UNIVERSIDADE FEDERAL DO CEARÁ

NPDM
Núcleo de Pesquisa e Desenvolvimento de Medicamentos

~~MTA0034 - BIOTECNOLOGIA COMPUTACIONAL (2026.2 - T01)~~

Bioinformática aplicada à medicina

ALLYSSON ALLAN

PROFESSOR VISITANTE

PPG EM MEDICINA TRANSLACIONAL

NPDM / UFC

ALLYSSON@UFC.BR

LABBAT.NPDM.UFC

PPGMTUFC

FORTALEZA

SETEMBRO 2026

CRONOGRAMA 2026.2

24/08 - CONCEITOS BÁSICOS DE GENÉTICA;
26/08 - CONCEITOS BÁSICOS DE PROGRAMAÇÃO EM R;
31/08 - BIOINFORMÁTICA EM R;
02/09 - GENÔMICA E TRANSCRIPTÔMICA;
09/09 - BANCOS DE DADOS DE ÔMICAS E CLÍNICOS;
14/09 - ONCOLOGIA DE PRECISÃO;
16/09 - DOENÇAS RARAS;
21/09 - MEDICINA DE PRECISÃO E IMPLEMENTAÇÃO CLÍNICA.

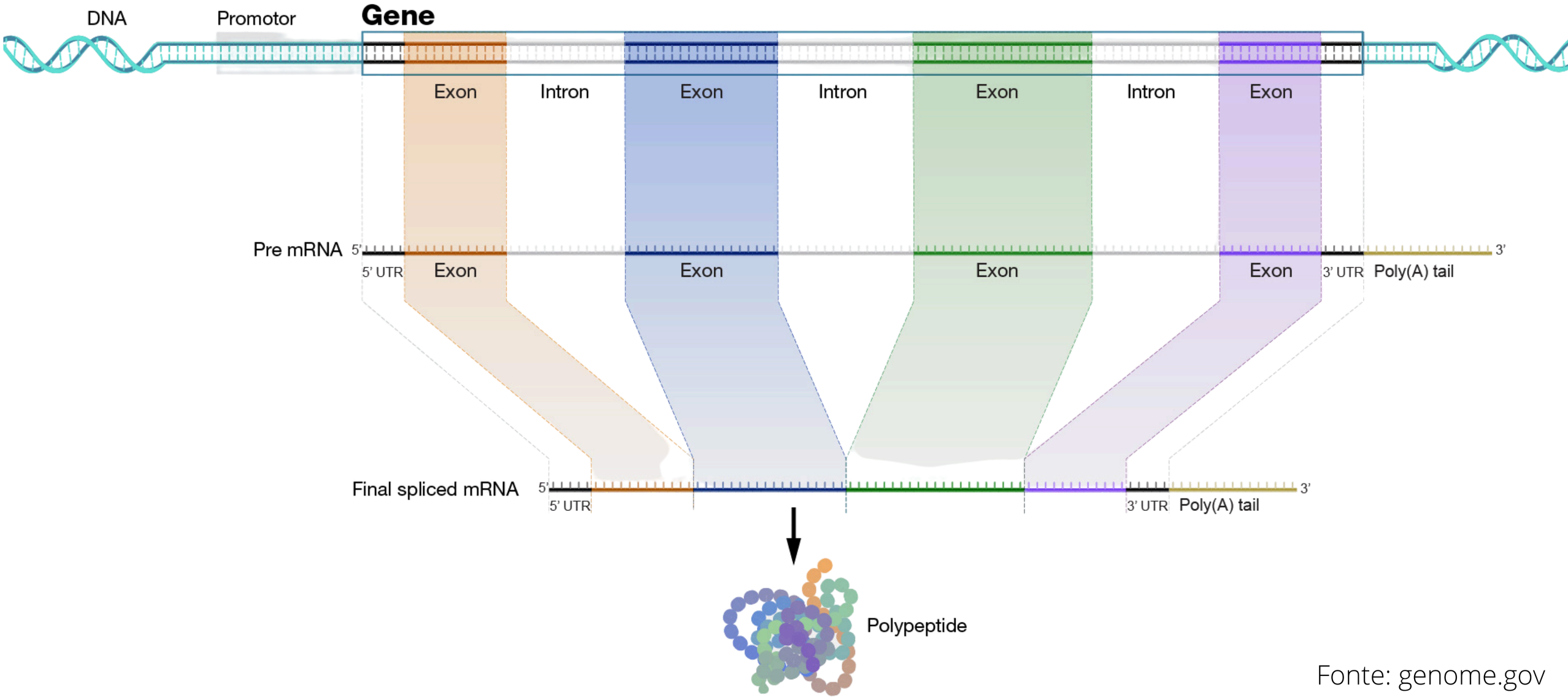
TABELA DE HORÁRIOS

8H - ÔMICAS, O QUE SÃO?;
9H30 - VISITA AOS SEQUENCIADORES NA CEGENBIO;
10H15 - IDENTIFICANDO DADOS ÔMICOS NO R;
10H45 - EXERCÍCIO 1;
11H45 - DÚVIDAS.

OBJETIVOS

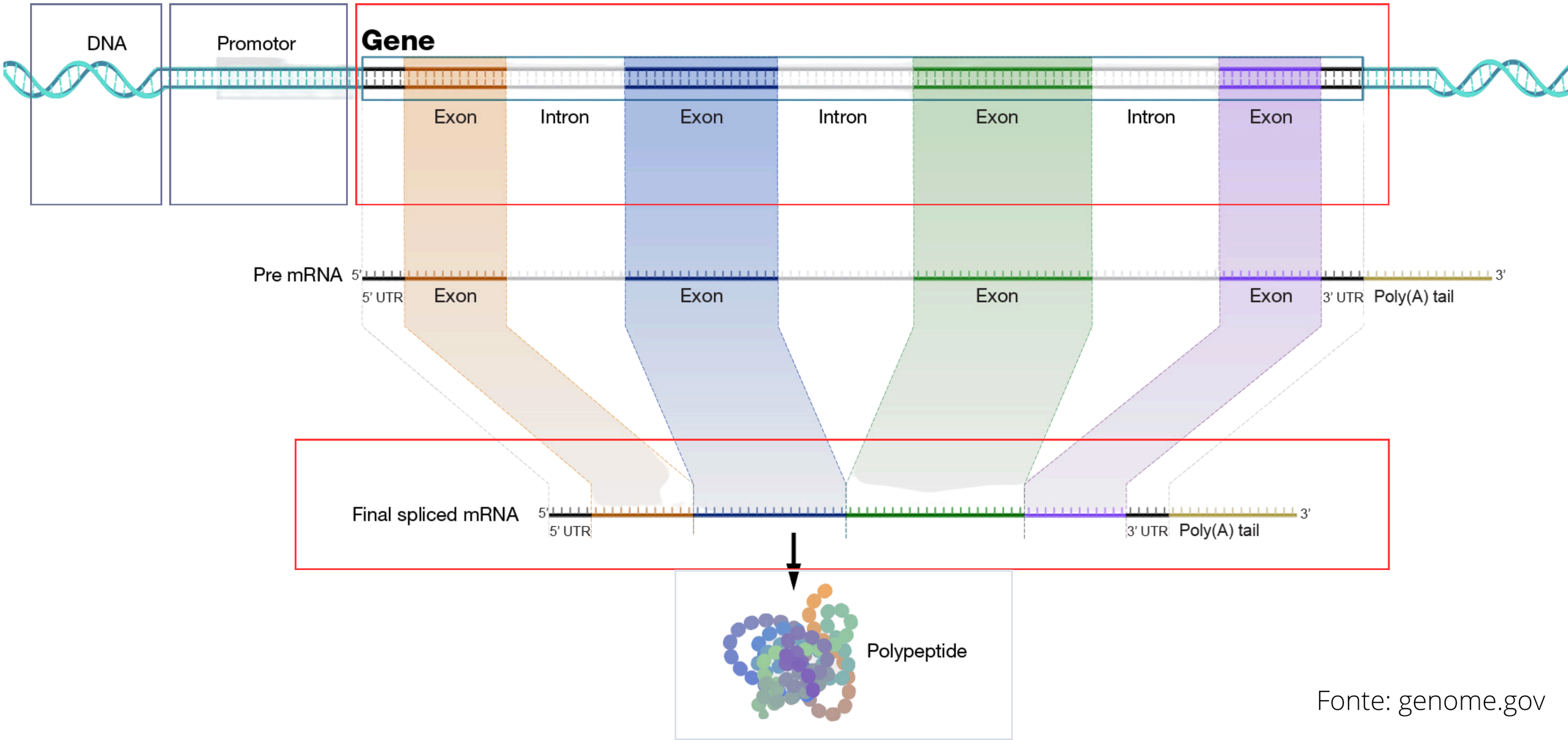
- COMPREENDER O FLUXO DO SEQUENCIAMENTO DE NOVA GERAÇÃO;
- COMPREENDER COMO FUNCIONA O MANEJO DE DADOS ÔMICOS PRODUZIDOS EM UM SEQUENCIADOR DE NOVA GERAÇÃO;
- EXPLORAR DADOS DE GENÔMICA E TRANSCRIPTÔMICA USANDO O R.

