**SUPPLEMENT**

**COMPLETE METHODS**

**Animals**

*Col4a5* knockout mice, B6.Cg-*Col4a5tm1Yseq*/J (Stock#006183), was developed by Dr. Yoav Segal using a targeting vector containing a loxP site flanked neomycin resistance gene and a G213T point mutation was introduced into exon. The construct was electroporated into 129X1/SvJ (Stock#000691 derived ESVJ-1182 embryonic stem (ES) cells. ES cells containing the point mutation were injected into C57BL/6J (Stock#000664). The neo cassettes were removed by crossing to FVB/N-Tg(ACTB-cre) 2Mrt/J (Stock#3376) mice to and then backcrossed to C57BL/6J for 15 generations. As a result regions immediately surrounding the point mutation contain residual 129X1/SvJ markers.

*Col4a5* knockout mice for this experiment were rederived from cryopreservation at The Jackson Laboratory, with the females maintained as heterozygous for the *Col4a5* mutation and hemizygous in males. 100 heterozygous female *Col4a5* mutant mice were crossed with 100 unique diversity outbred males, J:DO (JR#009376), to generate a cohort of 100 males and 100 female F1 animals, where females were heterozygous of the *Col4a5* mutation and males were homozygous. Each of the 200 F1 animals carries a single copy of C57BL/6J at each chromosome, while the other copy is contributed form on DO background. The DO background is a unique mixture of eight founder strains, which include five classical inbred strains (129S1/SvImJ, A/J, C57BL/6J, NOD/ShiLtJ, NZO/HILtJ) and three wild-derived strains (CAST/EiJ, PWK/PhJ and WSB/EiJ). Thus at a given locus each F1 animal will have a C57BL6/J allele, and one of eight founder strain allele.

**Albuminuria analysis**

Spot urine was collected at 6, 10, and 15 weeks of age for urinary albumin and creatinine measurements. Both urinary albumin and creatinine concentrations were determined using Synchron CX5 Chemistry Analyzer (Beckman Coulter), and the amount of albuminuria was determined with albumin to creatinine ratio (mg/g).

**Glomerular filtration rate analysis**

Glomerular filtration rate (GFR) was measured at 14 weeks of age. Mice were weighed one week prior to testing to establish dosage of FITC-inulin. A 5% FITC-inulin (Sigma, F3272) in 0.85% NaCl was prepared and dialyzed using a dialysis membrane (Spectrum labs, MWCO 1KD 132636) for 24 hours protected from light, and filtered using a 0.2 uM syringe filter (VWR, 28145-477). Animals were anesthetized with isoflurane prior to retro-orbital injection with FITC-inulin at a dose of 3.74ul x body weight (g) rounded to the nearest 10ul. Serial blood samples were taken at precise time points (0, 3, 5, 7, 10, 15, 35, 56, and 75 minutes post injection) from a nick in the tail tip. All blood was collected for a maximum during of one minute with a maximum quantity of 25ul. Blood samples were spun down and 5ul of serum were aliquoted in triplicates into a 384 well plate and read on a Spectramax i3 fluorescent plate reader (Molecular devices) with emission and excitation wavelengths set at 484nm and 535nm respectively. Triplicate readings were taken and assessed for technical precision using a 10% CV cutoff. We used a 2 compartment model (y = A\*exp(-B\*x) + C\*exp(-D\*x) + noise) for GFR quantification{Hall:1977vx}. GFR was determined using the initial fluorescent intensity, which was measured using a time 0 serum with added FITC-inulin corrected for dilution factor, divided by the area under the curve. A tool has been developed to automate this calculation, which can be found at <https://github.com/simecek/GFRcalc>.

**Kidney collection**

Right kidneys were collected at 15weeks after last urine collection, and the renal capsule containing perinephritic adipose tissue was removed before it was immediately flash frozen in liquid nitrogen. Each kidney was ground using a ceramic mortal and pestle on dry ice into frozen homogenate and separated into 3 aliquots for downstream analysis.

**DNA isolation and genotyping**

A non-phenol-chlorofrom based DNA isolation protocol was used to obtain higher quality samples, than compared to standard laboratory practices. Tail tips were collected at wean (4 weeks) and digested using proteinase K overnight. Samples were cooled to room temperature before protein precipitation solution containing 5M of ammonium acetate was added, vortexed, and incubated on ice for 30mins. The samples were spun at 14,000 rpm (applies to the rest of this protocol), and supernatants were pipetted into a clean tube. Isopropanol was used to precipitate DNA, and solution was centrifuged to a pellet. Then 70% ethanol was used to further desalt and precipitate DNA and centrifuged once more. The ethanol was discarded leaving a pellet of DNA. The samples were left on a bench top covered with a paper towel to dry. Once no liquid was visible, DNA was re-suspended in 100ul of ddH20 and incubated at 65C for 5 mins. DNA concentration and purity were measured using NanoDrop 2000 (Thermo Scientific). Samples were genotyped for *Col4a5* mutation using forward primer 5’GCATAACCGGGACACTCA-CT3’ and reverse primer 5’GAGGACTTACCGCAGCCTCT3’ to capture knockin construct located in exon 1. Samples for GigaMUGA genotyping met stringent quality standards of A260/280 ratio between 1.7 and 2.1. A minimum aliquot of 20ul at 20ng/ul concentrations were sent for array genotyping.

