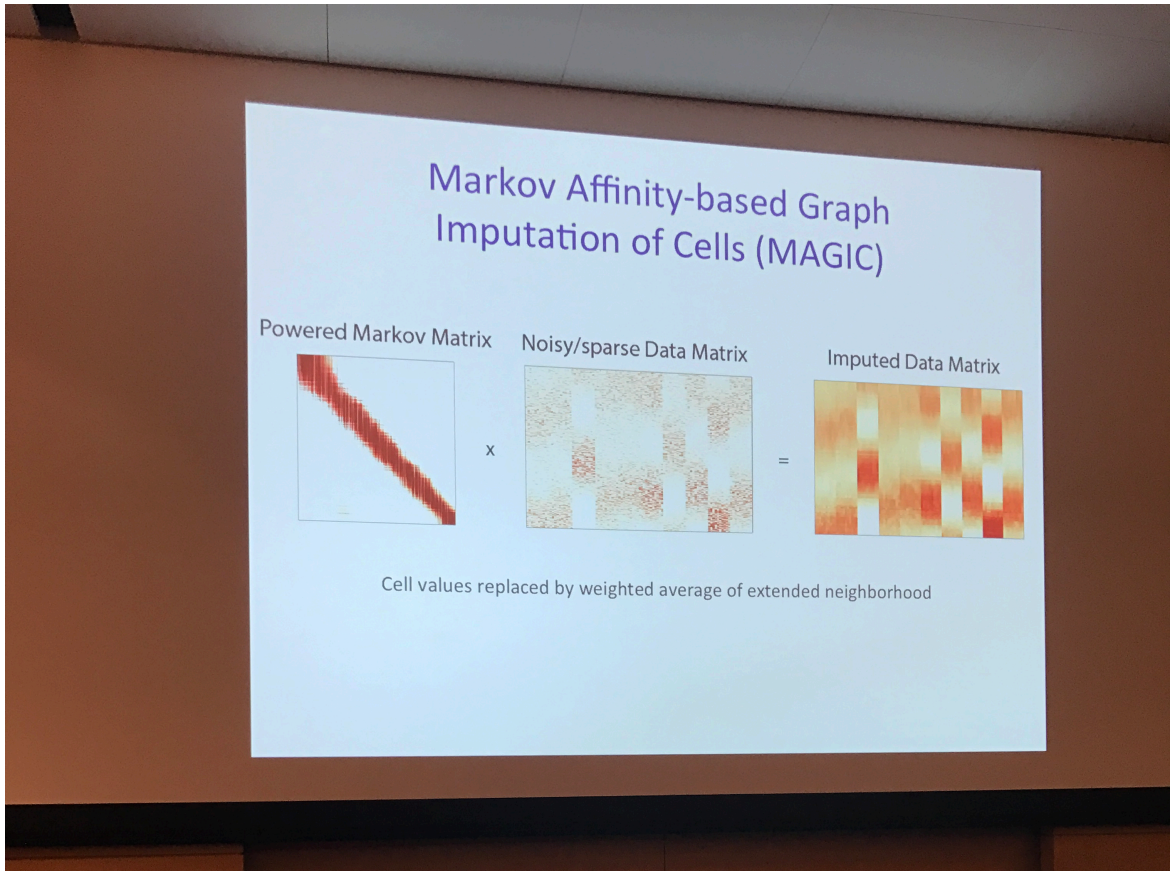
RSGDREAM 17 debrief

JZ



Dana Pe'er Lab of Computational Systems Biology

e organization, function and evolution of molecular networks.
nse multiple signals from the environment, robustly process an
rchestrate the regulation of hundreds of genes and proteins to
able functionality occurs through diverse mechanisms including
tic changes, translation, degradation, post-translational
ent of high throughput genomic and proteomic technologies is
n of new experimental data, quantitatively measuring the
genome-wide scale.

ethods to integrate diverse high throughput data and unravel a
ell. We elucidate the principles by which a cell robustly
n or response to environmental stimuli. Some of the question we
er between cell-type, individual and species? How does the
es of stimuli and over the course of evolution? How do small
ropagate and manifest in phenotypic diversity and changes to
gulation lead to disease such as cancer?

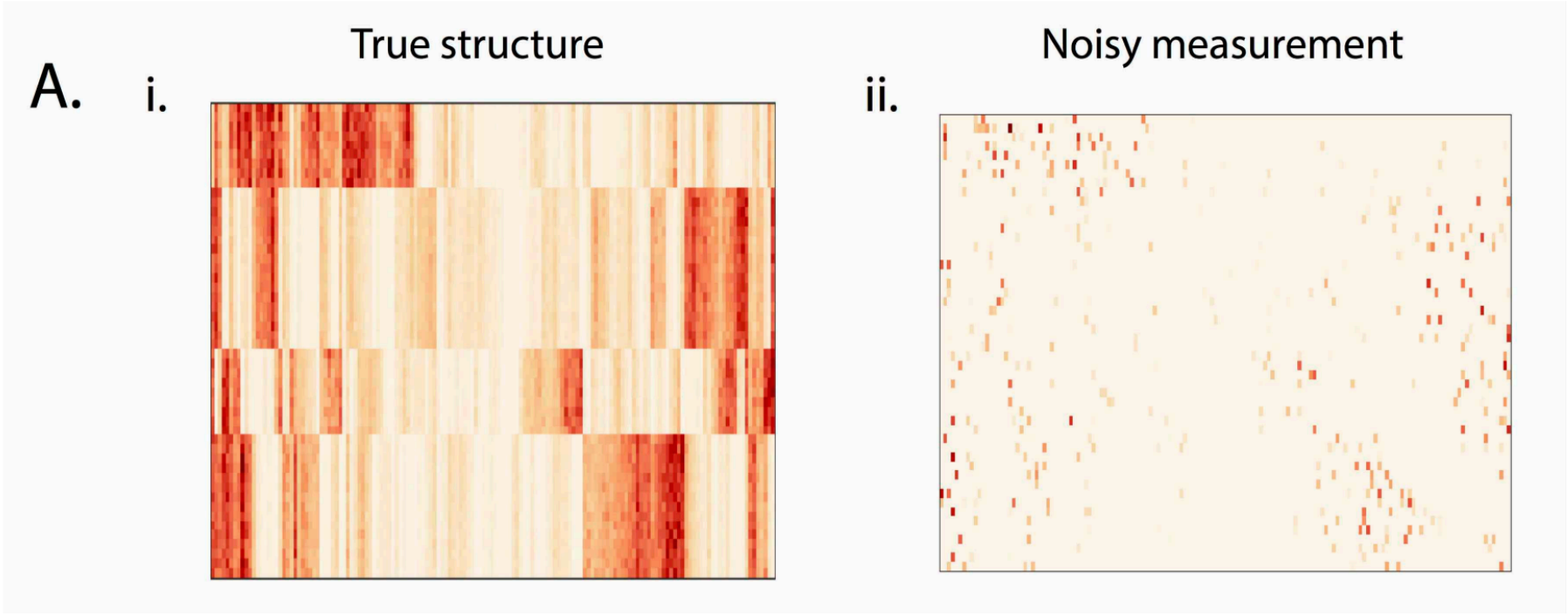
Phone:
(212) 854-4397
Email:
dpeer@biology.columbia.edu

