

# RNA-Seq Reveals Allele-Specific Expression of Somatic Mutations in Neuroblastoma

**Xiaoqing Li**

Chongqing Medical University

**Lan Sun** (✉ [ls\\_000@usc.edu](mailto:ls_000@usc.edu))

University of Southern California

**Qian Liu**

Children's Hospital of Philadelphia

**Lingli Tu**

University of Southern California

**Andres Stucky**

University of Southern California

**Jingyi Wang**

Chongqing Medical University

**Zhongjun Wu**

Chongqing Medical University

**Xi Zhang**

University of Southern California

**Kai Wang**

Children's Hospital of Philadelphia

**Jiang F Zhong**

University of Southern California

---

## Research Article

**Keywords:** RNA-Seq, somatic mutation, neuroblastoma, whole-exome sequencing, allele-specific expression

**Posted Date:** December 14th, 2020

**DOI:** <https://doi.org/10.21203/rs.3.rs-119906/v1>

**License:** © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

---

# Abstract

Neuroblastoma (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. Somatic mutations in several genes such as *ALK* have been associated with neuroblastoma progression and can facilitate the discovery of novel therapeutic strategies. However, differential expression of mutated and wild-type alleles on the transcriptome level is not well studied. In this study, we analyzed 219 whole-exome sequencing datasets (with somatic mutations detected by MuTect from paired normal and tumor samples) and prioritized mutations in eight candidate genes as potential driver mutations. Meanwhile, we analyzed 127 RNA-seq samples (of which 85 also had DNA-seq data available) for allele-specific expression levels of each mutation. Our integrated analysis of somatic mutations and allele-specific expression levels confirmed the presence of allele-specific expression of somatic mutations in neuroblastoma including *MYCN*, *ALK* and *PTPN22*. The allele-specific expression of mutations suggests that the same somatic mutation may have different effects on clinical outcomes of tumors. Our study also suggests possible involvement of *ZNF44* as a candidate driver gene for neuroblastoma. In summary, this study demonstrates the value of examining allele-specific expression levels of somatic mutations through the analysis of RNA-Seq data to assess the effects of somatic mutations in different patients. Improved understanding of allele-specific expression of somatic mutations can facilitate development of personalized treatment for neuroblastoma in precision medicine.

## Introduction

With the development of precision medicine, discriminating genomic factors have gained powerful prognostic and therapeutic implications. Nowadays, numerous gene mutations, including somatic and germline mutations, are identified using whole-exome, genome or transcriptome sequencing and have both improved our understanding of carcinogenesis and influenced the development of treatment plans for cancers including neuroblastoma. Neuroblastoma (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 <sup>1</sup>. Although genomic studies have revealed specific biological features of neuroblastoma, the genes mutated in high-risk NB and their complicated molecular pathways remain unclear <sup>2</sup>.

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 <sup>3</sup>. It is now routinely used and has revealed some rare and common gene variants in NB <sup>4</sup>. The high-level amplification of *MYCN* on chromosome 2p24 was found in NB previously <sup>5</sup>. 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 <sup>6</sup>. Mutations in the tyrosine kinase domain of anaplastic lymphoma kinase (ALK) have been identified in NB as well <sup>7</sup>. A series of ALK tyrosine kinase inhibitors (TKIs) have been approved for use in ALK-driven cancers including NB <sup>8,9</sup>. ALK proteins can mediate different signaling outputs due to various properties such as subcellular

localization and protein stability<sup>10</sup>. Although most ALK-driven tumors dramatically respond to ALK-TKIs, most patients eventually develop drug resistance<sup>11</sup>. 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. Patients carrying the same mutations may express the mutated alleles at different levels. Therefore, it is essential to investigate the expression levels of specific alleles at the mRNA level. This is because proteins are made from mRNAs. Mutations in the DNA may result in very different levels of allele expression patterns, ranging from no expression to dominance of one allele or the other. Whole transcriptome sequencing (WTS, also known as RNA-sequencing or RNA-seq) is an emerging tool for profiling gene transcription and has received wide adoption in cancer genetics, with significant prognostic and therapeutic implications<sup>12</sup>. Therefore, we conducted this study to explore allelic expression of somatic mutations in NB. The mutations of neuroblastoma identified through DNA-seq and RNA-seq were compared to identify novel mutations and their allelic expression patterns, enabling us to propose new risk factors and potential therapeutic targets for NB.

## Methods

### Datasets

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>) and downloaded from The Cancer Genome Project (TCGA) Genomic Data Commons (GDC) Data Portal (<https://docs.gdc.cancer.gov>) using the GDC data transfer tool (<https://gdc.cancer.gov/access-data/gdc-data-transfer-tool>). 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 MuTect2<sup>13</sup> were downloaded for the patients on which WXS had been performed, in order to compare to our own analysis of the raw sequencing files.

### Variant analysis for WXS samples

As shown in Fig 1a, for each VCF file from the WXS samples, we used ANNOVAR<sup>14</sup> to annotate variants. For each annotated VCF, we filtered those variants which were not exonic, synonymous SNVs, and those for which the maximum frequency in the population is >0.001 in gnomAD<sup>15</sup>. We also filtered out those variants which had flags such as 'alt\_allele\_in\_normal', 'panel\_of\_normals', or 'germline\_risk' in MuTect2<sup>13</sup>, 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 8 genes.

