

PDF hosted at the Radboud Repository of the Radboud University Nijmegen

The following full text is a publisher's version.

For additional information about this publication click this link.

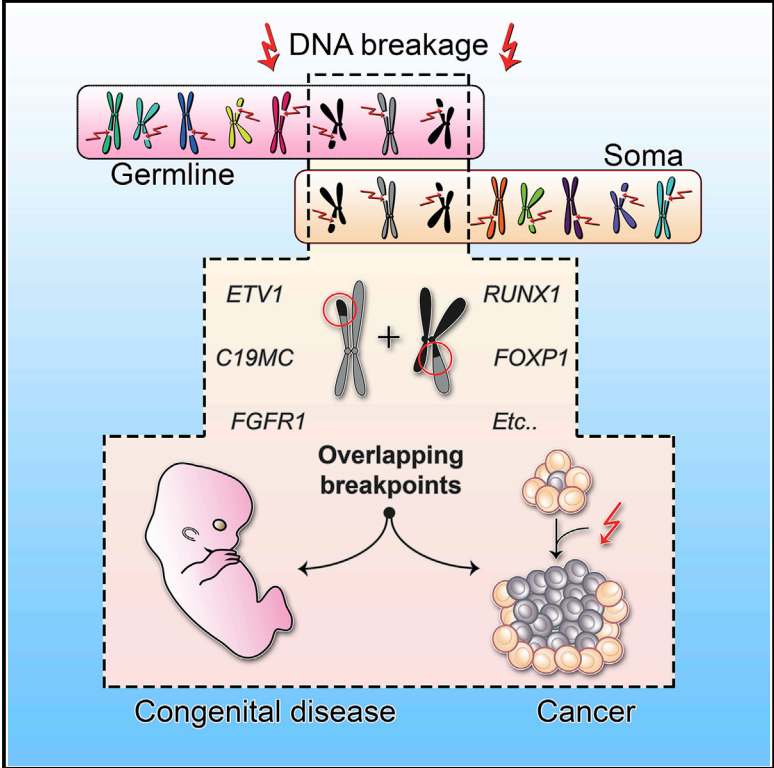
<http://hdl.handle.net/2066/139109>

Please be advised that this information was generated on 2017-12-05 and may be subject to change.

Cell Reports

Genomic and Functional Overlap between Somatic and Germline Chromosomal Rearrangements

Graphical Abstract



Authors

Sebastiaan van Heesch, Marieke Simonis, ..., Edwin Cuppen, Wigard P. Kloosterman

Correspondence

e.cuppen@hubrecht.eu (E.C.), w.kloosterman@umcutrecht.nl (W.P.K.)

In Brief

Chromosomal rearrangements are a frequent cause of human disease, including congenital defects and cancer. van Heesch et al. show that germline rearrangements can target the same genes and genomic regions as rearrangements in cancer. They highlight two patients with congenital disease caused by germline rearrangements involving the cancer genes *ETV1*, *FOXP1*, and microRNA cluster *C19MC*.

Highlights

- Germline structural variation breakpoints overlap with somatic cancer breakpoints
- De novo germline genomic rearrangements target genes rearranged in cancer
- De novo germline breakpoints involve activation of cancer genes via gene fusion

Genomic and Functional Overlap between Somatic and Germline Chromosomal Rearrangements

Sebastiaan van Heesch,^{1,7} Marieke Simonis,^{1,7} Markus J. van Roosmalen,² Vamsee Pillalamarri,³ Harrison Brand,³ Ewart W. Kuijk,¹ Kim L. de Luca,¹ Nico Lansu,¹ A. Koen Braat,⁴ Androniki Menelaou,² Wensi Hao,¹ Jeroen Korving,¹ Simone Snijder,⁵ Lars T. van der Veken,² Ron Hochstenbach,² Alida C. Knecht,⁵ Karen Duran,² Ivo Renkens,² Najla Alekozai,² Myrthe Jager,² Sarah Vergult,⁶ Björn Menten,⁶ Ewart de Bruijn,¹ Sander Boymans,¹ Elly Ippel,² Ellen van Binsbergen,² Michael E. Talkowski,³ Klaske Lichtenbelt,² Edwin Cuppen,^{1,2,*} and Wigard P. Kloosterman^{2,*}

¹Hubrecht Institute-KNAW and University Medical Center Utrecht, Uppsalalaan 8, 3584 CT Utrecht, the Netherlands

²Department of Medical Genetics, Center for Molecular Medicine, University Medical Center Utrecht, Universiteitsweg 100, 3584 CG Utrecht, the Netherlands

³Center for Human Genetic Research, Massachusetts General Hospital, Boston, MA 02114, USA

⁴Department of Cell Biology, Center for Molecular Medicine, University Medical Center Utrecht, Heidelberglaan 100, 3584 CX Utrecht, the Netherlands

⁵Department of Clinical Genetics, Academic Medical Center, Meibergdreef 9, 1105 AZ Amsterdam, the Netherlands

⁶Center for Medical Genetics, Ghent University Hospital, De Pintelaan 185, 9000 Ghent, Belgium

⁷Co-first author

*Correspondence: e.cuppen@hubrecht.eu (E.C.), w.kloosterman@umcutrecht.nl (W.P.K.)

<http://dx.doi.org/10.1016/j.celrep.2014.11.022>

This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/3.0/>).

SUMMARY

Genomic rearrangements are a common cause of human congenital abnormalities. However, their origin and consequences are poorly understood. We performed molecular analysis of two patients with congenital disease who carried de novo genomic rearrangements. We found that the rearrangements in both patients hit genes that are recurrently rearranged in cancer (*ETV1*, *FOXP1*, and microRNA cluster C19MC) and drive formation of fusion genes similar to those described in cancer. Subsequent analysis of a large set of 552 de novo germline genomic rearrangements underlying congenital disorders revealed enrichment for genes rearranged in cancer and overlap with somatic cancer breakpoints. Breakpoints of common (inherited) germline structural variations also overlap with cancer breakpoints but are depleted for cancer genes. We propose that the same genomic positions are prone to genomic rearrangements in germline and soma but that timing and context of breakage determines whether developmental defects or cancer are promoted.

INTRODUCTION

De novo germline genomic rearrangements are a common cause of congenital disease, including mental retardation and neurodevelopmental delay (Cooper et al., 2011; Stankiewicz and Lupski, 2010). Germline rearrangements can be classified in two major categories. One category arises through nonallelic homologous recombination via genomic repeats (Cooper et al., 2011; Hastings et al., 2009a; Vissers and Stankiewicz, 2012) primarily re-

sulting in copy-number changes (CNVs). Most of these CNVs are recurrent and give rise to recognizable phenotypes known as microdeletion and microduplication syndromes, which can result from dosage effects of one or more genes within the CNV interval (Golzio et al., 2012; Luo et al., 2012). The second category contains sporadic (nonrecurrent) genomic rearrangements and comprises more diverse rearrangement types including CNVs, translocations, inversions and complex events. These rearrangements are primarily caused by nonhomologous modes of DNA repair, such as direct end joining of free DNA ends (Lieber, 2010) or template-switching following replication fork stalling (Hastings et al., 2009b). Also, ultracomplex rearrangements resulting from the shattering of chromosomes in a single event, termed chromothripsis, arise through nonhomologous DNA repair (Chiang et al., 2012; Kloosterman et al., 2012; Stephens et al., 2011).

Despite the knowledge of repair mechanisms that may facilitate genomic rearrangements, the molecular basis and genomic context of sporadic de novo rearrangements is not fully understood. Moreover, for the majority of patients with de novo germline rearrangements—particularly complex ones—the actual cause of disease remains unclear because of the uniqueness of the breakpoints and the multiple possible effects on gene expression and function (Luo et al., 2012; Weischenfeldt et al., 2013).

Here, we gained deep insight into the effects of sporadic de novo genomic rearrangements in patients with congenital disease by performing integrated genomic, transcriptomic, small RNA, and chromatin immunoprecipitation (ChIP) profiling of blood from parent-offspring families. We identify gene fusions that resemble those found in cancer. In addition, we find an overlap in breakpoint positions for common and de novo germline rearrangements and somatic cancer rearrangements. These data suggest a common mechanistic origin for germline and somatic breakpoints.

