



**HAL**  
open science

## Genome-wide Characterization of Exons as Regulatory Elements

Jean-Christophe Mouren, Antoinette van Ouwerkerk, Magali Torres, Frédéric Gallardo, Salvatore Spicuglia, Benoit Ballester

► **To cite this version:**

Jean-Christophe Mouren, Antoinette van Ouwerkerk, Magali Torres, Frédéric Gallardo, Salvatore Spicuglia, et al.. Genome-wide Characterization of Exons as Regulatory Elements. JOBIM 2024 - Journées ouvertes en biologie, informatique, et mathématiques 2024, Jun 2024, Toulouse, France. hal-04633459

**HAL Id: hal-04633459**

**<https://hal.science/hal-04633459v1>**

Submitted on 3 Jul 2024

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



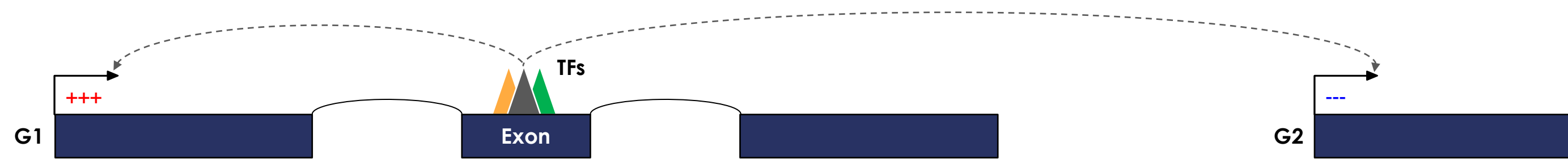
Jean-Christophe Mouren<sup>1</sup>, Antoinette Van Ouwerkerk<sup>1</sup>, Magali Torres<sup>1</sup>, Frederic Gallardo<sup>1</sup>,

Salvatore Spicuglia<sup>1</sup> and Benoit Ballester<sup>1</sup>

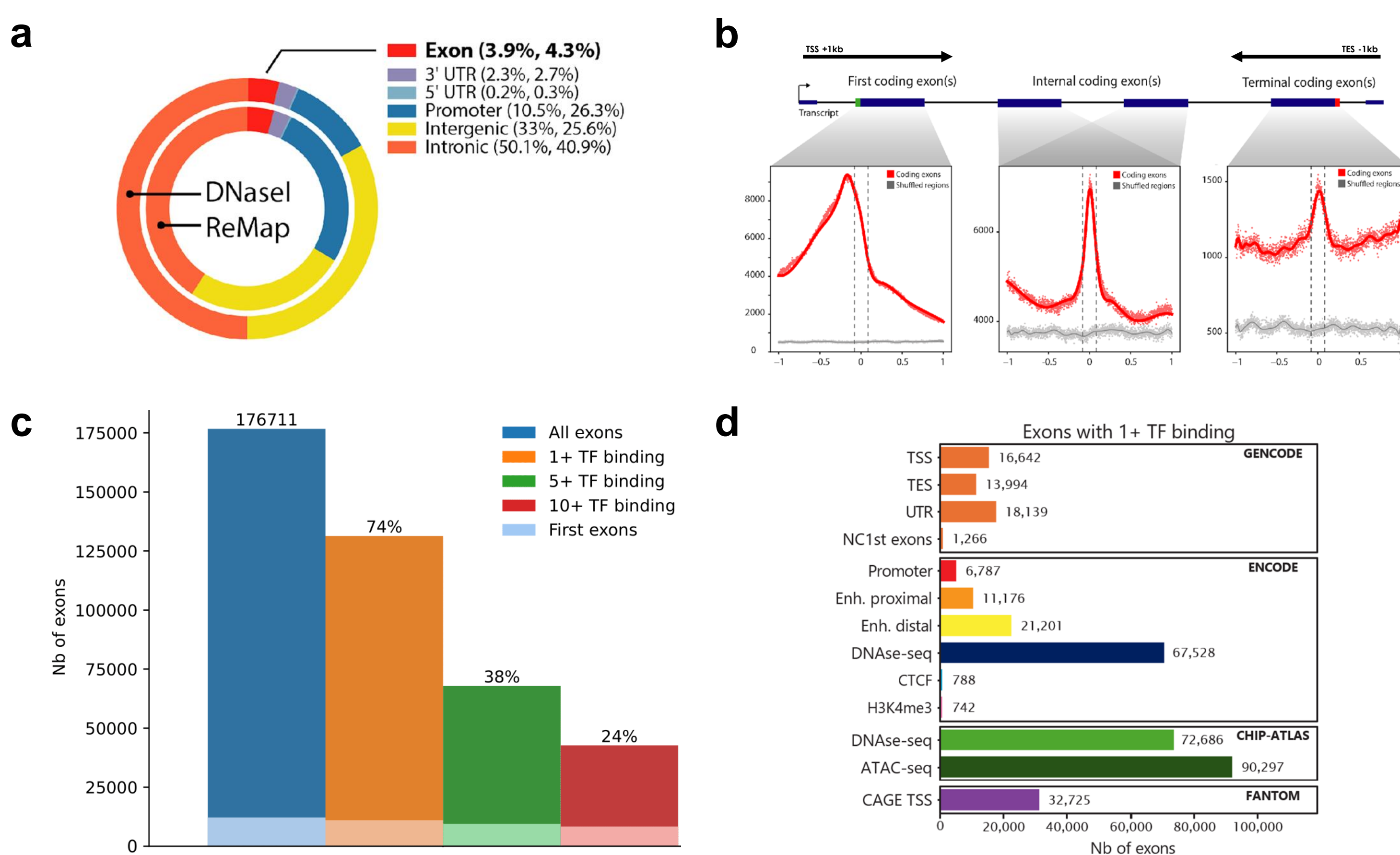
<sup>1</sup> TAGC Inserm U1090, 163 avenue de Luminy, 13288, Marseille 09, France

