

BridgePRS: A powerful trans-ancestry Polygenic Risk Score method

Clive Hoggart^{1*}, Shing Wan Choi^{1,2}, Michael Preuss³
and Paul O'Reilly^{1,2}

¹Department of Genetics and Genomic Sciences, Icahn School of Medicine, Mount Sinai, 1 Gustave L. Levy Pl, New York, 10029, NY, USA.

²MRC Social, Genetic and Developmental Psychiatry Centre, Institute of Psychiatry, Psychology and Neuroscience, King's College London, De Crespigny Park, Denmark Hill, London, SE5 8AF, London, UK.

³The Charles Bronfman Institute for Personalized Medicine, Icahn School of Medicine, Mount Sinai, 1 Gustave L. Levy Pl, New York, 10029, NY, USA.

*Corresponding author(s). E-mail(s): clive.hoggart@mssm.edu;
Contributing authors: shingwan.choi@mssm.edu;
michael.preuss@mssm.edu; paul.oreilly@mssm.edu;

2 BridgePRS: A powerful trans-ancestry Polygenic Risk Score method

	European		African			South Asian		
	Train	Test	Train	Test	Valid.	Train	Test	Valid.
Standing height	257675	10000	4862	1215	1214	6666	1666	1665
Body mass index	257327	10000	4853	1213	1212	6653	1662	1662
Platelet count	250404	10000	4657	1163	1163	6538	1634	1633
C-reactive protein	176422	10000	3281	820	819	4581	1145	1144
Neutrophil count	249945	10000	4642	1160	1159	6515	1628	1627
Mean corpuscular volume	250359	10000	4660	1164	1164	6541	1634	1634
Apolipoprotein A	161594	10000	3029	756	756	4182	1045	1044
Reticulocyte percentage	246092	10000	4531	1132	1131	6367	1591	1590
Alkaline phosphatase	177485	10000	3293	823	822	4611	1152	1152
LDL direct	177383	10000	3291	822	821	4603	1150	1150
Monocyte count	249782	10000	4639	1159	1158	6503	1625	1625
Red blood cell distribution width	249871	10000	4645	1161	1160	6523	1630	1630
Urea	177343	10000	3289	821	821	4603	1150	1149
Triglycerides	177317	10000	3289	821	821	4603	1150	1149
Basophil percentage	249368	10000	4623	1155	1154	6495	1623	1622
Total protein	162440	10000	3045	761	760	4190	1047	1046
Sodium in urine	250253	10000	4734	1183	1182	6497	1624	1623
IGF-1	176643	10000	3272	817	817	4583	1145	1144
Eosinophil count	249594	10000	4629	1157	1156	6503	1625	1624

Table 1 UK Biobank train, test and validation sample sizes for each ancestry group across the 19 traits analysed.

	African	South Asian
Standing height	4061	560
Body mass index	4084	560
LDL direct	1998	217
Mean corpuscular volume	2412	360
Platelet count	2459	368
Monocyte count	2229	283
Neutrophil count	2226	284
Eosinophil count	2204	268
Red blood cell distribution width	2398	366

Table 2 BioMe Biobank sample sizes for individuals of African and South Asian ancestry.

		Mean R^2		$BridgePRS$ > $PRS-CSx$
		$BridgePRS$	$PRS-CSx$	
UKB (All traits)	AFR	0.0335	0.0350	37%
	SAS	0.0666	0.0747	26%
UKB (Biome traits)	AFR	0.0522	0.0534	44%
	SAS	0.0831	0.0906	22%
Biome	AFR	0.0380	0.0272	78%
	SAS	0.0682	0.0564	33%

Table 3 Mean phenotypic variance explained (R^2) of $BridgePRS$ and $PRS-CSx$ in the UKB and Biome for samples of African (AFR) and South Asian (SAS) ancestry using HapMap variants. Also shown is the proportion of times R^2 of $BridgePRS$ is greater than $PRS-CSx$. UKB results are shown across all 19 traits and the 9 overlapping Biome traits.