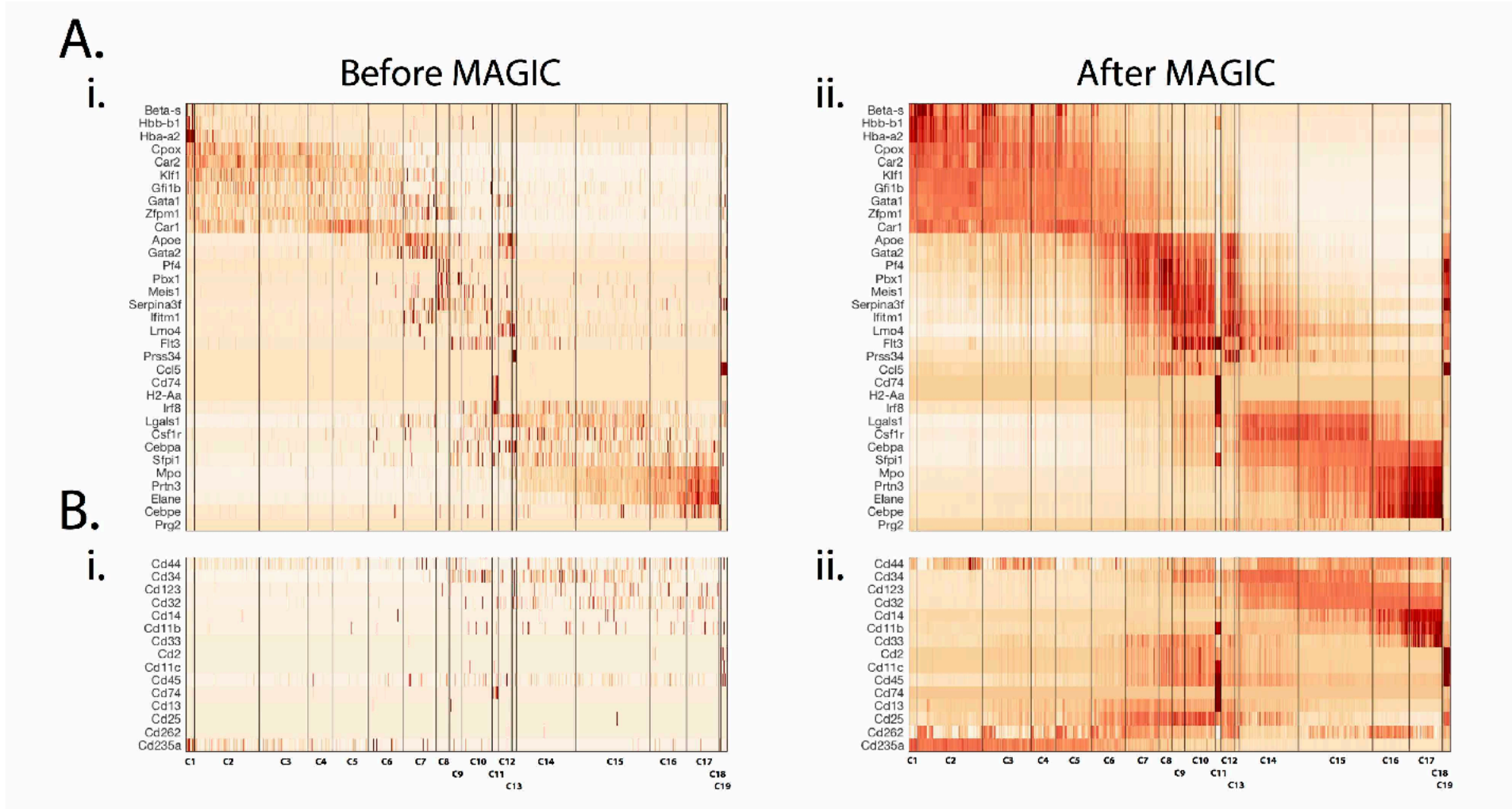
New Results

MAGIC: A diffusion-based imputation method reveals gene-gene interactions in single-cell RNA-sequencing data

David van Dijk, Juozas Nainys, Roshan Sharma, Pooja Kathail, Ambrose J Carr, Kevin R Moon, Linas Mazutis, Guy Wolf, Smita Krishnaswamy, Dana Pe'er

doi: https://doi.org/10.1101/111591




A.
i.
Before MAGIC
ii.
After MAGIC
Beta-s
Hbb-b1
Hba-a2
Cpox
Car2
Klf1
Gfi1b
Gata1
Zfpm1
Car1
Apoe
Gata2
Pf4
Pbx1
Meis1
Serpina3f
Ifitm1
Lmo4
Flt3
Prss34
Ccl5
Cd74
H2-Aa
Irf8
Lgals1
Csf1r
Cebpa
Sfpi1
Mpo
Prtn3
Elane
Cebpe
Prg2
B.
i.
ii.
Cd44
Cd34
Cd123
Cd32
Cd14
Cd11b
Cd33
Cd2
Cd11c
Cd45
Cd74
Cd13
Cd25
Cd262
Cd235a
C1
C2
C3
C4
C5
C6
C7
C8
C9
C10
C11
C12
C13
C14
C15
C16
C17
C18
C19

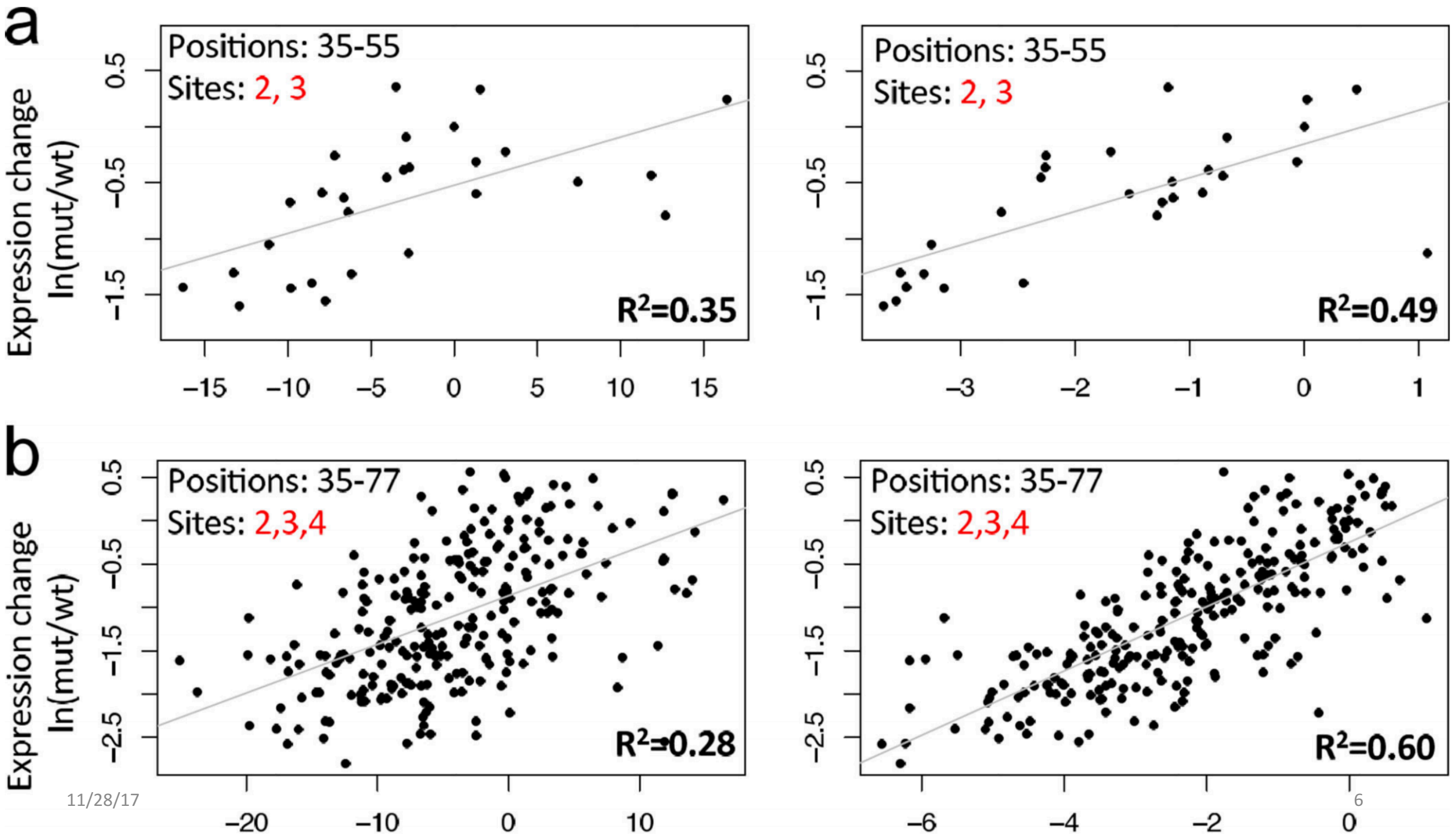
Published in final edited form as:
Res Comput Mol Biol. 2017 May ; 10229: 336–352. doi:10.1007/978-3-319-56970-3_21.

