


TYPE 2 DIABETES
KNOWLEDGE PORTAL




# T2D Knowledge Portal Platform Webinar & Workshop

*November 14, 2019*

# *TODAY: LocusZoom & the T2DKP*

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

SCHOOL OF PUBLIC HEALTH
UNIVERSITY OF MICHIGAN

TYPE 2 DIABETES KNOWLEDGE PORTAL

# *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

TYPE 2 DIABETES KNOWLEDGE PORTAL

# *What will you learn at future webinars?*

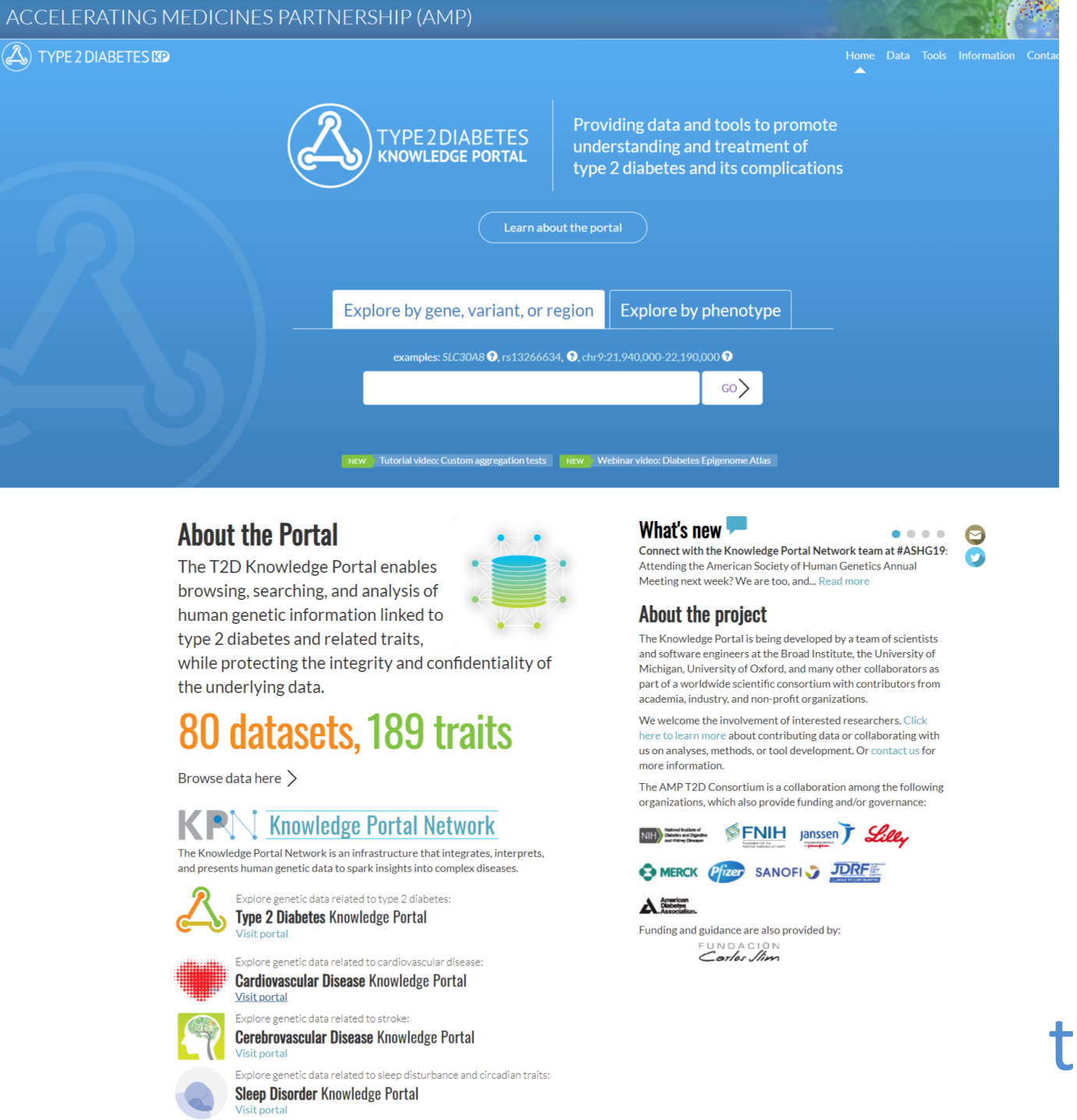
- How to access the KPN and T2DKP results programmatically- Smart APIs
- What aggregation tests can you run in real time on the site?
- How are we expanding to metabolic disorders?
- What new workflows are available for gene/variant prioritization?
- What methods are run on the GWAS datasets and how 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?