**Genotyping with GigaMuga**

All 200 mice were fully genotyped for 143,259 SNPs by GeneSeek (Neogen Genomics) using the Giga Mouse Universal Genotyping Array, GigaMUGA, built on an Illumina Infinium platform. Genotype calls of A, B, H, or N were generated using Illumina’s BeadStudio algorithm, whereby A represents homozygous reference allele, B represents homogygous for the alternate allele, H represents heterozygous genotype, and N represents “no call” at marker.

**Whole-genome diplotype probability construction using GigaMUGA**

Each chromosome pair of a B6.Cg-*Col4a5tm1Yseq*/J and a J:DO F1 animal is composed of a C57BL/6J haploid and a haploid containing unique mosaic of founder haplotypes. Here we refer to the haplotype at a given locus as a diplotype, where each diplotype consists of a haplotype from each parent. In the F1 mouse model there are 8 possible diplotypes, 1 homozygous and 7 heterozygous diplotypes. Gatti et al., has developed a hidden Markov model to reconstruct the diplotypes by generating a probabilistic estimate of the diplotype state at each SNP marker locus for all 200 animals{Gatti:2014ko}. To ensure quality of construction, 182 samples with call rates of 90% and over were kept.

**RNA extraction and quality control**

One homogenized kidney aliquot was sent to Genome Technologies, a scientific research service available at the Jackson Laboratory, for RNA extraction and RNA-seq library prep. Kidney samples were further lysed and homogenized in TRIzol Reagent (Ambion), and total RNA was extracted using miRNeasy Mini Kit (Qiagen), according to manufacturer’s protocols, including the optional DNase digest step. Sample concentration and quality were accessed using Nanodrop 2000 spectrophotometer (Thermo Scientific) and the RNA 6000 Nano LabChip assay (Agilent Technologies) respectively. RNA quality criteria for library construction were RIN of ≥ 8.0 and a 260/280 ratio of ≥ 1.7.

**Library construction and RNA sequencing**

Poly(A) RNA-seq libraries were constructed using TruSeq RNA Library Prep Kit v2 (Illumina), including the addition of unique barcode sequences to multiplex sequencing, and were quantified using quantitative PCR (Kapa Biosystems). Libraries were pooled and sequenced at 100bp single-end on the HiSeq 2500 (Illumina) using TruSeq SBS Kit v4 at the New York Genome Center. (Metrics needed for insert size, and average # of reads per sample)

**Allele specific expression analysis and whole-genome diplotype reconstruction using RNA-seq**

Both calculations for the allele specific expression analysis and whole-genome diplotype reconstruction were performed using a combination of Expectation-Maximization algorithm for Allele Specific Expression (EMASE) and Genotyping By RNA-Seq (GBRS) software respectively (manuscript in preperation). EMASE was used to align multi-parent allele-specific expression and gene expression simultaneously from RNA-seq data, and the diploid BAM files were used as input for GBRS. GBRS was used to quantify multiway allele specificity taking into account DO generation, which quantified expression contribution from each haplotype in a diplotype. The quantified multiway gene transcript per million (TPM) count was used to reconstruct genome probabilities along with an established reference transcriptome probability file that corresponds to the samples DO generation and sex.

In order to accurately compare and use the reconstructed genome probabilities with that of the genome probabilities from GeneSeek, we interpolated the output file in a 64k SNP grid to a suitably spaced-grid used for GeneSeek using GBRS’s interpolate tool.

**Improving accuracy of whole-genome diplotype probability construction**

Initial steps were taken using the DOQTL R package{Gatti:2014ko} to create kinship probability plots of the GigaMUGA genome probability construction to visualize and confirm heterogeneity of the F1 samples. The expectation is to see complete heterogeneity of kinship, as DO sire contributing to the F1 is genetically unique from one another. We were able to confirm 13 samples that were closely related to each other, 11 of which was due to sample switching with another Neogen Genomics customer and 2 samples that were duplicates.

Secondary steps were taken by cross-comparing each GigaMuga genome construction to their RNA-Seq reconstruction. Through this we found 12 samples that did not correlate with each other. All 25 samples that were identified to be problematic GigaMUGA constructs were replaced with RNA-seq reconstructions and additional samples that did not have sufficient GigaMUGA call rates were also replaced, giving a total of 192 samples for downstream analysis.

**Quantitative trait loci analysis**

Unlike simple inbred cross designs, quantitative trait loci (QTL) mapping for a F1 model with DO background requires the use of a mixed-linear regression model accounting for kinship{Gatti:2014ko}. DOQTL R package was used to perform additive and full QTL models for both haplotypes and SNP calls. Haplotypes QTL models compute allelic dosage of founders at a given haplotype block associated to a founder to determine founder effects at a given locus. Haplotype QTL was used to create QTL maps showing LOD scores for each autosomal chromosome, founder effect plots, as well as calculation of Bayesian intervals. Used in conjunction with haplotype QTL models, SNP call models compute the probabilistic imputation of the genotype at every known SNP locus genome-wide, total of 143,259 SNPs, similar to that of genome wide association studies (GWAS) well established in human studies. SNP call models were used to identify LOD scores of individual SNPs within a Bayesian interval. Analyzing both prior mentioned QTL models allows for high resolution mapping to narrow down candidate genes. All QTL modelss analyzed for GFR and Albumin at all time points accounted for sex as an additive covariate, and creatinine was also accounted for in Albumin QTLs for normalization.

**Additional materials**

Codes to all figures and analyses can be found at <https://github.com/TheJacksonLaboratory/1415-Col4a5xDO-Project>. (Will have to clean up repo or create a new one for public).

All animal experiments were performed in accordance with the National Institutes of Health Guide for the Care and Use of Laboratory Animals (National Research Council) and were approved by The Jackson Laboratory’s Animal Care and Use Committee.