1 Supplementary Data and Methods:

1.1 Supplementary Methods:

PKD ELV biomarker discovery by label free mass spectrometry: Coomassie–stained gel sections, in 1 mm cubes, were de-stained with 45% isopropanol in 150 mM Tris, pH 8.2, reduced with DTT for 30 min at 60°C, and alkylated with iodoacetamide for 30 min at room temperature in the dark. Dehydrated gel pieces were digested with 25μl of 4 ng/μl porcine trypsin (Promega, Sequencing grade), overnight at 37°C. Peptides were first extracted with 50% acetonitrile in 2% TFA for 30 min, then with acetonitrile for 30 min. The two extracts were combined, dried, and stored at -80°C until LC–MS/MS analysis. nLC–MS/MS analysis for label–free differential MS were run on a LTQ–Orbitrap Classic (Thermo Fisher Scientific, San Jose, CA) interfaced with an Eksigent nanoLC–2D liquid chromatograph (Redwood City, CA). Nano-scale liquid chromatography (nLC) used a 25cm by 75 m i.d. packed spray tip (PicoFrit, New Objective, Woburn, MA) packed with Michrom Magic C18AQ, 3μm, 200 Å(Michrom Bioresources, Auburn, CA). A 75 minute LC method (3–40%B in 50 min.) was used in conjunction with data-dependent acquisition of MS2 spectra. Survey scans at 60,000 resolving power (AGC=1E6), using lock masses at m/z 445.12002 and 371.10123, were followed by LTQ–CID spectra of the top 3 doubly or triply charged ions using gas phase fractionation where the precursor m/z range was divided into six regions to increase MS/MS depth of coverage. Six Xcalibur instrument methods, varying only in the precursor range used for selecting MS2 experiments, were used for collecting data across 46 gel lanes, so that each of the six precursor m/z ranges were used at least seven times within each gel section. Validated peptide identities from each precursor m/z range were mapped to MS1 data across all samples via m/z and LC retention time (RT) parameters using the Rosetta Elucidator [1] software package (ver. 3.3, Ceiba Solutions, Boston, MA). Elucidator software was used to extract LC/MS peaks from MS1 data, align chromatograms of data files within each gel section, and perform MASCOT searches against the Swiss–Prot 2011 ver11 database with isoforms, filtered for human and *E. coli* proteins plus an appended decoy database of the reversed entries for each protein. Peptide-spectrum matches (PSM) were validated using the Peptide Prophet [1] and Protein Prophet [2] algorithms implemented 11 in Elucidator that used PSMs against the decoy database entries to model the distributions of correct and incorrect PSMs. Data tables with protein abundance for each sample, for each gel section, were exported for evaluation of differences. Where proteins were detected in multiple gel sections, the gel section with highest abundance was used as representative for that protein.

1. Keller, A., Nesvizhskii, A.I., Kolker, E. Aebersold, R. Empirical statistical model to estimate the accuracy of peptide identifications made by MS/MS and database search. Anal Chem 74, 5383–5392

(2002). 2. Nesvizhskii, A.I., Keller, A., Kolker, E. Aebersold, R. A statistical model for identifying proteins by tandem mass spectrometry. Anal Chem 75, 4646–4658 (2003).

1.2 Supplemental Database 1:

Database of proteins and corrected intensities: Proteins sorted by gel slice then uniprot ID. Protein wise intensities corrected using fastlo normalization, a nonlinear normalization algorithm similar to cyclic Loess. Columns H–T are corrected intensities from individual PKD1 individuals (Red) with HtTKV and mutations, columns U–AL are from normal individuals (Light Blue). Columns AM, AN and AO are the PKD1 mean intensity and the sd and cv of the intensities respectively (Red). Columns AP, AQ and AR are the Normal mean intensity and the sd and cv of the intensities respectively (Blue). AS is the uncorrected p–value for Welchs *t*–test and AT is the Bonferroni adjusted p–value for that gel slice, column AU is the q–value derived from the slice wise unadjusted p–value and column AR is the PKD1 mean intensity divided by the normal intensity. The protein intensity data was also transformed to the natural log of the value and the Welch test AX, the Bonferroni corrected p–values AY and q–values AZ were computed