Recapitulando



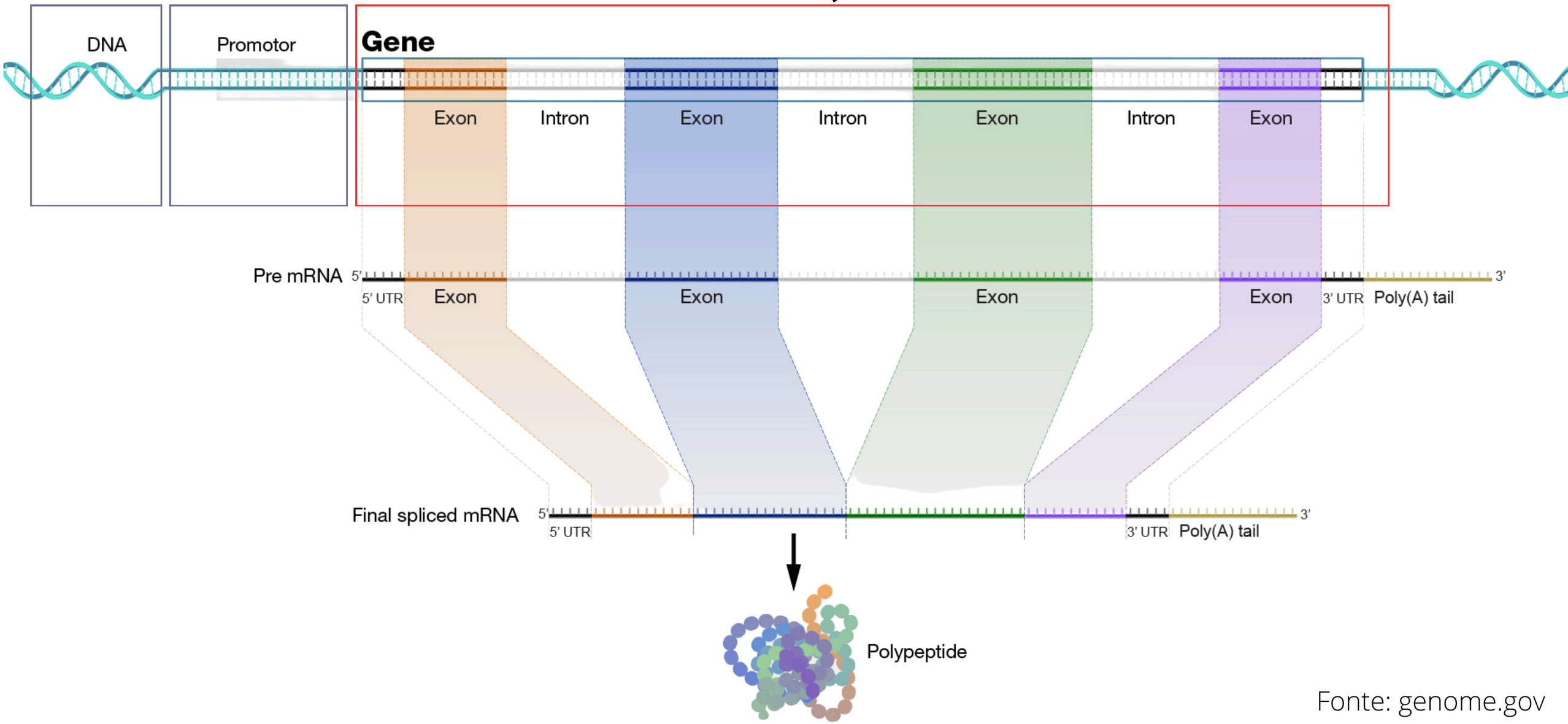
Fonte: genome.gov

Qual molécula responde o seu problema?



