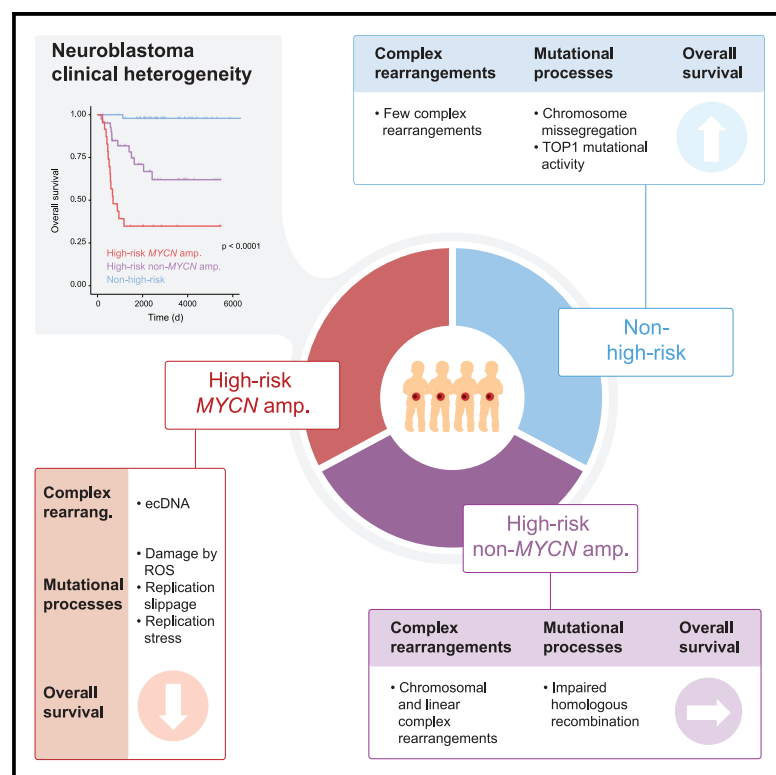


# Mutational topography reflects clinical neuroblastoma heterogeneity

## Graphical abstract



## Authors

Elias Rodriguez-Fos, Mercè Planas-Fèlix, Martin Burkert, ..., Kerstin Haase, Johannes H. Schulte, Anton G. Henssen

## Correspondence

henssenlab@gmail.com

## In brief

Rodriguez-Fos et al. reveal that the strong clinical heterogeneity observed in neuroblastoma patients is associated with differences in the activity of specific mutational processes. Their work suggests that replication slippage/stress, homologous recombination deficiency, and chromosome mis-segregation have important implications in neuroblastoma evolution and survival.

## Highlights

- Clinical neuroblastoma heterogeneity is linked to differences in mutational processes
- HR MYCN-amp neuroblastomas are characterized by signs of replication slippage and stress
- HR non-MYCN-amp neuroblastomas are defined by HRD signatures
- Non-HR neuroblastomas are marked by chromosome mis-segregation and TOP1 activity



## Resource

# Mutational topography reflects clinical neuroblastoma heterogeneity

Elias Rodriguez-Fos,<sup>1,2,11</sup> Mercè Planas-Fèlix,<sup>1,10,11</sup> Martin Burkert,<sup>3</sup> Montserrat Puiggròs,<sup>4</sup> Joern Toedling,<sup>2</sup> Nina Thiessen,<sup>6</sup> Eric Blanc,<sup>6</sup> Annabell Szymansky,<sup>2</sup> Falk Hertwig,<sup>2</sup> Naveed Ishaque,<sup>6</sup> Dieter Beule,<sup>6</sup> David Torrents,<sup>4,5</sup> Angelika Eggert,<sup>1,2</sup> Richard P. Koche,<sup>7</sup> Roland F. Schwarz,<sup>3,8,9</sup> Kerstin Haase,<sup>1,10,11</sup> Johannes H. Schulte,<sup>2,11</sup> and Anton G. Henssen<sup>1,2,3,6,10,11,12,\*</sup>

<sup>1</sup>Experimental and Clinical Research Center (ECRC) of the MDC and Charité Berlin, Berlin, Germany

<sup>2</sup>Department of Pediatric Oncology and Hematology, Charité – Universitätsmedizin Berlin, Corporate Member of Freie Universität Berlin, Humboldt-Universität zu Berlin, Berlin, Germany

<sup>3</sup>Berlin Institute for Medical Systems Biology (BIMSB), Max Delbrück Center for Molecular Medicine in the Helmholtz Association (MDC), Berlin, Germany

<sup>4</sup>Barcelona Supercomputing Center, Joint Barcelona Supercomputing Center – Center for Genomic Regulation – Institute for Research in Biomedicine Research Program in Computational Biology, Barcelona, Spain

<sup>5</sup>Institució Catalana de Recerca i Estudis Avançats (ICREA), Barcelona, Spain

<sup>6</sup>Berlin Institute of Health at Charité – Universitätsmedizin Berlin, Digital Health Center, Berlin, Germany

<sup>7</sup>Center for Epigenetics Research, Memorial Sloan Kettering Cancer Center, New York, NY, USA

<sup>8</sup>Center for Integrated Oncology (CIO), Cancer Research Center Cologne Essen (CCCE), Faculty of Medicine and University Hospital Cologne, University of Cologne, Cologne, Germany

<sup>9</sup>BIFOLD – Berlin Institute for the Foundations of Learning and Data, Berlin, Germany

<sup>10</sup>German Cancer Consortium (DKTK), Partner Site Berlin, and German Cancer Research Center (DKFZ), Heidelberg, Germany

<sup>11</sup>These authors contributed equally

<sup>12</sup>Lead contact

\*Correspondence: [henssenlab@gmail.com](mailto:henssenlab@gmail.com)  
<https://doi.org/10.1016/j.xgen.2023.100402>

## SUMMARY

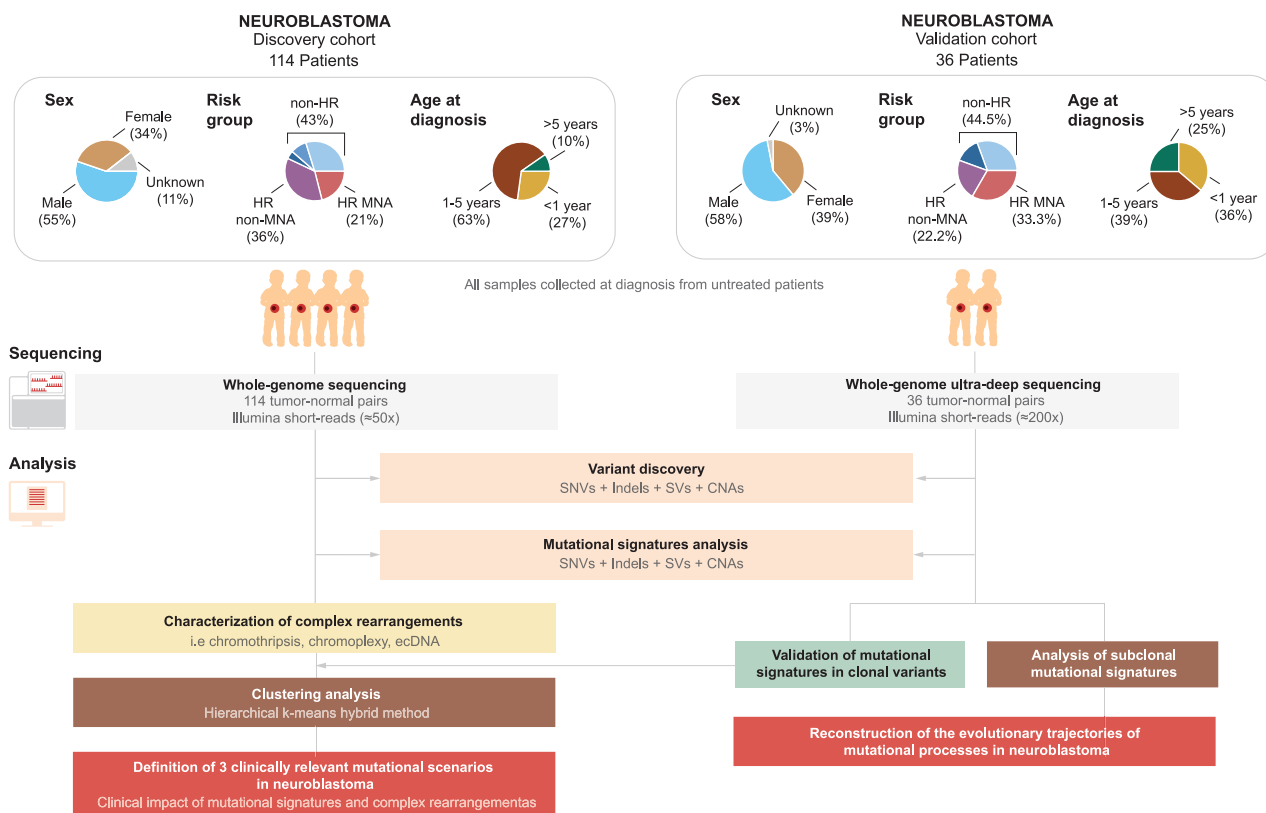
Neuroblastoma is a pediatric solid tumor characterized by strong clinical heterogeneity. Although clinical risk-defining genomic alterations exist in neuroblastomas, the mutational processes involved in their generation remain largely unclear. By examining the topography and mutational signatures derived from all variant classes, we identified co-occurring mutational footprints, which we termed mutational scenarios. We demonstrate that clinical neuroblastoma heterogeneity is associated with differences in the mutational processes driving these scenarios, linking risk-defining pathognomonic variants to distinct molecular processes. Whereas high-risk *MYCN*-amplified neuroblastomas were characterized by signs of replication slippage and stress, homologous recombination-associated signatures defined high-risk non-*MYCN*-amplified patients. Non-high-risk neuroblastomas were marked by footprints of chromosome mis-segregation and *TOP1* mutational activity. Furthermore, analysis of subclonal mutations uncovered differential activity of these processes through neuroblastoma evolution. Thus, clinical heterogeneity of neuroblastoma patients can be linked to differences in the mutational processes that are active in their tumors.

## INTRODUCTION

The presence of somatic mutations is a hallmark of cancer genomes.<sup>1</sup> Diverse types of somatic mutations are a result of different endogenous and/or exogenous mutational processes, including replication errors, exposure to DNA-damaging agents, expression of developmentally restricted recombinases,<sup>2</sup> and errors in DNA-repair mechanisms.<sup>3</sup> These processes imprint characteristic mutational patterns in the genome, termed mutational signatures.<sup>4</sup> Recent analyses in multiple cancer types have extracted signatures associated with single-nucleotide variants (SNVs),<sup>5,6</sup> small insertions and deletions (indels),<sup>6</sup> copy-number alterations (CNAs),<sup>7,8</sup>

and structural variants (SVs).<sup>9</sup> Some signatures are linked to known biological processes active in cancer, whereas others have yet-unknown etiologies.<sup>5,6,10</sup> In contrast with mutational signatures derived from SNVs and indels, those derived from CNAs and SVs remain difficult to classify, and their etiologies remain largely unknown. Various complex SV classes have recently been defined on the basis of their topography, e.g., circular extrachromosomal DNA<sup>11</sup> (ecDNA), chromothripsis,<sup>12</sup> chromoplexy,<sup>13</sup> templated insertion chains (TICs),<sup>14</sup> breakage-fusion-bridge cycles<sup>15</sup> (BFBs), complex non-cyclic amplicons<sup>16</sup> (CnCs), tyfona, rigma, and pyrgo,<sup>17</sup> among others. Even though the study of mutational signatures and structural variant patterns has advanced in recent years, it remains





**Figure 1. Cohort characteristics and analysis strategy**

Top: description of the distribution (in percentage) of the 150 neuroblastoma samples from the discovery ( $n = 114$ ; see also Figure S1) and validation ( $n = 36$ ) cohorts, within sex groups (female, male, and unknown), risk groups (HR MNA, HR non-MNA, and non-HR: low-risk; low-risk stage 4S; intermediate-risk), and age at diagnosis groups (<1 year old, 1–5 years old, and >5 years old). Middle: summary of the sequencing datasets available for our analysis. WGS matched tumor-normal pairs. Bottom: description of the main steps carried out in our study, starting with the variant discovery and the mutational signatures analysis, followed by validation of the signatures and subclonal SNV-based signatures extraction, characterization of the complex rearrangements present in our samples, unsupervised clustering, and definition of the three clinically relevant neuroblastoma mutational scenarios presented in this work.

largely unclear how different mutational processes and variant topographies are mechanistically linked.

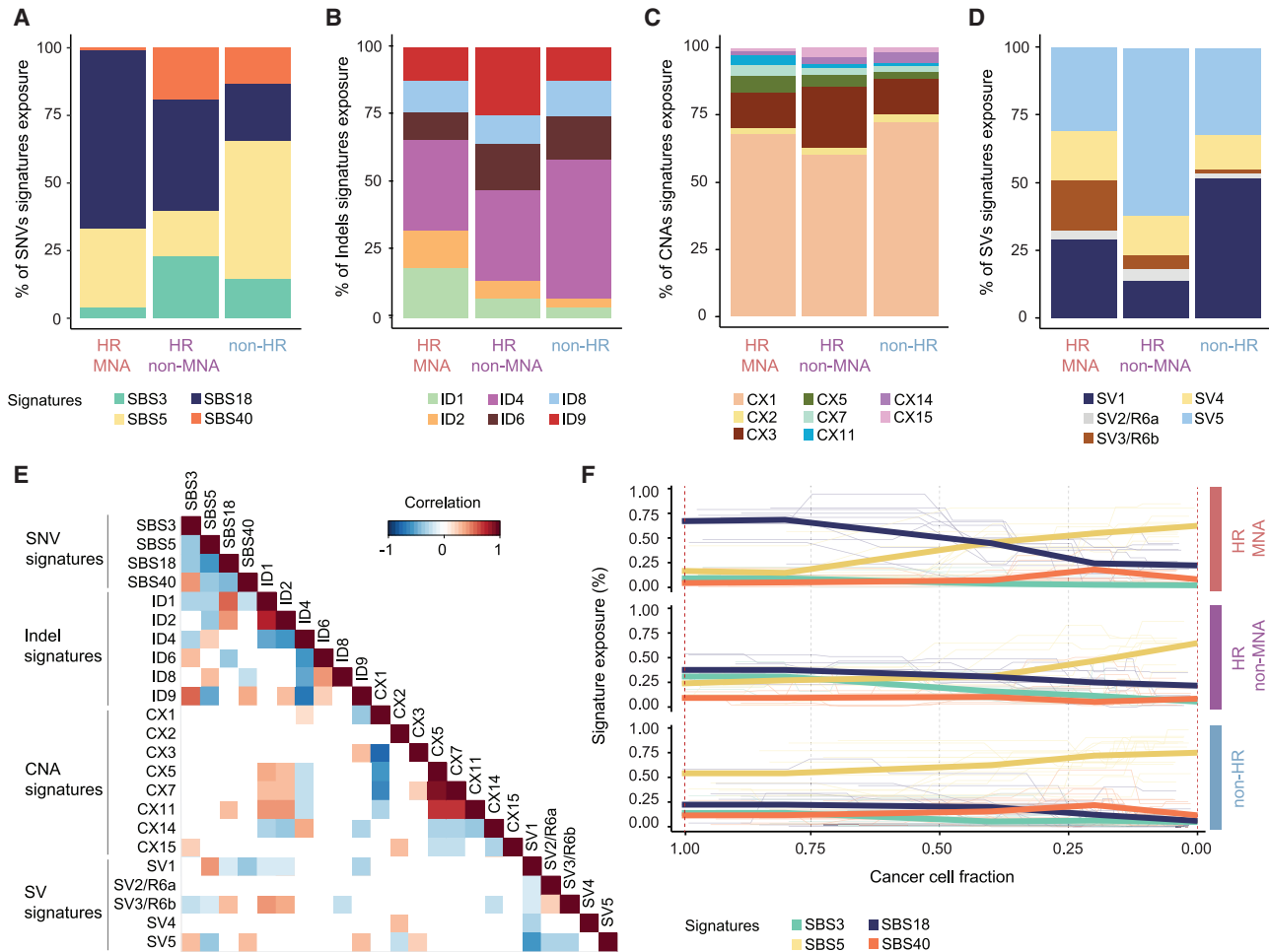
Neuroblastoma is a pediatric solid tumor characterized by strong clinical heterogeneity.<sup>18–20</sup> Patients can be classified into different disease risk groups depending on their clinical, genetic, and molecular characteristics, namely low-, intermediate- or high-risk groups, with the latter presenting poor prognosis despite intensive therapy and surgery.<sup>21</sup> Cytogenetic and genomic studies have identified many neuroblastoma risk-group-specific alterations<sup>19,22–26</sup> such as ecDNA and segmental chromosomal gains/losses in high-risk patients or whole chromosomal aberrations in low-risk patients.<sup>27–29</sup> Complex rearrangements involving oncogene amplicons are particularly frequent in neuroblastoma.<sup>30,31</sup> Recent studies have linked ecDNA to other complex variants such as chromothripsis and BFBs<sup>32,33</sup> and have identified circular recombination as a potential means of ecDNA evolution in neuroblastomas.<sup>33</sup> What causes these pathognomonic alterations remains largely unknown. We hypothesized that co-occurrence analyses of mutational signatures derived from independent variant types may identify new principles of neuroblastoma mutagenesis that

explain the differences in variant patterns observed across clinical risk groups.

## RESULTS

### SNV-based mutational signatures differ between neuroblastoma risk groups

To explore the link between different mutational patterns from independent variant types in neuroblastoma, we analyzed two publicly available, previously published cohorts of whole genomes from 103 matched tumor-normal pairs as well as 11 unpublished whole genomes, which were all derived from neuroblastoma patients treated according to the same clinical protocol. All sex, age, and disease stages were represented in this dataset (Figures 1 and S1). Pathognomonic chromosomal and driver gene alterations were identified at a comparable frequency in our cohort as described in other neuroblastoma cohorts<sup>26,30,34</sup> (Figure S1A). Recurrently mutated genes were comparable to those found in other cohorts (Figure S1A), evidencing that this cohort was representative for clinically heterogeneous neuroblastomas.



**Figure 2. Distribution and correlation of SNV-, indel-, SV-, and CNA-associated signatures in neuroblastoma clinical risk groups**

(A) Exposure (in percentage) of the four SNV-associated signatures (SBSs) identified in our neuroblastoma discovery cohort by clinical risk group (n = 114). Each color displays a different signature: SBS3, SBS5, SBS18, and SBS40 (see also Figures S2A–S2C). Columns are ordered by neuroblastoma clinical risk classification.

(B) Exposure (in percentage) of the six indel signatures (ID; insertions and deletions <50 bp) identified in our cohort by clinical risk group. Each color displays a different signature: ID1, ID2, ID4, ID6, ID8, and ID9 (see also Figures S2D–S2F). Columns are ordered by neuroblastoma clinical risk classification.

(C) Exposure (in percentage) of the eight CNA signatures (CX; gains, losses, amplifications, and homozygous deletions) extracted in our cohort by clinical risk group. Each color displays a different signature: CX1, CX2, CX3, CX5, CX7, CX11, CX14, and CX15 (see also Figures S2G and S2H). Columns are ordered by neuroblastoma clinical risk classification.

(D) Exposure (in percentage) of the six SV signatures (SV; deletions, duplications, translocations, and inversions) identified in our cohort by clinical risk group. Each color displays a different signature: SV1–SV6. SV2, and SV3 correspond to reference signatures R6b and R6a, respectively (see also Figures S3A–S3D). Columns are ordered by neuroblastoma clinical risk classification.

(E) Heatmap depicting the positive (red) and negative (blue) correlations between the signatures associated with different types of mutations (SNV, indel, SV, and CNA). Colors display the Spearman correlation coefficient. Only significant correlations are included (p < 0.05, false discovery rate [FDR] correction).

(F) Activity trajectories of the four SNV-associated signatures (SBSs) per cancer cell fraction by clinical risk group in the validation cohort (n = 36). Thick lines correspond to average exposure for all samples. Thin lines correspond to per-sample exposure (see also Figure S4).

To provide a comprehensive overview of the mutational processes involved in neuroblastoma, we first analyzed the SNV trinucleotide context (Figures 2A, 2E, and S2A–S2C; Tables S1, S2, and S3). We recurrently identified four SNV-based signatures<sup>6</sup> (SBSs) in this cohort, indicating that associated mutational processes are active in neuroblastoma. In line with previous reports,<sup>35</sup> SBS3 and SBS18, signatures experimentally linked to defective homologous recombination repair (HRR) and reactive

oxygen species (ROS),<sup>36,37</sup> respectively, were recurrently identified in this cohort (Figures 2A, S2A, and S2B). Whereas SBS18 was predominantly observed in high-risk MYCN-amplified patients (p < 1.9 × 10<sup>−4</sup>), SBS3 was observed to a significantly lower degree in this risk group (p < 5.4 × 10<sup>−5</sup>), displaying a negative correlation with signature SBS18, and was more prevalent in high-risk non-MYCN-amplified patients. Similarly, SBS40 and SBS5, clock-like signatures associated with patient



age,<sup>10,38</sup> were highly anti-correlated with SBS18 in this cohort. Clock-like signature SBS5 was inversely associated with clinical risk and most prevalent in non-high-risk patients ( $p < 3.3 \times 10^{-4}$ ). Thus, SNV-based mutational signatures attributed to defective HRR and ROS are prevalent and differentially active across neuroblastoma risk groups and serve as the basis for co-occurrence analyses to link mutational etiologies across variant types.

