Advances in Fusion Gene Research and Fusion Gene Families in Hematological Malignancies

Xue Chen¹, Fang Wang¹, Tong Wang¹,², Lili Yuan¹, Hongxing Liu¹,²,³

¹Pathology & Laboratory Medicine Division, Hebei Yanda Lu Daopei Hospital, Langfang, 065201, China
²Pathology & Laboratory Medicine Division, Beijing Lu Daopei Hospital, Beijing, 100176, China
³Beijing Lu Daopei Institute of Hematology, Beijing, 100176, China

Corresponding Author: Hongxing Liu, MD, Professor, Beijing Lu Daopei Institute of Hematology, 22 South Tongji Road, Beijing, 100076, China. Tel: +86-13693165205, Email: starliu@pku.edu.cn

Received January 13, 2019; Accepted May 8, 2019; Online Published June 15, 2019

Abstract

Fusion genes (FGs) are major molecular biological abnormalities in hematological malignancies and have become well-established molecular markers for disease classification, risk stratification, and targeted therapies. The long tail phenomenon is observed in the distribution of FGs, which means that except for dozens of the most common FGs, the positive rates of the other FGs are all below 1%, even if they have been frequently reported in the literature. However, the total positive rate of these singly relatively rare FGs is actually not low due to their wide variety and numerous members. We therefore put forward the conception “fusion gene family, FG-FM” to describe fusions that share one common protagonist gene with different partner genes. Most of the FGs in hematological malignancies discovered to date can be classified into about 20 major families. Based on the characteristics of the currently reported FGs, it can be expected that although a great many FGs may be identified in the future, the total number of FG-FMs is rather limited. FGs in the same FG-FM often have similarities in pathogenicity, clinical features, and treatment outcomes. Classifying FGs according to FG-FMs will aid in understanding their pathogenicity and optimizing detection schemes.

Keywords: Gene Fusion, Fusion Gene Family, Long Tail Phenomenon


Introduction

Chromosomal translocations in hematological malignancies often result in the generation of fusion genes (FGs) and fusion proteins which are usually effective therapeutic targets.¹ Sometimes the translocations result in abnormally high expression of certain genes without the formation of fusion transcripts or fusion proteins.²,³ Once a pathogenic FG exists, it is generally the most important molecular abnormality and has definite clinical implications. A total of 98 FGs were listed in the 2016 revision to the World Health Organization (WHO) classification of tumors of hematopoietic and lymphoid tissues, most of which can serve as molecular markers for the diagnosis and classification of hematological malignancies.⁴,⁵ Once FGs of pathological significance are detected, they will be deemed as the major molecular etiology of the hematological malignancies and will be monitored by quantitative polymerase chain reaction (qPCR) with high sensitivity. Therefore, FGs can also be used as important molecular markers for evaluating prognosis, selecting targeted therapy or chemotherapy, and monitoring minimal residual disease (MRD). Today, screening multiple FGs simultaneously and then quantitatively monitoring the positive ones has become a routine clinical application.⁶-⁸

Traditional Methods and Limitations of FGs Identification

The traditional FG identification model first identifies the corresponding recurrent chromosomal translocation through karyotype analysis (such as the discovery of the Philadelphia chromosome), and then focuses on the genes which are involved in the chromosomal region and have been broken and misassembled. The most common FGs with characteristic chromosomal translocations in hematological malignancies, such as BCR-ABL1 and PML-RARA, were identified with this strategy.⁹,¹⁰ However, it usually takes a long process to identify novel FGs using this method, and it has brought difficulties to applications in ascertaining the FGs caused by rare chromosomal translocations in the past. Moreover, FGs caused by cryptic chromosomal translocations or submicroscopic insertions/deletions without abnormal karyotypes (such as FIP1L1-PDGFRα) are difficult to find and identify with this traditional strategy.¹¹

FGs are mutually exclusive, which means that most patients are positive for only one FG. In this study, 36 commonly reported FGs in hematological malignancies in a large cohort

Copyright © 2019 The Author(s). This is an open-access article distributed under the terms of the Creative Commons Attribution License (http://creativecommons.org/licenses/by/4.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.
of Chinese acute leukemia patients were retrospectively analyzed. It was found that 23 FGs were detected in 1292 (41.21%) of 3135 acute myeloid leukemia (AML) cases, and 17 FGs were detected in 712 (28.72%) of 2479 acute lymphoblastic leukemia (ALL) cases. The co-occurrence of two FGs was observed in only 0.22% of AML and 0.24% of ALL patients.

