



**HAL**  
open science

## Discovery of New Fusion Transcripts in a Cohort of Pediatric Solid Cancers at Relapse and Relevance for Personalized Medicine

Célia Dupain, Anne C Harttrampf, Yannick Boursin, Manuel Lebeurrier, Windy Rondof, Guillaume Robert-Siegwald, Pierre Khoueiry, Birgit Georger, Liliane Massaad-Massade

### ► To cite this version:

Célia Dupain, Anne C Harttrampf, Yannick Boursin, Manuel Lebeurrier, Windy Rondof, et al.. Discovery of New Fusion Transcripts in a Cohort of Pediatric Solid Cancers at Relapse and Relevance for Personalized Medicine. *Molecular Therapy*, 2019, 27 (1), pp.200 - 218. 10.1016/j.ymthe.2018.10.022 . hal-03285971

**HAL Id: hal-03285971**

**<https://hal.science/hal-03285971>**

Submitted on 13 Jul 2021

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# Discovery of New Fusion Transcripts in a Cohort of Pediatric Solid Cancers at Relapse and Relevance for Personalized Medicine

Célia Dupain,<sup>1,2,3,7</sup> Anne C. Harttrampf,<sup>1,2,3,6,7</sup> Yannick Boursin,<sup>4</sup> Manuel Lebourrier,<sup>4</sup> Windy Rondof,<sup>1,2,3,4</sup> Guillaume Robert-Siegwald,<sup>4</sup> Pierre Khoueir,<sup>5</sup> Birgit Georger,<sup>1,2,3,6</sup> and Liliane Massaad-Massade<sup>1,2,3</sup>

<sup>1</sup>Université Paris-Sud 11, Laboratoire de Vectorologie et Thérapeutiques Anticancéreuses, UMR 8203, Villejuif 94805, France; <sup>2</sup>CNRS, Villejuif, Laboratoire de Vectorologie et Thérapeutiques Anticancéreuses, UMR 8203, Villejuif 94805, France; <sup>3</sup>Gustave Roussy, Laboratoire de Vectorologie et Thérapeutiques Anticancéreuses, UMR 8203, Villejuif 94805, France; <sup>4</sup>Gustave Roussy, Bioinformatics Platform, AMMICA, INSERM US23/CNRS UMS3655, Villejuif 94805, France; <sup>5</sup>American University of Beirut, Faculty of Medicine, Department of Biochemistry and Molecular Genetics, P.O. Box 11-0236 DTS 419-B, Bliss Street, Beirut, Lebanon; <sup>6</sup>Gustave Roussy, Department of Pediatric and Adolescent Oncology, Villejuif 94805, France

**We hypothesized that pediatric cancers would more likely harbor fusion transcripts. To dissect the complexity of the fusions landscape in recurrent solid pediatric cancers, we conducted a study on 48 patients with different relapsing or resistant malignancies. By analyzing RNA sequencing data with a new in-house pipeline for fusions detection named Chim-Comp, followed by verification by real-time PCR, we identified and classified the most confident fusion transcripts (FTs) according to their potential biological function and druggability. The majority of FTs were predicted to affect key cancer pathways and described to be involved in oncogenesis. Contrary to previous descriptions, we found no significant correlation between the number of fusions and mutations, emphasizing the particularity to study pre-treated pediatric patients. A considerable proportion of FTs containing tumor suppressor genes was detected, reflecting their importance in pediatric cancers. FTs containing non-receptor tyrosine kinases occurred at low incidence and predominantly in brain tumors. Remarkably, more than 30% of patients presented a potentially druggable high-confidence fusion. In conclusion, we detected new oncogenic FTs in relapsing pediatric cancer patients by establishing a robust pipeline that can be applied to other malignancies, to detect and prioritize experimental validation studies leading to the development of new therapeutic options.**

## INTRODUCTION

Fusion transcripts, or chimeric RNA, result from the juxtaposition of two genes, previously located separately from each other, due to chromosomal or non-chromosomal events. They can be the consequence of structural chromosomal rearrangements or be the product of alternative (*cis*-, *trans*-) splicing or transcriptional read-throughs.<sup>1</sup> Fusion transcripts can lead to the activation of proto-oncogenes or inactivation of tumor suppressor genes (TSGs), and they are considered as one of the leading mechanisms responsible for carcinogenesis.<sup>2</sup>

They present a great interest in clinics as they are increasingly explored as potential therapeutic targets fueled by the discovery of a drug, imatinib mesylate, able to target BCR-ABL fusion oncoprotein.<sup>3</sup> Several studies have aimed at targeting fusion products in various neoplasms, including (1) inhibitors of anaplastic lymphoma kinase (ALK) in *EML4-ALK*-positive lung cancer or *NPM-ALK*-positive anaplastic large-cell lymphomas, (2) larotrectinib in *NTRK*-containing fusions in various solid tumors,<sup>4</sup> and (3) FGFR inhibitors to target FGFR-TACC fusion detected in 3% of IDH1/2 wild-type glioblastoma and other malignancies.<sup>5</sup>

Fusion transcripts are considered as strong diagnostic biomarkers with several examples in this regard. Indeed, *EWS-FLI1* found in Ewing's sarcoma or *PAX3-FOXO1* responsible for 55%–70% of alveolar rhabdomyosarcoma are characteristic disease-defining fusion transcripts. The detection of *DNAJB1-PRKCA* gene fusion in fibrolamellar hepatocellular carcinoma (FLHCC) serves as illustrative example for both the powerful detection at unprecedented resolution by RNA sequencing (RNA-seq) and its successful diagnostic use in extremely rare tumors.<sup>6</sup> Therefore, the discovery of additional actionable fusion transcripts in cancers is attracting more attention.

High-throughput RNA-seq is a powerful tool to detect so far undiscovered fusion events. On the other hand, as reflected by the dramatic increase in reported gene fusions since its introduction, this will likely include rearrangements that represent secondary passenger events, due to the genomically unstable nature of cancer, and, in individual patients, non-recurrent events whose biological significance is

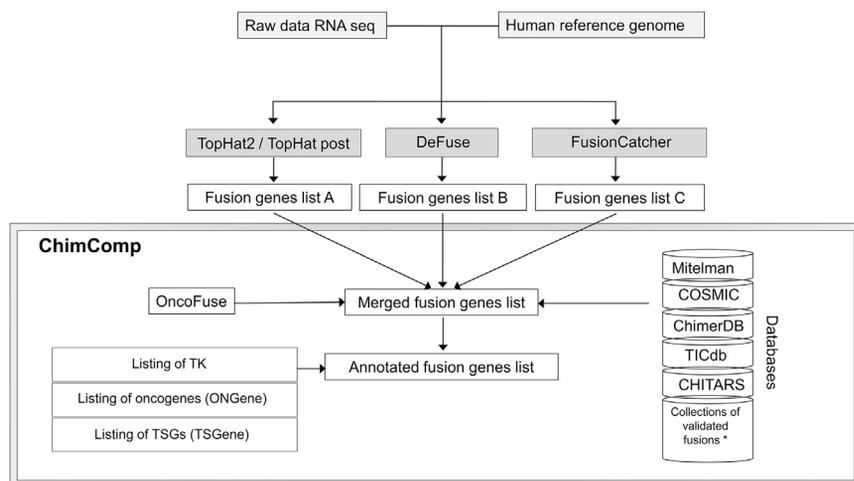
---

Received 23 May 2018; accepted 26 October 2018;  
<https://doi.org/10.1016/j.ymthe.2018.10.022>.

<sup>7</sup>These authors contributed equally to this work.

Correspondence: Liliane Massaad-Massade, PhD, Université Paris-Sud 11, Laboratoire de Vectorologie et Thérapeutiques Anticancéreuses, UMR 8203, 114 rue Edouard Vaillant, Villejuif 94805, France.

E-mail: [liliane.massade@gustaveroussy.fr](mailto:liliane.massade@gustaveroussy.fr)



**Figure 1. ChimComp Pipeline Workflow**

Raw data from RNA-seq were mapped on the human reference genome GRCh37/hg19 (Genome Reference Consortium Human Build 37/human genome 19). Three software tools (TopHat2 followed by TopHat post, deFuse, and FusionCatcher) were used to generate fusion gene lists A, B, and C. ChimComp processed the three lists in order to provide a unique merged fusion gene list per patient. The listing was analyzed by Oncofuse (left) to generate the oncoscores and compared to databases of known fusions (right). TK genes, oncogenes, and tumor suppressor genes were searched in the final annotated fusion genes list. \*Collection of fusions is according to Babak Alaei-Mahabadi et al.<sup>70</sup> TK, tyrosine kinases; TSGs, tumor suppressor genes.

currently unclear.<sup>1</sup> Hence, it is important to keep the balance between stringent selection parameters to avoid false positives without missing relevant oncogenic fusions.<sup>7,8</sup>

To date there exist more than 30 different bioinformatic tools for fusion detection,<sup>9</sup> each differing in terms of sensitivity and specificity. They majorly differ by their filtering approaches and algorithms regarding, among others, the distance between pair-end reads, the number of nucleotides mapping each side of a fusion breakpoint, the presence of readthrough transcripts, the presence of duplicated reads, and the number of reads mapping to homologous or repetitive regions. In numerous benchmarking studies, none of the existing tools has clearly outperformed;<sup>7,10,11</sup> therefore, a meta-caller approach by combining different tools has been shown to increase both specificity and sensitivity.<sup>10</sup>

Additionally, to translate bioinformatic results into meaningful biology, predictive algorithms such as Oncofuse<sup>12</sup> have been developed to predict the oncogenicity of the detected fusions. This can help to select candidates with higher confidence for further experimental exploration, thus reducing financial and human burdens.

Solid pediatric cancers represent around 60% of all pediatric neoplasms and include the cases of malignancies diagnosed under 14 years old. Around 80% of pediatric patients reach long-term remission after treatment while 20% of them die due to treatment failure. It is well-known that pediatric cancers differ from those of adults, as many arise from embryonal rather than epithelial cells<sup>13</sup> and as they particularly harbor low mutation rates, mostly due to the lower implication of exogenous toxic effects.<sup>6</sup> The exploration of the fusion landscape in these pathologies could, therefore, help to unravel the particularity of these malignancies and provide new therapeutic targets.

Given the limited data in solid pediatric cancer patients at relapse, we conducted our study using previously published RNA-seq data of

48 pediatric patients with different malignancies, generated as part of the Molecular Screening for Cancer Treatment Optimization (MOSCATO-01, ClinicalTrials.gov: NCT01566019) trial conducted at Gustave Roussy.<sup>14</sup> Our aim was to identify new fusion transcripts in recurrent pediatric cancers and to predict their probable oncogenic potential and targetability. For this, we established a reliable *in silico* detection method named ChimComp that integrates information from three different fusion detection algorithms, along with their corresponding oncogenic effects and targetability. We then confirmed 91% of the tested candidates and identified new potential targetable oncogenic fusions.

## RESULTS

### Patient Characteristics

Clinical characteristics of patients are provided in Table S1. Briefly, the transcriptome of 48 patients included in the MOSCATO-01 clinical trial with 22 different tumor entities at relapse or resistance was previously sequenced using RNA-seq.<sup>14</sup> The median age was 10.9 years (range, 0.8–23.3 years) and the sex ratio was equal to 0.71 (28 males). Patients were classified into 3 categories based on their tumor types: sarcomas (40%), brain tumors (35%), and other types (25%). All patients had undergone antineoplastic treatment (median number of lines, 2; range, 1–8). The median number of tumor cells in the analyzed specimen was high (around 80%).

### ChimComp Is a Powerful Pipeline to Detect Fusions from RNA-Seq Data

Pediatric cancers are known to harbor a low mutational rate, which suggests that other genomic alterations, including, among others, fusion transcripts and structural variants, may be causative. To increase our ability to detect true positive events from RNA-seq data, we devised a computational approach called ChimComp (Figure 1) that combines fusion transcript calls from three vastly used algorithms: deFuse,<sup>15</sup> FusionCatcher,<sup>16</sup> and TopHat<sup>17</sup> (Materials and Methods). All three tools have been used previously and showed diverse specificity and sensitivity in calling fusion transcript from RNA-seq data.

