

# Molecular characterization reveals subclasses of 1q gain in intermediate-risk Wilms tumors

Ianthe A.E.M. van Belzen<sup>1</sup>, Marc van Tuil<sup>1</sup>, Shashi Badloe<sup>1</sup>, Eric Strengman<sup>1</sup>, Alex Janse<sup>1</sup>, Eugène T.P. Verwiel<sup>1</sup>, Douwe F.M. van der Leest<sup>1</sup>, Sam de Vos<sup>1</sup>, John Baker-Hernandez<sup>1</sup>, Alissa Groenendijk<sup>1</sup>, Ronald de Krijger<sup>1</sup>, Hindrik H.D. Kerstens<sup>1</sup>, Jarno Drost<sup>1,2</sup>, Marry M. van den Heuvel-Eibrink<sup>1,3</sup>, Bastiaan B.J. Tops<sup>1</sup>, Frank C.P. Holstege<sup>1</sup>, Patrick Kemmeren<sup>1,4\*</sup>, Jayne Y. Hehir-Kwa<sup>1,\*5</sup>

<sup>1</sup>Princess Máxima Center for Pediatric Oncology, Utrecht, The Netherlands

<sup>2</sup>Oncode Institute, Utrecht, The Netherlands

<sup>3</sup>UMCU-Wilhelmina Children's Hospital - Child Health, Utrecht, The Netherlands

<sup>4</sup>Center for Molecular Medicine, UMC Utrecht and Utrecht University, Utrecht, The Netherlands

<sup>5</sup>Lead contact

Corresponding authors:

Patrick Kemmeren

p.kemmeren@prinsesmaximacentrum.nl

Jayne Hehir-Kwa

j.y.hehirkwa@prinsesmaximacentrum.nl

## Summary

Chromosomal alterations have recurrently been identified in Wilms tumors (WTs) and some are associated with poor prognosis. Gain of 1q (1q+) is of special interest given its high prevalence and is currently actively studied for its prognostic value. However, the underlying mutational mechanisms and functional effects remain unknown.

For 30 primary WTs, we integrated somatic SNVs, CNs and SVs with expression data and distinguished four clusters characterized by affected biological processes: muscle differentiation, immune system, kidney development and proliferation. We identified 1q+ in eight tumors that differ in mutational mechanisms, subsequent rearrangements and genomic contexts. 1q+ tumors were present in all four expression clusters and individual tumors overexpress different genes on 1q. Through integrating CNs, SVs and gene expression, we identified subgroups of 1q+ tumors reflecting differences in the functional effect of 1q gain, indicating that expression data is likely needed for further risk stratification of 1q+ WTs.

## Keywords

Pediatric cancer, Wilms tumor, structural variation, chromosomal alterations, 1q gain, WGS, RNA-seq, cancer genomics.

## Introduction

Structural variants (SVs) and other large chromosomal alterations are often identified in pediatric cancer genomes, but understanding their role in disease etiology is challenging. Gain of chromosome 1q (1q+) is of particular interest given its prevalence in embryonal tumors and associations with poor prognosis<sup>1</sup>. Also in Wilms tumors (WTs), the most prevalent kidney tumor during childhood, 1q gain has been recurrently identified and suggested as biomarker for risk stratification<sup>2-4</sup>. WTs are thought to arise from disrupted embryonic kidney development and typically consist of three histological components (stromal, epithelial and blastemal) in varying proportions and degrees of differentiation<sup>5,6</sup>. In addition to this phenotypic heterogeneity, somatic mutations have been reported in more than 40 genes, as well as chromosomal gains, losses and loss of heterozygosity (LOH) in a subset of WTs. Recent sequencing efforts have identified mutations beyond the classical Wnt pathway activating mutations (*WT1*, *CTNNB1* and *WTX*), affecting genes involved in developmental processes such as miRNA processing genes (*DICER*, *DROSHA*), kidney development transcription factors (*SIX1/2*) and genes involved in epigenetic remodeling or histone modification (*BCOR*, *CREBBP*)<sup>7</sup>. However, most of these single nucleotide variants (SNVs) occur with a low prevalence (<10%), and WTs individually carry few SNVs while some have no known pathogenic variants at all. As a result, it remains challenging to identify which alterations and biological processes are clinically relevant and contribute to tumorigenesis, suggesting a role for structural variants (SVs) and copy number alterations (CNs).

Risk stratification of WTs is mostly based on a combination of clinical and histological features of tumor cells remaining after preoperative chemotherapy as per SIOP-RTSG guidelines<sup>5,6</sup>. Fully necrotic tumors are stratified as low risk and tumors with blastemal predominance or diffuse anaplasia as high-risk. All other histological subtypes (stromal, epithelial, mixed/triphasic and regressive) are regarded as intermediate risk. The only

molecular event currently used in clinical practice for risk stratification is combined 1p LOH and 16q LOH, which is associated with poor prognosis but only occurs in ~10% of relapsed tumors<sup>3,8</sup>. Since a large number of relapses occur outside of the currently defined high-risk group, using molecular markers to further stratify the largest, intermediate-risk group for more intensive treatment would likely contribute to reducing the relapse rate<sup>5,6</sup>.

Recurrent chromosomal alterations are promising biomarkers and can help to gain insight into tumor biology, but a deeper understanding of their functional effects and underlying mutational mechanisms is needed to improve their clinical utility. Gain of chromosome 1q is highly prevalent in primary WTs (27-45%) and relapsed tumors (up to 75%) and evidence is accumulating for its association with poor prognosis<sup>2-4,9</sup>. However, it has been difficult to establish its prognostic value and the role of individual genes has not yet been elucidated. First, 1q+ can both occur as an early event in tumor development or acquired later, requiring up to three samples to detect its subclonal presence<sup>10</sup>. Second, the genomic context seems to matter, for example Haruta *et al.* reported that 12+ is associated with favorable outcomes also in the presence of 1q+<sup>11</sup>, but Gadd *et al.* recently reported the acquisition of 12+ in WT relapses<sup>9</sup>. Finally, 1q+ and 1p- can be the result of a single event (e.g. isochromosome or translocation) making it difficult to assess the prognostic value of 1q+ independently from 1p/16q LOH. These examples highlight that 1q+ tumors are a heterogeneous group, limiting the prognostic value of 1q+ without further exploration of its genomic context and functional effects.

Here, we investigated the genomic landscape and functional effects of somatic CN alterations and SVs in Wilms tumor patients using WGS data integrated with RNA-seq data. By focussing on primary diagnostic samples of intermediate and high-risk tumors we identified subgroups within the 1q+ patient population. Using genome-wide CN profiles, we identified three groups of tumors with different degrees of genomic instability: tumors with co-occurring 1p-/1q+, with multiple chromosomal gains and CN neutral tumors. Orthogonal

to this, gene expression data grouped tumors with different somatic alterations in four clusters indicating similar functional effects in terms of biological processes affected. Intermediate-risk tumors split into three clusters: muscle differentiation, immune system, and early kidney development whilst all high-risk tumors clustered together. Tumors with 1q+ were present in all gene expression clusters demonstrating the differences in the functional effect of 1q gain. Furthermore, we identified distinctly different 1q+ mutational mechanisms and subsequent rearrangements among individual tumors. Our results show that combined analysis of CNs, SVs and gene expression data can identify subgroups of 1q+ tumors that converge on biological processes. This may therefore be promising for risk stratification.

## Results

### **Copy number profiles cluster tumors into three groups reflecting different degrees of genomic instability**

To study the genomic landscape of Wilms tumor, we analyzed WGS data of 30 paired tumor-normal kidney tissue samples acquired after preoperative chemotherapy (Table S1). In this dataset we identified SVs, CNs, SNVs and indels. Matching gene expression was quantified using RNA-seq for a subset of 22 patients. Overall, WTs carry few mutations with a median of 6 somatic non-synonymous coding SNVs/indels (0.14/Mbp) and 5 SVs (Figure S1a) and display genetic heterogeneity reflected in few recurrent alterations across cancer-related genes (Figure 1a). *CTNNB1*, *WT1* and *AMER1* (*WTX*) are the most often recurrently mutated genes, in keeping with previous studies<sup>7</sup>. Most prevalent chromosomal alterations are gain of chromosome 1q (1q+), 11p loss of heterozygosity (LOH), 8+ and 12+ which occur in different genomic contexts (Figure 1b). WTs are not only genetically heterogeneous, but also several histopathological subgroups can be distinguished.

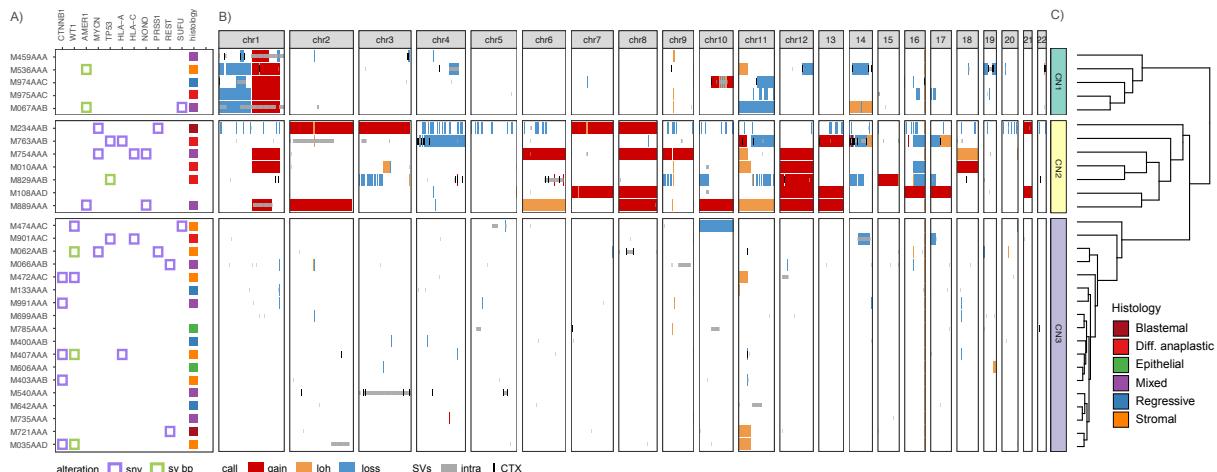
Within the cohort, seven patients had been stratified as high-risk and the remainder as intermediate-risk (stage I-IV, n=21; stage V, n=2), according to SIOP-RTSG definitions<sup>5,6</sup>. The high-risk blastemal and diffuse anaplastic tumors have more unstable genomes than intermediate-risk tumors, they carry more chromosome (arm) gains and losses (mean 3.9 vs 1.2) and overall have a larger fraction of their genome altered by CN alterations (mean 15% vs 5%) (Figure S1a, Table S2). SNV/indel burden correlates with age ( $r=0.55$  Pearson correlation) and clock-like signature (SBS40,  $r=0.66$ ) (Figure S1b,c). Furthermore, we identified platinum-associated signature (SBS31) in one tumor (M536AAA, stage V disease), and this sample was indeed the only one obtained after carboplatin treatment (Figure S1d). In contrast, there was no correlation between SV burden and histological risk group or age.

