Seminars in Cell & Developmental Biology 20 (2009) 856–862

Contents lists available at ScienceDirect

Seminars in Cell & Developmental Biology

journal homepage: www.elsevier.com/locate/semcdb

Invited review

A systems biology approach to understanding *cis*-regulatory module function

Danuta M. Jeziorska[1], Kate W. Jordan[1], Keith W. Vance*

Departments of Systems Biology and Biological Sciences, University of Warwick, Biomedical Research Institute, Gibbet Hill, Coventry CV4 7AL, UK

ARTICLE INFO

Article history:
Available online 4 August 2009

Keywords:
cis-Regulatory module
Transcription
Chromatin
Gene regulatory networks
Imaging

ABSTRACT

The genomic instructions used to regulate development are encoded within a set of functional DNA elements called *cis*-regulatory modules (CRMs). These elements determine the precise patterns of temporal and spatial gene expression. Here we summarize recent progress made towards cataloguing and characterizing the complete repertoire of CRMs. We describe CRMs as genomic information processing devices containing clusters of transcription factor binding sites and we position CRMs as nodes within large gene regulatory networks. We define CRM architecture and describe how these genomic elements process the information they encode to their target genes. Furthermore, we present an overview describing high-throughput techniques to identify CRMs genome wide and experimental methodologies to validate their function on a large scale. This review emphasizes the advantages and power of a systems biology approach which integrates computational and experimental technologies to further our understanding of CRM function.

Contents

1. Introduction 856
2. Defining CRMs 857
   2.1. CRM characterization 857
   2.2. Mode of CRM action 858
3. CRMs as components of genetic networks 858
4. High throughput identification of CRMs 859
5. Analysis of CRM function 860
   5.1. Model systems to validate CRM function 860
   5.2. Measuring CRM function in real-time 860
      5.2.1. Imaging live cells 860
      5.2.2. Imaging live organisms 860
6. Systems approach 860
   Acknowledgements 861
   References 861

1. Introduction

Development of multi-cellular organisms is crucially dependent on the correct temporal and spatial control of gene expression. With proper regulation a single cell can develop into a fully functioning, complex organism that has homeostatic mechanisms for dealing with the daily onslaught of environmental conditions. Furthermore, aberrant gene regulation is becoming increasingly linked to conditions affecting our morbidity and mortality.

The correct expression of a gene is controlled by genomic information processing devices called *cis*-regulatory modules (CRMs) which encode the precise instructions for transcriptional activation and repression. Understanding how CRMs operate involves the elucidation of CRM architecture, identification of the factors which modulate CRM activity and measurement of their ability to regulate gene transcription. Unlike genes and proteins, no database for the full gamut of CRMs has been created, and identifying and evaluating CRMs remains challenging. Any analysis of CRMs has inherent complexity and we propose that the more fully integrated the approach to studying CRMs becomes, the more information

* Corresponding author. Tel.: +44 2476 528 380.
E-mail address: k.w.vance@warwick.ac.uk (K.W. Vance).
[1] These authors contributed equally to this work.

1084-9521/$ – see front matter © 2009 Elsevier Ltd. All rights reserved.
doi:10.1016/j.semcdb.2009.07.007

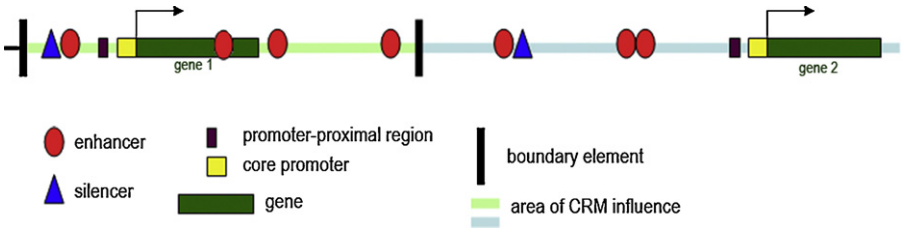

Fig. 1. CRM characterization. CRMs are DNA sequence elements, such as enhancers, silencers, promoters, boundary elements and locus control regions, which modulate gene transcription. CRMs can be located in the 5′, 3′, intronic, or coding regions of their cognate and surrounding genes. The effects of CRMs are combinatorial and the influence of CRMs can be unique to one gene or shared among many, with zones of influence largely dictated by the presence of boundary elements.

these studies will yield. For this reason a systems biology approach, combining several of the techniques presented here, can provide a powerful toolkit to investigate *cis*-regulatory information processing. As such, this review focuses on the areas of gene regulation that have most directly benefited from the advent of advanced, high-throughput technologies, such as CRM discovery and characterization, and the most current methods to analyze CRM function on a large scale. Indeed, recent progress in the field has allowed us to contextualize CRMs such that we now understand them to account for a large percentage of both the genetic landscape and genetic variability.

2. Defining CRMs

The definition of a *cis*-regulatory module has evolved from the original definition of an enhancer as a *cis*-acting DNA sequence that increases the rate of transcription from a linked promoter in a manner independent of its location or orientation relative to the RNA start site (reviewed in [1]). However, as our understanding of gene regulation increases a more expansive definition is required, which should focus on the ability of CRMs to regulate the spatial and temporal expression pattern of genes in capacities beyond simply promoting transcription initiation. A CRM is therefore defined here as a DNA sequence containing transcription factor binding sites (TFBSs) clustered into modular structures (typically 50 bp to 1.5 kbp in size) that act to regulate various aspects of the transcription process: in this classification CRMs include, but are not limited to, enhancers, silencers, promoters, locus control regions (LCRs), and insulators (extensively reviewed in [2–8]) (Fig. 1). These CRMs have the ability to act across huge landscapes of the genome and their position relative to their target genes varies enormously. CRMs can be located a few kb to many hundred kb upstream of the basal promoter [9,10], in the coding sequence [11], within introns [12], or in the downstream 3′ UTR sequences [13,14], as well as on different chromosomes where a CRM is able to function in *trans* [15–18] (Fig. 1).

