TYPE 2 DIABETES
KNOWLEDGE PORTAL

T2D Knowledge Portal Platform Webinar & Workshop

January 16, 2020

Today: Accessing the T2DKP Platform programmatically

TYPE 2 DIABETES KNOWLEDGE PORTAL

KPN

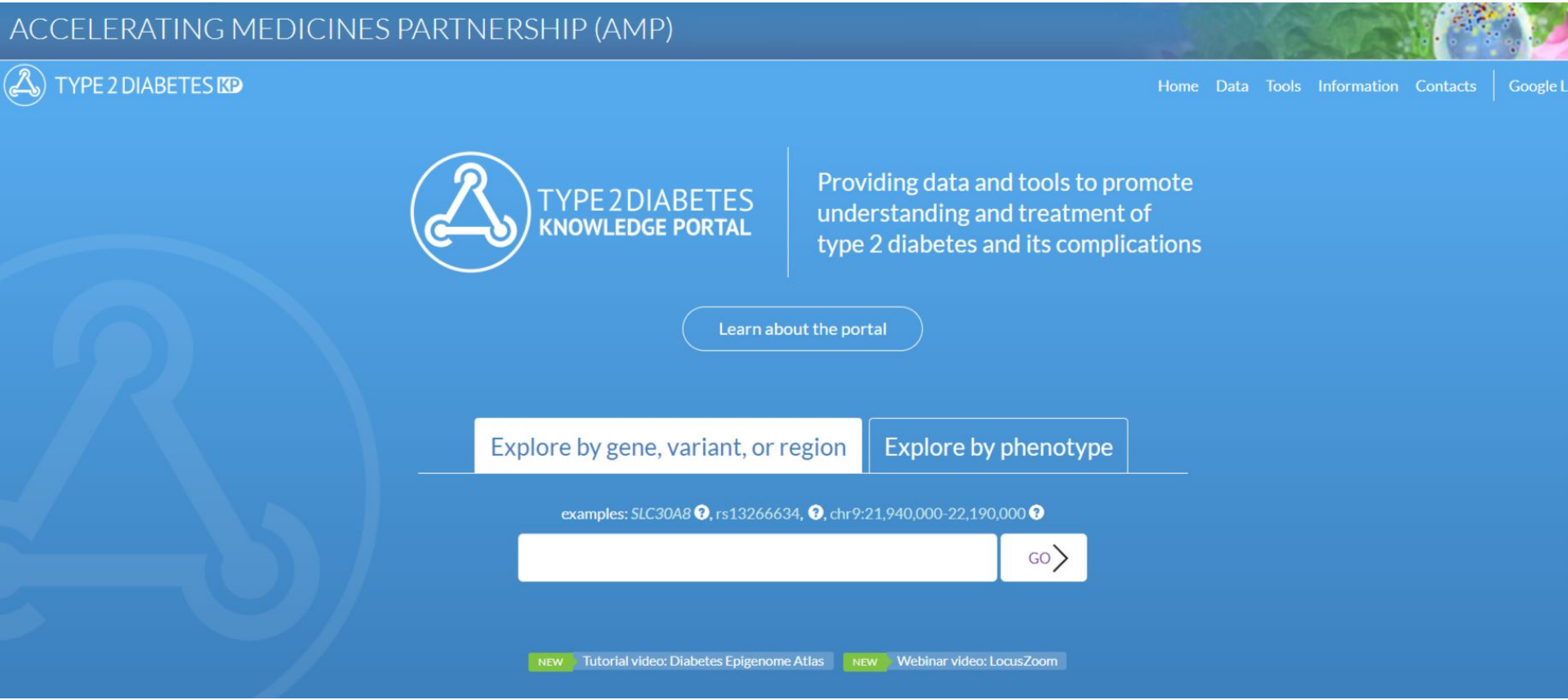
- Maria Costanzo- Gap & value of APIs
- Marc Duby- APIs: what & how?
- Oliver Ruebenacker- Practical use
- The team- Q & A

But first… what's new since October

- We applied to renew the T2DKP for 5 more years!
- We had 2 production releases
- Our sister site had 3 data releases

TYPE 2 DIABETES KNOWLEDGE PORTAL

Clean & streamlined home page



T2DKP Datasets

84 datasets, 191 traits

Filter Datasets (Click one to start)

Data type

GWAS | Whole genome sequencing | Exome sequence analysis | Exome chip | Show all

Phenotype

ANTHROPOMETRIC | ATRIAL FIBRILLATION | CARDIOVASCULAR | GLYCEMIC | HEPATIC | LIPIDS | METABOLITE | MUSCULOSKELETAL | OTHER | PSYCHIATRIC | RENAL | Show all

New Datasets (Click dataset for description)

