Title

Allele-specific expression reveals genes with recurrent cis-regulatory alterations in high-risk neuroblastoma

Permalink

https://escholarship.org/uc/item/1jn8g811

Journal

Genome Biology, 23(1)

ISSN

1474-760X

Authors

Sen, Arko
Huo, Yuchen
Elster, Jennifer
et al.

Publication Date

2022

DOI

10.1186/s13059-022-02640-y

Peer reviewed

Genome Biology

RESEARCH

Open Access

Allele-specific expression reveals genes with recurrent cis-regulatory alterations in high-risk neuroblastoma

Arko Sen[1], Yuchen Huo[2], Jennifer Elster[2,3], Peter E. Zage[2,3] and Graham McVicker[1*]

*Correspondence: gmcvicker@salk.edu
[1] Integrative Biology Laboratory, Salk Institute for Biological Studies, La Jolla, California, USA
Full list of author information is available at the end of the article

Abstract

Background: Neuroblastoma is a pediatric malignancy with a high frequency of metastatic disease at initial diagnosis. Neuroblastoma tumors have few recurrent protein-coding mutations but contain extensive somatic copy number alterations (SCNAs) suggesting that mutations that alter gene dosage are important drivers of tumorigenesis. Here, we analyze allele-specific expression in 96 high-risk neuroblastoma tumors to discover genes impacted by cis-acting mutations that alter dosage.

Results: We identify 1043 genes with recurrent, neuroblastoma-specific allele-specific expression. While most of these genes lie within common SCNA regions, many of them exhibit allele-specific expression in copy neutral samples and these samples are enriched for mutations that are predicted to cause nonsense-mediated decay. Thus, both SCNA and non-SCNA mutations frequently alter gene expression in neuroblastoma. We focus on genes with neuroblastoma-specific allele-specific expression in the absence of SCNAs and find 26 such genes that have reduced expression in stage 4 disease. At least two of these genes have evidence for tumor suppressor activity including the transcription factor *TFAP2B* and the protein tyrosine phosphatase *PTPRH*.

Conclusions: In summary, our allele-specific expression analysis discovers genes that are recurrently dysregulated by both large SCNAs and other cis-acting mutations in high-risk neuroblastoma.

Background

Neuroblastoma is an extracranial solid tumor of the peripheral sympathetic nervous system which accounts for approximately 8% of all childhood cancers and 15% of childhood cancer mortality [1–6]. Compared to other pediatric malignancies, neuroblastomas harbor few recurrent somatic mutations, and most tumors lack identifiable driver mutations in protein-coding genes at the time of initial diagnosis [7]. Instead, neuroblastoma tumors are characterized by frequent somatic copy number alterations (SCNAs). The most common focal SCNA is amplification of the chromosome 2p24 region, including

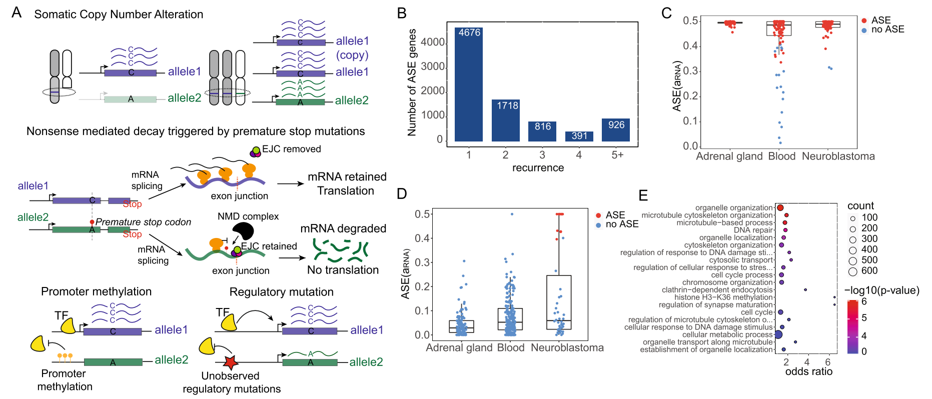

Fig. 1 Allele-specific expression analysis in neuroblastoma. **A** Four example mechanisms that can cause allele-specific expression (ASE) of genes: (i) somatic copy number alterations (deletions or duplications), (ii) premature stop mutations that trigger nonsense mediated decay (NMD), (iii) promoter DNA methylation, and (iv) cis-acting mutations that affect gene regulatory sequences such as enhancers. These examples are not exhaustive and there are additional genomic alterations that can cause ASE. **B** Recurrence of ASE across 96 neuroblastoma samples for genes with ASE in at least one sample (FDR < 0.1). **C** Estimates of ASE (a_{RNA}) for *H19*, an established imprinted gene, in neuroblastoma tumor samples compared to normal adrenal gland and whole blood samples from GTEx. **D** Estimates of ASE for *KIF1B*, a known neuroblastoma tumor-suppressor gene. **E** Gene Ontology analysis of 1043 genes with recurrent ASE in neuroblastoma (NB-ASE genes)

the *MYCN* oncogene, which is associated with high-risk disease and adverse treatment outcomes [8, 9]. Other common SCNAs span tens of megabases and include loss of distal chromosome arms 1p, 3p, and 11q and duplication of the distal arm of chromosome 17q [7, 9–12]. These large SCNAs may drive tumorigenesis by altering the expression of multiple tumor suppressors or oncogenes. For example, chromosome 1p deletions affect many potential tumor suppressors including *CHD5*, *CAMTA1*, *KIF1B*, *CASZ1*, *UBE4B*, and *MIR34A* [13–21]. In addition to the common SCNAs described above, neuroblastoma tumors also contain a patchwork of less common SCNAs or loss of heterozygosity (LOH) regions. A major challenge in interpreting large SCNAs is that they span dozens of genes, making it difficult to distinguish between driver and passenger genes.

Prior studies of genes with altered dosage in neuroblastoma have largely focused on functional characterization of genes affected by SCNAs while disregarding other dosage-altering mutations. Discovery of important driver genes dysregulated by non-SCNA mutations has been limited because only a small number of whole genome sequences for neuroblastoma are available, and it is difficult to determine which noncoding variants affect gene regulation. We hypothesized that genome-wide analysis of allele-specific expression (ASE) could illuminate dysregulated genes in neuroblastoma tumors.

ASE quantifies the difference in expression of two alleles of a gene and can be measured using RNA-seq reads that align to heterozygous sites. Compared to standard differential gene expression analysis, ASE is insensitive to environmental or trans-acting factors, which generally affect both alleles equally. This makes ASE a powerful tool for revealing genes that are affected by cis-acting mutations, including noncoding regulatory mutations that affect sequences such as promoters, enhancers, and insulators as well as protein-coding or splicing mutations that result in nonsense-mediated decay (NMD) (Fig. 1A). Another advantage of ASE is that it is detectable even when the identity of the pathogenic variants causing dysregulation is unknown and it can reveal the effects of rare germline or somatic mutations [22, 23]. Thus, ASE is a powerful tool for the identification of genes with altered gene dosage due to cis-acting genome alterations.

In addition to somatic mutations, ASE can also be caused by common germline polymorphisms [24–26], imprinting [27] or random monoallelic expression [28, 29]; however, these factors are less likely to be involved in tumorigenesis. To discover genes which are dysregulated by pathogenic events, ASE in disease tissue can be compared to either paired-normal tissue or to a large panel of normal tissues to identify cancer-specific gene dysregulation [22, 23, 30, 31]. Genome-wide analysis of ASE therefore has the potential to reveal novel tumor suppressor and oncogenes both within and outside SCNAs.

Results

To discover genes with ASE in neuroblastoma tumors, we obtained exome-seq and RNA-seq data for 96 neuroblastoma tumor samples from the NCI Therapeutically Applicable Research to Generate Effective Treatments (TARGET) project. To estimate ASE in these samples, we implemented a statistical model that utilizes allele-specific read counts at heterozygous sites within the exons of genes, while accounting for genotyping errors, sequencing errors, and overdispersion of RNA-seq read counts. This model estimates allele imbalance (a_{RNA}) for each gene, which is how far the reference allele proportion differs from the expected value of 0.5.

With this method, we identify 8527 genes with ASE in at least one tumor sample under a false discovery rate (FDR) of 10% (likelihood ratio test). Most genes exhibit ASE in only a single sample (4676 out of 8527); however, 3851 genes have ASE in more than one sample, and many genes show highly recurrent ASE in neuroblastoma (926 genes have ASE in 5 or more samples) (Fig. 1B and Additional file 2: Table S1). Since recurrent ASE can result from non-pathogenic factors including common germline polymorphisms [24–26], imprinting [27], or random monoallelic expression [28, 29, 32], we compared the frequency of ASE in neuroblastoma to that of normal tissues, obtained from the genotype tissue expression project (GTEx). Specifically, we compared neuroblastoma ASE estimates to those from normal adrenal gland and whole-blood tissues. These tissues were chosen because the adrenal cortex is the tissue of origin for most neuroblastoma tumors and whole-blood has by far the largest number of available samples in GTEx. To illustrate the utility of comparing normal and tumor tissues, we examined ASE for a well-established imprinted gene, *H19* [33], and for a tumor suppressor gene, *KIF1B*, which is located on chromosome 1p and is frequently deleted in neuroblastoma [16, 34]. As expected, *H19* has very strong ASE in almost all normal and tumor samples (Fig. 1C), whereas ASE of *KIF1B* is observed exclusively in neuroblastoma samples (Fig. 1D).

To define a set of genes with neuroblastoma-specific ASE (NB-ASE), we used two filtering criteria: (a) genes that are testable for ASE in at least 10 neuroblastoma and 10 adrenal gland or blood samples and (b) genes with significant ASE in $\geq$ 3 neuroblastoma tumors and $\leq$1 normal tissue (Additional file 2: Table S2). These criteria resulted in 1043 NB-ASE genes for downstream analysis. We performed a Gene Ontology analysis of these genes and found that they are enriched in biological processes frequently dysregulated during tumorigenesis including microtubule-based process (GO:0007017, p

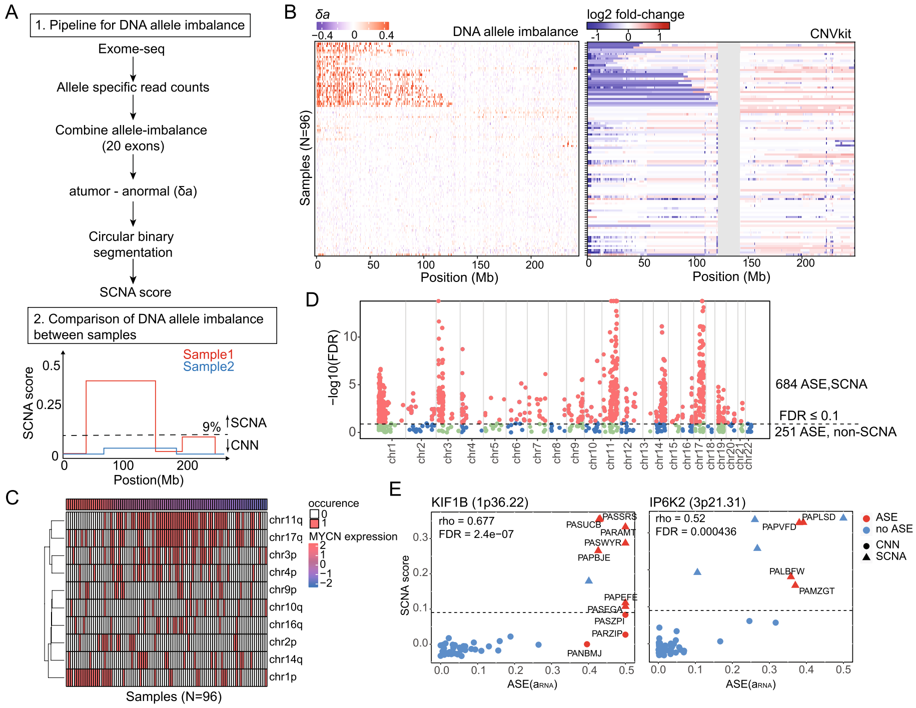

Fig. 2 Detecting somatic copy number alterations in neuroblastoma. **A** Schematic of the DNA allelic imbalance method for detecting somatic copy number alterations (SCNAs). **B** Difference in DNA allelic imbalance between normal and tumor tissues (δ_a) for 96 neuroblastoma patients across chromosome 1 (left panel). Results are compared to log2 fold-change in normalized read coverage between tumor and normal tissues estimated by CNVkit (right panel). **C** Occurrence matrix of the 10 most frequent SCNAs detected using the DNA-imbalance approach. SCNAs were filtered using SCNA score ≥ 0.09 and annotated based on their cytoband location. Neuroblastoma patients are grouped by *MYCN* z-score normalized gene expression. **D** Manhattan plot for Spearman's correlation analysis between ASE (a_{RNA}) and SCNA score for 945 NB-ASE genes. **E** Spearman's rank correlation between allele-specific expression (a_{RNA}) and SCNA score for *KIF1B* (left panel), a tumor suppressor in the chromosome 1p deletion region, and *IP6K2* (right panel), a putative tumor suppressor within the 3p deletion region

value = 2.6e−06), DNA repair (GO:0006281, p-value = 2.5e−05), and cellular metabolic process (GO:0044237, p-value = 0.00074) (Fig. 1E and Additional file 2: Table S3).

