Systems biology in the cell nucleus

Stanislaw Gorski and Tom Misteli

National Cancer Institute, NIH, 41 Library Drive, Bethesda, MD 20892, USA

Authors for correspondence (e-mail: gorskis@mail.nih.gov; mistelit@mail.nih.gov)

Accepted 20 July 2005

Journal of Cell Science 118, 4083-4092 Published by The Company of Biologists 2005

doi:10.1242/jcs.02596

Summary

The mammalian nucleus is arguably the most complex cellular organelle. It houses the vast majority of an organism's genetic material and is the site of all major genome regulatory processes. Reductionist approaches have been spectacularly successful at dissecting at the molecular level many of the key processes that occur within the nucleus, particularly gene expression. At the same time, the limitations of analyzing single nuclear processes in spatial and temporal isolation and the validity of generalizing observations of single gene loci are becoming evident. The next level of understanding of genome function is to integrate our knowledge of their sequences and the molecular mechanisms involved in nuclear processes with our insights into the spatial and temporal organization of the nucleus and to elucidate the interplay between protein and gene networks in regulatory circuits. To do so, catalogues of genomes and proteomes as well as a precise understanding of the behavior of molecules in living cells are required. Converging technological developments in genomics, proteomics, dynamics and computation are now leading towards such an integrated biological understanding of genome biology and nuclear function.

Introduction

The sequencing of whole genomes represents a landmark in modern biology (Adams et al., 2000; Goffeau et al., 1996; Lander et al., 2001; Venter et al., 2001; Waterston et al., 2002), revolutionizing the way genes are found, classified and analyzed. It has also brought about a shift in how biological problems are approached. It has encouraged us to move beyond the limitations of understanding single processes and inspired us to ask how biological events occur at the systems level; rather than analyzing the behavior of a single signaling kinase, for example, we now seek to understand how all elements in a signaling pathway interact; rather than asking how a gene responds to an extracellular cue, we now want to know how the genome as a whole responds. These types of global approach represent the first steps towards understanding biological functions as they really occur – in the context of biological systems.

While the sequencing of genomes has influenced virtually all fields of biology, it has had the most profound impact on the study of gene expression itself. This is particularly true because the availability of genome sequence information has coincided with the development of microarray analysis, which allows us to interrogate gene expression at a system level (Schena et al., 1995). Although most genome-wide analysis methods are largely descriptive and generally only provide lists of what parts of the genome undergo changes in activity, we can now routinely mine genome-wide expression data, using computational tools to predict biological pathways involved in a particular physiological response or to group samples, for example, normal and malignant tissues, according to expression patterns (Slonim, 2002). Despite these successes, a key realization has been that even these comprehensive approaches cannot answer some of the most fundamental questions in genome biology, such as why more advanced organisms do not necessarily use more genes to achieve complexity; how gene expression programs are defined in a tissue- and cell-type-specific manner; and how the spatial and temporal organization of transcription, RNA processing, RNA degradation, export, DNA replication or DNA repair affects genome function. It seems clear now that, apart from genome sequence, additional aspects of genome biology must be considered if we are to understand how genomes actually work.

A key emerging contributor to genome function and regulation is the spatial and temporal arrangement of the genome and gene expression processes in nuclear space (Misteli, 2001; Spector, 2003). Dramatic developments in high-resolution and live-cell imaging have revealed that the cell nucleus is a highly heterogeneous and complex organelle, and that global genome architecture changes during processes such as differentiation and development (Misteli, 2001; Spector, 2003). One unique feature of the mammalian cell nucleus is the presence of structural and functional domains that lack membrane boundaries (Lamond and Earnshaw, 1998; Matera, 1999). Spatial organization of the genome is achieved by a non-random arrangement of chromosomes in the interphase nucleus, with chromosomes occupying preferential intranuclear positions (Cremer and Cremer, 2001; Parada et al., 2004). Chromosomes, genome domains and gene loci may congress in space to form functional chromatin neighborhoods, such as transcriptionally silent heterochromatin regions or clusters of active genes. In addition to the non-random spatial arrangement of genomes, the nucleus also contains numerous proteinaceous domains, such as the nucleolus and splicing factor compartments, which represent distinct structural, and probably functional, subcompartments (Misteli, 2005). Although the full contribution of the spatial organization of the genome and nuclear proteins is still unknown, it seems clear that we must understand genome function in the context of this

architectural organization to answer some of the key questions in genome biology.

The ultimate goal of a systems biology view of the cell nucleus is to understand genome function within the architectural framework of the nucleus. Gaining such a systems view of nuclear function will involve several steps. First, we must generate proteomic and genomic inventories of nuclear components, including genome sequence elements, epigenetic modifications, higher-order chromatin structure, chromatin-binding complexes and nuclear compartments. Second, we must understand the cell biological properties of nuclear processes in living cells in terms of characteristics such as their spatial organization and dynamic properties. Third, we must integrate experimental data on process dynamics and spatial organization, using computational approaches (Fig. 1). Several technological developments now make this a realistic, although still highly challenging, goal. On the one hand, proteomic and genomic approaches can provide complete lists of components present in particular nuclear structures, and identify what transcription factors are bound where throughout the genome, and what coding and non-coding genome regions are active at any given time. On the other hand, quantitative in vivo microscopy methods provide the first glimpse of how DNA, RNA and proteins behave inside the nuclei of living cells. Combining these methods and mining the data by emerging computational approaches will eventually lead to a realistic picture of gene expression. We review here how far we have progressed on the long road towards achieving a systems biology view of the cell nucleus and of genome function.

Nuclear inventories

Proteomics in the cell nucleus

The mammalian cell nucleus contains an estimated 20% of all cellular proteins (Simpson et al., 2000). Although many exist freely in the nucleoplasm, numerous proteins are enriched in particular nuclear compartments. A crucial first step towards an integrated view of the nucleus and its compartments is the generation of a protein inventory for each domain (Fig. 2). Proteomic approaches involving biochemical purification and mass spectrometry are now beginning to yield this information. The first nuclear structure to be analyzed in this manner was the interchromatin granule cluster (IGC), which constitutes the major fraction of the splicing factor compartments (Mintz et al., 1999; Saitoh et al., 2004). Previously, these structures were thought to contain mostly proteins involved in pre-mRNA splicing, but the detection of >350 proteins, many of which have no apparent function in the splicing reaction, in purified IGCs suggests they have additional functions (Mintz et al., 1999; Saitoh et al., 2004). Proteomic analysis has now revealed a similarly astonishing number of proteins associated with the nucleolus. The latest count for mammalian nucleolar proteins is ~600 proteins, and the plant nucleolus appears to have at least 200 proteins (Andersen et al., 2005; Pendle et al., 2005; Scherl et al., 2002). These remarkably high numbers of constituents, and the fact that many are novel proteins, provide a clear indication of how little we know about the functions of nuclear domains as a whole.

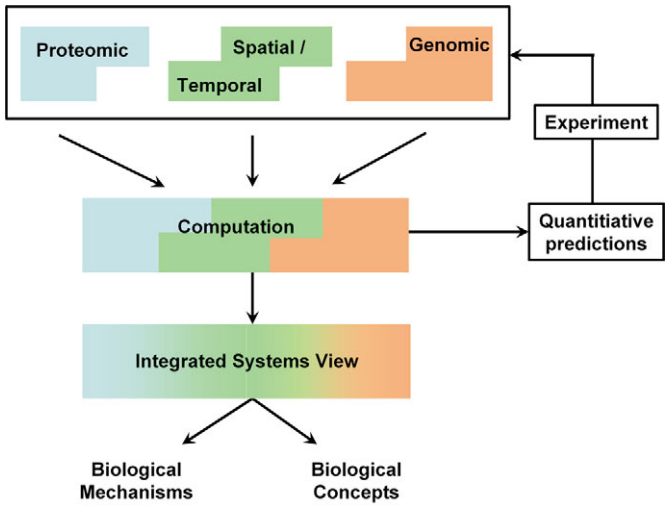

Fig. 1. Requirements and goals of a systems approach in the nucleus. Systems approaches require the combination of experimental genomic and proteomic information with principles of spatial and temporal organization. Computational methods allow hypothesis to be tested through quantitative predictions that can be verified experimentally. Integration of such data leads to a comprehensive view of nuclear processes within the context of nuclear architecture. The results of these interdisciplinary approaches are novel insights into molecular mechanisms and discovery of novel concepts.

