

Encyclopedia Update

August 26, 2016

Michael Purcaro, Henry Pratt, Jill Moore, and Zhiping Weng

Weng Lab

UMass Medical School

Part I –Visualizer 4 Alpha

Alpha release

Try it out!

<http://zlab-annotations-v4.umassmed.edu/>

- Please send comment/problems/suggestions to purcaro@gmail.com
 - Need more visualization ideas!
- Alpha version
 - Has rankings for 3 cell types (HeLa-S3, GM12878, K562)
 - UCSC Genome Browser visualization
 - Has heatmaps for ranks and gene expression
 - Venn diagrams to compare rank similarity between cell types
 - Shopping cart to group and visualize subsets of results

Visualizer 4 Alpha: Results Page

Candidate RE Viz (alpha version) hg19 HeLa-S3 chr14:35792756-35819812 Search 0 ENCODE

Assembly

hg19

Cell types


HeLa-S3

Chromosome

chr14

Coordinates

coordinates: 35792756 - 35819812



TF peak intersection

Enter TF:

Distance to Genes

Candidate RE Search **Nearby Gene Expression** Comparison

Show 10 entries

accession	-log(p)	chr	start	end	nearest gene	nearest protein-coding gene	cart	genome browsers
EE4140684	324.0	chr14	35,799,625	35,799,775	PSMA6	PSMA6		UCSC WashU Ensembl
EE0172427	324.0	chr14	35,800,980	35,801,130	PSMA6	PSMA6		UCSC WashU Ensembl
EE2849498	324.0	chr14	35,801,420	35,801,570	PSMA6	PSMA6		UCSC WashU Ensembl
EE4140654	324.0	chr14	35,802,300	35,802,450	PSMA6	PSMA6		UCSC WashU Ensembl
EE1256909	324.0	chr14	35,803,160	35,803,310	PSMA6	PSMA6		UCSC WashU Ensembl
EE3523940	324.0	chr14	35,803,625	35,803,775	PSMA6	PSMA6		UCSC WashU Ensembl
EE3523944	324.0	chr14	35,804,640	35,804,790	PSMA6	PSMA6		UCSC WashU Ensembl
EE1257002	324.0	chr14	35,804,800	35,804,950	PSMA6	PSMA6		UCSC WashU Ensembl
EE4111457	324.0	chr14	35,805,005	35,805,155	PSMA6	PSMA6		UCSC WashU Ensembl

- Facets for:
- Genome
 - Cell types
 - Genomic position
 - Candidate RE
 - Distance to nearest gene
 - Ranks

- Results table:
- Top 10 candidate REs in region
 - Genome browser links
 - UCSC
 - TODO: WashU + Ensembl
 - Shopping cart links to group interesting results

7 Next

Visualizer 4 Alpha: Candidate RE Details

Assembly
hg19

Cell types
HeLa-S3

Chromosome
chr14

Coordinates
coordinates: 35792756 - 35819812

TF peak intersection
Enter TF:

Distance to Genes

Ranks

← EE4140684

Top-ranked cell types

promoter-like:	K562	GM12878	HeLa-S3
enhancer-like:	K562	GM12878	HeLa-S3
DNase:	GM12878	K562	HeLa-S3
CTCF:	K562	GM12878	HeLa-S3

Nearest genes

symbol	distance (bp)
RALGAPA1	207,933
IGBP1P1	390,082
EAPP	790,859
PTCSC3	805,689
MBIP	968,145

Nearest SNPs

symbol	distance (bp)
rs56791327	3

Nearest candidate REs

symbol	distance (bp)
EE1595817	372,535
EE1582606	2,381,145
EE1582600	2,383,185
EE1230393	2,923,875
EE1589520	2,933,745

Intersecting TF peaks

K562	1 CHD4, 1 CHD7, 1 WHSC1, 1 KAT2B, 1 NCOR1, 1 ZBTB33, 1 CBX3, 1 SETDB1, 1 CREBBP
H1-hESC	1 HDAC2, 1 SIRT6, 1 CHD7, 1 HDAC6, 1 SUZ12, 1 EP300
osteoblast1	EP300

Intersecting histone peaks

HepG2	1 H4K20me1, 1 H3K4me2, 1 H3K4me1, 1 H2AFZ, 1 H3K27me3
A549	1 H3K4me2, 1 H3K4me1, 1 H3K27ac, 1 H2AFZ, 1 H3K36me3
keratinocyte	1 H3K9me1, 1 H2AFZ, 1 H3K36me3
K562	1 H4K20me1, 1 H3K9me3
GM12878	1 H4K20me1, 1 H3K9ac

Candidate RE shopping cart

Candidate RE Viz (alpha version) 4

Assembly

Cell types

Chromosome

Coordinates
coordinates:

