DiGSNP: A web tool for disease-gene-SNP prioritization

Carmen Navarro[1], Carlos Cano[1]
Armando Blanco[1], Fernando García-Alcalde[2]

[1] *University of Granada, Spain*
[2] *Max Planck Institute for Infection Biology, Berlin, Germany*

November, 2012

Index

Motivation and Objectives

Context

- Next-Generation Sequencing: decreasing cost, raising throughput.
- Increasing availability of large amounts genome information.

Motivation

- **Personalized medicine**: improve diagnosis, disease prevention and treatment based on individual genomical information.
- Aim: Relate **diseases - individual genomical information**.
- Required: Efficient computational tools that produce useful, summarized results.

Motivation and Objectives

Objectives

Support development of personalized medicine, by:

- Helping **discover putative mutations** and genes related to a query disease.
- **Reducing search space** to a reliable, manageable set of candidate genes and SNPs.
- Guiding further research on the **most promising** hypotheses.

Background

Disease-gene prioritization

- Many methods for disease-gene prioritization.
- Most **do not consider** genome variations (i.e. SNPs, insertions, deletions).
- Variations are **cause** for many diseases.

Background

SNP prioritization

- Relating variations directly to diseases: gene relationship with disease is **left out** or implicit.
- Many require experimental evidence, such as GWAS studies:

  difficult to **start research** on **new or rare** diseases.
- Most focused on **coding** regions.
- Important regulatory regions:

  SNPs altering TFBSs.

DiGSNP

Features

- Relates diseases, genes and SNPs **at once**.
- **Two-level hierarchy**:
  - Disease-gene.
  - Gene-SNP.
- Disease-gene prioritization using **ProphNet**.
- SNP scoring with **IntuitSNP** intuitionistic approach.
- Gene-SNP relations: dbSNP.

ProphNet

Description

- Used for disease-gene prioritization.
- **Network-based prioritization** tool.
- Integrates data from an **arbitrary** set of sources.
- Prioritizes a Target Set based on how related this is to a Query Set.
- Used sources: OMIM, HPRD, DOMINE, InterDom, Pfam.

