THE CHRONIC KIDNEY DISEASE DATABASE (CKDdb)

*Marco Fernandes and Holger Husi

Institute of Cardiovascular and Medical Sciences (ICAMS), University of Glasgow, Glasgow, G12 8TA, UK

*mail to m.fernandes.1@research.gla.ac.uk

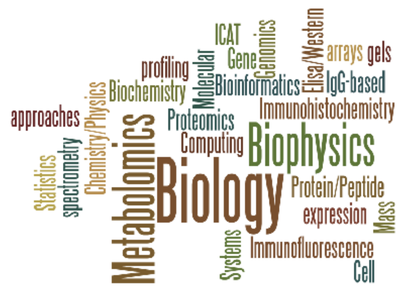
Background

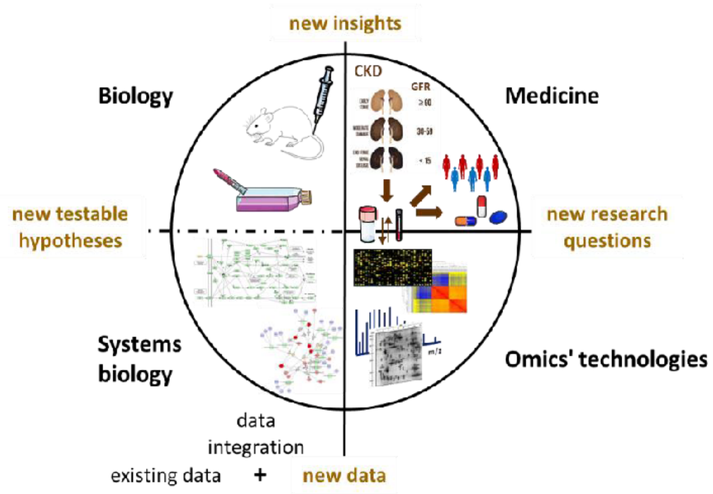
The Chronic Kidney Disease database CKDdb is an integrated and clustered information resource that covers multi-omic studies (microRNA, genomics, peptidomics, proteomics and metabolomics) of chronic kidney disease (CKD), disease-related and diseases leading to this trait.

This resource was built by mining the existing literature and afterwards was performed a manual curation of the collected data. In order to deal with the high heterogeneity and diversity of the gathered data, a specific ontology was applied to tie together and harmonise multi-level omic studies based on gene and protein clusters (CluSO) and mapping of orthologous genes (OMAP) across species.

This database was intentionally built to allow disease pathway analysis by data-driven using a system approach, which is much more amenable and has a greater potential to provide an unbiased and novel testable hypothesis as an end-result.

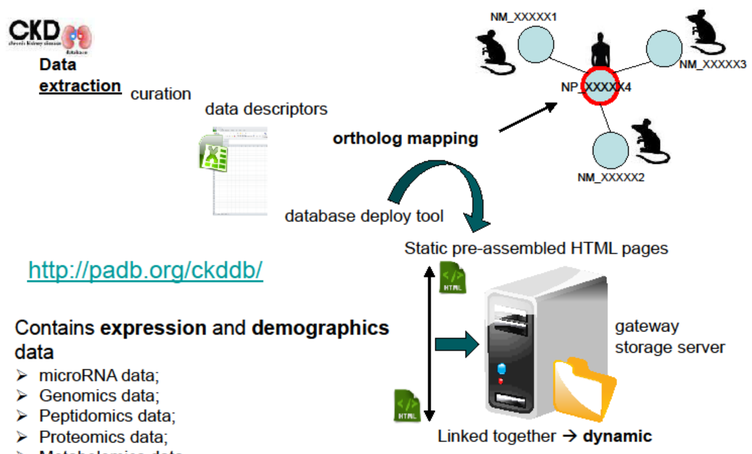
The CKDdb database is a component of the work-package #1 (WP1) of the iMODE-CKD consortium that aims for the identification of the molecular determinants of established CKD. The database will incorporate existing knowledge and de novo generated data from our research partners, in order to yield biological meaning by integrating all available information and therefore has the potential to unravel and allow an in-depth understanding of the key molecular events that modulate CKD progression.





