Report on the ESF exploratory workshop EW06-069 "Computational approaches to the role of epigenetic marks in transcription regulation"

Erik van Nimwegen

January 2008

Executive Summary

The ESF exploratory workshop entitled "Computational approaches to the role of epigenetic marks in transcription regulation", organized by Erik van Nimwegen (Biozentrum, University of Basel) and Nikolaus Rajewsky (Max-Delbrück Center, Berlin), took place on the 18th and the 19th of October, at the Biozentrum, Basel. Recently there has been a significant increase in large-scale experimental approaches to mapping chromatin states. At the same time a combination of recent theoretical and experimental developments has made it possible to map transcription factor binding sites genome-wide in model organisms and several theoretical groups have initiated attempts to model genome-wide transcriptional regulatory networks using such mappings. However, it has become clear that epigenetic changes to the state of the chromatin play an essential role in transcription regulation. The main motivation for the workshop was thus to explore how theoretical/computational researchers and experimental researchers could come together to collaborate on making progress on understanding the role of epigenetic marks in transcription regulation.

As intended, the workshop brought together researchers from different backgrounds working on widely varying aspects of the field. This included researchers working on the dynamics of the chromatin fibre, researchers working to elucidate the molecular mechanisms of chromatin modification and epigenetic inheritance, researchers investigating large-scale organization of chromosomes in the nucleus, researchers studying the genome-wide dynamics of chromatin modifications in different developmental processes, theorists working on the genome-wide mapping of regulatory sites and on the association of chromatin modification with disease, and researchers studying the role of small RNAs in chromatin modifications. The participants were selected in particular because of their interest in and experience with combined theoretical/experimental approaches.

The presentations at the meeting served to provide a timely overview of the current state of the art in these different areas, and allowed the participants to see what the key questions are that are being addressed in different groups and what approaches they are taking to address them. To give

a few examples, the two presentations by Renato Paro and Marc Rehmsmeier showed how joint computational and experimental approaches can be taken to identify targets of polycomb repressor proteins and their role in cellular differentiation. Current large-scale approaches to the analysis of chromatin marks were discussed by Dirk Schübeler, David Gifford, Denise Barlow and Antoine Peters who u are studying the molecular events underlying cellular differentiation during various developmental processes. They use oligonucleotide microarrays and deep sequencing technologies to analyze time courses of various genome-wide observables (expression, histone state, DNA methylation state, and binding of transcription factors), with the aim of understanding the dynamics of gene expression and its regulation by epigenetic marks. A strong emphasis was placed on the necessity of developing computational infrastructure and algorithms to deal with this type of data, some efforts in this direction being described by Christoph Bock. Emphasis on the development of combined experimental-computational technologies for accurate expression profiling was also put by Arndt Benecke. Heterochromatin formation in yeast was described by Genevieve Thon and Susan Gasser. The chromosome structure at a very large scale was discussed by Ana Pombo, who presented evidence for the existence of "transcription factories", 50-nanometer structures containing clusters of polymerases, of chromosome territories and inter-chromosomal interactions. Bas van Steensel presented work on the factors that appear to define boundaries of "chromosomal territories" in fly and human. At a very fine scale, there has been considerable work in modeling DNA as a polymer. The performance and limitations of this type of models were presented by Dieter Heermann and Susan Gasser. Finally, very recent work in both yeast and mammals revealed the involvement of small RNAs in the regulation of chromatin states. From this perspective, Mihaela Zavolan described the involvement of micro RNAs in the de novo DNA methylation in mouse embryonic development.

Important for the objectives of the workshop, each day was concluded by a discussion session, which participants found particularly stimulating scientifically, because the format encouraged them to share their diverse perspectives on this rather new field, and to be more speculative in discussing the directions that the field may take. Several participants indicated that they felt the workshop was unique in that it brought together a wide variety of researchers working in this rapidly developing field, and that it was especially valuable to be able to learn about the view points of people with such diverse backgrounds. It was felt that the field would benefit from more broad discussions such as in this workshop. In addition, participants agreed that it was strategically important to stimulate the joining of forces between theoretical and experimental researchers already at the level of formulating the research approach and experimental design. Of particular interest would be further investigations about the ways in which the new deep sequencing platforms can be used in combined theoretical/experimental approaches. Finally, traditionally epigenetics has mainly focused on DNA but in the meeting it became clear that more and more RNA is recognized as having important roles, too. Therefore, beyond the need of linking computational and experimental groups, it would be good foster collaborations between epigenetics groups and RNA biology groups.