Two remarkable commonalities between the proteomic analyses of IGCs and the nucleolus are noteworthy. First, neither revealed the presence of conserved targeting signals in resident proteins, which suggests that the enrichment of protein subsets is largely determined by the functions they have in a particular compartment rather than by specific targeting mechanisms (Dundr and Misteli, 2002). Second, proteins that one would not expect to be present in these compartments are abundant (Andersen et al., 2005; Saitoh et al., 2004). For example, histones are associated with IGCs although these compartments are insensitive to DNAse and are generally assumed to be chromatin free (Saitoh et al., 2004). Conversely, numerous pre-mRNA splicing factors can be detected in the nucleolus, despite the fact that the ribosomal genes found within the nucleolus do not produce spliceable pre-mRNAs (Andersen et al., 2005). Although contaminating proteins may contribute to some of these surprising findings, it seems more probable that the unexpected appearance of proteins in certain compartments reflects their ability to diffuse relatively freely throughout the nuclear space and stochastically associate with nuclear compartments non-specifically (Dundr and Misteli, 2002).

The behavior of single proteins in response to changes in cellular conditions is traditionally followed by in vivo microscopy. Proteomic approaches now allow us to track changes in the composition of entire subnuclear organelles. Proteins can be tagged with multiple metabolic labels, which allows us to trace the compositions of entire organelles over time (Andersen et al., 2005). When applied to the nucleolus, this approach demonstrates that, upon inhibition of transcription, nucleolar components can behave very differently: some accumulate in the morphologically altered nucleolus, whereas others dissociate (Andersen et al., 2005). Remarkably, functionally related proteins or proteins known to be present in a complex, such as pre-ribosomes, often exhibit similar kinetics, which suggests that this method can also be used to probe the dynamics of protein complexes and

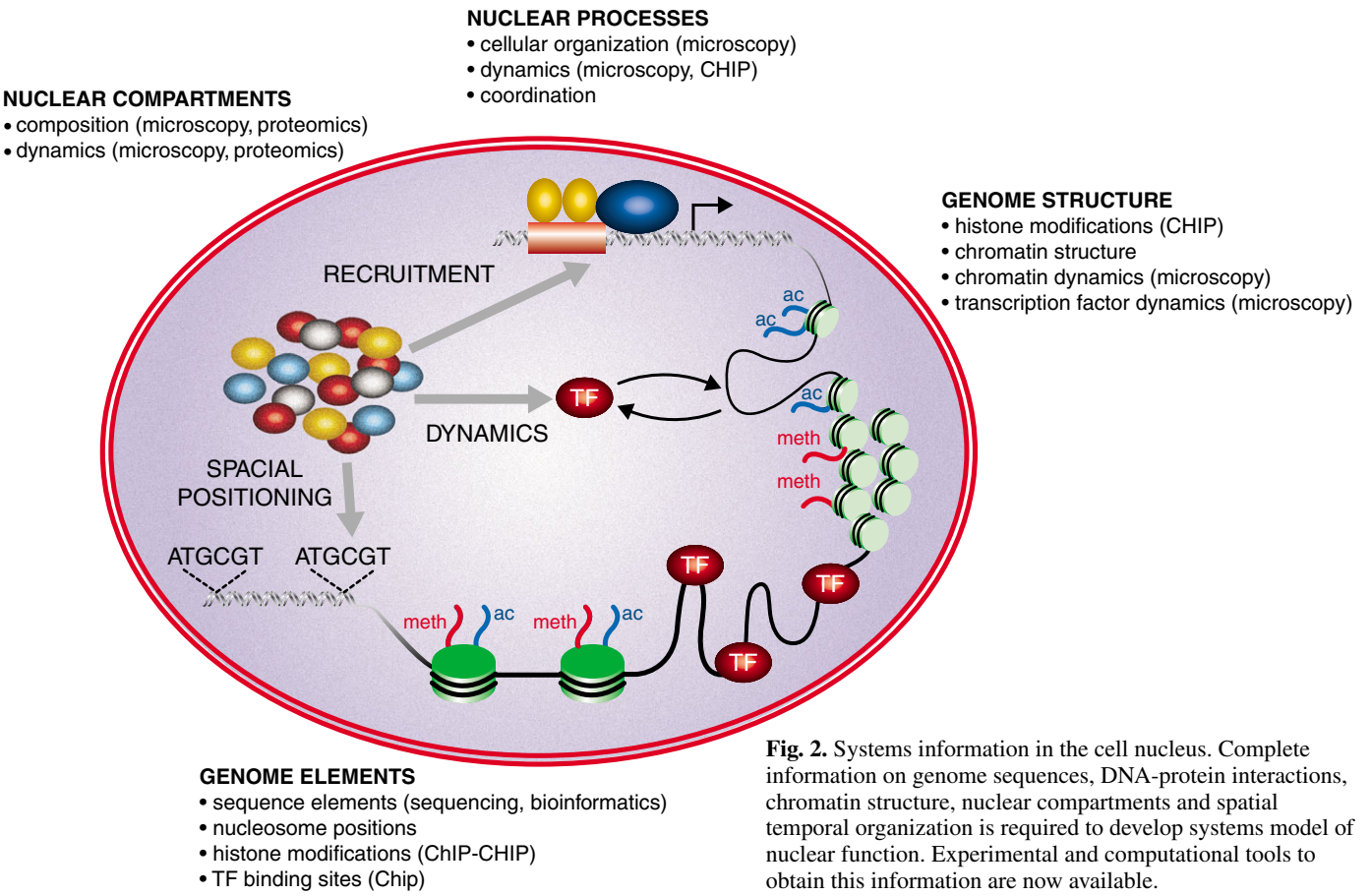

Fig. 2. Systems information in the cell nucleus. Complete information on genome sequences, DNA-protein interactions, chromatin structure, nuclear compartments and spatial temporal organization is required to develop systems model of nuclear function. Experimental and computational tools to obtain this information are now available.

functionally related groups of proteins. These observations highlight the dynamic nature of the nucleolus, and probably of nuclear compartments in general. More importantly, the ability to follow the fate of all components of a nuclear domain under changing physiological conditions is a powerful tool that will help us to understand how nuclear architecture responds at the molecular level as a system to signaling cues.

The components of various other nuclear structures have recently also been cataloged. Several hundred proteins that purify with the nuclear envelope have been identified and the composition of the nuclear pore complex (NPC) has been similarly characterized (Rout et al., 2000; Schirmer et al., 2003). Since nuclear lamins and associated nuclear envelope proteins have been implicated in various human diseases, this former effort has provided leads to numerous novel putative disease genes and mechanisms (Schirmer et al., 2003). The proteomic analysis of the NPC has allowed the spatial mapping of all NPC components within the three-dimensional structure of the pore and, on the basis of homology analysis, functional roles for the various NPC components have been proposed, which can now be experimentally tested (Rout et al., 2003).

Proteomic approaches are not limited to the analysis of architectural features of the nucleus, but can similarly be applied to nuclear protein complexes. Indeed, the protein compositions of pre-ribosomes and spliceosomes have been determined (Bassler et al., 2001; Harnpicharnchai et al., 2001; Rappsilber et al., 2002; Zhou et al., 2002). Analysis of such complexes is facilitated by their particle-like nature, which makes them more amenable to purification than nuclear compartments. Use of versatile protein-tagging systems now also allows identification of complexes that are generally more difficult to purify. A powerful demonstration of such an approach is the recent purification and proteomic analysis of protein complexes that bind to and define the boundary between silenced chromatin regions at *Saccharomyces cerevisiae* telomeres and the neighboring subtelomeric active regions (Tackett et al., 2005). In this approach, tagged proteins were crosslinked to their binding sites and mapped throughout the genome. Intriguingly, complexes with compositions similar to the one at the telomeric boundary were found at various locations throughout the genome, allowing the possibility that these sites have similarly, yet unappreciated, boundary functions. The ability to analyze not only soluble complexes and nuclear compartments but also chromatin-bound large macromolecular assemblies, will be crucial in global mapping of protein complexes along the genome.