## Introduction

Protein-coding DNA makes up only 2% of the human genome, with the rest containing key regulatory information. Current methods often miss the regulatory roles of coding DNA. Using computational and experimental methods, we identified 13,481 coding exons with strong regulatory signatures in human. We integrated DNase catalogs and STARR-seq experiments to explore these exons regulatory roles and used a promoter capture Hi-C compendium and the ENCODE-rE2G model to identify their gene targets. Using variants catalogs and luciferase assays, we explored coding mutations impact in regulatory exons.



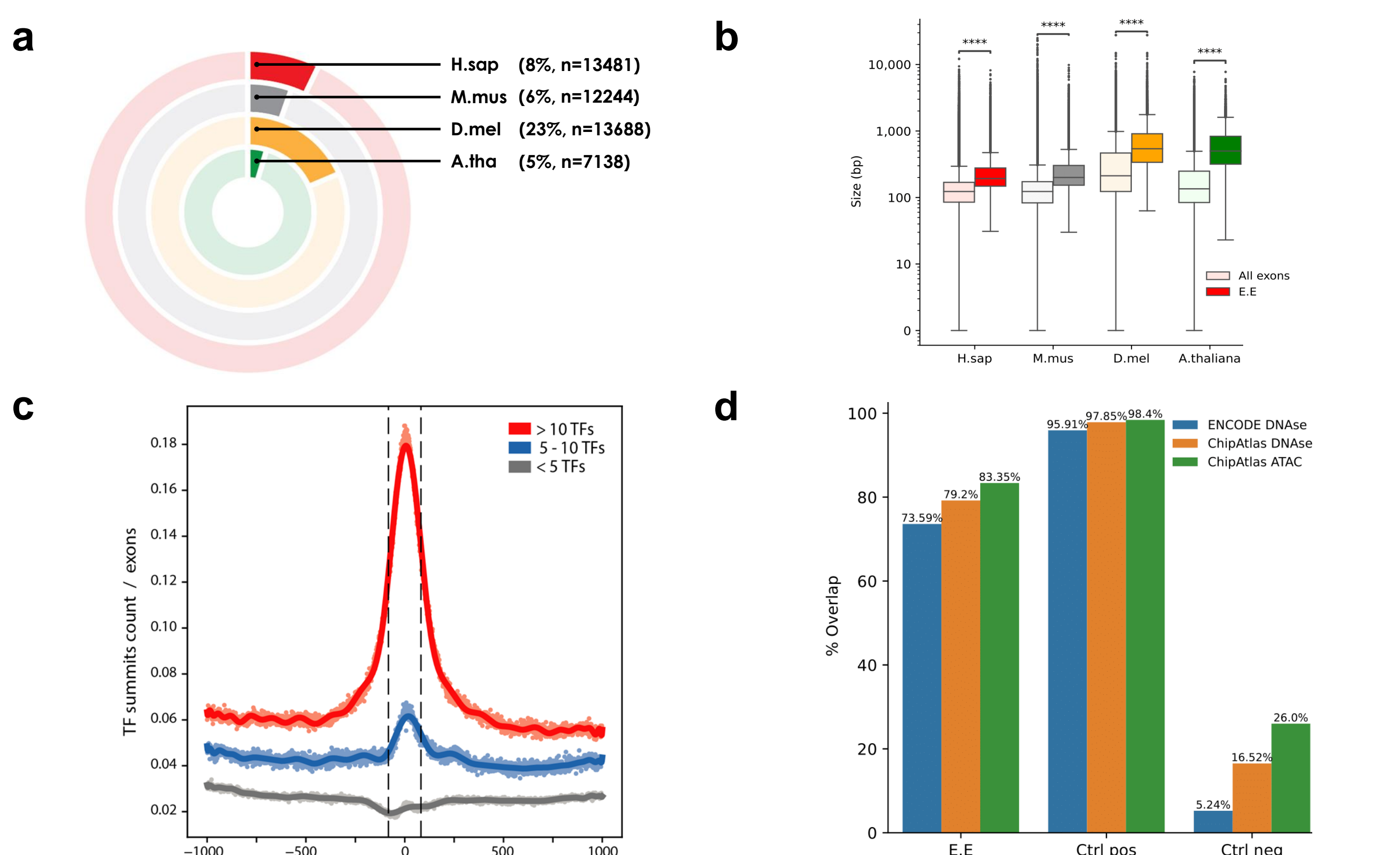
## Regulatory features overlap with exons