TYPE 2 DIABETES KNOWLEDGE PORTAL

# *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

KPN TYPE 2 DIABETES KNOWLEDGE PORTAL




# The AMP-T2D Knowledge Portal

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

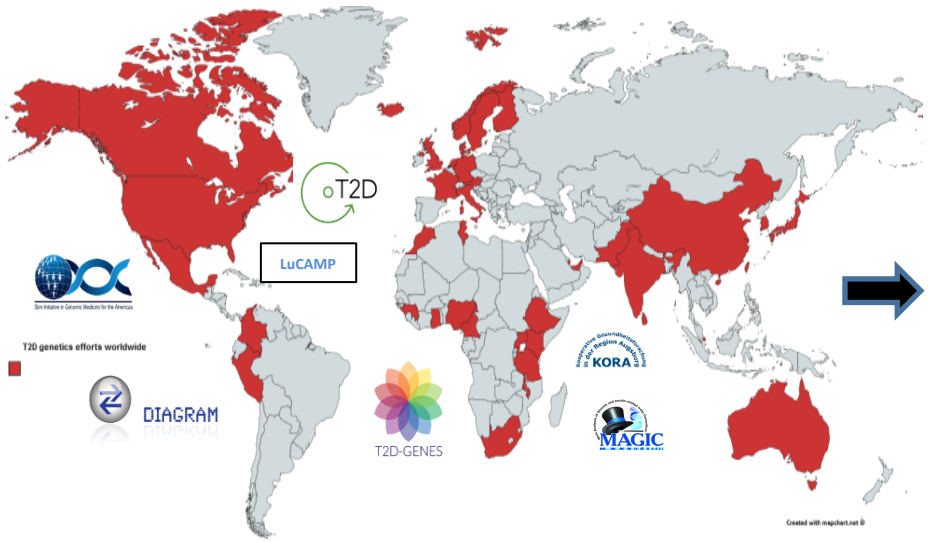
type2diabetesgenetics.org

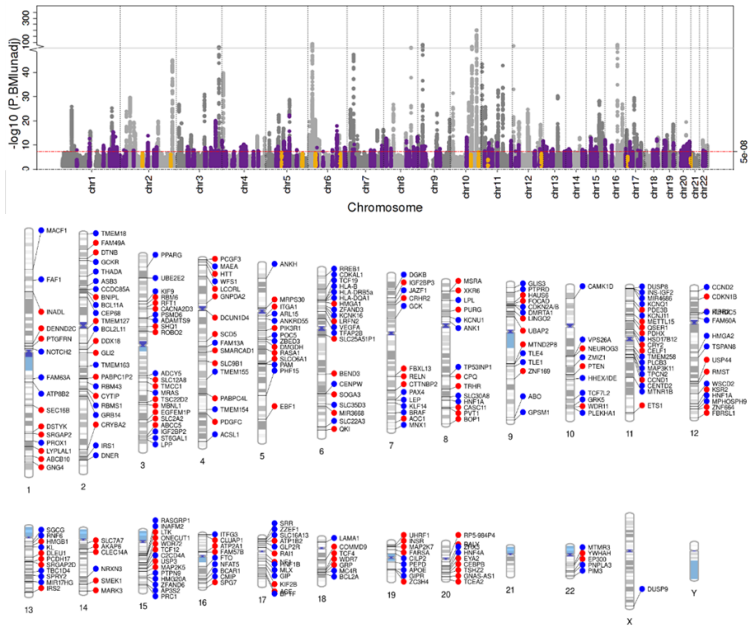
*A simple idea:*

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

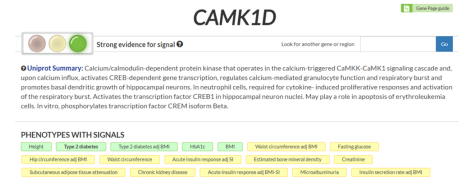
TYPE 2 DIABETES KNOWLEDGE PORTAL

# *Addressed the first opportunity & challenge*

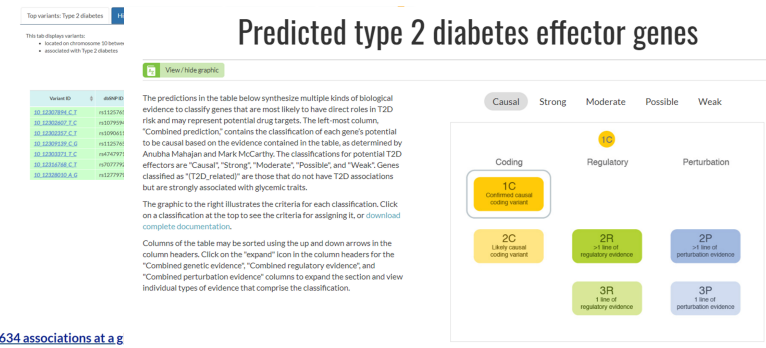


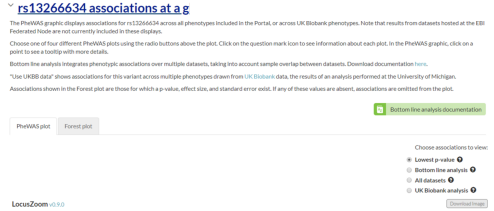



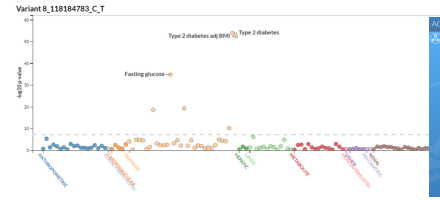

Mahajan *et al*. *Nat Gen*. 2018

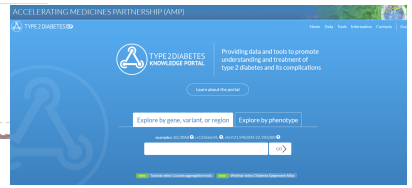



80 datasets, 189 traits



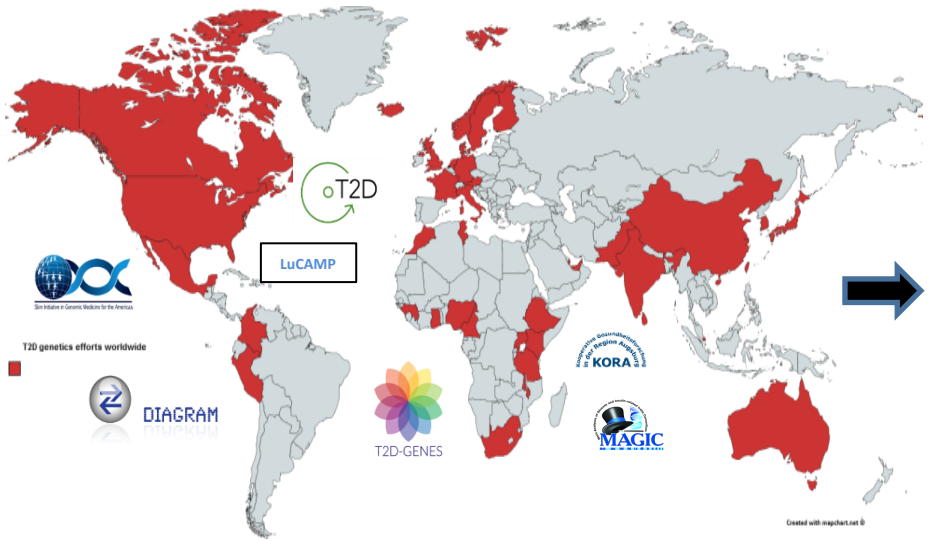


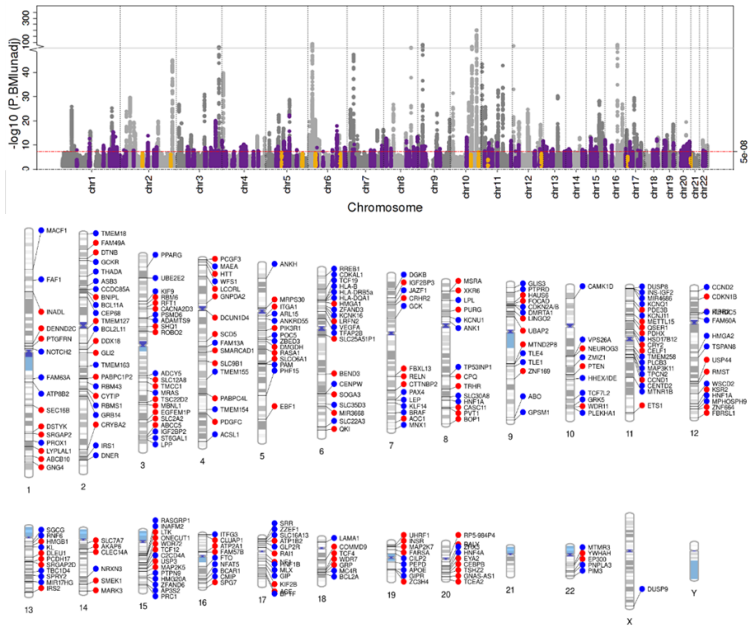

Motivated T2D Genetics Research community

Identification of over 403 loci

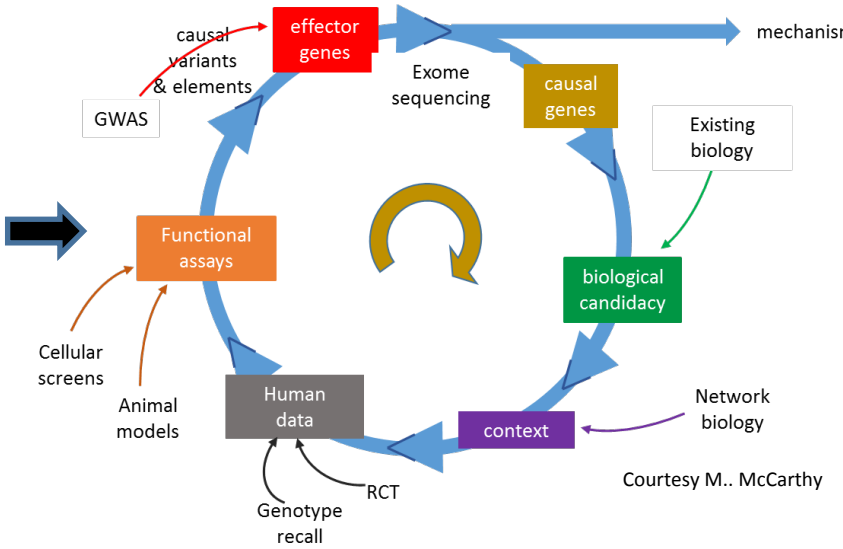
TYPE 2 DIABETES KNOWLEDGE PORTAL

# *A shift in the coming years…a new opportunity & challenge*




Mahajan *et al*. *Nat Gen*. 2018




Courtesy M.. McCarthy

Motivated T2D Genetics Research community

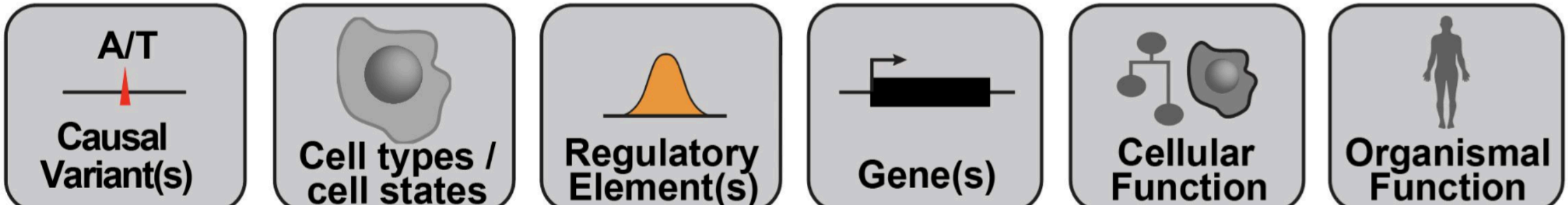
Identification of over 403 loci

Mechanism, magnitude, markers for T2D genes

TYPE 2 DIABETES KNOWLEDGE PORTAL

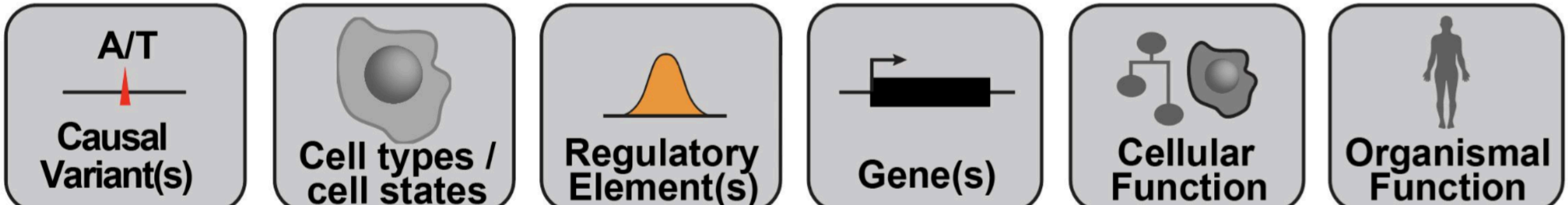
# *Shift in focus over the coming years- function*

- 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?

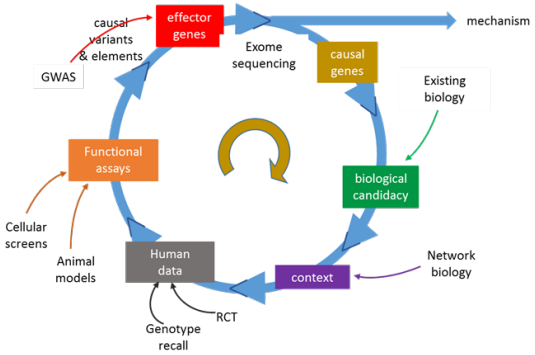


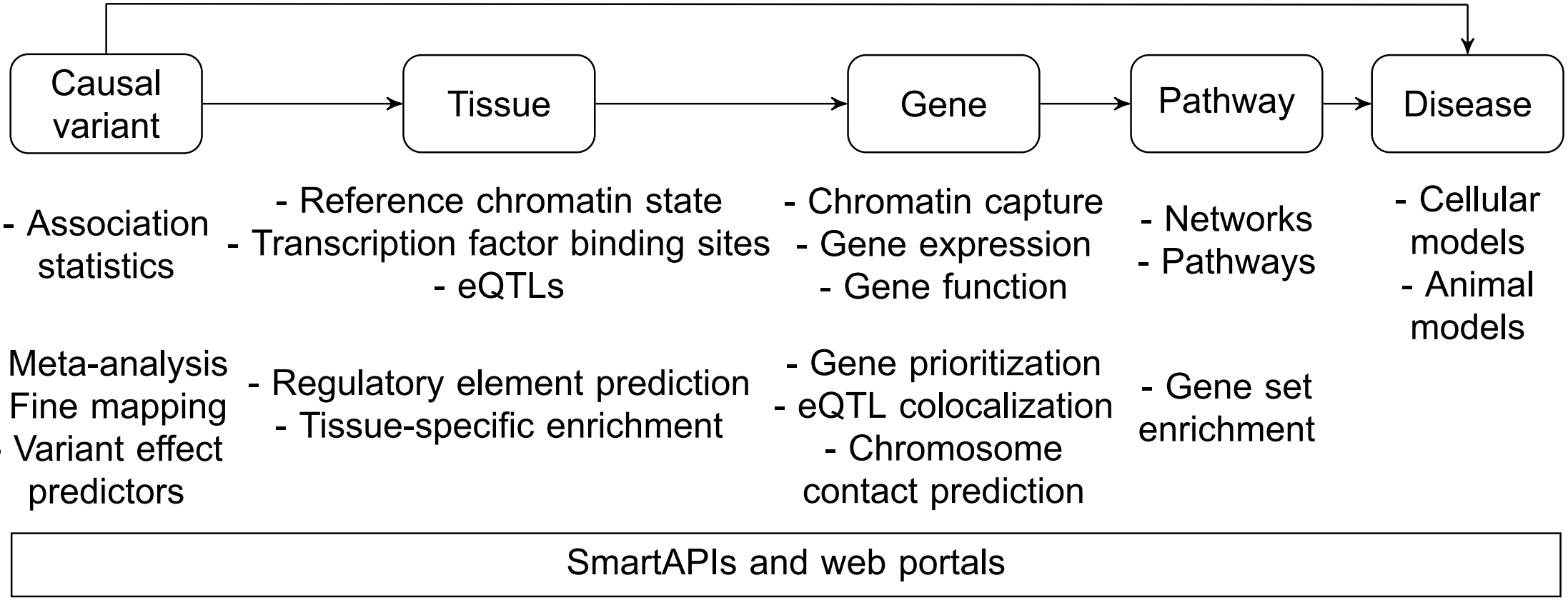
# *New set of gaps to address with this opportunity*