The long tail phenomenon is observed in the distribution of FGs. In AML, only 3 FGs (RUNX1-RUNXI1T1, PML-RARA and CBFB-MYH11) have a positive rate of more than 5%, and the total positive rate of the other 20 FGs is 13.85%. Similarly, only 2 FGs (BCR-ABL1 and ETv6-RUNX1) in ALL have a positive rate of more than 5%, and the total positive rate of the other 15 FGs is 10.93%. Except for the most common 13 FGs, the individual positive rates of the other FGs are all below 1% even if they have been frequently reported in the literature. The long tail distribution phenomenon of FGs has also been supported by several other systematic research reports.

Another feature of the long tail phenomenon is that the total positive rate of all singly relatively rare FGs is actually not low due to their wide variety and numerous members. To date, the exact population of FGs undetected in hematological malignancies and the total positive rate of them remain unknown. For most FGs with pathological significance, even if the positive rate of one single FG is relatively low, it still has definite significance in clinical diagnosis, treatment guidance, and MRD monitoring. Thus, the detection of these rare FGs is also clinically significant. For example, FIP1L1-PDGFRα is rare in hematological malignancies; however, the FIP1L1-PDGFRα fusion protein is 100 times more sensitive than BCR-ABL1 to imatinib. Even in the routine detection of FGs that are already well known, false negative results may occur because of variant fusion transcripts, such as BCR-ABL1 and their variant and rare isotopes. Therefore, how to effectively identify rare and variant FGs poses great challenges for detection methods and data analysis.

**Technological Advances and New Discoveries**

The rapid development of sequencing technology and the decline of sequencing costs in recent years have made next generation sequencing (NGS), including whole genome sequencing (WGS), whole exome sequencing (WES), and whole transcriptome sequencing (WTS), more accessible. WGS can identify breakpoints and splicing sites at the genomic level, while WTS can directly identify abnormal transcripts. WGS and WTS can not only analyze known FGs, but also have unique advantages for identification of unknown rare and variant FGs. In a study of 231 pediatric patients with ALL using WTS, 58 putative functional and predominant FGs were identified in 54.1% of patients, 31 of which were newly discovered. Thanks to the promising application of NGS, a large number of newly identified FGs have been reported in recent years, and new classification patterns of hematological malignancies have been suggested. BCR-ABL1-like ALL, which is now incorporated into the 2016 revision of the WHO classification of neoplastic diseases of the hematopoietic and lymphoid tissues as a provisional entity, is a subgroup of ALL characterized by a gene-expression profile similar to that of BCR-ABL1-positive ALL, alterations of lymphoid transcription factor genes, and a poor outcome. A wide range of genetic alterations that dysregulate several cytokine receptors and kinase signaling pathways (predominantly ABL and Janus kinases class), including gene rearrangements involving genes including ABL1, ABL2, CRLF2, CSF1R, EPOR, JAK2, NTRK3, and PDGFRB, are found in BCR-ABL1-like ALL. It is a common subtype of B-ALL, representing 7%-25% of new diagnoses, but the incidence of a specific FG is rather rare. So far, hundreds of kinase-related FGs in BCR-ABL1-like ALL have been reported, yet most of them are only seen in single cases in reports. Therefore, it is difficult to study and classify this subgroup of diseases based on the conventional strategy of single gene abnormality identification. Considering that BCR-ABL1-like ALL could benefit from tyrosine kinase inhibitor (TKI) therapy, the application of integrated genomic analysis to improve diagnoses is of vital clinical significance.

Gu et al recently reported their systematic study on a large cohort of 1988 B-ALL subjects to identify chromosomal rearrangements, gene expression profiles, and large-scale copy number alterations. They found that totally 7.4% patients have PAX5 alterations including rearrangements, intragenic amplifications, or non-PAX5 P80R mutations. Children in this subtype were more commonly classified as high risk rather than standard risk. 38.5% of them harbored PAX5 rearrangements involving 24 partner genes, the most frequent of which were PAX5-ETV6, PAX5-NOLAL, PAX5-AUTS2, and PAX5-CBFA2T3. The long tail phenomenon was also observed in the distribution of PAX5-FGs.

