Genetic changes associated with relapse in favorable histology Wilms tumor: A Children’s Oncology Group AREN03B2 study

Highlights

- Methods for measuring 1q gain, seen in 75% of relapse samples, merit optimization
- Drivers of progenitor proliferation (SIX1 and MYCN) affect tumorigenesis and relapse
- Mutations in DIS3 and in the TERT promoter are newly identified
- Chromosomal gains and losses often occur prior to mutations in Wilms tumor

Authors

Samantha Gadd, Vicki Huff, Andrew D. Skol, ..., Sheena Aris, Quy H. Phung, Elizabeth J. Perlman

Correspondence
eperlman@luriechildrens.org

In brief

Wilms tumors have many potential driver mutations. Gadd et al. focus on samples from relapse tumors revealing that over 40% contain mutations in SIX1 or in MYCN network genes, drivers of progenitor proliferation, and that 75% have 1q gain, supporting optimization of new tools for identifying 1q gain.
SUMMARY

Over the last decade, sequencing of primary tumors has clarified the genetic underpinnings of Wilms tumor but has not affected therapy, outcome, or toxicity. We now sharpen our focus on relapse samples from the umbrella AREN03B2 study. We show that over 40% of relapse samples contain mutations in SIX1 or genes of the MYCN network, drivers of progenitor proliferation. Not previously seen in large studies of primary Wilms tumors, DIS3 and TERT are now identified as recurrently mutated. The analysis of primary-relapse tumor pairs suggests that 11p15 loss of heterozygosity (and other copy number changes) and mutations in WT1 and MLLT1 typically occur early, but mutations in SIX1, MYCN, and WTX are late developments in some individuals. Most strikingly, 75% of relapse samples had gain of 1q, providing strong conceptual support for studying circulating tumor DNA in clinical trials to better detect 1q gain earlier and monitor response.

INTRODUCTION

The goal of this study is to identify genomic characteristics of relapse in Wilms tumors to better identify those at risk of relapse and to better understand the biology of relapse. Wilms tumor (WT; nephroblastoma) is the most common pediatric renal tumor; approximately 95% are of favorable histology (FHWT), and these are the focus of this study. The remaining 5% show histologic evidence of anaplasia, commonly associated with mutations or deletions of TP53. Individuals with FHWT are treated with a chemotherapy backbone including vincristine and actinomycin (stage I and II individuals); doxorubicin, cyclophosphamide, and etoposide may be added for stage III/IV individuals, and advanced-stage individuals also receive radiation to sites of disease. Although individuals with FHWT enjoy an overall survival rate of ~90%, this comes at a considerable cost, particularly for those with advanced disease. The priority is to identify biological factors that would improve our ability to predict relapse, better tailor relapse therapy, and reduce the significant toxicity associated with relapse therapies.

Two genetic loci have long been associated with the pathogenesis of FHWT, the WT1 gene located at 11p13 and the IGF2/H19 imprinted region on 11p15; each locus is associated with syndromes when present in the germline (reviewed in Gadd et al.1). Loss of imprinting (LOI) or loss of heterozygosity (LOH) of 11p15 is observed in the considerable majority of all WTs and results in overexpression of IGF2. Although 11p15 imprinting abnormalities clearly play a critical role in WT development, the observation of 11p15 LOH in normal tissue from some individuals with WT1 and the lack of tumors arising in mutant mice with LOI of the imprint control region imply that biallelic expression of IGF2 alone is insufficient for tumor development.

Over the last decade, a number of investigators have reported next-generation sequencing of large numbers of WTs. These studies indicate that WTs typically arise after acquisition of more than one genetic event. Rather than a
limited number of shared driver mutations, WTs have a large number of candidate driver genes that have in common functional involvement in early renal development, often through epigenetic regulation of transcription (chromatin modifications, transcription elongation, and microRNAs [miRNAs]). The most highly represented mutations have been identified in WT1, DROSHA, DGCR8, SIX1/SIX2, CTNNB1, FAM123B (WTX and AMERT), and MYCN. However, only half of individuals with FHWTs have mutations in one of these genes, and many FHWTs lack clear driver mutations. This observation prompted analysis of high-resolution SNP arrays of large numbers of WTs to identify additional regions recurrently gained and lost.14–16 This revealed several recurrent genetic regions of gain or loss, but the underlying pathogenetic genes and/or pathways remain elusive for most loci.

A molecular feature that has been used to stratify treatment of FHWTs in clinical trials is LOH of chromosomes 1p and 16q.17 Intensification of therapy for individuals with combined LOH 1p and 16q improves the outcome in all stages of FHWT.17 Although highly specific for predicting relapse, 1p/16q LOH is present in only 4.6% of FHWTs and in only 9.4% of relapses.18,19 More recently, gain of chromosome 1q has been associated with inferior event-free survival (EFS) and overall survival (OS) of individuals with WTs.20,21 This was subsequently validated by the National Wilms Tumor Study-5 and the International Society of Paediatric Oncology (SIOP) WT 2001 Trial. Both studies identified 1q gain in 28% of individuals overall and demonstrated diminished EFS for individuals with 1q gain.22,23 Upcoming protocols will determine whether modifying the initial therapy based on 1q gain will improve outcomes.

All of these studies have largely relied on randomly selected samples taken at the time of diagnosis, samples that may not contain the clonal events resulting in poor outcome. The current study seeks to determine whether examining relapse samples can provide further information regarding the pathogenesis, progression, and therapeutic responsiveness for individuals with FHWTs. These studies are possible because of the AREN03B2 umbrella biology and classification study that served as the entry portal to all Children’s Oncology Group (COG) individuals registered on therapeutic protocols for renal tumors from 2006 through 2017; it now includes banked samples from over 6,000 individuals. The overall goal for the current study is to analyze samples from individuals registered as FHWT on AREN03B2 who relapsed as FHWT.

RESULTS

Clinical samples
Individuals with currently valid and verified consent who relapsed with FHWT and who had samples banked at the Biopathology Center (BPC) were considered eligible. To gain maximal information from as many samples as possible, independent discovery and validation sets were defined.

Discovery set
Relapse and germline samples from 51 unique individuals passed the quality control steps. Two individuals had samples from two different relapse episodes (PAWPUL and PATEIS). In 45 of 51 individuals, DNA was also available from the primary tumor sample that likewise passed quality control and therefore represent complete trios. RNA sequencing was performed on 49 of 51 relapse samples and 12 of 45 primary samples for which an adequate sample was available. Adequate samples were available for miRNA extraction for all 51 relapse samples and 12 paired primary samples.

Validation set
Independent of the discovery set, 31 additional individuals with relapse samples that passed DNA quality control were eligible for the validation set but not for the discovery set (STAR Methods). For the majority of these, the samples consisted of two unstained slides and an H&E slide.

Somatic variants in the discovery set

Relapse samples
Whole genomic sequencing (WGS) was performed on 53 relapse samples from 51 individuals. This resulted in 3,846 small variants, 301 of which passed the filtering criteria (STAR Methods). The details of all 301 variants are provided in Table S1. Sixteen genes were affected by these 301 somatic variants in more than one individual, and these are provided in Table 1; all were verified by RNA sequencing (RNA-seq) or Sanger sequencing. Nine of these 16 genes involved genes and variants previously identified and fully described in previous large studies of primary FHWTs.9–13 In particular, the SIX1 177 Q/R hotspot mutation was identified in 7 of 51 individuals, and the MYCN 44 P/L hotspot mutation was identified in 5 of 51 individuals. The remaining 7 genes have not been reported previously to be mutated in WTs (MGA, TCF12, RBL1, HCFC1, MAPKBP1, COBLL1, and DIS3), each identified in two individuals.

Recurrence structural variants were also identified and are provided in Tables 1 and S2. These include tandem duplications of MYCN (7 individuals), deletions involving WT1 (2 individuals), and deletions including all or part of the WTX gene (7 individuals). WTX also contained a non-recurrent nonsense SNP mutation (Table 1).

Primary samples
WGS was also performed on 45 available paired primary samples, which resulted in 1,804 small variants, 249 of which passed the same filtering criteria applied to the relapse samples (STAR Methods). The details of all 249 are provided in Table S1; structural variants are provided in Table S2.

The tumor mutation burden (TMB) per megabase (assuming 25.8-Mb non-redundant coding regions in WGS) was calculated for the primary and relapse samples of all discovery set individuals. This demonstrated a low TMB for all tumors (ranging from 0.04–0.89 in the primary samples and 0.08–1.74 in the relapse samples), provided in Figure S1. Genes recurrently involved in more than three individuals in primary or relapse samples are illustrated in Figure 1, and those involved in more than two individuals are illustrated in Figure S1.

Germline mutations of the discovery set
Peripheral blood samples (43) or normal kidney samples (8) of the 51 discovery set individuals were examined for small variants in genes recognized to be predisposing to WTs24 and genes identified in individuals with WTs known to predispose to adult tumors.25 These include BLM, BRCA2, BUB1B, CDC73,
CHEK2, CTR9, DICER1, DIS3L2, GPC3, GPC4, MUTFYH, PALB2, PIK3CA, PMS2, REST, TP53, TRIM37, and WT1. This germline predisposition analysis revealed a pathogenic variant in DICER1 in an individual who also had a different somatic DICER1 mutation in the primary tumor (epithelial predominant) and relapse (blastemal predominant) samples (PAUSLU). A likely pathogenic variant in CHEK2 (rs587782471, associated with predisposition to breast cancer) was also identified. Variants of unknown significance were identified in WT1 (1), BLM (2), REST (1), and TRIM28 (1); none demonstrated a reduction to homozygosity in the primary and relapse samples.

We also searched the germline for variants in the genes showing recurrent somatic variants listed in Table 1. Germline DIS3 variants of unknown significance were identified in 4 individuals (in addition to the two individuals with somatic DIS3 hotspot 488 D/N mutations). The rs141067458 stop-loss germline variant was identified in two individuals (PASYKN and PAVBXS); two individuals had the rs35288597 coding-change variant at amino acid 438 (PAUGMT and PAYTJD). Last, germline variants of unknown significance were detected in CHD4 (rs372219150) and RBL1 (rs761881234), each in one individual. Details of germline variants are provided in Table S2; all were identified in peripheral blood samples.

