FROM RESEARCH TO INDUSTRY

VARIANT PRIORITIZATION WITH GENOWAP TOOL

ThinkLab | Thomas Riquelme

10 SEPTEMBRE 2018

SUMMARY

1) Intro – the necessity to prioritize GWAS identified variants
2) Choice of GenoSuite tools
3) GenoCanyon - General-functional annotation tool
4) GenoSkyline - Tissue-specific functional Annotation
5) GenoWAP - GWAS Signal Prioritization
6) GenoWAP pipeline
7) Comparison GWAS SNP pval VS GWAS SNP prioritized scores

INTRODUCTION

INTRODUCTION

- GWAS have identified more than 10,000 SNPs associated with numerous traits/diseases.
- Variant prioritization techniques are crucial for post-GWAS analysis
- To reveal truly functional sites within each significant locus
- To enhance signals at some loci to identify risk loci among all the SNPs known to be associated with a trait.

INTRODUCTION

⇒ Choice of GenoSuite tools (Lu et al.)

- prioritization based on the functionality of a DNA region
- Functional regions = regions which can regulate gene expression
- e.g. promoter, enhancer, silencer, insulator, transcription factor binding site, chromatin regulators

METHOD

GENOCANYON - General-functional annotation

- Whole-genome functional annotation approach
- Predicts the functional potential at each nucleotide : gives a scores (0-1)
- Integrates 22 annotation signals :
  - 2 genomics conservations measures (GERP, PhyloP),
  - 2 indicators of open-chromatine (DnaseI, FAIRE),
  - 8 histone modifications (H3K4me3, H3K9ac…)
  - 10 TFBS (CEBPB, CTCF, FOS, GATA2...)

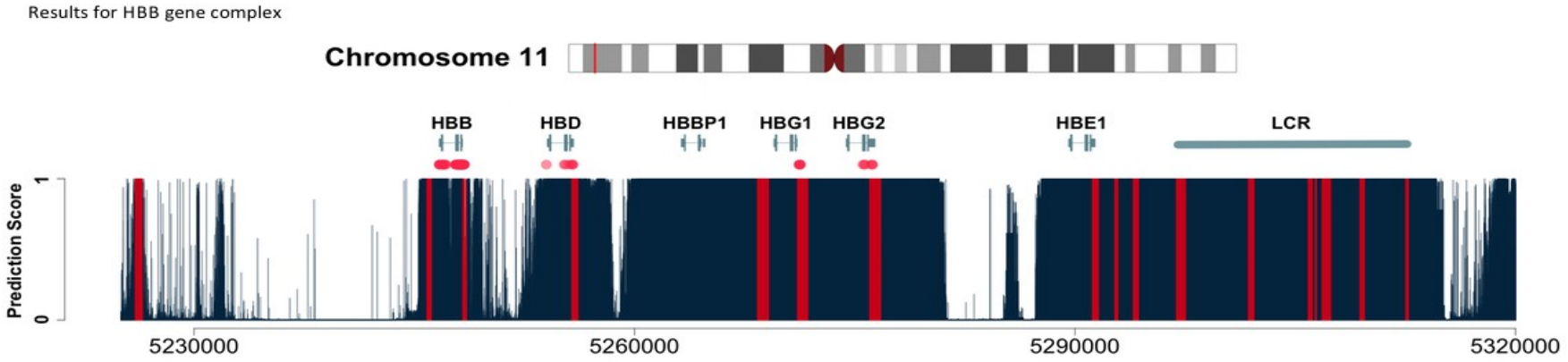


Fig1: Example with the prediction for cis-regulatory Modules in the β -globin (HBB) gene complex on chromosome 11
In red : already discovered cis regualtory modules.
In dark blue : prediction score at each location.

Picture from Lu et al. (2015). A statistical framework to predict functional non-coding regions in the human genome through integrated analysis of annotation data. Scientific reports, 5, 10576.

