Published in final edited form as:
Cancer Lett. 2013 November 1; 340(2): . doi:10.1016/j.canlet.2013.01.011.

Fusion genes and their discovery using high throughput sequencing

M. Annala[1], **B. Parker**[2], **W. Zhang**[2], and **M. Nykter**[1]

[1]Tampere University of Technology, Tampere, Finland [2]The University of Texas M.D. Anderson Cancer Center, Houston, TX, USA

Abstract

Fusion genes are hybrid genes that combine parts of two or more original genes. They can form as a result of chromosomal rearrangements or abnormal transcription, and have been shown to act as drivers of malignant transformation and progression in many human cancers. The biological significance of fusion genes together with their specificity to cancer cells has made them into excellent targets for molecular therapy. Fusion genes are also used as diagnostic and prognostic markers to confirm cancer diagnosis and monitor response to molecular therapies. High-throughput sequencing has enabled the systematic discovery of fusion genes in a wide variety of cancer types. In this review, we describe the history of fusion genes in cancer and the ways in which fusion genes form and affect cellular function. We also describe computational methodologies for detecting fusion genes from high-throughput sequencing experiments, and the most common sources of error that lead to false discovery of fusion genes.

1. Introduction

1.1. Fusion genes in cancer

Somatic fusion genes are regarded as one of the major drivers behind cancer initiation and progression (reviewed in [1]). The first signs of fusion genes in human cancer were identified in 1960 when a reciprocal translocation between the q-arms of chromosomes 9 and 22 was discovered in over 95% of chronic myelogenous leukemia patients [2, 3]. After two decades the translocation was understood to produce a chimeric *BCR-ABL1* transcript that encoded a constitutively active form of the ABL kinase [4]. At the same time, Burkitt's lymphoma was found to harbor activating fusions between immunoglobulin genes and *MYC* [5, 6, 7]. These initial findings were promptly followed by the discovery of dozens of new fusion genes in human cancers (Table 1). Among hematological malignancies, the identification of a *PML-RARA* fusion in acute promyelocytic leukemia paved the way for an effective tretinoin-based molecular therapy [8, 9], while a *RUNX1-ETO* chimeric protein was found to characterize a subtype of acute myeloid leukemia with prolonged median survival [10]. Success stories among solid cancers included the early discovery of fusions between *EWSR1* and members of the *ETS* transcription factor family in Ewing's sarcoma [11, 12], and the discovery of characteristic *SS18-SSX* fusions in synovial sarcoma [13, 14, 15]. In myxoid liposarcoma, *FUS-DDIT3* and *EWSR1-DDIT3* fusions were found to be

pathognomonic for the disease [16, 17, 18]. Despite these discoveries, fusion positive cases only accounted for a tiny fraction of all solid cancers. This changed in 2005 when fusion genes juxtaposing *TMPRSS2* and members of the *ETS* transcription factor family were found in 70% of prostate cancers [19]. Subsequent discoveries in solid cancers included the discovery of *EML4-ALK* fusions and *CHD7* rearrangements in non-small cell lung cancer [20, 21, 22], *KIAA1549-BRAF* fusions in pediatric glioma [23], *FGFR3-TACC3* fusions in glioblastoma [24, 25], and R-spondin fusions in colon cancer [26]. Some cancers were found to associate with multiple fusion genes that presented in a mutually exclusive manner. For instance, the fusions *TMPRSS2-ERG* and *TMPRSS2-ETV1* are common in prostate cancer, but almost never co-occur in a single tumor [19]. Similarly, the fusion genes SS18-SSX1 and SS18-SSX2 are found in 70% and 30% of synovial sarcoma patients, but never co-occur [27]. In some cases, fusion genes also exhibit mutual exclusivity or co-occurrence with other types of genomic aberrations, as exemplified by the mutual exclusivity of *ETS* fusions and *SPINK1* overexpression in prostate cancer [28]. Mutual exclusivity between two genomic alterations usually implies that the two alterations confer similar contributions to the malignant phenotype, and therefore oncogenic selection ceases after one alteration has been acquired.

Some fusion genes are found recurrently in multiple cancers. The *BCR-ABL1* fusion gene is found recurrently in both chronic myelogenous leukemia [3] and acute lymphocytic leukemia [29], and isolated cases have been reported in other leukemias. *TPM3-ALK* fusions provide an example of a fusion gene found in cancer cells of completely different lineages. *TPM3-ALK* is found in 15% of cases of anaplastic large cell lymphoma, a hematological malignancy of T-cell origin [30], and in 50% of inflammatory myofibroblastic tumors, solid cancers of myofibroblast origin [31]. More fusion genes involving alternative fusion partners of *ALK* are found in other cancers, including *EML4-ALK* in non-small cell lung cancer [32] and NPM1-ALK in anaplastic large cell lymphoma [33].

Because somatic fusion genes are only found in cancer cells, they are excellent targets for therapeutics and personalized medicine. Indeed, many known fusion genes are already used as drug targets. Examples include the treatment of *BCR-ABL1* positive leukemia patients with the ABL kinase inhibitor imatinib [34], and the treatment of *EML4-ALK* positive non-small cell lung cancer patients with ALK inhibitor crizotinib [32]. However, it must be noted that existing drugs do not target fusion proteins specifically, but instead only target protein domains of one of the genes participating in a fusion. This means that even the latest targeted drugs can have off-target effects on healthy cells that express the target proteins. Fusion genes have also been employed as diagnostic and prognostic markers. For example, detection of *BCR-ABL1* transcripts is used to confirm chronic myelogenous leukemia diagnoses, and transcript levels are followed throughout treatment to monitor for loss of therapeutic response [35].

1.2. Biological impact of fusion genes

Fusion genes can affect cell function through a number of mechanisms. One common mechanism is the overexpression of an oncogene through promoter exchange. In such cases, the 3′ gene participating in the fusion is overexpressed when a chromosomal rearrangement brings the 3′ gene's expression under the control of the 5′ gene's promoter. For example, the overexpression of ETS transcription factors in prostate cancer is caused by their fusion with the androgen regulated *TMPRSS2* promoter [19]. The overexpressed ETS proteins migrate to the nucleus and drive an anaplastic transformation by dysregulating the expression of genes associated with normal prostate epithelial differentiation [36]. Similarly, B cell lymphomas are characterized by chromosomal abnormalities where the promoter of an immunoglobulin heavy locus is fused with the *MYC* proto-oncogene [7]. A fusion event can

also change the expression level of an oncogene by replacing its 3′-UTR, leading to altered regulation of the 5′ gene when the original 3′-UTR microRNA binding sites are lost. *MYB-NFIB* fusions in adenoid cystic carcinoma produce elevated MYB protein levels due to the loss of miR-15a/16 and miR-150 binding sites in *MYB-NFIB* transcripts, and ultimately lead to the activation of MYB target genes and oncogenic pathways [37]. In glioblastoma, the loss of miR-99a binding sites in *FGFR3-TACC3* transcripts allows for significantly higher expression of *FGFR3-TACC3* than wild type *FGFR3* [25].

Another mechanism by which fusion genes alter cellular function is through the formation of chimeric proteins. Altered protein structure may render a chimeric protein constitutively active, lead it to activate alternative downstream targets, or sabotage a critical cellular function. For example, *ALK* fusion genes in anaplastic large cell lymphoma involve 5′ partner genes that harbor dimerization domains that promote ALK dimerization and autophosphorylation, rendering ALK constitutively active. The autophosphorylated ALK kinases then activate oncogenic pathways such as the MAPK, JAK3-STAT3 and PI3K-AKT pathways (reviewed in [38]). In leukemias, *BCR-ABL1* fusions constitutively activate the ABL1 kinase by enabling BCR-ABL1 oligomerization via the coiled coil domain present in BCR [39]. In glioblastoma, chimeric FGFR3-TACC3 proteins display constitutive phosphorylation and trigger aneuploidy by interfering with mitotic fidelity [24].

Not all fusion genes necessarily have biological impact. Cancer genomes are often heavily rearranged and contain pairs of genes that have fused together at random. Fusions involving an inactive 5′ promoter create fusion genes that are not transcribed. Such fusions can be phenotypically neutral if the 3′ gene is inactive or encodes a redundant protein. Alternatively, a fusion event between two genes can disrupt the structure or expression of a gene, leading to significant loss of function [40, 41].

