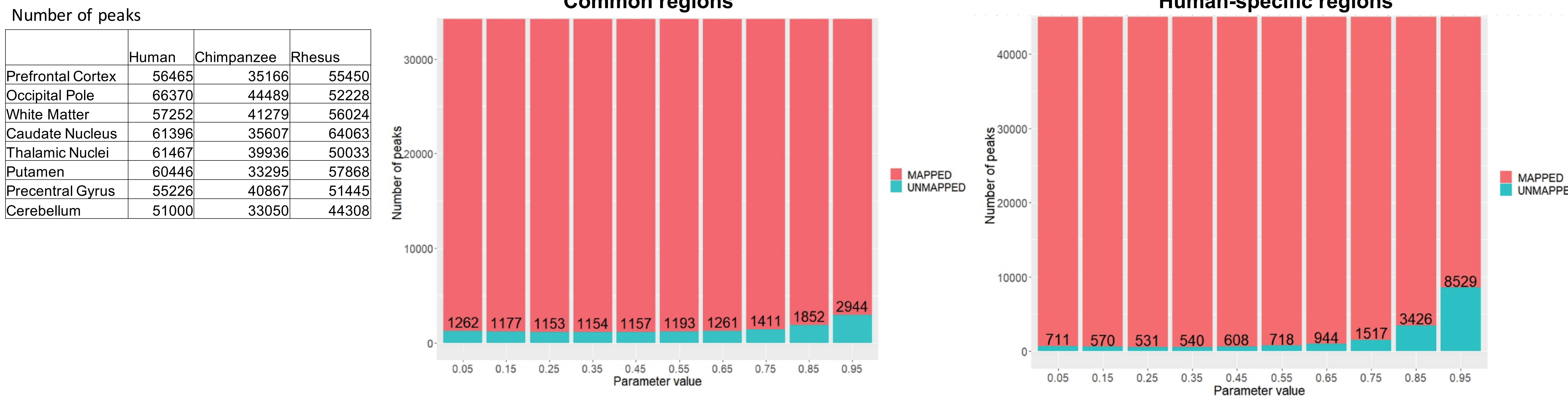# Evolutionary analysis of sequence divergence of cis-regulatory modules associated with changes in chromatin and gene expression in brain tissue of primates.

Abusaid Shaimardanov, Dmitry Svetlichnyy
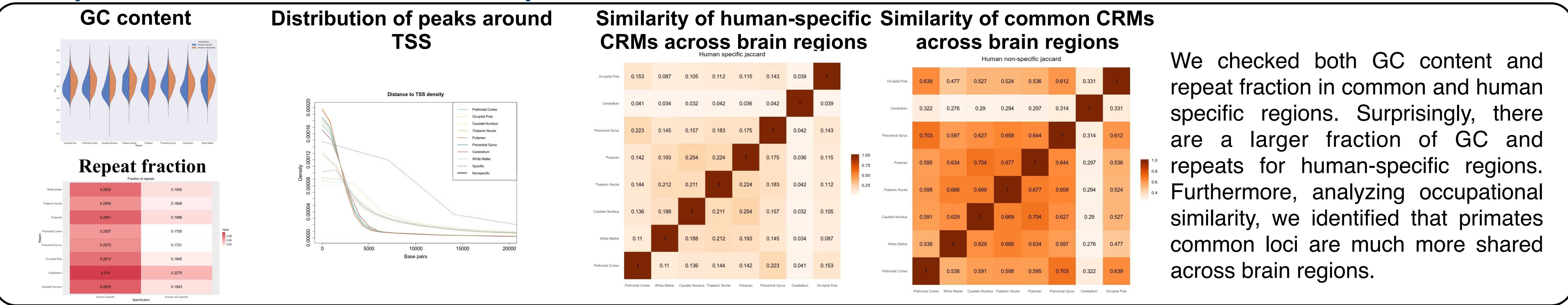National Research University Higher School of Economics

Introduction

Investigating gene expression evolution in brain tissue together with the divergence of genomic regulatory elements in a phylogenetic context is important to decipher genes that might shape phenotypic differences. In this study, we have investigated evolution of gene expression level in the context of sequence changes in the cis-regulatory elements.

Overview of the data

Number of peaks

	Human	Chimpanzee	Rhesus
Prefrontal Cortex	56465	35166	55450
Occipital Pole	66370	44489	52228
White Matter	57252	41279	56024
Caudate Nucleus	61396	35607	64063
Thalamic Nuclei	61467	39936	50033
Putamen	60446	33295	57868
Precentral Gyrus	55226	40867	51445
Cerebellum	51000	33050	44308

Cerebellum Common regions

Cerebellum Human-specific regions

Properties of the human specific CRMs in brain

GC content

Repeat fraction

Distribution of peaks around TSS

Similarity of human-specific CRMs across brain regions

Similarity of common CRMs across brain regions

We checked both GC content and repeat fraction in common and human specific regions. Surprisingly, there are a larger fraction of GC and repeats for human-specific regions. Furthermore, analyzing occupational similarity, we identified that primates common loci are much more shared across brain regions.

Identification of the regulatory proteins

Motifs

HOMER correlation matrix

ReMap correlation matrix

Non Specific

We performed a detailed investigation of the DNA regulatory motifs to identify transcription factors that work in different regions of the H3K27Ac coming from eight brain areas.

Feature importance

Enrichment results

H3K27Ac peaks -> foreground

Orthologous regions -> background

Similarity of found motifs across areas of brain

HOMER correlation matrix
Rhesus background

DCN for CRM classification

Model	Accuracy best	Accuracy worst/average
M1	0.88	0.56
M2	0.83	0.67
M3	0.83	0.71

M1 - binary classification
M2 - multi-label classification
M3 - multi-label classification with orthologus regions

We also investigated human-specific regulatory regions for evolution of binding sites for transcription factors. A low number of TFs show enrichment of the binding motifs compared to rhesus orthologus regions in primates

We applied machine learning approach based on convolutional neural networks in order to develop a classifier that allows getting insight in DNA features that were learned from the primary sequence and distinguishes regions of the H3K27Ac between different brain areas and species.