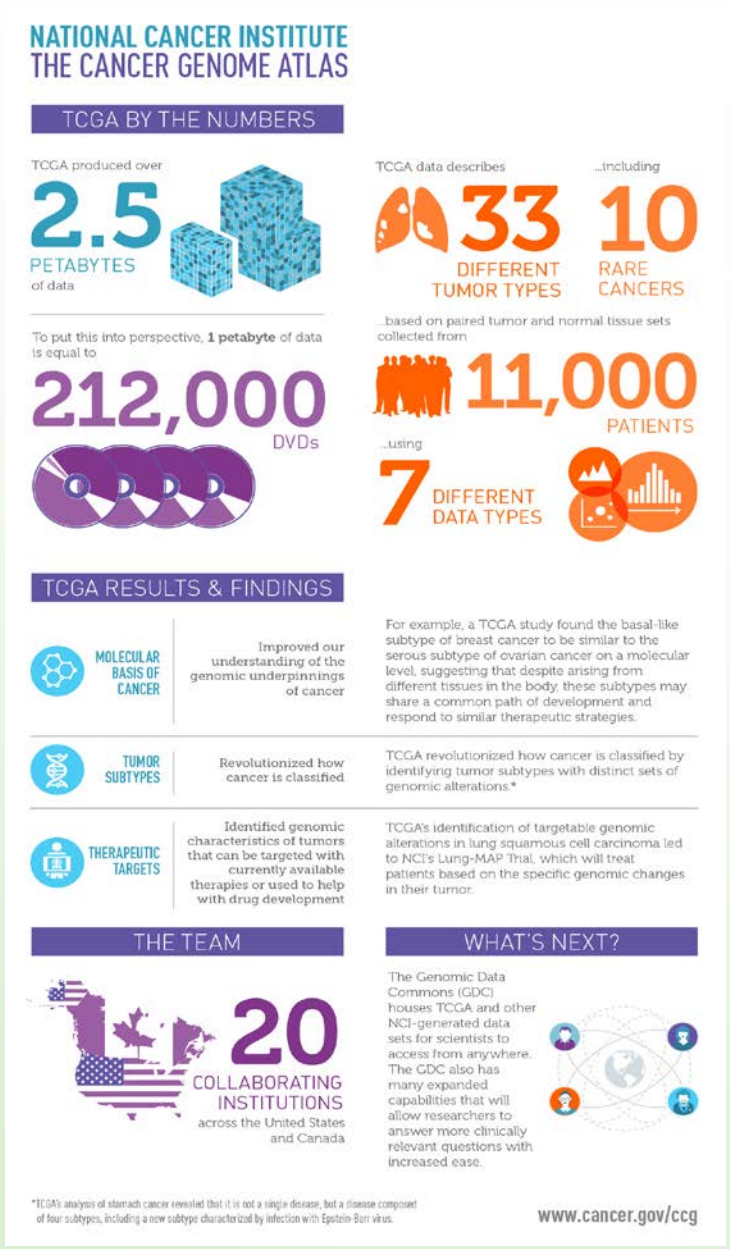
ML LAB

3. YOUR TASK

FRANCESCA M. BUFFA

GENOMICS BIG DATASETS – THE