



CrossMark
click for updates

Introduction

Cite this article: van Heyningen V, Bickmore W. 2013 Regulation from a distance: long-range control of gene expression in development and disease. *Phil Trans R Soc B* 368: 20120372.
<http://dx.doi.org/10.1098/rstb.2012.0372>

One contribution of 12 to a Discussion Meeting Issue 'Regulation from a distance: long-range regulation of gene expression'.

Subject Areas:

developmental biology, molecular biology, genetics

Keywords:

gene expression, development, genome

Author for correspondence:

Veronica van Heyningen
e-mail: veronica.vanheyningen@igmm.ed.ac.uk

Regulation from a distance: long-range control of gene expression in development and disease

Veronica van Heyningen and Wendy Bickmore

MRC Human Genetics Unit, Institute of Genetics and Molecular Medicine, University of Edinburgh, Edinburgh, UK

Regulated gene expression is a major requirement for all living organisms. The requirement for complex spatio-temporal regulation is most obvious during development and differentiation, when precise gene switching choreographs the generation of many different cell types, at the right time and the right place, from a single fertilized cell. When this process goes awry, deciphering the genetic cause can provide detailed insight into mechanisms. While chromatin structure and the recruitment of the transcriptional machinery to proximal promoters are well understood, how far-distant enhancers direct the correct spatial and temporal control of transcription is less clear. This concept prompted us to organize a Royal Society Discussion Meeting on this topic in October 2012. The timeliness of the debate was highlighted by the publication of results from the prominently heralded ENCODE project published just a month before the meeting (<http://www.nature.com/encode/#/threads>). This highlighted the unexpectedly large expanse of the human genome that appears to harbour regulatory elements [1,2]. Here, we present papers from some of the speakers at this lively and exciting meeting.

Simple Mendelian genetic analyses in mouse developmental models and human genetic diseases first pointed to the importance of regulation from elements—termed enhancers—that can be located up to 2 million base pairs away from the affected genes [3]. These mutations were often large structural chromosomal aberrations that either completely removed an enhancer from the genome (deletion) or that separated the enhancer from its target gene, e.g. as a consequence of a chromosome translocation or large inversion. However, with the advent of new genome analysis technologies and whole genome sequencing in large cohorts of control cases [4], more subtle sequence changes, that may be potentially pathological and affect regulatory elements, can now be explored with greater confidence.

Much effort has been invested recently in genome-wide association studies to identify genomic variants, usually single nucleotide polymorphisms (SNPs), that track with complex and common human disease. Variants within gene-coding regions lend themselves immediately to further functional analysis. However, it now appears that more than 85 per cent of the variants identified as associated with disease traits map outside the coding region of annotated genes [5]. Some of these regulatory SNPs are in nearby gene-flanking or intronic regions; others map in regions termed 'gene deserts' a long way from any recognized gene, which makes the recognition of target genes difficult [6]. A proportion is associated with non-coding RNAs [7]. A better understanding of the mechanisms of gene regulation will undoubtedly bring new insight into human genetic variation and so into the predisposition to common diseases, including cancer incidence and progression.

Other technological developments that have made genome-wide studies more informative include expression analysis [8] and chromatin immunoprecipitation (ChIP) to identify the genomic regions where transcription factors and chromatin proteins bind [9] or where histones carry post-translational modifications that are indicative of enhancer activity [10]. Another approach to define genomic regions harbouring actively transcribed genes in different cell types is DNase hypersensitivity analysis, which reveals regions of disrupted

nucleosome structure [2]. The possible physical associations between regulatory regions and target genes can be assessed by chromosome-conformation capture (3C) methods [11,12], comprising a number of techniques based on formaldehyde fixation of protein–protein and protein–DNA complexes, removal of the unattached DNA fragments and isolation of the cross-linked distant fragments to determine which regions are brought together by the cross-linked proteins.

The evolutionary conservation of putative enhancers, or indeed their rapid evolution in concert with morphological change, provides a complementary perspective from which to identify where enhancers come from and how they function.

