

UC Irvine

UC Irvine Previously Published Works

Title

RNA-Sequencing Combined With Genome-Wide Allele-Specific Expression Patterning Identifies ZNF44 Variants as a Potential New Driver Gene for Pediatric Neuroblastoma

Permalink

<https://escholarship.org/uc/item/51p0s0t2>

Authors

Sun, Lan
Li, Xiaoqing
Tu, Lingli
[et al.](#)

Publication Date


2023-04-01

DOI

10.1177/10732748231175017

Peer reviewed

RNA-Sequencing Combined With Genome-Wide Allele-Specific Expression Patterning Identifies ZNF44 Variants as a Potential New Driver Gene for Pediatric Neuroblastoma

Cancer Control
Volume 30: 1–10
© The Author(s) 2023
Article reuse guidelines:
sagepub.com/journals-permissions
DOI: 10.1177/10732748231175017
journals.sagepub.com/home/ccx


Lan Sun, MD-PhD^{1,*}, Xiaoqing Li^{1,*}, Lingli Tu², Andres Stucky, PhD², Chuan Huang¹, Xuelian Chen², Jin Cai, MD³, and Shengwen Calvin Li, PhD^{4,5} 

Abstract

Introduction: Neuroblastoma (NB) is one of the children's most common solid tumors, accounting for approximately 8% of pediatric malignancies and 15% of childhood cancer deaths. Somatic mutations in several genes, such as *ALK*, have been associated with NB progression and can facilitate the discovery of novel therapeutic strategies. However, the differential expression of mutated and wild-type alleles on the transcriptome level is poorly studied.

Methods: This study analyzed 219 whole-exome sequencing datasets with somatic mutations detected by MuTect from paired normal and tumor samples.

Results: We prioritized mutations in 8 candidate genes (*RIMS4*, *RUSC2*, *ALK*, *MYCN*, *PTPNI1*, *ALOX12B*, *ZNF44*, and *CNGB1*) as potential driver mutations. We further confirmed the presence of allele-specific expression of the somatic mutations in NB with integrated analysis of 127 RNA-seq samples (of which 85 also had DNA-seq data available), including *MYCN*, *ALK*, and *PTPNI1*. The allele-specific expression of mutations suggests that the same somatic mutation may have different effects on the clinical outcomes of tumors.

Conclusion: Our study suggests 2 novel variants of *ZNF44* as a novel candidate driver gene for NB.

Keywords

RNA-seq, somatic mutation, neuroblastoma, whole-exome sequencing, allele-specific expression pattern, ZNF44

Received November 18, 2022. Received revised March 24, 2023. Accepted for publication April 23, 2023.

¹Department of Oncology, Bishan Hospital of Chongqing Medical University, The People's Hospital of Bishan District, Chongqing, China

²Department of Otolaryngology, Keck School of Medicine, University of Southern California, Los Angeles, CA, USA

³Department of Oral and Maxillofacial Surgery, Zhuhai People's Hospital, Zhuhai Hospital Affiliated with Jinan University, Zhuhai, China

⁴Neuro-Oncology and Stem Cell Research Laboratory, Center for Neuroscience Research, CHOC Children's Research Institute, Children's Hospital of Orange County (CHOC), Orange, CA, USA

⁵Department of Neurology, University of California, Irvine School of Medicine, Orange, CA, USA

*These authors contributed equally to this work.

Corresponding Authors:

Jin Cai, Department of Oral and Maxillofacial Surgery, Zhuhai People's Hospital, Zhuhai Hospital Affiliated with Jinan University, Zhuhai 519000, China.
Email: cajin70@163.com

Shengwen Calvin Li, PhD, Neuro-oncology and Stem Cell Research Laboratory, Center for Neuroscience Research, **CHOC Children's Research Institute**, Children's Health of Orange County or CHOC, 1201 W. La Veta Ave., Orange, CA 92868-3874, USA. Department of Neurology, Irvine School of Medicine, University of California, Irvine, CA, United States. Tel: 714-509-4964; Fax: 714-509-4318.

Email: shengwen.li@alumni.mssm.edu



Creative Commons Non Commercial CC BY-NC: This article is distributed under the terms of the Creative Commons Attribution-NonCommercial 4.0 License (<https://creativecommons.org/licenses/by-nc/4.0/>) which permits non-commercial use, reproduction and distribution of the work without further permission provided the original work is attributed as specified on the SAGE and

Open Access pages (<https://us.sagepub.com/en-us/nam/open-access-at-sage>).

Introduction

With the development of precision medicine, discriminating genomic factors have gained prognostic and therapeutic implications. Many gene mutations, including somatic and germline mutations, are identified using whole-exome, genome, or transcriptome sequencing. Both have improved our understanding of carcinogenesis and influenced the development of cancer treatment plans, including neuroblastoma (NB). NB is one of the most common solid tumors in children, accounting for approximately 8% of all pediatric malignancies and 15% of childhood cancer deaths.¹ A four-center case-control study indicated that LIN28A SNPs (single nucleotide polymorphisms), especially rs34787247 G>A, may increase NB risk.² In contrast, a multi-center case-control study using 263 cases and 715 controls to examine the association of *NRAS* gene rs2273267 A>T polymorphism reduces NB risk in ethnic Chinese children.³ Nucleotide excision repair ERCC1/XPF genes surfaced in 4 polymorphisms for the genetic variations predisposed to NB risk upon screening Chinese children's 393 cases and 812 controls.⁴ Although these genomic studies have revealed specific features of the mutated genes in high-risk NB, their complicated molecular pathways remain unclear in allele-specific expression patterning of the protein-coding regions of genes.⁵

Whole-exome sequencing (WXS, also known as WES) is a genomic technique that is gradually being optimized to identify mutations in increasing proportions of the protein-coding regions of genes.⁶ It is now routinely used and has revealed some rare and common gene variants in NB.⁷ The high-level amplification of *MYCN* on chromosome 2p24 was found in NB previously.⁸ It occurs in approximately 20% of NB patients and indicates aggressive disease progress and a poor prognosis. Inhibitors that downregulate *MYCN*/*MYC* proteins can suppress NB tumor growth.⁹ Mutations in the tyrosine kinase domain of anaplastic lymphoma kinase (*ALK*) have also been identified in NB.¹⁰ A series of *ALK* tyrosine kinase inhibitors (TKIs) have been approved for use in *ALK*-driven cancers, including NB.^{11,12}

