HUGIn: a handy online visualization tool for across-tissue chromatin spatial organization data

Tailored specifically for **GWAS fine mapping**:

- **target gene(s) identification**
- **functional prioritization**

Huginn and Muninn

From Wikipedia, the free encyclopedia

Huginn and Muninn sit on Odin's shoulders in an illustration from an 18th-century Icelandic manuscript

In Norse mythology, **Huginn** (from Old Norse "thought"[1]) and **Muninn** (Old Norse "memory"[2] or "mind"[3]) are a pair of ravens that fly all over the world, Midgard, and bring information to the god Odin. Huginn and Muninn are attested in the *Poetic Edda*, compiled in the 13th century from earlier traditional sources: the *Prose Edda* and *Heimskringla*, written in the 13th century by Snorri Sturluson; in the *Third Grammatical Treatise*, compiled in the 13th century by Óláfr Þórðarson; and in the poetry of skalds. The names of the ravens are sometimes modernly anglicized as **Hugin** and **Munin**.

HUGIN AND MUNIN

Hugin and Munin (pronounced "HOO-gin" and "MOO-nin"; Old Norse *Huginn* and *Muninn*, the meaning of which will be discussed below) are two ravens in Norse mythology who are helping spirits of the god Odin. According to the medieval Icelandic historian Snorri Sturluson,

*Two ravens sit on his (Odin's) shoulders and whisper all the news which they see and hear into his ear; they are called Huginn and Muninn. He sends them out in the morning to fly around the whole world, and by breakfast they are back again. Thus, he finds out many new things and this is why he is called 'raven-god' (*hrafnaguð*).*[1]

A warrior, likely Odin, flanked by two ravens on an Iron Age helmet from what is now Sweden

HUGIn

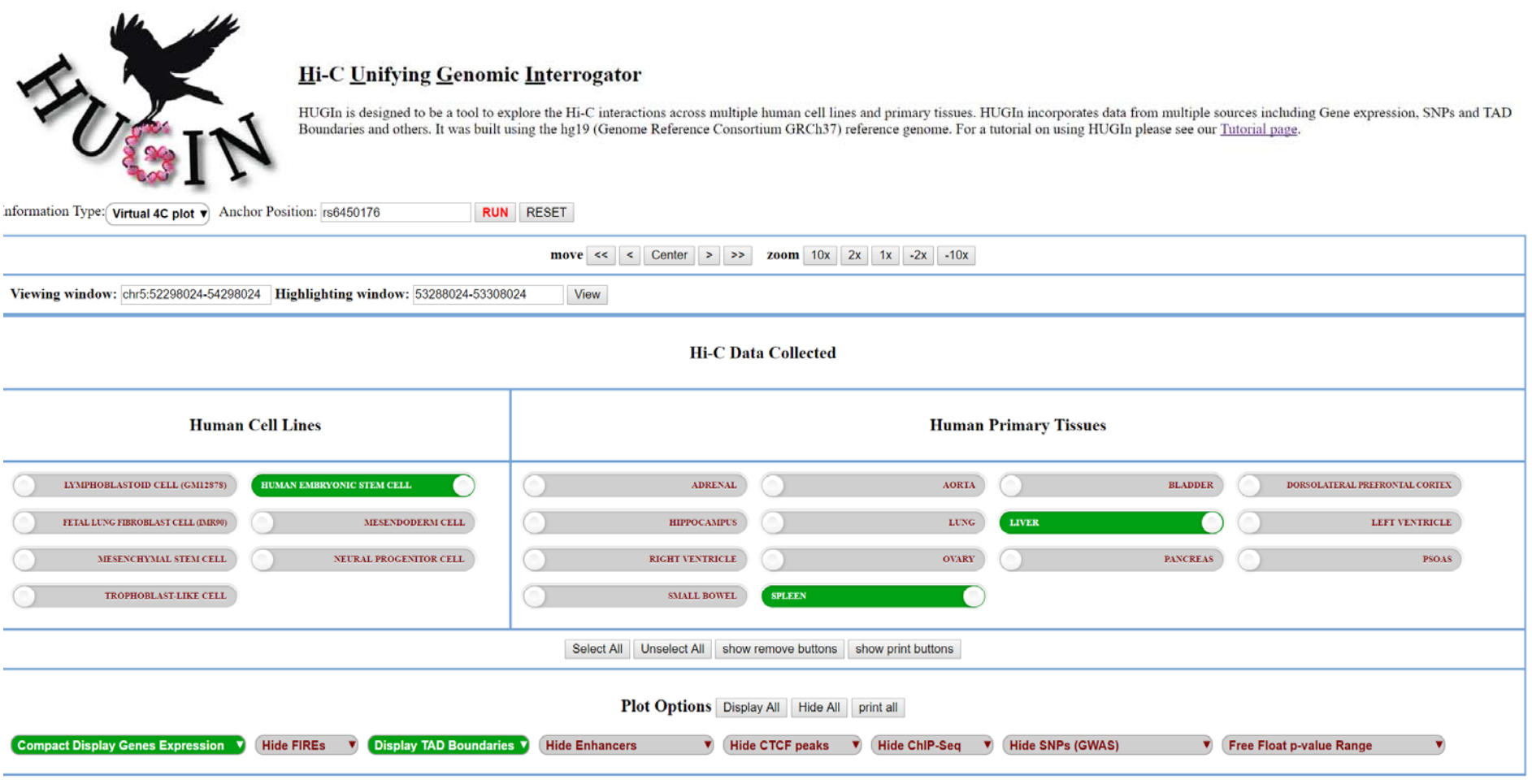
Bioinformatics, 2017, 1–3
doi: 10.1093/bioinformatics/btx359
Advance Access Publication Date: 5 June 2017
Applications Note

Genome analysis

HUGIn: Hi-C Unifying Genomic Interrogator

Joshua S. Martin[1], Zheng Xu[2,3], Alex P. Reiner[4,5], Karen L. Mohlke[1], Patrick Sullivan[1,6], Bing Ren[7], Ming Hu[8,*] and Yun Li[1,9,10,*]

HUGIn: Hi-C Unifying Genomic Interrogator

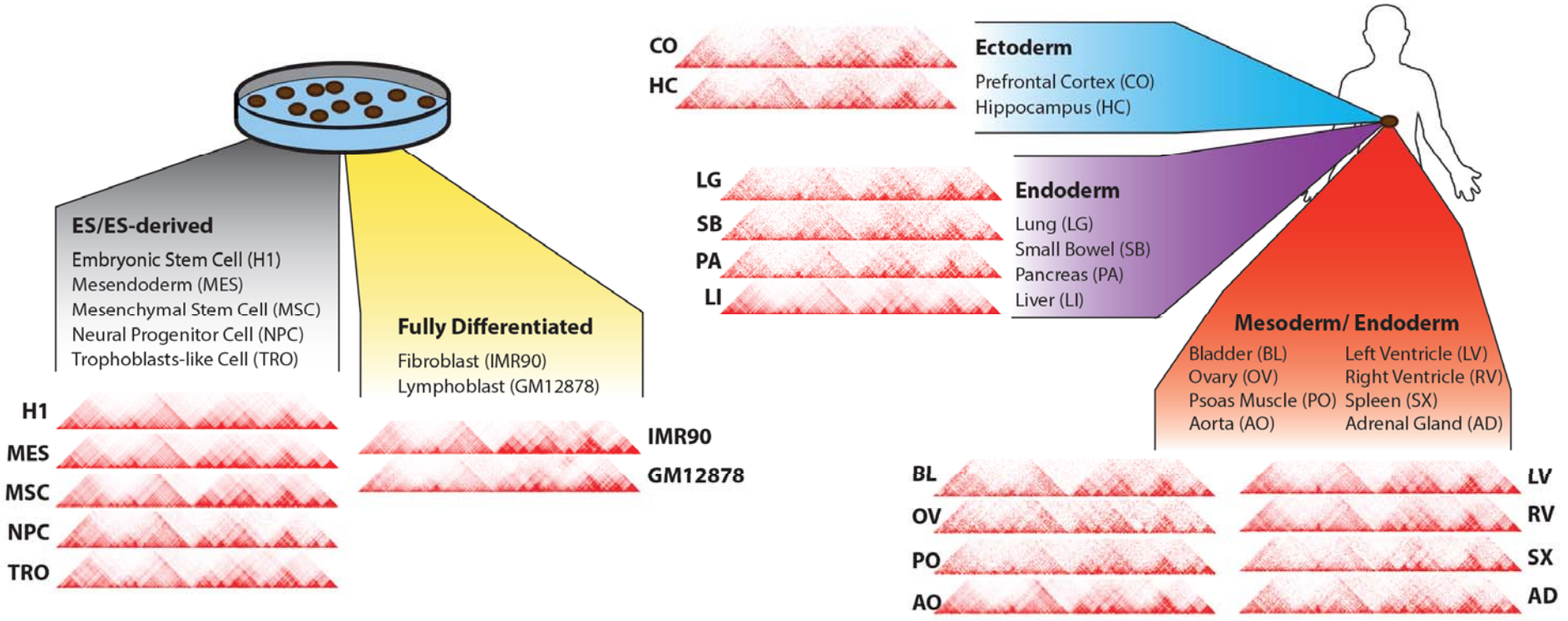


http://yunliweb.its.unc.edu/HUGIn/

A Compendium of Hi-C Data

A Compendium of Chromatin Contact Maps Reveals Spatially Active Regions in the Human Genome

Anthony D. Schmitt,[1,2,12,13] Ming Hu,[3,12,14,*] Inkyung Jung,[1,15] Zheng Xu,[4,10,11] Yunjiang Qiu,[1,5] Catherine L. Tan,[1,13] Yun Li,[4] Shin Lin,[6] Yiing Lin,[7] Cathy L. Barr,[8] and Bing Ren[1,9,16,*]



7 human cell lines

14 human primary tissues

Highlight by *Nature Reviews Genetics*

RESEARCH HIGHLIGHTS

Nature Reviews Genetics | Published online 13 Dec 2016; doi:10.1038/nrg.2016.161

The majority of disease-associated genetic variants identified by genome-wide asso- a new type of unexplored regulatory element that may need to be considered when genetic

In their study, Schmitt *et al.* carried out Hi-C across 21 diverse primary human tissues and cell types. Analysing an average of 214 million unique chromosome contacts per tissue type, they noticed that some regions displayed particularly high local contact frequencies, which they termed frequently interacting regions (FIREs). FIREs were distinct from previously defined types of chromosome domains such as A/B compartments, topologically associated domains (TADs) and loops, although in general they occurred towards the centre of TADs, partook in numerous intra-TAD interactions and were contained within broader regions of A-compartment active chromatin. Further analyses, including integration with profiles of histone modifications and transcription, revealed that FIREs are highly tissue-type-dependent, frequently occur near (and transcriptionally regulate) cell-identity genes and overlap substantially with chromatin features of active enhancers. Indeed, the

