



T2D Knowledge Portal Platform Webinar & Workshop

November 14, 2019

ACCELERATING MEDICINES PARTNERSHIP (AMP)

TYPE 2 DIABETES

TODAY: LocusZoom & the T2DKP



- Ryan Welch
- Andy Boughton
- Q & A
– *The team*



What will you learn today?

- Preview of our future directions
- How to find LocusZoom in the T2DKP
- How do you use LocusZoom in the T2DKP and what type of feature are dynamic
- What is to come for the tool
- Have your questions addressed directly by the Developers



What will you learn at future webinars?

- How to access the KPN and T2DKP results programmatically- Smart APIs
- What aggregation tests can you run in real time on the site?
- How are we expanding to metabolic disorders?
- What new workflows are available for gene/variant prioritization?
- What methods are run on the GWAS datasets and how to access the results?
- How will we add to the list of predicted effector genes for T2D and its complications?
- How does the T2DKP represent data residing in other geographic locations?

What other events are coming up?

- Production release December 2019
 - More data, enhanced features and tools
- New Videos tutorials
 - LocusZoom, Data intake pipeline, Using the Causal Variant Query tool, API access, and more
- Booth and Talk at Endocrine Society 2020, ADA 2020
- Read out from our User Survey



TYPE 2 DIABETES
KNOWLEDGE PORTAL



TYPE 2 DIABETES
KNOWLEDGE PORTAL

Providing data and tools to promote
understanding and treatment of
type 2 diabetes and its complications

Learn about the portal

Explore by gene, variant, or region

Explore by phenotype

examples: SLC30A8 rs13266634 chr9:21,940,000-22,190,000

GO >

NEW Tutorial video: Custom aggregation tests NEW Webinar video: Diabetes Epigenome Atlas

About the Portal

The T2D Knowledge Portal enables browsing, searching, and analysis of human genetic information linked to type 2 diabetes and related traits, while protecting the integrity and confidentiality of the underlying data.



80 datasets, 189 traits

Browse data here >

KPN Knowledge Portal Network

The Knowledge Portal Network is an infrastructure that integrates, interprets, and presents human genetic data to spark insights into complex diseases.



Explore genetic data related to type 2 diabetes:
Type 2 Diabetes Knowledge Portal
Visit portal



Explore genetic data related to cardiovascular disease:
Cardiovascular Disease Knowledge Portal
Visit portal



Explore genetic data related to stroke:
Cerebrovascular Disease Knowledge Portal
Visit portal



Explore genetic data related to sleep disturbance and circadian traits:
Sleep Disorder Knowledge Portal
Visit portal

What's new

Connect with the Knowledge Portal Network team at #ASHG19: Attending the American Society of Human Genetics Annual Meeting next week? We are too, and... [Read more](#)

About the project

The Knowledge Portal is being developed by a team of scientists and software engineers at the Broad Institute, the University of Michigan, University of Oxford, and many other collaborators as part of a worldwide scientific consortium with contributors from academia, industry, and non-profit organizations.

We welcome the involvement of interested researchers. [Click here to learn more](#) about contributing data or collaborating with us on analyses, methods, or tool development. Or [contact us](#) for more information.

The AMP T2D Consortium is a collaboration among the following organizations, which also provide funding and/or governance:



Funding and guidance are also provided by:



The AMP-T2D Knowledge Portal

*An open access resource
providing data & tools to
promote understanding
of type 2 diabetes & its
complications*

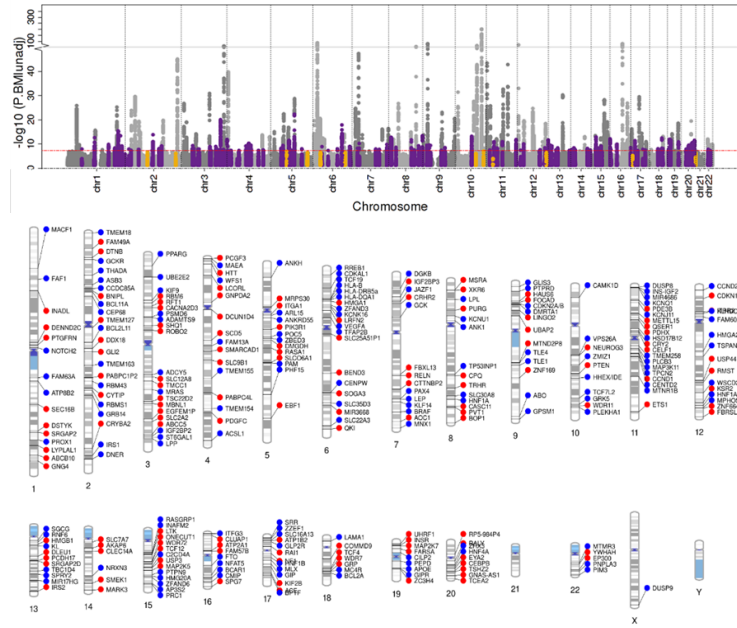
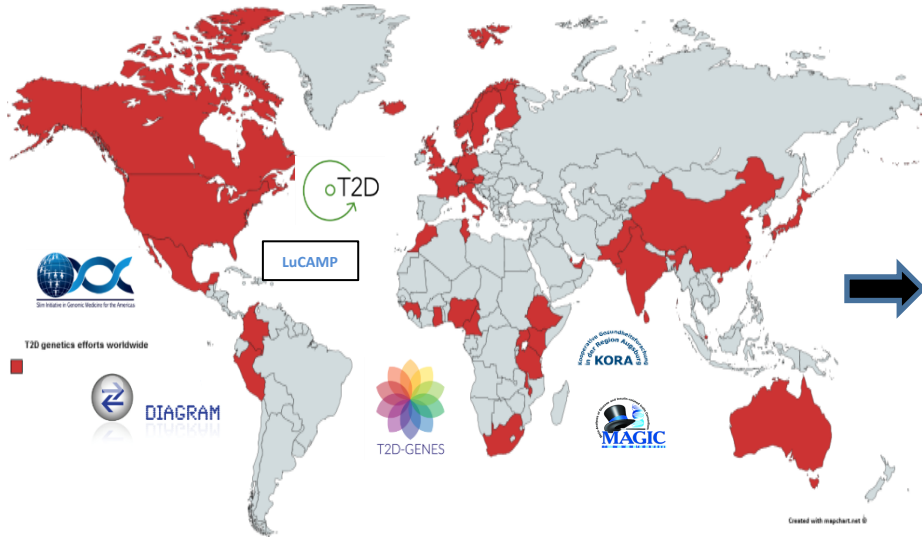
type2diabetesgenetics.org



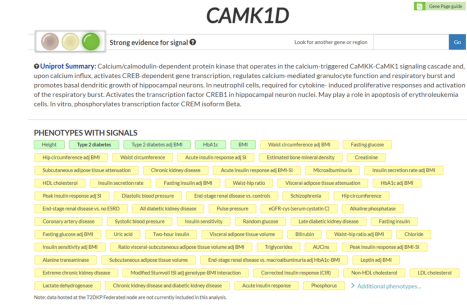
A simple idea:

Making genetic & related genomic data more broadly accessible & useful could have a significant impact on our ability to understand or treat human disease

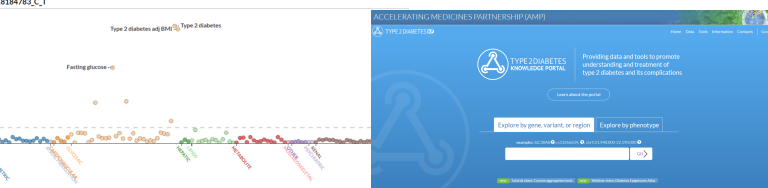
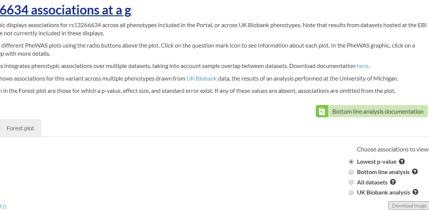
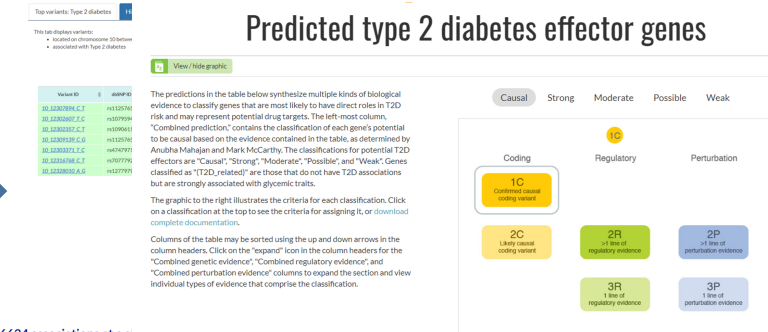
Addressed the first opportunity & challenge