Scientific content of the event

One of the most puzzling questions concerning transcription regulation in eukaryotes is how its specificity is realized. Because the eukaryotic genomes are considerably larger compared to bacterial genomes, and the specific of eukaryotic transcription factors is typically lower, i.e. frequently involving 6-8 nucleotides (compared to 20 in bacteria), the number of putative binding sites for any given eukaryotic transcription factor is very large. Thus, it is difficult to imagine how a transcription factor finds its targets in a cell-type specific manner. One attractive solution that has been proposed is that the *accessibility* of the binding sites is modulated in a cell-type specific manner through the chromatin structure. This of course then begs the question of how changes in chromatin structure are being brought about and regulated.

Much of the knowledge that we currently have about epigenetics comes from work in fruit fly. The meeting started out with Renato Paro presented a review of this work, and more generally the current understanding of the role of chromatin marks in transcription regulation. The second part of the presentation focused on recent results regarding the role of polycomb response elements (PRE) in switching off the activity of their target genes.

In the past few year, genome tiling arrays and deep sequencing technologies made possible the genome-wide profiling of various types of marks: RNA polymerase, DNA methylation, histone modifications (methylations, acetylations at various positions), etcetera. Several groups have started to obtain time courses of such genome-wide marks in order to study the dynamics of chromatin during different cellular differentiation processes. An important challenge from a computational point of few is how to abstract, from these enormous data sets, the regulatory logic that governs the dynamics of the chromatin state, and its effects on gene expression. David Gifford (Massachussetts Institute of Techology) and Dirk Schübeler (Friedrich-Miescher Institute) have pioneered computational and experimental approaches, respectively, to profiling of chromatin states, and presented on these in the meeting. David Gifford studied the process of mouse embryonic stem cell differentiation into motor neurons, and showed that based on time courses of chromatin marks and dynamic maps of the location of key transcription factors, one can start to infer causal relationships between these different types of marks. Dirk Schübeler uses much the same approaches, and his presentation focused on the dynamics of DNA methylation during pyramidal neuron differentiation from mouse embryonic stem cells. Denise Barlow (Center for Molecular Medicine of the Austrian Academy of Science) discussed similar issues in the context of chromatin formation at imprinted loci, while Antoine Peters (Friedrich Miescher Institute) focused on the epigenetic control of mouse pre-implantation development. This last presentation brought up the very interesting question of memory in epigenetic control, particularly in the context of the extensive remodeling of chromatin that takes places during the development of male germ cells.

A paradigm that emerged from the genome-wide studies of chromatin states during early development is the notion of bivalent marks. That is, during early development, many genes acquire both “active” and “silent” chromatin marks, and which of these marks is ultimately kept depends on subsequent events in the cell, and is cell lineage-specific. However, it is currently unclear what

precisely these events are, and how the boundaries between domains of active and inactive chromatin are being defined. This was one of the points taken up during discussions later in the day. Another paradigm that was discussed is the continuous cross-talk between chromatin state and transcription, meaning that changes in transcriptional state change local chromatin state and vice versa.

Finally, a currently open question is the role of non-coding transcripts in the establishment of chromatin states. Several examples of non-coding RNA involvement are known, such as the X chromosome inactivation, dependent on the Xist non-coding RNA, and the silencing of transposable elements in fly, dependent on the Piwi protein-associated RNAs, but it is believed that many more examples will be discovered in the near future. A very new example of the involvement of regulatory RNAs in the regulation of chromatin state was presented by Mihaela Zavolan (Biozentrum, Basel), who indicated that embryonic micro RNAs may indirectly regulate the activity of *de novo* DNA methylases. Dirk Schübeler uses much the same approaches, and his presentation focused on the dynamics of DNA methylation during pyramidal neuron differentiation from mouse embryonic stem cells. Denise Barlow (Center for Molecular Medicine of the Austrian Academy of Science) discussed similar issues in the context of chromatin formation at imprinted loci, while Antoine Peters (Friedrich Miescher Institute) focused on the epigenetic control of mouse pre-implantation development. This brought up the very interesting question of memory in epigenetic control, particularly in the context of the extensive remodeling of chromatin that takes places during the development of male germ cells.

From the computational side contributions span the spectrum from data analysis tools to conceptual computational models and both have been shown to be instrumental in the interpretation of the experimental data. On the data analysis side, David Gifford described an approach based on hidden Markov models to identification of chromatin domains, and a "joint binding deconvolution" algorithm to identification of peaks where individual factors appear to be bound. Similarly, Bas van Steensel described his joint work with Harmen Bussemaker (Columbia University) on segmentation algorithms for identification of chromosome domains. Genevieve Thon (Niels Bohr Institute, Denmark) presented insights into the establishment of chromosomal regions obtained through computational models (jointly developed with Kim Sneppen (Niels Bohr Institute, Denmark)). In particular, it was found that stable and heritable alternative chromatin states can be obtained epigenetically, but that robust bi-stability requires cooperativity between neighboring nucleosomes and also that the modification effects induced by one nucleosome extend beyond the nucleosome boundaries.

