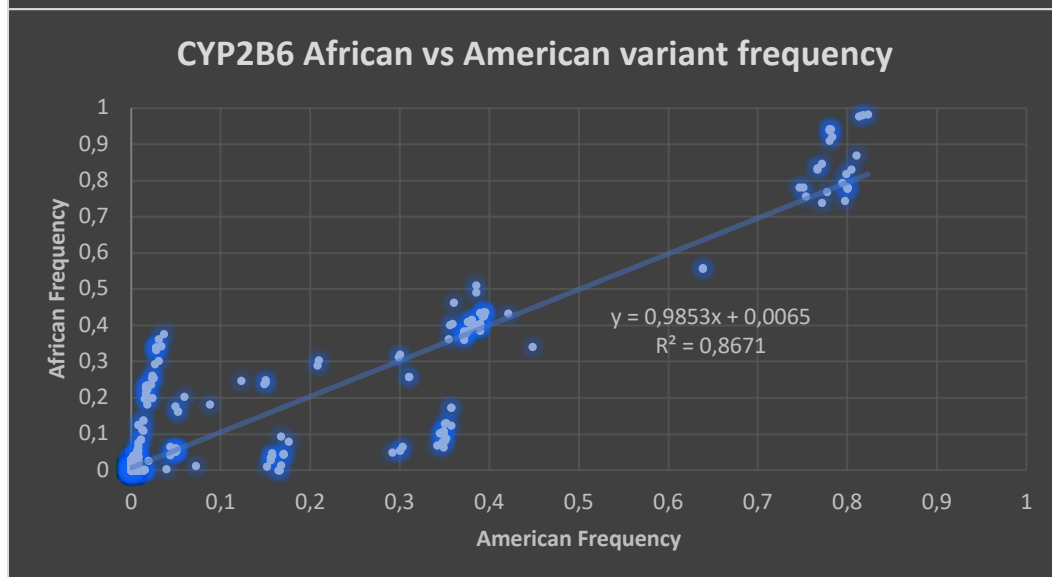
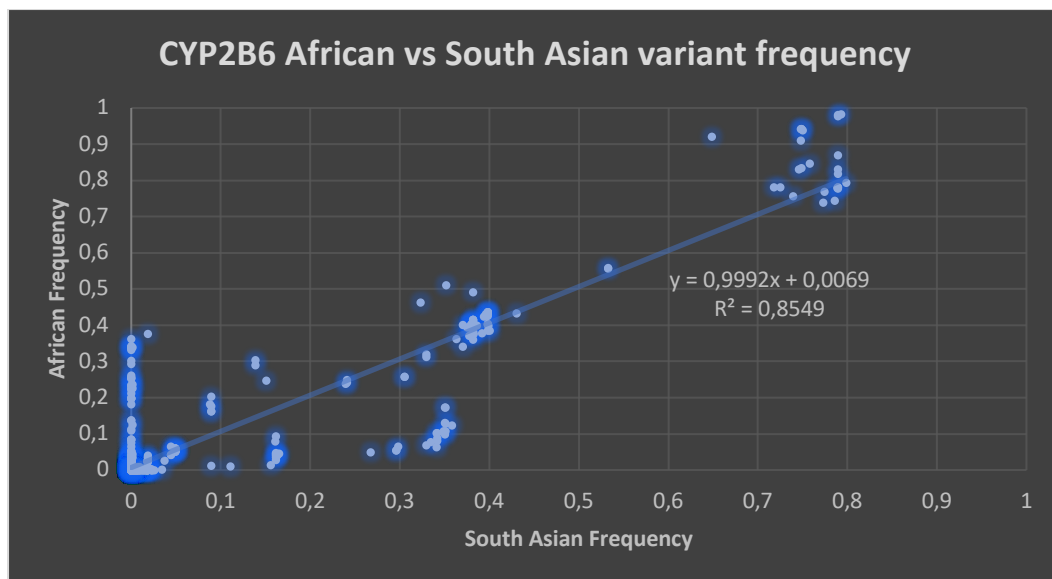


Figure S1 – African CYP2A6 variant frequencies compared to comparison super populations.

Footnote:

A linear trendline (blue) has been calculated to describe the relation between variant frequency in Africans as a product of previously studied superpopulations. Variance was calculated and indicated (R^2).