Mahajan et al. Nat Gen. 2018



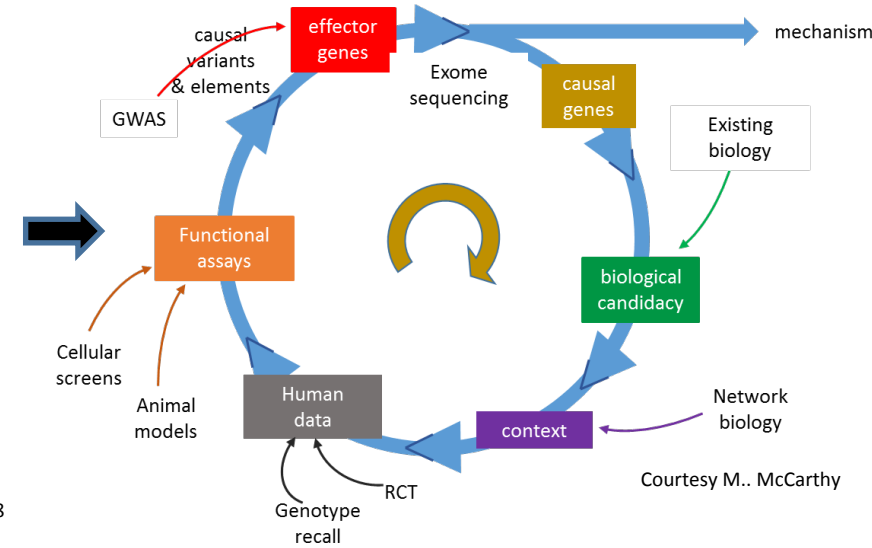
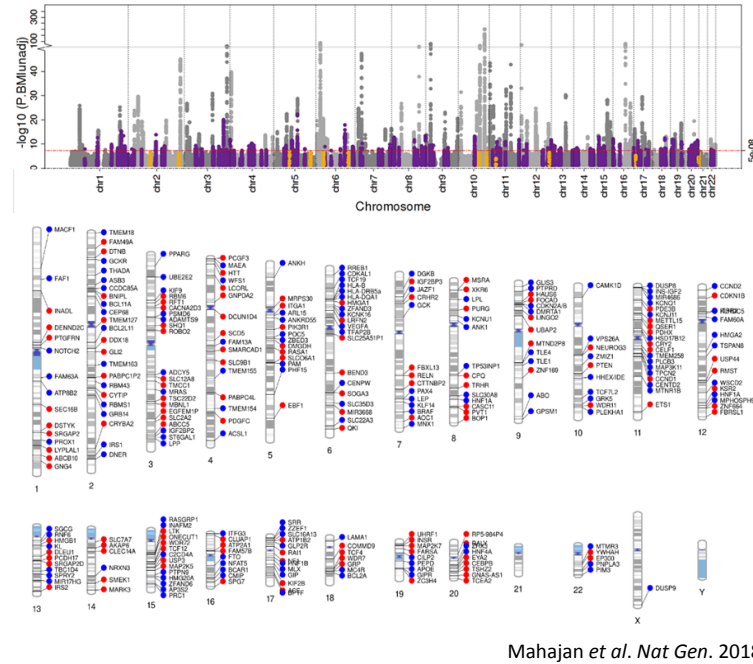
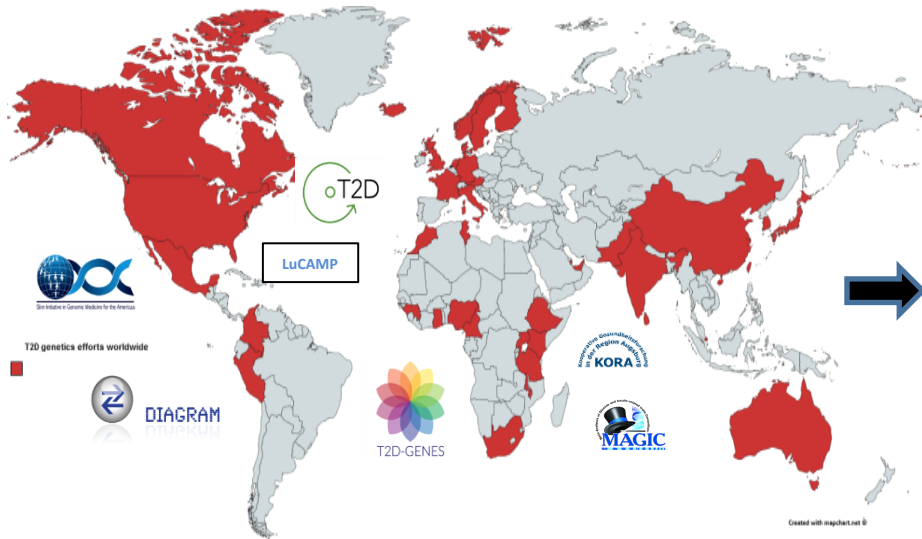
80 datasets, 189 traits



Motivated T2D Genetics Research community

Identification of over 403 loci

A shift in the coming years...a new opportunity & challenge



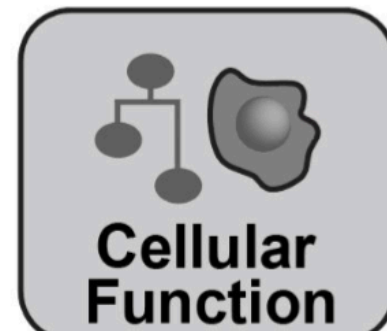
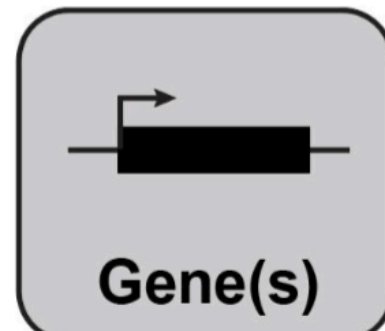
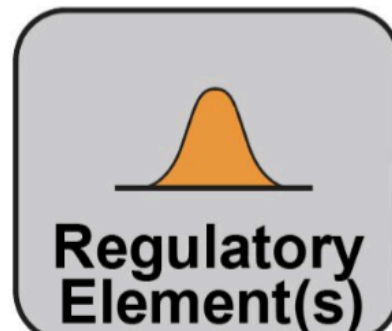
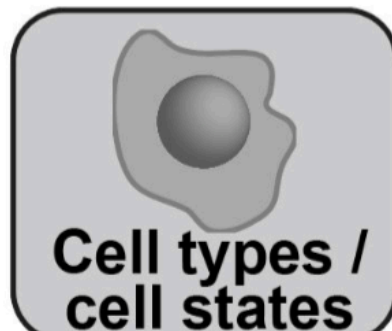
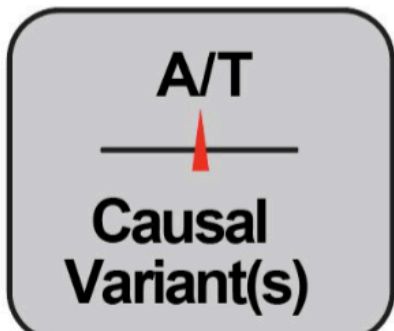
Motivated T2D Genetics Research community

Identification of over 403 loci

Mechanism, magnitude, markers for T2D genes

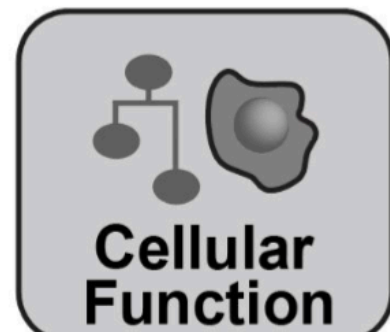
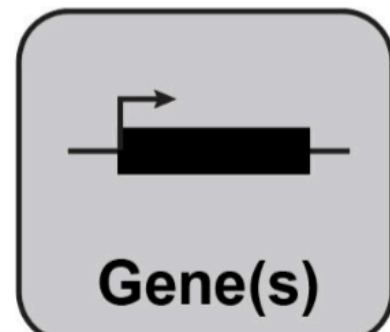
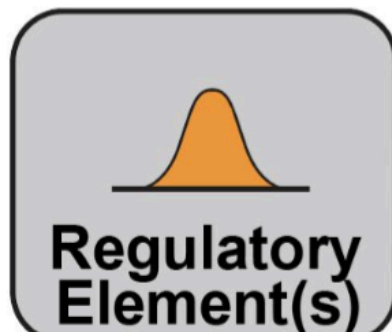
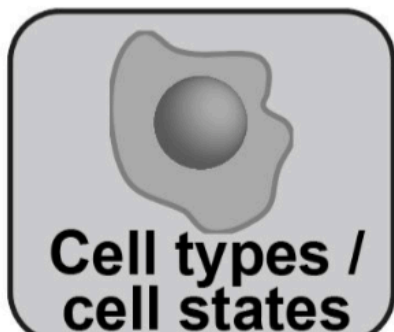
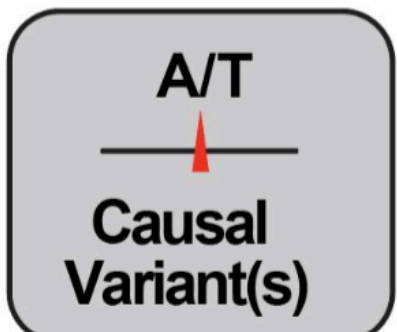
Shift in focus over the coming years- function

- What is the variant?
- What is the regulatory effect and in what tissue?
- What is the gene?
- What is the pathway?
- What is the mechanism?

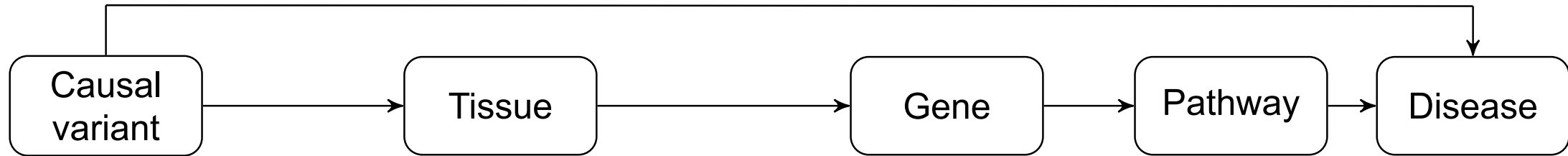
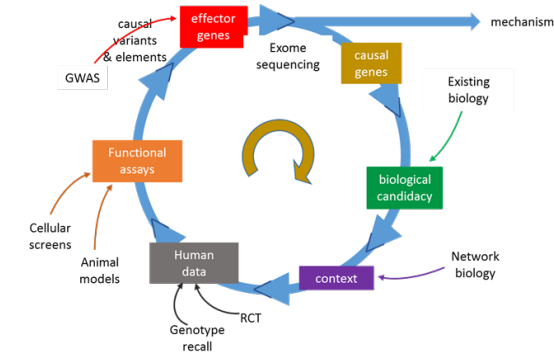


New set of gaps to address with this opportunity

- What are the needed datasets/types (validation)?
- What information must be captured/retained & represented?
- What methods need to be run and how are they validated?
- How do you express relationships between these outputs?
- How do you represent results from experimental work in a computational framework/open access resource?