Genomics in the cell nucleus

Beyond coding regions

The most basic component of a systems view of the nucleus is complete genome expression profiles under various physiological conditions (Fig. 2). Although microarray technology is now a standard tool, most such analyses have been limited to the <10% of total genome information that represents coding regions. The recent development of high-

density tiling arrays that also contain non-coding regions has started to reveal the true complexity of the metazoan transcriptome (Bernstein et al., 2005; Cawley et al., 2004; Cheng et al., 2005; Kampa et al., 2004). For example, an analysis of ten human chromosomes at 5 bp resolution identified as much as 32% of cellular transcripts to come from intergenic, non-coding regions, and >60% of the transcripts that hybridize with unannotated sequences show transcription from both genomic strands (Cheng et al., 2005). Systematic analysis of the function of these non-coding RNAs is particularly important in light of the emerging roles of mirco-RNAs in key physiological processes. These regulatory molecules represent nearly 1% of all genes in higher eukaryotes (Bartel, 2004), and have been implicated in developmental timing, neuronal patterning (Lee et al., 1993; Wightman et al., 1993), apoptosis (Brennecke et al., 2003) and cellular proliferation (Hatfield et al., 2005), and can be used in cancer profiling (Lu et al., 2005). These observations make it clear that, just as we are beginning to be able to explore gene expression globally, we must realize that a much larger percentage of the genome is being transcribed than is presently appreciated. Determination of the functional impact of transcription from these non-coding regions will be crucial to our understanding of genome function.

Transcription factor binding

A further key piece of information is the location of binding sites for the multitude of transcription factors bound along the genome (Fig. 2). Whereas in-silico bioinformatics has yielded some of this information by mapping consensus sites throughout the genome, approaches involving chromatin immunoprecipitation (ChIP) coupled with microarray analysis, (the so called ChIP-on-Chip approach), can reveal the actual binding sites within the genome for a particular protein under a given condition. In ChIP-on-Chip, proteins are crosslinked to DNA in vivo, they are then immunoprecipitated, the bound DNA amplified, and all recovered fragments are identified on microarrays and mapped onto the genome sequence. Such methods are beginning to reveal several surprising mechanistic principles of how proteins interact with the genome.

Bioinformatic analyses generally assume that the presence of a particular target sequence for a transcription factor means that it is used; genome-wide analysis by ChIP-on-Chip questions this view. In one study, mapping of binding sites for the transcriptional activator Gal4 showed that not all consensus binding sequences are occupied when *S. cerevisiae* cells are grown in the presence of galactose. This indicates that sequence itself is not sufficient for a protein to bind (Ren et al., 2000). It is not entirely clear what determines whether a specific binding site is occupied, but it seems probable that the combinatorial binding of several proteins as well as the status of higher-order chromatin structure contributes significantly.

A concept that has emerged from such analyses is how a single factor can regulate the expression of different sets of genes. This behavior is exemplified by the transcription factor Ste12, in *S. cerevisiae*. Upon the addition of pheromone, Ste12p activates genes involved in mating, whereas during starvation the same factor activates genes in the filamentous growth pathway. Genome-wide location analysis showed that binding of Ste12p to genes that regulate filamentous growth requires the transcription factor binding partner Tec1, but the binding of Ste12p to mating genes does not. Tec1 is negatively regulated by a pheromone-activated MAP kinase, which results in selective binding of Ste12p to mating genes because activation of filamentous genes by the Ste12p-Tec1 complex is inhibited (Zeitlinger et al., 2003). Therefore, regulation of gene expression by common factors may occur through modulation of the activities of their binding partners rather than solely through their interaction with the target sequence.

Genomic approaches have also proven important in identifying the targets of transcription factors that coordinate central cellular processes such as proliferation, growth, differentiation and apoptosis (Cawley et al., 2004; Fernandez et al., 2003; Levens, 2003; Orian et al., 2003). For example, the proto-oncogene Myc has been shown to bind to and modulate the levels of up to 10% of human genes (Fernandez et al., 2003; Orian et al., 2003). A further surprise was the location of the binding sites for important transcription factors such as Myc, Sp1 and p53 in relation to coding regions (Cawley et al., 2004). Analysis using tiling arrays across chromosomes 21 and 22 showed that these proteins, and presumably others, occupy many more sites than previously thought. Moreover, <25% of these binding sites are at the 5′ end of known genes, but >36% of binding sites are located in or proximal to the 3′ regions. Interestingly, many of these sites coincide with regions containing non-coding RNAs, which suggests the existence of an as-yet-unappreciated level of genome regulation by transcription factors through control of non-coding RNA expression (Cawley et al., 2004).

Chromatin organization and modifications

One way the usage of a transcription-factor-binding site can be controlled is by packaging of DNA within chromatin. Indeed, transcription-factor-binding sites within promoters have decreased nucleosome occupancy in yeast (Dion et al., 2005; Lee et al., 2004; Yuan et al., 2005). Interestingly, an inverse correlation between gene activity and nucleosome occupation is observed when global gene expression is modulated by heat shock (Lee et al., 2004). In addition to local DNA accessibility, higher-order chromatin structure also affects genome function. Consistent with this view, analysis of a small number of genes more than two decades ago suggested that genes that are actively transcribed exist in an open conformation, whereas inactive genes are maintained in a condensed chromatin structure (Kimura et al., 1983). Recent genome-wide mapping of open and closed chromatin regions, and their correlation with gene activity, paints a more complicated picture. Using fluorescence in situ hybridization and DNA microarray analysis with probes for either open or condensed chromatin regions, a direct correlation was identified between chromatin compaction and gene density, but not necessarily gene activity (Gilbert et al., 2004). Gene-rich chromosomes, such as human chromosome 22, contain large regions of chromatin in an open conformation compared with gene-poor regions within chromosome 1. Despite this structural heterogeneity, a good correlation with gene transcription was not observed because many silenced genes are also located in these open chromatin regions (Gilbert et al., 2004). These results suggest that the role of chromatin in the regulation of gene expression occurs on a more local basis. Such a notion is consistent with results from

genome-wide binding studies demonstrating that chromatin-remodeling and -modifying complexes that can change the local structure of chromatin and expose binding sites are recruited to promoters of many genes upon activation (Lusser and Kadonaga, 2003; Robert et al., 2004).

The functional status of chromatin often broadly correlates with post-translational modifications of the core histones and these modifications have been proposed to constitute an epigenetic code (Strahl and Allis, 2000). Genomic analyses can identify the patterns of these modifications and link them to features of the genome such as regulatory sequences and to gene activity (Fig. 2) (Bernstein et al., 2005; Dion et al., 2005; Kurdistani et al., 2004; Roh et al., 2005; Schubeler et al., 2004). For example, in human T cells the distribution and relative quantities of acetylated histones H3 and H4 have been mapped across the genome (Roh et al., 2005). This method was able to identify islands of acetylation that preferentially occur at known promoters and regulatory sequences rather than in the main body of genes.

Similar studies are beginning to reveal patterns of different modifications in relation to gene activity and organization (Bernstein et al., 2005; Schubeler et al., 2004). In yeast, for example, histone acetylation has a combinatorial effect (Kurdistani et al., 2004). When the distribution of the acetylation status of eleven different lysines in the same histone octamer region was determined, distinct combinations of acetylated and deacetylated lysines were observed at specific genomic locations. Intriguingly, gene regions that have similar modification patterns are co-regulated, which supports the notion of a combinatorial code (Kurdistani et al., 2004). However, the specific roles these acetylation patterns play in recruiting histone-associated proteins are still not clear. It has been a long standing view that acetylation of histones could play a non-specific role in maintenance of chromatin structure and dynamics (Horn and Peterson, 2002). Indeed, an analysis of combinations of four different modifications in the same octamer unit, demonstrated that three of the residues do not play a specific role in transcription but pointed towards a general cumulative effect, potentially on chromatin structure (Dion et al., 2005). Despite these differences, the power of a system-wide analysis of histone modifications should lead to elucidation of the functional link between these modifications and gene activity.