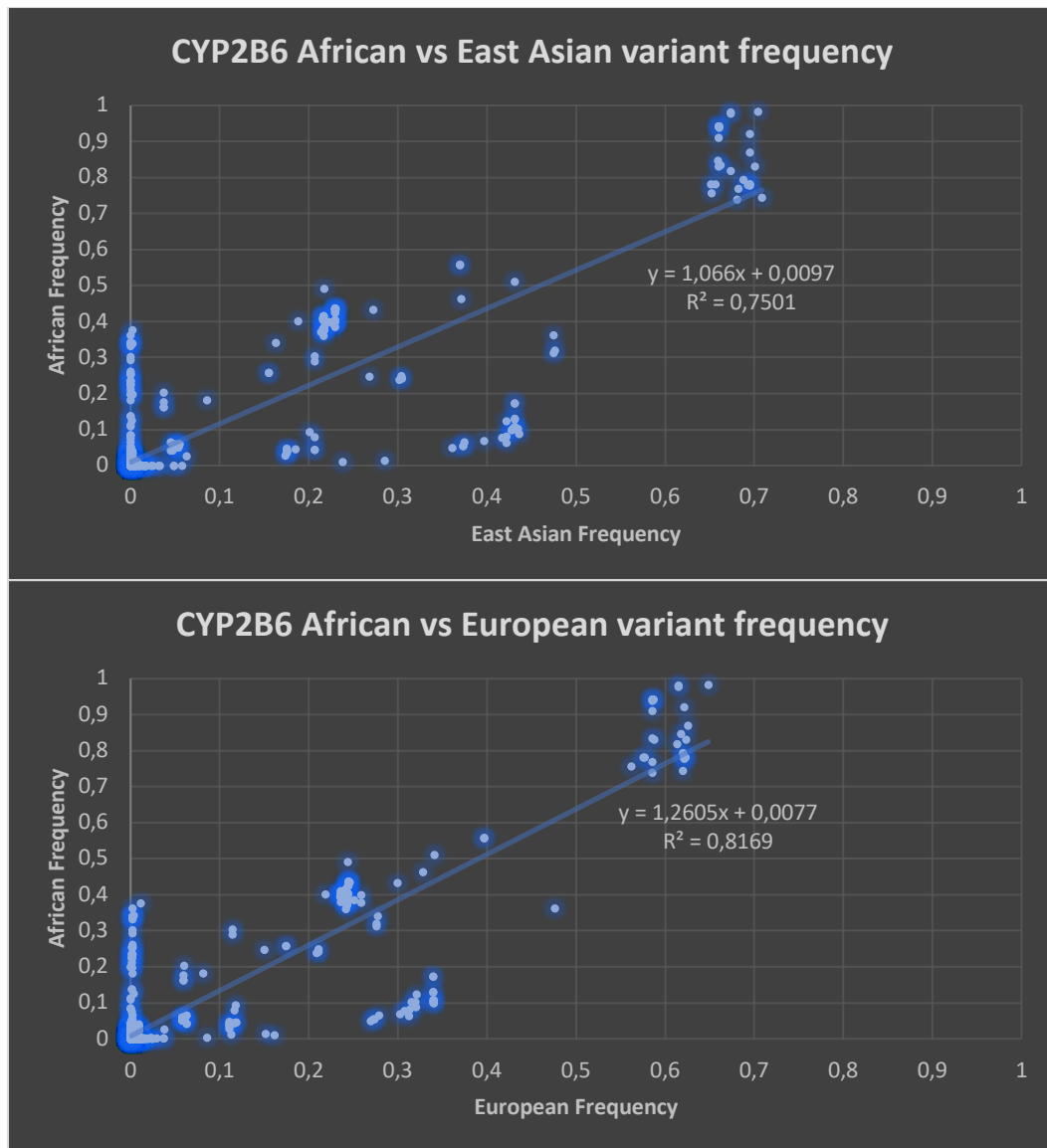
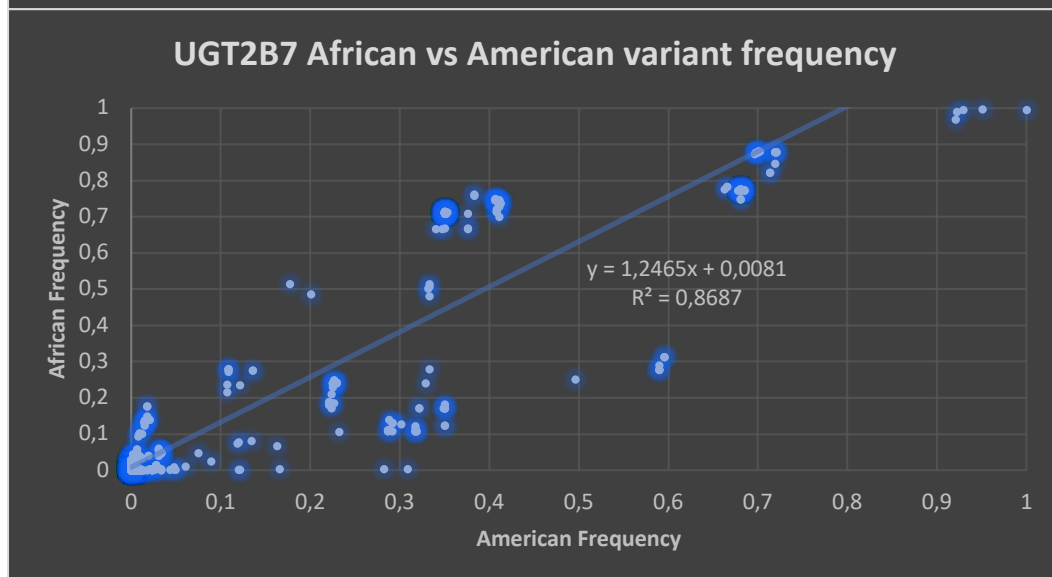
