Poly-Gene Fusion Transcripts and Chromothripsis in Prostate Cancer

Chunxiao Wu,[1] Alexander W. Wyatt,[1] Andrew McPherson,[2] Dong Lin,[1,3] Brian J. McConeghy,[1] Fan Mo,[1] Robert Shukin,[1] Anna V. Lapuk,[1] Steven J. M. Jones,[4] Yongjun Zhao,[4] Marco A. Marra,[4] Martin E. Gleave,[1] Stanislav V. Volik,[1] Yuzhuo Wang,[1,3] S. Cenk Sahinalp,[2] and Colin C. Collins[1]*

[1]Vancouver Prostate Centre and Department of Urologic Sciences, University of British Columbia, Vancouver, BC, Canada
[2]School of Computing Science, Simon Fraser University, Burnaby, BC, Canada
[3]Department of Experimental Therapeutics, BC Cancer Agency, Vancouver, BC, Canada
[4]Canada's Michael Smith Genome Sciences Centre, BC Cancer Agency, Vancouver, BC, Canada

Complex genome rearrangements are frequently observed in cancer but their impact on tumor molecular biology is largely unknown. Recent studies have identified a new phenomenon involving the simultaneous generation of tens to hundreds of genomic rearrangements, called chromothripsis. To understand the molecular consequences of these events, we sequenced the genomes and transcriptomes of two prostate tumors exhibiting evidence of chromothripsis. We identified several complex fusion transcripts, each containing sequence from three different genes, originating from different parts of the genome. One such poly-gene fusion transcript appeared to be expressed from a chain of small genomic fragments. Furthermore, we detected poly-gene fusion transcripts in the prostate cancer cell line LNCaP, suggesting they may represent a common phenomenon. Finally in one tumor with chromothripsis, we identified multiple mutations in the p53 signaling pathway, expanding on recent work associating aberrant DNA damage response mechanisms with chromothripsis. Overall, our data show that chromothripsis can manifest as massively rearranged transcriptomes. The implication that multigenic changes can give rise to poly-gene fusion transcripts is potentially of great significance to cancer genetics.

INTRODUCTION

Complex genome rearrangements, involving two or more breakpoint junctions are frequently observed in cancer (Volik et al., 2003,2006; Stephens et al., 2009; Berger et al., 2011). The resulting structural variation is of great functional consequence due to the potential for disruption of tumor suppressor genes, activation of oncogenes, and generation of fusion transcripts. Fusion transcripts can result in mutual inactivation, bring an oncogenic gene under different regulatory machinery or create fusion proteins, and are considered to be a defining feature of many cancers (Rubin et al., 1988; Maher et al., 2009b). In prostate cancer, multigenic events leading to the creation of key fusion genes which drive disease, such as *TMPRSS2-ERG* have been observed (Tomlins et al., 2005; Berger et al., 2011). These events involved closed chains of breakage and rejoining (CCBR), which were balanced and resulted in minor DNA loss (tens to hundreds of bp). The creation of driver fusions such as *TMPRSS2-ERG*, which is associated with early tumorgenesis (Tomlins et al., 2008; Clark and Cooper, 2009), suggests that CCBR can occur early in prostate cancer development. Interestingly, one such event generated a complex gene fusion involving *TMPRSS2*, *FKPB5*, and *ERG* (Pflueger et al., 2011).

Recent evidence suggests that tens to hundreds of genome rearrangements can be acquired in a single cell cycle, during a process equivalent to the shattering and reassembly of one or more chromosome arms (Stephens et al., 2011). Termed "chromothripsis," this cataclysmic event can promote the development of cancer through simultaneous deletion of tumor suppressors and creation of fusion genes. Chromothripsis has an incidence of 2–3% in a wide range of tumor types, and appears to be associated with poor outcome, at least in multiple myeloma and acute myeloid leukemia (Kloosterman et al., 2011b;

Additional Supporting Information may be found in the online version of this article.

Supported by: Canadian Institutes of Health Research, Prostate Cancer Canada, Prostate Cancer Foundation, The Canadian Prostate Cancer Genome Network, Coalition to Cure Prostate Cancer.

Chunxiao Wu and Alexander W. Wyatt have contributed equally to this work.

*Correspondence to: Colin C. Collins, Vancouver Prostate Centre, 2660 Oak Street, Vancouver, British Columbia, Canada, V6H 3Z6. E-mail: ccollins@prostatecentre.com

Received 28 May 2012; Accepted 26 July 2012

DOI 10.1002/gcc.21999

Published online 25 August 2012 in Wiley Online Library (wileyonlinelibrary.com).

Magrangeas et al., 2011; Stephens et al., 2011; Maher and Wilson, 2012; Rausch et al., 2012). Interestingly, constitutionally acquired complex rearrangements share many features with chromothripsis (Kloosterman et al., 2011a; Liu et al., 2011), albeit with extensive chromosomal duplication and triplication not observed in cancer chromothripsis (Maher and Wilson, 2012). A greater understanding of chromothripsis is therefore likely to have importance beyond cancer genetics.

Although considerable effort is now being expended to understand the mechanisms underpinning chromothripsis (Maher and Wilson, 2012), the impact such a complex genome rearrangement has on the transcriptome is unknown.

We recently reported the first case of chromothripsis in a prostate tumor. In this tumor, we detected expression of two complex fusion transcripts from the chromosomes rearranged during chromothripsis (Lapuk et al., 2012). Each complex transcript contained sequence from three different parts of the genome, but limited clinical sample availability prevented further characterization of the genomic breakpoints and associated transcriptome, and the association remained intriguing speculation. To explore further the fusion gene complexity generated by chromothripsis, we applied a novel algorithm specifically designed to predict fusion transcripts arising from multigenic events to whole genome and transcriptome paired-end sequencing data from a second prostate tumor with suspected chromothripsis.

MATERIALS AND METHODS

Tumor Specimen Collection

Collection of the mouse xenograft tumors LTL-313H and LTL-313B was previously described (Andersen et al., 2010; Watahiki et al., 2011). Tumor 890 was collected from a patient during a radical prostatectomy and snap frozen at the Vancouver General Hospital (Lapuk et al., 2012). The patient signed a formal consent form approved by the Ethics Board. The cell line LNCaP was obtained from Pfizer Global Research and Development, La Jolla Laboratories (Lapuk et al., 2012).

Paired-end Genome and Transcriptome Sequencing

Genomic DNA and total RNA were extracted from tumor tissue and the LNCaP cell line, and the quality of the RNA was assessed with the Agilent Bioanalyzer. Genome and transcriptome sequencing of the samples was performed at BCCA Michael Smith Genome Sciences Centre, Vancouver, BC according to established protocols essentially as described in Shah et al. (2009). Sequence read details are provided in Supporting Information, Table S1. Sequence data are available at NCBI SRA accession number SRP013021.

To generate genomic copy number (CN) profiles, DNA-Seq reads were mapped to the NCBI 36.1 (hg18) human genome reference sequence using MAQ 0.7.174. Genomic CN changes were approximated as the ratio of the average sequencing depth in a given window to the average sequencing depth across the genome, where the average sequencing depth across a window is calculated from the total number of sequenced bases in the given genomic window (of size 10 or 30 kb). These values were transformed into $\log_2$ space. CN profiles were visualized using the Nexus Copy Number software package. Note that for the cell line LNCaP, the CN profile used in Supporting Information, Figure S4 is from a publicly available dataset (Taylor et al., 2010), but visualized with the same software.