Systematic mapping of DNA replication sites

The structural state of chromatin has also been implicated in nuclear functions other than gene expression. There is a long-recognized link between replication timing and chromatin compaction: frequently euchromatic regions replicate early whereas heterochromatic regions replicate later during S phase (Gilbert, 2002). Genome-wide analysis of sequence-dependent replication timing in higher organisms such as *Drosophila* and humans has begun to reveal some relationships between genome structure and activity, and added more levels of complexity (Jeon et al., 2005; Schubeler et al., 2002; White et al., 2004; Woodfine et al., 2004). As previously appreciated, for many early-replicating genes replication timing correlates with their transcriptional activity, and most repetitive heterochromatic regions replicate late during S phase. However, other proximal heterochromatic regions replicate early, and some portions of euchromatic chromosome regions also replicate late in S phase (Schubeler et al., 2002). Intriguingly, tiling arrays have shown that highly transcribed sequences outside annotated genes also replicate early during S phase (White et al., 2004) and that there is a relationship between gene density and replication timing, most dense regions replicating in the first 4 hours of S phase (Jeon et al., 2005). These first glimpses into the global relationship between replication timing, gene activity, gene organization and chromatin status make it clear that these are complex links that are difficult to understand through analysis of single loci, and that the regulatory mechanisms and the precise nature of the local chromatin structure at these anomalous regions remain to be discovered.

Dynamics

Genome-wide mapping of sites at which proteins interact with chromatin provides a detailed picture of the steady-state distribution of proteins within the genome at a specific time. However, temporal aspects of their behavior add an additional level of complexity that must be taken into account (Fig. 2). Time-lapse microscopy experiments coupled with photobleaching protocols have demonstrated that most chromatin proteins can move freely and rapidly through the nuclear space by passive diffusion and, more importantly, that most protein-chromatin interactions are transient, including those involving structural components of chromatin (Agresti et al., 2005; Cheutin et al., 2003; Festenstein et al., 2003; Kimura and Cook, 2001; Lever et al., 2000; McNally et al., 2000; Misteli, 2001; Misteli et al., 2000; Phair et al., 2004b). The rapid exchange kinetics – typical residence times of proteins on chromatin are of the order of seconds – suggests that chromatin is highly dynamic (Bustin et al., 2005). This property is probably essential for the genome to respond rapidly to changing environmental conditions. It has implications for how proteins interact with the genome globally as well as locally and thus for the system-level behavior of the nucleus and the genome.

The ability of proteins to move rapidly throughout the nucleus ensures their availability throughout the genome (Misteli, 2001). In addition, the short residence times on chromatin allow proteins to search the genome for specific binding sites by three-dimensional scanning as a molecule briefly binds to chromatin, dissociates and rapidly binds to another site (Misteli, 2001). Upon a chance encounter with one of its specific sites, the molecule resides for a longer period of time and exerts its full biological activity. Some molecules may also function at non-specific binding sites. For example, chromatin-remodeling factors may continuously remodel chromatin stochastically, regardless of the nature of the binding site, and in that way contribute significantly to the overall dynamic nature of chromatin.

At the local level, dynamic binding significantly affects the behavior of proteins because it provides the basis for regulatory events through combinatorial, competitive networks of interactions. The transient binding of a given protein means that each time a molecule dissociates from its binding site, the vacant site can be occupied by another protein. Since this might have a distinct function, the fate of the region depends on the competitive binding of these factors. An example of this is the

binding of the linker histone H1 and the high mobility group (HMG) proteins (Bustin et al., 2005; Catez et al., 2002). H1 is generally considered a transcriptional repressor, whereas HMG proteins generally act as activators. The binding sites for several HMG proteins overlap with that of H1, and experimental titration of HMG into living cells results in increased occupancy of DNA by HMG proteins through competition for the same binding site (Catez et al., 2004). Similarly, the relevance of dynamic interaction networks has become apparent from a systematic analysis of the assembly dynamics of the RNA polymerase I transcription machinery with endogenous ribosomal genes (Dundr et al., 2002). GFP fusion proteins that contain pre-initiation, assembly factors and different polymerase I subunits have distinct binding kinetics, which suggests that their interactions at the promoter occur individually as part of a dynamic interaction network. It will be important to determine whether changes in the interaction dynamics are responsible for changes in the transcriptional output of the polymerase.

Computation

Analysis of complex biological processes in the cell nucleus requires the use of computational tools. In addition to the well-established methods for mining gene expression data to identify gene networks and pathways, simulation is becoming indispensable for the analysis of the kinetics of nuclear processes (Phair et al., 2004a; Slepchenko et al., 2002; Sprague and McNally, 2005). In these approaches, experimental data are used to determine numerical values for kinetic parameters and these parameters then serve as constraints in computational models that describe a particular process (Phair et al., 2004a; Slepchenko et al., 2002; Sprague and McNally, 2005). Such models, which often consist of systems of differential equations, can then be used to make quantitative predictions, which in turn can be tested experimentally. Not only does this iterative process provide quantitative information and predictions, but the failure of a given model to account for experimentally observed properties often points to novel conceptual aspects of a process. These types of computational approach are now being applied to the analysis of nuclear processes.

The most advanced system-level analysis of any nuclear function is probably the in-silico analysis of nuclear protein import (Gorlich et al., 2003; Riddick and Macara, 2005; Smith et al., 2002). Here, the authors microinjected cells with fluorescently labeled cargo and measured nuclear import rates under various experimental conditions, as well as using in-vitro analysis to determine several crucial parameters such as on- and off-rates for various import intermediates. Developing computational models of the import pathway, they were able to examine the contribution of the import-driving Ran gradient as well as that of importin (Gorlich et al., 2003; Riddick and Macara, 2005; Smith et al., 2002). This system demonstrated the value of modeling approaches because it generates several counterintuitive predictions, including a limiting role of importin α and an inhibitory role of excess importin β – both of which were confirmed when tested experimentally (Riddick and Macara, 2005). The model for nuclear import has thus not only provided the first quantitative assessment of many of the import steps but also given novel mechanistic insights and made quantitative predictions. Furthermore, it illustrates that the behavior of complex processes such as nuclear import is difficult to predict intuitively and that quantitative models are required to do so.

Cao and Parker have similarly combined quantitative experimental data and in-silico modeling to analyze normal mRNA-turnover and nonsense-mediated decay in *S. cerevisiae* (Cao and Parker, 2001; Cao and Parker, 2003). Apart from generating estimates of quantitative values for key steps in the degradation process such as de-adenylation and de-capping, they also used the model to test the hypothesis that preferential degradation of RNAs containing nonsense codons near the 5′ end is due to leaky mRNA surveillance in which the recognition of nonsense codons near the 5′ end of the mRNA is more efficient than recognition of a 3′ nonsense codon (Ishigaki et al., 2001). However, their model reveals that the leaky surveillance cannot account for the observed degradation kinetics (Cao and Parker, 2003). Rather, the experimental data can only be explained by a mechanism in which all nonsense-codon-containing mRNAs are recognized equally well but their decapping and subsequent degradation rates differ. This quantitative analysis of NMD has now given rise to a new, testable model of NMD referred to as position-independent efficient surveillance (Cao and Parker, 2003).

This type of computational analysis has also been applied to the assembly and elongation behavior of a transcription complex (Dundr et al., 2002). Dundr et al. fluorescently tagged the majority of the subunits of the RNA polymerase I complex and analysed their association and elongation kinetics on endogenous ribosomal genes, using photobleaching microscopy. This allowed them to estimate the in vivo RNA polymerase I elongation rate and show that the subunit behavior is consistent with a model in which the polymerase assembles in a stepwise fashion from its subunits via formation of metastable intermediates (Dundr et al., 2002).