SCNAs are a common cause of ASE in tumors [35], and we hypothesized that many NB-ASE genes would be attributable to large-scale SCNAs that dominate the genetic landscape of neuroblastoma [2, 3, 6, 7]. To determine which NB-ASE genes can be attributed to SCNAs, we adapted our ASE framework to identify SCNAs, which are detectable as large genome segments with allelic imbalance of DNA sequencing reads [36]. While several existing tools leverage read depth to predict SCNAs, these methods have limited precision and report many false positive focal SCNAs [37, 38]. To detect SCNAs, we estimated DNA allelic imbalance from heterozygous sites in windows consisting of 20 consecutive exons for tumor (a_{tumor}) and normal (a_{normal}) samples. We then computed the difference in their absolute values (δ_a) and performed circular binary segmentation (CBS) [39] to obtain DNA allelic imbalance for continuous segments which we refer to as the SCNA score (Fig. 2A and Additional file 2: Table S4).

To test our allelic imbalance approach for SCNA discovery, we applied it to chromosome 1, which has distal p arm deletions in ~30% of neuroblastoma tumors [7, 10, 12].

We compared our predictions to those made by CNVkit, which utilizes read depth at exome capture targets to infer copy number and has better sensitivity compared to other methods for SCNA discovery [40]. SCNA breakpoints detected by our method are consistent with those detected by CNVkit, but our predictions are considerably less noisy (Fig. 2B). We also compared our results to those from high density single-nucleotide polymorphism (SNP) arrays, which were available for 33 out of the 96 neuroblastoma tumors [41], and found them to be highly concordant (Additional file 1: Fig. S1). We obtained similar results for other common SCNAs such as the chromosome 11q deletion region (Additional file 1: Fig. S2).

To further examine the SCNAs in neuroblastoma, we partitioned samples based on *MYCN* expression and found known patterns of SCNA co-occurrence [7]. For example, chromosome 1p and 11q deletions occur most frequently in samples with high and low expression of *MYCN*, respectively (Fig. 2C). In addition to the well-characterized SCNAs, we detected less frequent SCNAs across all chromosomes (Fig. 2C) including loss of 16q in 16 neuroblastoma tumors (Additional file 1: Fig. S3). This SCNA has not been extensively studied but has been previously reported by comparative genomic hybridization in familial neuroblastomas and some other pediatric cancers such as Wilm's tumor [42–44]. Our deletion predictions for 16q appear to be true positives because they are concordant with both SNP-array predictions and patterns of ASE (Additional file 1: Fig. S3). In combination, these results indicate that DNA allelic imbalance is a powerful approach for the detection of SCNAs in cancer genomes.

To determine whether general patterns of ASE in neuroblastoma can be attributed to SCNAs, we computed Spearman's correlation between ASE and SCNA score, restricting our analysis to 935 NB-ASE genes that are located within SCNA segments in at least one neuroblastoma sample. Under an FDR of 10%, 65% (684 out of 1043) of NB-ASE genes are significantly correlated with SCNAs, and of these, 59% (401 out 684) are located on the chromosomes with the most frequent SCNAs (chromosomes 1, 3, 11, and 17) (Fig. 2D and Additional file 2: Table S5).

The chromosome 1p, 3p, and 11q deletion regions are hypothesized to contain tumor suppressor genes; however, the identities of the tumor suppressors are difficult to determine because the deletions are large and contain hundreds of genes. We reasoned that, in the absence of large deletions, tumor suppressors within these regions may be affected by other types of genome alterations that affect dosage, and that the effects of these alterations would be detectable by ASE. We examined the relationship between ASE and SCNAs for two genes, *IP6K2* and *KIF1B*, which have been previously identified as potential tumor suppressors located within the chromosome 3p and 1p deletion regions. Knockdown of *IP6K2* impairs apoptosis in colorectal cancer cells [45], and its deletion or low expression is associated with adverse clinical outcomes in aerodigestive tract carcinoma and breast cancer [46, 47]. Overexpression of *KIF1B* in neuroblastoma cell lines causes apoptotic cell death and its knockdown enhances tumor formation in mouse models [34]. In the case of *IP6K2*, we found that every sample with significant ASE also has a high SCNA score (Fig. 2E, right panel), indicating that ASE of *IP6K2* is solely attributable to overlapping chromosome 3p deletions. The pattern exhibited by *KIF1B* is different. While ASE of *KIF1B* is correlated with SCNA score (Spearman's rho = 0.68, FDR corrected *p*-value = 2.4e−07), several samples have strong ASE in the absence of a chromosome 1p deletion (SCNA score $\leq$ 0.09) (Fig. 2E, left panel). Thus, in some samples, ASE of *KIF1B* is caused by factors other than large-scale SCNAs. Several other putative tumor suppressors in the chromosome 1p or 11q deletion regions including *CHD5* [15, 48], *UBE4B* [13], *CADM1* [49], and *ATM* [50], have patterns that are similar to *KIF1B*, where a subset of samples exhibit strong ASE in the absence of deletions (Additional file 1: Fig. S4). These results indicate that both SCNA and non-SCNA genome alterations affect the expression of these genes.

Previous studies have demonstrated that ASE can be caused by nonsense-mediated decay (NMD) [51–53], an evolutionarily conserved mechanism that degrades transcripts with premature termination codons [54–56] (Fig. 1A). We hypothesized that genes that exhibit ASE in neuroblastoma in the absence of SCNAs may contain mutations that cause NMD. To identify somatic mutations that are likely to cause NMD, we analyzed paired tumor-normal exome-seq data with variant effect predictor (VEP), which collectively labels missense, frameshift, and nonsense mutations that are likely to cause NMD as "high-impact." We identified 12,122 unique high-impact mutations in the 96 tumor samples, 886 of which are located within 490 NB-ASE genes. Most of these mutations (788 out of 890) are stop-gain mutations (Fig. 3A and Additional file 2: Table S6) and map to 452 NB-ASE genes.

To determine if stop-gain mutations are enriched within NB-ASE genes, we examined their frequency in three different gene sets: (a) NB-ASE genes, (b) randomly selected genes with at least one somatic mutation, and (c) genes with ASE observed in both neuroblastoma and normal tissues. To create a null distribution, we sampled 100 genes from each gene set 500 times and counted the number of genes carrying stop-gain mutations each sampling iteration. A substantially greater number of NB-ASE genes carry stop-gain mutations, indicating that NMD is an important driver of neuroblastoma-specific ASE (Fig. 3B).

We observed that many genes with correlated ASE and SCNA scores also contain stop-gain mutations in some samples (Fig. 3C), leading us to hypothesize that NMD is an important mechanism that alters gene dosage in samples lacking SCNAs. To test this hypothesis, we partitioned neuroblastoma samples for each gene into three categories: non-ASE samples, ASE samples with SCNAs (ASE_SCNA), and ASE samples without SCNAs (ASE_non-SCNA) (Fig. 3D). We then calculated the rate of stop-gain mutations across all genes and samples in each of the three categories. To generate a null distribution of rates that controls for gene lengths and mutation rate heterogeneity, we permuted the category labels for each gene 1000 times. This analysis revealed that stop-gain mutations occur at a substantially higher rate in ASE_non-SCNA samples compared to other categories (Fig. 3E). This supports the hypothesis that gene expression is often altered by NMD-causing mutations in the samples that lack SCNAs.

An example of a gene which is dysregulated by both NMD and SCNAs is Pleckstrin Homology and RhoGEF Domain Containing G5 (*PLEKHG5*). *PLEKHG5* is located in cytoband 1p36.31 which is frequently deleted in neuroblastoma [57]. Two samples have strong ASE in the absence of SCNAs, one of which is heterozygous for a C>A mutation that introduces a premature stop codon (Fig. 3F). The cause of ASE in the other ASE_non-SCNA sample is unknown and could potentially be discovered by analysis of whole-genome sequencing data.

