View metadata, citation and similar papers at core.ac.uk
brought to you by CORE
provided by Aberdeen University Research Archive

Biology **2013**, *2*, 64-84; doi:10.3390/biology2010064

OPEN ACCESS

biology

ISSN 2079-7737

www.mdpi.com/journal/biology

Review

Understanding the Dynamics of Gene Regulatory Systems; Characterisation and Clinical Relevance of *cis*-Regulatory Polymorphisms

Philip Cowie, Ruth Ross and Alasdair MacKenzie *

School of Medical Sciences, Institute of Medical Sciences, University of Aberdeen, Aberdeen, Scotland, AB25 2ZD, UK; E-Mails: p.cowie@abdn.ac.uk (P.C.); r.ross@abdn.ac.uk (R.R.)

* Author to whom correspondence should be addressed; E-Mail: alasdair.mackenzie@abdn.ac.uk; Tel.: +44-122-476-7380; Fax: +44-122-476-7399.

Received: 1 November 2012; in revised form: 21 December 2012 / Accepted: 4 January 2013 / Published: 9 January 2013

Abstract: Modern genetic analysis has shown that most polymorphisms associated with human disease are non-coding. Much of the functional information contained in the non-coding genome consists of *cis*-regulatory sequences (CRSs) that are required to respond to signal transduction cues that direct cell specific gene expression. It has been hypothesised that many diseases may be due to polymorphisms within CRSs that alter their responses to signal transduction cues. However, identification of CRSs, and the effects of allelic variation on their ability to respond to signal transduction cues, is still at an early stage. In the current review we describe the use of comparative genomics and experimental techniques that allow for the identification of CRSs building on recent advances by the ENCODE consortium. In addition we describe techniques that allow for the analysis of the effects of allelic variation and epigenetic modification on CRS responses to signal transduction cues. Using specific examples we show that the interactions driving these elements are highly complex and the effects of disease associated polymorphisms often subtle. It is clear that gaining an understanding of the functions of CRSs, and how they are affected by SNPs and epigenetic modification, is essential to understanding the genetic basis of human disease and stratification whilst providing novel directions for the development of personalised medicine.

Keywords: gene regulation; *cis*-regulatory variation; non-coding DNA; chromatin; signal transduction; drug response stratification; cell specificity; context dependency; ENCODE consortium

1. Introduction

The importance of gene regulation cannot be overstated; the evolution of complex multicellular organisms whose cells possess identical genomes, yet exhibit phenotypic and functional diversity, coincides with the evolution of complex gene regulatory systems capable of controlling differential gene expression [1,2]. Further, multicellular life must have the ability to regulate its transcriptome in response to extracellular signals from the environment, and surrounding cells if it is to develop, adapt and survive. To this end eukaryotes have evolved a repertoire of extracellular signals and receptors which activate diverse signal transduction pathways ultimately resulting in the regulation of specific genes through recruitment of transcription factor (TF) complexes [3]. Central to this process in many genes is the involvement of *cis*-regulatory sequences (CRSs); non-coding functional regions of DNA which mediate TF binding and regulate transcription [4].

Interest in *cis*-regulatory sequences has intensified since the human genome sequence was first mapped [5,6] and subsequently shown to only contain 20,000–25,000 protein coding genes [7]; far fewer than was previously anticipated, leaving ~97% of the genome with no predicted coding function. Consequently, comparative genomics [8,9] has been used to demonstrate that conservation of non-coding DNA regions between evolutionarily divergent species is a powerful tool for the prediction of *cis*-regulatory sequences [10–13] including promoter and enhancer regions, insulators and locus control regions (reviewed; [14]). More recently, the international consortium ENCODE published a series of papers highlighting that 80.4% of the human genome functions in some form of biological process, and conservative estimates suggest that there may be 4.5 times more functional information within the genome than that which encodes proteins [15].

Given the fundamental role CRSs play in gene regulation, and the necessity for precise regulation to orchestrate correct development and function, it comes as no surprise that variation within CRSs is emerging as a major source of disease susceptibility in human populations [16]. Meta-analysis of multiple genome wide association (GWA) studies [17,18] indicates that 88% of disease-associated single nucleotide polymorphisms (SNPs) lie in intronic or intergenic regions [19]. More specifically, 71% of disease-associated SNPs (including SNPs in linkage disequilibrium) lie in non-coding regulatory regions identified by ENCODE [15]. Hence polymorphisms of non-coding regulatory regions are disproportionately linked to human disease likely through mechanisms involving aberrant gene regulation. In principle, these gene regulation aberrations will not only impact an individual's susceptibility to disease but also their response to drug treatments as a result of underlying biochemical differences.

A significant challenge for molecular genetics is therefore to: (1) determine the tissue-specific nature of *cis*-regulatory relationships within 3 dimensional paradigms; (2) locate interacting partners of CRSs (3) apply computational and experimental approaches to understand how they function in

regulatory networks; (4) evaluate the effect of endogenous CRS variation in the context of cellular signalling and (5) determine the role that CRS variation plays in human disease and drug response stratification.

2. The Importance of Non-Coding DNA

As a prerequisite to understanding developments within the field of CRS research we have outlined some basic aspects of eukaryotic transcription with respect to transcriptional machinery and *cis*-regulatory functions (Figure 1). To appreciate the value of studying non-coding DNA, and its role in gene regulation, we must evaluate its importance with respect to evolution and development and determine its pathological potential.

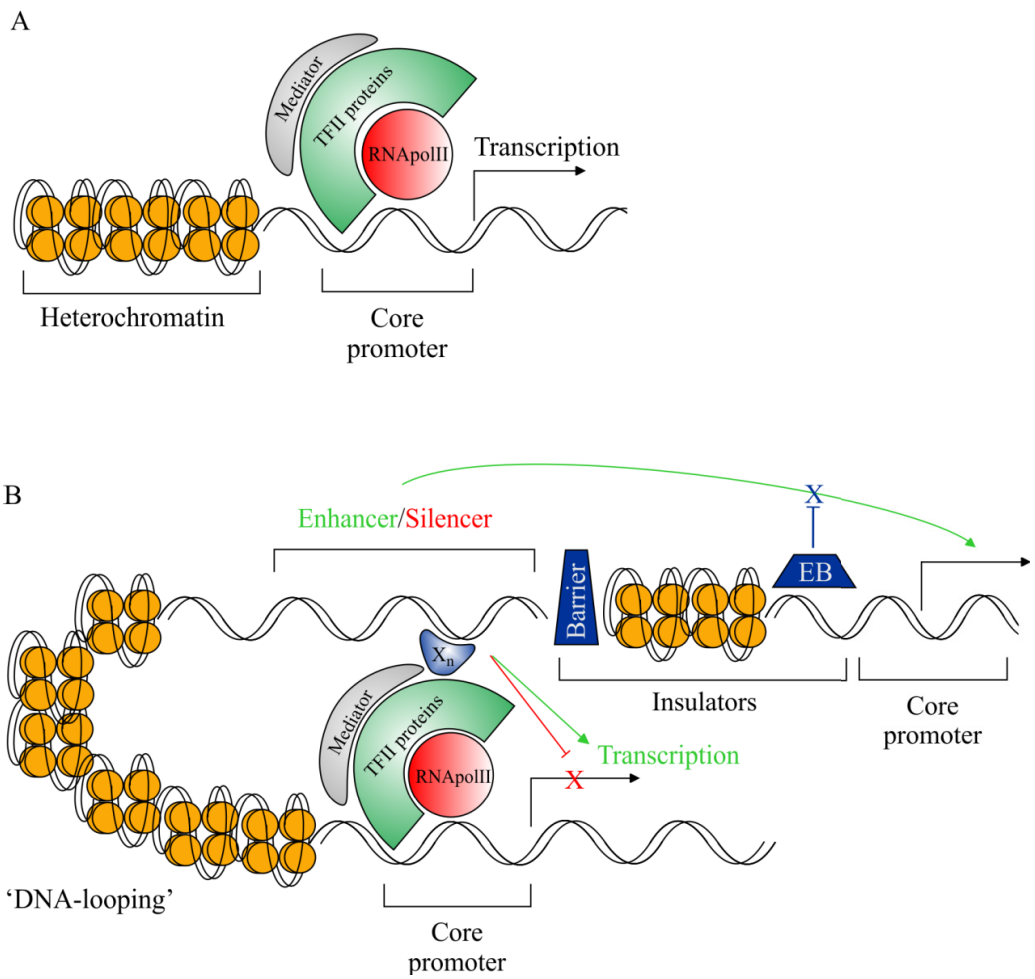
2.1. cis-Regulatory Sequences Have Shaped Human Evolution and Development

A critical feature of CRSs is the modular nature by which they regulate gene expression [20]. Thus tissue-specific (spatial) and developmental stage-specific (temporal) gene expression can be controlled by specific CRS-mediated TF-complex binding. The apolipoprotein E (*APOE*) locus is a well characterised example of a gene that is regulated by multiple flanking CRSs that direct differential expression to liver cells [21,22] or skin cells [23,24], or astrocytes, macrophages and adipocytes [24,25]. Consequently, the effects of mutations in CRSs can be limited to particular cell types or developmental stages making them less pleiotropic than coding mutations. The relative lack of pleiotrophism makes CRSs strong candidates for driving evolution through mutation as well as inducing susceptibility to late onset disease. For example a CRS SNP located upstream of the *DARC* promoter, which codes a human receptor important for the reception of immune system signals [26–28], abolishes expression of the receptor in erythrocytes [29,30]. This SNP confers complete resistance to malaria [31,32] by preventing *Plasmodium spp.* parasites entry to erythrocytes due to the lack of the *DARC*-coded receptor [33,34]. Importantly, the SNP has little or no deleterious effects in other *DARC* expression domains. Another example is *HACNS1*, a highly conserved non-coding sequence, which has been identified to contain human-specific polymorphisms that result in the differential limb patterning observed between humans and non-human primates [35].