CANCER GENOME ATLAS



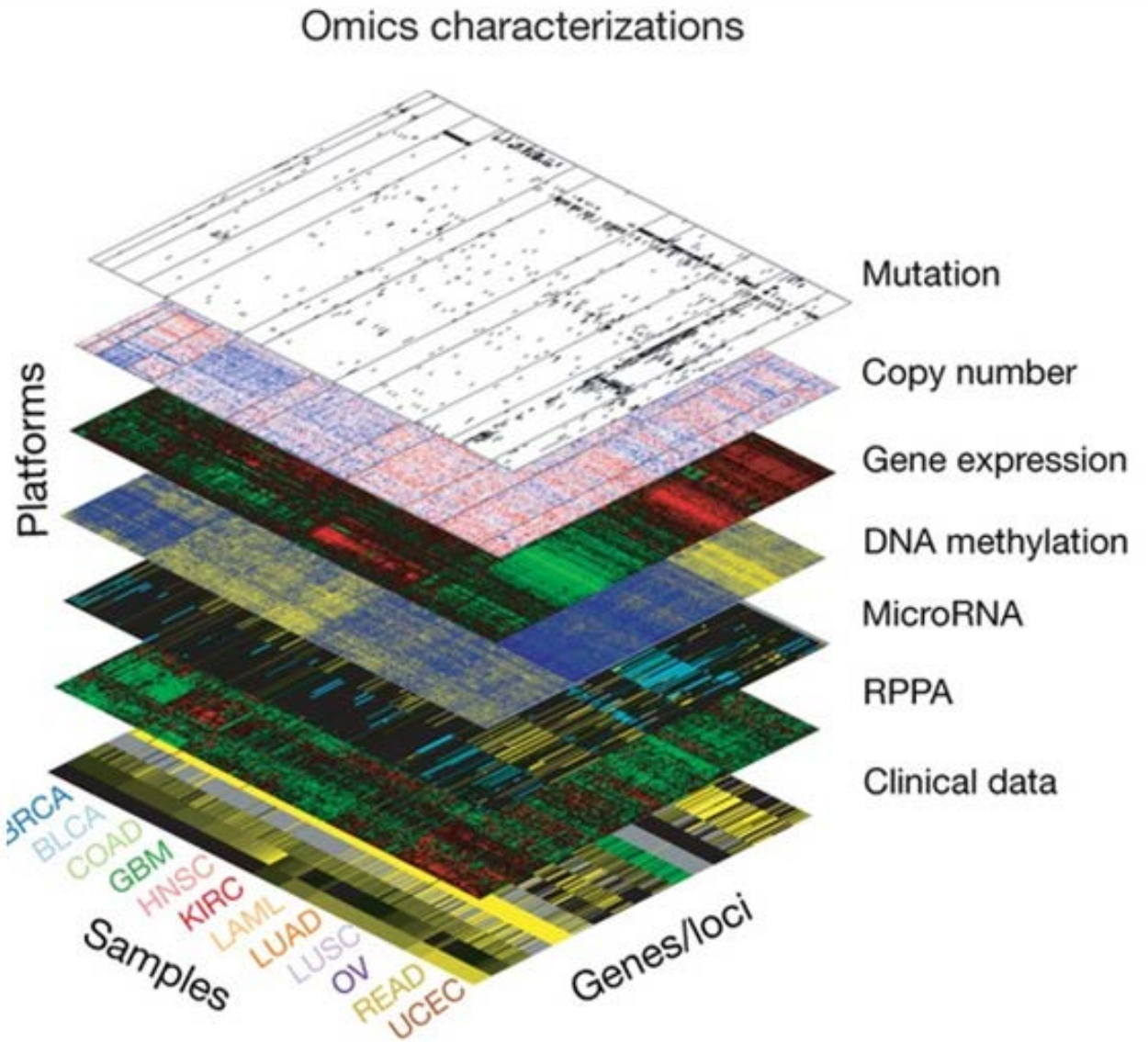
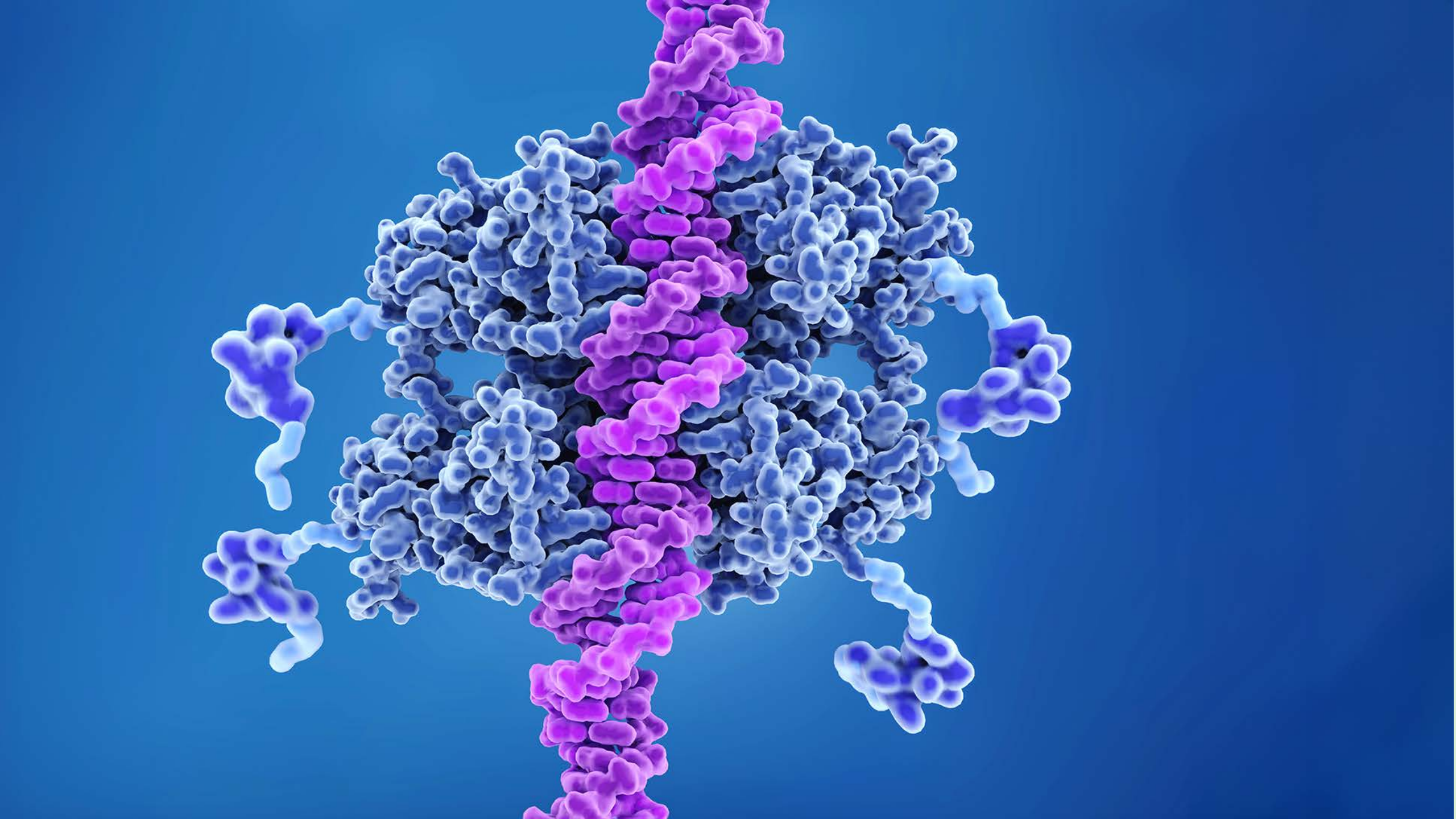


