



# SNPsea: an algorithm to identify cell types affected by risk loci

Kamil Slowikowski, Xinli Hu, Soumya Raychaudhuri



## Abstract

SNPsea is an algorithm to identify cell types and pathways likely to be affected by risk loci. It requires a list of SNP identifiers and a matrix of genes and conditions.

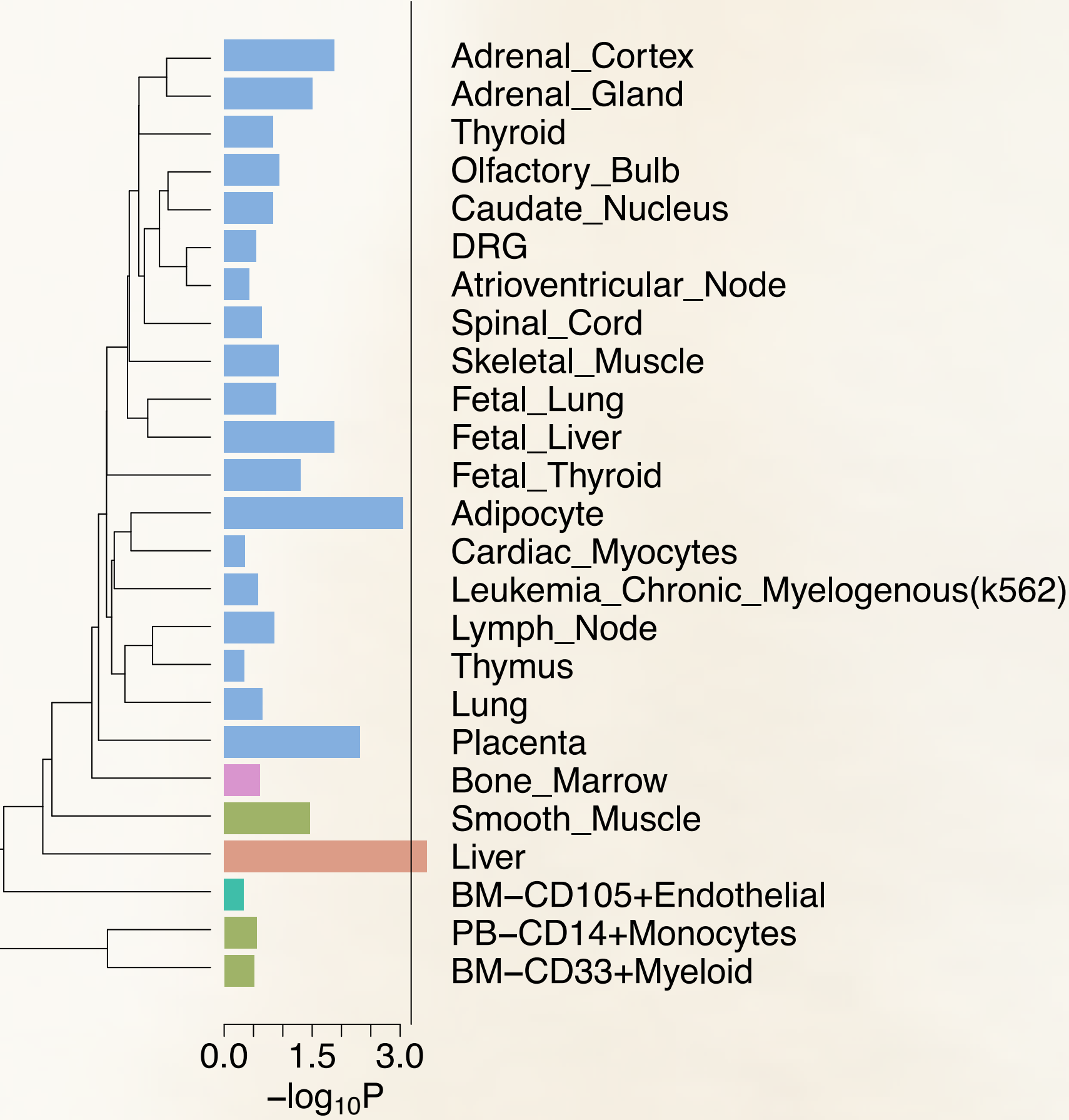
Genome-wide association studies (GWAS) have found many genomic loci associated with risk for different types of disease. SNPsea provides a simple way to determine cell types influenced by genes in these risk loci. Those cell types are likely to be relevant to the etiology of the given disease.

