Distilling a network of interactions to uncover genes involved in Retinitis Pigmentosa disease

Sergio Castillo-Lara[1], Daniel Boloc[1], Gemma Marfany[123], Roser Gonzàlez-Duarte[123], Josep F Abril[12]

1 Departament de Genètica, Universitat de Barcelona, Av. Diagonal 643, 08028 Barcelona, Catalonia, Spain;
2 Institut de Biomedicina (IBUB), Universitat de Barcelona, Av. Diagonal 643, 08028 Barcelona, Catalonia, Spain;
3 CIBERER, Instituto de Salud Carlos III, Barcelona, Catalonia, Spain.

RETINITIS PIGMENTOSA

Autosomal dominant	30-40%
Autosomal recessive	50-60%
X-Linked	5-15%

Rod
(dim light vision)

Cone
(daylight)

Loss of central vision

Night blindness

Peripheral constriction of visual fields

Normal view

RP affected view

Known RP Genes:

- Phototransduction cascade: RHO, PDE6A, PDE6B, CNGA1, CNGB1, SAG.
- Vitamin A metabolism: ABCA4, RLBP1, RPE65, LRAT, RGR.
- Structural or cytoskeletal: RDS, ROM1, FSCN2, TULP1, CRB1, RP1.
- Signaling, cell-cell interactions, or synaptic interaction: SEMA4A, USH2A, RP2.
- RNA intron-splicing factor: PRPF31, PRPF8, PRPF6, PRPF3, RP9.
- Maintenance of ciliated cells: TTC8, RPGR.
- pH regulation: CA4.
- Phagocytosis: MERTK.
- Other/unknown functions: CERKL, IMPDH1.

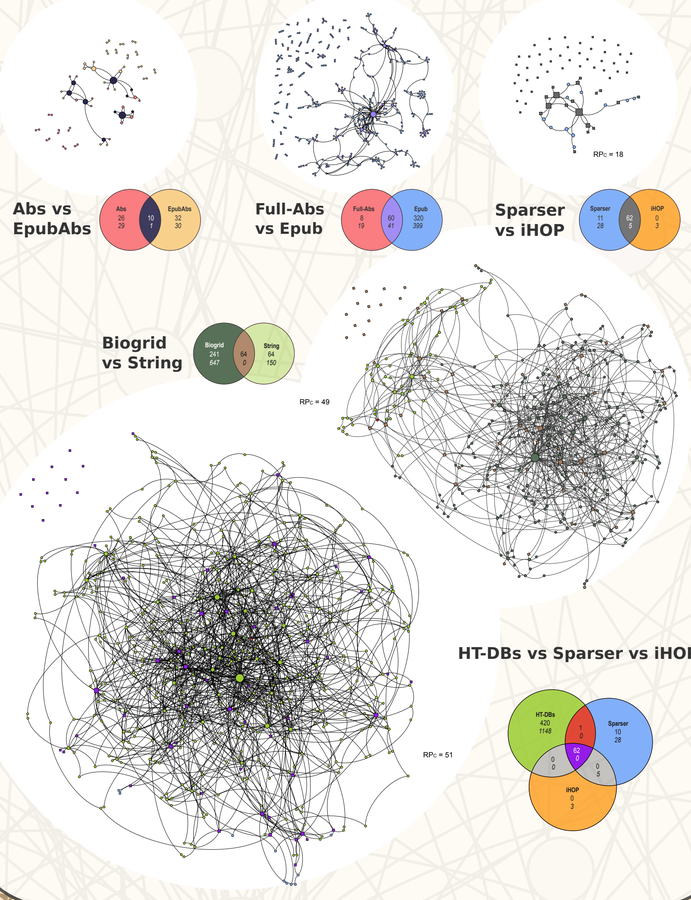
ABSTRACT

Retinitis Pigmentosa (RP) is a highly heterogeneous genetic disorder with more than 60 known causative genes. The identification of those genes, which are also closely related to non-syndromic retinal dystrophies, has increased steadily during the last decade. However, about 30% of the cases described for RP remain unassigned. Some of the known RP genes, such as CERKL, are still poorly under-stood at molecular level to properly describe their function and how this relates to the observed phenotype in retinal cells. Although a considerable amount of genetic and functional data on single RD genes and mutations has been gathered, a compre-hensive view is still missing. In contrast to the analysis of individual isolated genes, finding the networks linking disease genes provides powerful ethiopathological insights.