A CRM exerts its effect by performing a computation in which the input, the concentration of active transcription factors and associated co-factors expressed in a cell at a given time, is read and integrated into a transcriptional output. Additionally, a single gene may contain multiple CRMs and the ability of these CRMs to function in a combinatorial manner is necessary to generate precise four dimensional expression patterns. Furthermore, gene expression is not solely dictated by primary DNA sequence. The eukaryotic genome is packaged into chromatin and the regulation of chromatin structure and nuclear organization plays an important role in the control of CRM function (reviewed in [19–24]).

While evidence has accumulated to display the different types of CRMs in the genome, conflicting models have often been generated to explain the mechanisms of CRM action or elucidate how CRMs modulate transcription. Examining the large cohort of published data leads to the natural conclusion that most models have validity and different CRMs will have different modes of action and diverse effects. In this way each new contribution to the field extends our knowledge of *cis*-regulatory information processing in a way that often does not nullify previous hypotheses. Here we summarize our current understanding of the regulatory genome emphasizing that the field will continue to expand as new methods and technologies are created.

2.1. CRM characterization

Current classification of CRMs can be based on the type of information processing they encode and the arrangement of TFBSs within the cluster, dividing CRMs into highly cooperative and coordinated units, called enhanceosomes, and more flexible functional units, termed billboards (reviewed in [25]). In the enhanceosome model the assembly of the structural complex and the exact/highly cooperative arrangement of TFBSs is critical. Disruption or displacement of a single binding site in the enhanceosome complex can nullify functionality. The enhanceosome, for which the virus inducible interferon-β (INF-β) enhancer is a paradigm [26], acts as a central processing unit, merging multiple inputs from bound transcription factors and giving a single output to the target gene. In contrast, the billboard class of CRM has been used to describe many developmentally active regulatory elements. Billboard elements have a flexible organization of TFBSs and their transcriptional output is measured as the sum effect of the bound transcription factors.

CRMs are further characterized by the ways in which they affect the overall probability, proportion and rate of gene transcription. Conventionally enhancers were thought to increase the rate of initiation of transcription from a linked promoter (the rheostatic model); however, single-cell assays have revealed enhancers can increase the probability that a gene will be activated, with little or no effect on the transcription rate (the binary model) [27].

In the binary response model a CRM will act as part of an "on/off" switch for transcription. In this all or nothing model the CRM would increase or decrease the proportion of cells transcribing a gene, but not the rate of transcription initiation. By contrast, the rheostatic response model describes CRMs as regulating transcription rates quantitatively, such that an enhancer would increase the rate of initiation of transcription of its cognate gene. The rheostatic model has traditionally been the most widely used to describe CRM affect on transcription [28], and quantitative control of the transcription rate by CRMs is intuitively more synergistic with genes that are constantly transcribed at a basal rate. However, categorical adherence to one model may be erroneous as the necessity for both sharp boundaries of gene expression in space and time, as well as the continuous need for certain genes to be expressed, lend themselves to alternate forms of control.

Notably, population based measurements of transcription activity are not sufficient to distinguish whether enhancers modulate the transcription rate by an "on or off" or graded mechanism. Single-cell assays such as flow cytometry and live cell imaging are needed to acquire the necessary resolution to accurately establish the exact response of a CRM. Binary or rheostatic mechanisms of transcrip-

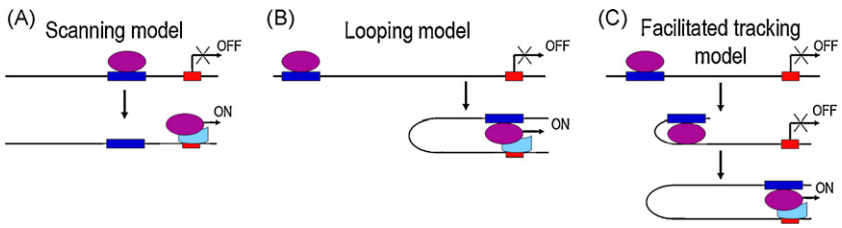

Fig. 2. CRM modes of action. (A) In the scanning model the transcription factor-cofactor complex (purple oval) forms at a CRM (dark blue box) and then moves continuously along the DNA sequence until it encounters the general transcription machinery (light blue shape) at the promoter (red box). (B) In the looping model transcription factors bound to a CRM interact with the promoter through looping out the intervening DNA. (C) The facilitated tracking model combines elements of the scanning and looping models, where the transcription factor complex assembled on the CRM moves across the intervening sequence to the promoter via small steps forming a stable looped configuration.

tional regulation are not mutually exclusive. A study by Rossi et al. established with single-cell resolution that transcription from a reporter construct was rheostatic in response to a single extracellular input when either activator or repressor proteins alone were present in the cell. However, when both activator and repressor proteins were competing for the same regulatory element the transcriptional response was converted to a binary on/off switch [29]. Such a model takes into account a fundamental feature of CRMs: their function is dictated by the transcription factors bound to them, and the presence or absence of these transcription factors in the cells is dictated by the expression of other genes, the cellular environment, extracellular signaling, and so forth. In some contexts a CRM might function to increase the amount of transcript being produced, while in others the control of even the smallest amount of transcript must be tightly regulated in an on/off way. The dual functionality of CRMs could be an important mechanism through which the same gene participates in multiple pathways and developmental contexts.

The description of a CRM as binary or rheostatic does not preclude it being stochastic. Random fluctuations in the level of gene expression can either be intrinsic, resulting from variations of independent expression of unlinked genes in the same cell, or extrinsic, where factors effect the expression of all genes in the same cell equally, but create differences between cells (for a review see [30]). Transcriptional noise can lead to the generation of cellular heterogeneity within clonal population of cells and can have profound implications in the control of developmental decisions (reviewed in [31]).

