HHS Public Access

Author manuscript

Nat Genet. Author manuscript; available in PMC 2018 February 21.

Published in final edited form as:

Nat Genet. 2017 October ; 49(10): 1487–1494. doi:10.1038/ng.3940.

A Children's Oncology Group and TARGET Initiative Exploring the Genetic Landscape of Wilms Tumor

Samantha Gadd[1], **Vicki Huff**[2], **Amy L. Walz**[3], **Ariadne H.A.G. Ooms**[1,4], **Amy E. Armstrong**[3], **Daniela S. Gerhard**[5], **Malcolm A. Smith**[6], **Jaime M. Guidry Auvil**[5], **Daoud Meerzaman**[7], **Qing-Rong Chen**[7], **Chih Hao Hsu**[7], **Chunhua Yan**[7], **Cu Nguyen**[7], **Ying Hu**[7], **Leandro C. Hermida**[5], **Tanja Davidsen**[7], **Patee Gesuwan**[5], **Yussanne Ma**[8], **Zusheng Zong**[8], **Andrew J. Mungall**[8], **Richard A. Moore**[8], **Marco A. Marra**[8,9], **Jeffrey S. Dome**[10], **Charles G. Mullighan**[11], **Jing Ma**[11], **David A. Wheeler**[12], **Oliver A. Hampton**[12], **Nicole Ross**[13], **Julie M. Gastier-Foster**[13], **Stefan T. Arold**[14], and **Elizabeth J. Perlman**[1,*]

[1]Department of Pathology and Laboratory Medicine, Ann & Robert H. Lurie Children's Hospital of Chicago, Northwestern University's Feinberg School of Medicine and Robert H. Lurie Cancer Center, Chicago, Illinois, 60611, USA [2]Department of Genetics, The University of Texas MD Anderson Cancer Center, Houston, Texas, 77030, USA [3]Division of Hematology-Oncology and Transplantation, Ann & Robert H. Lurie Children's Hospital of Chicago, Northwestern University's Feinberg School of Medicine, Chicago, Illinois, 60611, USA [4]Department of Pathology, Princess Maxima Center for Pediatric Oncology, Utrecht, The Netherlands [5]Office of Cancer Genomics, National Cancer Institute, Bethesda, Maryland, 20892, USA [6]Cancer Therapy Evaluation Program, National Cancer Institute, Bethesda, Maryland, 20892, USA [7]Center for Biomedical Informatics and Information Technology, National Cancer Institute, Bethesda, Maryland, 20892, USA [8]Canada's Michael Smith Genome Sciences Centre, British Columbia Cancer Agency (BCCA), Vancouver, British Columbia, V5Z 4S6, Canada [9]Department of Medical Genetics, University of British Columbia, Vancouver, BC, V6H 3N1, Canada [10]Division of Pediatric Hematology/Oncology, Children's National Medical Center, Washington, DC, 20010, USA [11]Department of Pathology, St. Jude Children's Research Hospital, Memphis, TN, 38105, USA [12]Department of Molecular and Human Genetics, Baylor College of Medicine, Houston, Texas 77030, USA [13]Department of Pathology and Laboratory Medicine, Nationwide Children's Hospital, Ohio State University College of Medicine, Columbus, Ohio, 43205, USA [14]King Abdullah University of Science and Technology, Computational Bioscience Research Center, Division of Biological and Environmental Sciences and Engineering, Thuwal, 23955-6900, Saudi Arabia

Correspondence: E.J. Perlman, 225 East Chicago Avenue, Box 17, Chicago, Illinois 60611, eperlman@luriechildrens.org.

Author Contributions: D.S.G., J.M.G.A., M.A.S., P.G., L.H., and T.M.D. oversaw the administrative and data management aspects of the TARGET project; D.M., C.H.H., C.N., C.Y., Q-R.C., C.N., and Y.H. provided the bioinformatic analysis; S.T.A. provided the in silico protein structure analysis; J.M.G-F. and N.R. performed the specimen processing and quality control; J.D. and V.H. provided the samples and the clinical data; C.G.M. and J.M. performed the copy number analysis through level 1; S.G. performed the level 4 analyses of all platforms, wrote the paper with contributions from all other authors; A.L.W. performed the miRNA analysis; A.E.A. performed the level 3 and 4 copy number analysis; A.H.A.G.O. performed the TP53 analysis; Y.M., Z.Z., A.J.M., R.A.M., and M.A.M. performed the target capture sequencing and mRNA sequencing; D.A.W. and O.A.H. performed the whole exomic sequencing; EJP designed and oversaw all aspects of the study.

Competing Financial Interests Statement: The authors declare no competing financial interests.

Abstract

Genome-wide sequencing, mRNA and miRNA expression, DNA copy number and methylation analyses were performed on 117 Wilms tumors, followed by targeted sequencing of 651 Wilms tumors. In addition to genes previously implicated in Wilms tumors (*WT1, CTNNB1, FAM123B, DROSHA, DGCR8, XPO5, DICER1, SIX1, SIX2, MLLT1, MYCN, and TP53*), mutations were identified in genes not previously recognized as recurrently involved in Wilms tumors, the most frequent being *BCOR, BCORL1, NONO, MAX, COL6A3, ASXL1*, *MAP3K4*, and *ARID1A*. DNA copy number changes resulted in recurrent 1q gain, *MYCN* amplification, *LIN28B* gain, and *let-7a* loss. Unexpected germline variants involved *PALB2* and *CHEK2.* Integrated analyses support two major classes of genetic changes that preserve the progenitor state and/or interrupt normal development.

Wilms Tumor is the most common malignant tumor of the kidney in childhood. Discoveries surrounding its pathogenesis have provided important conceptual advances in our knowledge of both oncogenesis and renal development. Until recently, knowledge of the genetic underpinnings of Wilms tumor was largely limited to aberrations of *WT1*, abnormalities of 11p15 methylation, and Wnt activating mutations involving *CTNNB1* and *FAM123*B (also known as WTX, AMER1) [1]. To address this, recent comprehensive genomic analyses of Wilms tumor identified novel mutations involving miRNA processing genes *DROSHA, DGCR8*, and *DICER1*, renal developmental genes *SIX1* and *SIX2*, and *MYCN* [2-4]. Despite these efforts, recurrent driver mutations were not identified in most Wilms tumors. The National Cancer Institute's “Therapeutically Applicable Research to Generate Effective Treatments” (TARGET) initiative enabled the comprehensive characterization of a discovery set of 117 high-risk Wilms tumor cases, defined as those with either favorable histology (FHWT) that relapsed, or those with diffuse anaplasia (DAWT, characterized by the presence of enlarged, hyperchromatic nuclei and atypical mitoses [5]). The variant frequency in a population-based set of Wilms tumor cases was then established in a validation set of 651 tumors (533 FHWT, 118 DAWT) consistently treated on the National Wilms Tumor Study-5 (NWTS-5) clinical trial. Thus far, this effort has reported recurrent mutations involving *DROSHA, DGCR8, XPO5, DICER, SIX1*, and *SIX2* in *FHWT* [6], novel mutations involving *MLLT1* in FHWT [7], and the clinical significance of *TP53* abnormalities in DAWT [8]. The current study provides the consolidated data for all TARGET Wilms tumors (FHWT and DAWT), identifying many additional recurrent novel somatic and germline mutations, recurrent copy number changes (including gain of *LIN28B*, and loss of *let-7a*), and global mRNA and miRNA expression as well as DNA methylation patterns. Through the integration of these analyses, we propose a genetic framework for Wilms tumor within the context of early renal development.

Results

Somatic Mutations in Discovery Set

Bioinformatic analysis of 117 discovery set Wilms tumors identified 1808 high-quality somatic, non-synonymous small exonic variants (mean of 13.6 and 19.7 variants/patient in FHWT and DAWT, respectively). All verified somatic missense and in-frame variants

predicted to be damaging and not identified in 1000 genomes series 3, all known COSMIC mutations, and all verified nonsense and frameshift mutations are provided in Supplementary Table 1; variants in genes recurrently involved, and those involving other selected genes, are provided in Table 1. Genes most impacted were *CTNNB1, DGCR8, DROSHA*, *MLLT1*, *MYCN, SIX1*, *SIX2, TP53, WT1*, and *FAM123B*, with frequencies in keeping with previous publications. An exception was fewer *WT1* variants due to the restriction of the discovery set to high-risk cases (the majority of Wilms tumor with *WT1* mutation are not high risk [1]). Three patients had p. Arg60Gln *MAX* mutations (NM_002382.4: c.179G>A), previously reported in a single Wilms tumor [3]. Genes with variants in more than two tumors and not previously reported in Wilms tumor include *BCOR* (5, all truncating), *NONO* (3, all either p. Arg75Cys [NM_007363.4: c. 223C>T] or p. Arg75His [NM_007363.4: c.224G>A]), and *ACTB*. An additional 16 genes had variants in two Wilms tumors (Table 1). One discovery patient had two somatic *DICER1* mutations, one truncating and one missense (discussed further below).

Bioinformatic analysis for structural variants was performed on 80 tumors analyzed by whole genomic sequencing (WGS). Subsequent DNA copy number analysis or mRNA-Seq verified 54 deletions, 17 tandem duplications, and 11 junctions (Supplementary Table 2). The deletions varied from 229-679 bases (average 412), and most occurred within short interspersed repeats and were randomly distributed across the genome. Deletions were identified involving *WT1* (2), *TP53* (3), *FAM123*B (3), *DIS3L2* (1), and *ACTB* (1). On the other hand, 10/17 duplications encompassed the *MYCN* locus, nine of which involved *NBAS;* all showed *MYCN* gain by copy number analysis.

