Supplementary Appendix

This appendix has been provided by the authors to give readers additional information about their work.

Supplement to: Steers NJ, Li Y, Drace Z, et al. Genomic mismatch at *LIMS1* locus and kidney allograft rejection. N Engl J Med 2019;380:1918-28. DOI: 10.1056/NEJMoa1803731

SUPPLEMENTAL INFORMATION:

Genomic Mismatch at *LIMS1* Locus is Associated with Kidney Allograft Rejection

Nicholas J. Steers, Yifu Li, Zahida Drace, Justin A. D'Addario, Clara Fischman, Lili Liu, Katherine Xu, Young-Ji Na, Y. Dana Neugut, Jun Y. Zhang, Roel Sterken, Olivia Balderes, Drew Bradbury, Nilgun Ozturk, Fatih Ozay, Sanya Goswami, Karla Mehl, Jaclyn Wold, Fatima Zahra Jelloul, Mersedeh Rohanizadegan, Christopher E. Gillies, Elena-Rodica M. Vasilescu, George Vlad, Yi-An Ko, Sumit Mohan, Jai Radhakrishnan, David J. Cohen, Lloyd Rattner, Francesco Scolari, Katalin Susztak, Matthew G. Sampson, Silvia Deaglio, Yasar Caliskan, Jonathan Barasch, Aisling E. Courtney, Alexander P. Maxwell, Amy Jayne McKnight, Iuliana Ionita-Laza, Stephan J.L. Bakker, Harold Snieder, Martin H. de Borst, Vivette D'Agati, Antonio Amoroso, Ali G. Gharavi and Krzysztof Kiryluk

Table of Contents

SUPPLEMENTAL TEXT:

Supplemental Results 3
Supplemental Methods 4
Brief Glossary 9

SUPPLEMENTAL FIGURES:

Figure S1. Rejection-free allograft survival by recipient genotype at rs893403 10
Figure S2. Rejection-free allograft survival by donor genotype, stratified by recipient's genotype 11
Figure S3. Rejection-free allograft survival in GG recipients by all donor genotype classes 12
Figure S4. Rejection Events by Banff Type and Grade 13
Figure S5. Allograft survival by the "genomic collision" status 14
Figure S6. The distribution of rs893403 alleles across worldwide HGDP populations 15
Figure S7. CNVR915.1 deletion breakpoints based on 1000G WGS data 16
Figure S8. Genomic location of rs893403 and CNVR915.1 and their functional annotation 17
Figure S9. FUN-LDA scoring of rs893403 and its proxies by genomic location 18
Figure S10. LIMS1 mRNA gene expression and cis-eQTL effect of rs893403 19
Figure S11. Effect of rs893403 on LIMS1 mRNA gene expression across GTEx tissues 20
Figure S12. GCC2 mRNA gene expression and cis-eQTL effect of rs893403 21
Figure S13. Effect of rs893403 on GCC2 mRNA gene expression across GTEx tissues 22
Figure S14. LIMS1 Protein Expression in Multiple Human Tissues by Immunohistochemistry 23
Figure S15. Human kidney tissue immunohistochemistry for LIMS1 24
Figure S16. Human kidney tissue mRNA expression of *LIMS1* by snRNA-seq 25
Figure S17. Confirmation of *LIMS1* mRNA expression in kidney tubules by RNAScope® 26
Figure S18. Hypoxia induces LIMS1 expression in human kidney cells 27
Figure S19. Human kidney tissue immunohistochemistry for GCC2 28
Figure S20. Protoarray intensity for LIMS1 and LIMS2 by genotype-phenotype group 29
Figure S21. ROC curves for anti-LIMS1 IgG, IgG2, IgG3, and IgG4 30
Figure S22. LIMS1 protein detection in HEK-293 and HRCE cells 31
Figure S23. Anti-LIMS1 Ab induces cytotoxicity in cultured human kidney epithelial cells 32

SUPPLEMENTAL TABLES:

Table S1. The study power 33
Table S2. Baseline clinical characteristics of the Columbia Discovery Cohort 34
Table S3. The deletion-tagging SNPs typed in the discovery cohort 35
Table S4. Baseline clinical characteristics of the Replication Cohorts 36
Table S5. Recipients-only CPH analysis of the risk genotype at rs893403 37
Table S6. Comparison of alternative genetic models 38
Table S7. Comparison of alternative statistical models 39
Table S8. Comparison of MAF for rs893403-G risk allele between 1000G populations 40
Table S9. ANNOVAR annotation of known variants in LD with rs893403 41
Table S10. IHC staining patterns for LIMS1 and GCC2 in human kidney tissue compartments 42

SUPPLEMENTAL REFERENCES:

Supplemental References 43

SUPPLEMENTAL RESULTS:

Genomic Annotation of the 2q12.3 Locus: The top SNP, rs893403, was initially selected as it tags a common 1.5-kb deletion CNVR915.1 downstream of the *LIMS1* gene (r^2=0.98 in the CEU population). CNVR915.1 was originally annotated to intersect *LOC100288532,* a pseudogene in hg18 that was removed in subsequent releases of the human genome. Using pooled sequencing read data for the European participants of the 1000 Genomes Project, we first re-mapped the precise breakpoints of this deletion (**Figure S7**). We next interrogated the deleted sequence for presence of microRNAs, pseudogenes, or regulatory elements. We confirmed the presence of several repeated elements as well as a single conserved retroposed gene (*retro-COX7B*), but no other coding elements. Additionally, interrogation of the ENCODE and Roadmap datasets revealed no obvious regulatory elements within the region. Because neither one of these datasets includes adult kidney cells, we also performed histone tail modification analysis of human kidney proximal tubule cell line that suggested the possibility of a weak tubule-specific enhancer spanning across this region, but no other functional segments (**Figure S8**). Using our new tissue-specific FUN-LDA scoring method based on 127 tissues and cell types[1], we performed analyses of the deleted sequence and all known variants in LD (r2>0.8) with rs893403. We did not detect any potentially functional variants within the deletion region by FUN-LDA. However, there were several potentially functional variants outside of the deletion region (**Figure S9, Table S9**), including rs10084199, which resides within the *LIMS1* transcription start site and has FUN-LDA posterior probability 1.0 across all 127 tissues.

Analysis of eQTL Effects: We next tested for expression QTL effects of rs893403. Although kidney tissue eQTL data is not available in Genotype-Tissue Expression (GTEx) project, data supported strong effect of rs893403 on the mRNA levels of *LIMS1* and nearby *GCC2* gene across multiple tissues (**Figures S10-S13**). In each case, the risk allele was associated with lower mRNA levels for both genes. To test for a similar effect in kidney tissue, we used transcriptomic data from laser-microdissected human kidney tissue compartments of N=166 NEPTUNE study participants[2]. We detected a direction-consistent cis-eQTL effect with the deletion-tagging allele associated with reduced *LIMS1* mRNA levels in the tubulointerstitium (Beta -0.28, P=0.014, **Figure S10c)**. Notably, LIMS1 transcript represented the most significant association among all genes within 1-Mb window of the rs893403 in the tubulointerstitial compartment, including *GCC2* (Beta -0.09, P=0.19, **Figure S12c**). Compared to rs893403, the TSS variant prioritized by our FUN-LDA analysis exhibited a stronger effect on *LIMS1* expression in the tubulointerstitium (rs10084199, Beta -0.29, P=0.012). In summary, our annotations suggest that rs893403 or another variant in LD, such as rs10084199, is associated with gene expression of *LIMS1* in human kidney tubules and other organs.

Induction of *in vitro* Cytotoxicity in Human Kidney Cells by Anti-LIMS1 Antibodies: By immunofluorescence staining, we demonstrated that LIMS1 protein is present in both HEK-293 cell line and primary human renal cortical epithelial (HRCE) cells (**Figure S22**). Overnight culture of the HRCE cells with the mouse anti-human LIMS1 antibodies disrupted the normal arrangement of actin filaments in the epithelial cells compared to the control antibody (**Figure S23a**) and was cytotoxic to kidney epithelial cells compared to the control antibody as determined by measuring LDH in the supernatant (8.2 $\pm$ 3.1% compared to the control antibody 1.3 $\pm$ 0.3%, $p < 0.01$). Similar cytotoxic effect was also observed in HEK-293 cells (8.7 $\pm$ 2.6% compared to the control antibody 3.2 $\pm$ 1.2%, $p < 0.01$) (**Figure S23b**).

Cell Surface Detection of LIMS1 Protein *in vitro* under Hypoxic Conditions: Previous report demonstrated that *LIMS1* mRNA expression was upregulated by hypoxia in arterial endothelial cells[3]. We hypothesized that under similar conditions, we would be able to detect the LIMS1 protein on the cell surface in cultured kidney cells. Following culture of the cells in hypoxic conditions, there

was a significant increase in Mean Florescent Intensity (MFI) of LIMS1 measured by Flow Cytometry on the surface of the HEK-293 cells (isotype control MFI 585, normal culture conditions MFI 695 $\pm$ 52, hypoxic conditions MFI 978 $\pm$ 17, P<0.01, **Figure S18**) compared to the control conditions. This demonstrates LIMS1 protein was detected on the cell membranes under hypoxic conditions. Under normal conditions, LIMS1 appears to be absent on the cell surface as there was not a significant increase in the specific LIMS1 MFI compared to the isotype control. A similar trend was observed for HRCE cells, although not statistically significant.

SUPPLEMENTAL METHODS:

Study Design: The study was designed in two stages. Stage 1 (the discovery phase) involved a genome-wide screen of 50 high priority common copy number polymorphisms (CNPs) in 705 kidney allograft recipients transplanted at Columbia University Irving Medical Center (CUIMC); signals reaching nominal P-value < 0.05 were advanced to the replication phase. We carried out power calculations for the discovery cohort under the assumption of a recessive model, nominal replication threshold alpha = 0.05, and perfect LD between a tag-SNP and a deletion allele (**Table S1**). This calculation demonstrates limited power at MAFs below 10%, motivating our MAF > 10% criterion for the selection of candidate deletions. Stage 2 (the replication phase) involved genotyping of the top signals from Stage 1 in additional cohorts, totalling N=2,004 full DR pairs.

Clinical Outcomes and Statistical Methods: The primary outcome was time-to-first-rejection, defined from the date of first transplant to the date of first biopsy demonstrating a rejection event, including both antibody mediated rejections (ABMR) and T-cell mediated rejections (TCMR). We used Kaplan-Meier survival analysis (for univariate analysis) and Cox proportional hazards model (for multivariate analysis) to model these outcomes. Multivariate modeling of clinical covariates was done using variables that were nominally associated with outcome on univariate analysis ($P < 0.05$). Such covariates were then subjected to stepwise selection using a BIC-guided selection. In the recipient-only analysis, the multivariate model was used to test recipient's genotype coded under a recessive model, with risk genotype defined by homozygosity for the deletion-tagging allele. In the full DR pair analysis, the risk genotype ("collision genotype") was defined by recipient homozygosity in the absence of donor homozygosity for the deletion-tagging allele. Under this coding, any DR pair with a donor genotype homozygous for the deletion-tagging allele was coded as non-risk regardless of the paired recipient genotype, including when recipient genotype was not available. Similar to recipient-only analyses, the "collision genotype" was used as one of the predictors in the multivariate Cox proportional hazards model to derive adjusted effect estimates and P-values. Statistical analyses were performed using R base and survival packages (R v3.4, CRAN).

Stage 1 (Discovery) Methods: The clinical characteristics of the CUIMC discovery cohort are summarized in **Table S2**. The association screen was performed using a tag-SNP approach based on the filtering strategy depicted in **Figure 1.** In total, we identified 50 common deletions perfectly tagged by a SNP at $r^2>0.8$. These 50 high priority candidate SNPs were genotyped in a cohort of 705 kidney transplant recipients recruited by the Columbia Chronic Kidney Disease (CKD) Bio-bank; genomic DNA was extracted from whole blood (QuickGene-610L, Kurabo) and individuals SNPs were typed using KASP (Kompetitive Alelle Specific PCR) assay by LGC Genomics. Strict genotype quality control (QC) analysis was performed, including per-SNP and per-individual genotyping rates >95% and elimination of SNPs deviating from Hardy-Weinberg equilibrium within each ethnic group. In total, 44 SNPs passed all QC filters (**Figure 1**). To test for effects of deletion homozygosity, we used a time-to-event survival analysis under a recessive model with and without

adjustment for relevant clinical covariates. We applied a Bonferroni-corrected alpha to declare statistical significance.

Stage 2 (Replication) Methods: The replication cohorts included the Belfast cohort (N=387 full DR pairs), the TransplantLines Genetics cohort[4] (N=833 full DR pairs, trial registration number NCT03272841) and the Torino cohort (N=784 full DR pairs), providing a total of 2,004 full donor-recipient pairs for analysis. The clinical characteristics of the replication cohorts are presented in **Table S4**. Targeted genotyping was performed using KASP assay (TransplantLines cohort), Sequenom iPLEX MassARRAY® (Belfast cohort) and by direct Sanger sequencing (Torino cohort). The QC assessment included genotyping rate >95% in the entire cohort and passing Hardy-Weinberg equilibrium test (P>0.05) in both donor and recipient groups separately. In addition, we compared genotype frequencies of rs893403 between donors and recipients within and across all replication cohorts. As summarized in **Table S4**, genotype frequencies were nearly identical in donors and recipients within each European cohort, and comparable across cohorts, assuring against potential genotyping errors or bias. The statistical models applied to the replication cohorts were the same as in the discovery analysis. The combined statistical analyses across replication cohorts were stratified by cohort membership. Because in our replication cohorts the majority of rejection events occurred within the first year post-transplant, we also compared alternative statistical models that do not rely on the assumption of proportional hazards. This included non-parametric (log-rank) and logistic-regression-based tests, as summarized in **Table S7.** These additional analyses demonstrate that our model choices have no effect on the overall conclusions of our tests, and our results remain statistically significant regardless of the specific model assumptions.

Deletion Breakpoint Mapping and Sequence Motif Analysis: We fine-mapped the CNVR915.1 deletion breakpoints using whole genome sequence data from 503 Europeans sequenced through the 1000 Genomes Project[5]. Briefly, based on the genotype of rs893403, we classified each individual into AA (N=176), AG (N=245), or GG (N=82) groups. We randomly selected 50 individuals for each group and merged all individual BAM files into a single BAM file for each group to visualize the coverage using Integrative Genomics Viewer (IGV)[6]. Using BEDtools (coverage command)[7], we extracted the region and defined the breakpoints by analyzing depth of coverage in the rs893403-GG group. **Figure S7** depicts sequence alignments by rs893403 genotype and the presence of 1.5-kb deletion in the region. The deletion was further confirmed by 3-primer PCR analysis of individuals with AG and GG genotype, and the precise boundaries were mapped to chr2: 109310555-109312110 (hg19, 1556 base pairs deleted). The deleted sequence was next interrogated for various regulatory motifs and features using UCSC Genome Browser[8] and a variety of other data types, including Human mRNAs[9] , Human ESTs[10], predicted Retroposed Genes (UCSC Genes V5), C/D and H/ACA Box snoRNAs, scaRNAs, and microRNAs from snoRNABase[11] and miRBase[12], tRNA Genes predicted by using tRNAscan-SE v.1.23[13]; TargetScan miRNA Regulatory Sites[14]; repeat sequences by RepeatMasker (http://www.repeatmasker.org); conservation metrics such as GERP[15]; and Transcription Factor Binding Sites computed with the Transfac Matrix Database (v7.0) from Biobase (http://gene-regulation.com/pub/databases.html).

PCR-based Deletion Confirmation: The CNVR915.1 deletion was confirmed by genomic DNA quantitative PCR in participant samples with rs893043-AG and GG genotypes and primer walking PCR method was adapted to confirm deletion boundaries based on the presence of a PCR product between close primer pairs. The below primer pair detected a single band PCR product and Sanger sequencing of this amplicon using the same primer pair identified the deletion breakpoints:

- CNV915.1-F: 5'-AAAGACCTCAAATCAATAGCCTG-3'
- CNVR915.1-R: 5'- GGACATTTAGGCTGCTTCTG-3'

The precise deletion boundaries were mapped to chr2:109,310,555-109,312,110 (hg19, 1,556 base pairs deleted). Next, a simple 3-primer PCR was designed to enable rapid deletion typing:

- CNP-F1: 5'-TGTTTGTTGTTAAGGTCTCTATTG-3'
- CNP-F2: 5'-ATACAGATGTGCAAATACCTCTTACAG-3'
- CNP-R: 5'-AAATGACAGTGGTAATCCTTACCTATC-3'

Functional Annotation: We used a broad range of methods to perform functional annotations of the *LIMS1* locus. First, we annotated rs893403, all variants in strong linkage disequilibrium with rs893403 (defined by r^2>0.8 based on Europeans in 1000 Genomes phase 3), and the CNVR915.1 deletion region using tissue-specific annotations available through the ENCODE and Roadmap Epigenomes projects[16-18], including conservation, chromatin state segmentation, DNAse hypersensitivity sites and protein-DNA binding sites; we used HaploReg2 to manually query individual variants[19]. Because neither ENCODE nor Roadmap includes adult kidney cells, we performed histone tail modification analysis of human kidney proximal tubule cell line (GSE49637) followed by ChromHMM segmentation[17] to define tubule-specific regulatory elements (**Figure S8**). Second, we used our recently proposed unsupervised tissue-specific functional scoring method for non-coding variants[1]. This method, called FUN-LDA, is based on the Latent Dirichlet Allocation (LDA) model, a generative probabilistic model widely used in the topic modeling literature that allows joint modeling of data from multiple tissues. Using FUN-LDA, we computed the posterior probability for each variant in the region to be functional separately for 127 different tissues and cell types. Based on FUN-LDA analysis of all Roadmap and ENCODE data, we detected no predicted functional variants within the CNVR915.1 deletion region. In contrast, in the analysis of all variants in LD with rs893403 (**Figure S9)** we detected several variants with high probability of being functional, including rs10084199 with high scores across nearly all 127 tissues and types (avg. posterior FUN-LDA probability 1.0).