2.2. Mode of CRM action

CRMs can regulate transcription over large genomic distances. An overview of the literature reveals that three models have been proposed to describe CRM communication with their target promoter (Fig. 2).

The DNA scanning model (Fig. 2A) proposes that transcription factors first bind to CRMs. The protein-complex then moves continuously along the DNA sequence until it encounters the promoter of its cognate gene (reviewed in [1]). This model is compatible with the function of boundary elements as an insulator would block the march of the protein complex prior to contact with the promoter. However, the DNA-scanning model is inconsistent with transvection-like phenomena, where a CRM regulates transcription from its target promoter on the homologous chromosome, and could only work in a scenario where the movement of the protein complex was short. If the complex encountered, for example, a promoter of an intervening gene, it would lead to transcription of an unspecific target. While DNA scanning has not been discounted, the evidence of more sophisticated methods of CRM regulation in eukaryotes continues to accumulate. In the looping model (Fig. 2B) CRMs communicate with their target genes through direct protein–protein interactions between transcription factors bound to the CRM and components of the general transcription machinery assembled at the core promoter, with the intervening DNA looping out [32,33]. Experiments using the chromatin conformation capture (3C) assay [34] have provided direct experimental evidence of both looping interactions between distant genomic regulatory regions and interchromosomal interactions in eukaryotic cells. Furthermore, these genomic interactions have been shown to mediate both activation and repression of transcription [17,35–38]. Molecular imaging methods have also provided evidence of regulatory DNA looping interactions. Fluorescent *in situ* hybridization (FISH) studies have visualized trans-homolog enhancer–promoter interactions of the *Drosophila* Abd-B Hox locus [39] while RNA tagging and recovery of associated protein (RNA-TRAP), a modification of the RNA-FISH technique, showed that the H2S enhancer of the LCR of β-globin locus is in close proximity to an actively transcribed HBB (β-globin) gene located 50 kb away *in vivo* [40]. A further model, facilitated tracking (Fig. 2C), integrates elements of the looping and scanning mechanisms. In this model the complex of transcription factors and cofactors formed at a CRM moves unidirectionaly along the intervening sequence in small steps forming looped chromatin structures (reviewed in [1]). Chromatin immunoprecipitation (ChIP) experiments in combination with DNAse-I and micrococcal nuclease (MNAse) analysis of the nucleosome structure of the HNF-α locus suggested that HNF-α expression is controlled by a facilitated tracking mechanism during enterocyte differentiation. These experiments showed that formation of a transcription factor complex on an upstream enhancer is followed by the tracking of the entire DNA–protein complex along the intervening sequences to the HNF-α promoter and spreading of histone hyperacetylation along the locus [41]. This model is consistent with boundary element function and long-distance and orientation-independent transcription activation. However, like DNA scanning this mechanism cannot explain interchromosomal CRM–promoter interactions.

Models of CRM–promoter communication have been extended by recent work mapping the dynamic organization of the nucleus. These studies showed that gene transcription occurs in bursts at discrete nuclear foci. These foci, termed transcription factories, colocalize with high concentrations of immobilized RNA polymerase II and can contain multiple actively transcribed genes [42–45]. Taken together, these experiments reveal that any model of CRM information processing must be viewed within the three dimensional structure of the nucleus.

3. CRMs as components of genetic networks

CRMs do not function in isolation but are components of large interconnected gene regulatory networks (Fig. 3). These networks contain the instructions to carry out specific developmental decisions. A single gene within this network can contain many CRMs, each of which contribute to the final transcription rate. Interactions

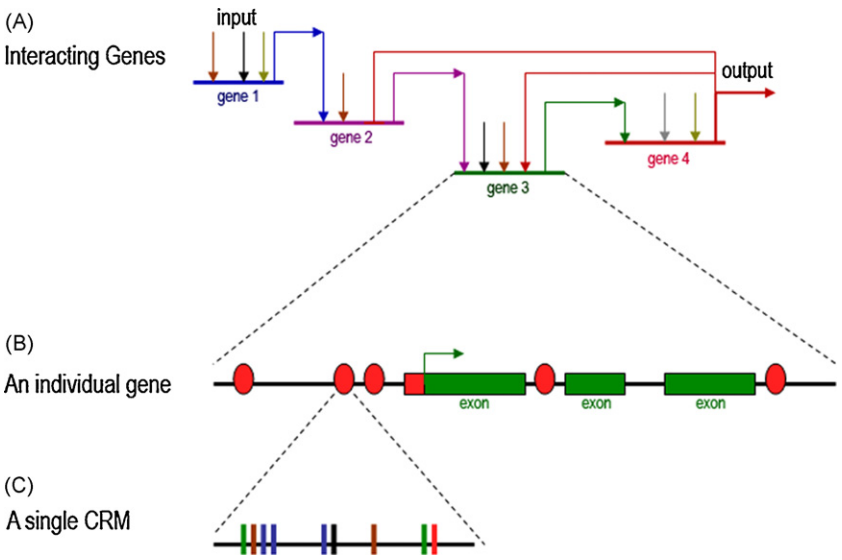

Fig. 3. Gene regulatory network organization. (A) Genes are components of regulatory networks, where a gene receives an input from the network and provides an output that regulates different nodes of the network. (B) An individual gene contains several CRMs (red ovals), each of which contribute to the final transcription rate and regulate the spatial and temporal expression pattern (exons indicated as green boxes; promoter as a red box). (C) An individual CRM contains clusters of TFBSs (presented as boxes of different colors).

