Supplemental material to:

Nonsynonymous SNPs in *LPA* homologous to plasminogen deficiency mutants represent novel null apo(a) alleles

Benjamin M. Morgan[1], Aimee N. Brown[1], Nikita Deo[1,2], Tom W.R. Harrop[1], George Taiaroa[1] Peter D. Mace[1,2], Sigurd M. Wilbanks[1], Tony R. Merriman[1,2], Michael J.A. Williams[3] and Sally P.A. McCormick[1,2*]

[1]Department of Biochemistry, School of Biomedical Sciences, University of Otago, Dunedin, New Zealand.

[2] Maurice Wilkins Centre for Molecular Biodiscovery, Private Bag 92019, Auckland 1142, New Zealand.

[3]Department of Medicine, Dunedin School of Medicine, University of Otago, Dunedin, New Zealand.

* To whom correspondence should be addressed: Sally McCormick, Department of Biochemistry, University of Otago, 710 Cumberland Street, Dunedin 9016, New Zealand. Phone: (64 3) 479 7840; Fax: (64 3) 479 7866; E-mail: sally.mccormick@otago.ac.nz

Supplemental methods

Bioinformatic Analysis

The paired end read pools were exported and demultiplexed into sample pools using Illumina bcl2fastq script, which separates reads using the unique Illumina identification index and then into sample amplicon pools. A custom in-house script was used to remove Illumina index sequences from the end of reads. The script simultaneously removed bases with low quality call scores (<0.99) from the ends of reads and discarded significantly shortened reads. The remainder of the bioinformatics analysis was performed in Geneious (1). Paired-end reads were matched and then merged using FLASH (2). Sample amplicon pools were mapped directly onto each amplicon reference sequence, as derived from the NCBI *LPA* gene reference sequence DQ452068. When mapping to KIV-2, the 'reference' sequence was derived from the first KIV-2 repeat. Standard variant calling was performed for single copy regions of *LPA*. Reads encompassing KIV-2 underwent a MUSCLE (3) alignment before variant calling. This grouped reads based on sequence similarity, aiding in the calling of low frequency variants. Variants were called down to a frequency of 0.01 to allow theoretical calling of a SNP present in a single repeat within a sample carrying the maximum number of 80 repeats.

Nomenclature of Variants

The SNPs in the non-repeat kringle domains were assigned numbering based on the *LPA* reference sequence DQ452068 and the corresponding protein sequence NP_005568 which contain six KIV-2 repeats. Bases are given according to the reverse strand, which is the coding strand of *LPA*. With respect to the KIV-2 repeat, the numbering scheme previously used by Coassin et al (4) was employed in this study, which numbers variants from the first amino acid that is fully encoded by exon 1 of the first KIV-2 repeat.

Supplemental tables

Supplemental Table S1: Demographic information for the Otago LPA population

Variable	
n	705
Europeans	89%*
% male	39%
Age (mean, years)	46.5 (19)
Heart disease (%)	7.5% (96%)
Diabetes (%)	2.7% (79%)
Hypertension (%)	16.5% (89%)
Dyslipidemia, (%)	32% (70%)
Total cholesterol (mean, mmol/L)	5.03 (1.16)
HDL cholesterol (mean, mmol/L)	1.58 (0.43)
LDL cholesterol (mean, mmol/L)	2.95 (1.01)
Lp(a), (median, mg/dL)	14.6 (40)
Lp(a), <3 mg/dL, n	79
Lp(a), 3-50 mg/dL, n	435
Lp(a), >50 mg/dL, n	191

Numbers in brackets indicate standard deviation, except in the case of the 4 disease categories in which case they represent the percentage of individuals on medication. The most common medications being used were statins, aspirin and β-blockers for heart disease; aspirin, ACE inhibitors, metformin and statins for diabetes; ACE inhibitors, statins and aspirin for hypertension and statins, aspirin and ACE inhibitors for dyslipidemia.
*The remaining 11% consisted of mainly Asian and Maori ethnicity.

Supplemental Table S2: Selection criteria for null samples

Population	LPA	GOUT
Size (n)	705	288
Selection Criteria		
1. Lp(a) <3 mg/dL (n)	79	51
2. Null on apo(a) phenotyping (n)	21	6
NGS sequenced (n)	5	6
3. SNPs associated with low Lp(a)	2	4

Supplemental Table S3: Primer sequences for R1771C apo(a)-GFP vector mutagenesis and apo(a) qPCR analysis

Primer name	Sequence 5'-3'	Tm C°
R1771C Forward	GGT CTG GAA AAA AAT TAC TGC TGT AAC CCT GAT GGT GAC A	60
R1771C Reverse	TGT CAC CAT CAG GGT TAC AGC AGT AAT TTT TTT CCA GAC C	60
Exon 33/34 Forward	TGA CAT CAA TGG TCC CTG GTG	49
Exon 33/34 Reverse	ACT TGA GGC TTC CCA CAA TCA	47
Neo Forward	TCC GGT GCC CTG AAT GAA CT	49
Neo Reverse	TAC TTT CTC GGC AGG AGC AA	47

Supplemental Table S4: R1771C mutagenesis protocol

Reagent			**Concentration**
dNTPs			300 μM
Primer (F/R)			0.1 μM (each primer)
DNA			75 ng
Pfu 10x Buffer			1x
Pfu polymerase			3U
RNAse-free H_2O			20 μL final vol.
Cycling conditions			
Step	Temperature (C°)	Duration	n Cycles
Initial denaturation	95	2 minutes	1
Denaturation	95	40 seconds	
Annealing	55	50 seconds	
Extension	68	20 minutes	25
Final extension	68	10 minutes	1

Supplemental Table S5: Effect of rs41259144 on Lp(a) levels in previously published populations

SNP	Gene	cDNA	Protein	GMAF	Cohort	No.	Ethnicity	MAF	Effect	Significance	Study
rs41259144	g.70301G>A	c.2969G>A	p.Arg990Gln	0.0052	EPIC Norfolk cohort	4101	European	0.01	-35%(-49,-21)[a]	3.38E-05	Erqou 2010 (5)
					Colaus	6182	Swiss Caucasian	0.006-0.012	-0.94(0.203)[b]	3.73E-06	Mack et al. 2017 (6)
					KORAF3	3076	European (Augsburg Germany)		-0.79(0.165)[b]	2.02E-06	
					KORAF4	2926			-0.79(0.118)[b]	2.35E-11	
					YFS Study	2103	Finnish		-1.08(0.256)[b]	2.57E-05	
					Single SNP Meta analysis				-0.85(0.084)[b]	4.45E-24	
					Joint SNP Meta analysis				-0.6(0.085)[c]	2.01E-12	
					FINRISK	2690	Finnish	0.0068	-0.57(0.0968)[d]	3.95E-09	Zekavat et al. 2018 (7)
					Copenhagen Heart Study			0.011	-12.8(1.2)[e]		Burgess et al. 2018 (8)
					Copenhagen General Population Study				-9.6(0.8)[f]		
					EPIC-CVD PROSPER WOSCOPS	48333	European			p<5E-08	

Gene= NG_016147.1, cDNA=NM_005577.2, Protein=NP_005568.2, a=Effect on natural log-transformed Lp(a) (95%CI), b=β inverse normal transformed Lp(a) (SE), c= β Inverse normal transformed Lp(a) adjusted for 49 additional independent SNPS (SE), d=β Inverse-rank normalized Lp(a) conditional on KIV-2 CNV (SE), e= β Lp(a) mg/dL-1 (SE) marginal association, f=β Lp(a) mg/dL-1 (SE) association conditional on all other variants in the published analysis.

