

The Polygenic Score (PGS) Catalog

Resource & submission overview

Sam Lambert, PhD

Assistant Professor, University of Cambridge

Visiting Researcher, EMBL-EBI

Polygenic Scores: Terminology

Polygenic Score (PGS)

A single value that quantifies an individual's genetic predisposition to a trait

Polygenic Risk Score (PRS)

A subset of PGS that is used to estimate the risk of disease or other clinically relevant outcomes (binary or discrete)

Other Terms

Genome-wide polygenic score (GPS), Genetic/genomic risk score (GRS), Genetic/genomic score (GS)

PGS Development & Evaluation is Highly Variable

Often multiple scores for the same trait, developed using overlapping datasets.

Metrics are **not directly comparable between** scores and studies

Reduced accuracy in populations dissimilar to development cohorts (e.g. across ancestry groups) Martin et al. (2017, 2019)

Variant information needed to apply a PGS to new samples is often **not shared**

Method / Publication	# of SNPs	Population Ancestry		Metrics: Evaluation
		Training	Evaluation	
<i>Lassosum</i> Elliott (2020)	40,079	Cardiogram GWAS	UKB 98% European, 2% other ancestries	<i>HR</i> = 1.32 [1.30 - 1.34]
<i>metaGRS</i> Inouye (2018)	1,745,180		UKB 95% European, 5% other ancestries	<i>HR</i> = 1.71 [1.68 - 1.73]
<i>LDpred</i> Khera (2018)	6,630,150		UKB European	AUROC = 0.81
<i>GRS</i> Tada (2015)	50	Europeans	MDC Europeans	<i>HR</i> = 1.23 [1.18 - 1.28]

The PGS Catalog

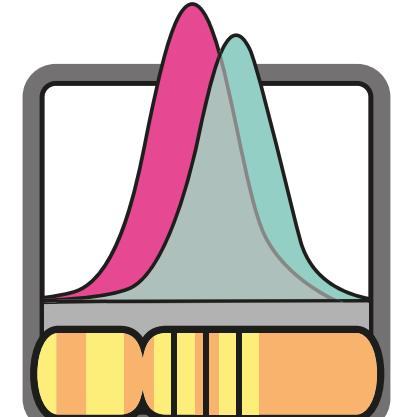
www.PGSCatalog.org

Open sharing of PGS would go beyond good reporting to enable downstream re-use and comparisons of scores.

- Resources like the GWAS Catalog have been transformative in their ability to organise and share usable genetic data.

An open database of **polygenic scores (PGS)** and the relevant metadata required for accurate application and evaluation

Developed with the
NHGRI-EBI GWAS Catalog



The screenshot shows the PGS Catalog website at pgscatalog.org. The header includes the PGS Catalog logo, navigation links for Home, Browse, Downloads, and Documentation, and a note about the latest release (April 18, 2024). Below the header is a search bar and a section for the 'Available tool: pgsc_calc'. A pink banner highlights this tool as a reproducible workflow for calculating PGS and custom polygenic scores. The main content area is titled 'Explore the Data' and features three teal boxes: 'Polygenic Scores' (4,723), 'Traits' (652), and 'Publications' (607). On the right side, there are sections for 'What is a PGS?' and 'About the PGS Catalog', along with a 'Submit a PGS' button.

The Polygenic Score (PGS) Catalog

An open database of polygenic scores and the relevant metadata required for accurate application and evaluation.

Search the PGS Catalog

Examples: breast cancer, glaucoma, BMI, EFO_0001645

Available tool: **pgsc_calc**

A reproducible workflow to calculate both PGS Catalog and custom polygenic scores. [See more information](#)

Explore the Data

In the current PGS Catalog you can **browse** the scores and metadata through the following categories:

Polygenic Scores **4,723**

Traits **652**

Publications **607**

Submit a PGS

What is a **PGS**?

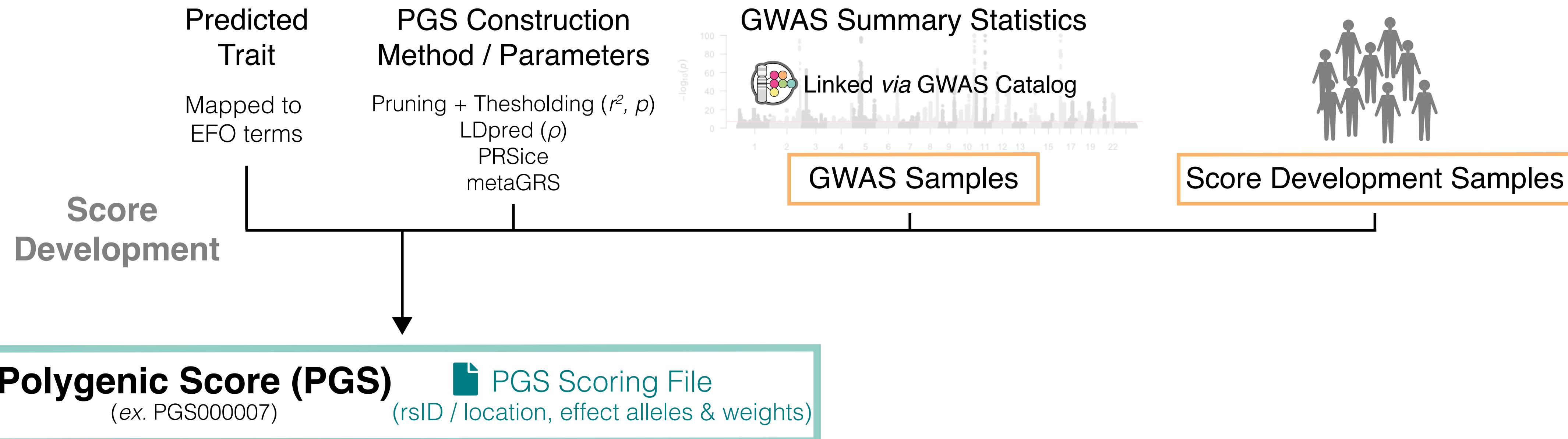
About the **PGS Catalog**

The **PGS Catalog** is an open database of published **polygenic scores (PGS)**. Each **PGS** in the **Catalog** is consistently annotated with relevant metadata; including scoring files (variants, effect alleles/weights), annotations of how the **PGS** was developed and applied, and evaluations of their predictive performance.

PGS nomenclature is heterogeneous: they can also be referred to as **genetic scores** or **genomic scores**, and as **polygenic risk scores (PRS)** or **genomic risk scores (GRS)** if they predict a discrete phenotype, such as a disease.