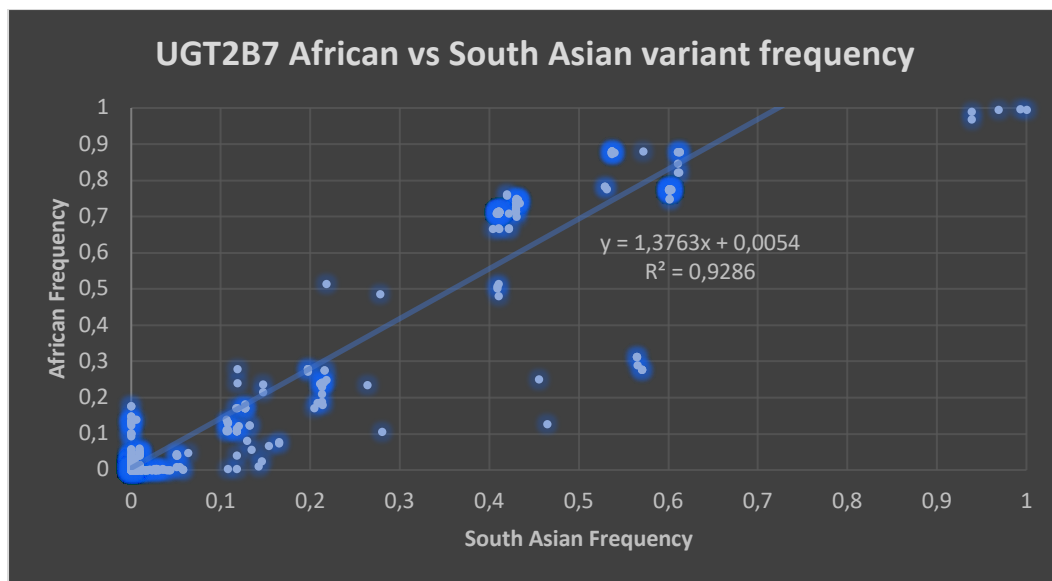