2. Characteristics of fusion genes

2.1. Formation of fusion genes through chromosomal rearrangement

The formation of fusion genes in cells can occur through multiple mechanisms. In the most common scenario, a fusion gene is formed via somatic chromosomal rearrangement. The four basic types of chromosomal rearrangement are deletions, translocations, tandem duplications, and inversions (Fig. 1). A fusion gene can arise via deletion when a genomic region between two genes located on the same strand is deleted (Fig. 1). The *TMPRSS2-ERG* fusion in prostate cancer is an example of a fusion that results from a 2.7 Mb deletion on chromosome 21 [42]. Interestingly, fusion genes can also arise from tandem duplication, a type of chromosomal rearrangement where a genomic region is duplicated one or more times, and the copies are tiled next to the original region. When the amplicon breakpoints are situated near existing genes, this can result in the formation of a fusion gene at the junction of the copied and original region (Fig. 1). Examples of fusion genes formed through tandem duplication include *KIAA1549-BRAF* fusions in pilocytic astrocytoma [23], *FGFR3-TACC3* fusions in glioblastoma [25], and *C2orf44-ALK* fusions in colorectal cancer [43]. A tandem duplication or deletion is likely the cause when two genes located on the same chromosomal strand are fused. The order of the two genes in the fusion transcript is also a helpful clue, as tandem duplication creates chimeric transcripts where the genes are in reverse order relative to their positions on the strand.

Occasionally fusion genes arise via inversion events where chromosomal segments are flipped around (Fig. 1). For example, the *EML4-ALK* fusion gene in non-small cell lung cancer results from a 12 Mb inversion on chromosome 2 [20]. If a fusion gene involves two genes located on opposite strands of a chromosome, there is suitable cause to suspect an inversion event. The genes can face inward or outward; an inversion in either scenario can

lead to a fusion gene. A characteristic feature of this class of fusion is the formation of reciprocal fusion genes at both ends of the inversion (Fig. 1) [20, 44]. However, depending on the properties of the promoters involved, one or both reciprocal fusions may not be transcribed, rendering them impossible to detect through transcriptome sequencing.

In addition to chromosomal rearrangements involving genes on the same chromosome, many fusion genes involve genes located on separate chromosomes. Such fusions are always caused by a translocation of some kind, whether it involves the translocation of a small genomic fragment to a new locus, or a reciprocal translocation involving the swapping of entire chromosome arms (Fig. 1). Examples of fusion genes caused by translocations include the *BCR-ABL1* fusion, formed by a reciprocal translocation between 9q and 22q [4] More complex rearrangements are also possible but less frequent [45].

The genomic breakpoints of fusion genes usually occur in intronic or intergenic regions, and rarely disrupt coding sequences. This phenomenon may be partly explained by introns being 35 times longer than exons on average [46]. Oncogenic selection may also play a role, as fusions that disrupt an exon have a two-in-three chance of creating a frameshifted protein with little effect on cellular function. Conversely, intronic breakpoints often lead to in-frame chimeric proteins because exons tend to terminate at codon boundaries [47, 48, 49]. Despite this bias for intronic breakpoints, isolated cases of exon disrupting breakpoints have been reported [25, 50, 51].

A characteristic feature of many fusion-generating chromosomal rearrangements is the presence of sequence microhomology at rearrangement breakpoints. A study of 40 RAF gene fusions in low-grade glioma found that 85% harbored microhomology at or near the breakpoints [45]. The microhomologies ranged in length between 1–6 bp and were significantly more common than expected by chance. This pattern is characteristic of microhomology-mediated break-induced replication (MMBIR), implying that MMBIR may be a major causative mechanism behind many fusion events [45]. Another study that looked at *TMPRSS2-ETS* breakpoints in prostate cancer also found evidence of microhomology, but implicated non-homologous end joining (NHEJ) as the driving mechanism behind the chromosomal rearrangements [52].

2.2. Read-through fusions and splicing

A particular class of fusion genes known as *read-through* chimeras can arise in the absence of any DNA level alterations. This type of fusion gene forms when an RNA polymerase does not properly terminate transcription at the end of a gene, but instead continues transcribing until the end of the next gene (Fig. 2). The chimeric pre-mRNA is spliced to produce a fusion transcript. In almost all cases, the resulting chimeric mRNA will lack the last exon of the upstream gene, and the first exon of the downstream gene. This phenomenon occurs because the last exon of a gene lacks a splicing donor site that is required for spliceosome function. Similarly, the first exon of a gene lacks a splicing acceptor site (Fig. 2). Due to the lack of these splicing sites, both exons are spliced out of the mRNA transcript [53]. Since the stop codon of a protein-coding gene is usually found in the last exon, the splicing of the last and first exons can lead to the formation of a functional chimeric protein (Fig. 2). The reason for the stop codon's preferential localization to the last exon of a gene is the avoidance of non-sense mediated decay, a cellular safety mechanism that degrades mRNAs whose coding sequence terminates prematurely before the last exon [54].

Last and first exon skipping can also occur with fusion genes that arise from chromosomal rearrangements. In this way a rearrangement can produce a functional fusion protein even though one or both genomic breakpoints localize to intergenic regions. Consider a case where two genes A and B are located on the same chromosomal strand, and a deletion event

removes the region between the two genes. Further, consider that the breakpoint in the upstream gene A is located in an intron, while the other breakpoint is located 20 kb upstream of gene B. Surprisingly, such a fusion gene can encode a functional chimeric protein, as the first exon of gene B is spliced out of the pre-mRNA (Fig. 3). Similar reasoning applies to the case where one breakpoint is located downstream of gene A, and the other breakpoint in an intron of gene B (Fig. 3). In fact, a functional fusion protein may arise even if both breakpoints are located in intergenic regions outside genes A and B. Actual examples of exon skipping in fusion genes caused by chromosomal rearrangement include first exon skipping in *BCR-ABL1* fusions [55] and last exon skipping in *FGFR3-TACC3* fusions [25].

3. Fusion gene discovery through sequencing

3.1. Identification of fusion genes via transcriptome sequencing

High throughput sequencing has transformed the field of cancer genomics by enabling affordable sequencing of entire cancer genomes and transcriptomes. Current methods of high throughput sequencing are based on an approach where DNA is sheared into short fragments that are sequenced in millions of parallel chemical reactions. Highly accurate instruments track the reactions and report them as millions of short nucleotide strings, also known as reads. Computational algorithms are then used to assemble reads into longer contiguous sequences, quantify reads originating from different genomic regions, or identify evidence for putative genomic alterations. In 2009, Maher et al. published two reports on the application of single end and paired end transcriptome sequencing to the problem of fusion gene discovery in human cancers [56, 57]. In paired end sequencing, double stranded DNA fragments are sequenced at both ends, producing paired end reads consisting of two mate sequences. Paired end sequencing is great for identifying chromosomal rearrangements because paired end reads have an effective length equal to the fragment size, which is often far longer than the combined length of the mates. The initial reports by Maher et al. were followed by a cascade of studies exploring the presence of fusion genes in cancers using high throughput sequencing. The amount of interest on this topic has led to the development of multiple open source software tools and pipelines that simplify the computational task of identifying novel fusion genes amidst millions of sequencing reads [58, 59, 60, 61, 62, 63] (Table 2).

Most algorithms for fusion gene discovery are based on the idea of using paired end reads to identify cDNA fragments that combine parts from two genes. Such algorithms typically begin with a pre-filtering step where paired end reads are aligned against a reference genome and transcriptome. A paired end read is said to align *concordantly* when the mates align within a short distance of one another and in the correct orientation relative to one another (Fig. 4). When a paired end read aligns concordantly to the sequence of a chromosome or transcript, the read is considered to originate from ordinary transcriptional activity and is discarded from further analysis. The remaining *discordantly* aligned mate pairs fall into two groups: ones where the mates align to distant sites, and ones where one or both mates fail to align. Some algorithms then trim the unaligned mates to a shorter length and realign them against the genome, with the goal of increased sensitivity [62] (Fig. 4).

At this point, the fusion gene discovery algorithm has compiled a list of discordantly aligned mate pairs. The next step is to use the discordant pairs to nominate fusion candidates, and then to validate them by searching for individual reads that overlap the fusion junction (Fig. 4). The implementation of the validation step varies between algorithms. The algorithms Defuse [60] and Comrad [61] look at unaligned reads whose mate aligned near a putative junction, and try to align these reads against the junction's neighborhood using dynamic programming. FusionSeq [58] builds a list of all possible junction sequences and realigns against them. ChimeraScan [62] and ShortFuse [59] use existing transcriptome annotations

to identify the most likely exon-exon junctions and then realign against their sequences. Since ChimeraScan and ShortFuse look only at splice sites, they cannot find junction-overlapping reads for fusion genes that disrupt exons. The identification of junction-overlapping reads allows a junction's location to be identified with single-base precision and provides important evidence about the validity of a fusion candidate.