Catalog & integrate these resources & visualize these together



- Association statistics
 - Meta-analysis
 - Fine mapping
 - Variant effect predictors
- Reference chromatin state
 - Transcription factor binding sites
 - eQTLs
 - Regulatory element prediction
 - Tissue-specific enrichment
- Chromatin capture
 - Gene expression
 - Gene function
 - Gene prioritization
 - eQTL colocalization
 - Chromosome contact prediction
- Networks
 - Pathways
 - Gene set enrichment
- Cellular models
 - Animal models

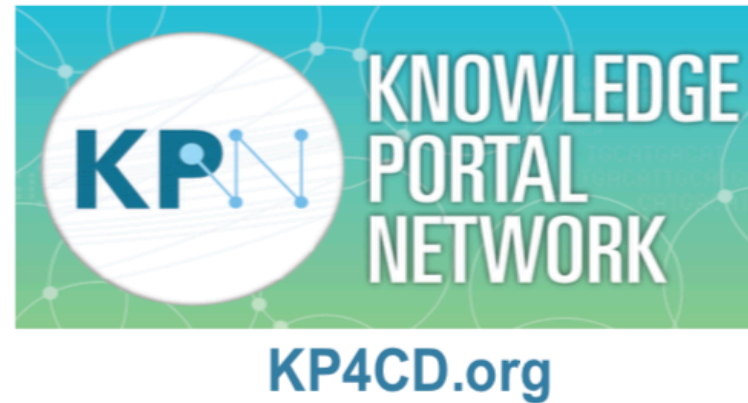
SmartAPIs and web portals

Onward...

- We are writing for our renewal and your ideas are vital!
- The next phase of the T2DKP will marry the best of the *community driven data representation platform & software architecture* we have built for human genetics data with the *new datatypes, methods & experimental approaches* from the *Functional Genomics community* to present within a fused platform to deliver insight for mechanism, magnitude & markers for T2D effector genes.
- Happy World Diabetes Day!



Finding LocusZoom in the Knowledge Portals



Gene/Region Page

PPARG

Gene Page guide



Strong evidence for signal ?

Look for another gene or region

Go

Uniprot Summary: Receptor that binds peroxisome proliferators such as hypolipidemic drugs and fatty acids. Once activated by a ligand, the receptor binds to a promoter element in the gene for acyl-CoA oxidase and activates its transcription. It therefore controls the peroxisomal beta-oxidation pathway of fatty acids. Key regulator of adipocyte differentiation and glucose homeostasis. Acts as a critical regulator of gut homeostasis by suppressing NF-kappa-B- mediated proinflammatory responses.

PHENOTYPES WITH SIGNALS

Phenotypes with signals for PPARG:

- Type 2 diabetes
- Type 2 diabetes adj BMI
- Height
- Total cholesterol
- Estimated bone mineral density
- BMI
- Waist-hip ratio adj BMI
- Waist-hip ratio
- LDL cholesterol
- Triglycerides
- Hip circumference adj BMI
- HbA1c
- Potassium
- Creatinine
- eGFR-creat (serum creatinine)
- Waist circumference
- HDL cholesterol
- Fasting insulin
- Modified Stumvoll ISI adj genotype-BMI interaction
- Chronic kidney disease
- Blood urea nitrogen
- Body fat percentage
- Modified Stumvoll ISI adj age-sex-BMI
- Microalbuminuria
- Waist circumference adj BMI
- End-stage renal disease vs. no ESRD
- Hip circumference
- HbA1c adj BMI
- Bone fracture
- End-stage renal disease vs. controls
- Acute insulin response
- Peak insulin response adj SI
- All diabetic kidney disease adj HbA1c-BMI
- Systolic blood pressure
- Insulin sensitivity adj BMI
- Peak insulin response adj BMI-SI
- Visceral adipose tissue volume adj BMI
- Acute insulin response adj SI
- Acute insulin response adj BMI-SI
- Fasting insulin adj BMI
- Pericardial adipose tissue volume adj height-weight
- Coronary artery disease
- Peak insulin response
- Bilirubin
- Chylomicrons and XXL-VLDL cholesterol
- Urinary albumin-to-creatinine ratio
- Serum albumin
- Insulin sensitivity

[Additional phenotypes...](#)

Note: data hosted at the T2DKP Federated node are not currently included in this analysis.

Top variants: Type 2 diabetes

High-impact variants: Type 2 diabetes

Credible sets: Type 2 diabetes

Genes in region: Type 2 diabetes



Gene/Region Page: Top variants tab

Top variants: Type 2 diabetes

High-impact variants: Type 2 diabetes

Credible sets: Type 2 diabetes

Genes in region: Type 2 diabetes

This tab displays variants:

- located on chromosome 3 between 12228817 and 12575905
- associated with [Add / Subtract Data](#)

Variant ID	dbSNP ID	Major allele	Minor allele	Predicted impact	p-Value	Effect	MAF	Data set
3_12336507_G_A	rs11709077	G	A	intron	1.60e-27	0.896	0.12	DIAMANTE (European) T2D GWAS
3_12329783_C_T	rs17036160	C	T	intron	2.80e-27	0.896	0.12	DIAMANTE (European) T2D GWAS
3_12344730_C_G	rs11712037	C	G	intron	7.00e-27	0.905	0.12	DIAMANTE (European) T2D GWAS
3_12351521_T_G	rs35000407	T	G	intron	8.80e-27	0.905	0.12	DIAMANTE (European) T2D GWAS
3_12365308_G_T	rs13083375	G	T	intron	1.10e-26	0.905	0.12	DIAMANTE (European) T2D GWAS
3_12369401_C_T	rs13064760	C	T	intron	1.20e-26	0.905	0.12	DIAMANTE (European) T2D GWAS
3_12375956_C_T	rs2012444	C	T	intron	1.70e-26	0.905	0.12	DIAMANTE (European) T2D GWAS
3_12386337_C_T	rs4684847	C	T	intron	6.30e-26	0.905	0.12	DIAMANTE (European) T2D GWAS
3_12396913_G_A	rs71304101	G	A	intron	1.30e-25	0.905	0.12	DIAMANTE (European) T2D GWAS
3_12393125_C_G,A	rs1801282	C	G	missense	1.70e-25	0.905	0.12	DIAMANTE (European) T2D GWAS

To add a new track, select a phenotype, then a dataset

1 Phenotype

2 Dataset

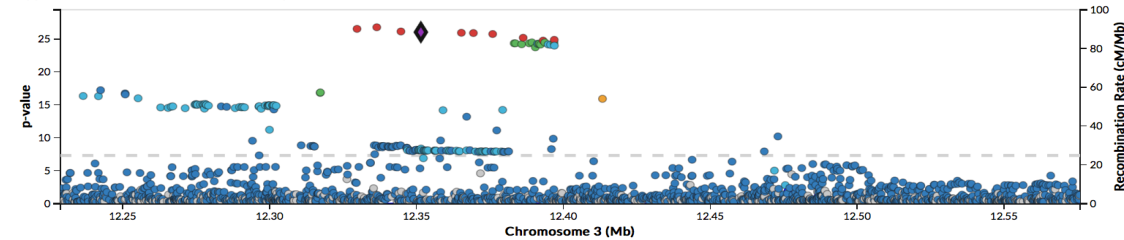
Tissues

LocusZoom v0.9.0

Model

Download Image

Type 2 diabetes (DIAMANTE (European) T2D GWAS)



adipose tissue

pancreatic islets

Gene/Region Page: Credible sets tab

Top variants: Type 2 diabetes High-impact variants: Type 2 diabetes **Credible sets: Type 2 diabetes** Genes in region: Type 2 diabetes

Credible sets in this range ⓘ

10041

Click on a variant ID to see more information, and to access a link for tissue selection. →

		3,123,447,90_CG	3,123,497,783_CT	3,123,999,943_GA	3,123,950,7_GA	3,123,921,245_CG	3,123,898,988_AG	3,123,859,988_CT	3,123,899,801_CT	3,123,919,21_TG	3,123,959,56_CT
annotation	Coding										
	Splice site										
	UTR										
association	Posterior probability	0.086	0.064	0.10	0.045	0.050	0.045	0.042	0.051	0.058	0.066
	P value	5e-8	7e-8	5e-8	1e-7	9e-8	1e-7	1e-7	9e-8	8e-8	7e-8
tissue	adipose tissue	■									
	hASC-t4 adipose stem cells			■	■						
	hASC-t3 adipose stem cells			■	■						
	hASC-t1 adipose stem cells			■	■						
	NHLF lung fibroblast primary cells				■	■					
	hASC-t2 adipose stem cells				■	■					
	rectal mucosa				■	■					
	HMEC mammary epithelial prim...				■	■					

■ Genic enhancer ■ Active enhancer 1 ■ Active enhancer 2 ■ Weak enhancer

To add a new track, select a phenotype, then a dataset

1 Phenotype ▾

2 Dataset ▾

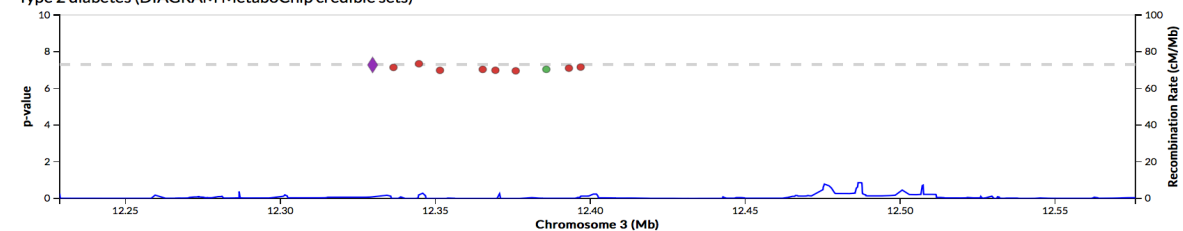
Tissues ▾

LocusZoom v0.9.0

Model

Download Image

Type 2 diabetes (DIAGRAM MetaboChip credible sets)



Variant Page

rs11709077 summary

This variant is located on chromosome 3 at position 12336507 in the human genome build hg19. It lies in the gene [PPARG](#). The nucleotide at this position in the reference sequence is G, and the variant allele is A. All associations shown in the Portal for rs11709077 refer to the variant allele.