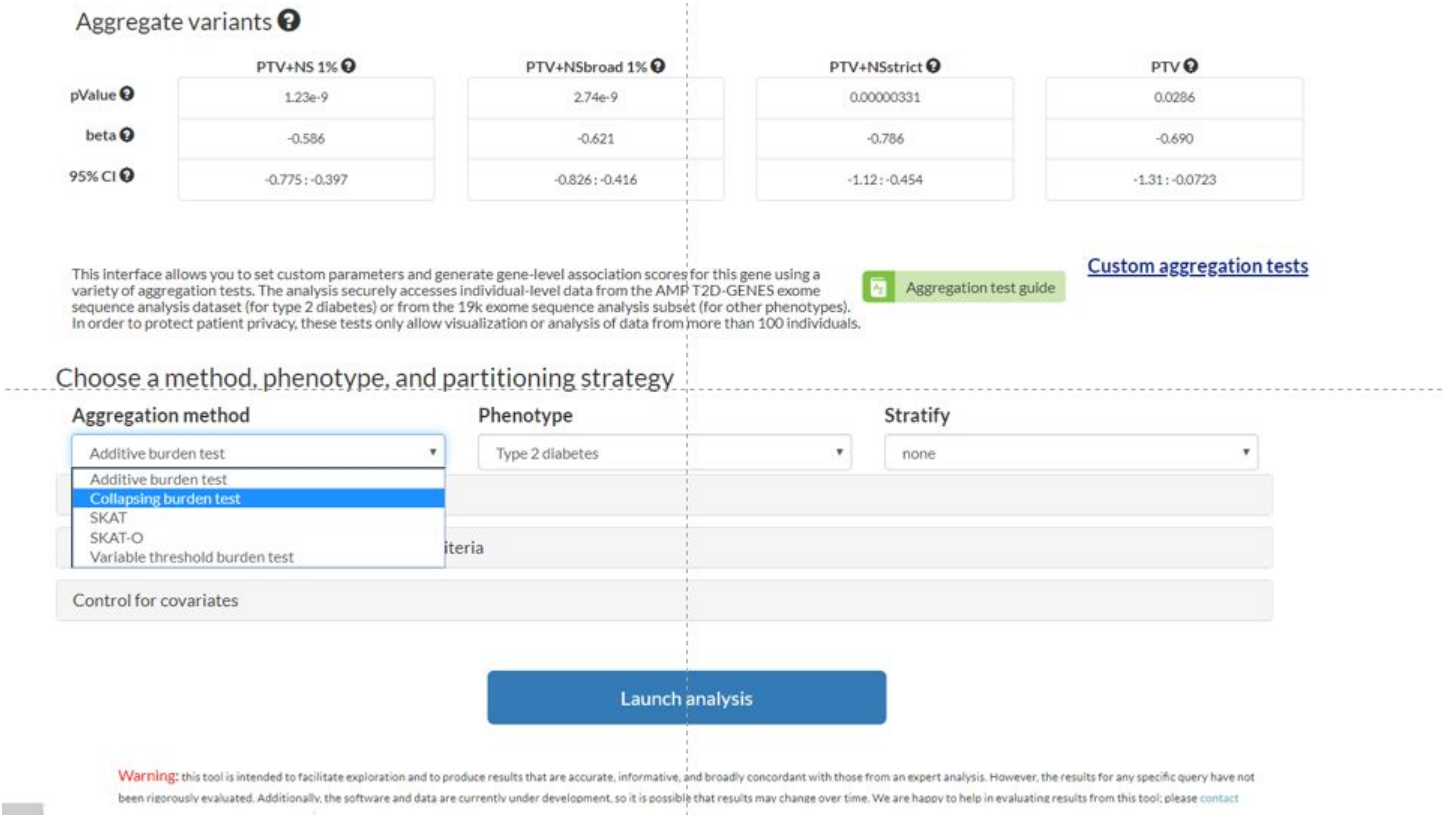
Dataset	Access	Samples	Ancestry	Data type
HERMES Heart Failure GWAS	Open access	972,032	European	GWAS
Diabetic Cohort - Singapore Prospective Study - SEED - Living Biobank GWAS	Pre-publication	12,109	Mixed	GWAS
Diabetic Cohort - Singapore Prospective Study - SEED GWAS	Pre-publication	10,248	Mixed	GWAS
Hong Kong Diabetes Register GWAS	Pre-publication	6,742	East Asian	GWAS
Singapore Prospective Study - Living Biobank GWAS	Pre-publication	3,515	East Asian	GWAS
Singapore Prospective Study Program GWAS	Pre-publication	1,896	East Asian	GWAS

Shared with publication!

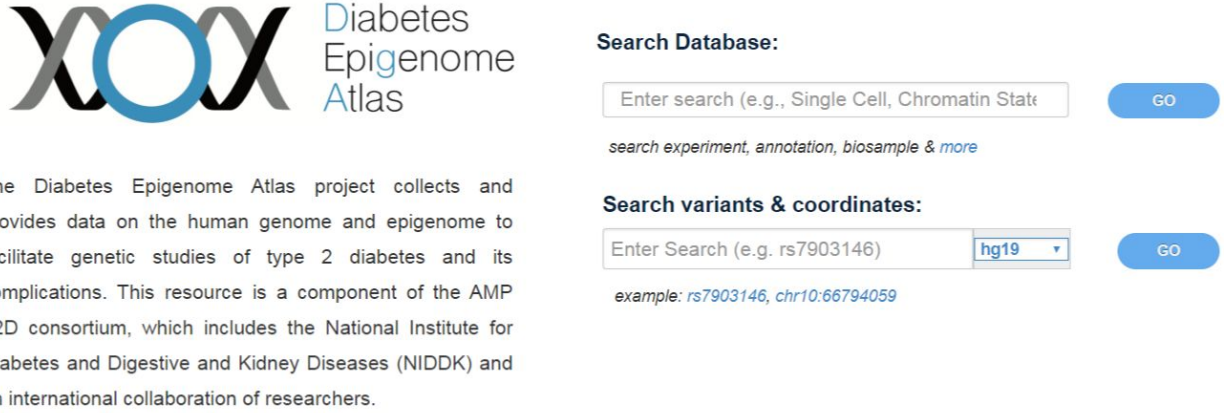
Datasets (Click dataset for description)

Dataset	Access	Samples	Ancestry	Data type

Improved interactive tests



Major DGA data release, integrated in T2DKP for February



What other events are coming up?

