Supplementary Materials to

A novel but frequent variant in LPA KIV-2 is associated with a pronounced Lp(a) and cardiovascular risk reduction

Stefan Coassin [1], Gertraud Erhart [1], Hansi Weissensteiner [1], Mariana Eca Guimarães de Araújo [2], Claudia Lamina [1], Sebastian Schönherr [1], Lukas Forer [1], Margot Haun [1], Jamie Lee Losso [1], Anna Köttgen [3], Konrad Schmidt [1], Gerd Utermann [4], Annette Peters [5,6,7], Christian Gieger [6,8], Konstantin Strauch [9,10], Armin Finkenstedt [11], Reto Bale [12], Heinz Zoller [11], Bernhard Paulweber [13], Kai-Uwe Eckardt [14], Alexander Hüttenhofer [15], Lukas A. Huber [2], Florian Kronenberg [1]

[1] Division of Genetic Epidemiology, Department of Medical Genetics, Molecular and Clinical Pharmacology, Medical University of Innsbruck, Schoepfstrasse 41, 6020 Innsbruck, Austria
[2] Division of Cell Biology, Biocenter, Medical University of Innsbruck, Innrain 80-82, 6020 Innsbruck, Austria
[3] Division of Genetic Epidemiology, Division of Genetic Epidemiology - Faculty of Medicine and Medical Center - University of Freiburg, Hugstetter Strasse 49, 79106 Freiburg, Germany
[4] Division of Human Genetics, Department of Medical Genetics, Molecular and Clinical Pharmacology, Medical University of Innsbruck, Peter-Mayr-Strasse 1, 6020 Innsbruck, Austria
[5] German Center for Diabetes Research (DZD e.V.), Ingolstaedter Landstrasse 1, 85764 Neuherberg, Germany
[6] Institute of Epidemiology II, Helmholtz Zentrum München - German Research Center for Environmental Health, Ingolstaedter Landstrasse 1, 85764 Neuherberg, Germany
[7] Munich Heart Alliance, German Center for Cardiovascular Disease Research, Ingolstaedter Landstrasse 1, 85764 Neuherberg, Germany
[8] Research Unit of Molecular Epidemiology, Helmholtz Zentrum München - German Research Center for Environmental Health, Ingolstaedter Landstrasse 1, 85764 Neuherberg, Germany
[9] Institute of Genetic Epidemiology, Helmholtz Zentrum München - German Research Center for Environmental Health, Ingolstaedter Landstrasse 1, 85764 Neuherberg, Germany
[10] Institute of Medical Informatics, Biometry and Epidemiology, Chair of Genetic Epidemiology, Ludwig-Maximilians-Universität, Marchioninistrasse 15, 81377 Munich, Germany
[11] Internal Medicine I, Medical University Innsbruck, Anichstrasse 35, 6020 Innsbruck, Austria
[12] Section of Interventional Oncology - Microinvasive Therapy (SIP), Department of Radiology, Medical University Innsbruck, Anichstrasse 35, 6020 Innsbruck, Austria
[13] First Department of Internal Medicine, Paracelsus Private Medical University, Müllner Hauptstrasse 48, 5020 Salzburg, Austria
[14] Department of Nephrology and Hypertension, University of Erlangen-Nürnberg, Ulmenweg 18, 91054 Erlangen, Germany
[15] Division of Genomics and RNomics, Biocenter, Medical University of Innsbruck, Innrain 80-82, 6020 Innsbruck, Austria

Address of correspondence:
Florian Kronenberg, MD, Division of Genetic Epidemiology, Department of Medical Genetics, Molecular and Clinical Pharmacology, Medical University of Innsbruck, Schöpfstr. 41, A-6020 Innsbruck, AUSTRIA; Phone: (+43) 512 9003-70560; E-mail: Florian.Kronenberg@i-med.ac.at

Supplementary Methods

Study populations

Informed consent was obtained from all participants and study protocols were approved by the respective Institutional Review Boards.

SAPHIR study[1]: The Salzburg Atherosclerosis Prevention Program in subjects at High Individual Risk (SAPHIR) is an observational study conducted in the years 1999-2002 involving healthy unrelated subjects: 645 females from 39 to 67 years of age and 1093 males from 39 to 66 years of age. Study participants were recruited by health screening programs in large companies in and around the city of Salzburg as described recently[1]. All individuals were of West-Eurasian origin. Subjects with established coronary artery, cerebrovascular or peripheral arterial disease, congestive heart failure, valvular heart disease, chronic alcohol (more than three drinks a day) or drug abuse, severe obesity (BMI>40kg/m²) and pregnant women were excluded. Informed consent was obtained from each participant. At baseline all study participants were subjected to a comprehensive screening examination. A detailed personal and family history was assessed via standardized questionnaires. A physical examination included measurement of anthropometric parameters such as weight, height, waist circumference and percentage body fat. Blood samples were collected after an overnight fasting period. Full Lp(a) phenotyping was available for n=1,522 participants. Details are given in[1]**.**

The German Chronic Kidney Disease study (GCKD): The design and methodology of the GCKD study has previously been reported in detail[2,3]. Briefly, the GCKD study is an ongoing prospective observational national cohort study including 5,217 Caucasian patients with CKD of moderate severity from a broad etiologic spectrum. The study was approved by the institutional ethics committees and subjects gave written informed consent. Patients were included based on a moderately reduced kidney function (estimated GFR (eGFR) 30-60ml/min per 1.73m^2, corresponding to Kidney Disease Improving Global Outcomes (KDIGO) CKD stage 3) or overt proteinuria in the presence of an eGFR >60ml/min per 1.73m^2. Serum and urine creatinine was measured using the CREA plus assay and urine albumin using the ALBU-XS assay (both Roche/Hitachi Diagnostics GmbH, Mannheim, Germany). The eGFR was calculated using the CKD-EPI equation[4]. Full Lp(a) phenotyping and genotypes was available for n=4,949 participants.

Prevalence of cardiovascular disease (CVD) (n=987 from 4949 ≙ 20%) was defined as having had at least one of the following events [5]: history of non-fatal myocardial infarction (n=559), coronary artery bypass grafting (n=329), percutaneous transluminal coronary angioplasty (n=705), stroke

(n=400), and interventions at the carotid arteries (carotid endarterectomy and/or carotid balloon angioplasty or stent implantation; n=114).

KORA F4: The KORA cohorts (Cooperative Health Research in the Region of Augsburg, KOoperative Gesundheitsforschung in der Region Augsburg) are several cohorts representative of the general population in Augsburg, Germany and two surrounding counties that were initiated as part of the WHO MONICA Study. Ten years age-sex strata have been sampled from the 25 to 74 year old population with a stratum size of 640 subjects. In the KORA S4 in total 4,261 subjects have been examined (response rate 67%). 3,080 individuals participated in a KORA F4 follow-up examination performed in 2004/2005 that is subject of this investigation. Details are given in[6,7]. Full Lp(a) phenotyping and G4925A genotype was available for n=2,892 participants.

Study design for extreme-phenotype discovery phase

Since the association between LMW apo(a) isoforms and high Lp(a) concentrations is well established, individuals carrying a LMW apo(a) isoform should have rather high Lp(a) values. Therefore, we expected that there might be additional not-yet known genetic variants in those LMW carriers with unexpectedly low Lp(a) values, which shift their Lp(a) levels down. Therefore, in LMW carriers, we have chosen individuals from those within the lowest 20% of Lp(a) levels (extreme phenotype group). Controls were selected from the remaining individuals with LMW apo(a) isoforms.

Likewise, individuals who carry HMW apo(a) isoforms are expected to have rather low Lp(a) values. According to the same rationale as above, we have chosen individuals from the HMW carriers within the individuals with highest 20% of Lp(a) values (extreme phenotype group). The controls were selected from the remaining individuals with HMW apo(a) isoforms.

The other extreme phenotype groups (no isoforms detectable, multiple bands etc., see Table 1) were added to this analysis since we expected an enrichment of mutations in this group.

These extreme phenotype and corresponding control groups were selected from the SAPHIR and GCKD study and serve as a discovery set. The numbers in each group and characteristics can be found in Table 1.

Lp(a) phenotyping

In brief, Lp(a) concentration was measured by a double-antibody ELISA using an affinity-purified polyclonal apo(a) antibody for coating and a monoclonal antibody 1A2 directed against KIV-2 for detection[8]. Apo(a) isoform phenotyping from plasma was performed by SDS-agarose electrophoresis using 150 ng of Lp(a) protein, followed by Western blot with 1A2 antibody. The isoform nomenclature always refers to the *total* number of all *KIV* (not KIV-2) domains for each allele according to [9] (Supplementary Fig. 2). The number of KIV-2 repeats can be deduced by subtracting 9 from the isoform designation. Total KIV-2 repeat number quantification by qPCR is described below.

KIV-2 copy determination by qPCR

The sum of KIV-2 repeats present on DNA level was estimated using the exon 4 qPCR protocol of Lanktree et al, 2009[10] with minor modifications (Supplementary Table 16). All measurements were done in quadruplicates in 384 well plates on an ABI 7900HT RT-PCR System (ThermoFisher Scientific, Waltham, MA, USA) using FAM-labeled *LPA* oligonucleotides and RNaseP (TaqMan® Copy Number Reference Assay, ThermoFisher Scientific) as 2n reference. A CV among replicates <2.5 % was required to pass quality control. Average CVs of the Ct values in the discovery set was CV_{LPA}=0.6% (±0.3 SD) and CV_{2nRef}=0.33% (±0.18 SD).

Data was analyzed using the 2^{ddCT} method[11] and 2^{ddCT} values were converted to KIV-2 copies using a heterozygous calibrator DNA from a healthy, G4925A-negative donor expressing both isoforms in plasma (alleles: 25/30). Importantly, the primers of Lanktree et al. bind only in the KIV-2 repeat and in KIV-1. Therefore the number of targets is lowered by 8 target sites per allele (-8 KIV per allele, i.e. KIV-3 to -10) compared to the Western Blot nomenclature, which reports the number of all KIV[9]. Therefore the 2^{ddCT} values must not be multiplied with the WB result of the calibrator sample but 16 must be subtracted (8 per allele, i.e. KIV-3 to KIV-10). Therefore a number of 39 copies was used for conversion of 2^{ddCt} values to number of copies.

Batch amplification

The amplicon used for NGS sequencing encompassed 92% of each KIV-2 repeat (about 5,100 bp) and was amplified using an adapted protocol from Noureen et al. with the primers 421U_2 and 422L11 (an overview is provided in Supplementary Fig. 3).

The PCR protocol and primer sequences are given in Supplementary Tables 3 and 4. PCR products were purified (QIAquick PCR Purification Kit, Qiagen, Hilde, Germany) and quantified (Qubit 3.0 fluorometer and Qubit HS Kit, ThermoFisher Scientific). 200 ng were fragmented to an average fragment length of about 700 bp (663±75 (mean±SD)) using NEBNext® dsDNA Fragmentase® (NEB,

Ipswich, MA, USA) (median digestion time 9 min) and subjected to Illumina TruSeq Nano workflow (Illumina, Inc., San Diego, CA, USA). Next Generation Sequencing (NGS) was performed on an Illumina MiSeq system using reagent kit v2 500 cycles.

NGS data analysis

Data analysis was performed by aligning all reads to one single kringle repeat. For variants occurring only in a subset of KIV-2 repeats, this generates a situation similar to "artificial mitochondrial heteroplasmies"[12] as any variant present only in a subset of KIV-2 repeats will be present only in a subfraction of NGS reads. Alignment and low level variant calling was done using the mtDNA-Server[13] (https://mtdna-server.uibk.ac.at/) with slight adaptions. The mtDNA-server pipeline provides highly accurate and specific detection of low level variants by applying several metrics and test statistics to each position. Filters prior to variant calling were a mapping quality score of >20, an alignment score of >30 and a base quality score >20. A minimal coverage of 10X per strand and a minimal amount of 2 reads for both alleles were required. The minor variant allele was required to exceed 1% on one of the two strands. A maximum likelihood model[14] taking sequencing errors into account as well as a strand bias model was applied to reduce false-positives as previously described[13].

The sixth KIV-2 repeat in the human genome reference hg19 as defined by the binding sites of the primers 421U_2 and 422L (Supplementary Table 4) was used as reference sequence (chr6:161,033,785-161,038,888), since this is the longest KIV-2 repeat in hg19. This reduces the need of insertion calling, which is still difficult using NGS data[15]. Accordingly the non-standard annotation "G4925A" relates to the variant position within this PCR product. To allow context sequence determination, we provide arbitrary HGVS annotation of this variant to the first KIV-2 repeat of the human reference genome as NM_005577.2:c.853G>A.

The LPA specific KIV-2 low level variant calling pipeline will be available at https://lpa-server.i-med.ac.at (manuscript in preparation).

1000G and Genome Austria data analysis

For the exome sequencing data only samples with >1000X total coverage on the variant positions were analyzed (n=1,589). The mean coverage on the variant position in the high coverage whole genome data was to 555X±154. Thirty high coverage whole genome and 2,535 whole exome data from the 1000 Genome project[16] were downloaded from the NCBI FTP Site (ftp://ftp-trace.ncbi.nih.gov/1000genomes/ftp) and all reads mapping to KIV-2 extracted by Samtools[17] with the coordinates on chr6:161,033,785-161,066,618 (hg19).

Additionally, 8 high coverage (average: 514X±199 at variant position) whole genome sequences from GenomeAustria were downloaded (http://genomaustria.at/unser-genom/ and www.personalgenomes.org/austria) and processed accordingly. Results are provided in Supplementary Tables 8, 9, 10, 11).

castPCR

In brief, a castPCR assay is an allele-specific qPCR assay consisting of two allele-specific reactions (wild type detection assay and mutation detection assay), which are run in parallel on each sample using the same DNA input amount. Both assays contain a common locus-specific Taqman probe and a common reverse primer. An allele-specific primer located on the variant position provides allele specificity. The other allele is blocked by a competing blocking oligo to avoid unspecific amplification. This provides thermodynamic advantage to the probe perfectly matching the target allele and allows detection of somatic mutations down to at least 0.1% level[18]. Assay was validated by independently typing the NGS-sequenced discovery set.

Procedure

Reactions were performed in a 384 well plate system on a QuantStudio™ 6 Flex Real-Time PCR System (ThermoFisher Scientific) using 2 ng genomic DNA as input. To control for PCR failure or false-negative non-template controls, a commercial Internal Process Control reagent (TaqMan® Mutation Detection IPC Reagent Kit, ThermoFisher Scientific) was included in 22 samples and two non-template controls for each master mix. Cycling was performed according to manufacturer instructions.

Data was analyzed using the MutationDetector™ software (ThermoFisher Scientific) according to manufacturer guidelines and using a conservative deltaCt cut-off of 9.0 (corresponding to a sensitivity of 0.2%) to reduce the likelihood of any variants just marginally passing the threshold due to random noise. Assay was validated by genotyping the whole discovery set. Classification performance figures from assay validation are reported in Supplementary Table 12.

Sequence used for design of G4925A castPCR assay

"N" represents bases marked to be avoided in assay design because variants were observed at these positions in ultra-deep NGS. G4925A is highlighted in red.

>422_G4925A_SSSNP

GGCACCGGTAATGACCTTGTTGCAGCACAAAGGAGAGAGTGTGGGGTGCCCCTGCATGTTGTCCCACCTCTTG
TGACGTGTATCGTTTTGGAATTTCCAGTGGCTTGATCATGAACTACTGCAGGAATCCAGATGCTGTGGCAGCTC
CTTATTGTTATACGAGGGATCCCGGTGTCAGGTGGGAGTACTGCAACCTGACGCAATGCTCAGACGCAGAAG
GGACTGCCGTCGCGCCTCCGACTGTTACCCCGGTTCCAAGCCTAGAGGCTCCTTCCGAACAA[G/A]GTAAGGA
GTCTGTGGCCAGACATCTACACGCTTCGATGCTGGNATGAAAAGCCATGGAAATTCCCACTGATGCAGCCGCC
TTCAATGGTAAACGGATGCTCGAGTGTTGCCNGAGTTCTANCATGTNGGNGGAAGCCTCCGTGNACTCTCTG
GGGGAGCCAGCGGAGTGATTTCTGGTGCAACGTGGTTGGGCTTTGTCTTTAGGATGGGCACAAACCCTCCAG
GGGGATCGACTTCAAAATTCACCTTGTTGTAAAACGGGCTACCTCAGTGTCCCAG

castPCR quality control

To exclude mistyping we applied extensive quality control measures as follows:

(1) Prior to population screening, the assay was validated on all samples (n=123) from the discovery set, showing excellent classification performance (Supplementary Table 12).