Figure S2 – African CYP2B6 variant frequencies compared to comparison super populations.

Footnote:

A linear trendline (blue) has been calculated to describe the relation between variant frequency in Africans as a product of previously studied superpopulations. Variance was calculated and indicated (R^2).



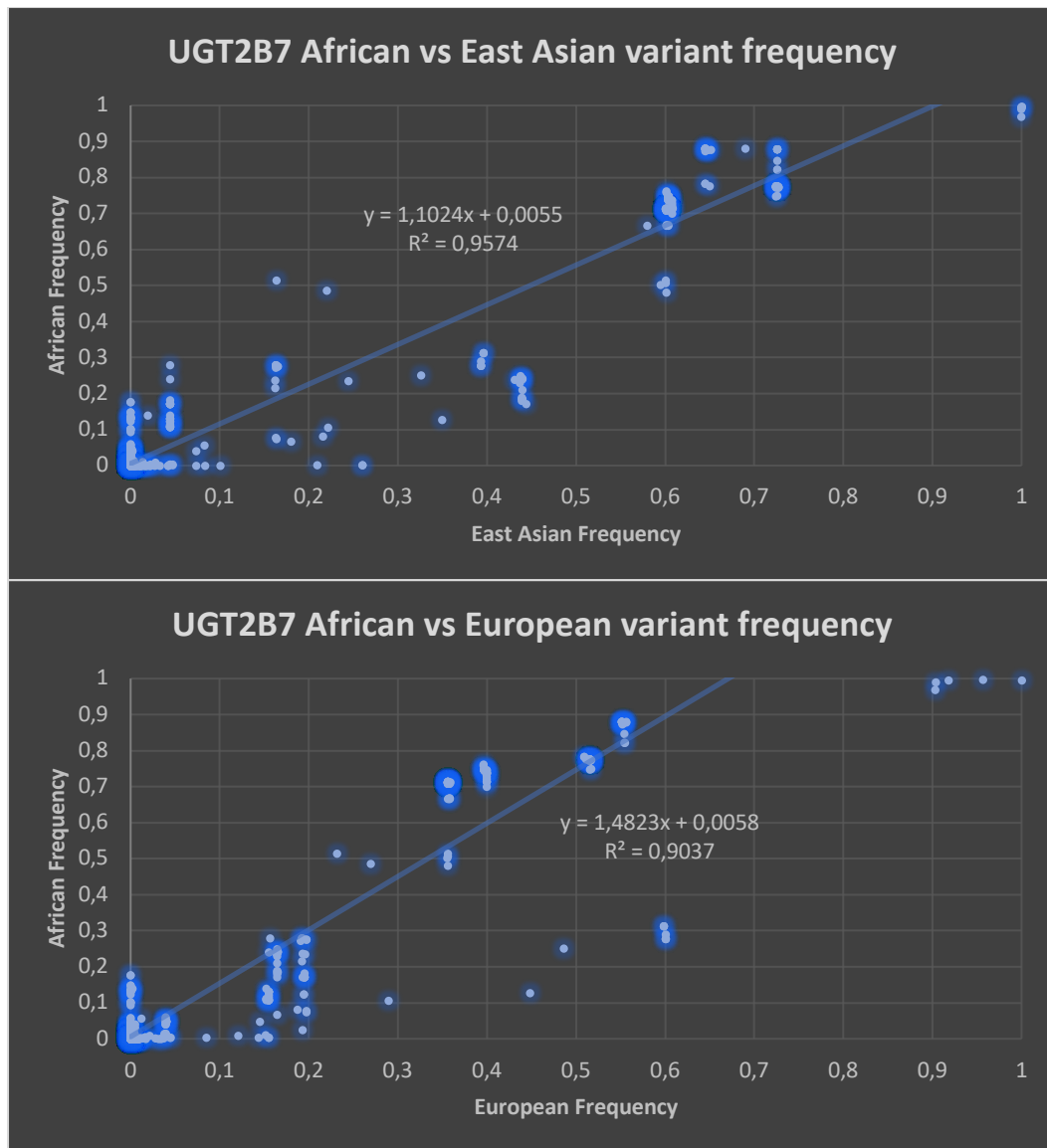


Figure S3 - African UGT2B7 variant frequencies compared to comparison super populations.

Footnote:

A linear trendline (blue) has been calculated to describe the relation between variant frequency in Africans as a product of previously studied superpopulations. Variance was calculated and indicated (R^2).