- 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*




*J. Flannick*

# *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!


TYPE 2 DIABETES KNOWLEDGE PORTAL

# *Finding LocusZoom in the Knowledge Portals*


type2diabetesgenetics.org

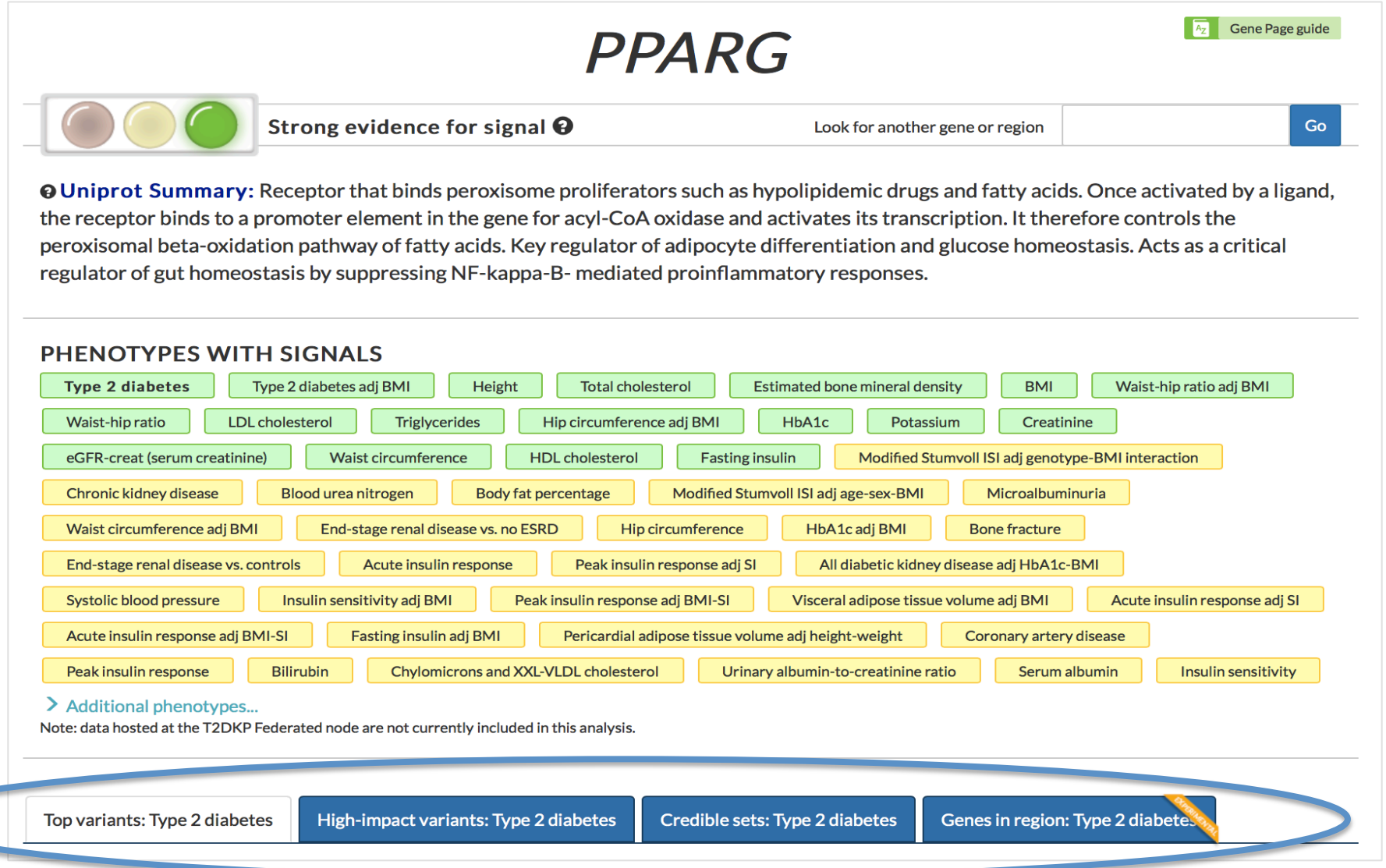

broadcvdi.org


KP4CD.org


cerebrovascularportal.org


sleepdisordergenetics.org

# Gene/Region Page



# *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 (EXPERIMENTAL)

This tab displays variants:

- located on chromosome 3 between 12228817 and 12575905
- associated wit Add / Subtract Data

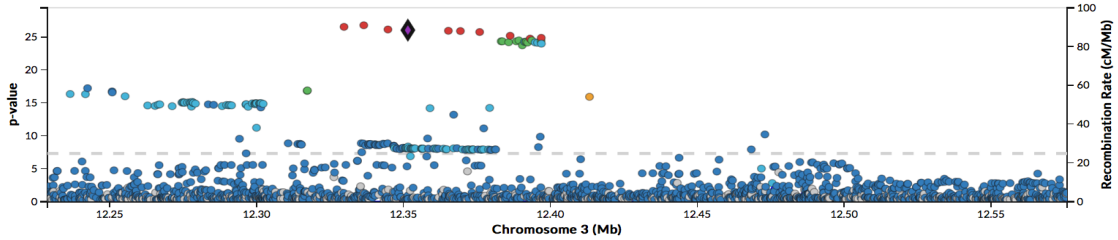
| Variant ID | dbSNP ID | Major allele | Minor allele | Predicted impact | p-Value | Effect | MAF | Data set All |
|---|---|---|---|---|---|---|---|---|
| 3_12336507_G_A | rs11709077 | G | A | intron | 1.60e-27 | 0.896 | 0.12 | DIAMANTE (European) T2D GWAS |
| 3_12329783_C_T | rs17036160 | C | T | intron | 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

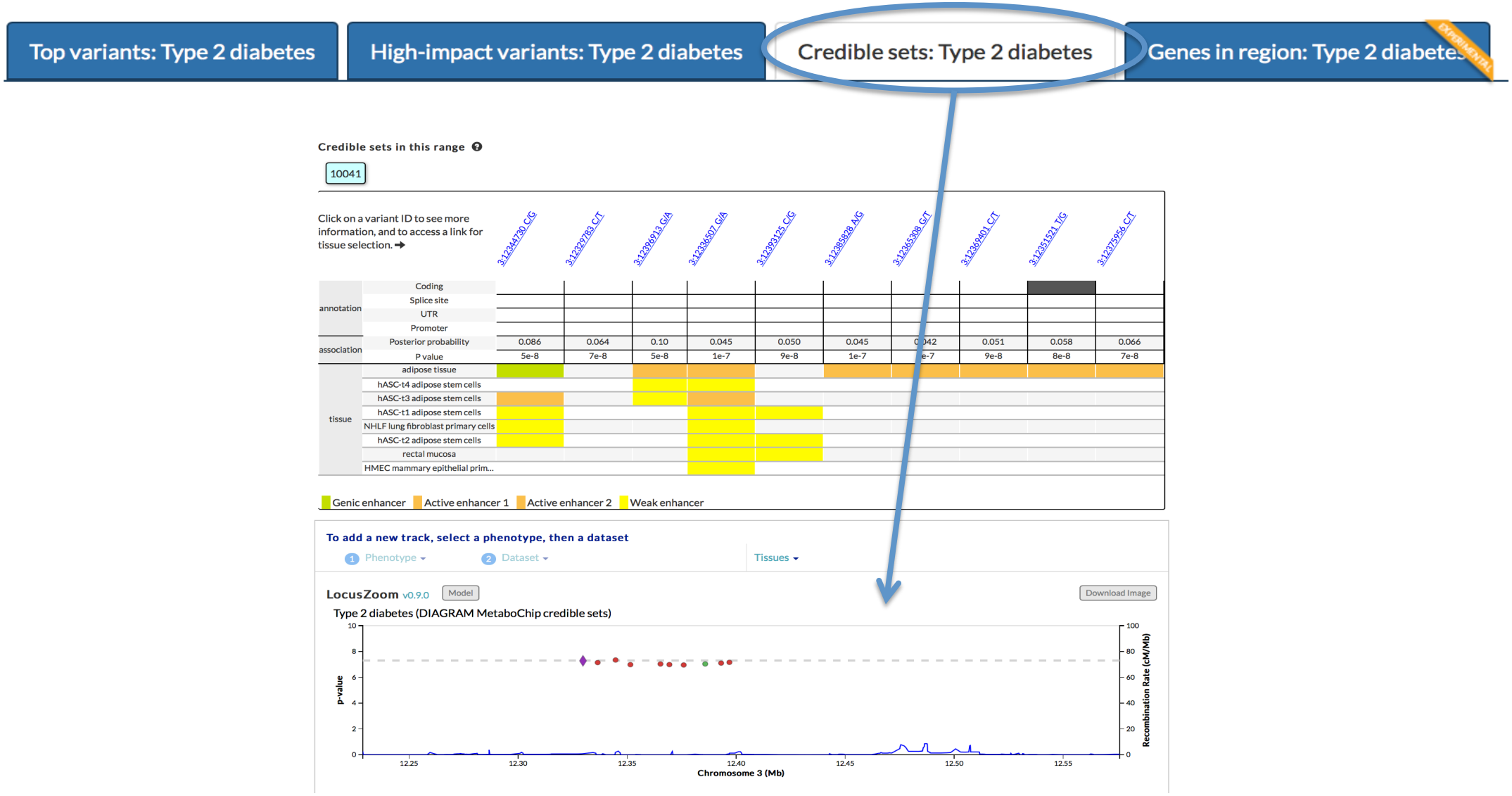
Type 2 diabetes (DIAMANTE (European) T2D GWAS)