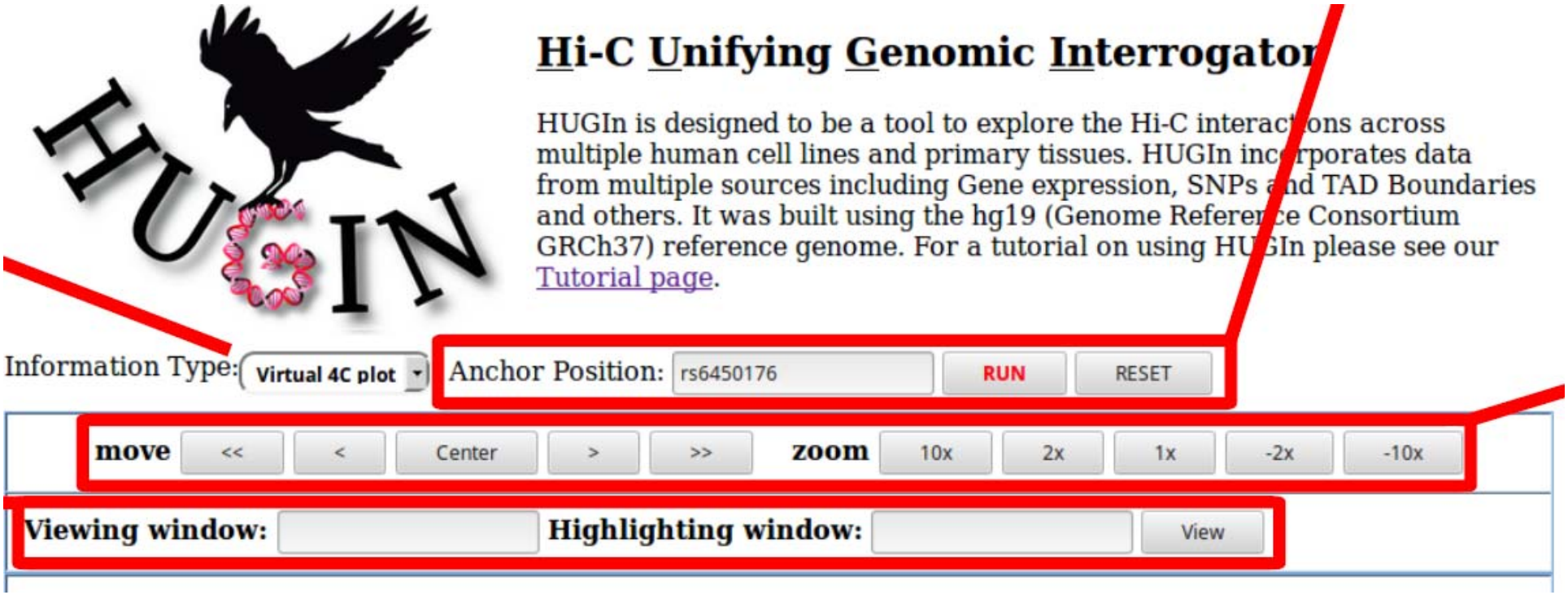
journals, and are available online.

Darren J. Burgess

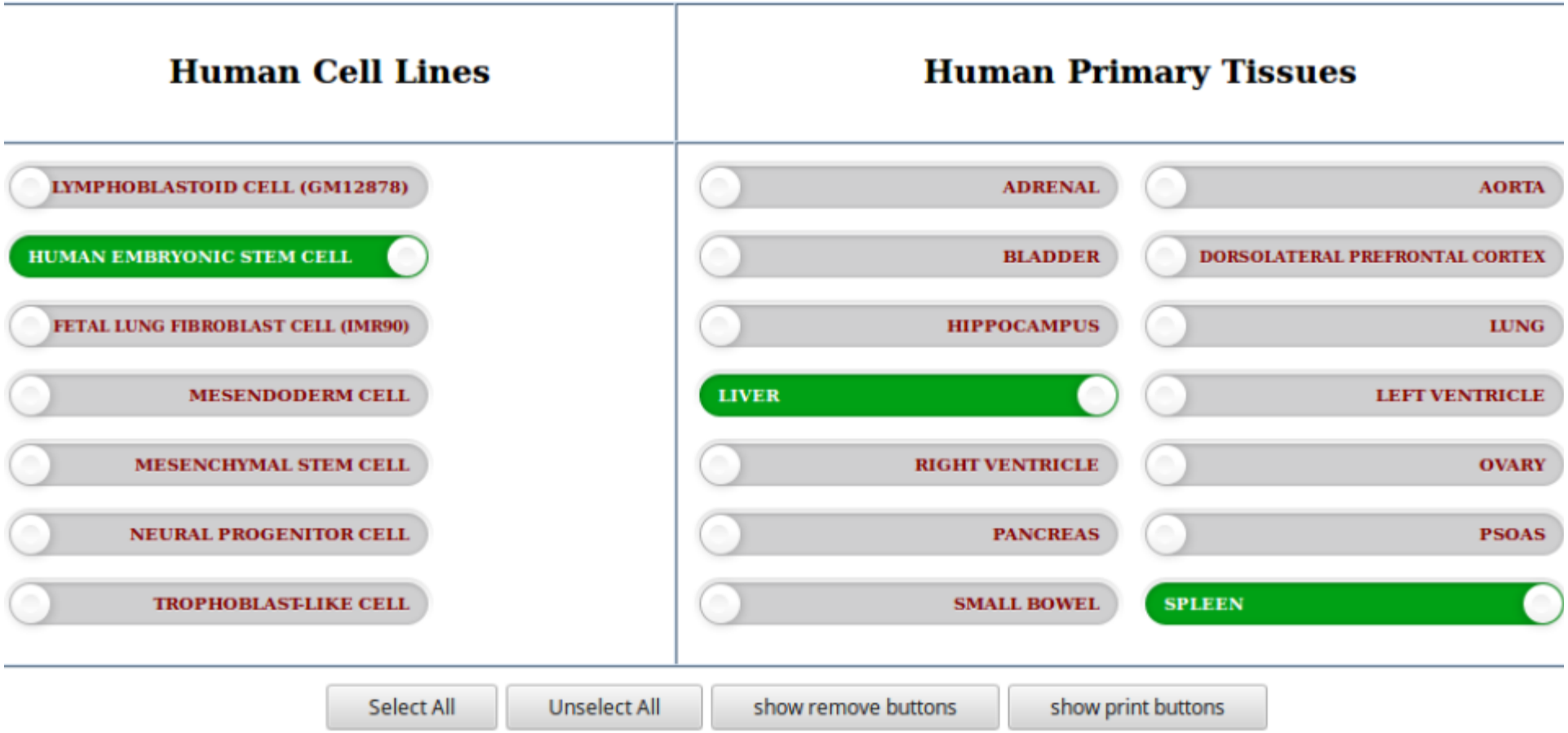
ORIGINAL ARTICLES Javierre, B. M. *et al.* Lineage-specific genome architecture links enhancers and non-coding disease variants to target gene promoters. *Cell* **167**, 1369–1384 (2016) | Schmitt, A. D. *et al.* A compendium of chromatin contact maps reveals spatially active regions in the human genome. *Cell Rep.* **17**, 2042–2059 (2016)

FURTHER READING Bonev. B. & Cavalli, G. Organization and function of the 3D genome. *Nat. Rev. Genet.* **17**, 661–678 (2016) | Stricker, S. H., Köferle, A. & Beck, S. From profiles to function in epigenomics. *Nat. Rev. Genet.* http://dx.doi.org/10.1038/nrg.2016.138 (2016)