### Variant analysis for RNA-seq samples

As shown in Fig 1b, we wrote a python pipeline to call variants from 127 RNA-seq samples with BAM files as input. We first used picard<sup>16</sup> to mark duplicates and then used GATK<sup>17</sup> to split Reads with N in Cigar, generated and applied recalibration table for Base Quality Score Recalibration (BQSR), and used HaplotypeCaller to call variants. After that, variants were filtered using GATK<sup>17</sup> 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 ANNOVAR<sup>14</sup> for further analysis. Similarly, we filtered out non-exonic or synonymous SNV variants with >0.001 maximum frequency in population.

Additional methods are needed, for example, how to assess the allele-specific or allele-imbalanced gene expression levels, how to examine the alignment files to manually examine candidate locations, where to get the GTEx data, where to get the cbiportal data for comparative analysis, etc.

## 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 WTS-identified group, with 85 patients for which we were able to 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.

Table 1  
Distribution of clinical characteristics in WXS- and WTS-identified NB samples.

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

## Whole-exome Sequencing Identifies Candidate Genes

As shown in Fig. 1a, we performed variant analysis for whole-exome sequencing on 219 NB samples after variant calling by MuTect2<sup>13</sup>. 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 of  $\leq 0.001$  in gnomAD<sup>15</sup>, since these variants might be more interpretable and perhaps more likely to be disease-associated. Ultimately, we identified nine variants in eight 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; and 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 2

A list of 9 mutations that are detected from MuTect2, its functional impacts (to protein), its population frequency, its number of occurrence 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 whole exome sequencing, 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 have different effects on cellular function. 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 = shown in Table 3. It can be seen that *MYCN* and *ALK*, two well-known NB genes, had more variants in RNA-seq samples, as expected. Among them,

the *ALK* mutation at chr2:29209798/ C > T was predominant (7/127, 5.5%). In addition to the 7 samples with chr2:29209798/C > T (Fig. 2a and Suppl Fig. 4), there were 2 samples with chr2:29220829/G > T (Fig. 2b). *MYCN* has a variant at chr2:15942195 which occurred in 3 WXS samples and 2 RNA-seq samples (Suppl Fig. 5). In particular, the two 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 < 0.001, exonic, not synonymous\_SNV). However, besides those variants which also occurred in WXS, all other variants only occurred once in the RNA-seq data (Table 3). Interestingly, in one patient (TARGET-30-PATGJU) who harbored two 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.

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	4,100
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



Besides, among these mutated sites, we observed two novel variants of *ZNF44* which were not reported previously as associated with NB. Importantly, the prevalence of the *ZNF44* mutated allele at chr19: 12,273,632/ C > CA (Fig. 3) was much higher (8/127, 6.3%) than other sites. After analysis of clinical data, the average of tumor percentage in all samples was 80%, ranging from 60–90%, and the average of stroma ratio was approximately 20%, ranging from 10–40%. Thus, the change of *ZNF44* expression was comparable and reliable across patients. As shown in Table 3, the allelic depth of 8 patients (normal vs. tumor) was 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

Neuroblastoma 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<sup>1</sup>. In this study, we analyzed both WXS and RNA-seq data from neuroblastoma patients to identify somatic mutations and their allele-specific expression. Because the majority of somatic mutations are identified from DNA-seq techniques such as whole exome sequencing, the allelic expression of those mutations is often not known. Proteins, the functional units of a live cell, are made from mRNA, so a somatic mutation may have very different effects on cellular function that vary with its allelic expression profile. In our study, we confirmed multiple known neuroblastoma mutations and also identified *ZNF44* (zinc finger protein 44) as a potentially actionable somatic mutation.

Our study explored two cohorts of neuroblastoma patients with either WXS or WTS data available. The overlapping rates of the two cohorts were high, 38.8% (85/219) of WXS group and 66.9% (85/127) of WTS 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* (0.83%) are significant in neuroblastoma<sup>18</sup>. 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, *ALK* (3.2%, 7/219) and *PTPN11* (1%, 2/219) mutation rates were lower than previously reported<sup>19</sup>. Notably, in our study, *RUSC2* presented the second highest mutation frequency (3.7%, 8/219) in analysis of WXS, and was identified by WTS as well. *RUSC2* is a protein-coding gene which has 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 neuroblastoma, they both presented similar mutation frequencies (1%, 2/219) to that of *MYCN* in the WXS analyzed in this study. *ZNF44* encodes a zinc finger protein also known as gonadotropin-inducible transcription factor (GIOT-2). *ZNF44* is expressed in human organs and tissues at various levels (Suppl Fig. 1). Mutations in *ZNF44* have been detected cancers of the uterus, stomach, ovaries, and more (Suppl Fig. 2). However, mutations of this gene have not been reported in neuroendocrine tumors before; to the best of our knowledge, it is the first report of *ZNF44* mutation in neuroblastoma (Suppl Fig. 3). *ZNF44* is reported to be involved in epilepsy susceptibility, and binds a factor which is abundant in developing nervous tissue<sup>20</sup>. Thus, *ZNF44* may play an as-yet-undetected role in neuroendocrine tumors like