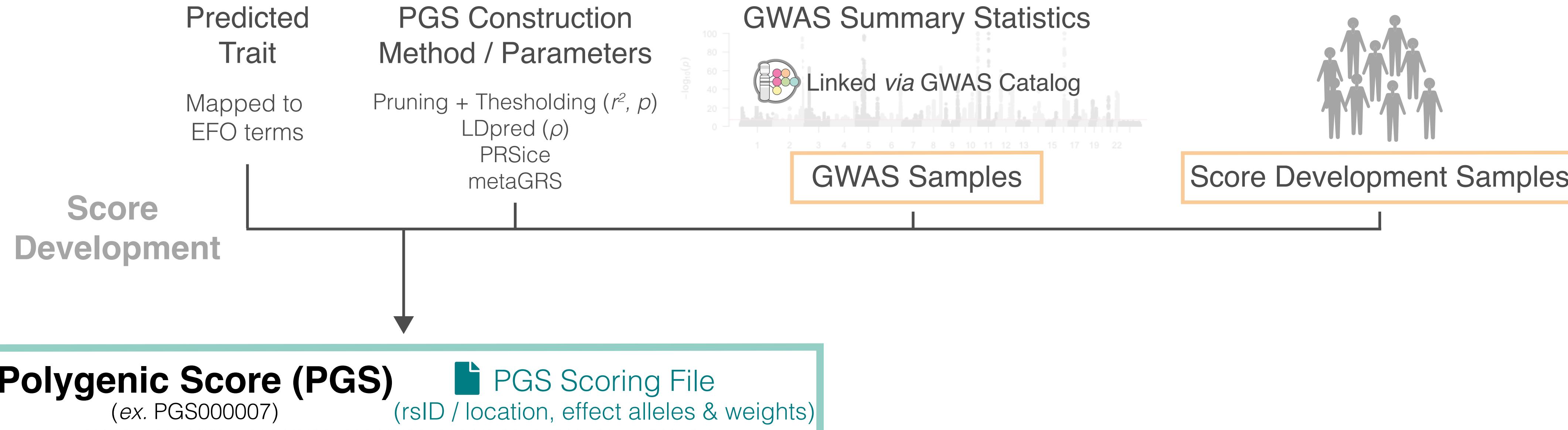
Lambert et al. The Polygenic Score Catalog as an open database for reproducibility and systematic evaluation. *Nature Genetics* (2021)

PGS Catalog (meta)data: Scores



Samples are consistently described with:
Sample size (case/control)
Ancestry
Age, sex, followup time
Cohort(s)

PGS Catalog (meta)data: Scoring files



```
PGS CATALOG SCORING FILE - see www.pgscatalog.org/downloads/#dl_ftp for additional information
## POLYGENIC SCORE (PGS) INFORMATION
# PGS ID = PGS000018
# Reported Trait = Coronary artery disease
# Original Genome Build = hg19
# Number of Variants = 1745180
## SOURCE INFORMATION
# PGP ID = PGP000007
# Citation = Inouye M et al. J Am Coll Cardiol (2018). doi:10.1016/j.jacc.2018.07.079
rsID chr_name chr_position effect_allele reference_allele effect_weight
rs2843152 1 2245570 G C -2.76009e-02
rs35465346 1 22132518 G A 2.39340e-02
rs28470722 1 38386727 G A -1.74935e-02
rs11206510 1 55496039 T C 2.93005e-02
rs9970807 1 56965664 C T 4.70027e-02
rs61772626 1 57015668 A G -2.71202e-02
rs7528419 1 109817192 A G 2.91912e-02
rs1277930 1 109822143 A G 2.60105e-02
rs11102000 1 110298166 G C 2.45969e-02
rs11810571 1 151762308 G C 2.09215e-02
rs6689306 1 154395946 G A -1.97906e-02
rs72702224 1 154911689 G A -2.81310e-02
rs3738591 1 155764808 C G 4.23731e-02
rs2789422 1 159892088 G A 1.67160e-02
rs2820315 1 201872264 T C 1.84269e-02
rs67180937 1 222823743 G T 3.33613e-02
rs75082168 1 223132028 T A -5.30459e-02
rs16986953 2 19942473 G A -3.63843e-02
rs312983 2 21378580 C A 2.49857e-02
rs13420649 2 43463637 T C -2.68940e-02
rs11126366 2 72554292 C G 3.15943e-02
rs11126367 2 72167844 S T 3.28901e-02
```

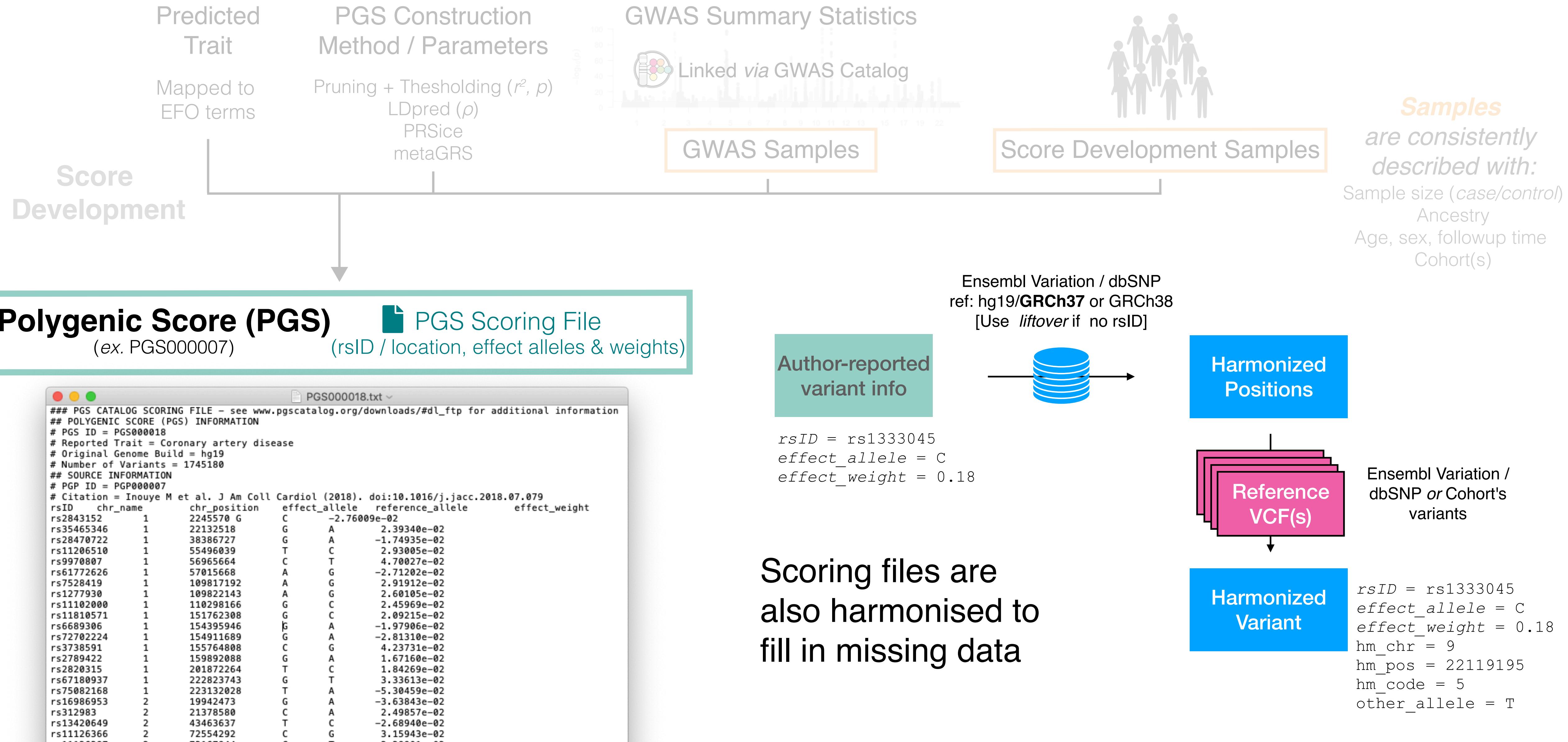
Consistently formatted scoring files with author-reported variant information are provided to **enable PGS calculation on new samples.**