### Copy number changes in the discovery set highlight 1q gain

Segmental copy number analysis for the relapse sample and the paired primary tumor sample (when available) was computed from WGS data by the GDC. Regions reported previously as gained or lost in WTs 14–16 were evaluated and are provided in Table 2. These data confirm numerous gains and losses of entire chromosomes or chromosomal arms in WTs, particularly gain of 1q, 6, 7q, and 12 and loss of 1p, 16q, and 22. When comparing the copy number changes identified in the relapse sample with the primary sample (when available), the only copy number change that was significantly different was gain of 1q. In relapse samples, 38 of 51 (74.5%) demonstrated 1q gain compared with 21 of 45 (47%, p = 0.008, Fisher’s exact test) of the available primary tumor samples. The rate identified in primary tumors (47%) and relapse tumors (74.5%) in the current study is also greater than the overall rate of 28% identified previously in all WTs. The number of male and female individuals with 1q gain in the discovery set (15 of 21 females and 23 of 30 males) was not significantly different (p = 0.7499, Fisher’s exact test). Comparing the copy-neutral LOH (CNLOH) or LOH for 1p or 16q within the relapse sample with the primary sample did not demonstrate significance.

### Table 1. Genes recurrently involved with mutations in relapse samples

| Gene/locus | No. of somatic variants in discovery relapse samples (51 individuals) | No. of variants in validation set relapse samples (31 individuals) | Percentage of all 82 relapse individuals | Reported previously in WTs | No. in TARGET (n = 533 except where noted) | % in TARGET | Fisher’s exact p value |
|------------|---------------------------------------------------------------------|------------------------------------------------------------------|-----------------------------------------|--------------------------|---------------------------------------------|------------|------------------------|
| SNVs, indels | | | | | | | |
| SIX1 (14q23) | 7 hotspot | 4 hotspot | 13.4 | yes | 23 | 4 | 0.0026 |
| CTNNB1 (3p22) | 5 (4 individuals) | 4 | 9.8 | yes | 86 | 16 | 0.255 |
| MYCN (2p24) | 5 hotspot | 2 hotspot | 8.5 | yes | 22 | 4 | 0.0914 |
| WT1 (11p13) | 3 | 2 | 6.1 | yes | 40 | 7.5 | 0.8208 |
| MLLT1 (19p13) | 2 hotspot | 3 | 6.1 | yes | 19 | 3.6 | 0.3509 |
| DGCR8 (22q11) | 4 hotspot | 0 | 4.9 | yes | 22 | 4 | 0.7667 |
| DROSHA (5p13) | 2 | N/D | 3.9 | yes | 61 | 11.4 | 0.1514 |
| CHD4 (12p13) | 2 | 1 | 3.7 | yes | 6 | 1.2 | 0.1057 |
| MAX (14q23) | 2 hotspot | 0 | 2.4 | yes | 11 | 2.1 | 0.6878 |
| MGA (15q15) | 2 | 3 | 6.1 | no | | | |
| HCFC1 (Xq28) | 2 | 3 | 6.1 | no | | | |
| COBLL1 (2q24) | 2 | 2 | 4.9 | no | | | |
| RBL1 (20q11) | 2 | 1 | 3.7 | no | | | |
| TCF12 (15q21) | 2 | 0 | 2.4 | no | | | |
| MAPKBP1 (15q15) | 2 | 0 | 2.4 | no | | | |
| DIS3 (13q21) | 2 | 0 | 7.3 | no | | | |
| WTX (Xq11) | 1 | 0 | 2.4 | yes | 34 | 6.4 | 0.0707 |
| TERT (5p15) | 3 | 0/25 | 3.9 | no | 1/56 | 1.8 | 0.643 |
| Structural variants | | | | | | | |
| MYCN (tandem duplication) | 7 | 4 | 13.4 | yes | 4/56 | 7.1 | 0.2801 |
| WTX (deletion) | 7 | 3 | 11.0 | yes | 4/56 | 7.1 | 0.4005 |
| WT1 (deletion) | 3 (2 individuals) | 3 | 6.1 | yes | 2/56 | 3.6 | 0.4278 |

Indel, insertion or deletion; N/D, not done.
(p = 0.3151 for 1p and p = 0.6399 for 16q). Segmental copy number changes identified in relapse and primary samples (when available) for each individual are provided in Table S3, where co-occurrence may be evaluated further. Contingency tables comparing patterns of gain of 1q and loss of 1p or 16q revealed no patterns that were statistically different. In addition to 1q gain, two other chromosomal gains (chromosomes 12 and 18) showed a significantly higher frequency in the relapse sample of the current study compared with previously reported studies of overall primary WTs (28 of 51 versus 9 of 50, p < 0.0001 for chromosome 12 and 17 of 51 versus 5 of 50, p = 0.0072 for chromosome 18, Fisher’s exact test) (Table 2). Of the 38 relapse samples with 1q gain, 25 also had gain of 12 (p = 0.0106), and 16 also had gain of 18 (p = 0.0384, Fisher’s exact test).

Comparison of primary-relapse tumor pairs reveals late acquisition of some recurrent mutations

Of the 51 discovery set individuals, 45 had available DNA from the primary tumor and were thus evaluable for comparison with the relapse tumor. In several individuals, a mutation was present in the relapse sample but not in the paired primary tumor. This discordancy was observed for SIX1 (3 of 6 evaluable individuals), MYCN (2 of 7 evaluable individuals), and WTX (3 of 8 evaluable individuals), illustrated in Figure 1 for genes with 3 or more mutations and in Figure S1 for genes with 2 or more mutations. To further verify the absence of the mutation in the primary sample, the unfiltered data of all evaluable discordant tumor sets were searched, and no variants were identified that were filtered out because of one of the criteria. For other genes, all evaluable tumor sets were concordant: WT1 (5), DROSHA (3), CTNNB1 (3), CHD4 (2), MLLT1 (2), and DGCR8 (1). All genes identified in relapsed WTs for the first time in this study (MGA, TCF12, RBL1, HCF1, MAPKBP1, COBLII, and DIS3) were discordant in at least one individual.

These observations indicate that there are 17 of 45 individuals whose primary tumor sample lacked a recurrent mutation from Table 1. This prompted a number of analyses of these samples with the aim of identifying underlying pathogenic variants in these 17 individuals. First, we identified variants that have been detected in FHWTs in prior studies that may have been filtered out as non-recurrent in this study. This identified a mutation in NONO that was observed in primary and relapse samples of PAVLIN. Second, we analyzed the raw variant calling format (VCF) files for recurrent mutations in non-coding regions. This resulted in identification of a promoter mutation in TERT (rs1242535815 G>A) in four individuals, including 3 of 51 individuals with relapse samples and 3 of 45 samples from primary tumors. (One relapse tumor with TERT mutation lacked a primary tumor, and the TERT mutation was present only in the primary tumor of one individual.) This mutation is a G>A change 124 bp upstream of the TERT transcription start site. The details of the TERT mutations are included in Tables 1 and S1. Retrospective analysis of the 56 FHWTs that underwent WGS in the Therapeutically Applicable Research to Generate Effective Treatments (TARGET) analysis revealed one tumor that carried this variant.13 Gene expression of TERT in the 5 samples with promoter

**Figure 1. Recurrent alterations in WT discovery set individuals**

Genetic alterations identified in at least 3 individuals within the primary (P) and relapse (R) tumors are illustrated in this OncoPrint. The numbers on the right provide the percentage of P or R samples that have alterations in the relevant gene. These data are also expanded in Figure S1, which provides all genes recurrently involved in each designated individual.
Mutations who had RNA available was significantly higher than those lacking promoter mutations (p = 0.0067) (Figure 2).

Individuals lacking mutations in their primary samples often have multiple copy number changes

The 16 remaining available primary tumors that lacked evidence of recurrent SNVs or structural variants were examined for copy number changes. We found that 11 of 16 demonstrated CNLOH of 11p15. In each case, the germline sample lacked 11p15 CNLOH. Excluding changes on 11p, the 16 samples had an average of 4.5 segmental copy number changes per tumor in the primary sample; only one individual lacked copy number changes (PAUWCD). Figure S1 illustrates the key mutations and copy number changes for each individual and provides the co-occurrence of those mutations and copy number changes.

The data in Figure S1 were also analyzed to identify tumors that had an identifiable stable clone (genetic changes present in primary and relapse samples) and evidence of clonal evolution (an additional change in the relapse sample). Of the 18 evaluable individuals, the changes most frequently identified in the stable clone included 11p15 CNLOH (7), gain of 12 (8), and gain of 18 (5). The most frequent additional genetic change identified only in the relapse sample included 1q gain (11), gain of 12 (4), gain of 18 (4), and mutations in WTX (4), MYCN (2), and SIX1 (2).

Also of interest was a paucity of copy number changes in individuals with WT1 and MLLT1 mutations.

Table 2. Segmental copy number changes

| Chr coordinates | Prevalence in literature | Copy call primary | Copy call relapse | Implicated genes in prior studies | Gene-level CN primary | Gene-level CN relapse |
|-----------------|--------------------------|-------------------|-------------------|----------------------------------|-----------------------|------------------------|
| Chr, chromosome, CNLOH, copy-neutral loss of heterozygosity.

The 16 remaining available primary tumors that lacked evidence of recurrent SNVs or structural variants were examined for copy number changes. We found that 11 of 16 demonstrated CNLOH of 11p15. In each case, the germline sample lacked 11p15 CNLOH. Excluding changes on 11p, the 16 samples had an average of 4.5 segmental copy number changes per tumor in the primary sample; only one individual lacked copy number changes (PAUWCD). Figure S1 illustrates the key mutations and copy number changes for each individual and provides the co-occurrence of those mutations and copy number changes.

The data in Figure S1 were also analyzed to identify tumors that had an identifiable stable clone (genetic changes present in primary and relapse samples) and evidence of clonal evolution (an additional change in the relapse sample). Of the 18 evaluable individuals, the changes most frequently identified in the stable clone included 11p15 CNLOH (7), gain of 12 (8), and gain of 18 (5). The most frequent additional genetic change identified only in the relapse sample included 1q gain (11), gain of 12 (4), gain of 18 (4), and mutations in WTX (4), MYCN (2), and SIX1 (2). Also of interest was a paucity of copy number changes in individuals with WT1 and MLLT1 mutations.

Targeted sequencing of validation set

Mutations in the genes identified in Table 1 were evaluated in the independent 31-individual validation set using targeted
sequencing; the location of all targets analyzed are provided in Table S4. 35 variants passed quality control parameters, and all are provided in Table 1. (TERT, discovered later in this study, was only able to be analyzed in 25 of 31 validation set individuals because of consumption of the sample; no mutations were identified). Details of these mutations in each tumor are provided in Table S4.