After refining the list of discordantly aligned mate pairs to a list of fusion candidates with varying levels of supporting evidence, additional filtering steps are applied to discard candidates that do not represent true fusion genes. Such filters are necessary because the human genome contains vast amounts of repetitive sequence that can complicate read alignment and result in thousands of falsely reported fusion genes. One approach is to filter fusion candidates based on the number of mate pairs and reads that span a fusion junction, as fusion candidates with few supporting reads often represent sequencing errors or misalignment [62]. The number of supporting reads can also be compared with the expression level of the involved genes, in order to filter out fusion candidates where the fusion junction is supported by more reads than would be expected based on gene expression [58]. Some algorithms filter out fusion candidates for which the supporting reads are not aligned evenly on both sides of a fusion junction, or do not overlap a sufficiently large region on one side of the junction [60, 63].

3.2. Identification of fusion genes via genome sequencing

Fusion genes arising from chromosomal rearrangements can also be identified using whole genome sequencing, although this approach has not been widely adopted due to a significantly higher cost per sample. A major benefit of whole genome sequencing is that it can detect fusion genes where only the promoter region of one gene is fused to the exons of another gene. However, this approach has the downside that it cannot detect read-through fusions and cannot determine the level at which a fusion gene is transcribed into chimeric transcripts. Some studies have adopted combined genome and transcriptome sequencing in order to achieve the best of both worlds [64]. The fusion gene discovery software Comrad [61] was designed for use with such combined sequencing data. The use of genome sequencing enables the direct identification of genomic breakpoints, a task that previously required careful primer design followed by capillary sequencing.

3.3. Technical artifacts that mimic fusion genes

The construction of a complementary DNA (cDNA) library for transcriptome sequencing is a complex process that involves multiple steps. Some of the steps are known to cause technical artifacts such as chimeric cDNA sequences that combine parts of two unrelated RNA sequences. One source of false chimeras is the reverse transcription step where cDNA is synthesized from RNA templates. Reverse transcriptase enzymes are prone to template switching, an event where the enzyme jumps to another template without terminating DNA synthesis [65]. Template switching has been proposed as an explanation for the anomalous chimeric transcripts that show up in transcriptome sequencing but are not supported by DNA level alterations [65]. Another potential source of false chimeras is the PCR amplification step where cDNA fragments are amplified to increase the amount of DNA available for sequencing. PCR chimeras have been proposed to arise when incomplete elongation occurs during a PCR cycle and the incomplete product partially hybridizes with an unrelated template, followed by chimeric elongation (Fig. 5) (reviewed in [66]). False chimeras are enriched among highly transcribed genes such as ribosomal RNA (rRNA), small nuclear RNA (snRNA), and small nucleolar RNA (snoRNA). For this reason, many fusion gene discovery algorithms include a filtering step where fusions involving blacklisted, highly expressed genes are filtered out [58]. Many algorithms also attempt to filter out PCR

chimeras by discarding candidate fusions where all supporting mate pairs are identical PCR duplicates and do not properly cover both sides of the fusion junction [58, 60].

3.4. Polymorphic fusion genes

Some fusion genes are present in the germline of a subset of the human population. Examples of such fusion genes include the *TFG-GPR128* fusion that is formed by a 111 kb tandem duplication and was found to be present in 2% of the healthy human population [67]. Such fusion genes are called *polymorphic*. It should be noted that even though polymorphic fusion genes in the current population are rare, the literature contains numerous examples of proteins that have originated from gene fusions somewhere along the evolutionary history of humans [68, 69] and other species [70].

3.5. Inter-sample contamination

Another rare but potential issue in fusion gene discovery is the impact of nucleic acid contamination between samples. One of our analyses using whole transcriptome sequencing data from the Cancer Genome Atlas GBM project identified two *FGFR3-TACC3* positive samples and a batch of samples that had been contaminated with *FGFR3-TACC3* transcripts from one true fusion positive sample. Twenty-one samples in the batch presented reads overlapping the fusion junction, but twenty of the samples showed thousand-fold less *FGFR3-TACC3* expression than the one true fusion positive sample. None of the twenty contaminated samples showed overexpression of *FGFR3* or *TACC3* exons, whereas the two true fusion positive samples did (Fig. 6). Additionally, samples within the contaminated batch exhibited evidence of only one fusion variant, although *FGFR3-TACC3* fusions are known to exhibit heterogeneity with regards to fusion structure and involved exons [24, 25]. In line with the reported heterogeneity, the fusion positive sample in the non-contaminated batch harbored an alternative, longer fusion variant (Fig. 6). If the level of contamination is low, false fusion discoveries can be avoided by discarding fusion candidates with an insufficient amount of overlapping reads.

4. Conclusions and future perspectives

The study of fusion genes in the context of human cancer has resulted in many important discoveries, some of which have subsequently led to the development of novel and effective molecular therapeutics. High throughput sequencing has made it possible to characterize all DNA and RNA level alterations in cancer cells, enabling the efficient cataloging of fusion genes that drive human cancers. The discovery of rare fusion genes in subpopulations of cancer patients is paving the way for a more personalized form of medicine where treatments are tailored to the molecular characteristics of each individual patient. Due to their specificity to cancer cells, fusion genes represent ideal targets for such tailored molecular therapy. Additionally, if a fusion gene is discovered in multiple cancers of different lineages, existing drugs and molecular therapies can be quickly adopted for the treatment of the new cancer.

Most current studies aimed at fusion gene discovery have based their results on whole transcriptome sequencing, as this option is cheaper and more sensitive at identifying fusion genes involving exons from two different genes. However, chromosomal rearrangements that swap the promoter of a gene cannot be detected through transcriptome sequencing. Thus considerable amounts of new biology may yet be discovered as the price of whole genome sequencing falls to a level where more laboratories will begin employing it in large scale characterization of cancers.

Acknowledgments

This review was funded by the Sigrid Juselius Foundation, and the Finnish Funding Agency for Technology and Innovation Finland Distinguished Professor programme. We apologize to investigators whose reports were not cited in this manuscript due to space limitation.

References

1. Mitelman F, Johansson B, Mertens F. The impact of translocations and gene fusions on cancer causation. Nature Reviews Cancer. 2007; 7:233–245.
2. Nowell PC, Hungerford DA. A minute chromosome in human chronic granulocytic leukemia. Science. 1960; 142:1497.
3. Rowley JD. A new consistent chromosomal abnormality in chronic myelogenous leukaemia identified by quinacrine fluorescence and Giemsa staining. Nature. 1973; 243:290–293. [PubMed: 4126434]
4. Shtivelman E, Lifshitz B, Gale RP, Canaani E. Fused transcript of abl and bcr genes in chronic myelogenous leukaemia. Nature. 1985; 315:550–554. [PubMed: 2989692]
5. Manolov G, Manolova Y. Marker band in one chromosome 14 from Burkitt lymphomas. Nature. 1972; 237:33–34. [PubMed: 4113130]
6. Zech L, Haglund U, Nilsson K, Klein G. Characteristic chromosomal abnormalities in biopsies and lymphoid-cell lines from patients with Burkitt and non-Burkitt lymphomas. International Journal of Cancer. 1976; 17:47–56.
7. Dalla-Favera R, Bregni M, Erikson J, Patterson D, Gallo RC, Croce CM. Human c-myc oncogene is located on the region of chromosome 8 that is translocated in Burkitt lymphoma cells. Proceedings of the National Academy of Science. 1982; 79:7824–7827.
8. Borrow J, Goddard AD, Sheer D, Solomon E. Molecular analysis of acute promyelocytic leukemia breakpoint cluster region on chromosome 17. Science. 1990; 249:1577–1580. [PubMed: 2218500]
9. Warrell RP, Frankel SR, Miller WH, Scheinberg DA, Itri LM, Hittelman WN, Vyas R, Andreeff M, Tafuri A, Jakubowski A, et al. Differentiation therapy of acute promyelocytic leukemia with tretinoin (all-trans retinoic acid). New England Journal of Medicine. 1991; 324:1385–1393. [PubMed: 1850498]
10. Erickson P, Gao J, Chang KS, Look T, Whisenant E, Raimondi S, Lasher R, Trujillo J, Rowley J, Drabkin H. Identification of breakpoints in t(8;21) acute myelogenous leukemia and isolation of a fusion transcript, AML1/ETO, with similarity to Drosophila segmentation gene, runt. Blood. 1992; 80:1825–1831. [PubMed: 1391946]
11. Turc-Carel C, Philip I, Berger MP, Philip T, Lenoir GM. Chromosomal translocations in Ewing's sarcoma. New England Journal of Medicine. 1983; 309:497–498.
12. Aurias A, Rimbaut C, Buffe D, Dubousset J, Mazabraud A. Chromosomal translocations in Ewing's sarcoma. New England Journal of Medicine. 1983; 309:496–497.
13. Turc-Carel C, Dal Cin P, Limon J, Rao U, Li FP, Corson JM, Zimmerman R, Parry DM, Cowan JM, Sandberg AA. Involvement of chromosome X in primary cytogenetic change in human neoplasia: nonrandom translocation in synovial sarcoma. Proc Natl Acad SciUSA. 1987; 84:1981–1985.
14. Smith S, Reeves BR, Wong L, Fisher C. A consistent chromosome translocation in synovial sarcoma. Cancer Genetics and Cytogenetics. 1987; 26:179–180. [PubMed: 3030536]
15. Clark J, Rocques PJ, Crew AJ, Gill S, Shipley J, Chan AM-L, Gusterson BA, Cooper CS. Identification of novel genes, SYT and SSX, involved in the t(X;18)(p11.2;q11.2) translocation found in human synovial sarcoma. Nature Genetics. 1994; 7:502–508. [PubMed: 7951320]
16. Crozat A, Aman P, Mandahl N, Ron D. Fusion of CHOP to a novel RNA-binding protein in human myxoid liposarcoma. Nature. 1993; 363:640–644. [PubMed: 8510758]
17. Rabbitts TH, Forster A, Larson R, Nathan P. Fusion of the dominant negative transcription regulator CHOP with a novel gene FUS by translocation t(12;16) in malignant liposarcoma. Nature Genetics. 1993; 4:175–180. [PubMed: 7503811]