between these CRMs modulate the expression of a gene through time and space and provide a unique code for when, where, and at which level a gene is transcribed. Each CRM can also have a variety of functions in different cellular conditions. A CRM can simultaneously contain enhancing and repressing TFBSs, the combinatorics of which dictate its overall transcriptional output (for reviews see [2,46]). The regulation of the *endo*16 gene of *Strongylocentrotus purpuratus* provides a classic example. The *endo16* gene is regulated during development by a 2.3 kb region that can be broken into distinct modules containing TFBSs that stimulate expression at differing time points in development (CRM A–F). CRM A controls initial *endo16* endodermal expression, while module B leads to later expression in the midgut and modules F, E, and DC have a negative regulatory effect and prevent aberrant expression [47,48]. Furthermore, each individual CRM can be seen a logic-processing unit [49] whose efficacy is controlled by the cellular context and the stoichiometry, affinity, spacing and arrangements of TFBSs within them [50]. In this way genes and their CRMs themselves act as "nodes" in genetic networks (reviewed in [51]). Given the human genome encodes only 25,000 genes [52] generating 2000 transcription factors [53,54], epistasis, polygeny, and pleiotropy have all been attributed to the interconnection of CRM networks [46,55]. The diversity achieved in the cells, tissues and organs of an organism is accomplished only when genes properly resolve the signals they are receiving and act appropriately. This situation is largely dictated by CRMs and exemplified by the role of CRMs in the gene regulatory networks controlling sea urchin development (reviewed in [46]).

4. High throughput identification of CRMs

Identification and characterization of CRMs within the vertebrate genome remains challenging. In contrast to the coding sequence, where analysis of cDNAs, expression sequence tags (ESTs) and translated proteins define a vocabulary that significantly facilitates protein annotation, our current knowledge about the grammar of regulatory modules remains incomplete. In recent years the development of large scale experimental technologies, along with the sequencing of a growing number of vertebrate genomes and the generation of powerful bioinformatic tools, has made significant steps towards decoding the sequences that orchestrate gene regulation. A systems biology approach intersecting computational as well as *in vitro* or *in vivo* analysis, ranging from single-cell examinations to studies of entire model organisms, can lead to a comprehensive characterization of CRMs as components of genetic regulatory networks.

Genomic sequences with the potential to have a regulatory function can be identified and characterized on a large scale *in silico* using computational and bioinformatic techniques. These offer sophisticated alternatives to classical experimental deletion analyses probing limited regulatory information on a single gene and allow for a rigorous examination of the set of CRMs prior to any laboratory experimentation. Methods of *in silico* CRM identification include phylogenetic footprinting, trained models, and advanced transcription factor binding site predictions, among others [56–58]. Continually increasing examples have shown *in silico* methods for CRM discovery are an impressive tool whose predictive powers are enhanced by incorporating as many factors as possible into the modeling: for example, an examination of conservation of non-coding regions of DNA across different species is more likely to find functional CRMs if the evolutionary distance between the species is increased and/or transcription factor binding predictions are incorporated [59–61].

The advent of microarray and massively parallel sequencing based experimental techniques such as ChIP-chip, ChIP-Seq, DNAse-Seq and DNAse-chip has made the large scale characterization of the regulatory genome a possibility. These techniques have created a platform to identify potential regulatory regions [62,63], to define the genome-wide targets of specific transcription factors [64–68] and cofactors [69] and to rapidly explore the chromatin modification status and nucleosome positioning of the genome [70–75]. Using these approaches the number of potential CRMs has increased by orders of magnitude in the matter of only a few years. Furthermore, studies using 3C based methods have proposed that the genome is organized into a three dimensional network of physical and functional interactions between different loci [34]. Development of high throughput 3C based techniques (4C and 5C) in which genomic interactions are identified using microarray and/or next generation sequencing methods [76–80] can be used to significantly increase the resolution of CRM identification and allow genome wide mapping of regulatory interactions. Intersecting these different technologies provides a platform to cat-

alogue the entire range of genomic functional elements and develop our understanding of the *cis*-regulatory grammar encoded in the genome.

5. Analysis of CRM function

Functional validation of the large number of identified putative CRMs remains a major challenge in the post genomics era. In this section we will present a short overview of the use of model organisms and cell lines in culture to systematically measure and manipulate CRM activity in a high throughput manner.

5.1. Model systems to validate CRM function

Studies using cells in culture, in which the ability of cloned CRMs to drive expression of fluorescent or luciferase reporters is measured, are ideal for the large scale rapid testing of CRM function [81]. Such reporter gene assays can be combined with biochemical techniques to precisely define CRM architecture. Data acquisition using flow cytometry and/or molecular imaging allows CRM activity to be measured with single-cell resolution and the CRM mechanism of action to be determined. Using this approach, huge steps have been made towards an exact functional annotation and classification of these genomic regulatory elements.

While studies using cell lines are able to explore CRM activity under desired cellular conditions, for example, in cancer cells in culture or during differentiation to a specific cell type, *in vivo* transgenics allow the role of a CRM to be followed throughout the development of entire organisms. Invertebrates, lower vertebrates, birds, and mammals are commonly used in the development of genetically modified organisms (GMOs). Invertebrates have been used to rapidly generate maps of entire networks of gene regulation in development and define the operating principles of CRM logic (reviewed in [46]). Additionally as invertebrates are at the base of the chordate phylum they are ideal organisms to examine the evolution of gene regulation. Comparatively, the utility of using fish species as GMOs relies heavily on their early divergence from mammals and their compact genome [82]. Zebrafish (*Danio rerio*) are a favored GMO due to the transparency they maintain throughout development and their fecundity [83]. The use of frogs (*Xenopus laevis* and *X. tropicalis*) as GMOs has rapidly increased as protocols for generating transgenics have improved allowing them to be used for high-throughput analysis of potential CRMs [84,85]. Chicken transgenics are attractive given their historical role in elucidating key genes and pathways in development, as well as the ease of generating transgenics via electroporation [86]. Mice have emerged as the foremost animal model for most bio-medical research due to their recent divergence from humans (approximately 70 million years ago), the high homology between mouse and human genomes, the establishment of extensive GMO mouse strains, and the unique ability to create transgenics using embryonic stem (ES) cells [87]. ES cells offer the ability to engineer DNA as desired *ex vivo* then generate mice from the cells *in vivo* [88].

