





# T2D Knowledge Portal Platform Webinar & Workshop

# January 16, 2020

ACCELERATING MEDICINES PARTNERSHIP (AMP)

**TYPE 2 DIABETES** 

Today: Accessing the T2DKP Platfor programmatically

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





ACCELERATING MEDICINES PARTNERSHIP (AMP)

TYPE 2 DIABETES





# 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



### Clean & streamlined home page

### ACCELERATING MEDICINES PARTNERSHIP (AMP)

### 🛆 TYPE 2 DIABETES 😰

Home Data Tools Information Contacts Goo



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 (9, rs13266634, (9, chr9:21,940,000-22,190,000 (9



### **T2DKP** Datasets

### 84 datasets, 191 traits

### Filter Datasets (Click one to start)

### Data type

Dataset

GWAS	Whole genor	me sequencing	Exome	sequence analysis	Exome chip	Sho	ow all				
Phenoty	pe										
ANTHRO	POMETRIC	ATRIAL FIBRIL	LATION	CARDIOVASCUL	AR GLYCE	MIC	HEPATIC	LIPIDS	METABOLITE	MUSCULOSKELETAL	OTHER
PSYCHIA		IAL Show all									

### New Datasets (Click dataset for description)

Dataset	Access	Samples	Ancestry	Data type	
HERMES Heart Failure GWAS	Open access	972,032	European	GWAS	Shared with
Diabetic Cohort - Singapore Prospective Study - SEED - Living					nublication
Biobank GWAS	Pre-publication	12,109	Mixed	GWAS	publication
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	

### Datasets (Click dataset for description)

|--|

### Improved interactive tests



### Choose a method, phenotype, and partitioning strategy



been rigorously evaluated. Additionally, the software and data are currently under development, so it is possible that results may change over time. We are happy to help in evaluating results from this tool; please contact

# Major DGA data release, integrated in T2DKP for February



The Diabetes Epigenome Atlas project collects and provides data on the human genome and epigenome to facilitate genetic studies of type 2 diabetes and its complications. This resource is a component of the AMP T2D consortium, which includes the National Institute for Diabetes and Digestive and Kidney Diseases (NIDDK) and an international collaboration of researchers.

# Search Database: Enter search (e.g., Single Cell, Chromatin State GO search experiment, annotation, biosample & more Search variants & coordinates: Enter Search (e.g. rs7903146) hg19 r GO example: rs7903146, chr10:66794059 GO GO

### News View all news...

### January 2020 Data Release

January 15, 2020

74 Annotations for accessible chromatin sites of diabetese relevant datasets from GEO & ENCODE and chromatin state from ROADMAP released Read more

### December 2019 Data release

December 9, 2019

We are pleased to announce relevant experimental data from 4 latest publications

Read more

# 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



# The T2DKP Platform: enhancing accessibility





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

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

Maria Costanzo

#### ACCELERATING MEDICINES PARTNERSHIP (AMP)

Gene Page guide



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

GO>

Explore by gene, variant, or region Explore by phenotype

futorial video: Diabetes Epigenome Atlas NEW Webinar video: LocusZoom

### rs13266634 summary

This variant is located on chromosome 8 at position 118184780 in the human genome build hg19. It lies in the gene SLC30A8. The nucleotide at this position in the reference sequence is C. and the variant allele is T.AI associations shown in the Portal for rs12266634 refer to the variant allele.

#### Transcripts

Transcript	ENST00000427715	ENST00000456015	ENST00000519688	ENST00000521243
Protein change	p.R276W	p.R325W	p.R276W	p.R276W
Consequence	missense_variant	missense_variant	missense_variant	missense,variant
PolyPhen prediction	benign	benign	benign	benign
SIFT prediction	D.D.T.D	D.D.T.D	D.D.T.D	D.D.T.D

