Analysis of Transcription Factor Interaction in Humans, Chimpanzees and Rhesus Macaques

Graphs and Biological Networks – Practical Course

Felix Kühnl

University of Leipzig

August 2013 – organized by
Stefano Berto
Katja Nowick
Konstantin Klemm

Introduction

- Human vs. Chimp: only about 1% difference in gene sequences
- but: many differences in phenotype, especially in brain and mental capabilities
- cause: different expression levels of genes in different species

↷ controlled by transcription factors (TFs)

Transcription Factors

- TFs: proteins that control expression of genes
- key to understand how the different phenotypes arise
- Questions:
    - How do TFs interact to regulate gene expression?
    - Does TF interaction differ between the species?
    - Is the TF interaction dependent on individual selection?

Analysis of TF Interaction

- dataset: expression level of 12,100 genes (incl. 280 TFs) for 12 individuals of each species (human, chimp and rhesus macaque)
- for any TF i and any gene j in the dataset let $a_{ij} = \rho_{ij} = cor(x_i, y_j) \in [-1, 1]$, where $x_i, y_j \in \mathbb{R}^{12}$ are the expression levels and $cor()$ is Spearman's ranked correlation coefficient

↷ interpret $(a_{ij}) = A \in [-1, 1]^{i \times j}$ as adjacency matrix of TF–gene interaction network (possibility: set $a_{ij} < \rho^*$ to 0)

- to mesure interaction between TFs i, j in species s, calculate topological overlap

$$\omega_{ij}^{s} = \frac{\sum_{u \in \text{Genes}} a_{iu} a_{ju} + a_{ij}}{\min(k_i, k_j) + 1 - |a_{ij}|} \in [-1, 1]$$

where $k_i = \sum_{j \in \text{Genes}} |a_{ij}|$ is the connectivity of TF i

Dependence on Individual Selection

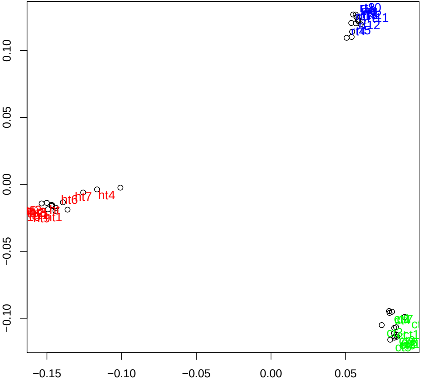
- analyse dependence of TF interaction: *leave-one-out* approach
- ↷ for each individual $x \in \{1, \ldots, 12, \emptyset\}$ calculate ω_{ij}^{sx} *without* expression data for this individual
- result: 3·12=36 new + 3 old = 39 TF interaction networks Ω^{sx}
- comparison of these: calculate distance

$$d_{\Omega^{sx}\Omega^{s'x'}} = \frac{1}{n^2} \sum_{ij} |\omega_{ij}^{sx} - \omega_{ij}^{s'x'}|$$

