



Article

Characterization of Alternative Splicing in High-Risk Wilms' Tumors

Yaron Trink¹, Achia Urbach², Benjamin Dekel³, Peter Hohenstein⁴ , Jacob Goldberger¹ and Tomer Kalisky^{1,*}

- ¹ Faculty of Engineering and Bar-Ilan Institute of Nanotechnology and Advanced Materials (BINA), Bar-Ilan University, Ramat Gan 5290002, Israel; trinkya@biu.ac.il (Y.T.); jacob.goldberger@biu.ac.il (J.G.)
² The Mina and Everard Goodman Faculty of Life Sciences, Bar-Ilan University, Ramat Gan 5290002, Israel; achia.urbach@biu.ac.il
³ Pediatric Stem Cell Research Institute and Division of Pediatric Nephrology, Edmond and Lily Safra Children's Hospital, Sheba Tel-HaShomer Medical Centre, Ramat Gan 5262000, Israel; binyamin.dekel@sheba.health.gov.il
⁴ Department of Human Genetics, Leiden University Medical Center, 2300 RC Leiden, The Netherlands; p.hohenstein@lumc.nl
* Correspondence: tomer.kalisky@biu.ac.il; Tel.: +972-3-738-4656

Abstract: The significant heterogeneity of Wilms' tumors between different patients is thought to arise from genetic and epigenetic distortions that occur during various stages of fetal kidney development in a way that is poorly understood. To address this, we characterized the heterogeneity of alternative mRNA splicing in Wilms' tumors using a publicly available RNAseq dataset of high-risk Wilms' tumors and normal kidney samples. Through Pareto task inference and cell deconvolution, we found that the tumors and normal kidney samples are organized according to progressive stages of kidney development within a triangle-shaped region in latent space, whose vertices, or "archetypes", resemble the cap mesenchyme, the nephrogenic stroma, and epithelial tubular structures of the fetal kidney. We identified a set of genes that are alternatively spliced between tumors located in different regions of latent space and found that many of these genes are associated with the epithelial-to-mesenchymal transition (EMT) and muscle development. Using motif enrichment analysis, we identified putative splicing regulators, some of which are associated with kidney development. Our findings provide new insights into the etiology of Wilms' tumors and suggest that specific splicing mechanisms in early stages of development may contribute to tumor development in different patients.

Keywords: Wilms' tumors; pareto task inference; alternative mRNA splicing; cell deconvolution



Citation: Trink, Y.; Urbach, A.; Dekel, B.; Hohenstein, P.; Goldberger, J.; Kalisky, T. Characterization of Alternative Splicing in High-Risk Wilms' Tumors. *Int. J. Mol. Sci.* **2024**, *25*, 4520. <https://doi.org/10.3390/ijms25084520>

Academic Editors: Athanasios G. Papavassiliou and Antonio Rescifina

Received: 22 February 2024
Revised: 5 April 2024
Accepted: 15 April 2024
Published: 20 April 2024



Copyright: © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Wilms' tumors, also referred to as nephroblastomas, are malignancies affecting the kidney that primarily occur in children below the age of five. The tumorigenesis of Wilms' tumors is believed to be closely associated with fetal nephron development since they often contain "blastemal" cells that histologically resemble fetal nephron progenitors. Moreover, Wilms' tumors often arise from or in the vicinity of "nephrogenic rests"—poorly differentiated structures that resemble early embryonic kidney precursor cells not usually seen in the postnatal kidney [1–3]. These observations suggest that genetic and epigenetic aberrations during various stages of fetal kidney development lead to Wilms' tumor formation, and that these aberrations result in extensive heterogeneity among patients.

Wilms' tumors are histologically and clinically diverse. Most cases are classified as favorable histology Wilms' tumors (FHWTs), which typically contain "stromal", "blastemal", and "epithelial" components. The stromal component resembles the nephrogenic stroma of the developing fetal kidney and may also contain cells resembling smooth or skeletal muscle, fat, cartilage, bone, and glial cells. The blastemal component histologically resembles the cap mesenchyme, a transient cell compartment of the fetal developing kidney which

contains nephron progenitors [4]. In the fetal kidney, these nephron progenitors undergo a mesenchymal-to-epithelial transition (MET) and differentiate into the various epithelial tubular segments of the nephron [5]. The epithelial component typically contains early non-functional epithelial structures from early stages of nephron differentiation. Tumors containing all three components are classified as “mixed” type or “triphasic”, whereas those containing a single dominant histology are classified as “blastemal”, “stromal”, or “epithelial” tumors. The blastemal component is presumed to be most malignant, and tumors containing a large blastemal component that survived preoperative chemotherapy are regarded as “high-risk” tumors and require more aggressive treatment [4,6,7]. Anaplastic Wilms’ tumors are characterized by multipolar mitoses, nuclear enlargement, and hyperchromatic nuclei [8], and are classified as “focal” (FAWT) or “diffuse” (DAWT) depending on the geographical distribution of the anaplastic cells within the tumor [6,9]. Diffuse anaplastic Wilms’ tumors are associated with lower event-free survival and lower overall survival estimates when compared to favorable histology Wilms’ tumors, and are also regarded as “high-risk” tumors [6,7,9]. Another class of high-risk tumors are favorable histology Wilms’ tumors that relapsed [6,10].

There are two main protocols for treating Wilms’ tumors. The Children’s Oncology Group (COG) in North America recommends initial surgery followed by chemotherapy, whereas the International Society of Pediatric Oncology (SIOP), followed by European countries, favors initial chemotherapy prior to surgery [6,11]. While efforts from both of these collaborative groups have significantly improved treatment efficacy and survival rates, considerable challenges remain. First, clinical outcomes for high-risk patients remain significantly worse when compared to low-risk patients [12]. In addition, it is crucial to improve risk stratification in order to minimize unnecessary therapeutic exposure and possible subsequent long-term complications for low-risk patients. Advances in the understanding of tumorigenesis have significant implications for this goal [13].

Previous analyses of Wilms’ tumors based on DNA microarrays or high-throughput sequencing methods have been used to identify subsets of tumors based on gene expression patterns [14,15], to find differentially expressed genes between Wilms’ tumors and normal kidney samples [16,17], and to search for genetic markers associated with relapse [18,19]. In a previous study [20], we observed that favorable histology Wilms’ tumors (treated according to the COG protocol) form a triangular shaped continuum in gene expression latent space, and that the vertices of this triangle, which represent tumor “archetypes,” have blastemal, stromal, and epithelial characteristics, corresponding to the three main lineages of the developing fetal kidney. In a consecutive study [21], we found that this geometry is also conserved in high-risk tumors that were treated with chemotherapy prior to surgery (according to the SIOP protocol) but still contained a significant amount of remaining viable blastema [15], and used a probabilistic generative model [22] to represent each tumor as a mixture of three latent biological “topics” with stromal, epithelial, and blastemal characteristics.

While these studies focused solely on gene expression in Wilms’ tumor subtypes, there is widespread evidence that alternative splicing also plays a major role in the tumorigenesis and phenotypes of tumors, many of which exhibit significant and widespread splicing abnormalities when compared to their normal counterparts [23]. The dysregulation of RNA splicing has been found to directly affect some of the hallmarks of cancer such as metastasis, one of the most difficult issues affecting cancer therapy [24]. Many alternative splicing events have been found to play a role in the epithelial-to-mesenchymal transition (EMT), whereby polarized epithelial cells undergo a series of transitions to assume a mesenchymal phenotype. This phenotype, which exhibits migratory capability and increased invasiveness, confers a metastatic ability upon cells [25,26]. For example, the skipping of exon 11 in the gene RON has been shown to enhance tumor migratory abilities [27], and splice isoform switching from the epithelial to mesenchymal isoform of the gene FGFR2 has been shown to induce EMT [28–32]. Likewise, RNA binding proteins known to regulate mRNA splicing were found to be abnormally expressed or somatically mutated in

cancer [33–37]. The RNA binding protein epithelial splicing regulatory protein 1 (ESRP1), for example, has been shown to regulate the splicing of CD44, and knockdown of ESRP1 was shown to suppress lung cancer metastasis [38]. Specifically in Wilms' tumor, a recent study showed that the splicing regulator ESRP2 is repressed by DNA methylation, whereas the overexpression of ESRP2 in Wilms' tumor cell lines promotes alternative splicing and inhibits cell proliferation both in vitro and in vivo [39].

In a recent work, we also found that fetal kidney cells in an early developmental stage have a mesenchymal splice-isoform profile that is similar to that of blastemal-predominant Wilms' tumor xenografts [40]. However, a comprehensive analysis of alternative splicing in high-risk Wilms' tumors is lacking, and could provide insights into the molecular processes driving tumor development and their relation to kidney development. It could also provide insights into the functional role of alternative splicing events, specifically with regard to the possible role of EMT in establishing different tumor phenotypes. Finally, such an analysis could pave the way for targeting biomarkers for tumor subtypes, an urgently needed clinical effort [41,42].

In this study, we set out to comprehensively characterize alternative splicing in high-risk Wilms' tumors. We first downloaded a publicly available dataset consisting of RNA sequences collected from favorable histology Wilms' tumors (FHWTs) that relapsed, diffuse anaplastic Wilms' tumors (DAWT), and associated normal kidney samples [10]. These tumors were treated according to the COG protocol and collected by the NCI TARGET ("Therapeutically Applicable Research to Generate Effective Treatments") initiative. Using Pareto task inference [43], Gene Ontology enrichment analysis, and cell deconvolution [44], we found that the tumors and normal kidney samples are organized according to progressive stages of kidney development within a triangle-shaped region in latent space, whose vertices, or "archetypes", correspond to cell states resembling the cap mesenchyme, the nephrogenic stroma, and epithelial tubular structures of the fetal kidney. We next identified genes that are alternatively spliced between samples located near the three archetypes and found that many of these genes are related to EMT and muscle structure and development. Finally, we used motif enrichment analysis for known RNA binding proteins to identify putative splicing regulators, some of which were previously found in kidney development. We anticipate that these findings will contribute to a better understanding of the role of alternative mRNA splicing in the formation of Wilms' tumors in diverse patient populations.

