Supplemental Information

Genetic background is a key determinant of glomerular extracellular matrix composition and organization

Michael J Randles,[1,2] Adrian S Woolf,[2] Jennifer L Huang,[3] Adam Byron,[4,] Jonathan D Humphries,[1] Karen L Price,[3] Maria Kolatsi-Joannou,[3] Sophie Collinson,[1] Thomas Denny,[1,2] David Knight,[1] Aleksandr Mironov,[1] Toby Starborg,[1] Ron Korstanje,[5] Martin J Humphries[1], David A Long[3] and Rachel Lennon[1,2].

[1]Wellcome Trust Centre for Cell-Matrix Research, Faculty of Life Sciences, University of Manchester, Manchester, M13 9PT, UK; [2]Institute of Human Development, Faculty of Medical and Human Sciences, University of Manchester, Manchester, M13 9PT UK; [3]Developmental Biology and Cancer Programme, Institute of Child Health, University College London, London, WC1N 1EH, UK; [4]Edinburgh Cancer Research UK Centre, Institute of Genetics and Molecular Medicine, University of Edinburgh, Edinburgh, EH4 2XR UK; [5]The Jackson Laboratory, Bar Harbor, Maine, ME 04609 USA.

Corresponding author:

Dr Rachel Lennon

Wellcome Trust Centre for Cell-Matrix Research, Michael Smith Building,

University of Manchester, Manchester M13 9PT, UK

Phone: 0044 (0) 161 2755498 Fax: 0044 (0) 161 2755082

Email: Rachel.Lennon@manchester.ac.uk

Table of Contents

Supplemental methods

Global microarray

Eighteen-week-old male and female B6 and FVB mice (n = 3 in each group) were anesthetized and perfused with $1x10^8$ Dynabeads (Invitrogen, Paisley, UK) through the left ventricle of the heart.[1] Kidneys were removed, decapsulated, minced, and digested,[1] and the glomeruli containing Dynabeads were gathered by a magnetic particle concentrator and RNA prepared using RNeasy kit (Qiagen, Crawley, UK). RNA quality was assessed on the Bioanalyser 2100 (Agilent Technologies, Palo Alto, CA); subsequent cDNA and cRNA synthesis was performed and hybridized to mouse MOE430 2.0 GeneChips. Signal values for transcripts on the Affymetrix array were calculated using the MAS 5.0 algorithm to generate .chp files and these were exported to GeneSpring 9.0 (Agilent Technologies, Wokingham, Berkshire, UK) for further analysis. The MAS 5.0–generated values were log2 transformed, normalized to the median within each array (to control for array loading), and these values were then baseline transformed to the median value of each transcript. Transcripts were filtered to exclude genes where expression did not reach a threshold value for reliable detection (based on the relaxed Affymetrix MAS 5.0 probability of detection; Pp0.1) in at least 1 of the 12 arrays assessed. To determine genes modified by strain and sex, a two-way analysis of variance analysis was performed. Transcripts that were significantly altered by either strain or sex ($P < 0.01$ after applying the Benjamini and Hochberg false discovery multiple testing correction) were used in subsequent analysis. Genes encoding extracellular matrix (ECM) proteins were selected from the microarray in the same manner as proteins were selected as ECM, as described below.

Western blotting

Following SDS-PAGE, resolved proteins were transferred to nitrocellulose membrane (Whatman, Maidstone, UK) as described previously.[2] Membranes were blocked with casein blocking buffer (Sigma-Aldrich, Poole, UK) and probed with primary antibodies diluted in

blocking buffer. Membranes were washed with Tris-buffered saline (10 mM Tris-HCl, pH 7.4, 150 mM NaCl) containing 0.05% (v/v) Tween 20 and incubated with species-specific fluorescent dye–conjugated secondary antibodies diluted in blocking buffer containing 0.05% (v/v) Tween 20. Membranes were washed in the dark, then scanned using the Odyssey infrared (IR) imaging system (LI-COR Biosciences, Cambridge, UK) to visualize bound antibodies.

qRT-PCR

Independent glomerular extracts were obtained from 18-week-old male and female B6 and FVB/N mice (*n*=3-4 in each group) for qPCR. cDNA was prepared and used to examine *Ntn4, Fgf2, Tnc, Col1a1, Mep1a and Mep1b* using previously described methods[3] with *Hprt* as a housekeeping gene. All measurements were performed in duplicate; primer details are available on request. Fold changes in expression are expressed relative to B6 female mice where average expression was given an arbitrary value of 1.

Functional annotation and enrichment analysis

Proteins identified in at least two of the three biological replicates were included for further analysis. GO annotations were downloaded using the online resource DAVID (Huang *et al*, 2009a)(Huang *et al*, 2009b). The GO cellular compartment annotation chart (GOTERM_CC_FAT) was selected, and proteins annotated in the extracellular region cluster were further cross-referenced with the mouse matrisome project,[4] and annotated as extracellular region. These ECM proteins were manually further divided into basement membrane, other structural and ECM-associated proteins. Proteins with no evidence of extracellular localization, were removed from the ECM dataset. This selection enabled the definition of the B6 female, B6 male, FVB female and FVB male glomerular ECM proteomes, as presented in Table S2.

Principal component analysis

Principal component analysis was performed using the commercial software package MATLAB (version 7.14.0).