Testing for eQTL Effects of rs893403: Using Genotype-Tissue Expression (GTEx) data, we tested for significant effect of rs893403 on mRNA levels of all genes located within 1Mb window of the index SNP. We detected a strong and significant cis-eQTL effects of rs893403 on two genes, *LIMS1* and *GCC2* across many tissues (**Figures S10-13**). Because GTEx does not include kidney tissue, we next tested rs893403 against transcriptomic data from manually micro-dissected human glomerular and tubulointerstitial compartments in 166 participants of the Nephrotic Syndrome Study Network (NEPTUNE) study[2,20,21]. Briefly, transcript quantification was performed with Affymetrix 2.1 ST arrays; after normalization of expression levels across genes using robust multi-array average (RMA)[22] and derivation of PEER factors as previously described[23,24], rs893403 was tested for eQTL effects with transcripts within a 1-Mb window using linear regression under additive genotype coding with adjustments for age, sex, PEER factors and the first 4 principal components of ancestry.

Detection of Anti-LIMS1 Alloantibodies by Protein Arrays: Sera were screened against 9,114 human proteins displayed on the Human Protein Microarray (ProtoArray®, Invitrogen) according to manufacturer's instructions. Serum was diluted at 1:250. The reactivity of the serum to proteins on the ProtoArray were detected using Alexa Fluor 647 goat anti-human IgG antibody, and Alexa Fluor® 647-labeled anti-V5 antibody was used to detect the control protein gradients that are printed on each of the sub-arrays. The arrays were read on a Molecular Devices GenePix 4000B scanner and analyzed using Genepix Pro7 software and protein prospector. Standard pre-processing was applied to the microarrays to account for technical variability and protein prospector was used to normalize the samples using a linear model and calculate the M-scores. Z-scores were calculated as the number of standard deviations of the signal derived from the signal of the mean and a Z score >2.5 was considered positive, on the condition this was concurrent with both protein spots on the microarray.

Detection of Anti-LIMS1 Alloantibodies by ELISA: LIMS1 protein (ab116807) was diluted to 2 μg/11ml in carbonate bicarbonate buffer and coated on an Immulon H2 plate overnight at 4°C. Plates were washed x3 with 200 μl of washing buffer (1xPBS, 0.05% tween) and blocked for two hours with blocking buffer (1xPBS, 0.05% tween, 1% fish gelatin). Plates were washed x3 with 200 μl of washing buffer, serum was diluted 1:1000 in washing buffer, 100 μl was added per well and incubated for two hours at room temperature. Plates were washed x7 with 200 μl of washing buffer, before the addition of the detection antibodies diluted according to manufacturer's instructions [anti-Human IgG-HRP (ab97160) 1:20,000, anti-human IgG1-biotin (Sigma B6775) 1:5,000, anti-human IgG2-biotin (Sigma B3393) 1:20,000, anti-human IgG3-biotin (Sigma B3523) 1:20,000, and anti-human IgG4-biotin (Sigma B3648) 1:20,000]. Plates were washed x7 with 200 μl of washing buffer. For the development of the total IgG plates, TMB peroxidase substrate and Peroxidase substrate Sol B were mixed at a 1:1 ratio at room temperature and 100 μl was added to the plate, and the reaction was stopped by adding 100 μl of 2M H_2SO_4, left on the bench for 20 minutes before reading at 450 nm. For the IgG subclasses, we added anti-biotin-HRP antibody (ab19221) diluted 1:20000 in 1xPBS, 0.05% tween and incubated at room temperature for 1 hour, then the plates were washed x7 with 200 μl of washing buffer. To develop the plates TMB peroxidase substrate and Peroxidase substrate Sol B were mixed at a 1:1 ratio and 100 μl was added to the plate at room temperature, and the reaction was stopped by adding 100 μl of 2M H2SO4, left on the bench for 20 minutes before reading at 450 nm. We used two sets of controls; for the first control to ensure that the LIMS1 protein bound to the plate, we used control antibody, mouse anti-LIMS1 (LSBio LS-C169391) in a serial dilution, that was developed using anti-mouse IgG1-HRP labeled (ab97240). Our second control was serum taken from normal healthy controls which were non-reactive to LIMS1. These samples were used as normalization controls between the plates. The reactivity of individual serum samples was measured as a fold-change in OD compared to the average for the normalization controls (**Figure 3**).

Western Blots: The protein blots were performed using standard procedures - 0.25 ug of reduced LIMS1 protein was run on a 10% SDS-PAGE gel and then transferred to nitrocellulose. The membrane was blocked with fish gelatin, before probed with subject's plasma (1:1000 dilution). Membranes were washed 5 times in 1 x PBS 0.1% tween, before the addition of the secondary antibody (anti-human IgG, 1:10,000) incubated for 1 hour before washed 5 times in 1 x PBS 0.1% tween, and developed using Immobilon Western HRP substrate Luminol Reagent (Millipore).

Immunohistochemistry: Tissue antibody staining was performed with the use of mouse monoclonal (IgG1) to human LIMS1 (LSBio LS-C169391) as well as rabbit polyclonal IgG antibody to human GCC2 (Genetex GTX51372) on paraffin-embedded tissues with the use of heat-induced antigen retrieval (**Figures S14, S15, and S19**). The following human tissues were sectioned and examined: kidney, liver, heart, lung, pancreas, and skin.

RNAscope® in Situ Hybridization: RNA in situ hybridization was performed using the RNAscope® 2.5 HD Duplex (Chromogenic) Detection Kit for Human (Advanced Cell Diagnostics, Cat. No. 322435) on formalin-fixed paraffin-embedded (FFPE) nephrectomy tissue sections. The 5μM sections were cut from tumor-free regions of the nephrectomy, which underwent 15 minutes of warm ischemia followed by 2 hours of cold ischemia from handling procedures. In situ hybridization was performed according to the manufacturer's protocol, RNAscope® 2.5 HD Duplex Kit User Manual for FFPE human samples. The following probes were used for both single- and dual-channel detection: Hs-LIMS1-C2 (Cat. No. 546401-C2), Hs-AQP2-C1 (Cat No. 434861), and Hs-SLC26A4-C1 (Cat No. 423311). Bright field images were captured using the Olympus IX73 Inverted Microscope under low (100x) and high (600x) magnifications.

Flow cytometry: Cells were stained using standard flow cytometry protocols. Briefly, for the staining of LIMS1, cells were resuspended in flow buffer (1 x PBS, 2 % FBS and blocked using Human TruStain FcX (BioLegend), before incubation with the anti-LIMS1 antibody (LSBio LS-C169391, dilution 1:200, manufacturer's recommendation), and incubated for 40 minutes at 4^{o}C. Cells were washed twice in flow buffer, and incubated with anti-mouse IgG1 (Alexa Fluor 488, BioLegend) for 30 minutes at 4^{o}C. Cells were washed twice with flow buffer and fixed using 1% formaldehyde in PBS. Cells were analyzed on an LSR II flow cytometer (BD Biosciences) and using FCS Express 6 Flow (De Novo Software).

Microscopic Evaluation of LIMS1 Protein in Human Kidney Cells: Human Renal Cortical Epithelial Cells (HRCE, Lonza, catalog no: CC-2554) and HEK-293 cells (ATCC® CRL-1573™) were grown on to sterile cover slips overnight and were washed x3 in PBS, before being fixed in 4% paraformaldehyde solution (methanol free) for 15 minutes at room temperature. Cells were washed x3 in PBS, followed by 20 min incubation in PBS, 0.1% Triton x100, and washed x3 in PBS. Cells were initially incubated with anti-LIMS1 antibody (1:200 dilution in PBS, 1% BSA, manufacturer's instructions) for 60 minutes, and washed x2 in PBS. Cells were incubated with anti-mouse IgG1 (Alexa Fluor 488) and DAPI for 45 minutes, before washed 4 times in PBS. Cells were attached to a slide using vector shield (Vector Labs) and stored at 4^{o}C. Images were taken using Nikon A1 confocal microscope at 600x, images were processed using ImageJ (NIH).

Microscopic Evaluation of the effect anti-LIMS1 antibodies on kidney cells. Human Renal Cortical Epithelial Cells and HEK-293 cells were grown on to sterile cover slips. Cells were treated with anti-LIMS1 antibody (3 μg/ml) or a control antibody (mouse IgG, 3 μg/ml) overnight. Both antibodies were filter washed using sterile PBS before use on the cells. Cells were washed x3 in PBS, and fixed in 4% paraformaldehyde solution (methanol free) for 15 minutes at room temperature. Cells were washed x3 in PBS, followed by 20 min incubation in PBS, 0.1% Triton x100, and washed x3 in PBS. Cells were stained with phalloidin (Alexa Fluor 594, Invitrogen) and DAPI (BioLegend) for 45 minutes at room temperature, before washed 4 times in PBS. Cells were attached to a slide using vector shield (Vector Labs) and stored at 4^{o}C. Images were taken on an Olympus IX73 microscope at 600x, images were processed using ImageJ (NIH).

Cytotoxicity Assay: Cell cytotoxicity was measured using the LDH-Cytotoxicity Assay (Abcam) following manufacturer's instructions. Briefly, HEK-293 cells and HRCE cells and were treated with anti-LIMS1 antibody (3 μg/ml) or a control antibody (mouse IgG, 3 μg/ml) overnight. Both antibodies were filter washed using sterile PBS before use on the cells. Cell supernatants were removed from the culture plate, centrifuges at 250g to remove cellular debris and used in the LDH-Cytotoxicity Assay. The OD was measured using a Bio-Tek Powerwave XS reader, absorbance of samples was measured at 490 nm and the reference wave length was 630 nm.

Contributions: KK and AG conceived and designed the study, wrote the manuscript and decided to publish the findings; NJS designed and performed the immunologic and cell culture experiments; VDA performed immunohistochemistry studies and pathology scoring. The primary data was generated by YL, ZD, JAD, CF, NJS, YC, KX, and JB. The genetic, statistical, and bioinformatic analyses were performed by KK, IIL, ZD, JAD, CF, LL, YJN, YDN, JYZ, and RS. Functional annotations were performed by KK, IIL, Y-AK, CEG, KS, and MGS. The Columbia cohort was recruited and characterized by OB, DB, NO, FO, SG, KM, JW, FZJ, MR, E-RMV, GV, SM, JR, DJC, LR and KK. The Belfast cohort was recruited and characterized by AEC, APM and AJMK. The TransplantLines Cohort was recruited and characterized by SJLB, HaS and MHB. The Torino Cohort was recruited and characterized by ZD, SD, FS and AA. All authors have read and approved the manuscript before submission; NJS, AG, and KK jointly vouch for the presented results.

Brief Glossary:

- **Single Nucleotide Variant (SNV):** genetic variation at a single base pair position in the genome; usually involves a substitution of one base pair for another.
- **Single Nucleotide Polymorphism (SNP):** Refers to SNV that is common in a given population, conventionally defined by a population frequency greater than 1%.
- **Copy Number Variant (CNV):** is a phenomenon in which a segment of the genome is either missing (deletion) or is repeated (duplication); the number of copies (segment repeats) can also vary between individuals.
- **Copy Number Polymorphism (CNP):** refers to a CNV that is common in a given population, conventionally defined by a population frequency greater than 1%.
- **Deletion Polymorphism:** refers to a deletion type of CNP; i.e. common deletion of a genomic segment with population frequency greater than 1%.
- **Expression Quantitative Trait Locus (eQTL):** a genetic variant that is associated with mRNA expression levels; cis-eQTLs (or local eQTLs) are genetic variants associated with transcript levels of nearby genes; trans-eQTLs (or distant eQTLs) are associated with transcript levels of distant genes (e.g. located on a different chromosome).

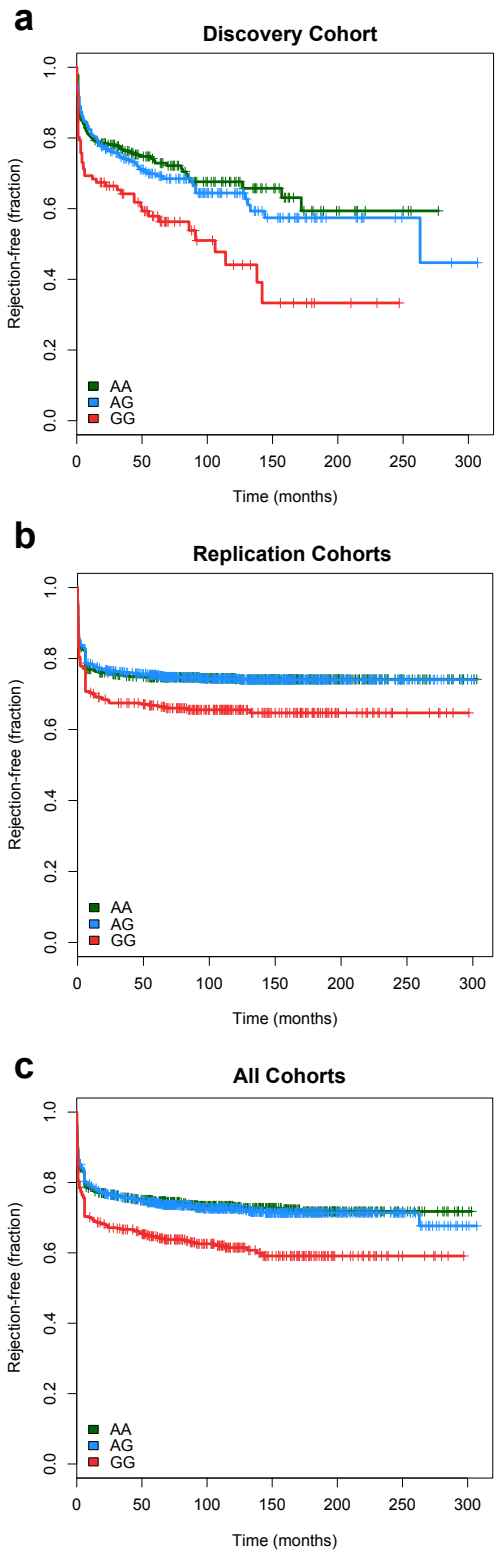


Figure S1. Rejection-free allograft survival by recipient genotype at rs893403: (a) The discovery phase (N=705 Columbia kidney transplants); **(b)** The replication phase: stratified analysis of 3 cohorts (Belfast, TransplantLines, and Torino; N=2,004 kidney transplants); **(c)** All cohorts: stratified analysis of 4 cohorts, N=2,709 kidney transplants.

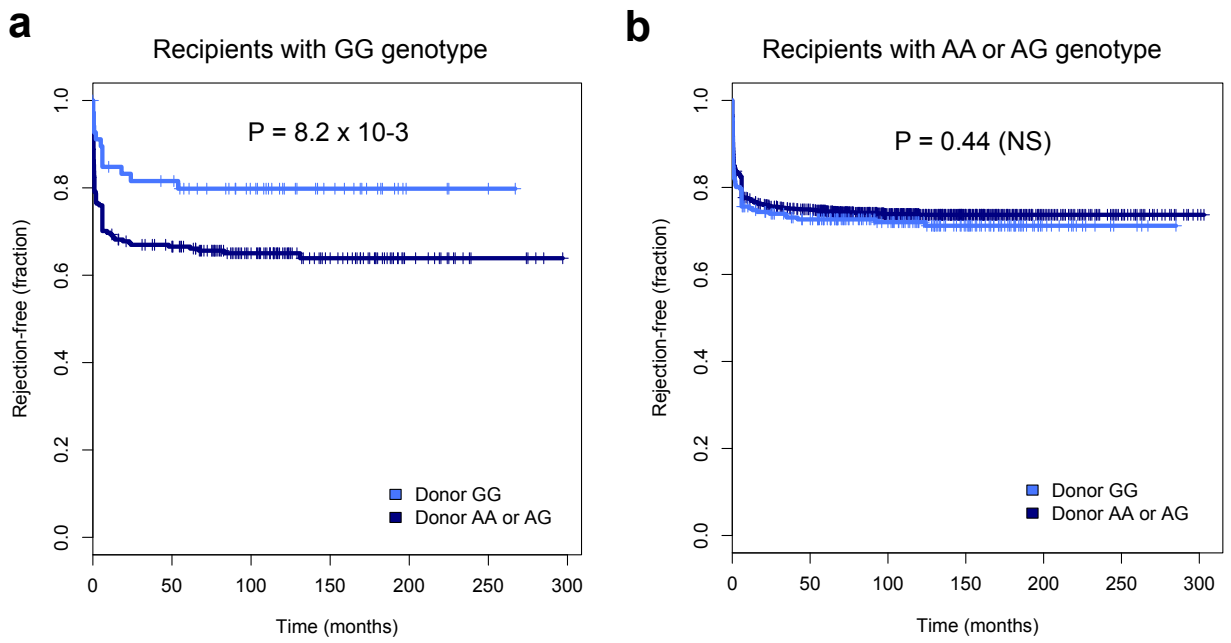


Figure S2. Rejection-free allograft survival by donor risk genotype stratified by recipient risk genotype in the replication cohorts: (a) in recipients with rs893403-GG genotype, donor's rs893403-GG genotype is associated with improved rejection-free allograft survival (Log Rank P = 8.2 x 10^{-3}); **(b)** in recipients with rs893403-AA or AG genotype, there are no significant differences in rejection-free allograft survival by donor's genotype (Log Rank P = 0.44). Only replication cohorts with available donor genotype data (N=2,004) are included in this analysis.