2. Results

2.1. High-Risk Wilms' Tumors and Normal Kidney Samples Form a Triangle-Shaped Continuum in Latent Space That Is Bounded by Archetypes with Blastemal, Stromal, and Epithelial Characteristics

We first downloaded an RNA sequencing dataset from the NCI TARGET [10] study containing 130 Wilms' tumor samples and six associated normal kidney samples. The tumors were all classified as "high-risk" since they were either relapsed favorable histology Wilms' tumors (FHWTs) or diffuse anaplastic Wilms' tumors (DAWTs). We then performed sequence alignment and obtained a gene expression matrix (Table S1). Using Principal Component Analysis (PCA), we found that high-risk Wilms' tumors form a continuum in gene expression latent space rather than discrete well-separated clusters [20,21] (Figure 1A–F, Supplementary Figures S1–S6). To better understand this continuous heterogeneity, we used Pareto task inference [43,45] to calculate the vertices of the best fitting polytope encompassing all tumors and normal kidney samples in latent space. These vertices, or "archetypes", presumably represent idealized cell types from which all samples within the continuum are composed, where the precise transcriptional state of each sample determines its position in latent space relative to each biological archetype [20,43,46]. This technique allows us to summarize the gene expression profile of each sample as a combination of different archetypal, or pure, biological states, based on a theoretical evolutionary principle [47]. We tested different numbers of archetypes and found that the biological heterogeneity of our dataset could be described by both three and four archetypes. However,

closer inspection of the fourth archetype showed that most of the variability captured by this archetype is likely related to RNAseq library size (Supplementary Figure S11). We therefore decided to focus on the polytope summarized by three archetypes which best summarize the biological heterogeneity of our dataset.

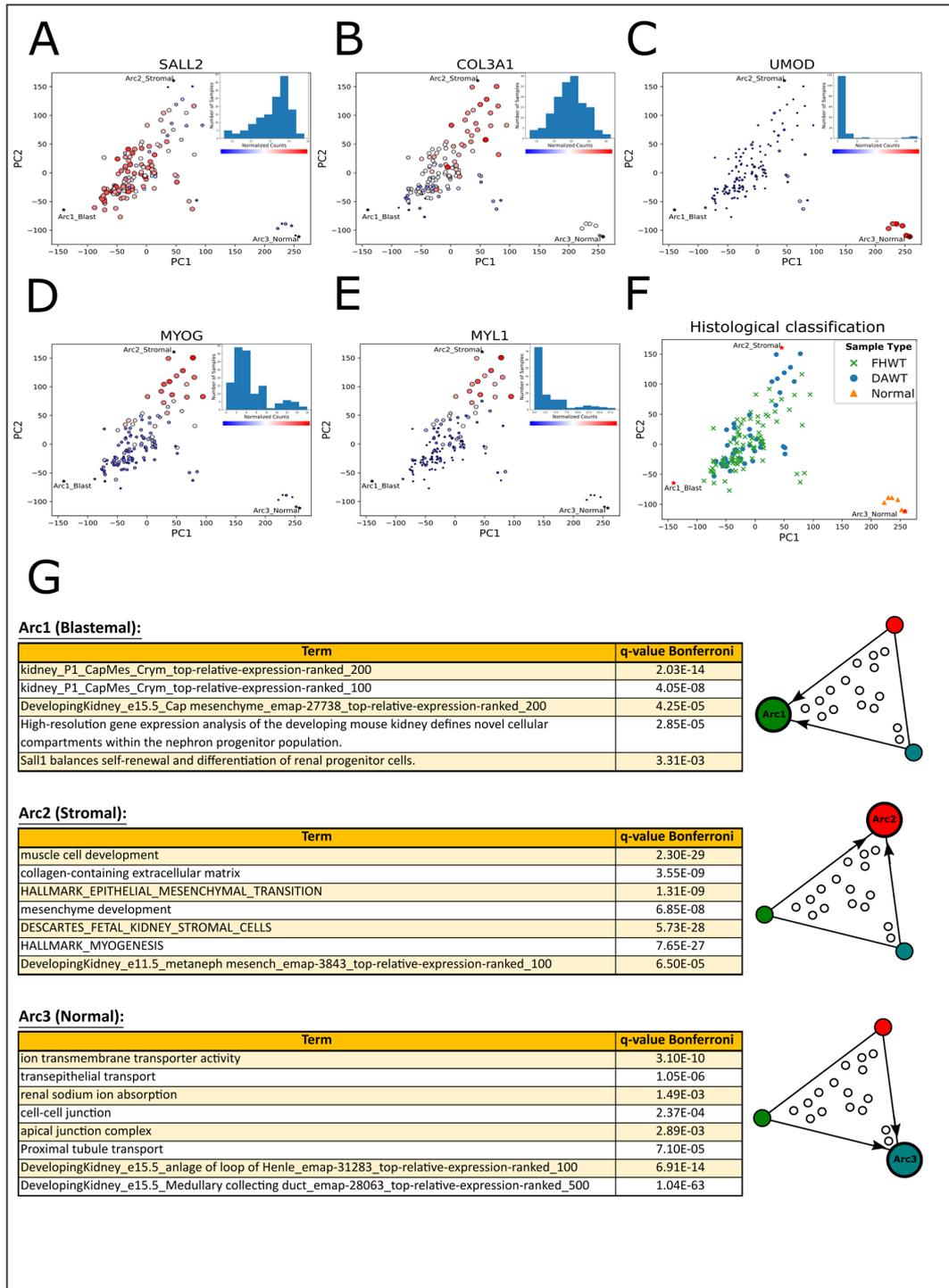


Figure 1. High-risk Wilms’ tumors and normal kidney samples form a triangle-shaped continuum in latent space that is bounded by archetypes with stromal, blastemal, and epithelial characteristics that resemble the nephrogenic stroma, the cap mesenchyme, and early epithelial structures of the fetal kidney, respectively.

(A–C) Shown are PCA plots of RNAseq gene expression profiles from 130 high-risk Wilms’ tumors (relapsed favorable histology or diffuse anaplastic Wilms’ tumors) and six normal kidneys from the US NCI TARGET project. Three archetypes, denoted in the plots by asterisks and labeled “Arc1_Blast”, “Arc2_Stromal”, and “Arc3_Normal”, form the vertices of the best fitting polytope which encompasses all data points. To identify the archetypes, the size and color of each sample (dot) are drawn according to expression levels of known genes marking the cell populations of the fetal developing kidney (large red—high expression; small blue—low expression). Histograms of gene expression levels are included in each PCA subplot. The gene SALL2, which marks the cap mesenchyme, is highly expressed near Arc1, which is the “blastemal” archetype. The gene COL3A1, which marks the nephrogenic stroma, is highly expressed near Arc2, which is the “stromal” archetype. Likewise, UMOD, a gene which is highly expressed in the epithelial cells of the loop of Henle, is highly expressed in the normal samples located near Arc3, which is the “normal” archetype that is predominantly epithelial. (D,E) MYOG, a transcription factor that can induce myogenesis, and MYL1, a gene that encodes a muscle motor protein, are highly expressed near the “stromal” archetype (Arc2_Stromal). (F) A PCA plot with each sample marked according to its histological classification. We did not observe a significant separation between relapsed favorable histology Wilms’ tumors (FHWTs) and diffuse anaplastic Wilms’ tumors (DAWTs) in this dataset. (G) Genes overexpressed ($\log_2FC > 3$) in each of the three archetypes with respect to both of the other two were used as input for Gene Ontology enrichment analysis. Arc1, the “blastemal” archetype, overexpresses genes related to the cap mesenchyme and nephron progenitor cells. Arc2, the “stromal” archetype, overexpresses genes that are related to the extracellular matrix, the nephrogenic stroma (the un-induced metanephric mesenchyme), and muscle development. Arc3, the “normal” archetype, overexpresses genes that are related to the structure and function of epithelial lineages of the kidney.

To infer the biological identity of the three archetypes, we checked the expression levels of selected genes that are known to mark the different lineages in the developing kidney (Figures 1A–C and 2A). We found that the gene SALL2, a marker for the cap mesenchyme, is overexpressed near the first archetype; hence, we named it the “blastemal” archetype. Similarly, the gene COL3A1, which is a marker of the nephrogenic stroma, is overexpressed in the vicinity of the second archetype, which we named the “stromal” archetype, and the gene UMOD, which is a marker for renal epithelial cells, is overexpressed in the normal samples near the third archetype, which we named the “normal” archetype. We also observed that MYOG, a muscle-specific transcription factor with myogenesis-inducing capabilities, and MYL1, a motor protein expressed in muscle cells, are overexpressed near the stromal archetype (Figure 1D,E). Consistent with the findings of the original paper by the TARGET initiative [10], we did not observe an overall significant separation in gene expression latent space between tumors with favorable histology (FHWT) and diffuse anaplasia (DAWT) (Figure 1F).

We further characterized the identity of the archetypes by performing Gene Ontology (GO) enrichment analysis (Figure 1G, Table S2). We chose genes that are overexpressed ($\log_2FC > 3$) in each of the three archetypes with respect to the other two and used these as inputs to ToppGene [48]. We observed that genes that are overexpressed in the blastemal archetype are related to the cap mesenchyme and nephron progenitors, genes that are overexpressed in the stromal archetype are related to the nephrogenic stroma (the un-induced mesenchyme) and muscle development, and genes that are enriched in the normal archetype are related to more differentiated epithelial structures of the nephron such as the proximal tubule, the loop of Henle (LOH), and the collecting duct.