	$BridgePRS$			$PRS-CSx$	
	Genotype	HM	Imputed	Genotype	Imputed
Standing height	87207	123823	524480	133043	366805
Body mass index	73546	103404	413878	133043	723684
Platelet count	68802	94484	402481	126055	723684
C-reactive protein	50483	67547	297829	133043	366858
Neutrophil count	66037	91474	346048	126055	366858
Mean corpuscular volume	64825	88758	382345	133043	366827
Apolipoprotein A	57951	78725	302515	133043	723684
Reticulocyte percentage	63607	88514	373447	126055	366834
Alkaline phosphatase	56736	76852	337408	133043	366858
LDL direct	50861	68668	300849	133043	723684
Monocyte count	65317	89455	377269	126055	366827
Red blood cell distribution width	59016	80272	341538	133043	723684
Urea	51303	69593	311268	133043	366858
Triglycerides	56507	77052	204873	126055	723684
Basophil percentage	48714	64182	289141	126055	366858
Total protein	58973	79396	339961	133043	723684
Sodium in urine	52591	71585	318576	126055	723684
IGF-1	61699	85475	333743	133043	723684
Eosinophil count	61913	83947	333743	133043	723684

Table 4 Number of variants included in the African PRS for $BridgePRS$ and $PRS-CSx$ for models estimated using genotyped or imputed variants. For $PRS-CSx$ number of variants available is the intersection with HapMap variants. HM – number of variants used by $BridgePRS$ using imputed HapMap variants.

	<i>BridgePRS</i>			<i>PRS-CSx</i>	
	Genotype	HM	Imputed	Genotype	Imputed
Standing height	115839	202665	975776	133043	590342
Body mass index (BMI)	96980	172878	847652	138560	723684
Platelet count	91872	155287	756870	138560	723684
C-reactive protein	68484	111978	561922	133043	723684
Neutrophil count	84696	141890	697828	138560	723684
Mean corpuscular volume	86943	147231	718429	133043	723684
Apolipoprotein A	77345	128441	627506	133043	723684
Reticulocyte percentage	84815	144224	690814	133043	723684
Alkaline phosphatase	76249	126993	625959	133043	590414
LDL direct	68326	112841	558028	138560	723684
Monocyte count	86408	143988	700602	133043	723684
Red blood cell distribution width	79156	131042	640731	133043	723684
Urea	70494	119000	600278	133043	723684
Triglycerides	75240	124776	364552	138560	590324
Basophil percentage	65220	105071	533637	133043	723684
Total protein	78864	128951	628154	138560	723684
Sodium in urine	70863	120127	607771	138560	723684
IGF-1	82545	140889	690321	133043	590352
Eosinophil count	83117	137221	667700	133043	723684

Table 5 Number of variants included in the South Asian PRS for *BridgePRS* and *PRS-CSx* for models estimated using genotyped or imputed variants. For *PRS-CSx* number of variants available is the intersection with HapMap variants. HM – number of variants used by *BridgePRS* using imputed HapMap variants.

	<i>BridgePRS</i>			<i>PRS-CSx</i>	
	Genotype	HM	Imputed	Genotype	Imputed
Standing height	99.3	99.7	99.3	93.5	99.6
Body mass index	99.3	99.7	99.3	93.5	96.8
Platelet count	99.3	99.8	99.3	99.5	96.8
Neutrophil count	99.3	99.7	99.3	99.5	99.6
Mean corpuscular volume	99.4	99.8	99.3	93.5	99.6
LDL direct	99.3	99.8	99.3	93.5	96.8
Monocyte count	99.4	99.7	99.3	99.5	99.6
Red blood cell distribution width	99.3	99.7	99.4	93.5	96.8
Eosinophil count	99.3	99.7	99.3	93.5	96.8

Table 6 Proportion of variants in the African PRS models available in the Biome data.

	<i>BridgePRS</i>			<i>PRS-CSx</i>	
	Genotype	HM	Imputed	Genotype	Imputed
Standing height	92.6	98.9	96.9	97.3	99.3
Body mass index	92.6	98.9	96.9	99.2	98.2
Platelet count	92.5	98.9	96.8	99.2	98.2
Neutrophil count	92.6	98.9	96.9	99.2	98.2
Mean corpuscular volume	92.6	98.9	96.7	97.3	98.2
LDL direct	92.1	98.8	96.7	99.2	98.2
Monocyte count	92.7	98.9	96.8	97.3	98.2
Red blood cell distribution width	92.4	98.9	96.7	97.3	98.2
Eosinophil count	92.6	98.9	96.8	97.3	98.2

Table 7 Proportion of variants in the South Asian PRS models available in the Biome data.

