HHS Public Access

Author manuscript

Cancer Discov. Author manuscript; available in PMC 2015 June 10.

Published in final edited form as:

Cancer Discov. 2014 February ; 4(2): 216–231. doi:10.1158/2159-8290.CD-13-0639.

Comprehensive genomic analysis of rhabdomyosarcoma reveals a landscape of alterations affecting a common genetic axis in fusion-positive and fusion-negative tumors

Jack F. Shern[1,*], **Li Chen**[1,*], **Juliann Chmielecki**[2,3], **Jun S. Wei**[1], **Rajesh Patidar**[1], **Mara Rosenberg**[2], **Lauren Ambrogio**[2], **Daniel Auclair**[2], **Jianjun Wang**[1], **Young K. Song**[1], **Catherine Tolman**[1], **Laura Hurd**[1], **Hongling Liao**[1], **Shile Zhang**[1], **Dominik Bogen**[1], **Andrew S. Brohl**[1], **Sivasish Sindiri**[1], **Daniel Catchpoole**[4], **Thomas Badgett**[1], **Gad Getz**[2], **Jaume Mora**[5], **James R. Anderson**[6], **Stephen X. Skapek**[7], **Frederic G. Barr**[8], **Matthew Meyerson**[2,3,9], **Douglas S. Hawkins**[10], and **Javed Khan**[1,#]

[1]Pediatric Oncology Branch, Oncogenomics Section, Center for Cancer Research, National Institutes of Health, Gaithersburg, MD, 20877, USA

[2]Broad Institute of MIT and Harvard, Cambridge, MA 02142

[3]Dana-Farber Cancer Institute, Medical Oncology and Center for Cancer Genome Discovery, Boston, MA 02215

[4]The Tumour Bank, The Children's Cancer Research Unit, The Children's Hospital at Westmead, Locked Bag 4001, Westmead, NSW 2145, Australia

[5]Department of Oncology. Hospital Sant Joan de Deu de Barcelona, Spain

[6]University of Nebraska Medical Center, Omaha, Nebraska

[7]Department of Pediatrics, Division of Hematology/Oncology, University of Texas Southwestern Medical Center, Dallas, TX 75390

[8]Laboratory of Pathology, National Cancer Institute, Bethesda, MD 20892

[9]Harvard Medical School, Department of Pathology, Boston, MA 02215

[10]Department of Pediatrics, Seattle Children's Hospital, Fred Hutchinson Cancer Research Center, University of Washington, Seattle WA

[#]corresponding author: Reprint Requests: Javed Khan MD, National Cancer Institute, Pediatric Oncology Branch, Advanced Technology Center, 8717 Grovemont Circle, Gaithersburg, MD 20877, khanjav@mail.nih.gov.

[*]Authors contributed equally

The authors disclose no potential conflicts of interest.

Author Contributions: All sequence data have been uploaded to dbGAP (http://www.ncbi.nlm.nih.gov/gap). Reprints and permissions information is available on request. JK, JS and MM conceived and led the project. JS, LC, JSW and JK wrote and edited the manuscript. LC performed whole-genome analysis and the related statistical tests and JS interpreted the data. JS, JC, and RP performed SNP array analysis. RP and JJW performed analysis and interpretation of SOLID exome sequencing. JS, LH, YS performed verification sequencing. JS, RP, SZ, SS performed RNAseq analysis. JC, MR, GG, MM performed exome capture and analyzed Illumina sequencing results. LA and DA facilitated transfer, sequencing, and analysis of the samples. JM and DC contributed samples. TB, DB and AB provided expertise in data analysis and edited the manuscript. DH, JA, SS and FB selected and characterized samples, provided disease-specific expertise, and edited the manuscript.

Competing financial interests: The authors declare no competing financial interests.

Abstract

Despite gains in survival, outcomes for patients with metastatic or recurrent rhabdomyosarcoma (RMS) remain dismal. In a collaboration between the National Cancer Institute, Children's Oncology Group, and Broad Institute, we performed whole-genome, whole-exome and transcriptome sequencing to characterize the landscape of somatic alterations in 147 tumor/normal pairs. Two genotypes are evident in RMS tumors; those characterized by the PAX3 or PAX7 fusion and those that lack these fusions but harbor mutations in key signaling pathways. The overall burden of somatic mutations in RMS is relatively low, especially in tumors that harbor a PAX3/7 gene fusion. In addition to previously reported mutations of *NRAS, KRAS, HRAS, FGFR4, PIK3CA, CTNNB1*, we found novel recurrent mutations in *FBXW7*, and *BCOR* providing potential new avenues for therapeutic intervention. Furthermore, alteration of the receptor tyrosine kinase/*RAS/PIK3CA* axis affects 93% of cases providing a framework for genomics directed therapies that might improve outcomes for RMS patients.

Keywords

Rhabdomyosarcoma; whole genome sequencing; whole exome sequencing; RNAseq

Introduction

Rhabdomyosarcoma (RMS) is a myogenic cancer that is the most common soft tissue sarcoma of childhood (1). With the development of multimodal chemotherapy regimens, relapse-free survival rates have improved to 70-80% in patients with localized disease, albeit with significant toxicity (2). Unfortunately, despite aggressive therapy, the 5 year survival rate for patients with metastatic disease remains only 30% (3). Currently, RMS tumors are classified by histology into two major subtypes: alveolar (ARMS) and embryonal (ERMS) which have distinct molecular and clinical profiles. ARMS carry a poor prognosis and tend to occur in adolescents. Genetically, ARMS is defined in the majority of cases by a characteristic fusion between the *PAX3* or *PAX7* and *FOXO1* genes (reviewed in (4)). The ERMS subtype typically affects younger children and portends a good prognosis when localized. Previous reports have identified a wide range of genetic aberrations in ERMS including loss of heterozygosity at 11p15.5 (5) as well as mutations in *TP53* (6), *NRAS, KRAS, HRAS* (7), *PIK3CA, CTNNB1* (8) and *FGFR4* (9).

Despite an increasing understanding of the molecular mechanisms underlying these tumors, few novel agents have made their way past early phase clinical trials and gains in survival have mainly been made through optimization of a cytotoxic chemotherapy regimen (10). Further characterization of the genetic events underlying this tumor type is critical to the development of more effective diagnostic, prognostic and therapeutic strategies. Here, we report a collaborative effort between the National Cancer Institute, the Children's Oncology Group, and the Broad Institute using a combination of whole-genome, whole-exome and whole-transcriptome sequencing along with high resolution SNP arrays to characterize the landscape of somatic alterations in 147 tumor/normal pairs. Our findings describe the landscape of genetic events that occur in RMS and provide a map for future studies of targeted molecular therapies for this tumor type.

Results

A set of 44 RMS tumors with matched normal leukocyte DNA was sequenced with whole-genome paired-end sequencing (WGS) and served as a discovery set. WGS generated an average of 294 gigabases (Gb) of sequence per sample to a mean depth of 105X. This depth of coverage allowed high quality calls covering 97% of the genome (Supplementary Table S1). To extend and validate our findings, we also performed whole-exome sequencing (WES) and high resolution SNP arrays on 103 additional tumors and their matched germlines (147 tumors in total with clinical data summarized in Supplementary Table S2). Eighty of the tumors were analyzed by whole-transcriptome sequencing (WTS) allowing us to evaluate the expression changes associated with the observed genomic alterations.

PAX gene rearrangements in RMS tumors

As expected, the defining genomic alteration seen across the entire cohort was recurrent t(2;13) or t(1;13), that resulted in a fusion of the N-terminus of *PAX3* or *PAX7* to the C-terminus of *FOXO1* (11, 12) (35 had *PAX3-FOXO1* and 15 *PAX7-FOXO1*) (Figure 1a and Figure 1b). The fusions discovered in WGS or WTS were confirmed by reverse transcription-polymerase chain (RT-PCR) reaction when adequate RNA was available. In addition to these classic fusions, three tumors that were histologically classified as ARMS but did not have the classical *PAX3/7-FOXO1* fusion by RT-PCR, were found to have alternative *PAX* fusions as detected by WGS or transcriptome sequencing. Cases RMS235 and RMS2031 harbored a *PAX3-NCOA1* fusion that resulted from an intra-chromosomal rearrangement previously described as having similar oncogenic properties as the *PAX3-FOXO1* (13). We also uncovered a novel *PAX* fusion in a region of massive rearrangement of chromosome 2q in RMS2046 (Figure 1c and 2a). This rearrangement resulted in a fusion of the N-terminus of PAX3 (first seven exons) and the C-terminus of INO80D, a subunit of the ATP dependent chromatin remodeling complex. RNA sequencing of RMS2046 showed in-frame expression of the novel fusion transcript (Figure 2b). Unsupervised clustering using the whole-transcriptome sequencing data showed clear separation between tumors that harbored the rearrangement of a PAX gene from those that did not. Of note, the tumors with the alternative PAX gene fusions clustered closely to the other ARMS that harbored the classical *PAX3/7-FOXO1* fusions expression profiles (Figure 2c). Aside from the three tumors that carried a novel PAX gene rearrangement, within this group, there were seven additional fusion-negative alveolar histology tumors that had no PAX gene alteration but a somatic mutation and expression profile more consistent with embryonal tumors (Figure 2d and Supplementary Figure S1A-D).

Recurrent chromosomal structural rearrangement in RMS

Beyond the rearrangement of the *PAX3/PAX7* genes, WGS identified 553 somatic structural variations (SV) affecting 419 genes in 44 RMS tumor genomes (Supplementary Table S3). High-resolution SNP arrays corroborated 90% of high confidence SV when a copy number change was present (see Methods). Forty-eight genes were recurrently affected by SV, including genes previously implicated in RMS pathology (*MIR17HG*, *CNR1*, *CDKN2A*) (14-16), tyrosine kinase signaling (*ERBB4*, *RPTOR*, *FRS2*, *CACNA1A*) and muscle development (*NRG1*, *FOXP2*) (Supplementary Table S4). Frequently (341/553, 61%),

junction events occurred in areas of complex rearrangement or tandem duplications most often associated with regions of high copy number amplification. Ten percent of the junctions were predicted to result in deletions. When a junction event occurred at the DNA level between two genes, a fusion transcript was produced at the RNA level 19% (55/296) of the time. Among these events, fusion of the PAX genes account for one third of the fusion transcripts and no additional recurrent fusions were detected (Supplementary Table S3).