TF peak intersection
Enter TF:

Distance to Genes

Ranks

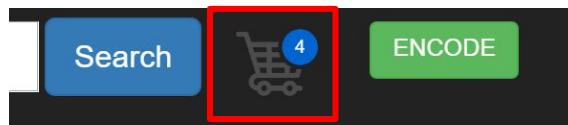
Candidate RE Search Nearby Gene Expression Comparison

Show entries

accession	-log(p)	chr	start	end	nearest gene	nearest protein-coding gene	cart	genome browsers
EE4140684	324.0	chr14	35,799,625	35,799,775	PSMA6	PSMA6		UCSC WashU Ensembl
EE0172427	324.0	chr14	35,800,980	35,801,130	PSMA6	PSMA6		UCSC WashU Ensembl
EE2849498	324.0	chr14	35,801,420	35,801,570	PSMA6	PSMA6		UCSC WashU Ensembl
EE4140654	324.0	chr14	35,802,300	35,802,450	PSMA6	PSMA6		UCSC WashU Ensembl
EE1256909	324.0	chr14	35,803,160	35,803,310	PSMA6	PSMA6		UCSC WashU Ensembl
EE3523940	324.0	chr14	35,803,625	35,803,775	PSMA6	PSMA6		UCSC WashU Ensembl
EE3523944	324.0	chr14	35,804,640	35,804,790	PSMA6	PSMA6		UCSC WashU Ensembl
EE1257002	324.0	chr14	35,804,800	35,804,950	PSMA6	PSMA6		UCSC WashU Ensembl
EE4111457	324.0	chr14	35,805,005	35,805,155	PSMA6	PSMA6		UCSC WashU Ensembl
EE0385318	324.0	chr14	35,805,740	35,805,890	PSMA6	PSMA6		UCSC WashU Ensembl



Candidate RE shopping cart



Clicking cart link takes user to cart page with details

Clicking icons adds or removes regulatory elements from cart

gene	cart	genome browsers
		UCSC WashU Ensembl
		UCSC WashU Ensembl
		UCSC WashU Ensembl
		UCSC WashU Ensembl
		UCSC WashU Ensembl
		UCSC WashU Ensembl
		UCSC WashU Ensembl
		UCSC WashU Ensembl
		UCSC WashU Ensembl
		UCSC WashU Ensembl
		UCSC WashU Ensembl

Candidate RE shopping cart: details page

Summary

[Compare in WashU](#)

Button to visualize in WashU browser (other browsers to come)

Show 10 entries

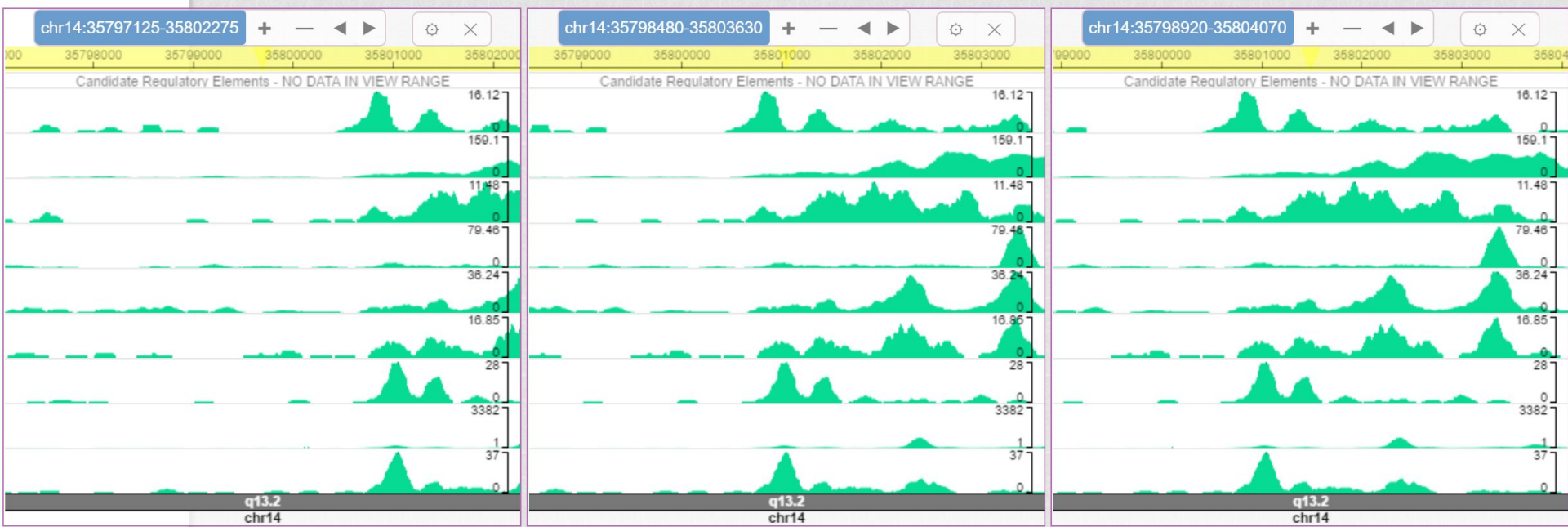
accession	-log(p)	chr	start	end	nearest gene	nearest protein-coding gene	genome browsers
EE2849498	324.0	chr14	35,801,420	35,801,570	PSMA6	PSMA6	UCSC WashU Ensembl
EE4140654	324.0	chr14	35,802,300	35,802,450	PSMA6	PSMA6	UCSC WashU Ensembl
EE0172427	324.0	chr14	35,800,980	35,801,130	PSMA6	PSMA6	UCSC WashU Ensembl
EE4140684	324.0	chr14	35,799,625	35,799,775	PSMA6	PSMA6	UCSC WashU Ensembl