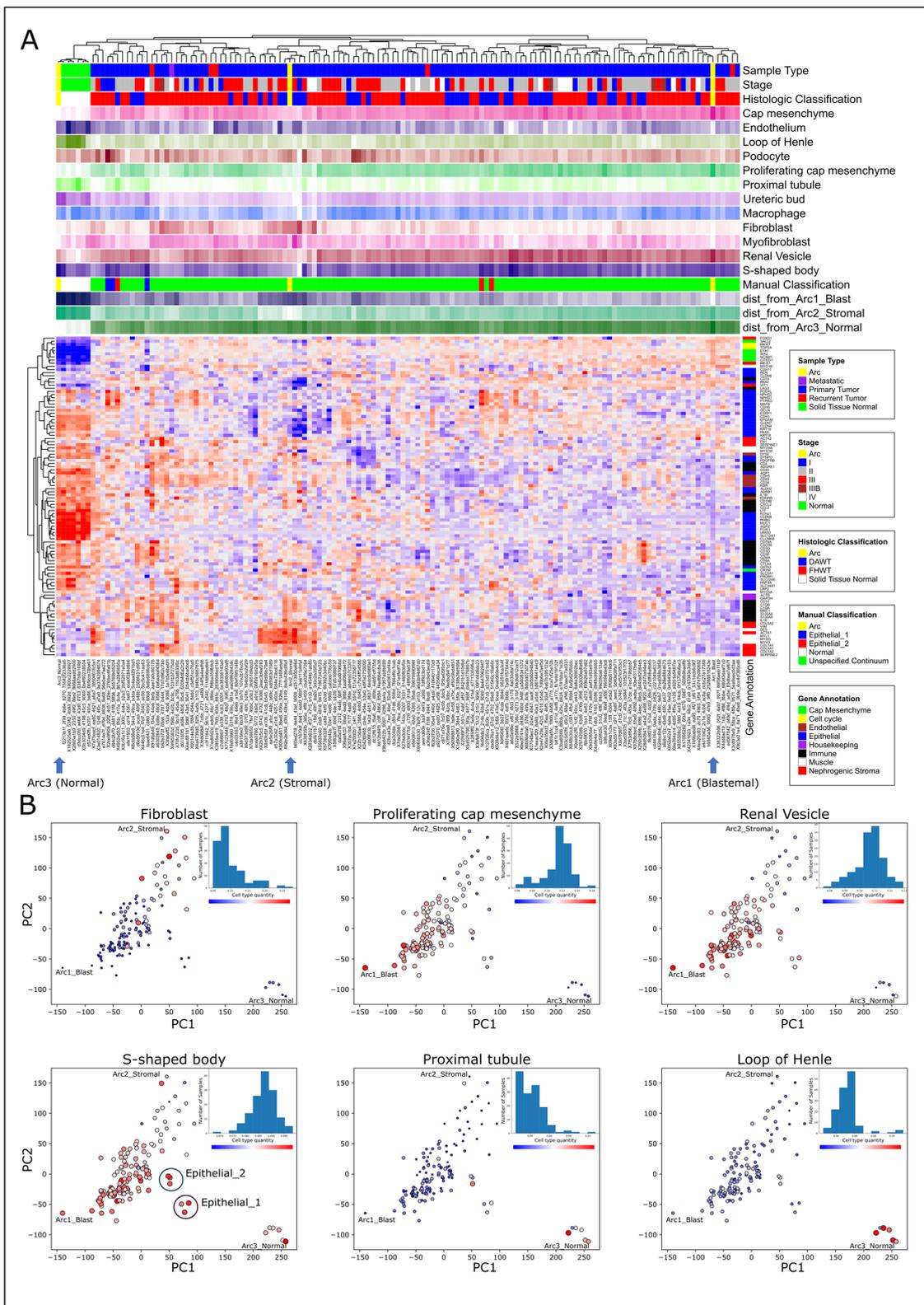


Figure 2. Cell deconvolution shows that high-risk Wilms’ tumors are organized in latent space according to progressive stages of kidney development. **(A)** A gene expression heatmap of 102 selected genes that are known from the literature to be involved in kidney development. The top panels show the clinical and histological information for each sample, the proportions of cell types within each sample as predicted by cell deconvolution, and the distances from each of the three archetypes (for cell deconvolution and distances to archetypes: light color—low; dark color—high).

It can be seen that genes characteristic to the cap mesenchyme (e.g., *CITED1*, *EYA1*, and *SALL2*) are overexpressed near the blastemal archetype. Genes characteristic of the nephrogenic stroma (e.g., *COL3A1*, *COL5A2*, and *VIM*), as well as muscle structure and development (e.g., *DES*, *MYL1*, and *MYOG*), are overexpressed in the vicinity of the stromal archetype. Likewise, genes marking epithelial tubular structures, mainly the loop of Henle and the distal tubule (e.g., *CDH1*, *KCNJ1*, *MUC1*, *SLC12A1*, and *UMOD*), are overexpressed near the normal archetype. **(B)** PCA plots marking proportions of fetal kidney cell populations within each sample, as inferred by cell deconvolution. The size and color of each data point are drawn according to the proportion of the selected population from the predicted cellular composition (large red—high proportion; small blue—low proportion). A histogram of the predicted cell type proportions for all cells is included in each PCA plot. It can be seen that the samples are arranged in latent space according to progressive stages of the developing kidney: cells resembling the cap mesenchyme are predominant in tumors located near the blastemal archetype, whereas cells resembling early nephron differentiation such as the renal vesicle and S-shaped body are found at progressively higher proportions in the more epithelial-like tumors that are located towards the normal archetype (“Epithelial 1” and “Epithelial 2”). Similarly, cells resembling fully differentiated epithelial structures such as the proximal tubule and the loop of Henle are predominant in the normal samples. Likewise, we observed that tumors located near the stromal archetype have a high proportion of cells resembling renal fibroblasts, which are the predominant component of the nephrogenic stroma.

2.2. High-Risk Wilms’ Tumors Are Organized in Latent Space According to Progressive Stages of Kidney Development

In order to better understand the relationship between Wilms’ tumor heterogeneity and kidney development, we performed hierarchical clustering using expression levels of 102 selected genes that are known from the literature to be involved in kidney development (Figure 2A, Table S3, Supplementary Figures S2–S4). We observed that, indeed, the tumors near the stromal archetype have a high expression of genes related to the nephrogenic stroma (e.g., *COL3A1*, *COL5A2*, and *VIM*) and muscle structure and development (e.g., *DES*, *MYL1*, and *MYOG*), and that tumors located near the blastemal archetype have a high expression of genes marking the cap mesenchyme (e.g., *CITED1*, *EYA1*, and *SALL2*). Likewise, we observed that the normal samples have a high expression of renal epithelial markers, mainly for the loop of Henle and the distal tubule (e.g., *CDH1*, *KCNJ1*, *MUC1*, *SLC12A1*, and *UMOD*). Finally, we identified six “epithelial-like” tumors that are located distinctly from the others in latent space (marked “Epithelial 1” and “Epithelial 2” in Figure 2B) and that also express some renal epithelial markers (e.g., *CDH6*, *KRT18*, and *KRT19*, Supplementary Figure S6).

The bulk expression profiles in this dataset are the average expression of the cell types present in the sample. To enhance the resolution of our analysis, we next used cell deconvolution [44] to predict the proportions of different cell types in each sample. Since Wilms’ tumors contain cells resembling those of the fetal kidney, we used a publicly available atlas of single-cell gene expression from the human fetal kidney [49] as a reference. After deconvolving each sample, we plotted the proportions of each predicted fetal cell subpopulation (e.g., “fibroblast” and “proximal tubule”) for each sample in gene expression latent space (Figure 2B and Supplementary Figures S7–S10). We observed that the tumors are arranged according to progressive developmental stages in latent space. For example, high proportions of cells resembling the cap mesenchyme are found within tumors located near the blastemal archetype, whereas cells resembling early stages of nephron differentiation (“renal vesicle” and “S-shaped body”) are found in progressively higher proportions in the more epithelial-like tumors that are located towards the normal archetype (marked as “Epithelial 1” and “Epithelial 2” in Figure 2B). Cells resembling fully differentiated epithelial structures (“proximal tubule” and “loop of Henle”) were found to be predominant in the normal samples. Likewise, we observed that tumors located near

the stromal archetype have a high proportion of cells resembling renal fibroblasts. We performed similar analysis on a dataset of high-risk blastemal-type Wilms' tumors (treated according to SIOP protocols) published by Wegert et al. [15] and found similar results [21] (see a detailed comparison in Supplementary Figures S7–S10).

2.3. A Significant Number of RNA Transcripts Related to EMT and Muscle Development Are Alternatively Spliced between High-Risk Wilms' Tumors Located in Different Regions of Latent Space

We next set to identify genes that are alternatively spliced between the different tumors and normal kidney samples located in different regions of latent space. We first chose three groups of samples, each containing five samples located near one of the three archetypes (Supplementary Figure S12). We then used rMATS [50] to perform three comparisons between these three groups. The results returned by rMATS (Figure 3A,B; Tables S4 and S5; and Supplementary Figures S13–S81) include a set of transcripts that are significantly (FDR = 0) alternatively spliced between tumors representing each of the three transcriptional archetypes. Gene Ontology (GO) enrichment analysis for alternatively skipped exons (FDR = 0, $|\Delta\psi| > 0.1$) showed that a significant fraction of these transcripts are related to the epithelial-to-mesenchymal transition (EMT) and muscle development (Figure 4A, Table S4).

For example, the gene *FGFR2*, a gene that is known to have epithelial and mesenchymal splice variants [28,51–53], is expressed in its mesenchymal form in the majority of tumors. Similarly, the gene *PRMT2*, a gene that is involved in cancer invasion, growth, and presumably EMT, and whose splice variants have been linked to cancer [54,55], contains an exon that is overexpressed near the stromal and normal archetypes, as well as some of the epithelial-like tumors. Likewise, the gene *TPM2*, whose splice variants and differential expression have been associated with EMT [56,57], cancer [58,59], and muscle diseases [60], is expressed in its muscle-specific isoform [58,61] near the stromal archetype. Another example is the gene *FLNB* whose exon 30 was found to be skipped in the majority of tumors (Supplementary Figure S32), which is consistent with previous findings that the skipping of this exon induces EMT, is associated with basal-like breast cancer, and is regulated by the RNA binding proteins QKI and RBFOX1 [62]. Similarly, in the gene *ATP2B1* (*PMCA1*), an isoform that is overexpressed specifically in skeletal and heart muscle but not in kidney tissues [63,64], was found to be overexpressed near the stromal archetype (Supplementary Figure S72). Another notable example is the gene *MEF2D* [65], whose muscle-specific isoform is overexpressed near the stromal archetype (Supplementary Figures S17 and S76). The muscle-specific isoform for this gene was previously found to be promoted by the splicing regulators RBFOX1/2 [66,67].

2.4. Motif Enrichment Analysis Reveals Putative RNA Binding Proteins (RBPs) Regulating Alternative Splicing between Wilms' Tumors Located in Different Regions of Latent Space

Alternative RNA splicing is controlled by RNA binding proteins (RBPs) which bind to pre-mRNA upstream or downstream of the regulated exons. GO enrichment analysis for transcripts that are alternatively spliced between samples located near the three archetypes suggests that many of them are known targets for the splicing regulators RBFOX2, ESRP1, and ESRP2 (Figure 4A). Therefore, in order to identify more putative splicing regulators, we used rMAPS [68,69] to perform enrichment analysis for RNA binding motifs belonging to known RNA binding proteins (Figure 4B,C; Table S6; and Supplementary Figures S82–S100). We found a set of putative splicing regulators whose binding motifs are enriched upstream or downstream of exons that are alternatively spliced, and that are also differentially expressed between the samples in different locations of latent space. The most prominent motif enrichment was found for the RNA binding proteins RBFOX1 [62,70] and QKI [62,70,71], which indicates that in the more stromal tumors, these splicing regulators are overexpressed and tend to bind downstream of their target exons and promote their inclusion. We note that RBFOX1 and QKI were previously shown to physically interact with each other in order to regulate alternative splicing in mammary epithelial cells, thereby promoting an epithelial-to-mesenchymal transition (EMT) [62], and have also been associated with the mesenchymal-to-epithelial transition (MET) that occurs during fetal kidney development [40].

