

Polygenic Risk Score – Predict for Array module Release Notes

In BaseSpace™ Sequence Hub

June 2023

Introduction

Polygenic Risk Score – Predict module provides comprehensive, easy to implement and versatile cloud-based analysis for polygenic risk score (PRS) predictions using microarrays. The easy-to-use interface on BaseSpace Sequence Hub simplifies the process of setting up, running, and reviewing analysis results.

Highlights include:

COMPREHENSIVE

The only commercially available genotype-to-risk prediction toolkit.

EASY TO IMPLEMENT

User friendly, out-of-the-box software simplifies the work typically done by bioinformaticians. Automated PRS calculations based on high-performing algorithms and published data.

VERSATILE

Enables a variety of research needs by supporting multiple disease areas of study. Scalable to enable small- and large-scale studies.

These Release Notes detail the key features and changes to software components for the release of the Predict module. For information on how to use the system for PRS analysis, see the [Polygenic Risk Score Software Online Help](#). Predict module is an efficient solution for array analysis, including features such as:

- Simple two-step process for analysis kick-off in BaseSpace Sequence Hub
 - Includes Configuration and Sample Selection
- iScan™ integration allows data to be sent directly to cloud platform removing the need for manual upload
- New Polygenic Risk Score – Predict Analysis Type to create PRS predictions.
 - User defined name and project to identify the analysis
- Create new Configuration settings
 - Select from commercially available products or upload semi-custom or custom product files
 - Upload a Polygenic Score ID File that contains up to 24 custom or built-in PRS to calculate
- Select samples by either importing a sample sheet or selecting BeadChips
 - Selecting BeadChips makes analysis setup easy and removes the need for manual sample sheet generation

- View analysis results on BaseSpace or download to review locally.
- Demo data available on the Demo Data tab in BaseSpace as an example of expected results.

NEW FEATURES IN DETAIL

- Addition of Polygenic Risk Score - Predict version 1.1.0 pipeline
 - Key outputs include PRS CSV, PRS PDF, SNV VCF (imputed per batch and un-imputed per sample), Genotype Summary File and PRS log file
 - Utilizes Illumina Connected Analytics (ICA) version 2 platform
- Ability to run Predict with built-in PRS, custom PRS, or one downloaded from the PGS catalog
 - When running custom PRS, a custom_results.csv is generated (no PDF report is generated)
- Includes access to over 2700 PRS representing over 550 different traits from the public PGS Catalog (www.pgscatalog.org)

⊖ Includes scores released in the PGS Catalog on or before Oct 19, 2022 with the following exceptions:

PGS000126	PGS002645	PGS002502	PGS002646	PGS002662
PGS000192	PGS002650	PGS002632	PGS002648	PGS002665
PGS000941	PGS002660	PGS002633	PGS002649	PGS002669
PGS000966	PGS002663	PGS002635	PGS002652	PGS002670
PGS002627	PGS002666	PGS002639	PGS002655	PGS002671
PGS002630	PGS002672	PGS002640	PGS002656	PGS002673
PGS002631	PGS002674	PGS002642	PGS002657	
PGS002643	PGS002746	PGS002644	PGS002659	

- Please note, PGS obtained from the Catalog should be cited appropriately, and used in accordance with any licensing restrictions set by the authors. See EBI Terms of Use (<https://www.ebi.ac.uk/about/terms-of-use/>) for additional details.
- Includes access to 11 Allelica specific scores. Please review further details of these scores including their Polygenic Score ID in Appendix A.
 - Coronary artery disease
 - Breast cancer (available for females only)
 - Prostate cancer (available for males only)

- Diabetes Type 2
 - Polygenic hypercholesterolemia (LDL Cholesterol)
 - Alzheimer's (new with Predict 1.1)
 - Atrial fibrillation
 - Early menopause (available for females only) (new with Predict 1.1)
 - Hypertension
 - Triglycerides
 - High-density lipoprotein (HDL) cholesterol
- The following products have product files, including manifests (*.csv and *.bpm formats) and cluster file (*.egt), already available within the system:
 - Global Diversity Array with PRS content
 - Global Diversity Array
 - Global Diversity Array with enhanced PGx
 - Global Screening Array version 3

KNOWN ISSUES AND LIMITATIONS

- Manifests (*.csv and *.bpm) and cluster files (*.egt) may take several minutes or longer to upload. Slow or instable network connections may lead to system timeout or failed uploads.
- Samples with the same Sample ID cannot be run in the same analysis. This will result in error and prevent analysis from proceeding.
- The Polygenic Score ID File should be saved and uploaded in CSV format. Files in UTF-8 CSV format will result in error.