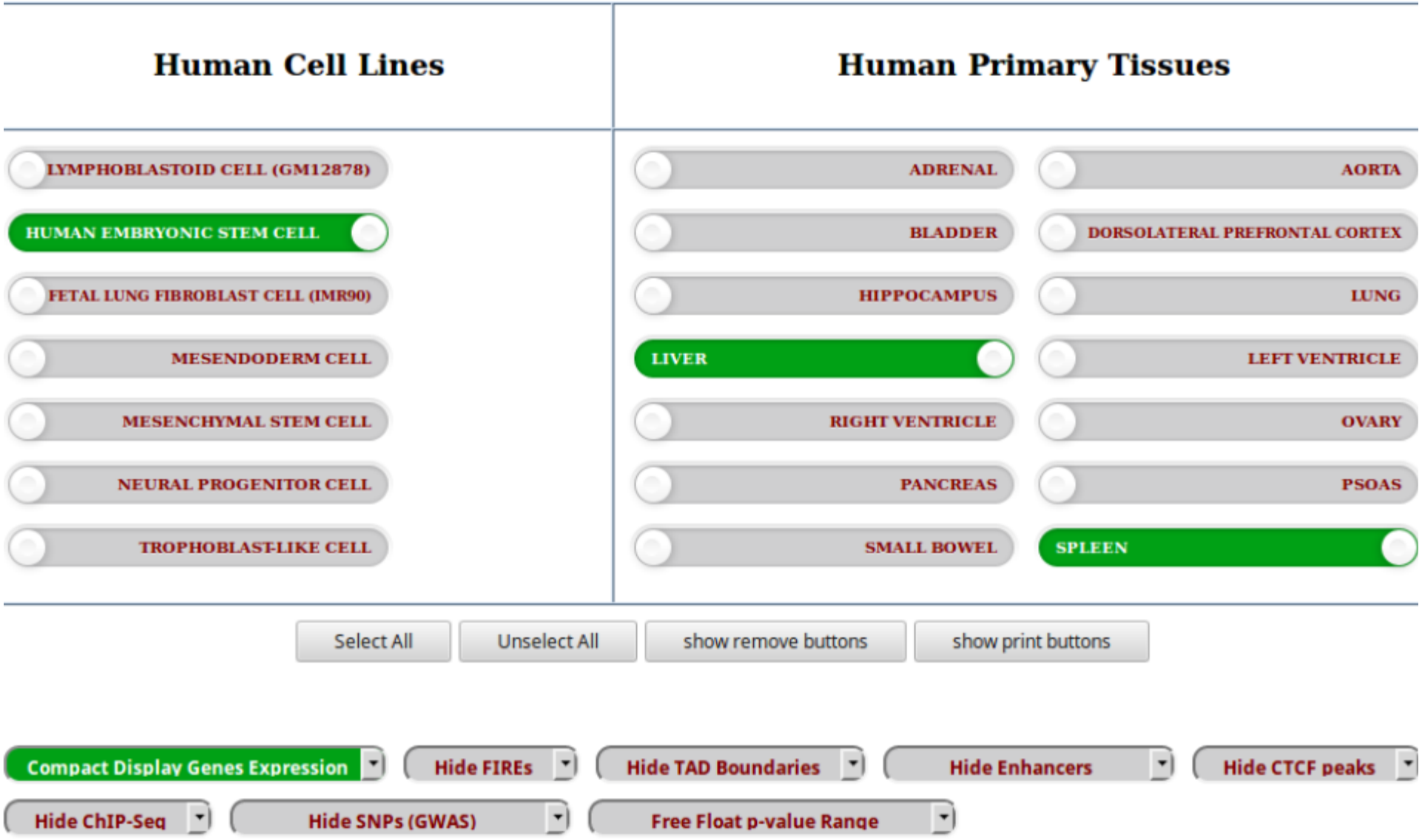
HUGIn Interface



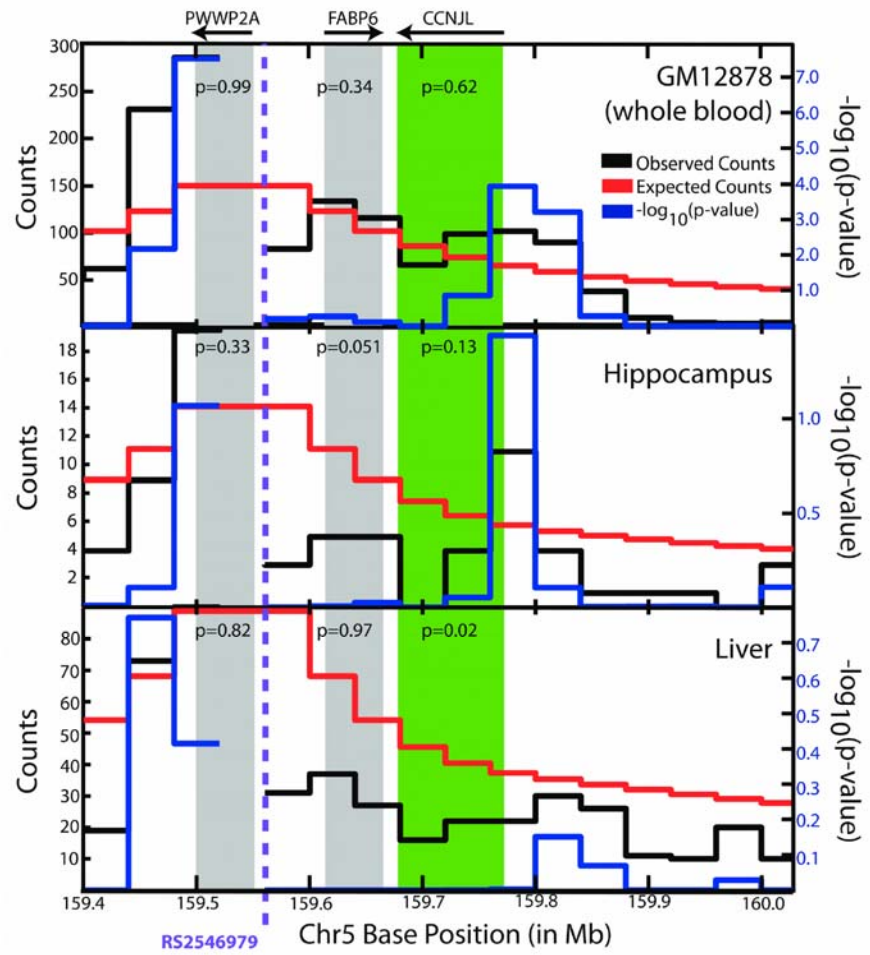
Data Availability Snapshot



Data Availability Snapshot

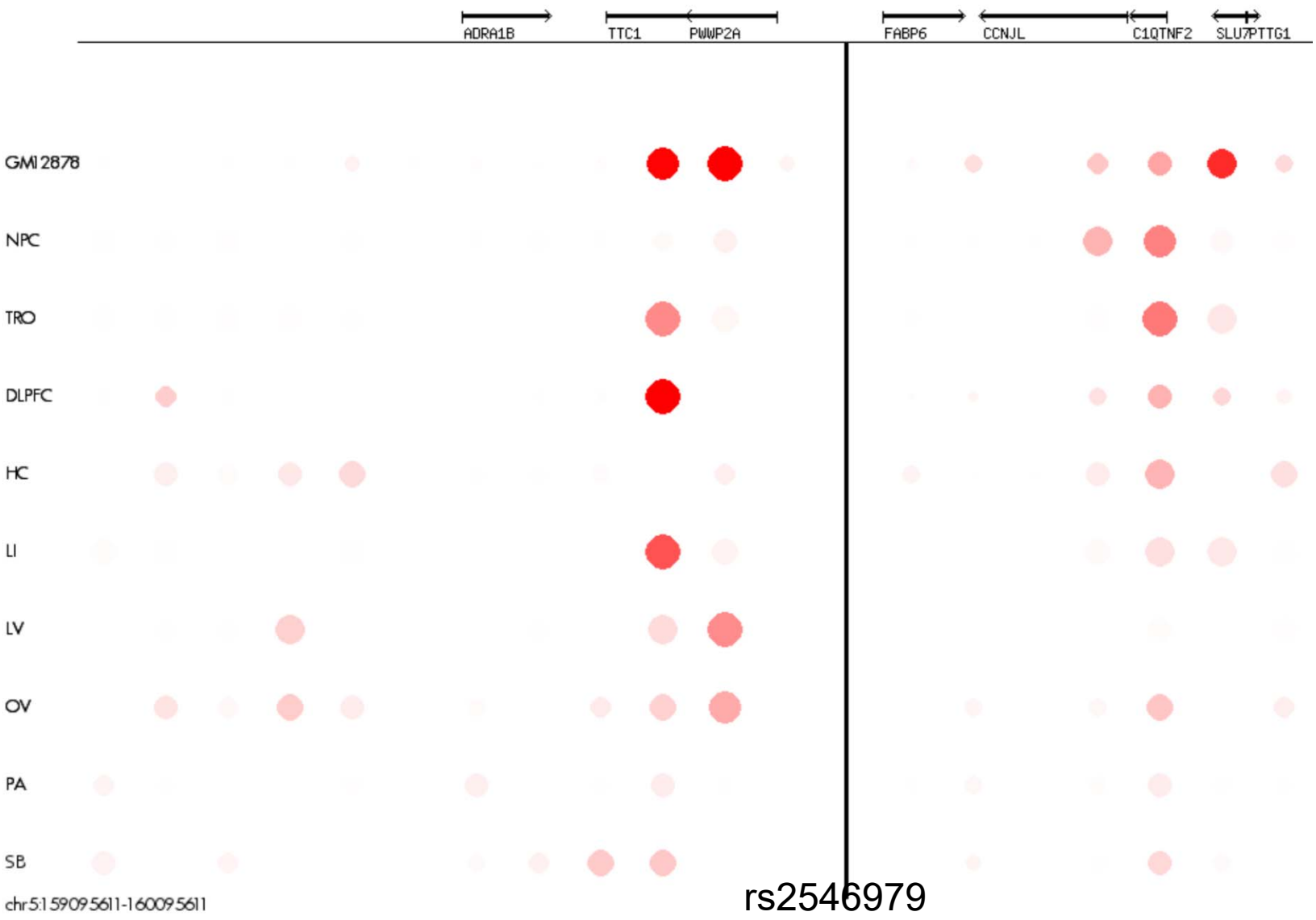


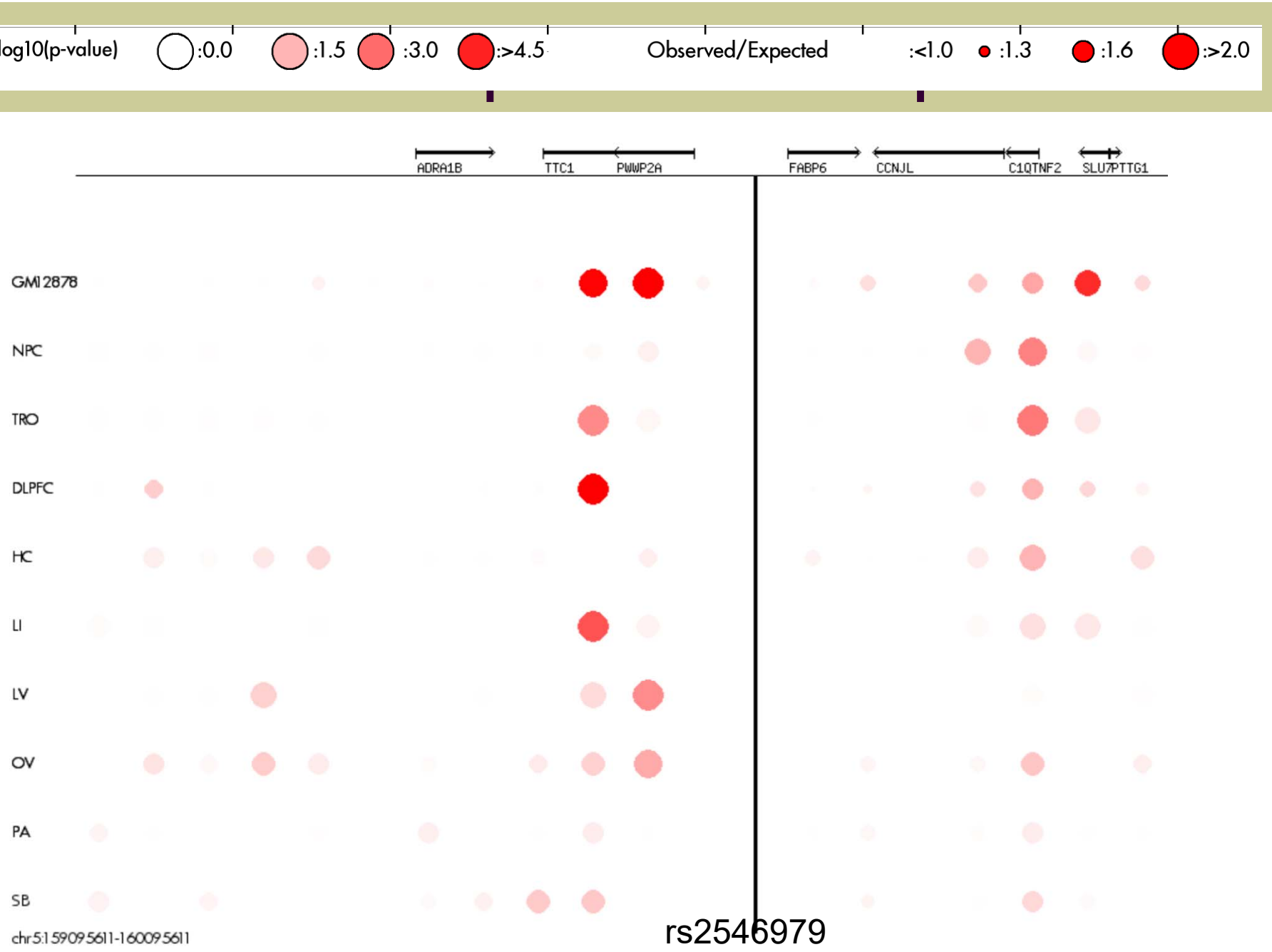
One Example: Virtual 4C Plot



- Iotchkova et al (*Nature Genetics* 2016) reported interaction between platelet count associated SNP rs2546979 and gene *CCNJL*.
- Evidence of physical interaction from BLUEPRINT Epigenomic project.
- **We found**: SNP rs2546979
  - shows long range interaction with gene *CCNJL* in GM12878 cell line and much weaker in hippocampus, but not in liver tissue.
  - near a GM12878 typical enhancer.
  - in a Hippocampus typical enhancer
- Using GTEx
  - In whole blood, rs2546979 is NOT eQTL of *CCNJL*, *FABP6* or *PWWP2A*.
  - only eQTL for *FABP6* in Skin-Sun Exposed (lower leg) (*P*=7.3e-7), arguably the wrong gene or wrong tissue, or both.