Using our ChimComp approach, we detected a total of 1,374 fusions in the 48 patients analyzed. The median number of fusions per patient detected was 49.5 fusions (range, 3–223) (Table 1). Of those 1,374 fusions, 787 (57.2%) were intra-chromosomal and 588 (42.8%) were inter-chromosomal. Only 8% (107/1,374) were already known and annotated in at least one of the databases, suggesting that the majority of fusion transcripts called here are *de novo* and potentially positive.

We then assessed the presence of known oncogenes and TSGs from publicly available databases<sup>18–20</sup> (Figure 1). A total of 91/1,374 (7%) fusions were found to contain an oncogene (Table S2) and 137/1,374 (11%) a TSG (Table S3). Interestingly, every patient had at least one fusion containing either a known oncogene or TSG (Table 1), except patient 45 who had very few fusions detected. Nervous system tumors were the ones with the highest rate of oncogenes (ependymoma, high-grade glioma [HGG], neuroblastoma, medulloblastoma, and diffuse intrinsic pontine glioma [DIPG]), while tumors with the highest number of TSGs were of various origins. A strong correlation between the total number of fusions and the number of fusions containing an oncogene or a TSG per patient was found (respectively, Spearman coefficient  $r = 0.4637$ ,  $p = 0.001$  and  $r = 0.6948$ ,  $p < 0.0001$ ).

When considering the output of each software independently, deFuse predicted the largest number of fusions followed by TopHat, whereas FusionCatcher software was the most conservative. A large number of fusions were exclusively detected by either deFuse or TopHat (29.8% and 9.5%, respectively), and, when combined, these 2 tools shared the highest number of fusion gene overlap (46.4%) (Figure 2A).

To define a high-confidence group of fusion transcripts, we combined the calls of the three algorithms by considering intersecting events. This identified a set of 118 fusions (9%, 118/1,374), of which 107 (90%) have not been previously reported (Table S4). The repartition of fusions as intra-chromosomal or inter-chromosomal for the common set followed the same tendency observed in the total fusions, with 66/118 (55.9%) intra-chromosomal fusions and 52/118 (44.1%) inter-chromosomal fusions. The median spanning read number (representing the reads aligning to the junction sequence) was 12.7 per fusion, emphasizing the high confidence of call in this group as they were mainly supported by high read numbers. 45 fusions (38%) were predicted in-frame by FusionCatcher, and 47/118 (39%) fusions had one or both genes known to be involved in tumorigenicity, suggesting a fundamental role on oncogenicity and tumor progression (Table S4). Interestingly, among the cancer pathologies included in the cohort, medulloblastoma and glioblastoma were highly represented in the high-confidence fusion candidates detected, which comprised 2 patients with highly rearranged transcriptome (16 and 48). Moreover, all the known fusions characterizing a cancer pathology (e.g., *EWS/FLI-1* in EWS, patients 11 and 46; or *HEY1-NCOA2* in mesenchymal chondrosarcoma, patient 33) were detected by our pipeline, reflecting the robustness of the ChimComp approach.

To identify fusions predicted to have a potential driving role, we scored detected fusions for their oncogenic potential using Onco-

fuse,<sup>12</sup> and we categorized all 1,374 fusions into three groups: fusions with high oncoscore (oncoscore  $> 0.7$ ; 157/1,374 or 11.5%), fusions with a mild oncoscore ( $0.35 < \text{oncoscore} < 0.7$ ; 45/1,374 or 3.3%), and fusions with a low oncoscore (oncoscore  $< 0.35$ ; 258/1,374 or 18.8%) (Figure 2B). A total of 66.5% fusions (914/1,374) did not have an oncoscore, most likely due to limitations of Oncofuse. Interestingly, the percentage of fusions with oncoscores significantly increased ( $Z$  score  $< 1.96$ ) when we considered the set of 118 high-confidence fusions (Figure 2B). For instance, 36.4% (43/118) of fusions jointly identified by the 3 tools exhibited a high oncoscore compared to 11.5% only (157/1,374), when all fusions were considered (Figure 2B). Similarly, 30.5% and 9.3% of fusions from the joint set belonged to categories of low and mild oncoscore, respectively, compared to 18.7% (258/1,374) and 3.3% (45/1,374), respectively, when all fusions were considered (Figure 2B). Regarding fusions containing an oncogene, 29/91 (31.8%) had a high oncoscore ( $> 0.7$ ) and 47/91 (51.6%) had no oncoscore. For the TSG-containing fusions, 30/148 (20.3%) had a high oncoscore and 76/148 (51.4%) had no oncoscore.

Finally, in order to get functional insights for the identified fusions, we predicted the biological function of each fusion, and we classified them according to the hallmarks of cancer described by Hanahan and Weinberg.<sup>21</sup> The major biological alterations identified were sustainability to proliferative signaling (18.1%), resistance to cell death (10.5%), and evasion to growth suppressors (10.5%) (Figure 2C). This suggests the involvement of these fusions in oncogenesis. Interestingly, 6.8% of fusions could affect several categories of biological functions, leading to possible multiple consequences. For a large proportion of fusions (17.3%), no classification could be attributed. Remarkably, those fusions were found to have a low or no oncoscore.

The above analysis showed that combining different fusion calling tools associated with a valid scoring scheme (Oncoscore) increases our chances in calling potentially active and actionable fusion transcripts.

#### The Occurrence of Fusions in Relapsed Pediatric Cancer Patients Is Not Related to the Number of Genomic Breakpoints or Mutations

Previous studies reported that the number of fusion events and genomic instability were positively correlated while the number of mutations and fusion events were negatively correlated.<sup>1,22,23</sup> Taking advantage of the RNA-seq, whole-exome sequencing (WES), and comparative genomic hybridization (CGH) data available for our samples,<sup>14</sup> we asked whether our cohort of pediatric cancer patients re-called known correlations or not.

The mutational burden (number of somatic non-synonymous mutations) as well as the number of genomic breakpoints was calculated from WES and CGH arrays, respectively<sup>14</sup> (Materials and Methods). Consistent with previous findings in pediatric tumor patients,<sup>24</sup> the majority of our samples harbored a low mutation rate, with a median of somatic non-synonymous mutations of 81 per patient

**Table 1. Fusion and Genomic Alteration Metrics for the 48 Pediatric Patients Studied**

Diagnosis <sup>a</sup>	Tumor Type <sup>b</sup>	Total Fusions Detected	No. of Oncogenes Detected among Fusions	No. of TSGs Detected among Fusions	No. of Fusions Detected by 3 Tools	No. of Fusions Detected by deFuse	No. of Fusions Detected by FusionCatcher	No. of Fusions Detected by TopHat	No. of Inter-chromosomal Fusions	No. of Intra-chromosomal Fusions	Somatic Non-synonymous Mutations	Somatic Non-synonymous Mutation Rate (Mb)	Genomic Breakpoints
Nephroblastoma	O	53	0	3	2	46	9	30	33	20	600	13.04	NA <sup>c</sup>
Sertoli Leydig granulosa tumor	O	41	2	2	2	34	6	27	27	14	91	1.98	77
Astroblastoma	B	40	2	4	1	35	9	19	21	19	45	0.98	14
ASPS	S	53	3	3	1	47	4	22	33	20	1,030	22.39	22
Nephroblastoma	O	64	3	4	1	57	0	32	34	30	81	1.76	56
PNET	B	49	2	2	1	40	3	31	36	13	794	17.26	16
Hepatoblastoma	O	33	2	2	0	29	8	13	18	15	49	1.07	41
eRMS	S	49	1	4	3	44	12	27	27	22	66	1.43	48
FL-HCC	O	59	2	6	3	52	13	34	33	26	38	0.83	55
eRMS	S	23	1	1	0	19	4	12	16	7	50	1.09	31
Ewing sarcoma	S	46	4	2	3	40	6	22	28	18	66	1.43	18
Ewing-like sarcoma	S	31	1	2	1	28	6	18	21	10	78	1.70	23
eRMS	S	48	5	2	3	45	12	26	24	24	58	1.26	38
aRMS	S	52	2	3	3	47	13	25	28	24	65	1.41	29
Epithelioid sarcoma	S	51	2	5	3	47	12	28	26	25	55	1.20	94
Medulloblastoma	B	153	6	19	24	144	33	115	78	75	142	3.09	398
Osteosarcoma	S	92	2	6	7	83	18	55	45	47	79	1.72	99
Ependymoma	B	65	5	6	3	62	2	35	32	33	39	0.85	22
SETTLE	O	23	0	1	0	19	3	12	13	10	505	10.98	15
eRMS Li-Fraumeni	S	58	0	10	7	45	15	46	33	25	68	1.48	173
Ependymoma	B	93	3	6	14	84	37	58	40	53	80	1.74	213
HGG	B	51	5	7	1	44	10	23	25	26	470	10.22	19
DIPG	B	69	6	1	4	56	10	47	43	26	49	1.07	29
Hepatoblastoma	O	44	0	5	1	37	9	20	23	21	28	0.61	42
LGG	B	67	3	2	1	59	15	36	33	34	31	0.67	67
Neuroblastoma	O	90	5	8	5	85	23	41	47	43	352	7.65	218
Ependymoma	B	47	2	4	2	40	12	22	19	28	69	1.50	114
aRMS	S	62	3	2	2	56	16	25	29	33	NW <sup>d</sup>	NW <sup>d</sup>	135

(Continued on next page)

Table 1. Continued

Diagnosis <sup>a</sup>	Tumor Type <sup>b</sup>	Total Fusions Detected	No. of Oncogenes Detected among Fusions	No. of TSGs Detected among Fusions	No. of Fusions Detected by 3 Tools	No. of Fusions Detected by deFuse	No. of Fusions Detected by FusionCatcher	No. of Fusions Detected by TopHat	No. of Inter-chromosomal Fusions	No. of Intra-chromosomal Fusions	Somatic Non-synonymous Mutations	Somatic Non-synonymous Mutation Rate (Mb)	Genomic Breakpoints
aRMS	S	37	3	4	2	36	7	25	30	7	107	2.33	119
Medulloblastoma	B	58	2	4	1	52	17	26	24	34	NW <sup>d</sup>	NW <sup>d</sup>	121
HGG (glioblastoma)	B	74	3	7	2	68	21	32	31	43	130	2.83	253
Medulloblastoma	B	48	2	5	3	44	17	23	22	26	115	2.50	163
Chondrosarcoma	S	34	2	1	0	29	6	16	18	16	184	4.00	129
Ependymoma	B	46	3	3	1	39	14	21	21	25	124	2.70	109
eRMS	S	33	4	3	0	29	8	16	12	21	128	2.78	164
Neuroblastoma	O	44	9	1	0	23	6	35	12	32	137	2.98	116
Medulloblastoma	B	50	3	4	1	48	15	20	19	31	182	3.96	169
aRMS	S	83	3	9	5	76	16	39	38	45	72	1.57	143
DSRCT	S	15	2	2	1	12	5	7	7	8	105	2.28	94
Osteosarcoma	S	82	3	7	12	70	25	59	40	42	126	2.74	227
MNST	O	25	0	3	2	20	5	14	18	7	128	2.78	136
Medulloblastoma	B	50	1	1	0	42	13	30	17	33	41	0.89	237
Abrikossoff tumor	O	36	2	2	1	31	7	11	15	21	81	1.76	35
Neuroblastoma	O	80	0	11	0	71	15	43	40	40	221	4.80	322
Medulloblastoma	B	3	0	0	0	2	0	2	2	1	70	1.52	342
Ewing's sarcoma	S	41	2	0	1	39	7	16	21	20	45	0.98	73
eRMS	S	25	1	1	3	19	5	20	17	8	92	2.00	261
Glioblastoma	B	223	15	23	18	199	61	159	102	121	125	2.72	423