GENOCANYON - General-functional annotation

- Method : unsupervised statistical learning
- Uses Bayes formula to compute the posterior probability that a locus is functional given the annotations by taking into account the probability of presence of these annotations when the locus is known to be functional.

$$P(Z = 1|\boldsymbol{A}) = \frac{\pi f(\boldsymbol{A}|Z = 1)}{\pi f(\boldsymbol{A}|Z = 1) + (1 - \pi) f(\boldsymbol{A}|Z = 0)}$$

Formula1: Posterior probability calculation that a locus is functional given the annotations
Z= measure of functionnality
A= vector of annotations = (A1, A2 … A22)

Formula from Lu et al. (2015). A statistical framework to predict functional non-coding regions in the human genome through integrated analysis of annotation data. Scientific reports, 5, 10576.

GENOSKYLINE - Tissue-specific Functional Annotation

- Declination of GenoCanyon
- Predicts tissue-specific functional scores at each nucleotide
- Based on 111 epigenomes database from the Epigenomics Roadmap Project
- Each tissue type is a clustering of relevant samples in order to contain at least these histone modifications : H3k4me1, H3k4me3, H3k36me3, H3k27me3, H3k9me3, H3k27ac, H3k9ac, and DNase I Hypersensitivity
- 7 unique tissue clusters : Brain, GI, Lung, Heart, Blood, Muscle, Epithelium
- Extension of the available tissues with GenoSkyline-Plus : Bone, Breast, Fat, Muscle, Kidney, Liver, Ovary, Pancreas, Skin, Spleen, Thymus, Vascular...
- Useful application : identify if GWAS SNPs are enriched in functional region specific for a tissue

