Human Islet Genotyping Initiative (HIGI) Pipeline

Last updated: 01/29/26

Stanford MEDICINE | Translational Genomics of Diabetes Lab
In the Division of Pediatric Endocrinology & Diabetes

Translational Genomics of Diabetes Lab

INTEGRATED ISLET DISTRIBUTION PROGRAM
IIDP
HIGI
HUMAN ISLET GENOTYPING INITIATIVE

Translational Genomics of Diabetes Lab

https://med.stanford.edu/genomics-of-diabetes.html

- The team is based in the Division of Endocrinology in the Department of Pediatrics at Stanford University. We aim to understand the genetic basis of diabetes and related metabolic conditions and to use this to leverage a better understanding of what causes diabetes and how we can improve treatment options for patients. Our work is predominantly focused on understanding what causes pancreatic islets to release insufficient insulin to control blood glucose levels after a meal in patients with type 2 diabetes, but often extends to efforts to relate this to metabolic dysfunction in other relevant tissues such as fat and liver.
- We have a long-standing interest in genetically characterizing human islet donors and relating their genotypes to islet gene expression and function. We have worked with the Alberta Islet Core (https://www.humanislets.com/#/)for many years. Since our relocation from Oxford, UK to the USA we have been able to lend our experience with these pipelines to the NIDDK funded Human Islet Research Network (HIRN) where we play a role in two of their initiatives. The Human Pancreas Analysis Program (HPAP) and the Integrated Islet Distribution Program (IIDP).
- Our role in these initiatives is to support the genetic characterization of islets which are distributed for research by the IIDP and the pancreas donors who are phenotyped within the HPAP program.
- We have worked with all 3 programs to establish an interoperable pipeline for sample processing and analysis and harmonised visualisations of data.

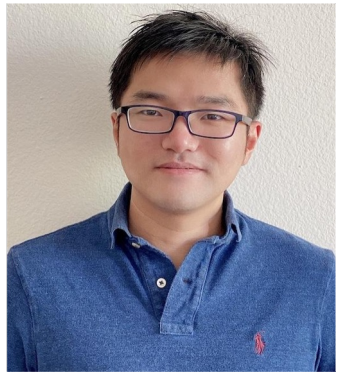
Translational Genomics of Diabetes Lab

INTEGRATED ISLET DISTRIBUTION PROGRAM IIDP HIGI HUMAN ISLET GENOTYPING INITIATIVE

The Current HIGI Team

Anna Gloyn
Principal Investigator

Swaraj Thaman
Research Assistant



Dr. Han Sun
Computational Biologist

Dr. Lu Zhang
Postdoc

HIGI Alumni

Dr. Alok Jha
Computational Biologist

Dr. Seth Sharp
Postdoc

Varsha Rajesh
Research Assistant

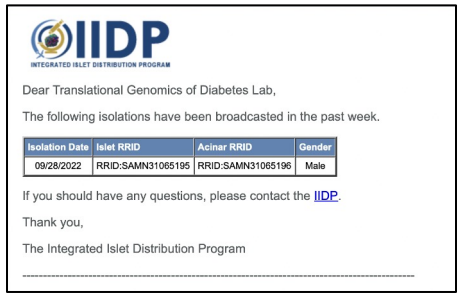
Translational Genomics of Diabetes Lab

INTEGRATED ISLET DISTRIBUTION PROGRAM
IIDP
HIGI
HUMAN ISLET GENOTYPING INITIATIVE

Pipeline Overview 1

Sample distribution, preparation and genotyping.

Swaraj Thaman



IIDP
INTEGRATED ISLET DISTRIBUTION PROGRAM

Dear Translational Genomics of Diabetes Lab,

The following isolations have been broadcasted in the past week.

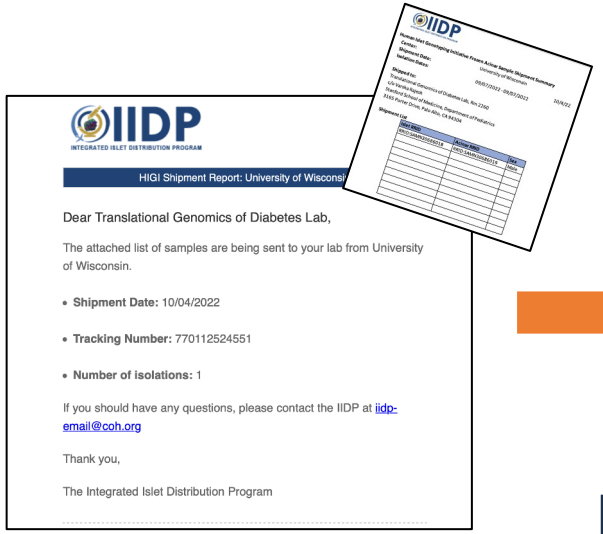
Isolation Date	Islet RRID	Acinar RRID	Gender
09/28/2022	RRID:SAMN31065195	RRID:SAMN31065196	Male

If you should have any questions, please contact the IIDP.

Thank you,

The Integrated Islet Distribution Program

Notification of islets Isolated



IIDP
INTEGRATED ISLET DISTRIBUTION PROGRAM

HIGI Shipment Report: University of Wisconsi

Dear Translational Genomics of Diabetes Lab,

The attached list of samples are being sent to your lab from University of Wisconsin.

- **Shipment Date:** 10/04/2022
- **Tracking Number:** 770112524551
- **Number of isolations:** 1

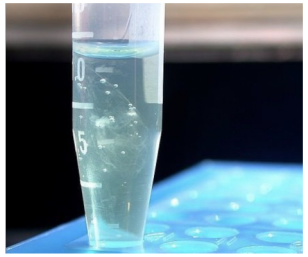
If you should have any questions, please contact the IIDP at iidp-email@coh.org

Thank you,

The Integrated Islet Distribution Program

Receive and barcode acinar samples from isolation centers

260/280: ~1.8
260/230: ~2.0 or above
Conc by qubit: >50 ng/uL
>500 ng total DNA for genotyping



- Resuspend acinar and extract DNA using Qiagen DNeasy Blood and Tissue Kit from a fraction of the acinar sample
- QC DNA sample by nanodrop and Qubit

Submit 15 mL of sample (diluted to 50 ng/mL) to the Stanford Functional Genomics Core for genotyping

protocols.io

Detailed protocols available on protocol.io

DNA Extraction Protocol
DNA Quantification Protocol
HIGI Processing Workflow

Translational Genomics of Diabetes Lab

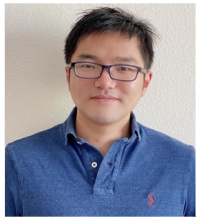
INTEGRATED ISLET DISTRIBUTION PROGRAM
IIDP
HIGI
HUMAN ISLET GENOTYPING INITIATIVE