### Mutational processes change during high-risk neuroblastoma evolution

We and others recently described that some complex rearrangements occur at different times throughout neuroblastoma development.<sup>39,40</sup> To link mutational processes driving SNVs to such complex rearrangement types as well as other variant types, we determined the timing of mutational processes during neuroblastoma evolution. To do so, we performed ultra-deep sequencing at an average coverage of 200× in an independent cohort of 36 neuroblastomas also derived from patients at the time of diagnosis (Figure 1). We validated 69.6% of the signatures extracted in the discovery cohort (SNV-, indel-, CNA-, and SV-based;  $n = 114$ ). The frequency of the mutational signatures across risk groups was significantly linearly correlated between both cohorts ( $r^2 = 0.61$ ,  $p = 4.4 \times 10^{-5}$ ; Figures S3E and S3F; Tables S2 and S3). Thus, the signatures identified in this study are robustly and reproducibly detectable in independent neuroblastoma cohorts. Next, we reconstructed the evolutionary trajectories of mutational processes based on the density of mutation frequencies and changes in mutational signature activities (Figures 2F and S5; Table S4). Through the reconstruction of changes in signature activity across different cancer cell fractions (CCFs), this approach enables the inference of mutational evolution in tumors.<sup>41</sup> Signatures SBS18 and SBS3 were predominantly found at higher CCFs in high-risk patients ( $p < 3.14 \times 10^{-2}$  and  $p < 2.7 \times 10^{-4}$ , respectively), whereas SBS5 activity was dominant in their subclones. This indicates that mutations resulting from ROS or HRR deficiency are more frequent in early high-risk neuroblastoma development, while cell-intrinsic mutational processes accumulate over time and occur later in neuroblastoma evolution. Non-high-risk patients, on the other hand, did not show any changes in mutational trajectories, exhibiting a predominance of SBS5-associated mutations across CCFs ( $p < 4.8 \times 10^{-3}$ ), i.e., a stable frequency of mutagenic processes throughout tumor evolution. This indicates that mutational processes involved in neuroblastoma initiation and progression are distinct and differ between risk groups. While non-high-risk neuroblastomas are characterized by continuous exposure to cell-intrinsic mutational processes, mutational processes in high-risk neuroblastomas seem to switch between early and late stages of tumor evolution.

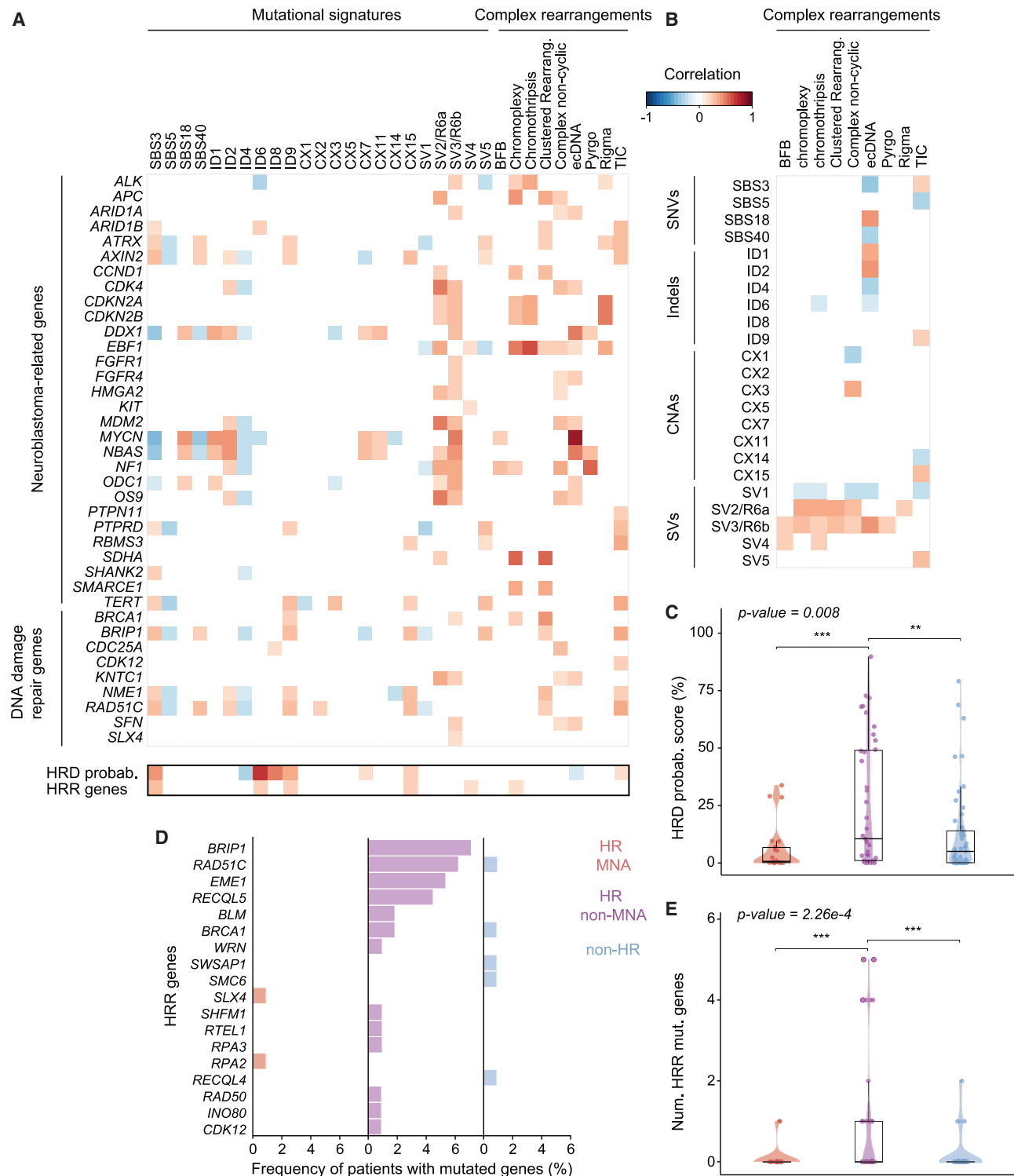
### Indel-based signatures confirm differences in mutational processes active in neuroblastoma risk groups

Co-occurrence analyses of mutational signatures from independent variant types that are experimentally linked to molecular etiologies such as SNVs and indels are crucial to determining the mutational processes active in a tumor.<sup>3,5,6</sup> Based on the insertion and deletion lengths and genomic context, including repet-

itive sequences and microhomology, we identified six indel-based signatures<sup>6</sup> (IDs) in neuroblastoma genomes (Figures 2B, 2E, and S2D–S2F; Tables S1 and S3). ID1 and ID2, both characterized by 1 bp insertions and deletions at long thymine homopolymers attributed to replication slippage and found in most cancer entities,<sup>6</sup> were also recurrently identified in neuroblastoma genomes. These signatures were significantly more frequent in patients with higher clinical risk, particularly in high-risk *MYCN*-amplified patients ( $p < 2 \times 10^{-4}$ ). ID1 and ID2 correlated positively with SNV signature SBS18, which indicates co-occurrence of DNA damage by ROS and replication slippage in *MYCN*-amplified neuroblastomas. ID6, characterized by larger deletions (>5 bp) with larger microhomology, was also recurrently observed in our cohort. In agreement with their common etiology associated with defective HRR, ID6 followed the same prevalence distribution as SBS3, corroborating the predominance of defective HRR in high-risk non-*MYCN*-amplified neuroblastomas ( $p < 1.8 \times 10^{-2}$ ). ID4, a signature enriched for deletions (>1 bp) at repeats and microhomology, recently associated experimentally with TOP1 mutational activity in cancer and healthy cells,<sup>42</sup> on the other hand was significantly more prevalent in non-high-risk patients ( $p < 1.68 \times 10^{-3}$ ). ID4 displayed an association with SBS5 and high anti-correlation with all the other indel signatures, notably with the ones related to SBS3 (ID6, ID8, ID9) or SBS18 (ID1, ID2). This is in line with previous works in other cancer entities describing ID4 as mutually exclusive with signatures ID1 and ID2<sup>6</sup> and points to functional impairment of topoisomerase 1 as a source of mutagenesis in non-high-risk neuroblastoma. The co-occurrence of SNV-based and indel-based signatures associated with the same mutational processes further supports that distinct mutational processes contribute to mutagenesis in different neuroblastoma risk groups.

### High-risk *MYCN*-amplified neuroblastoma genomes are defined by replication-stress-related CNA patterns

Whereas the mutational processes involved in SNV and indel generation have been extensively explored, less is known about the origin of CNAs and SVs. Recent reports suggest that CNA patterns can be grouped into signatures, offering more insights into molecular processes involved in their generation.<sup>7,8</sup> As neuroblastomas from different risk groups are known to harbor different CNA patterns,<sup>19,22,23,26–29</sup> we reasoned that co-occurrence analysis of CNA signatures with SNV and indel signatures may uncover processes linked to CNA genesis in neuroblastoma. Thus, we evaluated the presence of recently established CNA-associated signatures<sup>8</sup> (CXs) in our cohort (Figures 2C, 2E, S2G, and S2H; Tables S1 and S3). CX1 was the most active CNA signature found in this cohort (Figure 2C). CX1 together with CX14, both associated with chromosomal arm changes potentially caused by chromosome mis-segregation, were most prevalent in non-high-risk patients ( $p < 9.1 \times 10^{-4}$ ), in line with previous works associating whole-chromosome alterations with lower clinical risk.<sup>43</sup> CX3, linked to defective HRR, was most frequent in high-risk non-*MYCN*-amplified patients ( $p < 2.1 \times 10^{-2}$ ), further corroborating the role of deficient HRR as a risk-group-defining mutational process in neuroblastoma. CX11, a signature attributed to replication-stress-mediated focal



**Figure 3. Co-occurrence of mutational signatures, complex rearrangements, and cancer-related gene alterations in neuroblastoma**

(A) Heatmap showing the correlations between mutated neuroblastoma driver genes and DNA-damage-repair genes and all the mutational signatures and complex rearrangement types identified in our cohort ( $n = 114$ ). Below are rows showing the correlation between HRD probability score, mutations in HRR genes, and the mutational signature exposures and complex rearrangements identified in our cohort. In both heatmaps, colors display the Spearman correlation coefficient. Only significant correlations are included ( $p < 0.05$ , FDR correction).

(legend continued on next page)

amplifications, and SBS18 significantly co-occurred in high-risk neuroblastomas, raising the possibility that ROS-induced replication stress may contribute to the generation of focal amplifications in these tumors. CX5, CX7, and CX11 were most frequent in high-risk *MYCN*-amplified patients ( $p < 4.4 \times 10^{-3}$ ). Notably, CX5 and CX11, both attributed to replication stress, were significantly associated with the presence of indel-based signatures ID1 and ID2, which result from replication slippage. This further highlights replication-stress-associated mutational processes as contributors to mutagenesis in high-risk neuroblastomas. Thus, co-occurrence analyses of CNA, SNV, and indel signatures revealed that whereas focal amplifications typically found in *MYCN*-driven neuroblastomas are linked to signs of replication stress and slippage, CNAs in high-risk non-*MYCN*-amplified neuroblastomas are associated with defective HRR, and chromosome mis-segregation patterns are a general feature of all neuroblastomas across risk groups.

### De novo structural variant signature analysis identifies neuroblastoma-specific patterns

We and others have shown that complex SVs are prevalent in neuroblastomas.<sup>30,33</sup> The mechanisms involved in their generation, however, are currently largely unknown. Co-occurrence analyses with mutational signatures from other variant types facilitate the investigation of etiologies associated with distinct SV patterns. To this end, we performed *de novo* discovery of SV signatures<sup>9</sup> (SVs) based on the type, size, and clustering of SVs (Figures 2D, 2E, and S3A–S3D; Tables S1, S2, and S3). Comparing the SV signatures we extracted to SV signatures identified in other tumor entities,<sup>9</sup> we determined that three signatures were specific to this cohort (SV1, SV4, and SV5), whereas two SV-based signatures (SV2 and SV3, corresponding to R6a and R6b) had already been identified in other cancer types. SV1, defined by simple deletions smaller than 1 kb, was predominantly detected in non-high-risk patients ( $p < 3.3 \times 10^{-2}$ ) and showed a negative correlation with all the other SV-based signatures. It was positively correlated with SBS5, a clock-like signature, suggesting it may have similar origins. SV2/R6a and SV3/R6b, characterized by clustered intrachromosomal rearrangements around 1–10 Mb and larger than 10 Mb, respectively, were linked to the presence of focal oncogene amplifications (Figures 3A and 3B). This is in line with results from previous reports in other tumor entities<sup>9</sup> and indicates that these signatures could represent footprints of the molecular processes involved in the generation of high-level oncogene amplification. Consequently, these signatures were highly prevalent in high-risk *MYCN*-amplified neuroblastomas ( $p < 2.5 \times 10^{-2}$ ), which contain focal amplifications. SV3/R6b was also correlated with the exposure of replication

slippage signature (Figure 2E), suggesting a role of this mutational process in the generation of complex amplicons in neuroblastoma, as previously proposed in other tumor entities.<sup>44–46</sup> Notably, the overall frequency of clustered SV-based signatures (SV2/R6a, SV3/R6b, and SV4) was correlated with clinical risk. Thus, distinct complex SV patterns prevalent in high-risk neuroblastomas can be categorized using SV-based mutational signatures and significantly co-occur with signatures based on other variant types, raising the possibility that similar mutagenic processes contribute to their generation.

### Complex structural rearrangement topography is linked to distinct mutational processes across neuroblastoma risk groups

Recent reports have reclassified complex rearrangements topographically, considering both CNA- and SV-associated features.<sup>14,17,47</sup> Even though a subset of these rearrangement patterns has been observed in some neuroblastomas, e.g., ecDNA and chromothripsis,<sup>30,48</sup> their prevalence across neuroblastoma risk groups and their co-occurrence with mutational signatures has not yet been determined. To examine the topography of complex rearrangements in neuroblastoma, we used three state-of-the-art complementary algorithms, enabling the identification and reconstruction of nine complex variant classes: (1) chromothripsis; (2) BFBs; (3) ecDNA; (4) CnCs; (5) chromoplexy; (6) TICs; (7) rigma, a cluster of simple deletions; (8) pyrgo, a cluster of tandem duplications (Figure 4); and (9) unclassified regions with high SV density, termed clustered rearrangements. The fraction of SVs assigned to complex variant patterns corresponded to 69.41% and differed between risk groups ( $p < 2.39 \times 10^{-9}$ ; Figures 4A and S5D), suggesting that structural variation in high-risk neuroblastomas is predominantly complex. In line with previous observations,<sup>30</sup> ecDNA was the most frequently observed SV pattern in neuroblastomas, detectable in 31.65% of the patients (Figures 4B and 4C; Table S6). The predominance of ecDNA defined high-risk *MYCN*-amplified patients (Figures 4D, S5A, and S5E–S5N;  $p < 1.3 \times 10^{-7}$ ), in line with the fact that *MYCN* amplifications most often occur in the form of ecDNA.<sup>30,49</sup> *MYCN* amplifications detected in whole-genome sequencing (WGS) were confirmed cytogenetically using fluorescence *in situ* hybridization (FISH) and classified as either ecDNA or homogeneously staining regions (Table S5). A subset of ecDNA-harboring neuroblastomas also contained instances of chromothripsis and/or BFBs, in line with previous reports showing that these processes can contribute to ecDNA generation and evolution.<sup>32,33</sup> Most ecDNAs, however, did not co-occur with any other complex rearrangement (Figure 4B), suggesting that they are a result of other processes. Simple

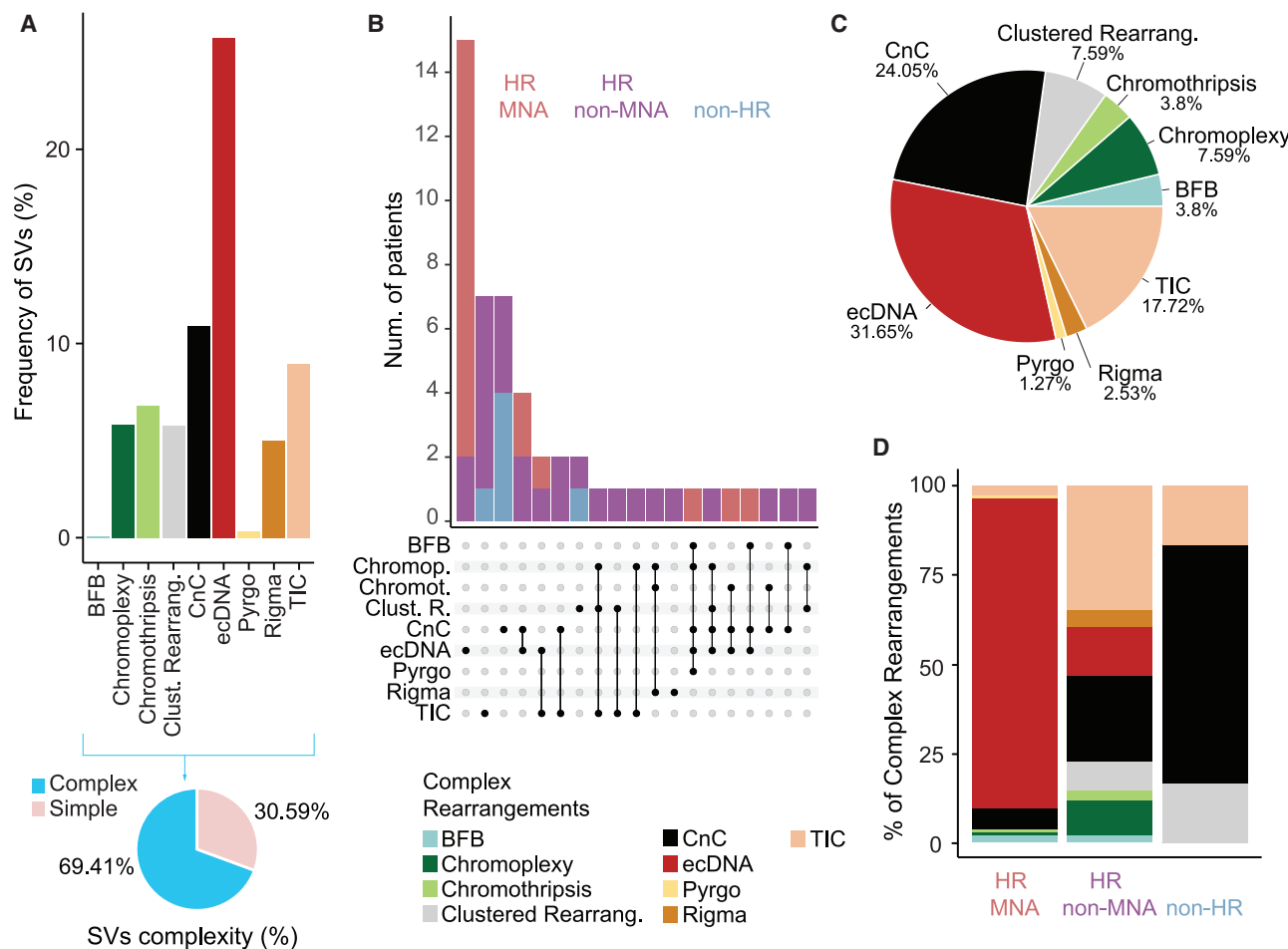
(B) Heatmap depicting the correlations between the signatures associated with different variant types (SNV, indel, SV, and CNA) and the nine complex rearrangement classes identified in our cohort ( $n = 114$ ).

(C) Box plot comparing the distribution of HRD probability scores across neuroblastoma risk groups (HR MNA, HR non-MNA, and non-HR). Each dot represents a patient. To assess whether there are differences between risk groups, we used the non-parametric Kruskal-Wallis test ( $p$  value in the upper-left corner).

(D) Frequency of patients with mutated HRR genes across the three neuroblastoma risk groups.

(E) Box plot comparing the distribution of HRR mutated genes across neuroblastoma risk groups (HR MNA, HR non-MNA, and non-HR). Each dot represents a patient. To assess whether there are differences between risk groups, we used the non-parametric Kruskal-Wallis test ( $p$  value in the upper-left corner).

Pairwise comparisons were done using the non-parametric Wilcoxon rank-sum test. Significance: \* $p < 0.1$ , \*\* $p < 0.05$ , \*\*\* $p < 0.01$ . All analyses were performed in the discovery cohort ( $n = 114$ ).



**Figure 4. Co-occurrence and distribution of complex rearrangements in neuroblastoma**

(A) Top: frequency of SVs involved in each complex rearrangement type, in percentage. Bottom: pie chart showing the percentage of SVs involved in complex and simple events in our cohort.