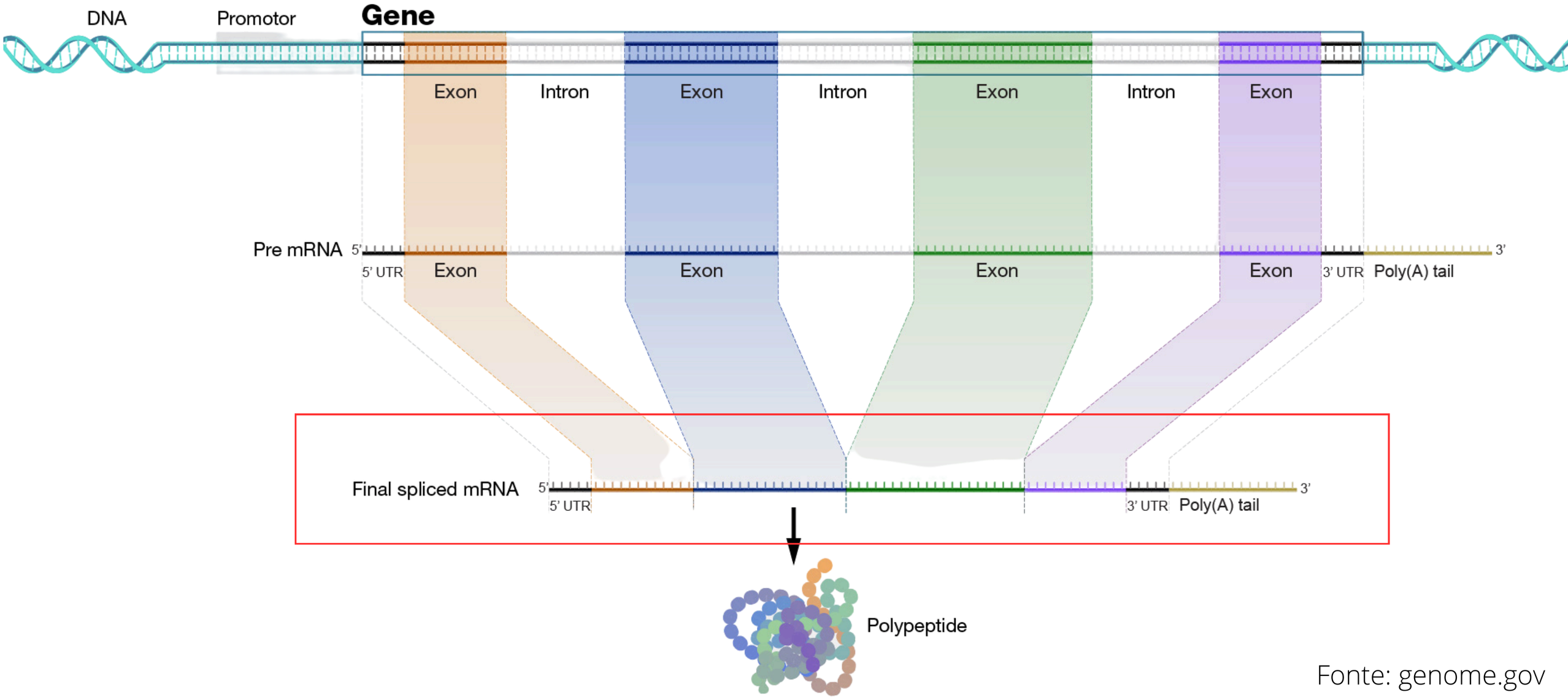
Fonte: genome.gov

Variação: painéis, amplicons, WES, short-read WGS, long-read WGS, cfDNA.