#### rs13266634 associations at a glance

The PheWAS graphic displays associations for rs13266634 across all phenotypes included in the Portal, or across UK Biobank phenotypes. Note that results from datasets hosted at the EB Federated Node are not currently included in these displays. Choose one of four different PheWAS plots using the radio buttors above the plot. Click on the question mark icon to see information about each plot. In the PheWAS graphic, click on a

ottom line analysis integrates phenotypic associations over multiple datasets, taking into account sample overlap between datasets. Do "UK Biobank analysis" shows associations for this variant across multiple phenotypes drawn from UK Biobank data, the results of an analys

for which a p-value, effect size, and standard error exist. If any of these values are abse

PheWAS plot Forest plot Choose associations to vie · Lowest purplue O Bottom line analysis O All datasets O UK Biobank analysis 9 LocusZoom v0.9.0 Variant 8\_118184783\_C\_T Type 2 diabetes adj BMI Type 2 diabetes Fasting glucose -0 a alla hand alla alla Parlona Pan de Panantale



### Manhattan plots

### SLC30A8

Strong evidence for signal O

#### CURATED SUMMARY

SLC30AB encodes the zinc transporter ZnT8 and is expressed almost exclusively in pancreatic islets. ZnT8 carries zinc ions from the cytoplasm of cells into insulin granules. The zinc is then secreted along with insulin and acts as an "endogenous molecular switch" that slows hepatic insulin clearance in response to food intake. A common missense variant in SLC30AB raises type 2 diabetes risk; however, rare loss-of function variants in the gene protect against the disease. Inhibiting ZnT8 might therefore treat or prevent type 2 diabetes. click to expand

Look for another gene or region

#### PHENOTYPES WITH SIGNALS

Type 2 diabetes Type 2 diabetes adj BMI Fasting glucose HbA1c Random glucose Proinsulin levels Menopause
Total cholesterol Chronic kidney disease and diabetic kidney disease BMI HOMA-B Peak insulin response Acute insulin response
Acute insulin response adj BMI-SI Acute insulin response adj SI Peak insulin response adj BMI-SI Macroalbuminuria vs. controls Triglycerides
Late diabetic kidney disease Fasting insulin Peak insulin response adj SI HbA1c adj BMI Macroalbuminuria vs. controls adj HbA1c-BMI H
Two-hour glucose Walst circumference eGFR-creat (serum creatinine) Estimated bone mineral density Microalbuminuria
End-stage renal disease vs. controls Disposition index Disposition index adj BMI Late diabetic kidney disease adj HbA1c-BMI
Pericardial adipose tissue volume Gamma-glutamyl transferase Chronic kidney disease adj HbA1c-BMI
Chronic kidney disease and diabetic kidney disease adj HbA1c-BMI LDL cholesterol End-stage renal disease vs. non-ESRD adj HbA1c-BMI Waist-hip ratio
Chronic kidney disease in type 2 diabetics Extreme chronic kidney disease adj HbA1c-BMI Chronic kidney disease Alcohol consumption
Blood urea nitrogen End-stage renal disease vs. no ESRD End-stage renal disease vs. macroalbuminuria > Additional phenotypes
ote: 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			s in region: Type 2 diabete
This tab displays varia located on ch associated wi	ints: romosome 8 bi th Type 2 diabe	rtween 11 rtes	7862462 ar	nd 118289003						Add / Su
Variant ID 0	dbSNP ID 0	Major allele	Minor 0	Predicted impact	0 p-Valu	Effect (	MAF 0	All	Data set	•) <sup>0</sup>
8 118185025 G A	rs3800	2177	G	A	3' UTR		6.30e-5	5 0.896	0.31	DIAMANTE (European) T2D GWAS
8 118184783 C T	rs132	56634	с	т	missense		9.90e-5	5 0.896	0.31	DIAMANTE (European) T2D GWAS
8 118191475 C.T	rs3585	59536	С	т	downstream v	ariant	1.20e-5	3 0.896	0.32	DIAMANTE (European) T2D GWAS
8 118204020 C T	rs9650	0069	с	т	intergenic var	iant	1.20e-5	3 0.896	0.31	DIAMANTE (European) T2D GWAS
8 118185733 A G	rs115	58471	A	G	3' UTR		1.60e-5	3 0.896	0.32	DIAMANTE (European) T2D GWAS
8.118217915 G.A	rs430	0038	G	A	intergenic var	lant	5.50e-5	3 0.896	0.31	DIAMANTE (European) T2D GWAS
8 118220270 T C	rs117	74700	т	с	Intergenic var	lant	1.40e-3	8 0.913	0.31	DIAMANTE (European) T2D GWAS
8 118185063 C G	rs246e	5294	С	G	3' UTR		6.40e-2	7 0.933	0.49	DIAMANTE (European) T2D GWAS
8 118190393 A.G	rs246	5291	A	G	downstream v	ariant	3.40e-2	6 0.934	0.49	DIAMANTE (European) T2D GWAS
	100		~	-			F 00- 0			

