ChipSeq

Technique and science

The genome wide dynamics of the binding of Ldb1 complexes during erythroid differentiation

Wilfred van IJcken
Sequencing Seminar Illumina September 8
Scheme of embryonic erythropoiesis
E. Soler, F. Grosveld

pluripotent stem cells
multipotent progenitors / committed precursors
mature cells

HSC

Ldb1
Runx1
Tal1
Lmo2

BFU-E

Ldb1
Gata2
Tal1
PU.1

CFU-E

Ldb1
Gata1
Tal1
Gfi1-b

Meg/E

Eto2 ?

Ldb1
Gata1
Tal1
Lmo4
Eklf

erythrocytes

Proerythroblast
“Erythrocyte-like”

+DMSO

MEL

Tagged transcription factors

Affinity purification
Mass spectrometry

TF complex composition

ChIP
Massively parallel sequencing
Map to genome

Bioinformatic analysis
Identification of target genes
Gene regulation mechanisms

Microarrays
Expression analysis
Basic Questions

- Multiprotein complex composition (Mass Spectrometry)
- Identification of target genes (ChIP-Seq, Microarray)
- Mechanisms of gene regulation

I. TF Networks
   Mass Spectrometry
   ChIP-Seq

II. Functional Analysis of TF Networks
   Microarrays, ChIP-Seq, Anchor-Away

III. Long-Range Interactions
    3C-Seq
Single-step protein complex capture

Mouse cells or tissue

Expression of tagged TF

Streptavidin Dynabeads

Capture with magnetic beads

On-beads trypsin digest

LC-MS/MS analysis

Validation (Co-IP, IF, MO ...)

Protein partners Identification

BirA

biotin

TF

Bio Flag V5

Pres

TEV

Capture with magnetic beads
The essential factor Ldb1

- Mouse Ldb1 KO dies between E9.5 and E10.5 from a series of developmental defects, including absence of hematopoiesis

- Ldb1 is an adaptor protein mediating interactions between transcription factors and their co-regulators

- Ldb1 is thought to mediate long range interactions on the chromatin fiber
Physical interactions in erythroid cells

de Boer et al, PNAS 2003; Rodriguez et al, EMBO J 2005; Meier et al, Development 2006

2nd round of tagging / MS analysis
Physical interactions in erythroid cells

Chromatin remodeling complexes
(ACF/WCRF, MeCP1/NuRD, BHC)

TARGET GENES ?
The essential factor Ldb1

- Mouse *Ldb1 KO* dies between E9.5 and E10.5 from a series of developmental defects, including absence of hematopoiesis

- Ldb1 is an *adaptor protein* mediating interactions between transcription factors and their co-regulators

- Ldb1 is thought to mediate long range interactions on the chromatin fiber

Target genes identification by ChIP-Sequencing

de Boer *et al*, PNAS 2003
Rodriguez *et al*, EMBO J 2005
Meier *et al*, Development 2006
ChIP-Seq procedure

1- Crosslinking (formaldehyde)

2- Random DNA shearing (sonication)

3- IP (antibody)

4- Decrosslinking and DNA purification

5- Ultra-high throughput sequencing
   $\approx 8,000,000$ seq per lane

6- Alignment of the sequence reads to the reference genome

Identification of unknown binding sites
Chip optimization

• Sonication of Fixed Cells
  – DNA fragment size 200 - 1,000bp range optimal
  – Overshearing DNA fragments (<200bp) -> low – yield
  – Undershearing of DNA will result in inefficient immunoprecipitation.

• Antibody Qualification / Titration
  – effectively immunoprecipitate protein-DNA complexes
  – target-specific signal in Western blot or gel-shift assay is necessary but not sufficient
  – the amount of antibody need to be empirically determined.

• PCR amplification of ChIP DNA
  – PCR amplification on a known binding-site region for that protein
ChIP-Seq data visualization
B.Lenhard, JC.Bryne, S.Thongjuea

11,163,593 seq. mapped to the genome in Ldb1 ChIP-Seq
24,810,454 seq. mapped to the genome in Gata1 ChIP-Seq
4,574,247 seq. mapped to the genome in Tal1 ChIP-Seq
31,462,144 seq. mapped to the genome in Eto2 ChIP-Seq
61,194,252 seq. mapped to the genome in Mtgr1 ChIP-Seq
Gfi1b, LSD1, CoREST, Runx1, p300, CTCF…..