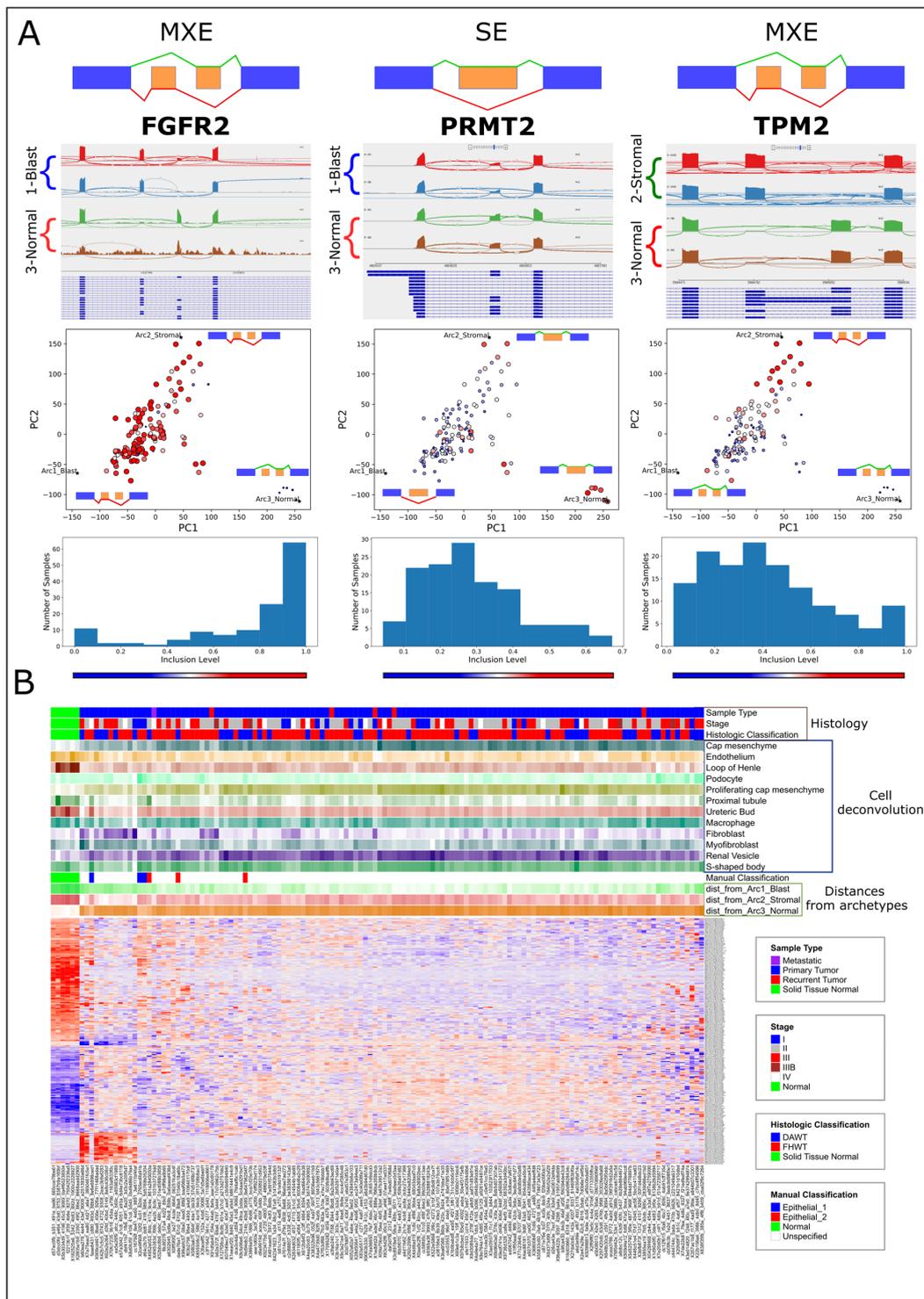


Figure 3. RNA transcripts related to EMT and muscle development are alternatively spliced between high-risk Wilms' tumors located in different regions of latent space. **(A)** Sashimi plots (top), feature plots (middle), and inclusion-level histograms (bottom) for the alternatively spliced genes FGFR2, PRMT2, and TPM2. The sashimi plots show representative samples that are located near the archetypes (denoted by asterisks). The feature plots show the isoform inclusion levels in the different tumors and samples in latent space, where the size and color of each sample are drawn according to the inclusion levels of a particular isoform (large red—high; small blue—low). **(B)** A heatmap showing the inclusion levels of 277 significantly skipped/included exons (SEs) and mutually exclusive exons (MXEs) for transcripts that were found to be significantly alternatively spliced in latent space. The rows are the union of all significant alternative splicing

events ($FDR = 0, |\Delta\psi| > 0.2$) from each of the three comparisons between the five nearest neighbors of the three archetypes. We plotted only SE (single-exon) and MXE (mutually exclusive exon) splicing events since these were the majority. The top panels show clinical and histological information, the proportions of cell types in each sample as predicted by cell deconvolution, our manual classification based on tumor location in latent space, and the distances to each of the three archetypes (for cell deconvolution and distances to archetypes: light color—low; dark color—high).

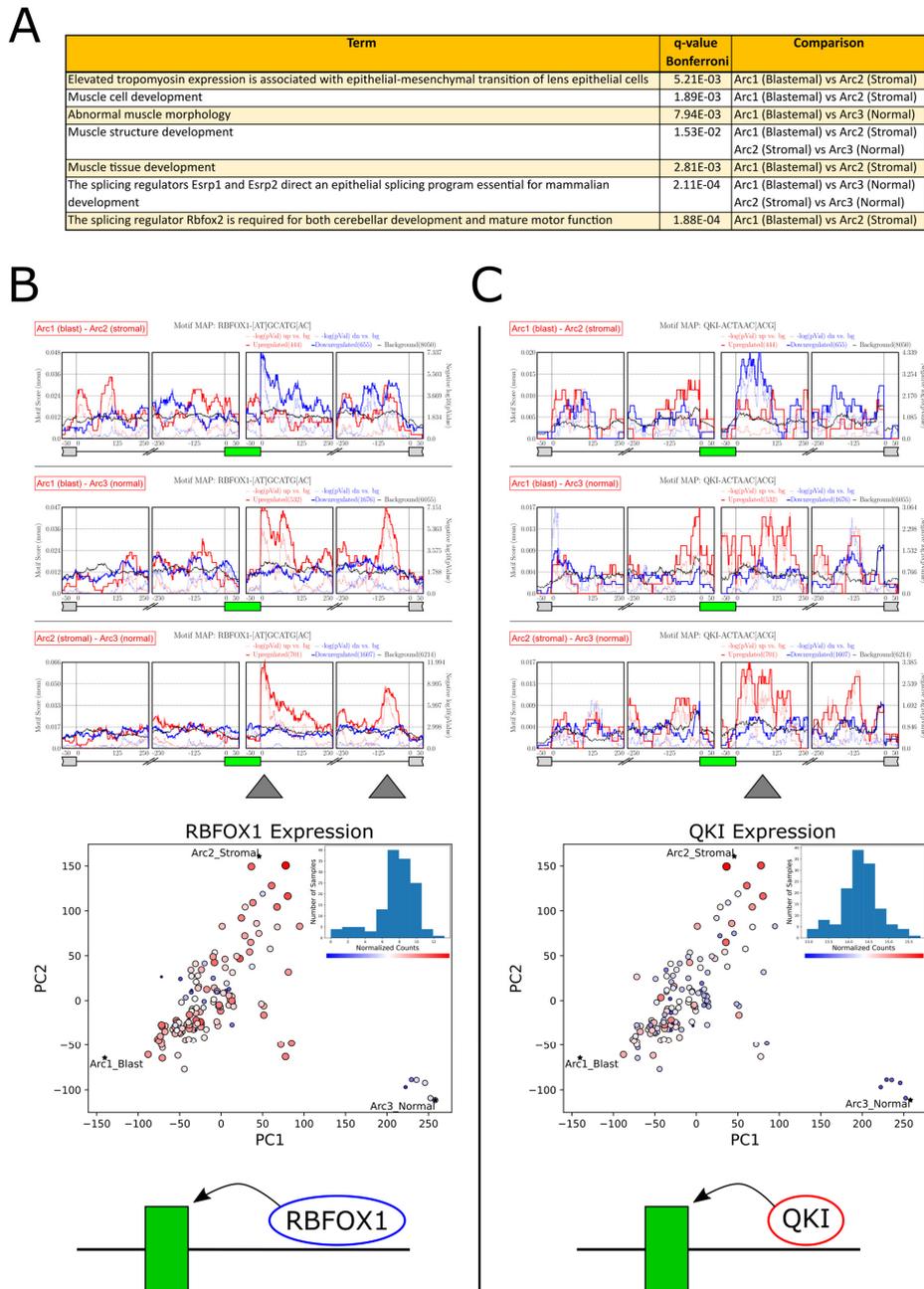


Figure 4. Motif enrichment analysis reveals putative RNA binding proteins (RBPs) regulating alternative mRNA splicing between Wilms’ tumors located in different regions of latent space. (A) Gene Ontology (GO) enrichment analysis for exons that were found to be significantly alternatively spliced ($FDR = 0, |\Delta\psi| > 0.1$) in latent space, as identified from the three comparisons between the five

nearest neighbors of the three archetypes. A significant fraction of these transcripts are related to the epithelial-to-mesenchymal transition (EMT) and muscle development. (B,C) Motif enrichment diagrams (top) and expression feature plots (middle) for the splicing regulators RBFOX1 and QKI. Histograms of RBFOX1 and QKI expression levels are included in each PCA subplot (archetypes denoted by asterisks). It can be seen that the expression levels of these RNA binding proteins are elevated in the blastemal, and even more in the stromal tumors. Independently, we observed that exons that are elevated in these tumors are enriched for binding motifs of these regulators at their downstream introns. This indicates that in these tumors, RBFOX1 and QKI bind to the mRNA downstream of the mesenchymal-associated cassette exons and promote their inclusion (bottom sketch).

3. Discussion

In this study, we characterized the continuous heterogeneity of high-risk Wilms' tumors using gene expression and mRNA splicing. We first showed that high-risk Wilms' tumors and normal kidney samples from different patients form a continuum in gene expression latent space that is bound by stromal, blastemal, and epithelial archetypes that resemble the nephrogenic stroma (the un-induced metanephric mesenchyme), the cap mesenchyme, and the early epithelial structures of the fetal kidney, respectively. Our previous studies have yielded archetypes with similar identities for blastemal-type Wilms' tumors treated according to SIOP protocols [21], and for a meta-analysis of hundreds of Wilms' tumors from a series of studies by the COG [20]. Our results here show that the identity of the archetypes is conserved also for an additional set of high-risk Wilms' tumors [10] (DAWT and relapsed FHWT). We next explored the alternative mRNA splicing landscape of high-risk Wilms' tumors and found that many transcripts in these tumors are alternatively spliced in different regions of gene expression latent space. Moreover, we found that a significant fraction of these transcripts are associated with EMT and muscle development. In particular, we observed an elevated expression of muscle-specific isoforms in tumors located near the stromal archetype, which we did not discuss in previous studies [20,21].

The development of Wilms' tumors has been tightly linked to aberrations in fetal kidney development. Two central processes in embryonic nephrogenesis are the induction of the metanephric mesenchyme, resulting in the condensation of cells from the nephrogenic stroma around the ureteric tip to form the cap mesenchyme, and the mesenchymal-to-epithelial transition (MET), by which cells from the cap mesenchyme progressively differentiate through a series of transformations to form the various epithelial segments of the nephron. Both the kidney metanephric mesenchyme and muscle tissues originate from the fetal mesoderm, but whereas the kidney is generally accepted to be derived from the intermediate mesoderm, muscles are paraxial-mesoderm-derived. Interestingly, although much is known about the intermediate mesoderm origin of the nephrogenic lineage, the origin of the stromal lineage is much less clear, and some lines of evidence even suggest that the renal stroma could be derived from the paraxial mesoderm [72].