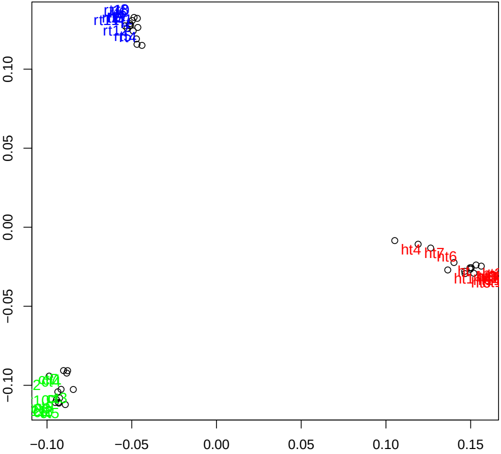
where n is the number of TFs

At a glance, ρ unfiltered

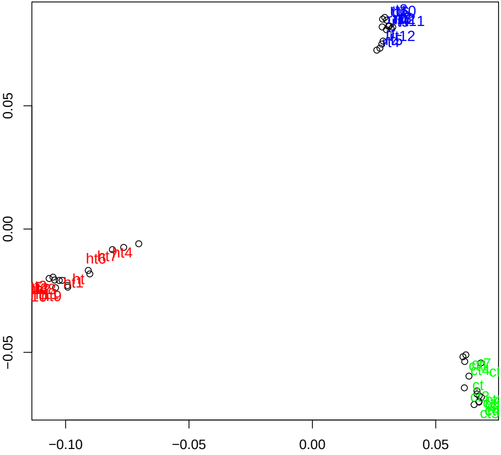
- *classical multi-dimensional scaling* of distance matrix d
- *labels:* red: human, green: chimp, blue: rhesus macaque
- dense clusters of original and *leave-one-out* networks for each species, but huge interspecies distances



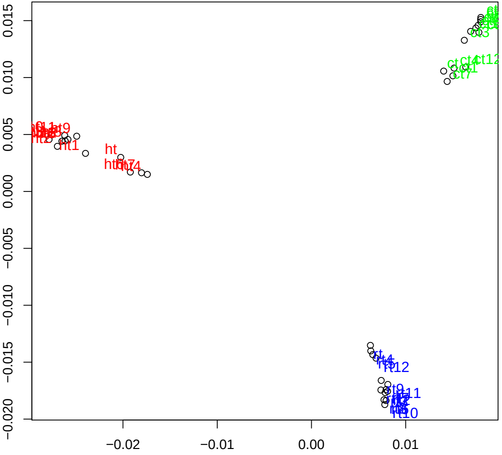
At a glance, $\rho^* = 0.4$



At a glance, $\rho^* = 0.6$

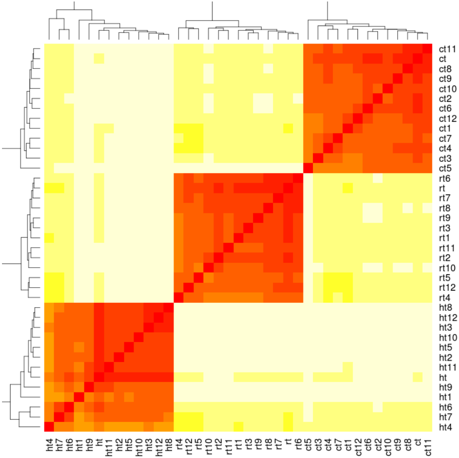


At a glance, $\rho^* = 0.8$

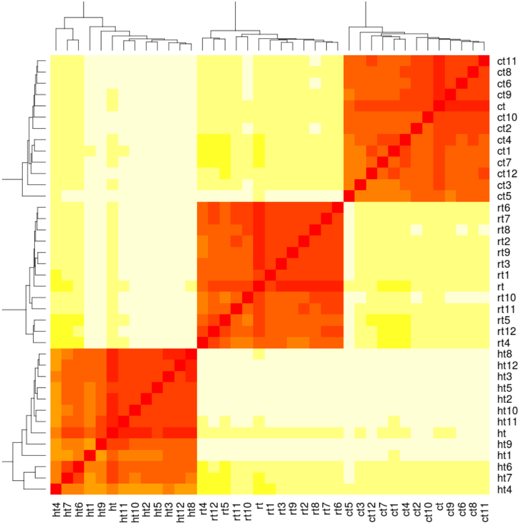


At a glance

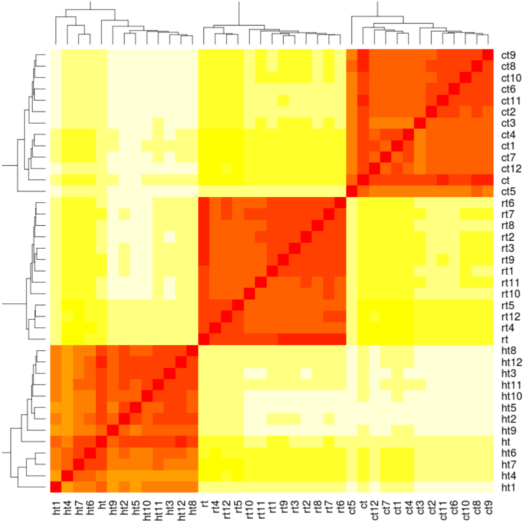
- *heatmap* of distances, dark color indicates low distance
- ht: human, ct: chimp, rh: rhesus macaque, number: left out individual



