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

**Foreword**

**Table of Contents**

**Abstract**

1. **Introduction**
   1. Understanding biological interactions using networks (subject and interest of work)
   2. 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 foundation for reproducible methods for network analysis of high-throughput genomic data (summary of 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. Initial project setup
      1. Setup script and project structure
      2. Usability and reproducibility
   3. Network generation algorithm
      1. Initial parsing and processing
      2. Filtering of target genes
      3. Distinction of activator and repressor matrices
      4. Data integration by quantile ranking
      5. Combination of ranked matrices for each data type
      6. Extraction of activating or repressing interactions from DESeq data
      7. Generation of the interaction table defining the final network
   4. Unit testing to guarantee integrity of code base
   5. Visualization of generated interaction data in Cytoscape
      1. Loading the interaction table and z-score table
      2. Enhancing visualization by style configurations
      3. AllegroLayout plugin and the Fruchterman-Reingold algorithm to calculate the final network layout
3. **Results**
   1. Generated interactions from test data and characteristics of the produced network
      1. Available high-throughput genomic test data
      2. Basic network statistics (via Cytoscape analysis)
   2. Changes in general network topology with varying parameters
      1. Z-score filtering and its effect on node clustering
      2. The influence of a confidence score cutoff on network interactions
   3. Gene enrichment analysis using literature curated list of Th17 relevant genes (mmc2.xlsx list provided on Cell, precision/ recall of reported transcription factor interactions, aucPR)
   4. Statistics of transcription factor interactions
      1. Interactions with target genes
      2. Interactions between core transcription factors
   5. Quantitative comparison of produced network with an example network provided by original authors (using Cytoscape analysis tools)
   6. Effects of using custom DESeq data
4. **Discussion**
   1. General implementation differences to original method
   2. The effects of score filtering on network topology
      1. What does z-score filtering achieve? Does node clustering significantly change?
      2. What is the ideal confidence score cutoff?
   3. Retrieval of literature curated Th17 relevant genes
      1. Predictive value for transcription factor interactions
      2. Possible improvements to increase prediction value
   4. Quantitative network comparison using Cytoscape tools
   5. Custom DESeq data
5. **Conclusion**
   1. Summary of results & added value
   2. Future work

**Bibliography**

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