Review

Structural mutations in cancer: mechanistic and functional insights

Koichiro Inaki[1,2] and Edison T. Liu[1,2]

[1] Cancer Biology and Pharmacology, Genome Institute of Singapore, 60 Biopolis Street, Genome, Singapore 138672, Singapore
[2] The Jackson Laboratory, 600 Main Street, Bar Harbor, ME 04609, USA

Next-generation sequencing (NGS) has enabled the comprehensive and precise identification of many somatic structural mutations in cancer. Analyses integrating point mutation information with data on rearrangements and copy number variation have revealed a higher-order organization of the seemingly random genetic events that lead to cancer. These meta-analyses provide a more refined view of the mutational mechanisms, genomic evolution, and combinations of mutations that contribute to tumorigenesis. Structural mutations, or genome-scale rearrangements of segments of DNA, may play a hitherto unappreciated role in cancer through their ability to move blocks of adjacent genes simultaneously, leading to concurrent oncogenic events. Moreover, whole-genome sequencing (WGS) data from tumors have revealed global rearrangements, such as those seen in the tandem duplicator phenotype and in chromothripsis, suggesting that massive rearrangements are a specific cancer phenotype. Taken together, the emerging data suggest that the chromosome structure itself functions as a systems oncogenic organizer.

Genomic changes in cancer

Cancer arises from normal cells that have progressively accumulated mutations that circumvent cellular regulatory controls [1]. Although epigenetic alterations change gene expression and contribute significantly to the oncogenic process, genetic changes represent the 'hard-wiring' of the cancer genome. These mutations span from single nucleotide alterations to copy number changes of entire chromosomes (Figure 1). The consequences either augment the action of an oncoprotein and increase its expression, or silence a tumor suppressor or gatekeeper function. Historically, much of the knowledge about these gene mutations has been dependent on the power and resolution of molecular techniques. Earlier approaches, such as PCR and Sanger sequencing, enabled the identification of point mutations and small indels. Analyses of large structural mutations (SMs), such as focal deletions, single tandem duplications (TDs), amplifications, translocations, and inversions, were difficult to detect when the technologies were limited to cytogenetics, cloning, and Southern blot analysis (listed in the Mitelman Database, http://cgap.nci.nih.gov/Chromosomes/Mitelman) [2,3]. However, the completion of the human genome sequence and the development of NGS techniques radically changed the discovery landscape [4,5]. Single-tag reads can readily assess point mutations and copy number changes, and the development of mate-pair sequencing of fragment ends extended the coverage of analysis to encompass more complex structural variations (Figure 2a–d). Such structural changes required

Corresponding author: Liu, E.T. (liue@gis.a-star.edu.sg), (Edison.Liu@JAX.org).
Keywords: Cancer genome; structural mutation; genomic rearrangement; tandem duplication; chromosome structure; mutational mechanisms.

Glossary

Chromothripsis: a genomic configuration in a tumor where a chromosome appears to have been shattered and reformed in a disorganized manner, probably arising from a single catastrophic event.
Conductor mutation: a SM that does not interrupt a gene or change a regulatory region at or near the break points, but is thought to facilitate subsequent downstream structural events leading to the acquisition of driver-oncogenic functions, such as amplification of oncogenes or deletion of TSGs.
Deletion: in sequence mapping, a deletion occurs when the sequences on both sides of a break point are mapped to the same chromosome and strand, and the 5′ and 3′ mappings are in the correct order but the interval between both mapped sequences are longer than found in the reference genome.
Fork stalling and template switching: one proposed mechanism of MMEJ. Stalled replication forks induced by DNA secondary structure find nearby single-stranded DNA regions based on microhomology, which then mediate a template switch and resumption of DNA replication.
Homologous recombination (HR): uses regions with close sequence homology, such as identical sequences on the sister chromatids, which results in complete, high-fidelity repair. When the chromosomal homolog, rather than the sister chromatid, is used, the outcome is allelic recombination resulting in a gene conversion event. When other misaligned DNA sources are involved, such as SDs or low copy repeats, the subsequent repair is termed 'non-allelic homologous recombination' and is a major source of normal genomic structural variation.
Inversion: in sequence mapping, an inversion occurs when both sides of a break point are mapped to the same chromosome but on opposite strands. An inversion is unpaired if the complementing end of the inversion is not detected by sequencing, and it is usually found when the SMs are in a highly complex and inverted in orientation. When a genomic region has been duplicated but the two copies are in opposite orientation from the break point, the structure is termed a 'fold-back' inversion and is thought to be caused by breakage–fusion–bridge cycles [25,28,57].
Microhomology-mediated break-induced replication (MMBIR): when a replication fork encounters a nick on the template, it collapses and a DSB can occur. Through MMBIR, the collapsed replication fork can find and inappropriately invade nearby single-stranded DNA regions with the help of microhomology, switch templates, and resume DNA replication to repair the DSB.
Microhomology mediated end-joining (MMEJ): a process of rejoining DNA ends with nucleic acid overhangs. Fused ends as a result of MMEJ have overlapping microhomology [>1 bp and usually (>96%) <8 bp].
Non-homologous end-joining (NHEJ): a process to rejoin broken ends in DSBs without the need for homologous templates. NHEJ can cause small insertions with new sequence between the break points or deletions flanking the break.
Tandem duplication (TD): in sequence mapping, a TD occurs when both sides of a break point map to the same chromosome and strand but the 5′ mapped segment is located downstream of the 3′ segment.
TD phenotype: a genomic configuration in which there are many (often over 50–100) individual TDs seen in the cancer genomic sequence.
Translocation: in sequencing mapping, a translocation occurs when the sequences on either side of a break point are mapped to different chromosomes.