The opportunity for detailed discussion was welcomed by the capacity crowd of participants, because it was a great occasion to compare technologies and debate the parallels and inconsistencies between results obtained using different approaches. One of the best characterized, but still not fully understood regulatory regions, associated with human, mouse and cat limb abnormalities, was described by Hill & Lettice [13] showing how a limb regulatory element for sonic hedgehog expression was identified through genetics, about a megabase upstream of *SHH/Shh*. Complex studies in mouse transgenics and biochemical approaches have deciphered quite a few mysteries surrounding this regulatory element, such as gain of function mutations that create new binding sites for specific transcription factors, but they also highlight how many facets of long-range regulation are still to be explored further. Another well-studied region, the Hox clusters, is an example of multiple clustered genes regulated by a complex of control elements and was discussed by Montavon & Duboule [14]. This example illustrates how mouse models, coupled to chromatin analysis, can be used to dissect the functions and interactions of multiple loci within a region and the role that dynamic chromatin architecture plays in the spatio-temporal regulation of Hox expression during embryonic development. Doug Higgs presented the beautiful analyses of another developmentally regulated gene cluster—the α -globin locus. Here, a great deal has been learnt from the spectrum of human α -thalassaemia mutations, but mouse models and transgenic studies are still essential for detailed analysis of the system. By the development of a 3C approach, they describe a high-resolution analysis of all the long-range interactions within and around the α -globin locus, and compare this data with the known regulatory elements identified for this region [15].

Transgene and reporter assays have been important for exploring the regulatory potential of candidate genomic elements, but these have often been conducted element by individual element. More large-scale *in vivo* and *in vitro* approaches are now being developed to address questions from human disease such as what are the functional consequences of subtle sequence variants in candidate enhancer elements, or how can such elements be pinpointed in the vast regulatory landscapes that surround some genes? Len Pennacchio discussed the functional anatomy of enhancer elements and asked whether assays can distinguish if the multiple transcription factor binding sites within these elements act as individual units or as functional modules, and the extent to which there might be functional redundancy within enhancers [16]. They highlight that there may not be ‘one rule fits all’ for the way in which enhancers work, so that it might be very hard to come up with algorithms for comprehensive enhancer prediction.

While assays that clone potential enhancers next to reporter genes are a powerful way to assay the regulatory potential of specific sequences, they may not completely and accurately reflect the precise pattern of control that an enhancer elicits in its endogenous genomic context. Spitz showed how he has adapted a transposon as a sensor to pick up endogenous regulatory activity within broad regulatory landscapes in the mouse genome. The output of this *in vivo* assay is compared with that of more conventional reporter assays [17], and they discuss how surprisingly widespread regulatory elements are in contrast to the original models of rather discrete individual elements, and the complex patterns of tissue-specific expression that this assay reveals. Importantly, they highlight how easily promoters inserted into the mouse genome seem to be able to respond to the surrounding regulatory genomic landscape they find themselves in. They remind us therefore that, for example, just because a non-coding transcript in such a region shows a tissue-specific pattern of expression, does not automatically imply a functional significance to the expression of that transcript.

Wouter de Laat is a master of chromatin conformation capture and its analysis. A protein that has consistently come up as associated with topological features of chromatin is CTCF. CTCF protein is a sequence-specific DNA-binding protein that has been proposed to mark insulator sites between genes, but also clearly acts as a transcription factor under some circumstances and can mediate chromatin looping between distant elements, including enhancers and promoters. Here, Holwerda & de Laat [18] discuss these many faces of CTCF, whether there are any unifying themes, and to what extent the possible functions of CTCF depend on other interacting or nearby proteins. Complex regulation is required not only during development and in different cell types but also in the differentiation of functionally distinct inflammatory cells that need to work together to control the response to environmental stimuli. Ghisletti & Natoli [19] gave a very nice introduction to the concept of environmental signals impacting on the *cis*-regulatory landscape in macrophage. In particular, the concept of binding by pioneer transcription factors is discussed, and how this can shape the regulatory landscape to be able to subsequently respond to external stimuli.

