

Improving efficiency in epistasis detection with a gene-based analysis using functional filters

Diane Duroux^{1,✉}, Héctor Climente-González^{2,3,4,✉}, Aldo Camargo¹, Lars Wienbrandt⁵, David Ellinghaus⁵,
Chloe-Agathe Azencott^{4,2,3}, Kristel Van Steen¹

[✉] Equally contributing authors, ¹ BIO3 - GIGA-R Medical Genomics, Liege, Belgium; ² Institut Curie, PSL Research University, F-75005 Paris, France; ³ INSERM, U900, F-75005 Paris, France; ⁴ MINES ParisTech, PSL Research University, CBIO-Centre for Computational Biology, F-75006 Paris, France; ⁵ Institute of Clinical Molecular Biology, Christian-Albrechts-University of Kiel, Germany

Motivation

Epistasis analysis can complement oversimplified GWAS to understand complex traits.

Available biological knowledge is growing. Many choices can be made to use one of the sources or a combinations of several ones. There is a need to study the associated benefits.

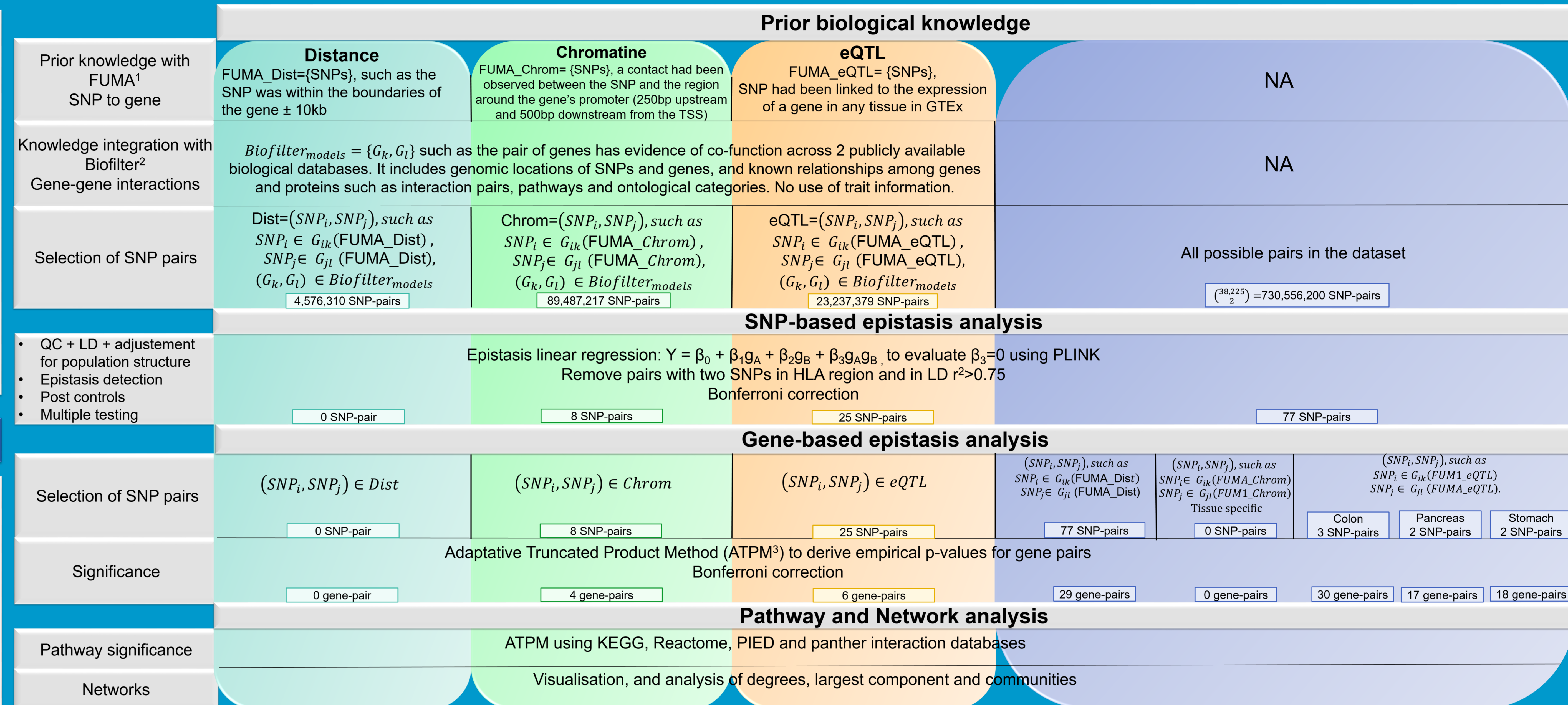
Genes are the natural units of analysis to interpret findings as they can indicate relevant pathways and biological processes. However in the context of epistasis, gene-based methods are underinvestigated.

