Summary

Application of next generation sequencing to the analysis of evolutionary changes in gene expression in primates

Michael Dannemann

Understanding the evolutionary basis for human-specific phenotypes such as complex speech and language, advanced cognition or the unique preparation of their food is a topic of broad interest. Approaches focusing on comparisons of the genomic DNA (deoxyribonucleic acid) or RNA (ribonucleic acid) sequence between species, individuals or tissues allow for the identification of evolutionary sequence changes, some of these changes may underlie differences in phenotypes. In addition, differences in when, where and how much of a particular gene is present may also contribute to functional changes and therefore also to phenotypic differences.

The resources to make such comparisons using genetic data are now available. The genome sequences of a number of outgroups: all living great apes, as well two archaic humans, are now publically available. Studying gene expression on the RNA level - a precursor of the protein expression - is considerably easier and cheaper than the measurement of expression of the protein itself. It has been shown that the RNA and protein expression levels are well correlated and therefore measuring RNA levels provides a good proxy for the expression of the protein. Using high-throughput sequencing techniques, relatively unbiased expression comparison is now possible because the RNA from any species can be sequenced directly, rather than being captured on arrays which are designed based on a particular reference sequence.

The aim of this research was to use gene expression as a molecular phenotype to identify changes relevant to human-specific biology and study the difference between humans and their closest living relatives to understand patterns and differences in the gene expression and in gene expression regulation in multiple tissues in primates using high-throughput sequencing techniques. In my thesis, I describe two analyses to address open questions in the field of gene expression and genes expression regulation in humans.

In the first part I will analyze how the effect of different diets impact gene expression using a mouse model. Two key components of the human diet that differ substantially from the diet of other primates, the frequent use of meat of many humans and the cooking of their food which is common for almost all

human populations, are modeled in the experiment. I tested for their impact on liver gene expression. I found that both the differences in food substrates - meat and tuber - as well as in their preparation affect gene expression in mice significantly. The effect is bigger between food substrates than between methods of preparation. Differentially expressed genes between food substrates and food preparation were predominantly related to metabolic functions. In addition, immune-genes showed differential expression between the comparisons of raw meat to both, raw tuber and cooked meat, respectively. The results indicate that different food substrates and food preparations activate different metabolic pathways and that the cooking of food and particularly of meat has an influence on the immune also changes immune-reactions of the body. I showed that expression differences in these mice are correlated with the differences observed between humans and other primates, and that there is evidence that adaptation to these diets dates to more than 300.000 years. Finally, I showed that transcription factors play in important role in regulation of gene expression with respect to different food preparation.

In the second part I analyzed the expression of one key regulator of gene expression: microRNAs (miRNAs). Using miRNA expression data from multiple primate species and for multiple tissues I found that expression differences vary between tissues. While heart and brain show only few expression differences between primates, other tissues are more variable in expression. The most extreme expression differences in all three primate species were found in the brain, which may reflect the importance of miRNAs in the regulation of gene expression in the brain. Expression differences in testis were significantly larger between humans and macaques than between chimpanzees and macaques, indicating that miRNAs evolved differently in human compared to chimpanzees. MiRNA expression differences were correlated with expression differences of their target genes genome-wide which underlines the regulatory importance of miRNAs. I also showed that differentially expressed miRNAs between species/tissues preferentially targeted transcription factors, which are important gene expression regulators as well. This finding that suggests complex regulatory pathways involving both miRNAs and transcription factors in the control of gene expression. Finally, I used the miRNA sequencing data to annotate new miRNAs in primates and was able to increase the number of annotated miRNAs substantially, especially for the non-human primates which were previously not extensively annotated. The overlap of miRNAs annotated in multiple primate species thereby also increased which will support future studies to investigate the evolutionary changes of miRNAs between these primates.