neuroblastoma. In our study, *ZNF44* mutations were discovered in neuroblastoma 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 average tumor and stroma percentages in samples were highly consistent at 80%/20%, respectively, the *ZNF44* normal allele was 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 this was not detected by WTS. Notably, more variants in known disease-related genes *ALK* (Suppl Fig. 4) and *MYCN* (Suppl Fig. 5) were detected in the WTS group, but far fewer were found by WXS, suggesting WTS provided more information on gene variation. It can be understood that WXS and WTS are both useful bio-technological methods with their own advantages. WXS is considered the gold standard method and is routinely used in oncology <sup>4</sup>. However, it cannot reflect gene expression levels. WTS has been hailed as a promising approach that presents distinct advantages, especially for determining transcriptome characteristics <sup>21</sup>. However, WTS is not suitable for discovery of DNA mutations. Thus, the combination of WXS and WTS can provide complementary perspectives on gene mutations. In our study, an interesting finding from RNA-seq analysis was that, in one patient sample harboring two 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 completely opposite, suggesting that the allelic expression levels of the two SNVs were not due to different ratios of normal and tumor cells in the sample, but due to allele-specific expression. Proteins play their bio-functional roles via both biological structure and expression levels. Therefore, both the mutation locations and their allelic expression levels may be related to response to targeted therapy. However, the limitations of our study include the small sample size, limited clinical information, lack of original raw FASTQ files to confirm indel alignment errors, and lack of original samples for clinical validation. A further, well-designed study with a larger number of samples and clinical details is planned for the future.

In summary, to date few studies have explored gene mutations in neuroblastoma using both WXS and WTS. Our study revealed that these two 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 like *MYCN* and *ALK*, but also discovering novel candidate genes such as *ZNF44*. Importantly, mutations identified in this study in genes such as *ZNF44*, which have not previously been reported in neuroblastoma, may provide new opportunities for diagnosis and treatment that are worthy of further investigation.

## Declarations

## Acknowledgements

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>.

## Competing interests

The authors declare that they have no competing interests.

## Funding

This study is supported in part by the National Institutes of Health (NCI, R01 CA197903 & R01 CA251848), the National Natural Science Foundation of China (81703063), and the Natural Science Foundation of Chongqing, China (cstc2020jcyj-msxmX1063).

## Authors' contributions

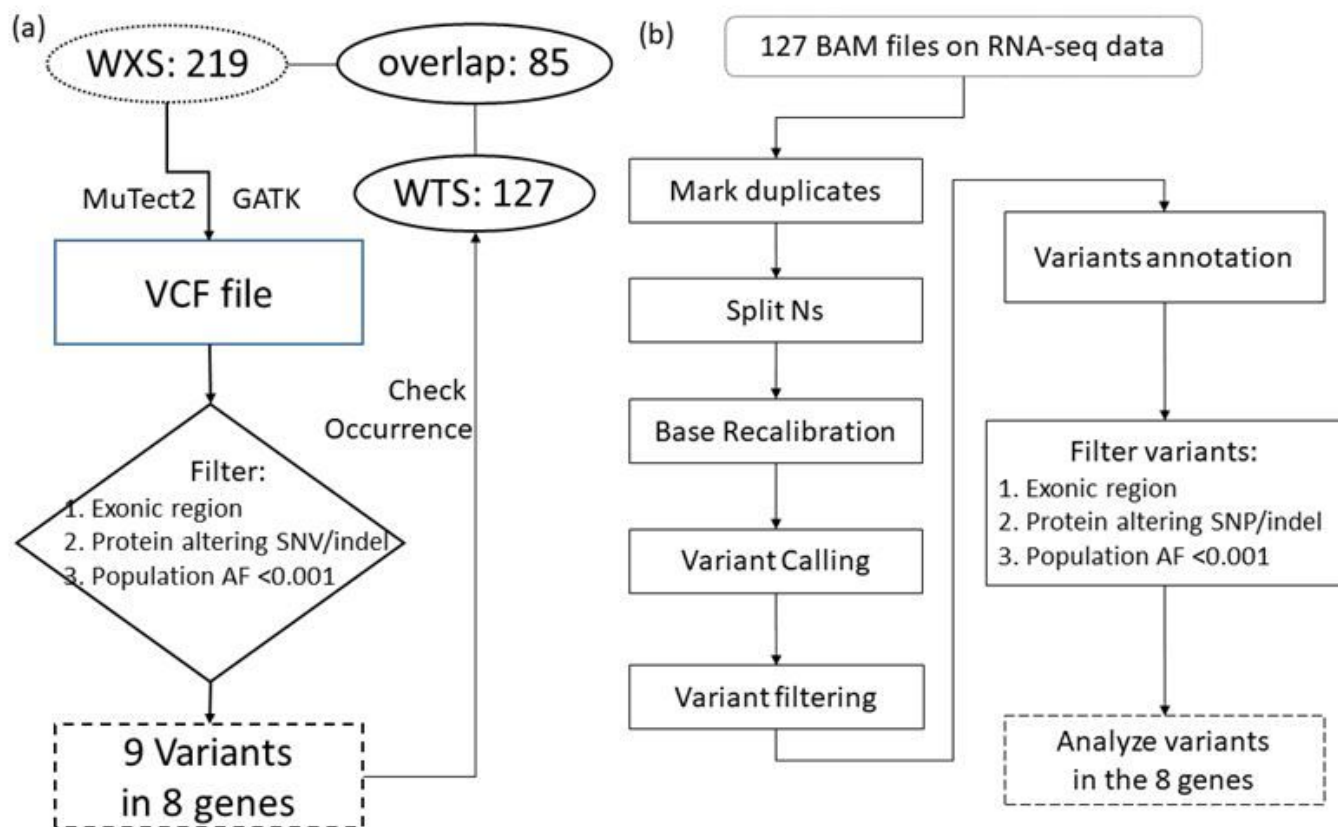
XL and QL performed the study, analyzed and interpreted the data. ZW and LS designed and conceptualized the study. LT, AS, JW, XZ, KW and JFZ prepared Tables, Figures and supplementary information files. All authors wrote and approved the final manuscript.