The finding of elevated muscle-specific isoforms in tumors near the stromal archetype is therefore also relevant for the identification of the developmental stage and cell type of origin of Wilms' tumors. Generally, Wilms' tumors are believed to originate from the nephron progenitor cells in the cap mesenchyme, explaining the blastemal and epithelial components found in Wilms' tumors [2]. Stromal components in the tumors might be more difficult to explain from these cells as the cell of origin. Possibly, different types of Wilms' tumors have different stages and cell-types of origin, leading to their different histological characteristics. A subset of tumors shows clearly recognizable ectopic muscle development, and these are usually *WT1*-mutant cases [73,74]. Data from mouse models [75], human iPSC-derived organoids [76], and a recent analysis of bilateral Wilms' tumors (also often *WT1*-mutant; [77]) all support an earlier stage of origin of the tumors than other cases, for instance, before or at the stage where the intermediate and paraxial mesoderm separate. If the renal stroma would indeed be of paraxial origin, this would offer a different developmental trajectory for these tumors. It should be noted, however, that *WT1*-mutant cases are

favorable histology Wilms' tumors, not high-risk cases as described in the present study. A final possibility to explain the finding of muscle-specific isoform expression in tumors near the stromal archetype could be that after initial differentiation into an early kidney lineage, they had undergone dedifferentiation or differentiation to a muscle-like lineage, or even direct transdifferentiation from the nephrogenic lineage to a stromal identity with more characteristics of early myoblasts. A comparable shift has been observed in the nephron progenitor cell-specific loss of *Pax2* [78]. Future functional studies will be required to test these possibilities in the context of Wilms' tumor mutations.

Our analysis allowed us to model each tumor as a convex mixture of archetypal gene expression profiles, each of which reflects an "extreme" or "idealized" biological state. The observed configuration of the tumors in gene expression latent space allowed us to infer splicing differences between neighbors of these archetypal tumors. This analysis makes a strong assumption, namely that each tumor can be effectively summarized as a simple combination of k different archetypes, where k is a subjective hyperparameter. The number of chosen archetypes must reflect the heterogeneity and complexity of the data while remaining sufficiently low as to allow the interpretation of each archetype's biological identity with confidence. We found that using either three or four archetypes allowed us to interpret each archetype confidently.

We note that in our study, we analyzed a small sample size of 130 tumors, and that the tumors within this study were restricted to high-risk cases as defined by the COG treatment protocol. Our dataset also included only five recurrent samples, limiting the power of a differential analysis between primary and recurrent patients.

We also note that we used a fetal kidney single-cell dataset for the deconvolution procedure. A recent analysis generated single-cell datasets from Wilms' tumor patients to examine the differences in cellular signals between treatment-naïve and post-chemotherapy Wilms' tumors [79]. However, such single-cell Wilms' tumor datasets are few and do not fully reflect the transcriptional heterogeneity between Wilms' tumor patients.

Recent studies have investigated the possible clinical implications for the link between alternative splicing and cancer [80]. First, alternatively spliced variants have the potential to be used as biomarkers for cancer diagnosis [81]. Second, there are potential applications for the targeted treatment of aberrantly spliced RNA to restore the normal function of impacted genes. For example, a recent study examined neurons affected by the group of neurodegenerative diseases known as TDP-43 proteinopathies (such as ALS) [42]. Depletion of the RNA binding protein TDP-43 from the nucleus can lead to the expression of aberrant RNA transcripts of the gene *STMN2* which exhibit so-called cryptic exons, a class of exons found within noncoding intronic regions whose inclusion into mature RNA can lead to a premature stop codon associated with these diseases. The authors reported that treatment with antisense oligonucleotides (ASOs) restored normal RNA splicing, *STMN2* expression, and axonal regeneration capacity in cultured motor neurons, suggesting the potential for therapeutic intervention aimed at targeting cryptic exons to mitigate RNA mis-splicing and its associated diseases [82]. Other studies have investigated potential applications for targeted treatment using a variety of new technologies to inhibit the splicing of aberrantly spliced genes [83], to affect the function of splicing factors [84], and to design novel antigens not recognized by the immune system [85,86]. We believe that our results may help facilitate the development of novel therapies aimed at targeting specific mRNA splicing mechanisms associated with Wilms' tumors, as well as assist in pinpointing the developmental trajectory of Wilms' tumors in different patients.

4. Methods

4.1. Datasets and Preprocessing

A total of 136 BAM files from the TARGET Wilms' Tumor study [10] were downloaded from the Genomics Data Commons (GDC) data portal using the GDC Data Transfer Tool Client (accessed on 1 December 2020). Each BAM file was sorted and converted to a paired-end fastq file using SAMtools (version 1.19.2) [87] and realigned to a reference genome

(hg38) using STAR (version 2.7.3a) [88] to produce a gene expression counts matrix. Clinical data were manually downloaded from the GDC website. The raw gene expression counts were normalized using DESeq2 (version 1.40.2) [89] and then a modified log-transform was performed ($\log_2(1 + x)$). Genes with zero counts in all samples were removed from the analysis.

4.2. Data Visualization, GO Enrichment Analysis, and Clustering

PCA was performed using the “Scikit-learn” python package (version 1.2.2) [90] and PCA plots were drawn using the Matplotlib python package (version 3.7.1) [91]. Gene Ontology enrichment analysis was performed using ToppGene [48]. Hierarchical clustering was performed using the “ComplexHeatmap” R package (version 2.16.0) [92], with standardized rows (=genes). For the clustering of gene expression, we used the Pearson correlation distance metric with complete linkage. For the clustering of mRNA splicing inclusion levels, we used the Euclidean distance metric with average linkage.

4.3. Archetype Analysis

The “ParTI_lite” MATLAB function (<https://www.weizmann.ac.il/mcb/UriAlon/download/ParTI>, last accessed on 22 May 2023) [43] was used to find the best fitting polytope encompassing the data points in latent space, where each data point is the normalized gene expression vector of a specific sample. The vertices of the best fitting polytope represent archetypes—idealized tumors/tissues or cell types which specialize in a particular biological task. The “performance” of each sample in this task is determined by its geometrical distance from each of the archetypes, and the identity of the biological tasks that characterize each archetype can be inferred from genes enriched in a nearby latent space [43].

4.4. Cell Deconvolution

For cell deconvolution, we used the CPM [44] algorithm as implemented in the “scBio” R package (version 0.1.6). We used the human fetal kidney single-cell RNAseq dataset from the Kidney Cell Atlas [49] as a reference single-cell gene expression matrix and a UMAP [93] embedding of this dataset as the cell state space. We used the default values for the “modelSize”, “minSelection”, and “neighborhoodSize” parameters, and we set the parameter “quantifyTypes” = T, which instructs the algorithm to quantify the proportions of different reference cell sub-populations in each bulk sample, in addition to the abundance values of each reference single-cell.

4.5. Alternative Splicing and RNA Binding Motif Enrichment Analysis

We used rMATS (version 4.0.2) [50] to find mRNA splicing events with significant inclusion-level differences between groups of samples located near each of the archetypes. We manually inspected top alternatively spliced transcripts in the IGV genome browser [94]. We then used the tables output by rMATS as input to rMAPS (version 2.2.0) [68,69] in order to identify putative splicing regulators by searching for binding motifs belonging to known RNA binding proteins (RBPs) that are also enriched in the vicinity of alternatively spliced exons. Additionally, in order to draw the heatmap of inclusion levels for all samples in the dataset, we reran rMATS for all samples with the command-line option “-cstat 0”.

The list of RNA binding proteins was obtained from the rMAPS website (<http://rmaps.cecsresearch.org/Help/RNABindingProtein>, accessed on 14 April 2024). In addition to the RNA binding motifs that were tested using the default settings on the rMAPS website, we also conducted tests on additional UGG-enriched motifs that have been previously identified as binding sites for the RNA binding proteins ESRP1 [95,96] and ESRP2 [97] (Table S6). Following Yang et al. [71] and the CISBP-RNA database [98] (<http://cisbp-rna.ccb.utoronto.ca>, accessed on 14 April 2024), we assumed that the proteins RBFOX1 and RBFOX2 both bind to the same mRNA motif ([AT]GCATG[AC]).

Supplementary Materials: The supporting information can be downloaded at <https://www.mdpi.com/article/10.3390/ijms25084520/s1>. References [10,15,34,35,54,55,61–67,71,95–101] are cited in the supplementary materials.

Author Contributions: Study initiation and conception, Y.T. and T.K.; data analysis, Y.T. and T.K.; other intellectual contributions, A.U., B.D., P.H. and J.G.; manuscript writing, Y.T., P.H. and T.K. All authors have read and agreed to the published version of the manuscript.

Funding: Y.T. and T.K. were supported by the Israel Science Foundation (ICORE no. 1902/12 and Grant nos. 1634/13, 2017/13, and 1814/20), the Israel Ministry of Health (Grant no. 3-10146), the EU-FP7 (Marie Curie International Reintegration Grant no. 618592), the Data Science Institute at Bar-Ilan University (seed grant), the ICRF (Grant no. 19-101-PG), the Israel Ministry of Science (Grant no. 3-16220), and the Israel Cancer Association (Grant no. 20240114). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The results published here are in whole or in part based upon data generated by the Therapeutically Applicable Research to Generate Effective Treatments (TARGET) initiative, phs000218, managed by the NCI. The data used for this analysis are available through the GDC (<https://gdc.cancer.gov/about-data/publications#/?groups=&years=&programs=TARGET&order=desc>). Information about TARGET can be found at <https://www.cancer.gov/ccg/research/genome-sequencing/target/about>, both accessed on 15 November 2020.

Acknowledgments: We wish to thank Yochai Israelashvili, Naomi Pode-Shakked, Iddo Ben-Dov, Thomas Tuschl, and all members of our lab for helpful comments and suggestions.

Conflicts of Interest: The authors have declared that no competing interests exist.