To investigate patterns of genomic instability, we clustered tumors based on their genome-wide CN profiles, resulting in three groups: tumors with co-occurring 1p-/1q+ (CN1), genetically unstable tumors carrying multiple chromosomal gains (CN2) and copy number neutral tumors (CN3) (Figure 1b,c). Comparing the CN clusters to histological classification shows that most epithelial and stromal WTs are copy number neutral and occur in CN3, which contains all tumors with mutations in both *CTNNB1* and *WT1*. Concurrent *CTNNB1/WT1* mutation was mutually exclusive to copy number gains and losses, but does co-occur with 11p LOH. High-risk tumors tend to be more genetically unstable (CN1/CN2) as previously observed<sup>12</sup>, although they occur in all three CN clusters. Integration with SV breakpoints suggests the CN clusters reflect differences in underlying mutational mechanisms (Figure 1b). Most full chromosome (arm) gains in CN2 tumors lack underlying SVs and likely result from mitotic errors, whereas segmental CN changes with nearby SV breakpoints were observed in both CN1 and CN2 tumors. In conclusion, we identified three CN patterns within the cohort and identified 1q gain either co-occurring with 1p loss (CN1) or multiple other chromosomal gains (CN2).

Next, we investigated whether genes located in regions that are gained or lost in three or more tumors show changes in gene expression levels. Comparing gene expression of tumors with these recurrent CN alterations to CN neutral tumors resulted in a set of 756 differentially expressed genes (DEG) ( $>10\%$  fold change,  $|l2fc| > +/- 0.138$ , false discovery rate (FDR)  $<0.2$ ) (Table S3). This set also contains WT genes of interest such as *AMER1* ( $-0.76 l2fc$ , three tumors) and *WT1* ( $-0.87 l2fc$ , four tumors) (Figure S2). Overall, only 9% of the recurrently gained and 10% of recurrently lost genes showed significant differences in their expression relative to CN neutral tumors. This uncoupling of gene expression and copy number has previously been related to additional regulatory mechanisms (e.g. epigenetic changes) or differences in underlying mutational mechanisms<sup>13</sup> and could reflect differences in functional consequences of CN alterations.

## Figure 1. Copy number profiles cluster tumors into three groups reflecting different degrees of genomic instability

Unsupervised analysis of copy number (CN) data identified three clusters of tumors with distinct genome-wide CN patterns. Tumors that have co-occurring 1p loss and 1q gain (CN1), multiple chromosomal gains (CN2) or that are copy number neutral (CN3). **A)** Oncoplot with cancer genes recurrently altered by SNVs (purple) or disrupted by SV breakpoints (green). Genes are ordered by the number of tumors they are mutated in. Tumors are annotated by their histological subtype. **B)** Genome-wide CN profiles with gains (red), losses (blue) and copy number neutral loss of heterozygosity (gold), overlaid by SVs with translocations (black lines) and intrachromosomal variants (gray). **C)** Dendrogram resulting from hierarchical clustering based on Euclidean distance of the CN profiles.



## Different somatic alterations converge on four gene expression clusters

To relate the heterogeneity in genomic alterations to affected downstream processes and pathways, we explored patterns in gene expression as a proxy for functional effect. Unsupervised analysis of matched RNA-seq data resulted in the identification of four

expression profiles (sets of genes with similar expression patterns) (Figure 2a) and corresponding clusters of tumors (Figure S3). Each expression profile shows enrichment for distinct biological processes ( $qvalue < 0.05$ ) (Figure 2b, Table S4). All high-risk tumors are assigned to cluster EX4 characterized by high expression of genes involved in mitosis or proliferation. The other three expression clusters consist of intermediate-risk tumors of various histological subtypes associated with muscle differentiation (EX1), immune system (EX2) or early kidney development (EX3). Next, we further characterized these expression clusters to investigate whether they can provide insights into the heterogeneous intermediate-risk group.

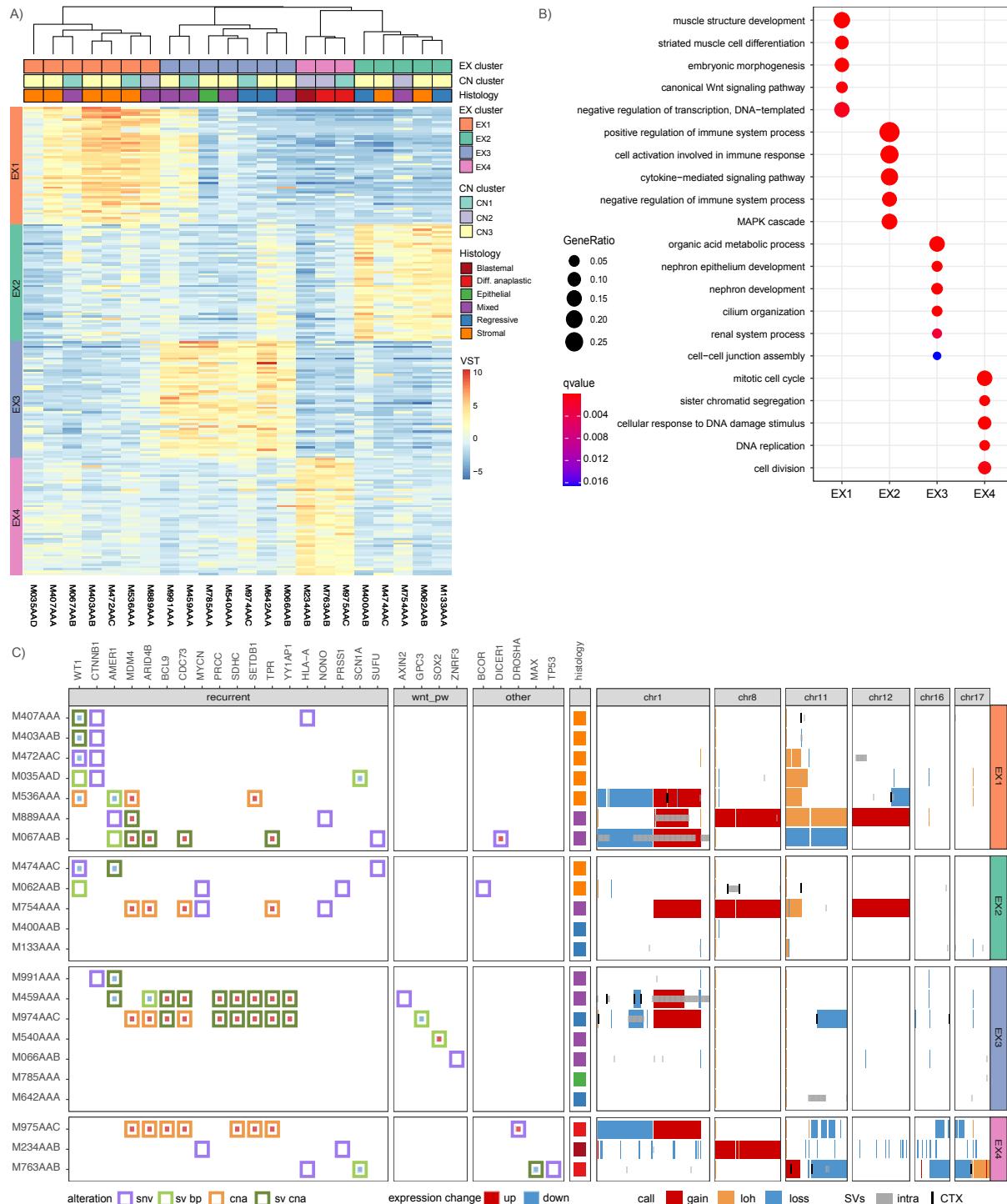
Biological processes reflecting aberrant kidney development are enriched in expression profiles from both clusters EX1 and EX3. Cluster EX1 ( $n=7$ ) is characterized by increased expression of genes related to muscle cell differentiation, canonical Wnt signaling and mesenchymal stem cells and consists of tumors with stromal or mixed histology. In keeping with the expression pattern indicating Wnt pathway activation, this cluster contains all four tumors with co-occurring *CTNNB1/WT1* mutations as well as three tumors without these characteristic mutations that instead carry a 1q gain (Figure 2c). Also, 11p LOH or loss is most prevalent in EX1 (five out of seven tumors) and absent in EX3 or EX4 tumors. Cluster EX3 ( $n=7$ ) is enriched for terms related to early kidney development, cilia organization and other genes related to epithelium or apical membrane. The only tumor with epithelial histology is present in EX3 together with mixed and regressive tumors. Of note, three tumors show significantly reduced *MIRLET7A* expression (M459AAA, M974AAC, M066AAB), suggestive of a differentiation block leading to preservation of the progenitor state<sup>14</sup>. Some EX3 tumors carry mutations in Wnt pathway genes, which suggests that Wnt pathway activation could be present in both EX1 and EX3 tumors albeit via different genetic alterations.

In contrast, clusters EX2 and EX4 do not show enrichment of the typical kidney developmental processes. Instead, cluster EX2 (n=5) is characterized by overexpression of genes related to immune system activation but also suppression, and signaling pathways including the *MAPK*, *ERK1* and *ERK2* cascade. Expression of immune genes previously linked<sup>15</sup> to a suppressive tumor microenvironment in WTs (*TGFB1*, *IL10*) were highly expressed in all EX2 tumors whilst activating genes (*TNF*, *IL6*) show a more restricted expression pattern (Figure S4). EX2 consists of tumors carrying mutations in *MYCN* and *NONO*, that have also been related to preservation of the progenitor state<sup>14</sup>, as well as epigenetic regulators/methylation (*BCOR*, *GNAS*, *KDM6A*) (Figure S5). Tumors in EX4 (n=3) are stratified as high-risk based on histology and as expected, we identified *TP53* and *DROSHA* mutations in this group as well as deletions in 16q and 17p. Furthermore, we also identified *MYCN* mutations in EX4 tumors as well as alterations affecting other signaling pathways e.g. *JAK-STAT* and *Notch* (Figure S5). Taken together, the presence of an immunosuppressive expression signature might be characteristic of a relatively more high-risk subgroup within the current histological intermediate-risk group, but follow up analyses in a larger cohort are needed to further explore this relationship.