### Gene/region page

### Variant page

# 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??

### Variant Finder

This versatile tool lets you specify multiple search criteria to find genetic variants meeting those criteria. You can choose search criteria on either or both of the tabs below. Add multiple criteria until you have specified all your criteria of interest, then submit the search request. The variants that meet all of your criteria will be returned.



For more information on how to use the Variant Finder, see our tutorial. For definitions of phenotypes, see our phenotype reference guide.

Select phenotypes and datasets Additional search options

Start by choosing a phenotype or trait, then select a dataset, enter any additional parameters, and click "Add criteria." See the Data page for a description of each dataset.

Trait or disease of interest	Type 2 diabetes	Choose a phenotype to act as the basis of a search
Dataset	GODarts exome chip analysis	Choose a dataset from which variants may be found
	UK Biobank 12D (WAS (DIAMANTE-Europeans Sept 2018) FUSION Metabochip BioMe AMP 12D GWAS FUSION GWAS	l
	Diabetic Cohort - Singapore Prospective Study - SEED GWAS DIAMANTE (European) T2D GWAS	
	CAMP GWAS DIAGRAM Transethnic meta-analysis 70KforT2D GWAS	NOWN
	GoDarts Illumina Infinium GWAS GoT2D WGS + replication BioBank Japan GWAS METSIM GWAS	
	Joint 12D-CHD GWAS ExText2D exome chip analysis DIAGRAM 1000G GWAS GoDarts Metabochip GWAS	
	AMP T20-GENES T2D exome sequence analysis GoT2D WGS GWAS SIGMA	
	EXTEND GWAS AGEN GWAS MEDIA T2D GWAS	

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

Select phenotypes and datasets	Additional search options			Select phenotypes and datasets	Additional search options		
Start by choosing a phenotype or trait	then select a dataset, enter any	additional parameters, and click "Add crit	teria." See the Data page for a description of each dataset.	Make selections in any or all of these criteria, click "Add criteria."	three options: 1) choose a dataset; 2) specify the genon	nic location of variants; 3) choose the predicted effe	ct of the variants on proteins. After choosing
Trait or disease of interest	Type 2 diabetes		Choose a phenotype to act as the basis of a search	Dat	aset (Choose a dataset from which variants may be found)	\$	
Dataset	DIAMANTE (Europe	ean) T2D GWAS	Choose a dataset from which variants may be found				
P-value	< \$	5e-8	Examples: 0.005, 5.0E-4	Gen B <sup>6</sup>	iomic location of variants ene (e.g. SLC30A8, HDAC9)	± flanking sequence (nt)	
Credible set ID	< \$			8	-	- or	1
Cumulative posterior probability	< \$						
Posterior probability	< \$			• a	all effects   protein-truncating  missense  sy	nonymous coding 💿 non-coding	
Effective sample size	< \$						Add criteria
Odds ratio	< \$						
Minor allele frequency	< \$						
Effect allele frequency	< \$			Regi	on: 8:117862	462-1182890	)03
Effect allele frequency (cases)	< \$						
Effective sample size	< \$						
Minor allele frequency (cases)	< \$						
		Add criteria					