Overview of Data Availability

- Data are released as specific "**data freezes or versions**" as we genotype the IIDP samples in batches, analyse and then re-analyse all available IIDP samples together. This allows us to present the results relative to one another so you can identify the donors at highest and lowest risk.
- We **currently have 2 data releases** which are available on the IIDP website. The **default setting** is our **most recent analysis (version 2)** and we display the key data available (genetic predicted ancestry, a genetic risk score for type 1 diabetes (also partitioned for HLA and non-HLA contributions) and a multi-ancestry weighted polygenic risk score for T2D which we present with options for both soft and hard clustering to explore the contributions of different underlying pathways.
- Users can also currently view the **previous release (version 1)** for the donors by selecting it from a drop down menu. Note that version 1 differs in the total number of donors in the analysis and the methods we have used to generate the scores which we have updated for version 2 to capture new variants identified in large multi-ancestry studies and updated clustering (partitioning) methods.
- The **default genetic data report available through the RDR** returns the information shown for each donor on the website but there are also **options to request an extended genetic report** which will return additional outputs we have calculated that some users might find useful.

Pipeline Overview 2

Genotyping, data cleaning and genetic variable generation.



Dr. Han Sun

Dr. Lu Zhang

CURRENT DEFAULT
VERSION 2

Genotype Array
Data QC

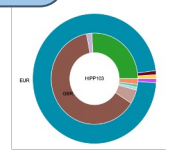
Repeat
failing
samples

Omni2.5Exome Array (2.5M variants)
V1.5 / V1.6 (from Oct' 24)
Minor Allele Frequency > 1%
SNP missing genotypes < 1%
Sample missing genotypes <3%
Reported and genetic sex matching

BioData CATALYST
TOPMed Imputation Server

NIH TOPMED
Imputation

Variants called to GRch38
Variants aligned to + strand
Imputation quality R^2>0.3
>41M high-quality variants output

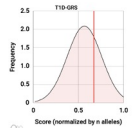


Genetic Ancestry
Prediction

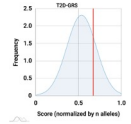
Machine Learning models
trained with 1000 Genomes Data

GAP

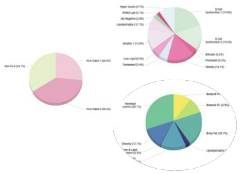
Type 1 Diabetes
Genetic Risk Score[1]



Type 2 Diabetes
Genetic Risk Score[2]



Partitioned Risk
Scores[1,2,,3]



DATA POWERED BY
IHIGI
HUMAN ISLET GENOTYPING INITIATIVE

PRS edm

Polygenic Risk Score extension for diabetes mellitus

References

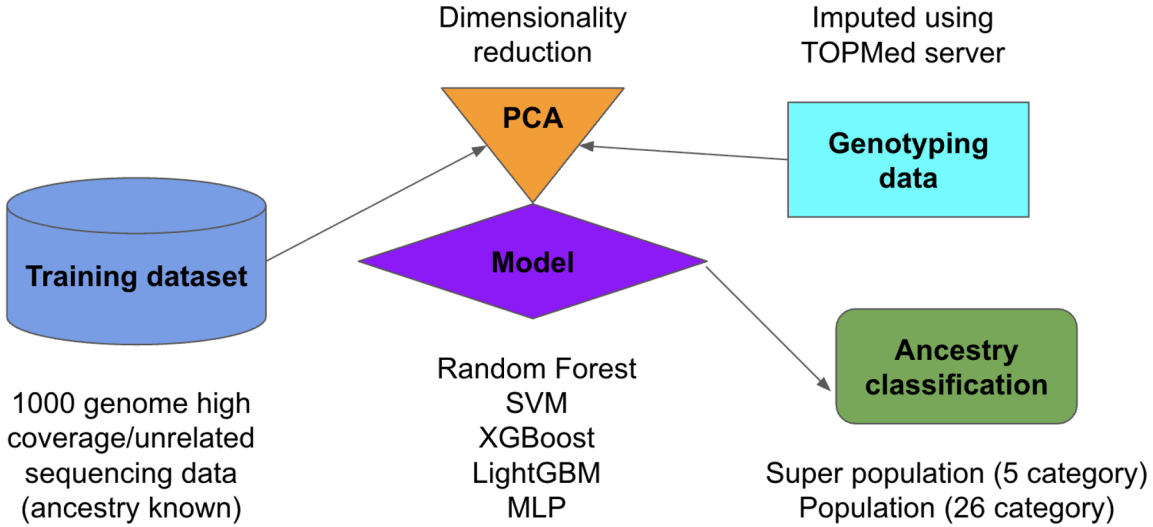
1. Luckett AM, Oram RA, Deutsch AJ, Ortega HI, Fraser DP, Ashok K, et al. Standardized Measurement of Type 1 Diabetes Polygenic Risk Across Multi-ancestry Population Cohorts. Diabetes Care. 2025;48(6):e81–e3. doi: 10.2337/dc25-0142. PubMed PMID: 40267362; PubMed Central PMCID: PMC12094190.
2. Suzuki K, Hatzikotoulas K, Southam L, Taylor HJ, Yin X, Lorenz KM, et al. Genetic drivers of heterogeneity in type 2 diabetes pathophysiology. Nature. 2024;627(8003):347–57. Epub 20240219. doi: 10.1038/s41586-024-07019-6. PubMed PMID: 38374256; PubMed Central PMCID: PMC10937372
3. Smith K, Deutsch AJ, McGrail C, Kim H, Hsu S, Huerta-Chagoya A, et al. Multi-ancestry polygenic mechanisms of type 2 diabetes. Nat Med. 2024;30(4):1065–74. Epub 20240305. doi: 10.1038/s41591-024-02865-3. PubMed PMID: 38443691; PubMed Central PMCID: PMC11175990.