ALK proteins can mediate different signaling outputs due to various properties, such as subcellular localization and protein stability.¹³ Although most *ALK*-driven tumors respond dramatically to *ALK*-TKIs, most patients develop drug resistance.¹⁴ However, the details of the resistance mechanism are not very clear. One thing for sure is that this phenomenon is not entirely due to the status of DNA mutation. That is to say, individual patients with the same somatic mutation can respond differently to targeted therapy. One of the potential reasons for variation in treatment response is the allele-specific expression of somatic mutations to model allele-specific expressions at the gene and SNP levels. Patients carrying the same mutations may express mutated alleles at different levels. Therefore, it is essential to investigate the expression levels of specific alleles at the mRNA level. This pattern is because proteins are made from mRNAs. Mutations in the

DNA may result in very different allele expression patterns, ranging from no expression to dominance of one allele. Whole transcriptome sequencing (WTS, also known as RNA-sequencing or RNA-seq) is an emerging tool for profiling gene transcription and has received broad adoption in cancer genetics, with significant prognostic and therapeutic implications.¹⁵ Therefore, we conducted this study to explore the allelic expression of somatic mutations in NB. The mutations of NB identified through DNA-seq, and RNA-seq were compared to confirm novel mutations in their allelic expression patterns, enabling us to propose new risk factors and potential therapeutic targets for NB.

Methods

Datasets

All of the sequencing data were obtained from the National Cancer Institute (NCI) Office of Cancer Genomics Therapeutically Applicable Research To Generate Effective Treatments (TARGET) NB project (<https://ocg.cancer.gov/programs/target>, assessed on 6 October 2020). The datasets were downloaded from The Cancer Genome Project (TCGA) Genomic Data Commons (GDC) Data Portal (<https://docs.gdc.cancer.gov>, assessed on 6 October 2020) using the GDC data transfer tool (<https://gdc.cancer.gov/access-data/gdc-data-transfer-tool>, assessed on 6 October 2020). A total of 127 RNA-seq and 219 WXS files were downloaded from the site. Of the downloaded sequences, both RNA-seq and WXS were available for 85 patients. In addition, 219 variant call format files (.vcf, specifically WXS.mutect2.raw_somatic_mutations.vcf) created using MuTect2¹⁶ were downloaded for the patients on which WXS was performed to compare to our analysis of the raw sequencing files.

Variant Analysis for WXS Samples

Each VCF file from the WXS samples, ANNOVAR,¹⁷ annotated variants (Figure 1A) was created. For each annotated VCF, we filtered the variants that were not exonic, synonymous SNVs, and those for which the maximum frequency in the population is > .001 in gnomAD.¹⁸ We also filtered out the variants that had flags such as "alt_allele_in_normal," "panel_of_normals," or "germline_risk" in MuTect2¹⁶ and required the remaining variants to have 'PASS' flags in more than one of the 219 WXS samples. Ultimately, using these criteria, we obtained 9 variants of the 8 genes.

Variant Analysis for RNA-Seq Samples

As shown in Figure 1B, we wrote a python pipeline to call variants from 127 RNA-seq samples with BAM files as input [Binary Alignment Map (BAM) is the comprehensive raw data of genome sequencing; BAM is the compressed binary representation of SAM (Sequence Alignment Map),

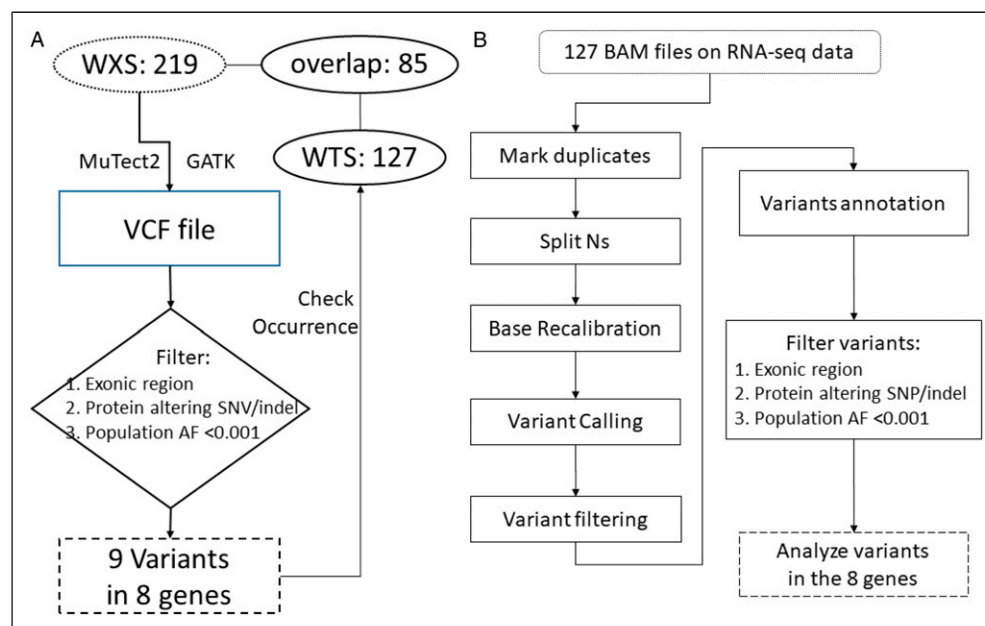


Figure 1. Workflow of the search for variants through filtering and analysis. The whole transcriptome sequencing (WTS): RNA-seq data. Exome sequencing (also known as whole-exome sequencing, WES, or WXS) is a technique for sequencing all the expressed genes in a genome. Population AF is the maximum allele frequency in the population obtained from GnomAD (the genome aggregation database). (A) Pipeline for filtering variants in WXS exome data from 219 patients. (B) Pipeline to call variants from 127 WTS RNA-seq datasets to analyze variants in the 8 genes for which variants meeting the criteria were found in (A). Shapes in dots: Inputs; boxes in dashes: outputs.

a compact and index-able representation of nucleotide sequence alignments]. We first used Picard¹⁹ to mark duplicates and then used Genome Analysis Toolkit (GATK)²⁰ to split Reads with N in Cigar, generated and applied a recalibration table for Base Quality Score Recalibration (BQSR), and used HaplotypeCaller to call variants. After that, variants were filtered using GATK²⁰ with the options “-window 35 -cluster 3 -filterName Filter -filter ‘QD <2.0’ -filterName Filter -filter ‘FS >30.0.’” The filtered variants were annotated with ANNOVAR¹⁷ for further analysis. Similarly, we filtered out non-exonic or synonymous SNV variants with $>.001$ maximum frequency in the population.

