# The Kidney-Genetics Documentation

## Bernt Popp, Nina Rank, Constantin Wolff, Jan Halbritter

## 2023-06-22

## Contents

P	refac	${ m e}$
	Obj	ective
	Met	hods
	0.1	Results
	0.2	Conclusion
	0.3	Outlook
1	Ana	alyses result tables
	1.1	Main table: Merged analyses sources
	1.2	Result table: PanelApp
	1.3	Result table: Literature
	1.4	Result table: Diagnostic panels
	1.5	Result table: HPO in rare disease databases
	1.6	Result table: PubTator
<b>2</b>	Ana	alyses plots
	2.1	UpSet plot of merged analyses sources
	2.2	Bar plot of PanelApp results
	2.3	Bar plot of Literature results
	2.4	Bar plot of Diagnostic panels results
	2.5	Bar plot of HPO in rare disease databases results
	2.6	Bar plot of PubTator results

## Preface

This documentation is intended to describe the Kidney-Genetics<sup>1</sup> project.

## Objective

How can we address the lack of a unified and standardized database of kidney disease-associated genes, which hampers diagnosis, treatment, and research comparability in the field of kidney diseases?

### Methods

To create a comprehensive and standardized database of kidney-related genes, we employed the following methods:

1. Utilized data from Genomics England and Australia PanelApp.

 $<sup>^{1} \</sup>rm https://github.com/halbritter-lab/kidney-genetics$ 

- 2. Conducted a comprehensive literature review of published gene lists.
- 3. Collected information from clinical diagnostic panels for kidney disease.
- 4. Performed a Human Phenotype Ontology (HPO)-based search in rare disease databases (OMIM, Orphanet).
- 5. Employed a PubTator API-based automated literature extraction from PubMed.

We also developed an evidence-scoring system to differentiate highly confirmed disease genes from candidate genes.

## 0.1 Results

The "Kidney-Genetics" database currently includes detailed information on 2,906 kidney-associated genes. Notably, 439 genes (15.1%) are present in three or more of the analyzed information sources, indicating high confidence and their potential for diagnostic use.

To ensure currency, Kidney-Genetics will be regularly and automatically updated. We will also provide phenotypic and functional clustering results to facilitate gene grouping.

### 0.2Conclusion

Kidney-Genetics is a comprehensive and freely accessible database that researchers can use to analyze genomic data related to kidney diseases. The database is regularly updated through a standardized pipeline and an automated system, ensuring it remains up-to-date with the latest advancements in kidney research and diagnostics.

By utilizing Kidney-Genetics, clinicians and researchers can enhance their understanding of the genetic aspects of kidney disorders.

#### 0.3Outlook

Future goals include manual curation and the assignment of diagnostic genes to specific nephrological disease groups, such as syndromic vs. isolated, adult- vs. pediatric-onset, and cystic vs. nephrotic, among others.

# Analyses result tables 1 Main table: Merged analyses sources This table shows the merged results of all analyses files as a wide table with summarized information. approved\_symbol hgnc\_id evidence\_count list\_count 01\_PanelApp 02\_Literature 03\_DiagnosticPanels 04\_HPO 05\_PubTator All All All All

### 1.2Result table: PanelApp

This table shows results of the first analysis searching kidney disease associated genes from the PanelApp project in the UK and Australia.

approved_symbol	hgnc_id	gene_name_	_reported s	ource s	source_count	source_evidence
All	All	All	Α	All [	All	All