Translational
Genomics
of Diabetes
Lab

INTEGRATED ISLET DISTRIBUTION PROGRAM
IIDP
IHIGI
HUMAN ISLET GENOTYPING INITIATIVE

Genetic Ancestry Interpretation

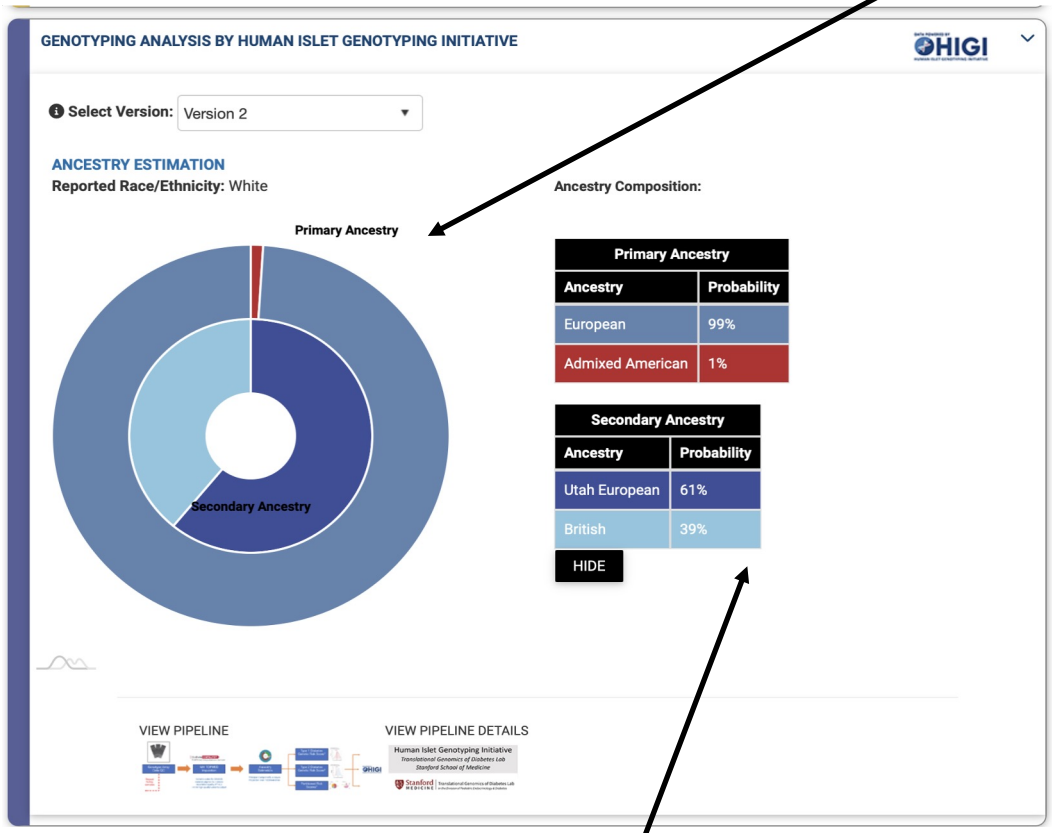
Overview of method



Details of the method available here

https://github.com/gloynlab/GeneticAncestry
DOI: 10.5281/zenodo.18157870

Primary ancestry probabilities are shown on the outer circle, aligned with 1000 Genomes primary populations.



Secondary ancestry (sub-ancestry) probabilities are shown on the inner circle

Translational Genomics of Diabetes Lab

References
Population descriptions - https://www.internationalgenome.org/data-portal/population

INTEGRATED ISLET DISTRIBUTION PROGRAM
IIDP HIGI
HUMAN ISLET GENOTYPING INITIATIVE

Genetic Risk for Type 1 Diabetes Interpretation

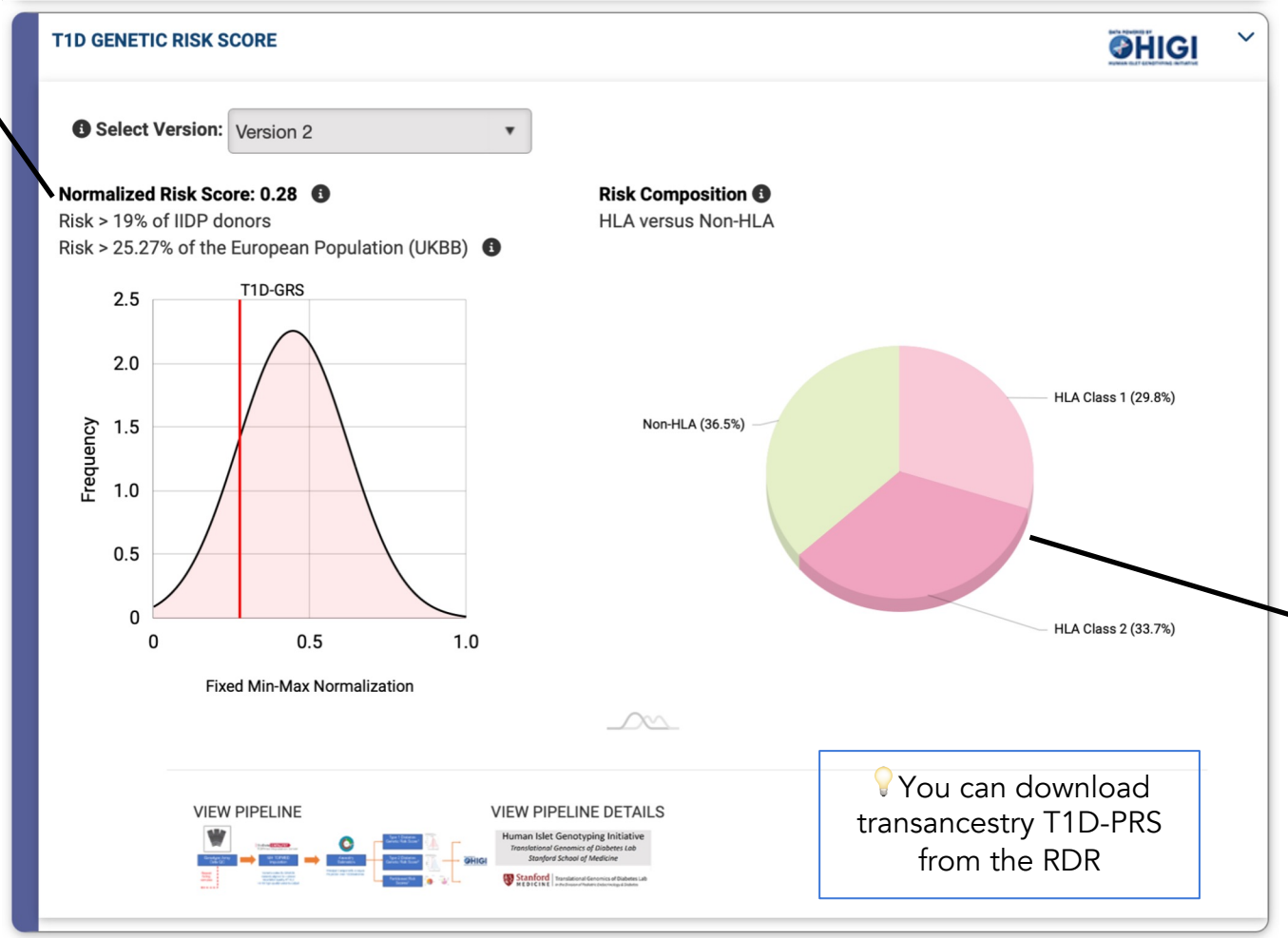
Risk score normalization applies a static normalization based on the theoretical minimum and maximum genetic risk contributions ranging from no risk alleles to all risk alleles present.

This procedure maps PRS values onto a 0–1 scale.