Additional methods included assessing the allele-specific or allele-imbalanced gene expression levels and manually examining the alignment files to locate candidates. We also obtained the GTEx data and the cBioPortal (The cBioPortal for Cancer Genomics website, <https://www.cbioportal.org/>, assessed on 6 October 2020) for comparative analyses.

Results

Clinical Characteristics

We extracted 219 WXS-identified samples and 127 WTS-identified samples from the TARGET database. There were 134 independent NB patients in the WXS-identified group and 42 independent individuals in the WTS-identified group, with 85 patients for which we could obtain both WTS and WXS data. As shown in Table 1, clinical characteristics such as

gender, race, and ploidy were similarly distributed in each group. However, *MYCN* status, COG risk, stage, and age at diagnosis were quite differently distributed. All 85 patients for whom we obtained both WTS and WXS results had stage 4 disease classified as high risk and were older than 18 months at diagnosis.

Whole-Exome Sequencing Identifies Candidate Genes

As shown in Figure 1A, we performed a variant analysis for whole-exome sequencing (WXS) on 219 NB samples after variant calling by MuTect2.¹⁶ The purpose was to identify a list of potential disease-predisposing variants and genes. We focused on the list of non-synonymous SNVs and indels in exonic regions with a maximum frequency in the population on Page: $4 \leq .001$ in gnomAD¹⁸ since these variants might be more interpretable and perhaps more likely to be disease-associated. Ultimately, we identified 9 variants in 8 candidate genes: *RIMS4*, *RUSC2*, *ALK*, *MYCN*, *PTPN11*, *ALOX12B*, *ZNF44*, and *CNGB1*. Among them, one variant of *RIMS4* mutated at Chr 20:44758168/C>G was shared by 19 patients. One variant of *RUSC2* mutated at Chr 9:35560530/A>G was shared by 8 patients. Two *ALK* variants mutated at Chr 9:29209798/C>T and Chr 9:29220829/G>T were shared by 6 and 7 patients, respectively. The remaining variants of *MYCN*, *PTPN11*, *ALOX12B*, *ZNF44*, and *CNGB1* were shared by 2 to 3 patients each (Table 2). We further assessed their allelic expression levels in NB with RNA-Seq analysis.

Table 1. Clinical Characteristics in WXS- and WTS-Identified NB Samples.

Characteristics	WXS	WTS	Overlap
	Number of Patients N = 219, (%)	Number of Patients N = 127, (%)	Number of Patients N = 85, (%)
Gender			
Male	136 (62.1)	74 (58.3)	49 (57.6)
Female	83 (37.9)	53 (41.7)	36 (42.3)
Race			
White	162 (74.0)	90 (70.9)	61 (71.8)
Non-white	57 (26.0)	37 (29.1)	24 (28.2)
Ploidy			
Diploid	103 (47.0)	53 (41.7)	42 (49.4)
Hyperdiploid	116 (53.0)	74 (58.3)	43 (50.6)
MYCN status			
Amplified	75 (34.2)	27 (21.3)	20 (23.5)
Non-amplified	144 (65.8)	100 (78.7)	65 (76.5)
Age at diagnosis			
<18 months	0 (0)	25 (19.7)	0 (0)
≥18 months	219 (100)	102 (80.3)	85 (100)
Stage			
4	219 (100)	104 (81.9)	85 (100)
Not 4	0 (0)	23 (18.1)	0 (0)
COG risk			
High risk	219 (100)	104 (81.9)	85 (100)
Not high risk	0 (0)	23 (18.1)	0 (0)

Table 2. A List of 9 Mutations Detected from MuTect2, Its Functional Impacts (To Protein), Population Frequency, Number of Occurrences in the Current Data Set, and the Respective Alt/Ref Reads Count.

Gene	Chr	Position	Ref	Alt	Patients number
RIMS4	20	44,758,168	C	G	19
RUSC2	9	35,560,530	A	G	8
ALK	2	29,209,798	C	T	7
ALK	2	29,220,829	G	T	6
MYCN	2	15,942,195	C	T	3
PTPN11	12	112,450,394	G	A	2
ALOX12B	17	8,072,868	T	G	2
ZNF44	19	12,273,632	CA	C	2
CNGB1	16	57,962,594	G	GAGCTAGGGGAAGTTGAGGGC	2

RNA-Sequencing Analysis of Candidate Gene Allelic Expression

After these candidate genes were found by WXS, we further used RNA-seq to check their expression levels in the patients for which we had RNA-seq results and to check whether other non-overlapping samples had somatic mutations from the RNA-seq data. Somatic mutations detected in WXS could be expressed at various levels and affect cellular function differently. Therefore, we further investigated the

allelic expression of those somatic mutations by analyzing a cohort of 127 NB patients with RNA-seq data, including 85 patients for whom we also have WXS data. We used GATK to detect variants and then ran ANNOVAR on mutations from the RNA-Seq VCF files to prioritize variants that occurred in the 8 genes listed in [Table 2](#), yielding the results in [Table 3](#). It could be seen that *MYCN* and *ALK*, 2 well-known NB genes, had more variants in RNA-seq samples, as expected. The *ALK* mutation at chr2:29209798/C>T was

Table 3. Variants in the 8 Genes in 127 RNA-seq Neuroblastoma Samples After Filtering. Ref and Alt Have the Same Meanings.