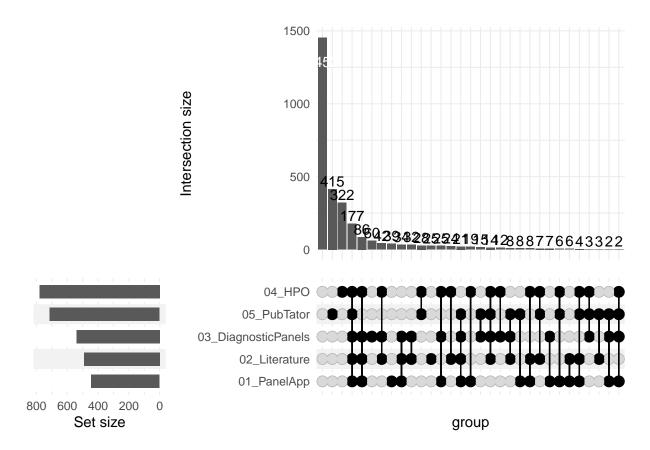
### Result table: Literature 1.3

This table shows results of the second analysis searching kidney disease associated genes from various publications.

All	All	All	All	All	All	
1.4 Resu	ılt table: Dia	gnostic panel	$\mathbf{s}$			
	ows results of the dney disease.	third analysis sear	ching kidney disea	se associated	genes from clinical diag	gnostic
approved_	_symbol hgnc_	id gene_name	_reported sour	ce source_	_count source_evid	lence
All	All	All	All	All	All	
		v	s searching kidney e disease database		ociated genes from a Frphanet).	Iuman
This table s	hows results of th	ne fourth analysis	searching kidney	disease asso	ociated genes from a H	Iuman
Phenotype (	Ontology (HPO)-b	ased search in rare	e disease database	es (OMIM, O	<u> </u>	
Phenotype (	Ontology (HPO)-b	ased search in rare	e disease database	es (OMIM, O	rphanet).	
Phenotype ( approved_ All	Ontology (HPO)-b	ased search in rar- id gene_name_ All	e disease database _reported sour	es (OMIM, O	rphanet). _count source_evid	
Phenotype (approved_All  1.6 Results table ships API-based approved_All	Intology (HPO)-be- symbol hgnc_ All  alt table: Publishes results of the strength of the stren	id gene_name_ All  Dator  ne fifth analysis sore extraction from	reported source [All earching kidney dan PubMed.	es (OMIM, O ce source_ All	count source_evice_All	lence
Phenotype (approved_All 1.6 Results table shaped approved_All 1.6 API-based approved_All 1.6 Results table shaped approved_All 1.6 Results table s	Intology (HPO)-be- symbol hgnc_ All  alt table: Publishes results of the strength of the stren	id gene_name_ All  Dator  ne fifth analysis sore extraction from	reported source [All earching kidney dan PubMed.	es (OMIM, O ce source_ All	rphanet)count source_evid	lence
All  1.6 Results table s API-based a approved_	Intology (HPO)-b  symbol hgnc  All  alt table: Pub  hows results of the  utomated literatu  symbol hgnc	id gene_name_ All  Tator  ne fifth analysis sore extraction from id gene_name_	e disease database _reported source All earching kidney dearching learching	es (OMIM, O ce source_ All lisease associ ce source_	count source_evice_ated genes from a Pul	lence
Phenotype (approved_All All All All All All All Approved_All All All Approved_All All Approved_All All All Approved_All All All All All All All All All All	Intology (HPO)-b  symbol hgnc  All  alt table: Pub  hows results of the  utomated literatu  symbol hgnc	id gene_name_ All  Tator  ne fifth analysis sore extraction from id gene_name_	e disease database _reported source All earching kidney dearching learching	es (OMIM, O ce source_ All lisease associ ce source_	count source_evice_ated genes from a Pul	lence

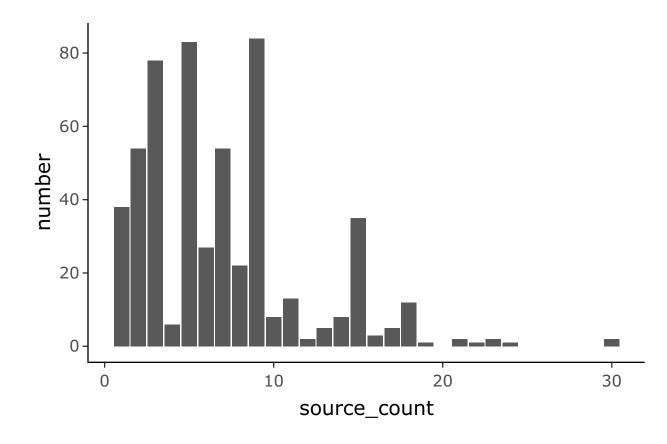
## 2.1 UpSet plot of merged analyses sources

UpSet plot of the merged analyses.