References

1. Beckwith, J.B.; Kiviat, N.B.; Bonadio, J.F. Nephrogenic rests, nephroblastomatosis, and the pathogenesis of Wilms' tumor. *Pediatr. Pathol.* **1990**, *10*, 1–36. [[CrossRef](#)] [[PubMed](#)]
2. Hohenstein, P.; Pritchard-Jones, K.; Charlton, J. The yin and yang of kidney development and Wilms' tumors. *Genes. Dev.* **2015**, *29*, 467–482. [[CrossRef](#)]
3. Park, S.; Bernard, A.; Bove, K.E.; Sens, D.A.; Hazen-Martin, D.J.; Garvin, A.J.; Haber, D.A. Inactivation of WT1 in nephrogenic rests, genetic precursors to Wilms' tumour. *Nat. Genet.* **1993**, *5*, 363–367. [[CrossRef](#)] [[PubMed](#)]
4. Popov, S.D.; Sebire, N.J.; Vujanic, G.M. Wilms' Tumour—Histology and Differential Diagnosis. In *Wilms Tumor*; Van den Heuvel-Eibrink, M.M., Ed.; Codon Publications: Brisbane, Australia, 2016.
5. Little, M.; Georgas, K.; Pennisi, D.; Wilkinson, L. Chapter Five—Kidney Development: Two Tales of Tubulogenesis. In *Current Topics in Developmental Biology*; Koopman, P., Ed.; Academic Press: Cambridge, MA, USA, 2010; pp. 193–229. [[CrossRef](#)]
6. Dome, J.S.; Fernandez, C.V.; Mullen, E.A.; Kalapurakal, J.A.; Geller, J.I.; Huff, V.; Gratias, E.J.; Dix, D.B.; Ehrlich, P.F.; Khanna, G.; et al. Children's Oncology Group's 2013 Blueprint for Research: Renal Tumors. *Pediatr. Blood Cancer* **2013**, *60*, 994–1000. [[CrossRef](#)]
7. Szychot, E.; Apps, J.; Pritchard-Jones, K. Wilms' tumor: Biology, diagnosis and treatment. *Transl. Pediatr.* **2014**, *3*, 12–24. [[CrossRef](#)]
8. Murphy, A.J.; Axt, J.R.; de Caestecker, C.; Pierce, J.; Correa, H.; Seeley, E.H.; Caprioli, R.M.; Newton, M.W.; de Caestecker, M.P.; Lovvorn, H.N. Molecular characterization of Wilms tumor from a resource-constrained region of sub-Saharan Africa. *Int. J. Cancer* **2012**, *131*, E983–E994. [[CrossRef](#)]
9. Daw, N.C.; Chi, Y.-Y.; Kim, Y.; Mullen, E.A.; Kalapurakal, J.A.; Tian, J.; Khanna, G.; Geller, J.I.; Perlman, E.J.; Ehrlich, P.F.; et al. Treatment of stage I anaplastic Wilms' tumour: A report from the Children's Oncology Group AREN0321 study. *Eur. J. Cancer* **2019**, *118*, 58–66. [[CrossRef](#)]
10. 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. A Children's Oncology Group and TARGET initiative exploring the genetic landscape of Wilms tumor. *Nat. Genet.* **2017**, *49*, 1487–1494. [[CrossRef](#)] [[PubMed](#)]
11. Vujančić, G.M.; Sandstedt, B. The pathology of Wilms' tumour (nephroblastoma): The International Society of Paediatric Oncology approach. *J. Clin. Pathol.* **2010**, *63*, 102–109. [[CrossRef](#)]
12. Ortiz, M.V.; Koenig, C.; Armstrong, A.E.; Brok, J.; de Camargo, B.; Mavinkurve-Groothuis, A.M.C.; Herrera, T.B.V.; Venkatramani, R.; Woods, A.D.; Dome, J.S.; et al. Advances in the clinical management of high-risk Wilms tumors. *Pediatr. Blood Cancer* **2023**, *70*, e30153. [[CrossRef](#)]
13. Bhutani, N.; Kajal, P.; Sharma, U. Many faces of Wilms Tumor: Recent advances and future directions. *Ann. Med. Surg.* **2021**, *64*, 102202. [[CrossRef](#)]

14. Gadd, S.; Huff, V.; Huang, C.-C.; Ruteshouser, E.C.; Dome, J.S.; Grundy, P.E.; Breslow, N.; Jennings, L.; Green, D.M.; Beckwith, J.B.; et al. 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. [[CrossRef](#)] [[PubMed](#)]
15. Wegert, J.; Ishaque, N.; Vardapour, R.; Geörg, C.; Gu, Z.; Bieg, M.; Ziegler, B.; Bausenwein, S.; Nourkami, N.; Ludwig, N.; et al. Mutations in the SIX1/2 Pathway and the DROSHA/DGCR8 miRNA Microprocessor Complex Underlie High-Risk Blastemal Type Wilms Tumors. *Cancer Cell* **2015**, *27*, 298–311. [[CrossRef](#)] [[PubMed](#)]
16. Lu, J.; Tao, Y.-F.; Li, Z.-H.; Cao, L.; Hu, S.-Y.; Wang, N.-N.; Du, X.-J.; Sun, L.-C.; Zhao, W.-L.; Xiao, P.-F.; et al. Analyzing the gene expression profile of anaplastic histology Wilms' tumor with real-time polymerase chain reaction arrays. *Cancer Cell Int.* **2015**, *15*, 44. [[CrossRef](#)] [[PubMed](#)]
17. Takahashi, M.; Yang, X.J.; Lavery, T.T.; Furge, K.A.; Williams, B.O.; Tretiakova, M.; Montag, A.; Vogelzang, N.J.; Re, G.G.; Garvin, A.J.; et al. Gene Expression Profiling of Favorable Histology Wilms Tumors and Its Correlation with Clinical Features. *Cancer Res.* **2002**, *62*, 6598–6605.
18. Abdel-Monem, M.M.; El-Khawaga, O.Y.; Awadalla, A.A.; Hafez, A.T.; Ahmed, A.E.; Abdelhameed, M.; Abdelhalim, A. Gene expression analysis and the risk of relapse in favorable histology Wilms' tumor. *Arab. J. Urol.* **2023**, *21*, 45–51. [[CrossRef](#)]
19. Maschietto, M.; Piccoli, F.S.; Costa, C.M.L.; Camargo, L.P.; Neves, J.I.; Grundy, P.E.; Brentani, H.; Soares, F.A.; de Camargo, B.; Carraro, D.M. Gene expression analysis of blastemal component reveals genes associated with relapse mechanism in Wilms tumour. *Eur. J. Cancer* **2011**, *47*, 2715–2722. [[CrossRef](#)] [[PubMed](#)]
20. Trink, A.; Kanter, I.; Pode-Shakked, N.; Urbach, A.; Dekel, B.; Kalisky, T. Geometry of Gene Expression Space of Wilms' Tumors From Human Patients. *Neoplasia* **2018**, *20*, 871–881. [[CrossRef](#)] [[PubMed](#)]
21. Trink, Y.; Urbach, A.; Dekel, B.; Hohenstein, P.; Goldberger, J.; Kalisky, T. Characterization of Continuous Transcriptional Heterogeneity in High-Risk Blastemal-Type Wilms' Tumors Using Unsupervised Machine Learning. *Int. J. Mol. Sci.* **2023**, *24*, 3532. [[CrossRef](#)]
22. Carbonetto, P.; Sarkar, A.; Wang, Z.; Stephens, M. Non-negative matrix factorization algorithms greatly improve topic model fits. *arXiv* **2022**, arXiv:2105.13440.
23. Bradley, R.K.; Anczuków, O. RNA splicing dysregulation and the hallmarks of cancer. *Nat. Rev. Cancer* **2023**, *23*, 135–155. [[CrossRef](#)] [[PubMed](#)]
24. Qian, C.-N.; Mei, Y.; Zhang, J. Cancer metastasis: Issues and challenges. *Chin. J. Cancer* **2017**, *36*, 38. [[CrossRef](#)] [[PubMed](#)]
25. Kalluri, R.; Weinberg, R.A. The basics of epithelial-mesenchymal transition. *J. Clin. Investig.* **2009**, *119*, 1420–1428. [[CrossRef](#)]
26. Mittal, V. Epithelial Mesenchymal Transition in Tumor Metastasis. *Annu. Rev. Pathol.* **2018**, *13*, 395–412. [[CrossRef](#)] [[PubMed](#)]
27. Ghigna, C.; Giordano, S.; Shen, H.; Benvenuto, F.; Castiglioni, F.; Comoglio, P.M.; Green, M.R.; Riva, S.; Biamonti, G. Cell motility is controlled by SF2/ASF through alternative splicing of the Ron protooncogene. *Mol. Cell* **2005**, *20*, 881–890. [[CrossRef](#)] [[PubMed](#)]
28. Warzecha, C.C.; Sato, T.K.; Nabet, B.; Hogenesch, J.B.; Carstens, R.P. ESRP1 and ESRP2 are epithelial cell-type-specific regulators of FGFR2 splicing. *Mol. Cell* **2009**, *33*, 591–601. [[CrossRef](#)]
29. Amann, T.; Bataille, F.; Spruss, T.; Dettmer, K.; Wild, P.; Liedtke, C.; Mühlbauer, M.; Kiefer, P.; Oefner, P.J.; Trautwein, C.; et al. Reduced Expression of Fibroblast Growth Factor Receptor 2IIIb in Hepatocellular Carcinoma Induces a More Aggressive Growth. *Am. J. Pathol.* **2010**, *176*, 1433–1442. [[CrossRef](#)]
30. Roy Burman, D.; Das, S.; Das, C.; Bhattacharya, R. Alternative splicing modulates cancer aggressiveness: Role in EMT/metastasis and chemoresistance. *Mol. Biol. Rep.* **2021**, *48*, 897–914. [[CrossRef](#)]
31. Acevedo, V.D.; Ittmann, M.; Spencer, D.M. Paths of FGFR-driven tumorigenesis. *Cell Cycle* **2009**, *8*, 580–588. [[CrossRef](#)]
32. Silva, A.L.; Carmo, F.; Bugalho, M.J. RAC1b overexpression in papillary thyroid carcinoma: A role to unravel. *Eur. J. Endocrinol.* **2013**, *168*, 795–804. [[CrossRef](#)]
33. Dvinge, H.; Kim, E.; Abdel-Wahab, O.; Bradley, R.K. RNA splicing factors as oncoproteins and tumour suppressors. *Nat. Rev. Cancer* **2016**, *16*, 413–430. [[CrossRef](#)] [[PubMed](#)]
34. Gonçalves, V.; Matos, P.; Jordan, P. Antagonistic SR proteins regulate alternative splicing of tumor-related Rac1b downstream of the PI3-kinase and Wnt pathways. *Human Mol. Genet.* **2009**, *18*, 3696–3707. [[CrossRef](#)] [[PubMed](#)]
35. Sebestyén, E.; Singh, B.; Miñana, B.; Pagès, A.; Mateo, F.; Pujana, M.A.; Valcárcel, J.; Eyra, E. Large-scale analysis of genome and transcriptome alterations in multiple tumors unveils novel cancer-relevant splicing networks. *Genome Res.* **2016**, *26*, 732–744. [[CrossRef](#)] [[PubMed](#)]
36. Seiler, M.; Peng, S.; Agrawal, A.A.; Palacino, J.; Teng, T.; Zhu, P.; Smith, P.G.; The Cancer Genome Atlas Research Network; Buonamici, S.; Yu, L. Somatic Mutational Landscape of Splicing Factor Genes and Their Functional Consequences across 33 Cancer Types. *Cell Rep.* **2018**, *23*, 282–296.e4. [[CrossRef](#)] [[PubMed](#)]
37. Zhang, J.; Manley, J.L. Misregulation of Pre-mRNA Alternative Splicing in Cancer. *Cancer Discov.* **2013**, *3*, 1228–1237. [[CrossRef](#)] [[PubMed](#)]
38. Yae, T.; Tsuchihashi, K.; Ishimoto, T.; Motohara, T.; Yoshikawa, M.; Yoshida, G.J.; Wada, T.; Masuko, T.; Mogushi, K.; Tanaka, H.; et al. Alternative splicing of CD44 mRNA by ESRP1 enhances lung colonization of metastatic cancer cell. *Nat. Commun.* **2012**, *3*, 883. [[CrossRef](#)] [[PubMed](#)]
39. Legge, D.; Li, L.; Moriarty, W.; Lee, D.; Szemes, M.; Zahed, A.; Panousopoulos, L.; Chung, W.Y.; Aghabi, Y.; Barratt, J.; et al. The epithelial splicing regulator ESRP2 is epigenetically repressed by DNA hypermethylation in Wilms tumour and acts as a tumour suppressor. *Mol. Oncol.* **2022**, *16*, 630–647. [[CrossRef](#)] [[PubMed](#)]