Presence or absence of a PAX gene fusion defines two distinct tumor genotypes

Of note, when PAX-gene fusion positive (PFP) tumors were compared to PAX-gene fusion negative (PFN) ones, we found a significantly increased mutation burden in the PFN population (Figure 3a). On average, PFN tumors had significantly more verified somatic non-synonymous mutations per tumor than PFP tumors (17.8 and 6.4 respectively) ($P=2\times10^{-4}$). In contrast to the PFP samples, the PFN samples had an overall increase in aneuploidy ($P=1\times10^{-5}$) (Figure 3b). One remarkable PFP tumor (RMS224) from a 3-month-old, had no protein coding somatic alterations with the exception of the *PAX3-FOXO1* fusion and copy neutral loss of heterozygosity (LOH) on chromosome 11p (Figure 1b). Interestingly, both of the PFP and PFN genotypes appear to have a distinct relationship between mutational frequency and age, with an increasing number of somatic mutations with older age of diagnosis and a steeper slope of curve in PFN tumors (Figure 3c).

Genes recurrently affected by mutation in RMS

In total, we identified 542 somatic mutations (including missense, nonsense, splice site, and small insertions/deletions) altering 495 genes (40 recurrent) in the discovery set of 44 tumors (Supplementary Table S5); 58% of these alterations were predicted to be deleterious by SIFT analysis (17). These genes were selected for further verification and validation across the entire cohort and ranked using recurrence, background mutation rate, gene size and nonsynonomous:synonomous ratio (Table 1). The list contained genes previously reported as altered in RMS including, *HRAS*, *KRAS*, *NRAS* (7), *FGFR4* (9), *PIK3CA*, and *NF1* (18), as well as genes not previously implicated in RMS such as *FBXW7* and *BCOR* (Figure 4).

RAS/PIK3CA/Tyrosine Kinase Mutations predominately affect PAX fusion-negative tumors

Mutations affecting the receptor tyrosine kinase/RAS/PIK3CA pathway were the most common mutations observed in the study. Alterations in RAS genes, *NRAS* (Representative genome Figure 1d; PFN frequency, 11.7%), *KRAS* (PFN 6.4%), *HRAS* (PFN 4.3%) affected the oncogenic codons 12, 13 or 61, and were predominantly found in the ERMS subtype as previously described (19), however one tumor (RMS2051) with “fusion-negative” ARMS histology carried a *NRAS* mutation. No RAS mutations were found in PFP tumors. Mutations in immediate effectors of RAS were also found, including alterations in the tumor suppressor *NF1* (PFN 5.3% mutated, 17q11.2 LOH 9% Supplementary Figure S2) and one tumor with an oncogenic mutation of *BRAF* at codon V600E (RMSS013). *PIK3CA* mutations (PFN 7.4%) occurred at the known oncogenic codons Q546 or H1047 affecting the helical and the kinase domain, respectively. Interestingly, two samples (RMS2028, RMS217) had concurrent mutation of *PIK3CA* and a RAS family gene. Despite a

predilection for ERMS tumors (6/7), one fusion positive ARMS tumor (RMS244) also harbored a mutation in *PIK3CA*. Direct effectors of PI3K were also found to be altered including a predicted damaging mutation in *PIK3CD* (RMS2107) and homozygous deletion of *PTEN* (RMS2117). Across the whole population, several tyrosine kinase genes were found to be recurrently mutated including *FGFR4* (PFP 0%, PFN 9.6%), *PDGFRA* (1.4%), and *ERBB2* (1.4%).

Mutations of Cell Cycle Genes and other key pathways

Genes that control the cell cycle were also frequently mutated in the study population. *FBXW7*, an E3 ubiquitin ligase, was mutated in 7.4% of PFN tumors. All mutations in this gene occurred in the PFN subtype at conserved arginine residues (R387P, R441G, R367P) within the WD40 repeat regions involved in substrate recognition. Mutations in the WNT signaling molecule *CTNNB1* are known drivers in colorectal cancer and medulloblastoma and have recently been described in RMS (8). In this study, we found three tumors (PFN 3%) with alterations at the known oncogenic codons S33 (n=1) and T41 (n=2) one of which occurred in a fusion negative ARMS tumor. Somatic mutation of *TP53* occurred exclusively in PFN tumors (PFN 5.3%; Representative genome Figure 1e) and 12% of all tumors have LOH of 17p13.1 which includes *TP53* (Supplementary Figure S2). One patient (RMS 212) was found to have a germline pathogenic mutation in *TP53* at R248 (Supplementary Figure S3A-C). Other mitotic cell cycle checkpoint genes were mutated at low frequencies, including *BUB1B* (1.4%), *FOXM1* (1.4%), *CCND1* (1%), *CCND2* (1%). A notable finding was the recurrent alteration of *BCOR*, located on chromosome Xp11.4, in 7% of all RMS cases. Among these alterations of *BCOR*, 9 were found in PFN tumors (7 mutations, 2 focal homozygous deletions) and 1 small indel was found in a PFP tumor (Table 1 and Supplementary Figure S4).

RNA sequencing highlights expression of candidate oncogenes

To further enrich the analysis for potential oncogenes and targetable mutations, we performed mutational analysis of the RNAseq data of 80 tumors (29 PFP and 51 PFN) to determine which of the somatic mutations found at the DNA level were also expressed in the transcriptome. Fifty-eight percent of the verified somatic mutations discovered at the DNA level had evidence of RNA expression (Supplementary Figure S5A and S5B). Each tumor harbored a median of five expressed somatic mutations (Range 0-26; PFN median, 9 mutations; PFP median, 2.5 mutations; Supplementary Table S6). Thirty-three genes were found to recurrently harbor expressed mutations including *PTPN11* (0 PFP *vs.* 2 PFN) (20), the DNA repair gene *ATM* (2.5%; 1 PFP *vs.* 1 PFN), a BRCA1-interacting protein *ZNF350* (2.5%; 1 PFP *vs.* 1 PFN), and a *MYCN*-interacting protein *TRPC4AP* (2.5%; 0 PFP *vs.* 2 PFN). In addition, we discovered expressed singleton mutations in *FOXO1* and *ARID1A* (Figure 5a) not previously observed in rhabdomyosarcoma. By Gene Ontology, the expressed mutations were markedly enriched for genes involved in cell cycle (*P*=2e-6), protein phosphorylation (*P*=6.9e-5) DNA damage (P=1.3e-4), muscle cell differentiation (*P*=3.3e-4), regulation of MAP kinase activity (P=3.3e-4), chromatin modification (P=9e-4) and induction of apoptosis (P=2.8e-3) (Supplementary Table S7). Many of the tumors appear to accumulate multiple genetic hits within these pathways (Figure 5b).

Copy Number Alterations

To evaluate somatic copy number alterations (CNAs) important in RMS, high resolution (2.5 or 5 million) SNP array analyses were performed on all tumors and recurrent focal amplifications and deletions were analyzed by frequency in the study population (Supplementary Figure S6). LOH of 11p15.5 was found in 50% (16 PFP *vs.* 59 PFN) of the surveyed tumors. The minimum common region of overlap encompassed a region of 11p15.5 (Supplementary Figure S7) that includes the paternally imprinted gene *IGF2*. Further evidence of insulin receptor signaling alterations in RMS were observed with focal amplification of *IGF1R* in 2.7% (1PFP *vs.* 3PFN) of cases (Supplementary Figure S8a) (21) and one somatic indel in the 3′ untranslated region of *IGF2* (RMS2037; Supplementary Table S5). Consistent with previous reports, 9.7% of the tumors displayed amplification of chromosomal region 12q13-q14, which has been shown to be associated with worse overall survival in RMS independent of gene fusion status (22). The 12q13-q14 amplicon was found predominantly in PFP tumors (12 PFP *vs.* 1 PFN). The minimum amplicon size (Supplementary Figure S8b) included 25 genes, including the cyclin-dependent kinase *CDK4*. Recurrent focal amplification of 12q15 (9%; Supplementary Figure S8c) which encompassed the genes *FRS2* and *MDM2* occurred predominantly in PFN tumors (9 PFN *vs.* 1 PFP). Amplification of 2p24 involving *MYCN* (5%) occurred predominantly in PFP tumors (8 PFP *vs.* 1 PFN) (Supplementary Figure S8d); amplification of the *PAX7-FOXO1* fusion gene occurred in 12/15 *PAX7-FOXO1* tumors (Supplementary Figure S8e) and amplification of 13q31-32 including the *MIR17HG* locus occurred exclusively in PFP tumors (4.5%) (Supplementary Figure S8f). Homozygous deletion of the tumor suppressor *CDKN2A* was found in 3% of samples and LOH at this locus (9p21.3) occurred in 9% (1 PFP *vs.* 13 PFN) of the study population. This allelic loss rate was lower than the previously reported frequency of 25% (15). As previously described (23), recurrent gain of chromosome 8 was seen in 46% of the PFN population. Other chromosome level events included recurrent gains of chromosomes 2, 7, 11 and 13 and the recurrent loss of chromosome 1p, 9 and 16 (Supplementary Figure S9).

Pathway analysis integrating mutations, copy number changes and structural variations implicates alteration of FGFR signaling

To identify dysregulated pathways relevant to RMS pathology, analysis incorporating structural variations, copy number changes and somatic mutations found in the WGS discovery set was performed. Using the 2,119 genes found to be somatically altered in the discovery cohort (Supplementary Table S8), Reactome (24) over-representation analysis indicated that FGFR signaling was the most significantly altered pathway ($P = 4.6\times10^{-5}$) with 29/112 candidate genes represented. Remarkably mutations in this pathway (Figure 6a) were found in 88% of PFN samples (22/25 tumors) which were analyzed by WGS. When examined separately, the genes altered in PFP tumors (435 out of 2119) had no canonical pathways significantly enriched.

PAX3-FOXO1 model system reveals alteration of a common genetic axis in fusion-positive and fusion-negative tumors

Of note, several genes found altered in PFN tumors including *MYOD1*, *MET*, *CNR1* and *FGFR4*, are known downstream targets of PAX3 and PAX3-FOXO1 (P=1.54×10^{-3}) (25), leading us to hypothesize that mutations in PFN tumors may be enriched for genes regulated downstream of the PAX fusion proteins. To experimentally test this hypothesis, we constructed a human fibroblast cell line stably expressing PAX3-FOXO1 (cell line 7250_PF) and used expression arrays to compare it with the isogenic control (26). This analysis identified 444 genes that had greater than four-fold change when PAX3-FOXO1 is expressed. Top up-regulated genes included multiple genes that were found to be mutated in PFN tumors such as *FGFR4*, *CCND2*, and *IGF2* (Supplementary Table S9). As confirmation of the model system, these differentially expressed genes were also over-represented in the 76 genes recently reported as PAX3-FOXO1 targets (P=1.7×10^{-3}) using ChIPseq (27). Remarkably, the 2,119 somatically altered genes identified in our whole genome sequencing samples were significantly enriched in the differentially expressed genes modulated by the PAX3-FOXO1 in the fibroblast cell line experiment using gene set enrichment analysis (GSEA) (P=3×10^{-3}) (Figure 6b). The observed enrichment was more prominent for the mutated genes from PFN tumors (P=7×10^{-3}) than those from PFP samples (P=0.08). To further validate this hypothesis, we repeated GSEA analyses using published data derived from a transgenic mouse model expressing the PAX3-FOXO1 fusion gene in the developing forelimb or somite (28). Consistent with our *in vitro* results, mutated genes in the PFN tumors were significantly enriched in both the forelimb and somite datasets (P=0.001 and 0.01) (Figure 6c and 6d), whereas there was no enrichment for those in the PFP tumors (P=0.159 and 0.543). A set of 116 common genes, including *FGFR4*, was found in the leading edge of all three PAX-fusion model systems in the GSEA analyses (Supplementary Table S10).

