Summary

Gene regulatory factors in the evolutionary history of humans

B.Sc. M.Sc. Alvaro Perdomo-Sabogal

Changes in *cis*- and *trans*-regulatory elements are among the prime sources of genetic and phenotypical variation at species level. The introduction of *cis*- and *trans* regulatory variation, as evolutionary processes, has played important roles in driving evolution, diversity and phenotypical differentiation in humans. Therefore, exploring and identifying variation that occurs on *cis*- and *trans* regulatory elements becomes imperative to better understanding of human evolution and its genetic diversity.

In this research, around 3360 gene regulatory factors in the human genome were catalogued. This catalog includes genes that code for proteins that perform gene regulatory activities such DNA-depending transcription, RNA polymerase II transcription cofactor and co-repressor activity, chromatin binding, and remodeling, among other 218 gene ontology terms. Using the classification of DNA-binding GRFs (Wingender et al. 2015), we were able to group 1521 GRF genes (~46%) into 41 different GRF classes. This GRF catalog allowed us to initially explore and discuss how some GRF genes have evolved in humans, archaic humans (Neandertal and Denisovan) and non-human primates species. It is also discussed which are the likely phenotypical and medical effects that evolutionary changes in GRF genes may have introduced into the human genome are; for instance, speech and language capabilities, recombination hotspots, and metabolic pathways and diseases.

In addition, by exploring genome-wide scan data for detecting selection, we built a list of GRF candidate genes that may have undergone positive selection in three human populations: Utah Residents with Northern and Western Ancestry (CEU),

Han Chinese in Beijing (CHB), and Yoruba in Ibadan (YRI). We think this set gathers genes that may have contributed in shaping the phenotypical diversity currently observed in these three human populations, for example by introducing regulatory diversity at population-specific level. Out of the 41 DNA-binding GRF classes, six groups evidenced enrichment for genes located on regions that may have been target of positive selection: C2H2 zinc finger, KRAB-ZNF zinc finger, Homeo domain, Tryptophan cluster, Fork head/winged helix and, and High-mobility HMG domain. We additionally identified three KRAB-ZNF gene clusters, in the chromosomes one, three, and 16, of the Asian population that exhibit regions with extended haplotype homozygosity EHH (larger than 100 kb). The presence of this EHH suggests that these three regions have undergone positive selection in CHB population. Out of the 22 GRF genes located within these three KRAB-ZNF clusters, seven C2H2-ZNF GRF genes (*ZNF695, ZNF646, ZNF668, ZNF167, ZNF35, ZNF502,* and *ZNF501*) carry nonsynonymous SNPs that code nonsynonymous SNPs that change the amino acid sequence in their protein domains (linkers and cystine-2 histidine-2 amino-acid sequence motifs). Six GRF genes located on the EHH region on the chromosome 16 of CHB have been associated with obesity (*KAT8*, *ZNF646*, *ZNF668*, *FBXL19)* and blood coagulation (*STX1B* and *VKORC1*) in humans. In addition, we also detected genetic changes at GRF sequence level that may have resulted in subtle regulatory changes in metabolic pathways associated with glucose and insulin metabolism at population-specific level.

Finally, acknowledging that a representative fraction of the phenotypic diversity we observed between humans and its closely related species are likely explained by changes in *cis*-regulatory elements (CREs), putative binding sites of the transcription factor GABPa were identified and investigated. GABPa is GRF protein member of the E-twenty six DNA-binding proteins class. GABPs control gene expression of many genes that play key roles at cellular level, for instance, in cell migration and differentiation, cell cycle control and fate, hormonal regulation and apoptosis. Using ChIP-Seq data generated from a human cell line (HEK293T), we found 11,619 putative GABPa CREs were found, of which 224 are putative human-specific. To experimentally validate the transcriptional activity of these human-specific GABPa CREs, reporter gene essays and knock-down experiments were performed. Our

results supported the functionality of these human-specific GABPa CREs and suggest that at least 1,215 genes are primary targets of GABPa. Finally, further analyses of the data gathered depict scenarios that bring together transcriptional regulation by GABPa with the evolution of particular human speciation and traits for instance, cognitive abilities, breast morphology, and lipids and glucose metabolic pathways and the regulation of human-specific genes.

By studying genetic changes in *cis-* and *trans-* regulatory elements in humans in two different evolutionary time frames, species evolution and population genetics, we were able to show how genome regulatory innovations and genetic variation may have contributed to the evolution of human- and population specific traits. Here, we conclude that human-specific changes in regulatory elements are likely introducing subtle regulatory variation in key pathways at physiological level, for instance, in glucose/insulin, lipids metabolism and cognitive abilities. Some of these changes may have resulted in adaptive responses that left signatures of positive selection at human population specific level.