Supplemental Table S6: Effect of rs139145675 on Lp(a) levels in previously published populations

SNP	Gene	cDNA	Protein	GMAF	Cohort	No.	Ethnicity	MAF	Effect	Significance	Study
rs139145675	g.125849C>T	c.5311C>T	p.Arg1771Cys	0.0044	Jackson Heart Study	2895	African American	0.00981	-0.47(0.12)[a]	1.35E-04	Li et al. 2015 (9)
					Jackson Heart Study	3418	African American	0.00971	-1.12(0.117)[b]	8.45E-22	Zekavat et al. 2018 (7)
						9272 Lp(a) levels from JHS and and Finnish imputation dataset					
					Copenhagen Heart Study			0.001	-8.6(3.7)[c]		Burgess et al. 2018 (8)
					Copenhagen General Population Study				-22.5(2.4)[d]		
					EPIC-CVD PROSPER WOSCOPS	48333	European			p<5E-08	

Gene= NG_016147.1, cDNA=NM_005577.2, Protein=NP_005568.2, a=β natural log-transformed Lp(a)(SE), b=β Inverse-rank normalized Lp(a) conditional on KIV-2 CNV (SE), c= β Lp(a) mg/dL-1 (SE) marginal association d=β Lp(a) mg/dL-1 (SE) association conditional on other variants in the published analysis.

Supplemental Table S7: Effect of rs41272114 on Lp(a) levels in previously published populations

SNP	Gene	cDNA	Protein	GMAF	Cohort	No.	Ethnicity	MAF	Effect	Significance	Study
rs41272114	g.86331G>A	c.4289+1G>A	+1 splice	0.0234	EPIC Norfolk cohort	4050	European	0.035	-26%(-35,-17)[a]	4.82E-07	Erqou 2010 (5)
					Jackson Heart Study	2895	African American	0.0107	-0.8(0.11)[b]	6.50E-12	Li et al. 2015 (9)
					Colaus	6182	Swiss Caucasian	0.023-0.048	-0.87(0.092)[c]	2.87E-21	Mack et al. 2017 (6)
					FamHS	1200 families (~6000 individuals)			-0.82(0.112)[c]	2.81E-13	
					KORAF3	3076	German		-0.73(0.083)[c]	4.17E-18	
					KORAF4	2926	German		-0.76(0.092)[c]	1.21E-16	
					YFS Study	2103	Finnish		-0.75(0.071)[c]	1.48E-25	
					Single SNP Meta analysis				-0.78(0.04)[c]	4.74E-86	
					Joint SNP Meta analysis				-0.48(0.043)[d]	7.98E-29	
					FINRISK	2690	Finnish	0.0482	-0.652(-0.037)[e]	2.27E-69	Zekavat et al. 2018 (7)
					Jackson Heart Study	3418	African American	0.009	-0.697(0.119)[e]	4.49E-09	
						9272 Lp(a) levels from JHS and and Finnish imputation dataset					

Gene= NG_016147.1, cDNA=NM_005577.2, Protein=NP_005568.2, a=Effect on natural log-transformed Lp(a) (95%CI), b=β natural log-transformed Lp(a)(SE), c=β inverse normal transformed Lp(a) (SE), d= β Inverse normal transformed Lp(a) adjusted for 49 additional independent SNPS (SE), e= β Inverse-rank normalized Lp(a).

Supplemental Table S8. Individuals from LPA study population carrying R990Q mutation

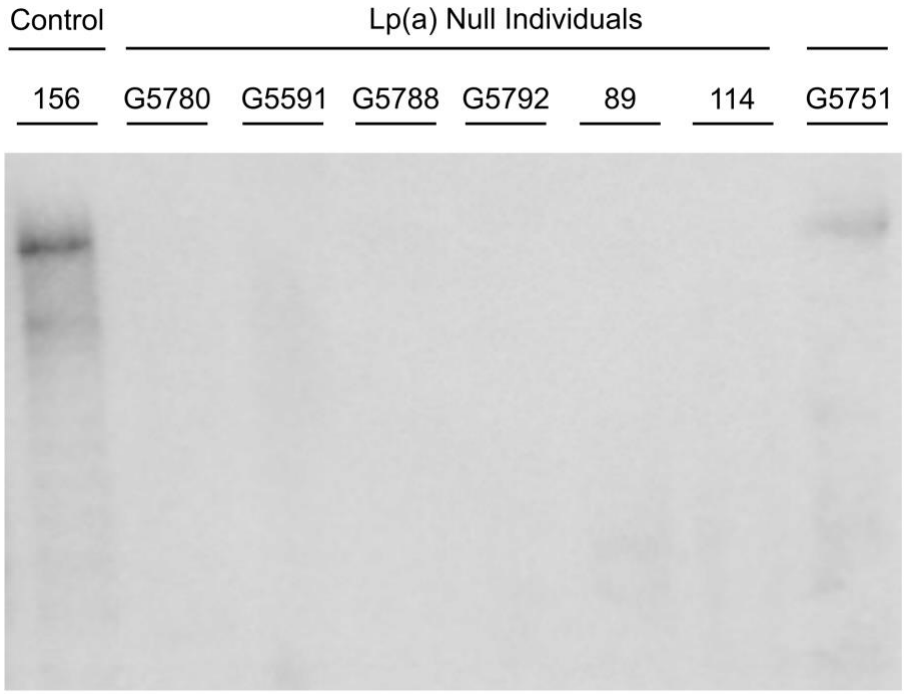
Participant	**338**	**268**	**163**	**154**	**91**
Age	54	51	63	66	47
Sex	F	M	F	M	F
CVD	No	No	No	Yes	Yes
CVD-family history	No	Yes	Yes	Yes	Yes
TC	5.9	5.7	5.5	3.9	7.2
TG	1	1.5	0.7	0.9	1.1
HDL-C	1.52	1.17	2.59	1.31	1.23
LDL-C	3.9	3.9	2.6	2.2	5.5
Lp(a)(mg/dL)	117	182	657	649	380
Apo(a) phenotype	24/0	25/0	19/0	20/0	22/0

Supplemental Table S9: qRT-PCR data for apo(a) exon 33/34 primer pair*

Sample	Target ct	Mean ct	Reference ct	Mean ct	Tgtct-Refct (Δct)
WT01	22.01		25.2		
WT02	21.94		25.2		
WT03	22.04	21.99666667	25.1	25.16666667	-3.17
R1771C01	22.26		25.66		
R1771C02	22.62		25.56		
R1771C03	22.46	22.44666667	25.64	25.62	-3.173333333
ΔR1771Cct-ΔWTct (ΔΔct)					-0.003333333
Fold change (2^-ΔΔct)					1.002313162

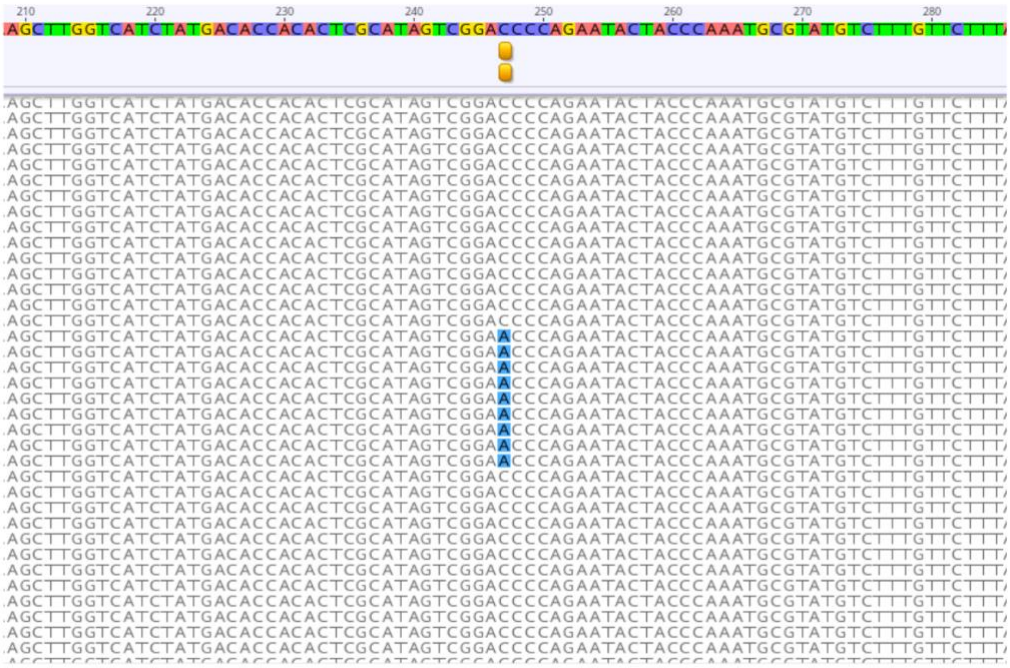
Samples are for three technical repeats using cDNA extracted from a single transfection experiment.
* In order to target dissimilar regions of apo(a) and prevent amplification of multiple regions of sequence, primers were designed to bind to regions spanning the junction between exons 33 and 34 of the protease domain.

