Trends in Cancer

CellPress

Review

Intrinsic Molecular Processes: Impact on Mutagenesis

Byungho Lim,[1] Jihyeob Mun,[2,3] and Seon-Young Kim[2,3,*]

Mutations provide resources for genome evolution by generating genetic variability. In addition, mutations act as a driving force leading to disease pathogenesis, and thus have important implications for disease diagnosis, prognosis, and treatment. Understanding the mechanisms underlying how mutations occur is therefore of prime importance for elucidating evolutionary and pathogenic processes. Recent genomics studies have revealed that mutations occur non-randomly across the human genome. In particular, the distribution of mutations is highly associated with intrinsic molecular processes including transcription, chromatin organization, DNA replication timing, and DNA repair. Interplay between intrinsic processes and extrinsic mutagenic exposure may thus imprint a characteristic mutational landscape on tumors. We discuss the impact of intrinsic molecular processes on mutation acquisition in cancer.

Cancer Genomics: Beyond Somatic Mutations

Mutation is an essential evolutionary and pathogenic driver. Mutations generate genetic variability both within and among species by introducing DNA structural and sequence alterations [1]. Mutations cause **gain-of-function**, **loss-of-function**, and **switch-of-function** alterations (see Glossary) to genes and regulatory mechanisms at the basis of cancer pathogenesis, and are important in disease diagnosis, prognosis, and treatment [2]. Global consortia (Box 1) have identified comprehensive mutational landscapes [3–14] and tumor type-specific mutational signatures in multiple cancers [15,16]. Beyond the cataloging of mutations, recent cancer genomics studies have focused on: (i) **intratumoral heterogeneity** and tumor evolution [17,18], (ii) the development of therapeutic strategies for **personalized precision medicine** [19,20], and (iii) the impact of molecular processes – genomic, epigenomic, and transcriptomic events – on mutagenesis.

Mutations are acquired through interactions between intrinsic (e.g., genetic) and extrinsic (e.g., lifestyle, environment) factors. Recent studies have begun to reveal that intrinsic molecular processes are important in mutagenesis across the human genome. In particular, intrinsic processes such as gene expression [21–23], **chromatin organization** [24–31], DNA replication timing [32–35], protein binding [36–40], and DNA repair [41–45] were found to correlate with the non-random distribution of mutations across the genome. Therefore, mutagenesis is not merely a random event but is to some extent a programmed and unavoidable phenomenon that depends on genomic position-specific intrinsic processes. This review discusses the impact of intrinsic molecular processes on mutagenesis and how they generate the characteristic distributions of mutations in normal and cancer genomes.

Molecular Events Leading to the Acquisition of Somatic Mutations in Normal Cells

The acquisition of mutations in normal cells has not been well elucidated. Unlike tumor cells, normal cells generally do not undergo explosive **clonal expansion** [46]. As a result, the

Trends

Identifying somatic mutations from large cohorts is an active research field of great interest in current cancer genomics.

Beyond the cataloging of somatic mutations, recent studies are now moving towards a multifaceted understanding of these mutations. For example, by tracking mutations in temporally ordered multiregion samples, we can understand longitudinal and spatial tumor heterogeneity and clonal evolution. Examination of mutual exclusivity and co-occurrence between mutations will assist the development of therapeutic strategies.

A more fundamental question is how mutations are acquired throughout the genome. Recent integrative analyses encompassing the genome, epigenome, and transcriptome have revealed that intrinsic molecular processes are deeply involved in the generation of mutations.

[1]Research Center for Drug Discovery Technology, Division of Drug Discovery Research, Korea Research Institute of Chemical Technology, Daejeon, Korea
[2]Personalized Genomic Medicine Research Center, Korea Research Institute of Bioscience and Biotechnology (KRIBB), Daejeon, Korea
[3]Department of Functional Genomics, University of Science and Technology, Daejeon, Korea

*Correspondence: kimsy@kribb.re.kr (S.-Y. Kim).

Box 1. Global Consortia to Study Genetic Variation

After the era of the Human Genome Project, researchers collaborated to delineate genetic variations among human populations. The International HapMap Project[iii] (launched in 2002) generated a haplotype map of the human genome, thereby revealing ethnic differences in genetic variations and enabling the discovery of disease- or phenotype-associated loci through linkage disequilibrium mapping. The 1000 Genomes Project[iv] (launched in 2008) created a global reference for human genetic variation by sequencing at least 1000 participants from various populations [103]. These projects surveyed genetic variations that originate in the germline.

