technologies in genomics, transcriptomics, or epigenomics have upgraded the resolution from bulk tissue level to single-cells level, which excludes the potential confounders deriving from tissue heterogeneity, allows scientists to address many new and longstanding questions (4). Especially, this brand-new sequencing scope enables us to investigate the specific transcriptional changes and their corresponding regulating mechanisms inside each single-cell with higher precision inside bulk tumor tissue, which used to be considered as tumor micro-environments (TME) by bulk sequencing data, existed unsolvable tissue heterogeneity issues before the emergence of single-cells sequencing data.

In this section, I compare those different high-throughput sequencing platforms throughout recent history, and it is important to understand how to combine the different omics data from online databases, whose data might be generated from various genotyping platforms.

1.1.1 Next generation sequencing technology (short reads)

Before the NGS platform came out, array technology had shown several limitations in high-throughput genotyping. Fortunately, the shotgun-sequencing-based NGS methods offered a number of advantages compared to the previous array-based methods. For example, WESeq, WGSeq, and RNA-Seq technologies enable faster profiling and deeper investigation of the exome, genome and transcriptome, respectively (5–7). NGS platform also equips the ability in detecting novel sequences which the array platform lacks. Besides, NGS platform can work with broader dynamic range, better specificity and even with higher sensitivity. Especially, by applying different sequence library constructing methods, various types of high-throughput epigenetics data can also be collected for different investigating purposes (8). In the following lines, the most widely used NGS platforms are going to be introduced:

- Illumina:

Illumina sequencing machines came out by combining the flow cells to shotgun sequencing technology. Inside the machine, thousands to millions of sequencing reactions are produced in parallel, where the synthesis reagents consist of primers, DNA polymerase, and four differently labeled reversible terminator nucleotides (3). During each cycle of the nucleotide incorporation, the nucleotide type is identified by its corresponding fluorescent color, when the 3' terminator on the base and the fluorophore are removed (8). It is worth noting that through relying on the various types of NGS libraries which are prepared by different biochemical protocols, the machine can further acquire the differently modified genomic or epigenomic signals to fulfill different research purposes.

- 10X Genomics Chromium and Smart-seq2:

Combined with Illumina sequencing platform, several single-cell sequencing methods have been developed for addressing new biological questions, such as Smart-seq, Drop-seq, CEL-seq, MARS-seq, SCRB-seq and others (9). Among them, 10X Genomics Chromium (10X) approach and the plate-based Smart-seq2 full-length approach are two frequently used scRNA-Seq methods (10). The droplet-based 10X Genomics Chromium technology can generate transcriptional RNA, genomic DNA profiles and epigenetic profiles from the same single-cell simultaneously. By combining unique molecular identifiers (UMIs) with cell barcodes, the gel beads in emulsion (GEM) technology being built on a novel chip architecture enables high-throughput analysis with huge sample cell numbers. In contrast, the plate-based Smart-seq2 full-length technology can generate relatively deeper read coverages and more complete mRNA sequence information for higher sensitivity in gene detections for each cell, but it shows lower throughput of sample cell numbers (9,10).

1.1.2 Single molecular sequencing technology (long reads)

- Nanopore/Pacbio:

Different from the previous shotgun sequencing methods, the molecular detector here can directly read the single DNR/RNA molecule as a long unsheared nucleotide sequence. So, this single molecular sequencing technology is also known as long-read molecular sequencing. By the nanosized pore, the shifts in electrical conductivity can be measured, when DNA/RNA strands pass through its biological pore (11,12). Such single molecular technology is a completely new sequencing method, and it has been currently under active development since the 2020s (13). These platforms enabled the allele-specific genotype phasing and also enabled the investigations and identifications of repeat sequences inside human genomes, such as characterizing or classifying the *de novo*[1] mutations in centromere, telomere, and transposable elements (TEs) inserted regions.

1.2 Big Data analysis emerged as the fourth paradigm of science

Thanks to the huge revolutions in genetic sequencing platforms, the lower cost in high-throughput sequencing technologies has led to an unrivaled explosion in the accumulated amount of multi-omics data, and this escalation has collaterally raised the advanced challenges in archiving, integrating, analyzing, and sharing these Big Data (14). Inevitably, this is also why the integrative approaches to utilizing online databases have already become necessary skills for genomic researchers.

1.2.1 Big Data in multi-genomics sequences

[1] The term "*De novo"* means "from the beginning", and it is used to describe a change in the DNA sequence of a gene that is seen for the first time in a person and has not appeared in previous generations. Therefore, a *de novo* mutation can occur in an egg or sperm cell of a parent, in the fertilized egg soon after the egg and sperm unite, or in another type of cell during embryo development. More details can be referred from https://www.cancer.gov/publications/dictionaries/cancer-terms/def/de-novo-mutation

Currently, many human disease researchers have started with the method of Big Data. To better understand what Big Data means, we should understand that Big Data equips the following characteristics, such as huge volume, high velocity, diverse variety, exhaustive scopes, fine resolutions, and so on, as Rob Kitchin and Gavin McArdle have introduced (15).

Also, one of those Big Data advocates, Whilst Jim Gray, has envisaged Big Data as the fourth paradigm of science, which can be data-intensive and a radically new extension of the established scientific method (16). In other words, Big Data has been expected to construct a new research paradigm across multiple disciplines, which could be currently classified as experimental science, theoretical science, computational science and exploratory science, respectively (16,17). Regarding how those methods may be applied in manipulating Big Data, they have also reframed the key questions about the constitution of knowledge, the processes of research, the engagement with information, and the nature and categorization of reality, and so on, as mentioned by Danah boyd and Kate Crawford (18).

Especially, since the Twenty-First Century, the era of Big Data in human genetics has begun. Advances in high-throughput molecular biology and electronic clinical health records are coupled with increasing computer capabilities (19), which have popularized the utilization of Big Data in healthcare systems, have promised better personalized and precision medicine for patients. Such revolutions have created a radical shift in medical researches (19). First and last, scientists are currently inevitable to strike a better balance between those data-driven researches and hypothesis-driven researches (20).

1.2.2 Online genomics data banks

Online resources are useful for those people who want to start their biological researches from bioinformatics, and it is especially for those people who would like to further study disease etiologies in ways of system biology (21). Nowadays, various types of high-throughput sequencing libraries have already accumulated huge amounts of data for bioinformaticians in online databases. For example, DNA Data Bank of Japan (DDBJ) (22) is a very famous and professional institute all over the world. DDBJ provides various types of sequence data in accompany with several online analyzing tools for researchers. Besides DDBJ, there are also many other online databases for storing multi-genomics-based sequence data, such as ICGC, NCBI, EBI, DDBJ, SRA, ENCODE, ROADMAP, BIG, SCBIT, CKB, and so on, (22–28). Those databases are constructed for different purposes of usage, and they are supported by different countries around the world, and they are all openly available for worldwide genome researchers. For my studies, Genotype-Tissue Expression (GTEx) and The Cancer Genome Atlas (TCGA) have been utilized to investigate the processes of cancer progressions (23,24).

1.3 Cancers

Cancers acting as complex diseases are my main research topic. In fact, cancerization includes complicated progressions. From the normal cells, benign tumor cells, precancerous cells and then into malignant cancer cells, multiple steps are required for cell transformation. Nevertheless, in contrast with benign tumors which are identified by their abnormal cell growth without spreading, cancers additionally show the potential to invade or spread to other parts of the body, which is called metastasis. For more details, cancers can be further classified into different types by the diverse origins of tumor cells, such as carcinoma (from epithelial cells), sarcoma (from connective tissue), lymphoma and leukemia (from blood-forming cells), germ cell tumor (from pluripotent cells), or blastoma (from immature precursor cells or embryonic tissue), and so forth, as being indicated by Claudette G. Varricchio (29).

1.3.1 Epidemiology and etiology of cancers

To understand cancers more, we have to realize that this type of disease has comprised a bewildering assortment of illnesses which may kill several millions of people each year. For example, according to online data in 2018, around 18.1 million new cases of cancer and 9.6 million deaths were reported globally (30). It was estimated that about 20% of males and 17% of females might have gotten cancer at some lifetime point while 13% of males and 9% of females would even die from it (30).

In fact, many types of cancers being diagnosed in the human population show an age-dependent incidence, which implicates some extent of stochastic but rate-limiting events (31). In other words, the associations between aging and oncogenesis have been reported, which are usually attributed to the accumulated errors in DNA over a lifetime (32). Based on those observations in human cancer cell lines and animal models (33,34), scientists begin to argue that tumor development proceeds via a process formally analogous to Darwinian selection inside the body, in which a succession of genetic changes may stochastically acquire growth advantage (35). Those accumulated genetic/epigenetics changes are supposed to be the keys leading to the progressive conversion of normal cells into cancer cells (36). Even though such events seem to be inevitable for every individual during their lifetime development, the triggering time points of the oncogenesis can still be quite different between each individual, which is actually hard to be explained rigorously based on current scientific knowledge.

1.3.2 Carcinogenesis and cancer progression

From the other view, cancers are supposed to be induced by the cumulative effects of both environmental carcinogens and spontaneous random genetic variations until the triggering point. In a real situation, a large number of cancer cells originate from solid tumor cells until they can grow to invade other parts of the body, and such transformation step may involve serial procedures for their development from normal tissues into malignant tumors, as the following timelines being introduced by Douglas Hanahan and Robert A. Weinberg (37): “cell growth and division without the proper signals”, “continuous growth and division without controls”, “avoidance of programmed cell death”, “limitless number of cell proliferation”, “angiogenesis for constructing blood vessels”, “tissue invasion”, and “metastases formation” (37).

Nevertheless, cancer patients typically do not die from the effects of the primary tumor but from the metastases (4). Despite decades of researches, the key question of how metastasis originates has still not been fully answered yet, and even the universal key for the derivation of cancer progenitor cells is still unclear, also. Even now, many scientists are trying to demythologize its veiled causality by analyzing those mutations located around cancer-associated genes, which might be either cancer driver genes or passenger genes.

1.3.3 Cancer driver genes and passenger genes

The Cancer Gene Census (CGC) has an ongoing effort to catalog those genes whose mutations might be causally implicated in cancer (38), and The Catalogue of Somatic Mutations in Cancer (COSMIC) is specifically focusing on those somatic mutation sites which are associated with cancer patients. Based on their accumulated evidence, CGC-COSMIC has reported several hundreds of cancer-associated genes as CGC genes, which present relatively higher mutation rates in cancer tissues, can be basically classified into the following categories: “Oncogenes”, which have the function of activating cell proliferation, belong to cancer driver genes. “Tumor suppressor genes”, which have the function of inhibiting cellular proliferation, also belong to cancer driver genes. “Passenger genes”, whose mutation or malfunction is not directly related to driving carcinogenesis at all. But currently, scientists have not yet found the golden criteria to precisely identify them from those cancer-associated genes.

1. Oncogenes:

Most of the oncogenes are derived from proto-oncogenes, which are genes involved in growth factors, growth factor receptors, signal transduction molecular, and nuclear transcription factors (39). Their mutations are usually "gain of function" type. They are dominant at the cellular level, and this also implies that only a single copy of the mutated allele might be able to lead to tumor progression (40,41).

2. Tumor suppressor genes (TSG):

In general, when compared to those mutations in oncogenes, those occurrences in tumor suppressor genes are much more frequently observed (42). For tumor suppressor genes, their mutations are usually “loss of function” type, and they are recessive at the cellular level (43). Only when the second hit happens, which can be either another mutation or loss of heterozygosity (LOH) in the other normal allele, can make both alleles lose their functions for tumor suppressing to trigger cancer progression in the most cases. Nevertheless, the incomplete penetrance of this type of mutation at the individual patient level has still been observed, which can somehow be explained by the protein amount being quantitatively taken over by the mutant type for enhancing the disease phenotype according to Alfred G. Knudson (44). Furthermore, we have to notice that this type of mutation may be dominant at the individual patient level among those germline mutation carriers (45), because there are millions of cells having the chance to acquire the mutation for the other allele, as mentioned by the textbook records (44,46).

3. Passenger genes:

The biological functions or malfunctions of passenger genes are not directly related to driving tumor initiation or cancer progression, which is opposed to oncogenes or TSGs. In other words, those mutations inside passenger genes do not benefit the fitness of those cancer cells or precancerous cells at all. But those mutations in passenger genes still randomly occur in a cell that coincidentally or subsequently acquired another driver mutation, and therefore they are able to be simultaneously found inside cancer cells which simultaneously carry the mutated cancer driver genes by chance.

1.3.4 Old and new hypotheses for how carcinogenesis is triggered

According to the “two genetic hits hypothesis” of carcinogenesis proposed by Alfred G. Knudson in the 1970s, tumorigenesis in a precancerous cell could only be initiated when it completely acquired both functionally damaged alleles of the TSG (47), which was typically considered in a recessive way for leading to the mutant phenotype (48). In other words, those individuals who inherited only one copy of the mutant allele were still not ready for triggering cancer progression, and the additional events of somatic mutations falling on the other normal allele were expected for provoking the inherited carcinoma (49,50). Similarly, for those common individuals without any inherited pathological alleles, at least two somatic mutation events were supposed to respectively occur in both alleles of the same TSG in order to start up sporadic carcinoma (51–53)

However, driven by the recent advances in next generation sequencing (NGS) technologies, large amounts of tumor genomics data have been accumulated (3,37). Many well-known cancer driver genes carrying somatic mutations were reported to still remain in heterozygous status in the DNA level inside those tumor tissues, which the Knudson two genetic hits paradigm cannot well explain (54–57). Based on these observations, a recently new continuum model for redescribing the mechanisms in tumor suppression and progression was further proposed by Alfred G. Knudson group (44), which notifies that even only one hit of somatic mutation inside either allele of the TSG might still show the chance to trigger the cancer developments by taking the effective tumor suppression as a functional protein dosage dependent result (58).

Until now, it still remains uncertain why only partial heterozygous TSG coding mutation carriers have finally developed into cancers among those individuals harboring similar genomic mutation profiles in exomes (43). Although several epigenetic modifications have been observed as the second hit for a few cases to turn on or turn off the mutant allele expression which may be involved in DNA repair pathways (homologous recombination and mismatch repairs pathways) (59–61). But, it is still difficult to explain how heterozygous mutation carriers of cancer driver genes specifically lift up their expression amount in disease functional proteins specifically for enhancing malignancy (44) from the pan-adenocarcinoma wide scope. For this, my hypothesis has supposed that allelic imbalance mechanisms may have roles.

1.3.5 Bulk sequencing data reflects the tumor micro-environment (TME) and single-cells sequencing data resolves the issue of its tissue heterogeneity

According to "The Human Protein Atalas", various types of human gland tissues can be mainly composed by three different types of cells. They belong to stromal cells, peripheral blood mononuclear cells (PBMCs) and various epithelial cells according to different glands. These three different main clusters of cells can be usually found in the UMAP plots by the scRNA-Seqs data from normal human glands (62). Basically, parenchymal cells (epithelial cells) are main functional cells, which constitute building blocks for glandular tissues. And mesenchymal cells (stromal cells) facilitate the functions of those parenchymal cells, which support the glandular structure. Normal functions of those two types of cells are necessary for the normal gland tissues.

As for adenocarcinomas, cancer cells are mainly derived from epithelial cells according to cytology (63). On the contrary, stroma is made up of the non-malignant cells. But this does not mean the stromal cells are not important for tumor development. In normal tissues, glands comprise epithelial cells & stromal cells, that communicate with each other through extracellular matrix (ECM) (64). In tumor tissues, the interactions between cancer cells and stromal cells also play major roles for building tumor microenvironment (TME) (65). Cancer cells cannot even survive without the growth factors which are secreted by stromal cells (66,67). Under certain conditions, stromal cells can be converted into tumor-associated stromal cells (TASCs) inside TME for promoting cancer progressions (68,69).

In fact, cancer developments exist series of cell selections for shaping disease-preferred TME (70). Inside TME those cancer cells and TASCs may carry somatic mutations which those normal cells do not have. Purity correction steps may remove those infiltrated reads from normal cells inside bulk tumor tissue, but the issue of intra-tumor heterogeneity may still exist (71). Therefore, when analyzing TCGA data, we should regard the purity corrected bulk sequencing data as TME, and then supposing that all types of cells express the gene with similar level because of technical limitations, although the mixed reads from those "tumor cells and tumor-associated stromal cells" may confound the analysis. Finally, single-cell RNA-Seqs are further required for confirming how different types of cells contributed to the observed expressional changes inside TME to resolve bulk tissue heterogeneity.

2 CHAPTER TWO: MAIN RESEARCHES ON ALLELIC IMBALANCE IN TRANSCRIPTIONAL EFFICIENCY

2.1 Introductions

Recent advances in next-generation sequencing (NGS) technologies have led to the production of large amounts of cancer genomic data (3,37). Many well-known cancer driver genes carrying somatic mutations were reported to be still heterozygous in tumors without loss of heterozygosity (LOH), a phenomenon that is not readily explained by the Knudson two genetic hits paradigm (54–57). Although some breast cancer cases have suggested that allele-specific epigenetic modifications at promoter regions may act as the second hit for *BRCA1/2* (59–61), for other genes and cancer types it is not clear why partial heterozygous tumor suppressor gene (TSG) mutation carriers can still lead to cancer development in Pan-adenocarcinomas (43). As the number of observations of the allelic imbalance effect in tumor transcriptomes has accumulated (72,73), the amounts of RNA produced by normal and mutant alleles have been shown to be imbalanced inside bulk tumor tissue (being regarded as TME), which is called RAI effect here, (Fig 1A). Based on this, differences in functional protein dosage can be supposed to be regulated by allele-specific expression of the mutant allele, and the roles of allelic imbalance in RNA products (RAI effect) in facilitating cancer progression have also attracted considerable attention (58,74,75) for being a possible explanation. Such allelic imbalance effect is not only for TSG, but an analogous situation also arises in the case of proto-oncogenes, which enhance the biological effects caused by oncogenic mutations in tumor cells (40).

In previous cancer studies, the cause of widely observed RAI events at somatic mutation sites was unclear (72,73) and could be attributed either to nearby putative cis-regulatory elements (76–79) or to somatic insertion/deletion (indel) events in chromatin, SCNA events (80–82), the frequent genetic recombination events caused by genome instability (83). Allele-specific SCNA events confound the analysis of the mechanisms by which cis-regulatory elements control allele-specific expression (ASE). By combining RNA-Seq data with SNP-array and WESeq data from the same aliquot of each homogenized tissue extract, my approach is going to clarify whether most RAI cases were not due to the dosage changes caused by SCNAs in exonic regions (84) but rather due to the effects from putative cis-regulatory elements, which is further defined as allelic imbalance in transcriptional efficiency (AITE effect) inside bulk tumor tissue (being regarded as TME), (Fig 1B).

My analyses in pan-adenocarcinoma have suggested that allelic imbalance events in transcriptome widely presented among cancer-associated mutation sites, in most cases of which the allele-specific SCNAs did not overlap with the corresponding exonic regions inside tumors. Based on this evidence, my study confirmed that the observed RAI effects in cancers can be mainly regulated by AITE effects, in which the confounders from imbalanced gene dosage between two alleles at the DNA level were excluded in advance. Using both genomic and transcriptomics data, the unknown cis-regulatory elements and their corresponding biological roles in enhancing cancer driving effects were carefully identified and discussed.

Before this study, several papers have already compared the transcriptomes and the genomes from cancer tissues for allelic imbalance (85–88). In previous methods reported by Anelia Horvath, Nathan Edwards, et al. (87,88) and Tae-Min Kim, et al. (86), they directly compared the variant allele fractions (VAF) which were directly derived from the raw RNA and DNA sequence data of tumors. Although this way might still be able to screen out the potential effects from putative cis-regulatory elements for RNA allelic imbalance, we need to notice that the contamination reads from nearby non-tumor transcriptomes/genomes may confound the real VAF of RNA and DNA data from the collected tumor samples. Especially, the allele frequency difference (AFD) between RNA and DNA data without carefully removing outs those allele-specific SCNA regions will dilute the actual AFD values, and only a small proportion of the genes could be detected with significant bias with diluted AFD values (86,87). My method enables scientists to focus on those AITE genes with more significant and clearer bias sides, and the special bias patterns according to different gene functions are supposed to exist biological meaning and to be connected with selection signals, which has not been investigated in previous papers.

To better focus on the cancer genome, the Catalogue of Somatic Mutations in Cancer (COSMIC) Cancer Gene Census (CGC) (38) has already reported a group of cancer-associated genes which show the higher mutation frequency among cancer patients. Previous studies have put more focus in comparing the differences of VAF value between the CGC genes (cancer-associated genes) and non-CGC genes, where the p-value is not that strong even for nonsynonymous SNVs (88). Therefore, I am suspecting that among those CGC genes, they might be still required to be further grouped into two types of genes which may exist different bias sides of allelic imbalance in transcriptional efficiency (AITE), potentially reflect their roles as cancer driver ones or non-cancer driver (passenger) ones. Especially, the bias patterns among those CGC genes have not yet been carefully discussed in detail.

To better distinguish cancer driver mutations from passenger mutations has actually become a major challenge for the data mining in pan-cancer studies. In fact, not all of those CGC genes are functionally connected to carcinogenesis. Thus far, scientists have distinguished cancer driver mutations from non-cancer driver mutations among cancer-associated genes by analyzing the gene functions and the distribution patterns of mutation sites (38) with the reference from their corresponding dN/dS ratio for positive selection signals (89). Combined with my ideas in allelic imbalance bias extent, the different bias sides of CGC genes are supposed to further suggest the actual cancer driver genes with higher resolution and confidence.

2.2 Materials and methods

2.2.1 Integrating data from different online databases

TCGA and GTEx online databases were mainly utilized in my Pan-adenocarcinoma study. Those high-throughput sequence data based on Illumina short-reads platform and Affymetrix arrays platform were downloaded under authorization from dbGaP.

According to different sampling sources, the collected tissue types could be classified into three different groups for phenotype comparisons. They were tumor tissues (green), paired-normal-like tissues (blue) and control-normal tissues (orange), respectively, (S1 Fig)

2.2.1.1 Cancer samples were provided by the TCGA consortium

The Cancer Genome Atlas (TCGA) program (90,91) is a landmark in cancer genomics studies, and it has collected different kinds of NGS data of over 20,000 primary tumors, paired-normal-like tissues, and paired-peripheral-blood tissues, from patients with over 33 types of cancer. Other supplementary data, such as SNP-arrays, tissues biopsies, and clinical information are also provided. A complete list of those TCGA cancer type

abbreviations is provided at the following URL: https://gdc.cancer.gov/resources-tcga-users/tcga-code-tables/tcga-study-abbreviations. The detailed information about each downloaded data file can be viewed using the Genomic Data Commons Data Portal (https://portal.gdc.cancer.gov/) and the Genomic Data Commons Legacy Archive (https://portal.gdc.cancer.gov/legacy-archive/search/f). To download those access-controlled files, users must obtain authorization from the Database of Genotypes and Phenotypes (dbGaP, http://www.ncbi.nlm.nih.gov/gap).

2.2.1.2 Control-normal samples were provided by GTEx project

The Genotype-Tissue Expression (GTEx) program (92,93) is another landmark in transcriptomics studies, studying human gene expression, transcriptional regulation, and the relationship between these factors and genetic variations. This project has collected multiple human tissues from cancer-free donors for both high-throughput sequencing and SNP-array-based analysis. Because the sequencing data come from the tissues of cancer-free individuals, they can be used as control-normal samples in our cancer studies. The detailed list of tissue sampling sites is provided in the following URL: https://gtexportal.org/home/anatomogramPage. To download these access-controlled files in SRA format, users must obtain authorization from the Database of Genotypes and Phenotypes (dbGaP, http://www.ncbi.nlm.nih.gov/gap).

2.2.1.3 The selected reference genomes

All alignments for NGS reads were performed using the human reference genome GRCh38.d1.vd1. In this version, decoy viral sequences are included in the reference genome to prevent reads from aligning erroneously, and they can attract those reads from viruses that are known to be present in human samples. This reference sequence is also used by the GDC and can be directly downloaded from https://gdc.cancer.gov/about-data/data-harmonization-and-generation/gdc-reference-files.

2.2.2 Dataset selection

For each candidate cancer patient, the TCGA consortium provided a complete set of RNA-Seq, WESeq, and SNP-array data from the same homogenized tumor tissue blocks, but a few RNA-Seqs are along with the WESeqs and SNP-array data from blood. Among them, 10 major types of adenocarcinomas were selected for the pan-adenocarcinoma analyses. They are liver hepatocellular (LIHC), stomach (STAD), colon (COAD), lung (LUAD), prostate (PRAD), kidney renal clear cell carcinoma (KIRC), kidney renal papillary cell carcinoma (KIRP), rectum (READ), thyroid (THCA), and uterine corpus endometrial tissue (UCEC). In total, data from 3,256 cancer patients were used in the analysis, (S2 Table).

To investigate the RAI effect in all heterozygous genes across the genome and the tumor specific events, only those cases which provide a complete set of RNA-Seq, WESeq, and SNP-array data from both the tumor tissue and the paired-normal-like tissue were collected. A total of 152 patients with the 10 major types of adenocarcinomas from the TCGA consortium were used for further analyses, (S3 Table).

GTEx individuals with complete RNA-Seq and WESeq data were used as the control-normal tissues group (n = 50), (S4 Table).

2.2.3 Getting corrected allelic-specific read coverages based on the estimated tumor purity

Before utilizing the tumor and paired-normal-like tissues to screen those differentially expressed genes, we need to understand that the cellular contaminations inside tumor samples by the surrounding paired-normal-like tissues might exist (94,95). When clinicians collected the tumor block, it is difficult to remove those infiltrated normal-like cells inside the tumor before putting them into a sequencing machine. Such contaminations

can be critical sources of noise and potential confounding factors when conducting downstream analysis. Especially, for the analysis in allelic imbalance, it becomes more sensitive to cause systematic bias. For example, if we are analyzing those SNV sites which are specifically carried by those cancer cells inside the tumor tissues, those contamination reads coming from the paired-normal-like tissues will then make the allele fraction of those tumor-specific SNV sites become biased.

Cancer genomes can be characterized by numerous types of sequence changes compared with the paired-normal-like genomes, which may range in size from single base changes to large chromosomal fragments of insertions/deletions or even to whole-genome duplications, which has mainly been checked based on comparative genomic hybridizatoin (CGH). Fortunately, such sequence differences also make it possible to distinguish the sequence reads from cancer cells and paired-normal-like cells, which may be all mixed together inside a tumor tissue. Currently, several methods have been proposed for detecting copy number variations (CNVs)[2] inside a tumor based on its genomic DNA (96), which can be detected by SNP-array or by WGSeq data. In my study, SNP-arrays were used to estimate the tumor purity by comparing the tumor sample to the paired-normal-like sample from the same cancer patient (80).

After acquiring the estimated tumor purity based on SNP-array data through ASCAT software, the RNA-Seq read coverages at heterozygous sites inside tumor tissues can be carefully corrected for the downstream analysis of allele fraction for allelic imbalance studies. The detailed formula is shown below:

(1) For the SNVs as germline alteration:

$$R_{CorrectedCount} = R_{RawCount} - (R_{RawCount} + A_{RawCount}) \times \frac{(1-P)\times(1+1)}{P\times(N_R+N_A)+(1-P)\times(1+1)} \times \frac{1}{2}$$

$$A_{CorrectedCount} = A_{RawCount} - (R_{RawCount} + A_{RawCount}) \times \frac{(1-P)\times(1+1)}{P\times(N_R+N_A)+(1-P)\times(1+1)} \times \frac{1}{2}$$

(2) For the SNVs as somatic mutation:

$$R_{CorrectedCount} = R_{RawCount} - (R_{RawCount} + A_{RawCount}) \times \frac{(1-P)\times(1+1)}{P\times(N_R+N_A)+(1-P)\times(1+1)}$$

$$A_{CorrectedCount} = A_{RawCount}$$

In the above formula, N_R indicates the estimated copy number of reference allele in tumor cell. N_A indicates the estimated copy number of alternate allele in tumor cell. P indicates the estimated tumor purity in tumor tissue. R indicates the reference allele and A indicates the alternate allele for the read counts before and after the read correction steps.

Here, we have to notice that the situations and methods being used in read correction steps are different between those germline-derived variant sites and somatically derived variant sites. This is because those somatically derived variants are supposed to only exist in cancer cells; In contrast, those germline-derived variants are supposed to exist in both cancer cells and paired-normal-like cells.

Still, there might be few somatically derived variants which occurred in early embryo tumor developments, which may be carried by both partial cancer cells and paired-normal-like cells, but such mosaic situations will make their allelic fraction become abnormal, and it is better to firstly ignore them in the allelic imbalance analysis. For the examples of read coverages before and after the read correction step based on the above formula, the detailed data for two typical cancer driver genes which carry the heterozygous SNV somatic mutations sites can be further viewed in (S5 Table).

2 In tumors, most detected CNVs were mainly derived from somatic copy number alterations (SCNA) events.

Nevertheless, even with the tumor purity correction steps, issues about tumor heterogeneity may still exist for some somatically derived variants. As we know, even inside a pure tumor tissue, different sub-colonies of cancer cells might still exist, and those cells will carry sub-colony specific SNV sites which enable the cell to equip distinct morphological and phenotypic profiles. However, considering all the cancer cells should start from few cancer stem cells (CSCs) in the very beginning of the tumor development, the majority of the tumor-specific SNVs should still be commonly carried by every cancer cell, which should allow the allele fraction in what SNP-arrays have predicted based on SCNAs stay consistent with whole exome sequence data. In other words, those tumor sub-colony specific SNV sites may exist but stay only a few, and we have better to first ignore them in the allelic imbalance analysis also, (S2 Fig F). The allele-specific read coverages after the above discussed read correction steps can be finally viewed in (S6 Table), which will be discussed in more detail in the result part.

According to (S2 Fig A), for those somatically derived variant sites, the RNA allelic imbalances (RAIs) were identified by statistical methods after those raw RNA read coverages had been applied to the read coverage correction formula. Simultaneously, their corresponding genotypes were also referred from the SNP-array data, and this would tell us whether those RAI cases were caused by allele-specific somatic copy number alteration (SCNA) or by Allelic Imbalance in Transcription Efficiency (AITE), (Fig 2B)

2.2.3.1 Collecting the RNA-Seq and WESeq data of tumor and paired-normal-like tissues and peripheral blood

For each selected cancer patient, Illumina short-read sequences derived from homogenized primary tumor tissues and paired-normal-like tissues were collected. For each selected tissue aliquot, both of the exomes (WESeq) or transcriptomes (RNA-Seq) sequence libraries were aligned to GRCh38.d1.vd1 (hg38.d1.v1) following the recommended protocols of the International Cancer Genome Consortium (ICGC) (97).

Illumina short-read sequences derived from the peripheral blood were also collected. Their exomes (WESeq) were aligned to GRCh38.d1.vd1 (hg38.d1.v1) following the recommended protocols issued of the ICGC.

2.2.3.2 Collecting the RNA-Seq and WESeq data of control-normal tissues from GTEx individuals

For the control-normal group data, Illumina short-read sequences derived from the homogenized target tissues taken from the GTEx individuals were downloaded in SRA format from the Database of Genotypes and Phenotypes (dbGaP, http://www.ncbi.nlm.nih.gov/gap). The WESeq of peripheral blood and the transcriptome sequences of target tissues were converted into ".bam" files.

In order to remove potential batch effects between the ".bam" files downloaded from different databases, the GTEx data in ".bam" file format were converted to FASTQ format and then were realigned to GRCh38.d1.vd1 (hg38.d1.v1) following the protocols recommended by the ICGC. All parameters were set exactly to the same values as those used in the ".bam" files from TCGA (98–100).

To realign the WESeq reads, reads from quality-controlled ".fastq" files were aligned to the reference genome using the BWA-MEM (101,102) algorithm with the reference genome, GRCh38.d1.vd1, which included GCA_000001405.15_GRCh38_no_alt_analysis_set, Sequence Decoys (GenBank Accession GCA_000786075), and several virus sequences, as mentioned earlier. Picard-Tools (103) was used for sorting, merging, and marking duplicate reads of the ".bam "files (104). These steps were important for the elimination of persistent PCR artifacts and for preventing downstream analysis errors. These steps were consistent with the TCGA DNA-Seq analysis

pipeline (98) with more details in https://docs.gdc.cancer.gov/Data/Bioinformatics_Pipelines/DNA_Seq_Variant_Calling_Pipeli ne/.

For RNA transcriptome sequencing (RNA-Seq) reads, reads from quality-controlled ".fastq" files were aligned to the reference genome using the STAR 2 pass method (105). For this step, the reference genome was also GRCh38.d1.vd1, and the gene annotation file was GDC.h38_GENCODE _v22_GTF, which was consistent with the one used by TCGA, provided in the following URLs: https://gdc.cancer.gov/about-data/data-harmonization-and-generation/gdc-reference-files. These RNA alignment steps were also consistent with the TCGA mRNA analysis pipeline (99), with more details in https://docs.gdc.cancer.gov/Data/Bioinformatics_Pipelines/Expression_mRNA_Pipeline/.

For the read quality control of those ".bam" files, RNA-Seq and WESeq libraries from both TCGA samples and GTEx samples were input to Qualimap (106,107) software for checking the read lengths, read coverages, and mapping qualities.

2.2.3.3 Identification of genome-wide heterozygous SNV sites

GATK-HaplotypeCaller (108) was used for calling germline-derived heterozygous variants. The input ".bam" files were realigned by us, and the reference file was GRCh38.d1.vd1 with annotations from b151_GRCh38p7_common_all.vcf, provided by dbGaP. For the raw output SNV sites, GATK-VariantFiltration was used for further filtration. In the INFO and QUAL columns of VCF files, sites with "QUAL < 50," "QD < 2.0," "FS > 60.0," "MQ < 40.0," MQRankSum < −12.5," and "ReadPosRankSum < −8.0" were removed. In the FORMAT column of the VCF files, sites with "GQ < 20" and "isHomVar = 1" were removed. Finally, GATK-SelectVariants was used to selected target heterozygous SNV sites, which had high quality in the read mapping step. The final output was in VCF format.

2.2.3.4 Identifying somatic mutation SNVs in primary tumor tissues

The somatic mutation sites in tumor from each cancer patient were detected using the Mutect2 (109,110) software with tumor WESeq and paired blood WESeq as the input. The output was in VCF format with annotation, following the recommended protocols issued by ICGC (100,111).

2.2.3.5 Calculating the raw read coverages of every SNV site

The allele-specific read coverages at each SNV site were called by the GATK-ASERead counter (111) with the input from the sample tissue WESeq or the sample tissue RNA-Seq, along with the VCF file from each individual for the DNA or RNA part.

2.2.3.6 Estimation of tumor purity and detection of allelic somatic copy number alterations segments

For the tumor tissue from each cancer patient, the paired ".cel" files (based on Affymetrix SNP-array 6.0 platform (112)) from the tumor and blood tissue derived from the same individual were used for genotyping. Those ".cel" files were firstly input to pennCNV (113) to calculate the log R ratio and B Allele frequency value, which were then piped to ASCAT (80,114) for downstream tumor purity estimation and SCNA joint segmentation detection using a circular binary method (115,116). Reference SNP sites of hg38 for heterozygous detection were provided by Affymetrix.

For cross validation, our output data were compared to those ASCAT output data, which have been published in part by the TCGA portal and the COSMIC website, for the same tumor samples. The results were confirmed to be highly consistent.

2.2.3.7 Calculating the corrected allele-specific read coverages based on estimated tumor purity and SCNA data

The read coverage at every somatic mutation or germline-derived variation site from each tumor was corrected according to the tumor purity, taking into account the contamination and infiltration of the surrounding normal tissue. By utilizing the estimated tumor purity and allelic SCNA segment information reported by ASCAT (SNP-array based), we can apply the above mentioned formula to the raw read coverage data reported by the GATK-ASERead counter for each tumor, to obtain the correct read coverage.

2.2.4 Bioinformatic methods to detect RNA allelic imbalance (RAI) and allelic imbalance in transcriptional efficiency (AITE)

There are so many mutations which may be identified inside the tumor tissues, some of them might be inherited from their parents (germline-derived variants) and some of them might be derived from the random events during life after born (somatically derived variants). Because the methods to identify those germline-derived variants and somatically derived variants are quite different, the reads correction steps should also be different correspondently as the above formula has shown. For those two different types of variant sites, they are analyzed separately.

The pipelines designed in this study are going to first detect RAI sites, and then focusing in AITE effects for those somatically derived variants and germline-derived variants. For those Pan-adenocarcinoma, the customized algorithms written by R scripts were carefully designed and constructed by myself for detecting RAI and AITE in those somatically derived variant sites and germline-derived variant sites separately under the consideration of tumor purity. The detailed codes can be referred in (AppledixCode), and the visualized workflows for the programming pipelines are shown in (S2 Fig).

2.2.4.1 Detection of RAI genes by those expressed SNV site

The input requires the GATK-ASERead counter-based read coverage data at each target SNV site, which can be reported either by Mutect2 for somatic mutations or by GATK-haplotype caller for genome-wide heterozygous sites. The read coverage data from tumor tissues requires a read coverage correction step, as discussed above.

For those target SNV sites, if their RNA read coverage was lower than 20, they were removed for the threshold control of the gene expression level and enough RNA read coverage. The corrected RNA read coverage data were used for calculating the variant allele fraction (VAF) under the formula of “alternate allele read coverage / total read coverage” at each target SNV site.

Binominal statistical tests were applied to examine whether the VAF deviated from 0.5 for each target SNV site. The Benjamini and Hochberg FDR was used to remove false-positive SNVs by adjusting the p-values by considering multiple test bias, and the threshold of the adjusted p-value was set to 0.05 to indicate significance.

For each gene, I calculated how many target SNV sites have passed the threshold control. Among those SNV sites that passed threshold control, I further checked how many of them had VAF significantly difference from 0.5. Finally, RAI genes were those ones which carry more than 50% of SNV sties showing significantly biased VAF.

2.2.4.2 Detection of the AITE genes by those expressed SNV site inside RAI genes

For those somatic SNV mutation sites, WESeq and SCNA segments which overlapped with the target somatic SNV mutation sites inside the RAI gene were screened for the allelic balance at the genomic level. After this step, the remained RAI genes without genomic level allelic balance could therefore all be attributed to AITE

effect. Furthermore, if the p-value of the Z proportion test for the VAF between WESeq and RNA-Seq was lower than 0.05, then that AITE gene will be given a special color mark in (Fig 3 & S6 Fig & S7 Fig).

For genome-wide heterozygous SNV sites, I excluded all of the regions with allelic imbalanced SCNA segments and also applied binomial tests to the allele-specific read coverage of target SNV sites in paired WESeq to remove deviated VAF in advance to ensure the allelic balance at the genomic level. After this step, the remained RAI genes without genomic level allelic balance could therefore all be attributed to AITE effect.

2.2.5 Differential expression analysis (DEA)

The raw RNA read counts of the selected pan-adenocarcinoma samples output by HTseq (117) workflows were downloaded from the TCGA database. The DESeq2 (118,119) R package was used for removing the batch effects, normalizing read counts, conducting principal component analysis, detecting the differentially expressed genes, and plotting the MA plots.

2.2.6 Gene Ontology (GO) analysis

The Cytoscape (120) software was used for visualization, and the GeneMANIA (121,122) package was applied for predicting the function of the target gene sets. The parameters were set as follows: max resultant genes was 10, max resultant attributes was 5, network weighting was GO biological process based, and the chosen interaction network was set as default.

2.2.7 Statistical analysis and visualization of the data output of RAI and AITE analysis

The integration, statistical analysis, and visualization of the data output from the DEA and biased allelic examination were conducted using R, Python, and Bash shell scripts. The details of algorithms can be referred in (AppledixCode).

2.2.8 Detecting the *de novo* insertion of TEs

The WGSeq data of the tumor tissue and paired-solid-normal tissues were downloaded from the TCGA database, and I only analyzed 27 samples from those 3256 selected cancer patients, because of the issue of small available sample size for WGSeq from paired-solid-normal tissues. These files were converted back to ".fastq" format as the input to the ERVcaller (123) software. Published data were used as the control-normal data, where ERVcaller has published demonstration results that are based on the WGSeq of non-cancer individuals collected by the 1000 genome project. Based on the output results, the average counts of *de novo* TE insertion[3] events were calculated between the tumor tissue group, paired-solid-normal tissue group, and the control-normal group.

2.2.9 Overlapping the *de novo* TE insertion sites in the noncoding regions around AITE genes

R script was developed for this task, which can be referred in the (AppledixCode). Input files were ".vcf" files from different tissue types by ERVcaller, which detect the *de novo* TE insertion site. Additionally, the tables

[3] "*de novo* TE insertion" is used to describe a transposable element insertion event which is seen for the first time in a person and has not appeared in previous generations. Therefore, a *de novo* TE insertion can happen in an egg or sperm cell of a parent, in the fertilized egg soon after the egg and sperm unite, or in another type of cell during embryo development. More details can be referred from https://www.cancer.gov/publications/dictionaries/cancer-terms/def/de-novo-mutation. As for "tumor specific *de novo* TE insertion", it indicate for those transposable element insertion events which only happened inside tumor tissues but not inside other tissues such as paired-normal-like or blood tissues for a specific individual.

recording those AITE sites and detailed information are also integrated. Furthermore, these 2D relationships are visualized by using the R packages, such as "VariantAnnotation," "GenomicRanges," "rtracklayer," and "Gviz," among others. Around the target AITE genes, by comparing those TE insertion sites detected inside tumors, paired-normal-like and blood tissues based on same cancer patient, the tumor-specific *de novo* TE insertion sites can be found.

2.2.10 Calculating the TE insertion density around all expressed heterozygous genes which may or may not show AITE effect

The program was written in R, and the details of algorithms can be referred in (AppledixCode). Input files were the ".vcf" files recording TE insertion sites and those tables recording AITE genes. The investigation window for calculating the *de novo* TE insertion density was set with the range in plus and minus 0.5 Mbp around the targeted expressed heterozygous SNV sites (whether showing AITE or not) (S12 Fig B).

2.2.11 Visualizing the epigenetic profiles around the *de novo* tumor specific TE insertion sites

The customized R scripts was programmed for this task, and the codes can be accessed in (AppledixCode). Herein, those multi-genomics and epigenetics tracks are constructed inside a figure. The input files are listed in the following lines: (1) tables recording the target AITE gene positiion, (2) corresponding ".vcf" file which records the structure variant sites (Here we used *de novo* TE insertion sites reported by ERVcaller softwares, and (3) epigenetics data which includes ATAC-Seq ".bam" file and methylation array.txt file. Then, by using R packages, such as "VariantAnnotation," "GenomicRanges," "rtracklayer," and "Gviz," among others, the epigenetic signals around the AITE gene and its noncoding regions are visualized in 2 dimensions, (S13 Fig).

2.2.12 Detecting ATAC-Seq peak signals at *de novo* TE insertion sites or at the non-coding regions arounds AITE genes

Raw ATAC-Seq ".bam" files for each tumor sample was downloaded from TCGA data portal (https://portal.gdc.cancer.gov/) which were deposited by the published paper (124). The methods for normalizing those aligned ATAC-Seq reads were also referred to the published paper (125), and the matrix of normalized ATAC-seq insertion counts within the pan-cancer peak set were downloaded from (https://gdc.cancer.gov/about-data/publications/ATACseq-AWG).

Next, customized R script was programmed for the following task:

1. To detect ATAC-Seq signals at the non-coding regions around those expressed heterozygous genes:

   Each bin was set into 500bp wide, and the normalize ATAC-Seq read counts (signal) inside each bin was converted into "score". Next, a threshold for the "score" was decided as 0, and if the "score" is above the threshold, then that bin would be claimed for the existence of ATAC-Seq peak. Next, an investigation window was also set around the expressed heterozygous gene in plus/minus (+/-) 0.5 Mbp wide for calculating density of the ATAC-Seq peaks, or for detecting the existence of ATAC-Seq peak around the gene. Finally, those expressed heterozygous genes were grouped into AITE group and no_AITE group. Then, Chi-squared test was performed to test whether the occurrence of any ATAC-Seq peak arounds the gene and the AITE effect at the gene were independent or not, (S14 Fig A,B).

2. To detect ATAC-Seq signals at TE insertion sites:

   The method for setting the bin and the threshold detect the ATAC-Seq peaks were stay the same as above, but this time the investigation window was changed to be set at the TE insertion site with

plus/minus (+/-) 1000 bp wide for calculating density of the ATAC-Seq peaks, or for detecting the existence of ATAC-Seq peak around, (S14 Fig C,D).

2.2.13 Detecting positive selection signals through dN/dS methods for all the somatic mutations from the selected Pan-adenocarcinomas

The raw somatic mutation sites with annotations from all of those selected 3,256 cancer patients were downloaded from TCGA data portal (https://portal.gdc.cancer.gov/), and those downloaded.vcf files were merged and inputted into R scripts for the utilization of R package, "dndscv", which provided the methods for detecting the positive selection signals among all the somatic mutation sites. Next, the customized R script was programmed for the quality controls and for the statistically investigating the relationships between "dN/dS ratio" versus "Average of alternate allele fraction in RNA" versus "Proportion of cases showing alternate allele bias" among those COSMIC genes.

2.3 Results

2.3.1 Characterizing allelic imbalance in transcriptional efficiency of somatic mutations among selected pan-adenocarcinomas cases

Ten major types of human adenocarcinoma samples provided by The Cancer Genome Atlas (TCGA) (126,127) were used for large-scale RAI analyses and investigations into AITE (Fig 1). These carcinomas were liver hepatocellular (LIHC), stomach (STAD), colon (COAD), lung (LUAD), prostate (PRAD), kidney renal clear cell (KIRC), kidney renal papillary cell (KIRP), rectum (READ), thyroid (THCA), and uterine corpus endometrial tissue (UCEC) (S1 Table). Because most types of adenocarcinomas were defined as neoplasia tissues with mixed stromal cells and exocrine epitheliums (glandular/specialized/squamous epithelial cells) according to the "The Human Protein Atalas" website[4] being published under (62), common biological mechanisms might occur in driving cancer cell transformation to some extent. These adenocarcinomas were merged as pan-adenocarcinoma cohort in order to increase the statistical power of the study.

Using the National Institute of Genetics (NIG) supercomputer system (128,129), I performed RAI analyses based on RNA sequencing (RNA-Seq) data from 3,256 pan-adenocarcinoma samples, along with paired genomic DNA information provided by whole exome sequencing (WESeq) and SNP-array format for each specific sample from the different tissue groups (S1 Fig). In total, 182,245 somatic mutations were identified. After excluding insertions and deletions, 151,574 single-nucleotide variants (SNV) of somatic mutations remained. Of these, 31,816 SNVs demonstrated sufficient RNA read coverage depth for downstream analysis.

I established RAI analysis pipelines to investigate somatic SNV mutations detected by the somatic SNV caller Mutect2 (109,110,130), and it is especially for those SNVs occurring in cancer-associated genes that were reported by the COSMIC consortium (38,131), (S2 Fig A,B). For each tumor tissue, purity correction steps were conducted in advance, to avoid the issue of read contamination (S2 Fig F), and the validity of the corrected read coverage was confirmed by the consistency between the estimated mutant allele fraction (MAF) in WESeq data and the estimated allele-specific copy number in SNP-array data (80,132,133) (S5 Table). This step is critical to estimate the correct MAF in RNA-Seq and WESeq data in cancer tissues (81,130,134,135). Using binomial statistical tests, SNV sites whose alternate allele fractions were biased from 0.5 were screened from the

[4] More details can be referred from the following "The Human Protein Atalas" website: https://www.proteinatlas.org/humanproteome/single+cell+type

transcriptome data to identify the existence of allelic imbalance (73,76,86,136,137), RAI effect (Fig 1A). The genomic allele-specific copy number information derived from WESeq and SNP-arrays were used to check whether those RAI sites overlapped with the indel regions with SCNA (72,82). If this was not the case, those RAI sites could be attributed to imbalanced transcriptional efficiency between alleles, AITE effect (Fig 1B). After going through my whole bioinformatic pipelines (S2 Fig A,B,D,E), the statistical data about the genome-wide RAI prevalence among the most frequently mutated genes were collected, and the cases that showed allelic imbalance likely through transcriptional efficiency were identified (Table 1).

The landscape of the mutation frequency of each gene, and the counts of cases with somatic SNVs and their expressed SNVs were visualized for the selected pan-adenocarcinoma cases (Fig S4). For each gene, only those RAI cases without the allele-specific SCNA events were kept and characterized as AITE cases. The detailed case counts under different conditions during each processing step were shown in (Fig 2A,B,C,D).

2.3.2 Most RNA allelic imbalance events were mostly derived from the regulations at transcriptional efficiency rather than gene dosage effect

Changes in the gene dosage at the DNA level might lead to changes in gene expression levels, so allelic expression level changes could be due to the allele-specific alterations in genomic copy number where SCNAs exist (72,82). For each gene, I further calculated the proportion of genes with existing allelic imbalances in exonic DNA dosage caused by SCNA events. Surprisingly, only a few genes, such as *TP53*, *KEAP1*, and *SMAD4*, showed a relatively high proportion of allele-specific SCNA events. *TP53* had 88%, *KEAP* had 68.2%, and *SMAD4* had 67.5% (Fig 2B), among which more than 90% of the SCNA events finally led to loss of heterozygosity (LOH). In contrast, most of the other genes carrying mutations did not show allele-specific SCNA overlapping with the corresponding mutation site, (Fig 2E). Therefore, the remaining RAI cases were attributed to AITE (Fig 2C,D,E).

2.3.3 The distribution of bias sides is not random

For each gene, after only RAI cases being caused by AITE were kept, only nonsynonymous mutation carriers were further remained for the allelic imbalance bias sides analysis, because they are the specific mutation type which is supposed to cause functional changes to the mutant protein, (Fig 2D).

It could be supposed that if the choice of bias sides is pure random for each cancer-associated genes, then we should observe the gene counts showing the bell shape as a normal distribution, where most genes tend to show the proportion of cases with alternate allele bias side arounds fifty percent. However, such middle peak normal distribution was not observed; Instead, two tail peaks were observed at the extreme values for the case proportion distribution in (Fig 2F).

This type of distribution means that most genes exist obvious and quite consistent bias sides for their nonsynonymous mutation carriers. Either ninety percent of the mutation carriers show the alternate allele bias side (preferring alternate/mutant allele), or only ten percent of the mutation carriers show the alternate allele bias side (preferring reference/normal allele) for the majority of the genes. After excluding those genes whose numbers of AITE cases are small (less than or equal to five), such distribution pattern with two extreme tails can still be observed (S10 Fig A).

2.3.4 Evidence of AITE in well-known cancer driver mutations

I ranked the genes according to the frequency of somatic mutations among the selected pan-adenocarcinoma cases. Among the cancer-associated genes with frequent somatic mutation events, some showed very low expression in the tumor tissues, according to differential expression analysis (DEA) using the FPKM method and DESeq2 software (119) (S3 Fig A). For genes such as *MUC16*, *LRP1B*, *CSMD3* and some other

genes, which belong to family genes, may exist mapping error for their corresponding high somatic mutation frequency, were excluded (138). Also, those genes with low expression levels were also excluded, because genes without producing significant protein amount may be difficult to conduct physiological effects in cancer cells. Finally, the remained genes which carry expressed heterozygous SNV mutations can then be available for investigating their allelic imbalance, (S4 Fig).

We can easily find out the representative cancer driver genes from those top frequent ones, including *TP53* (139), *CTNNB1* (140), *PIK3CA* (141), *BRAF* (142), *KRAS* (143), *ATM* (144) and *PTEN* (145). In principle, since these genes are well-known cancer drivers, their functions should be directly related to the control of cell proliferation. Most of them were observed in our data to have frequent RAI among their mutation carriers, (Fig 2A,E & S4 Fig E). After their genotypes were confirmed, a high frequency in acquisition of AITE were also found, (Fig 2C), and this is especially for those missense mutation carriers which show quite consistent bias towards alternate (mutant) allele, (Fig 2D, F).

TP53, which has been investigated in many studies, is a well-known TSG, which functions in cell growth arrest and apoptosis. Expression of nonsynonymous (missense) SNV mutations of *TP53* tend to be biased toward the mutant allele (Welch's t-test, $p = 4.86 \times 10^{-3}$) (Fig 3A & S6 Fig).

KRAS, which encodes a protein that is a member of the small GTPase superfamily, is part of the RAS/MAPK pathway and is a well-known proto-oncogene. Expression of nonsynonymous SNV mutations of *KRAS* also showed a bias toward the mutant allele in statistically significant ways (Welch's t-test, $p = 3.20 \times 10^{-12}$) (Fig 3B & S7 Fig).

CTNNB1 and *PTEN,* are known to function as a proto-oncogene and a tumor suppression gene, respectively. *CTNNB1* encodes a protein that is involved in adherens junctions (AJs), which are necessary for the creation and maintenance of epithelial cell layers through the regulation of cell growth and adhesion. *PTEN* is responsible for negatively regulating the Akt/PKB signaling pathway. Nonsynonymous SNV mutations in these two genes also show a significantly higher MAF in RNA-Seq than that in WESeq (Welch's t-test, $p = 9.75 \times 10^{-4}$ for *CTNNB1*; $p = 3.80 \times 10^{-3}$ for *PTEN*) (S5 Fig A).

Synonymous mutations in *KRAS* and *CTNNB1* were not observed to be significantly different between the average MAF in RNA-Seq and that in WESeq (Welch's t-test, $p = 0.983$ for *KRAS*; $p = 0.978$ for *CTNNB1*) (S5 Fig C). Nonsense mutations that produced an early stop of transcription in *TP53* and *PTEN* showed a bias toward the normal allele probably due to nonsense decay (Welch's t-test, $p = 6.66 \times 10^{-3}$ for *TP53*; $p = 6.51 \times 10^{-6}$ for *PTEN*) (S5 Fig C)(146).

2.3.5 AITE enhanced the expression of alternate alleles without greatly increasing the total expressional level of cancer driver genes

I performed the DEA analysis for pan-adenocarcinoma and observed no significant difference in the total expression level between the tumors and paired-solid-normal tissues for several cancer driver genes, including *BRAF*, *ATM*, and *ARID1A* (147,148), (S3 Fig B). Although there were significant differences for genes such as *TP53* and other several tumor suppressor genes, the total expression level did not decrease but increased instead (S3 Fig B). This observation led us to consider allele-specific FPKM for those somatic mutation carriers.

When a missense mutation occurs in one allele of a heterozygous diploid carrier, the expression level of the normal protein is expected to decrease by 50%, if both alleles are transcribed at the same rate without AITE. Then, the "assumed allele-specific FPKM" is supposed to be half of the total FPKM, (S6 Fig E & S7 Fig E). However, when taking into account AITE effects, the real "allele-specific FPKM" for those somatic mutation

carriers should be calculated by multiplying its alternate allele fraction with its corresponding total FPKM of the gene for each AITE case. According to our observations in those cancer driver genes showing AITE effects, the absolute transcriptional amount of the tumor suppressor protein from the normal allele is significantly decreased, and that the absolute transcriptional amount of the oncogenic protein from the mutant allele is significantly increased. In other words, the "allele-specific FPKM" of cancer driver genes were changed under the AITE effect with bias toward the alternate allele. Here one example for the typical TSG (*TP53*) and typical oncogene (*KRAS*) were listed respectively, (Fig 3A,B).

2.3.6 AITE in non-cancer driver genes tended to bias the expression toward the normal allele inside TME

There are some other cancer-associated genes reported in COSMIC, such as *COL3A1* (149), *CDH11* (150), *MYH11* (151), *PDGFRB* (152), *TNC* (153), and *PTPRC* (154), whose functions are not directly related to driving carcinogenesis, but also show a high frequency of acquisition of somatic mutations among pan-adenocarcinoma patients, have expression inside tumors, which have been confirmed with single-cells RNA-seq data, (S14 Fig D). Their expression implies that they should still have physiological roles for TME (bulk tumor tissue).

The nonsynonymous SNV mutations of these genes were significantly biased toward the normal allele, with the MAF in RNA-Seq significantly lower than that in WESeq (Welch' t-test, $p = 2.33 \times 10^{-20}$ for *COL3A1*; $p = 1.52 \times 10^{-5}$ for *CDH11*; $p = 5.82 \times 10^{-23}$ for *MYH11*; $p = 2.59 \times 10^{-5}$ for *PDGFRB*; $p = 4.41 \times 10^{-3}$ for *TNC*; $p = 1.01 \times 10^{-12}$ for *PTPRC*). Additionally, the MAF in RNA-Seq approached zero in most cases in those genes, (S5 Fig B).

Similarly, I calculated the "allele-specific FPKM," which has taken account of the AITE effect for each case. For *COL3A1* and *PDGFRB*, which act as typical passenger genes. It was observed that the AITE increased the ratio of the normal allele and reduced that of the mutant allele, so the absolute expression level of the normal allele could still be maintained at a normal or even higher level for the TME (Fig 3C,D). However, according to scRNA-Seq data, most of these non-cancer driver genes are supposed not to be expressed inside epithelial cells, and the cancer cells are mainly derived from those epithelial cells, so there exists the possibility that the allelic imbalance among those non-cancer driver inside TME might be caused by the various expression levels of different cell types, where the mutant allele cannot be expressed by mutated epithelial cells as much as the surrounding non-malignant cells do inside TME (S16 Fig D). Still, this phenomenon is supposed to guarantee the normal functioning of the passenger genes and to benefit the survival of the cancer cells by guaranteeing the support from TASCs inside TME.

2.3.7 Genes set enrichment analysis for groups of genes showing the same bias in AITE

There are more than 500 cancer-associated genes listed in COSMIC that were carried by only a few patients but have also been observed to exhibit AITE. Gene set enrichment analysis (GSEA) was performed for those genes with AITE biased toward the same allele side, either mutant (alternate) or wild (reference) type (S8 Fig A). For the AITE genes showing a bias toward the mutant allele, specific Gene Ontology (GO) terms with a significantly high Q-value were enriched including fibroblast growth factor receptor signaling pathway (GO:0008543, Q-value = 1.9×10^{-17}), epidermal growth factor receptor signaling pathway (GO:0007173, Q-value = 3.0×10^{-14}), and regulation of cellular response to stress (GO:0080135, Q-value = 1.9×10^{-17}) (S8 Fig

C). Most of these functions overlap with the functions of the well-known oncogenes and TSGs implying that functional dysregulation of cell growth and cell stress response are preferred by the tumor tissue.

Among several COSMIC-identified cancer-associated genes, there are some non-cancer driver genes that act like passenger genes and have also been demonstrated to exhibit normal allele bias tendency for AITE. Their functions are not believed to be directly connected to the triggering of carcinogenesis, but their expression might still be critical for the survival of the cancer cell inside the TME (bulk tumor tissue). Based on these premises, such genes with nonsynonymous mutations are believed to present normal allele side bias for AITE and to be enriched in carcinogenesis owing to their unrelated biological effects. For the group of AITE genes mainly showing normal allele bias, the specific GO terms with significantly high Q-value included positive regulation of cell motility (GO:2000147, Q-value = 5.8×10^{-13}), positive regulation of cell migration (GO:0030335, Q-value = 5.8×10^{-13}), regulation of leukocyte activation (GO:0002694, Q-value = 3.5×10^{-11}), regulation of lymphocyte activation (GO:0051249, Q-value = 7.6×10^{-11}), T-cell activation (GO:0042110, Q-value = 6.3×10^{-9}), blood vessel development (GO:0001568, Q-value = 3.1×10^{-8}), and angiogenesis (GO:0001525, Q-value = 4.0×10^{-8}) (S8 Fig B). These results imply that preserving the normal functions of cell movement, immunity, and blood supply are preferred by the TME, and such bias seems to be caused by the higher expression from those non-epithelial cells, such as stromal cells, fibroblasts, endothelial cells, immune cells, blood cells, and so on.

2.3.8 AITE dominant sides act as indicators for cancer drivers which is consistent with mutation frequency distribution patterns and positive selection signals of each gene

Among those cancer-associated genes, somatic mutations located inside cancer driver genes are known to show positive selections, and those functional selection events will finally produce a few peaks at specific exon positions that have a relatively higher chance of accumulating somatic mutations among tumors, where the strong positive selection signals can be detected by high dN/dS ratios inside the amino acids. In contrast, the mutations located inside passenger genes, the positive selections cannot be observed at their exon positions; additionally, frequency of somatic mutation at SNV positions appears to be evenly distributed among tumors along the gene, where the positive selection signals are hard to be detected by their relatively low dN/dS ratio inside the amino acids, the details of those data were shown in (S9 Fig D).

For example, *TP53* is considered as the typical tumor suppressor gene, for which the loss of function mutations in specific motif mainly defects transactivation capacity of p53 protein. Those positions concentrate in a specific region and act as a key factor for the selection of missense mutations. Therefore, we can observe several selection peaks inside this specific region, which leads to a relatively higher chance in the accumulation of somatic mutations in the tumors and also higher dN/dS ratio at those SNV sites and also *TP53* gene itself, (Fig 3A). Besides, if we examine the AITE bias sides among its somatic mis-sense mutation carrier cases, the "proportion of cases showing alternate allele bias" goes to 95% which reveals the consistent tendency to express mutant alleles (red side) for all the cases, (Fig 4).

For another example, *KRAS* is considered as a typical proto-oncogene, for which the gain of function mutations in codon 12, 13, or 61 among the ras genes (H-ras, K-ras and N-ras) convert these genes into active oncogenes. These few positions are acting as a key factor for the selection of missense mutations. Thus, we can observe several selection peaks at these few specific sites, which also means the relatively higher chance in accumulating somatic mutations among tumors, (Fig 3B). Examination of their AITE bias sides reveals a

tendency to express mutant alleles (red side) with "proportion of cases showing alternate allele bias" going to 97%, (Fig 4).

On the contrary, *COL3A1* is considered as non-cancer driver gene, which encodes a fibrillar collagen which is found in extensible connective tissues, belongs to a passenger gene. No position acts as a key factor for the positive selection of missense mutations inside tumors. Therefore, we cannot observe any selection peak along its exonic regions, which indicates that any site is randomly accumulating somatic mutations among tumors, (Fig 3C). An examination of their AITE bias sides reveals that there is a tendency to express the reference allele (orange side) with "proportion of cases showing alternate allele bias" going to 0%, (Fig 4).

The abovementioned pattern is true for other potential cancer driver genes and non-cancer driver genes. AITE preferences can easily be observed in the reference bias side for those genes which show the relatively low dN/dS ratio. In contrast, we can easily observe the AITE preference in mutant bias side for those genes which shows the relatively high dN/dS ratio (Fig 4A), if the available cases number is high enough. For the corresponding spearman's correlation coefficients and p-values, they were provided in (S10 Fig).

Furthermore, if we calculate the "average of alternate allele fraction in RNA" among the somatic missense mutation carrier cases of each gene, those genes with high dN/dS ratio, such as *TP53, KRAS, PIK3CA, CTNNB1, RHOA, BRAF* and others, show the relative high value. In contrast, for those genes with low dN/dS ratio, such as *COL3A1, MYH11, PTPRC, FBLN2, PRF1, ACKR3, TNC* and others, they show the relative low value in "average of alternate allele fraction in RNA", (Fig 4B).

Considering the "proportion of cases showing alternate allele bias" and "average of alternate allele fraction in RNA" also present clear positive correlation (Fig 4C), If we can simultaneously compare the "dN/dS ratio" to "proportion of cases showing alternate allele bias" and "average of alternate allele fraction in RNA" in 3D plot, then the positive correlations among those values can be observed, if the available cases number (sample size) is in enough amount, (S11 Fig).

2.3.9 AITE occurred more frequently in tumor tissues than in paired-normal-like tissues among genome-wide heterozygous genes

Besides those somatic mutation sites which have been discussed already, I am also going to investigate whether AITE effect could also be observed in other heterozygous genes in genome-wide (S2 Fig A,C,D,E), and to ask whether AITE effect could exist more in cancer genome as a tumor specific abnormal situation.

Firstly, the expressed genome-wide heterozygous genes were counted for each sample. And then, among those genes, the AITE genes were identified and counted for each sample. In order to calculate the AITE occurring rate for each sample genome, the count of AITE genes (Fig 5B) is divided by the count of expressed heterozygous genes (Fig 5A) in genome-wide. Finally, all the samples from different tissue types were calculated and differently grouped for calculating the average in (Fig 5C).

Different sample genomes existed the different AITE occurring rate. In average, the tumor tissues showed significantly higher rates in AITE occurrence than the paired-normal-like (paired-solid-normal) tissues did among those Pan-adenocarcinoma patients (Welch' t-test, $p < 2.2 \times 10^{-16}$). Also, the paired-normal-like tissues from patients showed a significantly higher rate in AITE occurrence than the control-normal tissues from non-cancer-affected individual patients did (Welch' t-test, $p < 2.2 \times 10^{-16}$). One-way Analysis of Variance and Tukey multiple pairwise comparisons also showed the significant differences of AITE occurring rates among those three groups of different tissue types (One-way ANOVA, $p < 2.2 \times 10^{-16}$). Together, these results imply

that the occurrence of AITE is not a normal situation, but it is a biological misregulation specifically occurring in tumors, (Fig 5C).

2.3.10 *De novo* TE insertion sites were more detectable in tumors than normal tissues

It has been reported that the insertion of transposable elements (TEs) may change the expression profiles of nearby genes (155,156), (S12 Fig A). In cell lines, haplotype structural variants caused by the insertion of TEs have been demonstrated to cause chromatin loop alterations in the allelic interactions of promoters and enhancers, causing allele-specific changes in the expression of nearby genes (157). Such biological profiles have not yet been fully elucidated in human cancer studies. Since genome-wide elevation of AITE effects in tumors was observed, those allelic imbalance sites were believed to be associated with *de novo* TE insertion events.

Among those selected 3256 cases, 152 cases simultaneously provide tumor tissues and paired-normal-like tissues, and only 27 of them further provide analyzable whole genome sequence (WGSeq) from both tumor tissues and paired-solid-normal tissues for ERVcaller. After inputting WGSeq data into ERVcaller, I have detected a higher frequency of TE insertion events in tumor tissues than that in paired-normal-like tissues ($n = 27$, Welch' t-test, $p = 0.0456$, but not so significant) (Fig 6A), where the control-normal groups from 1000 genome samples exhibit the even lower frequency (123).

2.3.11 *De novo* TE insertions were detected with higher frequency around AITE genes

I compared the average in TE insertion counts (density) around the AITE SNV sites to that around the no_AITE SNV sites for all the available 18 cancer patients (based on those previously available 27 cancer patients, and 5 samples were excluded because of the non-termination of the software running), which was shown in (Fig 6B). Subsequently, it was able to elucidate that the average in TE insertion counts around the AITE genes showed the highest value (the green bar) in tumor tissue for many cancer patient cases. This implied the pattern that the positions around AITE genes tended to get more *de novo* TE insertions inside tumor tissues.

2.3.12 Averages in counts of *de novo* TE insertion sites were suggested to be positively associated with genome-wide AITE occurring rate among those cancer patients

The average count of *de novo* TE insertion sites and genome-wide AITE occurrence rate appear to show a positive association (Fig 6C). In other words, for a specific cancer patient, if there are more *de novo* TE insertion sites exist in its tissue, then the more AITE events are supposed to be discovered in the genome-wide heterozygous gene. However, such positive correlation only indicates a considerably low R square value, and the clearer evidences in the direct association between the *de novo* TE insertion sites and the AITE occurrence are still required.

2.3.13 Tumor-specific insertions of *de novo* transposable elements were detected around those somatic mutation sites with AITE with specific epigenetic signals

In the vicinity of the genes carrying somatic mutations presenting the AITE effect, some of the *de novo* TE insertion sites were observed to be overlapped with the open chromatin regions where the peak signals from

ATAC-Seq data were detected, (S14 Fig C,D). Also, for those expressed heterozygous genes, the occurrence of AITE effect and the detected ATAC peaks around the non-coding regions are significantly related to each other under the Chi-square test (P=0.00475 with odds-ratio = 1.07), (S14 Fig B).

The multi-omics data from TCGA-44-6146 was obtained for the demonstration of tumor-specific *de novo* TE insertion locating on open chromatin region, (S13 Fig). Examination of *USP6NL* revealed that the downstream of the gene showed higher ATAC-Seq signals than the upstream of the gene (left side), and the *de novo* TE insertion sites existed more toward the downstream of the gene (right side). Additionally, intergene comparisons of *USP6NL* and *B4GAT1* revealed that *B4GAT1* acquired more *de novo* TE insertions sites than USP6NL did within the same range size (1 Mb) of the investigation window, where *B4GAT1* also showed more intensive ATAC-Seq signals than *USP6NL* in corresponding.

2.4 Discussions

This study mainly investigated the heterozygous somatic mutations sites. Those expressed somatically mutated SNV sites inside cancer-associated genes show the allelic imbalance in RNA (RAI), and quite high proportions of them (around 80% in average) are caused by allelic imbalance in transcriptional efficiency (AITE) rather than by allele-specific somatic copy number alterations (SCNAs). This suggests that there may exist disturbances at those putative cis-regulatory elements which potentially drive the AITE effects.

As for the biological roles in facilitating cancers, AITE effect is found to enhance the mutant allele expression level for those functional missense mutations inside cancer driver genes, and maintaining the normal allele expression level for those functional missense mutations inside non-cancer driver genes (most of them can be passenger genes) for the bulk tumor tissue. The bias sides of AITE genes are consistent with the distribution patterns of somatic mutation frequencies inside each gene, which further indicate whether a gene exists positive functional selections with few peaks in frequency by referring to those dN/dS ratio data.

In fact, in average CGC genes have already shown the relatively higher dN/dS ratios than those non-CGC genes, but if you focus on those AITE genes with normal allele bias sides, their dN/dS ratios are still relatively much lower than those driver ones with mutant allele bias sides). According to this, if a gene exists positive functional selections inside tumors (with higher dN/dS ratio), it usually belongs to cancer driver genes, and AITE of that gene will show the bias towards mutant allele also. In contrast, if a gene does not show any obvious positive functional selection inside tumors (with lower dN/dS ratio), it usually belongs to non-cancer driver genes, and AITE of that gene tends to show the bias towards normal allele.

According to the GO results, for expressed genes carrying somatic mutations with AITE bias sides towards the alternate (mutant) allele, whose malfunctions can be suggested to facilitate cancer progression. This phenomenon is explained to either decrease the expressional proportion of normal protein for tumor suppressor genes or increase the expressional proportion of mutant protein for oncogenes. On the contrary, for those genes with AITE bias sides towards the reference (normal) allele, their functions are not supposed to be associated with cancer progression directly, but their expression can still be important for the precancerous cells to survive, so that is why the normal alleles are preferred inside TME**.** This phenomenon is explained to guarantee the enough expression of normal protein to maintain the necessary physiological functions provided by those passenger genes for the TME (bulk tumor tissue). Overall, the patterns in AITE bias sides give suggestions to the roles of genes during disease development.

It has been reported that those *de novo* insertions of transposable elements (TEs) may change the expression profiles of nearby genes (155). For example, the expression of human endogenous retroviruses (HERVs) in normal tissues is controlled by epigenetic mechanisms (158,159). Recently, abnormal reactivations

of HERV transcription have also been reported in various types of cancers (160–164). The genome-wide elevation of HERV expressions inside tumors caused by abnormal epigenetic regulations is further presumed to increase *de novo* TE insertion events, in which the newly inserted HERVs or other types of TEs are believed to associate with disturbing the expressions of adjacent genes in cis-acting ways. Also, in cell lines, haplotype specific structural variants caused by the TE insertions have been demonstrated to induce the allele-specific chromatin loop alterations for the promoters and enhancers interactions, leading to allele-specific expressional changes in nearby genes (157). However, such biological profiles have not yet been fully elucidated in human cancer studies. Since genome-wide elevation of AITE effects in tumors was observed in this study, those allelic imbalance sites were believed to be associated with the elevation of *de novo* TE insertion events in tumors.

My data has confirmed that the tumors show more *de novo* TE insertion events than the paired-normal-like tissues do. Although, currently only several tumor-specific *de novo* TE insertions can be found around those cancer driver genes which carry heterozygous somatic mutations with AITE effect. Around those AITE genes the density of *de novo* TE insertions seems to be increased. Moreover, the density of *de novo* TE insertions is in positive correlation with the genome-wide AITE occurring rate. Therefore, above evidences suggest that *de novo* TE insertions might associate with the disturbances in putative cis-regulatory elements for driving AITE effect inside tumors. It's worth noting that the causal relationship still needs to be validated inside the animal models or cell lines, and my current data analysis is still difficult to tell whether the AITE effects happen first, or the *de novo* TE insertions happen first, or the epigenetic modification at insertion sites (open chromatin signals) happen first.

2.5 Others

2.5.1 Accession numbers

For those controlled data being provided by dbGaP, the data accession was approved by TCGA under the “Project #18492: Investigating the effect of allele-specific-expressing profile for human diseases” and “Project #15126: Systematic identification of reactivated human endogenous retroviruses in cancers.” All of the sequence data from cancer patients were downloaded from The Cancer Genome Atlas (TCGA) (phs000178.v11.p8.c1), and the sequence data from non-cancer individuals were downloaded from Genotype-Tissue Expression (GTEx) (phs000424.v8.p2.c1) for general research use.

For those controlled data being provided by ICGC, the data accession was approved by Data Access Compliance Office (DACO) under the “Project #DACO-7309”. Those sequence data from cancer patients were downloaded from ICGC as supplements materials.

2.5.2 Code Availability

The Bash script and R code used in this study is available in appendix part. Furthermore, my personal online Github repository can be viewed under the proper request by sending email to xji6486@gmail.com. Those uploaded codes can subsequently be accessed from following URL: https://github.com/xji6486. Or they can be referred in the Appendix C.

3 CHAPTER THREE: ADDITIONAL ANALYSIS BY RIBOSOME PROFILING AND SINGLE-CELLS DATA

The previous chapter has demonstrated that the allele-specific regulations at the transcriptional level may be associated with the facilitation in cancer progressions of TME (bulk tumor tissue). This chapter is further going to check whether the regulations in translation levels buffer the AITE effect or not, (Fig 7A).

On the other hand, the previous chapter has mainly analyzed the bulk tissue data from TCGA database. Even though the purity correction steps have been involved to remove those sequence reads from infiltrated normal-like cells inside the TME (bulk tumor tissue), the different transcriptional levels between different types of cancer cells and normal-like cells may still confound the RNA read coverage correction steps for each gene. Furthermore, the tumor tissue after purity correction steps may still exist transcriptional heterogeneity among those remaining cancer cells. Therefore, it is critical to confirm the intra-tissue heterogeneity of cancer-associated gene expression profiles by single-cells RNA-Seqs from both tumor tissue and normal tissue.

Besides, the AITE effect observed among cancer-associated genes inside bulk tissues was believed to be caused by cis-regulatory elements (79,165), which may or may not lead to the non-stochastic process of allele-specific expression (ASE) in single-cells. In fact, TCGA provides the traditional bulk sequencing data which is unable to tell whether the specific gene exists the effect of mono-allelic expression inside each cell, because those reads come from each single-cell are mixed together, which can only show the allele ratio of those mixed reads from multiple cells. Therefore, only single-cell sequencing data alone can distinguish whether the AITE of a specific gene is caused by “non-stochastic allele-specific expression (non-stochastic ASE)” or by “allele biased expression (ABE)” from the pool of single-cells inside a specific bulk tissue sample. This point can be important for revealing the actual mechanisms of AITE effects in higher resolution, (Fig 8E). Currently, there is a paper showing that each single-cell has been observed with a stochastic process to choose to express only one allele or only the other allele at one time for each specific gene, which is called as mono-allelic expression effect inside normal fibroblasts. Although there are only a very minority of those single-cells can simultaneously express both alleles for the most of the genes, exceptions still exist in those genes with extremely high expressions, which can be observed with bi-allelic expression effect inside most single-cells (166). In order to know how those cancer driver genes or non-cancer driver genes express inside each single-cell inside tumors, the higher read coverages of the single-cells data which is available for statistically analyzing the allele-specific transcriptional signals is required.

3.1 Materials and Workflows

3.1.1 Ribo-Seq in cell-line Huh-7.5.1

3.1.1.1 Data collections

The cell-line Huh-7.5.1 was selected for the investigation in translational regulation. Their WGSeq, RNA-seq and Ribo-Seq are provided from different published papers respectively. SRA-Toolkit (167) was used

to download those sequence data which were deposited on dbGap, and those file are converted into ".fastq" format for downstream analysis.

The HuH-7 cell-line was established in 1982 from a well differentiated hepatocyte derived cellular carcinoma cell line that was originally taken from a liver tumor in a 57-year-old Japanese male. For downloading RNA-seq and Ribo-Seq sequences, the accession number provided by dbGap is PRJNA560349 (168). For downloading WGSeq sequences, the accession number provided by dbGap is PRJDB7928 (169).

The available samples for cell-line Huh-7.5.1 are six totally from the database. There are 3 replicates for the control Huh-7.5.1 (SRR9971641, SRR9971640, SRR9971639), and 3 replicates for the hepatitis B virus (HBV) infected Huh-7.5.1 (SRR9971638, SRR9971637, SRR9971636). And all of those six samples are provided with RNA-seq and paired Ribo-Seq, which is described in details in the paper (168) where I downloaded the data.

3.1.1.2 Read alignments for WGSeq, RNA-seq, and Ribo-Seq

3. Whole genome sequence data (WGSeq):

Trimmomatic-0.36 was used for trimming the adapters and low quality bases of the reads from the ".fastq" files. Then, the trimmed reads in ".fastq" files were aligned to GRCh38.d1.vd1.fa by bwa-mem (bwakit-0.7.15_x64-linux). Samtools and picard-2.6.0 were used to mark duplicated reads, to sort the reads and to index the ".bam" files.

4. Whole Transcriptome sequence data (RNA-seq):

Trimmomatic-0.36 was used for trimming the adapters and low quality bases of the reads from the ".fastq" files. Then, the trimmed reads in ".fastq" files were aligned to GRCh38.d1.vd1.fa by STAR (STAR_2.4.2a) with parameters: --runThreadN 30 --outFilterMultimapScoreRange 1 --outFilterMultimapNmax 20 --outFilterMismatchNmax 10 --alignIntronMax 500000 --alignMatesGapMax 1000000 --sjdbScore 2 --alignSJDBoverhangMin 1 --genomeLoad NoSharedMemory --readFilesCommand zcat --outFilterMatchNminOverLread 0.33 --outFilterScoreMinOverLread 0.33 --sjdbOverhang 100 --outSAMstrandField intronMotif --outSAMtype None --outSAMmode None. Finally, samtools and picard-2.6.0 were used to mark duplicated reads, to sort the reads and to index the ".bam" files.

5. Ribosome profiling - polysomal RNA sequence data (Ribo-Seq):

The software fastx_clipper and fastx_trimmer were used with parameter -Q33 to trim the adapters and low quality bases of the reads from the ".fastq" files. Then, the trimmed reads in ".fastq" files were aligned to rRNA sequences by bowtie to remove those rRNA reads with parameter --seedlen=23. Next, those unaligned reads which show non-rRNA reads are further aligned to GRCh38.d1.vd1.fa by STAR (STAR_2.4.2a). Finally, samtools was used to mark duplicated reads, to sort the reads and to index the ".bam" files.

3.1.1.3 Quality controls for WGSeq, RNA-seq, and Ribo-Seq

The trimmed ".fastq" files were statistically viewed by fastqc, and next the aligned ".bam" files were statistically viewed by qualimap with bamqc and rnaseq function respectively.

3.1.1.4 Call of heterozygous variants sites

Those genome-wide heterozygous SNV sites were called from WGSeq by GATK-HaplotypeCaller, and the input files include dbsnp_146.hg38.vcf, GRCh38.d1.vd1.fa, and WGSeq ".bam" files.

Then, GATK-VariantFiltration was used with the parameter: --genotype-filter-expression "GQ < 20" & --genotype-filter-expression "isHomVar == 1" & --filter-expression "QUAL < 50" "QD < 2.0" "FS > 60.0" "MQ < 2.0" "MQRankSum < -12.5" "ReadPosRankSum < -8.0".

Finally, GATK-SelectVariants was used to selected target threshold passed variant sites with the parameter: -select 'vc.isNotFiltered()' -selectExpressions 'vc.getGenotype(0).isHet()' --set-filtered-gt-to-nocall. Those variants are recorded in ".vcf" file.

3.1.1.5 Counting read coverages at target SNV sites

Firstly, GATK-SelectVariants was used to get those heterozygous SNV sites from ".vcf" file which was previously called from WGSeq. GATK-IndexFeatureFile was used to index the ".vcf" file which carries those heterozygous SNV sites.

Next, GATK-ASEReadCounter was used for calling the read coverage. The input files can be WGSeq, RNA-seq, and Ribo-Seq, and those targe position were recorded in ".vcf" file which carries those heterozygous SNV sites. The output ".table" file can be read by customized R scripts for downstream analysis including the step in the detection of AITE sites. The distribution of read coverages in WGSeq, RNA-seq, and Ribo-Seq are shown in (S15 Fig C)

3.1.1.6 Detection of AITE sites and the sites with translational regulations

The step was written in R scripts, and the input ".table" file which has recorded the allele-specific read coverages at each heterozygous SNV sites.

There is a threshold control step for the read coverages at each heterozygous SNV site, and only those reads with enough read coverage in DNA, RNA and Ribo-Seq (WGSeq > 20, RNA-seq > 10, Ribo-Seq > 0[5]) will be available. Next, Binomial test was applied to both DNA and RNA data to detect those AITE sites, where the RNA showed the alternate allele fractions being biased from 0.5 and DNA did not show the alternate allele fractions being biased from 0.5. Finally, for those AITE sites, Z-proportion test was applied to both RNA and Ribo-Seq data to detect those translation inconsistent sites, which showed the inconsistency in alternate allele fractions between RNA and Ribo-Seq data, (S15 Fig B). And those translation inconsistent sites imply the regulations in translational levels.

3.1.2 10X genomics based single-cells RNA-Seqs from normal livers and liver tumors

The single-cells RNA-Seq (scRNA-Seq) sequencing data from normal livers were downloaded from "GSE115469" (170), which provides 5 sample sites from 5 non-cancer individuals. The single-cells RNA-Seq (scRNA-Seq) sequencing data from liver tumors were downloaded from "GSE112271" (171), which provides 7 sample sites from 2 cancer patients. All of those single-cells RNA-Seq libraries were all prepared by 10X genomics platform and being sequenced by Illumina short reads sequencer.

The downloaded scRNA-Seq data were converted into ".fastq" format, and next the "cellranger aggr" software provided by 10X genomics was used to align those raw reads from adapter-trimmed.fastq files. The output ".bam" files were then imported into R packages, "Seurat" and "SeuratData", to perform the quality controls, identify the single-cell clusters, and analyze differential expression. The target single-cells cluster can also be extracted by "subset-bam" software for the sliced ".bam" file. Then, the cluster-specific heterozygous sites can be called by GATK-HaplotypeCaller (108) and GATK-SelectVariants. Finally, the allele-specific read coverages at each SNV site were called by the GATK-ASERead counter (111).

---

[5] It is fine to set the lowest read coverage threshold as 0, because if the number of reads at such SNV site goes too small, the corresponding P value cannot go small enough to be significant, and these sites will be filtered out automatically.

3.1.3 Smart-seq2 based single-cells RNA-Seqs from liver tumors and paired normal-like tissues

The single-cells RNA-Seq (scRNA-Seq) sequencing data from human livers were downloaded from "SRP273101" (172), which provides the tumors samples and paired normal-like samples from 6 liver cancer patients, respectively. All of those single-cells RNA-Seq libraries were all prepared by Smart-seq2 protocols and being sequenced by Illumina short reads sequencer (173,174).

The downloaded scRNA-Seq data were converted into ".fastq" format, and next the "STAR" software was used to align those raw reads from adapter-trimmed.fastq files. The output ".bam" files were input for the heterozygous sites calling by GATK-HaplotypeCaller (108) and GATK-SelectVariants. Finally, the allele-specific read coverages at each SNV site were called by the GATK-ASERead counter (111).

Also, the above output ".bam" files were imported into R packages, "Seurat" and "SeuratData", to perform the classification for those single-cells, to identify each single-cell clusters by their enriched gene expression, and to analyze the differential expression.

3.2 Results

3.2.1 Ribo-Seqs suggest that transcriptional regulations can be more prevalent than translational regulations for allelic imbalance

Analysis of allelic imbalance in translational efficiency compares the alternate allele fraction between the RNA-Seq and the paired Ribo-Seq, which can directly reflect the protein amount under translations. If the ratios do not change, that means the allele-specific expression of proteins are mainly regulated in transcriptional level, (Fig 7).

Among those thresholds passed heterozygous SNV sites, only a few of them show the significant inconsistent alternate allele fraction between RNA-Seq and Ribo-Seq data. For the more details, only 4.2% of those expressed heterozygous sites show the allelic imbalance in translational efficiency (Fig 7C), and only 5.5% of those AITE sites show the allelic imbalance in translational efficiency. These are the average value of the six available samples, (S15 Fig D).

3.2.2 Single-cells RNA-Seq analysis reveals the intra-tumor heterogeneity in the transcriptional profiles among cancer-associated genes

All the single-cells RNA-Seqs from different liver tissues (including tumors and normal tissues) were merged and clustered by uniform manifold approximation and projection (UMAP) for dimension reduction according to the transcriptional profile of each single-cell. Different clusters were identified by referring to their top up-regulated genes (S16 Fig B), and each cluster was manually annotated in (S16 Fig A). Through machine-learning way each single-cell was also automatically annotated in (S16 Fig E) by referring to the "Blueprint Encode" database.

According to the detailed UMAP in (S16 Fig A,L), several big groups of clusters can be overall found: (1) There is one single big group of clusters which belong to parenchymal cells (epithelial cells like), are mainly hepatocytes and cholangiocytes. This group is basically composed by "cluster 0", "cluster 3". (2) In contrast, another single big group of clusters, which belong to non-parenchymal cells (stromal cells like), are mainly hepatic stellate cells, fibroblasts, smooth muscle cells, pericytes, endothelial cells, Kupffer cells, sinusoidal cells and so on. This group is basically composed by "cluster 2", "cluster 8". (3) As for the remaining clusters which

belong to peripheral blood mononuclear cells (PBMCs), can be gathered into two big groups. One is composed by ”cluster 5”, ”cluster 12”, ”cluster 15” and ”cluster 16”, and the other one is composed by ”cluster 4” and ”cluster 6”. These single-cells gene expressions and their clusters distribution are quite consistent with the normal liver tissue data from “The Human Protein Atlas” database (62), and my manually annotated clusters are also consistent with the automatically predicted cell types of each single-cells in (S16 Fig E).