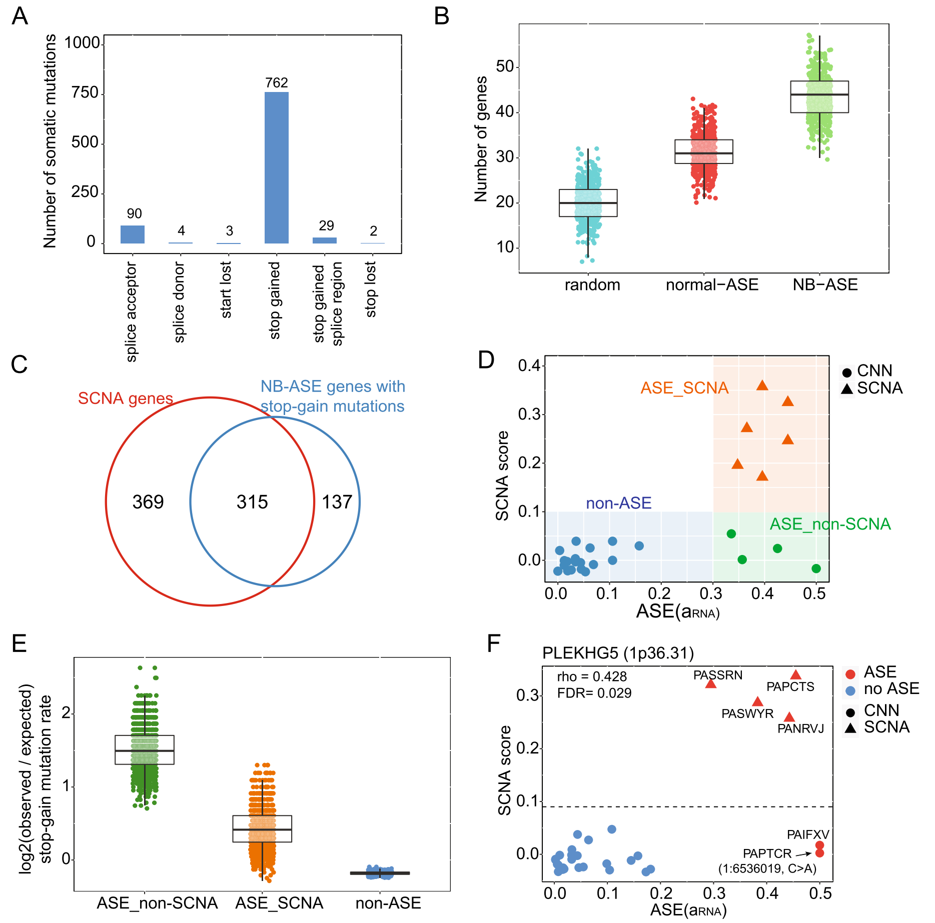

Fig. 3 Genes with recurrent allele-specific expression are enriched for stop-gain mutations that cause nonsense-mediated decay. **A** The number of variants annotated as high-impact by variant effect predictor (VEP) in genes with recurrent allele-specific expression in neuroblastoma (NB-ASE genes) grouped by functional consequence. **B** NB-ASE genes more frequently carry stop-gain mutations. To create the distributions, we performed 500 sampling iterations and counted the number of unique genes carrying stop-gain mutations each iteration. In each iteration, we sampled 100 genes from three gene sets: a random panel of control genes (random), genes with ASE in both neuroblastoma and normal tissues (normal-ASE), and genes with ASE that is specific to neuroblastoma (NB-ASE). **C** Overlap between genes with correlated ASE and somatic copy number alteration (SCNA) score and NB-ASE genes which contain at least one stop-gain mutation. **D** Graphical representation of three categories of neuroblastoma samples we define for a given gene: samples without ASE (non-ASE), samples with ASE and SCNA (ASE_SCNA), and samples with ASE but no SCNA (ASE_non-SCNA). **E** Samples with ASE in the absence of SCNAs are enriched for stop-gain mutations. The boxplots show the log2 ratio of observed to expected stop-gain mutation rate for the three sample categories. Expected rates are estimated by permuting category labels within each gene; points are observed/expected ratios computed across genes after each permutation. **F** Spearman's correlation between ASE (a_{RNA}) and SCNA score for *PLEKHG5*, a chromosome 1p deletion gene. Two samples have high ASE but low SCNA scores and one of these samples, PAPTCR, carries a premature stop mutation (1:6536019, C>A)

Outside of SCNAs, 108 NB-ASE genes are located in genome regions that are copy neutral across all tumor samples, including *PHOX2B* which is a target of recurrent germline mutations in neuroblastoma [58, 59]. In addition, 251 genes lack significant correlations between ASE and SCNA score even though they overlap SCNAs in one or more samples (Additional file 2: Table S5). Thus, 34% of the NB-ASE genes (359 out of 1043) are not associated with SCNAs, suggesting that other mutational events that alter gene dosage are common in neuroblastoma genomes.

We reasoned that the overall expression of NB-ASE genes could be examined in a larger gene expression dataset where ASE measurements are unavailable. We asked whether the 316 non-SCNA ASE genes are associated with neuroblastoma progression and metastasis, by analyzing the SEQC/MAQC-III Consortium dataset, which contains clinical and microarray expression data for 498 neuroblastoma tumors [60]. Under an FDR of 5% (Student's t test) and absolute log2 fold-change $\geq$ 0.5, 34 genes have significantly different gene-expression in stage 4 or metastatic disease compared to other stages (Fig. 4). Among them, 8 genes have increased expression and 26 genes have decreased expression in stage 4 disease. Most notably, *MAP7, PTPRH, TFAP2B*, and *SLC18A1* have more than a 2-fold decrease in expression in stage 4 tumors. We hypothesized that these genes may be important tumor suppressors in neuroblastoma, even though they lie outside of the common SCNA regions of the genome, and we performed further functional analysis of *TFAP2B* and *PTPRH*.

We first investigated *TFAP2B*, which is a retinoic acid-induced transcriptional activator that mediates noradrenergic neuronal differentiation of neuroblastoma cells in vitro [61,

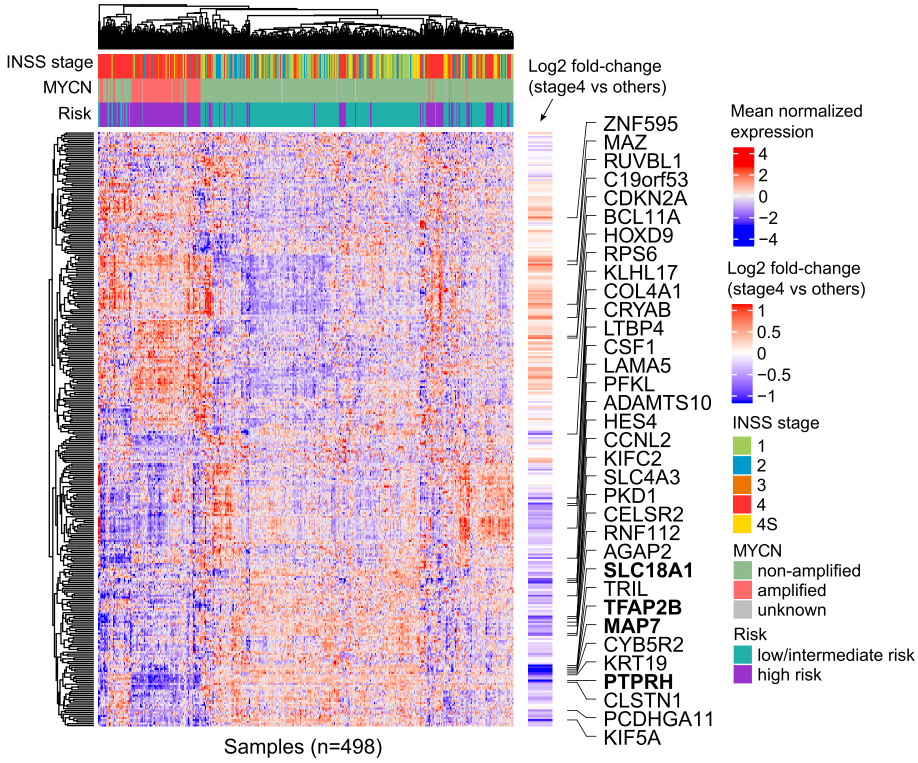

Fig. 4 Expression of genes with recurrent allele-specific expression that is not associated with somatic copy number alterations. The heatmap shows hierarchically clustered gene expression z-scores from 498 neuroblastoma tumors in the SEQC/MAQC-III Consortium. Of the 359 NB-ASE genes that were not associated with somatic copy number alterations (SCNAs), 316 genes were covered by at least one expression probe. Samples are labeled by clinical characteristics: *MYCN* amplification status (amplified, non-amplified and unknown); high-risk tumor (yes or no); International Neuroblastoma Staging System (INSS) (1, 2, 3, 4, and 4S). Difference in gene-expression between stage 4 and all other stages is indicated as log2 fold-change (log2FC). The 34 genes with absolute log2FC $\geq$ 0.5 and FD-corrected p-value $\leq$ 0.05 from Student's t test are labeled

62]. *TFAP2B* has ASE in 3 out of 31 testable neuroblastoma samples, has no evidence of ASE in adrenal gland tissues (0 out of 12 testable samples), is not expressed in whole blood, and is copy neutral in all patient samples (Additional file 1: Fig. S5A). Consistent with the above observations, the samples with ASE of *TFAP2B* have strong allelic imbalance of RNA-seq reads at heterozygous sites, but no allelic imbalance of exome-seq reads, indicating that the ASE is not due to SCNAs (Additional file 1: Fig. S5B). Dysregulation of *TFAP2B* in neuroblastoma cells has previously been associated with aberrant promoter-methylation [62], so we investigated DNA methylation as a potential mechanism. Using estimates of promoter methylation computed from the Human Methylation 450K array, we found that *TFAP2B* is one of the NB-ASE genes with the strongest correlations between ASE and promoter-methylation, although this correlation is not significant under an FDR threshold of 10% (Spearman's correlation coefficient = 0.60, FDR-corrected *p*-value = 0.116) (Additional file 1: Figs. S5C-E, S6 and Additional file 2: Table S7). Furthermore, one patient sample (PASNZU) has near-complete methylation (>75%) of the *TFAP2B* promoter, which is associated with loss-of-expression of both alleles (Additional file 1: Fig. S5C-E and Additional file 2: Table S8). In the SEQC/MAQC-III Consortium data, *TFAP2B* expression is decreased in stage 4 or metastatic neuroblastomas (Additional file 1: Fig. S5F) and low expression of *TFAP2B* is associated with worse event-free survival outcomes in non-*MYCN* amplified neuroblastoma patients (Additional file 1: Fig. S5G). Collectively these observations are consistent with earlier findings [62] and strongly suggest that *TFAP2B* is a tumor-suppressor in neuroblastoma with decreased expression in the presence of promoter methylation.

We next investigated *PTPRH* (Protein Tyrosine Phosphatase Receptor Type H), which is a member of a large family of receptor tyrosine phosphatases and a critical regulator of apoptosis and cell motility [63, 64]. In neuroblastoma, 5 out of 68 testable samples exhibit ASE, compared to 1 out 139 testable adrenal gland tissues from GTEx (Fisher's exact test *p*-value = 0.015) (Fig. 5A); *PTPRH* is not expressed in the whole blood. *PTPRH* is located on chromosome 19q, which rarely undergoes SCNA in neuroblastoma, and ASE of *PTPRH* is not correlated with SCNA score (Fig. 5B). In addition, the RNA-seq reads in ASE samples exhibit strong allelic imbalance but there is no allelic imbalance in exome-seq reads, confirming that ASE of *PTPRH* is not attributable to large or focal SCNAs (Fig. 5C). ASE of *PTPRH* is negatively correlated with gene-expression (Spearman's correlation coefficient = −0.58, FDR-corrected *p*-value = 2.3e−05) indicating that ASE reflects loss of expression of one allele, potentially due to regulatory or other cis-acting mutations (Fig. 5D and Additional file 2: Table S8). Gene expression of *PTPRH* is substantially reduced in stage 4 tumors, and reduced expression of this gene is associated with worse event-free survival outcomes in non-MYCN amplified tumors (Fig. 5E, F). To further test the function of *PTPRH*, we performed shRNA knockdown experiments in the neuroblastoma cell lines SK-N-SH and SK-N-BE(2), which are *MYCN* non-amplified and amplified, respectively (Additional file 1: Fig. S7). Knockdown of *PTPRH* increases proliferation in both cell lines and cellular migration in the SK-N-SH cell line (Fig. 6A, B). These results support the hypothesis that *PTPRH* is a MYCN-independent tumor suppressor.

