Polygenic Risk Score – Predict for Array module Release Notes

In BaseSpace™ Sequence Hub

December 2022

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 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 1.0.0 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
- 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 9 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)
    - Atrial fibrillation

  - 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.
- The following PRS are missing a data label:

PGS000085	PGS002794	PGS002798	PGS002802
PGS001018	PGS002795	PGS002799	PGS002803
PGS002745	PGS002796	PGS002800	PGS002804
PGS002752	PGS002797	PGS002801	PGS002805

-
- The following PRS will show a duplicated entry in the final csv:

PGS001185	PGS001611	PGS001644	PGS001677	PGS001710	PGS001743	PGS001776
PGS001192	PGS001612	PGS001645	PGS001678	PGS001711	PGS001744	PGS001777
PGS001199	PGS001613	PGS001646	PGS001679	PGS001712	PGS001745	PGS001778
PGS001200	PGS001614	PGS001647	PGS001680	PGS001713	PGS001746	PGS001779
PGS001218	PGS001615	PGS001648	PGS001681	PGS001714	PGS001747	PGS001780
PGS001219	PGS001616	PGS001649	PGS001682	PGS001715	PGS001748	PGS001781
PGS001220	PGS001617	PGS001650	PGS001683	PGS001716	PGS001749	PGS001782
PGS001225	PGS001618	PGS001651	PGS001684	PGS001717	PGS001750	PGS001783

PGS001226	PGS001619	PGS001652	PGS001685	PGS001718	PGS001751	PGS001784
PGS001227	PGS001620	PGS001653	PGS001686	PGS001719	PGS001752	PGS001785
PGS001228	PGS001621	PGS001654	PGS001687	PGS001720	PGS001753	PGS001786
PGS001229	PGS001622	PGS001655	PGS001688	PGS001721	PGS001754	PGS001787
PGS001230	PGS001623	PGS001656	PGS001689	PGS001722	PGS001755	PGS001788
PGS001591	PGS001624	PGS001657	PGS001690	PGS001723	PGS001756	PGS001789
PGS001592	PGS001625	PGS001658	PGS001691	PGS001724	PGS001757	PGS001790
PGS001593	PGS001626	PGS001659	PGS001692	PGS001725	PGS001758	PGS001791
PGS001594	PGS001627	PGS001660	PGS001693	PGS001726	PGS001759	PGS001792
PGS001595	PGS001628	PGS001661	PGS001694	PGS001727	PGS001760	PGS001793
PGS001596	PGS001629	PGS001662	PGS001695	PGS001728	PGS001761	PGS001794
PGS001597	PGS001630	PGS001663	PGS001696	PGS001729	PGS001762	PGS001795
PGS001598	PGS001631	PGS001664	PGS001697	PGS001730	PGS001763	PGS001796
PGS001599	PGS001632	PGS001665	PGS001698	PGS001731	PGS001764	PGS001797
PGS001600	PGS001633	PGS001666	PGS001699	PGS001732	PGS001765	PGS001798
PGS001601	PGS001634	PGS001667	PGS001700	PGS001733	PGS001766	PGS001799
PGS001602	PGS001635	PGS001668	PGS001701	PGS001734	PGS001767	PGS001800
PGS001603	PGS001636	PGS001669	PGS001702	PGS001735	PGS001768	PGS001801
PGS001604	PGS001637	PGS001670	PGS001703	PGS001736	PGS001769	PGS001802
PGS001605	PGS001638	PGS001671	PGS001704	PGS001737	PGS001770	PGS001803
PGS001606	PGS001639	PGS001672	PGS001705	PGS001738	PGS001771	PGS001804
PGS001607	PGS001640	PGS001673	PGS001706	PGS001739	PGS001772	PGS001805
PGS001608	PGS001641	PGS001674	PGS001707	PGS001740	PGS001773	PGS001806
PGS001609	PGS001642	PGS001675	PGS001708	PGS001741	PGS001774	PGS001807
PGS001610	PGS001643	PGS001676	PGS001709	PGS001742	PGS001775	PGS001808

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	- 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 *medrkiv* - 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	- 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	- 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	- 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	- Michailidou et al 2017 *Nature* - Prive et al 2019 *AJHG* - Bolli et al 2019 biorxiv - Busby et al 2021 medrkiv - 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	- 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	- 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)	Lassos um 322,564	European	0.75 (0.75-0.76)	1.84 (1.82-1.85)	59,800 233,486	- 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)	Lassos um 68,800	European	0.68 (0.67-0.68)	1.65 (1.64-1.66)	62,157 231,129	- 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	