GMOs can be used individually (examining a single species) or in parallel for comparison of function of CRM sequences across species and evolution. Moreover transgenic animals can be generated using sequences from their own genome as well as from others. As an illustration, Japanese pufferfish (*Takifugu rubripes*) have a minute genome and potential CRMs conserved between *fugu* and humans can be prioritized as likely to have function then analyzed using mouse transgenics; using this system 45% of potential regions from *fugu* tested functioned as enhancers in embryonic day 11.5 (e 11.5) mice [61].

5.2. Measuring CRM function in real-time

5.2.1. Imaging live cells

Live cell imaging offers an extremely powerful approach to measure and decipher the dynamics of CRM function. CRMs can be used to drive expression of fluorescent/luciferase reporters in cells in culture and their activity quantified in real time with single-cell resolution. In this way, measurement noise inherent in averaging over a heterogeneous population of cells is eliminated and gene expression is visualized at the level in which transcriptional decisions are made.

A few selected examples measuring the dynamics of oscillatory gene expression highlight the power of applying a live cell imaging approach to validate CRM function. The Hes1 transcription factor is a component of the somite segmentation clock. Real-time imaging of luciferase reporters under control of the Hes1 promoter showed that Hes1 expression oscillates with a periodicity of approximately 2 h in individual fibroblast and neural precursor cells [89]. These oscillations could not be detected in asynchronous populations of cells. In addition, time lapse fluorescent microscopy measurements of the activity of the Bmal1 promoter, a basic helix loop helix transcription factor that is part of the circadian clockwork, have been used to determine the period length of circadian oscillations in individual NIH3T3 cells [90]. This work has opened the way for a biochemical and genetic dissection of the mammalian circadian clock. An added advantage of a live cell imaging approach is that mathematical modeling of high resolution single-cell reporter measurements can be applied to reconstitute transcription profiles when reporter mRNA and reporter protein degradation rates are calculated [91].

Furthermore, advances in automated instrumentation such as the High Content Screening (HCS) platform developed at Cellomics Inc. [92], along with the development of image analysis tools such as those established by Alon and co-workers to identify and track individual cells [93–95], will permit the acquisition of high throughput single-cell measurements essential for the characterization of genome function.

5.2.2. Imaging live organisms

Confocal and multi-photon (such as two-photon excitation or 2PE) microscopy have recently been adapted to time-lapse studies of entire model organisms. The term *in toto* is now applied to tracking individual cells and tissues across both time and space in the developing organism [96]. Previous analyses have been marred by a lack of penetration into deep regions of the tissue, slow data accumulation, low spectral resolution, as well as phototoxicity to the organism. Furthermore, imaging of an entire organism was developed to study a sample that had been culled and for temporal information the GMO had to be culled at each time point and individually analyzed. Laser scanning microscopy, where only specific and focused fluorophores are excited, is being adapted to study live GMOs. Recent reports have shown entire live embryos of several species can be imaged for extended periods of time [97,98]. However, this approach still leads to the ultimate demise of the specimen and great care must be taken in the choice of fluorophore as well as assuring the optimal conditions for immobilization and maintenance of the specimen during imaging.

6. Systems approach

With increasing awareness of the various roles CRMs can play in gene expression, as well as rapidly advancing techniques for high-throughput assays, the utility of a systems approach comes into sharp focus. Just as an individual CRM can be seen as a node in a network, so to can the individual techniques and methods used to

probe CRMs. Each methodology has benefits and limitations, and a combination of traditional laboratory-based molecular and cellular biology with statistical, mathematic and bioinformatic modeling and data analysis allows for a much larger scope of work to be undertaken. Continuing the network parallel: nodes of networks are highly interconnected with many shared branches, just as many techniques lend themselves to obvious combinations as they yield complementary data. In the -omics era the value of collaborative and integrative efforts have yielded impressive results, and with the use of a systems biology approach we can begin to collate generations of research into the regulatory mechanisms guiding cellular behaviors as well as rapidly advance the current research being undertaken into all aspects of CRMs discussed in this review.

Acknowledgements

We thank Dr Robert Old for critically reading the manuscript.