In the era of next-generation sequencing, international consortia have produced comprehensive, multidimensional datasets across multiple cancer types. The Cancer Genome Atlas (TCGA)[ii] has extensively examined genomic, epigenomic, transcriptomic, and proteomic alterations across 33 cancer types [104]. The International Cancer Genome Consortium (ICGC)[v] coordinated a large number of research projects to obtain a compendium of genomic aberrations from 50 different cancer types. These projects contribute to the cancer genomics field by accelerating our understanding of the molecular basis of cancer, developing data analysis pipelines, and encouraging unconstrained data access and analyses.

frequency of acquired mutations is generally low in normal tissues unless mutations arise in adult stem cells that continuously propagate genomic features to daughter cells. For this reason it is difficult to identify somatic mutations in normal tissues by using conventional sequencing technology. The best method to detect low-frequency mutations is single-cell sequencing. However, sequencing of at least hundreds of single cells is necessary to obtain the representative mutational landscape of normal tissues [47].

Instead, mutational processes operating in normal cells can be indirectly inferred by considering the distributions of mutations that are shared by most cancers. If a mutational process continually operates during the lifetime of an organism, the extent of mutation will be proportional to the age of the cancer patient at diagnosis. Analysis of 10 250 cancer genomes across 36 cancer types revealed that the number of C > T mutations at NpCpG trinucleotides is significantly correlated with age at cancer diagnosis [46]. This process is referred to as a 'clock-like' mutational process because it is age-dependent and occurs at a constant rate during postnatal life [46]. Consistently, other studies have shown that the number of C > T mutations is strongly positively correlated with age [48–50]. The molecular process underlying this signature is spontaneous deamination of 5-methylcytosine (5mC) at CpG dinucleotides [51]. The resulting T:G mismatches lead to C > T mutations if the unrepaired DNA is replicated. Notably, cancer types with high rates of this signature originate from epithelial cells with high cell turnover. For example, stomach cancer (23.7 mutations/Gb/year) and colorectal cancer (23.4 mutations/Gb/year) originate from epithelial cells and show substantially higher rates of the 'clock-like' mutational signature than do other cancer types such as myeloma (3.1 mutations/Gb/year) and pilocytic astrocytoma (0.65 mutations/Gb/year) [46]. This difference is because high mitotic rates of epithelial cells may bypass complete DNA repair during DNA replication.

Because the mutational landscape in normal tissues is primarily established by somatic mutations that arise in adult stem cells, a recent study examined the mutational distribution in adult stem cells using *in vitro* **organoid** culture [52]. Clonal expansion through organoid cultures facilitated the sequencing of adult stem cells, and this approach identified 79 790 mutations from 45 independent clonal organoid cultures of three tissue types (colon, small intestine, and liver). From these data it was found that adult stem cells gradually accumulate mutations as a function of age at a rate of ~40 novel mutations per year. Remarkably, the mutational spectrum of cancer driver genes (e.g., *APC*, *TP53*, *SMAD4*, and *CTNNB1*) in adult stem cells was highly homologous to the mutational spectrum of those genes in cancer cells. Thus, mutational processes operating in adult stem cells seem to be similar to those in cancer cells. At the same time, this observation suggests that mutational processes in adult stem cells

Glossary

Abortive transcription: a normal process of early transcription in which RNA polymerase repeatedly synthesizes and releases short RNAs.
APOBEC signature: strand-coordinated clustered or dispersed C-to-T or C-to-G (C-to-A in rare cases) mutations in TCW motifs where C is the base mutated and W is A or T.
Chromatin organization: the multifaceted complex conformation of a genome; it consists of interactions between DNA, RNA, and proteins.
Clonal expansion: the generation of a population of identical cells by the proliferation of a single cell.
Gain-of-function: abnormal activation of a gene (typically of oncogenes in cancer).
Hi-C: a sequencing-based high-throughput technique that detects local and long-range interactions of chromatin.
Hypersensitivity to DNase I cleavage: exposed genomic regions that are sensitive to cleavage by the DNase I endonuclease (typically open chromatin).
Intratumoral heterogeneity: intermingling of cancer cells that have distinct molecular and phenotypic characteristics within a tumor.
Kataegis: hypermutations clustered within genomic regions (≥2 mutations within a 10 kb genomic window).
Loss-of-function: abnormally abolished function of a gene (typically of tumor suppressors in cancer).
Mutational asymmetry: a biased, distinct mutational landscape between the two DNA strands.
MutSα: a heterodimer of mismatch repair components MSH2 and MSH6.
Organoid: *in vitro* production of 3D organ mimetics.
Personalized precision medicine: tailored therapeutics that specifically target abnormal molecular alterations of each patient.
POLE: the DNA polymerase ε catalytic subunit, which has DNA repair activity through its proofreading capacity.
Purifying selection: a phenomenon that selectively removes deleterious alleles during evolution.