(2) Reproducibility was assessed by typing duplicates of 22 discovery samples. Only one duplicate was discordant giving one false-positive call. Inspection of raw data revealed a considerable higher dCt value between mutation and wild type target compared to the true positive (8.53 vs. 6.66 (average) ±0.79 SD of true positives). The sample had passed the cut-off value by only 0.43 Ct and was >2 SD apart from the other samples. This can occur due to a) unspecific amplification slightly passing the cutoff or b) template amount lower by $2^{\Delta dCt}$ fold, i.e. $2^{1.87}$= 3.86 fold. The latter would reflect also in the Ct value of the wild type assay since the mutation level reported by NGS indicates that only approximately 1 or two KIV-2 are affected by the mutation and thus every sample does contain most KIV-2 as wild type sequence. No shift in the Ct value of the wild type assay was observed (Ct wt assay = 21.91 vs. average 21.67 ±1.01 SD). Thus the positive signal was likely a false-positive amplification slightly passing the threshold. Accordingly a repeated measurement of the discordant sample gave the expected negative call. Based on this observation all positive signals were controlled for dCt values being >2 SD from the average dCt of all positive samples and raw data of such samples was manually reviewed.

(3) A positive and a negative control sample (G4925A carrier status determined by NGS) was included on each 384 well plate (n=16). All the 16 repeated measurement were concordant. CV of the Ct values of wildtype and mutant target was only 0.97% and 1.18%, showing excellent pipetting accuracy and reproducibility of the high throughput setup and assay chemistry.

(4) In KORAF4 n=216 samples (7.1%) were genotyped in duplicates for quality control. Discordance rate was 0.64% (n=2 samples).

(5) Overall call rate, defined as positive wild type assay amplification, was 98.9%.

Pulsed field gel electrophoresis (PFGE)

PFGE was performed as described previously[19] using a using a Bio-Rad CHEF Mapper Pulsed Field Electrophoresis System. Allele sizes were determined using Lambda PFGE Ladder (NEB). Allele separation and isolation was carried out as described[19].

Separated alleles were then tested by castPCR to determine whether G4925A is truly located on the short allele. Moreover, comparison of PFGE results on genomic DNA and Western blot results on plasma allows determining whether G4925A causes a truncated protein. By deleting KIV-3 to -10, the KV and the final protease domain (Fig. 1) a truncation within KIV-2 would cause a protein smaller by the equivalent of at least 11 KIV, if the truncation occurs in the last KIV-2 repeat (approximation without taking glycosylation into account).

pSPL3 minigene generation

A PCR fragment spanning the short intron between two KIV-2 repeats as well as the two boundary exons (with up and downstream intronic portions; Supplementary Fig. 3) was subcloned into a pCR-XL-TOPO vector using the TOPO® XL PCR Cloning Kit (ThermoFisher Scientific) in NEB Turbo E. coli chemically competent cells. G4925A mutation was introduced by site-directed PCR mutagenesis using the Pfu-Ultra II Fusion HS DNA polymerase (Agilent Technologies, Inc., Santa Clara, CA, USA) and DpnI digestion (8 u; QuikChange II Site-Directed Mutagenesis Kit, Agilent Technologies) in 40 µl PCR reaction (37°C, 1 h). Positive clones were identified by colony PCR and confirmed by Sanger sequencing.

Insert was amplified from pXL-TOPO-XL plasmids using NotI-HF (NEB) and BglII (NEB)-tagged primer and sequential digestion (both NEB). 1 µg vector was digested using NotI-HF and BamHI-HF (NEB), dephosphorylated by USB rSAP (USB, Affymetrix, Santa Clara, CA, USA), ligated by T4 ligase (NEB) and transformed into NEB Turbo E. coli. To properly test for induced non-sense mediated

mRNA decay, reading frame was adjusted to in-frame by insertion of two bases at position 9 and 10 of LPA exon 2 using site directed mutagenesis. Due to the large plasmid and insert size (6031 bp and 2645 bp), a multiple cloning site-depleted pEGFP-C1 eGFP cassette was inserted into pSPL3 for transfection control. Multiple cloning site was deleted as described in [20] (Supplementary Table 6 and 17). pSPL3 vector was opened by HpaI digestion, dephosphorylated by rSAP and ligated with PCR-amplified and T4PNK-phosporylated eGFP cassette. Expression was driven by SV40 promoter for the minigene and by CMV for the eGFP cassette.

PCR protocols and primer sequences are given in Supplementary Tables 5 and 6.

Minigene assays

Supplementary Fig. 8 gives an explanation of the minigene assay. Constructs were transfected in HepG2 human hepatoblastoma cells purchased from the American Type Culture Collection, Manassas, VA (ATCC_HB-8065). The original cells were subcloned and selected for hepatocyte-like morphology by the formation of bile-canaliculi. The derived cell clone has been routinely tested for mycoplasma contamination. Minigene expression was driven by SV40 promoter (CMV for the eGFP cassette). Parallel incubation with and without Puromycin was used to assess occurrence of nonsense-mediated mRNA decay[21] (10 µg/mL final concentration). Cells were cultivated in DMEM high glucose supplemented 2 mM L-glutamine (Sigma-Aldrich, St. Louis, MO, USA), 10% fetal calf serum and 1% final concentration PenStrep (100 u/mL penicillin and 0.1 mg/mL Streptomycin; Sigma-Aldrich) at 37°C in a 5% CO2 atmosphere. 24 hours prior to transfection, cells were seeded into 6-well plates at 4 x 10^5 cells/well. Transfection was carried out in two technical and two biological replicates using 2.5 µg plasmid DNA and Lipofectamine® LTX Plus (Thermo Fisher Scientific) according to manufacturer protocol. The newly introduced eGFP cassette served for transfection control purposes.

Puromycin was added 5 hours before harvesting, were applicable. Total RNA was isolated 48 h after transfection using RNAzol (Sigma-Aldrich) and 1 µg were reverse transcribed using AMV Reverse Transcriptase and oligo-dT primer (Promega). Splicing patterns were assessed by PCR using primer binding in the constitutive exons flanking the insert (Supplementary Fig. 8, Supplementary Table 5 and 6). PCR products were finally isolated from agarose gel using Wizard® SV Gel and PCR Clean-Up System (Promega) and analyzed by Sanger sequencing.

RT-PCRs and castPCR on liver biopsy mRNA

Liver biopsies were obtained from four patients undergoing stereotaxic radiofrequency ablation of hepatocellular carcinoma at the Department of Radiology of the Medical University of Innsbruck (1 G4925A carrier). Computed tomography guidance was used to place the biopsy needle in non-tumorous liver tissue. Informed consent was obtained from all participants and study protocol was approved by the Institutional Review Board of the Medical University of Innsbruck. Tissue was immediately frozen in liquid nitrogen, mRNA was isolated using Rneasy Plus Micro Kit and reverse transcribed using ThermoFisher Maxima H Minus First Strand cDNA Synthesis Kit with dsDNase treatment and random hexamer primers. G4925A carrier status was determined by castPCR in gDNA isolated from whole blood. Two batch RT-PCR products were designed to screen for exon skipping and for detection of adenosines at the splice site by NGS in mature mRNA (Supplementary Tables 18 and 19; Supplementary Fig. 10). NGS was performed as above. RNA secondary structure modeling was performed using mFold[22] (http://unafold.rna.albany.edu/?q=mfold/RNA-Folding-Form) with standard settings.

Amplicons of the G4925A context sequence from pre-mRNA were amplified as given in Supplementary Tables 18-19, quantified by Qubit fluorometer, normalized to 2 ng/µl and diluted 1:50,000 and 1:500,000 to ensure amplification of the castPCR wild type probe within the optimal Ct range as specified by the assay manufacturer. CastPCR results from both dilutions corresponded. Presence of genomic DNA in the pre-mRNA PCR was excluded by amplification of an amplicon spanning over intron 6 of *GAPDH* from the same RNA extracts. Only *GAPDH* amplicons corresponding to spliced product were observed (113 bp instead of 205 bp expected on gDNA; Supplementary Fig. 14), confirming efficient removal of genomic DNA by dsDNAse treatment.

Statistical Methods

Prevalence of G4925A in the discovery set phenotypic groups was compared by Fisher exact test (two-sided) and relative contribution of the smaller band in the Western blot to the total signal in Western blot was compared by Wilcoxon test.

The association of G4925A with Lp(a) was tested in the entire KORA F4 study and in LMW/HMW subgroups by linear regression analysis. Since the distribution of Lp(a) concentrations is highly skewed, Lp(a) concentrations were inverse normal-transformed to enforce a normal distribution. Linear regression on the original scale of Lp(a) was additionally used to get a sense of the effect size. All regression analyses were additionally adjusted for the smaller apo(a) isoform (which commonly determines Lp(a) level and CVD risk[23]), respectively the only isoform present. LMW status was

defined according to previous work[24] (at least one isoform <23 KIV repeats). The phenotypic variance of inverse normally transformed Lp(a) explained by G4925A was obtained from the regression models, either from the variant alone and also in addition to a model, which already contained the (smaller or only) isoform. In both studies, KORA F4 and GCKD, regression analysis on Lp(a) was also performed for the SNP rs75692336, which tags G4925A (see paragraph "Detection of tagging SNP for G4925A in the non-repetitive portion of LPA" in Supplementary Methods). The SNP was coded dominantly and regression analyses were performed as described above for G4925A. Finally, effects of G4925A on CVD prevalence in the GCKD study were tested by a logistic regression. Since the information on the carrier status of G4925A was not available for the complete study, the tagging SNP rs75692336 was used as a proxy to assess whether G4925A confers CVD risk reduction induced to LMW carriers compared to HMW carriers. This logistic regression model was performed using the whole GCKD study group and also stratified by the carrier status for at least one rare allele of rs75692336. In addition, adjusted models were performed, taking into account gender, age, Type 2 Diabetes, LDL cholesterol, triglycerides (which was log-transformed), and the kidney function parameters eGFR and albumin-creatinine-ratio (UACR), which was log-transformed. Since about halve of the patients in the GCKD study group take lipid-lowering medications, LDL cholesterol was corrected accordingly for those individuals taking lipid-lowering medication by dividing LDL cholesterol by 0.7 to approximate pre-medication LDL levels. Furthermore, each common cholesterol measurement method implicitly also measures a part of Lp(a)[25]. Therefore, LDL cholesterol values, that were partly already corrected for medication were additionally corrected for Lp(a) by substracting 0.3*Lp(a).

Detection of tagging SNP for G4925A in the non-repetitive portion of *LPA*

A direct evaluation of the effect of G4925A on CVD is hampered by the fact that variants in the KIV-2 are not contained in imputed whole-genome SNP datasets, commonly available in large, genetic-epidemiological studies on cardiovascular diseases. However, both 1000G-imputed whole-genome SNP data (Affymetrix Axiom)[7,26] and G4925A carrier status was available in KORA F4, allowing searching for proxies of G4925A among the SNP presented in the imputed data set.

Genome-wide genotype data have been imputed with the software IMPUTE using 1000 Genomes phase 1, version 3[27]. All SNPs having a minor allele frequency of >1%, a imputation quality of ≥0.4 (IMPUTE info) and a p-value of the HWE-test ≥1e-06 were included individually in a linear

regression model on inverse normal transformed Lp(a), additionally adjusted for isoforms. 932 SNPs showed a genome-wide significant p-value (<5e-08). Each of these SNPs was analyzed for its association with G4925A using Pearson correlation coefficient. The maximum r^2 was found to be 0.82. rs75692336 was selected as tagging SNP for G4925A (r^2=0.82, D'=0.99, minor allele frequency MAF=13%, imputation quality info=0.985). In subsequent association analysis of this SNP with Lp(a) concentrations, rs75692336 was coded dominantly, since heterozygous and homozygous G4925A carrier status cannot be discerned by castPCR.

The association of rs75692336 with CVD events was evaluated in the GCKD study. In this study, genome-wide genotyping data was performed by Illumina HumanOmni2.5-8Exome v1.0 microarrays, which were imputed using the 1000 Genomes phase 3 reference panel in the software Minimac3[28] . rs75692336 was derived from these imputed genotype data (MAF=0.13; imputation quality r^2=0.998) and coded dominantly.

Supplementary Tables

Supplementary Table S1: Baseline characteristics of contributing studies

	KORA F4 (n=2,892)	GCKD (n=4,949)	SAPHIR (n=1,522)
Age (in years)	56.2 ± 13.3 [44, 56, 67]	60.1 ± 12.0 [53, 63, 70]	51 ± 6.0 [46, 52, 55]
Women, n (%)	1490 (51.5%)	1969 (39.8%)	490 (32.7%)
Lp(a) (mg/dL), all subjects	21.8 ± 24.7 [5.2, 11.7, 30.6]	24.6 ± 30.5 [5.0, 11.5, 33.5]	24.0 ± 27.4 [5.4, 11.6, 36.3]
Lp(a) (mg/dL), LMW isoforms	47.7 ± 31.1 [22.6, 46.7, 66.2]	56.3 ± 40.1 [25.6 , 53.3, 77.6]	48.6 ± 35.3 [13.1, 46.4, 68.9]
Lp(a) (mg/dL), HMW isoforms	13.3 ± 14.1 [4.1, 8.7, 17.1]	14.2 ± 16.4 [3.8, 8.5, 17.7]	13.9 ± 14.1 [4.3, 8.7, 19.1]
LMW isoforms, n (%)	716 (24.8%)	1223 (24.7%)	440 (29.0%)
Short Isoform according to Western blot	26.8 ± 6.0 [23, 26, 31]	27.1 ± 6.0 [23, 27, 31]	26.9 ± 6.2 [22, 26, 31]
LDL cholesterol (mg/dL)	136.2 ± 34.9 [112, 134, 158]	118.4 ± 44.0 [89.2, 113.8, 142.7]	145.0 ± 35.6 [120, 144, 168]
HDL cholesterol (mg/dL)	55.9 ± 14.5 [45, 54, 65]	52.0 ± 18.1 [39.4, 48.4, 61.4]	59.1 ± 15.5 [48, 57, 68]
Total cholesterol (mg/dL)	216.2 ± 39.7 [188, 214, 240]	211.2 ± 52.9 [176.4, 207.1, 239.4]	226.5 ± 38.8 [199, 224, 251]
Triglycerides (mg/dL)	125.1 ± 89.1 [72, 105, 151]	198.1 ± 126.1 [118.0, 168.2, 239.0]	126.0 ± 89.3 [72, 101, 151]
Lipid-lowering medication, n (%)	367 (12.7%)	2520 (50.9%)	63 (4.2%)
Type 2 Diabetes, n (%)	209 (7.1%)	1739 (35.1%)	48 (3.2%)
eGFR (mL/min/1.73m^2)	--	49.4 ± 18.1 [37, 46, 57]	--
UACR (mg/g)	--	431.4 ± 960.4 [9.6, 50.9, 391.6]	--

Continuous variables are shown as mean ± SD and [25%,50%,75%] percentiles
n: Sample size of analysis dataset with available Lp(a)-values (and genotypes in KORA F4 and GCKD)

Supplementary Table S2: Description of discovery samples.

Column G4925A reports the mutation level according to deep sequencing. Minor allele carriers for the known null mutations are boldfaced.