## References

1. Li, J., Thompson, T. D., Miller, J. W., Pollack, L. A. & Stewart, S. L. Cancer incidence among children and adolescents in the United States, 2001–2003. *Pediatrics* **121**, e1470-1477, doi:10.1542/peds.2007-2964 (2008).
2. Esposito, M. R. *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* **143**, 2525–2536, doi:10.1002/ijc.31748 (2018).
3. Belkadi, A. *et al.* Whole-genome sequencing is more powerful than whole-exome sequencing for detecting exome variants. *Proc Natl Acad Sci U S A* **112**, 5473–5478, doi:10.1073/pnas.1418631112 (2015).
4. van Wezel, E. M. *et al.* Whole-genome sequencing identifies patient-specific DNA minimal residual disease markers in neuroblastoma. *J Mol Diagn* **17**, 43–52, doi:10.1016/j.jmoldx.2014.09.005 (2015).
5. Schwab, M. *et al.* Chromosome localization in normal human cells and neuroblastomas of a gene related to c-myc. *Nature* **308**, 288–291, doi:10.1038/308288a0 (1984).
6. Taylor, J. S. *et al.* Down-regulation of MYCN protein by CX-5461 leads to neuroblastoma tumor growth suppression. *J Pediatr Surg* **54**, 1192–1197, doi:10.1016/j.jpedsurg.2019.02.028 (2019).