It was clear from many of the talks at this meeting, and of course from the spirited discussion too, that computational and bioinformatic approaches are required for many steps in the genome-wide and comparative study of long-range gene control. Ovcharenko discussed the complex analyses that can be undertaken to predict novel enhancers and to assess, for example, whether long-range enhancers differ from proximal promoters, whether a tissue-specific code is embedded in promoters, and if so whether such a code could be used to search for more distant enhancers. Here, by comparing enhancers active in heart between mouse and human, it is suggested that species-specific regulatory activity can be acquired and lost rather readily in evolution and that conserved enhancers have a stronger effect on their target genes than species-specific enhancers [20]. Wysocka discussed human regulatory variation in neural crest cells that are implicated in face development. Facial morphology is quite strongly inherited but varies significantly between individuals in different families. Can regulatory variation account for this? Using several different model systems, including *in vitro* differentiated human embryonic stem cells, it was found that a specific transcription factor in

cooperation with nuclear receptors bind to enhancers regulating expression in the neural crest lineage, and that sequence variation that affects nuclear receptor binding sites has knock on effects on transcription factor binding [21]. Freedman's group describes some new approaches that are designed to deconvolute complex tissues into their component cell types, both using laboratory methods and bioinformatic analysis. This is very important as we know that every tissue consists of multiple cell types, and different regulatory elements are required for the functions of each. Studies such as ChIP or transcriptome profiling carried out on a cell mixture only give a cell-averaged view, but it will be important to determine what is happening in each component cell type. This is particularly true for cancer studies. In breast cancers, expression

quantitative trait loci (eQTLs) were examined in the regions surrounding genetic risk loci for these diseases to try to determine associations in different cell populations [22].

Finally, we end this wide sweep across the regulatory landscapes by a discourse from Dermitzakis, who was one of the most active discussants at the meeting. Nica & Dermitzakis [23] described their spare and elegant approaches to deciphering the regulation of quantitative trait loci using a combination of laboratory measurements and computational approaches to assess the role of genome variation on transcriptional output. His enthusiasm was infectious, conforming well to the intensely argumentative sessions that have characterized discussions at the Royal Society since its earliest days.