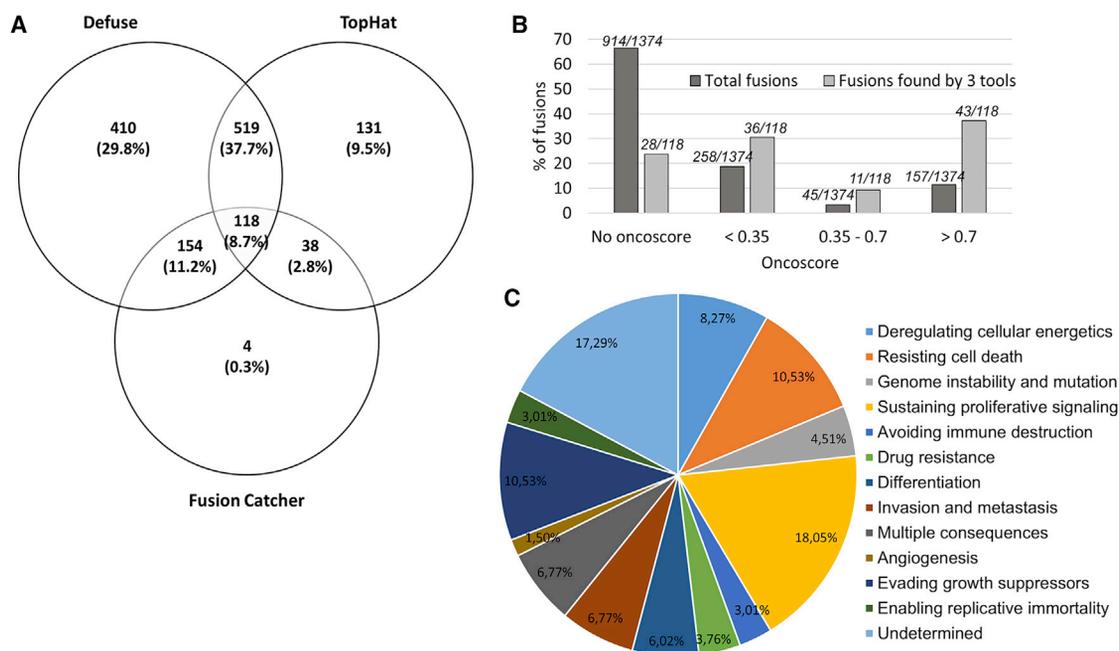
48 pediatric patients were enrolled in the previously described molecular profiling program MOSCATO-01 conducted at Gustave Roussy. Sequencing data and clinical record of these patients were used for our study. For each patient, identification number, associated diagnosis, and tumor type are reported. Fusions metrics and type of chromosomal rearrangements were also assessed thanks to ChimComp pipeline. Finally, the associated mutational rate and breakpoint numbers were calculated from WES and CGH array, respectively.

<sup>a</sup>ASPS, alveolar soft part sarcoma; PNET, primitive neuroectodermal tumor; eRMS, embryonal rhabdomyosarcoma; FL-HCC, fibrolamellar hepatocellular carcinoma; aRMS, alveolar rhabdomyosarcoma; SETTLE, spindle epithelial tumor with thymus-like differentiation; HGG, high-grade glioma; DIPG, diffuse intrinsic pontine glioma; LGG, low-grade glioma; DSRCT, desmoplastic small-round-cell tumor; MNST, malignant nerve sheath tumor; TSGs, tumor suppressor genes.

<sup>b</sup>S, sarcoma; B, brain tumors; O, others.

<sup>c</sup>NA, not able to be calculated.

<sup>d</sup>NW, no WES performed.



**Figure 2. Analysis of the Fusion Transcripts Found by ChimComp Pipeline**

(A) Venn diagram of the non-redundant 1,374 fusions detected from RNA-seq data of 48 pediatric patients analyzed by 3 tools (deFuse, TopHat, and FusionCatcher). (B) Repartition of fusions according to oncoscore. The dark gray bars represent the total fusions ( $n = 1,374$ ) and the light gray bars represent the high-confidence group found by 3 tools ( $n = 118$ ). (C) Distribution of predicted biological alterations of the high-confidence candidate group. Classification of the fusion transcripts was achieved according to the hallmarks of cancer described by Hanahan and Weinberg.<sup>21</sup>

(1.8 mutations/Mb)<sup>14</sup> (Table 1) and a mean number of chromosomal breakpoints of 670 per patient. To link the mutational rate and breakpoint number to the number of fusions, we used Spearman correlation coefficient at  $\alpha = 5\%$ . No correlation was found between the number of somatic mutations and fusions (Figure 3A) (Spearman coefficient  $r = 0.027$  and  $p = 0.857$ ) or between the number of breakpoints and fusions (when discarding the outlier patient number 48, Spearman coefficient  $r = 0.246$ ,  $p = 0.099$ ) (Figure 3B).

#### In Vivo Validation Identified a Set of De Novo Fusion Transcripts

To validate our approach for fusion transcription identification, we aimed at checking by qRT-PCR a list of 42 fusion transcripts chosen from among the high-confidence set (detected by the 3 tools) and among fusions of lower confidence (Table 2). For this, we randomly chose a set of 35 predicted fusion transcripts in addition to 7 fusion transcripts previously identified.<sup>25–30</sup> Interestingly, 38/42 (90.5%) fusion transcripts were successfully validated in patient biopsies, including the 7 known cases. Among the transcripts identified, only 14.4% displayed a low mean spanning read number ( $<5$ ), strengthening the reliability of the high-confidence group defined. Remarkably, when we considered the high-confidence set tested cases (29 fusions tested in total, 13 with high and intermediate oncoscore and 16 with low or no oncoscore), 100% of fusions with high and intermediate oncoscore (oncoscore  $> 0.35$ ) were validated (13/13), whereas in the group with low or no oncoscore ( $<0.35$ ), 3/16 (18%) could not be confirmed by qRT-PCR (*SAP30BP-CDRT4*, *ABR-DNAH2*, and

*C15orf57-CBX3*; Table 2, top). *C15orf57-CBX3*, observed in 21 patients with a low oncoscore, was not validated, thereby most likely representing a false positive. The other 2 transcripts that could not be confirmed had  $>30$  spanning reads but a low or no oncoscore (*ABR-DNAH2* and *SAP30BP-CDRT4*, respectively; Table 2).

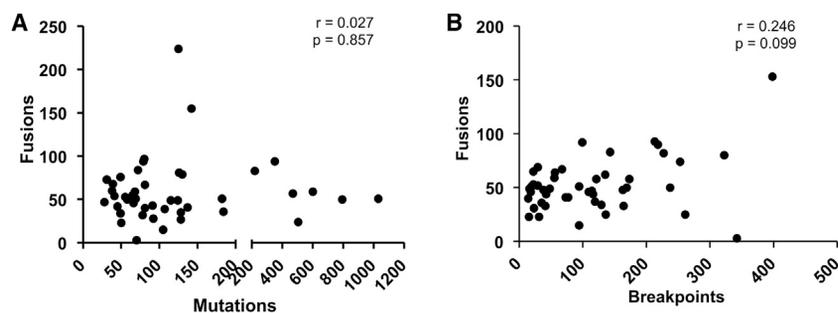
In addition to the high-confidence candidates, 8 fusions detected by less than 3 tools and harboring a high oncoscore, including one control healthy fusion (*LPP-TPRG1*), were also tested by qRT-PCR. Seven of the eight tested fusions (88%) were successfully validated (Table 2, middle).

The fact that 100% of high-confidence fusions exhibiting a high oncoscore were validated positively indicates that our approach of intersecting call from different fusion transcript algorithms coupled with a filtering based on oncoscore is a successful procedure to detect valid fusion transcripts.

#### Fusion Oncogenes Involving TK Domains Are Mainly Detected in Brain Tumors

Targetable and actionable fusions generally include tyrosine kinase (TK) genes. We checked for the presence of TK genes among the list of 1,374 fusions identified.

Only 5/1,374 fusions (0.36%) involved known TKs (Table S5). They corresponded to *FES-MAN2A2*, *CCM2L-HCK*, *JAK2-PTPRD*,



**Figure 3. Relationship between Fusions and Mutations or Genomic Breakpoints in Relapsed Pediatric Cancer Patients**

The Spearman correlation coefficient was calculated to assess (A) the relation between total fusion number (y axis) and mutations (x axis) and (B) the relation between total fusion number (y axis) and genomic breakpoints (x axis). No correlation was found between the number of breakpoints and fusions (B). *r*, Spearman correlation coefficient; *p*, *p* value of Spearman test.

*MANBAL-SRC*, and *TYRO3-KLHL18* (Figure 4), and they were mainly detected by 2 tools, except for *JAK2-PTPRD*, detected by 3 tools and belonging to the high-confidence set. All those fusions were validated by qRT-PCR followed by Sanger sequencing (Table 2). Of interest, they were mainly detected in brain tumors and with a high oncoscore (>0.7), except for *FES-MAN2A2*, which was found in a neuroblastoma and had no oncoscore. Moreover, within the same patient, these fusion oncogenes were found to have the highest oncogenic score compared to the other fusions (Table S5).

#### Fusions of Potential Therapeutic Interest with Currently Available Drugs

19 fusions belonging to the high-confidence list and affecting 33% (16/48) of patients were identified as potentially targetable or contributing to drug resistance, being thus important for drug choice in the individual patient (Table 3). 15 of 19 fusions had a very high oncoscore (>0.89). The total list contains four well-known disease-defining transcripts as well as two others that have been listed in fusion databases before. The remaining 13 transcripts constitute individual events, occurring in 8 patients.

Five transcripts (5/19) occurred in patient 16 with a highly rearranged medulloblastoma, three of which concerned the ERBB-signaling axis through different mechanisms, including two via *MGMT*. Further, two patients with neuroblastoma (patient 26) and alveolar rhabdomyosarcoma (patient 14), presented two potentially targetable transcripts, respectively. Eight of 19 transcripts (42%) were part of the validation study and were thus confirmed, including 6 newly identified fusions that were not described before.

Potentially druggable effects for these 19 transcripts include the following: (1) changes in drug sensitivity (e.g., *MGMT* and temozolomide; transcriptional overexpression of *ABCBI*, leading to multidrug resistance), (2) overexpression and/or overactivation of receptor TKs and downstream pathways activation (e.g., *MET* upregulation by chimeric transcript factor generated by *ASPSCR1-TFE3*; overexpression of ERBB receptor family ligands as *NRG2*), and (3) synthetic lethality (e.g., inhibition of *WEE1* TK in *TP53*-deficient cells). The classification according to the biological function of each oncogenic fusion transcript was done in regard to the literature (J. Fazal-Salom et al., 2017, *Cancer Res.*, abstract).<sup>31–49</sup>

Further, a patient with osteosarcoma (patient 17) was selected for validation of 3 potentially targetable fusions based on the presence of genes that, with their related pathways, are strongly linked to osteosarcoma biology (*TP53-ZNF565*, *UBE4B-CTNNBIP1*, and *NF2-KIAA0368*). These fusion transcripts had an oncoscore > 0.9, and, interestingly, they were found by three, two, or just one tool. They were also successfully validated by qRT-PCR (Table 2), and they are discussed in the Supplemental Materials and Methods.

#### DISCUSSION

With the aim to discover new fusion oncogenes in pediatric patients, we carried out a retrospective study in a cohort of relapsed or resistant patients included in the MOSCATO-01 clinical trial.<sup>14</sup> The in-house analysis package ChimComp was created to reliably detect fusion transcripts in a unique cohort of pediatric patients with solely solid tumors.

We primarily concentrated on fusion candidates from the high-confidence group detected by three tools representing 118/1,374 fusion transcripts (8.6%). To differentiate potential driving from passenger fusions, they were classified into four groups according to their oncoscore values. Interestingly, among them a misbalance in favor of high or low oncoscore (53.8% and 37%, respectively) was observed when compared to the distribution of all detected fusions. Remarkably, 100% of the selected candidates with high (>0.7) and intermediate (>0.35–0.7) oncoscores could be validated by qRT-PCR, whereas 19% of selected fusion transcripts with low or no oncoscore could not. This observation seems to hold also for the group of fusions found by less than three tools with a validation rate of 87.5%. However, it must be considered that the selection of these fusions was mainly driven by biological interest and did not include fusions with a low oncoscore or no oncoscore, except the *LPP-TPRG1* transcript, which is known in healthy tissues.