- Production release February 2020
    - New portal design & first version metabolic disease portal along with data for complications & T2D predictions
- New Video tutorials
    - LocusZoom, Data intake pipeline, Using the Causal Variant Query tool, API access, and more
- dkNET Webinar: February 24
- Exhibitor Booth & Talk at Endocrine Society 2020, ADA 2020

KPN

TYPE 2 DIABETES KNOWLEDGE PORTAL

The T2DKP Platform: enhancing accessibility

Data ingest

Data analysis & integration

Data visualization

Individual-level data*

Association statistics

'omics data to support interpretation

Methods

Method implementation

Pipeline APIs

Distributed APIs

Loamstream semi-automated pipeline for QC, association analysis

Association results

MySql

MySql

MySql

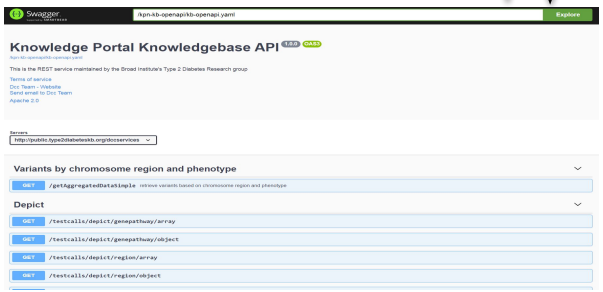
MySql

Neo4J "Knowledge graph" with links among biological entities, integrating 'omics data and results of computational methods (LD score regression, METAL, etc.)

Data sharing via federation

DGA

EBI



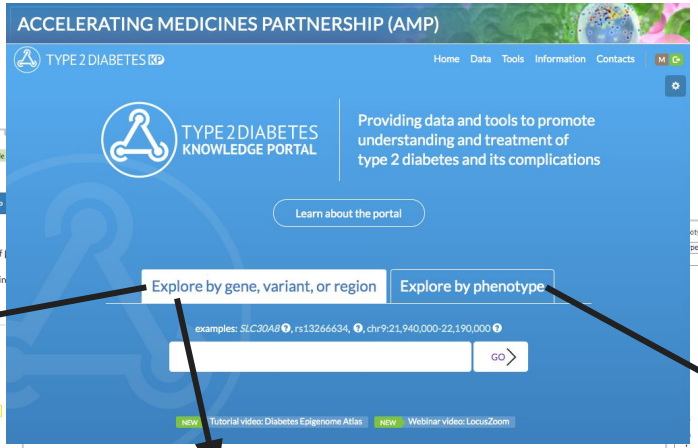
- Javascript libraries draw data from REST services
- Results and content largely configurable by text "metadata"
- All Portals deployed from same codebase

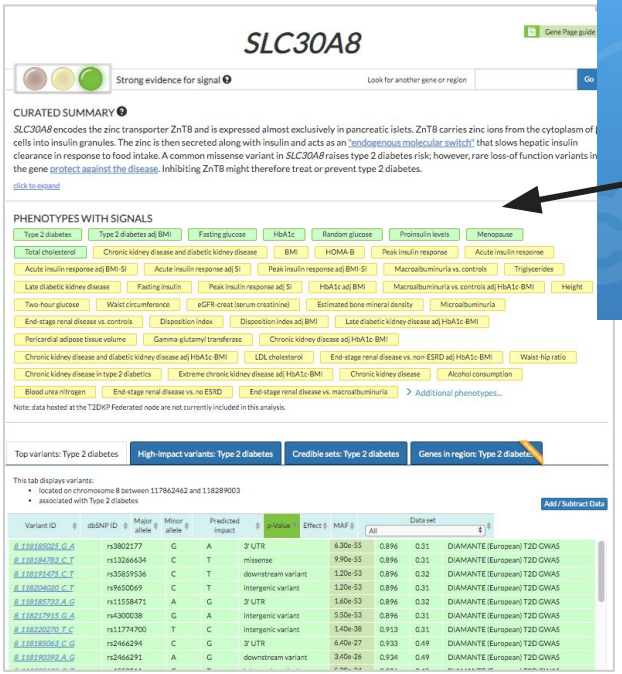
- Results shared via REST services
- Enables secure interaction with data that may not be transferred from federated site

**Individual-level data are not directly accessible to users.*

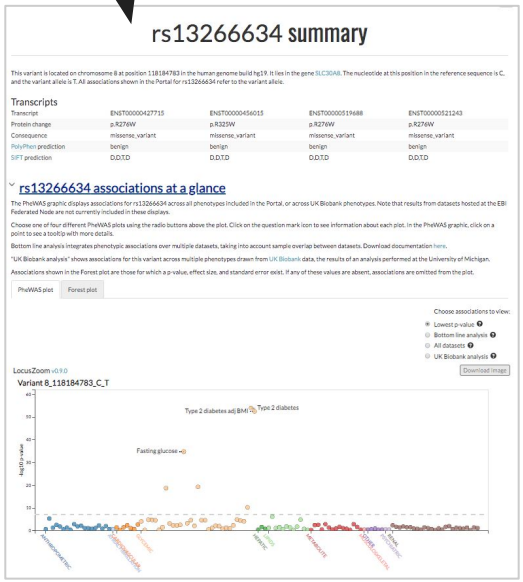
Why would you need programmatic access to the Knowledge Portal database?