18. Antonescu CR, Tschernyavsky SJ, Decuseara R, Leung DH, Woodruff JM, Brennan MF, Bridge JA, Neff JR, Goldblum JR, Ladanyi M. Prognostic impact of P53 status, TLS-CHOP fusion transcript structure, and histological grade in myxoid liposarcoma: a molecular and clinicopathologic study of 82 cases. Clinical Cancer Research. 2001; 7:3977–3987. [PubMed: 11751490]

19. 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. Recurrent fusion of TMPRSS2 and ETS transcription factor genes in prostate cancer. Science. 2005; 310:644–648. [PubMed: 16254181]

20. Soda M, Choi YL, Enomoto M, Takada S, Yamashita Y, Ishikawa S, Fujiwara S, Watanabe H, Kurashina K, Hatanaka H, Bando M, Ohno S, Ishikawa Y, Aburatani H, Niki T, Sohara Y, Sugiyama Y, Mano H. Identification of the transforming EML4-ALK fusion gene in non-small-cell lung cancer. Nature. 2007; 448:561–566. [PubMed: 17625570]

21. Rikova K, Guo A, Zeng Q, Possemato A, Yu J, Haack H, Nardone J, Lee K, Reeves C, Li Y, Hu Y, Tan Z, Stokes M, Sullivan L, Mitchell J, Wetzel R, Macneill J, Ren JM, Yuan J, Bakalarski CE, Villen J, Kornhauser JM, Smith B, Li D, Zhou X, Gygi SP, Gu TL, Polakiewicz RD, Rush J, Comb MJ. Global survey of phosphotyrosine signaling identifies oncogenic kinases in lung cancer. Cell. 2007; 131:1190–1203. [PubMed: 18083107]

22. Pleasance ED, Stephens PJ, O'Meara S, McBride DJ, Meynert A, Jones D, Lin ML, Beare D, Lau KW, Greenman C, Varela I, Nik-Zainal S, Davies HR, Ordonez GR, Mudie LJ, Latimer C, Edkins S, Stebbings L, Chen L, Jia M, Leroy C, Marshall J, Menzies A, Butler A, Teague JW, Mangion J, Sun YA, McLaughlin SF, Peckham HE, Tsung EF, Costa GL, Lee CC, Minna JD, Gazdar A, Birney E, Rhodes MD, McKernan KJ, Stratton MR, Futreal PA, Campbell PJ. A small-cell lung cancer genome with complex signatures of tobacco exposure. Nature. 2010; 463:184–190. [PubMed: 20016488]

23. Jones DT, Kocialkowski S, Liu L, Pearson DM, Bäcklund LM, Ichimura K, Collins VP. Tandem duplication producing a novel oncogenic BRAF fusion gene defines the majority of pilocytic astrocytomas. Cancer Research. 2008; 68:8673–8677. [PubMed: 18974108]

24. Singh D, Chan JM, Zoppoli P, Niola F, Sullivan R, Castano A, Liu EM, Reichel J, Porrati P, Pellegatta S, Qiu K, Gao Z, Ceccarelli M, Riccardi R, Brat DJ, Guha A, Aldape K, Golfinos JG, Zagzag D, Mikkelsen T, Finocchiaro G, Lasorella A, Rabadan R, Iavarone A. Transforming fusions of FGFR and TACC genes in human glioblastoma. Science. 2012; 337:1231–1235. [PubMed: 22837387]

25. Parker BC, Annala MJ, Cogdell DE, Granberg KJ, Sun Y, Ji P, Li X, Gumin J, Zheng H, Hu L, Yli-Harja O, Haapasalo H, Visakorpi T, Liu X, Liu C-G, Sawaya R, Fuller GN, Chen K, Lang FL, Nykter M, Zhang W. The tumorigenic fusion FGFR3-TACC3 escapes miR-99a regulation in glioblastoma. Journal of Clinical Investigation. 2012 in press.

26. Seshagiri S, Stawiski EW, Durinck S, Modrusan Z, Storm EE, Conboy CB, Chaudhuri S, Guan Y, Janakiraman V, Jaiswal BS, Guillory J, Ha C, Dijkgraaf GJ, Stinson J, Gnad F, Huntley MA, Degenhardt JD, Haverty PM, Bourgon R, Wang W, Koeppen H, Gentleman R, Starr TK, Zhang Z, Largaespada DA, Wu TD, de Sauvage FJ. Recurrent R-spondin fusions in colon cancer. Nature. 2012; 488:660–664. [PubMed: 22895193]

27. Crew AJ, Clark J, Fisher C, Gill S, Grimer R, Chand A, Shipley J, Gusterson BA, Cooper CS. Fusion of SYT to two genes, SSX1 and SSX2, encoding proteins with homology to the Kruppel-associated box in human synovial sarcoma. The EMBO Journal. 1995; 14:2333–2340. [PubMed: 7539744]

28. Tomlins SA, Rhodes DR, Yu J, Varambally S, Mehra R, Perner S, Demichelis F, Helgeson BE, Laxman B, Morris DS, Cao Q, Cao X, Andren O, Fall K, Johnson L, Wei JT, Shah RB, Al-Ahmadie H, Eastham JA, Eggener SE, Fine SW, Hotakainen K, Stenman UH, Tsodikov A, Gerald WL, Lilja H, Reuter VE, Kantoff PW, Scardino PT, Rubin MA, Bjartell AS, Chinnaiyan AM. The role of SPINK1 in ETS rearrangement-negative prostate cancers. Cancer Cell. 2008; 13:519–528. [PubMed: 18538735]

29. Westbrook CA, Hooberman AL, Spino C, Dodge RK, Larson RA, Davey F, Wurster-Hill DH, Sobol RE, Schiffer C, Bloomfield CD. Clinical significance of the BCR-ABL fusion gene in adult

acute lymphoblastic leukemia: a Cancer and Leukemia Group B study. Blood. 1992; 80:2983–2990. [PubMed: 1467514]