Fusion Gene and Genomic Rearrangement Detection

Fusion transcripts and associated genomic breakpoints were identified from RNA-Seq and DNA-Seq using deFuse version 0.4.2 and Comrad version 0.1.3 (McPherson et al., 2011a,b). We used NCBI 36 (hg18) human reference genome and Ensemble release 54 gene annotations for both deFuse and Comrad. Next, we applied the nFuse pipeline to identify complex rearrangements that produce fusion transcripts (McPherson et al., 2012).

PCR Validation

To detect genomic fusion junctions, primers were designed that flanked the predicted fusion position, and PCR reactions were performed to amplify the fusion fragments. The same approach was taken to validate fusion transcripts. All amplification products were sequenced with ABI PRISM 310 Genetic Analyzer using standard techniques to confirm identity. Oligonucleotide primer sequences are listed in Supporting Information, Table S2.

Single Nucleotide Variant Detection

Single nucleotide variants (SNVs) were identified in both LTL-313H and LTL-313B. Each

xenograft line was originally derived from the same patient tumor, but cultured separately since the initial grafting (Watahiki et al., 2011). Therefore, SNVs common to both lines are likely to have arisen in the patient, not during culture in the mouse.

Reads were first aligned onto the hg19 genome by burrows-wheeler aligner (BWA) (DNA-Seq) (Li and Durbin, 2010) or the splicing-aware aligner Tophat (RNA-Seq) (Trapnell et al., 2009). For RNA-Seq, reads with unmapped mate were filtered out (both ends of paired reads have to be mappable). PCR duplicates and optical duplicates were removed by Picard (http://picard.sourceforge.net/) and local realignment around InDels and base quality recalibration was performed by the Genome Analysis Toolkit (GATK). Unifiedgenotyper in GATK was applied to identify SNVs/InDels, and then filtered to obtain high-confidence sites. Because the filtering model in GATK was not perfectly suitable for RNA-Seq data, in parallel we used mpileup in Samtools (Li et al., 2009) to call variants and only considered overlapping sites from both tools as high-confidence. The annotation of variant sites were made by tool AnnoVar (Wang et al., 2010).

To remove sequencing artifacts and common variants, SNV calls were initially filtered against seven benign libraries sequenced in parallel (five transcriptome and two genome). These libraries were derived from in-house adjacent-benign prostate tissue. We chose not to filter against the single nucleotide polymorphism database (dbSNP) to retain rare variants that may contribute to disease. To further select only high confidence calls, we required that an SNV be predicted in both a DNA and RNA sequence library. Network analysis was performed using Ingenuity (IPA) Knowledge Base 9 (Ingenuity®Systems, www.ingenuity.com). The effect of SNVs on protein function was assayed using Condel, SIFT, and Polyphen2 (Ng and Henikoff, 2001; Adzhubei et al., 2010; Gonzalez-Perez and Lopez-Bigas, 2011).

RESULTS

Chromothripsis in Patient-Derived Prostate Tumor Xenograft LTL-313H

During an aCGH screen of chromosome alterations in a panel of five patient-derived prostate tumor xenografts (http://www.livingtumorlab.com), we identified a tumor (LTL-313H) which exhibited evidence of chromothripsis on the CN level [Supporting Information, Fig. S1; see Maher and Wilson's 'signature of chromothripsis' (Maher and Wilson, 2012)]. LTL-313H is a TMPRSS2-ERG+, PTEN-xenograft line derived from a needle biopsy specimen collected pre-treatment (Andersen et al., 2010; Watahiki et al., 2011). Using the Illumina Genome Analyzer II, we sequenced the genome (3.7× haploid coverage) and transcriptome (~71× coverage, Supporting Information, Table S1) of LTL-313H and integrated RNA-Seq and DNA-Seq data using the novel "nfuse" algorithm to predict fusion transcripts resulting from genome rearrangements (McPherson et al., 2012). We combined all predicted fusion transcripts together with the DNA-Seq derived CN profile (Supporting Information, Fig. S1) to produce a map of the LTL-313H tumor (Fig. 1A).

Three salient features in the genome and transcriptome of LTL-313H revealed by paired-end sequencing were consistent with chromothripsis (Stephens et al., 2011; Maher and Wilson, 2012). First, we detected multiple complex rearrangements affecting focal genomic regions. Using PCR and Sanger sequencing, we independently validated 42 genomic rearrangements (which underlie predicted fusion transcripts), informing 84 genomic breakpoints, involving just four Chromosomes: 4, 8, 12, and 20 (Fig. 2B). Second, these genomic breakpoints exhibited pronounced clustering, even between different chromosomes. Breakpoints were tightly associated with CN alterations (Fig. 2B), with over 84% (71/84) of breakpoints mapped to within just 100 kbp of a putative change in CN. In fact, 67% (56/84) of breakpoints were within 25 kbp, despite CN changes occurring on average only once every 3 Mbp. Third, affected chromosomal regions alternated between only two predominant CN states. Within these regions, two clear CN states (high and low) were discernible, with 74% (62/84) breakpoints either mapping to the edges of the high CN state or within the resolution window (10 kbp) of the technology (examples in Supporting Information, Fig. S2). As Stephens et al. (2011) demonstrated these characteristic genomic features are extremely difficult to reconcile with a model where rearrangements are acquired progressively, and potentially arose in a single "catastrophic" event.

We reconstructed an extensive chain of 11 genomic fragments from Chromosomes 8 and 12, joined together by validated genomic rearrangements (Fig. 3). The size of genomic fragments in this predicted chain was variable, ranging from

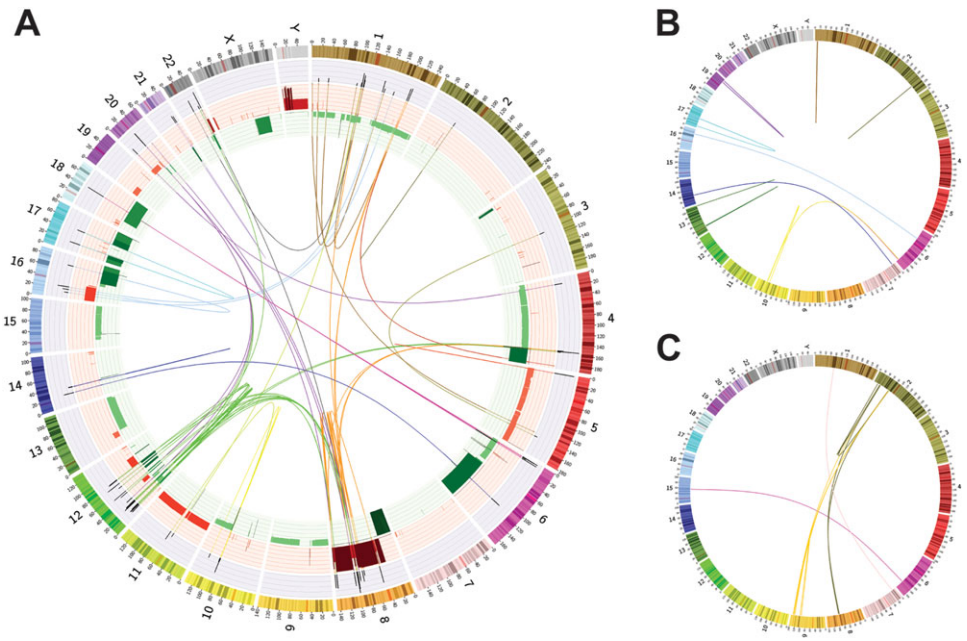

Figure 1. The landscape of transcriptome rearrangement in studied tumors. A) LTL-313H, (B) LNCaP, (C) patient tumor 890. Chromosomes are arranged circularly end-to-end with each chromosome's cytobands marked in the outer ring (Krzywinski et al., 2009). The inner ring in (A) displays CN data inferred from genome sequencing with red indicating gains and green indicating losses. Within the circle, all predicted fusion transcripts are shown as arcs.