Image source: TCGA Research Network et al. Nature Genetics 2013

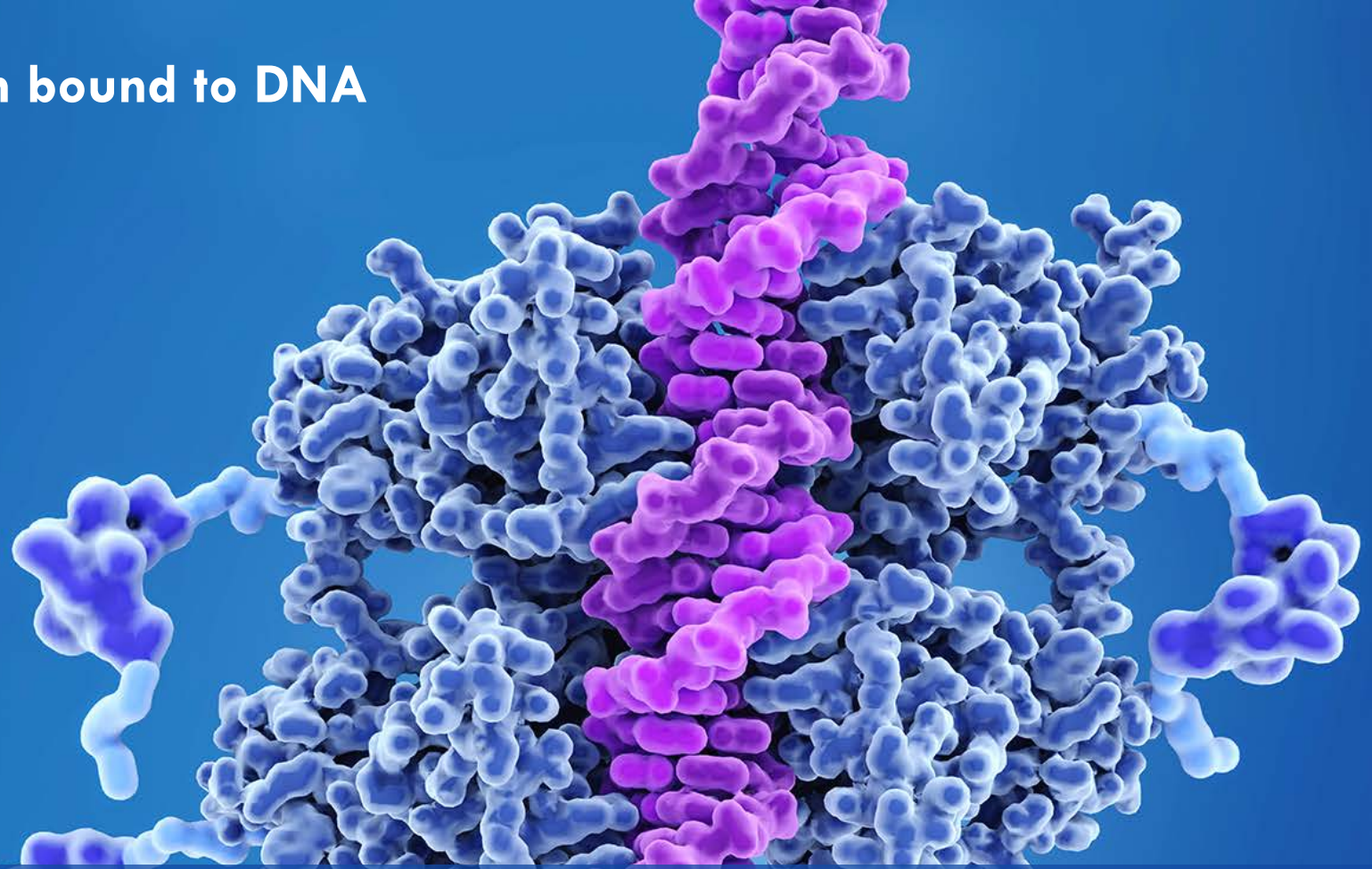
DATABASES - CELL LINES

https://sites.broadinstitute.org/ccle/

https://www.nature.com/articles/s41586-019-1186-3



p53 protein bound to DNA



p53 binds to DNA and activates genes responsible for many functions, it is a gene vital to many forms of life, including humans. It has been called "the guardian of the genome". It codes for a protein that acts a like a guard whose job is to stop a cell dividing when it detects DNA damage. Elefants have 20 copies of this genes! The p53 gene is the most frequently mutated gene (>50%) in human cancer. By knowing more about this protein and other proteins that are mutated in cancer, we can look for new ways of treating these diseases.

TRANSCRIPTION FACTORS

- Transcription factors(TFs) are the interpreter of the genome and perform the first step in gene expression.

- They can recognize specific DNA sequences, they bind to them forming a complex system that controls transcription.

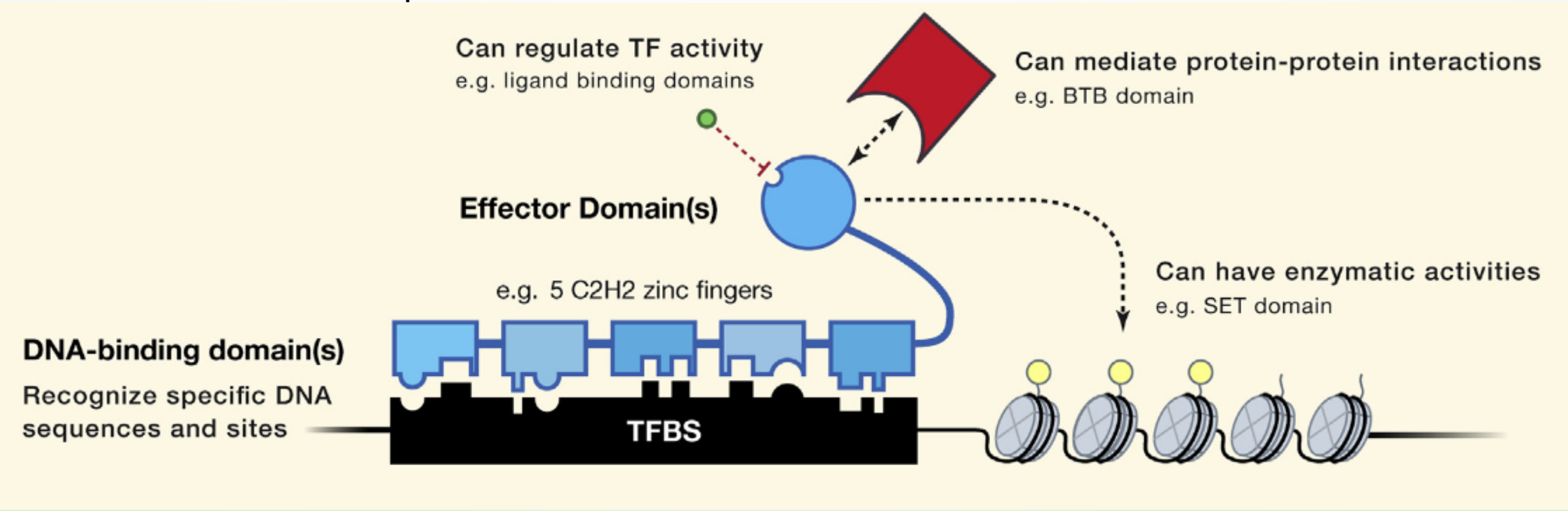


Figure adapted from Lambert et al, Cell, 2018

TRANSCRIPTION FACTORS

- TFs **DNA-binding domains** have preference for specific binding sequences ("motifs")
- TFs bind to other regulatory proteins via **effector domains** to interact with the transcriptional machinery, interact with other TFs, and recruit histone and chromatin modifying enzymes.