Discussion

To our knowledge this study represents the most comprehensive characterization yet reported of the genomic alterations that underlie RMS. We find that sub-categorization by the presence or absence of a PAX gene fusion more accurately captures the true genomic landscape and biology of RMS than the traditional ARMS/ERMS histological distinction. This finding is consistent with the clinical observation that presence or absence of a *PAX3/7-FOXO1* gene fusion is a crucial prognostic indicator in this disease (29, 30) and that fusion-negative ARMS seems to mimic the clinical course of ERMS in the majority of patients (23). Despite this, our findings indicate that there is a subpopulation of fusion-negative ARMS that harbor a rearrangement of the *PAX3* gene with a cryptic partner; a finding that may have important clinical ramifications for the proper therapeutic stratification of patients.

Overall, the low somatic mutation rate that we observed is consistent with large sequencing efforts of other pediatric solid tumors and presents an enormous correlative and clinical challenge (31-34). In RMS, this was particularly evident in tumors that harbored a translocation oncogene (0.1 protein coding changes per megabase). This finding underscores the importance of the *PAX* gene fusion as the dominant driver in this subtype, which through

its transcriptional reprogramming alters a host of downstream targets. However, it is important to note that multiple genetic model systems have shown that PAX3-FOXO1 by itself cannot cause RMS and that a coexisting genetic lesion is necessary (35, 36). Experimentally validated cooperating lesions in a mouse model of ARMS include *TP53* and *Ink4a/ARF* loss (37). Our data demonstrate that most commonly the cooperating event is due to genetic amplification (such as *MYCN*, *CDK4*, *MIR-17-92*) or deletion (*CDKN2A*, LOH of Chr11p15.5), and only in the rare case can an additional candidate somatic driver mutation be nominated. In contrast, fusion-negative tumors appear to have accumulated a higher degree of aneuploidy and mutational burden at the time of clinical presentation.

Despite the relatively low mutation rate, RMS tumors do harbor a significant array of alterations including chromosomal rearrangement, amplification, deletion and mutation of recurrent drivers and novel candidate therapeutic targets. Many of the genetic alterations identified in this study including *FGFR4, IGF1R, PDGFRA, ERBB2/4, MET, MDM2, CDK4* and *PIK3CA* are targeted by approved or late stage therapeutics that could immediately inform clinical trials in RMS (Figure 7). In this study, we find that the RAS pathway (including *FGFR4*, RAS, *NF1*, *PIK3CA*) is mutationally activated in at least 45% of PFN tumors. While directly targeting constitutively active RAS remains challenging, the recent success of the MEK1/2 inhibitor, Trametinib, in melanomas with mutated *NRAS* demonstrates the utility of inhibiting the effector pathways altered by the mutation (38). Early preclinical evidence has found efficacy of this method in RMS (39) and further efforts to precisely dissect the RAS effector pathways that are critical in RMS are currently underway.

A novel finding in this study is the discovery of recurrent mutations in *BCOR* affecting 7.4% of PFN tumors. BCOR is a transcription repressor that has been shown to interact with both class I and class II histone deacetylases (40) and somatic mutations in *BCOR* have been described in other pediatric tumors including AML (32), retinoblastoma (33) and medulloblastoma (34). Our discovery of its recurrent alteration in RMS reinforces this chromatin modifier (41) as a potential therapeutic target. Further functional validation of the discovered mutations of *BCOR, FBXW7, ARID1A, ZNF350, TRPC4AP* and others may provide targets for novel treatments in patients with RMS. Incorporation of the discovered genes into prospective, well annotated clinical trials will be crucial in extending these findings utility as therapeutic biomarkers.

Despite the challenges of low frequency of recurrence, the genetic study of pediatric cancer provides remarkable insight into the likely drivers of tumorigenesis by reducing the background of passenger mutations that naturally occur during aging. The observation that PFP and PFN genotypes appear to have a distinct relationship between mutational frequency and age, with a steeper slope in the PFN tumors, may have interesting implications. This finding suggests that PFN tumors require the accumulation of mutations prior to presentation, whereas malignant transformation of PFP tumors requires few somatic alterations beyond the occurrence of the fusion gene. This observation may also be due to differences in the respective tumor type's cell of origin, proliferation and apoptotic rate or an underlying DNA repair deficit. Our observation that 58% of the verified somatic mutations discovered at the DNA level had evidence of RNA expression is a higher proportion than the

36% rate observed in adult cancers such as breast (42) or lymphoma (43) and may reflect an enrichment of driver mutations or the presence of fewer accumulated passenger mutations in these pediatric patients. In many cases, the expression of a mutated gene appears to be relatively increased and favor the variant allele. This finding, at least in theory, provides tractable genetic targets against which therapies could be developed.

Finally, our integrative analysis demonstrates that despite remarkable genetic and molecular heterogeneity, RMS tumors appear to hijack a common receptor tyrosine kinase/RAS/PIK3CA genetic axis. This occurs through two alternative mechanisms – either by rearrangement of a PAX gene or accumulation of mutations in genes that are downstream targets of the PAX fusion protein. Evidence for alteration of this common genetic axis can be found in 93% (41/44) of the tumors surveyed by WGS and appears to hinge around the fibroblast and insulin receptor pathways. These observations are consistent with previous proteomic studies of RMS (44, 45) and warrant continued biologic investigation and pharmacologic targeting of this axis as crucial to expanding the available therapeutic options. In conclusion, we report here the most comprehensive analysis of the genomic landscape of rhabdomyosarcoma. Our discoveries provide a rational framework for new avenues of translational research including molecular sub-classification and developing novel therapeutic strategies for children suffering with rhabdomyosarcoma.

Methods

Sample Selection

All patient sample collection was approved by the institutional review board of the participating facility. Samples were assembled from collections at the Pediatric Oncology Branch of the National Cancer Institute, Children's Oncology Group, the Tumour Bank at The Children's Hospital at Westmead (New South Wales, Australia), and the Department of Oncology St. Joan de Deu De Barcelona (Barcelona, Spain). All tumors were collected at initial diagnosis and prior to any therapy with the exception of samples NCI0040 and NCI0080 which were collected at relapse. Samples were de-identified and histologic diagnosis and clinical information were compiled. The selected tumors were >70% tumor:normal tissue on histology review when available. Quality control genotyping for the whole genome samples was performed to ensure the match of tumor normal pairs.

Nucleic Acid Extraction and whole genome amplification

DNA was isolated from 10-25mg of tumor or 1 ml of whole blood using QIAamp DNA Mini Kits (Qiagen, Valencia, CA) or Agencourt Genfind v2 kits (Bechman Coulter, CA) respectively according to the manufactures protocol. For whole genome sequencing, approximately 6ug (Range 5-10ug) of native genomic DNA was sequenced according to the Complete Genomics (CG) method (46). Whole genome amplified genomic DNA using high-fidelity Phi29 polymerase (Qiagen REPLI-g) was used for the whole exome validation cohort according to the manufactures protocol (Qiagen, Valencia, CA). Quantification of DNA was performed using the Quanti-iT DNA assay (Life Technology, Grand Island, NY). Each DNA sample was examined by electrophoresis on a 1% agarose gel to ensure high

quality. RNA extraction was accomplished with Qiagen RNeasy micro kits according to the manufactures protocol (Qiagen, Valencia, CA).

Calculation of Background mutation rate

The background mutation rate was calculated using the method described by Zhang (47). Briefly, the background mutation rate is the silent mutation rate in coding region adjusted by silent-to-non-silent ratio (estimated to be 0.350 by the TCGA Consortium) across the coding regions.

Small Variant Discovery for Whole Genome Sequencing

Small variants, including single-nucleotide variants and indels were called using cgatools (http://cgatools.sourceforge.net/docs/1.6.0/) in build hg19. Somatic variants were determined first by comparison of the tumor with matched leukocyte normal DNA. To remove artifacts specific to the sequencing platform, we eliminated any somatic variants also found in normal subjects other than RMS patients (50 in-house normal samples and 69 Complete Genomics samples (http://www.completegenomics.com/public-data/69-Genomes/). The Somatic Score (48-50) http://info.completegenomics.com/rs/completegenomics/images/Cancer_Application_Note) is based on a Bayesian model and takes account of read depth, base call quality, mapping/alignment probabilities, and measured priors on sequencing error rate for both the germline variants and the tumor variants. Using an independent platform (SOLiD exome sequencing) for verification of somatic variants from the whole genome sequencing, we found an optimal balance between sensitivity and specificity by selecting variants with somatic score≥0 (Supplementary Figure S10A and S10B). Finally, small variants within regions that have significant similarity to other regions in the genome, taken from the “Self Chain” track of UCSC genome browser (http://genome.ucsc.edu/cgi-bin/hgTracks?org=human), were removed as they are likely due to mapping errors.

The somatic variants were then annotated using ANNOVAR(51) which details the synonomous/non-synonomous nature of the alteration, the corresponding amino acid alteration as well as presence or absence of the alteration in dbSNP135 and 1000 Genome Project. SIFT (http://sift.jcvi.org/www/SIFT_chr_coords_submit.html) and Polyphen (http://genetics.bwh.harvard.edu/pph2/bgi.shtml) scores were used to determine the potential impact of a SNP variant. Oncotator (http://www.broadinstitute.org/oncotator/) was used to add cancer specific annotations from COSMIC and TCGA.

Verification of WGS predicted somatic mutations

Verification of small variants with somatic score no less than 0 predicted by WGS was accomplished by comparing overlapping exome sequencing which was done on 30/44 tumor samples. Each somatic position was examined for the identical change as well as coverage in the exome sequencing data. Verification of the mutation was called when greater than 3 exome reads supported the WGS read. Additional site verification was performed with barcoded DNA libraries made from the 44 WGS tumors using a designed Custom AmpliSeq Cancer Panel and AmpliSeq Library Kit 2.0 with sequencing on the Ion Torrent Personal Genome Machine. Multiplex PCR library preparation, ePCR template preparation, and

semiconductor sequencing were performed according to the manufacturer's protocol. Sequencing generated 868Mb high quality bases with average amplicon coverage of 206X. Using this method, sensitivity is calculated at 84% (assuming the small variants reported by CG whole genome sequencing includes all the true positive variants) and a specificity of 93%. Additional verification of mutations reported in Figure 1, Table 1 was accomplished by PCR amplification of genomic DNA using a uniform annealing temperature of 65°C followed by standard Sanger sequencing and analysis using Sequencher 4.10 software (Gene Codes Corporation).

