Integration of proteomic data and clinical annotations reveals ciliopathy mechanisms

K. Koutroumpas[1], T.J.P. van Dam[2], G. Toedt[3], Q. Lu[4], J. van Reeuwijk[5], K. Boldt[6], M. Ueffing[6], T. Gibson[3], R. Russell[4], R. Roepman[5], M. Huynen[2], M. Elati[1], F. Képès[1].

1. institute of Systems and Synthetic Biology, Genopole, CNRS, Université d'Evry, France, 2. Centre for Molecular and Biomolecular Informatics, Radboud University Medical Centre, The Netherlands, 3. European Molecular Biology Laboratory, Germany, 4. BioQuant, University of Heidelberg, Germany, 5. Department of Human Genetics, Radboud University Nijmegen Medical Centre, The Netherlands, 6. Institute for Ophthalmic Research, Centre for Ophthalmology, University of Tübingen, Germany

Motivation

Understanding the molecular and cellular mechanisms of diseases is vital for dissecting pathogenesis, identifying appropriate therapeutic targets and designing effective treatments. Recent advances in DNA sequencing technology have provided a torrent of genetic data that can now be used to elucidate the genetic basis of human diseases. A consensus has emerged among biologists that to fully exploit the available data, they have to be correlated with additional type of data. This is especially important for rare-disease genetics, mainly because of the small number of available patients.

Data

Tandem Affinity Purification experiments: 453 experiments using 202 baits.

Proteomic Data

Gene phenotype associations from HPO

DNA Sequencing

Phenotype relations from HPO

Clinical Data

Methodology

Algorithm

Input : Proteomic (TAP) data, Clinical data
(*i.e* set *S* of genes associated with a phenotype)

1 : Create *M* random permutations of TAP data.
2 : Use a scoring system to create weighted graphs from the TAP data (experimental and permuted).
3 : Calculate enrichment scores (*ES*) for the experimental data
4 : Estimate the statistical significance of an observed *ES* by comparing it with a set of enrichment scores ES_{null} computed by the permuted data.
5 : Construct a disease/phenotype network from the known and predicted disease/phenotype associated genes.

Computational Methods

Enrichment score computation

Input : Weighted graph $G=(V,E,w)$, set $S \subseteq V$
1 : for $\boldsymbol{p} \in V$ do
2 : $W = \langle w_{(\boldsymbol{p},p')} : (\boldsymbol{p},p') \in E \rangle$ (decreasing order)
3 : for $i \in \left[1, |W|\right]$ do
4 : $L_i = \left\{ p' \in V : (\boldsymbol{p},p') \in E, w_{(\boldsymbol{p},p')} \geq w_i \right\}$
5 : $F(S,i) = \sum_{p' \in L_i \cap S} \frac{w_{(\boldsymbol{p},p')}}{\sum_{g \in S} w_{(\boldsymbol{p},g)}} - \sum_{p' \in L_i \setminus S} \frac{1}{|L| - |S|}$
7 : end
6 : $ES(G,S,\boldsymbol{p}) = \max\left(\max_i F(S,i), 0\right)$
8 : end