adipose tissue

pancreatic islets

# *Gene/Region Page: Credible sets tab*



# *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

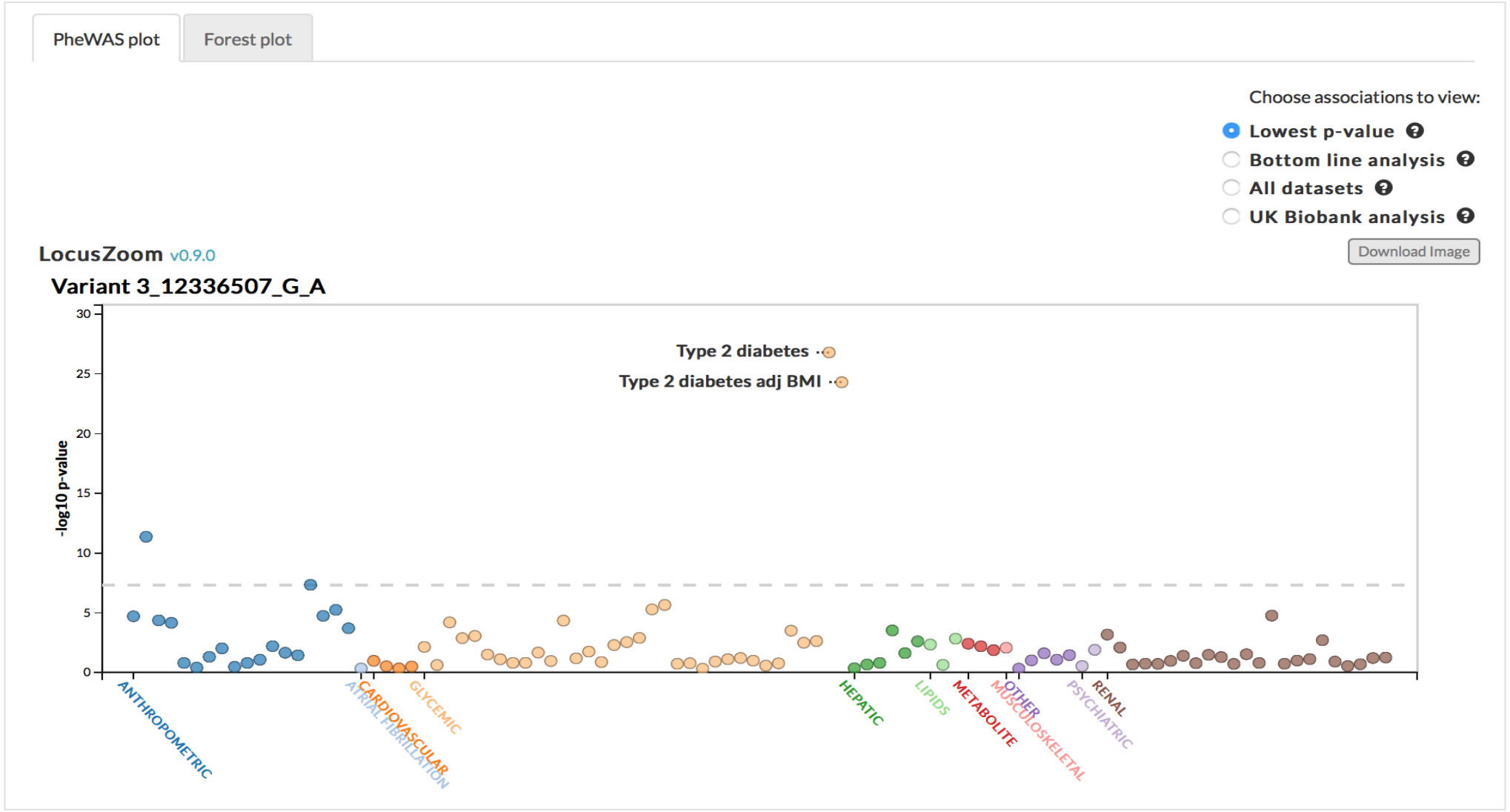
LocusZoom: view and analyze variants around rs11709077

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.

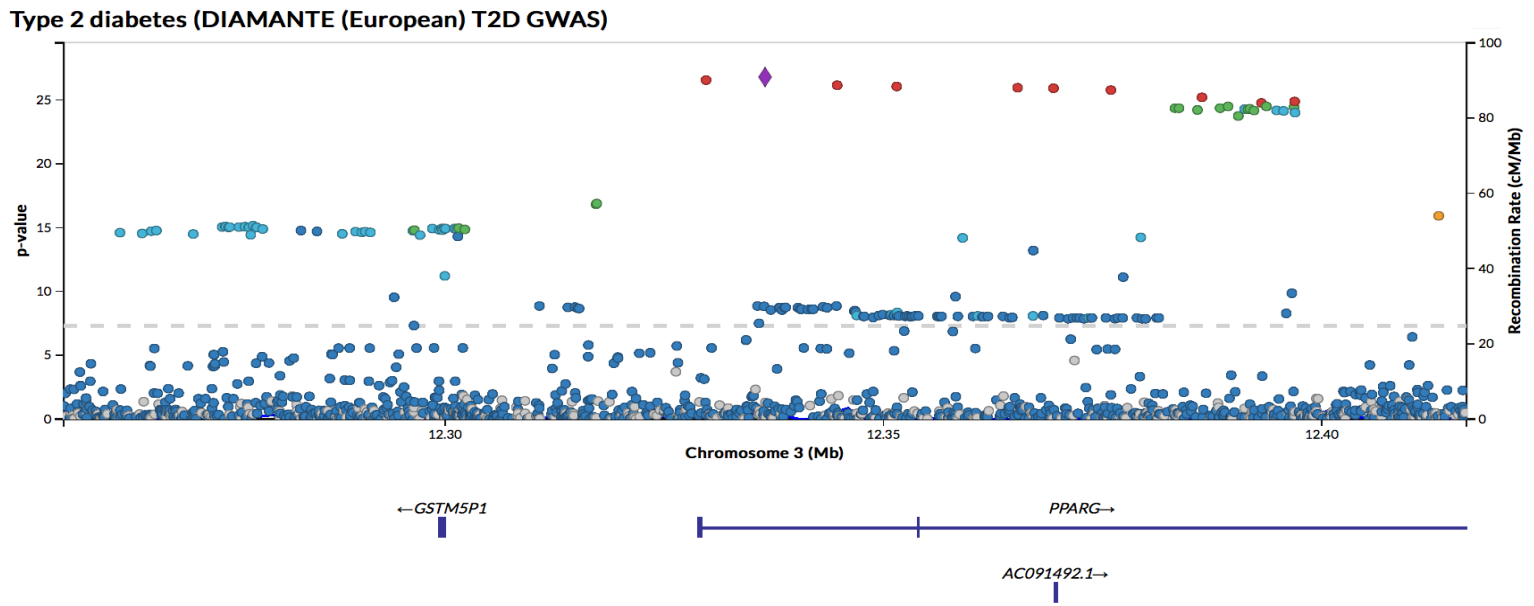
Linkage disequilibrium ($r^2$) values are based on the 1000G ALL reference panel (including every sample available in the 1000 Genomes project) and are supplied by the Michigan Imputation Server.