Interestingly, several signals involved in chromatin formation can be predicted based on the genome sequence. This is the case for polycomb response elements, which Marc Rehmsmeier (University of Bielefeld) predicted with high accuracy in a number of fly genomes in joint work with Renato Paro(Department of Biosystems Science and Engineering, Basel). A puzzling observation is that these regulatory elements often appear to show patterns of very rapid evolution, in spite of the fact that they are presumably very important for the organism's survival and fitness. This issue, i.e. the plasticity of regulatory sequences in DNA, has been under active discussion by

(mostly computational) researchers for quite some time. On the one hand, functionally important regulatory sites often tend to be well-conserved, whereas on the other hand, several studies suggest that gene regulation evolves much faster than the genes themselves and that regulatory DNA is very plastic. The mechanisms underlying these observations are currently not well understood and clearly deserve further investigation.

Christoph Bock (Max Planck Institute for Bioinformatics, Saarbrücken) also discussed computational approaches to identification important regulatory features directly from genome sequence, focusing mostly on CpG islands. He also described computational efforts supporting the analysis of links between chromatin modifications and disease states, in particular in cancer.

It is important to point out that the spectrum of topics presented at the workshop covered very different scales, from the organization of DNA into chromosome fibers, a topic discussed by Dieter Heermann (University of Heidelberg) and Susan Gasser (Friedrich Miescher Institute), to the chromosome and nuclear organization discussed by Ana Pombo (Imperial College, UK), Bas van Steensel and several others. At the smallest scales polymer-chain based models have been relatively successful in representing the chromatin fiber but there still some important open questions, such as determining the the unit scale for chromatin structure, as was brought up during the discussion session. On much larger scales, the presentations by Ana Pombo and Bas van Steensel showed that various regions of chromosome associate in an almost deterministic manner with different nuclear structures such as nuclear lamina. This raises the question of what factors contribute to the specificity of such interactions, and how these interactions subsequently determine the activity of genes in various areas of the chromosome.

The workshop was concluded with a discussion that aimed to identify the most important current questions in epigenetic research, and to point out less conspicuous areas where combined computational-experimental approaches are likely to foster progress in the field. Some of the open questions that were actively discussed are:

1. To what extent are the chromatin changes that accompany cellular differentiation irreversible? And if indeed some of these changes are effectively irreversible, what is the origin of this irreversibility? In physics the key concept in irreversible processes is entropy. Would a similar conceptual frame work apply to irreversible cellular changes?

2. What are the mechanisms by which chromatin marks spread along the genome and how are the domain boundaries set? Might transcription itself play a role in these processes?

3. Recent deep sequencing studies suggest a low level of "background transcription" coming from almost every in the genome. Is this process just the result of noise or is it regulated, and what role might it play in chromatin dynamics? Is there such a thing as a true "off" state?

4. To what extent is the concept of a "histone code" useful? Are epigenetic marks carrying and processing regulatory information that is separate and distinct from the information in the DNA and in the gene expression state? Are epigenetic marks simply providing a memory of

previous regulatory 'decisions'? Or do epigenetic marks play the role of amplifying regulatory pathways so as to give them more robustness?

5. How important are non-coding RNAs in epigenetics?

It is clear that in the immediate future we will be in an exploratory phase, in which much of the computational work will focus on analysis of large-scale datasets. Nonetheless, one conclusion that was shared by participants is that computation should not be done in a post-hoc manner, but that theoretical and experimental should be an integral part of planning of the research the experiment design, to ensure that the data that is ultimately obtained can be interpreted.

Assessment of the results

First, as an exploratory workshop the objectives were mainly to explore how computational and experimental methods can provide synergy. In this regard it was particularly stimulating for the theorists and experimentalists to exchange ideas about what the current key research questions are. In addition, the presentations provided a excellent overview of the current state of the art in the role of epigenetic marks in transcription regulation. The combination of an overview of the current approaches taken by different groups, and the discussions among experimentalists and theorists regarding interesting outstanding questions, has importantly shaped the planning for future research of all participants. This is one of the first and foremost results.