Transcripts

Transcript	ENST00000309576	ENST00000397010	ENST00000397012	ENST00000397015	ENST00000397026	ENST0000039702
Protein change	p--	p--	p--	p--	p--	p--
Consequence	intron_variant	intron_variant	intron_variant	intron_variant	intron_variant	intron_variant
PolyPhen prediction	-	-	-	-	-	-
SIFT prediction	-	-	-	-	-	-



rs11709077 associations at a glance

> Associations across all datasets

> Epigenomic annotations

> Genetic Association Interactive Tool



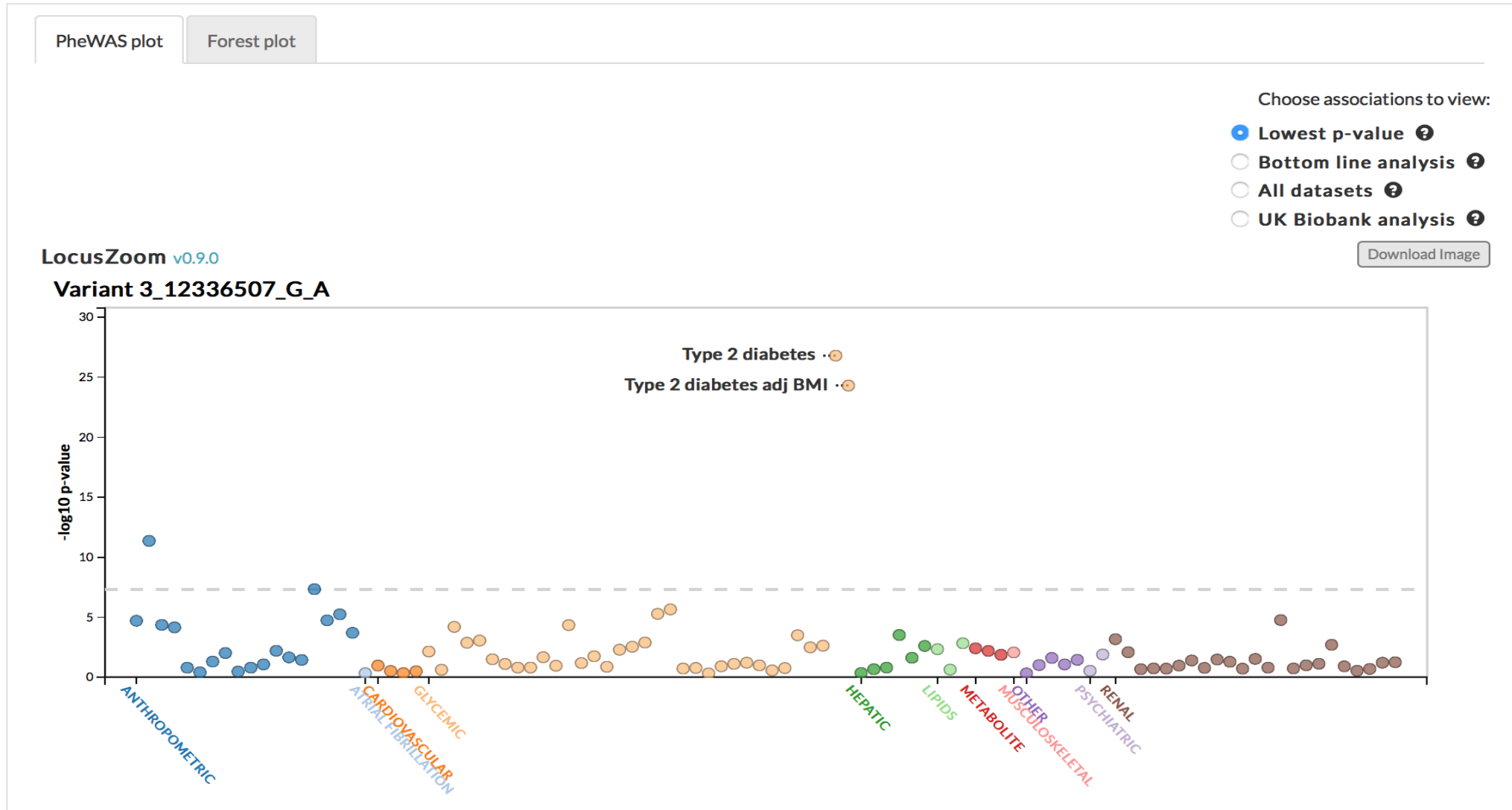
> LocusZoom: view and analyze variants around rs11709077

> How common is rs11709077?

External resources

[GWAS Catalog](#) [GTEx](#) [PheWeb](#) [SNPedia](#) [Diabetes Epigenome Atlas](#) [UCSC Genome Browser](#)

Variant Page: associations at a glance



Variant Page: LocusZoom

✓ LocusZoom: view and analyze variants around rs11709077

Use the [LocusZoom](#) interactive visualization to explore associations of the variants in this region. The reference variant for this page is shown in purple. First select a phenotype, then select a dataset to add a panel displaying those associations. To zoom in or out, hold the Shift key while scrolling. Click on a gene to see its Ensembl Gene and Transcript IDs and link to its [gnomAD](#) page.

"Dynamic" datasets are drawn from individual-level data and allow you to re-analyze the association data while conditioning on any selected variant. "Static" datasets are drawn from summary association statistics, and association data are conditioned on the reference variant.

In genomic regions where there are many variant associations, LocusZoom displays only the most significantly associated variants in order to improve performance.

Linkage disequilibrium (r^2) values are based on the 1000G ALL reference panel (including every sample available in the [1000 Genomes project](#)) and are supplied by the [Michigan Imputation Server](#).

Linkage disequilibrium (r^2) with the reference variant: ● 1 - 0.8 ● 0.8 - 0.6 ● 0.6 - 0.4 ● 0.4 - 0.2 ● 0.2 - 0 ● no information ● reference variant

To add a new track, select a phenotype, then a dataset

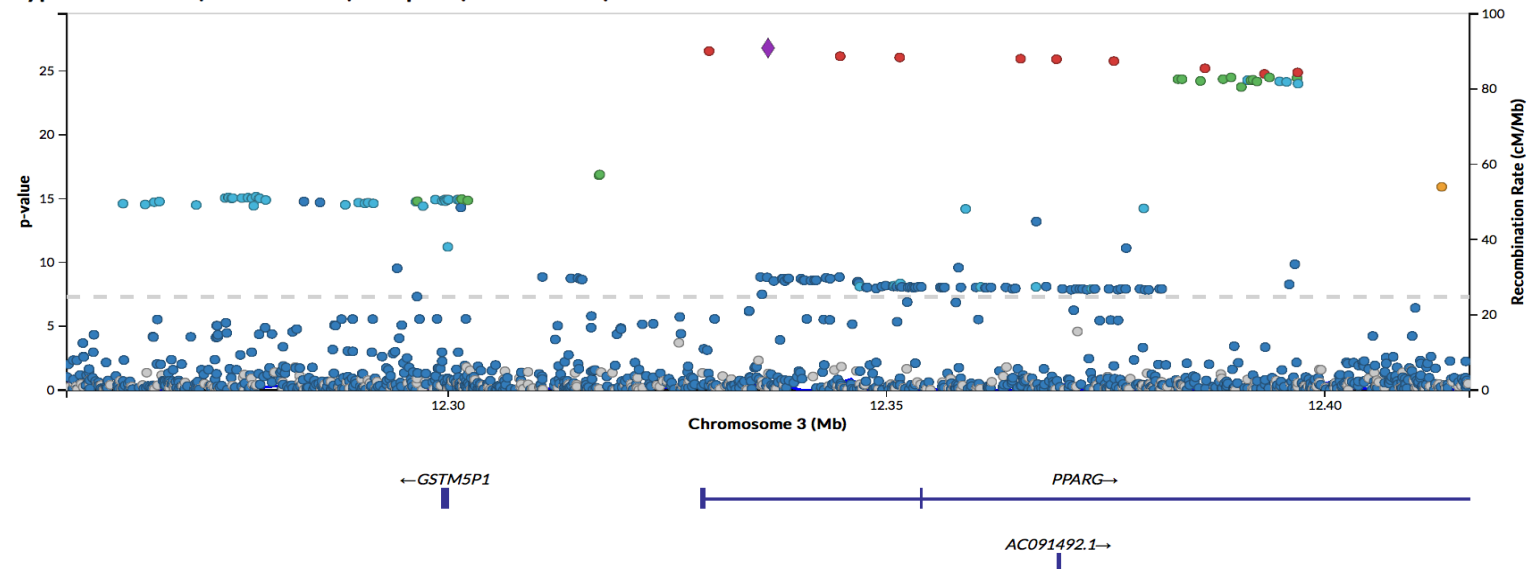
1 Phenotypes ▾ 2 Datasets ▾ Region: 3:12256507-12416507

LocusZoom v0.9.0

Model

Download Image

Type 2 diabetes (DIAMANTE (European) T2D GWAS)



Overview of LocusZoom.js in the T2D Knowledge Portal

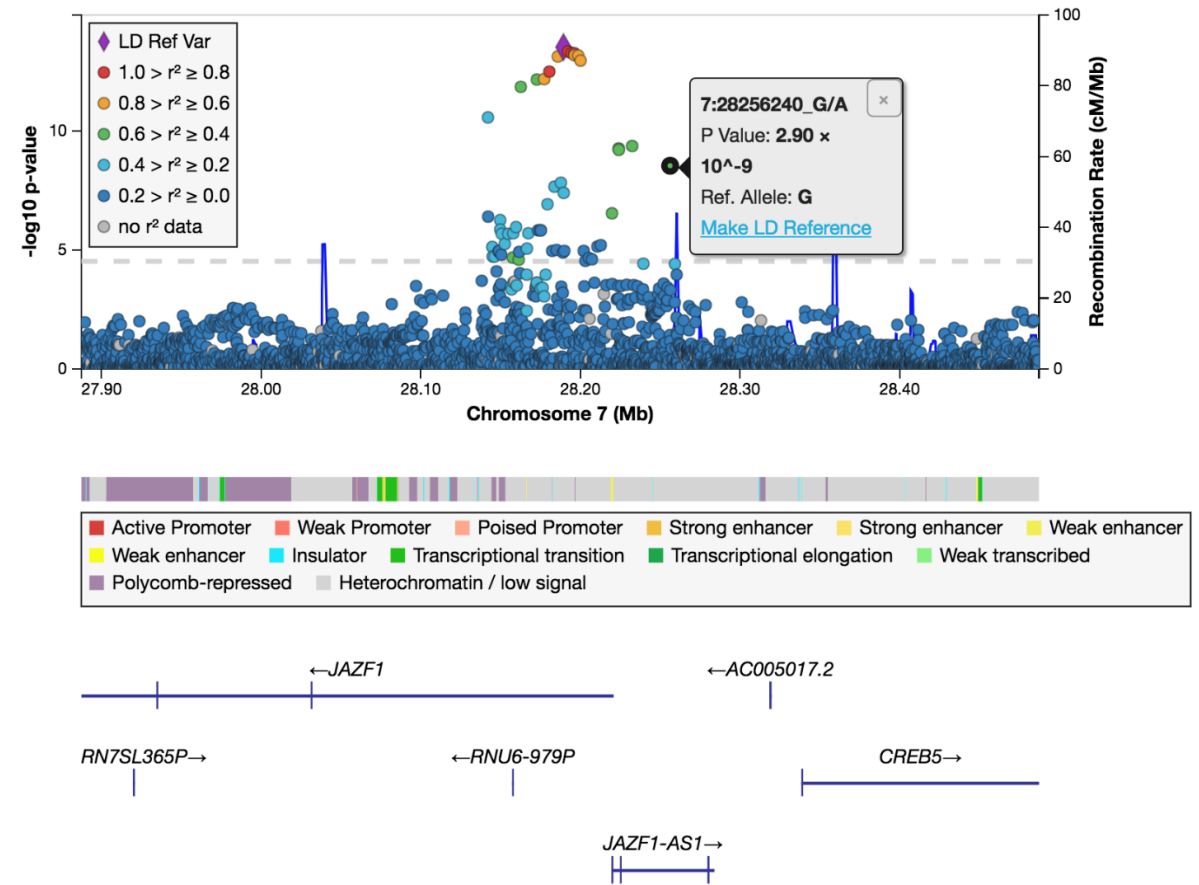
Ryan Welch and Andy Boughton
University of Michigan Center for
Statistical Genetics