**Figure 1 :** a. Repartition of ReMap ChIP-seq and ENCODE DNase-seq in the genome. b. Distribution of ReMap non-redundant ChIP-seq summits on coding exons and 1kb around (0 = middle of exons). c. Overlap of coding exons with ReMap non-redundant ChIP-seq summits. d. Overlap of coding exons with various genomic features.

## Genomic characteristic of exonic enhancers

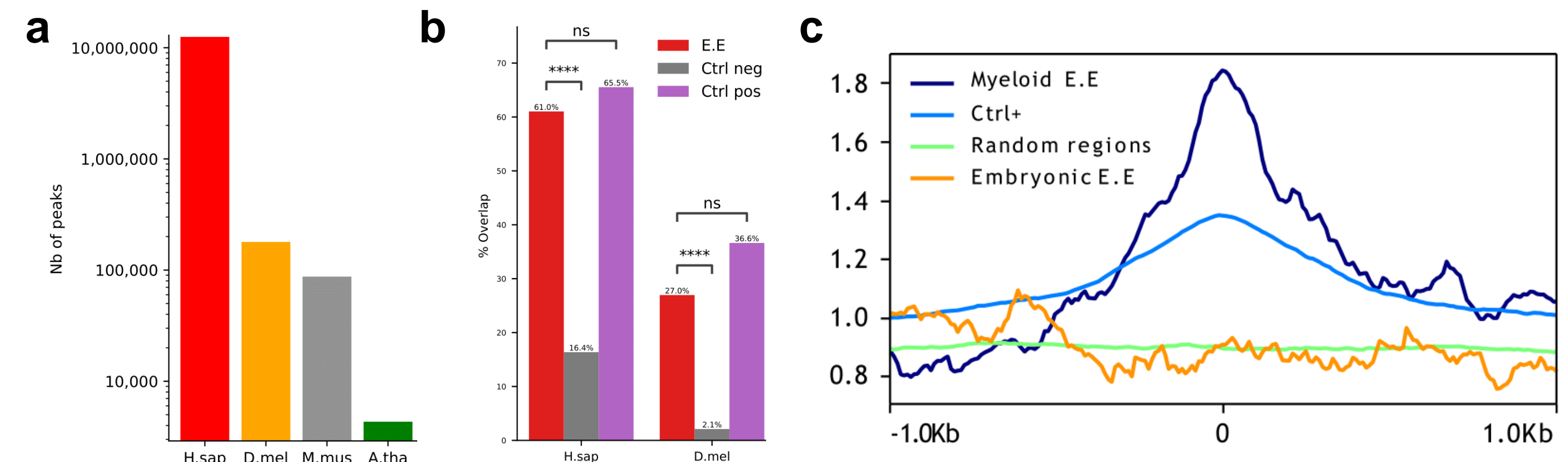
We selected regulatory exons candidates in 4 species based on our previous observations : +10 TF peaks, internal exons only and no promoter signatures.



**Figure 2 :** a. Percentages and numbers of regulatory exons in 4 specie's genomes. b. Mean size of regulatory exons compared to other exons. c. Distribution of ReMap non-redundant ChIP-seq summits in H.sap on normal (gray, blue) and regulatory exons (red), with 1kb around (0 = middle of exons). d. Overlap of H.sap regulatory exons with DNase-seq and ATAC-seq catalogs, compared to positive control (ENCODE distal cCREs) and negative controls (coding exons with no ReMap ChIP-seq peaks).