400 to 12,000 bp. Similarly, although 16 different gene loci contained >3 neighbouring breakpoints, there was no consistent spacing between them, again implying variable fragment size. Therefore, the genomic landscape of tumor LTL-313H resembled an assortment of different size genomic fragments from four chromosomes joined together, but with the loss of some fragments to the cell (giving rise to the "low" CN states).

LTL-313H Expresses Complex "Poly-gene" Fusion Transcripts

Because we were interested primarily in the impact of chromothripsis on the transcriptome, our analysis focused only on those genome rearrangements predicted to give rise to expressed fusion transcripts. Remarkably, in three different instances, reverse-transcription PCR and Sanger sequencing of the predicted fusion transcripts identified short sequences (<500 bp) interposed between the fusion point identified by paired-end sequencing (Fig. 2A). These can be thought of as "transcriptomic shards," similar to the genomic "shards" previously detected in complex rearranged cancer genomes (also here, e.g., G12, G25: see Supporting Information, Dataset 1) (Stephens et al., 2009). Two of these fusion transcripts were composed of sequence from three different genomic loci, whereas the third contained sequence from four independent loci (Fig. 2). Furthermore, F134 and F132 were mapped to a predicted chain of joined genomic fragments (Fig. 3; Supporting Information, Fig. S3), and indeed were the "scaffolds" that helped validate the chains. Each recombined locus giving rise to a "poly-gene fusion transcript" also gave rise to several other "conventional" two-gene fusion transcripts. Where the underlying genomic rearrangements had left them intact, exon-exon boundaries were preserved, and multiple isoforms of each transcript were expressed. The junction sequences of all validated fusion transcripts and genomic rearrangements (including a further 19 not involving Chromosomes 4,8,12, and 20, e.g., *TMPRSS2-ERG*) are provided in Supporting Information, Dataset 1.

Poly-Gene Fusion Transcripts may be Common in Prostate Cancer

To determine if poly-gene fusion transcripts are a common occurrence, we searched for poly-gene fusion candidates in matched DNA-Seq and RNA-Seq data from a patient's primary prostate tumor (890). The genome of this tumor also bore the hallmarks of chromothripsis (Fig. 1), identified during a next-generation sequencing study of five

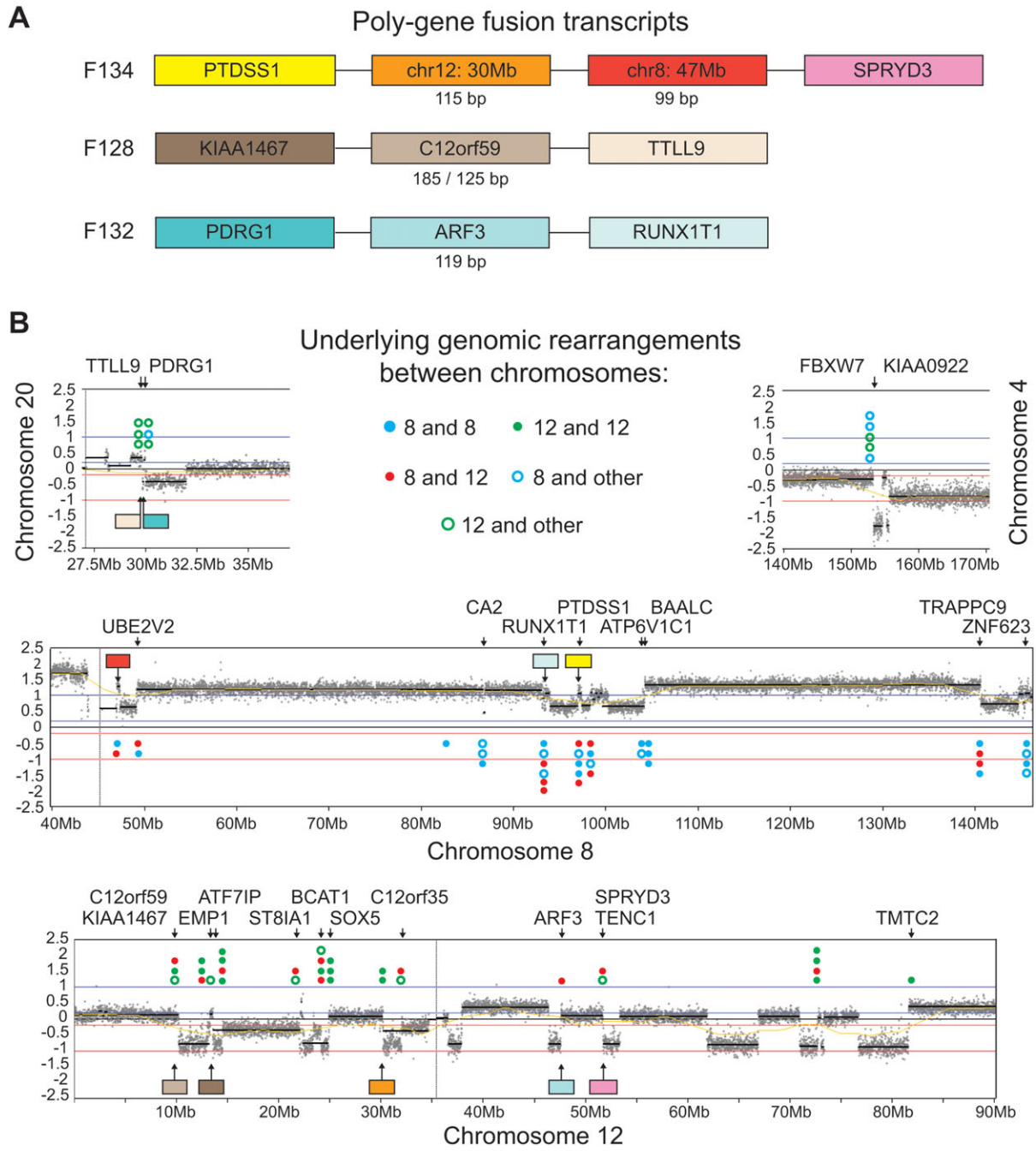

Figure 2. Poly-gene fusion transcripts and their underlying genomic rearrangements identified in the prostate tumor LTL-313H. A) Three fusion transcripts containing sequence mapping to three or more genomic loci were detected. Because the middle sequence in each transcript spanned <220bp, prediction of a conventional "2 gene" fusion transcript was possible from paired-end sequence reads. The middle sequence was then detected at the PCR validation stage. B) The 43 validated genomic rearrangements between Chromosomes 4, 8, 12, and 20, which lead to expressed fusion transcripts are annotated on a DNA-Seq derived CN profile of LTL-313H. Rearrangements, indicated by different color circles, predominantly map to the edges of CN breakpoints. Multiple rearrangements mapped to seemingly the same gene loci, although focal analysis confirmed spatial separation (Supplementary Figure S2). The normal location of sequences expressed in the three poly-gene fusion transcripts are illustrated by colored bars. The Y-axis on chromosome CN profiles indicates $\log_2$ signal intensity.