Acknowledgements

- **Michigan:** Daniel Taliun, Peter VandeHaar, Chris Clark, Matthew Flickinger, Sarah Gagliano Taliun, Gonçalo Abecasis, Mike Boehnke
- **T2D Portal Team:** **Ben Alexander**, Jeffrey Massung, Marc Duby, Maria Costanzo, Noel Burtt, Jason Flannick

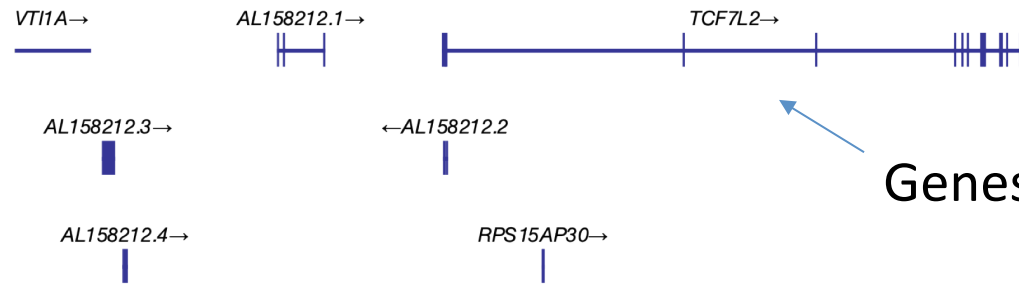
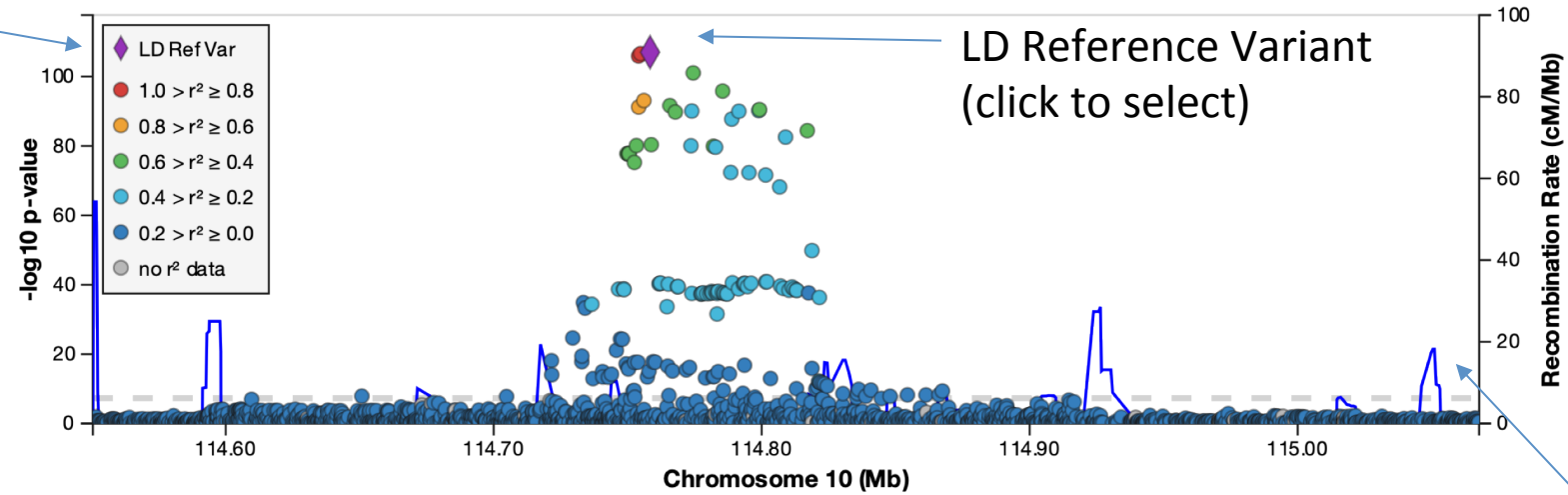
LocusZoom: History and Purpose

- Designed to visualize GWAS summary statistics in context
- A command line version has been used for ~10 years, and is widely established
- New interactive version supports exploration and data sharing



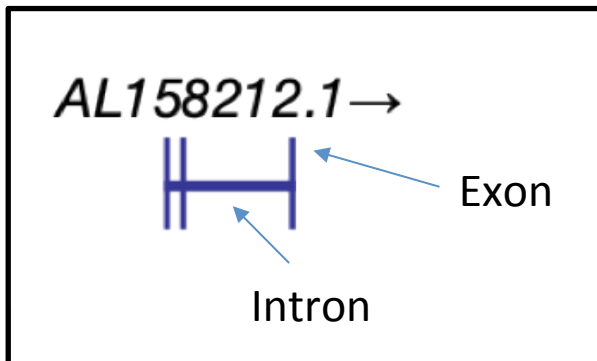
How to read a LocusZoom Plot

Color by Linkage Disequilibrium (LD)

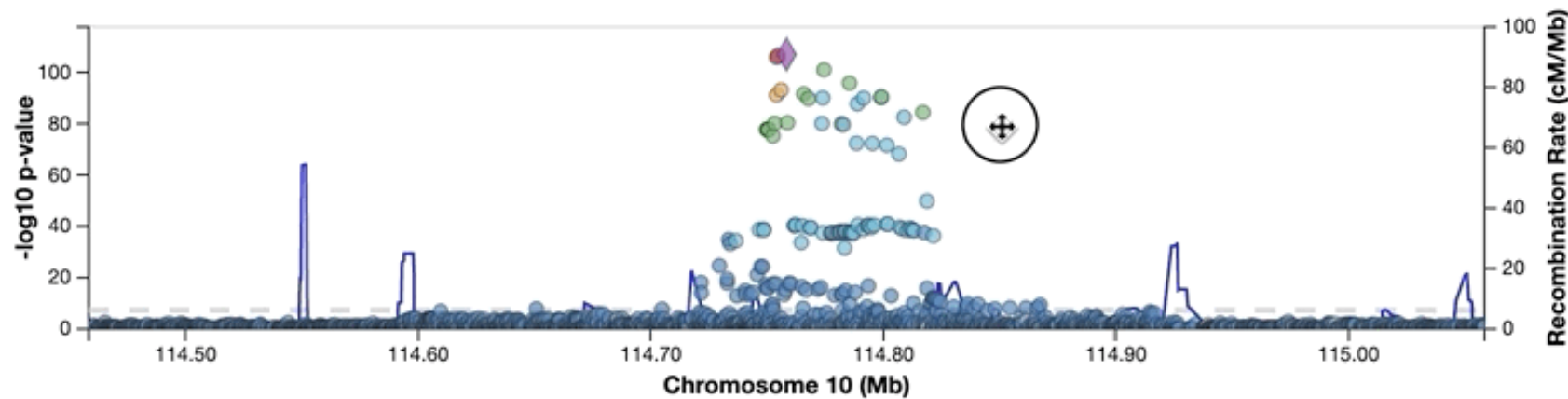


Recombination rate

Genes in viewing region





Interactively control the view




9:22134253_G/A
P Value: 5.72×10^{-87}
Ref. Allele: G
[Expanded view](#)
[Tissues with overlapping enhancer regions](#)
[Make LD Reference](#)

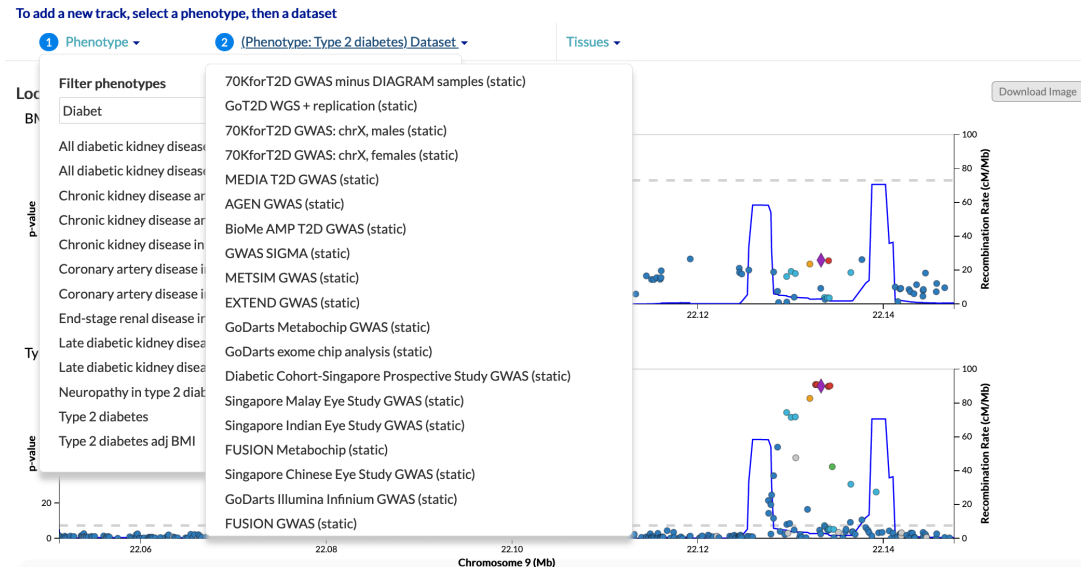


Drag to pan 

 Scroll to zoom

Click to interact 

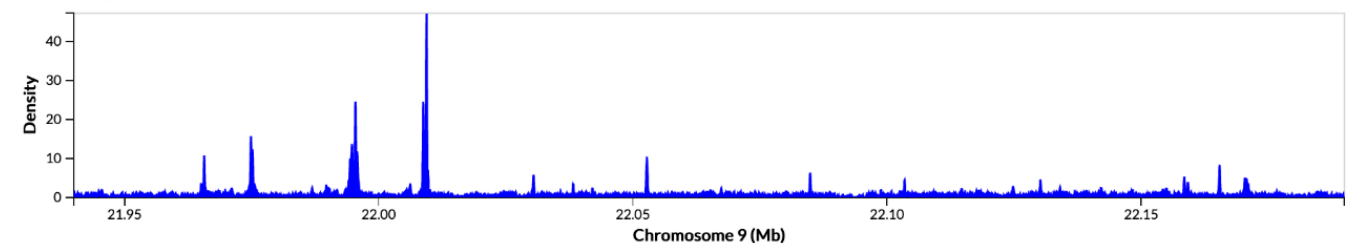
Explore a **wide range** of datasets in depth