We observed a tendency toward successful validation of candidates with a high oncoscore rather than high read numbers. This observation is in line with emerging evidence that a fusion's true existence cannot solely rely on the number of reads supporting it, this number being dependent on mapping parameters and sequencing depth.<sup>50</sup> However, we noticed certain limits of Oncofuse: (1) an inability to score the majority of fusion transcripts (herein 914 fusions, 66%), (2) a variation of oncoscore according to the orientation of both

**Table 2. Description, Recurrence, Validation, and Function Prediction of the Fusion Transcripts Tested by qRT-PCR after Detection by ChimComp on Tumor RNA-Seq**

Fusion Gene 1	Fusion Gene 2	Type of Rearrangement	No. of Tools	No. of Patients Carrying Fusion	Patient No.	Validated by PCR	Oncoscore	No. of Databases Reporting the Fusion	Read No. (mean)	Forward Primer Sequence (5'-3')	Reverse Primer Sequence (5'-3')	Oncogene	Tumor Suppressor Gene	Functional Prediction
High-Confidence Candidates Found by 3 Tools														
CDKN2A	ITCH	inter	3	1	41	yes	-	0	23	GGTTTT CGTGGT TCACATC	ATTGCT CCCTGGT CTTCTAG	-	-	inhibition of ubiquitination and loss of CDKN2A tumor suppressor gene
NF2	AK131325	intra	3	1	40	yes	-	0	74	GAGCTT CAGCTC TCTCAAG	CCACAGC ATAACT TCCAAC	-	NF2	loss of NF2 tumor suppressor, negative regulator of cell cycle and apoptosis
SAP30BP	CDRT4	intra	3	1	21	no	-	0	111	ACAGTAG ACAGTCG GAAGATG	TGAGAG GTATAG GTGACA TAGG	-	-	inhibition of cell death
SEMA3D	ABCB1	intra	3	1	26	yes	-	0	8.0	CCCTCTC CCTGAA AAGG	CGGATTG ACTGAA TGCTG	-	-	increased MDR (multidrug resistance)
TET1	TTC12	inter	3	1	21	yes	-	0	22	CAGAGTT TGGCTAC ACGATTAG	GTGTTAG GGGTAAC AAGTTTGG	-	-	aberrant DNA methylation
ABR	DNAH2	intra	3	1	40	no	0.00	0	94	CGGGAAA CAGGGTA GAAAGG	AGCGGGG GAAAATA TGTCG	-	-	may have an influence on cell cycle and activation of protein kinases
FLVCR1	TATDN3	intra	3	1	17	yes	0.01	0	150	TGCAGCA GCATCTC TTCTG	CGTGAGT GCACATT TACAGG	-	-	may have an influence on cell metabolism or differentiation
VPS53	NXN	intra	3	1	30	yes	0.03	0	168	ATCTCTG GCGAAC ATAGACG	GTTCTCG CACACAA CCTTC	-	VPS53	activation of Wnt signaling involved in cell growth and differentiation
FBXL7	TERT	intra	3	1	2	yes	0.11	0	51	ATCACAC GCCCACT AAAGC	CGACGAC GTACACA CTCATC	-	-	loss of cell-cycle arrest by ubiquitination of Aurora kinases and overactivation of telomerase
MAMLD1	BEND2	intra	3	1	3	yes	0.11	0	209	GCTGTTG CCATCT GTTTGC	GCCTTCA CCACCAT CTCTG	-	-	chromatin restructuring
TFG	GPR128 <sup>a</sup>	intra	3	1	18	yes	0.15	6	152	ACTTGGG GAGGAT ATTCGG	TGTAGGG GTGCTTG ATGAG	TFG	-	found in healthy individuals
KIAA1324L	TMEM243	intra	3	1	16	yes	0.19	0	17	CGGTGTC TTTGATC TATGC	ACGGTTT TGGAGGT AGTTG	-	-	may be involved in drug resistance

(Continued on next page)

Table 2. Continued

Fusion Gene 1	Fusion Gene 2	Type of Rearrangement	No. of Tools	No. of Patients Carrying Fusion	Patient No.	Validated by PCR	Oncoscore	No. of Databases Reporting the Fusion	Read No. (mean)	Forward Primer Sequence (5'-3')	Reverse Primer Sequence (5'-3')	Oncogene	Tumor Suppressor Gene	Functional Prediction
HMGCLL1	BAI3	intra	3	1	21	yes	0.21	0	88	ACCAGCT TCCATCA CTTG	GTGGTG TTACCC CTGAATG	-	-	may be involved in differentiation in neurons
DNAJC1	RERGL	inter	3	1	47	yes	0.26	0	7	GAGACCT GGAGTTG TTTGAC	CCCGGAT TCTGTAG ATCAG	-	-	may have consequences on protein synthesis
HERC4	TMPRSS15	inter	3	1	8	yes	0.46	0	7.0	GTATTCC CCCTTCT GAAAG	AGACCCT GGCATA ACTC	-	-	inhibition of ubiquitination, stabilization of proteins
IP6K1	RYBP	intra	3	1	13	yes	0.47	0	21.3	CGCAGAC GAAGGG TTTTG	CACGCAG GGGTACT TGAAG	-	-	increased kinase activity or apoptosis resistance
DMTF1	ABCB1	intra	3	1	15	yes	0.89	0	8.0	GGACCCA AAACAGG AAAGAG	CGAGAA ACTGCG AAACAGG	-	DMTF1	increased MDR (multidrug resistance)
NRG2	ZNF208	inter	3	1	16	yes	0.92	0	25.0	CAACGGC AGAAAGA ACTC	AGGACAC ACAGCA GTAAGG	-	-	activation of TK receptor
ERRFI1	CUBN	inter	3	1	16	yes	0.95	0	4	GGCACAA TGTCAT AGCAG	GGAAGG ACACAAA AGTCAAG	-	ERRFI1	activation of EGF
TP53	ZNF565	inter	3	1	17	yes	0.96	0	11.5	TCCCATG TGCTCAA GACTG	GTTCCAG GCATTTC CATTC	-	TP53	inactivation of p53 function
EIF2AK1	OGDH	intra	3	1	16	yes	0.98	0	20.0	CTCTTCC CGTTTCT GTG	GGGAGG GCATTCT TCTC	-	-	illegitimate activation of kinase
CTSB	BACE2	inter	3	1	11	yes	0.98	0	129.7	CCACCAT CAAAGA GATCAG	CCACCAA GACAAGA CTAC	-	-	degradation of extracellular matrix and overexpression of BACE2
MYB	AHI1	intra	3	1	13	yes	0.99	2	5.0	TCTGGAA TTGTTGC TGAGTTTC	TGAGAA GAGGTG TTGTG TGG	MYB, AHI1	-	aberrant proliferation and dedifferentiation of cells through MYB oncogene activation
ATRX	PASD1	intra	3	1	27	yes	1.00	0	20.0	TCGGCAAG TAACTAA GCAG	TGGTCG TCTGAG ATAGCAG	-	-	stabilization of DNA replication via chromatin remodeling
LMO3	BORCS5	intra	3	1	11	yes	1.00	0	17.0	GCAACCG AAAGATC AAGG	CATTGTG TGGGGA AGTCTG	LMO3	-	over-activation of oncogene LMO3

(Continued on next page)

Table 2. Continued

Fusion Gene 1	Fusion Gene 2	Type of Rearrangement	No. of Tools	No. of Patients Carrying Fusion	Patient No.	Validated by PCR	Oncoscore	No. of Databases Reporting the Fusion	Read No. (mean)	Forward Primer Sequence (5'-3')	Reverse Primer Sequence (5'-3')	Oncogene	Tumor Suppressor Gene	Functional Prediction
C11orf95	RELA <sup>a</sup>	intra	3	2	21, 27	yes	1.00	1	119.0	GTCCCCA CCAGAA CTG	CCCTCGC ACTTGT AGC	-	-	aberrant NF-kappaB transcription program
EWSR1	FLI1 <sup>a</sup>	inter	3	2	46, 11	yes	1.00	7	93.7	AGTTACCC ACCCCAA ACTGG	CCAAGGG GAGGACT TTTGTT	EWSR1, FLI1	-	regulates the expression of a number of genes important for cancer progression
KANSL1	ARL17A <sup>a</sup>	intra	3	9	40, 47, 5, 6, 13, 18, 20, 21, 23	yes	-	2	72	GGCTCATG TTTCTGAC TTG	GACTTC TACCTC CTTTCCAC	-	-	cancer predisposition, linked to histone acetylation
C15orf57	CBX3 <sup>a</sup>	inter	3	21	38, 40, 41, 43, 47, 48, 2, 12, 14, 16, 17, 18, 20, 22, 23, 25, 27, 28, 31, 32, 34	no	0.01	3	16	TCCTGTC TGCCAA ATCCTG	GGAAAG CCACCG ACTTCAC	-	-	aberrant epigenetic gene repression
Fusions Found by Less Than 3 Tools with Biological Relevance														
LPP	TPRG1 <sup>a</sup>	intra	2	1	15	yes	-	2	12.5	TCCTCTGA TTCCCTG TGTC	CCTTCT CCTTGTC TCCTGTC	-	-	healthy
RASSF4	ZNF22	intra	1	1	13	yes	0.86	0	41.0	GAGTACC CGCTGAT TTCC	CGCTTTA GGCTTT GCTAAC	-	RASSF4	disruption of RASSF4 tumor suppressor, which promotes apoptosis and cell-cycle arrest
CUBN	SMAP2	intra	2	1	16	no	1.00	0	3.5	AGACTGT TCTTGGG TCATTC	CCTTTA GACTGG CAATCTG	-	-	propable deregulation in cell metabolism
MPP3	PYY	intra	2	1	17	yes	0.77	0	1.5	AGAACC CACAGC ACTTTG	TGACTTC ACTCCAC CAATG	-	-	increased cell proliferation
UBE4B	CTNNBIP1	intra	2	1	17	yes	0.90	1	5.3	GACCGA GTTGGA ATAGAGG	GGAAAC GAGGAA TCTTAGG	-	UBE4B, CTNNBIP1	loss of UBE4B and CTNNBIP1 tumor suppressor genes
NF2	KIAA0368	inter	1	1	17	yes	0.99	0	16.5	AACTCTC TAGCACC CAAGAAG	TAAGCC TGACAC TGAAATCC	-	NF2	loss of NF2 tumor suppressor gene

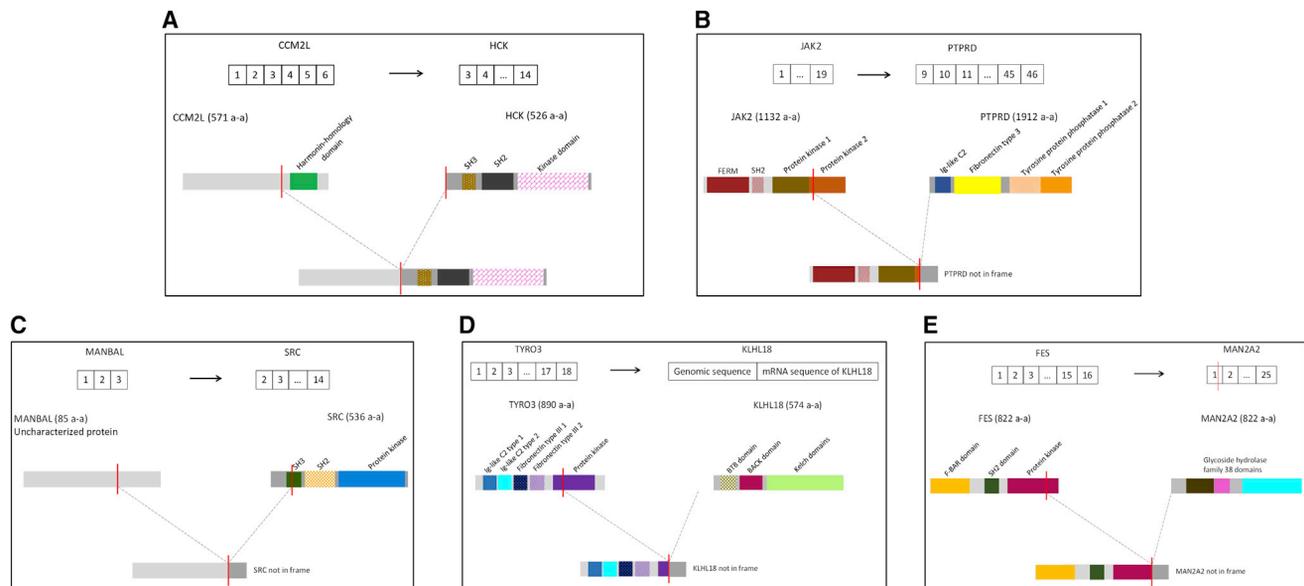
(Continued on next page)

