Editorial

The biological processes necessary for the development and survival of all living organisms rely on control systems that allow for the precisely regulated expression of genetic information. The term ‘gene’ was first invoked in 1909 to represent the physical entities within cells underlying discrete and heritable traits, such as those that had first been revealed by the experiments of Mendel half a century earlier [1]. In the 1940s, it was shown that different gene mutations lead to defects in specific steps in metabolic pathways, forming the basis for the concept ‘one gene, one polypeptide’ [2]. This perspective was further reinforced when, following the solution of DNA structure and the deciphering of the genetic code instructing the translation of RNA transcripts, Francis Crick [3] introduced the ‘Central Dogma’, which depicted the flow of genetic information as proceeding from DNA to RNA to protein.

Using the primary definition of a gene as a DNA unit that encodes information for producing protein, it was therefore surprising, upon completion of the sequencing of the human genome, to find that the total number of predicted, protein-coding sequences was far fewer than anticipated (currently corresponding to ~23,000 ‘genes’, similar to the number found for the mouse and fly genomes), and that these sequences only accounted for 1–2% of the genomic DNA [4, 5]. However, reassessment of the non-coding ‘intergenic’ regions (originally considered ‘junk’ DNA) began to emerge as a result of several studies showing that in fact >90% of the entire genome is transcribed into primary nuclear transcripts, many of which would not be predicted to encode protein [6–8]. In addition, the 400 annotated genes analyzed by ENCODE were shown to be capable of producing, on an average, 5.4 transcript variants per locus, only 1.7 of which were predicted to encode protein. While the functional significance of many of these non-coding transcripts remains to be determined, a plethora of recent work has firmly established a central role for microRNAs and other classes of non-coding RNAs in fundamental biological phenomena, such as the cell cycle, and cell fate and differentiation decisions, through their ability to modulate global gene expression programs [9]. Thus, the apparent blurring of the boundaries between previously designated genic- and intergenic regions, and the preponderance of potentially functionally important non-coding transcripts have forced a reconsideration of what constitutes a ‘gene’ [10].

In addition to an expanded view of the definition of a gene, these and other studies have indicated that the nature and organization of the regulatory components determining transcriptional activity within mammalian genomes are far more complex than was originally appreciated. For example, transcription of a given locus can involve alternate promoter usage, multiple transcription start sites and the production of antisense- or overlapping transcripts [6]. For the vast majority of cases, it remains to be determined how the production of these variants is regulated, and whether the overlapping or antisense products play a modulatory role. It has also become apparent that much of cell type-specific gene transcription may rely on the activity of distal enhancer elements that are embedded within non-coding ‘intergenic’ regions, and that overall transcriptional output depends on the interplay among a number of additional elements such as boundary elements and insulators [11, 12]. While high throughput approaches have made dramatic progress in determining the genomic locations of these elements, and the ubiquitous- or cell type-specific transcription factors (TFs) that bind them, most of these remain as of yet uncharacterized.

Since the ability of cells to dynamically and selectively activate (or repress) specific gene subsets is a fundamental determinant of cellular phenotype and function, development and responses to environmental cues, insights into the underlying transcriptional regulatory mechanisms are essential to our understanding of these processes. One approach that is central to this effort is the construction of gene regulatory networks (GRNs) [13]. GRNs are tools for modeling, on a global scale, all of the molecular components and interactions comprising different regulatory circuits for a range of biological processes, cell types and species, and aim to reveal the regulatory mechanisms that are employed to achieve
precise temporal- or spatial-specific gene expression patterns determining unique cellular phenotypes or responses to signaling molecules. To this end, GRNs create a conceptual framework in which all of the various functional molecular entities, including the genes, DNA regulatory elements (e.g. promoters, enhancers, silencers, insulators), TFs, cofactors and chromatin-modifying enzymes, signal transduction components, microRNAs and other non-coding RNAs, that are important to the generation of gene expression patterns of a given cell type or process are identified and assembled into hierarchical subsets according to their physical or functional interactions and effects on transcriptional output. The assembly of GRNs therefore calls for the development and implementation of a wide spectrum of novel and complementary technologies and will be best served by the analyses of divergent model organisms and cell types.

In this special issue of Briefings in Functional Genomics, we highlight a number of different approaches that are being brought to bear for the construction of gene/transcription regulatory networks and examples of some of the insights that GRNs can provide. Firstly, the interaction of TFs with regulatory DNAs is central to understanding GRNs. Although ChIP-based technologies are widely used for this purpose, not all relevant TFs have necessarily been identified or are amenable to this analysis. Marion Walhout discusses a complementary, gene-centered functional approach to identify TF-target DNA interactions and model GRNs in Caenorhabditis elegans. Pinay and Andrews describe an alternate, functional genomics approach, Reporter Synthetic Genetic Array analysis, which allows the quantitative assessment of a range of genetic perturbations on reporter gene activity and the discovery of components of transcriptional regulatory pathways. Savas and Tanese instead review advances in identification of the protein and nucleic acid components of machineries regulating post-transcriptional processes and the challenges of identifying the nucleic acid targets of these complexes. It is also important to consider that the activity of the proteins mediating all of these events can be modulated by signal transduction cascades and post-translational modifications, and Lin and colleagues discuss approaches for the global identification of phosphoproteins and the construction of phosphorylation networks. Puig and Mattila present an overview of the use of RNAi screens for the system-wide identification of GRN components, and Ni and Lee demonstrate that such RNAi screens can be implemented to identify molecular pathways that are related to complex biological processes such as aging. In an additional application of the GRN approach to understand complex biological phenomena, Nowick and Stubbs discuss how the expansion and diversification of TF-encoding genes may create functionally altered, species-specific variants whose incorporation into existing GRNs may modify the GRN and drive evolution and species differences. Finally, the construction of a complete GRN requires that the vast number of molecules and interactions identified in these types of studies be integrated and organized such that the network architecture and dynamics can be easily visualized. Fisher and Piterman discuss efforts to create computer programs that can achieve this and may also have the potential to predict new interactions or outcomes within the biological system. Although it is impossible to cover every aspect of GRNs, this special issue aims to provide a representative cross-section of the varied and exciting approaches and perspectives of this dynamic field.

Lisa Dailey

References