#	Lp(a) [mg/dL]	Isoform 1	Isoform 2	Sum Western blot	rs41272114 (null allele KIV-8)[29,30]	rs143431368 (null allele KIV-10)[29]	qPCR [Sum copies]	G4925A
GCKD study – Controls with HMW isoforms and expected Lp(a) concentrations								
1	6.2	37	-	37	**CT**	AA	54	
2	4.7	25	-	25	CC	AA	35	**2.9%**
3	11.9	26	38	64	CC	AA	47	
4	7.2	37	38	75	CC	AA	57	
5	10.1	34	36	70	CC	AA	48	
6	5.9	36	-	36	CC	AA	46	
7	6.5	37	-	37	CC	AA	60	
8	9.8	34	-	34	CC	AA	42	
9	0.2	34	37	71	CC	AA	48	
10	0.8	39	-	39	CC	AA	47	
11	1.0	36	-	36	CC	AA	64	
12	1.3	28	38	66	CC	AA	53	
13	1.3	27	-	27	CC	AA	50	
GCKD study – Controls with LMW isoforms and expected Lp(a) concentrations								
14	58.8	14	-	14	CC	AA	23	
15	30.1	20	31	51	CC	AA	39	
16	52.9	20	-	20	CC	AA	35	
17	35.0	20	32	52	CC	AA	36	
18	52.3	16	-	16	CC	AA	44	
19	55.0	20	-	20	CC	AA	36	
20	55.7	19	37	56	CC	AA	37	
GCKD study – Extreme phenotype approach: high Lp(a) despite HMW isoform								
21	76.1	29	34	63	CC	AA	42	
22	66.0	31	37	68	CC	AA	54	
23	65.4	32	36	68	CC	AA	54	
24	66.3	25	29	54	CC	AA	44	
25	71.9	27	39	66	CC	AA	56	
26	81.6	28	33	61	CC	AA	39	
27	98.4	29	-	29	CC	AA	43	
28	129.7	26	34	60	CC	AA	37	
29	76.7	30	37	67	CC	AA	51	
30	81.2	34	-	34	CC	AA	51	
31	68.9	25	27	52	CC	AA	40	
32	77.8	29	-	29	CC	AA	40	
33	78.5	25	-	25	CC	AA	40	
34	73.9	29	-	29	CC	AA	42	
35	68.6	28	35	63	CC	AA	45	

… to be continued

Supplementary Table S2: continued

#	Lp(a) [mg/dL]	Isoform 1	Isoform 2	Sum Western blot	rs41272114 (null allele KIV-8)[29,30]	rs143431368 (null allele KIV-10)[29]	qPCR [Sum copies]	G4925A
GCKD study – Extreme phenotype approach: low Lp(a) despite LMW isoform								
36	3.6	20	31	51	CC	AA	41	**6.1%**
37	2.2	20	-	20	**CT**	AA	44	**6.4%**
38	4.0	16	27	43	CC	AA	34	**7.4%**
39	8.2	21	32	53	CC	AA	44	**3.0%**
40	2.0	37	-	37	CC	AA	64	-
41	6.7	21	37	58	CC	AA	40	**2.7%**
42	2.5	20	-	20	CC	AA	40	**5.3%**
43	2.5	22	29	51	CC	AA	35	**2.9%**
44	8.8	14	34	48	CC	AA	35	**7.0%**
45	4.9	20	37	57	CC	AA	45	**5.4%**
46	4.6	20	31	51	CC	AA	36	**6.3%**
47	4.2	17	31	48	CC	AA	34	**6.1%**
48	3.9	22	37	59	CC	AA	36	-
49	2.7	28	-	28	CC	AA	49	-
50	3.2	20	-	20	CC	AA	45	**5.5%**
51	8.8	19	-	19	**CT**	AA	36	**3.2%**
52	7.9	19	37	56	CC	AA	38	**5.2%**
53	6.3	20	25	45	CC	AA	37	**7.4%**
GCKD study – Extreme phenotype approach: no isoform detectable in Western blot								
54	0.8	-	-	-	CC	AA	56	-
55	0.6	-	-	-	**CT**	AA	51	-
56	0.2	-	-	-	CC	AA	58	-
57	0.1	-	-	-	**TT**	AA	46	-
GCKD study – Extreme phenotype approach: other (only 10% of small allele)								
58	35.8	20	28	48	CC	AA	37	**6.4%**
GCKD study – Extreme phenotype approach: other (multiple bands in Western blot)								
59	85.8	18	23	41	CC	AA	32	**3.3%**

… to be continued

Supplementary Table S2: continued

#	Lp(a) [mg/dL]	Isoform 1	Isoform 2	Sum Western blot	rs41272114 (null allele KIV-8)[29,30]	rs143431368 (null allele KIV-10)[29]	qPCR [Sum copies]	G4925A
SAPHIR study – Controls with HMW isoforms and expected Lp(a) concentrations								
1	5.3	35	-	-	CC	AA	49	-
2	1.8	30	36	66	CC	AA	51	-
3	3.4	30	36	66	CC	AA	43	-
4	1.3	35	-	-	**CT**	AA	37	-
5	4.3	35	-	-	CC	**AG**	47	-
6	3.7	26	-	-	CC	AA	40	-
7	2.9	32	-	-	CC	AA	46	-
8	1.4	32	-	-	CC	AA	44	**4.6%**
9	2.1	31	-	-	CC	AA	49	-
10	1.4	36	-	-	CC	AA	42	-
11	3.4	36	-	-	CC	AA	50	**4.3%**
12	3.3	25	-	-	**CT**	AA	38	**3.0%**
13	2.5	30	-	-	CC	AA	45	-
14	2.3	30	-	-	CC	**AG**	44	-
15	2.0	30	-	-	CC	AA	38	-
SAPHIR study – Controls with LMW isoforms and expected Lp(a) concentrations								
16	140.3	15	-	-	CC	AA	31	-
17	51.2	15	-	-	CC	AA	32	-
18	88.9	15	-	-	CC	AA	23	-
19	96.9	14	-	-	CC	AA	31	-
20	77.1	15	-	-	CC	AA	30	-
21	40.3	15	-	-	CC	AA	30	-
22	69.2	15	-	-	CC	AA	26	**4.6%**
23	162.5	15	-	-	CC	AA	35	-
SAPHIR study – Extreme phenotype approach: high Lp(a) despite HMW isoform								
24	38.5	40	-	-	CC	AA	53	-
25	55.4	30	-	-	**CT**	AA	45	-
26	21.9	36	-	-	CC	AA	52	-
27	59.6	33	-	-	CC	AA	42	-
28	60.0	29	33	62	CC	AA	42	-
29	72.7	30	39	69	CC	AA	47	-
30	55.4	31	33	64	CC	AA	39	-
31	54.0	34	-	-	CC	AA	47	-
32	55.9	30	35	65	CC	AA	44	-
33	62.0	29	-	-	CC	AA	38	**4.0%**
34	50.8	34	39	73	CC	AA	47	-
35	64.5	30	-	-	CC	AA	42	-
36	25.8	36	-	-	CC	AA	51	-
37	66.5	29	-	-	CC	AA	43	-
38	51.6	33	-	-	CC	AA	37	-
39	64.2	30	36	66	CC	AA	38	-
40	69.4	29	34	63	CC	AA	41	-
41	49.6	33	36	69	CC	AA	43	-

… to be continued

Supplementary Table S2: continued

#	Lp(a) [mg/dL]	Isoform 1	Isoform 2	Sum Western blot	rs41272114 (null allele KIV-8)[29,30]	rs143431368 (null allele KIV-10)[29]	qPCR [Sum copies]	G4925A
SAPHIR study – Extreme phenotype approach: low Lp(a) despite LMW isoform								
42	1.9	12	20	32	CC	AA	41	-
43	6.4	21	-	-	CC	AA	37	**2.9%**
44	7.5	17	26	43	CC	AA	34	-
45	5.4	16	35	51	CC	AA	36	-
46	4.6	20	31	51	CC	AA	39	-
47	3.2	20	26	46	CC	AA	33	**6.5%**
48	3.7	22	-	-	CC	AA	41	**2.5%**
49	7.1	22	-	-	CC	AA	39	**3.0%**
50	3.3	22	-	-	CC	AA	48	-
51	2.9	21	-	-	CC	AA	45	**4.7%**
52	1.2	17	26	43	**CT**	AA	33	-
53	9.3	20	-	-	CC	AA	49	**2.3%**
54	7.7	22	-	-	CC	AA	39	**3.1%**
SAPHIR study – Extreme phenotype approach: no isoform detectable in Western blot								
55	0.6	-	-	-	**CT**	AA	59	-
56	0.9	-	-	-	CC	AA	60	-
57	0.8	-	-	-	CC	AA	39	**1.1%**
58	1.0	-	-	-	CC	AA	65	-
59	2.6	-	-	-	CC	AA	41	-
60	0.1	-	-	-	CC	AA	59	-
61	1.2	-	-	-	CC	AA	58	-
62	0.6	-	-	-	CC	AA	34	**3.3%**
63	0.1	-	-	-	**TT**	AA	50	-
64	0.4	-	-	-	CC	AA	56	-

Supplementary Table S3: PCR protocols for mutation screening

	Batch amplification PCR for NGS SAPHIR	Batch amplification PCR for NGS GCKD	Sequencing of rs143431368 / NM_005577.2: c.4974-2A>G	Sequencing of rs41272114 / NM_005577.2: c.4289+1G>A
Product length (based on hg19)	5,104 *	5,104 *	729	448
Reaction vol. [µl]	25	25	10	10
Enzyme	Qiagen long range PCR kit	LongAmp® Taq DNA Polymerase	Agilent Herc-ulase II Fusion	Qiagen HotStar Taq
Initial denaturation	93 °C, 2 min	94 °C, 3 min	95 °C, 2 min	94 °C, 3 min
Denaturation	93 °C, 15 s	94 °C, 30 s	95 °C, 20 s	94 °C, 30 s
Annealing	64.5 °C, 30 s	64 °C, 30 s	70 °C, 20 s	60 °C, 30 s
Extension	68 °C, 5 min	65 °C, 5 min	72 °C, 40 s	72 °C, 30 s
Final extension	68 °C, 10 min	65 °C, 10 min	72 °C, 3 min	72 °C, 10 min
Number of cycles	35	35	30	30
Primer fw **, ***	421U_2	421U_2	LPA_c.4974_fw2	LPA_rs41272114_f2
Primer rv **, ***	422L	422L	LPA_c.4974_rv1	LPA_rs41272114_r1
Final primer conc. [µM each]	0.6	0.4	0.25	0.5
Final dNTP concentration [mM]	0.6	0.4	0.25	0.2
Enzyme amount [µl]	0.25	1	0.2	0.5
DNA input [ng]	60	50	40	20

* Based on the sixth KIV-2 in hg19 reference sequence. Actual length of products in batch PCR mixture might slightly vary.

** For sequences see Supplementary Table S4

*** For rs143431368 and rs41272114 the PCR primer were used also for Sanger sequencing

Supplementary Table S4: Primer sequences for mutation screening.

Name	Sequence 5'-3'
421U_2[19]	TCAGGATGCAGGGCATGAG
421L[19]	TTTTTCTGACAATCGGAATATAC
422U[19]	AGAAACAAACCTACTAAACCTGACAG
422L[19]	CACCAGAAATCACTCCGCTG
LPA_c.4974_fw2	GAAACAGGCACCTTCTTCACAGGC
LPA_c.4974_rv1	CCTTGGTCATGGCAGAACCTCAAC
LPA_rs41272114_f2	GGGTCCAGGACTGCTACCGA
LPA_rs41272114_r1	CTTTAGTGGGCATTGTGGAATGATAC

Supplementary Table S5: PCR protocols for minigene assays and clones generation

	Batch amplification PCR for TOPO XL cloning	TOPO XL mutagenesis	Subcloning of eGFP cassette into pSPL3	Subcloning of *LPA* KIV-2 in pSPL3	pSPL3 minigene-assay PCR	pSPL3 mutagenesis for frame correction
Product length	2,645 *	6,164	1,723	2,645	606	10,424
Reaction vol. [µl]	25	50	10	10	50	10
Enzyme	Agilent Herculase II Fusion	Agilent PfuUltra II Fusion HotStart DNA Polymerase	Agilent Herculase II Fusion	Agilent Herculase II Fusion	Agilent Herculase II Fusion	Agilent PfuUltra II Fusion HotStart DNA Polymerase
Initial denaturation	95 °C, 2 min	92 °C, 2 min	95°C, 2 min	95 °C, 2 min	95 °C, 2 min	92 °C, 2 min
Denaturation	95 °C, 20 s	92 °C, 10 s	95 °C, 20 s	95 °C, 20 s	95 °C, 20 s	92 °C, 10 s
Annealing	56 °C, 20 s	67 °C, 20 s	58 °C, 20 s	58 °C, 20 s	63 °C, 20 s	65 °C, 20 s
Extension	72 °C, 1:45 min	72 °C, 2 min	72 °C, 2:40 min	72 °C, 1:20 min	72 °C, 30 s	68 °C, 5.5 min
Final extension	72 °C, 3 min	68 °C, 5 min	72 °C, 3 min	72 °C, 3 min	72 °C, 3 min	68 °C, 5 min
Number of cycles	30	18	30	30	30	18
ID Primer fw **	422U	LPA422-Mutg182aU	EGFP-fw-HpaI	422U-NotIHF	SD6	pSPL3_Lpa_Mut_fw
ID Primer rv **	421L	LPA422-Mutg182aL	EGFP-rv-HpaI	421L_BglII	SA2	pSPL3_Lpa_mut_rv
Final primer conc. [µM each]	0.25	0.20	0.25	0.25	0.25	0.20
Final dNTP conc. [mM]	0.25	0.25	0.25	0.25	0.25	0.25
Enzyme amount [µl]	0.5	1	0.2	0.2	1	0.20
DNA input [ng]	100	34	5	5	1.5	7

* Based on the sixth KIV-2 in hg19 reference sequence. Actual length of products in batch PCR mixture might slightly vary.

** For sequences see Supplementary Table S6

Supplementary Table S6: Primer sequences for clone generation and minigene assays. Bases for mutageneses are shown in red letters.

Name	Sequence 5'-3'
422U	AGAAACAAACCTACTAAACCTGACAG
421L	TTTTTCTGACAATCGGAATATAC
pEGFP-C1-del-rv	TTATGATCATAACTTGTACAGCTCGTCCATGCCGAGAGTGATCC
pEGFP-C1-del-fw	CTGTACAAGTTATGATCATAATCAGCCATACCACATTTGTAGAG
EGFP-fw-HpaI	TTGGATGTTAACTGATTCTGTGGATAACC
EGFP-rv-HpaI	TAATATGTTAACTATCTCGGTCTATTCTTTTGAT
422U-NotIHF	TAGGATGCGGCCGCAGAAACAAACCTACTAAACCTGACAG
421L_BglII	TGGTTTAGATCTTTTTTCTGACAATCGGAATATAC
SD6[31]	TCTGAGTCACCTGGACAACC
SA2[31]	ATCTCAGTGGTATTTGTGAGC
422U_2(GE)-BglII	TAATAAAGATCTAGAGATGTGCAGAGAGGATTAGTG
422L-MluI	TAATAAACGCGTCACCAGAAATCACTCCGCTG
LPA422-Mutg182aU	GGCTCCTTCCGAACAAAGTAAGGAGTCTGTGGC
LPA422-Mutg182aL	GCCACAGACTCCTTACTTTGTTCGGAAGGAGCC
pSPL3_Lpa_Mut_fw	GGAATTTCCAGTGGCTTGATCTCATGAACTACTGCAG
pSPL3_Lpa_mut_rv	CTGCAGTAGTTCATGAGATCAAGCCACTGGAAATTCC

Supplementary Table S7: Frequency of known splice site defects in KIV-7 (rs41272114) and KIV-10 (rs143431368) in the discovery set.

	Group	n	rs41272114 CT/TT*	rs143431368 AG/GG*
SAPHIR study				
HMW Isoform, expected Lp(a)	Controls	15	2/0	2/0
LMW Isoform, expected Lp(a)	Controls	8	0/0	0/0
High Lp(a) despite HMW isoform	Extreme phenotype	18	1/0	0/0
Low Lp(a) despite LMW isoform	Extreme phenotype	13	1/0	0/0
No detectable isoform	Extreme phenotype	12	1/1	0/0
GCKD study				
HMW Isoform, expected Lp(a)	Controls	13	1/0	0/0
LMW Isoform, expected Lp(a)	Controls	7	0/0	0/0
High Lp(a) despite HMW isoform	Extreme phenotype	15	0/0	0/0
Low Lp(a) despite LMW isoform	Extreme phenotype	18	2/0	0/0
No detectable isoform & other	Extreme phenotype	6	1/1	0/0

* Heterozygotes and homozygotes for the rare allele

Summary: Together 3 heterozygote carriers in the control group (from 43) and 6 heterozygotes and 2 homozygote carriers of the rare allele in the extreme phenotype group (from 78); p-value Fishers exact test = 0.87

Supplementary Table S8: Number of G4925A carriers in the 1000 Genome exome data. Only exome data with >1000X coverage on variant position was used (n=1,589).