Showing 1 to 4 of 4 entries

Previous 1 Next

Data table shows summary details displayed on search results page

WashU browser visualization




TODOs

- Add more cell types
- Incorporate TF data
- More visualizations!
 - Self-organizing maps? (Sheffield, et., Genome Res 2012)

TODO


Filter Candidate REs by intersection with one or more TFs

Candidate RE Viz (alpha version) 

Assembly:

Cell types:

Chromosome:

Coordinates:


TF peak intersection

Enter TF:

ctcf

Distance to Genes

Ranks

Candidate RE Search Nearby Gene Expression Comparison

Show entries

accession	$-\log(p)$	chr	start	end	nearest gene	nearest protein-coding gene	cart	genome browsers
EE4140684	324.0	chr14	35,799,625	35,799,775	PSMA6	PSMA6		<input type="button" value="UCSC"/> <input type="button" value="WashU"/> <input type="button" value="Ensembl"/>
EE0172427	324.0	chr14	35,800,980	35,801,130	PSMA6	PSMA6		<input type="button" value="UCSC"/> <input type="button" value="WashU"/> <input type="button" value="Ensembl"/>
EE2849498	324.0	chr14	35,801,420	35,801,570	PSMA6	PSMA6		<input type="button" value="UCSC"/> <input type="button" value="WashU"/> <input type="button" value="Ensembl"/>
EE4140654	324.0	chr14	35,802,300	35,802,450	PSMA6	PSMA6		<input type="button" value="UCSC"/> <input type="button" value="WashU"/> <input type="button" value="Ensembl"/>
EE1256909	324.0	chr14	35,803,160	35,803,310	PSMA6	PSMA6		<input type="button" value="UCSC"/> <input type="button" value="WashU"/> <input type="button" value="Ensembl"/>
EE3523940	324.0	chr14	35,803,625	35,803,775	PSMA6	PSMA6		<input type="button" value="UCSC"/> <input type="button" value="WashU"/> <input type="button" value="Ensembl"/>
EE3523944	324.0	chr14	35,804,640	35,804,790	PSMA6	PSMA6		<input type="button" value="UCSC"/> <input type="button" value="WashU"/> <input type="button" value="Ensembl"/>
EE1257002	324.0	chr14	35,804,800	35,804,950	PSMA6	PSMA6		<input type="button" value="UCSC"/> <input type="button" value="WashU"/> <input type="button" value="Ensembl"/>
EE4111457	324.0	chr14	35,805,005	35,805,155	PSMA6	PSMA6		<input type="button" value="UCSC"/> <input type="button" value="WashU"/> <input type="button" value="Ensembl"/>
EE0385318	324.0	chr14	35,805,740	35,805,890	PSMA6	PSMA6		<input type="button" value="UCSC"/> <input type="button" value="WashU"/> <input type="button" value="Ensembl"/>

Showing 1 to 10 of 69 entries

Previous Next

TODO

(After discussions w/ Ross Hardison, Barbara Wold, and Ben Hitz)

- Convert rank table to heatmap across all cell types (or tissues...)

Candidate RE Viz (alpha version) hg19 HeLa-S3 chr14:35792756-35819812 Search ENCODE

Candidate RE Search **RE Details** Nearby Gene Expression Comparison

← **EE0172427**

Assembly

hg19

Cell types

HeLa-S3

Chromosome

chr14

Coordinates

coordinates: 35792756 - 35819812

Distance to Genes

Ranks

Top-ranked cell types

promoter-like:	K562	GM12878	HeLa-S3
enhancer-like:	HeLa-S3	GM12878	K562
DNase:	GM12878	K562	HeLa-S3
CTCF:	K562	GM12878	HeLa-S3