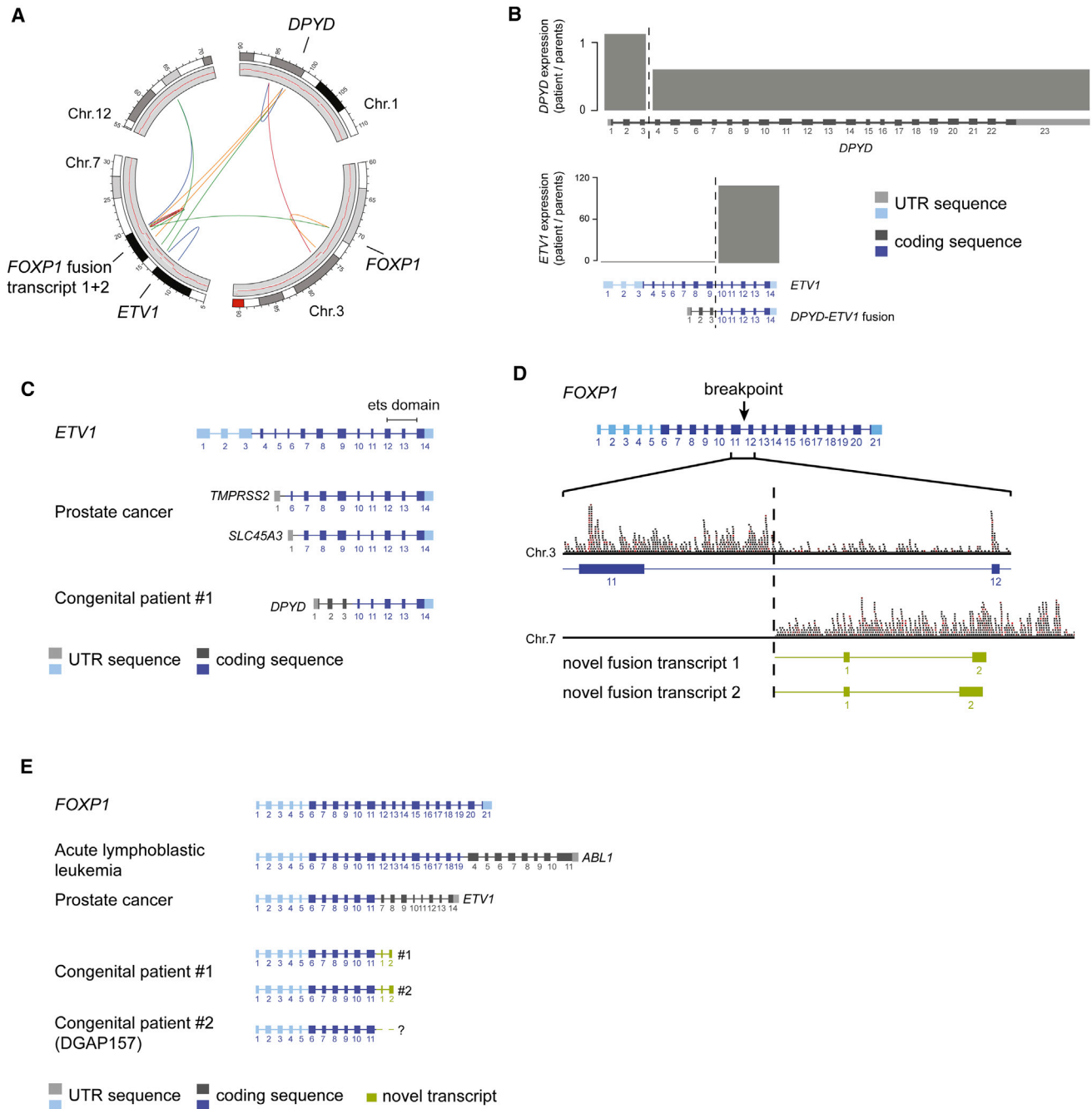


Figure 1. A De Novo Chromothripsis Rearrangement Results in *ETV1* and *FOXP1* Fusion Transcripts

(A) Circos plot of the 13 breakpoint junctions forming the chromothripsis rearrangement. The outer circle displays the chromosome ideogram. The inner circle represents the copy-number profile as based on read-depth measurements relative to the parents. The colored lines indicate breakpoint junctions. Blue, tail-to-head; green, head-to-tail; red, head-to-head inverted; yellow, tail-to-tail inverted. The locations of relevant genes are indicated. Chromosome coordinates are in megabases.

(B) Visualization of the *DPYD-ETV1* fusion gene and the transcriptional consequences thereof. RNA-seq reads within the genomic intervals from the start of the genes to the breakpoint (dashed vertical line) and from the breakpoint to the end of the gene were normalized for the total amount of reads per sample. The plot visualizes the ratios of normalized reads in the patient versus the average of the parents.

(C) Diagram showing the full-length *ETV1* gene, examples of *ETV1* fusion genes as observed in cancer (Hermans et al., 2008; Tomlins et al., 2007) and the *DPYD-ETV1* fusion in our patient.

(legend continued on next page)

RESULTS

Family-Based Molecular Analysis Reveals Fusion Transcripts Involving *ETV1* and *FOXP1* as a Result of Germline Chromothripsis

We employed an *in vivo* family-based molecular profiling approach to characterize the effects of *de novo* structural genomic rearrangements in two independent patients with multiple congenital abnormalities and intellectual disability (MCA/ID; Table S1).

In one patient with speech delay, psychomotor retardation, dysmorphic facial appearance, and doubling of one of the thumbs, we identified a *de novo* chromothripsis rearrangement. This germline chromothripsis rearrangement involves 17 breakpoints divided over four chromosomes (1, 3, 7, and 12; Figures 1A and S1A; Table S2) (Kloosterman et al., 2012). To study the effects of chromothripsis on gene expression, we performed RNA sequencing (RNA-seq) on peripheral blood mononuclear cells (PBMCs) of this patient and both parents. First, we examined the expression levels of 11 genes that reside within three large *de novo* genomic deletions caused by the chromothripsis. Four of these genes were expressed in PBMCs and showed a clear decrease in expression levels relative to the parents (Figure S1B). In addition to these deleted genes, six genomic breakpoints were located within a gene, thus splitting up the coding sequence (Table S2). Of the six genes disrupted by breakpoints, three are transcriptionally active in PBMCs. Two of them (*DPYD* and *FOXP1*) showed a decrease in expression following the breakpoint (Figures 1B and S1C). In contrast, for the third gene (*ETV1*) the C-terminal part showed elevated expression in the patient relative to the parents, whereas the N-terminal part was not expressed. Examination of the breakpoint junctions involving *ETV1* revealed a genomic fusion between the first three exons of the *DPYD* gene and exons 10–14 of the *ETV1* gene (Figure 1B). *DPYD* encodes for dihydropyrimidine dehydrogenase, an essential factor for uracil and thymidine catabolism that is ubiquitously expressed. As a result, the genomic fusion between *DPYD* and *ETV1* leads to high expression of the 3' part of *ETV1* in patient blood, whereas the parents do not express *ETV1*.

ETV1 is a member of the ETS (E-twenty-six) family of transcription factors that modulate target genes involved in cell differentiation, proliferation, migration, and apoptosis (Oh et al., 2012). *ETV1* gene fusions are frequently found in Ewing sarcoma and prostate cancer but have not been described as drivers of congenital disease (Hermans et al., 2008; Jeon et al., 1995; Kumar-Sinha et al., 2008; Tomlins et al., 2005, 2007). Remarkably, the topology of the *DPYD-ETV1* fusion in this patient resembles that of the *ETV1* fusion genes found in cancer (Figure 1C), albeit the patient has not been diagnosed with cancer at this point. Both in this patient and in cancer the 3' part of *ETV1*, which contains the ETS transcription activation domain, becomes ectopically expressed by fusion to the 5' part of an actively transcribed

gene (Kumar-Sinha et al., 2008). We constructed a cDNA gene mimicking the *DPYD-ETV1* fusion and overexpressed it in HEK293 cells to determine functionality of the protein. Although we can detect sporadic protein product in these cells (~1/50, Figure S1D), we could not detect a stable fusion protein on western blot analysis, suggesting that this product is only stable and/or translated under specific conditions.