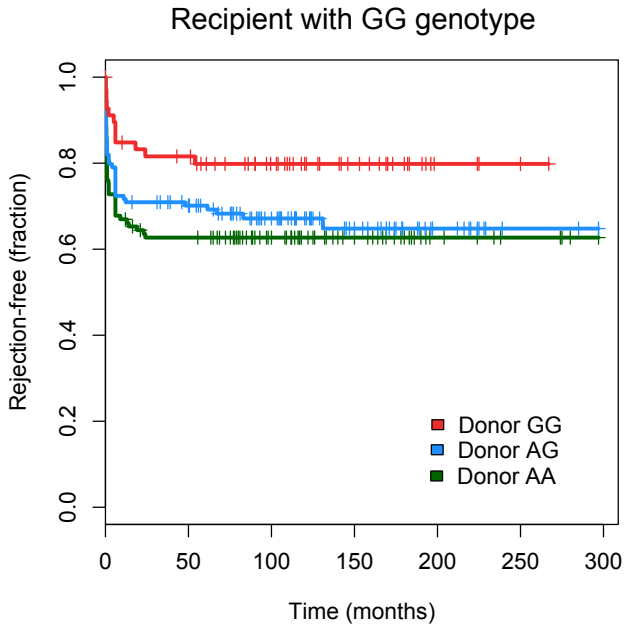


Figure S3. Rejection-free allograft survival in rs893403-GG recipients by all three donor genotype classes. Similar to Figure S2a, but demonstrating that in recipients with GG genotype, donor AA and AG genotypes convey similar risk, while donor GG genotype is protective. Only replication cohorts with available donor genotype data (N=2,004) are included in this analysis.

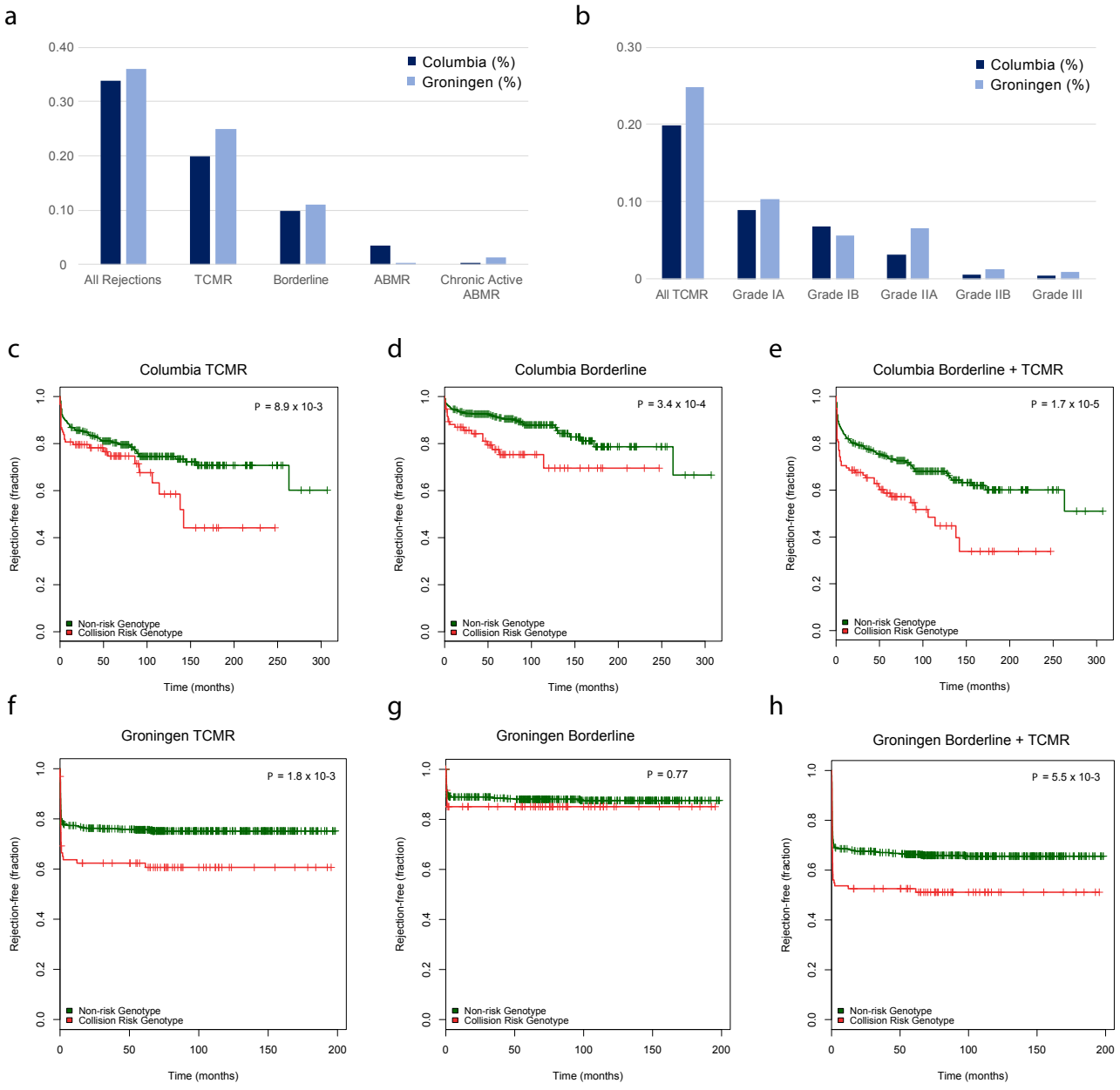


Figure S4. Rejection Events by Banff Type and Grade. Comparison of rejection frequency distributions between the Columbia (discovery) and the Groningen (replication) cohort by **(a)** rejection type and **(b)** Banff grade; TCMR represents the most common rejection type, while ABMR represents a rare event in both cohorts; Banff data were not available for other replication cohorts. The Columbia (discovery) cohort rejection free-survival restricted to the most common rejection types: **(c)**TCMR, **(d)** Borderline, and **(e)** TCMR or Borderline rejection events. The Groningen (replication) cohort rejection free-survival restricted to **(f)** TCMR, **(g)** Borderline, and **(h)** TCMR or Borderline rejection events; TCMR = T-cell Mediated Rejection; ABMR = Antibody Mediated Rejection. P-values correspond to a two-sided Log-Rank Test.

.

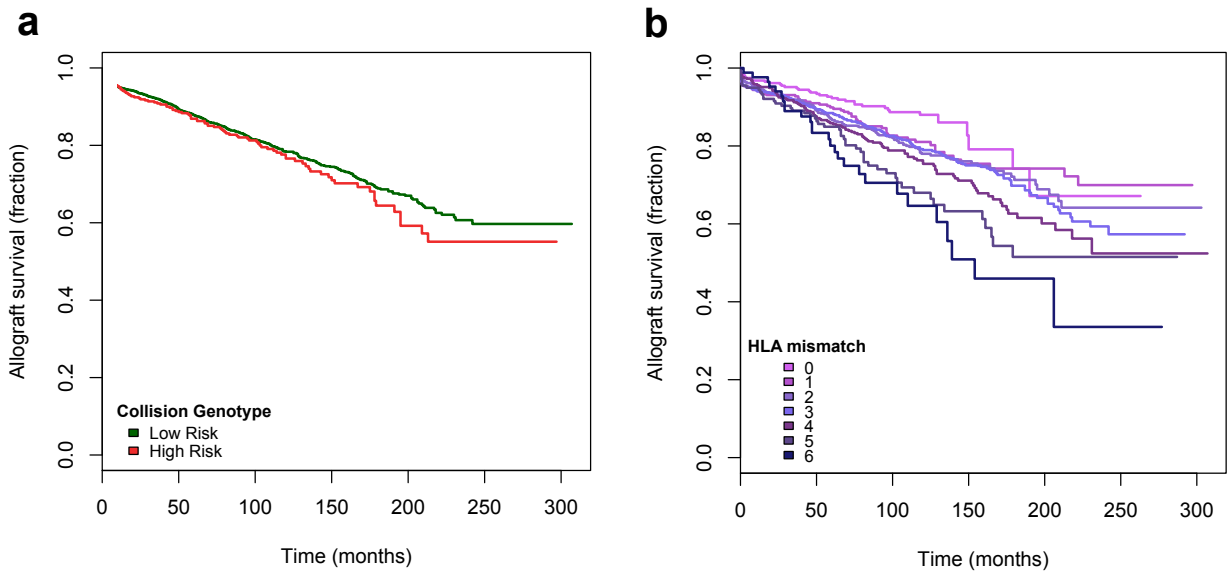


Figure S5. Kidney allograft survival analysis for all four cohorts combined (N=2709): (a) Allograft survival analysis demonstrates a trend for worse allograft survival in the high risk "collision genotype" group (adjusted HR=1.12, 95%CI: 0.90-1.39); the allograft survival model was adjusted for cohort, recipient age, donation type, and HLA mismatch; **(b)** Comparative analysis of HLA mismatch in the same dataset demonstrates per-allele adjusted HR=1.17 (95%CI: 1.09-1.25); the allograft survival model was adjusted for cohort, recipient age, donation type, and collision genotype.

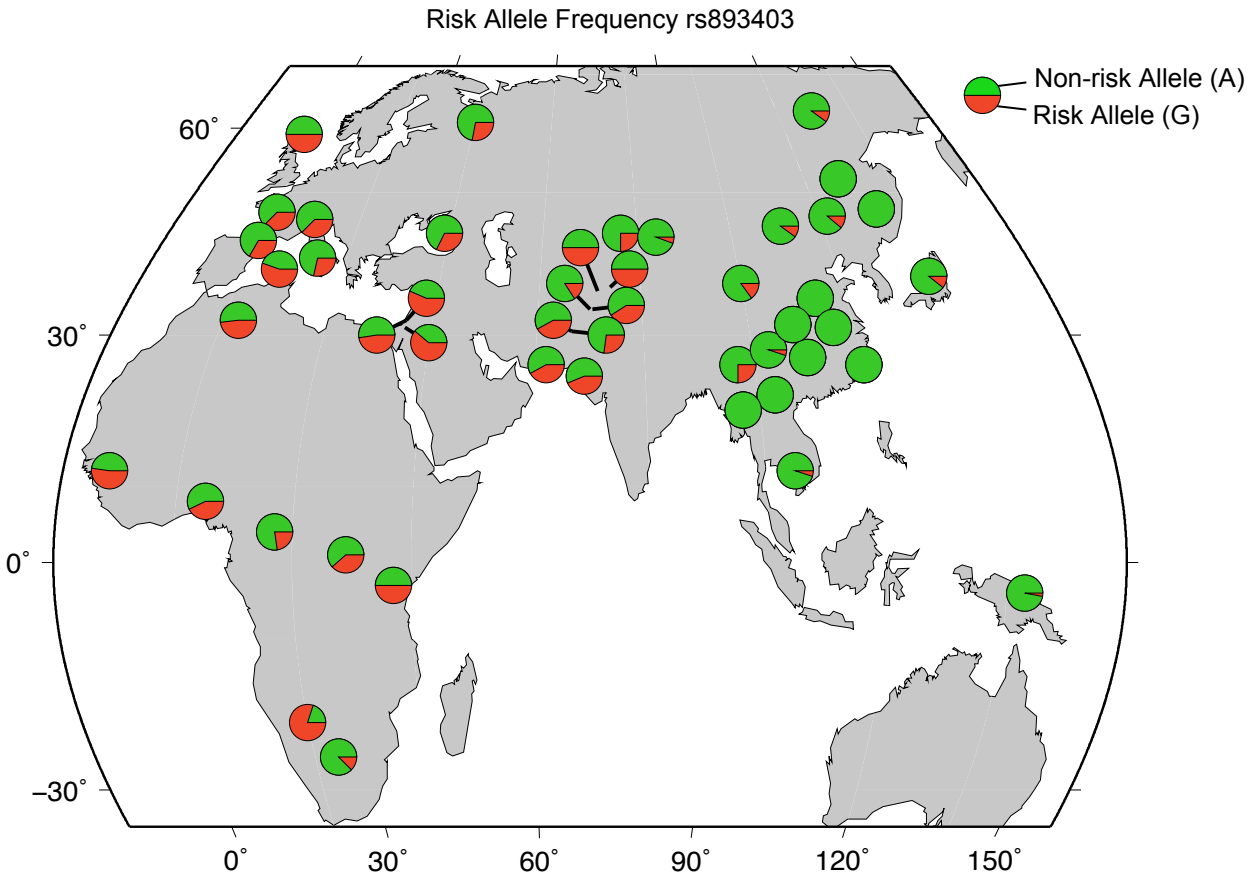


Figure S6. The distribution of rs893403 alleles across worldwide HGDP populations. The risk allele (red) is common in European and African populations, but has low frequency in East Asian populations.

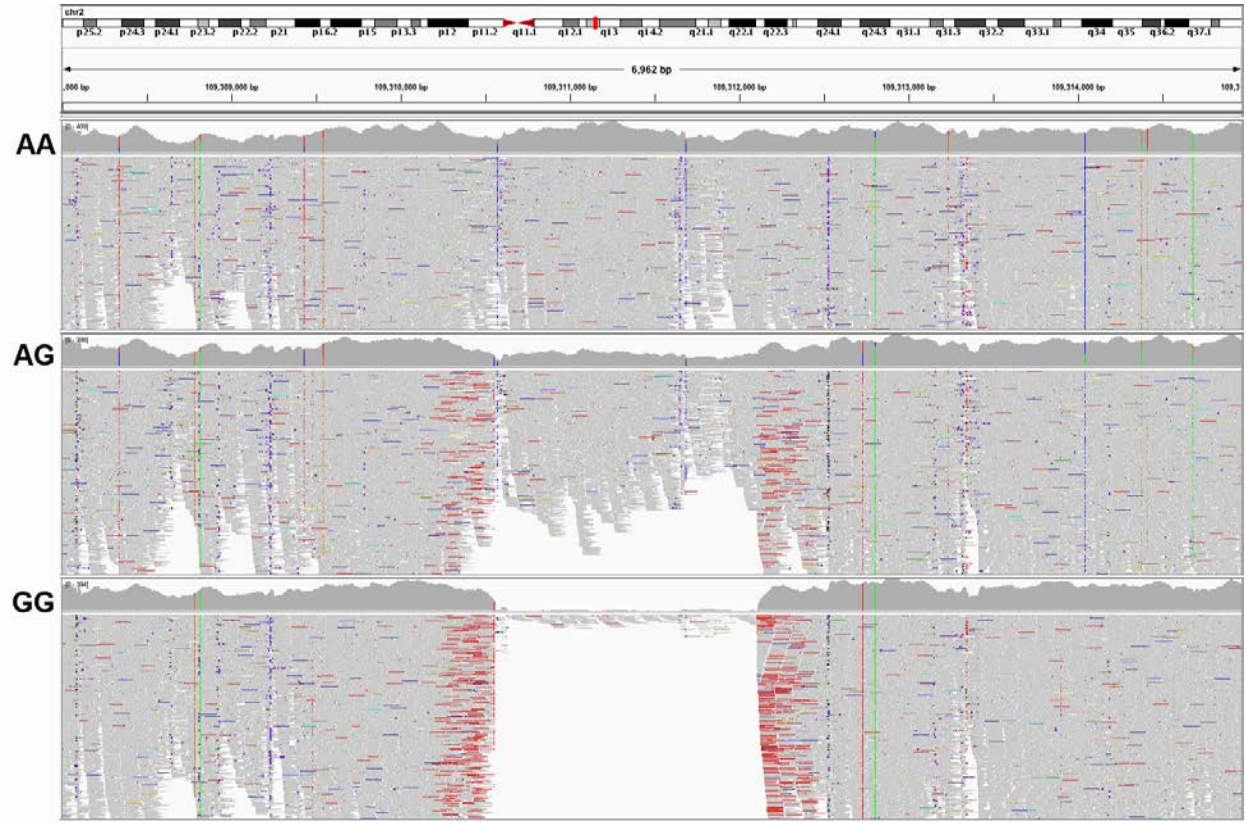


Figure S7. CNVR915.1 deletion breakpoints based on WGS data for 150 Europeans (N=50 per rs893403 genotype class). Read alignments were visualized in IGV based on pooled analysis of primary BAM files downloaded the 1000 Genomes phase 3 project website. The deletion was confirmed by PCR analysis of individuals with AG and GG genotype and mapped to chr2: 109,310,555-109,312,110 (hg19, 1,556 bp deleted)

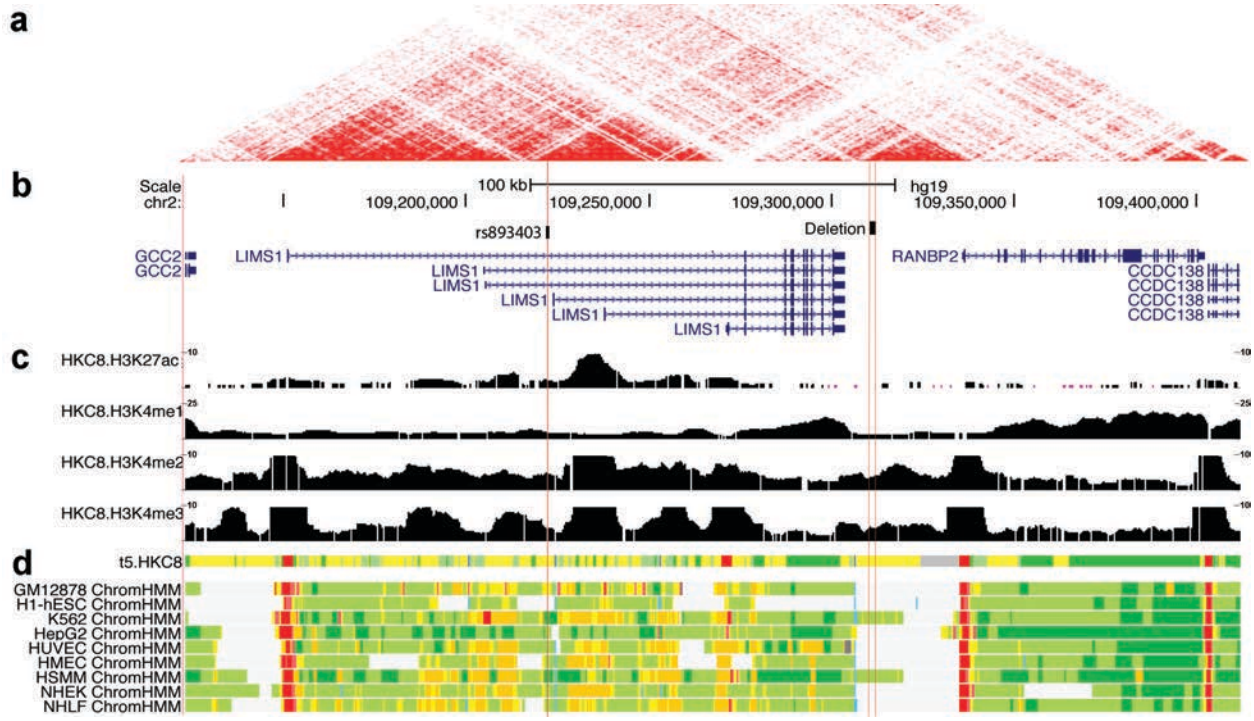


Figure S8. Genomic location of rs893403 and CNVR915.1 and their functional annotation: (a) Reference heatmap for Hi-C analysis of GM12878 from 3D Genome Browser; **(b)** Genomic coordinates and annotated transcripts in the region from the UCSC Genome Browser (hg19); **(c)** Histone modification marks for cultured human kidney tubule (HKC8) cells: H3K27ac, H3K4me1, H3K4me2, and H3K4me3; **(d)** ChromHMM tracks for HKC8 and ENCODE tissues; red = promoter, dark yellow = strong enhancer; yellow = weak enhancer; blue = insulator; green = transcribed regions; the weak enhancer intersected by the deletion appears specific to human kidney tubules.

