Fusion Genes and Their Detection through Next Generation Sequencing in Malignant Hematological Diseases and Solid Tumors

Jiaqi Liu1, 2, 3*, Linqian Weng3, Yue Ming3, Sen Liu1, 2, Nan Wu1, 2, Zhihong Wu1, 2*

1Department of Orthopedic Surgery, Peking Union Medical College Hospital, Peking Union Medical College and Chinese Academy of Medical Sciences, Beijing, PR China
2Beijing Key Laboratory for Genetic Research of Skeletal Deformity, Peking Union Medical College Hospital, Peking Union Medical College and Chinese Academy of Medical Sciences, Beijing, PR China
3Department of Breast Surgical Oncology, Cancer Hospital of Chinese Academy of Medical Sciences, Beijing, PR China

*The PET-CT Center, Cancer Hospital of Chinese Academy of Medical Sciences, Beijing, PR China

Corresponding author: Zhihong Wu, Beijing Key Laboratory for Genetic Research of Skeletal Deformity; Department of Central Laboratory, Peking Union Medical College Hospital, Peking Union Medical College and Chinese Academy of Medical Sciences, No. 1 Shuailiyuan, Dongcheng District, Beijing 100730, PR China; Tel: +8601069156081; Fax: +8601069156081; E-mail: orthoscience@126.com

Nan Wu, Department of Orthopaedic Surgery, Peking Union Medical College Hospital, Peking Union Medical College and Chinese Academy of Medical Sciences; Beijing Key Laboratory for Genetic Research of Skeletal Deformity, No.1 Shuailiyuan, Dongcheng District, Beijing 100730, PR China; Tel: +8601069156081; E-mail: dr.wunan@pumch.cn

Received date: January 02, 2016; Accepted date: February 20, 2016; Published date: February 26, 2016

Copyright: ©2016 Liu J, et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Abstract

Fusion genes are neoplasia-associated mutations, which play a particularly significant role in tumorigenesis and exhibit great importance for clinical applications in malignant hematological diseases and solid tumors. Simultaneously with copy number variants (CNVs), gene fusions are resulting from balanced and unbalanced chromosomal rearrangements. Thus, understanding the mutagenesis and instability of CNV, as well as the underlying molecular mechanisms of chromosomal rearrangements will improve our comprehension of gene fusions. Recently, next generation sequencing (NGS), especially transcriptome sequencing or RNA-Sequencing (RNA-seq), has become a very useful tool to identify gene alterations in cancer and a powerful approach for investigating the tumorigenesis. However, we are still facing with the challenge of minimizing false positives in results of RNA-seq. Whole-genome sequencing (WGS) is also used for the fusion gene detection, which provides us a more comprehensive and integrative way to detect structural variants. WGS may correct the false-negative results from RNA-seq. Additionally, many computational tools with more sensitivity and specificity have been developed for the detection of fusion transcripts from NGS datas. In the future, multi-omics analysis, third-generation sequencing and liquid-biopsy technique all provide opportunities to comprehensively interpret gene fusions and understand the biology of cancer genomes.

Keywords: Fusion gene; Next generation sequencing (NGS); Transcriptome sequencing or RNA-seq (RNA-Sequencing); Whole genome sequencing (WGS)

Introduction

Fusion genes, also called chimeric genes or hybrid genes, are neoplasia-associated mutations arising from structural chromosome rearrangements, such as chromosomal insertion, deletion, translocation or inversion that juxtaposes two separated genes [1,2]. They have been reported to be important genomic events in human cancer because their fusion gene products can drive the development of cancer, and thus are potential prognostic markers or therapeutic targets in cancer treatment. On the basis of transposons studies, human cancers could also be the result of the translocations and chromosome rearrangements which lead to the abnormal expression of genes located at breakpoints [3]. Up to now, the current next generation sequencing (NGS)-based approaches for detection, such as transcriptome sequencing or RNA-Sequencing (RNA-seq) and whole genome sequencing (WGS), have become a very useful tool to identify new tumor-associated gene fusions and investigate their impact on tumorigenesis [4]. In this study, we then comprehensively reviewed NGS studies to detect gene fusions in malignant hematological diseases and solid tumors, to update our knowledge about the advances and challenges in the gene fusion detection through NGS, especially RNA-seq.