Germline mutations in Discovery Set

The paired normal sample was evaluated for germline variants within genes identified and listed in Table 1 (see Supplementary Table 1 for details). Germline variants likely of clinical significance (defined as verified nonsense or frameshift variants and missense single nucleotide variants (SNVs) categorized by Clinvar as likely pathogenic or pathogenic) were identified in *XPO5* (1), *TP53* (1), *WT1* (1), *CHEK2* (1), and *PALB2* (1). Additional germline novel SNVs predicted to be damaging by ≥4 predictors, and in-frame variants predicted to be damaging by PROVEAN (see Methods), include *PALB2* (2), *CHEK2* (1), *HDAC4* (1), *WT1* (1), and *EP300* (1). Additional germline SNVs of unclear significance were identified in *DICER1* (2 in the same patient*), CHEK2, DROSHA, NF1, ARID1A*, and *MLLT1*. Of note is a germline splice site variant in *DIS3L2*, a *let-7* miRNA processing factor responsible for Perlman syndrome [9].

Mutations in Validation Set Tumor Samples

Targeted capture sequencing for 37 genes identified from the discovery set (and listed in Table 1) was performed on the validation set of 651 Wilms tumor samples. As described in the Methods section, the discovery and validation sets are both based upon NWTS-5 samples, and are not independent. Therefore, all frequencies refer only to the validation set. Overall, 440/651 carried mutations in the genes analyzed that were predicted to be damaging as defined in online Methods (Table 1, Supplementary Table 3 for details). Our experience with the discovery set would suggest that the remaining tumors had driver mutations in

genes not detected as recurrent in the discovery set, and therefore not analyzed in the validation set, or had copy number gains or losses involving important genes, discussed further below. Mutation distribution by histology and concurrent mutations are provided in Supplementary Table 4. Mutations in *TP53* were significantly associated with DAWT histology (56/118 DAWT and 9/533 FHWT; Fisher's exact test $p < 1\times10e\text{-}10$). These were identified concurrent with all high-frequency mutations (Table 1), most often with a lower allelic fraction, consistent with the role of *TP53* as a secondary mutation in Wilms tumor [8, 10]. *CTNNB1* represents the gene most commonly mutated, with 95 mutations (62 and 20 involving exons 3 and 8, respectively) in 88 validation set patients. Five patients had the same pair of variants involving nearby positions 41266133 (NM_001904.3: c.130C>G; COSMIC 17661) and 41266136 (NM_001904.3: c.133T>C; COSMIC 5663); both variants occurred on the same strand, suggesting they resulted from the same mutational event. Mutations in *CTNNB1* were significantly associated with favorable histology (86/533 FHWT and 2/118 DAWT; Fisher's exact test $p= 2.2E\text{-}06$). Analysis of genes concurrently mutated (Supplementary Table 4) revealed significant mutational co-occurrence of *WT1* with *CTNNB1* (34/88 patients with *CTNNB1* mutation also had *WT1* mutation and 34/46 patients with *WT1* mutation also had *CTNNB1* mutation; Fisher's exact test $p=0.0001$), as previously reported [11], and *DROSHA* with *SIX1/2* mutations (10/67 patients with *DROSHA* mutation also had mutation in *SIX1* or *SIX2* and 10/43 patients with *SIX1* or *SIX2* mutations also had *DROSHA* mutation; Fisher's exact test $p=0.0012$). *CTNNB1* mutation was concurrent with many recurrent mutations, and all cases except *FAM123B* had the same preference for exon 3. In four tumors with both *CTNNB1* and *FAM123B* mutations, *CTNNB1* exons 3, 7, 9, and 13 were involved. Mutation location relative to known protein domains are illustrated in Supplementary Figure 1.

Sixteen validation patients had 21 *DICER1* mutations. Wilms tumors containing two *DICER1* mutations have previously been reported [12,13]; therefore, *DICER1* mutations were examined in detail within the validation set. Only two validation set patients (one of which was also in the discovery set) had the truncating *DICER1* mutations combined with missense mutations in the RNAse III domain described in patients with Pleuropulmonary Blastoma Syndrome [14,15]. The *DICER1* RNAse III domain encodes a metal-binding site [16], critical for microRNA cleavage. In the remaining three tumors with two *DICER1* mutations, all had RNAse III missense mutations, accompanied by missense mutations of unknown significance in either the PAZ domain, the DEXD/H domain, or a splice site (details in Supplementary Table 3). Three additional cases had one missense mutation in the RNAse III domain with high allelic frequencies of this mutation (0.63, 0.75 and 0.97), suggesting loss of heterozygosity at this locus, similar to a previous report [13]. Therefore, biallelic involvement of *DICER1* (either somatic or germline) may be seen in up to 8/651 (1%) of patients with Wilms tumor.

Eight tumors had missense *CHEK2* mutations: four were rs17879961 variants (NM_007194.3: c.470T>C) and were classified as likely pathogenic by Clinvar. Eight *PALB2* mutations were identified in eight tumors: two frameshift deletions, two in-frame deletions, one nonsense variant, and 3 SNVs classified as uncertain or likely benign by Clinvar. *PALB*2 variants were preferentially identified in DAWT (3/533 FHWT and 5/118 DAWT; Fisher's exact test $p=0.006$) (Supplementary Table 3). Six tumors carried *EP300*

variants, five were the same germline in-frame deletion identified in the discovery set. The majority of the 19 *NF1* variants were missense mutations judged damaging by 2-3 of five algorithms and of unclear functional significance by Clinvar. *DIS3L2* was not included in the target capture sequencing.

Computational structure analysis

Protein structural models of genes with recurrent hot-spot mutations not previously analyzed were retrieved from the Protein Data Bank (PDB), or obtained by homology modelling (Supplementary Figure 2). MAX binds DNA as a heterodimer with MYCN or MYCC. Arg60 of MAX, located in the C-terminus of the basic helix-loop-helix domain, interacts non-specifically with the DNA phosphate backbone. The loss of the positive charge due to the p. Arg60Gln substitution is expected to weaken DNA interactions. NONO forms dimers with two RNA recognition motifs (RRM1 and 2) arranged in a particular constellation required for the 3D recognition of structured RNA substrates [17]. The NONO p. Arg75Cys and p. Arg75His substitutions (located in the hinge region between RRM1 and RRM2) are expected to disrupt the interactions between the two domains, leading to increased flexibility and protein destabilization, and thus may affect RNA recognition. MAP3K4 Gly1366 is located in a tight loop between the 2nd and 3rd beta strand of the kinase domain. Although Gly1366 is over 20Å away from the kinase active site, its replacement by arginine may allosterically change or deregulate the kinase activity. Within ACTB (beta actin), all mutations except p. Ile282del (NM_001101.3: c.846_848delCAT) cluster on the same surface region located toward the plus end. This region is not directly implicated in F-actin formation, but is accessible to F-actin and may constitute a binding site for actin modulating factors. P. Ile282del and p. Lys326del (NM_001101.3: c.977_979delAGA) are predicted to affect loops implicated in F-actin formation and hence likely affect F-actin stability. ACTB is a component of the BAF and MLL5-L chromatin remodeling complexes, and interacts with XPO6, Ran and ERBB2. Lack of known homologous structures for BRD7 precludes structure-based investigations.

Segmental chromosomal gains and losses

Analysis of the discovery set confirms previous reports that Wilms tumors characteristically have gains and losses of entire chromosomes or chromosomal arms, particularly gain of 1q, 6, and 12, and loss of 4q, 16q, 17p, 14, 11, and 22 (Supplementary Table 5). Gain of 1q was identified in 56/117, a frequency in keeping with previous reports [18,19]. Gain of 1q was not preferentially concurrent with mutations in Table 1, consistent with its suspected role as a secondary event. Amplification at 2p24 including *MYCN* (defined as log2 segment mean >2.0 and/or presence of structural variants) was identified in 19 tumors (9/78 FHWT and 10/39 DAWT), similar to frequencies previously reported [20]. Loss of 17p significantly correlated with *TP53* mutation as well as loss of 4q and 14q, as previously reported within TARGET tumors [8]. Genes in Table 1 were reviewed for copy gains and/or losses. Unexpectedly, gain (log2 mean >0.5) of chromosomal segments containing 6q16, the *LIN28B* location, was identified in 29/117 tumors, with 24/29 showing a log2 segment mean >1; the mechanism in all cases was gain of the entire chromosome 6. DNA copy changes for the chromosomal locations of *let-7a* family members revealed loss of *let-7a1* (9q22) in 6/117 (2/78 FHWT, 4/39 DAWT), *let-7a2* (11q24) in 22/117 (13/78 FHWT, 9/39 DAWT), and

let-7a3 (22q13) in 26/117 (8/78 FHWT, 18/39 DAWT), annotated in aggregate for FHWT in Figure 1.