The ZNF384 gene is located close to the telomeric end of the short arm of chromosome 12, making identification of translocations involving ZNF384 challenging by G-band analysis. WTS found that the total frequency of ZNF384-FGs was 4.1% in 291 B-ALL patients, and TCF3-ZNF384 was the most recurrent with a frequency of 2.4%. Patients harboring ZNF384-FGs have a common immunophenotype of weak CD10 and abnormal CD13 and/or CD33 expression. The gene expression signature of TCF3-ZNF384-positive patients was similar to that of patients with other ZNF384-FGs, but it was significantly different from that of TCF3-PMX1-positive and ZNF384-FGs-negative ALL patients. However, the clinical features of TCF3-ZNF384-positive patients were markedly different from those of EP300-ZNF384-positive patients, exhibiting higher cell counts and a younger age at presentation. Patients harboring TCF3-ZNF384 showed a poor response to steroid and a high frequency of relapse, and some of them had additional activating mutations in RAS signaling pathway genes. Another research showed that EP300-ZNF384 and CREBBP-ZNF384 fusions resulted in loss of histone lysine acetyltransferase activity and increased sensitivity of leukemia cells to histone deacetylase inhibitors. EP300-ZNF384 and CREBBP-ZNF384 may cause epigenetic deregulation with the potential for therapeutic targeting. These studies indicate that ZNF384-FGs-positive ALL has characteristic immunophenotype and can be clustered into...
one subtype of ALL. On the other hand, the clinical features, therapy, and prognosis may be influenced by the functional properties of individual fusion partners.\textsuperscript{21,32}

FGs involving \textit{MEF2D} have recently been identified in B-ALL, and the incidence of \textit{MEF2D}-FGs overall was 2.4\% in consecutive B-ALL patients in a single clinical trial.\textsuperscript{33,34} \textit{MEF2D}-FGs-positive patients frequently presented a cytoplasmic \(\mu\) chain-positive, pre-B immunophenotype and often expressed an aberrant CD5 antigen. Elevated GATA3 expression was also a characteristic feature of \textit{MEF2D}-FGs-positive patients, and \textit{PHF6} mutations showed an unexpectedly high frequency (50\%) in them. \textit{MEF2D}-FGs-positive patients were older (median age 9 years) with elevated WBC counts at the time of diagnosis and were mostly classified as NCI (National Cancer Institute) high risk. Although they responded well to steroid treatment, \textit{MEF2D}-FGs-positive patients showed a significantly worse outcome with more than half of them suffering relapse and death. Relapse was frequently associated with \textit{CDKN2A/CDKN2B} deletions. Therefore, \textit{MEF2D}-FGs represent a distinct subgroup of B-ALL with characteristic immunophenotype and gene expression signatures and poor prognoses.\textsuperscript{33-35}

\textbf{Classification and Recognition of FGs as “Fusion Gene Family”}

For a better understanding of the incidences of FGs and their pathological characteristics, we put forward the conception of “fusion gene family, FG-FM” to describe fusions which involve one protagonist gene and multiple fusion partners (Figure 1). Hundreds of FGs have been reported in hematological malignancies so far, and novel FGs may be reported at any time; most of them are rare in individuals.\textsuperscript{4,5,19} The study of their functions and pathogenic mechanisms as well as the choice of detection methods are often confusing. On the other hand, family aggregation phenomenon is quite common, and FGs in the same family often have commonalities in pathogenicity, clinical features, and treatment outcomes.\textsuperscript{28-30} Therefore, classifying FGs according to FG-FMs will aid in better understanding their pathological significance and suggesting new classification patterns of hematological malignancies.