Besides, there are tumor tissues specific clusters, which are ”cluster 1”, ”cluster 17” and ”cluster 18”, (S16 Fig A,E), whose expressions profiles can be more stromal cells like, are basically similar to those endothelial cells, mesenchymal cells and myofibroblasts according to their relative positions in UMAP map, (S16 Fig A). Also, according to “The Human Protein Atlas” (62), those stromal cells specific genes were mainly expressed by hepatic stellate cells inside normal livers (175), so these tumor tissues specific clusters have a high chance to be hepatic stellate cells.

3.2.3 Some cancer-associated genes which only expressed in specific cell types inside normal tissues can become disrupted inside tumor tissues

The clusters specific (cell-type specific) expressions for those cancer driver genes and non-cancer driver genes were checked in (S16 Fig C,D). For the most well-known cancer driver genes, such as *CTNNB1, KRAS, TP53, PTEN* and others, they were found to be expressed in all clusters including parenchymal cells group, stromal cells group and PBMCs groups, and their expression can be detected in both normal livers and liver tumor tissues with certain amounts.

In contrast, there are few cancer driver genes, such as *FAT1, MET, WNK2* and others, which were found to be specifically expressed in the clusters of parenchymal cells group in normal livers. Aa for some non-cancer driver genes, such as *COL3A1, FBLN2, ACKR3, MYH11* and others, they were found to be specifically expressed in the clusters of stromal cells group in the normal livers. Also, for some other non-cancer driver genes, such as *PTPRC, PRF1, CARD11* and others, they were found to be specifically expressed in the clusters of PBMCs group in the normal livers. The above-described expression patterns in normal livers are quite consistent with the single-cell data from “The Human Protein Atlas” database (62), which only provided normal tissue data. By further comparing the same clusters from normal livers to that from liver tumor tissues, we can see the expressions of those above-mentioned genes were disrupted in some certain inside tumors, and some clusters have changed to express those genes that were supposed not to be expressed in normal situations. Therefore, more different types of cell clusters seem to be able to simultaneously express the cancer driver genes and non-cancer driver genes inside tumors, (S16 Fig C,D).

3.2.4 High-throughput droplet-based 10X-genomics single-cells RNA-Seqs benefit the allelic imbalance analysis with huge cells population

Considering different genes with different functions may present the different expression level according to different cell types inside bulk sample tissue, which may confound the existence of allelic imbalance and their actual bias sides, I am going to reconfirm the existence RNA allelic imbalance in the pseudo-bulk RNA-Seq based on the cell type specific clusters.

Firstly, I picked up the parenchymal cells specific clusters (”cluster 0”, ”cluster 3”, ”cluster 10”) and the stromal cells specific clusters (”cluster 8”, ”cluster 17”) from a cancer patient (Pt13b) and also from a non-cancer individual (P2TLH), respectively (S16 Fig F), and those specific groups of clusters can further stand for the pseudo-bulk tissues being respectively composed by cancer cells, by tumor associated by stromal cells, by epithelial cells or by stromal cells without the issues of tissue heterogeneity (Fig 8A & S16 Fig L). Among those

pseudo-bulk tissues, we can observe some expressed heterozygous genes show the bias towards the alternate allele, and some other genes show the bias towards the reference allele (S16 Fig I,J).

Especially for the pseudo-bulk RNA-Seq data being derived from cancer cell clusters, those cancer cells were found to express those non-cancer driver genes which were not supposed to be expressed in normal epithelium (Fig 8B & S16 Fig C,D). If we further check the bias sides of those expressed heterozygous genes, whose average of RNA alternate allele fractions from the "single-cells based pseudo-bulk data (this chapter)" show the consistence with that from the "bulk-tissue based population data (previous chapter-TCGA data)" (Fig 8C,D).

3.2.5 FACS-based Smart-Seq single-cells RNA-Seqs benefit the allelic imbalance analysis with higher read coverage

According to the previously published single-cell data, the stochastic choosing of only one of the two alleles to express has been observed in most of the genes with medium expression levels in single cells from bulk normal tissue (176,177). It was believed that the observed allelic imbalance in transcriptional frequency in bulk tumor tissues can be regarded as the accumulated effect of non-stochastic choosing of the alleles for expression inside each cancer cell. That is, even though the gene dosages between the two alleles remain the same at the DNA level, if there are more cancer cells that tend to choose one specific allele for expression, this non-stochastic choosing process will make the observed alternate allele fraction in the cell population, as reflected in bulk tissue, become biased away from 0.5, a phenomenon that have been called AITE in previous chapter. Such mechanisms could be further stratified by deconvolution of the cellular composition of bulk RNA-Seq based on scRNA-Seq (170,171), to obtain higher resolution inside tumor tissue.

Considering the read coverages in high-throughput droplet-based single-cells RNA-Seqs are not high enough to investigate the allelic imbalance with statistical significance. For the deep single cell RNA-seq analysis, each scRNA-Seq library were prepared in 96-wells plate with Smart-Seq2 protocol, which can provide much higher read coverage and more complete RNA molecular information for each gene, has excluded the issues of sampling loss during sequence library construction. Through comparing the heterozygous proportion of those expressed SNVs in genome-wide, "liver tumor tissue (pseudo bulk scRNA-Seq - by 10X genomics)" showed 59.3%, "liver cancer cell-line single colony (bulk RNA-Seq - by Illumina true-seq)" showed 54.0%, and "Single liver cancer cell (deep scRNA-Seq - by SmartSeq2)" showed 12% around, (S17 Fig). The lower detectable expressed heterozygous SNVs suggest the mono allelic expression inside each single cell. And such non-stochastic allele-specific expressions (ASE) are observed in single-cells for contributing allelic imbalance, whose bias sides are consistent to bulk tissues based data after considering the tissue heterogeneity, (Fig 8D,E).

3.3 Discussions

In the previous chapter, the prevalence of AITE effect inside tumor tissues is suggested for their importance in allele-specific regulations in transcriptional level for facilitating cancer progression. In this chapter, I have further confirmed that the allele-specific regulation in translational level cannot be detected for the most of AITE cases based on the paired Ribo-Seq. In the future, if the read coverages of Ribo-Seq become deeper, more events of allele-specific translational regulation will be detected. At that time, we have to notice the shift side of ratio change in Ribo-Seq. If the regulations in translation are going to increase the allelic imbalances which were observed in transcription, that means the gene prefers the allelic imbalances with enhancing effects. In contrast, if the regulations in translation are going to decrease the allelic imbalances which were observed in transcription, that means the gene dislike the allelic imbalances with buffering effects, where the translation may

have roles. However, the read coverage depths of the Ribo-Seq might affect the number of available sites as current confounders, and results may change accordingly with more sensitive ribosome profiling technology, but the proportion of the allele-specific regulations in translational level can be quite small, as (S15 Fig) has shown. Therefore, I conclude that AITE effect can still be the most critical mechanism being utilized by tumors to enhance the carcinogenic phenotype which lifts up the mutant allele expression for those heterozygous somatic mutations in cancer driver genes.

Besides, in the previous chapter, the prevalence of AITE effects inside tumors was considered in TCGA provided bulk tissues, which lacks the consideration of intra-tissue heterogeneity of the cancer-associated genes expression. Through different tissue types of single-cells RNA-Seqs, it has become possible to clarify which types of cells mainly contributed to the observed allelic imbalance inside the TME (bulk tumor tissue) for each specific gene. As for how intra-tissue heterogeneity may confound the purity correction steps in chapter 2. We have to first clarify which kind of genes may show the tissue specific expression: (1) For most of those cancer associate genes showing alternate (mutant) allele bias sides, such as *CTNNB1, KRAS, TP53, PTEN* and others, their expression can be detected in almost every cell cluster inside normal livers and other normal tissue types which were confirmed by "The Human Protein Atlas" database (62). As for those few exceptions, such as *FAT1, MET, WNK2* and others, which are only expressed in epithelial cells in normal situations. Those cells can be gastric mucus-secreting cells (stomach), squamous epithelial cells (esophagus), alveolar cells / club cells (lung), basal prostatic cells / prostatic glandular cells / club cells (prostate), thymic epithelial cells (thymus), intestinal goblet cells (rectum), endometrial ciliated cells / glandular cells / luminal cells (endometrium), distal enterocytes / enteroendocrine cells/ paneth cells (colon), proximal tubular cells / collecting duct cells (kidney), and so on. (2) In contrast, for those cancer associate genes showing reference (normal) allele bias sides, such as *COL3A1, CDH11, FBLN2, ACKR3, MYH11, PTPRC, PRF1, CARD11* and others, they were found to only be expressed either only in stromal cells or only in PBMCs inside normal livers and other normal tissue types by referring to "The Human Protein Atlas" database (62).

Based on these facts, people may argue this is why the frequencies of the mutant alleles in passenger genes were very low in bulk tissue based RNA-Seq, since the expressions of these non-cancer driver genes in epithelial cells were originally supposed to be absent in normal tissues which may confound the purity correction step in bulk tissue and lead the bias side toward the reference allele of RNA reads. To validate this viewpoint, my single-cells study has further discovered that some clusters inside tumor bulk tissues have changed to express those genes which are supposed not to be expressed inside normal bulk tissue. For example: (1) For several stromal cells specific genes, such as *COL3A1, CDH11, FBLN2, ACKR3, MYH11* and others, or PBMCs specific genes, such as *PTPRC, PRF1, CARD11* and others, the "cluster 0" and "cluster 3", which are supposed to be parenchymal cells, have changed to be able to express these non-cancer driver genes inside liver tumor tissues. Based on this evidence, we can claim that some parenchymal-like cells from tumor tissues have changed to simultaneously express cancer driver genes and non-cancer driver genes as the "cluster 0" and "cluster 3" showed. (2) On the other hand, "cluster 8" and "cluster 17" are supposed to be stromal-like cells without expressing parenchymal cells specific genes, such as *FAT1, MET, WNK2* and others, but their expression can be detected inside tumor bulk tissues. Also, "cluster 8" is supposed to be myofibroblast without expressing PBMCs specific genes, such as *PTPRC, PRF1, CARD11* and others, but their expression can be detected inside tumors. Based on this evidence, we can claim that stromal like cells can also simultaneously express most non cancer driver genes and cancer driver genes inside liver tumor tissues, and those stromal like cells belong to hepatic stellate cells in livers according to the published papers (62,175,178).

Taken together, both the abnormal parenchymal cells and the abnormal hepatic stellate cells inside liver tumors seem to be able to simultaneously express the cancer driver genes and non-cancer driver genes (passenger genes). Considering the cancer cells may develop from either of them, we may safely claim that the observed allelic imbalance bias sides actually come from the cancer cells which may reflect their preference. Furthermore, I have also referred my single-cells data back to those TCGA bulk-tissue data, and their consistency in their allelic imbalance bias sides can be found.
(175,178)

4 CHAPTER FOUR: SUMMARY

4.1 Conclusions

I have discovered and then demonstrated that the existence of AITE plays critical roles for increasing the transcriptional level of mutant alleles for well-known cancer driver genes. Such allele-specific control does not exist so much in translational level which was validated by Robo-Seq data of liver cancer cell-line, (Fig 7A & S15 Fig). Therefore, the prevalence of AITE effect inside tumor tissues is suggested for their importance in allele-specific regulations of mutated cancer driver proteins for facilitating cancer progression.

For cancer driver genes, the increased ratio of products of mutant alleles can enhance the disease phenotype without markedly altering the total expression level of the gene itself, which is suggested to allow the cell to avoid the transcriptional surveillance system (146,179). In this way, the biological effect of wildtype tumor suppressing protein will be diluted and that of the mutated proto-oncoprotein will also be enhanced, without changing the total amount of gene transcription to be detected.

Among COSMIC reported cancer-associated genes, there are still many genes that do not functionally act as disease drivers. Their biological functions are not directly related to cell proliferation, but to the maintenance of basic cell physiological functions or the extracellular environment, such as *COL3A1* and *COL1A1*, which encode the pro-alpha1 chains of type III collagen; *MYH11*, which encodes smooth muscle myosin; *PDGFRA* and *PDGFRB*, which encode platelet-derived growth factor; *PTPRC*, which regulates T-cell and B-cell antigen receptor signals; and *CDH11*, which encodes type II classical cadherin. The production of a sufficient amount of normal protein to maintain the usual biological functions is suggested to be important for cell survival inside bulk tumor tissue and they are also believed to be important for cancer cells. Among those genes, I have detected the frequent presence of AITE, but with the bias toward the normal allele. In this way, the increased ratios of products of normal alleles have guaranteed the sufficient amount of normal functional protein. Such benefit may explain the observed bias sides by those passenger or non-cancer driver genes.

From the scope of molecular biology, how these putative cis-regulatory elements allele-specifically control the transcriptional efficiency requires further investigation. Based on the current epigenetic data, it was believed that the mechanisms might be linked to allele-specific chromatin remodeling between the loose and compact forms of chromatin (180,181), since transcription factors and RNA polymerases can access the promoter more easily in allelic chromatins with an open form (182).

As for the identities of those putative cis-regulatory elements which control the AITE effect of expressed somatic mutations, some papers have claimed that the heterozygous genomic variants located in the noncoding regions may allele-specifically disrupt the normal functions of promoters or enhancers (183,184). Based on extensive exome and WGSeq data, the genetic factors underlying many human diseases, including cancers, have been widely investigated using GWASs (185). However, a majority of disease-associated variants are located within noncoding regions and do not have a clearly identified functional role yet (186). These GWAS variants are thought to be located in the expression quantitative trait locus (eQTL) of the disease driving genes (187–189), and they can be candidates in putative cis-regulatory elements (190). In our opinion, the allele-specific insertion

of *de novo* TEs can also be the candidate genomic factors to induce transcriptional changes (158,191,192) and to produce the AITE effect (193,194). Many papers have demonstrated the upregulation of the activities of TEs in tumors (156,194–197), and I have also observed many tumor-specific TE insertion events around the AITE genes (S12 Fig). Allele-specific TE insertion events are supposed to be associated with allele-specific chromatin openness and to lead to non-stochastic ASE among single cells, (S13 Fig & S14 Fig).

Finally, there is one thing having to be noticed. The numbers of *de novo* TE insertions are about 1,500 on average inside each tumor sample. However, if we consider those tumor specific ones, the reported tumor-specific (somatic) TE insertion is not quite frequent in cancer genomes according to the Pan-Cancer Analysis of Whole Genomes (PCAWG) consortium data, which only shows tens to several hundreds of events inside each tumor sample. To answer this question, we have to realize the difference between *de novo* TE insertions and tumor-specific (somatic) *de novo* TE insertions. *De novo* TE insertions specifically indicate those TE sequences which are not recorded in the human reference genome (hg38 or CHM13), so such insertion events may occur at any time point in the patient life. In other words, after their birth people may keep accumulating TE insertion events in their somatic cells, and someone may get more or less depending on each situation, so it may not be strange at all if you can detect around 1,500 *de novo* TE insertion events for an individual. As for tumor-specific (somatic) *De novo* TE insertions, which specifically indicate to those TEs sequence only existing in tumors but not existing in normal-like tissues. Therefore, we should find less events occurring in tumors stage when compared to those events accumulating in lifelong.

Same for the genome-wide AITE occurring rate, the *de novo* TE insertions counts also show relatively high values in the normal-like tissues than the control-normal (123). This result implies that the *de novo* TE insertions may also happen in the early stage so the tumor-specific (somatic) *de novo* TE insertions can be observed in relatively low proportion. This makes sense that genome-wide AITE occurring rate and *de novo* TE insertions counts are both detected in higher values in the normal-like tissues than the control-normal tissues, which may actually make cancer patients become different from those normal individuals (without cancers), have not yet been noticed in previous researches.

4.2 Future perspectives

We need large scale data from long read sequences for cancers to do the haplotype phasing work. It is important to realize how TE insertions actually act to their nearby genes inside the same DNA molecular.

Also, to investigate the priority between TE insertions and chromatin opening, we need to collect the sample for the paired ATAC-Seq and WGSeq at different time points from a specific tumor tissue. But this can be quite difficult for the tissue collection step, because this might require large amount of sample tissues, and multiple surgery times may also be considered. Instead, designing a cell experiment for validation can be the better solution.

Most important of all, this large-scale population based investigation in pan-adenocarcinoma is conducted by the short-read sequence libraries from bulk tissues, and the resolutions cannot be compared to those sequence libraries from single-cells, if we want to further investigate how the different heterogenous cell lineages have been selected inside tumor (TME) and also how AITE effect was caused by the accumulated effects from each single-cells inside the tumor. In the near future, with the improved single molecular sequencing technology, the large-scale population based single-cells sequencing data for pan-adenocarcinoma are expected to be deposited on the online data bases. And how the allele-specific *de novo* TE insertions can affect the non-stochastic allele-specific expression inside single-cells can be the next research targets from the scope of human population.

MAIN FIGURES AND TABLES

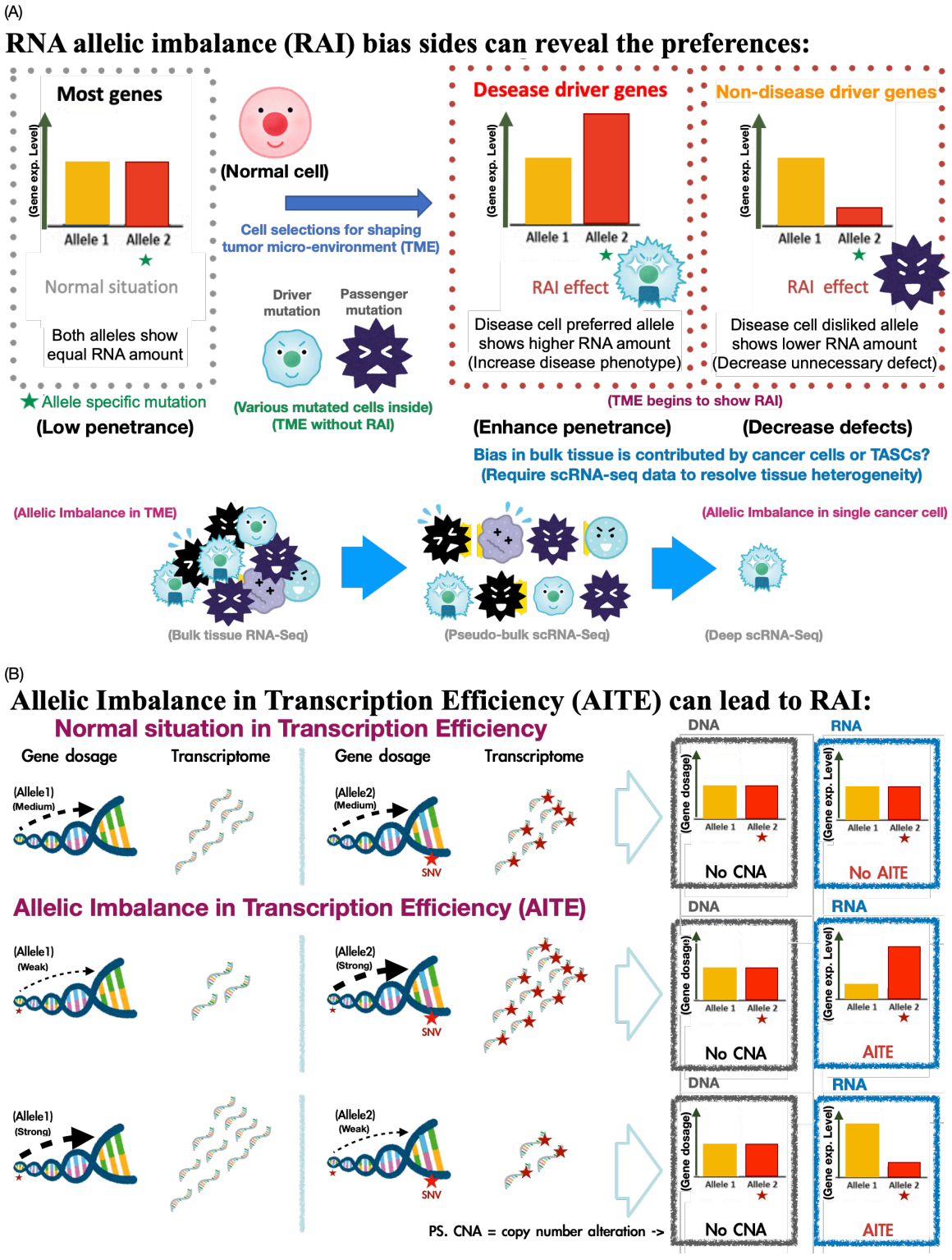


Figure 1. Allelic Imbalance in Transcriptional Efficiency can lead to RNA Allelic Imbalance effect, which is supposed to control disease penetrance.

(A) RNA Allelic Imbalance (RAI) effect indicates the observed alternate (mutant) allele fraction in RNA is biased from the supposed 0.5. More specifically, two alleles of heterozygously mutated disease driver gene can present different expression amount which can be detected as RAI effect, even though such effect can be attributed to many different regulating mechanisms in gene expression. Furthermore, the benefit allele in disease phenotype is supposed to be observed in higher amount. Take cancer for example, in either type of cancer driver genes (TSG or oncogenes), the increased expression ratios of mutant proteins are supposed to enhance disease phenotype of the tissue. But for passenger genes, the decreased expression ratios of mutant proteins are supposed to decrease unnecessary defects for the tumor micro-environment (TME). However, for the effect being observed inside tumor by bulk sequencing data, the tumor heterogeneity inside TME needs to be resolved by single-cell sequencing data.

(B) This figure illustrates a specific phenomenon called Allelic Imbalance in Transcriptional Efficiency (AITE). The above part shows the normal situations where transcriptional efficiency is balanced between the two alleles of a gene; In contrast, the below part shows the situations where transcriptional efficiency has become imbalanced, which will be called allelic imbalance in transcriptional efficiency, is abbreviated as (AITE). Among those illustrations, the left regions show the illustrations for the allele-specific gene dosage and the amount of allele-specific transcripts, and the right regions show the imaginary allele-specific reads coverages for DNA (from WESeq or WGSeq) and for RNA (from RNA-Seq), where the allelic imbalance in RNA is supposed to be discovered but it is not in DNA under the situation of AITE.

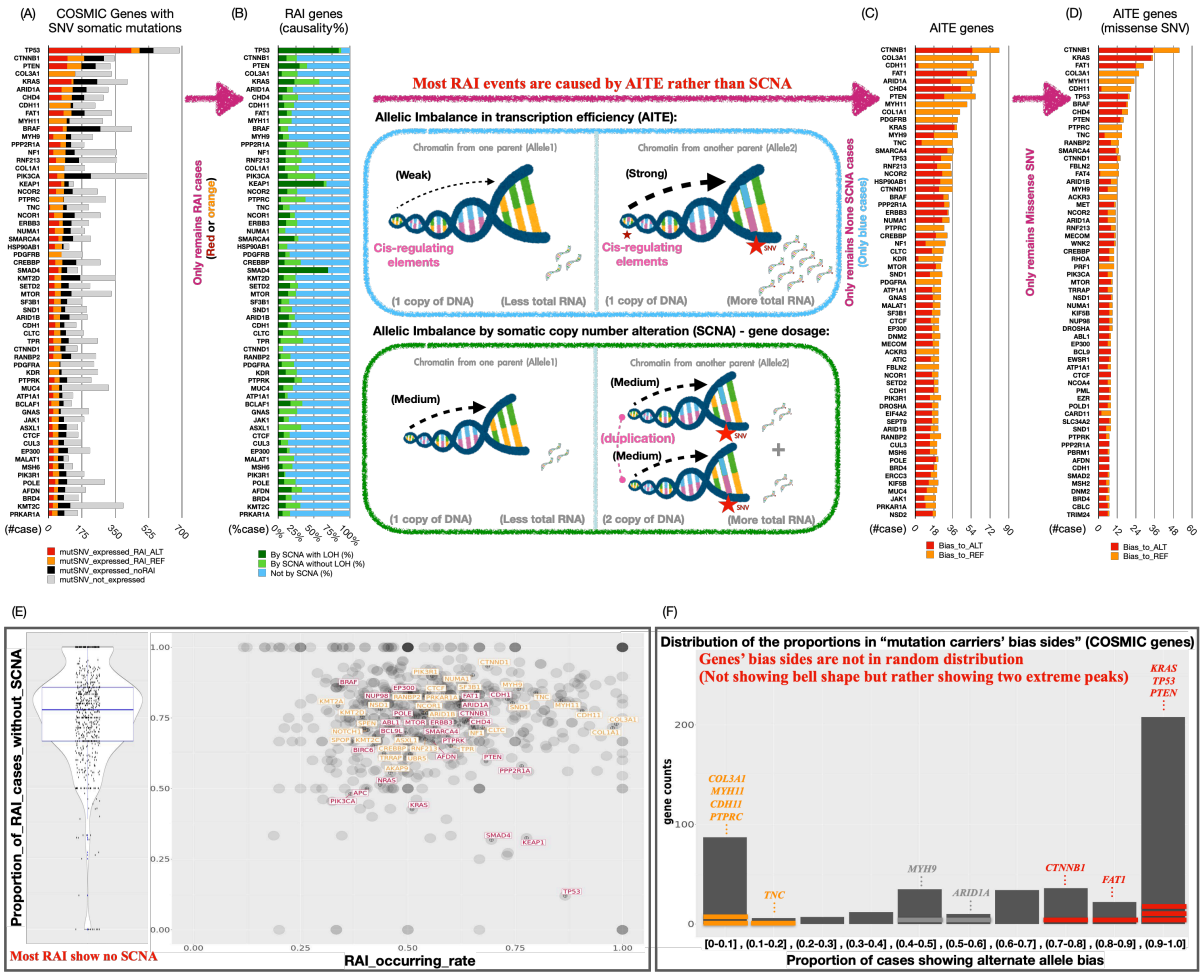


Figure 2. Investigation of the allelic imbalance for those somatic mutations inside COSMIC genes among selected Pan-adenocarcinoma cases (n=3256).

(A) After the allele-specific read coverages correction step based on each tumor purity, those COSMIC genes carrying somatic SNV mutations were detected for their RNA Allelic Imbalance (RAI) effect by RNA-Seq (expression-based data). Here only shows for the top-60 frequent ones.

(B) Among those cases with expressed heterozygous mutations showing the RAI effects, those Allelic Imbalance in Transcription Efficiency (AITE) cases were further identified (in blue color) by referring to WESeq and SNP-array (genotype-based data), and the remaining Somatic Copy Number Alterations (SCNA) cases (in green colors) could also be identified and excluded.

(C) Counts of AITE cases which carry expressed somatic SNV mutations with RAI but without SCNA were calculated for each gene. Here only shows for the top-60 frequent ones (re-ranked).

(D) Counts of AITE cases which carry expressed missense somatic SNV mutations with RAI but without SCNA were calculated for each gene.

(E) Here shows the prevalence of AITE events (showing no SCNA in genotype) in the cases carrying somatic SNV mutations versus RAI occurring rate for each gene, and those top-60 frequent genes (points being marked with gene name) can be viewed in (Figure 2A, 2B).

(F) For each COSMIC gene (cancer-associated gene), the heterozygous somatic nonsynonymous (missense) mutation carriers with AITE effects are finally remained, and those top-60 frequent genes (with most AITE cases counts) can be viewed in (Figure 2D). Here, the X-axis has grouped the genes with different proportion values which shows the percentage of cases showing the alternate allele bias side, and the Y-axis shows the gene counts. For example, in the right most bar, it shows there are arounds 200 cancer-associated genes which all present more

than 90 percent of their mutation carriers showing the alternate bias. In contrast, in the left most bar, it shows there are arounds 100 cancer-associated genes which all present more than 90 percent of their mutation carriers showing the reference bias (less than 10 percent for the alternate side). As for those middle bars, there are relatively less genes showing the ambiguous bias sides among their mutation carriers. If we further set the threshold (more than 5) for the number of samples with mutations showing AITE effect among those genes to decrease noise, the result can be further referred in (S10 Fig A, B).

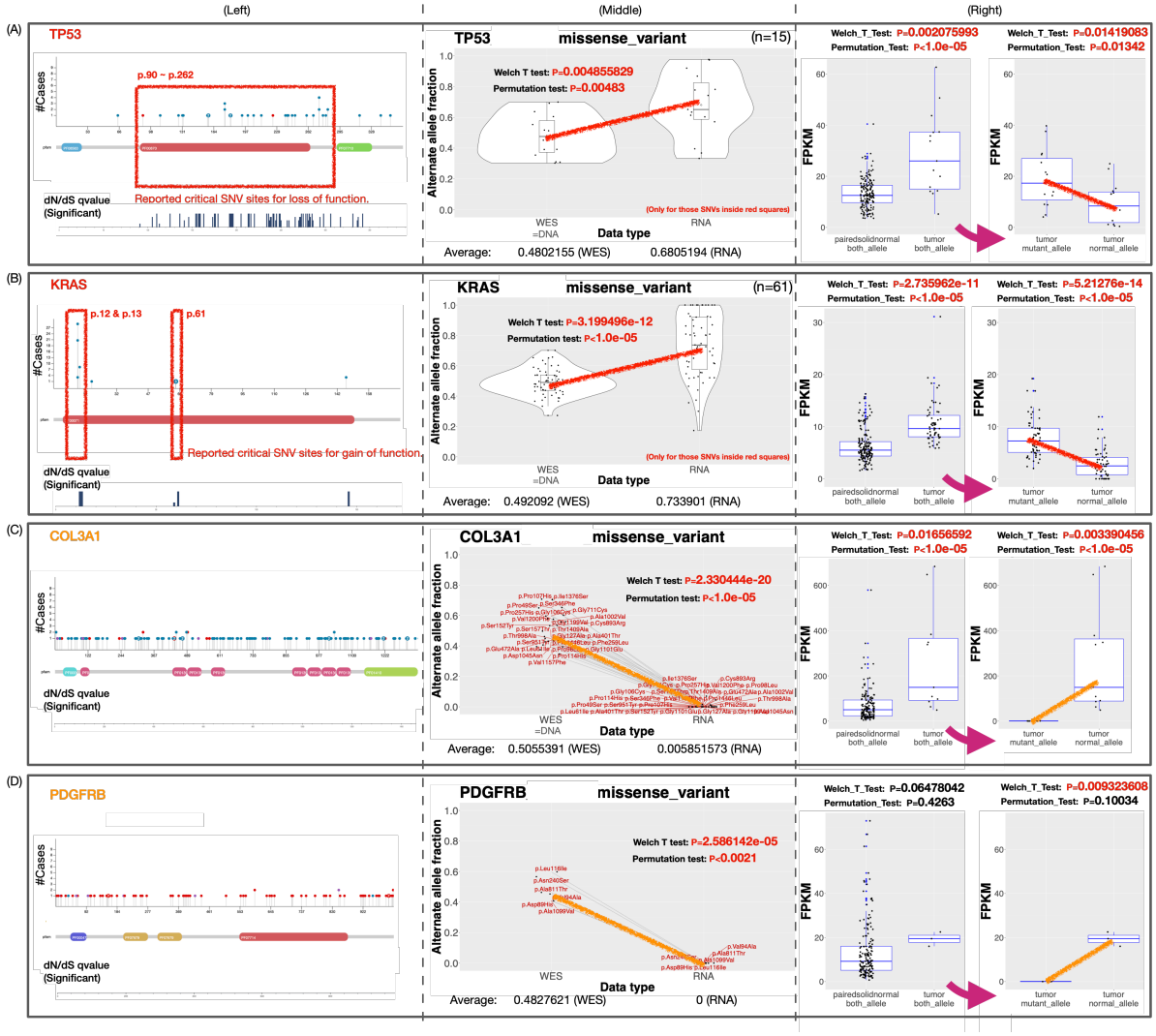


Figure 3. Examples for the AITE effects being observed in typical cancer driver genes and passenger genes, which show the imbalanced allele fraction and imbalanced specific allele expression level.

(Left) The somatic mutation frequency and the related positive selection signals (dN/dS based) were shown for each SNV site along the target gene.

(Middle) Evidences of Allelic Imbalance in Transcriptional Efficiency (AITE) for the well-known cancer driver genes / passenger genes were shown here. Among those RAI cases, after excluding those cases showing mutation sites being overlapped with SCNA (including LOH) region, the remained cases were supposed to be the AITE cases and being visualized here. Each point stands for each sample case. For those supposed AITE cases, the average values of alternate (mutant) allele fractions were counted for WESeq and RNA-Seq respectively, and their fractions were further compared by Welch t-test and permutation test.

(Right) How AITE effects impact the absolute expression level of alternate allele for well-known cancer driver genes / passenger genes were shown here. As described in (middle), those remained missense somatic SNV mutation cases showing AITE were identified from those RAI cases without overlapping their exons with SCNA (including LOH) regions. And in these sub-figures, it presented the FPKM in order to quantify the total gene expression level difference between the paired-normal-like tissues and tumor tissues under the AITE effect. And next, the "allele-specific FPKM" for those supposed AITE cases were further calculated by multiplying its alternate allele fraction with the corresponding total FPKM for each case, whose values were shown as "tumor mutant allele" and "tumor normal allele". In figures, each point stood for each sample case, and P values were calculated by Welch t-test and permutation test for comparing the average values between different groups.

(A) *TP53* acts as the typical tumor suppressor gene

(B) *KRAS* acts as the typical oncogene.

(C & D) *COL3A1* and *PDGFRB* act as the typical passenger genes.

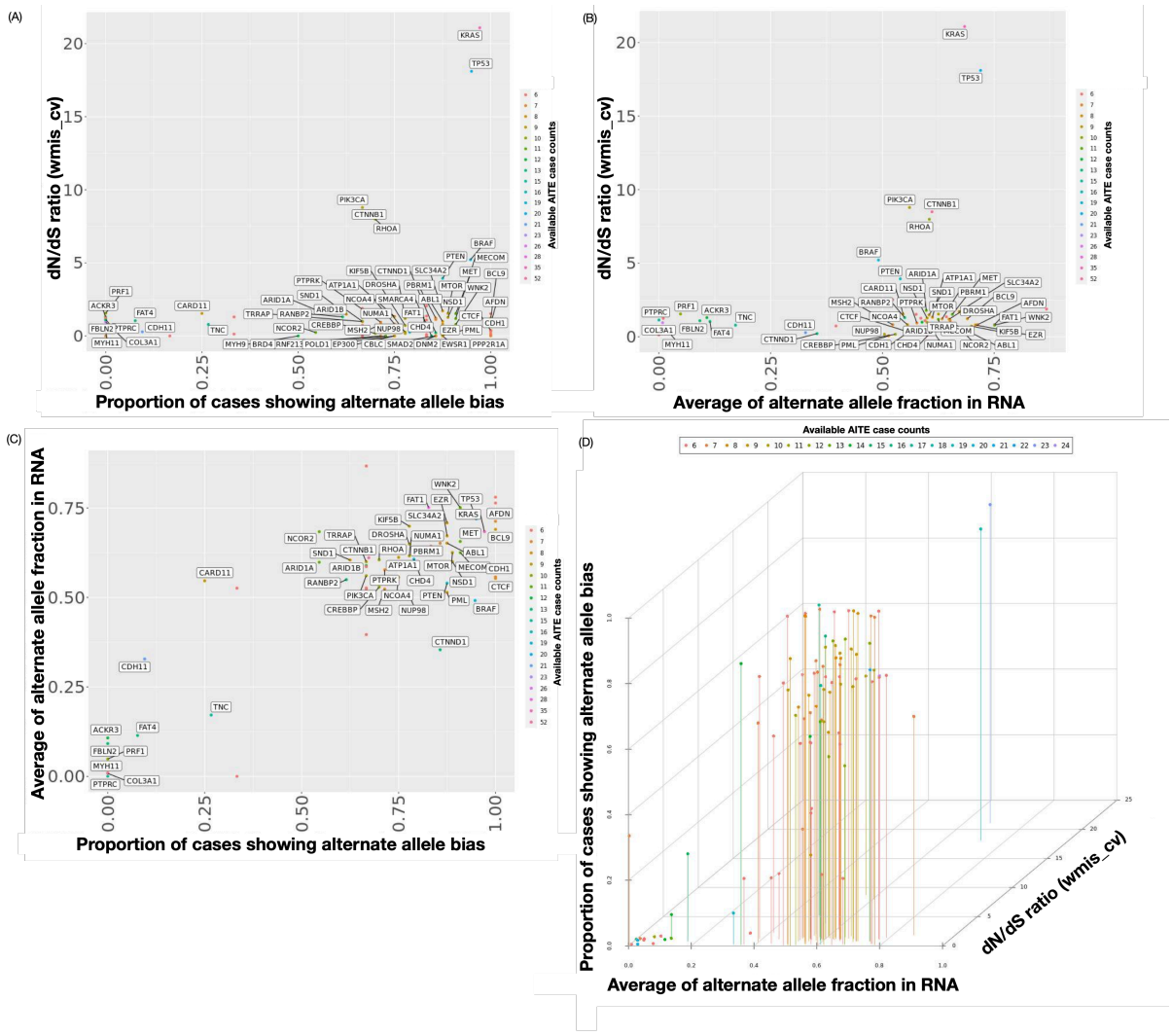


Figure 4. Correlations between the allelic imbalance bias extent and positive selection signals (dN/dS) for those COSMIC genes with AITE effects.

The Spearman's correlation co-efficiencies and the corresponding P-values of each figure can be referred in (S10 Fig).

(A) "Proportion of cases showing alternate allele bias" versus "dN/dS ratio".

(B) "Average of alternate allele fraction in RNA" versus "dN/dS ratio".

(C) "Proportion of cases showing alternate allele bias" versus "Average of alternate allele fraction in RNA".

(D) "Average of alternate allele fraction in RNA" versus "dN/dS ratio" versus "Proportion of cases showing alternate allele bias".

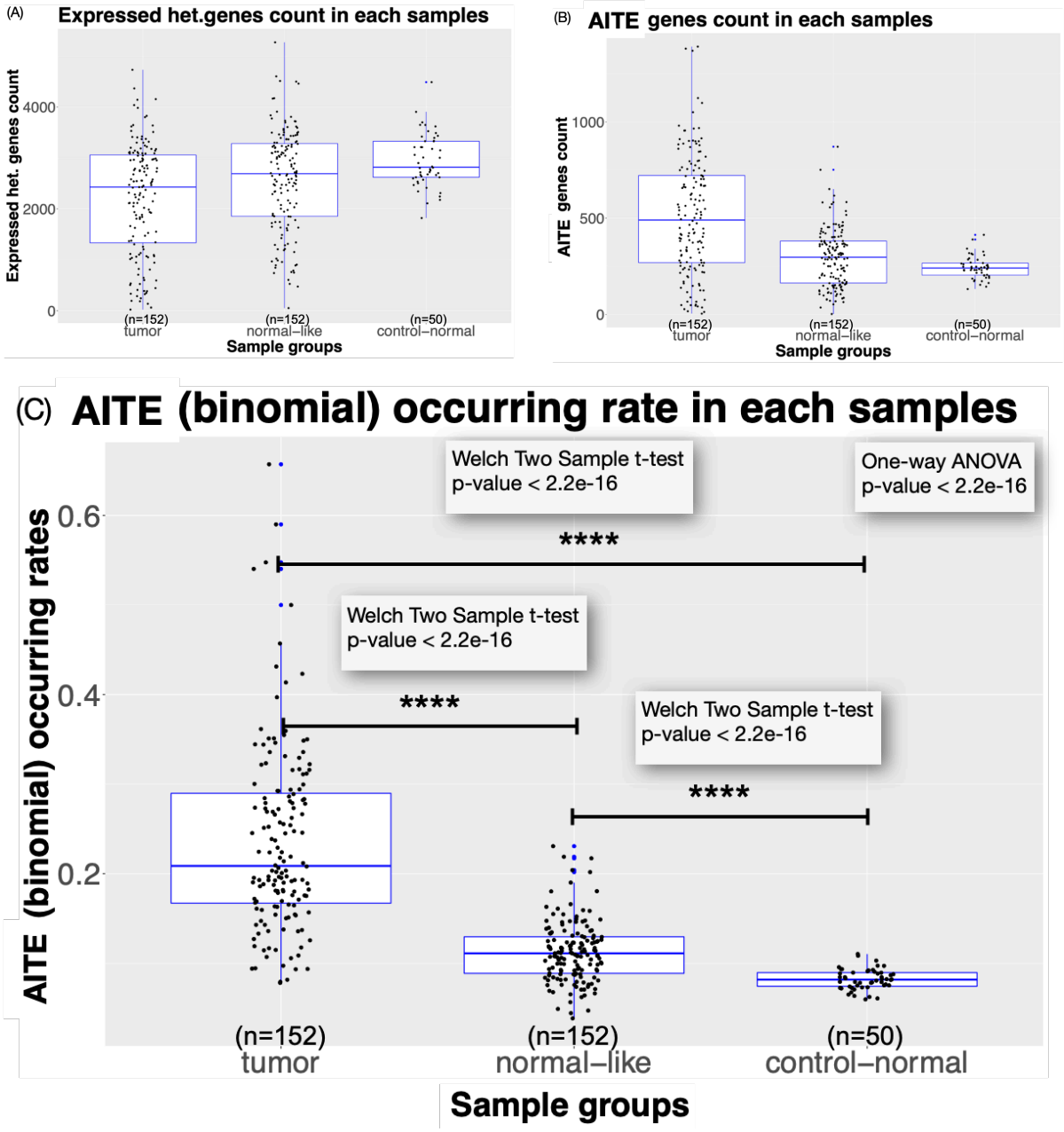


Figure 5. Genome-wide occurring rate of AITE for all expressed heterozygous genes in genome-wide.

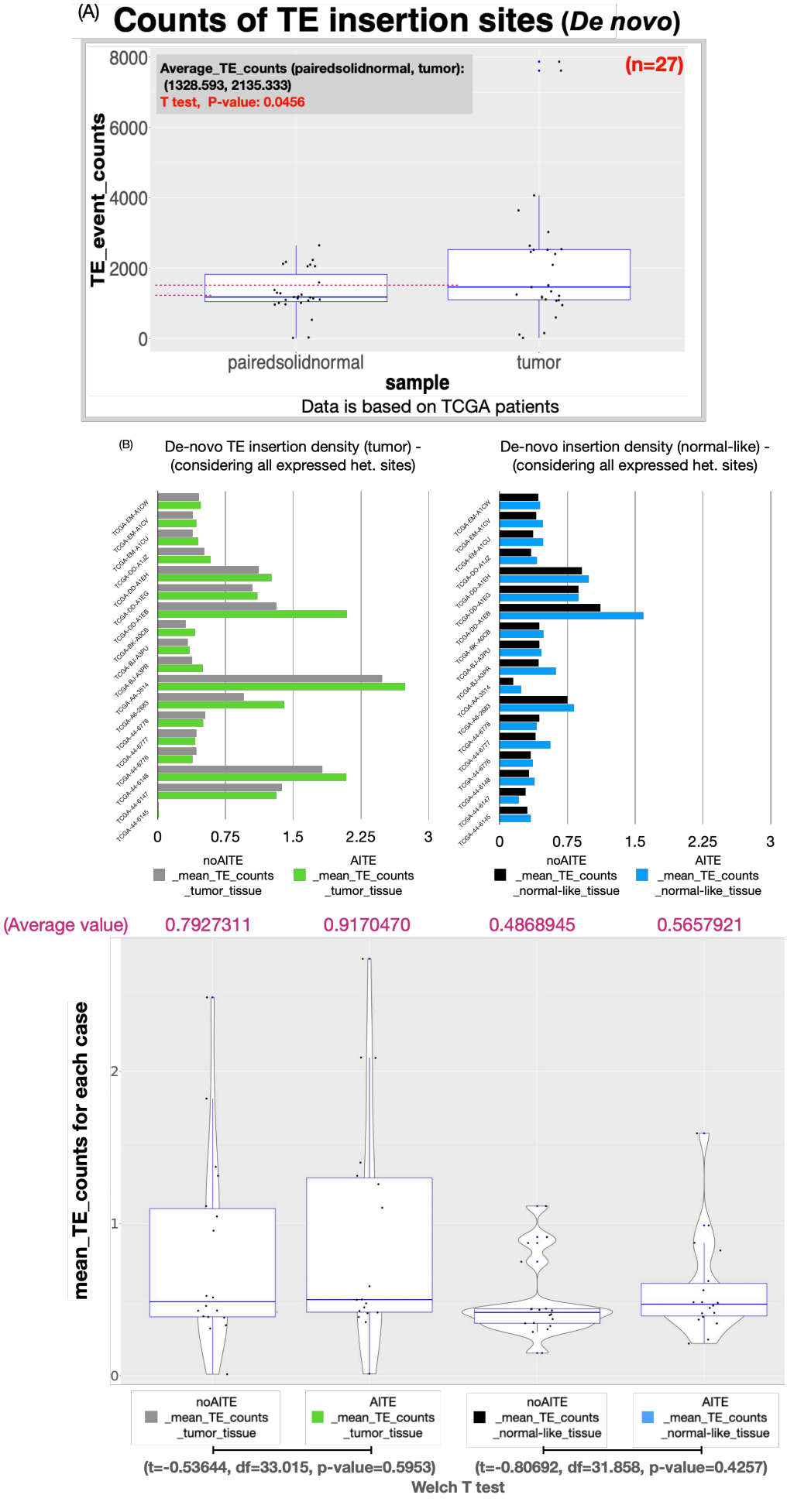
For the tumor tissues and normal tissues, each point stands for a cancer patient, and only the genes located in those genome regions showing no allelic imbalance in DNA (no SCNA event) were involved. For the group of control-normal tissues, each point represents an individual without cancer. The average counts of the genes in different groups for each individual in the different tissue groups are visualized. The AITE occurrence rates were calculated by dividing the AITE gene counts by the expressed heterozygous gene counts for each sample point.

(A) Counts of the expressed heterozygous genes in those genomic regions without DNA allelic imbalance (no SCNA event).

(B) Counts of the genes showing RNA allelic imbalances in those genomic regions without DNA allelic imbalance (no SCNA event).

(C) Genome-wide occurrence rates of AITE for heterozygous genes in those genomic regions without DNA allelic imbalance (no SCNA event) were calculated. Average occurrence rates for each sample group were calculated. Welch's t-tests and Tukey's multiple pairwise comparisons were utilized for comparisons between

any two of the three sample groups, respectively. One-way ANOVA tests were also used for direct comparisons for those three sample groups.


(A)
Counts of TE insertion sites (De novo)
Average_TE_counts (pairedsolidnormal, tumor):
(1328.593, 2135.333)
T test, P-value: 0.0456
(n=27)
TE_event_counts
8000
6000
4000
2000
0
pairedsolidnormal
tumor
sample
Data is based on TCGA patients
(B)
De-novo TE insertion density (tumor) -
(considering all expressed het. sites)
De-novo insertion density (normal-like) -
(considering all expressed het. sites)
0
0.75
1.5
2.25
3
noAITE
_mean_TE_counts
_tumor_tissue
AITE
_mean_TE_counts
_tumor_tissue
noAITE
_mean_TE_counts
_normal-like_tissue
AITE
_mean_TE_counts
_normal-like_tissue
(Average value)
0.7927311
0.9170470
0.4868945
0.5657921
mean_TE_counts for each case
2
1
0
noAITE
_mean_TE_counts
_tumor_tissue
AITE
_mean_TE_counts
_tumor_tissue
noAITE
_mean_TE_counts
_normal-like_tissue
AITE
_mean_TE_counts
_normal-like_tissue
(t=-0.53644, df=33.015, p-value=0.5953)
(t=-0.80692, df=31.858, p-value=0.4257)
Welch T test

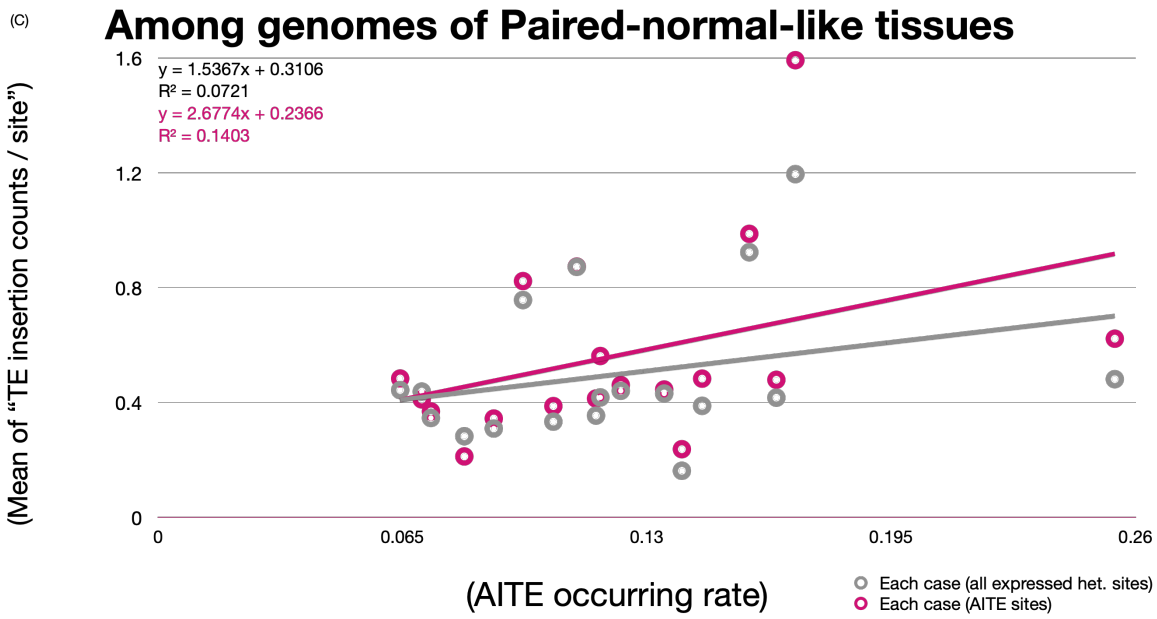
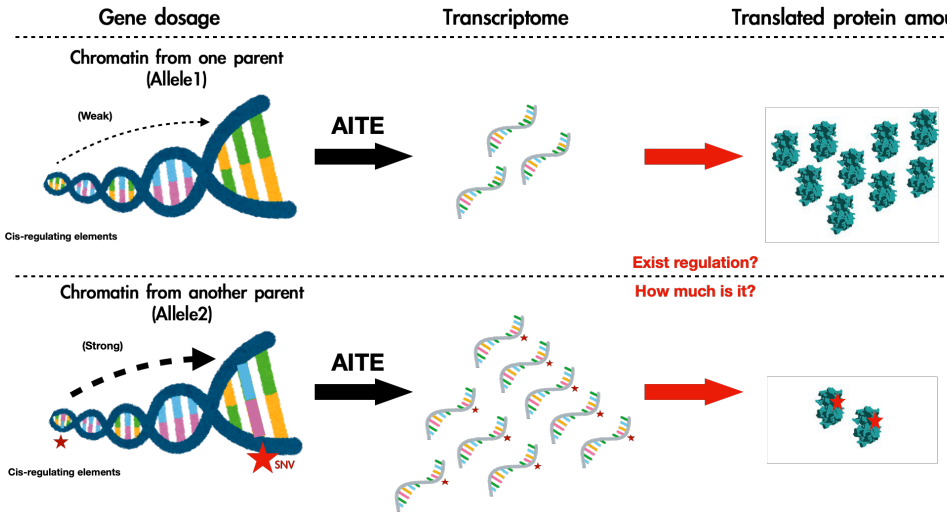


Figure 6. The detection of *de novo* transposable element insertions and their associations with AITE events. **(A)** Figure shows the detection of *de novo* transposable element insertion events in different samples groups. Among the 3,256 selected pan-adenocarcinoma patients, only 27 patients had whole genome sequence (WGSeq) data from the tumor tissues and the paired-normal-like tissues with sufficient read coverage. Therefore, the sample size used for the analysis of *de novo* transposable element insertions were only based on those individuals. Finally, counts of observed *de novo* TE insertion events in tumor tissues and normal tissues were calculated and compared by Welch's t-tests. (Although the permutation test is better for those datasets which exist extreme values (not outliers), the sample size is not big enough to perform it.)

(B) Average in TE insertion counts inside investigation windows (1 Mbp, TE density) for those expressed heterozygous SNV sites (with AITE or noAITE) of different tissues were calculated for each cancer patient (available n = 18, tumor vs. paired-normal-like). The *de novo* TE insertion density was calculated for each targeted position (for those expressed heterozygous SNV sites without allele-specific SCNA). The targeted positions could be further grouped into AITE sites, or for noAITE sites. Such calculations were further applied to tumor tissue and paired-normal-like tissue respectively. Therefore, there were finally four different situations. Finally, for each cancer patient and each situation, the average *de novo* TE insertion density was calculated and showed as each bar value in different colors. And in the right most box figures, the average *de novo* TE insertion density in all available cancer patients were calculated for each situation. (For comparing the differences between each sample group, permutation test is better for these datasets which exist extreme values (not outliers), but the sample size is not big enough to perform it.)

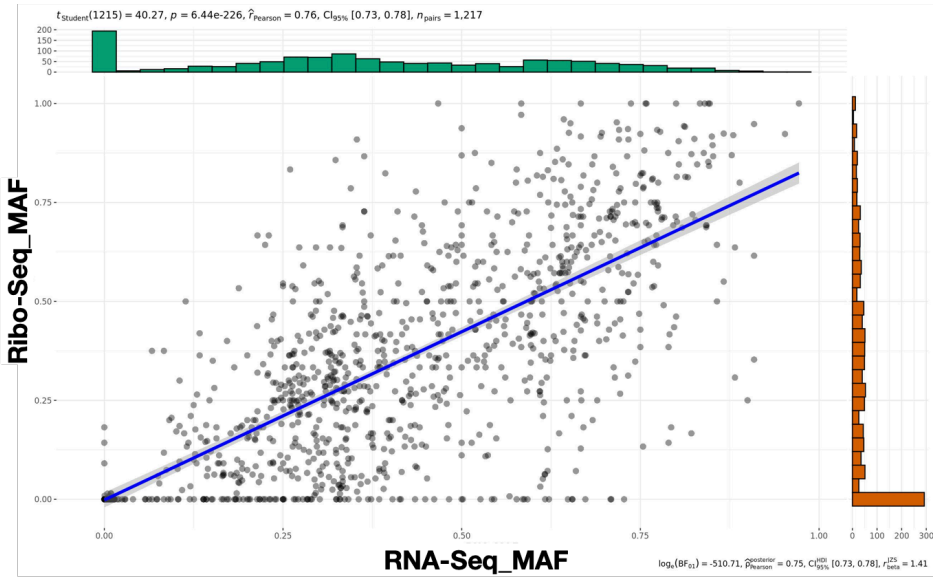
(C) Association between average TE insertions counts and AITE occurring rate in paired-normal-like tissues from different cancer patients were demonstrated here. For each normal-like tissues genome (each sample point), the average *de novo* TE insertion density was calculated for the targeted positions (for those expressed heterozygous SNV sites without allele-specific SCNA). The targeted positions were further settled for the all expressed heterozygous SNV sites (grey ones), or only for those AITE sites (pink ones). The positive correlation can be observed between the genome-wide AITE occurring rates and the average *de novo* TE insertion density among those normal-like tissue genomes (both in grey and in pink).

(A) **AITE effect may be buffered by translational level for allelic proteins amounts:**



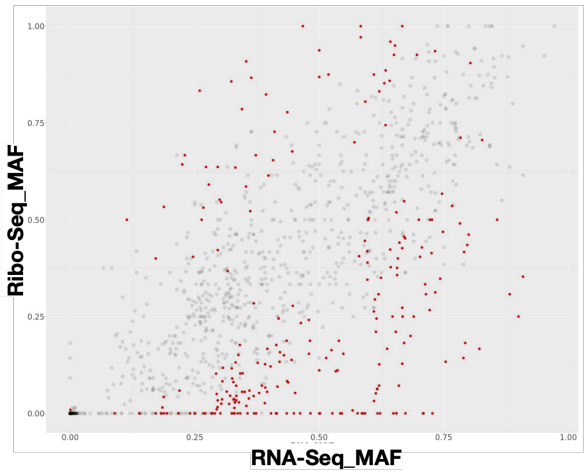
(Allelic Imbalance in translation efficiency: The alternate allele fraction is different between RNAseq and Riboseq.)

(B)

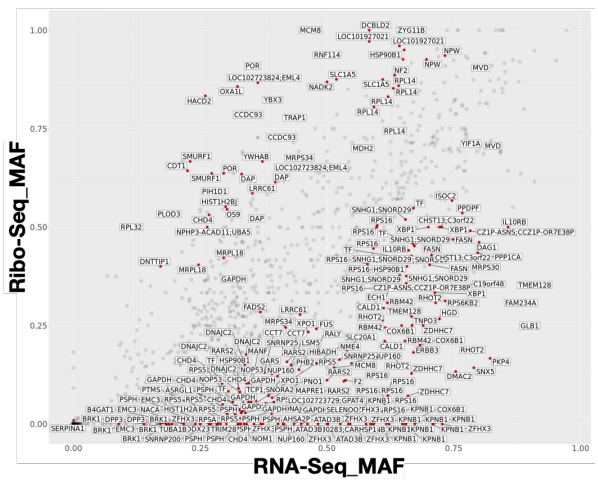


PS. Mutant Allele Fraction = MAF

(C)



(D)



Around 4.2% of the expressed het. SNV sites' alternate allelic ratios
(Red-marked) are significantly different between RNA-Seq & Ribo-Seq

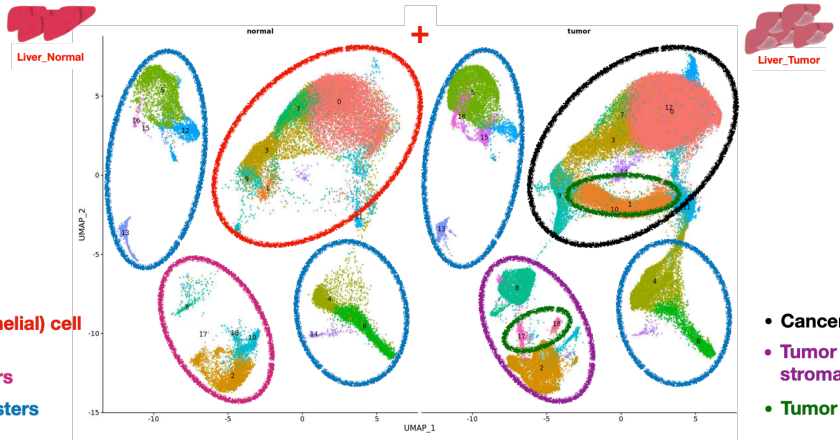
Figure 7. Ribo-Seq reveals the regulations for allelic imbalance in translation efficiency.

(A) This figure illustrates the definition in allelic Imbalance in translation efficiency, which belongs to those positions showing the different alternate allele fraction between RNA-Seq and Ribo-Seq. Here, RNA-Seq is supposed to reveal the transcriptome profile, and Ribo-Seq is supposed to reveal the proteome profile. If the allelic Imbalance in translation efficiency exist, the unidentified post transcriptional regulation should also exist.
(B) The correlations between the mutant allele frequency (MAF) of RNA-Seq and Ribo-Seq. Each point stands for each expressed heterozygous gene in liver cancer cell-line.
(C) This figure shows the mutant allele frequency (MAF) for RNA-Seq and Ribo-Seq, and each point stands for each expressed heterozygous gene. For those genes whose MAFs are significantly different between RNA-Seq and Ribo-Seq, they are marked as red points (4.2% of the sample points).
(D) This figure shows the gene names for those genes whose MAFs are significantly different between RNA-Seq and Ribo-Seq.

(A)

Pseudo-bulk single-cell RNA-Seqs:

(10X genomics)

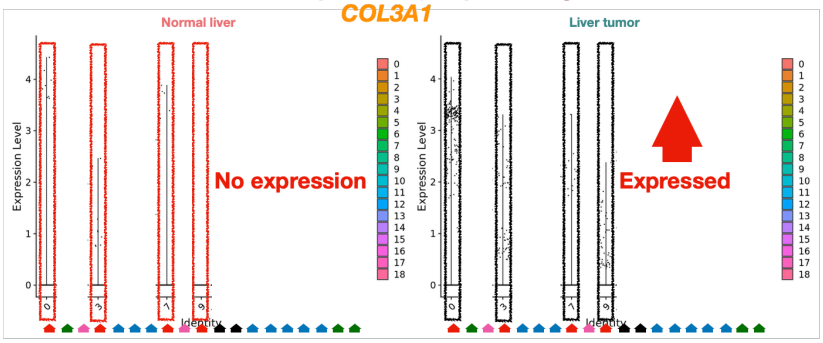


- Parenchymal (Epithelial) cell clusters
- Stromal cell clusters
- PBMC specific clusters

- Cancer cell clusters
- Tumor associated stromal cell clusters
- Tumor specific clusters

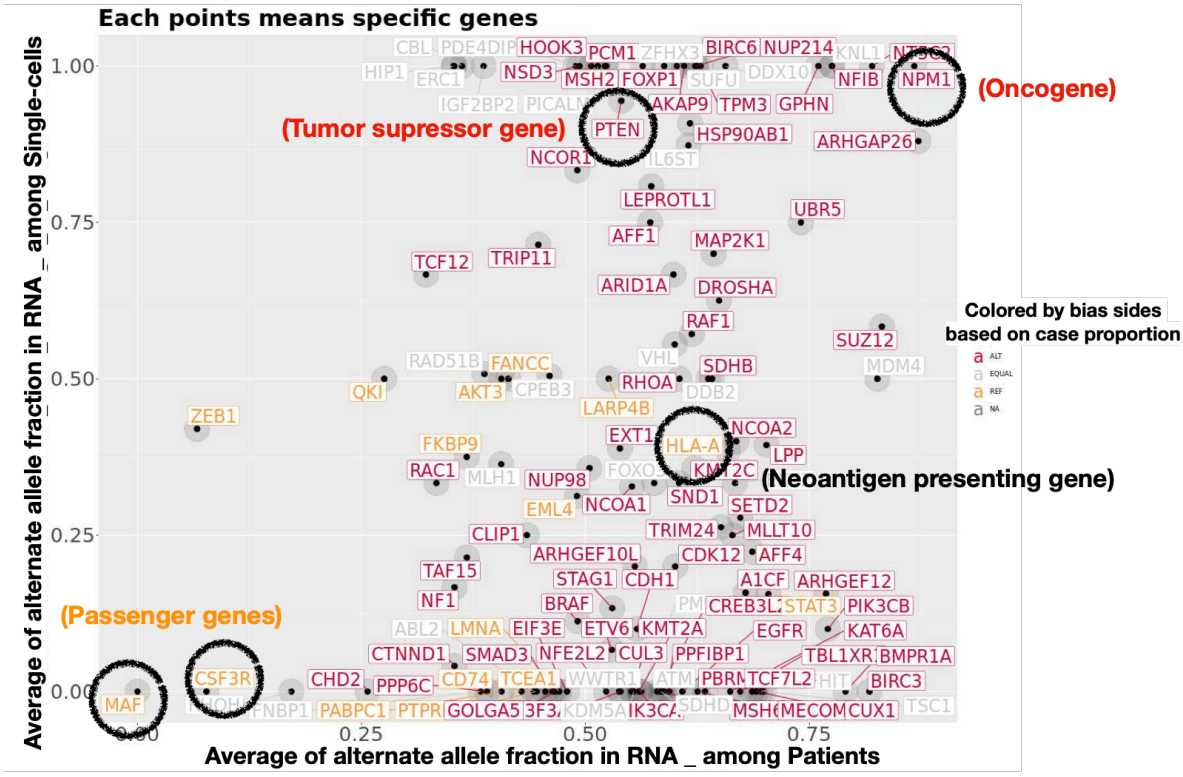
(B)

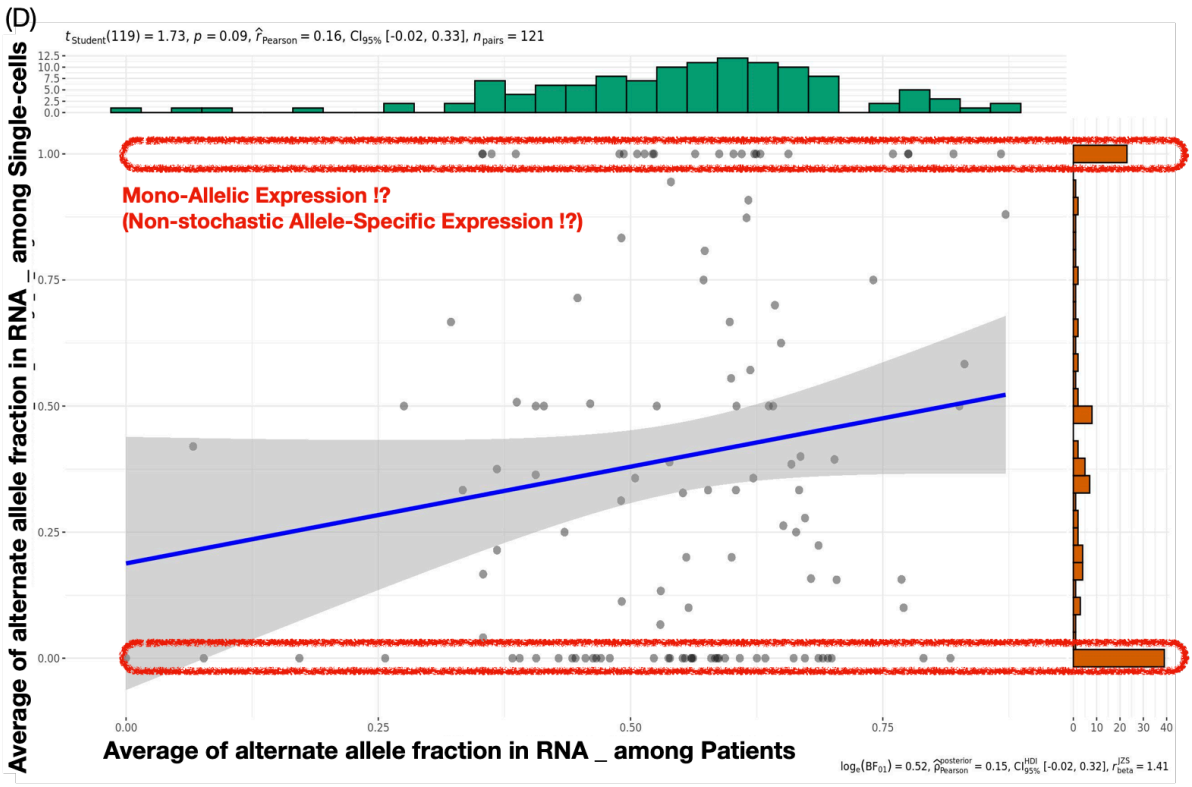
Stromal cell specifically expressed genes



Tumor cells change to express those non-cancer driver genes which are not expressed in normal epithelium!

(C) **(Cancer cells)** cluster_0_3_10_Pt13b (tumor) (44675 cells)





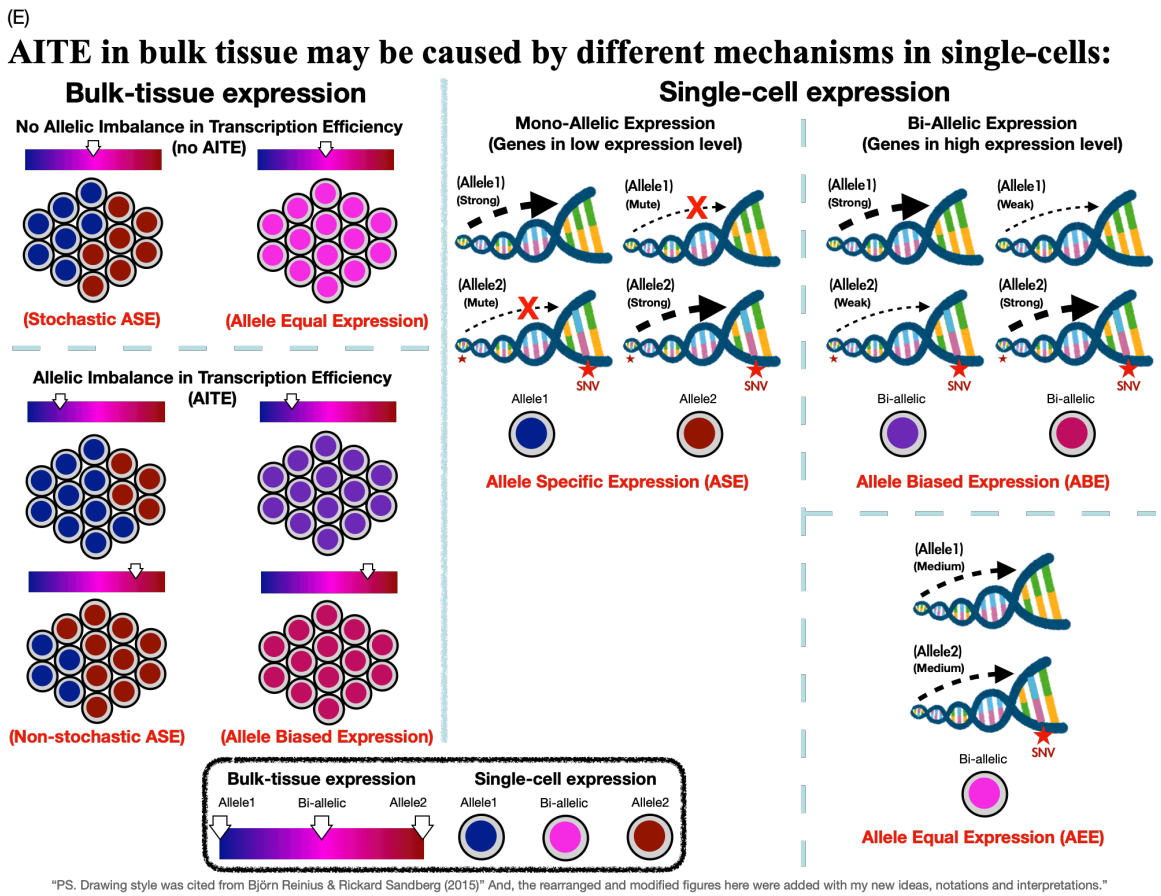


Figure 8. Pseudo-bulk scRNA-Seqs reveal the issues of tissue heterogeneity inside tumor micro-environments (TME) (10X genomics platform).

(A) Different cell clusters were identified by up-regulated genes and their sample sources, and those clusters were further grouped into by different cell types, which include "Parenchymal (Epithelial) cell clusters", "Cancer cell

clusters", "Stromal cell clusters", "Tumor associated stromal cell clusters", "PBMC specific clusters" and "Tumor specific clusters".

(B) COL3A1 is supposed to be only expressed by those stromal cells and not to be expressed by epithelial cells inside normal liver. But those cancer cells which were developed from the epithelial cells, can be detected for the COL3A1 expression for non-cancer driver functions inside liver tumors.

(C) Those scRNA-Seqs for the "Cancer cell clusters" were picked up and merged as pseudo-bulk RNA-Seq. Then, for those expressed heterozygous genes, alternate allele fractions were further calculated for each gene based on the pseudo-bulk RNA-Seq in Y-axis, and the average alternate allele fractions were also calculated for those bulk tissues based on TCGA-provided RNA-Seqs in X-axis.

(D) Based on (C), the correlations for the RNA alternate allele fractions between single cells derived data and bulk tissues derived data were calculated, and positive correlations can be observed for the gene bias sides consistency. Those genes showing mono-allelic expressions were circled by red lines, and their non-random bias sides among single cells suggest the occurrence of non-stochastic allele-specific expression.

(E) The expression profiles being observed in a single bulk tissue can be attributed to the accumulated effects from each single-cell inside that tissue. Here, it demonstrates how the allelic imbalance in transcriptional efficiency (AITE) effect inside bulk tissues might be caused by each single-cells in different ways. Firstly, for a single-cell, there exist three different situations in how a heterozygous gene can be expressed. For the situation in mono-allelic expression, allele-specific expression (ASE) will stochastically choose one allele to express, and this happens in those genes with low or middle expression level. For the situation in bi-allelic expression, both alleles will be expressed, and this happens in those genes with high expression level. However, in bi-allelic expression there might further exist allele equal expression (AEE) or allele biased expression (ABE). Now, let us further consider the bulk tissue. If ASE is under the process of stochastically choosing alleles inside each single-cell, or if AEE exist in each single-cell, then AITE will not be observed in bulk tissue. In contrast, if ASE is choosing alleles non-stochastically inside each single-cell, or if ABE exists in those single-cells, then AITE will be observed in bulk tissue.

Table 1. Result of AITE tests for the identified RNA allelic imbalance (RAI) genes which carry somatic SNV mutations.

Each gene whose somatic SNV mutation carriers showed allelic imbalance in RNA but with no allelic imbalance in DNA (no SCNA and no LOH, based on SNP-array data) are shown in the "RAI_no_SCNA (also no LOH)" column. The WESeq read coverage was used to confirm consistency with the SNP-array data, and if it also showed no significant bias from the expected 0.5 in the alternate allele fraction according to binomial tests in WESeq, then they were retained for AITE testing and were added to the "RAI_no_SCNA cons. transcription effi." column. They were further separated into "RAI_no_SCNA cons. transcription effi._ALT" or "RAI_no_SCNA cons. transcription effi._REF," based on the observed bias of the allele fraction. The top 60 most frequent genes showing RNA Allelic Imbalance were chosen for listing out here. This table only list top 60 frequent genes, and more complete data can be referred from S6 Table.

gene_ID	mut. SNV	expressed	RAI	RAI_ALT	RAI_REF	RAI _SCNA	RAI _SCNA_LOH	RAI _SCNA (%)	RAI _SCNA_LOH (%)	RAI _no_SCNA (also no LOH)	RAI_no_SCNA cons. transcription effi._ALT	RAI_no_SCNA cons. transcription effi._REF	
CTNNB1	210	200	127	68	60	33	15	25.98%	11.81%	52	35	17	ALT
KRAS	351	217	112	108	5	66	25	58.93%	22.32%	35	34	1	ALT
FAT1	154	85	48	33	18	7	3	14.58%	6.25%	28	24	5	ALT
COL3A1	61	56	56	0	56	13	2	23.21%	3.57%	26	0	26	REF
MYH11	71	43	39	0	39	6	1	15.38%	2.56%	23	0	23	REF
CDH11	75	56	48	2	46	12	6	25.00%	12.50%	21	2	19	REF
TP53	353	318	297	284	15	268	259	90.24%	87.21%	20	19	1	ALT
BRAF	182	137	50	42	8	8	2	16.00%	4.00%	19	18	1	ALT
CHD4	75	70	40	23	17	8	2	20.00%	5.00%	19	15	4	ALT
PTEN	134	118	71	55	17	34	24	47.89%	33.80%	16	14	2	ALT
PTPRC	81	46	44	0	44	19	3	43.18%	6.82%	15	0	15	REF
TNC	70	41	33	6	27	5	3	15.15%	9.09%	15	4	11	REF
RANBP2	106	73	37	16	24	7	1	18.92%	2.70%	13	8	5	ALT
SMARCA4	70	61	31	23	8	10	8	32.26%	25.81%	13	11	2	ALT
CTNND1	52	44	26	17	12	3	0	11.54%	0.00%	13	12	2	ALT
FBLN2	64	26	23	0	23	5	4	21.74%	17.39%	13	0	13	REF
FAT4	318	31	22	2	20	5	1	22.73%	4.55%	13	1	12	REF
ARID1B	85	67	42	19	23	14	9	33.33%	21.43%	12	8	4	ALT
MYH9	40	37	27	8	19	2	1	7.41%	3.70%	12	6	6	EQUAL
ACKR3	35	26	22	0	22	2	1	9.09%	4.55%	12	0	12	REF
MET	77	49	34	23	11	14	1	41.18%	2.94%	11	10	1	ALT
NCOR2	81	59	32	12	21	7	3	21.88%	9.38%	11	6	5	ALT
ARID1A	66	51	29	14	15	6	3	20.69%	10.34%	11	6	5	ALT
RNF213	91	50	24	12	12	6	4	25.00%	16.67%	11	8	3	ALT
MECOM	55	41	22	12	10	2	1	9.09%	4.55%	11	10	1	ALT
WNK2	55	31	16	11	6	1	1	6.25%	6.25%	11	10	1	ALT
CREBBP	85	65	26	15	11	9	4	34.62%	15.38%	10	7	3	ALT
RHOA	21	21	19	13	6	5	3	26.32%	15.79%	10	7	3	ALT
PRF1	46	16	15	0	15	3	1	20.00%	6.67%	10	0	10	REF
PIK3CA	261	117	38	22	16	18	4	47.37%	10.53%	9	6	3	ALT
MTOR	117	74	34	20	14	12	8	35.29%	23.53%	9	8	1	ALT
TRRAP	139	70	29	11	18	9	1	31.03%	3.45%	9	6	3	ALT
NSD1	86	55	26	12	14	6	3	23.08%	11.54%	9	8	1	ALT
NUMA1	33	29	21	9	12	2	0	9.52%	0.00%	9	7	2	ALT
KIF5B	37	34	20	11	10	2	0	10.00%	0.00%	9	7	2	ALT
NUP98	58	47	20	15	5	4	2	20.00%	10.00%	9	7	2	ALT
DROSHA	47	35	18	9	10	5	2	27.78%	11.11%	9	7	2	ALT
ABL1	51	45	23	11	12	6	4	26.09%	17.39%	8	7	1	ALT
EP300	70	42	20	11	9	5	4	25.00%	20.00%	8	6	2	ALT
BCL9	55	44	19	8	11	9	0	47.37%	0.00%	8	8	0	ALT
EWSR1	27	25	19	10	9	3	1	15.79%	5.26%	8	7	1	ALT
ATP1A1	26	26	18	7	12	4	3	22.22%	16.67%	8	6	2	ALT
CTCF	47	42	18	13	5	3	1	16.67%	5.56%	8	8	0	ALT
NCOA4	35	32	17	9	8	4	2	23.53%	11.76%	8	6	2	ALT
PML	25	21	15	10	5	4	2	26.67%	13.33%	8	7	1	ALT
EZR	20	20	14	11	3	3	1	21.43%	7.14%	8	7	1	ALT
POLD1	36	29	14	9	5	1	0	7.14%	0.00%	8	6	2	ALT
CARD11	66	19	13	3	10	3	0	23.08%	0.00%	8	2	6	REF
SLC34A2	34	20	13	10	3	3	3	23.08%	23.08%	8	7	1	ALT
SND1	27	16	13	5	8	0	0	0.00%	0.00%	8	5	3	ALT
PTPRK	71	54	26	18	8	10	7	38.46%	26.92%	7	5	2	ALT
PPP2R1A	30	29	23	18	5	11	1	47.83%	4.35%	7	7	0	ALT
PBRM1	59	42	22	11	11	7	4	31.82%	18.18%	7	6	1	ALT
AFDN	49	41	20	11	9	5	3	25.00%	15.00%	7	7	0	ALT
CDH1	27	25	20	14	6	3	0	15.00%	0.00%	7	7	0	ALT
SMAD2	35	31	20	11	9	5	4	25.00%	20.00%	7	6	1	ALT
MSH2	33	25	17	11	7	6	2	35.29%	11.76%	7	5	2	ALT
DNM2	23	21	14	10	4	2	0	14.29%	0.00%	7	6	1	ALT
BRD4	37	29	11	5	6	1	0	9.09%	0.00%	7	5	2	ALT
CBLC	17	13	10	8	2	2	0	20.00%	0.00%	7	6	1	ALT

6 REFERENCES

1. Watson JD, Crick FHC. Molecular Structure of Nucleic Acids: A Structure for Deoxyribose Nucleic Acid. Nature 1953 171:4356 [Internet]. 1953 Apr 25 [cited 2022 Aug 12];171(4356):737–8. Available from: https://www.nature.com/articles/171737a0
2. Wetterstrand KA. DNA sequencing costs: data from the NHGRI Genome Sequencing Program (GSP) [Internet]. National Human Genome Research Institute. [cited 2022 Jan 4]. Available from: http://www.genome.gov/sequencingcosts
3. Goodwin S, McPherson JD, McCombie WR. Coming of age: ten years of next-generation sequencing technologies. Nat Rev Genet. 2016;17(6):333–51.
4. Shapiro E, Biezuner T, Linnarsson S. Single-cell sequencing-based technologies will revolutionize whole-organism science. Vol. 14, Nature Reviews Genetics. 2013. p. 618–30.
5. Bentley DR, Balasubramanian S, Swerdlow HP, Smith GP, Milton J, Brown CG, et al. Accurate whole human genome sequencing using reversible terminator chemistry. Nature. 2008 Nov;456(7218):53–9.
6. Pease J, Sooknanan R. A rapid, directional RNA-seq library preparation workflow for Illumina® sequencing. Nat Methods. 2012 Mar 28;9(3):i–ii.
7. Barbitoff YA, Polev DE, Glotov AS, Serebryakova EA, Shcherbakova I v., Kiselev AM, et al. Systematic dissection of biases in whole-exome and whole-genome sequencing reveals major determinants of coding sequence coverage. Sci Rep. 2020 Dec 6;10(1):2057.
8. van Dijk EL, Auger H, Jaszczyszyn Y, Thermes C. Ten years of next-generation sequencing technology. Trends in Genetics [Internet]. 2014;30(9):418–26. Available from: https://www.sciencedirect.com/science/article/pii/S0168952514001127
9. Ziegenhain C, Vieth B, Parekh S, Reinius B, Guillaumet-Adkins A, Smets M, et al. Comparative Analysis of Single-Cell RNA Sequencing Methods. Mol Cell. 2017 Feb 16;65(4):631-643.e4.
10. Wang X, He Y, Zhang Q, Ren X, Zhang Z. Direct Comparative Analyses of 10X Genomics Chromium and Smart-seq2. Genomics Proteomics Bioinformatics. 2021 Apr 1;19(2):253–66.
11. Lu H, Giordano F, Ning Z. Oxford Nanopore MinION Sequencing and Genome Assembly. Vol. 14, Genomics, Proteomics and Bioinformatics. Beijing Genomics Institute; 2016. p. 265–79.
12. Rhoads A, Au KF. PacBio Sequencing and Its Applications. Genomics Proteomics Bioinformatics. 2015 Oct;13(5):278–89.
13. Bleidorn C. Third generation sequencing: technology and its potential impact on evolutionary biodiversity research. Syst Biodivers [Internet]. 2016 Jan 2;14(1):1–8. Available from: https://doi.org/10.1080/14772000.2015.1099575
14. Tripathi R, Sharma P, Chakraborty P, Varadwaj PK. Next-generation sequencing revolution through big data analytics. Front Life Sci [Internet]. 2016 Apr 2 [cited 2022 Aug 2];9(2):119–49. Available from: https://www.tandfonline.com/action/journalInformation?journalCode=tfls21

15. Kitchin R, McArdle G. What makes Big Data, Big Data? Exploring the ontological characteristics of 26 datasets: http://dx.doi.org/101177/2053951716631130 [Internet]. 2016 Feb 17 [cited 2022 Aug 11];3(1). Available from: https://journals.sagepub.com/doi/10.1177/2053951716631130
16. Kitchin R. Big Data, new epistemologies and paradigm shifts: http://dx.doi.org/101177/2053951714528481 [Internet]. 2014 Apr 1 [cited 2022 Apr 13];1(1). Available from: https://journals.sagepub.com/doi/10.1177/2053951714528481
17. Hey T, Tansley S, Tolle K. The Fourth Paradigm: Data-Intensive Scientific Discovery [Internet]. The Fourth Paradigm: Data-Intensive Scientific Discovery. Microsoft Research; 2009. Available from: https://www.microsoft.com/en-us/research/publication/fourth-paradigm-data-intensive-scientific-discovery/
18. boyd danah, Crawford K. CRITICAL QUESTIONS FOR BIG DATA. Inf Commun Soc. 2012 Jun;15(5):662–79.
19. McCue ME, McCoy AM. The Scope of Big Data in One Medicine: Unprecedented Opportunities and Challenges. Front Vet Sci. 2017 Nov 16;4.
20. Shih W, Chai S. Data-Driven vs. Hypothesis-Driven Research: Making sense of big data. Academy of Management Proceedings. 2016 Jan;2016(1):14843.
21. Likić VA, McConville MJ, Lithgow T, Bacic A. Systems Biology: The Next Frontier for Bioinformatics. Adv Bioinformatics. 2010 Feb 9;2010:1–10.
22. Mashima J, Kodama Y, Fujisawa T, Katayama T, Okuda Y, Kaminuma E, et al. DNA Data Bank of Japan. Nucleic Acids Res [Internet]. 2017 Jan 4;45(D1):D25–31. Available from: https://doi.org/10.1093/nar/gkw1001
23. Weinstein JN, Collisson EA, Mills GB, Shaw KRM, Ozenberger BA, Ellrott K, et al. The Cancer Genome Atlas Pan-Cancer analysis project. Nat Genet. 2013 Oct 26;45(10):1113–20.
24. The GTEx Consortium atlas of genetic regulatory effects across human tissues. Science (1979). 2020 Sep 11;369(6509):1318–30.
25. Abascal F, Acosta R, Addleman NJ, Adrian J, Afzal V, Aken B, et al. Perspectives on ENCODE. Nature. 2020 Jul 30;583(7818):693–8.
26. Zhang J, Bajari R, Andric D, Gerthoffert F, Lepsa A, Nahal-Bose H, et al. The International Cancer Genome Consortium Data Portal. Nat Biotechnol. 2019 Apr 15;37(4):367–9.
27. Sayers EW, Beck J, Bolton EE, Bourexis D, Brister JR, Canese K, et al. Database resources of the National Center for Biotechnology Information. Nucleic Acids Res. 2021;49(D1):D10–7.
28. Kundaje A, Meuleman W, Ernst J, Bilenky M, Yen A, Heravi-Moussavi A, et al. Integrative analysis of 111 reference human epigenomes. Nature. 2015 Feb 19;518(7539):317–30.
29. Varricchio CG. A cancer source book for nurses. Boston: Jones and Bartlett Publishers, editor. 2004.
30. Latest global cancer data: Cancer burden rises to 18.1 million new cases and 9.6 million cancer deaths in 2018 [Internet]. 2018. Available from: http://gco.iarc.fr/,
31. Balmain A, Barrett JC, Moses H, Renan MJ. How many mutations are required for tumorigenesis? implications from human cancer data. Mol Carcinog. 1993;7(3):139–46.
32. Lee S, Schmitt CA. The dynamic nature of senescence in cancer. Nat Cell Biol. 2019 Jan 2;21(1):94–101.
33. Kinzler KW, Vogelstein B. Lessons from Hereditary Colorectal Cancer. Cell. 1996 Oct;87(2):159–70.