## STARR-seq analysis

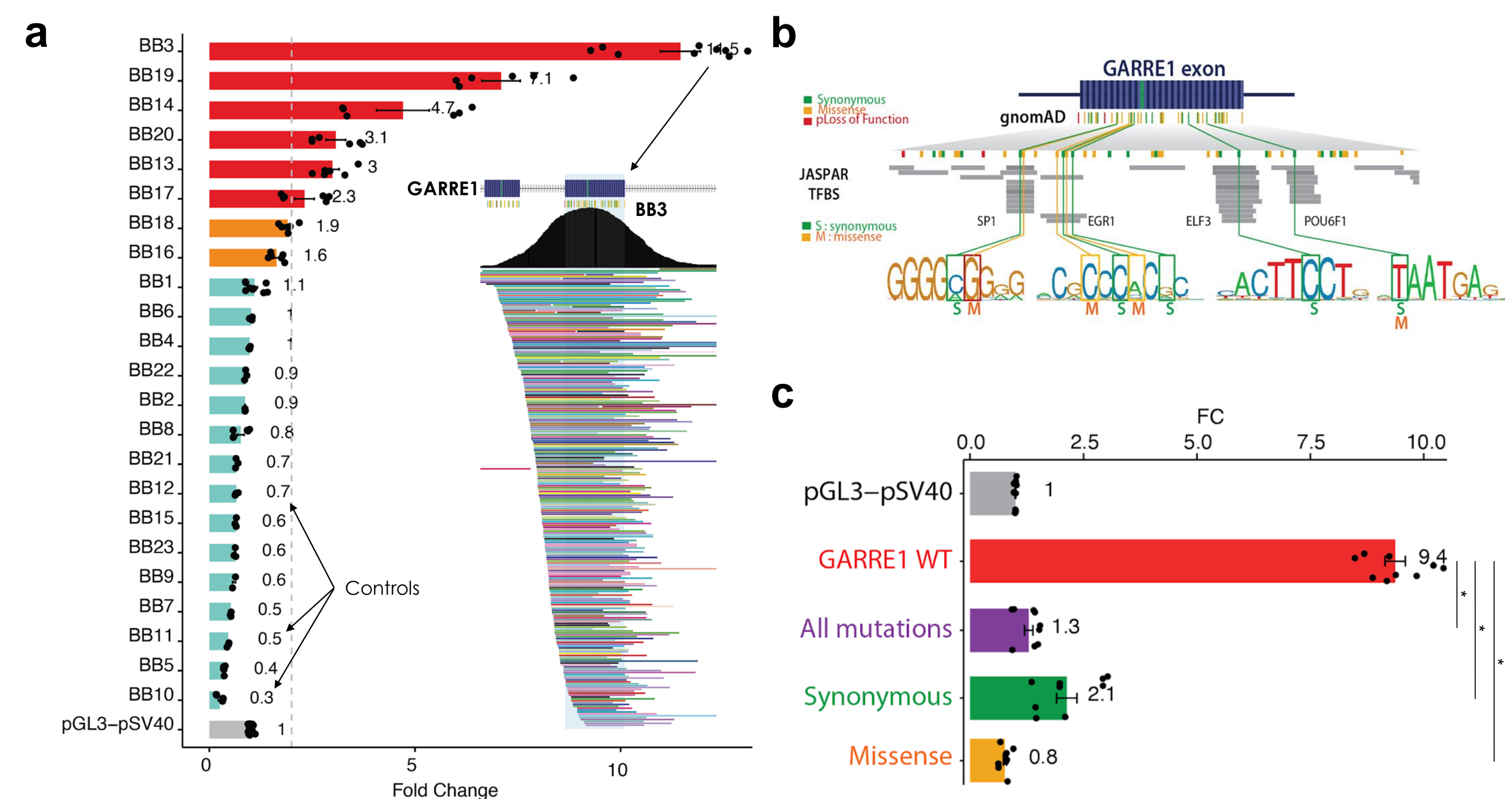
We created a STARR-seq catalog for 4 species.



**Figure 3 :** a. Size of STARR-seq catalogs. b. Overlap of exons candidates and control datasets with the STARR-seq catalogs. c. Signal of an ENCODE STARR-seq experiment in K-562 on H.sap regulatory exons with a specific ChIP-seq myeloid and embryonic signature.