Gene expression analysis

Unsupervised hierarchical analysis of the 76 FHWT with available gene expression data resulted in k=6 clusters (non-negative Matrix Factorization [NMF] cophenetic correlation 0.92). Figure 1 illustrates the most common somatic mutations, copy changes, and pathologic features ordered according to the NMF gene expression clustering. The most commonly recurrent variants revealed preferential cluster membership, with *LIN28B* gain and *let-7a* loss residing in cluster 1, *DROSHA, DGCR8, SIX1*, and *SIX2* mutations residing in cluster 2, and *MLLT1, WT1, CTNNB1*, and *WTX* mutations residing in clusters 3 and 4. Similarly, blastemal predominant tumors and perilobar nephrogenic rests predominated in cluster 2, while triphasic (mixed) tumors and intralobar nephrogenic rests predominated in clusters 3 and 4. Gene set enrichment analysis (GSEA) comparing tumors of each cluster with all other tumors yielded significantly enriched gene lists (Supplementary Table 6). Genes present in both the 100 top differentially expressed genes identified by GSEA for each cluster and in the significantly enriched gene lists for that cluster are illustrated in Figure 1. In cluster 1, genes encoding targets of E2F transcription factors, mitotic spindle assembly proteins, G2/M checkpoint proteins, and 1q genes were significantly enriched. E2F3 activation has previously been reported in Wilms tumors [21]. In cluster 4, an overlapping set of genes involved in myogenesis, epithelial-to-mesenchymal transition, and components of the apical junction complex were enriched, in concordance with previous reports showing a subset of high-risk Wilms tumors characterized by low *WT1* expression and increased muscle differentiation (most Wilms tumor with low *WT1* expression and muscle differentiation are low risk) [1]. Cluster 5 tumors showed enrichment of genes involved in oxidative phosphorylation (Figure 1).

Examination of representative genes critical to renal development revealed up-regulation of pre-induction metanephric mesenchyme (MM) genes and low expression of post-induction MM genes in the majority of Wilms tumors, whereas genes up-regulated following Wnt activation (including *DKK2, WIF1, AXIN2, CCND1* and *MYC*) were expressed predominately in those clusters containing Wnt activating mutations (Figure 1). Finally, increased expression of *LIN28B* correlated with the prevalence of *LIN28B* DNA copy gain in cluster 1. Increased *LIN28B* expression was also identified in clusters 3 and 4; of interest, Wnt activation has been shown to result in increased mRNA expression of *LIN28B* [22]. *LIN28A* was not expressed in most tumors; the functions of *LIN28A* and *LIN28B* have been reported to be different [23].

Global methylation analysis

In 78 discovery FHWT, unsupervised analysis of methylation probes that significantly correlated with gene expression (unpaired *t*-test $p < 0.05$, both negative and positive), resulted in k=4 clusters (NMF cophenetic correlation 0.965). Tumors with *MLLT1* mutations dominated methylation cluster 1, whereas tumors with *WT1* and *DROSHA/SIX* variants were in methylation clusters 2 and 3, respectively (Supplementary Figure 3 and annotation of Figure 1). Genes with localized, coordinately differentially methylated islands

were identified (see methods; Supplementary Table 7). Differences in methylation of these islands were small, and no clear pattern could be identified. Genes previously reported to have methylation changes in Wilms tumors [24], and the genes in Table 1, showed only isolated probes that correlated with gene expression. The exception was methylation of the ICR1 region of 11p15, which significantly correlated with *IGF2* expression (General Linear Model FDR 9.33E-18). Loss of heterozygosity (LOH), loss of imprinting (LOI), and retention of imprinting (ROI) were identified in 35%, 41%, and 23%, respectively, of the 116 evaluable discovery set tumors (Figure 2A, annotated in Figure 1).

MicroRNA analysis

Mutations involving miRNA processing genes within Wilms tumor have been previously shown to result in significantly decreased miRNAs, particularly *let-7a*, the *miR-200* family, and *mir-181* [3,6]. To determine the overall patterns of miRNA expression within Wilms tumors, unsupervised analysis of miRNAseq performed on 78 discovery FHWTs resulted in four clusters (NMF cophenetic coefficient=0.9511, Supplementary Figure 4). Only tumors with *MLLT1* mutations clearly clustered (methylation cluster 1). Comparing the miRNA expression between FHWTs with *MLLT1* mutations (N=7) and those without (N=71), revealed striking up-regulation of *mir-10a* (fold-change 3.3) and down-regulation of *mir-10b* (fold-change 0.4) (Figure 2B). The mir-10 family is co-localized and co-regulated with Hox genes, and is dysregulated in a variety of cancers [25]. *Let-7a* expression assessed by both miRNA sequencing within this study, and by quantitative RT-PCR performed previously [6] showed significantly increased *let7a* expression in tumors of gene expression clusters 3 and 4 compared with those of clusters 1 and 2 (Figure 2C). Lastly, one puzzling aspect is the lack of clustering of *DROSHA*-mutant tumors with *DGCR8*-mutant tumors by miRNA expression analysis. Possible explanations lie within the recent appreciation that the functions of the DROSHA protein extend beyond miRNA processing, and include the recognition and cleavage of mRNAs and potentially ribosomal RNAs, as well as modification of its own expression [26-28].

The above gene expression, methylation, and miRNA analyses performed on FHWT were also performed on both the combination of FHWT with DAWT, and on DAWT alone. These analyses did not yield further data of biologic relevance, likely due to the inclusion of tumors with *TP53* mutation and accompanying genomic instability.

Discussion

Comprehensive sequencing has illustrated that pediatric embryonal neoplasms typically arise following a limited number of genetic aberrations, an observation we confirm in Wilms tumors. In addition, our study shows that Wilms tumors 1) commonly arise through more than one genetic event, 2) show differences in gene expression and methylation patterns based upon different genetic aberrations, 3) have a large number of candidate driver genes, most of which are mutated in <5% of Wilms tumors, and 4) have recurrently mutated genes with common functions, with the majority involved in either early renal development or epigenetic regulation of transcription (chromatin modifications, transcription elongation, and miRNAs).

Early renal progenitor cells actively proliferate under the influence of genes such as *Eya1, Six1, Six2, Sall1*, and *Cited1*, and resist Wnt-directed differentiation [29,30]. A segment of this population becomes susceptible to the influence of Wnt9b (secreted by the ureteric bud), resulting in a switch from bivalently modified histones to permissive histone marks on *Wnt4, Fgfr8, Pax8, Cdh4, Lef1*, gain of repressive histone marks on renewal genes *Six2, Osr1* [31], and ultimately mesenchymal-to-epithelial transition (MET), known as induction. At least some of these cellular mechanisms are directed by Wt1 which binds the Wnt4 transcription start site and recruits Crebbp and Ep300, proteins involved in histone acetylation [32]. This Wt1/Crebbp/Ep300 protein complex is required for a shift from repressive to active chromatin at the *Wnt4* promoter, and for a change in the phosphorylation states of RNA polymerase II, allowing for the switch from paused transcription to transcription elongation [33]. (CBP and p300 also promote H3 acetylation at the *WT1* intronic enhancer, thereby increasing *WT1* transcription, while HDAC4 represses *WT1* transcription [34]).

Given this context, we demonstrate robust expression of pre-induction developmental genes, and low expression of post-induction genes in most Wilms tumors, consistent with the expected phenotype of pre-induction renal progenitor cells (Figure 1). This is consistent with previous studies showing expression of genes characteristic of early renal development within Wilms tumor blastemal cells and within self-renewing human Wilms tumor stem cells [35,36]. In Wilms tumor, the differentiation arrest is not complete, allowing for maturing lineages in varying proportions [37]. We also demonstrate recurrent mutations in several genes known to be critical to early renal development, including *SIX1/2, WT1, EP300, CREBBP*, and *MYCN*, while mutations in other genes encode proteins that mediate key histone modifications important in renal development (*MLLT1, BCOR, MAP3K4, BRD7, HDAC4*) (Table 1). Evidence of Wnt signaling is largely confined to those groups of Wilms tumors having Wnt activating mutations. These different genetic events are associated with two processes, preservation of the progenitor state (exemplified by clusters 1 and 2) and abnormal induction (exemplified by clusters 3 and 4). For clarity, these two patterns are discussed separately, although the regulatory processes governing the transition from the progenitor state through induction are complex and intertwined.

With regard to preservation of the progenitor state, *SIX1* and *SIX2* are critical for progenitor renewal. Loss of *Six2* leads to epithelial differentiation and rapid exhaustion of nephron progenitors [38]. The overwhelming predominance of the specific p. Gln177Arg *SIX1/2* mutations in Wilms tumors, the observation that these mutations result in up-regulated cell-cycle genes [6], and the structural analysis of these mutations previously reported [3], all support an activating function for these mutations that are thus far specific to Wilms tumor. MYCN also mediates proliferation of nephron progenitor cells through several mechanisms [39, 40]. SIX2 and EYA1 directly bind and dephosphorylate T58 of MYCN, preventing MYCN degradation and exit from the cell cycle [41]. We found both *MYCN* amplification and activating p. Pro44Leu/His mutations (NM_005378.4: c.131C>T), identified previously in Wilms tumor by others [6, 20]. Mutations in *MAX* and *NONO* may result in cellular impacts similar to that seen with *MYCN* over-expression. 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 [42]. The *MAX* p. Arg60Gln mutation has also

been identified in colon cancer, leukemia, and glioma, and lies within the conserved bHLH-Zip domain involved in protein–protein interactions and DNA binding (COSM166665). NONO, an RNA-binding protein, binds to MYCN leading to post-transcriptional up-regulation of *MYCN* mRNA and protein expression [43].