34. Hahn WC, Counter CM, Lundberg AS, Beijersbergen RL, Brooks MW, Weinberg RA. Creation of human tumour cells with defined genetic elements. Nature. 1999 Jul 29;400(6743):464–8.
35. Greaves M, Maley CC. Clonal evolution in cancer. Nature [Internet]. 2012 Jan 19 [cited 2022 Apr 13];481(7381):306–13. Available from: https://pubmed.ncbi.nlm.nih.gov/22258609/
36. Nowell PC. The Clonal Evolution of Tumor Cell Populations. Science (1979). 1976 Oct;194(4260):23–8.
37. Hanahan D, Weinberg RA. The Hallmarks of Cancer. Cell. 2000 Jan;100(1):57–70.
38. Sondka Z, Bamford S, Cole CG, Ward SA, Dunham I, Forbes SA. The COSMIC Cancer Gene Census: describing genetic dysfunction across all human cancers. Vol. 18, Nature Reviews Cancer. Nature Publishing Group; 2018. p. 696–705.
39. Hnisz D, Weintraub AS, Day DS, Valton AL, Bak RO, Li CH, et al. Activation of proto-oncogenes by disruption of chromosome neighborhoods. Science (1979). 2016 Mar 25;351(6280):1454–8.
40. Croce CM. Oncogenes and Cancer. New England Journal of Medicine. 2008 Jan 31;358(5):502–11.
41. Lynn Jorde, John Carey, Michael Bamshad. Medical Genetics. sixth. Elsevier; 2019.
42. Futreal PA, Coin L, Marshall M, Down T, Hubbard T, Wooster R, et al. A census of human cancer genes. Nat Rev Cancer. 2004 Mar;4(3):177–83.
43. Goodrich DW. The retinoblastoma tumor-suppressor gene, the exception that proves the rule. Oncogene. 2006 Aug 28;25(38):5233–43.
44. Berger AH, Knudson AG, Pandolfi PP. A continuum model for tumour suppression. Vol. 476, Nature. 2011. p. 163–9.
45. Carlo MI, Mukherjee S, Mandelker D, Vijai J, Kemel Y, Zhang L, et al. Prevalence of Germline Mutations in Cancer Susceptibility Genes in Patients With Advanced Renal Cell Carcinoma. JAMA Oncol [Internet]. 2018 Sep 1 [cited 2022 Aug 12];4(9):1228–35. Available from: https://jamanetwork.com/journals/jamaoncology/fullarticle/2686813
46. Uzman A, Lodish H, Berk A, Zipursky L, Baltimore D. Molecular Cell Biology (4th edition) New York, NY, 2000, ISBN 0-7167-3136-3. Biochemistry and Molecular Biology Education [Internet]. 2000 [cited 2022 Aug 11];29(3):Section 1.2The Molecules of Life. Available from: http://linkinghub.elsevier.com/retrieve/pii/S1470817501000236
47. Knudson AG. Two genetic hits (more or less) to cancer. Nat Rev Cancer. 2001 Nov;1(2):157–62.
48. Sager R. Tumor suppressor genes: the puzzle and the promise. Science. 1989 Dec 15;246(4936):1406–12.
49. Huang KL, Mashl RJ, Wu Y, Ritter DI, Wang J, Oh C, et al. Pathogenic Germline Variants in 10,389 Adult Cancers. Cell. 2018;173(2):355-370.e14.
50. Gonzalez-Perez A, Tamborero D, Lopez-Bigas N, Margolin A, Ding L, Gordenin D, et al. Thread 1: Mutational drivers. Nat Genet. 2013 Oct 17;
51. Bailey MH, Tokheim C, Porta-Pardo E, Sengupta S, Bertrand D, Weerasinghe A, et al. Comprehensive Characterization of Cancer Driver Genes and Mutations. Cell. 2018 Apr;173(2):371-385.e18.
52. Ciriello G, Miller ML, Aksoy BA, Senbabaoglu Y, Schultz N, Sander C. Emerging landscape of oncogenic signatures across human cancers. Nat Genet. 2013 Oct;45(10):1127–33.
53. Martincorena I, Campbell PJ. Somatic mutation in cancer and normal cells. Science (1979). 2015 Sep 25;349(6255):1483–9.

54. Ally A, Balasundaram M, Carlsen R, Chuah E, Clarke A, Dhalla N, et al. Comprehensive and Integrative Genomic Characterization of Hepatocellular Carcinoma. Cell. 2017 Jun 15;169(7):1327-1341.e23.
55. Cancer Genome Atlas Research Network. Comprehensive molecular characterization of gastric adenocarcinoma. Nature. 2014 Sep 11;513(7517):202–9.
56. Cancer Genome Atlas Research Network. Comprehensive molecular profiling of lung adenocarcinoma. Nature. 2014 Jul 31;511(7511):543–50.
57. Suda K, Nakaoka H, Yoshihara K, Ishiguro T, Tamura R, Mori Y, et al. Clonal Expansion and Diversification of Cancer-Associated Mutations in Endometriosis and Normal Endometrium. Cell Rep. 2018;24(7):1777–89.
58. McKean DM, Homsy J, Wakimoto H, Patel N, Gorham J, DePalma SR, et al. Loss of RNA expression and allele-specific expression associated with congenital heart disease. Nat Commun. 2016 Nov 27;7(1):12824.
59. Wu HC, Do C, Andrulis IL, John EM, Daly MB, Buys SS, et al. Breast cancer family history and allele-specific DNA methylation in the legacy girls study. Epigenetics [Internet]. 2018 Mar 4 [cited 2023 Jan 20];13(3):240–50. Available from: https://www.tandfonline.com/doi/abs/10.1080/15592294.2018.1435243
60. Downs B, Wang SM. Epigenetic changes in BRCA1-mutated familial breast cancer. Cancer Genet. 2015 May 1;208(5):237–40.
61. Stefansson OA, Esteller M. Epigenetic Modifications in Breast Cancer and Their Role in Personalized Medicine. Am J Pathol. 2013 Oct 1;183(4):1052–63.
62. Karlsson M, Zhang C, Méar L, Zhong W, Digre A, Katona B, et al. A single–cell type transcriptomics map of human tissues. Sci Adv [Internet]. 2021 Jul 1 [cited 2022 Nov 16];7(31). Available from: https://www.science.org/doi/10.1126/sciadv.abh2169
63. Feng J, Zhu R, Yin Y, Wang S, Zhou L, Lv F, et al. Re-Recognizing the Cellular Origin of the Primary Epithelial Tumors of the Liver. J Hepatocell Carcinoma [Internet]. 2021 Dec 7 [cited 2023 Jan 17];8:1537–63. Available from: https://www.dovepress.com/re-recognizing-the-cellular-origin-of-the-primary-epithelial-tumors-of-peer-reviewed-fulltext-article-JHC
64. Wiseman BS, Werb Z. Development: Stromal effects on mammary gland development and breast cancer. Science (1979) [Internet]. 2002 May 10 [cited 2023 Jan 17];296(5570):1046–9. Available from: https://www.science.org/doi/10.1126/science.1067431
65. Bussard KM, Mutkus L, Stumpf K, Gomez-Manzano C, Marini FC. Tumor-associated stromal cells as key contributors to the tumor microenvironment. Breast Cancer Research [Internet]. 2016 Aug 11 [cited 2023 Jan 17];18(1):84. Available from: https://mdanderson.elsevierpure.com/en/publications/tumor-associated-stromal-cells-as-key-contributors-to-the-tumor-m
66. Shi Y, Du L, Lin L, Wang Y. Tumour-associated mesenchymal stem/stromal cells: emerging therapeutic targets. Nature Reviews Drug Discovery 2016 16:1 [Internet]. 2016 Nov 4 [cited 2023 Jan 17];16(1):35–52. Available from: https://www.nature.com/articles/nrd.2016.193
67. Lin LY, Du LM, Cao K, Huang Y, Yu PF, Zhang LY, et al. Tumour cell-derived exosomes endow mesenchymal stromal cells with tumour-promotion capabilities. Oncogene. 2016 Nov 17;35(46):6038–42.
68. Guo Z, Zhang H, Fu Y, Kuang J, Zhao B, Zhang L, et al. Cancer-associated fibroblasts induce growth and radioresistance of breast cancer cells through paracrine IL-6. Cell Death Discovery 2023 9:1

[Internet]. 2023 Jan 13 [cited 2023 Jan 17];9(1):1–10. Available from: https://www.nature.com/articles/s41420-023-01306-3

69. Abubakar M, Zhang J, Ahearn TU, Koka H, Guo C, S. LawrencScottM. Lawrenc, et al. Tumor-associated stromal cellular density as a predictor of recurrence and mortality in breast cancer: Results from ethnically diverse study populations. Cancer Epidemiology Biomarkers and Prevention [Internet]. 2021 Jul 1 [cited 2023 Jan 17];30(7):1397–407. Available from: https://aacrjournals.org/cebp/article/30/7/1397/671007/Tumor-Associated-Stromal-Cellular-Density-as-a

70. Quail DF, Joyce JA. Microenvironmental regulation of tumor progression and metastasis. Nature Medicine 2013 19:11 [Internet]. 2013 Nov 7 [cited 2023 Jan 17];19(11):1423–37. Available from: https://www.nature.com/articles/nm.3394

71. Dentro SC, Leshchiner I, Haase K, Tarabichi M, Wintersinger J, Deshwar AG, et al. Characterizing genetic intra-tumor heterogeneity across 2,658 human cancer genomes. Cell. 2021 Apr 15;184(8):2239-2254.e39.

72. Bielski CM, Donoghue MTA, Gadiya M, Hanrahan AJ, Won HH, Chang MT, et al. Widespread Selection for Oncogenic Mutant Allele Imbalance in Cancer. Cancer Cell. 2018 Nov 12;34(5):852-862.e4.

73. Palin K, Pitkänen E, Turunen M, Sahu B, Pihlajamaa P, Kivioja T, et al. Contribution of allelic imbalance to colorectal cancer. Nat Commun. 2018 Dec 10;9(1):3664.

74. Santoni FA, Stamoulis G, Garieri M, Falconnet E, Ribaux P, Borel C, et al. Detection of Imprinted Genes by Single-Cell Allele-Specific Gene Expression. Am J Hum Genet. 2017 Mar 2;100(3):444–53.

75. Jiang Y, Zhang NR, Li M. SCALE: Modeling allele-specific gene expression by single-cell RNA sequencing. Genome Biol. 2017 Apr 26;18(1).

76. Nakaoka H, Gurumurthy A, Hayano T, Ahmadloo S, Omer WH, Yoshihara K, et al. Allelic Imbalance in Regulation of ANRIL through Chromatin Interaction at 9p21 Endometriosis Risk Locus. PLoS Genet. 2016 Apr;12(4):e1005893.

77. Ongen H, Andersen CL, Bramsen JB, Oster B, Rasmussen MH, Ferreira PG, et al. Putative cis-regulatory drivers in colorectal cancer. Nature. 2014;512(1):87–90.

78. Przytycki PF, Singh M. Differential Allele-Specific Expression Uncovers Breast Cancer Genes Dysregulated by Cis Noncoding Mutations. Cell Syst. 2020;10(2):193-203.e4.

79. Cheng Z, Vermeulen M, Rollins-Green M, DeVeale B, Babak T. Cis-regulatory mutations with driver hallmarks in major cancers. iScience. 2021 Mar 19;24(3).

80. van Loo P, Nordgard SH, Lingjærde OC, Russnes HG, Rye IH, Sun W, et al. Allele-specific copy number analysis of tumors. Proc Natl Acad Sci U S A. 2010 Sep 28;107(39):16910–5.

81. Zack TI, Schumacher SE, Carter SL, Cherniack AD, Saksena G, Tabak B, et al. Pan-cancer patterns of somatic copy number alteration. Nat Genet. 2013 Oct;45(10):1134–40.

82. Stamoulis G, Garieri M, Makrythanasis P, Letourneau A, Guipponi M, Panousis N, et al. Single cell transcriptome in aneuploidies reveals mechanisms of gene dosage imbalance. Nat Commun. 2019 Dec 1;10(1).

83. Knijnenburg TA, Wang L, Zimmermann MT, Chambwe N, Gao GF, Cherniack AD, et al. Genomic and Molecular Landscape of DNA Damage Repair Deficiency across The Cancer Genome Atlas. Cell Rep. 2018 Apr 3;23(1):239-254.e6.

84. Fehrmann RSN, Karjalainen JM, Krajewska M, Westra HJ, Maloney D, Simeonov A, et al. Gene expression analysis identifies global gene dosage sensitivity in cancer. Nat Genet. 2015 Feb 12;47(2):115–25.
85. Shlien A, Raine K, Fuligni F, Arnold R, Nik-Zainal S, Dronov S, et al. Direct Transcriptional Consequences of Somatic Mutation in Breast Cancer. Cell Rep. 2016 Aug 16;16(7):2032–46.
86. Rhee JK, Lee S, Park WY, Kim YH, Kim TM. Allelic imbalance of somatic mutations in cancer genomes and transcriptomes. Sci Rep [Internet]. 2017 Dec 1 [cited 2022 Apr 11];7(1). Available from: https://pubmed.ncbi.nlm.nih.gov/28490743/
87. Spurr L, Li M, Alomran N, Zhang Q, Restrepo P, Movassagh M, et al. Systematic pan-cancer analysis of somatic allele frequency. Scientific Reports 2018 8:1 [Internet]. 2018 May 16 [cited 2022 May 19];8(1):1–12. Available from: https://www.nature.com/articles/s41598-018-25462-0
88. Restrepo P, Movassagh M, Alomran N, Miller C, Li M, Trenkov C, et al. Overexpressed somatic alleles are enriched in functional elements in Breast Cancer. Scientific Reports 2017 7:1 [Internet]. 2017 Aug 15 [cited 2022 May 19];7(1):1–10. Available from: https://www.nature.com/articles/s41598-017-08416-w
89. Martincorena I, Raine KM, Gerstung M, Dawson KJ, Haase K, van Loo P, et al. Universal Patterns of Selection in Cancer and Somatic Tissues. Cell. 2017 Nov 16;171(5):1029-1041.e21.
90. ICGC/TCGA Pan-Cancer Analysis of Whole Genomes Consortium. Pan-cancer analysis of whole genomes. Nature. 2020;578(7793):82–93.
91. Liu J, Lichtenberg T, Hoadley KA, Poisson LM, Lazar AJ, Cherniack AD, et al. An Integrated TCGA Pan-Cancer Clinical Data Resource to Drive High-Quality Survival Outcome Analytics. Cell. 2018 Apr 5;173(2):400-416.e11.
92. GTEx Consortium, Laboratory DA &Coordinating C (LDACC)—Analysis WG, Statistical Methods groups—Analysis Working Group, Enhancing GTEx (eGTEx) groups, NIH Common Fund, NIH/NCI, et al. Genetic effects on gene expression across human tissues. Nature. 2017;550(7675):204–13.
93. GTEx Consortium. Erratum: Genetic effects on gene expression across human tissues. Nature. 2018;553(7689):530.
94. Ji Y, Yu C, Zhang H. ContamDE-lm: Linear model-based differential gene expression analysis using next-generation RNA-seq data from contaminated tumor samples. Bioinformatics. 2020 Apr 15;36(8):2492–9.
95. Shen Q, Hu J, Jiang N, Hu X, Luo Z, Zhang H. ContamDE: Differential expression analysis of RNA-seq data for contaminated tumor samples. Bioinformatics. 2016 Mar 1;32(5):705–12.
96. Oros KK, Arcand SL, Bayani J, Squire JA, Mes-Masson AM, Tonin PN, et al. Analysis of genomic abnormalities in tumors: a review of available methods for Illumina two-color SNP genotyping and evaluation of performance. Cancer Genet. 2013 Apr 1;206(4):103–15.
97. Meldrum C, Doyle MA, Tothill RW. Next-generation sequencing for cancer diagnostics: a practical perspective. Clin Biochem Rev. 2011 Nov;32(4):177–95.
98. Bioinformatics Pipeline: DNA-Seq Analysis - GDC Docs [Internet]. [cited 2022 Apr 11]. Available from: https://docs.gdc.cancer.gov/Data/Bioinformatics_Pipelines/DNA_Seq_Variant_Calling_Pipeline/
99. Bioinformatics Pipeline: mRNA Analysis - GDC Docs [Internet]. [cited 2022 Apr 11]. Available from: https://docs.gdc.cancer.gov/Data/Bioinformatics_Pipelines/Expression_mRNA_Pipeline/

100. Analysis Pipelines Overview | ICGC ARGO Docs [Internet]. [cited 2022 Apr 11]. Available from: https://docs.icgc-argo.org/docs/analysis-workflows/analysis-overview
101. Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics [Internet]. 2009 Jul [cited 2022 Apr 11];25(14):1754–60. Available from: https://pubmed.ncbi.nlm.nih.gov/19451168/
102. Li H, Durbin R. Fast and accurate long-read alignment with Burrows-Wheeler transform. Bioinformatics [Internet]. 2010 Jan 15 [cited 2022 Apr 11];26(5):589–95. Available from: https://pubmed.ncbi.nlm.nih.gov/20080505/
103. Picard Tools - By Broad Institute [Internet]. [cited 2022 Apr 11]. Available from: https://broadinstitute.github.io/picard/
104. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, et al. The Sequence Alignment/Map format and SAMtools. Bioinformatics [Internet]. 2009 Aug [cited 2022 Apr 11];25(16):2078–9. Available from: https://pubmed.ncbi.nlm.nih.gov/19505943/
105. Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, et al. STAR: Ultrafast universal RNA-seq aligner. Bioinformatics. 2013 Jan;29(1):15–21.
106. García-Alcalde F, Okonechnikov K, Carbonell J, Cruz LM, Götz S, Tarazona S, et al. Qualimap: Evaluating next-generation sequencing alignment data. Bioinformatics. 2012 Oct;28(20):2678–9.
107. Okonechnikov K, Conesa A, García-Alcalde F. Qualimap 2: advanced multi-sample quality control for high-throughput sequencing data. Bioinformatics [Internet]. 2016 Jan 15 [cited 2022 Apr 11];32(2):292–4. Available from: https://pubmed.ncbi.nlm.nih.gov/26428292/
108. 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. Available from: https://doi.org/10.1101/201178
109. Benjamin D, Sato T, Cibulskis K, Getz G, Stewart C, Lichtenstein L. Calling Somatic SNVs and Indels with Mutect2. bioRxiv [Internet]. 2019 Dec 2 [cited 2022 Jun 10];861054. Available from: https://www.biorxiv.org/content/10.1101/861054v1
110. do Valle ÍF, Giampieri E, Simonetti G, Padella A, Manfrini M, Ferrari A, et al. Optimized pipeline of MuTect and GATK tools to improve the detection of somatic single nucleotide polymorphisms in whole-exome sequencing data. BMC Bioinformatics [Internet]. 2016 Nov 8 [cited 2022 Jun 10];17(12):27–35. Available from: https://bmcbioinformatics.biomedcentral.com/articles/10.1186/s12859-016-1190-7
111. Castel SE, Levy-Moonshine A, Mohammadi P, Banks E, Lappalainen T. Tools and best practices for data processing in allelic expression analysis. Genome Biol. 2015 Sep 17;16(1).
112. Nishida N, Koike A, Tajima A, Ogasawara Y, Ishibashi Y, Uehara Y, et al. Evaluating the performance of Affymetrix SNP Array 6.0 platform with 400 Japanese individuals. BMC Genomics [Internet]. 2008 Sep 22 [cited 2022 Apr 11];9. Available from: https://pubmed.ncbi.nlm.nih.gov/18803882/
113. Wang K, Li M, Hadley D, Liu R, Glessner J, Grant SFA, et al. PennCNV: An integrated hidden Markov model designed for high-resolution copy number variation detection in whole-genome SNP genotyping data. Genome Res. 2007 Nov;17(11):1665–74.
114. Raine KM, van Loo P, Wedge DC, Jones D, Menzies A, Butler AP, et al. ascatNgs: Identifying somatically acquired copy-number alterations from whole-genome sequencing data. Curr Protoc Bioinformatics. 2016;2016:15.9.1-15.9.17.

115. Olshen AB, Venkatraman ES, Lucito R, Wigler M. Circular binary segmentation for the analysis of array-based DNA copy number data. Biostatistics. 2004 Oct;5(4):557–72.
116. Diskin SJ, Li M, Hou C, Yang S, Glessner J, Hakonarson H, et al. Adjustment of genomic waves in signal intensities from whole-genome SNP genotyping platforms. Nucleic Acids Res. 2008;36(19).
117. Anders S, Pyl PT, Huber W. HTSeq--a Python framework to work with high-throughput sequencing data. Bioinformatics [Internet]. 2015 Jan 15 [cited 2022 Apr 11];31(2):166–9. Available from: https://pubmed.ncbi.nlm.nih.gov/25260700/
118. Love M, Anders S, Huber W. Differential analysis of count data-the DESeq2 package. 2014.
119. Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol. 2014 Dec 5;15(12).
120. Cline MS, Smoot M, Cerami E, Kuchinsky A, Landys N, Workman C, et al. Integration of biological networks and gene expression data using cytoscape. Nat Protoc. 2007;2(10):2366–82.
121. Warde-Farley D, Donaldson SL, Comes O, Zuberi K, Badrawi R, Chao P, et al. The GeneMANIA prediction server: Biological network integration for gene prioritization and predicting gene function. Nucleic Acids Res. 2010 Jun 21;38(SUPPL. 2).
122. Mostafavi S, Ray D, Warde-Farley D, Grouios C, Morris Q. GeneMANIA: a real-time multiple association network integration algorithm for predicting gene function [Internet]. Vol. 9. 2008. Available from: http://genomebiology.com/2008/9/S1/S4
123. Chen X, Li D. ERVcaller: Identifying polymorphic endogenous retrovirus and other transposable element insertions using whole-genome sequencing data. Bioinformatics. 2019 Oct 15;35(20):3913–22.
124. Corces MR, Granja JM, Shams S, Louie BH, Seoane JA, Zhou W, et al. The chromatin accessibility landscape of primary human cancers. Science (1979). 2018 Oct 26;362(6413).
125. Grishin D, Gusev A. Allelic imbalance of chromatin accessibility in cancer identifies candidate causal risk variants and their mechanisms. Nature Genetics 2022 54:6 [Internet]. 2022 Jun 13 [cited 2022 Jun 23];54(6):837–49. Available from: https://www.nature.com/articles/s41588-022-01075-2
126. Campbell PJ, Getz G, Korbel JO, Stuart JM, Jennings JL, Stein LD, et al. Pan-cancer analysis of whole genomes. Nature 2020 578:7793 [Internet]. 2020 Feb 5 [cited 2022 Jun 10];578(7793):82–93. Available from: https://www.nature.com/articles/s41586-020-1969-6
127. Weinstein JN, Collisson EA, Mills GB, Shaw KRM, Ozenberger BA, Ellrott K, et al. The Cancer Genome Atlas Pan-Cancer analysis project [Internet]. Vol. 45, Nature Genetics. Nature; 2013 [cited 2022 Jun 10]. p. 1113–20. Available from: https://www.nature.com/articles/ng.2764
128. Ogasawara O, Kodama Y, Mashima J, Kosuge T, Fujisawa T. DDBJ Database updates and computational infrastructure enhancement. Nucleic Acids Res [Internet]. 2020 Jan 1 [cited 2022 Apr 11];48(D1):D45–50. Available from: https://pubmed.ncbi.nlm.nih.gov/31724722/
129. Kaminuma E, Mashima J, Kodama Y, Gojobori T, Ogasawara O, Okubo K, et al. DDBJ launches a new archive database with analytical tools for next-generation sequence data. Nucleic Acids Res [Internet]. 2010 Oct 22 [cited 2022 Apr 11];38(Database issue). Available from: https://pubmed.ncbi.nlm.nih.gov/19850725/
130. Cibulskis K, Lawrence MS, Carter SL, Sivachenko A, Jaffe D, Sougnez C, et al. Sensitive detection of somatic point mutations in impure and heterogeneous cancer samples. Nat Biotechnol. 2013 Mar;31(3):213–9.

131. Bamford S, Dawson E, Forbes S, Clements J, Pettett R, Dogan A, et al. The COSMIC (Catalogue of Somatic Mutations in Cancer) database and website. Br J Cancer [Internet]. 2004 Jul 19 [cited 2022 Apr 11];91(2):355–8. Available from: https://pubmed.ncbi.nlm.nih.gov/15188009/

132. Oesper L, Mahmoody A, Raphael BJ. THetA: Inferring intra-tumor heterogeneity from high-throughput DNA sequencing data. Genome Biol. 2013;14(7).

133. Snijders AM, Nowak N, Segraves R, Blackwood S, Brown N, Conroy J, et al. Assembly of microarrays for genome-wide measurement of DNA copy number. Nat Genet. 2001;29(3):263–4.

134. Aran D, Sirota M, Butte AJ. Systematic pan-cancer analysis of tumour purity. Nat Commun. 2015 Dec 4;6.

135. Carter SL, Cibulskis K, Helman E, McKenna A, Shen H, Zack T, et al. Absolute quantification of somatic DNA alterations in human cancer. Nat Biotechnol. 2012 May;30(5):413–21.

136. Mayba O, Gilbert HN, Liu J, Haverty PM, Jhunjhunwala S, Jiang Z, et al. MBASED: Allele-specific expression detection in cancer tissues and cell lines. Genome Biol. 2014 Aug 7;15(8).

137. Wood DLA, Nones K, Steptoe A, Christ A, Harliwong I, Newell F, et al. Recommendations for Accurate Resolution of Gene and Isoform Allele-Specific Expression in RNA-Seq Data. PLoS One [Internet]. 2015 May 12 [cited 2022 Apr 11];10(5). Available from: https://pubmed.ncbi.nlm.nih.gov/25965996/

138. Lawrence MS, Stojanov P, Polak P, Kryukov G v, Cibulskis K, Sivachenko A, et al. Mutational heterogeneity in cancer and the search for new cancer-associated genes. Nature. 2013 Jul 11;499(7457):214–8.

139. Liu Y, Chen C, Xu Z, Scuoppo C, Rillahan CD, Gao J, et al. Deletions linked to TP53 loss drive cancer through p53-independent mechanisms. Nature [Internet]. 2016 Mar 24 [cited 2022 Apr 11];531(7595):471–5. Available from: https://pubmed.ncbi.nlm.nih.gov/26982726/

140. Chan EF, Gat U, McNiff JM, Fuchs E. A common human skin tumour is caused by activating mutations in beta-catenin. Nat Genet [Internet]. 1999 Apr [cited 2022 Apr 11];21(4):410–3. Available from: https://pubmed.ncbi.nlm.nih.gov/10192393/

141. Lindhurst MJ, Parker VER, Payne F, Sapp JC, Rudge S, Harris J, et al. Mosaic overgrowth with fibroadipose hyperplasia is caused by somatic activating mutations in PIK3CA. Nat Genet [Internet]. 2012 Aug [cited 2022 Apr 11];44(8):928–33. Available from: https://pubmed.ncbi.nlm.nih.gov/22729222/

142. Davies H, Bignell GR, Cox C, Stephens P, Edkins S, Clegg S, et al. Mutations of the BRAF gene in human cancer. Nature [Internet]. 2002 Jun 27 [cited 2022 Apr 11];417(6892):949–54. Available from: https://pubmed.ncbi.nlm.nih.gov/12068308/

143. Almoguera C, Shibata D, Forrester K, Martin J, Arnheim N, Perucho M. Most human carcinomas of the exocrine pancreas contain mutant c-K-ras genes. Cell [Internet]. 1988 May 20 [cited 2022 Apr 11];53(4):549–54. Available from: https://pubmed.ncbi.nlm.nih.gov/2453289/

144. Matsuoka S, Ballif BA, Smogorzewska A, McDonald ER, Hurov KE, Luo J, et al. ATM and ATR substrate analysis reveals extensive protein networks responsive to DNA damage. Science [Internet]. 2007 May 25 [cited 2022 Apr 11];316(5828):1160–6. Available from: https://pubmed.ncbi.nlm.nih.gov/17525332/

145. Li J, Yen C, Liaw D, Podsypanina K, Bose S, Wang SI, et al. PTEN, a putative protein tyrosine phosphatase gene mutated in human brain, breast, and prostate cancer. Science [Internet]. 1997 Mar 28 [cited 2022 Apr 11];275(5308):1943–7. Available from: https://pubmed.ncbi.nlm.nih.gov/9072974/

146. Maniatis T, Reed[2] R. An extensive network of coupling among gene expression machines [Internet]. Vol. 416, NATURE. 2002. Available from: www.nature.com
147. Mathur R, Alver BH, San Roman AK, Wilson BG, Wang X, Agoston AT, et al. ARID1A loss impairs enhancer-mediated gene regulation and drives colon cancer in mice. Nat Genet [Internet]. 2017 Jan 31 [cited 2022 Apr 11];49(2):296–302. Available from: https://pubmed.ncbi.nlm.nih.gov/27941798/
148. Helming KC, Wang X, Wilson BG, Vazquez F, Haswell JR, Manchester HE, et al. ARID1B is a specific vulnerability in ARID1A-mutant cancers. Nat Med [Internet]. 2014 [cited 2022 Apr 11];20(3):251–4. Available from: https://pubmed.ncbi.nlm.nih.gov/24562383/
149. Zhao B, Pritchard JR. Inherited Disease Genetics Improves the Identification of Cancer-Associated Genes. PLoS Genet [Internet]. 2016 Jun 1 [cited 2022 Apr 11];12(6). Available from: https://pubmed.ncbi.nlm.nih.gov/27304678/
150. Warde N. Inflammation: Cadherin 11: a key mediator of fibroblast inflammation. Nat Rev Rheumatol [Internet]. 2011 Jul [cited 2022 Apr 11];7(7):374. Available from: https://pubmed.ncbi.nlm.nih.gov/21691325/
151. Morano I, Chai GX, Baltas LG, Lamounier-Zepter V, Lutsch G, Kott M, et al. Smooth-muscle contraction without smooth-muscle myosin. Nat Cell Biol. 2000 Jun 12;2(6):371–5.
152. Andrae J, Gallini R, Betsholtz C. Role of platelet-derived growth factors in physiology and medicine. Genes Dev [Internet]. 2008 May 15 [cited 2022 Apr 11];22(10):1276–312. Available from: https://pubmed.ncbi.nlm.nih.gov/18483217/
153. Bhattacharyya S, Wang W, Morales-Nebreda L, Feng G, Wu M, Zhou X, et al. Tenascin-C drives persistence of organ fibrosis. Nat Commun [Internet]. 2016 Jun 3 [cited 2022 Apr 11];7. Available from: https://pubmed.ncbi.nlm.nih.gov/27256716/
154. Higgs R. PTPRC mutation associated with response to anti-tNF therapy in rheumatoid arthritis. Nat Rev Rheumatol [Internet]. 2010 Jun [cited 2022 Apr 11];6(6):311. Available from: https://pubmed.ncbi.nlm.nih.gov/20527672/
155. Payne GS, Bishop JM, Varmus HE. Multiple arrangements of viral DNA and an activated host oncogene in bursal lymphomas. Nature [Internet]. 1982 [cited 2022 Apr 11];295(5846):209–14. Available from: https://pubmed.ncbi.nlm.nih.gov/6276760/
156. Kong Y, Rose CM, Cass AA, Williams AG, Darwish M, Lianoglou S, et al. Transposable element expression in tumors is associated with immune infiltration and increased antigenicity. Nat Commun. 2019 Dec 1;10(1).
157. Merkenschlager M, Odom DT. CTCF and cohesin: linking gene regulatory elements with their targets. Cell [Internet]. 2013 Mar 14 [cited 2022 Apr 11];152(6):1285–97. Available from: https://pubmed.ncbi.nlm.nih.gov/23498937/
158. Slotkin RK, Martienssen R. Transposable elements and the epigenetic regulation of the genome. Nat Rev Genet [Internet]. 2007 Apr [cited 2022 Apr 11];8(4):272–85. Available from: https://pubmed.ncbi.nlm.nih.gov/17363976/
159. Deniz Ö, Frost JM, Branco MR. Regulation of transposable elements by DNA modifications. Nat Rev Genet [Internet]. 2019 Jul 1 [cited 2022 Apr 11];20(7):417–31. Available from: https://pubmed.ncbi.nlm.nih.gov/30867571/

160. Rooney MS, Shukla SA, Wu CJ, Getz G, Hacohen N. Molecular and genetic properties of tumors associated with local immune cytolytic activity. Cell [Internet]. 2015 Jan 15 [cited 2022 Apr 11];160(1–2):48–61. Available from: https://pubmed.ncbi.nlm.nih.gov/25594174/

161. Smith CC, Beckermann KE, Bortone DS, Cubas AA, Bixby LM, Lee SJ, et al. Endogenous retroviral signatures predict immunotherapy response in clear cell renal cell carcinoma. J Clin Invest [Internet]. 2018 Nov 1 [cited 2022 Apr 11];128(11):4804–20. Available from: https://pubmed.ncbi.nlm.nih.gov/30137025/

162. Solovyov A, Vabret N, Arora KS, Snyder A, Funt SA, Bajorin DF, et al. Global Cancer Transcriptome Quantifies Repeat Element Polarization between Immunotherapy Responsive and T Cell Suppressive Classes. Cell Rep [Internet]. 2018 Apr 10 [cited 2022 Apr 11];23(2):512–21. Available from: https://pubmed.ncbi.nlm.nih.gov/29642008/

163. Panda A, de Cubas AA, Stein M, Riedlinger G, Kra J, Mayer T, et al. Endogenous retrovirus expression is associated with response to immune checkpoint blockade in clear cell renal cell carcinoma. JCI Insight [Internet]. 2018 Aug 23 [cited 2022 Apr 11];3(16). Available from: https://pubmed.ncbi.nlm.nih.gov/30135306/

164. Attig J, Young GR, Hosie L, Perkins D, Encheva-Yokoya V, Stoye JP, et al. LTR retroelement expansion of the human cancer transcriptome and immunopeptidome revealed by de novo transcript assembly. Genome Res [Internet]. 2019 [cited 2022 Apr 11];29(10):1578–90. Available from: https://pubmed.ncbi.nlm.nih.gov/31537638/

165. Selvarajan I, Toropainen A, Garske KM, López Rodríguez M, Ko A, Miao Z, et al. Integrative analysis of liver-specific non-coding regulatory SNPs associated with the risk of coronary artery disease. The American Journal of Human Genetics. 2021 Mar 4;108(3):411–30.

166. Borel C, Ferreira PG, Santoni F, Delaneau O, Fort A, Popadin KY, et al. Biased allelic expression in human primary fibroblast single cells. Am J Hum Genet. 2015 Jan 8;96(1):70–80.

167. SRA-Toolkit [Internet]. [cited 2022 Apr 11]. Available from: https://hpc.nih.gov/apps/sratoolkit.html

168. Yuan S, Liao G, Zhang M, Zhu Y, Xiao W, Wang K, et al. Multiomics interrogation into HBV (Hepatitis B virus)-host interaction reveals novel coding potential in human genome, and identifies canonical and non-canonical proteins as host restriction factors against HBV. Cell Discov [Internet]. 2021 Dec 1 [cited 2022 Apr 19];7(1). Available from: https://pubmed.ncbi.nlm.nih.gov/34725333/

169. Kawamoto M, Yamaji T, Saito K, Shirasago Y, Satomura K, Endo T, et al. Identification of Characteristic Genomic Markers in Human Hepatoma HuH-7 and Huh7.5.1-8 Cell Lines. Front Genet [Internet]. 2020 Oct 9 [cited 2022 Apr 19];11. Available from: https://pubmed.ncbi.nlm.nih.gov/33193621/

170. MacParland SA, Liu JC, Ma XZ, Innes BT, Bartczak AM, Gage BK, et al. Single cell RNA sequencing of human liver reveals distinct intrahepatic macrophage populations. Nature Communications 2018 9:1 [Internet]. 2018 Oct 22 [cited 2022 Jun 12];9(1):1–21. Available from: https://www.nature.com/articles/s41467-018-06318-7

171. Losic B, Craig AJ, Villacorta-Martin C, Martins-Filho SN, Akers N, Chen X, et al. Intratumoral heterogeneity and clonal evolution in liver cancer. Nature Communications 2020 11:1 [Internet]. 2020 Jan 15 [cited 2022 Jun 12];11(1):1–15. Available from: https://www.nature.com/articles/s41467-019-14050-z

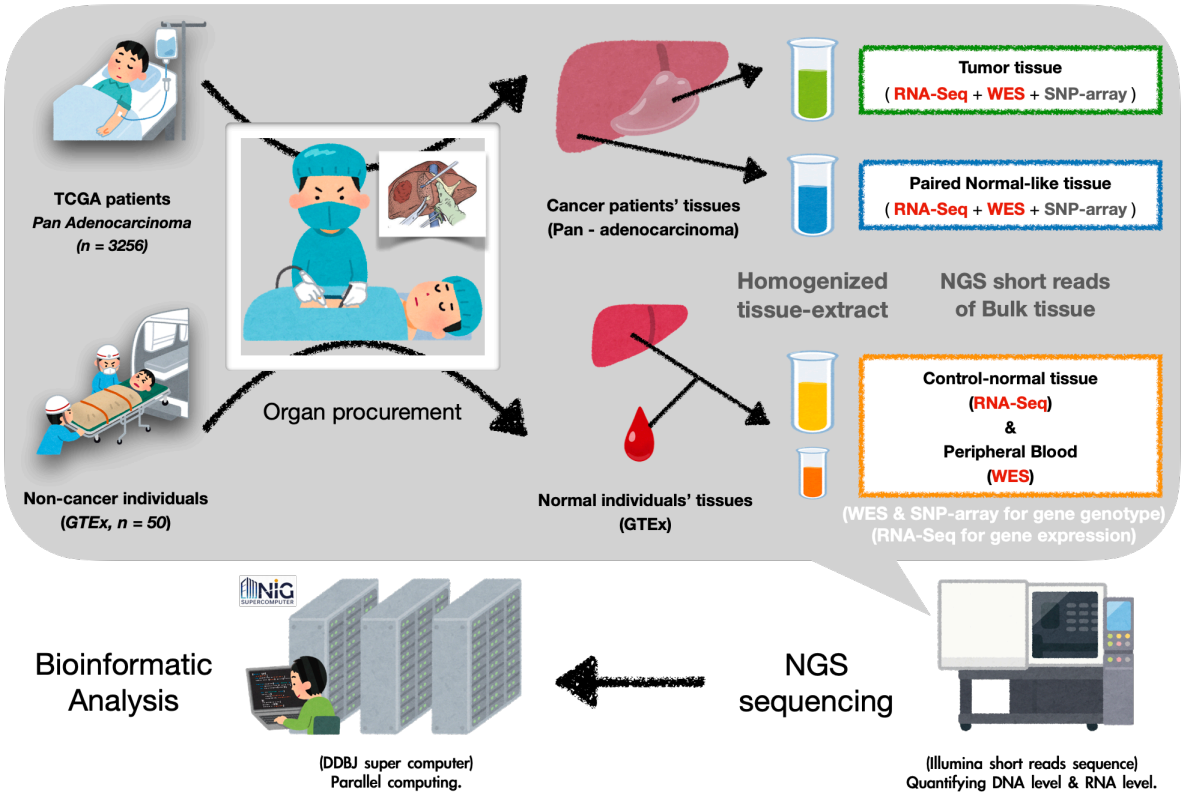
172. Dong X, Wang F, Liu C, Ling J, Jia X, Shen F, et al. Single-cell analysis reveals the intra-tumor heterogeneity and identifies MLXIPL as a biomarker in the cellular trajectory of hepatocellular

carcinoma. Cell Death Discovery 2021 7:1 [Internet]. 2021 Jan 18 [cited 2022 Dec 12];7(1):1–13. Available from: https://www.nature.com/articles/s41420-021-00403-5

173. Hagemann-Jensen M, Ziegenhain C, Chen P, Ramsköld D, Hendriks GJ, Larsson AJM, et al. Single-cell RNA counting at allele and isoform resolution using Smart-seq3. Nat Biotechnol. 2020 Jun 1;38(6):708–14.
174. Picelli S, Faridani OR, Björklund ÅK, Winberg G, Sagasser S, Sandberg R. Full-length RNA-seq from single cells using Smart-seq2. Nat Protoc. 2014 Jan;9(1):171–81.
175. Sato M, Suzuki S, Senoo H. Hepatic Stellate Cells: Unique Characteristics in Cell Biology and Phenotype. Cell Struct Funct. 2003;28(2):105–12.
176. Borel C, Ferreira PG, Santoni F, Delaneau O, Fort A, Popadin KY, et al. Biased Allelic Expression in Human Primary Fibroblast Single Cells. The American Journal of Human Genetics. 2015 Jan 8;96(1):70–80.
177. Choi K, Raghupathy N, Churchill GA. A Bayesian mixture model for the analysis of allelic expression in single cells. Nature Communications 2019 10:1 [Internet]. 2019 Nov 15 [cited 2022 Jun 12];10(1):1–11. Available from: https://www.nature.com/articles/s41467-019-13099-0
178. Barry AE, Baldeosingh R, Lamm R, Patel K, Zhang K, Dominguez DA, et al. Hepatic Stellate Cells and Hepatocarcinogenesis. Front Cell Dev Biol. 2020 Aug 5;8:709.
179. Orphanides G, Reinberg D. A Unified Theory of Gene Expression. Cell. 2002 Feb 22;108(4):439–51.
180. Rao SSP, Huntley MH, Durand NC, Stamenova EK, Bochkov ID, Robinson JT, et al. A 3D Map of the Human Genome at Kilobase Resolution Reveals Principles of Chromatin Looping. Cell. 2014 Dec 18;159(7):1665–80.
181. Khoury A, Achinger-Kawecka J, Bert SA, Smith GC, French HJ, Luu PL, et al. Constitutively bound CTCF sites maintain 3D chromatin architecture and long-range epigenetically regulated domains. Nat Commun. 2020 Dec 1;11(1).
182. Tang Z, Luo OJ, Li X, Zheng M, Zhu JJ, Szalaj P, et al. CTCF-Mediated Human 3D Genome Architecture Reveals Chromatin Topology for Transcription. Cell. 2015 Dec 17;163(7):1611–27.
183. Chen H, Li C, Peng X, Zhou Z, Weinstein JN, Caesar-Johnson SJ, et al. A Pan-Cancer Analysis of Enhancer Expression in Nearly 9000 Patient Samples. Cell. 2018 Apr 5;173(2):386-399.e12.
184. Hnisz D, Abraham BJ, Lee TI, Lau A, Saint-André V, Sigova AA, et al. Super-Enhancers in the Control of Cell Identity and Disease. Cell. 2013 Nov;155(4):934–47.
185. Burton PR, Clayton DG, Cardon LR, Craddock N, Deloukas P, Duncanson A, et al. Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. Nature 2007 447:7145 [Internet]. 2007 Jun 7 [cited 2022 Jun 12];447(7145):661–78. Available from: https://www.nature.com/articles/nature05911
186. Risch N, Merikangas K. The Future of Genetic Studies of Complex Human Diseases. Science (1979) [Internet]. 1996 [cited 2022 Jun 12];273(5281):1516–7. Available from: https://www.science.org/doi/abs/10.1126/science.273.5281.1516
187. Morloy M, Molony CM, Weber TM, Devlin JL, Ewens KG, Spielman RS, et al. Genetic analysis of genome-wide variation in human gene expression. Nature 2004 430:7001 [Internet]. 2004 Jul 21 [cited 2022 Jun 12];430(7001):743–7. Available from: https://www.nature.com/articles/nature02797

188. Cookson W, Liang L, Abecasis G, Moffatt M, Lathrop M. Mapping complex disease traits with global gene expression. Nature Reviews Genetics 2009 10:3 [Internet]. 2009 Mar [cited 2022 Jun 12];10(3):184–94. Available from: https://www.nature.com/articles/nrg2537
189. Schadt EE, Monks SA, Drake TA, Lusis AJ, Che N, Colinayo V, et al. Genetics of gene expression surveyed in maize, mouse and man. Nature 2003 422:6929 [Internet]. 2003 Mar 20 [cited 2022 Jun 12];422(6929):297–302. Available from: https://www.nature.com/articles/nature01434
190. French JD, Edwards SL. The Role of Noncoding Variants in Heritable Disease. Trends in Genetics. 2020 Nov 1;36(11):880–91.
191. Chuong EB, Elde NC, Feschotte C. Regulatory activities of transposable elements: from conflicts to benefits. Nature Reviews Genetics 2016 18:2 [Internet]. 2016 Nov 21 [cited 2022 Jun 12];18(2):71–86. Available from: https://www.nature.com/articles/nrg.2016.139
192. Kunarso G, Chia NY, Jeyakani J, Hwang C, Lu X, Chan YS, et al. Transposable elements have rewired the core regulatory network of human embryonic stem cells. Nature Genetics 2010 42:7 [Internet]. 2010 Jun 6 [cited 2022 Jun 12];42(7):631–4. Available from: https://www.nature.com/articles/ng.600
193. Zhang Y, Li T, Preissl S, Amaral ML, Grinstein JD, Farah EN, et al. Transcriptionally active HERV-H retrotransposons demarcate topologically associating domains in human pluripotent stem cells. Nat Genet. 2019 Sep 1;51(9):1380–8.
194. Shen JZ, Qiu Z, Wu Q, Finlay D, Garcia G, Sun D, et al. FBXO44 promotes DNA replication-coupled repetitive element silencing in cancer cells. Cell. 2021 Jan 21;184(2):352-369.e23.
195. Ito J, Kimura I, Soper A, Coudray A, Koyanagi Y, Nakaoka H, et al. Endogenous retroviruses drive KRAB zinc-finger protein family expression for tumor suppression [Internet]. Vol. 6, Sci. Adv. 2020. Available from: http://advances.sciencemag.org/
196. Lynch-Sutherland CF, Chatterjee A, Stockwell PA, Eccles MR, Macaulay EC. Reawakening the Developmental Origins of Cancer Through Transposable Elements. Front Oncol. 2020 May 5;10:468.
197. Rodriguez-Martin B, Alvarez EG, Baez-Ortega A, Zamora J, Supek F, Demeulemeester J, et al. Pan-cancer analysis of whole genomes identifies driver rearrangements promoted by LINE-1 retrotransposition. Nat Genet. 2020 Mar 1;52(3):306–19.

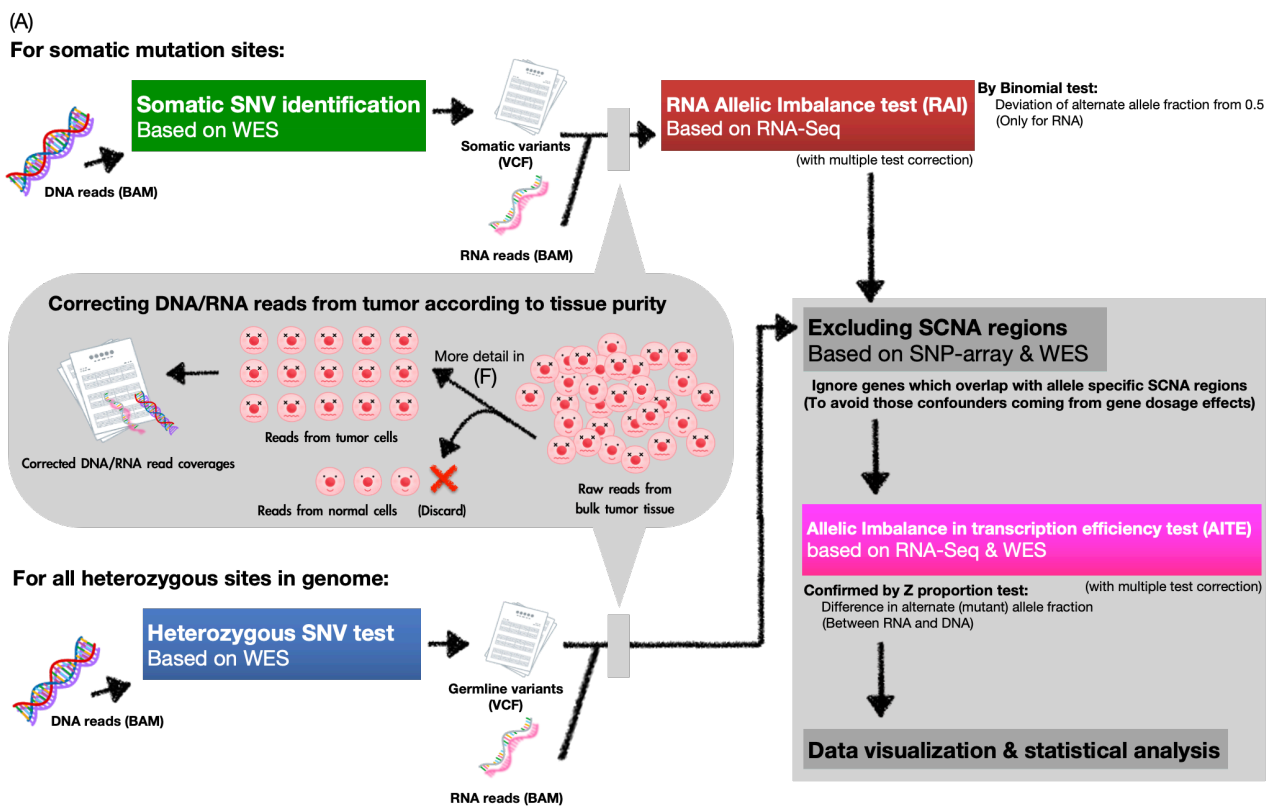
APPENDIX A: SUPPLEMENTARY FIGURES



Supplement Figure 1. Methods for data collecting and sample grouping by tissue type.

Various types of sequence data were downloaded from different online databases and being grouped into tumor tissues, paired-normal-like (paired-solid-normal) tissues, and control-normal tissues groups for statistical comparisons. Among them, tumor tissues and paired-normal-like tissues were collected from the same cancer patients by TCGA consortium; and the control-normal tissues were collected from non-cancer individuals by GTEx consortium. For the downstream bioinformatics analyzing steps, the downloaded Illumina short-read sequence data and SNP-array data were fed into computer pipelines under the NIG super computer platform for the parallel computations.

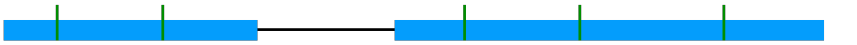
(A)



(B) **Somatic SNV identification** in tumor tissue:

1. (a) Finding the somatic SNVs positions:
(Based on WES reads from tumor and paired blood tissue**)**

Somatic SNVs positions by MuTect2:

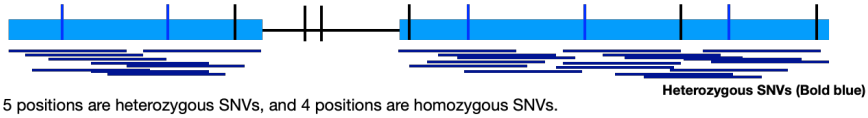


PS. All of the alternate SNVs detected here are somatically derived. Somatically derived SNVs (Bold green)

(C) **Heterozygous SNV test** for human common SNVs positions in each type of tissue:

1. (b) Finding the het. positions by checking whether the alternate allele fraction is biased from 0.5:
(Based on WES reads from tumor and paired normal-like tissue respectively**)**

Binomial test: **P>0.05** **P>0.05** P<0.05 P<0.05 **P>0.05** **P>0.05** P<0.05 **P>0.05** P<0.05



Heterozygous SNVs (Bold blue)

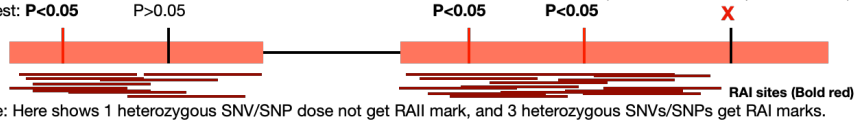
Here shows 5 positions are heterozygous SNVs, and 4 positions are homozygous SNVs.
PS. Theoretically, most of the alternate SNVs being detected here should be inherited from germ-line.

(D) **RNA Allelic Imbalance test** for those somatic SNVs / heterozygous SNVs positions:

2. Threshold control for checking the RNA read coverages in those target SNV sites (expression check):
For those candidate SNVs, if their RNA read coverage is lower than 20, they will be removed.
3. Check whether the alternate allele fraction is deviated away from 0.5 for each SNV sites:
(Based on RNA-Seq reads)

PS. **Benjamini & Hochberg FDR** is used to remove false positive SNVs through adjusting the P-values.

Statistical test: **P<0.05** P>0.05 **P<0.05** **P<0.05** No enough coverage (This het. SNV doesn't pass Thr. control)



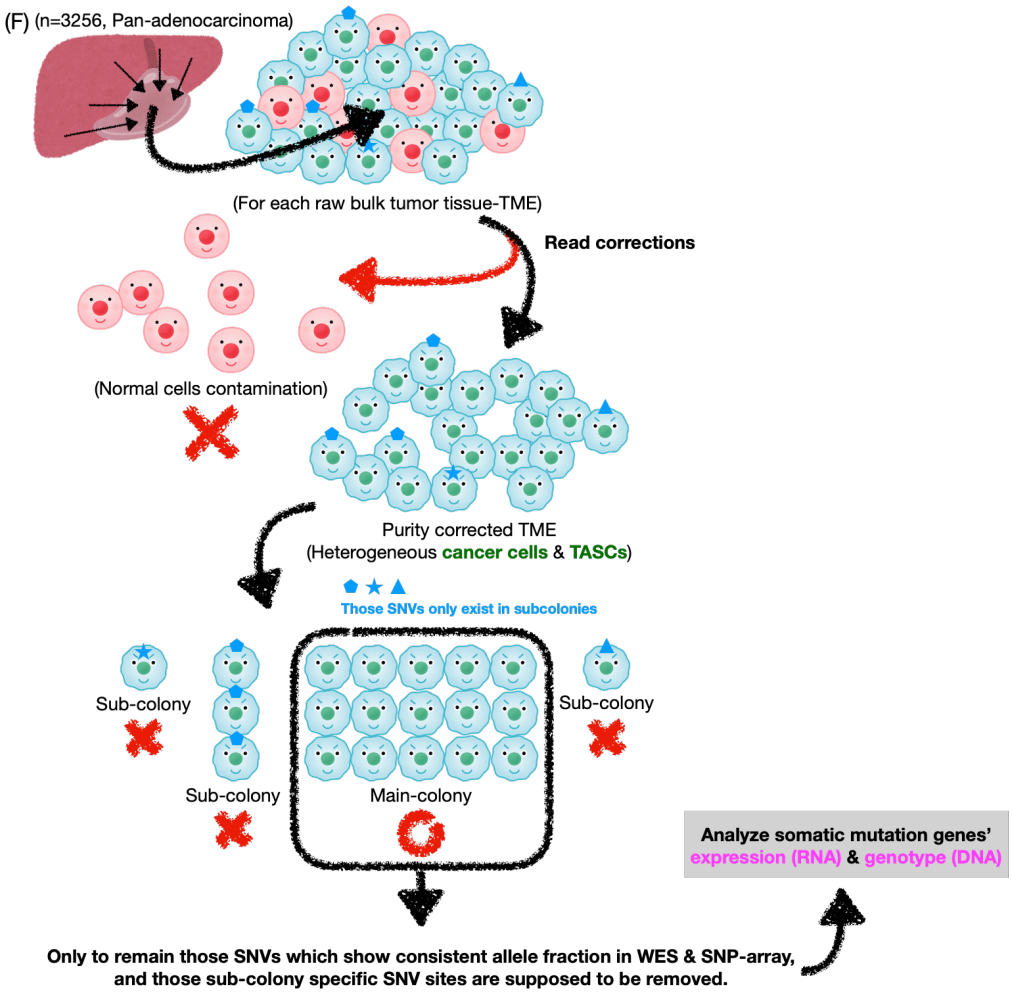
RAI sites (Bold red)

For example: Here shows 1 heterozygous SNV/SNP dose not get RAII mark, and 3 heterozygous SNVs/SNPs get RAI marks.

4. Calculated the percentage of RAI marked SNVs among those Thr. Control passed SNVs:
For example, the above gene shows 75% (3/4) of sites are in marks, which is above 50%. Therefore, it is supposed to be RAI gene.
(In fact, for most of genes, only one expressed heterozygous site can pass the threshold with enough read coverage.)

(E) **Allelic Imbalance in Transcriptional Efficiency (AITE)** confirmation for RAI genes:

5. Excluding those RAI genes which show SCNA (may further exist LOH) reported by SNP array, and then check allelic ratio difference.
Z proportion test is applied to double check the inconsistency of "alternate allele fraction" between RNA and DNA reads.



Supplement Figure 2. Bioinformatics workflow for detecting RAI effect and AITE effect in TCGA bulk sequencing data.

(A) Here shows the bioinformatics workflows to detect above described RAI, AITE, and SCNA. Firstly, somatically derived mutation sites were detected using mutect2 software. Only those single-nucleotide variant (SNV) sites were retained for further analysis. In contrast, germline-derived heterozygous sites were detected using the GATK-haplotype caller, and all false-positive sites caused by poor mapping quality were filtered out. Only validated SNV sites were retained for further analysis. Next, by referring to the paired SNP-array data of each tumor sample, the read coverage of both WESeq and RNA-Seq at every target SNV site were corrected according to the tumor purity and SCNA information reported by ASCAT software. Finally, after obtaining the corrected read coverage at each target SNV sites. RAI was detected using binomial tests with a BH multiple test correction step. The RAI sites overlapping with DNA allelic imbalance regions (existing SCNA and LOH) were further excluded from the next step of AITE investigation.

(B) The somatically derived mutation sites were detected by mutect2 software. Only those single nucleotide variants (SNV) sites were remained for next step analysis.

(C) The germline-derived heterozygous sites were detected by GATK haplotype caller. All of those false positive sites caused by poor mapping quality had been filtered out, and also only those single nucleotide variants (SNV) sites were remained for next step analysis.

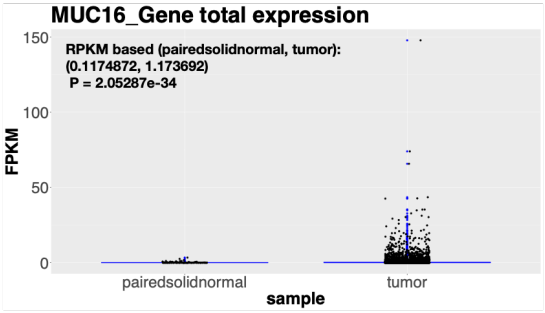
(D) By referring to the paired SNP-array data of each tumor sample, the read coverages of both WESeq and RNA-Seq at every target SNV sites were corrected according to the tumor purity and SCNA information reported by

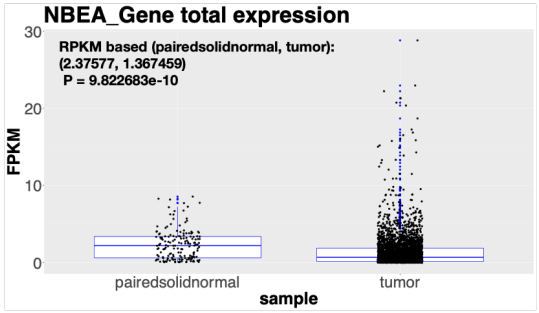
ASCAT software. After getting the corrected allele-specific reads coverages at each target SNV sites. Allelic Imbalance in RNA (RAI) was detected through the Binomial test along with BH multiple test correction.

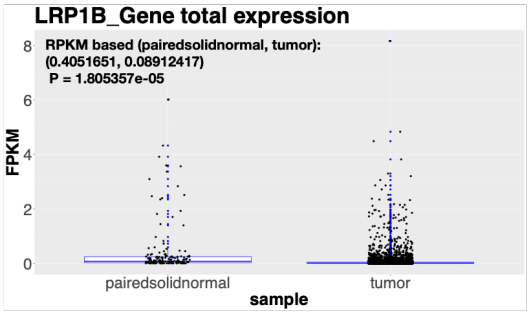
(E) Then, among those RNA allelic imbalance (RAI) sites overlapping with DNA allelic imbalance regions (existing SCNA or LOH), they were further excluded for the step of allelic imbalance in transcription efficiency (AITE) examination, here.

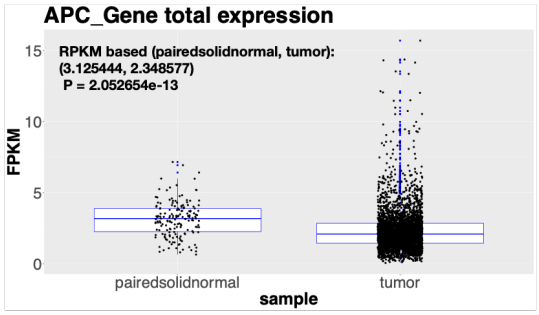
(F) More details in how to deal with those few somatic mutation sites carried by sub-colonies when correcting the read coverages.

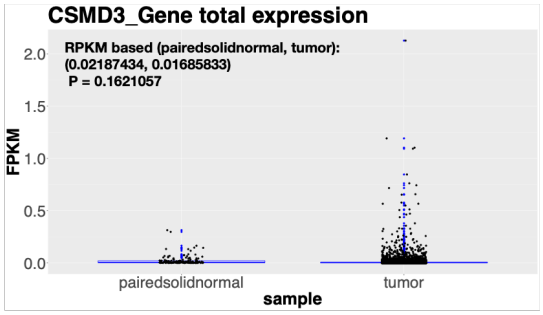
(A)

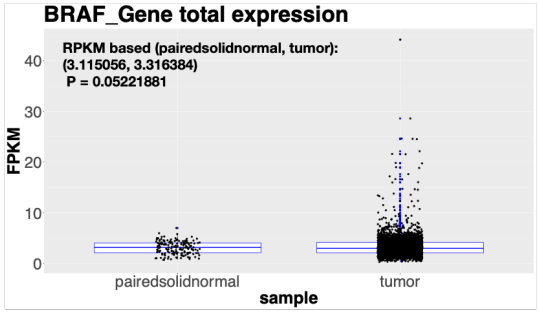

MUC16_Gene total expression
RPKM based (pairedsolidnormal, tumor):
(0.1174872, 1.173692)
P = 2.05287e-34
FPKM
150
100
50
0
pairedsolidnormal
tumor
sample

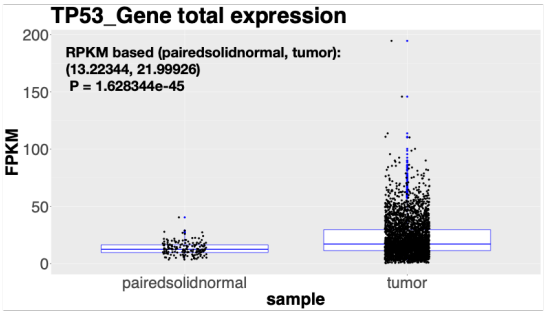

NBEA_Gene total expression
RPKM based (pairedsolidnormal, tumor):
(2.37577, 1.367459)
P = 9.822683e-10
FPKM
30
20
10
0
pairedsolidnormal
tumor
sample

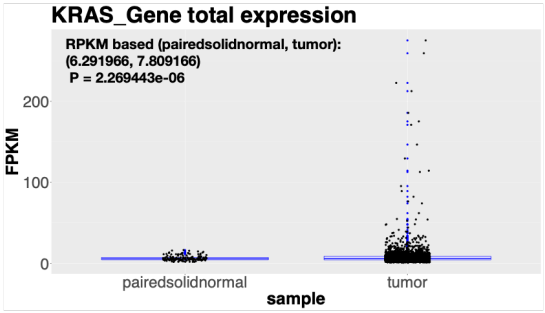

LRP1B_Gene total expression
RPKM based (pairedsolidnormal, tumor):
(0.4051651, 0.08912417)
P = 1.805357e-05
FPKM
8
6
4
2
0
pairedsolidnormal
tumor
sample

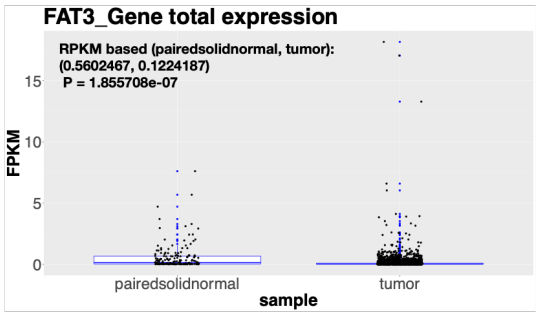

APC_Gene total expression
RPKM based (pairedsolidnormal, tumor):
(3.125444, 2.348577)
P = 2.052654e-13
FPKM
15
10
5
0
pairedsolidnormal
tumor
sample

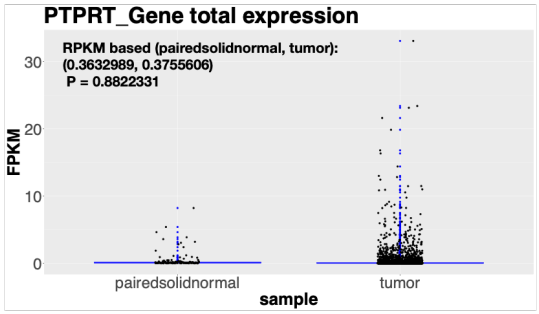

CSMD3_Gene total expression
RPKM based (pairedsolidnormal, tumor):
(0.02187434, 0.01685833)
P = 0.1621057
FPKM
2.0
1.5
1.0
0.5
0.0
pairedsolidnormal
tumor
sample

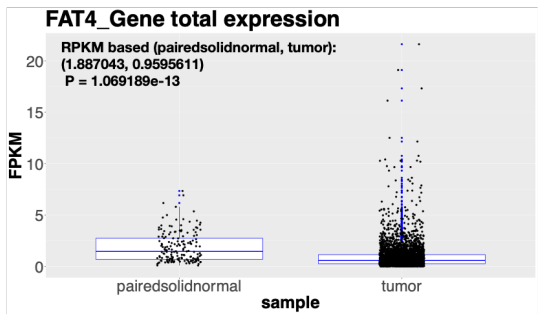

BRAF_Gene total expression
RPKM based (pairedsolidnormal, tumor):
(3.115056, 3.316384)
P = 0.05221881
FPKM
40
30
20
10
0
pairedsolidnormal
tumor
sample

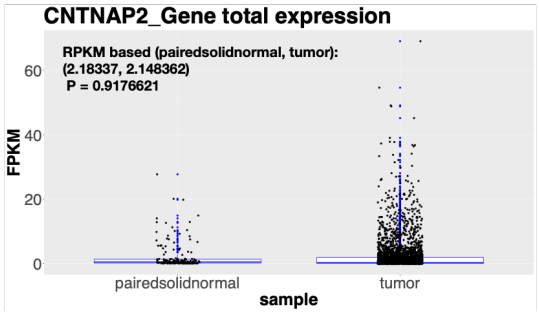

TP53_Gene total expression
RPKM based (pairedsolidnormal, tumor):
(13.22344, 21.99926)
P = 1.628344e-45
FPKM
200
150
100
50
0
pairedsolidnormal
tumor
sample

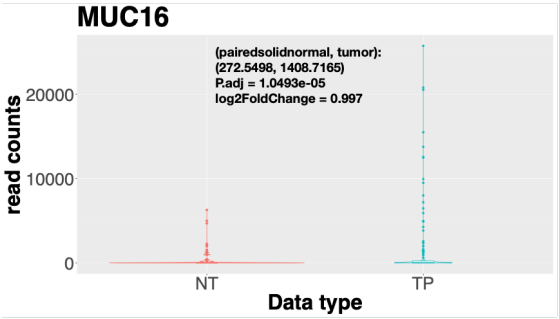

KRAS_Gene total expression
RPKM based (pairedsolidnormal, tumor):
(6.291966, 7.809166)
P = 2.269443e-06
FPKM
200
100
0
pairedsolidnormal
tumor
sample

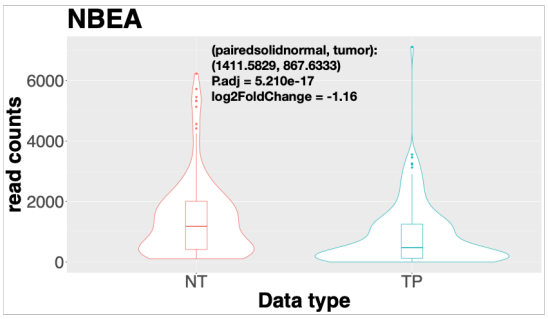

FAT3_Gene total expression
RPKM based (pairedsolidnormal, tumor):
(0.5602467, 0.1224187)
P = 1.855708e-07
FPKM
15
10
5
0
pairedsolidnormal
tumor
sample

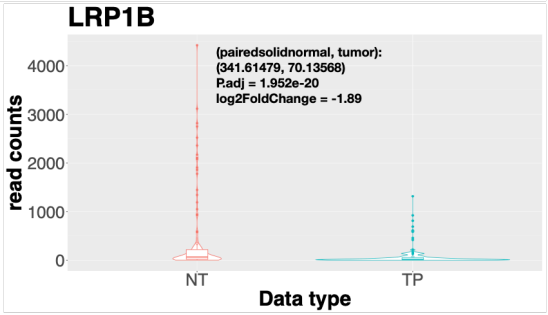

PTPRT_Gene total expression
RPKM based (pairedsolidnormal, tumor):
(0.3632989, 0.3755606)
P = 0.8822331
FPKM
30
20
10
0
pairedsolidnormal
tumor
sample

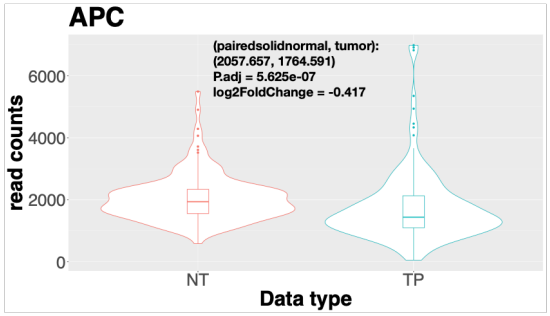

FAT4_Gene total expression
RPKM based (pairedsolidnormal, tumor):
(1.887043, 0.9595611)
P = 1.069189e-13
FPKM
20
15
10
5
0
pairedsolidnormal
tumor
sample

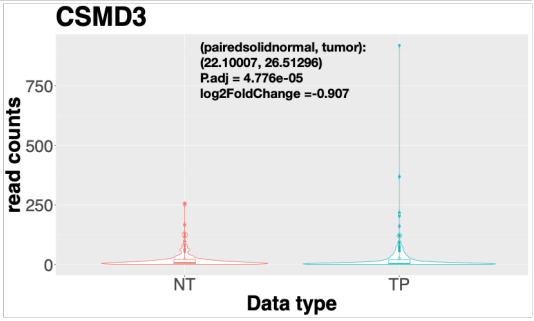

CNTNAP2_Gene total expression
RPKM based (pairedsolidnormal, tumor):
(2.18337, 2.148362)
P = 0.9176621
FPKM
60
40
20
0
pairedsolidnormal
tumor
sample

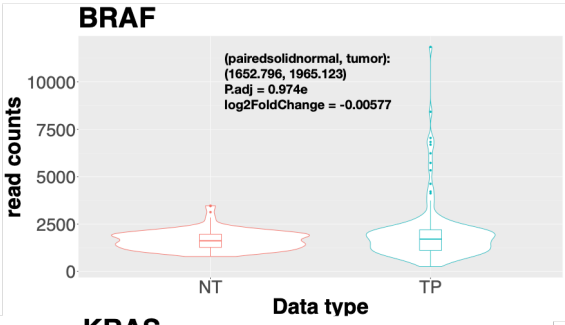

MUC16
(pairedsolidnormal, tumor):
(272.5498, 1408.7165)
P.adj = 1.0493e-05
log2FoldChange = 0.997
read counts
20000
10000
0
NT
TP
Data type

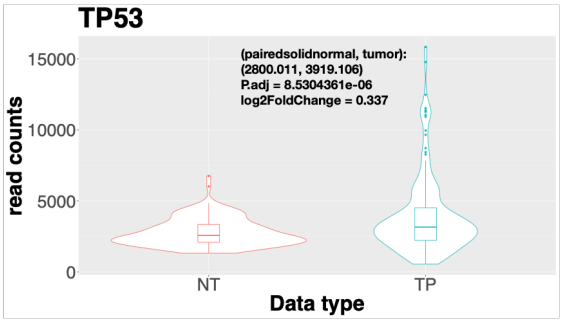

NBEA
(pairedsolidnormal, tumor):
(1411.5829, 867.6333)
P.adj = 5.210e-17
log2FoldChange = -1.16
read counts
6000
4000
2000
0
NT
TP
Data type

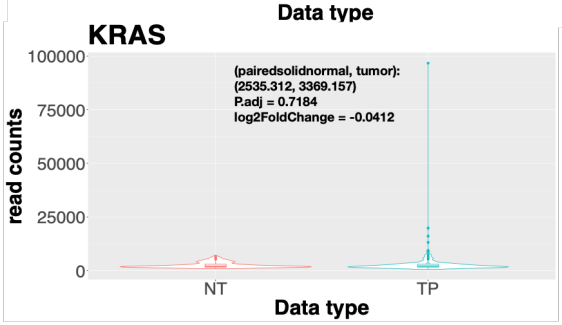

LRP1B
(pairedsolidnormal, tumor):
(341.61479, 70.13568)
P.adj = 1.952e-20
log2FoldChange = -1.89
read counts
4000
3000
2000
1000
0
NT
TP
Data type

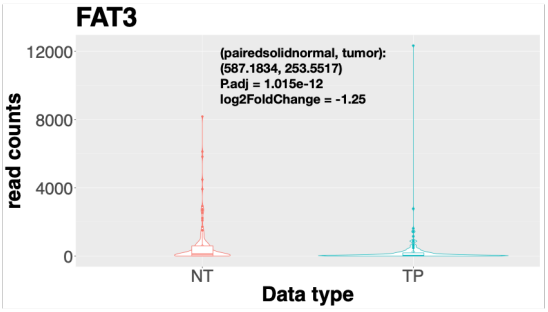

APC
(pairedsolidnormal, tumor):
(2057.657, 1764.591)
P.adj = 5.625e-07
log2FoldChange = -0.417
read counts
6000
4000
2000
0
NT
TP
Data type

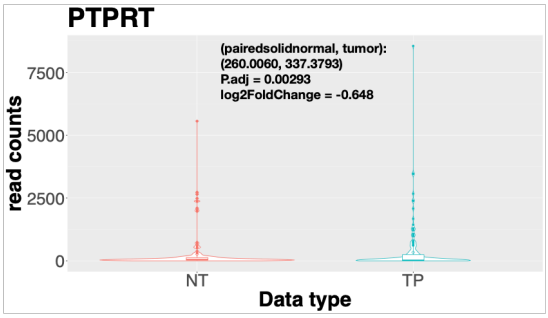

CSMD3
(pairedsolidnormal, tumor):
(22.10007, 26.51296)
P.adj = 4.776e-05
log2FoldChange =-0.907
read counts
750
500
250
0
NT
TP
Data type

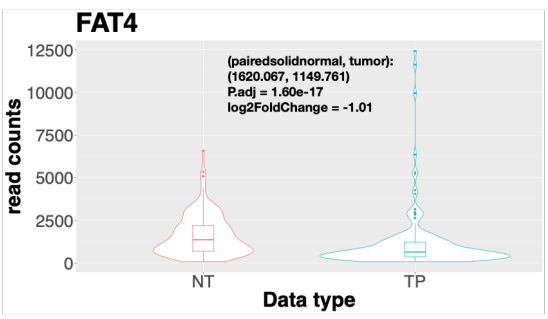

BRAF
(pairedsolidnormal, tumor):
(1652.796, 1965.123)
P.adj = 0.974e
log2FoldChange = -0.00577
read counts
10000
7500
5000
2500
0
NT
TP
Data type

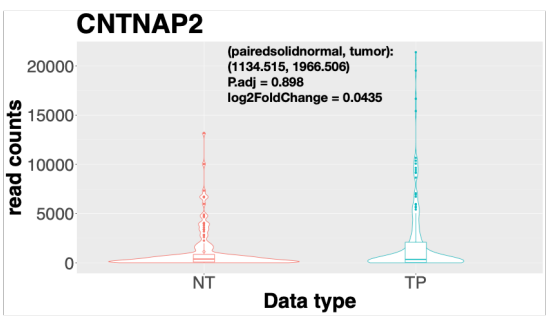

TP53
(pairedsolidnormal, tumor):
(2800.011, 3919.106)
P.adj = 8.5304361e-06
log2FoldChange = 0.337
read counts
15000
10000
5000
0
NT
TP
Data type


KRAS
(pairedsolidnormal, tumor):
(2535.312, 3369.157)
P.adj = 0.7184
log2FoldChange = -0.0412
read counts
100000
75000
50000
25000
0
NT
TP
Data type

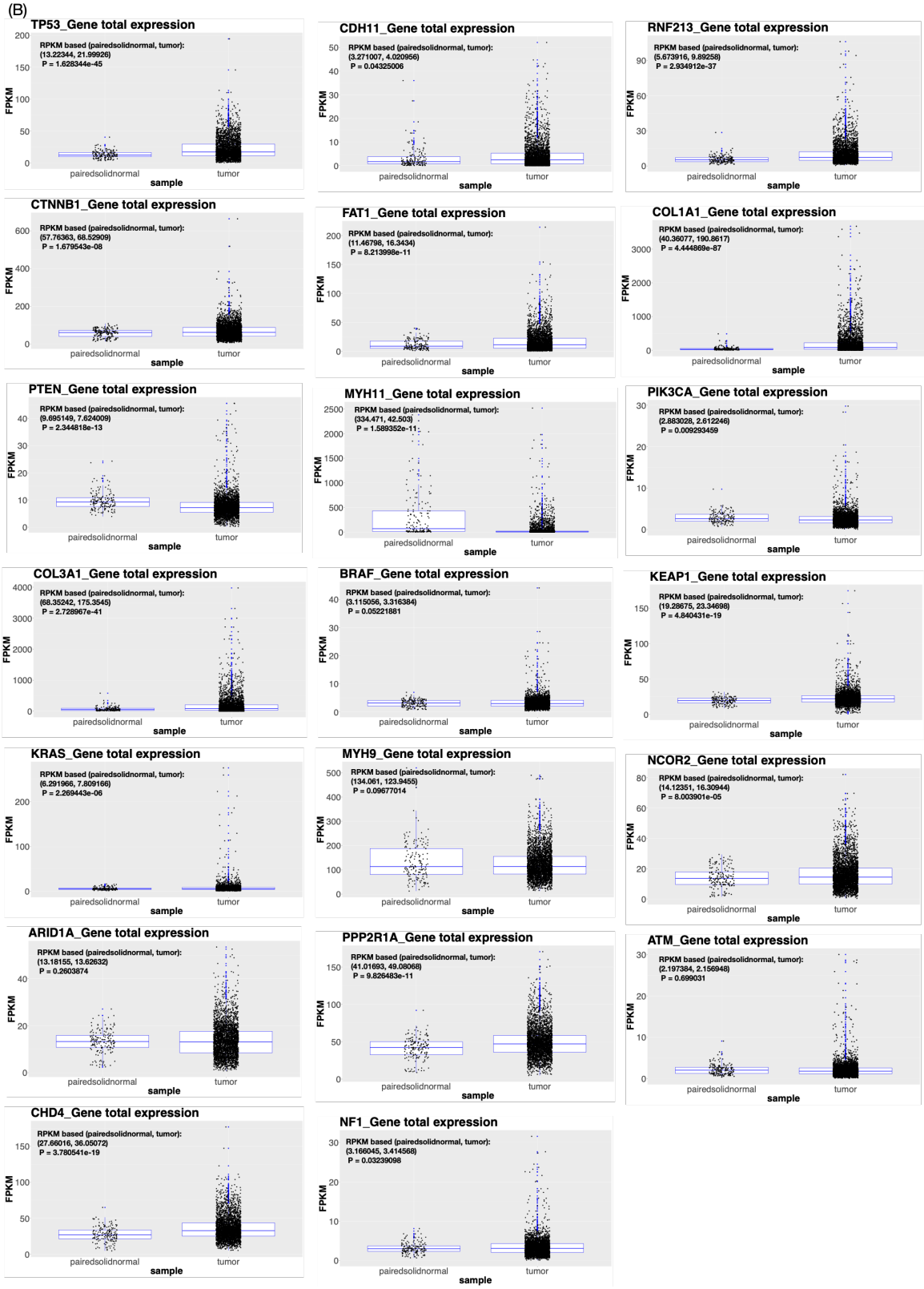

FAT3
(pairedsolidnormal, tumor):
(587.1834, 253.5517)
P.adj = 1.015e-12
log2FoldChange = -1.25
read counts
12000
8000
4000
0
NT
TP
Data type


PTPRT
(pairedsolidnormal, tumor):
(260.0060, 337.3793)
P.adj = 0.00293
log2FoldChange = -0.648
read counts
7500
5000
2500
0
NT
TP
Data type


FAT4
(pairedsolidnormal, tumor):
(1620.067, 1149.761)
P.adj = 1.60e-17
log2FoldChange = -1.01
read counts
12500
10000
7500
5000
2500
0
NT
TP
Data type


CNTNAP2
(pairedsolidnormal, tumor):
(1134.515, 1966.506)
P.adj = 0.898
log2FoldChange = 0.0435
read counts
20000
15000
10000
5000
0
NT
TP
Data type

(B)

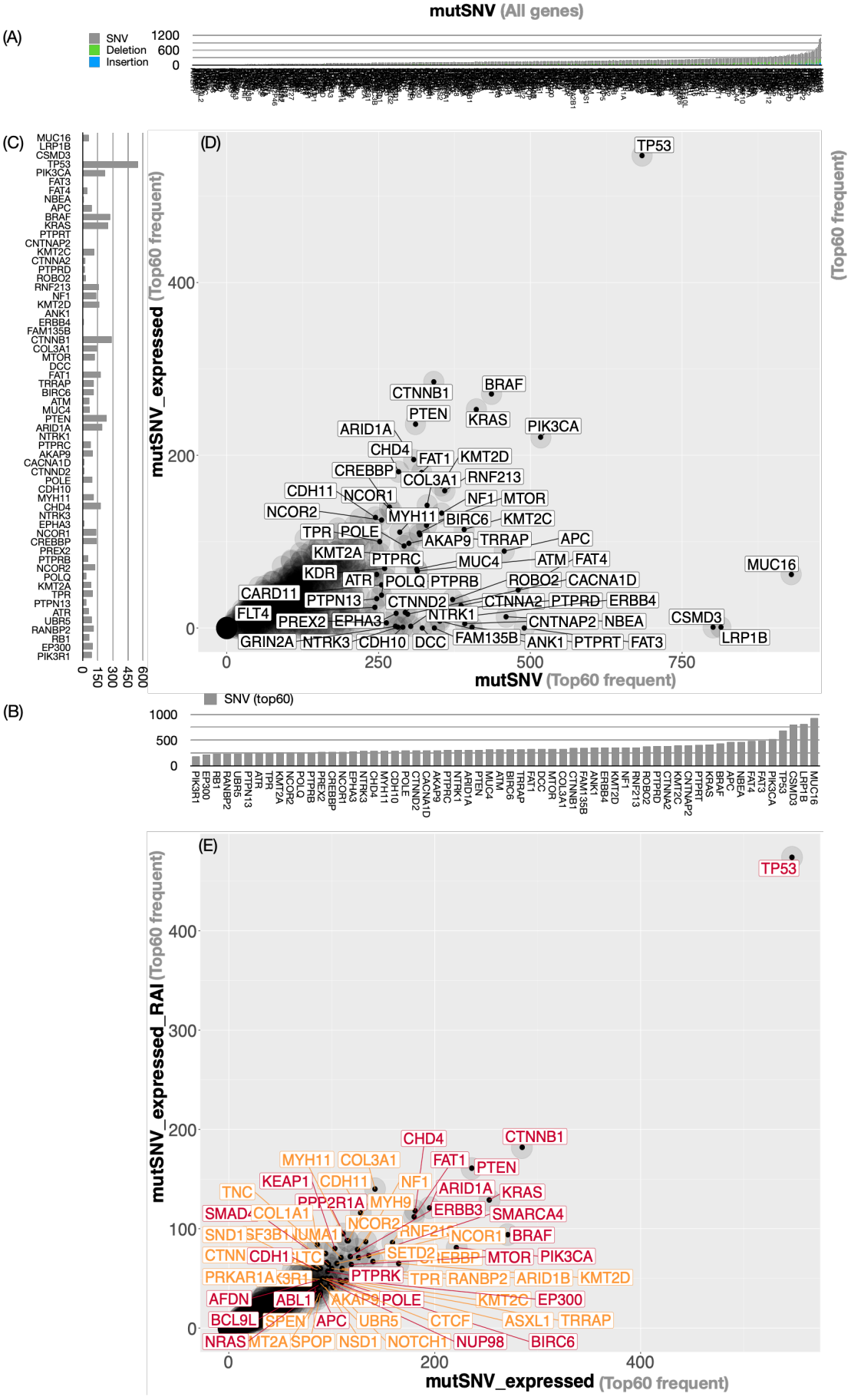

TP53_Gene total expression
RPKM based (pairedsolidnormal, tumor):
(13.22344, 21.99926)
P = 1.628344e-45
CDH11_Gene total expression
RPKM based (pairedsolidnormal, tumor):
(3.271007, 4.020956)
P = 0.04325006
RNF213_Gene total expression
RPKM based (pairedsolidnormal, tumor):
(5.673916, 9.89258)
P = 2.934912e-37
CTNNB1_Gene total expression
RPKM based (pairedsolidnormal, tumor):
(57.76363, 68.52909)
P = 1.679543e-08
FAT1_Gene total expression
RPKM based (pairedsolidnormal, tumor):
(11.46798, 16.3434)
P = 8.213998e-11
COL1A1_Gene total expression
RPKM based (pairedsolidnormal, tumor):
(40.36077, 190.8617)
P = 4.444869e-87
PTEN_Gene total expression
RPKM based (pairedsolidnormal, tumor):
(9.695149, 7.624009)
P = 2.344818e-13
MYH11_Gene total expression
RPKM based (pairedsolidnormal, tumor):
(334.471, 42.503)
P = 1.589352e-11
PIK3CA_Gene total expression
RPKM based (pairedsolidnormal, tumor):
(2.883028, 2.612246)
P = 0.009293459
COL3A1_Gene total expression
RPKM based (pairedsolidnormal, tumor):
(66.35242, 175.3545)
P = 2.728967e-41
BRAF_Gene total expression
RPKM based (pairedsolidnormal, tumor):
(3.115056, 3.316384)
P = 0.05221881
KEAP1_Gene total expression
RPKM based (pairedsolidnormal, tumor):
(19.28675, 23.34698)
P = 4.840431e-19
KRAS_Gene total expression
RPKM based (pairedsolidnormal, tumor):
(6.291966, 7.809166)
P = 2.269443e-06
MYH9_Gene total expression
RPKM based (pairedsolidnormal, tumor):
(134.061, 123.9455)
P = 0.09677014
NCOR2_Gene total expression
RPKM based (pairedsolidnormal, tumor):
(14.12351, 16.30944)
P = 8.003901e-05
ARID1A_Gene total expression
RPKM based (pairedsolidnormal, tumor):
(13.18155, 13.62632)
P = 0.2603874
PPP2R1A_Gene total expression
RPKM based (pairedsolidnormal, tumor):
(41.01693, 49.08068)
P = 9.826483e-11
ATM_Gene total expression
RPKM based (pairedsolidnormal, tumor):
(2.197384, 2.156948)
P = 0.699031
CHD4_Gene total expression
RPKM based (pairedsolidnormal, tumor):
(27.66016, 36.05072)
P = 3.780541e-19
NF1_Gene total expression
RPKM based (pairedsolidnormal, tumor):
(3.166045, 3.414568)
P = 0.03239098
FPKM
pairedsolidnormal
tumor
sample

Supplement Figure 3. Differential expression analysis (DEA) for those gene with top frequent somatic mutations among selected Pan-adenocarcinoma cases (n=3256).