References

[1] Blackwood EM, Kadonaga JT. Going the distance: a current view of enhancer action. Science 1998;281:60–3.
[2] Wray GA, Hahn MW, Abouheif E, Balhoff JP, Pizer M, Rockman MV, et al. The evolution of transcriptional regulation in eukaryotes. Mol Biol Evol 2003;20:1377–419.
[3] Bushey AM, Dorman ER, Corces VG. Chromatin insulators: regulatory mechanisms and epigenetic inheritance. Mol Cell 2008;32:1–9.
[4] Juven-Gershon T, Hsu JY, Theisen JW, Kadonaga JT. The RNA polymerase II core promoter—the gateway to transcription. Curr Opin Cell Biol 2008;20:253–9.
[5] Li Q, Peterson KR, Fang X, Stamatoyannopoulos G. Locus control regions. Blood 2002;100:3077–86.
[6] Sandelin A, Carninci P, Lenhard B, Ponjavic J, Hayashizaki Y, Hume DA. Mammalian RNA polymerase II core promoters: insights from genome-wide studies. Nat Rev Genet 2007;8:424–36.
[7] Valenzuela L, Kamakaka RT. Chromatin insulators. Annu Rev Genet 2006;40:107–38.
[8] Wallace JA, Felsenfeld G. We gather together: insulators and genome organization. Curr Opin Genet Dev 2007;17:400–7.
[9] Bien-Willner GA, Stankiewicz P, Lupski JR. SOX9cre1, a cis-acting regulatory element located 1.1 Mb upstream of SOX9, mediates its enhancement through the SHH pathway. Hum Mol Genet 2007;16:1143–56.
[10] Jack J, Dorsett D, Delotto Y, Liu S. Expression of the cut locus in the *Drosophila* wing margin is required for cell type specification and is regulated by a distant enhancer. Development 1991;113:735–47.
[11] Tumpel S, Cambronero F, Sims C, Krumlauf R, Wiedemann LM. A regulatory module embedded in the coding region of Hoxa2 controls expression in rhombomere 2. Proc Natl Acad Sci USA 2008;105:20077–82.
[12] Zimmerman L, Parr B, Lendahl U, Cunningham M, McKay R, Gavin B, et al. Independent regulatory elements in the nestin gene direct transgene expression to neural stem cells or muscle precursors. Neuron 1994;12:11–24.
[13] Palmer MB, Majumder P, Green MR, Wade PA, Boss JM. A 3′ enhancer controls snail expression in melanoma cells. Cancer Res 2007;67:6113–20.
[14] Yochum GS, Cleland R, Goodman RH. A genome-wide screen for beta-catenin binding sites identifies a downstream enhancer element that controls c-Myc gene expression. Mol Cell Biol 2008;28:7368–79.
[15] Bacher CP, Guggiari M, Brors B, Augui S, Clerc P, Avner P, et al. Transient colocalization of X-inactivation centres accompanies the initiation of X inactivation. Nat Cell Biol 2006;8:293–9.
[16] Lomvardas S, Barnea G, Pisapia DJ, Mendelsohn M, Kirkland J, Axel R. Interchromosomal interactions and olfactory receptor choice. Cell 2006;126:403–13.
[17] Spilianakis CG, Lalioti MD, Town T, Lee GR, Flavell RA. Interchromosomal associations between alternatively expressed loci. Nature 2005;435:637–45.
[18] Xu N, Tsai CL, Lee JT. Transient homologous chromosome pairing marks the onset of X inactivation. Science 2006;311:1149–52.
[19] Cremer T, Cremer C. Chromosome territories, nuclear architecture and gene regulation in mammalian cells. Nat Rev Genet 2001;2:292–301.
[20] Dillon N. Gene regulation and large-scale chromatin organization in the nucleus. Chromosome Res 2006;14:117–26.
[21] Fedorova E, Zink D. Nuclear architecture and gene regulation. Biochim Biophys Acta 2008;1783:2174–84.
[22] Fraser P, Bickmore W. Nuclear organization of the genome and the potential for gene regulation. Nature 2007;447:413–7.
[23] Jiang C, Pugh BF. Nucleosome positioning and gene regulation: advances through genomics. Nat Rev Genet 2009;10:161–72.
[24] Kouzarides T. Chromatin modifications and their function. Cell 2007;128:693–705.
[25] Arnosti DN, Kulkarni MM. Transcriptional enhancers: intelligent enhanceosomes or flexible billboards? J Cell Biochem 2005;94:890–8.
[26] Thanos D, Maniatis T. Virus induction of human IFN beta gene expression requires the assembly of an enhanceosome. Cell 1995;83:1091–100.
[27] Walters MC, Fiering S, Eidemiller J, Magis W, Groudine M, Martin DI. Enhancers increase the probability but not the level of gene expression. Proc Natl Acad Sci USA 1995;92:7125–9.
[28] Fiering S, Whitelaw E, Martin DI. To be or not to be active: the stochastic nature of enhancer action. Bioessays 2000;22:381–7.
[29] Rossi FM, Kringstein AM, Spicher A, Guicherit OM, Blau HM. Transcriptional control: rheostat converted to on/off switch. Mol Cell 2000;6:723–8.
[30] Raser JM, O'Shea EK. Noise in gene expression: origins, consequences, and control. Science 2005;309:2010–3.
[31] Raj A, van Oudenaarden A. Nature, nurture, or chance: stochastic gene expression and its consequences. Cell 2008;135:216–26.