Maria Costanzo

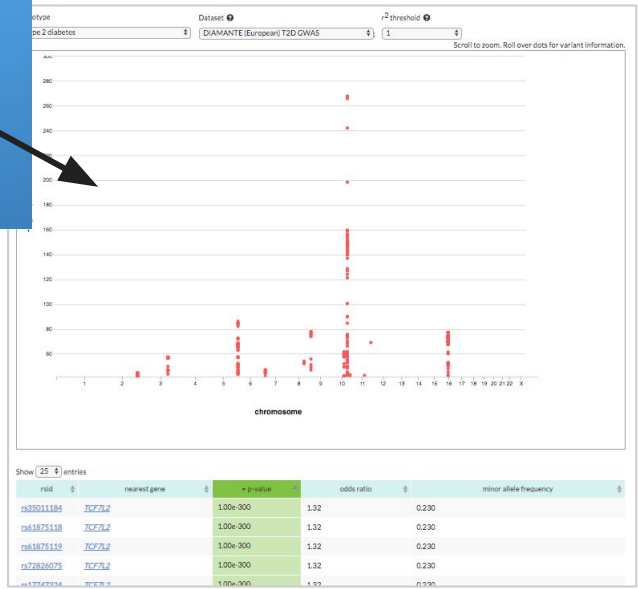




Gene/region page



Variant page

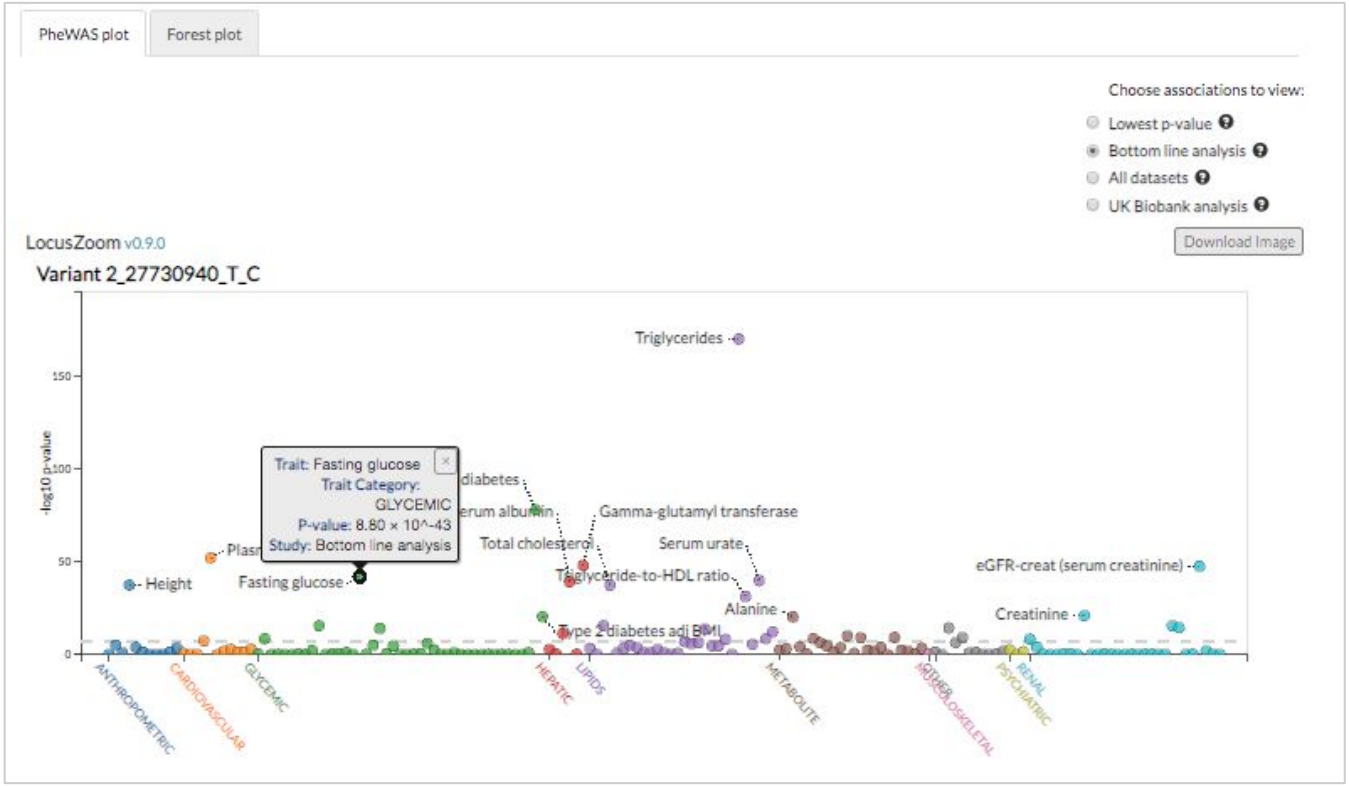


Manhattan plots

Use case 1: find all bottom-line p-values for a variant of interest

Bottom-line analysis:

- Trans-ethnic meta-analysis across multiple datasets
- Accounts for sample overlap between datasets
- Performed using METAL (U. Michigan)