Quantifying the Impact of Non-coding Variants on Transcription Factor-DNA Binding

Jingkang Zhao[1,2,†], **Dongshunyi Li**[3,†], **Jungkyun Seo**[2], **Andrew S. Allen**[1,3], and **Raluca Gordân**[1,3,4]

[1]Center for Genomic and Computational Biology, Duke University, Durham NC 27708, USA

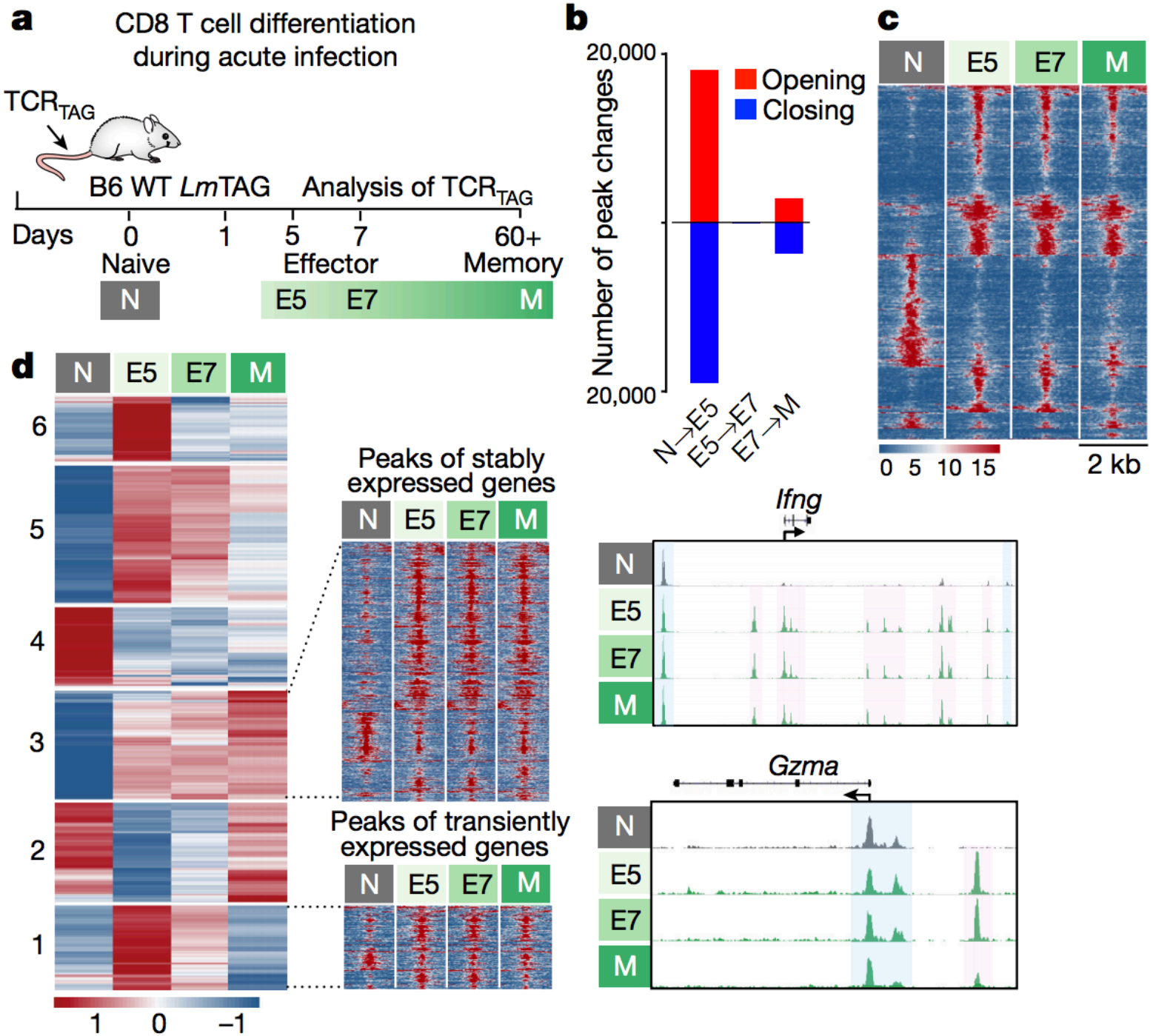
We use ordinary least squares (OLS) to estimate the parameters of the binding model for each TF, and we show that our predictions of TF-binding changes due to DNA mutations correlate well with measured changes in gene expression.