We also studied the transcriptional consequences of the breakpoint in *FOXP1* in more detail. RNA-seq analysis showed readthrough transcription from exon 11 of the *FOXP1* gene to a genomic segment on chromosome 7 (Figure 1D). No annotated coding gene was present as 3' fusion partner of *FOXP1*, but cDNA analysis of the readthrough transcripts showed two differentially spliced transcripts fused to the 11th exon of *FOXP1* (Figure 1D; Figure S1E). Mutations in *FOXP1* have been frequently associated with developmental disease (O'Roak et al., 2011; Talkowski et al., 2012), but newly generated fusion transcripts as a result of translocation to a noncoding region have not been observed previously. The fusion transcripts identified here resemble *FOXP1* gene fusions that are observed in cancer (Figure 1E) (Ernst et al., 2011; Hermans et al., 2008). In particular, the intron targeted for translocation is identical to the introns targeted in many *FOXP1* gene fusions in cancer. Furthermore, we identified a second patient with a *de novo* germline breakpoint in the same intron in *FOXP1*, indicating that this is a recurrently rearranged region in both cancer and germline (Talkowski et al., 2012). The two transcript isoforms identified in the patient with chromothripsis add 24 and 46 amino acids, respectively, to the *FOXP1* open reading frame. Upon expression in HEK293 cells, both transcripts result in stable protein products (Figure S1F).

These results provide insight into the molecular effects of germline chromothripsis rearrangements and show that chromosome shattering can lead to transcriptional activation in addition to gene disruption. The mechanisms of gene activation are very similar to those of somatic rearrangements in cancer genomes. Our data demonstrate that spliced transcripts resulting in stable proteins can be formed through a germline chromothripsis rearrangement.

A *De Novo* Germline Duplication Activates a Cluster of Oncogenic MicroRNAs

The second patient with congenital defects analyzed using molecular phenotyping carries a *de novo* 424.5 kb tandem duplication on chromosome 19, resulting in macrocephaly and severe psychomotor retardation (Figures 2A and S2A; Table S1). The most predictable effect of a genomic duplication is elevated gene expression due to an increase in gene copy number. Indeed, RNA-seq analysis performed on peripheral blood mononuclear cells (PBMCs) demonstrates that three duplicated genes are expressed significantly higher in the patient compared to both the parents and an unaffected sibling (Figure S2B). Closer examination of the breakpoints of the duplication revealed

(D) Raw RNA-seq reads depicting *FOXP1* transcription across the chromosome 3 to chromosome 7 breakpoint (dashed vertical line), resulting in two fusion transcripts (green).

(E) Diagram showing the full-length *FOXP1* gene, examples of *FOXP1* fusion genes as observed in cancer, and the *FOXP1* breakpoints in two patients in our data set (Chiang et al., 2012; Ernst et al., 2011; Hermans et al., 2008).

See also Figure S1 and Table S1.

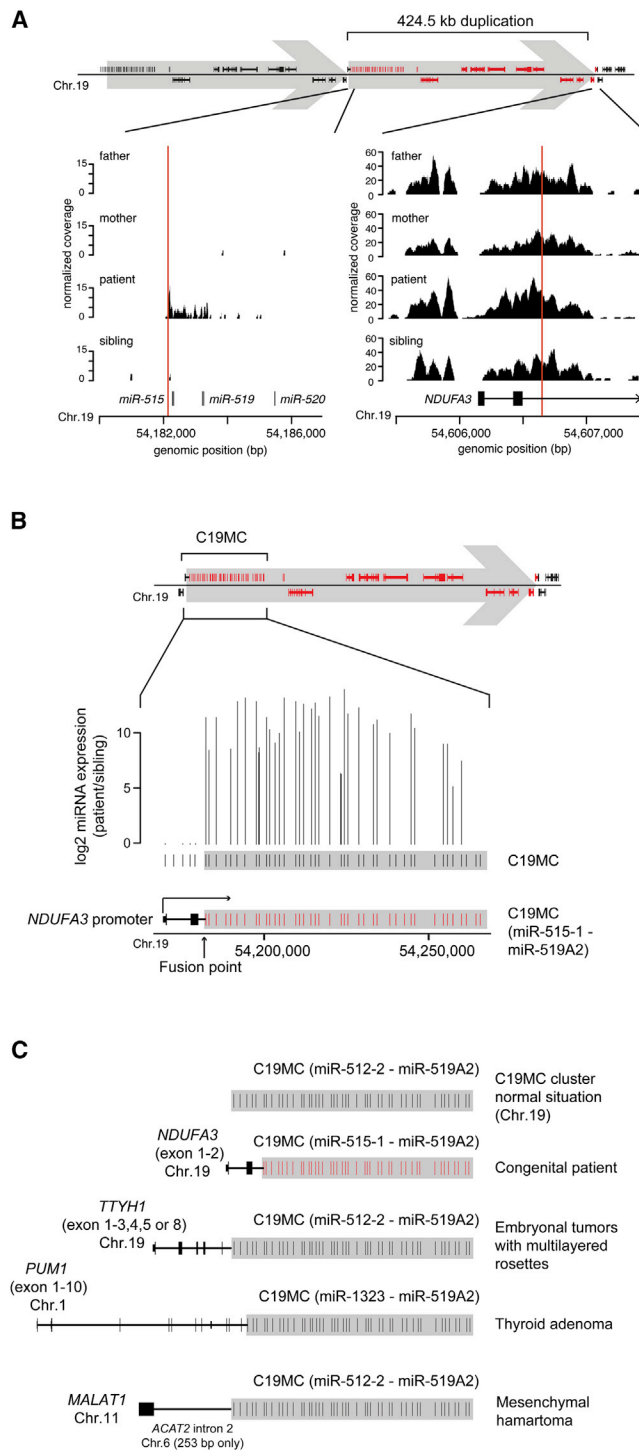


Figure 2. A De Novo 424.5 kb Tandem Duplication Activates C19MC Expression

(A) H3K4me3 ChIP-seq results for the promoter of the *NDUFA3* gene in the father, mother, patient, and healthy sibling. The upper panel shows a schematic representation of the duplication. The gray arrows show which part of the chromosome is duplicated. Genes within the duplication are depicted in red. The lower-left panel shows H3K4me3 signals for a region surrounding the 5' breakpoint of the tandem duplication. The lower-right panel shows

unexpected additional molecular effects. The 5' breakpoint of the duplicated region is located within the chromosome 19 microRNA cluster (C19MC) and the 3' breakpoint disrupts the *NDUFA3* gene. The tandem duplication repositioned a major part of the C19MC miRNA cluster immediately downstream of the promoter of *NDUFA3*. This prompted us to investigate the presence of active promoter elements in the rearranged locus by H3K4me3 chromatin immunoprecipitation sequencing (ChIP-seq). The *NDUFA3* promoter was found to have high H3K4me3 levels in all samples, and this H3K4me3 signal was found to extend into the C19MC cluster downstream of the 5' duplication breakpoint in the patient (Figure 2A). In addition, small RNA-seq revealed that the C19MC miRNAs positioned downstream of the *NDUFA3* promoter were highly expressed (Figure 2B), whereas they are nonexpressed in both parental samples and the unaffected sibling. The part of the C19MC cluster that is not repositioned by the duplication was not expressed in the patient (Figure 2B) and miRNAs elsewhere on the genome were also unaffected (Figure S2C). The miRNA encoded by the *MIR371* gene, which is also located in the duplication but is driven by its own promoter, also shows no upregulation (Figure S2B). Endogenous expression of C19MC is exclusive to embryonic stem cells and tumors, which suggests that normal differentiation and development could be disturbed upon ectopic expression of this cluster (Bar et al., 2008; Flor and Bullerdiek, 2012). The *NDUFA3* gene, which encodes a subunit of a mitochondrial protein complex, is broadly expressed and therefore expected to drive C19MC miRNA expression in many tissues in the patient.