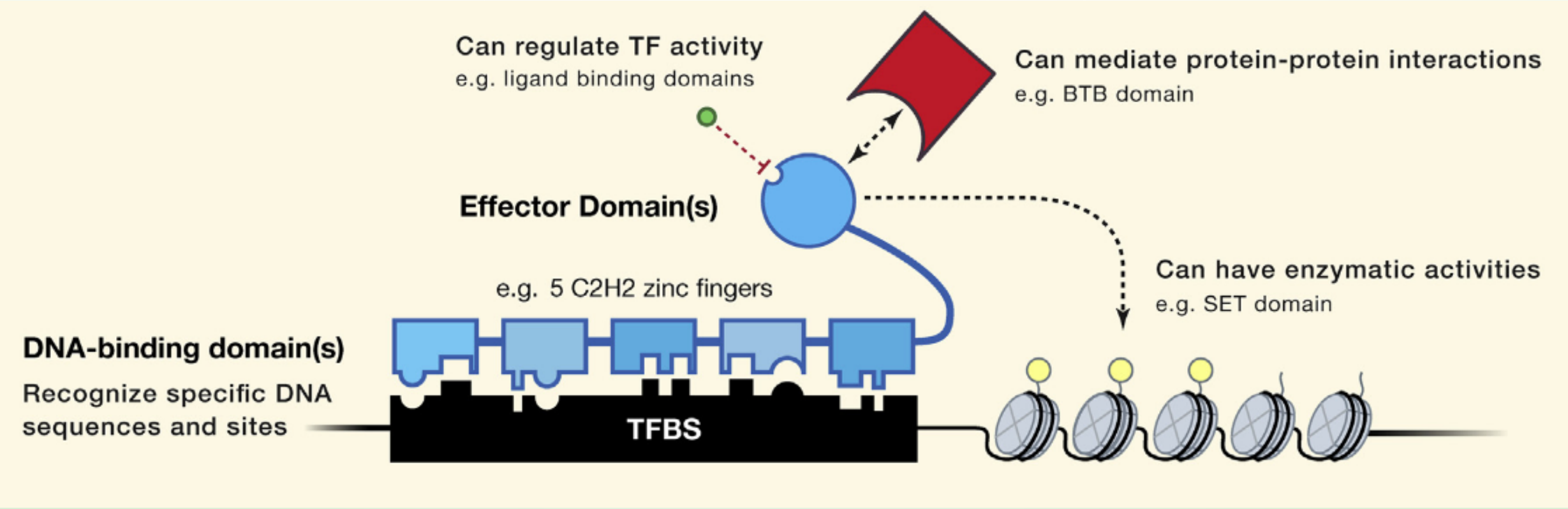
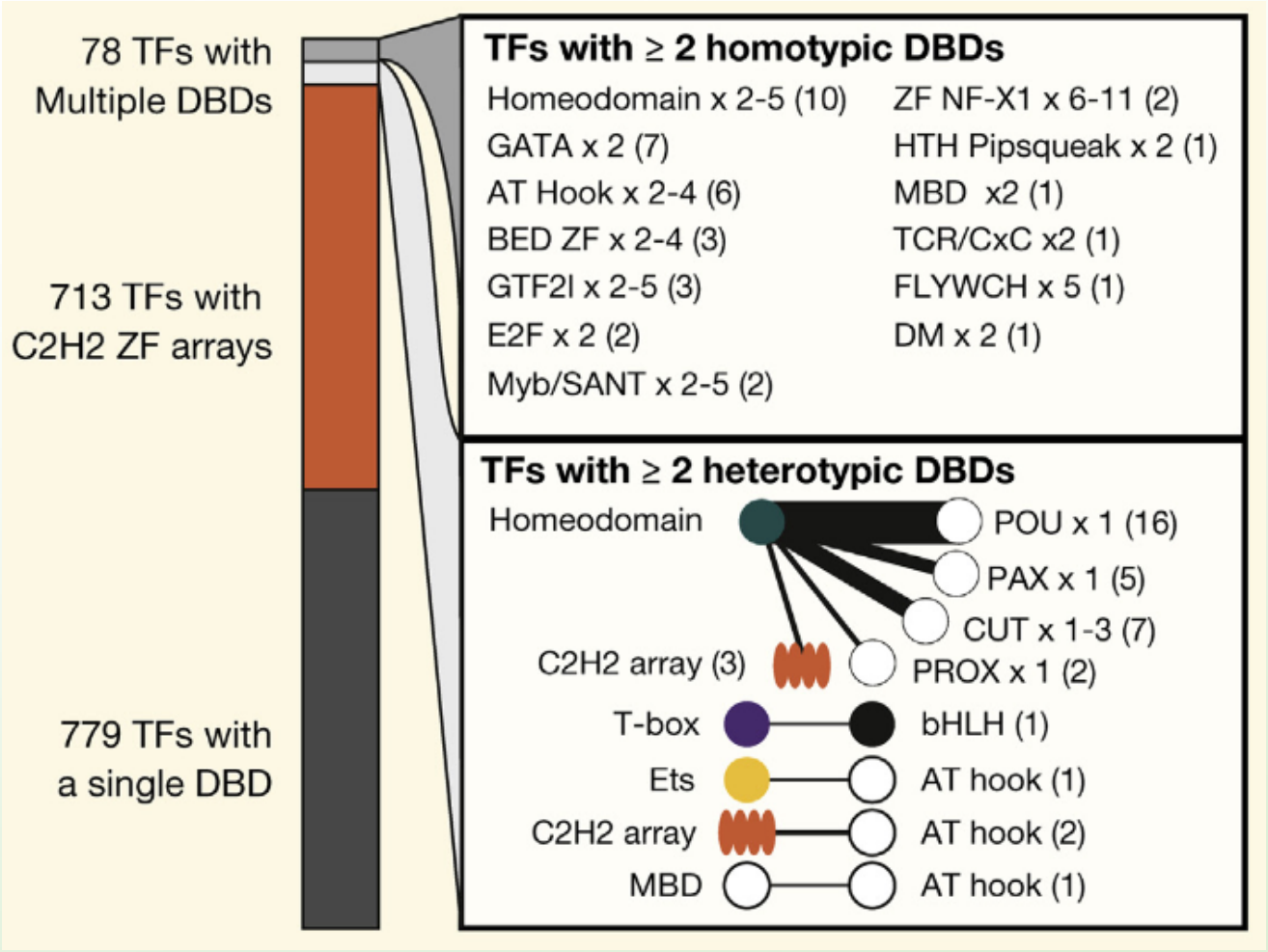


Figure adapted from Lambert et al, Cell, 2018

Human transcription factors

- Human TFs differ in their evolution, expression and function
- Understanding how TFs control gene expression is not complete
- Challenges in determining how DNA binding sites are specified and affect transcription
- 1,600 likely human TFs and DNA-binding domains (DBD) catalog by recent review (http://humantfs.ccbr.utoronto.ca/)