Hierarchical clustering analysis

Z-transformed mean normalized intensities or normalized spectral counts were used for hierarchical clustering of microarray data and proteomic data respectively. Agglomerative hierarchical clustering was performed using MultiExperiment Viewer (version 4.8.1).[5] Protein hits were hierarchically clustered on the basis of Euclidean distance, and distances between hits were computed using a complete-linkage matrix. Clustering results were visualized using MultiExperiment Viewer (version 4.8.1).

Protein interaction network analysis

Protein interaction network analysis was performed using Cytoscape (version 2.8.1) (Shannon *et al*, 2003). Proteins identified in at least two biological replicates were mapped onto a merged human, mouse and rat interactome built from Protein Interaction Network Analysis platform *Homo sapiens* network (release date, 10 December 2012), *Mus musculus* network (release date, 10 December 2012) and the *Rattus norvegicus* network (release date, 10 December 2012),[6] the ECM interactions database MatrixDB (release date, 20 April 2012),[7] and a literature-curated database of integrin-based adhesion–associated proteins.[8] For networks where enrichment is presented Z transformed normalized spectral counts were used. Topological parameters were computed using the NetworkAnalyzer plug-in.[9]

GO enrichment and Ingenuity pathway analysis

To generate Figure 7A and Figure E6A, GO annotations were downloaded using the online resource DAVID[10] and loaded into the Cytoscape plug-in, Enrichment map v1.2,[11] with the following settings: P-value cut-off 0.005, Q-value cut-off 0.05 and overlap coefficient 0.6. The generated enrichment maps were clustered using the Markov cluster algorithm. For

Figures 7B and 7C datasets were uploaded to IPA (Ingenuity® Systems, www.ingenuity.com), the 'overlapping canonical pathways' tool was used to identify pathways predicted to be active in B6 or FVB. The data was then visualized using Cytoscape (version 2.8.1).

References

1. Takemoto, M, Asker, N, Gerhardt, H, Lundkvist, A, Johansson, BR, Saito, Y, Betsholtz, C: A new method for large scale isolation of kidney glomeruli from mice. *The American journal of pathology,* 161**:** 799-805, 2002.

2. Lennon, R, Byron, A, Humphries, JD, Randles, MJ, Carisey, A, Murphy, S, Knight, D, Brenchley, PE, Zent, R, Humphries, MJ: Global Analysis Reveals the Complexity of the Human Glomerular Extracellular Matrix. *J Am Soc Nephrol*, 2014.

3. Price, KL, Long, DA, Jina, N, Liapis, H, Hubank, M, Woolf, AS, Winyard, PJ: Microarray interrogation of human metanephric mesenchymal cells highlights potentially important molecules in vivo. *Physiological genomics,* 28**:** 193-202, 2007.

4. Naba, A, Clauser, KR, Hoersch, S, Liu, H, Carr, SA, Hynes, RO: The matrisome: in silico definition and in vivo characterization by proteomics of normal and tumor extracellular matrices. *Molecular & cellular proteomics : MCP,* 11**:** M111 014647, 2012.

5. Saeed, AI, Sharov, V, White, J, Li, J, Liang, W, Bhagabati, N, Braisted, J, Klapa, M, Currier, T, Thiagarajan, M, Sturn, A, Snuffin, M, Rezantsev, A, Popov, D, Ryltsov, A, Kostukovich, E, Borisovsky, I, Liu, Z, Vinsavich, A, Trush, V, Quackenbush, J: TM4: a free, open-source system for microarray data management and analysis. *BioTechniques,* 34**:** 374-378, 2003.

6. Wu, J, Vallenius, T, Ovaska, K, Westermarck, J, Makela, TP, Hautaniemi, S: Integrated network analysis platform for protein-protein interactions. *Nature methods,* 6**:** 75-77, 2009.

7. Chautard, E, Ballut, L, Thierry-Mieg, N, Ricard-Blum, S: MatrixDB, a database focused on extracellular protein-protein and protein-carbohydrate interactions. *Bioinformatics,* 25**:** 690-691, 2009.

8. Zaidel-Bar, R, Geiger, B: The switchable integrin adhesome. *Journal of cell science,* 123**:** 1385-1388, 2010.

9. Assenov, Y, Ramirez, F, Schelhorn, SE, Lengauer, T, Albrecht, M: Computing topological parameters of biological networks. *Bioinformatics,* 24**:** 282-284, 2008.

10. Huang da, W, Sherman, BT, Lempicki, RA: Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nature protocols,* 4**:** 44-57, 2009.

11. Merico, D, Isserlin, R, Stueker, O, Emili, A, Bader, GD: Enrichment map: a network-based method for gene-set enrichment visualization and interpretation. *PLoS One,* 5**:** e13984, 2010.