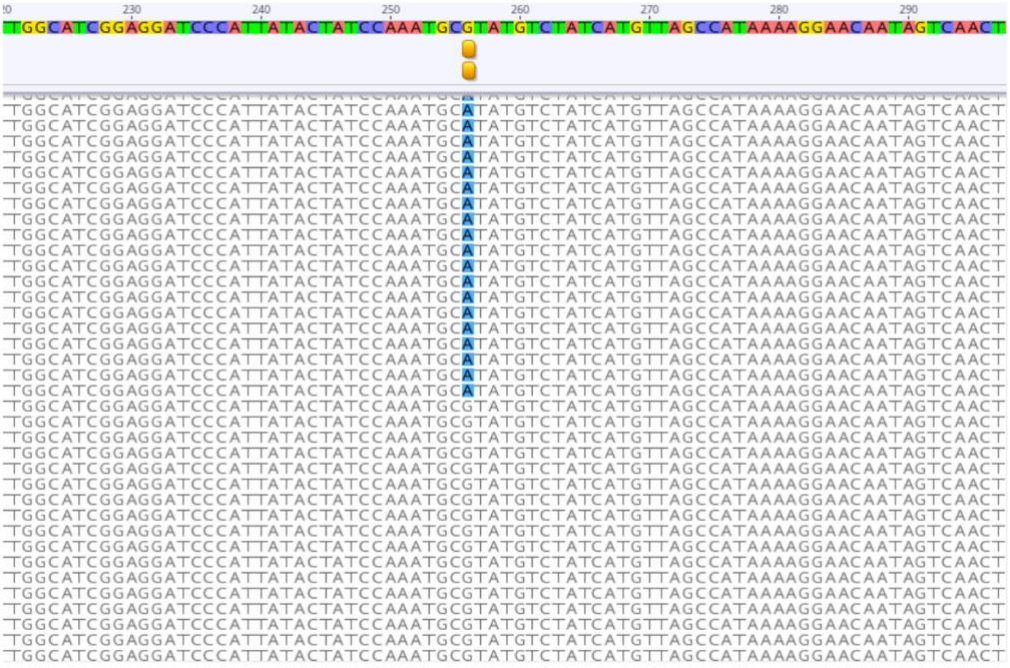
Supplemental Figures



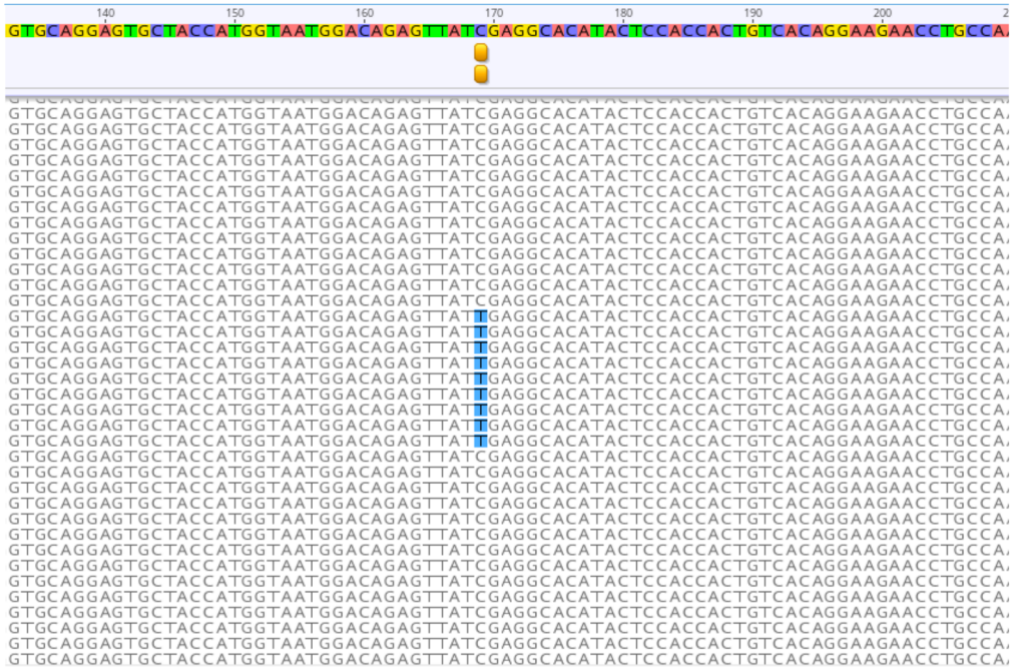
Supplemental Fig. S1: Apo(a) phenotyping of individuals showing no detectable Lp(a).

Plasma from individuals with no detectable Lp(a) was mixed with reducing buffer and subject to electrophoresis on 4% polyacrylamide gels. Proteins were transferred overnight to nitrocellulose membrane and incubated with a polyclonal anti-Lp(a) antibody. Following incubation with a HRP-labelled secondary antibody, membranes were imaged using ECL. Lane 1 contains a sample that exhibits a single apo(a) isoform size of 21 and was loaded to 590 ng of Lp(a). Lanes 2-8 each contain 7.5 µL of plasma samples from individuals classified as having an Lp(a) level of <3 mg/dL.

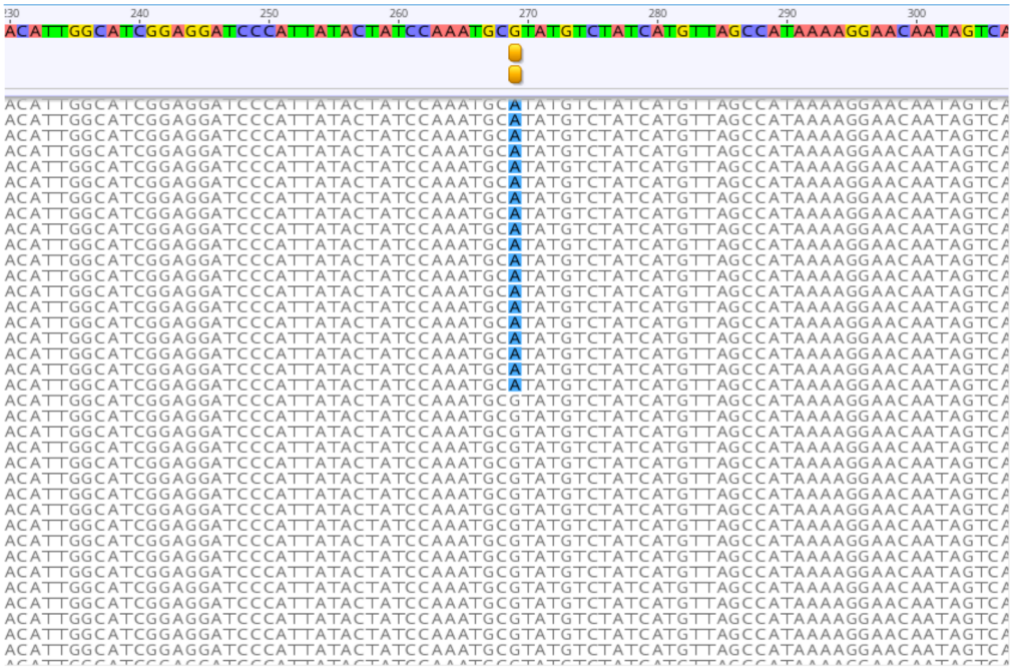




Supplemental Fig. S2: NGS reads of nonsynonymous SNPs in *LPA* identified in subject G5788.