The expression level is based on different normalization methods by following the TCGA standard protocol. (Left) part is by FPKM, "pairedsolidnormal" stands for the paired-normal-like tissue group, and "tumor" stands for the tumor tissue group. And (right) part is by DESeq2 R package, "NT" stands for the paired-normal-like tissue group, and "TP" stands for the tumor tissue group.

(A) All the genes were ranked and shown for those top ones. Especially for those genes, such as *MUC16*, *LRP1B*, *CSMD3* and others, which show high somatic mutation frequency but with low expression level (or no expression) are shown.

(B) Only those expressed genes were ranked and shown here.



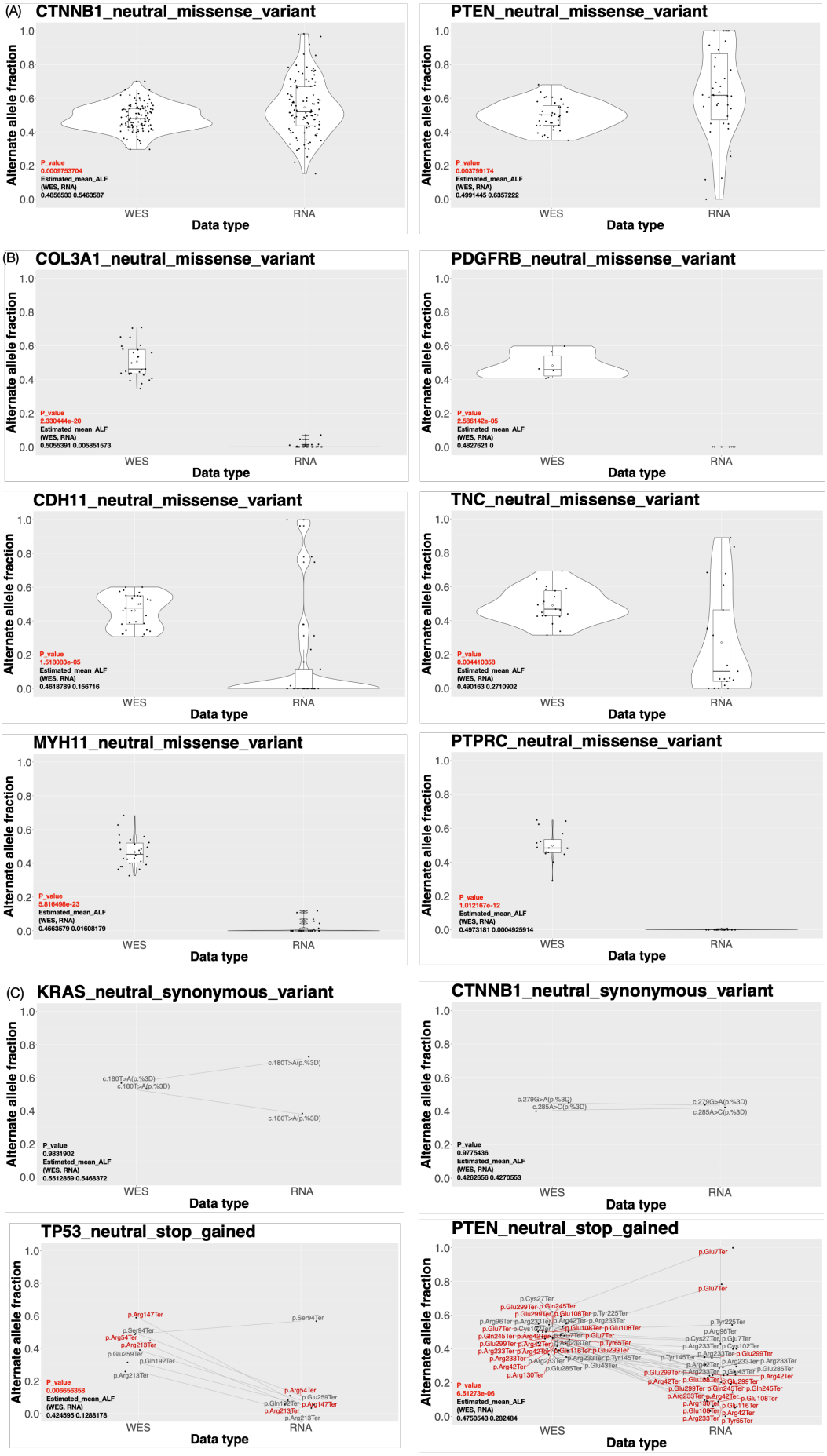
Supplement Figure 4. Frequency distribution of genes with expressed/unexpressed somatic mutations among selected Pan-adenocarcinoma cases (n=3256).

Here it shows the landscape of the somatic mutation frequencies among the selected Pan-adenocarcinoma cases. **(A)** For each gene, counts of cases carrying somatic mutation were showed in bar-plots, which were classified into different mutation types with different colors in each bar. Each color represents corresponding mutation types: Insertions were blue, Deletions were green, and SNVs were grey. Those cancer associating genes reported by COSMIC database were chosen for the further investigation in the sub-table (B), and it was only for the top-60 frequent ones.

(B&C) Among them, those top-60 frequent somatic mutations carrying genes with SNV mutation types were magnified in these two sub-tables. The bottom sub-table (B) showed the counts of mutation carrying cases for each gene. The left sub-table (C) showed the counts of the mutation carrying cases with expression for each gene. Those cases with expressed somatic mutations were the main candidates for the RAI investigation and then for the AITE check.

(D) The dispersion plot for the count of cases with somatic mutations (X-axis) and the count of cases with expressed somatic mutations (Y-axis) among the selected Pan-Adenocarcinoma cases for each COSMIC reported gene was shown for those top-60 ones according to event frequency. The somatic mutations carrying genes with expressions are those ones presenting along the diagonal lines, such as *TP53*, *BRAF*, *KRAS*, *CTNNB1*, *PTEN*, and so on. And the somatic mutations carrying genes without expressions are those ones presenting in the right bottom corners, such as *MUC16*, *LRP1B*, *CSMD3*, and so on.

(E) Here it shows the landscape of the RAI occurring frequency in those somatic SNV mutations carrying genes among the selected Pan-adenocarcinoma cases. The dispersion plot for the count of cases with expressed somatic mutations (X-axis) and the count of cases with expressed somatic mutations harboring RNA Allelic Imbalance (Y-axis) among the selected Pan-Adenocarcinoma cases for each COSMIC reported gene was shown for those top-60 ones according to event frequency. The red marks were for those RAI genes showing mutant allele bias (alternate allele fraction is greater than 0.5), and the orange marks were for those ones showing normal allele bias (alternate allele fraction is smaller than 0.5).



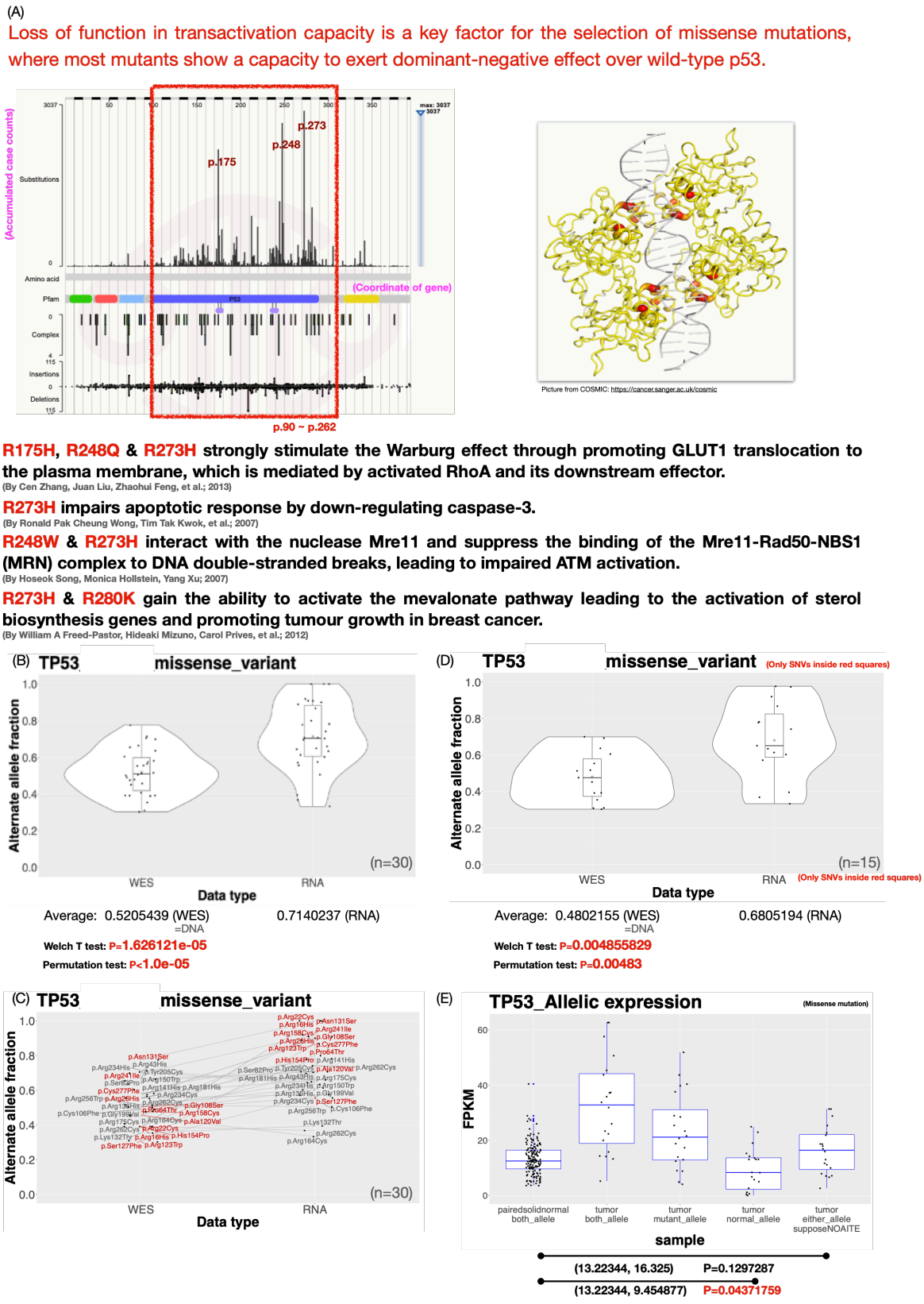
Supplement Figure 5. Shifts in alternate allele fractions between DNA and RNA for the heterozygous genes carrying different somatic mutation type.

Evidence of Allelic Imbalance in Transcriptional Efficiency (AITE) for the genes with expressed heterozygous somatic mutations was shown here. Among those RAI cases, after excluding those cases showing mutation sites being overlapped with SCNA (including LOH) region, the remained cases were supposed to exist AITE and being visualized here. Each point stands for each sample case. For those supposed AITE cases, the average values of alternate (mutant) allele fractions were counted for WESeq and RNA-Seq respectively, and they were further compared by Welch t test and permutation test.

(A) Other examples for typical cancer driver genes carrying missense mutations, *CTNNB1* and *PTEN*.

(B) Examples for typical passenger or non-cancer driver genes carrying missense mutations, *COL3A1*, *MYH11*, *PDGFRB*, *TNC*, *PTPRC*, and *CDH11*.

(C) Examples for typical cancer driver genes carrying non-missense mutations, *TP53*, *KRAS*, *PTEN* and *CTNNB1*.



Supplement Figure 6. The reported functional mutations in *TP53*.

(A) *TP53* is a typical tumor suppressor gene, and its cancer driver mutations mostly act as loss of function mutations which concentrate in a small region of the coding sequences. The left small figure shows the distribution of accumulated somatic mutation counts among cancer patients along the *TP53* coding region, and several peaks can be observed in a specific protein domain (p.90~p.262) of p53 protein, which is visualized and marked in red in the right small figure. Among those peaks, R175H, R248Q, R273H, R248W, R280K are those

reported mutations for their roles in driving cancer progressions. (This subfigure is modified screenshot from website https://cancer.sanger.ac.uk/cosmic)

(B) Comparison of the alternate allele fractions between WESeq and RNA-Seq for all non-synonymous SNV mutations inside *TP53*.

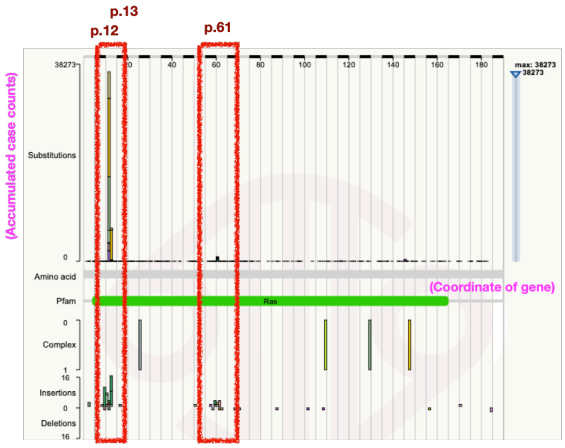
(C) Same as (B) but with detailed annotation for each case.

(D) Same as (B) but only for those ones inside the positive functional selection domain of *TP53*.

(E) Allele-specific expression level which is based on multiplying allelic ratio with total gene expression.

(A)

Gain of function mutations in codon 12, 13, or 61 among ras genes (H-ras, K-ras & N-ras), convert these genes into active oncogenes, which is a key factor for the selection of missense mutations.



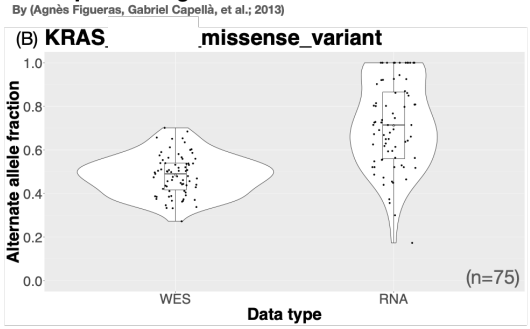
G12D, G12V, G12C, G12A, G12S, G12R, G13D, G13C & Q61H belong to highly recurrent ones.
(By S. A. Forbes, D. Beare, P. J. Campbell, et al.; 2016)

G12, G13 & Q61 cause Ras to accumulate in the active GTP bound state by impairing intrinsic GTPase activity and conferring resistance to GTPase activating proteins.
(By J. L. Bos; 1989)

G12D are resistant to the proapoptotic actions of TRAIL.
(By Ravi P. Sahu, Sanjay K. Srivastava, et al.; 2011)

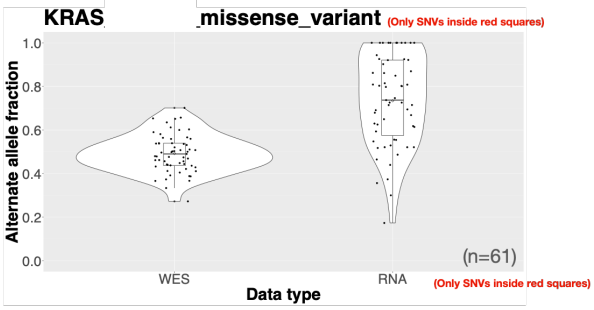
G12C present a highly glycolytic phenotype.
(By Pedro Vizan, Marta Cascante, et al.; 2005)

G13D present higher levels of VEGF-A.
By (Agnès Figueras, Gabriel Capellà, et al.; 2013)



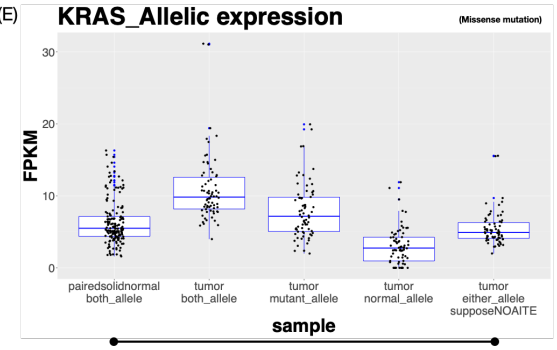
Average: 0.4851127 (WES) 0.7118938 (RNA)
=DNA

Welch T test: P=1.356997e-14
Permutation test: P<1.0e-05



Average: 0.492092 (WES) 0.733901 (RNA)
=DNA

Welch T test: P=3.199496e-12
Permutation test: P<1.0e-05



(6.291966, 5.461749) P=0.0260243
(6.291966, 7.906094) P=0.00238465

Supplement Figure 7. The reported functional mutations in *KRAS*.

(A) *KRAS* is a typical oncogene, and its cancer driver mutations mostly act as gain of function mutations which concentrate in few positions of the coding region. The figure shows the distribution of accumulated somatic mutation counts among cancer patients along the *KRAS* coding region, and few peaks (p.12, p.13, p.61) can be observed. Among those peaks, G12D, G12V, G12C, G12A, G12S, G12R, G13D, G13C and Q61H are those

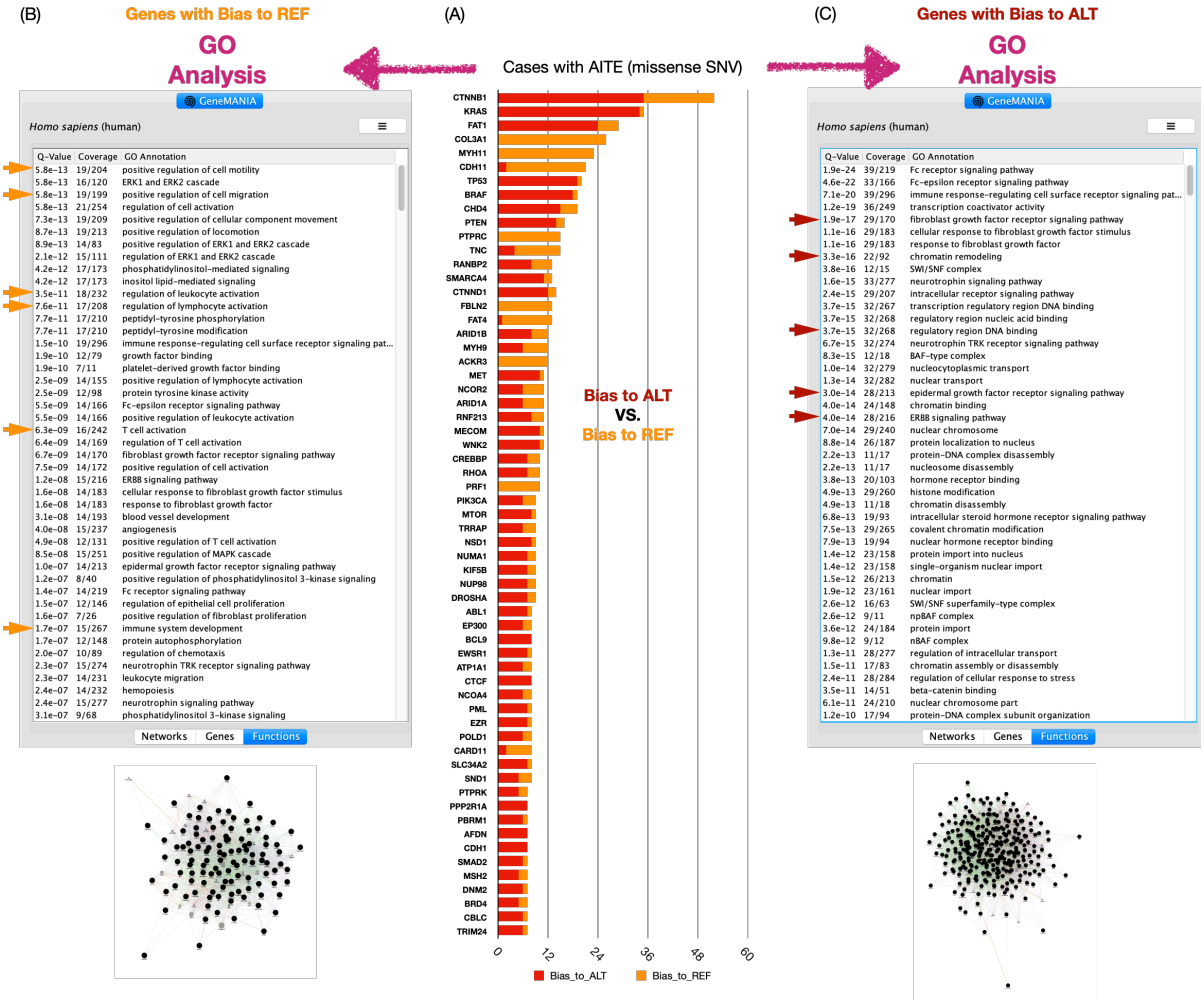
reported mutations for their roles in driving cancer progressions. (This subfigure is modified screenshot from website https://cancer.sanger.ac.uk/cosmic)

(B) Comparison of the alternate allele fractions between WESeq and RNA-Seq for all non-synonymous SNV mutations inside *KRAS*.

(C) Same as (B) but with detailed annotation for each case.

(D) Same as (B) but only for those ones inside the positive functional selection domain of *KRAS*.

(E) Allele-specific expression level which is based on multiplying allelic ratio with total gene expression.



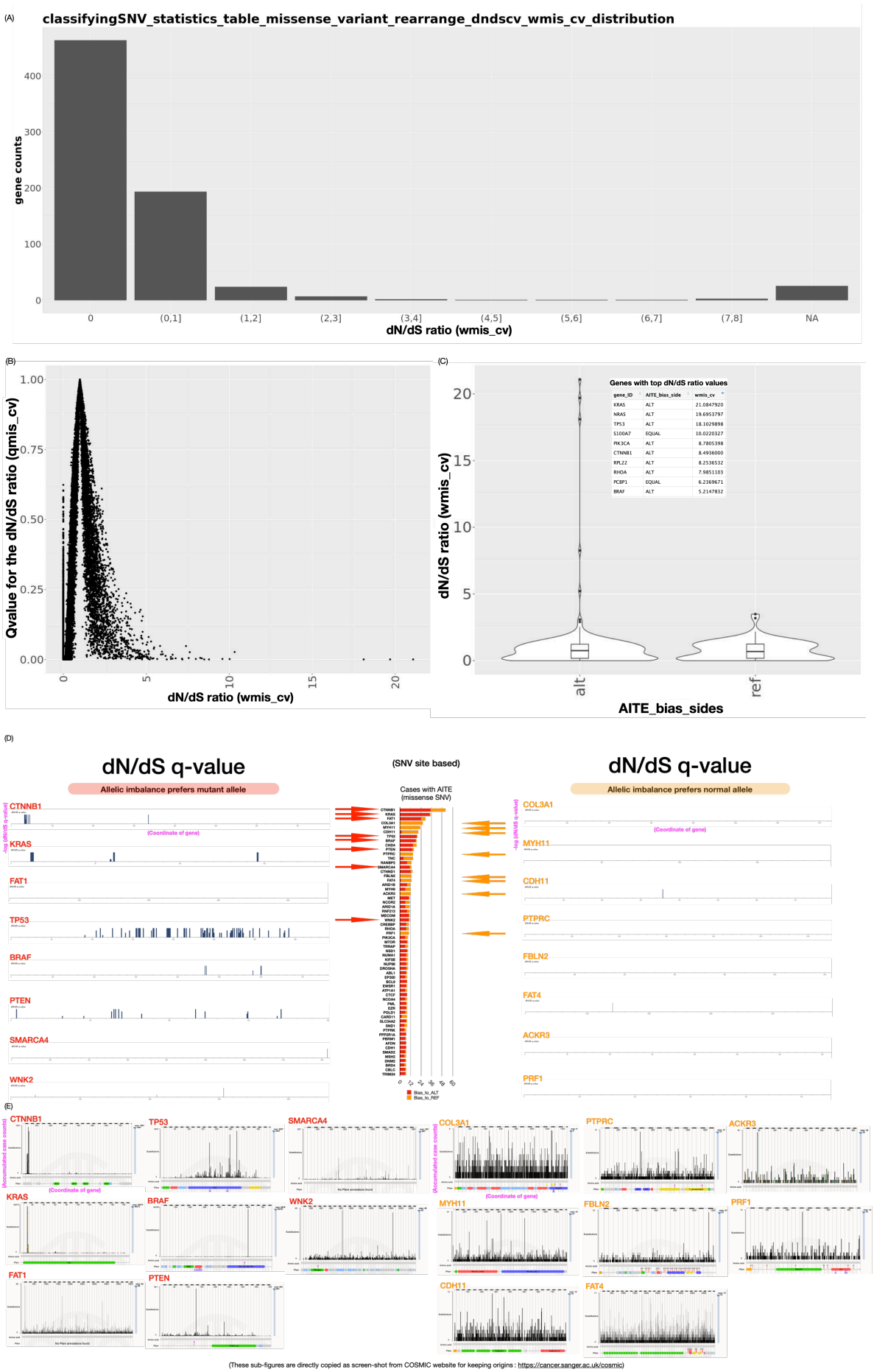
Supplement Figure 8. GO analysis for those AITE genes with the same allele bias side (ref. or alt.).

Here shows the genes set enrichment analysis (GSEA) for those AITE genes carrying missense somatic mutation with consistent bias sides.

(A) AITE genes carrying missense somatic mutations were grouped into two sets by their bias sides for the genes set enrichment analysis. Here only shows the top 60 frequent ones, but a total of around 700 CGC genes were input into GO analysis.

(B) AITE genes carrying missense somatic with bias side towards reference allele enriched in some specific biological functions.

(C) AITE genes carrying missense somatic with bias side towards alternate allele enriched in some specific biological functions.



Supplement Figure 9. Investigation of the positive selection signals through dN/dS methods for those somatic SNV mutations (all genes included) inside Pan-adenocarcinoma (n=3256).

Sub-figures (A), (B), (C) were calculated based on each gene. And sub-figures (D) and (E) were calculated based on each SNV site.

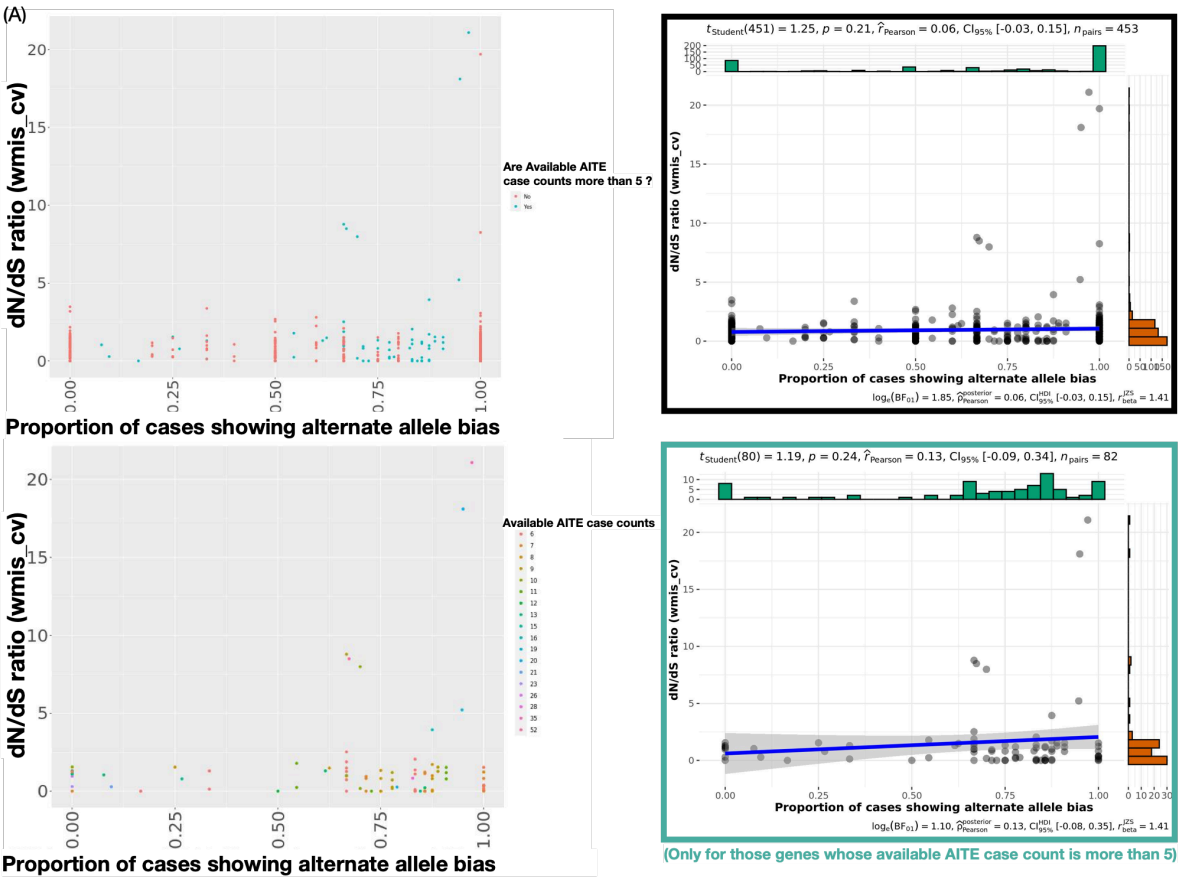
(A) Frequency distribution of the dN/dS ratio among all the genes.

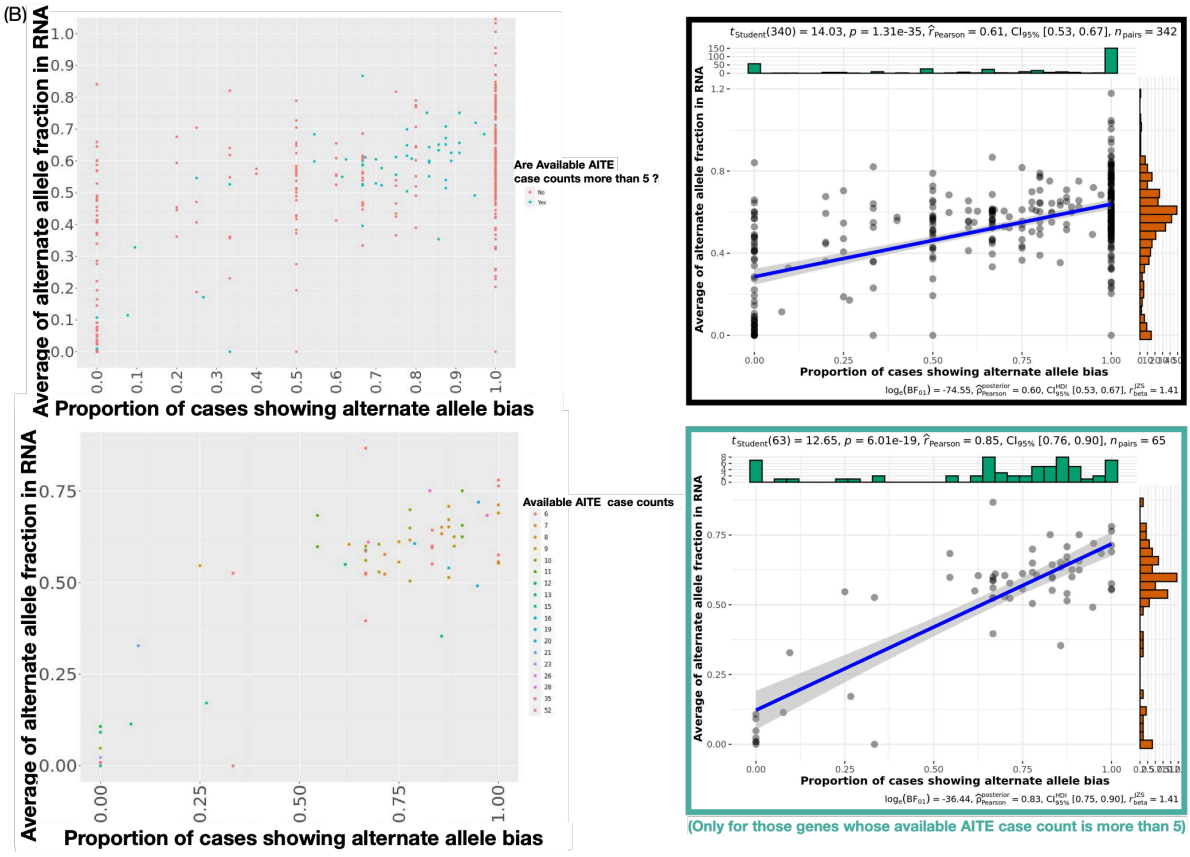
(B) Relationships between the q values for positive selection signals and the corresponding dN/dS ratio among all the genes.

(C) Distribution of the dN/dS ratio between the two different bias sides of AITE genes (only COSMIC genes were remained for visualization).

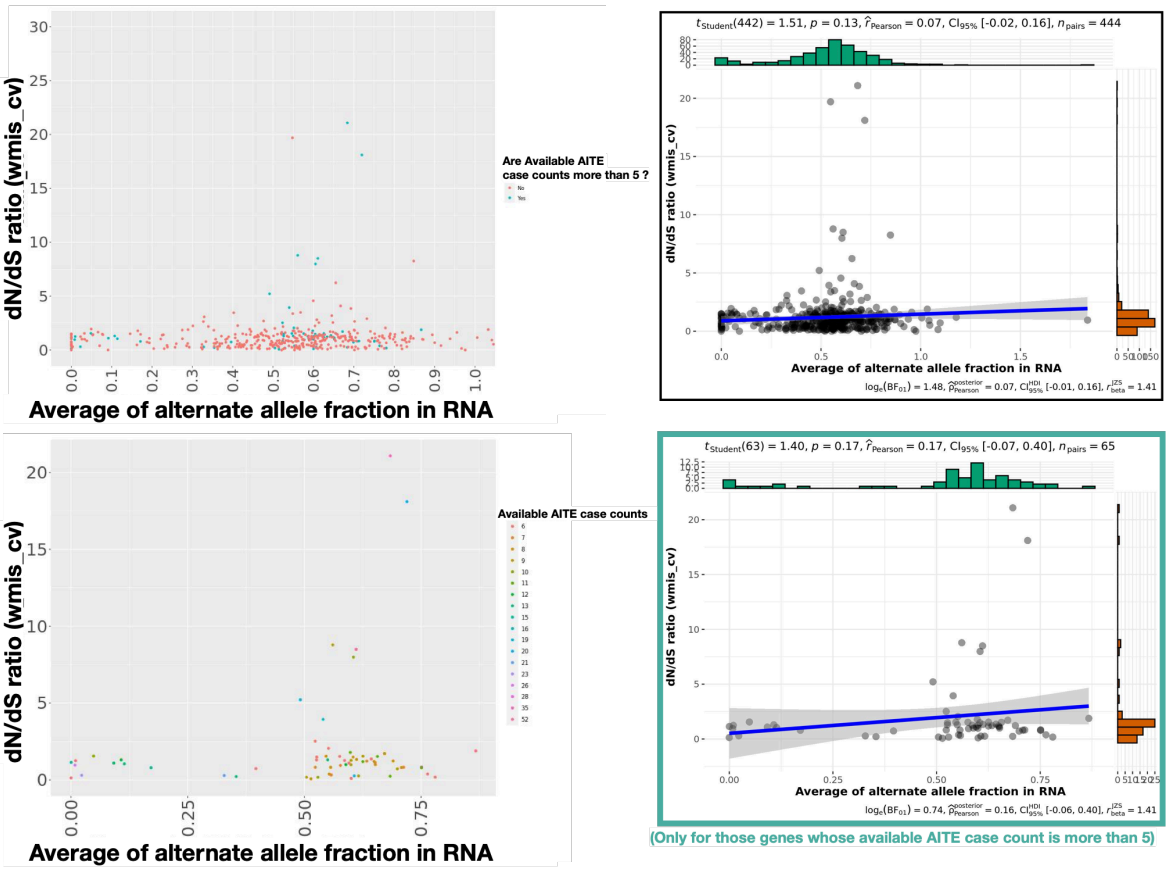
(D) Investigation of the positive selection signals through dN/dS at each somatic SNV mutation site was shown, which was based on (E). Each subfigure stands for each gene, and the X-axis shows the coordinate of the gene, and the Y-axis shows the calculated q-value (multiple test corrected p-value) of each SNV site inside each specific gene. For each SNV site, the higher dN/dS ratio will present a higher dN/dS q-value, and suggesting stronger positive selection signals. These sub-figures are directly copied as screen-shot from COSMIC website for keeping origins : https://cancer.sanger.ac.uk/cosmic.

(E) The counts of somatic mutation carriers at each somatic heterozygous SNV mutation site were shown in each subfigure for each gene. The X-axis shows the coordinate of the gene, and the Y-axis shows the accumulated case numbers. These sub-figures are directly copied as screen-shot from COSMIC website for keeping origins : https://cancer.sanger.ac.uk/cosmic.


(A)
dN/dS ratio (wmis_cv)
Proportion of cases showing alternate allele bias
Are Available AITE case counts more than 5 ?
No
Yes
Available AITE case counts
$t_{Student}(451) = 1.25$, $p = 0.21$, $\hat{r}_{Pearson} = 0.06$, $CI_{95\%}$ [-0.03, 0.15], $n_{pairs} = 453$
$t_{Student}(80) = 1.19$, $p = 0.24$, $\hat{r}_{Pearson} = 0.13$, $CI_{95\%}$ [-0.09, 0.34], $n_{pairs} = 82$
(Only for those genes whose available AITE case count is more than 5)


(B)
Average of alternate allele fraction in RNA
Proportion of cases showing alternate allele bias
Are Available AITE case counts more than 5 ?
No
Yes
Available AITE case counts
$t_{Student}(340) = 14.03$, $p = 1.31e-35$, $\hat{r}_{Pearson} = 0.61$, $CI_{95\%}$ [0.53, 0.67], $n_{pairs} = 342$
$t_{Student}(63) = 12.65$, $p = 6.01e-19$, $\hat{r}_{Pearson} = 0.85$, $CI_{95\%}$ [0.76, 0.90], $n_{pairs} = 65$
(Only for those genes whose available AITE case count is more than 5)

(C)

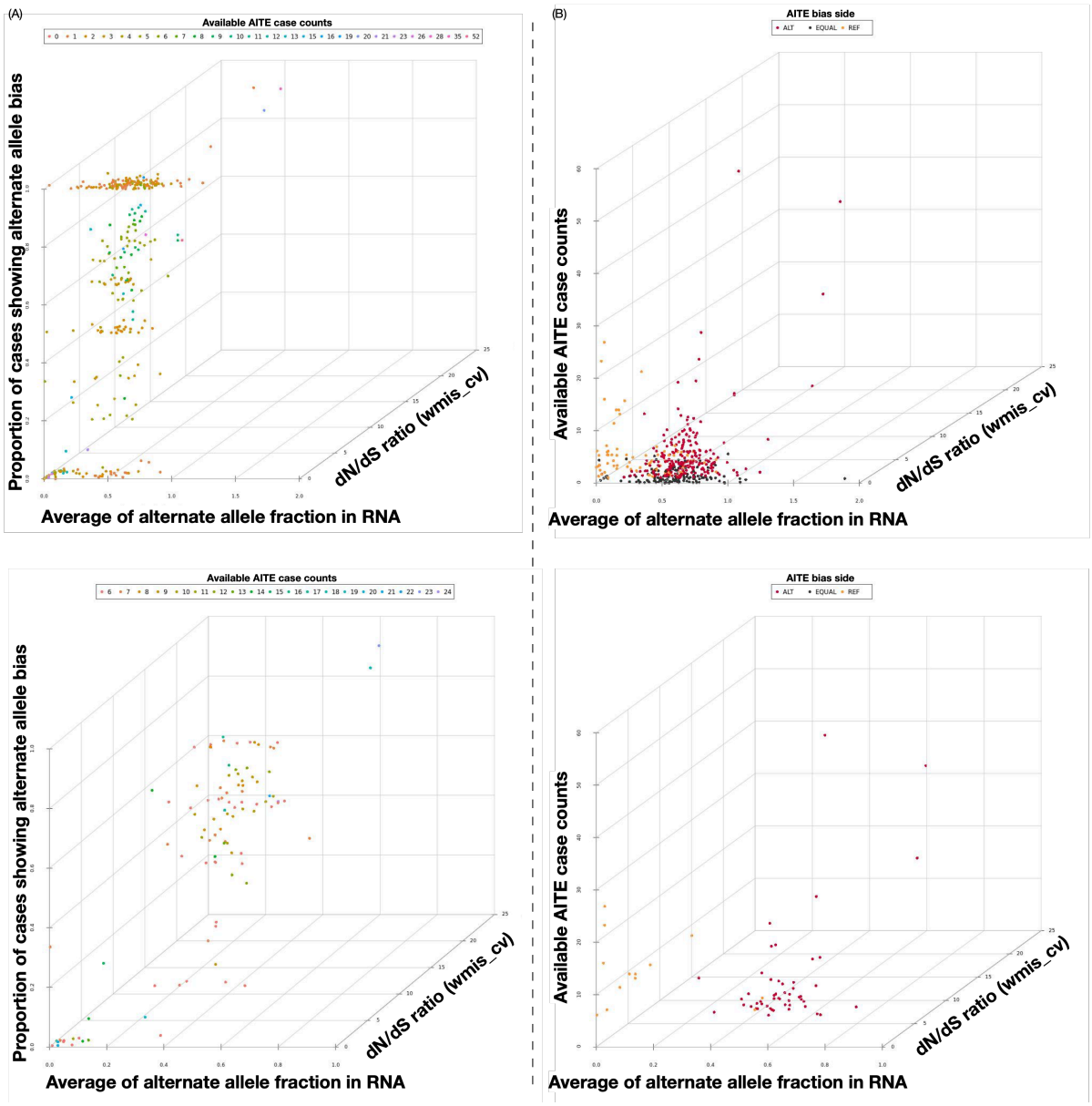


Supplement Figure 10. How the available case counts can impact the correlations between the allelic imbalance bias extent and positive selection signals (dN/dS) for those COSMIC genes with AITE effects.

(A) "Proportion of cases showing alternate allele bias" versus "dN/dS ratio".

(B) "Proportion of cases showing alternate allele bias" versus "Average of alternate allele fraction in RNA".

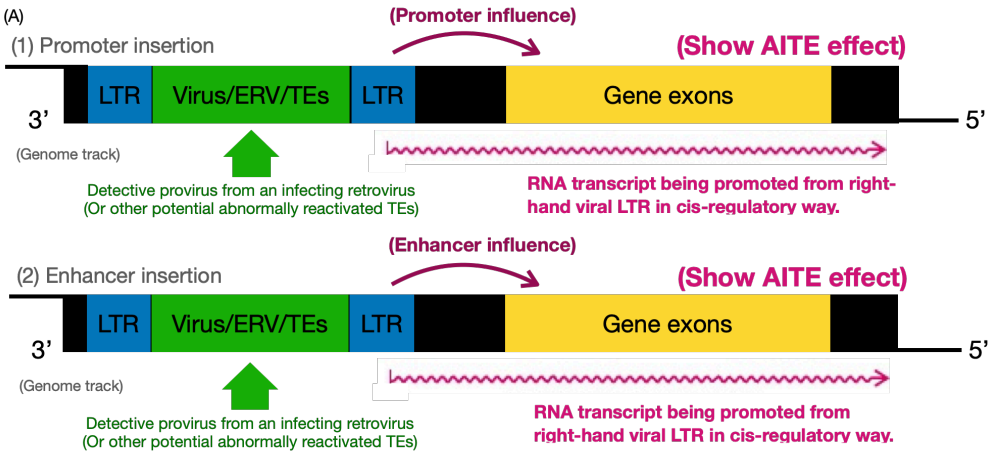
(C) "Average of alternate allele fraction in RNA" versus "dN/dS ratio".



Supplement Figure 11. How the available case counts can impact the correlations between the allelic imbalance bias extent and positive selection signals (dN/dS) for those COSMIC genes with AITE effects in 3D.

(A) "Average of alternate allele fraction in RNA" versus "dN/dS ratio" versus "Proportion of cases showing alternate allele bias".

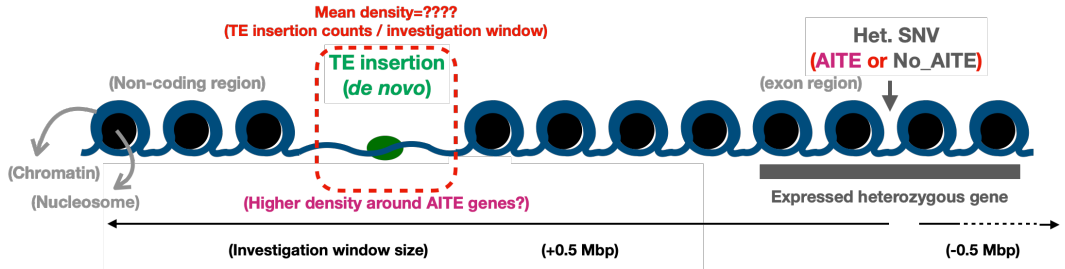
(B) "Average of alternate allele fraction in RNA" versus "dN/dS ratio" versus "Available AITE case counts".



"PS. Original figure was cited from Gregory S. Payne, Harold E. Varmus, et al. (1982)" And, the redrawn and modified figures here were added with my new ideas, notations and interpretations."

(B)

Higher *De-novo* TE insertions density can be detected around AITE genes:

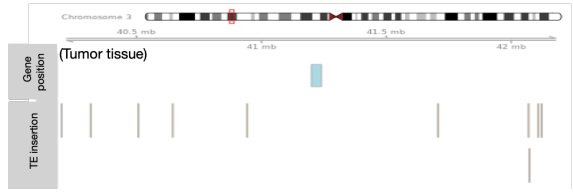


Expressed heterozygous SNV sites (already excluding those allele specific SCNV regions)

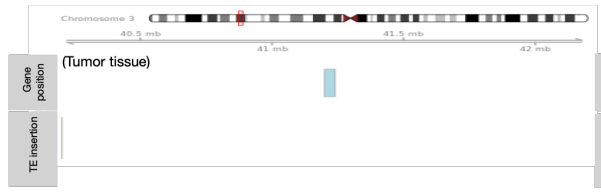
(C)

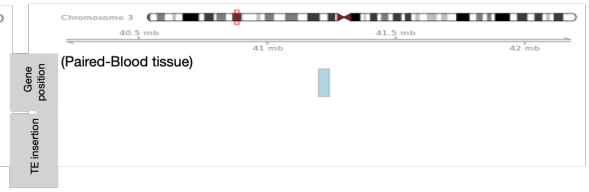
Tumor-specific *de novo* TE insertions occurred around AITE genes:

caseID	gene_ID	contig	position	variant_type	CSQ_HGVSc	CSQ_HGVSp	tumor_purity	WESref	WESref_corrected	WESalt	RNAref	RNAref_corrected	RNAalt	RNA_binomial_P_BH_corrected	Z_Proportion_P	+-1Mb_TE_counts (Tumor/SolidNormal)
TCGA-44-4112	CTNNB1	chr3	41233359	missense	c.1100G>T	p.Gly367Val	0.4	41	17.43	14	744	206.14	511	1.11E-29	3.61E-03	10/0

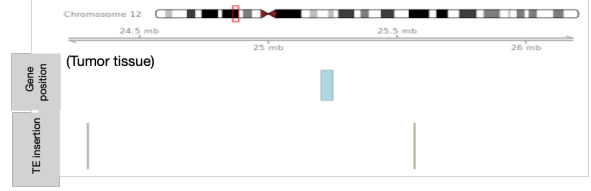


caseID	gene_ID	contig	position	variant_type	CSQ_HGVSc	CSQ_HGVSp	tumor_purity	WESref	WESref_corrected	WESalt	RNAref	RNAref_corrected	RNAalt	RNA_binomial_P_BH_corrected	Z_Proportion_P	+-1Mb_TE_counts (Tumor/Blood)
TCGA-A6-2682	CTNNB1	chr3	41227331	Downstream gene variant	None	None	0.55	56	22.70	18	73	19.45	46	0.0029588	0.01076	1/0





caseID	gene_ID	contig	position	variant_type	CSQ_HGVSc	CSQ_HGVSp	tumor_purity	WESref	WESref_corrected	WESalt	RNAref	RNAref_corrected	RNAalt	RNA_binomial_P_BH_corrected	Z_Proportion_P	+-1Mb_TE_counts (Tumor/Blood)
TCGA-50-5941	KRAS	chr12	25245350	Missense	C.35G>C	p.Gly12Ala	0.22	33	14.51	22	13	-1.46	30	1.31E-08	2.48E-04	2/1



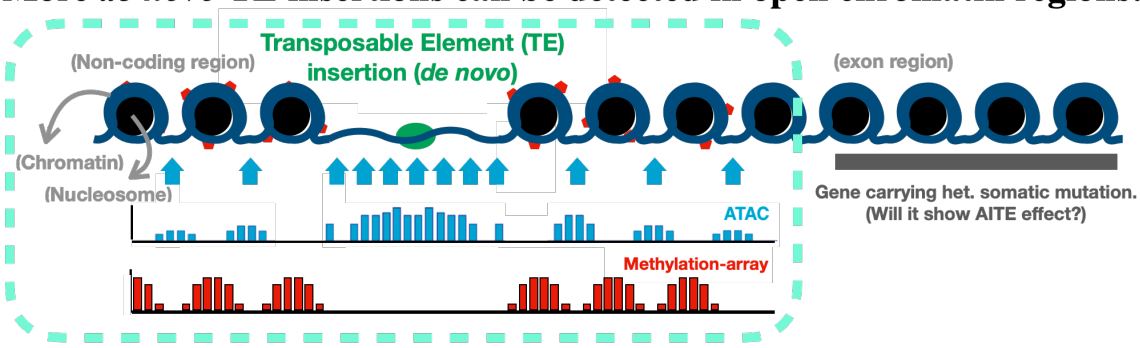
Supplement Figure 12. *De novo* transposable elements (TEs) are supposed to act as *de novo* cis-regulatory elements to be associated with AITE occurrence.

(A) *De novo* TE insertions act as newly derived enhancer/promoters. In non-coding regions around AITE genes, the allele-specific insertion of virus genes, *de novo* ERV or other *de novo* transposable elements (TEs) without LTR are supposed to act as *de novo* cis-regulatory elements, which may bring the *de novo* promoter or enhancers to change the allele-specific expression level of the nearby gene.

(B) Here shows the investigation window for calculating *de novo* TE insertion density. It is illustrated how the *de novo* TE insertion density is calculated for those heterozygous genes without allele-specific SCNA, and the investigated genes can be grouped into AITE group or no_AITE group. With the range in plus 0.5 Mbp and minus 0.5 Mbp of the targeted gene site, the counts of *de novo* TE insertion sites were calculated.

(C) Here shows the detections of tumor-specific *de novo* TE insertions nearby AITE gene. Some examples of well-known cancer driver genes showing AITE at heterozygous somatic mutation sites were listed here. For each event, the above tables show the detailed data for AITE genes, and the below figure shows the position of the target AITE gene and those *de novo* TE insertion sites in tumor and non-tumor tissue, where the tumor-specific *de novo* TE insertion sites can be further observed.

(A) **More *de novo* TE insertions can be detected in open chromatin regions:**

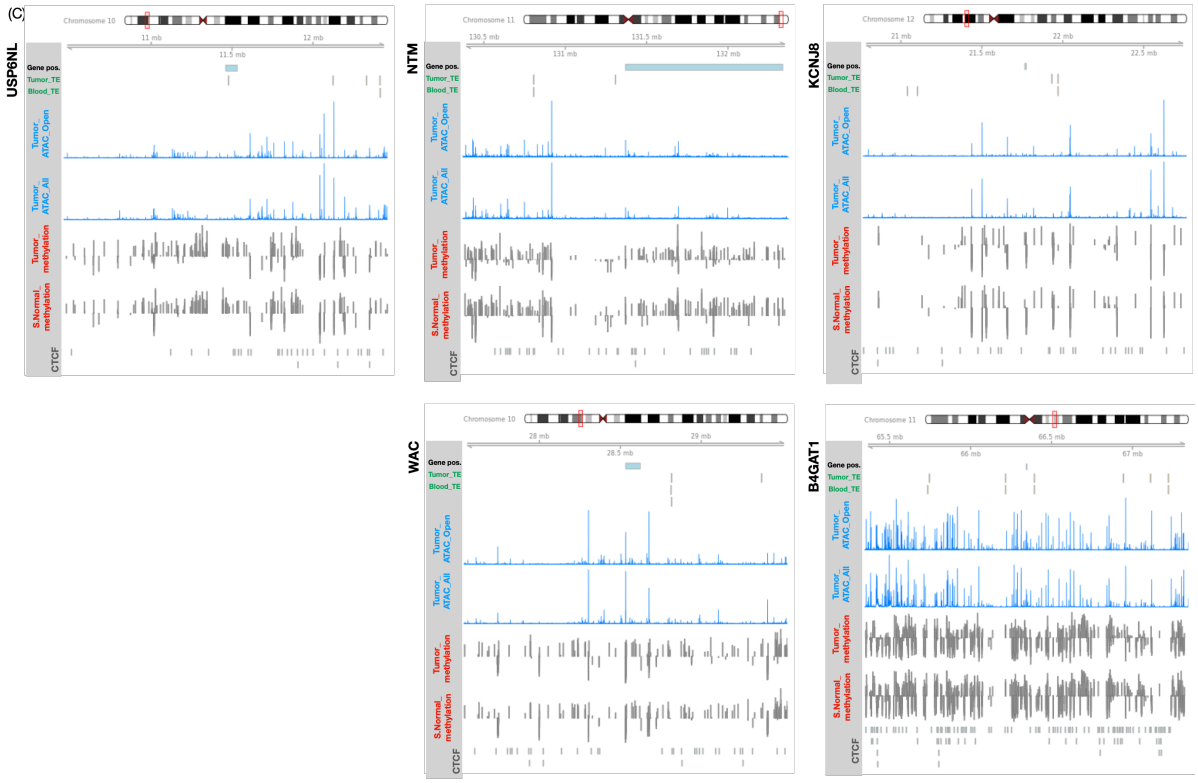


"PS. Original figure was cited from Maria Tsompana & Michael J Buck (2014)" This redrawn and modified figures were added with my new ideas and interpretations."

(B) **AITE genes _ carrying expressed het. somatic mutation:**

	seqnames	ranges	strand	symbol	dominant	alt_bias_snp	ref_bias_snp	unidentified_bias_snp	P_BH_corrected	snp_position	allelic_fraction_dR_dA_rR_rA	purity_considered_percent_dA_rA	P_uncorrected
	<Rle>	<IRanges>	<Rle>	<character>	<character>	<character>	<character>	<character>	<character>	<character>	<character>	<character>	<character>
[1]	chr10	11460509-11532280	*	USP6NL	alt_bias	1	0	0	:0.000306817618138667	:11462483	:55,40,32,36	:23,40,9,36	:6.5746632458285 7e-05
[2]	chr10	28532587-28623112	*	WAC	ref_bias	0	1	0	:1.10685713552369e-17	:28611860	:87,33,236,31	:46,33,145,31	:7.9061223965978e-19
[3]	chr11	66345371-66347690	*	B4GAT1	alt_bias	1	0	0	:0.0142734012333676	:66347032	:22,10,30,28	:11,10,10,28	:0.00509764329763 13
[4]	chr11	131370475-132336822	*	NTM	ref_bias	0	1	0	:5.340576171874 99e-05	:132146298	:54,20,28,0	:29,20,18,0	:7.6293945312499 9e-06
[5]	chr12	21764954-21774821	*	KCNJ8	ref_bias	0	1	0	:0.0068359375	:21773480	:72,31,15,0	:37,31,10,0	:0.001953125

(C)



Supplement Figure 13. Epigenetic signals and *de novo* TE insertions around each somatic mutation sites with AITE effect (TCGA-44-6146 for Demo).

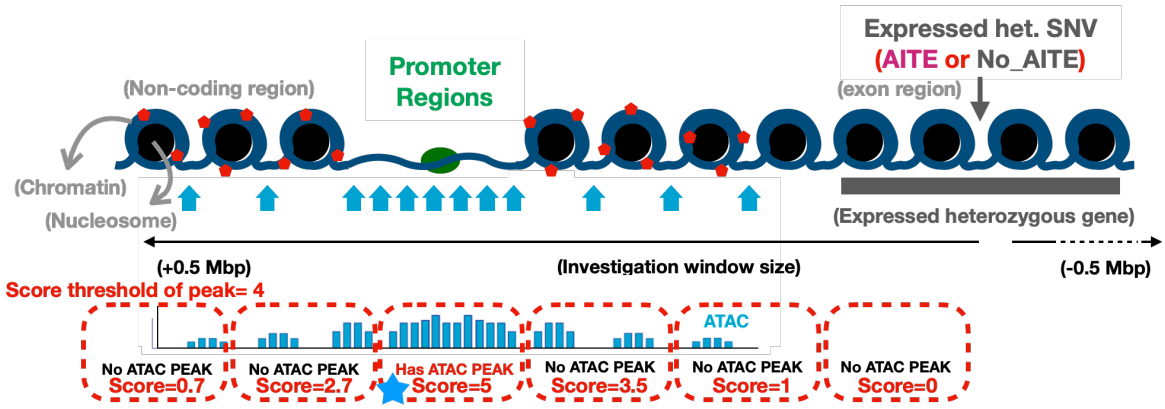
(A) Haplotype specific variants or epigenetic changes may be connected to AITE effect. AITE effect is controlled by the allele-specific regulations in transcription efficiency, which is supposed to be under the effects caused by those cis-regulatory elements. And those cis-regulatory elements usually exist in non-coding regions, which might show some corresponding changes in epigenetic markers. For example, methylation array data can reveal the methylated chromatin regions, and ATAC-Seq data can reveal the open regions of the chromatin. Both of them provide powerful epigenetic signals for detecting the potential locations of those cis-regulatory elements.

(B & C) For a specific individual (TCGA-44-6146 for Demo here), those AITE genes which carried expressed heterozygous somatic mutations were picked up for visualization. The top table shows the detailed data for how AITE was calculated. For the below five figures, those five genes were visualized in each specific figure with the genomic range of 1Mbp wide. Additionally, in each figure there were 8 tracks, the 1st track showed the gene

region (the one where the target variants located); the 2nd track showed the observed structure variants (TE insertion) in tumor; the 3rd track showed the observed structure variants (TE insertion) in paired-solid-normal; the 4th and 5th track showed the ATAC-Seq signal in tumor; the 6th track showed the methylation signals in tumor; the 7th track showed the methylation signals in paired-solid-normal; the 8th track showed the annotations where CTCF binding sites located.

(A)

ATAC-Seq peak signals around AITE genes



1. To Set bin into 500bp wide.
2. Then, to normalize ATACseq read counts (signal) inside each bin into "score".
3. Next, to set a threshold for the "score". If the "score" is above the threshold, then I claim that bin exist ATACseq peak.
4. Finally, to count how many ATACseq peaks can be detected (score > threshold) inside the investigation window.

(1 Peaks inside the investigation window for that specific SNV)

(B)

Odds ratio suggest more ATAC-seq peak signals around AITE SNV sites:

Pearson's Chi-squared Test for Count Data - test of independence:

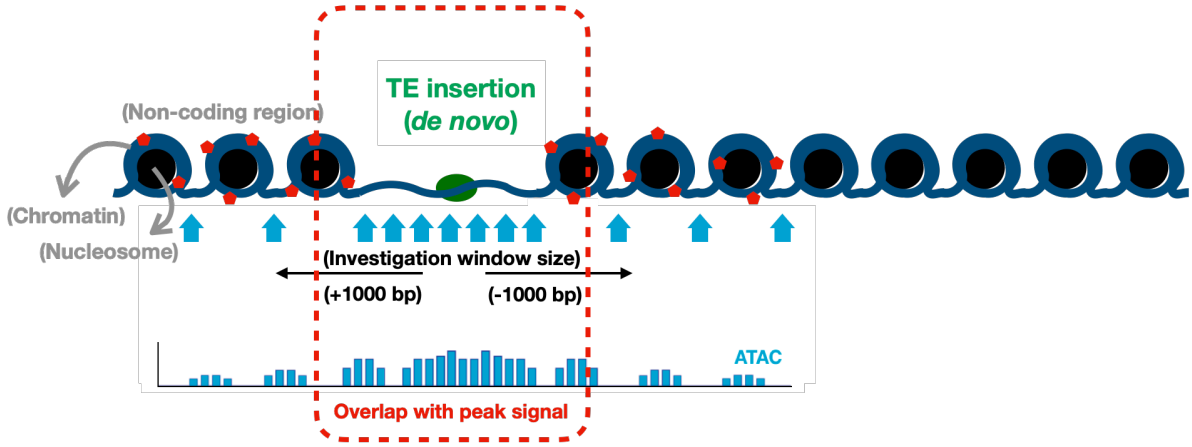
(score > 0) (All site)	AITE	No AITE
ATAC_peak	4854	5205
No ATAC_peak	9291	10679

#data: matrix(c(4854, 9291, 5205, 10679), ncol = 2)
#Chi-squared = 7.9707, df = 1, p-value = 0.004754

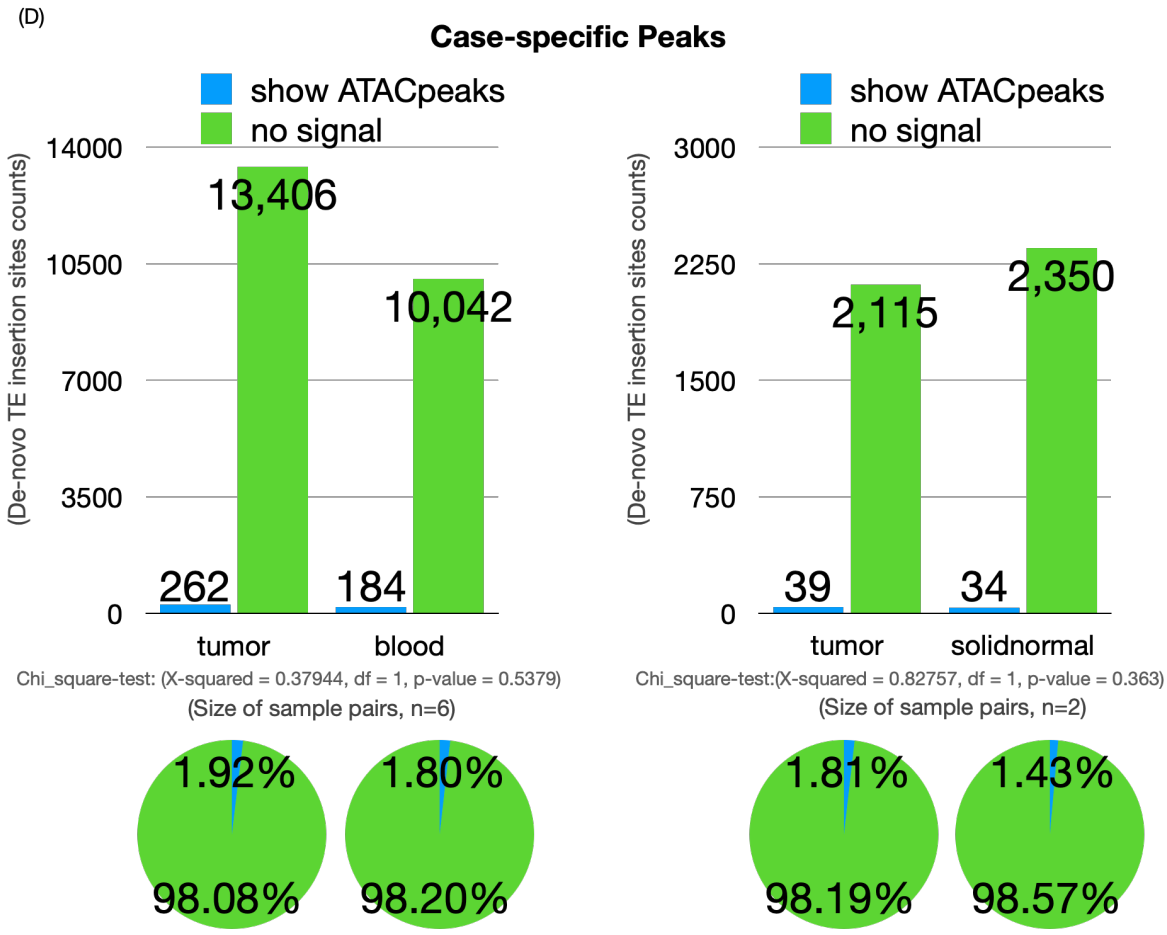
odds-ratio=1.07
(Although significant but not obvious)

(C)

ATAC-Seq peak signals at *de novo* TE insertion sites



How many TE insertion sites overlap with ATACseq peaks?



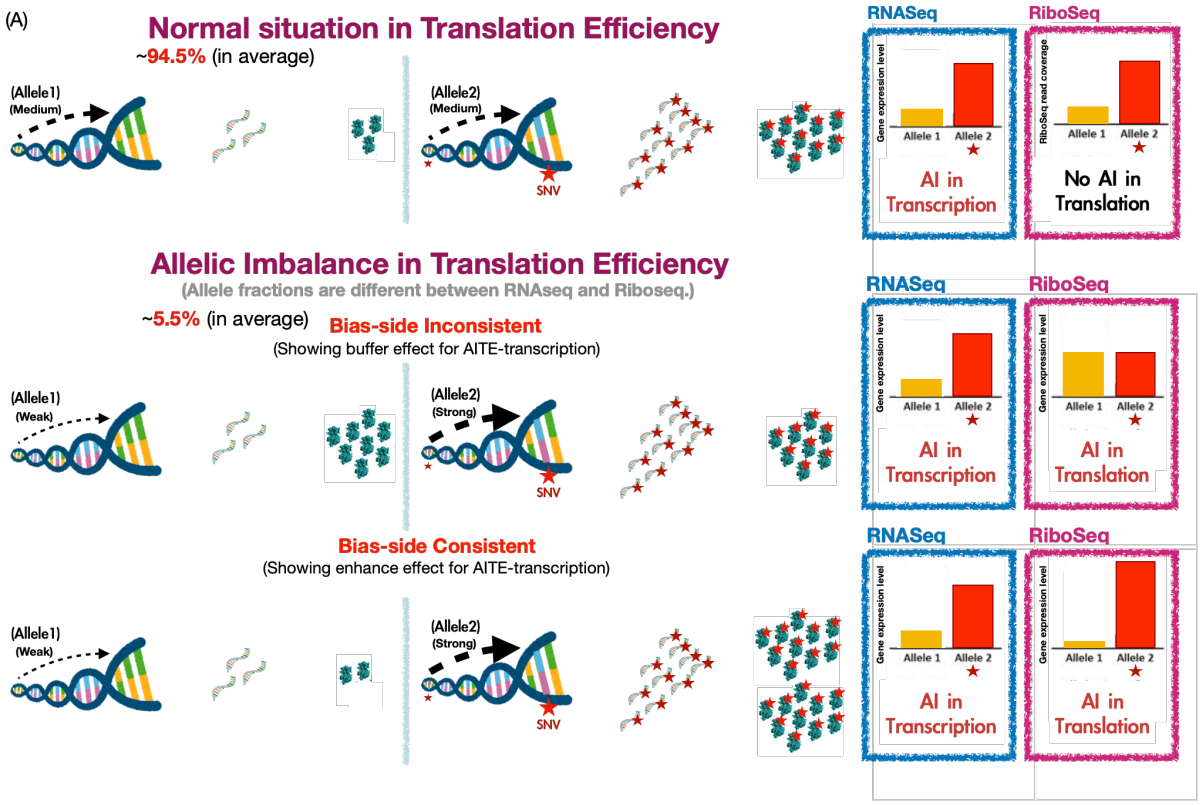
Supplement Figure 14. ATAC-Seq signals at *de novo* TE insertion sites or at the non-coding regions arounds AITE genes.

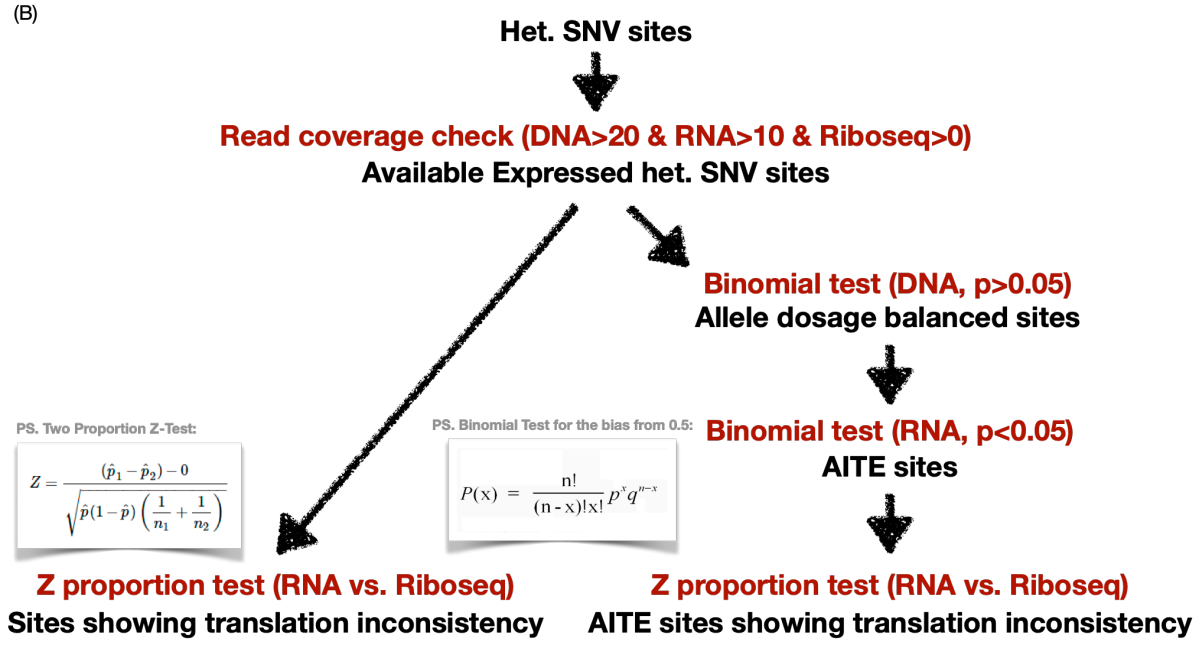
(A) Methods for investigating ATAC-Seq signals at the non-coding regions around AITE genes.

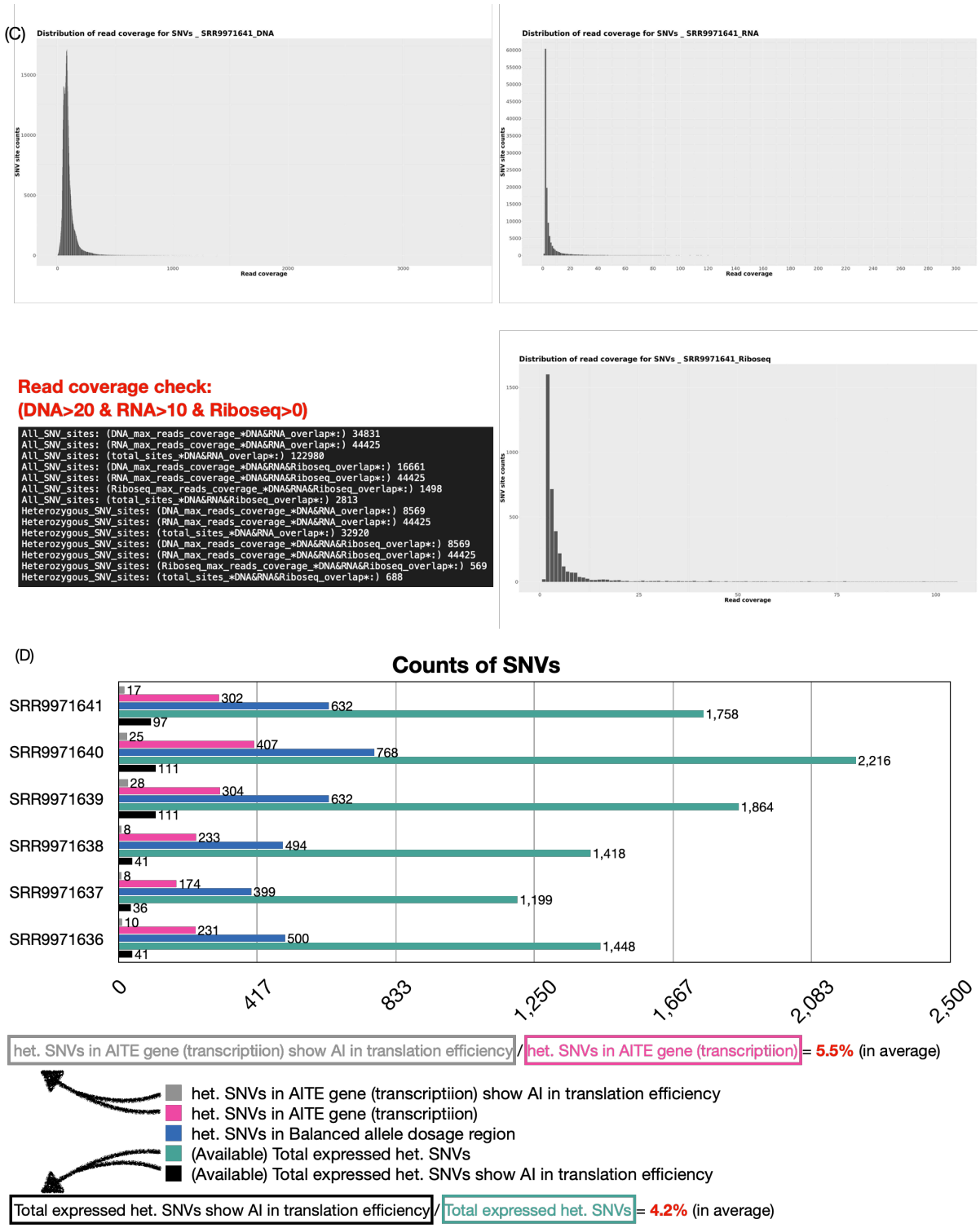
(B) Chi-square test for examing the independence of between chromatin openness (ATAC-Seq peaks) and AITE effect occurrence.

(C) Methods for investigating ATAC-Seq peak signals at *de novo* TE insertion sites.

(D) Proportion of the detection of chromatin openness (ATAC-Seq peaks) at *de novo* TE insertion sites.


(A)
Normal situation in Translation Efficiency
~94.5% (in average)
(Allele1)
(Medium)
(Allele2)
(Medium)
SNV
RNASeq
RiboSeq
Gene expression level
RiboSeq read coverage
Allele 1
Allele 2
AI in Transcription
No AI in Translation
Allelic Imbalance in Translation Efficiency
(Allele fractions are different between RNAseq and Riboseq.)
~5.5% (in average)
Bias-side Inconsistent
(Showing buffer effect for AITE-transcription)
(Allele1)
(Weak)
(Allele2)
(Strong)
SNV
AI in Translation
Bias-side Consistent
(Showing enhance effect for AITE-transcription)


(B)
Het. SNV sites
Read coverage check (DNA>20 & RNA>10 & Riboseq>0)
Available Expressed het. SNV sites
Binomial test (DNA, p>0.05)
Allele dosage balanced sites
Binomial test (RNA, p<0.05)
AITE sites
PS. Two Proportion Z-Test:
$Z = \frac{(\hat{p}_1 - \hat{p}_2) - 0}{\sqrt{\hat{p}(1-\hat{p})\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}$
PS. Binomial Test for the bias from 0.5:
$P(x) = \frac{n!}{(n-x)!x!} p^x q^{n-x}$
Z proportion test (RNA vs. Riboseq)
Sites showing translation inconsistency
Z proportion test (RNA vs. Riboseq)
AITE sites showing translation inconsistency



Supplement Figure 15. Investing the existence of Allelic Imbalance in translational efficiency for buffering AITE effect.

(A) It is described in details for how allelic imbalance in translation efficiency exist in those genes with allelic imbalance in transcription efficiency effect. And it also shows how the allele-specific read coverages will change under different situations in RNA-Seq and Ribo-Seq.

(B) The detailed bioinformatics work flow to detect allelic imbalance in translation efficiency is visualized here. Read coverage threshold for the available expressed heterozygous SNV sites are set. Binomial test is applied to detect those allelic imbalance in transcription efficiency (AITE) effects among those threshold passed sites. And

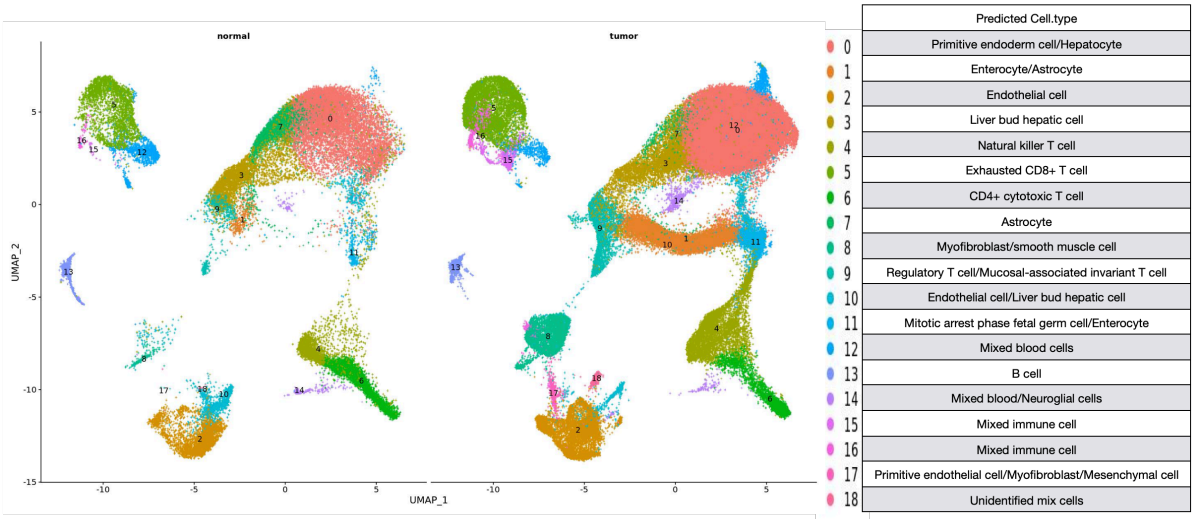
Z-proportion test is applied to detect those allelic imbalance in translation efficiency effects among those threshold passed sites.

(C) Quality control through read coverages in WESeq, RNA-Seq and Ribo-Seq. Here only took SRR9971641 as an example.

(D) Here calculates about the occurring rate of allelic imbalance in translational efficiency. After going through the workflows which is designed for detecting allelic imbalance in transcription efficiency (AITE) and for detecting allelic imbalance in translation efficiency (AI in translation efficiency), among those heterozygous SNVs we can know which sites are expressed (shallow green), which sites show the AITE (pink), which sites show the AI in translation efficiency (black), and also which sites simultaneously show the AITE and AI in translation efficiency (grey). The above figure shows the accumulated counts of SNVs under different situations for each sample genome, and the below part shows the formula for calculating the occurring rate of allelic imbalance in translation efficiency. And the occurrence rates of allelic imbalance in translation efficiency were calculated for the expressed heterozygous genes and for the AITE genes, separately.