Copy Number Discovery from Whole genome sequencing

CG copy number segments (based on 2-kb window) profile was used to call amplifications (>=5 copies) and homozygous deletions. Copy number alterations were divided into two groups: 1) focal amplification or deletions less than one arm in length and; 2) whole-arm or whole-chromosome events.

Junction Discovery

Based on the high-confidence junction reports for the tumor sample and the paired germline sample, we called the somatic junctions as those only present in tumor samples. Somatic junctions that are present in other normal samples (50 in-house germline samples and 69 CG baseline germline samples) were removed to reduce systematic artifact. Comparison of the predicted somatic junctions with the corresponding SNP array copy number data shows 90% of the junctions had changes in copy number state or allelic ratio at the predicted break point.

Circos Plots

Circos plots were generated for each sample using the circos plotting software provided by Complete Genomics (http://www.completegenomics.com/analysis-tools/cgatools), with in-house customized modifications.

RT-PCR of PAX-FOXO gene fusion

We determined the *PAX3-FOXO1* or *PAX7-FOXO1* fusion status using reverse transcription polymerase chain reaction (RT-PCR) of tumor RNA using specific oligonucleotide primers according to the published method (52).

RNA sequencing

PolyA selected RNA libraries were prepared for RNA sequencing on Illumina HiSeq2000 using TruSeq v3 chemistry according to the manufactures protocol (Illumina®, San Diego, CA). 100 bases long paired-end reads were assessed for quality and reads were mapped using CASAVA (Illumina®, San Diego, CA). The generated fastq files were used as input for TopHat2(53). Using Samtools (http://samtools.sourceforge.net/) the produced BAM files were compared with the sites found somatically mutated in DNA and total coverage and variant allele frequency were calculated. Expressed fusion transcripts were detected by tophat-fusion 0.1.0(54) and deFuse 0.4.3(55) with hg19 human genome assembly.

RNAseq expression analysis and unsupervised clustering

Cufflinks (http://cufflinks.cbcb.umd.edu/) (56) was used to assemble and estimate the relative abundances of transcripts mapped with TopHat2 at the gene and transcript level (FPKM). FPKM values were log2 transformed. Samples were clustered based on Wards algorithm based on Euclidean distance.

SOLID Exome Sequencing and data analysis

We constructed sequencing libraries and performed target enrichment using Agilent SureSelect Human All Exon kits designed to target 37.8 Mb regions of all human exons according to the manufacturer's instruction (Agilent, Santa Clara, CA). The PCR-amplified libraries were sequenced on SOLiD™ 4 systems using the 50x35bp paired-end sequencing protocol (Applied Biosystems, Foster City, CA). Sequencing was used to evaluate 120 tumor/normal pairs for coding sequence alterations. An average of 3.3 Gb of non-redundant sequence was mapped on-target per sample to hg19 using BFAST version 0.7.0a(57). Duplicates were removed using Picard (http://picard.sourceforge.net/command-line-overview.shtml), normal/tumor bam files were used as input for GATK version 2.1-11 (http://www.broadinstitute.org/gatk/)(58-60). Local realignment and base quality recalibration were performed using default parameters. Single nucleotide polymorphisms and indels were called using GATK UnifiedGenotyper (http://www.broadinstitute.org/gatk/gatkdocs/org_broadinstitute_sting_gatk_walkers_genotyper_UnifiedGenotyper.html). Variants which passed quality score greater than 50, coverage in tumor and normal greater than 10, VAF (Variant Allele Frequency) in the tumor greater than 15% and VAF in normal of 0% were further annotated with ANNOVAR, SIFT, PPH2 and COSMIC.

Illumina Exome Method

100ng of tumor and tumor DNA underwent shearing, end repair, phosphorylation, and ligation to barcoded sequencing adaptors. The ligated DNA was size-selected for fragments between 200-350 bp. This prepared DNA underwent exome capture using SureSelect v2 Exome bait (Agilent, Santa Clara, CA). Captured DNA was multiplexed and sequenced on Illumina HiSeq flowcells. Exome analysis was performed using Broad Institute pipelines(61, 62). MuTect and MutSig algorithms were used to call somatic mutations and determine statistical significance, respectively.

Comparison of PFP mutation rate with PFN mutation rate

We hypothesize that there are less somatic nonsynonymous mutations in PFP than those in PFN RMS (Figure 3). To test this hypothesis, we compared the number of verified somatic nonsynonymous mutations in PFP with that in PFN RMS patients. Random Permutation Test (RPT) was performed by permuting the group label of the patients, to avoid any assumption on the unknown distribution of the number of mutations as well as the bias brought by the assumption. The summary statistic is defined as the between-group variability divided by the within-group variability, to measure the difference between the two patient groups while considering the variability within each patient group.

Quantifying the relation between mutational frequency and age

We observed that the PFP and PFN RMS have distinct but consistent relationships between mutational frequency and age (Figure 3c). Therefore, we applied linear regression to model the relationship between the number of WGS SNVs (dependent variable) and the corresponding age at diagnose (explanatory variable) for PFP (16 patients with age information) and PFN patients (24 patients with age information), respectively. The goodness of fitting of each linear regression model is measured via the *t*-test on Pearson's correlation coefficient between the mutation number reported by WGS and that predicted from the linear regression model (one-tailed test due to the pre-known positive correlation).

Method of determining statistically significant genes

The 621 genes harboring small variants with somatic score no less than 0 in the WGS study were applied to the exome validation samples (103 total tumor/normal pairs – 90 SOLiD/13 Illumina). Included in the recurrence calculation is any small variant which was called using the aforementioned SOLiD or Illumina GATK analysis pipeline. Total recurrence was then calculated as the number of mutations in a gene in the WGS data in addition to the WES data. Ranking of this gene list was done by employing the binomial method reported by Wei et al (63) to calculate the significance of a gene mutation. This method considers the recurrence of mutation in the observed gene, the length of the gene coding region, and the background mutation rate as well as the synonomous:nonsynonomous ratio. Significant genes were selected based on false discovery rate ≤ 0.05(64).

SNP Array

Illumina Omni® 2.5M (97 paired plus 30 unmatched tumors) or 5M (10 paired samples) were performed according to the standard procedure from the manufacturer (Illumina, San Diego, CA) at the National Cancer, Cancer Genomics Research Laboratory. When available matched normal:tumor paired arrays were analyzed. For Copy number analysis the row data was processed and normalized in Illumina GenomeStudio (http://www.illumina.com/Documents/products/technotes/technote_infinium_genotyping_data_analysis.pdf). Final reports are exported and imported into Nexus BioDiscovery (http://www.biodiscovery.com/downloads/pdfs/SimplifyingDataInterpretationWithNexusCopyNumber.pdf) software in paired Mode. In Nexus, the data is corrected for GC content and segmented using SNP-FASST2. Frequency across the whole population and according to the fusion status was analyzed using the STAC algorithm (65). High copy amplicons were plotted using the row probe level Log Relative Ratios.

Pathway Analysis

Reactome Pathway analysis (http://www.reactome.org/) was performed as previously described (24). Over-representation analysis of the somatically altered genes was performed. Fisher's exact test was used to calculate a *P*-value determining the probability that the association between the genes in the dataset and the observed pathway is explained by chance alone. Further gene interaction analysis was performed through the use of IPA (Ingenuity® Systems, www.ingenuity.com). Functional Analysis was performed, in which the biological functions most significant to the data set were extracted. A right-tailed

Fisher's exact test was used to calculate a *P*-value determining the probability that each biological function assigned to that data set is due to chance alone.

Association between the genes altered in PFN RMS and the PAX-FOXO1 binding genes

We observed that many genes frequently altered in FN RMS tumors are PAX-FOXO1 binding genes. To test whether the overlap between these two gene groups is by chance or not, we compared the genes with somatic mutations, copy number amplification, copy number homozygous-deletion, or structural variants (1957 genes reported by WGS) in 25 FN RMS tumors, to the genes recently reported as significantly altered in a chromatin precipitation identification of PAX3-FOXO1 binding sites (76 genes reported in Cao (28)). Fisher's Exact Test was performed to assess the association between these two gene groups, and $P=4.5\times10^{-3}$, rejecting the null hypothesis (the significance threshold is set as 0.05). This result indicates that the genes altered in PFN RMS patients are significantly associated with the PAX-FOXO1 binding genes.

7250_PF Cell line

The stably transfected cell line expressing PAX3-FOXO1 was constructed as previously described from the parent cell line CRL7250 (ATCC) (26, 66). The cell lines were validated by the National Cancer Institute Division of Cancer Epidemiology and Genetics using short tandem repeat DNA fingerprinting. The expected expression of the PAX3-FOXO1 fusion oncogene was evaluated with RT-PCR (Supplementary Figure S11). All cells were grown in 85% DMEM with 300ug/ml of G418 and 10% fetal bovine serum under identical conditions and harvested at 80-85% confluency. Total cellular RNA was purified using Qiagen AllPrep mini-kit according to the manufacturer's protocol. Microarray expression analysis was performed using the human Genome U133 Plus 2.0 array (Affymetrix) and the data was normalized together using RMA (Affymetrix). This generates expression values for each probe in log2 space. We then calculated the absolute value of the relative fold change score (Supplementary Table S8) as the following: absolute value [(7250_PF RMA signal)-(7250_PF Nil RMA signal)]. These values were further analyzed using the GSEA algorithm (see method below). (http://www.broad.mit.edu/gsea/).

Mouse Model of PAX3-FOXO1 expression in somite or forelimb

Expression data for *PAX3-FOXO1* expression in mouse somite and forelimb were derived from the experiments previously described (28).