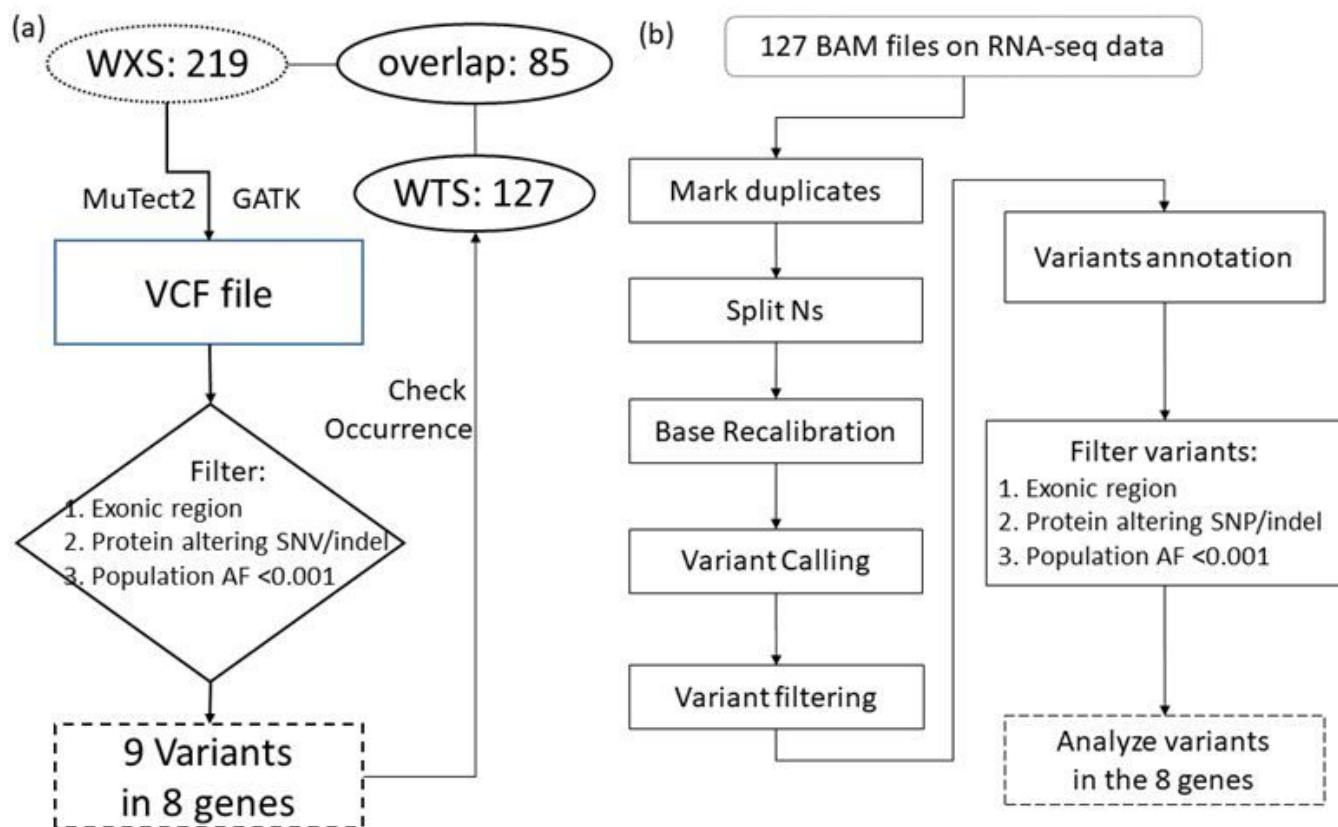
7. Janoueix-Lerosey, I. *et al.* Somatic and germline activating mutations of the ALK kinase receptor in neuroblastoma. *Nature* **455**, 967–970, doi:10.1038/nature07398 (2008).
8. Umapathy, G., Mendoza-Garcia, P., Hallberg, B. & Palmer, R. H. Targeting anaplastic lymphoma kinase in neuroblastoma. *APMIS* **127**, 288–302, doi:10.1111/apm.12940 (2019).
9. Versteeg, R. & George, R. E. Targeting ALK: The Ten Lives of a Tumor. *Cancer Discov* **6**, 20–21, doi:10.1158/2159-8290.CD-15-1411 (2016).
10. Hallberg, B. & Palmer, R. H. The role of the ALK receptor in cancer biology. *Ann Oncol* **27 Suppl 3**, iii4–iii15, doi:10.1093/annonc/mdw301 (2016).
11. Peters, S. & Zimmermann, S. Management of Resistance to First-Line Anaplastic Lymphoma Kinase Tyrosine Kinase Inhibitor Therapy. *Curr Treat Options Oncol* **19**, 37, doi:10.1007/s11864-018-0553-x (2018).
12. Gianfelici, V. *et al.* RNA sequencing unravels the genetics of refractory/relapsed T-cell acute lymphoblastic leukemia. Prognostic and therapeutic implications. *Haematologica* **101**, 941–950, doi:10.3324/haematol.2015.139410 (2016).
13. Benjamin, D. *et al.* Calling Somatic SNVs and Indels with Mutect2. *bioRxiv*, 861054, doi:10.1101/861054 (2019).
14. Wang, K., Li, M. & Hakonarson, H. ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. *Nucleic acids research* **38**, e164–e164, doi:10.1093/nar/gkq603 (2010).
15. Karczewski, K. J. *et al.* The mutational constraint spectrum quantified from variation in 141,456 humans. *Nature* **581**, 434–443, doi:10.1038/s41586-020-2308-7 (2020).
16. Picard toolkit ("Broad Institute", 2019).
17. Poplin, R. *et al.* Scaling accurate genetic variant discovery to tens of thousands of samples. *bioRxiv*, 201178, doi:10.1101/201178 (2018).
18. Pugh, T. J. *et al.* The genetic landscape of high-risk neuroblastoma. *Nat Genet* **45**, 279–284, doi:10.1038/ng.2529 (2013).
19. Trigg, R. M. & Turner, S. D. ALK in Neuroblastoma: Biological and Therapeutic Implications. *Cancers (Basel)* **10**, doi:10.3390/cancers10040113 (2018).
20. Bassuk, A. G. *et al.* Deletions of 16p11.2 and 19p13.2 in a family with intellectual disability and generalized epilepsy. *Am J Med Genet A* **161A**, 1722–1725, doi:10.1002/ajmg.a.35946 (2013).
21. Xu, J. *et al.* Comprehensive Assessments of RNA-seq by the SEQC Consortium: FDA-Led Efforts Advance Precision Medicine. *Pharmaceutics* **8**, doi:10.3390/pharmaceutics8010008 (2016).