- Choose from GWAS summary statistics for a wide range of phenotypes and study datasets
- Stack panels vertically and make custom comparisons
- Add tissue-specific annotations and ATAC-seq data within the plot

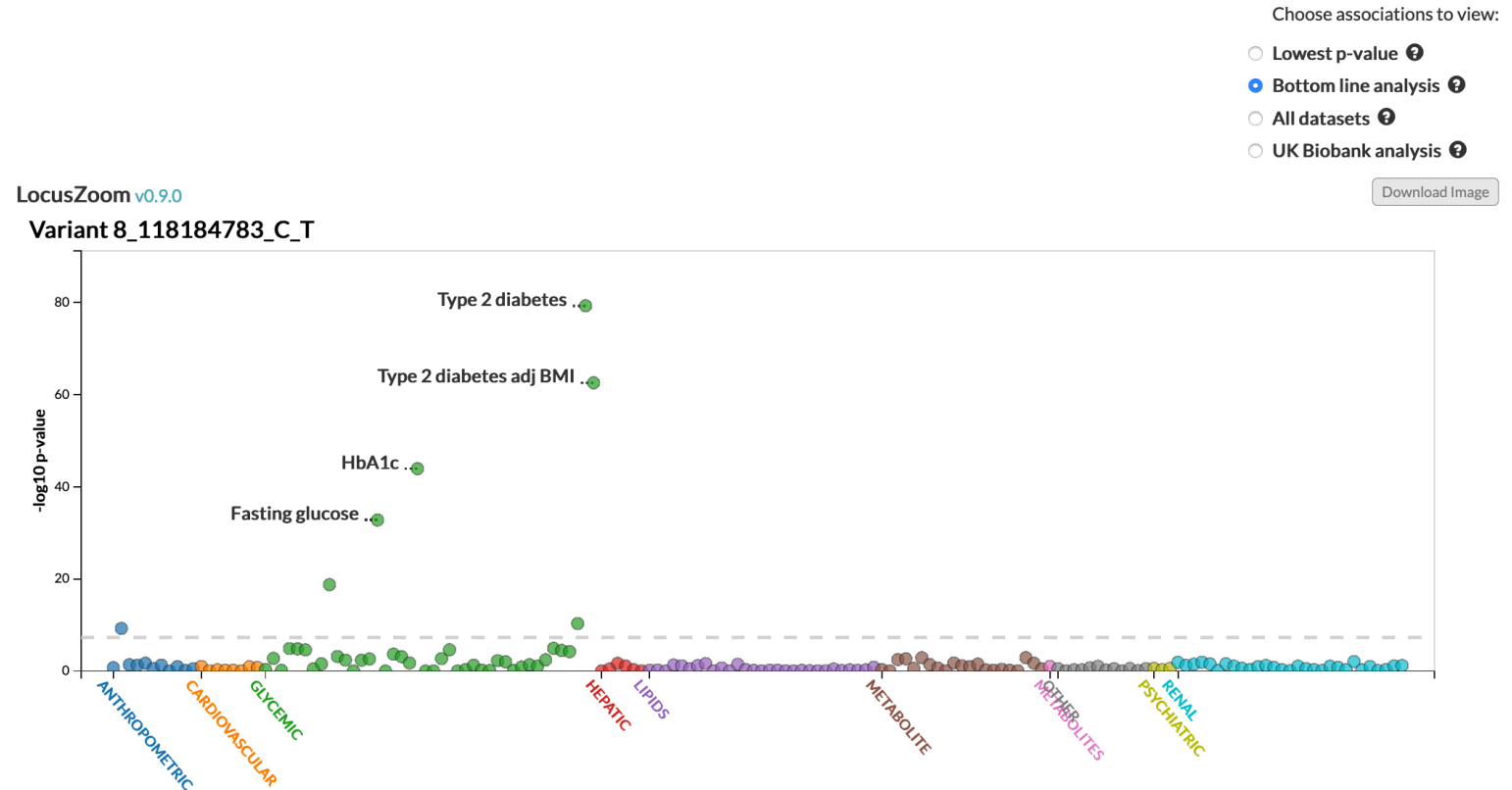
adipose tissue

Reads in pancreatic islets

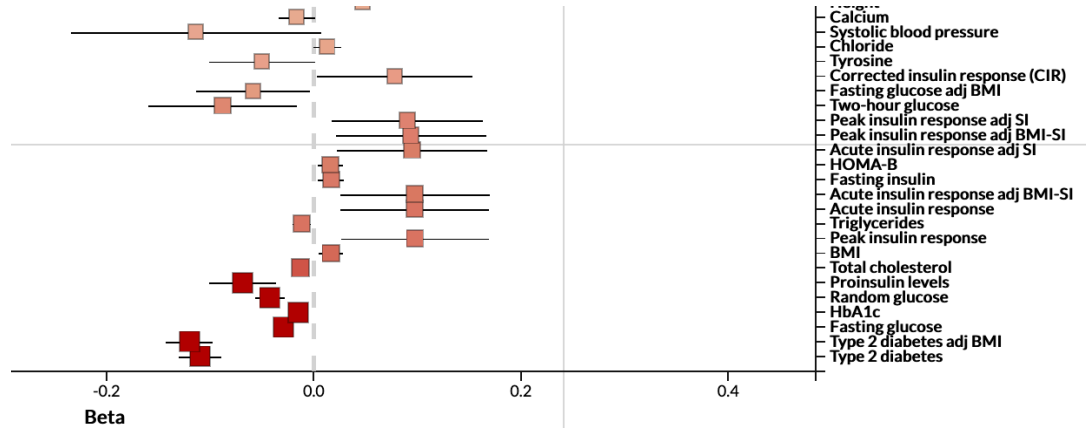


Showing many studies: PheWAS views

- The Portal aggregates data from many studies
- PheWAS plots enable wide comparisons
- Compare selected data, or see results across many UKBB phenotypes



PheWAS: Multiple Visualization options

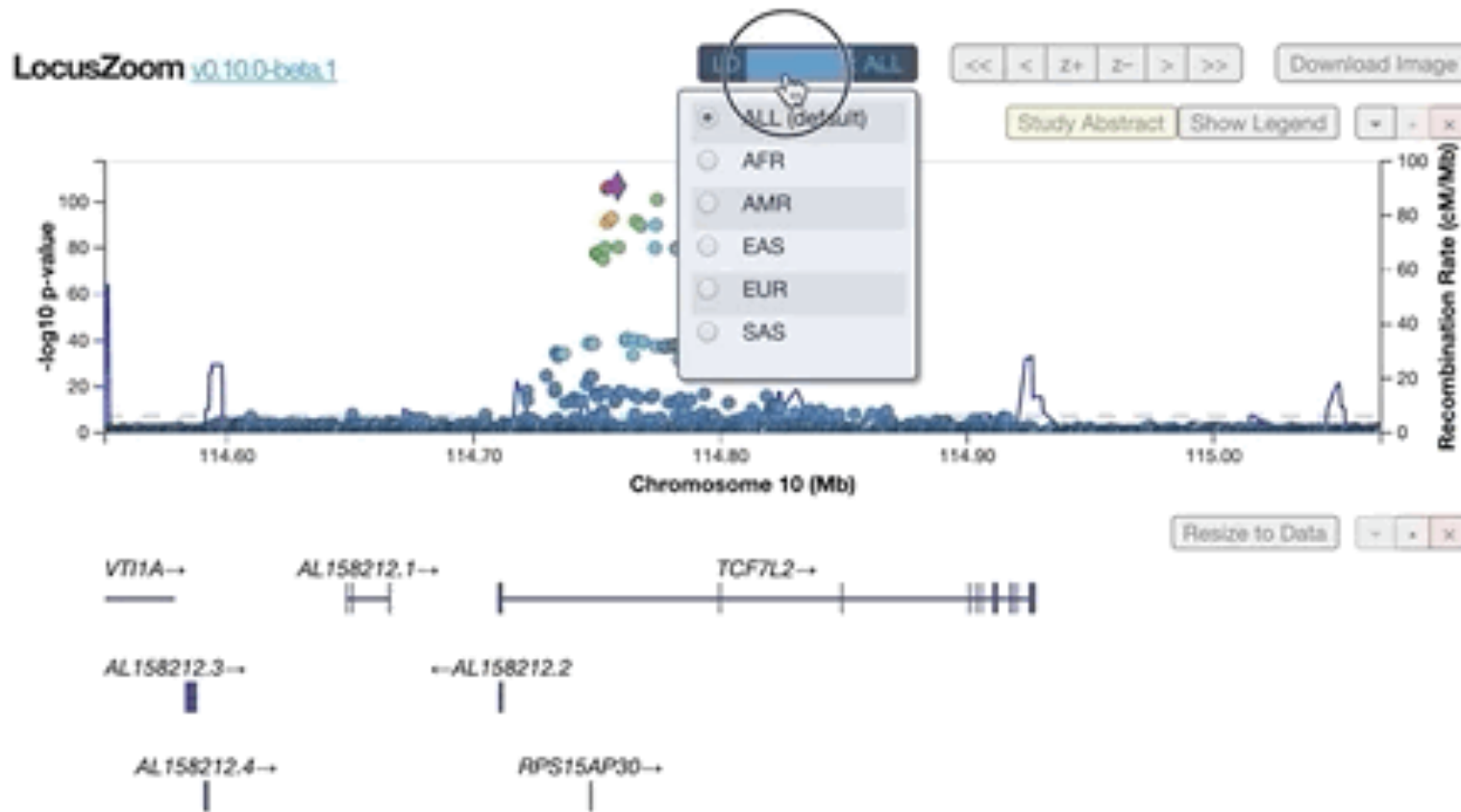


- PheWAS comparisons can be shown in multiple ways
- Forest plot view shows effect size, confidence intervals, and $-\log_{10} p$ across studies or phenotypes

