# Reproducible methods for network analysis of high-throughput genomic data

**Foreword** (Motivation, personal context)

**Table of Contents**

**Abstract** (main objective, result and scope)

1. **Introduction** (Objectives, global approach)
   1. Motivation to work with biological networks (Subject and interest of work)
   2. Biological background in immunology
      1. The role of CD4+ T helper cells in the immune system
      2. Transcription factors and their regulatory interactions
   3. High-throughput data and the analysis of transcription factor activity
      1. ChIP-seq for analysis of direct transcription factor targets
      2. RNA-seq / DESeq for analysis of functional targets
   4. State of Research (foundation for project)
      1. The basis for reproducible and reusable methods for network analysis of high-throughput genomic data [Ciofiani 2012 paper]
      2. Processing called peaks from ChIP-seq analysis (Poisson model)
      3. Combination of different NGS data types (idea, purpose)
2. **Methods**
   1. Software environment (R, git, bash)
   2. Usability and reproducibility
      1. Project setup and structure (setup-script)
      2. NCBI GEO data download and data folders
   3. Network generation algorithm
      1. NGS input data
      2. Initial parsing and processing
      3. Filtering of target genes
      4. Distinction of activator and repressor matrices
      5. Data integration by quantile ranking
      6. Combination of ranked data type matrices
      7. Application of sign matrix from DESeq data to indicate activating or repressing interactions
      8. Generation of the interaction table defining the final network
      9. Confidence score cutoff
   4. Testing to guarantee integrity of code base
   5. Visualization of generated data in Cytoscape
      1. Loading the interaction table and z-score table
      2. How style configuration enables visualization
      3. AllegroLayout plugin to calculate the network layout
3. **Results**
   1. Characteristics of the produced network
   2. Note network reactions to
      1. z-score filtering (layout when clustering?)
      2. confidence score cutoff in final step
   3. Statistics of transcription factor effects in final network?
      1. Percentage of repressive or activating interactions for each transcription factor
      2. Interactions of core transcription factors as expected?
   4. Error rate of reported transcription factor interactions compared to interactions predicted by literature or experience based biological functions
      1. aucPR values
   5. Effects of replacing DESeq data from GEO with custom data
   6. Quantitative comparison of produced network with example network provided by original authors (using Cytoscape analysis tools)
4. **Discussion**
   1. Computational differences to original method
   2. Discussion of quantitative network comparison using Cytoscape tools
   3. The effects of score filtering on network topology
      1. What does z-score filtering achieve? Leave only genes with significant differential expression?
      2. What does confidence score cutoff achieve?
   4. How close are we to literature or experience based biological functions of transcription factors?
5. **Conclusion** (summary of results, further work and scope)
   1. Summary of added value
   2. Future work

**Bibliography**

**Appendices** (Source code, etc.)