

Data Dictionary – GRS and Ancestry TGD Lab Pipeline

IID – Sample identifier TGD

RRID – Sample identifier from original study

SuperPopulationClass – classified primary super population ancestry

PopulationClass – classified primary subpopulation ancestry

1_AFR – probability of super population ancestry (AFR=African)

1_AMR – probability of super population ancestry (AMR=Admixed Americans)

1_EAS – probability of super population ancestry (EAS=East Asians)

1_EUR – probability of super population ancestry (EUR=Europeans)

1_SAS – probability of super population ancestry (SAS=South Asians)

2_ACB – probability of population ancestry (ACB=African Carribean)

2_ASW – probability of population ancestry (ASW=African Ancestry Southwest US)

2_BEB – probability of population ancestry (BEB=Bengali)

2_CDX – probability of population ancestry (CDX=Dai Chinese)

2_CEU – probability of population ancestry (CEU=Utah European)

2_CHB – probability of population ancestry (CHB=Han Chinese)

2_CHS – probability of population ancestry (CHS=Southern Han Chinese)

2_CLM – probability of population ancestry (CLM=Colombian)

2_ESN – probability of population ancestry (ESN=Esan Nigerian)

2_FIN – probability of population ancestry (FIN=Finnish)

2_GBR – probability of population ancestry (GBR=British)

2_GIH – probability of population ancestry (GIH=Gujurati Indian)

2_GWD – probability of population ancestry (GWD=Gambian Mandinka)

2_IBS – probability of population ancestry (IBS=Iberian Spanish)

2_ITU – probability of population ancestry (ITU=Telugu Indian)

2_JPT – probability of population ancestry (JPT=Japanese)

2_KHV – probability of population ancestry (KHV=Kinh Vietnamese)

2_LWK – probability of population ancestry (LWK=Luhya Kenyan)

2_MSL – probability of population ancestry (MSL=Mende Sierra Leone)

2_MXL – probability of population ancestry (MXL=Mexican American)

2_PEL – probability of population ancestry (PEL=Peruvian)

2_PJL – probability of population ancestry (PJB=Punjabi)

2_PUR – probability of population ancestry (PUR=Puerto Rican)

2_STU – probability of population ancestry (STU=Tamil Sri Lankan)

2_TSI – probability of population ancestry (TSI=Toscani Italian)

2_YRI – probability of population ancestry (YRI=Yoruba Nigerian)

HLA_DR/DQ – Sum of T1D risk at HLA-DR/DQ haplotypes with interaction model

HLA_Class_1 – Sum of T1D risk at HLA class 1 alleles

HLA_Class_2 – Sum of T1D risk at HLA class 2 excluding DR/DQ haplotypes

Non-HLA – Sum of T1D risk at from variants outside the HLA region, genome-wide

T1D-GRS – Complete T1D genetic risk score (67 variants, Sharp et al)

T2D-GRS – Complete T2D genetic risk score (338 variants, Mahajan et al)

BETA_CELL – Partitioned risk score cluster encoding beta cell variants (DiCorpo et al)

PROINSULIN – Cluster encoding proinsulin action variants (DiCorpo et al)

OBESITY – Cluster encoding obesity related variants (DiCorpo et al)

LIPODYSTROPHY – Cluster encoding lipodystrophy related variants (DiCorpo et al)

LIVER_LIPID – Cluster encoding liver / lipid related variants (DiCorpo et al)

T1D-GRS% – Centile of risk T1D-GRS, derived relative to all samples in the file

T2D-GRS% – Centile of risk T2D-GRS, derived relative to all samples in the file

T1D-GRS_norm – T1D-GRS min-max scaled relative to all samples in the file

T2D-GRS_norm – T2D-GRS min-max scaled relative to all samples in the file