Step-by-step guide for KidneyGPS GPS-tab

Introduction: The GPS tab of KidneyGPS can provide customized lists of 'prioritized' genes. It comprises three filter sections that are applied to the full and comprehensive GPS table (5,906 genes for 594 signals), which is shown by default: the first section, called "Signal-Filtering", enables to restrict the full GPS-table to signals with specific properties, such as statistical fine-mapping support or Cys/BUN-validation of the signal; the second section, called "Variant-to-Gene mapping", enables to further restrict to genes that are mapped by variants with specific properties, such as protein-relevance or regulatory variants; and the third section, called "Gene-to-phenotype mapping", enables to further restrict to genes with known kidney phenotype in mice, human or to druggable genes.

Three examples: how to

(1) generate a list of genes that are mapped to a signal with high statistical fine-mapping support for pinpointing the causal variant (credible set contains ≤5 variants or contains one variant with PPA≥50%), where the variant is protein-altering (using variant-effect-predictor; VEP) or an eQTL/sQTL in kidney tissue,

(2) further restrict this list to genes with evidence of a kidney phenotype in human (e.g. OMIM) or with evidence of being druggable for any disease, and

(3) further restrict to genes where the genetic effect is stronger in individuals with diabetes compared to individuals without diabetes.

Step-by-step instruction **example 1**:
To obtain the first gene-list, in the "Signal-filtering" section, we set "The signal's credible set contains a variant with posterior probability of association (PPA) of: 50%" and "signal has a small credible set (1-5 variants)":

1. Signal-filter:

Restrict to signals with any of the following properties regarding the strength of statistical support:

☑ The signal's credible set contains a variant with posterior probability of association (PPA) of:

    ○ PPA >99% (most likely causal variants)

    ◉ PPA >50% (variants with high probability of being causal)

    ○ PPA >10% (excluding variants with little probability of being causal)

☑ signal has a small credible set (1-5 variants)

☐ no filter (show all signals)

➔ This will restrict the full GPS-table to genes located in signals satisfying at least one of those criteria.

In the second section "Variant-to-gene mapping", we must then select all blue options and the orange options "eQTL in glomerular tissue", "eQTL in tubulo-interstitial tissue"," eQTL in kidney cortex tissue", and "sQTL in kidney cortex tissue":

2. Variant-to-gene mapping:

Restrict to genes which are mapped to a credible set variant by any of the following properties:

Gene contains a credible set variant that is protein-relevant

☑ stop-gained, stop-lost, non-synonymus

☑ canonical splice, noncoding change, synonymous, splice site

☑ other functional consequence with high predicted deleteriousness (CADD-Phred ≥ 15)

Gene maps to a credible set variant that modulates gene expression (eQTLs) or splicing (sQTLs), FDR < 5%:

☑ eQTL in glomerular tissue (NEPTUNE, or Sheng et al [Nat.Genet. 2021])

☑ eQTL in tubulo-interstitial tissue (NEPTUNE, or Sheng et al [Nat.Genet. 2021])

☑ eQTL in kidney cortex tissue (GTEx)

☐ eQTL in other tissue (GTEx)

☑ sQTL in kidney cortex tissue (GTEx)

☐ sQTL in other tissue (GTEx)

➔ This will further restrict our list of genes to genes mapped by a credible set variant that is protein-altering or by a credible set variant that is an eQTL or sQTL in kidney tissue.

If any blue or orange feature is chosen, you can further restrict the mapping based on the strength of statistical support of the mapped variant itself:

Restrict mapping to credible set variants with any of the following properties regarding strength of statistical support:
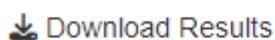
☑ Posterior probability of association (PPA):
  ○ PPA >99%
  ◉ PPA >50%
  ○ PPA >10%

☑ variant is contained in small credible set (1-5 variants)

☐ no filter (show all genes mapped by any credible set variant)

➔ The default is that the mapped variant must have a PPA > 50% or must be included in a small credible set, which fits with our request.

Finally, we click on "Go prioritize!" to generate our first gene-list:

The displayed GPS-Table is now restricted to genes which meet our applied criteria. To save this list, we can use the "Download Results" button, which will store the list as an Excel file. Below the download button, we can see that our list included 110 different genes from 86 independent eGFR signals:

Cave: filtering via column headers has no impact on the downloaded table

110 genes from 86 signals meet the applied filter criteria.

*Step-by-step instruction **example 2***:

To further restrict the list of genes to genes with evidence of a kidney phenotype in human (e.g., OMIM) and/or with evidence of being druggable for any disease we can use the "Gene-to-phenotype mapping" section. The second option of "Kidney phenotypes" is required to restrict to genes with evidence of a kidney phenotype in human. Additionally, both options of "Drug Information" must be selected to show genes that are drug targets of any disease:

3. Gene-to-phenotype mapping:

Restrict to genes with any of the following properties:

Kidney phenotypes
- ☑ Gene has known kidney phenotypes in mouse models (Mouse Genome Informatics, MGI)
- ☐ Gene is described to cause a human disease with kidney phenotype (online mendelian inheritance in man, OMIM, Groopman et al. [N.Engl.J.Med,2019], or Wopperer et al. [Kidney Int.,2022])

Drug information
- ☑ Gene is a known drug target for kidney diseases (Therapeutic Drug Database)
- ☑ Gene is a known drug target for any other diseases or with unspecific indication (Therapeutic Drug Database)

➔ After applying these three selections and again clicking "Go prioritize!", the GPS-Table view is further restricted to 21 genes from 20 signals.

Step-by-step instruction **example 3**:

To even further restrict the view of genes fulfilling the criteria from example 1 and 2, to genes mapping to eGFR signals that differ between individuals with/without diabetes, one can use the "Restrict further to: Signals with differential association in individuals with diabetes and those without diabetes" option in the "Signal-Filtering" section:

Restrict further to:

Signals located in a eGFRcys/BUN validated locus
○ yes ● no

Restrict further to:

Signals with differential association in individuals with diabetes vs. without diabetes
● yes ○ no

Signals with association with eGFR decline
○ yes ● no

After clicking "Go prioritize!", the GPS-Table view is restricted to *UMOD*, which is the only gene fulfilling all the exemplary criteria. This approach is recommended (over using filtering via the GPS table header) for download because the filtering options in the header line of the GPS-

table are not preserved in the downloaded data set (one would have to open the file in Excel and re-apply the filtering in Excel).