## Figures



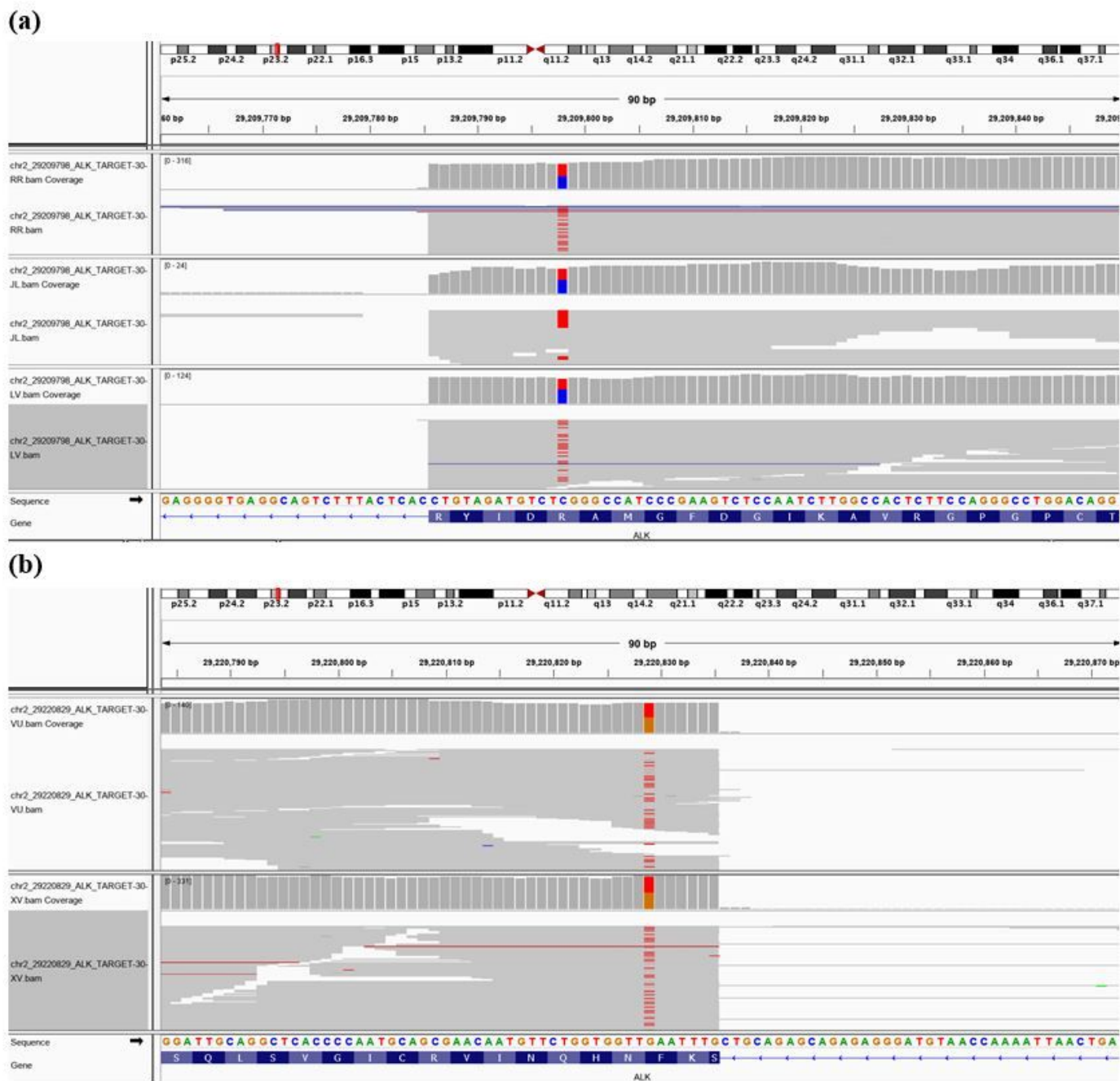
**Figure 1**

Work flow of variant filtering and analysis. WTS: RNA-seq data. WXS: Whole-exome sequencing. Population AF is the maximum allele frequency in the population obtained from GnomAD. (a) Pipeline for filtering variants in WXS data from 219 patients; (b) Pipeline to call variants from 127 WTS 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.