FIGURE S1

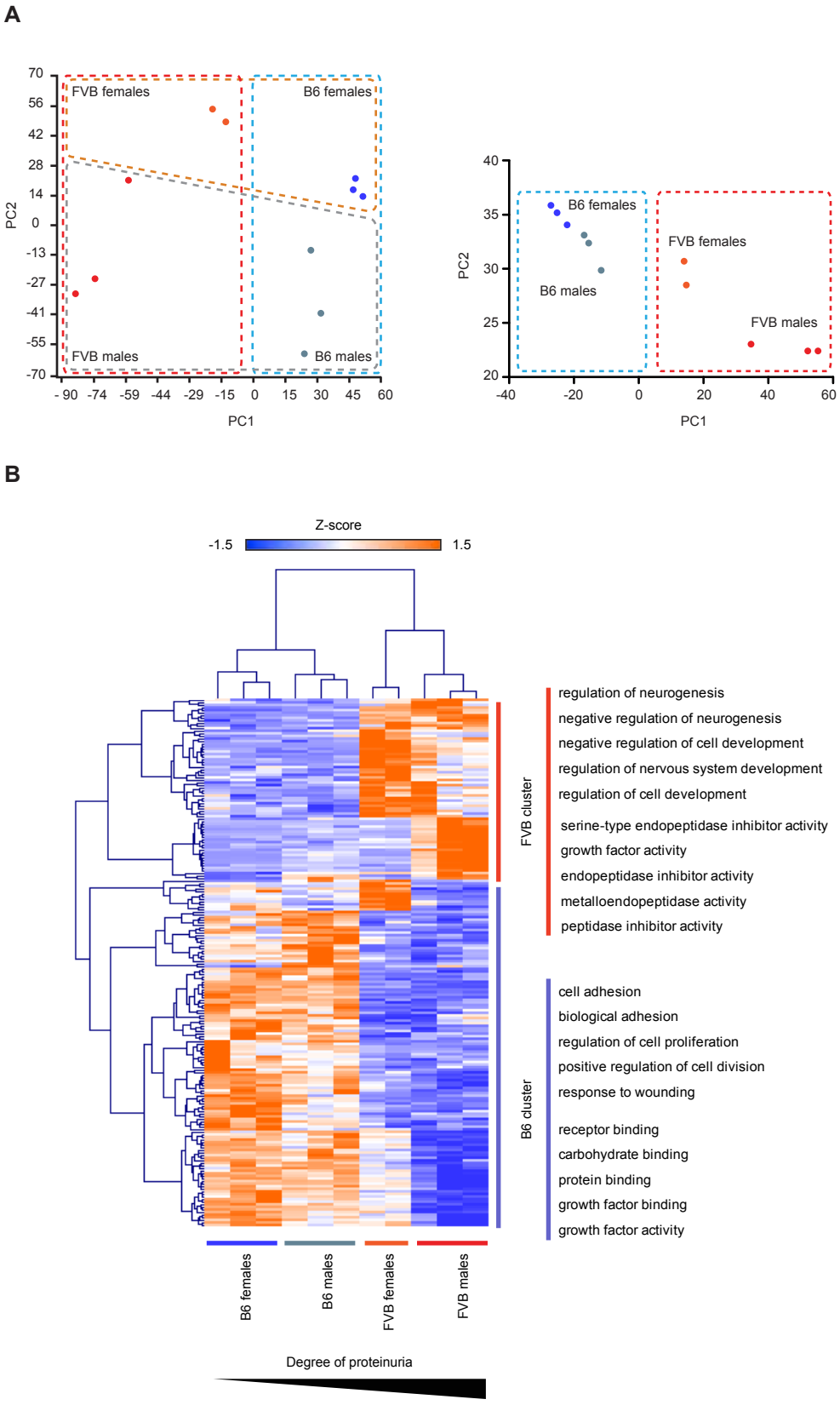


Figure S1. Transcript analysis of glomeruli. (A) Principal component analysis of all transcripts from whole glomerular microarray (left panel) and ECM transcripts (right panel). B6 females, blue dots; B6 males, grey dots; FVB females, orange dots; FVB males, red dots. Orthogonality between the principal components is highlighted by dashed quadrilaterals (B) Unsupervised hierarchical clustering of ECM transcripts. The heat map displays Z-transformed normalized intensities, and the associated dendrogram displays clustering on the basis of Euclidean distance. The top five enriched GO biological processes and molecular function terms are annotated onto the B6 and FVB clusters.