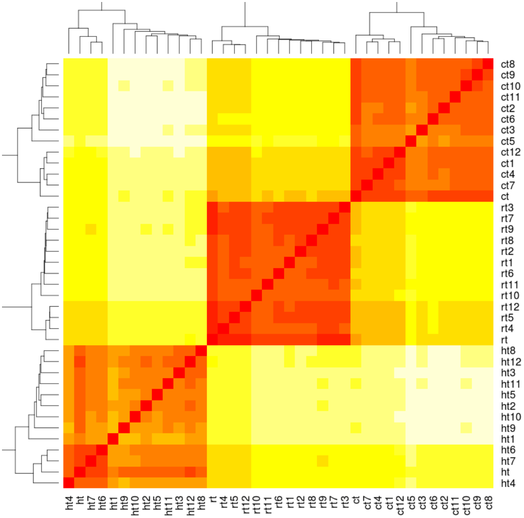
At a glance, $\rho^* = 0.4$



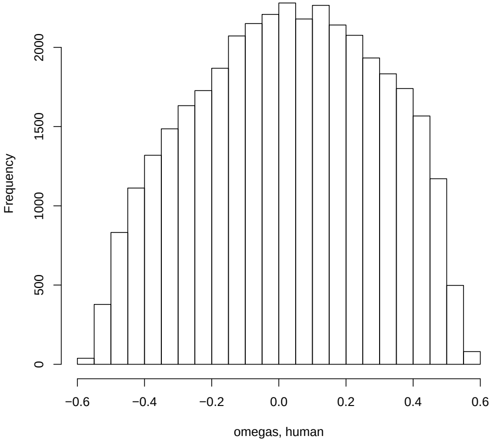
At a glance, $\rho^* = 0.6$



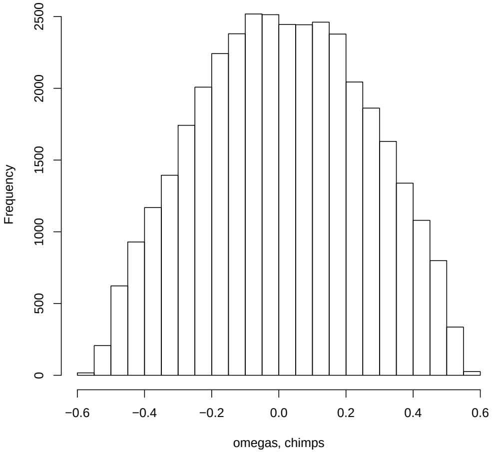
At a glance, $\rho^* = 0.8$



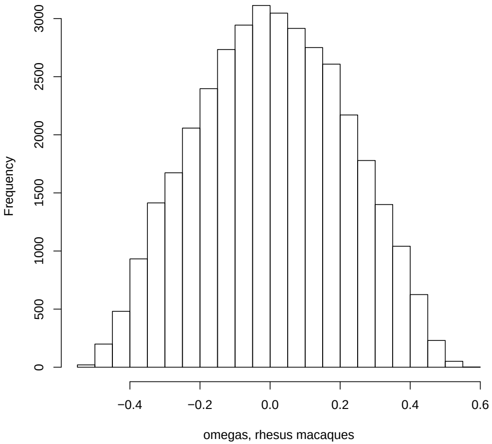
Distribution of ω, Human



Distribution of ω, Chimp



Distribution of ω, Rhesus Macaque



References

- gene expression data by Stefano Berto
- distance formulae by Katja Nowick