https://pubmed.ncbi.nlm.nih.gov/29425488/

Figure adapted from Lambert et al, Cell, 2018

TISSUE SPECIFICITY OF TF EXPRESSION

Some human TFs are expressed across tissues

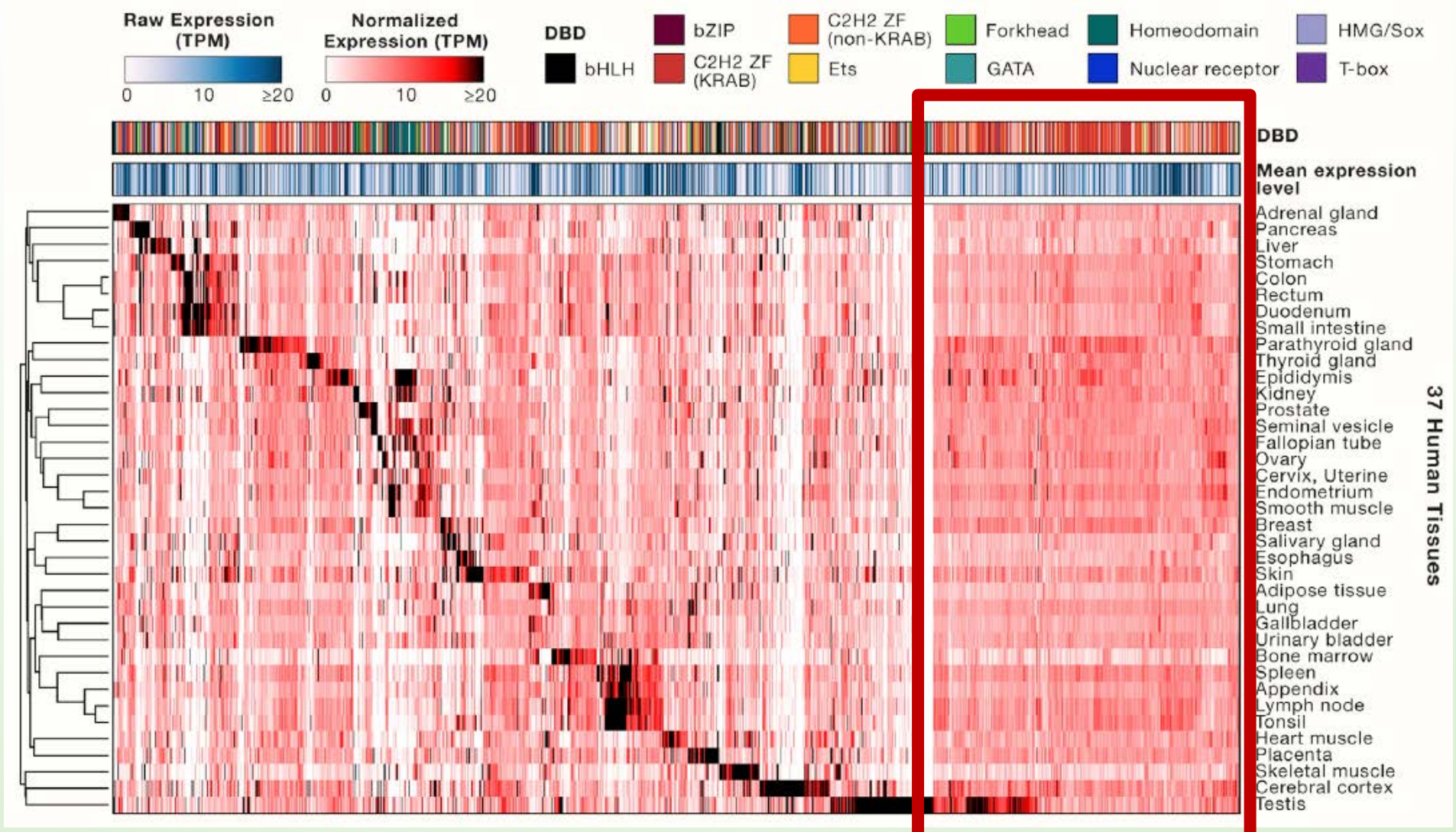


Figure adapted from Lambert et al, Cell, 2018

TISSUE SPECIFICITY OF TF EXPRESSION

Other TFs display tissue-specific expression

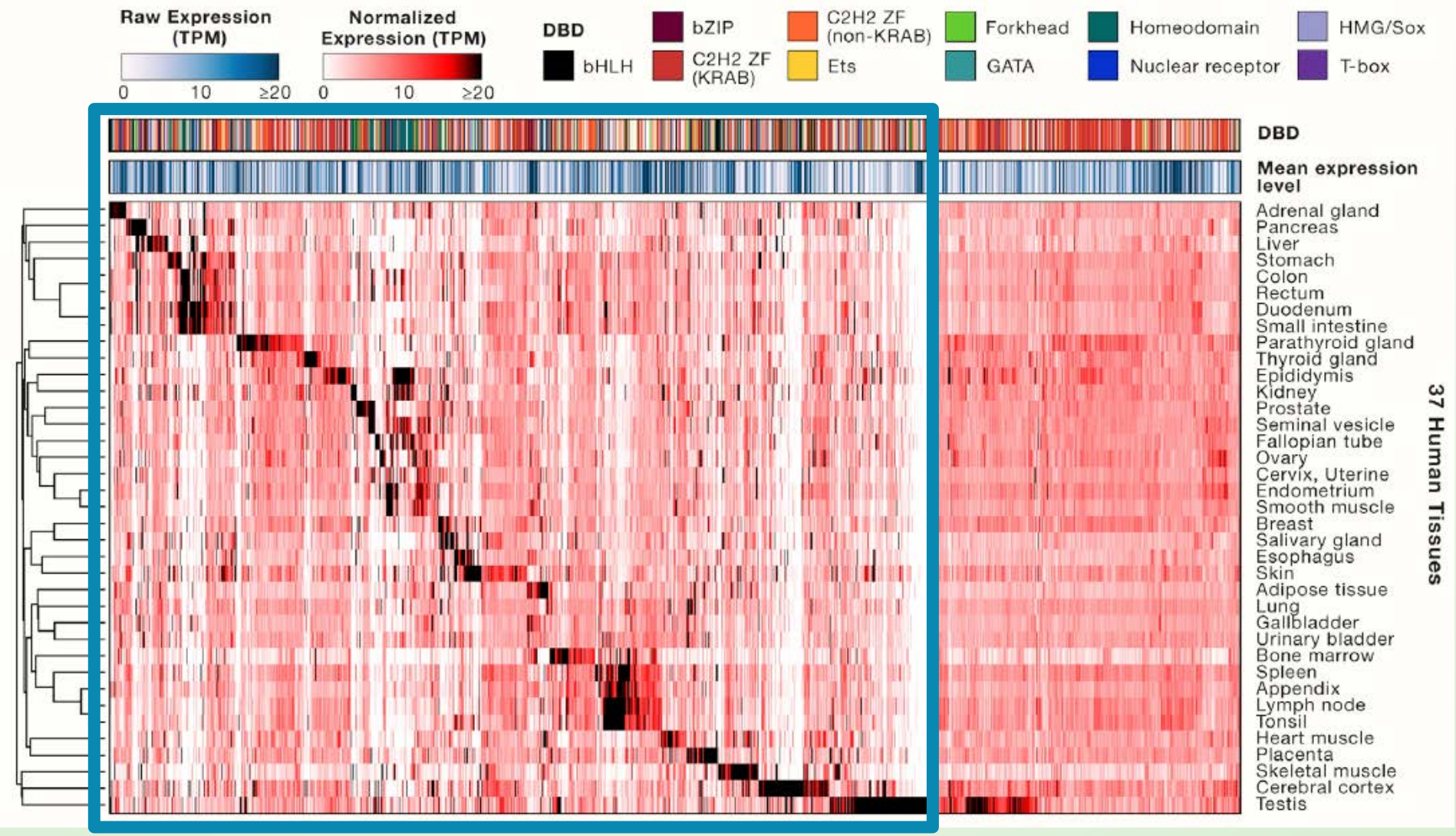
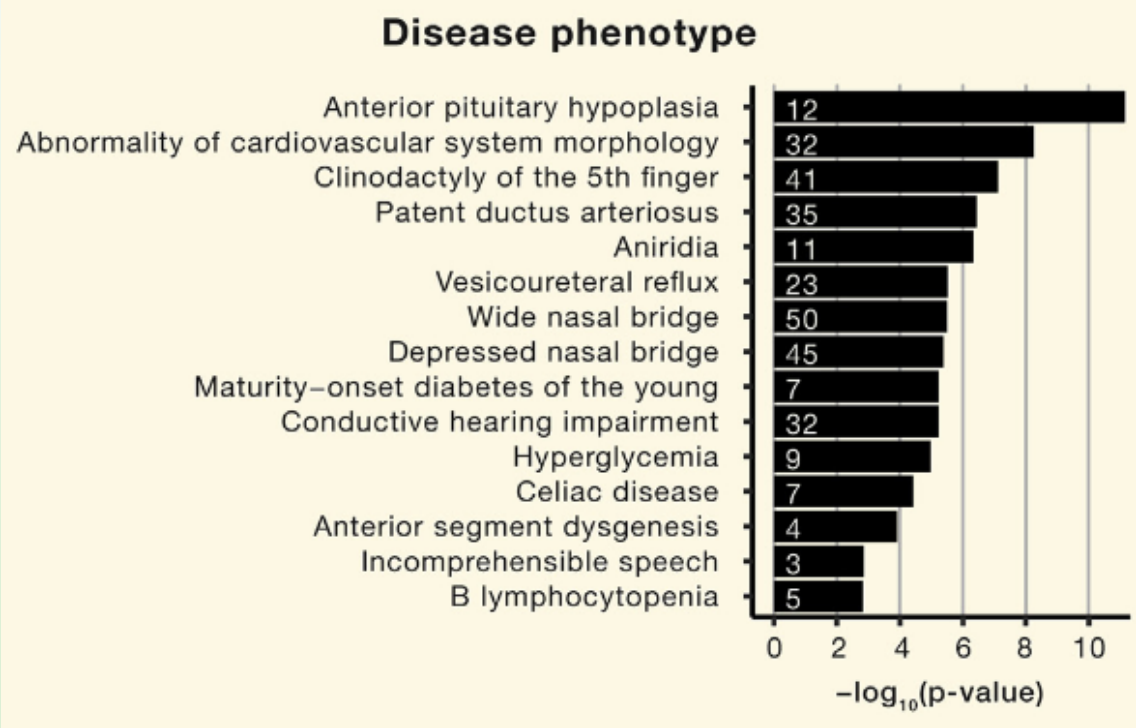


