Cell, Volume *166*

Supplemental Information

The Genetics of Transcription Factor DNA Binding Variation

Bart Deplancke, Daniel Alpern, and Vincent Gardeux

Table S1. Comparison of Computational Methods for Assessing the Impact of Regulatory Variants Based on DNA Sequence

Method	Type	Input	Statistical assessment	Software	Motif catalog	Reference
RAVEN	PWM	Gene, SNV, indel	N	Web	JASPAR	(Andersen et al., 2008)
sTRAP	PWM	SNV, indel	Y	Web	JASPAR, TRANSFAC	(Manke et al., 2010)
is-rSNP / is-rINDEL	PWM	SNV, indel	Y	Web / Web	JASPAR, TRANSFAC	(Macintyre et al., 2010)
rSNP-MAPPER	HMM	Gene, SNV	N	Web	MAPPER, TRANSFAC, JASPAR	(Riva, 2012)
regSNPs	PWM[1]	SNP	Y	R package	TRANSFAC	(Teng et al., 2012)
PMCA	PWM[2]	SNP	Y	NA	Genomatix	(Claussnitzer et al., 2014)
motifbreakR	PWM	SNV	Y	R package	MotifDb package, user-defined	(Coetzee et al., 2015)
BayesPI-BAR	PWM[3]	SNV, indel	Y	Batch	ENCODE, user-defined	(Wang and Batmanov, 2015)
atSNP	PWM	SNV	Y	R package	ENCODE, JASPAR, user-defined	(Zuo et al., 2015)
deltaSVM	SVM	SNV, indel	N	Batch	No Motif. k-mer vocabulary from ENCODE DNaseI-seq, ChIP-seq	(Lee et al., 2015)
GERV	PR[4]	SNV, indel	N	Batch	No Motif. k-mer vocabulary from ENCODE DNaseI-seq, ChIP-seq	(Zeng et al., 2015)
DeepBind	NN	SNV, indel	N	Batch	built from PBM, SELEX, ChIP-seq, CLIP-seq	(Alipanahi et al., 2015)
DeepSea	NN	SNV, indel	Y	Web	No Motif. ENCODE & Roadmap Epigenomics (TF, DHS, chromatin mark)	(Zhou and Troyanskaya, 2015)
HaploReg	PWM	SNP, indel	N	Web	TRANSFAC, JASPAR, built from PBM	(Ward and Kellis, 2016)

PWM: Position Weight Matrix; SNP: Single Nucleotide Polymorphism; SNV: Single Nucleotide Variation; indel: insertion and deletion; TF: Transcription Factor; DHS: DNase I Hypersensitivity Site; NA: Not Available; [Machine Learning Approaches]: PR: Poisson Regression; HMM: Hidden Markov Model; NN: Neural Network; SVM: Support Vector Machine;

[1]PWM differential score is then mixed with a gene prioritization score based on pathway analysis

[2]PWMs are used for identifying conserved TFBS across species. Sequence context of SNPs are scored based on conservation of TFBS modules

[3]Differential Binding Affinity score (ðdbA) also incorporates protein concentration and chemical properties from ChIP-seq data

[4]Regression predicts the ChIP-seq signal at every nucleotide position, then scoring is computed as the difference between predicted peaks

References

Alipanahi, B., Delong, A., Weirauch, M.T., and Frey, B.J. (2015). Predicting the sequence specificities of DNA- and RNA-binding proteins by deep learning. Nat Biotech *33*, 831-838.

Andersen, M.C., Engstrom, P.G., Lithwick, S., Arenillas, D., Eriksson, P., Lenhard, B., Wasserman, W.W., and Odeberg, J. (2008). In silico detection of sequence variations modifying transcriptional regulation. PLoS Comput Biol *4*, e5.

Claussnitzer, M., Dankel, Simon N., Klocke, B., Grallert, H., Glunk, V., Berulava, T., Lee, H., Oskolkov, N., Fadista, J., Ehlers, K.*, et al.* (2014). Leveraging Cross-Species Transcription Factor Binding Site Patterns: From Diabetes Risk Loci to Disease Mechanisms. Cell *156*, 343-358.

Coetzee, S.G., Coetzee, G.A., and Hazelett, D.J. (2015). motifbreakR: an R/Bioconductor package for predicting variant effects at transcription factor binding sites. Bioinformatics *31*, 3847-3849.

Lee, D., Gorkin, D.U., Baker, M., Strober, B.J., Asoni, A.L., McCallion, A.S., and Beer, M.A. (2015). A method to predict the impact of regulatory variants from DNA sequence. Nat Genet *advance online publication*.

Macintyre, G., Bailey, J., Haviv, I., and Kowalczyk, A. (2010). is-rSNP: a novel technique for in silico regulatory SNP detection. Bioinformatics *26*, i524-530.

Manke, T., Heinig, M., and Vingron, M. (2010). Quantifying the effect of sequence variation on regulatory interactions. Hum Mutat *31*, 477-483.

Riva, A. (2012). Large-scale computational identification of regulatory SNPs with rSNP-MAPPER. BMC genomics *13 Suppl 4*, S7.

Teng, M., Ichikawa, S., Padgett, L.R., Wang, Y., Mort, M., Cooper, D.N., Koller, D.L., Foroud, T., Edenberg, H.J., Econs, M.J.*, et al.* (2012). regSNPs: a strategy for prioritizing regulatory single nucleotide substitutions. Bioinformatics *28*, 1879-1886.

Wang, J., and Batmanov, K. (2015). BayesPI-BAR: a new biophysical model for characterization of regulatory sequence variations. Nucleic Acids Research.

Ward, L.D., and Kellis, M. (2016). HaploReg v4: systematic mining of putative causal variants, cell types, regulators and target genes for human complex traits and disease. Nucleic Acids Res *44*, D877-881.

Zeng, H., Hashimoto, T., Kang, D.D., and Gifford, D.K. (2015). GERV: a statistical method for generative evaluation of regulatory variants for transcription factor binding. Bioinformatics.

Zhou, J., and Troyanskaya, O.G. (2015). Predicting effects of noncoding variants with deep learning-based sequence model. Nat Meth *12*, 931-934.

Zuo, C., Shin, S., and Keles, S. (2015). atSNP: transcription factor binding affinity testing for regulatory SNP detection. Bioinformatics *31*, 3353-3355