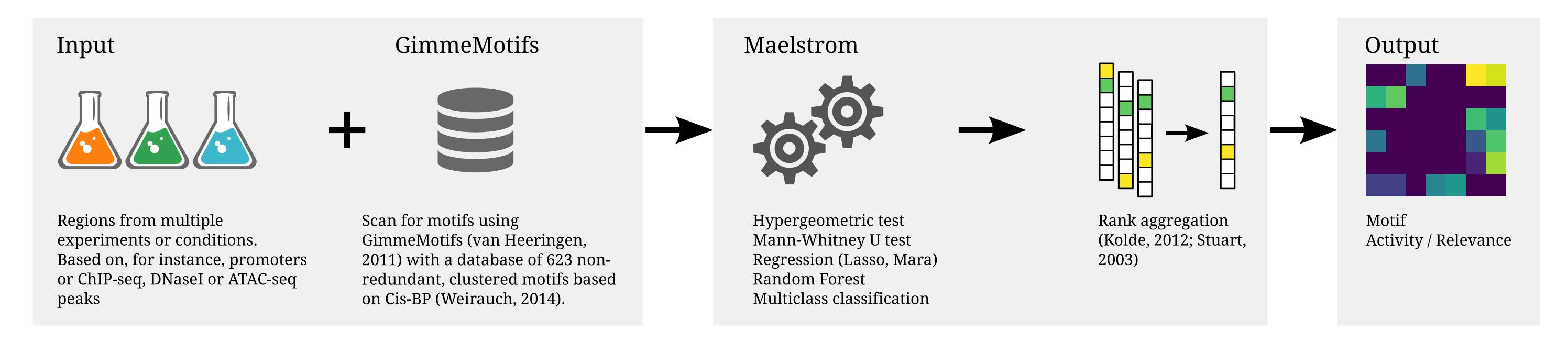
# Motif activity by ensemble learning

# for multi-experimental transcription factor motif analysis

Simon J. van Heeringen

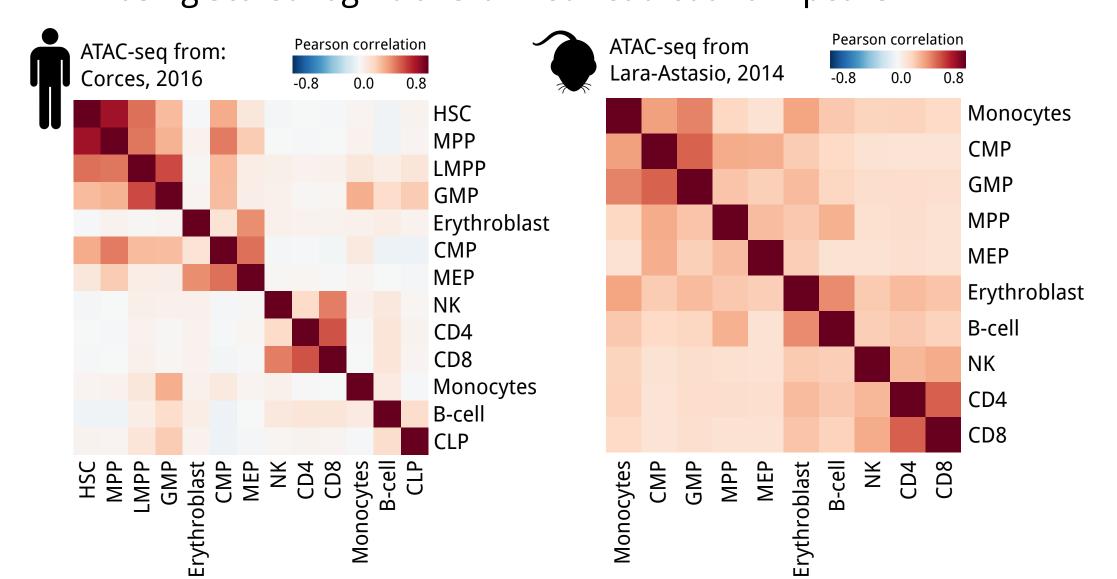


# Maelstrom: an ensemble method to reliably predict relevant regulatory motifs



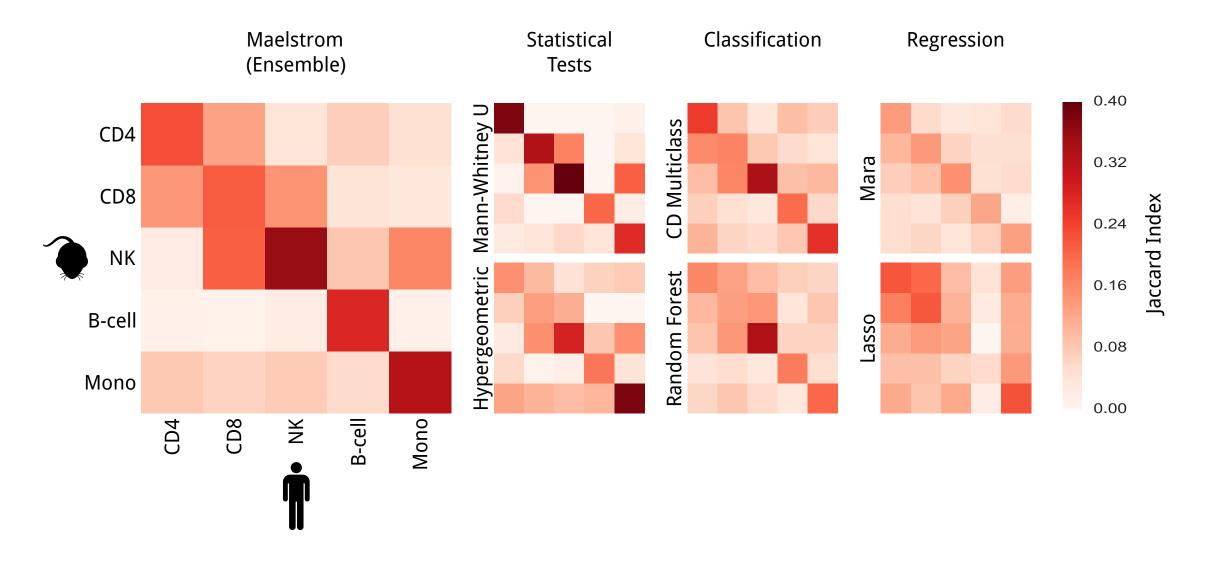
#### Example data: ATAC-seq in hematopoiesis