FIGURE S2

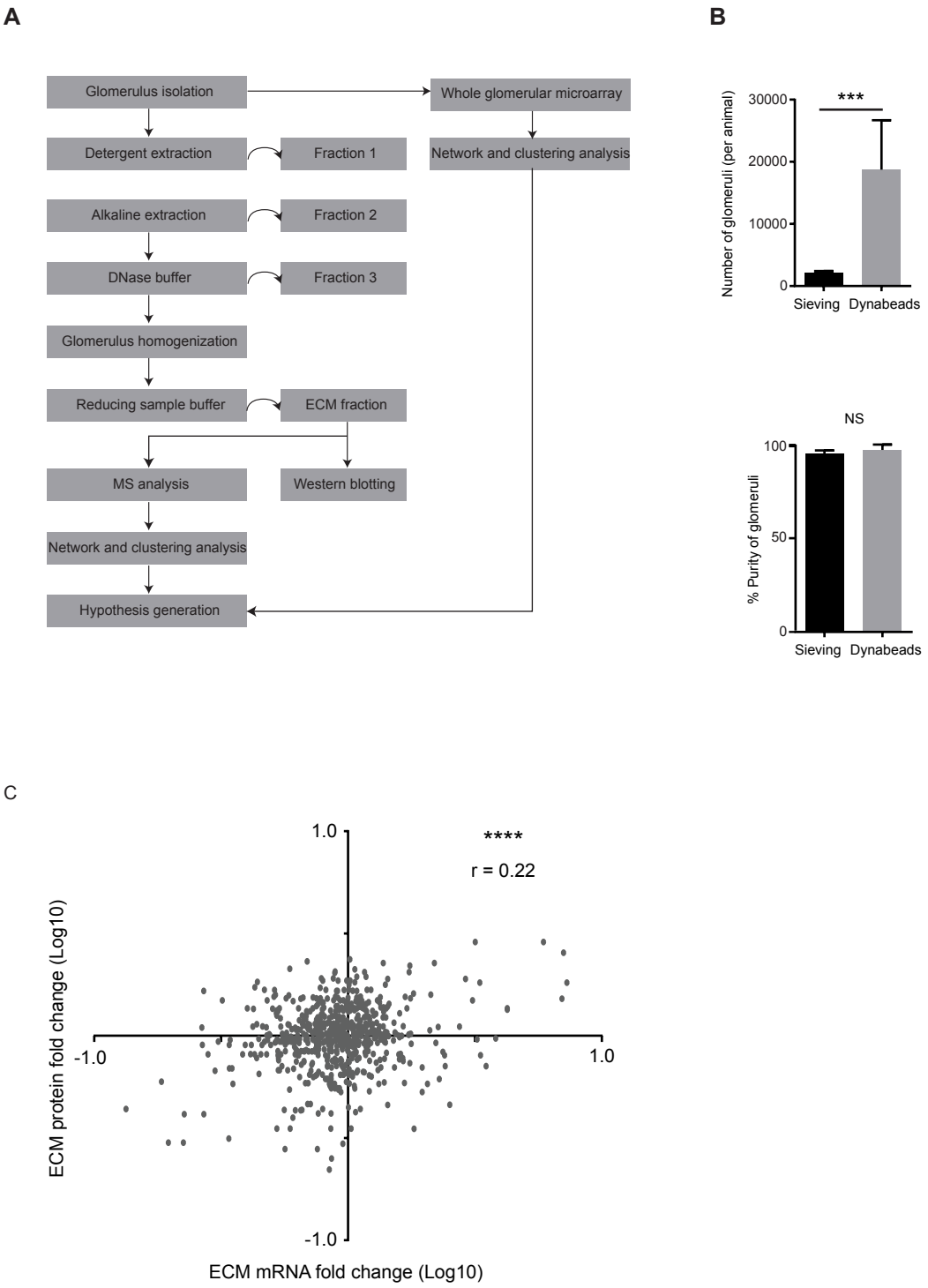


Figure S2. Workflow of glomerular ECM enrichment. (A) A proteomic workflow for the isolation of enriched glomerular ECM by fractionation (see methods for further details). (B) Mouse glomeruli were isolated by Dynabead perfusion and magnetic concentration. The Dynabead approach yielded significantly greater numbers of glomeruli, *** $p < 0.001$, (18744 ± 1590, n = 5) than glomerular isolation by differential sieving (2148 ± 238, n = 5) top panel. Glomerular purity was not significantly different between Dynabead isolation (97.00 ± 0.59, n = 5) and sieving isolation (95.02 ± 0.57, n = 5) methods, bottom panel. (C) Correlation of mRNA detected by microarray and protein detected by mass spectrometry. Significance was determined by two-tailed Pearson test. **** $p< 0.0001$.

