SUPPLEMENTAL INFORMATION

Somatic structural variation targets neurodevelopmental genes and identifies SHANK2 as a tumor suppressor in neuroblastoma

Gonzalo Lopez1-3+, Karina L. Conkrite1,2+, Miriam Doepner1,2, Komal S. Rathi3, Apexa Modi1,2,11, Zalman Vaksman1,2,4, Lance M. Farra1,2, Eric Hyson1,2, Moataz Noureddine1,2, Jun S. Wei5, Malcolm A. Smith10, Shahab Asgharzadeh6,7, Robert C. Seeger6,7, Javed Khan5, Jaime Guidry Auvil9, Daniela S. Gerhard9, John M. Maris1,2,8,12, Sharon J. Diskin1,2,4,8,12*

1Division of Oncology, Children’s Hospital of Philadelphia, Philadelphia, PA, USA.
2Center for Childhood Cancer Research, Children’s Hospital of Philadelphia, Philadelphia, PA, USA.
3Department of Genetics and Genomic Sciences and Icahn Institute for Data Science and Genomic Technology, Icahn School of Medicine at Mount Sinai, New York, NY 10029, USA.
4Department of Biomedical and Health Informatics, Children’s Hospital of Philadelphia, Philadelphia, PA, USA.
5Oncogenomics Section, Genetics Branch, Center for Cancer Research, National Cancer Institute, Bethesda, MD, USA.
6Division of Hematology, Oncology and Blood and Marrow Transplantation, Keck School of Medicine of the University of Southern California, Los Angeles, CA, USA.
7The Saban Research Institute, Children’s Hospital of Los Angeles, Los Angeles, CA, USA.
8Department of Pediatrics, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA, USA.
9Office of Cancer Genomics, National Cancer Institute, Bethesda, MD, USA.
10Cancer Therapy Evaluation Program, National Cancer Institute, Bethesda, MD, USA.
11Genomics and Computational Biology, Biomedical Graduate Studies, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA, USA.
12Abramson Family Cancer Research Institute, Perelman School of Medicine at the University of Pennsylvania, Philadelphia, PA, USA.

* Equal contribution

* Corresponding Author:
Sharon J. Diskin, Ph.D.
3501 Civic Center Boulevard, CTRB 3026
Philadelphia, PA 19104
Phone: (610) 639-7754
Email: diskin@email.chop.edu
I. SUPPLEMENTAL METHODS

Copy number segmentation, visualization (IGV) and recurrence analysis
CGI "somaticCnvDetailsDiploidBeta" files provide information on estimated ploidy and tumor/blood coverage ratio for every 2-kb along the genome. We used custom scripts to reformat coverage data to be processed with “copynumber” R bioconductor package \(^1\). We then utilized Winsorization (winsorize) data smoothing and segmentation with piecewise constant segmentation (pcs) algorithm with attributes kmin=2 and gamma=1000. Segmented data was visualized with IGV and used as input to GISTIC2.0\(^2\). GISTIC attributes were as follows: -v 30 -refgene hg19 -genegistic 1 -smallmem 1 -broad 1 -twoside 1 -brlen 0.98 -conf 0.90 -armpeel 1 -savegene 1 -gcm extreme -js 2 -rx 0.

Filtering of CGI SV calls
The CGI Cancer Pipeline 2.0 produces a full report including quality control, variant calling and CNV analyses. The “somaticAllJunctionsBeta” files provide unfiltered information for individual junctions detected in a tumor genome that were absent in the corresponding normal genome. We filtered high-confidence somatic variants in entire TARGET repertoire of tumor datasets including neuroblastoma (NBL, n=135), Acute Lymphoblastic Leukemia (ALL, n=165), Acute Myeloid Leukemia (AML, n=183), Osteosarcoma (OS, n=18) and Wilms Tumors (WT, n=80). We then removed germline variants that passed CGI filters as well as artifacts and low confidence variants often observed across tumor datasets. We used the Database of Genomic Variants (DGV v. 2016-05-15, GRCh37) to remove SVs which reciprocal overlap with DGV annotated common events was higher than 50%. We only filtered variants which type matched in both CGI SV set and DGV database. All tumor variants (before and after filtering) are accessible in the manuscript GitHub repository: https://github.com/diskin-lab-chop/NB_structural_variants/.