Chr	Position	Ref	Alt	Gene	Pts	Allele Depth: (Ref1, Alt1; Ref2, Alt2; ...)
2	15,940,647	G	T	MYCN	1	49, 24
2	15,942,047	C	T	MYCN	1	87, 177
2	15,942,195	C	T	MYCN	2	30, 93; 119, 171
2	15,945,762	G	A	MYCN	1	480, 173
2	29,193,868	C	T	ALK	1	76, 51
2	29,209,798	C	T	ALK	7	27, 31; 36, 21; 49, 38; 41, 14; 39, 10; 9, 8; 55, 49
2	29,209,873	A	G	ALK	1	30, 40
2	29,214,009	A	C	ALK	1	41, 38
2	29,220,829	G	T	ALK	2	56, 55; 147, 150
2	29,220,830	A	C	ALK	1	232, 196
2	29,220,831	A	T	ALK	1	44, 47
2	29,222,350	A	T	ALK	1	24, 52
2	29,222,362	A	C	ALK	1	72, 54
2	29,275,222	C	T	ALK	1	34, 14
2	29,920,310	G	C	ALK	1	41, 45
9	35,547,078	G	C	RUSC2	1	28, 3
9	35,547,698	A	T	RUSC2	1	4100
9	35,556,042	G	A	RUSC2	1	144, 131
9	35,560,298	G	A	RUSC2	1	295, 240
9	35,560,525	C	A	RUSC2	1	296, 114
9	35,561,311	G	C	RUSC2	1	166, 173
16	57,884,228	G	A	CNGB1	1	171, 194
16	57,917,374	T	C	CNGB1	1	86, 68
19	12,273,099	C	T	ZNF44	1	103, 86
19	12,273,632	C	CA	ZNF44	8	16, 5; 34, 10; 70, 16; 18, 6; 26, 10; 83, 21; 28, 8; 19, 9

predominant (7/127, 5.5%). In addition to the 7 samples with chr2:29209798/C>T (Figure 2A and Supplementary Figure S4), there were 2 samples with chr2:29220829/G>T (Figure 2B). *MYCN* has a variant at chr2:15942195, which occurred in 3 WXS samples and 2 RNA-seq samples (Supplementary Figure S5). In particular, the 2 RNA-seq samples both had higher levels of variant alleles than normal alleles (~3-fold for one sample and ~1.5-fold for the other sample), as shown in Table 3, indicating allele-specific expression of these variants. Besides *MYCN* and *ALK*, *RUSC2* (RUN and SH3 Domain Containing 2), *CNGB1* (Cyclic Nucleotide Gated Channel Subunit Beta 1), and *ZNF44* (Zinc Finger Protein 44) also had variants that met our criteria (AF_popmax<.001, exonic, not synonymous_SNV). However, besides those variants in WXS, all other variants only appeared once in the RNA-seq data (Table 3). Interestingly, in 1 patient (TARGET-30-PATGJU) who harbored 2 different *RUSC2* variants, the allelic expression of normal vs mutated *RUSC2* at chr 9:35547078/G>C was 28 vs 3, which is quite different from another SNV at chr 9: 35547698/A>T with a ratio of 4 vs 100. Therefore, the allele-specific expression of mutated SNVs cannot be explained by the ratio of normal to tumor cells.

Moreover, among these mutated sites, we observed 2 novel variants of *ZNF44*, which were not previously reported as associated with NB. Notably, the prevalence of the *ZNF44* mutated allele at chr19: 12, 273, 632/C>CA (Figure 3) was much higher (8/127, 6.3%) than at other sites. After analyzing the clinical data, we found that the average tumor percentage in all samples was 80%, ranging from 60% to 90%. The average stroma ratio was approximately 20%, ranging from 10% to 40%, consistent with the previous report.²¹ Thus, the change in *ZNF44* expression was comparable and reliable across patients. As shown in Table 3, the allelic depths of 8 patients (normal vs tumor) were 16/5, 34/10, 70/16, 18/6, 26/10, 83/21, 28/8, and 19/9. The average fold change (normal vs tumor) was 3.0, ranging from 2.1 to 4.4.

Discussion

NB is a solid tumor that can develop from immature nerve cells in several areas of the body. It most commonly affects children and rarely occurs in adults.¹ This study analyzed WXS and RNA-seq data from NB patients to identify somatic mutations and their allele-specific expression. As most somatic mutations are identified from DNA-seq techniques such as WXS, the allelic expression of those mutations is often unknown. Proteins, the functional units of a live cell, are made from mRNA, so a somatic

