Supplemental Information for

PGG.MHC: towards understanding the diversity of major histocompatibility complexes in global populations

This PDF file contains:
Supplementary Table S1
Supplementary Table S2
Supplementary Table S3
Supplementary Figure S1

Table S1. Detailed information about *PGG*.MHC data sources

Data source	Abbreviation	Data type	Sample size	Coverage	Genomic coordinate
The 1000 Genomes Project	KGP	WGS	1,453	35.69	GRCh38
The Human Genome Diversity Project	HGDP	WGS	761	36.75	GRCh38
Simons Genome Diversity Project	SGDP	WGS	279	42.32	GRCh37
HuaBiao	HuaBiao	WES	5,002	73.7	GRCh38
PGG	PGG	WGS	3,561	~23.03-35.65	GRCh37/ GRCh38
		Genotyping	42,198	Not available	GRCh37

Table S2. HLA typing results and benchmark typing results for KGP samples

Sample ID	Source	HLA-A_1	HLA-A_2	HLA-B_1	HLA-B_2	HLA-C_1	HLA-C_2	HLA-DQB1_1	HLA-DQB1_2	HLA-DRB1_1	HLA-DRB1_2
NA18528	Genotyping	A*11:01	A*31:01	B*13:01	B*35:03	C*03:04	C*06:02	DQB1*05:02	DQB1*05:03	DRB1*14:01	DRB1*14:04
	WGS	A*11:01	A*31:01	B*13:01	B*35:03	C*03:04	C*06:02	DQB1*05:02	DQB1*05:03	DRB1*14:01	DRB1*14:04
	WES	A*11:01	A*31:01	B*13:01	B*35:03	C*03:04	C*06:02	DQB1*05:02	DQB1*05:02	DRB1*14:01	DRB1*14:54
	Benchmark	A*11:01	A*31:01	B*13:01	B*35:03	C*03:04	C*06:02	DQB1*05:02	DQB1*05:03	DRB1*14:04	DRB1*14:04
NA18538	Genotyping	A*24:02	A*33:03	B*15:01	B*58:01	C*03:02	C*04:01	DQB1*03:02	DQB1*06:09	DRB1*04:06	DRB1*13:02
	WGS	A*24:02	A*33:03	B*15:01	B*58:01	C*03:02	C*04:01	DQB1*03:02	DQB1*06:09	DRB1*04:06	DRB1*13:02
	WES	A*24:02	A*33:03	B*15:01	B*58:01	C*03:02	C*04:01	DQB1*03:02	DQB1*06:09	DRB1*04:06	DRB1*13:02
	Benchmark	A*24:02	A*33:03	B*15:01	B*58:01	C*03:02	C*04:01	DQB1*03:02	DQB1*06:09	DRB1*04:06	DRB1*13:02
NA18539	Genotyping	A*02:01	A*26:01	B*39:01	B*54:01	C*01:02	C*07:02	DQB1*04:01	DQB1*06:01	DRB1*04:05	DRB1*08:03
	WGS	A*02:01	A*26:01	B*39:01	B*54:01	C*01:02	C*07:02	DQB1*04:01	DQB1*06:01	DRB1*04:05	DRB1*08:03
	WES	A*02:01	A*26:01	B*39:01	B*54:01	C*01:02	C*07:02	DQB1*04:01	DQB1*06:01	DRB1*04:05	DRB1*08:03
	Benchmark	A*02:01	A*26:01	B*39:01	B*54:01	C*01:02	C*07:02	DQB1*04:01	DQB1*06:01	DRB1*04:05	DRB1*08:03
NA18541	Genotyping	A*11:01	A*30:04	B*14:01	B*51:01	C*08:02	C*14:02	DQB1*04:02	DQB1*05:02	DRB1*04:04	DRB1*16:02
	WGS	A*11:01	A*30:04	B*14:01	B*51:01	C*08:02	C*14:02	DQB1*05:02	DQB1*04:02	DRB1*04:04	DRB1*16:02
	WES	A*11:01	A*30:04	B*14:01	B*51:01	C*08:02	C*14:02	DQB1*04:02	DQB1*05:02	DRB1*04:04	DRB1*16:02
	Benchmark	A*11:01	A*30:04	B*14:01	B*51:01	C*08:02	C*14:02	DQB1*05:02	DQB1*04:02	DRB1*04:04	DRB1*16:02
HG00403	Genotyping	A*02:01	A*31:01	B*13:01	B*35:01	C*03:03	C*03:04	DQB1*03:01	DQB1*06:02	DRB1*12:02	DRB1*15:01
	WGS	A*02:01	A*31:01	B*13:01	B*35:01	C*03:03	C*03:04	DQB1*03:01	DQB1*06:02	DRB1*12:02	DRB1*15:01
	WES	A*02:01	A*31:01	B*13:01	B*35:01	C*03:03	C*03:04	DQB1*03:01	DQB1*06:02	DRB1*15:01	DRB1*15:01
	Benchmark	A*02:01	A*31:01	B*13:01	B*35:01	C*03:03	C*03:04	DQB1*03:01	DQB1*06:02	DRB1*12:02	DRB1*15:01
HG00559	Genotyping	A*24:02	A*33:03	B*55:02	B*58:01	C*01:02	C*03:02	DQB1*02:01	DQB1*03:03	DRB1*03:01	DRB1*09:01
	WGS	A*24:02	A*33:03	B*55:02	B*58:01	C*01:02	C*03:02	DQB1*02:01	DQB1*03:03	DRB1*03:01	DRB1*09:01
	WES	A*24:02	A*33:03	B*55:02	B*58:01	C*01:02	C*03:02	DQB1*02:01	DQB1*03:03	DRB1*03:01	DRB1*09:01
	Benchmark	A*24:02	A*33:03	B*55:02	B*58:01	C*01:02	C*03:02	DQB1*02:01	DQB1*03:03	DRB1*03:01	DRB1*09:01
HG00565	Genotyping	A*02:03	A*33:03	B*40:01	B*40:01	C*03:04	C*07:02	DQB1*05:02	DQB1*05:02	DRB1*16:02	DRB1*16:02
	WGS	A*02:03	A*33:03	B*40:01	B*40:01	C*03:04	C*07:02	DQB1*05:02	DQB1*05:02	DRB1*16:02	DRB1*16:02
	WES	A*02:03	A*33:03	B*40:01	B*40:01	C*03:04	C*07:02	DQB1*03:22	DQB1*05:02	DRB1*16:01	DRB1*16:02
	Benchmark	A*02:03	A*33:03	B*40:01	B*40:01	03:04	C*07:02	DQB1*05:02	DQB1*05:02	DRB1*16:02	DRB1*16:02
HG00566	Genotyping	A*02:07	A*24:02	B*46:01	B*58:01	C*01:02	C*03:02	DQB1*02:01	DQB1*03:03	DRB1*03:01	DRB1*09:01
	WGS	A*02:07	A*24:02	B*46:01	B*58:01	C*01:02	C*03:02	DQB1*02:01	DQB1*03:03	DRB1*03:01	DRB1*09:01
	WES	A*02:07	A*24:02	B*46:01	B*58:01	C*01:02	C*03:02	DQB1*02:01	DQB1*03:03	DRB1*03:01	DRB1*09:01
	Benchmark	A*02:07	A*24:02	B*46:01	B*58:01	C*01:02	C*03:02	DQB1*02:01	DQB1*03:03	DRB1*03:01	DRB1*09:01