[32] Ptashne M. Gene regulation by proteins acting nearby and at a distance. Nature 1986;322:697–701.
[33] Wang JC, Giaever GN. Action at a distance along a DNA. Science 1988;240:300–4.
[34] Dekker J, Rippe K, Dekker M, Kleckner N. Capturing chromosome conformation. Science 2002;295:1306–11.
[35] Cleard F, Moshkin Y, Karch F, Maeda RK. Probing long-distance regulatory interactions in the *Drosophila melanogaster* bithorax complex using Dam identification. Nat Genet 2006;38:931–5.
[36] Jing H, Vakoc CR, Ying L, Mandat S, Wang H, Zheng X, et al. Exchange of GATA factors mediates transitions in looped chromatin organization at a developmentally regulated gene locus. Mol Cell 2008;29:232–42.
[37] Liu Z, Garrard WT. Long-range interactions between three transcriptional enhancers, active Vkappa gene promoters, and a 3′ boundary sequence spanning 46 kilobases. Mol Cell Biol 2005;25:3220–31.
[38] Tolhuis B, Palstra RJ, Splinter E, Grosveld F, de Laat W. Looping and interaction between hypersensitive sites in the active beta-globin locus. Mol Cell 2002;10:1453–65.
[39] Ronshaugen M, Levine M. Visualization of trans-homolog enhancer-promoter interactions at the Abd-B Hox locus in the *Drosophila embryo*. Dev Cell 2004;7:925–32.
[40] Carter D, Chakalova L, Osborne CS, Dai YF, Fraser P. Long-range chromatin regulatory interactions in vivo. Nat Genet 2002;32:623–6.
[41] Hatzis P, Talianidis I. Dynamics of enhancer–promoter communication during differentiation-induced gene activation. Mol Cell 2002;10:1467–77.
[42] Chubb JR, Trcek T, Shenoy SM, Singer RH. Transcriptional pulsing of a developmental gene. Curr Biol 2006;16:1018–25.
[43] Iborra FJ, Pombo A, Jackson DA, Cook PR. Active RNA polymerases are localized within discrete transcription "factories' in human nuclei. J Cell Sci 1996;109(Pt 6):1427–36.
[44] Osborne CS, Chakalova L, Brown KE, Carter D, Horton A, Debrand E, et al. Active genes dynamically colocalize to shared sites of ongoing transcription. Nat Genet 2004;36:1065–71.
[45] Ragoczy T, Bender MA, Telling A, Byron R, Groudine M. The locus control region is required for association of the murine beta-globin locus with engaged transcription factories during erythroid maturation. Genes Dev 2006;20:1447–57.
[46] Ben-Tabou de-Leon S, Davidson EH. Gene regulation: gene control network in development. Annu Rev Biophys Biomol Struct 2007;36:191.
[47] Yuh CH, Davidson EH. Modular cis-regulatory organization of Endo16, a gut-specific gene of the sea urchin embryo. Development 1996;122:1069–82.
[48] Yuh CH, Moore JG, Davidson EH. Quantitative functional interrelations within the cis-regulatory system of the *S. purpuratus* Endo16 gene. Development 1996;122:4045–56.
[49] Kulkarni MM, Arnosti DN. Information display by transcriptional enhancers. Development 2003;130:6569–75.
[50] Kulkarni MM, Arnosti DN. cis-Regulatory logic of short-range transcriptional repression in *Drosophila melanogaster*. Mol Cell Biol 2005;25:3411–20.
[51] Levine M, Davidson EH. Gene regulatory networks for development. Proc Natl Acad Sci USA 2005;102:4936–42.
[52] Stein LD. Human genome: end of the beginning. Nature 2004;431:915–6.
[53] Messina DN, Glasscock J, Gish W, Lovett M. An ORFeome-based analysis of human transcription factor genes and the construction of a microarray to interrogate their expression. Genome Res 2004;14:2041–7.
[54] van Nimwegen E. Scaling laws in the functional content of genomes. Trends Genet 2003;19:479–84.
[55] Gibson G. Epistasis and pleiotropy as natural properties of transcriptional regulation. Theor Popul Biol 1996;49:58–89.
[56] Pennacchio LA, Rubin EM. Genomic strategies to identify mammalian regulatory sequences. Nat Rev Genet 2001;2:100–9.
[57] Vavouri T, Elgar G. Prediction of cis-regulatory elements using binding site matrices—the successes, the failures and the reasons for both. Curr Opin Genet Dev 2005;15:395–402.
[58] Wasserman WW, Sandelin A. Applied bioinformatics for the identification of regulatory elements. Nat Rev Genet 2004;5:276–87.
[59] de la Calle-Mustienes E, Feijoo CG, Manzanares M, Tena JJ, Rodriguez-Seguel E, Letizia A, et al. A functional survey of the enhancer activity of conserved non-coding sequences from vertebrate Iroquois cluster gene deserts. Genome Res 2005;15:1061–72.
[60] Nobrega MA, Ovcharenko I, Afzal V, Rubin EM. Scanning human gene deserts for long-range enhancers. Science 2003;302:413.
[61] Pennacchio LA, Ahituv N, Moses AM, Prabhakar S, Nobrega MA, Shoukry M, et al. In vivo enhancer analysis of human conserved non-coding sequences. Nature 2006;444:499–502.