GENOSKYLINE - Tissue-specific Functional Annotation

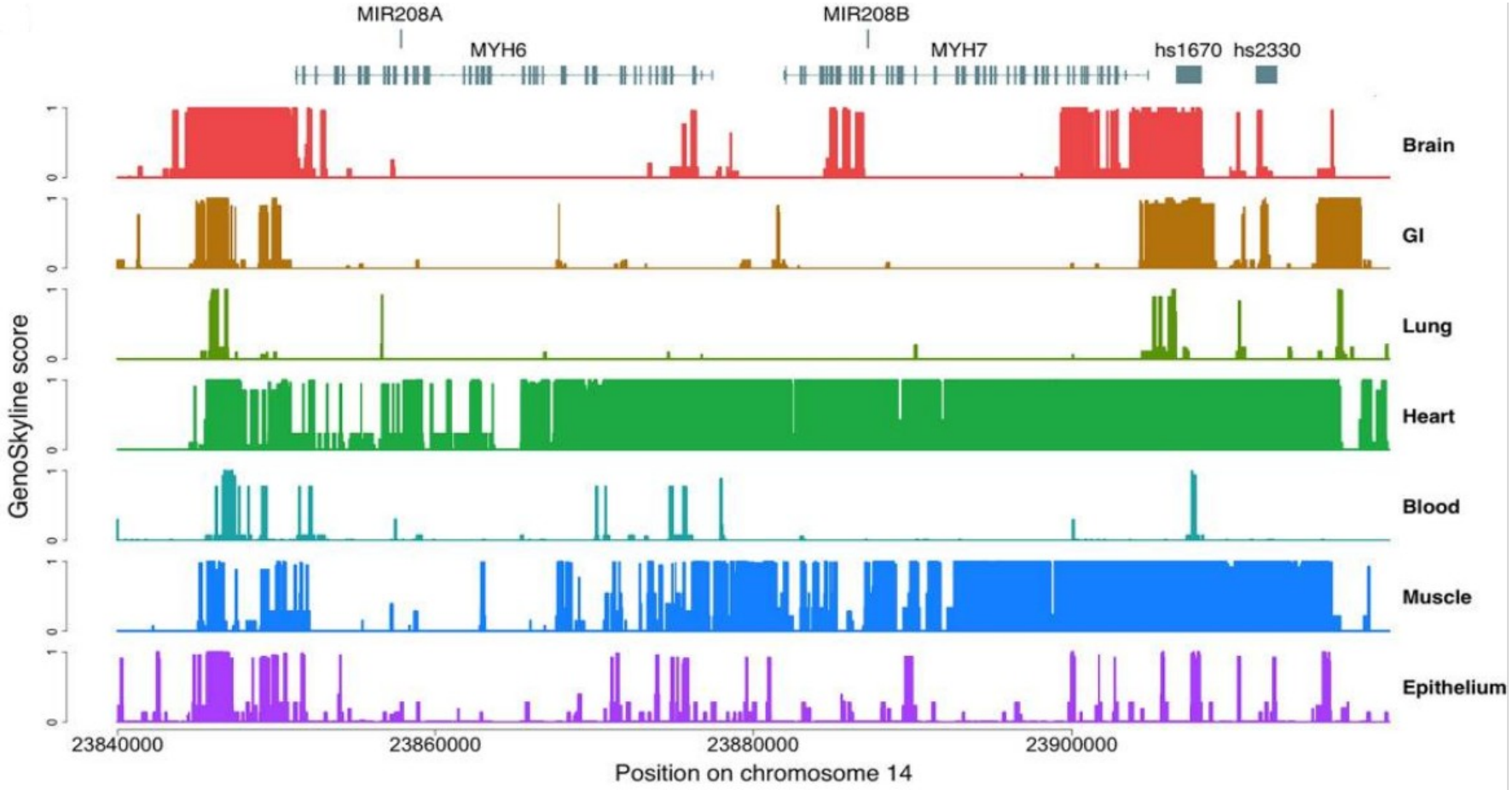


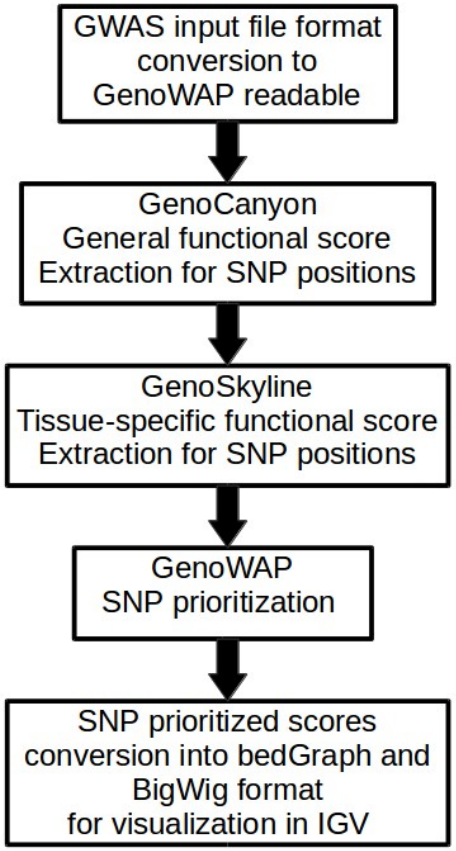
Fig1: Visualisation of the functionality tracks for seven tissues in the genomic region surrounding MYH6 and MYH7

Picture from Lu et al. (2016). Integrative tissue-specific functional annotations in the human genome provide novel insights on many complex traits and improve signal prioritization in genome wide association studies. PLoS genetics, 12(4), e1005947.