2.2. cis-Regulatory Sequences are Implicated in Human Pathologies

With respect to human pathologies it was shown that a non-coding regulatory SNP located near the α-globin gene cluster creates a new TF consensus sequence for GATA-1 augmenting the activation of the gene cluster and causing Thalassemia in affected individuals [36]. Further, very recent data concerning the transcription factor 7-like (*TCF7L2*) locus has utilised the results of GWA studies, identifying variation within the *TCF7L2* intronic regions as highly associated with risk of type 2 diabetes, and shown that the associated variation is located within a *cis*-regulatory region [37]. Moreover it has been discovered that Hirschsprung disease risk is associated with variation within an enhancer region of the receptor tyrosine kinase *RET* [38,39]. While coding mutations in *RET* were causative in a small portion of cases the authors also found that variation within a CRS of *RET* intron 1 resulted in a significant decrease in *RET* expression [38,39].

Figure 1. Graphic representation of eukaryotic transcriptional machinery. (**A**) Basal eukaryotic transcriptional machinery; members of the transcription factor II (TFII) family of proteins associate with RNA polymerase II (RNApolII) in an ordered manner to form the pre-initiation complex. The core promoter region, containing transcription factor binding sites (TFBS) and the transcriptional start site, is bound by the pre-initiation complex and RNApolII is directed to begin transcription of target genes. (**B**) *cis*-regulatory DNA sequences modulating eukaryotic transcription. Distant *cis*-regulatory sequences (CRSs), such as enhancers and silencers (located up to 1Mbp from the target promoter), associate with additional TFs (X_n) and form indirect interactions with the target promoter. Subsequently, transcriptional outputs are modified depending on the nature of the associated CRS; increases in transcript quantity (enhancer function—green arrows) or reduction/abolition of transcription (silencer function—red T-bars). In order for enhancer/silencer sequences to interact with target promoters DNA must be modified to "loop out" the interspaced DNA. Other recognised classes of regulatory sequences include insulators: Barrier-form insulators prevent chromatin condensation from repressing active regulatory regions setting up regulatory boundaries; Enhancer-blocking (EB) insulators maintain the specificity of CRS interactions by blocking regulatory sequences from impinging on neighbouring genes. Finally, locus control regions are described as regions containing multiple CRSs, they function in concert to confer correct temporal and/or spatial specificity of the target gene.



2.3. Rationale for cis-Regulatory Sequence Research

It is clear from these examples that CRSs play a vital role in evolution, development and human disease, indeed preeminent conjectures concerning the importance of CRSs to evolution and development through gene regulation were made ~40 years ago by Jacob and Monod [40], Britten and Davidson [41,42] and King and Wilson [43]. However, despite the wealth of evidence which has been mounting in recent years CRSs remain relatively poorly understood. This is due in part to decades of exon-focused research, which by comparison has more easily definable and testable entities. Intriguingly, computational analysis has shown that 87% of the conserved genome between humans and mice (>70% identity over 100 bp) is non-coding which highlights the potentially massive pool of unexamined functional DNA present within the genome [44]. One of the major challenges to examining CRSs is their identification and publication of the human genome sequence [5,6] has proved enormously helpful in addressing this issue. Moreover the collaborative efforts of the ENCODE project has marked a huge step towards elucidating the functional regulatory landscape of the human genome through systematic CRS identification using a number of well characterised computational and experimental paradigms which we have summarised below [15].

3. *cis*-Regulatory Sequence Identification—Comparative Genomics

Comparative genomics has emerged as a powerful tool for the discovery of CRSs and relies on the basic principle that regulatory functional sequences are under purifying selection and cross-species sequence comparisons can highlight this conservation. It is important to note that, while many CRSs regulate target gene expression through TF binding and recruitment to promoters, predicted TF binding motifs do not represent reliable candidate sequence motifs for the identification of CRSs due to their high degeneracy and wide-spread distribution in the genome. Instead we may broadly consider two approaches assessing genome-wide sequence conservation: evolutionary distant species comparisons and evolutionarily related species comparisons.

3.1. Evolutionary Distant Species Comparisons

In the first case, the availability of genome sequences from birds, fish and reptiles allow researchers to identify putative CRSs with functions critical to vertebrate development by way of pair-wise comparison to mammalian genomes. This approach has been highly successful for identifying CRSs, even prior to the availability of genome sequences for so many vertebrates [45], such as those involved in the tissue-specific expression of embryogenesis genes related to: cardiac development [46]; limb patterning [13,47,48] and brain development [13,48,49]. Indeed a common feature of CRSs identified by this method is that they are non-randomly located in gene deserts [12,50] adjacent to genes with developmental functions [49].

Unquestionable then is the potential importance of distant comparative approaches, clearly capable of locating vertebrate developmental gene-related CRSs, but there are a number of important caveats to consider. Firstly, altering the parameters of this strategy has been shown to cause estimations of CRS numbers to vary between 1,400 [49] and 5,700 [51], suggesting that the method is insensitive and misses many CRSs since these estimations are an order of magnitude lower than the predicted number

of human genes [7]. Additionally, such "deep" conservation is likely to be the result of a shared biological process between the species under comparison; hence this method is unable to identify CRSs involved in processes which evolved subsequent to the divergence of the species in question. Finally, if such comparisons are used between less divergent species such as human-rodent the relaxed parameters (>70% identity over 100 bp) will throw up large numbers of false positive results.

3.2. Evolutionary Related Species Comparisons

In the second case, researchers can identify CRSs more likely related to higher vertebrate health by comparing less distant species with more stringent conservations parameters. Specifically, typical conservation parameters between human-chicken or human-frog comparisons are >70% identity over 100 bp. However Bejerano *et al*., (2004) explored the use of human-rodent comparisons at parameters of 100% identity over 200 bp [52]. Unsurprisingly, they found a smaller set of putative CRSs as compared to Woolfe *et al*., (250 [52] and 1,400 [49] respectively), however investigation of some of these "ultra-conserved" sequences has proved, in principle, that the method is capable of identifying modulators of gene transcription [48,53,54]. Interestingly, the method was further assessed in combination with human-fugu comparisons, whereby the authors were able to predict enhancer activity of sequences very successfully (~60% of identified sequences showed enhancer capacity) by coupling "deep" conservation (human-fugu) with ultra-conservation (human-rodent, described above) [13]. However, subsequent investigation into ultra-conservation comparisons has lead some researchers to conclude that overall sequence conservation, as opposed to ultra-conservation, is a good predictor of CRSs functionality [55].

Consequently, ultra-conservation comparison techniques do suffer as a product of their design; they are likely to identify only small subsets of CRSs, and not only miss numerous other CRSs but also cannot be utilised as a large-scale prediction method [11]. Further, the parameters required to fulfil the "ultra-conservation" label mean that many predicted CRSs are also identified by evolutionary divergent comparisons [11]. Likely, even with the manifestation of well characterised highly accurate computation models to predict CRSs, we must acknowledge that computational data alone cannot provide extensive evidence as to biological function. Consequently, parallel experimental approaches have been developed to complement computational prediction of CRSs to good effect.

4. *cis*-Regulatory Sequence Identification—Experimental Approaches

In response to the stated drawbacks of computational conservation-based CRS prediction methods, well developed strategies now exist which allow researchers to identify CRSs in a conservation-independent manner (reviewed [56]). One particular reason for this is the observation that ~50% of experimentally validated CRSs do not show sequence conservation [57], and, depending on the tissue type under investigation, enhancers can been significantly non-conserved [58].

4.1. Transcriptional Associations: Chromatin Immunoprecipitation Techniques

A number of the experimental paradigms for CRS identification originate from the exploitation of an indirect physical association between the CRS and its target promoter via TF-complexes and

transcriptional co-activators such as p300 [14,59]. Researchers begin determining these interactions by cross-linking chromatin with formaldehyde, capturing endogenous DNA-protein interactions within the nucleus, and subsequently shearing it into smaller pieces by sonication or enzymatic digest. Samples are enriched for DNA showing an association with specific TF's, co-activators or histone-modifications associated with enhancers (e.g., H3K4me1) or silencers by immunoprecipitation with antibodies specific to the TF, co-activator or histone-modification. The principle technique is called chromatin immunoprecipitation (ChIP), and the resultant enriched samples can be analysed by hybridisation to microarrays (ChIP-chip) [60,61] or by deep sequencing the entire enriched DNA sample (ChIP-seq) [62,63]. Results are analysed for DNA sequences which are over represented in the enriched samples, demonstrating that they are likely associated with TF's and/or co-activators and therefore involved in transcriptional regulation. This method can also be used on restricted cell populations by initially micro-dissecting specific tissue regions, ChIP results then provide an immediate indication of the tissue-specific activity of identified CRSs [64].