high-risk primary prostate tumors (Lapuk et al., 2012). Significantly, we identified two transcripts containing sequence mapping to three genes from different loci from Chromosomes 2 and 9 (Supporting Information, Dataset 1). Although limited archival tumor material and nucleic acid degradation prevented systematic identification of all associated genomic rearrangements, we detected

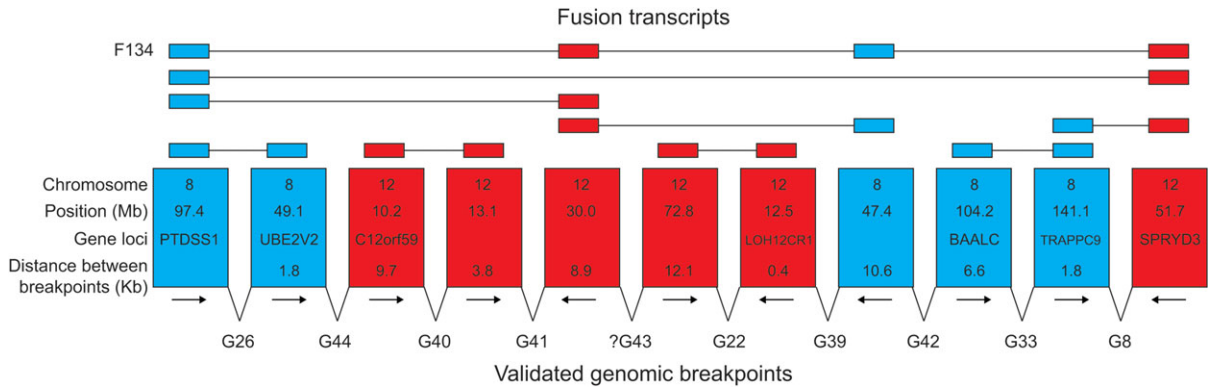

Figure 3. A chain of joined genomic fragments reconstructed using fusion transcripts and their associated genomic breakpoints in LTL-313H. The poly fusion transcript F134 is expressed from this chain. The chromosomal position from which each genomic fragment originated is annotated, together with any gene that mapped to that locus. Predicted fragment lengths are provided, and each fragment's orientation is indicated by an arrow. Note that high sequence ambiguity between two regions of Chromosome 12 prevented validation of G43. [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

multiple isoforms of each poly-gene fusion transcript. Tumor 890 also expressed a fusion involving the E-Twenty Six (ETS) transcription factor *ETV1* (Supporting Information, Dataset 1), which is likely to result in overexpression of an N-terminal truncated isoform of ETV1, similar to previously observed rearrangements (Attard et al., 2008). This meant in total, two out of the 10 independent prostate tumors samples screened [five high-risk patients (Lapuk et al., 2012); five patient-derived xenografts] exhibited evidence of chromothripsis and expressed poly-gene fusion transcripts.

In parallel, we applied our approach to transcriptome sequencing data from the prostate cancer cell line LNCaP (Supporting Information, Table S1), known to express 11 fusion transcripts resulting from genomic rearrangements (Maher et al., 2009a; McPherson et al., 2011b). Our unique fusion gene analysis revealed a previously undetected poly-gene fusion transcript arising from a cluster of breakpoints and rearrangements within six genes from loci of Chromosome 6 and 10 (Fig. 4). This fusion transcript includes exons from *SNX9*, a gene previously reported to be involved in a genomic fusion to UNC5C on Chromosome 4, although no fusion transcript was detected (Mao et al., 2011). Our transcript contained exons from *SNX9* fused to an exon from *CYP2C19* which in turn was fused to exons from *PDE6C* (Fig. 4). The genomic breakpoints underlying this transcript were also validated by PCR and Sanger sequencing (Supporting Information, Table S2; Supporting Information, Dataset 1). Although multiple isoforms were expressed, and a novel exon was generated from a fusion of *CYP2C19* and *PDE6C* sequence, a reading frame was not maintained and the isoforms were not highly expressed. Interestingly, the cluster of CN changes and genomic breakpoints which gave rise to the triple fusion in LNCaP bore focal similarity to chromothripsis (Supporting Information, Fig. S4), although the rest of the LNCaP genome does not share this resemblance. Therefore, although chromothripsis was initially detected in 2–3% of cancers (Stephens et al., 2011), it is possible that very focal chromothriptic events leading to the creation of poly-gene fusion transcripts are much more common, but difficult to detect.

Nonsynonymous SNVs in the p53 Signaling Pathway in LTL-313H

A recent study linked *TP53* mutations with chromothripsis in patients with Sonic Hedgehog medulloblastoma brain tumors (Rausch et al., 2012). To explore the spectrum of mutations in LTL-313H, we searched for nonsynonymous SNVs in our sequencing data. To mitigate concern over SNVs arising subsequent to xenografting, we sequenced the genome and transcriptome of a sister xenograft line known as LTL-313B. Critically, LTL-313B was originally derived from the same primary prostate tumor as LTL-313H (Watahiki et al., 2011), and exhibits the identical chromothripsis scars (Supporting Information, Fig. S1). We detected 643 nonsynonymous SNVs, falling in 555 genes, which were common to both lines and present in both a DNA and RNA library (therefore, detected with high confidence) (Supporting Information, Table S3). Of the 643 SNVs, 538 were present in dbSNP (version 132), while 105 were novel.