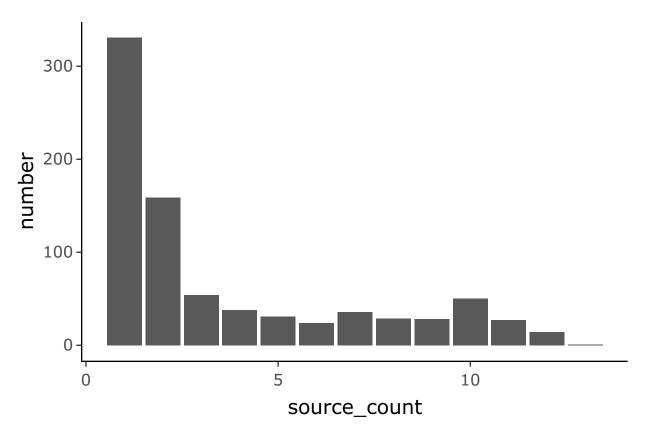
## 2.2 Bar plot of PanelApp results

Bar plot of the PanelApp analysis.



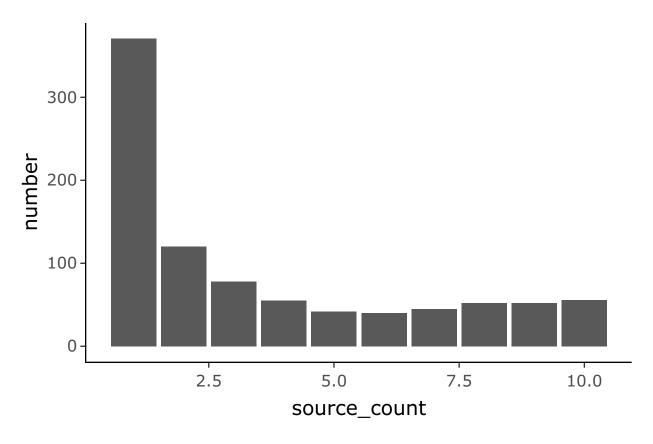
# 2.3 Bar plot of Literature results

Bar plot of the Literature analysis.



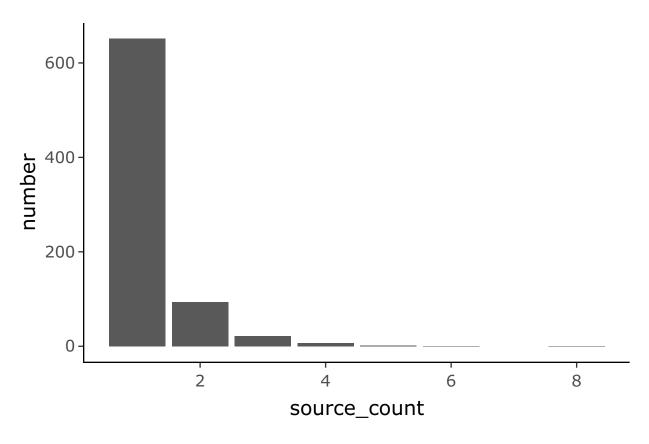
# 2.4 Bar plot of Diagnostic panels results

Bar plot of the Diagnostic panels analysis.



# 2.5 Bar plot of HPO in rare disease databases results

Bar plot of the HPO in rare disease databases analysis.



# 2.6 Bar plot of PubTator results

Bar plot of the PubTator analysis.