FIGURE S3

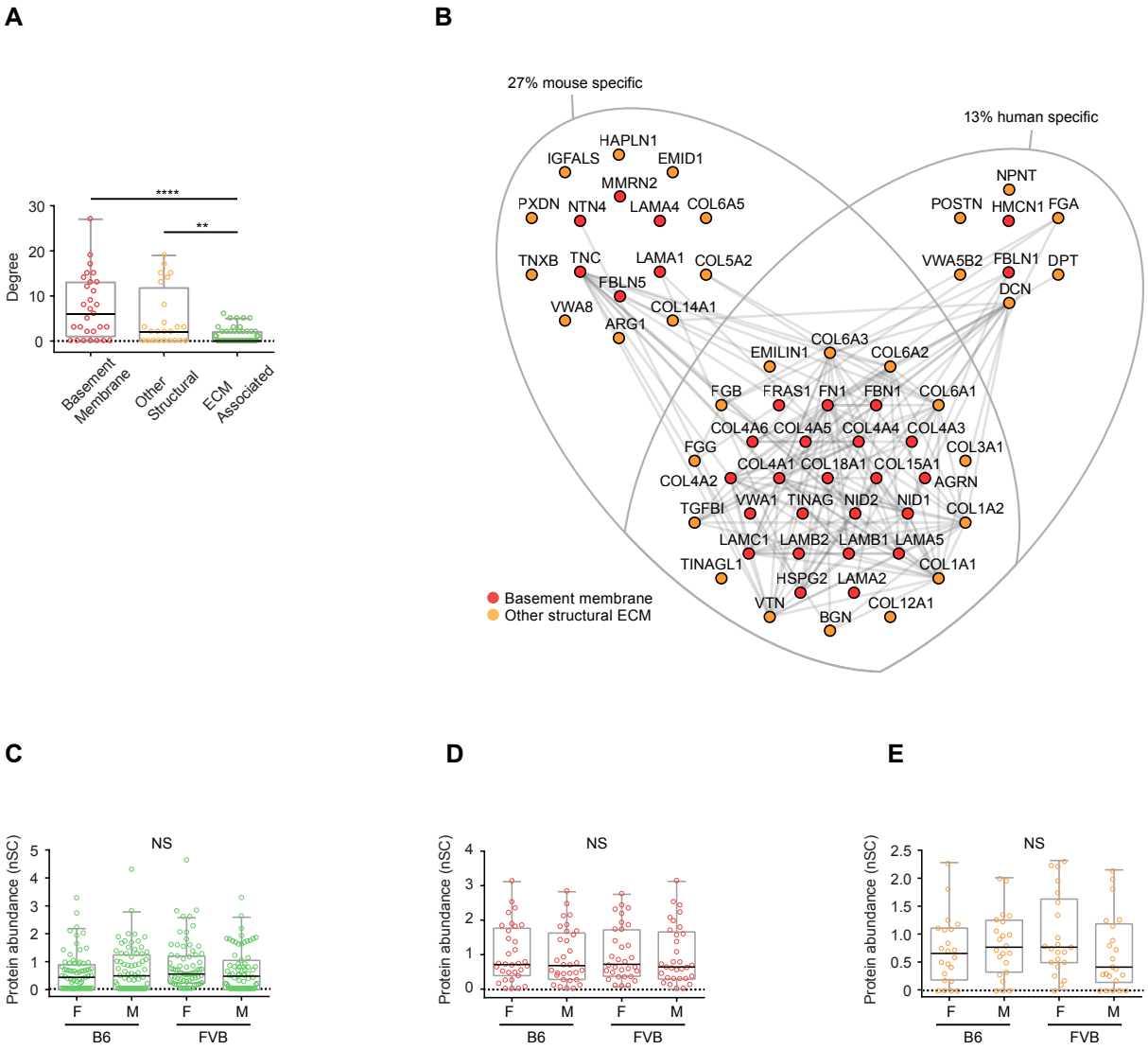
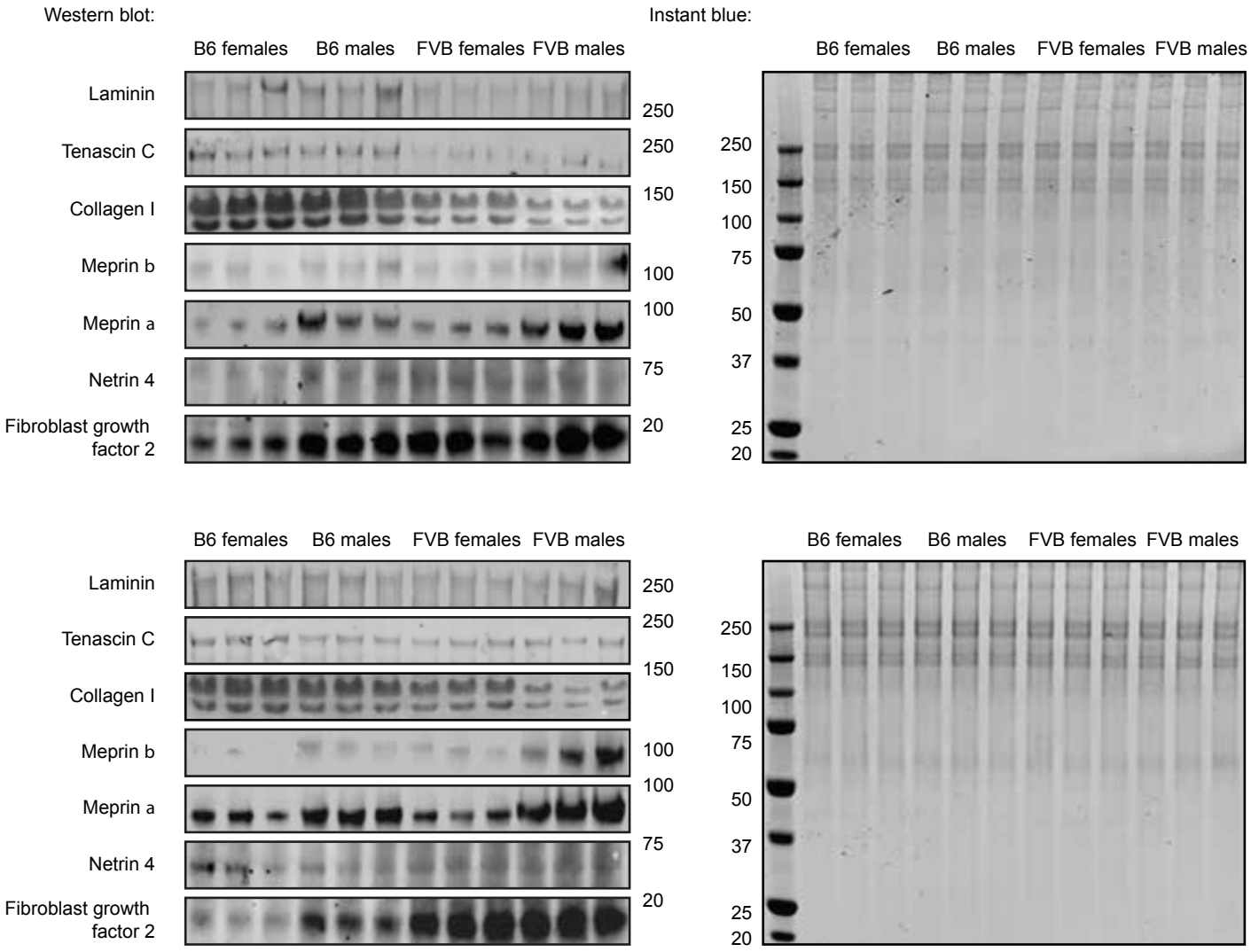


Figure S3. Glomerular ECM protein interaction topological network analysis.(A) Topological analysis of the number of protein-protein interactions (degree) made by the distinct categories of ECM proteins. ****, $p< 0.0001$ **, $p< 0.005$; NS, $p\geq 0.05$. (B) Overlap of mouse and human basement membrane and other structural ECM proteins. Protein interaction network constructed from enriched glomerular ECM proteins identified by MS. Nodes (circles) represent proteins and edges (grey lines) represent reported protein–protein interactions. ECM proteins were categorized as basement membrane or other structural ECM and were coloured accordingly. Nodes are labelled with gene names for clarity. Protein abundance differences in (C) basement membrane, (D) other structural ECM and (E) ECM associated proteins in each of the groups of animals.