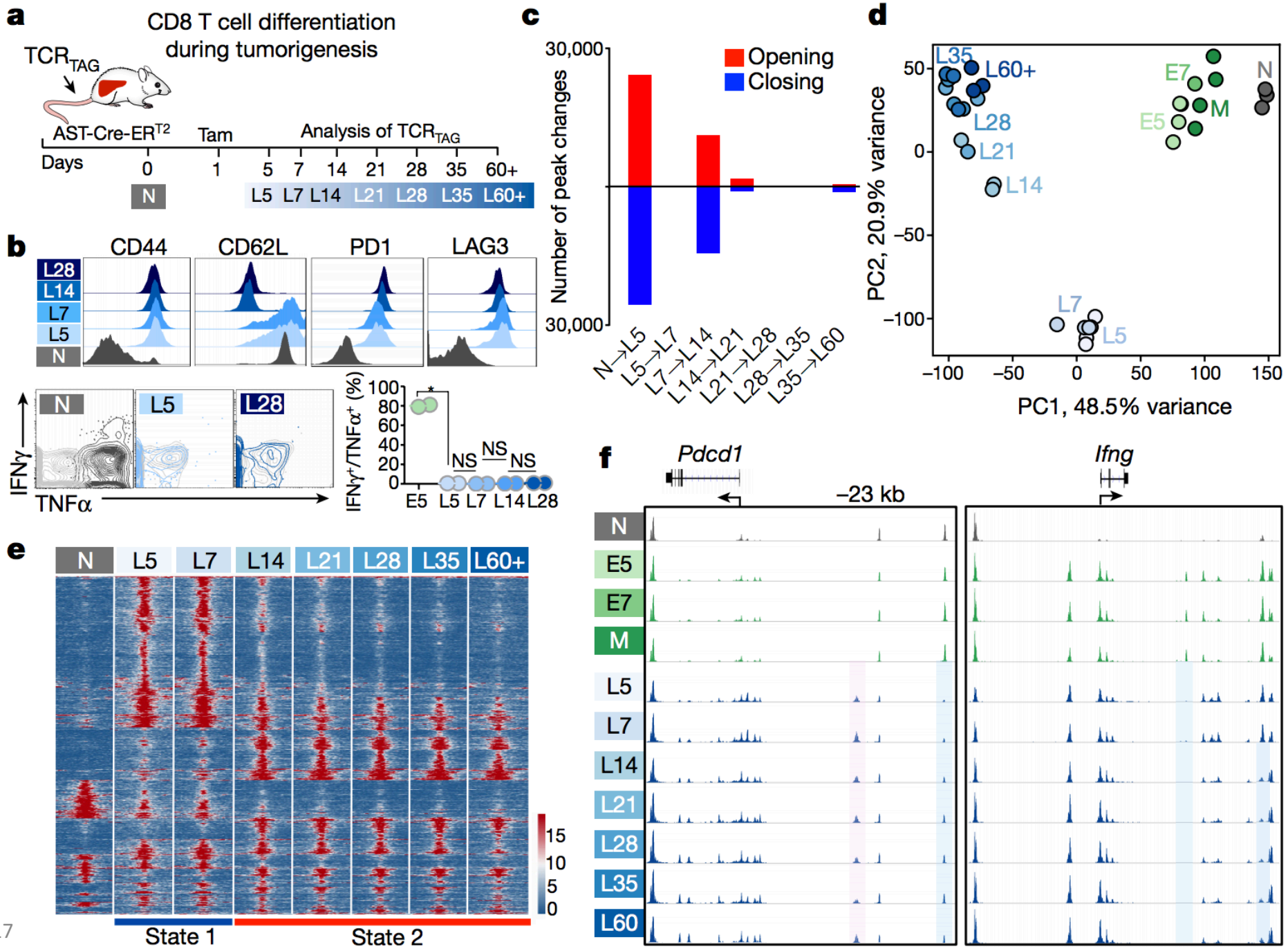

a
Positions: 35-55
Sites: 2, 3
Expression change
ln(mut/wt)
R²=0.35
Positions: 35-55
Sites: 2, 3
R²=0.49
b
Positions: 35-77
Sites: 2,3,4
R²=0.28
Positions: 35-77
Sites: 2,3,4
R²=0.60

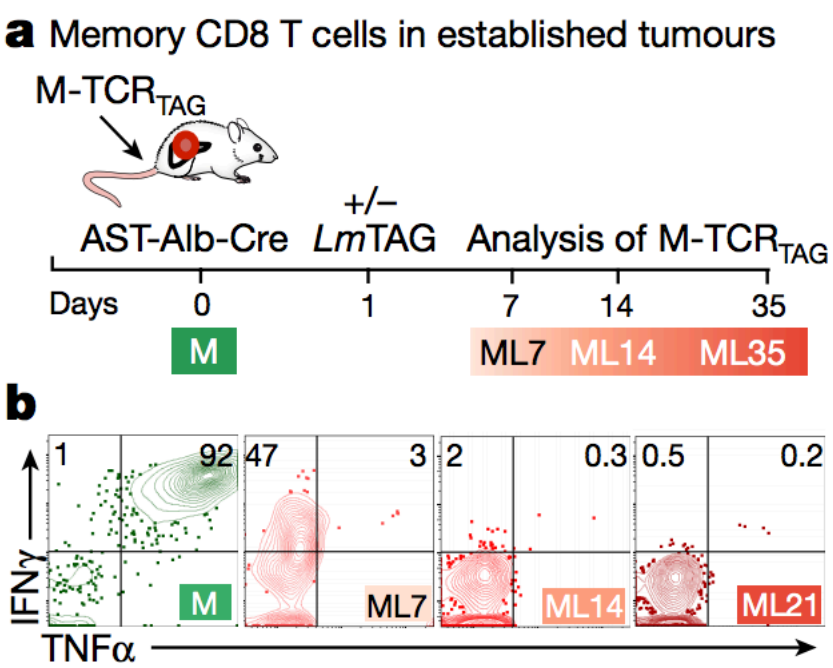
Chromatin states define tumour–specific T cell dysfunction and reprogramming

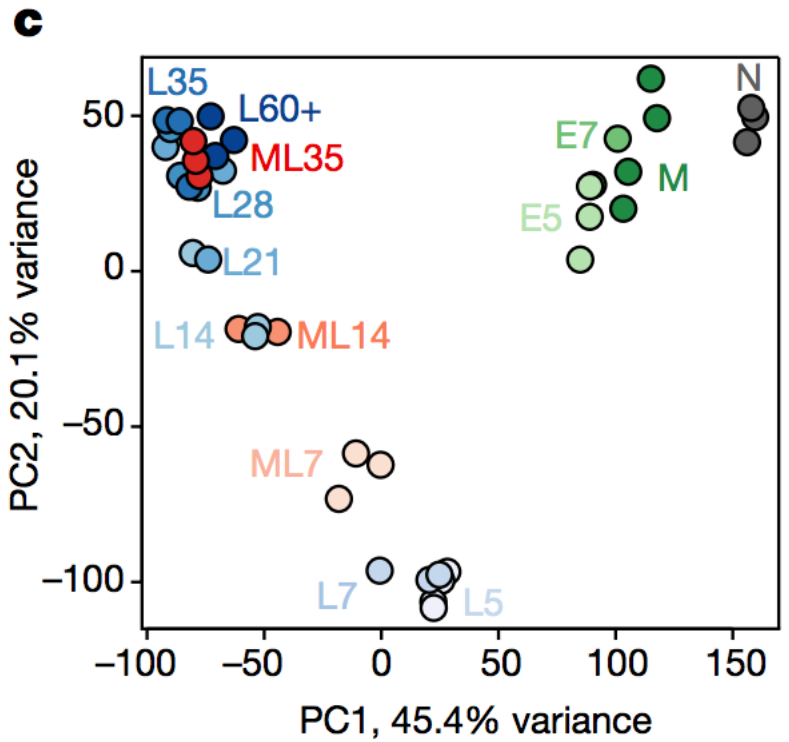
Mary Philip[1], Lauren Fairchild[2,3], Liping Sun[4], Ellen L. Horste[1], Steven Camara[1], Mojdeh Shakiba[1,5], Andrew C. Scott[1,5], Agnes Viale[4], Peter Lauer[6], Taha Merghoub[5,7], Matthew D. Hellmann[5,8], Jedd D. Wolchok[5,7,9], Christina S. Leslie[2] & Andrea Schietinger[1,5]

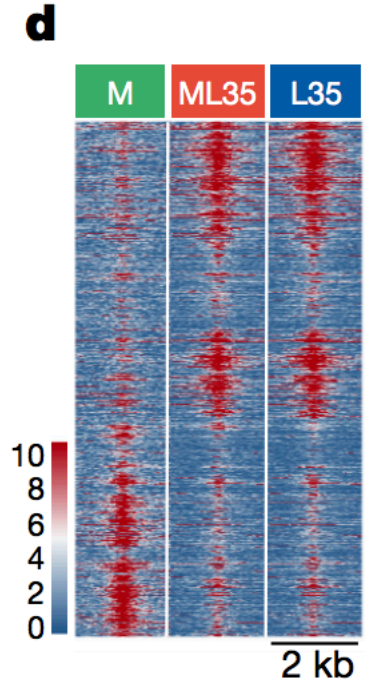
- *DNaseq 20m cells, attic-sea 50k cells, chromatin state to distinct the dysfunction state, 11.17 Nature*
- *changes of chromatin accessibility is corresponding to gene expression changes*
- *compare the changes of TF binding using the imputed network (ISMB?)*