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



# Overview of LocusZoom.js in the T2D Knowledge Portal

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

TYPE 2 DIABETES KNOWLEDGE PORTAL

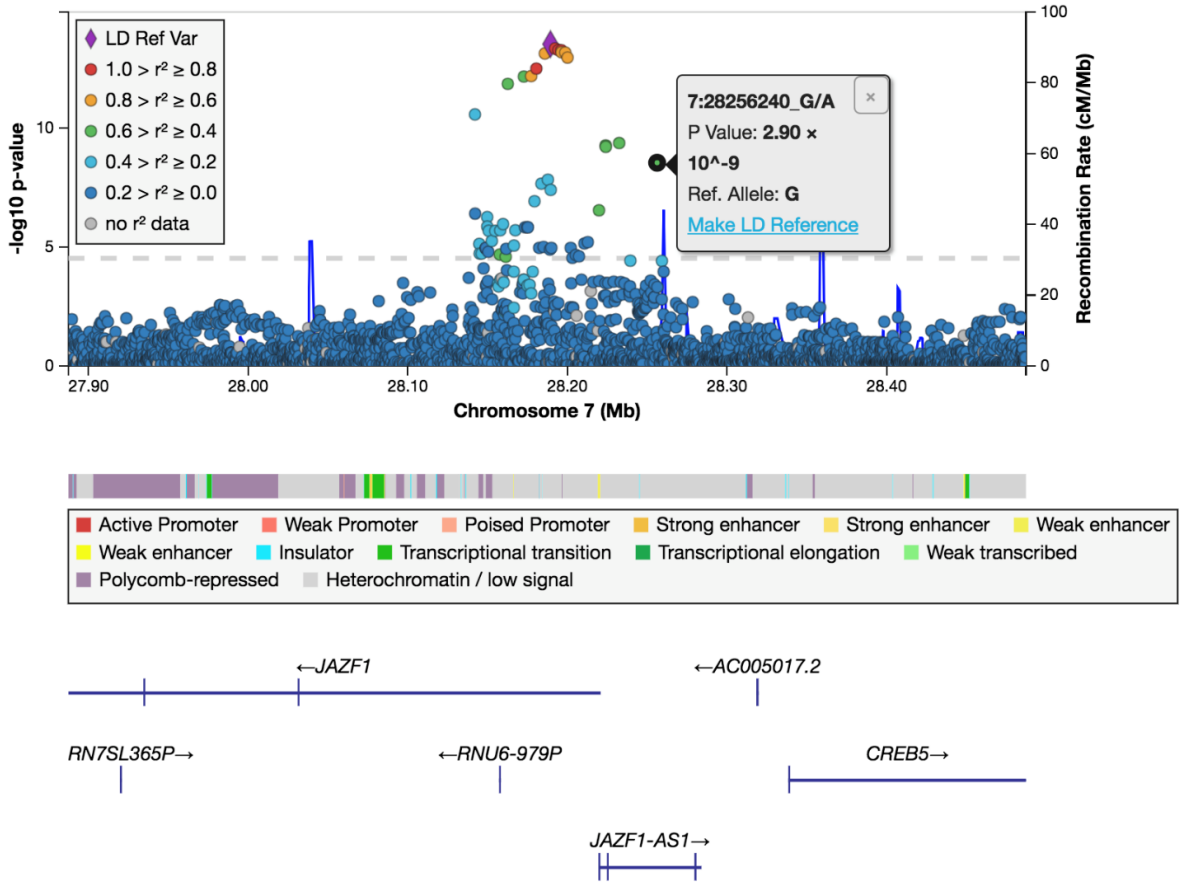
SCHOOL OF PUBLIC HEALTH
BIOSTATISTICS
UNIVERSITY OF MICHIGAN

# 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

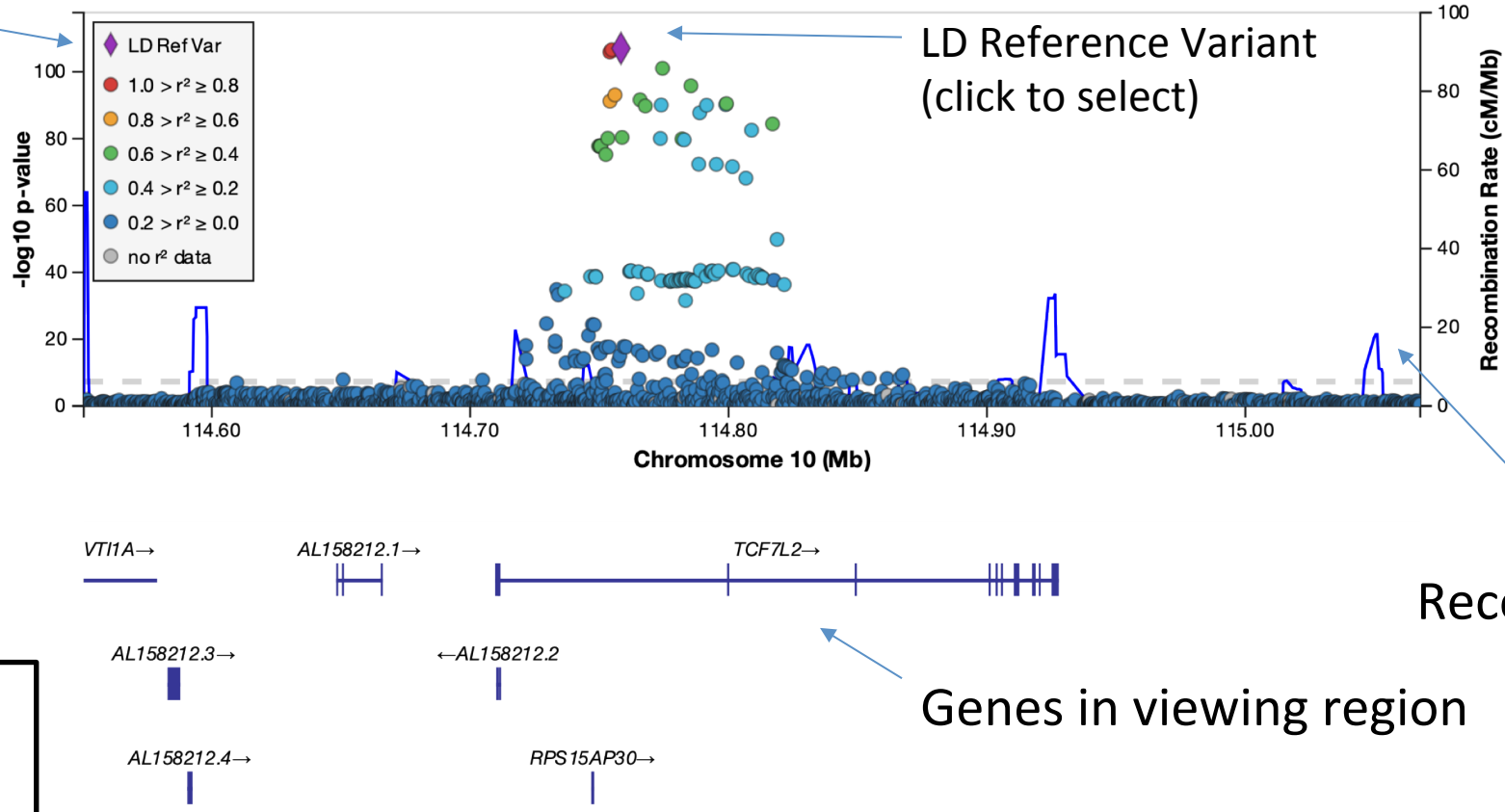



https://doi.org/10.1093/bioinformatics/btq419

https://github.com/statgen/locuszoom

# How to read a LocusZoom Plot

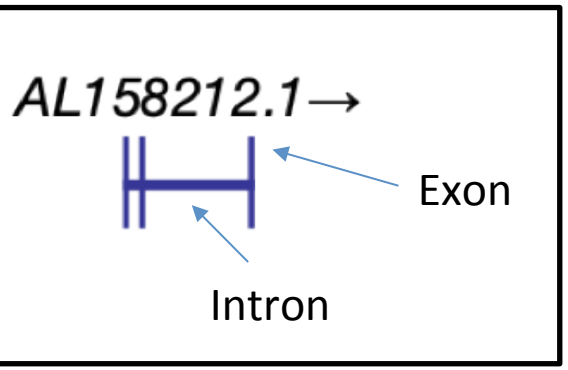
Color by Linkage Disequilibrium (LD)



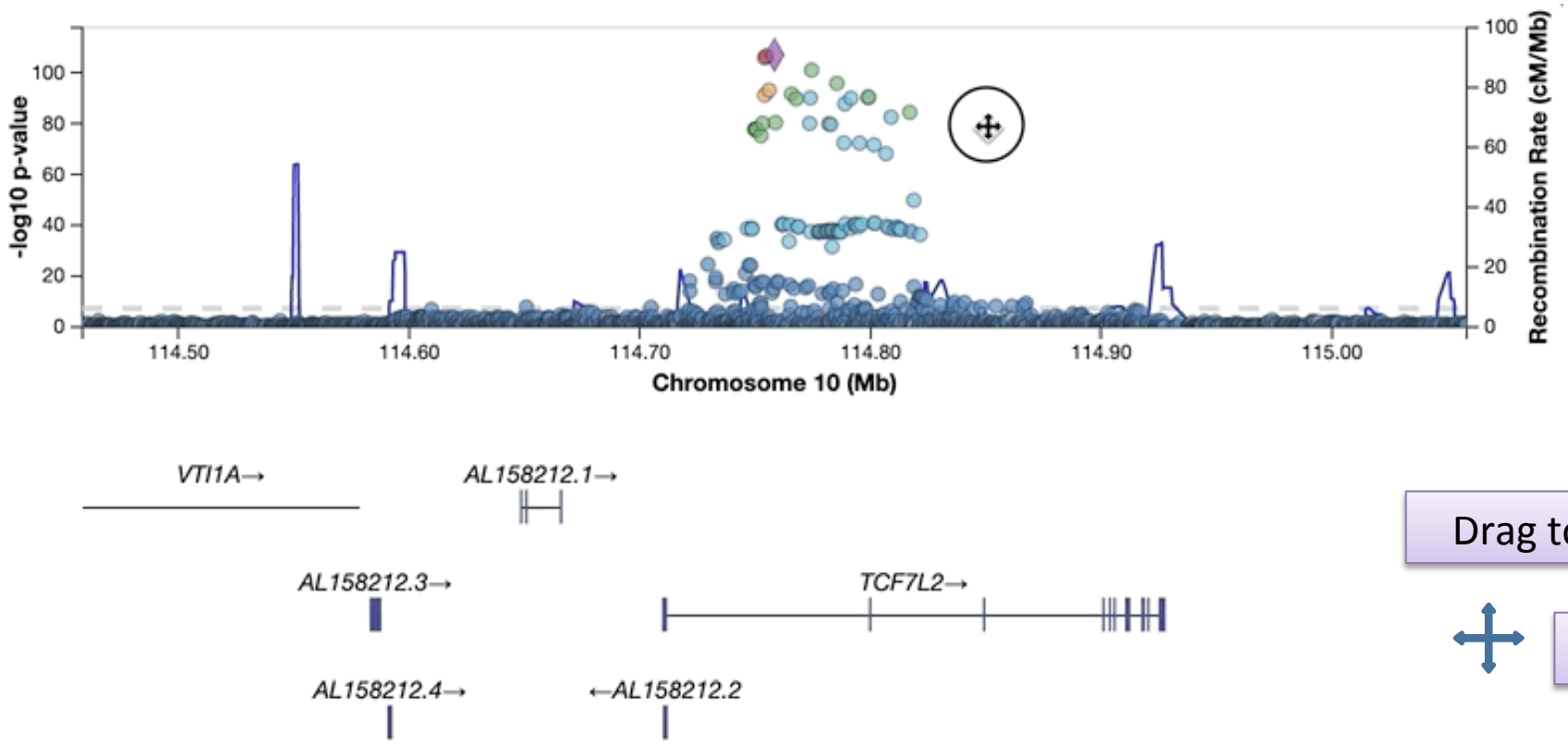

LD Reference Variant (click to select)