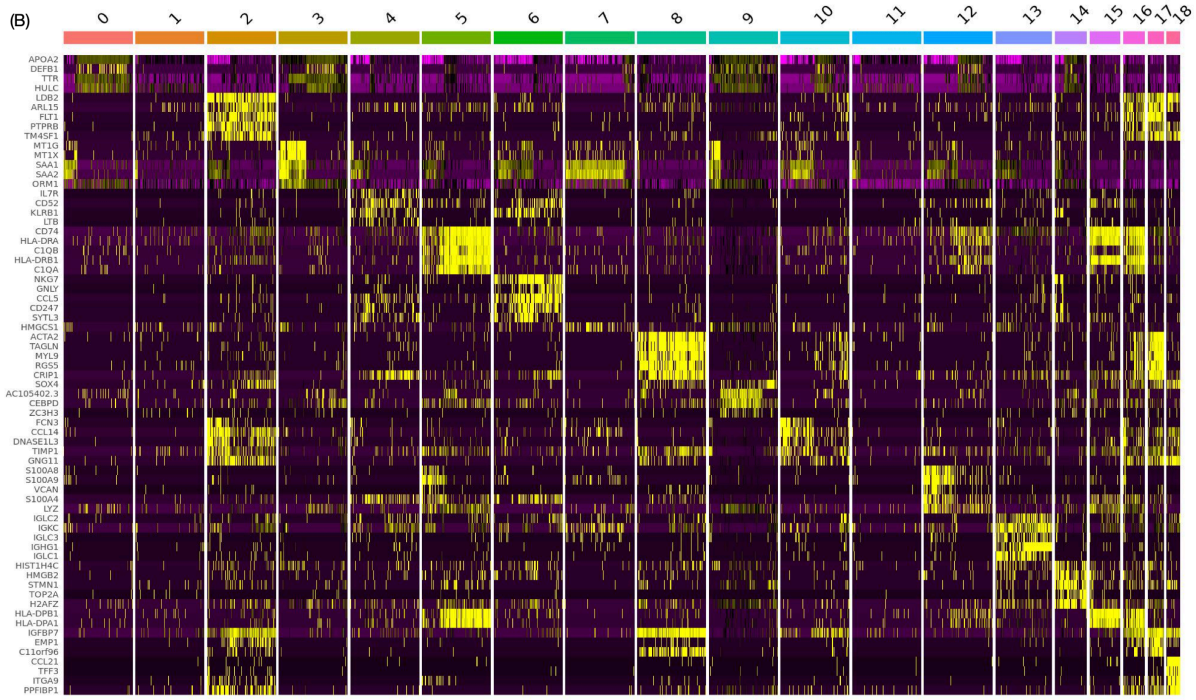
(A)

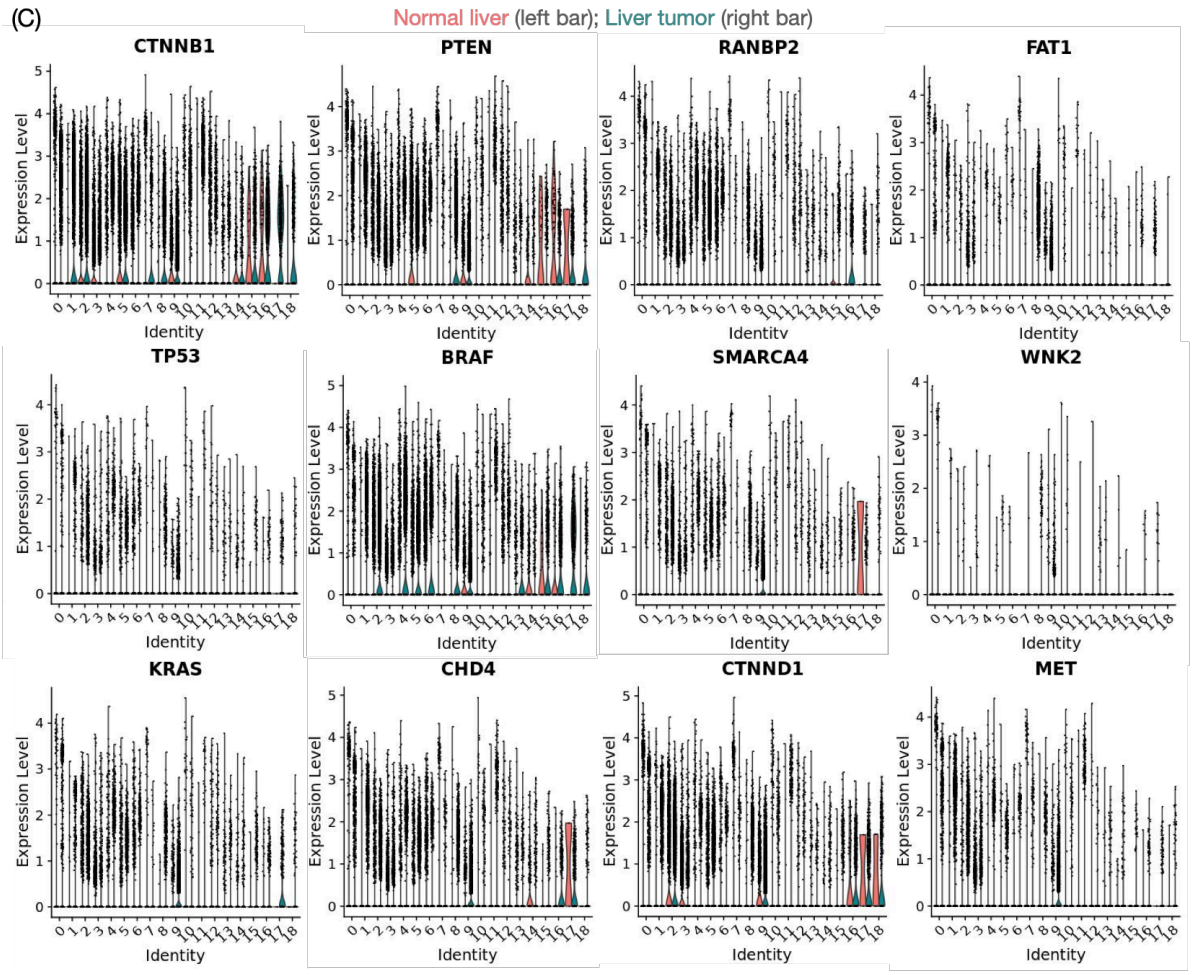
Single-cells RNA-Seq from normal liver and liver tumor tissues:

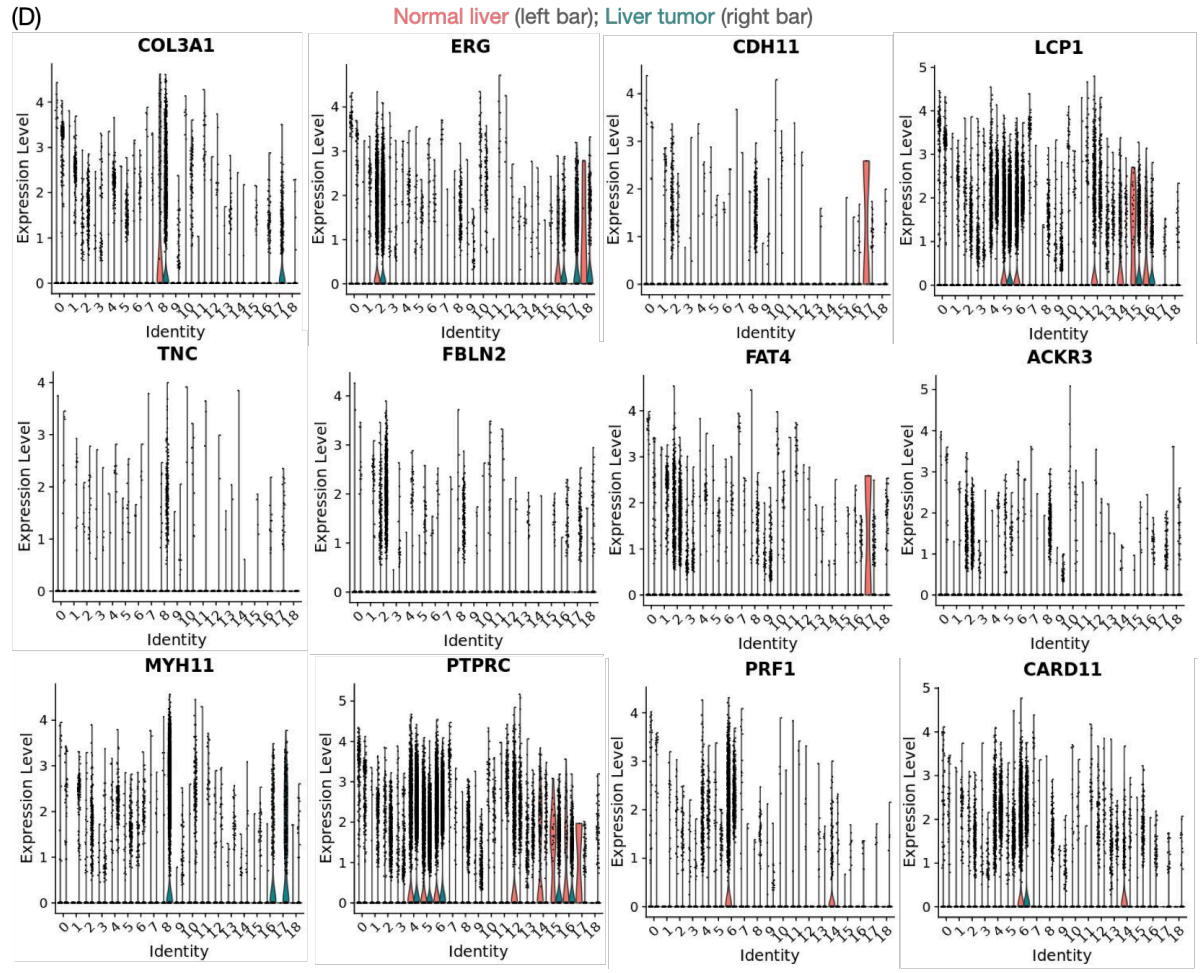


	Predicted Cell.type
0	Primitive endoderm cell/Hepatocyte
1	Enterocyte/Astrocyte
2	Endothelial cell
3	Liver bud hepatic cell
4	Natural killer T cell
5	Exhausted CD8+ T cell
6	CD4+ cytotoxic T cell
7	Astrocyte
8	Myofibroblast/smooth muscle cell
9	Regulatory T cell/Mucosal-associated invariant T cell
10	Endothelial cell/Liver bud hepatic cell
11	Mitotic arrest phase fetal germ cell/Enterocyte
12	Mixed blood cells
13	B cell
14	Mixed blood/Neuroglial cells
15	Mixed immune cell
16	Mixed immune cell
17	Primitive endothelial cell/Myofibroblast/Mesenchymal cell
18	Unidentified mix cells

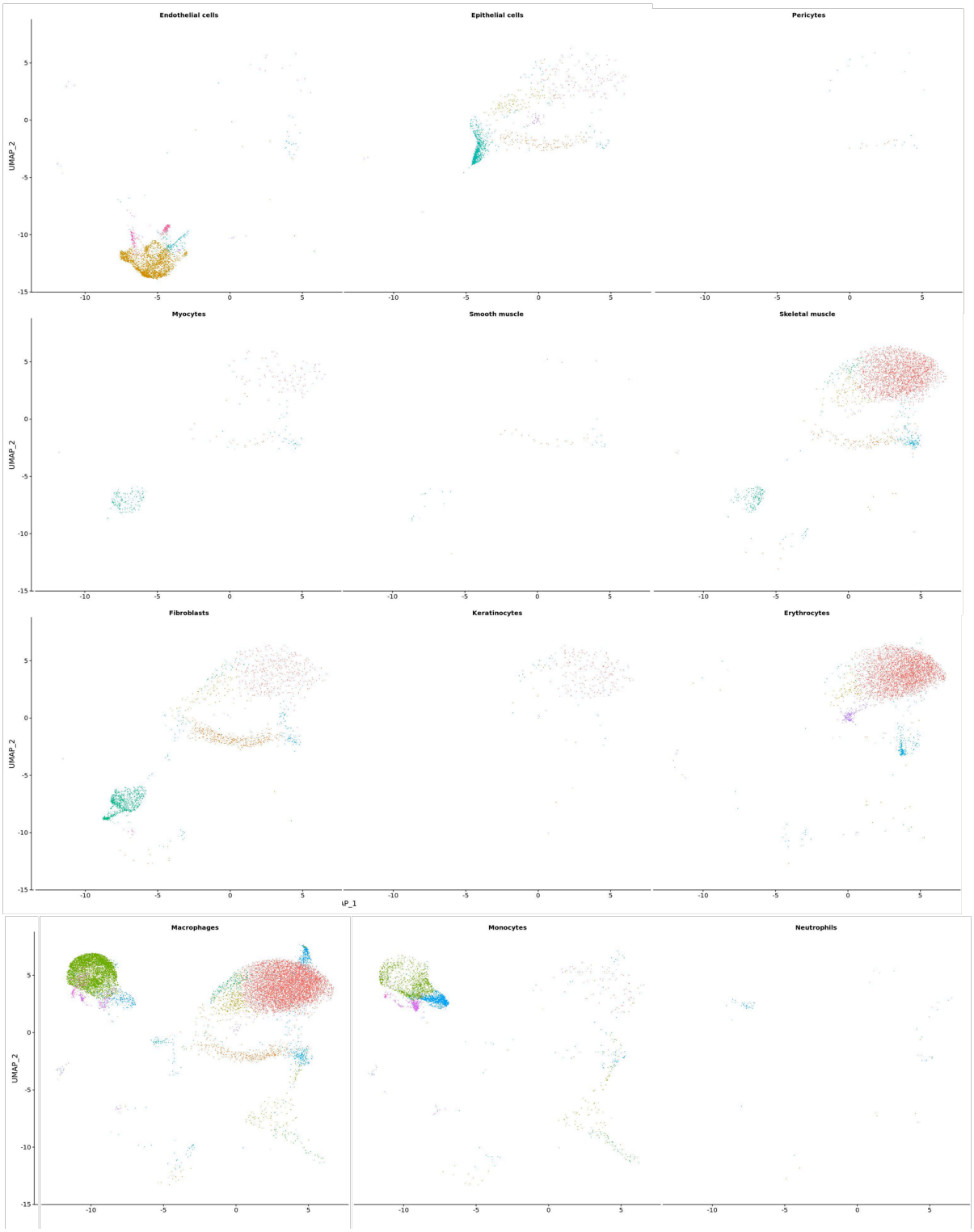
(B)

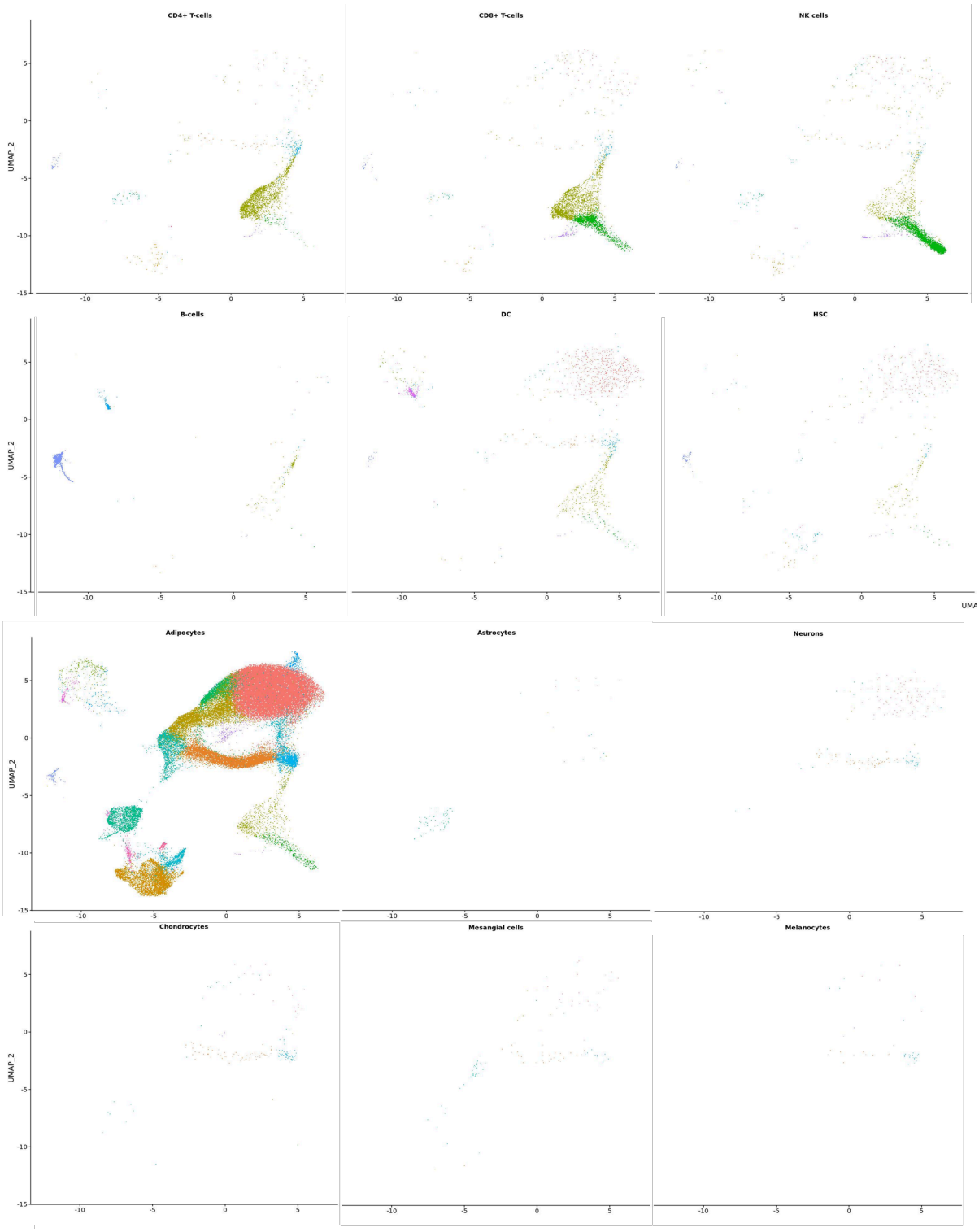


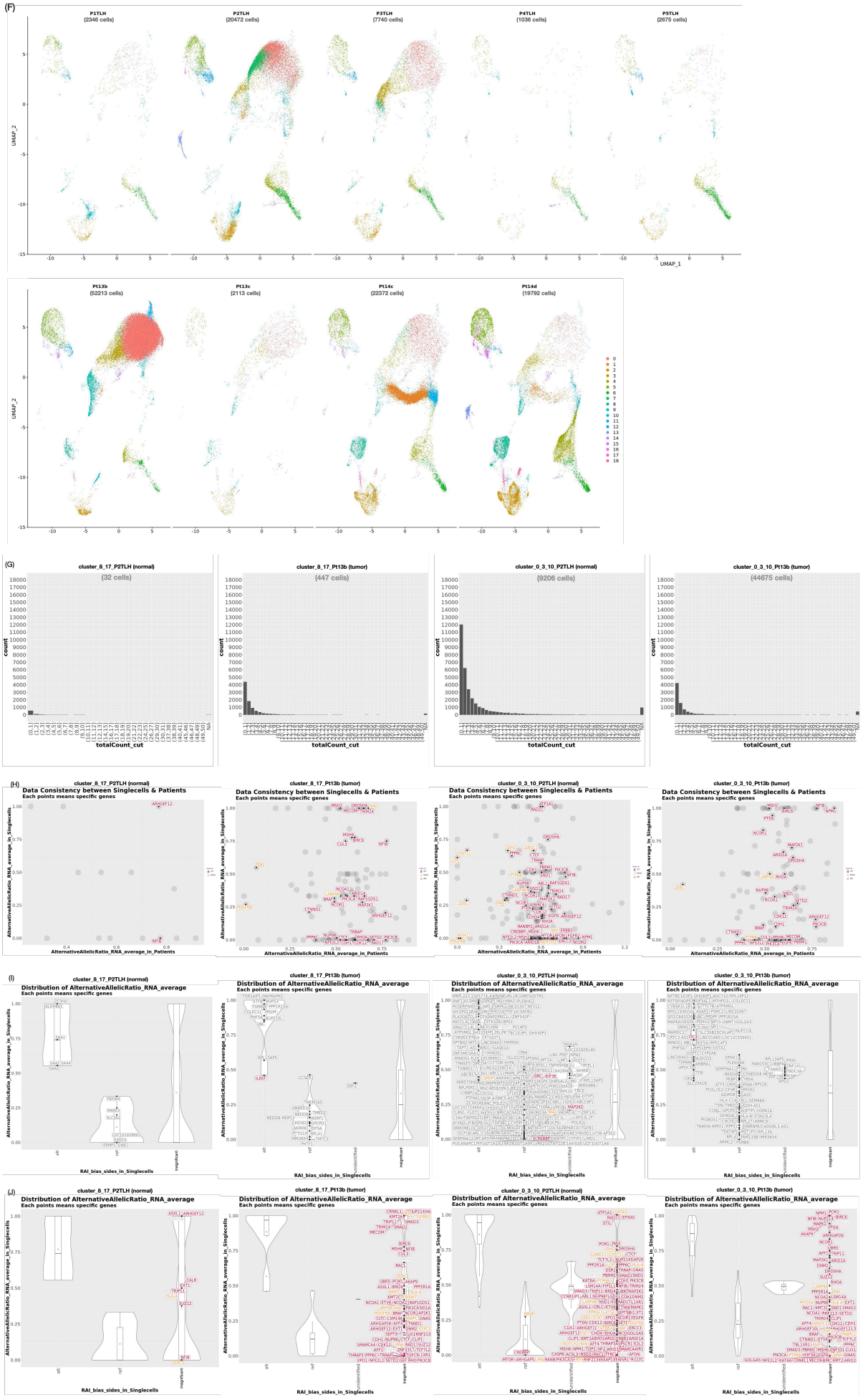

(C)
Normal liver (left bar); Liver tumor (right bar)
CTNNB1
PTEN
RANBP2
FAT1
TP53
BRAF
SMARCA4
WNK2
KRAS
CHD4
CTNND1
MET
Expression Level
Identity

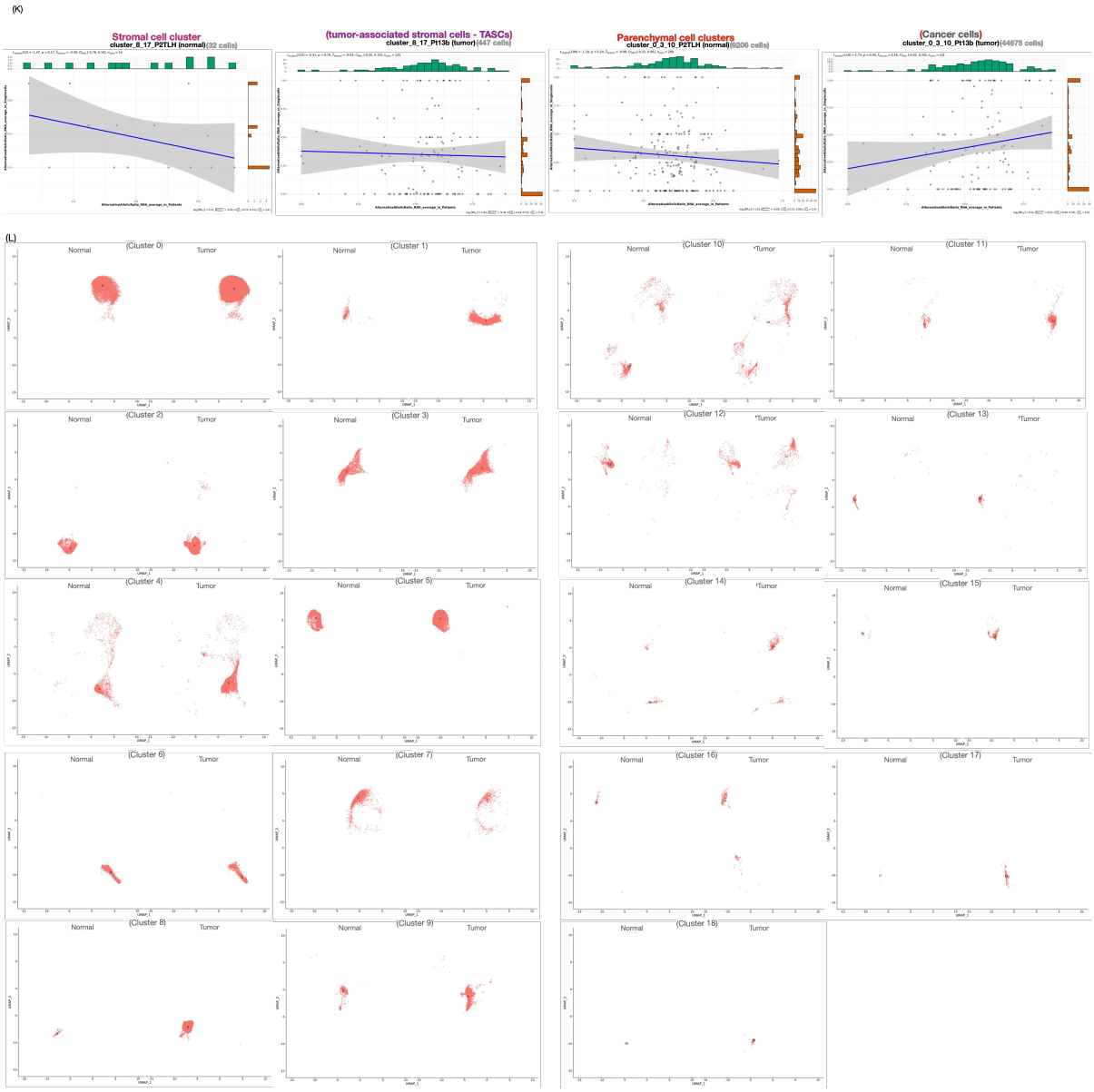

(D)
Normal liver (left bar); Liver tumor (right bar)
COL3A1
ERG
CDH11
LCP1
TNC
FBLN2
FAT4
ACKR3
MYH11
PTPRC
PRF1
CARD11
Expression Level
Identity

(E)


Endothelial cells
Epithelial cells
Pericytes
Myocytes
Smooth muscle
Skeletal muscle
Fibroblasts
Keratinocytes
Erythrocytes
Macrophages
Monocytes
Neutrophils
UMAP_2
UMAP_1


CD4+ T-cells
CD8+ T-cells
NK cells
B-cells
DC
HSC
Adipocytes
Astrocytes
Neurons
Chondrocytes
Mesangial cells
Melanocytes
UMAP_2


(F)
P1TLH (2346 cells)
P2TLH (20472 cells)
P3TLH (7740 cells)
P4TLH (1036 cells)
P5TLH (2675 cells)
Pt13b (52213 cells)
Pt13c (2113 cells)
Pt14c (22372 cells)
Pt14d (19792 cells)
UMAP_1
UMAP_2
(G)
cluster_8_17_P2TLH (normal)
(32 cells)
cluster_8_17_Pt13b (tumor)
(447 cells)
cluster_0_3_10_P2TLH (normal)
(9206 cells)
cluster_0_3_10_Pt13b (tumor)
(44675 cells)
count
totalCount_cut
(H)
Data Consistency between Singlecells & Patients
Each points means specific genes
AlternativeAllelicRatio_RNA_average_in_Singlecells
AlternativeAllelicRatio_RNA_average_in_Patients
(I)
Distribution of AlternativeAllelicRatio_RNA_average
Each points means specific genes
RAI_bias_sides_in_Singlecells
(J)
Distribution of AlternativeAllelicRatio_RNA_average
Each points means specific genes
RAI_bias_sides_in_Singlecells



Supplement Figure 16. Single-cells RNA-Seqs analysis for the normal liver and liver tumor tissues (10X genomics platform).

(A) UMAP clustering for the single-cells RNA-Seq from normal liver and liver tumor tissues under 10X genomics platform.

(B) Hot-map for the top 5 up-regulated genes in each identified cluster.

(C) Gene expression levels for the single-cells in each identified cluster for those AITE genes with alternate allele bias side. Y-axis indicates gene expression levels, and the X-axis indicates the clusters. For each cluster, the cells from normal tissues are in the left bar and the cells from tumor tissues are in the right bar.

(D) Gene expression levels for the single-cells in each identified cluster for those AITE genes with reference allele bias side. Y-axis indicates gene expression levels, and the X-axis indicates the clusters. For each cluster, the cells from normal tissues are in the left bar and the cells from tumor tissues are in the right bar.

(E) Predicting the cell clusters by artificial intelligent software under the information being provided by "Blueprint Encode" Data.

(F) The single cells in UMAP classified clusters from different tissue samples, where the single cells in specific clusters from sample "P2TLH" and sample "Pt13b" were selected for down-stream analysis.

(G) Quality control for the read coverages in specific single-cells merged pseudo-bulk RNA-Seqs.

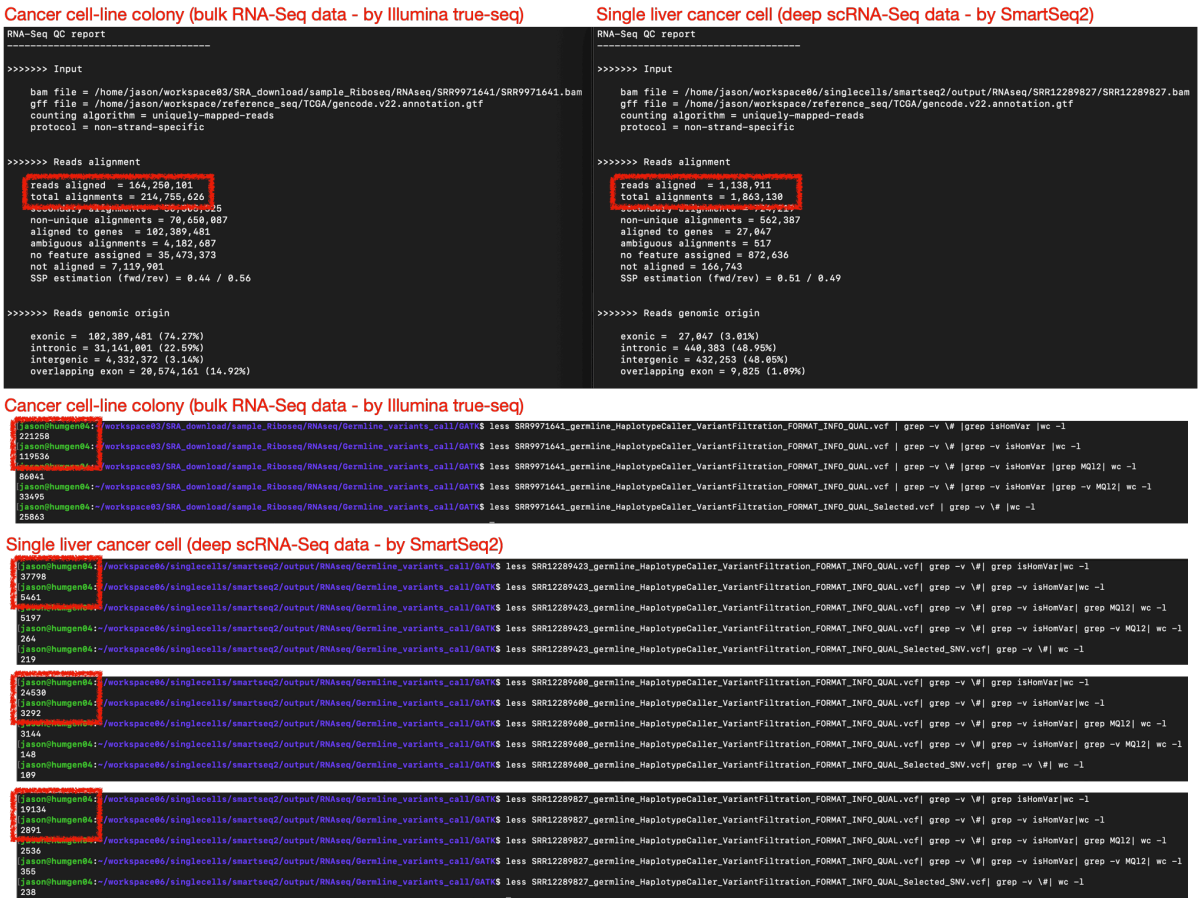
(H) Correlation between the bulk tissues based data and single-cells based data for the alternative allelic ratio of each expressed heterozygous genes.

(I) The biased sides of RNA allelic imbalance in single cells (the significant ones).

(J) The biased sides of RNA allelic imbalance in single cells (the non-significant ones).

(K) The corresponding spearman's correlation coefficients and p-values for (H).

(L) Details for the positions of different cell clusters in UMAP map for the single-cells RNA-Seqs from normal liver and liver tumor tissues.



Supplement Figure 17. Deep single-cells RNA-Seq analysis for the liver tumors and paired normal-like tissues (Smart-seq2 platform).

The heterozygous rates among those expressed heterozygous SNV sites were calculated in different samples, which includes “liver cancer cell-line single colony (bulk RNA-Seq - by Illumina true-seq)”, “liver tumor tissue (pseudo bulk scRNA-Seq - by 10X genomics)” and “Single liver cancer cell (deep scRNA-Seq - by SmartSeq2)”.

APPENDIX B: SUPPLEMENTARY TABLES

Supplement Table 1. Collected sample size for each type of Pan-adenocarcinomas.

Among those Pan-adenocarcinomas patients registered in TCGA database, the simultaneous provision of tumor tissues and nearby paired normal-like tissues can be found in only a few of the cancer patients, because of the hardness in normal-like tissues collection. Among Pan-adenocarcinomas, collected tumor sample counts in each cancer type were showed in the third column. Among them, sample counts in tumors with its paired normal-like tissues were further shown in the fourth column, which are in around 0.5% of the collected cases.

Cancer Types	ID	Tumor sample counts	Paired normal-like sample counts
Liver Hepatocellular Carcinoma	**(LIHC)**	320	13
Stomach Adenocarcinoma	**(STAD)**	336	18
Colon Adenocarcinoma	**(COAD)**	382	20
Lung Adenocarcinoma	**(LUAD)**	404	17
Prostate Adenocarcinoma	**(PRAD)**	433	37
Rectum Adenocarcinoma	**(READ)**	140	4
Kidney Renal Clear Cell Carcinoma	**(KIRC)**	88	3
Kidney Renal Papillary Cell Carcinoma	**(KIRP)**	226	10
Thyroid Carcinoma	**(THCA)**	418	18
Uterine Corpus Endometrial Carcinoma	**(UCEC)**	509	12
Total cases		3256	152

Supplement Table 2. The selected cases of Pan-adenocarcinoma (n=3256).

TCGA-AX-A3GI, TCGA-AX-A05Z, TCGA-BG-A0LW, TCGA-AJ-A3BI, TCGA-EO-A1Y7, TCGA-EY-A5W2, TCGA-AP-A0L9, TCGA-DI-A1BU, TCGA-D1-A17K, TCGA-BS-A0UA, TCGA-BG-A0MI, TCGA-D1-A1NX, TCGA-BK-A26L, TCGA-B5-A11L, TCGA-AX-A2HG, TCGA-B5-A11M, TCGA-B5-A0JY, TCGA-A5-A0G3, TCGA-B5-A1MV, TCGA-DI-A2QT, TCGA-BS-A0TG, TCGA-EO-A2CG, TCGA-BG-A222, TCGA-B5-A11F, TCGA-D1-A174, TCGA-BS-A0TE, TCGA-PG-A915, TCGA-B5-A0K7, TCGA-EO-A3AY, TCGA-A5-A0GP, TCGA-AP-A05D, TCGA-AX-A06F, TCGA-4E-A92E, TCGA-BG-A0YU, TCGA-B5-A5OC, TCGA-D1-A2G0, TCGA-BS-A0V7, TCGA-EY-A1GO, TCGA-FI-A2EX, TCGA-BG-A0W1, TCGA-AJ-A3BH, TCGA-D1-A17Q, TCGA-E6-A2P8, TCGA-AX-A0IS, TCGA-D1-A3DA, TCGA-D1-A17M, TCGA-D1-A167, TCGA-A5-A2K7, TCGA-AX-A0IZ, TCGA-PG-A5BC, TCGA-EY-A1GS, TCGA-EY-A1GR, TCGA-B5-A3FB, TCGA-BK-A139, TCGA-DF-A2KV, TCGA-BG-A0VW, TCGA-B5-A0JZ, TCGA-AP-A051, TCGA-B5-A0JU, TCGA-B5-A11V, TCGA-AJ-A23N, TCGA-EY-A1GK, TCGA-BS-A0T9, TCGA-EY-A1GI, TCGA-BS-A0V6, TCGA-PG-A7D5, TCGA-BK-A6W4, TCGA-AX-A064, TCGA-D1-A1O0, TCGA-D1-A1NW, TCGA-B5-A11I, TCGA-D1-A0ZS, TCGA-B5-A11N, TCGA-D1-A1NY, TCGA-EO-A22Y, TCGA-EO-A3AS, TCGA-AP-A1E0, TCGA-AP-A0LM, TCGA-AX-A05Y, TCGA-D1-A179, TCGA-AP-A0L8, TCGA-AP-A5FX, TCGA-D1-A2G5, TCGA-A5-A0GA, TCGA-AP-A054, TCGA-AJ-A3BK, TCGA-AP-A059, TCGA-D1-A16S, TCGA-D1-A16E, TCGA-AJ-A3NH, TCGA-AX-A0IU, TCGA-D1-A17D, TCGA-D1-A101, TCGA-BS-A0WQ, TCGA-A5-A0GX, TCGA-BG-A0MG, TCGA-D1-A17F, TCGA-BK-A0CB, TCGA-D1-A17U, TCGA-A5-A0RA, TCGA-EY-A1H0, TCGA-FI-A2EW, TCGA-FI-A2F9, TCGA-AJ-A8CW, TCGA-BG-A0W2, TCGA-AP-A0LF, TCGA-BG-A0VV, TCGA-AJ-A3EJ, TCGA-B5-A11U, TCGA-AP-A052, TCGA-AJ-A3OK, TCGA-AP-A1E4, TCGA-AP-A0LP, TCGA-BG-A0M7, TCGA-D1-A16I, TCGA-D1-A17B, TCGA-B5-A11W, TCGA-EO-A2CH, TCGA-B5-A0K8, TCGA-AJ-A5DW, TCGA-B5-A1MY, TCGA-AP-A1DP, TCGA-AX-A3FV, TCGA-AX-A05T, TCGA-D1-A16N, TCGA-BK-A13C, TCGA-DF-A2KU, TCGA-AX-A3FS, TCGA-BS-A0U7, TCGA-A5-A7WK, TCGA-BS-A0TD, TCGA-EY-A1GH, TCGA-QS-A5YQ, TCGA-E6-A1LX, TCGA-D1-A1NS, TCGA-B5-A11O, TCGA-E6-A2P9, TCGA-D1-A176, TCGA-BG-A0M4, TCGA-D1-A1NU, TCGA-AX-A3G6, TCGA-D1-A161, TCGA-BS-A0UJ, TCGA-B5-A0K0, TCGA-FI-A2D5, TCGA-BK-A0CA, TCGA-FI-A2F4, TCGA-D1-A175, TCGA-B5-A0JT, TCGA-A5-A0G2, TCGA-AX-A3G8, TCGA-BG-A0M3, TCGA-A5-A0R7, TCGA-FI-A2D0, TCGA-AJ-A3I9, TCGA-BG-A0LX, TCGA-D1-A17N, TCGA-A5-A3LO, TCGA-D1-A17A, TCGA-AX-A1CP, TCGA-A5-A0R9, TCGA-EY-A212, TCGA-AX-A3G9, TCGA-AJ-A2QK, TCGA-AP-A1DR, TCGA-EY-A1GT, TCGA-EY-A1GW, TCGA-BG-A0M9, TCGA-A5-A1OJ, TCGA-D1-A0ZN, TCGA-A5-A0R8, TCGA-BG-A3PP, TCGA-EY-A72D, TCGA-EY-A1GP, TCGA-FI-A2CY, TCGA-D1-A0ZV, TCGA-AP-A1E1, TCGA-AP-A1DH, TCGA-A5-A2K2, TCGA-BG-A0MO, TCGA-EY-A2OP, TCGA-AP-A0LG, TCGA-D1-A2G7, TCGA-AP-A1DQ, TCGA-BG-A0M0, TCGA-AP-A0LJ, TCGA-B5-A11E, TCGA-KJ-A3U4, TCGA-EY-A1GD, TCGA-D1-A168, TCGA-FI-A2D4, TCGA-5B-A90C, TCGA-D1-A16B, TCGA-BG-A0MU, TCGA-A5-A0GB, TCGA-B5-A11R, TCGA-B5-A0JN, TCGA-AX-A2IN, TCGA-A5-A0GQ, TCGA-AX-A06L, TCGA-AX-A3G3, TCGA-D1-A1NZ, TCGA-EO-A3KW, TCGA-AX-A05W, TCGA-AP-A1DV, TCGA-A5-A3LP, TCGA-BG-A18B, TCGA-EY-A2OQ, TCGA-AX-A1CA, TCGA-QS-A8F1, TCGA-DF-A2KS, TCGA-BK-A4ZD, TCGA-B5-A11S, TCGA-PG-A916, TCGA-AP-A0LI, TCGA-B5-A0JR, TCGA-EO-A3L0, TCGA-QF-A5YT, TCGA-5S-A9Q8, TCGA-D1-A0ZQ, TCGA-AX-A06B, TCGA-B5-A0JX, TCGA-B5-A0JV, TCGA-FI-A3PX, TCGA-E6-A1LZ, TCGA-DI-A0WH, TCGA-EY-A548, TCGA-D1-A15W, TCGA-D1-A103, TCGA-AJ-A3QS, TCGA-EO-A3AV, TCGA-AX-A1C4, TCGA-QF-A5YS, TCGA-JU-AAVI, TCGA-B5-A1MW, TCGA-AX-A2HH, TCGA-B5-A3FA, TCGA-BG-A18C, TCGA-AX-A2HK, TCGA-B5-A3FC, TCGA-B5-A3F9, TCGA-AX-A1C9, TCGA-DF-A2KR, TCGA-AP-A0LD, TCGA-AJ-A23O, TCGA-AW-A1PO, TCGA-BG-A187, TCGA-D1-A3DH, TCGA-EY-A1GF, TCGA-D1-A17T, TCGA-AJ-A8CV, TCGA-B5-A11P, TCGA-KP-A3W1, TCGA-D1-A0ZU, TCGA-BG-A220, TCGA-EO-A3KX, TCGA-B5-A11Z, TCGA-AJ-A2QN, TCGA-D1-A16R, TCGA-BG-A0YV, TCGA-AX-A3G4, TCGA-DF-A2KY, TCGA-BG-A0M2, TCGA-B5-A11J, TCGA-AP-A0LS, TCGA-BG-A0MS, TCGA-A5-A1OF, TCGA-EC-A24G, TCGA-AX-A06H, TCGA-D1-A16O, TCGA-D1-A2G6, TCGA-A5-A0G5, TCGA-D1-A0ZZ, TCGA-BS-A0V8, TCGA-AX-A06D, TCGA-AJ-A3OL, TCGA-BG-A0MK, TCGA-BG-A221, TCGA-AX-A3FX, TCGA-D1-A163, TCGA-A5-A7WJ, TCGA-B5-A11X, TCGA-BS-A0TC, TCGA-EY-A1GM, TCGA-AJ-A3EL, TCGA-A5-A1OK, TCGA-AX-A063, TCGA-EO-A22X, TCGA-A5-A0GW, TCGA-A5-A0GD, TCGA-AP-A05H, TCGA-A5-A0GE, TCGA-B5-A0K1, TCGA-BS-A0TI, TCGA-AX-A1C8, TCGA-A5-A0VQ, TCGA-BS-A0TA, TCGA-BG-A0MQ, TCGA-AJ-A8CT, TCGA-D1-A3JQ, TCGA-SL-A6JA, TCGA-BK-A13B, TCGA-AX-A1C5, TCGA-B5-A11G, TCGA-A5-A0GH, TCGA-KP-A3VZ, TCGA-B5-A1N2, TCGA-BS-A0UF, TCGA-D1-A0ZP, TCGA-D1-A16G, TCGA-BG-A2L7, TCGA-AX-A3G7, TCGA-B5-A0K3, TCGA-PG-A6IB, TCGA-D1-A17C, TCGA-FI-A2D6, TCGA-AP-A056, TCGA-AP-A05N, TCGA-EO-A1Y8, TCGA-A5-A0GR, TCGA-B5-A1MX, TCGA-A5-A0G9, TCGA-D1-A16D, TCGA-A5-A0GM, TCGA-EY-A2OM, TCGA-BG-A0VX, TCGA-DI-A1BY, TCGA-B5-A1MZ, TCGA-BG-A0RY, TCGA-A5-A2K4, TCGA-BK-A6W3, TCGA-EY-A54A, TCGA-D1-A16J, TCGA-AX-A05S, TCGA-A5-A2K5, TCGA-AX-A0J0, TCGA-EY-A3L3, TCGA-BG-A2AE, TCGA-AP-A0LO, TCGA-DF-A2KZ, TCGA-AP-A0LV, TCGA-EO-A3KU, TCGA-AJ-A3NG, TCGA-AP-A0LL, TCGA-FI-A2F8, TCGA-EY-A1G7, TCGA-D1-A16Q, TCGA-BG-A0MC, TCGA-A5-A2K3, TCGA-AP-A053, TCGA-BG-A18A, TCGA-AJ-A3OJ, TCGA-PG-A914, TCGA-B5-A1MS, TCGA-A5-A0GV, TCGA-EY-A2OO, TCGA-SL-A6J9, TCGA-AX-A06J, TCGA-AJ-A23M, TCGA-AP-A0LH, TCGA-BK-A56F, TCGA-A5-A0G1, TCGA-AP-A1E3, TCGA-BG-A0M8, TCGA-D1-A160, TCGA-D1-A165, TCGA-AP-A05P, TCGA-KP-A3W0, TCGA-AJ-A5DV, TCGA-FI-A2EY, TCGA-B5-A0K9, TCGA-DF-A2KN, TCGA-D1-A102, TCGA-AP-A05O, TCGA-AX-A3FZ, TCGA-EY-A1GL, TCGA-B5-A0K6, TCGA-FI-A2D2, TCGA-A5-A1OG, TCGA-AX-A3G1, TCGA-BG-A186, TCGA-EY-A214, TCGA-EY-A1GV, TCGA-D1-A17R, TCGA-EY-A2ON, TCGA-EC-A1QX, TCGA-AJ-A3NF, TCGA-AJ-A3BG, TCGA-EY-A1GQ, TCGA-BK-A0CC, TCGA-EY-A3QX, TCGA-A5-A0R6, TCGA-D1-A16Y, TCGA-A5-A0GG, TCGA-EO-A3AZ, TCGA-B5-A3FD, TCGA-B5-A5OE, TCGA-KP-A3W4, TCGA-AP-A0LE, TCGA-EY-A1GE, TCGA-AX-A1CR, TCGA-D1-A3DG, TCGA-BS-A0UL, TCGA-A5-A0GJ, TCGA-A5-A0GN, TCGA-BS-A0VI, TCGA-BG-A0VZ, TCGA-DF-A2L0, TCGA-BS-A0U8, TCGA-AJ-A2QM, TCGA-AX-A1CE, TCGA-D1-A17S, TCGA-E6-A8L9, TCGA-BS-A0UM, TCGA-EO-A1Y5, TCGA-D1-A16V, TCGA-EO-A3AU, TCGA-KP-A3W3, TCGA-AX-A0J1, TCGA-EY-A4KR, TCGA-B5-A3FH, TCGA-B5-A11Y, TCGA-D1-A1O7, TCGA-BS-A0V4, TCGA-AX-A1CN, TCGA-AX-A2H7, TCGA-E6-A1M0, TCGA-AP-A0LN, TCGA-AX-A1C7, TCGA-A5-A0VP, TCGA-AP-A05J, TCGA-AX-A2HJ, TCGA-EO-A3B0, TCGA-AP-A1DK, TCGA-D1-A0ZO, TCGA-AX-A1CC, TCGA-AX-A05U, TCGA-AJ-A3NC, TCGA-B5-A0KB, TCGA-AP-A1DM, TCGA-BG-A0VT, TCGA-AJ-A3NE, TCGA-SJ-A6ZI, TCGA-AX-A2HF, TCGA-BG-A0M6, TCGA-QS-A5YR, TCGA-BS-A0TJ, TCGA-B5-A1MR, TCGA-B5-A11Q, TCGA-AJ-A3TW, TCGA-D1-A1O5, TCGA-D1-A15V, TCGA-AJ-A2QL, TCGA-FI-A2EU, TCGA-B5-A3S1, TCGA-BG-A0MT, TCGA-BS-A0UT, TCGA-AX-A3FT, TCGA-EO-A22U, TCGA-B5-A0JS, TCGA-AX-A0IW, TCGA-EO-A3B1, TCGA-B5-A0K4, TCGA-EY-A1GC, TCGA-AJ-A3BF, TCGA-BS-A0U5, TCGA-A5-AB3J, TCGA-D1-A17L, TCGA-AP-A1DO, TCGA-EY-A1G8, TCGA-QS-A744, TCGA-K6-A3WQ, TCGA-FI-A3PV, TCGA-FI-A2CX, TCGA-B5-A121, TCGA-EY-A215, TCGA-H5-A2HR, TCGA-EY-A1GX, TCGA-EY-A1GU, TCGA-AP-A05A, TCGA-BS-A0UV, TCGA-A5-A0GU, TCGA-B5-A11H, TCGA-B5-A1MU, TCGA-EC-A1NJ, TCGA-D1-A17H, TCGA-AX-A2IO, TCGA-D1-A15Z, TCGA-AJ-A3IA, TCGA-SJ-A6ZJ, TCGA-D1-A3JP, TCGA-D1-A0ZR, TCGA-B5-A0K2, TCGA-EY-A547, TCGA-AX-A062, TCGA-2E-A9G8, TCGA-AX-A3FW, TCGA-EY-A210, TCGA-AJ-A2QO, TCGA-BK-A0C9, TCGA-D1-A1O8, TCGA-AX-A3GB, TCGA-AP-A3K1, TCGA-D1-A16X, TCGA-D1-A169, TCGA-AP-A0LT, TCGA-D1-A177, TCGA-B5-A5OD, TCGA-D1-A15X, TCGA-ET-A3DR, TCGA-DJ-A3UT, TCGA-H2-A3RH, TCGA-EM-A1CS, TCGA-J8-A3YH, TCGA-DJ-A2PP, TCGA-EM-A1CV, TCGA-DJ-A3UV, TCGA-GE-A2C6, TCGA-J8-A3YF, TCGA-EL-A3GO, TCGA-EM-A4FU, TCGA-FK-A3SD, TCGA-EL-A3CZ, TCGA-DJ-A3VK, TCGA-DJ-A3VE, TCGA-EM-A22P, TCGA-E8-A417, TCGA-FY-A4B0, TCGA-BJ-A0YZ, TCGA-CE-A484, TCGA-ET-A40P, TCGA-FE-A22Z, TCGA-EL-A3CP, TCGA-EL-A3CW, TCGA-E8-A414, TCGA-CE-A3MD, TCGA-DJ-A1QL, TCGA-EM-A2CN, TCGA-ET-A3BU, TCGA-BJ-A0Z0, TCGA-E8-A415, TCGA-ET-A3BT, TCGA-FY-A3R8, TCGA-EL-A4KG, TCGA-EM-A2OY, TCGA-DE-A0XZ, TCGA-DJ-A3V7, TCGA-DJ-A2PV, TCGA-ET-A3DS, TCGA-DJ-A4UP, TCGA-BJ-A0ZB, TCGA-DJ-A3UY, TCGA-EL-A3D4, TCGA-EM-A4FK, TCGA-FY-A40L, TCGA-DE-A2OL, TCGA-EM-A3AO, TCGA-BJ-A192, TCGA-DJ-A2PY, TCGA-DO-A2HM, TCGA-ET-A39N, TCGA-J8-A3O2, TCGA-EM-A2OX, TCGA-BJ-A45J, TCGA-EM-A3O8, TCGA-ET-A39I, TCGA-EM-A22K, TCGA-FY-A3I5, TCGA-DJ-A3VJ, TCGA-DJ-A2PN, TCGA-EM-A2OW, TCGA-DJ-A3VL, TCGA-ET-A39L, TCGA-DJ-A4UQ, TCGA-E8-A44K, TCGA-DJ-A1QO, TCGA-BJ-A2NA, TCGA-MK-A4N6, TCGA-BJ-A2N8, TCGA-CE-A485, TCGA-EM-A1CU, TCGA-EM-A3FL, TCGA-DJ-A2PT, TCGA-FY-A3ON, TCGA-DE-A4MB, TCGA-ET-A2MZ, TCGA-BJ-A28Z, TCGA-DJ-A2PU, TCGA-EM-A22Q, TCGA-EM-A3AP, TCGA-IM-A4EB, TCGA-EM-A4FH, TCGA-EL-A3GV, TCGA-EM-A3FJ, TCGA-DJ-A1QM, TCGA-DJ-A4UL, TCGA-ET-A39S, TCGA-FY-A3R7, TCGA-DJ-A4UT, TCGA-ET-A3BO, TCGA-DJ-A1QD, TCGA-DE-A7U5, TCGA-FK-A3SE, TCGA-E8-A433, TCGA-DJ-A13P, TCGA-ET-A3BN, TCGA-DJ-A3VG, TCGA-BJ-A45I, TCGA-DJ-A13U, TCGA-BJ-A0Z9, TCGA-EM-A3FP, TCGA-J8-A3YG, TCGA-DJ-A2Q5, TCGA-DJ-A4V2, TCGA-FY-A40K, TCGA-DE-A4M8, TCGA-EM-A3AK, TCGA-DJ-A1QQ, TCGA-FY-A40N, TCGA-BJ-A0ZH, TCGA-DJ-A3VA, TCGA-BJ-A0ZG, TCGA-DJ-A3V5, TCGA-EM-A1YB, TCGA-BJ-A18Y, TCGA-EM-A22O, TCGA-BJ-A28T, TCGA-DO-A1JZ, TCGA-EL-A3CV, TCGA-E3-A3E2, TCGA-DJ-A3VM, TCGA-E8-A418, TCGA-FY-A3YR, TCGA-DE-A0Y2, TCGA-DJ-A2PQ, TCGA-E8-A44M, TCGA-FE-A3PC, TCGA-BJ-A45H, TCGA-MK-A84Z, TCGA-FY-A3BL, TCGA-DJ-A2QC, TCGA-DJ-A2Q0, TCGA-DJ-A3UQ, TCGA-BJ-A4O8, TCGA-EL-A3H4, TCGA-DJ-A13T, TCGA-DO-A1K0, TCGA-EM-A2OV, TCGA-FY-A2QD, TCGA-BJ-A191, TCGA-ET-A25I, TCGA-BJ-A18Z, TCGA-BJ-A45D, TCGA-FE-A230, TCGA-DJ-A4UR, TCGA-FY-A3R6, TCGA-J8-A3YE, TCGA-E8-A413, TCGA-DJ-A3V2, TCGA-FY-A3W9, TCGA-DJ-A1QH, TCGA-EL-A3D5, TCGA-DJ-A2QA, TCGA-DJ-A3UN, TCGA-DJ-A3UM, TCGA-DE-A4MD, TCGA-BJ-A3PT, TCGA-BJ-A45G, TCGA-DJ-A4V4, TCGA-ET-A39J, TCGA-BJ-A3F0, TCGA-BJ-A45K, TCGA-E3-A3E1, TCGA-ET-A3BS, TCGA-EM-A3OA, TCGA-L6-A4EQ, TCGA-EM-A3SU, TCGA-DJ-A3V6, TCGA-BJ-A45E, TCGA-ET-A3DO, TCGA-L6-A4EU, TCGA-DJ-A13O, TCGA-EM-A4FR, TCGA-DE-A4MA, TCGA-IM-A420, TCGA-BJ-A0ZJ, TCGA-J8-A3NZ, TCGA-FE-A231, TCGA-DJ-A2Q2, TCGA-DJ-A13S, TCGA-ET-A4KN, TCGA-EL-A4KI, TCGA-CE-A27D, TCGA-J8-A3O1, TCGA-FY-A3WA, TCGA-EM-A3AL, TCGA-EM-A22N, TCGA-DE-A3KN, TCGA-EM-A2CT, TCGA-EM-A22M, TCGA-FK-A3S3, TCGA-DJ-A1QN, TCGA-DJ-A3UR, TCGA-IM-A3EB, TCGA-E8-A436, TCGA-ET-A25G, TCGA-E8-A419, TCGA-BJ-A2N7, TCGA-DJ-A13R, TCGA-FY-A4B3, TCGA-DJ-A13M, TCGA-FE-A3PD, TCGA-EM-A3AJ, TCGA-ET-A3BP, TCGA-DJ-A1QI, TCGA-DJ-A2PR, TCGA-DJ-A2Q3, TCGA-DJ-A2PW, TCGA-BJ-A3PR, TCGA-EM-A4FN, TCGA-EL-A3H8, TCGA-EM-A4FO, TCGA-ET-A3BQ, TCGA-EL-A3CU, TCGA-ET-A25K, TCGA-E8-A3X7, TCGA-DJ-A2Q1, TCGA-J8-A42S, TCGA-CE-A3ME, TCGA-EM-A3OB, TCGA-ET-A25R, TCGA-DJ-A3VD, TCGA-DJ-A4UW, TCGA-DE-A0Y3, TCGA-DJ-A3UZ, TCGA-FK-A3SG, TCGA-EM-A3O7, TCGA-EL-A3D6, TCGA-BJ-A2P4, TCGA-FE-A237, TCGA-J8-A4HW, TCGA-EL-A3CX, TCGA-DJ-A3V8, TCGA-EL-A3GS, TCGA-EL-A3CM, TCGA-BJ-A28V, TCGA-CE-A13K, TCGA-DJ-A3UW, TCGA-FE-A233, TCGA-FK-A4UB, TCGA-QD-A8IV, TCGA-EL-A3GP, TCGA-EL-A3H5, TCGA-DJ-A1QE, TCGA-ET-A2N0, TCGA-EM-A2CM, TCGA-EL-A3CY, TCGA-DJ-A2PX, TCGA-DJ-A13L, TCGA-DJ-A3US, TCGA-E3-A3E0, TCGA-ET-A40Q, TCGA-EL-A3CS, TCGA-FK-A3SH, TCGA-EM-A3SZ, TCGA-FY-A3TY, TCGA-EL-A3D0, TCGA-DJ-A2QB, TCGA-EL-A3D1, TCGA-DJ-A2Q9, TCGA-EL-A3CO, TCGA-DJ-A3V0, TCGA-BJ-A28S, TCGA-DJ-A2Q4, TCGA-FE-A3PB, TCGA-ET-A2N3, TCGA-EM-A3FO, TCGA-EM-A2CQ, TCGA-EM-A4FQ, TCGA-EM-A2CK, TCGA-ET-A39M, TCGA-DE-A69K, TCGA-DJ-A3V9, TCGA-E8-A242, TCGA-E8-A416, TCGA-EL-A3GX, TCGA-FY-A3R9, TCGA-H2-A3RI, TCGA-FE-A234, TCGA-EM-A2OZ, TCGA-EL-A3GW, TCGA-E3-A3E5, TCGA-E8-A432, TCGA-BJ-A3EZ, TCGA-EM-A1CW, TCGA-E3-A3DY, TCGA-FK-A3SB, TCGA-IM-A41Z, TCGA-EM-A2CR, TCGA-FE-A23A, TCGA-EM-A22J, TCGA-DJ-A3VB, TCGA-EM-A3FK, TCGA-CE-A482, TCGA-DJ-A3V3, TCGA-DJ-A2PZ, TCGA-EM-A4G1, TCGA-ET-A25O, TCGA-DE-A4M9, TCGA-EL-A3GR, TCGA-ET-A3DU, TCGA-EM-A3O6, TCGA-EM-A4FM, TCGA-EM-A4FF, TCGA-ET-A39R, TCGA-E3-A3E3, TCGA-H2-A422, TCGA-ET-A3DQ, TCGA-EL-A3GY, TCGA-EM-A3AN, TCGA-L6-A4ET, TCGA-CE-A481, TCGA-EM-A3FR, TCGA-FE-A3PA, TCGA-DJ-A3VF, TCGA-EM-A4FV, TCGA-FE-A232, TCGA-EM-A3FN, TCGA-DJ-A2PS, TCGA-ET-A40R, TCGA-DJ-A2Q7, TCGA-DJ-A3UO, TCGA-DJ-A3UP, TCGA-BJ-A0ZF, TCGA-EL-A3GQ, TCGA-EM-A1CT, TCGA-EM-A1YA, TCGA-BJ-A45F, TCGA-EL-A3H3, TCGA-CE-A483, TCGA-L6-A4EP, TCGA-EL-A3CT, TCGA-EM-A3FQ, TCGA-DJ-A3V4, TCGA-DJ-A4V5, TCGA-E3-A3DZ, TCGA-BJ-A0ZA, TCGA-BJ-A28R, TCGA-BJ-A3PU, TCGA-DJ-A13W, TCGA-IM-A3U2, TCGA-IM-A3U3, TCGA-DE-A69J, TCGA-BJ-A0ZC, TCGA-IM-A41Y, TCGA-BJ-A290, TCGA-EM-A3AQ, TCGA-EM-A2CL, TCGA-BJ-A2N9, TCGA-DJ-A3UX, TCGA-ET-A3BW, TCGA-EM-A22I, TCGA-EM-A3SX, TCGA-FY-A76V, TCGA-EM-A2P0, TCGA-FE-A236, TCGA-DJ-A4V0, TCGA-FY-A4B4, TCGA-BJ-A0ZE, TCGA-ET-A40S, TCGA-DJ-A1QG, TCGA-DJ-A2PO, TCGA-ET-A3BX, TCGA-EM-A3AI, TCGA-EM-A22L, TCGA-DE-A4MC, TCGA-ET-A39K, TCGA-EM-A2P2, TCGA-MK-A4N7, TCGA-FY-A3NM, TCGA-EL-A3CR, TCGA-DJ-A13V, TCGA-FE-A235, TCGA-ET-A40T, TCGA-DJ-A13X, TCGA-EL-A4KH, TCGA-E8-A2EA, TCGA-BJ-A4O9, TCGA-ET-A3BV, TCGA-FY-A3RA, TCGA-ET-A4KQ, TCGA-DJ-A3UU, TCGA-ET-A39O, TCGA-H2-A2K9, TCGA-EM-A3O3, TCGA-EL-A3CL, TCGA-H2-A421, TCGA-MK-A4N9, TCGA-FY-A3NN, TCGA-BJ-A0Z2, TCGA-4C-A93U, TCGA-ET-A2N4, TCGA-DJ-A2Q6, TCGA-FY-A3I4, TCGA-J8-A3O0, TCGA-EM-A2CU, TCGA-H2-A26U, TCGA-DJ-A3VI, TCGA-IM-A3ED, TCGA-EM-A3AR, TCGA-E8-A434, TCGA-EM-A2CO, TCGA-EM-A2P3, TCGA-FY-A40M, TCGA-ET-A39P, TCGA-EM-A3FM, TCGA-EM-A2CJ, TCGA-EM-A2CP, TCGA-DJ-A1QF, TCGA-EM-A3SY, TCGA-ET-A39T, TCGA-FY-A3NP, TCGA-EL-A3CN, TCGA-BJ-A0Z3, TCGA-DJ-A3UK, TCGA-BJ-A28X, TCGA-J8-A3YD, TCGA-VQ-A91X, TCGA-BR-8676, TCGA-BR-8680, TCGA-CD-5813, TCGA-IP-7968, TCGA-HU-8249, TCGA-D7-A6EX, TCGA-FP-A9TM, TCGA-RD-A8NB, TCGA-BR-8589, TCGA-BR-6710, TCGA-HU-8608, TCGA-CD-A4MG, TCGA-VQ-A91K, TCGA-CD-5801, TCGA-VQ-A91E, TCGA-RD-A7BS, TCGA-BR-8289, TCGA-CD-5799, TCGA-BR-6705, TCGA-CD-5800, TCGA-

BR-8080, TCGA-CD-8531, TCGA-MX-A666, TCGA-BR-8592, TCGA-BR-6801, TCGA-IN-A7NU, TCGA-VQ-A8E3, TCGA-BR-8373, TCGA-HU-8238, TCGA-RD-A8N9, TCGA-HU-A4G8, TCGA-CG-4305, TCGA-FP-A8CX, TCGA-BR-A4CS, TCGA-IN-7806, TCGA-BR-8284, TCGA-CD-8533, TCGA-VQ-A8PH, TCGA-BR-A4J9, TCGA-D7-6524, TCGA-SW-A7EA, TCGA-CG-4469, TCGA-HF-A5NB, TCGA-D7-8579, TCGA-BR-A4J8, TCGA-VQ-AA64, TCGA-VQ-A924, TCGA-HF-7134, TCGA-CD-A486, TCGA-HU-8604, TCGA-CD-5803, TCGA-BR-8487, TCGA-CG-4440, TCGA-R5-A7ZE, TCGA-VQ-A8P5, TCGA-D7-A6EV, TCGA-BR-6564, TCGA-D7-6822, TCGA-R5-A7ZR, TCGA-BR-A44U, TCGA-VQ-A91Y, TCGA-VQ-A922, TCGA-BR-6802, TCGA-RD-A7BT, TCGA-F1-A72C, TCGA-VQ-A91S, TCGA-VQ-A8E0, TCGA-HU-A4H8, TCGA-BR-6455, TCGA-BR-8291, TCGA-CD-A48A, TCGA-R5-A7O7, TCGA-CD-8535, TCGA-VQ-A8PJ, TCGA-VQ-A8PB, TCGA-IN-A7NT, TCGA-IN-A6RS, TCGA-BR-8591, TCGA-HU-8244, TCGA-CD-8532, TCGA-BR-8058, TCGA-BR-8059, TCGA-CG-4462, TCGA-BR-A4J6, TCGA-BR-7716, TCGA-IN-A6RI, TCGA-BR-8371, TCGA-BR-6457, TCGA-D7-A6F2, TCGA-BR-6566, TCGA-CG-4442, TCGA-D7-6818, TCGA-BR-8366, TCGA-D7-6527, TCGA-BR-8381, TCGA-BR-6707, TCGA-HU-A4GH, TCGA-HU-A4GQ, TCGA-F1-6874, TCGA-RD-A8N4, TCGA-KB-A93J, TCGA-HU-A4GU, TCGA-HU-A4H3, TCGA-CD-8530, TCGA-CG-4460, TCGA-HU-A4HB, TCGA-B7-A5TK, TCGA-D7-A748, TCGA-BR-8678, TCGA-VQ-A91D, TCGA-BR-7197, TCGA-R5-A7ZF, TCGA-BR-8367, TCGA-BR-A4IV, TCGA-VQ-A8E2, TCGA-VQ-A8P2, TCGA-VQ-AA6J, TCGA-VQ-A94T, TCGA-R5-A7ZI, TCGA-D7-8575, TCGA-D7-8578, TCGA-VQ-A94O, TCGA-IN-AB1X, TCGA-B7-A5TJ, TCGA-RD-A7C1, TCGA-VQ-A8PD, TCGA-BR-6709, TCGA-BR-7723, TCGA-HU-A4H2, TCGA-BR-7717, TCGA-CG-5719, TCGA-B7-5818, TCGA-VQ-A8PF, TCGA-BR-8690, TCGA-VQ-A8PE, TCGA-VQ-AA6D, TCGA-BR-8485, TCGA-VQ-A8DZ, TCGA-VQ-A91V, TCGA-VQ-A928, TCGA-BR-7958, TCGA-BR-A4CR, TCGA-VQ-AA68, TCGA-HU-A4G9, TCGA-BR-6452, TCGA-D7-6815, TCGA-CG-4465, TCGA-CD-A48C, TCGA-BR-A4J7, TCGA-BR-7707, TCGA-RD-A8N0, TCGA-BR-A4J5, TCGA-D7-6519, TCGA-CD-A4MH, TCGA-VQ-A91U, TCGA-RD-A8N1, TCGA-D7-6525, TCGA-CG-4443, TCGA-D7-8573, TCGA-BR-8590, TCGA-D7-A4YU, TCGA-CD-8527, TCGA-FP-8099, TCGA-HF-7133, TCGA-FP-7829, TCGA-MX-A5UG, TCGA-CG-4436, TCGA-BR-8380, TCGA-HU-A4GJ, TCGA-VQ-A925, TCGA-BR-8483, TCGA-SW-A7EB, TCGA-KB-A6F7, TCGA-CG-4477, TCGA-D7-5577, TCGA-BR-A4PF, TCGA-IN-7808, TCGA-VQ-A8P3, TCGA-BR-8484, TCGA-HJ-7597, TCGA-CG-4444, TCGA-BR-8679, TCGA-VQ-A8PU, TCGA-BR-8686, TCGA-VQ-A8PQ, TCGA-VQ-A94U, TCGA-VQ-A8PX, TCGA-D7-A4Z0, TCGA-MX-A663, TCGA-D7-A747, TCGA-EQ-8122, TCGA-FP-7998, TCGA-HU-A4GY, TCGA-CG-4304, TCGA-BR-8297, TCGA-KB-A93H, TCGA-VQ-A8DV, TCGA-D7-6528, TCGA-HU-A4GP, TCGA-D7-8572, TCGA-FP-8211, TCGA-D7-8576, TCGA-BR-8369, TCGA-CD-8526, TCGA-BR-8372, TCGA-HU-8602, TCGA-BR-8060, TCGA-BR-8295, TCGA-BR-8077, TCGA-HU-A4H6, TCGA-HU-A4GC, TCGA-CD-8534, TCGA-D7-8574, TCGA-IN-AB1V, TCGA-VQ-A91N, TCGA-CD-5798, TCGA-IN-A6RJ, TCGA-BR-6563, TCGA-D7-6522, TCGA-BR-8382, TCGA-D7-A4YX, TCGA-VQ-A927, TCGA-RD-A7BW, TCGA-VQ-A8DU, TCGA-BR-6458, TCGA-CD-8529, TCGA-VQ-A8PO, TCGA-IN-A6RL, TCGA-HU-A4G3, TCGA-BR-8081, TCGA-CD-5804, TCGA-VQ-A91A, TCGA-CG-5718, TCGA-IN-A7NR, TCGA-CG-4301, TCGA-IN-8663, TCGA-VQ-A8P8, TCGA-BR-7959, TCGA-HU-A4G2, TCGA-D7-6520, TCGA-HF-7131, TCGA-BR-8486, TCGA-BR-8286, TCGA-B7-A5TI, TCGA-D7-6521, TCGA-D7-6526, TCGA-HU-8610, TCGA-CG-4437, TCGA-BR-8365, TCGA-CD-8528, TCGA-VQ-A8PM, TCGA-CG-5717, TCGA-VQ-AA6K, TCGA-HU-A4GF, TCGA-VQ-A8PP, TCGA-BR-6565, TCGA-FP-7735, TCGA-RD-A8N5, TCGA-F1-6177, TCGA-IN-A6RO, TCGA-CG-4438, TCGA-CG-5716, TCGA-BR-7851, TCGA-BR-8682, TCGA-CD-A489, TCGA-D7-8570, TCGA-MX-A5UJ, TCGA-ZQ-A9CR, TCGA-3M-AB47, TCGA-R5-A805, TCGA-BR-7722, TCGA-IN-A6RN, TCGA-VQ-A8E7, TCGA-BR-A4J4, TCGA-VQ-A91Z, TCGA-BR-6852, TCGA-CD-A487, TCGA-RD-A8MV, TCGA-ZA-A8F6, TCGA-VQ-AA6G, TCGA-VQ-A94R, TCGA-VQ-A91Q, TCGA-VQ-A923, TCGA-D7-A6EZ, TCGA-VQ-AA6A, TCGA-CG-4306, TCGA-HU-A4HD, TCGA-VQ-A8DT, TCGA-BR-8364, TCGA-VQ-AA6F, TCGA-BR-7957, TCGA-RD-A8MW, TCGA-F1-A448, TCGA-FP-8631, TCGA-FP-A4BF, TCGA-HF-7132, TCGA-CD-8525, TCGA-3M-AB46, TCGA-D7-5578, TCGA-BR-7715, TCGA-D7-A6EY, TCGA-CG-4441, TCGA-BR-8677, TCGA-IN-A6RR, TCGA-CG-4476, TCGA-KB-A93G, TCGA-BR-8368, TCGA-HU-A4GX, TCGA-VQ-A8PK, TCGA-CD-8524, TCGA-BR-7196, TCGA-RD-A8N6, TCGA-VQ-A92D, TCGA-CG-4475, TCGA-FP-7916, TCGA-VQ-A8PC, TCGA-VQ-A94P, TCGA-BR-6803, TCGA-BR-8384, TCGA-RD-A8N2, TCGA-HU-A4GD, TCGA-D7-A74A, TCGA-BR-7704, TCGA-BR-8683, TCGA-CG-4466, TCGA-B7-A5TN, TCGA-D7-A6F0, TCGA-BR-8296, TCGA-VQ-AA69, TCGA-IN-8462, TCGA-BR-6456, TCGA-BR-A4QL, TCGA-BR-8361, TCGA-BR-8588, TCGA-HU-A4GT, TCGA-AG-3583, TCGA-DY-A1DF, TCGA-AG-A025, TCGA-EI-6511, TCGA-AG-A01N, TCGA-AH-6644, TCGA-AG-3890, TCGA-AF-3911, TCGA-F5-6861, TCGA-AG-A00C, TCGA-CI-6622, TCGA-EI-6512, TCGA-EI-6885, TCGA-AG-3902, TCGA-AG-3885, TCGA-EF-5830, TCGA-AG-4008, TCGA-F5-6814, TCGA-DC-6155, TCGA-AG-3896, TCGA-AG-3999, TCGA-AG-3909, TCGA-AG-4007, TCGA-AG-3892, TCGA-AG-A02G, TCGA-AF-A56L, TCGA-EI-6884, TCGA-AG-4001, TCGA-BM-6198, TCGA-CL-5918, TCGA-AG-3726, TCGA-EI-7004, TCGA-AG-A00Y, TCGA-G5-6235, TCGA-F5-6813, TCGA-G5-6233, TCGA-F5-6864, TCGA-AG-A011, TCGA-AG-3584, TCGA-AG-3598, TCGA-AF-2692, TCGA-G5-6641, TCGA-EI-6882, TCGA-AG-4015, TCGA-F5-6571, TCGA-AG-3599, TCGA-AG-3882, TCGA-G5-6572, TCGA-AG-3727, TCGA-AF-6136, TCGA-AF-6655, TCGA-AG-3893, TCGA-AG-3605, TCGA-DY-A1DD, TCGA-AF-4110, TCGA-AG-A01L, TCGA-F5-6464, TCGA-EI-6509, TCGA-EI-6514, TCGA-EI-6881, TCGA-DC-5869, TCGA-AG-4021, TCGA-CI-6621, TCGA-F5-6810, TCGA-AG-3592, TCGA-AG-3887, TCGA-F5-6702, TCGA-EI-6507, TCGA-AG-A015, TCGA-EI-6917, TCGA-DC-6158, TCGA-CI-6619, TCGA-EI-6510, TCGA-AF-2690, TCGA-CI-6624, TCGA-AF-5654, TCGA-DC-4749, TCGA-AG-3580, TCGA-AG-A026, TCGA-EI-7002, TCGA-DC-6681, TCGA-DC-6160, TCGA-AH-6897, TCGA-AG-3901, TCGA-AH-6549, TCGA-AG-A014, TCGA-AG-A02X, TCGA-AG-A032, TCGA-EI-6883, TCGA-DC-6157, TCGA-DC-6682, TCGA-F5-6863, TCGA-DY-A0XA, TCGA-AG-3878, TCGA-AG-A008, TCGA-AG-3574, TCGA-DT-5265, TCGA-AF-3400, TCGA-AG-3594, TCGA-AG-3581, TCGA-AG-3602, TCGA-AG-3894, TCGA-F5-6811, TCGA-AG-4005, TCGA-CI-6623, TCGA-F5-6812, TCGA-AG-A002, TCGA-AF-2687, TCGA-AG-4022, TCGA-DY-A1DC, TCGA-AF-6672, TCGA-AG-A00H, TCGA-DY-A1DG, TCGA-AG-3591, TCGA-DC-6683, TCGA-AG-A016, TCGA-AH-6903, TCGA-AG-A023, TCGA-AF-2691, TCGA-AF-A56K, TCGA-CL-4957, TCGA-CL-5917, TCGA-AH-6544, TCGA-AF-A56N, TCGA-AG-A01J, TCGA-EI-6506, TCGA-DC-6154, TCGA-EI-6508, TCGA-AG-3728, TCGA-AG-3883, TCGA-F5-6465, TCGA-AG-3575, TCGA-AG-3881, TCGA-DC-5337, TCGA-DC-4745, TCGA-DY-A1H8, TCGA-EI-6513, TCGA-AG-3898, TCGA-AF-2693, TCGA-CI-6620, TCGA-J9-A8CK, TCGA-J4-A67R, TCGA-CH-5738, TCGA-EJ-7315, TCGA-YL-A8SC, TCGA-G9-6384, TCGA-V1-A9OT, TCGA-EJ-5530, TCGA-J9-A52C, TCGA-CH-5748, TCGA-CH-5751, TCGA-HC-7209, TCGA-EJ-A6RC, TCGA-G9-6329, TCGA-EJ-A7NM, TCGA-G9-6365, TCGA-J4-A6G1, TCGA-CH-5743, TCGA-EJ-5514, TCGA-FC-A8O0, TCGA-YL-A8SJ, TCGA-XJ-A83G, TCGA-YL-A8HM, TCGA-CH-5788, TCGA-HC-A48F, TCGA-G9-6363, TCGA-HC-7752, TCGA-EJ-A65M, TCGA-EJ-A65D, TCGA-ZG-A9L9, TCGA-XK-AAJP, TCGA-VP-A87D, TCGA-H9-A6BX, TCGA-QU-A6IP, TCGA-V1-A9ZG, TCGA-J4-A83L, TCGA-G9-6338, TCGA-EJ-7314, TCGA-ZG-A9MC, TCGA-HC-7079, TCGA-EJ-5506, TCGA-V1-A9OA, TCGA-ZG-A9LB, TCGA-ZG-A9N3, TCGA-J4-A83N, TCGA-J4-A67Q, TCGA-EJ-5521, TCGA-HC-8259, TCGA-V1-A9OY, TCGA-YL-A8SB, TCGA-EJ-A8FO, TCGA-J4-AATV, TCGA-HI-7169, TCGA-KC-A4BL, TCGA-HC-7081, TCGA-HC-A6AQ, TCGA-M7-A71Z, TCGA-4L-AA1F, TCGA-YL-A8HJ, TCGA-VP-AA1N, TCGA-EJ-AB20, TCGA-EJ-5496, TCGA-EJ-5503, TCGA-EJ-A8FP, TCGA-V1-A9O7, TCGA-G9-7510, TCGA-EJ-8470, TCGA-EJ-5542, TCGA-J4-8200, TCGA-HC-8264, TCGA-KC-A7F5, TCGA-FC-A66V, TCGA-EJ-A65G, TCGA-G9-6342, TCGA-XJ-A9DI, TCGA-WW-A8ZI, TCGA-HC-A6AO, TCGA-HC-7820, TCGA-CH-5792, TCGA-2A-AAYF, TCGA-J4-A67M, TCGA-YL-A9WH, TCGA-EJ-A8FU, TCGA-Y6-A8TL, TCGA-V1-A9Z7, TCGA-V1-A9OF, TCGA-G9-6373, TCGA-EJ-A46I, TCGA-HC-A6AS, TCGA-J4-AAU2, TCGA-ZG-A9LU, TCGA-EJ-A65B, TCGA-YL-A8SK, TCGA-M7-A71Y, TCGA-XJ-A9DX, TCGA-J4-A6G3, TCGA-ZG-A8QW, TCGA-QU-A6IO, TCGA-EJ-A46F, TCGA-VN-A88M, TCGA-EJ-7783, TCGA-ZG-A9L2, TCGA-ZG-A9L6, TCGA-EJ-A7NF, TCGA-V1-A9OL, TCGA-G9-6369, TCGA-VP-A876, TCGA-VN-A88Q, TCGA-HC-7821, TCGA-QU-A6IN, TCGA-VP-A875, TCGA-EJ-A65E, TCGA-CH-5752, TCGA-G9-6385, TCGA-EJ-8474, TCGA-HC-8258, TCGA-VN-A88R, TCGA-KC-A7F6, TCGA-HC-7738, TCGA-EJ-7794, TCGA-EJ-7317, TCGA-EJ-5526, TCGA-ZG-A9KY, TCGA-QU-A6IL, TCGA-G9-6333, TCGA-EJ-7792, TCGA-J4-A67S, TCGA-V1-A9ZR, TCGA-KC-A7FD, TCGA-EJ-A65F, TCGA-G9-A9S4, TCGA-EJ-A46E, TCGA-J4-A67N, TCGA-G9-6371, TCGA-CH-5739, TCGA-CH-5737, TCGA-M7-A724, TCGA-ZG-A9NI, TCGA-V1-A8WS, TCGA-VN-A88O, TCGA-HC-A6HX, TCGA-VN-A88N, TCGA-ZG-A9M4, TCGA-G9-7525, TCGA-G9-6366, TCGA-YL-A8SO, TCGA-XK-AAJA, TCGA-EJ-7318, TCGA-XK-AAJT, TCGA-J9-A52B, TCGA-FC-A4JI, TCGA-HC-7748, TCGA-J9-A8CM, TCGA-EJ-5498, TCGA-EJ-5511, TCGA-G9-7523, TCGA-ZG-A8QZ, TCGA-VN-A88K, TCGA-EJ-5502, TCGA-HC-8262, TCGA-M7-A720, TCGA-CH-5789, TCGA-EJ-A7NH, TCGA-TP-A8TV, TCGA-H9-A6BY, TCGA-VP-A87B, TCGA-ZG-A9LS, TCGA-CH-5754, TCGA-ZG-A8QX, TCGA-EJ-5494, TCGA-YL-A8SA, TCGA-VN-A943, TCGA-HC-7230, TCGA-EJ-5532, TCGA-EJ-7781, TCGA-G9-7519, TCGA-EJ-A6RA, TCGA-EJ-5525, TCGA-HC-A9TE, TCGA-G9-6364, TCGA-HC-7213, TCGA-HC-7750, TCGA-XK-AAJU, TCGA-G9-6499, TCGA-VN-A88L, TCGA-EJ-5522, TCGA-HI-7171, TCGA-HC-7077, TCGA-J9-A8CN, TCGA-EJ-7125, TCGA-MG-AAMC, TCGA-2A-A8VL, TCGA-XK-AAJ3, TCGA-HC-7742, TCGA-HC-8216, TCGA-J9-A8CL, TCGA-HC-8213, TCGA-J4-A6M7, TCGA-2A-A8VT, TCGA-G9-6353, TCGA-TP-A8TT, TCGA-EJ-AB27, TCGA-YJ-A8SW, TCGA-HI-7168, TCGA-HC-7740, TCGA-EJ-5508, TCGA-V1-A9OX, TCGA-J4-A83M, TCGA-G9-6332, TCGA-J4-A83J, TCGA-KC-A7FE, TCGA-HC-A9TH, TCGA-CH-5790, TCGA-V1-A9OH, TCGA-EJ-5531, TCGA-EJ-7328, TCGA-XK-AAIV, TCGA-2A-A8VV, TCGA-VP-A87H, TCGA-M7-A725, TCGA-CH-5753, TCGA-V1-A9ZI, TCGA-G9-6347, TCGA-VN-A88P, TCGA-EJ-8469, TCGA-X4-A8KQ, TCGA-EJ-A7NG, TCGA-EJ-7312, TCGA-VP-A87J, TCGA-EJ-5497, TCGA-V1-A8MG, TCGA-V1-A8MM, TCGA-EJ-A7NJ, TCGA-HC-A76X, TCGA-V1-A8WV, TCGA-V1-A9Z9, TCGA-EJ-5518, TCGA-EJ-5524, TCGA-M7-A722, TCGA-HI-7170, TCGA-ZG-A8QY, TCGA-G9-A9S7, TCGA-J4-A67O, TCGA-HC-A631, TCGA-J4-A83K, TCGA-EJ-5499, TCGA-H9-7775, TCGA-M7-A721, TCGA-HC-8256, TCGA-EJ-5512, TCGA-G9-6339, TCGA-ZG-A9L1, TCGA-EJ-5505, TCGA-EJ-A65J, TCGA-J4-A67T, TCGA-EJ-7115, TCGA-VP-A87E, TCGA-XJ-A9DK, TCGA-HC-7212, TCGA-HC-7745, TCGA-EJ-A46B, TCGA-G9-6367, TCGA-EJ-5517, TCGA-V1-A8X3, TCGA-EJ-7330, TCGA-2A-A8VO, TCGA-G9-6351, TCGA-HC-A8D0, TCGA-G9-6348, TCGA-HC-7232, TCGA-V1-A8ML, TCGA-G9-7521, TCGA-EJ-5527, TCGA-KC-A7FA, TCGA-G9-6498, TCGA-G9-6361, TCGA-V1-A8MF, TCGA-TK-A8OK, TCGA-EJ-8468, TCGA-ZG-A9L5, TCGA-XK-AAIW, TCGA-V1-A8WW, TCGA-HC-7818, TCGA-EJ-A7NK, TCGA-G9-A9S0, TCGA-HC-8265, TCGA-V1-A8MU, TCGA-2A-AAYO, TCGA-EJ-7788, TCGA-XJ-A9DQ, TCGA-EJ-7331, TCGA-EJ-5504, TCGA-G9-6496, TCGA-VP-A879, TCGA-HC-A632, TCGA-EJ-A8FN, TCGA-YL-A8HK, TCGA-V1-A9ZK, TCGA-XQ-A8TA, TCGA-EJ-7123, TCGA-EJ-5515, TCGA-ZG-A9L4, TCGA-XK-AAJR, TCGA-HC-7080, TCGA-FC-7961, TCGA-CH-5750, TCGA-YL-A8SH, TCGA-YL-A8SR, TCGA-VP-A87K, TCGA-HC-A8CY, TCGA-EJ-7784, TCGA-HC-8260, TCGA-G9-6343, TCGA-EJ-7325, TCGA-EJ-7782, TCGA-XK-AAK1, TCGA-EJ-5510, TCGA-CH-5740, TCGA-HC-7233, TCGA-J9-A8CP, TCGA-V1-A8WN, TCGA-G9-6356, TCGA-EJ-7218, TCGA-EJ-7785, TCGA-G9-6336, TCGA-QU-A6IM, TCGA-HC-A6AL, TCGA-CH-5794, TCGA-YL-A8S8, TCGA-HC-7817, TCGA-CH-5744, TCGA-HC-8257, TCGA-HC-7210, TCGA-G9-6377, TCGA-G9-7522, TCGA-HC-8266, TCGA-V1-A9O9, TCGA-YL-A8S9, TCGA-HC-A6HY, TCGA-KC-A4BN, TCGA-HC-7744, TCGA-EJ-5507, TCGA-VP-A872, TCGA-2A-A8W3, TCGA-CH-5746, TCGA-2A-A8W1, TCGA-HC-A6AN, TCGA-XJ-A83H, TCGA-XQ-A8TB, TCGA-G9-6354, TCGA-V1-A9OQ, TCGA-EJ-A46D, TCGA-CH-5741, TCGA-EJ-7321, TCGA-G9-6494, TCGA-G9-6362, TCGA-XK-AAIR, TCGA-ZG-A9LM, TCGA-HC-7736, TCGA-M7-A723, TCGA-HC-A76W, TCGA-VN-A88I, TCGA-YL-A9WI, TCGA-HC-7231, TCGA-V1-A9Z8, TCGA-YL-A8SP, TCGA-XA-A8JR, TCGA-HC-7078, TCGA-FC-7708, TCGA-G9-7509, TCGA-J4-AATZ, TCGA-HC-A4ZV, TCGA-HC-A8D1, TCGA-2A-AAYU, TCGA-ZG-A9LN, TCGA-YL-A8SL, TCGA-G9-6379, TCGA-EJ-5509, TCGA-EJ-5495, TCGA-YL-A8HL, TCGA-ZG-A9LY, TCGA-HC-A6AP, TCGA-J4-8198, TCGA-ZG-A9L0, TCGA-EJ-A46G, TCGA-V1-A8MK, TCGA-HC-7747, TCGA-X4-A8KS, TCGA-ZG-A9ND, TCGA-YL-A9WX, TCGA-EJ-7791, TCGA-EJ-A46H, TCGA-EJ-7786, TCGA-CH-5791, TCGA-EJ-5519, TCGA-EJ-7793, TCGA-YL-A9WY, TCGA-ZG-A9LZ, TCGA-J9-A52E, TCGA-G9-6370, TCGA-EJ-8472, TCGA-YL-A9WJ, TCGA-YL-A8SI, TCGA-VP-A878, TCGA-Y6-A9XI, TCGA-J4-A83I, TCGA-V1-A9O5, TCGA-FC-A5OB, TCGA-EJ-7789, TCGA-EJ-7797, TCGA-2A-A8VX, TCGA-G9-6378, TCGA-YL-A8SQ, TCGA-HC-7819, TCGA-V1-A8WL, TCGA-SU-A7E7, TCGA-KC-A7F3, TCGA-EJ-5516, TCGA-FC-A6HD, TCGA-HC-8261, TCGA-VP-A87C, TCGA-KC-A4BV, TCGA-KC-A4BR, TCGA-HC-7075, TCGA-EJ-7327, TCGA-V1-A8MJ, TCGA-J4-A67L, TCGA-CH-5745, TCGA-XJ-A83F, TCGA-J4-A67K, TCGA-YL-A9WL, TCGA-EJ-A8FS, TCGA-EJ-5501, TCGA-62-A46O, TCGA-J2-A4AE, TCGA-55-8511, TCGA-69-7764, TCGA-93-A4JN, TCGA-05-4425, TCGA-97-A4M1, TCGA-05-4432, TCGA-44-5644, TCGA-62-A46S, TCGA-MP-A4TI, TCGA-55-8090, TCGA-67-6215, TCGA-55-7911, TCGA-44-2666, TCGA-64-1676, TCGA-L9-A5IP, TCGA-97-7546, TCGA-55-7283, TCGA-44-6777, TCGA-05-5715, TCGA-62-A472, TCGA-50-8459, TCGA-97-A4M3, TCGA-55-8299, TCGA-75-5146, TCGA-62-8402, TCGA-55-8205, TCGA-44-3918, TCGA-05-4426, TCGA-78-7633, TCGA-53-7813, TCGA-55-8085, TCGA-35-5375, TCGA-95-8039, TCGA-L9-A50W, TCGA-NJ-A4YF, TCGA-44-2656, TCGA-44-7660, TCGA-97-8174, TCGA-64-5775, TCGA-86-8673, TCGA-69-A59K, TCGA-MP-A4T8, TCGA-62-A46Y, TCGA-44-6775, TCGA-80-5611, TCGA-55-6543, TCGA-64-5774, TCGA-44-2662, TCGA-55-A57B, TCGA-MP-A5C7, TCGA-44-6148, TCGA-78-8662, TCGA-78-7146, TCGA-38-A44F, TCGA-44-A47B, TCGA-MN-A4N4, TCGA-05-4403, TCGA-MP-A4TH, TCGA-50-5068, TCGA-86-8280, TCGA-78-7156, TCGA-44-5645, TCGA-67-4679, TCGA-73-A9RS, TCGA-83-5908, TCGA-93-A4JO, TCGA-55-8094, TCGA-62-A46R, TCGA-35-3615, TCGA-55-7994, TCGA-67-3771, TCGA-44-6147, TCGA-44-6146, TCGA-64-1677, TCGA-97-7554, TCGA-55-6712, TCGA-64-1680, TCGA-50-5051, TCGA-64-5815, TCGA-44-2659, TCGA-69-7763, TCGA-55-8514, TCGA-55-8208, TCGA-05-4424, TCGA-44-7662, TCGA-50-5066, TCGA-91-8497, TCGA-44-A47A, TCGA-44-6778, TCGA-44-7671, TCGA-97-8177, TCGA-86-8073, TCGA-55-7227, TCGA-78-7150, TCGA-44-2665, TCGA-44-5643, TCGA-55-7913, TCGA-86-8668, TCGA-78-7154, TCGA-55-7725, TCGA-50-8460, TCGA-73-7498, TCGA-J2-A4AD, TCGA-53-A4EZ, TCGA-44-3396, TCGA-91-6840, TCGA-44-8117, TCGA-44-3919, TCGA-86-8281, TCGA-55-8203, TCGA-44-4112, TCGA-78-7155, TCGA-95-7043, TCGA-MP-A4T6, TCGA-MP-A4SY, TCGA-91-A4BC, TCGA-55-7570, TCGA-35-4123, TCGA-95-7562, TCGA-95-7567, TCGA-97-A4M5, TCGA-69-7765, TCGA-J2-8192, TCGA-MP-A4TF, TCGA-75-5125, TCGA-69-8254, TCGA-44-2668, TCGA-78-7152, TCGA-55-8302, TCGA-05-4398, TCGA-44-A4SS, TCGA-78-8648, TCGA-50-5055, TCGA-67-3772, TCGA-97-A4M7, TCGA-78-7158, TCGA-05-5428, TCGA-NJ-A7XG, TCGA-86-6562, TCGA-55-7576, TCGA-05-4422, TCGA-55-8620, TCGA-55-5899, TCGA-69-8255, TCGA-75-6203, TCGA-95-A4VP, TCGA-L4-