APPENDIX A – INFORMATION ON ALLELICA DEVELOPED SCORES

Polygenic Score File ID	Disease/Trait	PRS method (number of variants)	Ancestry	AUC (95%CI)	ORxSTD (95%CI)	Number (cases/total)	Literature Summary
ALLELICA_CAD_2019	Coronary Artery Disease (CAD)	SCT + metaPRS 1,926,521	European	0.79 (0.79-0.8)	1.77 (1.75-1.79)	16,257 293,286	<ul style="list-style-type: none"> - Nikpay et al 2015 Nature Genetics - Bolli et al 2019 biorxiv - Bolli et al 2021 Circulation - Prive et al 2019 AJHG - Inouye et al 2018 JACC - Khera et al 2018 Nature Genetics - Weale et al 2021 Am J Cardiol - Mujwara et al 2021 medrxiv - Riveros-Mckay et al 2021 Circ Genom Precis Med - Sun et al 2021 PLoS Med
			African	0.67 (0.62-0.73)	1.24 (1.14-1.36)	222 7,646	
			South Asian	0.8 (0.78-0.82)	1.81 (1.71-1.92)	903 9,417	
			East Asian	0.73 (0.6-0.85)	1.27 (1.09-1.5)	39 1,503	
ALLELICA_AF_2020	Atrial Fibrillation (AF)	SCT 445,014	European	0.76 (0.76-0.77)	1.53 (1.51-1.56)	14,063 286,192	<ul style="list-style-type: none"> - Christopherson et al 2017 Nat Gen - Prive et al 2019 AJHG
			African	0.76 (0.73-0.80)	1.13 (0.93-1.34)	147 7,554	
			South Asian	0.76 (0.73-0.79)	1.38 (1.20-1.60)	194 7,551	
			East Asian	0.77 (0.66-0.88)	1.31 (0.81-2.09)	19 1,484	
ALLELICA_LDL_2020	Low Density Lipoprotein (LDL)	Lassosum 3,036	European	0.72 (0.71-0.72)	2.23 (2.2-2.25)	18,828 279,815	<ul style="list-style-type: none"> - Willer et al 2013 Nature Genetics - Mak et al 2017 Genetic Epidemiology - Sinnott-Armstrong et al 2021 Nature Genetics - Bolli et al 2021 Circulation - Medeiros and Bourbon 2021 Curr Opin Lipidol
			African	0.68 (0.63-0.72)	1.91 (1.76-2.1)	236 7,161	
			South Asian	0.66 (0.62-0.71)	1.82 (1.7-1.96)	307 8,927	
			East Asian	0.65 (0.53-0.76)	1.82 (1.57-2.11)	44 1,425	