Methods

The literatures searching was conducted on PubMed, ScienceDirect and Google Scholar comprehensively, using keywords included "gene fusion", "RNA-seq"/"whole-genome sequencing"/"next generation sequencing" AND "cancer"/"tumor"/"leukemia"/"lymphoma". After the relevant literatures were carefully read and analyzed, we found 71 publications directly related to our study purpose by this searching method.

Fusion genes: tumorigenesis, biomarker and therapeutic target

Fusion genes play a particularly significant role in tumorigenesis, which has been identified with great importance for clinical...
applications [5]. Gene fusion events can be observed in cancer samples more frequently than benign samples. They are present in approximately 20% of all human neoplasms. Although the functional outcomes of many gene fusions are still under exploration, it is well established that most of them will lead to tumorigenesis. Since a strong correlation can be found between recurrent gene fusions and tumor types, gene fusion detection has been suggested to be used for screening of common tumor types. Subtypes identification provides a roadmap for targeted therapies. Although recent studies have thus far defined a large quantity of gene fusions that involve different cancer related genes, which constitute an important diagnostic and prognostic parameter especially malignant hematological diseases and sarcomas, gene fusions in solid tumors have rather limited clinical and biological impact [5].

The BCR-ABL1 fusion gene in the well-known Philadelphia (Ph) chromosome is the prototypic fusion oncogene, which is associated with chronic myeloid leukemia (CML). It is now used as a biomarker during diagnosis and monitoring patient response to treatment. As some morphologically homogeneous malignancies are heterogeneous because of gene fusion status, they play an important role in treatment stratification, such as different MLL fusions in AML or fusion-positive versus fusion-negative Alveolar rhabdomyosarcoma (ARMS) [6,7]. Nowadays, many technologies have already been used in detecting gene fusions and other genetic aberrations, such as chromosome banding analysis, reverse transcriptase-polymerase chain reaction (RT-PCR) and Sanger sequencing etc [8]. Except for the hematological malignancies, large amount of data emerges from the studies of malignant solid tumors, including most sarcomas and a few carcinomas.

Ewing’s sarcoma is defined by a recurrent chromosomal translocation between the EWSR1 gene and various ETS genes, and EWS-FLI1 is the most common gene fusion in Ewing's sarcoma, which present in 85% of cases [9]. In the study carried out by Saravanan et al. [10], genes such as CLKI, CASP3, PPFBP1 and TERT, which potentially participate in oncogenesis, are alternatively spliced by EWS-FLI1. Thus EWS-FLI1 can be used as a diagnostic biomarker for Ewing's sarcoma. While, there are still many important questions to be solved to understand the molecular mechanism of EWS-FLI1 and its potential value for cancer therapy.

As is known, the oncogenic potential of ETS-related gene (ERG) is involved in Ewing's sarcoma and leukemia. However, in the past decades, ERG is found to be highly associated with prostate cancer [11]. It is showed by Tomlins et al. [12] that ERG is overexpressed in most prostate carcinomas because of a gene fusion with the androgen-driven promoter of TMPRSS2 gene. Many other studies have also shown TMPRSS2-ERG gene rearrangements to be the most commonly found TMPRSS2:ETS family pairing in prostate cancer, demonstrating the specificity of TMPRSS2:ERG for prostate cancer and a role for TMPRSS2-ERG in the development and progression of prostate cancer [13]. The TMPRSS2-ERG is showed to trigger carcinogenesis by inhibiting apoptosis of prostate gland cells and at the same time, increasing cell proliferation [14]. The proto-oncogenes ETV1, were also found to be highly expressed in a subset of prostate cancers [12]. Recently, it is recommended that the classification of prostate cancer can be divided by distinct molecular subtypes, which includes mutually exclusive ETS fusions (ETS-positive), SPINK1-overexpressing, and CHD1-loss cancers etc [15]. In this way, a simple molecular barcode (includes ETS/SPINK1/SPOP/CHD1/RAS-RAF/PTE/TP53 status) can be used in molecular prostate cancer subtypes, and thus may allow stratification of patients for different management strategies in the future.