Same Example: Heatmap View



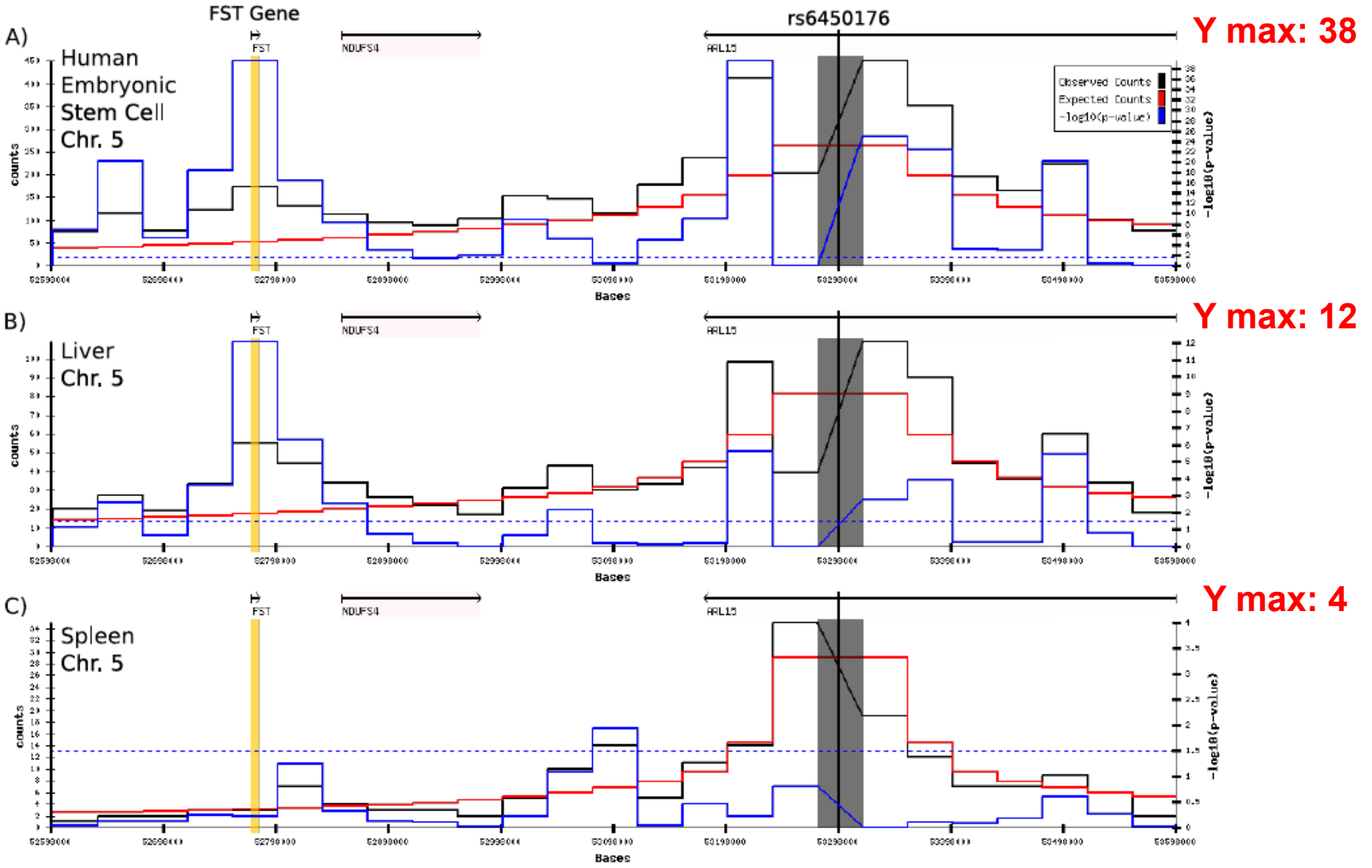

-log10(p-value)
:0.0
:1.5
:3.0
:>4.5
Observed/Expected
:<1.0
:1.3
:1.6
:>2.0
ADRA1B
TTC1
PWWP2A
FABP6
CCNJL
C1QTNF2
SLU7PTTG1
GM12878
NPC
TRO
DLPFC
HC
LI
LV
OV
PA
SB
chr5:159095611-160095611
rs2546979

Another Example: rs6450176

- rs6450176
  - Associated with risk of T2D
  - Associated with adiponectin
  - Located within *ARL15* gene

Potential Target Gene: ??

Rs6450176 Virtual 4C Plot



Heatmap for rs6450176

Another Example: rs6450176

- rs6450176
  - Associated with risk of T2D
  - Associated with adiponectin
  - Located within *ARL15* gene

  - **Now with multi-tissue/cell-line Hi-C data:**

Potential Target Gene: ??

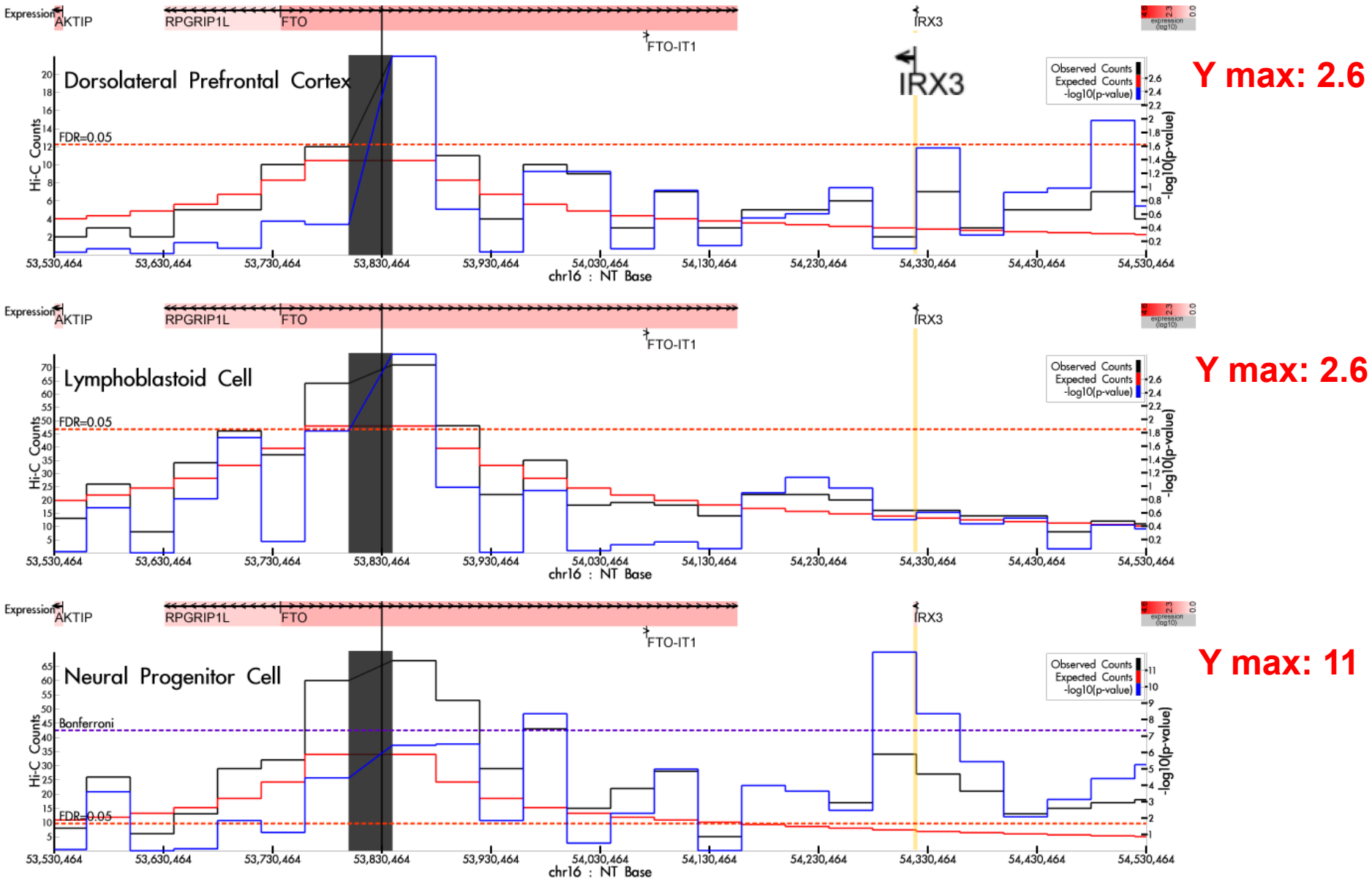
Another Example: rs6450176

- rs6450176
  - Associated with risk of T2D
  - Associated with adiponectin
  - Located within *ARL15* gene

  - **Now with multi-tissue/cell-line Hi-C data**
  - eQTL for *FST* but not *ARL15* in subcutaneous adipose tissue

