The International Journal of Biochemistry & Cell Biology xxx (2014) xxx–xxx

Contents lists available at ScienceDirect

The International Journal of Biochemistry & Cell Biology

journal homepage: www.elsevier.com/locate/biocel

Review

Sequential combination of karyotyping and RNA-sequencing in the search for cancer-specific fusion genes☆

Ioannis Panagopoulos[a,b,*], Jim Thorsen[a,b], Ludmila Gorunova[a,b], Francesca Micci[a,b], Sverre Heim[a,b]

[a] Section for Cancer Cytogenetics, Institute for Cancer Genetics and Informatics, The Norwegian Radium Hospital, Oslo University Hospital, Oslo, Norway
[b] Centre for Cancer Biomedicine, Faculty of Medicine, University of Oslo, Oslo, Norway

ARTICLE INFO

Article history:
Received 15 January 2014
Received in revised form 30 April 2014
Accepted 9 May 2014
Available online xxx

Keywords:
Karyotyping
RNA-sequencing
Cancer-specific fusion genes

ABSTRACT

Cancer-specific fusion genes are often caused by cytogenetically visible chromosomal rearrangements such as translocations, inversions, deletions or insertions, they can be the targets of molecular therapy, they play a key role in the accurate diagnosis and classification of neoplasms, and they are of prognostic impact. The identification of novel fusion genes in various neoplasms therefore not only has obvious research importance, but is also potentially of major clinical significance. The "traditional" methodology to detect them began with cytogenetic analysis to find the chromosomal rearrangement, followed by utilization of fluorescence in situ hybridization techniques to find the probe which spans the chromosomal breakpoint, and finally molecular cloning to localize the breakpoint more precisely and identify the genes fused by the chromosomal rearrangement. Although laborious, the above-mentioned sequential approach is robust and reliable and a number of fusion genes have been cloned by such means. Next generation sequencing (NGS), mainly RNA sequencing (RNA-Seq), has opened up new possibilities to detect fusion genes even when cytogenetic aberrations are cryptic or information about them is unknown. However, NGS suffers from the shortcoming of identifying as "fusion genes" also many technical, biological and, perhaps in particular, clinical "false positives," thus making the assessment of which fusions are important and which are noise extremely difficult. The best way to overcome this risk of information overflow is, whenever reliable cytogenetic information is at hand, to compare karyotyping and sequencing data and concentrate exclusively on those suggested fusion genes that are found in chromosomal breakpoints.

This article is part of a Directed Issue entitled: Rare Cancers.

Contents

1. Introduction

Molecular studies of cancer-associated chromosomal structural rearrangements, in particular translocations and inversions but also in some cases deletions, have shown that they often exert their effect through deregulation, usually overexpression, of a gene in one breakpoint or creation of a chimeric gene through melting together of parts of two genes, one in each breakpoint. The first

Abbreviations: FISH, fluorescence in situ hybridization; NGS, next generation sequencing; RNA-Seq, RNA sequencing; WES, whole exome sequencing; WGS, whole genome sequencing.

☆ This article is part of a Directed Issue entitled: Rare Cancers.

* Corresponding author at: Section for Cancer Cytogenetics, Institute for Cancer Genetics and Informatics, The Norwegian Radium Hospital, Oslo University Hospital, Montebello, Oslo 0310, Norway. Tel.: +47 22934431; fax: +47 22935477.
E-mail address: ioannis.panagopoulos@rr-research.no (I. Panagopoulos).

http://dx.doi.org/10.1016/j.biocel.2014.05.018
1357-2725/

mechanism is common in lymphatic malignancies whereas the second one, i.e., formation of fusion genes, has been described in many hematological disorders, malignant lymphomas, and solid tumors (Heim and Mitelman, 2009).

Because the chromosomal rearrangement leading to a fusion gene is often seen as the sole aberration at cytogenetic analysis, it is assumed to be a primary tumorigenic event. The gene-level change they lead to will hence also be a primary one, present in all cells that belong to the neoplastic parenchyma if we are dealing with a monoclonal tumorigenic process (Teixeira and Heim, 2011). The identification and characterization of novel fusion genes and the study of their effects on cellular processes are therefore of obvious research interest inasmuch as they provide insights into the pathogenetic process behind the neoplastic growth. The knowledge obtained may also be of clinical importance as experience tells us (Heim and Mitelman, 2009), even prior to the discovery of any drugs directed specifically against the primary molecular rearrangements that unleash the pathogenetic process.