Figure adapted from Lambert et al, Cell, 2018

TRANSCRIPTION FACTORS IN HUMAN DISEASE

- TFs are associated with a varieties of diseases and phenotypes
- Human disease phenotypes are enriched for mutations within or near genes encoding TFs
- Genome-wide association studies has shown association between diseases and loci-encoding TFs



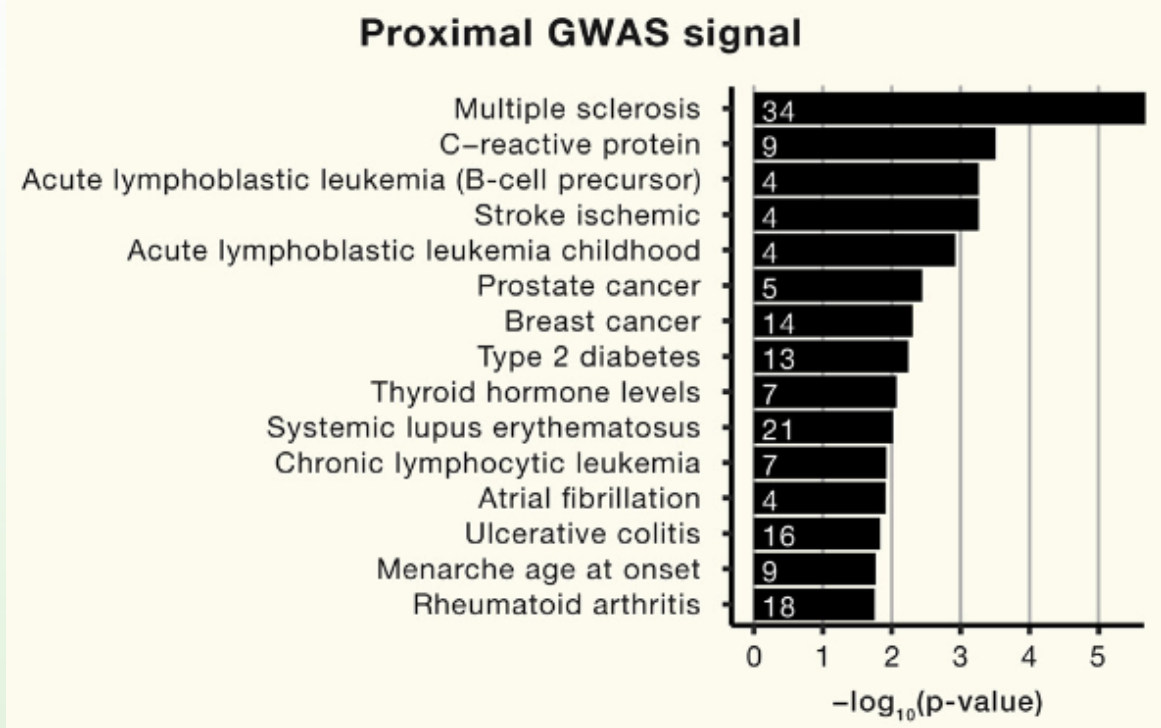


Figure adapted from Lambert et al, Cell, 2018

Transcription factor networks

- Tissue-specific function of TF is not solely regulated by differences in expression
- The same TF can regulate different genes in different cell types
- TFs regulation of gene expression and the networks of genes regulated ("regulons") are dynamic
- Important to determine how TFs are assembled in different ways to recognize binding sites and control transcription
- Important to assess in different context/tissue

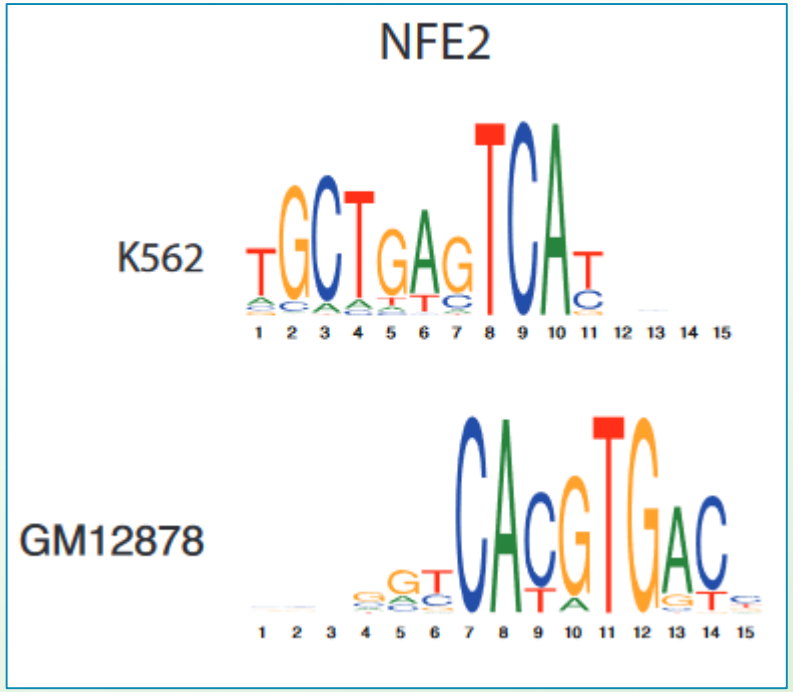


Figure adapted from Lin et al, Nucleic Acids Research, 2019

ChIP-seq

- Sequence preferences and binding sites of TFs can be assessed by a wide variety of techniques in-vitro and in-vivo
- Chromatin immunoprecipitation (ChIP) assays can be combined with sequencing (ChIP-seq))
- Powerful for identifying genome-wide DNA binding sites
- DNA-bound protein is immuno-precipitated using a specific antibody
- The bound DNA is then co-precipitated, purified, and sequenced