Suppose disease-associated alleles influence a small number of pathogenic cell types. **We hypothesize that genes with critical functions in those cell types are likely to be within risk loci for that disease.** We assume that a gene's specificity to a cell type is a reasonable indicator of importance to the unique function of that cell type.

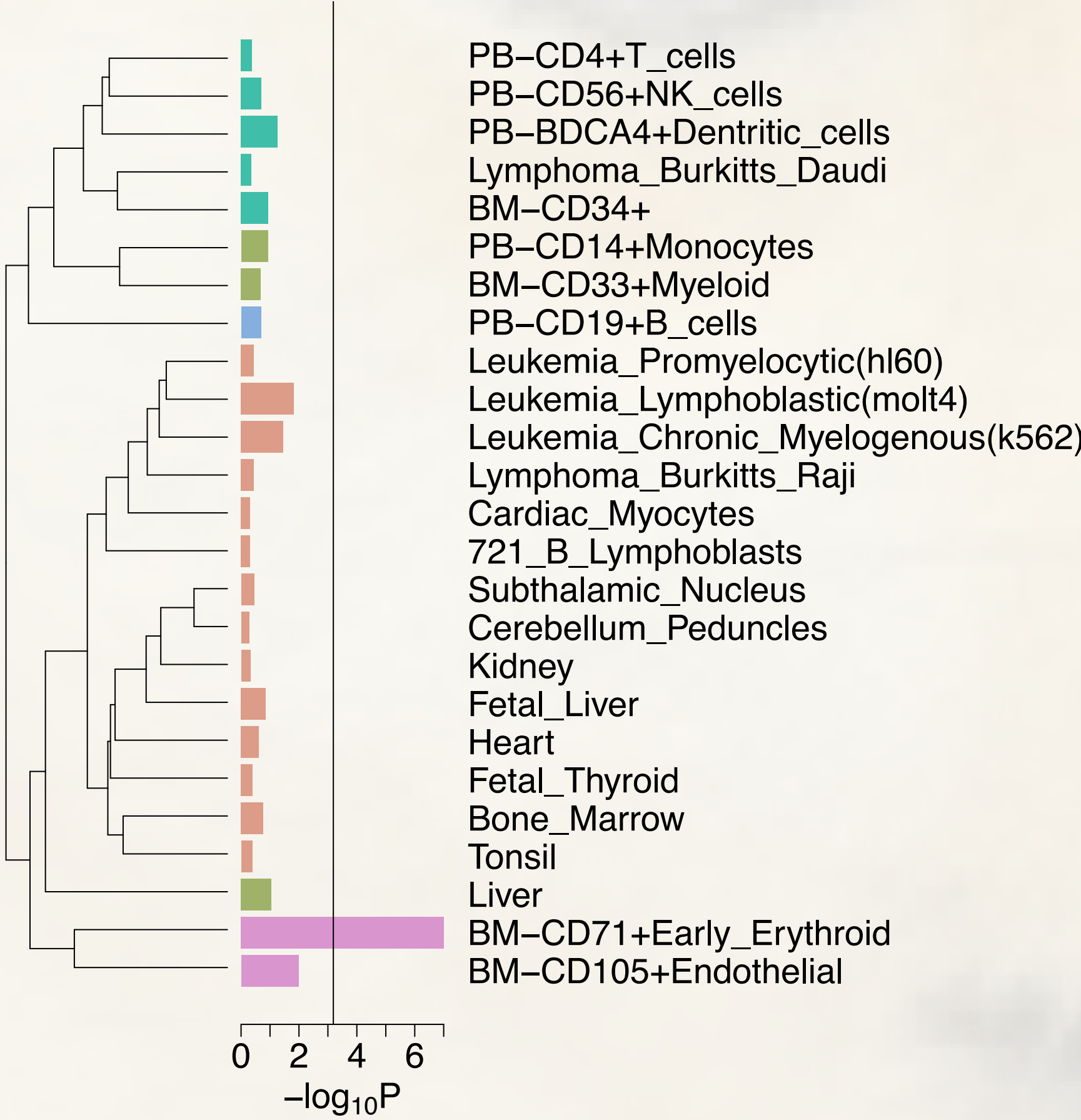
<http://broadinstitute.org/mpg/snpsea>

## Results

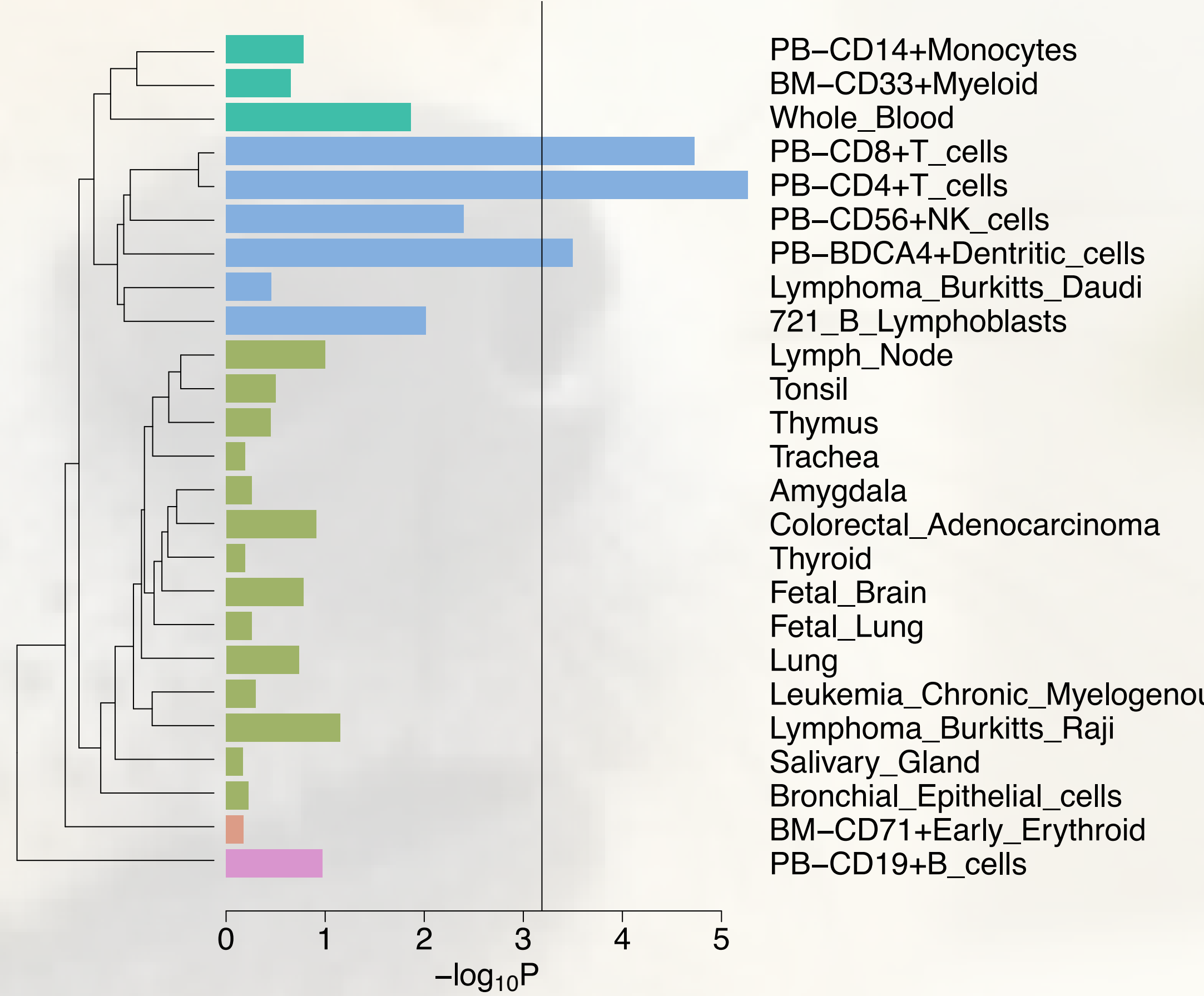
### HDL Cholesterol (46 SNPs)