Risk is presented relative to the current data freeze of IIDP donors (e.g. >19%)

This is why the number could change from version to version as the IIDP resource grows

For European Donors T1D risk is also presented relative to the UK BioBank (e.g. >25.27%).



You can download transancestry T1D-PRS from the RDR

The donor's risk total genetic risk for T1D is broken down into risk from non-HLA and different HLA classes

DR3/DR4 haplotype can be downloaded from the RDR

Details of the method available here

https://github.com/sethsh7/PRSedm

DOI: https://zenodo.org/records/17903985

References

1. Sharp SA. sethsh7/PRSedm: v1.1.0. v1.1.0 ed: Zenodo; 2025.
2. Luckett AM, Oram RA, Deutsch AJ, Ortega HI, Fraser DP, Ashok K, et al. Standardized Measurement of Type 1 Diabetes Polygenic Risk Across Multiancestry Population Cohorts. Diabetes Care. 2025;48(6):e81–e3. doi: 10.2337/dc25-0142. PubMed PMID: 40267362; PubMed Central PMCID: PMC12094190.
3. Sharp SA, Rich SS, Wood AR, Jones SE, Beaumont RN, Harrison JW, et al. Development and Standardization of an Improved Type 1 Diabetes Genetic Risk Score for Use in Newborn Screening and Incident Diagnosis. Diabetes Care. 2019;42(2):200–7. doi: 10.2337/dc18-1785. PubMed PMID: 30655379; PubMed Central PMCID: PMC6341291.

Translational Genomics of Diabetes Lab

INTEGRATED ISLET DISTRIBUTION PROGRAM IIDP HIGI HUMAN ISLET GENOTYPING INITIATIVE

Genetic Risk for Type 2 Diabetes Interpretation - Part I

The score is now based on 1,232,226 variants in the full genome wide multi-ancestry analysis from Suzuki et al 2024 [5].

Risk score normalization applies a static normalization based on the theoretical minimum and maximum genetic risk contributions ranging from no risk alleles to all risk alleles present.

This procedure maps PRS values onto a 0–1 scale.

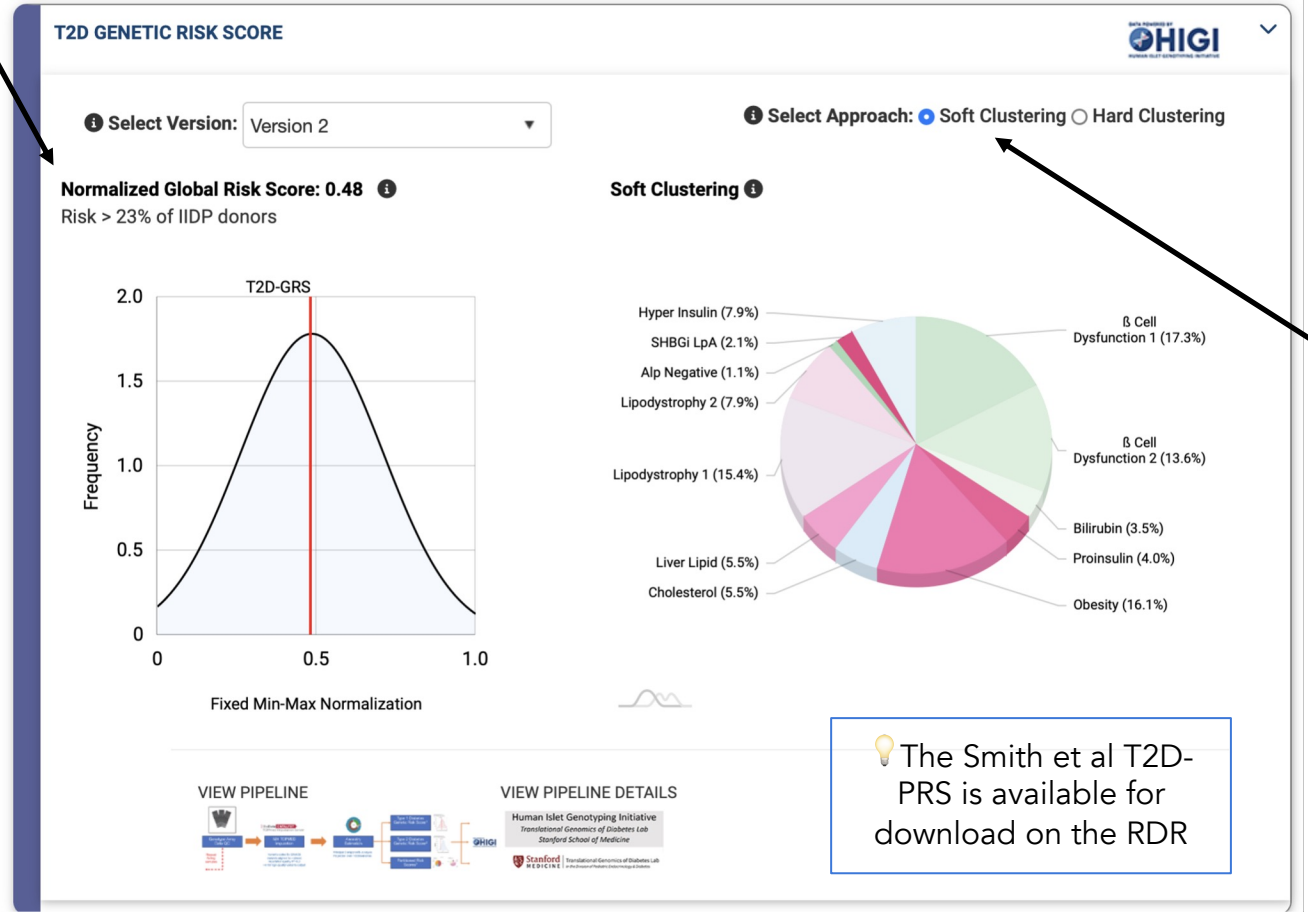
Risk is presented relative to the current data freeze of IIDP donors (e.g. >23%)

This is why the number could change from version to version as the IIDP resource grows

Details of the method available here

https://github.com/sethsh7/PRSedm

DOI: https://zenodo.org/records/17903985



The Smith et al T2D-PRS is available for download on the RDR

We offer IIDP users the choice of hard or soft clustering for partitioned genetic risk scores for T2D diabetes using the latest underlying T2D-GWAS data set available [5].

Soft Clustering: These contain multi-ancestry PRS and pPS constructed using T2D index variants defined in the soft clustering framework from Smith et al [4].

Under the soft clustering framework, genetic variants can contribute to multiple clusters with different weights, allowing for overlap and continuity across genetic mechanisms. The corresponding pPS are computed specifically under this soft clustering framework

Translational Genomics of Diabetes Lab

References

4. Smith K, Deutsch AJ, McGrail C, Kim H, Hsu S, Huerta-Chagoya A, et al. Multi-ancestry polygenic mechanisms of type 2 diabetes. Nat Med. 2024;30(4):1065–74. Epub 20240305. doi: 10.1038/s41591-024-02865-3. PubMed PMID: 38443691; PubMed Central PMCID: PMC11175990.