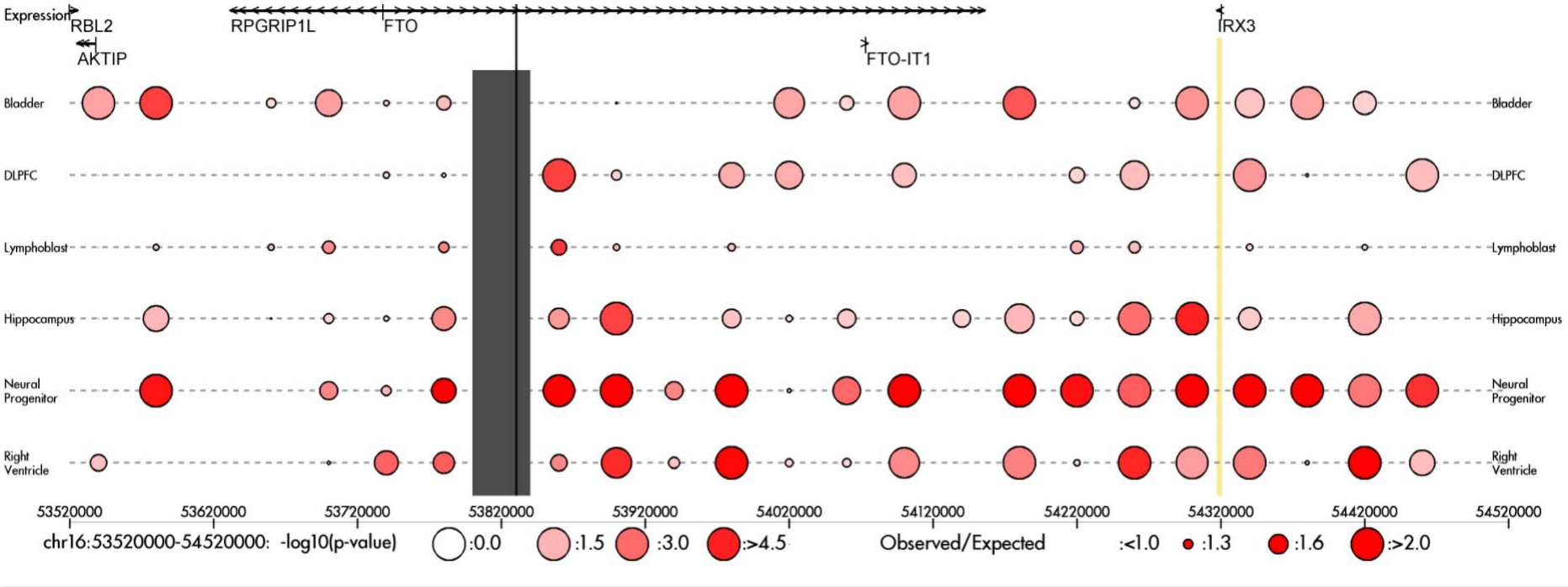
Potential Target Gene: ??

A 3rd example: FTO rs9930506

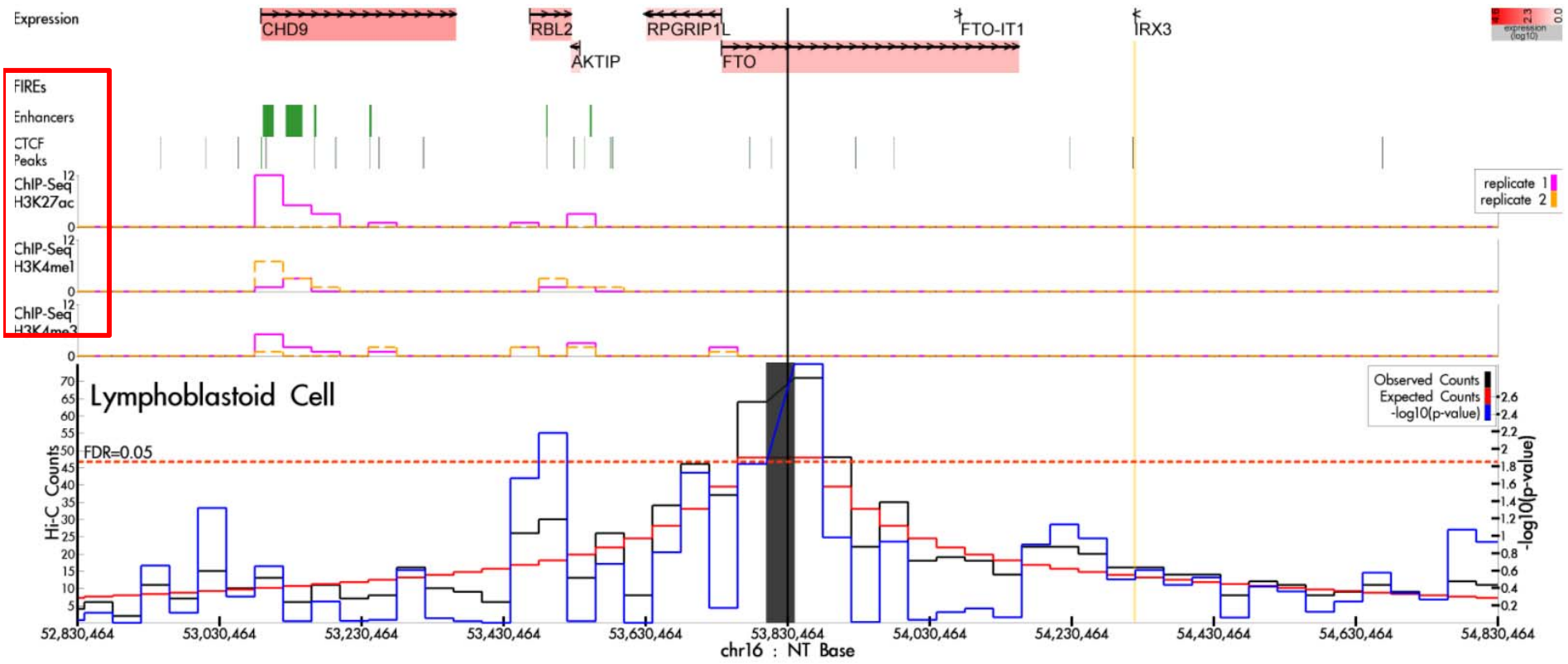


Smemo *et al* 2014 *Nature* 507: 371-5.

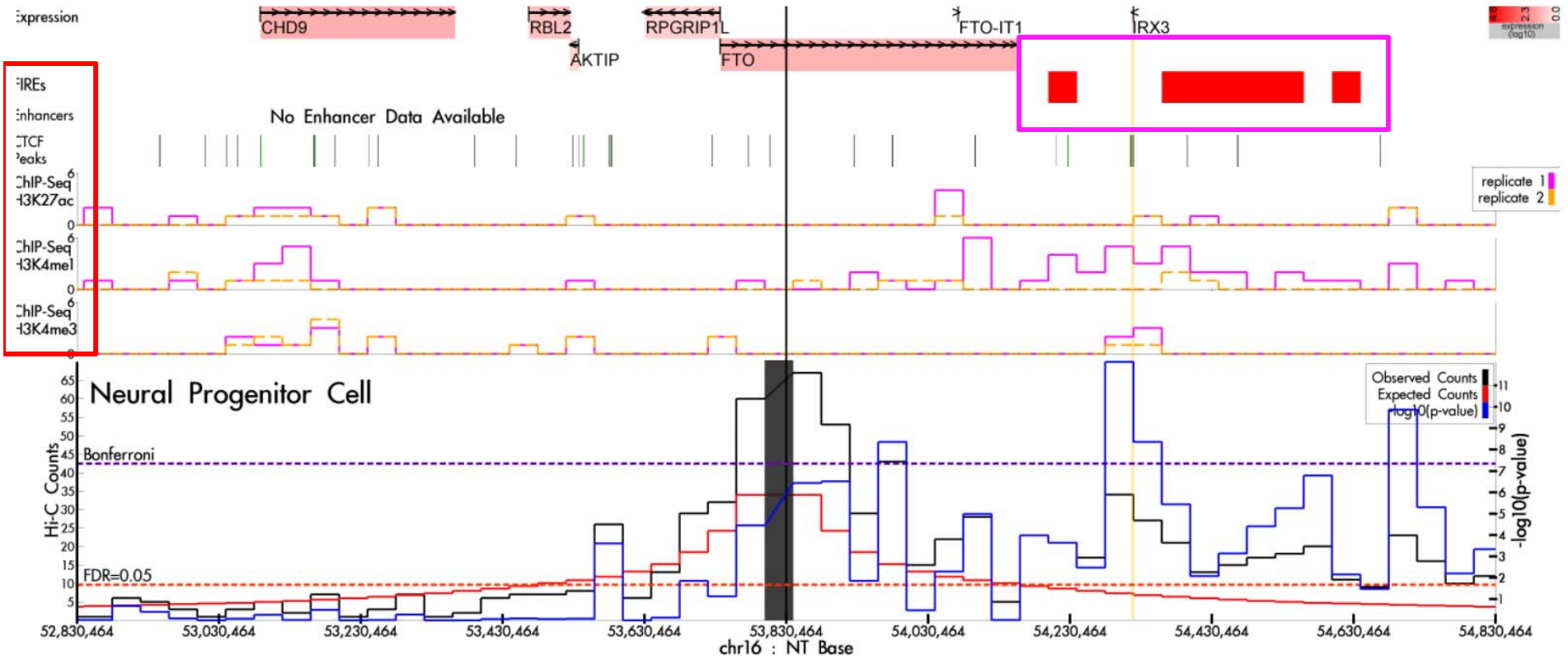
FTO rs9930506 heatmap view



FTO rs9930506 w/ additional info. in LCL



FTO rs9930506 w/ additional info. in NPC



Frequently **I**nteracting **Re**gions (**FIRE**s) (Schmitt et al 2016 *Cell Rep*):

- are tissue/cell-type specific (in contrast to TADs);
- are enriched of tissue/cell-type specific enhancers;
- are enriched of tissue/cell-type specifically expressed genes;
- are conserved in human and mouse
- are enriched of GWAS SNPs for relevant traits

Better Gene Set Enrichment

Disease	# of GWAS SNPs	# of HUGIn genes	# (%) of HUGIn genes being the closest to GWAS SNPs	# of GWAS catalog genes	# (%) of GWAS catalog genes being the closest to GWAS SNPs	# (%) of HUGIn genes overlapping with GWAS catalog genes
Schizophrenia	407	471	35 (7.4%)	676	120 (17.8%)	69 (14.6%)
Leukemia	56	56	2 (3.6%)	82	19 (23.2%)	7 (12.5%)
Alzheimer's disease	203	228	14 (6.1%)	343	56 (16.3%)	31 (13.6%)
Autism	37	36	2 (5.6%)	61	6 (9.8%)	4 (11.1%)
Depression	133	172	7 (4.1%)	226	31 (13.7%)	18 (10.5%)
Type 1 diabetes	38	57	2 (3.5%)	68	19 (27.9%)	6 (10.5%)
Type 2 diabetes	124	120	12 (10.0%)	180	46 (25.6%)	17 (14.2%)