Superpopulation/ Population	Description	Non-carriers	carriers	Total	Carrier frequency
AFR	**African**	**514**	**14**	**528**	**0.027**
ACB	African Caribbeans in Barbados	58	3	61	0.049
ASW	Americans of African Ancestry in SW USA	54	3	57	0.053
ESN	Esan in Nigeria	98	1	99	0.010
GWD	Gambian in Western Divisions in the Gambia	108	5	113	0.044
LWK	Luhya in Webuye, Kenya	40	0	40	0.000
MSL	Mende in Sierra Leone	84	1	85	0.012
YRI	Yoruba in Ibadan, Nigeria	72	1	73	0.014
AMR	**Ad Mixed American**	**219**	**37**	**256**	**0.145**
CLM	Colombians from Medellin, Colombia	52	13	65	0.200
MXL	Mexican Ancestry from Los Angeles USA	46	13	59	0.220
PEL	Peruvians from Lima, Peru	43	4	47	0.085
PUR	Puerto Ricans from Puerto Rico	78	7	85	0.082

… to be continued

Supplementary Table S8: continued

Superpopulation/ Population	Description	Non-carriers	carriers	Total	Carrier frequency
EAS	**East Asian**	**148**	**2**	**150**	**0.013**
CDX	Chinese Dai in Xishuangbanna, China	5	0	5	0.000
CHB	Han Chinese in Bejing, China	19	0	19	0.000
CHS	Southern Han Chinese	23	0	23	0.000
JPT	Japanese in Tokyo, Japan	65	2	67	0.030
KHV	Kinh in Ho Chi Minh City, Vietnam	36	0	36	0.000
EUR	**European**	**146**	**46**	**192**	**0.240**
CEU	Utah Residents (CEPH) with Northern and Western Ancestry	32	12	44	0.273
FIN	Finnish in Finland	25	5	30	0.167
GBR	British in England and Scotland	28	8	36	0.222
IBS	Iberian Population in Spain	23	8	31	0.258
TSI	Toscani in Italia	38	13	51	0.255
SAS	**South Asian**	**400**	**63**	**463**	**0.136**
BEB	Bengali from Bangladesh	78	6	84	0.071
GIH	Gujarati Indian from Houston, Texas	68	9	77	0.117
ITU	Indian Telugu from the UK	86	17	103	0.165
PJL	Punjabi from Lahore, Pakistan	84	12	96	0.125
STU	Sri Lankan Tamil from the UK	84	19	103	0.184

Supplementary Table S9: 1000 Genomes Exome set.

Individual data for G4925A carriers from all 1000G exome samples (>1000X) at variant position.

SampleID	POS in reference	Change	Mutation level	Coverage Forward	Coverage Reverse	Total Coverage	Pop
AFR							
HG02014	4925	G>A	0.0343	3347	10123	13470	ACB
HG02501	4925	G>A	0.0433	890	1280	2170	ACB
HG02554	4925	G>A	0.0352	634	1726	2360	ACB
NA19703	4925	G>A	0.0289	498	1646	2144	ASW
NA20276	4925	G>A	0.0625	922	2772	3694	ASW
NA20341	4925	G>A	0.0345	1262	3867	5129	ASW
HG02971	4925	G>A	0.0299	1062	2713	3775	ESN
HG02595	4925	G>A	0.0649	2193	4818	7011	GWD
HG02807	4925	G>A	0.0275	2456	5186	7642	GWD
HG02840	4925	G>A	0.0270	1037	2626	3663	GWD
HG02852	4925	G>A	0.0258	1261	3314	4575	GWD
HG03028	4925	G>A	0.0189	2748	6118	8866	GWD
HG03559	4925	G>A	0.0287	712	858	1570	MSL
NA18881	4925	G>A	0.0282	1009	2602	3611	YRI
AMR							
HG01112	4925	G>A	0.024	846	3117	3963	CLM
HG01269	4925	G>A	0.0265	1424	3981	5405	CLM
HG01280	4925	G>A	0.0224	883	2464	3347	CLM
HG01341	4925	G>A	0.0302	670	2008	2678	CLM
HG01360	4925	G>A	0.0435	834	2314	3148	CLM
HG01431	4925	G>A	0.0310	1068	2678	3746	CLM
HG01432	4925	G>A	0.0313	932	2163	3095	CLM
HG01440	4925	G>A	0.0193	617	1866	2483	CLM
HG01455	4925	G>A	0.0315	1314	3834	5148	CLM
HG01464	4925	G>A	0.0224	680	1959	2639	CLM
HG01488	4925	G>A	0.0262	624	1823	2447	CLM
HG01489	4925	G>A	0.0292	692	2045	2737	CLM
HG01498	4925	G>A	0.0513	1282	3570	4852	CLM
NA19658	4925	G>A	0.0279	579	1787	2366	MXL
NA19661	4925	G>A	0.0263	582	1777	2359	MXL
NA19664	4925	G>A	0.0208	707	2076	2783	MXL
NA19717	4925	G>A	0.0318	612	1837	2449	MXL
NA19731	4925	G>A	0.0320	1401	3938	5339	MXL
NA19732	4925	G>A	0.0316	1527	4104	5631	MXL
NA19759	4925	G>A	0.0336	1417	4175	5592	MXL
NA19762	4925	G>A	0.0342	545	2434	2979	MXL
NA19780	4925	G>A	0.0279	735	2773	3508	MXL
NA19782	4925	G>A	0.0301	1024	3492	4516	MXL
NA19783	4925	G>A	0.0230	1275	3978	5253	MXL
NA19785	4925	G>A	0.0272	1213	3833	5046	MXL
NA19792	4925	G>A	0.0334	1161	2494	3655	MXL

… to be continued

Supplementary Table S9: continued

SampleID	POS in reference	Change	Mutation level	Coverage Forward	Coverage Reverse	Total Coverage	Pop
AMR continued							
HG01927	4925	G>A	0.0197	3010	8914	11924	PEL
HG01967	4925	G>A	0.0313	4030	12057	16087	PEL
HG01976	4925	G>A	0.0239	5202	14556	19758	PEL
HG02298	4925	G>A	0.0305	1236	3907	5143	PEL
HG00734	4925	G>A	0.0262	2595	11170	13765	PUR
HG01092	4925	G>A	0.0405	722	1007	1729	PUR
HG01098	4925	G>A	0.0281	946	3853	4799	PUR
HG01102	4925	G>A	0.0223	931	4304	5235	PUR
HG01190	4925	G>A	0.0236	1457	5082	6539	PUR
HG01204	4925	G>A	0.0249	1924	8194	10118	PUR
HG01325	4925	G>A	0.0231	1160	1476	2636	PUR
EAS							
NA18970	4925	G>A	0.0257	1388	4016	5404	JPT
NA19076	4925	G>A	0.0256	929	3254	4183	JPT
EUR							
NA06984	4925	G>A	0.0223	783	2171	2954	CEU
NA06994	4925	G>A	0.0643	907	2856	3763	CEU
NA07048	4925	G>A	0.0244	1834	6993	8827	CEU
NA07347	4925	G>A	0.0283	897	4129	5026	CEU
NA11920	4925	G>A	0.0255	1412	8189	9601	CEU
NA12272	4925	G>A	0.0765	1241	3491	4732	CEU
NA12282	4925	G>A	0.0657	1068	3465	4533	CEU
NA12286	4925	G>A	0.0319	1315	3636	4951	CEU
NA12718	4925	G>A	0.0246	1386	4093	5479	CEU
NA12761	4925	G>A	0.0331	777	2272	3049	CEU
NA12842	4925	G>A	0.0655	1645	11325	12970	CEU
NA12890	4925	G>A	0.0329	2278	8170	10448	CEU
HG00190	4925	G>A	0.0233	1443	4812	6255	FIN
HG00272	4925	G>A	0.0254	655	2657	3312	FIN
HG00306	4925	G>A	0.0688	482	2063	2545	FIN
HG00367	4925	G>A	0.0361	553	2386	2939	FIN
HG00380	4925	G>A	0.0239	629	2465	3094	FIN
HG00117	4925	G>A	0.0269	1296	4984	6280	GBR
HG00123	4925	G>A	0.0217	1184	4288	5472	GBR
HG00131	4925	G>A	0.0263	1298	4759	6057	GBR
HG00145	4925	G>A	0.0296	1139	4840	5979	GBR
HG00155	4925	G>A	0.0261	1733	7070	8803	GBR
HG00159	4925	G>A	0.0379	730	3074	3804	GBR
HG00244	4925	G>A	0.0258	1354	5131	6485	GBR
HG00265	4925	G>A	0.0527	834	3173	4007	GBR

… to be continued

Supplementary Table S9: continued

SampleID	POS in reference	Change	Mutation level	Coverage Forward	Coverage Reverse	Total Coverage	Pop
EUR continued							
HG01515	4925	G>A	0.0232	1175	4869	6044	IBS
HG01518	4925	G>A	0.0231	1271	5228	6499	IBS
HG01519	4925	G>A	0.0238	1207	5102	6309	IBS
HG01521	4925	G>A	0.0354	2072	8594	10666	IBS
HG01697	4925	G>A	0.0224	40873	116560	157433	IBS
HG01707	4925	G>A	0.0209	3359	9439	12798	IBS
HG01708	4925	G>A	0.0244	3139	8765	11904	IBS
HG01766	4925	G>A	0.0378	408	1072	1480	IBS
NA20763	4925	G>A	0.0398	863	2103	2966	TSI
NA20771	4925	G>A	0.0247	683	2110	2793	TSI
NA20778	4925	G>A	0.0196	703	2108	2811	TSI
NA20786	4925	G>A	0.0148	885	2488	3373	TSI
NA20800	4925	G>A	0.0612	1031	3466	4497	TSI
NA20802	4925	G>A	0.0242	1250	5039	6289	TSI
NA20803	4925	G>A	0.1070	1237	4335	5572	TSI
NA20804	4925	G>A	0.0251	707	3714	4421	TSI
NA20805	4925	G>A	0.0325	809	3963	4772	TSI
NA20807	4925	G>A	0.0431	668	1975	2643	TSI
NA20808	4925	G>A	0.0360	813	2632	3445	TSI
NA20813	4925	G>A	0.0341	1023	3082	4105	TSI
NA20814	4925	G>A	0.0291	1000	3191	4191	TSI
SAS							
HG03615	4925	G>A	0.0332	1169	3112	4281	BEB
HG03805	4925	G>A	0.0282	915	2527	3442	BEB
HG03902	4925	G>A	0.0182	1488	3844	5332	BEB
HG04144	4925	G>A	0.0164	745	1877	2622	BEB
HG04173	4925	G>A	0.0270	1049	1465	2514	BEB
HG04182	4925	G>A	0.0191	871	1166	2037	BEB
NA20852	4925	G>A	0.0235	2895	8358	11253	GIH
NA20872	4925	G>A	0.0208	3304	9057	12361	GIH
NA20878	4925	G>A	0.0218	2171	6369	8540	GIH
NA20893	4925	G>A	0.0239	2247	6290	8537	GIH
NA20894	4925	G>A	0.0654	2063	5921	7984	GIH
NA21093	4925	G>A	0.0267	1918	5786	7704	GIH
NA21094	4925	G>A	0.0675	1207	3801	5008	GIH
NA21098	4925	G>A	0.0231	1110	3347	4457	GIH
NA21124	4925	G>A	0.0334	1266	2711	3977	GIH
HG03729	4925	G>A	0.0286	1545	4229	5774	ITU
HG03731	4925	G>A	0.0287	1392	3595	4987	ITU
HG03775	4925	G>A	0.0248	2098	5902	8000	ITU
HG03784	4925	G>A	0.0246	1789	5035	6824	ITU
HG03787	4925	G>A	0.0357	1122	3105	4227	ITU

… to be continued

Supplementary Table S9: continued

SampleID	POS in reference	Change	Mutation level	Coverage Forward	Coverage Reverse	Total Coverage	Pop
SAS continued							
HG03863	4925	G>A	0.0244	1192	2908	4100	ITU
HG03871	4925	G>A	0.0210	916	1270	2186	ITU
HG03872	4925	G>A	0.0289	1943	4278	6221	ITU
HG03976	4925	G>A	0.0307	1159	3017	4176	ITU
HG04001	4925	G>A	0.0269	888	1233	2121	ITU
HG04056	4925	G>A	0.0132	1116	1460	2576	ITU
HG04060	4925	G>A	0.0309	1312	2017	3329	ITU
HG04070	4925	G>A	0.0734	726	1957	2683	ITU
HG04080	4925	G>A	0.0358	811	1088	1899	ITU
HG04118	4925	G>A	0.0347	862	1273	2135	ITU
HG04198	4925	G>A	0.0281	1036	3133	4169	ITU
HG04225	4925	G>A	0.0226	1052	2579	3631	ITU
HG01593	4925	G>A	0.0301	786	1106	1892	PJL
HG02603	4925	G>A	0.0214	2479	5828	8307	PJL
HG02652	4925	G>A	0.0297	1077	2969	4046	PJL
HG02657	4925	G>A	0.0536	1742	3953	5695	PJL
HG02658	4925	G>A	0.0579	2107	4613	6720	PJL
HG02690	4925	G>A	0.0296	835	1905	2740	PJL
HG02727	4925	G>A	0.0208	2095	4860	6955	PJL
HG02731	4925	G>A	0.0304	1208	1590	2798	PJL
HG02774	4925	G>A	0.0233	890	2371	3261	PJL
HG03624	4925	G>A	0.0245	1280	1822	3102	PJL
HG03663	4925	G>A	0.0254	2240	5193	7433	PJL
HG03706	4925	G>A	0.0292	1033	1366	2399	PJL
HG03643	4925	G>A	0.0251	2028	5746	7774	STU
HG03645	4925	G>A	0.0251	1273	2985	4258	STU
HG03689	4925	G>A	0.0225	1147	2950	4097	STU
HG03733	4925	G>A	0.0366	1001	2606	3607	STU
HG03740	4925	G>A	0.0468	1276	3617	4893	STU
HG03746	4925	G>A	0.0178	1911	5330	7241	STU
HG03750	4925	G>A	0.0248	1604	4047	5651	STU
HG03755	4925	G>A	0.0266	1822	4045	5867	STU
HG03836	4925	G>A	0.0174	1055	2731	3786	STU
HG03838	4925	G>A	0.0567	2381	6271	8652	STU
HG03844	4925	G>A	0.0226	1573	4182	5755	STU
HG03854	4925	G>A	0.0725	721	1625	2346	STU
HG03857	4925	G>A	0.0275	1104	2681	3785	STU
HG03886	4925	G>A	0.0539	1202	3361	4563	STU
HG03887	4925	G>A	0.0270	1765	4935	6700	STU
HG03894	4925	G>A	0.0250	1216	1579	2795	STU
HG03986	4925	G>A	0.0256	1895	5067	6962	STU
HG03995	4925	G>A	0.0208	1268	3347	4615	STU
HG04229	4925	G>A	0.0261	834	2236	3070	STU

Supplementary Table S10: 1000 Genomes high coverage data set (PCR-free whole genome sequencing data).

Individual data for G4925A carriers from all 1000G high coverage data set. Please note, that the overall coverage in the 1000G high coverage data set is lower than the coverage of the same samples in the exome data (Supplementary Table S9). Thus this analysis has to be seen complementary, but with lower confidence than the exome data set. No sample achieved a coverage >1000X at the G4925A position (555 ± 156 (mean ± SD)). HG01112 is confirmed also in the exome data in Supplementary Table S9, albeit with a considerably higher coverage (3,963X), while HG01565 is not included in Supplementary Table S9 since coverage in the exome data set was only 201X. However, using coverage-unrestricted analysis, HG01565 is indeed detected as G4925A carrier also in the 1000G exome data at a mutation level of 0.0398.

SampleID	POS in reference	Change	Mutation level	Coverage Forward	Coverage Reverse	Total Coverage	Pop
HG01112	4925	G>A	0.038	197	275	472	CLM
HG01565	4925	G>A	0.014	140	199	339	PEL

Supplementary Table S11: Individual data for G4925A carriers in the GenomeAustria project (PCR-free whole genome sequencing data; n=8 whole genome sequenced individuals from Austria).

All other samples were non-carriers.

SampleID	POS in reference	Change	Mutation level	Coverage Forward	Coverage Reverse	Total Coverage
PGA1.K4_6	4925	G>A	0.035	193	208	401
PGA9.K4_6	4925	G>A	0.056	135	150	285

Supplementary Table S12: Classification performance of castPCR assay.