Data

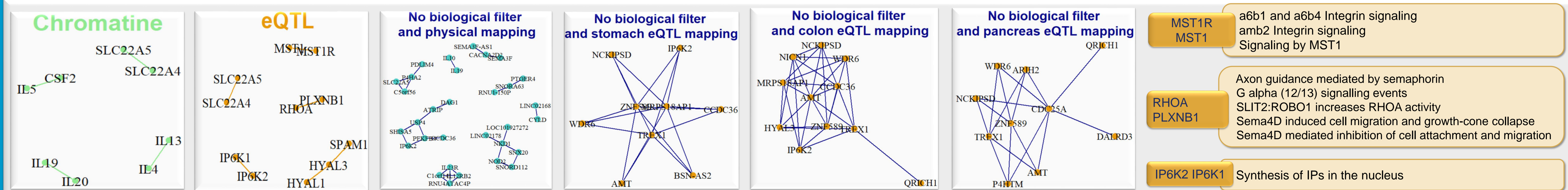
Inflammatory bowel disease is characterized by chronic inflammation of the gastrointestinal tract. Crohn's disease and ulcerative colitis are the principal types of IBD.

IBD data from the IBD consortium: 66,280 samples, corresponding to 32,622 cases and 33,658 controls and 130,071 variants .

Methodology



Results



- In this application, the chromatine filter does not seem well balanced in terms of number of tests and number of significant pairs detected because it requires more tests than eQTL and highlights fewer gene pairs.
- eQTL has a reduced number of tests, leads to more significant gene pairs than the hypothesis free analysis and identifies significant pathways.
- Hypothesis-free screening is more cumbersome but detects a higher number of significant SNP-pairs and gene-pairs.

Discussion & Conclusion

Gene-based epistasis analysis is recommended to understand the biological mechanism of diseases because it highlights biological processes and pathways. The hypothesis-free screening allows the identification of more gene-pairs and communities, whereas the eQTL filter is more suited to detect significant pathways.

- ✓ The combination of epistasis networks and SNP aggregation enhances identification and interpretation of SNP-based epistasis findings via communities and pathway detection.
- ✓ This pipeline not only allows to find high-dimensional interactions at the gene level, but also gives a global visualization with neighbors, and highly connected substructures.
- ✓ We discovered novel susceptibility genes and gene-pairs for IBD. For instance and to our knowledge, HYAL3, LINC02168, and SNORD112 have never been associated to IBD before.

Permutation analysis validate type I error and current knowledge about IBD was largely retrieved. Further simulations will be necessary to study the influence of the number of tests on the number of false positives. Therefore, additional tests of this protocol with other analytic tools will be performed.

References

- 1 Watanabe, Kyoko, et al. "Functional mapping and annotation of genetic associations with FUMA." *Nature communications* 8.1 (2017): 1826
- 2 Pendergrass, Sarah A., et al. "Genomic analyses with biofilter 2.0: knowledge driven filtering, annotation, and model development." *BioData mining* 6.1 (2013): 25.
- 3 Sheng, Xuguang, and Jingyun Yang. "An adaptive truncated product method for combining dependent p-values." *Economics letters* 119.2 (2013): 180-182.

Acknowledgement

This project has received funding from the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement No 813533.



Contact: diane.duroux@uliege.be