It should be noted that 6 of 35 variants (2 each involving COBLL1 and MGA, and one each involving RBL1 and HCFC1) are in the National Center for Biotechnology Information’s Single Nucleotide Polymorphism Database (dbSNP) and have an allelic fraction of ~50% and are therefore suspected to be rare germline variants. They are retained because of their predicted effect according to the sorting intolerant from tolerant (SIFT) and/or Polyphen tools for predicting whether an amino acid substitution will affect protein function. Such variants would not have been identified in the discovery set because of their presence in the paired germline DNA. The discovery set tumors were evaluated for these germline variants, and none were identified. In Table 1, the combined frequency of mutations in the discovery and validation sets was compared with the frequency identified in the TARGET study (using the appropriate total tumor number of 533 FHWTs in the TARGET validation set and 56 FHWTs in the TARGET discovery set that were analyzed by WGS for structural variants.) Only SIX1, identified in 11 of 82 (13.4%) individuals, showed a frequency of mutations significantly higher compared with the frequency seen in tumor samples in TARGET (23 of 533 FHWTs in the TARGET study (using the appropriate total tumor number of 533 FHWTs in the TARGET validation set and 56 FHWTs in the TARGET discovery set that were analyzed by WGS for structural variants.) Only SIX1, identified in 11 of 82 (13.4%) individuals, showed a frequency of mutations significantly higher compared with the frequency seen in tumor samples in TARGET (23 of 533 [4%], p = 0.0026). The number of male and female individuals with SIX1 mutation in the combined discovery and validation sets (6 of 36 females and 5 of 46 males) was not significantly different ($p = 0.5227$, Fisher’s exact test).

**Pharmacogenomic analysis**

Of the 226 unique pharmacogenomic polymorphisms analyzed (STAR Methods), four were excluded (one was not in gnomAD, and three had missing calls in more than 10% of samples). None of the remaining 222 polymorphisms were significantly different (adjusted $p < 0.05$, binomial test) in the entire sample set compared with the gnomAD general population. A separate analysis comparing all individuals of European descent ($N = 35$) with the non-Finnish European gnomAD population similarly yielded no significantly different polymorphisms. Other ethnicity-specific comparisons were not run because of low sample numbers: admixed American (AMR), 7; African (AFR), 6; Asian, 1; unknown, 2. Analysis for reduction to homozygosity performed in the paired relapse sample and (when available) the primary sample did not demonstrate a significant shift in the polymorphism frequencies in the tumor samples.

**DISCUSSION**

Studies comprehensively reporting the genomic analysis of WTs over the last decade have greatly increased our knowledge of the underlying genetic underpinnings of WTs. Despite this, identification of therapeutic targets has been limited. To sharpen our focus, this study analyzes WT relapse samples, which should contain clonal molecular features contributing to relapse.

**SIX1 and MYCN mutations are more frequently identified in relapsed WTs**

Prior studies have illustrated that mutations in many of the genes involved in renal development also play a key role in development of WTs, including SIX1, WT1, MYCN, WTX, MLLT1, and CHD4. The current study emphasizes that some of the same genetic mutations are also increased in prevalence at relapse, particularly those whose role is preserving the progenitor state. In particular, the highly homologous SIX1 and SIX2 genes are required for maintaining the progenitor state; the identical 177 Q/R hotspot mutations in SIX1 and SIX2 (so far specific to WTs) have been identified previously in about 5% of primary FHWTs. Previous reports of structural analysis of the SIX1 Q177R mutations suggest an effect on the DNA binding site, likely altering DNA binding specificity. This mutation is accompanied by up-regulation in cell cycle genes, supporting an activating function. We demonstrate the same SIX1 hotspot mutation in 11 of 82 (13.4%) of relapse samples in the current study, a significantly higher frequency compared with the 4%
of primary tumors found in the TARGET dataset (N = 533, p = 0.0026). It is noteworthy that SIX1 mutations were identified in the relapse sample but not in the primary sample in 3 of 6 of evaluable individuals, suggesting that it is not required for tumor development in many individuals. Supporting this is the report of increased allelic fraction of SIX1 mutations in relapse compared with primary samples in a study of 8 primary-relapse pairs. An increased frequency of SIX mutations was also identified in post-therapy WTs that were blastemal predominant compared with other histologic subtypes, suggesting that SIX mutations may confer resistance to chemotherapeutic agents. Prior studies point toward multiple functional roles of SIX1 that may contribute to its increased prevalence in relapsed individuals in our study; these include alteration of DNA binding specificity, up-regulation of cell cycle genes, and resistance to chemotherapy.

The MYCN network is also involved in preservation of the progenitor state in the kidneys. The activating MYCN 44 P/L hot-spot mutation and MYCN tandem duplications have together been reported previously in ~15% of primary WTs in studies that report SNVs and structural variants, and NOXO (1 individuals). MAX binds DNA as a heterodimer with MYCN, leading to post-transcriptional up-regulation of MYCN mRNA and protein expression. We identified the hotspot mutation only in the relapse sample and not in the primary sample in 2 of 3 evaluable individuals, whereas all four evaluable individuals with MYCN tandem duplication in the relapse sample also had this mutation in the primary sample. Additional individuals in this study had variants in other members of the MYC transcription factor network that are expected to result in cellular effects similar to MYCN overexpression, including MAX (2 individuals), MGA (5 individuals), and NONO (1 individuals). MAX binds DNA as a heterodimer with MYCN or MYCC, and this MYC-MAX transcription activator is involved in all known oncogene functions of MYC. MGA, likewise frequently mutated in cancer, binds to MAX and regulates target gene expression. NONO, an RNA-binding protein, binds to MYCN, leading to post-transcriptional up-regulation of MYCN mRNA and protein expression. In total, the relapse samples of 25 of 82 individuals (30%) included in this study had evidence of mutations involving the MYCN network. Studies suggest that the mechanisms underlying the increased relapse rate in individuals with activation of N-MYC may be linked to interacting partners, including PEG10, YEATS2, FOXX1, CBLL1, and MCRS1, all of which correlate positively with MYCN expression in WTs. FOXK1 in particular is known to regulate cancer initiation, development, angiogenesis, and drug resistance. Knockdown of YEATS2 in lung cancer cells results in growth suppression and reduced survival, all of which are key MYC functions. The interaction of MYC-N with YEATS2 may therefore contribute to oncogenesis by supporting cell growth and survival. Although all MYCN variants reported here were somatic events, germline MYCN duplication has been identified in a family predisposed to WTs.

### Recurrent DIS3 germline and somatic mutations and TERT promoter mutations

We identified recurrent somatic mutations in two functionally important genes that have not been recognized previously in WTs: DIS3 and TERT.

### DIS3 mutations

Mutations of miRNA processing genes are an important category of mutations in WTs; these result in global reduction of mature miRNAs, but in particular let-7a. Decreased let-7a may also result from up-regulation of LIN28B, which specifically binds pri/pre-let-7 miRNAs, triggering their degradation by Ds3L2, an exoribonuclease that is also rarely mutated in WTs. Rare genmine mutations in Ds3L2 result in Perlman syndrome, associated with increased risk of WTs. A paralog of Ds3L2 is Ds3, a protein not reported previously to be mutated in WT but recognized as a recurrent mutation resulting in multiple myeloma. A Ds3 somatic mutation (488 D/N) involving the catalytic domain of ribonuclease II (RNII domain) was identified in two individuals; this has been reported previously in individuals with multiple myeloma. Two different germline variants of unknown significance were also identified in two individuals each. In particular, the rs141067458 stop-loss germline variant has been reported to result in lower Ds3 expression and to be associated with familial multiple myeloma, although its role in WT development is unknown. Ds3 knockdown in Drosophila severely disrupts development of wing imaginal discs by regulating a small subset of microRNAs, in particular miR-252 and miR-982, miRNAs with no known human orthologs.

### TERT promoter mutations

The current study also provides the first report of TERT promoter mutations in WTs. Telomerase maintains telomere length, thereby maintaining self-renewal potential. Somatic mutations in the promoter of TERT, the catalytic subunit of telomerase, have been reported in several tumor types and are predicted to increase promoter activity and TERT transcription 2- to 6-fold (reviewed by Ackerman and Fischer). These promoter mutations were associated with high expression of TERT in the current study (Figure 2). Analysis of TERT expression in 78 FHWTs demonstrated a significant association between expression of TERT and relapse in univariate and multivariate analyses. This significant association was verified in a subsequent study of 244 NWTS-5 individuals (96 relapse, 148 without relapse).

Additional genes recurrently mutated in the relapse samples of small numbers of WTs have likewise not been reported previously, although their significance remains unclear. These include mutations in (1) the basic helix-loop-helix (bHLH) binding domain of TCF12, reported previously to be associated with an aggressive tumor phenotype in anaplastic oligodendroglioma; (2) HCFC1, whose loss results in proliferation of neural progenitor cells at the expense of differentiation; (3) RBL1, a member of the retinoblastoma tumor suppressor family that modulates E2F transcription factor activity; and (4) MAPKBP1, one of over 20 genes linked to development of nephroblastomatisis; and (5) COBLL1, a gene associated with age-related macular degeneration.

### Gain of 1q is highly prevalent in WT relapse

Observations highlighted by the current study, but certainly documented previously by others, include the important role of copy number change in WTs. Although some of the regions gained or lost have some degree of data supporting the role of individual genes, most do not, despite a great deal of
effort over the last decade. The most striking finding of this study is the prevalence of 1q gain in the relapse samples of WTs (75%) compared with the primary samples (47%) and compared with the overall prevalence of 1q gain reported previously in primary WTs (28%). The increased prevalence of 1q gain in individuals with increased stage and increased age,22 and the increased allelic fraction seen in smaller studies of relapsed individuals,27 strengthens the growing consensus that 1q gain is often associated with progression and solidifies its role in guiding therapy.