All 123 samples of the NGS discovery phase were re-typed using the castPCR assay as described in the Online Methods of the main manuscript. In brief, the same input amount of each sample was typed in parallel using the wild-type and the mutation detection assay and data was analyzed according manufacturer protocol. Samples showing a deltaCt cut-off <9 were counted as positive carriers. The NGS result was considered as gold standard.

Calls [n]	**123**
True Positives [n]	31
False Positives [n]	0
True Negatives [n]	93
False Negatives [n]	1

Sensitivity [%]	96.9%
Specificity [%]	100.0%
Accuracy [%]	99.2%
Precision (PPV) [%]	100.0%
False Discovery Rate [%]	0.0%
Negative Predictive Value [%]	98.9%

Supplementary Table S13: Crosstable for G4925A and rs75692336 carrier status.

rs75692336 is coded dominantly, since heterozygous and homozygous G4925A carrier status cannot be discerned by castPCR

	rs75692336 non-carrier (CC)	rs75692336 carrier (CA/AA)
G4925A non-carrier	**2189**	**63**
G4925A carrier	**7**	**633**

Supplementary Table S14: Comparison of association with Lp(a) for G4925A and rs75692336

	G4925A: KORAF4			rs75692336: Dominant model in KORAF4 study			rs75692336: Dominant model in GCKD study		
Group	**$\beta_{original}$**	**p[#]**	**R^2 in %**	**$\beta_{original}$**	**p[#]**	**R^2 in %**	**$\beta_{original}$**	**p[#]**	**R^2 in %**
Unadjusted model									
All participants	-8.1	0.006	0.2	-2.3	0.195	0.02	-2.3	0.002	0.2
LMW only	-31.3	5.57e-38	20.6	-21.7	6.10e-21	11.4	-29.5	5.6e-50	16.4
HMW only	-2.5	0.231	0.02	+0.19	7.45E-04	0.5	+0.4	2.04e-08	0.8
Isoform-adjusted model									
All participants	-21.3	1.05E-55	6.1 (*)	-15.7	3.99e-31	3.3 (*)	-19	9.09e-45	2.9 (*)
LMW only	-30.6	3.0e-36	19.3 (*)	-20.9	7.60e-19	10.0 (*)	-29.0	5.98e-47	15.3 (*)
HMW only	-10	1.69e-10	1.6 (*)	-6.5	4.78e-05	0.7 (*)	-6.6	3.39E-03	0.2 (*)

(*) variance explained by G4925A/rs75692336 additionally to apo(a) isoforms

[#] The p-value refers to linear regression analysis on inverse-normal transformed Lp(a) concentrations; the β–estimate is based on the original scale of Lp(a)

Supplementary Table S15: Comparison of *LPA* genotype determined by Western Blot in plasma and by pulsed field gel electrophoresis (PFGE) on genomic DNA level.

N.B.: Since the PFGE fragment assessed in PFGE exceeds 100 kb for any isoform >27 KIV repeats, the differences between WB and PFGE in the larger isoform are expected and are due to the more difficult separation of larger fragments, respectively the optimization of the PFGE protocol for smaller isoforms.

A truncation of the protein at the last KIV-2 repeat would, however, result in a molecular weight difference corresponding to about 11 KIV repeats. Such a large difference is not visible for any of the G4925A carriers, indicating that no truncated protein form occurs.

#	Sample ID	G4925A carrier status	Lp(a) (mg/dL)	WB Isoform 1	WB Isoform 2	PFGE Isoform 1	PFGE Isoform 2
1	AK5	No	7.1	35	39	30	35
2	AK7	No	45.2	19	32	21	28
3	AK10	No	12.3	24	31	24	29
4	AK12	No	1.5	21	28	23	26
5	AK19	No	41.7	20	30	21	27
6	AK29	No	16.5	28	32	25	28
7	AK17	**Yes**	9.6	24	38	24	32
8	AK23	**Yes**	11.2	24	28	24	27
9	AK28	**Yes**	8.1	24	36	23	30
10	AK33	**Yes**	3.5	20	23	22	23
11	AK38	**Yes**	14.3	22	38	23	33

Supplementary Table S16: qPCR Protocol

Primer fw (5'-3')	GTGGCAGCTCCTTATTGTTATACG
Primer rv (5'-3')	CGACGGCAGTCCCTTCTG
Probe (5'-3')	TACTGCAACCTGACGCAATGCTCAGAC
Reporter/ Quencher	FAM / BHQ1
Initial denaturation	95 °C, 10 min
Denaturation	95 °C, 15 s
Annealing / Extension	60 °C, 1 min
Number of cycles	40
Manual Ct threshold setting	0.2
Reaction mix	Type-it Fast SNP Probe PCR Kit qPCR Mastermix (Qiagen, Hilden, Germany)

Primers and probes were purchased from Microsynth AG, Balgach, CH

Supplementary Table S17: PCR protocol for deletion of MCS from pEGFP-C1 according to [20]

Product length (based on hg19)	4,731 (4,652 with deleted MCS)
Reaction vol. [µl]	50
Enzyme	Agilent Herculase II Fusion
ID Primer fw *	pEGFP-C1-del-fw
ID Primer rv *	pEGFP-C1-del-rv
Final primer conc. [µM each]	0.25 µM
Final dNTP concentration [mM]	0.25 mM
Enzyme amount [µl)	1
DNA input [ng]	5
PCR cycling protocol	
Initial denaturation	95 °C, 4 min
Cycling block 1, denaturation	95 °C, 20 s
Cycling block 1, annealing	66 °C, 40 s
Cycling block 1, extension	72 °C, 2 min 40 s
Number of cycles, block 1	20
Cycling block 2, denaturation	95 °C, 20 s
Cycling block 2, annealing	66 °C, 40 s
Cycling block 2, extension	72 °C, 2 min 40 s
Number of cycles, block 2	1

* For sequences see Supplementary Table S6

Supplementary Table S18: PCR protocols for RT-PCRs on liver biopsy

	RT-PCR for NGS	RT-PCR for exon skipping detection	GAPDH PCR for pre-mRNA	LPA PCR for G4925A context sequence in pre-mRNA
Product length (based on Ensembl 75, ENST00000316300)	275	335 (152 if exon is skipped)	113 bp spliced, 205 bp unspliced	324 bp
Reaction vol. [µl]	20	20	20	20
Enzyme	Qiagen Hotstar Taq DNA Polymerase	Qiagen Hotstar Taq DNA Polymerase	Qiagen Hotstar Taq DNA Polymerase	Qiagen Hotstar Taq DNA Polymerase
Initial denaturation	95 °C, 15 min	95 °C, 15 min	95 °C, 15 min	95 °C, 15 min
Denaturation	94 °C, 30 s	94 °C, 30 s	94 °C, 30 s	94 °C, 30 s
Annealing	60 °C, 30 s	60 °C, 30 s	60 °C, 30 s	60 °C, 30 s
Extension	72 °C, 15 s	72 °C, 30 s	72 °C, 60 s	72 °C, 60 s
Final extension	72 °C, 10 min	72 °C, 10 min	72 °C, 10 min	72 °C, 10 min
Number of cycles	30	35	40	40
Primer fw	LPA_Ex4_fw	LPA_exon-skip_fw1	GAPDH for	LPA Ex4fw
Primer rv	LPA_Ex5_rv	LPA_ exon-skip _rv1	GAPDH rev	422L
Final primer conc. [µM each]	0.5 µM	0.5 µM	0.5 µM	0.5 µM
Final dNTP concentration [mM]	0.2 mM	0.2 mM	0.2 mM	0.2 mM
Enzyme amount [u]	2.5	2.5	0.5	0.5
cDNA input [µl]	1	1	1	1

Supplementary Table S19: Primer sequences protocols for RT-PCRs on liver biopsy

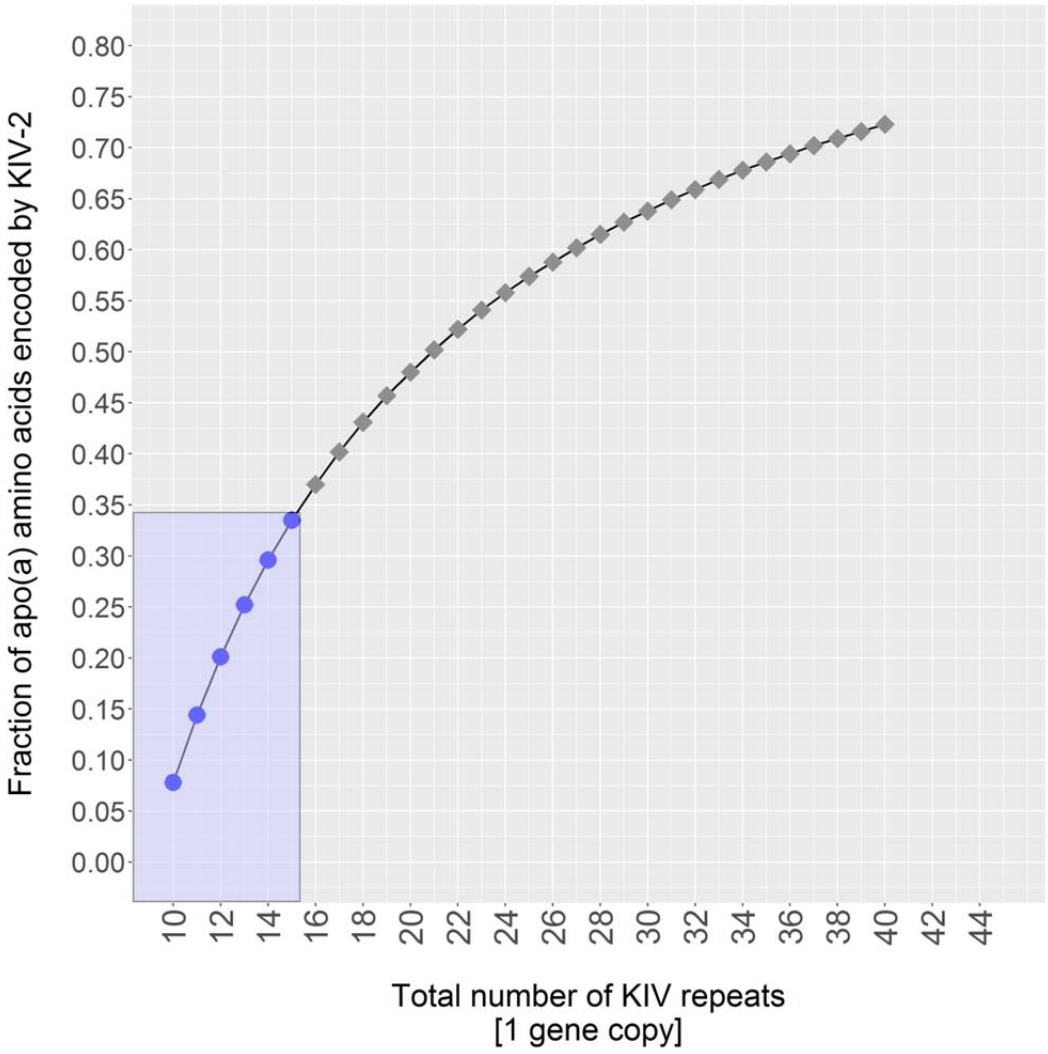
Name	Sequence 5'-3'
LPA_ exon-skip _fw1	GGACAGAGTTATCGAGGCACATACT
LPA_ exon-skip _rv1	TGGTAGCACTCCTGCACCCCA
GAPDH for	CATGAGAAGTATGACAACAGCCT
GAPDH rev	AGTCCTTCCACGATACCAAAGT
LPA Ex4fw	GTGGCAGCTCCTTATTGTTATACG
422L	CACCAGAAATCACTCCGCTG
LPA_Ex5_fw	GGACAGAGTTATCGAGGCACATACT
LPA_Ex5_rv_skipping	TGGTAGCACTCCTGCACCCCA

Supplementary Figures

Supplementary Figure S1

Percentage of apo(a) amino acids encoded by KIV-2 depending on the number of total repeats present (calculation for one gene allele). The blue dots represent the fraction of amino acids present in the gene annotation according to hg38, which contains 15 KIV copies (i.e. 6 KIV-2 repeats). In hg38, KIV-2 encompasses only the shaded fraction of total apo(a) amino acids. In the general population, however, up to 45 KIV copies can be present (grey data points), encoding up to 70 percent of the protein (based on one gene allele).

The calculation is based on the contribution of 114 amino acids by each KIV-2 repeat, while the invariant apo(a) region consists of 1,356 amino acids.



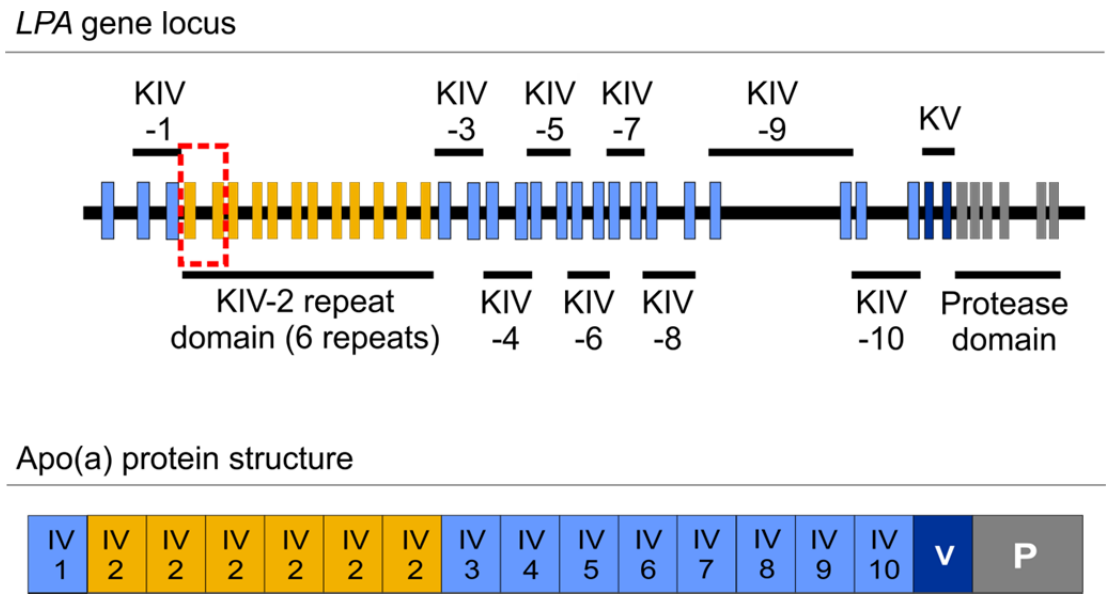
Supplementary Figure S2

Structure of *LPA* and apo(a)

Top panel. Structure of the *LPA* gene locus.

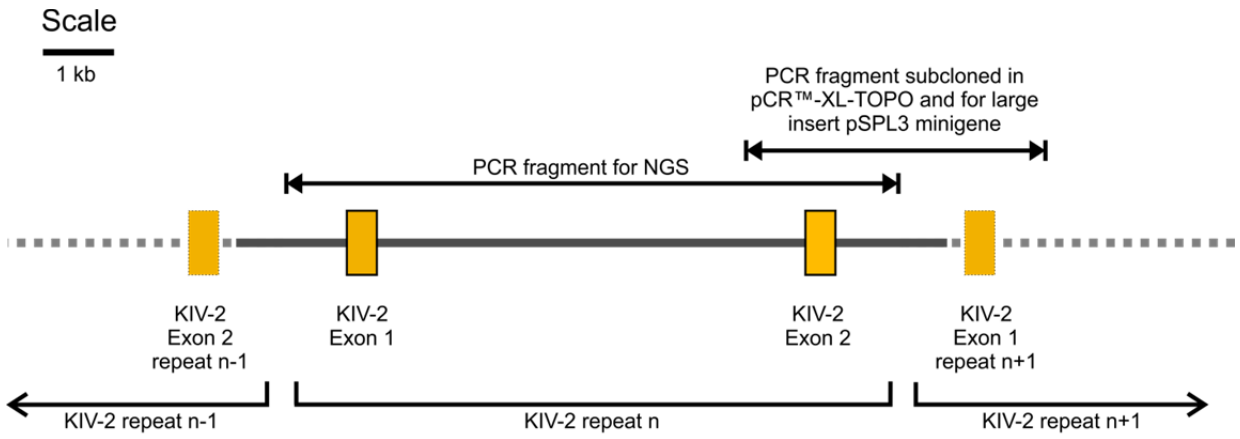
The gene structure, which is actually located on the minus strand, has been reverse-complemented for better comprehensibility and the kringle domains annotated above and below the gene structure. Each KIV repeat consists of two exons separated by a large (4 kb) intron. A small intron (1.2 kb) separates two kringles. The repetitive KIV-2 domain is colored in orange and given with 6 repeats as present in hg38. The G4925A variant is located at the last exonic base of the second exon of the KIV-2 repeat. The batch sequencing amplification primers encompass each KIV-2 element (red dotted box shows one element; see also Supplementary Fig. 3). Two previously described splice site mutations are located in KIV-7 (rs41272114[30]) and KIV-10 (rs143431368[29]).

