Leading Edge
Essay

Cell

The Cell Biology of Genomes: Bringing the Double Helix to Life

Tom Misteli[1,*]
[1]National Cancer Institute, National Institutes of Health, Bethesda, MD 20892, USA
*Correspondence: mistelit@mail.nih.gov
http://dx.doi.org/10.1016/j.cell.2013.02.048

The recent ability to routinely probe genome function at a global scale has revolutionized our view of genomes. One of the most important realizations from these approaches is that the functional output of genomes is affected by the nuclear environment in which they exist. Integration of sequence information with molecular and cellular features of the genome promises a fuller understanding of genome function.

It was a moment of scientific amazement in 1953 when Watson and Crick revealed the structure of DNA. The magnificence of the double helix and its elegant simplicity were awe inspiring. But more than just being beautiful, the double helix immediately paved the way forward; its structure implied fundamental biological processes such as semiconservative replication and the notion that chemical changes in its composition may alter heritable traits. The linear structure of DNA laid the foundation for the concept that a string of chemical entities could encode the information that determines the very essence of every living organism. The beauty of the double helix was the promise that, if the sequence of bases in the genome could be mapped and decoded, the genetic information that underlies all living organisms would be revealed and the secret of biological systems would be unlocked.

The idea of linearly encoded genetic information has been spectacularly successful, culminating in the recent development of powerful high-throughput sequencing methods that now allow the routine reading of entire genomes. The conceptual elegance of the genome is that the information contained in the DNA sequence is absolute. The order of bases can be determined by sequencing, and the result is always unequivocal. The ability to decipher and accurately predict the behavior of genome sequences was appealing to the early molecular biologists, has given rise to the discipline of molecular genetics, and has catalyzed the reductionist thinking that has driven and dominated the field of molecular biology since its inception.

But the apparent simplicity and deterministic nature of genomes can be deceptive. One of the most important lessons learned from our ability to exhaustively sequence DNA and to probe genome behavior at a global scale by mapping chromatin properties and expression profiling is that the sequence is only the first step in genome function. In intact living cells and organisms, the functional output of genomes is modulated, and the hard-wired information contained in the sequence is often amplified or suppressed. While mutations are an extreme case of genome modulation, most commonly occurring changes in genome function are more subtle and consist of fluctuations in gene expression, temporary silencing, or temporary activation of genes. Although not caused by mutations, these genome activity changes are functionally important.

Several mechanisms modulate genome function (Figure 1). At the transcription level, the limited availability of components of the transcription machinery at specific sites in the genome influences the short-term behavior of genes and may make their expression stochastic. Epigenetic modifications are capable of overriding genetically encoded information via chemical modification of chromatin. Similarly, changes in higher-order chromatin organization and gene positioning within the nucleus alter functional properties of genome regions.

The existence of mechanisms that modulate the output of genomes makes it clear that a true understanding of genome function requires integration of what we have learned about genome sequence with what we are still discovering about how genomes are modified and how they are organized in vivo in the cell nucleus.

The Stochastic Genome

The genome is what defines an organism and an individual cell. It is therefore tempting to assume that identical genomes behave identically in a population of cells. We now know that this is not the case. Individual, genetically identical cells can behave very differently even in the same physiological environment. It is rare to find a truly homogeneous population of cells even under controlled laboratory conditions, as anyone who has tried to make a cell line stably expressing a transgene knows. Much of the variability in biological behavior between individual cells comes from stochastic activity of genes (Raj and van Oudenaarden, 2008).

Genes are by definition low-copy-number entities, as each typically only exists in two copies in the cell. Similarly, many transcription factors are present in relatively low numbers in the cell nucleus. The low copy number of genes and transcription factors makes gene expression inherently prone to stochastic effects (Raj and van Oudenaarden, 2008). Numerous observations make it clear that gene expression is stochastic in vivo. For example, dose-dependent increases in gene expression after treatment of cell populations with stimulating

ligands, such as hormones, are often brought about by high expression of target genes in a relatively small number of cells in the population rather than by a uniform increase in the activity in all cells. Stochastic gene behavior is most evident in single-cell imaging approaches, and mapping by fluorescence in situ hybridization of multiple genes, which according to population-based PCR analysis are active in a given cell population, shows that only a few cells transcribe all "constitutively active" genes at any given time. Most cells only express a subset of genes, and the combinations vary considerably between individual cells. These observations suggest that many genes blink on and off and are expressed in bursts rather than in a continuous fashion (Larson et al., 2009).