This study relies on retrospective analysis of prospectively obtained tumor samples from individuals registered on COG studies, which have historically relied on a single randomly selected tumor sample. This practice enables collection of the highest quality of sample (a fresh tumor collected shortly after surgery). However, this leaves large areas of a tumor unsampled and does not allow selection based on histology. Concern regarding the effect of tumor heterogeneity has therefore grown. To address this, Cresswell et al.6 collected 70 tumor samples from 24 tumors in 20 individuals and demonstrated striking heterogeneity in their ability to detect 1q gain in multi-sampled tumors. In fact, had they only collected a single sample per tumor, 1q gain would have been detected in only about a third of the cases in which it was present. They estimated that at least three samples per tumor were needed to ensure that more than 95% of tumors with 1q gain would be detected. Sampling bias resulting from reliance on a single random tumor sample is the largest limitation of the current study. To correct this sampling bias in clinical trials is remarkably difficult for a number of practical reasons. Simply taking three samples from each tumor at the time of nephrectomy will not address all situations, particularly those relying on initial biopsy. To address this concern, efforts have recently focused on detecting circulating tumor DNA (ctDNA), and the possibilities and pitfalls have been discussed in the setting of pediatric cancer.56,57 In a recent study of ctDNA in 50 individuals with high-stage WTs, only individuals with detectable ctDNA experienced relapse or died from disease.58 Although the presence of 1q gain in the tumor predicted its presence in the serum, a number of individuals showed 1q gain in the serum but lacked 1q gain in the randomly selected tumor sample.58 This supports the concept that measuring ctDNA at the time of diagnosis in individuals with WTs may enable detection of clonal 1q gain anywhere in the entire tumor burden. Studies examining the clinical utility of ctDNA have been included in the next therapeutic protocols for WTs. Copy number changes of other chromosomes, such as chromosomes 12 and 18, within the tumor may also be independently useful in predicting relapse and may augment our understanding of the development of WTs. These data will be easily captured because microarrays will be utilized in the next COG protocols to comprehensively evaluate copy number change in the tumor.

The power of this study is that it represents a comprehensive analysis of the largest number of relapse samples of WTs reported to date. In addition, the availability of primary and normal tissue from many individuals enables us to gain some insight into the temporal acquisition of mutations in WTs. This observation adds to an accumulating set of evidence that suggests that genetic variants may play important roles throughout tumor evolution, although there is not yet evidence to support a particular sequence of genetic events. In fact, the combinations of mutations or structural changes may be critical, rather than the temporal order of their accumulation. In particular, the co-occurrence of mutations in genes supporting continued progenitor proliferation with those preventing differentiation may be most important. Examples include SIX with DROSHA and WT1 with CTNNB1.13

Limitations of the study
The limitations of this study reside in the reliance on a single random sample at each episode (reviewed above) combined with the reliance on the relatively low coverage provided by WGS, precluding assessment of clonal evolution in this study. Another limitation of this study is the relatively small number of cases analyzed, given the nature of the tools applied, which generally require a large number of samples to achieve confidence. This limits the type of conclusions that can be confidently drawn in this study, particularly for pharmacogenomic variants. It also limits our ability to provide biologic verification using the RNA and miRNA expression patterns.

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 SUBJECT DETAILS
- METHOD DETAILS
  - Eligibility for this study included the following
  - Specimen processing and quality control
  - Sequencing methods
- VALIDATION SET TARGETED SEQUENCING
  - Project analytic methods
- PHARMACOGENOMIC ANALYSIS
- QUANTIFICATION AND STATISTICAL ANALYSIS

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at https://doi.org/10.1016/j.xcrm.2022.100644.

ACKNOWLEDGMENTS

This project was funded by the NCI Molecular Profiling to Predict Response to Treatment (MP2PRT) Program: Retrospective Characterization and Analysis of Biospecimens Collected from NCI-Sponsored Trials of the NCTN and NCORP. The project was also supported by the COG Chair’s Grant (U10CA098543), NCTN Network Group Operations Center Grant (U10CA180886), Statistics & Data Center Grant (U10CA098413), NCTN Statistics & Data Center (U10CA180899), Human Specimen Banking in NCI-Sponsored Clinical Trials (U24CA114766), Human Specimen Banking in NCI-Sponsored Clinical Trials (U24-CA196173), and supported by St. Baldrick’s Foundation. This clinical study was supported in whole or in part by funding from the National Cancer Institute (NCI) under the Cancer MoonshotTM Initiative. This project was
funded in whole or in part from federal funds from the National Cancer Institute, National Institutes of Health, under contract HHSN261201500003I. The content of this publication does not necessarily reflect the views or policies of the Department of Health and Human Services, nor does mention of trade names, commercial products, or organizations imply endorsement by the US government. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health. The authors thank the Children’s Oncology Group (COG) protocol coordinators, research coordinators, clinical research assistants, and other health professionals who contributed to acquiring samples used in this study. The authors in particular wish to thank the extended AREN03B2 team, who contributed to the central review and clinical annotation of these samples, as well as Jeff Dome, chair of the COG renal tumor disease committee during acquisition of some of the samples utilized in this study. We also acknowledge the excellent efforts of project managers from the BPC (Yvonne Moyer and Kae Tegtmeier), as well as support from Kristina Tracy from the Genomics Platform of the Broad Institute of MIT and Harvard. Most importantly, we are very grateful to the individuals and their families for consenting to sample deposition at the Broad Institute of MIT and Harvard. Most importantly, we are very grateful to the individuals and their families for consenting to sample deposition at the Broad Institute of MIT and Harvard. Most importantly, we are very grateful to the individuals and their families for consenting to sample deposition through AREN03B2 to the BPC.

AUTHOR CONTRIBUTIONS
Conceptualization: E.J.P., C.V.F., E.A.M., V.H., and N.R.; methodology, software, formal analysis, and investigation: S.G., A.D.S., C.D.J., K.A.H., E.J.P., and K.L.Y.; resources: C.V.F., E.A.M., V.H., and N.C.R.; data curation: Tegtmeier), as well as support from Kristina Tracy from the Genomics Platform of the Broad Institute of MIT and Harvard. Most importantly, we are very grateful to the individuals and their families for consenting to sample deposition through AREN03B2 to the BPC.

DECLARATION OF INTERESTS
The authors declare no competing interests.

REFERENCES
1. Ooms, A.H., Gadd, S., Gerhard, D.S., Smith, M.A., Guidry Auvil, J.M., Meerzaman, D., Chen, Q.R., Hu, S.H., Yan, C., Nguyen, C., et al. (2016). Significance of TP53 mutation in Wilms tumors with diffuse anaplasia: a report from the children’s Oncology group. Clin Cancer Res 22, 5582–5591. https://doi.org/10.1158/1078-0432.ccr-16-0985.
2. Dome, J.S., Mullen, E.A., Dix, D.B., Gratias, E.J., Ehrlich, P.F., Daw, N.C., Geller, J.I., Chintagumpala, M., Khanna, G., Kalapurakal, J.A., et al. (2021). Impact of the first generation of children’s Oncology group clinical trials on clinical practice for Wilms tumor. J. Natl. Compr. Canc. Netw. 19, 978–985. https://doi.org/10.6004/jnccn.2021.7070.
3. Termuhlen, A.M., Tersak, M.J., Liu, Q., Yasui, Y., Stovali, M., Weathers, R., Deutsch, M., Sklar, C.A., Oeffinger, K.C., Armstrong, G., et al. (2011). Twenty-five-year follow-up of childhood Wilms tumor: a report from the childhood cancer survivor study. Pediatr Blood Cancer 57, 1210–1216. https://doi.org/10.1002/pbc.23090.
4. 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., and Perlman, E.J. (2012). Clinically relevant subsets identified by gene expression pattern support a revised ontogenic model of Wilms tumor: a children’s Oncology group study. Neoplasia 14, 742–756. https://doi.org/10.1593/neo.12714.
5. Grundy, P., Telzerow, P., Moksness, J., and Breslow, N.E. (1996). Clinico-pathologic correlates of loss of heterozygosity in Wilms’ tumor: a preliminary analysis. Med. Pediatr. Oncol. 27, 429–433. https://doi.org/10.1002/(sici)1096-911x(199611)27:5<429:aaid-mpo7>3.0.co;2-o.
6. Cresswell, G.D., Apps, J.R., Chagati, T., Misfud, B., Bentley, C.C., Maschietto, M., Popov, S.D., Weeks, M.E., Olsen, Ø.E., Sebire, N.J., et al. (2016). Intra-tumor genetic heterogeneity in Wilms tumor: clonal evolution and clinical implications. EBioMedicine 9, 120–129. https://doi.org/10.1016/j.ebiom.2016.05.029.
7. Chao, L.Y., Huff, V., Tomlinson, G., Riccardi, V.M., Strong, L.C., Saunders, G.F., et al. (1993). Genetic mosaicism in normal tissues of Wilms’ tumour patients. Nat. Genet. 3, 127–131. https://doi.org/10.1038/ng0293-127.
8. Huang, L., Koppakati, S., Hu, Q, Ruteshouser, E.C., Hicks, M.J., and Huff, V. (2016). Nephron progenitor but not stromal progenitor cells give rise to Wilms tumors in mouse models with p-Catenin activation or Wt1 ablation and Igf2 upregulation. Neoplasia 18, 71–81. https://doi.org/10.1016/j.neo.2015.12.001.
9. Torreznan, G.T., Ferreira, E.N., Nakahata, A.M., Barros, B.D.F., Castro, M.T.M., Correa, B.R., Krepischi, A.C.V., Olivieri, E.H.R., Cunha, I.W., Tabori, U., et al. (2014). Recurrent somatic mutation in DROSHA induces microRNA profile changes in Wilms tumour. Nat. Commun. 5, 4039. https://doi.org/10.1038/ncomms5039.
10. Wegrz, J., Ishaque, N., Vardapour, R., Geog, C., Gu, Z., Biesig, M., Bausenwein, S., Nourani, K., Ludvig, 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. https://doi.org/10.1016/j.ccell.2015.01.002.
11. Rakheja, D., Chen, K.S., Liu, Y., Shukla, A.A., Schmid, V., Chang, T.C., Khokhar, S., Wickiser, J.E., Karandikar, N.J., Matser, J.L., et al. (2014). Somatic mutations in DROSHA and DICER1 impair microRNA biogenesis through distinct mechanisms in Wilms tumours. Nat. Commun. 5, 4802. https://doi.org/10.1038/ncomms5802.
12. Walz, A.L., Ooms, A.H., Gadd, S., Gerhard, D.S., Smith, M.A., Guidry Auvil, J.M., Meerzaman, D., Chen, Q.R., Hu, S.H., Yan, C., et al. (2015). Recurrent DGC8, DROSHA, and SIX homeodomain mutations in favorable histology Wilms tumors. Cancer Cell 27, 426–487. https://doi.org/10.1016/j.ccell.2015.02.008.
13. 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. https://doi.org/10.1038/ng.3940.
14. Ciceri, S., Montiváio-de-Azevedo, R., Tajbakhsh, A., Bertolotti, A., Spagnuolo, R.D., Boschetti, L., Capasso, M., D’Angelo, P., Serra, A., Diomedi-Camassi, F., et al. (2021). Analysis of the mutational status of SIX1/2 and microRNA processing genes in paired primary and relapsed Wilms tumors and association with relapse. Cancer Gene Ther. 28, 1016–1024. https://doi.org/10.1038/s41417-020-00268-3.
15. Krepsich, 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, 20. https://doi.org/10.1186/s13039-016-0227-y.
16. Williams, R.D.O., Al-Saadi, R., Chagati, T., Popov, S., Messeheli, B., Sebire, N., Gessler, M., Weger, J., Graf, N., Leuschner, I., et al. (2010). Subtype-specific FBXW7 mutation and MYCN copy number gain in Wilms’ tumor. Clin. Cancer Res 16, 2036–2045. https://doi.org/10.1158/1078-0432.ccr-09-2890.
17. Dix, D.B., Fernandez, C.V., Chi, Y.Y., Anderson, J., Mullen, E.A., Geller, J.I., Gratias, E.J., Khanna, G., Kalapurakal, J.A., Seibel, N.L., et al. (2019). Augmentation of therapy for combined loss of heterozygosity 1p and 16q in favorable histology Wilms tumors: a Children’s Oncology group AREN0532 and AREN0533 study report. J. Clin. Oncol. 37, 2769–2777. https://doi.org/10.1200/jco.18.01972.
18. Grundy, P.E., Telzerow, P.E., Breslow, N., Moksness, J., Huff, V., and Paterson, M.C. (1994). Loss of heterozygosity for chromosomes 16q and 1p in Wilms’ tumors predicts an adverse outcome. Cancer Res. 54, 2331–2333.