(B) Upset plot depicting the co-occurrence of the different types of complex SV patterns within patients. The number of patients with each combination of rearrangements is shown in the top histogram (colors display the risk group for each patient).

(C) Pie chart showing the frequency in percentage of each complex rearrangement type in the whole neuroblastoma cohort.

(D) Relative frequency in percentage of the nine different complex rearrangement types identified in our cohort by clinical risk group (HR MNA, HR non-MNA, and non-HR).

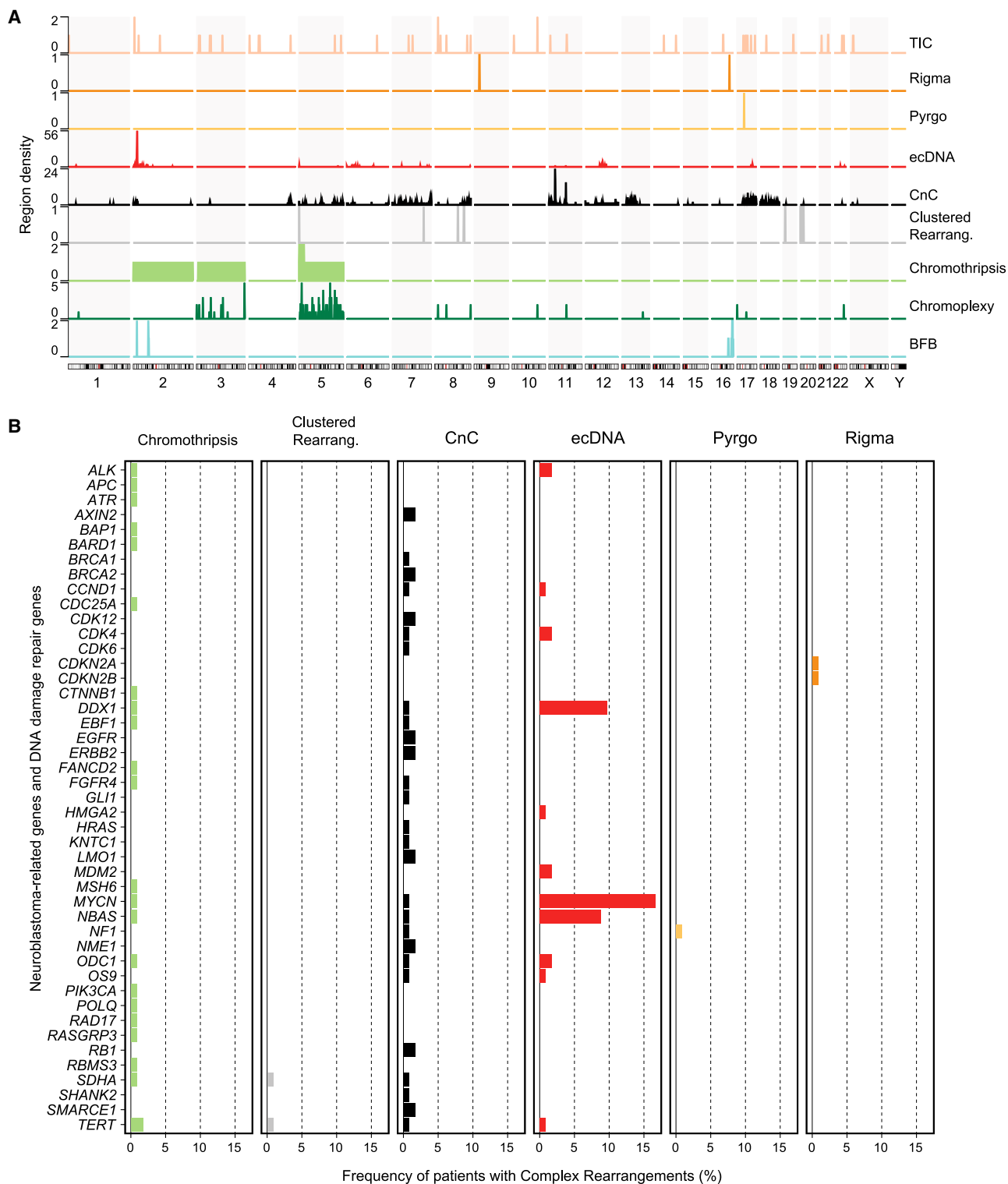
All analyses were performed in the discovery cohort ( $n = 114$ ). See also Figure S5.

DNA circularization due to replication slippage can lead to ecDNA generation,<sup>44</sup> which would be consistent with our co-occurrence analyses results (Figure 3B). High-risk non-MYC-amplified patients, on the other hand, exhibited a high variety of mostly linear chromosomal complex rearrangements (Figures 4B, 4D, S9A, and S9E–S9N), including unclassified clustered rearrangements. Through manual reconstruction of these unclassified events, we identified two complex SV patterns of local  $n$ -jumps<sup>14</sup> named duplication-inverted-triplication-duplication (Dup-Trp-Dup) and duplications linked by inverted segments (Dup-invDup) (Figure S5O), which had not been previously described in neuroblastoma.

Exploring the potential etiologies of complex SV patterns, we found that signatures characterized by clustered SVs (SV2/R6a, SV3/R6b, and SV4) were correlated with the presence of

complex amplicon topography such as chromothripsis, BFBs, and ecDNA (Figure 3B). Furthermore, SV5, characterized by simple deletions and duplications along with non-clustered translocations, was linked to TICs, based on its SV pattern composition and the significant co-occurrence with this complex rearrangement pattern (Figures 3B and S3C). SV5 was correlated with signatures of HRR deficiency, suggesting its role in the generation of TICs.

Complex SVs did not occur uniformly throughout the genome (Figures 5 and S5C). Known neuroblastoma-related genes such as *MYCN*, *TERT*, *ODC1*, *CDK4*, and *MDM2* were recurrently affected by different classes of complex rearrangements including ecDNA, chromothripsis, and CnCs. Thus, topographically distinct complex variant patterns are more common in neuroblastoma than previously anticipated and recurrently occur at



**Figure 5. Genomic distribution of topographically defined complex rearrangement patterns in neuroblastoma**

(A) Density plot showing the number of regions affected by complex rearrangements per chromosome across the whole human genome. Each row/color corresponds to a different type of rearrangement.

(B) Frequency of patients with complex rearrangements affecting neuroblastoma driver genes or DNA-damage-repair genes. Each column and color corresponds to a different type of rearrangement.

All analyses were performed in the discovery cohort (n = 114). See also [Figure S5](#).



sites of cancer-related loci, suggesting their functional oncogenic importance.

### Risk-group-defining mutational scenarios emerge through integration of mutational signatures and complex rearrangement patterns in neuroblastoma

We hypothesized that the co-occurrence of complex variant patterns and mutational signatures from independent variant types with similar molecular etiologies may be linked to the clinical heterogeneity observed in patients suffering from neuroblastoma. Unsupervised clustering based only on the frequency of the different types of mutational signatures (SNV-, indel-, CNA-, and SV-based) and complex rearrangement patterns grouped tumors into three different clusters, which we termed mutational scenarios (Figures 6A–6C and Table S7). In line with our hypothesis, mutational scenarios exhibited concordance of more than 80% with neuroblastoma risk group classification. Mutational scenario #1 was enriched in high-risk *MYCN*-amplified neuroblastomas and was defined by footprints of DNA damage by ROS (SBS18), replication slippage (ID1/ID2), replication stress (CX5/CX11), and SV-based signature SV3/R6b, association with clustered SVs, and focal oncogene amplification in the form of ecDNA. This nominates replication-associated mutagenesis as the underlying mechanism of variants occurring in this risk group, e.g., ecDNA (Figures 6A–6C). The second mutational scenario (scenario #2) was characterized by signs of impaired HRR (SBS3/ID6/CX3) and single base deletions (ID9). It was associated with a higher variety of linear chromosomal complex rearrangements such as CnCs, and TICs (SV5). This mutational scenario was enriched in high-risk non-*MYCN*-amplified neuroblastomas, suggesting that the high number of CNAs and SVs found in this risk group might be caused by the portrayed HRR deficiency, which is in line with recent reports in other cancer types.<sup>50</sup> Supporting these findings, non-*MYCN*-amplified patients showed an enrichment in mutations affecting genes from the HRR pathway such as *BRIP1* and *RAD51C* (Figures 3D and 3E;  $p = 2.26 \times 10^{-4}$ ). Interestingly, the presence of ecDNA was nearly mutually exclusive, with defective homologous recombination-associated signatures defining scenario #2 (Figures 3A and 3B). Consistently, ecDNA-harboring, *MYCN*-amplified neuroblastomas exhibited the lowest HRDetect probability scores, an alternative measure of HRR activity,<sup>51</sup> across our cohort, with a median score of 1% and no patient with a score >70% (Figures 3A, 3C, and 3E; Table S8). Scenario #3 was linked to high prevalence of clock-like mutations (SBS5), topoisomerase-associated mutational activity (ID4), small simple deletions (SV1), and signs of chromosome mis-segregation (CX1/CX14). This mutational scenario was enriched in non-high-risk neuroblastomas. Concordance between mutational scenarios and clinical risk groups was also reflected by overall survival rates (e.g., non-high-risk and scenario #3 patients; Figures 6D and 6E;  $p = 8.7 \times 10^{-4}$  and  $p < 1 \times 10^{-4}$ , respectively, by log-rank test). Thus, clinical neuroblastoma heterogeneity is significantly associated with risk-group-defining mutational footprints, which not only offers new insights into the etiology of disease group-defining genomic variants but also raises the possibility that differences in mutational processes during malignant transformation and tumor progression may contribute to inter-tumor phenotypic and clinical differences.

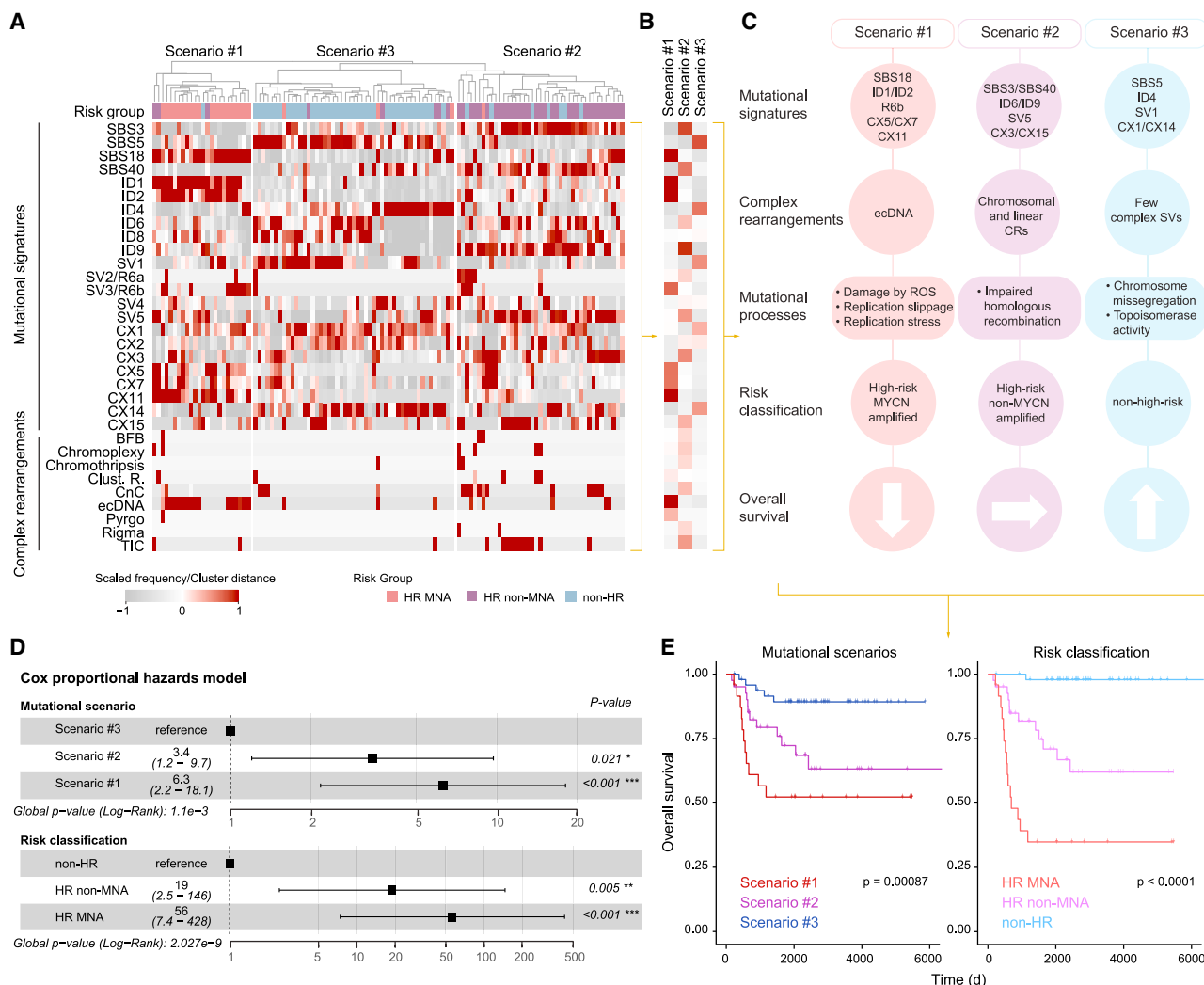
### DISCUSSION

This study analyzed neuroblastoma genomes from clinically heterogeneous patients and classified the mutational processes and genomic rearrangement patterns based on all genomic variant classes. This led to the identification of three distinct mutational scenarios defined by the co-occurrence of mutational signatures and complex rearrangement patterns associated with similar mutational processes independent of variant classes. Scenario #1 was driven by DNA damage caused by ROS, replication slippage, and stress and was enriched in high-risk *MYCN*-amplified, ecDNA-harboring patients with low overall survival. Scenario #2 was driven by HRR-associated signatures and characterized by linear, chromosomal, complex rearrangements, frequently observed in high-risk non-*MYCN*-amplified patients with low survival. Scenario #3 was characterized by footprints of chromosome mis-segregation, topoisomerase activity, and high presence of clock-like signature SBS5, and was enriched in non-high-risk patients exhibiting good prognosis. Our findings demonstrate that clinical neuroblastoma heterogeneity is associated with differences in mutational footprints across genomic variant classes, offering a new perspective on the mutational processes contributing to neuroblastoma genesis and evolution.

Even though the identified mutational scenarios are linked to specific mutational processes,<sup>36,37,52–56</sup> what causes these processes to be active in neuroblastoma currently remains unknown. Some molecular features in neuroblastomas can activate mutational processes. For example, *MYCN* can induce ROS, replication stress, and fork stalling.<sup>57,58</sup> Thus, high *MYCN* expression in neuroblastomas harboring *MYCN* amplifications may explain increased footprints of ROS, replication slippage, and stress observed in mutational scenario #1, which was enriched in *MYCN*-amplified neuroblastomas. Recent reports suggest that the presence of ecDNA itself may cause high replication stress in cancer cells,<sup>59</sup> indicating that some of the complex variants observed may themselves fuel mutational activity.

Deficiencies in HRR can cause the generation of simple and complex SVs,<sup>60</sup> such as BFBs and chromothripsis.<sup>61</sup> For example, mutations in *BRCA1/2* resulting in HRR deficiency are associated with high prevalence of certain SVs, termed *BRCA*ness.<sup>62</sup> Although *BRCA1/2* mutations are not frequent in neuroblastoma,<sup>63</sup> we identified mutational signatures related to HRR deficiency in high-risk non-*MYCN*-amplified patients. These tumors were also marked by higher homologous recombination deficiency (HRD) probability scores (HRDetect) and higher prevalence of mutated HRR genes, including the ones in the PROfound clinical trial.<sup>64</sup> Thus, alterations other than *BRCA1/2* mutations may cause the apparent HRR deficiency in neuroblastomas. In line with our work, recent pediatric pan-cancer studies also observed signatures associated with HRR deficiency in neuroblastoma.<sup>10</sup> The high prevalence of these footprints in high-risk compared to non-high-risk patients indicates that HRR deficiency may not only represent a mere feature of these tumors but could also contribute to their aggressive clinical behavior, similar to what was observed in HRR-deficient pancreatic cancers and gliomas.<sup>65</sup> This may be especially relevant for





**Figure 6. Characterization and definition of mutational scenarios linked to clinical heterogeneity**

(A) Unsupervised clustering analysis (hkmeans method;  $k = 3$ ) obtaining three mutational scenarios from the scaled mutational signatures exposure and complex rearrangements. All mutational signatures from different variant types and complex rearrangements detected in our cohort are included in the analysis. Color scale indicates frequency.

(B) Summary of the defining features for each of the three mutational scenarios/clusters. Color gradation corresponds to clustering distance.

(C) Summary of the three mutational scenarios described in our neuroblastoma cohort, including different characteristic features such as mutational signatures, complex rearrangements, mutational processes, and risk classification correspondence associated with each of them. CRs, complex rearrangements.

(D) Univariate Cox proportional hazards model. Forest plot shows the proportional risk of the three mutational scenarios and the three neuroblastoma risk groups.

(E) Kaplan-Meier survival curves showing the clinical impact of the three mutational scenarios and the three neuroblastoma risk groups. ( $p = 0.00087$  and  $p < 0.0001$ , respectively by log-rank test). Colors on Kaplan-Meier plot display each condition.

All analyses were performed in the discovery cohort ( $n = 114$ ).

future clinical trial designs, as HRR deficiency is associated with altered response to certain chemotherapeutic agents<sup>66</sup> as well as increased sensitivity to pharmacological PARP (poly(ADP-ribose) polymerase) trapping.<sup>67</sup> Thus, the identification of mutational patterns associated with HRR deficiency in high-risk neuroblastomas may have important implications for neuroblastoma risk and therapy stratification.

Changes in mutational process activities can inform the causes of tumor evolution.<sup>68,69</sup> Our study revealed that in high-risk neuroblastomas ROS and HRR deficiency predominantly

contribute to mutagenesis in early tumor development, while replication-associated mutations are accumulated throughout evolution. This indicates that the difference between high-risk and non-high-risk neuroblastomas may already be determined by the mutagenic processes involved in their early development and/or the initial transformation of their non-cancerous cell of origin.

We also provide a complete catalog of recently described complex variant patterns<sup>17</sup> in neuroblastoma, including chromothripsis, BFBs, ecDNA, CnCs, chromoplexy, TICs, rigma,

pyrgo, Dup-Trp-Dup, and Dup-invDup, and reveal their distinct co-occurrence patterns with mutational signatures, indicating differences in the mutagenic processes active during their generation. Moreover, we linked *de novo* signature SV5 with TICs, suggesting that recently topographically defined variant types can be reflected by novel SV signatures.

Extrachromosomal DNA is one of the most prevalent genomic alterations in cancer.<sup>70</sup> We and others recently demonstrated that it not only serves as a vehicle for oncogene amplification<sup>30</sup> but is also the most recurrent site of complex structural rearrangements.<sup>16,30,31</sup> Indeed, footprints of BFBs, chromothripsis, and kataegis on ecDNA in other cancer entities nominated these processes as mechanisms of ecDNA generation.<sup>32,33</sup> In contrast, we only observed signs of BFBs and chromothripsis on ecDNA in a subset of cases and did not observe footprints indicative of APOBEC3-driven kataegis to the same extent as observed elsewhere.<sup>71</sup> Based on our most recent observations in single neuroblastoma cells<sup>72</sup> and consistent with our mutational signature co-occurrence analyses, we propose that other processes such as replication slippage may create structurally simpler ecDNA in neuroblastomas in which BFBs and chromothripsis were not detected. Such simple ecDNAs may then evolve and gain in complexity through further mutagenesis. For example, recently described ecDNA reintegration and circular recombination may contribute to ecDNA evolution and result in similarly complex structures as those created through BFBs or chromothripsis.<sup>30,33</sup> The relative contributions of these different mechanisms in the structural shaping of ecDNA are still largely unresolved. Combined detection of the mutational scenarios defined in this study with longitudinal single-cell sequencing may facilitate the investigation of these open questions.

In summary, our study provides a comprehensive classification of active mutational processes in neuroblastoma, offering new insights into the origin of genomic alterations involved in neuroblastoma genesis and progression. The three mutational scenarios presented here not only refine our understanding of neuroblastoma's clinical heterogeneity but may also improve our understanding of how mutational processes contribute to the generation of different variant classes in cancer in general.

### Limitations of the study

While the unsupervised clustering of neuroblastoma patients in the three mutational scenarios was in high concordance (>80%) with the clinical risk group classification, we encountered some patients who were considered as outliers. These patients, although clinically classified in a specific risk group, showed genomic features associated with other risk groups. Owing to the limitation in sample size, we were not able to draw significant conclusions that could explain these discrepancies between the genomic and clinical features. We also did not include transcriptomic data in this study; such data could be helpful in future projects for the understanding of the mentioned outlier divergence and the contribution of specific gene expression to the mutational scenarios that we describe.

### STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- RESOURCE AVAILABILITY
  - Lead contact
  - Materials availability
  - Data and code availability
- EXPERIMENTAL MODEL AND STUDY PARTICIPANT DETAILS
- METHOD DETAILS
  - Sequencing data
  - Variant calling
  - Identification of variants affecting genes and chromosome arms
  - Mutational signatures analyses
  - Subclonal signatures analysis
  - Homologous recombination deficiency analysis
  - Complex rearrangements calling and classification
  - Functional analysis of complex rearrangements
  - Kataegis analysis
  - Mutational scenarios analysis
- QUANTIFICATION AND STATISTICAL ANALYSIS

### SUPPLEMENTAL INFORMATION

Supplemental information can be found online at <https://doi.org/10.1016/j.xgen.2023.100402>.

### ACKNOWLEDGMENTS

We thank F. Dubois and O. Shoshani for critical discussions. A.G.H. is a participant in the Berlin Institute of Health-Charité Clinical Scientist Program funded by the Charité – Universitätsmedizin Berlin and the Berlin Institute of Health. A.G.H. is supported by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation, 398299703) and the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation program (grant agreement no. 949172). A.G.H. is supported by the Deutsche Krebshilfe (German Cancer Aid) Mildred Scheel Professorship program – 70114107. This project received funding from the NIH/CRUK (398299703, the eDynamic Cancer Grand Challenge). J.H.S., K.H., and M.P.-F. are supported by the German Cancer Consortium (DKTK). E.R.-F. is supported by the Alexander von Humboldt Foundation. This project was supported by the Berlin Institute of Health at Charité – Universitätsmedizin Berlin within the collaborative research projects TERMINATE-NB (CRG04) and PeDiOn. We thank the patients and their parents for granting access to the tumor specimens and clinical information that were analyzed in this study. We thank B. Hero, H. Düren, and N. Hemstedt of the Neuroblastoma Biobank and Neuroblastoma Trial Registry (University Children's Hospital Cologne) of the German Society of Pediatric Oncology and Hematology for providing samples and clinical data. Computation was performed on the HPC for Research cluster of the Berlin Institute of Health and the MareNostrum of the Barcelona Supercomputing Center.

### AUTHOR CONTRIBUTIONS

All authors contributed to the study design and the collection and interpretation of the data. F.H., A.S., and J.H.S. collected and prepared patient samples. E.R.-F. and K.H. analyzed Illumina WGS data from the discovery cohort. M.P.-F. and K.H. analyzed Illumina WGS data from the validation cohort. M.B., M.P., J.T., N.T., E.B., and N.I. performed experiments and analyzed data. D.B., D.T., A.E., R.P.K., and R.F.S. contributed to study design.

E.R.-F., M.P.-F., K.H., J.H.S., and A.G.H. led the study design, performed data analysis, and wrote the manuscript, to which all authors contributed.

## DECLARATION OF INTERESTS

The authors declare no competing interests.

## INCLUSION AND DIVERSITY

We support inclusive, diverse, and equitable conduct of research.

Received: April 28, 2023

Revised: June 13, 2023

Accepted: August 11, 2023

Published: September 7, 2023

## REFERENCES

- Greenman, C., Stephens, P., Smith, R., Dalgleish, G.L., Hunter, C., Bignell, G., Davies, H., Teague, J., Butler, A., Stevens, C., et al. (2007). Patterns of somatic mutation in human cancer genomes. *Nature* 446, 153–158. <https://doi.org/10.1038/nature05610>.
- Henssen, A.G., Koche, R., Zhuang, J., Jiang, E., Reed, C., Eisenberg, A., Still, E., MacArthur, I.C., Rodríguez-Fos, E., Gonzalez, S., et al. (2017). PGBD5 promotes site-specific oncogenic mutations in human tumors. *Nat. Genet.* 49, 1005–1014. <https://doi.org/10.1038/ng.3866>.
- Jeggo, P.A., Pearl, L.H., and Carr, A.M. (2016). DNA repair, genome stability and cancer: a historical perspective. *Nat. Rev. Cancer* 16, 35–42. <https://doi.org/10.1038/nrc.2015.4>.
- Koh, G., Degasperi, A., Zou, X., Momen, S., and Nik-Zainal, S. (2021). Mutational signatures: emerging concepts, caveats and clinical applications. *Nat. Rev. Cancer* 21, 619–637. <https://doi.org/10.1038/s41568-021-00377-7>.
- Australian Pancreatic Cancer Genome Initiative, ICGC Breast Cancer Consortium, ICGC MMML-Seq Consortium, ICGC PedBrain; Alexandrov, L.B., Nik-Zainal, S., Wedge, D.C., Aparicio, S.A.J.R., Behjati, S., Biankin, A.V., et al. (2013). Signatures of mutational processes in human cancer. *Nature* 500, 415–421. <https://doi.org/10.1038/nature12477>.
- PCAWG Mutational Signatures Working Group; Consortium, P.C.A.W.G., Alexandrov, L.B., Kim, J., Haradhvala, N.J., Huang, M.N., Tian Ng, A.W., Wu, Y., Boot, A., Covington, K.R., et al. (2020). The repertoire of mutational signatures in human cancer. *Nature* 578, 94–101. <https://doi.org/10.1038/s41586-020-1943-3>.
- Macintyre, G., Goranova, T.E., De Silva, D., Ennis, D., Piskorz, A.M., Eldridge, M., Sie, D., Lewsley, L.-A., Hanif, A., Wilson, C., et al. (2018). Copy number signatures and mutational processes in ovarian carcinoma. *Nat. Genet.* 50, 1262–1270. <https://doi.org/10.1038/s41588-018-0179-8>.
- Drews, R.M., Hernando, B., Tarabichi, M., Haase, K., Lesluyes, T., Smith, P.S., Morrill Gavarró, L., Couturier, D.-L., Liu, L., Schneider, M., et al. (2022). A pan-cancer compendium of chromosomal instability. *Nature* 606, 976–983. <https://doi.org/10.1038/s41586-022-04789-9>.
- Degasperi, A., Amarante, T.D., Czarnecki, J., Shooter, S., Zou, X., Glodzik, D., Morganella, S., Nanda, A.S., Badja, C., Koh, G., et al. (2020). A practical framework and online tool for mutational signature analyses show inter-tissue variation and driver dependencies. *Nat. Can. (Ott.)* 1, 249–263. <https://doi.org/10.1038/s43018-020-0027-5>.
- Thatikonda, V., Islam, S.M.A., Autry, R.J., Jones, B.C., Gröbner, S.N., Warsaw, G., Hutter, B., Huebschmann, D., Fröhling, S., Kool, M., et al. (2023). Comprehensive analysis of mutational signatures reveals distinct patterns and molecular processes across 27 pediatric cancers. *Nat. Can. (Ott.)*. <https://doi.org/10.1038/s43018-022-00509-4>.
- Turner, K.M., Deshpande, V., Beyter, D., Koga, T., Rusert, J., Lee, C., Li, B., Arden, K., Ren, B., Nathanson, D.A., et al. (2017). Extrachromosomal oncogene amplification drives tumour evolution and genetic heterogeneity. *Nature* 543, 122–125. <https://doi.org/10.1038/nature21356>.
- Stephens, P.J., Greenman, C.D., Fu, B., Yang, F., Bignell, G.R., Mudie, L.J., Pleasance, E.D., Lau, K.W., Beare, D., Stebbings, L.A., et al. (2011). Massive Genomic Rearrangement Acquired in a Single Catastrophic Event during Cancer Development. *Cell* 144, 27–40. <https://doi.org/10.1016/j.cell.2010.11.055>.
- Baca, S.C., Prandi, D., Lawrence, M.S., Mosquera, J.M., Romanel, A., Drier, Y., Park, K., Kitabayashi, N., MacDonald, T.Y., Ghandi, M., et al. (2013). Punctuated evolution of prostate cancer genomes. *Cell* 153, 666–677. <https://doi.org/10.1016/j.cell.2013.03.021>.
- Li, Y., Roberts, N.D., Wala, J.A., Shapira, O., Schumacher, S.E., Kumar, K., Khurana, E., Waszak, S., Korbel, J.O., Haber, J.E., et al. (2020). Patterns of somatic structural variation in human cancer genomes. *Nature* 578, 112–121. <https://doi.org/10.1038/s41586-019-1913-9>.
- Zakov, S., Kinsella, M., and Bafna, V. (2013). An algorithmic approach for breakage-fusion-bridge detection in tumor genomes. *Proc. Natl. Acad. Sci. USA* 110, 5546–5551. <https://doi.org/10.1073/pnas.1220977110>.
- Kim, H., Nguyen, N.-P., Turner, K., Wu, S., Gujar, A.D., Luebeck, J., Liu, J., Deshpande, V., Rajkumar, U., Namburi, S., et al. (2020). Extrachromosomal DNA is associated with oncogene amplification and poor outcome across multiple cancers. *Nat. Genet.* 52, 891–897. <https://doi.org/10.1038/s41588-020-0678-2>.
- Hadi, K., Yao, X., Behr, J.M., Deshpande, A., Xanthopoulos, C., Tian, H., Kudman, S., Rosiene, J., Darmofal, M., DeRose, J., et al. (2020). Distinct Classes of Complex Structural Variation Uncovered across Thousands of Cancer Genome Graphs. *Cell* 183, 197–210. <https://doi.org/10.1016/j.cell.2020.08.006>.
- Bown, N. (2001). Neuroblastoma tumour genetics: clinical and biological aspects. *J. Clin. Pathol.* 54, 897–910. <https://doi.org/10.1136/jcp.54.12.897>.
- Brodeur, G.M. (2003). Neuroblastoma: biological insights into a clinical enigma. *Nat. Rev. Cancer* 3, 203–216. <https://doi.org/10.1038/nrc1014>.
- Capasso, M., and Diskin, S.J. (2010). Genetics and Genomics of Neuroblastoma. In *Cancer Genetics Cancer Treatment and Research*, B. Pasche, ed. (Springer US), pp. 65–84. [https://doi.org/10.1007/978-1-4419-6033-7\\_4](https://doi.org/10.1007/978-1-4419-6033-7_4).
- Cohn, S.L., Pearson, A.D.J., London, W.B., Monclair, T., Ambros, P.F., Brodeur, G.M., Faldut, A., Hero, B., Iehara, T., Machin, D., et al. (2009). The International Neuroblastoma Risk Group (INRG) classification system: an INRG Task Force report. *J. Clin. Oncol.* 27, 289–297. <https://doi.org/10.1200/JCO.2008.16.6785>.
- Brodeur, G.M., Maris, J.M., Yamashiro, D.J., Hogarty, M.D., and White, P.S. (1997). Biology and Genetics of Human Neuroblastomas. *J. Pediatr. Hematol. Oncol.* 19, 93–101. <https://doi.org/10.1097/00043426-199703000-00001>.
- Maris, J.M., and Matthay, K.K. (1999). Molecular Biology of Neuroblastoma. *J. Clin. Oncol.* 17, 2264. <https://doi.org/10.1200/JCO.1999.17.7.2264>.
- Valentijn, L.J., Koster, J., Zwiijnenburg, D.A., Hasselt, N.E., van Sluis, P., Volckmann, R., van Noesel, M.M., George, R.E., Tytgat, G.A.M., Molenaar, J.J., et al. (2015). TERT rearrangements are frequent in neuroblastoma and identify aggressive tumors. *Nat. Genet.* 47, 1411–1414. <https://doi.org/10.1038/ng.3438>.
- Cheung, N.-K.V. (2012). Association of Age at Diagnosis and Genetic Mutations in Patients With Neuroblastoma. *JAMA* 307, 1062. <https://doi.org/10.1001/jama.2012.228>.
- Pugh, T.J., Morozova, O., Attiyeh, E.F., Asgharzadeh, S., Wei, J.S., Auclair, D., Carter, S.L., Cibulskis, K., Hanna, M., Kiezun, A., et al. (2013). The genetic landscape of high-risk neuroblastoma. *Nat. Genet.* 45, 279–284. <https://doi.org/10.1038/ng.2529>.
- Caron, H., van Sluis, P., de Kraker, J., Bökörker, J., Egeler, M., Laureys, G., Slater, R., Westerveld, A., Voûte, P.A., and Versteeg, R. (1996). Allelic

- Loss of Chromosome 1p as a Predictor of Unfavorable Outcome in Patients with Neuroblastoma. *N. Engl. J. Med.* 334, 225–230. <https://doi.org/10.1056/NEJM199601253340404>.
28. Plantaz, D., Mohapatra, G., Matthay, K.K., Pellarin, M., Seeger, R.C., and Feuerstein, B.G. (1997). Gain of chromosome 17 is the most frequent abnormality detected in neuroblastoma by comparative genomic hybridization. *Am. J. Pathol.* 150, 81–89.
29. Vandesompele, J., Van Roy, N., Van Gele, M., Laureys, G., Ambros, P., Heimann, P., Devalck, C., Schuurin, E., Brock, P., Otten, J., et al. (1998). Genetic heterogeneity of neuroblastoma studied by comparative genomic hybridization. *Genes Chromosomes Cancer* 23, 141–152. [https://doi.org/10.1002/\(SICI\)1098-2264\(199810\)23:2<141::AID-GCC7>3.0.CO;2-2](https://doi.org/10.1002/(SICI)1098-2264(199810)23:2<141::AID-GCC7>3.0.CO;2-2).
30. Koche, R.P., Rodriguez-Fos, E., Helmsauer, K., Burkert, M., MacArthur, I.C., Maag, J., Chamorro, R., Munoz-Perez, N., Puiggròs, M., Dorado Garcia, H., et al. (2020). Extrachromosomal circular DNA drives oncogenic genome remodeling in neuroblastoma. *Nat. Genet.* 52, 29–34. <https://doi.org/10.1038/s41588-019-0547-z>.
31. Helmsauer, K., Valieva, M.E., Ali, S., Chamorro González, R., Schöpflin, R., Röefzaad, C., Bei, Y., Dorado Garcia, H., Rodriguez-Fos, E., Puiggròs, M., et al. (2020). Enhancer hijacking determines extrachromosomal circular MYCN amplicon architecture in neuroblastoma. *Nat. Commun.* 11, 5823. <https://doi.org/10.1038/s41467-020-19452-y>.
32. Shoshani, O., Brunner, S.F., Yaeger, R., Ly, P., Nechemia-Arbely, Y., Kim, D.H., Fang, R., Castillon, G.A., Yu, M., Li, J.S.Z., et al. (2021). Chromothripsis drives the evolution of gene amplification in cancer. *Nature* 591, 137–141. <https://doi.org/10.1038/s41586-020-03064-z>.
33. Rosswog, C., Bartenhagen, C., Welte, A., Kahlert, Y., Hemstedt, N., Lorenz, W., Cartolano, M., Ackermann, S., Perner, S., Vogel, W., et al. (2021). Chromothripsis followed by circular recombination drives oncogene amplification in human cancer. *Nat. Genet.* <https://doi.org/10.1038/s41588-021-00951-7>.
34. Peifer, M., Hertwig, F., Roels, F., Dreidax, D., Gartlgruber, M., Menon, R., Krämer, A., Roncalioli, J.L., Sand, F., Heuckmann, J.M., et al. (2015). Telomerase activation by genomic rearrangements in high-risk neuroblastoma. *Nature* 526, 700–704. <https://doi.org/10.1038/nature14980>.
35. Brady, S.W., Liu, Y., Ma, X., Gout, A.M., Hagiwara, K., Zhou, X., Wang, J., Macias, M., Chen, X., Easton, J., et al. (2020). Pan-neuroblastoma analysis reveals age- and signature-associated driver alterations. *Nat. Commun.* 11, 5183. <https://doi.org/10.1038/s41467-020-18987-4>.
36. Zámbrorsky, J., Szikriszt, B., Gervai, J.Z., Pipek, O., Póti, Á., Krzystanek, M., Ribli, D., Szalai-Gindl, J.M., Csabai, I., Szallasi, Z., et al. (2017). Loss of BRCA1 or BRCA2 markedly increases the rate of base substitution mutagenesis and has distinct effects on genomic deletions. *Oncogene* 36, 746–755. <https://doi.org/10.1038/ncr.2016.243>.
37. Kucab, J.E., Zou, X., Morganello, S., Joel, M., Nanda, A.S., Nagy, E., Gomez, C., Degasper, A., Harris, R., Jackson, S.P., et al. (2019). A Compendium of Mutational Signatures of Environmental Agents. *Cell* 177, 821–836.e16. <https://doi.org/10.1016/j.cell.2019.03.001>.
38. Alexandrov, L.B., Jones, P.H., Wedge, D.C., Sale, J.E., Campbell, P.J., Nik-Zainal, S., and Stratton, M.R. (2015). Clock-like mutational processes in human somatic cells. *Nat. Genet.* 47, 1402–1407. <https://doi.org/10.1038/ng.3441>.
39. Schmelz, K., Toedling, J., Huska, M., Cwikla, M.C., Kruetzfeldt, L.-M., Proba, J., Ambros, P.F., Ambros, I.M., Boral, S., Lodrini, M., et al. (2021). Spatial and temporal intratumour heterogeneity has potential consequences for single biopsy-based neuroblastoma treatment decisions. *Nat. Commun.* 12, 6804. <https://doi.org/10.1038/s41467-021-26870-z>.
40. Körber, V., Stainczyk, S.A., Kurilov, R., Henrich, K.-O., Hero, B., Brors, B., Westermann, F., and Höfer, T. (2023). Neuroblastoma arises in early fetal development and its evolutionary duration predicts outcome. *Nat. Genet.* <https://doi.org/10.1038/s41588-023-01332-y>.
41. Harrigan, C.F., Rubanova, Y., Morris, Q., and Selega, A. (2020). TrackSig-Freq: subclonal reconstructions based on mutation signatures and allele frequencies. *Pac Symp Biocomput* 25, 238–249.
42. Reijns, M.A.M., Parry, D.A., Williams, T.C., Nadeu, F., Hindshaw, R.L., Rios Szwed, D.O., Nicholson, M.D., Carroll, P., Boyle, S., Royo, R., et al. (2022). Signatures of TOP1 transcription-associated mutagenesis in cancer and germline. *Nature*. <https://doi.org/10.1038/s41586-022-04403-y>.
43. on behalf of the Société Française des Cancers de l'Enfant (SFCE); Schleiermacher, G., Michon, J., Huon, I., d'Enghien, C.D., Klijanienko, J., Brisse, H., Ribeiro, A., Mosseri, V., Rubie, H., et al. (2007). Chromosomal CGH identifies patients with a higher risk of relapse in neuroblastoma without MYCN amplification. *Br. J. Cancer* 97, 238–246. <https://doi.org/10.1038/sj.bjc.6603820>.
44. Paulsen, T., Kumar, P., Koseoglu, M.M., and Dutta, A. (2018). Discoveries of Extrachromosomal Circles of DNA in Normal and Tumor Cells. *Trends Genet.* 34, 270–278. <https://doi.org/10.1016/j.tig.2017.12.010>.
45. Gaubatz, J.W. (1990). Extrachromosomal circular DNAs and genomic sequence plasticity in eukaryotic cells. *Mutation Research/DNAging* 237, 271–292. [https://doi.org/10.1016/0921-8734\(90\)90009-G](https://doi.org/10.1016/0921-8734(90)90009-G).
46. Noer, J.B., Hørsdal, O.K., Xiang, X., Luo, Y., and Regenber, B. (2022). Extrachromosomal circular DNA in cancer: history, current knowledge, and methods. *Trends Genet.* 38, 766–781. <https://doi.org/10.1016/j.tig.2022.02.007>.
47. Campbell, P.J., Yachida, S., Mudie, L.J., Stephens, P.J., Pleasance, E.D., Stebbings, L.A., Morsberger, L.A., Latimer, C., McLaren, S., Lin, M.-L., et al. (2010). The patterns and dynamics of genomic instability in metastatic pancreatic cancer. *Nature* 467, 1109–1113. <https://doi.org/10.1038/nature09460>.
48. Molenaar, J.J., Koster, J., Zwiijnenburg, D.A., van Sluis, P., Valentijn, L.J., van der Ploeg, I., Hamdi, M., van Nes, J., Westerman, B.A., van Arkel, J., et al. (2012). Sequencing of neuroblastoma identifies chromothripsis and defects in neuritogenesis genes. *Nature* 483, 589–593. <https://doi.org/10.1038/nature10910>.
49. Cox, D., Yuncen, C., and Spriggs, A.I. (1965). Minute chromatin bodies in malignant tumours of childhood. *Lancet* 286, 55–58. [https://doi.org/10.1016/S0140-6736\(65\)90131-5](https://doi.org/10.1016/S0140-6736(65)90131-5).
50. Webster, A.L.H., Sanders, M.A., Patel, K., Dietrich, R., Noonan, R.J., Lach, F.P., White, R.R., Goldfarb, A., Hadi, K., Edwards, M.M., et al. (2022). Genomic signature of Fanconi anaemia DNA repair pathway deficiency in cancer. *Nature*. <https://doi.org/10.1038/s41586-022-05253-4>.
51. Davies, H., Glodzik, D., Morganello, S., Yates, L.R., Staaf, J., Zou, X., Ramakrishna, M., Martin, S., Boyault, S., Sieuwerts, A.M., et al. (2017). HRDetect is a predictor of BRCA1 and BRCA2 deficiency based on mutational signatures. *Nat. Med.* 23, 517–525. <https://doi.org/10.1038/nm.4292>.
52. Blokzijl, F., de Ligt, J., Jager, M., Sasselli, V., Roerink, S., Sasaki, N., Huch, M., Boymans, S., Kijik, E., Prins, P., et al. (2016). Tissue-specific mutation accumulation in human adult stem cells during life. *Nature* 538, 260–264. <https://doi.org/10.1038/nature19768>.
53. Meier, B., Cooke, S.L., Weiss, J., Bailly, A.P., Alexandrov, L.B., Marshall, J., Raine, K., Maddison, M., Anderson, E., Stratton, M.R., et al. (2014). *C. elegans* whole-genome sequencing reveals mutational signatures related to carcinogens and DNA repair deficiency. *Genome Res.* 24, 1624–1636. <https://doi.org/10.1101/gr.175547.114>.
54. Drost, J., van Boxtel, R., Blokzijl, F., Mizutani, T., Sasaki, N., Sasselli, V., de Ligt, J., Behjati, S., Grolleman, J.E., van Wezel, T., et al. (2017). Use of CRISPR-modified human stem cell organoids to study the origin of mutational signatures in cancer. *Science* 358, 234–238. <https://doi.org/10.1126/science.aao3130>.
55. Olivier, M., Weninger, A., Ardin, M., Huskova, H., Castells, X., Vallée, M.P., McKay, J., Nedelko, T., Muehlbauer, K.-R., Marusawa, H., et al. (2014). Modelling mutational landscapes of human cancers in vitro. *Sci. Rep.* 4, 4482. <https://doi.org/10.1038/srep04482>.



