

Parameter

b - bed file, optionally with activity column(s) for cell types / metacells

g - fasta-file with the genome sequence

p - PSEMs of TFs **<OR>** ***s*** - PSCM as transfac and ***y*** - GC content

a - gene annotation as gtf-file

o - output prefix, creates a folder

optional:

n - activity column(s), start counting at 1, e.g. for metacells

c - number of cores to use for computation

x - bed-file with regions to exclude

w - window size centred at 5' TSS (default 50kb for 'Gene window' and 5MB for 'ABC-Scoring')

e - scale for distance (default True), only w/o ABC-mapping

f - folder with normalized Hi-C contacts

k - bin-size of Hi-C files, required for ABC-Scoring

t - cut-off for ABC-scored interactions (default 0.02)

q - use the adapted ABC-score (default True)

m - window size for -q adaptation (5MB, set at least to -w)

d - add pseudocount to Hi-C contacts (default True)

r - ABC-Score file, if already calculated in advance

$$af_{g,tf} = \sum_{r \in R_g} \frac{af_{r,tf}}{ml_{tf}} \cdot scaler$$

scaler depends on approach:

Gene window

$$A_r \cdot e^{-\frac{d_{r,g}}{d_0}} \quad (f)$$

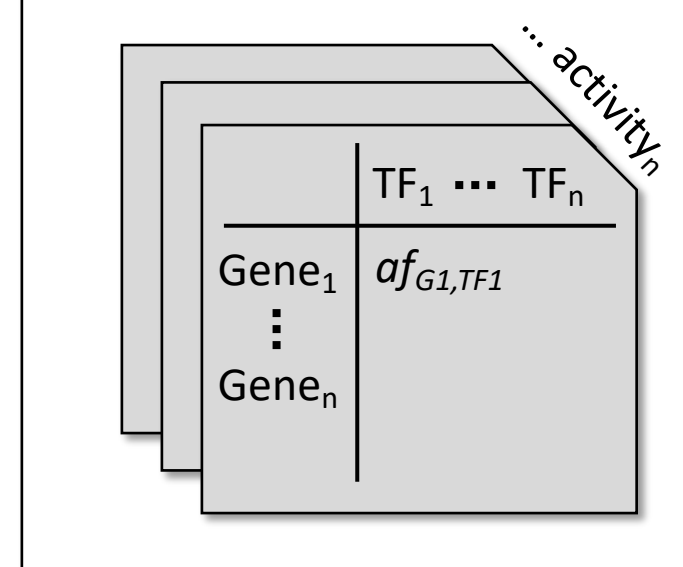
adapted \overline{ABC} -scoring (*q*)

$$\begin{cases} A_r \cdot e^{-\frac{d_{r,g}}{d_0}}, & \text{if } d_{r,g} \leq 2500bp \\ A_{r,g}, & \text{otherwise} \end{cases}$$

ABC-scoring

$$\begin{cases} A_r \cdot e^{-\frac{d_{r,g}}{d_0}}, & \text{if } d_{r,g} \leq 2500bp \\ A_r \cdot \frac{C_{r,g}}{C_{max}}, & \text{otherwise} \end{cases}$$

Gene-TF matrices



$af_{g,tf}$: affinity score of TF *tf* to *g*

R_g : set of regions mapped to *g*

$af_{r,tf}$: affinity of *tf* in *r*

ml_{tf} : motif length of *tf*

A_r : activity of *r*

$A_{r,g}$: adapted activity of *r* to *g*

$d_{r,g}$: distance of *r* to *g*

d_0 : distance constant of 5000 bp

$C_{r,g}$: contact of *r* with *g*

C_{max} : maximum $C_{r,g}$



[1] adapted from: Florian Schmidt, Fabian Kern, Peter Ebert, Nina Baumgarten, Marcel H Schulz, TEPIIC 2—an extended framework for transcription factor binding prediction and integrative epigenomic analysis, *Bioinformatics*, Volume 35, Issue 9, 1 May 2019, Pages 1608–1609, <https://doi.org/10.1093/bioinformatics/bty856>; <https://github.com/SchulzLab/TEPIIC>

[2] adapted from: Fulco CP, Nasser J, Jones TR, Munson G, Bergman DT, Subramanian V, Grossman SR, Anyoha R, Doughty BR, Patwardhan TA, Nguyen TH, Kane M, Perez EM, Durand NC, Lareau CA, Stamenova EK, Aiden EL, Lander ES & Engreitz JM. Activity-by-contact model of enhancer–promoter regulation from thousands of CRISPR perturbations. *Nat. Genet.* 51, 1664–1669 (2019). <https://www.nature.com/articles/s41588-019-0538-0> <https://github.com/broadinstitute/ABC-Enhancer-Gene-Prediction>