[62] Boyle AP, Davis S, Shulha HP, Meltzer P, Margulies EH, Weng Z, et al. High-resolution mapping and characterization of open chromatin across the genome. Cell 2008;132:311–22.
[63] Shi B, Guo X, Wu T, Sheng S, Wang J, Skogerbo G, et al. Genome-scale identification of *Caenorhabditis elegans* regulatory elements by tiling-array mapping of DNase I hypersensitive sites. BMC Genomics 2009;10:92.
[64] Boyer LA, Lee TI, Cole MF, Johnstone SE, Levine SS, Zucker JP, et al. Core transcriptional regulatory circuitry in human embryonic stem cells. Cell 2005;122:947–56.
[65] Kidder BL, Yang J, Palmer S. Stat3 and c-Myc genome-wide promoter occupancy in embryonic stem cells. PLoS One 2008;3:e3932.
[66] Kim TH, Abdullaev ZK, Smith AD, Ching KA, Loukinov DI, Green RD, et al. Analysis of the vertebrate insulator protein CTCF-binding sites in the human genome. Cell 2007;128:1231–45.
[67] Robertson G, Hirst M, Bainbridge M, Bilenky M, Zhao Y, Zeng T, et al. Genome-wide profiles of STAT1 DNA association using chromatin immunoprecipitation and massively parallel sequencing. Nat Methods 2007;4:651–7.
[68] Visel A, Blow MJ, Li Z, Zhang T, Akiyama JA, Holt A, et al. ChIP-seq accurately predicts tissue-specific activity of enhancers. Nature 2009;457:854–8.
[69] Boyer LA, Plath K, Zeitlinger J, Brambrink T, Medeiros LA, Lee TI, et al. Polycomb complexes repress developmental regulators in murine embryonic stem cells. Nature 2006;441:349–53.
[70] Barski A, Cuddapah S, Cui K, Roh TY, Schones DE, Wang Z, et al. High-resolution profiling of histone methylations in the human genome. Cell 2007;129:823–37.
[71] Bernstein BE, Humphrey EL, Erlich RL, Schneider R, Bouman P, Liu JS, et al. Methylation of histone H3 Lys 4 in coding regions of active genes. Proc Natl Acad Sci USA 2002;99:8695–700.
[72] Heintzman ND, Hon GC, Hawkins RD, Kheradpour P, Stark A, Harp LF, et al. Histone modifications at human enhancers reflect global cell-type-specific gene expression. Nature 2009;459:108–12.
[73] Wang Z, Zang C, Rosenfeld JA, Schones DE, Barski A, Cuddapah S, et al. Combinatorial patterns of histone acetylations and methylations in the human genome. Nat Genet 2008;40:897–903.
[74] Yuan GC, Liu YJ, Dion MF, Slack MD, Wu LF, Altschuler SJ, et al. Genome-scale identification of nucleosome positions in *S. cerevisiae*. Science 2005;309:626–30.
[75] Zhang Y, Shin H, Song JS, Lei Y, Liu XS. Identifying positioned nucleosomes with epigenetic marks in human from ChIP-Seq. BMC Genomics 2008;9:537.
[76] Dostie J, Dekker J. Mapping networks of physical interactions between genomic elements using 5C technology. Nat Protoc 2007;2:988–1002.
[77] Dostie J, Richmond TA, Arnaout RA, Selzer RR, Lee WL, Honan TA, et al. Chromosome conformation capture carbon copy (5C): a massively parallel solution for mapping interactions between genomic elements. Genome Res 2006;16:1299–309.
[78] Simonis M, Klous P, Splinter E, Moshkin Y, Willemsen R, de Wit E, et al. Nuclear organization of active and inactive chromatin domains uncovered by chromosome conformation capture-on-chip (4C). Nat Genet 2006;38:1348–54.
[79] Wurtele H, Chartrand P. Genome-wide scanning of HoxB1-associated loci in mouse ES cells using an open-ended Chromosome Conformation Capture methodology. Chromosome Res 2006;14:477–95.
[80] Zhao Z, Tavoosidana G, Sjolinder M, Gondor A, Mariano P, Wang S, et al. Circular chromosome conformation capture (4C) uncovers extensive networks of epigenetically regulated intra- and interchromosomal interactions. Nat Genet 2006;38:1341–7.
[81] Birney E, Stamatoyannopoulos JA, Dutta A, Guigo R, Gingeras TR, Margulies EH, et al. Identification and analysis of functional elements in 1% of the human genome by the ENCODE pilot project. Nature 2007;447:799–816.
[82] Muller F, Blader P, Strahle U. Search for enhancers: teleost models in comparative genomic and transgenic analysis of cis regulatory elements. Bioessays 2002;24:564–72.
[83] Amsterdam A, Becker TS. Transgenes as screening tools to probe and manipulate the zebrafish genome. Dev Dyn 2005;234:255–68.
[84] Allende ML, Manzanares M, Tena JJ, Feijoo CG, Gomez-Skarmeta JL. Cracking the genome's second code: enhancer detection by combined phylogenetic footprinting and transgenic fish and frog embryos. Methods 2006;39:212–9.
[85] Khokha MK, Loots GG. Strategies for characterising cis-regulatory elements in *Xenopus*. Brief Funct Genomic Proteomic 2005;4:58–68.
[86] Uchikawa M, Ishida Y, Takemoto T, Kamachi Y, Kondoh H. Functional analysis of chicken Sox2 enhancers highlights an array of diverse regulatory elements that are conserved in mammals. Dev Cell 2003;4:509–19.
[87] Glasser SW, Eszterhas SK, Detmer EA, Maxfield MD, Korfhagen TR. The murine SP-C promoter directs type II cell-specific expression in transgenic mice. Am J Physiol Lung Cell Mol Physiol 2005;288:L625–32.
[88] Meissner A, Eminli S, Jaenisch R. Derivation and manipulation of murine embryonic stem cells. Methods Mol Biol 2009;482:3–19.
[89] Masamizu Y, Ohtsuka T, Takashima Y, Nagahara H, Takenaka Y, Yoshikawa K, et al. Real-time imaging of the somite segmentation clock: revelation of unstable oscillators in the individual presomitic mesoderm cells. Proc Natl Acad Sci USA 2006;103:1313–8.
[90] Nagoshi E, Saini C, Bauer C, Laroche T, Naef F, Schibler U. Circadian gene expression in individual fibroblasts: cell-autonomous and self-sustained oscillators pass time to daughter cells. Cell 2004;119:693–705.
[91] Finkenstadt B, Heron EA, Komorowski M, Edwards K, Tang S, Harper CV, et al. Reconstruction of transcriptional dynamics from gene reporter data using differential equations. Bioinformatics 2008;24:2901–7.
[92] Taylor DL. Past, present, and future of high content screening and the field of cellomics. Methods Mol Biol 2007;356:3–18.
[93] Cohen AA, Geva-Zatorsky N, Eden E, Frenkel-Morgenstern M, Issaeva I, Sigal A, et al. Dynamic proteomics of individual cancer cells in response to a drug. Science 2008;322:1511–6.
[94] Sigal A, Milo R, Cohen A, Geva-Zatorsky N, Klein Y, Alaluf I, et al. Dynamic proteomics in individual human cells uncovers widespread cell-cycle dependence of nuclear proteins. Nat Methods 2006;3:525–31.
[95] Sigal A, Milo R, Cohen A, Geva-Zatorsky N, Klein Y, Liron Y, et al. Variability and memory of protein levels in human cells. Nature 2006;444:643–6.
[96] Megason SG, Fraser SE. Digitizing life at the level of the cell: high-performance laser-scanning microscopy and image analysis for in toto imaging of development. Mech Dev 2003;120:1407–20.
[97] Carvalho L, Heisenberg CP. Imaging zebrafish embryos by two-photon excitation time-lapse microscopy. Methods Mol Biol 2009;546:273–87.
[98] Passamaneck YJ, Di Gregorio A, Papaioannou VE, Hadjantonakis AK. Live imaging of fluorescent proteins in chordate embryos: from ascidians to mice. Microsc Res Technol 2006;69:160–7.