Second, the discussions were instrumental in generating and crystallizing out important research questions such as listed in the previous section. Third, the fact that each participant got to know experts in related, yet not identical fields of research, is likely to stimulate new research projects and yield collaborations in the near future. Some of the themes that were either touched upon and that are envisioned for such collaborations are:

1. The development of more accurate methods for large scale profiling of gene activity, especially including transcription starts and ends.

2. Development of computational methods to interpret large-scale profiling data on chromatin marks, especially with regards to mapping boundaries (i.e. the segmentation of the chromatin into different domains).

3. Developing new methods for making optimal use of the power of recent deep-sequencing technology, especially with regards to transcription regulation and the dynamics of chromatin state.

4. Methods for going beyond correlations between different chromatin modification and gene activity and infer causal relationships. For examples through the development of joint computational/experimental approaches that use perturbation of differentiating systems to probe regulatory logic.

5. Development of computational frameworks to infer gene interaction networks using multiple large-scale datasets.

Finally, the discussions also lead to a few conclusions of a more strategic nature. First and foremost, it is important the collaborations between experimental and theoretical researchers are being stimulated. In particular the different backgrounds leads theorists and experimentalists to have rather different points of view on the scientific questions at hand and it is important that they already collaborate at the stage of problem formulation and experimental design. Second, it is the organizers' opinion that the progress of the field would be catalyzed by the organization of future specialized, small-scale workshops, design to discuss in detail particular methods or technologies, and to resolve outstanding problems within small groups of specialists. For example a workshop on how deep sequencing technology can be used for investigating the dynamics of chromatin marks and their role in transcription regulation would be particularly welcome. Finally, small RNAs are starting to be recognized as important players in the regulation of chromatin structure and it would be strategically opportune to stimulate links between epigenetics groups and groups studying RNA.

Participant Statistics

Country of origin:

- Austria: 3
- Denmark: 1
- France: 1
- Germany: 3 + 1 convenor)
- Netherlands: 1
- Switzerland: 5 + 1 convenor
- United Kingdom: 1
- USA: 1

Gender Male: 13 Female: 5
Age bracket

- 20-30 years: 2
- 30-40 years: 8
- 40-50 years: 6
- 50-60 years: 2

ESF Workshop (Exploratory Workshop)

Computational Approaches to the Role of Epigenetic Marks in Transcription Regulation

Date: 17. - 20. October 2007
Location: Biozentrum, University of Basel – Switzerland (Room 1067)
Organizers: Erik van Nimwegen, Nikolaus Rajewsky

Program

Thursday	October 18
08.45-09.00	Welcome and introduction.
Session I	**Epigenetic regulation by polycomb proteins**
09.00-09.40	Renato Paro D-BSSE, ETH-Zurich, Switzerland *Epigenetic profiles of ON/OFF gene expression states*
09.40-10.20	Ana Pombo MRC Clinical Sciences Centre, London, UK *Poised transcription complexes in epigenetics and genome organisation*
10.20-10.30	Break
10.30-11.10	Marc Rehmsmeier CeBiTec, Universität Bielefeld, Germany *Evolutionary plasticity of regulatory DNA*
11.10.11.50	Christoph Bock MPI für Informatik, Saarbrücken, Germany *Realizing the medical potential of epigenomics by tailored algorithms and software*
12.00-13.30	Lunch
Session II	**Modeling chromatin state and dynamics**
13.30-14.10	Dieter Heermann Ruprecht-Karls-Universität Heidelberg, Germany *Polymer Models for Chromosomes*
14.10-14.50	Arndt Benecke Institut des Hautes Etudes Scientifiques, Bures sur Yvette, France *Linking molecular mechanisms to the organism's physiology using transcriptomics*
14.50-15.00	Break
15.00-15.40	David K Gifford MIT, Cambridge, USA *Chromatin changes and transcriptional regulation during motor neuron development*
15.40-16.00	Break
16.00-18.00	Discussion session
20.00-....	Workshop dinner: Restaurant Rollerhof, Münsterplatz 20