Ectopic Expression of Oncogenic C19MC MicroRNAs Drives Defects in Embryonic Development

Genomic rearrangements in the 150 kb common breakpoint cluster on the long arm of chromosome 19 are known to affect C19MC expression in thyroid adenomas, epithelial tumors, and embryonal brain tumors (Belge et al., 2001; Kleinman et al., 2014; Rippe et al., 2010). Previous reports have shown aberrant expression of part of the C19MC cluster in cancer resulting from the repositioning of an active promoter (Kleinman et al., 2014; Rajaram et al., 2007; Rippe et al., 2010) (Figure 2C). Other studies have shown that expression of the C19MC cluster in cancer cells is an important driver of tumorigenesis, tumor invasion, and metastasis (Hu et al., 2011; Huang et al., 2008), with eight C19MC members directly targeting *p21* (*CDKN1A*) and C19MC being a transcriptional target of TP53 (Flor and Bullerdiek, 2012; Fornari et al., 2012; Wu et al., 2010).

H3K4me3 signals for the *NDUFA3* promoter region. The vertical red lines indicate the position of the duplication breakpoints.

(B) Normalized log₂ expression ratios of microRNAs in C19MC for the patient versus healthy sibling (control). The duplicated fraction of C19MC is colored red. An arrow depicts the breakpoint junction that fuses exon 2 of *NDUFA3* to C19MC.

(C) Examples of chromosomal rearrangements activating C19MC in cancer. The germline rearrangement activating C19MC is depicted followed by three previously described rearrangements in embryonal brain tumors (Kleinman et al., 2014), thyroid adenoma (Rippe et al., 2010), and mesenchymal hamartoma (Rajaram et al., 2007).

See also Figure S2 and Table S1.

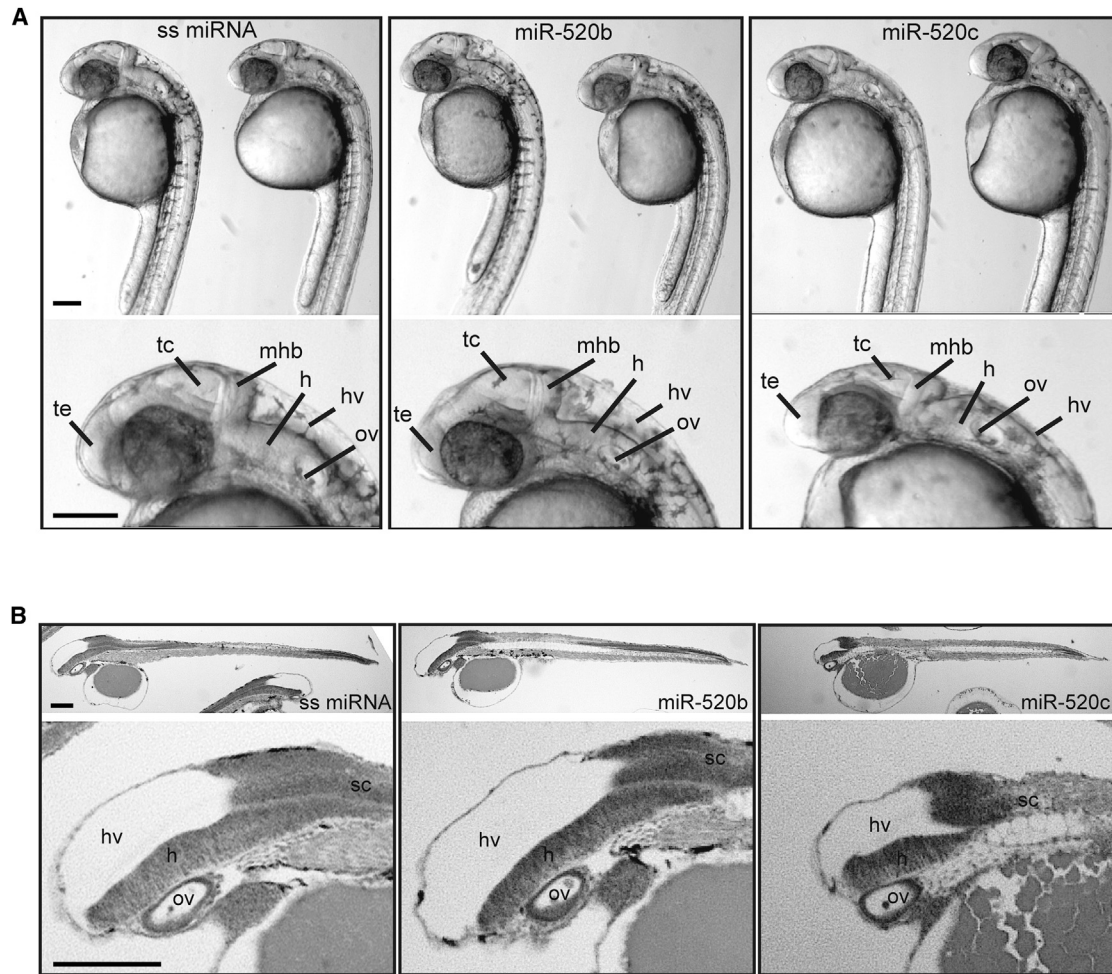


Figure 3. Ectopic Expression of C19MC MicroRNAs in Zebrafish Embryos

(A) Whole-mount bright-field images of 24 hr postfertilization (hpf) zebrafish embryos derived from control injections (single-stranded miRNAs) and injections with miR-520b and miR-520c duplexes. Zoomed views of the zebrafish are displayed in the lower panel, with annotation of the hindbrain (h), hindbrain ventricle (hv), otic vesicle (ov), telencephalon (te), midbrain-hindbrain boundary (mhb), and tectum (tc).

(B) Sagittal sections of embryos derived from the same experiments as in (A). Hindbrain (h), hindbrain ventricle (hv), otic vesicle (ov), and spinal cord (sc) are annotated. Scale bars, 200 μ m.

See also [Figure S2](#).

We selected two of these cancer-related miRNAs, miR-520b and miR-520c, to study the effects of overexpression of the mature miRNA duplex on zebrafish embryonic development ([Figures S2D–S2F](#)). These two miRNAs were selected based on homology with zebrafish miRNAs ([Figure S2F](#)), presence of the same miRNA seed sequence among several other C19MC miRNAs ([Figure S2E](#)), oncogenic potential as shown by previous studies ([Hu et al., 2011](#); [Huang et al., 2008](#)), and de novo expression in the patient. Injection of single-stranded miRNA controls and miR-520b duplexes in the 1-cell stage embryo did not result in a noticeable phenotype at 24 hr postfertilization (the miRNA is stable up to \sim 30 hr after injection). However, for miR-520c, we detected specific developmental malformations ([Figure 3A](#)). Particularly, the head of miR-520c-injected embryos is smaller and displays a reduced fore- and hindbrain ventricle and an altered morphology of the midbrain-hindbrain boundary ([Fig-](#)

[ure 3B](#)). These results demonstrate that overexpression of specific oncogenic C19MC miRNAs can disturb normal embryonic development and are therefore likely to have contributed to the neurodevelopmental defects in the patient.