Mutations involving the miRNA processing genes also appear to perpetuate the progenitor state. They result in global reduction of mature miRNAs, in particular *let-7a*, which is a potent mediator of differentiation during early development [3,5,6]. Decreased *let-7a* expression may also result from up-regulation of *LIN28B*, which encodes a protein that specifically binds pri/pre-*let-7* miRNAs, preventing their maturation [23,44,45]. Gain of *LIN28B* has been mechanistically tied to Wilms tumor development: over-expression of *LIN28B* within the pre-induction MM of mice results in expansion of progenitor cells (due to sustained proliferation and delay of maturation), and formation of Wilms tumor, a process rescued by *let-7* over-expression [44]. We document increased DNA copy number of *LIN28B* in 25% of Wilms tumor and DNA copy loss of *let-7a* in 46%. Of interest, while mutations in miRNA processing genes were identified preferentially in gene expression cluster 2, gain of *LIN28B* and loss of *let-7a* were more frequent in cluster 1. This separate clustering by gene expression was not accompanied by separate clustering by global miRNA expression. This common pattern of miRNA expression, despite the different pathogeneses, may result from similar direct impacts on a subset of miRNAs (such as *let-7a*); alternatively, this may reflect common down-stream effects on miRNA expression within early renal progenitor cells. The differences in gene expression between clusters 1 and 2 may be due to tumorigenesis during different developmental windows. This is suggested by the observation that clusters 1 and 2 differ in the types of precursor lesions (nephrogenic rests) as well as dominant histologic features observed. These differences will be important to investigate in the future.

With regard to abnormal induction, during renal development induction requires 1) binding of WT1, CREBBP, and EP300 on the *Wnt4* promoter, 2) changes in histone modifications, and 3) changes in RNA polymerase II phosphorylation, and results in the switch from paused to transcriptional elongation of *Wnt4* [33]. Strikingly, clusters 3 and 4 contain the majority of the mutations involving *WT1* as well as those involving *MLLT1*, a gene encoding a key component of the super-elongation-complex whose dominant role is to regulate transcriptional elongation. MLLT1 functions to bring the PTEFb complex (which phosphorylates RNA polymerase II) to the PAF complex (which governs chromatin modifications and is bound to RNA polymerase II), resulting in transcriptional elongation [46]. Genetic aberrations of genes encoded by PAF complex components *CTR9* and *CDC73* are also rare causes of familial Wilms tumor (reviewed in reference 7). In addition to WT1, CREBBP, EP300, and MLLT1, other proteins involved in transcriptional elongation whose genes are recurrently mutated in Wilms tumor include MAP3K4, responsible for activating CREBBP [47], and BRD7, a component of a SWI/SNF chromatin remodeling complex that interacts with EP300 and mediates histone H3 acetylation at Lys-9 (H3K9ac), an important permissive modification during renal development [48,49]. It has been proposed that MLLT1 (and the highly homologous AF9) serve as reader proteins that specifically recognize H3K9ac [50]. Of particular interest, recurrent truncating mutations involving *BCOR*, which encodes a protein that binds to MLLT1 and results in transcriptional repression [51]. Loss-

of-function mutations in *BCOR* may therefore have the same impact as gain-of-function *MLLT1* mutations seen in Wilms tumor. *BCOR* mutations result in increased H3K4 and H3K36 methylation (a modification characteristic of transcription elongation) accompanied by expression of genes normally silenced in mesenchymal stem cells [52]. Recurrent truncating mutations in the related transcriptional corepressor, *BCORL1*, were also identified in Wilms tumor. In summary, 30-50% of mutations in Wilms tumors appear to converge upon the process of transcriptional elongation. Investigations of the functional and mechanistic details represent a promising area for future research.

CTNNB1, the gene most commonly mutated in Wilms tumor, encodes a protein critical for Wnt4 and Wnt9 mediated induction. Following induction, down-regulation of CTNNB1 is required for normal nephron differentiation [53,54]. Therefore, Wnt activating *CTNNB1* mutations in Wilms tumors likely result in abnormal induction and/or continued progenitor proliferation. Mutations in *FAM123B*, encoding a protein responsible for the degradation of CTNNB1, may similarly result in Wnt activation [55]. However, the paucity of *FAM123B* alterations concurrent with *WT1* mutations, compared with concurrent *CTNNB1* mutations identified in 80% of *WT1*-mutant Wilms tumors, suggests that *FAM123B* and *CTNNB1* mutations may have different functional activities.

Mutations in *FBXW7* have been described in 2/104 post-therapy Wilms tumors, and were associated with epithelial differentiation [56]. *FBXW7* mutations and small segmental copy losses were not identified in our discovery set. Wilms tumors with exclusive epithelial histology are rare, are commonly low stage, and seldom relapse if completely resected [57]. Such tumors were therefore likely not present in the TARGET discovery set. *FBXW7* mutations may also occur late, and therefore are more likely to be identified post-therapy or in response to therapy.

Germline mutations were identified in at least 10% of patients. These include not only *WT1*, *TP53, DICER1*, and *DIS3L2*, germline mutations already recognized in Wilms tumor, but also variants known to predispose to breast cancer, including *PALB2* (previously recognized to rarely result in Fanconi anemia and Wilms tumor [58]), and *CHEK2*, not previously recognized in Wilms tumors. Other novel recurrent germline mutations identified, including *EP300* and *ARID1A*, will require functional validation.

In conclusion, comprehensive genomic analyses suggest that many different genetic changes converge into a limited number of developmental pathways resulting in oncogenesis. One such pathway is regulated by miRNA biogenesis (which promotes the progenitor state), and another is transcriptional elongation that prevents normal induction. Both rely on the epigenetic regulation of transcription during early renal development, which represents a fruitful area of future research. The large number of genes with driver mutations identified in Wilms tumor combined with the relatively small number of gene expression patterns suggests that future studies that attempt to target common processes or pathways may be more efficient than targeting individual gene mutations.

Online Methods

The TARGET initiative supports the long-term maintenance of the data files, methods, and quality control steps involved in the comprehensive genomic analysis of TARGET samples. Sequencing FASTQ and BAM data files are deposited in the Sequence Read Archive at the National Center for Biotechnology Information, and are accessible through the Database of Genotypes and Phenotypes (dbGaP, TARGET URL is https://www.ncbi.nlm.nih.gov/projects/gap/cgi-bin/study.cgi?study_id=phs000218). The Wilms tumor data can be found under the study accession number phs000471. Gene and miRNA expression, methylation, chromosome segmental copy number, results of sequence analysis (e.g. MAF and summary files), and clinical information are available through the TARGET Data Matrix (https://ocg.cancer.gov/programs/target/data-matrix). All datasets include MIAME- and MINSEQE-compliant MAGE-TAB metadata [59] files (using v1.1 specification supporting both next-generation sequencing and microarrays, http://fged.org/projects/mage-tab/) and fully describe the methods, the specimen processing details, and the quality control parameters. Summaries of the methods are provided below. Studies were performed with the approval of the Ann & Robert H. Lurie Children's Hospital Institutional Review Board.

Clinical Samples

Discovery Set—All patients registered on the NWTS-5 protocol with obtained informed consent were eligible if sufficient frozen tumor from the primary, pre-treatment tumor and normal kidney or peripheral blood was available and if they met the criteria for high risk. This includes 78 FHWT that relapsed following chemotherapy, and 39 DAWT demonstrating anaplasia in at least 50% of the slides reviewed centrally, for a total of 117 Wilms tumors. This sample size is associated with a >99% probability of detecting mutations present in 5% of Wilms tumors.

Validation Set—From the 1473 FHWT enrolled on NWTS-5 with available samples, a random sample of 33% was taken. To this group, all cases who relapsed were added, resulting in a total of 600 patients, 30% of whom relapsed. This approach allows for efficient evaluation of molecular markers in a tumor type characterized by a low rate of relapse. The 533/600 patients that had sufficient tumor DNA and passed quality control were included. In addition, all patients enrolled on NWTS-5 and classified as DAWT on central pathology review for which tumor DNA that passed quality control was available (118 DAWT) were included.

Sequencing

Of the 117 discovery set cases, 80 were characterized by whole genome sequencing (WGS) and 37 by whole exome sequencing (WES). WGS libraries were constructed and sequenced by Complete Genomics Inc. (CGI) [60]. Sequencing and alignment of reads to the NCBI Build 37 reference human genome assembly was performed according to the CGI Cancer Sequencing service analytic pipeline. Somatic, non-synonymous exonic variants with a somatic score >-10, somatic rank ≥0.1, and Fisher's Exact Test score ≥13 were retained. Structural variants with discordant mate pair alignments ≥25, successful local de novo assembly, frequency = 0 in the baseline genome set, no overlap with known

underrepresented repeats, and length ≥70 for left and right sections and that occurred within or contained genes were retained. WES was performed on the Illumina HiSeq platform. Variant calling used two approaches which were then combined, retaining all variants. The first approach applied both ATLAS and SAMtools followed by the SACBE annotation pipeline [61,62]. In the second approach, the BAM files were independently analyzed using Bambino Version 1.05 [63]. All non-silent variants from both WGS and WES were combined into a single file and annotated with the Oncotator program (http://www.broadinstitute.org/oncotator/). Missense variants were classified as polymorphisms (identified within dbSNP build 142), known COSMIC mutations, or novel variants and were assessed by five prediction algorithms: SIFT [64], PolyPhen Version 2 [65], dbSNF LR [66], Mutation Taster [67], and Mutation Assessor [68]. Variants considered to be damaging were COSMIC mutations, novel SNVs predicted to be damaging by at least 2 algorithms, SNVs present in dbSNP predicted to be damaging by at least 3 algorithms, in-frame insertion-deletion variants predicted to be damaging by Provean Version 1.1.3 [69], and all frameshift and nonsense mutations.

Cases with available tumor mRNA (78 FHWT and 37 DAWT) were characterized by mRNAseq and miRNAseq. Libraries were prepared following a strand-specific, paired-end protocol as previously described [70]. For mRNA analysis, variants were detected on positive- and negative-split BAMs separately and annotated with SnpEff [71] and SnpSift [72]. Normalized fractional read counts were calculated for each exon to quantify gene and exon expression. For miRNA analysis, reads aligning to known miRNAs in miRBase v20 were summed and normalized to a million miRNA-aligned reads.