Table 2. Continued

Fusion Gene 1	Fusion Gene 2	Type of Rearrangement	No. of Tools	No. of Patients Carrying Fusion	Patient No.	Validated by PCR	Oncoscore	No. of Databases Reporting the Fusion	Read No. (mean)	Forward Primer Sequence (5'-3')	Reverse Primer Sequence (5'-3')	Oncogene	Tumor Suppressor Gene	Functional Prediction
ARHGGEF7	MYO16	intra	2	1	32	yes	1.00	0	4.0	CAGACTG TGCTTTC AACG	AAGGGGA AGGGACT CTAG	-	-	loss of ARHGGEF7, a Ras-like family of Rho proteins acting as a pro-apoptosis factor
HEY1	NCOA2 <sup>a</sup>	intra	2	1	33	yes	1.00	4	72.5	TGAGTTC GGCTCT AGGTTC	TGATTGG TGGGAA AGGTC	-	-	aberrant expression of genes
Fusions with TK Coding Gene														
TYRO3 <sup>b</sup>	KLHL18	inter	2	1	18	yes	0.78	0	7	CCAAACT GCCTGTC AAGTG	GGGTAG CAGGGA TGTCTTC	-	-	disruption of TYRO3 domain
JAK2 <sup>b</sup>	PTPRD	intra	3	1	21	yes	1.00	0	22	AAAGAAC CTGGTGA AAGTCC	ACGTCT CCCTTG AGTTAGC	JAK2	JAK2, PTPRD	disruption of PTPRD that regulates cell growth, differentiation, and mitotic cycle
FES <sup>b</sup>	MAN2A2	intra	2	1	26	yes	-	0	11	GACAAGT CCCCGTG AAGTG	TTTCCGT GCCTGC TCTATG	FES	-	disruption of the end of FES kinase domain
CCM2L	HCK <sup>b</sup>	intra	2	1	30	yes	1.00	0	7	GCCTATG ATGCCGA CTTCAG	CCTGGTG TGTTGC TGTTGTG	-	-	probably overexpression of HCK; increase of cell survival, proliferation, and migration
MANBAL	SRC <sup>b</sup>	intra	2	1	48	yes	1.00	0	9	GCTACGG TACGGAC TCTTC	CCTGGC TCTGTC TCTCATAG	SRC	-	SRC kinase not in frame

To validate our approach, we first focused on the fusions from the high-confidence group (detected by 3 tools), then we divided the list into three categories, based on oncoscore results: oncoscore <0.35, 0.35–0.7, and >0.7. Of those, 25% of fusions were selected for validation by qRT-PCR in each category, resulting in 29 fusions in total (top). Additionally, 8 fusions detected by less than 3 tools harboring a high oncoscore, except for one control healthy fusion (*LPP-TPRG1*), were tested by qRT-PCR (middle). Finally, 5 potentially targetable fusions harboring tyrosine kinase-coding genes were also tested (bottom). A total of 42 fusions found by ChimComp underwent qRT-PCR assay. As positive controls, some of the well-known fusion oncogenes detected by ChimComp were also included, namely, *C11orf95-RELA*, *EWSR1-FLI1*, and *HEY1-NCOA2*. A total of 38/42 fusions were successfully validated. For each fusion, the type of chromosomal rearrangement potentially involved, the number of tools detecting the fusion, the number of patients carrying the fusion and their identification numbers, the result of PCR, the oncoscore provided by ChimComp as well as the number of databases reporting the fusion also provided by ChimComp, the primer sequences used, the presence of oncogenes or tumor suppressor genes, and the functional prediction are enlisted. Intra, intrachromosomal; inter, interchromosomal; read number (mean), mean number of junction/spanning reads of all tools.

<sup>a</sup>Known fusions.  
<sup>b</sup>With tyrosine kinase activity.



**Figure 4. Graphical Reconstitution of the mRNA and Protein of 5 Fusions Involving Tyrosine Kinase Genes**

The NCBI BLAST (<https://blast.ncbi.nlm.nih.gov/Blast.cgi>) was used to obtain the putative mRNA sequence, which was reconstituted with the use of nucleotide (NCBI). When the coding sequence was obtained and predicted, Uniprot Knowledge base (<https://www.uniprot.org/>) was used to reconstitute the predicted fusion protein domains. (A) CCM2L-HCK, (B) JAK2-PTPRD, (C) MANBAL-SRC, (D) TYRO3-KLHL18, and (E) FES-MAN2A2 are shown.

genes, and (3) an undervaluation of affected TSGs. Regarding possible reasons for not validating events, this might include alignment errors, showing that the use of meta-caller pipeline alone does not allow one to reach a perfectly reliable analysis. A visualization tool appears to be one of the solutions to support a fusion's real existence at the highest *in silico* confidence level. Our fusion detection method also reached its limits as (1) results would need to be compared to the respective healthy tissue of the individual patient, and (2) shifts in the open reading frames could not be reliably predicted on the bioinformatic level.

Interestingly, as seen for *SAP30BP-CDRT4* and *ABR-DNAH2* fusions, a high read number (>50) did not guarantee biological validation. As highlighted by Gao et al.,<sup>51</sup> the use of 3 algorithms therefore gives more confidence to fusions with lower levels of read support that might otherwise have been discarded. Another important observation that has been made regarding the three non-validated fusions is their high mapping quality (MAPQ > 45), suggesting that this parameter cannot be considered as a criterion for non-validation.

The total number of fusions detected per patient was high. Moreover, we found no significant correlation between the number of fusions and genomic breakpoints, supporting the hypothesis that the observed fusions could be mainly due to balanced events or non-genomic events rather than unbalanced events or copy number changes, as it was described for adult tumors.<sup>1,23</sup> The same studies described a negative correlation between fusions and mutations, which was not observed in our cohort. All these observations could

be explained in part by the particular status of our cohort, enrolling pre-treated patients with resistant tumors having acquired a succession of genomic events to become more aggressive and that might harbor more chaotic genomes compared to diagnosis. In addition, at advanced tumor stages, the clonal diversity and intratumoral heterogeneity is higher, which can distort the analysis.<sup>52</sup>

In this cohort, 57% of all detected transcripts resulted from intra-chromosomal events, with an almost identical ratio in the high-confidence group, whereas 75% of fusions newly discovered by deep sequencing techniques were found to be intra-chromosomal, often resulting from subtle inversions or deletions.<sup>1</sup> Possible explanations could be (1) the exclusion of leukemias in our cohort; (2) the rarity of studies searching for fusion oncogenes, especially in pediatric solid tumors; and (3) the unique biology of pediatric tumors. However, this result should be confirmed at the genomic level to exclude *trans*-splicing mechanisms.

Both oncogenes and TSGs were frequently found in our analysis. Usually, fusion oncogenes are of great interest because of their potential targetability. In our study, more than one-third of events for the high-confidence group was predicted to be in-frame, potentially giving rise to proteins with newly acquired or enhanced oncogenic functions. Most of these fusions present a low oncoscore; nevertheless, the function of newly created oncoproteins should be investigated by functional studies. Remarkably, a considerable proportion of fusion transcripts containing TSGs was detected. This reflects the importance of TSGs in pediatric cancers. In fact, it is estimated that at least 10% of pediatric cancers are caused by hereditary genetic events,

**Table 3. Potentially Targetable Fusion Alterations with Associated Putative Biological Effect and Predicted Impact on Drug Sensitivity**

Patient No.	Gene 1	Gene 2	Fusion Validated by qRT-PCR	Oncoscore	Known Fusions	Presumptive Fusion Reconstruction	Predicted Biological Effect	Additional Genomic Information from WES <sup>b</sup>	Possible Drug	Mechanism	Reference
2	FBXL7	TERT <sup>a</sup>	yes	0.109682436		loss of cell-cycle arrest mechanism by ubiquitination of Aurora kinases (FBXL7) + overactivation of telomerase (TERT)	inactivation of FBXL7 and activation of TERT	-	TERT inhibitor, AURKA inhibitor	maintenance of telomere ends and cell cycle deregulation	31,32,68
4	ASPSCR1 (OG) <sup>a</sup>	TFE3 (OG) <sup>a</sup>	ND	0.954511534	x	known fusion with generation of a chimeric transcription factor, leading to MET upregulation	activation of chimeric transcription factor	-	MET inhibitor	MET upregulation	33
13	MYB (OG) <sup>a</sup>	AHI1 (TSG) <sup>a</sup>	yes	0.988687802	x	loss of the major AHI1 (SH3) domain, MYB coding sequence almost completely retained	inactivation of AHI1	strong focal amplification of MYB and AHI1	IGF1R inhibitor	amplification of this locus has been linked to increased drug sensitivity	34
14	ARHGEF7	ATP11A <sup>a</sup>	ND	0.99049634	x	loss of ATP11A regulation by truncation of long noncoding RNA (lncRNA)	activation of ATP11A	ATP11A amplification	Farnesyltransferase inhibitors	drug resistance	35
14	PAX7 <sup>a</sup>	FOXO1 (OG+TSG) <sup>a</sup>	ND	0.97944093	x	known fusion with generation of a chimeric transcription factor	activation	PAX7, FOXO1 amplification	inhibitors of FGFR2/4, ALK, MET, IGF1R, BET	transcriptional activation of target genes	36,69
15	DMTF1 (TSG)	ABCB1 <sup>a</sup>	yes	0.89356784		loss of complete coding sequence of DMTF and fusion of DMTF promoter to ABCB1	inactivation of DMTF and activation of ABCB1	-	chemotherapy and receptor tyrosine kinase inhibitors	drug resistance	37
16	ERFF1 (TSG) <sup>a</sup>	CUBN	yes	0.951965411		loss of complete coding sequence of ERFF1	inactivation of ERFF1	ERFF1 amplification	EGFR inhibitor	increased drug sensitivity	38
16	ESCO1	CNDP2 (TSG) <sup>a</sup>	ND	0.087563145		loss of only domain (dimerization domain) of CNDP2	inactivation of CNDP2	-	MEK inhibitor	activation of mitogen-activated protein kinase (MAPK) pathway	39

(Continued on next page)

Table 3. Continued

Patient No.	Gene 1	Gene 2	Fusion Validated by qRT-PCR	Oncoscore	Known Fusions	Presumptive Fusion Reconstruction	Predicted Biological Effect	Additional Genomic Information from WES <sup>b</sup>	Possible Drug	Mechanism	Reference
16	ASAP2	MGMT <sup>a</sup>	ND	0.999883057		truncation in DNA methyltransferase domain of MGMT	inactivation of MGMT	–	Temozolomide	increased drug sensitivity	40
16	MGMT <sup>a</sup>	C8orf34	ND	0.933256552		loss of 3/4 of MGMT coding sequence	inactivation of MGMT	–	Temozolomide	increased drug sensitivity	40
16	ZNF208	NRG2 <sup>a</sup>	ND	0.915020583		complete coding sequence of NRG2 is kept, resulting in overexpression of NRG2	activation	–	tyrosine kinase inhibitors targeting ErbB receptors	activation of ErbB receptor tyrosine kinases	41
17	TP53 (TSG) <sup>a</sup>	ZNF565	yes	0.956287477		loss of complete coding sequence of TP53	inactivation	heterozygous deletion of TP53 and ZNF565	inhibitors of WEE1	WEE1 sensitizes cells to DNA-damaging agents via G2 checkpoint inhibition in p53-deficient cells	42,43
21	JAK2 (OG)	PTPRD (TSG) <sup>a</sup>	ND	0.999997433		loss of PTPRD, out-of-frame	inactivation of PTPRD	heterozygous PTPRD deletion	STAT3 inhibitor, AURKA inhibitor, BET inhibitors	PTPRD dephosphorylates STAT3 and destabilizes AURKA and MYCN	44,45
26	ATRX <sup>a</sup>	PASD1	ND	0.997390308		truncation in helicase C domain	inactivation	heterozygous ATRX deletion	PARP inhibitors	PARP overexpression	Fazal-Salom et al., 2017, Cancer Res., abstract
26	SEMA3D	ABCB1 <sup>a</sup>	yes	–		probably overexpression of ABCB1 due to SEMA3D promoter activity or loss of regulatory sequences in 5' UTR of ABCB1	activation	ABCB1 amplification	chemotherapy and receptor tyrosine kinase inhibitors	drug resistance	37
32	MYO16 <sup>a</sup>	IRS2 <sup>a</sup>	ND	0.999756812		loss of all domains of IRS2 and of the last two domains of MYO16	inactivation of IRS2	6-kB amplicon including both genes with high-level amplification	IGF1R inhibitor	amplification of this locus has been linked to increased drug sensitivity	34
40	NF2 (TSG) <sup>a</sup>	AKI131325	yes	–		loss of NF2, which acts as negative regulator of HIPPO/YAP pathway and regulates activity of several receptor tyrosine kinases	inactivation	–	inhibitors of pathways regulated by HIPPO/YAP- pathway including AKT/mTOR inhibitors and FAK inhibitors	pathway activation of major oncogenic pathways	46,47