Intersection of De Novo Germline Breakpoints with Cancer-Related Genes and Breakpoints

The two clinical cases described above both carried breakpoints at positions close to genomic rearrangement positions and genes broken in cancer ([Figure 4A](#)). Triggered by these findings, we set out to systematically analyze a large set of 552 de novo germline chromosomal rearrangements (DN), in comparison to a set of 28,844 common germline (CG) structural variants commonly present in the population and a comprehensive set of 68,018 breakpoints from somatic cancer rearrangements (SC; [Supplemental Experimental Procedures](#); [Table S2](#)). All

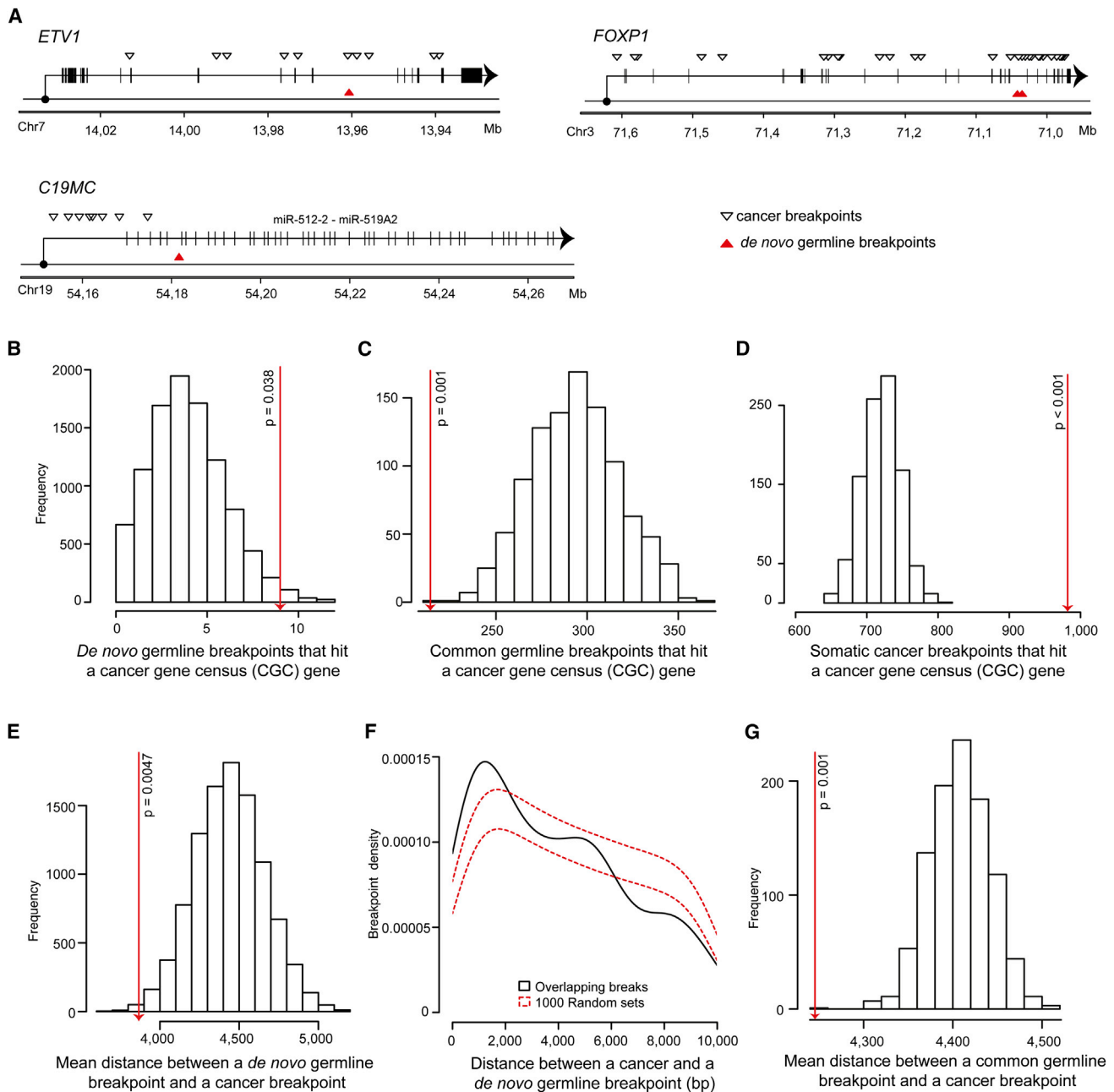


Figure 4. Overlap of De Novo and Common Germline Breakpoints with Cancer Gene Census Genes and Somatic Cancer Breakpoints

(A) Positions of cancer breakpoints (white triangles) and germline breakpoints (red triangles) in the *ETV1* and *FOXP1* genes and C19MC microRNA cluster. (B–D) Overlap of de novo germline (DN) breakpoints (B), common germline (CG) breakpoints from the 1000 Genomes project (C), and somatic cancer (SC) breakpoints (D) with recurrently translocated genes derived from the Cancer Gene Census (CGC) database.

(E) Histogram showing the mean distance of DN breakpoints relative to the nearest SC breakpoint (red line), as compared to randomly permuted control sets. (F) Density plot showing the distance between DN breakpoints and SC breakpoints within a 10 kb window (black line). The area between dotted red lines represents the mean distance between breakpoints ± 1 SD computed over 1000 simulation sets. This plot highlights that the distribution of DN breakpoints is skewed toward shorter interbreakpoint distances (<2 kb) as compared to the random breakpoint sets.

(G) Histogram showing the mean distance of CG breakpoints (1000 Genomes) relative to the nearest SC breakpoint (red line), as compared to randomly permuted control sets.

See also [Figures S3](#) and [S4](#) and [Tables S2](#), [S3](#), and [S4](#).

following analyses exclude breakpoints from the original two patients and are thus based on an independent DN breakpoint data set ($n = 533$).

First, we intersected DN breakpoints with genes listed in the Cancer Gene Census (CGC) database of genes recurrently rearranged in cancer (Table S3) (Futreal et al., 2004). This revealed nine independent breakpoints targeting CGC genes (Table S3). These nine CGC hits represent on average 2.03-fold enrichment relative to matched randomly simulated breakpoint sets (Figure 4B, $p = 0.0383$; Experimental Procedures). This enrichment of DN breakpoints for CGC genes does not result from general enrichment for protein-coding genes, because the overlap of DN breakpoint with protein-coding genes is not significantly deviating from randomly simulated control sets (Figure S3A). To substantiate this, we tested for association of the DN breakpoint positions with CGC genes and all protein coding genes and observed a positive association of DN breakpoints with CGC genes independent of the overlap of DN breakpoints with all protein-coding genes (logistic regression coefficient = 0.59; p value = 0.0469). In contrast to DN breakpoints, we found that the 28,844 CG breakpoints show 1.35-fold depletion for CGC genes ($p < 0.001$; Figure 4C), but this is largely explained by a general depletion for protein coding genes (Figure S3B) (Mills et al., 2011). This was confirmed by regression analysis, which did not reveal significant association between CGC genes and CG breakpoints when both CGC genes and all protein coding genes were added as predictive variables (logistic regression p value = 0.07). As expected, the set of 68,018 SC breakpoints shows a strong enrichment (1.36-fold) for CGC genes (Figure 4D; p value < 0.001).

In the DN set, we found a breakpoint in the CGC gene *FGFR1*, contributing to an in-frame gene fusion. Recently, the transforming activity of recurrent *FGFR1* fusions was reported for glioblastoma (Singh et al., 2012). We did not observe an overall enrichment for in-frame gene fusions among de novo breakpoint junctions in our data as compared to matching randomly generated sets. However, a much larger fraction of in-frame gene fusions was found for the DN set (2.5%) than for the CG set (0.2%, $p = 1.8 \times 10^{-8}$).

Triggered by the observed enrichment of both DN and SC breakpoints for CGC genes, we investigated the overlap between DN breakpoints and the 68,018 SC breakpoints. We captured all DN breakpoints within a distance of 10 kb from an SC breakpoint and calculated the mean distance to the SC breakpoint for this set. We found that the mean distance of a DN breakpoint to an SC breakpoint is on average 1.15-fold smaller than for random control sets ($p = 0.0047$) (Figures 4E and 4F). Similar results were obtained for the overlap of CG with SC breakpoints ($p = 0.001$) (Figures 4G, S3C, and S3D). Despite the overlap of breakpoint positions between germline and soma, suggestive of local predisposition to genome fragility, we found no common genome characteristics that could explain this observation (Figure S4). Future larger DN data sets may contribute to more insight in the genomic basis of the fragility.