**Figure 2. Gene expression data groups tumors with diverse genetic alterations that affect similar biological processes**

Tumors were clustered in four groups (EX1-EX4) based on their resemblance to the four expression profiles identified by unsupervised analysis of the 10k most variably expressed genes (see methods). **A**) Tumors (columns) of each expression cluster show upregulation of the genes (rows) of the corresponding expression profile compared to tumors of other expression clusters. Shown are the top 50 genes of each expression profile sorted by log2 fold change (l2fc). Tumors are annotated by their histological subtype. **B**) Biological processes enriched in expression profiles (qvalue<0.05). For each expression profile, five

representative ontology terms were selected to prevent overlap within each expression cluster. **C**) Oncoplot of tumors grouped by their expression cluster with alterations (left) and CN profiles with SVs (right), similar to Figure 1. Tumors are annotated by their histological subtype as in A). The Oncoplot displays alterations affecting recurrently altered cancer genes, Wnt pathway genes and WT associated genes from Treger *et. al.* (2019)<sup>7</sup>. Alteration types included are: SNVs (purple) and SV breakpoints (light green), as well as nearby or overlapping SVs (dark green) and CNs (orange) in case the tumor also has a significant change in gene expression (nzscore  $>+/-1.98$ ). Genes are ordered by the number of tumors they are mutated in. The CNs and SVs are displayed for recurrently altered chromosomes 1, 8, 11, 12, 16 and 17, depicted as in Figure 1b. See Table S5 for gene alteration data and Figure S5 and S6 for full sized figures.



## CNs/SVs result in Wnt pathway activation

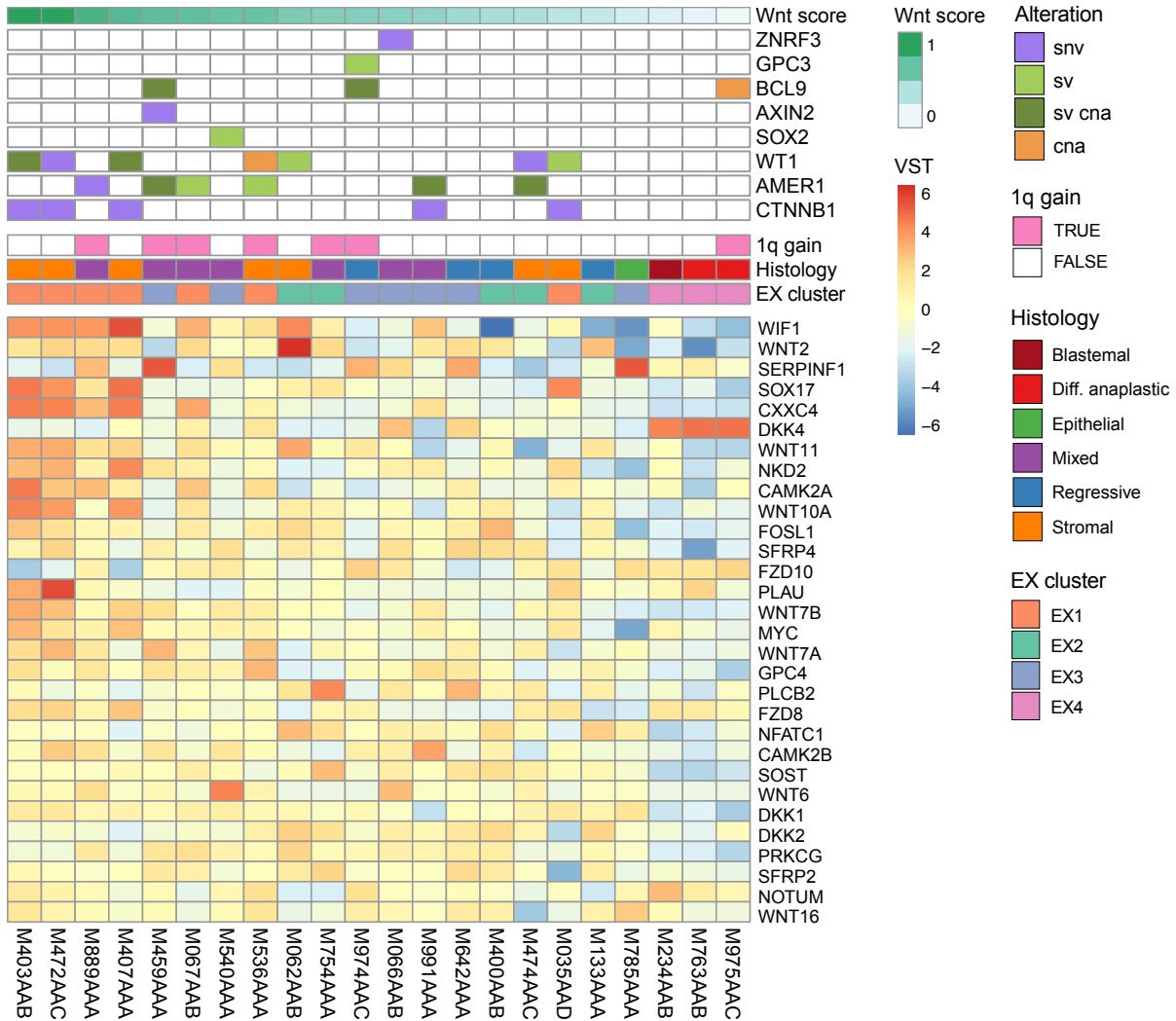
Activation of the Wnt signaling pathway is a well-established mechanism contributing to WT formation. Expression cluster EX1 was enriched for ‘canonical Wnt pathway signaling’, and indeed tends to show increased expression of downstream signaling molecules as well as Wnt antagonists relative to the rest of the cohort (Figure 3). To compare the Wnt signaling pathway activity in individual tumors, we used the mean expression of genes (Wnt score). As expected, the four *CTNNB1/WT1* mutated tumors of cluster EX1 show the strongest activation. Three tumors carry a hotspot gain of function *CTNNB1* mutation (p.S45F, p.S45P, p.T41A) and M035AAD carries p.H36P which is less common but also associated with nuclear accumulation of beta-catenin (protein product of *CTNNB1*)<sup>16</sup>. In each of these tumors, we identified multiple alterations in *WT1* indicative of disruption of both alleles irrespective of their disruption via either SNV or SV breakpoints. In one tumor (M472AAC) we identified a deleterious SNV whilst the other three tumors have SVs affecting *WT1*, often accompanied by loss of expression (Figure 2c, Table S6). Although the underlying mutations are quite different among tumors, the similarity of their expression profiles suggests that they all have a disruptive effect on *WT1*. This highlights the importance of considering SVs as potentially pathogenic events, otherwise the *WT1* disruption would have been missed in three out of four *CTNNB1/WT1* mutated tumors.

Overall, we observed a range of Wnt pathway activation throughout the cohort reflected in coordinated upregulation of signaling genes (Figure 3, Wnt score). Tumors with hotspot mutations in *CTNNB1* and disrupted *WT1* have the strongest Wnt pathway activation, followed by a subset of 1q+ tumors in EX1 and EX3. However, these 1q+ tumors do not have a *CTNNB1* mutation or evidence of biallelic *WT1* disruption. Although we identified SNV/SVs in *AMER1*, these events alone are unlikely to be sufficient for Wnt pathway activation, especially when *AMER1* expression itself is unaffected (Figure 2c, Table S6). We hypothesize that 1q gain could contribute to the Wnt pathway activation, for example via

overexpression of *BCL9*, which is significantly higher expressed in 1q+ tumors and most pronounced in the EX3 1q+ tumors. Furthermore, in five of seven EX3 tumors, we identified alterations in genes previously associated with Wnt pathway activation (*SOX2*, *AXIN2*, *ZNRF3*, *BCL9*, *GPC3*) and moderate Wnt scores. In contrast, most tumors from EX2 and EX4 lack evidence of Wnt pathway activation. Taken together, in addition to the known concurrent SNVs in *CTNNB1/WT1*, we identified SNVs, CNs and SVs in EX1 and EX3 tumors that likely result in Wnt pathway activation.

**Figure 3. Wnt pathway activation resulting from distinct somatic alterations**

Tumors (columns) display a gradient of Wnt pathway activation, ordered from left to right by the normalized mean expression of all Wnt pathway genes (Wnt score). Tumors are annotated with the alterations they carry (colors as in Figure 2c), and according to their 1q gain status, histological subtype and expression cluster membership. We observed the strongest activation in EX1 tumors with hotspot *CTNNB1* mutations co-occurring with alterations that disrupt *WT1*, followed by 1q+ tumors of EX1 and tumors of EX3 that carry other Wnt pathway alterations. The genes (rows) displayed are the top 30 most variably expressed Wnt pathway genes across the cohort (MSigDB M39669).