\textit{KMT2A-FM} is a large FG-FM which has been systematically studied and reported in hematological malignancies.\textsuperscript{36} \textit{KMT2A-FM} has lots of members and some of them are more likely to occur in certain types of leukemia. For instance, \textit{KMT2A-ELL} is found almost exclusively in AML, \textit{KMT2A-AFF1} is more likely to occur in ALL, while \textit{KMT2A-MLLT3} is prone to occur in AML and infant ALL.\textsuperscript{4,5} New and rare \textit{KMT2A}-FGs are constantly being reported.\textsuperscript{20} Meyer et al. identified 11 novel partner genes of \textit{KMT2A} in 2345 acute leukemia cases with \textit{KMT2A} rearrangement.\textsuperscript{36} They summarized that a total of 135 different \textit{KMT2A} rearrangements have been identified so far, of which 94 \textit{KMT2A}-FGs have been characterized at the molecular level. Thirty-five of the 94 \textit{KMT2A}-FGs occur recurrently, while 9 of them account for more than 90\% of all FGs. The distribution of \textit{KMT2A-FM} members also reveals a prominent long tail characteristic.

Because \textit{KMT2A} has a large amount of known and an even much greater number of unknown partners, it is difficult to accurately calculate the positive rate of \textit{KMT2A}-FGs. It is not uncommon to find chromosome translocations involving 11q23, where the \textit{KMT2A} gene is located, but without

\begin{figure}[h]
  \centering
  \includegraphics[width=\textwidth]{figure1.png}
  \caption{Summary View of the Key Fusion Families (FG-FMs) in BCR-ABL1-like ALL. Each colored region represents an FG-FM. The lines connecting gene symbols indicate fusion partners. The thickness of each line reflects the frequency of the observed fusion.}
\end{figure}
detectable relatively common KMT2A FGs due to the long tail effect. The translocations may result in rare known or unknown KMT2A FGs or occasionally involve other genes (e.g., ZBTB16) rather than KMT2A in this chromosome region.\textsuperscript{37}

Another FG-FM with numerous family members in hematological malignancies is PDGFRB-FM. About 30 PDGFRB-FGs have been recorded in the 2016 revision to the WHO classification of tumors of hematopoietic and lymphoid tissues, with TEL-PDGFRB being the most common one. Most of them are sensitive to TKIs such as imatinib.\textsuperscript{4} Some PDGFRB-FGs are seen in BCR-ABL1-like B-ALL and are also sensitive to TKIs.\textsuperscript{3,4} However, most of the PDGFRB-FGs are reported only in single cases, thus exerting great difficulties in detection and applying targeted therapy.

According to the concept of “FG-FM” and recent literature, the 98 FGs listed in the 2016 revision to the WHO classification of tumors of hematopoietic and lymphoid tissues can be classified into 18 FG-FMs, such as KMT2A-FM, RARA-FM, PDGFRB-FM, and PDGFRB-FMs, with only 12 FGs like CBFB-MYH11. STIL-TAL1 cannot be classified into any family.\textsuperscript{1,2} Along with the MEF2D-FM, ZNF384-FM, and PAX5-FM, which were recently reported in the literature,\textsuperscript{4,3,35} most of the FGs in hematological malignancies discovered to date can be clustered into about 20 major FG-FMs (Table 1). Furthermore, the content of FGs that can be covered in these FG-FMs far exceeds the known FGs in de facto. Based on the characteristics of the currently reported FGs, it can be expected that although numerous FGs may be identified in the future, the total number of FG-FMs is rather limited.

### Challenges in the Detection and Application of FGs

Genomic analysis has now provided generic methods for FG identification; however, high experimental costs and analytical thresholds are still being faced. Moreover, WGS or WTS generate a lot of background sequences and many abnormal splicing sequences without pathological significance.\textsuperscript{21} How to effectively analyze and differentiate the key FG sequences from multitudinous background sequences is a great challenge to the analyst's professional background and analytical capabilities. Care should also be taken to avoid false negative results caused by methodologies or analysis procedures.

FGs in the same FG-FM have commonalities in pathological and clinical features in general, and the concept of FG-FM will help to speculate on and study the functions of newly identified FGs. Notably, attention should also be paid to the influence of partner genes on the pathogenicity of FGs. If both partners of one FG are protagonist genes of one FG-FM, the FG should be classified mainly according to its major pathogenicity and corresponding clinical features. For example, the gene expression signature and clinical manifestations of PAX5-JAK2-positive ALL are similar to that of BCR-ABL1-like ALL, so PAX5-JAK2 should be classified into the JAK2-FM rather than the PAX5-FM.\textsuperscript{1,2,22}