Mapping and annotation of filtered alignment-based SVs (SJ-BP)
RefSeq gene definitions for hg19 were downloaded from UCSC Genome Browser (10/31/2018) in order to map SV calls to nearby genes. We used two approaches to map variants. First, numerical changes (tandem-duplications and deletions size <2Mb) containing whole genes were used to define copy number alterations. Second, we mapped breakpoints relative to gene exonic coordinates. SVs were considered ‘disrupting’ when
either one of the breakpoints localized between transcription start and ends of any of isoforms of a gene and ‘proximal’ when localized within 100-kb upstream or 25-kb downstream the most distal isoform of each gene. In addition, we considered ‘intronic’ SVs as those in which both breakpoints mapped to the same intron.

**Mapping and annotation of DNA amplifications, deep deletions and copy number breakpoints from WGS (RD-BP) and SNP arrays (CN-BP)**

Breakpoints were called from segmentation profiles. Due to artifacts observed at subtelomeric and pericentromeric regions, these regions were excluded. For the rest of the genome, a breakpoint was called when the absolute value of the copy number log-ratio difference between contiguous segments is higher than 0.152 for SNP arrays and 0.304 for WGS; both cutoffs account for 10% and 20% copy number change respectively (i.e for diploid regions, ΔCN = 0.2 and ΔCN = 0.4 respectively). For the WGS dataset, we called amplifications when CN >=8 and deep deletions when CN <= 0.5. For SNP arrays, we used less stringent cutoffs for amplification (CN >= 4.5) and deep deletions (CN <= 0.9). These relaxed criteria were selected due to SNP arrays having a narrower dynamic range and lower resolution. We used RefSeq gene definitions for hg19 downloaded from UCSC Genome Browser (October 31st, 2018 version) to map copy number alterations and breakpoints to nearby genes. Genes were considered amplified or deep deleted when all isoforms were contained within the altered segment boundaries. Breakpoints were considered ‘disrupting’ when the breakpoint localized between transcription start and ends of any isoform of a gene and proximal when localized at within 100-kb upstream or 25-kb.

**Co-localization of breakpoints derived from WGS (SJ-BP & RD-BP) and SNP arrays (CN-BP)**

We studied the co-localization of breakpoints derived from alternative measurements. First, we compared SJ-BPs and RD-BPs in each of the 135 WGS samples; overall, 30.5% of SJ-BPs co-localize with a RD-BPs (Supplemental Fig. S3A) whereas 62% RD-BPs matched with SJ-BPs (Supplemental Fig. S3B). We next evaluated the co-localization of breakpoints across WGS and SNP platforms within the subset of 52 overlapping samples. 50.2% of CN-BPs from SNP arrays co-localized with SJ-BPs from the WGS dataset (Supplemental Fig. S3C) whereas only 8.2% of SJ-BPs co-localize with CN-BPs (Supplemental Fig. S3D). Furthermore, when comparing dosage-based breakpoints across platforms (RD-BP and CN-BP), 23.6% RD-BPs where found co-localizing CN-BP (Supplemental Fig. S3E) whereas 66.6% CN-BP co-localized with RD-BPs (Supplemental
Finally, we performed a randomized test by sample shuffling ($N=1000$) in order to evaluate whether each of the co-localization percentages listed above could arise by chance or due to recurrence of SVs across samples. All randomized percentage distributions range between 0.7% and 2.3%; in all cases the null hypothesis was discarded ($p$-value < 0.001, Supplemental Fig. 3G-L).

**Tumor mutational burden analyses**
We obtained measures representative of the burden of different mutation types under study (including SNVs, SVs and BPs). To this end, the density of mutations of every type is calculated as the average number of mutations in a given sample per sequence window (10Mb for SVs and BPs and 1Mb for SNVs). Instead of a single density value per sample we measure mutational densities for each chromosomal arm, excluding short arms with very low mappability (Chrs 13p, 14p, 15p, 21p, 22p and Y). The remaining 41 chromosomal arms in each sample represent single sample distributions of mutational densities from which quantiles are obtained. We used the interquartile mean (IQM) since it offered a measure robust against outliers while conserving the variability across samples even in low-density breakpoint samples.