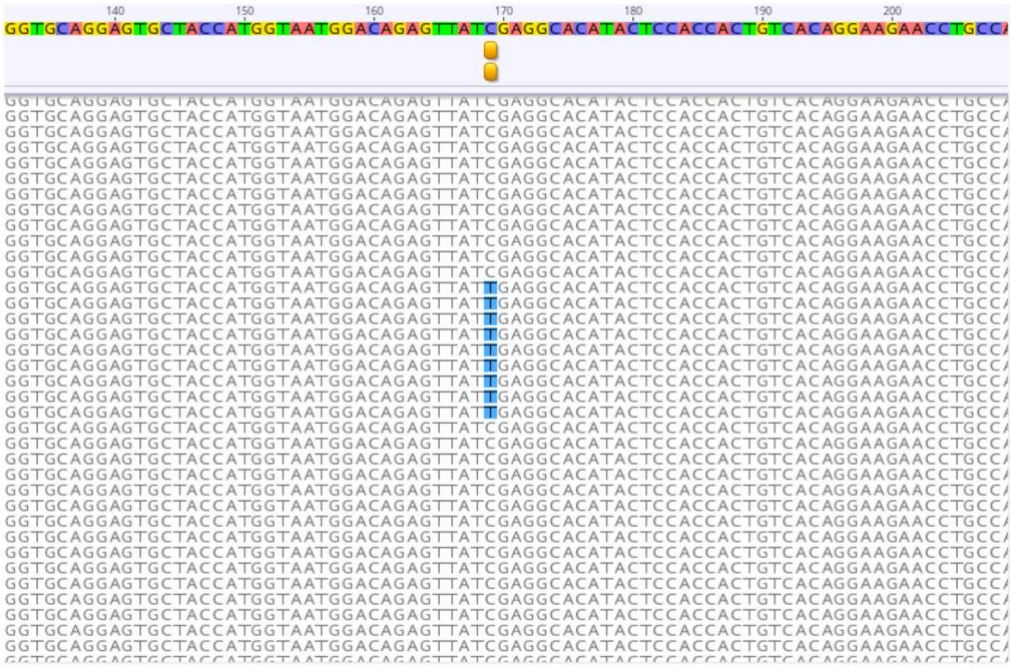


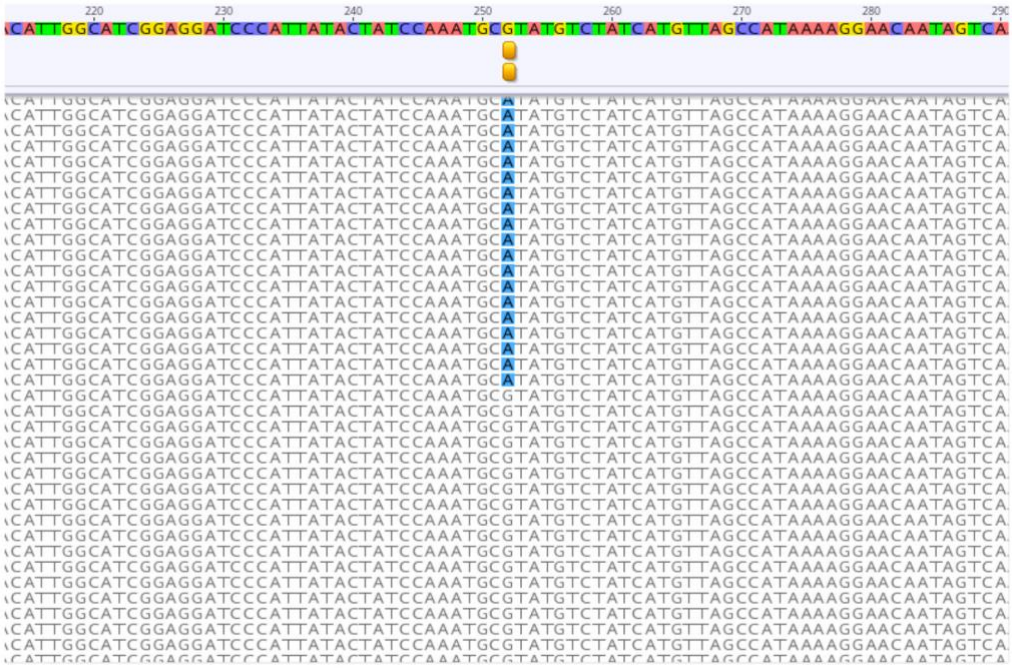
R20X (KIV-2)



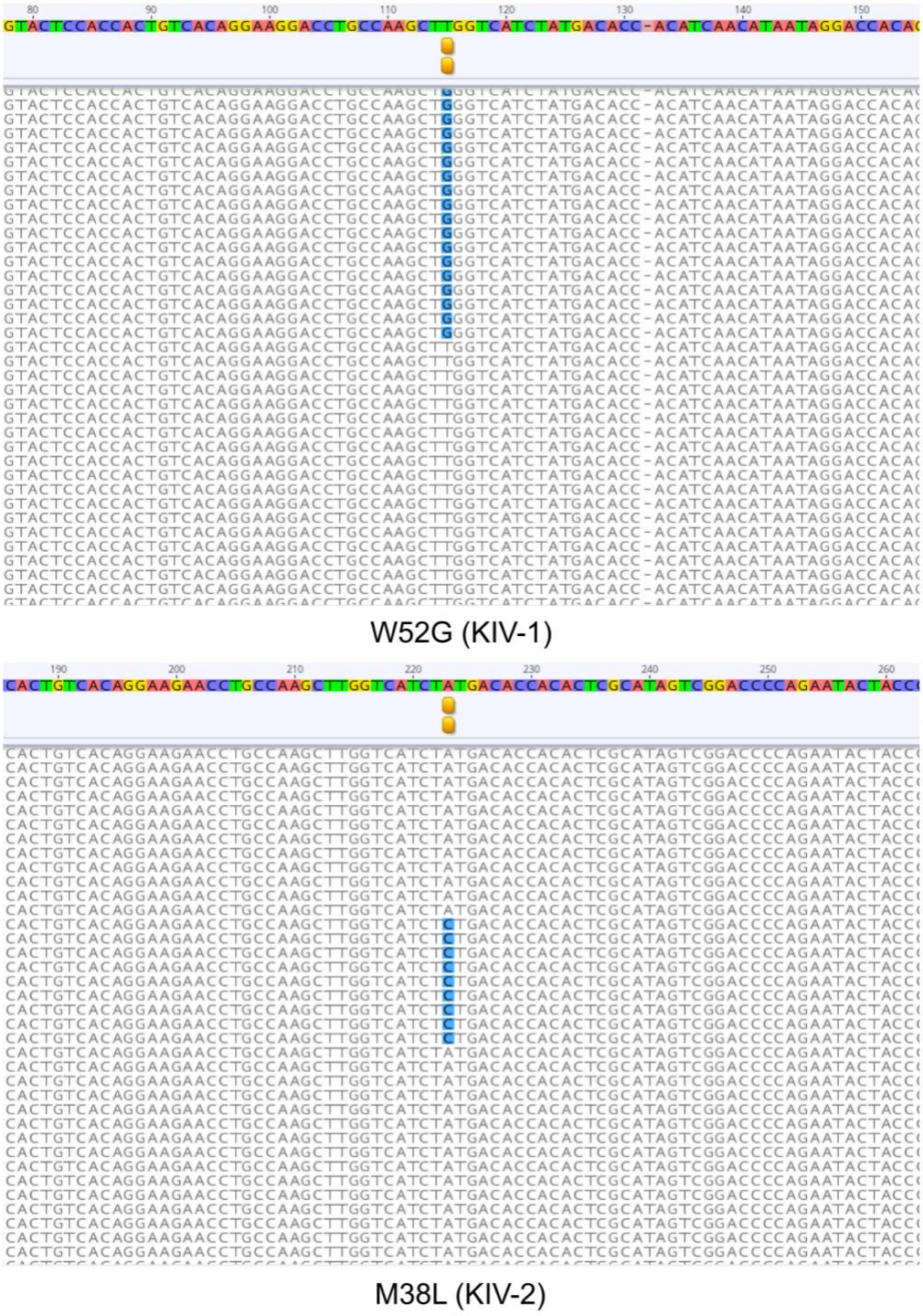
rs41272114 (KIV-8)

Supplemental Fig. S3: NGS reads of nonsynonymous SNPs in *LPA* identified in subject G5792.