## 1q gain is associated with overexpression of different gene sets

Following our observations that CNs/SVs affect the same biological processes as known pathogenic SNVs, we investigated the contribution of CNs/SVs to the expression clusters. Hereto we analyzed patterns of co-occurring increases in gene expression with underlying CN gains or SV breakpoints within 1 Mbp (Figure S7). For each expression cluster, we selected the genes from the expression profile that are recurrently upregulated in two or more tumors assigned to that cluster ( $nzscore > 1.5$ ,  $n=2,498$  genes). Next, we focussed on the genes recurrently altered by CN/SV in two or more tumors within a cluster ( $n=272$ ). Combining these criteria, we identified 51 genes that are both recurrently overexpressed and

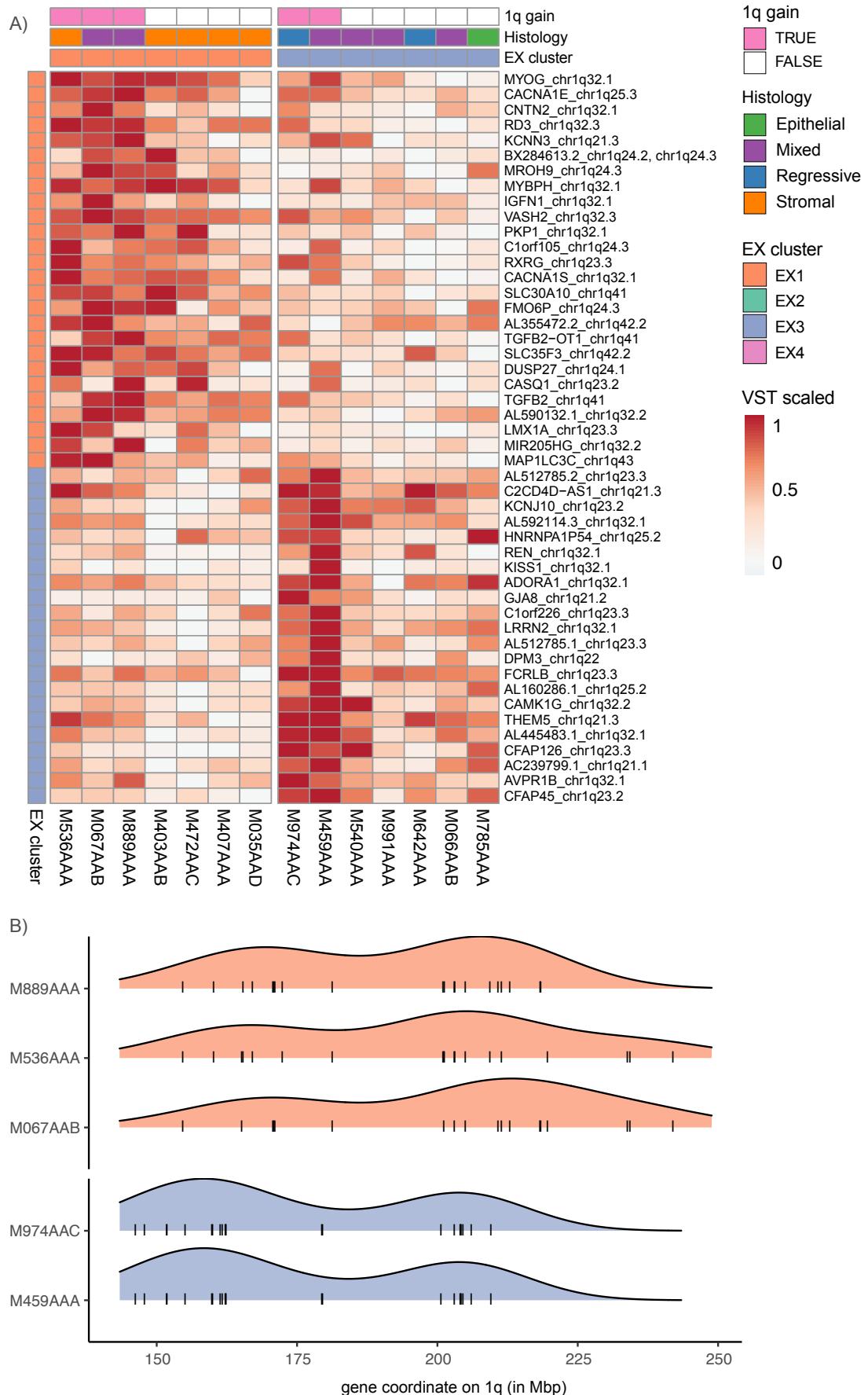
recurrently altered, representing 19% of genes with a recurrent CN/SV within the expression clusters (51/272, Table S7). Vice versa, 2% of recurrently upregulated genes have a recurrent CN/SV (51/2498), likely due to low recurrence of CNs/SVs within the expression clusters. This is in agreement with our earlier observations that tumors group differently based on copy number data and expression data (Figure 2c), e.g. 8+ and 12+ are split over the expression clusters and 1q+ is present in all. When considering alterations in individual tumors, we could identify underlying gains or nearby SVs in at least one tumor for 23% of recurrently upregulated genes (581/2498). Therefore, it seems that CNs/SVs contribute to the expression profiles, but their effects are not consistent and can differ among tumors.

Gain of 1q is recurrent within EX1 and EX3 but associated with overexpression of different gene sets that reflect the biological processes characteristic for these clusters. We found that 1q gain is the underlying alteration for 48 of the 51 genes that are both recurrently altered and recurrently overexpressed (Table S6). These genes are assigned to EX1 (n=26) or EX3 (n=22) suggesting that 1q gain can have different effects on gene expression (Figure 4a). In the 1q+ tumors of EX1, the recurrently upregulated genes are associated with GO terms "striated muscle cell differentiation" and "muscle contraction" (e.g. *MYOG*, *ACTA1*, *MYBPH*, *ACTN2*, *CACNA1S*). In contrast, 1q+ tumors of EX3 upregulated genes associated with "kidney development" and "epithelial cell differentiation" (e.g. *NPHS2*, *FLG*, *HRNR*, *REN*). Furthermore, these EX1 and EX3 genes are distributed across the chromosome arm and we did not observe preferential localization of either gene set (Figure 4b). This demonstrates that 1q gain can affect different biological processes even if the same region is altered among tumors.

#### **Figure 4. Wilms tumors with a 1q gain upregulate distinct gene sets**

1q+ tumors of EX1 and EX3 upregulate distinct gene sets corresponding to their expression cluster. Of the 51 genes that are recurrently upregulated and recurrently altered by CNs/SVs

within an expression cluster, 48 are located on 1q and assigned to either the EX1 (n=26) or EX3 (n=22) expression profiles. **A)** The 1q+ tumors of EX1 overexpress genes assigned to expression profile EX1 (orange), and vice versa for 1q+ tumors of EX3 (blue). Expression values are scaled per row to highlight relative differences among tumors (columns) rather than between genes (rows). Tumors are annotated by their 1q gain status, histological subtype and expression cluster membership. **B)** The 48 recurrently gained and overexpressed genes (black lines) are distributed across the full 1q arm and the density of these genes is similar for EX1 (orange) or EX3 (blue).



## 1q gain arises through different mutational mechanisms

Gain of chromosome 1q is the most recurrent chromosomal alteration in our cohort (27%) and of particular interest because of its suggested prognostic value<sup>2</sup>. Across all 1q+ tumors we identified 1q21.1-q32.3 (chr1:146,335,568-211,517,567) as a shared region but in different genomic contexts, such as co-occurring with 1p loss or with multiple chromosome (arm) gains (Figure 1). Using CN data, we inferred that 1q gain was clonally present in all 1q+ tumors (Table S8). And for tumors carrying multiple chromosome arm-level alterations with the same abundance, these were likely acquired through a single event. By combined analysis of CNs and SVs, we resolved profound differences in underlying mutational mechanisms or subsequent rearrangements of the eight 1q+ tumors (Figure 5a).

First, we investigated the co-occurring 1p loss and 1q gain (five tumors), since previous studies suggested a mechanistic link<sup>3</sup>. Three tumors carry a stable full arm 1p-/1q+ which suggests they were acquired at the same time through isochromosome 1q (M536AAA, M067AAB, M975AAC). For the two tumors with focal 1p losses, we identified translocations and deletions that likely mediated these losses independently from their 1q gain (M459AAA, M974AAC, Table S9). Consistent with these differences in mechanism, tumors with full 1p loss downregulate genes across the chromosome arm, whilst tumors with focal losses tend to downregulate genes nearby their SV breakpoints (Figure S8a). These results show that the breakpoints of 1p-/1q+ CN gains and losses are important and that not all co-occurring 1p-/1q+ are the result of an isochromosome.

Next, we focussed on structural variants affecting 1q (Figure 5a, Table S9). In two tumors with partial 1q+, we identified underlying SVs that could have mediated the CN gain via a breakage-fusion-bridge cycle (M459AAA) and a large inverted duplication (M889AAA). Subsequent rearrangements were identified in three tumors, where we resolved translocations and large SVs that likely occurred after the isochromosome 1q or stable arm gain (M536AAA, M974AAC, M067AAB). In contrast, no large underlying or subsequent SVs

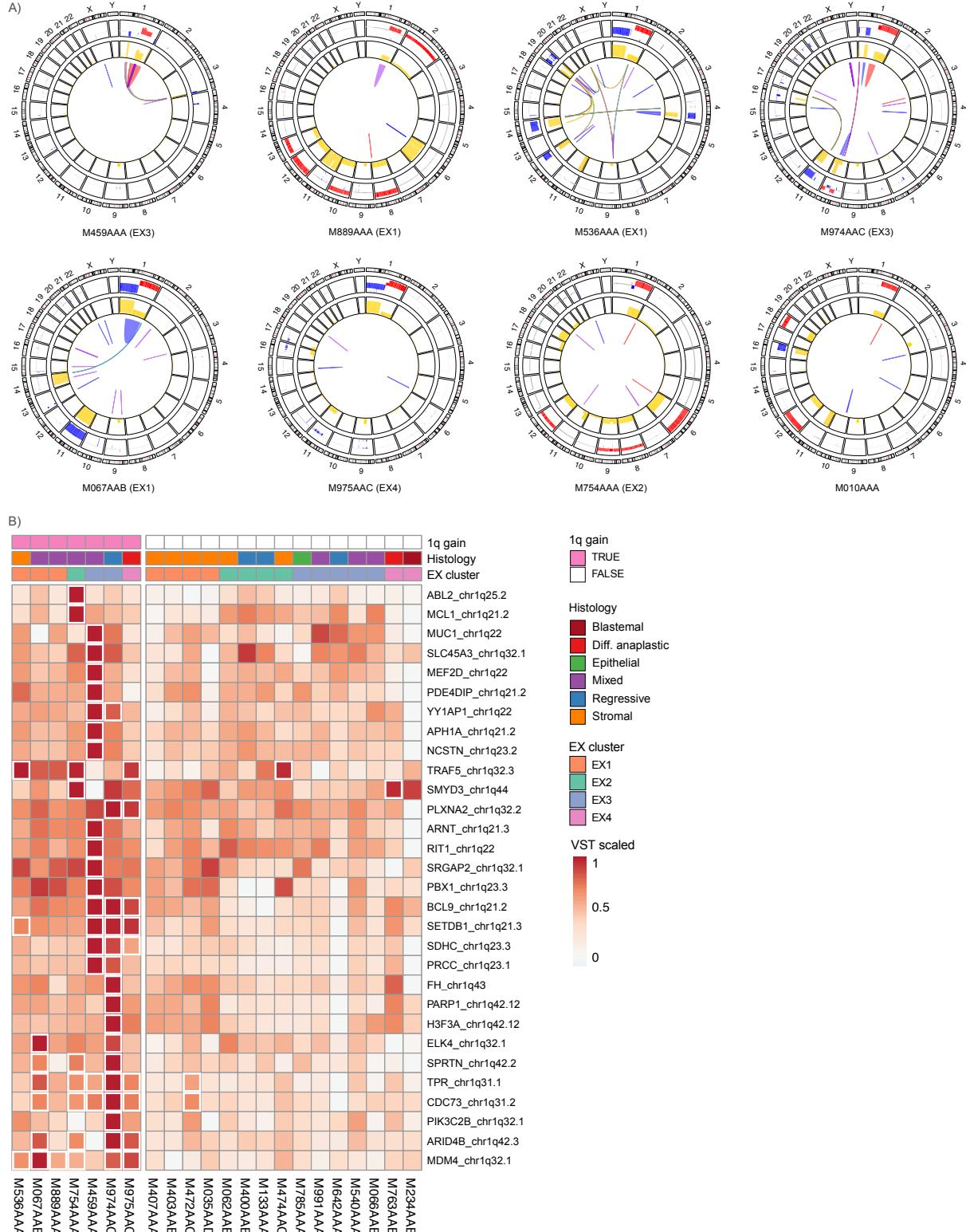
were identified in the remaining three 1q+ tumors (M975AAC, M754AAA, M010AAA). Our observations indicate that in all WTs analyzed the 1q+ was clonal regardless of the mutational mechanism resulting in 1q gain, and that diversification can take place due to subsequent SVs which is not visible from CN data alone.