Approximately 40%-70% of men with castration resistant prostate cancers have ERG rearrangements, which may respond better to anti-hormonal therapy than ERG-negative ones [13]. Currently, there are many studies targeting at the TMPRSS2-ERG fusion and its downstream signaling. It was shown that knockdown of the TMPRSS2-ERG fusion in a cancer cell line can lead to primary tumour growth inhibition, which made TMPRSS2-ERG a potential therapeutic target [16]. There is also study showing that targeting the most common and clinically significant alternatively spliced isoforms of the TMPRSS2-ERG mRNA with specific siRNAs via liposomal nanovectors is promising therapy for men with prostate cancer [17]. For example, siRNA has been used to target the BCR-ABL fusion successfully in CML and against the AML1-ETO in AML-M2 [18]. Because specific ETS factors could be found in many other solid tumor types, their downstream effectors are very likely to be in common, therefore providing more possible novel drug targets for treatment of these malignancies.

Relatively few recurrent gene fusion events have been associated with breast cancer. In a study of whole-transcriptome sequencing of 120 fresh-frozen primary breast cancer samples, six newly validated gene fusions were recurrent, including three in-frame and three out-frame ones [19]. A recurrent gene fusion, RPS6KB1 kinase, and EGFR, which is a therapeutically important receptor kinase and involving in the rapamycin signaling, was discovered in the analysis of 14 breast cancer cell lines [1]. Not only in common tumors, recent study also indicated a novel FNI-FGFR1 fusion gene might participate the tumorgenesis of phosphaturic mesenchymal tumors (PMTs), which typically cause hypophosphataemia and tumor-induced osteomalacia (TIO) [20].

As people getting more into the clinical importance of gene fusions and other types of genetic rearrangement, greater emphasis has been putting on genetic features in the classification of neoplasms. In the latest World Health Organization (WHO) classifications, translocation and/or gene fusion status is mandatory for the diagnosis of some types of tumors, such as “AML with t(8;21)(q22;q22), RUNX1-RUNXITT” and “B lymphoblastic leukemia/lymphoma with t(5;14)(q31:g32), IL3-IGH” [4]. And for other malignancies, the Xp11 translocation renal cell carcinomas (RCC) harbor gene fusions involving TFE3, which is among the MiT subfamily of transcription factors. Thus it was first officially recognized in the 2004 WHO renal tumor classification [21].

Therapeutic approaches based on oncogene addiction can offer significant anticancer benefit, among which the identification of anaplastic lymphoma kinase (ALK) rearrangements is a key aspect. For all the lung cancer patients, 4-8% can be detected with the EML4-ALK gene fusion, especially in light smokers and nonsmokers [22]. Crizotinib was the first approved medication for ALK-positive patients. In the phase III PROFILE 1014 study, crizotinib is associated with a median progression-free survival of 10.9 months when used as the first-line treatment [23]. Imatinib, the tyrosine kinase inhibitor, which was the first drug that was specifically designed to target a fusion gene, BCR-ABL1 in CML. There are various other common malignancies that have been shown to display various fusions involving kinase-encoding regions, e.g. BRAF, FGFR3, NTRK1, RET and ROS1 etc [4]. With more and more novel drugs under approval of FDA regarding to these gene fusions, stratification of diagnosis and treatment could be of great importance in clinical practice.
Chromosomal rearrangements: the origin of gene fusions

Chromosomal rearrangements are very pervasive in cancer, while their impacts are hard to characterize and interpret [24]. Gene fusions are resulting from balanced and unbalanced chromosomal rearrangements. Balanced changes are the prototypical mechanism behind gene fusions, including translocations, insertions and inversions. While gene fusions can also arise through unbalanced chromosomal rearrangements, such as interstitial deletions, as to a deletion of an interstitial chromosomal segment. Both balanced and unbalanced aberrations may lead to create a chimeric gene through the fusion of parts of the two genes from each side of breakpoint, or juxtapose the coding sequences in one gene with the regulatory sequences of another gene from the other breakpoint. Even there should be two derivative chromosomal segments and each of which may harbor the pathogenetic gene fusion through a balanced chromosomal rearrangement, usually only one of these genes will produce an in-frame fusion transcript [4]. However, genes in one of the breakpoints may also become truncated and lose their function as haploinsufficiency. As the gene fusions can upregulate or deregulate genes depending on the breakpoints, it may lead to tumorigenesis through activation of oncogene or inactivation of tumor suppressor gene.