(Continued on next page)

Table 3. Continued

Patient No.	Gene 1	Gene 2	Fusion Validated by qRT-PCR	Oncoscore	Known Fusions	Presumptive Fusion Reconstruction	Predicted Biological Effect	Additional Genomic Information from WES <sup>a,b</sup>	Possible Drug	Mechanism	Reference
11, 46	EWSR1 (OG)	FLI1 (TSG)	yes	0.999925015	x	known fusion with generation of a chimeric transcription factor	activation of chimeric transcription factor	-	PARP inhibitor, BET inhibitor	widespread transcriptional reprogramming of Ewing sarcoma cells required for proliferation and tumorigenesis	48,49
28, 29, 38	PAX3 (OG) <sup>a</sup>	FOXO1 (OG + TSG) <sup>a</sup>	ND	0.998640978	x	known fusion with generation of a chimeric transcription factor	activation	29: PAX3 amplification	inhibitors of EGFR2/4, ALK, MET, IGF1R, BET	transcriptional activation of target genes	36,69

Fusion transcripts representing potential targetable alterations by existing anti-cancer drugs were found among the high-confidence list candidates and are listed with the associated patient number, the result of PCR validation, the associated oncoscore provided by ChimComp, the presumptive fusion reconstruction and biological effect, the possible drugs to target the fusion products, the mechanism involved, and the references in literature used. Additional relevant genomic information from WES of patients was also added when relevant regarding the associated potentially targetable fusion. ND, not done.

<sup>a</sup>Gene whose alteration induced by the fusion has potential therapeutic implications.

<sup>b</sup>Genomic results from concomitant WES if available (Harttrampf et al.<sup>14</sup>).

mostly TSG inactivation.<sup>53,54</sup> Furthermore, the contribution of fusions leading to TSG inactivation might be generally underestimated.

Fusion oncogenes containing TK occurred at low incidence and predominantly were found in patients with brain tumors, involving mostly non-receptor TKs. Whereas for certain molecular subgroups of high-grade gliomas genetic aberrations affecting receptor TKs such as *PDGFRA*, *MET*, *EGFR*, *FGFR*, and *NTRK* have been well defined, this has not been found in the extensive molecular characterization of medulloblastomas.<sup>55</sup> For ependymomas, *EGFR* overexpression is common but only seldom related to copy number changes. *EGFR*-containing gene fusions have been reported in a subset of ependymomas.<sup>56</sup>

Given the therapeutic potential of a TK-containing fusion, we selected these 5 fusion oncogenes for validation and sequencing. The only in-frame transcript that we confirmed by sequencing was *CCM2L-HCK* transcript in patient 30 with medulloblastoma, who lacked aberrations in *TP53* and genes of *WNT* and *SHH* pathways. The complete coding sequence of *HCK* is preserved in the chimeric transcript. *HCK* belongs to the *SRC* family of non-receptor protein TKs, and increased *HCK* activity is capable of increasing proliferation and cell survival as well as associating with receptor TKs, such as *EGFR*, *FGFR*, and *PDGFR*, to amplify their oncogenic potential.<sup>57</sup> *HCK* itself is targetable by small molecule kinase inhibitors such as dasatinib, bosutinib, and saracatinib.

The sequencing of the other 4 fusions containing TK proved that the *MANBAL-SRC* fusion (patient 48) was out-of-frame. Of note, this patient with glioblastoma had several driving pathogenic mutations affecting *PI3KCA*, *NF1*, and *FGFR1* next to amplifications of *MDM2* and *CDK4* (already described by Harttrampf et al.<sup>14</sup>). The *FES-MAN2A2* (patient 26) and the *TYRO3-KLHL18* transcripts (patient 18) led to a disruption of TK domains, and, therefore, were of no therapeutic interest. The *JAK2-PTPRD* transcript identified in patient 21 with ependymoma truncates *JAK2* in the regulatory pseudo-kinase domain, with subsequent loss of the C-terminal TK domain. However, this leads to a shift of the reading frame of *PTPRD*, a gene whose role as tumor suppressor has been documented in many malignancies, including neuroblastoma and glioblastoma.<sup>45,58</sup> In these pathologies, loss of *PTPRD* activates *AURKA* and *STAT3*, respectively. Importantly, both *AURKA* and *STAT3* inhibitors are currently being investigated in clinical trials, rendering downstream effects of tumor suppressor inactivation as an attractive therapeutic target. Ortiz et al.<sup>58</sup> further provided *in vivo* proof that loss of *PTPRD* cooperates with *CDKN2A/B* deletion, which was also found in our patient, and promotes tumorigenesis even if *PTPRD* is heterozygously inactivated.

Next we extracted fusions from the high-confidence group that might be also therapeutically relevant. One-third of analyzed patients displayed a fusion transcript with potential druggability or involved in drug resistance to certain agents, three of which presented more than one of those. To which extent these fusions might have been present at diagnosis or have evolved over time under current treatment

regimens was not investigated, but they might be important contributors to disease progression. However, data correlated to putative effects exerted by the fusion are often preclinical and result from the analysis of other often adult tumors, as, for example, 2 transcripts that we confirmed in 2 different patients (*MYO16-IRS2* in patient 32 and *MYB-AH11* in patient 13), the latter of which has already been described in acute lymphoblastic leukemia and breast cancer.<sup>23,59</sup> Strong amplification of these loci has been linked to *in vitro* sensitivity to IGF1R inhibition.<sup>34</sup> These fusions were associated with a high-level focal amplification, further supporting a fusion event<sup>1,23</sup> and potentially sensitizing the tumor toward IGF1R inhibition that was not applied to the patients in the absence of an available drug.

Also interesting regarding potential targetability was transcript *CDKN2A-ITCH* (Table S5) in patient 41 with a malignant peripheral nerve sheath tumor. Next to its causal relation to neurofibromatosis type 1, recurrent genetic aberrations affecting tumor suppressor *CDKN2A* have been well documented in this malignancy.<sup>60</sup> This transcript was validated and sequenced and corresponded to p14<sup>ARF</sup>, which acts as a stabilizer of TP53 by sequestering MDM2. The breakpoint leads to the inclusion of exon 1 in the final putative protein, which contains the MDM2-binding domain. It has been experimentally shown that this proportion of the gene alone is sufficient to sustain its full TP53-stabilizing capacity, questioning the driving potential of this particular transcript.<sup>61–63</sup> We also validated three fusions with high oncoscore (*TP53-ZNF565*, *UBE4B-CTNBP1*, and *NF2-KIAA0368*) in one patient with osteosarcoma, because their involved genes were linked to the biology of the patient's cancer. To what extent events like these could be causative, occur due to antineoplastic treatment, occur due to progression, or are random bystanders of multiple copy number aberrations still needs to be defined.

In conclusion, we established a robust pipeline called ChimComp to detect fusion transcripts in cancer patients with 90.5% reliability and close to 100%, if fusions with a very low oncoscore were omitted. This might be especially useful in large screenings, as fusions that are found by three tools and are primarily provided with a high oncoscore most likely represent true events in a patient's tumor and can be considered for further investigation. However, before orientating a patient's treatment or initiating dedicated research projects on a single fusion event, experimental validation is indispensable. The knowledge of fusion genes' function and their involvement in different signaling pathways can provide new therapeutic options for patients and, therefore, add significantly to personalized treatments. Also, fusion transcripts occurring in a single patient, so-called private events, should not be primarily discarded.

## MATERIALS AND METHODS

### Patients

The cohort comprised 48 pediatric patients with 21 different relapsed or refractory solid neoplasms, which had undergone RNA-seq as part of the Molecular Screening for Cancer Treatment Optimization

genomic profiling program (MOSCATO-01, ClinicalTrials.gov: NCT01566019) run at Gustave Roussy.<sup>14,64</sup>

### Whole-Exome and RNA-Seq

Extraction of tumor DNA, RNA, and germline DNA, as well as library preparation, capture and WES, RNA-seq (detailed in the [Supplemental Materials and Methods](#)), and CGH array, was performed as part of the MOSCATO-01 clinical trial, as previously published.<sup>14</sup> In the same study, the mutational burden per patient was also calculated from WES data. Bioinformatic analysis was done by IntegraGen and Gustave Roussy Bioinformatics Core.

The total number of genomic breakpoints per patient was calculated from the results of the CGH array, which were converted into .cbs files containing the genome position of the single segments, the probes, and Log2 ratio value.

Raw data provided from RNA-seq are available on the European Genome-phenome Archive website (<https://www.ebi.ac.uk/ega/home>) under the accession number EGAS00001003236.

### Fusion Detection by ChimComp Pipeline

To optimize the detection of potential fusion transcripts, a new analysis pipeline called ChimComp was developed by Gustave Roussy Bioinformatics Core (Figure 1 details the single steps). ChimComp relies on 3 tools: TopHat version (v.)2.0.14<sup>17</sup> followed by TopHat post, deFuse v.0.6.2,<sup>15</sup> and FusionCatcher.<sup>16</sup> The strength of ChimComp is to reconcile fusions predicted by different algorithms from the same data in order to classify and annotate these predictions. The non-filtered output from deFuse (file "results.classify.tsv") was used.

The 3 tools give a listing of fusion genes obtained from the raw data of RNA-seq. A fusion is listed if (1) the anchor length (number of nucleotides overlapping each side of the breakpoint in spanning reads) is  $\geq 10$  nt; (2) when a fusion is detected by 1 tool, the minimum number of supported read pairs is 40; if found by  $>1$  tool, no read threshold is applied; and (3) at least one of the genes has a non-intergenic breakpoint. Then, for each fusion detected, an oncogenic potential score was computed by using the Oncofuse algorithm.<sup>12</sup> This results in a score between 0 (non-oncogenic) and 1 (highly oncogenic). In our study, it was calculated for both gene orientations and the highest oncoscore value was considered. The detected fusions were annotated with different fusion gene databases comprising Mitelman Database of Chromosome Aberrations and Gene Fusions in Cancer (<https://cgap.nci.nih.gov/Chromosomes/Mitelman>), TICdb,<sup>65</sup> COSMIC (<https://cancer.sanger.ac.uk/cosmic>), ChimerDB,<sup>66</sup> and ChiTaRS.<sup>67</sup> All the fusions identified among the cohort were tested for the presence of TKs, oncogenes, or TSGs by comparing our listing of genes to a set of 90 known TK genes, 803 known oncogenes, and 1,017 known human protein-coding TSGs<sup>18–20</sup> (Figure 1).

### Selection and Validation of Detected Fusions by ChimComp

To validate our approach, 42 fusions found by ChimComp underwent real-time qRT-PCR assay. We primarily focused on the fusions

that were found by three tools (referred to as the high-confidence group), then divided the list into three categories, based on oncoscore results (oncoscore <0.35, 0.35–0.7, and >0.7). Of those, 25% of fusions were selected for experimental validation by qRT-PCR in each category, resulting in 29 fusions in total. As a positive control, some of the well-known fusion oncogenes detected by ChimComp were also included (n = 3). Moreover, 8 fusions detected by less than 3 tools harboring a high oncoscore, except for one control healthy fusion (*LPP-TPRG1*), were tested by qRT-PCR. Finally, 5 potentially targetable fusions harboring TK-coding genes were also validated. It should be noticed that the fusions were selected according to the relevance of genes involved in the fusion and the availability of the biological material.