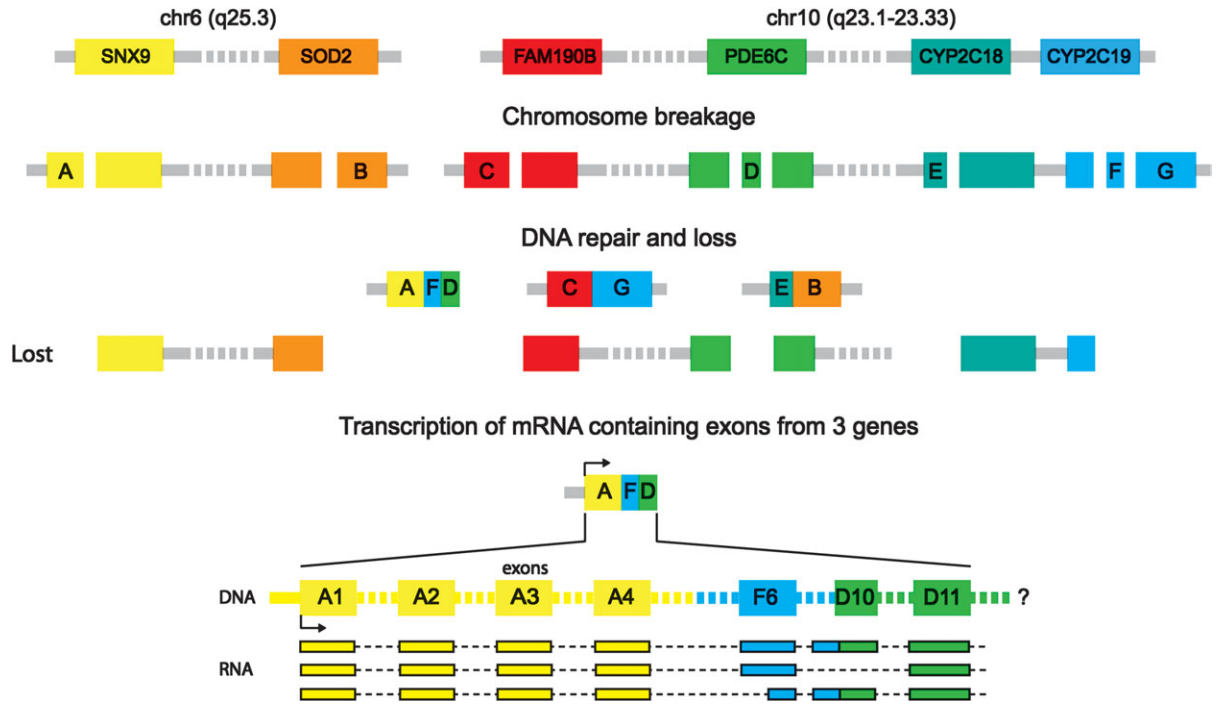

Figure 4. Poly-gene fusion transcripts expressed in the LNCaP prostate cancer cell line. Identification of fusion transcripts and their underlying genomic rearrangements allowed the inference of breakpoints within six genes on Chromosomes 6 and 10. We hypothesize that erroneous DNA repair lead to the fusion of three genomic fragments (A, F, and D) spawned by the breakpoints. Transcription initiation from *SNX9* (fragment A) then produced multiple isoforms of a transcript spanning the three fused fragments. Exon-exon junctions were maintained, and a novel exon was created from the fusion of fragment F to D. Conventional fusion transcripts arise from other fragments spawned by the breakpoints (e.g. expressed from the fusion of fragment C to G). The junction sequences for all fusion transcripts and underlying genomic rearrangements are provided in Supplemental Data. See Supplementary Figure S4 for the CN associated with rearranged regions.

TABLE 1. Nonsynonymous SNVs in the p53 Signaling Pathway in LTL-313H (and LTL-313B)

Gene	SNV location	Ref	Alt	Amino acid	snp132	SIFT	PPH2	Condel	Condel prediction
ATM	Chr11: 108124761	T	C	p.S707P	rs4986761	0.34	0	0.667	Deleterious
ATR	Chr3:142269075	C	T	p.V895M	rs28910271	0.31	0	0.692	Deleterious
ATR	Chr3:142274770	T	C	p.K700E	rs77208665	0.27	0.822	0.812	Deleterious
ATR	Chr3:142281298	C	T	p.V316I	rs28897764	0	0.984	0.992	Deleterious
BRCA1	Chr17:41247921	C	G	p.L204F	rs80357394	0.01	0.265	0.842	Deleterious
PIK3R3	Chr1:46532657	G	T	p.H141N	–	0.01	0.004	0.941	Deleterious
PRKDC[a]	Chr8:48841708	G	A	p.P695S	rs8178046	0.5	0.547	0.545	Deleterious
THBS1	Chr15:39882178	A	G	p.N700S	rs2228262	0.01	0.965	0.977	Deleterious
TOPBP1	Chr3:133368715	G	A	p.R388C	–	0	1	1	Deleterious
TP53[a]	Chr17:7577094	G	A	p.R282W	rs28934574	0	1	1	Deleterious

Scores for the amino acid substitution classification tools SIFT (deleterious < 0.05), Polyphen2 (possibly damaging 0.2–0.85; probably damaging >0.85), and Condel are provided (Adzhubei, et al. 2010; Gonzalez-Perez and Lopez-Bigas, 2011; Ng and Henikoff, 2001).
[a]Indicates that an SNV was homozygous.

Network analysis of the 555 altered genes revealed that the top enriched canonical pathways included p53 signaling ($P = 00164$) (Supporting Information, Table S4). Ten SNVs fell in eight genes in the p53 signaling pathway (Table 1), and although eight of these were present in dbSNP, five had a minor allele frequency of <1%, and two (in *TP53* and *BRCA1*) were flagged as "clinically associated." In fact, rs 28934574 in *TP53* is a somatic mutation in 579 tumors and a germline mutation in 15 families with Li-Fraumeni syndrome (IARC *TP53* database, R15) (Petitjean et al., 2007; Plon et al., 2008). Rs28934574 results in the substitution of an arginine for a tryptophan (p.R282W) in the DNA-binding core domain, causing reduced DNA-binding ability and loss of transactivation activity (Kato et al., 2003; Calhoun and Daggett, 2011).

Furthermore, because CN analysis shows that LTL-313H retains only one copy of *TP53*, there is no wild type *TP53*. Of the two novel nonsynonymous SNVs in the p53 signaling pathway, one fell in a BRCT domain of TOPBP1, a scaffold protein that coordinates activation of the DNA damage checkpoint response (Rappas et al., 2011). The other variant fell in the first SH2 domain of PIK3R3 (Table 1).

We noted that LTL-313H has a homozygous deletion of *FBXW7* (Supporting Information, Fig. S1), a putative tumor suppressor required for regulating levels of MYC in response to DNA damage, and involved in the degradation of the cell cycle G1/S transition protein cyclin-E (CCNE1) (Popov et al., 2007). *FBXW7* was also reported as deleted in the index case of chromothripsis (Stephens et al., 2011).

DISCUSSION

Our results suggest that chromothripsis can give rise to fusion transcripts containing sequence from as many as four genomic loci. Importantly, they were expressed in a patient's prostate tumor as well as two model systems, and their identification in a well-characterized cell line suggests that their presence may not be restricted to genomes bearing obvious hallmarks of chromothripsis.

Poly-gene fusion transcripts arising from chromothripsis appear markedly different to more classical fusion genes generated by balanced translocations and CCBR (Berger et al., 2011; McPherson et al., 2012; Wu et al., 2012). First, CCBRs result in only minor DNA loss (tens to hundreds of bp), compared to chromothripsis which is associated with the loss of large segments of DNA (kbp to Mbp). Second, CCBR is less complex than chromothripsis and involves genes scattered across the genome, rather than during chromothripsis where rearrangements are concentrated within a few chromosomes. Third, transcripts arising from balanced translocations and CCBR are frequently "in-frame," involve important oncogenes and play roles in tumor genesis and progression, while poly-fusion transcripts appear (at least in this study) to be indiscriminate of tumorgenicity or reading frame.