**Filtering and annotation of likely pathogenic somatic SNVs**
The CGI cancer pipeline 2.0 provides somatic variant calls for SNVs and small indels. Given the gapped nature of CGI reads which leads to high noise to signal ratio, we incorporated additional SNV filtering. We first annotated CGI SNV calls using Variant Effect Predictor (VEP) pipeline $^3$. Our filter follows two steps: 1) collect high quality somatic non-synonymous coding variants (Phred like Fisher’s exact test $P<0.001$) annotated as having a moderate or high functional impact; this set of variants was combine with COSMIC catalogue of pathogenic variants (release v84). 2) Hot-spot analysis of variants from our combined catalogue (step 1) to identify both clonal and low allele frequency pathogenic variants.
II. SUPPLEMENTAL TABLES

Supplemental Table S1: Neuroblastoma patient clinical and biological characteristics.

| Covariate                  | WGS (CGI)                  | SNP (Illumina)          | RNA-seq (Illumina)   | Expression array (Affy-HuEx) |
|----------------------------|----------------------------|-------------------------|----------------------|-----------------------------|
| INSS Stage                 | 105 (Stage 4)              | 650 (Stage 4)           | 214 (Stage 4)        | 105 (Stage 4)               |
|                            | 6 (Stage 3)                | 98 (Stage 3)            | 6 (Stage 3)          | 1 (Stage 3)                 |
|                            | 1 (Stage 2B)               | 26 (Stage 2B)           | 30 (Stage 1)         |                             |
|                            | 23 (Stage 4S)              | 16 (Stage 2A)           | 2 (Unknown)          |                             |
|                            |                            | 82 (Stage 1)            |                      |                             |
|                            |                            | 37 (Stage 4S)           |                      |                             |
|                            |                            | 5 (Unknown)             |                      |                             |
| MYCN amplification status  | 29 (Amp)                   | 241 (Amp)               | 31 (Amp)             | 60 (Amp)                    |
|                            | 106 (Not-Amp)              | 670 (Not-Amp)           | 121 (Not-Amp)        | 185 (Not-Amp)               |
|                            | 1 (Unknown)                | 3 (Unknown)             | 1 (Unknown)          | 2 (Unknown)                 |
| COG Risk Group             | 106 (High-risk)            | 695 (High-risk)         | 127 (High-risk)      | 215 (High-risk)             |
|                            | 14 (Inter-Risk)            | 70 (Inter-Risk)         | 13 (Inter-Risk)      | 30 (Low-risk)               |
|                            | 15 (Low-risk)              | 145 (Low-risk)          | 13 (Low-risk)        | 2 (Unknown)                 |
| Vital status               | 74 (Alive)                 | 454 (Alive)             | 75 (Alive)           | 105 (Alive)                 |
|                            | 61 (Dead)                  | 353 (Dead)              | 78 (Dead)            | 140 (Dead)                  |
|                            |                            | 107 (Unknown)           |                      | 2 (Unknown)                 |
| Age at diagnosis group     | 32 (< 18 months)           | 265 (< 18 months)       | 29 (< 18 mo/old)     | 32 (< 18 mo/old)            |
|                            | 103 (>18 months)           | 646 (>18 months)        | 124 (>18 mo/old)     | 215 (>18 mo/old)            |
|                            |                            | 3 (Unknown)             |                      |                             |
| Gender                     | 52 (Female)                | 408 (Female)            | 64 (Female)          | 106 (Female)                |
|                            | 83 (Male)                  | 506 (Male)              | 89 (Male)            | 141 (Male)                  |
Supplemental Table S2: Cohort clinical data and sample description (xls file)
A survey of all samples used in this study including clinical information and availability for each genomic profiling platform.

Supplemental Table S3: Filtered structural variants calls (xls file)
Resulting SVs in neuroblastoma tumors after additional quality control filtering.

Supplemental Table S4: Chromothripsis and high-breakpoint density chromosomes (xls file)
Quantification of different measures of structural variations per chromosomal arm used to infer chromothripsis in WGS samples.

Supplemental Table S5: Recurrently altered genes by alternative breakpoint analyses (xls file)
Survey of genes affected by structural variants (SVs) obtained by alternative breakpoint analyses and cohorts (SJ-BP, RD-BP and CN-BP). Altered genes are classified into different lists according to SV impact on “coding” or “non-coding” regions at each gene loci.

Supplemental Table S6: Orthogonally identified recurrently altered genes (xls file)
List of 77 genes disrupted by structural variants (SVs) with alignment (SJ-BP) and read-depth (RD-BP) based evidences from the WGS dataset.

