CroCo: The Cross-Species Network Conservation Framework

Robert Pesch and Ralf Zimmer

LFE Bioinformatik, Institut für Informatik, Ludwig-Maximilians-Universität München,

Amalienstraße 17, 80333 München, Germany, robert.pesch@bio.ifi.lmu.de

The comparison of networks for different species allows the investigation and visualization of cross-species conservation of network edges and the extension of sparse species networks by orthologous transfer of edges. Previously, we developed ConReg [1] a system for the storage and identification of transcription factor target conservations for eukaryotes. We applied ConReg to the 471 known regulatory relations for fruit fly (as reported in the manually curated resource REDfly [2]) and could confirm the conservation of 66 relations in at least one vertebrate model organism.

With CroCo we extend ConReg to a general-purpose analysis framework for the investigation of conservations in any type of biological interaction network. Currently, our system includes regulatory and protein interaction networks inferred with various methods for eight eukaryotic model organisms. The repository consists of static and condition-specific regulatory networks derived from the literature, predicted via binding sites and networks derived from data of the ENCODE projects (ENCODE, mouseENCODE and modENCODE). In addition, we integrated protein interaction networks from COIN-DB [3]. Besides the central and publicly available network repository, CroCo consists of user interfaces and an application-programming interface (API), which allows to access our data repository and to conduct network operations, or to integrate CroCo in further analysis pipelines.

As an example, we use CroCo to study the cell-specificity and conservation of transcription factor target networks for Human and Mouse. We follow the approach by Neph et al. [4] to infer regulatory networks by overlying binding site predictions with open chromatin data from the ENCODE project. We observe that the majority of interactions are condition specific for both species (Fig 1a). But also some regulations appear in all experiments. In addition, we show the conserved network between Human and Mouse for those relations that could be found in at least 75% of the experiments for the respective species. This network is dominated by few transcription factors such as SP1, a known house keeping factor, and CTCF, which is known to be strongly conserved (see Fig 1b and 1c).

With CroCo we provide a systems biology framework for the investigation of conserved networks and

for addressing a wide range of different questions. The system is designed such that it can easily be extended via additional networks and be integrated in other analysis pipelines. The system will be made available soon via our website (http://services.bio.ifi.lmu.de/croco).

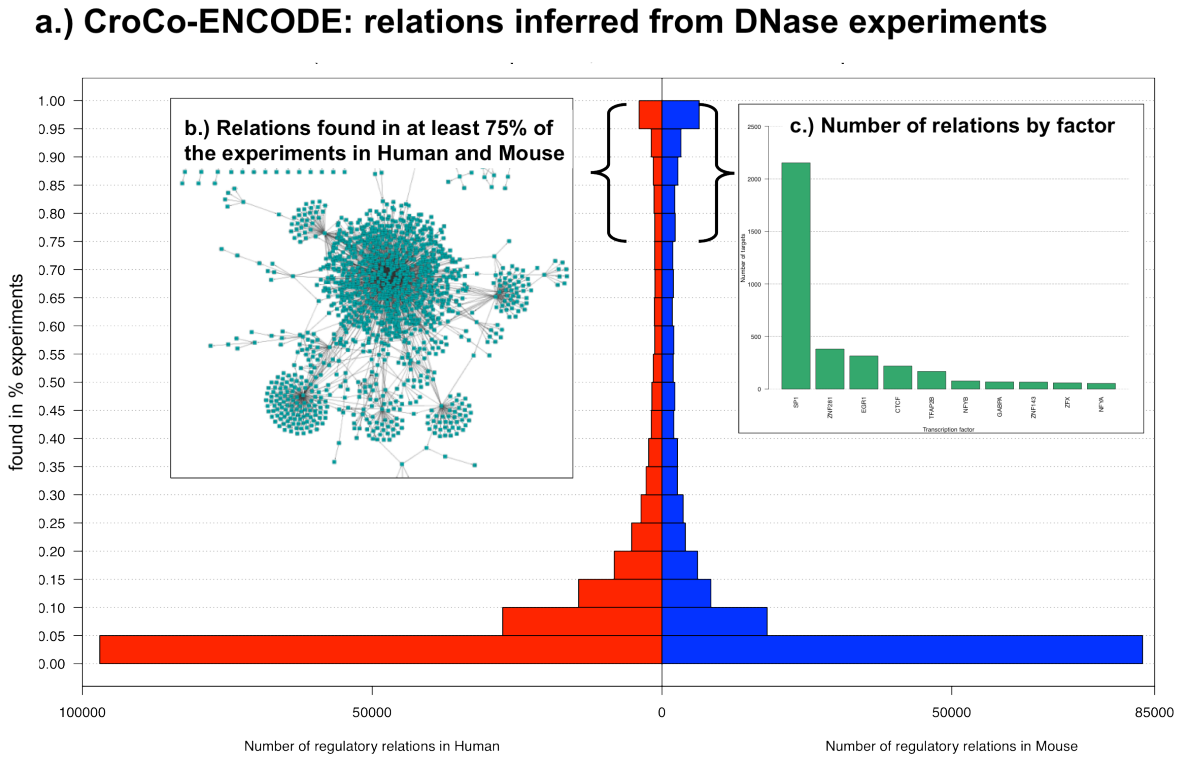

Fig. 1: The histogram shows the cell specificity of the inferred relations from 105 Human cell lines (red) and 42 Mouse cell lines (blue) [5]. b.) The inferred conserved network between the two species for those relations that are observed in at least 75% of the open chromatin experiments. c.) The number of relations for the most highly connected transcription factors in the conserved network.

1. Pesch et al. (2012), ConReg: Analysis and Visualization of Conserved Regulatory Networks in Eukaryotes, *In: Proceedings of the GCB, Open Access Series in Informatics of the Schloss Dagstuhl*, 69-81

2. Gallo et al. (2010), REDfly v3.0: Toward a comprehensive database of transcriptional regulatory elements in Drosophila, *Nucleic Acids Res*, **39**: D118-D123

3. Pesch and Zimmer (2013), Complementing the Eukaryotic Protein Interactome, *PLoSE ONE*, accepted

4. Neph et al. (2012), Circuitry and Dynamics of Human Transcription Factor Regulatory Networks, *Cell*, **150**:1274-1286

5. The ENCODE Project Consortium (2012), An integrated encyclopedia of DNA elements in the human genome, *Nature*, **489**:57–74