DISCUSSION

Through a combination of functional studies and large-scale genomic analyses of breakpoints from patients with congenital

disease, we made two important observations. First, breakpoints of germline rearrangements and somatic cancer rearrangements overlap each other. Second, de novo germline breakpoints in patients with congenital disease may hit cancer genes and lead to formation of fusion genes similar to those in cancer. In contrast, common germline breakpoints were depleted for cancer genes, which is likely a result of purifying selection of breaks involving protein coding genes (Mills et al., 2011).

The link between genes mutated in cancer and development has been noted before, among others for mutations in *FOXP1* (O’Roak et al., 2011; Talkowski et al., 2012), RAS/MAPK signaling (deregulated in Noonan syndrome) (Cirstea et al., 2010), and the PI3K pathway (deregulated in megalencephaly syndromes) (Fam, 2012). Here, we identify transcriptional deregulation involving three gene fusions involving C19MC microRNAs, *ETV1*, and *FOXP1*. In the latter two cases, the relevance of the fusions to the patient’s phenotype could not be entirely resolved with functional studies. Because *FOXP1* has previously been associated with neurodevelopmental disorders driven by de novo translocation breakpoints, CNVs, and point mutations (O’Roak et al., 2011; Talkowski et al., 2012), it is well possible that the loss of one functional allele of *FOXP1* and not a gain-of-function effect of the observed fusion transcripts in the above-described patient drive the patient’s phenotype.

The observation of a genomic parallel between cancer and germline rearrangements raises the question of cancer predisposition among individuals carrying de novo chromosome rearrangements. For example, rearrangements of the *MYCN* locus have been found to underlie childhood neuroblastoma (Lipska et al., 2013), germline rearrangement of *RUNX1* caused acute myeloid leukemia (Buijs et al., 2012), and 7q22 rearrangements are associated with myeloproliferative disorder (Forrest and Lee, 2002). Also, two patients within our data set contain germline rearrangements of the *RUNX1* gene. Both suffered from leukemia most likely as a result of the *RUNX1* rearrangement. The high incidence of morphological abnormalities among patients with childhood cancer further underscores a potential genetic and mechanistic link between cancer and congenital disease (Bleeker et al., 2014a, 2014b; Merks et al., 2008), possibly driven by involvement of genomic rearrangements in both types of disease, similar to what we describe here. Although the two patients that we phenotyped at the molecular level in this work (aged 25 and 8) do not suffer from cancer at this point, we cannot rule out a predisposition for developing cancer later in life. The phenotypic outcome may be determined by timing (germline or soma) and context (additional mutations) of the rearrangements. For example, somatic rearrangements causing C19MC overexpression were recently shown to drive embryonal tumors with multilayered rosettes (ETMR) (Kleinman et al., 2014). In these examples, the fusion partner is different from the one in patient 2 (*TTHY1* versus *NDUFA3*). Also, five microRNAs in the beginning of the C19MC cluster are not activated in the patient described here but are activated in ETMRs.

Altogether, our results set the stage for further efforts to characterize genome rearrangement mechanisms in human development and disease and show that applying multiple genomics approaches to analyze the in vivo molecular phenotypes provides insights in congenital disease etiology.

EXPERIMENTAL PROCEDURES

Please see [Supplemental Information](#) for [Supplemental Experimental Procedures](#)

Patient Material and Informed Consent

We obtained informed consent for the analysis of DNA and RNA from each patient and their parents. The genetic analysis was performed according to the guidelines of the Medical Ethics Committee of the University Medical Center Utrecht.

Small RNA Sequencing and Analysis

Total RNA was isolated from ± 5 M peripheral blood mononuclear cells (PBMCs). RNA libraries for SOLiD sequencing were prepared using either the RiboMinus Eukaryote Kit for RNA-seq (Life Technologies) (chromothripsis patient and family members) or Ambion Poly(A)Purist (Life Technologies) and mRNA only Eukaryote mRNA Isolation Kit (Epicenter) (C19MC duplication patient and family members). Small RNA Library preparation was done using the SOLiD Total RNA-seq Kit, following the guidelines for small RNA sequencing library preparation (Life Technologies). Sequencing reads were mapped using Burrows-Wheeler Alignment tool (BWA) (Li and Durbin, 2009), and differential expression analyses were performed with DEGseq (Wang et al., 2010).

ChIP Sequencing

H3K4me3 IPs were carried out using the MAGnify system (Invitrogen) following manufacturers' instructions. Sequencing libraries were prepared from double-fragmented DNA, as described by Mokry et al. (2010) and sequenced on SOLiD. Read mapping was done using BWA (Li and Durbin, 2009).

Expression of miRNAs and Fusion Genes in Zebrafish and Cell Lines

Duplex RNA oligonucleotides matching hsa-miR-520c-5p and hsa-miR-520b were injected in 1-cell stage zebrafish embryos at a concentration of 5 μ M.

Fusion genes were cloned into the mammalian expression vector Phage2-EF1alpha-IRES-Puro (Westburg) and transfected into HEK293FT cells. For immunofluorescence, fixed cells were stained with a rabbit polyclonal anti-HA tag antibody (Abcam) followed by a goat anti rabbit secondary antibody conjugated with Alexa Fluor 488 (Life Technologies).

Breakpoint Data from Patients with Congenital Disease and Cancer Samples

We obtained breakpoint data for germline (constitutional), de novo (DN) genomic rearrangements in 96 patients with congenital disorders from published studies and from our own genome sequencing efforts (Tables S2A and S2B). Somatic cancer (SC) breakpoints were derived from published studies (Table S2C). We obtained common germline (CG) deletion breakpoints from 1000 Genomes phase 1 (Mills et al., 2011) and GoNL (Francioli et al., 2014). If coordinates were in hg18, we used the UCSC liftOver tool to convert the coordinates to hg19. Per sample, the breakpoints were ordered by genomic position, and breaks from the same sample occurring within a genomic interval of 2 kb were merged, because these may represent the same double-stranded DNA break (Kloosterman et al., 2012).

Analysis of the Overlap between DN and CG Breakpoints with Protein Coding Genes, Cancer Gene Census Genes, and Breakpoints

To test whether DN breakpoints showed overlap with SC breakpoints and cancer gene census (CGC) genes, we generated 10,000 random breakpoint sets equal in size to the DN breakpoint set ($n = 533$, excluding breakpoints from the two patients described in detail in this study). The random breakpoints in these sets were only taken from positions in the genome that were amenable to structural variation breakpoint calling by next-generation mate-pair sequencing (Kloosterman et al., 2012). Therefore, we compiled a BAM file from six high-quality data sets and required at least 300 uniquely, unambiguously mapped reads (SAM flag $X0 \leq 1$) with no secondary mapping hits (SAM flag $X1 = 0$), no alignment gaps ($XO = 0$), and mapping quality > 0 in the region of 1 kb flanking each side of each simulated breakpoint. This elim-

inated 34% of the random breakpoints that mostly covered repetitive regions such as the centromeres. We matched the sample size and chromosomal distribution to the original DN breakpoint set. Thus, the number of rearrangements per patient, plus the sizes of simple deletions, inversions, or tandem duplications were maintained in the simulated set. Also, the interbreakpoint distance for (chromothripsis) breakpoint clusters were maintained in the simulations to control for nonindependent breaks within the patients. To derive an empirical p value, we computed the overlap of DN and corresponding random breakpoint sets with CGC genes (Futreal et al., 2004) and all protein-coding genes using BEDtools (Quinlan and Hall, 2010). We only counted the overlap of nonindependent breakpoints with CGC genes once. To determine the overlap between DN and SC breakpoints, we calculated the distance for DN breakpoints (and permuted breakpoints) to a nearest SC breakpoint. Subsequently, we captured all DN breakpoints (and permuted breakpoints) within a distance of 10 kb from a SC breakpoint and calculated the mean distance to a cancer breakpoint for this set. The same calculations as for DN breakpoints were also performed for CG breakpoints, based on 1,000 simulated data sets.

We used logistic regression analysis to test for association of CGC genes and protein coding genes with DN and CG breakpoints. Therefore, we used a set of 100,000 control breakpoints and performed regression analysis using standard glm functions in R.