Correlation between cell types using scaled log2-transformed read count in peaks

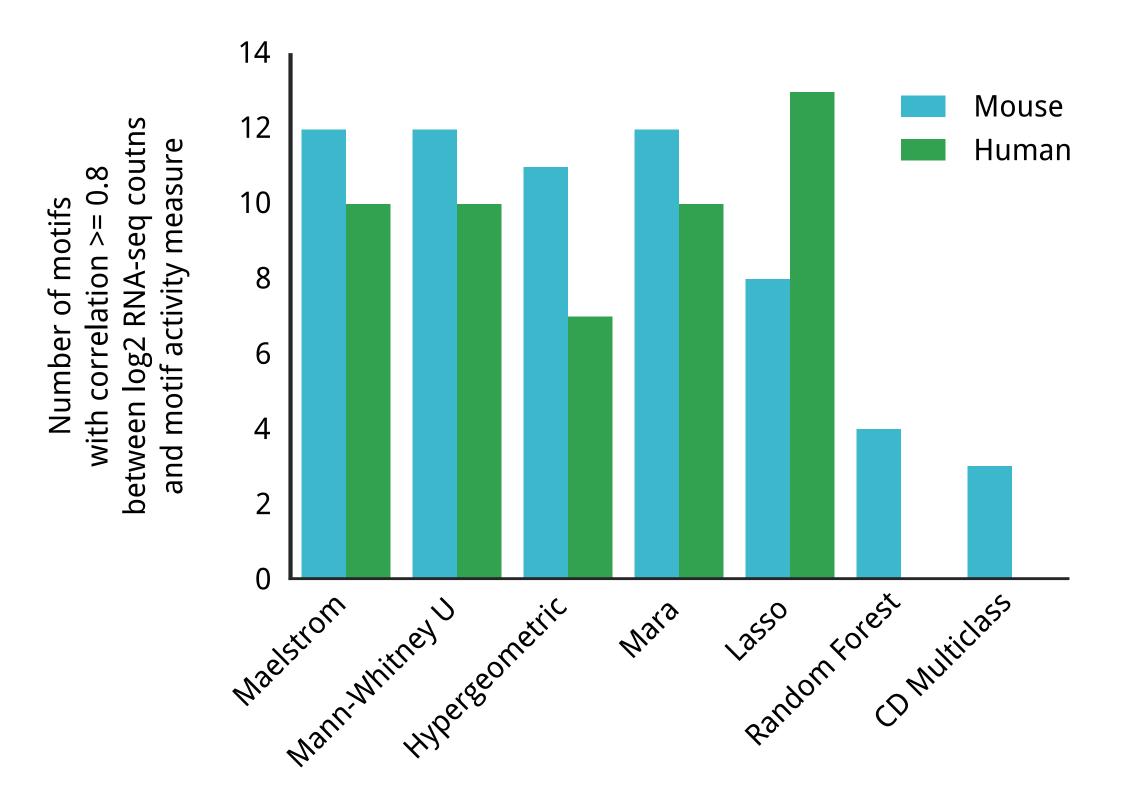


#### Motifs identified by maelstrom are conserved

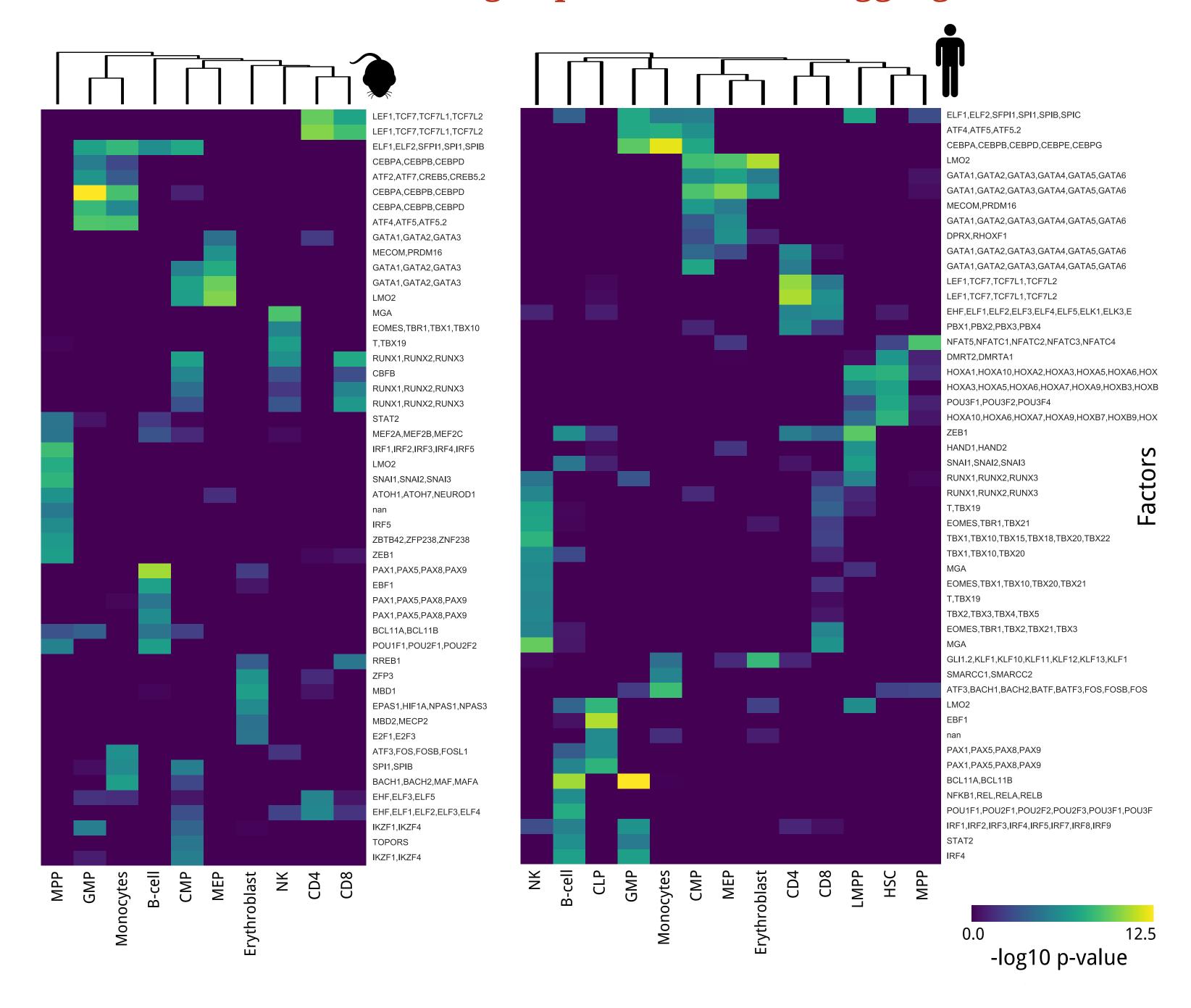
Common motifs between human and mouse (Jaccard index)



# Correlation of motifs with TF gene expression



# Maelstrom output: heatmap of motif relevance or activity (-log10 p-value of rank aggregation)



### Take-home points

- Maelstrom enables straightforward analysis of differential transcription factor motifs.
- Different methods to estimate relevant transcription factor methods are included; based on statistical tests, regression and classification.
- From differential ATAC-seq peaks, maelstrom identifies robust motifs in the hematopoietic lineage that are conserved between mouse and human.
- Different measures of motif activity or significance result in similar motif numbers that correlate with TF gene expression (except classificationbased methods)
- Development version of maelstrom is available: http://github.com/simonvh/gimmemotifs.
- Questions and feedback welcome! (github, e-mail, Twitter)

Affiliations: Radboud University, Department of Molecular Developmental Biology, Faculty of Science, Nijmegen, The Netherlands. Funding: the Netherlands Organization of Scientific Research (NWO-ALW)



http://github.com/simonvh/maelstrom-poster