19. Grundy, P.E., Breslow, N.E., Li, S., Perlman, E.J., Beckwith, J.B., 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 tumors: a report from the National Wilms Tumor Study Group. J. Clin. Oncol. 23, 7312–7321. https://doi.org/10.1200/JCO.2005.01.2799.

20. Gratas, 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. https://doi.org/10.1002/cncr.28239.

21. Gratas, E.J., Dome, J.S., Jennings, L.J., Chi, Y.Y., Tian, J., Anderson, J., Grundy, P., Mullen, E.A., Geier, J.L., Fernandez, C.V., and Perlman, E.J. (2016). Association of chromosome 1q gain with inferior survival in favorable-histology Wilms tumors: a report from the Children’s Oncology Group. J. Clin. Oncol. 34, 3189–3194. https://doi.org/10.1200.JCO.2015.66.1140.

22. Chagtai, T., Zill, C., Dainese, L., Wegert, J., Savola, S., Popov, S., Mifsud, W., Vujanic, G., Seibire, 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 tumors biology consortium study. J. Clin. Oncol. 34, 3195–3203. https://doi.org/10.1200.JCO.2015.66.0001.

23. O’Brien, L.L., Guo, Q., Lee, Y., Tran, T., Benazet, J.D., Whitney, P.H., Vaquero, A., and McMahon, A.P. (2016). Differential regulation of mouse and human nephron progenitors by the Six family of transcriptional regulators. Development 143, 595–608. https://doi.org/10.1242/dev.121715.

24. Spreafico, F., Ciceri, S., Gamba, B., Torri, F., Terenziani, M., Collini, P., Macciardi, F., Radice, P., and Perotti, D. (2016). Chromosomal anomalies at 1q, 3, 16q, and mutations of SIX1 and DROSHA genes underlie Wilms tumor recurrences. Oncotarget 7, 8908–8915. https://doi.org/10.18632/oncotarget.6950.

25. Mugrauer, G., Alt, F.W., and Ekkblom, P. (1988). N-myc proto-oncogene expression during organogenesis in the developing mouse as revealed by in situ hybridization. J. Cell Biol. 107, 1325–1333. https://doi.org/10.1083/jcb.107.4.1325.

26. Lauretti, E., Wilson, A., and Trumpp, A. (2009). Myc’s other life: stem cells and beyond. Curr. Opin. Cell Biol. 21, 844–854. https://doi.org/10.1016/j.ccell.2009.09.006.

27. Xu, J., Wong, E.Y., Cheng, C., Li, J., Sharkar, M.T., Xu, C.Y., Chen, B., Sun, J., Jing, D., and Xu, P.X. (2014). Eya1 interacts with Six2 and Myc to regulate expansion of the nephron progenitor pool during nephrogenesis. Dev. Cell 31, 434–447. https://doi.org/10.1016/j.devcel.2014.10.015.

28. Williams, R.D., Chagtai, T., Alcaide-German, M., Apps, J., Wegert, J., Popov, S., Vujanic, G., van Tinteren, H., van den Heuvel-Eibrink, M.M., Kool, M., et al. (2015). Multiple mechanisms of MYCN dysregulation in Wilms tumour. Oncotarget 6, 7232–7243. https://doi.org/10.18632/oncotarget.3377.

29. Jiménez Martín, O., Schlosser, A., Furtwängler, R., Wegert, J., and Gessler, M. (2021). MYCN and MAX alterations in Wilms tumor and identification of novel N-MYC interaction partners as biomarker candidates. Cancer Cell Int 21, 555. https://doi.org/10.1186/s12935-021-02259-2.

30. Grandori, C., Cowley, S.M., James, L.P., and Eisenman, R.N. (2000). The Myc/Max/Mad network and the transcriptional control of cell behavior. Ann. Rev. Cell Dev. Biol. 16, 653–699. https://doi.org/10.1146/annurev.cellbio.16.1.653.

31. Carroll, P.A., Freie, B.W., Mathyasaraya, H., and Eisenman, R.N. (2018). The MYC transcription factor network: balancing metabolism, proliferation and oncogenesis. Front. Med. 12, 412–425. https://doi.org/10.1186/s11684-018-0650-z.

32. Schaub, F.X., Dhankani, V., Berger, A.C., Trivedi, M., Richardson, A.B., Shaw, R., Zhao, W., Zhang, X., Ventura, A., Liu, Y., et al. (2018). Pan-cancer alterations of the MYC oncogene and its proximal network across the cancer genome atlas. Cell Syst 6, 282–300. https://doi.org/10.1016/j.cels.2018.03.003.

33. Liu, P.Y., Enriquez, D., Marshall, G.M., Tee, A.E., Polly, W., Miu, L., Bell, J.I., Zhang, X.D., Milazzo, G., et al. (2014). Effects of a novel long noncoding RNA, IncUSMycN, on N-Myc expression and neuroblastoma progression. J. Natl. Cancer Inst. 106, dju113, PMID: 24906397. https://doi.org/10.1093/jnci/dju113.

34. Fuji, Y., and Nakamura, M. (2010). FOXK2 transcription factor is a novel G/T-mismatch DNA binding protein. J. Biochem. 147, 705–709. https://doi.org/10.1093/jb/mvq004.

35. Sukonina, V., Ma, H., Zhang, W., Bartesaghi, S., Subhash, S., Heglin, M., Foy, H., Betz, M.J., Nilsson, D., Lidell, M.E., et al. (2019). FOXK1 and FOXK2 regulate aerobic glycolysis. Nature 566, 279–283. https://doi.org/10.1038/s41586-019-0900-5.

36. Mi, W., Guan, H., Lyu, J., Zhao, D., Xi, Y., Jiang, S., Andrews, F.H., Wang, X., Gagea, M., Wen, H., et al. (2017). YEATS2 links histone acetylation to tumorigenesis of non-small cell lung cancer. Nat. Commun. 8, 1088. https://doi.org/10.1038/s41467-017-01173-4.

37. Fievret, A., Belaud-Rotureau, M.A., Dugay, F., Abadie, C., Henry, C., Taque, S., Andrieux, J., Guyetant, S., Robert, M., Dubourg, C., et al. (2013). Involvement of germline DDX1-MYCN duplication in inherited nephroblastoma. Eur. J. Med. Genet. 56, 643–647. https://doi.org/10.1016/j.ejmg.2013.10.004.

38. Tsialikas, J., and Romer-Seibert, J. (2015). LIN28: roles and regulation in development and beyond. Development 142, 2397–2404. https://doi.org/10.1242/dev.117580.

39. Urbach, A., Yermalovich, A., Zhang, J., Spina, C.S., Zhu, H., Perez-Atayde, A.R., Shukrun, R., Charlot, J., Sebire, N., Mifsud, W., et al. (2014). LIN28 sustains early renal progenitors and induces Wilms tumor. Genes Dev. 28, 971–982. https://doi.org/10.1101/gad.237149.113.

40. Viswanathan, S.R., Powers, J.T., Einhorn, W., Hoshida, Y., Ng, T.L., Toftanin, S., O’Sullivan, M., Lu, J., Phillips, L.A., Lockhart, V.L., et al. (2009). LIN28 promotes transformation and is associated with advanced human malignancies. Nat. Genet. 41, 843–848. https://doi.org/10.1038/ng.392.

41. Chang, H.M., Triboulet, R., Thornton, J.E., and Gregory, R.I. (2013). A role for the Perlman syndrome exonuclease Dis3l2 in the Lin28-let-7 pathway. Biochem. Cell Biol. 16.1.653.
46. Chapman, M.A., Lawrence, M.S., Keats, J.J., Cibulskis, K., Sougnez, C., Schinzler, A.C., Harview, C.L., Brunet, J.P., Ahmann, G.J., Adl, M., et al. (2011). Initial genome sequencing and analysis of multiple myeloma. Nature 471, 467–472. https://doi.org/10.1038/nature09837.

47. Weißbach, S., Langer, C., Puppe, B., Nedeva, T., Bach, E., Kull, M., Bargou, R., Eisene, H., Rosenwald, A., Knop, S., and Leich, E. (2015). The molecular spectrum and clinical impact of DIS3 mutations in multiple myeloma. Br J Haematol. 169, 57–70. https://doi.org/10.1111/bjh.13256.

48. Pertesi, M., Vallée, M., Wei, X., Revuelta, M.V., Galia, P., Demangel, D., Oliver, J., Foll, M., Chen, S., Perial, E., et al. (2019). Exome sequencing identifies germline variants in DIS3 in familial multiple myeloma. Leukemia 33, 2324–2330. https://doi.org/10.1038/s41373-019-0452-6.

49. Towler, B.P., Jones, C.I., Viegas, S.C., Smalley, S.K., Arraino, C.M., and Newbury, S.F. (2013). The 3′–5′ exonuclease activity of DIS3 regulates the expression of specific microRNAs in Drosophila wing imaginal discs. RNA Biol. 12, 728–741. https://doi.org/10.1080/15476286.2015.1040978.

50. Ackermann, S., and Fischer, M. (2019). Telomere maintenance in pediatric Wilms’ tumor. Cancer Res. 79, 4301–4307. https://doi.org/10.1158/0008-5472.CAN-18-3335.