30. Lamant L, Dastugue N, Pulford K, Delsol G, Mariame B. A new fusion gene TPM3-ALK in anaplastic large cell lymphoma created by a (1;2)(q25;p23) translocation. Blood. 1999; 93:3088–3095. [PubMed: 10216106]
31. Lawrence B, Perez-Atayde A, Hibbard MK, Rubin BP, Dal Cin P, Pinkus JL, Pinkus GS, Xiao S, Yi ES, Fletcher CD, Fletcher JA. TPM3-ALK and TPM4-ALK oncogenes in inflammatory myofibroblastic tumors. The American Journal of Pathology. 2000; 157:377–384. [PubMed: 10934142]
32. Shaw AT, Yeap BY, Solomon BJ, Riely GJ, Gainor J, Engelman JA, Shapiro GI, Costa DB, ou SH, Butaney M, Salgia R, Maki RG, Varella-Garcia M, Doebele RC, Bang YJ, Kulig K, Selaru P, Tang Y, Wilner KD, Kwak EL, Clark JW, Iafrate AJ, Camidge DR. Effect of crizotinib on overall survival in patients with advanced non-small-cell lung cancer harbouring ALK gene rearrangement: a retrospective analysis. Lancet Oncology. 2011; 12:1004–1012. [PubMed: 21933749]
33. Morris SW, Kirstein MN, Valentine MB, Dittmer KG, Shapiro DN, Saltman DL, Look AT. Fusion of a kinase gene, ALK, to a nucleolar protein gene, NPM, in non-Hodgkin's lymphoma. Science. 1994; 263:1281–1284. [PubMed: 8122112]
34. Druker BJ, Tamura S, Buchdunger E, Ohno S, Segal GM, Fanning S, Zimmermann J, Lydon NB. Effects of a selective inhibitor of the Abl tyrosine kinase on the growth of Bcr-Abl positive cells. Nature Medicine. 1996; 2:561–566.
35. Hughes T, Deininger M, Hochhaus A, Branford S, Radich J, Kaeda J, Baccarani M, Cortes J, Cross NCP, Druker BJ, Gabert J, Grimwade D, Hehlmann R, Kamel-Reid S, Lipton JH, Longtine J, Martinelli G, Saglio G, Soverini S, Stock W, Goldman JM. Monitoring CML patients responding to treatment with tyrosine kinase inhibitors: review and recommendations for harmonizing current methodology for detecting BCR-ABL transcripts and kinase domain mutations and for expressing results. Blood. 2006; 108:28–37. [PubMed: 16522812]
36. Sun C, Dobi A, Mohamed A, Li H, Thangapazham RL, Furusato B, Shaheduzzaman S, Tan S-H, Vaidyanathan G, Whitman E, Hawksworth DJ, Chen Y, Nau M, Patel V, Vahey M, Gutkind JS, Sreenath T, Petrovics G, Sesterhenn IA, McLeod DG, Srivastava S. TMPRSS2-ERG fusion, a common genomic alteration in prostate cancer activates C-MYC and abrogates prostate epithelial differentiation. Oncogene. 2008; 27:5348–5353. [PubMed: 18542058]
37. Persson M, Andren Y, Mark J, Horlings HM, Persson F, Stenman G. Recurrent fusion of MYB and NFIB transcription factor genes in carcinomas of the breast and head and neck. Proc Natl Acad Sci USA. 2009; 106:18740–18744. [PubMed: 19841262]
38. Chiarle R, Voena C, Ambrogio C, Piva R, Inghirami G. The anaplastic lymphoma kinase in the pathogenesis of cancer. Nature Reviews Cancer. 2008; 8:11–23.
39. McWhirter JR, Galasso DL, Wang JY. A coiled-coil oligomerization domain of Bcr is essential for the transforming function of Bcr-Abl oncoproteins. Molecular and Cellular Biology. 1993; 13:7587–7595. [PubMed: 8246975]
40. Nothwang HG, Kim HG, Aoki J, Geisterfer M, Kübart S, Wegner RD, van Moers A, Ashworth LK, Haaf T, Bell J, Arai H, Tommerup N, Ropers HH, Wirth J. Functional hemizygosity of PAFAH1B3 due to a PAFAH1B3-CLK2 fusion gene in a female with mental retardation, ataxia and atrophy of the brain. Human Molecular Genetics. 2001; 10:797–806. [PubMed: 11285245]
41. Karenko L, Hahtola S, Päivinen S, Karhu R, Syrjä S, Kähkönen M, Nedoszytko B, Kytölä S, Zhou Y, Blazevic V, Pesonen M, Nevala H, Nupponen N, Sihto H, Krebs I, Poustka A, Roszkiewicz J, Saksela K, Peterson P, Visakorpi T, Ranki A. Primary cutaneous T-cell lymphomas show a deletion or translocation affecting NAV3, the human UNC-53 homologue. Cancer Research. 2005; 65:8101–8110. [PubMed: 16166283]
42. Perner S, Demichelis F, Beroukhim R, Schmidt FH, Mosquera J-M, Setlur S, Tchinda J, Tomlins SA, Hofer MD, Pienta KG, Kuefer R, Vessella R, Sun X-W, Meyerson M, Lee C, Sellers WR, Chinnaiyan AM, Rubin MA. TMPRSS2:ERG fusion-associated deletions provide insight into the heterogeneity of prostate cancer. Cancer Research. 2006; 66:8337–8341. [PubMed: 16951139]
43. Lipson D, Capelletti M, Yelensky R, Otto G, Parker A, Jarosz M, Curran JA, Balasubramanian S, Bloom T, Brennan KW, Donahue A, Downing SR, Frampton GM, Garcia L, Juhn F, Mitchell KC,

White E, White J, Zwirko Z, Peretz T, Nechushtan H, Soussan-Gutman L, Kim J, Sasaki H, Kim HR, Park SI, Ercan D, Sheehan CE, Ross JS, Cronin MT, Jänne PA, Stephens PJ. Identification of new ALK and RET gene fusions from colorectal and lung cancer biopsies. Nature Medicine. 2012; 18:382–384.

44. Ciampi R, Knauf JA, Kerler R, Gandhi M, Zhu Z, Nikiforova MN, Rabes HM, Fagin JA, Nikiforov YE. Oncogenic AKAP9-BRAF fusion is a novel mechanism of MAPK pathway activation in thyroid cancer. Journal of Clinical Investigation. 2005; 115:94–101. [PubMed: 15630448]
45. Lawson AR, Hindley GF, Forshew T, Tatevossian RG, Jamie GA, Kelly GP, Neale GA, Ma J, Jones TA, Ellison DW, Sheer D. RAF gene fusion breakpoints in pediatric brain tumors are characterized by significant enrichment of sequence microhomology. Genome Research. 2011; 21:505–514. [PubMed: 21393386]
46. Zhu L, Zhang Y, Zhang W, Yang S, Chen J-Q, Tian D. Patterns of exon-intron architecture variation of genes in eukaryotic genomes. BMC Genomics. 2009; 10:47. [PubMed: 19166620]
47. Long M, Deutsch M. Association of intron phases with conservation at splice site sequences and evolution of spliceosomal introns. Molecular Biology and Evolution. 1999; 16:1528–1534. [PubMed: 10555284]
48. Sverdlov AV, Rogozin IB, Babenko VN, Koonin EV. Evidence of splice signal migration from exon to intron during intron evolution. Current Biology. 2003; 13:2170–2174. [PubMed: 14680632]
49. Ruvinsky A, Eskesen ST, Eskesen FN, Hurst LD. Can codon usage bias explain intron phase distributions and exon symmetry? Journal of Molecular Evolution. 2005; 60:99–104. [PubMed: 15696372]
50. Martinelli G, Amabile M, Giannini B, Terragna C, Ottaviani E, Soverini S, Saglio G, Rosti G, Baccarani M. Novel types of bcr-abl transcript with breakpoints in BCR exon 8 found in Philadelphia positive patients with typical chronic myeloid leukemia retain the sequence encoding for the DBL- and CDC24 homology domains but not the pleckstrin homology one. Haematologica. 2002; 87:688–694. [PubMed: 12091118]
51. Tort F, Campo E, Pohlman B, Hsi E. Heterogeneity of genomic breakpoints in MSN-ALK translocations in anaplastic large cell lymphoma. Human Pathology. 2004; 35:1038–1041. [PubMed: 15297972]
52. Lin C, Yang L, Tanasa B, Hutt K, Ju BG, Ohgi K, Zhang J, Rose DW, Fu XD, Glass CK, Rosenfeld MG. Nuclear receptor-induced chromosomal proximity and DNA breaks underlie specific translocations in cancer. Cell. 2009; 139:1069–1083. [PubMed: 19962179]
53. Akiva P, Toporik A, Edelheit S, Peretz Y, Diber A, Shemesh R, Novik A, Sorek R. Transcription-mediated gene fusion in the human genome. Genome Research. 2006; 16:30–36. [PubMed: 16344562]
54. Chang YF, Imam JS, Wilkinson MF. The nonsense-mediated decay RNA surveillance pathway. Annu Rev Biochem. 2007; 76:51–74. [PubMed: 17352659]
55. Laurent E, Talpaz M, Kantarjian H, Kurzrock R. The BCR gene and Philadelphia chromosome-positive leukemogenesis. Cancer Research. 2001; 61:2343–2355. [PubMed: 11289094]
56. Maher CA, Kumar-Sinha C, Cao X, Kalyana-Sundaram S, Han B, Jing X, Sam L, Barrette T, Palanisamy N, Chinnaiyan AM. Transcriptome sequencing to detect gene fusions in cancer. Nature. 2009; 458:97–101. [PubMed: 19136943]
57. 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. Chimeric transcript discovery by paired-end transcriptome sequencing. Proc Natl Acad Sci USA. 2009; 106:12353–12358. [PubMed: 19592507]
58. Sboner A, Habegger L, Pflueger D, Terry S, Chen DZ, Rozowsky JS, Tewari AK, Kitabayashi N, Moss BJ, Chee MS, Demichelis F, Rubin MA, Gerstein MB. FusionSeq: a modular framework for finding gene fusions by analyzing paired-end RNA-sequencing data. Genome Biology. 2010; 11:R104. [PubMed: 20964841]