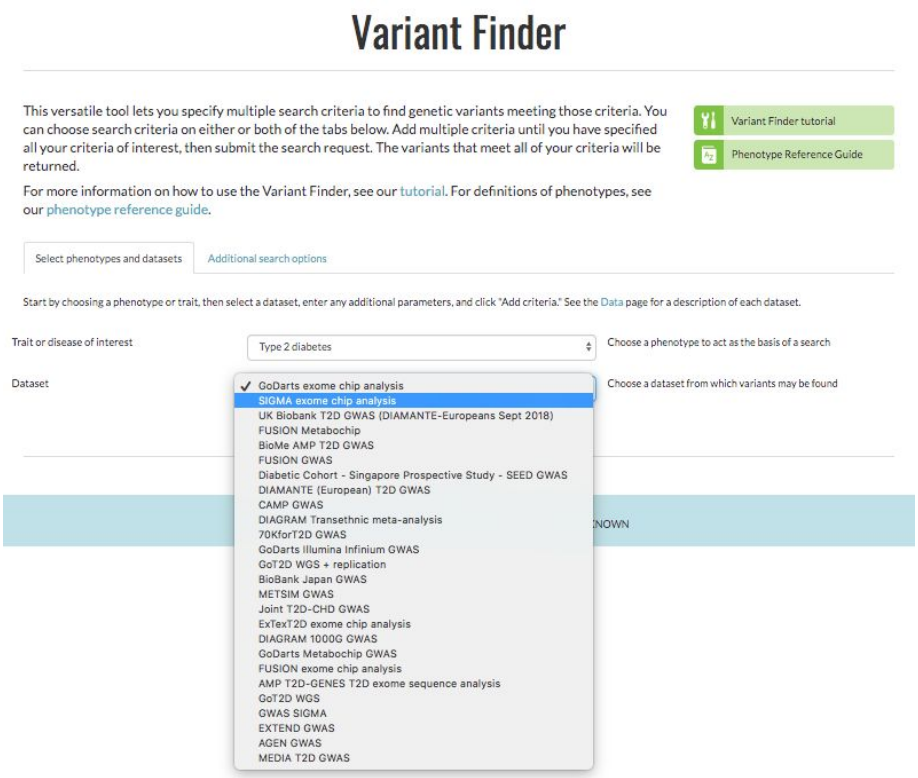
Use case 2: for a region of interest, find the set of variants most strongly associated with a phenotype

Example: find all variants

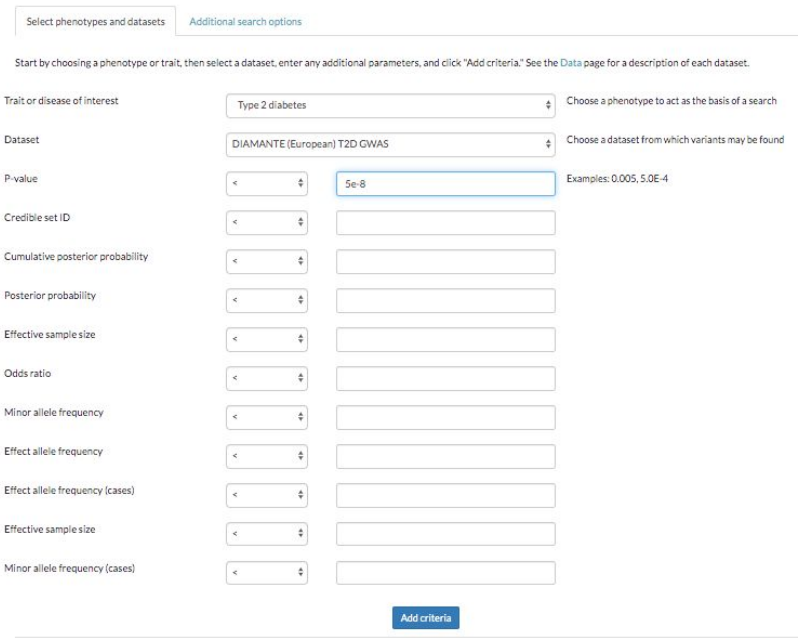
- associated with type 2 diabetes at genome-wide significance (p < 5 x 10e-8)
- on chromosome 8
- between coordinates 117862462-118289003

To do this using the Portal, build a Variant Finder query.

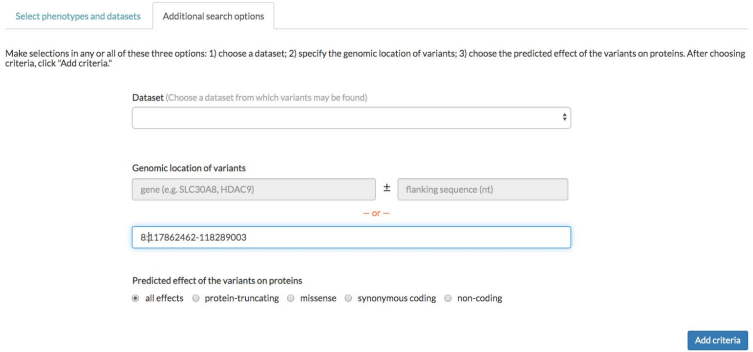
Which dataset to use??



Use case 2: for a region of interest, find the set of variants most strongly associated with a phenotype



Type 2 diabetes, DIAMANTE (European) T2D GWAS, p < 5 x 10e-8



Region: 8:117862462-118289003

Variant search results

Search criteria

Edit criteria/Start new search

Copy URL to share the search result

Results table guide

chromosome = 8

position > 117862462

position < 118289003

Type 2 diabetes[DIAMANTE (European) T2D GWAS]P-value<5e-8

For variants that do not reach significance, odds ratios may be unreliable.