Target Capture DNA Sequencing (Validation set)—Biologically relevant genes that were recurrently mutated, selected genes mutated in a single tumor, and selected genes previously reported to be mutated in Wilms tumors were analyzed by target capture. Biologic relevance was determined by features used to filter the discovery set (above), presence of the gene in the Sanger Consensus Gene List [73], and expression during renal development [74]. These features are provided in Supplemental Table 1 for each variant. Probes were designed and genomic DNA libraries were constructed and hybridized to the RNA baits as previously described [6]. Variants were called using Samtools v. 0.1.17 [75]. Variants with tumor coverage >10, tumor alternative base count >3, tumor alternative base allele frequency ≥0.05, and global minor allele frequency ≤0.01 were retained and annotated using Oncotator. Further filtering was performed based on the predicted effect of the variant by multiple algorithms as described above for the discovery set. For recurrent indels, a short indel reference set was created including flanking sequences to which we realigned all of the data to identify additional variants present below the Samtools mpileup thresholds; identified variants were manually reviewed in IGV to confirm their presence.

Variants of interest in the discovery set that were not verified by the above methods were verified by Sanger sequencing. Primers were generated with PrimerZ software (http://genepipe.ngc.sinica.edu.tw/primerz/) and PCR amplification was performed using Platinum® Taq DNA Polymerase and Invitrogen® dNTP Mix (Life Technologies Corporation, Carlsbad, CA) according to the manufacturer's protocol. Samples were sequenced using the Applied Biosystems DNA Sequencer (Model 3730, Life Technologies

Corporation, Carlsbad, CA) according to the manufacturer's protocol. The DNA sequence was analyzed with Chromas Lite 2.1 software (Technelysium, technelysium.com.au/).

Copy Number Analysis (Discovery Set)

DNA labelling, hybridization, and array scanning were performed with Affymetrix 6.0 SNP arrays (Affymetrix, Santa Clara, CA) according to the manufacturer's protocol and processed using Affymetrix Genotyping Console (GTC) 4.0 software. Data were processed as previously described [6]. When SNP 6.0 data was not available, copy number data were calculated based on the CGI relative coverage smoothed in 100-kb windows, corrected for the GC content, and normalized using composite baseline coverage from multiple healthy samples. Segments defined by a minimum of 8 probes with segment log2 mean <-0.5 or > +0.5 were considered lost and gained, respectively, unless otherwise indicated.

Gene Expression Analysis (Discovery set)

Analysis was performed on 76 FHWT and 38 DAWT with the Affymetrix U133+2 chip according to the manufacturer's protocol as previously described [6]. Gene Set Enrichment Analyses (http://www.broadinstitute.org/gsea), version 2.0.14 [76] was run using 1000 permutations and phenotype permutation. Significant enrichment within the GSEA Hallmark and positional gene sets was defined as FDR <20%, and *p*-value <5%. For unsupervised analysis, Non-negative Matrix Factorization Consensus Version 5 was utilized [77].

DNA methylation analysis (Discovery set)

Analysis was performed on 78 FHWT and 39 DAWT using Illumina Infinium Human Methylation 450K BeadChips (Illumina, San Diego, CA, USA) according to the manufacturer's protocol, as previously described [78]. To correlate methylation and gene expression data, methylation probes located within 10k base pairs upstream and downstream of a gene were identified. For each probe and gene pair, the correlation between methylation and gene expression (using the probe with the highest expression) was determined using the General Linear Model (GLM), implemented in R (http://www.R-project.org/). *P*-values were adjusted for multiple comparisons using the multi-test package in R, and correlations with an adjusted p-value <0.05 were considered significant. To identify particular methylation islands driving methylation patterns, the mean β value of each probe within each methylation cluster was compared with the mean β value of all remaining tumors. Those probes with a β value >0.1 in at least one comparison that were also significantly differentially methylated (q <0.05 and FC either >1.5 or <0.5) were retained. All remaining genes with at least 5 differentially methylated probes were annotated with the location of the probe within the gene and the relationship of the probe within a CpG island using UCSC genome build 37. Methylation levels for all probes in imprint control regions ICR1 (IGF2/H19) and ICR2 (KCNO1/CDKN1C) were averaged. ROI was defined as 0.3-0.7 for ICR1 and ICR2, LOI as 0.8-1 for ICR1 and 0.3-0.7 for ICR2, and LOH as 0.8-1 for ICR1 and 0-0.2 for ICR2. Tumors outside of these ranges were not classified.

Computational structural analysis of mutants

Structural models for MAX, NONO and ACTB were retrieved from the Protein Data Bank (http://www.rcsb.org/pdb). The MAP3K4 kinase domain was modelled based on its 31% homology to MAP4K4 (PDB ID 4u3y) using Swiss-Model [79]. Models were manually inspected, and mutations evaluated, using the Pymol program (pymol.org).

Statistical analysis

Fisher's exact test was used to evaluate the associations between categorical variables such as between histology and the occurrence of mutations in specific genes and the co-occurrence of mutations in specific genes. *P*-values for differential gene expression were calculated by using the unpaired t-test assuming unequal variance, and the two-tailed *p*-values were reported. The correlation between methylation and gene expression was determined using the General Linear Model (GLM), implemented in R (http://www.R-project.org/).

Data Availability Statement

Sequencing FASTQ and BAM data files are deposited in the Sequence Read Archive at the National Center for Biotechnology Information, and are accessible through the Database of Genotypes and Phenotypes (dbGaP, TARGET URL is https://www.ncbi.nlm.nih.gov/projects/gap/cgi-bin/study.cgi?study_id=phs000218). The Wilms tumor data can be found under the study accession number phs000471. Gene and miRNA expression, methylation, chromosome segmental copy number, results of sequence analysis (e.g. MAF and summary files), and clinical information are available through the TARGET Data Matrix (https://ocg.cancer.gov/programs/target/data-matrix).

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

The TARGET initiative is supported by NCI Grant U10 CA98543. Work performed under contracts from the NCI, US NIH within HHSN261200800001E includes specimen processing (the COG Biopathology Center), WGS (CGI, Inc.), WXS (Baylor College of Medicine), miRNAseq, RNAseq, and target capture sequencing (BCCA Genome Sciences Center). The content of this publication does not necessarily reflect the views or policies of the Department of Health and Human Services, nor does the mention of trade names, commercial products, or organizations imply endorsement by the U.S. Government. The authors thank the Clinical Applications of Core Technology Laboratory of the Hartwell Center for Bioinformatics and Biotechnology of St. Jude Children's Research Hospital for performing the copy number analysis, and the Northwestern University Genomic Core facility for performing the methylation analysis. This work is also supported by the American and Lebanese Syrian Associated Charities of St. Jude (J. Ma, C. Mullighan), the King Abdullah University of Science and Technology (S. Arold), and the Dutch Cancer Society (A. Ooms). The authors are grateful for the expertise of K. Novik, L. Monovich, P. Beezhold, D. Kersey, D. Turner, M. McNulty, and Y. Moyer. This work would not be possible without the dedication of all the experts within the many clinical disciplines at the local institutions and centrally within the COG and NWTSG, and the patients and their families.

Financial support: TARGET U10 CA98543 contract HHSN261200800001E (see acknowledgements); The Children's Oncology Group award numbers NIH U10CA180886, NIH U10CA180899; NIH U10CA098413; NIH U10CA42326 (EJP); U10CA98543 (JSD, EJP); U24 CA114766; UO1CA88131 (EJP); NCI T32 CA079447, (ALW, AEA).