Sneak peek at recent interactive capabilities and annotations

NEW FEATURES IN LOCUSZOOM.JS

Choose custom LD population

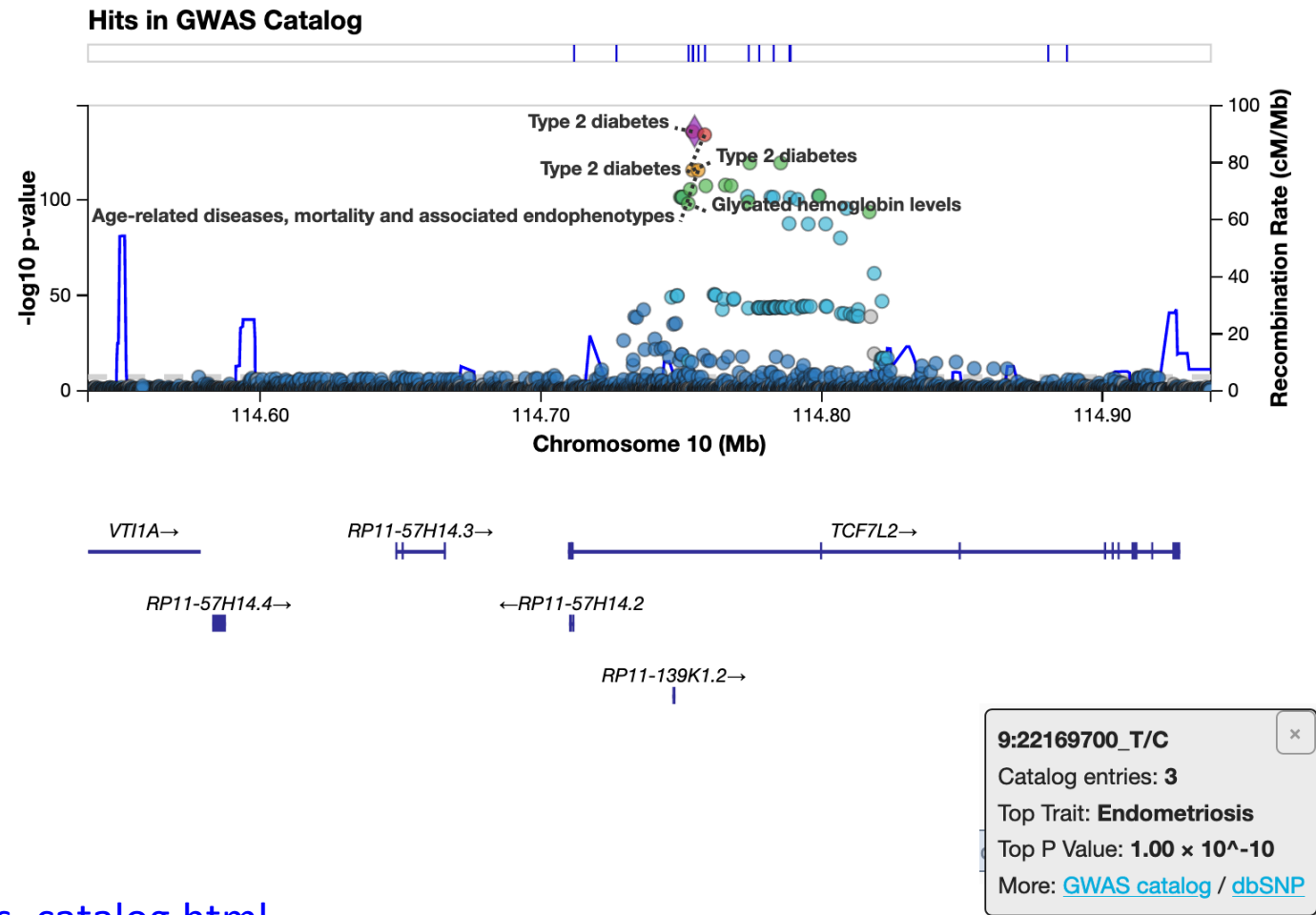


Preview: <https://statgen.github.io/locuszoom/>

Coming in LZ.js version 0.10. Powered by the new Michigan LDServer

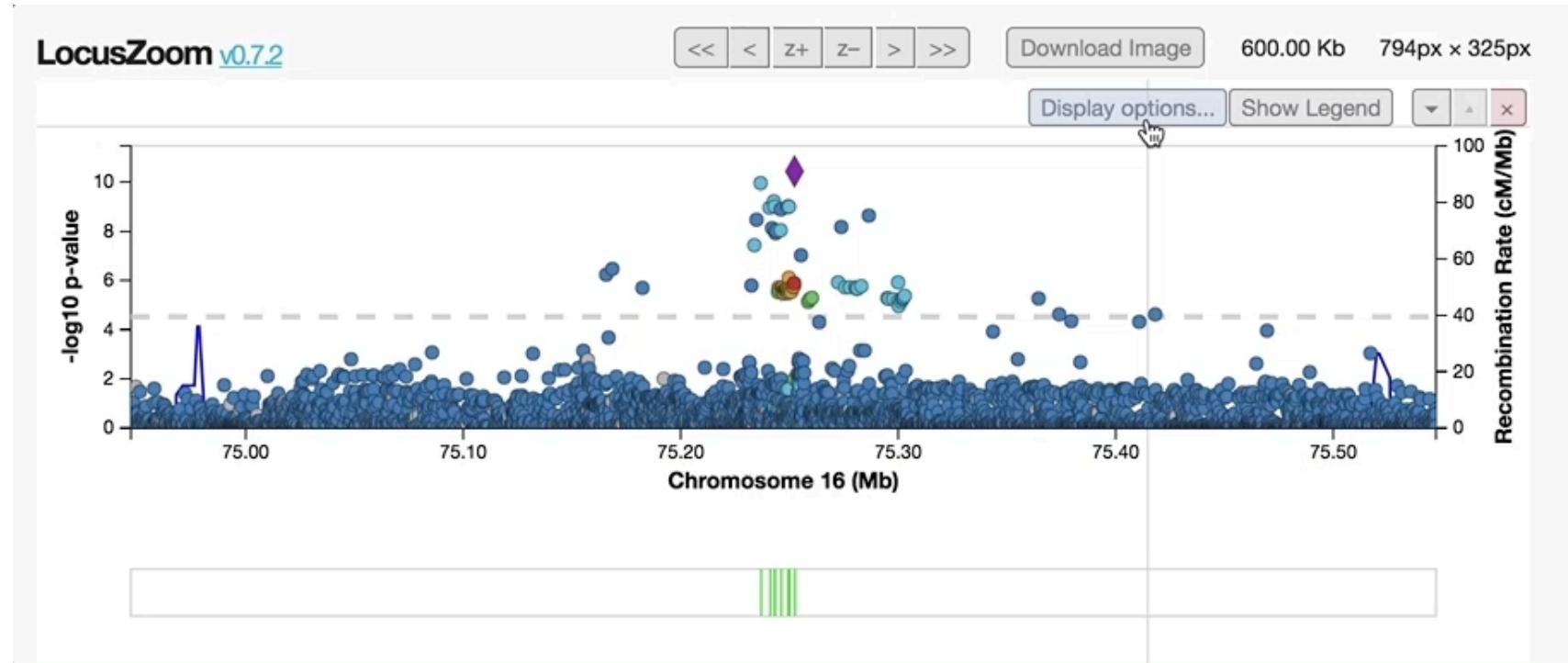
Draw comparisons to published studies

- Compare association results to known significant hits from EBI GWAS catalog
- Interactively show or hide labels
- Click a point in one panel to highlight matching elements in another panel



Enhance data with Credible sets

- Calculate 95% credible sets in the browser
- Dynamically recolor the plot to change display
- Modular method: Use just the calculations, or add visualization



Credible sets: on the Portal

Click on a variant ID to see more information, and to access a link for tissue selection. →

[9:22134094 T/C](#)

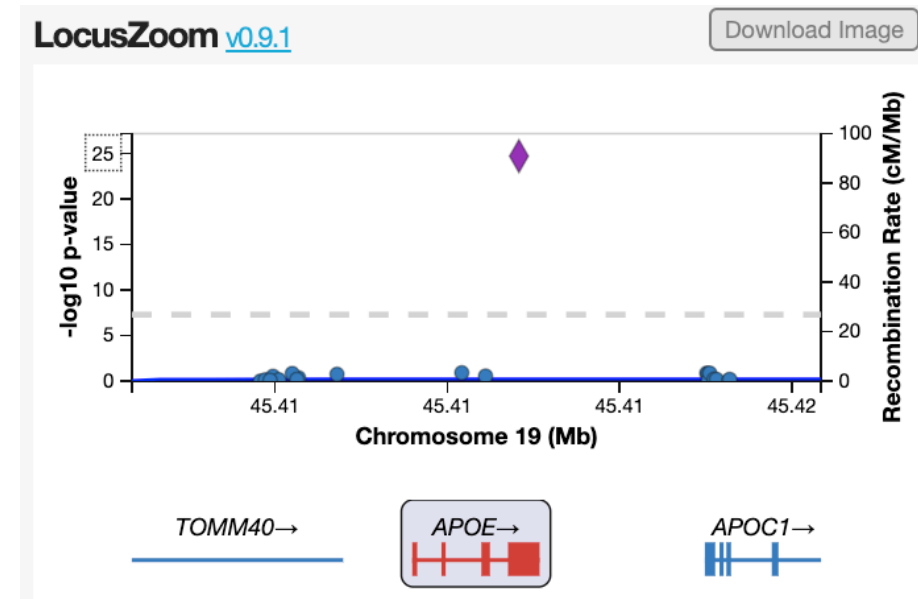
[9:22134068 G/A](#)

annotation	Coding		
	Splice site		
	UTR		
	Promoter		
association	Posterior probability	0.079	0.071
	P value	2e-42	3e-42
tissue	HMEC mammary epithelial prim...		