Bottom panel. Structure of the apo(a) protein. Color code of top panel is used. The 1A2 antibody employed in this work is directed against KIV-2.



Supplementary Figure S3

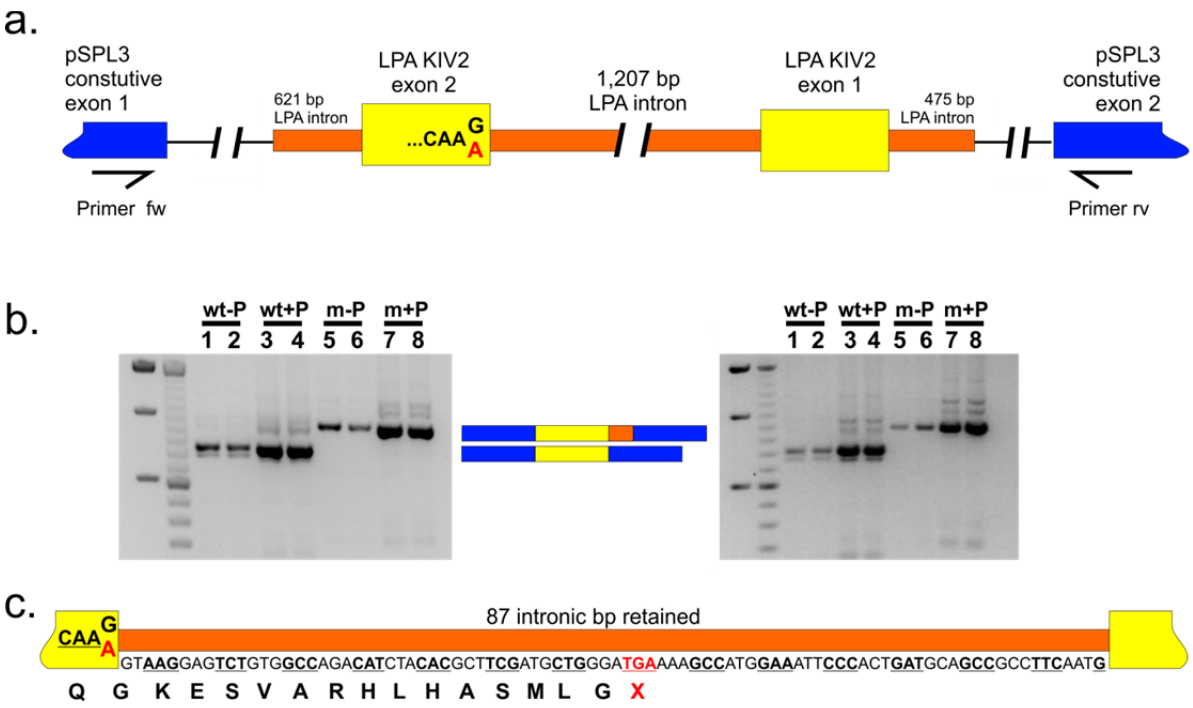
Schematic representation of KIV-2 structure and PCR fragments used for NGS batch sequencing and minigene assays. G4925A is located at the last 3' base of KIV-2 exon 2.



Supplementary Figure S4

Results of the pSPL3 minigene assays.

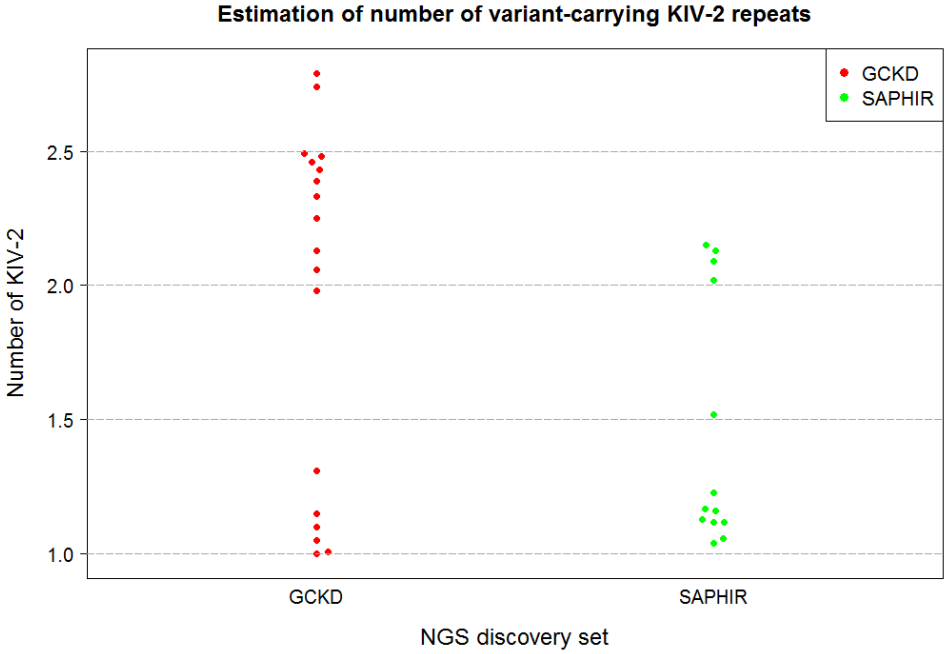
a. Structure of pSPL3 minigene. The insert encompasses the *LPA* KIV-2 exon 2 (G4925A shown in red letters) and 621 bp intronic upstream sequence, the full following intron (1.2 kb), the KIV-2 exon 1 of the next repeat element and 475 bp downstream sequence. **b.** RT-PCR image (agarose, 2.0%; both biological replicates with two technical replicates each are depicted) showing the effect of G4925A on exon splicing. wt: wild type, m: mutant, +/- P: puromycin addition for non-sense mediated mRNA decay inhibition. The mutant band shows a pronounced shift in length caused by 87 bp intron retention (**c.**) causing the product to shift from 606 to 693 bp. The intron sequence contains an opal stop codon (highlighted). The codons are underlined according to the reading frame. The translated product is given below. The gel image (b.) of the minigene has been cropped for concise presentation. Full gel images including all replicates are given in Supplementary Fig. 9.



Supplementary Figure S5

Estimation of number of KIV-2 repeats carrying G4925A

Based on total number of KIV-2 divided by the mutation level observed in NGS by the total number of KIV-2 repeats estimated by KIV-2 qPCR. Given the inherent imprecision of qPCR in estimation of the exact number of KIV-2 repeats, the values must be seen as approximations.

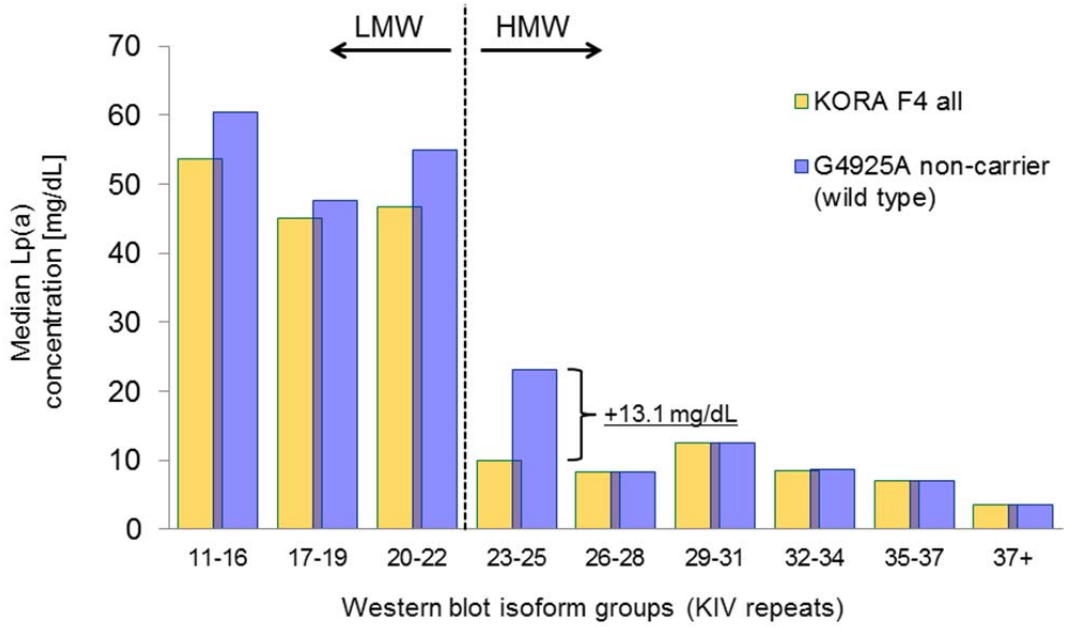


Supplementary Figure S6

Median values of Lp(a) concentrations for the different isoforms groups in KORA F4.

We have already previously reported, that the Lp(a) levels drop quite sharply at 23 KIV[32]. It is therefore intriguing, that this corresponds, to the isoform group with the largest percentage of G4925A carriers. This figure shows that by excluding the G4925A carriers from this evaluation, this reduction is mitigated and the median in the 23 KIV group increases by 13.1 mg/dL. This implies that G4925A might be a molecular cause for this pronounced reduction at KIV 23.

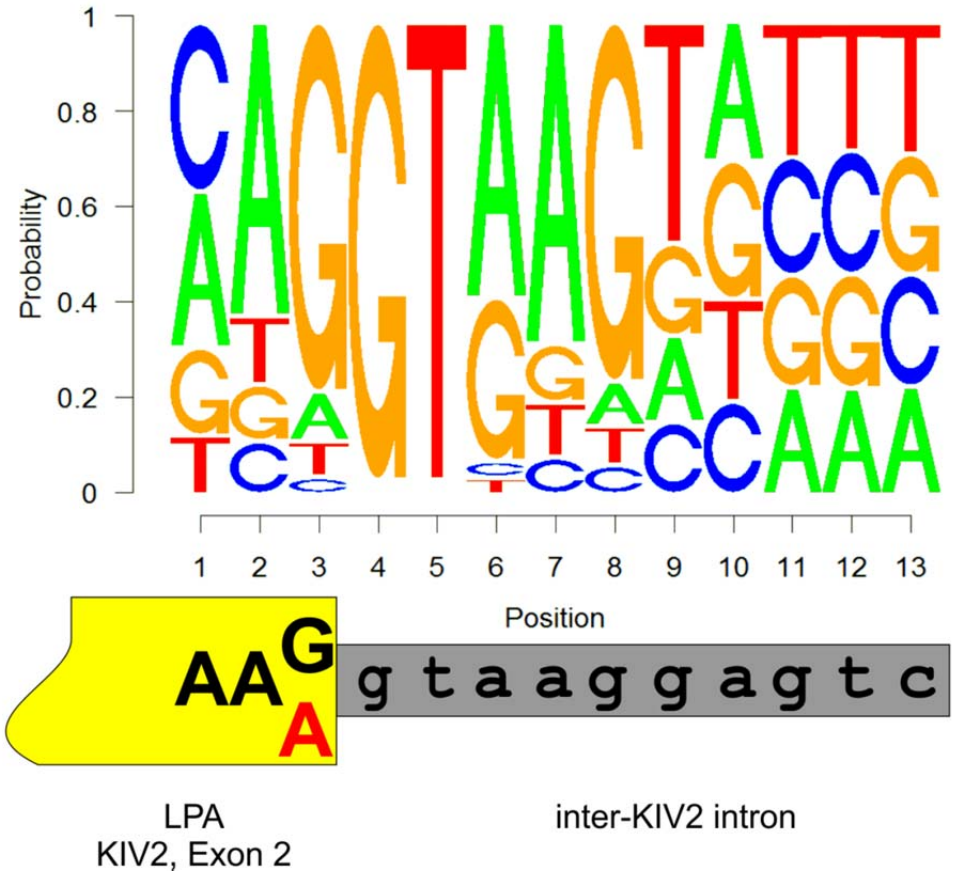
Group sizes were as follows ($n_{all}/n_{non.carrier}$): 11-16: 70/56, 17-19: 196/145, 20-22: 450/318, 23-25: 615/221, 26-28: 462/445, 29-31: 428/416, 32-34: 254/247, 35-37: 260/255, 37+:131/127



Supplementary Figure S7

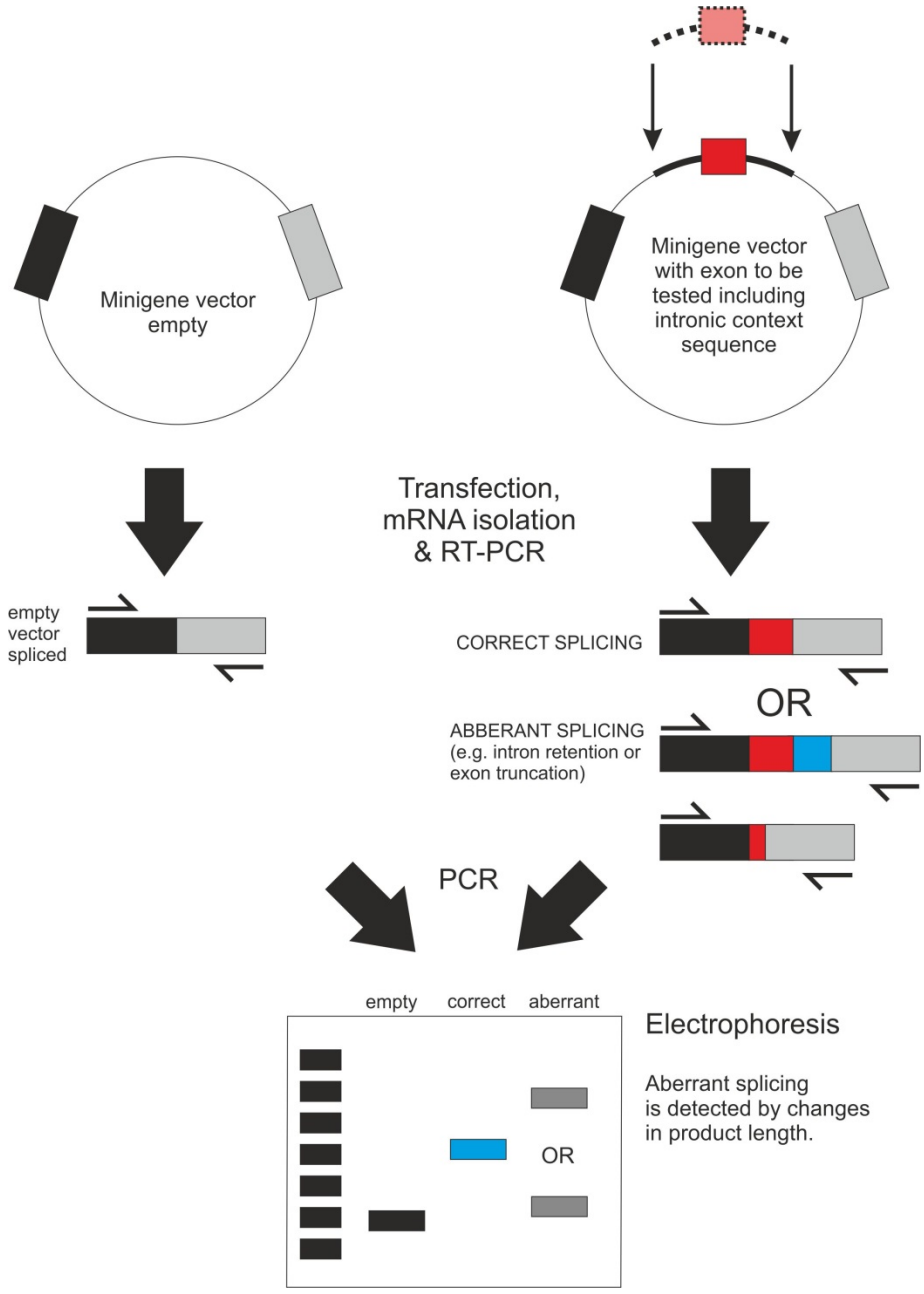
U1/U2 consensus splice site sequence. The gene structure and the actual sequence of the exon-intron boundary at the position of G4925A is shown below the sequence logo. The splice site at G4925A belongs to the U1/U2 subtype, which is characterized by presence of an adenine at the last exonic position in about 10% of all sites[33]. This hampers conclusive in silico prediction of biological effects of G4925A and requires in-vitro functional testing.