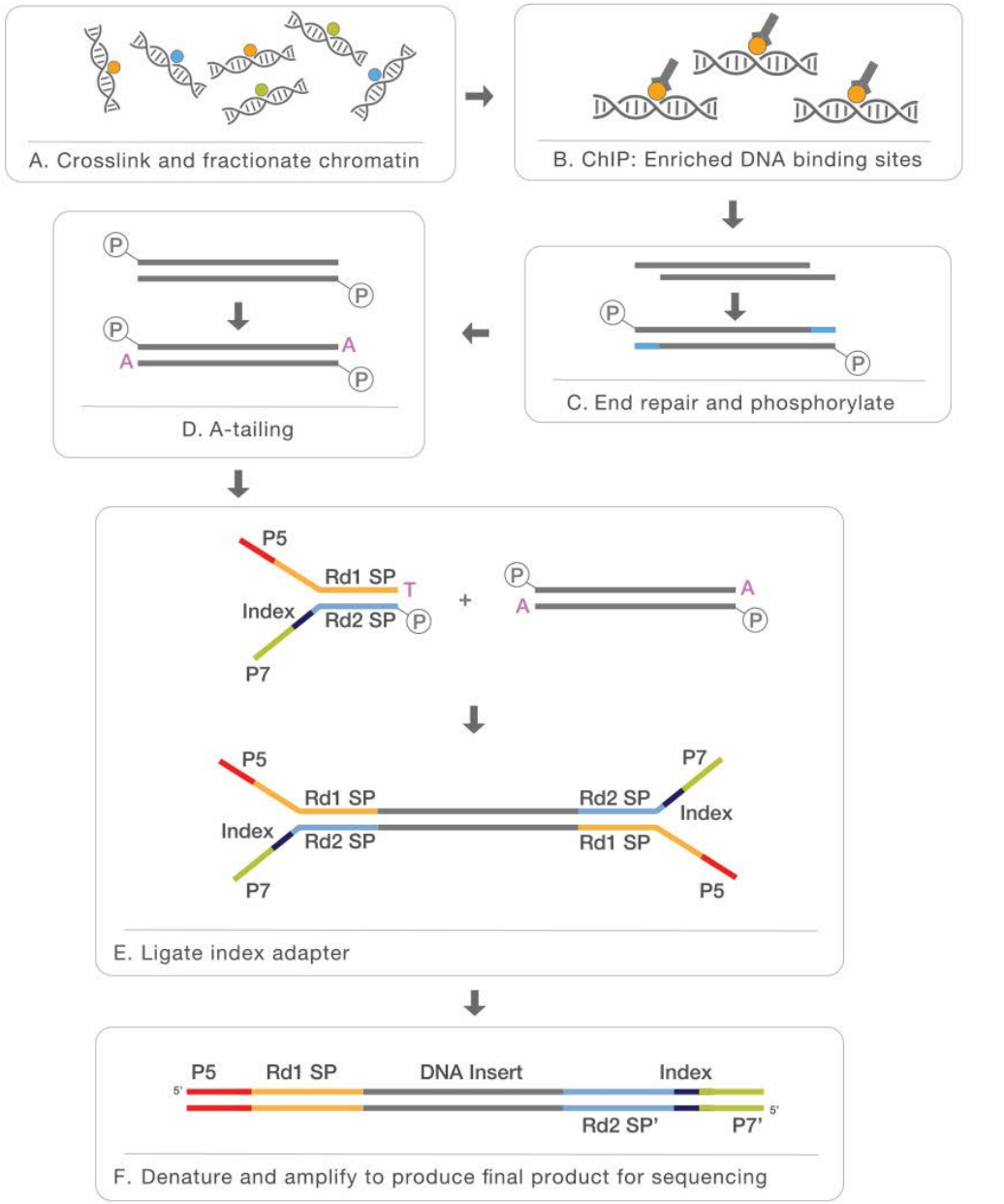
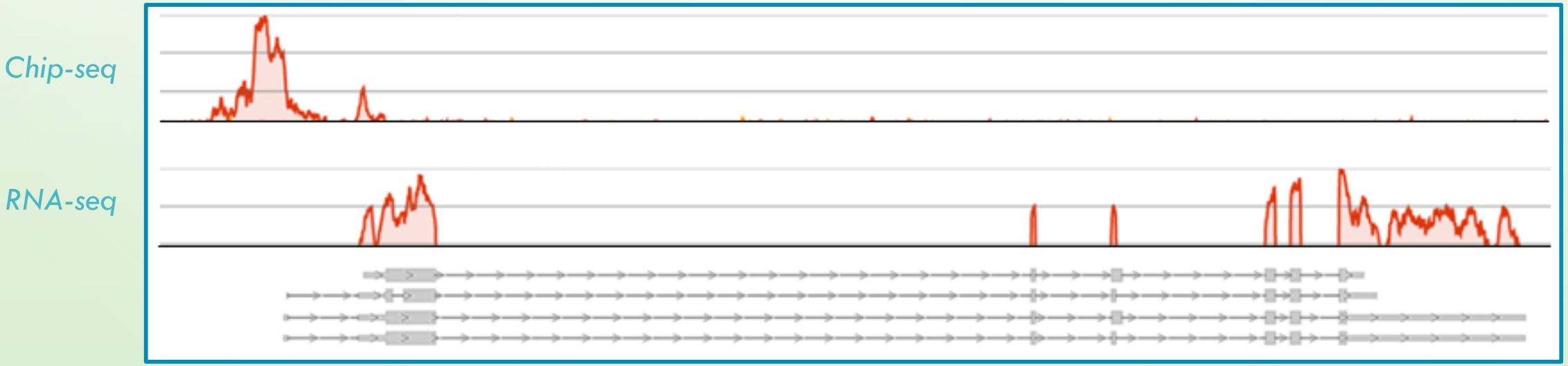