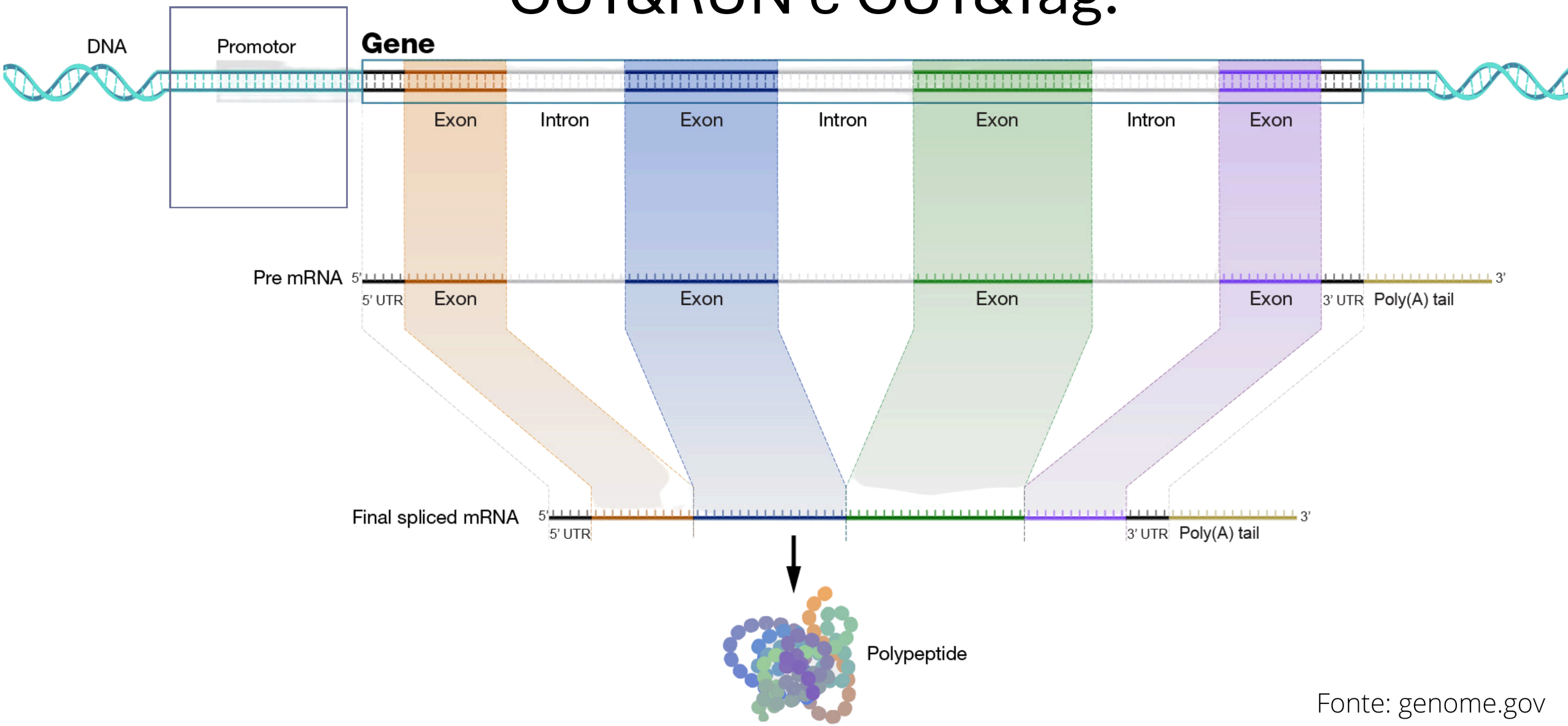
Fonte: genome.gov

Expressão gênica: bulk RNA-seq, small RNA-seq, long-read RNA-seq.