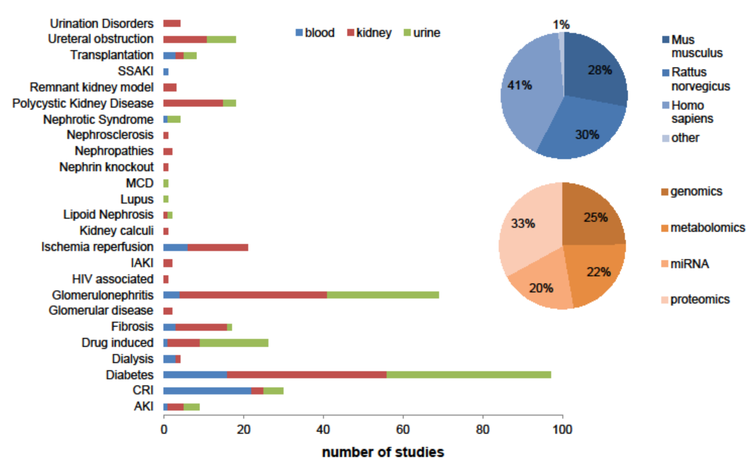
Construction & Content

CKDdb database development workflow overview



Contains **expression** and **demographics** data

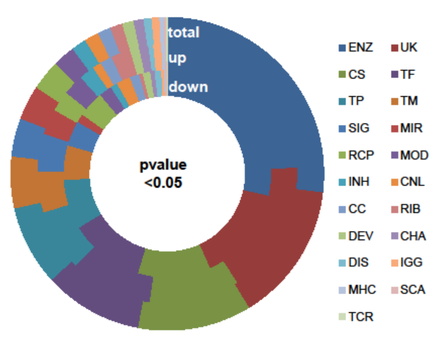
- microRNA data;
- Genomics data;
- Peptidomics data;
- Proteomics data;
- Metabolomics data.



Overlook of the main diseases (MeSH classification) related with CKD that are represented in CKDdb. Categorisation by tissue/fluid source, organism, and 'omic technology is described.

SSAKI: severe septic shock-associated kidney injury, MCD: minimal change disease, IAKI: ischemic acute kidney injury, CRI: chronic renal insufficiency, AKI: acute kidney injury.

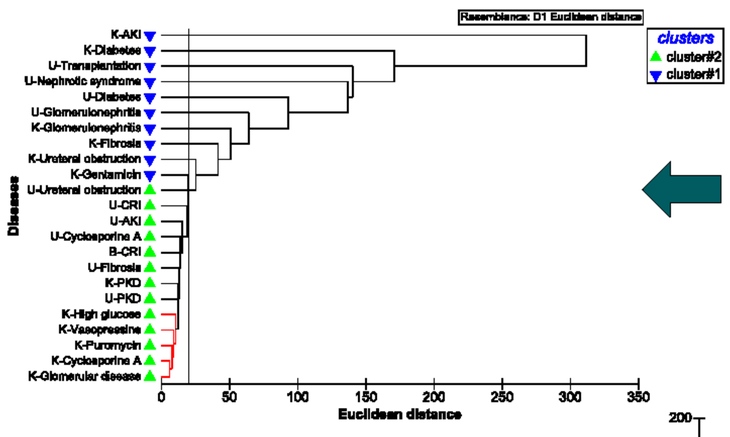
Functional tag clustering



Functional tag clustering of gene and proteins that are inside CKDdb. Tagging was originally built for PADB (www.padb.org) based on oversimplification of Swiss-Prot keywords. Grouping of molecules using this approach allow us in first instance to summarise all the studies containing gene and protein expression and therefore narrowing our research.

CC: cell cycle; CHA: chaperone, chaperonin; CNL: channel; CS: cell shape; DEV: development, cell growth, differentiation, morphogenesis; DIS: disease; ENZ: enzyme, enzymatic properties; IGG: Immunoglobulin; INH: inhibitor; MHC: major histocompatibility complex component/protein cluster; MIR: microRNA; MOD: modulator, regulator; RCP: receptor; RIB: ribosome; SCA: scaffold, docking, adaptor; SIG: signalling; TCR: T-cell receptor; TF: transcription and translation, gene regulation; TM: transmembrane; TP: transport, storage, endocytosis, exocytosis, vesicles; UK: unknown.

Utility & Exploratory Analysis



Hierarchical clustering of the molecular expression (miRNA, protein and metabolites) by group averaging of the Euclidean distances between renal diseases from multiple tissue and fluid sources.

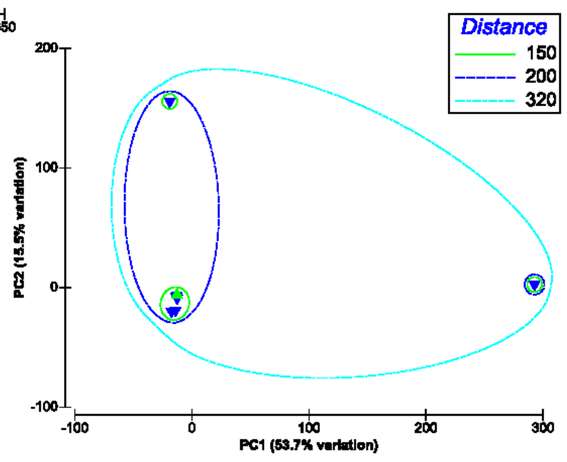
Using a similarity level of 20 we were able to distinguish two major clusters: #1 and #2.