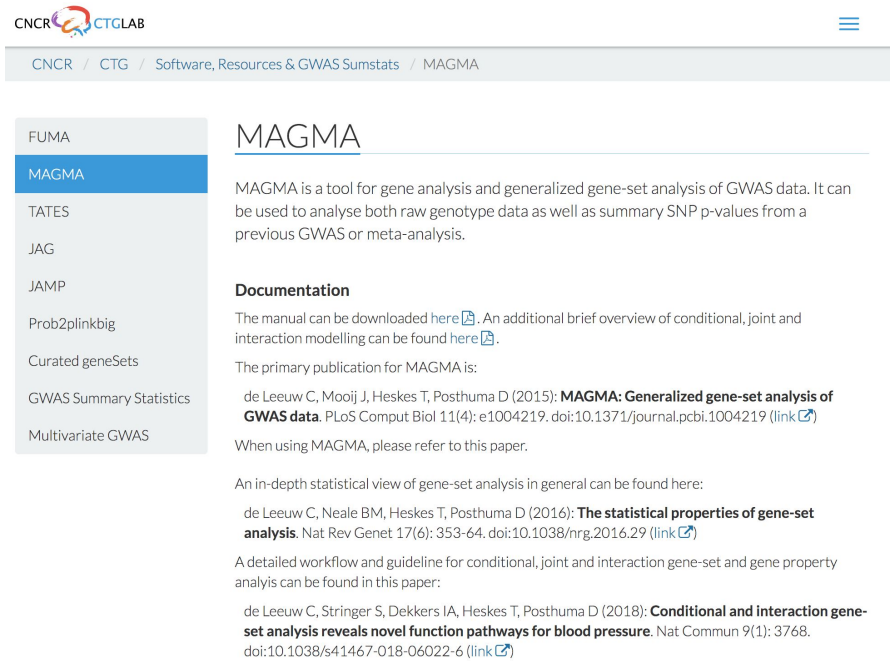
For more information on how to use this interactive table, see our Variant Results Table Guide.

Modify the table

Show 25 entries

Variant annotations									Type 2 diabetes DIAMANTE (European) T2D GWAS	
Variant ID	dbSNP ID	Chromosome	Position	Reference allele	Effect allele	Nearest gene	Protein change	Consequence	P-value	Odds ratio
8:118185025	rs3802177	8	118185025	G	A	SLC30A8		3' UTR variant	6.30e-55	0.896
8:118184783	rs13266634	8	118184783	C	T	SLC30A8	p.R325W	missense variant	9.90e-55	0.896
8:118191475	rs35859536	8	118191475	C	T	SLC30A8		downstream gene variant	1.20e-53	0.896
8:118204020	rs9650069	8	118204020	C	T	SLC30A8		intergenic variant	1.20e-53	0.896
8:118185733	rs11558471	8	118185733	A	G	SLC30A8		3' UTR variant	1.60e-53	0.896

Use case 3: retrieve results of a computational method



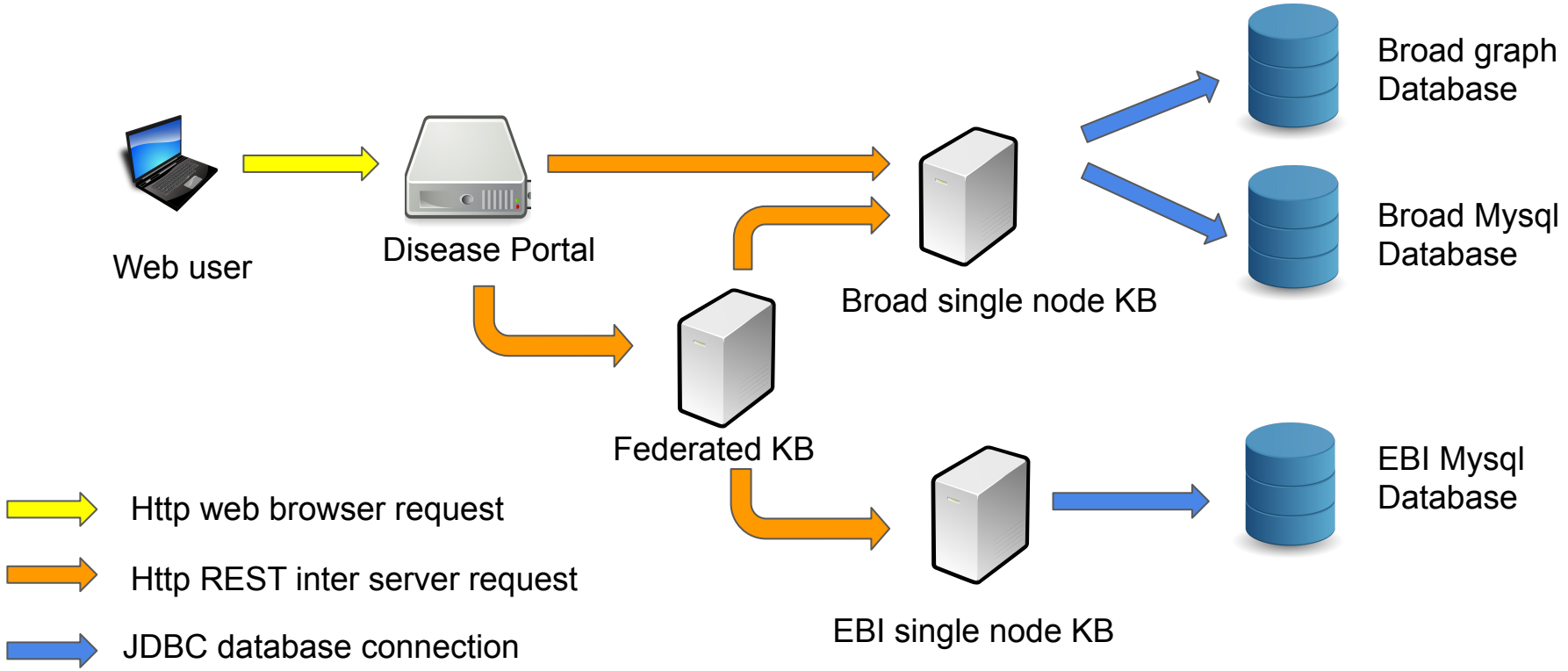
MAGMA (Multi-marker Analysis of GenoMic Annotation)