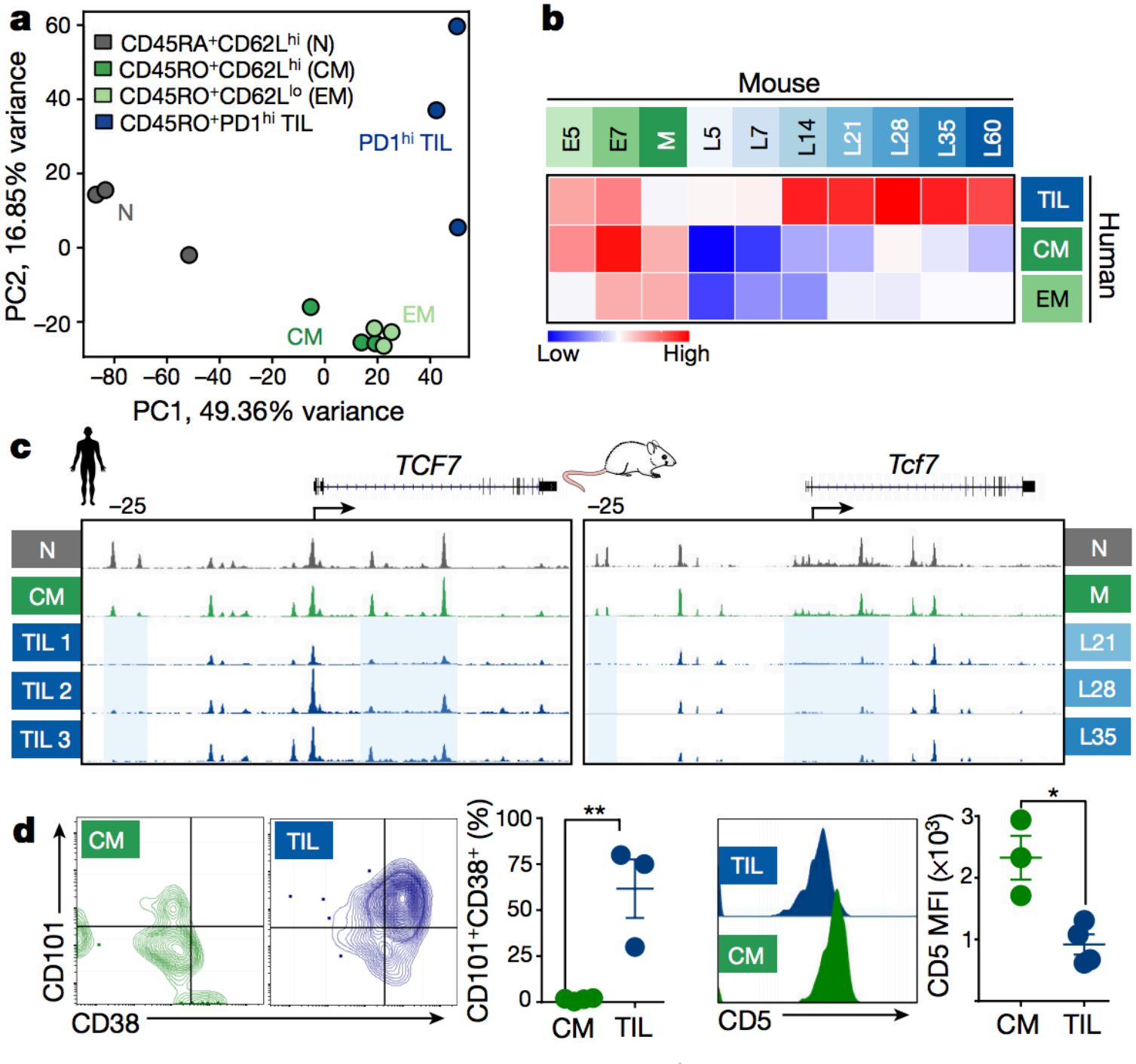

a
CD8 T cell differentiation during acute infection
TCR$_{TAG}$
B6 WT *Lm*TAG
Analysis of TCR$_{TAG}$
Days 0 1 5 7 60+
Naive
Effector
Memory
N
E5
E7
M
b
20,000
Number of peak changes
20,000
Opening
Closing
N→E5
E5→E7
E7→M
c
N
E5
E7
M
0 5 10 15
2 kb
d
N
E5
E7
M
6
5
4
3
2
1
1 0 −1
Peaks of stably expressed genes
Peaks of transiently expressed genes
Ifng
Gzma


a
CD8 T cell differentiation during tumorigenesis
TCR_TAG
AST-Cre-ER^T2
Tam
Analysis of TCR_TAG
Days
0 1 5 7 14 21 28 35 60+
N
L5 L7 L14 L21 L28 L35 L60+
b
CD44
CD62L
PD1
LAG3
L28
L14
L7
L5
N
IFNγ
TNFα
IFNγ+/TNFα+ (%)
NS NS NS
E5 L5 L7 L14 L28
c
Opening
Closing
Number of peak changes
30,000
N→L5
L5→L7
L7→L14
L14→L21
L21→L28
L28→L35
L35→L60
d
PC2, 20.9% variance
PC1, 48.5% variance
L35
L60+
L28
L21
L14
L7
L5
E7
M
E5
N
e
N L5 L7 L14 L21 L28 L35 L60+
State 1
State 2
15
10
5
0
f
Pdcd1
Ifng
−23 kb
N
E5
E7
M
L5
L7
L14
L21
L28
L35
L60