GSEA analysis

To test whether somatically altered genes in fusion negative RMS overlap with genes downstream of fusion positive RMS we performed gene set enrichment analysis (GSEA) (67) on three PAX gene fusion model systems. Gene expression from each model system was ranked according to absolute fold change expression over the corresponding control. GSEA analysis (http://www.broadinstitute.org/gsea/index.jsp) was performed using default parameter settings. *P*-values are calculated based on Kolmogorov-Smirnov statistic with random permutation test.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

The authors thank the Children's Oncology Group Soft Tissue Sarcoma Committee and the BioPathology Center, especially Dr. Julie Gastier-Foster, for their careful collection and annotation of clinical samples. We thank the staff of the NCI Cancer Genomics Research Laboratory and core RNA sequencing facility for their contribution to the study. This study utilized the high-performance computational capabilities of the Biowulf Linux cluster at the National Institutes of Health (http://biowulf.nih.gov), and was supported by the Intramural Research Program of the National Institutes of Health, National Cancer Institute, Center for Cancer Research. The content of this publication does not necessarily reflect the views or policies of the Department of Health and Human Services, nor does mention of trade names, commercial products, or organizations imply endorsement by the U.S. government. Additional funding was provided to MM by generous gifts from M. B. Zuckerman and Team Dragonfly of the Pan-Mass Challenge. JC is supported by an American Cancer Society AstraZeneca Postdoctoral Fellowship. DC is supported by a joint appointment with the Discipline of Pediatrics and Child Health, Sydney Medical School, University of Sydney and receives funding support through The Kids Cancer Project.

Financial Support – JFS, LC,JSW, RP, JW, YKS, CT, LH, HL, SZ, DB, AB, SS, TB, FB, JK, are supported by the Intramural Research Program of the National Institutes of Health, National Cancer Institute, Center for Cancer Research. MM is supported by M. B. Zuckerman and Team Dragonfly of the Pan-Mass Challenge. JC is supported by an American Cancer Society AstraZeneca Postdoctoral Fellowship. DC is supported by a joint appointment with the Discipline of Pediatrics and Child Health, Sydney Medical School, University of Sydney and receives funding support through The Kids Cancer Project.

References

1. Ognjanovic S, Linabery AM, Charbonneau B, Ross JA. Trends in childhood rhabdomyosarcoma incidence and survival in the United States, 1975-2005. Cancer. 2009; 115(18):4218–26. Epub 2009/06/19. [PubMed: 19536876]
2. Malempati S, Hawkins DS. Rhabdomyosarcoma: Review of the Children's Oncology Group (COG) soft-tissue Sarcoma committee experience and rationale for current COG studies. Pediatr Blood Cancer. 2012; 59(1):5–10. [PubMed: 22378628]
3. Breneman JC, Lyden E, Pappo AS, Link MP, Anderson JR, Parham DM, et al. Prognostic factors and clinical outcomes in children and adolescents with metastatic rhabdomyosarcoma--a report from the Intergroup Rhabdomyosarcoma Study IV. Journal of clinical oncology : official journal of the American Society of Clinical Oncology. 2003; 21(1):78–84. Epub 2002/12/31. [PubMed: 12506174]
4. Linardic CM. PAX3-FOXO1 fusion gene in rhabdomyosarcoma. Cancer Lett. 2008; 270(1):10–8. Epub 2008/05/07. [PubMed: 18457914]
5. Scrable H, Cavenee W, Ghavimi F, Lovell M, Morgan K, Sapienza C. A model for embryonal rhabdomyosarcoma tumorigenesis that involves genome imprinting. Proceedings of the National Academy of Sciences of the United States of America. 1989; 86(19):7480–4. Epub 1989/10/01. [PubMed: 2798419]
6. Taylor AC, Shu L, Danks MK, Poquette CA, Shetty S, Thayer MJ, et al. P53 mutation and MDM2 amplification frequency in pediatric rhabdomyosarcoma tumors and cell lines. Medical and pediatric oncology. 2000; 35(2):96–103. Epub 2000/08/05. [PubMed: 10918230]
7. Stratton MR, Fisher C, Gusterson BA, Cooper CS. Detection of point mutations in N-ras and K-ras genes of human embryonal rhabdomyosarcomas using oligonucleotide probes and the polymerase chain reaction. Cancer Res. 1989; 49(22):6324–7. [PubMed: 2680062]
8. Shukla N, Ameur N, Yilmaz I, Nafa K, Lau CY, Marchetti A, et al. Oncogene mutation profiling of pediatric solid tumors reveals significant subsets of embryonal rhabdomyosarcoma and neuroblastoma with mutated genes in growth signaling pathways. Clinical cancer research : an official journal of the American Association for Cancer Research. 2012; 18(3):748–57. Epub 2011/12/07. [PubMed: 22142829]

9. Taylor, JGt; Cheuk, AT.; Tsang, PS.; Chung, JY.; Song, YK.; Desai, K., et al. Identification of FGFR4-activating mutations in human rhabdomyosarcomas that promote metastasis in xenotransplanted models. J Clin Invest. 2009; 119(11):3395–407. Epub 2009/10/08. [PubMed: 19809159]
10. Maurer HM, Beltangady M, Gehan EA, Crist W, Hammond D, Hays DM, et al. The Intergroup Rhabdomyosarcoma Study-I. A final report. Cancer. 1988; 61(2):209–20. Epub 1988/01/15. [PubMed: 3275486]
11. Barr FG, Galili N, Holick J, Biegel JA, Rovera G, Emanuel BS. Rearrangement of the PAX3 paired box gene in the paediatric solid tumour alveolar rhabdomyosarcoma. Nat Genet. 1993; 3(2): 113–7. [PubMed: 8098985]
12. Davis RJ, D'Cruz CM, Lovell MA, Biegel JA, Barr FG. Fusion of PAX7 to FKHR by the variant t(1;13)(p36;q14) translocation in alveolar rhabdomyosarcoma. Cancer Res. 1994; 54(11):2869–72. [PubMed: 8187070]
13. Wachtel M, Dettling M, Koscielniak E, Stegmaier S, Treuner J, Simon-Klingenstein K, et al. Gene expression signatures identify rhabdomyosarcoma subtypes and detect a novel t(2;2)(q35;p23) translocation fusing PAX3 to NCOA1. Cancer Res. 2004; 64(16):5539–45. Epub 2004/08/18. [PubMed: 15313887]
14. Marshall AD, Lagutina I, Grosveld GC. PAX3-FOXO1 Induces Cannabinoid Receptor 1 to Enhance Cell Invasion and Metastasis. Cancer Research. 2011; 71(24):7471–80. [PubMed: 22037868]
15. Iolascon A, Faienza MF, Coppola B, Rosolen A, Basso G, DellaRagione F, et al. Analysis of cyclin-dependent kinase inhibitor genes (CDKN2A, CDKN2B, and CDKN2C) in childhood rhabdomyosarcoma. Gene Chromosome Canc. 1996; 15(4):217–22.
16. Reichek JL, Duan FH, Smith LM, Gustafson DM, O'Connor RS, Zhang CN, et al. Genomic and Clinical Analysis of Amplification of the 13q31 Chromosomal Region in Alveolar Rhabdomyosarcoma: A Report from the Children's Oncology Group. Clinical Cancer Research. 2011; 17(6):1463–73. [PubMed: 21220470]
17. Kumar P, Henikoff S, Ng PC. Predicting the effects of coding non-synonymous variants on protein function using the SIFT algorithm. Nature protocols. 2009; 4(7):1073–81. Epub 2009/06/30.
18. Paulson V, Chandler G, Rakheja D, Galindo RL, Wilson K, Amatruda JF, et al. High-Resolution Array CGH Identifies Common Mechanisms that Drive Embryonal Rhabdomyosarcoma Pathogenesis. Gene Chromosome Canc. 2011; 50(6):397–408.
19. Martinelli S, McDowell HP, Vigne SD, Kokai G, Uccini S, Tartaglia M, et al. RAS signaling dysregulation in human embryonal Rhabdomyosarcoma. Genes Chromosomes Cancer. 2009; 48(11):975–82. Epub 2009/08/15. [PubMed: 19681119]
20. Chen YY, Takita J, Hiwatari M, Igarashi T, Hanada R, Kikuchi A, et al. Mutations of the PTPN11 and RAS genes in rhabdomyosarcoma and pediatric hematological malignancies. Gene Chromosome Canc. 2006; 45(6):583–91.
21. Bridge JA, Liu J, Qualman SJ, Suijkerbuijk R, Wenger G, Zhang J, et al. Genomic gains and losses are similar in genetic and histologic subsets of rhabdomyosarcoma, whereas amplification predominates in embryonal with anaplasia and alveolar subtypes. Genes Chromosomes Cancer. 2002; 33(3):310–21. Epub 2002/01/25. [PubMed: 11807989]
22. Barr FG, Duan F, Smith LM, Gustafson D, Pitts M, Hammond S, et al. Genomic and clinical analyses of 2p24 and 12q13-q14 amplification in alveolar rhabdomyosarcoma: a report from the Children's Oncology Group. Genes Chromosomes Cancer. 2009; 48(8):661–72. Epub 2009/05/08. [PubMed: 19422036]
23. Williamson D, Missiaglia E, de Reynies A, Pierron G, Thuille B, Palenzuela G, et al. Fusion Gene-Negative Alveolar Rhabdomyosarcoma Is Clinically and Molecularly Indistinguishable From Embryonal Rhabdomyosarcoma. Journal of Clinical Oncology. 2010; 28(13):2151–8. [PubMed: 20351326]
24. Croft D, O'Kelly G, Wu G, Haw R, Gillespie M, Matthews L, et al. Reactome: a database of reactions, pathways and biological processes. Nucleic acids research. 2011; 39(Database issue):D691–7. Epub 2010/11/12. [PubMed: 21067998]