Figure adapted from https://www.illumina.com

ChIP-seq and RNAseq

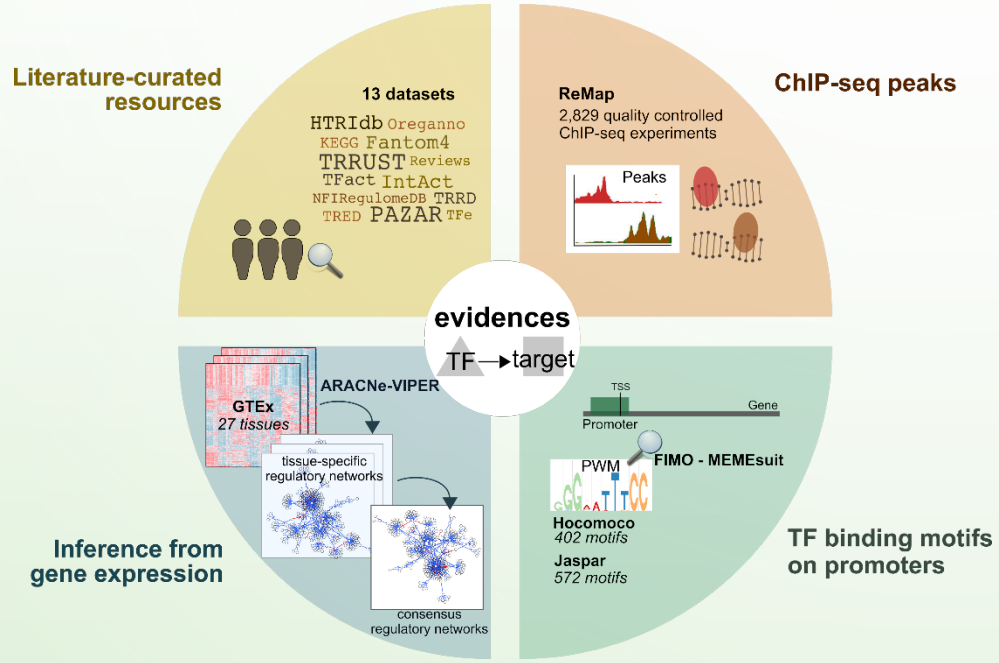
- The same samples assayed in ChIP-seq, can also be submitted for RNA-seq
- RNA provides information on gene expression (transcription)
- The advantage of combining RNA-seq and ChIP-seq in the same experiment is to link a change in occupancy with a change in transcription
- This allows inference of which peaks are functional binding sites



GENE EXPRESSION AS MEASURE OF TF ACTIVITY

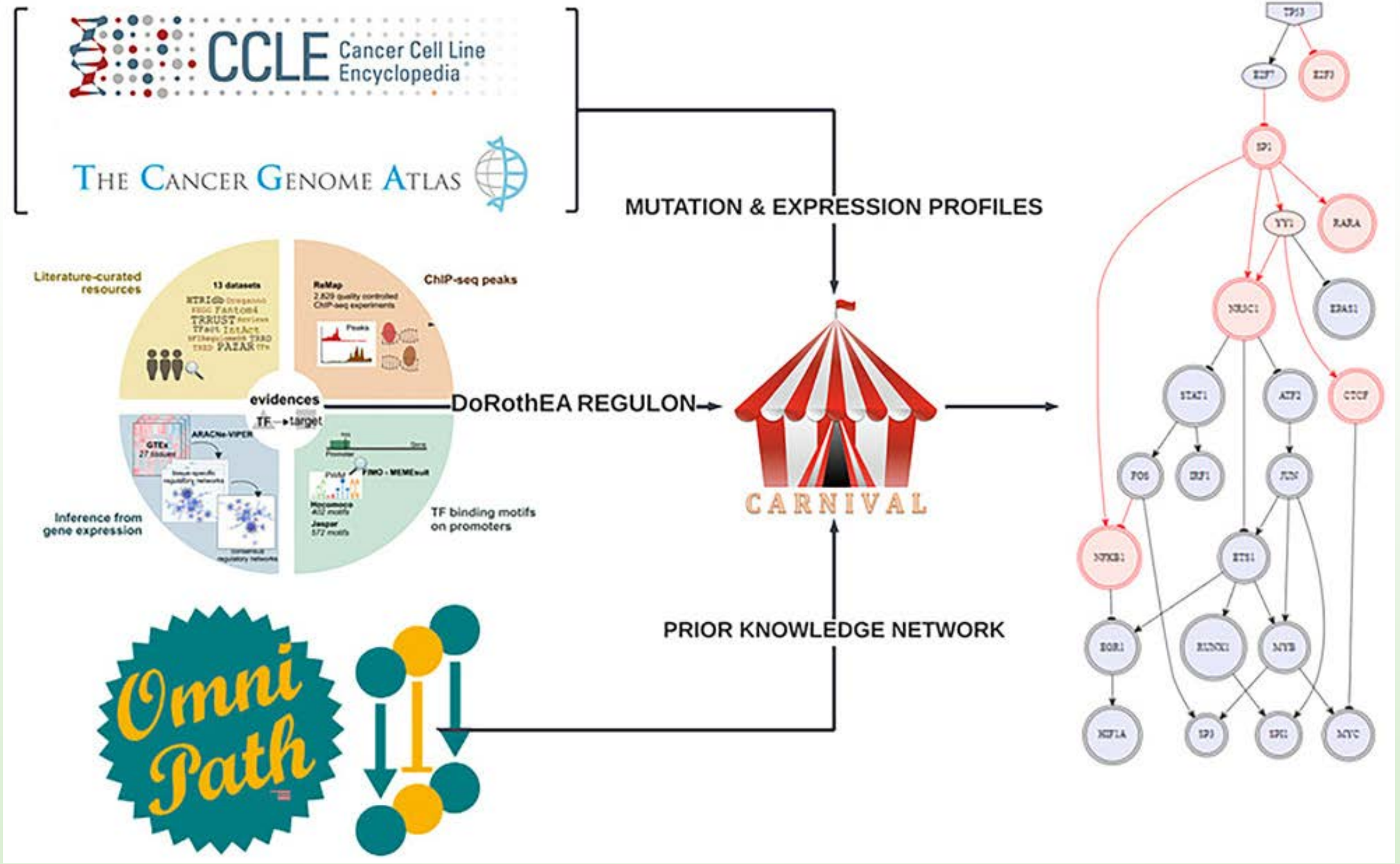
- Activity of TFs can also be estimated using their cumulative effects on expression of target genes (***regulon***). This allows to use gene expression data from clinical samples.

- Prior experimental or sequenced-based knowledge of target genes is required.



Example of sources that can be used to provide a collection of transcriptional targets of a TF (from Garcia-Alonso, Genome Research, 2019).

- As more than one TF (and other regulators) can act on the same target gene this can be noisy.
  - Consider ***global regulon signal*** not single targets
  - Consider ***large-scale analysis of all TFs***



https://www.sciencedirect.com/science/article/pii/S2589004223023684