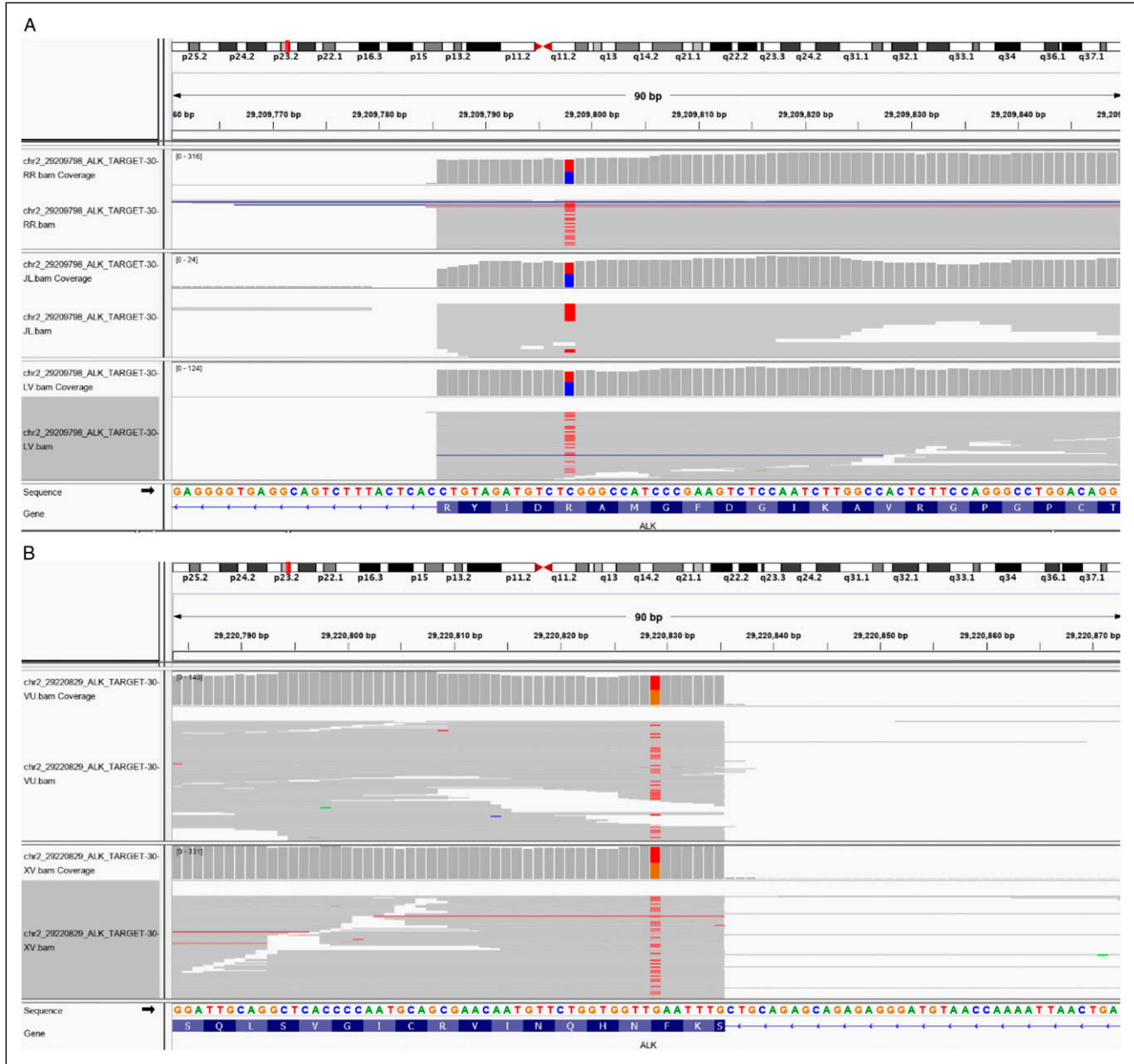


Figure 2. Patterns of 2 *ALK* variants in the *ALK* gene using IGV plot from RNA-seq data: (A) chr2:29209798/C>T with 3 out of 7 patient's samples (the other 4 samples are shown in [Supplementary Figure S4](#)); (B): chr2:29220829/G>T with 2 patient's samples.

mutation may have very different effects on the cellular function that vary with its allelic expression profile. Our study confirmed multiple known NB mutations and identified *ZNF44* (zinc finger protein 44) as a potentially actionable somatic mutation.

Our study explored 2 cohorts of NB patients with either WXS or WTS data available. The overlapping rates of the 2 cohorts were high: 38.8% (85/219) of the WXS group and 66.9% (85/127) of the WTS (Whole transcriptome sequencing) group. It has been reported that using WXS identification, mutation frequencies of somatic genes, including *ALK* (9.2% of cases), *PTPN11* (2.9%), *ATRX* (2.5%), *MYCN* (1.7%), and *NRAS* (.83%) are significant in NB.²² As expected, WXS

revealed mutations in some of these genes, including *ALK*, *MYCN*, and *PTPN11*. The *MYCN* mutation rate (1.4%, 3/219) was similar to a previous report. However, the *ALK* (3.2%, 7/219) and *PTPN11* (1%, 2/219) mutation rates were lower than previously reported.²³ Notably, in our study, *RUSC2* presented the second-highest mutation frequency (3.7%, 8/219) in the analysis of WXS and was identified by WTS as well. *RUSC2* is a protein-coding gene with mutations associated with mental retardation and microcephaly. *CNGB1* and *ZNF44* variants were revealed by both WXS and WTS as well. Although *CNGB1* and *ZNF44* are not commonly associated with NB, they both presented similar mutation frequencies (1%, 2/

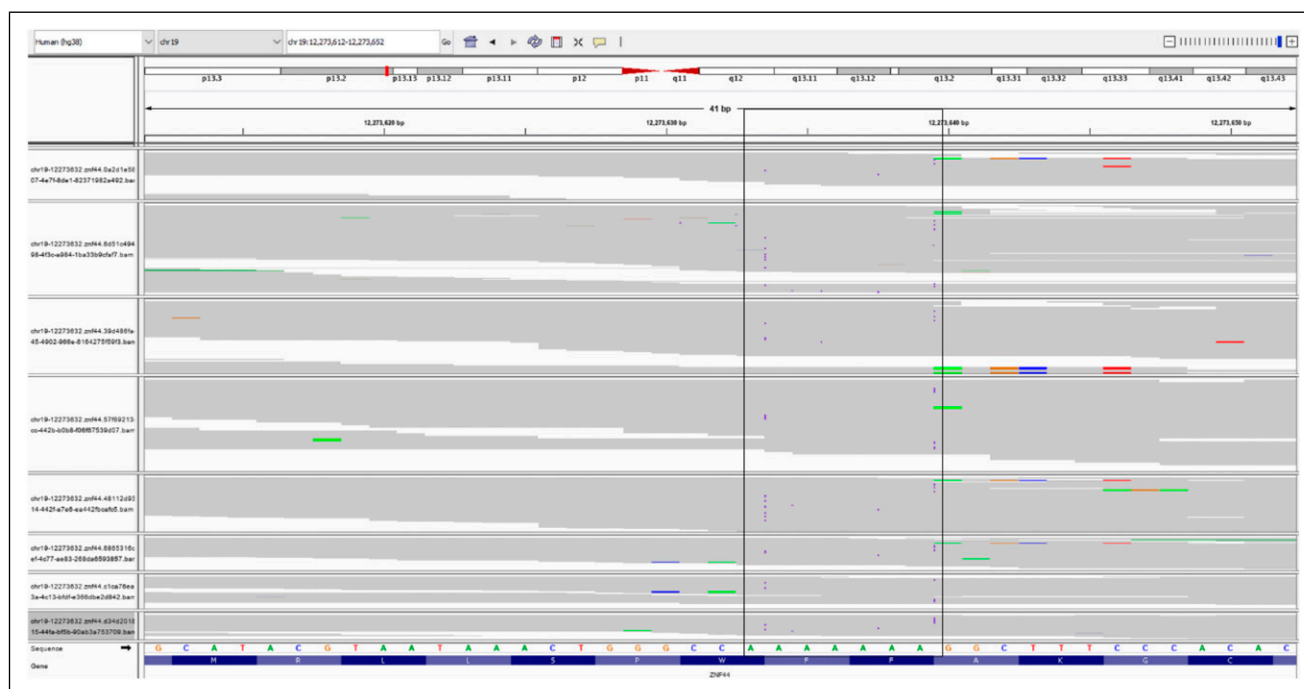


Figure 3. ZNF44 variant patterns by IGV plot for chr19:12,273,632/C>CA in the ZNF44 gene from 8 RNA-seq datasets. The box in black shows a region where insertions happen and which should result in the same insertion interpretation due to the polyA region.