2. Chromosome-level (cytogenetic) methods

During the last two decades, the typical method whereby fusion genes have been detected in tumors has begun with chromosome banding analysis of tumor cells resulting in the detection of a characteristic chromosomal rearrangement. The genomic breakpoints of the said rearrangement, typically a translocation, are then mapped to a distinct band on each chromosome using fluorescence in situ hybridization (FISH) techniques on abnormal metaphase plates with gradually smaller probes such as YACs, BACs, PACs, and Fosmids to find the smallest possible probe which spans the chromosomal breakpoint (the signal for this probe will be split between the two breakpoints). The third step (now already moved beyond the cytogenetic level to the level of genes and DNA primary structure) has involved molecular cloning (Southern blot and various types of PCR amplification) to localize the breakpoint more precisely and identify the genes fused by the chromosomal rearrangement. Although laborious, the above-mentioned sequential procedure has proved very robust and reliable and a plethora of fusion genes have been cloned by such means in various types of malignancies (Heim and Mitelman, 2009). A "short-cut" in the above-mentioned methodology has been the candidate gene approach, i.e., a cancer-relevant gene, possibly one which is already known to be involved in a fusion gene in another type of neoplasia, is considered to be a candidate for the pathogenetically involved gene also in other rearrangements targeting the same chromosomal band. An example is the *EWSR1* gene on 22q12 (Sankar and Lessnick, 2011). In mesenchymal tumors with a chromosomal rearrangement involving 22q, *EWSR1* is considered a candidate gene and its possible involvement is tested for immediately, often using FISH methodology, before other and more systematic investigations are undertaken. A detected rearrangement of *EWSR1* will be followed up with molecular techniques to identify the partner gene. The above-mentioned approach has been used to identify many fusion genes in neoplasms with simple structural chromosomal aberrations or when recurrent breakpoints are detected, but it cannot be used in neoplasms with a normal or complex banding karyotype. Submicroscopic rearrangements per definition as well as in practice leave no clues behind from which the method above can begin the cumbersome FISH-search that is necessary for success.

3. Next generation sequencing

The introduction of next generation sequencing (NGS, also called high throughput sequencing, deep sequencing etc.) techniques has opened up new possibilities in the field of DNA or RNA sequencing (Mardis, 2008) and has spearheaded new approaches in cancer genetics (Meyerson et al., 2010). It involves sequencing of stretches of DNA and alignment to a reference genome. Alignments with unexpected position, orientation or separation distance often reflect genomic rearrangements such as translocations, inversions or deletions. High throughput sequencing is therefore a powerful method for detecting structural rearrangements in an unbiased manner and can detect previously unknown fusions in neoplasms with highly complex karyotypes and cryptic fusions in cases with normal karyotype.

Whole genome sequencing (WGS), whole exome sequencing (WES), and RNA sequencing (RNA-Seq, also known as whole transcriptome sequencing) are the three NGS methodologies for fusion gene detection. WGS is a powerful sequencing technology and provides a comprehensive and unbiased characterization of genomic alterations in cancer genomes. Using WGS technology, a variety of fusion genes have been discovered (Bass et al., 2011; Totoki et al., 2011; Welch et al., 2011). WES selectively sequences the coding regions of the genome and is mostly used to identify somatic point mutations in various cancers (Liu et al., 2013). RNA-Seq provides a picture of the RNA present at a given moment in time enabling a search for alternative gene spliced transcripts, post-transcriptional changes, gene fusions, mutations/SNPs, and changes in gene expression (Chu and Corey, 2012). WGS and RNA-Seq are the two major NGS technologies for fusion gene detection. However, WGS has not been used as much as RNA-Seq for the detection of fusion genes because it requires a large amount of sequencing and intensive computational analysis. The whole process of WGS, from sample preparation to fusion identification and verification, may take months to complete (Welch et al., 2011) and is still expensive compared to RNA-Seq (Sboner et al., 2011; Welch et al., 2011) although the cost of NGS has decreased dramatically during the past few years. Finally, the importance of a fusion gene discovered using WGS relies on its effects on expression and on whether it produces fusion transcripts. Compared to WGS, RNA-Seq only sequences the regions of the genome that are transcribed (and spliced into mature mRNA), which is estimated at between 2% and 6% of the entire genome (Sboner et al., 2010). Another advantage that makes RNA-Seq ideal for the discovery of expressed fusion genes is that it allows for detection of multiple alternative splicing variants resulting from a single fusion event. These distinct features of RNA-Seq, together with its low cost and quick turnaround time, make RNA-Seq very popular in fusion gene detection. Thus, most of the published studies to detect fused oncogenes are based on RNA-Seq (Wang et al., 2013). The main limitation of RNA-Seq is that it cannot detect fusion events involving non-transcribed regions (Kim and Salzberg, 2011). Moreover, it is a challenge to differentiate fusions of interest from artifacts due to the prevalence of gene read-through events, new alternative exons or partially correct mRNA splicing.

For the discovery of fusion transcripts, various programs have been developed the computational features of which are reviewed by Wang et al. (2013). For a detailed description, interested readers are referred to the individual programs, for example FusionFinder, FusionMap, nFuse, and FusionSeq (Francis et al., 2012; Ge et al., 2011; McPherson et al., 2012; Sboner et al., 2010). All the programs for the discovery of fusion transcripts follow a three-step procedure: (i) mapping and filtering, (ii) fusion junction detection, and (iii) fusion gene assembly and selection (Wang et al., 2013). The sequence reads are mapped to the reference sequences. Those mapped concordantly to the sequence of a chromosome or transcript are considered to originate from ordinary transcriptional activity and filtered out. The remaining discordantly aligned reads are further filtered in order to discard those which are less likely to harbor fusion transcripts. The FusionSeq program for example contains three misalignment filters, a large scale sequence

Please cite this article in press as: Panagopoulos I, et al. Sequential combination of karyotyping and RNA-sequencing in the search for cancer-specific fusion genes. Int J Biochem Cell Biol (2014), http://dx.doi.org/10.1016/j.biocel.2014.05.018