Friday	October 19
Session III	**Genome-wide mapping of chromatin modifications**
09.00-09.40	Dirk Schübeler Friedrich Miescher Institute, Basel, Switzerland *Targets and function of DNA methylation in mammalian genomes*
9.40.10.20	Mathew A. Sloane / Florian M. Pauler / Denise Barlow Ce-M-M- Research Center for Molecular Medicine, Vienna, Austria *A simple method to identify histone modifications spreading along mammalian chromosomes*
10.20-10.30	Break
10.30-11.10	Antoine Peters Friedrich Miescher Institute, Basel, Switzerland *Investigating epigenetic control of mouse pre-implantation development*
11.10-11.50	Bas van Steensel Netherlands Cancer Institute, Amsterdam, the Netherlands *Chromatin domains in flies and humans*
12.00-13.00	Lunch
Session IV	**Small RNAs**
13.30-14.10	Mihaela Zavolan Biozentrum, Basel, Switzerland *Small RNAs in the regulation of mammalian gene expression*
14.10-14.50	Massimo Vergassola Pasteur Institute, Paris, France *Non-coding RNAs in bacteria*
14.50-15.00	Break
Session V	**Epigenetics in yeast**
15.00-15.40	Genevieve Thon Biocenter, University of Copenhagen, Denmark *Heterochromatin formation in fission yeast*
15.40-16.20	Susan Gasser Friedrich Miescher Institute, Basel, Switzerland *Spatial considerations in gene expression*
16.20-16.30	Break
16.30-18.00	Discussion & Planning of follow-up activities and collaborations

List of Participants

Denise BARLOW
Vienna Biocenter
Max F. Perutz Laboratories
Institute of Genetics
Dr. Bohr-Gasse 9/4
1030 Vienna
Austria
Email: denise.barlow@univie.ac.at

Arndt BENECKE
Institut des Hautes Etudes Scientifiques
Systems Epigenomics Group
35, route de Chartres
91440 Bures sur Yvette
France
Email: arndt@ihes.fr

Christoph BOCK
Max-Planck-Institut für Informatik
Department of Computational Biology
Campus E1
4/ 66123 Saarbrücken
Germany
Email: christoph@bock.name

Susan GASSER
Friedrich Miescher Institute for Biomedical Research
Maulbeerstrasse 66
4058 Basel
Switzerland
Email: susan.gasser@fmi.ch

David GIFFORD
Massachusetts Institute of Technology
Computational Genomics Research Group
Computer Science and Artificial Intelligence Laboratory
32 Vassar Street G540
Cambridge MA 02139
United States
Email: dkg@mit.edu

Dieter HEERMANN
Theoretical Biophysics Group
Institut für Theoretische Physik
Ruprecht-Karls-Universität Heidelberg
Philosophenweg 19
69120 Heidelberg
Germany
Email: heermann@tphys.uni-heidelberg.de

Erik VAN NIMWEGEN
Biozentrum
Division of Bioinformatics
University of Basel
Klingelbergstrasse 50/70
4056 Basel
Switzerland
Tel: +41 61 267 1576
Fax: +41 61 267 1584
Email: erik.vannimwegen@unibas.ch

Renato PARO
Department of Biosystems Science and Engineering
ETH Zurich
Mattenstrasse 26
4058 Basel
Switzerland
Email: renato.paro@bsse.ethz.ch

Florian PAULER
Max F. Perutz Laboratories
Institute of Genetics
Dr. Bohr-Gasse 9/4
1030 Vienna
Austria
Email: florian.pauler@univie.ac.at

Antoine PETERS
Friedrich Miescher Institute for Biomedical Research
epigenetics
Maulbeerstrasse 66
4058 Basel
Switzerland
Email: Antoine.Peters@fmi.ch

Ana POMBO
MRC Clinical Sciences Centre
Nuclear Organisation Group
Imperial College
Du Cane Road
London W12 0NN
United Kingdom
Email: ana.pombo@csc.mrc.ac.uk

Nikolaus RAJEWSKY
Max Delbrueck Center for Molecular Medicine
Robert-Rossle-Str. 10
13092 Berlin
Germany
Tel: +49 30 9406 2999
Email: rajewsky@mdc-berlin.de

Marc REHMSMEIER
CeBiTec
Universitaet Bielefeld
33594 Bielefeld
Germany
Email: marc@techfak.uni-bielefeld.de

Dirk SCHUEBELER
Friedrich Miescher Institute for Biomedical Research
epigenetics
Maulbeerstrasse 66
4058 Basel
Switzerland
Email: dirk.schubeler@fmi.ch

Mathew SLOANE
Max F. Perutz Laboratories
Institute of Genetics
Dr. Bohr-Gasse 9/4
1030 Vienna
Austria
Email: mathew.sloane@univie.ac.at

Genevieve THON
Department of Molecular Biology, BioCenter
University of Copenhagen
Ole Maaløes Vej 5
2200 Copenhagen
Denmark
Email: gen@my.molbio.ku.dk

Bas VAN STEENSEL
Netherlands Cancer Institute
Plesmanlaan 121
1066 CX Amsterdam
Netherlands
Email: b.v.steensel@nki.nl

Mihaela ZAVOLAN
Biozentrum
University of Basel
Klingelbergstrasse 50/70
5046 Basel
Switzerland
Email: mihaela.zavolan@unibas.ch