Finally, to complement the mechanistic differences, we set out to explore the effect of 1q gain on expression of cancer genes in more detail. Hereto, cancer genes on 1q were selected that are overexpressed in one or more 1q+ tumors relative to CN neutral tumors (nzscore>1.98, Figure 5b, Figure S8b, Table S10). We focussed on individual tumors given our previous findings that 1q gain was associated with overexpression of distinct gene sets in EX1 and EX3 tumors (Figure 4a). We identified cancer genes recurrently overexpressed across the cohort, e.g. *MDM4* in six and *CDC73* and *TPR* in five tumors respectively (Figure 5b). The 1q+ tumors of EX3 share overexpression of seven cancer genes e.g. *BCL9* and *SETDB1*, whilst EX1 1q+ tumors share only *MDM4* overexpression and overexpress fewer cancer genes overall. Furthermore, some cancer genes located in the shared 1q gain region are overexpressed only in a single 1q+ tumor, which stresses the importance of considering gene expression profiles of individual tumors. For example, *ABL2* and *MCL1* in M754AAA, or *PARP1* in M974AAC. Taken together, we show that 1q gain has different functional effects through selective overexpression of genes. Further subclassification of 1q+ tumors is therefore crucial to elucidate the effect of 1q gain on tumor biology and this may redefine its prognostic potential.

**Figure 5. 1q gain can arise through different mechanisms and result in overexpression of specific genes**

**A)** Tumors show different patterns in copy number data and structural variants affecting 1q. For two tumors, we identified underlying SVs: M459AAA has an amplification likely caused by a breakage-fusion-bridge cycle and M889AAA has a large inverted duplication. The other

six tumors carry a stable chromosome arm-level gain. For three of those we identified SVs that likely occurred subsequently, since the SV breakpoints do not correspond to the CN segments. M536AAA has a three-way translocation connecting chr1q25.1, chr9p13.3, chr22q13.31. M974AAC has a large duplication on 1q. M067AAB has a deletion connecting chr1p32.1-chr1q42.3 and a CTX t(1,14)(q44,q32.2). For the remaining three 1q+ tumors we did not identify large SVs or translocations: M975AAC, M754AAA and M010AAA. **B**) All 1q+ tumors (pink) upregulate cancer genes located on 1q but also show large differences in which genes are significantly overexpressed (nzscore >1.98, white border). Expression values are scaled per row to display relative differences among tumors rather than between genes. See Figure S8b for locations of SVs and overexpressed genes on 1q.



## Discussion

Gain of chromosome 1q is one of most prevalent chromosomal alterations in Wilms tumor and associated with poor prognosis in multiple pediatric cancers<sup>1</sup>. 1q gain is actively studied for its potential utility as biomarker for risk stratification of WTs<sup>2-4</sup>. However, the underlying mutational mechanisms or functional effects of 1q gain on tumor biology remain unknown<sup>9</sup>. In this study, we integrated SNVs/indels, SVs and CNs with expression data to elucidate the biological and mutational processes active in primary Wilms tumors. Tumors profoundly differ in genomic stability and can be grouped into WTs with co-occurring 1p-/1q+ (CN1), multiple chromosomal gains (CN2) or CN neutral tumors (CN3). 1q gain was found in different genomic contexts (CN1 and CN2) and combined analysis with SVs showed that the underlying genomic rearrangement differs as well. Clustering based on expression data identified four groups of tumors that carry distinct genetic alterations but activate similar biological processes: muscle differentiation (EX1), immune system (EX2), early kidney development (EX3) and proliferation (EX4). Tumors with 1q+ were present in all four EX clusters and differ in which 1q genes are overexpressed. Recurrent 1q gains could be attributed to the overexpression of muscle related genes in EX1 tumors and early kidney development genes in EX3 tumors. Furthermore, individual tumors overexpressing cancer genes located on 1q (e.g. *MDM4*, *ABL1*, *MCL1*) indicate these patients might benefit from targeted therapies in the future<sup>7</sup>. These observations highlight the importance of precision oncology approaches implementing gene expression assessment because 1q gain does not always have the same effect. In conclusion, we showed that integration of WGS and RNA-seq reveals mutational and functional heterogeneity of 1q gain, which could be resolved to affected biological processes.

Wilms tumors are characterized by mutations that affect different biological processes, including Wnt pathway activation, miRNA processing, transcription elongation and epigenetic regulation<sup>7,14</sup>. Previously, Gadd *et al.* identified clusters of WTs related to disrupted kidney

developmental processes and distinguished between preservation of the progenitor state or abnormal induction<sup>14</sup>. They identified upregulation of 1q genes as one of the characteristics associated with preservation of the progenitor state, but did not further specify candidate genes. Later, they also showed that especially alterations associated with preserving the progenitor state increased in prevalence at relapse<sup>9</sup>. Our EX3 cluster also contains tumors where expression data supports preservation of the progenitor state as they have upregulation of early kidney development genes and 1q genes as well as *MIRLET7A* downregulation. In contrast, cluster EX1 shows evidence of abnormal induction (mesenchymal- to-epithelial transition) reflected in Wnt pathway activation, overexpression of muscle differentiation genes and stromal or mixed histology. A subset of EX1 tumors carry *CTNNB1* hotspot mutations together with different *WT1* alterations, showing that disruption of *WT1* through either SNVs, SVs, or losses has a similar effect. We show that SVs in *WT1* are a potential explanation for tumors lacking SNVs in this gene but behave similarly to *WT1*-mutated tumors<sup>17</sup>. Three tumors in EX1 have no aberrations in *CTNNB1/WT1* and yet show Wnt pathway activation, which could be due to overexpression of *BCL9* and/or *Wnt9b* resulting from their 1q gain<sup>18</sup>. This association between *CTNNB1/WT1* mutations and muscle differentiation has been observed previously and suggested to be induced by chemotherapy or reflect developmental arrest in the mesenchymal stem cell-like precursor lineage<sup>19,20</sup>. Notably, the transcription factor *MYOG* located on 1q can induce terminal differentiation upon overexpression<sup>21</sup> and is highly expressed in both the 1q+ and *CTNNB1/WT1*-mutated tumors of cluster EX1. Our results show that the 1q+ tumors of EX1 resemble the expression profile of *CTNNB1/WT1*-mutated tumors, which could indicate that they also have a similar cell of origin or susceptibility to chemotherapy. Therefore, the EX1 1q+ tumors displaying muscle differentiation could represent a relatively lower risk subset compared to the 1q+ tumors in EX2-EX4 associated with preservation of the progenitor state, immune suppression or that resemble the high-risk tumors.

The 1q+ tumors in our cohort share a commonly altered region (1q21.1-1q32.3) but integration of CNs and SVs revealed different mutational mechanisms underlying the 1q gain. We observed both tumors with SVs resulting in 1q+ and tumors without explanatory SV breakpoints that likely acquired 1q+ via mitotic errors, such as gain of the full 1q arm or isochromosome 1q (1p-/1q+). In our cohort, 1q+ was always clonally present and in case of multiple clonal chromosomal alterations, an early catastrophic mitotic event provides the most parsimonious explanation. Simultaneous acquisition of multiple chromosomal gains followed by stabilization is also consistent with the observed uncoupling of gene expression with copy numbers, which is a known adaptation mechanism for tolerating aneuploidy<sup>13</sup>. Previous studies report 1q+ being both an early ancestral as well as a late subclonal event<sup>10</sup>, or acquired during relapse<sup>9</sup>. Krepischi *et al* identified a specific region of interest (1q21.1-1q23.2) which was associated with relapse<sup>22</sup>. Although the region overlaps, their candidate genes do not match our findings, apart from *CHD1L* overexpression in both EX3 1q+ tumors. These differences likely result from sample selection: they focussed on the blastemal component and relapse samples whilst we characterized 1q+ tumors across histological subtypes and disease stages. Alternatively, late acquisition of 1q gain could have different functional effects that contribute to tumor progression and relapse.