219) to *MYCN* in the WXS analyzed in this study. *ZNF44* encodes a zinc finger protein, also known as a gonadotropin-inducible transcription factor (GIOT-2). *ZNF44* is expressed in human organs and tissues at various levels (Supplementary Figure S1). Mutations in *ZNF44* have been detected in cancers of the uterus, stomach, ovaries, melanoma, cervical, colorectal, lung, bladder, GBM (glioblastoma multiforme), liver, prostate, sarcoma, invasive breast carcinoma, head and neck, etc. (Supplementary Figure S2). However, mutations of this gene have not been reported in neuroendocrine tumors before; to the best of our knowledge, this is the first report on *ZNF44* mutations found in neuroendocrine tumors (Supplementary Figure S3) and on 2 novel variants of *ZNF44* associated with NB [*ZNF44* mutated allele at chr19: 12, 273, 632/C>CA, Figure 3, Tables 2 and 3]. *ZNF44* is reported to be involved in epilepsy susceptibility and binds a factor that is abundant in developing nervous tissue.²⁴ Thus, *ZNF44* may play an as-yet-undetected role in neuroendocrine tumors such as NB. In our study, *ZNF44* mutations were discovered in NB by both WXS and RNA-seq. Specifically, in the analyzed RNA-seq data, the prevalence of *ZNF44* variant chr19: 12, 273, 632/C>CA, was 6.3%. Given that the samples' average tumor and stroma percentages were highly consistent at 80% and 20%, respectively, the *ZNF44* normal allele has expressed an average of 3.0-fold higher than the mutated allele, suggesting the *ZNF44* variant might be a potential disease-related site.

As expected, there were some differences between WTS and WXS analysis. WXS identified a gene *RIMS4* (Regulating Synaptic Membrane Exocytosis 4) with a high

mutation frequency (8.7%, 19/219), but WTS did not detect this. More variants in known disease-related genes *ALK* (Supplementary Figure S4) and *MYCN* (Supplementary Figure S5) were detected in the WTS group. Still, far fewer were found by WXS, suggesting that WTS provided more information on gene variation. WXS and WTS may both be practical biotechnological methods with their advantages. WXS is considered the standard gold method and is routinely used in oncology.⁷ However, it cannot reflect gene expression levels.

WTS has been hailed as a promising approach with distinct advantages, especially for determining transcriptome characteristics.²⁵ However, WTS is not suitable for the discovery of DNA mutations. Thus, combining WXS and WTS can provide complementary perspectives on gene mutations. In our study, an interesting finding from RNA-seq analysis was that in 1 patient sample harboring 2 different *RUSC2* variants, the allelic expression levels of the normal vs mutated SNVs were quite different. Although the samples were from the same patient, the results were opposite, suggesting that the allelic expression levels of the 2 SNVs were not due to different ratios of normal and tumor cells in the sample but due to allele-specific expression. Proteins play bio-functional roles via both biological structure and expression levels. Therefore, both the mutation locations and their allelic expression levels may be related to the response to targeted therapy. However, the limitations of our study include the small sample size, limited clinical information, lack of original raw FASTQ files (any variant) to confirm indel alignment errors, and lack of original

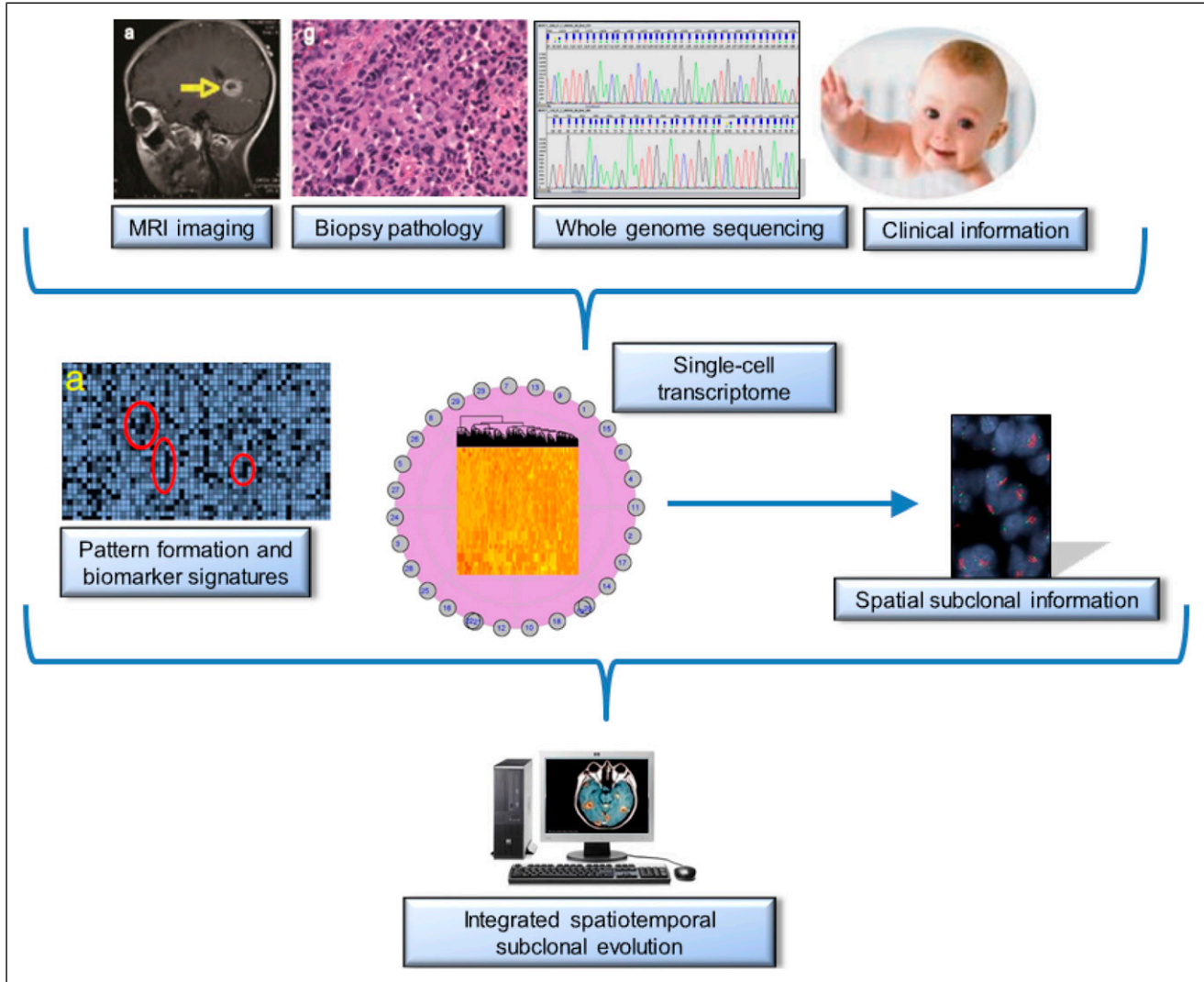


Figure 4. Proposed workflow of pediatric cancer for lifetime management. Collective information is based on the publications²⁶⁻²⁹ related to the single-cell subclonal evolution of pediatric tumors. This workflow might track 2 new variants of ZNF44 and validate them as a novel candidate driver gene for neuroblastoma as details in the discussions.

samples for clinical validation. A further, well-designed study with a more significant number of samples and clinical details is planned for the future (Figure 4).