References

1. Gadd S, et al. Clinically relevant subsets identified by gene expression patterns support a revised ontogenic model of Wilms tumor: A Children's Oncology Group study. Neoplasia. 2012; 14:742–756. [PubMed: 22952427]
2. Torrezan GT, et al. Recurrent somatic mutation in DROSHA induces microRNA profile changes in Wilms tumour. Nat Commun. 2014; 5:4039. [PubMed: 24909261]
3. Wegert J, et al. Mutations in the SIX1/2 pathway and the DROSHA/DGCR8 miRNA microprocessor complex underlie high-risk blastemal type WTs. Cancer Cell. 2015; 27:298–311. [PubMed: 25670083]
4. Rakheja D, et al. Somatic mutations in DROSHA and DICER1 impair microRNA biogenesis through distinct mechanisms in Wilms tumours. Nat Commun. 2014; 5:4802.
5. Beckwith JB, Palmer NF. Histopathology and prognosis of Wilms tumors: results from the First National Wilms Tumor Study. Cancer. 1978; 41:1937–48. [PubMed: 206343]
6. Walz AL, et al. Recurrent DGCR8, DROSHA, and SIX homeodomain mutations in favorable histology Wilms tumors. Cancer Cell. 2015; 27:286–97. [PubMed: 25670082]
7. Perlman EJ, et al. MLLT1 YEATS domain mutations in clinically distinctive favorable histology Wilms tumours. Nat Commun. 2015; 6:10013. [PubMed: 26635203]
8. Ooms AHAG, et al. Significance of TP53 mutation in Wilms umors with diffuse anaplasia: a report from the Children's Oncology Group. Clin Cancer Res. 2016; 22:5582–5591. [PubMed: 27702824]
9. Astuti D, et al. Germline mutations in DIS3L2 cause the Perlman syndrome of overgrowth and Wilms tumor susceptibility. Nat Genet. 2012; 44:277–84. [PubMed: 22306653]
10. Maschietto M, et al. TP53 mutational status is a potential marker for risk stratification in Wilms tumour with diffuse anaplasia. PLoS One. 2014; 9:e109924. [PubMed: 25313908]
11. Maiti S, Alam R, Amos CI, Huff V. Frequent association of B-catenin and WT1 mutations in Wilms tumors. Cancer Res. 2000; 60:6288–92. [PubMed: 11103785]
12. Wu MK, et al. Biallelic DICER1 mutations occur in Wilms tumours. J Pathol. 2013; 230:154–64. [PubMed: 23620094]
13. Palculict TB, et al. Identification of germline DICER1 mutations and loss of heterozygosity in familial Wilms tumour. J Med Genet. 2016; 53:385–8. [PubMed: 26566882]
14. Foulkes WD, et al. Extending the phenotypes associated with DICER1 mutations. Hum Mutat. 2011; 32:1381–4. [PubMed: 21882293]
15. Hill DA, et al. DICER1 mutations in familial pleuropulmonary blastoma. Science. 2009; 325:965. [PubMed: 19556464]
16. Heravi-Moussavi A, et al. Recurrent somatic DICER1 mutations in nonepithelial ovarian cancers. N Engl J Med 2012. 2012; 366:234–42.
17. Passon DM, et al. Structure of the heterodimer of human NONO and paraspeckle protein component 1 and analysis of its role in subnuclear body formation. Proc Natl Acad Sci USA. 2012; 109:4846–4850. [PubMed: 22416126]
18. Gratias EJ, et al. Association of chromosome 1q gain with inferior survival in favorable-histology Wilms tumor: a report from the Children's Oncology Group. J Clin Oncol. 2016; 34:3189–94. [PubMed: 27400937]
19. Chagtai TZ, et al. Gain of 1q as a pPrognostic 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. 2016; 34:3195–203. [PubMed: 27432915]
20. Williams RD, et al. Multiple mechanisms of MYCN dysregulation in Wilms tumour. Oncotarget. 2015; 6:7232–43. [PubMed: 25749049]
21. Kort EJ, et al. The E2F3-Oncomir-1 axis is activated in Wilms' tumor. Cancer Res. 2008; 68:4034–8. [PubMed: 18519660]
22. Cai WY, et al. The Wnt-β-catenin pathway represses let-7 microRNA expression through transactivation of Lin28 to augment breast cancer stem cell expansion. J Cell Sci. 2013; 126:2877–89. [PubMed: 23613467]

23. Tsialikas J, Romer-Seibert J. LIN28: roles and regulation in development and beyond. Development. 2015; 142:2397–2404. [PubMed: 26199409]
24. Maschietto M, et al. The IGF signalling pathway in Wilms tumours--a report from the ENCCA renal tumours biology-driven drug development workshop. Oncotarget. 2014; 5:8014–26. [PubMed: 25478630]
25. Tehler D, Hoyland-Kroghsbo NM, Lund AH. The miR-10 microRNA precursor family. RNA Biology. 2011; 8:728–734. [PubMed: 21881411]
26. Mechtler P, et al. The evidence for a microRNA product of human DROSHA gene. RNA Biol. 2017 Jun 30.:0.doi: 10.1080/15476286.2017.1342934
27. Johanson TM, Lew AM, Chong MM. MicroRNA-independent roles of the RNase III enzymes Drosha and Dicer. Open Biol. 2013; 3:130144. [PubMed: 24153005]
28. Lee D, Nam JW, Shin C. DROSHA targets its own transcript to modulate alternative splicing. RNA. 2017; 23:1035–1047. [PubMed: 28400409]
29. Xu J, Liu H, Park JS, Lan Y, Jiang R. Osr1 acts downstream of and interacts synergistically with Six2 to maintain nephron progenitor cells during kidney organogenesis. Development. 2014; 141:1442–52. [PubMed: 24598167]
30. Kobayashi A, et al. Six2 defines and regulates a multipotent self-renewing nephron progenitor population throughout mammalian kidney development. Cell Stem Cell. 2008; 3:169–81. [PubMed: 18682239]
31. Adli M, Parlak M, Li Y, El-Dahr SS. Epigenetic States of nephron progenitors and epithelial differentiation. Cell Biochem. 2015; 116:893–902.
32. Dutta R, Tiu B, Sakamoto KM. CBP/p300 acetyltransferase activity in hematologic malignancies. Mol Genet Metab. 2016; 119:37–43. [PubMed: 27380996]
33. Essafi A, et al. A Wt1-controlled chromatin switching mechanism underpins tissue-specific wnt4 activation and repression. Dev Cell. 2011; 21:559–74. [PubMed: 21871842]
34. Shao Y, et al. Reversible histone acetylation involved in transcriptional regulation of WT1 gene. Acta Biochim Biophys Sin (Shanghai). 2007; 39:931–8. [PubMed: 18064385]
35. Dekel B, et al. Multiple imprinted and stemness genes provide a link between normal and tumor progenitor cells of the developing human kidney. Cancer Res. 2006; 66:6040–9. [PubMed: 16778176]
36. Harari-Steinberg O, et al. Identification of human nephron progenitors capable of generation of kidney structures and functional repair of chronic renal disease. EMBO Mol Med. 2013; 5:1556–68. [PubMed: 23996934]
37. Pode-Shakked N, et al. Dissecting stages of human kidney development and tumorigenesis with surface markers affords simple prospective purification of nephron stem cells. Sci Rep. 2016; 6:23562. [PubMed: 27020553]
38. Self M, et al. Six2 is required for suppression of nephrogenesis and progenitor renewal in the developing kidney. EMBO J. 2006; 25:5214–28. [PubMed: 17036046]
39. Mugrauer G, Alt FW, Ekblom P. N-myc proto-oncogene expression during organogenesis in the developing mouse as revealed by in situ hybridization. J Cell Biol. 1988; 107:1325–35. [PubMed: 3049618]
40. Laurenti E, Wilson A, Trumpp A. Myc's other life: stem cells and beyond. Curr Opin Cell Biol. 2009; 21:844–54. [PubMed: 19836223]
41. Xu J, et al. Eya1 interacts with Six2 and Myc to regulate expansion of the nephron progenitor pool during nephrogenesis. Dev Cell. 2014; 31:434–47. [PubMed: 25458011]
42. Grandori C, Cowley SM, James LP, Eisenman RN. The Myc/Max/Mad network and the transcriptional control of cell behavior. Ann Rev Cell Dev Biol. 2000; 16:653–699. [PubMed: 11031250]
43. Liu PY, et al. Effects of a novel long noncoding RNA, lncUSMycN, on N-Myc expression and neuroblastoma progression. J Natl Cancer Inst. 2014; 106 pii: dju113.
44. Urbach A, et al. Lin28 sustains early renal progenitors and induces Wilms tumor. Genes Dev. 2014; 28:971–82. [PubMed: 24732380]

45. Viswanathan SR, et al. Lin28 promotes transformation and is associated with advanced human malignancies. Nat Genet. 2009; 41:843–8. [PubMed: 19483683]
46. He N, et al. Human Polymerase-Associated Factor complex (PAFc) connects the Super Elongation Complex (SEC) to RNA polymerase II on chromatin. Proc Natl Acad Sci U S A. 2011; 108:e636–45. [PubMed: 21873227]
47. Abell AN, et al. MAP3K4/CBP-regulated H2B acetylation controls epithelial-mesenchymal transition in trophoblast stem cells. Cell Stem Cell. 2011; 8:525–37. [PubMed: 21549327]
48. Kaeser MD, Aslanian A, Dong MQ, Yates JR, Emerson BM. BRD7, a novel PBAF-specific SWI/SNF subunit, is required for target gene activation and repression in embryonic stem cells. J Biol Chem. 2008; 283:32254–63. [PubMed: 18809673]
49. Sun H, et al. Solution structure of BRD7 bromodomain and its interaction with acetylated peptides from histone H3 and H4. Biochem Biophys Res Commun. 2007; 358:435–441. [PubMed: 17498659]
50. Li Y, et al. AF9 YEATS domain links histone acetylation to DOT1L-mediatied H3K79 Methylation. Cell. 2014; 159:558–571. [PubMed: 25417107]
51. Srinivasan RS, de Erkenez AC, Hemenway CS. The mixed lineage leukemia fusion partner AF9 binds specific isoforms of the BCL-6 corepressor. Oncogene. 2003; 22:3395–3406. [PubMed: 12776190]
52. Fan Z, et al. BCOR regulates mesenchymal stem cell function by epigenetic mechanisms. Nat Cell Biol. 2009; 11:1002–1009. [PubMed: 19578371]
53. Park JS, Valerius MT, McMahon AP. Wnt/beta-catenin signaling regulates nephron induction during mouse kidney development. Development. 2007; 134:2533–2539. [PubMed: 17537789]
54. Schmidt-Ot KM, et al. Beta-catenin/TCF/Lef controls a differentiation-associated transcriptional program in renal epithelial progenitors. Development. 2007; 134:3177–3190. [PubMed: 17693601]
55. Major MB, et al. Wilms tumor suppressor WTX negatively regulates WNT/β-catenin signaling. Science. 2007; 316:1043–1046. [PubMed: 17510365]
56. Williams RD, et al. Subtype-specific FBXW7 mutation and MYCN copy number gain in Wilms' tumor. Clin Cancer Res. 2010; 16:2036–45. [PubMed: 20332316]
57. Beckwith JB, Zuppan CE, Browning NG, Moksness J, Breslow NE. Histological analysis of aggressiveness and responsiveness in Wilms' tumor. Med Pediatr Oncol. 1996; 27:422–8. [PubMed: 8827069]
58. Reid S, et al. Biallelic mutations in PALB2 cause Fanconi anemia subtype FA-N and predispose to childhood cancer. Nat Genet. 2007; 39:162–164. [PubMed: 17200671]
59. Rayner TF, et al. A simple spreadsheet-based, MIAME-supportive format for microarray data: MAGE-TAB. BMC Bioinformatics. 2006; 7:489. [PubMed: 17087822]
60. Drmanac R, et al. Human genome sequencing using unchained base reads on self-assembling DNA nanoarrays. Science. 2010; 327:78–81. [PubMed: 19892942]
61. Bainbridge MN, et al. De novo truncating mutations in ASXL3 are associated with a novel clinical phenotype with similarities to Bohring-Opitz syndrome. Genome Med. 2013; 5:11. [PubMed: 23383720]
62. Lupski JR, et al. Exome sequencing resolves apparent incidental findings and reveals further complexity of SH3TC2 variant alleles causing CMT neuropathy. Genome Med. 2013; 5:57–70. [PubMed: 23806086]
63. Edmonson MN, et al. Bambino: a variant detector and alignment viewer for next-generation sequencing data in the SAM/BAM format. Bioinformatics. 2011; 27:865–866. [PubMed: 21278191]
64. McLaren W, et al. Deriving the consequences of genomic variants with Ensembl API and SNP Effect Predictor. Bioinformatics. 2010; 26:2069–2070. [PubMed: 20562413]
65. Adzhubei IA, et al. A method and server for predicting damaging missense mutations. Nat Methods. 2010; 7:248–9. [PubMed: 20354512]
66. Liu X, Jian X, Boerwinkle E. dbNSFP: a lightweight database of human non-synonymous SNPs and their functional predictions. Hum Mutat. 2011; 32:894–899. [PubMed: 21520341]