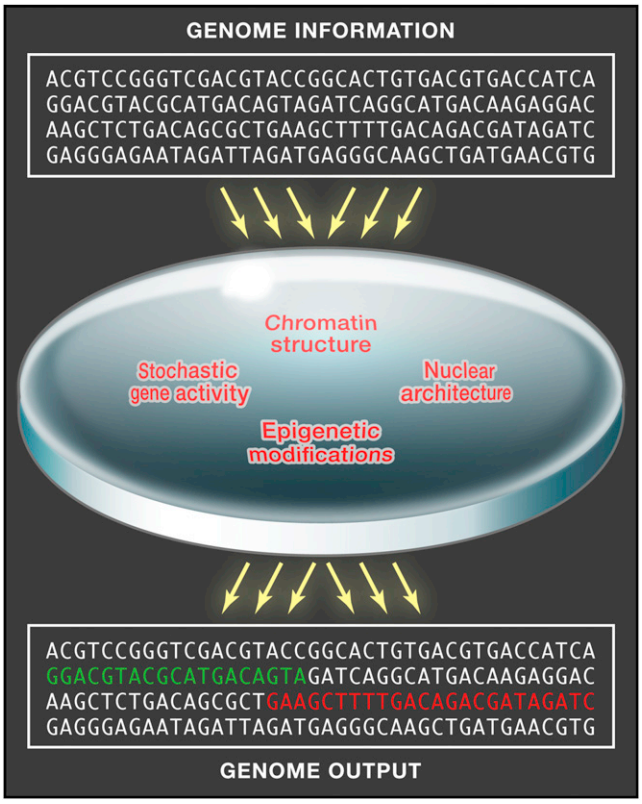

Figure 1. From Primary Sequence to Genome Output
The hard-wired primary information contained in the genome sequence is modulated at short or long timescales by several molecular and cellular events. Modulation may lead to activation (green) or silencing (red) of genome regions.

The molecular basis for stochastic gene expression is unknown. There are several candidate mechanisms, all of which are related to genome or nuclear organization. Most genes require some degree of chromatin remodeling for activity, which is thought to make regulatory regions accessible to the transcription machinery. Several observations suggest that chromatin remodeling contributes to the stochastic bursting of gene expression. Maybe most compelling is the finding that genes located near each other on the same chromosome show correlated blinking behavior, indicating that a local chromosome property, such as chromatin structure, drives stochastic behavior (Becskei et al., 2005). Furthermore, altering chromatin, for example by deletion of chromatin remodeling machinery, affects stochastic variability in yeast. It can be envisioned that the stochastic behavior of genes is caused by the requirement for cyclical opening of chromatin regions. Open chromatin has a limited persistence time, and maintaining chromatin in an open state requires the cyclical action of chromatin remodelers. Whether an "active" gene is transcribed at any given time may thus depend on the transient condensation status of its chromatin at a particular moment.

A second mechanism to impose nonuniform stochastic genome activity may be the local availability of the transcription machinery at a gene. Although transcription factors are able to relatively freely diffuse through the nuclear space, and in this way effectively scan the genome for binding sites, their availability and functionality at a given local site may undergo significant temporal fluctuations (Misteli, 2001). The local availability of transcription complexes may affect transcription frequency positive or negatively. On the one hand, it is possible that relatively stable preinitiation complexes persist on a given gene, where they may support multiple rounds of transcription and in this way boost initiation frequency. On the other hand, assembly of the full polymerase is a stochastic and relatively inefficient event itself. In order for a functional polymerase complex to assemble, individual transcription machinery components associate with chromatin in a step-wise fashion, and formation of the mature polymerase complex involves multiple partially assembled intermediates, many of which are unstable and disintegrate before a functionally competent complex is formed (Misteli, 2001). The inefficiency of polymerase assembly may create stochasticity at an individual locus.

A further contributor to stochastic gene expression may be the organization of transcription events in transcription factories. These hubs of transcription consist of accumulations of transcription factors to which multiple genes, often located on distinct chromosomes, are recruited (Edelman and Fraser, 2012). Typically only a few hundred such transcription factories are observed in a mammalian cell nucleus. It is possible that some genes need to physically relocate from nucleoplasmic locations to transcription factories. A nominally "active" gene locus that is not associated with a transcription factory may thus be stochastically silent. The relatively low number of transcription sites makes them a limiting factor in the transcription process and thus a potential mediator of stochastic gene expression.

Epigenetics—And When Epigenetics Is Not Epigenetics

Stochastic effects modulate genome output on short timescales. A mechanism to modulate the hardwired information of genomes on longer timescales is via epigenetics. The Greek-derived "Epi" means "over" or "above," and epigenetic effects are defined as heritable changes in genome activity caused by mechanisms other than changes in DNA sequence. Epigenetic events are mediated by