ChIP-Seq, RNA and genome sequencing in the IGBMC high-throughput sequencing platform

Céline Keime
Outline

- Presentation of the platform

- ChIP-Seq and 3’end RNA-Seq
  - Role of MITF in melanoma

- Whole RNA sequencing
Outline

- Presentation of the platform
- ChIP-Seq and 3’end RNA-Seq
  - Role of MITF in melanoma
- Whole RNA sequencing
IGBMC
microarray and sequencing platform

- Based in Strasbourg
- Created in 2000
- IBiSA National Platform
- ISO certified since 2007
- Microarray
  - Affymetrix
  - Agilent
  - 6 people team (lab : 5, biostatistics : 1)
- High Throughput Sequencing
  - Illumina GA II X (since october 2008)
  - Server : 4 Operon 4 cores, 64Go memory
  - 4 people team (lab : 2, bioinformatics & biostatistics : 2)

http://www-microarrays.u-strasbg.fr
Since the implementation of the platform (1,5 year)

- ~ 50 projects
- ~ 350 samples
- ~ 60 runs
  - From 18 to 72 bp
IGBMC sequencing platform

- Main applications

- 3'end RNA sequencing: 15%
- whole RNA sequencing: 5%
- small RNA sequencing: 10%
- whole genome sequencing (de novo): 1%
- whole genome re-sequencing: 9%
- targeted re-sequencing: 1%
- ChIPseq: 59%

Mutation analysis
From sample to sequences

Biological questions and experimental design

- Sample quality control
  - DNA: qubit
  - RNA: nanodrop + bioanalyzer

- Library preparation and quality control
  - As recommended by Illumina

- Illumina pipeline to derive sequences from images

- ~20-25 million sequences in 1 lane
Sequences quality control

Generally

Whole RNA sequencing

Kasper et al., 2010
Jun et al., 2010
Outline

- Presentation of the platform
- ChIP-Seq and 3’end RNA-Seq
  - Role of MITF in melanoma
- Whole RNA sequencing
Role of MITF in melanoma?

- **MITF**
  - Microphthalmia transcription factor
  - Essential for the survival of melanocytes
  - Regulates multiple aspects of normal melanocyte function
  - Important role in regulating the proliferative and invasive properties of melanoma cells

- Comprehensive view of MITF role in human melanoma cells
  - Generation of human melanoma cell lines stably expressing 3HA-tagged MITF
  - Anti-HA ChIP-Seq on the tagged and native cells
  - Use of siRNAs to knockdown MITF expression in melanoma cells
  - 3’end RNA-Seq on RNA from siMITF and control cells
Analysis of ChIP-Seq data

- Alignment: Eland (Illumina)
- Peak detection: MACS (Zhang et al., 2008)
- Peak annotation: GPAT (Krebs et al., 2008)
  - Web tool
  - Automatically extracts from public databases the annotations around the submitted positions
  - Access to the expression status of the corresponding genes (from public or user supplied data)
- Search for pattern in a ChIP-Seq dataset and comparison between different ChIP-Seq datasets: seqMINER (Ye, Krebs et al.)
  - JAVA software
Analysis of ChIPseq data

Ye, Krebs et al.
MITF occupancy of the TYR locus
Location of MITF occupied loci relative to TSS

- Preferential location ~150 bp upstream TSS, a region that is often nucleosome free in active promoters.
Comparison of read density around the MITF-occupied loci
Analysis of 3’end RNA sequencing data

- Tag identification
  - Comparison of experimentally obtained tags with a set of “virtual tags” computed from known sequences
- Normalization
  - Estimation of a size factor for each library (Anders et al.)
- Identification of significantly differentially expressed genes
  - Negative binomial model (Anders et al.)
  - Adjustment for multiple testing (Benjamini and Hochberg, 1995)
- Use of Gene Ontology and pathways to interpret results
Gene expression changes in siMITF cells

Gene Ontology
732 Up-regulated genes

<table>
<thead>
<tr>
<th>Category</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cytoplasm</td>
<td>21.7%</td>
</tr>
<tr>
<td>Signal</td>
<td>20.2%</td>
</tr>
<tr>
<td>Cytoskeleton</td>
<td>4.4%</td>
</tr>
</tbody>
</table>

Gene Ontology
613 Down-regulated genes

<table>
<thead>
<tr>
<th>Category</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Nucleus</td>
<td>37.5%</td>
</tr>
<tr>
<td>Cell cycle</td>
<td>8.3%</td>
</tr>
<tr>
<td>Cell division</td>
<td>5.2%</td>
</tr>
<tr>
<td>Mitosis</td>
<td>4.4%</td>
</tr>
</tbody>
</table>
Gene expression changes in siMITF cells

Pathways
732 Up-regulated genes

- Focal adhesion: p-value = 0.001
- ErbB pathway: p-value = 0.003
- Cell adhesion: p-value = 0.09
- Glioma: p-value = 0.006
- Colorectal cancer: p-value = 0.02
- TGF beta pathway: p-value = 0.08

Pathways
613 Down-regulated genes

- Cell cycle: p-value < 0.001
- Replication: p-value < 0.001
- Ub-med proteolysis: p-value = 0.06
- H-recombination: p-value = 0.001
- B-E repair: p-value = 0.01
- N-E repair: p-value = 0.04
High concordance between 3’RNAseq and qPCR data

$R^2 = 0.88$
Intersection between MITF-occupied genes from ChIP-Seq and MITF regulated genes following siMITF knockdown
Outline

- Presentation of the platform
- ChIP-Seq and 3’end RNA-Seq
  - Role of MITF in melanoma
- Whole RNA sequencing
Whole RNA sequencing

- Two strategies used for whole RNA sequencing
  - Purification of polyA RNA
  - Depletion in rRNA
    - But: not always complete
    - & highly variable between samples
Analysis of whole RNA sequencing data with a reference genome

- **Reads mapping**: Tophat / Bowtie (Trapnell et al., 2009)
  - Aligns reads to a genome & identify exon-exon splice junctions

- **Assembly of reads into transcripts**: Cufflinks (Trapnell et al., 2010)
  - Constructs a parsimonious set of transcripts that "explain" the reads observed in an RNAseq experiment

- **Estimation of expression level**: Cufflinks (Trapnell et al., 2010)
  - Uses a statistical model of sequencing experiments to derive a likelihood for the abundances of a set of transcripts given a set of reads
Whole RNA sequencing: transcript coverage

- Highly variable number of reads per transcript
  - Between the different transcripts
    - eg for a mouse RNAseq library sequenced in 1 lane (54 bp):
      - $1 \leq nb \leq 15,000$, median ~ 100
  - Within a transcript

Coverage

Exons / introns annotations
RNAseq : transcript coverage

- Global repartition along transcripts
Characterization of splicing events
Future directions

- Reduction of the amount of starting material
- Strand-specific RNA sequencing
- De novo transcriptome assembly
People involved

- **IGBMC sequencing platform**
  - Irwin Davidson, Bernard Jost, Céline Keime, Stéphanie Le Gras, Serge Vicaire

- **MITF project**
  - Thomas Strub, Sandy Giuliano, Tao Ye, Caroline Bonnet, Céline Keime, Dominique Kobi, Stéphanie Legras, Mireille Cormont, Robert Ballotti, Corine Bertolotto and Irwin Davidson


Zhang et al., Model-based analysis of ChIP-Seq (MACS). Genome Biology 2008;9 (9):R137.