67. Schwarz JM, Cooper DN, Schuelke M, Seelow D. MutationTaster2: mutation prediction for the deep-sequencing age. Nat Methods. 2014; 11:361–2. [PubMed: 24681721]
68. Reva B, Antipin Y, Sander C. Predicting the functional impact of protein mutations: application to cancer genomics. Nucleic Acids Res. 2011; 39:e118. [PubMed: 21727090]
69. Choi Y, Sims GE, Murphy S, Miller JR, Chan AP. Predicting the functional effect of amino acid substitutions and indels. PLoS ONE. 2012; 7:e46688. [PubMed: 23056405]
70. Chun HJ, et al. Genome-wide profiles of extra-cranial malignant rhabdoid tumors reveal heterogeneity and dysregulated developmental pathways. Cancer Cell. 2016; 29:394–406. [PubMed: 26977886]
71. Cingolani P, et al. Using Drosophila melanogaster as a model for genotoxic chemical mutational studies with a new program, SnpSift. Front Genet. 2012; 3:35. [PubMed: 22435069]
72. Cingolani P, et al. A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of Drosophila melanogaster strain w1118; iso-2; iso-3. Fly (Austin). 2012; 6:80–92. [PubMed: 22728672]
73. Futreal PA, et al. A census of human cancer genes. Nat Rev Cancer. 2004; 4:177–83. [PubMed: 14993899]
74. Brunskill EW, et al. Atlas of gene expression in the developing kidney at microanatomic resolution. Dev Cell. 2008; 5:781–91.
75. Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics. 2009; 25:1754–1760. [PubMed: 19451168]
76. Subramanian A, et al. Gene set enrichment analysis: A knowledge-based approach for interpreting genome-wide expression profiles. Proc Natl Acad Sci U S A. 2005; 102:15545–15550. [PubMed: 16199517]
77. Brunet JP, Tamayo P, Golub TR, Mesirov JP. Metagenes and molecular pattern discovery using matrix factorization. Proc Natl Acad Sci U S A. 2004; 101:4164–4169. [PubMed: 15016911]
78. Gooskens SL, et al. TCF21 hypermethylation in genetically quiescent clear cell sarcoma of the kidney. Oncotarget. 2015; 6:15828–41. [PubMed: 26158413]
79. Arnold K, Bordoli L, Kopp J, Schwede T. The SWISS-MODEL workspace: a web-based environment for protein structure homology modelling. Bioinformatics. 2006; 22:195–201. [PubMed: 16301204]

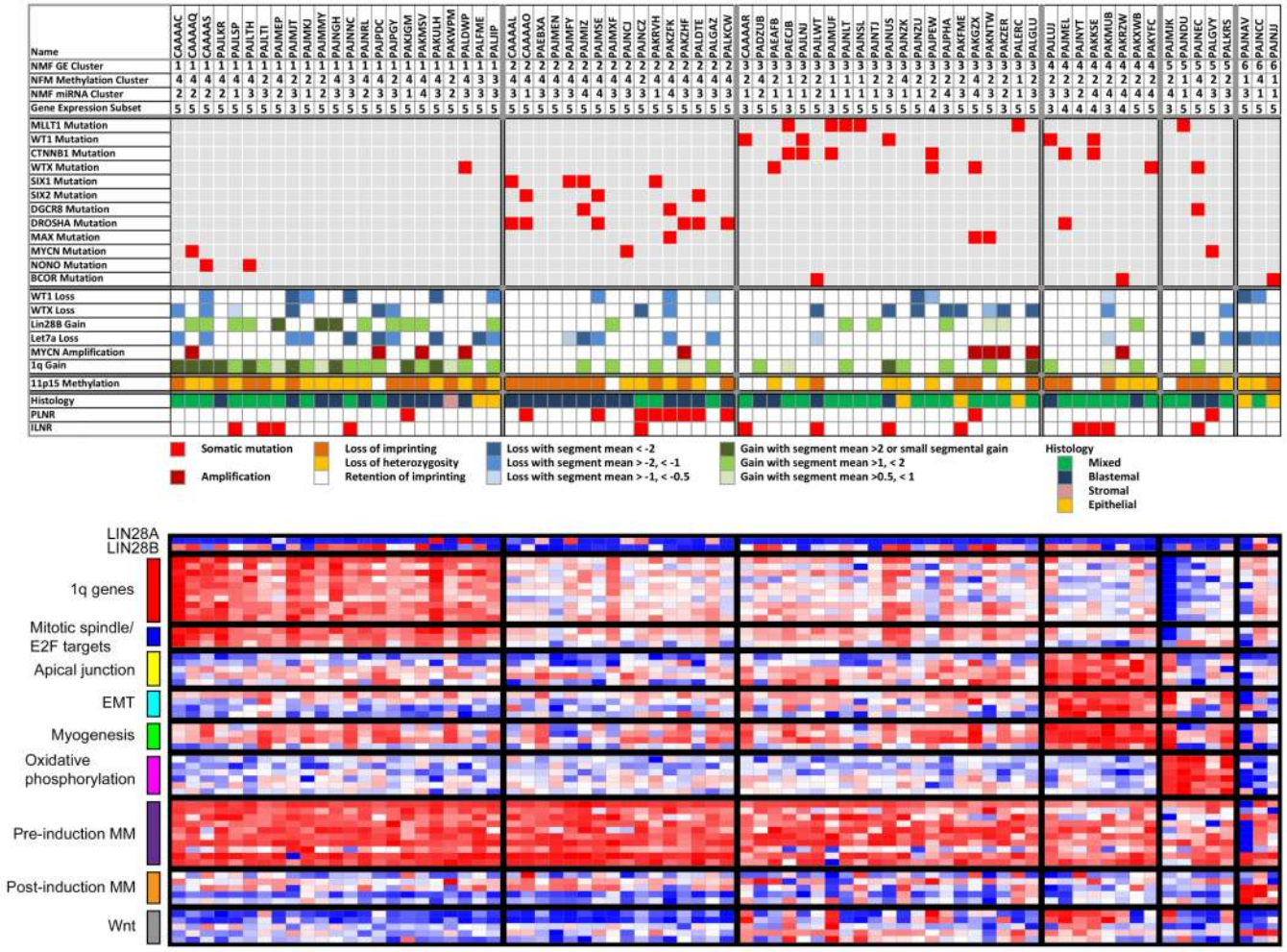

Figure 1. Genetic Landscape of Favorable Histology Wilms Tumor

Data variables of interest (rows) for 76 FHWT samples (columns) are separated into six clusters according to gene expression.

Upper panel: Rows 2, 3, 4 provide NMF cluster according to gene expression, DNA methylation, or microRNA expression, respectively. Row 5 provides gene expression subset according to expression patterns previously reported [1]. These are followed by the most highly recurrently mutated genes, copy number changes, and methylation status at 11p15 as determined by 450K data. The predominant pre-treatment histological classification of the tumor (blastemal, stromal, or epithelial for those showing these patterns in >66% of the tumor represented in all slides, and mixed for those lacking a predominant histologic pattern), and the presence of accompanying nephrogenic rests (perilobar, PLNR, and intralobar, ILNR) are also provided.

Lower panel: Illustrates the expression of genes of interest, with red and blue indicating relatively high and low expression, respectively. Genes present in the top 100 GSEA ranked list for each cluster that overlapped with genes from lists significantly enriched by GSEA in the same cluster are illustrated. Also provided are genes associated with the pre-induction

and post-induction metanephric mesenchyme (MM) and genes associated with Wnt signalling in the early developing kidney.