4.2. Active Chromatin Signatures: DNaseI Hypersensitivity and Formaldehyde-Assisted Identification of Regulatory Elements

Another approach to discovering CRSs employs the fact that functional non-coding sequences are associated with "active" chromatin conformations, induced through TF binding, making these stretches of DNA more sensitive to DNase I activity [65]. DNase I hypersensitivity (DHS) approaches can again be combined with microarrays or deep sequencing to identify regions of DNA with an "open" chromatin structure indicative of TF binding and presumed regulatory potential [66,67]. Of particular interest, this technique is capable of detecting hypersensitivity differences which result from polymorphisms within the genetic code, highlighting the potential for polymorphic variation in CRSs to impact gene regulation and by extension disease [68]. Further, DHS sites are known to be enriched for non-coding disease-associated genetic variants and commonly map to disease-associated loci [69]. Consequently, DHS data can be highly predictive of disease-associated regulatory networks including causative CRSs and interacting proteins [69,70]. FAIRE (formaldehyde-assisted identification of regulatory elements) is similar to the DNase I hypersensitivity technique, in that it exploits open chromatin's susceptibility to mechanical shearing after formaldehyde cross-linking to non-selectively identify functional regulatory DNA regions [71]. Both of these methods can provide researchers with fast, cost effective results. Combined with well organised comparative genomic analysis CRSs can often be inferred providing a reliable basis for further study.

4.3. Chromosome Interactions: Chromosome Conformation Capture Strategies

The above techniques identify either DNA which associates with transcriptional regulatory proteins (ChIP) or DNA which is putatively active in the binding of transcriptional regulatory proteins (DNase, FAIRE), but neither is able to remote chromatin interactions nor do they provide information relating to the 3-dimensional structure of the genome. Development of chromosome conformation capture (3C) [72], and derived techniques (4C, 5C and Hi-C [73] (see [74] and [75] for review)), overcome this hurdle on the premise that CRSs and promoters must indirectly interact across large regions of the genome. A consequence of these long distance interactions is that, following cross-linking and

shearing, DNA can be covalently ligated to sequences in close 3-dimensional proximity (proximity ligation). The experimental output then identifies interactions between DNA sequences, which may normally be separated by up to 1 Mb, being sequenced together more frequently as a result of a 3D chromatin interaction. A drawback of 3C, 4C and 5C is that they are all biased towards a particular locus, or set of loci, under investigation.

Conversely, Hi-C is both genome-wide and unbiased in its identification of long distance chromatin interactions; by incorporating biotinylated residues into the fragment ends after digestion of cross-linked DNA streptavidin can be used to select for sequences in close proximity which are subsequently analysed [73]. Further advancements towards the functional annotation of the genome have resulted in the development of the technique ChIA-PET (chromatin interaction analysis by paired-end tag sequencing) [76,77]. Similar in methodology to Hi-C, but requiring an interacting protein for sample enrichment by immunoprecipitation before proximity ligation, ChIA-PET is seen as a promising alternative to ChIP-Seq since it is capable of identifying both TFBSs and chromatin structure within purified sequences [77,78].

4.4. Towards a Map of the Genome's Regulatory Landscape: The ENCODE Consortium

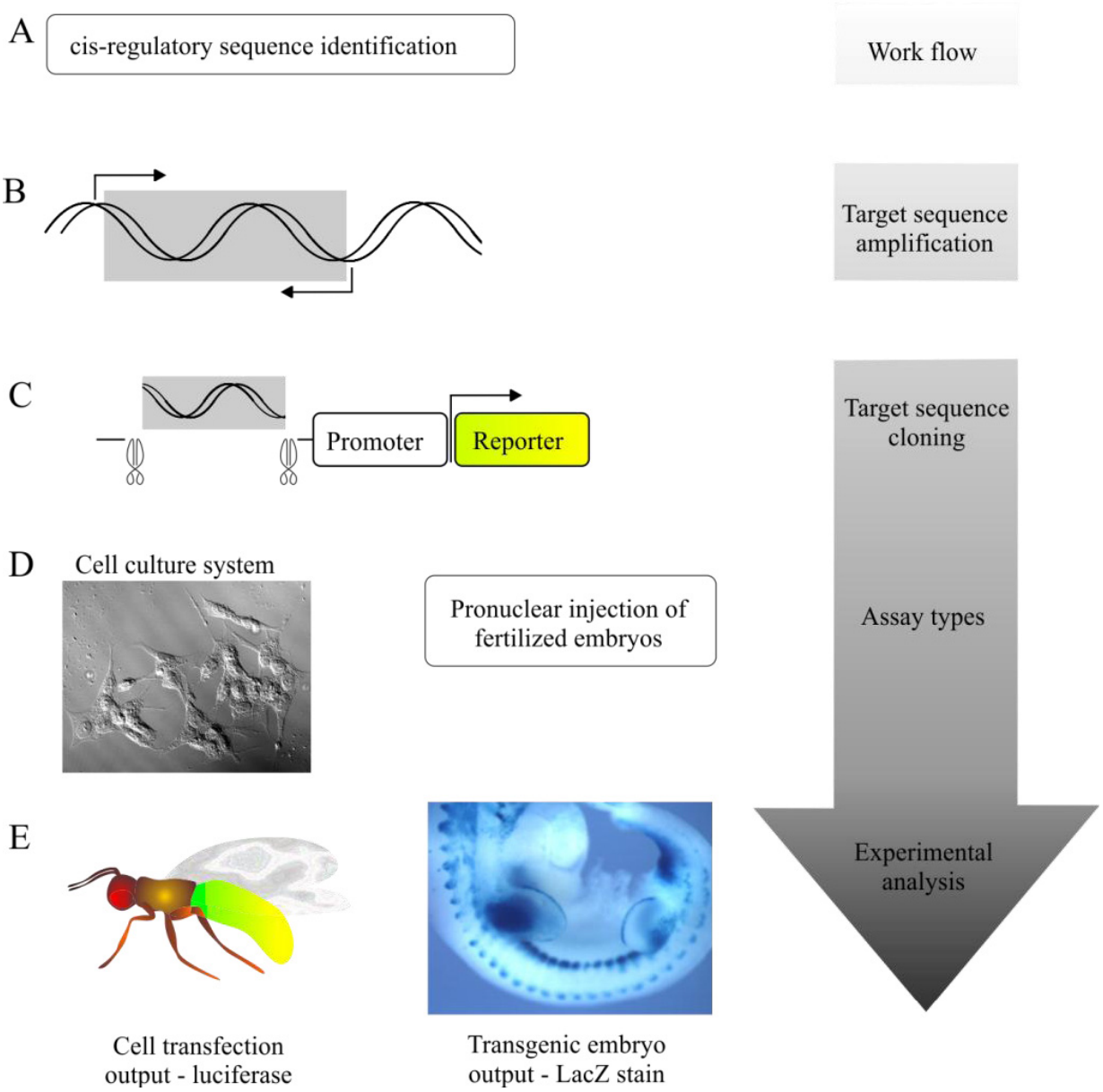
The ENCODE consortium represents an international project aimed at identifying all the functional elements in the human genome using a combination of computational and experimental approaches [15] (some of which are outlined above). Data generated by the project is available on the UCSC genome website [79,80]; customisable tracks can be selected to view chromatin modification signatures, DNase I hypersensitivity, FAIRE analysis, TF binding sites, transcriptional start sites and DNA methylation patterns for particular genomic regions within a number of different cell type. Consequently, ENCODE data is likely to represent the starting point for the majority of CRS investigations of the future; a vast database of the regulatory landscape of the genome will provide researchers with immediate indications of the regulatory capacity of selected regions. Further, work in progress by ENCODE to complete genome wide chromosome conformation maps will provide researchers with invaluable insights into long distance DNA sequences interactions.

However, we must highlight some caveats of ENCODE's three tiered cell type strategy [15]. The exclusion of many important primary cell types, such as neuronal cells, has undoubtedly resulted in many CRSs going undetected due to both the context dependent nature of CRSs and their inducibility by cellular signalling events (see: *A question of specificity?* for more information). This ultimately means that while ENCODE data at UCSC will serve as a platform for much CRS research the lack of positive functional information for many highly conserved sequences does not yet persuasively indicate that they are not regulatory but that the particular cell types or specific stimuli used to ascribe functionality have yet to be ascertained.