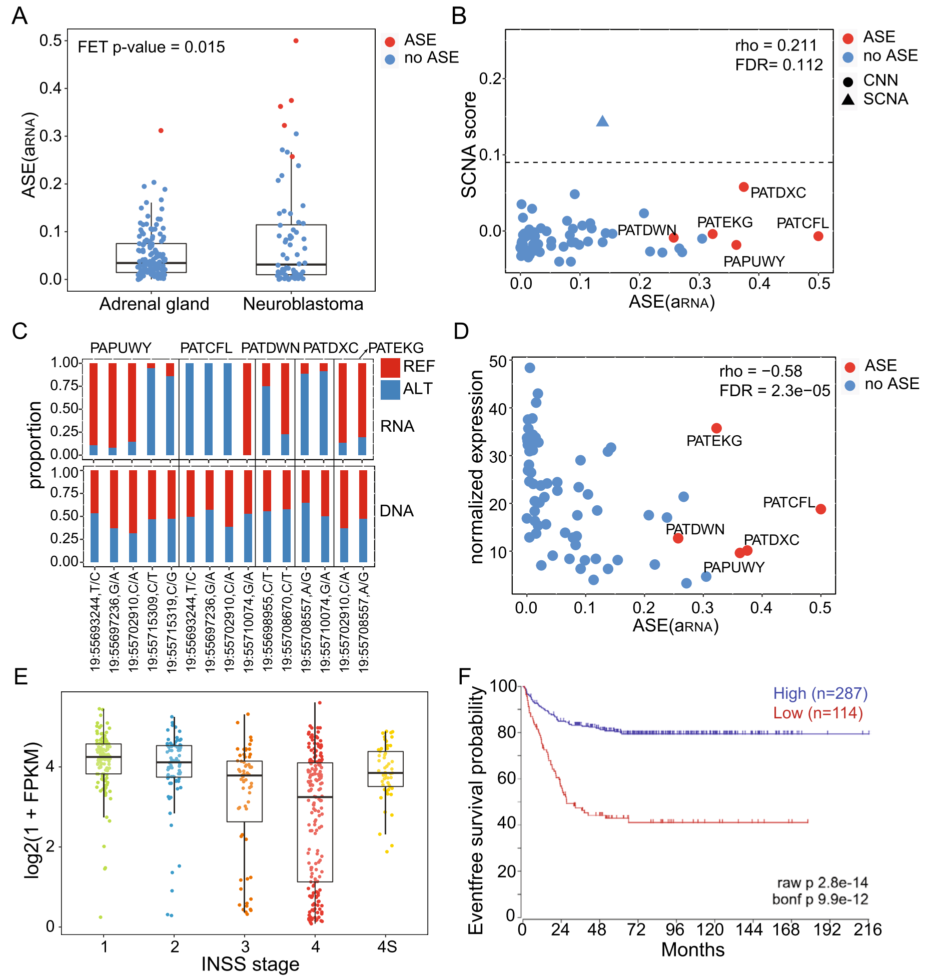

Fig. 5 Genomic profiling of *PTPRH*. **A** ASE (a_{RNA}) of *PTPRH* in neuroblastoma and adrenal gland tissues. *PTRPH* has detectable ASE in 5 out of 68 neuroblastoma patients and 1 out of 139 Adrenal gland tissues (Fisher's exact test *p*-value = 0.015). **B** Spearman's correlation between ASE (a_{RNA}) for *PTPRH* and SCNA score for the overlapping genomic segments. **C** Reference and alternate allele proportions for RNA-seq and exome-seq reads at heterozygous sites within the *PTPRH* gene in 5 samples with significant ASE. **D** Spearman's correlation between ASE (a_{RNA}) and gene expression for *PTPRH*. **E** Normalized gene expression of *PTPRH* across different disease stages for 498 neuroblastoma patients from the SEQC/MAQC-III Consortium data. **F** Kaplan-Meier survival analysis comparing MYCN non-amplified patients with high or low expression of *PTPRH* from the SEQC/MAQC-III Consortium data set

Discussion

Our study leveraged allelic imbalance of RNA and DNA sequencing reads to discover genes with recurrent ASE and delineate SCNA regions in neuroblastoma genomes. Neuroblastoma genomes contain a surprisingly large number of genes with recurrent ASE; however, the majority of ASE events can be attributed to SCNAs which are well-characterized and common genomic alterations in neuroblastoma that typically span tens of megabases.

Our ASE analysis revealed that, in some samples, genes within recurrent SCNA regions are dysregulated by non-SCNA events. Non-SCNA ASE events are present in genes which have been previously described as putative tumor suppressors

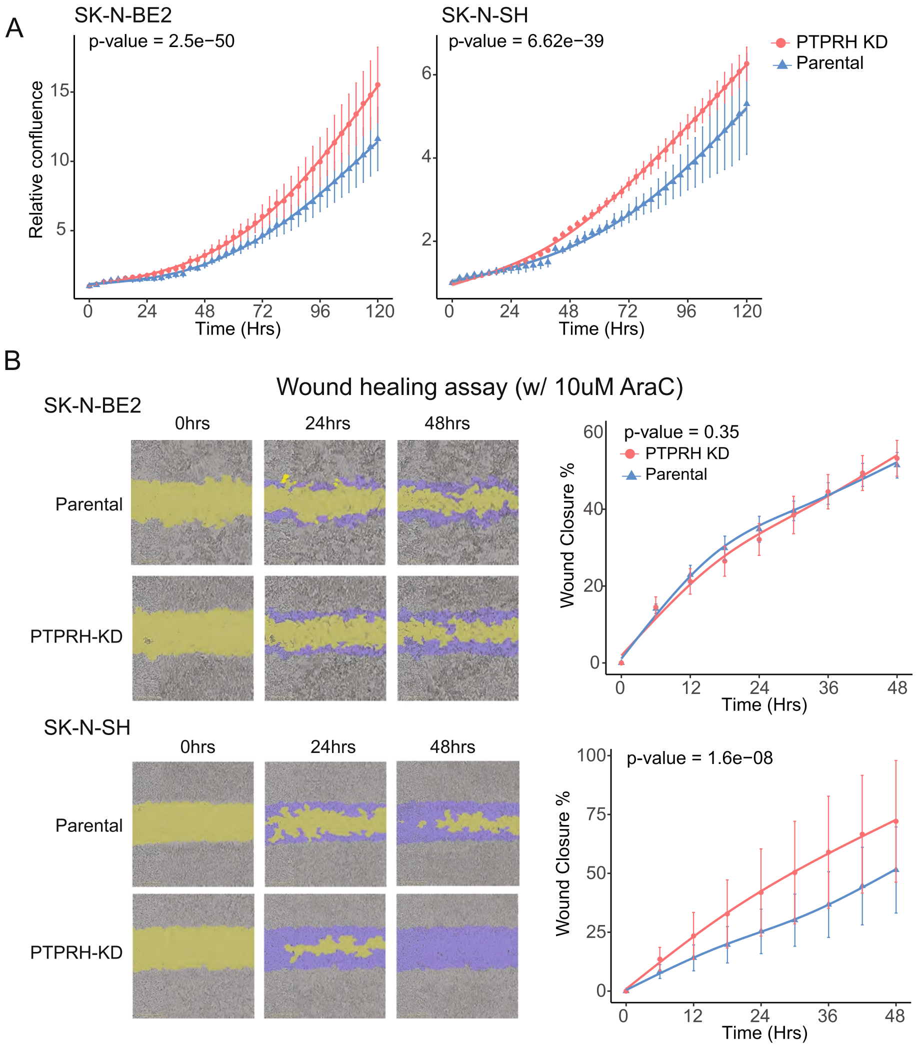

Fig. 6 PTPRH knockdown in neuroblastoma cell lines increases cellular proliferation and migration. **A** Change in relative confluence (i.e., confluence at each time point normalized to confluence at time 0) for parental and PTPRH knockdowns (KDs) in SK-N-BE(2) and SK-N-SH cells. Six replicates for each condition were used for the SK-N-BE(2) cell line and SK-N-SH cell line. **B** Left panels show representative images for scratch wound assays used to measure cellular migration for parental and PTPRH-KD in SK-N-BE(2) and SK-N-SH cells. Right panels show wound closure percentage for parental and PTPRH-KD in SK-N-BE(2) and SK-N-SH cells over 48 h. Each time point had 9 replicates. Error bars indicate +/− one standard deviation around the mean across replicates at each time point. The lines are fits from cubic spline regression using 3 knots, and the *p*-value is from an F-test for a difference in splines between parental and KD cells

and studied in the context of recurrent SCNAs including *KIF1B*, *PLEKHG5*, *UBE4B*, *CHD5*, *CADM1*, and *ATM*. With larger sample sizes, ASE could potentially be utilized to distinguish passenger genes from driver genes within recurrent SCNA regions. In some samples, ASE in the absence of SCNAs can be attributed to mutations that cause NMD; however, in other samples, the cause is unknown. In these cases, the cause of ASE could potentially be revealed by future studies. For example, whole genome sequencing and analysis of non-coding mutations near NB-ASE genes could illuminate cis-acting regulatory mutations that cause ASE.

Outside of recurrent SCNA regions, we discovered 359 genes that are recurrently dysregulated in neuroblastoma. These genes include *TFAP2B*, *MAP7*, *PTPRH*, and *SLC18A1*, which have substantially lower expression in stage 4 disease. *TFAP2B* is important for noradrenergic neuronal differentiation of neuroblastoma cells in vitro and is dysregulated by aberrant promoter methylation [62]. Our independent validation of this finding is additional evidence that *TFAP2B* is an important tumor suppressor in neuroblastoma. *PTPRH* belongs to a group of receptor tyrosine phosphotases which reduce phosphorylation of Akt and its cellular substrates such GSK-3α or GSK-3β [64]. *PTPRH* may inactivate Akt and promote apoptosis in cancer cells. In addition, overexpression of *PTPRH* has been demonstrated to disrupt actin-based cytoskeleton as well as inhibit cellular responses promoted by integrin-mediated cell adhesion, including cell spreading on fibronectin, growth factor-induced activation of extracellular signal-regulated kinase 2, and colony formation [63]. In our study, we found that (a) *PTPRH* exhibits recurrent ASE in neuroblastoma, (b) low expression of *PTPRH* is associated with adverse patient outcomes, and (c) knockdown of *PTPRH* increases proliferation and wound healing in neuroblastoma cell lines. Collectively, these observations suggest that *PTPRH* functions as a tumor suppressor in high-risk neuroblastomas. Confirmation that *PTPRH* acts as a tumor suppressor will require in vivo experiments that are beyond the scope of this study.

Conclusions

In summary, our study provides a framework for analysis of ASE and SCNAs in tumors. Using this framework, we study the impact of genomic alterations that affect gene expression in neuroblastoma and discover that multiple types of mutations work in concert to dysregulate gene expression. While most ASE in neuroblastoma is driven by large-scale SCNAs, many genes exhibit ASE in samples that lack SCNAs. These samples are enriched for mutations that are predicted to cause NMD. In addition, we identify some genes that have recurrent ASE outside of common SCNA regions, including *TFAP2B* and *PTPRH*, both of which have low expression in stage 4 disease and evidence for tumor suppressor activity.

Methods

Datasets

Next-generation sequencing (NGS) data for neuroblastoma patients were obtained from the Therapeutically Applicable Research to Generate Effective Treatments (TARGET) initiative [7]. Our dataset consisted of RNA sequencing for 143 tumors and paired tumor-normal exome sequencing for 97 neuroblastoma patients. Out of the 97 samples with both RNA-seq and exome-seq data, 87 also had Illumina Infinium Human Methylation 450K data and 33 had HumanHap 550K BeadChIP (SNP-array) data. We also obtained 175 adrenal gland and 369 whole blood samples from the GTEx Consortium and used them as a normal ASE reference set [65].

Quality control

To ensure that NGS data from the same patient are properly paired, we compared the RNA-seq and exome-seq data from 97 neuroblastoma tumors using NGSCheckMate [66]. Based on this analysis, we found that one sample had mismatched RNA-seq and exome-seq data and we removed this sample from the study. Our final dataset for ASE analysis consisted of RNA-seq and exome-seq data from 96 neuroblastoma patients.

Variant calling pipeline

We aligned exome-seq reads to the reference genome (hg19) using BWA-MEM with default parameters [67]. Then, we generated GVCF files for each sample using the GATK *HaplotypeCaller* (4.1.1) and performed joint genotyping using GATK *GenotypeGVCFs*. We extracted single nucleotide polymorphism (SNPs) using GATK *SelectVariants* command and recalibrated variant quality scores with GATK variant quality score recalibration (VQSR) pipeline. The filtered and processed SNPs were used for downstream analyses.

Somatic mutation discovery pipeline

We used Mutect2 from GATK (4.1.1) to compare the mutation profile from exome-seq data for 96 neuroblastoma tumor and normal whole blood samples [68]. We filtered somatic mutations from Mutect2 using the GATK recommended filtering pipeline (https://gatk.broadinstitute.org/hc/en-us/articles/360035531132). To determine the functional consequence of somatic mutations and to assign mutations to respective genes, we further analyzed individual somatic mutations in each sample using variant effect predictor (VEP) [69].