A4E6, TCGA-97-8179, TCGA-44-7669, TCGA-05-4420, TCGA-86-A4P7, TCGA-86-8674, TCGA-55-8616, TCGA-95-A4VN, TCGA-55-A490, TCGA-86-8278, TCGA-86-7954, TCGA-91-6828, TCGA-62-8399, TCGA-50-5045, TCGA-97-7938, TCGA-95-7947, TCGA-44-2655, TCGA-64-5778, TCGA-55-7726, TCGA-91-A4BD, TCGA-55-8096, TCGA-50-5072, TCGA-55-7724, TCGA-64-1681, TCGA-67-6217, TCGA-NJ-A55R, TCGA-62-8397, TCGA-97-8175, TCGA-L9-A443, TCGA-86-8359, TCGA-05-4384, TCGA-97-8547, TCGA-44-3917, TCGA-NJ-A55O, TCGA-55-8092, TCGA-05-5425, TCGA-62-8398, TCGA-55-8506, TCGA-55-A492, TCGA-95-7039, TCGA-75-6206, TCGA-62-A46V, TCGA-86-8585, TCGA-67-3774, TCGA-78-7220, TCGA-53-7624, TCGA-67-3770, TCGA-86-8056, TCGA-44-A4SU, TCGA-93-7347, TCGA-75-7025, TCGA-78-8660, TCGA-05-4415, TCGA-44-7667, TCGA-44-2661, TCGA-80-5607, TCGA-55-A4DG, TCGA-L9-A7SV, TCGA-95-7948, TCGA-55-7574, TCGA-91-8499, TCGA-05-5423, TCGA-55-8619, TCGA-62-8394, TCGA-O1-A52J, TCGA-05-4389, TCGA-93-7348, TCGA-97-A4M2, TCGA-67-6216, TCGA-MP-A4TK, TCGA-99-8033, TCGA-05-4402, TCGA-75-6211, TCGA-35-4122, TCGA-L9-A8F4, TCGA-05-4249, TCGA-64-1678, TCGA-55-8097, TCGA-97-A4M0, TCGA-97-7937, TCGA-86-8279, TCGA-86-8358, TCGA-44-7659, TCGA-95-A4VK, TCGA-MP-A4TD, TCGA-NJ-A55A, TCGA-55-A48Y, TCGA-97-A4LX, TCGA-05-4417, TCGA-55-8614, TCGA-44-7670, TCGA-55-8091, TCGA-55-8507, TCGA-50-5941, TCGA-73-7499, TCGA-55-8204, TCGA-86-7713, TCGA-05-4427, TCGA-44-3398, TCGA-86-A4P8, TCGA-75-6214, TCGA-97-A4M6, TCGA-86-7953, TCGA-67-3773, TCGA-97-7941, TCGA-55-7727, TCGA-78-7536, TCGA-86-8074, TCGA-J2-A4AG, TCGA-62-A471, TCGA-MN-A4N5, TCGA-05-4430, TCGA-93-A4JP, TCGA-71-6725, TCGA-44-8119, TCGA-75-6207, TCGA-64-1679, TCGA-NJ-A4YP, TCGA-55-A491, TCGA-86-8055, TCGA-50-5049, TCGA-86-8076, TCGA-05-5429, TCGA-NJ-A4YQ, TCGA-55-8089, TCGA-99-AA5R, TCGA-05-4382, TCGA-55-8206, TCGA-55-7281, TCGA-44-7661, TCGA-55-7903, TCGA-97-7552, TCGA-78-7539, TCGA-05-4397, TCGA-05-4390, TCGA-86-8669, TCGA-78-7145, TCGA-55-A48Z, TCGA-86-8075, TCGA-44-8120, TCGA-75-7030, TCGA-55-8621, TCGA-NJ-A4YI, TCGA-75-5122, TCGA-05-4410, TCGA-05-4405, TCGA-93-8067, TCGA-55-8512, TCGA-L4-A4E5, TCGA-75-6212, TCGA-55-8087, TCGA-69-7980, TCGA-05-4434, TCGA-MP-A4TJ, TCGA-78-7160, TCGA-86-8671, TCGA-71-8520, TCGA-99-8032, TCGA-86-7711, TCGA-64-5781, TCGA-95-7944, TCGA-55-A494, TCGA-97-8171, TCGA-55-8505, TCGA-86-A4D0, TCGA-44-A479, TCGA-55-8615, TCGA-86-8672, TCGA-55-7995, TCGA-J2-8194, TCGA-50-5044, TCGA-55-7914, TCGA-55-A48X, TCGA-78-7535, TCGA-75-7031, TCGA-97-8552, TCGA-44-6776, TCGA-L9-A444, TCGA-L9-A743, TCGA-55-A493, TCGA-4B-A93V, TCGA-75-5147, TCGA-55-7728, TCGA-53-7626, TCGA-05-4250, TCGA-69-7974, TCGA-69-7978, TCGA-86-8054, TCGA-50-5944, TCGA-05-4395, TCGA-86-6851, TCGA-78-7161, TCGA-50-5942, TCGA-MP-A4T7, TCGA-55-8207, TCGA-44-6145, TCGA-99-7458, TCGA-97-7547, TCGA-55-7815, TCGA-62-A470, TCGA-78-8655, TCGA-80-5608, TCGA-05-4244, TCGA-S2-AA1A, TCGA-MP-A4SV, TCGA-05-4433, TCGA-97-8176, TCGA-95-8494, TCGA-50-8457, TCGA-MP-A4T4, TCGA-86-7714, TCGA-91-8496, TCGA-55-A4DF, TCGA-44-2657, TCGA-69-7760, TCGA-97-7553, TCGA-44-A47G, TCGA-99-8025, TCGA-86-7955, TCGA-78-7148, TCGA-75-6205, TCGA-91-7771, TCGA-50-6597, TCGA-55-8508, TCGA-55-8301, TCGA-05-4396, TCGA-38-6178, TCGA-93-A4JQ, TCGA-MN-A4N1, TCGA-86-A4JF, TCGA-44-6774, TCGA-MP-A4TE, TCGA-55-7284, TCGA-55-7907, TCGA-55-8510, TCGA-62-A46P, TCGA-99-8028, TCGA-NJ-A4YG, TCGA-69-7761, TCGA-44-7672, TCGA-05-4418, TCGA-97-8172, TCGA-75-5126, TCGA-62-8395, TCGA-78-7147, TCGA-55-8513, TCGA-86-A456, TCGA-78-7537, TCGA-64-5779, TCGA-86-7701, TCGA-69-8253, TCGA-78-7153, TCGA-MP-A4SW, TCGA-78-7149, TCGA-69-7973, TCGA-ZP-A9CZ, TCGA-ZP-A9D2, TCGA-CC-A3M9, TCGA-GJ-A6C0, TCGA-CC-A123, TCGA-LG-A9QC, TCGA-DD-AAE2, TCGA-RC-A7SH, TCGA-DD-A1ED, TCGA-CC-A8HU, TCGA-XR-A8TG, TCGA-DD-AACA, TCGA-2Y-A9H9, TCGA-DD-A4NH, TCGA-EP-A12J, TCGA-DD-AADD, TCGA-DD-A4NP, TCGA-DD-AADQ, TCGA-G3-A7M6, TCGA-DD-AAEA, TCGA-RG-A7D4, TCGA-ED-A7PX, TCGA-FV-A3R3, TCGA-GJ-A3OU, TCGA-ED-A459, TCGA-HP-A5MZ, TCGA-DD-AAD6, TCGA-5C-A9VG, TCGA-DD-AAEH, TCGA-DD-A73A, TCGA-DD-AADG, TCGA-DD-A4NQ, TCGA-2V-A95S, TCGA-CC-A5UE, TCGA-2Y-A9GZ, TCGA-K7-AAU7, TCGA-2Y-A9H7, TCGA-G3-A5SK, TCGA-G3-AAV4, TCGA-DD-A1EA, TCGA-XR-A8TD, TCGA-DD-AACO, TCGA-DD-AAD2, TCGA-YA-A8S7, TCGA-G3-A25T, TCGA-G3-A7M8, TCGA-CC-A3MC, TCGA-G3-A5SL, TCGA-CC-5260, TCGA-MI-A75E, TCGA-DD-AAD3, TCGA-UB-AA0U, TCGA-DD-AACI, TCGA-FV-A3I1, TCGA-DD-AAD5, TCGA-DD-AAC9, TCGA-DD-AACF, TCGA-G3-A25S, TCGA-K7-A6G5, TCGA-DD-AA3A, TCGA-DD-AACG, TCGA-FV-A4ZQ, TCGA-DD-AADJ, TCGA-RC-A7SB, TCGA-G3-A3CI, TCGA-G3-AAV0, TCGA-DD-AAEK, TCGA-DD-AAVU, TCGA-DD-AADC, TCGA-DD-A1EE, TCGA-ED-A82E, TCGA-G3-AAV6, TCGA-2Y-A9GW, TCGA-CC-A5UD, TCGA-DD-AADK, TCGA-G3-A3CK, TCGA-DD-AAW3, TCGA-NI-A8LF, TCGA-DD-A73G, TCGA-CC-5261, TCGA-5R-AAAM, TCGA-DD-A4NG, TCGA-DD-AAEG, TCGA-G3-A25V, TCGA-CC-A9FU, TCGA-DD-AADA, TCGA-RC-A6M5, TCGA-DD-A73F, TCGA-FV-A2QQ, TCGA-BC-A8YO, TCGA-DD-A73D, TCGA-BW-A5NQ, TCGA-WX-AA47, TCGA-XR-A8TE, TCGA-G3-A6UC, TCGA-DD-AAW2, TCGA-DD-A11A, TCGA-DD-A1EL, TCGA-G3-A7M9, TCGA-EP-A2KC, TCGA-DD-A115, TCGA-ED-A8O5, TCGA-DD-AADB, TCGA-2Y-A9GX, TCGA-CC-5258, TCGA-LG-A9QD, TCGA-DD-A73B, TCGA-UB-A7MD, TCGA-ZP-A9CV, TCGA-RC-A6M6, TCGA-G3-AAV7, TCGA-DD-AAED, TCGA-DD-AACK, TCGA-QA-A7B7, TCGA-DD-AADR, TCGA-EP-A3JL, TCGA-DD-A4NS, TCGA-UB-A7ME, TCGA-DD-AADL, TCGA-UB-A7MC, TCGA-DD-A4NK, TCGA-DD-A1EF, TCGA-G3-AAUZ, TCGA-DD-AAVY, TCGA-G3-A7M5, TCGA-ZP-A9CY, TCGA-DD-AAD8, TCGA-UB-A7MF, TCGA-MR-A8JO, TCGA-ED-A7PZ, TCGA-K7-A5RG, TCGA-BW-A5NO, TCGA-DD-AAEE, TCGA-GJ-A9DB, TCGA-CC-A7IG, TCGA-DD-AACJ, TCGA-ZS-A9CD, TCGA-MI-A75G, TCGA-DD-AAVX, TCGA-WJ-A86L, TCGA-WX-AA46, TCGA-EP-A3RK, TCGA-DD-A11B, TCGA-DD-AAW1, TCGA-DD-AADY, TCGA-MI-A75H, TCGA-RC-A6M4, TCGA-BC-A69H, TCGA-G3-AAV1, TCGA-DD-A4NV, TCGA-KR-A7K0, TCGA-CC-A7IJ, TCGA-DD-AAVW, TCGA-UB-A7MB, TCGA-ZP-A9D4, TCGA-DD-AAC8, TCGA-2Y-A9H3, TCGA-CC-A7IK, TCGA-DD-AADW, TCGA-FV-A495, TCGA-G3-A5SJ, TCGA-CC-A9FW, TCGA-G3-A25Y, TCGA-ED-A627, TCGA-DD-A113, TCGA-DD-AAVZ, TCGA-CC-A7IL, TCGA-DD-A116, TCGA-2Y-A9H2, TCGA-LG-A6GG, TCGA-G3-A3CG, TCGA-DD-AAEI, TCGA-EP-A2KA, TCGA-KR-A7K2, TCGA-PD-A5DF, TCGA-2Y-A9H1, TCGA-2Y-A9H6, TCGA-BC-A217, TCGA-DD-AACQ, TCGA-DD-A119, TCGA-ZS-A9CE, TCGA-DD-AAVS, TCGA-ED-A97K, TCGA-HP-A5N0, TCGA-DD-AACB, TCGA-WQ-A9G7, TCGA-DD-AACH, TCGA-FV-A496, TCGA-5R-AA1C, TCGA-ZS-A9CG, TCGA-BC-A5W4, TCGA-DD-AACC, TCGA-MI-A75I, TCGA-2Y-A9HB, TCGA-CC-A9FS, TCGA-DD-AAE1, TCGA-KR-A7K7, TCGA-RC-A7S9, TCGA-DD-AACE, TCGA-CC-A3MB, TCGA-CC-A7IF, TCGA-BC-A3KF, TCGA-DD-AAE6, TCGA-2Y-A9GS, TCGA-ED-A66X, TCGA-DD-AAW0, TCGA-DD-A4NR, TCGA-CC-A7II, TCGA-DD-AACL, TCGA-ZP-A9D0, TCGA-DD-AADI, TCGA-ED-A66Y, TCGA-DD-AADM, TCGA-G3-A25U, TCGA-CC-5262, TCGA-WX-AA44, TCGA-DD-AACN, TCGA-DD-AAE9, TCGA-RC-A6M3, TCGA-CC-A3MA, TCGA-DD-AACY, TCGA-2Y-A9H8, TCGA-MR-A520, TCGA-ED-A4XI, TCGA-KR-A7K8, TCGA-T1-A6J8, TCGA-ED-A7XO, TCGA-G3-A5SI, TCGA-DD-AAE4, TCGA-ED-A7XP, TCGA-ED-A8O6, TCGA-DD-AAE3, TCGA-DD-A4NF, TCGA-BC-4073, TCGA-CC-A9FV, TCGA-ZS-A9CF, TCGA-DD-AAE7, TCGA-DD-A1EJ, TCGA-5C-A9VH, TCGA-2Y-A9H4, TCGA-2Y-A9H5, TCGA-DD-A4NJ, TCGA-UB-A7MA, TCGA-DD-A4NO, TCGA-DD-A73C, TCGA-DD-A1EG, TCGA-G3-AAV5, TCGA-DD-AACT, TCGA-DD-AADU, TCGA-DD-AADV, TCGA-DD-A1EK, TCGA-DD-AAVP, TCGA-CC-A8HT, TCGA-DD-AAEB, TCGA-RC-A7SF, TCGA-RC-A7SK, TCGA-2Y-A9GT, TCGA-DD-AAVQ, TCGA-ZP-A9D1, TCGA-NI-A4U2, TCGA-5R-AA1D, TCGA-DD-AADS, TCGA-DD-AACD, TCGA-G3-AAV3, TCGA-MI-A75C, TCGA-CC-A7IE, TCGA-DD-AADF, TCGA-DD-A1EB, TCGA-DD-AACZ, TCGA-2Y-A9GY, TCGA-CC-5264, TCGA-G3-A3CJ, TCGA-2Y-A9GU, TCGA-XR-A8TF, TCGA-DD-A4NN, TCGA-DD-A1EH, TCGA-G3-A5SM, TCGA-EP-A2KB, TCGA-2Y-A9HA, TCGA-K7-A5RF, TCGA-WQ-AB4B, TCGA-2Y-A9H0, TCGA-CC-5263, TCGA-DD-AAD0, TCGA-UB-AA0V, TCGA-EP-A26S, TCGA-DD-A73E, TCGA-DD-AACS, TCGA-3K-AAZ8, TCGA-DD-AADN, TCGA-G3-A7M7, TCGA-DD-AADO, TCGA-CC-5259, TCGA-DD-AAD1, TCGA-CC-A1HT, TCGA-DD-AAVV, TCGA-CC-A7IH, TCGA-CC-A5UC, TCGA-DD-AACP, TCGA-BW-A5NP, TCGA-DD-AACV, TCGA-DD-A114, TCGA-ED-A5KG, TCGA-ED-A7PY, TCGA-CC-A8HV, TCGA-2Y-A9GV, TCGA-O8-A75V, TCGA-DD-A4NI, TCGA-DD-AACU, TCGA-G3-A25Z, TCGA-DD-AADP, TCGA-DD-AACW, TCGA-DD-AAE0, TCGA-G3-AAV2, TCGA-5C-AAPD, TCGA-4R-AA8I, TCGA-BC-A69I, TCGA-BC-A3KG, TCGA-DD-AACX, TCGA-DD-AAVR, TCGA-GL-A9DC, TCGA-SX-A71U, TCGA-AL-3468, TCGA-B3-4103, TCGA-B1-5398, TCGA-G7-A8LD, TCGA-A4-7UZ, TCGA-A4-A6HP, TCGA-A4-8517, TCGA-SX-A71V, TCGA-DW-5561, TCGA-DZ-6133, TCGA-5P-A9K2, TCGA-Y8-A8S0, TCGA-G7-A4TM, TCGA-2Z-A9JI, TCGA-2Z-A9JK, TCGA-SX-A7SU, TCGA-IA-A40U, TCGA-A4-7734, TCGA-PJ-A5Z8, TCGA-G7-6792, TCGA-KV-A6GD, TCGA-G7-7502, TCGA-B1-A654, TCGA-UZ-A9PM, TCGA-B9-4115, TCGA-HE-A5NH, TCGA-GL-7966, TCGA-UZ-A9PL, TCGA-UZ-A9PQ, TCGA-B1-A47O, TCGA-Y8-A8S1, TCGA-2Z-A9J2, TCGA-B9-5156, TCGA-IA-A40X, TCGA-G7-6789, TCGA-5P-A9KH, TCGA-5P-A9KA, TCGA-4A-A93Y, TCGA-SX-A71S, TCGA-A4-7996, TCGA-UZ-A9PO, TCGA-5P-A9K4, TCGA-HE-A5NI, TCGA-5P-A9JZ, TCGA-B1-A655, TCGA-Y8-A8RZ, TCGA-AL-3466, TCGA-UZ-A9Q0, TCGA-O9-A75Z, TCGA-SX-A71R, TCGA-B3-A6W5, TCGA-B1-7332, TCGA-A4-7286, TCGA-2Z-A9JM, TCGA-UN-AAZ9, TCGA-HE-A5NK, TCGA-UZ-A9PV, TCGA-B1-A47N, TCGA-F9-A8NY, TCGA-DW-7842, TCGA-B9-A8YH, TCGA-HE-7130, TCGA-2Z-A9J5, TCGA-A4-7583, TCGA-5P-A9JW, TCGA-A4-8312, TCGA-SX-A7SS, TCGA-2Z-A9JR, TCGA-4A-A93X, TCGA-B9-A44B, TCGA-A4-7997, TCGA-IA-A40Y, TCGA-A4-A4ZT, TCGA-B9-5155, TCGA-Y8-A8RY, TCGA-5P-A9JU, TCGA-SX-A7SN, TCGA-MH-A560, TCGA-5P-A9JV, TCGA-B9-4117, TCGA-UZ-A9Q1, TCGA-5P-A9K8, TCGA-A4-A5XZ, TCGA-J7-8537, TCGA-AL-3473, TCGA-A4-7915, TCGA-GL-7773, TCGA-PJ-A5Z9, TCGA-B3-8121, TCGA-AL-3472, TCGA-B9-A5W7, TCGA-A4-8630, TCGA-A4-8310, TCGA-J7-6720, TCGA-UZ-A9PR, TCGA-A4-8518, TCGA-GL-8500, TCGA-AL-7173, TCGA-HE-A5NJ, TCGA-2Z-A9JQ, TCGA-GL-A9DD, TCGA-G7-6795, TCGA-EV-5903, TCGA-AL-3467, TCGA-GL-6846, TCGA-WN-AB4C, TCGA-B9-A5W9, TCGA-4A-A93W, TCGA-5P-A9K9, TCGA-A4-8515, TCGA-UZ-A9PZ, TCGA-GL-A59R, TCGA-B9-A5W8, TCGA-HE-7128, TCGA-2Z-A9J1, TCGA-IZ-8196, TCGA-2Z-A9J8, TCGA-B9-7268, TCGA-Q2-A5QZ, TCGA-SX-A71W, TCGA-DW-7963, TCGA-B3-3925, TCGA-IZ-A6M8, TCGA-GL-A9DE, TCGA-Y8-A896, TCGA-UZ-A9PS, TCGA-G7-A8LE, TCGA-2Z-A9JT, TCGA-DZ-6135, TCGA-2Z-A9J7, TCGA-HE-7129, TCGA-IZ-A6M9, TCGA-B9-4617, TCGA-2Z-A9JJ, TCGA-MH-A562, TCGA-KV-A6GE, TCGA-UZ-A9PU, TCGA-5P-A9JY, TCGA-DW-7841, TCGA-DW-7834, TCGA-2Z-A9JO, TCGA-KV-A74V, TCGA-A4-A48D, TCGA-Y8-A898, TCGA-2Z-A9J9, TCGA-2Z-A9JG, TCGA-UZ-A9PX, TCGA-V9-A7HT, TCGA-DZ-6134, TCGA-B1-A656, TCGA-DW-7836, TCGA-WN-A9G9, TCGA-F9-A97G, TCGA-EV-5901, TCGA-G7-6797, TCGA-Y8-A894, TCGA-G7-6796, TCGA-DW-5560, TCGA-G7-A8LB, TCGA-5P-A9KE, TCGA-SX-A7SL, TCGA-UZ-A9PJ, TCGA-F9-A7Q0, TCGA-Y8-A895, TCGA-A4-7584, TCGA-2Z-A9J6, TCGA-DW-7840, TCGA-2Z-A9JL, TCGA-5P-A9K6, TCGA-UZ-A9PP, TCGA-HE-A5NL, TCGA-2Z-A9JE, TCGA-MH-A855, TCGA-F9-A7VF, TCGA-A4-7287, TCGA-UZ-A9PK, TCGA-A4-A772, TCGA-A4-7828, TCGA-HE-A5NF, TCGA-5P-A9KC, TCGA-SX-A7SQ, TCGA-MH-A55Z, TCGA-2Z-A9J3, TCGA-AT-A5NU, TCGA-A4-A5DU, TCGA-A4-8098, TCGA-DW-7837, TCGA-SX-A7SO, TCGA-A4-A57E, TCGA-B9-A8YI, TCGA-G7-5902, TCGA-G7-6793, TCGA-G7-A8LC, TCGA-5P-A9K0, TCGA-MH-A857, TCGA-GL-A4EM, TCGA-SX-A7SP, TCGA-2Z-A9JS, TCGA-A4-7732, TCGA-B1-A657, TCGA-B9-4116, TCGA-2Z-A9JP, TCGA-5P-A9K3, TCGA-GL-A59T, TCGA-AL-A5DJ, TCGA-2Z-A9JD, TCGA-2K-A9WE, TCGA-MH-A854, TCGA-MH-A856, TCGA-A4-8311, TCGA-MH-A561, TCGA-MH-A55W, TCGA-G7-7501, TCGA-B9-A69E, TCGA-SX-A7SR, TCGA-F9-A4JJ, TCGA-2Z-A9JN, TCGA-B1-A47M, TCGA-SX-A7SM, TCGA-DW-7839, TCGA-IZ-8195, TCGA-G7-6790, TCGA-UZ-A9PN, TCGA-DW-7838, TCGA-B2-5633, TCGA-A3-A8OX, TCGA-AK-3453, TCGA-MM-A563, TCGA-EU-5906, TCGA-B4-5832, TCGA-EU-5907, TCGA-AK-3440, TCGA-AS-3778, TCGA-B8-5552, TCGA-B8-A8YJ, TCGA-DV-5576, TCGA-DV-5573, TCGA-DV-5574, TCGA-AK-3447, TCGA-A3-A6NL, TCGA-B8-5546, TCGA-B8-A54K, TCGA-6D-AA2E, TCGA-B4-5378, TCGA-B8-A54E, TCGA-B8-A54I, TCGA-B8-A54D, TCGA-AK-3427, TCGA-B4-5843, TCGA-B8-5550, TCGA-B8-5163, TCGA-B2-5635, TCGA-A3-A8OV, TCGA-DV-5575, TCGA-B8-5162, TCGA-3Z-A93Z, TCGA-MM-A564, TCGA-AS-3777, TCGA-G6-A8L8, TCGA-DV-5565, TCGA-G6-A8L7, TCGA-A3-A6NN, TCGA-AK-3465, TCGA-B8-A54F, TCGA-B2-5641, TCGA-DV-5568, TCGA-B2-5639, TCGA-B2-A4SR, TCGA-DV-A4W0, TCGA-T7-A92I, TCGA-B8-5551, TCGA-MW-A4EC, TCGA-B4-5836, TCGA-GK-A6C7, TCGA-AK-3443, TCGA-A3-A8OU, TCGA-B8-4151, TCGA-B8-A54G, TCGA-MM-A84U, TCGA-A3-A6NI, TCGA-EU-5905, TCGA-B8-A7U6, TCGA-A3-A8CQ, TCGA-B8-5159, TCGA-B8-4146, TCGA-B8-A54J, TCGA-B0-5399, TCGA-B8-5164, TCGA-B4-5835, TCGA-B4-5844, TCGA-B8-A54H, TCGA-DV-A4VX, TCGA-B4-5377, TCGA-B8-4621, TCGA-DV-A4VZ, TCGA-G6-A8L6, TCGA-DV-5566, TCGA-A3-A8OW, TCGA-B8-5165, TCGA-B0-5698, TCGA-AK-3455, TCGA-G6-A5PC, TCGA-B8-5553, TCGA-B4-5834, TCGA-B8-5545, TCGA-B8-4148, TCGA-A3-A6NJ, TCGA-B8-5158, TCGA-DV-5569, TCGA-B8-5549, TCGA-B4-5838, TCGA-EU-5904, TCGA-A6-2684, TCGA-DM-A0X9, TCGA-DM-A28C, TCGA-F4-6805, TCGA-AA-3949, TCGA-DM-A1DA, TCGA-AA-3532, TCGA-AY-4071, TCGA-A6-5656, TCGA-AA-3855, TCGA-AD-6889, TCGA-D5-6541, TCGA-DM-A282, TCGA-AA-3680, TCGA-F4-6569, TCGA-A6-2683, TCGA-AZ-4323, TCGA-AZ-4313, TCGA-AA-3548, TCGA-AZ-4682, TCGA-CA-6719, TCGA-AA-3679, TCGA-A6-6781, TCGA-AA-3976, TCGA-AA-3667, TCGA-5M-AAT6, TCGA-AA-3549, TCGA-CM-4748, TCGA-G4-6627, TCGA-A6-A567, TCGA-A6-5660, TCGA-AA-3986, TCGA-CK-5915, TCGA-A6-3807, TCGA-DM-A1D4, TCGA-CA-5256, TCGA-CM-5868, TCGA-CM-6161, TCGA-AA-A017, TCGA-CA-5255, TCGA-AD-6963, TCGA-D5-7000, TCGA-AA-3848, TCGA-AA-3534, TCGA-CK-5914, TCGA-AY-6197, TCGA-A6-3808, TCGA-AA-3833, TCGA-A6-A565, TCGA-AA-3966, TCGA-CK-5916, TCGA-AU-6004, TCGA-AA-3815, TCGA-NH-A50U, TCGA-D5-6535, TCGA-DM-A0XF, TCGA-AA-3684, TCGA-D5-5540, TCGA-CM-4752, TCGA-A6-4105, TCGA-AA-3560, TCGA-CM-6168, TCGA-AA-A02K, TCGA-AA-3693, TCGA-AD-6548, TCGA-AA-A004, TCGA-DM-A1D7, TCGA-A6-5665, TCGA-A6-6140, TCGA-AA-3837, TCGA-A6-5657, TCGA-F4-6459, TCGA-CA-6715, TCGA-AA-A02J, TCGA-AY-5543, TCGA-DM-A0XD, TCGA-CM-5344, TCGA-CM-6164, TCGA-G4-6309, TCGA-AY-6196, TCGA-G4-6315, TCGA-AA-A010, TCGA-CM-6163, TCGA-AA-3715, TCGA-AA-3975, TCGA-F4-6460, TCGA-D5-6929, TCGA-NH-A8F8, TCGA-D5-6536, TCGA-DM-A1D8, TCGA-AD-6890, TCGA-A6-A56B, TCGA-CM-6679, TCGA-A6-2677, TCGA-QG-A5Z1, TCGA-A6-2678, TCGA-A6-2679, TCGA-3L-AA1B, TCGA-AA-3692, TCGA-QG-A5Z2, TCGA-CK-5912, TCGA-A6-6651, TCGA-CK-4947, TCGA-AA-3821, TCGA-AA-3869, TCGA-4T-AA8H, TCGA-F4-6461, TCGA-AA-3984, TCGA-CM-6172, TCGA-A6-2686, TCGA-A6-4107, TCGA-AZ-4315, TCGA-AA-3538, TCGA-AA-3864, TCGA-AA-3517, TCGA-A6-2674, TCGA-A6-6650, TCGA-CK-6748, TCGA-G4-6297, TCGA-D5-6898, TCGA-AA-3688, TCGA-AA-A024, TCGA-CM-6169, TCGA-CM-4743, TCGA-G4-6321, TCGA-G4-6314, TCGA-AA-3531, TCGA-AA-3977, TCGA-AA-3681, TCGA-A6-3810, TCGA-AA-3989, TCGA-AD-6895, TCGA-AA-3980, TCGA-AA-3861, TCGA-A6-

3809, TCGA-AA-3968, TCGA-G4-6320, TCGA-4N-A93T, TCGA-G4-6586, TCGA-AA-A00N, TCGA-CK-4952, TCGA-AA-A02E, TCGA-G4-6626, TCGA-CK-4951, TCGA-AA-3870, TCGA-AA-A01K, TCGA-CK-4950, TCGA-AY-A54L, TCGA-A6-6649, TCGA-AA-3526, TCGA-AA-3851, TCGA-AA-3982, TCGA-G4-6628, TCGA-A6-6652, TCGA-D5-6926, TCGA-D5-6533, TCGA-A6-2671, TCGA-AA-3875, TCGA-G4-6294, TCGA-AA-3846, TCGA-AA-3556, TCGA-AA-3862, TCGA-A6-5666, TCGA-DM-A288, TCGA-NH-A8F7, TCGA-AY-A69D, TCGA-D5-6930, TCGA-A6-6654, TCGA-NH-A6GB, TCGA-AA-3696, TCGA-AZ-4684, TCGA-AA-3527, TCGA-NH-A5IV, TCGA-AY-A8YK, TCGA-AA-3514, TCGA-A6-2675, TCGA-AA-3866, TCGA-AA-3858, TCGA-AD-6964, TCGA-AA-3811, TCGA-D5-5539, TCGA-D5-6532, TCGA-AA-A01C, TCGA-AA-3666, TCGA-DM-A28K, TCGA-A6-6780, TCGA-AA-3844, TCGA-D5-6924, TCGA-QG-A5YW, TCGA-DM-A1D0, TCGA-G4-6303, TCGA-NH-A6GA, TCGA-G4-6304, TCGA-CM-5863, TCGA-A6-5667, TCGA-A6-6782, TCGA-AA-3950, TCGA-AD-5900, TCGA-CM-5861, TCGA-A6-2682, TCGA-AY-4070, TCGA-CM-4746, TCGA-AA-3867, TCGA-AA-3941, TCGA-AA-3872, TCGA-F4-6806, TCGA-CK-6751, TCGA-DM-A1D9, TCGA-CM-4747, TCGA-NH-A50V, TCGA-D5-6531, TCGA-D5-6539, TCGA-D5-6529, TCGA-AD-6899, TCGA-CM-5341, TCGA-AA-3956, TCGA-AA-3860, TCGA-A6-6142, TCGA-D5-6928, TCGA-CA-6718, TCGA-CA-6716, TCGA-A6-5659, TCGA-AA-3529, TCGA-D5-6920, TCGA-AA-3994, TCGA-AD-6965, TCGA-AA-3678, TCGA-G4-6293, TCGA-F4-6570, TCGA-AA-3972, TCGA-CK-6746, TCGA-A6-A566, TCGA-F4-6463, TCGA-DM-A1D6, TCGA-AA-3675, TCGA-AA-A02R, TCGA-AA-3852, TCGA-CM-6680, TCGA-CK-6747, TCGA-5M-AATE, TCGA-CM-6171, TCGA-AA-3955, TCGA-AY-6386, TCGA-AA-3831, TCGA-DM-A1DB, TCGA-AZ-4614, TCGA-CK-5913, TCGA-SS-A7HO, TCGA-AA-3519, TCGA-AA-3530, TCGA-G4-6310, TCGA-A6-5661, TCGA-A6-5662, TCGA-AA-3522, TCGA-CA-5797, TCGA-F4-6808, TCGA-RU-A8FL, TCGA-AA-3845, TCGA-F4-6856, TCGA-D5-6932, TCGA-D5-6538, TCGA-D5-6927, TCGA-AA-3850, TCGA-G4-6299, TCGA-A6-6648, TCGA-A6-6653, TCGA-D5-5541, TCGA-WS-AB45, TCGA-DM-A280, TCGA-D5-6923, TCGA-5M-AAT4, TCGA-AA-3673, TCGA-CM-6677, TCGA-AA-3877, TCGA-5M-AATA, TCGA-DM-A1HB, TCGA-CM-6675, TCGA-AA-3979, TCGA-AA-3812, TCGA-AA-A02H, TCGA-DM-A28F, TCGA-CM-6674, TCGA-D5-6931, TCGA-A6-2685, TCGA-G4-6306, TCGA-AA-3818, TCGA-D5-6537, TCGA-G4-6322, TCGA-AA-A02F, TCGA-AA-3971, TCGA-G4-6323, TCGA-AA-A01D, TCGA-D5-5538, TCGA-G4-6588, TCGA-AD-6901, TCGA-CM-5860, TCGA-F4-6807, TCGA-DM-A28M, TCGA-T9-A92H, TCGA-D5-5537, TCGA-CM-6167, TCGA-DM-A28H, TCGA-CM-6676, TCGA-NH-A6GC, TCGA-AA-3819, TCGA-CM-6162, TCGA-CM-6166, TCGA-G4-6298, TCGA-QG-A5YX, TCGA-CM-5348, TCGA-DM-A28A, TCGA-F4-6809, TCGA-F4-6854, TCGA-AA-3664, TCGA-DM-A1HA, TCGA-G4-6317, TCGA-AA-3854, TCGA-CK-4948, TCGA-CA-5796, TCGA-CM-4751, TCGA-CM-5862, TCGA-AM-5821, TCGA-AA-3947, TCGA-AD-A5EJ, TCGA-A6-6141, TCGA-AY-A71X, TCGA-AD-6888, TCGA-AA-3561, TCGA-AA-A00Z, TCGA-CM-6170, TCGA-AA-3856, TCGA-CM-5349, TCGA-AA-3544, TCGA-D5-6540, TCGA-CA-6717, TCGA-QG-A5YV, TCGA-AZ-4308, TCGA-G4-6311, TCGA-A6-2681, TCGA-AA-3952, TCGA-AA-3842, TCGA-G4-6295, TCGA-AA-3973, TCGA-AZ-5407, TCGA-AA-A00O, TCGA-AA-3841, TCGA-AA-3685, TCGA-AU-3779, TCGA-CM-6678, TCGA-AA-3814, TCGA-AA-3930, TCGA-AZ-4615, TCGA-AA-3672, TCGA-A6-6138, TCGA-D5-6534, TCGA-A6-5664, TCGA-QL-A97D, TCGA-AD-A5EK, TCGA-AA-A02Y, TCGA-F4-6703, TCGA-AA-A01I, TCGA-AA-3710, TCGA-F4-6855, TCGA-NH-A50T, TCGA-D5-6922, TCGA-CM-6165, TCGA-DM-A28E, TCGA-AA-A029, TCGA-AA-3524, TCGA-5M-AAT5, TCGA-DM-A28G, TCGA-AA-A02W, TCGA-G4-6302, TCGA-AZ-5403, TCGA-AM-5820, TCGA-CM-5864, TCGA-A6-6137, TCGA-D5-6530, TCGA-AZ-4616, TCGA-CA-5254, TCGA-G4-6307, TCGA-A6-2680

Supplement Table 3. Among those chosen cases in Pan-adenocarcinoma (n=3256), the cases which also provide paired-normal-like tissue are further listed here (n=152).

TCGA-44-6777, TCGA-44-2662, TCGA-44-6148, TCGA-44-5645, TCGA-44-6147, TCGA-44-6146, TCGA-44-6778, TCGA-44-2665, TCGA-44-3396, TCGA-44-2668, TCGA-91-6828, TCGA-44-2655, TCGA-44-2661, TCGA-44-3398, TCGA-44-6776, TCGA-44-6145, TCGA-44-2657, TCGA-B8-5552, TCGA-B2-5641, TCGA-B8-5549, TCGA-EM-A1CS, TCGA-EM-A1CV, TCGA-GE-A2C6, TCGA-BJ-A2NA, TCGA-BJ-A2N8, TCGA-EM-A1CU, TCGA-DO-A1JZ, TCGA-BJ-A2N7, TCGA-BJ-A3PR, TCGA-FY-A3TY, TCGA-EM-A1CW, TCGA-EM-A1CT, TCGA-BJ-A28R, TCGA-BJ-A3PU, TCGA-BJ-A290, TCGA-BJ-A2N9, TCGA-H2-A2K9, TCGA-BJ-A28X, TCGA-EJ-7315, TCGA-G9-6384, TCGA-G9-6365, TCGA-G9-6363, TCGA-HC-7752, TCGA-EJ-7314, TCGA-G9-6342, TCGA-EJ-7783, TCGA-HC-7738, TCGA-EJ-7794, TCGA-EJ-7317, TCGA-G9-6333, TCGA-EJ-7792, TCGA-EJ-7781, TCGA-G9-6499, TCGA-EJ-7125, TCGA-HC-7742, TCGA-HC-7740, TCGA-EJ-7328, TCGA-EJ-7115, TCGA-HC-7745, TCGA-EJ-7330, TCGA-G9-6351, TCGA-G9-6348, TCGA-EJ-7331, TCGA-G9-6496, TCGA-EJ-7123, TCGA-EJ-7784, TCGA-EJ-7782, TCGA-G9-6356, TCGA-EJ-7785, TCGA-G9-6362, TCGA-HC-7747, TCGA-EJ-7786, TCGA-EJ-7789, TCGA-HC-7819, TCGA-EJ-7327, TCGA-AX-A0IZ, TCGA-AX-A05Y, TCGA-AJ-A3NH, TCGA-BK-A0CB, TCGA-BK-A13C, TCGA-BG-A3PP, TCGA-BK-A4ZD, TCGA-AX-A0J0, TCGA-E6-A1M0, TCGA-AJ-A3NC, TCGA-AJ-A3NE, TCGA-AJ-A2QL, TCGA-IP-7968, TCGA-IN-7806, TCGA-BR-6802, TCGA-BR-7716, TCGA-BR-6457, TCGA-HU-A4GH, TCGA-HU-A4HB, TCGA-IN-AB1X, TCGA-BR-7717, TCGA-FP-7829, TCGA-HU-A4GY, TCGA-HU-A4GP, TCGA-HU-A4GC, TCGA-IN-8663, TCGA-FP-7735, TCGA-BR-7851, TCGA-BR-7715, TCGA-BR-7704, TCGA-A6-2684, TCGA-A6-2683, TCGA-AA-3534, TCGA-A6-5665, TCGA-A6-2678, TCGA-A6-2679, TCGA-A6-2686, TCGA-AA-3517, TCGA-AA-3531, TCGA-A6-2671, TCGA-AA-3527, TCGA-AA-3514, TCGA-A6-2675, TCGA-A6-5667, TCGA-A6-2682, TCGA-A6-5659, TCGA-A6-5662, TCGA-AA-3522, TCGA-A6-2685, TCGA-A6-2680, TCGA-EP-A12J, TCGA-FV-A3I1, TCGA-DD-A1EE, TCGA-DD-A11A, TCGA-EP-A3RK, TCGA-DD-A11B, TCGA-DD-A113, TCGA-DD-A116, TCGA-DD-A119, TCGA-DD-A1EG, TCGA-DD-A1EB, TCGA-DD-A1EH, TCGA-EP-A26S, TCGA-DZ-6133, TCGA-B9-4115, TCGA-GL-7966, TCGA-A4-A4ZT, TCGA-Y8-A8RY, TCGA-GL-6846, TCGA-GL-A59R, TCGA-GL-A9DE, TCGA-DZ-6134, TCGA-A4-A57E, TCGA-AF-2692, TCGA-AF-5654, TCGA-AF-3400, TCGA-AF-2691

Supplement Table 4. The selected cases of control-normal livers (GTEx, n=50).

subject	bloodWXSID	bloodWXSname	liverRNAID	liverRNAname
GTEX-13NZ9	SRR2156698	GTEX-13NZ9-0004-SM-5DU3T	SRR1383283	GTEX-13NZ9-1326-SM-5MR3V
GTEX-11NUK	SRR2156686	GTEX-11NUK-0004-SM-58Q87	SRR1349562	GTEX-11NUK-1226-SM-5P9GM
GTEX-146FH	SRR2163639	GTEX-146FH-0002-SM-5DU3S	SRR1313807	GTEX-146FH-1526-SM-5NQBU
GTEX-11UD2	SRR2163840	GTEX-11UD2-0001-SM-58Q8A	SRR1336314	GTEX-11UD2-1626-SM-5EQM3
GTEX-12KS4	SRR2164488	GTEX-12KS4-0003-SM-5ANGF	SRR1447547	GTEX-12KS4-1326-SM-5LUB3
GTEX-13N11	SRR2164605	GTEX-13N11-0003-SM-5JVB5	SRR1433476	GTEX-13N11-0926-SM-5IJG2
GTEX-11ZVC	SRR2164745	GTEX-11ZVC-0004-SM-58Q8O	SRR1376852	GTEX-11ZVC-0726-SM-5FQT9
GTEX-1192X	SRR2165103	GTEX-1192X-0004-SM-58Q7Q	SRR1476485	GTEX-1192X-1026-SM-5H12P
GTEX-13112	SRR2165106	GTEX-13112-0004-SM-5ANFX	SRR1324295	GTEX-13112-1426-SM-5EGH8
GTEX-13O1R	SRR2165156	GTEX-13O1R-0002-SM-5DU3N	SRR1316096	GTEX-13O1R-2026-SM-5KM3N
GTEX-13FLV	SRR2165424	GTEX-13FLV-0004-SM-5DU4T	SRR1444083	GTEX-13FLV-0326-SM-5N9DJ
GTEX-13PVR	SRR2165514	GTEX-13PVR-0003-SM-5GYDY	SRR1321877	GTEX-13PVR-0126-SM-5S2PY
GTEX-12WSG	SRR2165693	GTEX-12WSG-0001-SM-5ANGG	SRR1468314	GTEX-12WSG-0626-SM-5FQTQ
GTEX-1212Z	SRR2165753	GTEX-1212Z-0004-SM-58Q8P	SRR1353878	GTEX-1212Z-0226-SM-59HLF
GTEX-12ZZZ	SRR2166160	GTEX-12ZZZ-0002-SM-5ANGD	SRR1383763	GTEX-12ZZZ-1326-SM-59HKW
GTEX-145MF	SRR2166285	GTEX-145MF-0004-SM-5DU4B	SRR1479489	GTEX-145MF-0826-SM-5QGQA
GTEX-13QJC	SRR2166621	GTEX-13QJC-0003-SM-5GYCG	SRR1325290	GTEX-13QJC-0726-SM-5RQJK
GTEX-145MO	SRR2167259	GTEX-145MO-0001-SM-5GY9Z	SRR1478637	GTEX-145MO-2326-SM-5NQ9K
GTEX-11NV4	SRR2167397	GTEX-11NV4-0001-SM-58Q86	SRR1402590	GTEX-11NV4-1326-SM-5HL6V
GTEX-13113	SRR2167421	GTEX-13113-0003-SM-5ANGI	SRR1334866	GTEX-13113-1326-SM-5GCOI
GTEX-12WSI	SRR2167693	GTEX-12WSI-0004-SM-5ANGP	SRR1331488	GTEX-12WSI-0226-SM-5GCNA
GTEX-13SLX	SRR2167936	GTEX-13SLX-0003-SM-5DU5R	SRR1388214	GTEX-13SLX-1226-SM-5S2Q6
GTEX-132NY	SRR2170211	GTEX-132NY-0001-SM-5GYA3	SRR1407179	GTEX-132NY-0926-SM-5P9G3
GTEX-13NZB	SRR2170653	GTEX-13NZB-0004-SM-5DU5H	SRR1338103	GTEX-13NZB-0626-SM-5IFH6
GTEX-11WQC	SRR2170752	GTEX-11WQC-0003-SM-58Q8N	SRR1317532	GTEX-11WQC-0726-SM-5EQMR
GTEX-13FTZ	SRR2170783	GTEX-13FTZ-0004-SM-5DU3W	SRR1414515	GTEX-13FTZ-0726-SM-5IFFY
GTEX-132AR	SRR2155763	GTEX-132AR-0002-SM-5DU51	SRR1317554	GTEX-132AR-0426-SM-5IFH8
GTEX-12WSD	SRR2156704	GTEX-12WSD-0002-SM-5ANGA	SRR1338257	GTEX-12WSD-1426-SM-5GCN9
GTEX-11EQ9	SRR2164492	GTEX-11EQ9-0003-SM-58Q7V	SRR1414185	GTEX-11EQ9-0526-SM-5A5JZ
GTEX-131XE	SRR2165074	GTEX-131XE-0004-SM-5GYA4	SRR1354053	GTEX-131XE-0326-SM-5LZVO
GTEX-12WSM	SRR2165100	GTEX-12WSM-0001-SM-5ANFP	SRR1500661	GTEX-12WSM-0726-SM-5GCOW
GTEX-13VXU	SRR2165167	GTEX-13VXU-0004-SM-5DU5Q	SRR1403544	GTEX-13VXU-0926-SM-5IFFH
GTEX-11ZTS	SRR2165260	GTEX-11ZTS-0001-SM-58Q8K	SRR1376830	GTEX-11ZTS-1426-SM-5EQMM
GTEX-139YR	SRR2165430	GTEX-139YR-0004-SM-5DU3U	SRR1433729	GTEX-139YR-0226-SM-5IFEM
GTEX-11DXZ	SRR2165954	GTEX-11DXZ-0003-SM-58Q7X	SRR1324412	GTEX-11DXZ-0126-SM-5EGGY
GTEX-13PVQ	SRR2165940	GTEX-13PVQ-0003-SM-5GYCU	SRR1350002	GTEX-13PVQ-1526-SM-5IFEQ

GTEX-11TT1	SRR2165890	GTEX-11TT1-0001-SM-58Q8E	SRR1310433	GTEX-11TT1-1726-SM-5EQLJ
GTEX-13NYB	SRR2166282	GTEX-13NYB-0004-SM-5DU4R	SRR1390540	GTEX-13NYB-1026-SM-5IFH3
GTEX-11ZUS	SRR2166691	GTEX-11ZUS-0003-SM-58Q8H	SRR1322312	GTEX-11ZUS-2526-SM-59872
GTEX-13OW6	SRR2166865	GTEX-13OW6-0002-SM-5GYAF	SRR1323491	GTEX-13OW6-2626-SM-5IFF2
GTEX-13O3O	SRR2167027	GTEX-13O3O-0003-SM-5CV8I	SRR1485438	GTEX-13O3O-1826-SM-5IFGW
GTEX-13OVJ	SRR2167143	GTEX-13OVJ-0001-SM-5GYAL	SRR1346734	GTEX-13OVJ-1026-SM-5IFGI
GTEX-131XH	SRR2167284	GTEX-131XH-0003-SM-5DU5M	SRR1358143	GTEX-131XH-0626-SM-5LZWH
GTEX-13N2G	SRR2167365	GTEX-13N2G-0001-SM-5DU64	SRR1341877	GTEX-13N2G-0926-SM-5IFGJ
GTEX-13FTW	SRR2167440	GTEX-13FTW-0003-SM-5CV8M	SRR1412969	GTEX-13FTW-1126-SM-5J2NV
GTEX-131YS	SRR2167437	GTEX-131YS-0003-SM-5DU4F	SRR1414792	GTEX-131YS-1626-SM-5HL6C
GTEX-11GSP	SRR2167540	GTEX-11GSP-0004-SM-58Q83	SRR1312266	GTEX-11GSP-0626-SM-5986T
GTEX-11DXY	SRR2167953	GTEX-11DXY-0004-SM-58Q7Z	SRR1487890	GTEX-11DXY-0526-SM-5EGGQ
GTEX-12WSL	SRR2168744	GTEX-12WSL-0003-SM-5ANFZ	SRR1349456	GTEX-12WSL-0226-SM-5CVMJ
GTEX-11TUW	SRR2170534	GTEX-11TUW-0004-SM-58Q8F	SRR1473674	GTEX-11TUW-1726-SM-5BC5C

Supplement Table 5. The correction of read coverages at each target SNV site (Illumina short read based data), which is based on the SCNA and tumor purity being reported by ASCAT (SNP-array based data). GATK-ASERead Counter calculated the Illumina short reads coverage at each target SNV site. Because of the issues of tissue heterogeneity, the contamination reads caused by normal cell infiltrations were removed from the bulk sequence data by the provided formula in this paper. The read correction steps were applied to each target SNV site in the WESeq and RNA-Seq respectively according to the corresponding regional SCNAs and tumor purity. In the table, the allele-specific corrected WESeq read coverages are supposed to be consistent with allele-specific copy numbers reported by SNP-array data for those somatic mutations being carried by the main cell colony. Because of technical limitations, the purity corrected bulk sequence data can only reveal the purity corrected tumor micro-environments (TME). For those columns for purity corrected WESeq and RNA-Seq reads coverage, such cell heterogeneity inside TME is unavoidable, and it may sometimes make these values become negative, which will then be automatically regarded as zero. And for those columns for alternate allele fractions in purity corrected WESeq and RNA-Seq, if the ratio is greater than one, then those values will be automatically regarded as one.

(A) Example for the target sites which carry the somatic SNV mutation in *TP53*.

caseID	tumor purity	WESref	WESref corrected	WESalt	DNA ratio	RNAref	RNAref corrected	RNA alt	RNA ratio	ASCAT nMajor	ASCAT nMinor	Annotation_2	Annotation_1
TCGA-UB-A7MF	0.55	54	16.5882	52	0.75815	10	0.8235	16	0.9510	2	1	c.527G>T|MODERATE	missense_variant|protein_coding
TCGA-DD-A1EE	0.8	46	35.2000	8	0.18519	176	140.8000	0	0.0000	1	1	|MODIFIER	downstream_gene_variant|protein_coding
TCGA-CC-A3M9	0.2	77	7.2222	80	0.91720	16	-11.1111	45	1.3279	10	0	c.747G>T|MODERATE	missense_variant|protein_coding
TCGA-RC-A6M3	0.41	24	3.9253	17	0.81241	0	-9.7925	20	1.9593	3	0	c.96+393G>T|MODIFIER	intron_variant|protein_coding
TCGA-CC-A7IJ	0.47	22	-2.8907	36	1.08731	12	-58.8097	153	1.6244	3	0	c.190G>T|MODERATE	missense_variant|protein_coding
TCGA-DD-AAD5	0.52	28	-5.9048	61	1.10717	1	-14.2381	39	1.5750	3	0	|MODIFIER	downstream_gene_variant|processed_transcript
TCGA-G3-A5SJ	0.61	19	4.1800	19	0.81967	5	-5.9200	23	1.3466	2	0	c.21G>T|MODERATE	missense_variant|protein_coding
TCGA-FV-A4ZQ	0.59	30	4.1700	33	0.88781	12	-27.3600	84	1.4831	2	0	c.680G>T|MODERATE	missense_variant|protein_coding
TCGA-2Y-A9GY	0.5	47	3.5000	40	0.91954	1	-11.5000	24	1.9200	2	0	c.242G>T|MODERATE	missense_variant|protein_coding
TCGA-FV-A3I1	0.52	33	2.7600	30	0.91575	8	-8.8000	27	1.4835	2	0	c.182A>G|MODERATE	missense_variant|protein_coding
TCGA-DD-AAE0	0.66	20	1.6400	34	0.95398	2	-2.4200	11	1.2821	2	0	|MODIFIER	downstream_gene_variant|protein_coding
TCGA-DD-AACU	0.62	24	1.5800	35	0.95681	10	-10.9000	45	1.3196	2	0	|MODIFIER	upstream_gene_variant|protein_coding
TCGA-CC-A3MB	0.73	25	0.7000	65	0.98935	6	-11.2800	58	1.2414	2	0	|MODIFIER	upstream_gene_variant|protein_coding
TCGA-RG-A7D4	0.34	31	0.6400	15	0.95908	10	-20.3600	36	2.3018	2	0	c.824G>A|MODERATE	missense_variant|protein_coding
TCGA-RC-A7SK	0.61	9	0.0300	14	0.99786	2	-2.6800	10	1.3661	2	0	c.124A>T|MODERATE	missense_variant|protein_coding

TCGA-K7-AAU7	0.33	46	-0.2300	23	1.01010	33	-1.8400	19	1.1072	2	0	c.467G>C|MODERATE	missense_variant|protein_coding
TCGA-DD-A1EG	0.4	41	-0.4000	28	1.01449	7	-17.0000	33	2.0625	2	0	c.74T>G|MODERATE	missense_variant|protein_coding
TCGA-DD-A1EL	0.67	13	-2.1800	33	1.07073	5	-13.1500	50	1.3569	2	0	|MODIFIER	upstream_gene_variant|processed_transcript
TCGA-DD-AACB	0.6	17	-2.2000	31	1.07639	0	-4.4000	11	1.6667	2	0	c.366C>G|MODERATE	missense_variant|protein_coding
TCGA-BC-A217	0.77	4	-2.4400	24	1.11317	2	-5.5900	31	1.2200	2	0	c.422G>A|MODERATE	missense_variant|protein_coding
TCGA-DD-AADL	0.65	28	-3.1500	61	1.05445	1	-8.4500	26	1.4815	2	0	n.67+136A>G|MODIFIER	intron_variant&non_coding_transcript_variant|processed_transcript
TCGA-RC-A7SB	0.62	5	-3.3600	17	1.24633	2	-7.5000	23	1.4839	2	0	c.772G>A|MODERATE	missense_variant|protein_coding
TCGA-YA-A8S7	0.33	87	-3.4500	48	1.07744	9	-19.8100	34	2.3961	2	0	n.870A>C|MODIFIER	non_coding_transcript_exon_variant
TCGA-BC-A5W4	0.76	4	-5.1200	34	1.17729	0	-2.8800	12	1.3158	2	0	c.350C>G|MODERATE	missense_variant|protein_coding
TCGA-CC-A8HV	0.61	15	-6.0600	39	1.18397	3	-4.8000	17	1.3934	2	0	c.747G>T|MODERATE	missense_variant|protein_coding
TCGA-DD-AAVV	0.36	54	-6.1600	40	1.18203	4	-10.7200	19	2.2947	2	0	c.-9G>T|MODIFIER	5_prime_UTR_variant|protein_coding
TCGA-DD-AADD	0.66	6	-11.0000	44	1.33333	2	-9.2200	31	1.4233	2	0	c.263A>G|MODERATE	missense_variant|protein_coding
TCGA-BC-A3KG	0.77	8	-11.3200	76	1.17502	3	-4.3600	29	1.1769	2	0	|MODIFIER	downstream_gene_variant|protein_coding
TCGA-DD-AADW	0.83	60	40.5299	7	0.14728	101	69.3248	8	0.1035	1	0	c.411C>G|MODERATE	missense_variant|protein_coding
TCGA-DD-AADR	0.6	47	17.2857	5	0.22436	25	10.7143	0	0.0000	1	0	c.916C>T|HIGH	stop_gained|protein_coding
TCGA-DD-AAD3	0.62	44	9.3043	19	0.67128	12	-5.6232	20	1.3911	1	0	|MODIFIER	downstream_gene_variant|protein_coding
TCGA-CC-A7II	0.82	19	6.7966	21	0.75549	6	-20.8475	82	1.3409	1	0	c.747G>T|MODERATE	missense_variant|protein_coding
TCGA-ED-A459	0.69	60	5.5725	55	0.90800	0	-12.3053	26	1.8986	1	0	n.915A>G|MODIFIER	non_coding_transcript_exon_variant
TCGA-DD-AADB	0.64	28	3.6471	18	0.83152	17	-34.8824	81	1.7564	1	0	|MODIFIER	downstream_gene_variant|protein_coding
TCGA-ED-A7XO	0.44	40	1.9487	13	0.86964	7	-3.0513	7	1.7727	1	0	c.374T>C|MODERATE	missense_variant|protein_coding
TCGA-CC-A3MA	0.77	38	0.6016	62	0.99039	28	-9.3984	72	1.1501	1	0	|MODIFIER	upstream_gene_variant|protein_coding
TCGA-ZP-A9D2	0.66	29	0.0746	28	0.99734	17	-24.6119	65	1.6094	1	0	c.782+400G>A|MODIFIER	intron_variant|protein_coding
TCGA-MI-A75I	0.64	18	-0.0000	16	1.00000	0	-7.4118	14	2.1250	1	0	c.-5G>A|MODIFIER	5_prime_UTR_variant|protein_coding
TCGA-GJ-A3OU	0.31	53	-0.0769	12	1.00645	27	-5.6627	13	1.7718	1	0	|MODIFIER	downstream_gene_variant|processed_transcript
TCGA-G3-A25U	0.85	11	-0.7391	34	1.02222	3	-6.6522	34	1.2432	1	0	c.630G>T|MODERATE	missense_variant|protein_coding
TCGA-ED-A7XP	0.67	18	-0.8571	20	1.04478	10	-9.8496	30	1.4888	1	0	|MODIFIER	downstream_gene_variant|protein_coding

TCGA-2Y-A9GS	0.64	34	-0.9412	32	1.03030	13	-23.5294	56	1.7246	1	0	|MODIFIER	downstream_gene_variant|protein_coding
TCGA-G3-A7M6	0.68	21	-2.2727	27	1.09191	14	-8.3030	32	1.3504	1	0	c.644G>A|MODERATE	missense_variant|protein_coding
TCGA-CC-A8HT	0.5	29	-3.0000	19	1.18750	1	-15.0000	23	2.8750	1	0	c.315G>T|MODERATE	missense_variant|protein_coding
TCGA-DD-A114	0.34	104	-3.3494	31	1.12113	33	-30.6145	47	2.8684	1	0	|MODIFIER	downstream_gene_variant|protein_coding
TCGA-CC-A8HU	0.79	9	-4.5372	30	1.17819	1	-5.2479	17	1.4466	1	0	c.270G>T|MODERATE	missense_variant|protein_coding
TCGA-QA-A7B7	0.8	9	-4.6667	32	1.17073	5	-12.6667	48	1.3585	1	0	c.630G>T|MODERATE	missense_variant|protein_coding
TCGA-DD-AACL	0.46	31	-4.7662	20	1.31287	2	-7.8182	12	2.8696	1	0	c.316T>C|MODERATE	missense_variant|protein_coding
TCGA-UB-A7MB	0.83	11	-6.1453	48	1.14682	3	-18.5043	71	1.3525	1	0	c.614A>G|MODERATE	missense_variant|protein_coding
TCGA-DD-AACS	0.82	4	-11.2542	46	1.32390	5	-5.9831	31	1.2392	1	0	c.376T>A|MODERATE	missense_variant&splice_region_variant|protein_coding
TCGA-MI-A75G	0.66	18	-27.6716	72	1.62424	3	-11.2090	25	1.8128	1	0	c.299A>G|MODERATE	missense_variant|protein_coding

(B) Example for the target sites which carry the somatic SNV mutation in *CTNNB1*.

caseID	tumor_purity	WESref	WESref_corrected	WESalt	DNA_ratio	RNAref	RNAref_corrected	RNAalt	RNA_ratio	ASCAT_nMajor	ASCAT_nMinor	Annotation_2	Annotation_1
TCGA-G3-AAV1	0.67	149	112.625944	72	0.38998	286	219.83541	116	0.34541	4	1	c.74A>G|MODERATE	missense_variant|protein_coding
TCGA-ZS-A9CG	0.58	42	5.58228	95	0.94450	14	-32.51899	161	1.25310	4	0	c.100G>A|MODERATE	missense_variant|protein_coding
TCGA-DD-AADD	0.66	78	56.13065	50	0.47112	160	119.33668	78	0.39526	3	2	n.184C>A|MODIFIER	non_coding_transcript_exon_variant
TCGA-G3-AAV4	0.56	63	33.34783	61	0.64654	266	150.26087	218	0.59197	3	2	c.98C>G|MODERATE	missense_variant|protein_coding
TCGA-RG-A7D4	0.34	89	27.80795	51	0.64714	61	7.23841	62	0.89546	3	2	c.112T>C|MODERATE	missense_variant|protein_coding
TCGA-NI-A8LF	0.55	77	39.52055	75	0.65490	112	54.54795	121	0.68927	3	2	c.133T>C|MODERATE	missense_variant|protein_coding
TCGA-DD-AAEE	0.65	56	33.13924	73	0.68778	73	41.27848	106	0.71972	3	2	c.133T>C|MODERATE	missense_variant|protein_coding
TCGA-CC-A7IK	0.56	114	72.53846	33	0.31268	527	333.79487	158	0.32127	3	1	|MODIFIER	downstream_gene_variant|protein_coding
TCGA-ED-A7PZ	0.72	8	-12.58824	92	1.15852	2	-2.94118	22	1.15432	3	0	|MODIFIER	downstream_gene_variant|protein_coding
TCGA-DD-A116	0.56	135	79.71795	61	0.43349	12	4.10256	16	0.79592	2	2	c.76T>C|MODERATE	missense_variant|protein_coding
TCGA-DD-A1EL	0.67	66	44.85629	41	0.47754	280	178.23353	235	0.56869	2	2	|MODIFIER	downstream_gene_variant|protein_coding
TCGA-G3-A5SJ	0.61	19	11.73292	11	0.48388	110	59.37267	99	0.62511	2	2	c.977A>T|MODERATE	missense_variant|protein_coding
TCGA-2Y-A9HA	0.59	78	44.47799	52	0.53898	181	94.61635	154	0.61943	2	2	c.122C>T|MODERATE	missense_variant|protein_coding
TCGA-G3-AAV3	0.3	95	23.38462	38	0.61905	199	3.53846	164	0.97888	2	2	c.94G>T|MODERATE	missense_variant|protein_coding

TCGA-DD-A1EJ	0.62	74	39.04938	75	0.65761	236	123.40741	244	0.66411	2	2	|MODIFIER	downstream_gene_variant|protein_coding
TCGA-CC-A7IL	0.82	101	83.63830	35	0.29501	67	54.87234	28	0.33787	2	1	c.74A>T|MODERATE	missense_variant|protein_coding
TCGA-DD-AAC8	0.81	73	58.93594	31	0.34469	345	280.76512	130	0.31648	2	1	n.186G>C|MODIFIER	non_coding_transcript_exon_variant
TCGA-CC-A5UD	0.69	79	52.49442	36	0.40681	135	92.82156	48	0.34086	2	1	|MODIFIER	downstream_gene_variant|protein_coding
TCGA-XR-A8TF	0.78	82	61.89928	45	0.42096	103	81.63309	32	0.28161	2	1	c.98C>T|MODERATE	missense_variant|protein_coding
TCGA-UB-A7MD	0.42	72	21.19008	34	0.61605	87	14.14050	65	0.82132	2	1	|MODIFIER	downstream_gene_variant|protein_coding
TCGA-3K-AAZ8	0.72	55	32.35294	55	0.62963	53	26.64706	75	0.73785	2	1	c.113C>T|MODERATE	missense_variant|protein_coding
TCGA-5R-AA1C	0.75	68	39.45455	89	0.69285	151	77.00000	256	0.76877	2	1	c.113C>A|MODERATE	missense_variant|protein_coding
TCGA-G3-AAV2	0.78	70	42.93525	101	0.70170	101	56.52518	180	0.76102	2	1	c.133T>C|MODERATE	missense_variant|protein_coding
TCGA-DD-AAE9	0.74	58	32.18978	78	0.70787	21	6.95620	53	0.88398	2	1	c.121A>G|MODERATE	missense_variant|protein_coding
TCGA-DD-AAD1	0.41	61	23.24000	3	0.11433	357	145.78000	1	0.00681	2	0	c.1251C>G|LOW	synonymous_variant|protein_coding
TCGA-FV-A4ZQ	0.59	160	70.21000	59	0.45662	127	32.70000	103	0.75903	2	0	c.77C>G|MODERATE	missense_variant|protein_coding
TCGA-DD-AAW3	0.79	16	-5.63000	87	1.06919	9	-12.21000	92	1.15303	2	0	c.133T>C|MODERATE	missense_variant|protein_coding
TCGA-DD-A1EE	0.8	101	79.80000	5	0.05896	640	511.80000	1	0.00195	1	1	|MODIFIER	upstream_gene_variant|retained_intron
TCGA-DD-A1EE	0.8	106	83.40000	7	0.07743	668	534.40000	0	0.00000	1	1	None	None
TCGA-EP-A2KC	0.91	72	64.53000	11	0.14564	104	90.95000	41	0.31072	1	1	c.95A>G|MODERATE	missense_variant|protein_coding
TCGA-KR-A7K0	0.58	121	60.52000	23	0.27538	72	21.18000	49	0.69820	1	1	c.109T>G|MODERATE	missense_variant|protein_coding
TCGA-2Y-A9HB	0.66	96	53.84000	28	0.34213	153	88.06000	38	0.30144	1	1	c.121A>G|MODERATE	missense_variant|protein_coding
TCGA-K7-A6G5	0.61	75	37.56000	21	0.35861	80	27.74000	54	0.66063	1	1	|MODIFIER	downstream_gene_variant|protein_coding
TCGA-DD-A1EE	0.8	227	161.60000	100	0.38226	270	178.60000	187	0.51149	1	1	None	None
TCGA-EP-A3RK	0.44	131	41.40000	29	0.41193	77	14.28000	35	0.71023	1	1	c.80G>A|MODERATE	missense_variant|protein_coding
TCGA-DD-AACX	0.82	79	56.86000	44	0.43625	115	86.02000	46	0.34843	1	1	c.104T>G|MODERATE	missense_variant|protein_coding
TCGA-EP-A3JL	0.5	64	23.00000	18	0.43902	47	3.50000	40	0.91954	1	1	c.1161T>A|MODERATE	missense_variant|protein_coding
TCGA-G3-A5SL	0.73	120	72.21000	57	0.44114	131	61.34000	127	0.67431	1	1	c.95A>G|MODERATE	missense_variant|protein_coding
TCGA-DD-A73E	0.81	124	87.14000	70	0.44546	389	277.28000	199	0.41782	1	1	c.74A>T|MODERATE	missense_variant|protein_coding
TCGA-BW-A5NQ	0.82	126	89.82000	75	0.45504	80	41.30000	135	0.76574	1	1	c.89C>T|MODERATE	missense_variant|protein_coding
TCGA-DD-A73A	0.67	87	44.43000	42	0.48594	110	36.41000	113	0.75631	1	1	|MODIFIER	upstream_gene_variant|retained_intron

TCGA-4R-AA8I	0.74	97	55.66000	62	0.52694	129	64.78000	118	0.64558	1	1	|MODIFIER	downstream_gene_variant|protein_coding
TCGA-ZP-A9D4	0.76	63	37.56000	43	0.53376	124	63.28000	129	0.67090	1	1	c.83T>G|MODERATE	missense_variant|protein_coding
TCGA-EP-A26S	0.84	54	38.16000	45	0.54113	203	140.76000	186	0.56923	1	1	|MODIFIER	downstream_gene_variant|protein_coding
TCGA-DD-A4NF	0.7	145	74.80000	89	0.54335	361	158.80000	313	0.66342	1	1	c.94G>T|MODERATE	missense_variant|protein_coding
TCGA-ZP-A9CV	0.56	72	26.24000	32	0.54945	118	33.08000	75	0.69393	1	1	c.113C>T|MODERATE	missense_variant|protein_coding
TCGA-CC-5262	0.39	178	39.53000	49	0.55348	98	8.33000	49	0.85470	1	1	c.1149G>C|MODERATE	missense_variant|protein_coding
TCGA-LG-A9QD	0.56	91	32.04000	43	0.57303	30	4.48000	28	0.86207	1	1	c.134C>T|MODERATE	missense_variant|protein_coding
TCGA-O8-A75V	0.52	103	31.48000	46	0.59370	78	18.48000	46	0.71340	1	1	c.133T>C|MODERATE	missense_variant|protein_coding
TCGA-FV-A2QQ	0.59	75	26.21000	44	0.62669	85	28.01000	54	0.65846	1	1	c.1161T>G|MODERATE	missense_variant|protein_coding
TCGA-DD-AAW1	0.77	90	49.98000	84	0.62696	192	106.90000	178	0.62478	1	1	c.95A>G|MODERATE	missense_variant|protein_coding
TCGA-DD-A4NK	0.73	55	26.92000	49	0.64542	112	56.65000	93	0.62145	1	1	c.134C>A|MODERATE	missense_variant|protein_coding
TCGA-CC-A9FW	0.56	66	19.80000	39	0.66327	93	16.44000	81	0.83128	1	1	n.187G>T|MODIFIER	non_coding_transcript_exon_variant

Supplement Table 6. Investigation of the positive selection signals through dN/dS methods for those somatic SNV mutations (all genes included for dN/dS calculation but only COSMIC genes were shown here) among selected Pan-adenocarcinoma (n=3256).

For each gene whose somatic missense SNV mutation carriers showing allelic imbalance in RNA but with no allelic imbalance in DNA (no SCNA and no LOH, based on SNP-array data) was counted in "mutSNV_expressed_RAI_Neutral" column. Next, the WESeq reads coverage was further used for confirming the consistency to SNP-array data, and if it also showed no significant bias from the supposed 0.5 in the alternate allele fraction by binomial test in WESeq, then they were remained for allelic imbalance in transcriptional efficiency test and were counted in "mutSNV_expressed_RAI_Neutral_consistent" column. And based on the dominant bias sides in allele fraction, they were further separated into "mutSNV_expressed_RAI_Neutral_consistent_ALT" and "mutSNV_expressed_RAI_Neutral_consistent_REF", respectively. However, we have to notice that these data show "the carriers counts" but not "the SNV counts" for each gene under different situations, and only missense SNV mutation carriers were considered in this table. And, sometimes for a specific missense SNV mutation carrier, he may simultaneously show the InDel event for a specific gene, or multiple missense SNV mutations with different bias sides may occur for a specific gene, and this is the reason why the sum of ninth and tenth columns does not equal to the eighth column and so on for the fourteenth, fifteenth and sixteenth columns, and others. As for the columns inside the green window, they are about the information of dN/dS-based positive selection signals based on all the somatic SNV mutations from the selected Pan-adenocarcinoma cases, whose AITE bias sides tendencies can also be compared. "wmis_cv" stands for dN/dS ratio based on those missense SNV sites of the target gene, and "qmis_cv" stands for the q-value of the dN/dS value based on those missense SNV sites of the target gene. And they also show the information for nonsense, splicing sites, truncation sites, and so on, respectively.

gene_ID	Mut ALL	Mut IN	Mut DEL	Mut SNV	mutSNV_expressed	mutSNV_expressed_LOH	mutSNV_expressed_RAI	mutSNV_expressed_RAI_ALT	mutSNV_expressed_RAI_REF	mutSNV_expressed_RAI_BySCNV	mutSNV_expressed_RAI_ByLOH	mutSNV_expressed_RAI_Neutral	mutSNV_expressed_RAI_Neutral_consistent	mutSNV_expressed_RAI_Neutral_consistent_ALT	mutSNV_expressed_RAI_Neutral_consistent_REF	AITE_bias_side	mutSNV_expressed_RAI_Neutral_consistent_ALT_percentage	n_syn	n_mis	n_non	n_spl	wmis_cv	wnon_cv	wspl_cv	pmis_cv	ptrunc_cv	pallsubs_cv	qmis_cv	qtrunc_cv	qallsubs_cv	AlternativeAllelicRatio_DNA_average	AlternativeAllelicRatio_RNA_average	AlternativeAllelicRatio_RNAminusDNA_average
CTNNB1	210	0	0	210	200	17	127	68	60	33	15	94	52	35	17	ALT	0.673	6	136	0	1	8.4936	0.5688	0.5688	0.0000	0.5807	0.0000	0.0000	0.7171	0.0000	0.500	0.611	0.111
KRAS	351	0	0	351	217	30	112	108	5	66	25	46	35	34	1	ALT	0.971	0	50	1	0	21.0848	3.1346	3.1346	0.0000	0.3367	0.0000	0.0000	0.6858	0.0000	0.493	0.684	0.191
FAT1	154	0	0	154	85	5	48	33	18	7	3	41	28	24	5	ALT	0.828	38	77	8	2	0.8321	1.4181	1.4181	0.3614	0.3431	0.2638	0.5141	0.6858	0.4234	0.472	0.751	0.279
COL3A1	61	0	0	61	56	2	56	0	56	13	2	43	26	0	26	REF	0.000	19	51	5	5	0.9559	1.1310	1.1310	0.8687	0.7555	0.8901	0.9247	0.8270	0.9180	0.471	0.008	-0.462
MYH11	71	0	0	71	43	1	39	0	39	6	1	33	23	0	23	REF	0.000	4	9	1	0	0.2956	0.2780	0.2780	0.0000	0.1046	0.0000	0.0001	0.6858	0.0001	0.503	0.023	-0.481
CDH11	75	0	0	75	56	6	48	2	46	12	6	36	21	2	19	REF	0.095	7	4	0	0	0.2821	0.0000	0.0000	0.0099	0.1357	0.0136	0.0306	0.6858	0.0460	0.470	0.328	-0.142
TP53	353	0	0	353	318	264	297	284	15	268	259	29	20	19	1	ALT	0.950	3	106	22	15	18.1030	58.0957	58.0957	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.444	0.720	0.276
CHD4	75	0	0	75	70	2	40	23	17	8	2	32	19	15	4	ALT	0.789	2	7	2	0	0.2601	0.5781	0.5781	0.0000	0.3938	0.0000	0.0001	0.6858	0.0003	0.479	0.607	0.128
BRAF	188	6	0	182	137	4	50	42	8	8	2	42	19	18	1	ALT	0.947	4	53	3	2	5.2148	3.2358	3.2358	0.0000	0.0566	0.0001	0.0002	0.6858	0.0009	0.469	0.491	0.023

PTEN	134	0	0	134	118	29	71	55	17	34	24	37	16	14	2	ALT	0.875	2	20	9	2	3.9369	15.8810	15.8810	0.0134	0.0000	0.0000	0.0388	0.0074	0.0003	0.475	0.540	0.065
TNC	70	0	0	70	41	4	33	6	27	5	3	28	15	4	11	REF	0.267	18	33	4	0	0.7881	1.0909	1.0909	0.4273	0.8764	0.6556	0.5779	0.9147	0.7507	0.522	0.171	-0.350
PTPRC	81	0	0	81	46	3	44	0	44	19	3	25	15	0	15	REF	0.000	39	121	12	9	1.1359	1.3569	1.3569	0.4866	0.2686	0.5327	0.6339	0.6858	0.6618	0.493	0.000	-0.493
SMARCA4	70	0	0	70	61	10	31	23	8	10	8	21	13	11	2	ALT	0.846	0	0	0	0	0.0000	0.0000	0.0000	0.0000	0.0226	0.0000	0.0000	0.6858	0.0000	NA	NA	NA
CTNND1	52	0	0	52	44	0	26	17	12	3	0	23	13	12	2	ALT	0.857	3	3	1	0	0.2168	0.6589	0.6589	0.0004	0.6539	0.0017	0.0024	0.7570	0.0090	0.402	0.354	-0.049
FAT4	318	0	0	318	31	2	22	2	20	5	1	17	13	1	12	REF	0.077	64	180	5	1	1.0416	0.4996	0.4996	0.7794	0.0736	0.1366	0.8622	0.6858	0.2684	0.451	0.114	-0.337
FBLN2	64	0	0	64	26	4	23	0	23	5	4	18	13	0	13	REF	0.000	12	26	1	0	1.0895	0.4945	0.4945	0.8080	0.4557	0.6741	0.8828	0.6915	0.7636	0.479	0.092	-0.387
RANBP2	106	0	0	106	73	2	37	16	24	7	1	30	13	8	5	ALT	0.615	11	52	2	0	1.3030	0.4290	0.4290	0.1993	0.1638	0.1084	0.3323	0.6858	0.2264	0.505	0.550	0.044
MYH9	40	0	0	40	37	1	27	8	19	2	1	25	12	6	6	EQUAL	0.500	0	0	0	0	0.0000	0.0000	0.0000	0.0000	0.0095	0.0000	0.0000	0.6858	0.0000	NA	NA	NA
ARID1B	85	0	0	85	67	10	42	19	23	14	9	28	12	8	4	ALT	0.667	8	38	4	2	0.9930	2.2112	2.2112	0.9655	0.0859	0.2287	0.9814	0.6858	0.3855	0.504	0.589	0.084
ACKR3	35	0	0	35	26	3	22	0	22	2	1	20	12	0	12	REF	0.000	6	12	0	0	1.3002	0.0000	0.0000	0.6057	0.3899	0.5587	0.7333	0.6858	0.6820	0.511	0.107	-0.404
RNF213	91	0	0	91	50	5	24	12	12	6	4	18	11	8	3	ALT	0.727	0	0	0	0	0.0000	0.0000	0.0000	0.0000	0.0002	0.0000	0.0000	0.1378	0.0000	NA	NA	NA
NCOR2	81	0	0	81	59	3	32	12	21	7	3	25	11	6	5	ALT	0.545	3	10	0	0	0.2379	0.0000	0.0000	0.0000	0.0086	0.0000	0.0000	0.6858	0.0000	0.575	0.684	0.109
WNK2	55	0	0	55	31	1	16	11	6	1	1	15	11	10	1	ALT	0.909	18	29	1	0	0.7864	0.4058	0.4058	0.1777	0.2890	0.2298	0.3048	0.6858	0.3868	0.482	0.751	0.268
MECOM	55	0	0	55	41	1	22	12	10	2	1	20	11	10	1	ALT	0.909	7	21	1	1	1.1842	1.0265	1.0265	0.7030	0.9742	0.9209	0.8064	0.9841	0.9401	0.506	0.625	0.119
MET	77	0	0	77	49	1	34	23	11	14	1	20	11	10	1	ALT	0.909	9	33	2	4	1.5205	2.5206	2.5206	0.2581	0.0966	0.2304	0.4023	0.6858	0.3874	0.526	0.656	0.130
ARID1A	66	0	0	66	51	6	29	14	15	6	3	23	11	6	5	ALT	0.545	26	105	151	13	1.7856	36.0482	36.0482	0.0055	0.0000	0.0000	0.0192	0.0000	0.0000	0.487	0.598	0.112
CREBBP	85	0	0	85	65	5	26	15	11	9	4	17	10	7	3	ALT	0.700	3	6	2	1	0.1702	0.8636	0.8636	0.0000	0.7945	0.0000	0.0000	0.8552	0.0000	0.483	0.529	0.046
PRF1	46	0	0	46	16	1	15	0	15	3	1	12	10	0	10	REF	0.000	7	21	2	0	1.5512	3.3377	3.3377	0.3118	0.1848	0.3374	0.4612	0.6858	0.4985	0.581	0.048	-0.533
RHOA	21	0	0	21	21	3	19	13	6	5	3	14	10	7	3	ALT	0.700	1	21	1	0	7.9851	3.9703	3.9703	0.0002	0.2614	0.0011	0.0015	0.6858	0.0063	0.518	0.605	0.087
NUP98	58	0	0	58	47	3	20	15	5	4	2	16	9	7	2	ALT	0.778	2	4	0	2	0.1724	0.7226	0.7226	0.0000	0.6271	0.0000	0.0000	0.7410	0.0001	0.442	0.505	0.063
NUMA1	33	0	0	33	29	0	21	9	12	2	0	19	9	7	2	ALT	0.778	2	8	0	0	0.2553	0.0000	0.0000	0.0000	0.0082	0.0000	0.0000	0.6858	0.0000	0.466	0.616	0.150
KIF5B	37	0	0	37	34	0	20	11	10	2	0	18	9	7	2	ALT	0.778	5	10	0	0	0.7172	0.0000	0.0000	0.4152	0.0425	0.0934	0.5670	0.6858	0.2030	0.485	0.699	0.214

TRRAP	139	0	0	139	70	1	29	11	18	9	1	20	9	6	3	ALT	0.667	32	72	4	0	0.9829	0.5239	0.5239	0.9356	0.1859	0.3859	0.9647	0.6858	0.5418	0.479	0.600	0.120
DROSHA	47	0	0	47	35	2	18	9	10	5	2	13	9	7	2	ALT	0.778	4	23	1	1	1.1945	0.7224	0.7224	0.4125	0.6268	0.6308	0.5648	0.7408	0.7336	0.467	0.649	0.182
NSD1	86	0	0	86	55	6	26	12	14	6	3	20	9	8	1	ALT	0.889	9	45	1	1	1.2669	0.5472	0.5472	0.2019	0.3432	0.2360	0.3354	0.6858	0.3933	0.489	0.600	0.110
MTOR	117	0	0	117	74	8	34	20	14	12	8	22	9	8	1	ALT	0.889	41	164	12	3	1.5400	1.1875	1.1875	0.0103	0.5749	0.0311	0.0316	0.7150	0.0871	0.495	0.625	0.131
PIK3CA	261	0	0	261	117	10	38	22	16	18	4	20	9	6	3	ALT	0.667	5	119	1	0	8.7805	0.4630	0.4630	0.0000	0.3869	0.0000	0.0000	0.6858	0.0000	0.482	0.561	0.078
EP300	70	0	0	70	42	4	20	11	9	5	4	15	8	6	2	ALT	0.750	0	0	0	0	0.0000	0.0000	0.0000	0.0000	0.0079	0.0000	0.0000	0.6858	0.0000	NA	NA	NA
EWSR1	27	0	0	27	25	2	19	10	9	3	1	16	8	7	1	ALT	0.875	0	0	0	0	0.0000	0.0000	0.0000	0.0000	0.1172	0.0000	0.0002	0.6858	0.0003	NA	NA	NA
POLD1	36	0	0	36	29	1	14	9	5	1	0	13	8	6	2	ALT	0.750	0	0	0	0	0.0000	0.0000	0.0000	0.0000	0.0747	0.0000	0.0000	0.6858	0.0000	NA	NA	NA
PML	25	0	0	25	21	3	15	10	5	4	2	11	8	7	1	ALT	0.875	0	1	1	0	0.0671	0.9853	0.9853	0.0000	0.9882	0.0000	0.0000	0.9928	0.0002	0.484	0.514	0.030
CTCF	47	0	0	47	42	1	18	13	5	3	1	15	8	8	0	ALT	1.000	1	4	0	0	0.3763	0.0000	0.0000	0.0197	0.1454	0.0228	0.0533	0.6858	0.0685	0.425	0.553	0.128
EZR	20	0	0	20	20	2	14	11	3	3	1	11	8	7	1	ALT	0.875	1	7	0	0	0.8022	0.0000	0.0000	0.5450	0.1556	0.3037	0.6853	0.6858	0.4650	0.465	0.708	0.244
NCOA4	35	0	0	35	32	2	17	9	8	4	2	13	8	6	2	ALT	0.750	6	12	2	0	0.8149	1.0786	1.0786	0.6957	0.9275	0.8889	0.8012	0.9514	0.9172	0.490	0.556	0.066
ABL1	51	0	0	51	45	4	23	11	12	6	4	17	8	7	1	ALT	0.875	8	17	0	0	0.9904	0.0000	0.0000	0.9681	0.1089	0.2765	0.9828	0.6858	0.4366	0.445	0.652	0.207
BCL9	55	0	0	55	44	0	19	8	11	9	0	10	8	8	0	ALT	1.000	25	82	5	2	1.2257	1.5586	1.5586	0.3686	0.3212	0.5170	0.5211	0.6858	0.6485	0.469	0.690	0.221
ATP1A1	26	0	0	26	26	3	18	7	12	4	3	14	8	6	2	ALT	0.750	19	59	2	2	1.3276	0.7963	0.7963	0.2766	0.6722	0.3660	0.4227	0.7696	0.5239	0.485	0.612	0.126
SND1	27	0	0	27	16	0	13	5	8	0	0	13	8	5	3	ALT	0.625	2	20	3	1	1.4836	2.1854	2.1854	0.1184	0.1664	0.1572	0.2235	0.6858	0.2967	0.451	0.604	0.154
CARD11	66	0	0	66	19	0	13	3	10	3	0	10	8	2	6	REF	0.250	15	48	4	1	1.5416	1.5442	1.5442	0.1346	0.4198	0.3162	0.2464	0.6863	0.4780	0.495	0.547	0.052
SLC34A2	34	0	0	34	20	3	13	10	3	3	3	10	8	7	1	ALT	0.875	5	17	0	1	1.7112	1.1944	1.1944	0.1954	0.8630	0.4254	0.3274	0.9052	0.5755	0.446	0.672	0.225
PPP2R1A	30	0	0	30	29	1	23	18	5	11	1	12	7	7	0	ALT	1.000	0	0	0	0	0.0000	0.0000	0.0000	0.0000	0.1894	0.0001	0.0003	0.6858	0.0006	NA	NA	NA
SMAD2	35	0	0	35	31	5	20	11	9	5	4	15	7	6	1	ALT	0.857	0	0	0	0	0.0000	0.0000	0.0000	0.0008	0.1982	0.0016	0.0041	0.6858	0.0084	NA	NA	NA
DNM2	23	0	0	23	21	0	14	10	4	2	0	12	7	6	1	ALT	0.857	0	0	0	0	0.0000	0.0000	0.0000	0.0000	0.1045	0.0000	0.0000	0.6858	0.0000	NA	NA	NA
BRD4	37	0	0	37	29	0	11	5	6	1	0	10	7	5	2	ALT	0.714	0	0	0	0	0.0000	0.0000	0.0000	0.0000	0.0575	0.0000	0.0000	0.6858	0.0000	NA	NA	NA
CBLC	17	0	0	17	13	0	10	8	2	2	0	8	7	6	1	ALT	0.857	0	0	0	0	0.0000	0.0000	0.0000	0.0001	0.2222	0.0004	0.0011	0.6858	0.0025	NA	NA	NA