Better Gene Set Enrichment

Disease	# of GWAS SNPs	# of HUGIn genes	# (%) of HUGIn genes being the closest to GWAS SNPs	# of GWAS catalog genes	# (%) of GWAS catalog genes being the closest to GWAS SNPs	# (%) of HUGIn genes overlapping with GWAS catalog genes
Schizophrenia	407	471	35 (7.4%)	676	120 (17.8%)	69 (14.6%)
Leukemia	56	56	2 (3.6%)	82	19 (23.2%)	7 (12.5%)
Alzheimer's disease	203	228	14 (6.1%)	343	56 (16.3%)	31 (13.6%)
Autism	37	36	2 (5.6%)	61	6 (9.8%)	4 (11.1%)
Depression	133	172	7 (4.1%)	226	31 (13.7%)	18 (10.5%)
Type 1 diabetes	38	57	2 (3.5%)	68	19 (27.9%)	6 (10.5%)
Type 2 diabetes	124	120	12 (10.0%)	180	46 (25.6%)	17 (14.2%)

Better Gene Set Enrichment

Disease	# of GWAS SNPs	# of HUGIn genes	# (%) of HUGIn genes being the closest to GWAS SNPs	# of GWAS catalog genes	# (%) of GWAS catalog genes being the closest to GWAS SNPs	# (%) of HUGIn genes overlapping with GWAS catalog genes
Schizophrenia	407	471	35 (7.4%)	676	120 (17.8%)	69 (14.6%)
Leukemia	56	56	2 (3.6%)	82	19 (23.2%)	7 (12.5%)
Alzheimer's disease	203	228	14 (6.1%)	343	56 (16.3%)	31 (13.6%)
Autism	37	36	2 (5.6%)	61	6 (9.8%)	4 (11.1%)
Depression	133	172	7 (4.1%)	226	31 (13.7%)	18 (10.5%)
Type 1 diabetes	38	57	2 (3.5%)	68	19 (27.9%)	6 (10.5%)
Type 2 diabetes	124	120	12 (10.0%)	180	46 (25.6%)	17 (14.2%)

HUGIn output/information types

1. Virtual 4C plot
2. Heatmap
3. Text information
4. Association

Text information

Excerpt of Text Information

Fragment Start	Lymphoblastoid Cell Observed Count	Lymphoblastoid Cell Expected Count	Lymphoblastoid Cell - log10(pValue)	Neural Progenitor Cell Observed Count	Neural Progenitor Cell Expected Count	Neural Progenitor Cell -log10(pValue)
52840000	6	7.59521	0.11418	1	3.85092	0.00933
52880000	2	7.95852	0.00136	6	4.00916	0.66494
52920000	11	8.34400	0.65825	5	4.17524	0.39216
52960000	7	8.75363	0.11360	3	4.35348	0.09191
53000000	15	9.18937	1.31943	1	4.54898	0.00462
53040000	10	9.65219	0.30277	3	4.76687	0.06829
54680000	9	8.34400	0.34165	23	4.17524	9.86795
54720000	8	7.95852	0.26661	16	4.00916	5.29830
54760000	12	7.59521	1.07091	10	3.85092	2.19817
54800000	11	7.25233	0.93035	12	3.69954	3.32820

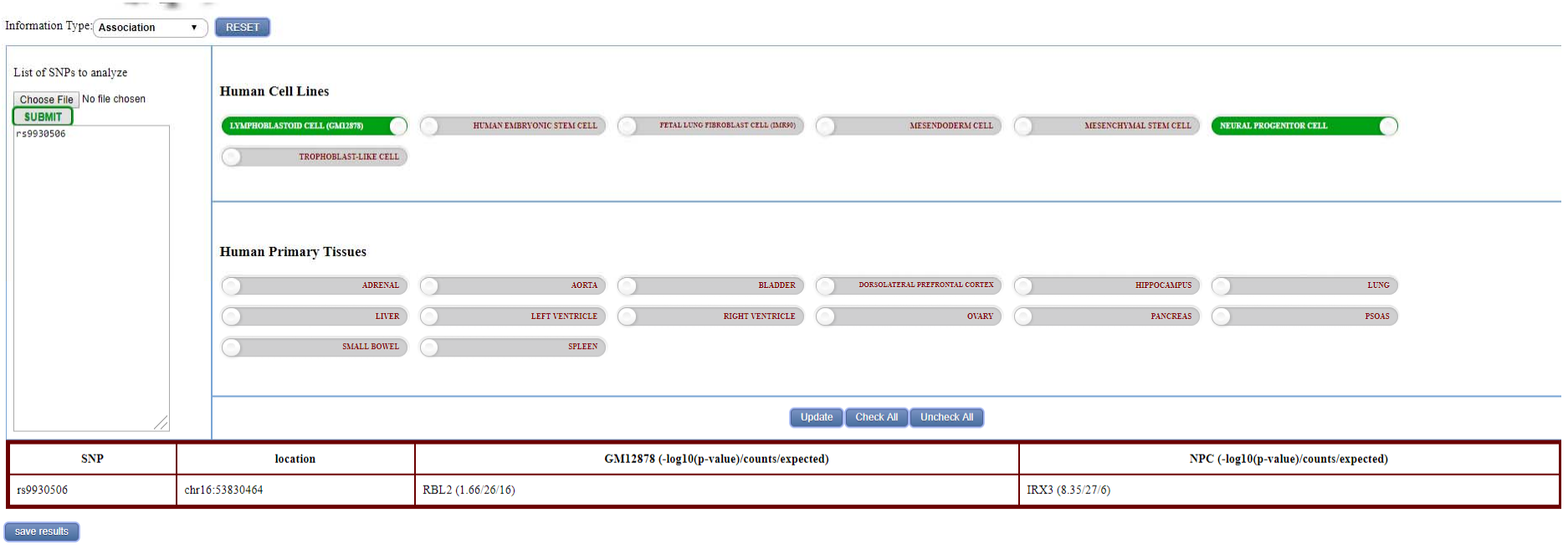
Save to File

Excerpt of Text Information

Fragment Start	Lymphoblastoid Cell Observed Count	Lymphoblastoid Cell Expected Count	Lymphoblastoid Cell - log10(pValue)	Neural Progenitor Cell Observed Count	Neural Progenitor Cell Expected Count	Neural Progenitor Cell -log10(pValue)
52840000	6	7.59521	0.11418	1	3.85092	0.00933
52880000	2	7.95852	0.00136	6	4.00916	0.66494
52920000	11	8.34400	0.65825	5	4.17524	0.39216
52960000	7	8.75363	0.11360	3	4.35348	0.09191
53000000	15	9.18937	1.31943	1	4.54898	0.00462
53040000	10	9.65219	0.30277	3	4.76687	0.06829
54680000	9	8.34400	0.34165	23	4.17524	9.86795
54720000	8	7.95852	0.26661	16	4.00916	5.29830
54760000	12	7.59521	1.07091	10	3.85092	2.19817
54800000	11	7.25233	0.93035	12	3.69954	3.32820

Save to File

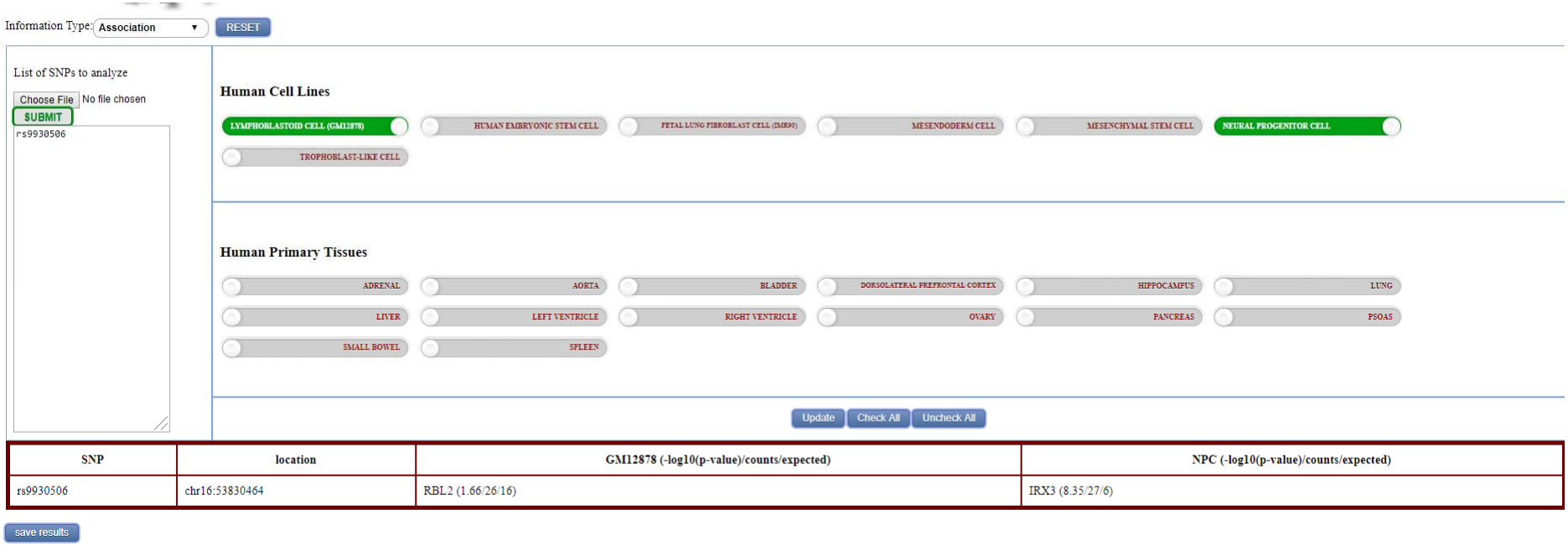
Association: rs9930506 example



SNP	location	GM12878 (-log10(p-value)/counts/expected)	NPC (-log10(p-value)/counts/expected)
rs9930506	chr16:53830464	RBL2 (1.66/26/16)	IRX3 (8.35/27/6)