40. Wineberg, Y.; Kanter, I.; Ben-Haim, N.; Pode-Shakked, N.; Bucris, E.; Bar-Lev, T.H.; Oriol, S.; Reinus, H.; Yehuda, Y.; Gershon, R.; et al. Characterization of alternative mRNA splicing in cultured cell populations representing progressive stages of human fetal kidney development. *Sci. Rep.* **2022**, *12*, 19548. [[CrossRef](#)] [[PubMed](#)]
41. Singh, B.; Eyra, E. The role of alternative splicing in cancer. *Transcription* **2017**, *8*, 91–98. [[CrossRef](#)]
42. Baughn, M.W.; Melamed, Z.; López-Erauskin, J.; Beccari, M.S.; Ling, K.; Zuberi, A.; Presa, M.; Gonzalo-Gil, E.; Maimon, R.; Vazquez-Sanchez, S.; et al. Mechanism of STMN2 cryptic splice-polyadenylation and its correction for TDP-43 proteinopathies. *Science* **2023**, *379*, 1140–1149. [[CrossRef](#)]
43. Hart, Y.; Sheftel, H.; Hausser, J.; Szekely, P.; Ben-Moshe, N.B.; Korem, Y.; Tendler, A.; E Mayo, A.; Alon, U. Inferring biological tasks using Pareto analysis of high-dimensional data. *Nat. Methods* **2015**, *12*, 233–235. [[CrossRef](#)] [[PubMed](#)]
44. Frishberg, A.; Peshes-Yaloz, N.; Cohn, O.; Rosentul, D.; Steuerman, Y.; Valadarsky, L.; Yankovitz, G.; Mandelboim, M.; Iraqi, F.A.; Amit, I.; et al. Cell composition analysis of bulk genomics using single-cell data. *Nat. Methods* **2019**, *16*, 327–332. [[CrossRef](#)] [[PubMed](#)]
45. Morup, M.; Hansen, L.K. Archetypal analysis for machine learning and data mining. *Neurocomputing* **2012**, *80*, 54–63. [[CrossRef](#)]
46. Korem, Y.; Szekely, P.; Hart, Y.; Sheftel, H.; Hausser, J.; Mayo, A.; Rothenberg, M.E.; Kalisky, T.; Alon, U. Geometry of the Gene Expression Space of Individual Cells. *PLoS Comput. Biol.* **2015**, *11*, e1004224. [[CrossRef](#)] [[PubMed](#)]
47. Shoval, O.; Sheftel, H.; Shinar, G.; Hart, Y.; Ramote, O.; Mayo, A.; Dekel, E.; Kavanagh, K.; Alon, U. Evolutionary Trade-Offs, Pareto Optimality, and the Geometry of Phenotype Space. *Science* **2012**, *336*, 1157–1160. [[CrossRef](#)] [[PubMed](#)]
48. Chen, J.; Xu, H.; Aronow, B.J.; Jegga, A.G. Improved human disease candidate gene prioritization using mouse phenotype. *BMC Bioinform.* **2007**, *8*, 392. [[CrossRef](#)] [[PubMed](#)]
49. Stewart, B.J.; Ferdinand, J.R.; Young, M.D.; Mitchell, T.J.; Loudon, K.W.; Riding, A.M.; Richoz, N.; Frazer, G.L.; Staniforth, J.U.L.; Braga, F.A.V.; et al. Spatiotemporal immune zonation of the human kidney. *Science* **2019**, *365*, 1461–1466. [[CrossRef](#)] [[PubMed](#)]
50. Shen, S.; Park, J.W.; Lu, Z.; Lin, L.; Henry, M.D.; Wu, Y.N.; Zhou, Q.; Xing, Y. rMATS: Robust and flexible detection of differential alternative splicing from replicate RNA-Seq data. *Proc. Natl. Acad. Sci. USA* **2014**, *111*, E5593–E5601. [[CrossRef](#)] [[PubMed](#)]
51. Hovhannisyann, R.H.; Warzecha, C.C.; Carstens, R.P. Characterization of sequences and mechanisms through which ISE/ISS-3 regulates FGFR2 splicing. *Nucleic Acids Res.* **2006**, *34*, 373–385. [[CrossRef](#)]
52. Ranieri, D.; Rosato, B.; Nanni, M.; Belleudi, F.; Torrisi, M.R. Expression of the FGFR2c mesenchymal splicing variant in human keratinocytes inhibits differentiation and promotes invasion. *Mol. Carcinog.* **2018**, *57*, 272–283. [[CrossRef](#)]
53. Warzecha, C.C.; Shen, S.; Xing, Y.; Carstens, R.P. The epithelial splicing factors ESRP1 and ESRP2 positively and negatively regulate diverse types of alternative splicing events. *RNA Biol.* **2009**, *6*, 546–562. [[CrossRef](#)] [[PubMed](#)]
54. Qin, J.; Xu, J. Arginine methylation in the epithelial-to-mesenchymal transition. *FEBS J.* **2022**, *289*, 7292–7303. [[CrossRef](#)] [[PubMed](#)]
55. Zhong, J.; Cao, R.-X.; Zu, X.-Y.; Hong, T.; Yang, J.; Liu, L.; Xiao, X.; Ding, W.; Zhao, Q.; Liu, J.; et al. Identification and characterization of novel spliced variants of PRMT2 in breast carcinoma. *FEBS J.* **2012**, *279*, 316–335. [[CrossRef](#)] [[PubMed](#)]
56. Kubo, E.; Hasanova, N.; Fatma, N.; Sasaki, H.; Singh, D.P. Elevated tropomyosin expression is associated with epithelial-mesenchymal transition of lens epithelial cells. *J. Cell. Mol. Med.* **2013**, *17*, 212–221. [[CrossRef](#)] [[PubMed](#)]
57. Shibata, T.; Shibata, S.; Ishigaki, Y.; Kiyokawa, E.; Ikawa, M.; Singh, D.P.; Sasaki, H.; Kubo, E. Tropomyosin 2 heterozygous knockout in mice using CRISPR-Cas9 system displays the inhibition of injury-induced epithelial-mesenchymal transition, and lens opacity. *Mech. Ageing Dev.* **2018**, *171*, 24–30. [[CrossRef](#)] [[PubMed](#)]
58. Assinder, S.J.; Au, E.; Dong, Q.; Winnick, C. A novel splice variant of the β -tropomyosin (TPM2) gene in prostate cancer. *Mol. Carcinog.* **2010**, *49*, 525–531. [[CrossRef](#)] [[PubMed](#)]
59. Mele, V.; Basso, C.; Governa, V.; Glaus Garzon, J.F.; Muraro, M.G.; Däster, S.; Nebiker, C.A.; Mechera, R.; Bolli, M.; Schmidt, A.; et al. Identification of TPM2 and CNN1 as Novel Prognostic Markers in Functionally Characterized Human Colon Cancer-Associated Stromal Cells. *Cancers* **2022**, *14*, 2024. [[CrossRef](#)] [[PubMed](#)]
60. Mroczek, M.; Kabzińska, D.; Chrzanowska, K.H.; Pronicki, M.; Kochański, A. A novel TPM2 gene splice-site mutation causes severe congenital myopathy with arthrogryposis and dysmorphic features. *J. Appl. Genet.* **2017**, *58*, 199–203. [[CrossRef](#)] [[PubMed](#)]
61. Gimona, M.; Watakabe, A.; Helfman, D.M. Specificity of dimer formation in tropomyosins: Influence of alternatively spliced exons on homodimer and heterodimer assembly. *Proc. Natl. Acad. Sci. USA* **1995**, *92*, 9776–9780. [[CrossRef](#)] [[PubMed](#)]
62. Li, J.; Choi, P.S.; Chaffer, C.L.; Labella, K.; Hwang, J.H.; Giacomelli, A.O.; Kim, J.W.; Ilic, N.; Doench, J.G.; Ly, S.H.; et al. An alternative splicing switch in FLNB promotes the mesenchymal cell state in human breast cancer. *eLife* **2018**, *7*, e37184. [[CrossRef](#)]
63. Keeton, T.P.; Burk, S.E.; Shull, G.E. Alternative splicing of exons encoding the calmodulin-binding domains and C termini of plasma membrane Ca^{2+} -ATPase isoforms 1, 2, 3, and 4. *J. Biol. Chem.* **1993**, *268*, 2740–2748. [[CrossRef](#)]
64. Stauffer, T.P.; Hilfiker, H.; Carafoli, E.; Strehler, E.E. Quantitative analysis of alternative splicing options of human plasma membrane calcium pump genes. *J. Biol. Chem.* **1993**, *268*, 25993–26003. [[CrossRef](#)] [[PubMed](#)]
65. Zhu, B.; Ramachandran, B.; Gulick, T. Alternative Pre-mRNA Splicing Governs Expression of a Conserved Acidic Transactivation Domain in Myocyte Enhancer Factor 2 Factors of Striated Muscle and Brain. *J. Biol. Chem.* **2005**, *280*, 28749–28760. [[CrossRef](#)]
66. Singh, R.K.; Xia, Z.; Bland, C.S.; Kalsotra, A.; Scavuzzo, M.A.; Curk, T.; Ule, J.; Li, W.; Cooper, T.A. Rbfox2-Coordinated Alternative Splicing of Mef2d and Rock2 Controls Myoblast Fusion during Myogenesis. *Mol. Cell* **2014**, *55*, 592–603. [[CrossRef](#)] [[PubMed](#)]
67. Runfola, V.; Sebastian, S.; Dilworth, F.J.; Gabellini, D. Rbfox proteins regulate tissue-specific alternative splicing of Mef2D required for muscle differentiation. *J. Cell Sci.* **2015**, *128*, 631–637. [[CrossRef](#)] [[PubMed](#)]