51. Dome, J.S., Chung, S., Bergemann, T., Umbricht, C.B., Saji, M., Carey, J., and Crompton, B.D. (2020). Chasing ctDNA as a biomarker in cancer. Cancer Discov. 10, 15476286.2015.1040978.

52. Abbou, S.D., Shulman, D.S., DuBois, S.G., and Crompton, B.D. (2019). Genomics in the Cloud: Using Docker, GATK, and WDL in Terra, 1st Edition (O’Reilly Media).

53. Chu, A., Robertson, G., Brooks, D., Mungall, A.J., Birol, I., Coope, R., Ma, Y., Jones, S., and Marra, M.A. (2016). Large-scale profiling of microRNAs in the cancer genome atlas. Nucleic Acids Res. 44, e3, PMID: 26271990. https://doi.org/10.1093/nar/gkw085.

54. Van der Auwera, G.A., and O’Connor, B.D. (2020). Next-generation sequencing: the association for molecular pathology. Genet. Med. 20, e2795. https://doi.org/10.1038/s41436-018-0207-4.

55. McMullen, J.R., Chang, E., Broome, S.G., and Crompton, B.D. (2020). Assessing the association between somatic mutations in pediatric cancer patients with sarcoma. Am. Soc. Clin. Oncol. Educ. Book. 40, e351–e360. https://doi.org/10.1200/edbk.820969.

56. Abbou, S.D., Shulman, D.S., DuBois, S.G., and Crompton, B.D. (2019). Assessment of circulating tumor DNA in pediatric solid tumors: the promise of liquid biopsies. Pediatr. Blood Cancer 66, e2795. https://doi.org/10.1002/bpc.27595.

57. Madanat-Harjuoja, L.M., Renfro, L., Kiega, K., Tornwall, B., Diller, L.R., Fernandez, C.V., et al. (2022). Circulating tumor DNA as a biomarker in favourable histology Wilms Tumor (WT) patients. J. Clin. Oncol. JCO 20.00098. Published on line May 17, 2022. In press. https://doi.org/10.1200/JCO.20.00098.

58. Nik-Zainal, S., Davies, H., Statt, J., Ramakrishna, M., Glodzik, D., Zou, X., Martincorenca, I., Alexandrov, L.B., Martin, S., Wedge, D.C., et al. (2016). Landscape of somatic mutations in 560 breast cancer whole-genome sequences. Nature 534, 47–54. https://doi.org/10.1038/nature17676.

59. 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.

60. Campbell, P.J., Stephens, P.P., Pleasance, E.D., O’Meara, S., Li, H., Santarius, T., Stebbings, L.A., Leroy, C., Edkins, S., Hardy, C., et al. (2008). Identification of somatically acquired rearrangements in cancer using genome-wide massively parallel paired-end sequencing. Nat. Genet. 40, 722–729. https://doi.org/10.1038/ng.128.

61. Chapman, M.A., Lawrence, M.S., Keats, J.J., Cibulskis, K., Sougnez, C., Schinzler, A.C., Harview, C.L., Brunet, J.P., Ahmann, G.J., Adl, M., et al. (2011). Initial genome sequencing and analysis of multiple myeloma. Nature 471, 467–472. https://doi.org/10.1038/nature09837.

62. Chapman, M.A., Lawrence, M.S., Keats, J.J., Cibulskis, K., Sougnez, C., Schinzler, A.C., Harview, C.L., Brunet, J.P., Ahmann, G.J., Adl, M., et al. (2011). Initial genome sequencing and analysis of multiple myeloma. Nature 471, 467–472. https://doi.org/10.1038/nature09837.
variation in 141,466 humans. Nature 581, 434–443. https://doi.org/10.1530/ey.17.14.3.

76. Whirl-Carrillo, M., McDonagh, E.M., Hebert, J.M., Gong, L., Sangkuhl, K., Thorn, C.F., Altman, R.B., and Klein, T.E. (2012). Pharmacogenomics knowledge for personalized medicine. Clin. Pharmacol. Ther. 92, 414–417. https://doi.org/10.1038/clpt.2012.96.

77. Lever, J., Barbarino, J.M., Gong, L., Huddart, R., Sangkuhl, K., Whaley, R., Whirl-Carrillo, M., Woon, M., Klein, T.E., and Altman, R.B. (2020). PGxMine: text mining for curation of PharmGKB. Pac. Symp. Biocomput. 25, 611–622.

78. Taliun, D., Chothani, S.P., Schönherr, S., Forer, L., Boehnke, M., Abecasis, G.R., and Wang, C. (2017). LASER server: ancestry tracing with genotypes or sequence reads. Bioinformatics 33, 2056–2058. https://doi.org/10.1093/bioinformatics/btx075.

79. Wickham, H. ggplot2: Elegant Graphics for Data Analysis. In Springer-Verlag New York. ISBN 978-3-319-24277-4, 2016
## STAR★METHODS

### KEY RESOURCES TABLE

| REAGENT or RESOURCE | SOURCE | IDENTIFIER |
|---------------------|--------|------------|
| **Biological samples** | | |
| Samples from normal blood or tissue, primary tumor, relapse tumor from patients with Wilms Tumor | Children’s Oncology Group | childrensoncologygroup.org |
| **Critical commercial assays** | | |
| AllPrep DNA/RNA kit | Qiagen | Cat# 80204 |
| mirVana miRNA Isolation Kit | Invitrogen | Cat# AM1560 |
| Roche High Pure miRNA Isolation Kit | Roche | Cat# 05080576001 |
| QiAamp DNA Blood Midi kit | Qiagen | Cat# 5183 |
| Quant-IT PicoGreen quantification assay kit | Invitrogen | Cat# P7589 |
| RNA 6000 NanoChip Kit | Agilent | Cat # 5067-1511 |
| Complex iPLEX Gold Genotyping Reagent | Agena | Cat #10158 |
| Extend Primer Mix | Integrated DNA Technologies (IDT) | N/A |
| Ampure XP SPRI Beads | Beckman Coulter | Cat# A63881 |
| Hyper Prep without amplification | KAPA Biosystems/Roche | KK8505 |
| Unique Dual-Index Adapters | KAPA Biosystems/Roche | Cat# KK8727 |
| Quantitative PCR kit (library quantification) | KAPA Biosystems/Roche | Cat# KK4835 |
| Stranded Total RNA Prep with RiboZero Gold | Illumina | Cat# 20020599 |
| IDT for Illumina – TruSeq RNA UD indexes | Illumina | Cat# 20022371 |
| NEXTflex Small RNA-Seq Kit v.3 | PerkinElmer | Cat# NOVA-5132-06 |
| HyperPrep Kit with Amplification | KAPA Biosystems/Roche | Cat # KK8504 |
| xGen™ UDI-UMI adapters | Integrated DNA Technologies | N/A |
| HiFi HotStart Ready Mix | KAPA Biosystems/Roche | KK2602 |
| p5 and p7 primers | Integrated DNA Technologies | N/A |
| Dual Index F&R primers | Integrated DNA Technologies | Cat# 100981K |
| xGen™ Hybridization and Wash Kit | Integrated DNA Technologies | Cat# 1080584 |
| xGen™ Universal Blockers | Integrated DNA Technologies | Cat# 1075476 |
| CustomPanel bait | Twist Biosciences | Custom |
| Platinum Hot-start PCR kit | Invitrogen | Cat# 13000012 |
| ExoSAP-IT | Applied Biosystems | Cat# 78200.200.UL |
| Quant-it dNTP quantification assay kit (Picogreen) | Invitrogen | Thermo Science catalogue Q33130 |
| BigDye X Terminator Purification Kit | Applied Biosystems | Cat# 4376486 |
| **Deposited data** | | |
| Raw and Analyzed Data (Project Publication Page: MP2PRT-WT) | NCI Genomic Data Commons | https://gdc.cancer.gov/about-data/publications/MP2PRT-WT-2022; https://portal.gdc.cancer.gov/projects/MP2PRT-WT |
| gnomAD | Karczewski et al.2020 | gnomAD (broadinstitute.org) |
| PharmGKB | Whirl-Carrillo et al.2012 | PharmGKB (pharmgkb.org) |
| PGxMine | Lever et al.2020 | PGxMine (pharmgkb.org) |
| **Oligonucleotides** | | |
| Primers for verification of variants involving MGA, WT1, MCFC1, RBL1, COBLL1 | This paper (Methods) | N/A |
| Primer Target locations for Validation set | This paper (Table S4) | N/A |

(Continued on next page)
**RESOURCE AVAILABILITY**

**Lead contact**
Further information and requests for resources and reagents should be directed to and will be fulfilled by the lead contact, Elizabeth J. Perlman, MD (eperlman@luriechildrens.org).

| REAGENT or RESOURCE | SOURCE | IDENTIFIER |
|---------------------|--------|------------|
| **Software and algorithms** | | |
| Picard | Broad Institute | https://broadinstitute.github.io/picard/ |
| CaVEMan | Nik-Zainal et al., 2016 | https://github.com/cancerit/CaVEMan |
| Pindel | Ye et al., 2009 | https://github.com/genome/pindel |
| BRASS | Campbell et al., 2008 | https://github.com/cancerit/BRASS |
| AscatNGS | Raine et al., 2016 | https://github.com/cancerit/ascatNgs |
| bc12fastq | Illumina | v.2.20.0 |
| Bioinformatics Pipeline: mRNA analysis | GDC | https://docs.gdc.cancer.gov/Data/Bioinformatics_Pipelines/Expression_mRNA_Pipeline/ |
| STAR two-pass | Dobin et al. 2013 | N/A |
| HTSeq | Anders et al. 2015 | v.0.6.1.p1 |
| Small RNA Quantification pipeline | Chu et al. 2015 | N/A |
| GATK Mutect 2 tumor-only pipeline | Van der Auwera et al.2020 | Mutect2 – GATK (broadinstitute.org) |
| Ensembl Variant Effect Predictor (VEP) | McLaren et al. 2016 | Variant Effect Predictor - Homo_sapiens - Ensembl genome browser 105 |
| samtools mpileup | Danecek et al. 2021 | N/A |
| Oncoprinter | Cerami et al., 2012 | cBioPortal for Cancer Genomics::Oncoprinter |
| cBioPortal | Gao et al., 2013 | cBioPortal for Cancer Genomics |
| DRAGEN Germline v.3 pipeline | Illumina | DRAGEN Germline (illumina.com) |
| GATK Haplotype Caller pipeline | Van der Auwera et al.2020 | HaplotypeCaller – GATK (broadinstitute.org) |
| Sequencing Analysis Software v 6.0 | ABI | N/A |
| TRACE/LASER | Taliun et al.2017 | LASER (umich.edu) |
| ggplot2 | Wickham 2016 | https://ggplot2.tidyverse.org |
| **Other** | | |
| FilterMax F5 Multi Mode Microplate Reader | Molecular Devices | N/A |
| Focused-ultrasonicator | Covaris | LE220-Plus |
| Bravo liquid handling platform | Agilent | N/A |
| Viia qPCR machine | Life Technologies (ABI) | N/A |
| NovaSeq 6000 | Illumina | N/A |
| Starlet Liquid Handling System | Hamilton | N/A |
| Tape Station System | Agilent | N/A |
| HiSeq 4000 | Illumina | N/A |
| Liquid Handling Workstation | SciClone | |
| Pippin Prep system | Sage Science Beverly, MA | N/A |
| SPE-DRY 96 | Biotage | N/A |
| Veriti thermal cycler | ABI | N/A |
| 3500XL Capillary sequencer | ABI | N/A |
| Bioanalyzer | Agilent | N/A |
| Agena’s MassARRAY™ System | Agena | N/A |
| High Seq X | Illumina | SY-301-2002 |
Materials availability
This study did not generate new unique reagents.