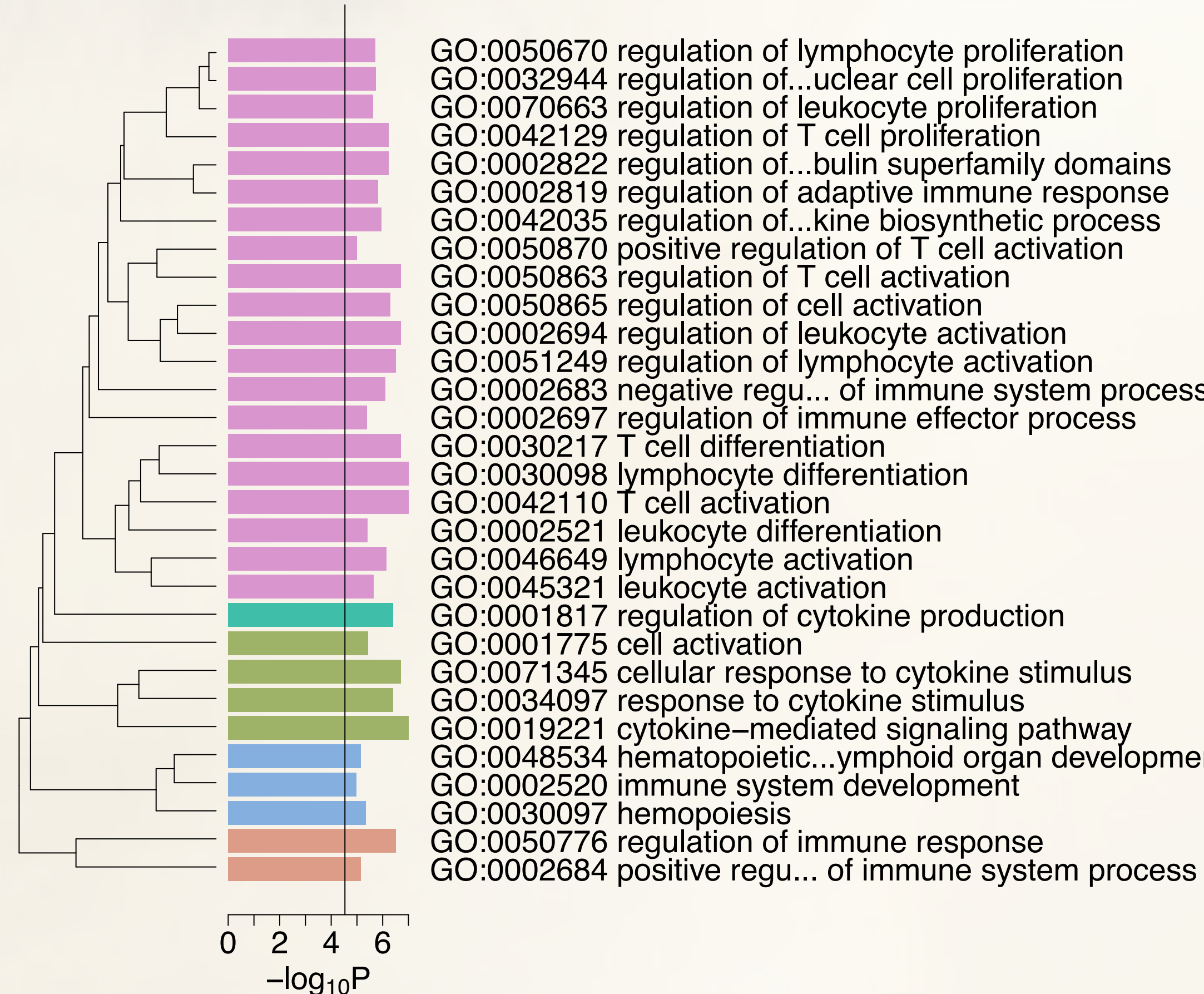