5. Analysis of *cis*-Regulatory Sequences

Two standard approaches used to evaluate putative CRSs are transgenic animal-based reporter gene assays and cell-based reporter gene assays. By providing qualitative and quantitative information (respectively) about CRSs of interest these techniques are widely used in the confirmation of putative regulatory sequences. A schematic representation of CRS research workflow summarises how Sections 3, 4 and 5 are commonly implemented (Figure 2).

Figure 2. General experimental workflow of *cis*-regulatory sequence studies. (**A**) Numerous well characterised methods for CRS identification exist including computational and experimental approaches (described in main text). (**B**) Identified target sequences (boxed—grey) are reliably amplified via polymerase chain reaction (PCR) using specific primers (arrows). (**C**) Target sequences (putative CRSs) are cloned into a variety of reporter plasmid constructs, including luciferase, *LacZ* and fluorescent protein derivatives (e.g., *GFP*). Typically reporter plasmids are sequenced to ensure sequence integrity. (**D**) Reporter plasmids may be introduced to cell culture-type systems by transfection or into animal embryos by cytoplasmic or pronuclear injection. (**E**) Depending on the assay type a number of experimental outputs are obtainable: cell culture assays can provide quantitative analysis of target CRSs via luminosity readings (e.g., luciferase) and are particularly useful for pharmacological studies (see Figure 3); animal/embryo studies can provide qualitative explanations of where and when the target CRS is active during development.



5.1. Transgenic Animal Reporter Assays

Using analysis of transgenic animals the CRS of interest is typically cloned upstream of a reporter gene such as *LacZ* [81] or GFP, and the resultant construct is injected into fertilized animal embryos typically derived from species such as zebrafish, Xenopus, chicken or mouse. Subsequently, animals containing the construct are assessed for β-galactosidase activity via X-Gal staining or *GFP* expression with fluorescent microscopes. This method provides the chance to assess the ability of the CRS of interest to drive tissue-specific expression of the reporter gene; a central requirement of CRSs in gene regulation.

Transgenic analysis is considered by many researchers to represent the "gold standard" for confirming the tissue specificity of a candidate CRS. A number of hugely successful examples of its use exist [13,48,49,55], in particular Pennacchio and colleagues examined 167 putative CRSs, identified through comparative genomics, and established that 45% of the candidate sequences supported tissue specific expression of *LacZ* in developing mouse embryos [13]. Indeed the majority of deeply conserved CRSs identified to date function in early development [35], and consequently *LacZ* expression is often assessed in embryonic mice [13]. Within our lab CRSs have also been tested for tissue-specific expression in adult mice where our focus relates to their impact in adult neuronal gene regulation as opposed to developmental programmes [82].

Transgenic animal reporter assays alone are not sufficient to confirm the identity of a target sequence as a specific regulator of the proposed target gene. Subsequent in-situ hybridisation or immunohistological staining are required to demonstrate that putative CRS-driven *LacZ* expression co-localises with the endogenous transcript or endogenous protein. Further it is noteworthy that pronuclear injection creates a random insertion of reporter constructs, consequently at least 2 different transgenic lines with corroborating expression patterns are required.

5.2. Cell-Based Reporter Gene Assays

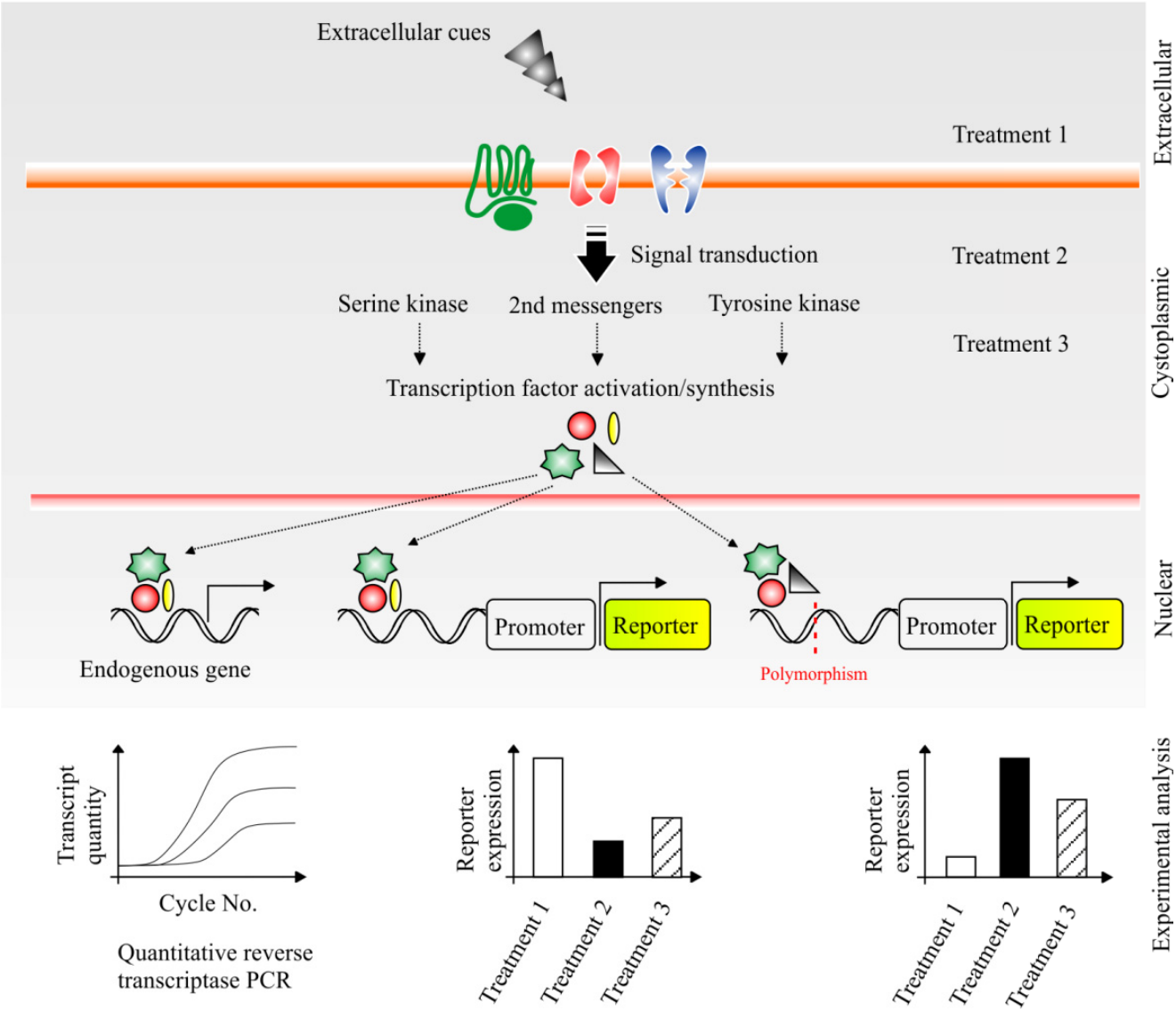
In addition to qualitative cell specific analysis it is useful to analyse the effects of SNPs or signal transduction cues on the quantitative activity of candidate CRSs. Putative CRSs are typically PCR amplified and cloned into reporter constructs, upstream of quantifiable reporter genes such as firefly luciferase. These constructs are then transfected into transformed cell lines or primary cell cultures. This method ultimately determines whether the CRS of interest is capable of eliciting a significant effect on the expression of the reporter gene, indicating its potential to function in gene regulation or to determine polymorphic effects.

We have used primary cell-based reporter gene assays to establish the presence of a highly conserved CRS (BE5.2) which functions as a silencer of the brain derived neurotropic factor (*BDNF*) promoter IV that plays a role in modulating mood [83]. Further, the quantitative nature of this method has been employed by our group to analyse the impact of allelic variation on CRS function; we have demonstrated significant allele-dependent changes in the activity of the *galanin* gene enhancer (GAL5.1) in primary hypothalamic neurons using luciferase reporter assays [82].

6. Beyond Identification: *cis*-Regulatory Sequence Characterisation

CRS characterisation studies are becoming increasingly pertinent in the wake of large scale, high-throughput, genome-wide identification projects (e.g., ENCODE). Vast CRS identification, even when coupled to the aforementioned methodologies, falls short of characterising the intricate signal transduction events which control CRS function. A molecular-level understanding of CRS functions is therefore essential if we hope to exploit them clinically and understand how regulatory polymorphisms impact susceptibility to many common human pathologies. The logic of CRS characterisation studies by pharmacological perturbations (as discussed below) is graphically represented (Figure 3).

Figure 3. Characterisation of *cis*-regulatory sequences.



6.1. Dissecting the Impact of Cellular Signalling

Due to its quantitative output cell-based reporter gene assays provide a means to investigate the cellular systems that modulate the activity of a given CRS through the manipulation of intracellular transduction pathways or ligand-receptor interactions by pharmacological means. The function of

CRSs depends on the availability and binding of TF's and co-activators [4], TF's are subject to regulation though mechanisms such as extracellular receptor activation, cytoplasmic serine kinase activation and intracellular proteolysis activity [84]. Consequently, cell cultures may be treated with a host of pharmacological agents to elucidate the precise biochemical requirements for CRS-mediated gene regulation. For example, we have previously demonstrated the ability of GAL5.1 to respond to PKC activation [82] and MAPkinase signalling as a necessary cue to the activation of a CRS contained within intron 2 of the CNR1 gene [85]. Similar work has been conducted by the Barolo laboratory as they set about defining the biochemical pathways which regulate the *Drosophila sparkling (spa)* enhancer [86]. Research of this nature is required to define the parameters of CRS function, without knowing the precise events which precede the involvement of a CRS in gene regulation we cannot begin to define their role in disease or produce clinical strategies based on their perturbation.