Many nuclear processes are tightly interconnected (Maniatis and Reed, 2002; Neugebauer, 2002). For example, pre-mRNA processing steps, including poly-adenylation and splicing, occur co-transcriptionally, and the RNA-processing machinery is physically associated with elongating transcription complexes. Similarly, the trigger for RNA export is intimately linked to completion of pre-mRNA splicing (Tange et al., 2004), and even the choice between alternative splice sites is influenced by transcriptional co-activators and promoter usage (Cramer et al., 1997). The next step in computational modeling, therefore, will be to interrogate combinations of multiple processes. This poses an experimental challenge because, ideally, information about multiple steps in the gene expression process should be extracted from a single experimental system. This is rarely possible, particularly in the context of nuclear architecture. However, Janicki et al. have developed a potentially powerful tool that holds much promise for analyzing sequential steps from transcription to protein synthesis (Janicki et al., 2004). They have generated an artificial gene array that allows visualization of changes in chromatin structure, binding of transcription factors, mRNA synthesis, RNA export and protein synthesis. This system has already been used to quantitatively track the movement of single mRNA-protein complexes (mRNPs) in living cells, demonstrating that mRNAs move rapidly by diffusion through the nucleus (Shav-Tal et al., 2004). Notwithstanding the caveat

of an artificial experimental system, the application of photobleaching methods in combination with computational analysis should provide a comprehensive quantitative picture of the coordination of gene expression steps.

Towards an integrated view of nuclear function

The ultimate goal of systems biology in the nucleus is a computational model that is able to describe all nuclear processes from genome replication, via maintenance of DNA integrity, to regulation of expression within the spatial and temporal framework of nuclear architecture. If this were not ambitious enough, the benchmark for such a model will be its ability to generate quantitative predictions in silico that can be tested experimentally. Achieving this goal will not be possible in a single step but will involve the development of gradually more complex, and presumably more accurate, computational models by iterative comparison of experimental and simulation data. Two complementary strategies should eventually achieve this goal (Fig. 3).

Genome and proteome information can be directly assembled into pathways and networks (Fig. 3). In these approaches, little attention is paid to cellular organization or the details of molecular interactions, but only functional relationships between pathway components are considered. A good example of this approach is the analysis of DNA-damage responses using microarray methods (Begley and Samson, 2004) Analysis of global changes in gene expression shows that expression patterns of genes involved in a wide variety of processes, such as protein degradation and synthesis, signal transduction, RNA metabolism and chromatin remodeling as well as transcription and DNA repair, are altered upon treatment with DNA damage reagents (Begley and Samson, 2004). It has also become clear that specific repair pathways share common members and even functions. An example of this is the base-excision-repair pathway, which is linked to the nucleotide excision, mismatch and recombinatorial repair pathways through various molecular interactions (Begley and Samson, 2004).

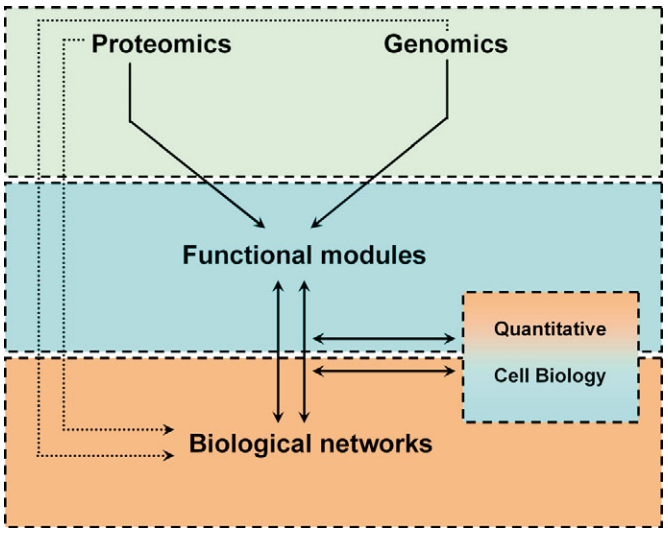

Fig. 3. Strategies for the development of comprehensive system models. Proteomic or genomic information can be directly used to develop pathway and network models (dashed lines). Alternatively, specific cellular processes such as transcription or replication can be modeled in the form of functional modules, using quantitative cell-biological methods including combined imaging and computational approaches. Linking models of functional modules yields an integrated model.

An extension of the traditional transcription profiling approach is the combination of genome-wide transcription-factor-binding information with bioinformatics to mine sequence elements in yeast (Harbison et al., 2004). This approach generates a molecular network of all potential transcription-factor-binding sites and motifs under various environmental conditions. Integrating the data on different transcription-factor-binding sites produces a 'transcriptional-potential-network' that provides a framework for modeling the mechanisms that lead to global gene expression (Harbison et al., 2004). Analyzing data in this manner is important because it allows us to identify genes that share similar combinations of transcription-factor-binding sites and regulatory motifs, and draw general conclusions concerning organization of gene sequences.

Although this type of network modeling provides insights into the overall properties of physiological processes, it is relatively low-resolution and generally ignores most aspects of how biological processes are organized in cells. To obtain a more realistic, complete system-level view we must model single processes in the greatest possible detail, taking into account the networks of molecular interactions, their dynamics and spatial organization (Fig. 3). Processes that have been explored in this way include transcription by RNA polymerase I machinery, nuclear import and NF-κB signaling (Dundr et al., 2002; Hoffmann et al., 2002; Nelson et al., 2004; Riddick and Macara, 2005; Smith et al., 2002). Each can be considered a functional module in a larger network, and ultimately the computational models for multiple such functional modules will be combined to give a comprehensive global view of nuclear function.

This type of bottom-up approach to systems biology is more difficult owing to the incorporation of more detail, but at the same time it has several important advantages. Analysis of limited functional modules allows development of relatively simple computational models that can be tested experimentally and are easier to grasp than large networks. For example, in the case of NF-κB signaling, Hoffmann et al. analyzed the effects of knocking-out each known inhibitor of NF-κB (Hoffmann et al., 2002). Quantitation of the response to stimulation and computational modeling identified specific roles for each inhibitor, and provided a global, yet detailed, view of how this particular pathway behaves (Hoffmann et al., 2002). In addition, modeling of distinct functional modules using quantitative imaging combined with computational analysis allows the incorporation of spatial and temporal information, i.e. the role of subcellular localization, local protein concentrations or wave-like propagation of a signal. In the case of NF-κB, quantitative live-cell imaging has shown that persistent asynchronous oscillations between the nucleus and cytoplasm are required for NF-κB target gene expression. The authors suggest, that akin to Ca^{2+} signaling, the consequences of NF-κB may depend on the number, period and amplitude of oscillations (Nelson et al., 2004). Such cell biological properties are often ignored in network modeling

approaches, despite their established importance in regulation of biological processes.

Once single functional modules are satisfactorily simulated, they can be linked into larger process networks. An example of this approach is a system-level view of sea urchin development. Davidson and co-workers first analyzed the spatial and temporal regulation of genes responsible for cell specification, using a bottom-up modular approach (Yuh et al., 1998). They combined quantitative experiments and computational analysis to identify the roles of different cis-regulatory sequences controlling expression of a protein involved in the development of the midgut (Yuh et al., 1998; Yuh et al., 2001). They then verified this approach by confirming in-silico-generated predictions experimentally (Yuh et al., 1998). The subsequent addition of more modules representing different spatial domains has allowed them to assemble a large regulatory network of the genes and regulatory sequences involved in endomesoderm development (Davidson et al., 2002a; Davidson et al., 2002b). Furthermore, they can now perform large-scale system perturbations and quantitatively describe the resulting changes in multiple phenotypic parameters (Davidson et al., 2002a; Davidson et al., 2002b).

The top-down and bottom-up modular strategies described above are not mutually exclusive but highly complementary. An impressive example in which they have been combined is computational analysis of apoptosis. In this case Bentele et al. modeled the well-studied parts of the network in great detail, using experimentally determined rate constants, whereas they modeled the less well-studied parts at low resolution, only taking into account the overall behavior of particular modules (Bentele et al., 2004).