ACCESSION NUMBERS

The mate pair, RNA, and ChIP-sequencing data are available from the European Nucleotide Archive (<http://www.ebi.ac.uk/ena/>) under accession numbers PRJEB5063 and PRJEB3030 (SAMEA1325278).

SUPPLEMENTAL INFORMATION

Supplemental Information includes Supplemental Experimental Procedures, four figures, and four tables and can be found with this article online at <http://dx.doi.org/10.1016/j.celrep.2014.11.022>.

AUTHOR CONTRIBUTIONS

S.v.H., M.S., E.C., and W.P.K. designed the study. S.v.H., W.P.K., E.d.B., and N.L. performed RNA-seq. W.P.K., K.D., E.d.B., M.J., N.L., I.R., V.P., H.B., and M.E.T. performed mate-pair seq and breakpoint validations. S.v.H., K.d.L., and N.L. performed H3K4Me3 ChIP-seq. W.H., S.v.H., and W.P.K. performed fusion gene cloning. E.W.K. performed fusion gene overexpression in cell lines, immunoblotting, and confocal microscopy. W.P.K. and N.A. performed zebrafish experiments. J.K. performed sectioning and staining of zebrafish embryos. A.K.B., S.v.H., and W.P.K. isolated PBMCs. M.J.v.R., W.P.K., S.v.H., A.M., M.S., and S.B. analyzed data. B.M., S.V., S.S., A.C.K., E.I., K.L., V.P., M.E.T., R.H., L.T.v.d.V., and E.v.B. provided patient data and material. S.v.H., M.S., E.C., and W.P.K. wrote the manuscript. All authors contributed to the final version of the manuscript.

ACKNOWLEDGMENTS

We would like to thank A. de Graaff and the Hubrecht Imaging Center for supporting the imaging, M. van Dooren, L. van Zutven, and A. Baas for providing access to patient data, and A. Marozzi for help with artwork. This work was supported by the Cancer Genomics Center and Netherlands Center for Systems Biology programs of the Netherlands Genomics Initiative, the Child Health Priority Program of the University Medical Center Utrecht, an NWO Vici grant (865.13.004) to E.C., a research grant from the Wilhelmina Children's Hospital to W.P.K., and the NIH (GM061354, HD065286, and MH095867 to M.E.T.).

Received: May 26, 2014

Revised: October 20, 2014

Accepted: November 15, 2014

Published: December 11, 2014

REFERENCES

- Bar, M., Wyman, S.K., Fritz, B.R., Qi, J., Garg, K.S., Parkin, R.K., Kroh, E.M., Bendoraitis, A., Mitchell, P.S., Nelson, A.M., et al. (2008). MicroRNA discovery and profiling in human embryonic stem cells by deep sequencing of small RNA libraries. *Stem Cells* 26, 2496–2505.
- Belge, G., Rippe, V., Meiboom, M., Drieschner, N., Garcia, E., and Bullerdiek, J. (2001). Delineation of a 150-kb breakpoint cluster in benign thyroid tumors with 19q13.4 aberrations. *Cytogenet. Cell Genet.* 93, 48–51.
- Bleeker, F.E., Hopman, S.M.J., Merks, J.H.M., Aalfs, C.M., and Hennekam, R.C.M. (2014a). Brain tumors and syndromes in children. *Neuropediatrics* 45, 137–161.
- Bleeker, F.E., Hopman, S.M., and Hennekam, R.C. (2014b). Co-occurrence in body site of malformations and cancer. *Eur. J. Med. Genet.* 57, 480–485.
- Buijs, A., Poot, M., van der Crabben, S., van der Zwaag, B., van Binsbergen, E., van Roosmalen, M.J., Tavakoli-Yaraki, M., de Weerd, O., Nieuwenhuis, H.K., van Gijn, M., et al. (2012). Elucidation of a novel pathogenomic mechanism using genome-wide long mate-pair sequencing of a congenital t(16;21) in a series of three RUNX1-mutated FPD/AML pedigrees. *Leukemia* 26, 2151–2154.
- Chiang, C., Jacobsen, J.C., Ernst, C., Hanscom, C., Heilbut, A., Blumenthal, I., Mills, R.E., Kirby, A., Lindgren, A.M., Rudiger, S.R., et al. (2012). Complex reorganization and predominant non-homologous repair following chromosomal breakage in karyotypically balanced germline rearrangements and transgenic integration. *Nat. Genet.* 44, 390–397, S1.
- Cirstea, I.C., Kutsche, K., Dvorsky, R., Gremer, L., Carta, C., Horn, D., Roberts, A.E., Lepri, F., Merbitz-Zahradnik, T., König, R., et al. (2010). A restricted spectrum of NRAS mutations causes Noonan syndrome. *Nat. Genet.* 42, 27–29.
- Cooper, G.M., Coe, B.P., Girirajan, S., Rosenfeld, J.A., Vu, T.H., Baker, C., Williams, C., Stalker, H., Hamid, R., Hannig, V., et al. (2011). A copy number variation morbidity map of developmental delay. *Nat. Genet.* 43, 838–846.
- Ernst, T., Score, J., Deinger, M., Hidalgo-Curtis, C., Lackie, P., Ershler, W.B., Goldman, J.M., Cross, N.C.P., and Grand, F. (2011). Identification of FOXP1 and SNX2 as novel ABL1 fusion partners in acute lymphoblastic leukaemia. *Br. J. Haematol.* 153, 43–46.
- Fam, H.K. (2012). Caught in the AKT: identification of a de novo pathway in MCAP and MPPH and its therapeutic implications. *Clin. Genet.* 82, 521–522.
- Flor, I., and Bullerdiek, J. (2012). The dark side of a success story: microRNAs of the C19MC cluster in human tumours. *J. Pathol.* 227, 270–274.
- Fornari, F., Milazzo, M., Chieco, P., Negrini, M., Marasco, E., Capranico, G., Mantovani, V., Marinello, J., Sabbioni, S., Callegari, E., et al. (2012). In hepatocellular carcinoma miR-519d is up-regulated by p53 and DNA hypomethylation and targets CDKN1A/p21, PTEN, AKT3 and TIMP2. *J. Pathol.* 227, 275–285.
- Forrest, D.L., and Lee, C.L.Y. (2002). Constitutional rearrangements of 7q22 in hematologic malignancies. a new case report. *Cancer Genet. Cytogenet.* 139, 75–77.
- Francioli, L.C., Menelaou, A., Pulit, S.L., van Dijk, F., Palamara, P.F., Elbers, C.C., Neerincx, P.B.T., Ye, K., Guryev, V., Kloosterman, W.P., et al.; Genome of the Netherlands Consortium (2014). Whole-genome sequence variation, population structure and demographic history of the Dutch population. *Nat. Genet.* 46, 818–825.
- Futreal, P.A., Coin, L., Marshall, M., Down, T., Hubbard, T., Wooster, R., Rahman, N., and Stratton, M.R. (2004). A census of human cancer genes. *Nat. Rev. Cancer* 4, 177–183.
- Golzio, C., Willer, J., Talkowski, M.E., Oh, E.C., Taniguchi, Y., Jacquemont, S., Reymond, A., Sun, M., Sawa, A., Gusella, J.F., et al. (2012). KCTD13 is a major driver of mirrored neuroanatomical phenotypes of the 16p11.2 copy number variant. *Nature* 485, 363–367.
- Hastings, P.J., Lupski, J.R., Rosenberg, S.M., and Ira, G. (2009a). Mechanisms of change in gene copy number. *Nat. Rev. Genet.* 10, 551–564.
- Hastings, P.J., Ira, G., and Lupski, J.R. (2009b). A microhomology-mediated break-induced replication model for the origin of human copy number variation. *PLoS Genet.* 5, e1000327.
- Hermans, K.G., van der Korput, H.A., van Marion, R., van de Wijngaart, D.J., Ziel-van der Made, A., Dits, N.F., Boormans, J.L., van der Kwast, T.H., van Dekken, H., Bangma, C.H., et al. (2008). Truncated ETV1, fused to novel tissue-specific genes, and full-length ETV1 in prostate cancer. *Cancer Res.* 68, 7541–7549.
- Hu, N., Zhang, J., Cui, W., Kong, G., Zhang, S., Yue, L., Bai, X., Zhang, Z., Zhang, W., Zhang, X., and Ye, L. (2011). miR-520b regulates migration of breast cancer cells by targeting hepatitis B X-interacting protein and interleukin-8. *J. Biol. Chem.* 286, 13714–13722.
- Huang, Q., Gumireddy, K., Schrier, M., le Sage, C., Nagel, R., Nair, S., Egan, D.A., Li, A., Huang, G., Klein-Szanto, A.J., et al. (2008). The microRNAs miR-373 and miR-520c promote tumour invasion and metastasis. *Nat. Cell Biol.* 10, 202–210.
- Jeon, I.S., Davis, J.N., Braun, B.S., Sublett, J.E., Roussel, M.F., Denny, C.T., and Shapiro, D.N. (1995). A variant Ewing's sarcoma translocation (7;22) fuses the EWS gene to the ETS gene ETV1. *Oncogene* 10, 1229–1234.
- Kleinman, C.L., Gerges, N., Papillon-Cavanagh, S., Sin-Chan, P., Pramatarova, A., Quang, D.-A.K., Adoue, V., Busche, S., Caron, M., Djambazian, H., et al. (2014). Fusion of TTYH1 with the C19MC microRNA cluster drives expression of a brain-specific DNMT3B isoform in the embryonal brain tumor ETMR. *Nat. Genet.* 46, 39–44.
- Kloosterman, W.P., Tavakoli-Yaraki, M., van Roosmalen, M.J., van Binsbergen, E., Renkens, I., Duran, K., Ballarati, L., Vergult, S., Giardino, D., Hansson, K., et al. (2012). Constitutional chromothripsis rearrangements involve clustered double-stranded DNA breaks and nonhomologous repair mechanisms. *Cell Reports* 1, 648–655.
- Kumar-Sinha, C., Tomlins, S.A., and Chinnaiyan, A.M. (2008). Recurrent gene fusions in prostate cancer. *Nat. Rev. Cancer* 8, 497–511.
- Li, H., and Durbin, R. (2009). Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* 25, 1754–1760.
- Lieber, M.R. (2010). The mechanism of double-strand DNA break repair by the nonhomologous DNA end-joining pathway. *Annu. Rev. Biochem.* 79, 181–211.
- Lipska, B.S., Koczkowska, M., Wierzbza, J., Ploszynska, A., Iliszko, M., Izycka-Swieszewska, E., Adamkiewicz-Drozynska, E., and Limon, J. (2013). On the significance of germline cytogenetic rearrangements at MYCN locus in neuroblastoma. *Mol. Cytogenet.* 6, 43.
- Luo, R., Sanders, S.J., Tian, Y., Voineagu, I., Huang, N., Chu, S.H., Klei, L., Cai, C., Ou, J., Lowe, J.K., et al. (2012). Genome-wide transcriptome profiling reveals the functional impact of rare de novo and recurrent CNVs in autism spectrum disorders. *Am. J. Hum. Genet.* 91, 38–55.
- Merks, J.H.M., Ozgen, H.M., Koster, J., Zwinderman, A.H., Caron, H.N., and Hennekam, R.C.M. (2008). Prevalence and patterns of morphological abnormalities in patients with childhood cancer. *JAMA* 299, 61–69.
- Mills, R.E., Walter, K., Stewart, C., Handsaker, R.E., Chen, K., Alkan, C., Abyzov, A., Yoon, S.C., Ye, K., Cheetham, R.K., et al.; 1000 Genomes Project (2011). Mapping copy number variation by population-scale genome sequencing. *Nature* 470, 59–65.
- Mokry, M., Hatzis, P., de Bruijn, E., Koster, J., Versteeg, R., Schuijers, J., van de Wetering, M., Guryev, V., Clevers, H., and Cuppen, E. (2010). Efficient double fragmentation ChIP-seq provides nucleotide resolution protein-DNA binding profiles. *PLoS ONE* 5, e15092.
- O'Roak, B.J., Deriziotis, P., Lee, C., Vives, L., Schwartz, J.J., Girirajan, S., Karakoc, E., Mackenzie, A.P., Ng, S.B., Baker, C., et al. (2011). Exome sequencing in sporadic autism spectrum disorders identifies severe de novo mutations. *Nat. Genet.* 43, 585–589.
- Oh, S., Shin, S., and Janknecht, R. (2012). ETV1, 4 and 5: An oncogenic subfamily of ETS transcription factors. *Biochim. Biophys. Acta - Rev. Cancer* 1826, 1–12.
- Quinlan, A.R., and Hall, I.M. (2010). BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* 26, 841–842.
- Rajaram, V., Knezevich, S., Bove, K.E., Perry, A., and Pfeifer, J.D. (2007). DNA sequence of the translocation breakpoints in undifferentiated embryonal