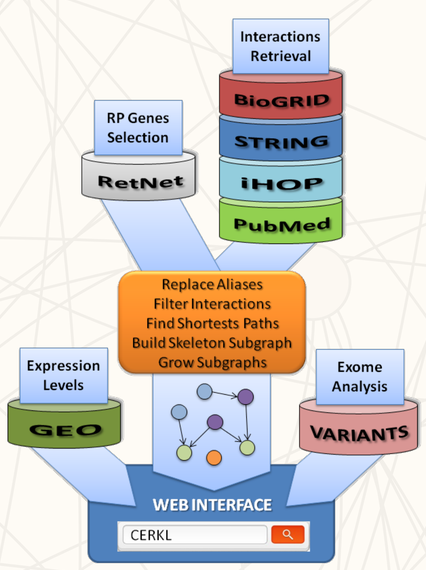
Our approach relies on the connectivity of the network generated by merging data from different sources: high-throughput data from BioGRID and STRING databases; manually curated data for interactions retrieved from iHOP; as well as interactions filtered out by syntactical parsing from up-to-date abstracts and full-text papers related to RP research field. We analyzed the paths emerging when known RP genes are used as bait over the whole interactome. In order to simplify the search space we kept the minimal number of connections among those genes and their closer neighbors. We are integrating tissue-specific expression levels and phenotypic data on top of this simplified network, centered on the RP causative genes, while providing an interactive interface for the molecular biologist to explore that network.

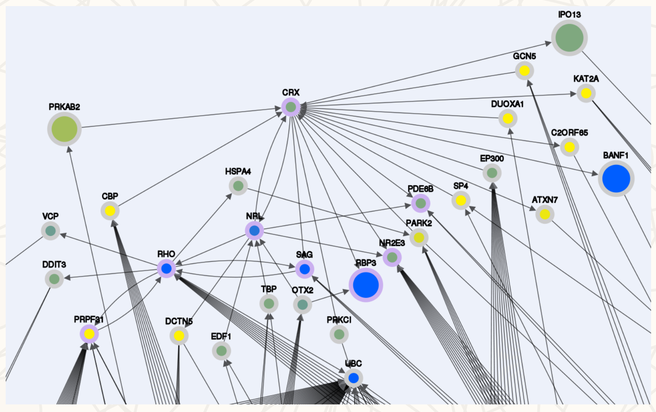
RECONSTRUCTING RP GENES NETWORK

A graph skeleton was produced for every interactions dataset separately, using the RP genes as bait to capture the paths connecting them. The overlap graph is shown for five different dataset combinations, *from left to right and from top to bottom*: 1) interactions filtered by Sparser from abstracts, having or not full text available (Abs and EpubAbs respectively); 2) Sparser interac-tions for all abstracts (FullAbs) compared with those retrieved from full text (Epub); 3) overlap between interactions filtered from iHOP and manually curated, from those provided by Sparser; 4) skeleton comparison for the two high-throughput (HT) databases, BioGRID and STRING; and finally, 5) the triple merge of skeleton graphs for iHOP curated interactions, Sparser extraction, and HT databases. Square nodes refer to the 62 RP genes used as bait to distill the subnetworks. RP_C value correspondsto the RP genes complement, those that were found on the network when distilling the skeleton subnetwork. This number increases when considering larger interactions sets, but varies when merging the individual sub-networks. The largest RP_C value (56) is retrieved when combining all the interactions from the different datasets in a single run, which was used to distill the skeleton graph underlying the RPGeNet browser. All graphs were edited with Gephi.