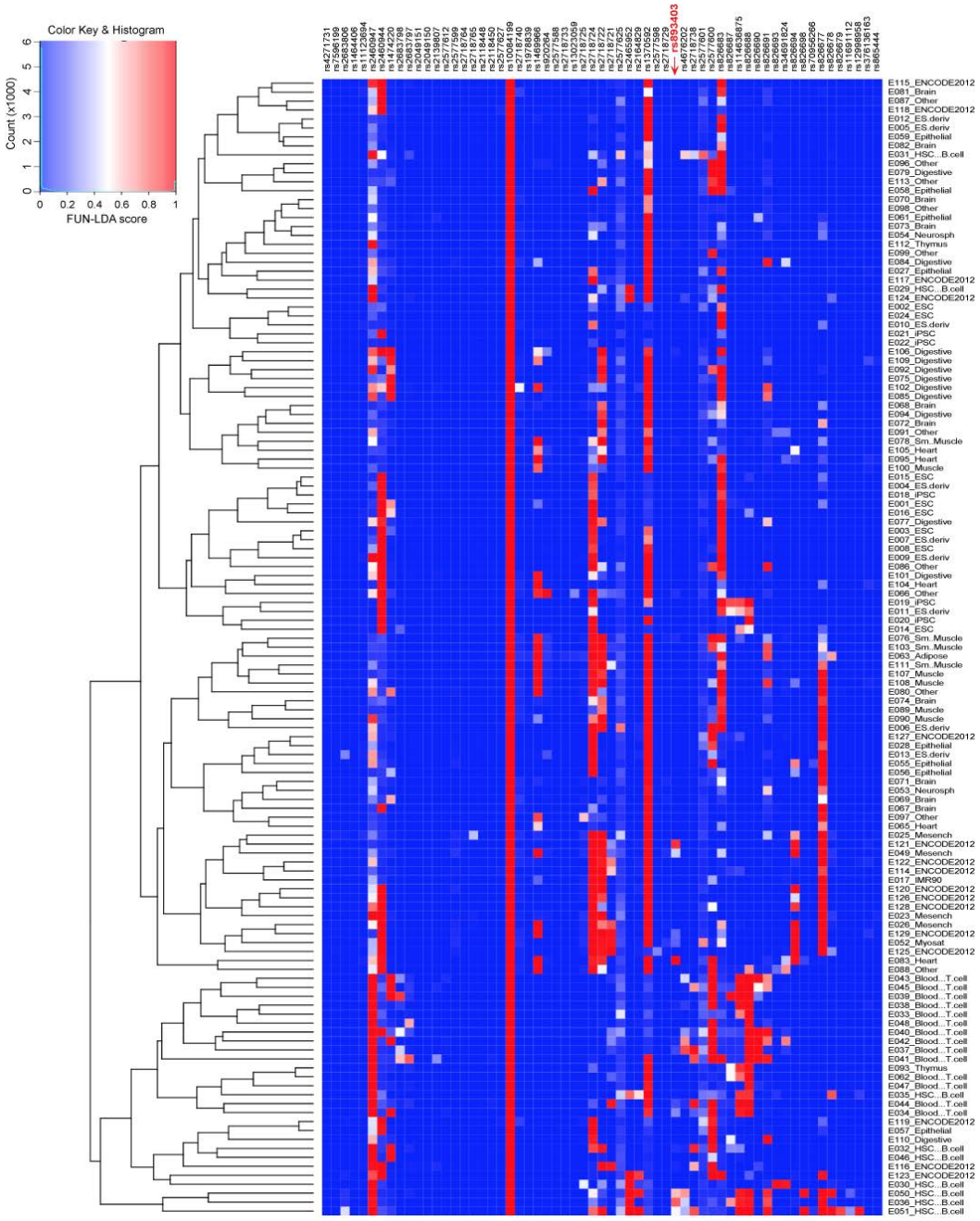


Figure S9. **FUN-LDA scoring of rs893403 and its proxies ordered by genomic location.** Red arrow indicates the index SNP (rs893403); the proxies were defined by r^2 > 0.8 (1000G phase 3 for Europeans); 127 Roadmap Epigenomics tissues were clustered based on the patterns of FUN-LDA scores. The top ranking SNP with predicted functional effect across all tissues was rs10084199 (average posterior probability 1.0) that intersects TSS of *LIMS1* gene. This SNP exhibits stronger and more significant eQTL effect on *LIMS1* mRNA expression (Beta=0.29, P=0.012) compared to rs893403 (Beta=0.28, P=0.014) in kidney tubulointerstitium of 166 NEPTUNE participants.

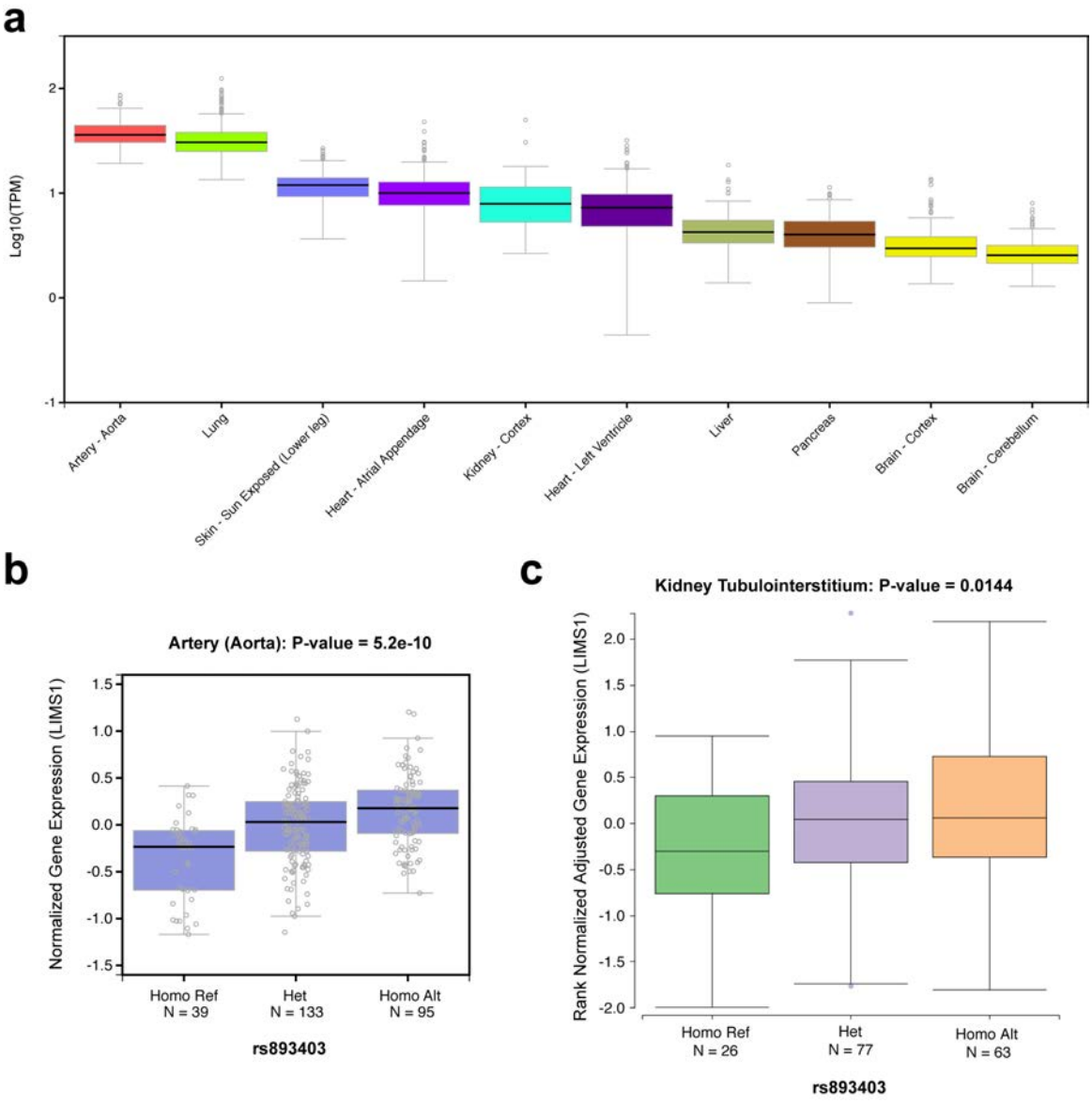


Figure S10. LIMS1 mRNA gene expression and cis-eQTL effect of rs893403: (a) *LIMS1* mRNA levels [log10(TPM)] in commonly transplanted tissues based on GTEx v.7 are consistent with protein staining by immunohistochemistry; aorta and brain are provided for reference as the GTEx tissues with highest and lowest *LIMS1* mRNA expression respectively; **(b)** The strongest cis-eQTL effect of rs893403 on *LIMS1* mRNA expression in GTEx is in the arterial (aorta) tissue (N=267, P=5.2x10^{-10}); the risk allele is associated with lower *LIMS1* mRNA levels; kidney tissue not available in GTEx; **(c)** Cis-eQTL effect of rs893403 in the NEPTUNE study confirms the association of risk genotype with lower *LIMS1* mRNA expression in kidney tubulointerstitium (N=166, P=0.014).

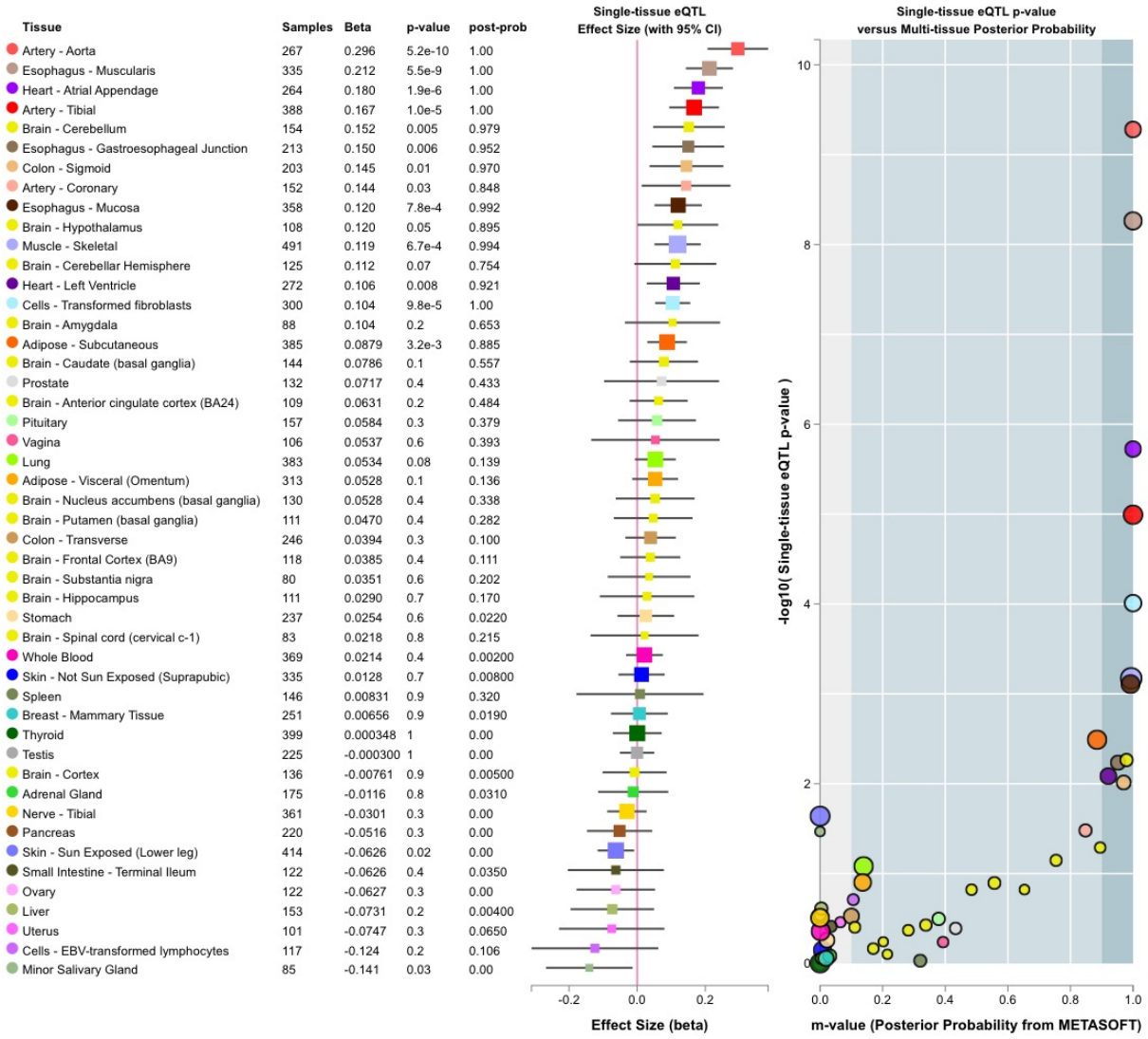


Figure S11. Effect of rs893403 on LIMS1 mRNA gene expression across GTEx tissues: the tissues are ordered based on cis-eQTL effect estimate from largest to smallest; kidney tissue is not included in GTEx (data from the latest GTEx release v.7).

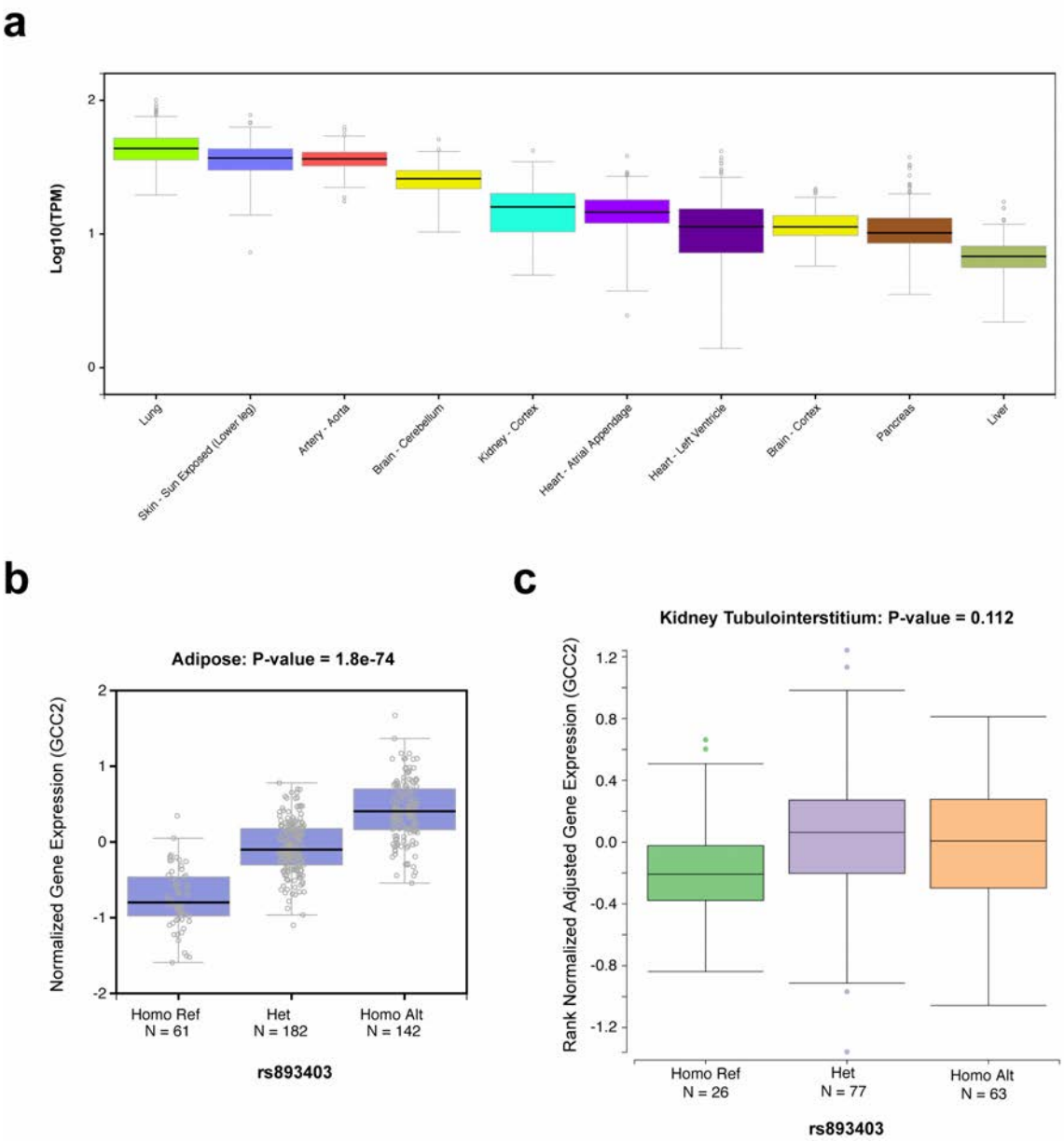


Figure S12. *GCC2* mRNA gene expression and cis-eQTL effect of rs893403: (a) *GCC2* mRNA levels [log10(TPM)] based on GTEx v.7 in the same tissues as in **Figure S4** for direct comparison with *LIMS1*; **(b)** The strongest cis-eQTL effect of rs893403 on *GCC2* mRNA expression in GTEx is in the adipose tissue (N=385, P=1.8x10^{-74}); similar to *LIMS1*, the risk allele is associated with lower *GCC2* expression; **(c)** Cis-eQTL effect of rs893403 on *GCC2* mRNA expression in the kidney tubulointerstitium is not statistically significant in the NEPTUNE study (N=166, P=0.11).

