Cancer Letters 340 (2013) 152–160

Contents lists available at SciVerse ScienceDirect

Cancer Letters

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

Mini-review

Cancer genome sequencing: Understanding malignancy as a disease of the genome, its conformation, and its evolution

Lalit R. Patel [a,b], Matti Nykter [c], Kexin Chen [d], Wei Zhang [b,*]

[a] The University of Texas Health Science Center, Houston, TX, USA
[b] Department of Pathology, The University of Texas MD Anderson Cancer Center, Houston, TX, USA
[c] Department of Signal Processing, Tampere University of Technology, Tampere 33101, Finland
[d] Department of Epidemiology and Biostatistics, Tianjin Medical University Cancer Institute and Hospital, Tianjin 300060, China

ARTICLE INFO

Keywords:
Next-generation sequencing
Cancer genomics
Transcriptomics
Chromosomal conformation sequencing
Bioinformatics
Tumor heterogeneity

ABSTRACT

Advances in cancer genomics have been propelled by the steady evolution of molecular profiling technologies. Over the past decade, high-throughput sequencing technologies have matured to the point necessary to support disease-specific shotgun sequencing. This has compelled whole-genome sequencing studies across a broad panel of malignancies. The emergence of high-throughput sequencing technologies has inspired new chemical and computational techniques enabling interrogation of cancer-specific genomic and transcriptomic variants, previously unannotated genes, and chromatin structure. Finally, recent progress in single-cell sequencing holds great promise for studies interrogating the consequences of tumor evolution in cancers presenting with genomic heterogeneity.

1. Introduction

Cancer is often described as a population of cells that have either lost genomic integrity or have altered the way they manifest their genome. Another way to say this is that cancer is a genomic and epigenomic disease. Advances in genomic profiling have accordingly benefitted our understanding of cancer [1,2]. Molecular aberrations identified in cancer using genomic profiling predict therapeutic response, providing a suite of diagnostic features that may help inform treatment selection and patient care [3,4]. Cancer genomics is thus at the heart of both basic and translational research aiming to eradicate this deadly disease.

A major limitation of the previous generation of genomics technologies was its reliance on hybridization and probe-based techniques. While highly informative with respect to copy number variations (CNVs) [5], single-nucleotide polymorphisms (SNPs) [6], and differential expression of known transcripts [7], microarrays by design capture only information about genomic variation that can be assessed by probes with predetermined sequences. Reviewed here are advances reflected in the current generation of cancer genomics, which are being propelled by the rapid evolution of high-throughput sequencing technology and concomitant bioinformatics [8]. This generation of genomic technologies is furthering our understanding of the cancer genome by providing insight into cancer-associated somatic sequence variants, genomic conformations, and expression of previously unannotated transcripts (Fig. 1).

Several cancers have been studied extensively using high-throughput sequencing. The Cancer Genome Atlas (TCGA) is an ongoing publicly funded endeavor invested in comprehensively studying the cancer genomes across a broad panel of cancer types [9]. Several whole-exome and a few whole-genome sequencing studies independent of the TCGA have also been reported [10–14].

2. Highly parallel shotgun sequencing

Shotgun sequencing is defined by the practice of sequencing many overlapping genomic fragments while relying on computational algorithms exploiting regions of sequence overlap to assemble fragments into a whole genome. This differs from clone-based sequencing where defined segments of the genome are cloned into and sequenced from bacterial or yeast artificial chromosomes [15]. A significant debate in the late 1990s over the feasibility of applying shotgun-sequencing to the genomes of higher order organisms focused on major concerns such as cost, logistics, and the computational challenges inherent to assembling large genomes from relatively short reads [16,17]. Nevertheless, a first draft of the human genome developed by using shotgun sequencing was published in 2001 [18].

* Corresponding author. Address: Cancer Genomics Core Laboratory, Department of Pathology, Unit 85, The University of Texas M.D. Anderson Cancer Center, 1515 Holcombe Blvd., Houston, TX 77030, USA. Tel.: +1 713 745 1103; fax: +1 713 792 5549.
E-mail address: wzhang@mdanderson.org (W. Zhang).

0304-3835/$ - see front matter © 2012 Elsevier Ireland Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.canlet.2012.10.018