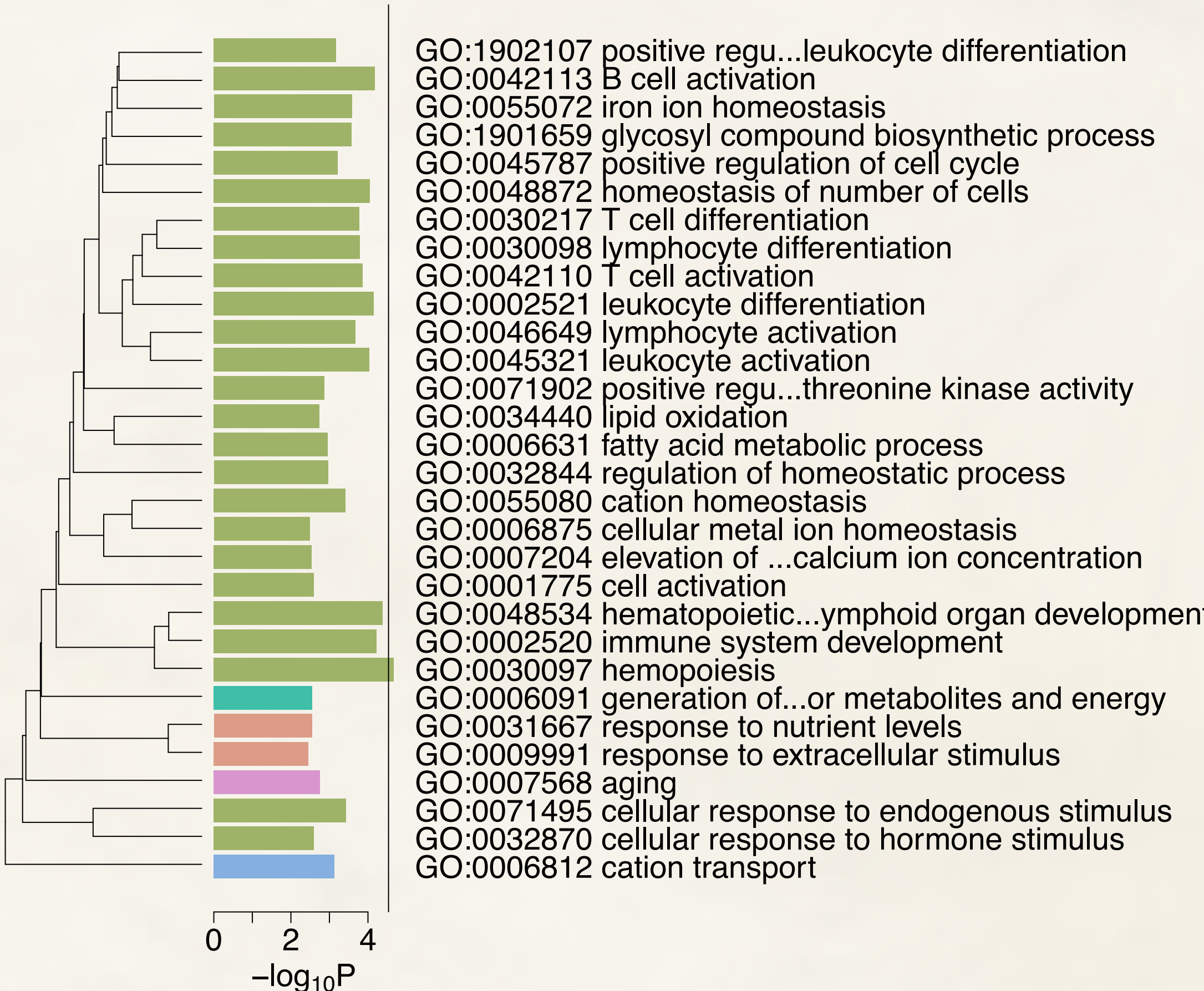
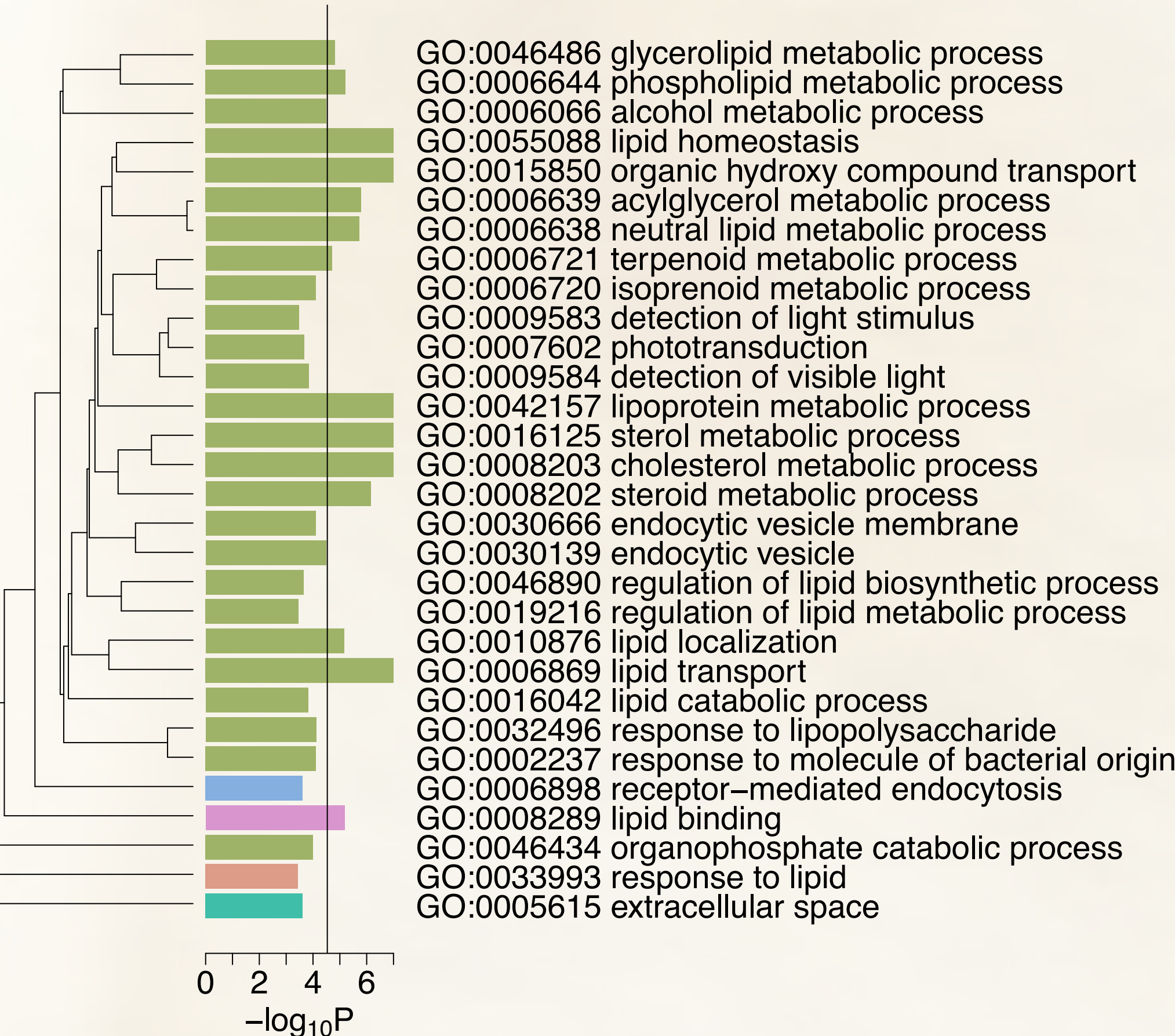
### Red Blood Cell-Count (45 SNPs)