- Generates gene-level p-values for association of a gene with a phenotype
- Run on largest dataset for each phenotype
- Results will be available in the Gene FOCUS table

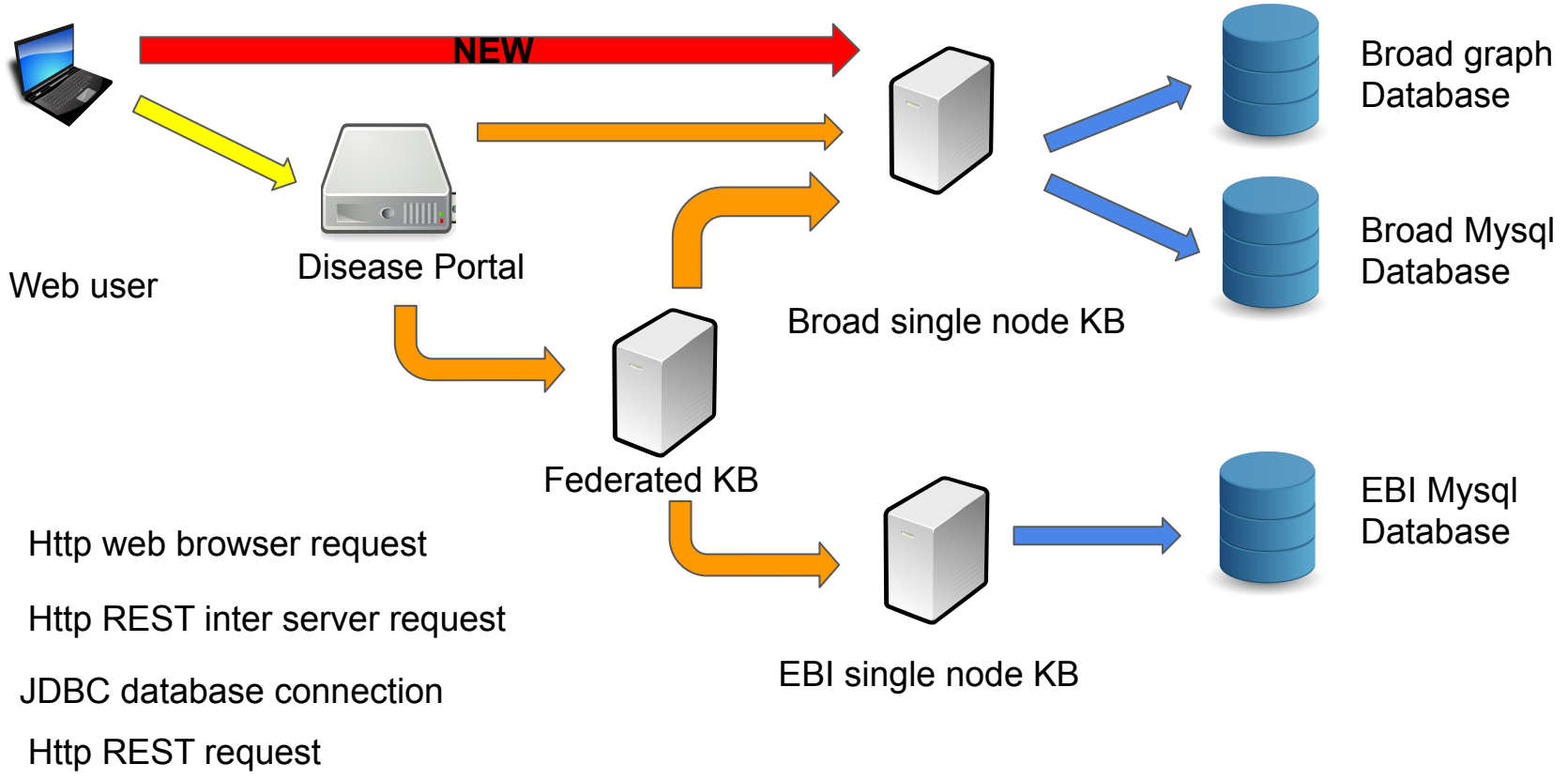
REST APIs

Marc Duby

Current Web Portal Services Architecture



Updated REST Services Architecture



Why This Architecture

- This architecture decouples the display of data from how it is stored
  - The portal can query multiple sources
    - Some outside of our organization
  - Enables the portal to choose how to combine the data
- Enables other systems/groups to query our data
  - Other portals
  - Individuals doing research

What is a REST call

- REST calls are programmatic http calls
  - An easy way to query multiple data sources at once
- Return JSON format
  - Standard hashmap based format
- Specific host
  - ie: public.type2diabeteskb.org
- Path
  - ie: /dccservices/graph/meta/variant/object
- Parameters
  - ie: var_id=8_118184783_A_G
- Example:
  - **http://public.type2diabeteskb.org/dccservices/graph/meta/variant/object?var_id=8_118184783_C_T**

Sample REST Call Result


```
{
   "data": [
      {
         "p_value": 3.11e-178,
         "std_err": 0.0034,
         "beta": -0.1066,
         "number_samples": 2237710,
         "chrom": "8",
         "pos": 118184783,
         "ref": "C",
         "alt": "T",
         "var_id": "8_118184783_C_T",
         "phenotype": "T2D",
         "phenotype_group": "GLYCEMIC"
      }
  ]
}
```