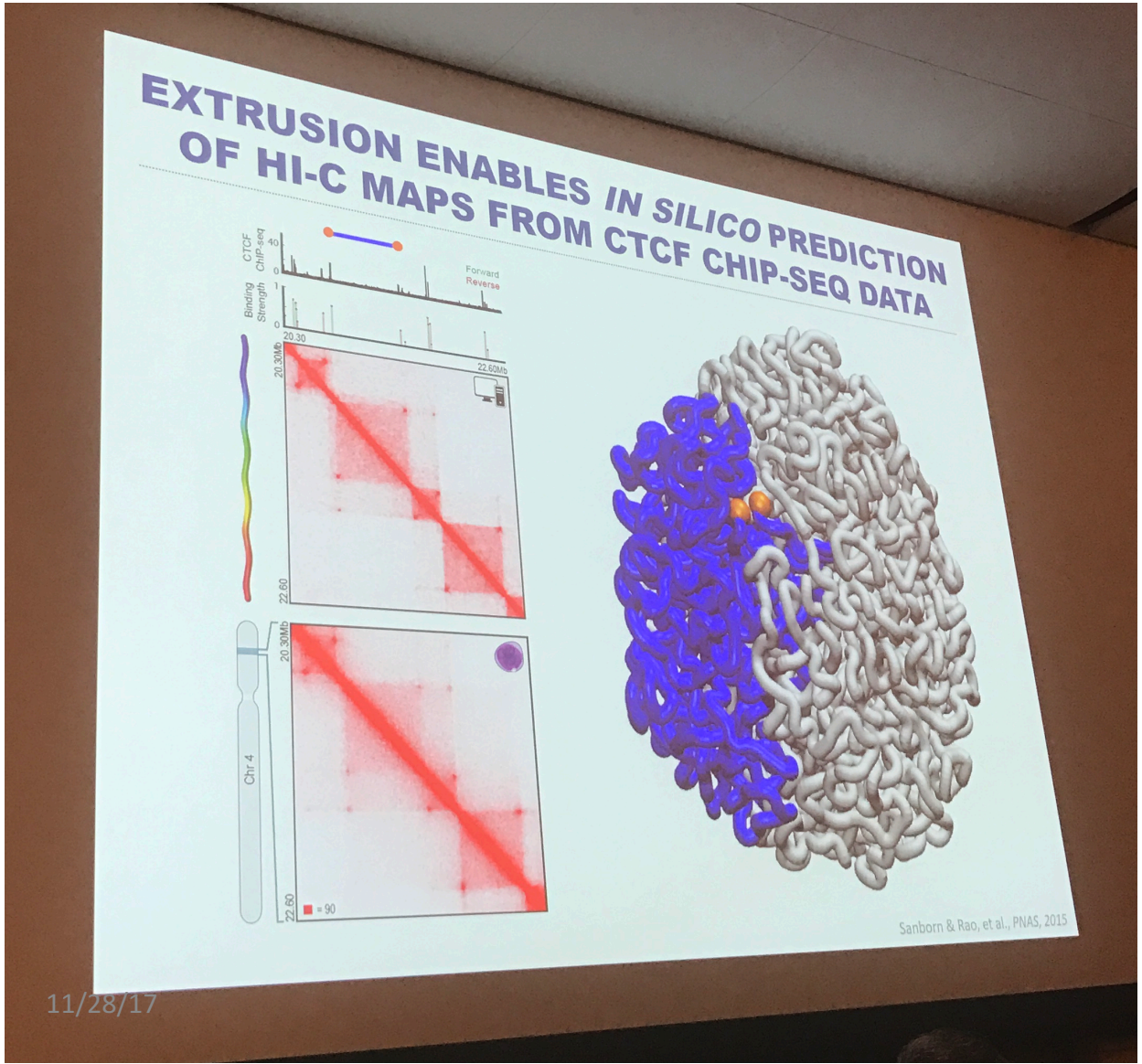

a Memory CD8 T cells in established tumours
M-TCR$_{TAG}$
+/–
AST-Alb-Cre
*Lm*TAG
Analysis of M-TCR$_{TAG}$
Days
0
1
7
14
35
M
ML7
ML14
ML35
b
1
92
47
3
2
0.3
0.5
0.2
IFNγ
TNFα
M
ML7
ML14
ML21


c
L35
L60+
ML35
L28
L21
L14
ML14
ML7
L7
L5
E7
E5
M
N
PC2, 20.1% variance
PC1, 45.4% variance
50
0
–50
–100
–100
–50
0
50
100
150

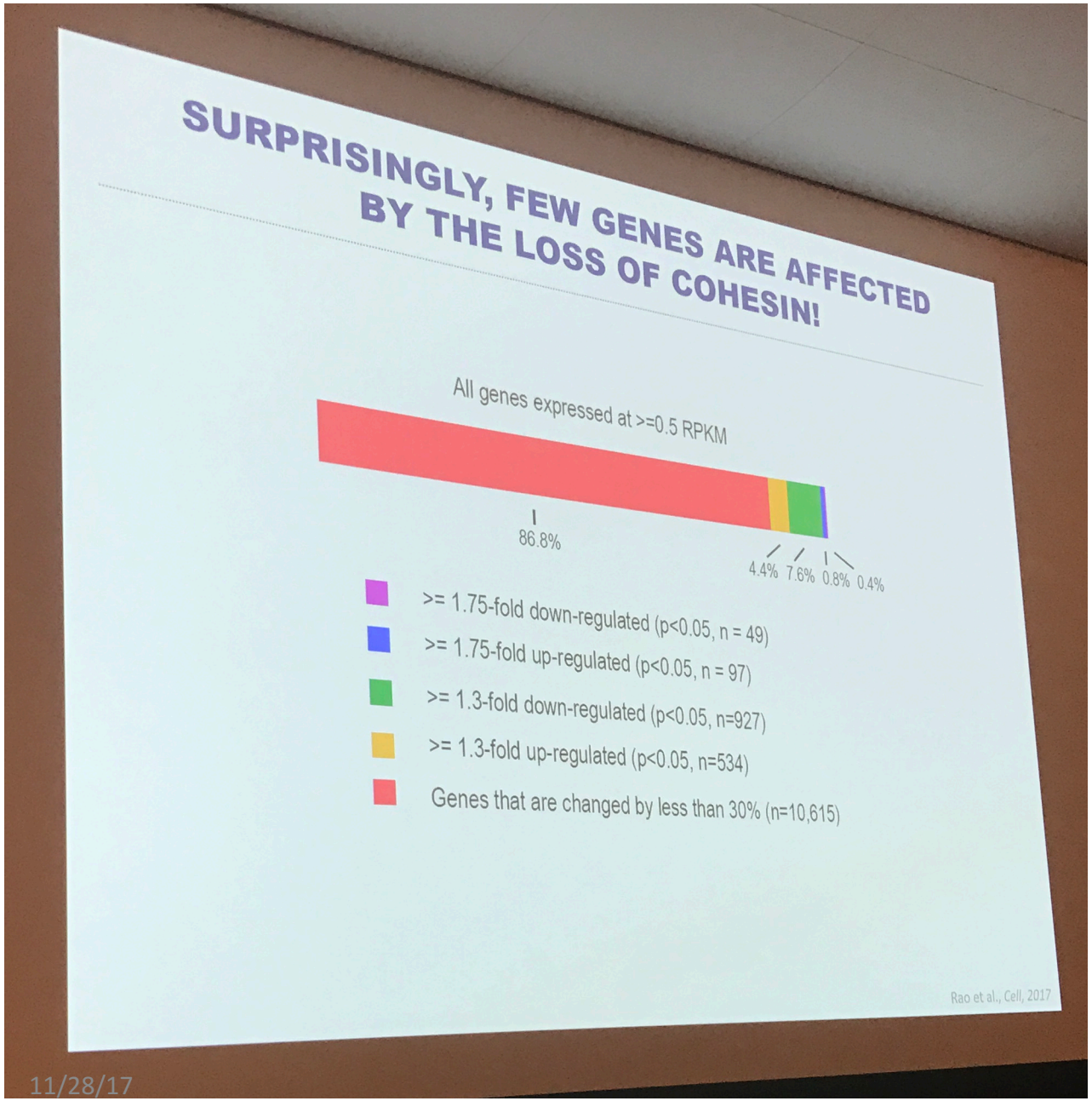

d
M
ML35
L35
10
8
6
4
2
0
2 kb


a
CD45RA+CD62Lhi (N)
CD45RO+CD62Lhi (CM)
CD45RO+CD62Llo (EM)
CD45RO+PD1hi TIL
PD1hi TIL
N
CM
EM
PC2, 16.85% variance
PC1, 49.36% variance
b
Mouse
E5
E7
M
L5
L7
L14
L21
L28
L35
L60
TIL
CM
EM
Human
Low
High
c
TCF7
Tcf7
−25
−25
N
CM
TIL 1
TIL 2
TIL 3
N
M
L21
L28
L35
d
CM
TIL
CD101
CD38
CD101+CD38+ (%)
**
CM
TIL
TIL
CM
CD5
CD5 MFI (×10³)
*
CM
TIL