References

- Neph S *et al.* 2012 An expansive human regulatory lexicon encoded in transcription factor footprints. *Nature* **489**, 83–90. (doi:10.1038/nature11212)
- Thurman RE *et al.* 2012 The accessible chromatin landscape of the human genome. *Nature* **489**, 75–82. (doi:10.1038/nature11232)
- Kleinjan DA, van Heyningen V. 2005 Long-range control of gene expression: emerging mechanisms and disruption in disease. *Am. J. Hum. Genet.* **76**, 8–32. (doi:10.1086/426833)
- 1000 Genomes Project Consortium, Abecasis GR, Auton A, Brooks LD, DePristo MA, Durbin RM, Handsaker RE, Kang HM, Marth GT, McVean GA. 2012 An integrated map of genetic variation from 1,092 human genomes. *Nature* **491**, 56–65. (doi:10.1038/nature11632)
- Pastinen T. 2010 Genome-wide allele-specific analysis: insights into regulatory variation. *Nat. Rev. Genet.* **11**, 533–538. (doi:10.1038/nrg2815)
- Ghousaini M *et al.*; UK Genetic Prostate Cancer Study Collaborators/British Association of Urological Surgeons' Section of Oncology; UK ProtecT Study Collaborators. 2008 Multiple loci with different cancer specificities within the 8q24 gene desert. *J. Natl Cancer Inst.* **100**, 962–966. (doi:10.1093/jnci/djn190)
- Kumar V *et al.* 2013 Human disease-associated genetic variation impacts large intergenic non-coding RNA expression. *PLoS Genet.* **9**, e1003201. (doi:10.1371/journal.pgen.1003201)
- Dong X *et al.* 2012 Modeling gene expression using chromatin features in various cellular contexts. *Genome Biol.* **13**, R53. (doi:10.1186/gb-2012-13-9-r53)
- Furey TS. 2012 ChIP-seq and beyond: new and improved methodologies to detect and characterize protein–DNA interactions. *Nat. Rev. Genet.* **13**, 840–852. (doi:10.1038/nrg3306)
- Rajagopal N, Xie W, Li Y, Wagner U, Wang W, Stamatoyannopoulos J, Ernst J, Kellis M, Ren B. 2013 RFECs: a random-forest based algorithm for enhancer identification from chromatin state. *PLoS Comput. Biol.* **9**, e1002968. (doi:10.1371/journal.pcbi.1002968)
- de Graaf CA, van Steensel B. 2012 Chromatin organization: form to function. *Curr. Opin. Genet. Dev.* (doi:10.1016/j.gde.2012.11.011)
- Stadhouders R *et al.* 2013 Multiplexed chromosome conformation capture sequencing for rapid genome-scale high-resolution detection of long-range chromatin interactions. *Nat. Protoc.* **8**, 509–524. (doi:10.1038/nprot.2013.018)
- Hill RE, Lettice LA. 2013 Alterations to the remote control of *Shh* gene expression cause congenital abnormalities. *Phil. Trans. R. Soc. B* **368**, 20120357. (doi:10.1098/rstb.2012.0357)
- Montavon T, Duboule D. 2013 Chromatin organization and global regulation of *Hox* gene clusters. *Phil. Trans. R. Soc. B* **368**, 20120367. (doi:10.1098/rstb.2012.0367)
- Hughes JR *et al.* 2013 High-resolution analysis of *cis*-acting regulatory networks at the α -globin locus. *Phil. Trans. R. Soc. B* **368**, 20120361. (doi:10.1098/rstb.2012.0361)
- Dickel DE, Visel A, Pennacchio LA. 2013 Functional anatomy of distant-acting mammalian enhancers. *Phil. Trans. R. Soc. B* **368**, 20120359. (doi:10.1098/rstb.2012.0359)
- Symmons O, Spitz F. 2013 From remote enhancers to gene regulation: charting the genome's regulatory landscapes. *Phil. Trans. R. Soc. B* **368**, 20120358. (doi:10.1098/rstb.2012.0358)
- Holwerda SJB, de Laat W. 2013 CTCF: the protein, the binding partners, the binding sites and their chromatin loops. *Phil. Trans. R. Soc. B* **368**, 20120369. (doi:10.1098/rstb.2012.0369)
- Ghisletti S, Natoli G. 2013 Deciphering *cis*-regulatory control in inflammatory cells. *Phil. Trans. R. Soc. B* **368**, 20120370. (doi:10.1098/rstb.2012.0370)
- Hsu C-H, Ovcharenko I. 2013 Effects of gene regulatory reprogramming on gene expression in human and mouse developing hearts. *Phil. Trans. R. Soc. B* **368**, 20120366. (doi:10.1098/rstb.2012.0366)
- Rada-Iglesias A, Prescott SL, Wysocka J. 2013 Human genetic variation within neural crest enhancers: molecular and phenotypic implications. *Phil. Trans. R. Soc. B* **368**, 20120360. (doi:10.1098/rstb.2012.0360)
- Seo J-H, Li Q, Fatima A, Eklund A, Szallasi Z, Polyak K, Richardson AL, Freedman ML. 2013 Deconvoluting complex tissues for expression quantitative trait locus-based analyses. *Phil. Trans. R. Soc. B* **368**, 20120363. (doi:10.1098/rstb.2012.0363)
- Nica AC, Dermitzakis ET. 2013 Expression quantitative trait loci: present and future. *Phil. Trans. R. Soc. B* **368**, 20120362. (doi:10.1098/rstb.2012.0362)