It is important to determine the relevance of pharmacological CRS manipulation to endogenous gene expression by assessing the effects of these pharmacological agents on the endogenous mRNA levels in parallel using quantitative reverse transcriptase PCR (qrtPCR). This combination of luciferase reporter gene assay and qrtPCR strengthens the argument for a CRS's capacity to regulate target gene expression. For example using qrtPCR we demonstrated the induction of the *TAC1* gene in primary dorsal root ganglia (DRG) cells by MAPkinase agonism or noxious stimulation by capsaicin. However, as assessed by luciferase reporter assay the *TAC1* promoter alone was unable to respond to these stimuli. Only by combining the *TAC1* promoter with a remote and highly conserved enhancer region called ECR2 could we induce a response from the *TAC1* promoter that was consistent with the response of the endogenous *TAC1* gene. This provides evidence of a requirement for enhancer-promoter synergy at the *TAC1* locus within DRG neurons following noxious induction [87,88].

Rapid development of CRS identification methods and collaborative efforts by the ENCODE consortium have placed an increasing emphasis on the characterisation of newly identified CRSs. Our schematic (Figure 3) shows the layers of a eukaryotic cell (from the extracellular to the nuclear) depicting a simplified cascade of cellular events from: extracellular cues binding to/transporting through cellular receptors; to intracellular transduction pathways; culminating in the production/activation of TFs and ultimately modulating gene transcription accordingly.

Using the previously discussed cell culture assays we highlight how pharmacological treatments aimed at specific cellular processes can potentially alter the activity of a CRS under investigation. For example, in the middle case treatment 2 has defined that the CRS in question is regulated by a particular signal transduction event. Further analysis would eventually determine the specific cellular conditions which precede the recruitment of this CRS to transcription of its target gene. Indeed, this scheme also highlights the potential of such an experimental paradigm to explore the impact of CRS polymorphisms (red line) on gene regulation. In the final case (right) the CRS polymorphism has altered the expression profile regulated by the CRS and perturbation with treatment 2 is now non-effective, a finding which may have clinical implications for individuals with this polymorphism (see: *cis-Regulatory Sequence Variation and Drug Response Stratification*). Finally, the first case (left) highlights the need for this experimental paradigm to include qrtPCR analysis in order to qualify that such changes in reporter gene quantities (either by treatments or by polymorphisms) are corroborated by changes in endogenous transcript quantity of the target gene. Demonstration of alterations in the

endogenous transcript quantity indicate the potential for alterations in biochemical events to be associated with the target genes product.

6.2. Embryonic Stem Cell Targeting

Despite high financial and time costs, embryonic stem cell targeting studies in mice are required to allow a full analysis of the role of CRSs in development and disease. Employing well defined strategies to knock-in or knock-out CRSs of interest, through the use of Cre-lox or Flp systems [89–93], researchers can define the effects of CRSs, and their polymorphisms on endogenous genes in an *in vivo* system that would be difficult to detect using the previously mentioned primary cell or transgenic strategies. In particular, the developmental role of a CRS may be assessed by knocking it out and analysing resultant changes in body plan, organ development or neuronal patterning. It is worth noting, however, that to date most CRSs are recognised as having modest effects on gene expression and therefore stable transgenic mouse models may only be used when the analysis of the effects of a SNP on CRS function is compelling and has been exhausted by the means described previously.

6.3. A Question of Specificity?

To date the majority of CRS studies utilising reporter constructs are conducted using exogenous promoters, and the use of transformed cell lines during analysis by reporter assay. Thus, a seriously underestimated but critical property of CRSs; namely, specificity in terms of promoter specificity and cell-type specificity is being overlooked in these cases.

The principle behind CRS-promoter specificity lies in the fact that CRSs may be located within or beyond neighbouring genes therefore the interaction (e.g., CRS-promoter) that takes place during CRS-mediated transcription relies on the CRS preferentially recognising its specific promoters. Indeed, there are examples of this phenomenon whereby the enhancer required to drive the expression of the *Sonic hedgehog (Shh)* gene in the developing limb bud is found in the intron of a gene lying 1 Mb from the *Shh* locus, called *Lmbr1*, which is also unaffected by its activity [94]. In addition, regulatory elements functioning in trans such as those found in *Drosophila* olfactory receptor genes serve as further evidence of this principle [95]. Whether CRS-promoter interactions are controlled and maintained by levels of chromatin flexibility [96], chromosomal location with the nucleus [97–100], the interaction of TFs and chromatin remodelling complexes [100], or perhaps a combination of these and undiscovered mechanisms does not alter the principle that CRS-promoter interactions must be specific for the appropriate regulation of their associated genes.

CRS specificity to particular cell types is well documented and a defining feature of their mode of action. Hence experimental approaches aimed at defining the impact of a CRS and/or endogenous CRS variation should also consider the impact that different cell types may have on the ability of the chosen CRS to function accurately. Both ECR1 of *TAC1* [10] and GAL5.1 of the *Galanin* gene [82] exhibit extreme cell-type dependent activity where they are only able to support reporter gene expression in a tiny subset of hypothalamic and amygdala and PVN (paraventricular nucleus) cells respectively [82]; representing a very small fraction of the total cells found within the animal. With this in mind it is essential that CRS characterisation studies include paradigms that most accurately reflect the expression of endogenous candidate genes in order to develop faithful models of CRS-mediated gene

regulation. Indeed, many of the reports of non-functionality of highly conserved sequences in the existing literature may stem from a failure to analyse these sequences within an appropriate *in vivo* or primary cell-derived model system in which the appropriate cellular components are active.

7. Novel Considerations of *cis*-Regulatory Sequence Polymorphic Variation

7.1. cis-Regulatory Sequence Variation and Drug Response Stratification

Variation in drug response within the human population represents an important barrier to clinical drug development by an increasingly pressured pharmaceutical industry. Referred to as drug response stratification, the outcome is often rejection of the drug based on a lack of a significantly positive or unpredictable response. We propose that CRS variation may be a major causative or contributing factor to drug response stratification. Firstly, consider that the effect on any drug is reliant on its perturbation of a targeted biochemical process or of a receptor function. Modulation of receptor function results in alterations of downstream signal transduction systems that, in turn changes gene expression through CRS activation. Changes in the activity of these CRS, as a result of polymorphic or epigenetic variation, may have important consequences for the downstream effects of these drugs thus contributing to drug response stratification. Indeed, research has indicated that stratified responses to glucocorticoid treatments can result from *cis*-regulatory polymorphisms located near glucocorticoid target genes [101]. Further, non-coding SNPs have been identified which significantly inpact the IC_{50} values and cytotoxicity of chemotherapeutic agents highlighting the potential for such SNPs to be used as markers for predicting drug responses. Characterisation of human genome variation may therefore allow genetic screening to determine the likelihood of a positive/negative drug response in advance of clinical trials. Implementation of this strategy will rely on detailed characterisation of CRSs and their variation in part by the techniques described above which are designed to dissect the precise biochemical events associated with CRS-mediated gene regulation.

7.2. Genetic and Epigenetic Interaction within CRSs and Disease Susceptibility

DNA methylation, the addition of methyl groups to CpG dinucleotides in the genomic sequence, is a heritable form of epigenetic gene regulation vital to cellular homeostasis and development [102]. The presence or absence of the methyl group has been shown to be affected by early life cues such as starvation or stress, and directly prevents TF-DNA binding thereby altering gene transcription. Furthermore, DNA methylation aberrations are associated with human disease [103]. If we consider this process with respect to CRSs that are critical to gene regulation, it is not unreasonable to conclude that CRSs methylation plays an important role in contributing to human pathologies. For example, it has been shown that methylation of a CRS involved in arginine vasopressin (AVP) gene expression can be altered by early life stress. This results in aberrant hormone secretion leading to changes in passive stress coping and memory [104]. We have also detected allelic variants within the GAL5.1 enhancer which renders it susceptible to DNA methylation through the introduction of a CpG sequence [82]. By contrast, analysis of the ECR1 sequence within *CNR1* intron 2 shows the presence of an allelic variant that confers resistance to DNA methylation [85]. Considering the role that the *Galanin* and *CNR1* genes play in appetite, mood and inflammatory pain these examples suggest the

presence of an interplay between genetic and epigenetic variation within CRSs that may have an important baring on our future ability to understand disease susceptibility.

Acknowledgements

We thank Geoffrey Marsh and Adam Osman for comments on the manuscript.