25. Davicioni E, Finckenstein FG, Shahbazian V, Buckley JD, Triche TJ, Anderson MJ. Identification of a PAX-FKHR gene expression signature that defines molecular classes and determines the prognosis of alveolar rhabdomyosarcomas. Cancer Res. 2006; 66(14):6936–46. [PubMed: 16849537]
26. Khan J, Simon R, Bittner M, Chen Y, Leighton SB, Pohida T, et al. Gene expression profiling of alveolar rhabdomyosarcoma with cDNA microarrays. Cancer Res. 1998; 58(22):5009–13. [PubMed: 9823299]
27. Cao L, Yu Y, Bilke S, Walker RL, Mayeenuddin LH, Azorsa DO, et al. Genome-wide identification of PAX3-FKHR binding sites in rhabdomyosarcoma reveals candidate target genes important for development and cancer. Cancer Res. 2010; 70(16):6497–508. Epub 2010/07/29. [PubMed: 20663909]
28. Lagha M, Sato T, Regnault B, Cumano A, Zuniga A, Licht J, et al. Transcriptome analyses based on genetic screens for Pax3 myogenic targets in the mouse embryo. Bmc Genomics. 2010; 11
29. Skapek SX, Anderson J, Barr FG, Bridge JA, Gastier-Foster JM, Parham DM, et al. PAX-FOXO1 fusion status drives unfavorable outcome for children with rhabdomyosarcoma: A children's oncology group report. Pediatr Blood Cancer. 2013; 60(9):1411–7. Epub 2013/03/26. [PubMed: 23526739]
30. Missiaglia E, Williamson D, Chisholm J, Wirapati P, Pierron G, Petel F, et al. PAX3/FOXO1 Fusion Gene Status Is the Key Prognostic Molecular Marker in Rhabdomyosarcoma and Significantly Improves Current Risk Stratification. J Clin Oncol. 2012; 30(14):1670–7. [PubMed: 22454413]
31. Pugh TJ, Morozova O, Attiyeh EF, Asgharzadeh S, Wei JS, Auclair D, et al. The genetic landscape of high-risk neuroblastoma. Nature Genetics. 2013; 45(3):279–84. [PubMed: 23334666]
32. Grossmann V, Tiacci E, Holmes AB, Kohlmann A, Martelli MP, Kern W, et al. Whole-exome sequencing identifies somatic mutations of BCOR in acute myeloid leukemia with normal karyotype. Blood. 2011; 118(23):6153–63. [PubMed: 22012066]
33. Zhang JH, Benavente CA, McEvoy J, Flores-Otero J, Ding L, Chen X, et al. A novel retinoblastoma therapy from genomic and epigenetic analyses. Nature. 2012; 481(7381):329–34. [PubMed: 22237022]
34. Pugh TJ, Weeraratne SD, Archer TC, Krummel DAP, Auclair D, Bochicchio J, et al. Medulloblastoma exome sequencing uncovers subtype-specific somatic mutations. Nature. 2012; 488(7409):106–10. [PubMed: 22820256]
35. Lagutina I, Conway SJ, Sublett J, Grosveld GC. Pax3-FKHR knock-in mice show developmental aberrations but do not develop tumors. Molecular and cellular biology. 2002; 22(20):7204–16. Epub 2002/09/21. [PubMed: 12242297]
36. Linardic CM, Naini S, Herndon JE, Kesserwan C, Qualman SJ, Counter CM. The PAX3-FKHR fusion gene of rhabdomyosarcoma cooperates with loss of p16(INK4A) to promote bypass of cellular senescence. Cancer Research. 2007; 67(14):6691–9. [PubMed: 17638879]
37. Keller C, Arenkiel BR, Coffin CM, El-Bardeesy N, DePinho RA, Capecchi MR. Alveolar rhabdomyosarcomas in conditional Pax3:Fkhr mice: cooperativity of Ink4a/ARF and Trp53 loss of function. Genes & development. 2004; 18(21):2614–26. Epub 2004/10/19. [PubMed: 15489287]
38. Ascierto PA, Schadendorf D, Berking C, Agarwala SS, van Herpen CML, Queirolo P, et al. MEK162 for patients with advanced melanoma harbouring NRAS or Val600 BRAF mutations: a non-randomised, open-label phase 2 study. Lancet Oncol. 2013; 14(3):249–56. [PubMed: 23414587]
39. Renshaw J, Taylor KR, Bishop R, Valenti M, De Haven Brandon A, Gowan S, et al. Dual blockade of the PI3K/AKT/mTOR (AZD8055) and RAS/MEK/ERK (AZD6244) pathways synergistically inhibits rhabdomyosarcoma cell growth in vitro and in vivo. Clinical cancer research : an official journal of the American Association for Cancer Research. 2013 Epub 2013/08/07.
40. Huynh KD, Fischle W, Verdin E, Bardwell VJ. BCoR, a novel corepressor involved in BCL-6 repression. Genes & development. 2000; 14(14):1810–23. [PubMed: 10898795]
41. Gearhart MD, Corcoran CM, Wamstad JA, Bardwell VJ. Polycomb group and SCF ubiquitin ligases are found in a novel BCOR complex that is recruited to BCL6 targets. Molecular and cellular biology. 2006; 26(18):6880–9. [PubMed: 16943429]

42. Shah SP, Roth A, Goya R, Oloumi A, Ha G, Zhao YJ, et al. The clonal and mutational evolution spectrum of primary triple-negative breast cancers. Nature. 2012; 486(7403):395–9. [PubMed: 22495314]
43. Morin RD, Mendez-Lago M, Mungall AJ, Goya R, Mungall KL, Corbett RD, et al. Frequent mutation of histone-modifying genes in non-Hodgkin lymphoma. Nature. 2011; 476(7360):298–303. [PubMed: 21796119]
44. Petricoin EF 3rd, Espina V, Araujo RP, Midura B, Yeung C, Wan X, et al. Phosphoprotein pathway mapping: Akt/mammalian target of rapamycin activation is negatively associated with childhood rhabdomyosarcoma survival. Cancer Res. 2007; 67(7):3431–40. [PubMed: 17409454]
45. Cen L, Arnoczky KJ, Hsieh FC, Lin HJ, Qualman SJ, Yu SL, et al. Phosphorylation profiles of protein kinases in alveolar and embryonal rhabdomyosarcoma. Modern Pathol. 2007; 20(9):936–46.
46. Drmanac R, Sparks AB, Callow MJ, Halpern AL, Burns NL, Kermani BG, et al. Human genome sequencing using unchained base reads on self-assembling DNA nanoarrays. Science. 2010; 327(5961):78–81. Epub 2009/11/07. [PubMed: 19892942]
47. Zhang J, Ding L, Holmfeldt L, Wu G, Heatley SL, Payne-Turner D, et al. The genetic basis of early T-cell precursor acute lymphoblastic leukaemia. Nature. 2012; 481(7380):157–63. Epub 2012/01/13. [PubMed: 22237106]
48. Reumers J, De Rijk P, Zhao H, Liekens A, Smeets D, Cleary J, et al. Optimized filtering reduces the error rate in detecting genomic variants by short-read sequencing. Nat Biotechnol. 2012; 30(1): 61–U103. [PubMed: 22178994]
49. Molenaar JJ, Koster J, Ebus ME, Zwijnenburg DA, van Sluis P, Lamers F, et al. The Genomic landscape of neuroblastoma. Cell Oncol. 2012; 35:S11–S.
50. Carnevali P, Baccash J, Halpern AL, Nazarenko I, Nilsen GB, Pant KP, et al. Computational Techniques for Human Genome Resequencing Using Mated Gapped Reads. J Comput Biol. 2012; 19(3):279–92. [PubMed: 22175250]
51. Wang K, Li MY, Hakonarson H. ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. Nucleic acids research. 2010; 38(16)
52. Sorensen PHB, Lynch JC, Qualman SJ, Tirabosco R, Lim JF, Maurer HM, et al. PAX3-FKHR and PAX7-FKHR gene fusions are prognostic indicators in alveolar rhabdomyosarcoma: A report from the children's oncology group. Journal of Clinical Oncology. 2002; 20(11):2672–9. [PubMed: 12039929]
53. Trapnell C, Pachter L, Salzberg SL. TopHat: discovering splice junctions with RNA-Seq. Bioinformatics. 2009; 25(9):1105–11. [PubMed: 19289445]
54. Kim D, Salzberg SL. TopHat-Fusion: an algorithm for discovery of novel fusion transcripts. Genome Biol. 2011; 12(8)
55. McPherson A, Hormozdiari F, Zayed A, Giuliany R, Ha G, Sun MGF, et al. deFuse: An Algorithm for Gene Fusion Discovery in Tumor RNA-Seq Data. Plos Comput Biol. 2011; 7(5)
56. Trapnell C, Williams BA, Pertea G, Mortazavi A, Kwan G, van Baren MJ, et al. Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. Nature Biotechnology. 2010; 28(5):511–U174.
57. Homer N, Merriman B, Nelson SF. BFAST: An Alignment Tool for Large Scale Genome Resequencing. Plos One. 2009; 4(11):A95–A106.
58. McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernytsky A, et al. The Genome Analysis Toolkit: A MapReduce framework for analyzing next-generation DNA sequencing data. Genome Res. 2010; 20(9):1297–303. [PubMed: 20644199]
59. DePristo MA, Banks E, Poplin R, Garimella KV, Maguire JR, Hartl C, et al. A framework for variation discovery and genotyping using next-generation DNA sequencing data. Nat Genet. 2011; 43(5):491-+. [PubMed: 21478889]
60. Barbieri CE, Baca SC, Lawrence MS, Demichelis F, Blattner M, Theurillat JP, et al. Exome sequencing identifies recurrent SPOP, FOXA1 and MED12 mutations in prostate cancer. Nat Genet. 2012; 44(6):685–U107. [PubMed: 22610119]

61. Stransky N, Egloff AM, Tward AD, Kostic AD, Cibulskis K, Sivachenko A, et al. The Mutational Landscape of Head and Neck Squamous Cell Carcinoma. Science. 2011; 333(6046):1157–60. [PubMed: 21798893]

62. Chapman MA, Lawrence MS, Keats JJ, Cibulskis K, Sougnez C, Schinzel AC, et al. Initial genome sequencing and analysis of multiple myeloma. Nature. 2011; 471(7339):467–72. [PubMed: 21430775]

63. Wei XM, Walia V, Lin JC, Teer JK, Prickett TD, Gartner J, et al. Exome sequencing identifies GRIN2A as frequently mutated in melanoma. Nat Genet. 2011; 43(5):442-+. [PubMed: 21499247]

64. Benjamini Y, Hochberg Y. Controlling the False Discovery Rate - a Practical and Powerful Approach to Multiple Testing. J Roy Stat Soc B Met. 1995; 57(1):289–300.

65. Diskin SJ, Eck T, Greshock J, Mosse YP, Naylor T, Stoeckert CJ, et al. STAC: A method for testing the significance of DNA copy number aberrations across multiple array-CGH experiments. Genome Res. 2006; 16(9):1149–58. [PubMed: 16899652]

66. Khan J, Bittner ML, Chen YD, Faller AJ, Saal LH, Azorsa DA, et al. Elucidation of the downstream targets of the PAX3-FKHR fusion oncogene found in aveolar rhabdomyosarcoma using cDNA microarrays. Pediatr Res. 1999; 45(4):148a–a. [PubMed: 9890624]

67. Subramanian A, Tamayo P, Mootha VK, Mukherjee S, Ebert BL, Gillette MA, 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(43):15545–50. [PubMed: 16199517]

Abbreviations

RMS	Rhabdomyosarcoma
WGS	Whole Genome Sequencing
WES	Whole exome sequencing
PFP	PAX gene fusion positive
PFN	PAX gene fusion negative

Statement of Significance

This is the most comprehensive genomic analysis of rhabdomyosarcoma to date. Despite a relatively low mutation rate, multiple genes were recurrently altered including *NRAS*, *KRAS*, *HRAS*, *FGFR4*, *PIK3CA*, *CTNNB1*, *FBXW7*, and *BCOR*. In addition, a majority of rhabdomyosarcoma tumors alter the tyrosine kinase/*RAS/PIK3CA* axis providing opportunity for genomics guided intervention.