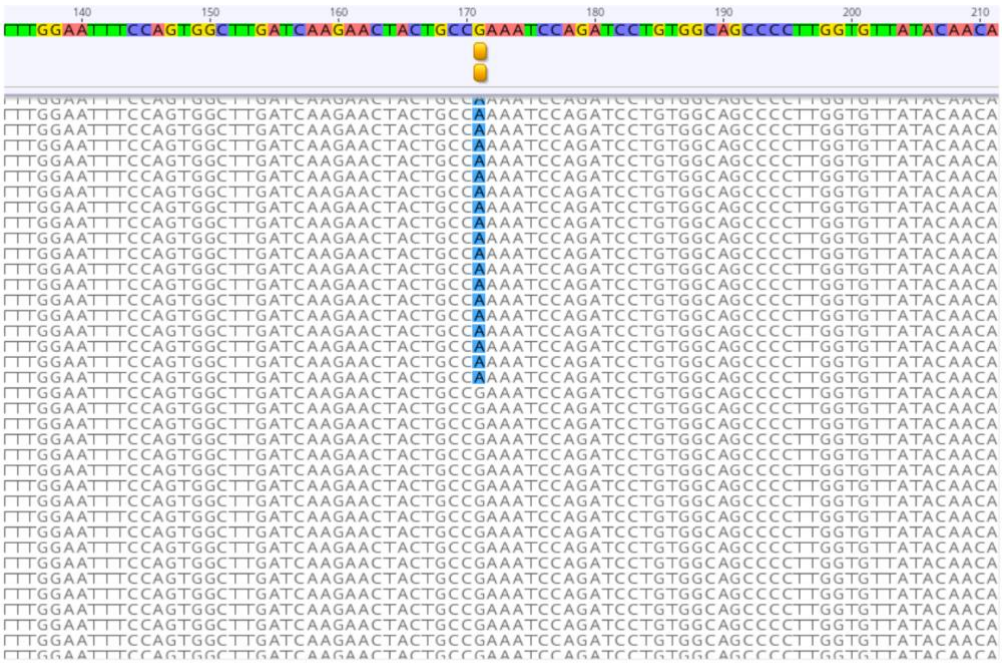




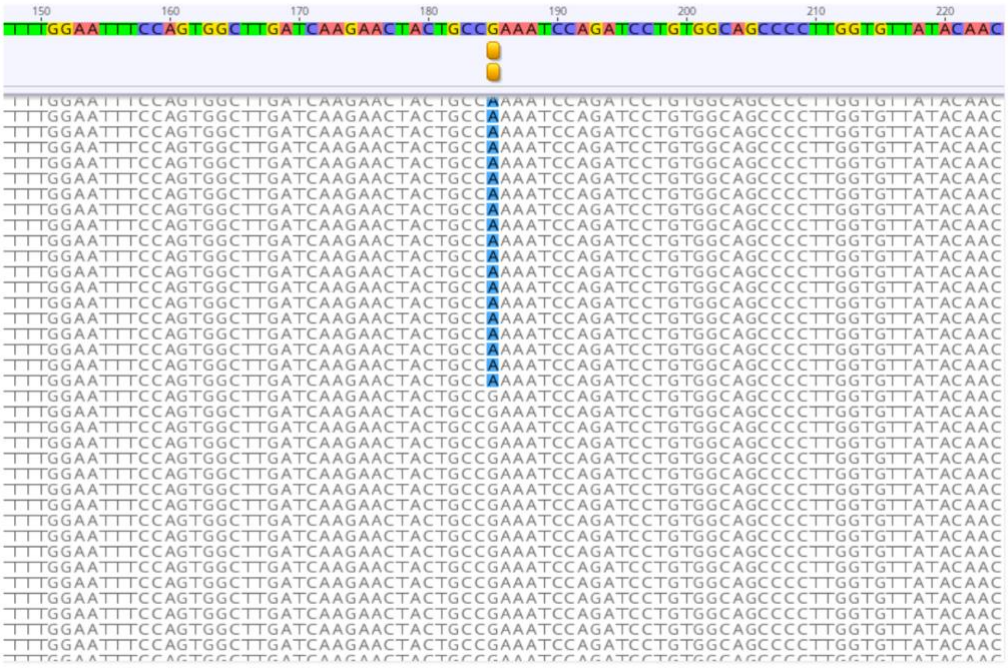
Supplemental Fig. S4: NGS reads of nonsynonymous SNPs in *LPA* identified in subject LPA114.



Supplemental Fig. S5: NGS reads of nonsynonymous SNPs in *LPA* identified in subject G5780.