A 3D Code in the Human Genome

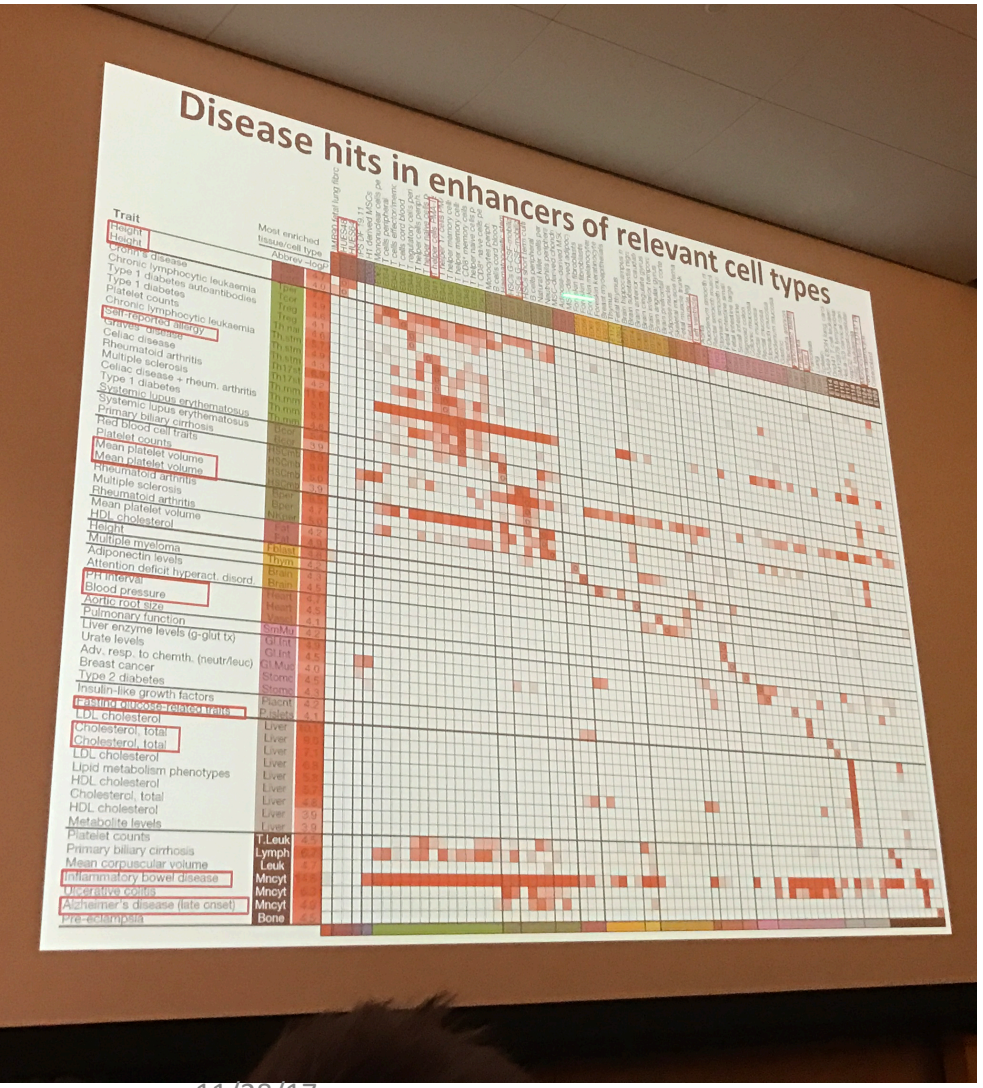


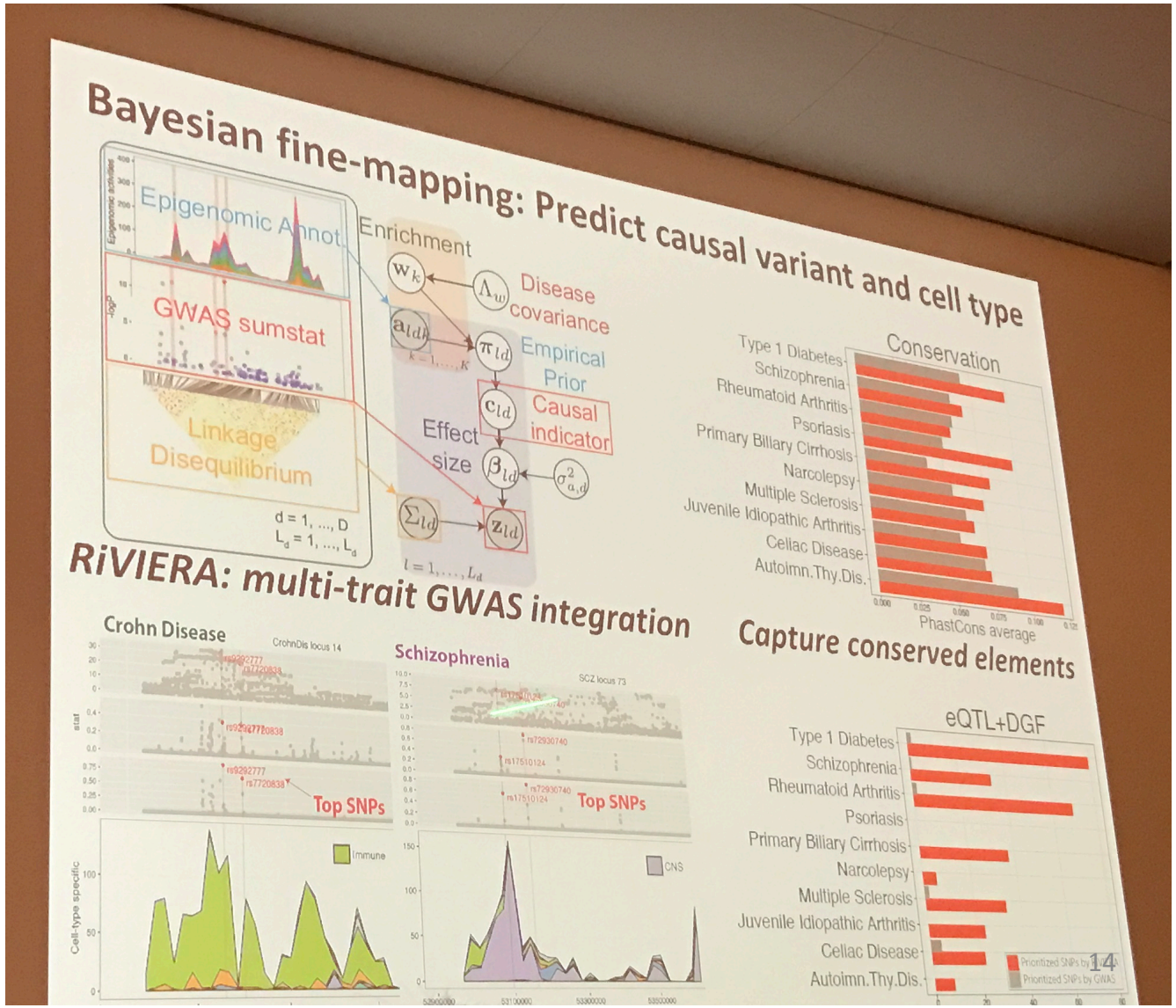
CTCF and Hi-C

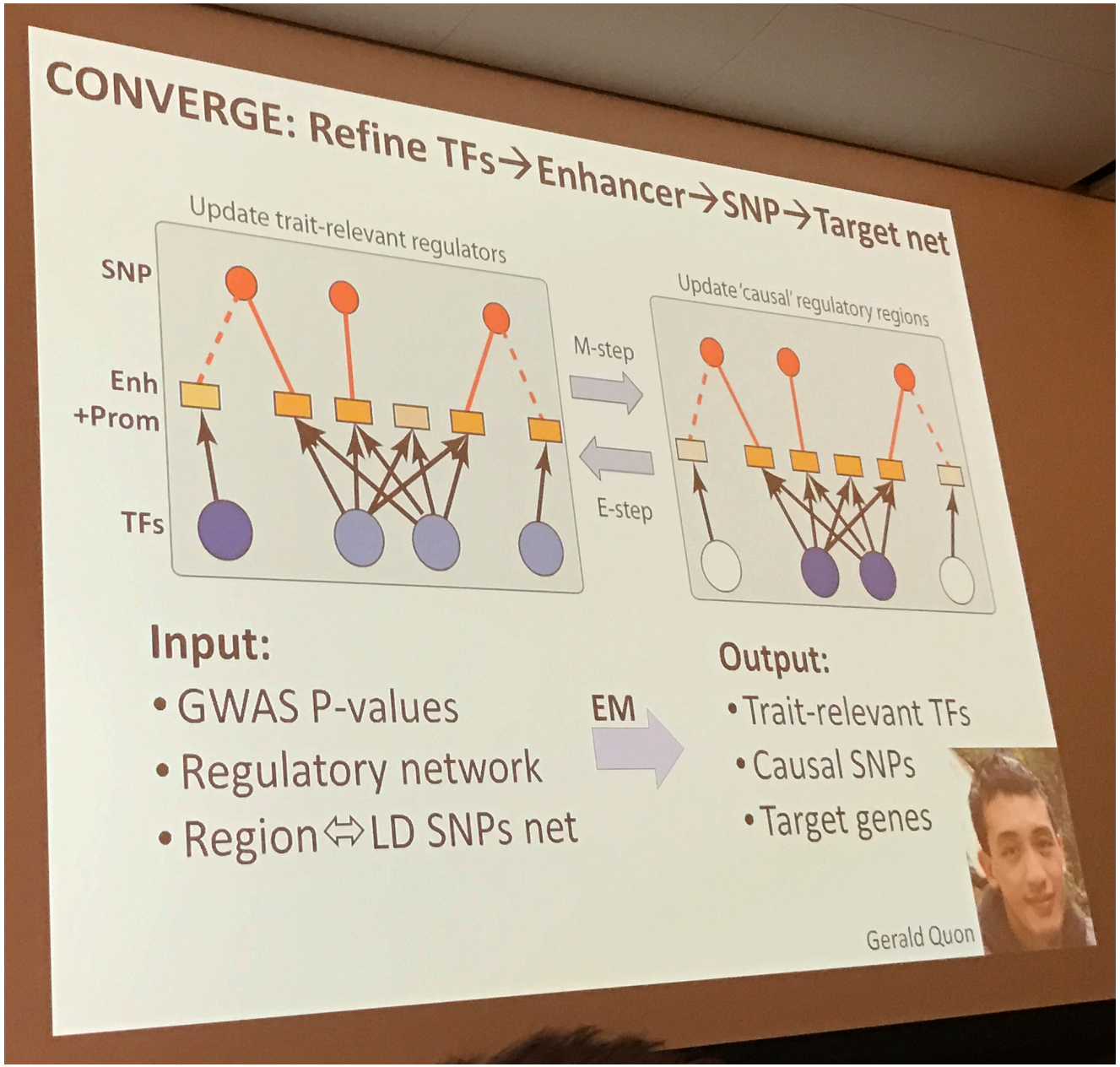


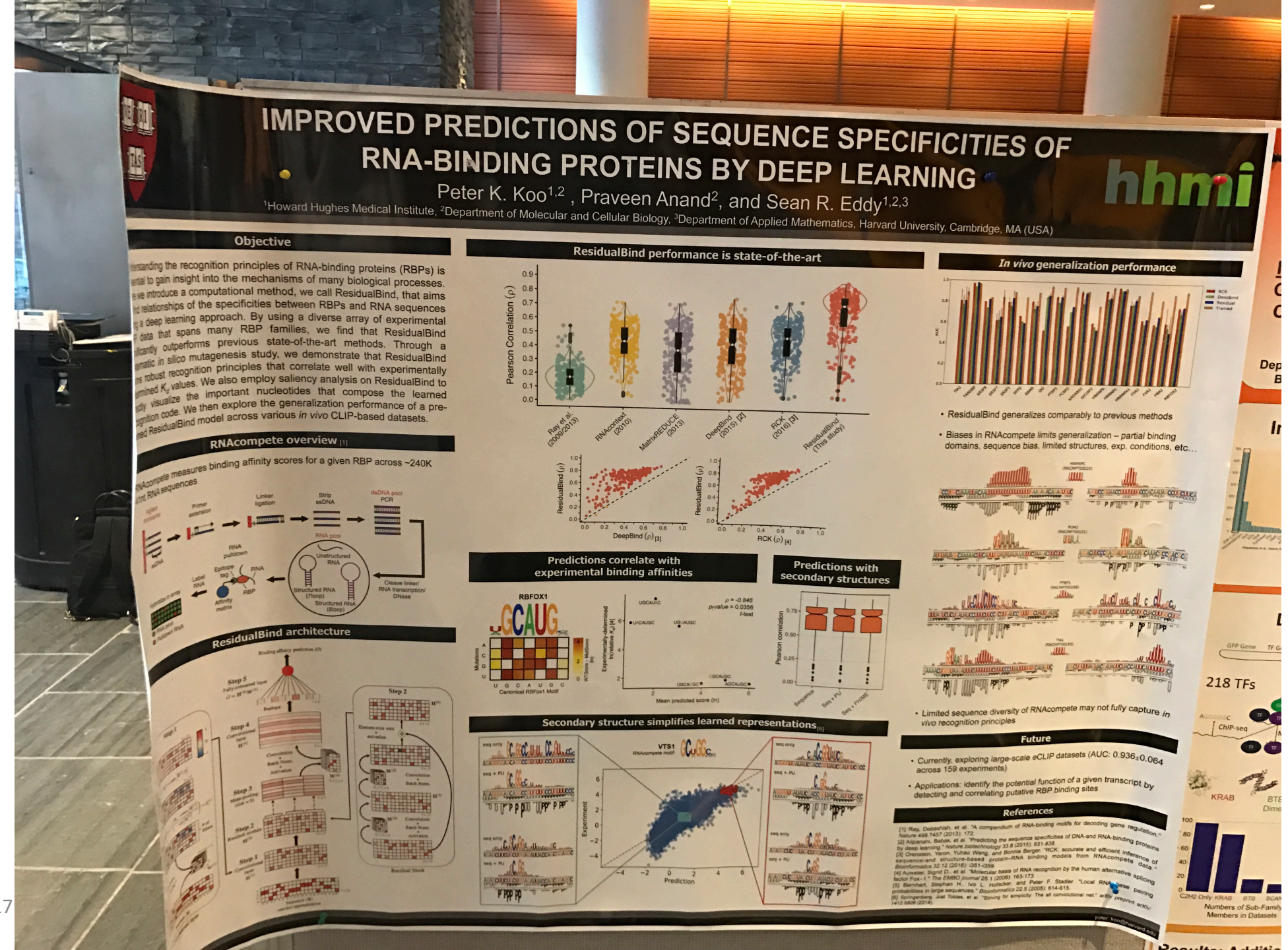
Hi-C is useless in predicting enhancer gene linkage?

From Genetics To Therapeutics: Uncovering And Manipulating The Circuitry Of Non-coding Disease Variants






CONVERGE: Refine TFs→Enhancer→SNP→Target net
Update trait-relevant regulators
SNP
Enh
+Prom
TFs
M-step
E-step
Update 'causal' regulatory regions
Input:
• GWAS P-values
• Regulatory network
• Region⇔LD SNPs net
EM
Output:
• Trait-relevant TFs
• Causal SNPs
• Target genes
Gerald Quon


IMPROVED PREDICTIONS OF SEQUENCE SPECIFICITIES OF RNA-BINDING PROTEINS BY DEEP LEARNING
Peter K. Koo[1,2] , Praveen Anand[2], and Sean R. Eddy[1,2,3]
[1]Howard Hughes Medical Institute, [2]Department of Molecular and Cellular Biology, [3]Department of Applied Mathematics, Harvard University, Cambridge, MA (USA)
hhmi
Objective
RNAcompete overview [1]
ResidualBind architecture
ResidualBind performance is state-of-the-art
Predictions correlate with experimental binding affinities
Predictions with secondary structures
Secondary structure simplifies learned representations [6]
In vivo generalization performance
Future
References