Read depth-based detection of somatic copy number alternations

SCNAs in neuroblastoma were detected using our DNA allelic imbalance method described below and a read depth-based method, CNVkit [40]. Briefly, CNVkit uses exome-seq reads to calculate log2 copy ratios across the genome for tumor-normal pairs. SCNAs for large chromosomal regions are then detected by combining log2 copy ratios across adjacent genomic regions using Circular Binary Segmentation (CBS). For this study, we processed aligned exome-seq reads with the *batch* option from CNVkit using the *--drop-low-coverage* parameter to control for low coverage exome targets. Heatmaps showing CNV calls for all samples were generated using CNVkit's *heatmap* script.

RNA-seq alignment and processing

We aligned RNA-seq reads end-to-end to the reference genome (hg19) using STAR (2.5.3a) [70]. The aligned reads were filtered to those with mapping quality ≥ 20 using samtools (1.9) [71]. Reads mapping to each gene were counted using featureCounts (1.6.3) for GENCODE (v28) genes [72]. Gene counts were converted to Fragments Per Kilobase Per Million (FPKM) using DESeq2 (1.22.2) [73]. Finally, the FPKM matrix was quantile normalized using the preprocessCore (1.44) package and z-score transformed.

Estimating DNA allelic imbalance using exome-seq

We realigned exome-seq reads from tumor and normal samples to the reference genome (hg19) using BWA-ALN and filtered the sequencing reads to remove mapping bias using WASP [74]. We obtained allele-specific read counts at heterozygous sites (excluding multiallelic sites) for normal tissues using the *CollectAllelicCounts* from GATK (4.1.1). We assumed that most heterozygous sites in normal tissues are germline polymorphisms and obtained allele-specific read counts at the shared positions for matched tumor samples. We analyzed shared heterozygous positions because this facilitates direct comparison of reference allele proportions between tumor and paired normal tissues.

To model DNA allelic imbalance over large genomic segments, we sorted exons by their genomic coordinates and grouped 20 consecutive exons into genomic bins. Next, we assigned the heterozygous sites which were covered by at least 10 reads to the genomic bins. To ensure robust regional DNA allele imbalance estimates, we retained genomic bins with at least 10 heterozygous sites for DNA allele imbalance analysis.

To model allele-specific read counts. we assumed that the reference allele count at heterozygous sites is beta-binomially distributed. The likelihood for the allelic imbalance parameter, *a*, and the dispersion parameter, *d*, at a single heterozygous site *i* is then defined by:

$$\mathcal{L}_i\left(a, d|x_{R,i}, n_i\right) = p_X(0.5+a, d) = \Pr\left(X = x_{R,i}|p = 0.5+a, d, n = n_i\right) = \frac{B\left(x_{R,i} + d(0.5+a), n_i - x_{R,i} + d(0.5-a)\right)\binom{n}{x_{R,i}}}{B(d(0.5+a), d(0.5-a))}$$

where $p_X()$ is the beta binomial probability mass function; $x_{R,\ i}$ is the observed reference allele count from overlapping reads at site *i*; n_i is the total count of overlapping reads matching the reference or alternate allele at site *i*; *p* is the reference allele proportion; *a* is the allelic imbalance parameter and is defined over the range [−0.5,0.5]; *d* is the beta binomial dispersion parameter; and *B*() is the beta function. To perform likelihood calculations, we used the beta-binomial probability mass function (dbetabinom) provided by the rmutil (1.1.5) package, which is described in the vignette (https://cran.r-project.org/web/packages/rmutil/rmutil.pdf).

We estimate a single-dispersion parameter across all heterozygous sites genome-wide. This is accomplished by fixing *a* to 0 and finding the value of *d* that maximizes the total likelihood across all heterozygous sites in either a normal or a tumor sample:

$$\hat{d} = \text{argmax}_d \prod_i \mathcal{L}_i\left(a = 0, d|x_{R,i}, n_i\right)$$

After estimating the dispersion parameter, we estimate the allelic imbalance and compute the likelihood for bins. In the case of DNA allelic imbalance, a bin consists of 20 consecutive exons as described above. In the case of RNA allelic imbalance, we utilize the set of heterozygous sites within the exons of the gene being considered as described below. To estimate the imbalance of a bin, we combine information across all of the heterozygous sites within the bin. Since we do not know the phasing of the alleles, we consider all possible phasings (i.e., haplotype configurations), when computing the likelihood. Under the assumption that all phasings are equally probable, the likelihood of the parameters for bin *B* is:

$$\mathcal{L}_B\left(a, d = \hat{d}|x_R, n\right) = \frac{1}{\|H_B\|}\prod_{h\in H_B}\prod_{i\in B}(1-h_i)\left(\mathcal{L}_i\left(a, d = \hat{d}|x_{R,i}, n_i\right)\right) + (h_i)\left(\mathcal{L}_i\left(a, d = \hat{d}|n_i - x_{R,i}, n_i\right)\right)$$

where H_B is the set of all possible phasings (i.e., hapolotype configurations) for bin B. A phasing is defined by a vector of 0s and 1s, with each element corresponding to a heterozygous site. Element i of the phase vector is set to $h_i = 0$ if the reference allele for site i is on "chromosome A" and set to $h_i = 1$ if the reference allele for site i is on "chromosome B". "Chromosome A" is defined as the chromosome that carries the reference allele for the first heterozygous site in the bin. In total there are $\|H_B\| = 2^{m-1}$ possible phasings, where m is the number of heterozygous sites within the bin. While the number of phasings grows exponentially with m, the number of heterozygous sites in the bin, the likelihood for the bin can be computed efficiently in linear time using dynamic programming:

```
$T[1] \leftarrow \mathcal{L}_1\left(a, d = \hat{d}|x_{R,1}, n_1\right)$

$for\ i$ in $2, \ldots, m\ do$

$T[i] \leftarrow (T[i-1])\left(\mathcal{L}_i\left(a, d = \hat{d}|x_{R,i}, n_i\right) + \mathcal{L}_i\left(a, d = \hat{d}|n_i - x_{R,i}, n_i\right)\right)$

$return\ \frac{T[m]}{\|H_B\|}$
```

where T is an array of length m that is used to compute the cumulative likelihood.

We performed one dimensional optimization to obtain maximum likelihood estimates of a for each bin. We perform this procedure separately for tumor and normal samples to obtain estimates of a for tumor (a_{tumor}) and normal samples (a_{normal}). We then calculate the difference in a between tumor and normal samples,δ_a as follows:

$$\delta_a = |a_{tumor}| - |a_{normal}|$$

Finally, to create contiguous segments of allelic imbalance, we performed Circular Binary Segmentation (CBS) on δ_a using the DNAcopy package (1.56.0) [39]. We defined the aggregate value for δ_a obtained from CBS as the "SCNA score". Plots for δ_a and SCNA scores were generated using Gviz (1.26.5), ComplexHeatmap (1.20.0), and gplots (3.0.1.1) [75].

Estimating allele-specific expression per gene using RNA-seq reads

We filtered RNA-seq reads for mapping bias and obtained allele specific read counts at heterozygous positions using WASP [74]. WASP uses random sampling to ensure that allele counts at nearby heterozygous sites are independent. Specifically, when a read overlaps multiple sites, the allelic count is incremented at only one of the sites, which is selected randomly. We observed that at most heterozygous sites overlapping RNA-seq reads only match the reference or alternate alleles. However, some sites also have reads that match neither allele, which we refer to as "other" reads. "Other" reads may reflect sequencing errors or mis-mapped reads so, prior to ASE analysis, we removed all heterozygous sites where the "other" read count was greater than two.

Genotyping errors can create a false signal of allelic imbalance. For the SCNA analysis above, DNA imbalance is estimated from many heterozygous sites within each bin and then estimates from multiple bins are combined with circular binary segmentation. Since many heterozygous sites are used for this analysis, genotyping errors can be ignored without a major effect. However, when we calculate ASE for a gene, we only consider heterozygous sites within the exons of the gene. Thus, when estimating ASE for a single gene, we often utilize a small number of heterozygous sites, so it is desirable to account for genotyping errors. To control for genotyping errors, we calculate the genotyping error rate, $\epsilon_{G,\ i}$, for each heterozygous site i directly from the genotype quality (GQ) scores provided by GATK:

$$\epsilon_{G,i} = 10^{-\frac{GQ_i}{10}}$$

When a genotyping error occurs, a heterozygous site can be homozygous reference (0/0) or alternate (1/1), and we assume these two possibilities are equally likely. When there is a genotyping error and the sample is homozygous, all sequencing reads that come from one of the two alleles must arise due to sequencing or mapping errors. We use the parameter ϵ_S to describe the frequency of these sequencing errors. The likelihood of the parameters for a single heterozygous site is then:

$$\begin{aligned}
\mathcal{L}_i\left(a_i, d, \epsilon_S | x_{R,i}, n_i, \epsilon_{G,i}\right) &= \left(1-\epsilon_{G,i}\right)p_X(0.5+a,d) + \epsilon_{G,i}\left(0.5\,p_X\left(\epsilon_S, d\right) + 0.5\,p_X\left(1-\epsilon_S, d\right)\right) \\
&= \left(1-\epsilon_{G,i}\right)\Pr\left(X = x_{R,i} | p = 0.5+a, d = d, n = n_i\right) \\
&+ \left(0.5\epsilon_{G,i}\right)\Pr\left(X = x_{R,i} | p = \epsilon_s, d = d, n = n_i\right) \\
&+ \left(0.5\epsilon_{G,i}\right)\Pr\left(X = x_{R,i} | p = 1-\epsilon_S, d = d, n = n_i\right)
\end{aligned}$$

where, as described above, $p_X()$ is the probability mass function for the beta binomial distribution.

To find the maximum likelihood estimate of d and ϵ_S, we fix a to 0 and optimize d and ϵ_S over all heterozygous sites overlapping exons:

$$\hat{d}, \hat{\epsilon_S} = \text{argmax}_{d,\epsilon_s} \prod_i \mathcal{L}_i\left(a_i, d, \epsilon_S | x_{R,i}, n_i, \epsilon_{G,i}\right)$$

For optimization, we used the L-BFGS-B algorithm implemented by the "*optim*" function provided by the stats package in R-4.0.1.

To estimate the ASE of a gene within a sample, we obtain a maximum likelihood estimate of a, keeping the dispersion and sequencing error rate fixed to their genome-wide estimates ($\hat{d}$ and $\hat{\epsilon_S}$). We combine information across all heterozygous sites within each gene. To combine information across heterozygous sites, we use the same approach described in the DNA imbalance section above. We group all of heterozygous sites that fall within the exons of a gene into a "bin" and compute likelihoods that consider all possible phasings of alleles. We use a likelihood ratio test to compare the alternative model (with a free a parameter) to the null model of no allelic imbalance (with a fixed to $a = 0$) and to compute p-values. We correct the p-values for multiple testing using the Benjamini-Hochberg method. To make it clear when we are referring to allelic imbalance in RNA instead of DNA, we refer to a for RNA-seq read as a_{RNA}.

Gene Ontology enrichment analysis

Gene Ontology (GO) enrichment analysis for Biological Processes (BP) was performed using topGO (2.34.0). Enrichment was calculated using Fisher's exact test, and all genes tested for ASE in neuroblastoma were used as the universe.

Correlation between DNA allele-imbalance and SNP-array predictions

We used GenomicRanges (1.34.0) to find overlaps between SCNAs detected using our DNA allelic-imbalance method and SCNA predictions obtained from TARGET which were generated using HumanHap 550K Beadchip (SNP-array) [41]. For segments which show at least a 50% overlap, we computed Spearman's correlation between segmented DNA allele imbalance (i.e., SCNA score) and corrected Log R ratio estimated using SNP-array data from TARGET [41].