Using SVs we explored suggested mechanistic links between 1q+ and 1p/16q LOH<sup>3</sup>. In three tumors, we identified 1p-/1q+ likely simultaneously arising via isochromosome 1q, and in contrast focal 1p deletions with underlying SVs in two tumors independent from their 1q events. We did not identify SVs linking chromosomes 1 and 16. Furthermore, we identified subsequent rearrangements in a subset of tumors where SVs likely occurred after acquisition of 1q+ or 1p-/1q+ through mitotic errors. Although we observed co-localization of SV breakpoints and differentially expressed genes for the SV-mediated focal 1p losses, no such relationship was apparent for 1q gain (Figure S8). Nevertheless, these findings demonstrate the great diversity in alterations affecting 1q which has not been reported before. Further molecular profiling of 1q gains in a larger cohort using whole genome

sequencing is required to fully understand the different 1q gain mechanisms and how the observed heterogeneity impacts functional effects.

Previous studies showed that 1q+ is associated with poor prognosis and more prevalent in relapse<sup>9</sup>. However, our results indicate that both the underlying mutational mechanisms and the biological processes affected by 1q gain can greatly vary among tumors. Therefore, the prognostic value of 1q+ is likely dependent on the genomic context and combination with other mutations<sup>9</sup> and there have been conflicting findings among studies regarding the role of specific chromosomes in relation to clinical outcomes<sup>9,11</sup>. Haruta *et al.* reported that 1q+ co-occurring with 12+ was associated with favorable prognosis and overexpression of candidate genes<sup>11</sup>. We also analyzed two tumors with 1q+/12+ and they each overexpress a distinct set of genes with only ~5% overlap. Both M754AAA (EX2) and M889AAA (EX1) overexpressed *MDM2*, but M889AAA also overexpressed four other candidate genes on 12+ associated with higher overall survival rate (*BRAP*, *KDM5A*, *SFSWAP*, *ZCCHC8*)<sup>11</sup>. So, whereas EX1 tumor M889AAA might have a better prognosis, this seems less likely for M754AAA despite both carrying an extra copy of chromosome 12. On the other hand, Gadd *et al.* reported not only a high prevalence of 1q+ in relapsed favorable histology WTs (75%) but also of 12+ and combined 1q+/12+ compared to primary tumors. However, they did not confirm the previously suggested link between 1p-/1q+ nor report candidate genes<sup>9</sup>. These apparently contradictory findings suggest that 1q gain by itself might not be sufficiently specific as a marker for poor prognosis. Here, we also described substantial heterogeneity in mutational mechanisms, genomic context and gene expression effects among individual 1q+ tumors. Our results suggest that gene expression needs to be incorporated along with 1q+ status for improved risk stratification.

Overall, we investigated the contribution of CNs and SVs in WT and showed that diverse genetic alterations converge on similar biological processes. Moreover, we reported distinct effects of 1q gain on gene expression, and we propose the inclusion of gene expression

data for further stratification of patients with 1q gain. While many studies include either only high-risk tumors or only favorable histology tumors, we characterized differences among mostly intermediate-risk tumors by analyzing primary samples after preoperative chemotherapy treatment. As our cohort consisted of recently collected diagnosis samples, follow up data to establish prognostic value are not mature as of yet. Given the low number of relapses, future results are needed from international studies with substantial cohorts, such as currently pursued by SIOP-RTSG and COG to determine the impact of 1q gain. In contrast to the first 1q gain studies utilizing targeted approaches for detection<sup>4</sup>, this should include expression data to distinguish subtypes of 1q+ tumors. Taken together, the functional effect and potential prognostic value of recurrent CNs/SVs including 1q gain can be improved by characterizing events at genomic level and integrating gene expression data. Genome-wide characterisation of the full spectrum of alterations in WT helps to acquire new insights into tumor biology and is essential for improving clinical utility of CNs/SVs.

## Acknowledgments

We acknowledge the willingness of all children and their families to donate their material and data for our research. We also want to thank the contributions from all pediatric oncologists, research nurses, lab technicians and biobank committee members for obtaining patient consent, sample collection and processing as well as general logistics. We acknowledge the fruitful discussions and feedback with fellow lab members as well as Ruben van Boxtel and Reek Manders. We gratefully acknowledge financial support provided by the Foundation Children Cancer Free (KiKa core funding) and Adessium Foundation. The funders had no role in the design of the study, as well as the collection, analysis and interpretation of the data or writing.

## Author contributions

Conceptualisation: I.A.E.M.vB., J.Y.H.K., P.K., F.C.P.H. and B.B.J.T.; supervision: P.K., J.Y.H.K, F.C.P.H. and B.B.J.T.; funding acquisition: P.K.; methodology: I.A.E.M.vB., J.Y.H.K. and P.K.; investigation and visualization I.A.E.M.vB.; software: I.A.E.M.vB., S.B., A.J., E.T.P.V., D.F.M.vdL., S.dV., J.B.H. and H.H.D.K.; resources: M.vT, E.S., A.G. and R.dK; data curation: I.A.E.M.vB., S.B., A.J., E.T.P.V., D.F.M.vdL., S.dV., J.B.H., H.H.D.K, A.G. and R.dK; writing - original draft: I.A.E.M.vB., J.Y.H.K. and P.K.; writing - review & editing: I.A.E.M.vB., J.Y.H.K., P.K., B.B.J.T., R.dK, A.G., J.D. and M.M.vdH.E.

## Declaration of interests

The authors declare no competing interests.

## Ethics approval and consent to participate

Informed consent has been obtained for all subjects involved in this study through the Máxima biobank informed consent procedure and corresponding protocol. The Máxima biobank protocol has been approved by the Medical Ethics Committee of the Erasmus

Medical Center in Rotterdam, The Netherlands, under reference number MEC-2016-739.

Approval for use of the subject's data within the context of this study has been granted by the Máxima biobank and data access committee, biobank request nr. PMCLAB2018.017.

## Methods

### Cohort selection and sequencing

Cohort selection was based on patients diagnosed with Wilms tumor for whom samples, subject to informed consent, were included in the Máxima biobank. Tumors were classified by a pediatric pathologist into histological subtypes and risk groups based on viable tumor tissue after preoperative chemotherapy, as per SIOP-RTSG guidelines <sup>5,6</sup>. Patients were eligible when whole-genome sequencing (WGS) data was available from matching tumor-normal samples of sufficient quality. Sequencing library preparation and data pre-processing was done via the institute's standardized pipelines and guidelines <sup>23–25</sup>. High quality WGS samples were selected requiring a minimum median coverage of 27x for normal samples and 80x for tumor samples. RNA sequencing (RNA-seq) data was included when obtained from the same biosource as the tumor WGS and had at least 30 million unique reads. In total, for 30 patients WGS data was available with a median coverage of 36x for the normal and 104x for the tumor samples of which 22 patients also had RNA-seq data representative of the tumor with a median of 96 million unique reads.

To check whether samples are representative for the tumor, the frozen tissue samples used for molecular analysis were subjected to histological verification by a pathologist subject to material availability. For 10 of 11 tumors, the histology of the sample was found to be representative of the tumor. However, not for tumor M829AAB, this diffuse anaplastic tumor sample lacked anaplasia and consisted of mostly stromal tissue with rhabdomyomatous differentiation, which was also reflected in the expression data (outlier) and therefore this RNA was excluded. Likewise, the RNA data was not included from three patients (M108AAD, M735AAA, M699AAB) where samples were taken prior to chemotherapy treatment.

### Variant calling

Somatic SNVs/indels were called using Mutect2 from GATK 4.1<sup>26</sup> and filtered to a set of high confidence variants with tumor allele fraction > 0.1, located on chromosomes 1-22 and X and not in ENCODE Blacklist poor mappability/high complexity regions<sup>27</sup>. Furthermore, we excluded variants in dbSNP unless they were also present in Cosmic. Pathogenicity was predicted by variant effect predictor (VEP) (version 104)<sup>28</sup>. Copy number (CN) alterations were identified with the GATK4 pipeline following their recommended best practices<sup>29</sup>. To assess the copy ratio and minor allele fraction levels of chromosome arms, cytobands and gene bodies, we used weighted averages of the CN segments across these genomic intervals. Relative abundance of chromosomes was ascertained by comparing copy ratio and minor allele fraction of each chromosome arm relative to the maximum levels observed in the sample.

Somatic SVs were called using Manta<sup>30</sup> (version 1.6), DELLY<sup>31</sup> (version 0.8.1) and GRIDSS<sup>32</sup> (version 2.7.2). First, we selected SVs with at least seven supporting reads and removed those that have >90% reciprocal overlap with common (>1%) population variants retrieved from the NCBI repository (nstd166<sup>33</sup>, nstd186<sup>34</sup>) and from DGV<sup>35</sup> (version 2020-02-25) accessed on 2021-03-11. Next, calls from the three tools were merged based on 50% reciprocal overlap and we required identification by at least two tools. To select variants that potentially have functional impact, we further filtered on allele fraction >0.1. Due to technical issues with running the tool, GRIDSS output was not available for two patients (M606AAA, M901AAC).

### **Mutation burden**

The number of nonsynonymous somatic SNVs and indels per megabase was calculated to compare the burden of damaging small variants between tumors. Nonsynonymous SNVs/indels were included with VEP impact moderate or high in protein coding genes on chromosomes 1-22 (excl. chrX). Likewise, the number of possible nonsynonymous

mutations (denominator) was calculated by summing all coding sequence bases in protein coding genes (~4.1 Mbp).

### **Mutational signatures**

The contribution of mutational signatures was assessed using the MutationalPatterns package (version 3.2.0)<sup>36</sup>. First, somatic SNVs were further filtered to prevent artifacts or products of selection. We required allelic depth >20 (DP>20) for both tumor and normal, removed variants with supporting reads in the normal, and removed all recurrent SNVs. Next, de novo signature extraction was performed on samples with >100 somatic SNVs (excl. M721AAA, M642AAA, M735AAA, M606AAA). As per recommended practices, we selected the number of signatures to extract based on the highest cophenetic correlation coefficient before it decreased (k=3, cophenetic = 0.98)<sup>36</sup>. These signatures were compared against the known COSMIC signatures based on cosine similarity (>85%) and refitted to the complete cohort to obtain their relative contribution to the somatic SNV profiles of all tumors.

### **Copy number clustering**

Unsupervised analysis and clustering of copy number profiles was done with CNpare (version 0.99.0)<sup>37</sup>. Segmentation files were mapped into 1 Mbp genomic bins, excluding gvar/stalk/acen regions. Next, hierarchical clustering of samples was performed based on the euclidean distance of these copy ratio profiles. The optimum number of clusters was derived from the elbow plot of within-cluster sum of squares (k=3).