5.Suzuki K, Hatzikotoulas K, Southam L, Taylor HJ, Yin X, Lorenz KM, et al. Genetic drivers of heterogeneity in type 2 diabetes pathophysiology. Nature. 2024;627(8003):347–57. Epub 20240219. doi: 10.1038/s41586-024-07019-6. PubMed PMID: 38374256; PubMed Central PMCID: PMC10937372.

INTEGRATED ISLET DISTRIBUTION PROGRAM
IIDP
HIGI
HUMAN ISLET GENOTYPING INITIATIVE

Genetic Risk for Type 2 Diabetes Interpretation - Part II

The score is now based on 1,232,226 variants in the full genome wide multi-ancestry analysis from Suzuki et al 2024 [5].

Risk score normalization applies a static normalization based on the theoretical minimum and maximum genetic risk contributions ranging from no risk alleles to all risk alleles present.

This procedure maps PRS values onto a 0–1 scale.

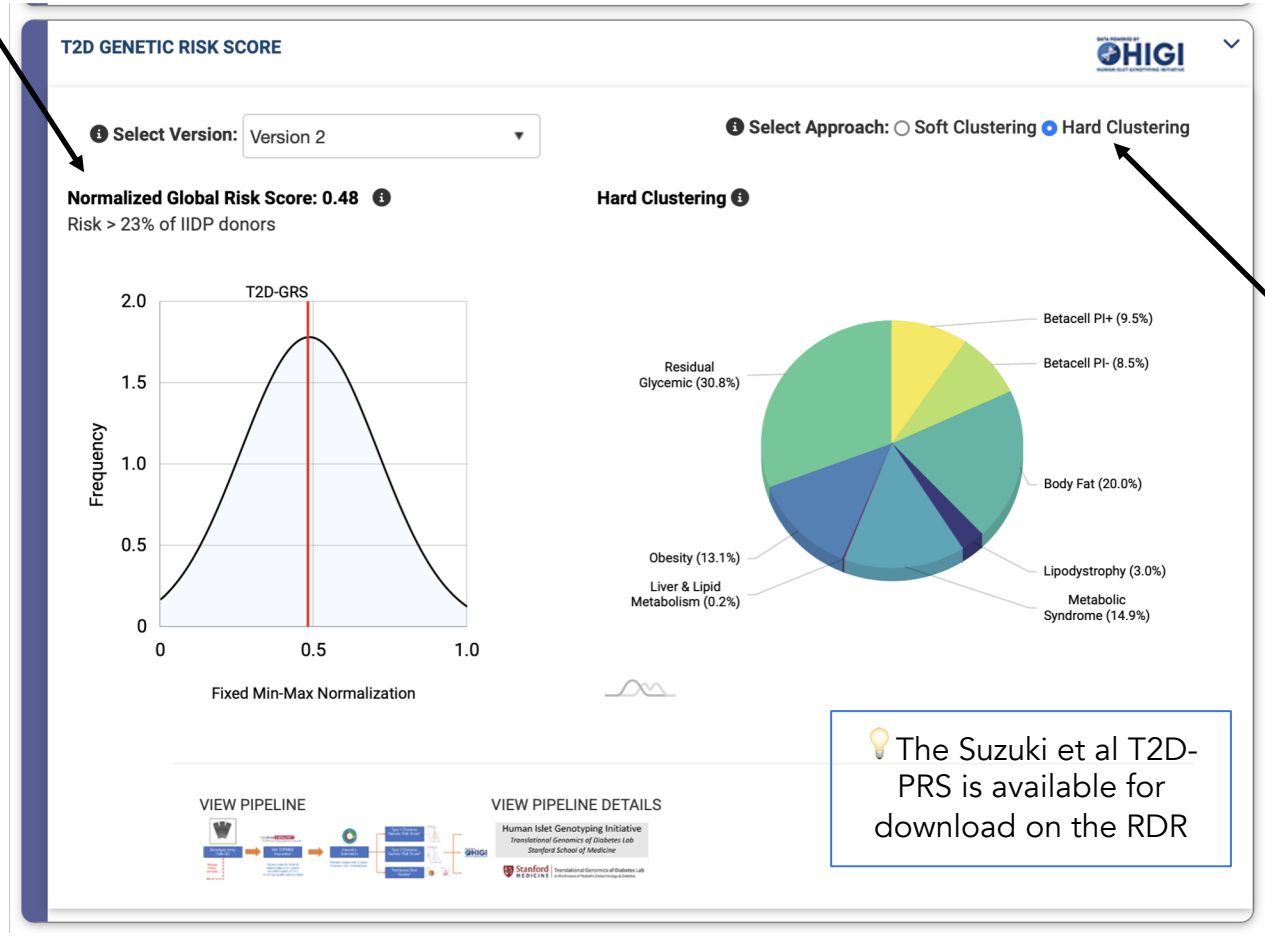
Risk is presented relative to the current data freeze of IIDP donors (e.g. >23%)

This is why the number could change from version to version as the IIDP resource grows

Details of the method available here

https://github.com/sethsh7/PRSedm

DOI: https://zenodo.org/records/17903985



The Suzuki et al T2D-PRS is available for download on the RDR

We offer IIDP users the choice of hard or soft clustering for partitioned genetic risk scores for T2D diabetes using the latest underlying T2D-GWAS data set available [5].

Hard Clustering: These contain multi-ancestry PRS and pPS constructed using T2D index variants defined in the hard clustering framework from Suzuki et al [5].

Under the hard clustering framework, each genetic variant is uniquely assigned to a single cluster, representing relatively discrete genetic mechanisms. The corresponding pPS are computed within this same clustering framework and are provided alongside the PRS.