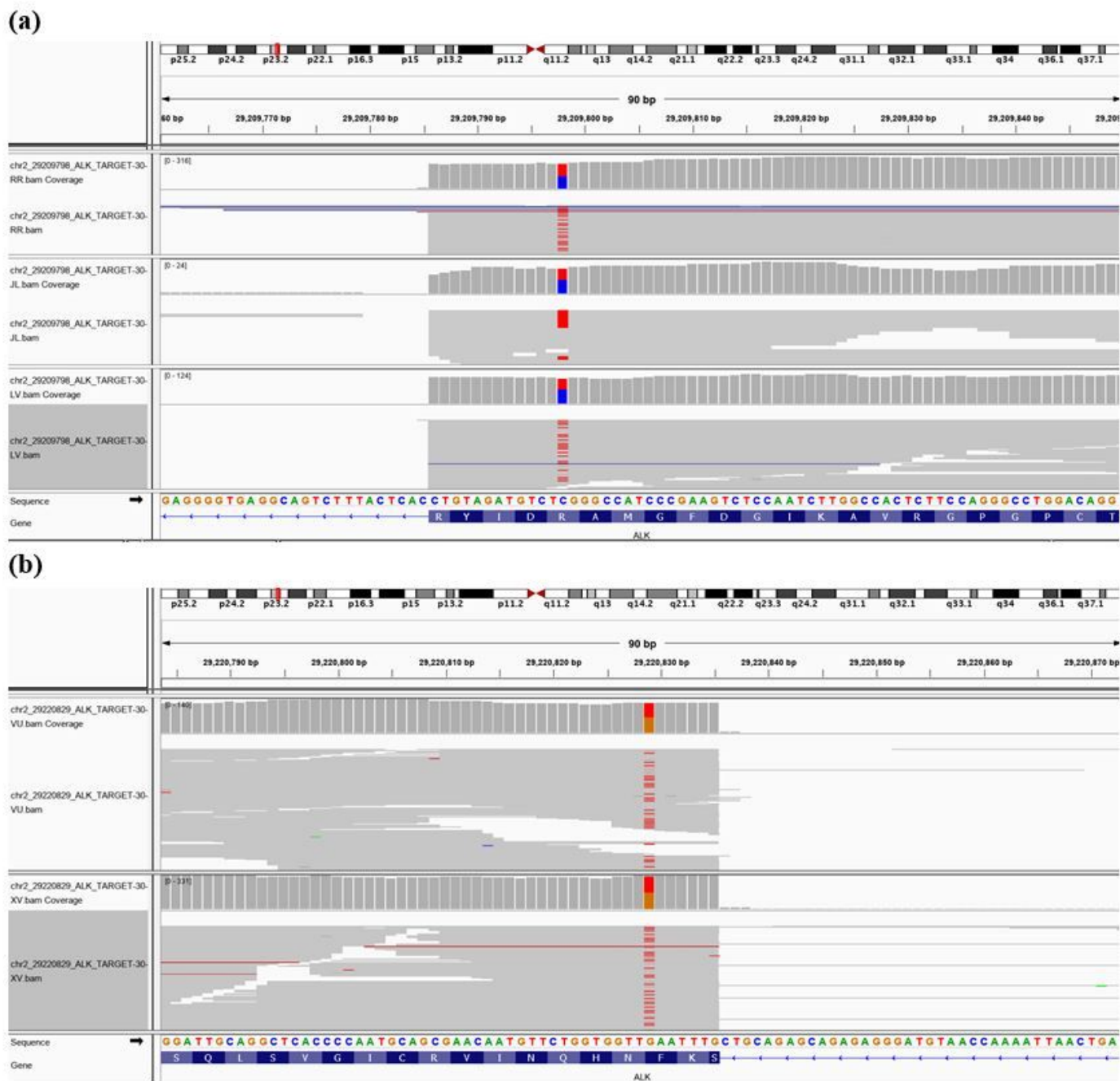
**Figure 1**

Work flow of variant filtering and analysis. WTS: RNA-seq data. WXS: Whole-exome sequencing. Population AF is the maximum allele frequency in the population obtained from GnomAD. (a) Pipeline for filtering variants in WXS data from 219 patients; (b) Pipeline to call variants from 127 WTS 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.



**Figure 2**

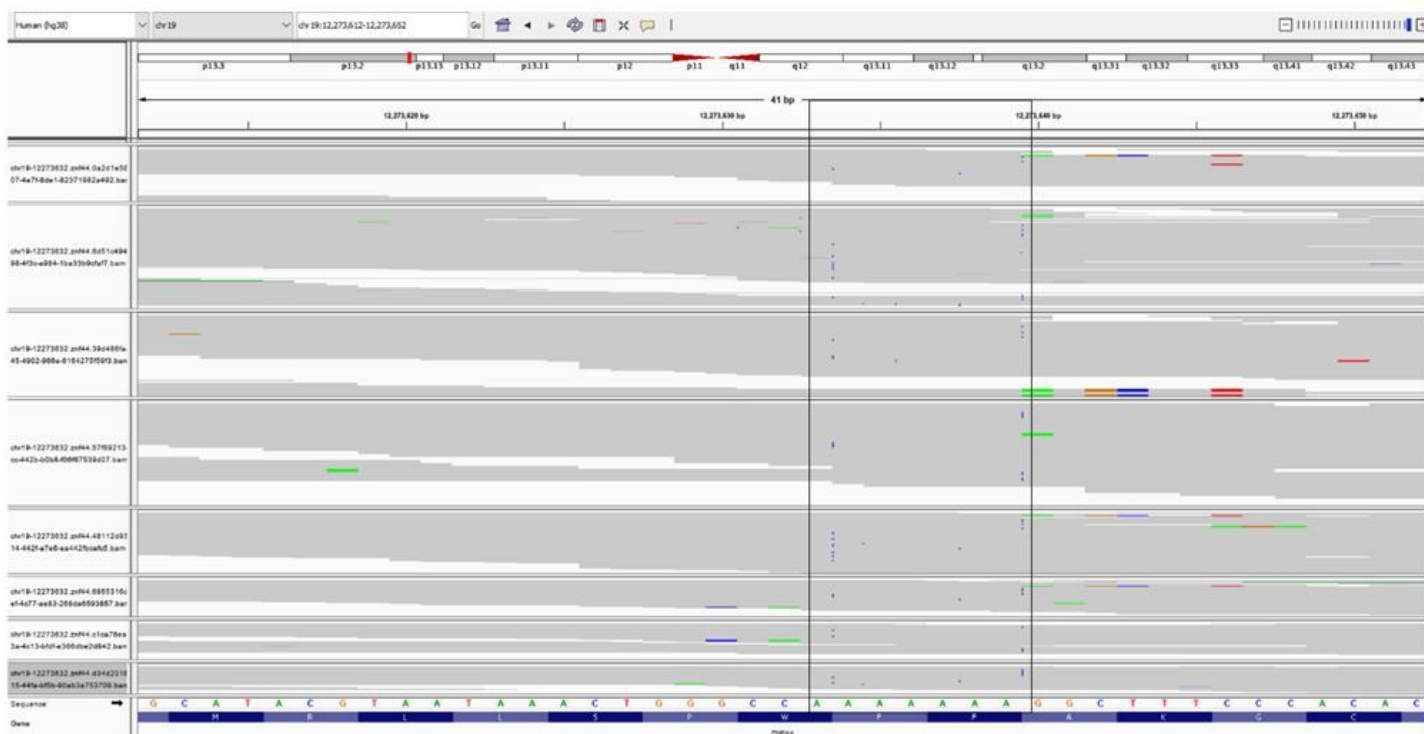
IGV plot for 2 variants in the ALK gene from RNA-seq data: (a): chr2:29209798/C>T with 3 samples out of the total of 7 samples; (b): chr2:29220829/G>T with 2 samples.