save results

Association: rs9930506 example

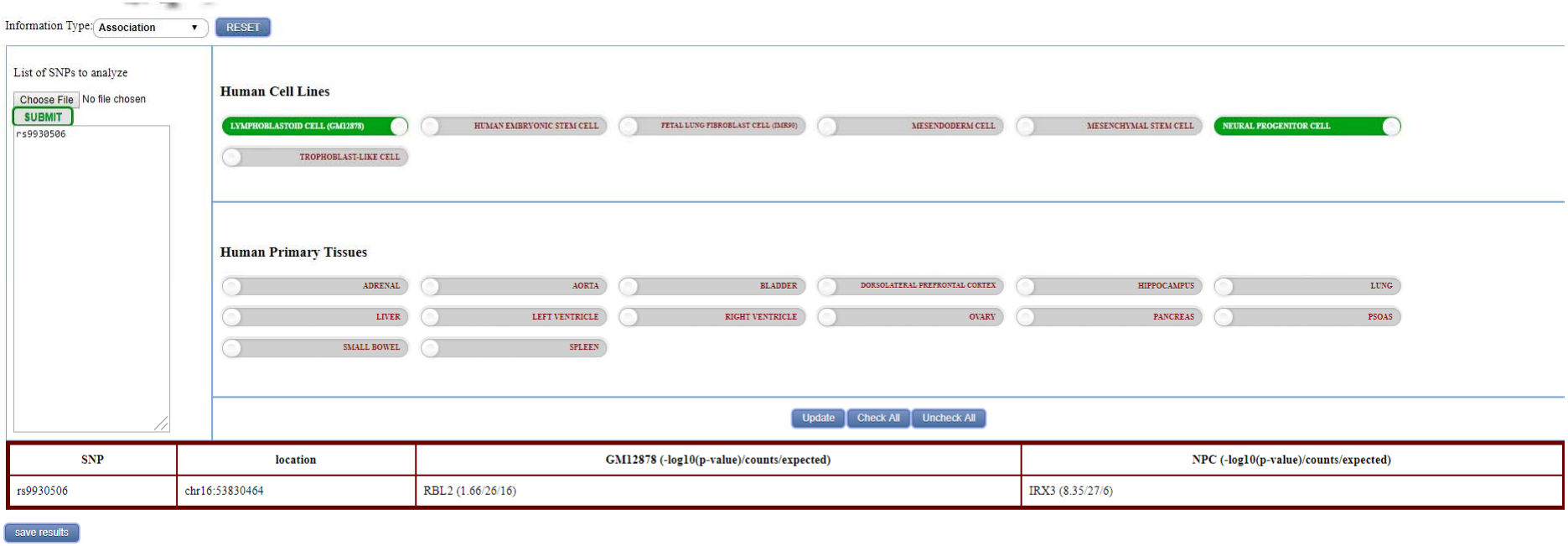


SNP	location	GM12878 (-log10(p-value)/counts/expected)	NPC (-log10(p-value)/counts/expected)
rs9930506	chr16:53830464	RBL2 (1.66/26/16)	IRX3 (8.35/27/6)

save results

SNP	location	GM12878 (-log10(p-value)/counts/expected)
rs9930506	chr16:53830464	RBL2 (1.66/26/16)

Association: rs9930506 example



SNP	location	GM12878 (-log10(p-value)/counts/expected)	NPC (-log10(p-value)/counts/expected)
rs9930506	chr16:53830464	RBL2 (1.66/26/16)	IRX3 (8.35/27/6)

SNP	location	GM12878 (-log10(p-value)/counts/expected)	NPC (-log10(p-value)/counts/expected)
rs9930506	chr16:53830464	RBL2 (1.66/26/16)	IRX3 (8.35/27/6)

Updates planned:

Hi-C Data currently being analyzed

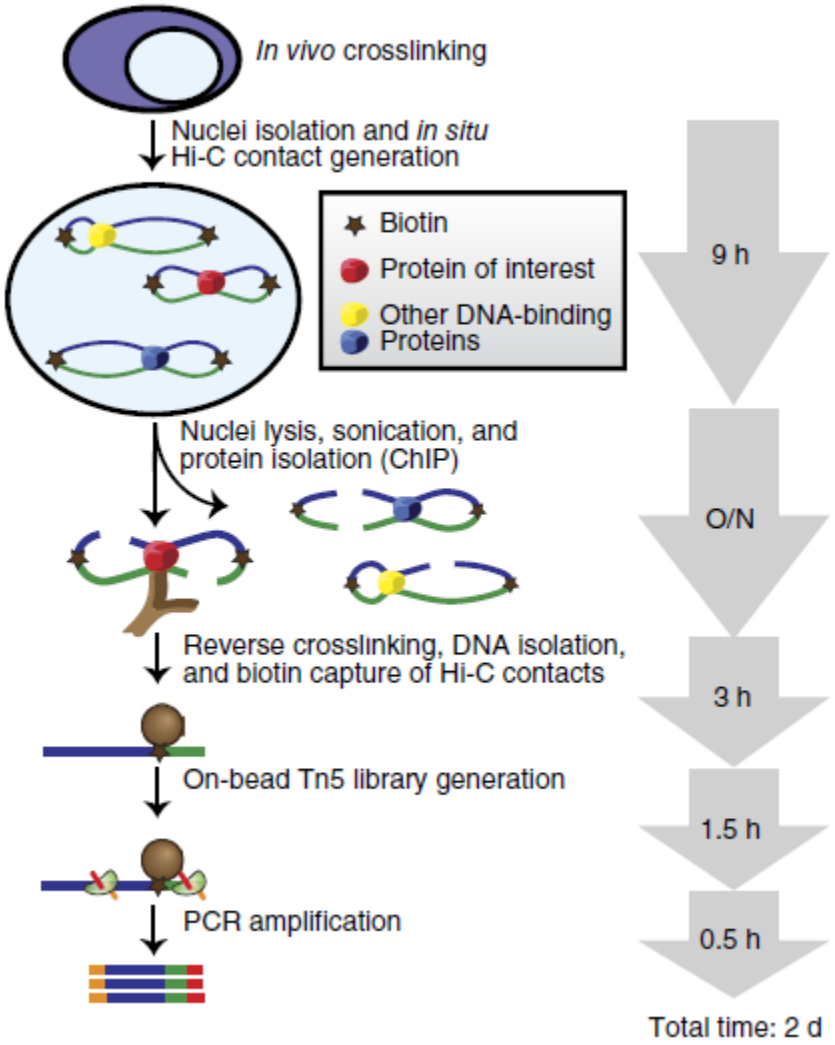
Tissue or Cell Line	SeqDepth (in millions of reads per sample)	Source
3 adult & 3 fetal human brain tissues	338-1,016	Sullivan lab#
GM12878, IMR90, h1 ES and 4 h1-derived CLs	600-1,079	Ren lab[22]
7 human CLs and 14 human primary tissues	56-541	Ren lab[22]
20 human LCLs	454-1,068	Ren lab#
Modified mES CL with 16p11.2 deletion (a human ASD locus)	~500	Pombo lab#

Updates planned:

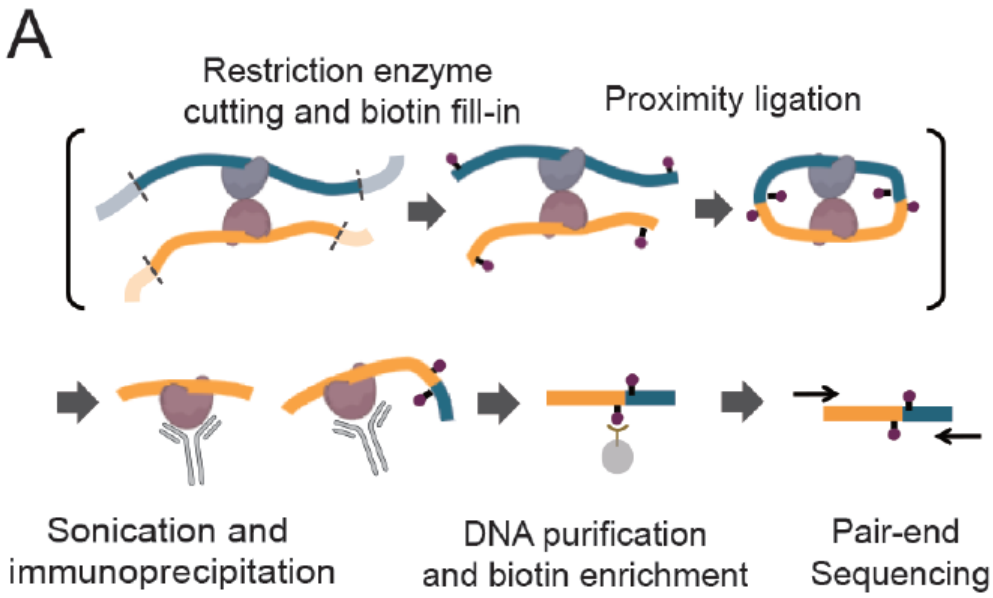
H/P Data currently being analyzed

Tissue or Cell Line[a]	Data Type	SeqDepth	Source
Mouse ES CL (FangData)	PLAC-Seq	~300	Ren lab[3]
Mouse brain tissue from 8 developmental stages	PLAC-Seq	~300	Ren lab#
8 mouse tissues at the same developmental stage	PLAC-Seq	~300	Ren lab#
h1 ES, h1 ES-derived CLs	PLAC-Seq	~300	Ren lab#
brain tissue from 20 mouse RIX[c] lines	PLAC-Seq	~300	Shen lab[f]#
Human NPC[d] and 4 differentiated CLs	PLAC-Seq	~300	Shen lab#
GM12878 Smc1, mES Oct4, mES cohesin	HiChIP	~300	Chang lab[8]

H/P Data: ~Hi-C + IP (Immuno-Precipitation)



Hi-ChIP



PLAC-Seq
proximity ligation-assisted ChIP-seq

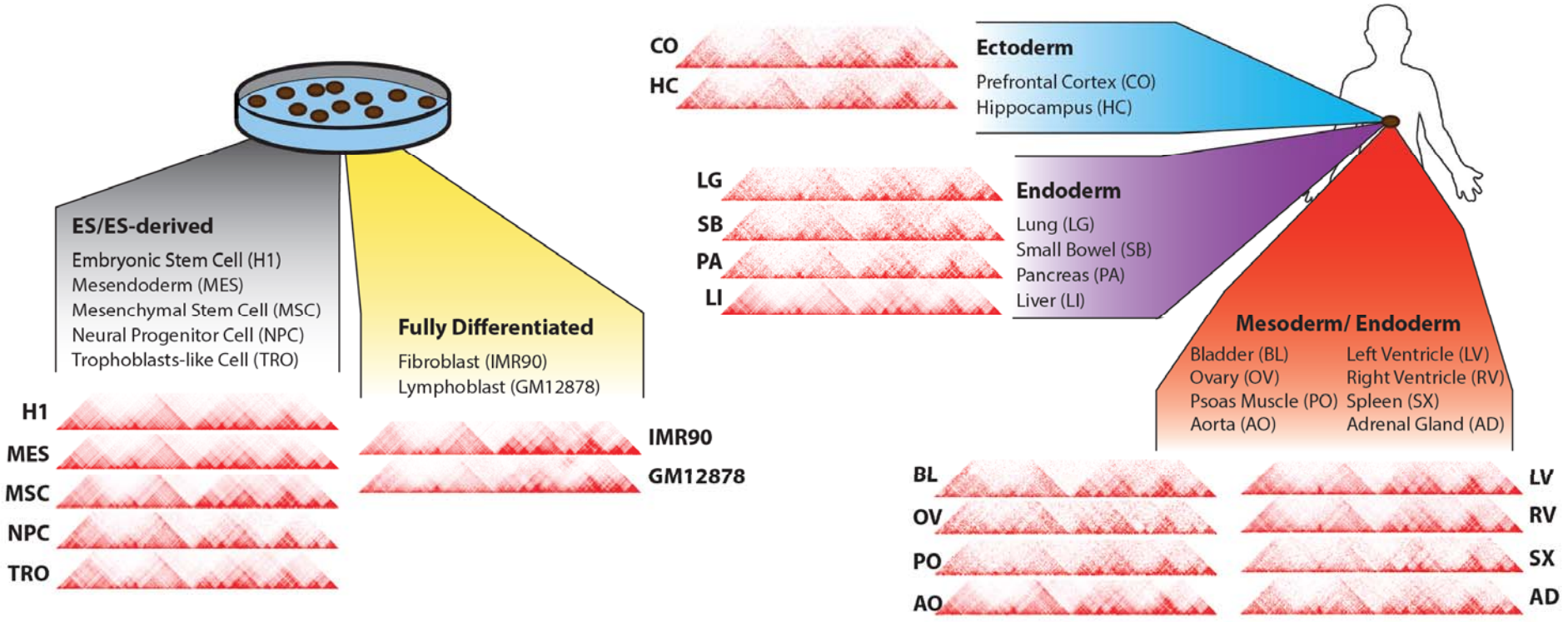
Fang et al 2016 *Cell res.* 26: 1345-8.