ALLELICA_T2D_2019	Type 2 Diabetes (T2D)	SCT 620,612	European	0.71 (0.71-0.72)	1.66 (1.64-1.68)	14,314 293,286	<ul style="list-style-type: none"> - Scott et al 2017 Diabetes - Prive et al 2019 AJHG - Vujkovic et al 2020 Nature Genetics - Mahajan et al 2018 Nature Genetics - Li et al 2021 Diabetes - Emdin et al 2017 JAMA - Udler et al 2019 Endocr Rev - Padilla-Martinez et al 2020 Int J Mol Sci 	
			African	0.69 (0.66-0.72)	1.3 (1.2-1.41)	751 7,646		
			South Asian	0.69 (0.67-0.71)	1.44 (1.38-1.5)	1,522 9,417		
			East Asian	0.69 (0.59-0.78)	1.11 (0.95-1.31)	75 1,503		
ALLELICA_BC_2019	Breast Cancer (BC)	SCT 577,113	European	0.67 (0.66-0.68)	1.71 (1.68-1.73)	10,986 278,323	<ul style="list-style-type: none"> - Michailidou et al 2017 Nature - Prive et al 2019 AJHG - Bolli et al 2019 biorxiv - Busby et al 2021 medrxiv - Mavaddat et al 2019 AJHG - Fahed et al 2020 Nature Communications - Mars et al 2020 Nature Medicine - Mars et al 2020 Nature Communications - Kramer et al 2020 AJHG - Kapoor et al 2021 J Natl Cancer Inst - Barnes et al 2020 Genet Med 	
			South Asian	0.65 (0.59-0.71)	1.55 (1.43-1.7)	218 9,351		
		Mavaddat 313	East Asian	0.63 (0.51-0.74)	1.45 (1.22-1.79)	45 937		
			African	0.6 (0.54-0.67)	1.41 (1.27-1.55)	161 4,360		
ALLELICA_PC_2019	Prostate Cancer (PC)	SCT 682,397	European	0.8 (0.79-0.8)	2.13 (2.09-2.17)	6,219 278,323		<ul style="list-style-type: none"> - Schumacher et al 2018 Nature Genetics - Prive et al 2019 AJHG - Conti et al 2021 Nature Genetics - Black et al 2020 Prostate - Karunamuni et al 2021 Prostate Cancer Prostatic Dis - Plym et al 2021 J Natl Cancer Inst - Darst et al 2021 Eur Urol - Szulkin et al 2015 Prostate - Zhang et al 2020 Nature Communications - Xu and Isaacs 2021 Eur Urol - Huynh-Le et al 2021 Nature Communications
			South Asian	0.8 (0.75-0.86)	2.78 (2.11-3.62)	93 9,351		
		Schumacher 149	East Asian	0.79 (0.37-0.95)	1.24 (1.08-1.5)	8 565		
			African	0.83 (0.78-0.86)	1.57 (1.41-1.74)	196 3,283		

ALLELICA_HT_2020	Hypertension (SBP>160 mmHg) (HT)	Lassosum 247,151	European	0.72 (0.71-0.72)	1.57 (1.56-1.58)	30,032 41,329	<ul style="list-style-type: none"> - Warren et al 2017 Nature Genetics - Mak et al 2017 Genetic Epidemiology - Padmanabhan and Dominiczak 2020 Nat Rev Cardiol - Sinnott-Armstrong et al 2021 Nature Genetics - Sakaue et al 2020 Nature Medicine
			African	0.67 (0.64-0.7)	1.23 (1.17-1.3)	797 6,572	
			South Asian	0.72 (0.69-0.75)	1.35 (1.28-1.43)	727 7,926	
			East Asian	0.74 (0.67-0.81)	1.43 (1.25-1.64)	102 1,263	

ALLELICA_HDL_2020	High Density Lipoprotein (HDL<=1.164 mmol/l)	Lasso sum 322,564	European	0.75 (0.75-0.76)	1.84 (1.82-1.85)	59,800 233,486	<ul style="list-style-type: none"> - Willer et al 2013 Nature Genetics - Mak et al 2017 Genetic Epidemiology - Sinnott-Armstrong et al 2021 Nature Genetics
			African	0.71 (0.69-0.74)	1.34 (1.26-1.42)	773 5,866	
			South Asian	0.73 (0.71-0.75)	1.69 (1.62-1.75)	2,070 6,103	
			East Asian	0.75 (0.69-0.82)	1.60 (1.42-1.82)	141 1,162	
ALLELICA_TG_2020	Triglycerides (TG>=2.258 mmol/l)	Lasso sum 68,800	European	0.68 (0.67-0.68)	1.65 (1.64-1.66)	62,157 231,129	<ul style="list-style-type: none"> - Willer et al 2013 Nature Genetics - Mak et al 2017 Genetic Epidemiology - Sinnott-Armstrong et al 2021 Nature Genetics
			African	0.63 (0.42-0.81)	1.29 (1.03-1.60)	23 7,150	
			South Asian	0.75 (0.69-0.80)	2.30 (2.10-2.52)	204 8,733	
			East Asian	0.76 (0.62-0.88)	1.88 (1.54-2.3)	26 1,400	