**Figure 2**

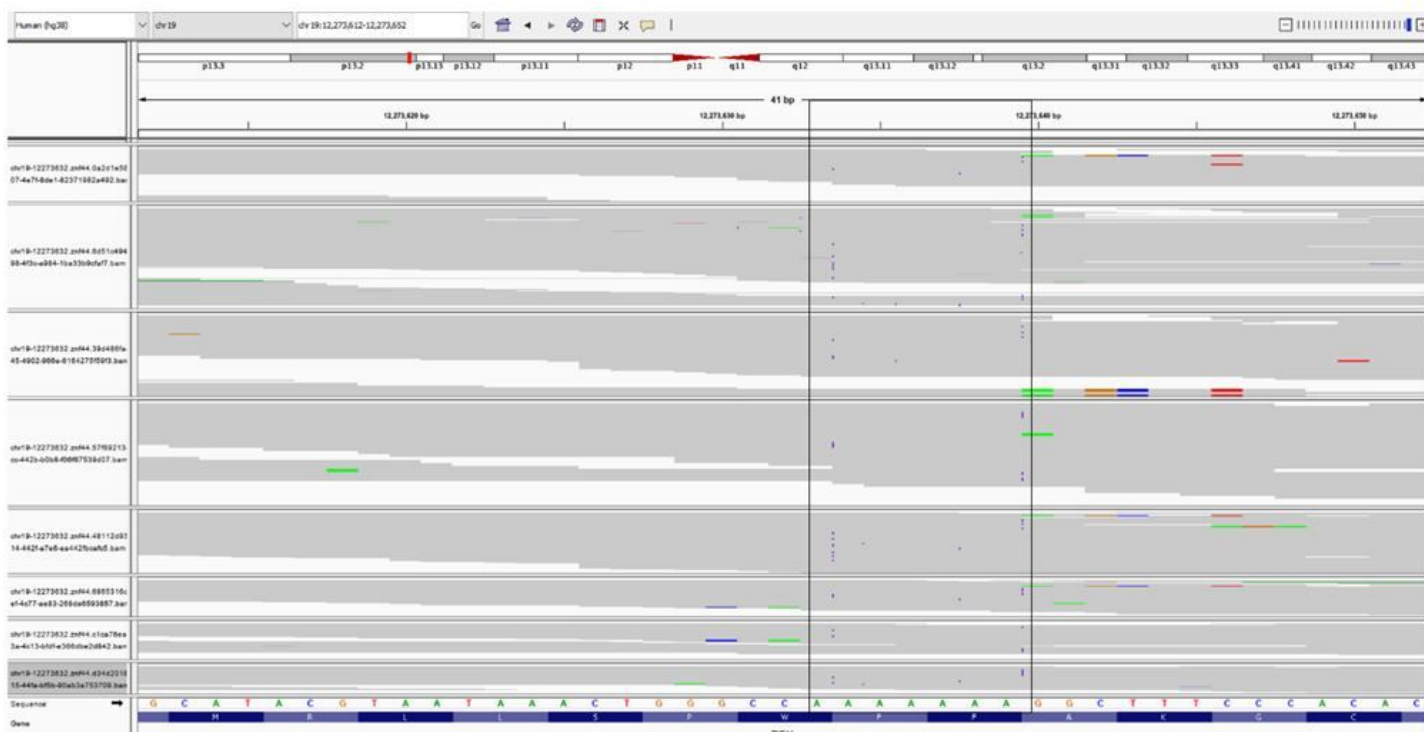
IGV plot for 2 variants in the ALK gene from RNA-seq data: (a): chr2:29209798/C>T with 3 samples out of the total of 7 samples; (b): chr2:29220829/G>T with 2 samples.





**Figure 3**

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.



**Figure 3**

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.

## Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [Supplimentary.docx](#)
- [Supplimentary.docx](#)