Nearest genes

symbol	distance (bp)
IGBP1P1	391,437
RNU7-41P	486,474
TTC6	2,264,072
AKAP6	2,500,563
RN7SL660P	2,889,287

Nearest SNPs

symbol	distance (bp)
rs140723156	26

Nearest candidate REs

symbol	distance (bp)
EE0092538	1,190,600
EE0093844	1,248,760
EE0089865	2,351,680
EE0093087	2,557,460
EE0080697	3,059,300

Intersecting TF peaks

K562	4 JUN, 2 CREB1, 1 CHD7, 1 JUND, 1 CBX3, 1 KDM1A, 1 KAT2B, 1 ATF3, 1 CHD4, 1 NCOR1, 1 WHSC1, 1 NFE2, 1 ATF1, 1 SETDB1, 1 CREBBP, 1 CREM
MCF-7	2 NR2F2, 2 ELF1, 2 GATA3, 1 HDAC2, 1 RAD21, 1 MAX, 1 FOSL2, 1 MYC, 1 EP300, 1 SIN3A
	2 JUN, 2 CREB1, 2 MAEK, 1

Intersecting histone peaks

A549	2 H3K4me2, 2 H3K4me1, 2 H3K27ac, 2 H2AFZ, 1 H3K36me3, 1 H3K9ac
keratinocyte	1 H3K4me3, 1 H3K4me2, 1 H3K4me1, 1 H3K36me3, 1 H3K27ac, 1 H2AFZ, 1 H3K9me1, 1 H3K9ac
HepG2	2 H3K4me3, 1 H3K4me2, 1 H3K4me1, 1 H2AFZ, 1 H3K27me3, 1 H3K9ac
	2 H3K4me1, 1 H4K20me1, 1

TODO

(After discussions w/ Ross Hardison, Barbara Wold, and Ben Hitz)

- Use consensus TAD calls to filter nearest genes?

Candidate RE Viz (alpha version) hg19 HeLa-S3 chr14:35792756-35819812 Search ENCODE

Assembly

hg19 ⊗

Cell types


HeLa-S3 ⊗

Chromosome

chr14 ⊗

Coordinates

coordinates: 35792756 - 35819812



Distance to Genes

Ranks

Candidate RE Search RE Details Nearby Gene Expression Comparison

← EE0172427

Top-ranked cell types

promoter-like: K562 GM12878 HeLa-S3

enhancer-like: HeLa-S3 GM12878K562

DNase: GM12878K562 HeLa-S3

CTCF: K562 GM12878 HeLa-S3

Nearest genes

symbol	distance (bp)
IGBP1P1	391,437
RNU7-41P	486,474
TTC6	2,264,072
AKAP6	2,500,563
RN7SL660P	2,889,287

Nearest SNPs

symbol	distance (bp)
rs140723156	26

Nearest candidate REs

symbol	distance (bp)
EE0092538	1,190,600
EE0093844	1,248,760
EE0089865	2,351,680
EE0093087	2,557,460
EE0080697	3,059,300

Intersecting TF peaks

K562

4 JUN, 2 CREB1, 1 CHD7, 1 JUND, 1 CBX3, 1 KDM1A, 1 KAT2B, 1 ATF3, 1 CHD4, 1 NCOR1, 1 WHSC1, 1 NFE2, 1 ATF1, 1 SETDB1, 1 CREBBP, 1 CREM

MCF-7

2 NR2F2, 2 ELF1, 2 GATA3, 1 HDAC2, 1 RAD21, 1 MAX, 1 FOSL2, 1 MYC, 1 EP300, 1 SIN3A

2 JUND, 2 CREB1, 2 MAEK, 1

Intersecting histone peaks

A549

2 H3K4me2, 2 H3K4me1, 2 H3K27ac, 2 H2AFZ, 1 H3K36me3, 1 H3K9ac

keratinocyte

1 H3K4me3, 1 H3K4me2, 1 H3K4me1, 1 H3K36me3, 1 H3K27ac, 1 H2AFZ, 1 H3K9me1, 1 H3K9ac

HepG2

2 H3K4me3, 1 H3K4me2, 1 H3K4me1, 1 H2AFZ, 1 H3K27me3, 1 H3K9ac

2 H3K4me1, 1 H4K20me1, 1

TODO

(After discussions w/ Ross Hardison, Barbara Wold, and Ben Hitz)

- Gene expression should use hierarchical clustering

Candidate RE Viz (alpha version) hg19 HeLa-S3 chr14:35819812-35819812 Search 0 ENCODE

Assembly
hg19

Cell types
HeLa-S3

Chromosome
chr14

Coordinates
coordinates: 35819812 - 35819812

TF peak intersection
Enter TF:

Distance to Genes

Ranks

Candidate RE Search Nearby Gene Expression Comparison

Gene	ENSG00000006704.6	ENSG00000010072.11	ENSG00000013619.9	ENSG00000081320.6	ENSG00000148384.11	ENSG00000148484.13	ENSG00000148672.7	ENSG00000148942.10	ENSG00000149571.6	ENSG00000110429.9
aortic smooth muscle cell (ENCSR000AAA, rep1)										
bladder microvascular endothelial cell (ENCSR000AAB, rep1)										
smooth muscle cell of bladder (ENCSR000AAC, rep1)										
bronchial epithelial cell (ENCSR000AAD, rep1)										
bronchial smooth muscle cell (ENCSR000AAE, rep1)										
endothelial cell of coronary artery (ENCSR000AAF, rep1)										
smooth muscle cell of the coronary artery (ENCSR000AAG, rep1)										
regular cardiac myocyte (ENCSR000AAH, rep1)										
dermis blood vessel endothelial cell (ENCSR000AAI, rep1)										
dermis lymphatic vessel endothelial cell (ENCSR000AAJ, rep1)										
dermis microvascular lymphatic vessel endothelial cell (ENCSR000AAK, rep1)										
nasal cavity respiratory epithelium epithelial cell of viscerocranial mucosa (ENCSR000AAL, rep1)										
pulmonary artery endothelial cell (ENCSR000AAM, rep1)										
smooth muscle cell of the pulmonary artery (ENCSR000AAN, rep1)										
fibroblast of lung (ENCSR000AAO, rep1)										
lung microvascular endothelial cell (ENCSR000AAP, rep1)										
renal cortical epithelial cell (ENCSR000AAQ, rep1)										
tracheal epithelial cell (ENCSR000AAR, rep1)										
smooth muscle cell of trachea (ENCSR000AAS, rep1)										
epithelial cell of umbilical artery (ENCSR000AAT, rep1)										
smooth muscle cell of the umbilical artery (ENCSR000AAU, rep1)										
uterine smooth muscle cell (ENCSR000AAV, rep1)										
GM12878 (ENCSR000AEC, rep1)										
GM12878 (ENCSR000AED, rep1)										
GM12878 (ENCSR000AEE, rep1)										
GM12878 (ENCSR000AEF, rep1)										
GM12878 (ENCSR000AEG, rep1)										

Part II – Linking Candidate REs with Genes

Discussion at ENCODE Meeting

What Genes Might Be Regulated by This Element?

Physical connectivity between elements and genes should be presented as part of the third layer (?)

The question of regulatory implications of a physical connection is key for our users, but not yet easy for us to arbitrate. How to handle? Call them EL0000n (Encode Link) or EI00000n (Encode Interaction)?

We therefore aim to provide all evidence for a given regulatory element (and vice versa for a gene) along with its statistical significance. How much of this is ready to go into version 1 of encyclopedia?

Explicitly state the type of evidence per element: chromatin interactions, eQTLs, or functional assays (knock-out)

Creating Gene Links: Proposed Plan

Candidate RE

EE000001

EL00045



Gene

EP300

Evidence supporting link:

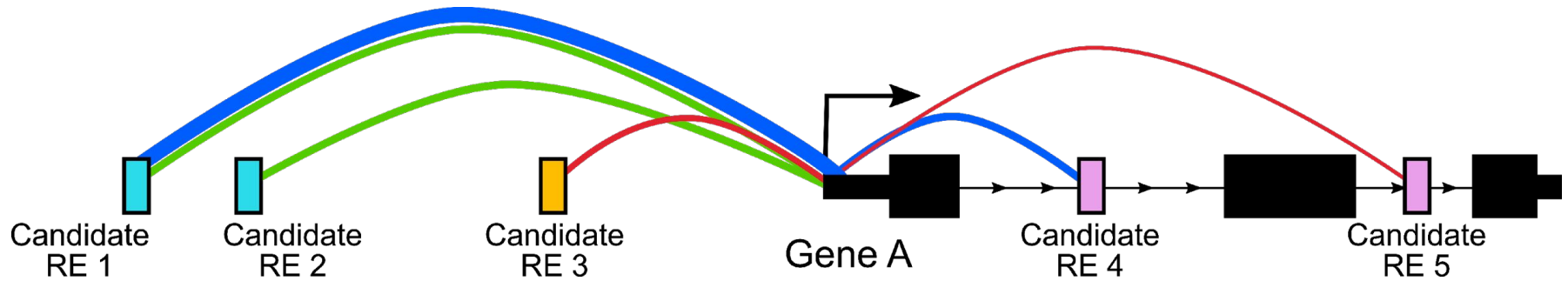
Method	Tissue/Cell Type	Confidence Score
CTCF ChIA-PET	GM12878	56.8
eQTLs	Whole blood	4.5
Computational Prediction	GM12878	10.3