### **Expression profile extraction and expression clustering**

Non-negative matrix factorisation (NMF, version 0.23) was used to derive expression profiles (or metagenes) that group genes sharing similar expression patterns throughout the cohort<sup>38</sup>.

First, read counts were normalized with variance stabilizing transformation (vst)<sup>39</sup> after which the 10k most variable genes were selected, excluding genes on chromosomes X, Y and mitochondrial genes. Following recommended practices, the optimum number of profiles to extract was based on the highest cophenetic correlation coefficient before it decreased (k=4, cophenetic = 0.98)<sup>38</sup>. Tumors were assigned to 'expression clusters' based on the highest relative contribution of these expression profiles. In addition, the expression clusters were verified with UMAP of the 10k most variable genes (Figure S3).

### **Gene set enrichment of expression profiles**

To compare the expression profiles in terms of the biological processes they reflect, we performed gene set enrichment analysis using CompareCluster<sup>40</sup>. Default settings were used to compare gene ontology terms of the top 1000 genes assigned to the expression profiles with the greatest log2 fold change. Significant terms were selected (qvalue < 0.05) for each ontology (biological process, molecular function, cellular component), relative to the 10k most variable genes corresponding to the gene set used as input for the expression clustering. The number of genes of each expression profile used for ontology assessment is respectively EX1 (780), EX2 (778), EX3 (729) and EX4 (610) because 24.5% of genes failed to map to entrez identifiers

### **Effects of copy number alterations on gene expression**

To assess the effect of copy number (CN) alterations on expression, we compared the expression values of tumors with a certain CN alteration against CN neutral tumors. We determined the copy number status of a gene by averaging the copy ratio values across the gene body and applying the default copy ratio threshold for inferring a gain or loss(+/- 0.2). We selected recurrently altered genes for which three or more tumors carried a CN alteration and assessed gain and loss events separately. The Wilcoxon rank sum test was used to assess whether the expression of CN altered genes was significantly different from CN

neutral genes. VST normalized counts were used as a measure of expression and p-values were adjusted for multiple testing with Benjamini and Hochberg, FDR <0.2). We also required a minimum log2 fold change (l2fc) corresponding to 10% of the mean or median to select potentially biologically relevant changes (l2fc>+/- 0.138).

For individual tumors, we assessed whether a gene is affected by a CN/SV by comparing the tumor's expression relative to those of CN neutral tumors. Hereto, a z-score of vst normalized counts (nzscore) was calculated:  $nzscore = (vst - vst\_mean\_neutral) / vst\_sd\_neutral$ .

### **Cancer genes**

Cancer gene datasets were retrieved from COSMIC (cancer gene census v92)<sup>41</sup>, OncoKB (accessed on 2021-04-14)<sup>42</sup> and Grobner *et al.* (2018)<sup>43</sup>. In addition, we obtained genes recurrently mutated in Wilms tumors from Table 1 from Treger *et al.* (2019)<sup>7</sup>.

### **Wnt signaling pathway analysis**

To assess Wnt signaling pathway activation, we considered the expression of the pathway genes as retrieved from MSigDB M39669 (version 7.4)<sup>44,45</sup>. To study relative differences among individual tumors, we defined the Wnt score. For this, we used the mean of vst normalized counts of all Wnt signaling pathway genes, further scaled and normalized to a range of 0-1.

### **Joint analysis of recurrent CNs/SVs and recurrent expression changes**

To investigate to what extent CNs/SVs contributed to the expression clusters, we analysed whether increases in gene expression co-occur with CN gains or nearby SV breakpoints within 1 Mbp. For example, the MYOG gene is assigned to profile EX1 and upregulated in EX1 tumors with a gain, therefore its upregulation could be CN/SV mediated. First, we

selected genes upregulated in two or more tumors assigned to that expression cluster ( $\text{nzscore} > 1.5$ ,  $n=2,498$  genes). We only consider upregulation because NMF extracts positive components (expression profiles). Next, we assessed whether these expression changes co-occur with gains or nearby SVs in individual tumors ( $n=581$ ) or in two or more tumors ( $n=51$ ). Furthermore, we compared the set of CN/SV mediated expression changes with the full set of (expressed) genes with gains or nearby SVs ( $n=272$ ) in two or more tumors.

### **Gene-level integration of SNVs, CNs and SVs**

To assess whether a gene is affected by SNV/indels, CNs or SVs, we combined evidence acquired from these alteration types as reported in Table S5 and the oncoplots of Figure 1 and 2. Each type of variant was first filtered to select the alterations that could affect gene function either directly (located in the gene body) or indirectly (nearby or spanning the gene). In case of the possible indirect effects, we also required an expression change relative to the rest of the cohort ( $\text{nzscore} >+/- 1.98$ ). For SNV/indels, we removed variants predicted as benign by SIFT/PolyPhen unless possibly pathogenic/ in COSMIC cancer gene census. For SVs we considered breakpoints inside genes to have a direct impact, and required evidence from expression data for SVs breakpoints within 1 Mbp of the gene or SVs spanning the gene. Also CNs were regarded as indirect variants requiring expression change, and the copy ratio average over the gene body was used as its CN status.

Finally, the alterations identified in a gene are summarized in the 'alteration' column: directly by SNV (snv) or SV bp (sv\_bp), indirectly by spanning SV or SV bp within 1 Mbp together with change in expression (sv\_indirect), or by CN gain or loss together with change in expression (gain or loss). These labels are simplified in 'alteration simple' where we distinguish only between snv, sv and cna. For example: downregulated gene with CN loss with SV breakpoints inside and nearby "sv\_bp sv\_indirect loss" and simplified as "sv cna".

## Visualization

Figures were generated using R and further edited with Adobe Illustrator. Packages used to generate figures: ggplot2 (version 3.3.5)<sup>46</sup> , ggpubr (version 0.4.0)<sup>47,48</sup>, pheatmap (version 1.0.12)<sup>47</sup>, rstatix <sup>49</sup>, circlize (version 0.4.13)<sup>50</sup>.

## Resource availability

The WGS and RNA-seq datasets supporting the conclusions of this article are currently being submitted to the EGA repository. Code is available through <https://github.com/princessmaximacenter/structuralvariation/> (tag v1.0).

## References

1. Byron, S.A., Hendricks, W.P.D., Nagulapally, A.B., Kraveka, J.M., Ferguson, W.S., Brown, V.I., Eslin, D.E., Mitchell, D., Cornelius, A., Roberts, W., et al. (2021). Genomic and Transcriptomic Analysis of Relapsed and Refractory Childhood Solid Tumors Reveals a Diverse Molecular Landscape and Mechanisms of Immune Evasion. *Cancer Res.* **81**, 5818–5832.
2. Gratias, E.J., Dome, J.S., Jennings, L.J., Chi, Y.Y., Tian, J., Anderson, J., Grundy, P., Mullen, E.A., Geller, J.I., Fernandez, C.V., et al. (2016). Association of Chromosome 1q Gain With Inferior Survival in Favorable-Histology Wilms Tumor: A Report From the Children's Oncology Group. *J. Clin. Oncol.* **34**.
3. Gratias, E.J., Jennings, L.J., Anderson, J.R., Dome, J.S., Grundy, P., and Perlman, E.J. (2013). Gain of 1q is associated with inferior event-free and overall survival in patients with favorable histology Wilms tumor: a report from the Children's Oncology Group. *Cancer* **119**, 3887–3894.
4. Chagtai, T., Zill, C., Dainese, L., Wegert, J., Savola, S., Popov, S., Mifsud, W., Vujanić, G., Sebire, N., Le Bouc, Y., et al. (2016). Gain of 1q As a Prognostic Biomarker in Wilms Tumors (WTs) Treated With Preoperative Chemotherapy in the International Society of Paediatric Oncology (SIOP) WT 2001 Trial: A SIOP Renal Tumours Biology Consortium Study. *J. Clin. Oncol.* **34**.
5. van den Heuvel-Eibrink, M.M., Hol, J.A., Pritchard-Jones, K., van Tinteren, H., Furtwängler, R., Verschuur, A.C., Vujanic, G.M., Leuschner, I., Brok, J., Rübe, C., et al. (2017). Position paper: Rationale for the treatment of Wilms tumour in the UMBRELLA SIOP-RTSG 2016 protocol. *Nat. Rev. Urol.* **14**.
6. Vujanić, G.M., on behalf of the International Society of Paediatric Oncology–Renal Tumour Study Group (SIOP–RTSG), Gessler, M., Ooms, A.H.A., Collini, P., Coulomb-

I'Hermine, A., D'Hooghe, E., de Krijger, R.R., Perotti, D., Pritchard-Jones, K., et al. (2018). The UMBRELLA SIOP-RTSG 2016 Wilms tumour pathology and molecular biology protocol. *Nature Reviews Urology* 15, 693–701.

7. Treger, T.D., Chowdhury, T., Pritchard-Jones, K., and Behjati, S. (2019). The genetic changes of Wilms tumour. *Nat. Rev. Nephrol.* 15, 240–251.
8. Grundy, P.E., Breslow, N.E., Li, S., Perlman, E., Bruce Beckwith, J., Ritchey, M.L., Shamberger, R.C., Haase, G.M., D'Angio, G.J., Donaldson, M., et al. (2005). Loss of Heterozygosity for Chromosomes 1p and 16q Is an Adverse Prognostic Factor in Favorable-Histology Wilms Tumor: A Report From the National Wilms Tumor Study Group. *J. Clin. Oncol.*
9. Gadd, S., Huff, V., Skol, A.D., Renfro, L.A., Fernandez, C.V., Mullen, E.A., Jones, C.D., Hoadley, K.A., Yap, K.L., Ramirez, N.C., et al. (2022). Genetic changes associated with relapse in favorable histology Wilms tumor: A Children's Oncology Group AREN03B2 study. *Cell reports. Medicine* 3.
10. Intra-Tumor Genetic Heterogeneity in Wilms Tumor: Clonal Evolution and Clinical Implications (2016). *EBioMedicine* 9, 120–129.
11. Haruta, M., Arai, Y., Okita, H., Tanaka, Y., Takimoto, T., Sugino, R.P., Yamada, Y., Kamijo, T., Oue, T., Fukuzawa, M., et al. (2019). Combined Genetic and Chromosomal Characterization of Wilms Tumors Identifies Chromosome 12 Gain as a Potential New Marker Predicting a Favorable Outcome. *Neoplasia* 21, 117.
12. Wegert, J., Ishaque, N., Vardapour, R., Geörg, C., Gu, Z., Bieg, M., Ziegler, B., Bausenwein, S., Nourkami, N., Ludwig, N., et al. (2015). Mutations in the SIX1/2 pathway and the DROSHA/DGCR8 miRNA microprocessor complex underlie high-risk blastemal type Wilms tumors. *Cancer Cell* 27, 298–311.