### Celiac Disease (35 SNPs)



We tested each SNP set for cell type-specific gene expression relative to 78 other cell types or tissues in the **Gene Atlas** (Su et al. 2004). Top 25 are shown.



We also tested each SNP set for enrichment of gene annotations with 1,751 **Gene Ontology** (GO) terms (Botstein et al. 2000). Top 30 are shown.

## References

Elizabeth J. Rossin, et al. Proteins encoded in genomic regions associated with immune-mediated disease physically interact and suggest underlying biology. *PLoS Genet*, 7(1):e1001273, January 2011.  
The 1000 Genomes Project Consortium. An integrated map of genetic variation from 1,092 human genomes. *Nature*, 491(7422):56–65, November 2012.  
Simon Myers, et al. A fine-scale map of recombination rates and hotspots across the human genome. *Science*, 310(5746):321–324, October 2005.  
Xinli Hu, et al. Integrating autoimmune risk loci with gene-expression data identifies specific pathogenic immune cell subsets. *The American Journal of Human Genetics*, 89(4):496–506, 2011.  
Hana Lango Allen, et al. Hundreds of variants clustered in genomic loci and biological pathways affect human height. *Nature*, 467(7317):832–838, October 2010.  
Belinda Phipson and Gordon K. Smyth. Permutation p-values should never be zero: calculating exact p-values when permutations are randomly drawn. *Statistical Applications in Genetics and Molecular Biology*, 9(1), 2010.  
Pim van der Harst, et al. Seventy-five genetic loci influencing the human red blood cell. *Nature*, 492(7429):369–375, December 2012.  
Andrew I. Su, et al. A gene atlas of the mouse and human protein-encoding transcriptomes. *Proceedings of the National Academy of Sciences of the United States of America*, 101(16):6062–6067, April 2004. PMID: 15075390 PMCID: PMC395923.  
The International Multiple Sclerosis Genetics Consortium & The Wellcome Trust Case Control Consortium. Genetic risk and a primary role for cell-mediated immune mechanisms in multiple sclerosis. *Nature*, 476(7359):214–219, August 2011.  
Gosia Trynka, et al. Dense genotyping identifies and localizes multiple common and rare variant association signals in celiac disease. *Nature Genetics*, 43(12):1193–1201, December 2011.  
Tanya M. Teslovich, et al. Biological, clinical and population relevance of 95 loci for blood lipids. *Nature*, 466(7307):707–713, August 2010.

## Citation

Kamil Slowikowski, et al. **SNPsea: an algorithm to identify cell types, tissues, and pathways affected by risk loci.** *Bioinformatics* (2014) 30 (17): 2496–2497. doi: 10.1093/bioinformatics/btu326