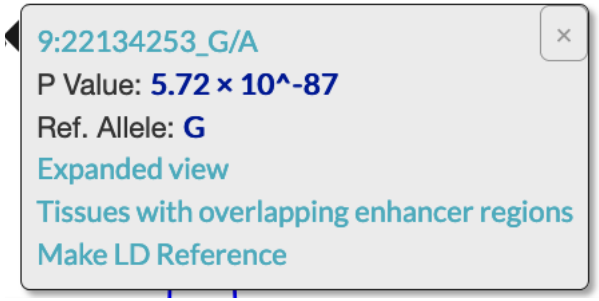
Recombination rate

Genes in viewing region



# Interactively control the view



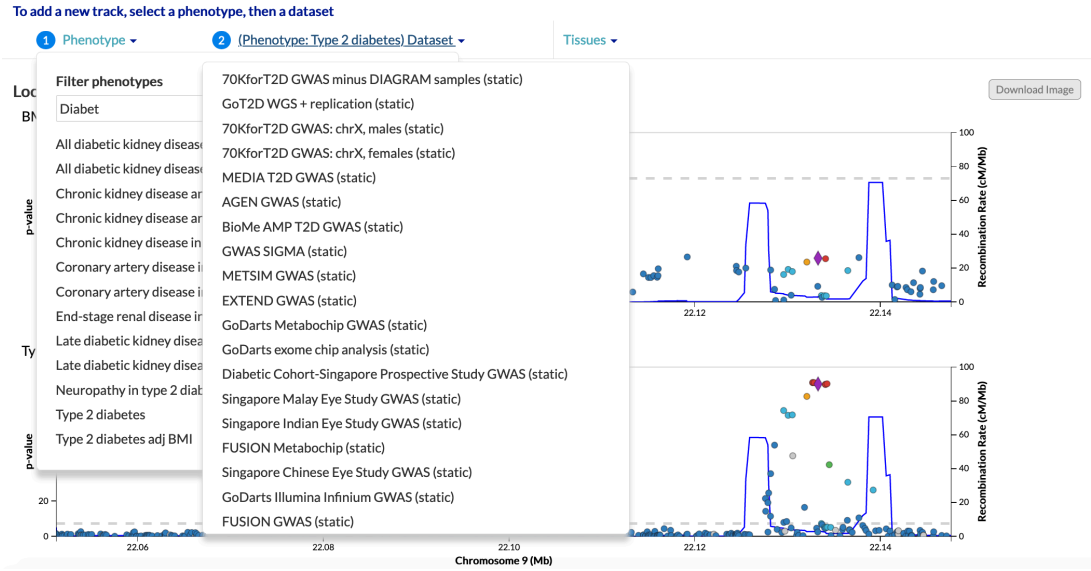




Drag to pan

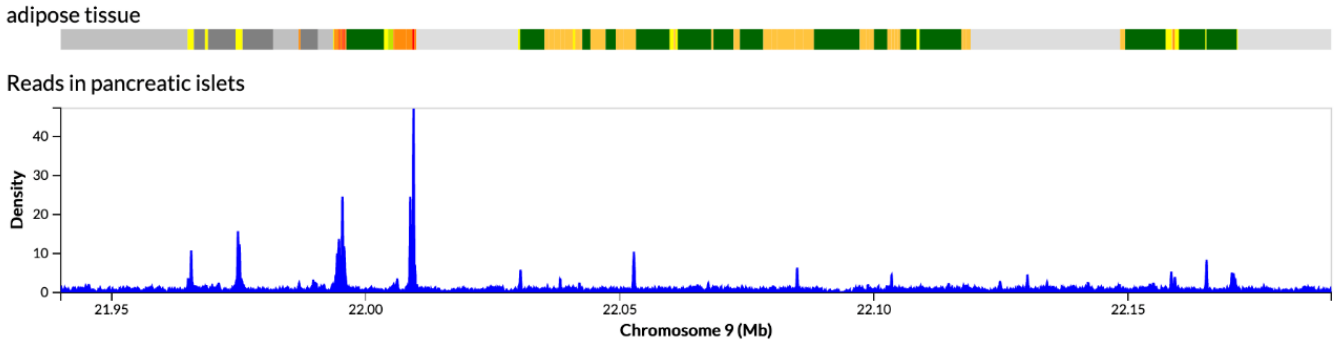
Scroll to zoom

Click to interact

# Explore a wide range of datasets in depth

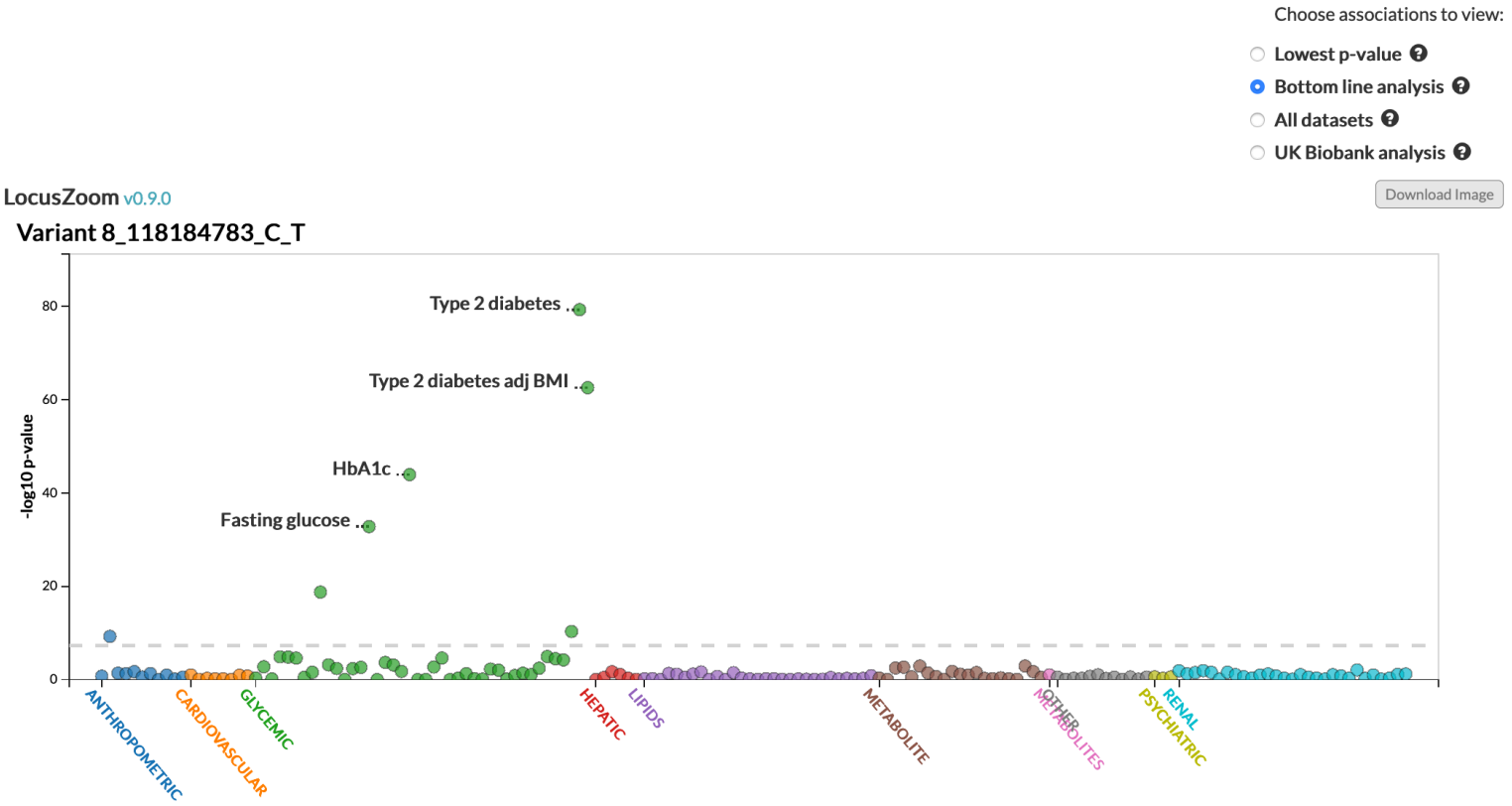



- Choose from GWAS summary statistics for a wide range of phenotypes and study datasets
- Stack panels vertically and make custom comparisons
- Add tissue-specific annotations and ATAC-seq data within the plot



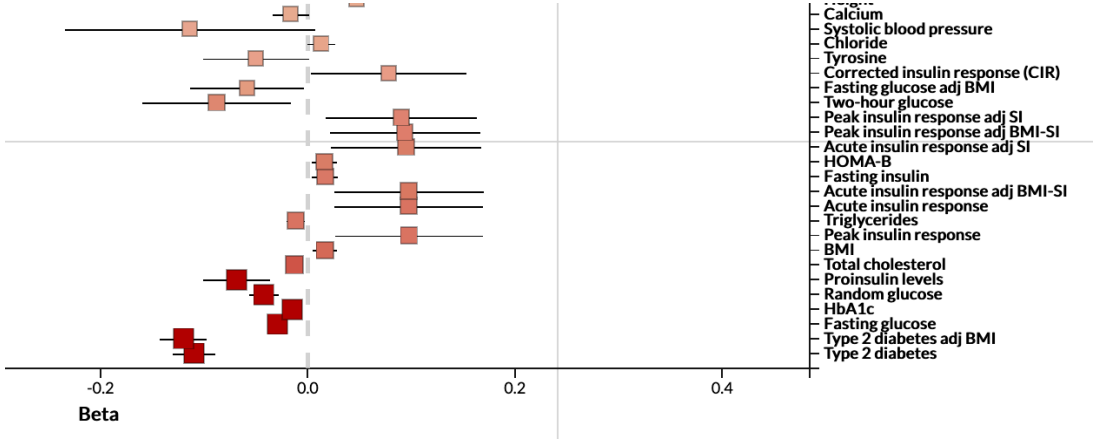
# Showing many studies: PheWAS views

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




Bottom line analysis method: https://genome.sph.umich.edu/wiki/METAL_Documentation#Sample_Overlap_Correction

# PheWAS: Multiple Visualization options



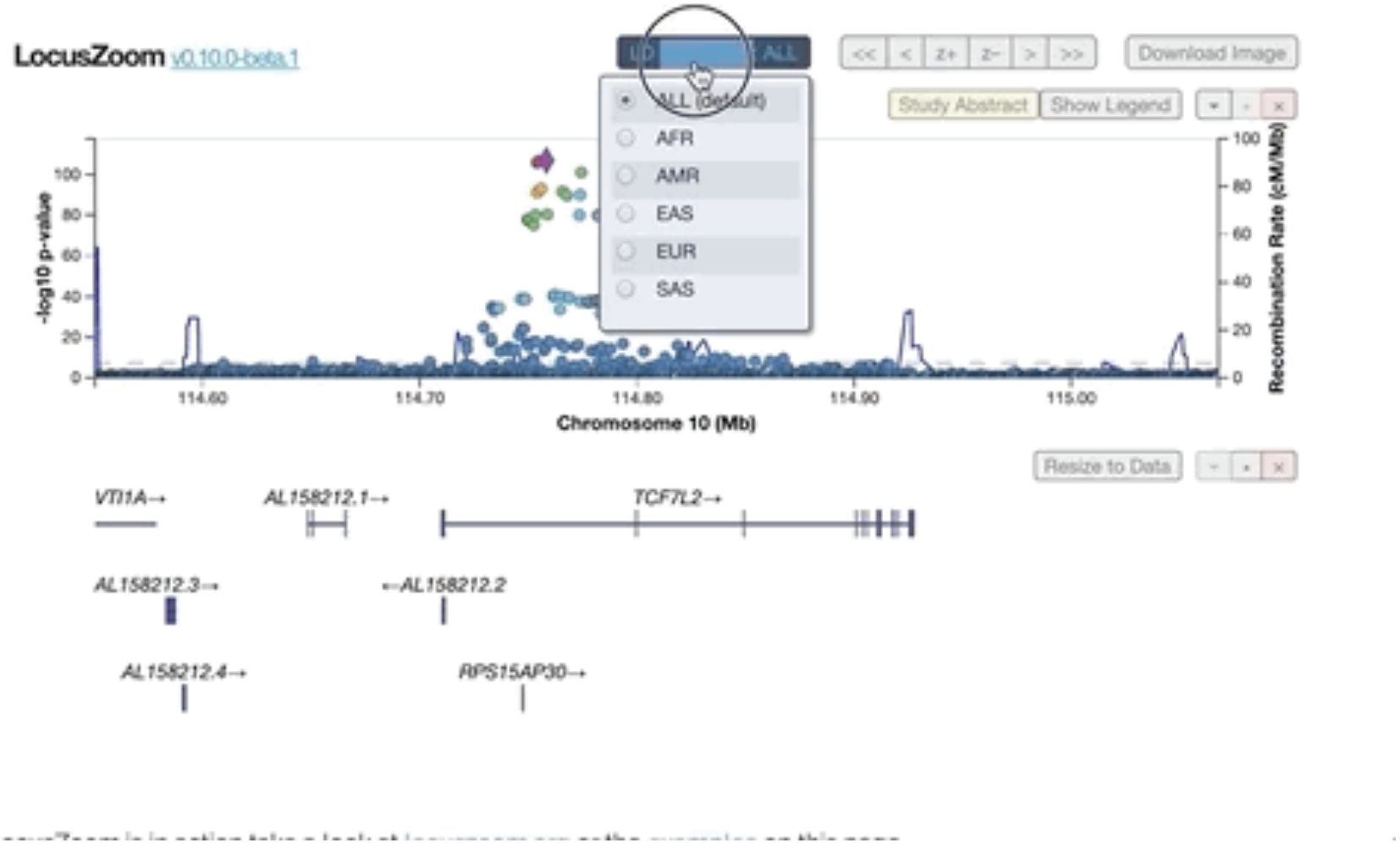

- PheWAS comparisons can be shown in multiple ways
- Forest plot view shows effect size, confidence intervals, and $-\log_{10} p$ across studies or phenotypes