All three tumor samples presented here were positive for an ETS fusion or rearrangement (Tomlins et al., 2007), apparently independent to the chromothripsis event. This is consistent with a recent study by Pflueger et al. (2011) suggesting that ETS fusion genes predispose tumors to further genomic rearrangements. Because the common ETS fusion *TMPRSS2-ERG* is associated with CCBR and early tumorgenesis (Tomlins et al., 2008; Clark and Cooper, 2009), it is likely that they occurred before the chromothriptic event in each tumor presented here. Interestingly, ERG overexpression induces DNA damage (Brenner et al., 2011). Furthermore, recent work by Rausch et al. (2012) associated germline and somatic *TP53* mutations with chromothripsis. In a separate study into early T-cell precursor acute lymphoblastic leukaemia, Zhang et al. (2012) suggested a link between mutations in DNA repair genes and complex genome rearrangements resembling chromothripsis. Therefore, it is possible that chromothripsis occurs in prostate tumor cells which have accrued mutations in the DNA repair pathway, perhaps inflicted by ETS transcription factor overexpression. As such chromothripsis may even have the same initial double strand break clustering as those observed in CCBR and balanced translocations, but a failure of DNA repair mechanisms in advanced tumors lead to a massive increase in the number of rearrangements and the loss of large segments of DNA. The detection of deleterious SNVs in the p53 signaling pathway in LTL-313H appears to support the link between *TP53* mutations and chromothripsis (Rausch et al., 2012); however, mutations in *TP53* are common in prostate cancer (Kumar et al., 2011), and analyses of more chromothripsis tumors is required to better elucidate this potential association. As an aside, several of the "deleterious" nonsynonymous SNVs identified here were present in dbSNP, but with a minor allele frequency of $<1\%$ or flagged as "clinically associated." This has clear implications for disease studies wishing to use dbSNP to filter out "neutral" variants.

It is unclear whether the poly-gene fusions presented here are biologically relevant, as they did not maintain a reading frame. However, exon boundaries were preserved and multiple isoforms of each transcript expressed, suggesting that they are recognized by the spliceosome. Even if no poly-gene fusion transcripts with coherent reading frames exist in these tumors, the possibility that they may have regulatory function, akin to pseudogenes and lncRNAs, remains intriguing (Poliseno et al., 2010; Prensner and Chinnaiyan, 2011). At the very least poly-gene fusion transcripts exist as a phenotypic marker of chromothripsis, and therefore may be important for

sequence-based personalized pathology and monitoring tumor response to therapies.

In conclusion, our data demonstrates that multigenic changes acquired potentially in a single event can give rise to novel transcripts containing exons from three or more different genes. This discovery may have important implications for our understanding of cancer genetics.

ACKNOWLEDGMENTS

We would like to thank Robert H Bell, Sonal Brahmbhatt, Anne Haegert, and Shawn Anderson for helpful discussion and manuscript review.

REFERENCES

Adzhubei IA, Schmidt S, Peshkin L, Ramensky VE, Gerasimova A, Bork P, Kondrashov AS, Sunyaev SR. 2010. A method and server for predicting damaging missense mutations. Nat Methods 7:248–249.

Andersen RJ, Mawji NR, Wang J, Wang G, Haile S, Myung JK, Watt K, Tam T, Yang YC, Banuelos CA, Williams DE, McEwan IJ, Wang Y, Sadar MD. 2010. Regression of castrate-recurrent prostate cancer by a small-molecule inhibitor of the amino-terminus domain of the androgen receptor. Cancer Cell17:535–546.

Attard G, Clark J, Ambroisine L, Mills IG, Fisher G, Flohr P, Reid A, Edwards S, Kovacs G, Berney D, Foster C, Massie CE, Fletcher A, De Bono JS, Scardino P, Cuzick J, Cooper CS. 2008. Heterogeneity and clinical significance of ETV1 translocations in human prostate cancer. Br J Cancer 99:314–320.

Berger MF, Lawrence MS, Demichelis F, Drier Y, Cibulskis K, Sivachenko AY, Sboner A, Esgueva R, Pflueger D, Sougnez C, Onofrio R, Carter SL, Park K, Habegger L, Ambrogio L, Fennell T, Parkin M, Saksena G, Voet D, Ramos AH, Pugh TJ, Wilkinson J, Fisher S, Winckler W, Mahan S, Ardlie K, Baldwin J, Simons JW, Kitabayashi N, MacDonald TY, Kantoff PW, Chin L, Gabriel SB, Gerstein MB, Golub TR, Meyerson M, Tewari A, Lander ES, Getz G, Rubin MA, Garraway LA. 2011. The genomic complexity of primary human prostate cancer. Nature 470:214–220.

Brenner JC, Ateeq B, Li Y, Yocum AK, Cao Q, Asangani IA, Patel S, Wang X, Liang H, Yu J, Palanisamy N, Siddiqui J, Yan W, Cao X, Mehra R, Sabolch A, Basrur V, Lonigro RJ, Yang J, Tomlins SA, Maher CA, Elenitoba-Johnson KS, Hussain M, Navone NM, Pienta KJ, Varambally S, Feng FY, Chinnaiyan AM. 2011. Mechanistic rationale for inhibition of poly(ADP-ribose) polymerase in ETS gene fusion-positive prostate cancer. Cancer Cell 19:664–678.

Calhoun S, Daggett V. 2011. Structural effects of the L145Q, V157F, and R282W cancer-associated mutations in the p53 DNA-binding core domain. Biochemistry 50:5345–5353.

Clark JP, Cooper CS. 2009. ETS gene fusions in prostate cancer. Nat Rev Urol 6:429–439.

Gonzalez-Perez A, Lopez-Bigas N. 2011. Improving the assessment of the outcome of nonsynonymous SNVs with a consensus deleteriousness score, Condel. Am J Hum Genet 88: 440–449.

Kato S, Han SY, Liu W, Otsuka K, Shibata H, Kanamaru R, Ishioka C. 2003. Understanding the function-structure and function-mutation relationships of p53 tumor suppressor protein by high-resolution missense mutation analysis. Proc Natl Acad Sci USA 100:8424–8429.

Kloosterman WP, Guryev V, van Roosmalen M, Duran KJ, de Bruijn E, Bakker SC, Letteboer T, van Nesselrooij B, Hochstenbach R, Poot M, Cuppen E. 2011a. Chromothripsis as a mechanism driving complex de novo structural rearrangements in the germline. Hum Mol Genet 20:1916–1924.

Kloosterman WP, Hoogstraat M, Paling O, Tavakoli-Yaraki M, Renkens I, Vermaat JS, van Roosmalen MJ, van Lieshout S, Nijman IJ, Roessingh W, van 't Slot R, van de Belt J, Guryev V, Koudijs M, Voest E, Cuppen E. 2011b. Chromothripsis is a common mechanism driving genomic rearrangements in primary and metastatic colorectal cancer. Genome Biol 12:R103.

Krzywinski M, Schein J, Birol I, Connors J, Gascoyne R, Horsman D, Jones SJ, Marra MA. 2009. Circos: An information aesthetic for comparative genomics. Genome Res 19:1639–1645.