Quick Postman Demo

DCC REST Call Path Structure

- For http://public.type2diabeteskb.org/dccservices/graph/meta/variant/object?var_id=8_118184783_C_T
- http://<host>/<root_path>/<internal_indentifier>/<format>
  - Public.type2diabeteskb.org
    - Host (immutable)
  - dccservices:
    - root path (immutable)
  - graph/meta/variant
    - Internal identifier (immutable)
  - object
    - Format <object/array> (set by user)
      - Will determine how the data is returned

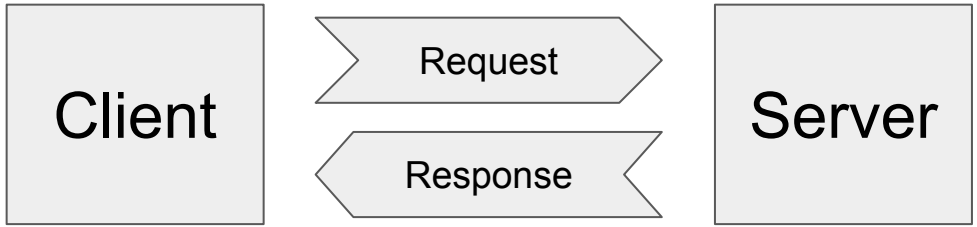
Uses and Demo

- Uses
  - Good for programmatic retrieval of real time data
  - Researcher can call multiple services and combine data as needed
- Demo
  - Postman queries
  - Jupyter notebook
- Note
  - There are hard per call limits imposed to avoid system crashes
- https://github.com/marcduby/MachineLearningPython/blob/master/DccKP/multipleMetaVariants.ipynb

OpenAPI
(aka Swagger)

Oliver Ruebenacker

Problem: keep Server, Client in sync

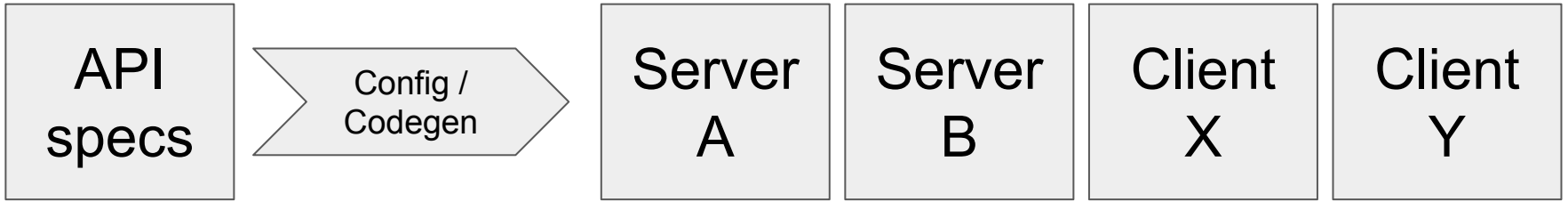


Server and client need to agree on what is valid request and response

Defining it separately for each is tedious and error-prone

Especially, if we want multiple clients in different languages

Solution: API specification



Write API specs detailing all valid requests and responses

Use specs to configure or generate code for servers and clients

OpenAPI (formerly Swagger)

Industry standard governed by Consortium

Specification for writing API specifications in JSON/YAML

Various tools to edit, validate specs

Code generation for various languages and platforms

Also, dynamic JavaScript client

DEMO!

What will you learn at future webinars?

- What aggregation tests can you run in real time on the site?
- What does the new portal design allow me to do?
- 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

Upcoming dates- all at 12 noon ET

- Thursday, March 12
- Thursday, May 14
- Thursday, July 16

TYPE 2 DIABETES KNOWLEDGE PORTAL

DCC and Knowledge Portal Team

Benjamin Alexander
Lizz Caulkins
Maria Costanzo
Marc Duby
Clint Gilbert
Quy Hoang
DK Jang
Alexandria Kluge
Ryan Koesterer
Jeffrey Massung
Oliver Ruebenacker
Preeti Singh
Marcin von Grotthuss

Noël Burtt
Jason Flannick
Jose Florez

Josep Mercader Miriam Udler
Analytical & clinical contributors

TYPE 2 DIABETES KNOWLEDGE PORTAL
type2diabetesgenetics.org

AMP T2D Knowledge Portal Development

T2DKP and DCC

Jose Florez
Jason Flannick
Noël Burtt

Ben Alexander
Lizz Caulkins
Maria Costanzo
Marc Duby
Clint Gilbert
Quy Hoang
DK Jang
Alexandria Kluge
Ryan Koesterer
Jeffrey Massung
Oliver Ruebenacker
Preeti Singh
Marcin von Grotthuss

Josep Mercader
Miriam Udler

BROAD INSTITUTE

Method and Tool Development Teams

AMP Type 2 Diabetes Knowledge (T2DK)

Daniel MacArthur
Benjamin Neale

Jonathan Bloom
Konrad Karczewski
Cotton Seed
Matthew Solomonson

BROAD INSTITUTE

AMP Enhanced Diabetes Portal (EDP)

Michael Boehnke
Gonçalo Abecasis

Andy Boughton
Christopher Clark
Matthew Flickinger
Daniel Taliun
Ryan Welch

SCHOOL OF PUBLIC HEALTH
UNIVERSITY OF MICHIGAN

AMP Federated Nodes

EBI Federated Node

Paul Flicek
Mark McCarthy
Gil McVean
Will Rayner

Thomas Keane
Dylan Spalding
Selva Kamatchinathan
Aravind Sankar

UNIVERSITY OF OXFORD

EMBL-EBI

DGA

Kyle Gaulton

Parul Kudtarkar
Ying Sun
Samuel Morabito

UC San Diego