- sarcoma arising in mesenchymal hamartoma of the liver harboring the t(11;19)(q11;q13.4) translocation. *Genes Chromosomes Cancer* 46, 508–513.
- Rippe, V., Dittberner, L., Lorenz, V.N., Drieschner, N., Nimzyk, R., Sendt, W., Junker, K., Belge, G., and Bullerdiek, J. (2010). The two stem cell microRNA gene clusters C19MC and miR-371-3 are activated by specific chromosomal rearrangements in a subgroup of thyroid adenomas. *PLoS ONE* 5, e9485.
- 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.
- Stankiewicz, P., and Lupski, J.R. (2010). Structural variation in the human genome and its role in disease. *Annu. Rev. Med.* 61, 437–455.
- Stephens, P.J., Greenman, C.D., Fu, B., Yang, F., Bignell, G.R., Mudie, L.J., Pleasance, E.D., Lau, K.W., Beare, D., Stebbings, L.A., et al. (2011). Massive genomic rearrangement acquired in a single catastrophic event during cancer development. *Cell* 144, 27–40.
- Talkowski, M.E., Rosenfeld, J.A., Blumenthal, I., Pillalamarri, V., Chiang, C., Heilbut, A., Ernst, C., Hanscom, C., Rossin, E., Lindgren, A.M., et al. (2012). Sequencing chromosomal abnormalities reveals neurodevelopmental loci that confer risk across diagnostic boundaries. *Cell* 149, 525–537.
- Tomlins, S.A., Rhodes, D.R., Perner, S., Dhanasekaran, S.M., Mehra, R., Sun, X.-W., Varambally, S., Cao, X., Tchinda, J., Kuefer, R., et al. (2005). Recurrent fusion of TMPRSS2 and ETS transcription factor genes in prostate cancer. *Science* 310, 644–648.
- Tomlins, S.A., Laxman, B., Dhanasekaran, S.M., Helgeson, B.E., Cao, X., Morris, D.S., Menon, A., Jing, X., Cao, Q., Han, B., et al. (2007). Distinct classes of chromosomal rearrangements create oncogenic ETS gene fusions in prostate cancer. *Nature* 448, 595–599.
- Vissers, L.E.L.M., and Stankiewicz, P. (2012). Microdeletion and microduplication syndromes. *Methods Mol. Biol.* 838, 29–75.
- Wang, L., Feng, Z., Wang, X., Wang, X., and Zhang, X. (2010). DEGseq: an R package for identifying differentially expressed genes from RNA-seq data. *Bioinformatics* 26, 136–138.
- Weischenfeldt, J., Symmons, O., Spitz, F., and Korbel, J.O. (2013). Phenotypic impact of genomic structural variation: insights from and for human disease. *Nat. Rev. Genet.* 14, 125–138.
- Wu, S., Huang, S., Ding, J., Zhao, Y., Liang, L., Liu, T., Zhan, R., and He, X. (2010). Multiple microRNAs modulate p21Cip1/Waf1 expression by directly targeting its 3' untranslated region. *Oncogene* 29, 2302–2308.