Supplemental Table S7: SNVs mapping to recurrently altered genes (xls file)
Pathogenic non-synonymous SNVs found in the coding regions and splice sites of recurrently altered genes. Mutation Annotation Format (MAF) data file.

Supplemental Table S8: Primer sequences used for Sanger sequencing validations (xls file)
The table informs about sequences used for Sanger validation of SVs of ALK, ATRX, DLG2, SHANK2 and TERT; includes Junction Id which relates to variants described in Supplementary Table 3.

Supplemental Table S9: Combined gene fusion transcripts and DNA structural variations (xls file)
List of gene fusions: junction location, gene fusion method used to identify the event including: STARfusion, fusion-CATCHER and DeFUSE (DCC). See Methods.

Supplemental Table S10: Functional enrichment analysis of recurrently altered genes (xls file)
Enrichment analysis results for genes harboring recurrent SVs. Analysis derived from ToppGene (see Methods). Separate tabs are included for each of the SV detection methods, and separately for coding and non-coding.
Supplemental Fig S1. CNV and GISTIC analyses for neuroblastomas tumors subtypes.

Integrated Genome Viewer (IGV) visualization of DNA copy number gains (red) and losses (blue) across neuroblastoma subtypes in the (A) WGS and (B) SNP datasets. (C-H) GISTIC q-plot analysis neuroblastoma subtypes indicating frequent deletions (blue) and gains (red) obtained from combined LOWR+INTR tumors.
profiled with (C) WGS and (D) SNP arrays, MNA group profiled with (E) WGS and (F) SNP arrays and HR-NA group profiled with (G) WGS and (H) SNP arrays.

Supplemental Fig S2. Identification of alignment based structural variants across TARGET datasets. (A) distribution of the number of discordantly aligned read pairs supporting each structural variant across TARGET histotypes. (B-D) Structural Variant genomic sizes for (B) deletions, (C) tandem-duplications and (D) inversions + probable-inversion.
Supplemental Fig S3. Orthogonal validation of breakpoint detection methods

(A-F) Cross platform co-localization of breakpoints across 52 samples with overlapping WGS and SNP profiles. Overall co-localization percentages at the right-side panel of each bar plot. (G-L) Randomized empirical distributions of co-localization percentages using sample shuffling; every null distribution allows calculating empirical p-values for the observed overall co-localization percentages panels A-F.
Supplemental Fig S4. TERT rearrangements and expression across neuroblastomas subtypes

(A) SVs at the TERT locus in 25 WGS samples based on RD-BPs combined with SJ-BPs and 15 tumors profiled by SNP array using CN-BPs. Nearest upstream TERT junctions derived by sequencing are highlighted in text; ‘S’ at the left of the panel indicates positive validation by Sanger sequencing. (B) Expression of TERT in primary tumors with different genetic backgrounds obtained from RNA-seq TARGET cohort. (C) Expression of TERT in primary tumors with different genetic backgrounds obtained from HumanExon arrays TARGET cohort.
Supplemental Fig S5. Sanger sequencing validation of rearrangements near TERT gene

Primers and additional details described in Supplemental Table S8.
Supplemental Fig S6. MYCN and ALK structural variants and Sanger validation

(A) IGV track visualization plot shows copy number segmentation data around the MYCN amplicon (Chr 2p.24) region. Three groups separated at the left margin: MNA (red), HR-NA (orange) and S4s (dark green). Left margin separates samples by gender within each tumor group: female (light blue) and male (light green).

(B) IGV track visualization plot shows copy number segmentation data comprising MYCN and ALK loci in cases with ALK associated SVs; variant position and type highlighted for ALK; “S” at the left of the panel indicates that Sanger sequencing validation is available.

(C) Sanger sequencing validation of rearrangements near ALK gene in 3
cases with available DNA (PARETE harbored multiple SVs). Primers and additional details described in Supplemental Table 8.