http://www.type2diabetesgenetics.org/variantInfo/variantInfo/rs13266634

Sneak peek at recent interactive capabilities and annotations

# NEW FEATURES IN LOCUSZOOM.JS

# Choose custom LD population

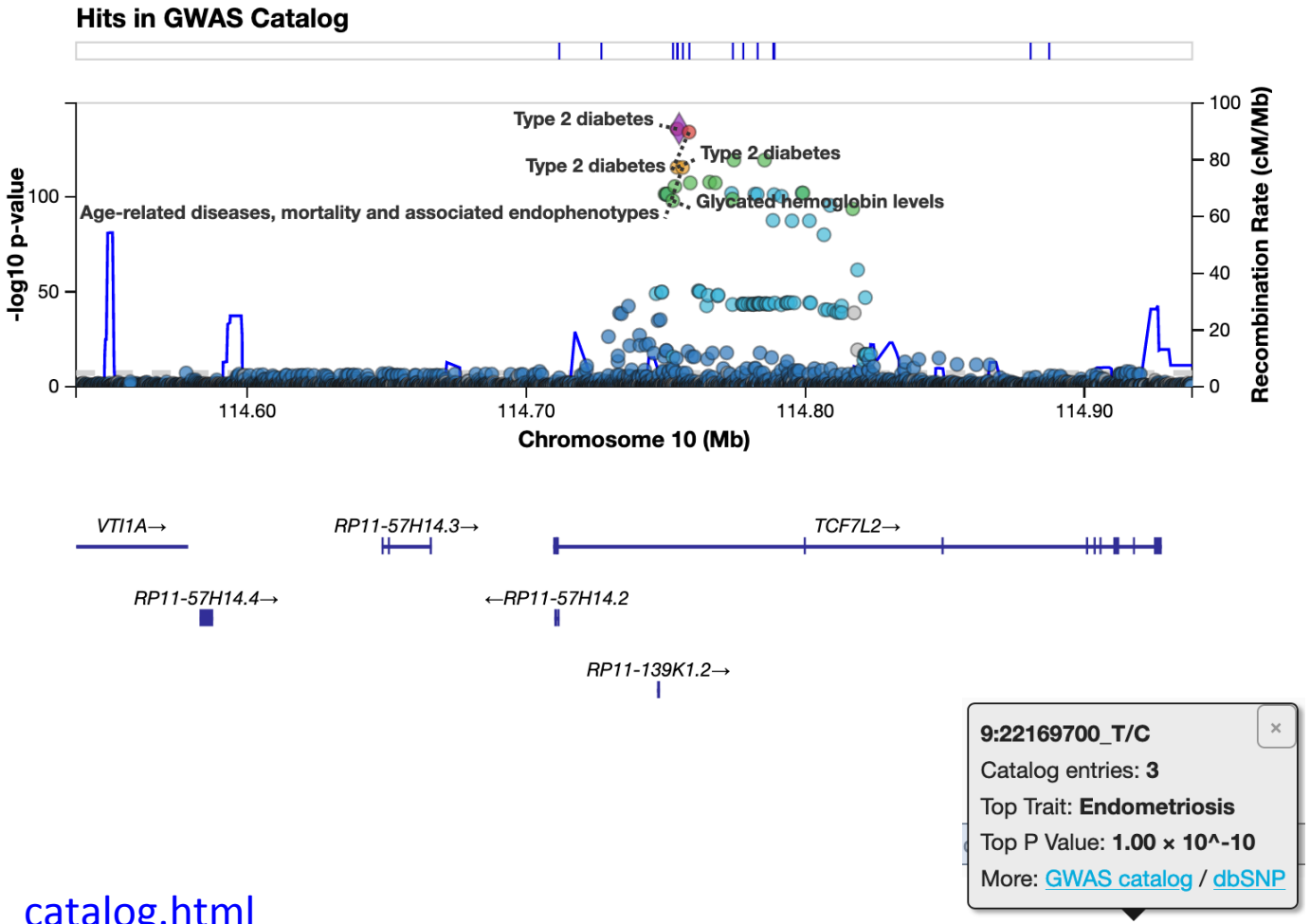



Preview: https://statgen.github.io/locuszoom/

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

# Draw comparisons to published studies

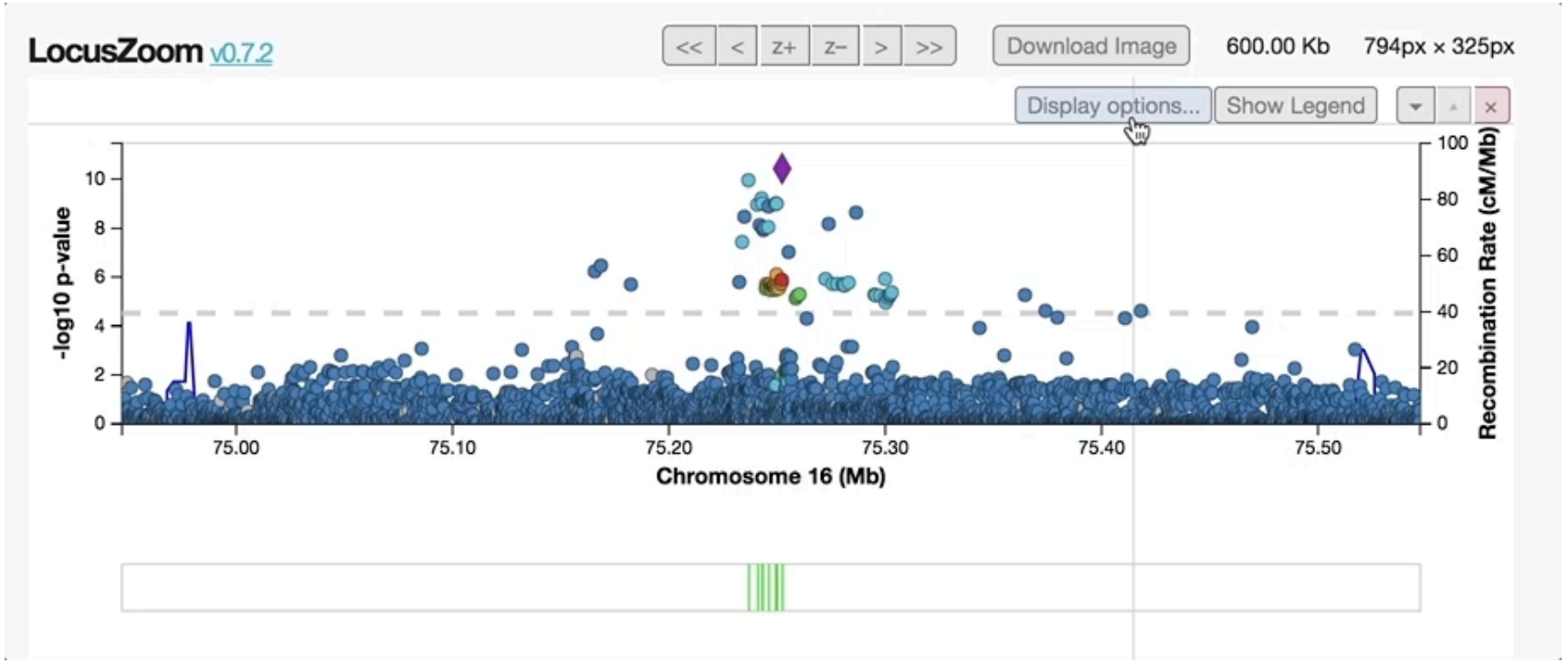
- Compare association results to known significant hits from EBI GWAS catalog
- Interactively show or hide labels
- Click a point in one panel to highlight matching elements in another panel




https://statgen.github.io/locuszoom/examples/gwas_catalog.html

# 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




https://github.com/statgen/gwas-credible-sets

Maller et al (2012) https://doi.org/10.1038/ng.2435

# Credible sets: on the Portal

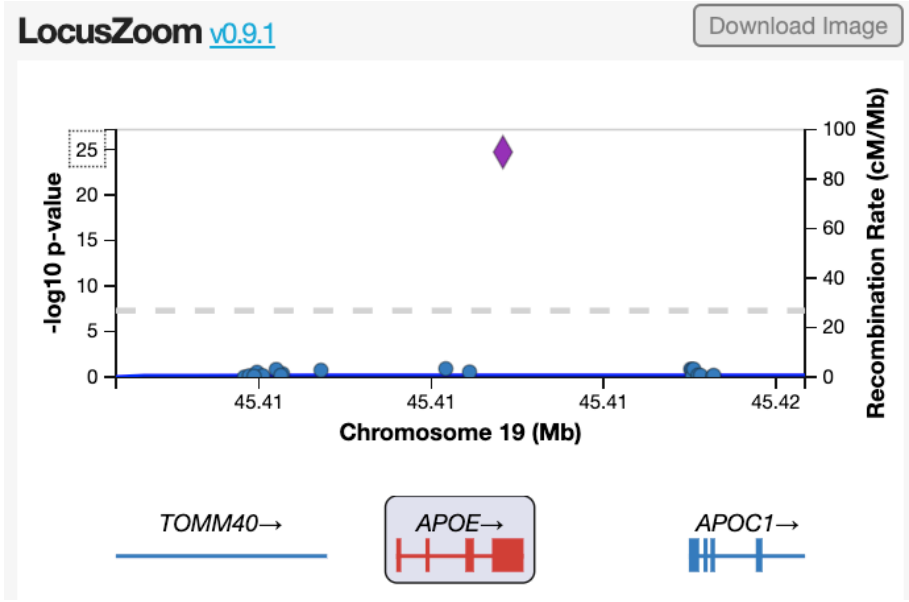
Click on a variant ID to see more information, and to access a link for tissue selection. ➜

| | | 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... | | | |


Available on the Gene Page (credible sets tab)

# 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

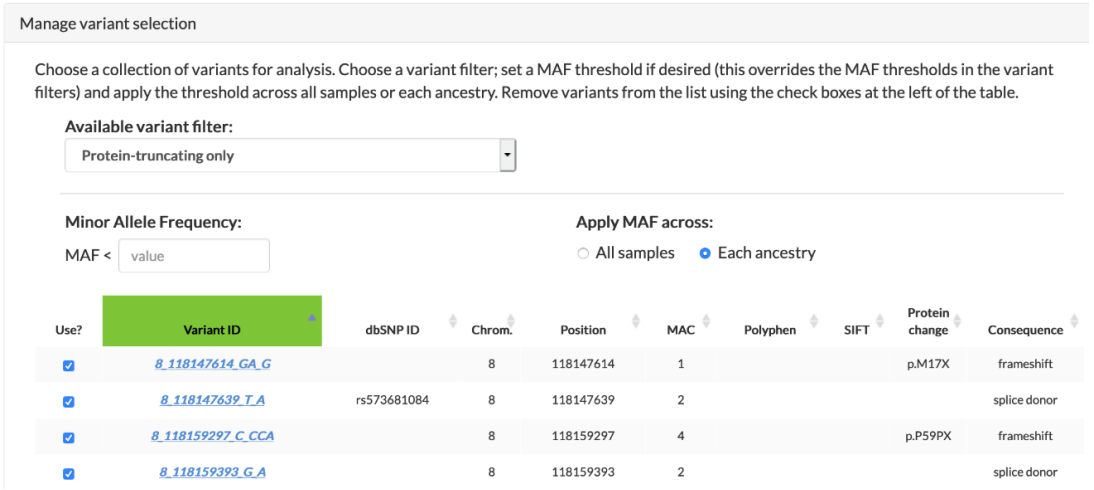
| Gene | Mask | # Variants | Test type | p-value |
|---|---|---|---|---|
| | filter column... | | filter column... | |
| APOE | Exonic SNVs from GENCODE genes with allele frequency < 5% | 3 | skat | 4.09 × 10^-24 |
| APOE | Exonic SNVs from GENCODE genes with allele frequency < 5% | 3 | burden | 1.44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