The majority of FGs with pathogenicity can be used as molecular markers for MRD monitoring after treatment. As the most common FG in hematological malignancies, the quantitative standardization of BCR-ABL1 has been carried out for more than ten years, but there are still many problems to be solved.\textsuperscript{3,4} Every FG-FM may consist of numerous but individually fairly rare FGs. How to effectively design suitable detection methods and how to promote the standardization of quantitative testing need further optimized solutions. For the quantitative detection of rare FGs, using internationally traceable standardized samples is not practical. Digital PCR technology can provide absolute quantification without calibrator and standard curve, yet a specific detection method still needs to be designed for each FG. Digital PCR technology is limited to transcript type, such as BCR-ABL1 e1a2 and e13a2.\textsuperscript{40}

### Conclusions

Massive genomic sequencing and analysis can aid the identification of FGs, especially the rare ones, and the further exploration of their molecular etiology in hematological malignancies. The concept of FG-FM simplifies the understanding of FGs based on their functional clustering and pathogenicity. At the same time, it also brings some new challenges to designing suitable detection methods, which needs persistent efforts together with technological progression and an understanding of the panoramic view of FGs and FG-FMs in hematological malignancies.

---

**Table 1. Gene Translocations Reported in WHO 2016 and Recent Literature**

<table>
<thead>
<tr>
<th>FG-FMs</th>
<th>FGs not Belonging to any FG-FMs</th>
<th>Gene Translocations Without Fusion Proteins</th>
</tr>
</thead>
<tbody>
<tr>
<td>ABL1-FM</td>
<td>BIRC3-MALT1</td>
<td>BCL2-TL</td>
</tr>
<tr>
<td>ABL2-FM</td>
<td>CBF2-T3-GLIS2</td>
<td>BCL6-TL</td>
</tr>
<tr>
<td>ALK-FM</td>
<td>CBFB-MYH11</td>
<td>CD274 (PDL1)-TL</td>
</tr>
<tr>
<td>CSF3R-FM</td>
<td>CTLA-CD28</td>
<td>CIITA-TL</td>
</tr>
<tr>
<td>ETV6-FM</td>
<td>DEK-NUP214</td>
<td>CRLF2-TL</td>
</tr>
<tr>
<td>FGFR3-FM</td>
<td>IFR5-SYK</td>
<td>DUSP2-TL</td>
</tr>
<tr>
<td>JAK2-FM</td>
<td>KMTA-CREBBP</td>
<td>DUX4-TL*</td>
</tr>
<tr>
<td>KMT2A-FM</td>
<td>MYB-CAT1A</td>
<td>EPOR-TL</td>
</tr>
<tr>
<td>MEF2D-FM*</td>
<td>NPM1-TYK2</td>
<td>IGH@-TL</td>
</tr>
<tr>
<td>NTRK3-FM</td>
<td>PICALM-MLL10</td>
<td>MYC-TL</td>
</tr>
<tr>
<td>NUP98-FM</td>
<td>RBM15-MKL1</td>
<td>PDCD1LG2 (FDL2)-TL</td>
</tr>
<tr>
<td>PAX5-FM*</td>
<td>STIL-TAL1</td>
<td>TCR@-TL</td>
</tr>
<tr>
<td>PDGFRB-FM</td>
<td></td>
<td></td>
</tr>
<tr>
<td>RARA-FM</td>
<td></td>
<td></td>
</tr>
<tr>
<td>RUNX1-FM</td>
<td></td>
<td></td>
</tr>
<tr>
<td>TBL1XR1-FM</td>
<td></td>
<td></td>
</tr>
<tr>
<td>TCF3-FM</td>
<td></td>
<td></td>
</tr>
<tr>
<td>TP63-FM</td>
<td></td>
<td></td>
</tr>
<tr>
<td>TYK2-FM</td>
<td></td>
<td></td>
</tr>
<tr>
<td>ZNF384-FM*</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Abbreviations: FG, fusion gene; FM, family; TL, translocation. * indicates the FG-FM or gene-TL not mentioned in the WHO 2016, but reported in recent literature.
Fusion Genes and Fusion Gene Families in Hematological Malignancies

Authors’ Contributions
All authors contributed equally to this study.

Conflict of Interest Disclosures
The authors have no conflicts of interest to disclose.

Acknowledgments
There are no particular acknowledgments to declare for this review article.

References