0168-9525/$ – see front matter http://dx.doi.org/10.1016/j.tig.2012.07.002 Trends in Genetics, November 2012, Vol. 28, No. 11

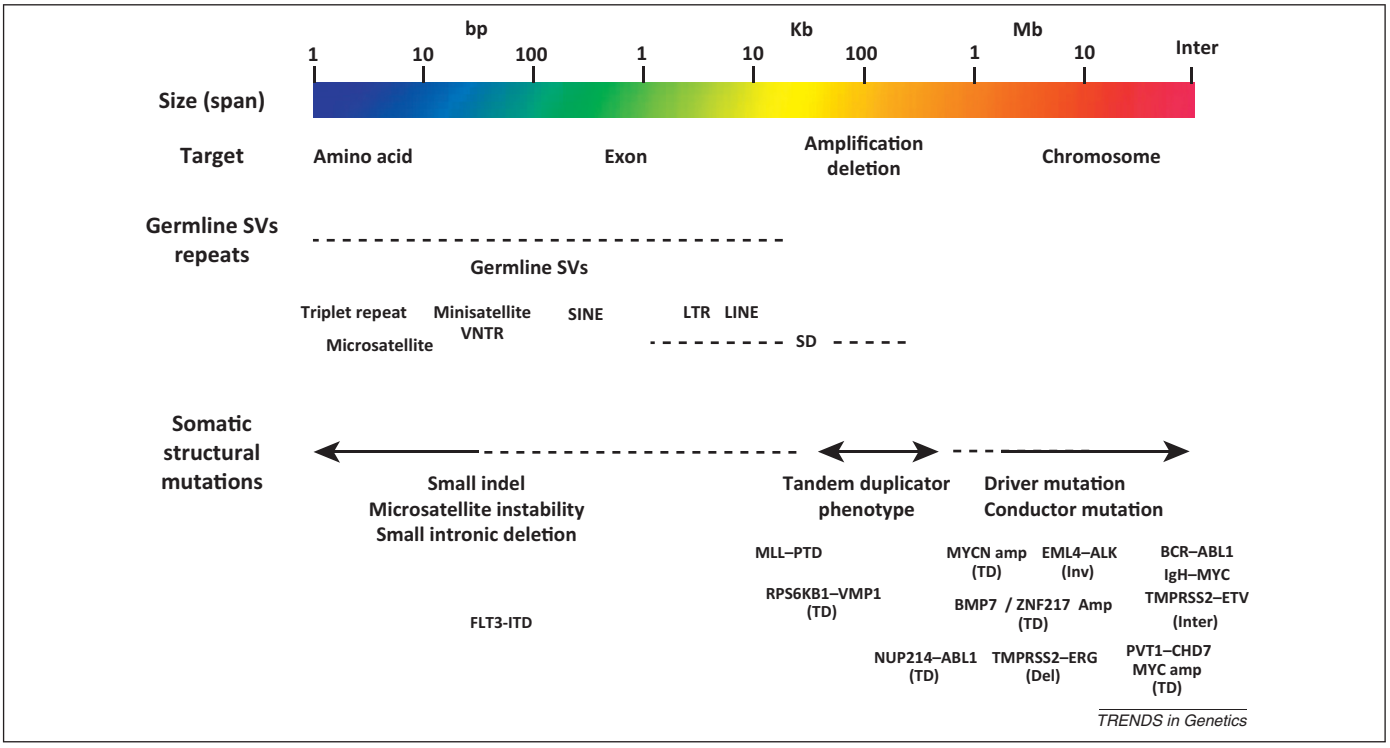

Figure 1. Size and/or span of germline structural variations (SVs), repeats, and somatic structural mutations (SMs). Figure shows the span of mutations (distance between break points) from 1 bp to infinity (different chromosome), targets of such changes, the range of most of germline SVs, well-known genomic repeats, and ranges of SMs with examples. Abbreviations: Amp, amplification; Del, deletion; Inter, interchromosomal; Inv, inversion; ITD, internal TD; LINE, long interspersed element; LTR, long terminal repeat; PTD, partial TD; SD, segmental duplication. SINE, short interspersed element; TD, tandem duplication; VNTR, variable number of tandem repeats.

deeper sequence coverage for statistical sampling, and the dramatic reduction in sequencing costs in recent years now allows for the statistical ascertainment of chromosomal break points in a cost-accessible manner.

With these advances in sequencing technologies, various types of SM have been reported (Table 1) that often overlap to form complex rearrangements (e.g., deletions within inversions) and copy number changes. New terms, such as chromothripsis and tandem duplicator phenotype, have emerged based solely on findings from WGS (Figure 2d).

Recently, large numbers of cancer samples (ranging from 100 to a few thousand clinical samples and cell lines) were analyzed by limited sequencing or array comparative genomic hybridization (CGH) to detect mutations or changes in copy number through multiple collaborations [6–16]. The availability of these data sets has enhanced significantly the process of oncogene and tumor suppressor gene (TSG) discovery.