List of Figures

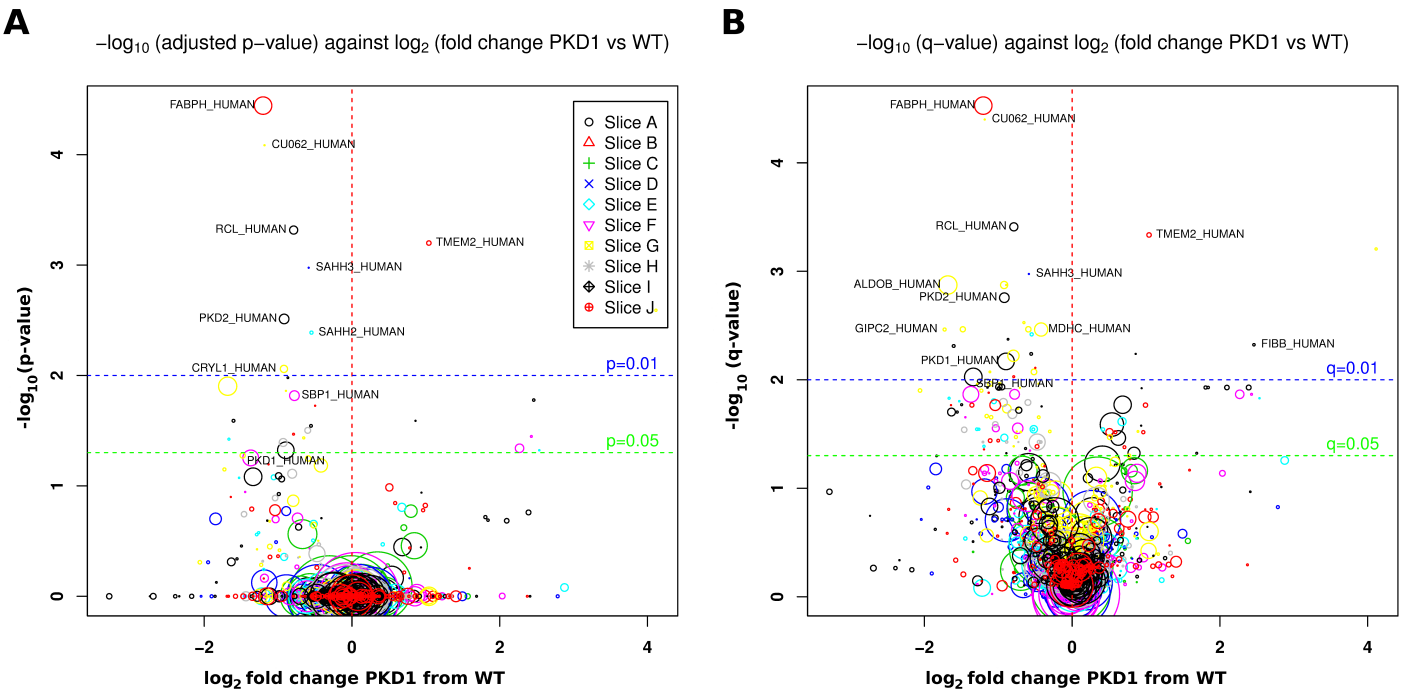

Figure 1: **Supplemental Figure 1: Volcano plots:** We $\log_e$ transformed protein intensity data in order to suppress the effect of outliers and to approximate a Guassian distribution. Proteins found at significantly different levels in normal and PKD1 individuals, color maps to the gel slice in which they were found, see key. X-axis $\log_2$ fold change between normal and PKD1 individuals, Y-axis$\log_{10}$ of A) p-value, B) q-value. Diameter of symbols varies as the squareroot of the mean ion intensity in the normal cohort. This shows that the data are robust to log transformation and that the same proteins found in the initial analysis were still significant in the $\log_e$ transformed data.