Highly relevant to our study, an Italian group aimed to determine the differential genetic landscapes between short survival (SS) and long survival (LS) in high-risk (HR) neuroblastoma (NB) (HR-NB) patients at stage M.⁵ The significant percentage of patients who demonstrate rapid disease progression despite multimodal treatment presents one of the biggest problems for oncologists treating high-risk (HR-NB) patients. About 60% of these HR-NBs develop fatal conditions within 5 years of diagnosis. They focused on a cohort of stage M NB patients from the Italian NB Registry with complete clinical data, and follow-up over 10 years was considered, including SS (n = 14) and LS (n = 15). They found ZNF44 mutations in only 2 SS patients (#1965, #2578), about

14%, but not in LS patients. The percentage of mutated ZNF44 in the total of patients at the M stage, including SS and LS, is about 6%. They pointed out, “In SS patients, 4 genes (*SMO*, *SMARCA4*, *ZNF44*, and *CHD2*), all known to be expressed in neural tissues, were recurrently mutated and group-specific and carried particularly deleterious variants.” Intriguingly, 5 genes (*CHD2*, *DIDO1*, *KRTAP4-8*, *ZNF44*, and *ZNF91*) with SS-specific recurrence in the Italian cohort were mutated with SS-specificity also in Pugh cohort patients.

Nevertheless, ZNF44 had been barely investigated in previous research, even though the identification of gain-of-function mutations in the *ALK* receptor tyrosine kinase gene (also in our gene list) as the most common cause of familial NB led to the identification of identical somatic mutations and the rapid development of *ALK* as a tractable therapeutic target.³⁰ Similarly, we proposed a workflow of pediatric cancer

for lifetime management with collective information based on the publications²⁶⁻²⁹ related to the single-cell subclonal evolution of pediatric tumors (Figure 4). This workflow might track 2 new variants of ZNF44 and validate them as a novel candidate driver gene for NB. However, the potential transcriptional control of downstream cascades by ZNF44 must be elucidated by either over-expression or knock-down of ZNF44 transcripts in vitro and in vivo. We can explore the DNA binding region or protein partners of ZNF44 by CHIPseq or pull-down assays. The potential loss-of-function of ZNF44 mutation could be illustrated by the 3D protein structure analysis using the alpha-fold-2 platform.

In summary, to date, few studies have explored gene mutations in NB, like simultaneously using both WXS and WTS. Our study revealed that these 2 methods present different perspectives and meaningful results. Specifically, we found that allele-specific expression assessed by RNA-seq can be quite different even for the same gene mutations, which underscores the importance of WTS in cancer research. Furthermore, we identified gene mutations through both methods, validating some well-known NB genes such as *MYCN* and *ALK* and discovering novel candidate genes such as *ZNF44* variants. Importantly, mutations identified in this study in genes such as *ZNF44* may provide new opportunities for diagnosis- and treatment-driven subclonal evolution,²⁷ which is worthy of further investigation (Figure 4). The candidate driver gene ZNF44 remains to be validated in the clinic.

Conclusion

We discovered 2 novel *ZNF44* variants as novel candidate genes in NB.

Acknowledgments

We thank Jiang F. Zhong for his guidance and critical reading of the manuscript.

Author Contributions

Writing—initial draft preparation, XL and LS. Writing—review and revision, SCL; Conceptualization, JC, SCL; Methodology and formal analysis, LT and CH; Software and data curation, AS and XC. All authors have read and agreed to the published version of the manuscript.

Declaration of Conflicting Interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This study

is partly supported by the Natural Science Foundation of Chongqing, China (cstc2020jcyj-msxmX1063). This work was also supported in part by the CHOC Children's-UC Irvine Child Health Research Awards #16004004, CHOC-UCI Child Health Research Grant #16004003, and CHOC CSO Grant #16986004.

Ethical Approval

The data sets came from a public database without ethical issues. Specifically, all of the sequencing data were obtained from the National Cancer Institute (NCI) Office of Cancer Genomics Therapeutically Applicable Research To Generate Effective Treatments (TARGET) neuroblastoma project (<https://ocg.cancer.gov/programs/target>, assessed on 6 October 2020). The datasets were downloaded from The Cancer Genome Project (TCGA) Genomic Data Commons (GDC) Data Portal (<https://docs.gdc.cancer.gov>, assessed on 6 October 2020) using the GDC data transfer tool (<https://gdc.cancer.gov/access-data/gdc-data-transfer-tool>, assessed on 6 October 2020).

Data Availability

The results published here are based on data generated by the Therapeutically Applicable Research to Generate Effective Treatments (<https://ocg.cancer.gov/programs/target>, assessed on 6 October 2020) initiative phs000467. The data used for this analysis are available at <https://portal.gdc.cancer.gov/projects>.

ORCID iD

Shengwen C. Li  <https://orcid.org/0000-0002-9699-9204>

Supplemental Material

The results published here are based upon data generated by the Therapeutically Applicable Research to Generate Effective Treatments (<https://ocg.cancer.gov/programs/target>) initiative, phs000467. The data used for this analysis are available at <https://portal.gdc.cancer.gov/projects>. Supplemental material for this article is available online.