The current data on somatic point mutations (COSMIC, http://www.sanger.ac.uk/genetics/CGP/cosmic/) [11–18] show some commonly mutated genes, such as *TP53*, *KRAS*, *PTEN*, *CTNNB1*, *ARID1A*, *PI3KCA*, *CDKN2A*, *NF1*, *APC*, *SMARCA4*, *EGFR*, and *BRAF*, across major cancer types; however, mutations in only a few signaling pathways, such as receptor tyrosine kinase (RTK), RAS, RB, TP53, WNT, and NOTCH, are significantly and commonly overrepresented in cancers. This apparently limited diversity of mutational targets may restrict the impact of single base point mutations in cancer evolution. By contrast, the universe of SMs is more varied and induces not only gene activation and silencing, but also transcriptional deregulation and the generation of chimeric genes. Given these observations and the new insights now arising from NGS, we focus here on the role of SMs in cancer.

SM and fusion genes

Type of SMs

Since 2008, more than ten studies have reported WGS by NGS of various types of clinical cancer samples and cell lines (Table 1) [19–32]. Although there is no common nomenclature for SMs, we classified the reported SMs into four categories based on discordant mapping against the reference genome (Figure 2a): deletions, TDs, inversions, and translocations (see Glossary).

A comparison of the number of SMs and germline structural variants (SVs) in healthy cells revealed that most deletions, inversions, and insertions are germline SVs. By contrast, cancer cells are marked by an overrepresentation of TDs, unpaired inversions, and translocations [28]. Furthermore, the span distributions of the intrachromosomal events (i.e., distance between two break points) of SVs and SMs are different, in that normal SVs have smaller span lengths than do cancer SMs (Figure 1) [19,28]. Some of the normal small SVs ranging up to 10 kb are thought to be caused by recombination at repeats in the genome, such as long interspersed elements (LINEs) and short interspersed elements (SINEs) [19].

Functional consequences of SMs

Examining a large number of SMs reveals ones whose break points directly generate an oncogenic element, such as a driver fusion transcript (e.g., *EML4–ALK*) or a deletion of a critical TSG (e.g., *TP53*) (Figure 3a). It has been