**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 129SvJ 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 129SvJ 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 carry 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.

**DNA isolation**

The DNA isolation protocol did not contain phenol-choloroform to obtain higher quality samples, than compared to standard lab practices. Tail tips were collected at wean (4 weeks) and digested using proteinase K overnight. Samples were cools to room temperature before protein precipitation solution containing 5M 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 once more and centrifuged. The ethanol was discarded leaving a pellet of DNA. The samples were left on a bench top covered with a paper towl to dry. Once no liquid is visible, DNA was re-suspended in 100ul of ddH20 and incubated at 65C for 5 mins. DNA concentrations and purity were measured using NanoDrop 2000 (Thermo Scientific). Samples for 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 genotyping.

**Genotyping with GigaMuga**

All 200 mice were fully genotyped for 143,259 SNPs by GeenSeek (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 heterozygosity, and N represents “no call”.

**RNA extraction and library prep**

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 one of which was sent for RNA-extraction.

* RNA extraction kit
* RNA quality QC
* cDNA synthesis and library prep
* Bcl2fastq tool to convert to fastq

**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. EMAS 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 in GBRS. GBRS was used to quantify multiway allele specificity taking into account DO generation and sex. 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.

**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 (reference).

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

Cross-comparisons between each whole-genome diplotype probability of 192 out of the 200 animals that had both RNA-Seq and GigaMUGA data were performed. Initial steps were taken using the DOQTL R package (reference) to create kinship probability plots of each set of genome probability construction, GigaMuga and RNA-Seq, to visualize and confirm heterogeneity of the F1 samples. The expectation is to see heterogeneity of kinship, as DO sire contributing to the F1 is genetically unique from one another.

**Albumin quantification**

**Glomerular filtration rate analysis**

**########################### Reference materials below**

(need Dan’s input here about genome probability construction)

Only 182 samples passed quality control (QC) from genome probability reconstruction.

Simultaneously we reconstructed the genome for the 192 mice we had RNA sequencing (RNA-seq) data for using the Expectation-Maximization algorithm for Allele Specific Expression (EMASE) software developed by the Churchill group (ref). EMASE algorithm takes diploid transcript tom alignment and estimates the expression abundance for each allele. The seemingly redundant effort was essential for our rigorous quality control for sample switching, deviation in allele frequency, and sample recover of the GigaMUGA data. Our comparison between the 182 mice with both GigaMUGA and RNA-seq data confirmed correct genotyping for 157 GigaMUGA and 24 samples that could not be confirmed. Upon further investigation revealed 11 of the 24 unconfirmed samples to be swapped with another Neogen Genomic customer. We were able to replace some missing samples using RNA-seq reconstruction, however the 8 samples that did not have RNA-seq data were removed from the study, as we could not verify their integrity. Through these stringent processes, we were able to confirm 192 quality samples for our analysis.

Albumin quantification

GFR calculation

GFR from Far2 paper

MME and glomerular filtration rate (GFR) were compared between knockout and wildtype mice at 6, 12, and 18 months of age (Figure 1C). At 6 months of age both groups had a low MME score (wildtype: 66%, knockout: 72%) and there was no significant difference between the two groups. However, at 12 months the MME score in the wildtype animals increased significantly, while this did not happen in the knockout animals (wildtype: 89%, knockout 68%, P=1.27x10-6). At 18 months, the knockout animals showed the same high amount (89%) of MME as the wildtype animals (88%). A significant difference (P=0.0118) was observed for GFR is at 6 months, with higher GFR (481±64 μl/min) in the knockout animals than the wildtype animals (381±79 μl/min) (Figure 1D).Therefore, our data shows that deletion of *Far2* leads to a delay in MME and an improvement of renal function at a young age.