E.Soler, C.Andrieu, E.de Boer

Genome Analyzer II
(C.Kockx)

Ldb1: 4982 binding sites
Gata1: 5205 binding sites
Tal1: 4173 binding sites

Your favorite factor is not where you think it is…
ChIP-Seq validation

- Identification of virtually all known target genes of the TF studied
- Good correlation between ChIP and ChIP-Seq
- Overlapping and Specific binding sites for the different factors of the Ldb1 complex

- Autoregulatory loops: all TF in the complex regulate their own expression
- Association of known DNA binding motifs with peaks

Gata1 only peaks  Ldb1/Gata1/Tal1/Eto2/Mtgr1 peaks

Further selection of positive hits

- Cross ChIP-Seq data with expression array data (RNAi for each factor)
- Functional classification of target genes (GO terms / KEGG pathways)
- Is there a specific function for each (sub)-complex?
The different Ldb1 complexes
The Ldb1/Eto2 complex
The Ldb1/Eto2 complex binds developmentally poised genes

- **Slc22a4**
  - Ldb1
  - Gata1
  - Tal1
  - Eto2

- **EpB4.2 (Band4.2)**
  - Ldb1
  - Gata1
  - Tal1
  - Eto2

- **Alas2**
  - Ldb1
  - Gata1
  - Tal1
  - Eto2

- **Gypa (Glycophorin A)**
  - Ldb1
  - Gata1
  - Tal1
  - Eto2
The Ldb1/Eto2 complex regulates developmentally poised genes

**Erythroid progenitors**

**Differentiated cells**

**Ldb1/Eto2 ratio**
The Ldb1/Eto2 complex regulates developmentally poised genes

Eto2 mediates the repression of poised erythroid genes \textit{in vivo}
Proerythroblast  Erythrocyte

Expression (log ratio)

1.7
-0.9

MEL
MEL Ind.
Genome-wide analysis

C. Steinhoff
The Ldb1/Eto2 complex regulates developmentally poised genes

- Ldb1
- Gata1
- Tal1
- Eto2
- Mtgr1

Peak position relative to TSS

Expression fold change

Gata1/Ldb1/Tal1/
Eto2/Mtgr1

p=0, chi-squared test

Ldb1/Gata1

Gata1 only

Peak position relative to TSS
Towards a biological function of the different complexes

Gene Ontology Analysis of Target Genes:

**Activation of the late erythroid program**
- Red cell membrane proteins
- Heme synthesis enzymes and Globin genes
- Erythroid transcription factors

Control of signal transduction pathways
Control of cell proliferation

Repression of alternative lineage genes
- Lymphoid genes

Towards a biological function of the different complexes

Gene Ontology Analysis of Target Genes:

**Activation of the late erythroid program**
- Red cell membrane proteins
- Heme synthesis enzymes and Globin genes
- Erythroid transcription factors

Control of signal transduction pathways
Control of cell proliferation

Repression of alternative lineage genes
- Lymphoid genes
Does ChIP-Seq tell anything about long-range interactions?
Ldb1 ChIP-Seq

Hbb-b1

Ldb1 ChIP-Seq

Klf3

MTGR1
Eto2
LMO
Ldb1
Lmol
Cdk9
E2.2
Ly1
E2A
Tal1
Gata1
Ldb1
Ldb1
Gata1
3C-Seq: Chromosome Conformation Capture-Seq

Klf3

210kb

3C-Seq: Chromosome Conformation Capture-Seq

Fetal liver cells → Crosslink → Digest → Ligate → PCR → Next Generation Sequencing
3C-Seq: Chromosome Conformation Capture-Seq

- Hind III digestion of cross-linked DNA
- Dilution
- Ligation of cross-linked fragments
- De-cross-link
- Circularize (ligate)
- PCR amplify
- Next generation sequencing
- Alignment to the reference genome
3C-Seq: Chromosome Conformation Capture-Seq

Fetal liver cells → Crosslink → Digest → Ligate → PCR → Next Generation Sequencing

Klf3

210kb

Klf3

Ldb1

ChIP-Seq
3C-Seq: Chromosome Conformation Capture-Seq

UCSC Genes Based on RefSeq, UniProt, GenBank, CGAP and Comparative Genomics

RefSeq Genes
Ensembl Gene Predictions
HindIII Restriction Site
Ldb1 fetal liver

KLFS 12.6 Significant of HindIII Fragment [Liver]

KLFS 12.6 Significant of HindIII Fragment [Brain]
To generate real networks, the dynamic changes in genetic and biochemical interactions need to be known. The networks will be incredibly complicated and will be very difficult to generate with the current tools, in particular, quantification is problematic.
Acknowledgements

Department of Cell Biology

Eric Soler, Frank Grosveld

Ralph Stadhouders
Charlotte Andrieu-Soler
Ernie de Boer
Robert-Jan Palstra
Ruud Jorna
Irem Baymaz
Mary Stevens

Soler et al, Genes Dev. 2010 Feb 1;24(3):277-89
Soler et al, Methods 2010 in press

Collaborations

Bergen Center for Computational Science
Boris Lenhard
Jan Christian Bryne
Supat Thongjuea

Research Institute of Molecular Pathology
Meinrad Busslinger

Christine Steinhoff

Francis Stewart

William Skarnes

Erasmus MC
University Medical Center Rotterdam

Cells into Organs

EuTRACC

Marie Curie Actions
Human resources and mobility