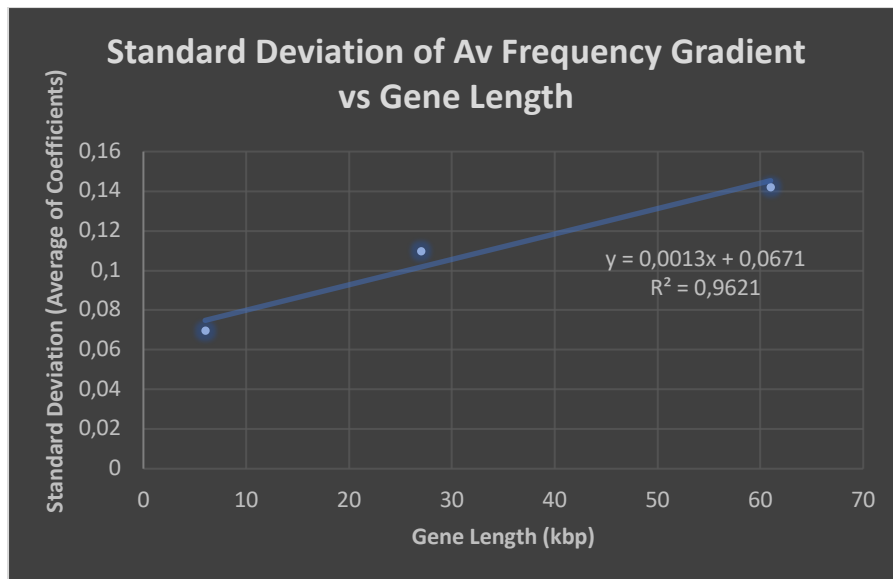


Figure S4 - Standard Deviation of Average Frequency Gradient vs Gene Length.

Footnote:

A trendline indicating the linear relationship between the average Standard Deviation observed in each gene when characterising variant frequency as a product of previously studied super populations, and the length of the Gene studied (In kilobase pairs).

DATASET:	DATASET COUNT:	SUPER POPULATION:	POPULATION COUNT:	SUBPOPULATION:	SAMPLE COUNT:
1000 GENOMES	2 504	AFR	661	ACB	96
				ASW	61
				ESN	99
				GWD	113
				LWK	99
				MSL	85
				YRI	108
		EUR	503	CEU	99
				FIN	99
				GBR	91
				IBS	107
				TSI	107
		AMR	347	CLM	94
				MXL	64
				PEL	85
				PUR	104
		EAS	504	CDX	93
				CHB	103
				CHS	105
				JPT	104
				KHV	99
AGVP	1 307	AFR	1 307	BEB	86
				GIH	103
				ITU	102
				PJL	96
				STU	102
				BAG	100
				BAN	100
				BAR	97
				ETH	107
				FUL	74
				GAA	100

				IGB	99
				JOL	79
				KAL	100
				KIK	99
				MAN	88
				SOT	86
				WOL	78
				ZUL	100
SAHGP	24	AFR	24	COL	8
				SOT	8
				XHO	8
ALL	3 835	AFR	1 992		
		AMR	347		
		EUR	503		
		EAS	504		

Table S1 - Sample composition by data source.

Footer:

Red blocks indicate subpopulations which were removed from analysis.

	Criteria	American	European	South Asian	East Asian
UGT2B6 CYP2A6	>=1	26	10	17	13
	<1	22	22	30	20
	TOTAL	48	32	47	33
UGT2B6 CYP2B6	>=1	50	71	20	70
	<1	40	43	49	39
	TOTAL	90	114	69	109
UGT2B7	>=1	16	18	15	18
	<1	7	6	6	5
	TOTAL	23	24	21	23

Table S2 - A partition breakdown of alleles identified with previously documented disease associations (E! Ensemble VEP disease association) as well as significant Fishers Exact Odds Ratios between African populations and each comparison population.

	Criteria	American	European	South Asian	East Asian
UGT2B6 CYP2A6	>=1	26	10	17	13
	<1	22	22	30	20
	TOTAL	48	32	47	33
UGT2B6 CYP2B6	>=1	50	71	20	70
	<1	40	43	49	39
	TOTAL	90	114	69	109
UGT2B7	>=1	16	18	15	18
	<1	7	6	6	5
	TOTAL	23	24	21	23

Table S3 – Alleles with CONDEL predictions and known phenotype associations found at a significantly ($\alpha = 0.05$) different frequency when compared to at least one comparison population.