References

1. Levine, M.; Tjian, R. Transcription regulation and animal diversity. *Nature* **2003**, *424*, 147–151.
2. Moore, M.J. From birth to death: The complex lives of eukaryotic mrnas. *Science* **2005**, *309*, 1514–1518.
3. Davidson, E. *The Regulatory Genome: Gene Regulatory Networks in Development and Evolution*; Academic Press: Burlington, San Diego, USA, London, UK, **2006**.
4. Ong, C.-T.; Corces, V.G. Enhancer function: New insights into the regulation of tissue-specific gene expression. *Nat. Rev. Genet.* **2011**, *12*, 283–293.
5. 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*. Initial sequencing and analysis of the human genome. *Nature* **2001**, *409*, 860–921.
6. 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*. The sequence of the human genome. *Science* **2001**, *291*, 1304–1351.
7. Collins, F.S.; Lander, E.S.; Rogers, J.; Waterson, R.H. Finishing the euchromatic sequence of the human genome. *Nature* **2004**, *431*, 931–945.
8. O'Brien, S.J.; Menotti-Raymond, M.; Murphy, W.J.; Nash, W.G.; Wienberg, J.; Stanyon, R.; Copeland, N.G.; Jenkins, N.A.; Womack, J.E.; Marshall Graves, J.A. The promise of comparative genomics in mammals. *Science* **1999**, *286*, 458–481.
9. Lindblad-Toh, K.; Garber, M.; Zuk, O.; Lin, M.F.; Parker, B.J.; Washietl, S.; Kheradpour, P.; Ernst, J.; Jordan, G.; Mauceli, E.; *et al*. A high-resolution map of human evolutionary constraint using 29 mammals. *Nature* **2011**, *478*, 476–482.
10. Davidson, S.; Miller, K.A.; Dowell, A.; Gildea, A.; MacKenzie, A. A remote and highly conserved enhancer supports amygdala specific expression of the gene encoding the anxiogenic neuropeptide substance-p. *Mol. Psychiatry* **2006**, *11*, 410–421.
11. Visel, A.; Bristow, J.; Pennacchio, L.A. Enhancer identification through comparative genomics. *Semi. Cell Dev. Biol.* **2007**, *18*, 140–152.
12. Boffelli, D.; Nobrega, M.A.; Rubin, E.M. Comparative genomics at the vertebrate extremes. *Nat. Rev. Genet.* **2004**, *5*, 456–465.
13. Pennacchio, L.A.; Ahituv, N.; Moses, A.M.; Prabhakar, S.; Nobrega, M.A.; Shoukry, M.; Minovitsky, S.; Dubchak, I.; Holt, A.; Lewis, K.D.; *et al. In vivo* enhancer analysis of human conserved non-coding sequences. *Nature* **2006**, *444*, 499–502.
14. Maston, G.A.; Evans, S.K.; Green, M.R. Transcriptional regulatory elements in the human genome. *Annu. Rev. Genomics Human Genet.* **2006**, *7*, 29–59.

15. The ENCODE Project Consortium. An integrated encyclopedia of DNA elements in the human genome. *Nature* **2012**, *489*, 57–74.
16. Singleton, A.B.; Hardy, J.; Traynor, B.J.; Houlden, H. Towards a complete resolution of the genetic architecture of disease. *Trends Genet.* **2010**, *26*, 438–442.
17. Hirschhorn, J.N.; Daly, M.J. Genome-wide association studies for common diseases and complex traits. *Nat. Rev. Genet.* **2005**, *6*, 95–108.
18. Wang, W.Y.S.; Barratt, B.J.; Clayton, D.G.; Todd, J.A. Genome-wide association studies: Theoretical and practical concerns. *Nat. Rev. Genet.* **2005**, *6*, 109–118.
19. Hindorff, L.A.; Sethupathy, P.; Junkins, H.A.; Ramos, E.M.; Mehta, J.P.; Collins, F.S.; Manolio, T.A. Potential etiologic and functional implications of genome-wide association loci for human diseases and traits. *Proc. Natl. Acad. Sci. USA* **2009**, *106*, 9362–9367.
20. Stern, D.L. Perspective: Evolutionary developmental biology and the problem of variation. *Evolution* **2000**, *54*, 1079–1091.
21. Simonet, W.S.; Bucay, N.; Lauer, S.J.; Taylor, J.M. A far-downstream hepatocyte-specific control region directs expression of the linked human apolipoprotein e and c-i genes in transgenic mice. *J. Biol. Chem.* **1993**, *268*, 8221–8229.
22. Allan, C.M.; Walker, D.; Taylor, J.M. Evolutionary duplication of a hepatic control region in the human apolipoprotein e gene locus. *J. Biol. Chem.* **1995**, *270*, 26278–26281.
23. Simonet, W.S.; Bucay, N.; Pitas, R.E.; Lauer, S.J.; Taylor, J.M. Multiple tissue-specific elements control the apolipoprotein e/c-i gene locus in transgenic mice. *J. Biol. Chem.* **1991**, *266*, 8651–8654.
24. Grehan, S.; Tse, E.; Taylor, J.M. Two distal downstream enhancers direct expression of the human apolipoprotein e gene to astrocytes in the brain. *J. Neurosci.* **2001**, *21*, 812–822.
25. Shih, S.-J.; Allan, C.; Grehan, S.; Tse, E.; Moran, C.; Taylor, J.M. Duplicated downstream enhancers control expression of the human apolipoprotein e gene in macrophages and adipose tissue. *J. Biol. Chem.* **2000**, *275*, 31567–31572.
26. Chaudhuri, A.; Zbrzezna, V.; Polyakova, J.; Pogo, A.O.; Hesselgesser, J.; Horuk, R. Expression of the duffy antigen in k562 cells. Evidence that it is the human erythrocyte chemokine receptor. *J. Biol. Chem.* **1994**, *269*, 7835–7838.
27. Horuk, R.; Chitnis, C.; Darbonne, W.; Colby, T.; Rybicki, A.; Hadley, T.; Miller, L. A receptor for the malarial parasite plasmodium vivax: The erythrocyte chemokine receptor. *Science* **1993**, *261*, 1182–1184.
28. Tournamille, C.; Blancher, A.; Le Van Kim, C.; Gane, P.; Apoil, P.; Nakamoto, W.; Cartron, J.; Colin, Y. Sequence, evolution and ligand binding properties of mammalian duffy antigen/receptor for chemokines. *Immunogenetics* **2004**, *55*, 682–694.
29. Iwamoto, S.; Li, J.; Sugimoto, N.; Okuda, H.; Kajii, E. Characterization of the duffy gene promotor: Evidence for tissue-specific abolishment of expression in fy(aâˆ’bâˆ’) of black individuals. *Biochem. Biophys. Res. Commun.* **1996**, *222*, 852–859.
30. Tournamille, C.; Colin, Y.; Cartron, J.P.; Le Van Kim, C. Disruption of a gata motif in the duffy gene promoter abolishes erythroid gene expression in duffy-negative individuals. *Nat. Genet.* **1995**, *10*, 224–228.

31. Hadley, T.J.; Peiper, S.C. From malaria to chemokine receptor: The emerging physiologic role of the duffy blood group antigen. *Blood* **1997**, *89*, 3077–3091.
32. Miller, L.H.; Mason, S.J.; Clyde, D.F.; McGinniss, M.H. The resistance factor to plasmodium vivax in blacks. *N. Engl. J. Med.* **1976**, *295*, 302–304.
33. Oscar Pogo, A.; Chaudhuri, A. The duffy protein: A malarial and chemokine receptor. *Semi. Hematol.* **2000**, *37*, 122–129.
34. Chaudhuri, A.; Polyakova, J.; Zbrzezna, V.; Pogo, A. The coding sequence of duffy blood group gene in humans and simians: Restriction fragment length polymorphism, antibody and malarial parasite specificities, and expression in nonerythroid tissues in duffy-negative individuals. *Blood* **1995**, *85*, 615–621.
35. Noonan, J.P.; McCallion, A.S. Genomics of long-range regulatory elements. *Annu. Rev. Genomics Hum. Genet.* **2010**, *11*, 1–23.
36. De Gobbi, M.; Viprakasit, V.; Hughes, J.R.; Fisher, C.; Buckle, V.J.; Ayyub, H.; Gibbons, R.J.; Vernimmen, D.; Yoshinaga, Y.; de Jong, P.; *et al.* A regulatory snp causes a human genetic disease by creating a new transcriptional promoter. *Science* **2006**, *312*, 1215–1217.
37. Savic, D.; Park, S.; Bailey, K.; Bell, G.; Nobrega, M. *In vitro* scan for enhancers at the *TCF7L2* locus. *Diabetologia* **2012**, *56*, 121–125.
38. Emison, E.S.; McCallion, A.S.; Kashuk, C.S.; Bush, R.T.; Grice, E.; Lin, S.; Portnoy, M.E.; Cutler, D.J.; Green, E.D.; Chakravarti, A. A common sex-dependent mutation in a ret enhancer underlies hirschsprung disease risk. *Nature* **2005**, *434*, 857–863.
39. Grice, E.A.; Rochelle, E.S.; Green, E.D.; Chakravarti, A.; McCallion, A.S. Evaluation of the ret regulatory landscape reveals the biological relevance of a hscr-implicated enhancer. *Hum. Mol. Genet.* **2005**, *14*, 3837–3845.
40. Monod, J.; Jacob, F. Teleonomic mechanisms in cellular metabolism, growth, and differentiation. *Cold Spring Harb. Symp. Quant. Biol.* **1961**, *26*, 389–401.
41. Britten, R.J.; Davidson, E.H. Gene regulation for higher cells: A theory. *Science* **1969**, *165*, 349–357.
42. Britten, R.J.; Davidson, E.H. Repetitive and non-repetitive DNA sequences and a speculation on the origins of evolutionary novelty. *Q. Rev. Biol.* **1971**, *46*, 111–138.
43. King, M.; Wilson, A. Evolution at two levels in humans and chimpanzees. *Science* **1975**, *188*, 107–116.
44. Davidson, S.; Starkey, A.; MacKenzie, A. Evidence of uneven selective pressure on different subsets of the conserved human genome; implications for the significance of intronic and intergenic DNA. *BMC Genomics* **2009**, *10*, 614.
45. Aparicio, S.; Morrison, A.; Gould, A.; Gilthorpe, J.; Chaudhuri, C.; Rigby, P.; Krumlauf, R.; Brenner, S. Detecting conserved regulatory elements with the model genome of the japanese puffer fish, fugu rubripes. *Proc. Natl. Acad. Sci. USA* **1995**, *92*, 1684–1688.
46. Miller, K.A.; Davidson, S.; Liaros, A.; Barrow, J.; Lear, M.; Heine, D.; Hoppler, S.; MacKenzie, A. Prediction and characterisation of a highly conserved, remote and camp responsive enhancer that regulates msx1 gene expression in cardiac neural crest and outflow tract. *Dev. Biol.* **2008**, *317*, 686–694.