Association between allele-specific expression and SCNAs

We assigned our candidate genes to genomic segments predicted to be SCNAs based on the location of their promoters (transcription start site +/− 1500bp) using GenomicRanges (1.34.0) [76] and computed Spearman's correlation between ASE (a_{RNA}) and SCNA score. We corrected the p-values for multiple testing using the Benjamini-Hochberg procedure. We only tested genes with non-zero variance in both SCNA scores and a_{RNA} and with an SCNA score in at least one sample ≥ 0.09. Manhattan plots for Spearman's correlation coefficient were generated using ggbio (1.30.0) [77].

Correlations between allele-specific expression and promoter methylation

In Human Methylation 450K BeadChIP array (HM450K) data, the ratios of intensities between methylated and unmethylated CpG probes are referred to as beta values (β) and range from 0 (unmethylated) to 1 (completely methylated). We downloaded a pre-computed β matrix for 87 neuroblastoma samples from TARGET and annotated the CpG probe positions based on GENCODE (version 28) genes. Then, we computed the mean β for promoter regions (transcription start site +/− 1500 bp) and computed Spearman's correlations between promoter methylation and ASE ($\mathrm{a_{RNA}}$) for 1043 NB-ASE genes. We corrected the p-values for multiple testing using the Benjamini Hochberg procedure. We only tested genes with at least 3 CpG probes within promoter regions.

Survival and expression analysis in neuroblastoma patients

We analyzed the SEQC/MAQC-III Consortium dataset consisting of 498 individuals using the R2: Genomics Analysis and Visualization Platform (http://r2.amc.nl) to generate Kaplan-Meier survival plots for neuroblastoma [60]. We also downloaded a normalized gene expression matrix (i.e., log (1 + FPKM)) for SEQC/MAQC-III Consortium dataset from Gene Expression Omnibus (GSE49711) [60] and generated gene expression heatmaps using ComplexHeatmap (1.20.0) and ggplot2 (3.2.1) [75, 78].

Cell culture and transfection

The SK-N-BE(2) and SK-N-SH cell lines were purchased from the American Type Culture Collection (www.atcc.org) and grown in a humidified chamber with 5% CO_2 in RPMI 1640 medium (Gibco, #11875119) supplemented with 10% fetal bovine serum

(FBS), 2mM L-glutamine, sodium pyruvate, non-essential amino acids, and 1% antibiotic antimycotic.

For stable transfection, the cell lines were seeded in 6-well plates and allowed to grow to 70% confluence. Cells were transfected with shRNAs purchased from Sigma-Aldrich (shPTPRH-373, #TRCN0000355579; shPTPRH-1136, #TRCN0000355581; shPTPRH-1947, #TRCN0000355580; shPTPRH-3265, TRCN0000355631; shPTPRH-3621, #TRCN0000002866) with jetOPTIMUS® DNA transfection Reagent (VWR, #76299-632) following the protocol provided by the manufacturer. The cells were selected in RPMI containing puromycin 24hrs after transfection (SK-N-BE(2), 1μg/ml puromycin and SK-N-SH, and 1.25μg/ml puromycin). The stably transfected cells were maintained in complete medium supplemented with puromycin until use.

Quantitative RT-PCR analysis

SK-N-BE(2) and SK-N-SH cells were seeded in 6-well plates and allowed to grow to 50% confluence. RNA was isolated from the cells with Zymo Quick-RNA Miniprep kit (VWR, #76299-632) according to the manufacturer's instructions. Total RNA was reversed transcribed into cDNA using High-Capacity cDNA Reverse Transcription Kit (Fisher Scientific, #4374966). Real-time PCR was performed using iTaq™ Universal SYBR Green Supermix (Bio-Rad Laboratories, #1725122) on a Bio-Rad CFX96 system. Gene expression was analyzed by the log2ΔΔCt method.

Immunoblotting

SK-N-BE(2) and SK-N-SH cells were seeded in 6-well plates and allowed to grow to 70–80% confluence. Cells were lysed in RIPA buffer, and the protein concentration was determined by a Pierce BCA protein assay (Life Technologies, #23225). Equal amounts of protein were loaded into 8% Bolt™ Bis-Tris Plus gels (Life Technologies, #NW00085BOX), separated by SDS-PAGE and then transferred to PVDF membranes. The membranes were incubated with primary antibodies (PTPRH antibody, 1:1000, Fisher Scientific, #PIPA531340) overnight at 4°C. The membranes were then probed with appropriate horseradish peroxidase-conjugated secondary antibodies (Goat Anti-Rabbit IgG(H+L)-HRP Conjugate, Bio-rad Laboratories, #170-6515). The immunoblots were visualized with SuperSignal West Pico Plus Chemiluminescent Substrate (Life Technologies, #PI34580).

Proliferation and migration assays

SK-N-BE(2) and SK-N-SH cells were plated in 96-well plates at a seeding density of 7500 cells/well and allowed to attach overnight. Cells were then monitored by continuous live-cell imaging in the IncuCyte® Zoom™ system (Essen Bioscience), and 10x phase contrast images were taken every 3h. Cell confluence in each image was calculated by the IncuCyte® analysis software.

For migration assays, SK-N-BE(2) and SK-N-SH cells were seeded in IncuCyte® Imagelock 96-well plates (Essen Bioscience, #4379) at seeding densities between 100,000 and 200,000 cells/well and allowed to grow to 100% confluence. Cells were treated with 10μM cytosine arabinoside (Sigma-Aldrich, #C1768) for 4h, and then, identical scratch wounds were made in each well with a 96-pin WoundMaker (Essen Bioscience). Wound closure was monitored by continuous live-cell imaging in the IncuCyte Zoom^{TM} (Essen Bioscience), and 10x phase contrast images were taken every 6h. The wound closure percentage was calculated using IncuCyte Scratch Wound Analysis Software.

To analyze the cellular proliferation and migration rates, we computed the mean confluence and mean wound closure at each time point across replicates. To allow for nonlinearity in both types of data, we performed cubic spline regression with 3 knots (two degrees of freedom) to describe the change in confluence or wound closure with time. To quantify differences in proliferation or migration between knockdown and parental cells, we included an interaction term between knockdown status and the time spline in the model. Specifically, we utilized the following linear model command in R, where “response” is the confluence or wound closure, “ns” is the cubic spline function, and “cond” is a factor giving knockdown status (parental or knockdown):

$$\mathrm{lm}(\mathrm{response} \sim \mathrm{cond} * \mathrm{ns}(\mathrm{time}, \mathrm{df} = 2))$$

Significance of the interaction term was determined with an F-test.

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s13059-022-02640-y.

Additional file 1: Fig. S1. Validation of SCNA scores using SNP-array data. A SCNA predictions based on DNA allelic imbalance compared to SNP-array predictions. Left panel shows SCNA scores across chromosome 1 estimated from DNA allelic imbalance from 33 neuroblastoma patients. Right Panel shows Corrected Log R ratio (or Corrected LRR) calculated from SNP array data for the same 33 patients. Corrected LRR is defined as aneuploidy corrected total probe intensity of a given genomic segment relative to a canonical set of normal controls and directly available from the TARGET. B Spearman’s rank correlation between SCNA score and absolute Corrected LRR for chromosome 1 (Spearman’s correlation coefficient = 0.614, *p*-value = 3.51e-06). Exome-seq and SNP arrays use different sets of SNPs to predict SCNAs. Therefore, the Circular Binary Segmentation (CBS) algorithm tends to output segments which do not share the same genomic start and end positions. To directly compare SCNA detection using DNA allele imbalance and SNP array, we first calculated the fraction overlap between genome segments identified by the respective methods. Next, we performed pairwise Spearman’s correlation between SCNA score and absolute Corrected LRR for genomic segments with fraction of overlap ≥ 0.5 or 50%. The points in the correlation scatter plot are colored by fraction overlap. The two points labelled PANRVJ correspond to two disjointed SCNAs spanning chr1:1922327-9171333 and chr1:49201909-120298048. These regions showed absolute Corrected LRR < 0.5 and were annotated as copy neutral by SNP array. We suspect that these segments may be copy-neutral loss of heterozygosity regions, which are not detectable using direct analysis of SNP-arrays in TARGET. **Fig. S2.** SCNA predictions for chromosome 11. A Comparison between DNA-imbalance SCNA predictions and CNVkit predictions for chr11. Left panel shows heatmap of δ" for 96 neuroblastoma patients. Right panel shows fold-change in normalized read coverage between tumor and normal tissues estimated using CNVkit. B Comparison between DNA-imbalance predictions and SNP-array predictions for chr11. Left panel shows SCNA score across 33 neuroblastoma patients with SNP-array data in TARGET. Right panel shows corrected LRR calculated array SNP-array available through TARGET. C Spearman’s rank correlation between SCNA score and absolute corrected LRR for chr11 (Spearman’s correlation coefficient = 0.565, *p*-value = 2.3e-05). The points are colored based on fraction of overlap between genomic regions detected by our method and genomic regions from SNP-array based predictions. D Spearman’s rank correlation between ASE (aRNA) and SCNA score for *MTMR2*, a gene located within a common deletion segment on cytoband 11q21 (Spearman’s correlation coefficient = 0.64, *p*-value = 0.0001). **Fig. S3.** Detection of rare SCNAs on chromosome 16. A Left panel shows the heatmap of δ" for chromosome 16. Right panel shows the log2 fold-change in normalized read coverage between tumor and normal tissues estimated using CNVkit for chromosome 16. B Comparison between SCNA score and SNP-array predictions (i.e. corrected LRR) for chromosome 16 for 33 neuroblastoma samples. C Spearman’s rank correlation between SCNA score and absolute corrected LRR for chromosome 16 (Spearman’s correlation coefficient=0.59, *p*-value = 5.04e-05) for overlapping genomic regions. The points are colored based on fraction overlap between genomic regions detected by our method and genomic regions from SNP-array based predictions. D Spearman’s rank correlation between ASE (aRNA) and SCNA score for *CDT1*, a gene located in the distal region of the q-arm (i.e., 16q24.3) (Spearman’s correlation coefficient = 0.58, FDR corrected *p*-value = 2e-06). **Fig. S4.** Haplo-insufficient tumor suppressors within common SCNAs may be dysregulated by secondary mechanisms. Spearman’s correlation between ASE (aRNA) and SCNA score for example chromosome 1p and chromosome 11q deletion genes: A *CHD5*, B *UBE4B*, C *CADM1*, and D *ATM*. Several samples show ASE in the absence of SCNAs. **Fig. S5.** Allele-specific expression, gene expression, promoter methylation, and survival for *TFAP2B*. A ASE (*a*RNA) of *TFAP2B* in neuroblastoma and adrenal gland tissues. B Reference and alternate allele proportion for RNA-seq and exome-seq reads at heterozygous sites which were used to estimate ASE for *TFAP2B*. C Correlation between ASE (*a*RNA) and promoter methylation for *TFAP2B*. DNA methylation data was missing for 1 neuroblastoma sample (PAMVRA). The two samples with significant ASE are among those with the greatest promoter methylation. ASE of *TFAP2B* is correlated with its promoter methylation, however this correlation is not significant under an FDR threshold of 10% (Spearman's rho= 0.604, FDR corrected *p*-value = 0.116). D Spearman's correlation between ASE (*a*RNA) and gene expression for *TFAP2B.* E Genomic distribution of HM450K β-values for *TFAP2B* locus. The *TFAP2B* promoter is highlighted (gold box). F Expression profile of *TFAP2B* across different stages of disease for 498 neuroblastoma patients obtained from SEQC/ MAQC-III Consortium data set. We observed loss of expression of *TFAP2B* in stage 4 or metastatic disease suggesting this gene might act as a tumor suppressor. G Kaplan Meier survival analysis for *MYCN* nonamplified patients from the SEQC/MAQC-III Consortium data set. **Fig. S6.** Quantile-quantile plot for Spearman's correlation analysis between ASE (aRNA) and promoter methylation for 1,043 NB-ASE genes. Under an FDR of 10% only the expression of *ODZ4* is significantly correlated with promoter methylation. **Fig. S7.** Knockdown of PTPRH in neuroblastoma cell lines. A, B *PTPRH* expression measured by qPCR in (A) SK-N-SH or (B) SK-N-BE(2) neuroblastoma cell lines stably transfected with 5 different shRNAs targeting *PTPRH*. Gene expression is plotted as 2-DDCt normalized to *HPRT1* expression. C, D Western blot of PTPRH and GAPDH protein expression for the same (C) SK-N-SH and (D) SK-N-BE(2) cell lines. shRNA shPTPRH-373 consistently reduced gene and protein expression of *PTPRH* in both SK-N-SH and SK-N-BE(2) cells and was used for all downstream experiments. E, F Uncropped versions of the western blots of PTPRH and GAPDH protein expression in (E) SK-N-SH and CHP212 cells, and (F) SK-N-BE(2) cells. Note that we only used SK-NSH and SK-N-BE(2) cells for the proliferation and wound healing assays shown in main text Fig. 6.

