



EPSILON: an eQTL prioritization framework using similarity measures derived from local networks

Lieven P.C. Verbeke¹, Piet Demeester¹, Jan Fostier¹, and Kathleen Marchal^{2,3}

¹IBCN - iMinds, Ghent University, Belgium, ²Department of Microbial and Molecular Systems, KU Leuven, Belgium,

³Department of Plant Biotechnology and Bioinformatics, Ghent University, Belgium

Introduction

- We used **SNP** data and **gene expression** data of ***S. cerevisiae*** to detect **eQTLs** using three different mapping methods
 - Non parametric regression
 - Mixed models
 - Elastic net
- Using a **physical interaction network**, we **prioritized** eQTLs spanning multiple genes to individual *causal* genes using a graph node **kernel** based approach
- Evaluating using **knockout pairs**, and when compared to random assignment or a shortest path approach, we obtained **superior results**
- Published in Bioinformatics:
 - Verbeke, L.P.C. et al. (2013) *EPSILON: an eQTL prioritization framework using similarity measures derived from local networks*. *Bioinformatics (Oxford, England)*, 29, 1308–16
 - Software available at bioinformatics.intec.ugent.be/epsilon

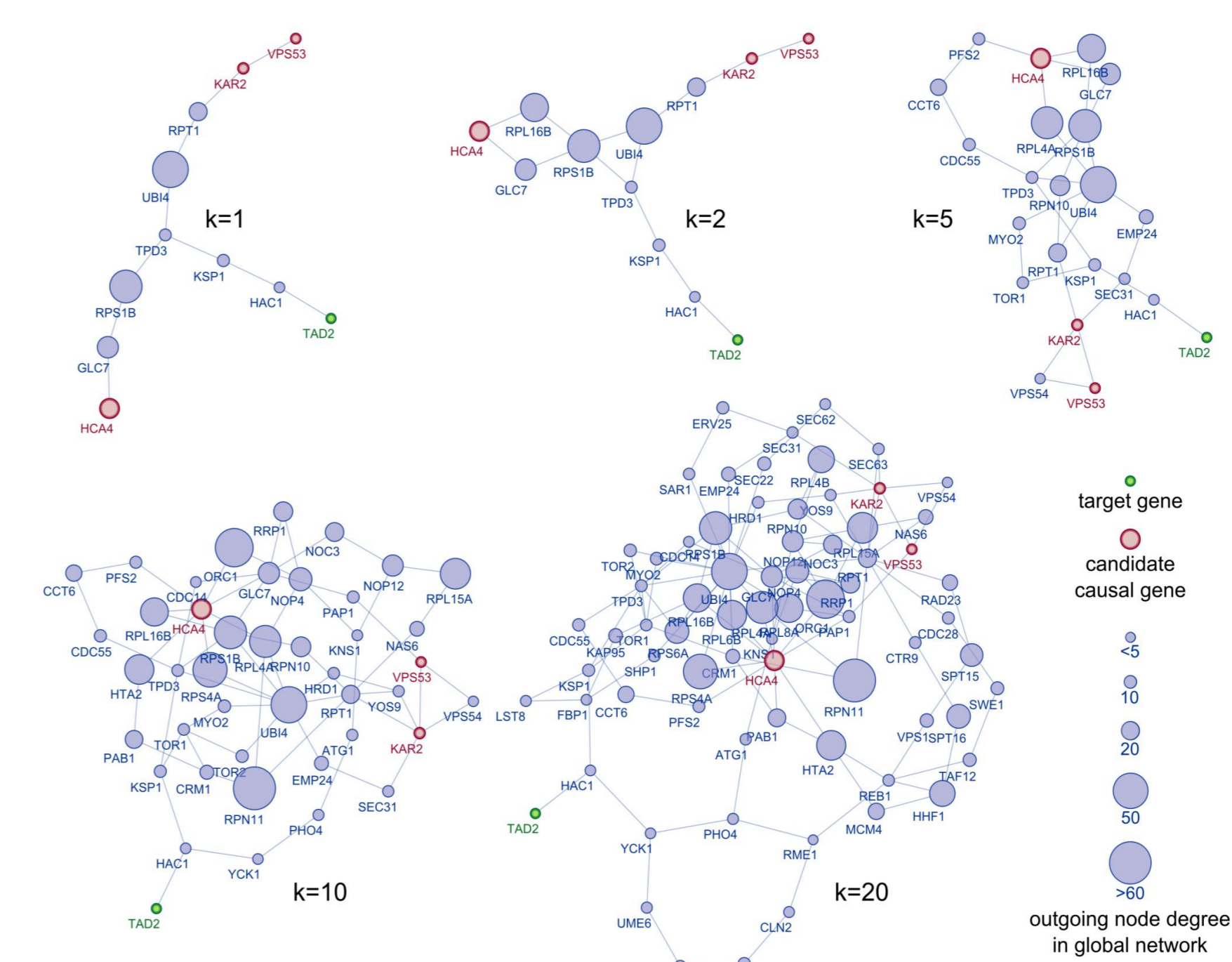
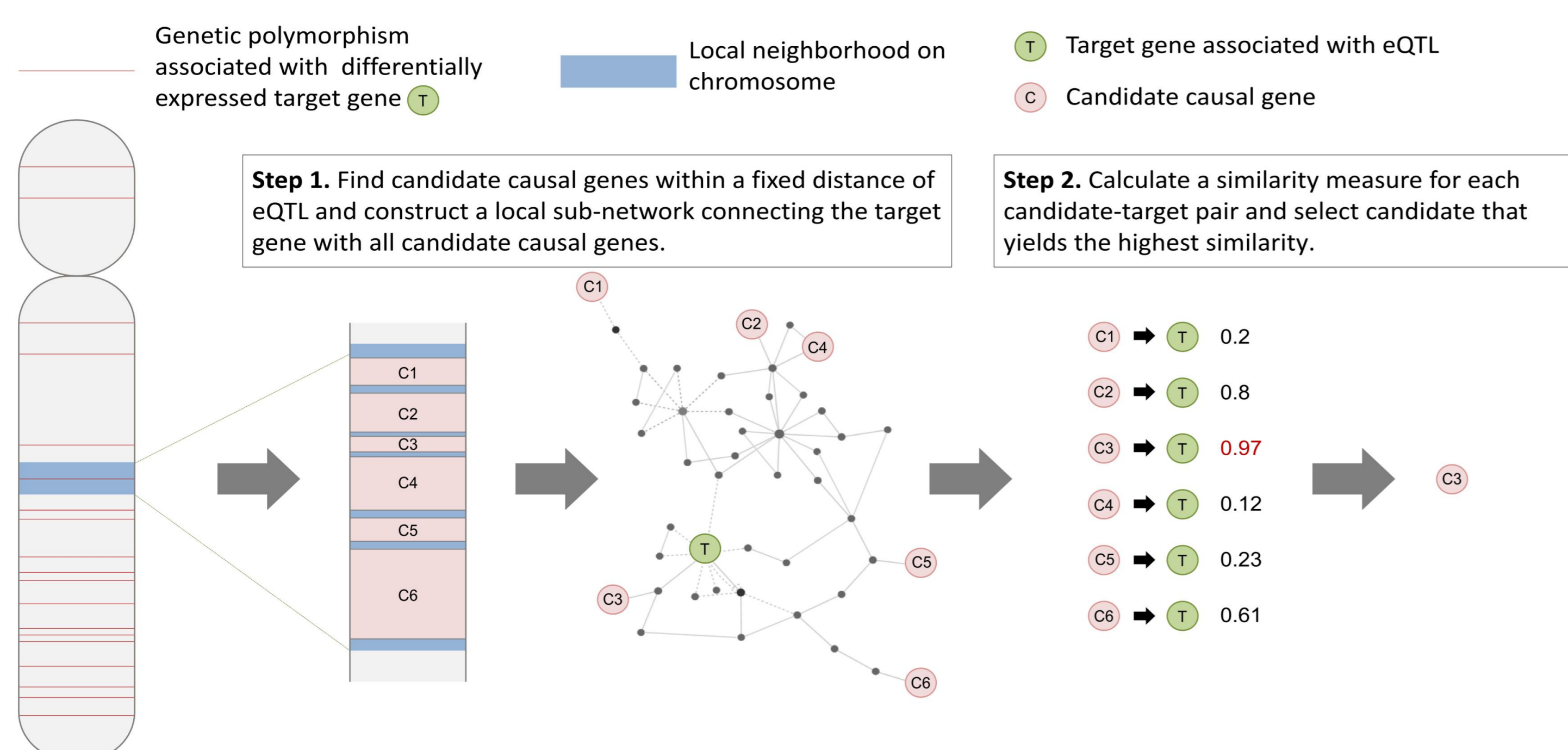
Kernel based similarity