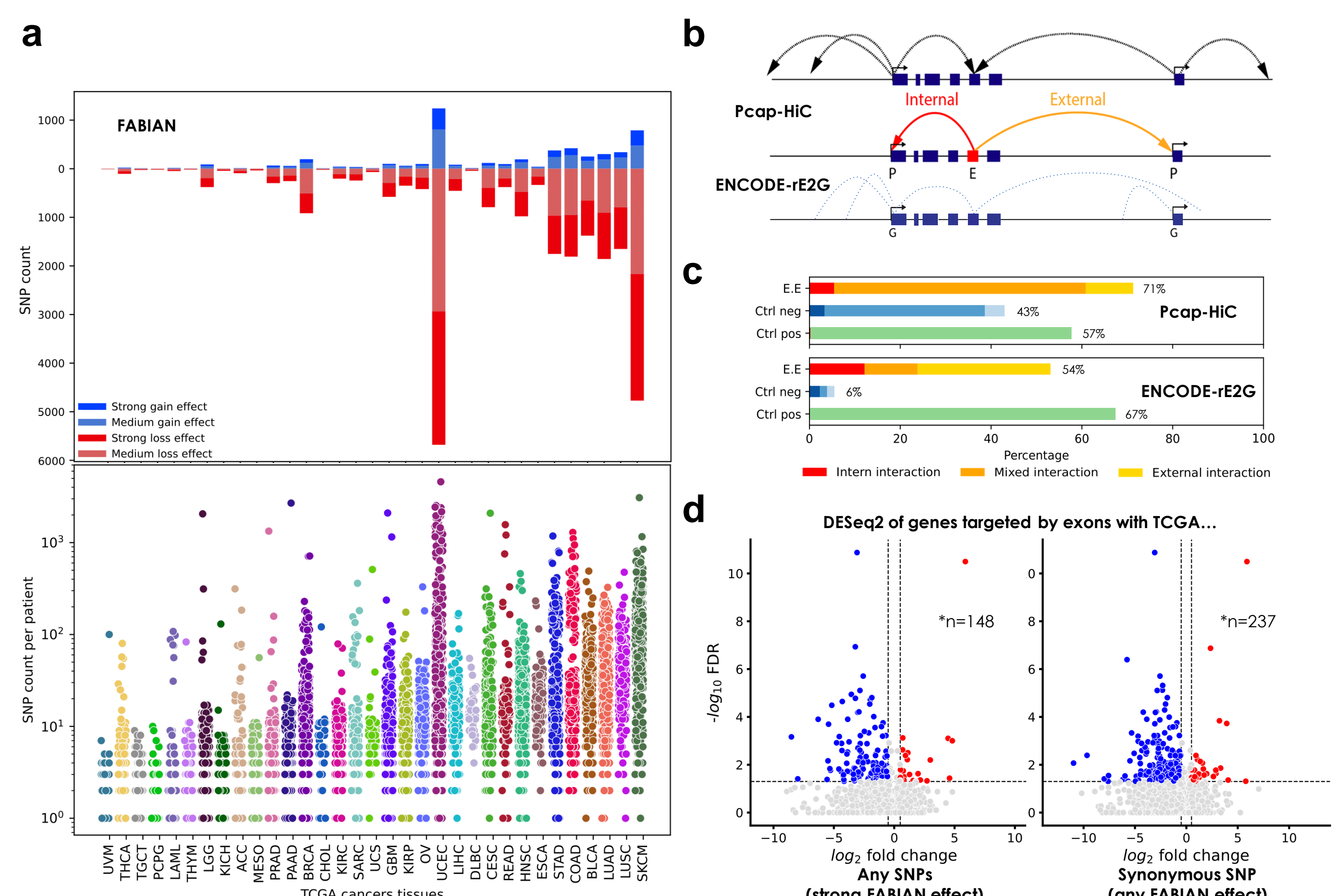
## Regulatory activity of exonic enhancers

Some coding exons show enhancer activities as demonstrated by luciferase assays.



**Figure 4 :** a. Log<sub>2</sub>FC of regulatory exons tested by luciferase assay in the K-562 cell line. b. gnomAD SNPs in BB3 and the REF allele affected within the corresponding cluster of JASPAR TFBS. c. Luciferase assay activity of BB3 sequence with all the highlighted gnomAD SNPs, or only the synonymous/misense mutations.

## Coding variant effects and interactions



**Figure 5 :** a. TCGA variants in regulatory exons and their predicted effect on TF binding by the FABIAN-variant tool, in presence of a ReMap peak and a JASPAR TFBS. b. Schema of the Pcap-HiC method applied for regulatory exons, and the ENCODE-rE2G model. c. Regulatory exons interacting with target genes based on the Pcap-HiC compendium and the ENCODE-rE2G model, compared to controls datasets. d. Differential gene expression analysis of genes targeted by regulatory exons with TCGA SNPs affecting TF binding (in cancer tissues).