Mumbach et al 2016 *Nat Methods* 13: 919-22.

Designs of selected datasets: 1

A Compendium of Chromatin Contact Maps Reveals Spatially Active Regions in the Human Genome

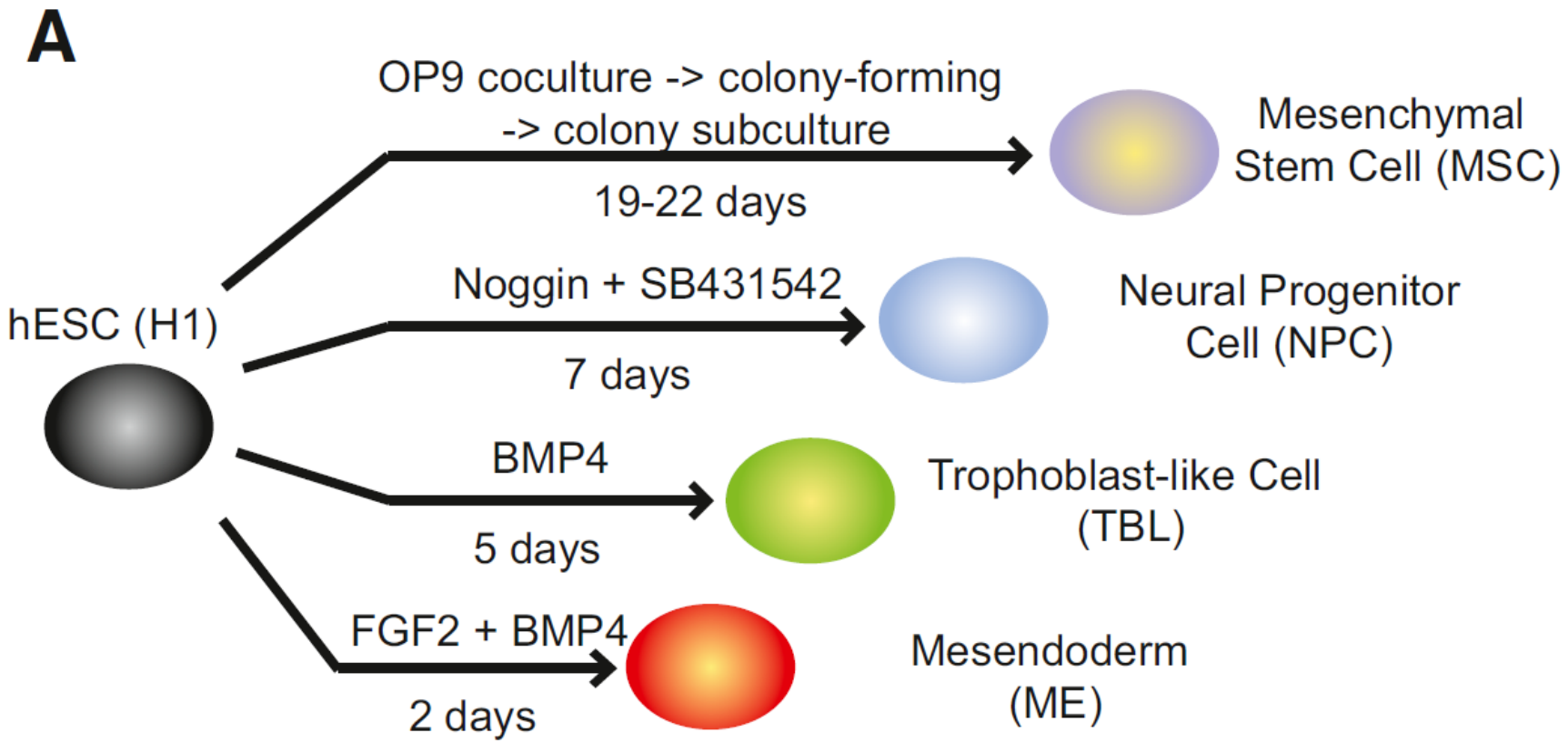
Anthony D. Schmitt,[1,2,12,13] Ming Hu,[3,12,14,*] Inkyung Jung,[1,15] Zheng Xu,[4,10,11] Yunjiang Qiu,[1,5] Catherine L. Tan,[1,13] Yun Li,[4] Shin Lin,[6] Yiing Lin,[7] Cathy L. Barr,[8] and Bing Ren[1,9,16,*]



7 human cell lines

14 human primary tissues

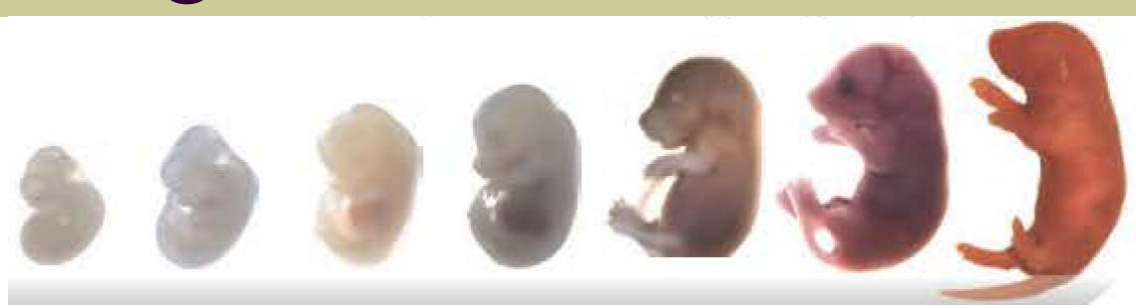
Designs of selected datasets: 2



GM12878, IMR90, h1 ES and 4 h1-derived CLs	600-1,079	Ren lab[22]

h1 ES, h1 ES-derived CLs	PLAC-Seq	~300	Ren lab#

Designs of selected datasets: 3



E11.5 E12.5 E13.5 E14.5 E15.5 E16.5 P0: r ɔrn P56: 56 days after

Forebrain

Midbrain

Hindbrain

Neural tube

Heart

Liver

Limb

Craniofacial prominence

8 different mouse tissues at the same development stage (12.5 days post conception)

For forebrain tissue, time-course data, from 11.5 days post conception to 56 days after birth.

Design of a multi-tissue and time course interactomic project.

Designs of selected datasets: 3

For forebrain tissue, time-

Mouse brain tissue from 8 developmental stages	**PLAC-Seq**	**~300**	**Ren lab#**
8 mouse tissues at the same developmental stage	**PLAC-Seq**	**~300**	**Ren lab#**

Design of a multi-tissue and time course interactomic project.

Brief Summary

- **HUGIn**: a handy online visualization tool
- **Primary goal**: help identify potential target gene(s) of regulatory SNPs in a tissue/cell-type specific manner
- 4 different output formats
- **Rich data behind**: 7 human cell lines and 14 human primary tissues. Will add more (analyzing!)
- Rigorous statistics w/ FDR & Bonferroni correction
- Caveats:
  - Different sequencing depths across tissues/cell-lines
  - Mostly 40Kb resolution. We are getting at 4-10Kb.

Acknowledgements

- **University of North Carolina at Chapel Hill**:

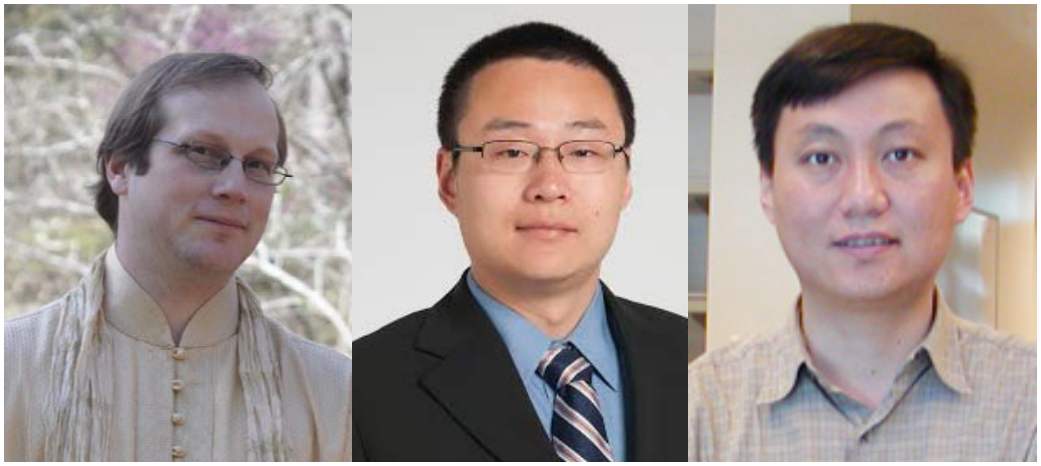
  Josh Martin, Zheng Xu, Karen L. Mohlke, Patrick F. Sullivan

- **Ludwig Institute for Cancer Research, UCSD**:

  Anthony Schmitt, **Bing Ren**

- **Cleveland Clinic**: **Ming Hu**

- **FHCRC**: Alex P. Reiner



Josh **Ming** **Bing**

URL:
http://yunliweb.its.unc.edu/hugin/

Please use it!!
We update & maintain!!

THANK YOU!!