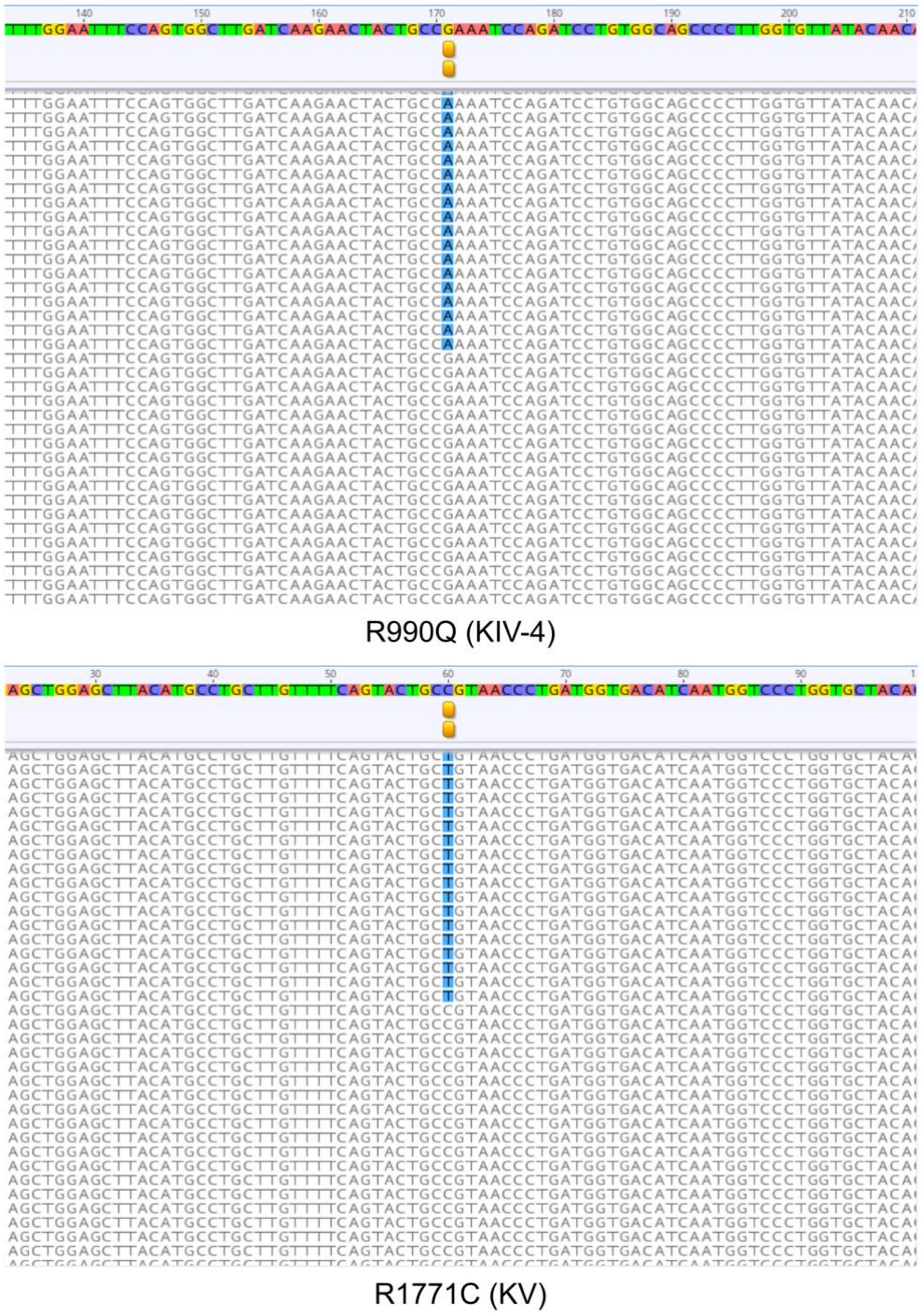


Supplemental Fig. S5 (cont): NGS reads of nonsynonymous SNPs in *LPA* identified in subject G5780.

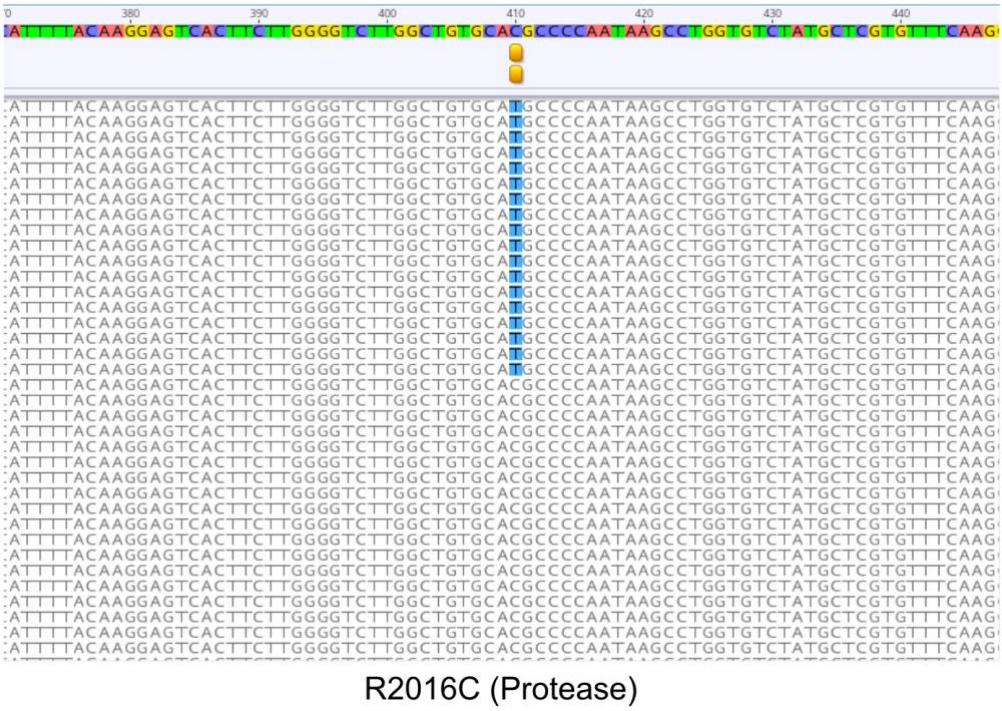


R990Q (KIV-4)

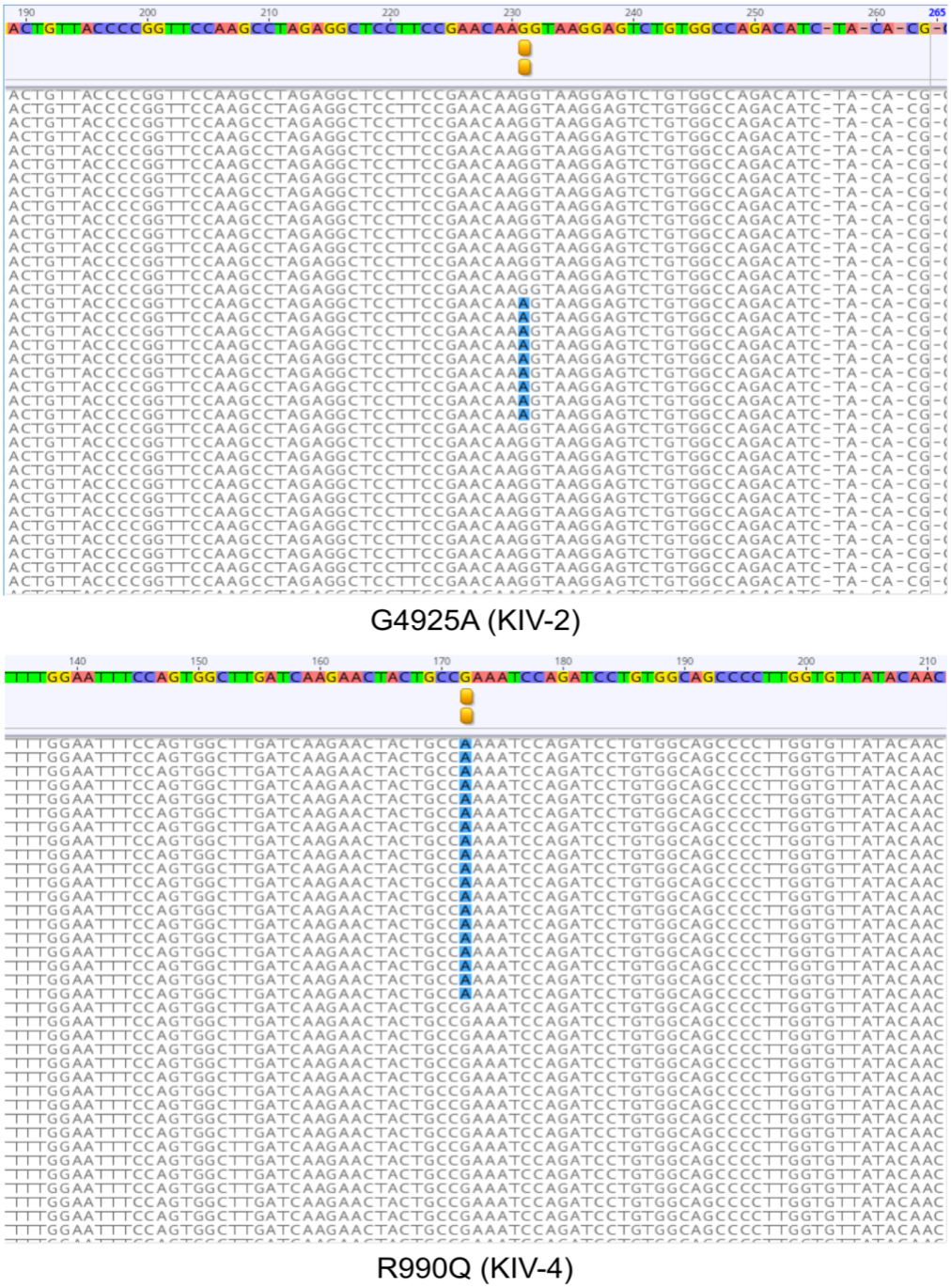
Supplemental Fig. S6: NGS reads of nonsynonymous SNPs in *LPA* identified in subject G5591.