47. Miller, K.A.; Barrow, J.; Collinson, J.M.; Davidson, S.; Lear, M.; Hill, R.E.; MacKenzie, A. A highly conserved wnt-dependent tcf4 binding site within the proximal enhancer of the anti-myogenic msx1 gene supports expression within pax3-expressing limb bud muscle precursor cells. *Dev. Biol.* **2007**, *311*, 665–678.
48. Nobrega, M.A.; Ovcharenko, I.; Afzal, V.; Rubin, E.M. Scanning human gene deserts for long-range enhancers. *Science* **2003**, *302*, 413.
49. Woolfe, A.; Goodson, M.; Goode, D.K.; Snell, P.; McEwen, G.K.; Vavouri, T.; Smith, S.F.; North, P.; Callaway, H.; Kelly, K.; *et al*. Highly conserved non-coding sequences are associated with vertebrate development. *PLoS Biol.* **2004**, *3*, e7.
50. Ovcharenko, I.; Loots, G.G.; Nobrega, M.A.; Hardison, R.C.; Miller, W.; Stubbs, L. Evolution and functional classification of vertebrate gene deserts. *Genome Res.* **2005**, *15*, 137–145.
51. Prabhakar, S.; Poulin, F.; Shoukry, M.; Afzal, V.; Rubin, E.M.; Couronne, O.; Pennacchio, L.A. Close sequence comparisons are sufficient to identify human cis-regulatory elements. *Genome Res.* **2006**, *16*, 855–863.
52. Bejerano, G.; Pheasant, M.; Makunin, I.; Stephen, S.; Kent, W.J.; Mattick, J.S.; Haussler, D. Ultraconserved elements in the human genome. *Science* **2004**, *304*, 1321–1325.
53. Poulin, F.; Nobrega, M.A.; Plajzer-Frick, I.; Holt, A.; Afzal, V.; Rubin, E.M.; Pennacchio, L.A. *In vivo* characterization of a vertebrate ultraconserved enhancer. *Genomics* **2005**, *85*, 774–781.
54. Sandelin, A.; Bailey, P.; Bruce, S.; Engstrom, P.; Klos, J.; Wasserman, W.; Ericson, J.; Lenhard, B. Arrays of ultraconserved non-coding regions span the loci of key developmental genes in vertebrate genomes. *BMC Genomics* **2004**, *5*, 99.
55. Visel, A.; Prabhakar, S.; Akiyama, J.A.; Shoukry, M.; Lewis, K.D.; Holt, A.; Plajzer-Frick, I.; Afzal, V.; Rubin, E.M.; Pennacchio, L.A. Ultraconservation identifies a small subset of extremely constrained developmental enhancers. *Nat. Genet.* **2008**, *40*, 158–160.
56. Maston, G.A.; Landt, S.G.; Snyder, M.; Green, M.R. Characterization of enhancer function from genome-wide analyses. *Annu. Rev. Genomics Hum. Genet.* **2012**, *13*, 29–57.
57. Birney, E.; Stamatoyannopoulos, J.; Dutta, A.; Guigó, R.; Gingeras, T.; Margulies, E.; Weng, Z.; Snyder, M.; Dermitzakis, E.; Thurman, R.; *et al.* Identification and analysis of functional elements in 1% of the human genome by the encode pilot project. *Nature* **2007**, *447*, 799–816.
58. Blow, M.J.; McCulley, D.J.; Li, Z.; Zhang, T.; Akiyama, J.A.; Holt, A.; Plajzer-Frick, I.; Shoukry, M.; Wright, C.; Chen, F.; *et al*. Chip-seq identification of weakly conserved heart enhancers. *Nat. Genet.* **2010**, *42*, 806–810.
59. Eckner, R.; Ewen, M.E.; Newsome, D.; Gerdes, M.; DeCaprio, J.A.; Lawrence, J.B.; Livingston, D.M. Molecular cloning and functional analysis of the adenovirus e1a-associated 300-kd protein (p300) reveals a protein with properties of a transcriptional adaptor. *Genes Dev.* **1994**, *8*, 869–884.
60. Iyer, V.R.; Horak, C.E.; Scafe, C.S.; Botstein, D.; Snyder, M.; Brown, P.O. Genomic binding sites of the yeast cell-cycle transcription factors sbf and mbf. *Nature* **2001**, *409*, 533–538.
61. Ren, B.; Robert, F.O.; Wyrick, J.J.; Aparicio, O.; Jennings, E.G.; Simon, I.; Zeitlinger, J.; Schreiber, J.R.; Hannett, N.; Kanin, E.; *et al*. Genome-wide location and function of DNA binding proteins. *Science* **2000**, *290*, 2306–2309.

62. Impey, S.; McCorkle, S.R.; Cha-Molstad, H.; Dwyer, J.M.; Yochum, G.S.; Boss, J.M.; McWeeney, S.; Dunn, J.J.; Mandel, G.; Goodman, R.H. Defining the creb regulon: A genome-wide analysis of transcription factor regulatory regions. *Cell* **2004**, *119*, 1041–1054.
63. Wei, C.-L.; Wu, Q.; Vega, V.B.; Chiu, K.P.; Ng, P.; Zhang, T.; Shahab, A.; Yong, H.C.; Fu, Y.; Weng, Z.; *et al*. A global map of p53 transcription-factor binding sites in the human genome. *Cell* **2006**, *124*, 207–219.
64. Visel, A.; Blow, M.J.; Li, Z.; Zhang, T.; Akiyama, J.A.; Holt, A.; Plajzer-Frick, I.; Shoukry, M.; Wright, C.; Chen, F.; *et al*. Chip-seq accurately predicts tissue-specific activity of enhancers. *Nature* **2009**, *457*, 854–858.
65. Wu, C. The 5' ends of drosophila heat shock genes in chromatin are hypersensitive to dnase i. *Nature* **1980**, *286*, 854–860.
66. Boyle, A.P.; Song, L.; Lee, B.-K.; London, D.; Keefe, D.; Birney, E.; Iyer, V.R.; Crawford, G.E.; Furey, T.S. High-resolution genome-wide *in vivo* footprinting of diverse transcription factors in human cells. *Genome Res.* **2011**, *21*, 456–464.
67. Crawford, G.E.; Holt, I.E.; Whittle, J.; Webb, B.D.; Tai, D.; Davis, S.; Margulies, E.H.; Chen, Y.; Bernat, J.A.; Ginsburg, D.; *et al.* Genome-wide mapping of dnase hypersensitive sites using massively parallel signature sequencing (mpss). *Genome Res.* **2006**, *16*, 123–131.
68. McDaniell, R.; Lee, B.-K.; Song, L.; Liu, Z.; Boyle, A.P.; Erdos, M.R.; Scott, L.J.; Morken, M.A.; Kucera, K.S.; Battenhouse, A.; *et al*. Heritable individual-specific and allele-specific chromatin signatures in humans. *Science* **2010**, *328*, 235–239.
69. Maurano, M.T.; Humbert, R.; Rynes, E.; Thurman, R.E.; Haugen, E.; Wang, H.; Reynolds, A.P.; Sandstrom, R.; Qu, H.; Brody, J.; *et al*. Systematic localization of common disease-associated variation in regulatory DNA. *Science* **2012**, *337*, 1190–1195.
70. Schadt, E.; Chang, R. A gps for navigating DNA. *Science* **2012**, *337*, 1179–1180.
71. Giresi, P.G.; Lieb, J.D. Isolation of active regulatory elements from eukaryotic chromatin using faire (formaldehyde assisted isolation of regulatory elements). *Methods* **2009**, *48*, 233–239.
72. Dekker, J.; Rippe, K.; Dekker, M.; Kleckner, N. Capturing chromosome conformation. *Science* **2002**, *295*, 1306–1311.
73. Lieberman-Aiden, E.; van Berkum, N.L.; Williams, L.; Imakaev, M.; Ragoczy, T.; Telling, A.; Amit, I.; Lajoie, B.R.; Sabo, P.J.; Dorschner, M.O.; *et al*. Comprehensive mapping of long-range interactions reveals folding principles of the human genome. *Science* **2009**, *326*, 289–293.
74. De Wit, E.; de Laat, W. A decade of 3c technologies: Insights into nuclear organization. *Genes Dev.* **2012**, *26*, 11–24.
75. Simonis, M.; Kooren, J.; de Laat, W. An evaluation of 3c-based methods to capture DNA interactions. *Nat. Meth.* **2007**, *4*, 895–901.
76. Fullwood, M.J.; Wei, C.-L.; Liu, E.T.; Ruan, Y. Next-generation DNA sequencing of paired-end tags (pet) for transcriptome and genome analyses. *Genome Res.* **2009**, *19*, 521–532.
77. Fullwood, M.J.; Ruan, Y. Chip-based methods for the identification of long-range chromatin interactions. *J. Cell. Biochem.* **2009**, *107*, 30–39.
78. Zhang, J.; Poh, H.M.; Peh, S.Q.; Sia, Y.Y.; Li, G.; Mulawadi, F.H.; Goh, Y.; Fullwood, M.J.; Sung, W.-K.; Ruan, X.; *et al.* Chia-pet analysis of transcriptional chromatin interactions. *Methods* **2012**, *58*, 289–299.