Interestingly, gene fusion always occurred simultaneously with CNVs, which also has a significant role in tumorigenesis in many cancers, such as gastric cancer [25], ovarian cancer [26], hepatocellular carcinoma [27], colorectal cancer [28], bladder cancer [29] and so on. CNV involves deletions, duplications and insertions of DNA segments larger than 1 kb, which is variable among individuals [30]. Many seemingly balanced translocations that result in gene fusions are accompanied by extensive deletions, duplications or amplifications among the breakpoints [31,32]. In most cases, CNV generates more than one breakpoint. When a breakpoint located between the functional elements of the two genes, a fusion gene may occur. Fusion partner genes can be found to contribute promoters (5’ UTR), coding sequences and 3’ UTRs. Consequently, genes affected by CNV are potential candidates for fusion events [4]. Thus, understanding the mutagenesis and instability of CNV, as well as the underlying molecular mechanisms of chromosomal rearrangements will improve our understanding of gene fusions.

In addition, transcript fusions may also originate from non-adjacent genes without a corresponding fusion at the DNA level, resulting in so-called transcription-induced gene fusions (TIGFs), including cis-TIGFs (neighbouring genes located on the same DNA strand) and trans-TIGFs (genes located far apart or on different chromosomes). Some cis-TIGFs have been identified associated with particular tumor types, which indicates TIGFs may play important roles in tumor development [33]. Although trans-TIGFs have been identified in human cells [34], no trans-TIGFs have yet been verified in any independent studies [4].

Next generation sequencing (NGS): a high-performing strategy for fusion gene discovery

Although cytogenetics and fluorescence in situ hybridization (FISH) approach will continue to be indispensable tools for fusion gene diagnostics in hematological diseases and solid tumors, the modern high-throughput NGS have showed their great impact to identify new tumor-associated gene fusions [35]. Recently, NGS has become a very useful tool to identify gene alterations in cancer and a powerful approach for investigating the tumorigenesis [36]. Chromosomal rearrangements, such as deletion, duplication, translocation, inversion and inversion, can be detected by paired-end information and apparent fragment length and orientation of NGS [37]. Additionally, chimeric RNA analysis can detect gene fusions and also reveal their breakpoints directly [37,38], and the de novo assembly approach can be used for some complex fusions [39]. Over the past few years, advances of NGS and affordable price provide an opportunity for detection of cancer transcriptomes, including the expressed fusion genes. The first NGS study to detect gene fusions in cancer were carried out on cell lines [37], and quickly extended to numerous investigations in different cancer types. As another landscape, Maher et al. [40] successfully re-discovered the BCR-ABL1 gene fusion in a CML cell line and the TMPRSS2-ERG gene fusion in a prostate cancer cell line and tissues through RNA-Seq. Yoshihara et al. [2] queried transcriptome data from 4,366 neoplasms from 13 different cancer types, which had been studied within the Cancer Genome Atlas (TCGA) network, and detected more than 8,600 different fusion transcripts. During only the past 3 years, more than 9,000 novel gene fusions have been identified mostly through NGS technologies [4], while most of them have now been described as probably passenger mutations which show little or no effect on tumorigenesis [2]. RNA-Seq is a useful tool for the discovery of gene fusions in cancer transcriptomes and has already become the primary technology for discovering gene fusions. Some open databases of gene fusions in cancer from RNA-seq data have been set up, including Fusion Cancer [41], while we are still faced with the challenge of minimizing false positives in RNA-seq result [19,40,42]. In addition, there are lower proportions (about 3%) of recurrent fusion genes detected by RNA-seq [4].