References

1. Li J, Thompson TD, Miller JW, Pollack LA, Stewart SL. Cancer incidence among children and adolescents in the United States, 2001-2003. *Pediatrics*. 2008;121(6):e1470-e1477.
2. Hua RX, Zhuo Z, Ge L, Zhu J, Yuan L, Chen C, et al. LIN28A gene polymorphisms modify neuroblastoma susceptibility: A four-centre case-control study. *J Cell Mol Med*. 2020;24(1):1059-1066.
3. Li S, Zhuo Z, Chang X, Ma Y, Zhou H, Zhang J, et al. NRAS rs2273267 A>T polymorphism reduces neuroblastoma risk in Chinese children. *Gene*. 2020;727:144262.
4. Zhuo ZJ, Liu W, Zhang J, Zhu J, Zhang R, Tang J, et al. Functional polymorphisms at ERCC1/XPF genes confer neuroblastoma risk in Chinese children. *EBioMedicine*. 2018;30:113-119.
5. Esposito MR, Binatti A, Pantile M, Coppe A, Mazzocco K, Longo L, et al. Somatic mutations in specific and connected subpathways are associated with short neuroblastoma patients' survival and indicate proteins targetable at onset of disease. *Int J Cancer*. 2018;143(10):2525-2536.

6. Belkadi A, Bolze A, Itan Y, Cobat A, Vincent QB, Antipenko A, et al. Whole-genome sequencing is more powerful than whole-exome sequencing for detecting exome variants. *Proc Natl Acad Sci USA*. 2015;112(17):5473-5478.
7. van Wezel EM, Zwijnenburg D, Zappeij-Kannegieter L, Bus E, van Noesel MM, Molenaar JJ, et al. Whole-genome sequencing identifies patient-specific DNA minimal residual disease markers in neuroblastoma. *J Mol Diagn*. 2015;17(1):43-52.
8. Schwab M, Varmus HE, Bishop JM, Grzeschik KH, Naylor SL, Sakaguchi AY, et al. Chromosome localization in normal human cells and neuroblastomas of a gene related to c-myc. *Nature*. 1984;308(5956):288-291.
9. Taylor JS, Zeki J, Ornell K, Coburn J, Shimada H, Ikegaki N, et al. Down-regulation of MYCN protein by CX-5461 leads to neuroblastoma tumor growth suppression. *J Pediatr Surg*. 2019;54(6):1192-1197.
10. Janoueix-Lerosey I, Lequin D, Brugieres L, Ribeiro A, de Pontual L, Combaret V, et al. Somatic and germline activating mutations of the ALK kinase receptor in neuroblastoma. *Nature*. 2008;455(7215):967-970.
11. Umapathy G, Mendoza-Garcia P, Hallberg B, Palmer RH. Targeting anaplastic lymphoma kinase in neuroblastoma. *AP-MIS*. 2019;127(5):288-302.
12. Versteeg R, George RE. Targeting ALK: The ten lives of a tumor. *Cancer Discov*. 2016;6(1):20-21.
13. Hallberg B, Palmer RH. The role of the ALK receptor in cancer biology. *Ann Oncol*. 2016;27(suppl 3):iii4-iii15.
14. Peters S, Zimmermann S. Management of resistance to first-line anaplastic lymphoma kinase tyrosine kinase inhibitor therapy. *Curr Treat Options Oncol*. 2018;19(7):37.
15. Gianfelici V, Chiaretti S, Demeyer S, Di Giacomo F, Messina M, La Starza R, et al. RNA sequencing unravels the genetics of refractory/relapsed T-cell acute lymphoblastic leukemia. Prognostic and therapeutic implications. *Haematologica*. 2016;101(8):941-950.
16. Benjamin D, Sato T, Cibulskis K, Getz G, Stewart C, Lichtenstein L. Calling somatic SNVs and indels with Mutect2. *bioRxiv* 2019. doi:10.1101/861054.
17. Wang K, Li M, Hakonarson H. ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. *Nucleic Acids Res*. 2010;38(16):e164.
18. Karczewski KJ, Francioli LC, Tiao G, Cummings BB, Alföldi J, Wang Q, Genome Aggregation Database Consortium, et al. The mutational constraint spectrum quantified from variation in 141,456 humans. *Nature*. 2020;581(7809):434-443.
19. Broad Institute. *Picard Toolkit, GitHub Repository*. Cambridge, MA: Broad Institute; 2019.
20. Poplin R, Ruano-Rubio V, DePristo MA, Fennell TJ, Carneiro MO, Van der Auwera GA et al. Scaling accurate genetic variant discovery to tens of thousands of samples. *bioRxiv* 2018. doi:10.1101/2011178.
21. Aran D, Sirota M, Butte AJ. Systematic pan-cancer analysis of tumour purity. *Nat Commun*. 2015;6:8971.
22. Pugh TJ, Morozova O, Attiyeh EF, Asgharzadeh S, Wei JS, Auclair D, et al. The genetic landscape of high-risk neuroblastoma. *Nat Genet*. 2013;45(3):279-284.
23. Trigg RM, Turner SD. ALK in neuroblastoma: Biological and therapeutic implications. *Cancers*. 2018;10(4):113.
24. Bassuk AG, Geraghty E, Wu S, Mullen SA, Berkovic SF, Scheffer IE, et al. Deletions of 16p11.2 and 19p13.2 in a family with intellectual disability and generalized epilepsy. *Am J Med Genet*. 2013;161A(7):1722-1725.
25. Xu J, Gong B, Wu L, Thakkar S, Hong H, Tong W. Comprehensive assessments of RNA-seq by the SEQC consortium: FDA-led efforts advance precision medicine. *Pharmaceutics*. 2016;8(1):8.
26. Chen Y, Millstein J, Liu Y, Chen GY, Chen X, Stucky A, et al. Single-cell digital lysates generated by phase-switch microfluidic device reveal transcriptome perturbation of cell cycle. *ACS Nano*. 2018;12(5):4687-4694.
27. Lee LX, Li SC. Hunting down the dominating subclone of cancer stem cells as a potential new therapeutic target in multiple myeloma: An artificial intelligence perspective. *World J Stem Cell*. 2020;12(8):706-720. doi: 10.4252/wjsc.v12.i8.706
28. Li SC, Kabeer MH. Spatiotemporal switching signals for cancer stem cell activation in pediatric origins of adulthood cancer: Towards a watch-and-wait lifetime strategy for cancer treatment. *World J Stem Cell*. 2018;10(2):15-22.
29. Li SC, Vu LT, Ho HW, Yin HZ, Keschrumrus V, Lu Q, et al. Cancer stem cells from a rare form of glioblastoma multiforme involving the neurogenic ventricular wall. *Cancer Cell Int*. 2012;12(1):41-54.
30. Tolbert VP, Coggins GE, Maris JM. Genetic susceptibility to neuroblastoma. *Curr Opin Genet Dev*. 2017;42:81-90. PMC5604862. doi:10.1016/j.gde.2017.03.008