Performing a Principal Component Analysis (PCA) allowed us to identify the components that account more for the change of variance.

The variables "K-AKI" and "K-Diabetes" are more correlated with PC1 and PC2, respectively.

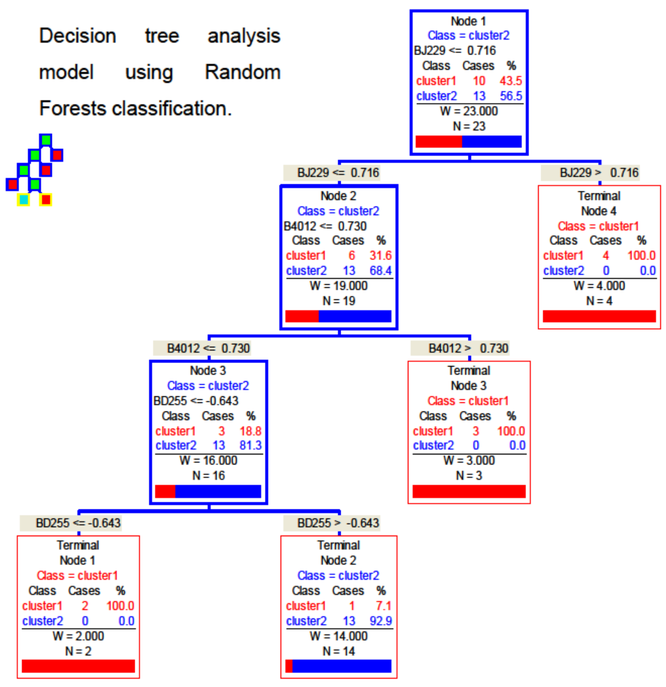
Both components account for ~69.2% of the cumulative variation of the data.

Using a resemblance level of 150, 200 and 300 allowed us to identify overlay clusters.

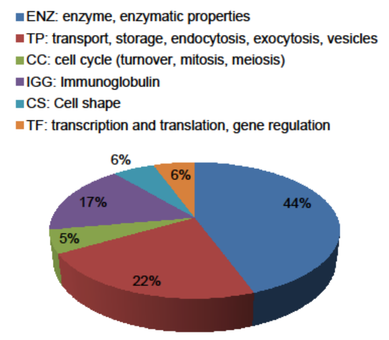


Random Forests tree classification

Decision tree analysis model using Random Forests classification.



Decision tree modelling using molecule expression (miRNA, protein and metabolites) of the clusters #1 and #2 allowed us to identify three relevant molecules: BJ229 (Ig heavy chain V-III region; IGH) B4012 (Cytoglobin; CYGB) and BD255 (Ras-related protein Rab-35; RAB35) with a AUC of 0.950 and a overall correctness of 92%.



Associated functional tags of the top 25% most important molecules of the assembled model.

Discussion & Conclusions

- To our knowledge this database is the most comprehensive molecular information resource in characterising CKD-related experiments and model systems;
- This database is primarily aimed to allow disease pathway analysis through a system approach in order to yield biological meaning by integrating all existing information and therefore has the potential to unravel and gain an in-depth understanding of the key events that modulate CKD;
- Multidimensional -omics data can be used to construct models of molecular interaction networks, using both prior and de novo knowledge, therefore linking genes with disease based on genome-wide association studies, miRNAs and mRNAs targets, protein-DNA interactions, protein-protein interactions, protein-substrate binding, metabolic pathway interactions and drug-target interactions, where these molecular entities are represented as nodes and their interactions as edges;
- The assembled tree model was able to distinguish the two initial formed disease clusters based in three main proteins: the Ig heavy chain V-III region protein, the cytoglobin and the ras-related protein Rab-35. This exploratory analysis of the CKDdb by data-driving using previous knowledge can be implemented in a further study for the description of a specific disease phenotype by their molecular elements.

Further Information

To download the PDF of this poster or to access similar studies:

http://www.padb.org/ckddb

References

http://www.padb.org/ckddb

Acknowledgments

The research leading to these results has received funding from the European Union's Seventh Framework Programme FP7/2007–2013 under grant agreement **FP7-PEOPLE-2013-ITN-608332**

52 ERA 222--FP

J1) Chronic Kidney Disease. Lab methods, GFR measurement, urine proteomics.

Marco Fernandes

DOI: 10.3252/pso.eu.52era.2015

ePosters supported by F. Hoffmann- La Roche Ltd.

52nd ERA-EDTA CONGRESS London 2015 UNITED KINGDOM MAY 28TH

PosterSessionOnline.com