Data and code availability
- Sequencing data, de-identified human clinical information, and sample quality control information have been deposited on the publication page of this study at the Genome Data Commons (https://portal.gdc.cancer.gov/projects/MP2PRT-WT) GDC and are publicly available as of the date of publication. Accession numbers are listed in the key resources table.
- This paper does not report original code.
- Any additional information required to reanalyze the data reported in this paper is available from the lead contact upon request.

EXPERIMENTAL MODEL AND SUBJECT DETAILS

All patients who were registered as Favorable Histology Wilms tumor (FHWT) on the Children’s Oncology Group AREN03B2 umbrella biology and classification study from 2006 to 2017 who relapsed, and who had samples banked at the Biopathology Center (BPC) with a valid and verified consent were eligible. A discovery set (51 patients, 30 M, 21 F) and an independent validation set (31 patients, 16 M, 15 F) were defined. The clinical and quality control details of all patients and samples, including gender and age can be found in https://portal.gdc.cancer.gov/projects/MP2PRT-WT GDC. All samples were de-identified and the Institutional Review Board of the Ann & Robert H. Lurie Children’s Hospital of Chicago approved the reported studies.

METHOD DETAILS

Eligibility for this study included the following
- Patients registered on AREN03B2 as FHWT who relapsed with FHWT (per central review) following therapy, with a current valid verified consent. (The development of contralateral disease following chemotherapy was not considered relapse for this study.)
- Patients received both vincristine and actinomycin following diagnosis. (Patients who relapsed after surveillance alone or single agent vincristine were not included).
- Patients have relapse sample banked by the Biopathology Center (BPC).

Of patients meeting the above criteria, several had limited availability of a germline or normal tissue comparator sample, or of a relapse sample. To maximally utilize relapse samples, a discovery and a validation set were defined. For the discovery set, the following additional criteria were included: 1) sufficient relapse sample banked at the BPC to perform whole genomic sequencing, and 2) a source of germline DNA (normal kidney or peripheral blood) banked by the BPC. Available samples from the primary tumor prior to therapy were not required for the discovery set, but when available were also included in the analysis. The validation set included all patients that met the first set of criteria but failed one or more of the second set of criteria.

Specimen processing and quality control
Quality control was performed on each tumor specimen from either a frozen section prepared by the BPC or from a permanent section taken from a formalin-fixed, paraffin-embedded (FFPE) block. Hematoxylin and eosin (H&E) stained sections were reviewed to confirm that the tumor specimen was histologically consistent with FHWT. Percent tumor nuclei, percent necrosis, and other pathology annotations were assessed. Tumor samples with \( \geq 60\% \) tumor nuclei and \( \leq 20\% \) necrosis and normal tissue samples with 0% tumor nuclei were submitted for nucleic acid extraction at the BPC.

RNA and DNA were extracted from tissue using a modification of the AllPrep DNA/RNA kit (Qiagen). The flow-through from the Qiagen DNA column was processed using a mirVana miRNA Isolation Kit (Invitrogen) for frozen tissue and High Pure miRNA Isolation Kit (Roche) for FFPE samples. DNA was extracted from blood using the QIAamp DNA Blood Midi Kit (Qiagen). RNA samples were quantified by measuring Abs260 with a UV spectrophotometer and DNA was quantified by Quant-iT PicoGreen Assay Kit using the FilterMax F5 Multi Mode Microplate Reader. DNA specimens were resolved by 1% agarose gel electrophoresis to confirm high molecular weight fragments. A custom SNP panel (using Complex iPLEX Gold Genotyping Reagent from Agena and Extend primary mix from Integrated DNA Technologies) verified that tumor DNA and germline DNA representing a case were derived from the same patient. RNA was analyzed via the RNA6000 Nano assay (Agilent) on the Agilent Bioanalyzer for determinations of an RNA Integrity Number (RIN) for frozen tissue and the DV200 values for FFPE samples. Cases yielding 1.0 \( \mu g \) of tumor DNA from FFPE, 1.2 \( \mu g \) of tumor DNA from frozen tissue, 2.0 \( \mu g \) RNA, and 1.0 \( \mu g \) of germline DNA were preferred in this study; samples with lower yields were also included if they passed all other quality control steps.

The BPC processed tumor samples from a total of 115 cases, of which 85 cases qualified, 11 requiring macrodissection. All qualified cases were sent for genomic analysis. Of the 30 cases that were disqualified, 19 cases failed pathology, 2 cases were too small to extract, 1 case did not have a germline sample available, and 8 did not meet molecular quality metrics.
Sequencing methods

**Discovery set whole genomic sequencing**

PCR-free Whole Genome Sequencing was performed at the Broad Institute of MIT and Harvard.

Genomic DNA (350 ng in 50 μL) was used as the input into DNA fragmentation with acoustic shearing performed using a Covaris focused-ultrasonicator, targeting 385 bp fragments. Following fragmentation, a clean-up step was performed using Ampure XP SPRI beads. Library preparation was performed using a commercially available kit (KAPA Biosystems Hyper Prep without amplification module), and with palindromic forked adapters with unique 8-base index sequences embedded within the adapter (Unique Dual Indexed Adapter Kits, Roche). Following sample preparation, libraries were quantified using quantitative PCR (KAPA Biosystems Quantitative PCR kit), with probes specific to the ends of the adapters using a ViiA7 qPCR machine, and automated using Agilent’s Bravo liquid handling platform. Based on qPCR quantification, libraries were normalized to 2.2 nM and pooled into 24-plexes. The pools were loaded onto NovaSeq 6000 S4 flowcells using the Hamilton Starlet liquid handling system, to produce 151 bp paired-end reads and each sample was sequenced to a coverage of 30x.

Output from Illumina software was processed by the Picard data-processing pipeline to yield CRAM files containing demultiplexed, aggregated aligned reads and submitted to the Genomics Data Commons (GDC). GDC identified somatic DNA variants for each primary and relapse tumor sample using the paired normal sample. Single nucleotide variants (SNVs) were called using CaVEMan, indels using Pindel, structural variants using BRASS, and copy number variation using ASCATngs. (Bioinformatics Pipeline: DNA-Seq Analysis - GDC Docs [cancer.gov]). The following files were generated: aligned harmonized BAM files, raw SNVs in VCF format, raw indels in VCF format, raw structural variants in VCF and browser extensible data paired-end (BEDPE) format, gene-level copy number data in TSV format, and genomic segmented copy number in TXT format (https://portal.gdc.cancer.gov/projects/MP2PRT-WTGDC).

**Discovery set RNA sequencing**

RNA sequencing was performed by the University of North Carolina. Fresh frozen RNA analytes were assayed for RNA integrity, concentration, and fragment size. Samples for total RNA-seq and small RNA-sequencing were quantified on a TapeStation system (Agilent, Inc. Santa Clara, CA). RNA Integrity score (RIN) averaged 8.1. Samples with RINs >8.0 were considered high quality. Input concentrations greater than 100 ng/μl were ideal and the amount of material ranged between 0.85 and 2.52 μg of RNA. Initial fragment size determined if additional fragmentation was needed.

For total RNA-sequencing, library construction was performed using the Stranded Total RNA Prep with RiboZero Gold protocol (Illumina) and Truseq RNA UD indexes (IDT for Illumina) following the manufacturer’s instructions. Libraries were prepared on an Agilent Bravo Automated Liquid Handling System. Quality control was performed at every step and the libraries were quantified using a TapeStation system. Indexed libraries were prepared and run on HiSeq4000 paired end 75 base pairs to generate a minimum of 150 million reads per sample library with a target of greater than 90% mapped reads. Typically, these were pools of three to four samples. The raw Illumina sequence data were demultiplexed and converted to fastq files with bcl2fastq v.2.20.0, and adapter and low-quality sequences were removed.

FASTQ files were submitted to the GDC where the files were processed according to their pipeline (https://docs.gdc.cancer.gov/Data/Bioinformatics_Pipelines/Expression_mRNA_Pipeline/). Briefly, fastq files were aligned to the human genome (hg38) using STAR two-pass, v.2.4.2a. The first pass generates splice junctions to help with the second pass final alignment. Genes were quantified using HTSeq, v.0.6.1p1. The following files were generated: genomic, transcriptomic, and chimeric BAM files, HTSeq raw read counts, fragments per kilobase of transcript per million mapped reads (FPKM), and FPKM-UQ (upper quartile) (https://portal.gdc.cancer.gov/projects/MP2PRT-WTGDC).

For miRNA-sequencing, miRNA-seq library construction used the NEXTflex Small RNA-Seq Kit (v3, PerkinElmer, Waltham, MA). Samples were bar-coded with individual tags following the manufacturer’s instructions. Libraries were prepared on Sciclone Liquid Handling Workstation Quality control was performed at every step, and the libraries were quantified using a TapeStation system and an Agilent BioAnalyzer using the Nextflex Small RNA analysis kit (Perkin Elmer). Pooled libraries were then size selected according to NEXTflex Kit specifications using a Pippin Prep system (Sage Science, Beverly, MA). Typical pool size was 20 libraries. Indexed libraries were loaded on the Hiseq4000 to generate a minimum of ~10 million reads per library with a minimum of 90% reads mapped. The raw Illumina sequence data were demultiplexed and converted to FASTQ files using bcl2fastq v.2.20.0. Resultant data were analyzed using a variant of the small RNA quantification pipeline developed for TCGA. Samples were assessed for the number of miRNAs called, species diversity, and total abundance. Samples passing quality control were uploaded to the GDC repository where the files were processed using the following pipeline: Bioinformatics Pipeline: miRNA Analysis - GDC Docs [cancer.gov].