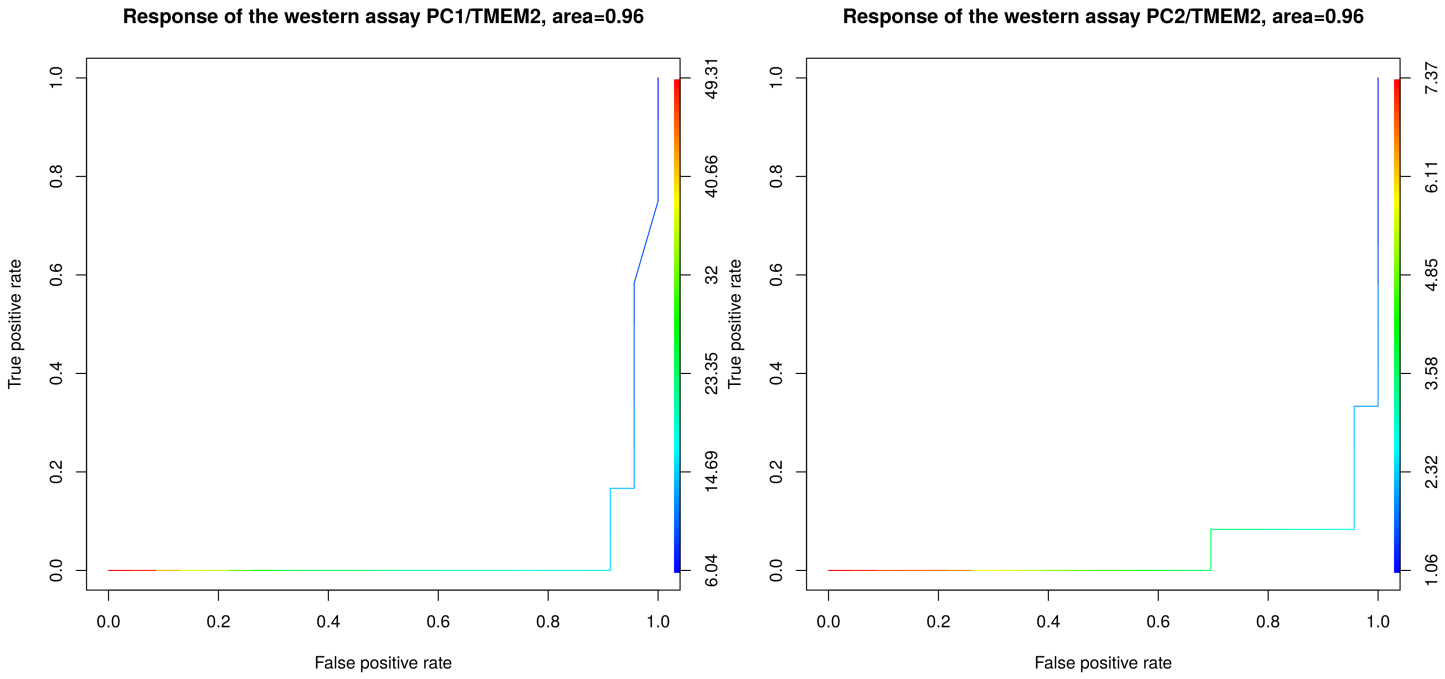

Figure 2: **Supplemental Figure 2:** ROC analysis: ROC curves for the PC1/TMEM2 and PC2/TMEM2 western assays, showing that for individuals with PKD1 mutations, they perform almost identically with a under curve area of 0.96.True positive=PKD1, false positive=normal n=35.

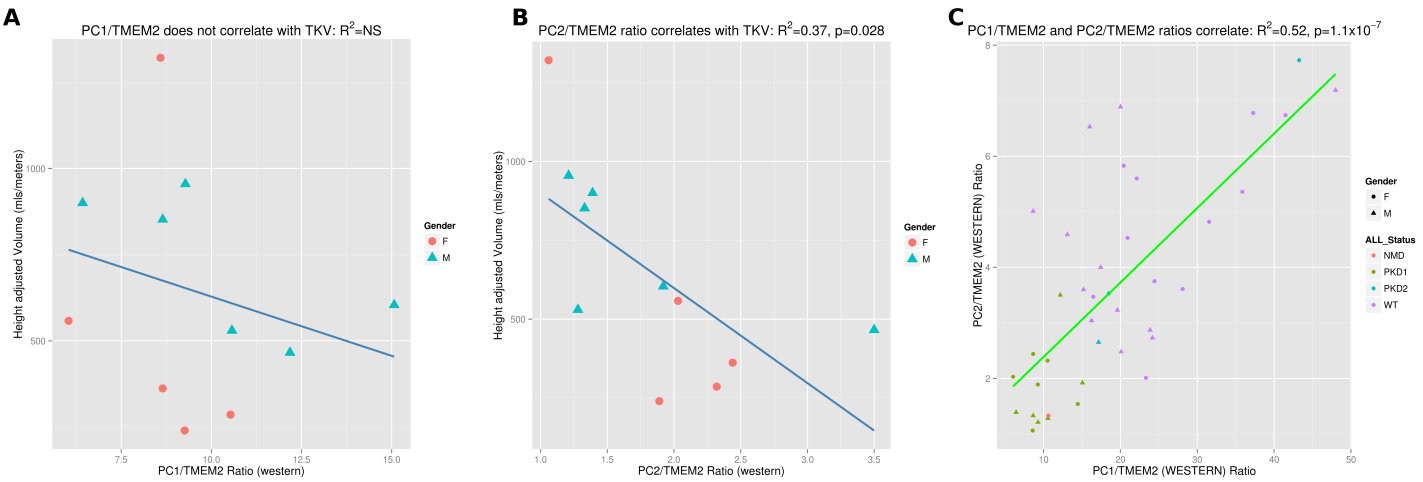

Figure 3: **Supplemental Figure 3:** A) HtTKV plotted against the western derived PC1/TMEM2 ratio R^2 NS, or B) the western derived PC2/TMEM2 ratio R^2 =0.37 p=0.028. C) All western derived PC2/TMEM2 ratios against all PC1/TMEM2 ratios, for normal, PKD1, PKD2 and NMD individuals, R^2 =0.52, p=1.1x10^{-7}.