Minimal specification:

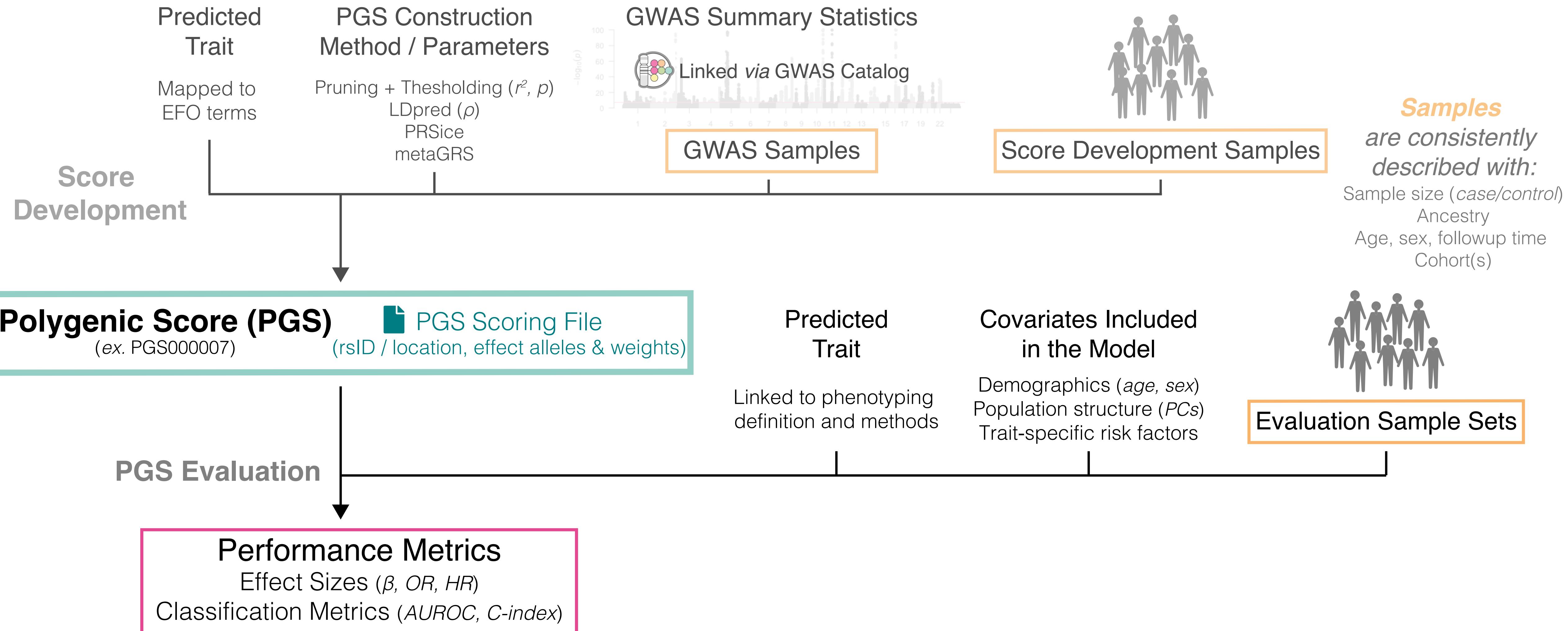
- *rsID or chromosomal location*
- *Effect allele and weight*

Samples are consistently described with:
Sample size (case/control)
Ancestry
Age, sex, followup time
Cohort(s)