56. Zou, X., Owusu, M., Harris, R., Jackson, S.P., Loizou, J.I., and Nik-Zainal, S. (2018). Validating the concept of mutational signatures with isogenic cell models. *Nat. Commun.* 9, 1744. <https://doi.org/10.1038/s41467-018-04052-8>.
57. King, D., Li, X.D., Almeida, G.S., Kwok, C., Gravells, P., Harrison, D., Burke, S., Hallsworth, A., Jamin, Y., George, S., et al. (2020). MYCN expression induces replication stress and sensitivity to PARP inhibition in neuroblastoma. *Oncotarget* 11, 2141–2159. <https://doi.org/10.18632/oncotarget.27329>.
58. Wang, T., Liu, L., Chen, X., Shen, Y., Lian, G., Shah, N., Davidoff, A.M., Yang, J., and Wang, R. (2018). MYCN drives glutaminolysis in neuroblastoma and confers sensitivity to an ROS augmenting agent. *Cell Death Dis.* 9, 220. <https://doi.org/10.1038/s41419-018-0295-5>.
59. Chowdhry, S., Garcia, S., Nguyen, N.-P., Celeste, A., Tse, E., Milutinovic, S., Wilkinson, D., Hassig, C., and Kasibhatla, S. (2022). Abstract 1520: Replication stress and the inability to repair damaged DNA, the potential “Achilles’ heel” of ecDNA+ tumor cells. *Cancer Res.* 82, 1520. <https://doi.org/10.1158/1538-7445.AM2022-1520>.
60. Currall, B.B., Chiangmai, C., Talkowski, M.E., and Morton, C.C. (2013). Mechanisms for Structural Variation in the Human Genome. *Curr Genet Med Rep* 1, 81–90. <https://doi.org/10.1007/s40142-013-0012-8>.
61. Piazza, A., and Heyer, W.-D. (2019). Homologous Recombination and the Formation of Complex Genomic Rearrangements. *Trends Cell Biol.* 29, 135–149. <https://doi.org/10.1016/j.tcb.2018.10.006>.
62. Byrum, A.K., Vindigni, A., and Mosammaparast, N. (2019). Defining and Modulating “BRCAness. *Trends Cell Biol.* 29, 740–751. <https://doi.org/10.1016/j.tcb.2019.06.005>.
63. Mersch, J., Jackson, M.A., Park, M., Nebgen, D., Peterson, S.K., Single-terry, C., Arun, B.K., and Litton, J.K. (2015). Cancers associated with *BRCA* 1 and *BRCA* 2 mutations other than breast and ovarian: *BRCA* and Other Cancers. *Cancer* 121, 269–275. <https://doi.org/10.1002/cncr.29041>.
64. de Bono, J., Mateo, J., Fizazi, K., Saad, F., Shore, N., Sandhu, S., Chi, K.N., Sartor, O., Agarwal, N., Olmos, D., et al. (2020). Olaparib for Metastatic Castration-Resistant Prostate Cancer. *N. Engl. J. Med.* 382, 2091–2102. <https://doi.org/10.1056/NEJMoa1911440>.
65. Heeke, A.L., Pishvaian, M.J., Lynce, F., Xiu, J., Brody, J.R., Chen, W.-J., Baker, T.M., Marshall, J.L., and Isaacs, C. (2018). Prevalence of Homologous Recombination-Related Gene Mutations Across Multiple Cancer Types. *JCO Precision Oncology*, 1–13. <https://doi.org/10.1200/PO.17.00286>.
66. Zhang, J. (2013). The role of *BRCA*1 in homologous recombination repair in response to replication stress: significance in tumorigenesis and cancer therapy. *Cell Biosci.* 3, 11. <https://doi.org/10.1186/2045-3701-3-11>.
67. Creeden, J.F., Nanavaty, N.S., Einloth, K.R., Gillman, C.E., Stanbery, L., Hamouda, D.M., Dworkin, L., and Nemunaitis, J. (2021). Homologous recombination proficiency in ovarian and breast cancer patients. *BMC Cancer* 21, 1154. <https://doi.org/10.1186/s12885-021-08863-9>.
68. Gundem, G., Levine, M.F., Roberts, S.S., Cheung, I.Y., Medina-Martínez, J.S., Feng, Y., Arango-Ossa, J.E., Chadoutaud, L., Rita, M., Asimomitis, G., et al. (2022). Clonal evolution during metastatic spread in high-risk neuroblastoma. *Genomics*. <https://doi.org/10.1101/2022.08.15.503973>.
69. Evolution, P.C.A.W.G., Heterogeneity Working Group; Consortium, P.C.A.W.G., Gerstung, M., Jolly, C., Leshchiner, I., Dentre, S.C., Gonzalez, S., Rosebrock, D., Mitchell, T.J., Rubanova, Y., et al. (2020). The evolutionary history of 2,658 cancers. *Nature* 578, 122–128. <https://doi.org/10.1038/s41586-019-1907-7>.
70. Verhaak, R.G.W., Bafna, V., and Mischel, P.S. (2019). Extrachromosomal oncogene amplification in tumour pathogenesis and evolution. *Nat. Rev. Cancer* 19, 283–288. <https://doi.org/10.1038/s41568-019-0128-6>.
71. Bergstrom, E.N., Luebeck, J., Petljak, M., Khandekar, A., Barnes, M., Zhang, T., Steele, C.D., Pillay, N., Landi, M.T., Bafna, V., et al. (2022). Mapping clustered mutations in cancer reveals APOBEC3 mutagenesis of ecDNA. *Nature*. <https://doi.org/10.1038/s41586-022-04398-6>.
72. González, R.C., Conrad, T., Stöber, M.C., Xu, R., Giurgiu, M., Rodríguez-Fos, E., Kasack, K., Brückner, L., Eric van Leen, Helmsauer, K., et al. Parallel sequencing of extrachromosomal circular DNAs and transcriptomes in single cancer cells. *Nat. Genet.* (in press).
73. FastQC.
74. Li, H. (2013). Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. Preprint at arXiv. <https://doi.org/10.1048550/arXiv.1303.3997>.
75. Danecek, P., Bonfield, J.K., Liddle, J., Marshall, J., Ohan, V., Pollard, M.O., Whitwham, A., Keane, T., McCarthy, S.A., Davies, R.M., et al. (2021). Twelve years of SAMtools and BCftools. *GigaScience* 10, giab008. <https://doi.org/10.1093/gigascience/giab008>.
76. Tischler, G., and Leonard, S. (2014). biobambam: tools for read pair collation based algorithms on BAM files. *Source Code Biol. Med.* 9, 13. <https://doi.org/10.1186/1751-0473-9-13>.
77. McKenna, A., Hanna, M., Banks, E., Sivachenko, A., Cibulskis, K., Kernysky, A., Garimella, K., Altshuler, D., Gabriel, S., Daly, M., et al. (2010). The Genome Analysis Toolkit: A MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res.* 20, 1297–1303. <https://doi.org/10.1101/gr.107524.110>.
78. Rausch, T., Zichner, T., Schlattl, A., Stutz, A.M., Benes, V., and Korbel, J.O. (2012). DELLY: structural variant discovery by integrated paired-end and split-read analysis. *Bioinformatics* 28, i333–i339. <https://doi.org/10.1093/bioinformatics/bts378>.
79. Wala, J.A., Bandopadhyay, P., Greenwald, N.F., O’Rourke, R., Sharpe, T., Stewart, C., Schumacher, S., Li, Y., Weischenfeldt, J., Yao, X., et al. (2018). SvABA: genome-wide detection of structural variants and indels by local assembly. *Genome Res.* 28, 581–591. <https://doi.org/10.1101/gr.221028.117>.
80. Chong, Z., Ruan, J., Gao, M., Zhou, W., Chen, T., Fan, X., Ding, L., Lee, A.Y., Boutros, P., Chen, J., et al. (2017). novoBreak: local assembly for breakpoint detection in cancer genomes. *Nat. Methods* 14, 65–67. <https://doi.org/10.1038/nmeth.4084>.
81. Van Loo, P., Nordgard, S.H., Lingjaerde, O.C., Russnes, H.G., Rye, I.H., Sun, W., Weigman, V.J., Marynen, P., Zetterberg, A., Naume, B., et al. (2010). Allele-specific copy number analysis of tumors. *Proc. Natl. Acad. Sci. USA* 107, 16910–16915. <https://doi.org/10.1073/pnas.1009843107>.
82. Fan, Y., Xi, L., Hughes, D.S.T., Zhang, J., Zhang, J., Futreal, P.A., Wheeler, D.A., and Wang, W. (2016). MuSE: accounting for tumor heterogeneity using a sample-specific error model improves sensitivity and specificity in mutation calling from sequencing data. *Genome Biol.* 17, 178. <https://doi.org/10.1186/s13059-016-1029-6>.
83. Ye, K., Schulz, M.H., Long, Q., Apweiler, R., and Ning, Z. (2009). Pindel: a pattern growth approach to detect break points of large deletions and medium sized insertions from paired-end short reads. *Bioinformatics* 25, 2865–2871. <https://doi.org/10.1093/bioinformatics/btp394>.
84. Nik-Zainal, S., Van Loo, P., Wedge, D.C., Alexandrov, L.B., Greenman, C.D., Lau, K.W., Raine, K., Jones, D., Marshall, J., Ramakrishna, M., et al. (2012). The life history of 21 breast cancers. *Cell* 149, 994–1007. <https://doi.org/10.1016/j.cell.2012.04.023>.
85. Quinlan, A.R., and Hall, I.M. (2010). BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* 26, 841–842. <https://doi.org/10.1093/bioinformatics/btq033>.
86. McLaren, W., Gil, L., Hunt, S.E., Riat, H.S., Ritchie, G.R.S., Thormann, A., Flicek, P., and Cunningham, F. (2016). The Ensembl Variant Effect Predictor. *Genome Biol.* 17, 122. <https://doi.org/10.1186/s13059-016-0974-4>.
87. Fantini, D., Vidimar, V., Yu, Y., Condello, S., and Meeks, J.J. (2020). Mut-Signatures: an R package for extraction and analysis of cancer

- mutational signatures. *Sci. Rep.* 10, 18217. <https://doi.org/10.1038/s41598-020-75062-0>.
88. Daniel Huebschmann, Z.G. (2017). YAPSA. <https://doi.org/10.1018129/B9.BIOC.YAPSA>.
89. Letouzé, E., Shinde, J., Renault, V., Couchy, G., Blanc, J.-F., Tubacher, E., Bayard, Q., Bacq, D., Meyer, V., Semhoun, J., et al. (2017). Mutational signatures reveal the dynamic interplay of risk factors and cellular processes during liver tumorigenesis. *Nat. Commun.* 8, 1315. <https://doi.org/10.1038/s41467-017-01358-x>.
90. Evolution, P.C.A.W.G., Heterogeneity Working Group; Consortium, P.C.A.W.G., Rubanova, Y., Shi, R., Harrigan, C.F., Li, R., Wintersinger, J., Sahin, N., Deshwar, A., and Morris, Q. (2020). Reconstructing evolutionary trajectories of mutation signature activities in cancer using TrackSig. *Nat. Commun.* 11, 731. <https://doi.org/10.1038/s41467-020-14352-7>.
91. Deshpande, V., Luebeck, J., Nguyen, N.-P.D., Bakhtiari, M., Turner, K.M., Schwab, R., Carter, H., Mischel, P.S., and Bafna, V. (2019). Exploring the landscape of focal amplifications in cancer using AmpliconArchitect. *Nat. Commun.* 10, 392. <https://doi.org/10.1038/s41467-018-08200-y>.
92. Cortés-Ciriano, I., Lee, J.J.-K., Xi, R., Jain, D., Jung, Y.L., Yang, L., Gordenin, D., Klimczak, L.J., Zhang, C.-Z., Pellman, D.S., et al. (2020). Comprehensive analysis of chromothripsis in 2,658 human cancers using whole-genome sequencing. *Nat. Genet.* 52, 331–341. <https://doi.org/10.1038/s41588-019-0576-7>.
93. Lex, A., Gehlenborg, N., Strobel, H., Vuilleumot, R., and Pfister, H. (2014). UpSet: Visualization of Intersecting Sets. *IEEE Trans. Vis. Comput. Graph.* 20, 1983–1992. <https://doi.org/10.1109/TVCG.2014.2346248>.
94. Kassambara, A., and Mundt, F. (2020). Factoextra: Extract and Visualize the Results of Multivariate Data Analyses.
95. Therneau, T.M., and Grambsch, P.M. (2000). *Modeling Survival Data: Extending the Cox Model* (Springer).
96. Wei, T., and Simko, V. (2021). R Package “Corrplot”: Visualization of a Correlation Matrix.
97. Lawrence, M., Huber, W., Pagès, H., Aboyoun, P., Carlson, M., Gentleman, R., Morgan, M.T., and Carey, V.J. (2013). Software for Computing and Annotating Genomic Ranges. *PLoS Comput. Biol.* 9, e1003118. <https://doi.org/10.1371/journal.pcbi.1003118>.
98. Wickham, H. (2016). *ggplot2: Elegant Graphics for Data Analysis*, 2nd ed (Springer International Publishing : Imprint: Springer). <https://doi.org/10.1007/978-3-319-24277-4>.
99. Wickham, H., François, R., Henry, L., Müller, K., and Vaughan, D. (2023). *Dplyr: A Grammar of Data Manipulation*.
100. Kassambara, A., Kosinski, M., Biecek, P., and Fabian, S. (2021). *Survminer: Drawing Survival Curves Using “Ggplot2”*.
101. Gu, Z., Eils, R., and Schlesner, M. (2016). Complex heatmaps reveal patterns and correlations in multidimensional genomic data. *Bioinformatics* 32, 2847–2849. <https://doi.org/10.1093/bioinformatics/btw313>.
102. Wickham, H. (2007). Reshaping data with the reshape package. *J. Stat. Software* 21.
103. Patil, I. (2019). *pairwiseComparisons: Multiple Pairwise Comparison Tests*.
104. Patil, I. (2021). Visualizations with statistical details: The “ggstatsplot” approach. *JOSS* 6, 3167. <https://doi.org/10.21105/joss.03167>.
105. Harrell, F., Jr., and Dupont, C. (2019). *Hmisc: Harrell Miscellaneous*.



## STAR★METHODS

### KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
<b>Deposited data</b>		
Neuroblastoma whole-genome sequencing data	European Genome-Phenome Archive <a href="https://ega-archive.org/">https://ega-archive.org/</a>	EGA: EGAS00001001308, <sup>34</sup> EGA: EGAS00001004022, <sup>30</sup> EGA: EGAS00001006983, EGA: EGAS00001007016, and EGA: EGAS00001007019
Original code for all the analyses	Henssen lab	Github: <a href="https://github.com/henssen-lab/mutsignsNBLpaper">https://github.com/henssen-lab/mutsignsNBLpaper</a>
Raw data results	This paper	Zenodo: <a href="https://doi.org/10.5281/zenodo.8032024">https://doi.org/10.5281/zenodo.8032024</a>
<b>Software and algorithms</b>		
FastQC v.0.11.8	Babraham bioinformatics <sup>73</sup>	<a href="https://www.bioinformatics.babraham.ac.uk/projects/fastqc/">https://www.bioinformatics.babraham.ac.uk/projects/fastqc/</a>
BWA-MEM v.0.7.15	Li et al. <sup>74</sup>	<a href="https://github.com/lh3/bwa">https://github.com/lh3/bwa</a>
SAMtools v.1.10	Danecek et al. <sup>75</sup>	<a href="https://github.com/samtools/samtools">https://github.com/samtools/samtools</a>
Biobambam v.2.0.87	Tischler et al. <sup>76</sup>	<a href="https://gitlab.com/german.tischler/biobambam2">https://gitlab.com/german.tischler/biobambam2</a>
Mutect2 v.4.1.8.1	McKenna et al. <sup>77</sup>	<a href="https://github.com/broadinstitute/gatk/releases">https://github.com/broadinstitute/gatk/releases</a>
GATK v.4.1.9.0	McKenna et al. <sup>77</sup>	<a href="https://github.com/broadinstitute/gatk/releases">https://github.com/broadinstitute/gatk/releases</a>
bedtools v.2.29.2	Quinlan et al. <sup>85</sup>	<a href="https://bedtools.readthedocs.io/en/latest/index.html">https://bedtools.readthedocs.io/en/latest/index.html</a>
ASCAT v.2.6	Van Loo et al. <sup>81</sup>	<a href="https://github.com/VanLoo-lab/ascats">https://github.com/VanLoo-lab/ascats</a>
Novobreak v.1.1.3	Chong et al. <sup>80</sup>	<a href="https://github.com/czc/nb_distribution">https://github.com/czc/nb_distribution</a>
SvABA v.1.1.0	Wala et al. <sup>79</sup>	<a href="https://github.com/walaj/svaba">https://github.com/walaj/svaba</a>
Delly2 v.0.7.7 and v.0.8.1	Rausch et al. <sup>78</sup>	<a href="https://github.com/dellytools/delly">https://github.com/dellytools/delly</a>
Ensembl VEP v.102.0	McLaren et al. <sup>86</sup>	<a href="https://www.ensembl.org/info/docs/tools/vep/index.html">https://www.ensembl.org/info/docs/tools/vep/index.html</a>
mutSignatures v.2.1.1	Fantini et al. <sup>87</sup>	<a href="https://github.com/dami82/mutSignatures">https://github.com/dami82/mutSignatures</a>
YAPSA v.1.16.0	Daniel Huebschmann et al. <sup>88</sup>	<a href="https://github.com/HiDiHlabs/YAPSA">https://github.com/HiDiHlabs/YAPSA</a>
Palimpsest v.2.0.0	Letouze et al. <sup>89</sup>	<a href="https://github.com/FunGeST/Palimpsest">https://github.com/FunGeST/Palimpsest</a>
CNA signatures (no name)	Drews et al. <sup>8</sup>	<a href="https://github.com/markowetzlab/Drews2022_CIN_Compendium">https://github.com/markowetzlab/Drews2022_CIN_Compendium</a>
JaBba v.1.0	Hadi et al. <sup>17</sup>	<a href="https://github.com/mskilab/JaBba">https://github.com/mskilab/JaBba</a>
Amplicon Architect v.1.2	Deshpande et al. <sup>91</sup>	<a href="https://github.com/virajbdeshpande/AmpliconArchitect">https://github.com/virajbdeshpande/AmpliconArchitect</a>
Shatterseek v.0.5	Cortés-Ciriano et al. <sup>92</sup>	<a href="https://github.com/parklab/ShatterSeek">https://github.com/parklab/ShatterSeek</a>
gGnome v.0.1	Imieliński, M.	<a href="https://github.com/mskilab/gGnome">https://github.com/mskilab/gGnome</a>
AmpliconClassifier v.0.4.6	Deshpande et al. <sup>16</sup>	<a href="https://github.com/jluebeck/AmpliconClassifier">https://github.com/jluebeck/AmpliconClassifier</a>
MuSE2.0 v.1.0rc	Fan et al. <sup>81</sup>	<a href="https://github.com/danielfan/MuSE">https://github.com/danielfan/MuSE</a>
Pindel v.0.2.5b9	Ye et al. <sup>82</sup>	<a href="https://github.com/genome/pindel">https://github.com/genome/pindel</a>
Battenberg v.2.2.9	Nik-Zainal et al. <sup>84</sup>	<a href="https://github.com/Wedge-lab/battenberg">https://github.com/Wedge-lab/battenberg</a>
TrackSigFreq	PCAWG Evolution and Heterogeneity Working Group et al. <sup>41</sup>	<a href="https://github.com/morrislab/TrackSigFreq">https://github.com/morrislab/TrackSigFreq</a>
R v.4.0.3 with packages:		N/A
ComplexUpset v.1.2.1	Lex et al. <sup>93</sup>	<a href="https://krassowski.github.io/complex-upset/index.html">https://krassowski.github.io/complex-upset/index.html</a>
corrplot v.0.90	Taiyun et al. <sup>96</sup>	<a href="https://github.com/taiyun/corrplot">https://github.com/taiyun/corrplot</a>
dplyr v.1.0.3	Wickham et al. <sup>99</sup>	<a href="https://dplyr.tidyverse.org/">https://dplyr.tidyverse.org/</a>
GenomicRanges v.1.42.0	Lawrence et al. <sup>97</sup>	<a href="https://bioconductor.org/packages/release/bioc/html/GenomicRanges.html">https://bioconductor.org/packages/release/bioc/html/GenomicRanges.html</a>
ggplot2 v.3.3.5	Wickham et al. <sup>98</sup>	<a href="https://ggplot2.tidyverse.org/">https://ggplot2.tidyverse.org/</a>
gUtils v.0.2.0	<a href="http://mskilab.com/gGnome/tutorial.html">http://mskilab.com/gGnome/tutorial.html</a>	<a href="https://github.com/mskilab/gUtils">https://github.com/mskilab/gUtils</a>
IRanges v.2.24.0	Lawrence et al. <sup>97</sup>	<a href="https://bioconductor.org/packages/release/bioc/html/IRanges.html">https://bioconductor.org/packages/release/bioc/html/IRanges.html</a>
ComplexHeatmap v.2.4.2	Gu et al. <sup>101</sup>	<a href="https://bioconductor.org/packages/release/bioc/vignettes/ComplexHeatmap/inst/doc/complex_heatmap.html">https://bioconductor.org/packages/release/bioc/vignettes/ComplexHeatmap/inst/doc/complex_heatmap.html</a>