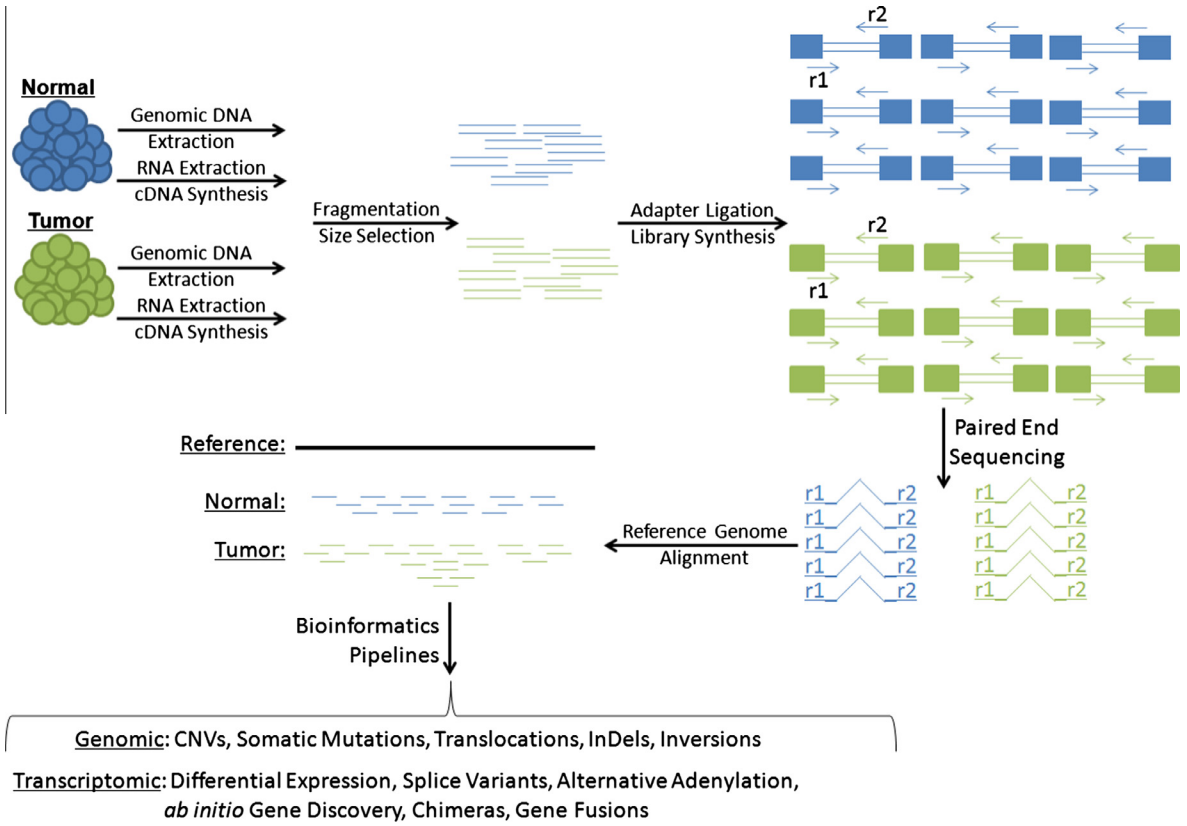

Fig. 1. Shotgun sequencing. Genomic DNA and cDNA synthesized from RNA are fragmented into a size range compatible with the specifications of massively parallel high-throughput sequencing machines. During repair of fragmented ends, distinct adapters are ligated to each end, permitting paired-end sequencing of DNA fragments. Read pairs are then aligned computationally against a reference genome. Read depth at specific loci can be used to assess genomic copy number or RNA expression of genes. Bioinformatic comparisons of normal, tumor, and reference genomes identify sequence variants associated with disease. Discordant read pairs where the ends of a pair align to separate genomic loci nominate structural genomic rearrangements and expressed chimeras. Evidence of junction spanning read pairs supports *ab initio* annotation of novel genes in previously unannotated loci.

While this draft was a critical step toward enabling cancer genome sequencing, the National Human Genome Research Institute estimated the cost of shotgun sequencing in 2001 at nearly $100 million per genome. During the decade since, sequencing technologies have significantly improved their throughput, reducing this cost into the range of thousands of dollars per genome [19]. These advances have made cancer genome sequencing studies economically feasible.

3. Genomic sequencing

Genomic sequencing can be divided into two major categories: whole-genome sequencing and targeted sequencing. Whole-genome sequencing refers to sequencing the full length of the genome, whereas targeted sequencing enables greater depth of coverage for higher resolution analysis of sequence variation within specific genomic elements. The most frequent case of targeted sequencing is whole-exome sequencing, where known exons are enriched and analyzed for mutations. In contrast, whole-genome sequencing identifies structural changes in the cancer genome, such as inversions, deletions, and other genomic rearrangements, in addition to mutations [8].

Targeted exome sequencing efforts have identified recurrent somatic mutations in diseases where oncogenic drivers were previously not well known. For instance, it was recently reported that the most frequent class of genomic aberrations in malignant melanomas is activating events in the MAPK pathway [20]. Interestingly, the same study showed differences between cutaneous, acral, uveal, and mucosal melanomas with respect to their proclivities for accumulating specific aberrations within this pathway. This suggests exploitation of a common pathway across known anatomical subtypes of disease while also laying a foundation for investigating differences in the molecular etiologies of each type. Similarly, the mutational landscape as determined through exome sequencing of high-grade serous ovarian cancer reveals that, while more than 95% of sequenced tumors demonstrate inactivating mutations in participants of the tumor-suppressive TP53-pathway, activating alterations (mutational and amplification) of oncogenes were diverse and rarely recurrent in more than 10% of samples [21]. In a study of colorectal carcinomas, more than 90% of tumor samples presented with activating events in the WNT signaling pathway, while the specific events responsible occurred at low frequencies across a broad panel of genes involved in the WNT pathway [22].

An added layer of complexity is emerging from early whole-genome studies. While the existence of recurrent copy-number alterations in cancer has been well documented using array-comparative genomic hybridization (CGH) studies [23], whole-genome sequencing studies have also revealed a breadth of genomic rearrangements previously underappreciated in many solid tumors [14,21,24]. While accepted as drivers of oncogenesis in hematologic malignancy, recurrent genomic rearrangements have recently also come to the fore as oncogenic events in solid tumors where recurrent functional gene fusions are being identified by sequencing [25].

The emerging theme of complex genetic backgrounds within cancers and differing mechanisms of oncogenesis across cancer