Conclusion

How close are we to a system-level view of the nucleus? Clearly we are still at the very beginning, but the tools that will make it possible to explore the nucleus on a more global scale are being put in place, and the first examples of large scale, complex systems analyses are emerging. The relatively simplistic computational models that have been applied have already demonstrated that quantitative systems analysis has the potential to provide new insights into the molecular mechanisms of nuclear process and also to generate novel concepts of how cellular processes are organized, coordinated and regulated. In the future we will probably see an increasing convergence of experimental and in-silico analysis, both of particular cellular processes as well as of process networks. These approaches are conceptually different from how biochemists and molecular biologist have traditionally thought about problems in the past and will require us to embrace multi-disciplinary approaches. While this would have been a serious hurdle only a few years ago, fortunately the sequencing of whole genomes has provided the experimental foundation, as well as the inspiration, for us to tackle cell biological problems, be it in the cell nucleus or elsewhere, at the systems level.

This research was supported by the Intramural Research Program of the NIH, National Cancer Institute, Center for Cancer Research. T.M. is a fellow of the Keith R. Porter Endowment for Cell Biology.

References

Adams, M. D., Celniker, S. E., Holt, R. A., Evans, C. A., Gocayne, J. D., Amanatides, P. G., Scherer, S. E., Li, P. W., Hoskins, R. A., Galle, R. F. et al. (2000). The genome sequence of Drosophila melanogaster. *Science* **287**, 2185-2195.

Agresti, A., Scaffidi, P., Riva, A., Caiolfa, V. R. and Bianchi, M. E. (2005). GR and HMGB1 interact only within chromatin and influence each other's residence time. *Mol. Cell* **18**, 109-121.

Andersen, J. S., Lam, Y. W., Leung, A. K., Ong, S. E., Lyon, C. E., Lamond, A. I. and Mann, M. (2005). Nucleolar proteome dynamics. *Nature* **433**, 77-83.

Bartel, D. P. (2004). MicroRNAs: genomics, biogenesis, mechanism, and function. *Cell* **116**, 281-297.

Bassler, J., Grandi, P., Gadal, O., Lessmann, T., Petfalski, E., Tollervey, D., Lechner, J. and Hurt, E. (2001). Identification of a 60S preribosomal particle that is closely linked to nuclear export. *Mol. Cell* **8**, 517-529.

Begley, T. J. and Samson, L. D. (2004). Network responses to DNA damaging agents. *DNA Repair* **3**, 1123-1132.

Bentele, M., Lavrik, I., Ulrich, M., Stosser, S., Heermann, D. W., Kalthoff, H., Krammer, P. H. and Eils, R. (2004). Mathematical modeling reveals threshold mechanism in CD95-induced apoptosis. *J. Cell Biol.* **166**, 839-851.

Bernstein, B. E., Kamal, M., Lindblad-Toh, K., Bekiranov, S., Bailey, D. K., Huebert, D. J., McMahon, S., Karlsson, E. K., Kulbokas, E. J., 3rd, Gingeras, T. R. et al. (2005). Genomic maps and comparative analysis of histone modifications in human and mouse. *Cell* **120**, 169-181.

Brennecke, J., Hipfner, D. R., Stark, A., Russell, R. B. and Cohen, S. M. (2003). bantam encodes a developmentally regulated microRNA that controls cell proliferation and regulates the proapoptotic gene hid in Drosophila. *Cell* **113**, 25-36.

Bustin, M., Catez, F. and Lim, J. H. (2005). The dynamics of histone H1 function in chromatin. *Mol. Cell* **17**, 617-620.

Cao, D. and Parker, R. (2001). Computational modeling of eukaryotic mRNA turnover. *RNA* **7**, 1192-1212.

Cao, D. and Parker, R. (2003). Computational modeling and experimental analysis of nonsense-mediated decay in yeast. *Cell* **113**, 533-545.

Catez, F., Brown, D. T., Misteli, T. and Bustin, M. (2002). Competition between histone H1 and HMGN proteins for chromatin binding sites. *EMBO Rep.* **3**, 760-766.

Catez, F., Yang, H., Tracey, K. J., Reeves, R., Misteli, T. and Bustin, M. (2004). Network of dynamic interactions between histone H1 and high-mobility-group proteins in chromatin. *Mol. Cell. Biol.* **24**, 4321-4628.

Cawley, S., Bekiranov, S., Ng, H. H., Kapranov, P., Sekinger, E. A., Kampa, D., Piccolboni, A., Sementchenko, V., Cheng, J., Williams, A. J. et al. (2004). Unbiased mapping of transcription factor binding sites along human chromosomes 21 and 22 points to widespread regulation of noncoding RNAs. *Cell* **116**, 499-509.

Cheng, J., Kapranov, P., Drenkow, J., Dike, S., Brubaker, S., Patel, S., Long, J., Stern, D., Tammana, H., Helt, G. et al. (2005). Transcriptional maps of 10 human chromosomes at 5-nucleotide resolution. *Science* **308**, 1149-1154.

Cheutin, T., McNairn, A. J., Jenuwein, T., Gilbert, D. M., Singh, P. B. and Misteli, T. (2003). Maintenance of stable heterochromatin domains by dynamic HP1 binding. *Science* **299**, 721-725.

Cramer, P., Pesce, C. G., Baralle, F. E. and Kornblihtt, A. R. (1997). Functional association between promoter structure and transcript alternative splicing. *Proc. Natl. Acad. Sci. USA* **94**, 11456-11460.

Cremer, T. and Cremer, C. (2001). Chromosome territories, nuclear architecture and gene regulation in mammalian cells. *Nat. Rev. Genet.* **2**, 292-301.

Davidson, E. H., Rast, J. P., Oliveri, P., Ransick, A., Calestani, C., Yuh, C. H., Minokawa, T., Amore, G., Hinman, V., Arenas-Mena, C. et al. (2002a). A genomic regulatory network for development. *Science* **295**, 1669-1678.

Davidson, E. H., Rast, J. P., Oliveri, P., Ransick, A., Calestani, C., Yuh, C. H., Minokawa, T., Amore, G., Hinman, V., Arenas-Mena, C. et al. (2002b). A provisional regulatory gene network for specification of endomesoderm in the sea urchin embryo. *Dev. Biol.* **246**, 162-190.

Dion, M. F., Altschuler, S. J., Wu, L. F. and Rando, O. J. (2005). Genomic characterization reveals a simple histone H4 acetylation code. *Proc. Natl. Acad. Sci. USA* **102**, 5501-5506.

Dundr, M. and Misteli, T. (2002). Nucleolomics: an inventory of the nucleolus. *Mol. Cell* **9**, 5-7.

