Karolinska Institutet

Institutionen för Cell och Molekylärbiologi

Inferring transcriptional regulation on the promoter level and its applications to diseases

AKADEMISK AVHANDLING
som för avläggande av medicine doktorsexamen vid Karolinska Institutet offentligen försvaras i Föreläsningssal Jacob Berzelius (Adam), Berzelius väg 3.

Fredagen den 24 maj, 2013, kl 09.30

av
Morana Vitezic

Huvudhandledare:
Dr. Carsten O. Daub
Karolinska Institutet
Institutionen för Biovetenskaper och Näringslära

Bihandledare:
Professor Björn Andersson
Karolinska Institutet
Institutionen för Cell och Molekylärbiologi

Fakultetsopponent:
Dr. Ali Mortazavi
Univeristy of California Irvine
Department of Developmental and Cell Biology

Betygsnämnd:
Professor Jussi Taipale
Karolinska Institutet
Institutionen för Biovetenskaper och Näringslära

Dr. Thomas Svensson
Stockholm Universitet
Department of Biochemistry and Biophysics

Dr. Peter Swoboda
Karolinska Institutet
Institutionen för Biovetenskaper och Näringslära

Stockholm 2013

ABSTRACT

Gene regulation is important in maintaining cell identity and in higher organisms is a very complex process with many layers of regulation. Genome-wide transcriptional studies that define gene expression across different cells and tissues give important insights into overall gene regulation of a cell as well as the impact of dysregulation in diseases. With the recent advances of high-throughput sequencing methods, it has become increasingly feasible to elucidate transcriptional regulation in the cell, under normal conditions or during cell perturbation.

The aim of this thesis is, using these genome-wide profiling methods, to study in depth the regulatory promoter regions.

In Paper I, we knocked down 4 transcription factors in the THP-1 cell line and applied Cap Analysis of Gene Expression (CAGE) with sequencing. We were able to elucidate *de-novo* the transcriptional binding motifs of these 4 transcription factors as well as build perturbation driven gene regulatory networks. In Paper II, we utilized a similar approach on DYX1C1, a dyslexia susceptibility gene. Using perturbation studies and gene expression profiling with microarrays, the perturbed genes corresponded to the previously described neuronal migration phenotype that was speculated to be linked to the function of this gene. Furthermore, using mass spectrometry, we were able to identify novel protein interacting partners for DYX1C1 and combining with already available data build protein level interaction network. In Paper III, relying on the post-mortem brain samples from the FANTOM5 project, and using CAGE in conjunction with a single molecule sequencer, we identified brain specific transcriptional start sites and brain specific alternative promoters. Additionally, we identified differences between adult and infant brain, interestingly noting many of them originating from alternative promoters. We also classified differences between 15 brain regions into 4 distinct groups and built underlying transcription factor interaction networks. In Paper IV, using the FANTOM5 database we investigated the promoter structure and expression of 3 genes implicated in Rett syndrome. We identified novel promoters, silencing of FOXG1 in the cerebellum, as well as the low correlation between MECP2 and FOXG1 expression. Interestingly, although expression of FOXG1 is limited to the brain and MECP2 is ubiquitous, MECP2 motif activity is significantly lower in the brain than in other tissues, while no differences were observed for FOXG1 motif activity.

In summary, our genome-wide studies employing quantitative gene expression measures on promoter level resolution let us describe how cells are different, let us obtain insights into likely underlying regulatory mechanisms as well as gave us the opportunity to explore diseases.

ISBN 978-91-7549-123-3