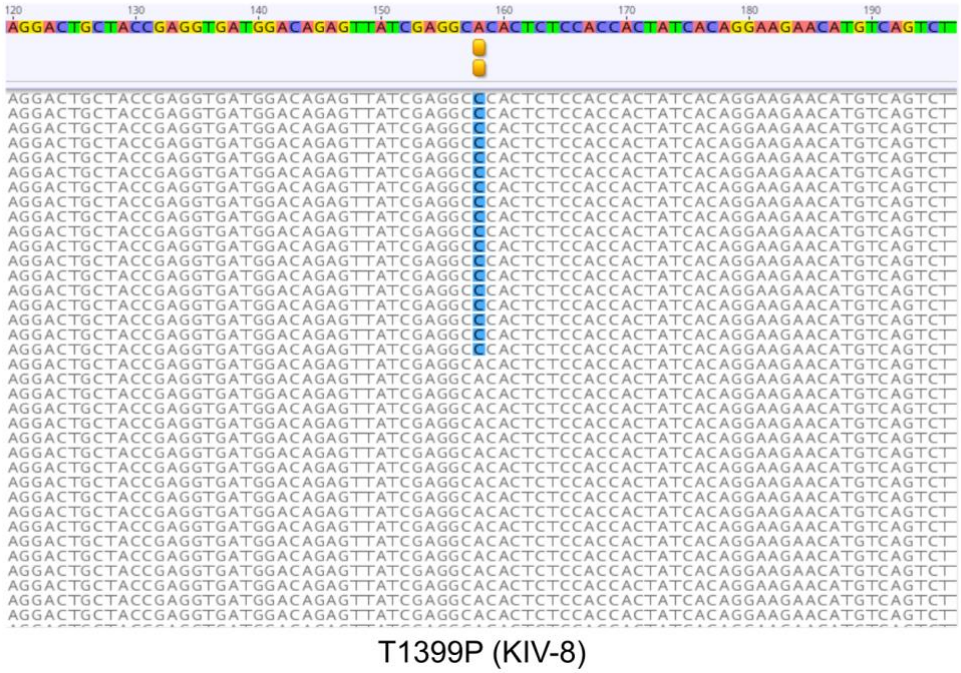
Supplemental Fig. S7: NGS reads of nonsynonymous SNPs in *LPA* identified in subject LPA089.



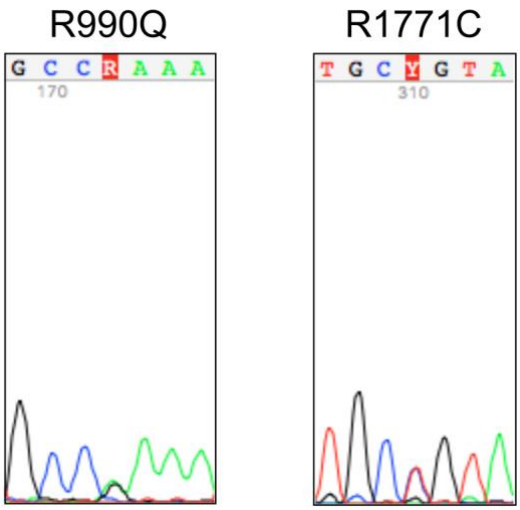
Supplemental Fig. S7 (cont): NGS reads of nonsynonymous SNPs in *LPA* identified in subject LPA089.



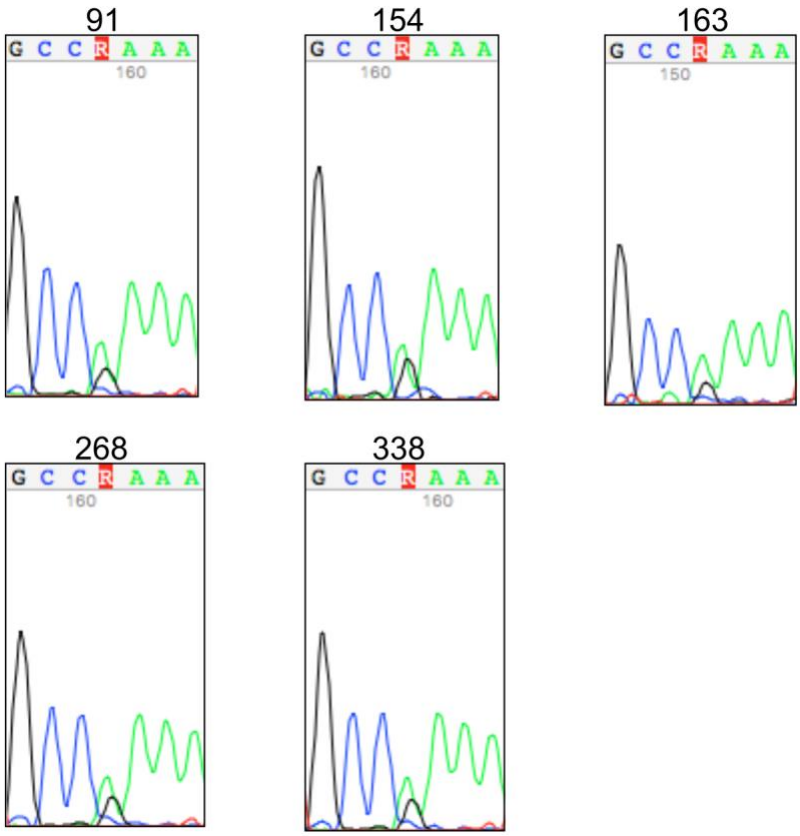
Supplemental Fig. S8: NGS reads of nonsynonymous SNPs in *LPA* identified in G5751.



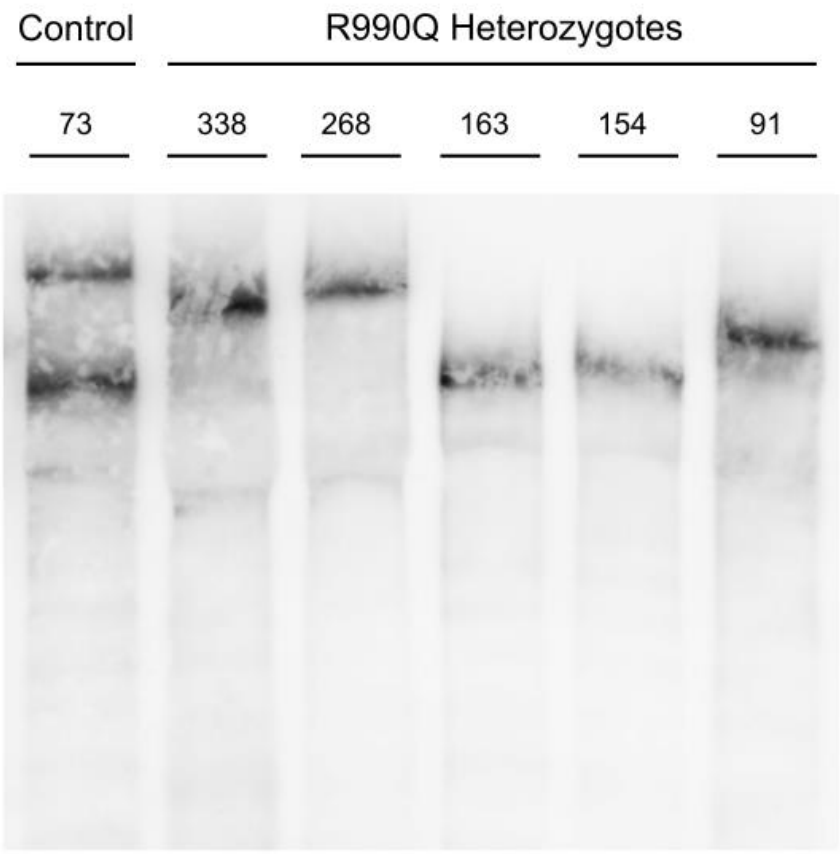
Supplemental Fig. S8 (cont): NGS reads of nonsynonymous *LPA* variants in subject G5751.