Translational Genomics of Diabetes Lab

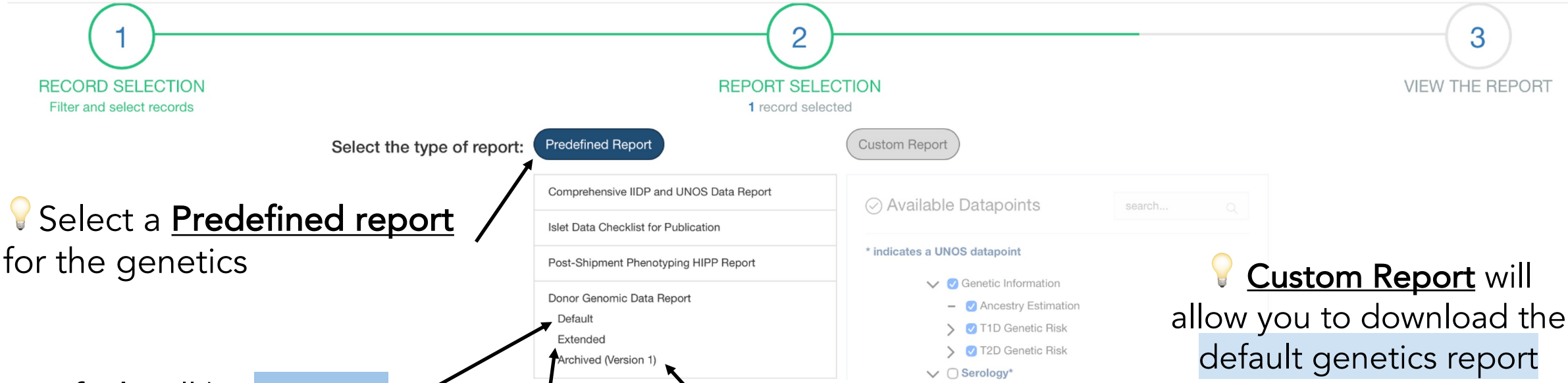
References

4. Smith K, Deutsch AJ, McGrail C, Kim H, Hsu S, Huerta-Chagoya A, et al. Multi-ancestry polygenic mechanisms of type 2 diabetes. Nat Med. 2024;30(4):1065–74. Epub 20240305. doi: 10.1038/s41591-024-02865-3. PubMed PMID: 38443691; PubMed Central PMCID: PMC11175990.

5.Suzuki K, Hatzikotoulas K, Southam L, Taylor HJ, Yin X, Lorenz KM, et al. Genetic drivers of heterogeneity in type 2 diabetes pathophysiology. Nature. 2024;627(8003):347–57. Epub 20240219. doi: 10.1038/s41586-024-07019-6. PubMed PMID: 38374256; PubMed Central PMCID: PMC10937372.

INTEGRATED ISLET DISTRIBUTION PROGRAM
IIDP
HIGI
HUMAN ISLET GENOTYPING INITIATIVE

Getting more from the RDR



💡 Select a **Predefined report** for the genetics

Default will be version 2 as seen on the web browser

Extended will be version 2 all available outputs (e.g. T2D-GRS across different ancestries, DR3/DR4 etc)

Archived will be version 1 all available outputs from the previous release

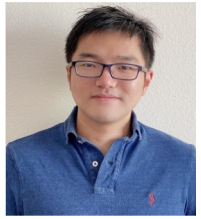
💡 **Custom Report** will allow you to download the default genetics report with other donor data points

Translational Genomics of Diabetes Lab

ARCHIVED Pipeline Overview

Genotyping, data cleaning and genetic variable generation.

Dr. Seth Sharp



Dr. Han Sun

VERSION 1
ARCHIVED PIPELINE

Genotype Array Data QC

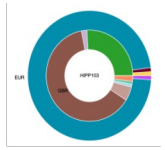
Repeat failing samples

Omni2.5Exome Array (2.5M variants)
V1.5 / V1.6 (from Oct' 24)
Minor Allele Frequency > 1%
SNP missing genotypes < 1%
Sample missing genotypes <3%
Reported and genetic sex matching

BioData CATALYST
TOPMed Imputation Server

NIH TOPMED Imputation

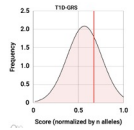
Variants called to GRch38
Variants aligned to + strand
Imputation quality R^2>0.3
>41M high-quality variants output



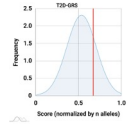
Genetic Ancestry Prediction

Machine Learning models trained with 1000 Genomes Data

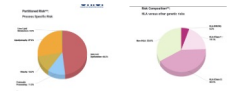
Type 1 Diabetes Genetic Risk Score[1]



Type 2 Diabetes Genetic Risk Score[2]



Partitioned Risk Scores[3]



DATA POWERED BY HIGI
HUMAN ISLET GENOTYPING INITIATIVE

References
1. Sharp, Seth A et al. "Development and Standardization of an Improved Type 1 Diabetes Genetic Risk Score for Use in Newborn Screening and Incident Diagnosis." *Diabetes care* vol. 42,2 (2019): 200-207. doi:10.2337/dc18-1785
2. Mahajan, Anubha et al. "Multi-ancestry genetic study of type 2 diabetes highlights the power of diverse populations for discovery and translation." *Nature genetics* vol. 54,5 (2022): 560-572. doi:10.1038/s41588-022-01058-3
3. DiCorpo, Daniel et al. "Type 2 Diabetes Partitioned Polygenic Scores Associate With Disease Outcomes in 454,193 Individuals Across 13 Cohorts." *Diabetes care* vol. 45,3 (2022): 674-683. doi:10.2337/dc21-1395