Table S3. Imputation reference panels in *PGG*.MHC

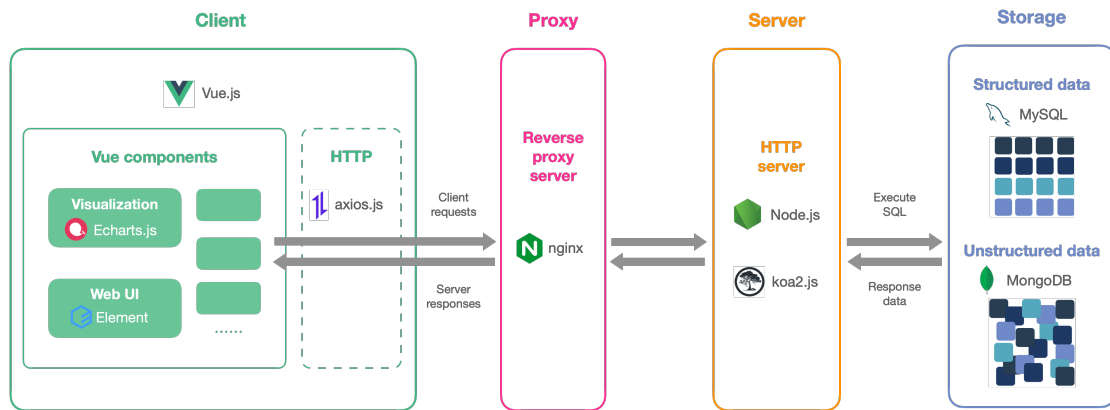| Reference panel	Sample region	Sample size	Number of classical HLA alleles	Number of SNPs in HLA region
1000G_REF	AFR, AMR, EAS, EUR, and SAS	2,504	465	5,067
Pan-Asian panel	EAS and SAS	530	273	6,173
Korean panel	EAS	413	251	5,858
HuaBiao_REF	EAS	5,002	485	4,393
PGG_REF	EAS	1,835	981	13,350

AFR (Africa), AMR (America), EAS (East Asia), SAS (South Asia), EUR (Europe)

Figure S1. Database architecture