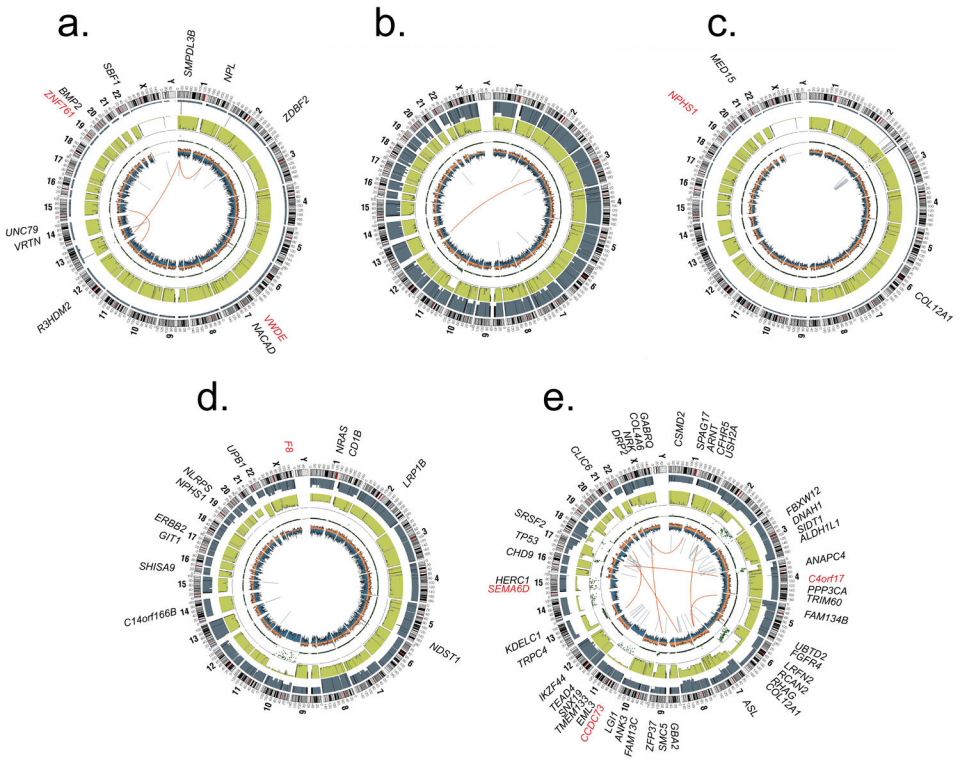
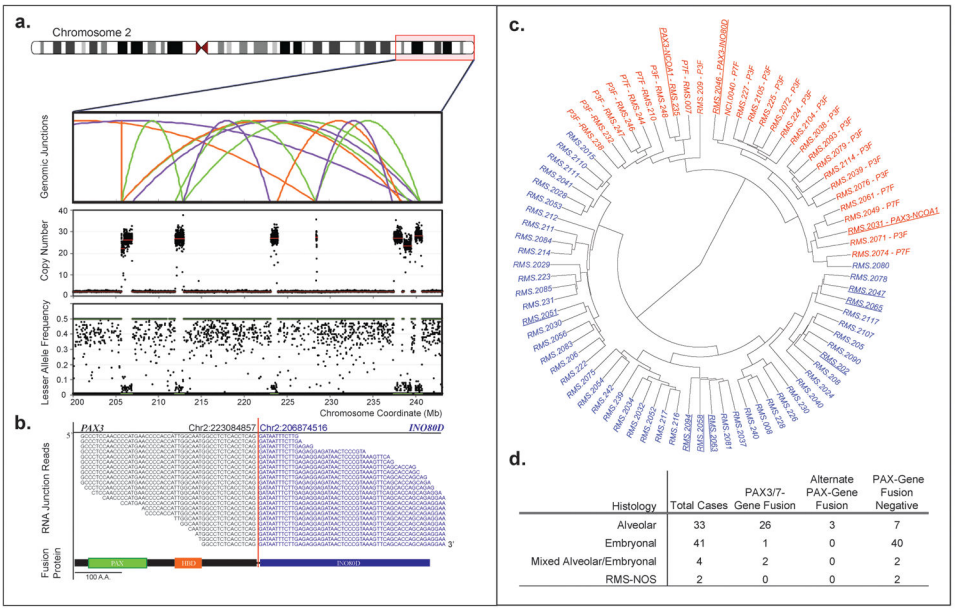


Figure 1.

Circos Plots of RMS represenative tumors. Circos plot tracks representing verified somatic mutations, from outside circle; mutated genes missense mutations (Black), Nonsense and indel mutations (Red); genomic location, genome wide copy number alterations (Gray), lesser allele frequency (Green) LOH (dotted track), density of heterozygous SNPs (Orange) homozygous SNPs (Blue). Intrachromasomal rearrangements (inner circle gray) and Interchromasomal rearrangements (Inner circle red). **a**, NCI-40: A PAX7-FOXO1 translocation noting the associated high level copy number gain. This tumor also has high level copy number gain of MYCN on chromosome 2. **b**, RMS224: Represenative *PAX3-FOXO1* fusion with no somatic point mutations. Note LOH on short arm of Chr11p. **c**, RMS2046: Multiple rearrangements on Chromosome 2 with corresponding junctions and copy number changes. This rearrangement produces a novel gene fusion of *PAX3-INO80D*. **d**, RMS216: Represenative PFN RMS. Note relative increase in point mutations including *NRAS* mutation on Chr1 and increase in aneuploidy including gain of Chr8. Complete loss of heterozygozity on Chr 10 and the short arm of Chr11 **e**, RMS2030: Multiple genome wide alterations in a tumor with *TP53* mutation. Point mutation of *FGFR4* on Chromosome 5.



Histology	Total Cases	PAX3/7-Gene Fusion	Alternate PAX-Gene Fusion	PAX-Gene Fusion Negative
Alveolar	33	26	3	7
Embryonal	41	1	0	40
Mixed Alveolar/Embryonal	4	2	0	2
RMS-NOS	2	0	0	2

Figure 2.

Rearrangement of chromosome 2 in RMS2046 produces *PAX3-INO80D* fusion. **a**, WGS junctions. Purple lines represent tail-to-head junction, green lines show head-to-tail junction (possibly tandem duplication) and orange lines show tail-to-tail junction or head-to-head junction (inversion). LogR ratio and VAF (Variant Allele Frequency) demonstrate two distinct copy number patterns with multiple breakpoints. **b**, RNA sequencing discovered 22 high quality reads spanning the junction of the *PAX3* and *INO80D* genes. The fusion joins exon 7 of *PAX3* with exon 9 of *INO80D*. The predicted fusion protein maintains the paired box domain (labeled PAX in white) and homeobox DNA binding domain (labeled HBD in white) of the PAX3 protein. The two predicted N-terminal DNA binding domains of INO80D are lost in the fusion. **c**, Unsupervised clustering of transcriptome expression data demonstrates a clear definition between fusion positive (Red) and fusion negative (Blue) tumors including the alternate PAX gene fusions with *NCOA1* and *INO80D*. “Fusion negative” alveolar histology tumors are underlined. RMS 2080 is a fusion negative embryonal tumor that appears to cluster amoung the fusion positive tumors however no rearrangement of the PAX gene was discovered by RT-PCR or transcriptome sequencing **d**, Histologic diagnosis of tumors evaluated by transcriptome sequencing reveals 10 tumors with “fusion negative” alveolar histology.

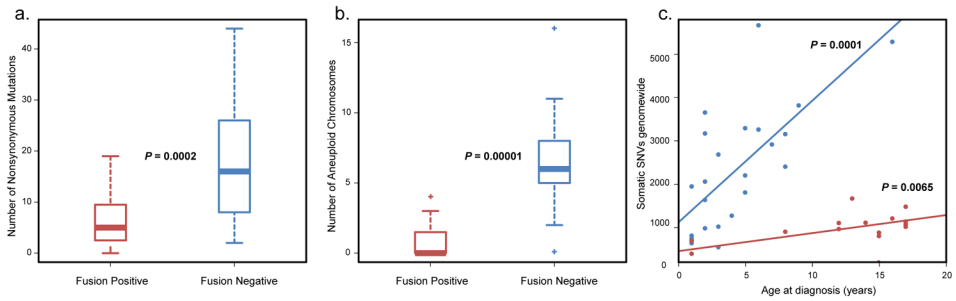


Figure 3.

Fusion-positive and fusion-negative RMS have distinct genotypes. **a**, Number of protein coding mutations in fusion-positive tumors (red) and fusion negative tumors (blue). **b**, Significant difference in the number of aneuploid chromosomes between fusion-positive tumors (red) and fusion-negative (blue) tumors. **c**, Age at diagnosis versus genome-wide mutations in fusion-positive (red) versus fusion negative (blue).

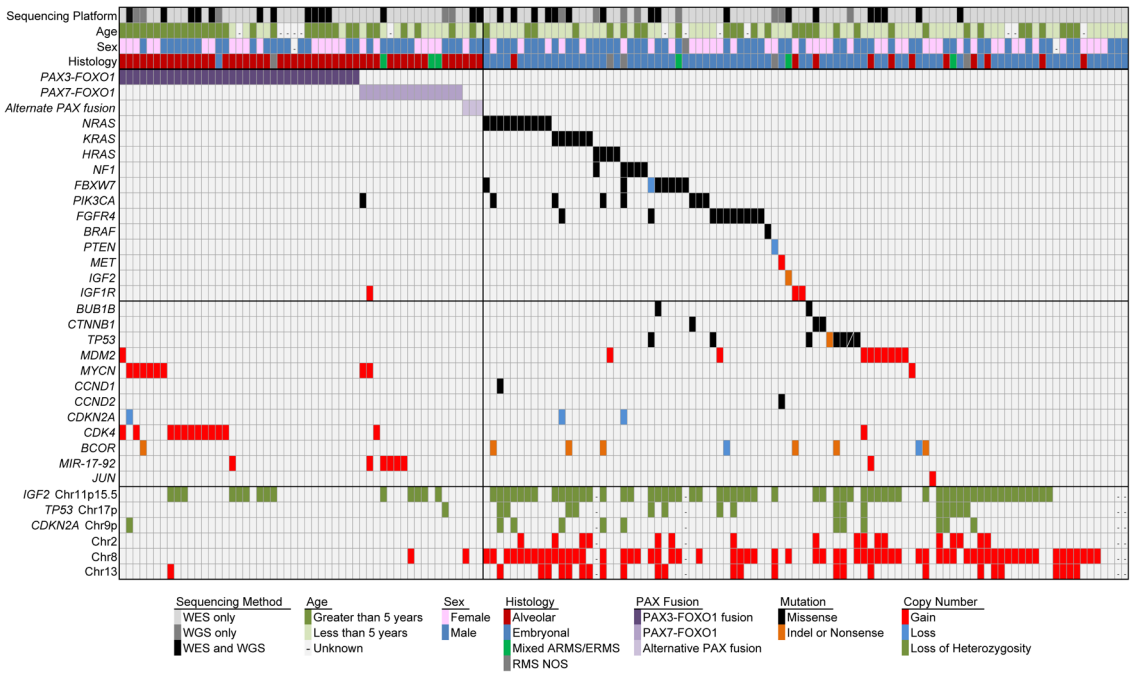


Figure 4.