79. Kent, W.J.; Sugnet, C.W.; Furey, T.S.; Roskin, K.M.; Pringle, T.H.; Zahler, A.M.; Haussler, D. The human genome browser at ucsc. *Genome Res.* **2002**, *12*, 996–1006.
80. Rosenbloom, K.R.; Dreszer, T.R.; Long, J.C.; Malladi, V.S.; Sloan, C.A.; Raney, B.J.; Cline, M.S.; Karolchik, D.; Barber, G.P.; Clawson, H. Encode whole-genome data in the ucsc genome browser. *Nucleic Acids Res.* **2010**, *38*, 620–625.
81. Kothary, R.; Clapoff, S.; Darling, S.; Perry, M.D.; Moran, L.A.; Rossant, J. Inducible expression of an hsp68-lacz hybrid gene in transgenic mice. *Development* **1989**, *105*, 707–714.
82. Davidson, S.; Lear, M.; Shanley, L.; Hing, B.; Baizan-Edge, A.; Herwig, A.; Quinn, J.P.; Breen, G.; McGuffin, P.; Starkey, A.; *et al*. Differential activity by polymorphic variants of a remote enhancer that supports galanin expression in the hypothalamus and amygdala: Implications for obesity, depression and alcoholism. *Neuropsychopharmacology* **2011**, *36*, 2211–2221.
83. Hing, B.; Davidson, S.; Lear, M.; Breen, G.; Quinn, J.; McGuffin, P.; MacKenzie, A. A polymorphism associated with depressive disorders differentially regulates brain derived neurotrophic factor promoter iv activity. *Biol. Psychiatry* **2012**, *71*, 618–626.
84. Brivanlou, A.H.; Darnell, J.E. Signal transduction and the control of gene expression. *Science* **2002**, *295*, 813–818.
85. Nicoll, G.; Davidson, S.; Shanley, L.; Hing, B.; Lear, M.; McGuffin, P.; Ross, R.; MacKenzie, A. Allele-specific differences in activity of a novel cannabinoid receptor 1 (cnr1) gene intronic enhancer in hypothalamus, dorsal root ganglia, and hippocampus. *J. Biol. Chem.* **2012**, *287*, 12828–12834.
86. Swanson, C.I.; Evans, N.C.; Barolo, S. Structural rules and complex regulatory circuitry constrain expression of a notch- and egfr-regulated eye enhancer. *Dev. Cell* **2010**, *18*, 359–370.
87. Shanley, L.; Davidson, S.; Lear, M.; Thotakura, A.K.; McEwan, I.J.; Ross, R.A.; MacKenzie, A. Long-range regulatory synergy is required to allow control of the tac1 locus by mek/erk signalling in sensory neurones. *Neurosignals* **2010**, *18*, 173–185.
88. Shanley, L.; Lear, M.; Davidson, S.; Ross, R.; MacKenzie, A. Evidence for regulatory diversity and auto-regulation at the tac1 locus in sensory neurones. *J. Neuroinflammation* **2011**, *8*, 10.
89. Sauer, B. Functional expression of the cre-lox site-specific recombination system in the yeast saccharomyces cerevisiae. *Mol. Cell. Biol.* **1987**, *7*, 2087–2096.
90. Sauer, B.; Henderson, N. Site-specific DNA recombination in mammalian cells by the cre recombinase of bacteriophage p1. *Proc. Natl. Acad. Sci. USA* **1988**, *85*, 5166–5170.
91. Orban, P.C.; Chui, D.; Marth, J.D. Tissue- and site-specific DNA recombination in transgenic mice. *Proc. Natl. Acad. Sci. USA* **1992**, *89*, 6861–6865.
92. Gu, H.; Zou, Y.-R.; Rajewsky, K. Independent control of immunoglobulin switch recombination at individual switch regions evidenced through cre-loxp-mediated gene targeting. *Cell* **1993**, *73*, 1155–1164.
93. Gu, H.; Marth, J.; Orban, P.; Mossmann, H.; Rajewsky, K. Deletion of a DNA polymerase beta gene segment in t cells using cell type-specific gene targeting. *Science* **1994**, *265*, 103–106.
94. Lettice, L.A.; Horikoshi, T.; Heaney, S.J.H.; van Baren, M.J.; van der Linde, H.C.; Breedveld, G.J.; Joosse, M.; Akarsu, N.; Oostra, B.A.; Endo, N.; *et al*. Disruption of a long-range cis-acting regulator for shh causes preaxial polydactyly. *Proc. Natl. Acad. Sci. USA* **2002**, *99*, 7548–7553.

95. Lomvardas, S.; Barnea, G.; Pisapia, D.J.; Mendelsohn, M.; Kirkland, J.; Axel, R. Interchromosomal interactions and olfactory receptor choice. *Cell* **2006**, *126*, 403–413.
96. Li, Q.; Barkess, G.I.; Qian, H. Chromatin looping and the probability of transcription. *Trends Genet.* **2006**, *22*, 197–202.
97. Jackson, D.A.; Hassan, A.B.; Errington, R.J.; Cook, P.R. Visualization of focal sites of transcription within human nuclei. *EMBO J.* **1993**, *12*, 1059.
98. Fraser, P.; Bickmore, W. Nuclear organization of the genome and the potential for gene regulation. *Nature* **2007**, *447*, 413–417.
99. Hu, Q.; Kwon, Y.-S.; Nunez, E.; Cardamone, M.D.; Hutt, K.R.; Ohgi, K.A.; Garcia-Bassets, I.; Rose, D.W.; Glass, C.K.; Rosenfeld, M.G.; *et al.* Enhancing nuclear receptor-induced transcription requires nuclear motor and lsd1-dependent gene networking in interchromatin granules. *Proc. Natl. Acad. Sci. USA* **2008**, *105*, 19199–19204.
100. Gondor, A.; Ohlsson, R. Chromosome crosstalk in three dimensions. *Nature* **2009**, *461*, 212–217.
101. Maranville, J.C.; Luca, F.; Richards, A.L.; Wen, X.; Witonsky, D.B.; Baxter, S.; Stephens, M.; di Rienzo, A. Interactions between glucocorticoid treatment and cis-regulatory polymorphisms contribute to cellular response phenotypes. *PLoS Genet.* **2011**, *7*, e1002162.
102. Robertson, K.D. DNA methylation and human disease. *Nat. Rev. Genet.* **2005**, *6*, 597–610.
103. Jaenisch, R.; Bird, A. Epigenetic regulation of gene expression: How the genome integrates intrinsic and environmental signals. *Nat. Genet.* **2003**, *33,* 245-254.
104. Murgatroyd, C.; Patchev, A.V.; Wu, Y.; Micale, V.; Bockmuhl, Y.; Fischer, D.; Holsboer, F.; Wotjak, C.T.; Almeida, O.F.X.; Spengler, D. Dynamic DNA methylation programs persistent adverse effects of early-life stress. *Nat. Neurosci.* **2009**, *12*, 1559–1566.

© 2013 by the authors; licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution license (http://creativecommons.org/licenses/by/3.0/).