PROTOCOL

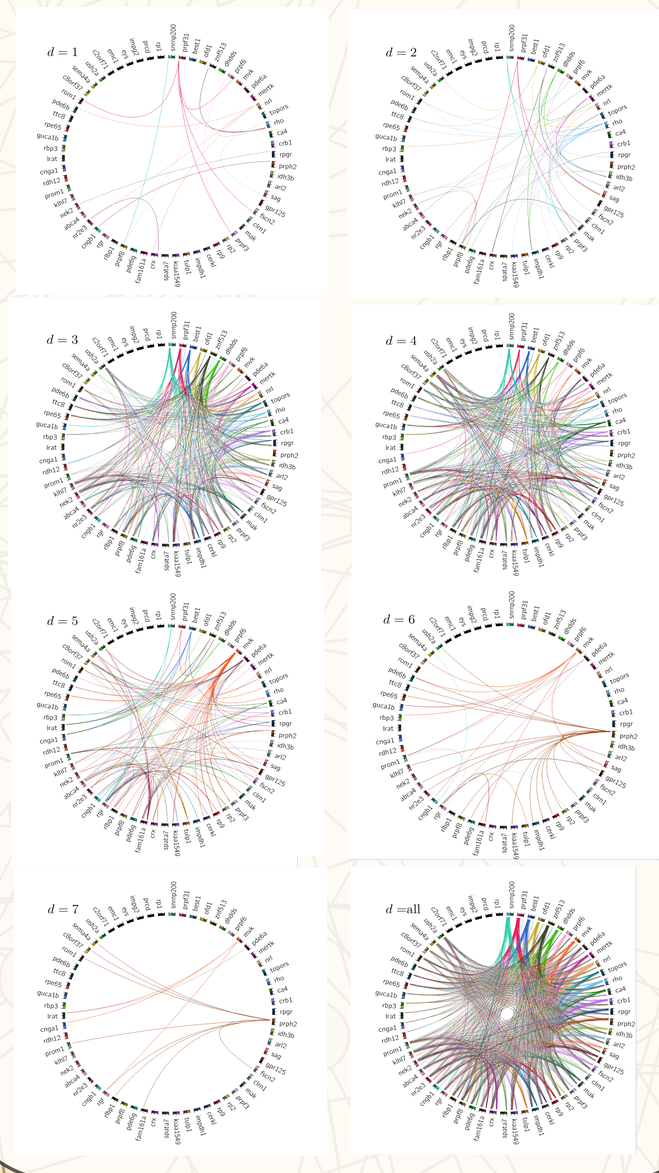




Network snapshot of the interaction partners retrieved from RPGeNet when querying for CRX, NRL and RHO genes. Level two neighbourhood nodes were requested upon the basic skeleton graph. RP genes are highlighted by a light violet border, in contrast to the grey colored border for the non-RP genes. Core color reflects the relative expression level in retina with respect to average expression over all tissues; shading from blue (over-expressed in retina) to yellow (under-expressed). As an example, a non-RP gene, ATXN7, which is connected to CRX, was initially described as one of the main factors causing human ataxia; it has been recently described that it plays a role in macular degeneration. Our approach can provide hypotheses to link genes/proteins to relevant retina physiological roles.

RP GENES PAIR-WISE CONNECTIVITY

Each of the diagrams show the pair-wise connectivity between RP genes or proteins from the interaction network generated, where d corresponds to the distance for the shortest directed path across the network between the pair of connected RP genes—maximum distance found was of 7 edges (or six internal nodes between the two RP genes)—. The last plot summarizes all the previous diagrams, showing all the possible connections among pairs of RP genes at different distances. On the outer ring we have depicted the standard symbols for the 62 RP genes used to build the network, they were sorted by the number of incoming connections, then by the outcoming ones. Each of the RP genes found to interact with other RP genes has a distinct color box, while those genes for which there was not interaction found are shown in black. The color boxes have two halves: one lighter grouping all the outgoing connections—so there is a directed path from current gene node—, and one darker gathering all the incoming connections—meaning that there is a directed path to current gene node—. By making this difference it is easy to spot genes that are upstream of many in the network—like snRNP200 and PRPF31—, mainly downstream of others—like USH2a and SEMA4a—, or even internal hubs—like CERKL, CRX and RP2—. All diagrams made using Circos.



References:

① **RetNet, the Retinal Information Network** http://sph.uth.edu/retnet/ ② Stark C, et al. **BioGRID: a general repository for interaction datasets**. NAR, 2006. ③ Franceschini A, et al. **STRING v9.1: Protein-protein interaction networks, with increased coverage and integration**. NAR, 2013. ④ Fernández JM, et al. **iHOP web services.** NAR, 2007. ⑤ Canese K, Weis S: **PubMed: The bibliographic database**. NCBI Handbook, 2013. ⑥ Mostafavi S, et al. **GeneMANIA: a real-time multiple association network integration algorithm for predicting gene function.** Genome Biol, 2008. ⑦ Krzywinski M, et al. **Circos: An information aesthetic for comparative genomics**. Gen Res 2009. ⑧ Bastian M, et al. **Gephi: an open source software for exploring and manipulating networks**. Int AAAI Conf on Weblogs and Social Media, 2009. ⑨ Marie-Catherine de Marneffe, et al. **Generating Typed Dependency Parses from Phrase Structure Parses**. LREC, 2006.

Acknowledgements

This work was supported by grants SAF2013-49069-C2-1-R (GM & RGD), BFU2010-15656 (GM, Ministerio de Ciencia e Innovación), SGR2014-0932 (RGD & GM, Generalitat de Catalunya), CIBERER (U718), Retina Asturias (RGD), ONCE (RGD), BFU2009-09102 (JFA, Ministerio de Ciencia e Innovación), and 2009-SGR-001018 (JFA, Generalitat de Catalunya).