Additional file 2: Table S1. Results from Allele-Specific Expression (ASE) analysis of RNA-seq data for 96 neuroblastoma samples from TARGET. **Table S2.** Number of significant (FDR ≤ 0.1 or 10%) and testable (i.e., at least 1 heterozygous site with ≥ 10 reads) samples for neuroblastoma tumors, adrenal gland, and whole-blood tissues for all significant ASE genes. Genes were considered to have neuroblastoma-specific ASE if they met these criteria; a) testable in ≥ 10 neuroblastoma and normal (i.e., adrenal-gland and whole-blood) samples and b) significant in ≥3 neuroblastoma and significant ≤ 1 normal sample. The prefix "r." indicates the number of samples showing significant ASE and "N." indicates the number of samples testable for ASE. **Table S3.** Top 20 Gene Ontology (Biological processes) categories enriched for 1,043 NB-ASE genes. **Table S4.** SCNA scores for 96 neuroblastoma patient samples. **Table S5.** Spearman's correlation between ASE (a_{RNA}) for 1,043 NB-ASE genes and SCNA score for overlapping genomic segments. **Table S6.** Table of high-impact somatic mutations mapping to NB-ASE genes detected using exome-seq data using Variant Effect Predictor (VEP) in 96 neuroblastoma tumors. **Table S7.** Spearman's correlation between ASE (aRNA) and mean promoter methylation for 1,043 NB-ASE genes. **Table S8.** Spearman's correlation between ASE (a_{RNA}) and gene expression (z-score normalized Fragments per Kilobase Per Million mapped) for 1,043 NB-ASE genes.

Additional file 3. Review history.

Acknowledgements

The results published here are in whole or part based upon data generated by the Therapeutically Applicable Research to Generate Effective Treatments (TARGET) initiative, phs000218, managed by the NCI. The data used for this analysis are available from dbGaP (phs000467.v1.p1). Information about TARGET can be found at http://ocg.cancer.gov/programs/target.

The Genotype-Tissue Expression (GTEx) Project was supported by the Common Fund of the Office of the Director of the National Institutes of Health, and by the NCI, NHGRI, NHLBI, NIDA, NIMH, and NINDS. The data used for the analyses described in this manuscript were obtained from dbGaP accession number phs000424.v7.p2 on July 16, 2018.

Review history

Review history for this manuscript is available as Additional file 3.

Peer review information

Barbara Cheifet and Stephanie McClelland were the primary editors of this article and managed its editorial process and peer review in collaboration with the rest of the editorial team.

Authors' contributions

G.M., J.E., and A.S. conceived of the project. G.M. and J.E. obtained funding for the project. Data processing and analysis were performed by A.S. A.S. wrote the initial draft, and A.S. and G.M. wrote the manuscript. P.Z. provided comments on the manuscript. G.M. supervised the analysis and data processing. Y.H. performed all experiments under the supervision of P.Z. The author(s) read and approved the final manuscript.

Funding

This research was supported by the National Cancer Institute funded Salk Institute Cancer Center (NIH/NCI CCSG: P30 014195), a grant from Padres Pedal the Cause/RADY #PTC2019 awarded to G.M. and J.E., and NIH/NHGRI grant HG011315 awarded to G.M. A.S. was supported by a Pioneer Fund Postdoctoral Scholar Award. G.M. was supported by the Frederick B. Rentschler Developmental Chair.

Availability of data and materials

The datasets analyzed in this study are available in the following repositories. Data from the Genotype-Tissue Expression (GTEx) Project are available from the Database of Genotypes and Phenotypes (dbGAP) via accession phs000424.v7.p2 (https://www.ncbi.nlm.nih.gov/projects/gap/cgi-bin/study.cgi?study_id=phs000424.v7.p2) [79]. Data from the Therapeutically Applicable Research to Generate Effective Treatments (TARGET) initiative are available from dbGAP via accession phs000218.v1.p1 (https://www.ncbi.nlm.nih.gov/projects/gap/cgi-bin/study.cgi?study_id=phs000218.v1.p1) [7]. Gene expression data from the SEQC/MAQC-III Consortium are available from the Gene Expression Omnibus (GEO) via accession GSE49711 (https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE49711) [80]. Survival data for this dataset are available from the R2: Genomics Analysis and Visualization Platform (http://r2.amc.nl). Source code for RNA and DNA allelic imbalance analyses are available from GitHub at https://github.com/Arkosen/Allele-imbalance-analysis-of-RNA-and-DNA and in archived form at https://doi.org/10.5281/zenodo.6229172.

Declarations

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Competing interests

No competing interests are declared.

Author details

[1]Integrative Biology Laboratory, Salk Institute for Biological Studies, La Jolla, California, USA. [2]Department of Pediatrics, Division of Hematology-Oncology, University of California San Diego, La Jolla, California, USA. [3]Peckham Center for Cancer and Blood Disorders, Rady Children's Hospital-San Diego, San Diego, California, USA.

Received: 16 July 2021 Accepted: 23 February 2022
Published online: 04 March 2022

References

1. Mlakar V, Jurkovic Mlakar S, Lopez G, Maris JM, Ansari M, Gumy-Pause F. 11q deletion in neuroblastoma: a review of biological and clinical implications. Mol Cancer. 2017;16:114.
2. Maris JM, Hogarty MD, Bagatell R, Cohn SL. Neuroblastoma. Lancet. 2007;369:2106–20.
3. Schleiermacher G, Janoueix-Lerosey I, Delattre O. Recent insights into the biology of neuroblastoma. Int J Cancer. 2014;135:2249–61.
4. Bagatell R, Cohn SL. Genetic discoveries and treatment advances in neuroblastoma. Curr Opin Pediatr. 2016;28:19–25.
5. Pinto NR, Applebaum MA, Volchenboum SL, Matthay KK, London WB, Ambros PF, et al. Advances in risk classification and treatment strategies for neuroblastoma. J Clin Oncol. 2015;33:3008–17.
6. Matthay KK, Maris JM, Schleiermacher G, Nakagawara A, Mackall CL, Diller L, et al. Neuroblastoma. Nat Rev Dis Primers. 2016;2:16078.
7. Pugh TJ, Morozova O, Attiyeh EF, Asgharzadeh S, Wei JS, Auclair D, et al. The genetic landscape of high-risk neuroblastoma. Nat Genet. 2013;45:279–84.
8. Cheung NK, Dyer MA. Neuroblastoma: developmental biology, cancer genomics and immunotherapy. Nat Rev Cancer. 2013;13:397–411.
9. Huang M, Weiss WA. Neuroblastoma and MYCN. Cold Spring Harb Perspect Med. 2013;3:a014415.
10. Maris JM, Weiss MJ, Guo C, Gerbing RB, Stram DO, White PS, et al. Loss of heterozygosity at 1p36 independently predicts for disease progression but not decreased overall survival probability in neuroblastoma patients: a Children's Cancer Group study. J Clin Oncol. 2000;18:1888–99.
11. Plantaz D, Mohapatra G, Matthay KK, Pellarin M, Seeger RC, Feuerstein BG. Gain of chromosome 17 is the most frequent abnormality detected in neuroblastoma by comparative genomic hybridization. Am J Pathol. 1997;150:81–9.
12. Attiyeh EF, London WB, Mosse YP, Wang Q, Winter C, Khazi D, et al. Chromosome 1p and 11q deletions and outcome in neuroblastoma. N Engl J Med. 2005;353:2243–53.
13. Zage PE, Sirisaengtaksin N, Liu Y, Gireud M, Brown BS, Palla S, et al. UBE4B levels are correlated with clinical outcomes in neuroblastoma patients and with altered neuroblastoma cell proliferation and sensitivity to epidermal growth factor receptor inhibitors. Cancer. 2013;119:915–23.
14. Garcia I, Mayol G, Rodriguez E, Sunol M, Gershon TR, Rios J, et al. Expression of the neuron-specific protein CHD5 is an independent marker of outcome in neuroblastoma. Mol Cancer. 2010;9:277.
15. Kolla V, Naraparaju K, Zhuang T, Higashi M, Kolla S, Blobel GA, et al. The tumour suppressor CHD5 forms a NuRD-type chromatin remodelling complex. Biochem J. 2015;468:345–52.
16. Yang HW, Chen YZ, Takita J, Soeda E, Piao HY, Hayashi Y. Genomic structure and mutational analysis of the human KIF1B gene which is homozygously deleted in neuroblastoma at chromosome 1p36.2. Oncogene. 2001;20:5075–83.
17. Liu Z, Yang X, Li Z, McMahon C, Sizer C, Barenboim-Stapleton L, et al. CASZ1, a candidate tumor-suppressor gene, suppresses neuroblastoma tumor growth through reprogramming gene expression. Cell Death Differ. 2011;18:1174–83.
18. Schlisio S, Kenchappa RS, Vredeveld LC, George RE, Stewart R, Greulich H, et al. The kinesin KIF1Bbeta acts downstream from EglN3 to induce apoptosis and is a potential 1p36 tumor suppressor. Genes Dev. 2008;22:884–93.
19. Fujita T, Igarashi J, Okawa ER, Gotoh T, Manne J, Kolla V, et al. CHD5, a tumor suppressor gene deleted from 1p36.31 in neuroblastomas. J Natl Cancer Inst. 2008;100:940–9.
20. Welch C, Chen Y, Stallings RL. MicroRNA-34a functions as a potential tumor suppressor by inducing apoptosis in neuroblastoma cells. Oncogene. 2007;26:5017–22.
21. Henrich KO, Bauer T, Schulte J, Ehemann V, Deubzer H, Gogolin S, et al. CAMTA1, a 1p36 tumor suppressor candidate, inhibits growth and activates differentiation programs in neuroblastoma cells. Cancer Res. 2011;71:3142–51.
22. Mohammadi P, Castel SE, Cummings BB, Einson J, Sousa C, Hoffman P, et al. Genetic regulatory variation in populations informs transcriptome analysis in rare disease. Science. 2019;366:351–6.