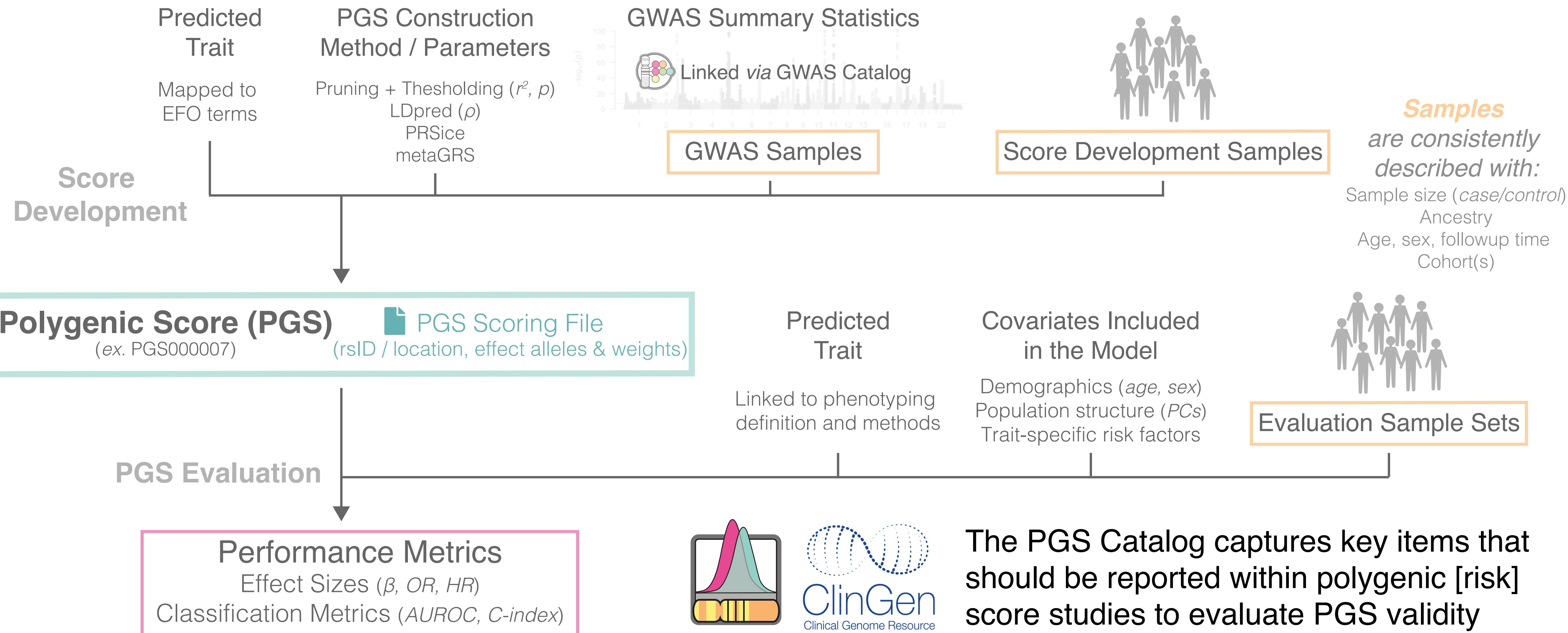
PGS Catalog (meta)data: Scoring files



PGS Catalog (meta)data: performance



PGS Catalog (meta)data: Reporting Standards



Polygenic Risk Score Reporting Standards (PRS-RS)
Wand, Lambert *et al. Nature* (2021)

PGSCatalog.org: Access

You can browse the **Scores**, **Publications** and **Traits** in the Catalog on the website

Data in the catalog can also be accessed through:

Bulk downloads of metadata describing PGS development and evaluation

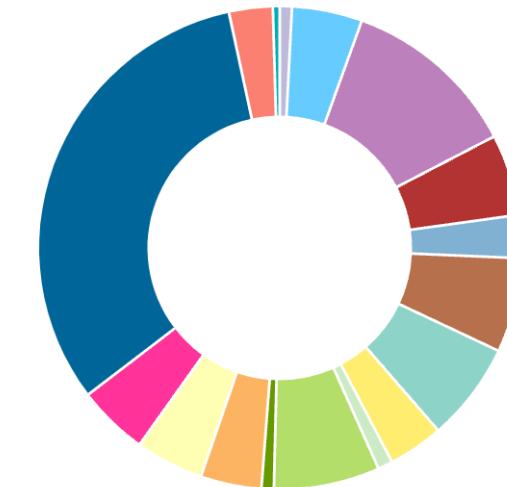
A **REST API** to programmatically access and query the Catalog

PGS Scoring Files from our **FTP**

Traits

Browse PGS by Trait Category 

 Reset view



Biological process	39 PGS
Body measurement	273 PGS
Cancer	702 PGS
Cardiovascular disease	317 PGS
Cardiovascular measurement	163 PGS
Digestive system disorder	370 PGS
Hematological measurement	388 PGS
Immune system disorder	212 PGS
Inflammatory measurement	50 PGS
Lipid or lipoprotein measurement	416 PGS
Liver enzyme measurement	40 PGS
Metabolic disorder	235 PGS
Neurological disorder	265 PGS
Other disease	275 PGS
Other measurement	1913 PGS
Other trait	166 PGS
Sex-specific PGS	18 PGS

Abdominal Aortic Aneurysm	6 PGS
Abnormal EKG	1 PGS
acute myocardial infarction	2 PGS
angina pectoris	6 PGS
aortic valve disease	1 PGS
atrial fibrillation	30 PGS
atrial flutter	4 PGS
brain aneurysm	6 PGS
Brugada syndrome	2 PGS
cardiomyopathy	1 PGS
cardiovascular disease	13 PGS
cerebrovascular disorder	1 PGS
congestive heart failure	2 PGS
cor pulmonale	2 PGS
coronary artery disease	58 PGS
coronary atherosclerosis	2 PGS
diabetic retinopathy	4 PGS

Traits

Trait (ontology term label)	Trait Identifier (ontology ID)	Trait Category	Number of Related PGS
Abdominal Aortic Aneurysm	EFO_0004214	Cardiovascular disease	6
Abnormal EKG	HP_0003115	Cardiovascular disease	1
acute myocardial infarction	EFO_0008583	Cardiovascular disease	2
angina pectoris	EFO_0003913	Cardiovascular disease	6
aortic valve disease	EFO_0009531	Cardiovascular disease	1
atrial fibrillation	EFO_0000275	Cardiovascular disease	30
atrial flutter	EFO_0003911	Cardiovascular disease	4
brain aneurysm	EFO_0003870	Cardiovascular disease	6
Brugada syndrome	MONDO_0015263	Cardiovascular disease	2

Browsing ancestry information

Trait: cardiovascular disease

Experimental Factor Ontology (EFO) Information	
Identifier	EFO_0000319
Description	A disease involving the cardiovascular system.
Trait category	Cardiovascular disease
Synonyms	76 synonyms
Mapped terms	26 mapped terms
Child trait(s)	38 child traits

[View in NHGRI-EBI GWAS Catalog](#)

Associated Polygenic Score(s)

Filter PGS by Participant Ancestry [i](#)

Individuals included in: [PGS Evaluation \[E\]](#)

List of ancestries includes: [African](#)

G - Source of Variant Associations (GWAS)
D - Score Development/Training
E - PGS Evaluation

Display options:

- Show European ancestry data [i](#)
- Show only Multi-ancestry data [i](#)

Ancestry legend [i](#)

Multi-ancestry (including European)	European
Multi-ancestry (excluding European)	Greater Middle Eastern
African	Hispanic or Latin American
East Asian	Additional Diverse Ancestries
South Asian	Not Reported
Additional Asian Ancestries	

Note: This table shows all PGS for "cardiovascular disease" and any child terms of this trait in the EFO hierarchy by default.

Include PGS Score(s) for child traits

Search [+ !\[\]\(0fad87bef43f637224b259f1cee77186_img.jpg\) !\[\]\(bc6c810ddcadd995a220ca721047856d_img.jpg\)](#)

Polygenic Score ID & Name	PGS Publication ID (PGP)	Reported Trait	Mapped Trait(s) (Ontology)	Number of Variants	Ancestry distribution
					GWAS Dev Eval
PGS000011 (GRS50)	PGP00004 » Tada H et al. Eur Heart J (2015)	Coronary artery disease	coronary artery disease	50	
PGS000013 (GPS_CAD)	PGP00006 » Khera AV et al. Nat Genet (2018)	Coronary artery disease	coronary artery disease	6,630,150	
PGS000016 (GPS_AF)	PGP00006 » Khera AV et al. Nat Genet (2018)	Atrial fibrillation	atrial fibrillation	6,730,541	
PGS000018 (metaGRS_CAD)	PGP00007 » Inouye M et al. J Am Coll Cardiol (2018)	Coronary artery disease	coronary artery disease	1,745,179	
PGS000039 (metaGRS_ischaemicstroke)	PGP000027 » Abraham G et al. Nat Commun (2019)	Ischemic stroke	stroke, Ischemic stroke	3,225,583	

Filter PGS by Participant Ancestry [i](#)

Individuals included in: [Any Stage \[G, D, E\]](#)

G - Source of Variant Associations (GWAS)
D - Score Development/Training

List of ancestries includes: [--](#)

Display options:

- Show European ancestry data [i](#)
- Show only Multi-ancestry data [i](#)

Ancestry legend [i](#)

Multi-ancestry (including European)	European
Multi-ancestry (excluding European)	Greater Middle Eastern
African	Hispanic or Latin American
East Asian	Additional Diverse Ancestries
South Asian	Not Reported
Additional Asian Ancestries	

This table shows all PGS for "cardiovascular disease" and any child terms of this trait in the EFO hierarchy by default.