(Continued on next page)

**Continued**

REAGENT or RESOURCE	SOURCE	IDENTIFIER
reshape v.0.8.8,	Wickham et al. <sup>102</sup>	<a href="https://cran.r-project.org/web/packages/reshape/index.html">https://cran.r-project.org/web/packages/reshape/index.html</a>
pairwiseComparisons v.3.1.2	Patil et al. <sup>103</sup>	<a href="https://cran.r-project.org/src/contrib/Archive/pairwiseComparisons/">https://cran.r-project.org/src/contrib/Archive/pairwiseComparisons/</a>
ggstatsplot v.0.6.8	Patil et al. <sup>104</sup>	<a href="https://indrajeetpatil.github.io/ggstatsplot/">https://indrajeetpatil.github.io/ggstatsplot/</a>
Hmisc v.4.4.2,	Harrell et al. <sup>105</sup>	<a href="https://cran.r-project.org/web/packages/Hmisc/index.html">https://cran.r-project.org/web/packages/Hmisc/index.html</a>
BSgenome.Hsapiens. UCSC.hg19	Team TBD	<a href="https://bioconductor.org/packages/release/data/annotation/html/BSgenome.Hsapiens.UCSC.hg19.html">https://bioconductor.org/packages/release/data/annotation/html/BSgenome.Hsapiens.UCSC.hg19.html</a>
survival v.3.2–11	Therneau et al. <sup>95</sup>	<a href="https://cran.r-project.org/web/packages/survival/index.html">https://cran.r-project.org/web/packages/survival/index.html</a>
survminer v.0.4.9	Kassambara et al. <sup>100</sup>	<a href="https://cran.r-project.org/web/packages/survminer/index.html">https://cran.r-project.org/web/packages/survminer/index.html</a>
signature.tools.lib v2.1.2	Degasperi et al. <sup>9</sup>	<a href="https://github.com/Nik-Zainal-Group/signature.tools.lib">https://github.com/Nik-Zainal-Group/signature.tools.lib</a>
factoextra v1.0.7	Kassambara et al. <sup>94</sup>	<a href="https://cran.r-project.org/web/packages/factoextra/index.html">https://cran.r-project.org/web/packages/factoextra/index.html</a>

## RESOURCE AVAILABILITY

### Lead contact

Further information and requests for resources should be directed to and will be fulfilled by the lead contact, Anton G. Henssen ([henssenlab@gmail.com](mailto:henssenlab@gmail.com)).

### Materials availability

This study did not generate new unique reagents.

### Data and code availability

- The WGS data that support the findings of this study have been deposited with the European Genome-phenome Archive (<https://www.ebi.ac.uk/ega/>) under accession nos. EGA: EGAS00001001308,<sup>34</sup> EGA: EGAS00001004022,<sup>30</sup> EGA: EGAS00001006983, EGA: EGAS00001007016, and EGA: EGAS00001007019 and are publicly available as of the date of publication.
- All original code used to analyze the sequencing data, perform the statistical analysis, and generate the plots have been deposited and is publicly accessible in GitHub Github: <https://github.com/henssen-lab/mutsignsNBLpaper> as of the date of publication.
- All the raw data required to support the conclusions reported in this manuscript, including the results from variant calling and mutational signature analysis, is available in Zenodo under DOI: Zenodo: <https://doi.org/10.5281/zenodo.8032024>.
- Any additional information required to reanalyze the data reported in this paper is available from the [lead contact](#) upon request.

## EXPERIMENTAL MODEL AND STUDY PARTICIPANT DETAILS

This study comprised the analyses of tumor and blood samples of 150 human patients diagnosed with neuroblastoma between 1991 and 2016. All stages, sex and age were included in this study (Figure 1). All samples were collected at diagnosis from untreated patients. Patients were registered and treated according to the trial protocols of the German Society of Pediatric Oncology and Hematology (GPOH). This study was conducted in accordance with the World Medical Association Declaration of Helsinki (2013) and good clinical practice; informed consent was obtained from all patients or their guardians. The collection and use of patient specimens was approved by the institutional review boards of Charité-Universitätsmedizin Berlin and the Medical Faculty, University of Cologne. Specimens and clinical data were archived and made available by Charité-Universitätsmedizin Berlin or the National Neuroblastoma Biobank and Neuroblastoma Trial Registry (University Children's Hospital Cologne) of the GPOH.

## METHOD DETAILS

### Sequencing data

This study is based on the analysis of two cohorts: (1) Discovery cohort, and (2) Validation cohort.

#### Discovery cohort

120 neuroblastoma primary tumor and matching control samples. Those samples are divided into two datasets. The first one, sequenced in Cologne, was completed in 2015,<sup>34</sup> and it is composed of WGS from 56 primary tumor and matching blood control samples (Illumina HiSeq). It was downloaded (<https://www.ebi.ac.uk/ega/>) under accession number EGA: EGAS00001001308.<sup>34</sup>

The second one, sequenced at the German Cancer Research Center, was completed in 2017–2019, and it includes WGS of 64 primary tumor and matching control samples (Illumina HiSeq X Ten). The data of 53 out of the total 64 patients was publicly available and downloaded (<https://www.ebi.ac.uk/ega/>) under accession number EGA: EGAS00001004022.<sup>30</sup> The data of the 11 remaining patients is now available (<https://www.ebi.ac.uk/ega/>) under accession number EGA: EGAS00001006983. The quality of the raw data was assured using FastQC.<sup>73</sup> Reads were 3' trimmed for both quality and adapter sequences, with reads removed if the length was shorter than 20 nucleotides. Burrows–Wheeler Aligner MEM<sup>74</sup> v.0.7.15 with default parameters was used to align<sup>75</sup> the reads to human reference assembly hg19. PCR and optical duplicates were marked with bammarkduplicates2 from BIOBAMBAM2<sup>76</sup> v2.0.87 (<https://github.com/gt1/biobambam2>). Six patients have been excluded from the analyses due to an abnormally high number of SVs detected (CB2044, NBL47, NBL49, NBL50, NBL53, NBL54). They present in average more than 5-fold SVs than the rest of the cohort (median of SVs of the cohort: 14 SVs per patient; median of SVs of the excluded patients: 70.5 SVs per patient).

#### **Validation cohort**

This dataset was completed in January 2022, and it is composed of ultra-deep WGS (x200) from 39 primary tumor and matching control samples (adrenal gland, blood, fat tissue, lymph node, muscle, skin) (Illumina NovaSeq S4). The data is available (<https://www.ebi.ac.uk/ega/>) under accession numbers EGA: EGAS00001007016 and EGA: EGAS00001007019. The quality of the raw data was assured using FastQC.<sup>73</sup> Reads were 3' trimmed for both quality and adapter sequences, with reads removed if the length was shorter than 20 nucleotides. Burrows–Wheeler Aligner MEM<sup>74</sup> v.0.7.15 with default parameters was used to align<sup>75</sup> the reads to human reference assembly hg19. PCR and optical duplicates were marked with bammarkduplicates2 from BIOBAMBAM2<sup>76</sup> v2.0.87 (<https://github.com/gt1/biobambam2>). Three patients have been excluded from the analyses due to a contamination on normal sample (A06R-NFBQNJ, A06R-NFVDJM, A06R-GNBIPE).

#### **Variant calling**

##### **SNVs and indels**

Somatic single-nucleotide variants and small insertions and deletions were detected using Mutect2 v.4.1.8.1 from the GATK<sup>77</sup> software package, with standard parameters according to GATK best practices recommendations. Variants were filtered using recommended filtering parameters, removing all germline mutations using a panel of normals.

##### **SVs**

In matched normal and tumor samples, somatic structural variants were called using three different variant callers with their corresponding complementary methodologies: Delly2<sup>78</sup> v0.7.7, SvABA<sup>79</sup> v.1.1.0, and Novobreak<sup>80</sup> v.1.1.3 with default parameters. To increase specificity, we followed the standard practices for SV consensus calling and intersected all the variants from the three callers keeping the ones detected at least in two of them. An exception has been made for insertions since we only identified them using SvABA. Variants with both breakpoints within a window of 500 bp were collapsed. To increase sensitivity in our consensus approach, each variant included in the final set passed all the standard filters in at least one of the required two callers to be considered in our analysis, not requiring the passing of all the standard filters in both callers.

##### **CNAs**

Allele-specific copy-number (CN) profiles were generated from tumor and normal B allele frequencies (BAFs) and LogR values using ASCAT<sup>81</sup> v2.6 with parameters adjusted for sequencing data input (segmentation penalty = 200 and aspcf function gamma = 1). The segmentation procedure from ASCAT was replaced with a custom implementation that only considers BAFs to determine start and endpoints of segments but still estimates the segment's coverage using the log coverage ratios. This approach avoids potential over-segmentation due to noisy LogR values. All CN profiles were inspected manually for quality control. For samples with an estimated tumor purity, below 60% CN calling was rerun with adjusted purity and ploidy values that were manually selected after inspection of the sunrise plots and in agreement with pathology estimates of tumor purity. We assigned different copy number states to all profiles considering the local copy number and overall ploidy estimates. CN gain was defined as  $\log_2((\text{Cn}_{\text{major}} + \text{Cn}_{\text{minor}})/\text{ploidy}) > 0.3$ ; CN loss as  $\log_2((\text{Cn}_{\text{major}} + \text{Cn}_{\text{minor}})/\text{ploidy}) < -0.3$ . CN amplifications and homozygous losses have been determined following COSMIC CNA parameters (<https://cancer.sanger.ac.uk/cosmic/help/cnv/overview>). In the case of amplifications, we added an additional filter of  $\log_2((\text{Cn}_{\text{major}} + \text{Cn}_{\text{minor}})/\text{ploidy}) > 1.25$  to increase specificity. For three patients no CNA results could be obtained (NBL31, NBL36, and NBL61).

#### **Variant calling in the validation cohort**

Somatic SNVs were detected using MuSE2.0<sup>82</sup> v1.0rc (<https://github.com/wwylab/MuSE>) in two-steps: Muse Sump and Muse call, using recommended parameters. SNVs were filtered using tier-based cutoffs from a sample-specific model computed by the method, removing Tier5 class. Indels were called using Pindel<sup>83</sup> v 0.2.5b9. with standard parameters and filtered following the next parameters;  $\text{VAF}=0.2$ ,  $\text{cov} = 20$ ,  $\text{hom} = 6$ . Somatic SVs were detected using Delly2<sup>78</sup> v0.8.1 with default parameters. SVs were processed and filtered using recommended filtering parameters, removing contamination and variants <50bp in length. Allele-specific copy-number (CN) profiles were generated using Battenberg<sup>84</sup> v2.2.9 with default parameters. All the reference files used on general static were downloaded from (<https://ora.ox.ac.uk/objects/uuid:2c1fec09-a504-49ab-9ce9-3f17bac531bc>) following the recommendations by the developers of the method.

## Identification of variants affecting genes and chromosome arms

### Chromosomal arms

Gains and losses of chromosomal arms have been determined intersecting the coordinates of the chromosomal regions with the CNA gains and losses, using bedtools<sup>85</sup> v2.29.2 with parameters `-wo -f 1 -F 0.60` (<https://bedtools.readthedocs.io/>). We considered an overlap of the whole CNA with at least 60% of the chromosomal arm.

### Genes

We predicted the variant effect for SNV and indel calls using Ensembl VEP<sup>86</sup> (Variant Effect Predictor) v.102.0. Only SNVs and indels classified as *missense* and *stop gained* overlapping genes, were considered. For SV calls, we considered all the genes affected by SV breakpoints. We classified them in two types: (1) SV close to gene, defined as genes with at least one SV breakpoint at a distance of  $\pm 20$  kb; (2) SV within gene, defined as genes overlapping at least one SV breakpoint. For CNA calls, we considered the genes overlapping amplifications or homozygous deletions.

## Mutational signatures analyses

The mutational signatures analyses have been performed using the same methodology for both the discovery and validation cohorts.

### SNV signatures (SBS)

To retrieve mutational signatures related to somatic single-nucleotide variants operating in neuroblastoma, we used the R package mutSignatures<sup>87</sup> v.2.1.1. We started pre-processing all the SNVs in coding and non-coding regions detected in our patients. We applied a filtering step for non-SNV variants, followed by the tri-nucleotide context extraction for each of the variants which are classified in 96 different mutation types, or features, and counted across samples. Next, we ran the *de-novo* signature extraction step where non-negative matrix factorization (NMF) is performed. Complying with the pipeline recommendations, we computed 500 iterations, first, evaluating different choice of ranks/number of signatures:  $k=4$ ,  $k=5$ , and  $k=6$ . We chose  $k=4$ , resulting in four *de-novo* signatures, based on the presence of clusters with consistent non-negative high silhouette scores. Then, we determined which known COSMIC v.3.2 signatures (94 SBS reference signatures; <https://cancer.sanger.ac.uk/signatures/sbs/>) matched our *de-novo* signatures, using a cosine similarity threshold of  $\geq 0.85$ . In the deconvolution step, we estimated the absolute and relative exposure of the COSMIC mutational signatures in each of the patients. Next, to increase specificity and reduce the false positive assignment we filtered all signatures showing an absolute exposure lower than 5% of the total number of mutations in our cohort. Finally, we obtained four COSMIC signatures: SBS3, SBS5, SBS18, and SBS40. The same signatures were extracted in the validation cohort.

### Indel signatures (ID)

Mutational signatures associated with small somatic insertions and deletions (<50 bp) have been extracted using the R package YAPSA<sup>88</sup> v.1.16.0, which includes PCAWG/COSMIC indel signatures (18 ID reference signatures; <https://cancer.sanger.ac.uk/signatures/id/>), following the standard pipeline. This pipeline starts with a pre-processing step of all the indels detected in our patients, in which we annotate the sequence context 10bp downstream and 60bp upstream of each variant, the variant type, and the length. In line with the SNV signature analysis, indels are classified in 83 different classes of features, and counted across samples. Next, we computed the supervised mutational signature analysis step, based on the linear combination decomposition function (LCD), obtaining the exposure for each known mutational signature in each patient. Signature exposure per neuroblastoma risk group was estimated from the exposure per patient. In order to reduce false positive calls, recommended signature-specific cutoffs for the indel-based PCAWG signatures have been applied, using the determined value of *optimal cost factor* = 3. As an additional step to increase specificity, we filtered all signatures showing an absolute exposure lower than 5% of the total number of indels in our cohort. From this analysis, we have obtained six PCAWG/COSMIC indel signatures: ID1, ID2, ID4, ID6, ID8, and ID9. Applying the same methodology to the validation cohort we obtained: ID2, ID4, ID5, ID6, and ID9 indel signatures.

### CNA signatures (CX)

To quantify the activity of copy number signatures in our samples, we used the 17 signatures that have been derived from over 6,000 tumors of the cancer genome atlas,<sup>8</sup> following the methodology used by Drews et al. (17 CX reference signatures; [https://github.com/markowetzlab/Drews2022\\_CIN\\_Compendium](https://github.com/markowetzlab/Drews2022_CIN_Compendium)). Using the raw absolute copy number as obtained from ASCAT (for the discovery cohort) or Battenberg (for the validation cohort) as input, very copy number quiet samples are removed and first 43 features describing the fundamental features of the copy number profiles are extracted from the remaining profiles. Afterward the exposure for each previously described signature in each of our samples is calculated based on the posterior distribution of the extracted features. As an additional step to increase specificity, we filtered all signatures showing an exposure lower than 5% of the total exposure in our cohort. From this analysis we obtained eight CNA-based signatures: CX1, CX2, CX3, CX5, CX7, CX11, CX14, and CX15. Applying the same methodology to the validation cohort we obtained: CX1, CX2, CX3, CX4, CX6, CX14, and CX15 CNA signatures.

### SV signatures (SV)

Mutational signatures associated with somatic structural variants have been extracted using the R package Palimpsest<sup>89</sup> v.2.0.0 following the standard pipeline for this variant type. This pipeline starts with a pre-processing step in which we annotate the type (deletions, inversions, tandem-duplications, and translocations), size (<1 kb, 1–10 kb, 10–100 kb, 100 kb–1 Mb, 1–10 Mb, and >10 Mb), and clustered nature of rearrangements (distinguish between clustered –  $\geq 10$  breakpoints within a 1 Mb window – and non-clustered events). SVs are classified into 38 different classes of features and counted across samples. Next, we computed the *de-novo* signature extraction step, based on NMF. Complying with the pipeline recommendations, we computed 500 iterations with *num\_of\_sigs* = *auto* (the appropriate rank/number of signatures was estimated from NMF metrics such as the cophenetic

distance, obtaining  $k=5$ ). In the final step of the pipeline, we estimated the absolute and relative exposure of each signature in each sample. Following what we did for the other variant types to increase specificity we filtered all signatures showing an absolute exposure lower than 5% of the total number of mutations in our cohort. As a result, we obtained five *de novo* signatures. In order to evaluate if the *de novo* SV-based signatures found in neuroblastoma matched any of the 21 reference SV signatures from previous pan-cancer studies<sup>9</sup> we used the *deconvolution\_compare* function of Palimpsest obtaining cosine similarities between signatures. Using the same cosine similarity threshold from the SNV analysis ( $\geq 0.85$ ) we identified two matches. *De novo* signatures SV2, and SV3 corresponded to reference signatures R6a (*MDM2*, *CDK4*, and 17q mutations), R6b (*MDM2*, and *CDK4* mutations).<sup>9</sup> Although SV1 and R2 (*TP53* mutations) also showed a cosine similarity  $\geq 0.85$  we could not confirm the match due to mutational profile distances (high presence of small deletions < 1kb in SV1). In addition, we described three *de-novo* signatures that did not match any reference signature. We named them based on the prevalence of SVs in the different features classes: (1) SV1, small deletions; (2) SV4, large simple intra and interchromosomal SVs + clustered translocations; (3) SV5, medium size simple intra and interchromosomal SVs. Applying the same methodology to the validation cohort we obtained 5 *de novo* signatures, 3 of which present a cosine similarity  $>0.75$  with the SV-based signatures from the discovery cohort.

From the patient's exposure, we assessed the contribution of each of the different mutational signatures from the different variant types within neuroblastoma risk groups.

### Subclonal signatures analysis

The analysis of subclonal signature trajectories has been performed with the method TrackSigFreq<sup>41</sup> based on the observed density of mutation frequencies and changes in mutational signature activities. This method has been used instead of its predecessor, TrackSig,<sup>90</sup> because it adds the advantage to identify distinct populations of mutations that share similar signature activities. TrackSigFreq has been executed using the SNV VCF file generated by MuSE2.0<sup>82</sup> as described in the *variant calling* section, and copy number and purity information from Battenberg.<sup>84</sup> We selected the mutational signatures detected by mutSignatures, as explained in the *mutational signatures analysis* section, to evaluate the changes in the trajectory of the mutational signature activity through the different cancer cell fractions. To avoid losing mutational signatures with low activity, we set the exposure filter to 0 to visualize the performance of all mutational signatures without restrictions. To compare the activity of the different signatures across the cancer cell fraction (CCF), we selected for each signature the exposures for the highest and lowest CCF, per patient. We used the same *statistical analysis* explained below, comparing the frequencies of signature activity per risk group, and per CCF.