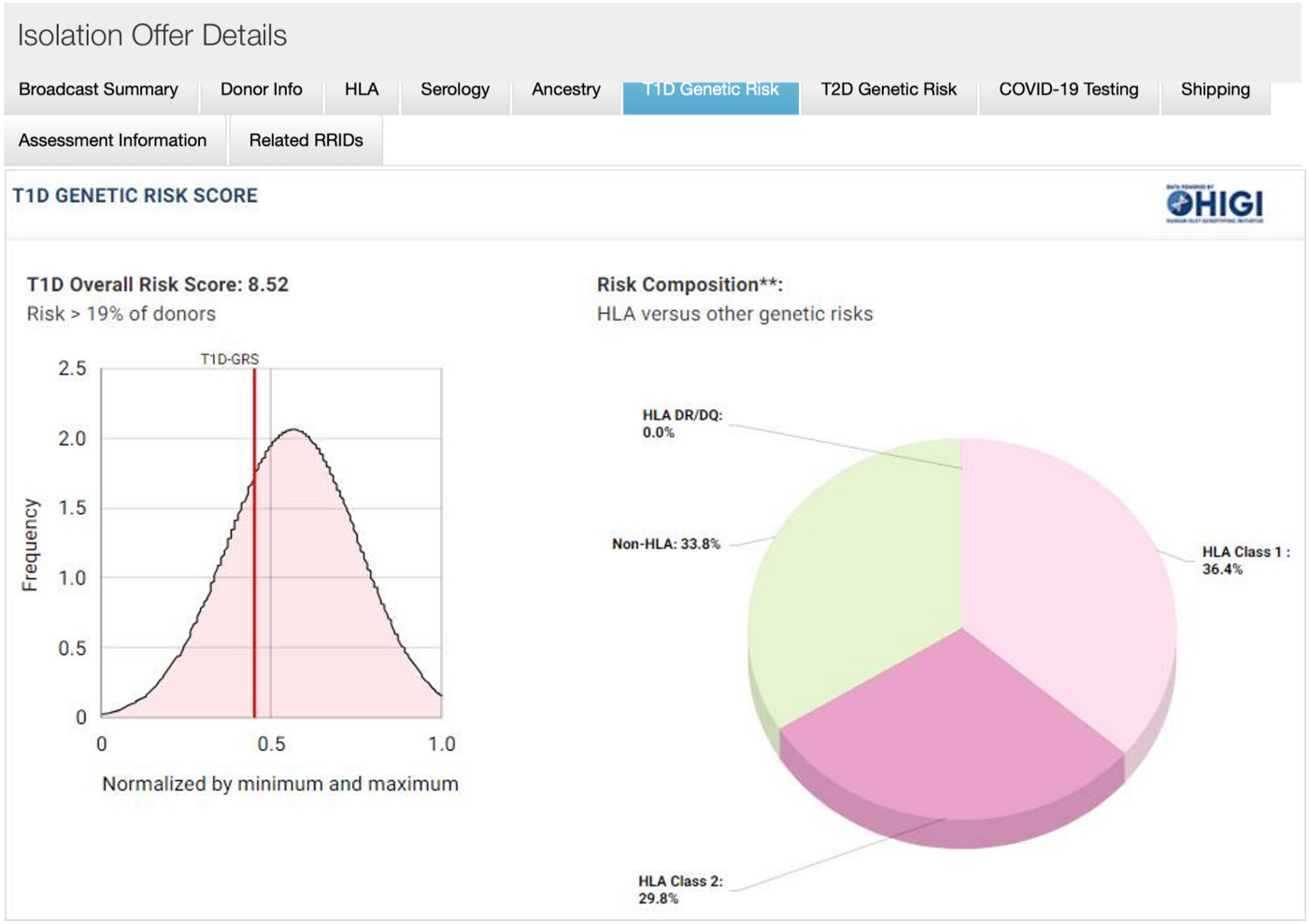
Translational Genomics of Diabetes Lab

INTEGRATED ISLET DISTRIBUTION PROGRAM
IIDP
HIGI
HUMAN ISLET GENOTYPING INITIATIVE

Interpretation

VERSION 1
ARCHIVED PIPELINE

Overall genetic risk is normalized to minimum and maximum possible value



The donor's risk total genetic risk for T1D is broken down into risk from non-HLA and different HLA classes

Non-HLA
HLA DR/DQ
HLA Class I
HLA Class II

References

1. Sharp, Seth A et al. "Development and Standardization of an Improved Type 1 Diabetes Genetic Risk Score for Use in Newborn Screening and Incident Diagnosis." *Diabetes care* vol. 42,2 (2019): 200-207. doi:10.2337/dc18-1785

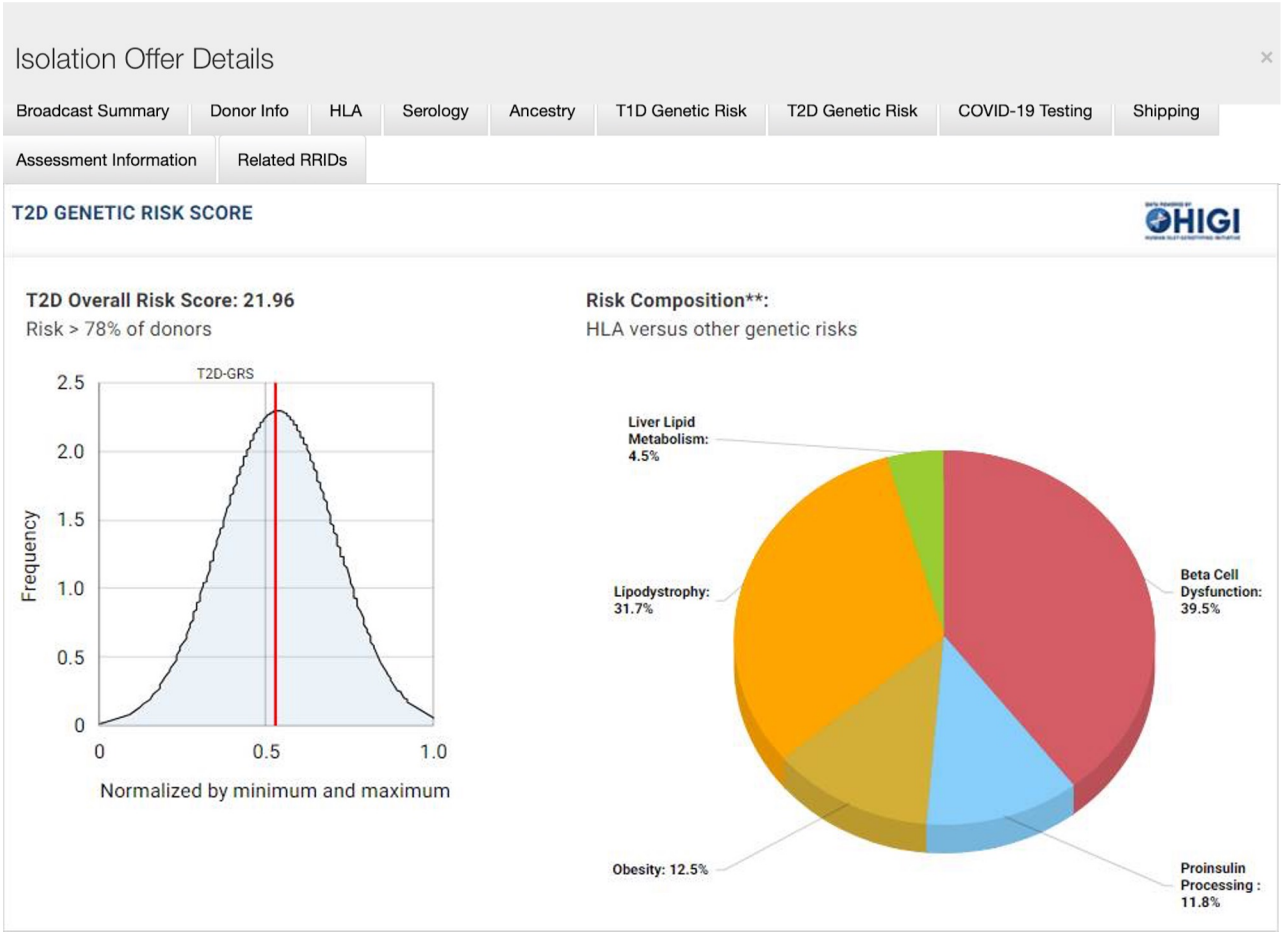
Translational Genomics of Diabetes Lab

INTEGRATED ISLET DISTRIBUTION PROGRAM
IIDP
HIGI
HUMAN ISLET GENOTYPING INITIATIVE

Interpretation

VERSION 1
ARCHIVED PIPELINE

The overall genetic risk for T2D is presented in relation to all other donors in IIDP



The donor's risk total genetic risk for T2D is broken down into risk for variants which contribute to different physiological processes representing different tissues of action and pathophysiology