Supplemental Fig S7. Chromothripsis in chromosome 2 associates with MYCN amplification

Circos plots representing copy number and large structural variants (<100Kb) in samples with high BP and SV density indicative of chromothripsis. Samples with MYCN amplification are highlighted in red text and TERT rearrangements in blue text.
Supplemental Fig S8. Chromothripsis on chromosome 5 associates with TERT rearrangements

Circos plots representing copy number and large structural variants (<100Kb) in samples with high BP and SV density indicative of chromothripsis as derived from Supplemental Table S3. Samples with TERT rearrangements are highlighted in blue text.
Supplemental Fig S9. Interchromosomal chromothriptic events absent MYCN and TERT variants

Circos plots representing copy number and large structural variants (<100Kb) in samples with high BP and SV density indicative of chromothripsis as derived from Fig. 3B.
Supplemental Fig S10. Variability of SV burden across subtypes

(A) Number of variants (top) and frequency (bottom) of SVs by variant type across different TARGET histotypes. (B) Number of variants (top) and frequency (bottom) of SVs by variant type across different neuroblastoma subtypes (C) Localization of structural variants in Chromosome 2 represented as a histogram overlaying MNA (red) and HR-NA (orange) tumors. (D) A comparison of the differential frequencies between MNA and HR-NA of structural variants by type when chromosome 2 is ignored (left) versus included (right); orange bars indicate higher frequency in HR-NA and red bars indicate higher frequencies in MNA; p-values calculated by Wilcoxon rank-sum test. (E-F) By chromosome comparison between MNA and HR-NA of the inter-quantile average
number of SVs including (E) inversions and (F) deletions (G-H) Mutational burden analysis for (G) SV and (H) SNVs using the Inter-quantile mean (IQM) burden across 41 chromosome arms.

Supplemental Fig S11. Schematic of SV breakpoint classification according to impact on known genes
Schematic representation of SVs derived from (A) junction breakpoints (SJ-BPs) classified according to their impact on known genes and (B) read-depth and copy number breakpoints (RD-BPs, CN-BPs) classified according to their impact on known genes.
Supplemental Fig S12. Recurrently altered genes: *PTPRD* and *ATRX*

Genomic visualization of sequence junction based structural variants in (A) *PTPRD* and (B) *ATRX* loci. Boxes represent the spanned region of deletions (blue), duplications (red) and inversions (yellow). Translocations are represented as vertical line followed by the sample names and genomic coordinate of the destination region. Coordinates of structural variants outside the plot genomic span also included in the labels.
Supplemental Fig S13. Sanger sequencing validation of ATRX deletions
Primers and additional details described in Supplemental Table S8.
Supplemental Fig S14. Chromosome 11 breakpoints frequently disrupt \textit{SHANK2} and \textit{DLG2}

Chromosome 11 segmentation plot for HR-NA tumors obtained from WGS highlights regions of high breakpoint density near \textit{SHANK2} and \textit{DLG2} loci.
Supplemental Fig S15. Recurrently altered genes: AUTC2 and CACNA2D3

(A) Sequence junction based structural variants in AUTC2 locus. Boxes represent the spanned region of deletions (blue) and duplications (red) and inversions (yellow). Translocations are represented as vertical line followed by the sample names and genomic coordinate of the destination region. Coordinates of structural variants outside the plot genomic span also included in the labels. (B) IGV copy number visualization of CACNA2D3 locus with sequence junctions overlay.
Supplemental Fig S16. Recurrently altered genes: LINC00910, CDKN2A and PLXDC

Genomic visualization of breakpoints implicated in eQTL associations of (A) LINC00910, (B) CDKN2A and (C) PLXDC1 loci. Information was derived from sequence junctions in (A) and read-depth copy number for (B) and (C). Regions in blue indicate deletions or copy number loss, regions in red indicate duplications and gains. Translocations are represented as vertical line followed by the sample names and genomic coordinate of the destination region.
Supplemental Fig S17. Sanger sequencing validation of \textit{SHANK2} structural variants

Primers and additional details described in Supplemental Table S8.
Supplemental Fig S18. Sanger sequencing validation of DLG2 structural variants

Primers and additional details described in Supplemental Table S8.
Supplemental Fig S19. Transcriptional effect of structural variants: eQTL and gene fusions

(A) Differential expression analysis of genes with samples harboring SVs compared to unaltered samples using a Wilcoxon rank sum test. Gene names colored in red map into Chr 2p24 region near MYCN; Genes colored in blue map into Chr 5p15 near TERT. Asterisk indicates the p-value cut off (*** = P < 0.001; ** = P<0.01; *= P<0.05); n.a. indicates that the test couldn’t be done due to lack of data points or because the gene was not included in the HumanExon array probeset design. (B) Overlap across three different gene fusion callers from 153 RNA-seq neuroblastoma samples. (C) Overlap across three different gene fusion callers from RNA-seq with evidence of translocation events in 89 samples with both RNA-seq and WGS profiles, see also Supplemental Table S9.
Supplemental Fig S20. Pathway and disease enrichment of genes altered by structural variants.