**VALIDATION SET TARGETED SEQUENCING**

Targeted sequencing was performed at the Broad Institute of MIT and Harvard. For covering variants within a gene, the exons plus 10 bp at exon–intron boundaries were sequenced and DNA variants were called using the GATK Mutect2 tumor-only calling pipeline and were filtered as described for the discovery set. Hotspot mutations underwent direct sequencing. Exon 3 of CTNNB1 was sequenced. Structural variants resulting in small copy number gains or losses (WTX and MYCN) were addressed by sequencing.
MYCN exon 2 and WTX exon 2, which were involved in the segmental copy number changes identified in both the current study and in the TARGET discovery sets. For WT1, all exons were sequenced. The location of all the primer targets analyzed are provided in Table S4.

For library construction, genomic DNA (50-200 ng in 50µL) was used as the input into DNA fragmentation performed acoustically using a Covaris focused-ultrasonicator, targeting 150 bp fragments. Library preparation was performed using a commercially available kit (KAPA HyperPrep Kit with Library Amplification) and IDT’s duplex UDI-UMI adapters. Unique 8-base dual index sequences embedded within the p5 and p7 primers (IDT) were added during PCR. Enzymatic clean-ups were performed using AMPure XP SPRI beads with elution volumes reduced to 30µL to maximize library concentration. Library quantification was performed using the Invitrogen Quant-It broad range dsDNA quantification assay kit with a 1:200 PicoGreen dilution. Following quantification, each library was normalized to a concentration of 35 ng/µL, using Tris-HCl, 10 mM, pH 8.0.

After library construction, hybridization and capture were performed using the relevant components of IDT’s XGen hybridization and wash kit and following the manufacturer’s suggested protocol, with several exceptions. A set of 12-plex pre-hybridization pools were created. These pre-hybridization pools were created by equimole pooling of the normalized libraries, Human Cot-1 (from the XGen hybridization and wash kit), and blocking oligos (xGen Universal blockers, IDT). The pre-hybridization pools underwent lyophilization using the Biotage SPE-DRY. Post lyophilization, the custom target bait (https://www.twistbioscience.com) along with hybridization mastermix was added to the lyophilized pool prior to resuspension. Library normalization and hybridization setup were performed on a Hamilton Starlet liquid handling platform, while target capture was performed on the Agilent Bravo automated platform. Post capture, PCR was performed to amplify the capture material using a mastermix containing HiFi HotStart Ready Mix (Kapa Biosystems), and dual index forward and reverse primers (IDT). Library pools were then quantified using qPCR (Quantitative PCR kit, KAPA Biosystems) on a Viia7 qPCR machine. Based on qPCR quantification, pools were normalized using a Hamilton Starlet to 2 nM. The pools were loaded onto lanes of the HiSeq X sequencer to produce 151 bp paired-end reads and samples sequenced to a coverage goal of 500x MTC. Output from Illumina software was processed by the Picard data-processing pipeline to yield BAM files containing demultiplexed, aggregated aligned reads. Samples passing quality control were uploaded to the GDC repository (GDCHttps://portal.gdc.cancer.gov/projects/MP2PRT-WT).

**Project analytic methods**

**Somatic variant analysis**

For analysis, the raw unannotated simple somatic mutation pindel files (for small indels) and CaVEMan VCF files (for SNVs) were obta

**Somatic copy number analysis**

Gene-level and segmented genomic copy number files from WGS data were obtained from the GDC (https://portal.gdc.cancer.gov/projects/MP2PRT-WT) GDC. Loci of interest based on prior large studies of copy number changes relevant to WT were evaluated. Those variants evaluated as benign or likely benign were filtered out. The general population allelic frequencies provided by ENSEMBL were verified in gnomAD gnomAD (broadinstitute.org); variants not present in gnomAD are labeled “Novel”.

**Germline analysis**

The germline (normal kidney or peripheral blood) CRAM files were processed by the DRAGEN v3 germline pipeline (DRAGEN Germline (illumina.com)) using default settings. The resulting GVCF files were filtered to include 1) 15 genes classified as Wilms tumor predisposition genes (BLM, BRCA2, BUB1B, CDC73, CTR9, DICER1, DIS3L2, GPC3, GPC4, PALB2, PIK3CA, REST, TP53, TRIM37, WT1), 2) three additional genes with germline mutations identified in patients with WT that have been associated with predisposition to adult tumors (CHEK2, MUTYH, PMS2) and 3) the variants contained in the genes identified in Table 1. These variants were annotated using Ensembl VEP, and filtered in the same manner as described for somatic variants. The clinical impact was designated in the categories of benign, likely benign, variant of unknown significance, and pathogenic or likely pathogenic according to American College of Medical Genetics and Genomics and the Association for Molecular Pathology germline variant classification guidelines. Those variants evaluated as benign or likely benign were filtered out. The general population allelic frequencies provided by ENSEMBL were verified in gnomAD gnomAD (broadinstitute.org); variants not present in gnomAD are labeled “Novel”.
Variant verification
For verification of recurrent somatic and germline variants, genomic RNAseq bam files were run through the GATK HaplotypCaller pipeline modified for RNAseq [RNAseq short variant discovery (SNPs + Indels) – GATK (broadinstitute.org)]. Missense and in-frame indel variants not detected in RNAseq data were considered not verified and removed. From this filtered list, those genes with somatic or germline variants in more than one patient were identified; those variants not verified by RNAseq underwent Sanger sequencing by the UNC McLendon Clinical Laboratory. In brief, custom primers were developed to amplify the regions flanking the variant of interest. The following custom primers were developed to amplify the regions flanking the five variants of interest: MGA F: 5’AGTTTAGTGCTTGCCACT-3’, MGA R: 5’GCTAGAGGAAGAAGACACCAAGA-3’, WT1 F: 5’CTTTCGGACTTGT-3’, WT1 R: 5’AACACATGGCTGACTCTCTCA-3’. HCFC1 F: 5’GCCCAACTTGCTGTGCTT-3’, HCFC1 R: 5’CAC GTGTTCCACTTGTGTG-3’, RBL1 F: 5’GTTTGGATCAATCGCTTACC-3’, RBL1 R: 5’TCCGAATCTGGGCTCAAGGC-3’, COB LL1 F: 5’CATGTAAGAAAAGCGAGACACAG-3’, COBLL1 R: 5’GGTACACTGCCTCATCCAAA-3’. PCR was performed on Platinum HOT START PCR (Invitrogen) using an ABI Veriti thermal cycler and amplicons cleaned using ExoSAP-IT. The amplicons were sequenced on an ABI 3500XL capillary sequencer using either the initial PCR primers or ones nested within the target region and BigDye XTerminator chemistry. Variants were confirmed using Sequencing Analysis Software Version 6.0 (ABI).

Analysis of targeted sequencing
The same parameters were used for filtering the discovery set WGS were applied to the targeted capture validation variant list, with the exception that 10% allelic fraction was required. Recurrent copy number variants for WTX, MYCN, and WT1 were evaluated for each tumor by 1) identifying control genes shown to have a stable copy number in all the discovery relapse samples (TBC1D1 and MDD); 2) determining the median read count of high-quality reads (mapping quality ≥60 and base quality ≥20) for TBC1D1 and MDD, and of the other genetic locus being tested; 3) normalizing the read count for each genetic locus using the average read count of the two control genes in that tumor; 4) establishing gain/loss calls by determining the median normalized read count for each gene across all samples and applying 25% gain or loss levels. The median normalized read count for WTX was determined separately for males and females.

PHARMACOGENOMIC ANALYSIS
The PharmGKB (PharmGKB) and PGxMine (PGxMine pharmgkb.org) databases were filtered to include only pharmacogenomic variants annotated as associated with dactinomycin, doxorubicin, or vincristine (n = 226 unique variants). The germline GVCF files were filtered to include only those 226 variants. Polymorphisms were retained if the genotype quality score was ≥20 and the missing rate was <10%. The frequency of the remaining polymorphisms within the normal sample of the discovery patients was compared with the general population frequency in gnomAD (gnomAD broadinstitute.org) using the binomial test in R. Ancestry-specific binomial tests were run using the estimated ancestries determined by TRACE. The polymorphisms were also evaluated for reduction to homozygosity in the tumor samples.

QUANTIFICATION AND STATISTICAL ANALYSIS
- Fisher’s exact test was used to compare primary tumor samples (n = 45) to the relapse tumor samples (n = 51). A p value < 0.05 was required for significance. These findings are provided in the Results section.
- Fisher exact test was used to compare the frequencies of mutations identified as recurrent in this study (Table 1) in relapse samples (n = 82) to the TARGET dataset (n = 533, except where n = 56 as noted in Table 1). A p value < 0.05 was required for significance. These results are provided in Table 1.
- Fisher exact test was used to determine the correlation between gain of chromosome 1q with either gain of chromosome 12 or gain of chromosome 18. A p value < 0.05 was required for significance. These findings are provided in the Results section.
- The binomial test was used to compare the allelic frequencies of polymorphisms of interest (n = 222) in the gnomAD general population to WT patients. Four different binomial comparisons were performed: (1) germline allelic frequencies for all WT patients in this study (n = 51 samples) compared to the gnomAD general population, (2) germline allelic frequencies for all WT in this study of European ancestry (n = 37 samples) to the gnomAD Non-Finnish European population, (3) primary tumor allelic frequencies (n = 45 samples) to the gnomAD general population, and (4) relapse tumor allelic frequencies (n = 51 samples) to the gnomAD general population. Multiple testing correction was performed using the Benjamini and Hochberg False Discovery Rate method. An adjusted p value < 0.05 was required for significance. This information is provided in the Results section.
- To evaluate TERT gene expression, htseq-count files for all samples with available RNAseq data (n = 64) were obtained from the GDC, imported into R (version 3.6.3), and normalized using variance stabilizing transformation. The boxplot comparing TERT gene expression in samples with TERT promoter mutation (n = 5) to samples lacking the TERT promoter mutation (n = 59) was generated using ggplot2. The Student’s t-test was used to compare TERT gene expression in these two groups. A p value < 0.05 was considered significant. This information is provided in Figure 2.