Score(s) for child traits

Search [+ !\[\]\(6dc81516652938c9144bbe5fc90f94ff_img.jpg\) !\[\]\(21a86f37ba901ff8293159aebe8c34ae_img.jpg\)](#)

Name	PGS Publication ID (PGP)	Reported Trait	Mapped Trait(s) (Ontology)	Number of Variants	Ancestry distribution	Scoring File (FTP L)
					GWAS Dev Eval	
PGP000003 » Mega JL et al. Lancet (2015)	PGP000003	Coronary heart disease	coronary artery disease	27	 	Terms/Licenses
PGP000004 » Tada H et al. Eur Heart J (2015)	PGP000004	Coronary artery disease	coronary artery disease	50	 	Terms/Licenses
PGP000005 » Abraham G et al. Eur Heart J (2016)	PGP000005	Coronary artery disease	coronary artery disease	49,310	 	Terms/Licenses
PGP000006 » Khera AV et al. Nat Genet (2018)	PGP000006	Coronary artery disease	coronary artery disease	6,630,150	 	Terms/Licenses
PGP000006 » Khera AV et al. Nat Genet (2018)	PGP000006	Atrial fibrillation	atrial fibrillation	6,730,541	 	Terms/Licenses
PGP000007 » Inouye M et al. J Am Coll Cardiol (2018)	PGP000007	Coronary artery disease	coronary artery disease	1,745,179	 	Terms/Licenses
PGP000009 » Abraham G et al. Nat Commun (2019)	PGP000009	Coronary artery				

Browsing PGS metadata

PGS Catalog | Home | Browse | Downloads | Documentation | Terms and Licenses | Search... | breast cancer, glaucoma, BMI, EFO_0001645

Predicted Trait

Reported Trait	Coronary artery disease
Mapped Trait(s)	coronary artery disease (EFO_0001645)

Released in PGS Catalog: Oct. 14, 2019

Score Details

Score Construction

PGS Name	metaGRS_CAD
Development Method	
Name	metaGRS
Parameters	metaGRS log(HR) mixing weights: GRS46K=0.1278, FDR202=0.2359 and 1000Genomes=0.2400
Variants	
Original Genome Build	hg19
Number of Variants	1,745,179
Effect Weight Type	NR

PGS Source

PGS Catalog Publication (PGP) ID	PGP000007
Citation (link to publication)	Inouye M et al. J Am Coll Cardiol (2018)

Ancestry Distribution

Source of Variant Associations (GWAS)

382,026 individuals (100%)

Score Development/Training

3,000 individuals (100%)

PGS Evaluation

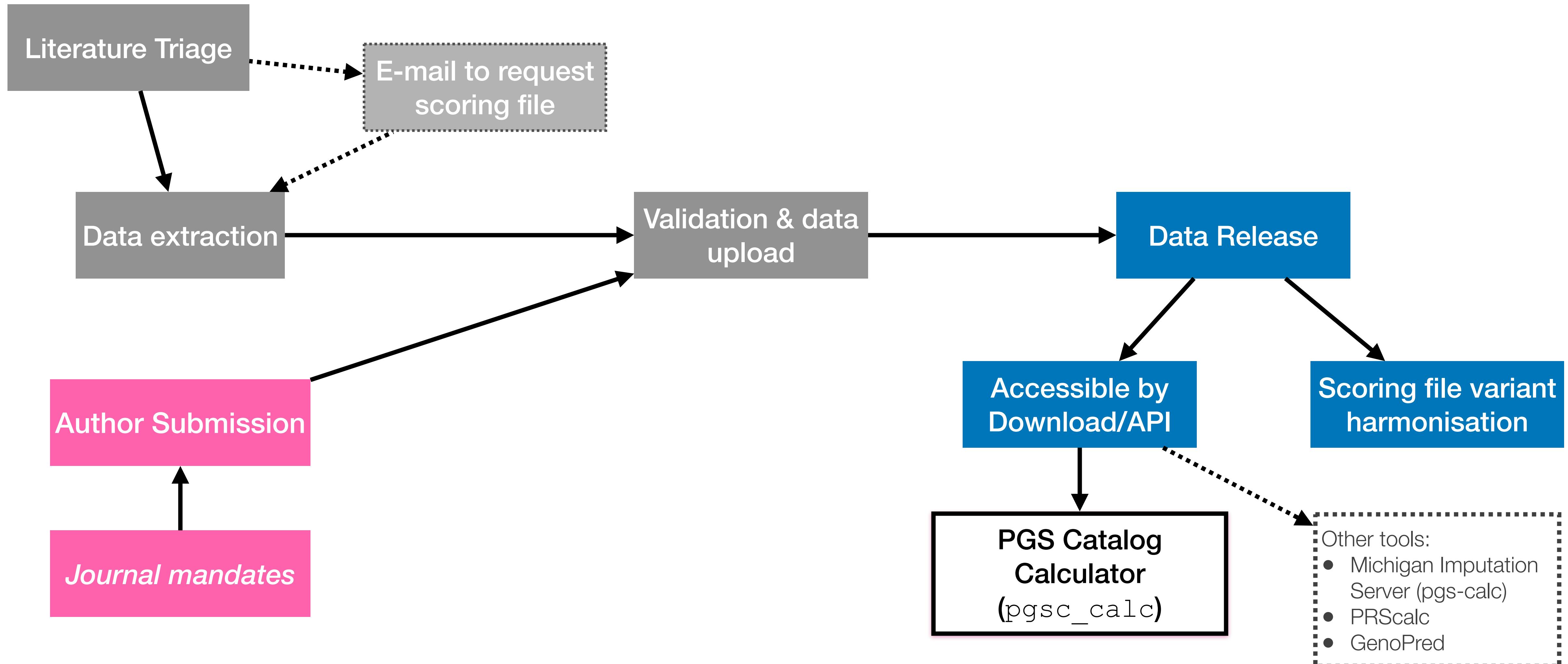
30 Sample Sets

Development Samples

Source of Variant Associations (GWAS)

Study Identifiers	Sample Numbers	Sample Ancestry	Cohort(s)
Europe PMC: 23202125	194,427 individuals	European, South Asian	34 cohorts
GWAS Catalog: GCST003116	2,268 individuals	Greater Middle Eastern (Middle Eastern, North African or Persian)	40 cohorts

PGS Catalog: FAIR data journey



How to submit data to the PGS Catalog.

www.PGSCatalog.org/submit

Submission: eligibility criteria

Newly developed PGS

- Variant information necessary to apply the PGS to new samples (variant rsID and/or genomic position, weights/effect sizes, effect allele, genome build).
- Information about how the PGS was developed (computational method, variant selection, relevant parameters).
- Descriptions of the samples used to develop and train the score, and samples used for external evaluation.
- Establishment of the score's analytic validity and predictive performance (e.g. effect sizes (beta, OR, HR), classification accuracy (AUC, C-index), proportion of the variance explained (R²)), and any covariates included in these models.

