

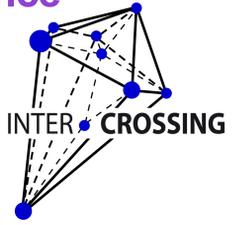
Signals enriched for positive selection

Alexandra I. Vatsiou, Eric Bazin, Oscar Gaggiotti

Laboratoire d'Ecologie Alpine, University Joseph Fourier, Grenoble, France

Era7 Bioinformatics, Granada, Spain

email: alex.vatsiou @ gmail.com



Introduction

Goal: Analyze a series of human data to examine how selection and adaptive events have influenced specific pathways. Testing a whole pathway could reveal much information rather than studying each locus or gene independently.

How: *Enrichment analysis* methods will be used to detect associations of networks of genes as the phenotype is usually the result of many interactions.

Data Used

- ◆ **SNP data:** Hapmap database phase II
http://hapmap.ncbi.nlm.nih.gov/downloads/phasing/2007-08_rel22/
3 populations (CEU, YRI, CHB+JTP)
- ◆ **Gene data:** Entrez NCBI database on the May 5/2014
<http://www.ncbi.nlm.nih.gov/gene/>
Final Number of genes: 27081
- ◆ **Gene Sets:** Biosystems database on the 6/5/2014.
<http://www.ncbi.nlm.nih.gov/biosystems>
Final number of pathways used: 2362

Enrichment analysis steps

XPCLR: Method to detect positive selection in pairwise populations. SNP data were used for 3 comparisons: CEU-YRI, CEU-CHB+JTP and YRI-CHB+JTP

Assignment of SNP scores to genes:

Assign a SNP to a gene, if the SNP is located within the gene transcript. Extend the start and end position by 50kb upstream or downstream away from the gene to account for intergenic region. Cluster the genes to bins according to the number of SNPs that they have. Normalize the z-score of each gene based on the distribution of the bin.

SUMSTAT score: Sum the scores of each gene in the pathway. We inferred empirical P-values for each gene set based on random gene set distribution.

Gowinda: Test for enrichment gene sets. Calculates the significance through permutation tests and the empirical FDR to account for multiple testing.

Significant Pathways

CEU-YRI (SUMSTAT)

1. Signal attenuation
2. txa2 pathway
3. Spliceosome, 35S U5-snRNP
4. Ubiquitin mediated proteolysis
5. Inactivation, recovery and regulation of the phototransduction cascade///The phototransduction cascade

YRI-JTP/CHB

1. Meiotic Recombination
2. Telomere Maintenance
3. Chromatin modifying enzymes

CEU-YRI (Gowinda)

1. Ubiquitin mediated proteolysis
2. Pertussis

CEU- JTP/CHB (Gowinda)

1. Glycolysis gluconeogenesis



Genes enriched for positive selection implicated with Diabetes/ Obesity and Metabolic Syndrome

Using Bio4j (<http://bio4j.com/>), an integrated open source bioinformatic data platform for protein information, we extracted 683 genes that could directly or indirectly be associated with obesity or metabolic syndrome or diabetes.

CEU-YRI

SUMO4
MEGF8

YRI-JTP/CHB

KDM3A

CEU-JTP/CHB

IFIH1
BLK
FTH1

MEGF8
SLC2A10

Future Directions

- Repeat the analysis with iHS and xp-EHH.

Acknowledgements

We thank the INTERCROSSING for funding this project and Raquel Tobes for extracting the genes list from Bio4j.

References

- Chen, H., Patterson, N., and Reich, D. (2010). Population differentiation as a test for selective sweeps. *Genome research* 20, 393-402.
- Kofler, R., and Schlotterer, C. (2012). Gowinda: unbiased analysis of gene set enrichment for genome-wide association studies. *Bioinformatics* 28, 2084-2085.
- Daub, J.T., Hofer, T., Cutivet, E., Dupanloup, I., Quintana-Murci, L., Robinson-Rechavi, M., and Excoffier, L. (2013). Evidence for polygenic adaptation to pathogens in the human genome. *Molecular biology and evolution* 30, 1544-1558.