Supplemental Fig. S9: Sanger sequencing traces of Subject LPA089 showing nonsynonymous R990Q and R1771C variants.

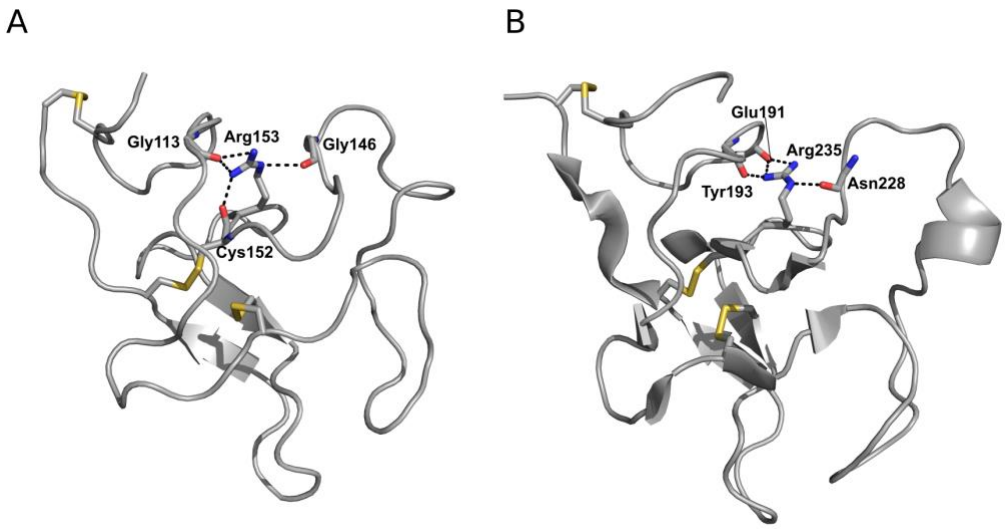


Supplemental Fig. S10: Sanger sequencing traces of five LPA study individuals identified as carrying nonsynonymous R990Q SNP.



Supplemental Fig. S11: Apo(a) phenotyping of individuals heterozygous for R990Q.

Plasma from individuals confirmed heterozygous for R990Q by Sanger sequencing was mixed with reducing buffer and subject to electrophoresis on 4% polyacrylamide gels. Proteins were transferred overnight to nitrocellulose membrane and incubated with a polyclonal anti-Lp(a) antibody. Following incubation with a HRP-labelled secondary antibody, membranes were imaged using ECL. Lanes 1 contains a control sample that exhibits a apo(a) isoform sizes of 20 and 26 and was loaded to 575 ng. Lanes 2-6 contain samples from individuals heterozygous for R990Q loaded to an equivalent level of Lp(a) (500ng).



Supplemental Fig. S12: Protein structures of KI and KII of plasminogen showing homologous arginine residues R153 and R235 respectively. (A) The X-ray crystal structure of KI (PDB 4cik) shows the gaunidinium sidechain of arginine residue 153 is positioned to interact once with the main chain carbonyl group of Gly146 in a 2.9Å contact, twice with the main chain carbonyl of Gly113 in 2.9Å and 3.0Å contacts, and once with the main chain carbonyl of Cys152 in a 2.6Å contact. **(B)** The NMR structure of KII (PDB 1b2i) shows the gaunidinium sidechain of arginine residue 235 is positioned to interact once with the main chain carbonyl of Asn228 in a 2.7Å contact, once with the main chain carbonyl of Tyr193 in a 2.6Å contact and twice with the main chain carbonyl of Glu191 in two 2.6Å polar contacts.

Supplemental references

1. Kearse, M., R. Moir, A. Wilson, S. Stones-Havas, M. Cheung, S. Sturrock, S. Buxton, A. Cooper, S. Markowitz, C. Duran, T. Thierer, B. Ashton, P. Meintjes, and A. Drummond. 2012. Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics* **28**: 1647-1649.

2. Magoc, T., and S. L. Salzberg. 2011. FLASH: fast length adjustment of short reads to improve genome assemblies. *Bioinformatics* **27**: 2957-2963.

3. Edgar, R. C. 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.* **32**: 1792-1797.

4. Coassin, S., S. Schonherr, H. Weissensteiner, G. Erhart, L. Forer, J. L. Losso, C. Lamina, M. Haun, G. Utermann, B. Paulweber, G. Specht, and F. Kronenberg. 2019. A comprehensive map of single-base polymorphisms in the hypervariable LPA kringle IV type 2 copy number variation region. *J. Lipid Res.* **60**: 186-199.

5. Erqou, S. (2010). Lipoprotein(a) and the risk of vascular disease (Doctoral thesis). https://doi.org/10.17863/CAM.16247

6. Mack, S., S. Coassin, R. Rueedi, N. A. Yousri, I. Seppala, C. Gieger, S. Schonherr, L. Forer, G. Erhart, P. Marques-Vidal, J. S. Ried, G. Waeber, S. Bergmann, D. Dahnhardt, A. Stockl, O. T. Raitakari, M. Kahonen, A. Peters, T. Meitinger, K. Strauch, K. O.-S. Group, L. Kedenko, B. Paulweber, T. Lehtimaki, S. C. Hunt, P. Vollenweider, C. Lamina, and F. Kronenberg. 2017. A genome-wide association meta-analysis on lipoprotein (a) concentrations adjusted for apolipoprotein (a) isoforms. J. Lipid Res. 58: 1834-1844.

7. Zekavat, S. M., S. Ruotsalainen, R. E. Handsaker, M. Alver, J. Bloom, T. Poterba, C. Seed, J. Ernst, M. Chaffin, J. Engreitz, G. M. Peloso, A. Manichaikul, C. Yang, K. A. Ryan, M. Fu, W. C. Johnson, M. Tsai, M. Budoff, V. S. Ramachandran, L. A. Cupples, J. I. Rotter, S. S. Rich, W. Post, B. D. Mitchell, A. Correa, A. Metspalu, J. G. Wilson, V. Salomaa, M. Kellis, M. J. Daly, B. M. Neale, S. McCarroll, I. Surakka, T. Esko, A. Ganna, S. Ripatti, S. Kathiresan, P. Natarajan, and N. T. L. W. Group. 2018. Deep coverage whole genome sequences and plasma lipoprotein(a) in individuals of European and African ancestries. Nat Commun 9: 2606.

8. Burgess, S., B. A. Ference, J. R. Staley, D. F. Freitag, A. M. Mason, S. F. Nielsen, P. Willeit, R. Young, P. Surendran, S. Karthikeyan, T. R. Bolton, J. E. Peters, P. R. Kamstrup, A. Tybjaerg-Hansen, M. Benn, A. Langsted, P. Schnohr, S. Vedel-Krogh, C. J. Kobylecki, I. Ford, C. Packard, S. Trompet, J. W. Jukema, N. Sattar, E. Di Angelantonio, D. Saleheen, J. M. M. Howson, B. G. Nordestgaard, A. S. Butterworth, J. Danesh, C. European Prospective Investigation Into, and C. Nutrition-Cardiovascular Disease. 2018. Association of LPA Variants With Risk of Coronary Disease and the Implications for Lipoprotein(a)-Lowering Therapies: A Mendelian Randomization Analysis. JAMA Cardiol 3: 619-627.

9. Li, J., L. A. Lange, J. Sabourin, Q. Duan, W. Valdar, M. S. Willis, Y. Li, J. G. Wilson, and E. M. Lange. 2015. Genome- and exome-wide association study of serum lipoprotein (a) in the Jackson Heart Study. J. Hum. Genet. 60: 755-761.