**https://github.com/statgen/raremetal.js**

https://statgen.github.io/locuszoom/examples/aggregation_tests.html

# Aggregation tests: on the Portal

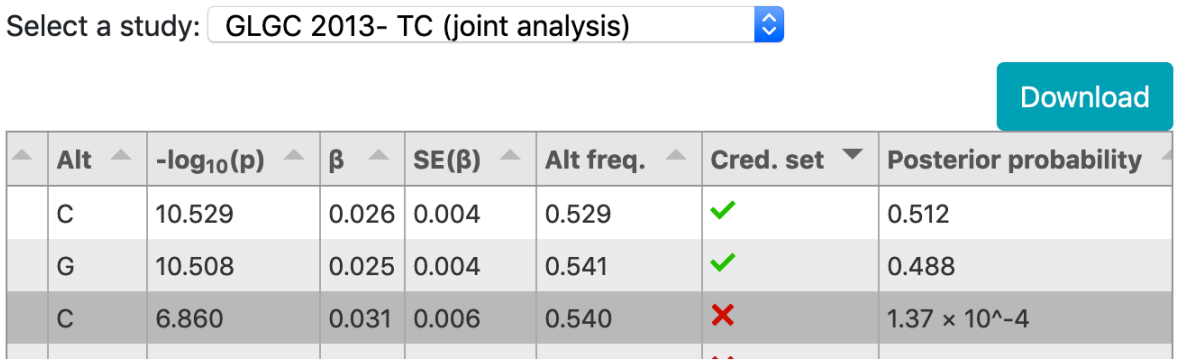
- Perform aggregation tests in a web browser, using only summary statistics (score statistics, genotype covariance matrices)
- Burden, SKAT, VT, and SKAT-O
- Soon:  conditional analysis
- Future:  meta-analysis of aggregation tests
- Explore pre-defined variant groups, or create your own via the portal




Available on the Gene Page (high impact variants tab → “Custom aggregation tests”)

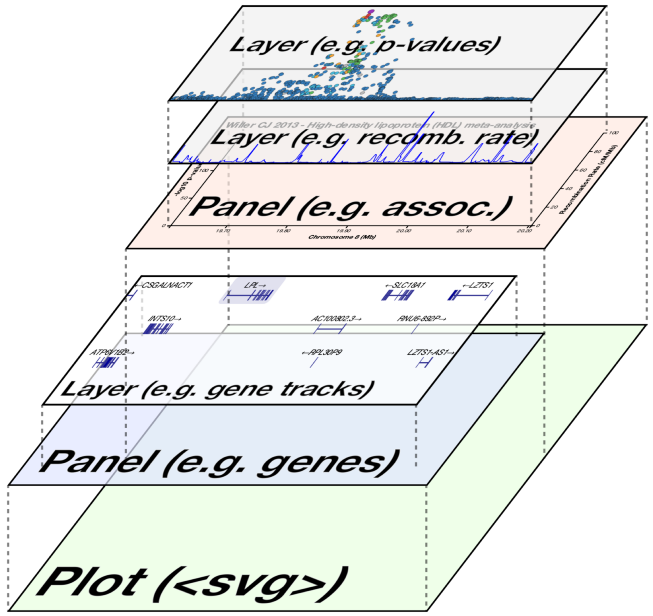
# Connect to other representations

- 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



# LocusZoom.js: Modular and reusable

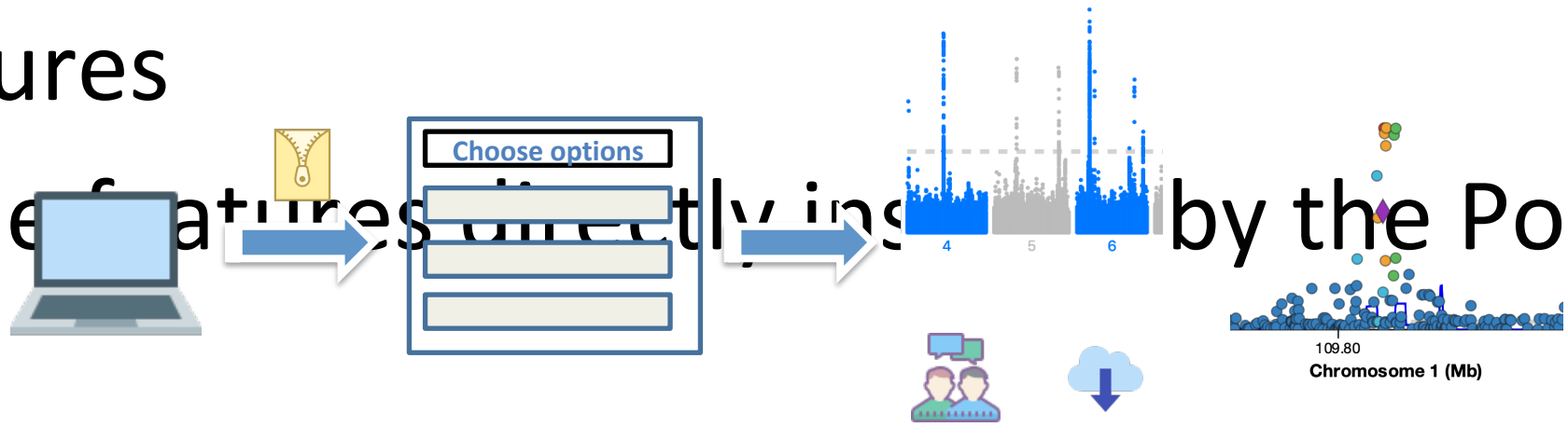
- LocusZoom.js can be embedded in your web page, and can draw data from a wide range of sources
- It is based on standard web technologies that work with many sites
- No special tooling is required, but it will work with advanced tools if you use them
- Our standard REST APIs provide common data (like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



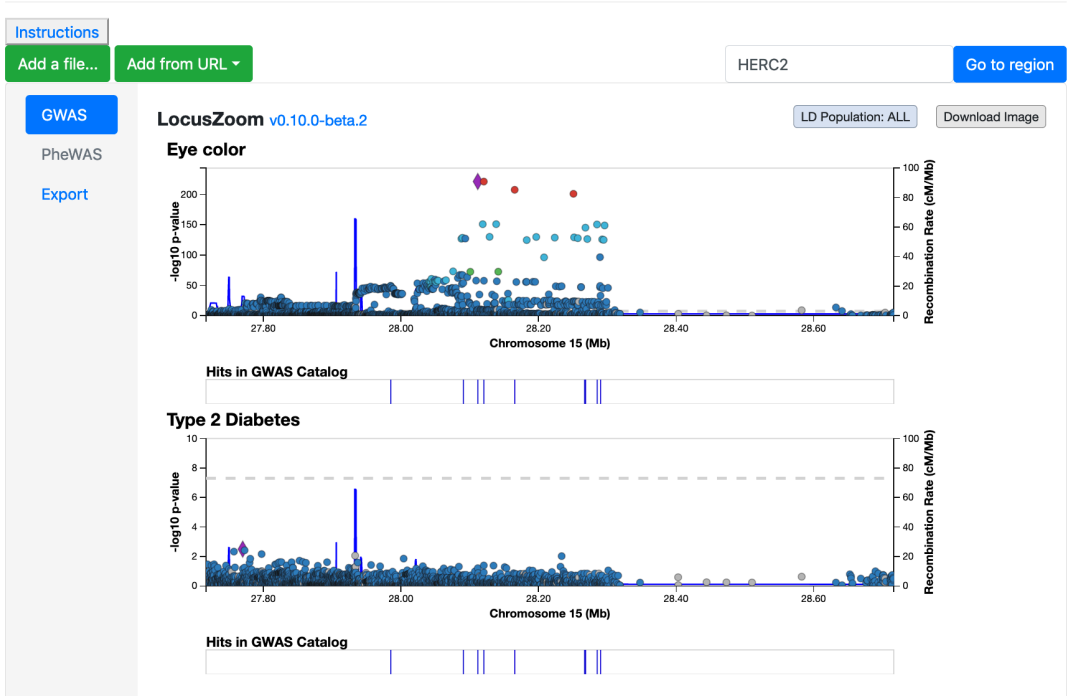

https://github.com/statgen/locuszoom/

# 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

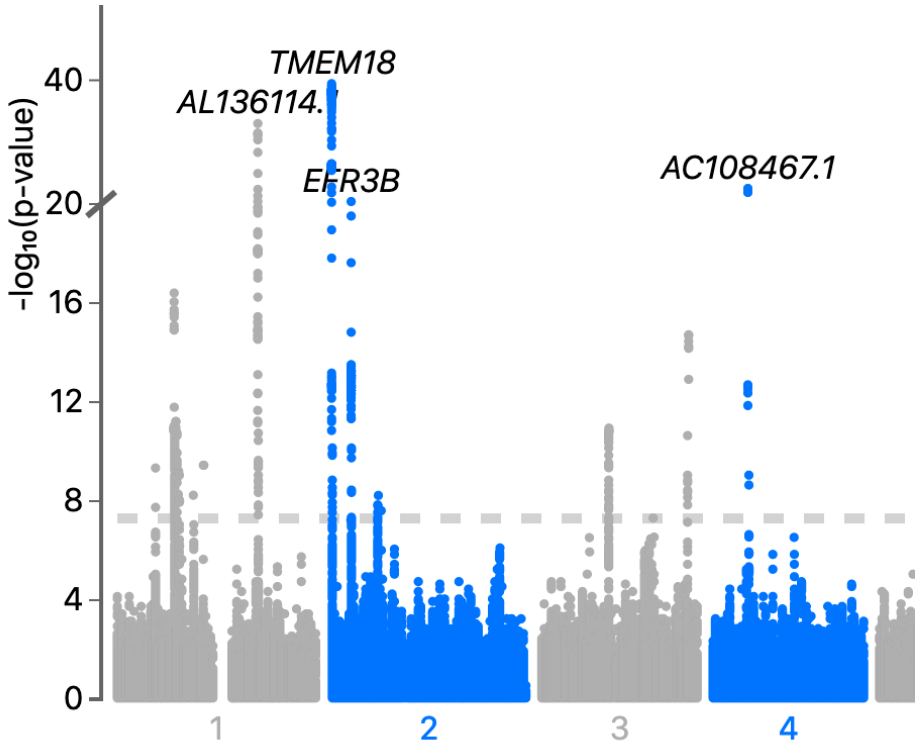



- Make a standard LocusZoom plot from tabixed GWAS summary statistics, in a web browser
- Full support for build GRCh37 and GRCh38
- Annotate with hits found in the GWAS catalog
- Perform region-based credible set calculations, in your web browser
- Explore hits for any SNP in a PheWAS of the UK BioBank

https://statgen.github.io/localzoom/

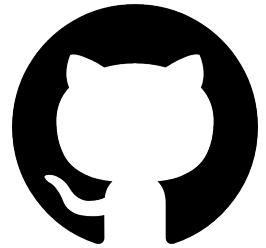
# Upload, analyze, and share: my.locuszoom.org

- 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




https://my.locuszoom.org/

# 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!

Support/ Mailing list: https://groups.google.com/forum/#!forum/locuszoom

User Survey: https://bit.ly/32ZClo8