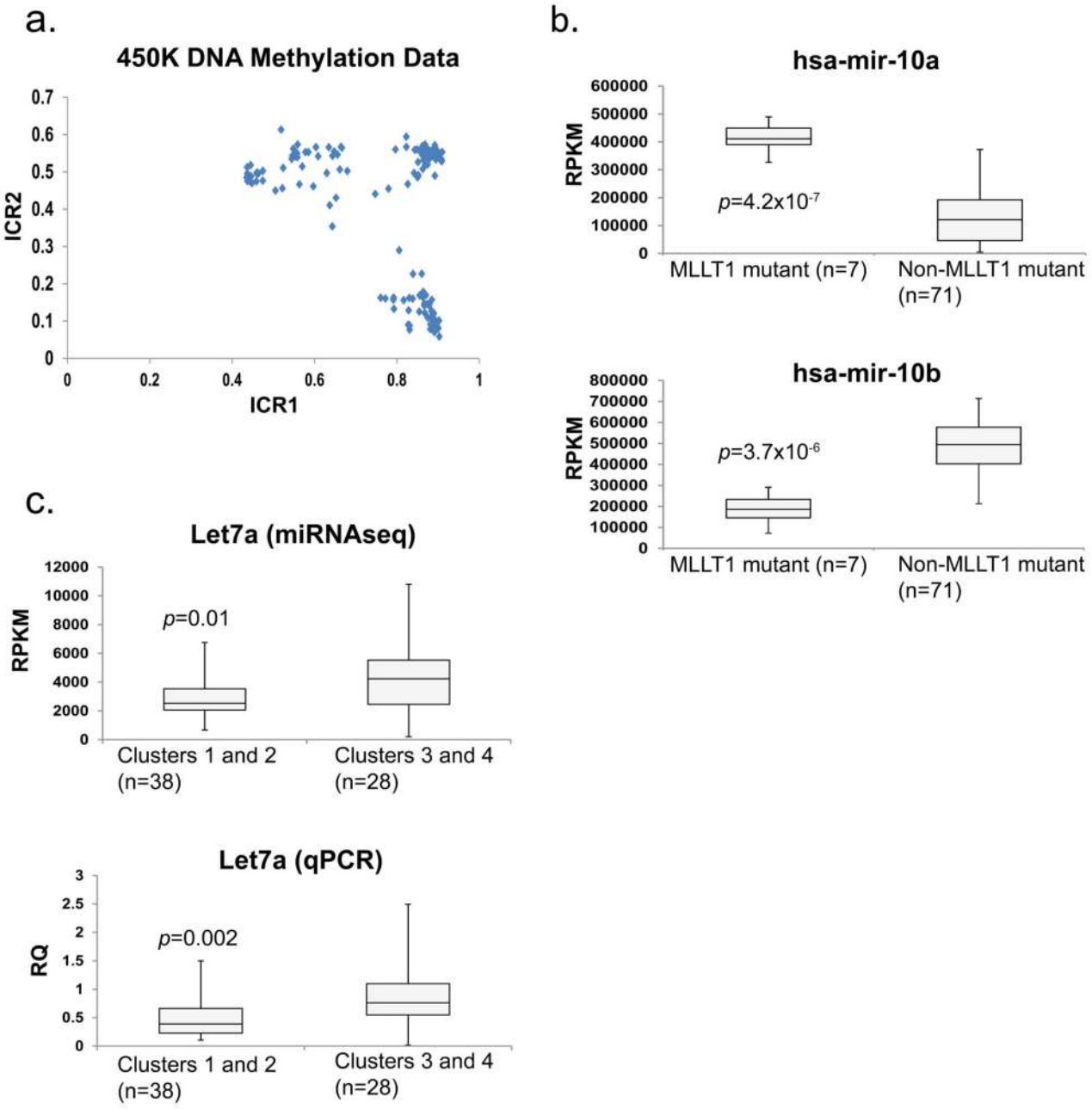

Figure 2. 11p15 ICR1 and ICR2 methylation and select miRNA expression patterns

A. 11p15 Imprinting Status: Graphical representation of the average DNA methylation beta values for 11p15 imprinting control region 1 (ICR1; x axis) and ICR2 (y axis) in 78 FHWT. Retention of imprinting (ROI) is defined as ICR1 and ICR2 average beta values of 0.3–0.7; loss of imprinting (LOI) is defined as 0.8–1 for ICR1 and 0.3–0.7 for ICR2; loss of heterozygosity (LOH) is defined as 0.8–1 for ICR1 and 0–0.2 for ICR2.

B. ***miRNA-10a* and *-10b* expression:** Box plots representing *miRNA-10a* (top panel) and *-10b* (bottom panel) expression in FHWT. The maximum (top whisker), the 75th, median,

and 25th percentiles (box), and minimum expression values (bottom whisker) were calculated for FHWT with *MLLT1* mutation (n = 7) and FHWT without *MLLT1* mutation (n = 71) by using miRNAseq reads per kilobase of transcript per million mapped reads (RPKM) data. Two-tailed *p*-values were calculated by using unpaired *t*-tests assuming unequal variance. Upper panel: $p = 4.2\times10^{-7}$, $t = 5.531$, $df = 76$. Lower panel: $p = 3.7\times10^{-6}$, $t = 4.987$, $df = 76$.

C. ***Let7a expression:*** Box plots representing *let7a* expression in FHWT in gene expression clusters 1 and 2 (N=38) compared with clusters 3 and 4 (N=28). The maximum, 75th percentile, median, 25th percentile, and minimum expression values were calculated by using microRNAseq RPKM data (top) or using RT-PCR to measure mature *let7a* normalized to *RNU44* previously reported [3] (bottom). Two-tailed p-values were calculated by using unpaired *t*-tests assuming unequal variance. Upper panel: $p = 0.01$, $t = 2.654$, $df = 64$. Lower panel: $p = 0.002$, $t = 3.22$, $df = 64$.

Table I

Selected mutations identified within the discovery and validation sets

GENE		DISCOVERY SET			VALIDATION SET		
Hugo Symbol	**Location**	**# patients with somatic variants**	**# patients with germline variants**	**# tumors with copy gain/ loss**	**# variants**	**# patients**	**% of 651 total**
TP53	**17p13**	26	1	6/31	70	65 (9 FHWT, 56 DAWT)	1.7% FH 47.5%, DA
CTNNB1	**3p22**	7	0	15/5	95 (62 exon 3, 20 exon 8)	88	13.5
DROSHA	**5p13**	8	2	7/5	68 (40 p. Glu1147Lys)	66	10.1
WT1	**11p13**	5	2	3/20	46	41	6.3
FAM123B	**Xq11**	7	0	6/27	38 (12 p. Arg358*)	38	5.8
DGCR8	**22q11**	4	0	5/20	29 (21 p. Glu518Lys)	29	4.5
SIX1	**14q23**	5	0	1/28	25 (24 p. Gln177Arg)	25	3.8
SIX2	**2p21**	4	0	15/3	19 (13 p. Gln177Arg)	18	2.9
MLLT1	**19p13**	8	2	14/7	21 (11 p. Asn115_Leu117dup)	20	3.7
MYCN	**2p24**	3	0	29/3	24 (23 p. Pro44Leu/His)	24	3.7
BCOR	**Xp11**	5	0	9/12	17	17	2.6
NONO	**Xq13**	3	0	7/13	13 (10 p. Arg75Cys/His)	13	2
MAX	**14q23**	3	0	1/31	11 (all p. Arg60Gln)	11	1.7
ACTB	**7p22**	3	0	11/21	8 (4 p. Gly146Val)	7	1.1
COL6A3	**2q37**	2	1	11/6	21	21	3.2
MAP3K4	**6q26**	2	0	28/6	11 (5 p. Gly1366Arg)	11	1.7
ASXL1	**20q11**	2	0	24/1	11	11	1.7
BRD7	**16q12**	2	0	3/36	10 (6 p. Pro456fs)	10	1.5
XPO5	**6p21**	2	1	28/8	13	10	1.5
FGFR1	**8p11**	2	0	27/6	9	9	1.4
CHD4	**12p13**	2	0	42/1	9	8	1.2
RLIM	**Xq13**	2	1	7/13	5	5	0.8
NF2	**22q12**	2	0	4/23	3	3	0.5
CREBBP	**16p13**	2	0	6/11	6	6	0.9
CSMD2	**1p35**	2	NA	7/13	NA		
TNRC18	**7p22**	2	NA	11/21	NA		

GENE		DISCOVERY SET			VALIDATION SET		
Hugo Symbol	**Location**	**# patients with somatic variants**	**# patients with germline variants**	**# tumors with copy gain/ loss**	**# variants**	**# patients**	**% of 651 total**
CAPN15	**16p13**	2	NA	3/16	NA		
ZC3H18	**16q24**	2	NA	3/42	NA		
ANKRD11	**16q24**	2	NA	3/42	NA		
RBM10	**Xp11**	2	NA	9/10	NA		
DICER1	**14q32**	1	1	4/24	21	16	2.5
BCORL1	**Xq26**	1	0	7/13	25	25	3.8
ARID1A	**1p36**	1	2	8/14	12	12	1.8
HDAC4	**2q37**	1	1	11/10	8	8	1.2
KRAS	**12p12**	1	0	43/1	3	3	0.5
SALL1	**16q12**	0	0	3/36	2	2	0.3
TERT	**5p15**	0	0	6/7	2 (both p. Glu441del)	2	0.3
LIN28A	**1p36.11**	0	0	8/14	1	1	0.2
LIN28B	**6q16**	0	0	29/5	0	0	0
DIS3L2	**2q37**	1	1	10/7	NA		
PALB2	**16p12**	0	3 (2.6%)	5/13	8	8	1.2
CHEK2	**22q12**	0	3(2.6%)	4/23	8 (4 p. Ile157Thr)	8	1.2
NF1	**17q11**	0	2 (1.7%)	8/17	19 (7 p. Asp176Glu)	19	2.9
EP300	**22q13**	0	1 (0.8%)	5/32	6 (5 p. Asn2209_Gln2213delinsLys)	6	0.9