ERG	37	0	0	37	10	1	9	0	9	2	1	7	7	0	7	REF	0.000	0	0	0	0	0.0000	0.0000	0.0000	0.0002	0.2466	0.0004	0.0012	0.6858	0.0029	NA	NA	NA
ERBB2	12	0	0	12	12	2	9	9	0	2	2	7	7	7	0	ALT	1.000	1	1	0	0	0.0550	0.0000	0.0000	0.0000	0.0470	0.0000	0.0000	0.6858	0.0000	NA	NA	NA
CDH1	27	0	0	27	25	0	20	14	6	3	0	17	7	7	0	ALT	1.000	2	4	2	1	0.3187	2.5535	2.5535	0.0048	0.1600	0.0070	0.0172	0.6858	0.0271	0.507	0.558	0.051
MSH2	33	0	0	33	25	2	17	11	7	6	2	11	7	5	2	ALT	0.714	10	21	3	1	0.8161	1.2370	1.2370	0.6088	0.7254	0.7141	0.7359	0.8046	0.7918	0.460	0.523	0.063
AFDN	49	0	0	49	41	4	20	11	9	5	3	15	7	7	0	ALT	1.000	5	22	5	2	0.8195	2.1534	2.1534	0.3343	0.0719	0.1242	0.4852	0.6858	0.2504	0.471	0.713	0.242
PTPRK	71	0	0	71	54	7	26	18	8	10	7	16	7	5	2	ALT	0.714	11	27	2	1	0.9401	0.7906	0.7906	0.8660	0.7140	0.9350	0.9224	0.7974	0.9503	0.478	0.577	0.099
TRIM24	30	0	0	30	23	0	10	6	4	0	0	10	7	6	1	ALT	0.857	6	16	0	0	1.1272	0.0000	0.0000	0.8074	0.0625	0.1579	0.8825	0.6858	0.2975	0.546	0.651	0.105
PBRM1	59	0	0	59	42	5	22	11	11	7	4	15	7	6	1	ALT	0.857	5	27	11	6	1.2184	5.7385	5.7385	0.3569	0.0000	0.0000	0.5097	0.0067	0.0001	0.532	0.634	0.101
ASXL1	48	0	0	48	33	0	17	5	12	3	0	14	6	5	1	ALT	0.833	0	0	0	0	0.0000	0.0000	0.0000	0.0000	0.0560	0.0000	0.0000	0.6858	0.0000	NA	NA	NA
MLLT1	32	0	0	32	25	1	15	7	8	2	0	13	6	5	1	ALT	0.833	0	0	0	0	0.0000	0.0000	0.0000	0.0000	0.2150	0.0001	0.0002	0.6858	0.0006	NA	NA	NA
SIRPA	37	0	0	37	23	2	15	4	11	4	2	11	6	1	5	REF	0.167	0	0	0	0	0.0000	0.0000	0.0000	0.0001	0.2663	0.0003	0.0008	0.6858	0.0022	NA	NA	NA
MAPK1	16	0	0	16	16	1	8	5	3	1	0	7	6	4	2	ALT	0.667	0	0	0	0	0.0000	0.0000	0.0000	0.0014	0.2989	0.0036	0.0065	0.6858	0.0160	NA	NA	NA
CDK12	59	0	0	59	35	3	17	13	4	8	3	9	6	5	1	ALT	0.833	0	2	1	0	0.0980	0.4889	0.4889	0.0000	0.4166	0.0000	0.0000	0.6861	0.0000	0.476	0.600	0.124
LCP1	20	0	0	20	18	1	15	2	13	4	0	11	6	2	4	REF	0.333	0	1	1	0	0.1274	0.9404	0.9404	0.0020	0.9505	0.0083	0.0084	0.9674	0.0311	0.410	0.000	-0.410
FGFR2	77	0	0	77	43	9	21	15	6	11	7	10	6	6	0	ALT	1.000	0	2	1	0	0.1735	0.7692	0.7692	0.0005	0.7837	0.0023	0.0029	0.8473	0.0114	0.455	0.780	0.325
ERBB3	31	0	0	31	29	3	14	12	2	4	2	10	6	6	0	ALT	1.000	2	7	1	0	0.3817	0.4401	0.4401	0.0024	0.3420	0.0064	0.0099	0.6858	0.0253	0.463	0.764	0.301
SYK	32	0	0	32	21	2	13	7	6	3	2	10	6	4	2	ALT	0.667	5	7	2	0	0.7333	1.8842	1.8842	0.4809	0.4771	0.5305	0.6284	0.6944	0.6601	0.468	0.396	-0.072
RAD21	15	0	0	15	14	2	9	7	2	2	1	7	6	5	1	ALT	0.833	3	9	0	0	1.0969	0.0000	0.0000	0.7900	0.1297	0.3057	0.8698	0.6858	0.4668	0.503	0.595	0.092
PDGFRB	16	0	0	16	14	3	14	0	14	3	3	11	6	0	6	REF	0.000	10	25	0	1	1.2490	0.5221	0.5221	0.5540	0.4999	0.5560	0.6923	0.6974	0.6803	0.482	0.010	-0.472
MSH6	31	0	0	31	24	2	13	5	8	3	2	10	6	4	2	ALT	0.667	10	35	4	0	1.2608	1.3915	1.3915	0.5181	0.5886	0.7790	0.6620	0.7209	0.8363	0.508	0.586	0.078
LARP4B	33	0	0	33	26	0	13	5	8	3	0	10	6	2	4	REF	0.333	5	14	5	0	1.2965	3.6941	3.6941	0.5745	0.0433	0.1167	0.7097	0.6858	0.2392	0.462	0.526	0.064
ATIC	16	0	0	16	15	0	11	6	6	2	0	9	6	5	1	ALT	0.833	2	11	2	0	1.3593	2.0275	2.0275	0.3497	0.3711	0.4797	0.5018	0.6858	0.6199	0.487	0.643	0.156
BCL6	26	0	0	26	20	0	14	7	7	0	0	14	6	4	2	ALT	0.667	6	20	0	0	1.4859	0.0000	0.0000	0.3953	0.1442	0.1498	0.5475	0.6858	0.2865	0.447	0.526	0.079

DCTN1	35	0	0	35	35	0	15	8	7	1	0	14	6	6	0	ALT	1.000	10	37	0	0	1.5257	0.0000	0.0000	0.2276	0.0215	0.0087	0.3663	0.6858	0.0322	0.518	0.576	0.058
NPM1	11	0	0	11	11	1	10	4	6	1	1	9	6	4	2	ALT	0.667	2	7	0	0	1.8802	0.0000	0.0000	0.3919	0.2682	0.2762	0.5443	0.6858	0.4362	0.513	0.867	0.354
CUL3	29	0	0	29	28	1	18	9	9	3	1	15	6	5	1	ALT	0.833	3	21	5	1	2.0501	4.2789	4.2789	0.0779	0.0123	0.0391	0.1603	0.6858	0.1044	0.435	0.552	0.117
NFE2L2	32	0	0	32	31	1	15	7	8	2	0	13	6	4	2	ALT	0.667	1	19	1	0	2.5155	1.3211	1.3211	0.0131	0.7902	0.0435	0.0383	0.8520	0.1133	0.498	0.523	0.026
SEPT9	14	0	0	14	12	0	8	6	2	1	0	7	6	5	1	ALT	0.833	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
MUC16	554	0	0	554	43	9	29	16	13	11	8	18	5	4	1	ALT	0.800	0	0	0	0	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0008	0.0000	NA	NA	NA
CLTC	35	0	0	35	33	2	19	9	10	7	2	12	5	4	1	ALT	0.800	0	0	0	0	0.0000	0.0000	0.0000	0.0000	0.0179	0.0000	0.0000	0.6858	0.0000	NA	NA	NA
TOP1.00	27	0	0	27	22	1	14	6	9	3	1	11	5	3	2	ALT	0.600	0	0	0	0	0.0000	0.0000	0.0000	0.0000	0.0621	0.0000	0.0001	0.6858	0.0001	NA	NA	NA
SPOP	53	0	0	53	40	2	13	6	7	4	2	9	5	2	3	REF	0.400	0	0	0	0	0.0000	0.0000	0.0000	0.0014	0.2623	0.0033	0.0065	0.6858	0.0150	NA	NA	NA
LSM14A	20	0	0	20	19	1	11	8	3	2	0	9	5	5	0	ALT	1.000	0	0	0	0	0.0000	0.0000	0.0000	0.0003	0.1920	0.0006	0.0018	0.6858	0.0037	NA	NA	NA
NF2	27	0	0	27	22	1	9	4	5	2	1	7	5	3	2	ALT	0.600	0	0	0	0	0.0000	0.0000	0.0000	0.0000	0.1250	0.0001	0.0003	0.6858	0.0006	NA	NA	NA
GNAS	22	0	0	22	9	0	7	5	2	1	0	6	5	4	1	ALT	0.800	0	0	0	0	0.0000	0.0000	0.0000	0.0005	0.2309	0.0011	0.0026	0.6858	0.0061	NA	NA	NA
NCOR1	76	0	0	76	60	10	32	12	20	8	7	24	5	4	1	ALT	0.800	1	4	2	1	0.1218	0.6701	0.6701	0.0000	0.4576	0.0000	0.0000	0.6916	0.0000	0.468	0.491	0.023
SMARCE1	11	0	0	11	11	1	7	7	0	1	1	6	5	5	0	ALT	1.000	0	1	0	0	0.1707	0.0000	0.0000	0.0129	0.1861	0.0190	0.0378	0.6858	0.0595	NA	NA	NA
MAP2K1	16	0	0	16	16	1	8	6	2	2	1	6	5	4	1	ALT	0.800	1	1	0	0	0.1867	0.0000	0.0000	0.0206	0.2625	0.0367	0.0554	0.6858	0.0993	0.409	0.643	0.234
HIF1A	20	0	0	20	18	0	12	2	10	1	0	11	5	2	3	REF	0.400	1	3	0	0	0.2843	0.0000	0.0000	0.0060	0.0954	0.0057	0.0204	0.6858	0.0231	0.437	0.575	0.138
IDH2	13	0	0	13	13	1	11	4	7	3	1	8	5	1	4	REF	0.200	0	2	0	0	0.2856	0.0000	0.0000	0.0254	0.2573	0.0433	0.0654	0.6858	0.1131	0.405	0.361	-0.043
ELN	21	0	0	21	10	0	8	0	8	3	0	5	5	0	5	REF	0.000	11	6	0	1	0.2973	0.3558	0.3558	0.0151	0.2552	0.0478	0.0430	0.6858	0.1220	0.498	0.003	-0.495
HSP90AA1	8	0	0	8	8	1	8	6	2	2	1	6	5	5	0	ALT	1.000	1	3	0	0	0.3117	0.0000	0.0000	0.0124	0.1268	0.0137	0.0365	0.6858	0.0460	0.420	0.500	0.079
NIN	63	0	0	63	23	3	11	3	8	4	2	7	5	1	4	REF	0.200	2	9	0	0	0.3242	0.0000	0.0000	0.0000	0.0054	0.0000	0.0003	0.6735	0.0001	0.513	0.676	0.163
ARHGEF12	39	0	0	39	21	2	9	4	5	2	2	7	5	4	1	ALT	0.800	1	7	0	0	0.3410	0.0000	0.0000	0.0005	0.0109	0.0001	0.0030	0.6858	0.0009	0.481	0.769	0.287
CHST11	29	0	0	29	15	0	12	0	12	1	0	11	5	0	5	REF	0.000	1	2	1	0	0.3843	2.8457	2.8457	0.1080	0.3728	0.1848	0.2076	0.6858	0.3315	0.529	0.070	-0.458
PTPRB	85	0	0	85	26	2	14	1	13	6	2	8	5	1	4	REF	0.200	6	12	0	0	0.4185	0.0000	0.0000	0.0004	0.0119	0.0001	0.0025	0.6858	0.0008	0.496	0.446	-0.050

FUBP1	10	0	0	10	10	1	6	5	1	1	1	5	5	5	0	ALT	1.000	14	30	4	4	0.8129	1.3851	1.3851	0.5334	0.4725	0.4182	0.6753	0.6944	0.5697	0.454	0.545	0.092
RAD17	27	0	0	27	19	1	9	5	4	1	1	8	5	4	1	ALT	0.800	0	7	2	0	0.8302	1.4994	1.4994	0.6114	0.5915	0.7610	0.7380	0.7217	0.8241	0.480	0.707	0.227
RAP1GDS1	30	0	0	30	28	2	15	8	7	4	2	11	5	3	2	ALT	0.600	6	12	1	1	0.8408	0.9432	0.9432	0.7399	0.9433	0.9430	0.8338	0.9629	0.9562	0.474	0.655	0.181
NOTCH2	57	0	0	57	30	2	14	3	11	4	1	10	5	1	4	REF	0.200	36	96	11	1	0.9820	1.3290	1.3290	0.9263	0.4059	0.6355	0.9597	0.6858	0.7370	0.487	0.594	0.107
PDGFRA	46	0	0	46	18	0	17	0	17	4	0	13	5	0	5	REF	0.000	19	46	3	1	1.0005	0.7374	0.7374	0.9987	0.5684	0.8254	0.9993	0.7138	0.8715	0.445	0.026	-0.418
TPR	51	1	0	50	39	1	21	8	13	8	1	13	5	3	2	ALT	0.600	39	132	10	6	1.0146	0.7542	0.7542	0.9369	0.3342	0.5037	0.9655	0.6858	0.6389	0.479	0.525	0.046
NOTCH1	70	0	0	70	43	7	15	4	12	5	3	10	5	2	3	REF	0.400	12	46	4	0	1.0670	1.3985	1.3985	0.6635	0.5252	0.7434	0.7777	0.7015	0.8111	0.492	0.559	0.067
NFIB	35	0	0	35	19	1	12	7	5	4	1	8	5	4	1	ALT	0.800	4	9	1	1	1.0674	1.7688	1.7688	0.9175	0.5337	0.8078	0.9543	0.7045	0.8587	0.453	0.775	0.322
SPEN	93	0	0	93	44	3	16	7	10	4	2	12	5	5	0	ALT	1.000	54	161	16	2	1.0853	1.3899	1.3899	0.6018	0.2403	0.5010	0.7300	0.6858	0.6368	0.490	0.632	0.142
EZH2	35	0	0	35	25	1	8	4	4	1	0	7	5	3	2	ALT	0.600	3	11	1	0	1.0901	0.6844	0.6844	0.7778	0.6856	0.8854	0.8613	0.7781	0.9147	0.546	0.557	0.011
NDRG1	11	0	0	11	11	0	10	1	9	4	0	6	5	1	4	REF	0.200	2	7	0	0	1.1635	0.0000	0.0000	0.6960	0.2107	0.4232	0.8013	0.6858	0.5735	0.385	0.454	0.068
BIRC6	113	0	0	113	41	2	18	12	6	7	2	11	5	4	1	ALT	0.800	33	112	9	5	1.1723	1.2735	1.2735	0.4185	0.4572	0.6588	0.5701	0.6915	0.7535	0.465	0.602	0.137
ERCC3	11	0	0	11	10	0	7	4	3	0	0	7	5	4	1	ALT	0.800	4	13	1	1	1.1819	1.5023	1.5023	0.5841	0.5899	0.7771	0.7169	0.7213	0.8352	0.504	0.646	0.142
SF3B1	25	0	0	25	22	0	15	6	9	1	0	14	5	3	2	ALT	0.600	10	34	1	1	1.1821	0.5300	0.5300	0.6440	0.3838	0.4427	0.7642	0.6858	0.5900	0.470	0.553	0.083
PIK3CB	25	0	0	25	16	1	8	5	3	1	1	7	5	5	0	ALT	1.000	4	16	1	0	1.1959	0.5106	0.5106	0.5317	0.4492	0.5917	0.6742	0.6913	0.7062	0.458	0.771	0.313
PPP6C	14	0	0	14	13	1	8	4	4	0	0	8	5	4	1	ALT	0.800	1	5	1	1	1.2300	3.7692	3.7692	0.6546	0.1240	0.2988	0.7715	0.6858	0.4601	0.429	0.390	-0.039
CSF1R	22	0	0	22	12	3	9	0	9	2	2	7	5	0	5	REF	0.000	10	25	3	0	1.2381	1.3973	1.3973	0.5700	0.6242	0.8140	0.7059	0.7393	0.8632	0.458	0.023	-0.435
SETD2	39	0	0	39	19	2	12	8	4	2	2	10	5	5	0	ALT	1.000	8	42	17	6	1.2827	5.7448	5.7448	0.2120	0.0000	0.0000	0.3472	0.0002	0.0000	0.484	0.673	0.189
RBM15	45	0	0	45	15	1	7	6	1	1	1	6	5	4	1	ALT	0.800	14	51	2	0	1.4201	0.7795	0.7795	0.2370	0.7347	0.3665	0.3771	0.8117	0.5244	0.512	0.790	0.278
FH	17	0	0	17	17	0	12	7	5	2	0	10	5	4	1	ALT	0.800	8	27	4	4	1.5386	4.3810	4.3810	0.2790	0.0050	0.0189	0.4256	0.6563	0.0593	0.457	0.520	0.063
ZEB1	34	0	0	34	14	0	8	0	8	2	0	6	5	0	5	REF	0.000	9	40	6	0	1.6575	2.1797	2.1797	0.1593	0.1579	0.2608	0.2805	0.6858	0.4202	0.476	0.067	-0.409
NCOA1	66	0	0	66	37	4	14	7	7	4	3	10	5	4	1	ALT	0.800	9	42	2	1	1.7749	1.1450	1.1450	0.1057	0.8420	0.2312	0.2041	0.8901	0.3881	0.473	0.552	0.079
EPAS1	17	0	0	17	14	0	9	0	9	1	0	8	5	0	5	REF	0.000	7	31	1	0	2.1736	0.7469	0.7469	0.0520	0.7785	0.0909	0.1162	0.8434	0.1988	0.498	0.340	-0.158

BCLAF1	32	0	0	32	31	2	18	8	11	6	2	12	5	3	2	ALT	0.600	2	29	0	0	2.2561	0.0000	0.0000	0.0148	0.0586	0.0024	0.0424	0.6858	0.0114	0.444	0.412	-0.032
EGFR	65	0	0	65	39	0	20	15	5	14	0	6	5	3	2	ALT	0.600	9	53	2	0	2.7996	0.9603	0.9603	0.0018	0.9587	0.0030	0.0077	0.9732	0.0139	0.473	0.608	0.135
KEAP1	41	0	0	41	35	20	29	24	5	20	19	9	4	3	1	ALT	0.750	0	0	0	0	0.0000	0.0000	0.0000	0.0000	0.2299	0.0000	0.0001	0.6858	0.0002	NA	NA	NA
RNF43	34	0	0	34	19	4	12	8	4	4	3	8	4	4	0	ALT	1.000	0	0	0	0	0.0000	0.0000	0.0000	0.0000	0.1409	0.0000	0.0000	0.6858	0.0001	NA	NA	NA
CCNE1	19	0	0	19	16	0	10	8	2	0	0	10	4	4	0	ALT	1.000	0	0	0	0	0.0000	0.0000	0.0000	0.0007	0.2474	0.0016	0.0035	0.6858	0.0082	NA	NA	NA
CIC	33	0	0	33	23	1	9	4	5	2	1	7	4	3	1	ALT	0.750	0	0	0	0	0.0000	0.0000	0.0000	0.0000	0.0269	0.0000	0.0000	0.6858	0.0000	NA	NA	NA
VAV1	44	0	0	44	14	2	9	1	8	0	0	9	4	0	4	REF	0.000	0	0	0	0	0.0000	0.0000	0.0000	0.0000	0.0663	0.0000	0.0000	0.6858	0.0000	NA	NA	NA
ASPSCR1	12	0	0	12	9	0	6	4	2	0	0	6	4	4	0	ALT	1.000	0	0	0	0	0.0000	0.0000	0.0000	0.0000	0.1825	0.0000	0.0002	0.6858	0.0005	NA	NA	NA
BAX	7	0	0	7	7	1	5	3	2	1	1	4	4	2	2	EQUAL	0.500	0	0	0	0	0.0000	0.0000	0.0000	0.0161	0.3944	0.0384	0.0452	0.6858	0.1031	NA	NA	NA
GOLGA5	29	0	0	29	27	1	11	6	5	2	1	9	4	3	1	ALT	0.750	1	1	0	0	0.1051	0.0000	0.0000	0.0004	0.0783	0.0004	0.0024	0.6858	0.0028	0.483	0.466	-0.017
CLIP1	43	0	0	43	27	2	10	6	4	4	1	6	4	3	1	ALT	0.750	1	2	0	0	0.1059	0.0000	0.0000	0.0000	0.0235	0.0000	0.0000	0.6858	0.0000	0.473	0.435	-0.038
POLE	49	0	0	49	24	1	7	5	2	0	0	7	4	4	0	ALT	1.000	3	4	0	1	0.1146	0.2754	0.2754	0.0000	0.1014	0.0000	0.0000	0.6858	0.0000	0.506	0.711	0.204
ETV6	25	0	0	25	18	1	7	6	1	1	0	6	4	4	0	ALT	1.000	0	1	0	0	0.1512	0.0000	0.0000	0.0064	0.2211	0.0114	0.0214	0.6858	0.0400	0.554	0.530	-0.025
KTN1	35	0	0	35	29	1	17	8	9	2	1	15	4	4	0	ALT	1.000	1	3	1	1	0.1848	0.6116	0.6116	0.0001	0.4489	0.0002	0.0005	0.6913	0.0017	0.514	0.566	0.052
MYO5A	78	0	0	78	6	0	6	1	5	1	0	5	4	1	3	REF	0.250	0	6	0	0	0.2435	0.0000	0.0000	0.0000	0.0077	0.0000	0.0001	0.6858	0.0000	NA	NA	NA
PABPC1	15	0	0	15	15	0	13	5	9	3	0	10	4	1	3	REF	0.250	7	5	0	0	0.2676	0.0000	0.0000	0.0266	0.0548	0.0295	0.0677	0.6858	0.0838	0.429	0.406	-0.022
STAT3	23	0	0	23	21	1	11	2	9	3	1	8	4	1	3	REF	0.250	0	3	0	0	0.2911	0.0000	0.0000	0.0073	0.0779	0.0058	0.0237	0.6858	0.0233	0.577	0.704	0.128
SUZ12	25	0	0	25	14	0	6	5	1	0	0	6	4	4	0	ALT	1.000	0	3	0	1	0.2969	0.7667	0.7667	0.0085	0.7810	0.0302	0.0270	0.8450	0.0853	0.427	0.831	0.404
EED	17	0	0	17	15	0	8	4	4	0	0	8	4	3	1	ALT	0.750	1	2	0	0	0.3420	0.0000	0.0000	0.0651	0.1883	0.0768	0.1392	0.6858	0.1751	0.592	0.366	-0.226
DDX10	35	0	0	35	21	1	10	3	7	2	1	8	4	2	2	EQUAL	0.500	0	4	0	0	0.3432	0.0000	0.0000	0.0094	0.0843	0.0077	0.0291	0.6858	0.0292	0.500	0.656	0.156
BIRC3	22	0	0	22	15	1	10	3	7	2	1	8	4	3	1	ALT	0.750	0	3	1	0	0.3968	1.1882	1.1882	0.0587	0.8669	0.1651	0.1281	0.9080	0.3071	0.429	0.817	0.388
CREB3L1	14	0	0	14	11	0	7	3	4	0	0	7	4	2	2	EQUAL	0.500	0	3	0	0	0.3991	0.0000	0.0000	0.0606	0.2262	0.0826	0.1313	0.6858	0.1850	0.327	0.000	-0.327
ALDH2	15	1	0	14	14	0	11	5	6	1	0	10	4	2	2	EQUAL	0.500	1	4	0	0	0.5129	0.0000	0.0000	0.1331	0.2251	0.1551	0.2443	0.6858	0.2937	0.452	0.429	-0.024

RPN1	10	0	0	10	10	0	9	6	3	1	0	8	4	3	1	ALT	0.750	2	5	0	0	0.5781	0.0000	0.0000	0.1775	0.2235	0.1920	0.3046	0.6858	0.3406	0.459	0.522	0.063
FGFR1	21	0	0	21	19	1	11	2	9	2	0	9	4	2	2	EQUAL	0.500	11	15	0	0	0.6516	0.0000	0.0000	0.2955	0.0377	0.0995	0.4441	0.6858	0.2124	0.431	0.193	-0.237
KIAA1549	87	0	0	87	30	3	12	7	5	3	2	9	4	4	0	ALT	1.000	16	24	2	0	0.7079	0.7784	0.7784	0.2403	0.7306	0.4955	0.3812	0.8089	0.6330	0.483	0.830	0.347
PRDM1	37	0	0	37	11	0	6	1	5	0	0	6	4	1	3	REF	0.250	7	10	0	0	0.7264	0.0000	0.0000	0.5317	0.1743	0.3649	0.6742	0.6858	0.5229	0.495	0.188	-0.307
UBR5	75	0	0	75	44	1	22	10	12	8	1	14	4	4	0	ALT	1.000	17	40	5	2	0.7674	0.9218	0.9218	0.3752	0.8561	0.6493	0.5280	0.9008	0.7462	0.500	0.741	0.241
SMAD3	29	0	0	29	17	3	11	10	1	4	3	7	4	4	0	ALT	1.000	0	6	0	0	0.9279	0.0000	0.0000	0.8529	0.2686	0.5331	0.9134	0.6858	0.6622	0.448	0.463	0.014
ARHGAP26	25	0	0	25	10	0	5	4	1	0	0	5	4	4	0	ALT	1.000	4	10	1	0	0.9676	0.6522	0.6522	0.9167	0.6455	0.8947	0.9537	0.7515	0.9215	0.527	0.872	0.345
KDR	53	0	0	53	15	3	10	1	9	4	2	6	4	0	4	REF	0.000	19	48	7	4	1.0162	1.8083	1.8083	0.9532	0.1316	0.2500	0.9745	0.6858	0.4090	0.444	0.222	-0.222
KAT6A	86	0	0	86	47	2	17	7	10	6	2	11	4	4	0	ALT	1.000	11	36	5	0	1.2069	1.4890	1.4890	0.5863	0.4755	0.7514	0.7184	0.6944	0.8170	0.468	0.662	0.194
XPO1	17	0	0	17	12	0	8	5	3	0	0	8	4	4	0	ALT	1.000	9	30	0	0	1.2228	0.0000	0.0000	0.5988	0.0153	0.0177	0.7279	0.6858	0.0565	0.449	0.694	0.245
CUX1	41	0	0	41	25	1	12	6	6	2	0	10	4	4	0	ALT	1.000	9	30	2	3	1.2397	2.1347	2.1347	0.2992	0.1321	0.2452	0.4479	0.6858	0.4033	0.472	0.691	0.218
SDHA	12	0	0	12	12	1	8	3	5	1	1	7	4	2	2	EQUAL	0.500	6	15	1	0	1.2836	0.7732	0.7732	0.6109	0.8077	0.7929	0.7376	0.8646	0.8474	0.476	0.372	-0.105
DDR2	57	0	0	57	15	0	14	1	13	6	0	8	4	0	4	REF	0.000	23	83	5	2	1.4253	1.1215	1.1215	0.1237	0.7932	0.2822	0.2310	0.8541	0.4425	0.469	0.078	-0.391
LMNA	12	0	0	12	10	1	8	1	7	2	1	6	4	1	3	REF	0.250	8	26	0	2	1.4682	1.4354	1.4354	0.3390	0.6614	0.6297	0.4900	0.7620	0.7327	0.512	0.471	-0.042
CASP8	29	0	0	29	19	2	11	5	6	3	1	8	4	3	1	ALT	0.750	0	10	2	0	1.5078	2.3304	2.3304	0.2242	0.2941	0.2935	0.3621	0.6858	0.4548	0.438	0.425	-0.014
IL7R	23	0	0	23	7	0	7	0	7	2	0	5	4	0	4	REF	0.000	8	25	2	0	1.5110	1.0906	1.0906	0.3056	0.9140	0.5648	0.4547	0.9423	0.6864	0.486	0.000	-0.486
LCK	15	0	0	15	8	0	6	0	6	1	0	5	4	0	4	REF	0.000	12	36	1	0	1.5119	0.3727	0.3727	0.2076	0.2767	0.1284	0.3424	0.6858	0.2566	0.499	0.146	-0.354
CXCR4	11	0	0	11	10	0	9	0	9	2	0	7	4	0	4	REF	0.000	2	7	0	0	1.5446	0.0000	0.0000	0.3669	0.4589	0.4782	0.5196	0.6916	0.6188	0.440	0.032	-0.407
ARHGEF10L	43	0	0	43	28	3	16	9	7	5	3	11	4	4	0	ALT	1.000	25	79	2	4	1.5584	1.2742	1.2742	0.0467	0.6039	0.1361	0.1064	0.7282	0.2677	0.486	0.555	0.070
GMPS	37	0	0	37	36	3	20	13	7	8	3	12	4	4	0	ALT	1.000	4	15	0	1	1.5733	0.7881	0.7881	0.4268	0.8304	0.5986	0.5774	0.8814	0.7106	0.481	0.471	-0.010
H3F3A	9	0	0	9	9	1	8	5	3	1	1	7	4	3	1	ALT	0.750	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	0.460	0.539	0.079
MAF	17	0	0	17	12	2	12	1	11	3	2	9	3	0	3	REF	0.000	2	0	0	0	0.0000	0.0000	0.0000	0.0001	0.4440	0.0002	0.0005	0.6908	0.0019	0.687	0.000	-0.687
AXIN2	33	0	0	33	19	2	9	4	5	3	1	6	3	2	1	ALT	0.667	0	0	0	0	0.0000	0.0000	0.0000	0.0000	0.1558	0.0000	0.0000	0.6858	0.0000	NA	NA	NA

PATZ1	19	0	0	19	17	1	9	4	5	3	1	6	3	2	1	ALT	0.667	0	0	0	0	0.0000	0.0000	0.0000	0.0000	0.2834	0.0000	0.0001	0.6858	0.0002	NA	NA	NA
NFATC2	67	0	0	67	19	0	8	0	8	1	0	7	3	0	3	REF	0.000	0	0	0	0	0.0000	0.0000	0.0000	0.0000	0.1356	0.0000	0.0000	0.6858	0.0000	NA	NA	NA
TMPRSS2	20	0	0	20	11	1	8	4	4	2	1	6	3	3	0	ALT	1.000	0	0	0	0	0.0000	0.0000	0.0000	0.0002	0.1977	0.0003	0.0011	0.6858	0.0025	NA	NA	NA
CCNB1IP1	10	0	0	10	10	2	7	7	0	2	2	5	3	3	0	ALT	1.000	0	0	0	0	0.0000	0.0000	0.0000	0.0055	0.3510	0.0137	0.0190	0.6858	0.0460	NA	NA	NA
DGCR8	11	0	0	11	8	0	7	3	4	1	0	6	3	2	1	ALT	0.667	0	0	0	0	0.0000	0.0000	0.0000	0.0000	0.1242	0.0000	0.0000	0.6858	0.0000	NA	NA	NA
APOBEC3B	14	0	0	14	9	1	6	5	1	2	1	4	3	3	0	ALT	1.000	0	0	0	0	0.0000	0.0000	0.0000	0.0007	0.2389	0.0016	0.0037	0.6858	0.0085	NA	NA	NA
MAX	13	0	0	13	11	1	6	4	2	1	1	5	3	3	0	ALT	1.000	0	0	0	0	0.0000	0.0000	0.0000	0.0260	0.3883	0.0578	0.0665	0.6858	0.1405	0.521	0.702	0.181
CRNKL1	9	0	0	9	9	1	5	3	2	0	0	5	3	3	0	ALT	1.000	0	0	0	0	0.0000	0.0000	0.0000	0.0000	0.0787	0.0000	0.0002	0.6858	0.0002	NA	NA	NA
CALR	6	0	0	6	5	0	4	3	1	0	0	4	3	3	0	ALT	1.000	0	0	0	0	0.0000	0.0000	0.0000	0.0004	0.2450	0.0011	0.0026	0.6858	0.0061	NA	NA	NA
ETNK1	15	0	0	15	9	1	4	3	1	0	0	4	3	3	0	ALT	1.000	1	0	0	0	0.0000	0.0000	0.0000	0.0023	0.2662	0.0052	0.0096	0.6858	0.0216	NA	NA	NA
ETV4	11	0	0	11	5	0	4	3	1	0	0	4	3	3	0	ALT	1.000	0	0	0	0	0.0000	0.0000	0.0000	0.0002	0.1873	0.0003	0.0011	0.6858	0.0024	NA	NA	NA
ZNF331	18	0	0	18	7	0	3	2	1	0	0	3	3	2	1	ALT	0.667	0	0	0	0	0.0000	0.0000	0.0000	0.0003	0.2834	0.0008	0.0018	0.6858	0.0047	NA	NA	NA
LATS2	53	0	0	53	23	2	9	2	7	3	1	6	3	1	2	REF	0.333	1	2	0	0	0.1155	0.0000	0.0000	0.0000	0.1450	0.0000	0.0000	0.6858	0.0001	0.554	0.230	-0.324
FANCA	23	0	0	23	11	1	6	4	2	2	0	4	3	2	1	ALT	0.667	0	3	0	0	0.1455	0.0000	0.0000	0.0000	0.0185	0.0000	0.0000	0.6858	0.0000	NA	NA	NA
ABI1	17	0	0	17	13	0	6	4	2	0	0	6	3	3	0	ALT	1.000	2	1	0	2	0.1597	2.5021	2.5021	0.0088	0.2612	0.0172	0.0278	0.6858	0.0552	0.438	0.561	0.122
FES	14	0	0	14	7	0	5	1	4	1	0	4	3	0	3	REF	0.000	0	2	0	0	0.1628	0.0000	0.0000	0.0003	0.0900	0.0003	0.0017	0.6858	0.0023	NA	NA	NA
KMT2A	90	0	0	90	29	0	11	5	6	2	0	9	3	2	1	ALT	0.667	4	10	3	0	0.1883	0.5507	0.5507	0.0000	0.2514	0.0000	0.0000	0.6858	0.0000	0.537	0.560	0.023
SFRP4	7	0	0	7	6	0	6	0	6	0	0	6	3	0	3	REF	0.000	12	5	1	1	0.2202	1.1191	1.1191	0.0028	0.8848	0.0070	0.0112	0.9215	0.0270	0.542	0.002	-0.540
CYLD	35	0	0	35	19	1	10	3	7	2	1	8	3	2	1	ALT	0.667	0	3	0	0	0.2505	0.0000	0.0000	0.0019	0.0871	0.0019	0.0081	0.6858	0.0095	0.514	0.465	-0.049
ERCC5	37	0	0	37	33	1	10	5	5	3	0	7	3	2	1	ALT	0.667	0	4	0	0	0.2506	0.0000	0.0000	0.0003	0.0505	0.0002	0.0021	0.6858	0.0019	0.550	0.547	-0.003
BCL9L	29	0	0	29	21	2	7	4	3	3	2	4	3	3	0	ALT	1.000	3	6	2	0	0.2667	1.5427	1.5427	0.0000	0.5674	0.0002	0.0004	0.7134	0.0014	0.494	0.696	0.202
ARHGAP5	39	0	0	39	20	1	11	5	6	5	1	6	3	2	1	ALT	0.667	1	5	1	0	0.2676	0.5738	0.5738	0.0002	0.5405	0.0007	0.0012	0.7055	0.0043	0.551	0.686	0.135
DICER1	75	0	2	73	40	2	16	7	9	3	1	13	3	3	0	ALT	1.000	2	7	0	1	0.2817	0.3425	0.3425	0.0000	0.1928	0.0001	0.0003	0.6858	0.0006	0.510	0.558	0.048

AXIN1	22	0	0	22	14	1	6	5	1	2	1	4	3	2	1	ALT	0.667	1	5	9	3	0.3546	11.0152	11.0152	0.0051	0.0000	0.0000	0.0181	0.0000	0.0000	0.658	0.565	-0.093
PIK3R1	22	0	0	22	16	0	8	5	3	1	0	7	3	3	0	ALT	1.000	7	6	7	1	0.3600	3.1937	3.1937	0.0737	0.0293	0.0004	0.1534	0.6858	0.0029	0.437	0.389	-0.048
TRIP11	62	0	0	62	25	1	6	2	5	0	0	6	3	2	1	ALT	0.667	0	10	2	0	0.4180	0.5660	0.5660	0.0013	0.3739	0.0037	0.0059	0.6858	0.0165	0.480	0.447	-0.032
ARID2	62	0	0	62	28	1	10	7	3	3	1	7	3	3	0	ALT	1.000	2	10	6	2	0.4344	3.2618	3.2618	0.0022	0.0052	0.0002	0.0092	0.6569	0.0015	NA	NA	NA
FLI1	24	0	0	24	9	1	8	0	8	2	1	6	3	0	3	REF	0.000	1	3	0	0	0.4569	0.0000	0.0000	0.1189	0.2612	0.1577	0.2240	0.6858	0.2972	0.426	0.000	-0.426
MB21D2	36	0	0	36	7	1	4	2	2	1	0	3	3	2	1	ALT	0.667	10	14	1	0	0.6577	0.7351	0.7351	0.3277	0.7603	0.6190	0.4782	0.8306	0.7249	0.460	0.604	0.144
LRIG3	26	0	0	26	17	2	10	6	4	5	2	5	3	3	0	ALT	1.000	1	10	0	0	0.6579	0.0000	0.0000	0.1547	0.0822	0.0802	0.2738	0.6858	0.1808	0.573	0.596	0.023
TNFAIP3	21	0	0	21	13	1	7	1	6	4	1	3	3	1	2	REF	0.333	7	11	1	1	0.7161	1.4110	1.4110	0.5076	0.6801	0.6261	0.6526	0.7742	0.7303	0.485	0.361	-0.124
PCM1	43	0	0	43	27	6	15	6	9	6	3	9	3	3	0	ALT	1.000	4	19	1	1	0.7428	0.5383	0.5383	0.1728	0.3289	0.2451	0.2981	0.6858	0.4032	0.416	0.522	0.106
TFG	15	0	0	15	15	0	9	3	6	0	0	9	3	3	0	ALT	1.000	1	4	1	0	0.7707	1.3891	1.3891	0.5860	0.7553	0.8213	0.7184	0.8268	0.8687	0.452	0.837	0.385
RAC1	5	0	0	5	5	0	4	2	2	0	0	4	3	2	1	ALT	0.667	1	2	0	0	0.7710	0.0000	0.0000	0.7007	0.4151	0.6663	0.8048	0.6861	0.7586	0.468	0.334	-0.134
TBL1XR1	12	0	0	12	10	0	6	5	1	0	0	6	3	3	0	ALT	1.000	2	5	1	1	0.7865	2.2490	2.2490	0.5760	0.3137	0.5134	0.7108	0.6858	0.6463	0.439	0.673	0.234
FIP1L1	14	0	0	14	12	0	6	4	2	0	0	6	3	2	1	ALT	0.667	5	10	0	1	0.7965	0.4850	0.4850	0.6940	0.4825	0.7781	0.8001	0.6955	0.8359	0.486	0.616	0.131
NFKB2	9	0	0	9	6	0	5	1	4	1	0	4	3	1	2	REF	0.333	3	11	0	0	0.7968	0.0000	0.0000	0.4337	0.0716	0.1452	0.5839	0.6858	0.2801	0.500	0.618	0.118
TCF7L2	41	0	0	41	26	3	10	9	1	3	3	7	3	3	0	ALT	1.000	1	6	2	1	0.8153	3.1137	3.1137	0.6045	0.0978	0.2220	0.7324	0.6858	0.3773	0.416	0.686	0.270
THRAP3	26	0	0	26	24	0	11	5	6	2	0	9	3	3	0	ALT	1.000	20	51	4	0	0.8219	0.5616	0.5616	0.4685	0.2641	0.5148	0.6171	0.6858	0.6472	0.475	0.539	0.065
NSD2	32	0	0	32	25	1	8	5	3	2	1	6	3	3	0	ALT	1.000	12	23	0	2	0.8293	0.6921	0.6921	0.6086	0.6163	0.8276	0.7358	0.7353	0.8729	0.548	0.758	0.210
AFF4	34	0	0	34	16	1	6	5	1	3	1	3	3	3	0	ALT	1.000	3	13	1	1	0.8433	1.0526	1.0526	0.5271	0.9427	0.8166	0.6703	0.9625	0.8650	0.517	0.686	0.170
FLT4	35	0	0	35	6	0	5	0	5	0	0	5	3	0	3	REF	0.000	21	35	1	3	0.8450	1.0620	1.0620	0.5484	0.9129	0.7935	0.6882	0.9416	0.8477	0.522	0.038	-0.483
STIL	43	0	0	43	16	1	8	4	4	1	0	7	3	3	0	ALT	1.000	26	65	6	3	0.8969	1.1740	1.1740	0.6444	0.6831	0.7212	0.7645	0.7760	0.7965	0.514	0.736	0.222
AFF1	45	0	0	45	19	3	4	3	1	0	0	4	3	3	0	ALT	1.000	6	16	2	0	0.9608	1.1058	1.1058	0.8722	0.8888	0.9775	0.9270	0.9240	0.9832	0.421	0.572	0.151
TGFBR2	32	0	0	32	21	1	12	4	8	2	1	10	3	1	2	REF	0.333	6	12	1	1	1.0015	2.0102	2.0102	0.9977	0.4260	0.6952	0.9988	0.6875	0.7793	0.483	0.357	-0.125
DAXX	9	0	0	9	8	1	4	3	1	1	1	3	3	3	0	ALT	1.000	3	11	1	0	1.0292	1.0224	1.0224	0.9243	0.9824	0.9952	0.9585	0.9888	0.9965	0.484	0.595	0.111

MUC4	31	0	0	31	6	0	5	3	2	0	0	5	3	3	0	ALT	1.000	26	78	1	0	1.0364	0.2690	0.2690	0.7534	0.0937	0.2336	0.8427	0.6858	0.3907	0.479	0.938	0.460
FGFR3	21	0	0	21	11	0	5	3	2	0	0	5	3	3	0	ALT	1.000	6	15	0	0	1.0692	0.0000	0.0000	0.7978	0.1218	0.2924	0.8751	0.6858	0.4537	0.487	0.840	0.354
GATA3	30	0	0	30	7	0	5	0	5	0	0	5	3	0	3	REF	0.000	10	17	2	0	1.0812	2.0003	2.0003	0.8469	0.4095	0.7113	0.9099	0.6858	0.7899	0.450	0.193	-0.257
ACSL3	15	0	0	15	12	0	5	3	2	1	0	4	3	2	1	ALT	0.667	2	10	1	2	1.0851	2.4504	2.4504	0.7989	0.1766	0.3932	0.8758	0.6858	0.5485	0.495	0.611	0.116
NUP214	23	0	0	23	18	2	9	5	4	3	1	6	3	3	0	ALT	1.000	9	32	1	2	1.1806	1.0527	1.0527	0.4082	0.9298	0.7102	0.5609	0.9533	0.7892	0.481	0.623	0.142
AKAP9	117	0	0	117	45	0	22	4	19	11	0	11	3	2	1	ALT	0.667	17	65	2	2	1.1943	0.4350	0.4350	0.5126	0.0991	0.0703	0.6571	0.6858	0.1637	0.473	0.610	0.137
TFRC	18	0	0	18	16	1	9	4	5	2	1	7	3	3	0	ALT	1.000	3	12	0	0	1.1979	0.0000	0.0000	0.5533	0.1298	0.2640	0.6918	0.6858	0.4235	0.487	0.694	0.206
EXT1	25	0	0	25	21	0	11	3	8	5	0	6	3	2	1	ALT	0.667	5	13	0	0	1.2100	0.0000	0.0000	0.6956	0.1512	0.2970	0.8011	0.6858	0.4582	0.462	0.539	0.077
JAK1	27	0	0	27	19	1	11	4	8	1	1	10	3	2	1	ALT	0.667	27	80	12	5	1.2120	2.2006	2.2006	0.3839	0.0147	0.0494	0.5362	0.6858	0.1250	0.476	0.459	-0.017
FAS	7	0	0	7	6	0	3	3	0	0	0	3	3	3	0	ALT	1.000	1	5	0	0	1.2315	0.0000	0.0000	0.6527	0.2822	0.5068	0.7704	0.6858	0.6413	0.469	0.610	0.141
MUC1	8	0	0	8	8	0	6	1	5	1	0	5	3	1	2	REF	0.333	4	8	1	1	1.2360	2.8715	2.8715	0.7388	0.2654	0.5346	0.8330	0.6858	0.6637	0.446	0.641	0.195
NCOA2	43	0	0	43	23	2	10	5	5	5	2	5	3	3	0	ALT	1.000	10	33	1	0	1.2519	0.3646	0.3646	0.5342	0.2722	0.2985	0.6762	0.6858	0.4599	0.488	0.668	0.180
FBXO11	20	0	0	20	14	0	5	3	2	1	0	4	3	3	0	ALT	1.000	7	25	2	0	1.2547	0.7088	0.7088	0.5994	0.6617	0.6473	0.7281	0.7621	0.7455	0.465	0.788	0.323
KMT2C	110	0	0	110	42	4	17	6	11	4	1	13	3	3	0	ALT	1.000	16	87	10	5	1.3800	2.0150	2.0150	0.0382	0.0165	0.0275	0.0907	0.6858	0.0792	0.487	0.622	0.135
ROBO2	105	0	0	105	19	2	8	0	8	3	1	5	3	0	3	REF	0.000	21	77	6	3	1.3913	1.2965	1.2965	0.1710	0.5237	0.3916	0.2957	0.7014	0.5472	0.512	0.053	-0.459
ITGAV	37	0	0	37	22	2	10	1	9	3	1	7	3	0	3	REF	0.000	1	20	4	1	1.4341	2.3770	2.3770	0.1346	0.0905	0.1089	0.2464	0.6858	0.2273	0.500	0.444	-0.057
HLA-A	10	0	0	10	10	0	9	0	9	2	0	7	3	0	3	REF	0.000	2	9	2	0	1.4399	3.7304	3.7304	0.3246	0.1262	0.2392	0.4751	0.6858	0.3965	0.476	0.659	0.184
RET	50	0	0	50	6	0	3	3	0	0	0	3	3	3	0	ALT	1.000	10	30	0	0	1.5113	0.0000	0.0000	0.2521	0.0778	0.0579	0.3951	0.6858	0.1405	0.446	0.752	0.306
PTPN13	104	0	0	104	23	4	12	5	7	5	3	7	3	2	1	ALT	0.667	8	47	6	1	1.5143	1.6127	1.6127	0.0667	0.2415	0.1645	0.1420	0.6858	0.3064	0.450	0.469	0.019
BAP1	11	0	0	11	7	2	4	3	1	1	1	3	3	3	0	ALT	1.000	4	17	7	2	1.5575	8.6685	8.6685	0.1991	0.0000	0.0001	0.3321	0.0453	0.0011	0.497	0.655	0.158
SET	10	0	0	10	9	1	6	3	3	1	1	5	3	1	2	REF	0.333	0	6	0	0	1.6257	0.0000	0.0000	0.2709	0.2664	0.2940	0.4169	0.6858	0.4553	0.538	0.821	0.283
RPL5	5	0	0	5	5	0	5	4	1	0	0	5	3	3	0	ALT	1.000	6	28	1	0	2.0184	0.5278	0.5278	0.1057	0.5283	0.0975	0.2042	0.7026	0.2092	0.503	0.560	0.057
ESR1	37	0	0	37	12	0	7	3	4	1	0	6	3	2	1	ALT	0.667	3	19	4	1	2.1013	6.8580	6.8580	0.0403	0.0018	0.0053	0.0944	0.5157	0.0218	NA	NA	NA

PAX8	8	0	0	8	6	0	6	4	2	1	0	5	3	3	0	ALT	1.000	2	15	0	0	2.2310	0.0000	0.0000	0.0373	0.2428	0.0396	0.0890	0.6858	0.1054	0.505	0.604	0.099
PRCC	9	0	0	9	9	1	6	5	2	1	1	5	3	3	0	ALT	1.000	5	23	1	0	2.5331	1.2727	1.2727	0.0478	0.8318	0.1249	0.1083	0.8823	0.2516	0.518	0.613	0.095
IDH1	17	0	0	17	15	0	10	6	4	2	0	8	3	3	0	ALT	1.000	1	16	0	0	2.8999	0.0000	0.0000	0.0151	0.2614	0.0146	0.0429	0.6858	0.0483	0.524	0.732	0.208
LEF1	15	0	0	15	10	0	7	5	2	0	0	7	3	3	0	ALT	1.000	2	16	0	1	3.0725	1.6458	1.6458	0.0183	0.6455	0.0590	0.0504	0.7515	0.1428	0.488	0.329	-0.159
FBXW7	81	0	0	81	42	3	12	9	3	5	3	7	3	1	2	REF	0.333	3	30	11	2	3.3776	10.2218	10.2218	0.0057	0.0000	0.0000	0.0196	0.0089	0.0003	0.458	0.549	0.091
SMAD4	40	0	0	40	34	19	22	20	2	17	17	5	2	2	0	ALT	1.000	0	0	0	0	0.0000	0.0000	0.0000	0.0001	0.1957	0.0003	0.0010	0.6858	0.0023	NA	NA	NA
SMARCB1	14	0	0	14	13	0	10	6	4	2	0	8	2	2	0	ALT	1.000	0	0	0	0	0.0000	0.0000	0.0000	0.0007	0.2610	0.0017	0.0037	0.6858	0.0090	NA	NA	NA
CNOT3	18	0	0	18	17	1	9	6	3	1	0	8	2	2	0	ALT	1.000	0	0	0	0	0.0000	0.0000	0.0000	0.0000	0.1268	0.0000	0.0000	0.6858	0.0001	NA	NA	NA
SDC4	14	0	0	14	14	1	8	5	3	4	1	4	2	2	0	ALT	1.000	0	0	0	0	0.0000	0.0000	0.0000	0.0149	0.4335	0.0381	0.0426	0.6885	0.1024	NA	NA	NA
ERCC2	17	0	0	17	13	1	7	4	3	3	1	4	2	1	1	EQUAL	0.500	0	0	0	0	0.0000	0.0000	0.0000	0.0000	0.1110	0.0000	0.0000	0.6858	0.0000	NA	NA	NA
AKT1	11	0	0	11	11	0	6	3	3	0	0	6	2	2	0	ALT	1.000	0	0	0	0	0.0000	0.0000	0.0000	0.0001	0.1844	0.0002	0.0006	0.6858	0.0014	NA	NA	NA
CRTC1	22	0	0	22	18	1	6	3	3	2	1	4	2	2	0	ALT	1.000	0	0	0	0	0.0000	0.0000	0.0000	0.0000	0.1977	0.0000	0.0001	0.6858	0.0002	NA	NA	NA
KAT7	21	0	0	21	15	1	6	4	2	2	1	4	2	2	0	ALT	1.000	0	0	0	0	0.0000	0.0000	0.0000	0.0000	0.1208	0.0001	0.0004	0.6858	0.0006	NA	NA	NA
MAFB	14	0	0	14	8	0	6	0	6	2	0	4	2	0	2	REF	0.000	0	0	0	0	0.0000	0.0000	0.0000	0.0004	0.5377	0.0015	0.0022	0.7053	0.0078	NA	NA	NA
PRKAR1A	19	0	0	19	16	1	6	2	4	1	0	5	2	1	1	EQUAL	0.500	0	0	0	0	0.0000	0.0000	0.0000	0.0009	0.2020	0.0018	0.0044	0.6858	0.0091	NA	NA	NA
BCR	20	0	0	20	15	1	5	2	3	1	1	4	2	2	0	ALT	1.000	0	0	0	0	0.0000	0.0000	0.0000	0.0000	0.0614	0.0000	0.0000	0.6858	0.0000	NA	NA	NA
JAK3	29	0	0	29	6	0	5	0	5	1	0	4	2	0	2	REF	0.000	0	0	0	0	0.0000	0.0000	0.0000	0.0000	0.0711	0.0000	0.0000	0.6858	0.0000	NA	NA	NA
SRC	13	0	0	13	10	0	5	3	2	1	0	4	2	2	0	ALT	1.000	0	0	0	0	0.0000	0.0000	0.0000	0.0000	0.1975	0.0001	0.0003	0.6858	0.0007	NA	NA	NA
TCF3	11	0	0	11	9	0	5	2	3	0	0	5	2	1	1	EQUAL	0.500	0	0	0	0	0.0000	0.0000	0.0000	0.0000	0.1447	0.0000	0.0001	0.6858	0.0001	NA	NA	NA
GNA11	6	0	0	6	6	1	4	4	0	1	1	3	2	2	0	ALT	1.000	0	0	0	0	0.0000	0.0000	0.0000	0.0003	0.3351	0.0010	0.0020	0.6858	0.0058	NA	NA	NA
KDSR	9	0	0	9	8	1	4	3	1	1	1	3	2	2	0	ALT	1.000	0	0	0	0	0.0000	0.0000	0.0000	0.0030	0.2765	0.0067	0.0117	0.6858	0.0263	NA	NA	NA
MAP2K2	5	0	0	5	4	0	4	2	2	0	0	4	2	2	0	ALT	1.000	0	0	0	0	0.0000	0.0000	0.0000	0.0003	0.2727	0.0008	0.0018	0.6858	0.0047	NA	NA	NA
TPM4	8	0	0	8	6	0	4	0	4	0	0	4	2	0	2	REF	0.000	0	0	0	0	0.0000	0.0000	0.0000	0.0034	0.2369	0.0068	0.0129	0.6858	0.0265	NA	NA	NA

COX6C	8	0	0	8	5	0	3	1	2	0	0	3	2	1	1	EQUAL	0.500	0	0	0	0	0.0000	0.0000	0.0000	0.1129	0.6630	0.2589	0.2151	0.7630	0.4184	NA	NA	NA
ELL	9	0	0	9	6	0	3	1	2	1	0	2	2	1	1	EQUAL	0.500	0	0	0	0	0.0000	0.0000	0.0000	0.0000	0.1905	0.0000	0.0001	0.6858	0.0002	NA	NA	NA
HRAS	8	0	0	8	8	0	3	1	2	0	0	3	2	0	2	REF	0.000	0	0	0	0	0.0000	0.0000	0.0000	0.0141	0.4974	0.0391	0.0407	0.6971	0.1045	NA	NA	NA
MRTFA	20	0	0	20	14	1	3	2	1	1	1	2	2	2	0	ALT	1.000	0	0	0	0	0.0000	0.0000	0.0000	0.0000	0.1242	0.0000	0.0000	0.6858	0.0000	NA	NA	NA
PDGFB	12	0	0	12	4	0	3	0	3	0	0	3	2	0	2	REF	0.000	0	0	0	0	0.0000	0.0000	0.0000	0.0033	0.3487	0.0086	0.0127	0.6858	0.0320	NA	NA	NA
PTK6	12	0	0	12	4	0	3	1	2	1	0	2	2	1	1	EQUAL	0.500	0	0	0	0	0.0000	0.0000	0.0000	0.0001	0.2562	0.0003	0.0008	0.6858	0.0022	NA	NA	NA
SETD1B	48	0	0	48	24	0	5	2	3	0	0	5	2	2	0	ALT	1.000	8	2	0	0	0.0634	0.0000	0.0000	0.0000	0.0415	0.0000	0.0000	0.6858	0.0000	0.385	0.566	0.181
RFWD3	22	0	0	22	14	2	7	5	2	2	2	5	2	2	0	ALT	1.000	0	1	1	0	0.1008	0.8493	0.8493	0.0003	0.8667	0.0013	0.0017	0.9079	0.0072	0.552	0.819	0.267
FANCF	10	0	0	10	7	0	2	1	1	0	0	2	2	1	1	EQUAL	0.500	1	1	0	0	0.1553	0.0000	0.0000	0.0075	0.4541	0.0211	0.0243	0.6913	0.0645	0.372	0.562	0.190
MAP2K4	8	0	0	8	5	2	4	4	0	2	2	2	2	2	0	ALT	1.000	0	1	1	1	0.1864	2.8788	2.8788	0.0205	0.2032	0.0304	0.0550	0.6858	0.0857	NA	NA	NA
TCF12	19	0	0	19	13	0	8	2	6	2	0	6	2	2	0	ALT	1.000	0	2	0	0	0.2171	0.0000	0.0000	0.0039	0.1060	0.0042	0.0146	0.6858	0.0183	0.468	0.322	-0.146
KDM5A	59	0	0	59	27	0	10	1	9	1	0	9	2	1	1	EQUAL	0.500	0	5	0	0	0.2198	0.0000	0.0000	0.0000	0.0148	0.0000	0.0001	0.6858	0.0000	0.519	0.554	0.035
LASP1	5	0	0	5	5	0	2	1	1	0	0	2	2	1	1	EQUAL	0.500	1	1	0	0	0.2217	0.0000	0.0000	0.0453	0.3382	0.0852	0.1039	0.6858	0.1894	NA	NA	NA
CHD2	40	0	0	40	22	1	9	4	5	5	1	4	2	2	0	ALT	1.000	1	6	0	0	0.2431	0.0000	0.0000	0.0000	0.0072	0.0000	0.0001	0.6858	0.0000	0.580	0.257	-0.323
PICALM	23	0	0	23	19	0	11	4	7	1	0	10	2	1	1	EQUAL	0.500	2	2	0	0	0.2474	0.0000	0.0000	0.0103	0.1503	0.0132	0.0316	0.6858	0.0450	0.500	0.514	0.014
FUS	6	0	0	6	6	0	4	2	2	1	0	3	2	2	0	ALT	1.000	1	2	0	0	0.2557	0.0000	0.0000	0.0128	0.1423	0.0154	0.0376	0.6858	0.0505	0.582	0.420	-0.162
PER1	13	0	0	13	7	1	5	2	3	2	1	3	2	1	1	EQUAL	0.500	0	5	0	0	0.2578	0.0000	0.0000	0.0001	0.0595	0.0001	0.0008	0.6858	0.0008	0.542	0.727	0.185
VTI1A	4	0	0	4	3	0	2	2	0	0	0	2	2	2	0	ALT	1.000	0	1	0	0	0.3089	0.0000	0.0000	0.1449	0.2766	0.1912	0.2610	0.6858	0.3396	NA	NA	NA
TBX3	37	0	0	37	10	0	5	1	4	0	0	5	2	1	1	EQUAL	0.500	1	4	0	0	0.3170	0.0000	0.0000	0.0046	0.2671	0.0097	0.0165	0.6858	0.0351	0.580	0.426	-0.154
NF1	61	0	0	61	26	2	13	4	9	4	2	9	2	2	0	ALT	1.000	2	12	0	0	0.3332	0.0000	0.0000	0.0000	0.0025	0.0000	0.0001	0.5558	0.0000	0.486	0.354	-0.132
TAF15	14	0	0	14	14	0	11	4	7	3	0	8	2	2	0	ALT	1.000	0	3	0	0	0.3610	0.0000	0.0000	0.0337	0.0797	0.0226	0.0821	0.6858	0.0680	0.483	0.368	-0.116
LHFPL6	11	0	0	11	7	0	5	0	5	2	0	3	2	0	2	REF	0.000	1	1	0	0	0.3729	0.0000	0.0000	0.2383	0.4994	0.3972	0.3788	0.6973	0.5525	0.529	0.286	-0.243
CBLB	44	0	0	44	29	1	10	4	6	2	1	8	2	2	0	ALT	1.000	8	9	0	0	0.4291	0.0000	0.0000	0.0910	0.0364	0.0490	0.1816	0.6858	0.1244	0.472	0.509	0.036

CD74	3	0	0	3	3	0	3	0	3	0	0	3	2	0	2	REF	0.000	2	2	0	0	0.4766	0.0000	0.0000	0.2319	0.3195	0.2982	0.3714	0.6858	0.4595	0.507	0.429	-0.078
SDHD	4	0	0	4	4	0	3	2	1	1	0	2	2	1	1	EQUAL	0.500	0	1	0	0	0.4774	0.0000	0.0000	0.3992	0.4764	0.5440	0.5512	0.6944	0.6706	0.578	0.584	0.006
PRKCB	47	0	0	47	5	0	5	0	5	2	0	3	2	0	2	REF	0.000	0	5	0	0	0.5092	0.0000	0.0000	0.0891	0.1455	0.0817	0.1787	0.6858	0.1835	0.543	0.000	-0.543
TFEB	16	0	0	16	11	0	6	0	6	2	0	4	2	0	2	REF	0.000	6	6	0	1	0.5258	0.9640	0.9640	0.2836	0.9730	0.5347	0.4307	0.9832	0.6637	0.474	0.410	-0.064
TRIM27	9	0	0	9	9	0	4	3	1	0	0	4	2	1	1	EQUAL	0.500	5	5	1	0	0.5389	1.1601	1.1601	0.2664	0.8949	0.5131	0.4117	0.9284	0.6461	0.490	0.573	0.083
NFKBIE	8	0	0	8	7	0	3	1	2	0	0	3	2	1	1	EQUAL	0.500	1	5	0	0	0.6036	0.0000	0.0000	0.2179	0.2631	0.2502	0.3545	0.6858	0.4092	0.524	0.417	-0.107
ARHGEF10	51	0	0	51	17	1	8	2	6	3	1	5	2	2	0	ALT	1.000	15	20	0	2	0.6293	0.5537	0.5537	0.1870	0.3991	0.3823	0.3167	0.6858	0.5383	0.477	0.515	0.038
HNF1A	10	0	0	10	3	0	2	2	0	0	0	2	2	2	0	ALT	1.000	0	6	1	2	0.6542	3.5195	3.5195	0.2634	0.0712	0.1051	0.4082	0.6858	0.2210	0.434	0.780	0.347
KAT6B	43	0	0	43	21	2	7	5	2	2	0	5	2	2	0	ALT	1.000	13	23	3	1	0.6760	1.1381	1.1381	0.2488	0.8239	0.4196	0.3913	0.8767	0.5708	0.484	0.527	0.044
FNBP1	23	0	0	23	11	1	4	1	3	1	0	3	2	1	1	EQUAL	0.500	1	6	0	0	0.7026	0.0000	0.0000	0.3582	0.1194	0.1950	0.5110	0.6858	0.3446	0.471	0.383	-0.088
WWTR1	9	0	0	9	7	1	4	2	2	1	1	3	2	1	1	EQUAL	0.500	1	5	0	0	0.8057	0.0000	0.0000	0.6162	0.2921	0.5063	0.7420	0.6858	0.6411	0.570	0.538	-0.033
PTPN11	37	0	0	37	30	3	10	7	3	4	3	6	2	2	0	ALT	1.000	1	7	0	0	0.8410	0.0000	0.0000	0.6370	0.1363	0.2950	0.7578	0.6858	0.4562	0.430	0.510	0.080
SGK1	9	0	0	9	5	0	4	0	4	0	0	4	2	0	2	REF	0.000	2	6	0	1	0.8430	1.1928	1.1928	0.6666	0.8640	0.8981	0.7799	0.9060	0.9243	0.482	0.412	-0.070
PPARG	12	0	0	12	8	1	3	3	0	1	1	2	2	2	0	ALT	1.000	5	9	1	0	0.8720	1.1582	1.1582	0.8145	0.8960	0.9487	0.8876	0.9292	0.9606	0.437	0.532	0.094
CCNC	11	0	0	11	10	0	6	2	4	2	0	4	2	2	0	ALT	1.000	1	3	2	0	0.8830	3.3024	3.3024	0.8258	0.1661	0.3609	0.8954	0.6858	0.5191	0.378	0.515	0.137
POLQ	108	0	0	108	17	2	6	4	2	2	1	4	2	2	0	ALT	1.000	20	54	2	4	0.8890	0.8313	0.8313	0.6595	0.6876	0.8835	0.7746	0.7796	0.9133	0.452	0.590	0.138
GPHN	35	0	0	35	19	2	10	6	4	2	2	8	2	2	0	ALT	1.000	2	9	1	0	0.9007	0.7136	0.7136	0.7494	0.7207	0.8914	0.8401	0.8015	0.9190	0.506	0.760	0.254
FOXO3	30	0	0	30	16	2	6	3	3	3	2	3	2	1	1	EQUAL	0.500	2	10	2	0	0.9112	3.6061	3.6061	0.7651	0.1345	0.3121	0.8508	0.6858	0.4739	0.506	0.577	0.071
NT5C2	12	0	0	12	11	2	5	4	1	2	2	3	2	2	0	ALT	1.000	0	7	0	0	0.9376	0.0000	0.0000	0.8633	0.1446	0.3400	0.9208	0.6858	0.5010	0.487	0.820	0.333
APC	67	0	0	67	11	0	5	2	3	0	0	5	2	2	0	ALT	1.000	14	42	114	7	0.9544	24.5769	24.5769	0.8820	0.0000	0.0000	0.9332	0.0000	0.0000	0.502	0.679	0.178
SUFU	21	0	0	21	7	0	2	1	1	0	0	2	2	1	1	EQUAL	0.500	2	7	0	0	0.9676	0.0000	0.0000	0.9301	0.1964	0.4324	0.9615	0.6858	0.5813	0.475	0.625	0.150
QKI	17	0	0	17	13	3	9	1	8	3	2	6	2	0	2	REF	0.000	6	12	0	0	1.0389	0.0000	0.0000	0.9410	0.1126	0.2469	0.9677	0.6858	0.4053	0.465	0.275	-0.189
MACC1	43	0	0	43	11	1	6	4	2	3	1	3	2	2	0	ALT	1.000	6	16	0	0	1.0728	0.0000	0.0000	0.8866	0.1126	0.2341	0.9359	0.6858	0.3913	0.438	0.828	0.390

MLF1.00	7	0	0	7	5	0	4	3	1	1	0	3	2	1	1	EQUAL	0.500	2	4	1	0	1.0929	1.5434	1.5434	0.9264	0.7426	0.9444	0.9597	0.8173	0.9573	0.367	0.789	0.423
PTCH1	59	0	0	59	16	0	7	0	7	1	0	6	2	0	2	REF	0.000	13	29	1	0	1.0968	0.4724	0.4724	0.7839	0.4215	0.6301	0.8657	0.6867	0.7330	0.447	0.602	0.154
BRD3	23	0	0	23	17	0	6	6	0	1	0	5	2	2	0	ALT	1.000	3	13	1	1	1.1306	2.1188	2.1188	0.6646	0.3451	0.5830	0.7787	0.6858	0.6999	0.531	0.667	0.136
A1CF	32	0	0	32	4	0	2	2	0	0	0	2	2	2	0	ALT	1.000	7	18	2	2	1.1393	1.9214	1.9214	0.7739	0.3207	0.6004	0.8578	0.6858	0.7119	0.490	0.679	0.189
PRDM16	73	0	0	73	6	0	3	1	2	1	0	2	2	1	1	EQUAL	0.500	43	97	6	1	1.1569	1.2077	1.2077	0.4240	0.6511	0.7078	0.5751	0.7553	0.7877	0.472	0.705	0.232
PDE4DIP	61	0	0	61	22	0	10	2	8	4	0	6	2	1	1	EQUAL	0.500	32	109	18	3	1.2273	1.6100	1.6100	0.3026	0.0976	0.2464	0.4516	0.6858	0.4047	0.459	0.489	0.030
FOXP1	40	0	0	40	25	6	11	6	5	5	4	6	2	2	0	ALT	1.000	6	16	2	2	1.3008	2.2952	2.2952	0.5881	0.2225	0.4750	0.7198	0.6858	0.6159	0.463	0.588	0.125
ACVR1	31	0	0	31	23	1	7	2	5	2	0	5	2	1	1	EQUAL	0.500	3	9	0	0	1.3103	0.0000	0.0000	0.5439	0.2331	0.3758	0.6845	0.6858	0.5325	0.501	0.501	0.000
PSIP1	18	0	0	18	16	2	11	4	8	3	2	8	2	2	1	ALT	0.667	2	9	3	1	1.3403	3.2258	3.2258	0.4735	0.0800	0.2115	0.6220	0.6858	0.3655	0.439	0.606	0.167
RAF1	12	0	0	12	11	2	5	5	0	2	1	3	2	2	0	ALT	1.000	2	12	0	0	1.3525	0.0000	0.0000	0.3378	0.1392	0.2041	0.4887	0.6858	0.3555	0.527	0.619	0.092
BCL11A	31	0	0	31	7	0	4	1	3	0	0	4	2	1	1	EQUAL	0.500	18	47	2	0	1.3600	1.0368	1.0368	0.2621	0.9616	0.5160	0.4067	0.9751	0.6478	0.548	0.565	0.017
HSP90AB1	13	0	0	13	12	0	11	5	6	1	0	10	2	2	0	ALT	1.000	3	14	1	0	1.3888	0.9340	0.9340	0.3547	0.9450	0.6317	0.5076	0.9640	0.7342	0.520	0.617	0.097
SFPQ	12	0	0	12	11	0	5	4	1	1	0	4	2	2	0	ALT	1.000	12	36	4	0	1.4305	1.5488	1.5488	0.2772	0.4680	0.5240	0.4234	0.6927	0.6543	0.493	0.627	0.135
NR4A3	52	0	0	52	6	0	5	0	5	0	0	5	2	0	2	REF	0.000	3	14	0	0	1.4414	0.0000	0.0000	0.2215	0.2284	0.2161	0.3587	0.6858	0.3709	0.495	0.054	-0.441
AKT3	41	0	0	41	10	0	8	0	8	3	0	5	2	0	2	REF	0.000	12	46	8	3	1.4534	2.1386	2.1386	0.2404	0.0759	0.1999	0.3812	0.6858	0.3494	0.452	0.406	-0.046
DEK	12	0	0	12	12	1	5	2	3	2	0	3	2	2	0	ALT	1.000	2	7	0	0	1.4593	0.0000	0.0000	0.4760	0.2331	0.3343	0.6241	0.6858	0.4958	0.440	0.614	0.174
PBX1	27	0	0	27	11	1	6	1	5	3	0	3	2	1	1	EQUAL	0.500	14	42	2	1	1.4680	0.8853	0.8853	0.2062	0.8468	0.3487	0.3406	0.8939	0.5088	0.466	0.511	0.045
CSF3R	25	0	0	25	4	0	3	0	3	0	0	3	2	0	2	REF	0.000	19	59	3	2	1.4774	1.2980	1.2980	0.1308	0.6141	0.3190	0.2410	0.7338	0.4807	0.477	0.077	-0.400
BMPR1A	33	0	0	33	19	0	8	4	4	1	0	7	2	2	0	ALT	1.000	1	11	4	2	1.5167	6.2647	6.2647	0.2250	0.0013	0.0056	0.3631	0.4343	0.0228	0.450	0.790	0.340
TRIM33	38	0	0	38	17	0	6	3	3	0	0	6	2	2	0	ALT	1.000	16	65	7	0	1.6785	1.6251	1.6251	0.0542	0.3029	0.1529	0.1203	0.6858	0.2907	0.457	0.688	0.231
FGFR4	16	0	0	16	10	2	9	8	1	4	2	5	2	2	0	ALT	1.000	8	26	0	0	1.7313	0.0000	0.0000	0.1661	0.1111	0.0520	0.2893	0.6858	0.1299	0.497	0.741	0.244
ARNT	20	0	0	20	12	0	6	0	6	1	0	5	2	0	2	REF	0.000	9	39	4	2	1.7354	1.9488	1.9488	0.1242	0.2238	0.2657	0.2316	0.6858	0.4253	0.476	0.483	0.007
KLF4	9	0	0	9	4	0	3	3	0	0	0	3	2	2	0	ALT	1.000	2	15	0	0	1.7763	0.0000	0.0000	0.0594	0.3011	0.0896	0.1293	0.6858	0.1967	0.477	0.549	0.072

SMO	9	0	0	9	7	0	4	1	3	0	0	4	2	1	1	EQUAL	0.500	3	23	4	0	1.8493	4.1478	4.1478	0.0274	0.0212	0.0174	0.0694	0.6858	0.0555	0.481	0.275	-0.206
ASXL2	60	0	0	60	17	1	6	3	3	2	0	4	2	2	0	ALT	1.000	8	40	5	1	2.0438	3.0188	3.0188	0.0527	0.0535	0.0775	0.1175	0.6858	0.1765	0.486	0.778	0.292
CDC73	43	0	0	43	23	0	11	7	4	5	0	6	2	2	0	ALT	1.000	9	50	8	6	2.1617	3.5234	3.5234	0.0236	0.0032	0.0100	0.0617	0.6059	0.0360	0.464	0.525	0.060
POT1	16	0	0	16	11	0	4	3	1	2	0	2	2	1	1	EQUAL	0.500	4	22	2	2	2.5469	3.3832	3.3832	0.0717	0.1021	0.1397	0.1504	0.6858	0.2727	0.424	0.648	0.224
HNRNPA2B1	6	0	0	6	6	0	5	2	3	1	0	4	2	1	1	EQUAL	0.500	2	12	2	1	2.6800	4.0318	4.0318	0.1162	0.1008	0.1898	0.2200	0.6858	0.3381	0.490	0.476	-0.013
CDKN2A	13	0	0	13	10	5	8	8	0	5	5	3	2	2	0	ALT	1.000	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	0.471	0.663	0.192
CARS	13	0	0	13	12	0	3	2	1	0	0	3	2	2	0	ALT	1.000	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	0.458	0.635	0.176
CHEK2	18	0	0	18	12	1	4	3	1	1	1	3	1	1	0	ALT	1.000	0	0	0	0	0.0000	0.0000	0.0000	0.0002	0.1571	0.0003	0.0012	0.6858	0.0022	NA	NA	NA
MALT1	21	0	0	21	11	2	4	1	3	1	0	3	1	1	0	ALT	1.000	0	0	0	0	0.0000	0.0000	0.0000	0.0000	0.0945	0.0000	0.0001	0.6858	0.0001	NA	NA	NA
PLCG1	13	0	0	13	9	1	4	2	2	2	1	2	1	1	0	ALT	1.000	0	0	0	0	0.0000	0.0000	0.0000	0.0000	0.0282	0.0000	0.0000	0.6858	0.0000	NA	NA	NA
BRIP1	51	0	0	51	10	1	3	2	1	1	1	2	1	1	0	ALT	1.000	0	0	0	0	0.0000	0.0000	0.0000	0.0000	0.0444	0.0000	0.0000	0.6858	0.0000	NA	NA	NA
COL1A1	3	0	0	3	3	1	3	0	3	1	1	2	1	0	1	REF	0.000	0	0	0	0	0.0000	0.0000	0.0000	0.0000	0.0180	0.0000	0.0000	0.6858	0.0000	NA	NA	NA
MN1	33	0	0	33	3	0	3	0	3	0	0	3	1	0	1	REF	0.000	0	0	0	0	0.0000	0.0000	0.0000	0.0000	0.1596	0.0000	0.0000	0.6858	0.0000	NA	NA	NA
PRKACA	7	0	0	7	7	0	3	1	2	0	0	3	1	0	1	REF	0.000	0	0	0	0	0.0000	0.0000	0.0000	0.0010	0.2941	0.0026	0.0050	0.6858	0.0125	NA	NA	NA
RUNX1	18	0	0	18	6	0	3	1	2	1	0	2	1	0	1	REF	0.000	0	0	0	0	0.0000	0.0000	0.0000	0.0001	0.3049	0.0002	0.0006	0.6858	0.0017	NA	NA	NA
ZNRF3	29	0	0	29	10	0	3	1	2	0	0	3	1	1	0	ALT	1.000	0	0	0	0	0.0000	0.0000	0.0000	0.0000	0.1573	0.0000	0.0000	0.6858	0.0000	NA	NA	NA
MGMT	4	0	0	4	4	0	2	1	1	0	0	2	1	0	1	REF	0.000	0	0	0	0	0.0000	0.0000	0.0000	0.0063	0.3818	0.0162	0.0211	0.6858	0.0527	0.500	0.091	-0.409
TFPT	5	0	0	5	4	0	2	1	1	0	0	2	1	1	0	ALT	1.000	0	0	0	0	0.0000	0.0000	0.0000	0.0040	0.3412	0.0100	0.0147	0.6858	0.0360	NA	NA	NA
YWHAE	2	0	0	2	2	0	2	0	2	0	0	2	1	0	1	REF	0.000	1	0	0	0	0.0000	0.0000	0.0000	0.0073	0.3305	0.0170	0.0237	0.6858	0.0547	NA	NA	NA
ZNF521	43	0	0	43	3	1	2	0	2	0	0	2	1	0	1	REF	0.000	0	0	0	0	0.0000	0.0000	0.0000	0.0000	0.0874	0.0000	0.0000	0.6858	0.0000	NA	NA	NA
BCL3	2	0	0	2	2	0	1	1	0	0	0	1	1	1	0	ALT	1.000	0	0	0	0	0.0000	0.0000	0.0000	0.0001	0.3121	0.0002	0.0005	0.6858	0.0014	NA	NA	NA
CD209	19	0	0	19	1	0	1	0	1	0	0	1	1	0	1	REF	0.000	0	0	0	0	0.0000	0.0000	0.0000	0.0011	0.2399	0.0025	0.0054	0.6858	0.0121	NA	NA	NA
CD79A	2	0	0	2	2	0	1	0	1	0	0	1	1	0	1	REF	0.000	0	0	0	0	0.0000	0.0000	0.0000	0.0096	0.3905	0.0243	0.0299	0.6858	0.0719	NA	NA	NA

CD79B	8	0	0	8	1	0	1	0	1	0	0	1	1	0	1	REF	0.000	0	0	0	0	0.0000	0.0000	0.0000	0.0096	0.4005	0.0245	0.0297	0.6858	0.0725	NA	NA	NA
DDX5	2	0	0	2	2	0	1	0	1	0	0	1	1	0	1	REF	0.000	0	0	0	0	0.0000	0.0000	0.0000	0.0001	0.1700	0.0001	0.0005	0.6858	0.0010	NA	NA	NA
FSTL3	5	0	0	5	1	0	1	1	0	0	0	1	1	1	0	ALT	1.000	0	0	0	0	0.0000	0.0000	0.0000	0.0013	0.3868	0.0039	0.0061	0.6858	0.0173	NA	NA	NA
ISX	18	0	0	18	2	0	1	1	0	0	0	1	1	1	0	ALT	1.000	0	0	0	0	0.0000	0.0000	0.0000	0.0091	0.4211	0.0242	0.0286	0.6867	0.0718	NA	NA	NA
LZTR1	9	0	0	9	3	0	1	1	0	0	0	1	1	1	0	ALT	1.000	0	0	0	0	0.0000	0.0000	0.0000	0.0000	0.1093	0.0000	0.0000	0.6858	0.0000	NA	NA	NA
TSC2	26	0	0	26	20	1	6	3	3	2	1	4	1	1	0	ALT	1.000	1	2	7	0	0.0720	2.7121	2.7121	0.0000	0.0235	0.0000	0.0000	0.6858	0.0000	0.609	0.635	0.025
CBFA2T3	23	0	0	23	4	0	2	2	0	1	0	1	1	1	0	ALT	1.000	2	1	0	0	0.0857	0.0000	0.0000	0.0001	0.1718	0.0001	0.0005	0.6858	0.0010	NA	NA	NA
BAZ1A	44	0	0	44	28	1	16	4	12	7	1	9	1	1	0	ALT	1.000	3	2	0	1	0.0978	0.3255	0.3255	0.0000	0.1681	0.0000	0.0000	0.6858	0.0000	0.487	0.486	-0.001
CRTC3	8	0	0	8	4	0	2	1	1	0	0	2	1	1	0	ALT	1.000	1	1	0	0	0.1130	0.0000	0.0000	0.0008	0.1581	0.0013	0.0039	0.6858	0.0070	NA	NA	NA
FLCN	6	0	0	6	4	1	3	2	1	1	1	2	1	1	0	ALT	1.000	0	1	0	0	0.1131	0.0000	0.0000	0.0008	0.1858	0.0014	0.0039	0.6858	0.0078	NA	NA	NA
ZMYM2	26	0	0	26	20	1	7	3	4	1	1	6	1	1	0	ALT	1.000	4	2	2	0	0.1178	0.8286	0.8286	0.0000	0.7837	0.0000	0.0001	0.8473	0.0003	0.482	0.611	0.128
MLLT6	8	0	0	8	4	0	2	2	0	1	0	1	1	1	0	ALT	1.000	1	2	0	1	0.1233	0.7316	0.7316	0.0000	0.7417	0.0000	0.0001	0.8168	0.0005	0.385	0.204	-0.181
GAS7	20	0	0	20	6	1	3	1	2	1	1	2	1	1	0	ALT	1.000	1	1	0	1	0.1337	1.0719	1.0719	0.0028	0.9453	0.0115	0.0111	0.9640	0.0401	NA	NA	NA
STAT6	13	0	0	13	12	0	6	2	4	4	0	2	1	1	0	ALT	1.000	3	2	0	1	0.1733	0.7305	0.7305	0.0005	0.7403	0.0023	0.0029	0.8159	0.0111	0.504	0.730	0.227
CCR7	17	0	0	17	5	1	2	0	2	1	1	1	1	0	1	REF	0.000	1	1	0	0	0.1771	0.0000	0.0000	0.0157	0.4584	0.0412	0.0444	0.6916	0.1087	NA	NA	NA
CIITA	27	0	0	27	11	1	5	0	5	1	1	4	1	0	1	REF	0.000	0	3	0	0	0.1772	0.0000	0.0000	0.0000	0.0815	0.0000	0.0003	0.6858	0.0004	NA	NA	NA
FOXO1	20	0	0	20	7	0	4	1	3	1	0	3	1	0	1	REF	0.000	1	2	0	0	0.1882	0.0000	0.0000	0.0011	0.3167	0.0031	0.0055	0.6858	0.0142	NA	NA	NA
CCND1	5	0	0	5	4	0	1	0	1	0	0	1	1	0	1	REF	0.000	0	1	0	0	0.2084	0.0000	0.0000	0.0347	0.3929	0.0746	0.0840	0.6858	0.1715	NA	NA	NA
BLM	12	0	0	12	4	1	1	1	0	0	0	1	1	1	0	ALT	1.000	3	4	0	0	0.2240	0.0000	0.0000	0.0001	0.0510	0.0001	0.0006	0.6858	0.0006	NA	NA	NA
FANCC	17	0	0	17	4	0	1	0	1	0	0	1	1	0	1	REF	0.000	2	2	2	2	0.2707	3.6611	3.6611	0.0185	0.0398	0.0064	0.0507	0.6858	0.0251	0.482	0.414	-0.068
CDKN1B	5	0	0	5	5	2	3	1	2	1	1	2	1	0	1	REF	0.000	0	1	2	0	0.2994	7.1161	7.1161	0.1320	0.0357	0.0354	0.2428	0.6858	0.0966	0.538	0.601	0.062
COL2A1	31	0	0	31	2	0	1	0	1	0	0	1	1	0	1	REF	0.000	2	7	1	1	0.3142	0.6516	0.6516	0.0002	0.5142	0.0006	0.0011	0.6986	0.0039	NA	NA	NA
KMT2D	92	0	0	92	41	0	9	5	4	2	0	7	1	1	0	ALT	1.000	9	25	2	0	0.3214	0.2845	0.2845	0.0000	0.0249	0.0000	0.0000	0.6858	0.0000	0.478	0.513	0.035

NRG1	29	0	0	29	2	0	2	0	2	0	0	2	1	0	1	REF	0.000	3	3	2	1	0.3281	2.9510	2.9510	0.0180	0.1198	0.0172	0.0495	0.6858	0.0551	0.339	0.000	-0.339
BRCA2	82	0	0	82	7	1	2	1	1	1	0	1	1	1	0	ALT	1.000	1	14	2	0	0.3398	0.4679	0.4679	0.0000	0.2190	0.0000	0.0000	0.6858	0.0001	0.495	0.598	0.104
PAFAH1B2	9	0	0	9	6	0	1	1	0	0	0	1	1	1	0	ALT	1.000	0	1	0	0	0.3420	0.0000	0.0000	0.1920	0.3685	0.2850	0.3228	0.6858	0.4457	NA	NA	NA
SHTN1	20	0	0	20	12	3	4	3	1	0	0	4	1	1	0	ALT	1.000	0	3	0	1	0.3739	0.6769	0.6769	0.0417	0.6765	0.1153	0.0972	0.7722	0.2372	NA	NA	NA
PPFIBP1	32	0	0	32	24	0	9	1	8	3	0	6	1	1	0	ALT	1.000	1	5	1	0	0.3838	0.4548	0.4548	0.0109	0.3647	0.0260	0.0330	0.6858	0.0758	0.446	0.594	0.147
SMARCD1	7	0	0	7	5	0	3	1	2	1	0	2	1	1	0	ALT	1.000	0	3	0	0	0.3986	0.0000	0.0000	0.0601	0.1550	0.0621	0.1306	0.6858	0.1489	0.501	0.557	0.056
MEN1	12	0	0	12	9	1	5	2	3	2	1	3	1	1	0	ALT	1.000	0	4	0	0	0.4097	0.0000	0.0000	0.0362	0.2084	0.0505	0.0869	0.6858	0.1272	0.329	0.237	-0.092
RARA	15	0	0	15	13	1	6	4	2	2	1	4	1	1	0	ALT	1.000	0	3	0	0	0.4226	0.0000	0.0000	0.0818	0.2970	0.1276	0.1669	0.6858	0.2555	0.605	0.425	-0.180
CHIC2	10	0	0	10	8	1	2	1	1	0	0	2	1	1	0	ALT	1.000	1	1	0	0	0.4305	0.0000	0.0000	0.3271	0.3799	0.4208	0.4777	0.6858	0.5716	0.496	0.807	0.310
CNTRL	50	0	0	50	10	1	5	3	2	1	1	4	1	1	0	ALT	1.000	9	13	3	1	0.4507	0.8551	0.8551	0.0009	0.7478	0.0039	0.0046	0.8213	0.0173	0.483	0.619	0.136
NSD3	45	0	0	45	31	2	14	3	11	6	2	8	1	1	0	ALT	1.000	10	13	2	0	0.4687	0.5570	0.5570	0.0718	0.4244	0.1903	0.1505	0.6871	0.3386	0.488	0.493	0.005
RB1	42	0	0	42	23	4	4	1	3	2	1	2	1	1	0	ALT	1.000	1	6	1	5	0.5028	2.8729	2.8729	0.0573	0.0278	0.0146	0.1256	0.6858	0.0483	0.464	0.324	-0.140
TET2	35	0	0	35	4	1	1	1	0	0	0	1	1	1	0	ALT	1.000	8	13	4	1	0.5226	2.0818	2.0818	0.0087	0.1686	0.0110	0.0275	0.6858	0.0390	0.524	1.046	0.522
FANCE	14	0	0	14	7	0	3	1	2	1	0	2	1	1	0	ALT	1.000	0	4	0	1	0.5249	1.2311	1.2311	0.1489	0.8407	0.3457	0.2661	0.8891	0.5061	0.516	0.608	0.092
SRSF3	3	0	0	3	3	0	3	1	2	1	0	2	1	1	0	ALT	1.000	2	2	0	0	0.5651	0.0000	0.0000	0.5464	0.3843	0.5871	0.6866	0.6858	0.7033	0.455	0.518	0.063
HOXD11	4	0	0	4	1	0	1	1	0	0	0	1	1	1	0	ALT	1.000	1	4	0	0	0.5720	0.0000	0.0000	0.2180	0.4858	0.3673	0.3547	0.6962	0.5252	0.497	0.260	-0.237
RALGDS	22	0	0	22	21	1	9	5	4	1	1	8	1	1	0	ALT	1.000	3	8	2	1	0.5756	2.3864	2.3864	0.0853	0.1880	0.0957	0.1726	0.6858	0.2066	0.493	0.518	0.024
RSPO3	12	0	0	12	2	0	2	0	2	1	0	1	1	0	1	REF	0.000	1	2	1	0	0.5793	1.6386	1.6386	0.3958	0.6482	0.6283	0.5477	0.7535	0.7318	NA	NA	NA
JAK2	44	0	0	44	14	0	5	1	4	2	0	3	1	0	1	REF	0.000	9	15	3	1	0.5869	1.1088	1.1088	0.2290	0.8658	0.3438	0.3680	0.9072	0.5043	0.523	0.373	-0.150
FKBP9	14	0	0	14	14	0	10	2	8	1	0	9	1	0	1	REF	0.000	1	5	0	0	0.5903	0.0000	0.0000	0.1964	0.2220	0.2059	0.3286	0.6858	0.3581	0.480	0.368	-0.112
LEPROTL1	9	0	0	9	2	0	1	1	0	0	0	1	1	1	0	ALT	1.000	1	1	0	0	0.6210	0.0000	0.0000	0.6049	0.5340	0.7210	0.7327	0.7047	0.7963	0.522	0.574	0.051
HOOK3	20	0	0	20	6	0	4	1	3	3	0	1	1	1	0	ALT	1.000	10	16	0	0	0.6239	0.0000	0.0000	0.2616	0.0044	0.0174	0.4064	0.6332	0.0555	0.496	0.507	0.010
GATA2	11	0	0	11	6	1	4	1	3	1	1	3	1	0	1	REF	0.000	4	5	0	2	0.6466	4.2962	4.2962	0.2931	0.1073	0.1446	0.4414	0.6858	0.2795	0.481	0.529	0.047