59. Kinsella M, Harismendy O, Nakano M, Frazer KA, Bafna V. Sensitive gene fusion detection using ambiguously mapping RNA-Seq read pairs. Bioinformatics. 2011; 27:1068–1075. [PubMed: 21330288]
60. 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. deFuse: an algorithm for gene fusion discovery in tumor RNA-Seq data. PLoS Computational Biology. 2011; 7:e1001138. [PubMed: 21625565]
61. McPherson A, Wu C, Hajirasouliha I, Hormozdiari F, Hach F, Lapuk A, Volik S, Shah S, Collins C, Sahinalp SC. Comrad: detection of expressed rearrangements by integrated analysis of RNA-Seq and low coverage genome sequence data. Bioinformatics. 2011; 27:1481–1488. [PubMed: 21478487]
62. Iyer MK, Chinnaiyan AM, Maher CA. ChimeraScan: a tool for identifying chimeric transcription in sequencing data. Bioinformatics. 2011; 27:2903–2904. [PubMed: 21840877]
63. Kim D, Salzberg SL. Tophat-Fusion: an algorithm for discovery of novel fusion transcripts. Genome Biology. 2011; 12:R72. [PubMed: 21835007]
64. Ju YS, Lee WC, Shin JY, Lee S, Bleazard T, Won JK, Kim YT, Kim JI, Kang JH, Seo JS. A transforming KIF5B and RET gene fusion in lung adenocarcinoma revealed from whole-genome and transcriptome sequencing. Genome Research. 2012; 22:436–445. [PubMed: 22194472]
65. Houseley J, Tollervey D. Apparent non-canonical trans-splicing is generated by reverse transcriptase in vitro. PLoS ONE. 2010; 5:e12271. [PubMed: 20805885]
66. Kanagawa T. Bias and artifacts in multitemplate polymerase chain reactions. Journal of Bioscience and Bioengineering. 2003; 96:317–323. [PubMed: 16233530]
67. Chase A, Ernst T, Fiebig A, Collins A, Grand F, Erben P, Reiter A, Schreiber S, Cross NC. TFG, a target of chromosome translocations in lymphoma and soft tissue tumors, fuses to GPR128 in healthy individuals. Haematologica. 2010; 95:20–26. [PubMed: 19797732]
68. Kipersztok S, Osawa GA, Liang LF, Modi WS, Dean J. POM-ZP3, a bipartite transcript derived from human ZP3 and a POM121 homologue. Genomics. 1995; 25:354–359. [PubMed: 7789967]
69. Paulding CA, Ruvolo M, Haber DA. The Tre2 (USP6) oncogene is a hominoid-specific gene. Proc Natl Acad Sci USA. 2003; 100:2507–2511. [PubMed: 12604796]
70. Pasek S, Risler JL, Brezellec P. Gene fusion/fission is a major contributor to evolution of multi-domain bacterial proteins. Bioinformatics. 2006; 22:1418–1423. [PubMed: 16601004]

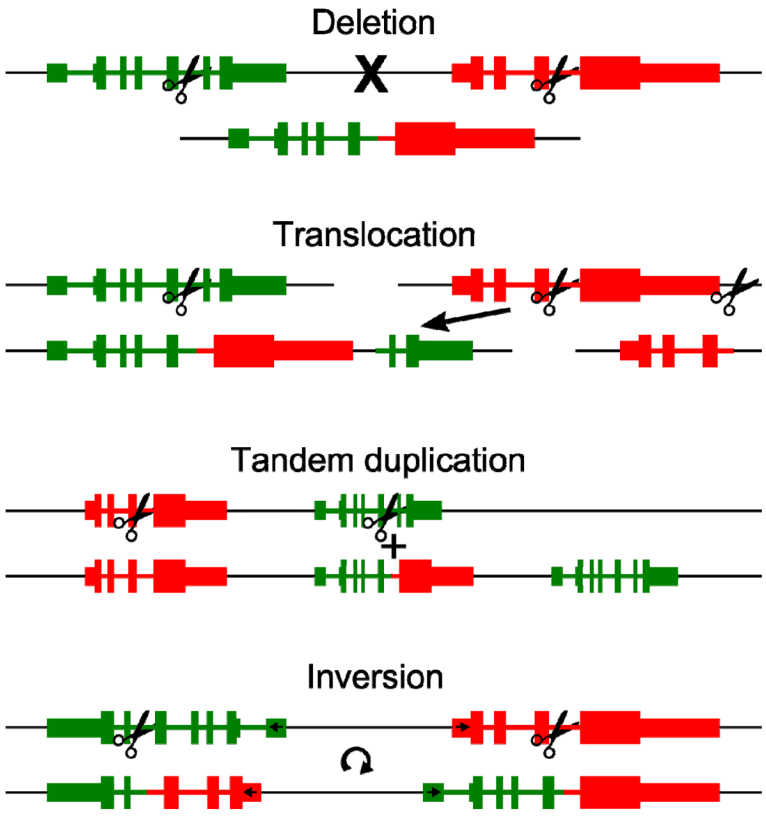

Figure 1.
An illustration of the four basic types of chromosomal rearrangement and how they lead to the formation of fusion genes. Original genomic layout is shown at the top, layout after rearrangement is shown at the bottom. Scissors indicate genomic breakpoints. A discontinuity in the black line indicates separate chromosomes.

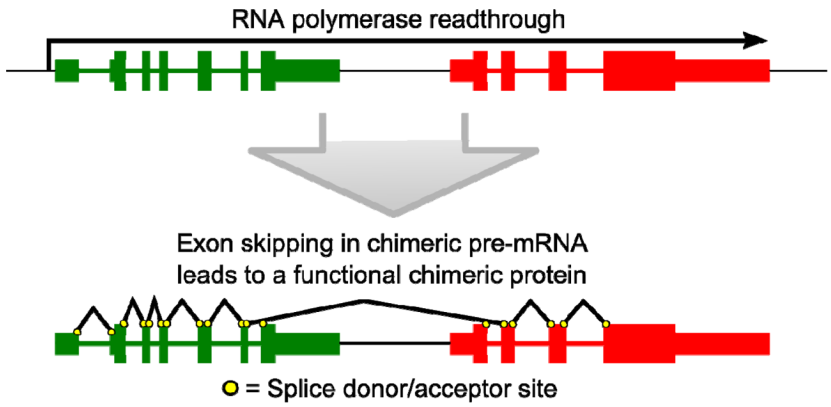

Figure 2.
A read-through fusion transcript is formed when an RNA polymerase continues transcribing beyond the end of a gene and transcription continues to an adjacent downstream gene. Exon skipping due to missing splice sites can give rise to a fusion transcript encoding a functional chimeric protein. Boxes indicate exons, thicker boxes indicate coding sequence.