FIGURE S4



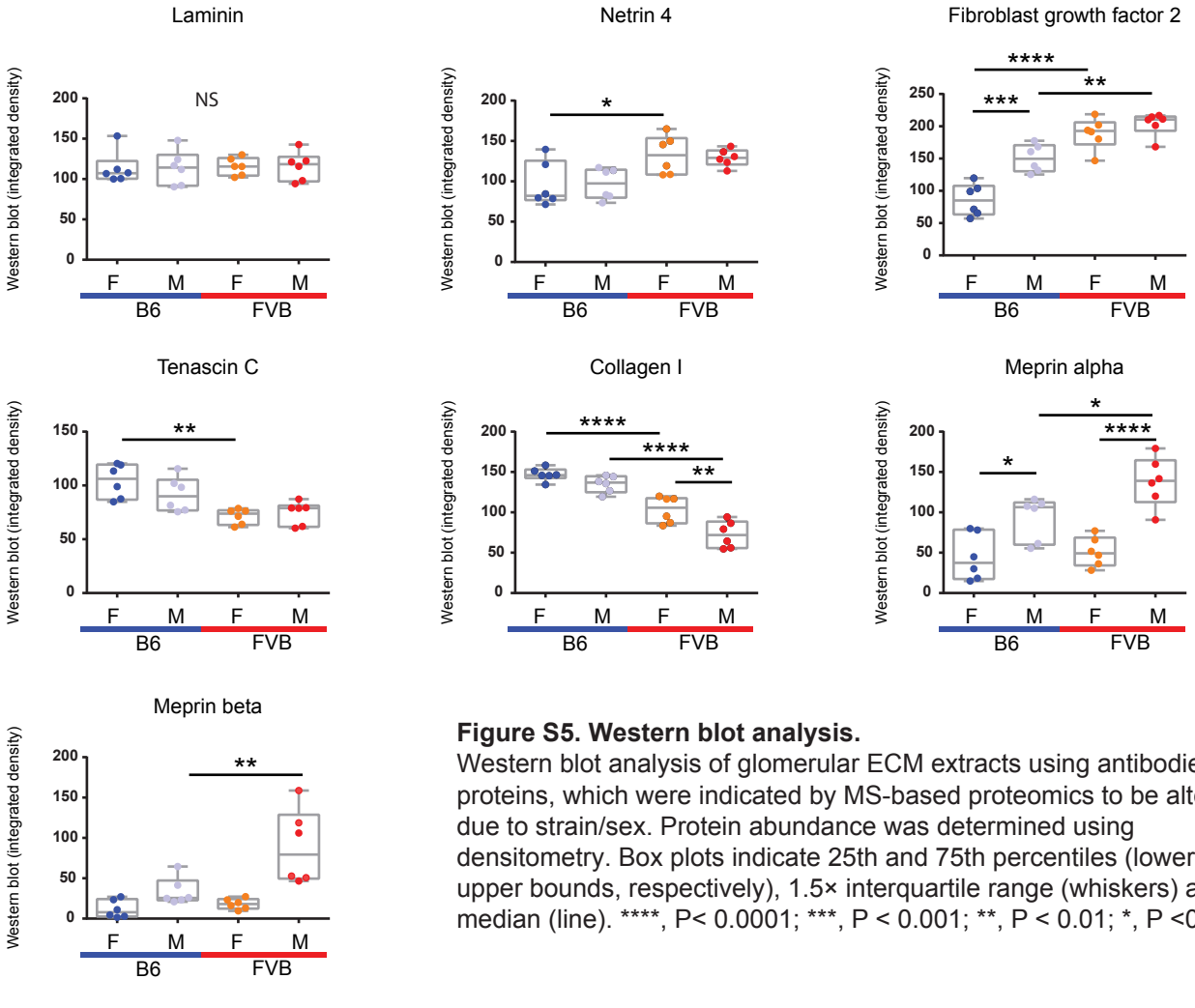


Figure S5. Western blot analysis.
Western blot analysis of glomerular ECM extracts using antibodies to proteins, which were indicated by MS-based proteomics to be altered due to strain/sex. Protein abundance was determined using densitometry. Box plots indicate 25th and 75th percentiles (lower and upper bounds, respectively), 1.5× interquartile range (whiskers) and median (line). ****, P< 0.0001; ***, P < 0.001; **, P < 0.01; *, P <0.05.