FANCD2	26	0	0	26	12	2	4	1	3	1	1	3	1	1	0	ALT	1.000	9	15	4	0	0.6672	1.1511	1.1511	0.3388	0.8182	0.4996	0.4898	0.8727	0.6358	0.470	0.764	0.294
HMGA1	1	0	0	1	1	0	1	1	0	0	0	1	1	1	0	ALT	1.000	0	1	0	0	0.6686	0.0000	0.0000	0.6662	0.5102	0.7335	0.7795	0.6981	0.8053	0.464	0.900	0.436
CCND2	8	0	0	8	4	1	4	0	4	2	1	2	1	0	1	REF	0.000	0	3	0	0	0.6830	0.0000	0.0000	0.4807	0.3852	0.5349	0.6282	0.6858	0.6638	NA	NA	NA
JUN	3	0	0	3	3	0	3	1	2	1	0	2	1	1	0	ALT	1.000	7	7	0	0	0.6896	0.0000	0.0000	0.4981	0.3915	0.5874	0.6444	0.6858	0.7033	0.419	0.498	0.079
NACA	10	0	0	10	4	0	4	1	3	0	0	4	1	1	0	ALT	1.000	0	2	0	0	0.7116	0.0000	0.0000	0.6099	0.3477	0.5649	0.7366	0.6858	0.6864	0.430	0.434	0.003
IKZF1	29	0	0	29	5	0	5	0	5	2	0	3	1	0	1	REF	0.000	16	22	0	0	0.7261	0.0000	0.0000	0.3403	0.0281	0.0795	0.4915	0.6858	0.1796	0.548	0.000	-0.548
NBN	31	0	0	31	24	1	8	5	4	5	1	3	1	1	0	ALT	1.000	7	13	3	1	0.7371	1.5383	1.5383	0.5341	0.5071	0.4541	0.6761	0.6976	0.5993	0.542	0.605	0.063
KIT	62	0	0	62	8	1	4	2	2	2	1	2	1	1	0	ALT	1.000	10	17	4	2	0.7551	2.4271	2.4271	0.4950	0.1088	0.0844	0.6419	0.6858	0.1881	0.373	0.000	-0.373
NCKIPSD	11	0	0	11	8	1	5	4	1	1	1	4	1	1	0	ALT	1.000	2	9	0	1	0.8145	0.9734	0.9734	0.5239	0.9784	0.8159	0.6675	0.9866	0.8644	0.499	0.556	0.057
POU2AF1	8	0	0	8	2	0	2	0	2	1	0	1	1	0	1	REF	0.000	0	3	0	0	0.8273	0.0000	0.0000	0.7346	0.3866	0.6490	0.8299	0.6858	0.7462	NA	NA	NA
CCDC6	11	0	0	11	7	0	3	3	0	1	0	2	1	1	0	ALT	1.000	3	6	1	0	0.8430	1.1149	1.1149	0.6667	0.9149	0.9063	0.7800	0.9427	0.9304	0.516	0.604	0.089
CHCHD7	3	0	0	3	2	0	1	1	0	0	0	1	1	1	0	ALT	1.000	0	1	0	0	0.8452	0.0000	0.0000	0.8627	0.6019	0.8598	0.9204	0.7271	0.8969	NA	NA	NA
NTRK3	99	0	0	99	1	0	1	1	0	0	0	1	1	1	0	ALT	1.000	4	10	0	0	0.8607	0.0000	0.0000	0.6265	0.1175	0.2609	0.7504	0.6858	0.4203	NA	NA	NA
TNFRSF14	3	0	0	3	2	0	2	0	2	0	0	2	1	0	1	REF	0.000	11	16	0	0	0.8682	0.0000	0.0000	0.7237	0.0670	0.1866	0.8229	0.6858	0.3340	0.475	0.478	0.003
DNMT3A	19	0	0	19	14	1	4	1	3	1	0	3	1	1	0	ALT	1.000	16	31	5	1	0.8868	1.6251	1.6251	0.7017	0.3314	0.4435	0.8055	0.6858	0.5906	0.474	0.627	0.153
TPM3	6	0	0	6	4	0	4	1	3	1	0	3	1	1	0	ALT	1.000	7	14	2	0	0.9056	0.7144	0.7144	0.8360	0.6690	0.9125	0.9029	0.7671	0.9346	0.467	0.629	0.162
N4BP2	48	0	0	48	8	1	1	0	1	0	0	1	1	0	1	REF	0.000	7	20	3	0	0.9113	1.3282	1.3282	0.6731	0.6388	0.8194	0.7842	0.7472	0.8672	0.489	0.474	-0.015
GNAQ	19	0	0	19	15	0	6	4	2	1	0	5	1	1	0	ALT	1.000	0	5	0	0	0.9611	0.0000	0.0000	0.9288	0.2586	0.5262	0.9609	0.6858	0.6563	0.473	0.383	-0.090
PTPRD	143	0	0	143	11	4	4	3	1	3	1	1	1	1	0	ALT	1.000	36	97	6	1	0.9869	0.6008	0.6008	0.9463	0.1931	0.3791	0.9710	0.6858	0.5353	0.497	0.798	0.301
CDK6	16	0	0	16	2	0	2	1	1	0	0	2	1	1	0	ALT	1.000	1	5	0	0	0.9973	0.0000	0.0000	0.9951	0.3124	0.6003	0.9973	0.6858	0.7119	0.328	0.870	0.542
ATR	69	0	0	69	24	1	8	3	5	4	1	4	1	1	0	ALT	1.000	13	39	5	3	1.0218	1.6059	1.6059	0.9470	0.3061	0.5222	0.9712	0.6858	0.6525	0.514	0.695	0.182
EML4	21	0	0	21	13	1	3	1	2	0	0	3	1	0	1	REF	0.000	5	14	1	2	1.0983	1.5320	1.5320	0.8154	0.5441	0.8302	0.8881	0.7067	0.8750	0.447	0.491	0.043
MLLT10	34	0	0	34	17	2	2	2	0	1	1	1	1	1	0	ALT	1.000	7	19	2	0	1.1024	0.8375	0.8375	0.8292	0.8239	0.9184	0.8977	0.8767	0.9384	0.446	0.664	0.218

PRDM2	39	0	0	39	20	0	8	1	7	1	0	7	1	0	1	REF	0.000	34	101	5	0	1.1209	0.6096	0.6096	0.5636	0.2729	0.3196	0.7004	0.6858	0.4812	0.482	0.588	0.106
GOPC	16	0	0	16	11	1	7	3	4	3	1	4	1	1	0	ALT	1.000	3	7	0	0	1.1455	0.0000	0.0000	0.8045	0.2113	0.4304	0.8802	0.6858	0.5796	0.403	0.601	0.198
MYB	8	0	0	8	3	0	2	2	0	0	0	2	1	1	0	ALT	1.000	5	13	1	0	1.1470	0.6546	0.6546	0.8006	0.6893	0.8318	0.8770	0.7808	0.8760	0.488	1.006	0.517
FCGR2B	17	0	0	17	8	0	8	0	8	1	0	7	1	0	1	REF	0.000	11	24	1	2	1.1555	1.3577	1.3577	0.6939	0.6502	0.8751	0.8001	0.7549	0.9075	0.475	0.000	-0.475
CACNA1D	108	0	0	108	12	2	5	4	1	3	2	2	1	1	0	ALT	1.000	11	37	4	1	1.1963	1.3784	1.3784	0.3759	0.4957	0.6007	0.5286	0.6964	0.7121	0.481	1.178	0.698
LIFR	45	0	0	45	8	0	5	1	4	4	0	1	1	0	1	REF	0.000	7	23	2	0	1.2036	0.7931	0.7931	0.6721	0.7702	0.7903	0.7834	0.8377	0.8454	0.466	0.570	0.104
ETV5	14	0	0	14	7	0	2	1	1	0	0	2	1	1	0	ALT	1.000	4	10	0	1	1.2279	0.8599	0.8599	0.7382	0.8928	0.9036	0.8326	0.9269	0.9280	0.523	0.595	0.072
TCEA1	9	0	0	9	9	0	4	0	4	1	0	3	1	0	1	REF	0.000	1	5	0	1	1.2380	1.5047	1.5047	0.6448	0.7020	0.8362	0.7648	0.7891	0.8794	0.406	0.455	0.049
IKBKB	11	0	0	11	6	1	3	2	1	1	1	2	1	1	0	ALT	1.000	5	13	1	1	1.2403	1.2056	1.2056	0.6899	0.8282	0.9227	0.7974	0.8797	0.9411	0.437	0.464	0.027
TEC	8	0	0	8	3	0	1	1	0	0	0	1	1	1	0	ALT	1.000	2	10	4	0	1.2598	3.0449	3.0449	0.4953	0.0724	0.1991	0.6421	0.6858	0.3496	0.502	0.487	-0.015
CTNNA2	108	0	0	108	7	0	1	0	1	0	0	1	1	0	1	REF	0.000	34	102	4	3	1.2752	0.7705	0.7705	0.2134	0.5189	0.2221	0.3488	0.7004	0.3774	0.534	0.841	0.306
CDKN2C	12	0	0	12	10	0	2	1	1	0	0	2	1	1	0	ALT	1.000	7	14	1	0	1.3148	1.5034	1.5034	0.5591	0.7182	0.8237	0.6962	0.8000	0.8703	0.530	0.482	-0.048
CASP9	11	0	0	11	9	2	3	3	0	1	1	2	1	1	0	ALT	1.000	9	23	1	1	1.3415	1.1865	1.1865	0.4555	0.8308	0.7561	0.6047	0.8818	0.8206	0.495	0.771	0.275
EPS15	15	0	0	15	9	1	5	2	3	1	1	4	1	1	0	ALT	1.000	9	31	1	0	1.3801	0.3106	0.3106	0.3924	0.1991	0.1426	0.5447	0.6858	0.2767	0.503	0.462	-0.041
CREB3L2	17	0	0	17	10	0	4	1	3	1	0	3	1	1	0	ALT	1.000	0	10	0	1	1.4119	1.2026	1.2026	0.3023	0.8579	0.5781	0.4514	0.9021	0.6962	0.421	0.580	0.159
EPHA3	116	0	0	116	11	1	7	1	6	3	0	4	1	1	0	ALT	1.000	18	65	5	4	1.4174	1.6316	1.6316	0.1815	0.2469	0.3373	0.3095	0.6858	0.4985	0.435	0.228	-0.207
CREB1	4	0	0	4	3	0	1	1	0	0	0	1	1	1	0	ALT	1.000	0	6	0	0	1.4422	0.0000	0.0000	0.3979	0.2744	0.3849	0.5499	0.6858	0.5411	0.437	0.580	0.143
ROS1	86	0	0	86	8	1	7	3	4	5	1	2	1	1	0	ALT	1.000	11	44	0	1	1.4559	0.2655	0.2655	0.2571	0.1051	0.0536	0.4013	0.6858	0.1328	0.553	0.802	0.249
STRN	24	0	0	24	11	2	2	2	0	1	1	1	1	1	0	ALT	1.000	5	18	1	3	1.5960	2.5772	2.5772	0.3555	0.1805	0.3895	0.5084	0.6858	0.5453	0.475	0.597	0.122
STAG1	39	0	0	39	18	0	4	1	3	0	0	4	1	1	0	ALT	1.000	3	26	1	1	1.6108	0.7676	0.7676	0.0855	0.6956	0.1635	0.1728	0.7851	0.3050	0.469	0.530	0.061
BARD1	15	0	0	15	6	0	1	1	0	0	0	1	1	1	0	ALT	1.000	5	19	1	0	1.6320	0.8263	0.8263	0.3287	0.8604	0.5161	0.4795	0.9037	0.6478	0.547	0.955	0.409
HOXA9	3	0	0	3	2	0	1	0	1	0	0	1	1	0	1	REF	0.000	3	8	1	0	1.6431	4.2330	4.2330	0.3835	0.2548	0.4439	0.5360	0.6858	0.5911	NA	NA	NA
SDHB	4	0	0	4	4	0	2	1	1	1	0	1	1	1	0	ALT	1.000	4	13	0	0	1.7096	0.0000	0.0000	0.3523	0.1684	0.1378	0.5047	0.6858	0.2699	0.466	0.641	0.175

MITF	30	0	0	30	4	0	2	1	1	0	0	2	1	1	0	ALT	1.000	3	13	2	0	1.7156	2.1356	2.1356	0.2137	0.3420	0.4026	0.3493	0.6858	0.5573	0.482	0.446	-0.037
PMS1	29	0	0	29	18	1	6	3	3	2	0	4	1	1	0	ALT	1.000	4	20	1	0	1.7281	0.7107	0.7107	0.2381	0.7173	0.3494	0.3785	0.7993	0.5095	0.493	0.563	0.070
PRRX1	7	0	0	7	4	0	3	0	3	1	0	2	1	0	1	REF	0.000	9	27	2	0	1.7847	1.5718	1.5718	0.1238	0.5835	0.3054	0.2311	0.7189	0.4665	0.502	0.165	-0.337
LPP	28	0	0	28	10	0	3	1	2	1	0	2	1	1	0	ALT	1.000	3	15	1	0	1.8139	1.2611	1.2611	0.1547	0.8233	0.3585	0.2738	0.8764	0.5170	0.594	0.702	0.108
EIF3E	17	0	0	17	16	0	9	2	7	4	0	5	1	1	0	ALT	1.000	0	11	2	0	1.8738	2.5564	2.5564	0.0881	0.2508	0.1678	0.1772	0.6858	0.3101	0.438	0.479	0.041
ACSL6	25	0	0	25	1	0	1	1	0	0	0	1	1	1	0	ALT	1.000	2	18	1	0	1.8969	0.7752	0.7752	0.0519	0.7903	0.1217	0.1161	0.8520	0.2469	0.507	1.033	0.526
BMP5	44	0	0	44	8	1	4	2	2	2	1	2	1	0	1	REF	0.000	1	12	3	0	1.9440	4.3799	4.3799	0.0935	0.0404	0.0756	0.1855	0.6858	0.1732	0.535	0.049	-0.486
CDKN1A	11	0	0	11	8	1	3	1	2	1	1	2	1	0	1	REF	0.000	2	9	0	0	3.1871	0.0000	0.0000	0.1325	0.4678	0.1733	0.2435	0.6927	0.3172	0.425	0.645	0.221
MYCL	13	0	0	13	5	1	3	0	3	1	1	2	1	0	1	REF	0.000	1	23	1	0	3.4855	2.0389	2.0389	0.0007	0.5241	0.0030	0.0035	0.7015	0.0140	0.453	0.568	0.114
RPL22	5	0	0	5	5	0	4	2	2	2	0	2	1	1	0	ALT	1.000	1	15	1	2	8.2537	13.8984	13.8984	0.0011	0.0050	0.0019	0.0054	0.6563	0.0098	0.472	0.848	0.376
NRAS	12	0	0	12	12	1	3	2	1	1	1	2	1	1	0	ALT	1.000	3	90	1	1	19.6954	3.7170	3.7170	0.0000	0.1913	0.0000	0.0000	0.6858	0.0000	0.484	0.548	0.064
CEP89	15	0	0	15	9	1	5	1	4	2	1	3	0	0	0	EQUAL	NA	0	0	0	0	0.0000	0.0000	0.0000	0.0000	0.0753	0.0000	0.0001	0.6858	0.0001	NA	NA	NA
DNAJB1	8	0	0	8	8	1	5	2	3	3	1	2	0	0	0	EQUAL	NA	0	0	0	0	0.0000	0.0000	0.0000	0.0008	0.3853	0.0024	0.0039	0.6858	0.0116	NA	NA	NA
MSI2	7	0	0	7	5	0	3	1	2	3	0	0	0	0	0	EQUAL	NA	0	0	0	0	0.0000	0.0000	0.0000	0.0016	0.2616	0.0037	0.0071	0.6858	0.0163	NA	NA	NA
PPM1D	10	0	0	10	8	1	3	2	1	1	0	2	0	0	0	EQUAL	NA	0	0	0	0	0.0000	0.0000	0.0000	0.0000	0.1723	0.0000	0.0002	0.6858	0.0005	NA	NA	NA
AKT2	10	0	0	10	9	0	2	0	2	1	0	1	0	0	0	EQUAL	NA	0	0	0	0	0.0000	0.0000	0.0000	0.0001	0.2178	0.0002	0.0007	0.6858	0.0018	NA	NA	NA
BCL2L12	4	0	0	4	2	0	2	0	2	1	0	1	0	0	0	EQUAL	NA	0	0	0	0	0.0000	0.0000	0.0000	0.0009	0.2856	0.0022	0.0043	0.6858	0.0109	NA	NA	NA
SH3GL1	5	0	0	5	5	0	2	2	0	0	0	2	0	0	0	EQUAL	NA	0	0	0	0	0.0000	0.0000	0.0000	0.0005	0.2624	0.0012	0.0028	0.6858	0.0069	NA	NA	NA
BCL2	7	0	0	7	1	0	1	0	1	0	0	1	0	0	0	EQUAL	NA	0	0	0	0	0.0000	0.0000	0.0000	0.0036	0.5729	0.0124	0.0137	0.7145	0.0427	NA	NA	NA
BCL7A	10	0	0	10	5	0	1	0	1	1	0	0	0	0	0	EQUAL	NA	0	0	1	0	0.0000	2.2628	2.2628	0.0102	0.4721	0.0285	0.0313	0.6944	0.0816	NA	NA	NA
CANT1	7	0	0	7	5	0	1	1	0	0	0	1	0	0	0	EQUAL	NA	0	0	0	0	0.0000	0.0000	0.0000	0.0002	0.3969	0.0008	0.0015	0.6858	0.0046	NA	NA	NA
CLTCL1	36	0	0	36	1	0	1	1	0	0	0	1	0	0	0	EQUAL	NA	0	0	0	0	0.0000	0.0000	0.0000	0.0000	0.0265	0.0000	0.0000	0.6858	0.0000	NA	NA	NA
HERPUD1	1	0	0	1	1	0	1	1	0	1	0	0	0	0	0	EQUAL	NA	0	0	0	0	0.0000	0.0000	0.0000	0.0009	0.2830	0.0023	0.0045	0.6858	0.0112	0.420	0.553	0.133

HLF	12	0	0	12	1	0	1	0	1	1	0	0	0	0	0	EQUAL	NA	0	0	0	0	0.0000	0.0000	0.0000	0.0031	0.4552	0.0094	0.0119	0.6915	0.0343	NA	NA	NA
HOXC13	16	0	0	16	1	0	1	0	1	0	0	1	0	0	0	EQUAL	NA	0	0	0	0	0.0000	0.0000	0.0000	0.0004	0.4598	0.0014	0.0023	0.6916	0.0077	NA	NA	NA
KCNJ5	36	0	0	36	1	0	1	0	1	0	0	1	0	0	0	EQUAL	NA	0	0	0	0	0.0000	0.0000	0.0000	0.0005	0.3955	0.0015	0.0027	0.6858	0.0081	NA	NA	NA
RABEP1	5	0	0	5	3	0	1	0	1	0	0	1	0	0	0	EQUAL	NA	0	0	0	0	0.0000	0.0000	0.0000	0.0000	0.0592	0.0000	0.0000	0.6858	0.0000	NA	NA	NA
SRSF2	2	0	0	2	2	0	1	1	0	0	0	1	0	0	0	EQUAL	NA	0	0	0	0	0.0000	0.0000	0.0000	0.0013	0.4694	0.0045	0.0062	0.6929	0.0191	NA	NA	NA
SS18	6	0	0	6	5	0	1	1	0	0	0	1	0	0	0	EQUAL	NA	0	0	0	0	0.0000	0.0000	0.0000	0.0010	0.1928	0.0019	0.0049	0.6858	0.0098	NA	NA	NA
STAT5B	6	0	0	6	5	0	1	0	1	0	0	1	0	0	0	EQUAL	NA	2	0	0	0	0.0000	0.0000	0.0000	0.0000	0.1056	0.0000	0.0000	0.6858	0.0001	0.467	0.521	0.054
STK11	5	0	0	5	2	1	1	1	0	1	1	0	0	0	0	EQUAL	NA	0	0	0	0	0.0000	0.0000	0.0000	0.0002	0.2684	0.0005	0.0012	0.6858	0.0032	NA	NA	NA
AMER1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	EQUAL	NA	0	0	0	0	0.0000	0.0000	0.0000	0.0000	0.0956	0.0000	0.0000	0.6858	0.0000	NA	NA	NA
AR	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	EQUAL	NA	0	0	0	0	0.0000	0.0000	0.0000	0.0000	0.1390	0.0000	0.0000	0.6858	0.0000	NA	NA	NA
ARAF	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	EQUAL	NA	0	0	0	0	0.0000	0.0000	0.0000	0.0000	0.1693	0.0000	0.0002	0.6858	0.0004	NA	NA	NA
ATP2B3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	EQUAL	NA	0	0	0	0	0.0000	0.0000	0.0000	0.0000	0.0818	0.0000	0.0000	0.6858	0.0000	NA	NA	NA
ATRX	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	EQUAL	NA	0	0	0	0	0.0000	0.0000	0.0000	0.0000	0.0027	0.0000	0.0000	0.5694	0.0000	NA	NA	NA
BCOR	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	EQUAL	NA	0	0	0	0	0.0000	0.0000	0.0000	0.0000	0.0479	0.0000	0.0000	0.6858	0.0000	NA	NA	NA
BCORL1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	EQUAL	NA	0	0	0	0	0.0000	0.0000	0.0000	0.0000	0.0445	0.0000	0.0000	0.6858	0.0000	NA	NA	NA
BTK	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	EQUAL	NA	0	0	0	0	0.0000	0.0000	0.0000	0.0000	0.1185	0.0001	0.0003	0.6858	0.0006	NA	NA	NA
C15orf65	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	EQUAL	NA	0	0	0	0	0.0000	0.0000	0.0000	0.0914	0.6153	0.2120	0.1823	0.7346	0.3662	NA	NA	NA
CEBPA	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	EQUAL	NA	0	0	0	0	0.0000	0.0000	0.0000	0.0001	0.5475	0.0004	0.0007	0.7080	0.0026	NA	NA	NA
DCAF12L2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	EQUAL	NA	0	0	0	0	0.0000	0.0000	0.0000	0.0001	0.3954	0.0003	0.0007	0.6858	0.0022	NA	NA	NA
DCC	106	0	0	106	0	0	0	0	0	0	0	0	0	0	0	EQUAL	NA	0	0	0	0	0.0000	0.0000	0.0000	0.0000	0.0260	0.0000	0.0000	0.6858	0.0000	NA	NA	NA
DDIT3	1	0	0	1	0	0	0	0	0	0	0	0	0	0	0	EQUAL	NA	0	0	0	0	0.0000	0.0000	0.0000	0.0307	0.4528	0.0730	0.0761	0.6913	0.1686	NA	NA	NA
DDX3X	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	EQUAL	NA	0	0	0	0	0.0000	0.0000	0.0000	0.0000	0.1333	0.0000	0.0001	0.6858	0.0002	NA	NA	NA
EIF1AX	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	EQUAL	NA	0	0	0	0	0.0000	0.0000	0.0000	0.0501	0.3932	0.1019	0.1127	0.6858	0.2162	NA	NA	NA

ELF4	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	EQUAL	NA	0	0	0	0	0.0000	0.0000	0.0000	0.0000	0.2453	0.0001	0.0003	0.6858	0.0007	NA	NA	NA
FAM47C	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	EQUAL	NA	0	0	0	0	0.0000	0.0000	0.0000	0.0000	0.1899	0.0000	0.0000	0.6858	0.0000	NA	NA	NA
FLNA	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	EQUAL	NA	0	0	0	0	0.0000	0.0000	0.0000	0.0000	0.0131	0.0000	0.0000	0.6858	0.0000	NA	NA	NA
FOXO4	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	EQUAL	NA	0	0	0	0	0.0000	0.0000	0.0000	0.0001	0.3114	0.0003	0.0008	0.6858	0.0024	NA	NA	NA
GATA1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	EQUAL	NA	0	0	0	0	0.0000	0.0000	0.0000	0.0007	0.3460	0.0021	0.0037	0.6858	0.0103	NA	NA	NA
GPC3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	EQUAL	NA	0	0	0	0	0.0000	0.0000	0.0000	0.0001	0.2223	0.0002	0.0007	0.6858	0.0018	NA	NA	NA
GPC5	53	0	0	53	0	0	0	0	0	0	0	0	0	0	0	EQUAL	NA	0	0	0	0	0.0000	0.0000	0.0000	0.0001	0.2025	0.0002	0.0006	0.6858	0.0015	NA	NA	NA
HMGA2	1	0	0	1	0	0	0	0	0	0	0	0	0	0	0	EQUAL	NA	0	0	0	0	0.0000	0.0000	0.0000	0.0634	0.4568	0.1353	0.1363	0.6915	0.2666	NA	NA	NA
IRS4	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	EQUAL	NA	0	0	0	0	0.0000	0.0000	0.0000	0.0000	0.1123	0.0000	0.0000	0.6858	0.0000	NA	NA	NA
KDM5C	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	EQUAL	NA	0	0	0	0	0.0000	0.0000	0.0000	0.0000	0.0287	0.0000	0.0000	0.6858	0.0000	NA	NA	NA
KDM6A	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	EQUAL	NA	0	0	0	0	0.0000	0.0000	0.0000	0.0000	0.0264	0.0000	0.0000	0.6858	0.0000	NA	NA	NA
KLK2	3	0	0	3	0	0	0	0	0	0	0	0	0	0	0	EQUAL	NA	0	0	0	0	0.0000	0.0000	0.0000	0.0092	0.4080	0.0239	0.0288	0.6858	0.0713	NA	NA	NA
LMO1	6	0	0	6	0	0	0	0	0	0	0	0	0	0	0	EQUAL	NA	0	0	0	0	0.0000	0.0000	0.0000	0.0305	0.5165	0.0780	0.0758	0.6996	0.1772	NA	NA	NA
LYL1	1	0	0	1	0	0	0	0	0	0	0	0	0	0	0	EQUAL	NA	0	0	0	0	0.0000	0.0000	0.0000	0.0017	0.5034	0.0057	0.0073	0.6976	0.0230	NA	NA	NA
MAP3K13	1	0	0	1	1	0	0	0	0	0	0	0	0	0	0	EQUAL	NA	0	0	0	0	0.0000	0.0000	0.0000	0.0000	0.0737	0.0000	0.0000	0.6858	0.0000	NA	NA	NA
MED12	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	EQUAL	NA	0	0	0	0	0.0000	0.0000	0.0000	0.0000	0.0067	0.0000	0.0000	0.6858	0.0000	NA	NA	NA
MSN	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	EQUAL	NA	0	0	0	0	0.0000	0.0000	0.0000	0.0000	0.1477	0.0001	0.0004	0.6858	0.0008	NA	NA	NA
MTCP1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	EQUAL	NA	0	0	0	0	0.0000	0.0000	0.0000	0.0652	0.5219	0.1488	0.1394	0.7010	0.2853	NA	NA	NA
NAB2	1	0	0	1	1	0	0	0	0	0	0	0	0	0	0	EQUAL	NA	1	0	0	0	0.0000	0.0000	0.0000	0.0001	0.2563	0.0002	0.0005	0.6858	0.0014	NA	NA	NA
NONO	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	EQUAL	NA	0	0	0	0	0.0000	0.0000	0.0000	0.0002	0.1517	0.0003	0.0013	0.6858	0.0025	NA	NA	NA
OLIG2	6	0	0	6	0	0	0	0	0	0	0	0	0	0	0	EQUAL	NA	0	0	0	0	0.0000	0.0000	0.0000	0.0003	0.5608	0.0012	0.0018	0.7119	0.0066	NA	NA	NA
PHF6	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	EQUAL	NA	0	0	0	0	0.0000	0.0000	0.0000	0.0028	0.1950	0.0050	0.0111	0.6858	0.0207	NA	NA	NA
POU5F1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	EQUAL	NA	1	0	0	0	0.0000	0.0000	0.0000	0.0013	0.2980	0.0033	0.0060	0.6858	0.0150	NA	NA	NA

PTPRT	139	0	0	139	1	0	0	0	0	0	0	0	0	0	0	EQUAL	NA	0	0	0	0	0.0000	0.0000	0.0000	0.0000	0.0317	0.0000	0.0000	0.6858	0.0000	NA	NA	NA
RBM10	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	EQUAL	NA	0	0	0	0	0.0000	0.0000	0.0000	0.0000	0.0638	0.0000	0.0000	0.6858	0.0000	NA	NA	NA
RMI2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	EQUAL	NA	0	0	0	0	0.0000	0.0000	0.0000	0.0150	0.6164	0.0458	0.0428	0.7353	0.1178	NA	NA	NA
RPL10	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	EQUAL	NA	0	0	0	0	0.0000	0.0000	0.0000	0.0103	0.4182	0.0269	0.0316	0.6863	0.0779	NA	NA	NA
SALL4	38	0	0	38	2	0	0	0	0	0	0	0	0	0	0	EQUAL	NA	0	0	0	0	0.0000	0.0000	0.0000	0.0000	0.1840	0.0000	0.0000	0.6858	0.0000	NA	NA	NA
SETBP1	11	0	0	11	0	0	0	0	0	0	0	0	0	0	0	EQUAL	NA	0	0	0	0	0.0000	0.0000	0.0000	0.0000	0.0895	0.0000	0.0000	0.6858	0.0000	NA	NA	NA
SMC1A	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	EQUAL	NA	0	0	0	0	0.0000	0.0000	0.0000	0.0000	0.0342	0.0000	0.0000	0.6858	0.0000	NA	NA	NA
SOCS1	3	0	0	3	1	0	0	0	0	0	0	0	0	0	0	EQUAL	NA	0	0	0	0	0.0000	0.0000	0.0000	0.0018	0.6361	0.0068	0.0078	0.7455	0.0266	0.667	0.976	0.309
SS18L1	9	0	0	9	3	0	0	0	0	0	0	0	0	0	0	EQUAL	NA	0	0	0	0	0.0000	0.0000	0.0000	0.0004	0.2112	0.0008	0.0021	0.6858	0.0047	NA	NA	NA
SSX1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	EQUAL	NA	0	0	0	0	0.0000	0.0000	0.0000	0.0238	0.4037	0.0549	0.0621	0.6858	0.1351	NA	NA	NA
SSX2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	EQUAL	NA	0	0	0	0	0.0000	0.0000	0.0000	0.0186	0.4080	0.0444	0.0508	0.6858	0.1150	NA	NA	NA
SSX4	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	EQUAL	NA	0	0	0	0	0.0000	0.0000	0.0000	0.0189	0.3820	0.0434	0.0516	0.6858	0.1132	NA	NA	NA
STAG2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	EQUAL	NA	0	0	0	0	0.0000	0.0000	0.0000	0.0000	0.0232	0.0000	0.0000	0.6858	0.0000	NA	NA	NA
TCL1A	2	0	0	2	0	0	0	0	0	0	0	0	0	0	0	EQUAL	NA	0	0	0	0	0.0000	0.0000	0.0000	0.0519	0.5187	0.1228	0.1161	0.7004	0.2485	NA	NA	NA
TFE3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	EQUAL	NA	0	0	0	0	0.0000	0.0000	0.0000	0.0000	0.1974	0.0001	0.0003	0.6858	0.0006	NA	NA	NA
TNFRSF17	3	0	0	3	1	0	0	0	0	0	0	0	0	0	0	EQUAL	NA	0	0	0	0	0.0000	0.0000	0.0000	0.0254	0.4964	0.0652	0.0654	0.6967	0.1547	NA	NA	NA
U2AF1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	EQUAL	NA	0	0	0	0	0.0000	0.0000	0.0000	0.0029	0.3145	0.0072	0.0115	0.6858	0.0276	NA	NA	NA
WAS	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	EQUAL	NA	0	0	0	0	0.0000	0.0000	0.0000	0.0001	0.1763	0.0002	0.0009	0.6858	0.0019	NA	NA	NA
ZMYM3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	EQUAL	NA	0	0	0	0	0.0000	0.0000	0.0000	0.0000	0.0404	0.0000	0.0000	0.6858	0.0000	NA	NA	NA
ZNF429	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	EQUAL	NA	0	0	0	0	0.0000	0.0000	0.0000	0.0001	0.1724	0.0001	0.0005	0.6858	0.0011	NA	NA	NA
ZRSR2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	EQUAL	NA	0	0	0	0	0.0000	0.0000	0.0000	0.0001	0.1301	0.0002	0.0010	0.6858	0.0018	NA	NA	NA
PALB2	8	0	0	8	2	0	0	0	0	0	0	0	0	0	0	EQUAL	NA	0	1	0	0	0.0681	0.0000	0.0000	0.0000	0.0748	0.0000	0.0000	0.6858	0.0001	0.555	0.760	0.205
CBL	37	0	0	37	6	1	1	1	0	1	0	0	0	0	0	EQUAL	NA	0	1	0	0	0.0809	0.0000	0.0000	0.0000	0.0942	0.0000	0.0003	0.6858	0.0004	0.664	0.362	-0.302

MAML2	48	0	0	48	14	1	3	1	2	1	1	2	0	0	0	EQUAL	NA	10	2	0	0	0.0832	0.0000	0.0000	0.0000	0.0443	0.0000	0.0001	0.6858	0.0002	0.565	0.325	-0.240
TSHR	31	0	0	31	1	0	1	1	0	0	0	1	0	0	0	EQUAL	NA	1	1	0	0	0.0986	0.0000	0.0000	0.0002	0.2113	0.0005	0.0015	0.6858	0.0033	NA	NA	NA
ZCCHC8	9	0	0	9	7	0	4	0	4	0	0	4	0	0	0	EQUAL	NA	1	1	0	0	0.1013	0.0000	0.0000	0.0003	0.1119	0.0004	0.0018	0.6858	0.0027	NA	NA	NA
POLG	9	0	0	9	7	1	2	2	0	1	1	1	0	0	0	EQUAL	NA	0	2	1	0	0.1085	0.5266	0.5266	0.0000	0.4728	0.0000	0.0000	0.6944	0.0001	0.510	0.526	0.015
FOXA1	2	0	0	2	0	0	0	0	0	0	0	0	0	0	0	EQUAL	NA	1	1	0	0	0.1174	0.0000	0.0000	0.0010	0.4403	0.0034	0.0050	0.6901	0.0155	NA	NA	NA
ETV1	20	0	0	20	2	0	2	0	2	0	0	2	0	0	0	EQUAL	NA	7	2	0	0	0.1188	0.0000	0.0000	0.0033	0.0499	0.0055	0.0126	0.6858	0.0226	0.429	0.621	0.193
CAMTA1	2	0	0	2	2	1	1	0	1	1	1	0	0	0	0	EQUAL	NA	3	3	1	0	0.1206	0.4482	0.4482	0.0000	0.3545	0.0000	0.0000	0.6858	0.0000	0.549	0.729	0.180
BRCA1	32	0	0	32	6	0	3	1	2	1	0	2	0	0	0	EQUAL	NA	2	3	0	0	0.1312	0.0000	0.0000	0.0000	0.0218	0.0000	0.0000	0.6858	0.0000	0.533	0.967	0.434
USP8	18	0	0	18	11	0	2	1	1	0	0	2	0	0	0	EQUAL	NA	1	2	0	0	0.1410	0.0000	0.0000	0.0000	0.0480	0.0000	0.0005	0.6858	0.0004	0.407	1.087	0.680
KNL1	42	0	0	42	9	2	4	1	3	3	1	1	0	0	0	EQUAL	NA	1	4	0	0	0.1419	0.0000	0.0000	0.0000	0.0128	0.0000	0.0000	0.6858	0.0000	0.461	0.776	0.314
BUB1B	17	0	0	17	13	1	1	1	0	1	1	0	0	0	0	EQUAL	NA	0	2	0	0	0.1488	0.0000	0.0000	0.0001	0.0412	0.0001	0.0008	0.6858	0.0006	0.472	0.447	-0.025
PRPF40B	14	0	0	14	7	1	4	1	3	1	1	3	0	0	0	EQUAL	NA	0	2	0	0	0.1573	0.0000	0.0000	0.0002	0.0627	0.0002	0.0012	0.6858	0.0014	0.521	0.393	-0.128
FLT3	25	0	0	25	0	0	0	0	0	0	0	0	0	0	0	EQUAL	NA	0	2	0	1	0.1574	0.5481	0.5481	0.0002	0.5040	0.0007	0.0013	0.6976	0.0044	NA	NA	NA
DDX6	10	0	0	10	8	1	1	0	1	0	0	1	0	0	0	EQUAL	NA	1	1	1	0	0.1588	1.2007	1.2007	0.0085	0.8591	0.0310	0.0270	0.9028	0.0870	NA	NA	NA
SOX21	5	0	0	5	0	0	0	0	0	0	0	0	0	0	0	EQUAL	NA	0	1	0	0	0.1606	0.0000	0.0000	0.0092	0.5804	0.0287	0.0286	0.7170	0.0820	NA	NA	NA
SOX2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	EQUAL	NA	4	1	0	0	0.1688	0.0000	0.0000	0.0131	0.5164	0.0373	0.0382	0.6995	0.1008	NA	NA	NA
CDX2	8	0	0	8	2	0	1	0	1	0	0	1	0	0	0	EQUAL	NA	1	1	0	0	0.1819	0.0000	0.0000	0.0181	0.4482	0.0459	0.0498	0.6913	0.1179	NA	NA	NA
PIM1	2	0	0	2	1	0	0	0	0	0	0	0	0	0	0	EQUAL	NA	2	1	1	0	0.1942	2.2359	2.2359	0.0250	0.4778	0.0631	0.0646	0.6944	0.1508	0.594	0.462	-0.132
SH2B3	4	0	0	4	2	0	1	0	1	0	0	1	0	0	0	EQUAL	NA	1	2	0	0	0.1952	0.0000	0.0000	0.0016	0.2439	0.0035	0.0071	0.6858	0.0156	0.399	0.324	-0.075
ZBTB16	17	0	0	17	2	0	1	0	1	0	0	1	0	0	0	EQUAL	NA	1	2	1	0	0.1964	1.3338	1.3338	0.0017	0.7834	0.0070	0.0075	0.8472	0.0270	NA	NA	NA
NTHL1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	EQUAL	NA	0	1	0	0	0.2016	0.0000	0.0000	0.0299	0.3526	0.0614	0.0744	0.6858	0.1474	0.573	0.607	0.034
ZFHX3	105	0	0	105	25	0	6	1	5	0	0	6	0	0	0	EQUAL	NA	1	11	0	0	0.2072	0.0000	0.0000	0.0000	0.0064	0.0000	0.0000	0.6858	0.0000	0.513	0.564	0.051
CCND3	1	0	0	1	1	0	1	1	0	0	0	1	0	0	0	EQUAL	NA	0	1	0	0	0.2155	0.0000	0.0000	0.0402	0.3843	0.0834	0.0943	0.6858	0.1863	NA	NA	NA

KNSTRN	4	0	0	4	2	0	0	0	0	0	0	0	0	0	0	EQUAL	NA	1	1	1	0	0.2300	1.4860	1.4860	0.0526	0.7102	0.1426	0.1173	0.7948	0.2767	0.470	0.695	0.225
BCL11B	49	0	0	49	0	0	0	0	0	0	0	0	0	0	0	EQUAL	NA	1	4	2	0	0.2309	2.8589	2.8589	0.0001	0.2058	0.0003	0.0009	0.6858	0.0020	NA	NA	NA
FOXR1	13	0	0	13	0	0	0	0	0	0	0	0	0	0	0	EQUAL	NA	1	1	0	0	0.2311	0.0000	0.0000	0.0536	0.2963	0.0900	0.1191	0.6858	0.1975	NA	NA	NA
RAD51B	8	0	0	8	1	0	0	0	0	0	0	0	0	0	0	EQUAL	NA	0	1	0	0	0.2316	0.0000	0.0000	0.0541	0.2471	0.0801	0.1201	0.6858	0.1806	0.479	0.387	-0.091
ZNF384	6	0	0	6	6	0	2	0	2	0	0	2	0	0	0	EQUAL	NA	0	2	0	0	0.2557	0.0000	0.0000	0.0129	0.2154	0.0210	0.0377	0.6858	0.0643	0.475	0.292	-0.182
GLI1	24	0	0	24	0	0	0	0	0	0	0	0	0	0	0	EQUAL	NA	0	4	1	0	0.2607	0.8276	0.8276	0.0006	0.8451	0.0025	0.0030	0.8927	0.0121	NA	NA	NA
USP6	31	0	0	31	1	1	1	0	1	1	1	0	0	0	0	EQUAL	NA	1	5	0	0	0.2635	0.0000	0.0000	0.0001	0.0314	0.0001	0.0010	0.6858	0.0007	NA	NA	NA
PWWP2A	10	0	0	10	4	0	0	0	0	0	0	0	0	0	0	EQUAL	NA	1	3	1	0	0.2654	1.3396	1.3396	0.0033	0.7804	0.0127	0.0126	0.8446	0.0436	0.452	0.616	0.164
NUTM1	47	0	0	47	0	0	0	0	0	0	0	0	0	0	0	EQUAL	NA	1	4	1	0	0.2674	0.7557	0.7557	0.0007	0.7689	0.0033	0.0039	0.8366	0.0149	NA	NA	NA
JAZF1	4	0	0	4	0	0	0	0	0	0	0	0	0	0	0	EQUAL	NA	1	1	0	1	0.2715	2.6730	2.6730	0.0967	0.3979	0.1761	0.1904	0.6858	0.3206	0.418	0.238	-0.179
NKX2-1	6	0	0	6	2	0	2	0	2	1	0	1	0	0	0	EQUAL	NA	0	2	0	0	0.2762	0.0000	0.0000	0.0209	0.4236	0.0504	0.0559	0.6869	0.1271	NA	NA	NA
ATF1	4	0	0	4	4	0	0	0	0	0	0	0	0	0	0	EQUAL	NA	0	1	0	0	0.2836	0.0000	0.0000	0.1116	0.3141	0.1699	0.2130	0.6858	0.3128	NA	NA	NA
MYOD1	25	0	0	25	0	0	0	0	0	0	0	0	0	0	0	EQUAL	NA	0	2	0	0	0.2843	0.0000	0.0000	0.0248	0.4871	0.0632	0.0640	0.6962	0.1510	NA	NA	NA
CD274	1	0	0	1	0	0	0	0	0	0	0	0	0	0	0	EQUAL	NA	1	1	0	1	0.2852	2.3486	2.3486	0.1136	0.4546	0.2163	0.2162	0.6915	0.3710	NA	NA	NA
CBFB	6	0	0	6	5	0	2	0	2	1	0	1	0	0	0	EQUAL	NA	0	1	0	0	0.3169	0.0000	0.0000	0.1560	0.3419	0.2327	0.2759	0.6858	0.3897	NA	NA	NA
SNX29	31	0	0	31	6	0	0	0	0	0	0	0	0	0	0	EQUAL	NA	0	4	0	0	0.3276	0.0000	0.0000	0.0062	0.0913	0.0057	0.0210	0.6858	0.0230	NA	NA	NA
MDM2	4	0	0	4	4	0	1	1	0	0	0	1	0	0	0	EQUAL	NA	1	2	0	0	0.3316	0.0000	0.0000	0.0562	0.1716	0.0634	0.1237	0.6858	0.1513	NA	NA	NA
CLP1	19	0	0	19	14	1	3	2	1	1	0	2	0	0	0	EQUAL	NA	1	2	0	0	0.3329	0.0000	0.0000	0.0572	0.3123	0.0984	0.1256	0.6858	0.2107	NA	NA	NA
FEN1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	EQUAL	NA	1	2	0	0	0.3707	0.0000	0.0000	0.0930	0.3774	0.1652	0.1847	0.6858	0.3073	0.499	0.539	0.040
TRAF7	7	0	0	7	5	1	2	2	0	2	1	0	0	0	0	EQUAL	NA	0	4	0	0	0.3854	0.0000	0.0000	0.0235	0.1327	0.0249	0.0616	0.6858	0.0733	0.481	0.738	0.257
CNBP	3	0	0	3	3	0	3	1	2	0	0	3	0	0	0	EQUAL	NA	0	1	2	0	0.3942	6.5314	6.5314	0.2710	0.0424	0.0696	0.4171	0.6858	0.1624	0.607	0.916	0.309
EXT2	19	0	0	19	17	0	7	2	5	2	0	5	0	0	0	EQUAL	NA	0	4	0	1	0.4022	0.9516	0.9516	0.0319	0.9601	0.0999	0.0784	0.9740	0.2131	0.467	0.667	0.200
BTG1	1	0	0	1	1	0	1	0	1	0	0	1	0	0	0	EQUAL	NA	0	1	0	0	0.4038	0.0000	0.0000	0.2858	0.5054	0.4532	0.4332	0.6976	0.5986	0.553	0.573	0.021

LMO2	7	0	0	7	6	0	3	0	3	1	0	2	0	0	0	EQUAL	NA	0	1	0	0	0.4080	0.0000	0.0000	0.2924	0.5384	0.4754	0.4406	0.7054	0.6163	0.468	0.467	-0.001
FEV	2	0	0	2	0	0	0	0	0	0	0	0	0	0	0	EQUAL	NA	1	2	0	0	0.4148	0.0000	0.0000	0.1450	0.5268	0.2830	0.2612	0.7022	0.4433	NA	NA	NA
SIX1	7	0	0	7	2	0	1	0	1	1	0	0	0	0	0	EQUAL	NA	0	2	0	0	0.4200	0.0000	0.0000	0.1518	0.4145	0.2567	0.2699	0.6861	0.4162	0.371	0.417	0.045
GRIN2A	114	0	0	114	1	0	0	0	0	0	0	0	0	0	0	EQUAL	NA	7	9	0	1	0.4246	0.7001	0.7001	0.0027	0.7046	0.0105	0.0109	0.7906	0.0373	NA	NA	NA
PTPN6	5	0	0	5	4	0	2	0	2	0	0	2	0	0	0	EQUAL	NA	0	4	0	0	0.4335	0.0000	0.0000	0.0522	0.1556	0.0554	0.1166	0.6858	0.1362	0.462	0.498	0.036
NUTM2D	11	0	0	11	1	1	0	0	0	0	0	0	0	0	0	EQUAL	NA	2	5	1	0	0.4386	1.0435	1.0435	0.0328	0.9663	0.1023	0.0801	0.9788	0.2167	NA	NA	NA
ATM	71	0	0	71	11	1	3	1	2	2	1	1	0	0	0	EQUAL	NA	2	17	0	4	0.4482	0.7222	0.7222	0.0001	0.4913	0.0005	0.0010	0.6962	0.0035	0.562	0.588	0.026
MNX1	4	0	0	4	0	0	0	0	0	0	0	0	0	0	0	EQUAL	NA	2	4	0	0	0.4545	0.0000	0.0000	0.0695	0.4123	0.1377	0.1467	0.6858	0.2697	0.383	0.601	0.218
PDCD1LG2	11	0	0	11	2	0	2	0	2	0	0	2	0	0	0	EQUAL	NA	4	3	0	1	0.4609	1.3914	1.3914	0.3279	0.7780	0.5023	0.4785	0.8430	0.6380	0.291	0.163	-0.129
CDK4	3	0	0	3	3	0	0	0	0	0	0	0	0	0	0	EQUAL	NA	0	2	0	0	0.4632	0.0000	0.0000	0.2121	0.2774	0.2546	0.3473	0.6858	0.4139	0.487	0.503	0.016
ERCC4	16	0	0	16	4	0	0	0	0	0	0	0	0	0	0	EQUAL	NA	1	6	2	1	0.4673	2.2162	2.2162	0.0329	0.2234	0.0490	0.0804	0.6858	0.1243	NA	NA	NA
SPEC C1	36	0	0	36	15	1	1	0	1	0	0	1	0	0	0	EQUAL	NA	1	7	0	0	0.4798	0.0000	0.0000	0.0269	0.0660	0.0159	0.0683	0.6858	0.0519	0.535	0.423	-0.112
NUTM2B	10	0	0	10	0	0	0	0	0	0	0	0	0	0	0	EQUAL	NA	0	6	0	0	0.4895	0.0000	0.0000	0.0471	0.1529	0.0501	0.1070	0.6858	0.1265	NA	NA	NA
DDB2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	EQUAL	NA	0	3	1	0	0.4945	1.5663	1.5663	0.1671	0.6763	0.3530	0.2908	0.7721	0.5124	0.443	0.637	0.194
USP44	23	0	0	23	0	0	0	0	0	0	0	0	0	0	0	EQUAL	NA	1	5	0	0	0.5487	0.0000	0.0000	0.1359	0.1498	0.1166	0.2483	0.6858	0.2391	NA	NA	NA
CYP2C8	15	0	0	15	0	0	0	0	0	0	0	0	0	0	0	EQUAL	NA	5	6	1	0	0.5518	0.8132	0.8132	0.3520	0.8486	0.6398	0.5045	0.8950	0.7404	0.430	0.571	0.141
HOXC11	8	0	0	8	1	0	0	0	0	0	0	0	0	0	0	EQUAL	NA	0	3	0	0	0.5621	0.0000	0.0000	0.2698	0.4051	0.3846	0.4156	0.6858	0.5407	0.511	0.455	-0.056
MAP3K1	50	0	0	50	15	0	3	0	3	0	0	3	0	0	0	EQUAL	NA	3	12	5	0	0.5668	2.3501	2.3501	0.0299	0.0943	0.0233	0.0744	0.6858	0.0697	0.447	0.494	0.047
HEY1	4	0	0	4	2	0	1	0	1	0	0	1	0	0	0	EQUAL	NA	1	3	1	0	0.5876	2.0426	2.0426	0.3123	0.5232	0.4895	0.4618	0.7012	0.6279	0.606	0.000	-0.606
WT1	13	0	0	13	0	0	0	0	0	0	0	0	0	0	0	EQUAL	NA	1	5	0	1	0.5877	1.4257	1.4257	0.1923	0.7377	0.4041	0.3231	0.8141	0.5581	NA	NA	NA
TERT	17	0	0	17	0	0	0	0	0	0	0	0	0	0	0	EQUAL	NA	14	16	0	0	0.5990	0.0000	0.0000	0.1610	0.0793	0.1057	0.2827	0.6858	0.2222	0.305	0.289	-0.017
ERC1	28	0	0	28	8	0	1	0	1	0	0	1	0	0	0	EQUAL	NA	3	9	0	0	0.5994	0.0000	0.0000	0.0932	0.0418	0.0308	0.1851	0.6858	0.0865	0.540	0.354	-0.186
FANCG	5	0	0	5	5	0	2	0	2	0	0	2	0	0	0	EQUAL	NA	2	5	0	0	0.6000	0.0000	0.0000	0.2120	0.1597	0.1708	0.3472	0.6858	0.3139	0.394	0.357	-0.037

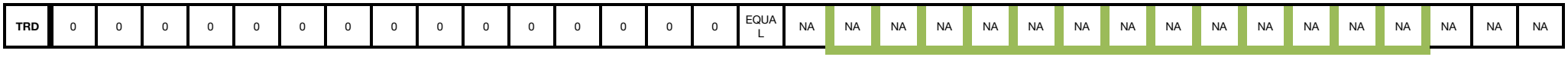
FAT3	258	0	0	258	0	0	0	0	0	0	0	0	0	0	0	EQUAL	NA	7	39	0	0	0.6369	0.0000	0.0000	0.0023	0.0030	0.0001	0.0096	0.6059	0.0011	NA	NA	NA
CYSLTR2	18	0	0	18	0	0	0	0	0	0	0	0	0	0	0	EQUAL	NA	0	3	0	0	0.6978	0.0000	0.0000	0.5074	0.5187	0.6518	0.6525	0.7004	0.7478	NA	NA	NA
XPC	8	0	0	8	4	1	1	1	0	1	1	0	0	0	0	EQUAL	NA	7	12	0	0	0.7128	0.0000	0.0000	0.4941	0.0655	0.1673	0.6412	0.6858	0.3093	0.433	0.549	0.116
HOXA11	2	0	0	2	1	0	1	0	1	0	0	1	0	0	0	EQUAL	NA	5	6	0	0	0.7143	0.0000	0.0000	0.5945	0.3661	0.6163	0.7245	0.6858	0.7233	0.601	1.098	0.497
TMEM127	9	0	0	9	9	1	4	2	2	1	1	3	0	0	0	EQUAL	NA	3	3	0	0	0.7234	0.0000	0.0000	0.6235	0.4826	0.6956	0.7480	0.6955	0.7795	0.486	0.803	0.317
TLX1	5	0	0	5	0	0	0	0	0	0	0	0	0	0	0	EQUAL	NA	6	7	0	1	0.7836	2.4382	2.4382	0.6713	0.4602	0.6283	0.7828	0.6916	0.7318	0.535	0.558	0.024
OMD	17	0	0	17	0	0	0	0	0	0	0	0	0	0	0	EQUAL	NA	2	4	0	0	0.7885	0.0000	0.0000	0.6207	0.3457	0.5671	0.7458	0.6858	0.6881	NA	NA	NA
CPEB3	26	0	0	26	1	0	0	0	0	0	0	0	0	0	0	EQUAL	NA	2	9	0	0	0.8043	0.0000	0.0000	0.4978	0.1830	0.3274	0.6443	0.6858	0.4892	0.375	0.460	0.085
NBEA	189	0	0	189	8	0	5	2	3	3	0	2	0	0	0	EQUAL	NA	8	31	4	2	0.8128	1.2279	1.2279	0.2320	0.6267	0.4349	0.3714	0.7408	0.5835	NA	NA	NA
SIX2	20	0	0	20	1	0	0	0	0	0	0	0	0	0	0	EQUAL	NA	12	16	0	0	0.8141	0.0000	0.0000	0.5975	0.1579	0.3521	0.7270	0.6858	0.5118	0.581	0.496	-0.085
PMS2	15	0	0	15	7	0	1	1	0	0	0	1	0	0	0	EQUAL	NA	6	11	1	0	0.8206	0.6292	0.6292	0.7090	0.6537	0.8779	0.8109	0.7569	0.9094	0.494	0.562	0.068
RECQL4	7	0	0	7	7	0	3	0	3	3	0	0	0	0	0	EQUAL	NA	5	17	0	1	0.8543	0.5852	0.5852	0.5050	0.5564	0.6735	0.6505	0.7102	0.7633	0.387	0.435	0.048
TSC1	41	0	0	41	13	1	4	1	3	2	1	2	0	0	0	EQUAL	NA	8	16	2	1	0.8551	1.4294	1.4294	0.7266	0.6110	0.7276	0.8247	0.7317	0.8008	0.460	0.699	0.240
ALK	32	0	0	32	0	0	0	0	0	0	0	0	0	0	0	EQUAL	NA	10	20	2	0	0.8579	0.8591	0.8591	0.4817	0.8255	0.7620	0.6291	0.8778	0.8247	NA	NA	NA
SDHAF2	4	0	0	4	3	0	3	0	3	0	0	3	0	0	0	EQUAL	NA	0	2	0	0	0.8588	0.0000	0.0000	0.8252	0.4755	0.7565	0.8951	0.6944	0.8210	NA	NA	NA
SBDS	2	0	0	2	2	0	1	0	1	0	0	1	0	0	0	EQUAL	NA	0	3	1	0	0.8623	2.4234	2.4234	0.7925	0.4403	0.7172	0.8712	0.6901	0.7939	0.484	0.736	0.252
SKI	6	0	0	6	2	0	1	0	1	0	0	1	0	0	0	EQUAL	NA	12	19	0	1	0.8713	0.8207	0.8207	0.7142	0.8456	0.9296	0.8152	0.8929	0.9462	0.465	0.537	0.072
MYCN	10	0	0	10	1	0	0	0	0	0	0	0	0	0	0	EQUAL	NA	11	17	0	0	0.8721	0.0000	0.0000	0.7292	0.1995	0.4343	0.8265	0.6858	0.5831	0.497	0.621	0.124
MDM4	9	0	0	9	2	0	0	0	0	0	0	0	0	0	0	EQUAL	NA	11	24	0	0	0.9050	0.0000	0.0000	0.7891	0.0121	0.0369	0.8692	0.6858	0.0999	0.642	0.826	0.184
FHIT	10	0	0	10	5	0	1	0	1	0	0	1	0	0	0	EQUAL	NA	1	2	0	0	0.9107	0.0000	0.0000	0.8932	0.4733	0.7663	0.9394	0.6944	0.8272	0.364	0.696	0.332
PREX2	23	0	0	23	0	0	0	0	0	0	0	0	0	0	0	EQUAL	NA	15	38	1	2	0.9167	0.5517	0.5517	0.7798	0.3184	0.6010	0.8625	0.6858	0.7123	0.417	1.042	0.625
ANK1	98	0	0	98	3	0	1	0	1	0	0	1	0	0	0	EQUAL	NA	29	57	5	1	0.9273	0.9232	0.9232	0.7433	0.8576	0.9455	0.8360	0.9019	0.9581	0.487	0.690	0.203
NTRK1	49	0	0	49	0	0	0	0	0	0	0	0	0	0	0	EQUAL	NA	28	54	1	4	0.9529	0.9967	0.9967	0.8374	0.9945	0.9773	0.9040	0.9969	0.9830	0.524	1.837	1.313

MPL	29	0	0	29	0	0	0	0	0	0	0	0	0	0	0	EQUAL	NA	17	35	3	1	0.9588	0.8543	0.8543	0.8884	0.7741	0.9591	0.9369	0.8404	0.9689	NA	NA	NA
SDHC	4	0	0	4	4	0	3	1	2	0	0	3	0	0	0	EQUAL	NA	4	6	0	2	0.9713	2.6795	2.6795	0.9657	0.2956	0.5136	0.9815	0.6858	0.6464	0.453	0.652	0.199
SETDB1	24	0	0	24	19	0	7	1	6	6	0	1	0	0	0	EQUAL	NA	27	73	9	1	0.9890	1.2448	1.2448	0.9612	0.5620	0.8014	0.9788	0.7122	0.8538	0.504	0.574	0.070
ITK	14	0	0	14	0	0	0	0	0	0	0	0	0	0	0	EQUAL	NA	7	16	1	0	0.9905	0.4332	0.4332	0.9837	0.3903	0.6512	0.9915	0.6858	0.7475	0.511	0.033	-0.478
WIF1	18	0	0	18	1	0	1	0	1	1	0	0	0	0	0	EQUAL	NA	0	5	0	0	0.9928	0.0000	0.0000	0.9870	0.2279	0.4833	0.9929	0.6858	0.6224	NA	NA	NA
ECT2L	38	0	0	38	0	0	0	0	0	0	0	0	0	0	0	EQUAL	NA	6	14	1	1	1.0033	1.0003	1.0003	0.9947	0.9997	1.0000	0.9971	0.9999	1.0000	NA	NA	NA
CD28	16	0	0	16	0	0	0	0	0	0	0	0	0	0	0	EQUAL	NA	2	3	0	0	1.0171	0.0000	0.0000	0.9766	0.4898	0.7875	0.9878	0.6962	0.8429	0.351	0.000	-0.351
CCR4	20	0	0	20	0	0	0	0	0	0	0	0	0	0	0	EQUAL	NA	3	5	0	0	1.0207	0.0000	0.0000	0.9790	0.4950	0.7842	0.9886	0.6962	0.8403	NA	NA	NA
PAX3	34	0	0	34	0	0	0	0	0	0	0	0	0	0	0	EQUAL	NA	7	13	0	1	1.0220	0.8841	0.8841	0.9640	0.9074	0.9896	0.9803	0.9382	0.9923	NA	NA	NA
FOXL2	2	0	0	2	0	0	0	0	0	0	0	0	0	0	0	EQUAL	NA	5	8	0	0	1.0381	0.0000	0.0000	0.9163	0.5210	0.8093	0.9536	0.7008	0.8600	NA	NA	NA
IGF2BP2	22	0	0	22	8	1	2	2	0	2	1	0	0	0	0	EQUAL	NA	4	9	1	1	1.0490	1.7685	1.7685	0.8868	0.4653	0.7659	0.9359	0.6921	0.8271	0.397	0.386	-0.010
MYC	17	0	0	17	16	0	8	3	5	5	0	3	0	0	0	EQUAL	NA	0	8	0	0	1.0718	0.0000	0.0000	0.8463	0.2954	0.5676	0.9097	0.6858	0.6885	0.457	0.847	0.390
FAM135B	51	0	0	51	0	0	0	0	0	0	0	0	0	0	0	EQUAL	NA	35	93	5	3	1.0925	0.9485	0.9485	0.6553	0.8922	0.8584	0.7718	0.9265	0.8957	NA	NA	NA
AFF3	61	0	0	61	3	2	2	1	1	1	1	1	0	0	0	EQUAL	NA	18	43	4	0	1.0955	1.0388	1.0388	0.7464	0.9455	0.9478	0.8382	0.9641	0.9602	0.513	0.745	0.232
MLH1	8	0	0	8	4	0	0	0	0	0	0	0	0	0	0	EQUAL	NA	2	11	2	3	1.1227	3.5854	3.5854	0.7065	0.0212	0.0692	0.8093	0.6858	0.1619	0.507	0.406	-0.101
WRN	33	0	0	33	6	1	1	0	1	0	0	1	0	0	0	EQUAL	NA	8	25	1	2	1.1274	0.9245	0.9245	0.7714	0.9081	0.9198	0.8560	0.9385	0.9393	0.451	0.754	0.303
RGPD3	102	0	0	102	11	0	11	0	11	2	0	9	0	0	0	EQUAL	NA	11	36	2	0	1.1540	0.5043	0.5043	0.6796	0.3395	0.4222	0.7899	0.6858	0.5728	NA	NA	NA
IL2	4	0	0	4	0	0	0	0	0	0	0	0	0	0	0	EQUAL	NA	0	2	1	2	1.1698	12.9716	12.9716	0.8288	0.0027	0.0097	0.8976	0.5694	0.0351	NA	NA	NA
B2M	4	0	0	4	4	1	4	1	3	1	1	3	0	0	0	EQUAL	NA	0	2	1	1	1.1852	8.7881	8.7881	0.8152	0.0238	0.0777	0.8880	0.6858	0.1768	NA	NA	NA
MUTYH	12	0	0	12	11	0	4	0	4	0	0	4	0	0	0	EQUAL	NA	12	32	1	4	1.2017	1.5361	1.5361	0.5884	0.4374	0.7192	0.7200	0.6900	0.7950	0.422	0.648	0.226
IL21R	29	0	0	29	4	0	2	0	2	1	0	1	0	0	0	EQUAL	NA	1	9	0	0	1.2165	0.0000	0.0000	0.5692	0.2416	0.4285	0.7052	0.6858	0.5782	NA	NA	NA
GRM3	67	0	0	67	0	0	0	0	0	0	0	0	0	0	0	EQUAL	NA	15	39	0	2	1.2190	0.9211	0.9211	0.5140	0.9124	0.7678	0.6582	0.9412	0.8282	NA	NA	NA
MLLT3	12	0	0	12	8	2	5	1	4	2	1	3	0	0	0	EQUAL	NA	3	9	0	0	1.2207	0.0000	0.0000	0.6418	0.2134	0.3921	0.7622	0.6858	0.5476	0.411	0.801	0.390

PHOX2B	9	0	0	9	0	0	0	0	0	0	0	0	0	0	0	EQUAL	NA	1	7	1	0	1.2284	3.5140	3.5140	0.5987	0.2981	0.5067	0.7278	0.6858	0.6413	NA	NA	NA
HOXA13	1	0	0	1	0	0	0	0	0	0	0	0	0	0	0	EQUAL	NA	3	10	2	1	1.2768	10.7064	10.7064	0.4591	0.0032	0.0129	0.6082	0.6059	0.0440	0.500	0.763	0.263
LATS1	36	0	0	36	10	2	1	0	1	0	0	1	0	0	0	EQUAL	NA	5	19	1	0	1.2894	0.6393	0.6393	0.5586	0.6288	0.6646	0.6960	0.7412	0.7576	0.461	0.596	0.135
TP63	45	0	0	45	0	0	0	0	0	0	0	0	0	0	0	EQUAL	NA	6	16	1	0	1.2897	0.7470	0.7470	0.6006	0.7816	0.7695	0.7291	0.8455	0.8295	NA	NA	NA
TENT5C	16	0	0	16	5	0	1	0	1	0	0	1	0	0	0	EQUAL	NA	7	17	2	0	1.2919	2.7838	2.7838	0.5726	0.2514	0.5057	0.7079	0.6858	0.6405	0.478	0.535	0.057
CSMD3	373	0	0	373	1	0	0	0	0	0	0	0	0	0	0	EQUAL	NA	83	320	32	9	1.2997	1.2539	1.2539	0.0295	0.2428	0.0926	0.0737	0.6858	0.2018	NA	NA	NA
EBF1	29	0	0	29	3	1	2	0	2	1	1	1	0	0	0	EQUAL	NA	11	28	2	1	1.3052	1.5260	1.5260	0.4537	0.5355	0.7024	0.6032	0.7048	0.7845	0.421	0.000	-0.421
SLC45A3	11	0	0	11	5	0	1	1	0	0	0	1	0	0	0	EQUAL	NA	14	34	1	0	1.3075	0.7375	0.7375	0.3966	0.7588	0.6083	0.5484	0.8294	0.7173	0.494	0.678	0.184
IL6ST	25	0	0	25	9	0	4	1	3	1	0	3	0	0	0	EQUAL	NA	2	15	0	2	1.3478	1.4147	1.4147	0.2923	0.6428	0.5414	0.4405	0.7493	0.6682	0.474	0.615	0.141
CASP3	14	0	0	14	12	1	4	2	2	1	0	3	0	0	0	EQUAL	NA	2	5	0	0	1.3601	0.0000	0.0000	0.6989	0.3497	0.5513	0.8035	0.6858	0.6766	0.474	0.475	0.001
TAL1	7	0	0	7	0	0	0	0	0	0	0	0	0	0	0	EQUAL	NA	7	16	0	0	1.3784	0.0000	0.0000	0.4820	0.3328	0.4300	0.6294	0.6858	0.5792	NA	NA	NA
RUNX1T1	59	0	0	59	1	0	0	0	0	0	0	0	0	0	0	EQUAL	NA	12	36	2	2	1.3807	1.3535	1.3535	0.3288	0.6114	0.6156	0.4796	0.7318	0.7227	0.356	0.225	-0.130
HOXD13	23	0	0	23	0	0	0	0	0	0	0	0	0	0	0	EQUAL	NA	2	9	0	1	1.3839	3.2159	3.2159	0.3559	0.3277	0.4175	0.5088	0.6858	0.5692	NA	NA	NA
SRGAP3	26	0	0	26	0	0	0	0	0	0	0	0	0	0	0	EQUAL	NA	3	23	1	0	1.4110	0.5897	0.5897	0.1313	0.5625	0.2579	0.2418	0.7122	0.4176	0.335	0.894	0.560
LRP1B	347	0	0	347	0	0	0	0	0	0	0	0	0	0	0	EQUAL	NA	74	317	25	14	1.4261	1.3755	1.3755	0.0045	0.1140	0.0172	0.0163	0.6858	0.0550	NA	NA	NA
CNBD1	41	0	0	41	0	0	0	0	0	0	0	0	0	0	0	EQUAL	NA	8	26	4	0	1.4329	1.6043	1.6043	0.3715	0.4580	0.6246	0.5240	0.6916	0.7294	NA	NA	NA
KLF6	5	0	0	5	4	1	0	0	0	0	0	0	0	0	0	EQUAL	NA	2	6	2	0	1.4334	5.1943	5.1943	0.4625	0.0769	0.2072	0.6112	0.6858	0.3598	0.531	0.433	-0.098
REL	27	0	0	27	4	0	2	1	1	1	0	1	0	0	0	EQUAL	NA	4	13	0	2	1.4456	1.8602	1.8602	0.5286	0.4995	0.7455	0.6717	0.6973	0.8128	0.502	0.521	0.020
IRF4	11	0	0	11	3	2	3	0	3	2	2	1	0	0	0	EQUAL	NA	8	20	1	0	1.4502	0.7014	0.7014	0.3738	0.7276	0.5409	0.5265	0.8063	0.6679	0.487	0.000	-0.487
PAX7	42	0	0	42	2	1	0	0	0	0	0	0	0	0	0	EQUAL	NA	15	42	1	1	1.4547	0.8787	0.8787	0.2060	0.8617	0.3793	0.3403	0.9045	0.5354	NA	NA	NA
RHOH	5	0	0	5	1	0	1	0	1	1	0	0	0	0	0	EQUAL	NA	2	4	1	0	1.4566	6.0908	6.0908	0.6167	0.1900	0.4225	0.7424	0.6858	0.5729	0.539	0.172	-0.367
ZNF479	71	0	0	71	0	0	0	0	0	0	0	0	0	0	0	EQUAL	NA	11	41	8	1	1.4810	2.8763	2.8763	0.2394	0.0248	0.0798	0.3799	0.6858	0.1802	NA	NA	NA
PAX5	24	0	0	24	0	0	0	0	0	0	0	0	0	0	0	EQUAL	NA	3	9	2	0	1.4853	3.4229	3.4229	0.3886	0.1606	0.3459	0.5410	0.6858	0.5062	0.371	0.000	-0.371

XPA	4	0	0	4	1	0	1	0	1	0	0	1	0	0	0	EQUAL	NA	1	6	0	0	1.4853	0.0000	0.0000	0.3838	0.2920	0.3798	0.5362	0.6858	0.5360	0.566	0.690	0.124
TLX3	8	0	0	8	0	0	0	0	0	0	0	0	0	0	0	EQUAL	NA	2	9	1	0	1.5487	4.6357	4.6357	0.2446	0.2208	0.2701	0.3864	0.6858	0.4303	NA	NA	NA
EIF4A2	2	0	0	2	2	1	2	0	2	1	1	1	0	0	0	EQUAL	NA	3	10	0	0	1.5649	0.0000	0.0000	0.5115	0.2161	0.2527	0.6562	0.6858	0.4118	0.472	0.709	0.238
MYD88	2	0	0	2	1	0	0	0	0	0	0	0	0	0	0	EQUAL	NA	3	8	1	0	1.6484	2.2026	2.2026	0.4754	0.5319	0.7187	0.6237	0.7036	0.7949	0.467	0.521	0.055
CTNND2	50	0	0	50	5	0	3	1	2	2	0	1	0	0	0	EQUAL	NA	24	79	4	4	1.6734	1.8671	1.8671	0.0222	0.1490	0.0587	0.0588	0.6858	0.1423	0.471	0.649	0.177
CNTNAP2	132	0	0	132	2	0	0	0	0	0	0	0	0	0	0	EQUAL	NA	22	86	6	5	1.6935	1.9930	1.9930	0.0218	0.0752	0.0485	0.0578	0.6858	0.1232	0.469	0.569	0.100
FCRL4	16	0	0	16	0	0	0	0	0	0	0	0	0	0	0	EQUAL	NA	12	45	4	3	1.7207	1.8247	1.8247	0.0845	0.2240	0.2017	0.1712	0.6858	0.3527	NA	NA	NA
HIP1	44	0	0	44	19	1	7	2	5	4	0	3	0	0	0	EQUAL	NA	5	27	0	2	1.8374	0.9350	0.9350	0.0567	0.9235	0.1273	0.1247	0.9484	0.2549	0.409	0.353	-0.055
CDH10	79	0	0	79	1	0	0	0	0	0	0	0	0	0	0	EQUAL	NA	19	88	7	2	1.8533	1.7733	1.7733	0.0102	0.1749	0.0356	0.0313	0.6858	0.0969	NA	NA	NA
TET1	107	0	0	107	2	0	0	0	0	0	0	0	0	0	0	EQUAL	NA	10	50	3	0	1.8559	1.2118	1.2118	0.0561	0.7624	0.1407	0.1235	0.8320	0.2740	0.529	0.701	0.172
EPHA7	75	0	0	75	8	0	3	0	3	1	0	2	0	0	0	EQUAL	NA	9	44	3	1	1.8864	1.3985	1.3985	0.0702	0.5886	0.1840	0.1477	0.7209	0.3308	0.418	0.289	-0.129
ACVR2A	30	0	0	30	12	1	3	1	2	1	1	2	0	0	0	EQUAL	NA	2	12	5	2	1.9018	8.4979	8.4979	0.1623	0.0005	0.0019	0.2843	0.2347	0.0097	0.478	0.588	0.110
ABL2	59	0	0	59	7	0	3	0	3	2	0	1	0	0	0	EQUAL	NA	16	76	5	1	1.9226	1.4195	1.4195	0.0121	0.4785	0.0399	0.0358	0.6944	0.1061	0.459	0.442	-0.017
WDCP	25	0	0	25	5	0	2	0	2	0	0	2	0	0	0	EQUAL	NA	2	18	1	0	2.0051	1.6793	1.6793	0.0506	0.6332	0.1474	0.1135	0.7435	0.2832	0.488	0.505	0.017
RGS7	42	0	0	42	0	0	0	0	0	0	0	0	0	0	0	EQUAL	NA	11	51	6	4	2.0192	2.4846	2.4846	0.0259	0.0437	0.0493	0.0664	0.6858	0.1248	NA	NA	NA
MLLT11	2	0	0	2	1	0	0	0	0	0	0	0	0	0	0	EQUAL	NA	3	7	0	0	2.2327	0.0000	0.0000	0.2529	0.5681	0.3928	0.3961	0.7137	0.5483	0.465	0.301	-0.164
CDH17	43	0	0	43	12	4	8	3	5	6	4	2	0	0	0	EQUAL	NA	5	27	4	2	2.4356	4.8767	4.8767	0.0540	0.0122	0.0318	0.1199	0.6858	0.0890	NA	NA	NA
ERBB4	125	0	0	125	1	0	1	0	1	0	0	1	0	0	0	EQUAL	NA	13	90	6	3	2.5764	1.9776	1.9776	0.0004	0.1292	0.0020	0.0026	0.6858	0.0101	NA	NA	NA
PLAG1	33	0	0	33	3	0	1	0	1	0	0	1	0	0	0	EQUAL	NA	1	17	0	0	2.6251	0.0000	0.0000	0.0165	0.3226	0.0227	0.0462	0.6858	0.0681	0.318	0.327	0.009
ELK4	5	0	0	5	2	0	2	0	2	1	0	1	0	0	0	EQUAL	NA	1	14	1	1	2.6411	4.8762	4.8762	0.0254	0.0756	0.0442	0.0654	0.6858	0.1146	0.470	0.702	0.232
TAL2	3	0	0	3	0	0	0	0	0	0	0	0	0	0	0	EQUAL	NA	1	4	0	0	2.7938	0.0000	0.0000	0.1897	0.7016	0.3705	0.3199	0.7889	0.5281	NA	NA	NA
FAM131B	16	0	0	16	0	0	0	0	0	0	0	0	0	0	0	EQUAL	NA	3	15	1	0	2.9088	1.6350	1.6350	0.0776	0.6851	0.1987	0.1599	0.7777	0.3492	NA	NA	NA
RSPO2	22	0	0	22	0	0	0	0	0	0	0	0	0	0	0	EQUAL	NA	1	10	3	0	2.9997	6.6699	6.6699	0.0620	0.0231	0.0561	0.1338	0.6858	0.1375	NA	NA	NA

ID3	3	0	0	3	3	1	1	0	1	1	1	0	0	0	0	EQUAL	NA	1	7	0	0	3.4593	0.0000	0.0000	0.0463	0.6161	0.1053	0.1057	0.7352	0.2214	0.436	0.402	-0.034
BCL10	9	0	0	9	6	1	2	2	0	0	0	2	0	0	0	EQUAL	NA	2	12	3	0	3.8469	7.1999	7.1999	0.0620	0.0394	0.0768	0.1339	0.6858	0.1753	0.464	0.693	0.228
ELF3	6	0	0	6	5	1	5	2	3	4	1	1	0	0	0	EQUAL	NA	2	23	2	0	4.0815	3.2075	3.2075	0.0017	0.1672	0.0073	0.0076	0.6858	0.0280	0.447	0.667	0.220
VHL	10	0	0	10	7	2	3	0	3	2	1	1	0	0	0	EQUAL	NA	1	18	5	3	4.5569	25.8561	25.8561	0.0009	0.0000	0.0000	0.0043	0.0003	0.0000	0.469	0.600	0.131
PCBP1	2	0	0	2	2	0	2	1	1	2	0	0	0	0	0	EQUAL	NA	1	34	1	0	6.2370	4.7660	4.7660	0.0000	0.2142	0.0000	0.0000	0.6858	0.0002	0.527	0.655	0.128
S100A7	21	0	0	21	2	0	0	0	0	0	0	0	0	0	0	EQUAL	NA	2	22	1	1	10.0220	15.4912	15.4912	0.0002	0.0182	0.0005	0.0012	0.6858	0.0032	NA	NA	NA
H3F3B	4	0	0	4	4	1	2	1	1	1	1	1	0	0	0	EQUAL	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
FGFR1OP	5	0	0	5	2	0	1	0	1	0	0	1	0	0	0	EQUAL	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	0.413	0.843	0.430
HIST1H3B	6	0	0	6	1	0	1	0	1	0	0	1	0	0	0	EQUAL	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
CRLF2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	EQUAL	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
DUX4L1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	EQUAL	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
HIST1H4I	10	0	0	10	0	0	0	0	0	0	0	0	0	0	0	EQUAL	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	0.460	1.254	0.794
HMGN2P46	1	0	0	1	0	0	0	0	0	0	0	0	0	0	0	EQUAL	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
IGH	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	EQUAL	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
IGK	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	EQUAL	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
IGL	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	EQUAL	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
MALAT1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	EQUAL	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	0.462	0.631	0.169
MDS2	2	0	0	2	0	0	0	0	0	0	0	0	0	0	0	EQUAL	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
P2RY8	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	EQUAL	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
SEPT5	9	0	0	9	3	1	0	0	0	0	0	0	0	0	0	EQUAL	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
SEPT6	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	EQUAL	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
TRA	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	EQUAL	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
TRB	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	EQUAL	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA


TRD
0
0
0
0
0
0
0
0
0
0
0
0
0
0
0
EQUAL
NA
NA
NA
NA
NA
NA
NA
NA
NA
NA
NA
NA
NA
NA
NA
NA
NA
NA

Supplement Table 6. Heterozygous somatic mutations sites which were screened for their allelic imbalance in transcription efficiency effect, were also screened for TE insertions.

Among the selected Pan-adenocarcinoma patients (n=27), their WGS data were analyzed. The table listed all those somatic SNV mutation sites locating in the regions without the existence of allele specific SCNA events. For each somatic SNV mutation sites, the detailed information about the AITE occurrence and also the counts of the detected *de novo* transposable elements (TE) insertions event in +-1Mbp around respectively in tumor and paired normal-like tissues were provided.

caseID	gene_ID	contig	position	somatic mut._type	CSQ_HGVSc	CSQ_HGVSp	tumor_purity	WESref	WESref_corrected	WESalt	RNAref	RNAref_corrected	RNAalt	RNA_binomial_P_BH_corrected	Z_Proportion_P	+-1Mb_TE_counts (Normal-like/Tumor)
TCGA-44-4112	CTNNB1	chr3	41233359	missense	c.1100G>T	p.Gly367Val	0.4	41	17.43	14	744	206.14	511	1.11E-29	3.61E-03	0/10
TCGA-44-6147	MSI2	chr17	57401444	synonymous	c.378A>C	c.378A>C(p.%3D)	0.38	63	16.50	12	96	7.96	46	5.32E-07	1.85E-04	2/2
TCGA-AA-3514	TMPRSS2	chr21	41473380	missense	c.844G>A	p.Gly282Arg	0.63	41	16.21	26	59	18.30	51	1.97E-04	2.63E-01	0/3
TCGA-44-6777	IL21R	chr16	27430089	synonymous	c.18C>A	c.18C>A(p.%3D)	0.37	29	6.95	6	32	11.84	0	1.31E-03	2.56E-02	0/0
TCGA-A6-3808	TGFBR2	chr3	30674172	missense	c.1322C>T	p.Ser441Phe	0.45	325	102.25	80	19	0.85	14	2.70E-03	6.47E-04	0/0
TCGA-44-4112	GPHN	chr14	67113112	missense	c.1567G>T	p.Val523Phe	0.4	87	38.14	27	10	0.57	12	7.25E-03	2.38E-03	1/6
TCGA-44-6778	CARS	chr11	3020306	synonymous	c.931C>T	c.931C>T(p.%3D)	0.27	46	12.09	13	43	15.98	4	2.87E-02	5.87E-02	0/0
TCGA-44-6778	PTPRK	chr6	127973818	missense	c.3997G>A	p.Gly1333Ser	0.27	58	11.28	6	27	7.29	0	3.61E-02	1.95E-01	0/1
TCGA-44-6147	PPM1D	chr17	60600832	missense	c.418G>T	p.Ala140Ser	0.38	71	18.30	14	31	3.72	13	7.62E-02	5.87E-02	0/4
TCGA-44-4112	FAT1	chr4	186617798	missense	c.8794G>T	p.Asp2932Tyr	0.4	92	35.43	40	16	4.00	12	1.13E-01	1.90E-01	0/4
TCGA-44-6776	KRAS	chr12	25245350	missense	c.35G>A	p.Gly12Asp	0.62	13	4.26	10	11	3.40	9	2.39E-01	1.00E+00	0/0
TCGA-44-6776	BCL9L	chr11	118901486	missense	c.2257A>T	p.Met753Leu	0.62	50	22.64	22	33	17.04	9	2.53E-01	3.58E-01	0/0
TCGA-44-6145	PDCD1LG2	chr9	5534764	synonymous	c.75C>A	c.75C>A(p.%3D)	0.41	59	19.47	8	14	5.15	1	2.63E-01	8.90E-01	1/1
TCGA-44-6778	ZFHX3	chr16	72795125	missense	c.4815G>T	p.Gln1605His	0.27	97	29.17	21	22	4.76	8	7.05E-01	3.44E-01	0/0
TCGA-44-6774	LATS2	chr13	20983299	missense	c.2407G>T	p.Gly803Cys	0.43	57	16.53	14	30	9.48	6	8.43E-01	9.89E-01	3/1
TCGA-44-6778	BAZ1A	chr14	34783871	missense	c.1792A>T	p.Thr598Ser	0.27	69	17.27	21	37	9.41	11	9.17E-01	1.00E+00	2/3
TCGA-44-6145	KRAS	chr12	25245350	missense	c.35G>T	p.Gly12Val	0.41	19	3.66	7	17	4.61	4	1.04E+00	6.84E-01	0/0

Supplement Table 7. Here shows the mean_TE_counts (density) inside investigation window (1Mbp) around those expressed heterozygous SNV sites (with AITE or no_AITE) of different tissues from each cancer patient (n=18, tumor vs. paired-normal-like).

The table showed the detailed information for the average in *de novo* TE insertion density (counts inside investigation window, 1Mbp) for each individual under different target position setting (sites with AITE or no_AITE). These were the exact values being used in (Figure 6B).

caseID	tumor_noAITE _sites	tumor_noAITE _mean_TE_counts _ERVcaller_tumor	tumor_AITE _sites	tumor_AITE _mean_TE_counts _ERVcaller_tumor	normal_noAITE _sites	normal_noAITE _mean_TE_counts _ERVcaller_normal	normal_AITE _sites	normal_AITE _mean_TE_counts _ERVcaller_normal
TCGA-EM-A1CW	3280	0.458231707317073	1294	0.475270479134467	4391	0.429059439763152	683	0.445095168374817
TCGA-EM-A1CV	3991	0.390628915058882	1367	0.425749817117776	4827	0.404391961881086	950	0.478947368421053
TCGA-EM-A1CU	3240	0.38641975308642	1383	0.449023861171367	5030	0.372166998011928	851	0.482961222091657
TCGA-DO-A1JZ	6715	0.51481757259866	1266	0.587677725118483	6895	0.3464829586657	909	0.413641364136414
TCGA-DD-A1EH	3513	1.11471676629661	708	1.25847457627119	2117	0.911195087387813	395	0.987341772151899
TCGA-DD-A1EG	1210	1.04628099173554	328	1.10365853658537	1197	0.87218045112782	150	0.873333333333333
TCGA-DD-A1EB	3297	1.3130118289354	755	2.09006622516556	3508	1.11431014823261	716	1.59217877094972
TCGA-BK-A0CB	2055	0.310948905109489	591	0.411167512690355	3066	0.440313111545988	211	0.483412322274882
TCGA-BJ-A3PU	2547	0.331762858264625	704	0.352272727272727	4874	0.43783340172343	684	0.460526315789474
TCGA-BJ-A3PR	3519	0.381926683716965	1262	0.499207606973059	2171	0.432980193459235	741	0.622132253711201
TCGA-AA-3514	551	2.48457350272232	179	2.73743016759777	4479	0.149810225496763	725	0.237241379310345
TCGA-A6-2683	63	0.952380952380952	15	1.4	735	0.749659863945578	79	0.822784810126582
TCGA-44-6778	1969	0.525139664804469	423	0.501182033096927	2970	0.43973063973064	224	0.410714285714286
TCGA-44-6777	1315	0.428136882129278	625	0.416	2003	0.398402396405392	267	0.561797752808989
TCGA-44-6776	2624	0.427210365853659	430	0.386046511627907	3361	0.344242784885451	263	0.368821292775665
TCGA-44-6148	4462	1.82025997310623	469	2.08742004264392	4988	0.326784282277466	586	0.387372013651877
TCGA-44-6147	1405	1.37224199288256	608	1.3125	2075	0.288674698795181	184	0.21195652173913
TCGA-44-6145	382	0.0104712041884817	365	0.0136986301369863	1275	0.305882352941176	125	0.344