### Homologous recombination deficiency analysis

The analysis of homologous recombination deficiency has been performed using two complementary approaches: 1) presence of somatic mutations in HRR genes, and 2) HRDetect algorithm.

#### HRR genes

The presence of mutations in HRR genes has been evaluated using the set of somatic variants described in the *Variant calling* section above, including SNVs, indels, SVs, and CNAs. We predicted the variant effect for SNV and indel calls using Ensembl VEP<sup>86</sup> (Variant Effect Predictor) v.102.0. Only SNVs and indels classified as *missense* and *stop gained* overlapping genes, were considered. For SV calls, we considered all the genes affected by: (1) SV close to gene; (2) SV within gene, defined in the *Identification of variants affecting genes and chromosome arms* section above. For CNA calls, we considered the genes overlapping amplifications or homozygous deletions. The set of HRR genes considered in this analysis correspond to 94 genes associated to the homologous recombination repair (HRR) pathway, including the 15 HRR genes from the PROfound clinical trial<sup>64</sup> (*BRCA1*, *BRCA2*, *ATM*, *BRIP1*, *BARD1*, *CDK12*, *CHEK1*, *CHEK2*, *FANCL*, *PALB2*, *PPP2R2A*, *RAD51B*, *RAD51C*, *RAD51D*, and *RAD54L*). We used the mutational status of these genes to compute the correlation with the activity of mutational signatures and complex rearrangements in our data.

#### HRDetect analysis

For this analysis we used the *signature.tools.lib* R package<sup>9</sup> (<https://github.com/Nik-Zainal-Group/signature.tools.lib>) for mutational signatures analysis, which includes the HRDetect pipeline that computes the HRDetect BRCAness probability score.<sup>51</sup> Following the best practices specified by the authors, we used the tri-nucleotide context from our previous mutational signature analysis along with the COSMIC v.3.2 reference signature catalog. We also supplied the indel, SV, and CNA calling from the *Variant calling* step described above, to compute indels classification, rearrangements, and the copy number-based score HRD-LOH. Then the method extracts six features: (1) proportion of deletions at microhomology; (2) SBS3 exposure; (3) R3 exposure; (4) SBS5 exposure; (5) HRD LOH index; (6) SBS8 exposure. Finally, the function returns the HRDetect BRCAness probability score for all the samples. We used this probability score to compute the correlation with the activity of mutational signatures and complex rearrangements in our data, along with the evaluation of differences of BRCAness probability between risk groups.

### Complex rearrangements calling and classification

To detect and reconstruct all linear and circular complex rearrangements types in neuroblastoma, we used three established complementary algorithms: JaBba<sup>17</sup> v1.0, Amplicon Architect<sup>91</sup> v.1.2, and the R package Shatterseek<sup>92</sup> v.0.5.

#### JaBba

JaBba<sup>17</sup> has been run using standard parameters and following the best practices pipeline, including the pre-processing steps using fragCounter and dryclean to correct for GC content and mappability and denoise the coverage data using a panel of normals. From



these steps, we obtained the coverage input that we use along our SV calls to run the pipeline. Then, following what has been done in previous studies,<sup>17</sup> the output from JaBba was run into gGnome v.0.1 to classify the different complex SVs into rigma, pyrgo, TIC, chromoplexy, chromothripsis, BFB, double minute (DM), and tyfonas.

### **AmpliconArchitect**

We used the copy number profiles obtained from ASCAT and the bam files of all tumor samples as input to PrepareAA<sup>91</sup> (<https://github.com/jluebeck/PrepareAA>), a wrapper function that handles the preprocessing and execution of AmpliconArchitect v.1.2.

We predicted the types of amplifications present in Amplicon Architect's<sup>91</sup> output using AmpliconClassifier<sup>16</sup> v.0.4.6 with standard parameters except for `-force = TRUE`. With this method we classified circular amplicons in ecDNA and BFB, and linear amplicons in complex non-cyclic amplicons (CnC).

### **ShatterSeek**

ShatterSeek<sup>92</sup> has been run from CNA and SV calls using standard parameters to detect chromothripsis. The filtering steps presented in the tutorial to obtain high confidence calls have been followed: at least 6 interleaved intrachromosomal SVs, 7 contiguous segments oscillating between 2 CN states, significant fragment join tests, and either significant chromosomal enrichment or exponential distribution of breakpoints test. In order to include chromothriptic candidates in our results that might affect highly amplified regions, we also called low-confidence chromothripsis having at least 6 interleaved intrachromosomal SVs and 7 contiguous segments. In these cases, a final step of visual inspection was performed.

The three complex rearrangement calling pipelines have been run using the hg19 reference genome. The consensus variant calling has been used for this analysis. Complex rearrangements detected by at least one method were considered. We evaluated each of the overlapping complex rearrangements per patient. All overlapping complex rearrangements detected in the same patient showing similar overlapping segments and junctions have been collapsed, keeping the one with most informative classification.

### **Clustered rearrangements identification**

With the aim of classifying all those clustered SVs not falling into the known complex rearrangement categories we developed a method following the clustering definition used in SV-based mutational signature analysis. Considering the consensus SV calling, we classified as clustered rearrangements the events exhibiting  $\geq 10$  SV breakpoints within a 10 Mb window of the genome. All clustered rearrangements overlapping known complex rearrangements detected in our cohort were filtered out.

### **Functional analysis of complex rearrangements**

Co-occurrence of the different complex rearrangement types was computed considering the presence of the different rearrangement classes per patient. The figure has been created using the R package ComplexUpset<sup>93</sup> v.1.2.1 (<https://github.com/krassowski/complex-upset>).

The ratio of SVs involved in complex rearrangements has been extracted establishing the ratio of the number of SVs overlapping segments or junctions of those rearrangements per patient.

For all detected complex rearrangement class in our cohort, we obtained the coordinates of each complex SV. Then, using 1 Mb windows of the genome, we plotted the density of regions affected by complex rearrangements in each chromosome across the whole human genome.

Using the coordinates information of complex rearrangements, we evaluated the overlap of the segments and junctions of these clustered events with cancer-related genes and DNA repair genes using *intersectBed* function from bedtools<sup>85</sup> v2.29.2 with parameters `-wo` (<https://bedtools.readthedocs.io/>). We performed this analysis for each complex rearrangement class separately to obtain which genes were affected by each class differentially.

### **Kataegis analysis**

We considered a cluster of kataegis when we called 6 or most consecutive SNVs with an intermutation distance of  $< 1$  kb. The same method has been used in previous studies.<sup>32</sup>

### **Mutational scenarios analysis**

The three mutational scenarios are generated using a hybrid between hierarchical and k-means unsupervised clustering methodology (*hkmeans* from the *factoextra* R package<sup>94</sup>). We set the rank/number of clusters at  $k=3$ , after evaluation using *fviz\_nbclust*, from the same package, which determines and visualize the optimal number of clusters using different methods such as gap statistics. The information fed to the clustering method was the scaled frequency of the different types of mutational signatures (SNVs, indels, CNAs, and SVs) and complex rearrangements for each single patient. No signatures were excluded. No risk group classification, mutational status of known driver genes (i.e., *ALK*, *MYCN*, *ATR*, etc.) or presence of neuroblastoma-associated chromosomal arms rearrangements (i.e., 17q, 11p, 1p, etc.) was used to compute the clustering. From the mutational signatures and complex rearrangements data, we obtained 3 different clusters which we named mutational scenarios, which when compared to the clinical risk group classification, showed a concordance of more than 80% of the samples. Scenario #1 was enriched in high-risk *MYCN*-amplified patients (20 HR MNA, 3 HR non-MNA, 1 non-HR), scenario #2 in high-risk non-*MYCN*-amplified patients (1 HR MNA, 33 HR non-MNA, 7 non-HR), and scenario #3 in non-high-risk patients (3 HR MNA, 5 HR non-MNA, 41 non-HR). From the results of the hybrid clustering analysis, we were able to extract the mean clustering distances for each feature (mutational signatures and complex



rearrangements) per scenario. Using this information, and considering a distance threshold of 0.3 to contemplate that the feature of interest has a moderate/strong association, we identified the defining features for each scenario.

## QUANTIFICATION AND STATISTICAL ANALYSIS

All comparisons between distributions in the different neuroblastoma risk groups were made using the non-parametric Wilcoxon rank-sum test. To assess if there are differences between risk groups, we used the non-parametric Kruskal-Wallis test. The relationship between different variables such as signatures, complex rearrangements, and mutated genes has been calculated using Spearman's correlation coefficient. All statistical analyses have been corrected by multiple testing when applicable using the false discovery rate (FDR) correction. All p values in the main text have been obtained using the Wilcoxon rank-sum test unless stated otherwise. The significance level has been established at  $p < 0.05$ .

Log rank tests were used for survival analysis across subgroups. To assess the clinical impact of the three defined mutational scenarios, we stratified our patients by scenario and clinical risk group classification. With this methodology we were able to evaluate the differences in overall survival between scenarios and risk groups. The subgroups for the analysis were: mutational scenario #1 ( $n = 24$ ; 20 HR MNA, 3 HR non-MNA, 1 non-HR), mutational scenario #2 ( $n = 41$ ; 1 HR MNA, 33 HR non-MNA, 7 non-HR), mutational scenario #3 ( $n = 49$ ; 3 HR MNA, 5 HR non-MNA, 41 non-HR), high-risk MYCN-amplified patients ( $n = 24$ ), high-risk non-MYCN-amplified patients ( $n = 41$ ), and non-high-risk patients ( $n = 49$ ). To assess the clinical impact of complex rearrangements in this tumor, we stratified our patients by the presence/absence of linear and circular complex SVs. The subgroups for the analysis were: Linear comp. rearrang./ecDNA<sup>-</sup> ( $n = 63$ ), Linear comp. rearrang./ecDNA<sup>-</sup> ( $n = 26$ ), Linear comp. rearrang./ecDNA<sup>+</sup> ( $n = 15$ ), and Linear comp. rearrang./ecDNA<sup>+</sup> ( $n = 10$ ).

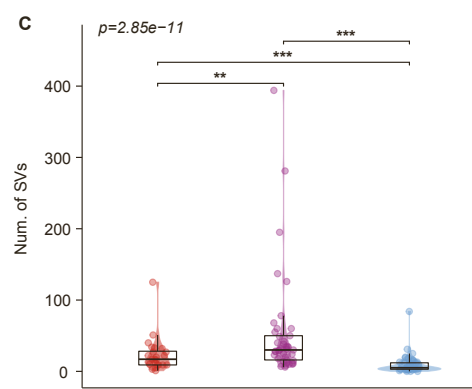
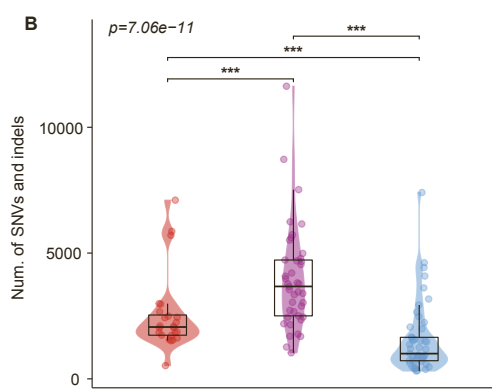
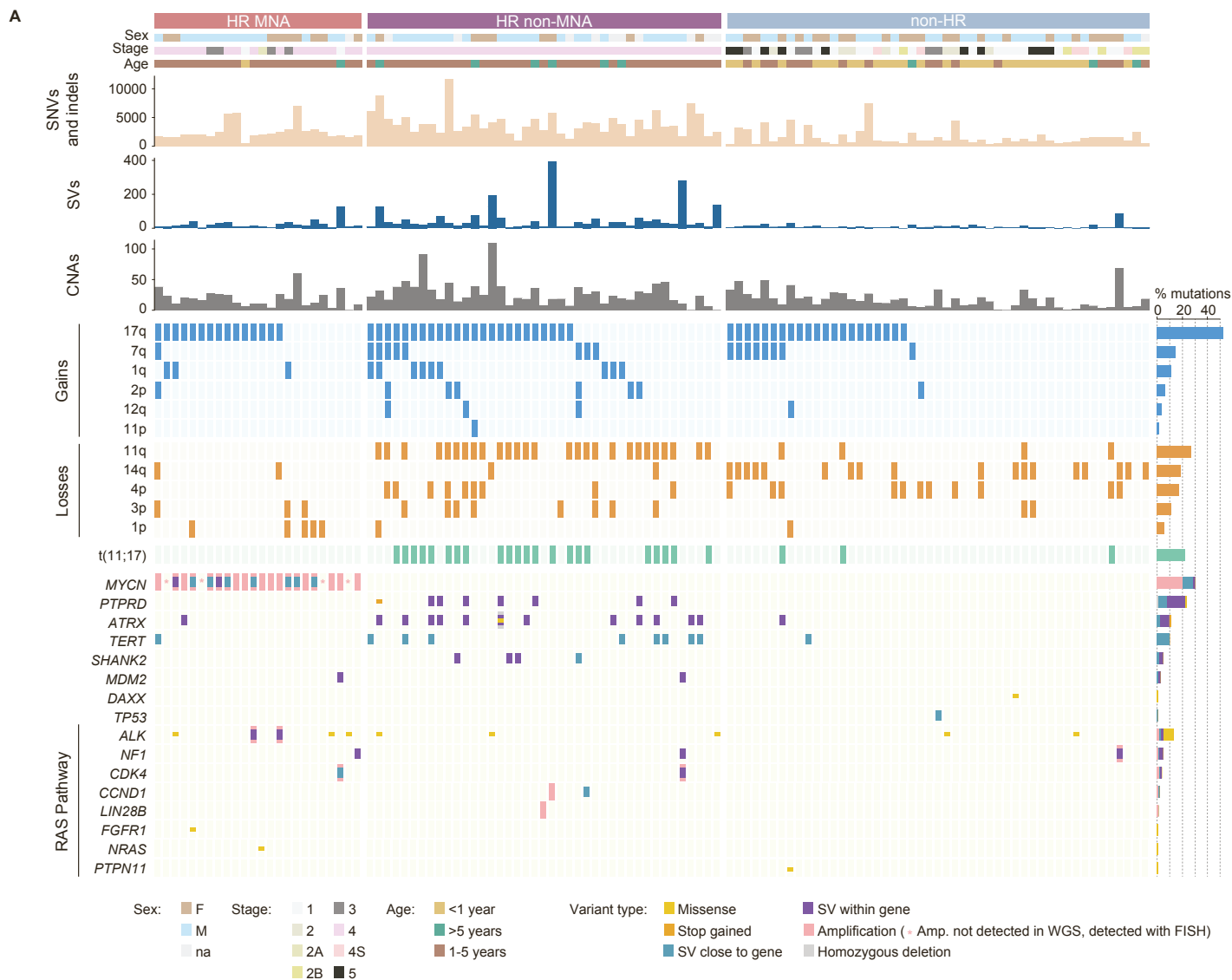
To assess the hazard risk associated to the mutational scenarios compared to neuroblastoma risk group classification, we computed two univariate Cox proportional hazards regression analysis using the *coxph* function from the *survival* R package.<sup>95</sup> Same subgroups used for the survival analysis, defined above, were used here. Hazard ratios, give the proportional clinical risk of belonging to each of the subgroups, including upper and lower 95% confidence intervals. Scenario #3 and non-high-risk patients have been used as reference in this analysis. Statistical significance was assessed through the Wald statistic value for each variable and through the Log rank test for the global statistical significance of the model.

Different additional R packages<sup>96–105</sup> (see KRT) have been used to compute and plot the results from these analyses.

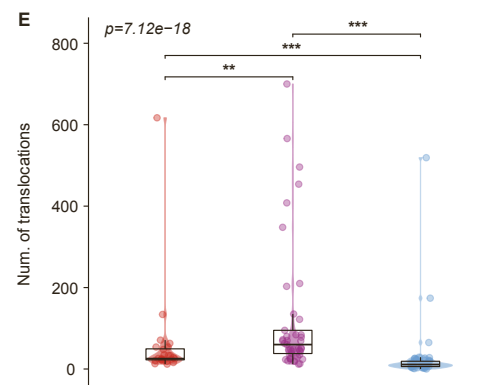
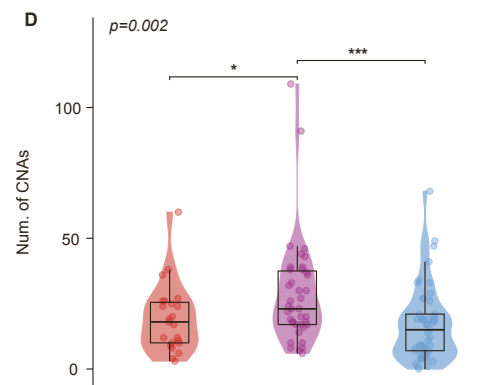
**Supplemental information**

**Mutational topography reflects  
clinical neuroblastoma heterogeneity**

**Elias Rodriguez-Fos, Mercè Planas-Fèlix, Martin Burkert, Montserrat Puiggròs, Joern Toedling, Nina Thiessen, Eric Blanc, Annabell Szymansky, Falk Hertwig, Naveed Ishaque, Dieter Beule, David Torrents, Angelika Eggert, Richard P. Koche, Roland F. Schwarz, Kerstin Haase, Johannes H. Schulte, and Anton G. Henssen**

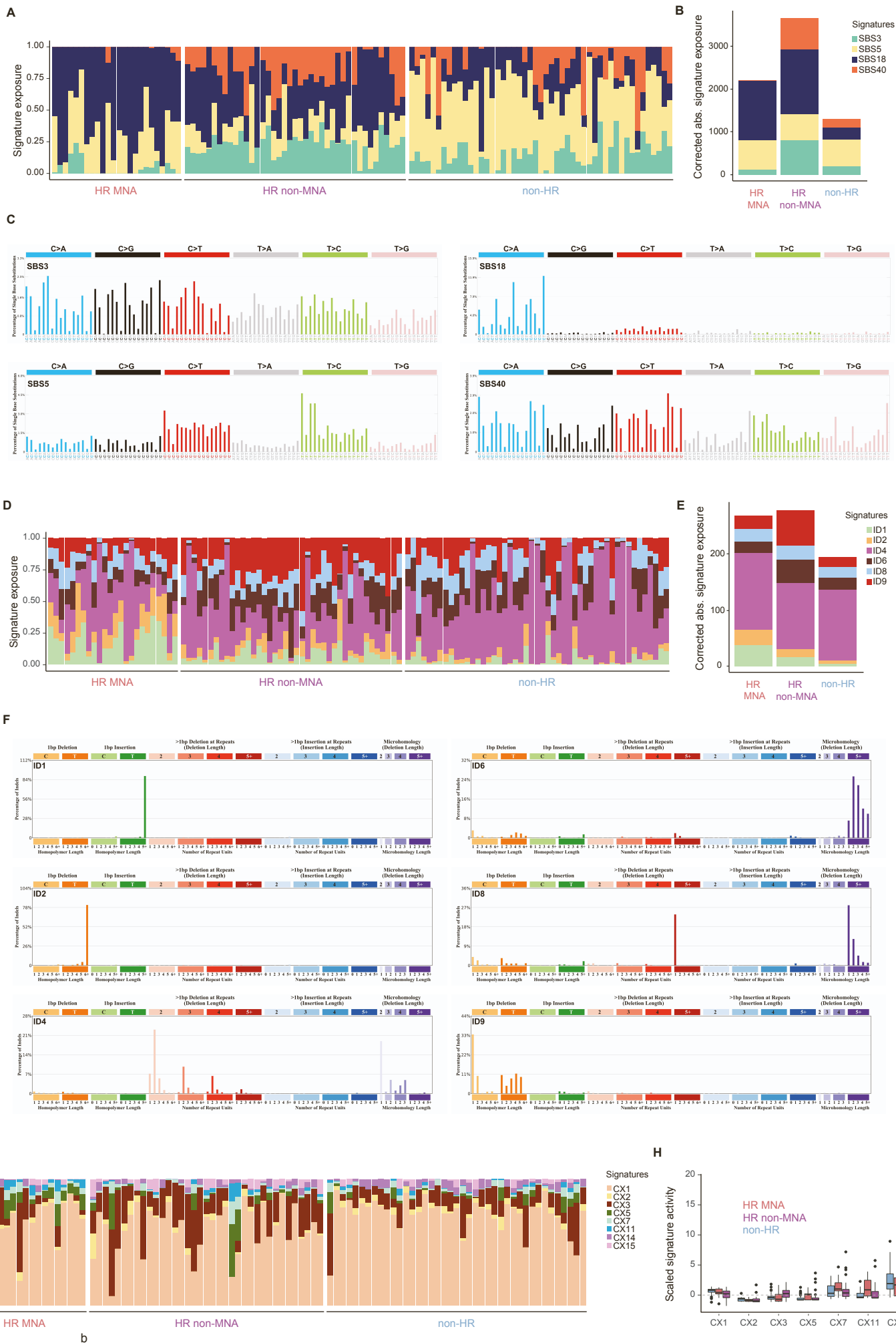


Risk groups: HR MNA HR non-MNA non-HR



**Figure S1. Landscape of somatic variants in clinically heterogenous neuroblastomas, related to Figure 1.**

(A) Top, risk group (HR MNA: high-risk *MYCN*-amplified,  $n = 24$ ; HR non-MNA: high-risk non-*MYCN*-amplified,  $n = 41$ ; non-HR: non-high-risk,  $n = 49$ ), sex, stage of the disease, and age. Next, number of SNVs and indels, SVs, and CNAs in each of the 114 neuroblastoma patients from the discovery cohort. Middle, presence of known neuroblastoma chromosomal arm gains (blue), losses (orange), and translocations between chromosome 11 and 17 (green) in each patient. Bottom, presence of mutations in known neuroblastoma driver genes, including genes of the RAS pathway in each patient. Genes are ranked based on their mutation frequency in the cohort. The variant types considered here are SNVs and indels classified as missense and stop gained variants, SVs close and within the gene, amplifications, and homozygous deletions. Color indicates the type of mutation (key at the bottom). Patients with more than one type of variant in a gene have multiple colors designated. The right histogram indicates the percentage of mutations affecting each genomic feature in the whole cohort. Variant type is indicated by color. (B-E) Comparison between distributions of different variant types in the three neuroblastoma risk groups (HR MNA, HR non-MNA, non-HR) from the discovery cohort ( $n = 114$ ). To assess if there are differences between risk groups, we used the non-parametric Kruskal-Wallis test (p-value in the upper left corner). The pairwise comparisons were done using the non-parametric Wilcoxon rank-sum test. (B) SNVs and indels, (C) SVs, (D) CNAs, and (E) translocations. Significance legend: \*: p-value < 0.1, \*\*: p-value < 0.05, \*\*\*: p-value < 0.01.