(A) Function enrichment analysis bar plots for genes recurrently altered based on SJ-BP analyses across 4 different TARGET tumors (ALL, AML, OS and WT). (B-D) Neuroblastoma enrichment analysis bar plots for genes recurrently altered based on proximal and intronic breakpoint analyses of (B) SJ-BPs, (C) RD-BPs and (D) CN-BPs.
Supplemental Fig S21. Neurodevelopmental pathways are down-regulated in high-risk neuroblastoma

(A-B) Gene Set Enrichment Analysis of high- versus low-risk tumors from the RNA-seq data showing enrichment of (A) autism spectrum disorder genes and (B) neuronal and synapse part genes. (C) Volcano plot showing differential expression between high- and low-risk highlighting genes with recurrent SVs and their functional classification (D) Subtype specific high- versus low-risk differential expression analysis of 77 recurrently altered genes from Fig. 3I shown as scatter plot (MNA = x-axis, HR-NA = y axis). This figure replicates Fig. 4D-G (Affymetrix HuEx data) using the RNA-seq dataset (n=153 samples).
Supplemental Fig S22. Low SHANK2 expression in neuroblastoma is associated with poor survival

(A, B) Clustering analysis of SHANK2 exon level expression from Affymetrix HumanExon arrays in MNa, S4na and S1 tumors. The heatmap (A) shows higher exon expression level in S1 compared to MNa and S4na. The correlation matrix (B) shows two well-defined clusters associated with the two known coding isoforms of the gene.

(C) SHANK2 long isoform (NM_012309) expression decreases in high INSS stage across 498 primary neuroblastomas (SEQC dataset; GSE62564). (D) Kaplan-Meier analysis of SHANK2 long isoform (NM_012309) expression shows association of low expression with poor outcome (SEQC dataset, all samples). (E) decreased SHANK2 expression (NM_012309) is associated with reduced overall survival within the low- and intermediate-risk subsets (excluding high-risk).
Supplemental Fig S23. Genes of the postsynaptic density are down-regulated in high-risk neuroblastoma

“Postsynaptic membrane” genes were obtained from MsigDB Gene Ontology Cellular Component gene set library and subjected to GSEA analysis of differential gene expression (DGE) signatures. DGE were obtained by comparing high-risk tumor’s expression against stage low and intermediate risk samples in different datasets: (A) TARGET Affymetrix Human Exon Stage 1 vs. High-Risk, (B) SEQC 498 RNA-seq Stage 1 vs. High-risk, (C) TARGET RNA-seq Stage 4S vs. High-risk and (D) SEQC 498 RNA-seq Stage 4S vs. High-risk.
Supplemental Fig S24. SHANK2 expression in neuroblastoma cell lines

SHANK2 RNA-seq expression (FPKM) across 38 neuroblastoma cell lines.
Supplemental Fig S25. SHANK2 accelerates differentiation of neuroblastoma cells

(A, B) Confluence measures across SHANK2 expressing neuroblastoma cells, vector controls, and vehicle controls included for (A) Be(2)C and (b) SY5Y cell lines. (C, D) Total neurite length measurement in all cells over time for (C) Be(2)C and (D) SY5Y neuroblastoma cell lines. Arrow indicates time of ATRA introduction. (E, F) Neurite outgrowth normalized to cell body area over time in (E) Be(2)C and (F) SY5Y cell lines.
IV. SUPPLEMENTAL REFERENCES

1. Nilsen, G. et al. Copynumber: Efficient algorithms for single- and multi-track copy number segmentation. *BMC Genomics* **13**, 591 (2012).

2. Mermel, C.H. et al. GISTIC2.0 facilitates sensitive and confident localization of the targets of focal somatic copy-number alteration in human cancers. *Genome Biol* **12**, R41 (2011).

3. McLaren, W. et al. The Ensembl Variant Effect Predictor. *Genome Biol* **17**, 122 (2016).