Dundr, M., Hoffmann-Rohrer, U., Hu, Q., Grummt, I., Rothblum, L. I.,

Phair, R. D. and Misteli, T. (2002). A kinetic framework for a mammalian RNA polymerase in vivo. *Science* **298**, 1623-1626.
Fernandez, P. C., Frank, S. R., Wang, L., Schroeder, M., Liu, S., Greene, J., Cocito, A. and Amati, B. (2003). Genomic targets of the human c-Myc protein. *Genes Dev.* **17**, 1115-1129.
Festenstein, R., Pagakis, S. N., Hiragami, K., Lyon, D., Verreault, A., Sekkali, B. and Kioussis, D. (2003). Modulation of heterochromatin protein 1 dynamics in primary Mammalian cells. *Science* **299**, 719-721.
Gilbert, D. M. (2002). Replication timing and transcriptional control: beyond cause and effect. *Curr. Opin. Cell Biol.* **14**, 377-383.
Gilbert, N., Boyle, S., Fiegler, H., Woodfine, K., Carter, N. P. and Bickmore, W. A. (2004). Chromatin architecture of the human genome: gene-rich domains are enriched in open chromatin fibers. *Cell* **118**, 555-566.
Goffeau, A., Barrell, B. G., Bussey, H., Davis, R. W., Dujon, B., Feldmann, H., Galibert, F., Hoheisel, J. D., Jacq, C., Johnston, M. et al. (1996). Life with 6000 genes. *Science* **274**, 546, 563-567.
Gorlich, D., Seewald, M. J. and Ribbeck, K. (2003). Characterization of Ran-driven cargo transport and the RanGTPase system by kinetic measurements and computer simulation. *EMBO J.* **22**, 1088-1100.
Harbison, C. T., Gordon, D. B., Lee, T. I., Rinaldi, N. J., Macisaac, K. D., Danford, T. W., Hannett, N. M., Tagne, J. B., Reynolds, D. B., Yoo, J. et al. (2004). Transcriptional regulatory code of a eukaryotic genome. *Nature* **431**, 99-104.
Harnpicharnchai, P., Jakovljevic, J., Horsey, E., Miles, T., Roman, J., Rout, M., Meagher, D., Imai, B., Guo, Y., Brame, C. J. et al. (2001). Composition and functional characterization of yeast 66S ribosome assembly intermediates. *Mol. Cell* **8**, 505-515.
Hatfield, S. D., Shcherbata, H. R., Fischer, K. A., Nakahara, K., Carthew, R. W. and Ruohola-Baker, H. (2005). Stem cell division is regulated by the microRNA pathway. *Nature* **435**, 974-978.
Hoffmann, A., Levchenko, A., Scott, M. L. and Baltimore, D. (2002). The IkappaB-NF-kappaB signaling module: temporal control and selective gene activation. *Science* **298**, 1241-1245.
Horn, P. J. and Peterson, C. L. (2002). Molecular biology. Chromatin higher order folding–wrapping up transcription. *Science* **297**, 1824-1827.
Ishigaki, Y., Li, X., Serin, G. and Maquat, L. E. (2001). Evidence for a pioneer round of mRNA translation: mRNAs subject to nonsense-mediated decay in mammalian cells are bound by CBP80 and CBP20. *Cell* **106**, 607-617.
Janicki, S. M., Tsukamoto, T., Salghetti, S. E., Tansey, W. P., Sachidanandam, R., Prasanth, K. V., Ried, T., Shav-Tal, Y., Bertrand, E., Singer, R. H. et al. (2004). From silencing to gene expression: real-time analysis in single cells. *Cell* **116**, 683-698.
Jeon, Y., Bekiranov, S., Karnani, N., Kapranov, P., Ghosh, S., Macalpine, D., Lee, C., Hwang, D. S., Gingeras, T. R. and Dutta, A. (2005). Temporal profile of replication of human chromosomes. *Proc. Natl. Acad. Sci. USA* **102**, 6419-6424.
Kampa, D., Cheng, J., Kapranov, P., Yamanaka, M., Brubaker, S., Cawley, S., Drenkow, J., Piccolboni, A., Bekiranov, S., Helt, G. et al. (2004). Novel RNAs identified from an in-depth analysis of the transcriptome of human chromosomes 21 and 22. *Genome Res.* **14**, 331-342.
Kimura, H. and Cook, P. R. (2001). Kinetics of core histones in living human cells: little exchange of H3 and H4 and some rapid exchange of H2B. *J. Cell Biol.* **153**, 1341-1353.
Kimura, T., Mills, F. C., Allan, J. and Gould, H. (1983). Selective unfolding of erythroid chromatin in the region of the active beta-globin gene. *Nature* **306**, 709-712.
Kurdistani, S. K., Tavazoie, S. and Grunstein, M. (2004). Mapping global histone acetylation patterns to gene expression. *Cell* **117**, 721-733.
Lamond, A. I. and Earnshaw, W. C. (1998). Structure and function in the nucleus. *Science* **280**, 547-553.
Lander, E. S., Linton, L. M., Birren, B., Nusbaum, C., Zody, M. C., Baldwin, J., Devon, K., Dewar, K., Doyle, M., FitzHugh, W. et al. (2001). Initial sequencing and analysis of the human genome. *Nature* **409**, 860-921.
Lee, C. K., Shibata, Y., Rao, B., Strahl, B. D. and Lieb, J. D. (2004). Evidence for nucleosome depletion at active regulatory regions genome-wide. *Nat. Genet.* **36**, 900-905.
Lee, R. C., Feinbaum, R. L. and Ambros, V. (1993). The C. elegans heterochronic gene lin-4 encodes small RNAs with antisense complementarity to lin-14. *Cell* **75**, 843-854.
Levens, D. L. (2003). Reconstructing MYC. *Genes Dev.* **17**, 1071-1077.
Lever, M. A., Th'ng, J. P., Sun, X. and Hendzel, M. J. (2000). Rapid exchange of histone H1.1 on chromatin in living human cells. *Nature* **408**, 873-876.
Lu, J., Getz, G., Miska, E. A., Alvarez-Saavedra, E., Lamb, J., Peck, D., Sweet-Cordero, A., Ebert, B. L., Mak, R. H., Ferrando, A. A. et al. (2005). MicroRNA expression profiles classify human cancers. *Nature* **435**, 834-838.
Lusser, A. and Kadonaga, J. T. (2003). Chromatin remodeling by ATP-dependent molecular machines. *BioEssays* **25**, 1192-1200.
Maniatis, T. and Reed, R. (2002). An extensive network of coupling among gene expression machines. *Nature* **416**, 499-506.
Matera, A. G. (1999). Nuclear bodies: multifaceted subdomains of the interchromatin space. *Trends Cell. Biol.* **9**, 302-309.
McNally, J. G., Muller, W. G., Walker, D., Wolford, R. and Hager, G. L. (2000). The glucocorticoid receptor: rapid exchange with regulatory sites in living cells. *Science* **287**, 1262-1265.
Mintz, P. J., Patterson, S. D., Neuwald, A. F., Spahr, C. S. and Spector, D. L. (1999). Purification and biochemical characterization of interchromatin granule clusters. *EMBO J.* **18**, 4308-4320.
Misteli, T. (2001). Protein dynamics: implications for nuclear architecture and gene expression. *Science* **291**, 843-847.
Misteli, T. (2005). Concepts in nuclear architecture. *BioEssays* **27**, 477-487.
Misteli, T., Gunjan, A., Hock, R., Bustin, M. and Brown, D. T. (2000). Dynamic binding of histone H1 to chromatin in living cells. *Nature* **408**, 877-881.
Nelson, D. E., Ihekwaba, A. E., Elliott, M., Johnson, J. R., Gibney, C. A., Foreman, B. E., Nelson, G., See, V., Horton, C. A., Spiller, D. G. et al. (2004). Oscillations in NF-kappaB signaling control the dynamics of gene expression. *Science* **306**, 704-708.
Neugebauer, K. M. (2002). On the importance of being co-transcriptional. *J. Cell Sci.* **115**, 3865-3871.
Orian, A., van Steensel, B., Delrow, J., Bussemaker, H. J., Li, L., Sawado, T., Williams, E., Loo, L. W., Cowley, S. M., Yost, C. et al. (2003). Genomic binding by the Drosophila Myc, Max, Mad/Mnt transcription factor network. *Genes Dev.* **17**, 1101-1114.
Parada, L. A., Sotiriou, S. and Misteli, T. (2004). Spatial genome organization. *Exp. Cell Res.* **296**, 64-70.
Pendle, A. F., Clark, G. P., Boon, R., Lewandowska, D., Lam, Y. W., Andersen, J., Mann, M., Lamond, A. I., Brown, J. W. and Shaw, P. J. (2005). Proteomic analysis of the Arabidopsis nucleolus suggests novel nucleolar functions. *Mol. Biol. Cell* **16**, 260-269.
Phair, R. D., Gorski, S. A. and Misteli, T. (2004a). Measurement of dynamic protein binding to chromatin in vivo, using photobleaching microscopy. *Methods Enzymol.* **375**, 393-414.
Phair, R. D., Scaffidi, P., Elbi, C., Vecerova, J., Dey, A., Ozato, K., Brown, D. T., Hager, G., Bustin, M. and Misteli, T. (2004b). Global nature of dynamic protein-chromatin interactions in vivo: three-dimensional genome scanning and dynamic interaction networks of chromatin proteins. *Mol. Cell. Biol.* **24**, 6393-6402.
Rappsilber, J., Ryder, U., Lamond, A. I. and Mann, M. (2002). Large-scale proteomic analysis of the human spliceosome. *Genome Res.* **12**, 1231-1245.
Ren, B., Robert, F., Wyrick, J. J., Aparicio, O., Jennings, E. G., Simon, I., Zeitlinger, J., Schreiber, J., Hannett, N., Kanin, E. et al. (2000). Genome-wide location and function of DNA binding proteins. *Science* **290**, 2306-2309.
Riddick, G. and Macara, I. G. (2005). A systems analysis of importin-{alpha}-{beta} mediated nuclear protein import. *J. Cell Biol.* **168**, 1027-1038.
Robert, F., Pokholok, D. K., Hannett, N. M., Rinaldi, N. J., Chandy, M., Rolfe, A., Workman, J. L., Gifford, D. K. and Young, R. A. (2004). Global position and recruitment of HATs and HDACs in the yeast genome. *Mol. Cell* **16**, 199-209.
Roh, T. Y., Cuddapah, S. and Zhao, K. (2005). Active chromatin domains are defined by acetylation islands revealed by genome-wide mapping. *Genes Dev.* **19**, 542-552.
Rout, M. P., Aitchison, J. D., Suprapto, A., Hjertaas, K., Zhao, Y. and Chait, B. T. (2000). The yeast nuclear pore complex: composition, architecture, and transport mechanism. *J. Cell Biol.* **148**, 635-651.
Rout, M. P., Aitchison, J. D., Magnasco, M. O. and Chait, B. T. (2003). Virtual gating and nuclear transport: the hole picture. *Trends Cell. Biol.* **13**, 622-628.
Saitoh, N., Spahr, C. S., Patterson, S. D., Bubulya, P., Neuwald, A. F. and Spector, D. L. (2004). Proteomic analysis of interchromatin granule clusters. *Mol. Biol. Cell* **15**, 3876-3890.
Schena, M., Shalon, D., Davis, R. W. and Brown, P. O. (1995). Quantitative monitoring of gene expression patterns with a complementary DNA microarray. *Science* **270**, 467-470.
Scherl, A., Coute, Y., Deon, C., Calle, A., Kindbeiter, K., Sanchez, J. C.,