Sequence logo was generated using R package seqLogo[34] using probability data from the SpliceRack Website[35] (http://katahdin.mssm.edu/splice/poster_data.html).



Supplementary Figure S8

Minigene assay[36]**.** The exon and the intronic context sequence, respectively the genomic region to be tested (red box and dotted line) is cloned into a vector containing two constitutive exons (black and grey box), spaced by an intron. The 3' constitutive exon contains a polyadenylation site to ensure mRNA production. After transfection in an appropriate cell line, mRNA is isolated and reverse transcribed. The recombinant mRNA is amplified using primers directed against the constitutive exons. The PCR fragment size reports any unexpected changes in the mRNA fragment length, which indicates changes in splicing behavior. The change is resolved by isolating and sequencing the products. This allows detection of any changes without prior knowledge. For mutation evaluation, both alleles have to be tested and compared.



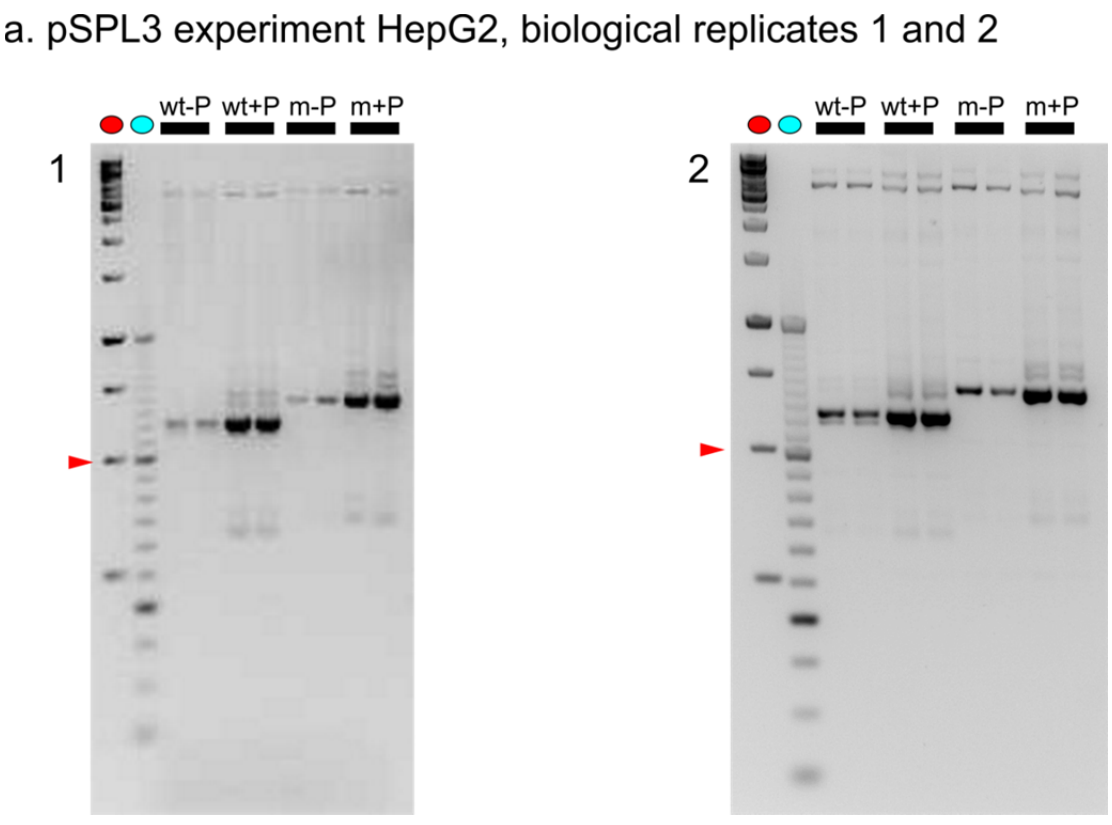
Supplementary Figure S9

Minigene experiments.

pSPL3 minigenes were transfected into HepG2 cells. Minigene structure and interpretation is shown in Figure 3 of the main manuscript. The experiment was carried out in two biological replicates with two technical replicates for each condition.

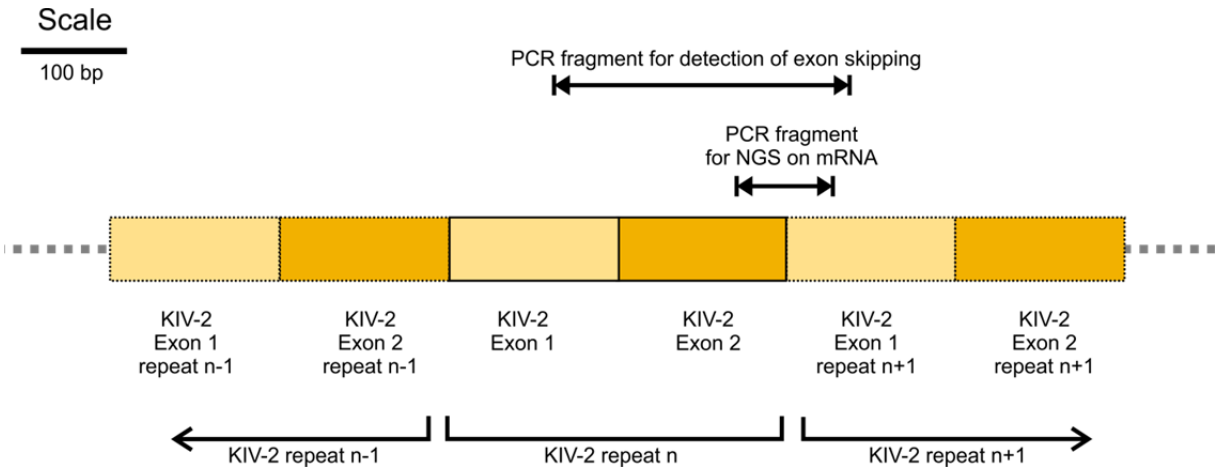
wt: wild type construct. m: mutant construct. +/- P: puromycin addition for non-sense mediated mRNA decay inhibition. Two different size markers are present on each gel: DNA-ladder 50 bp peqGOLD (lower molecular weight bands, blue circle, 50-1000 bp; VWR) and DNA-ladder 1 kb peqGOLD (250 bp – 10,000 bp; red circle VWR).

A red arrow marks the 500 bp marker band in each gel. The expected product size of the wild type fragment is 606 bp. No indication for pronounced non-sense mediated decay is observed, as primary band pattern is similar between the + and – puromycin group. In all experiments the same shift in the major band was observed, as determined by sequencing.



Supplementary Figure S10

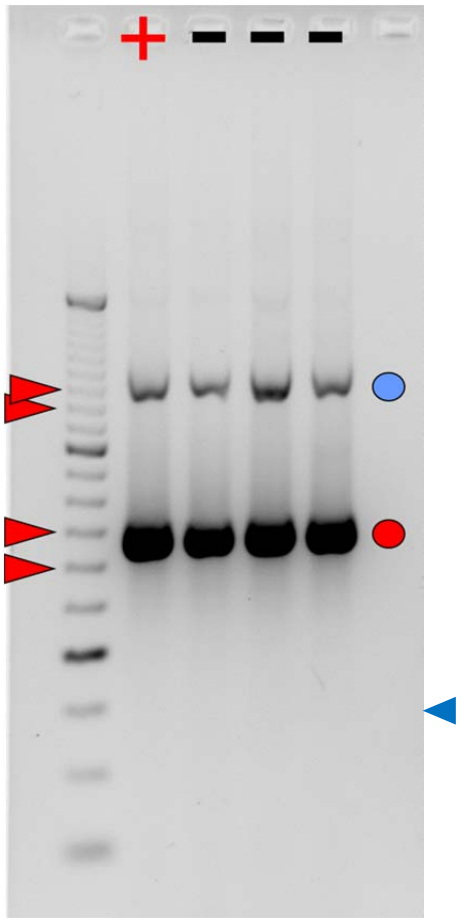
PCR fragment design on mRNA. For better comprehensibility only one fragment is depicted but, similar to the batch PCR on DNA, the primer bind in each repeat and thus amplify again all repeats simultaneously. G4925A is located at the last 3' base of KIV-2 exon 2.



Supplementary Figure S11

Agarose gel for a RT-PCR spanning KIV -2 exon 2 for 4 human liver biopsies (1 G4925A-positive sample (+), 3 controls (-))

The expected wild type product is 335 bp. In case of exon skipping, a 153 bp product would be expected. No exon skipping was observed. Most of the PCR product shows the expected length of 335 bp (red circle). Moreover a second, longer band at about ~650 was observed (blue circle). Since the primer bind in every repeat, this corresponds to a PCR product spanning two KIV-2 repeats. Each ladder band corresponds to a 50 bp step, starting at 50 bp. The bands at 300, 350, 600 and 650 bp are marked by a red triangle (from bottom to top). The approximate position of the expected exon skipping product is marked a blue triangle (Supplementary Figure S10)



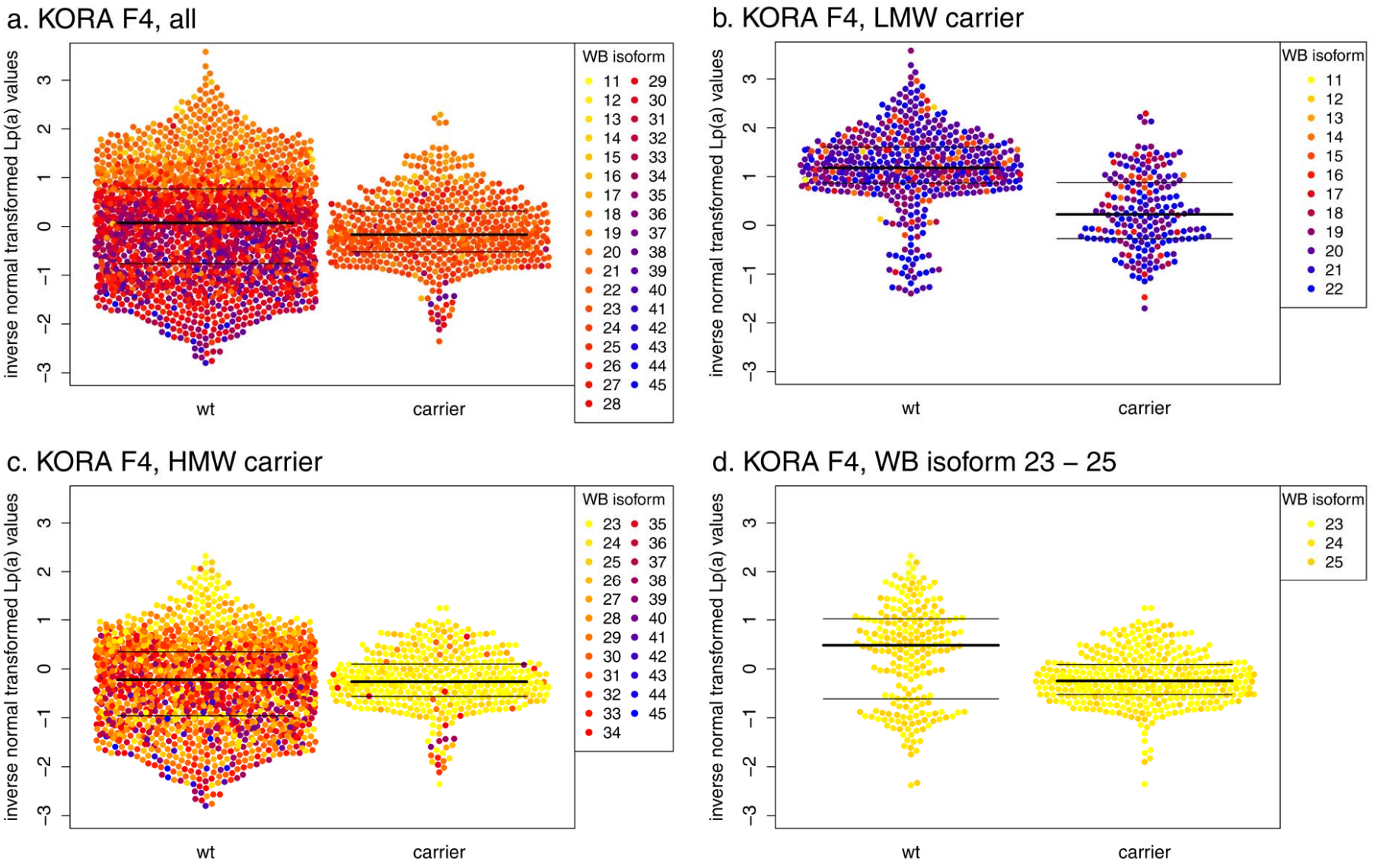
Supplementary Figure S12

Bee-swarm scatter plot of inverse normal transformed Lp(a) values by carrier status in **a)** the total KORA F4 study, **b)** LMW apo(a) isoform carriers, **c)** HMW apo(a) isoform carriers and **d)** subgroup of the HMW carriers with WB isoforms 23-25. Horizontal lines indicate the median level (bold line) and the 25th and 75th percentile. The color coding indicates the small (respectively only) isoform.

Especially for the HMW group, it becomes obvious that carriers are clustered at isoform 23. In the non-carriers, individuals with isoform 23 (yellow dots) present rather high Lp(a) values (note the yellow top tail in wild type samples (wt) in panel c), while in G4925A carriers the Lp(a) levels are reduced (carriers in panels c and d). This effect becomes detectable only after isoform adjustment, since the reduced levels present in G4925A are close to the group's median and therefore the reduction is missed when the two groups are compared without taking isoforms into account.

[Figure on next page]

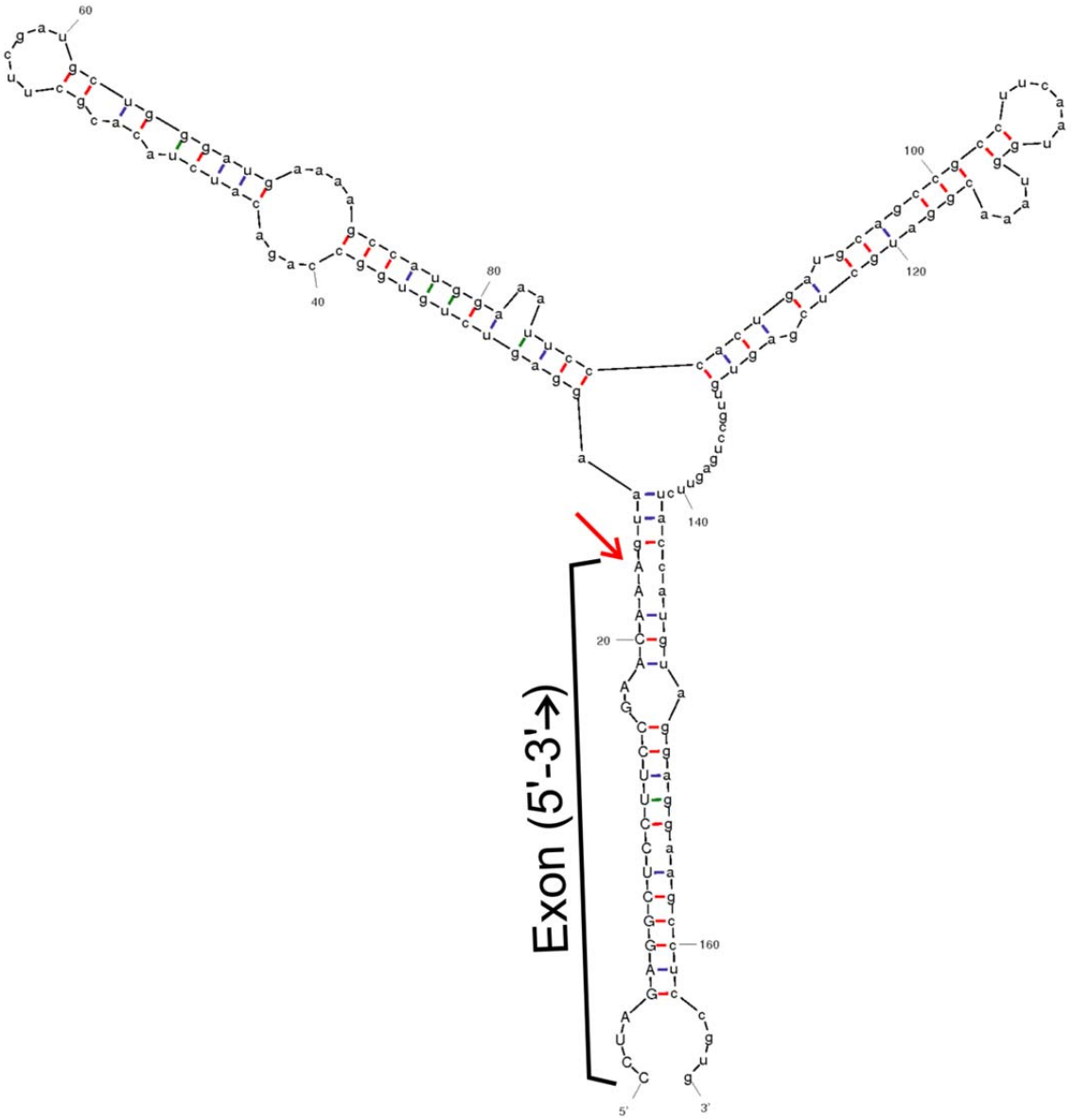
.


a. KORA F4, all
b. KORA F4, LMW carrier
c. KORA F4, HMW carrier
d. KORA F4, WB isoform 23 – 25
inverse normal transformed Lp(a) values
−3
−2
−1
0
1
2
3
wt
carrier
WB isoform
11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31 32 33 34 35 36 37 38 39 40 41 42 43 44 45

Supplementary Figure S13

RNA structure at the 3' end of the KIV-2 exon 2

Structure as provided by mFold[22] (http://unafold.rna.albany.edu/?q=mfold/RNA-Folding-Form) using standard settings. The region shows extensive double stranded structures, which are a prerequisite for RNA editing by ADARs. The position of the G4925A variant is marked by a red arrow.