68. Hwang, J.Y.; Jung, S.; Kook, T.L.; Rouchka, E.C.; Bok, J.; Park, J.W. rMAPS2: An update of the RNA map analysis and plotting server for alternative splicing regulation. *Nucleic Acids Res.* **2020**, *48*, W300–W306. [[CrossRef](#)]
69. Park, J.W.; Jung, S.; Rouchka, E.C.; Tseng, Y.-T.; Xing, Y. rMAPS: RNA map analysis and plotting server for alternative exon regulation. *Nucleic Acids Res.* **2016**, *44*, W333–W338. [[CrossRef](#)]
70. Lyu, J.; Cheng, C. Regulation of Alternative Splicing during Epithelial-Mesenchymal Transition. *Cells Tissues Organs* **2022**, *211*, 238–251. [[CrossRef](#)]
71. Yang, Y.; Park, J.W.; Bebee, T.W.; Warzecha, C.C.; Guo, Y.; Shang, X.; Xing, Y.; Carstens, R.P. Determination of a Comprehensive Alternative Splicing Regulatory Network and Combinatorial Regulation by Key Factors during the Epithelial-to-Mesenchymal Transition. *Mol. Cell. Biol.* **2016**, *36*, 1704–1719. [[CrossRef](#)]
72. Wilson, S.B.; Little, M.H. The origin and role of the renal stroma. *Development* **2021**, *148*, dev199886. [[CrossRef](#)]
73. Miyagawa, K.; Kent, J.; Moore, A.; Charlier, J.P.; Little, M.H.; Williamson, K.A.; Kelsey, A.; Brown, K.W.; Hassam, S.; Briner, J.; et al. Loss of WT1 function leads to ectopic myogenesis in Wilms' tumour. *Nat. Genet.* **1998**, *18*, 15–17. [[CrossRef](#)] [[PubMed](#)]
74. Schumacher, V.; Schuhen, S.; Sonner, S.; Weirich, A.; Leuschner, I.; Harms, D.; Licht, J.; Roberts, S.; Royer-Pokora, B. Two molecular subgroups of Wilms' tumors with or without WT1 mutations. *Clin. Cancer Res.* **2003**, *9*, 2005–2014. [[PubMed](#)]
75. Berry, R.L.; Ozdemir, D.D.; Aronow, B.; Lindström, N.O.; Dudnakova, T.; Thornburn, A.; Perry, P.; Baldock, R.; Armit, C.; Joshi, A.; et al. Deducing the stage of origin of Wilms' tumours from a developmental series of Wt1-mutant mice. *Dis. Models Mech.* **2015**, *8*, 903–917. [[CrossRef](#)]
76. Waehle, V.; Ungricht, R.; Hoppe, P.S.; Betschinger, J. The tumor suppressor WT1 drives progenitor cell progression and epithelialization to prevent Wilms tumorigenesis in human kidney organoids. *Stem Cell Rep.* **2021**, *16*, 2107–2117. [[CrossRef](#)] [[PubMed](#)]
77. Murphy, A.J.; Cheng, C.; Williams, J.; Shaw, T.I.; Pinto, E.M.; Dieseldorff-Jones, K.; Brzezinski, J.; Renfro, L.A.; Tornwall, B.; Huff, V.; et al. Genetic and epigenetic features of bilateral Wilms tumor predisposition in patients from the Children's Oncology Group AREN18B5-Q. *Nat. Commun.* **2023**, *14*, 8006. [[CrossRef](#)] [[PubMed](#)]
78. Naiman, N.; Fujioka, K.; Fujino, M.; Valerius, M.T.; Potter, S.S.; McMahon, A.P.; Kobayashi, A. Repression of Interstitial Identity in Nephron Progenitor Cells by Pax2 Establishes the Nephron-Interstitium Boundary during Kidney Development. *Dev. Cell* **2017**, *41*, 349–365.e3. [[CrossRef](#)] [[PubMed](#)]
79. Young, M.D.; Mitchell, T.J.; Custers, L.; Margaritis, T.; Morales-Rodriguez, F.; Kwakwa, K.; Khabirova, E.; Kildisiute, G.; Oliver, T.R.W.; de Krijger, R.R.; et al. Single cell derived mRNA signals across human kidney tumors. *Nat. Commun.* **2021**, *12*, 3896. [[CrossRef](#)] [[PubMed](#)]
80. Zhang, Y.; Qian, J.; Gu, C.; Yang, Y. Alternative splicing and cancer: A systematic review. *Signal Transduct. Target. Ther.* **2021**, *6*, 78. [[CrossRef](#)]
81. Kim, Y.-J.; Kim, H.-S. Alternative Splicing and Its Impact as a Cancer Diagnostic Marker. *Genom. Inform.* **2012**, *10*, 74–80. [[CrossRef](#)]
82. O'Brien, N.; Mizielińska, S. A cryptic clue to neurodegeneration? *Science* **2023**, *379*, 1090–1091. [[CrossRef](#)]
83. Ohe, K.; Hagiwara, M. Modulation of Alternative Splicing with Chemical Compounds in New Therapeutics for Human Diseases. *ACS Chem. Biol.* **2015**, *10*, 914–924. [[CrossRef](#)] [[PubMed](#)]
84. de Fraipont, F.; Gazzeri, S.; Cho, W.C.; Eymin, B. Circular RNAs and RNA Splice Variants as Biomarkers for Prognosis and Therapeutic Response in the Liquid Biopsies of Lung Cancer Patients. *Front. Genet.* **2019**, *10*, 390. [[CrossRef](#)] [[PubMed](#)]
85. Slansky, J.E.; Spellman, P.T. Alternative Splicing in Tumors—A Path to Immunogenicity? *N. Engl. J. Med.* **2019**, *380*, 877–880. [[CrossRef](#)] [[PubMed](#)]
86. Vauchy, C.; Gamonet, C.; Ferrand, C.; Daguindau, E.; Galaine, J.; Beziaud, L.; Chauchet, A.; Dunand, C.J.H.; Deschamps, M.; Rohrllich, P.S.; et al. CD20 alternative splicing isoform generates immunogenic CD4 helper T epitopes. *Int. J. Cancer* **2015**, *137*, 116–126. [[CrossRef](#)]
87. Danecek, P.; Bonfield, J.K.; Liddle, J.; Marshall, J.; Ohan, V.; Pollard, M.O.; Whitwham, A.; Keane, T.; McCarthy, S.A.; Davies, R.M.; et al. Twelve years of SAMtools and BCFtools. *Gigascience* **2021**, *10*, giab008. [[CrossRef](#)] [[PubMed](#)]
88. Dobin, A.; Davis, C.A.; Schlesinger, F.; Drenkow, J.; Zaleski, C.; Jha, S.; Batut, P.; Chaisson, M.; Gingeras, T.R. STAR: Ultrafast universal RNA-seq aligner. *Bioinformatics* **2013**, *29*, 15–21. [[CrossRef](#)] [[PubMed](#)]
89. Love, M.I.; Huber, W.; Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* **2014**, *15*, 550. [[CrossRef](#)]
90. Pedregosa, F.; Varoquaux, G.; Gramfort, A.; Michel, V.; Thirion, B.; Grisel, O.; Blondel, M.; Prettenhofer, P.; Weiss, R.; Dubourg, V.; et al. Scikit-Learn: Machine Learning in Python. *J. Mach. Learn. Res.* **2011**, *12*, 2825–2830.
91. Hunter, J.D. Matplotlib: A 2D Graphics Environment. *Comput. Sci. Eng.* **2007**, *9*, 90–95. [[CrossRef](#)]
92. Gu, Z.; Eils, R.; Schlesner, M. Complex heatmaps reveal patterns and correlations in multidimensional genomic data. *Bioinformatics* **2016**, *32*, 2847–2849. [[CrossRef](#)]
93. McInnes, L.; Healy, J.; Saul, N.; Großberger, L. UMAP: Uniform Manifold Approximation and Projection. *J. Open Source Softw.* **2018**, *3*, 861. [[CrossRef](#)]
94. Robinson, J.T.; Thorvaldsdóttir, H.; Winckler, W.; Guttman, M.; Lander, E.S.; Getz, G.; Mesirov, J.P. Integrative genomics viewer. *Nat. Biotechnol.* **2011**, *29*, 24–26. [[CrossRef](#)] [[PubMed](#)]

95. Bebee, T.W.; Park, J.W.; Sheridan, K.I.; Warzecha, C.C.; Cieply, B.W.; Rohacek, A.M.; Xing, Y.; Carstens, R.P. The splicing regulators Esrp1 and Esrp2 direct an epithelial splicing program essential for mammalian development. *eLife* **2015**, *4*, e08954. [[CrossRef](#)] [[PubMed](#)]
96. Dittmar, K.A.; Jiang, P.; Park, J.W.; Amirikian, K.; Wan, J.; Shen, S.; Xing, Y.; Carstens, R.P. Genome-Wide Determination of a Broad ESRP-Regulated Posttranscriptional Network by High-Throughput Sequencing. *Mol. Cell. Biol.* **2012**, *32*, 1468–1482. [[CrossRef](#)]
97. Bhate, A.; Parker, D.J.; Bebee, T.W.; Ahn, J.; Arif, W.; Rahan, E.H.; Chorghade, S.; Chau, A.; Lee, J.-H.; Anakk, S.; et al. ESRP2 controls an adult splicing programme in hepatocytes to support postnatal liver maturation. *Nat. Commun.* **2015**, *6*, 8768. [[CrossRef](#)]
98. Ray, D.; Kazan, H.; Cook, K.B.; Weirauch, M.T.; Najafabadi, H.S.; Li, X.; Gueroussov, S.; Albu, M.; Zheng, H.; Yang, A.; et al. A compendium of RNA-binding motifs for decoding gene regulation. *Nature* **2013**, *499*, 172–177. [[CrossRef](#)] [[PubMed](#)]
99. Bangru, S.; Arif, W.; Seimetz, J.; Bhate, A.; Chen, J.; Rahan, E.H.; Carstens, R.P.; Anakk, S. Alternative splicing rewires Hippo signaling pathway in hepatocytes to promote liver regeneration. *Nat. Struct. Mol. Biol.* **2018**, *25*, 928–939. [[CrossRef](#)] [[PubMed](#)]
100. Warzecha, C.C.; Carstens, R.P. Complex changes in alternative pre-mRNA splicing play a central role in the epithelial-to-mesenchymal transition (EMT). *Semin. Cancer Biol.* **2012**, *22*, 417–427. [[CrossRef](#)]
101. Pascual, M.; Vicente, M.; Monferrer, L.; Artero, R. The Muscleblind family of proteins: An emerging class of regulators of developmentally programmed alternative splicing. *Differentiation* **2006**, *74*, 65–80. [[CrossRef](#)]

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.