Journal of Cell Science

Greco, A., Hochstrasser, D. and Diaz, J. J. (2002). Functional proteomic analysis of human nucleolus. *Mol. Biol. Cell* **13**, 4100-4109.
Schirmer, E. C., Florens, L., Guan, T., Yates, J. R., 3rd. and Gerace, L. (2003). Nuclear membrane proteins with potential disease links found by subtractive proteomics. *Science* **301**, 1380-1382.
Schubeler, D., Scalzo, D., Kooperberg, C., van Steensel, B., Delrow, J. and Groudine, M. (2002). Genome-wide DNA replication profile for Drosophila melanogaster: a link between transcription and replication timing. *Nat. Genet.* **32**, 438-442.
Schubeler, D., MacAlpine, D. M., Scalzo, D., Wirbelauer, C., Kooperberg, C., van Leeuwen, F., Gottschling, D. E., O'Neill, L. P., Turner, B. M., Delrow, J. et al. (2004). The histone modification pattern of active genes revealed through genome-wide chromatin analysis of a higher eukaryote. *Genes Dev.* **18**, 1263-1271.
Shav-Tal, Y., Darzacq, X., Shenoy, S. M., Fusco, D., Janicki, S. M., Spector, D. L. and Singer, R. H. (2004). Dynamics of single mRNPs in nuclei of living cells. *Science* **304**, 1797-1800.
Simpson, J. C., Wellenreuther, R., Poustka, A., Pepperkok, R. and Wiemann, S. (2000). Systematic subcellular localization of novel proteins identified by large-scale cDNA sequencing. *EMBO Rep.* **1**, 287-292.
Slepchenko, B. M., Schaff, J. C., Carson, J. H. and Loew, L. M. (2002). Computational cell biology: spatiotemporal simulation of cellular events. *Annu. Rev. Biophys. Biomol. Struct.* **31**, 423-441.
Slonim, D. K. (2002). From patterns to pathways: gene expression data analysis comes of age. *Nat. Genet.* **32**, 502-508.
Smith, A. E., Slepchenko, B. M., Schaff, J. C., Loew, L. M. and Macara, I. G. (2002). Systems analysis of Ran transport. *Science* **295**, 488-491.
Spector, D. L. (2003). The dynamics of chromosome organization and gene regulation. *Annu. Rev. Biochem.* **72**, 573-608.
Sprague, B. L. and McNally, J. G. (2005). FRAP analysis of binding: proper and fitting. *Trends Cell. Biol.* **15**, 84-91.
Strahl, B. D. and Allis, C. D. (2000). The language of covalent histone modifications. *Nature* **403**, 41-45.
Tackett, A. J., Dilworth, D. J., Davey, M. J., O'Donnell, M., Aitchison, J. D., Rout, M. P. and Chait, B. T. (2005). Proteomic and genomic characterization of chromatin complexes at a boundary. *J. Cell Biol.* **169**, 35-47.
Tange, T. O., Nott, A. and Moore, M. J. (2004). The ever-increasing complexities of the exon junction complex. *Curr. Opin. Cell Biol.* **16**, 279-284.
Venter, J. C., Adams, M. D., Myers, E. W., Li, P. W., Mural, R. J., Sutton, G. G., Smith, H. O., Yandell, M., Evans, C. A., Holt, R. A. et al. (2001). The sequence of the human genome. *Science* **291**, 1304-1351.
Waterston, R. H., Lindblad-Toh, K., Birney, E., Rogers, J., Abril, J. F., Agarwal, P., Agarwala, R., Ainscough, R., Alexandersson, M., An, P. et al. (2002). Initial sequencing and comparative analysis of the mouse genome. *Nature* **420**, 520-562.
White, E. J., Emanuelsson, O., Scalzo, D., Royce, T., Kosak, S., Oakeley, E. J., Weissman, S., Gerstein, M., Groudine, M., Snyder, M. et al. (2004). DNA replication-timing analysis of human chromosome 22 at high resolution and different developmental states. *Proc. Natl. Acad. Sci. USA* **101**, 17771-17776.
Wightman, B., Ha, I. and Ruvkun, G. (1993). Posttranscriptional regulation of the heterochronic gene lin-14 by lin-4 mediates temporal pattern formation in C. elegans. *Cell* **75**, 855-862.
Woodfine, K., Fiegler, H., Beare, D. M., Collins, J. E., McCann, O. T., Young, B. D., Debernardi, S., Mott, R., Dunham, I. and Carter, N. P. (2004). Replication timing of the human genome. *Hum. Mol. Genet.* **13**, 191-202.
Yuan, G. C., Liu, Y. J., Dion, M. F., Slack, M. D., Wu, L. F., Altschuler, S. J. and Rando, O. J. (2005). Genome-scale identification of nucleosome positions in *S. cerevisiae*. *Science* **309**, 626-630.
Yuh, C. H., Bolouri, H. and Davidson, E. H. (1998). Genomic cis-regulatory logic: experimental and computational analysis of a sea urchin gene. *Science* **279**, 1896-1902.
Yuh, C. H., Bolouri, H. and Davidson, E. H. (2001). Cis-regulatory logic in the endo16 gene: switching from a specification to a differentiation mode of control. *Development* **128**, 617-629.
Zeitlinger, J., Simon, I., Harbison, C. T., Hannett, N. M., Volkert, T. L., Fink, G. R. and Young, R. A. (2003). Program-specific distribution of a transcription factor dependent on partner transcription factor and MAPK signaling. *Cell* **113**, 395-404.
Zhou, Z., Licklider, L. J., Gygi, S. P. and Reed, R. (2002). Comprehensive proteomic analysis of the human spliceosome. *Nature* **419**, 182-185.