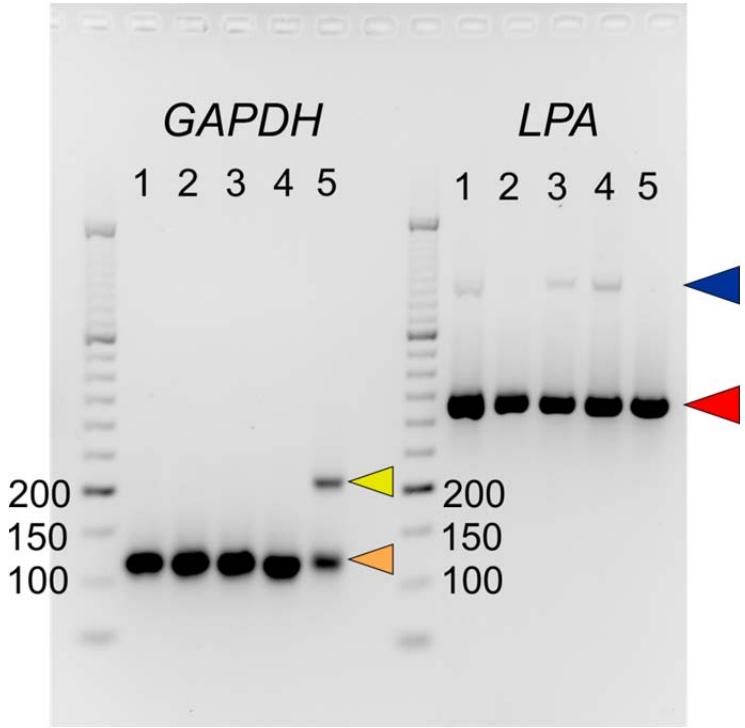


Supplementary Figure S14

PCR amplicons amplifying the G4925A context sequence from pre-mRNA and control reactions on *GAPDH* of the four liver biopsy RNA samples.

Sample 1: the G4925A carrier, samples 2-4: G4925A-negative controls; Sample 5: genomic DNA sample (control). All RNA templates were treated with dsDNase. LPA pre-mRNA product was readily amplifiable from RNA samples from all four liver biopsies tested (red arrow). The faint larger band (blue arrow) corresponds to an amplification spanning two KIV-2 repeats. 2 ng of these products were diluted 1:50,000 and 1:500,000 and tested by castPCR for G4925A.

To control for genomic DNA contamination, a control amplicon spanning intron 6 of GAPDH was amplified from each RNA sample. Only products corresponding to a spliced template are visible (113 bp; orange arrow). A reaction on genomic DNA (5) showed the expected, larger band at 205 bp containing the intron (yellow arrow). The smaller band at the same length as the mRNA product likely stems from the known *GAPDH* retrotransposed pseudogenes on chromosomes 4, 5 and 6 (Pseudogene.org accession numbers: PGOHUM00000245994, PGOHUM00000235837, PGOHUM00000243261 as reported in UCSC Genome Browser hg19). Each marker band corresponds to 50 bp (range: 50 to 1,000 bp).



References

1. Heid IM, Wagner SA, Gohlke H, Iglseder B, Mueller JC, Cip P, Ladurner G, Reiter R, Stadlmayr A, Mackevics V, Illig T, Kronenberg F, Paulweber B. Genetic architecture of the APM1 gene and its influence on adiponectin plasma levels and parameters of the metabolic syndrome in 1,727 healthy Caucasians. *Diabetes* 2006;**55**:375–384.

2. Eckardt K-U, Bärthlein B, Baid-Agrawal S, Beck A, Busch M, Eitner F, Ekici AB, Floege J, Gefeller O, Haller H, Hilge R, Hilgers KF, Kielstein JT, Krane V, Köttgen A, Kronenberg F, Oefner P, Prokosch H-U, Reis A, Schmid M, Schaeffner E, Schultheiss UT, Seuchter SA, Sitter T, Sommerer C, Walz G, Wanner C, Wolf G, Zeier M, Titze S. The German Chronic Kidney Disease (GCKD) study: design and methods. *Nephrol Dial Transplant* 2012;**27**:1454–1460.

3. Titze S, Schmid M, Köttgen A, Busch M, Floege J, Wanner C, Kronenberg F, Eckardt KU. Disease burden and risk profile in referred patients with moderate chronic kidney disease: Composition of the German Chronic Kidney Disease (GCKD) cohort. *Nephrol Dial Transplant* 2015;**30**:441–451.

4. Levey AS. A New Equation to Estimate Glomerular Filtration Rate. *Ann Intern Med* 2009;**150**:604.

5. Raschenberger J, Kollerits B, Titze S, Köttgen A, Bärthlein B, Ekici AB, Forer L, Schönherr S, Weissensteiner H, Haun M, Wanner C, Eckardt K-U, Kronenberg F, GCKD study Investigators. Association of relative telomere length with cardiovascular disease in a large chronic kidney disease cohort: the GCKD study. *Atherosclerosis* 2015;**242**:529–534.

6. Teslovich TM, Musunuru K, Smith A V, Edmondson AC, Stylianou IM, Koseki M, Pirruccello JP, Ripatti S, Chasman DI, Willer CJ, Johansen CT, Fouchier SW, Isaacs A, Peloso GM, Barbalic M, Ricketts SL, Bis JC, Aulchenko YS, Thorleifsson G, Feitosa MF, Chambers J, Orho-Melander M, Melander O, Johnson T, Li X, Guo X, Li M, Shin CY, Jin GM, Jin KY, et al. Biological, clinical and population relevance of 95 loci for blood lipids. *Nature* Center for Statistical Genetics, Department of Biostatistics, University of Michigan, Ann Arbor, Michigan 48109, USA; 2010;**466**:707–713.

7. Heid IM, Boes E, Müller M, Kollerits B, Lamina C, Coassin S, Gieger C, Döring A, Klopp N, Frikke-Schmidt R, Tybjaerg-Hansen A, Brandstätter A, Luchner A, Meitinger T, Wichmann H-E, Kronenberg F. Genome-wide association analysis of high-density lipoprotein cholesterol in the population-based KORA study sheds new light on intergenic regions. *Circ Cardiovasc Genet* 2008;**1**:10–20.

8. Kronenberg F, Kuen E, Ritz E, Junker R, König P, Kraatz G, Lhotta K, Mann JF, Müller GA, Neyer U, Riegel W, Reigler P, Schwenger V, Eckardstein A Von, Konig P, Kraatz G, Lhotta K, Mann JF, Muller GA, Neyer U, Riegel W, Reigler P, Schwenger V, Eckardstein A Von. Lipoprotein(a) serum concentrations and apolipoprotein(a) phenotypes in mild and moderate renal failure. *J Am Soc Nephrol* 2000;**11**:105–115.

9. Kraft HG, Lingenhel A, Bader G, Kostner GM, Utermann G. The relative electrophoretic mobility of apo(a) isoforms depends on the gel system: proposal of a nomenclature for apo(a) phenotypes. *Atherosclerosis* Institute for Medical Biology and Human Genetics, University of Innsbruck, Austria; 1996;**125**:53–61.

10. Lanktree MB, Rajakumar C, Brunt JH, Koschinsky ML, Connelly PW, Hegele RA. Determination of lipoprotein(a) kringle repeat number from genomic DNA: copy number variation genotyping using qPCR. *J Lipid Res* 2009;**50**:768–772.

11. Livak KJ, Schmittgen TD. Analysis of relative gene expression data using real-time quantitative PCR and the 2(-Delta Delta C(T)) Method. *Methods* Applied Biosystems, Foster City, California 94404, USA; 2001;**25**:402–408.

12. Payne BAI, Gardner K, Coxhead J, Chinnery PF. Deep resequencing of mitochondrial DNA. *Methods Mol Biol* 2015;**1264**:59–66.

13. Weissensteiner H, Forer L, Fuchsberger C, Schöpf B, Kloss-Brandstätter A, Specht G, Kronenberg F, Schönherr S. mtDNA-Server: next-generation sequencing data analysis of human mitochondrial DNA in the cloud. *Nucleic Acids Res* 2016;**44**:W64-9.

14. Ye K, Lu J, Ma F, Keinan A, Gu Z. Extensive pathogenicity of mitochondrial heteroplasmy in healthy human individuals. *Proc Natl Acad Sci* 2014;**111**:10654–10659.

15. Ratan A, Olson TL, Loughran TP, Miller W. Identification of indels in next-generation sequencing data. *BMC Bioinformatics* 2015;**16**:42.

16. Auton A, Abecasis GR, Altshuler DM, Durbin RM, Abecasis GR, Bentley DR, Chakravarti A, Clark AG, Donnelly P, Eichler EE, Flicek P, Gabriel SB, Gibbs RA, Green ED, Hurles ME, Knoppers BM, Korbel JO, Lander ES, Lee C, Lehrach H, Mardis ER, Marth GT, McVean GA, Nickerson DA, Schmidt JP, Sherry ST, Wang J, Wilson RK, Gibbs RA, Boerwinkle E, et al. A global reference for human genetic variation. *Nature* 2015;**526**:68–74.

17. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R. The Sequence Alignment/Map format and SAMtools. *Bioinformatics* 2009;**25**:2078–2079.

18. Morlan J, Baker J, Sinicropi D. Mutation detection by real-time PCR: a simple, robust and highly selective method. *PLoS One* 2009;**4**:e4584.

19. Noureen A, Fresser F, Utermann G, Schmidt K. Sequence Variation within the KIV-2 Copy Number Polymorphism of the Human LPA Gene in African, Asian, and European Populations. *PLoS One* 2015;**10**:e0121582.

20. Liu H, Naismith JH. An efficient one-step site-directed deletion, insertion, single and multiple-site plasmid mutagenesis protocol. *BMC Biotechnol* 2008;**8**:91.

21. Houdayer C, Dehainault C, Mattler C, Michaux D, Caux-Moncoutier V, Pages-Berhouet S, D'Enghien CD, Lauge A, Castera L, Gauthier-Villars M, Stoppa-Lyonnet D. Evaluation of in silico splice tools for decision-making in molecular diagnosis. *Hum Mutat* 2008;**29**:975–982.

22. Zuker M. Mfold web server for nucleic acid folding and hybridization prediction. *Nucleic Acids Res* 2003;**31**:3406–3415.

23. Kronenberg F, Utermann G. Lipoprotein(a): resurrected by genetics. *J Intern Med* 2013;**273**:6–30.

24. Erqou S, Thompson A, Angelantonio E Di, Saleheen D, Kaptoge S, Marcovina S, Danesh J. Apolipoprotein(a) isoforms and the risk of vascular disease: systematic review of 40 studies involving 58,000 participants. *J Am Coll Cardiol* 2010;**55**:2160–2167.

25. Kinpara K, Okada H, Yoneyama A, Okubo M, Murase T. Lipoprotein(a)-cholesterol: a significant component of serum cholesterol. *Clin Chim Acta* 2011;**412**:1783–1787.

26. Willer CJ, Schmidt EM, Sengupta S, Peloso GM, Gustafsson S, Kanoni S, Ganna A, Chen J, Buchkovich ML, Mora S, Beckmann JS, Bragg-Gresham JL, Chang HY, Demirkan A, Hertog HM Den, Do R, Donnelly LA, Ehret GB, Esko T, Feitosa MF, Ferreira T, Fischer K, Fontanillas P, Fraser RM, Freitag DF, Gurdasani D, Heikkila K, Hypponen E, Isaacs A, Jackson AU, et al. Discovery and refinement of loci associated with lipid levels. *Nat Genet* 2013;**45**:1274–1283.

27. Marchini J, Howie B, Myers S, McVean G, Donnelly P. A new multipoint method for genome-wide association studies by imputation of genotypes. *Nat Genet* 2007;**39**:906–913.

28. Fuchsberger C, Forer L, Schönherr S, Kronenberg F, Abecasis G. Imputation server: next generation genotype imputation service. *Present 64th Annu Meet Am Soc Hum Genet Oct 20, 2014, San Diego, CA, USA* 2014;

29. Lim ET, Würtz P, Havulinna AS, Palta P, Tukiainen T, Rehnström K, Esko T, Mägi R, Inouye M, Lappalainen T, Chan Y, Salem RM, Lek M, Flannick J, Sim X, Manning A, Ladenvall C, Bumpstead S, Hämäläinen E, Aalto K, Maksimow M, Salmi M, Blankenberg S, Ardissino D, Shah S, Horne B, McPherson R, Hovingh GK, Reilly MP, Watkins H, et al. Distribution and medical impact of loss-of-function variants in the Finnish founder population. *PLoS Genet* 2014;**10**:e1004494.

30. Ogorelkova M, Gruber A, Utermann G. Molecular basis of congenital lp(a) deficiency: a frequent apo(a) 'null' mutation in caucasians. *Hum Mol Genet* 1999;**8**:2087–2096.

31. Bottillo I, Luca A De, Schirinzi A, Guida V, Torrente I, Calvieri S, Gervasini C, Larizza L, Pizzuti A, Dallapiccola B. Functional analysis of splicing mutations in exon 7 of NF1 gene. *BMC Med Genet* IRCCS-CSS, San Giovanni Rotondo and CSS-Mendel Institute, Rome, Italy; 2007;**8**:4.

32. Laschkolnig A, Kollerits B, Lamina C, Meisinger C, Rantner B, Stadler M, Peters A, Koenig W, Stöckl A, Dähnhardt D, Böger CA, Krämer BK, Fraedrich G, Strauch K, Kronenberg F. Lipoprotein (a) concentrations, apolipoprotein (a) phenotypes, and peripheral arterial disease in three independent cohorts. *Cardiovasc Res* 2014;**103**:28–36.

33. Sibley CR, Blazquez L, Ule J. Lessons from non-canonical splicing. *Nat Rev Genet* 2016;**17**:407–421.

34. Bembom O. seqLogo: Sequence logos for DNA sequence alignments. 2016.

35. Sheth N. Comprehensive splice-site analysis using comparative genomics. *Nucleic Acids Res* 2006;**34**:3955–3967.

36. Gaildrat P, Killian A, Martins A, Tournier I, Frébourg T, Tosi M. Use of Splicing Reporter Minigene Assay to Evaluate the Effect on Splicing of Unclassified Genetic Variants. *Methods Mol Biol* 2010;**653**:249–257.