GWAS Signal Prioritization

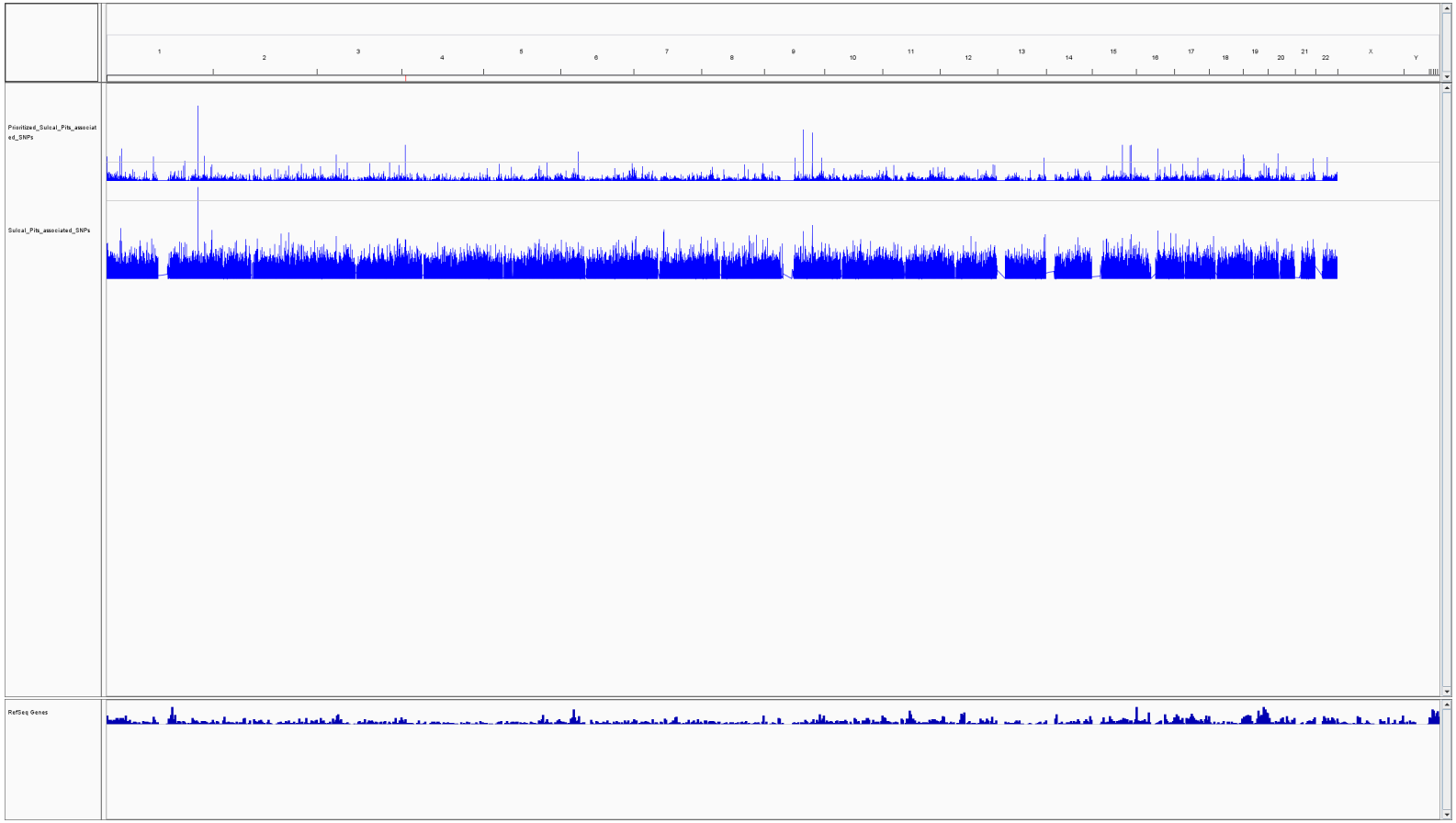
- Post-GWAS prioritization method
- Integrates genomic functional annotation and GWAS test statistics

GENOWAP PIPELINE



RESULTS

RESULT VISUALIZATION



Comparison GWAS SNP pval VS GWAS SNP prioritized scores (snapshot from IGV)

REFERENCES

- GenoCanyon paper :
  Lu, Q., Hu, Y., Sun, J., Cheng, Y., Cheung, K. H., & Zhao, H. (2015). A statistical framework to predict functional non-coding regions in the human genome through integrated analysis of annotation data. Scientific reports, 5, 10576.

- GenoSkyline paper :
  Lu, Q., Powles, R. L., Wang, Q., He, B. J., & Zhao, H. (2016). Integrative tissue-specific functional annotations in the human genome provide novel insights on many complex traits and improve signal prioritization in genome wide association studies. PLoS genetics, 12(4), e1005947.

- GenoWAP paper :
  Lu, Q., Yao, X., Hu, Y., & Zhao, H. (2015). GenoWAP: GWAS signal prioritization through integrated analysis of genomic functional annotation. Bioinformatics, 32(4), 542-548.

Commissariat à l'énergie atomique et aux énergies alternatives
Centre de Saclay | 91191 Gif-sur-Yvette Cedex

DRF/JOLIOT
NEUROSPIN
UNATI-Brainomics

Etablissement public à caractère industriel et commercial | RCS Paris B 775 685 019