Genic enhancer
 Active enhancer 1
 Active enhancer 2
 Weak enhancer

Summarize aggregation test results

- LocusZoom can run (many) gene-based tests in a region and visually summarize results
- Comparing to single-variant test results enables interpretation
- Click or interact with the plot to update a table of results: drill down to understand findings



DOWNLOAD

Gene	Mask	# Variants	Test type	p-value
APOE	Exonic SNVs from GENCODE genes with allele frequency < 5%	3	skat	4.09×10^{-24}
APOE	Exonic SNVs from GENCODE genes with allele frequency < 5%	3	burden	1.44×10^{-19}



Aggregation tests: on the Portal

- Perform aggregation tests in a web browser, using only summary statistics (score statistics, genotype covariance matrices)
- Burden, SKAT, VT, and SKAT-O
- Soon: conditional analysis
- Future: meta-analysis of aggregation tests
- Explore pre-defined variant groups, or create your own via the portal

Manage variant selection

Choose a collection of variants for analysis. Choose a variant filter; set a MAF threshold if desired (this overrides the MAF thresholds in the variant filters) and apply the threshold across all samples or each ancestry. Remove variants from the list using the check boxes at the left of the table.

Available variant filter:
Protein-truncating only


Minor Allele Frequency:
MAF <

Apply MAF across:
 All samples Each ancestry

Use?	Variant ID	dbSNP ID	Chrom.	Position	MAC	Polyphen	SIFT	Protein change	Consequence
<input checked="" type="checkbox"/>	8_118147614_GA_G		8	118147614	1			p.M17X	frameshift
<input checked="" type="checkbox"/>	8_118147639_TA	rs573681084	8	118147639	2				splice donor
<input checked="" type="checkbox"/>	8_118159297_C_CCA		8	118159297	4			p.P59PX	frameshift
<input checked="" type="checkbox"/>	8_118159393_G_A		8	118159393	2				splice donor

Connect to other representations

- LocusZoom.js can talk to other visualizations on the same web page, enabling rich comparisons that update as the view changes
- Interact with LocusZoom, drag plot region, and instantly see results update in a table
- Additional results can be computed in browser, then saved as a file for further analysis
- Plots can be saved and shared as high-resolution SVG images

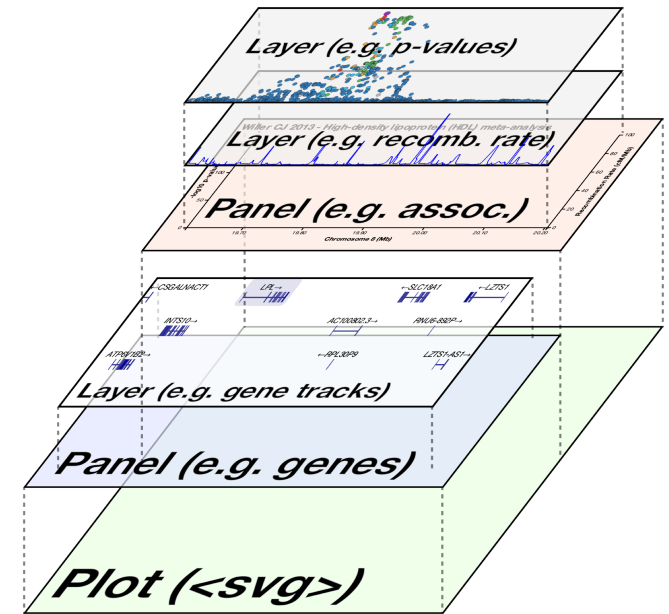
Select a study: GLGC 2013- TC (joint analysis) 

Download

▲	Alt ▲	-log ₁₀ (p) ▲	β ▲	SE(β) ▲	Alt freq. ▲	Cred. set ▼	Posterior probability ▲
	C	10.529	0.026	0.004	0.529	✓	0.512
	G	10.508	0.025	0.004	0.541	✓	0.488
	C	6.860	0.031	0.006	0.540	✗	1.37 × 10 ⁻⁴

LocusZoom.js: Modular and reusable

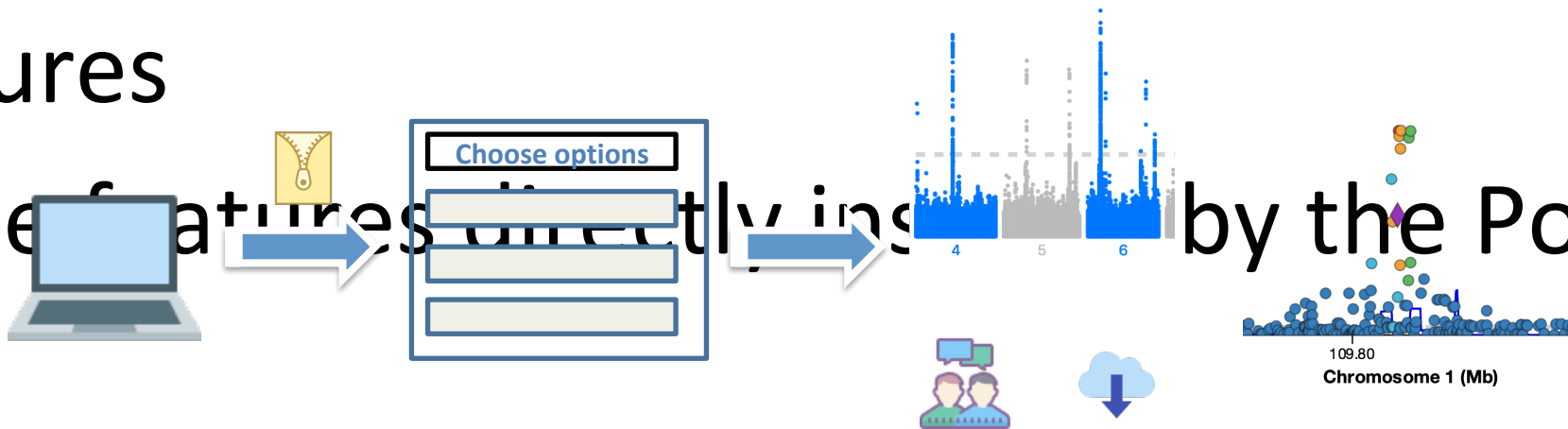
- LocusZoom.js can be embedded in your web page, and can draw data from a wide range of sources
- It is based on standard web technologies that work with many sites
- No special tooling is required, but it will work with advanced tools if you use them
- Our standard REST APIs provide common data (like genes) for builds GRCh37 and GRCh38



Making it easier to get started

- We are building new ways to use LocusZoom.js with your own data
- Serves as a testbed for new LocusZoom features

- Some features directly inspired by the Portal



LocalZoom: Quick Region Plots, without uploading

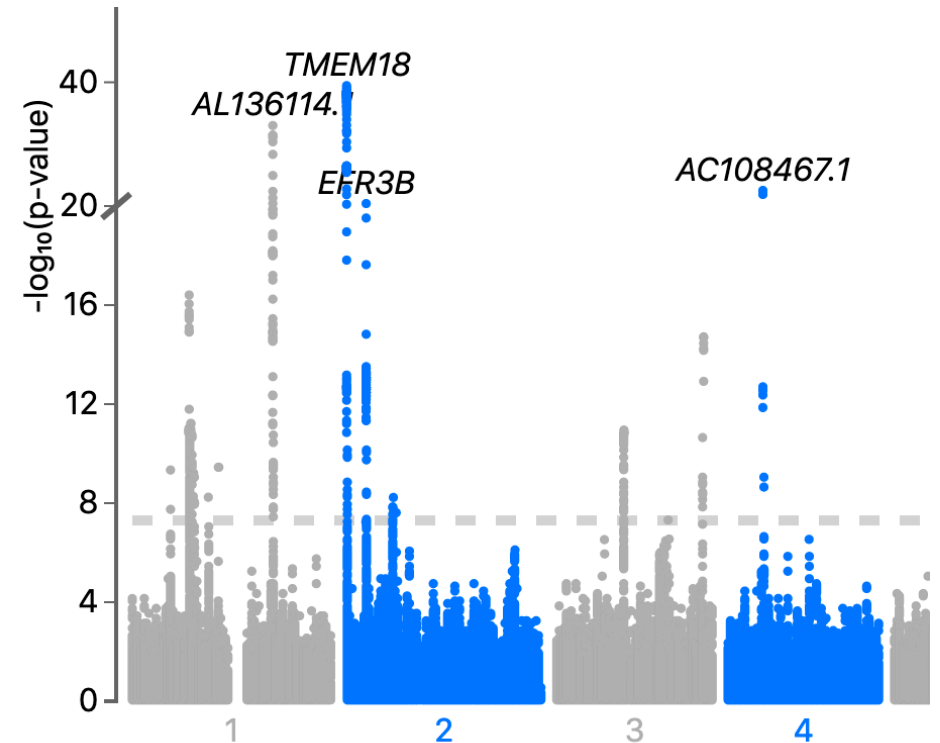
- Make a standard LocusZoom plot from tabixed GWAS summary statistics, in a web browser
- Full support for build **GRCh37** and **GRCh38**
- Annotate with hits found in the GWAS catalog
- Perform region-based credible set calculations, in your web browser
- Explore hits for any SNP in a PheWAS of the UK BioBank

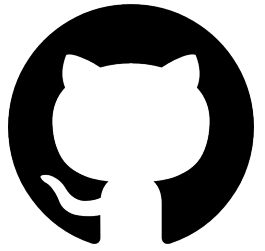


Upload, analyze, and share:

my.locuszoom.org

- Enhanced view with additional information: top loci, Manhattan plot
- Supports a variety of GWAS file formats (no need to run tabix first)
- New website allows you to make your data public, or preview an analysis privately and share with trusted collaborators
- Explore publicly shared results from other studies





Free, reusable tools

- Code, instructions, and examples are available for download
- LocusZoom.js: <https://github.com/statgen/locuszoom>
- Credible sets: <https://github.com/statgen/gwas-credible-sets>
- Aggregation tests: <https://github.com/statgen/raremetal.js>
- Michigan LDServer: <https://github.com/statgen/LDServer>
- PheWeb: <https://github.com/statgen/pheweb>

Summary

- LocusZoom provides deep context for a large number of datasets available on the Portal: single studies (region plots) and aggregate presentation (PheWAS plots)
- Interactive features help users to draw connections and explore novel hypotheses
- Available on the T2D Portal, but fully reusable on other sites to view your own data
- More annotation features are coming soon!