4 BridgePRS: A powerful trans-ancestry Polygenic Risk Score method

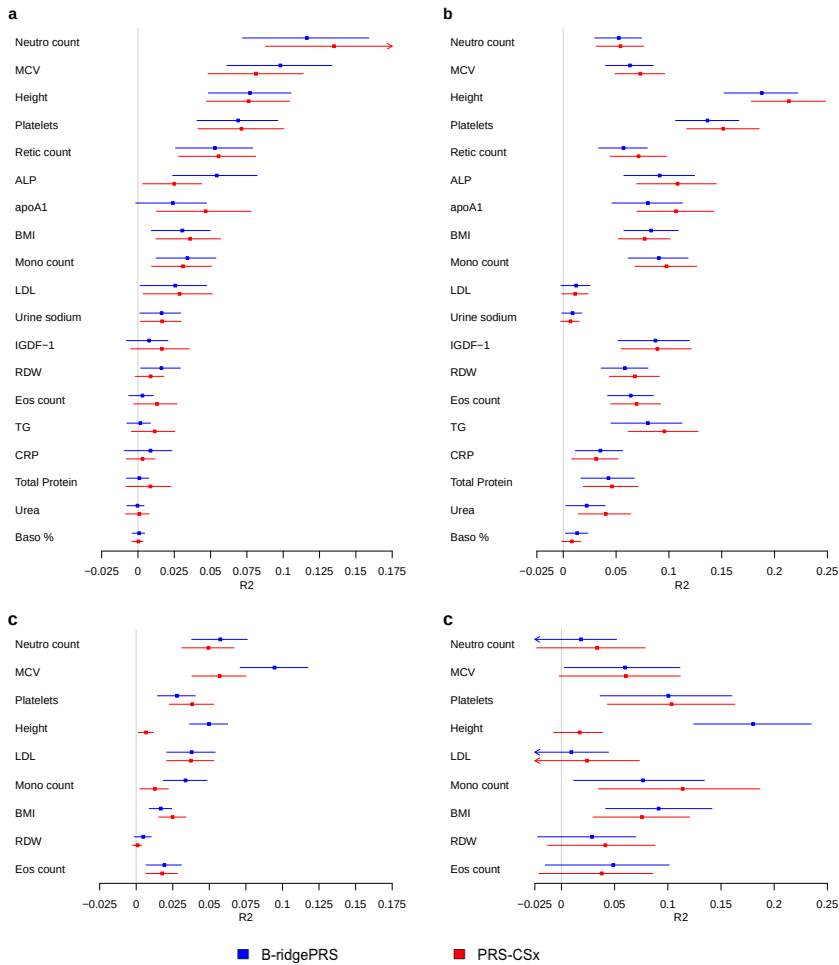


Fig. 1 Phenotypic variance explained (R^2 point estimates and 95% confidence intervals) by *BridgePRS* and *PRS-CSx* in samples of African and South Asian ancestry in the UK Biobank and Biome cohorts using HapMap variants. **A** African ancestry samples in UK Biobank. **B** South Asian ancestry samples in UK Biobank. **C** African ancestry samples in Biome. **D** South Asian ancestry samples in Biome. Neutro count=Neutrophill count, MCV=Mean corpuscular volume, Platelets=Platelet count, Retic count=Reticulocyte percentage, ALP=Alkaline phosphatase, Mono count=Monocyte count, apoA1=Apolipoprotein A, BMI=Body mass index, RDW=Red blood cell distribution width, Eos count=Eosinophill count, TG=Triglycerides, Baso %=Basophill percentage, CRP=C-reactive protein