13. Mohanty, V., Wang, F., Mills, G.B., CTD2 Research Network, and Chen, K. (2021). Uncoupling of gene expression from copy number presents therapeutic opportunities in aneuploid cancers. *Cell Rep Med* 2, 100349.
14. Gadd, S., Huff, V., Walz, A.L., Ooms, A.H.A.G., Armstrong, A.E., Gerhard, D.S., Smith, M.A., Auvil, J.M.G., Meerzaman, D., Chen, Q.-R., et al. (2017). A Children's Oncology Group and TARGET initiative exploring the genetic landscape of Wilms tumor. *Nat. Genet.* 49, 1487–1494.
15. Systematic review of the immunological landscape of Wilms tumors (2021). *Molecular Therapy - Oncolytics* 22, 454–467.
16. Austinat, M., Dunsch, R., Wittekind, C., Tannapfel, A., Gebhardt, R., and Gaunitz, F. (2008). Correlation between beta-catenin mutations and expression of Wnt-signaling target genes in hepatocellular carcinoma. *Mol. Cancer* 7.
17. Gadd, S., Huff, V., Huang, C.-C., Ruteshouser, E.C., Dome, J.S., Grundy, P.E., Breslow, N., Jennings, L., Green, D.M., Beckwith, J.B., et al. (2012). Clinically relevant subsets identified by gene expression patterns support a revised ontogenetic model of Wilms tumor: a Children's Oncology Group Study. *Neoplasia* 14, 742–756.
18. Corbin, M., de Reyniès, A., Rickman, D.S., Berrebi, D., Boccon-Gibod, L., Cohen-Gogo, S., Fabre, M., Jaubert, F., Faussillon, M., Yilmaz, F., et al. (2009). WNT/β-catenin pathway activation in Wilms tumors: A unifying mechanism with multiple entries? *Genes, Chromosomes and Cancer* 48, 816–827.
19. Royer-Pokora, B., Busch, M., Beier, M., Duhme, C., de Torres, C., Mora, J., Brandt, A., and Royer, H.-D. (2010). Wilms tumor cells with WT1 mutations have characteristic features of mesenchymal stem cells and express molecular markers of paraxial mesoderm. *Hum. Mol. Genet.* 19, 1651–1668.

20. Royer-Pokora, B., Beier, M., Brandt, A., Duhme, C., Busch, M., de Torres, C., Royer, H., and Mora, J. (2018). Chemotherapy and terminal skeletal muscle differentiation in WT1-mutant Wilms tumors. *Cancer Med.* 7, 1359.
21. Nayak, P., Colas, A., Mercola, M., Varghese, S., and Subramaniam, S. (2021). Temporal mechanisms of myogenic specification in human induced pluripotent stem cells. *Sci Adv* 7.
22. Krepischi, A.C.V., Maschietto, M., Ferreira, E.N., Silva, A.G., Costa, S.S., da Cunha, I.W., Barros, B.D.F., Grundy, P.E., Rosenberg, C., and Carraro, D.M. (2016). Genomic imbalances pinpoint potential oncogenes and tumor suppressors in Wilms tumors. *Mol. Cytogenet.* 9, 1–10.
23. Van der Auwera, G.A., Carneiro, M.O., Hartl, C., Poplin, R., Del Angel, G., Levy-Moonshine, A., Jordan, T., Shakir, K., Roazen, D., Thibault, J., et al. (2013). From FastQ data to high confidence variant calls: the Genome Analysis Toolkit best practices pipeline. *Curr. Protoc. Bioinformatics* 43, 11.10.1–11.10.33.
24. Kerstens, H.H.D., Hehir-Kwa, J.Y., van de Geer, E., van Run, C., Verwiel, E.T.P., van der Leest, D., Tops, B.B.J., and Kemmeren, P. (2020). Trecode: a FAIR eco-system for the analysis and archiving of omics data in a combined diagnostic and research setting. *bioRxiv*, 2020.11.13.363689.
25. van Belzen, I.A.E.M., Cai, C., van Tuil, M., Badloe, S., Strengman, E., Janse, A., Verwiel, E.T., van der Leest, D.F.M., Kester, L., Molenaar, J.J., et al. (2021). Systematic discovery of gene fusions in pediatric cancer by integrating RNA-seq and WGS. *bioRxiv*, 2021.08.31.458342.
26. Benjamin, D., Sato, T., Cibulskis, K., Getz, G., Stewart, C., and Lichtenstein, L. (2019). Calling Somatic SNVs and Indels with Mutect2. *bioRxiv*, 861054.

27. Amemiya, H.M., Kundaje, A., and Boyle, A.P. (2019). The ENCODE Blacklist: Identification of Problematic Regions of the Genome. *Sci. Rep.* 9.
28. McLaren, W., Gil, L., Hunt, S.E., Riat, H.S., Ritchie, G.R.S., Thormann, A., Flieck, P., and Cunningham, F. (2016). The Ensembl Variant Effect Predictor. *Genome Biol.* 17, 122.
29. Van der Auwera, G.A., Carneiro, M.O., Hartl, C., Poplin, R., Del Angel, G., Levy-Moonshine, A., Jordan, T., Shakir, K., Roazen, D., Thibault, J., et al. (2013). From FastQ data to high confidence variant calls: the Genome Analysis Toolkit best practices pipeline. *Curr. Protoc. Bioinformatics* 43, 11.10.1–11.10.33.
30. Chen, X., Schulz-Trieglaff, O., Shaw, R., Barnes, B., Schlesinger, F., Källberg, M., Cox, A.J., Kruglyak, S., and Saunders, C.T. (2016). Manta: rapid detection of structural variants and indels for germline and cancer sequencing applications. *Bioinformatics* 32, 1220–1222.
31. 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.
32. Cameron, D.L., Schröder, J., Penington, J.S., Do, H., Molania, R., Dobrovic, A., Speed, T.P., and Papenfuss, A.T. (2017). GRIDSS: sensitive and specific genomic rearrangement detection using positional de Bruijn graph assembly. *Genome Research* 27, 2050–2060.
33. Collins, R.L., Brand, H., Karczewski, K.J., Zhao, X., Alfoldi, J., Francioli, L.C., Khera, A.V., Lowther, C., Gauthier, L.D., Wang, H., et al. (2020). A structural variation reference for medical and population genetics. *Nature* 581.
34. dbVar <https://www.ncbi.nlm.nih.gov/dbvar/studies/nstd186/>.

35. MacDonald, J.R., Ziman, R., Yuen, R.K.C., Feuk, L., and Scherer, S.W. (2014). The Database of Genomic Variants: a curated collection of structural variation in the human genome. *Nucleic Acids Res.* 42, D986–92.
36. Manders, F., Brandsma, A.M., de Kanter, J., Verheul, M., Oka, R., van Roosmalen, M.J., van der Roest, B., van Hoeck, A., Cuppen, E., and van Boxtel, R. (2022). MutationalPatterns: the one stop shop for the analysis of mutational processes. *BMC Genomics* 23, 134.
37. Chaves-Urbano, B., Hernando, B., Garcia, M.J., and Macintyre, G. (2022). CNpare: matching DNA copy number profiles. *Bioinformatics*.
38. Brunet, J.P., Tamayo, P., Golub, T.R., and Mesirov, J.P. (2004). Metagenes and molecular pattern discovery using matrix factorization. *Proc. Natl. Acad. Sci. U. S. A.* 101.
39. Love, M.I., Huber, W., and Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* 15, 550.
40. Wu, T., Hu, E., Xu, S., Chen, M., Guo, P., Dai, Z., Feng, T., Zhou, L., Tang, W., Zhan, L., et al. (2021). clusterProfiler 4.0: A universal enrichment tool for interpreting omics data. *Innovation (Camb)* 2, 100141.
41. Tate, J.G., Bamford, S., Jubb, H.C., Sondka, Z., Beare, D.M., Bindal, N., Boutselakis, H., Cole, C.G., Creatore, C., Dawson, E., et al. (2019). COSMIC: the Catalogue Of Somatic Mutations In Cancer. *Nucleic Acids Res.* 47, D941–D947.
42. Chakravarty, D., Gao, J., Phillips, S.M., Kundra, R., Zhang, H., Wang, J., Rudolph, J.E., Yaeger, R., Soumerai, T., Nissan, M.H., et al. (2017). OncoKB: A Precision Oncology Knowledge Base. *JCO Precis Oncol* 2017.
43. Gröbner, S.N., Worst, B.C., Weischenfeldt, J., Buchhalter, I., Kleinheinz, K., Rudneva,

V.A., Johann, P.D., Balasubramanian, G.P., Segura-Wang, M., Brabetz, S., et al. (2018). The landscape of genomic alterations across childhood cancers. *Nature* 555, 321–327.

44. Martens, M., Ammar, A., Riutta, A., Waagmeester, A., Slenter, D.N., Hanspers, K., A Miller, R., Digles, D., Lopes, E.N., Ehrhart, F., et al. (2021). WikiPathways: connecting communities. *Nucleic Acids Res.* 49, D613–D621.

45. Liberzon, A., Birger, C., Thorvaldsdóttir, H., Ghandi, M., Mesirov, J.P., and Tamayo, P. (2015). The Molecular Signatures Database (MSigDB) hallmark gene set collection. *Cell systems* 1, 417.

46. Wickham, H. (2016). *ggplot2: Elegant Graphics for Data Analysis* (Springer).

47. Kolde, R. (2019). *pheatmap: Pretty Heatmaps*. <https://rdrr.io/cran/pheatmap/>.

48. *ggpubr: “ggplot2” Based Publication Ready Plots* (2020). <https://rdrr.io/cran/ggpubr/>.

49. *rstatix: Pipe-Friendly Framework for Basic Statistical Tests* (2021).  
<https://rdrr.io/cran/rstatix/>.

50. Gu, Z., Gu, L., Eils, R., Schlesner, M., and Brors, B. (2014). *circlize* implements and enhances circular visualization in R. *Bioinformatics* 30, 2811–2812.