IntuitSNP

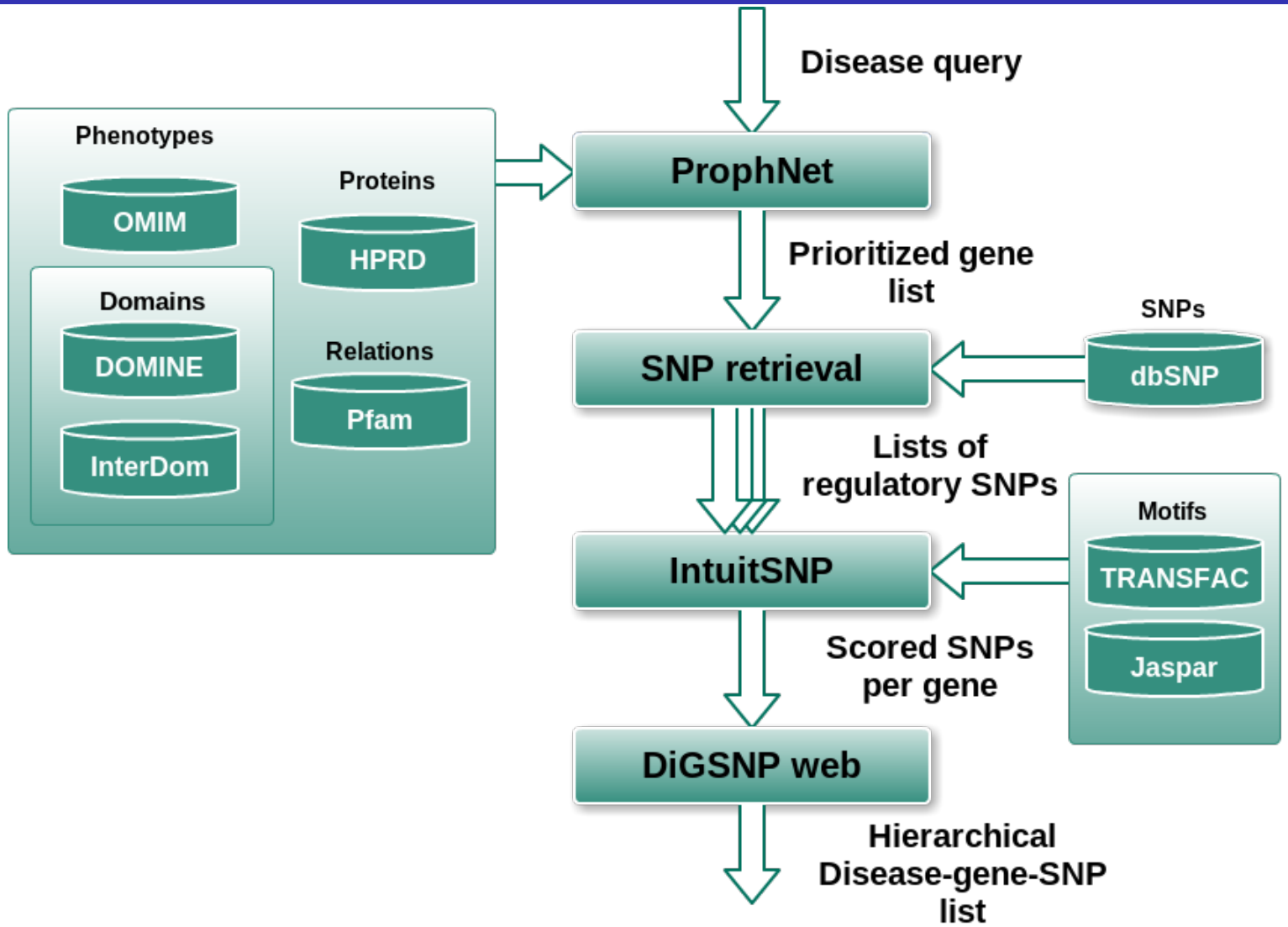
Description

- Calculation based on intuitionistic similarity score motif-sequence: SC_{intuit}
- SNPs that drastically alter a motif's binding affinity:

Top candidate regulatory SNPs.

- For each gene obtained in prioritized list:
  - Query dbSNP for SNPs in regulatory regions.
  - Obtain their IntuitSNP score.
  - Order them in descending order based on the difference between wild-type and mutated scores.

DiGSNP workflow



DiGSNP

Web tool

- DiGSNP is available at `http://genome2.ugr.es/digsnp`
- **Efficient**: Results within 2-3 minutes.
- **Intuitive**: Visually understandable results set.
- **Cross-referenced** with several information sources (e.g. dbSNP, Jaspar).

Results

Validation data

- Systematic validation difficult:

  Lack of variations-genes-diseases resources for validation.

	putative rSNPs	# cited	% cited
dbSNP	576.440	3994	**0,69**
HMGD (pub.)	576.440	1582	**0,27**
HMGD (subscr.)	576.440	2542	**0,44**

Results

Validation

- Currently gathering a larger dataset to better validate results:
  - dbSNP's **medical references**.
  - HMGD public available citations.
  - **RegulomeDB**: Recent curated database with regulatory information of SNPs.
- Preliminary results **consistent** with medical literature.

Results for Breast Cancer

Gene	ProphNet score	SNP id	Function	Motif	Motif score
RAD51	0,527	rs150213086	UTR5	C-MAF	0,335
		rs1801320	UTR5	Kid3	0,319
		rs183455067	DWN500B	Kid3	0,319
		rs184898629	UP2K	Kid3	0,319
		rs187471538	UP2K	Kid3	0,319
		rs7180135	UTR3	Kid3	0,319
BRCA2	0,519	rs145901536	UP2K	Kid3	0,319
		rs10492394	UP2K	ZNF333	0,300
		rs187284594	UP2K	ZNF333	0,300
		rs79681965	DWN500B	HMGIY	0,289
		rs185674638	UTR3	HMGIY	0,279
		rs55641815	UTR5	Churchill	0,278
BRCA1	0,452	rs191995002	DWN500B	C-MAF	0,327
		rs148196794	UTR5	Kid3	0,319
		rs80356827	UTR5	C-MAF	0,294

Conclusions

- DiGSNP is **helpful for early stage research**.
- Suggestions made based on genomic information, motifs and binding affinity.
- **No need** for previous experimental results (i.e. GWAS, previous publications).
- Focus on **regulatory variations**.
- Useful tool as a start point to **boost further research**.

Further work

- Integrate **coding** variations.
- Influence of variation function class to score.
- User-defined SNP ranking.