Creating Gene Links: Proposed Plan

- Users can search for links using candidate RE or gene
- Users will be able to select facets to filter their results. Such facets will include:
 - Cell type
 - Linking method (ChIA-PET, eQTLs, computation predictions etc.)
 - Confidence of link

Visualization Ideas

Genome Browser-like View

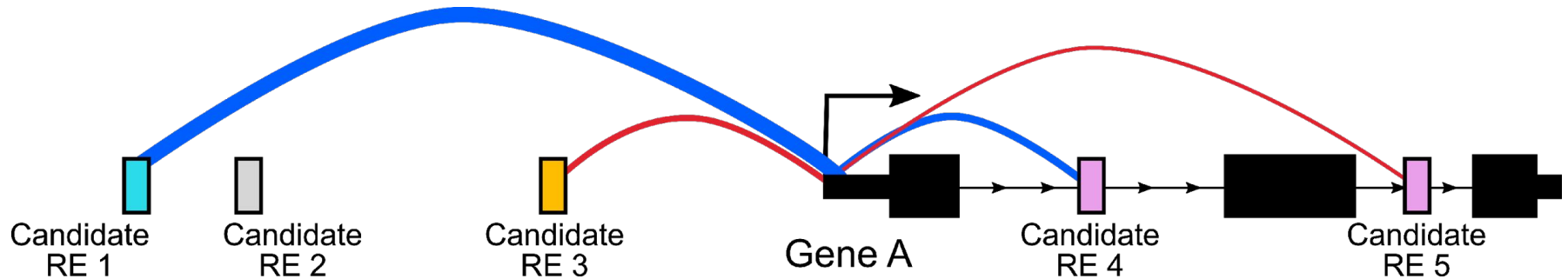


Tissue/Cell Type
GM12878
K562
H1-hESC

Method
ChIA-PET
eQTLs
Computational Predictions

Line thickness indicates confidence