ES Null distribution

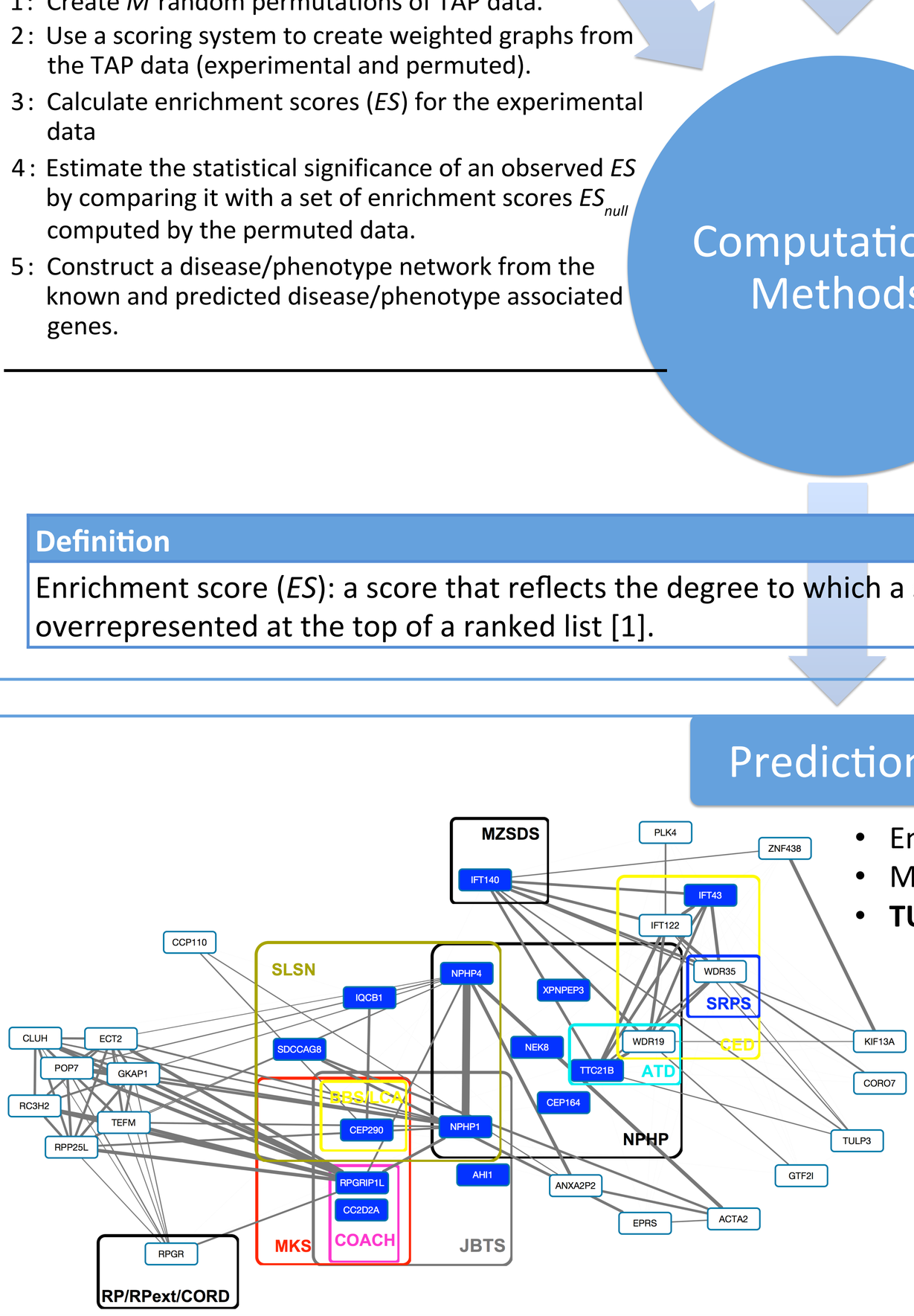
Input : Set of weighted graphs $\boldsymbol{G} = \{G_1,\ldots,G_N\}$ with $G_i = (V,E_i,w_i)$, set $S \subseteq V$
1 : $ES_{null} = \{\ \}$
2 : for $G_i \in \boldsymbol{G}$ do
3 : Compute Enrichment Score
4 : $ES_{null} = ES_{null} \cup \{ES(G_i,S,p)\ \forall\ p \in V\}$
5 : end

Definition

Enrichment score (*ES*): a score that reflects the degree to which a set *S* is overrepresented at the top of a ranked list [1].

Results - Conclusions

Predictions



- End-stage renal failure for a patient with **IFT122** mutation [2]
- Mutations in **WDR19** are associated with NPHP [3]
- **TULP3:**
  - Required for ciliary GPCR localization [4,5].
  - Gpr161 mutations phenocopy Tulp3 and IFT-A mutants, and cause increased Shh signalling [5].
  - Abnormal Hh signaling has been linked to NPHP [6,7]
  - NPHP proteins have been implied to ciliary GPCR localization [8].

➢ **Abnormal Hh signaling may be the cause of NPHP**
➢ **GPCR mislocalization is a possible way by which cilia defects affect Hh signaling.**

References:

[1] Subramanian, Aravind, et al. *Proceedings of the National Academy of Sciences* 102.43 (2005): 15545-15550. **[2]** Alazami, Anas M., et al. *Molecular genetics & genomic medicine* 2.2 (2014): 103-106. **[3]** Bredrup, Cecilie, et al. *The American Journal of Human Genetics* 89.5 (2011): 634-643. **[4]** Mukhopadhyay, Saikat, et al. *Genes & development* 24.19 (2010): 2180-2193. **[5]** Mukhopadhyay, Saikat, et al. *Cell* 152.1 (2013): 210-223. **[6]** Hildebrandt, Friedhelm, Massimo Attanasio, and Edgar Otto. *Journal of the American Society of Nephrology* 20.1 (2009): 23-35. **[7]** Hurd, Toby W., and Friedhelm Hildebrandt. *Nephron Experimental Nephrology* 118.1 (2010): e9-e14. **[8]** Sang, Liyun, et al. *Cell* 145.4 (2011): 513-528.

Contact:
Konstantinos Koutroumpas
institute of Systems and Synthetic Biology
Genopole, CNRS, University of Evry
91030 Evry, France
konstantinos.koutroumpas@issb.genopole.fr

Acknowledgements: Work supported by European Commission under the project SYSCILIA (FP7/2007-2013)

Syscilia