**Figure S2. SNV-, indel-, and CNA-based mutational signatures analysis, related to Figure 2.**

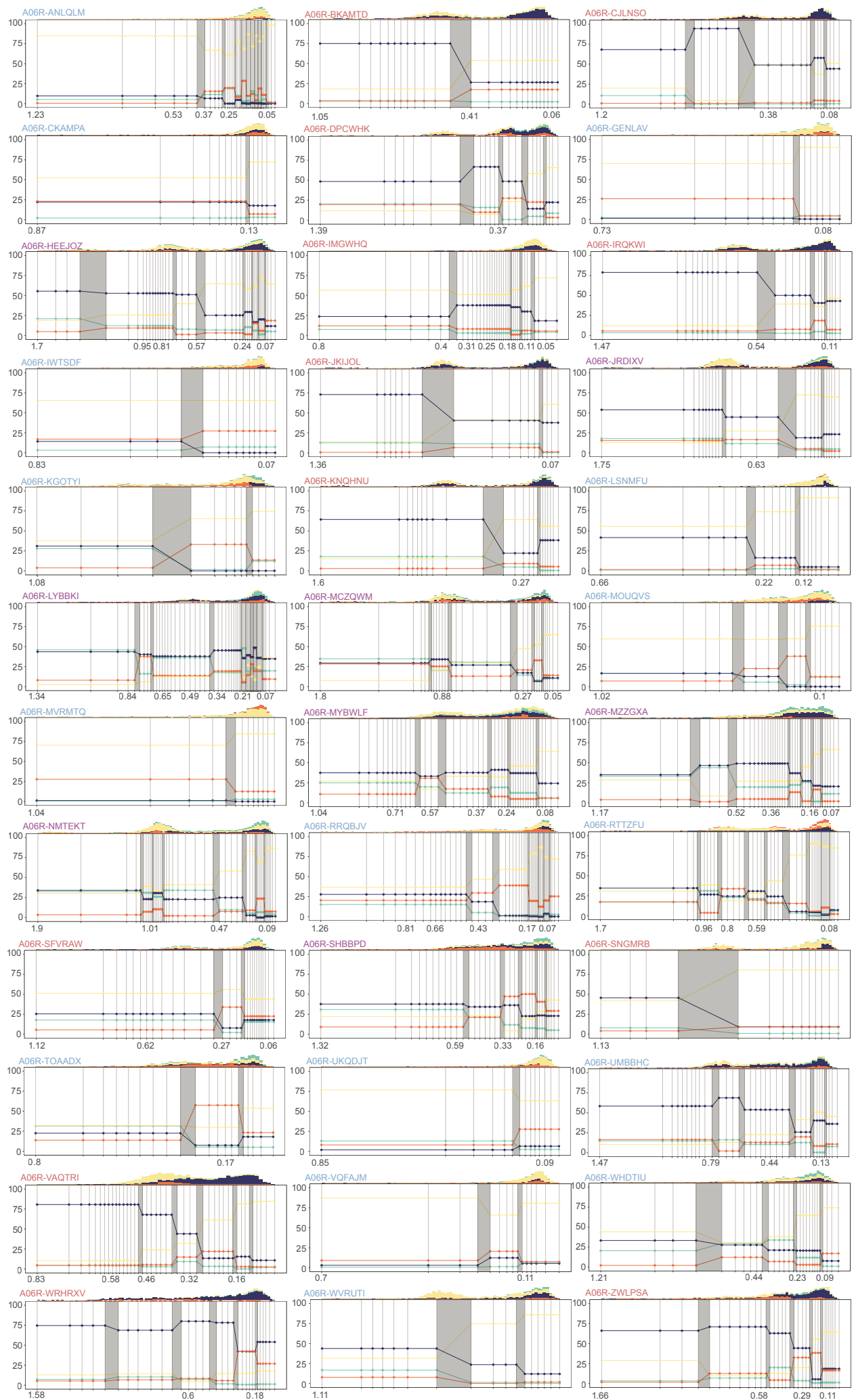
(A) Ratio of exposure of the six reference SNV-based signatures (SBS) identified in the discovery cohort for each of the 114 neuroblastoma patients. Patients have been grouped by risk (HR MNA, HR non-MNA, non-HR). Each color displays a different signature: SBS3, SBS5, SBS18, SBS40. (B) Absolute exposure of the four SNV-associated signatures (SBS) identified in the neuroblastoma discovery cohort by clinical risk group. Each color displays a different signature: SBS3, SBS5, SBS18, SBS40. (C) Mutational profiles for each of the reference SNV-based signatures detected in our cohort, reproduced from the COSMIC database (<https://cancer.sanger.ac.uk/signatures/sbs/>). (D) Ratio of exposure of the six reference indel-based signatures (ID) identified in the discovery cohort for each of the 114 neuroblastoma patients. Patients have been grouped by risk (HR MNA, HR non-MNA, non-HR). Each color displays a different signature: ID1, ID2, ID4, ID6, ID8, ID9. (E) Absolute exposure of the six indel-associated signatures (ID) identified in the neuroblastoma discovery cohort by clinical risk group. Each color displays a different signature: ID1, ID2, ID4, ID6, ID8, ID9. (F) Mutational profiles for each of the reference indel-based signatures detected in our cohort, reproduced from the COSMIC database (<https://cancer.sanger.ac.uk/signatures/id/>). (G) Ratio of exposure of the 8 CNA-associated signatures identified in the discovery cohort for each of the 114 neuroblastoma patients. Patients have been grouped by risk (HR MNA, HR non-MNA, non-HR). Each color displays a different signature: CX1, CX2, CX3, CX5, CX7, CX11, CX14, CX15. (H) Ratio of scaled signature activity the 8 CNA-associated signatures identified in our cohort for each of the 114 neuroblastoma patients. Scaling has been performed using PCAWG data.





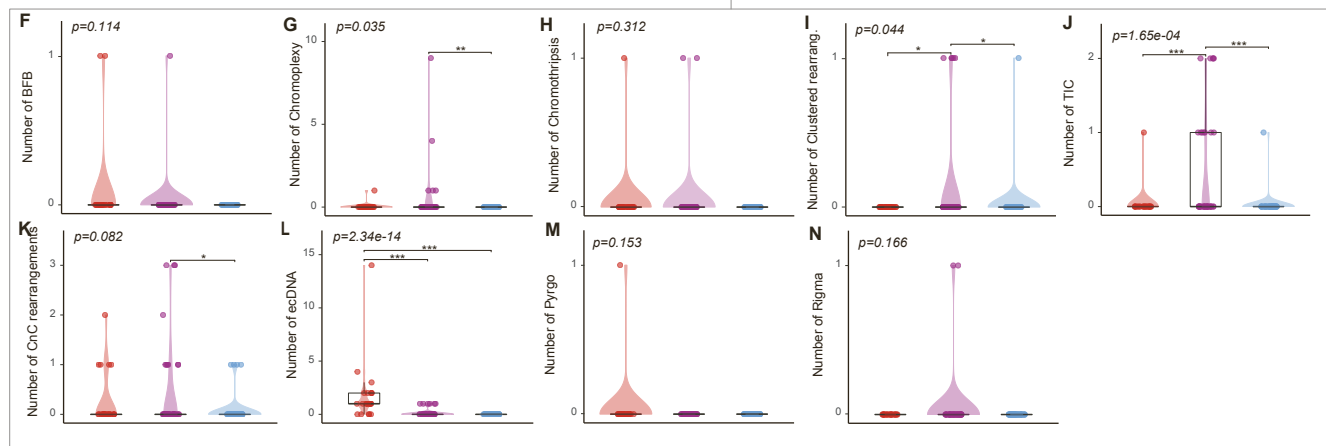
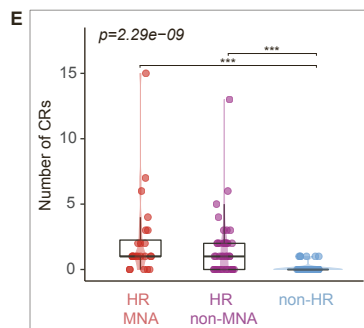
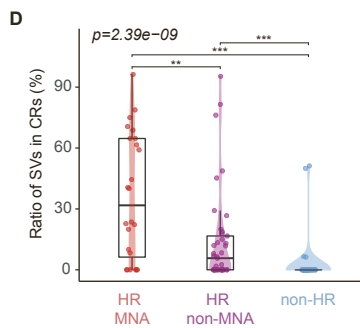
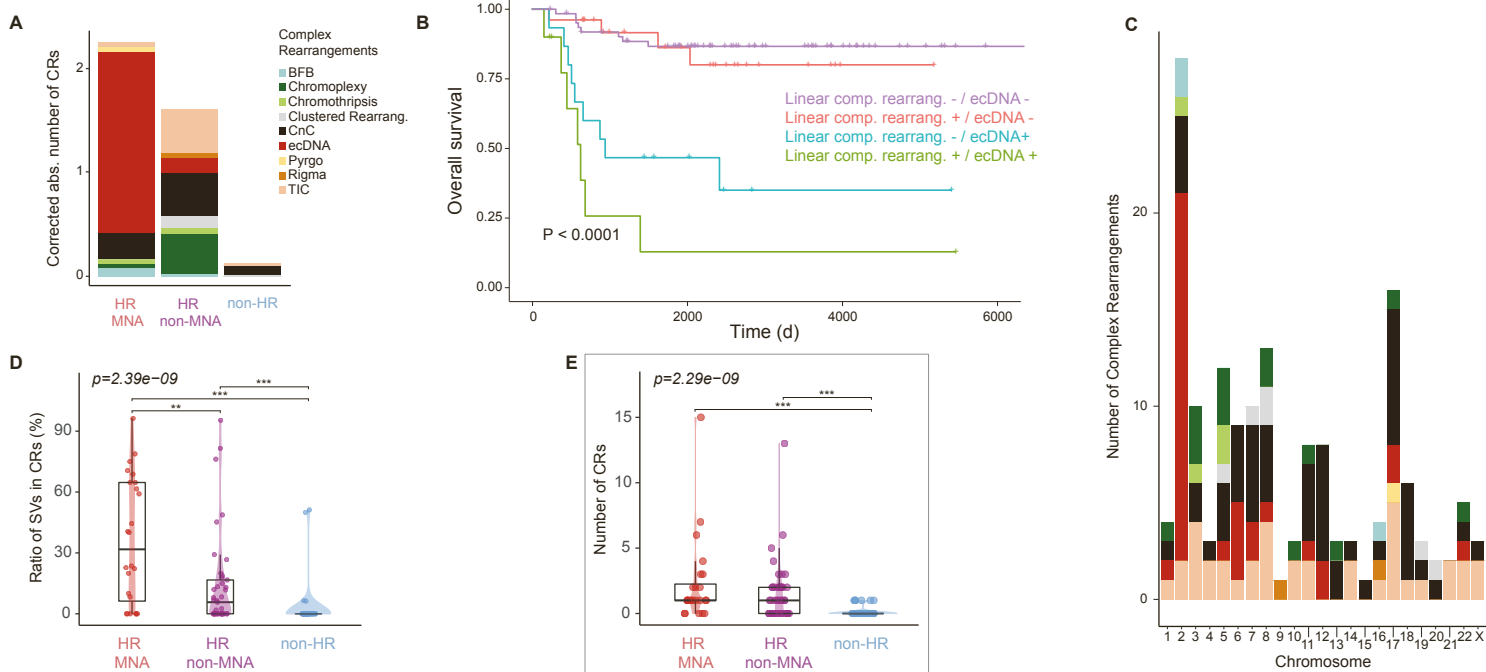
**Figure S3. SV-based mutational signatures analysis and comparison of mutational signatures from all the independent variant types in two different neuroblastoma cohorts, related to Figure 2.**

(A) Ratio of exposure of the 5 SV-based signatures identified in the discovery cohort for each of the 114 neuroblastoma patients. Patients have been grouped by risk (HR MNA, HR non-MNA, non-HR). Each color displays a different signature: SV1-SV5. (B) Absolute exposure of the 5 SV-based signatures identified in the discovery cohort for each of the 114 neuroblastoma patients. Patients have been grouped by risk (HR MNA, HR non-MNA, non-HR). Each color displays a different signature: SV1-SV5. (C) Mutational profiles for each of the five SV-based signatures detected in our cohort, including a short description of their defining characteristics. (D) Mutational profiles for reference SV-based signatures R6a and R6b, corresponding to the *de novo* identified signatures SV2 and SV3, respectively. Reproduced from the Signal database (<https://signal.mutationalsignatures.com/>). (E) Plot depicting the detection status of the different reference mutational signatures in both the discovery (n = 114) and the validation (n = 36) cohorts. Dark red corresponds to presence, light yellow to absence. (F) Dot plot showing the linear correlation ( $r^2 = 0.61$ ,  $P = 4.4 \times 10^{-5}$ ) between the frequencies of exposures of the different mutational signatures in each risk group for each cohort. Each dot represents the correlation between the frequencies in both cohorts. Colors represent risk group classification. Size of the dot represents statistical differences in the frequencies between the cohorts (non-parametric Wilcoxon rank-sum test).



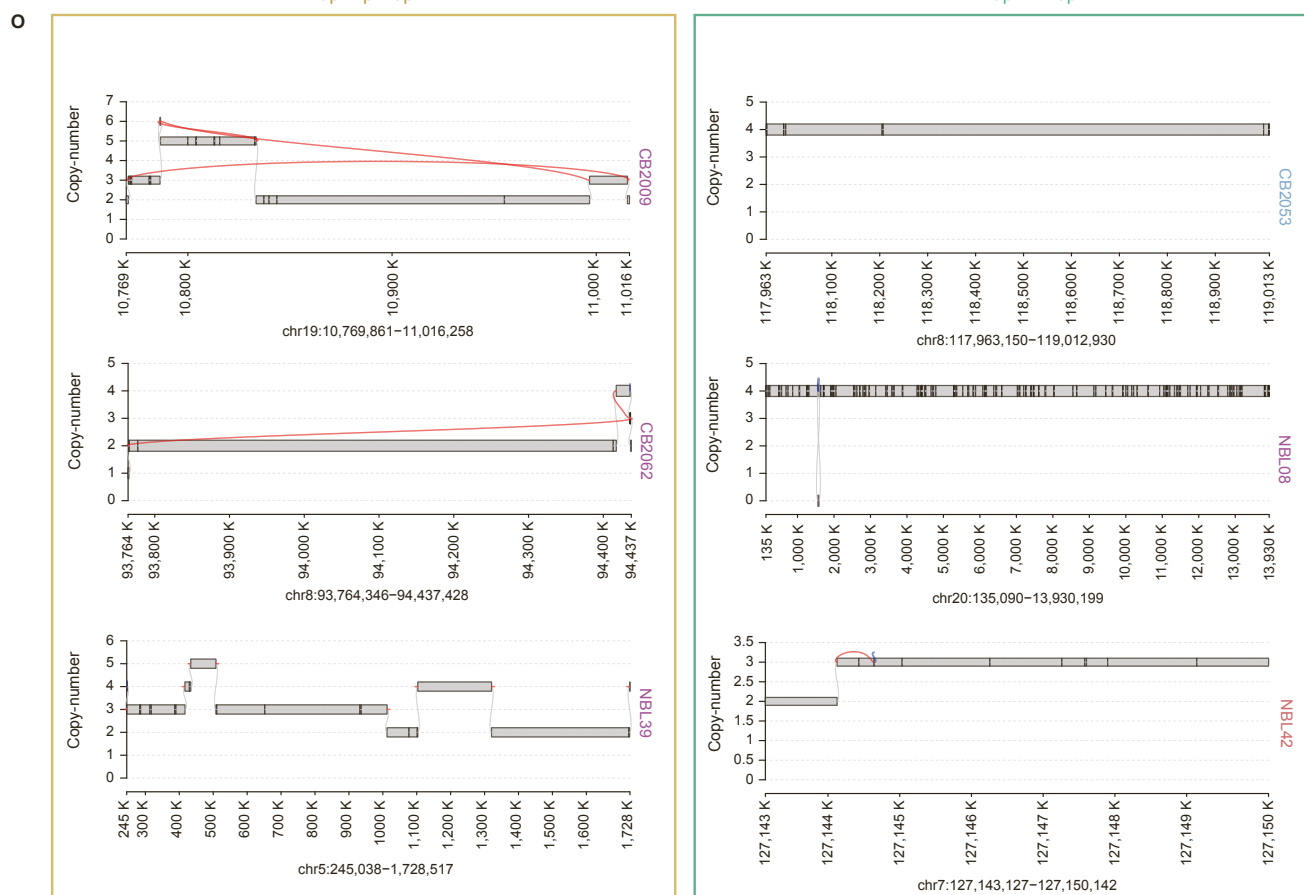
**Figure S4. Subclonal mutational signature trajectories by patient, related to Figure 2.**

Activity trajectories of the four SNV-associated signatures (SBS) per cancer cell fraction by patient from the validation cohort ( $n = 36$ ). For each plot, the y axis corresponds to the activity of mutational signatures (in %), and the x axis, the decreasing cancer cell fraction (CCF). Vertical grey lines indicate timepoints, the density of which increases with increasing mutation VAF density. Top histograms show different clusters of subclonal populations, with the frequency of each signature.



Dup-Trp-Dup

Dup-invDup





**Figure S5. General characteristics of complex rearrangements, and reconstruction of the unclassified *clustered rearrangements* identified in the discovery cohort, related to Figures 4 and 5.**

(A) Absolute count of the 9 complex rearrangement types identified in the discovery cohort for each of the 114 neuroblastoma patients. Patients have been grouped by risk (HR MNA, HR non-MNA, non-HR). Counts are corrected by the number of patients per risk group. Each color displays a complex rearrangement class. (B) Kaplan-Meier survival curve showing the overall survival of patients presenting ecDNA and linear complex rearrangements (green), only ecDNA (blue), only linear complex rearrangements (red) or no complex rearrangements (purple) ( $P < 1 \times 10^{-4}$  by Log-rank test). (C) Distribution of the number of different complex rearrangement classes per chromosome. (D) Box plot comparing the ratio in percentage of structural variants (SVs) involved in complex rearrangements for each of the three neuroblastoma risk groups (HR MNA, HR non-MNA, non-HR). (E)-(N) Box plots comparing the number of complex rearrangements across the three neuroblastoma risk groups (HR MNA, HR non-MNA, non-HR). (E) for all complex rearrangement types, (F) BFB, (G) chromoplexy, (H) chromothripsis, (I) *clustered rearrangements*, (J) TIC, (K) complex non-cyclic, (L) ecDNA, (M) pyrgo, and (N) rigma. Each dot represents a patient. To assess if there are differences between risk groups, we used the non-parametric Kruskal-Wallis test (p-value in the upper left corner). The pairwise comparisons were done using the non-parametric Wilcoxon rank-sum test. Significance legend: \*: p-value < 0.1, \*\*: p-value < 0.05, \*\*\*: p-value < 0.01. (O) Reconstruction of the *clustered rearrangements* detected in the neuroblastoma discovery cohort, based on the copy number of the segments and the junctions connecting these segments. Each plot corresponds to a different patient. The plots are classified in two columns by the complex SV type: duplication-inverted-triplication-duplication (Dup-Trp-Dup) and duplications linked by inverted segments (Dup-invDup).