- Kernels are functions calculated on graph nodes (each gene is a node in the interaction graph), producing node similarity matrices
- The kernels we use are typically used for recommendation tasks like
 - Customers who bought this also bought ...
 - People you may know ...
 - Web page importance ranking
- We are not the first to use kernels for prioritization (see e.g. Nitsch et al., 2010, *BMC bioinformatics*, 11, 460) but to our knowledge, this is a new application

Local networks

- Calculating kernels on the global interaction network yields bad performance
- Better to first find a local subnetwork, and then calculate similarity measures
- We use a variant of the Dijkstra shortest path algorithm to connect multiple candidate causal genes in an eQTL to a target gene: **k-trials shortest paths**
- The algorithm tries at trial k to find an alternative path, after making the path found at trial $k-1$ slightly more expensive

Method



Results

- Three different eQTL mapping methods
- Evaluation using Hughes (2000) knockout dataset
- Local kernel based prioritization consistently outperforms both reference methods
- Maximum number of retrievable knockout pairs depends on the eQTL mapping method and the network at hand
- Results are comparable or superior to results mentioned in literature, see e.g.
 - Suthram, S., Beyer, A., Karp, R. M., Eldar, Y., & Ideker, T. (2008). eQED: an efficient method for interpreting eQTL associations using protein networks. *Molecular systems biology*, 4(162)
 - Stojmirović, A. and Yu, Y.-K. (2009) ITM Probe: analyzing information flow in protein networks. *Bioinformatics (Oxford, England)*, 25, 2447–9.