WGS is also pervasively used for the fusion gene detection [43]. And it provides us a more comprehensive and integrative way to detect structural variants than RNA-seq, especially for de novo gene fusions. WGS would correct the false-negative results from RNA-seq [44,42,44]. As an example, WGS revealed a distinct phenomenon named “chromothripsis” [39], which means chromosomes in a tumor cell produce hundreds of clustered rearrangements [45]. This complicated rearrangement phenomenon was generated as distinct chromosomes or genomic regions shatter into many segments, which are then pieced together by DNA repair mechanisms inaccurately [46]. Recent WGS study suggested this genomic instability phenomenon in cancers co-segregated with inactivation of DNA maintenance genes, like BRCA1/2 [47], and increasing from patients with germline p53 mutations [48]. Some structural variants without producing fusion genes can also change the expression of nearby genes by changing the functional elements. Although RNA-Seq data can detect most of the transcriptional fusions of these genomic alterations [13], there are still much potential transcriptional consequences of structural variants to be further explored. Integrating data from RNA-Seq and WGS would disclose more genetic variants, as TIGF. However, up to now, there are only few studies comprehensively evaluate the transcriptional fusions from WGS and RNA-Seq [49].

Due to widespread applications of high-throughput NGS technologies, major advancements have been made in computational strategies for fusion gene discovery in recent years [50]. Several computational tools have also been developed for the detection of fusion transcripts using RNA-Seq data, such as MapSplice [51], ShortFuse [52], FusionHunter [44], FusionMap [53], SnowShoes-FTD [54], defuse [55], chimerascan [56], FusionCatcher [57], TopHatFusion [44], BreakFusion [58], EricScript [59], SOAPfuse [60],
FusionQ [61], PRADA [62] and JAFFA [63]. Liu et al. [64] performed a large-scale comparative study by applying these above 15 fusion transcript detection pipelines to 3 synthetic data sets and 3 real paired-end RNA-seq studies and developed a meta-caller algorithm to combine three top-performing methods (FusionCatcher, SOAPfusea and JAFFA). If possible, it is recommended to apply all three above pipelines and combine the results in applications. FusionMatcher (FuMa) is a recently designed fusion genes identical program which can automatically compare and summarize all combinations of two or more datasets in a single run and use one gene annotation, to avoid mismatches caused by tool specific gene annotations [65]. It’s believed that both WGS and RNA-seq have their limitations when used independently, and orthogonal validating both data could generate a more sensitive and specific gene fusion detection. To integrate both RNA-seq and WGS data, INTEGRATE was developed to analyze both data to reconstruct gene fusion junctions and genomic breakpoints by split-read mapping. As a result, it was confirmed to be a highly sensitive and accurate approach for detecting high-confidence gene fusion predictions [66]. However, developing the new generation of fusion genes identifying tools from RNA-seq or other NGS data with both sensitivity and specificity remains an important and open question.

Perspective

Gene fusions have strong association with CNVs and whole genomic instability in cancer, which makes it impossible that revealing the complete genomic consequence through only one strategy up to now. In the future, multi-omics analysis of molecular data, such as DNA sequence mutations, CNVs, RNA profiling, DNA methylation, protein expression and chromatin structure may be required to comprehensively interpret gene fusions in order to understand the biology of cancer genomes. Another integrated approach should be done to interpret gene fusions and identify their impact. It is better to combine the NGS result with high-throughput functional cellular assays and more functional data in cancer genomics. In addition, third-generation sequencing which can produce long read sequences is now attempted to clarify complicated genomic structures, including gene fusions, in cancer genome [67].

Nowadays, as the circulating tumor DNA (ctDNA) and circulating tumor cells (CTCs) are more frequently utilized in research and clinical medicine, the ‘liquid biopsies’ can provide the opportunity to promptly track cancer genome evolution of all cancerous lesions [68]. With the rapid development of highly sensitive and accurate technologies of NGS, it can not only predict the response to treatment, but also monitor minimal residual disease [69,70]. As an example, FGFR2 fusion in ctDNA was readily detectable by quantitative real-time reverse transcription-polymerase chain reaction and corroborated to be more sensitive and specific than previous biomarkers, such as CA125 [71]. It is promising that fusion genes can be detected by NGS in liquid biopsies, in the near future.

Competing Interests

The authors declare that they have no competing interests.

Authors’ Contributions

All authors participated in searching and reviewing literatures. JL and LW drafted the manuscript.

References