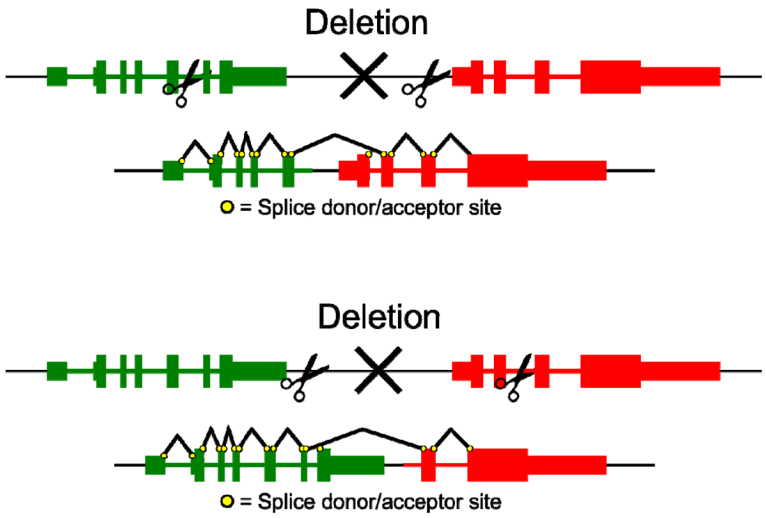

Figure 3.
A chromosomal rearrangement with intergenic breakpoints can result in a fusion gene encoding a functional chimeric protein. Illustration depicts two example scenarios. Boxes indicate exons, thicker boxes indicate coding sequence.

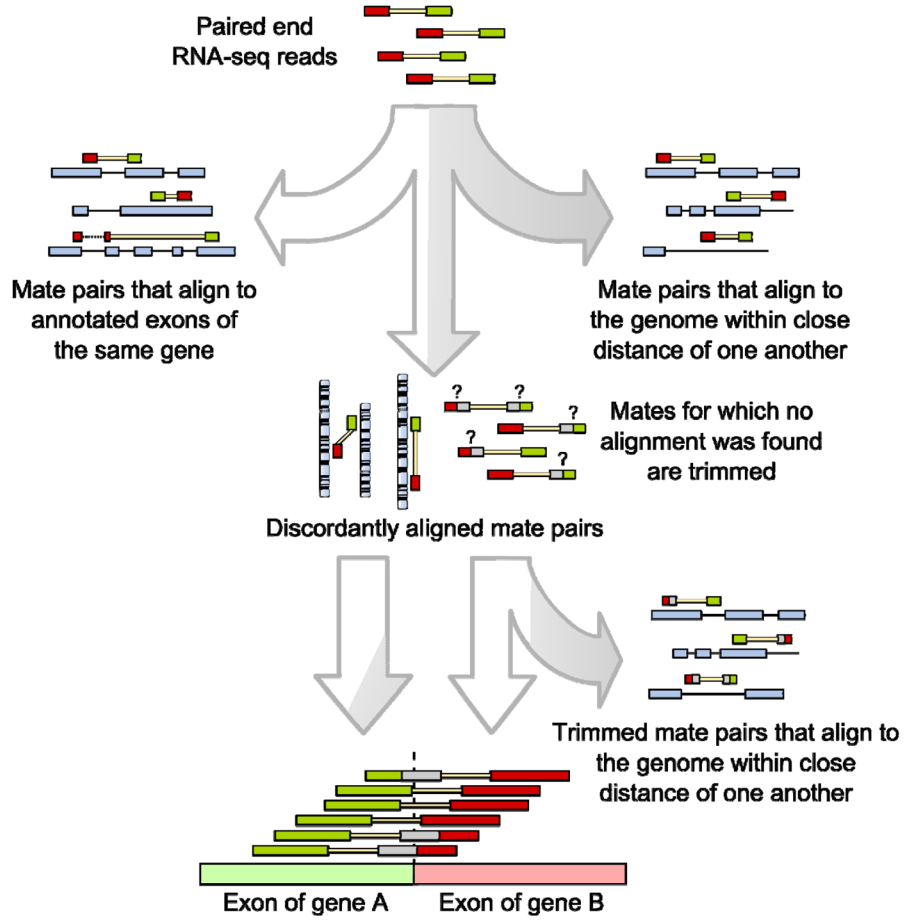

Figure 4.
Illustration of the typical workflow involved in fusion gene discovery. The process for mate trimming is shown as performed by the ChimeraScan algorithm [57].

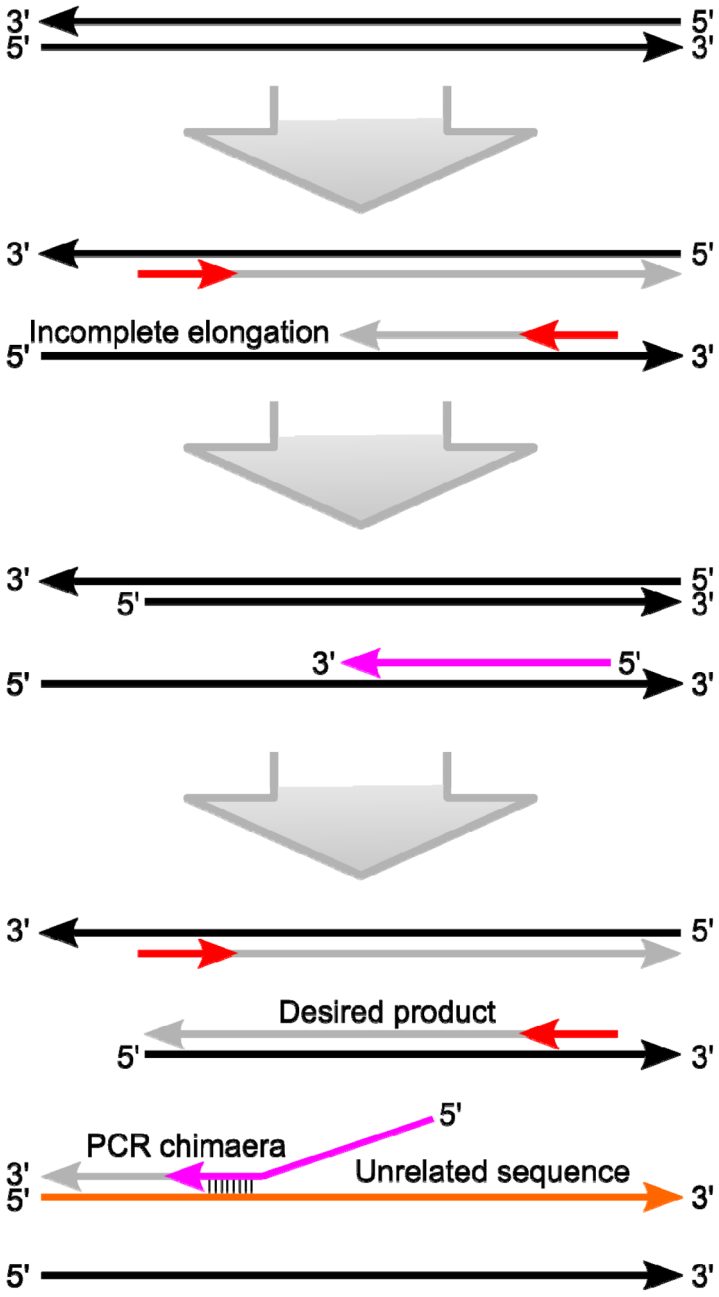

Figure 5.
An illustration of the "incomplete elongation" theory for the formation of PCR chimeras [61]. According to this theory, a PCR chimaera is formed when an incomplete elongation product (pink) of a PCR primer (red) hybridizes with an unrelated but partially homologous template (orange), followed by chimeric elongation.