Evaluation of a previously developed PGS

- Evaluation of score already in the catalog (or eligible for inclusion) using new samples not used for PGS development
- Same information required as for evaluation of a new PGS

Preparing a data submission

Requires two file types:

1. Scoring file(s)

1. With PGS Catalog columns names (<https://www.pgscatalog.org/downloads/>)

1. Minimal: chromosomal position or rsID, effect allele, weight

2. Strongly recommended: non-effect allele (makes variant matching much easier)

2. Metadata file ([excel file](#))

	A	B	C	D	E	F
1	rsID	chr_name	chr_position	effect_allele	other_allele	OR
2	rs7912495	10	11676714	G	A	1.06
3	rs543928	6	32317471	G	A	1.08
4	rs1060743	2	127068957	G	A	1.11
5	rs9373079	6	134068394	G	A	1.12
6	rs73223431	8	27362470	T	C	1.07
7	rs679515	1	207577223	T	C	1.14
8	rs6733839	2	127135234	T	C	1.19
9	rs3851179	11	86157598	T	C	0.89
10	rs11787077	8	27607795	T	C	0.9
11	rs35220752	11	85941686	T	C	0.86
12	rs2972178	8	1739731	T	C	0.9
13	rs1582763	11	60254475	A	G	0.91
14	rs12151021	19	1050875	A	G	1.12
15	rs11525457	11	60321826	T	G	1.11
16	rs12998662	2	207101667	C	G	1.13
17	rs13391802	2	54918863	C	G	1.07
18	rs35114168	2	127090354	A	G	1.15
19	rs598561	11	85941783	A	G	1.08
20	rs11111149	12	102072301	A	G	1.27
21	rs2632516	17	58331728	C	G	0.91
22	rs7564197	2	127123792	A	G	0.88
23	rs745717	2	127136321	C	G	1.19
24	rs11168036	5	140327854	T	G	1.08
25	rs12358692	10	11679103	C	T	0.92
26	rs12618593	2	127128795	C	T	1.13
27	rs9304690	19	49950060	T	C	1.05
28	rs889555	16	31111250	T	C	0.96
29	rs8025980	15	50701814	G	A	0.96
30	rs7908662	10	122413396	G	A	0.96
31	rs7767350	6	47517390	T	C	1.06

Preparing a submission template

1: Publication information

	A	B	C	D	E	F	G	H
1	PubMed ID (PMID)	doi	Journal	Publication Date (e.g. dd-mm-yyyy)	First Author Last Name (e.g. Lambert)	First Author Initials (e.g. S A)		
2	Publication details							
	37450379	10.1002/alz.13378	Alzheimers Dement	14/7/2023	Sleiman	S PM		
2	Publication Curation Notes <small>(Supporting information with any details about the publication that PGS Catalog curators should know - <u>not displayed in the Catalog</u>)</small>	New cohort added: NxC						
3								
4								
5								

Preprints are eligible for inclusion
(use *doi*).

Pre-publication data can be embargoed until publication and we will provide IDs that can be added to papers.

Preparing a submission template

2: Score(s)

	A	B	C	D	E	F	G	H	I	J	K
1	Score Name/ID (must be unique)	Polygenic Score Trait Information				Polygenic Score Development Details					Curation Notes (Supporting information with any details about the score that PGS Catalog curators should know about - <u>not displayed in the Catalog</u>)
2	Reported Trait (phenotype that the polygenic score predicts)	Additional Trait Information (info that should be attached to the PGS rather than samples)	EFO IDs (comma-separated list of EFO_ID from Experimental Factor Ontology <u>-can ignore: it will be extracted by curators</u>)	EFO Names (Names of suggested ontology terms <u>-can ignore: it will be extracted by curators</u>)	Score Development Method (e.g. LDpred, PRSice, P+T, lassosum)	Score Development Details (describe key parameters used in the score. e.g. LD/SNP r2 threshold, significance/p-value threshold, fraction of causal variants (p))	Genome Build (Original genome build the variants/PGS are associated with)	Number of variants included in polygenic score (used for cross-referencing with the score file)	Number of interaction terms included in polygenic score (used for cross-referencing with the score file)		
3	PRS74_AD	Alzheimer's disease	MONDO_0004975	Alzheimer disease	Genome-wide significant SNPs	MAF<0.03, r2=0.3	GRCh38	74.00	0.00		
4											
5											
6											

Genome build is essential to ensure proper variant harmonisation and reusability!

Preparing a submission template

3: Sample descriptions

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q
1	Associated Score Name(s) (comma separated list, must match score names from Score(s) sheet)	Study Stage (Variant associations / Score development /Testing)	Sample Set ID (Used to associate the testing samples and scores with the relevant performance metrics if performance is reported separately)	Source GWAS Catalog (GCST ID will extract blue fields automatically from the database)	PMID or DOI (If no GWAS is available or to cite sample collections/cohorts)	Number of Individuals	Number of Cases (if applicable)	Number of Controls (if applicable)	Percent of participants who are Male (if known)	Age of Study Participants (if known)	Broad Ancestral Category (select from drop down list)	Ancestry (e.g. French, Chinese)	Country of recruitment (comma separated list)	Additional Ancestry Description	Detailed Phenotype Descriptions (e.g. ICD/SNOMED codes used to identify cases/phenotypes in EHR data)	Participant Follow-up Time (e.g. median duration or range of participant follow-up in the case of a prospective study design)	Cohort(s) (Provide a comma (,) separated list of Cohort IDs (see reference sheet) used in each phase. If the cohort is not present in our sheet you can add it and it will be parsed into the database)
2	PRS74_AD	GWAS/Variant associations		GCST004246													
3	PRS74_AD	GWAS/Variant associations		GCST90027158													EADB, GR@ACE, EADI, GERAD, PERADES, DemGene, Bonn, RS, CCHS, NxC, UKB
4		Testing	test.Japan			2,000	1,000	1,000	40.00	Mean = 56.3 years; sd = 7.1	East Asian	Japanese	Japan				NCGG
5		Testing	test.East London			718	104	614	43.00	Mean = 58.1 years; sd = 6.6	South Asian	Pakistani, Bangladeshi	UK				G&H
6		Testing	test.Korea			696	157	539	54.00	Mean = 59.9 years; sd = 7.0	East Asian	Korean	South Korea				BICWALZS
7		Testing	test.UKB			229,265	2,923	226,342	52.00	Mean = 55.3 years; sd = 6.1	European	British	UK				UKB