FIGURE S5

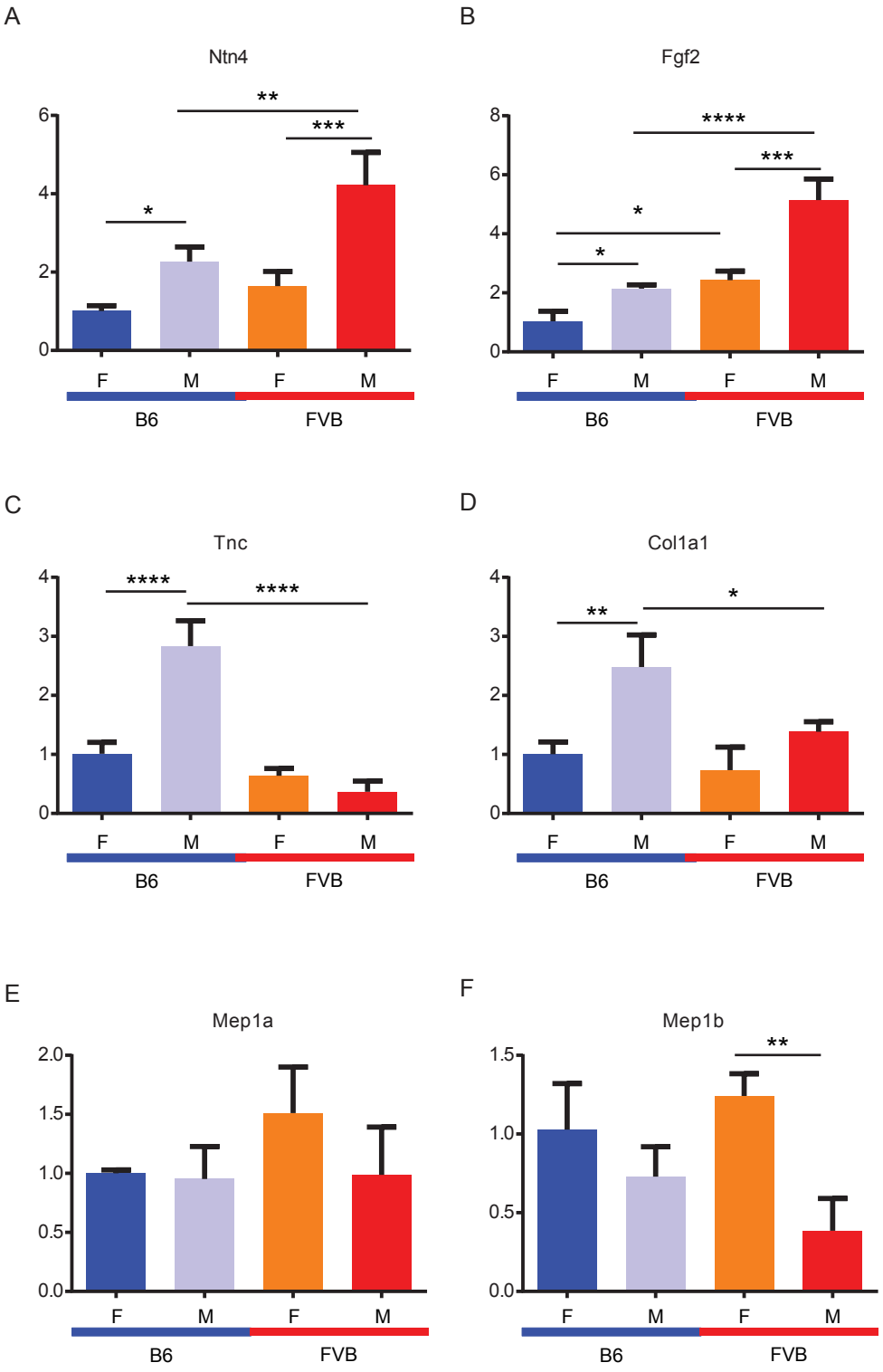


Figure S4. Analysis of glomerular ECM mRNA expression. Analysis glomerular ECM mRNA expression by qPCR. ****, P< 0.0001; ***, P < 0.001; **, P < 0.01; *, P < 0.05.