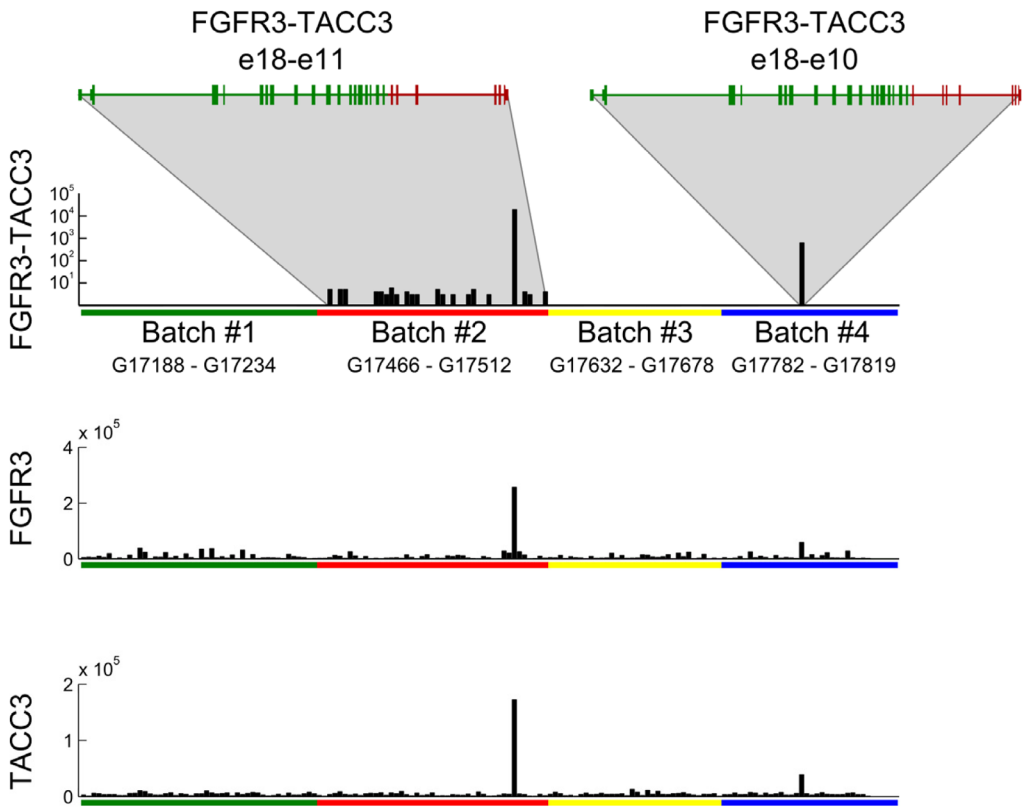

Figure 6.

Data showing inter-sample contamination in a batch of transcriptome sequenced samples from The Cancer Genome Atlas glioblastoma project. Batch #2 is contaminated with fusion transcripts from a single sample expressing high levels of the fusion. Y-axis represents the total number of mate pairs spanning the fusion junction. The top panel (FGFR3-TACC3) shows the number of reads overlapping the fusion junction. The middle and bottom panels show the number of reads aligned to FGFR3 and TACC3 transcript sequences.

Table 1

Fusion genes in human cancers.

	Cancer	Fusion gene	Frequency	Mechanism of formation	Biological impact	References
Hematological cancers	Acute lymphocytic leukemia	ETV6-RUNX1	25%	Interchromosomal translocation	Oncogenic chimeric protein	Golub et al. (1995), Romana et al. (1995)
		BCR-ABL1	15%	Interchromosomal translocation	Oncogenic chimeric protein	Westbrook et al. (1992)
	Acute myeloid leukemia	RUNX1-ETO	10–15%	Interchromosomal translocation	Oncogenic chimeric protein	Erickson et al. (1992)
		CBFB-MYH11	10–15%	Inversion	Oncogenic chimeric protein	Liu et al. (1993)
	Acute promyelocytic leukemia	PML-RARA	95%	Interchromosomal translocation	Oncogenic chimeric protein	Borrow et al. (1990), Warrell et al. (1991)
		PLZF-RARA	0–5%	Interchromosomal translocation	Oncogenic chimeric protein	Chen et al. (1993)
	Anaplastic large cell lymphoma	NPM1-ALK	75%	Interchromosomal translocation	Oncogenic chimeric protein	Morris et al. (1994)
		TPM3-ALK	15%	Interchromosomal translocation	Oncogenic chimeric protein	Lamant et al. (1999)
	Burkitt's lymphoma	IG@-MYC	90–100%	Interchromosomal translocation	Promoter exchange	Manolov et al. (1972), Dalla-Favera et al. (1982)
	Chronic myelogenous leukemia	BCR-ABL1	95–100%	Interchromosomal translocation	Oncogenic chimeric protein	Nowell et al. (1960), Shtivelman et al. (1985)
Solid cancers	Adenoid cystic carcinoma	MYB-NFIB	90–100%	Interchromosomal translocation	Loss of microRNA regulation	Persson et al. (2009)
	Bladder cancer	FGFR3-TACC3	0–10%	Tandem duplication	Oncogenic chimeric protein	Williams et al. (2012)
	Clear cell sarcoma	EWSR1-ATF1	90–100%	Interchromosomal translocation	Oncogenic chimeric protein	Bridge et al. (1990), Zucman et al. (1993)
	Colon cancer	PTPRK-RSPO3	5–10%	Inversion	Promoter exchange	Seshagiri et al. (2012)
		EIF3E3-RSPO2	0–5%	Deletion	Promoter exchange	Seshagiri et al. (2012)
	Congenital fibrosarcoma	ETV6-NTRK3	90–100%	Interchromosomal translocation	Oncogenic chimeric protein	Knezevich et al. (1998)
	Ewing sarcoma	EWSR1-FLI1	90%	Interchromosomal translocation	Oncogenic chimeric protein	Turc-Carel et al. (1983), Aurias et al. (1983)
	Follicular thyroid carcinoma	PAX8-PPARG	60%	Interchromosomal translocation	Oncogenic chimeric protein	Kroll et al. (2000)
	Glioblastoma	FGFR3-TACC3	0–5%	Tandem duplication	Oncogenic chimeric protein	Singh et al. (2012), Parker et al. (2012)
	Inflammatory myofibroblastic tumor	TPM3-ALK	50%	Interchromosomal translocation	Oncogenic chimeric protein	Lawrence et al. (2000)
	Mucoepidermoid carcinoma	MECT1-MAML2	60%	Interchromosomal translocation	Oncogenic chimeric protein	Tonon et al. (2003)
	Myxoid liposarcoma	FUS-DDIT3	90–100%	Interchromosomal translocation	Oncogenic chimeric protein	Crozat et al. (1993), Rabbits et al. (1993)
		EWSR1-DDIT3	0–5%	Interchromosomal translocation	Oncogenic chimeric protein	Panagopoulos et al. (1996)

Cancer	Fusion gene	Frequency	Mechanism of formation	Biological impact	References
Non-small cell lung cancer	EML4-ALK	0–10%	Inversion	Oncogenic chimeric protein	Soda et al. (2007), Rikova et al. (2007)
NUT midline carcinoma	BRD4-NUT	90–100%	Interchromosomal translocation	Promoter exchange	French et al. (2003)
Papillary thyroid carcinoma	CCDC6-RET	15%	Inversion	Oncogenic chimeric protein	Grieco et al. (1990)
	NCOA4-RET	15%	Complex rearrangement	Oncogenic chimeric protein	Santoro et al. (1994)
Pediatric renal cell carcinoma	PRCC-TFE3	20–40%	Interchromosomal translocation	Oncogenic chimeric protein	Weterman et al. (1996)
Pilocytic astrocytoma	KIAA1549-BRAF	70%	Tandem duplication	Oncogenic chimeric protein	Jones et al. (2008)
Prostate cancer	TMPRSS2-ERG	60%	Deletion	Promoter exchange	Tomlins et al. (2005)
	TMPRSS2-ETV1	0–5%	Interchromosomal translocation	Promoter exchange	Tomlins et al. (2005)
	TMPRSS2-ETV4	0–5%	Interchromosomal translocation	Promoter exchange	Tomlins et al. (2006)
Secretory breast carcinoma	ETV6-NTRK3	90%	Interchromosomal translocation	Oncogenic chimeric protein	Tognon et al. (2002)
Serous ovarian cancer	ESRRA-C11orf20	15%	Intrachromosomal translocation	Oncogenic chimeric protein	Salzman et al. (2011)
Synovial sarcoma	SS18-SSX1	70%	Interchromosomal translocation	Oncogenic chimeric protein	Turc-Carel et al. (1987), Clark et al. (1994)
	SS18-SSX2	30%	Interchromosomal translocation	Oncogenic chimeric protein	Crew et al. (1995)
	SS18-SSX4	0–5%	Interchromosomal translocation	Oncogenic chimeric protein	Skytting et al. (1999)

Cancer Lett. Author manuscript; available in PMC 2014 November 01.

Table 2

A comparison of widely used software packages for fusion gene detection.

Software	Installation requirements	Uses DNA-seq to identify genomic breakpoints?	Detects exon disrupting fusions?	Supports colorspace reads?	References
ChimeraScan	Python, Bowtie	No	No	No	Iyer et al. (2011)
Comrad	Perl, Bowtie, Blat	Yes	Yes	No	McPherson et al. (2011)
Defuse	Perl, Bowtie, Blat	No	Yes	No	McPherson et al. (2011)
Tophat-Fusion	Python, Bowtie	No	Yes	Yes	Kim et al. (2011)
ShortFuse	Python, Bowtie	No	Yes	No	Kinsella et al. (2011)

Cancer Lett. Author manuscript; available in PMC 2014 November 01.