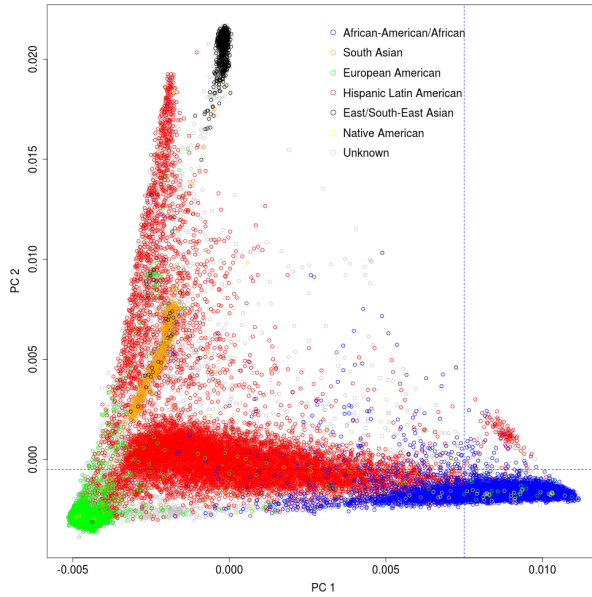


Fig. 2 Plot of PC 1 v PC 2 for Biome samples. Samples are coloured by self reported ancestry. Samples in bottom right corner indicated by blue dashed line were defined as African ancestry in our analyses.

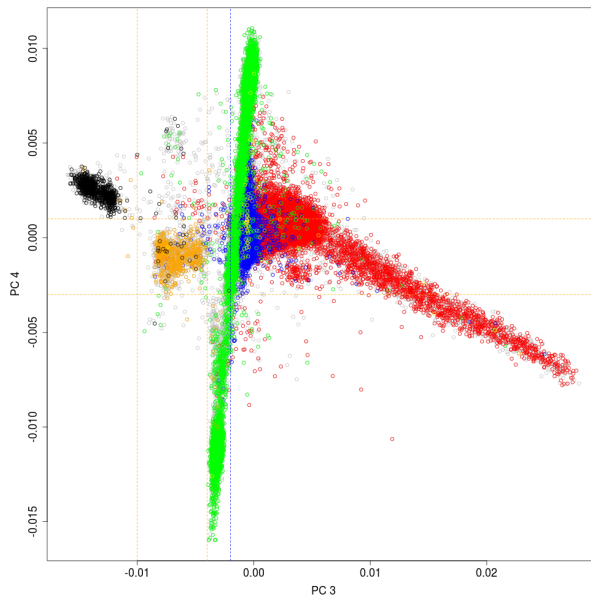


Fig. 3 Plot of PC 3 v PC 4 for Biome samples. Samples are coloured by self reported ancestry. Samples in box defined by orange lines were defined as South Asian ancestry in our analyses. Samples to left of blue vertical were excluded from African analyses.

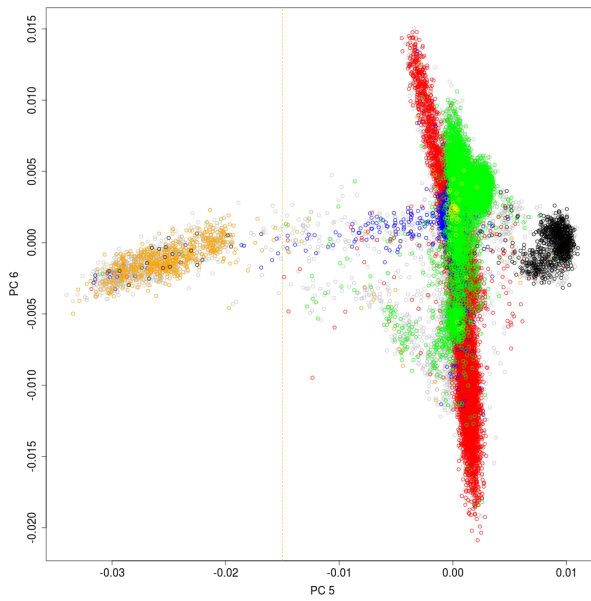


Fig. 4 Plot of PC 5 v PC 6 for Biome samples. Samples are coloured by self reported ancestry. Samples to right of orange vertical were excluded from South Asian analyses.