Multiple evaluation samples can be described



GWAS Catalog IDs link data and sample descriptions

Structured ancestry information and cohort annotations improve reusability

Preparing a submission template

4: Performance metrics

PGS Catalog Submissions

www.PGSCatalog.org/submit

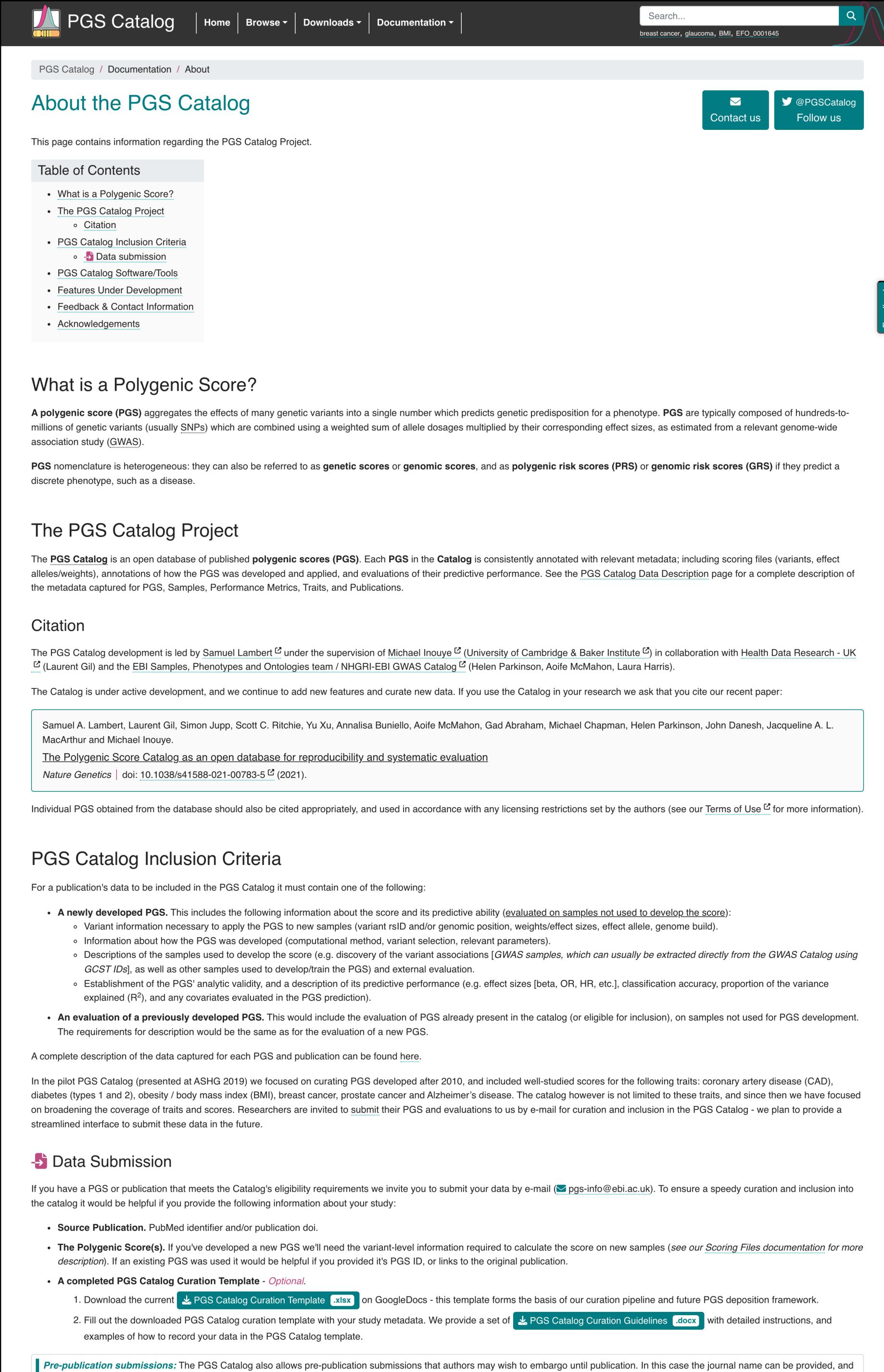
- Python package ([pgscatalog_utils](#)) provides tools for validating scoring files.

- More information about the PGS Catalog can be found in our documentation (www.PGSCatalog.org/about)

- Links to the metadata submission template ([excel file](#))

- We provide step-by-step guide to filling out the template and guidelines for each metadata field ([instructions documentation](#))

- Includes checklist, final submissions and questions can be sent to pgs-info@ebi.ac.uk



The screenshot shows the PGS Catalog website's "About" page. The header includes the PGS Catalog logo, navigation links for Home, Browse, Downloads, Documentation, and a search bar. On the right, there are social media links for Contact us (@PGSCatalog) and Follow us. A feedback link is also visible on the far right.

About the PGS Catalog

This page contains information regarding the PGS Catalog Project.

Table of Contents

- What is a Polygenic Score?
- The PGS Catalog Project
 - Citation
- PGS Catalog Inclusion Criteria
 - Data submission
- PGS Catalog Software/Tools
- Features Under Development
- Feedback & Contact Information
- Acknowledgements

What is a Polygenic Score?

A polygenic score (PGS) aggregates the effects of many genetic variants into a single number which predicts genetic predisposition for a phenotype. PGS are typically composed of hundreds-to-millions of genetic variants (usually SNPs) which are combined using a weighted sum of allele dosages multiplied by their corresponding effect sizes, as estimated from a relevant genome-wide association study (GWAS).

PGS nomenclature is heterogeneous: they can also be referred to as **genetic scores** or **genomic scores**, and as **polygenic risk scores (PRS)** or **genomic risk scores (GRS)** if they predict a discrete phenotype, such as a disease.

The PGS Catalog Project

The PGS Catalog is an open database of published polygenic scores (PGS). Each PGS in the Catalog is consistently annotated with relevant metadata; including scoring files (variants, effect alleles/weights), annotations of how the PGS was developed and applied, and evaluations of their predictive performance. See the PGS Catalog Data Description page for a complete description of the metadata captured for PGS, Samples, Performance Metrics, Traits, and Publications.

Citation

The PGS Catalog development is led by Samuel Lambert ¹ under the supervision of Michael Inouye ² (University of Cambridge & Baker Institute ²) in collaboration with Health Data Research - UK ² (Laurent Gil) and the EBI Samples, Phenotypes and Ontologies team / NHGRI-EBI GWAS Catalog ² (Helen Parkinson, Aoife McMahon, Laura Harris).