Kumar A, White TA, MacKenzie AP, Clegg N, Lee C, Dumpit RF, Coleman I, Ng SB, Salipante SJ, Rieder MJ, Nickerson DA, Corey E, Lange PH, Morrissey C, Vessella RL, Nelson PS, Shendure J. 2011. Exome sequencing identifies a spectrum of mutation frequencies in advanced and lethal prostate cancers. Proc Natl Acad Sci USA 108:17087–17092.

Lapuk AV, Wu C, Wyatt AW, McPherson A, McConeghy BJ, Brahmbhatt S, Mo F, Zoubeidi A, Anderson S, Bell RH, Haegert A, Shukin R, Wang Y, Fazli L, Hurtado-Coll A, Jones EC, Hach F, Hormozdiari F, Hajiresouliha I, Boutros PC, Bristow RG, Zhao Y, Marra MA, Fanjul A, Maher CA, Chinnaiyan AM, Rubin MA, Beltran H, Sahinalp SC, Gleave ME, Volik SV, Collins CC. 2012. From sequence to molecular pathology, and a mechanism driving the neuroendocrine phenotype in prostate cancer. J Pathol 227:286–297.

Li H, Durbin R. 2010. Fast and accurate long-read alignment with Burrows–Wheeler transform. Bioinformatics 26:589–595.

Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R. 2009. The sequence alignment/map format and SAMtools. Bioinformatics 25:2078–2079.

Liu P, Erez A, Nagamani SC, Dhar SU, Kolodziejska KE, Dharmadhikari AV, Cooper ML, Wiszniewska J, Zhang F, Withers MA, Bacino CA, Campos-Acevedo LD, Delgado MR, Freedenberg D, Garnica A, Grebe TA, Hernandez-Almaguer D, Immken L, Lalani SR, McLean SD, Northrup H, Scaglia F, Strathearn L, Trapane P, Kang SH, Patel A, Cheung SW, Hastings PJ, Stankiewicz P, Lupski JR, Bi W. 2011. Chromosome catastrophes involve replication mechanisms generating complex genomic rearrangements. Cell 146:889–903.

Magrangeas F, Avet-Loiseau H, Munshi NC, Minvielle S. 2011. Chromothripsis identifies a rare and aggressive entity among newly diagnosed multiple myeloma patients. Blood 118:675–678.

Maher CA, Kumar-Sinha C, Cao X, Kalyana-Sundaram S, Han B, Jing X, Sam L, Barrette T, Palanisamy N, Chinnaiyan AM. 2009a. Transcriptome sequencing to detect gene fusions in cancer. Nature 458:97–101.

Maher CA, Palanisamy N, Brenner JC, Cao X, Kalyana-Sundaram S, Luo S, Khrebtukova I, Barrette TR, Grasso C, Yu J, Lonigro RJ, Schroth G, Kumar-Sinha C, Chinnaiyan AM. 2009b. Chimeric transcript discovery by paired-end transcriptome sequencing. Proc Natl Acad Sci USA 106:12353–12358.

Maher CA, Wilson RK. 2012. Chromothripsis and human disease: Piecing together the shattering process. Cell 148:29–32.

Mao X, Boyd LK, Yanez-Munoz RJ, Chaplin T, Xue L, Lin D, Shan L, Berney DM, Young BD, Lu YJ. 2011. Chromosome rearrangement associated inactivation of tumour suppressor genes in prostate cancer. Am J Cancer Res 1:604–617.

McPherson A, Hormozdiari F, Zayed A, Giuliany R, Ha G, Sun MG, Griffith M, Heravi Moussavi A, Senz J, Melnyk N, Pacheco M, Marra MA, Hirst M, Nielsen TO, Sahinalp SC, Huntsman D, Shah SP. 2011a. deFuse: An algorithm for gene fusion discovery in tumor RNA-Seq data. PLoS Comput Biol 7:e1001138.

McPherson A, Wu C, Hajirasouliha I, Hormozdiari F, Hach F, Lapuk A, Volik S, Shah S, Collins C, Sahinalp SC. 2011b. Comrad: A novel algorithmic framework for the integrated analysis of RNA-Seq and WGSS data. Bioinformatics 27:1481–1488.

McPherson A, Wu C, Wyatt AW, Shah S, Collins C, Sahinalp SC. nFuse: Discovery of complex genomic rearrangements in cancer using high-throughput sequencing. Genome Res 2012.

Ng PC, Henikoff S. 2001. Predicting deleterious amino acid substitutions. Genome Res 11:863–874.

Petitjean A, Mathe E, Kato S, Ishioka C, Tavtigian SV, Hainaut P, Olivier M. 2007. Impact of mutant p53 functional properties on TP53 mutation patterns and tumor phenotype: Lessons from recent developments in the IARC TP53 database. Hum Mutat 28:622–629.

Pflueger D, Terry S, Sboner A, Habegger L, Esgueva R, Lin PC, Svensson MA, Kitabayashi N, Moss BJ, MacDonald TY, Cao X, Barrette T, Tewari AK, Chee MS, Chinnaiyan AM, Rickman DS, Demichelis F, Gerstein MB, Rubin MA. 2011. Discovery of non-ETS gene fusions in human prostate cancer using next-generation RNA sequencing. Genome Res 21:56–67.

Plon SE, Pirics ML, Nuchtern J, Hicks J, Russell H, Agrawal S, Zbuk K, Eng C, Hegde M, Chin EL. 2008. Multiple tumors in a child with germ-line mutations in TP53 and PTEN. N Engl J Med 359:537–539.

Poliseno L, Salmena L, Zhang J, Carver B, Haveman WJ, Pandolfi PP. 2010. A coding-independent function of gene and pseudogene mRNAs regulates tumour biology. Nature 465:1033–1038.

Popov N, Herold S, Llamazares M, Schulein C, Eilers M. 2007. Fbw7 and Usp28 regulate myc protein stability in response to DNA damage. Cell Cycle 6:2327–2331.

Prensner JR, Chinnaiyan AM. 2011. The emergence of lncRNAs in cancer biology. Cancer Discov 1:391–407.

Rappas M, Oliver AW, Pearl LH. 2011. Structure and function of the Rad9-binding region of the DNA-damage checkpoint adaptor TopBP1. Nucleic Acids Res 39:313–324.

Rausch T, Jones DT, Zapatka M, Stutz AM, Zichner T, Weischenfeldt J, Jager N, Remke M, Shih D, Northcott PA, Pfaff E, Tica J, Wang Q, Massimi L, Witt H, Bender S, Pleier S, Cin H, Hawkins C, Beck C, von Deimling A, Hans V, Brors B, Eils R, Scheurlen W, Blake J, Benes V, Kulozik AE, Witt O, Martin D, Zhang C, Porat R, Merino DM, Wasserman J, Jabado N, Fontebasso A, Bullinger L, Rucker FG, Dohner K, Dohner H, Koster J, Molenaar JJ, Versteeg R, Kool M, Tabori U, Malkin D, Korshunov A, Taylor MD, Lichter P, Pfister SM, Korbel JO. 2012. Genome sequencing of Pediatric medulloblastoma links catastrophic DNA rearrangements with TP53 mutations. Cell 148:59–71.