#### Reconstitution of the Fusion Sequence and Primer Design

The breakpoint position in mRNA sequences of both genes was predicted *in silico* (see the [Supplemental Materials and Methods](#)). The sequence around the breakpoint was used to design the primers necessary for qRT-PCR validation. Primers were selected with a size of about 20 nt, a melting temperature of 60°C, and an amplicon size of 100–300 bp, using Oligo Explorer 1.1.0 program (Kuopio University, Kuopio, Finland).

#### Biological Validation by Real-Time qRT-PCR

The same RNA used for RNA-seq was used to evaluate the presence of the different fusions detected. To verify the respective size of each amplicon, 10 µL of the amplified products was mixed with 2 µL gel loading dye (New England Biolabs, Evry, France) and loaded into 2% agarose gel mixed with DNA or RNA stain Midori Green Advance (Nippon Genetics Europe, Dueren, Germany). Then the gel was visualized under UV using Ingenius bioimaging system (Syngene, Cambridge, England). PCR products were sequenced after DNA extraction with NucleoSpin Gel and PCR clean-up extraction kit (Macherey-Nagel, Dueren, Germany) by Sanger technique (Eurofins, Cochin Institute, Paris, France).

#### Statistical Analysis

The numbers of fusion transcripts found in patients are presented as mean ± interquartile range. To compare the number of fusions with mutational load or genomic breakpoint number per patient, Spearman correlation coefficient was determined. To compare the average number of fusions per patient between the different groups of tumor types, Kruskal-Wallis followed by Dunn's tests were used. To compare the percentage of total fusion versus fusions found by 3 tools in the 4 categories of oncoscore, the Z score test was used. A Z score < 1.96 was considered as significant. All tests were performed by using GraphPad Prism 4 software; p values and adjusted p values < 0.05 were considered as a statistically significant level.

#### SUPPLEMENTAL INFORMATION

Supplemental Information includes five tables and Supplemental Materials and methods and can be found with this article online at <https://doi.org/10.1016/j.ymthe.2018.10.022>.

#### AUTHOR CONTRIBUTIONS

C.D. and A.C.H. designed and carried out the experiment and wrote the manuscript, with support from P.K., B.G., and L.M.-M. Y.B., M.L., W.R., G.R.-S., and P.K. performed bioinformatic studies and analyzed the data. B.G. contributed in the conception of the study and ensured the transfer of the clinic to the laboratory. L.M.-M. supervised the project and was in charge of overall direction and planning of the study. All authors provided critical feedback and helped shape the research, analysis, and manuscript.

#### CONFLICTS OF INTEREST

The authors declare no competing interests.

#### ACKNOWLEDGMENTS

We are grateful to all patients and their parents for participating in the trial. We acknowledge Catherine Richon (technical assistance), Damien Drubay (biostatistics platform of Gustave Roussy Institute), as well as Marc Gabriel and Bastien Job (bioinformatics core of Gustave Roussy). This study was supported by a public grant overseen by the French National Research Agency (ANR) as part of the “Investissements d’Avenir” program (Labex NanoSaclay, ANR-10-LABX-0035), the Fondation Gustave Roussy, the Fondation Philantropia (ACH), the Université Paris-Sud (CD), and the CNRS.

#### REFERENCES

- Mertens, F., Johansson, B., Fioretos, T., and Mitelman, F. (2015). The emerging complexity of gene fusions in cancer. *Nat. Rev. Cancer* 15, 371–381.
- Duesberg, P.H. (1987). Cancer genes generated by rare chromosomal rearrangements rather than activation of oncogenes. *Med. Oncol. Tumor Pharmacother.* 4, 163–175.
- Druker, B.J., Talpaz, M., Resta, D.J., Peng, B., Buchdunger, E., Ford, J.M., Lydon, N.B., Kantarjian, H., Capdeville, R., Ohno-Jones, S., and Sawyers, C.L. (2001). Efficacy and safety of a specific inhibitor of the BCR-ABL tyrosine kinase in chronic myeloid leukemia. *N. Engl. J. Med.* 344, 1031–1037.
- Nagasubramanian, R., Wei, J., Gordon, P., Rastatter, J.C., Cox, M.C., and Pappo, A. (2016). Infantile Fibrosarcoma With NTRK3-ETV6 Fusion Successfully Treated With the Tropomyosin-Related Kinase Inhibitor LOXO-101. *Pediatr. Blood Cancer* 63, 1468–1470.
- Singh, D., Chan, J.M., Zoppoli, P., Niola, F., Sullivan, R., Castano, A., Liu, E.M., Reichel, J., Porrati, P., Pellegatta, S., et al. (2012). Transforming fusions of FGFR and TACC genes in human glioblastoma. *Science* 337, 1231–1235.
- Dupain, C., Harttrampf, A.C., Urbinati, G., Geoerger, B., and Massaad-Massade, L. (2017). Relevance of Fusion Genes in Pediatric Cancers: Toward Precision Medicine. *Mol. Ther. Nucleic Acids* 6, 315–326.
- Carrara, M., Beccuti, M., Lazzarato, F., Cavallo, F., Cordero, F., Donatelli, S., and Calogero, R.A. (2013). State-of-the-art fusion-finder algorithms sensitivity and specificity. *BioMed Res. Int.* 2013, 340620.
- Wang, Q., Xia, J., Jia, P., Pao, W., and Zhao, Z. (2013). Application of next generation sequencing to human gene fusion detection: computational tools, features and perspectives. *Brief. Bioinform.* 14, 506–519.
- Kumar, S., Razaq, S.K., Vo, A.D., Gautam, M., and Li, H. (2016). Identifying fusion transcripts using next generation sequencing. *Wiley Interdiscip. Rev. RNA* 7, 811–823.
- Liu, S., Tsai, W.-H., Ding, Y., Chen, R., Fang, Z., Huo, Z., Kim, S., Ma, T., Chang, T.Y., Priedigkeit, N.M., et al. (2016). Comprehensive evaluation of fusion transcript detection algorithms and a meta-caller to combine top performing methods in paired-end RNA-seq data. *Nucleic Acids Res.* 44, e47.
- Fernandez-Cuesta, L., Sun, R., Menon, R., George, J., Lorenz, S., Meza-Zepeda, L.A., Peifer, M., Plenker, D., Heuckmann, J.M., Leenders, F., et al. (2015). Identification of

- novel fusion genes in lung cancer using breakpoint assembly of transcriptome sequencing data. *Genome Biol.* 16, 7.
12. Shugay, M., Ortiz de Mendibil, I., Vizmanos, J.L., and Novo, F.J. (2013). Oncofuse: a computational framework for the prediction of the oncogenic potential of gene fusions. *Bioinformatics* 29, 2539–2546.
  13. Marshall, G.M., Carter, D.R., Cheung, B.B., Liu, T., Mateos, M.K., Meyerowitz, J.G., and Weiss, W.A. (2014). The prenatal origins of cancer. *Nat. Rev. Cancer* 14, 277–289.
  14. Harttrampf, A.C., Lacroix, L., Deloger, M., Deschamps, F., Puget, S., Auger, N., Vielh, P., Varlet, P., Balogh, Z., Abbou, S., et al. (2017). Molecular Screening for Cancer Treatment Optimization (MOSCATO-01) in pediatric patients: a single-institutional prospective molecular stratification trial. *Clin. Cancer Res.* 23, 6101–6112.
  15. McPherson, A., Hormozdiari, F., Zayed, A., Giuliany, R., Ha, G., Sun, M.G.F., Griffith, M., Heravi Moussavi, A., Senz, J., Melnyk, N., et al. (2011). deFuse: an algorithm for gene fusion discovery in tumor RNA-Seq data. *PLoS Comput. Biol.* 7, e1001138.
  16. Nicoric, D., Satalan, M., Edgren, H., Kangaspeska, S., Murumagi, A., Kallioniemi, O., Virtanen, S., and Kilkku, O. (2014). FusionCatcher - a tool for finding somatic fusion genes in paired-end RNA-sequencing data. <https://doi.org/10.1101/011650>.
  17. Kim, D., Pertea, G., Trapnell, C., Pimentel, H., Kelley, R., and Salzberg, S.L. (2013). TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. *Genome Biol.* 14, R36.
  18. Robinson, D.R., Wu, Y.-M., and Lin, S.-F. (2000). The protein tyrosine kinase family of the human genome. *Oncogene* 19, 5548–5557.
  19. Liu, Y., Sun, J., and Zhao, M. (2017). ONGene: A literature-based database for human oncogenes. *J. Genet. Genomics* 44, 119–121.
  20. Zhao, M., Sun, J., and Zhao, Z. (2013). TSGene: a web resource for tumor suppressor genes. *Nucleic Acids Res.* 41, D970–D976.
  21. Hanahan, D., and Weinberg, R.A. (2011). Hallmarks of cancer: the next generation. *Cell* 144, 646–674.
  22. Kalyana-Sundaram, S., Shankar, S., Deroo, S., Iyer, M.K., Palanisamy, N., Chinnaiyan, A.M., and Kumar-Sinha, C. (2012). Gene fusions associated with recurrent amplicons represent a class of passenger aberrations in breast cancer. *Neoplasia* 14, 702–708.
  23. Yoshihara, K., Wang, Q., Torres-Garcia, W., Zheng, S., Vegesna, R., Kim, H., and Verhaak, R.G. (2015). The landscape and therapeutic relevance of cancer-associated transcript fusions. *Oncogene* 34, 4845–4854.
  24. Vogelstein, B., Papadopoulos, N., Velculescu, V.E., Zhou, S., Diaz, L.A., Jr., and Kinzler, K.W. (2013). Cancer genome landscapes. *Science* 339, 1546–1558.
  25. Chase, A., Ernst, T., Fiebig, A., Collins, A., Grand, F., Erben, P., Reiter, A., Schreiber, S., and Cross, N.C. (2010). TFG, a target of chromosome translocations in lymphoma and soft tissue tumors, fuses to GPR128 in healthy individuals. *Haematologica* 95, 20–26.
  26. Cachia, D., Wani, K., Penas-Prado, M., Olar, A., McCutcheon, I.E., Benjamin, R.S., Armstrong, T.S., Gilbert, M.R., and Aldape, K.D. (2015). C11orf95-RELA fusion present in a primary supratentorial ependymoma and recurrent sarcoma. *Brain Tumor Pathol.* 32, 105–111.
  27. Lin, P.P., Pandey, M.K., Jin, F., Xiong, S., Deavers, M., Parant, J.M., and Lozano, G. (2008). EWS-FLI1 induces developmental abnormalities and accelerates sarcoma formation in a transgenic mouse model. *Cancer Res.* 68, 8968–8975.
  28. Zhu, C., Lv, Y., Wu, L., Guan, J., Bai, X., Lin, J., Liu, T., Sang, X., Xue, C., and Zhao, H. (2016). The landscape of gene fusions in hepatocellular carcinoma. <https://doi.org/10.1101/055376>.
  29. Mitelman, F., Johansson, B., and Mertens, F. (2007). The impact of translocations and gene fusions on cancer causation. *Nat. Rev. Cancer* 7, 233–245.
  30. Wang, L., Motoi, T., Khanin, R., Olshen, A., Mertens, F., Bridge, J., Dal Cin, P., Antonescu, C.R., Singer, S., Hameed, M., et al. (2012). Identification of a novel, recurrent HEY1-NCOA2 fusion in mesenchymal chondrosarcoma based on a genome-wide screen of exon-level expression data. *Genes Chromosomes Cancer* 51, 127–139.
  31. Bell, R.J.A., Rube, H.T., Xavier-Magalhães, A., Costa, B.M., Mancini, A., Song, J.S., and Costello, J.F. (2016). Understanding TERT Promoter Mutations: A Common Path to Immortality. *Mol. Cancer Res.* 14, 315–323.
  32. Pilsworth, J.A., Cochrane, D.R., Xia, Z., Aubert, G., Färkkilä, A.E.M., Horlings, H.M., Yanagida, S., Yang, W., Lim, J.L.P., Wang, Y.K., et al. (2018). TERT promoter mutation in adult granulosa cell tumor of the ovary. *Mod. Pathol.* 31, 1107–1115.
  33. Tsuda, M., Davis, I.J., Argani, P., Shukla, N., McGill, G.G., Nagai, M., Saito, T., Laé, M., Fisher, D.E., and Ladanyi, M. (2007). TFE3 fusions activate MET signaling by transcriptional up-regulation, defining another class of tumors as candidates for therapeutic MET inhibition. *Cancer Res.* 67, 919–929.
  34. Pavlicek, A., Lira, M.E., Lee, N.V., Ching, K.A., Ye, J., Cao, J., Garza, S.J., Hook, K.E., Ozeck, M., Shi, S.T., et al. (2013). Molecular predictors of sensitivity to the insulin-like growth factor 1 receptor inhibitor Figitumumab (CP-751,871). *Mol. Cancer Ther.* 12, 2929–2939.
  35. Zhang, B., Groffen, J., and Heisterkamp, N. (2005). Resistance to farnesyltransferase inhibitors in Bcr/Abl-positive lymphoblastic leukemia by increased expression of a novel ABC transporter homologue ATP11a. *Blood* 106, 1355–1361.
  36. Marshall, A.D., and Grosveld, G.C. (2012). Alveolar rhabdomyosarcoma - The molecular drivers of PAX3/7-FOXO1-induced tumorigenesis. *Skelet. Muscle* 2, 25.
  37. Beretta, G.L., Cassinelli, G., Pennati, M., Zuco, V., and Gatti, L. (2017). Overcoming ABC transporter-mediated multidrug resistance: The dual role of tyrosine kinase inhibitors as multitargeting agents. *Eur. J. Med. Chem.* 142, 271–289.
  38. Borad, M.J., Champion, M.D., Egan, J.B., Liang, W.S., Fonseca, R., Bryce, A.H., McCullough, A.E., Barrett, M.T., Hunt, K., Patel, M.D., et al. (2014). Integrated genomic characterization reveals novel, therapeutically relevant drug targets in FGFR and EGFR pathways in sporadic intrahepatic cholangiocarcinoma. *PLoS Genet.* 10, e1004135.
  39. Zhang, Z., Miao, L., Xin, X., Zhang, J., Yang, S., Miao, M., Kong, X., and Jiao, B. (2014). Underexpressed CNDP2 participates in gastric cancer growth inhibition through activating the MAPK signaling pathway. *Mol. Med.* 20, 17–28.
  40. Wick, W., Weller, M., van den Bent, M., Sanson, M., Weiler, M., von Deimling, A., Plass, C., Hegi, M., Platten, M., and Reifenberger, G. (2014). MGMT testing—the challenges for biomarker-based glioma treatment. *Nat. Rev. Neurol.* 10, 372–385.
  41. Carraway, K.L., 3rd, Weber, J.L., Unger, M.J., Ledesma, J., Yu, N., Gassmann, M., and Lai, C. (1997). Neuregulin-2, a new ligand of ErbB3/ErbB4-receptor tyrosine kinases. *Nature* 387, 512–516.
  42. Hirai, H., Iwasawa, Y., Okada, M., Arai, T., Nishibata, T., Kobayashi, M., Kimura, T., Kaneko, N., Ohtani, J., Yamanaka, K., et al. (2009). Small-molecule inhibition of Wee1 kinase by MK-1775 selectively sensitizes p53-deficient tumor cells to DNA-damaging agents. *Mol. Cancer Ther.* 8, 2992–3000.
  43. Aarts, M., Sharpe, R., Garcia-Murillas, I., Gevensleben, H., Hurd, M.S., Shumway, S.D., Toniatti, C., Ashworth, A., and Turner, N.C. (2012). Forced mitotic entry of S-phase cells as a therapeutic strategy induced by inhibition of WEE1. *Cancer Discov.* 2, 524–539.
  44. Zhao, S., Sedwick, D., and Wang, Z. (2015). Genetic alterations of protein tyrosine phosphatases in human cancers. *Oncogene* 34, 3885–3894.
  45. Meehan, M., Parthasarathi, L., Moran, N., Jefferies, C.A., Foley, N., Lazzari, E., Murphy, D., Ryan, J., Ortiz, B., Fabius, A.W., et al. (2012). Protein tyrosine phosphatase receptor delta acts as a neuroblastoma tumor suppressor by destabilizing the aurora kinase A oncogene. *Mol. Cancer* 11, 6.
  46. Pettrilli, A.M., and Fernández-Valle, C. (2016). Role of Merlin/NF2 inactivation in tumor biology. *Oncogene* 35, 537–548.
  47. Shapiro, I.M., Kolev, V.N., Vidal, C.M., Kadariya, Y., Ring, J.E., Wright, Q., Weaver, D.T., Menges, C., Padval, M., McClatchey, A.I., et al. (2014). Merlin deficiency predicts FAK inhibitor sensitivity: a synthetic lethal relationship. *Sci. Transl. Med.* 6, 237ra68.
  48. Brenner, J.C., Feng, F.Y., Han, S., Patel, S., Goyal, S.V., Bou-Maroun, L.M., Liu, M., Lonigro, R., Prensner, J.R., Tomlins, S.A., and Chinnaiyan, A.M. (2012). PARP-1 inhibition as a targeted strategy to treat Ewing's sarcoma. *Cancer Res.* 72, 1608–1613.
  49. Loganathan, S.N., Tang, N., Fleming, J.T., Ma, Y., Guo, Y., Borinstein, S.C., Chiang, C., and Wang, J. (2016). BET bromodomain inhibitors suppress EWS-FLI1-dependent transcription and the IGF1 autocrine mechanism in Ewing sarcoma. *Oncotarget* 7, 43504–43517.
  50. Huang, Z., Jones, D.T.W., Wu, Y., Lichter, P., and Zapatka, M. (2017). confuse: High-Confidence Fusion Gene Detection across Tumor Entities. *Front. Genet.* 8, 137.