The Catalog is under active development, and we continue to add new features and curate new data. If you use the Catalog in your research we ask that you cite our recent paper:

Samuel A. Lambert, Laurent Gil, Simon Jupp, Scott C. Ritchie, Yu Xu, Annalisa Buniello, Aoife McMahon, Gad Abraham, Michael Chapman, Helen Parkinson, John Danesh, Jacqueline A. L. MacArthur and Michael Inouye.
[The Polygenic Score Catalog as an open database for reproducibility and systematic evaluation](#)
Nature Genetics | doi: 10.1038/s41588-021-00783-5 ² (2021).

Individual PGS obtained from the database should also be cited appropriately, and used in accordance with any licensing restrictions set by the authors (see our Terms of Use ² for more information).

PGS Catalog Inclusion Criteria

For a publication's data to be included in the PGS Catalog it must contain one of the following:

- A newly developed PGS. This includes the following information about the score and its predictive ability (evaluated on samples not used to develop the score):
 - Variant information necessary to apply the PGS to new samples (variant rsID and/or genomic position, weights/effect sizes, effect allele, genome build).
 - Information about how the PGS was developed (computational method, variant selection, relevant parameters).
 - Descriptions of the samples used to develop the score (e.g. discovery of the variant associations [GWAS samples, which can usually be extracted directly from the GWAS Catalog using GCST IDs], as well as other samples used to develop/train the PGS) and external evaluation.
 - Establishment of the PGS' analytic validity, and a description of its predictive performance (e.g. effect sizes [beta, OR, HR, etc.], classification accuracy, proportion of the variance explained (R^2), and any covariates evaluated in the PGS prediction).
- An evaluation of a previously developed PGS. This would include the evaluation of PGS already present in the catalog (or eligible for inclusion), on samples not used for PGS development. The requirements for description would be the same as for the evaluation of a new PGS.

A complete description of the data captured for each PGS and publication can be found [here](#).

In the pilot PGS Catalog (presented at ASHG 2019) we focused on curating PGS developed after 2010, and included well-studied scores for the following traits: coronary artery disease (CAD), diabetes (types 1 and 2), obesity / body mass index (BMI), breast cancer, prostate cancer and Alzheimer's disease. The catalog however is not limited to these traits, and since then we have focused on broadening the coverage of traits and scores. Researchers are invited to submit their PGS and evaluations to us by e-mail for curation and inclusion in the PGS Catalog - we plan to provide a streamlined interface to submit these data in the future.

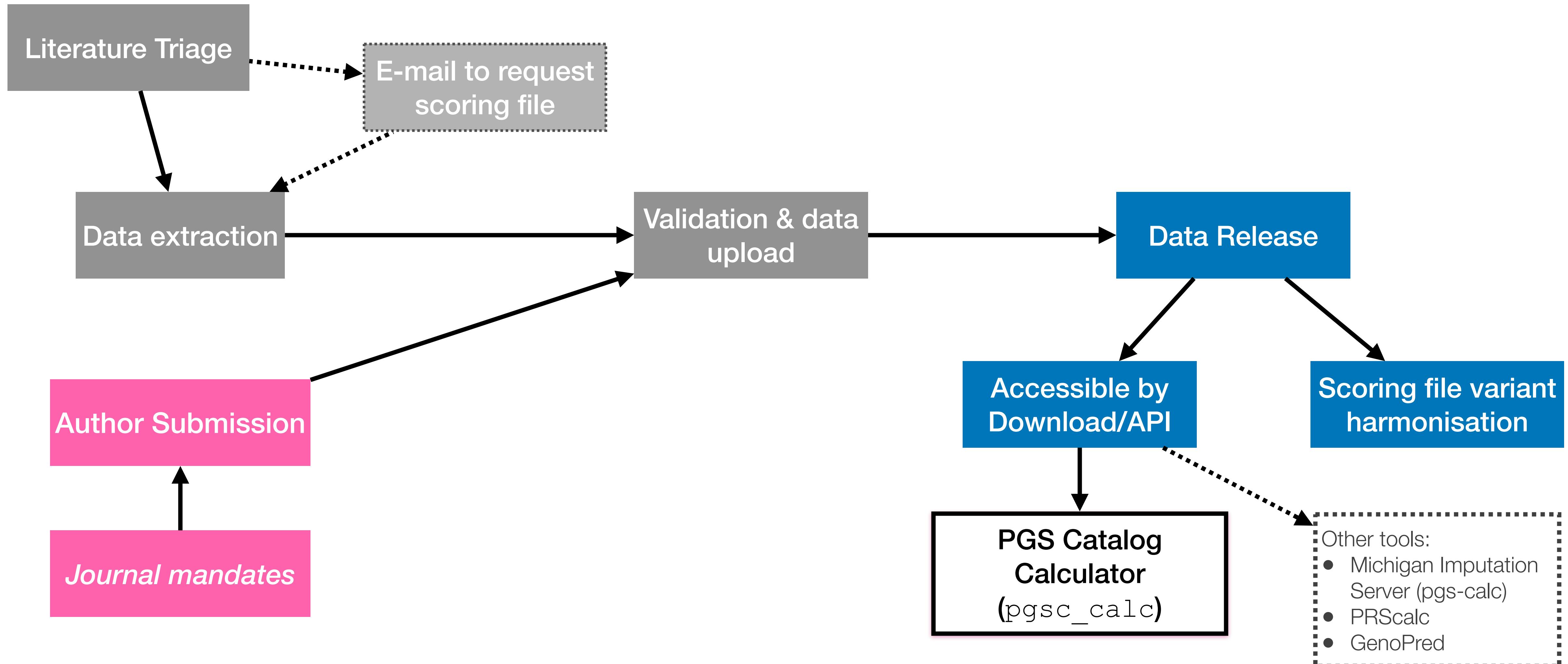
Data Submission

If you have a PGS or publication that meets the Catalog's eligibility requirements we invite you to submit your data by e-mail (pgs-info@ebi.ac.uk). To ensure a speedy curation and inclusion into the catalog it would be helpful if you provide the following information about your study:

- Source Publication. PubMed identifier and/or publication doi.
- The Polygenic Score(s). If you've developed a new PGS we'll need the variant-level information required to calculate the score on new samples (see our [Scoring Files documentation](#) for more description). If an existing PGS was used it would be helpful if you provided its PGS ID, or links to the original publication.
- A completed PGS Catalog Curation Template - *Optional*.
 1. Download the current [PGS Catalog Curation Template .xlsx](#) on GoogleDocs - this template forms the basis of our curation pipeline and future PGS deposition framework.
 2. Fill out the downloaded PGS Catalog curation template with your study metadata. We provide a set of [PGS Catalog Curation Guidelines .docx](#) with detailed instructions, and examples of how to record your data in the PGS Catalog template.

Pre-publication submissions: The PGS Catalog also allows pre-publication submissions that authors may wish to embargo until publication. In this case the journal name can be provided, and

PGS Catalog: FAIR data journey



Questions?

www.PGSCatalog.org & pgs-info@ebi.ac.uk