Liver Lipid Metabolism
Lipodystrophy
Obesity
Proinsulin processing
Beta Cell dysfunction

References

3. DiCorpo, Daniel et al. "Type 2 Diabetes Partitioned Polygenic Scores Associate With Disease Outcomes in 454,193 Individuals Across 13 Cohorts." *Diabetes care* vol. 45,3 (2022): 674-683. doi:10.2337/dc21-1395

Translational Genomics of Diabetes Lab

INTEGRATED ISLET DISTRIBUTION PROGRAM
IIDP
HIGI
HUMAN ISLET GENOTYPING INITIATIVE

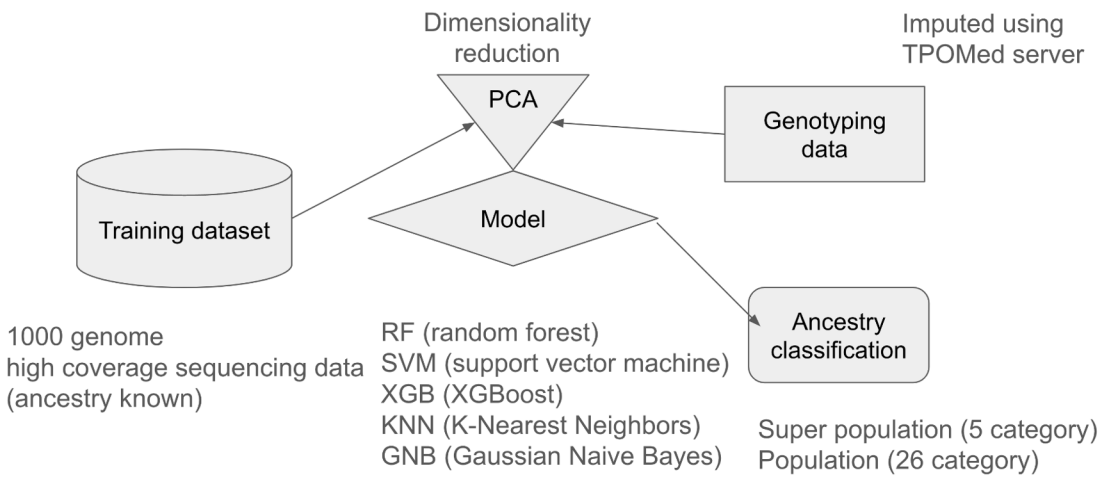
Interpretation

VERSION 1
ARCHIVED PIPELINE

Machine learning prediction of ancestry from the genotyping data

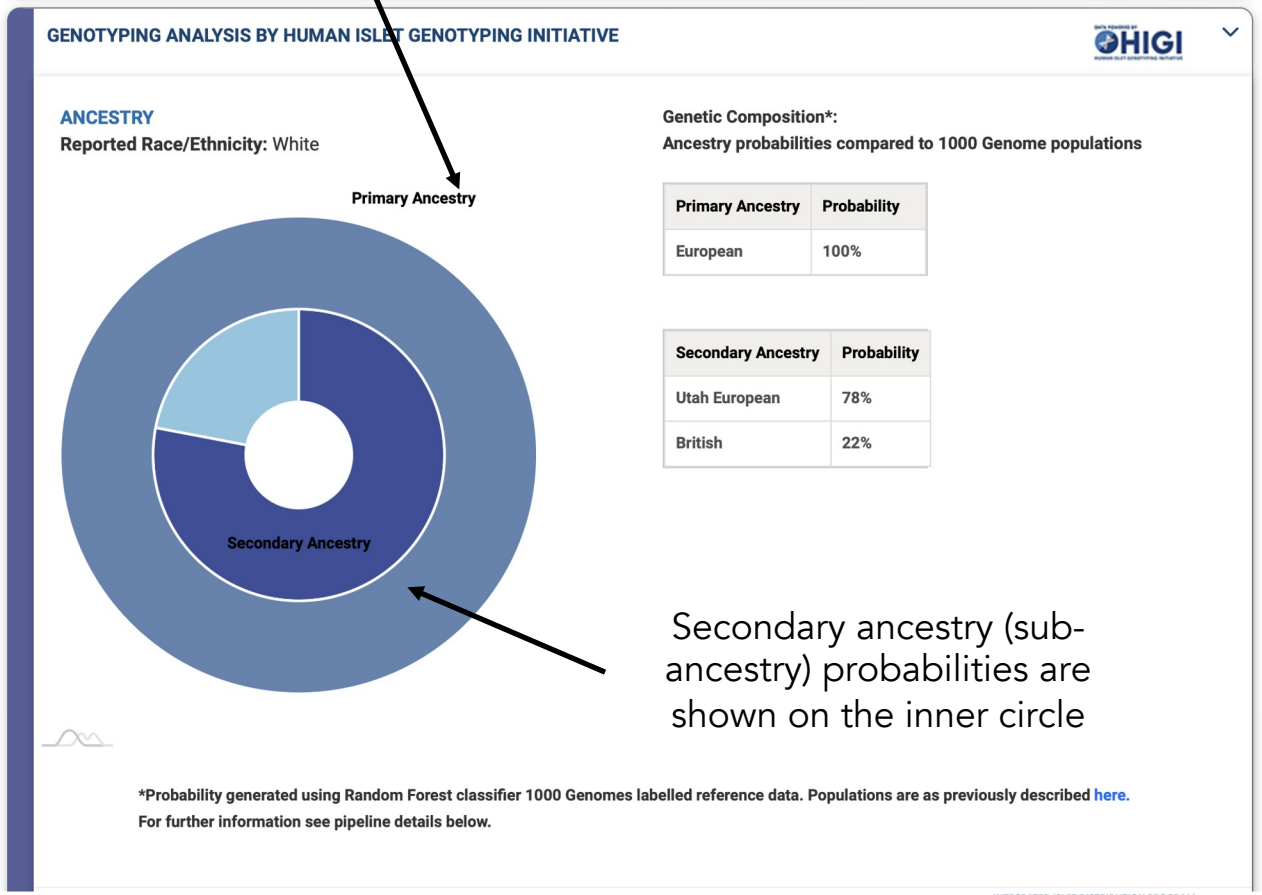
Han Sun, PhD, Gloyn Lab, Stanford University

Outline: the pipeline



https://github.com/gloynlab/GeneticAncestry

Primary ancestry probabilities are shown on the outer circle, aligned with 1000 Genomes primary populations.



Secondary ancestry (sub-ancestry) probabilities are shown on the inner circle

Translational Genomics of Diabetes Lab

References
Population descriptions - https://www.internationalgenome.org/data-portal/population

INTEGRATED ISLET DISTRIBUTION PROGRAM
IIDP
HIGI
HUMAN ISLET GENOTYPING INITIATIVE