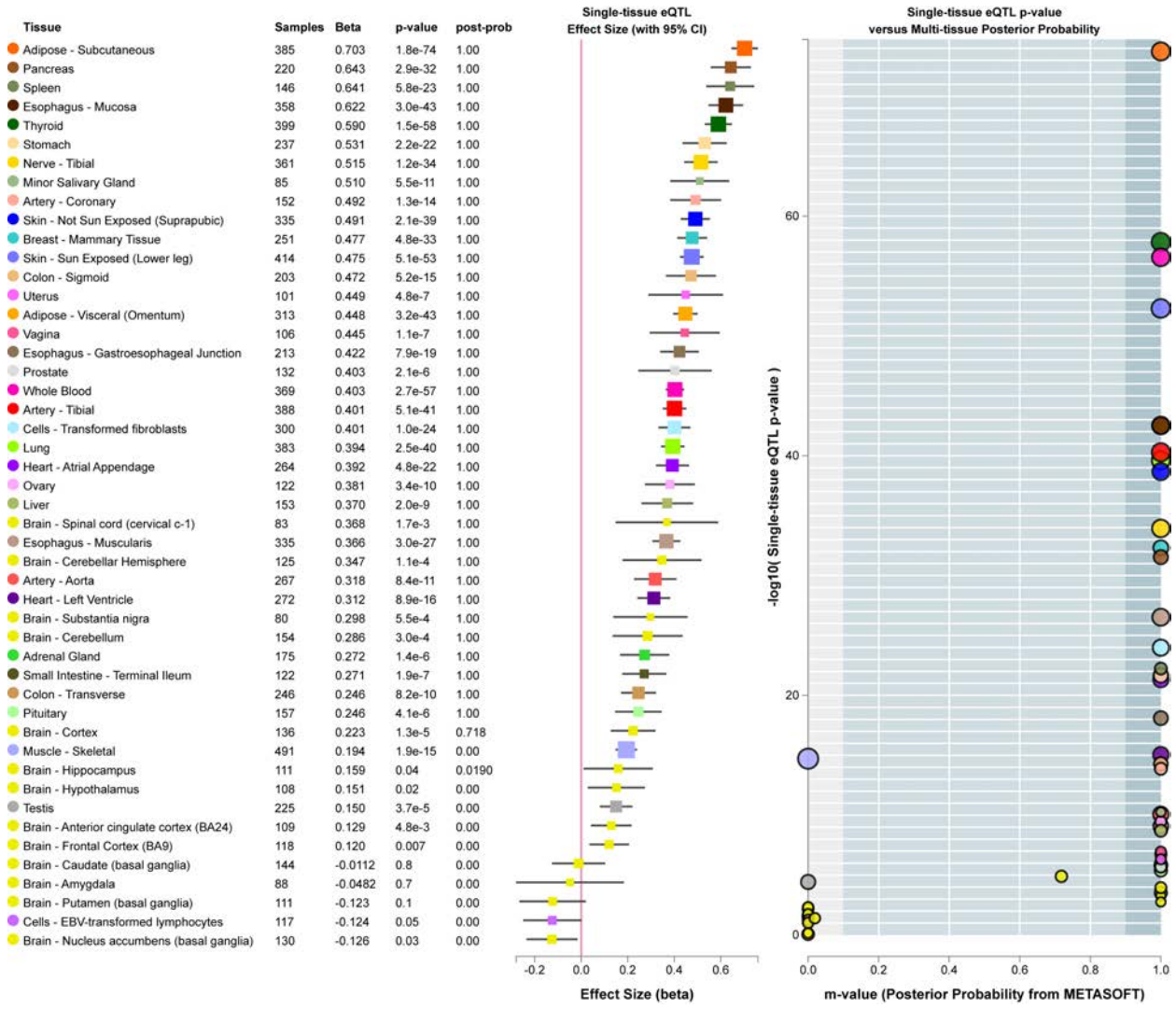


Figure S13. Effect of rs893403 on GCC2 mRNA gene expression across GTEx tissues: the tissues are ordered based on cis-eQTL effect estimate from largest to smallest; kidney tissue is not included in GTEx (data from the latest GTEx release v.7).

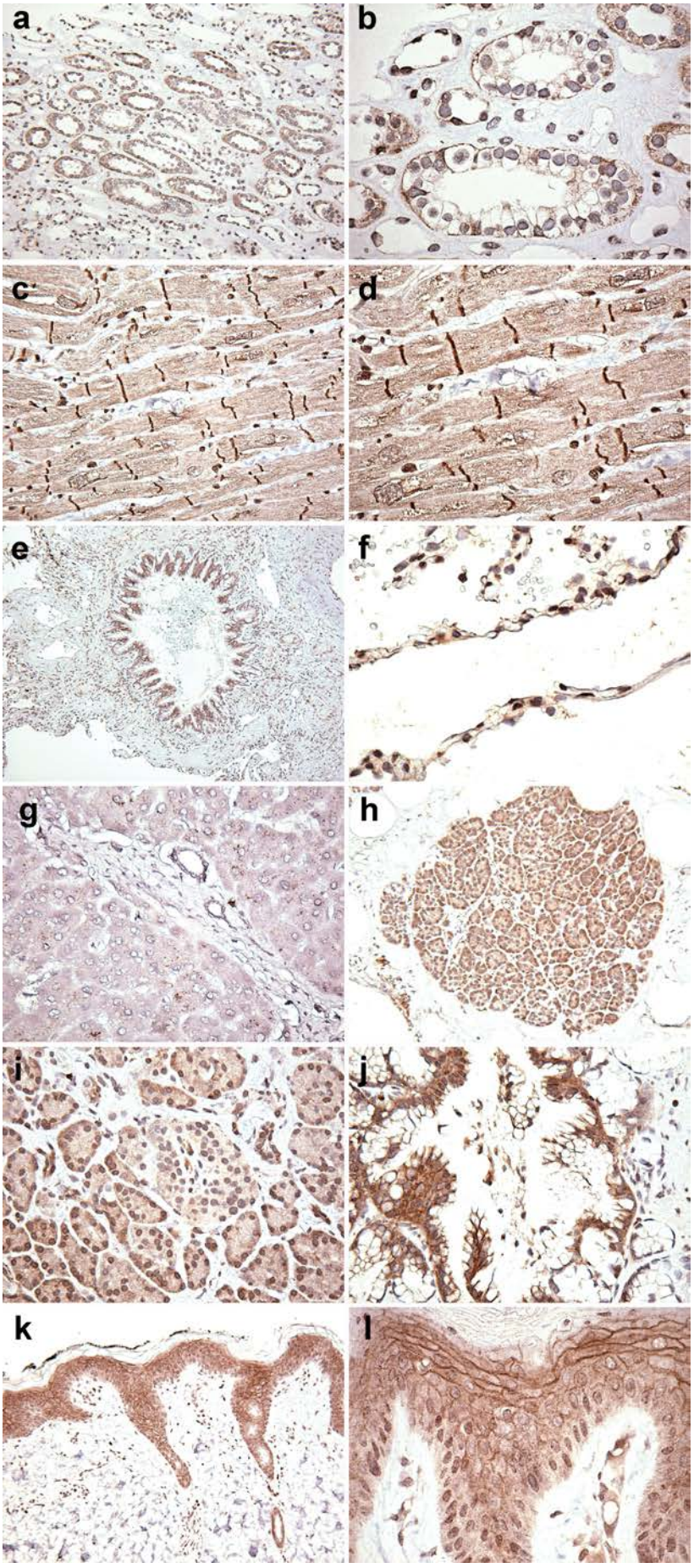


Figure S14. LIMS1 Protein Expression in Multiple Human Tissues by Immunohistochemistry: (a) low power view of kidney medulla with staining specific to medullary collecting tubules (200x); **(b)** higher power view of the kidney medulla with basolateral staining of collecting duct epithelial cells (400x); **(c)** medium power view of heart tissue (400x); **(d)** high power view of cardiac myocytes with staining of Z-bands (600x); **(e)** low power view of bronchus (100x) with staining of respiratory epithelium; **(f)** high power view of lung parenchyma with staining of pneumocytes (600x); **(g)** liver bile ducts section with low intensity staining of biliary epithelium (400x); **(h)** low power view of pancreas with staining of pancreatic acini (200x); **(i)** high power view of pancreatic acini and negatively stained pancreatic islet (400x); **(j)** pancreatic duct (400x) with positive staining of pancreatic duct epithelial cells; **(k)** low power view of human skin (200x) with strong epidermis staining; and **(l)** high power view of the epidermis demonstrating peripheral staining of keratinocytes and outlining desmosomes (600x).

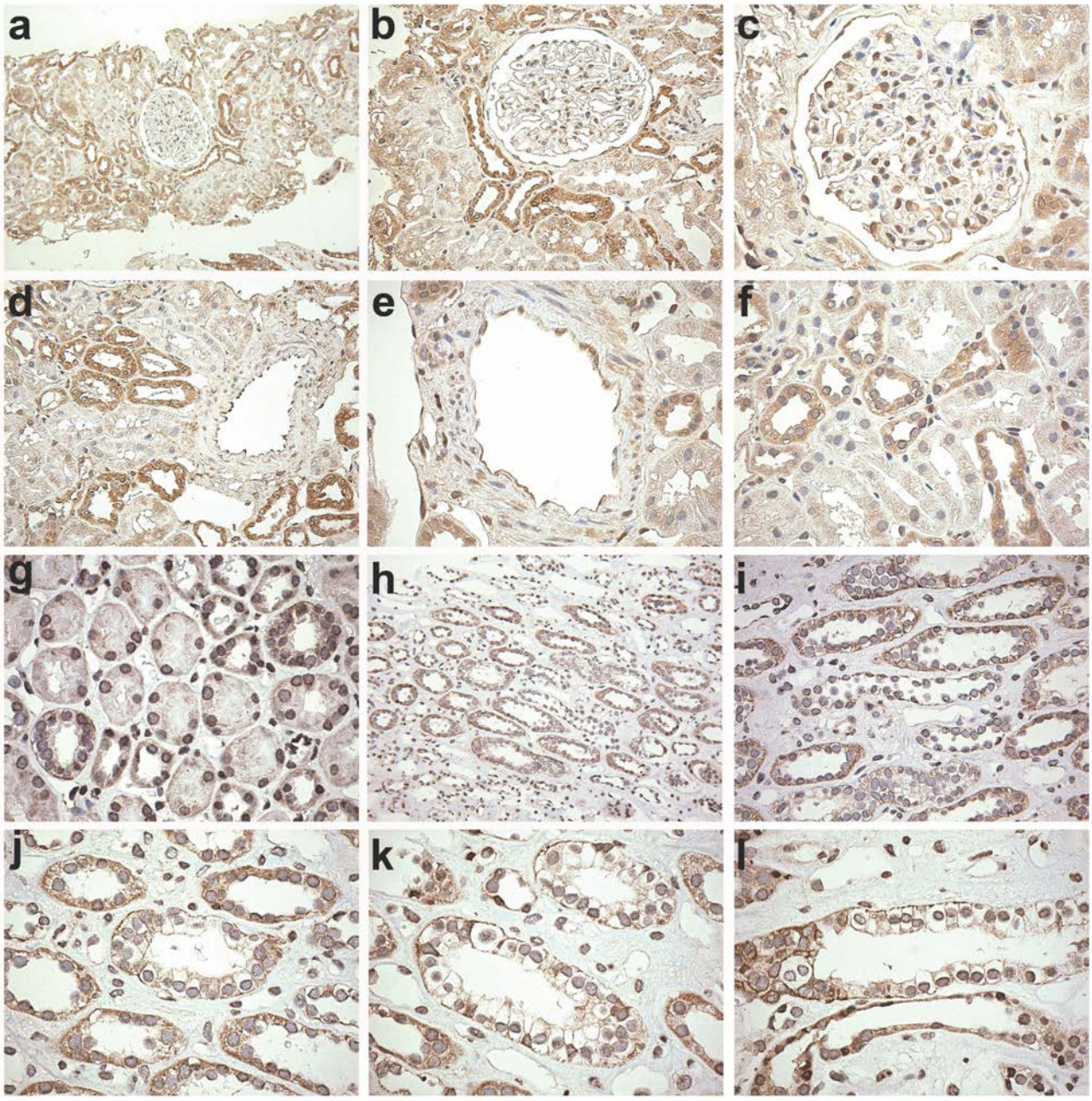


Figure S15. Human Kidney Tissue Immunohistochemistry for LIMS1: (a) kidney cortex (200x) with strong staining of distal tubules but negative proximal tubules; **(b)** kidney cortex at 400x highlights strong positivity of distal tubules; **(c)** high power view of the glomerulus (600x) reveals positive endothelial cells and parietal epithelial cells and weak podocyte staining; **(d)** positive distal tubules and arterial endothelium (400x); **(e)** higher power (600x) view of positive arterial endothelial cells; **(f)** high power (600x) view of proximal (negative) and distal (positive) tubules; **(g)** medullary ray (400x); **(h)** medulla (200x); **(i)** positive medullary tubules and thin limbs of Henle (400x); **(j-l)** high power view (600x) with positive basolateral staining pattern of medullary collecting ducts and negative interstitial staining.

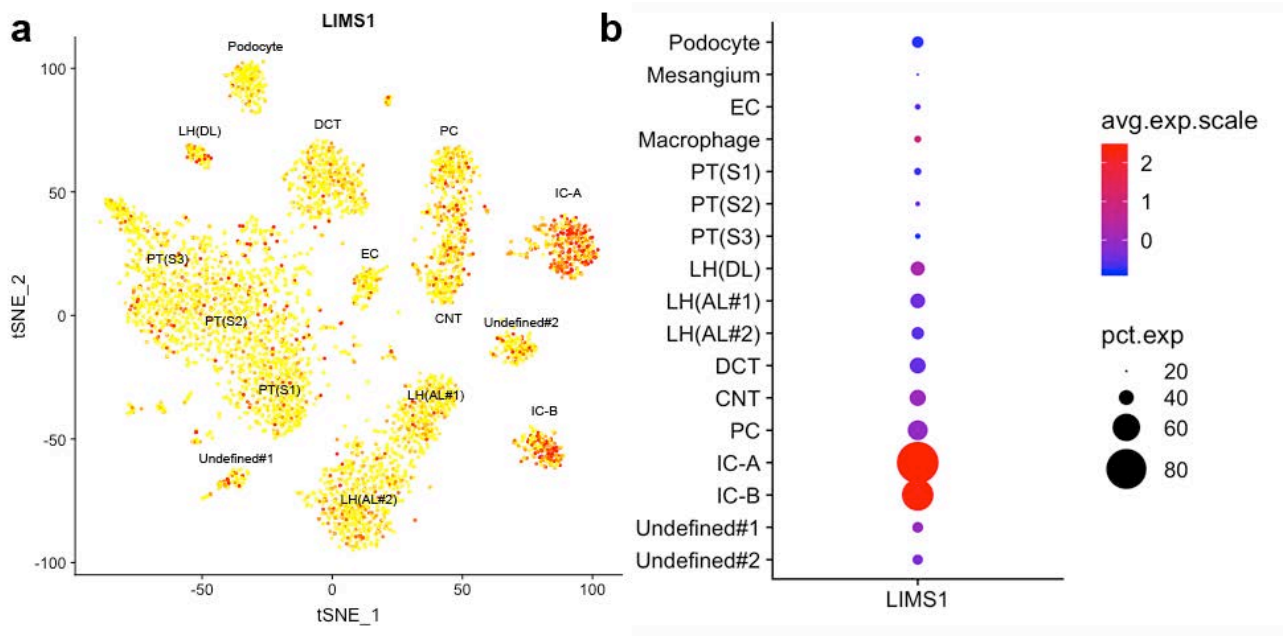


Figure S16. Human Kidney Tissue LIMS1 mRNA Expression by snRNA-seq: (a) tSNE plots for single nuclei RNA sequencing of normal human kidney; each point represents a single nucleus, each cluster corresponds to an individual kidney cell type; the cells expressing *LIMS1* are marked in red, cells negative for *LIMS1* are marked in yellow; **(b)** comparison of average mRNA expression levels per cell type demonstrates a gradient of *LIMS1* expression across the nephron, with relatively weak expression in proximal tubule segments and high expression in distal nephron, including highest levels in the collecting duct intercalated cells; snRNA-seq data was generated, analyzed and visualized by the Humphreys Lab and downloaded from the KIT (Kidney Interactive Transcriptomics) atlas at http://humphreyslab.com/SingleCell. EC=endothelial cell; PT=proximal tubule; LH=Loop of Henle; DCT=Distal Convoluted Tubule; PC=Principal Cells; IC-A=alpha-intercalated cells; IC-B=beta-intercalated cells.

