Diverse noncoding mutations contribute to deregulation of cis-regulatory landscape in pediatric cancers

Kai Tan
Associate Professor
ITCR Annual Meeting
Award: U01 CA226187
May 30, 2019
Vast majority of variants are non-coding
Transcriptional Enhancers

- Largest class of cis-regulatory elements
- High tissue-specificity
- Long-distance action
- Complicated regulatory grammar

VISTA database (http://enhancer.lbl.gov)
Many ways to break function of an enhancer

- SNV & small indels
- Duplication
- Deletion
- Translocation/Inversion

**PANGEA:** Predictive Analysis of Noncoding Genomic Enhancer/promoter Alterations

expression x noncoding mutation count matrix

<table>
<thead>
<tr>
<th>Patient ID</th>
<th>FPKM</th>
<th>Promoter</th>
<th>Enhancer 1</th>
<th>...</th>
<th>Enhancer k</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>10.4</td>
<td>2</td>
<td>0</td>
<td>2</td>
<td>0</td>
</tr>
<tr>
<td>2</td>
<td>22.3</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>3</td>
<td>0.13</td>
<td>0</td>
<td>1</td>
<td>3</td>
<td>0</td>
</tr>
<tr>
<td>...</td>
<td></td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
</tr>
<tr>
<td>n</td>
<td>11.7</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
</tr>
</tbody>
</table>

 Software available @ https://github.com/tanlabcode/PANGEA

Causal mutations disrupting enhancers & EP interactions
Genome in 3D

enhancer-promoter

Chromatin Loops (5kb resolution)

TADs (10kb resolution)

Compartments (50kb resolution)

Interchromosomal

Adapted from
Dekker et al, Nat. 2017;
IM-PET: Integrated Method for Predicting Enhancer Targets

He et al. *PNAS*. 2014

CD34+ HSPCs, CD4+ T cells, GM12878 HeLa, HMEC, HUVEC, IMR90, K562 MCF7, NHEK
Therapeutically Applicable Research To Generate Effective Treatment

Cancer types
- Acute lymphoblastic leukemia
- Acute myeloid leukemia
- Kidney tumor
- Neuroblastoma
- Osteosarcoma

Matched WGS & RNA-Seq
- 153 patients
- 163 patients
- 53 patients
- 100 patients
- 32 patients
Number of predicted causal noncoding mutations per patient
Worse clinical outcome associated with GFI1B enhancer SNVs in AML
Enhancer hijacking to CHD4 due to translocation in B-ALL

H3K4Me1
H3K27Ac
H3K4Me3
break points

EP300
Chr22

H3K4Me1
H3K27Ac
H3K4Me3
break points

CHD4
ZNF384
Chr12

TCF3
Chr19

H3K4Me1
H3K27Ac
H3K4Me3
break points

SMARCA2
Chr9

enhancer
enhancer
enhancer
CHD4 expression correlates with patient prognosis

**Graphs:**

- **Box plot (b):**
  - CHD4 FPKM
  - P-val = 7e-4
  - Patients w/ TRAN (12)
  - Patients w/o TRAN (151)

- **Kaplan-Meier survival curve (c):**
  - Patients w/ CHD4 Translocation
  - Patients w/o CHD4 Translocation
  - P-val = 0.0075
  - Time to relapse (days)
  - Relapse free
CHD4 is a member of NuRD complex & regulates B cell dev.

Dege & Hagman, Immunol Rev. 2014
CHD4 inhibition impairs cell growth of B-ALL leukemia

B-ALL cells → CHD4 sgRNA-GFP → WT-RFP

Monitoring of RFP+/GFP+ cells by FACS

CHD4-KO WT

NALM-6

Relative cell number (%) vs. Days post transduction

REH

Days post transduction

Chia-hui Chen
Translocation induces CHD4 over-expression and IL3 independent growth of Ba/F3 cells

Peng Gao
Majority of TF regulons are disrupted **NOT** by TF coding mutations
Level of TF regulon perturbation correlates with patient prognosis
Distinct sets of genes perturbed by coding vs. noncoding mutations
Contrasting features of genes affected by coding vs. noncoding mutations

<table>
<thead>
<tr>
<th>Genes affected by coding mutations</th>
<th>Genes affected by noncoding mutations</th>
</tr>
</thead>
<tbody>
<tr>
<td>Exon Length</td>
<td>Exon Conservation</td>
</tr>
<tr>
<td>Long</td>
<td>High</td>
</tr>
<tr>
<td>Short</td>
<td>Low</td>
</tr>
<tr>
<td>Replication timing</td>
<td>Enhancer Degree</td>
</tr>
<tr>
<td>Late</td>
<td>Low</td>
</tr>
<tr>
<td>Early</td>
<td>High</td>
</tr>
<tr>
<td>Enhancer Conservation</td>
<td>Enhancer Conservation</td>
</tr>
<tr>
<td>Low</td>
<td>Low</td>
</tr>
<tr>
<td>High</td>
<td>High</td>
</tr>
<tr>
<td>Enhancer Specificity</td>
<td></td>
</tr>
<tr>
<td>Low</td>
<td></td>
</tr>
<tr>
<td>High</td>
<td></td>
</tr>
</tbody>
</table>
Summary

- PANGEA: general framework to predict risk noncoding mutations of all classes (point mutations & SVs).

- Comprehensive analysis of risk noncoding mutations in 5 major pediatric cancers.

- Coding and noncoding mutations affecting distinct pathways.

- Degree of TF regulon perturbation as a way to stratify patient prognosis.
Genome in 3D

Chromatin Loops
(5kb resolution)

TADs
(10kb resolution)

Compartments
(50kb resolution)

Interchromosomal

Adapted from
Dekker et al, Nat. 2017;
3DOncoGenome Database

Welcome to 4DGenome, a general repository for chromatin interaction data.

Records in 4DGenome are compiled through comprehensive literature curation of experimentally-derived and computationally-predicted interactions. The current release contains 4,433,071 experimentally-derived and 3,605,176 computationally-predicted interactions in 5 organisms. Experimental data cover both high throughput datasets and individual focused studies.

All interaction data are freely available in a standardized file format. Records can be queried by genomic regions, gene names, organism, and detection technology.

Teng et al. *Bioinformatics* 2015
#SLC2A13

1 records from 5C data

2 records from ChIA-PET data

3 records from Hi-C data

20 records from IM-PET data
Acknowledgement

Tan lab
Changya Chen
Chia-hui Chen
Greg Chen
Yang-yang Ding
Kamyar Esmaeili
Peng Gao
Bing He
Hannah Kim
Tao Peng
Yasin Uzun
Wenbao Yu
Qin Zhu

CHOP
Steve Hunger
Sarah Tasian
Kathrin Bernt
John Maris
Sharon Diskin

Funding
NIH
ITCR U01 CA226187