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

### Variant search results



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

Modify the table

Tran 2 diabat

#### Show 25 \$ entries

									Type 2	2 diabetes
				Variant annotat	ions				DIAMANTE (Eur	ropean) T2D GWAS
Variant ID 🛛 🌲	dbSNP ID 👙	Chromosome 🍦	Position 🔶	Reference allele   🎍	Effect allele   🌲	Nearest gene   🌲	Protein change 👙	Consequence \$	P-value	Odds ratio 👙
<u>8:118185025</u>	rs3802177	8	118185025	G	А	SLC30A8		3' UTR variant	6.30e-55	0.896
<u>8:118184783</u>	rs13266634	8	118184783	с	Т	SLC30A8	p.R325W	missense variant	9.90e-55	0.896
<u>8:118191475</u>	rs35859536	8	118191475	с	т	SLC30A8		downstream gene variant	1.20e-53	0.896
<u>8:118204020</u>	rs9650069	8	118204020	с	Т	SLC30A8		intergenic variant	1.20e-53	0.896
<u>8:118185733</u>	rs11558471	8	118185733	А	G	<u>SLC30A8</u>		3' UTR variant	1.60e-53	0.896

### Use case 3: retrieve results of a computational method

	=
CNCR / CTG / Software	2, Resources & GWAS Sumstats / MAGMA
FUMA	MAGMA
MAGMA	MACMAA is a tool for some analysis and constrained some set analysis of CMAS data. It some
TATES	be used to analyse both raw genotype data as well as summary SNP p-values from a
JAG	previous GWAS or meta-analysis.
JAMP	Documentation
Prob2plinkbig	The manual can be downloaded here 🔁 . An additional brief overview of conditional, joint and interaction modelling can be found here 🖪 .
Curated geneSets	The primary publication for MAGMA is:
GWAS Summary Statistics	de Leeuw C, Mooij J, Heskes T, Posthuma D (2015): MAGMA: Generalized gene-set analysis of GWAS data. PLoS Comput Biol 11(4): e1004219. doi:10.1371/journal.pcbi.1004219 (link 🗷)
Multivariate GWAS	When using MAGMA, please refer to this paper.
	An in-depth statistical view of gene-set analysis in general can be found here:
	de Leeuw C, Neale BM, Heskes T, Posthuma D (2016): <b>The statistical properties of gene-set analysis</b> . Nat Rev Genet 17(6): 353-64. doi:10.1038/nrg.2016.29 (link ☑)
	A detailed workflow and guideline for conditional, joint and interaction gene-set and gene property analyis can be found in this paper:
	de Leeuw C, Stringer S, Dekkers IA, Heskes T, Posthuma D (2018): <b>Conditional and interaction gene</b>

doi:10.1038/s41467-018-06022-6 (link C)

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 <a href="http://public.type2diabeteskb.org/dccservices/graph/meta/variant/object?var\_id=8\_118184783\_C\_T">http://public.type2diabeteskb.org/dccservices/graph/meta/variant/object?var\_id=8\_118184783\_C\_T</a>
- http://<host>/<root\_path>/<internal\_indentifier>/<format>
  - Public.type2diabeteskb.org
    - Host (immutable)
  - dccservices:
    - root path (immutable)
  - o 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
- <u>https://github.com/marcduby/MachineLearningPython/blob/master/DccKP/multipleMetaVariants.ipynb</u>

# 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?





# Upcoming dates- all at 12 noon ET

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



### 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























Josep Mercader Miriam Udler Analytical & clinical contributors



### AMP T2D Knowledge Portal Development

Method and Tool Development Teams

### 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



AMP Type 2 Diabetes Knowledge (T2DK)

Daniel MacArthur Benjamin Neale

Jonathan Bloom Konrad Karczewski Cotton Seed Matthew Solomonson



AMP Enhanced Diabetes Portal (EDP)

Michael Boehnke Gonçalo Abecasis

Andy Boughton Christopher Clark Matthew Flickinger Daniel Taliun Ryan Welch



AMP Federated Nodes

EBI Federated Node

Paul Flicek Mark McCarthy Gil McVean Will Rayner

Thomas Keane Dylan Spalding Selva Kamatchinathan Aravind Sankar



Kyle Gaulton

DGA

Parul Kudtarkar Ying Sun Samuel Morabito

XOX UC San Diego

ACCELERATING MEDICINES PARTNERSHIP (AMP)

**TYPE 2 DIABETES**