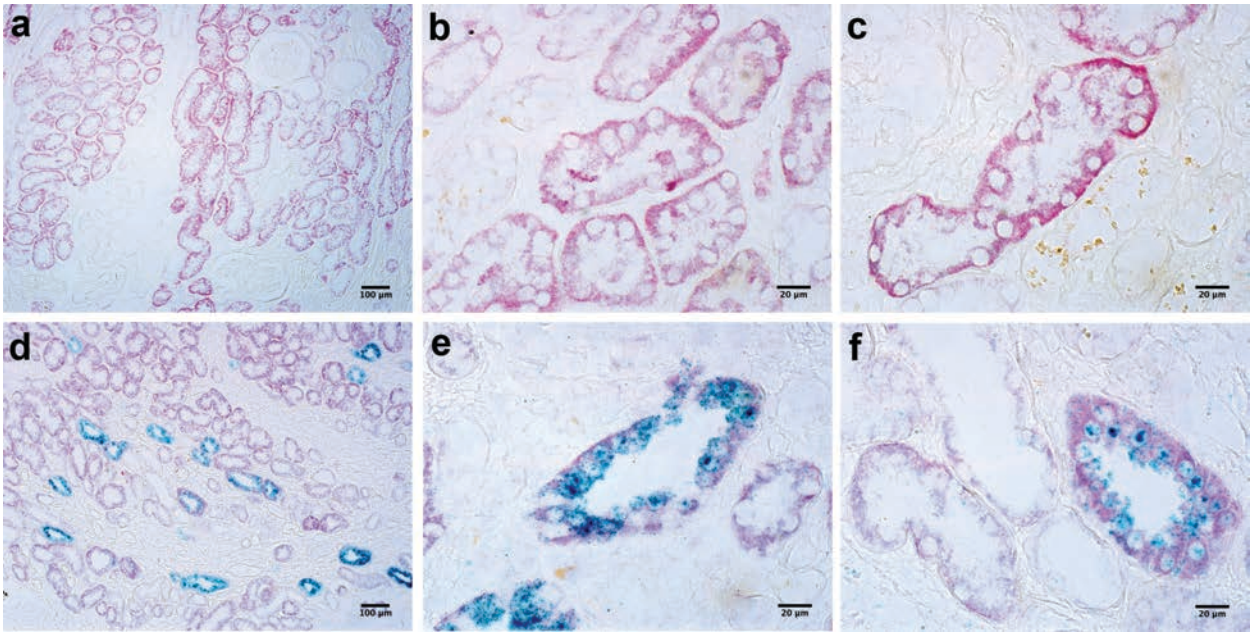
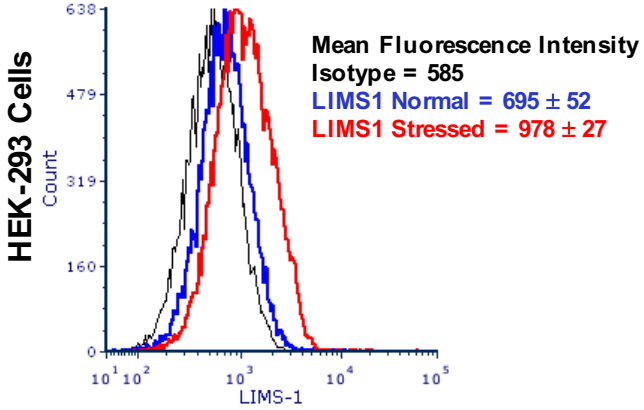


Figure S17. RNAScope® *in situ* hybridization confirms a gradient of *LIMS1* mRNA expression in human kidney tubules: (a-c) single stain for *LIMS1* (red, a-c) and (d-f) dual stain [*AQP2* (green) and *LIMS1* (red)]. The staining pattern is generally consistent with the immunohistochemistry staining: **(a)** low resolution (100x) view of the cortex demonstrates *LIMS1* mRNA expression in some, but not all kidney tubules; **(b)** high resolution (600x) view of *LIMS1* positive tubules in the cortex; **(c)** high resolution (600x) view of *LIMS1* positive tubules in the medulla; **(d)** low resolution (100x) view demonstrating that both *AQP2*-positive (collecting duct) and *AQP2*-negative epithelial cells express *LIMS1*; **(e)** high resolution (600x) view demonstrating strong *LIMS1* staining in the *AQP2*-positive cells; **(f)** weaker *LIMS1* staining is apparent in other AQP2-negative cells. *AQP2*: encodes Aquaporin 2, marker of the collecting duct principal cells.

a



b

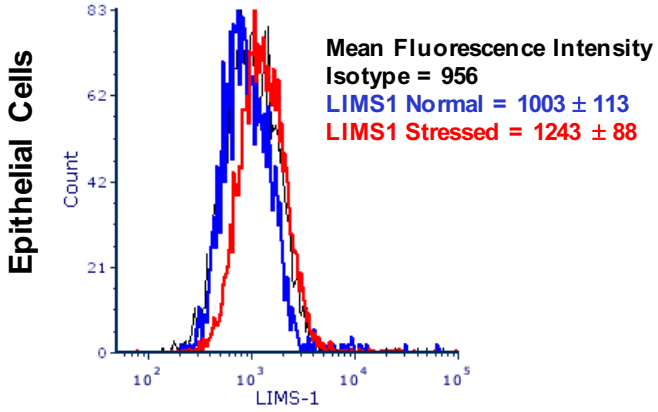


Figure S18. Hypoxia induces LIMS1 expression in human kidney cells. Flow cytometric analysis of the surface expression of LIMS1 protein under normal (blue) and stressed (hypoxic, red) conditions on cultured HEK-293 cells and cultured human kidney epithelial cells: **(a)** under hypoxic conditions, surface expression of LIMS1 was significantly increased in HEK-293 cells when compared to normal non-hypoxic control cells ($p<0.01$), and **(b)** a similar non-significant trend was observed for HRCE cells ($p=0.077$). The Y-axis represent the number of events and the X-axis represent the fluorescent intensity on a log scale.

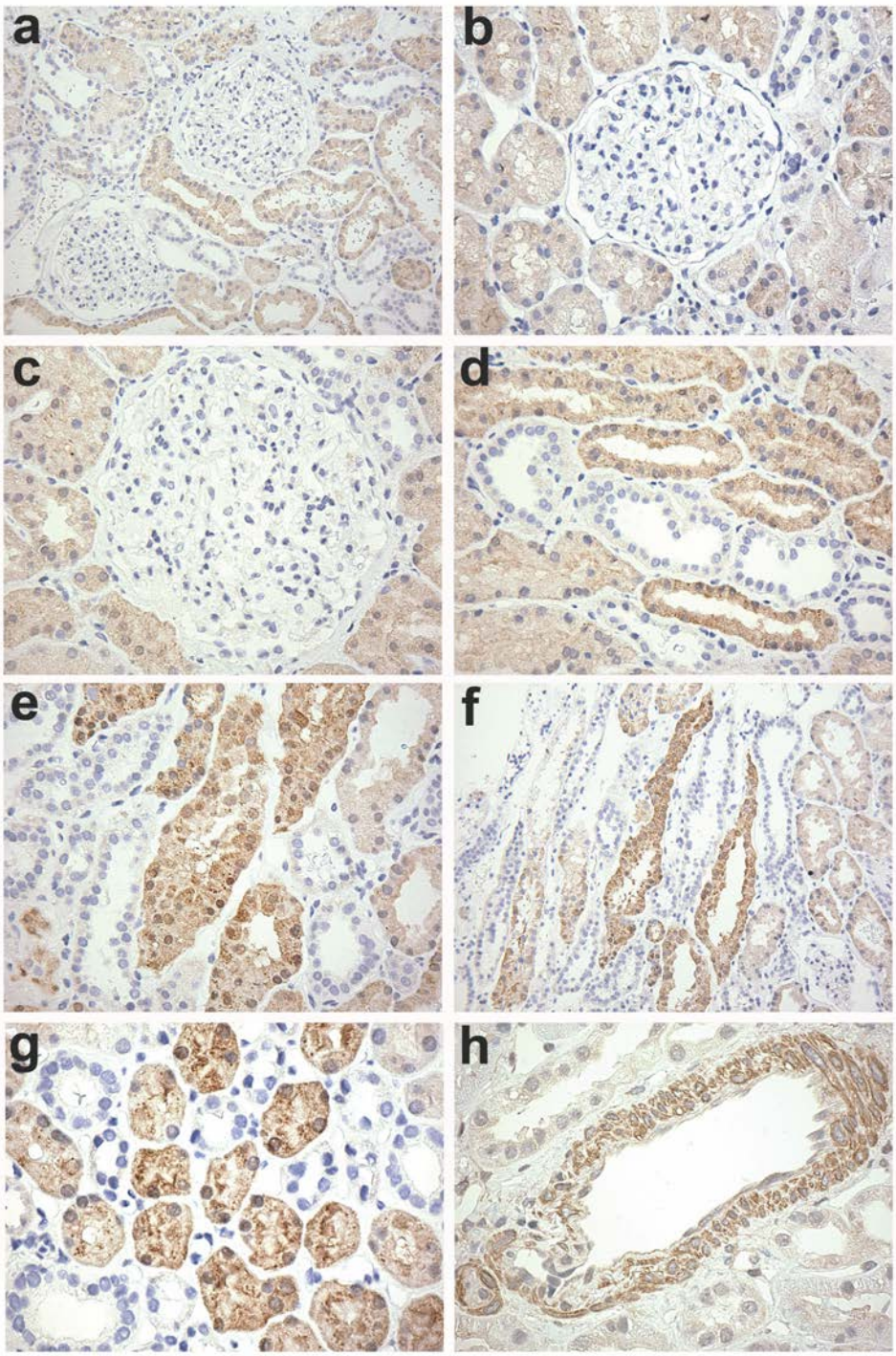


Figure S19. Human Kidney Tissue Immunohistochemistry for GCC2: kidney cortex at **(a)** 200x and **(b)** 400x with positive cytoplasmic staining of proximal tubules but no glomerular staining; **(c)** high power (600x) view of the glomerulus shows no intra-glomerular staining; **(d)** high power (600x) view of the cortical tubules reveals punctate cytoplasmic staining within proximal tubular cells and negative adjacent distal tubules; **(e)** higher power (600x) view of strongly positive S3 segments of proximal tubules; **(f)** low power (200x) view of positive S3 segment in medullary rays; **(g)** high power (600x) view of S3 in medullary rays; **(h)** high power (600x) view of the arterial vessel with positive staining of arterial medial myocytes and negative staining of endothelial cells.

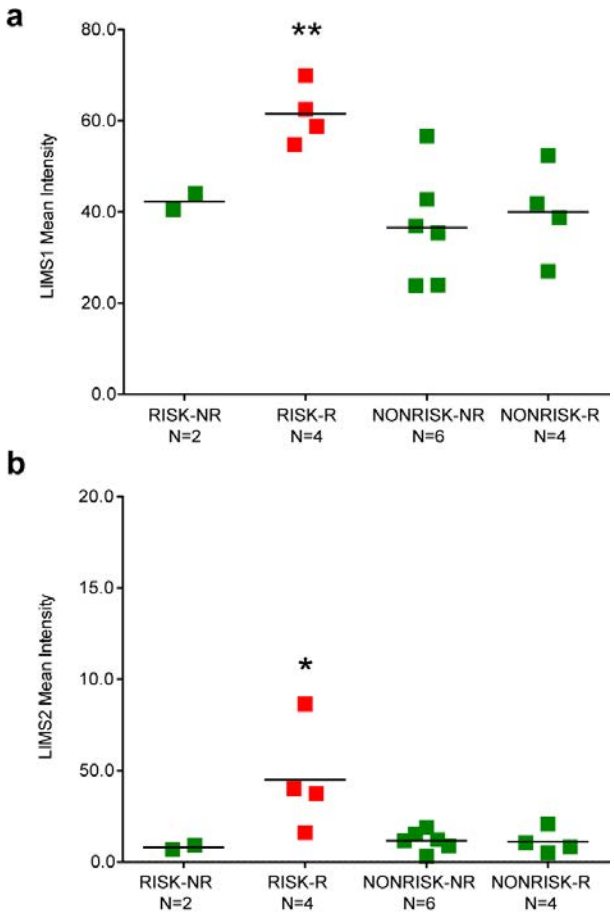


Figure S20. **Protoarray normalized intensity for LIMS1 and LIMS2 by specific genotype-phenotype group**: ** P=0.002 for LIMS1 mean intensity in high risk recipients with rejection versus all other groups; * P=0.007 for LIMS2 mean intensity in high risk recipients with rejection versus all other groups. Note that the mean intensity (Y-axis) for LIMS1 is 6-fold higher compared to LIMS2. These results suggest potential cross-reactivity of anti-LIMS1 antibodies.

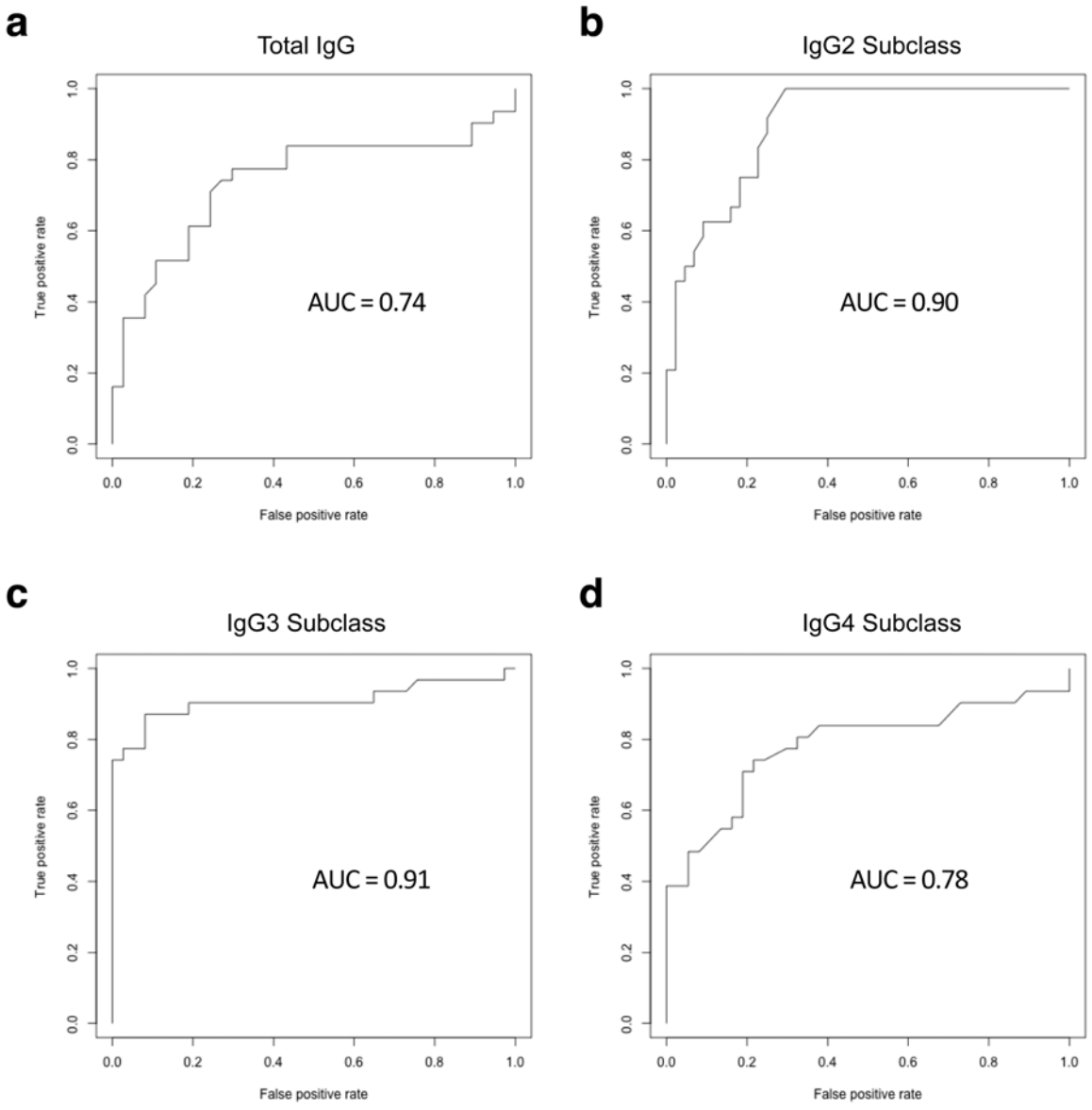


Figure S21. Receiver Operating Characteristic (ROC) curves for anti-LIMS1 IgG, IgG2, IgG3, and IgG4 demonstrate superior diagnostic performance of IgG2 and IgG3 subclasses: Kidney transplant recipients with rejection and a high risk genotype compared to all other groups. AUC = area under ROC curve.