Rubin CM, Carrino JJ, Dickler MN, Leibowitz D, Smith SD, Westbrook CA. 1988. Heterogeneity of genomic fusion of BCR and ABL in Philadelphia chromosome-positive acute lymphoblastic leukemia. Proc Natl Acad Sci USA 85:2795–2799.

Shah SP, Morin RD, Khattra J, Prentice L, Pugh T, Burleigh A, Delaney A, Gelmon K, Guliany R, Senz J, Steidl C, Holt RA, Jones S, Sun M, Leung G, Moore R, Severson T, Taylor GA, Teschendorff AE, Tse K, Turashvili G, Varhol R, Warren RL, Watson P, Zhao Y, Caldas C, Huntsman D, Hirst M, Marra MA, Aparicio S. 2009. Mutational evolution in a lobular breast tumour profiled at single nucleotide resolution. Nature 461:809–813.

Stephens PJ, McBride DJ, Lin ML, Varela I, Pleasance ED, Simpson JT, Stebbings LA, Leroy C, Edkins S, Mudie LJ, Greenman CD, Jia M, Latimer C, Teague JW, Lau KW, Burton J, Quail MA, Swerdlow H, Churcher C, Natrajan R, Sieuwerts AM, Martens JW, Silver DP, Langerod A, Russnes HE, Foekens JA, Reis-Filho JS, van 't Veer L, Richardson AL, Borresen-Dale AL, Campbell PJ, Futreal PA, Stratton MR. 2009. Complex landscapes of somatic rearrangement in human breast cancer genomes. Nature 462:1005–1010.

Stephens PJ, Greenman CD, Fu B, Yang F, Bignell GR, Mudie LJ, Pleasance ED, Lau KW, Beare D, Stebbings LA, McLaren S, Lin ML, McBride DJ, Varela I, Nik-Zainal S, Leroy C, Jia M, Menzies A, Butler AP, Teague JW, Quail MA, Burton J, Swerdlow H, Carter NP, Morsberger LA, Iacobuzio-Donahue C, Follows GA, Green AR, Flanagan AM, Stratton MR, Futreal PA, Campbell PJ. 2011. Massive genomic rearrangement acquired in a single catastrophic event during cancer development. Cell 144:27–40.

Taylor BS, Schultz N, Hieronymus H, Gopalan A, Xiao Y, Carver BS, Arora VK, Kaushik P, Cerami E, Reva B, Antipin Y, Mitsiades N, Landers T, Dolgalev I, Major JE, Wilson M, Socci ND, Lash AE, Heguy A, Eastham JA, Scher HI, Reuter VE, Scardino PT, Sander C, Sawyers CL, Gerald WL. 2010. Integrative genomic profiling of human prostate cancer. Cancer Cell 18:11–22.

Tomlins SA, Rhodes DR, Perner S, Dhanasekaran SM, Mehra R, Sun XW, Varambally S, Cao X, Tchinda J, Kuefer R, Lee C, Montie JE, Shah RB, Pienta KJ, Rubin MA, Chinnaiyan AM. 2005. Recurrent fusion of TMPRSS2 and ETS transcription factor genes in prostate cancer. Science 310:644–648.

Tomlins SA, Laxman B, Dhanasekaran SM, Helgeson BE, Cao X, Morris DS, Menon A, Jing X, Cao Q, Han B, Yu J, Wang L, Montie JE, Rubin MA, Pienta KJ, Roulston D, Shah RB, Varambally S, Mehra R, Chinnaiyan AM. 2007. Distinct classes of chromosomal rearrangements create oncogenic ETS gene fusions in prostate cancer. Nature 448:595–599.

Tomlins SA, Laxman B, Varambally S, Cao X, Yu J, Helgeson BE, Cao Q, Prensner JR, Rubin MA, Shah RB, Mehra R, Chinnaiyan AM. 2008. Role of the TMPRSS2-ERG gene fusion in prostate cancer. Neoplasia 10:177–188.

Trapnell C, Pachter L, Salzberg SL. 2009. TopHat: Discovering splice junctions with RNA-Seq. Bioinformatics 25:1105–1111.

Volik S, Zhao S, Chin K, Brebner JH, Herndon DR, Tao Q, Kowbel D, Huang G, Lapuk A, Kuo WL, Magrane G, De Jong P, Gray JW, Collins C. 2003. End-sequence profiling: Sequence-based analysis of aberrant genomes. Proc Natl Acad Sci USA 100:7696–7701.

Volik S, Raphael BJ, Huang G, Stratton MR, Bignel G, Murnane J, Brebner JH, Bajsarowicz K, Paris PL, Tao Q, Kowbel D, Lapuk A, Shagin DA, Shagina IA, Gray JW, Cheng JF, de Jong PJ, Pevzner P, Collins C. 2006. Decoding the fine-scale structure of a breast cancer genome and transcriptome. Genome Res 16:394–404.

Wang K, Li M, Hakonarson H. 2010. ANNOVAR: Functional annotation of genetic variants from high-throughput sequencing data. Nucleic acids research 38:e164–e164.

Watahiki A, Wang Y, Morris J, Dennis K, O'Dwyer HM, Gleave M, Gout PW. 2011. MicroRNAs associated with metastatic prostate cancer. PLoS One 6:e24950.

Wu C, Wyatt AW, Lapuk AV, McPherson A, McConeghy BJ, Bell RH, Anderson S, Haegert A, Brahmbhatt S, Shukin R, Mo F, Li E, Fazli L, Hurtado-Coll A, Jones EC, Butterfield YS, Hach F, Hormozdiari F, Hajirasouliha I, Boutros PC, Bristow RG, Jones SJ, Hirst M, Marra MA, Maher CA, Chinnaiyan AM, Sahinalp SC, Gleave ME, Volik SV, Collins CC. 2012. Integrated genome and transcriptome sequencing identifies a novel form of hybrid and aggressive prostate cancer. J Pathol 227:53–61.

Zhang J, Ding L, Holmfeldt L, Wu G, Heatley SL, Payne-Turner D, Easton J, Chen X, Wang J, Rusch M, Lu C, Chen SC, Wei L, Collins-Underwood JR, Ma J, Roberts KG, Pounds SB, Ulyanov A, Becksfort J, Gupta P, Huether R, Kriwacki RW, Parker M, McGoldrick DJ, Zhao D, Alford D, Espy S, Bobba KC, Song G, Pei D, Cheng C, Roberts S, Barbato MI, Campana D, Coustan-Smith E, Shurtleff SA, Raimondi SC, Kleppe M, Cools J, Shimano KA, Hermiston ML, Doulatov S, Eppert K, Laurenti E, Notta F, Dick JE, Basso G, Hunger SP, Loh ML, Devidas M, Wood B, Winter S, Dunsmore KP, Fulton RS, Fulton LL, Hong X, Harris CC, Dooling DJ, Ochoa K, Johnson KJ, Obenauer JC, Evans WE, Pui CH, Naeve CW, Ley TJ, Mardis ER, Wilson RK, Downing JR, Mullighan CG. 2012. The genetic basis of early T-cell precursor acute lymphoblastic leukaemia. Nature 481:157–163.