51. Gao, Q., Liang, W.-W., Foltz, S.M., Mutharasu, G., Jayasinghe, R.G., Cao, S., Liao, W.W., Reynolds, S.M., Wyczalkowski, M.A., Yao, L., et al.; Fusion Analysis Working Group; Cancer Genome Atlas Research Network (2018). Driver Fusions and Their Implications in the Development and Treatment of Human Cancers. *Cell Rep.* 23, 227–238.e3.
52. Burrell, R.A., and Swanton, C. (2014). Tumour heterogeneity and the evolution of polyclonal drug resistance. *Mol. Oncol.* 8, 1095–1111.
53. Zhang, J., Walsh, M.F., Wu, G., Edmonson, M.N., Gruber, T.A., Easton, J., Hedges, D., Ma, X., Zhou, X., Yergeau, D.A., et al. (2015). Germline Mutations in Predisposition Genes in Pediatric Cancer. *N. Engl. J. Med.* 373, 2336–2346.
54. Chang, W., Brohl, A.S., Patidar, R., Sindiri, S., Shern, J.F., Wei, J.S., Song, Y.K., Yohe, M.E., Gryder, B., Zhang, S., et al. (2016). Multi-Dimensional ClinOmics for Precision Therapy of Children and Adolescent Young Adults with Relapsed and Refractory Cancer: A Report from the Center for Cancer Research. *Clin. Cancer Res.* 22, 3810–3820.
55. Pajtler, K.W., Witt, H., Sill, M., Jones, D.T.W., Hovestadt, V., Kratochwil, F., Wani, K., Tatevosian, R., Punchihewa, C., Johann, P., et al. (2015). Molecular Classification of Ependymal Tumors across All CNS Compartments, Histopathological Grades, and Age Groups. *Cancer Cell* 27, 728–743.
56. Servidei, T., Meco, D., Muto, V., Bruselles, A., Ciolfi, A., Trivieri, N., Lucchini, M., Morosetti, R., Mirabella, M., Martini, M., et al. (2017). Novel *SEC61G-EGFR* Fusion Gene in Pediatric Ependymomas Discovered by Clonal Expansion of Stem Cells in Absence of Exogenous Mitogens. *Cancer Res.* 77, 5860–5872.
57. Poh, A.R., O'Donoghue, R.J.J., and Ernst, M. (2015). Hematopoietic cell kinase (HCK) as a therapeutic target in immune and cancer cells. *Oncotarget* 6, 15752–15771.
58. Ortiz, B., Fabius, A.W.M., Wu, W.H., Pedraza, A., Brennan, C.W., Schultz, N., Pitter, K.L., Bromberg, J.F., Huse, J.T., Holland, E.C., and Chan, T.A. (2014). Loss of the tyrosine phosphatase PTPRD leads to aberrant STAT3 activation and promotes gliomagenesis. *Proc. Natl. Acad. Sci. USA* 111, 8149–8154.
59. Atak, Z.K., Gianfelici, V., Hulselmans, G., De Keersmaecker, K., Devasia, A.G., Geerdens, E., Mentens, N., Chiaretti, S., Durinck, K., Uytendaele, A., et al. (2013). Comprehensive analysis of transcriptome variation uncovers known and novel driver events in T-cell acute lymphoblastic leukemia. *PLoS Genet.* 9, e1003997.
60. Brohl, A.S., Kahen, E., Yoder, S.J., Teer, J.K., and Reed, D.R. (2017). The genomic landscape of malignant peripheral nerve sheath tumors: diverse drivers of Ras pathway activation. *Sci. Rep.* 7, 14992.
61. Midgley, C.A., Desterro, J.M., Saville, M.K., Howard, S., Sparks, A., Hay, R.T., and Lane, D.P. (2000). An N-terminal p14<sup>ARF</sup> peptide blocks Mdm2-dependent ubiquitination *in vitro* and can activate p53 *in vivo*. *Oncogene* 19, 2312–2323.
62. Brennan, C.W., Verhaak, R.G., McKenna, A., Campos, B., Nounshahr, H., Salama, S.R., Zheng, S., Chakravarty, D., Sanborn, J.Z., Berman, S.H., et al.; TCGA Research Network (2013). The somatic genomic landscape of glioblastoma. *Cell* 155, 462–477.
63. Sturm, D., Bender, S., Jones, D.T.W., Lichter, P., Grill, J., Becher, O., Hawkins, C., Majewski, J., Jones, C., Costello, J.F., et al. (2014). Paediatric and adult glioblastoma: multifiform (epi)genomic culprits emerge. *Nat. Rev. Cancer* 14, 92–107.
64. Massard, C., Michiels, S., Ferté, C., Le Deley, M.-C., Lacroix, L., Hollebecque, A., Verlingue, L., Ileana, E., Rosellini, S., Ammari, S., et al. (2017). High-Throughput Genomics and Clinical Outcome in Hard-to-Treat Advanced Cancers: Results of the MOSCATO 01 Trial. *Cancer Discov.* 7, 586–595.
65. Novo, F.J., de Mendibil, I.O., and Vizmanos, J.L. (2007). TICdb: a collection of gene-mapped translocation breakpoints in cancer. *BMC Genomics* 8, 33.
66. Kim, N., Kim, P., Nam, S., Shin, S., and Lee, S. (2006). ChimerDB—a knowledgebase for fusion sequences. *Nucleic Acids Res.* 34, D21–D24.
67. Gorohovski, A., Tagore, S., Palande, V., Malka, A., Raviv-Shay, D., and Frenkel-Morgenstern, M. (2017). ChiTaRS-3.1—the enhanced chimeric transcripts and RNA-seq database matched with protein-protein interactions. *Nucleic Acids Res.* 45 (D1), D790–D795.
68. Coon, T.A., Glasser, J.R., Mallampalli, R.K., and Chen, B.B. (2012). Novel E3 ligase component FBXL7 ubiquitinates and degrades Aurora A, causing mitotic arrest. *Cell Cycle* 11, 721–729.
69. Gryder, B.E., Yohe, M.E., Chou, H.C., Zhang, X., Marques, J., Wachtel, M., Schaefer, B., Sen, N., Song, Y., Gualtieri, A., et al. (2017). PAX3-FOXO1 Establishes Myogenic Super Enhancers and Confers BET Bromodomain Vulnerability. *Cancer Discov.* 7, 884–899.
70. Alaei-Mahabadi, B., Bhadury, J., Karlsson, J.W., Nilsson, J.A., and Larsson, E. (2016). Global analysis of somatic structural genomic alterations and their impact on gene expression in diverse human cancers. *Proc Natl Acad Sci U S A.* 113, 13768–13773.