FIGURE S6

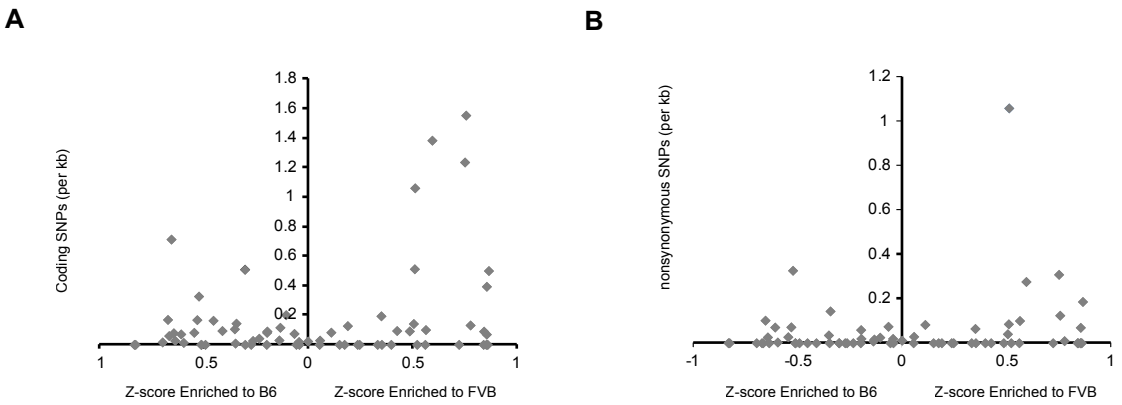


Figure S6. Variations in genetic sequence of glomerular ECM proteome. Genetic sequence variation identified between B6 and FVB glomerular ECM genes, as a function of relative protein abundance in FVB and B6 glomerular ECM (A) Total Coding SNPs (C) nonsynonymous SNPs/gene length. We identified 58 synonymous SNPs in 13 ECM proteins 1.5 fold enriched to B6: Tnc (1), Col1a1 (1), Col1a2 (1), Col18a1 (1), Angptl6 (1), Col3a1 (2), Itih2 (2), Arg1 (2), Col4a6 (3), Vwa8 (4), Col12a1 (5), Ren1 (6), Fras1 (29). We identified 16 non synonymous SNPs in 10 proteins enriched to B6: Col3a1 (1), Itih2 (1), Col4a6 (1), Ren1 (1), Itih1 (1), Apoe (1), Itih5 (2), Vwa8 (4), Mep1b (4), Fras1 (8). We identified 31 non synonymous SNPs in 7 ECM proteins that were greater than 1.5 fold enriched to FVB: Igfals (1), Ntn4 (1), Col6a5 (19), Pzp (3), Tfip11 (1), Apoa1 (2) and Cfhr2 (4). We identified 58 synonymous SNPs in 7 ECM proteins 1.5 fold enriched to FVB: Cpn2 (2), Calr (2), Igfals (3), Ntn4 (13), Col6a5 (32), Plscr1 (2) and Tfip11 (5)

FIGURE S7

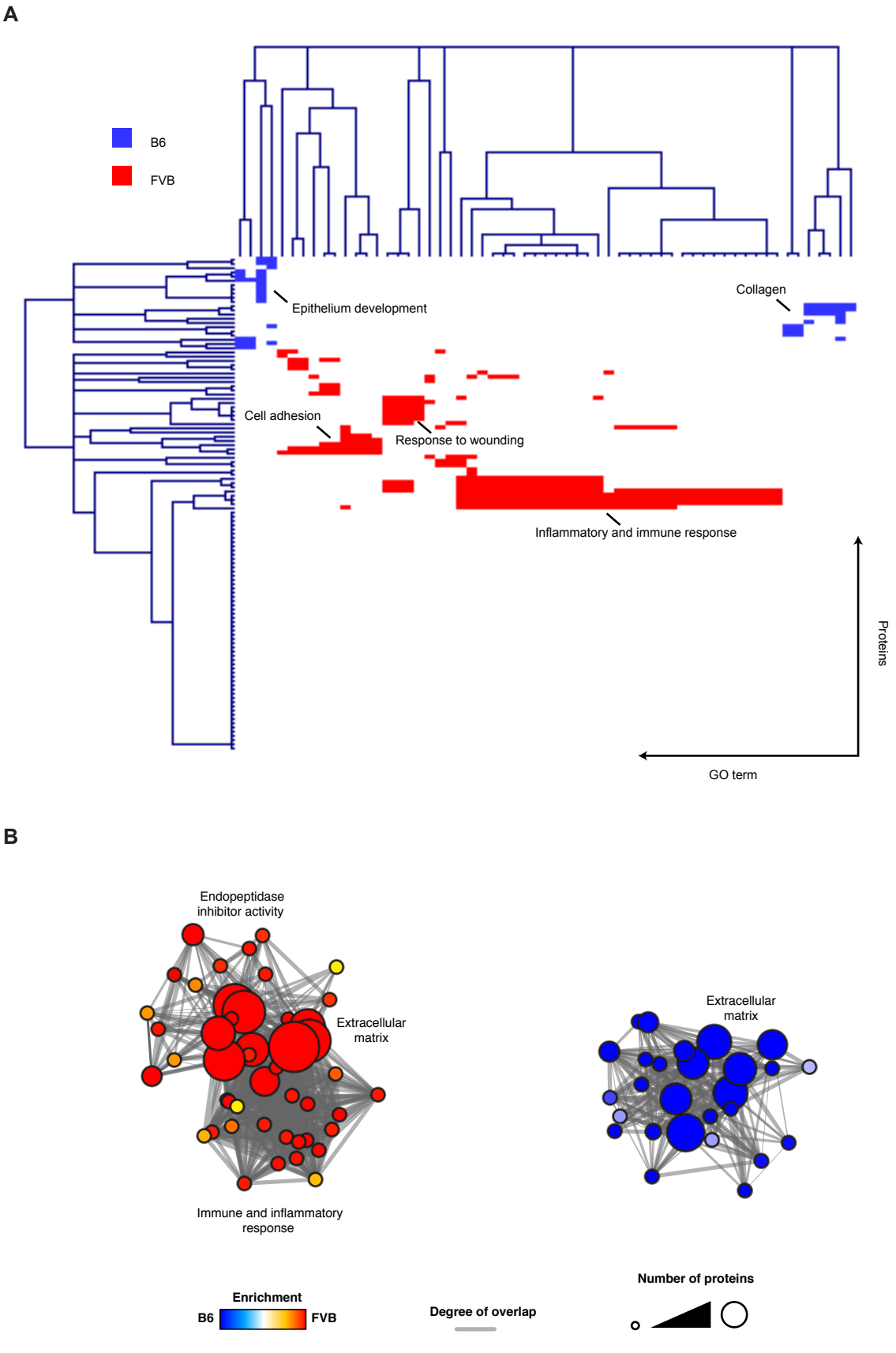


Figure S7. GO enrichment analysis of proteomic datasets. (A) Unsupervised hierarchical clustering of GO terms against proteins, (B) GO enrichment map of terms significantly enriched in FVB or B6 glomerular ECM. Nodes (circles) represent GO terms significantly enriched in the proteomic data-set, edges (lines) represent overlap of proteins within terms. Node size relates to the number of proteins which are allocated to a given term and node colour relates to the enrichment to either FVB (red) or B6 (Blue).