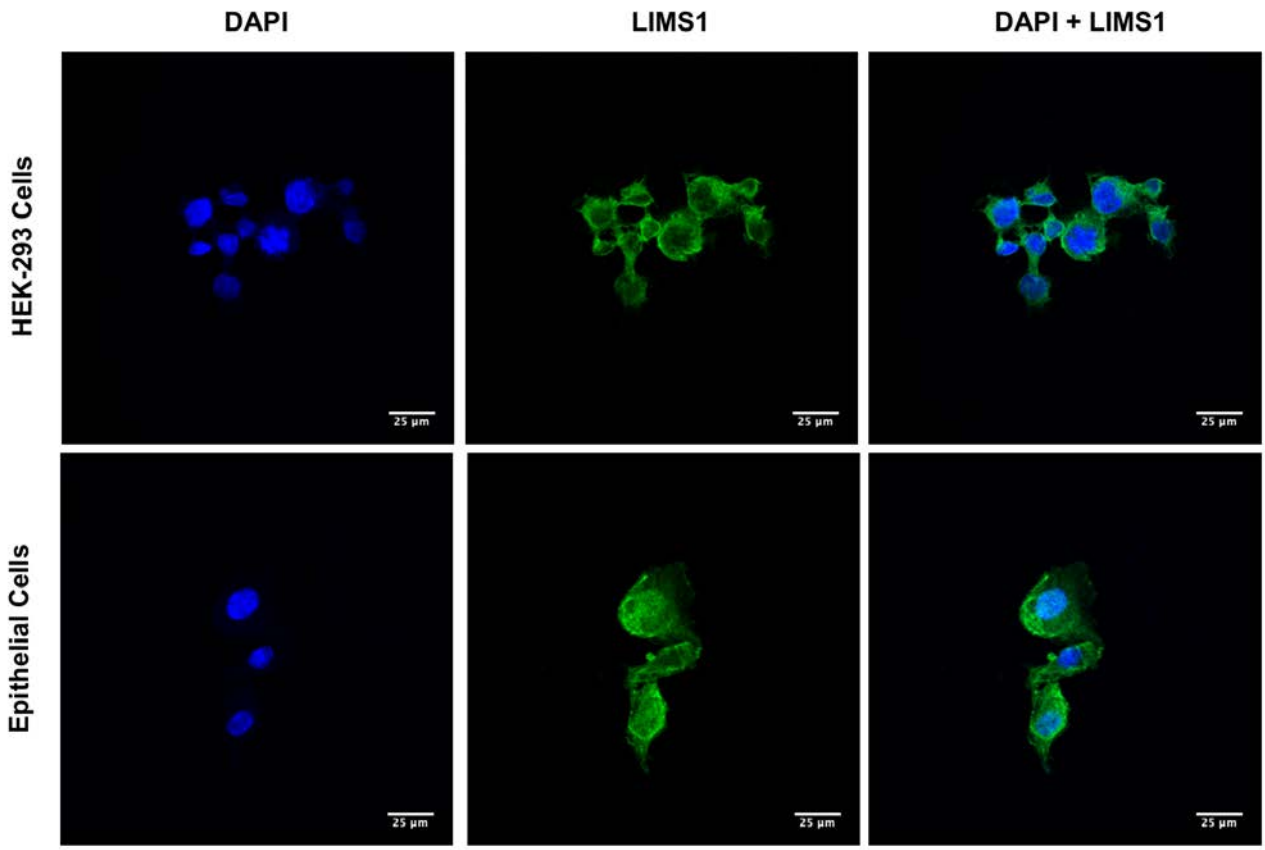


Figure S22. LIMS1 protein detection in HEK-293 cells (top), and in HRCE cells (bottom). Confocal microscopy images at 600x magnification; DAPI nuclear stain (blue); anti-LIMS1 antibody binding (green).

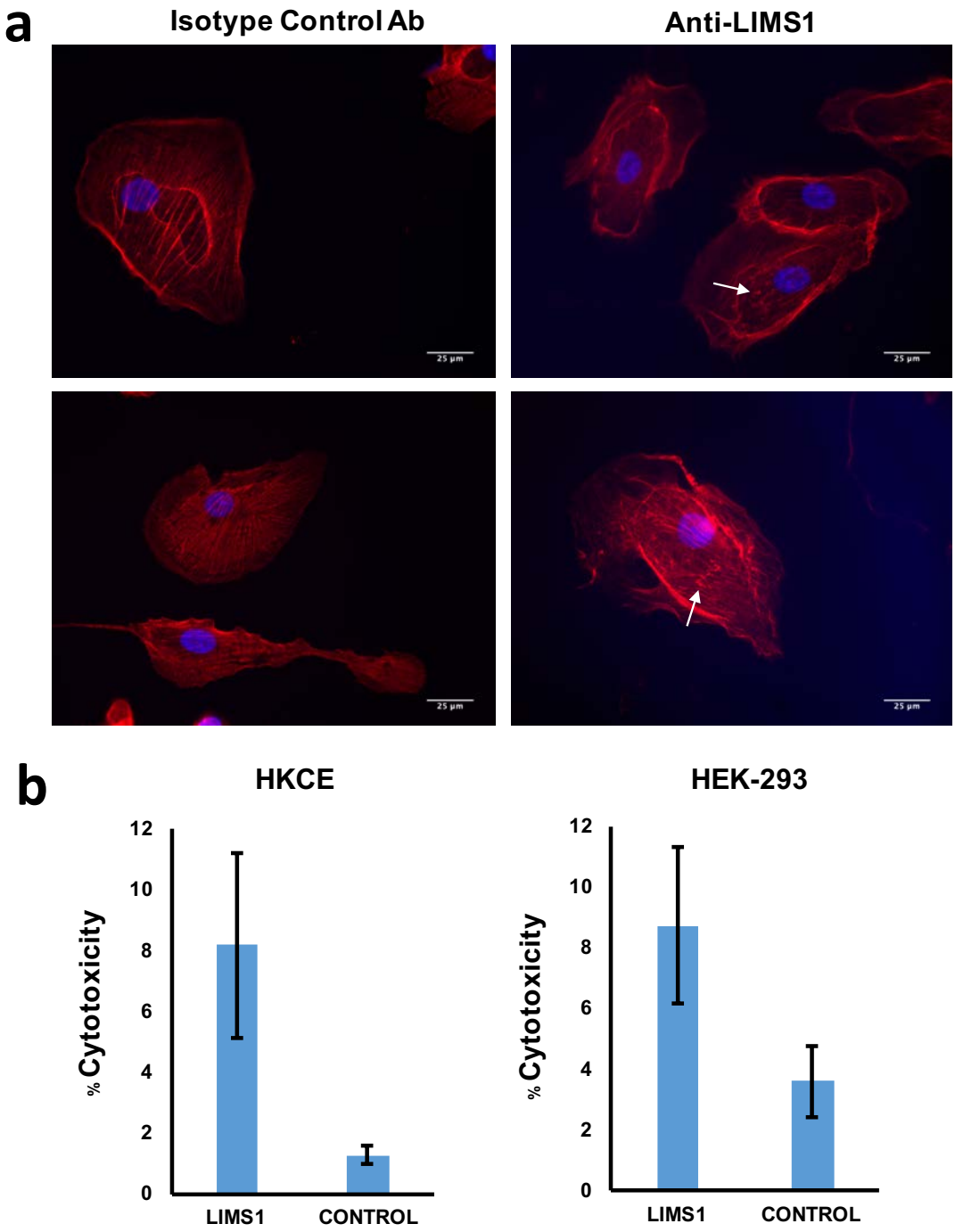


Figure S23. Anti-LIMS1 Ab induces cytotoxicity in cultured human kidney epithelial cells. (a) cultured human renal cortical epithelial (HRCE) cells were treated with control and mouse anti-human LIMS1 antibodies and labelled for DAPI (nuclei) and Phalloidin (F-actin binding protein) at 600x magnification; treatment with the anti-LIMS1 antibody disrupted the organization of the F-actin filaments compared to the non-specific control antibody; white arrows indicate abnormal clumping of actin filaments; **(b)** treatment of both HRCE and HEK-293 cells with anti-LIMS1 antibody had a significant cytotoxic effect by LDH assay as compared to the non-specific control antibodies (8.2 $\pm$ 3.1% *vs.* 1.3 $\pm$ 0.3%, $p<0.01$ and 8.7 $\pm$ 2.6% *vs.* 3.2 $\pm$ 1.2%, $p<0.01$, respectively).

Table S1. The study power: The power for the discovery phase was calculated for a range of expected effect sizes (HR 1.50-2.00) and MAFs (from 10% to 50%), assuming a recessive model, a perfectly tagged causal variant, average rejection rate of 35%, and a nominal replication threshold α=0.05 used for selection of markers for replication. The power for the joint analysis of discovery and replication was calculated using the Bonferroni-corrected significance threshold (α=1.1x10^{-3}), other assumptions as above. Only polymorphisms with MAF>10% were selected for genotyping based on these calculations.

	Discovery Phase Power (N=705)					
	HR=1.50	**HR=1.60**	**HR=1.70**	**HR=1.80**	**HR=1.90**	**HR=2.00**
MAF=0.10	8%	9%	11%	12%	14%	15%
MAF=0.20	18%	23%	28%	33%	38%	43%
MAF=0.30	33%	42%	51%	60%	67%	74%
MAF=0.40	49%	61%	72%	80%	86%	91%
MAF=0.50	62%	74%	84%	91%	95%	97%

	Discovery + Replication Power (N=2,709)					
	HR=1.50	**HR=1.60**	**HR=1.70**	**HR=1.80**	**HR=1.90**	**HR=2.00**
MAF=0.10	1%	2%	3%	4%	5%	7%
MAF=0.20	11%	18%	28%	38%	49%	60%
MAF=0.30	38%	57%	74%	85%	93%	97%
MAF=0.40	70%	87%	95%	98%	99%	100%
MAF=0.50	87%	96%	99%	100%	100%	100%

Table S2. Baseline characteristics of the Columbia Discovery Cohort: N=705 kidney transplant recipients recruited to the Columbia CKD Bio-bank in the years 2008-2012.

Cohort Characteristics	N (%)
Ethnicity (self-report):	
White (%)	336 (47.7%)
Black (%)	115 (16.3%)
Hispanic (%)	200 (28.4%)
East Asian (%)	40 (5.7%)
Male-to-Female ratio:	1.5
Mean Age (range):	46 (5-84)
Mean follow-up time (years):	8.6
High PRA (%):	79 (11.3%)
Family History of Renal Disease (%)	163 (24.7%)
Biopsy Diagnosis of Rejection (%)	234 (33.2%)
Borderline (%)	68 (9.6%)
T Cell Mediated Rejection (TCMR) (%)	138 (19.6%)
Grade IA (%)	62 (8.7%)
Grade IB (%)	47 (6.6%)
Grade IIA (%)	22 (3.1%)
Grade IIB (%)	4 (0.6%)
Grade III (%)	3 (0.4%)
Antibody Mediated Rejection (ABMR) (%)	24 (3.4%)
Chronic Active ABMR (%)	2 (0.3%)
History of Pregnancies (%)	196 (28.2%)
History of Transfusions (%)	304 (47.2%)
Previous Transplants (%)	129 (18.5%)
Donor Status:	
Living Related Donor (%)	273 (39.1%)
Living Unrelated Donor (%)	121 (17.3%)
Deceased Donor (%)	304 (43.6%)
Primary Diagnosis:	
Diabetic Nephropathy	123 (17.4%)
Hypertensive Nephropathy	97 (13.7%)
Glomerulonephritis	255 (36.2%)
Kidney Malformations	31 (4.4%)
Cystic Diseases	73 (10.4%)
Other	60 (8.5%)
Unknown	66 (9.4%)

Table S3. The deletions and their tag-SNPs genotyped in the discovery cohort

CHR	CNV id	Tag-SNP*	R2	MAF	DEL length#	INTERSECTED GENES
1	CNVR358.1	rs6693105	0.96	0.40	34481	LCE3B/LCE3C
1	CNVR117.1	rs11249248	0.95	0.44	63161	LOC100288240/RHCE/SDHDP7/TMEM50A
1	CNVR513.1	rs158736	1.00	0.34	6556	LOC100287974
1	CNVR217.1	rs11209948	1.00	0.38	45687	RPL31P12
1	CNVR360.1	rs11587012	1.00	0.34	11025	LCE1D/LCE1E/LOC100289267
1	CNVR90.1	rs10927864	1.00	0.35	3228	RPS16P1
1	CNVR459.1	rs7542235	0.97	0.22	114964	CFH/CFHR1/CFHR3/LOC100289145
2	CNVR915.1	rs893403	0.99	0.42	1456	LOC100288532
2	CNVR985.1	rs7419565	1.00	0.38	2668	TUBA3E
5	CNVR2719.1	rs2387715	0.88	0.32	57999	BTNL3/BTNL8/LOC100128762/LOC646227
5	CNVR2643.1	rs10053292	0.84	0.12	195944	HMGXB3/PDE6A/RPS20P4/SLC26A2/TIGD6
5	CNVR2637.1	rs7703761	1.00	0.44	2105	SPINK5L2
6	CNVR3057.1	rs17654108	0.96	0.15	3605	LOC260339
6	CNVR2845.6	rs36149991	1.00	0.13	13474	HLA-DRB5
7	CNVR3509.1	rs4729606	1.00	0.24	15353	ZAN
7	CNVR3673.2	rs4621754	1.00	0.10	8827	NCAPG2
7	CNVR3513.2	rs6943474	0.99	0.48	6015	EMID2
7	CNVR3354.1/CNVR3354.3	rs2160195	0.98	0.23	19835	TRGV1/TRGV2/TRGV3/TRGV4/TRGV5
8	CNVR3859.1	rs11985201	0.99	0.44	155757	ADAM3A/ADAM5P
8	CNVR3769.2	rs4543566	0.99	0.12	1807	DEFA10P
9	CNVR4426.1	rs1523688	1.00	0.20	7715	OR13C2/OR13C5
9	CNVR4456.1	rs2174926	0.99	0.45	1017	LOC442434
10	CNVR4893.1	rs10885336	0.94	0.43	4565	GUCY2G
10	CNVR4819.1	rs2342606	0.97	0.42	7898	LOC642521/LOC642538
10	CNVR4911.1	rs3793917	0.99	0.21	600	ARMS2
11	CNVR5178.5	rs11228868	1.00	0.10	10260	TRIM48
11	CNVR5165.1	rs4882017	1.00	0.37	3445	OR4A45P
11	CNVR5179.4	rs1944862	1.00	0.26	11502	OR4P1P
12	CNVR5430.1	rs1478309	0.99	0.19	37659	KLRC1/KLRC2/KLRC3
12	CNVR5433.1/CNVR5434.1	rs2256845	1.00	0.48	21783	TAS2R19/TAS2R31/TAS2R43/TAS2R64P
13	CNVR5826.1	rs9318648	0.86	0.24	15784	LOC374491
14	CNVR6117.1	rs11156875	1.00	0.15	9001	RPL23AP70
14	CNVR6079.1	rs8007442	0.99	0.40	792	TRAV14DV4
14	CNVR6209.2	rs8022070	0.99	0.12	1460	LOC731308
14	CNVR6294.26	rs4977155	0.98	0.18	137811	GOLGA4P1/IGHV3/IGHV4/IGHV7/IGHVII/IGHVIII
15	CNVR6302.2	rs8025963	0.83	0.25	58862	HERC2P6
16	CNVR6685.1	rs10521145	0.99	0.13	17516	SULT1A1
16	CNVR6730.1	rs2244613	0.83	0.16	27924	CES4
17	CNVR7096.1	rs8064493	1.00	0.27	10568	KRTAP9P1
17	CNVR7097.1	rs16966699	0.98	0.22	18679	KRT33A,KRT33B
19	CNVR7736.1	rs3810336	0.99	0.33	1939	GALP
19	CNVR7627.1	rs4806152	0.89	0.26	16347	FFAR3/GPR42P
19	CNVR7706.1	rs324121	1.00	0.14	2379	LOC400713
19	CNVR7722.1	rs103294	0.97	0.19	10965	LILRA3

* R2 estimated based on Europeans in HapMap; # deletion length in base pairs (bp).

Table S4. Baseline Clinical Characteristics of the Replication Cohorts

Cohort characteristics	TransplantLines Cohort (N=833)	Belfast Cohort (N=387)	Torino Cohort (N=784)
Mean Age (range)	47 (15-74)	41 (2-76)	49 (12-76)
Male-to-Female ratio	1.4	1.7	1.9
Mean follow-up time (years)	6.6 years	9.2 years	9.6 years
White/European Race	833 (100%)	387 (100%)	784 (100%)
Maximum PRA >30% (%)	111 (13.4%)	50 (13.0%)	107 (13.6%)
Biopsy Diagnosis of Rejection (%)	300 (36.0%)	92 (23.8%)	174 (22.2%)
History of Pregnancies (%)	NA	89 (23.0%)	165 (21.0%)
History of Transfusions (%)	NA	NA	399 (50.9%)
Previous Transplants (%)	94 (10.5%)	NA	46 (5.9%)
Donor Status:			
Deceased Donor (%)	678 (81.4%)	387 (100%)	784 (100%)
Living Donor (%)	155 (18.6%)	0 (0%)	0 (0%)
Primary Diagnosis:			
Diabetic Nephropathy	28 (3%)	NA	34 (4%)
Hypertensive Nephropathy	78 (9%)	NA	47 (6%)
Glomerulonephritis	192 (23%)	NA	286 (36%)
Kidney Malformations	19 (2%)	NA	25 (3%)
Cystic Diseases	138 (17%)	NA	141 (18%)
Other	275 (33%)	NA	172 (22%)
Unknown	103 (12%)	NA	79 (10%)
Risk Genotypes:			
Recipient rs893403 G allele frequency	0.402	0.390	0.450
Donor rs893403 G allele frequency	0.406	0.390	0.459
Recipient rs893403 GG genotype frequency	0.152	0.176	0.188
Donor-Recipient rs893403 collision frequency	0.120	0.142	0.149

Table S5. Recipients-only Cox proportional hazards association analysis of *LIMS1* risk genotype with time-to-first-rejection in the discovery, replication, and all cohorts combined.