The Genomic Landscape of pediatric RMS highlighting candidate alterations. Demographic characteristics, histologic subtypes and selected genes with copy number alterations or somatic mutations across 147 rhabdomyosarcoma cases. Unique sample identifier and sequencing platform. Sex, males in blue, females in pink. Age, years at diagnosis divided into less than 5 years and greater than 5 years. Histologic diagnosis, Red, Alveolar; Blue, Embryonal including Spindle and Botryoid subtypes; Gray, RMS not otherwise specified. Mixed alveolar and embryonal histology in green. Copy number gains and losses for selected genes. Blue, losses; red, gains; green, loss of heterozygozity. Selected genes with somatic mutations. Purple, fusion protein; black, missense; orange, nonsense/splice site/ indel mutations.

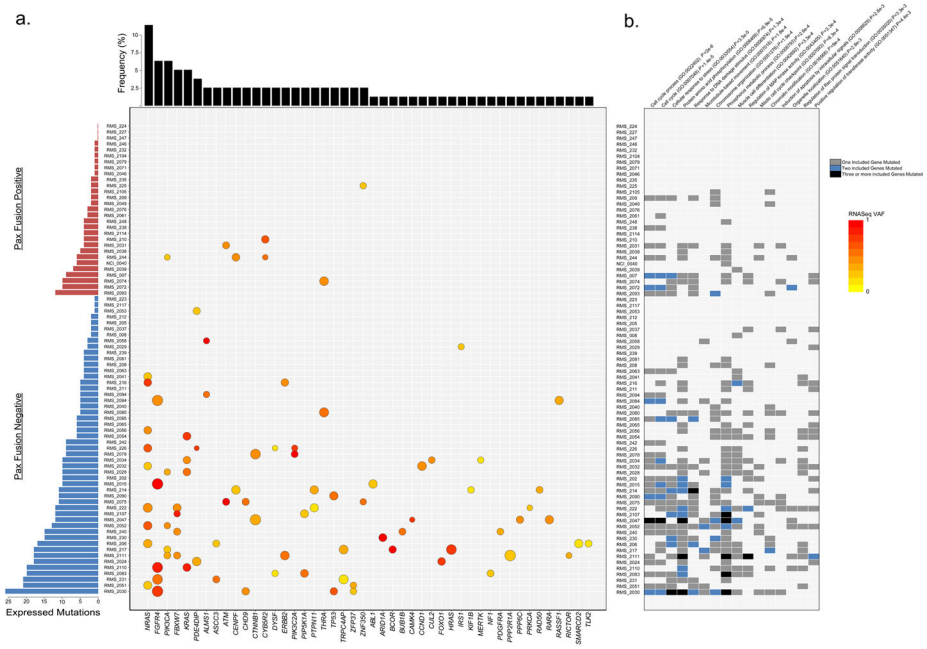


Figure 5.

Expressed mutations in 80 RMS tumors. **a**. Candidate somatic alterations found to be expressed in whole transcriptome sequencing and the discovered genes ranked by frequency. Top, the number of expressed mutations by sample; Blue, PAX gene fusion negative, Red, PAX gene fusion positive. The color (yellow to red) of the mark represents the variant allele frequency (VAF) with many mutations appearing to favor the mutant allele. The size of the circle is proportional to the fragments per kilobase of transcript per million mapped reads (FPKM). **b**. Gene Ontology analysis of the expressed mutations reveals multiple alterations of cell cycle, cellular response to stress, protein amino acid phosphorylation, response to DNA damage stimulus, microtubule-based movement, chromosome organization, muscle cell differentiation, regulation of MAP kinase activity, mitotic cell cycle checkpoint, chromatin modification, induction of apoptosis by intracellular signals, organelle localization, regulation of Rac protein signal transduction and regulation of transferase activity. Many tumors appear to accumulate multiple mutations in the same pathway (blue = 2, black = 3 or more).

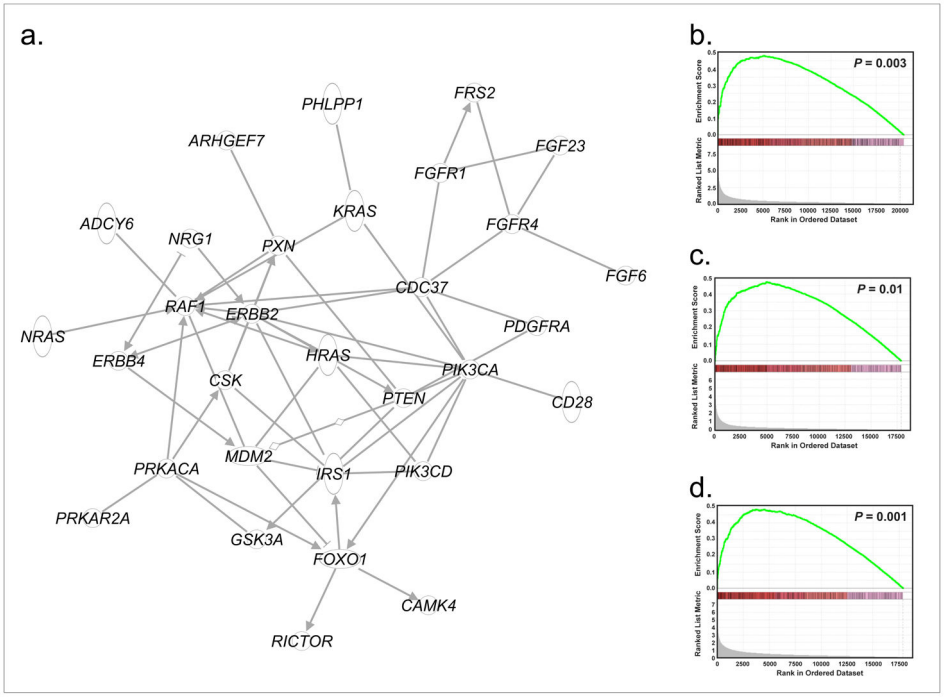


Figure 6.

a, Gene interaction map of Reactome pathway analysis that discovers alteration of FGFR signaling as the most altered pathway. 22/25 fusion negative tumors alter at least one gene in the pathway. **b**, GSEA enrichment plot of altered genes in fusion negative RMS tumors versus altered genes in the PAX3-FOXO1 expressing model cell line (7250_PF) with enrichment scores plotted for each gene moving down the ranked list of genes. Genes altered in the 7250_PF cell line show significant enrichment in the fusion negative tumors (FDR q-value = 0.004). GSEA enrichment plot of the altered genes in published mouse models of PAX3-FOXO1 in **c**, somite cells (FDR q-value = 0.009) and **d**, forelimb cells (FDR q-value 0.0009).

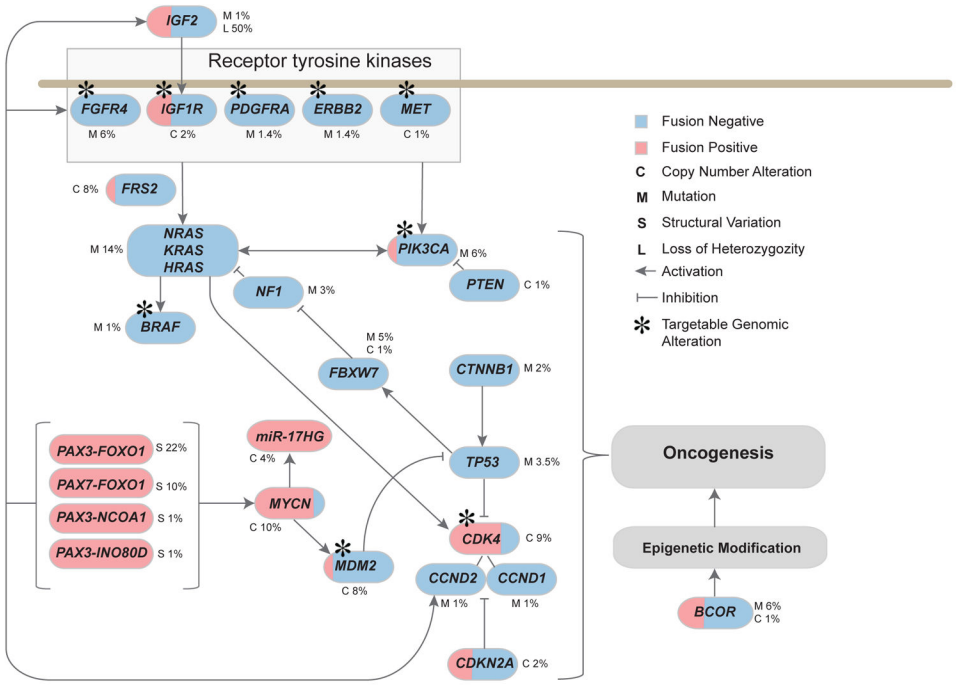


Figure 7.

Model pathway altered in Rhabdomyosarcoma. Genes colored red are found in fusion positive tumors while genes colored blue are found in tumors without a PAX gene fusion. Alterations and their frequency in the population include mutations and small indels (M), copy number deletions and amplifications (C), or structural variations (S) that affect the gene.

Table 1

Genes with significant frequency of somatic mutation across 147 Rhabdomyosarcomas

Gene	WGS	WES	RMS Recurrence	RMS Frequency	PAX-Fusion Positive Recurrence	PAX-Fusion Positive Frequency	PAX-Fusion Negative Recurrence	PAX-Fusion Negative Frequency	FDR
NRAS	4	7	11	7.5%	0	0	11	11.7%	5.10E-09
FGFR4	3	6	9	6.1%	0	0	9	9.6%	3.15E-12
PIK3CA	3	5	8	5.4%	1	1.9%	7	7.4%	5.86E-10
BCOR	3	5	8	5.4%	1	1.9%	7	7.4%	2.11E-08
FBXW7	3	4	7	4.8%	0	0	7	7.4%	2.11E-08
KRAS	3	3	6	4.1%	0	0	6	6.4%	5.51E-06
TP53	1	4	5	3.4%	0	0	5	5.3%	5.51E-06
NF1	3	2	5	3.4%	0	0	5	5.3%	2.06E-03
HRAS	2	2	4	2.7%	0	0	4	4.3%	1.12E-05

WGS, Whole Genome Sequencing; WES, Whole Exome Sequencing; FDR, False Discovery Rate