23. Liu Y, Li C, Shen S, Chen X, Szlachta K, Edmonson MN, et al. Discovery of regulatory noncoding variants in individual cancer genomes by using cis-X. Nat Genet. 2020;52:811–8.
24. Pastinen T. Genome-wide allele-specific analysis: insights into regulatory variation. Nat Rev Genet. 2010;11:533–8.
25. Skelly DA, Johansson M, Madeoy J, Wakefield J, Akey JM. A powerful and flexible statistical framework for testing hypotheses of allele-specific gene expression from RNA-seq data. Genome Res. 2011;21:1728–37.
26. Knowles DA, Davis JR, Edgington H, Raj A, Fave MJ, Zhu X, et al. Allele-specific expression reveals interactions between genetic variation and environment. Nat Methods. 2017;14:699–702.
27. Buckberry S, Bianco-Miotto T, Hiendleder S, Roberts CT. Quantitative allele-specific expression and DNA methylation analysis of H19, IGF2 and IGF2R in the human placenta across gestation reveals H19 imprinting plasticity. PLoS One. 2012;7:e51210.
28. Deng Q, Ramskold D, Reinius B, Sandberg R. Single-cell RNA-seq reveals dynamic, random monoallelic gene expression in mammalian cells. Science. 2014;343:193–6.
29. Chess A. Mechanisms and consequences of widespread random monoallelic expression. Nat Rev Genet. 2012;13:421–8.
30. Przytycki PF, Singh M. Differential allele-specific expression uncovers breast cancer genes dysregulated by cis noncoding mutations. Cell Syst. 2020;10:193–203 e194.
31. Calabrese C, Davidson NR, Demircioğlu D, Fonseca NA, He Y, Kahles A, et al. Genomic basis for RNA alterations in cancer. Nature. 2020;578:129–36.
32. Krueger C, Morison IM. Random monoallelic expression: making a choice. Trends Genet. 2008;24:257–9.
33. Rachmilewitz J, Goshen R, Ariel I, Schneider T, de Groot N, Hochberg A. Parental imprinting of the human H19 gene. FEBS Lett. 1992;309:25–8.
34. Munirajan AK, Ando K, Mukai A, Takahashi M, Suenaga Y, Ohira M, et al. KIF1Bbeta functions as a haploinsufficient tumor suppressor gene mapped to chromosome 1p36.2 by inducing apoptotic cell death. J Biol Chem. 2008;283:24426–34.
35. 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;15:405.
36. Van Loo P, Nordgard SH, Lingjaerde OC, Russnes HG, Rye IH, Sun W, et al. Allele-specific copy number analysis of tumors. Proc Natl Acad Sci U S A. 2010;107:16910–5.
37. Yao R, Zhang C, Yu T, Li N, Hu X, Wang X, et al. Evaluation of three read-depth based CNV detection tools using whole-exome sequencing data. Mol Cytogenet. 2017;10:30.
38. Zare F, Dow M, Monteleone N, Hosny A, Nabavi S. An evaluation of copy number variation detection tools for cancer using whole exome sequencing data. BMC Bioinformatics. 2017;18:286.
39. Olshen AB, Venkatraman ES, Lucito R, Wigler M. Circular binary segmentation for the analysis of array-based DNA copy number data. Biostatistics. 2004;5:557–72.
40. Talevich E, Shain AH, Botton T, Bastian BC. CNVkit: genome-wide copy number detection and visualization from targeted dna sequencing. PLoS Comput Biol. 2016;12:e1004873.
41. Attiyeh EF, Diskin SJ, Attiyeh MA, Mosse YP, Hou C, Jackson EM, et al. Genomic copy number determination in cancer cells from single nucleotide polymorphism microarrays based on quantitative genotyping corrected for aneuploidy. Genome Res. 2009;19:276–83.
42. Grundy PE, Breslow NE, Li S, Perlman E, Beckwith JB, Ritchey ML, et al. Loss of heterozygosity for chromosomes 1p and 16q is an adverse prognostic factor in favorable-histology Wilms tumor: a report from the National Wilms Tumor Study Group. J Clin Oncol. 2005;23:7312–21.
43. Uryu K, Nishimura R, Kataoka K, Sato Y, Nakazawa A, Suzuki H, et al. Identification of the genetic and clinical characteristics of neuroblastomas using genome-wide analysis. Oncotarget. 2017;8:107513–29.
44. Altura RA, Maris JM, Li H, Boyett JM, Brodeur GM, Look AT. Novel regions of chromosomal loss in familial neuroblastoma by comparative genomic hybridization. Genes Chromosom Cancer. 1997;19:176–84.
45. Koldobskiy MA, Chakraborty A, Werner JK Jr, Snowman AM, Juluri KR, Vandiver MS, et al. p53-mediated apoptosis requires inositol hexakisphosphate kinase-2. Proc Natl Acad Sci U S A. 2010;107:20947–51.
46. Morrison BH, Haney R, Lamarre E, Drazba J, Prestwich GD, Lindner DJ. Gene deletion of inositol hexakisphosphate kinase 2 predisposes to aerodigestive tract carcinoma. Oncogene. 2009;28:2383–92.
47. Sandstrom J, Balian A, Lockowandt R, Fornander T, Nordenskjold B, Lindstrom L, et al. IP6K2 predicts favorable clinical outcome of primary breast cancer. Mol Clin Oncol. 2021;14:94.
48. Koyama H, Zhuang T, Light JE, Kolla V, Higashi M, McGrady PW, et al. Mechanisms of CHD5 Inactivation in neuroblastomas. Clin Cancer Res. 2012;18:1588–97.
49. Michels E, Hoebeeck J, De Preter K, Schramm A, Brichard B, De Paepe A, et al. CADM1 is a strong neuroblastoma candidate gene that maps within a 3.72 Mb critical region of loss on 11q23. BMC Cancer. 2008;8:173.
50. Mandriota SJ, Valentijn LJ, Lesne L, Betts DR, Marino D, Boudal-Khoshbeen M, et al. Ataxia-telangiectasia mutated (ATM) silencing promotes neuroblastoma progression through a MYCN independent mechanism. Oncotarget. 2015;6:18558–76.
51. Montgomery SB, Lappalainen T, Gutierrez-Arcelus M, Dermitzakis ET. Rare and common regulatory variation in population-scale sequenced human genomes. PLoS Genet. 2011;7:e1002144.
52. MacArthur DG, Balasubramanian S, Frankish A, Huang N, Morris J, Walter K, et al. A systematic survey of loss-of-function variants in human protein-coding genes. Science. 2012;335:823–8.
53. Lappalainen T, Sammeth M, Friedlander MR, Hoen PA, Monlong J, Rivas MA, et al. Transcriptome and genome sequencing uncovers functional variation in humans. Nature. 2013;501:506–11.
54. Popp MW, Maquat LE. Nonsense-mediated mRNA decay and cancer. Curr Opin Genet Dev. 2018;48:44–50.
55. Chamieh H, Ballut L, Bonneau F, Le Hir H. NMD factors UPF2 and UPF3 bridge UPF1 to the exon junction complex and stimulate its RNA helicase activity. Nat Struct Mol Biol. 2008;15:85–93.
56. Kashima I, Yamashita A, Izumi N, Kataoka N, Morishita R, Hoshino S, et al. Binding of a novel SMG-1-Upf1-eRF1-eRF3 complex (SURF) to the exon junction complex triggers Upf1 phosphorylation and nonsense-mediated mRNA decay. Genes Dev. 2006;20:355–67.

57. Garcia-Lopez J, Wallace K, Otero JH, Olsen R, Wang YD, Finkelstein D, et al. Large 1p36 deletions affecting arid1a locus facilitate Mycn-driven oncogenesis in neuroblastoma. Cell Rep. 2020;30:454–64 e455.
58. Trochet D, Bourdeaut F, Janoueix-Lerosey I, Deville A, de Pontual L, Schleiermacher G, et al. Germline mutations of the paired-like homeobox 2B (PHOX2B) gene in neuroblastoma. Am J Hum Genet. 2004;74:761–4.
59. Bourdeaut F, Trochet D, Janoueix-Lerosey I, Ribeiro A, Deville A, Coz C, et al. Germline mutations of the paired-like homeobox 2B (PHOX2B) gene in neuroblastoma. Cancer Lett. 2005;228:51–8.
60. Zhang W, Yu Y, Hertwig F, Thierry-Mieg J, Zhang W, Thierry-Mieg D, et al. Comparison of RNA-seq and microarray-based models for clinical endpoint prediction. Genome Biol. 2015;16:133.
61. Luscher B, Mitchell PJ, Williams T, Tjian R. Regulation of transcription factor AP-2 by the morphogen retinoic acid and by second messengers. Genes Dev. 1989;3:1507–17.
62. Ikram F, Ackermann S, Kahlert Y, Volland R, Roels F, Engesser A, et al. Transcription factor activating protein 2 beta (TFAP2B) mediates noradrenergic neuronal differentiation in neuroblastoma. Mol Oncol. 2016;10:344–59.
63. Noguchi T, Tsuda M, Takeda H, Takada T, Inagaki K, Yamao T, et al. Inhibition of cell growth and spreading by stomach cancer-associated protein-tyrosine phosphatase-1 (SAP-1) through dephosphorylation of p130cas. J Biol Chem. 2001;276:15216–24.
64. Takada T, Noguchi T, Inagaki K, Hosooka T, Fukunaga K, Yamao T, et al. Induction of apoptosis by stomach cancer-associated protein-tyrosine phosphatase-1. J Biol Chem. 2002;277:34359–66.
65. Pirinen M, Lappalainen T, Zaitlen NA, Consortium GT, Dermitzakis ET, Donnelly P, et al. Assessing allele-specific expression across multiple tissues from RNA-seq read data. Bioinformatics. 2015;31:2497–504.
66. Lee S, Lee S, Ouellette S, Park WY, Lee EA, Park PJ. NGSCheckMate: software for validating sample identity in next-generation sequencing studies within and across data types. Nucleic Acids Res. 2017;45:e103.
67. Houtgast EJ, Sima VM, Bertels K, Al-Ars Z. Hardware acceleration of BWA-MEM genomic short read mapping for longer read lengths. Comput Biol Chem. 2018;75:54–64.
68. McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernytsky A, et al. The genome analysis toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. Genome Res. 2010;20:1297–303.
69. McLaren W, Gil L, Hunt SE, Riat HS, Ritchie GR, Thormann A, et al. The Ensembl variant effect predictor. Genome Biol. 2016;17:122.
70. Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, et al. STAR: ultrafast universal RNA-seq aligner. Bioinformatics. 2013;29:15–21.
71. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, et al. Genome Project Data Processing S: The sequence alignment/map format and SAMtools. Bioinformatics. 2009;25:2078–9.
72. Liao Y, Smyth GK, Shi W. featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. Bioinformatics. 2014;30:923–30.
73. Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol. 2014;15:550.
74. van de Geijn B, McVicker G, Gilad Y, Pritchard JK. WASP: allele-specific software for robust molecular quantitative trait locus discovery. Nat Methods. 2015;12:1061–3.
75. Gu Z, Eils R, Schlesner M. Complex heatmaps reveal patterns and correlations in multidimensional genomic data. Bioinformatics. 2016;32:2847–9.
76. Lawrence M, Huber W, Pages H, Aboyoun P, Carlson M, Gentleman R, et al. Software for computing and annotating genomic ranges. PLoS Comput Biol. 2013;9:e1003118.
77. Yin T, Cook D, Lawrence M. ggbio: an R package for extending the grammar of graphics for genomic data. Genome Biol. 2012;13:R77.
78. SEQC/MAQC-III Consortium. A comprehensive assessment of RNA-seq accuracy, reproducibility and information content by the sequencing quality control consortium. Nat Biotechnol. 2014;32:903–14.
79. Consortium GT, Laboratory DA, Coordinating Center -Analysis Working G, Statistical Methods groups-Analysis Working G, Enhancing Gg, Fund NIHC, Nih/Nci, Nih/Nhgri, Nih/Nimh, Nih/Nida, et al. Genetic effects on gene expression across human tissues. Nature. 2017;550:204–13.
80. Wang C, Gong B, Bushel PR, Thierry-Mieg J, Thierry-Mieg D, Xu J, et al. The concordance between RNA-seq and microarray data depends on chemical treatment and transcript abundance. Nat Biotechnol. 2014;32:926–32.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.