User Selects Facets to Filter Results



Tissue/Cell Type
GM12878
K562
H1-hESC

Method
ChIA-PET
eQTLs
Computational Predictions

Line thickness indicates confidence

Network View

Tissue/Cell Type

GM12878

K562

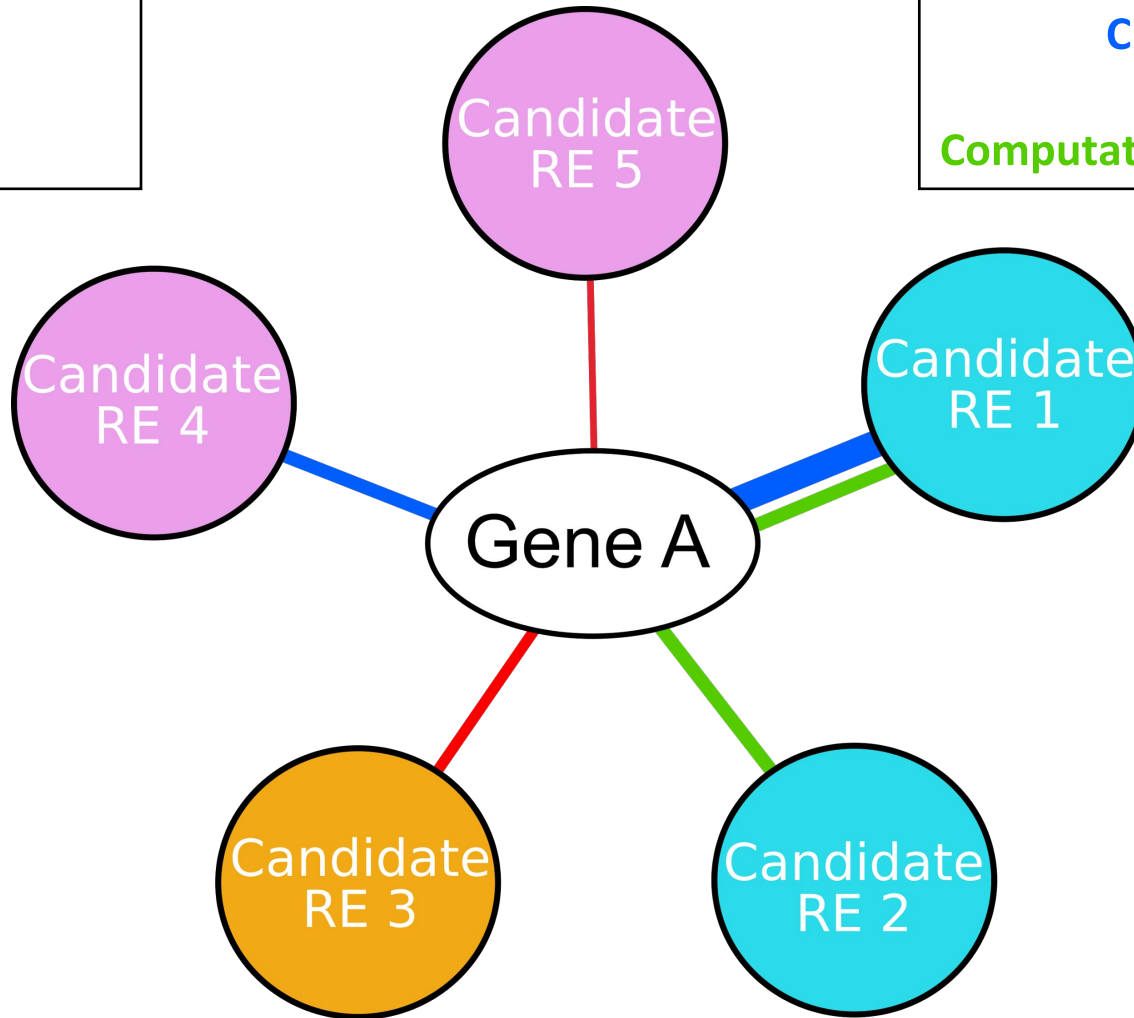
H1-hESC

Method

ChIA-PET

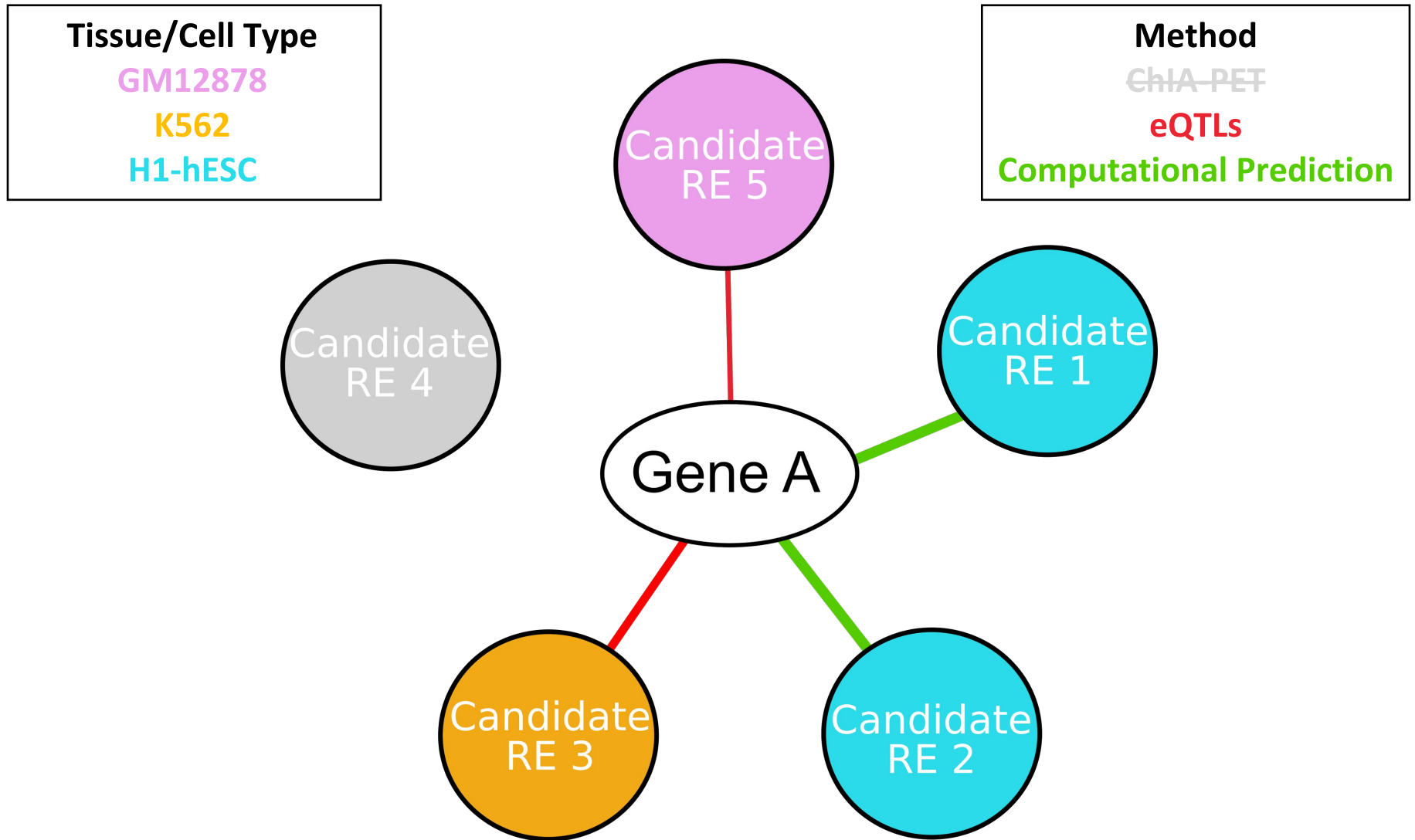
eQTLs

Computational Prediction



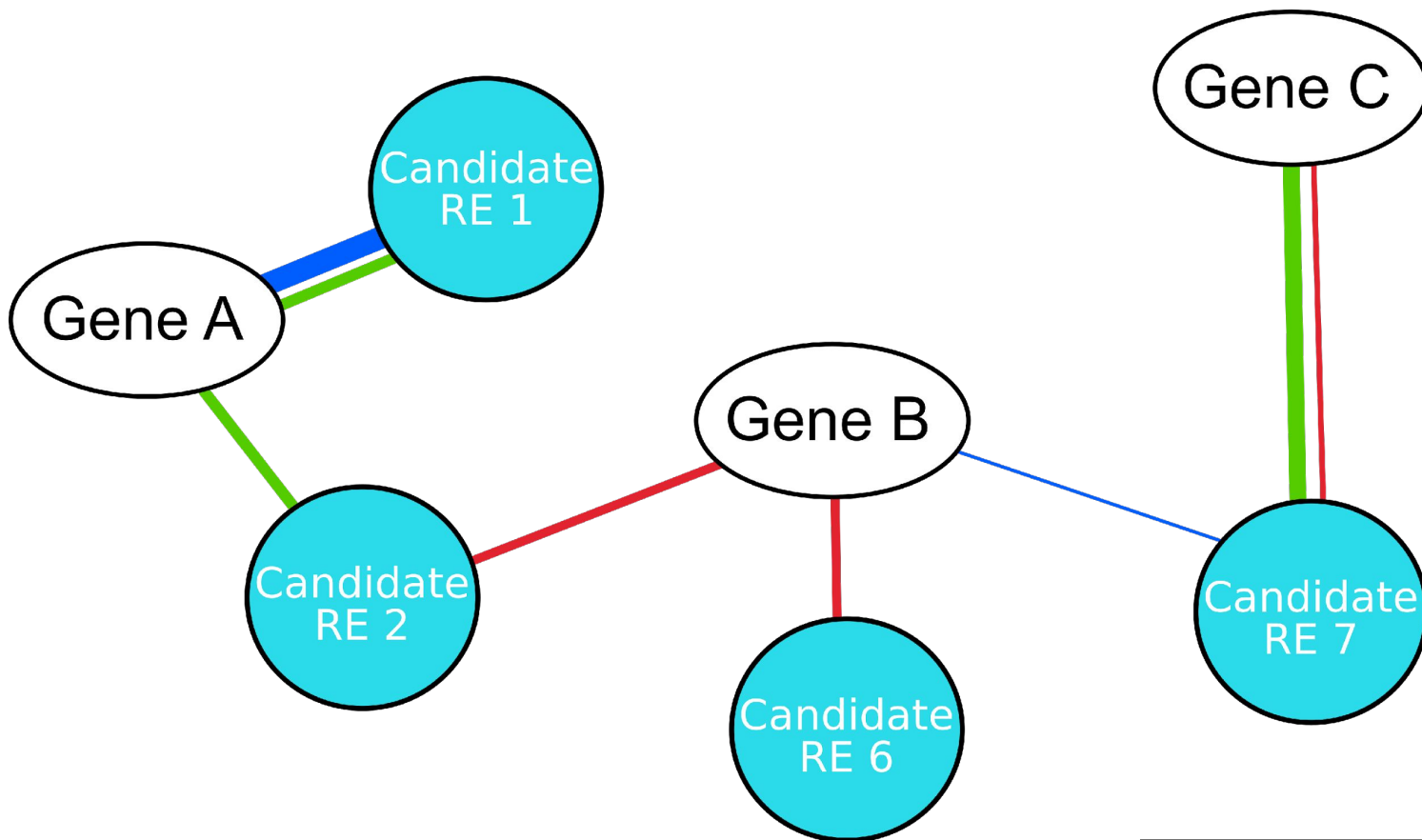
Line thickness indicates confidence

User Selects Facets to Filter Results



Line thickness indicates confidence

User Can Explore Network of Links Using Facets



Tissue/Cell Type

GM12878

K562

H1-hESC

Method

ChIA-PET

eQTLs

Computational Prediction

Line thickness indicates confidence

Moving Forward...

1. Assemble links for ChIA-PET and eQTL data
2. Continue developing Random Forest method:
 - Create new training/testing/validation sets for V4 Encyclopedia
 - Test method across cell-types

Optional open discussion on V4 alpha

<http://zlab-annotations-v4.umassmed.edu/>