Cohort	N	Minimally Adjusted Model*		Fully Adjusted Model**	
		HR (95%CI)	P-value	HR (95%CI)	P-value
Columbia Discovery	N=705 Recipients	1.84 (1.35-2.50)	9.77 x 10-5	1.76 (1.29-2.40)	3.80 x 10-4
Belfast Replication	N=387 Recipients	1.57 (0.97-2.54)	6.41 x 10-2	1.62 (1.00-2.62)	4.85 x 10-2
TransplantLines Replication	N=833 Recipients	1.41 (1.06-1.89)	1.87 x 10-2	1.48 (1.09-2.01)	1.12 x 10-2
Torino Replication	N=784 Recipients	1.18 (0.82-1.71)	0.375	1.20 (0.83-1.74)	0.325
All Replication	N=2004 Combined	1.36 (1.11-1.67)	3.2 x 10-3	1.41 (1.14-1.74)	1.59 x 10-3
All Cohorts	N=2709 Combined	1.48 (1.25-1.76)	6.21 x 10-6	1.50 (1.26-1.79)	5.42 x 10-6

* adjusted for cohort only (if applicable)

** adjusted for recipient's age, sex, race, ethnicity, HLA mismatch and cohort (if applicable)

Table S6. Comparison of alternative genetic models demonstrates the best fit of the "genomic collision" model. Note: A small number DR pairs with donors homozygous for rs893403-G but missing recipient genotype were classified as non-risk in the main study, but were excluded from this analysis in order to make the models comparable, i.e. all three models have identical number of observations, facilitating direct comparisons of statistical metrics of goodness-of-fit.

Model *	HR (95%CI)	P-value	R2	Log-likelihood	AIC
DR pairs, genomic collision^	1.65 (1.38-1.96)	3.15E-08	0.036	-6005.709	12019.42
Recipient only, recessive	1.48 (1.25-1.76)	6.21E-06	0.033	-6010.059	12028.12
Recipient only, additive	1.18 (1.07-1.31)	1.08E-03	0.030	-6014.206	12036.41

* All cohorts combined, excluding any transplant with missing recipient genotype and adjusted for cohort only

^ Genomic collision risk coded as recipient homozygosity for rs893403-G in the absence of donor homozygosity

Table S7. Comparison of alternative tests for association of collision genotype with rejection: non-parametric (log-rank) and logistic-regression-based statistical tests produce comparable results to Cox models.

Cohort	N	Log Rank Test, Unadjusted		Cox Proportional Hazards, Unadjusted Model		Logistic Regression, Unadjusted Model		Cox Proportional Hazards, Adjusted Model		Logistic Regression, Adjusted Model	
		Chisq	P-value	HR (95%CI)	P-value	OR (95%CI)	P-value	OR (95%CI)	P-value	OR (95%CI)	P-value
Columbia Discovery	N=705 Recipients	15.4, df=1	8.7 x 10-5	1.84 (1.35-2.50)	9.8 x 10-5	2.13 (1.40-3.23)	3.7 x 10-4	1.76 (1.29-2.40)	3.8 x 10-4	2.10 (1.36-3.26)	8.7 x 10-4
Belfast Replication	N=387 DR Pairs	4.2, df=1	4.1 x 10-2	1.70 (1.03-2.82)	4.0 x 10-2	1.87 (1.03-2.82)	4.5 x 10-2	1.77 (1.06-2.93)	2.8 x 10-2	1.96 (1.05-3.65)	3.3 x 10-2
TransplantLines Replication	N=833 DR Pairs	7.5, df=1	6.2 x 10-3	1.53 (1.13-2.09)	6.6 x 10-3	1.79 (1.17-2.74)	6.8 x 10-3	1.58 (1.14-2.19)	5.7 x 10-3	1.88 (1.19-2.98)	6.7 x 10-3
Torino Replication	N=784 DR Pairs	4.4, df=1	3.7 x 10-2	1.49 (1.02-2.16)	3.9 x 10-2	1.61 (1.03-2.50)	3.5 x 10-2	1.53 (1.05-2.23)	2.6 x 10-2	1.65 (1.06-2.59)	2.7 x 10-2
All Replication	N=2004 Combined	12.8, df=1	3.5 x 10-4	1.55 (1.25-1.93)	6.5 x 10-5	1.73 (1.32-2.28)	7.7 x 10-5	1.58 (1.27-1.97)	5.2 x 10-5	1.76 (1.32-2.34)	8.2 x 10-5
All Cohorts	N=2709 Combined	26.3, df=1	2.8 x 10-7	1.63 (1.37-1.95)	4.7 x 10-8	1.61 (1.47-2.32)	1.4 x 10-7	1.63 (1.36-1.95)	9.4 x 10-8	1.85 (1.46-2.34)	3.3 x 10-7

Table S8. Frequency of rs893403-G risk allele in the 1000 Genomes Populations.

1000G Population Name	Population Ancestry	Freq. of G allele	Freq. of GG genotype
AFR	African	0.423	0.175
EUR	European	0.407	0.163
SAS	South Asian	0.373	0.133
AMR	Admixed American	0.290	0.086
EAS	East Asian	0.024	0.000

Table S9. ANNOVAR annotation of all variants in LD (r2>0.8) with rs893403.

CHR	BP	SNPs	Functonal	Gene Annot
2	109310556	BI_GS_DEL1_B5_P0360_299	intergenic	LIMS1(dist=6854),RANBP2(dist=25381)
2	109138677	rs2258404	ncRNA_intronic	GCC2-AS1
2	109139209	rs2718694	ncRNA_intronic	GCC2-AS1
2	109316789	rs7575335	intergenic	LIMS1(dist=13087),RANBP2(dist=19148)
2	109066344	rs2460944	intronic	GCC2
2	109066799	rs1474220	intronic	GCC2
2	109067975	rs2683798	intronic	GCC2
2	109069008	rs2683797	intronic	GCC2
2	109070634	rs2718749	intronic	GCC2
2	109071433	rs3083173	intronic	GCC2
2	109076377	rs2049151	intronic	GCC2
2	109076457	rs2049150	intronic	GCC2
2	109076856	rs2139807	intronic	GCC2
2	109078616	rs9789664	intronic	GCC2
2	109079902	rs2577612	intronic	GCC2
2	109088868	rs1898557	intronic	GCC2
2	109089404	rs2243888	intronic	GCC2
2	109090111	rs2718704	intronic	GCC2
2	109095886	rs2577616	intronic	GCC2
2	109099477	rs2577613	intronic	GCC2
2	109099865	rs1542025	intronic	GCC2
2	109104464	rs10184417	intronic	GCC2
2	109107041	rs57919394	intronic	GCC2
2	109114691	rs2378149	intronic	GCC2
2	109116846	rs2003996	intronic	GCC2
2	109118132	rs2718761	intronic	GCC2
2	109120418	rs2438251	intronic	GCC2
2	109122204	rs3098319	intronic	GCC2
2	109217351	2:109217351:TCTCTTC:T	intronic	LIMS1
2	109217358	2:109217358:TCTTC:T	intronic	LIMS1
2	109260044	rs70956267	intronic	LIMS1
2	109065858	rs2460947	intronic	GCC2
2	109310555	rs10202224	intergenic	LIMS1(dist=6853),RANBP2(dist=25382)
2	109310556	BI_GS_DEL1_B5_P0360_299	intergenic	LIMS1(dist=6854),RANBP2(dist=25381)
2	109102534	rs2577586	intronic	GCC2
2	109108460	rs2577599	intronic	GCC2
2	109112655	rs2917988	intronic	GCC2
2	109088996	rs2577622	intronic	GCC2
2	109117975	rs2917983	intronic	GCC2
2	109128310	rs111963733	ncRNA_intronic	GCC2-AS1
2	109038026	rs13022595	intergenic	SULT1C4(dist=32600),GCC2(dist=27551)
2	109042040	rs59274398	intergenic	SULT1C4(dist=36614),GCC2(dist=23537)
2	109042493	rs1829601	intergenic	SULT1C4(dist=37067),GCC2(dist=23084)
2	109042578	rs1829599	intergenic	SULT1C4(dist=37152),GCC2(dist=22999)
2	109043961	rs4271731	intergenic	SULT1C4(dist=38535),GCC2(dist=21616)
2	109044008	rs1915487	intergenic	SULT1C4(dist=38582),GCC2(dist=21569)
2	109047386	rs10170784	intergenic	SULT1C4(dist=41960),GCC2(dist=18191)
2	109049305	rs35256991	intergenic	SULT1C4(dist=43879),GCC2(dist=16272)
2	109058119	rs2683808	intergenic	SULT1C4(dist=52693),GCC2(dist=7458)
2	109059760	rs2718759	intergenic	SULT1C4(dist=54334),GCC2(dist=5817)
2	109060047	rs2718758	intergenic	SULT1C4(dist=54621),GCC2(dist=5530)
2	109060094	rs2176959	intergenic	SULT1C4(dist=54668),GCC2(dist=5483)
2	109060227	rs2139811	intergenic	SULT1C4(dist=54801),GCC2(dist=5350)
2	109062290	rs1464406	intergenic	SULT1C4(dist=56864),GCC2(dist=3287)
2	109128129	rs62148145	ncRNA_intronic	GCC2-AS1
2	109048145	rs147976738	intergenic	SULT1C4(dist=42719),GCC2(dist=17432)
2	109097214	rs35817047	intronic	GCC2
2	109119111	rs2438253	intronic	GCC2
2	109119112	rs2438252	intronic	GCC2
2	109057701	rs59698498	intergenic	SULT1C4(dist=52275),GCC2(dist=7876)
2	109128319	rs112503499	ncRNA_intronic	GCC2-AS1
2	109127984	rs62148112	ncRNA_intronic	GCC2-AS1
2	109128217	rs200858383	ncRNA_intronic	GCC2-AS1
2	109312728	rs7606621	intergenic	LIMS1(dist=9026),RANBP2(dist=23209)
2	109128296	rs73954373	ncRNA_intronic	GCC2-AS1
2	109128231	rs201904559	ncRNA_intronic	GCC2-AS1
2	109128260	rs73954372	ncRNA_intronic	GCC2-AS1
2	109239869	rs826688	intronic	LIMS1
2	109045638	rs2139809	intergenic	SULT1C4(dist=40212),GCC2(dist=19939)
2	109048620	rs147362369	intergenic	SULT1C4(dist=43194),GCC2(dist=16957)
2	109057011	rs35997658	intergenic	SULT1C4(dist=51585),GCC2(dist=8566)
2	109217352	rs11123709	intronic	LIMS1
2	109084523	rs55803150	intronic	GCC2
2	109305944	rs865444	intergenic	LIMS1(dist=2242),RANBP2(dist=29993)
2	109305465	rs376136163	intergenic	LIMS1(dist=1763),RANBP2(dist=30472)
2	109145808	rs2118446	ncRNA_intronic	GCC2-AS1
2	109134387	rs2465951	ncRNA_intronic	GCC2-AS1
2	109150714	rs10084199	upstream	GCC2-AS1,LIMS1(dist=97)
2	109170306	rs1469966	intronic	LIMS1
2	109062693	rs2718755	intergenic	SULT1C4(dist=57267),GCC2(dist=2884)
2	109064563	rs11123694	intergenic	SULT1C4(dist=59137),GCC2(dist=1014)
2	109128632	rs2718764	ncRNA_intronic	GCC2-AS1
2	109042679	rs1829598	intergenic	SULT1C4(dist=37253),GCC2(dist=22898)
2	109045009	rs7596199	intergenic	SULT1C4(dist=39583),GCC2(dist=20568)
2	109049182	rs10200997	intergenic	SULT1C4(dist=43756),GCC2(dist=16395)
2	109060980	rs2683806	intergenic	SULT1C4(dist=55554),GCC2(dist=4597)
2	109131824	rs2953739	ncRNA_intronic	GCC2-AS1
2	109131828	rs2917971	ncRNA_intronic	GCC2-AS1
2	109050406	rs140686654	intergenic	SULT1C4(dist=44980),GCC2(dist=15171)
2	109166974	rs10084394	intronic	LIMS1

Table S10. Immunohistochemistry staining patterns for LIMS1 and GCC2 in human kidney tissue compartments. Supportive imaging data provided in Figures S3 and S4.

	Cell Type or Segment	LIMS1	GCC2
Glomerulus	Podocytes	Weakly Positive	Negative
	Endothelial Cells	Weakly Positive	Negative
	Mesangial Cells	Negative	Negative
Cortical Tubules	Proximal Tubules	Negative	Positive (S3 >> S1)**
	Distal Tubules	Strongly Positive*	Negative
Medullary Tubules	MTAL and LOH	Strongly Positive*	Negative
	Collecting Ducts	Strongly Positive*	Negative
Vascular & Interstitial	Arterial Endothelial Cells	Positive	Negative
	Arterial Smooth Muscle Cells	Negative	Positive
	Interstitium	Negative	Negative

* basolateral subcellular staining pattern
** punctate cytoplasmic subcellular distribution, strongest in S3 segment of the proximal tubule

Supplemental References:

1. Backenroth D, He Z, Kiryluk K, et al. FUN-LDA: A Latent Dirichlet Allocation Model for Predicting Tissue-Specific Functional Effects of Noncoding Variation: Methods and Applications. Am J Hum Genet 2018;102:920-42.
2. Gillies CE, Putler R, Menon R, et al. An eQTL Landscape of Kidney Tissue in Human Nephrotic Syndrome. Am J Hum Genet 2018;103:232-44.
3. Manalo DJ, Rowan A, Lavoie T, et al. Transcriptional regulation of vascular endothelial cell responses to hypoxia by HIF-1. Blood 2005;105:659-69.
4. Eisenga MF, Gomes-Neto AW, van Londen M, et al. Rationale and design of TransplantLines: a prospective cohort study and biobank of solid organ transplant recipients. BMJ Open 2018;8:e024502.
5. Genomes Project C, Abecasis GR, Auton A, et al. An integrated map of genetic variation from 1,092 human genomes. Nature 2012;491:56-65.
6. Robinson JT, Thorvaldsdottir H, Winckler W, et al. Integrative genomics viewer. Nat Biotechnol 2011;29:24-6.
7. Quinlan AR, Hall IM. BEDTools: a flexible suite of utilities for comparing genomic features. Bioinformatics 2010;26:841-2.
8. Fujita PA, Rhead B, Zweig AS, et al. The UCSC Genome Browser database: update 2011. Nucleic Acids Res 2011;39:D876-82.
9. Benson DA, Cavanaugh M, Clark K, et al. GenBank. Nucleic Acids Res 2013;41:D36-42.
10. Benson DA, Karsch-Mizrachi I, Lipman DJ, Ostell J, Wheeler DL. GenBank: update. Nucleic Acids Res 2004;32:D23-6.
11. Lestrade L, Weber MJ. snoRNA-LBME-db, a comprehensive database of human H/ACA and C/D box snoRNAs. Nucleic Acids Res 2006;34:D158-62.
12. Griffiths-Jones S, Saini HK, van Dongen S, Enright AJ. miRBase: tools for microRNA genomics. Nucleic Acids Res 2008;36:D154-8.
13. Lowe TM, Eddy SR. tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence. Nucleic Acids Res 1997;25:955-64.
14. Lewis BP, Burge CB, Bartel DP. Conserved seed pairing, often flanked by adenosines, indicates that thousands of human genes are microRNA targets. Cell 2005;120:15-20.
15. Davydov EV, Goode DL, Sirota M, Cooper GM, Sidow A, Batzoglou S. Identifying a high fraction of the human genome to be under selective constraint using GERP++. PLoS Comput Biol 2010;6:e1001025.
16. Roadmap Epigenomics C, Kundaje A, Meuleman W, et al. Integrative analysis of 111 reference human epigenomes. Nature 2015;518:317-30.
17. Ernst J, Kellis M. ChromHMM: automating chromatin-state discovery and characterization. Nature methods 2012;9:215-6.
18. Hoffman MM, Buske OJ, Wang J, Weng Z, Bilmes JA, Noble WS. Unsupervised pattern discovery in human chromatin structure through genomic segmentation. Nature methods 2012;9:473-6.
19. Ward LD, Kellis M. HaploReg: a resource for exploring chromatin states, conservation, and regulatory motif alterations within sets of genetically linked variants. Nucleic acids research 2012;40:D930-4.
20. Gadegbeku CA, Gipson DS, Holzman LB, et al. Design of the Nephrotic Syndrome Study Network (NEPTUNE) to evaluate primary glomerular nephropathy by a multidisciplinary approach. Kidney Int 2013;83:749-56.
21. Cohen CD, Frach K, Schlöndorff D, Kretzler M. Quantitative gene expression analysis in renal biopsies: a novel protocol for a high-throughput multicenter application. Kidney international 2002;61:133-40.

22. Lockstone HE. Exon array data analysis using Affymetrix power tools and R statistical software. Briefings in bioinformatics 2011:bbq086.
23. Stegle O, Parts L, Durbin R, Winn J. A Bayesian framework to account for complex non-genetic factors in gene expression levels greatly increases power in eQTL studies. PLoS Comput Biol 2010;6:e1000770.
24. Parts L, Stegle O, Winn J, Durbin R. Joint genetic analysis of gene expression data with inferred cellular phenotypes. PLoS Genet 2011;7:e1001276.