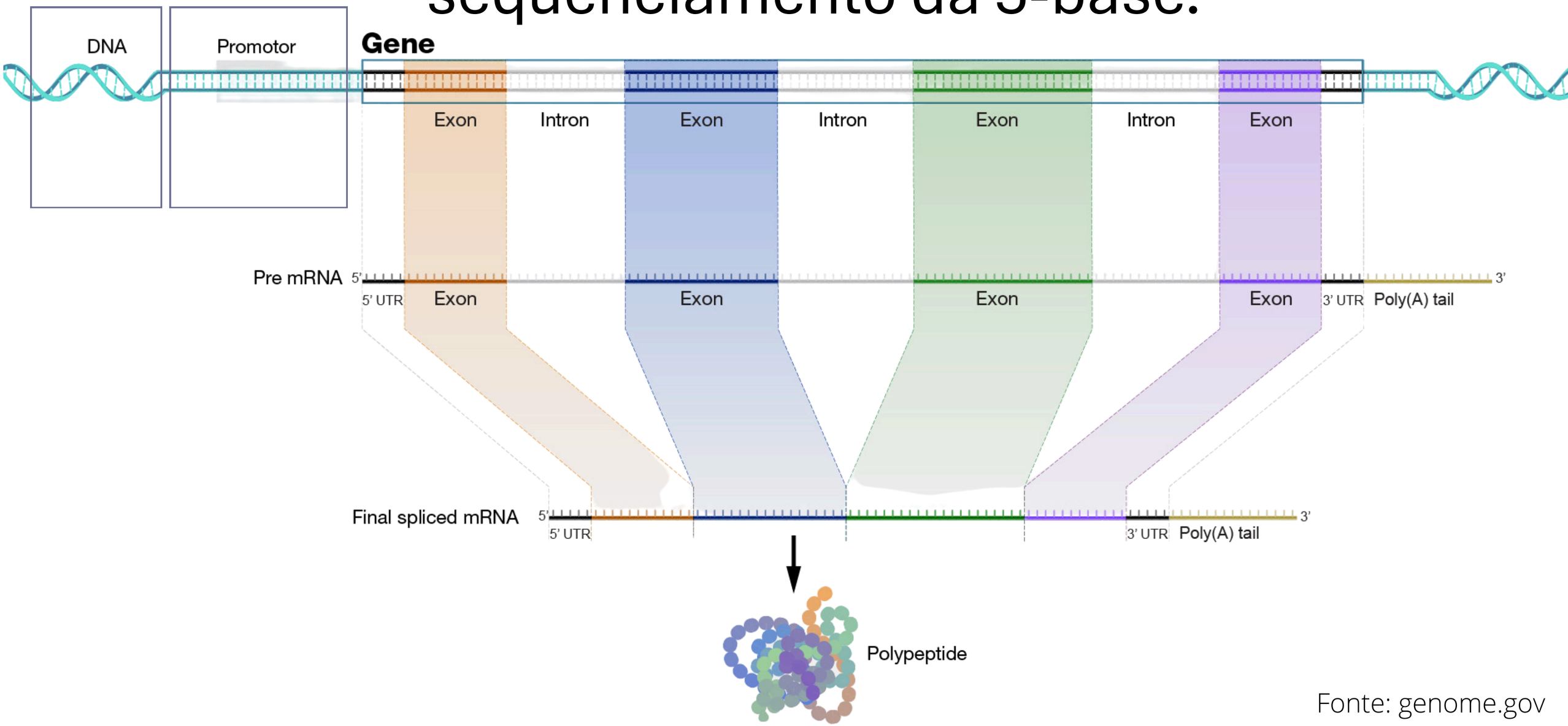
Fonte: genome.gov

Regulação gênica: ATAC-seq, ChIP-seq, CUT&RUN e CUT&Tag.



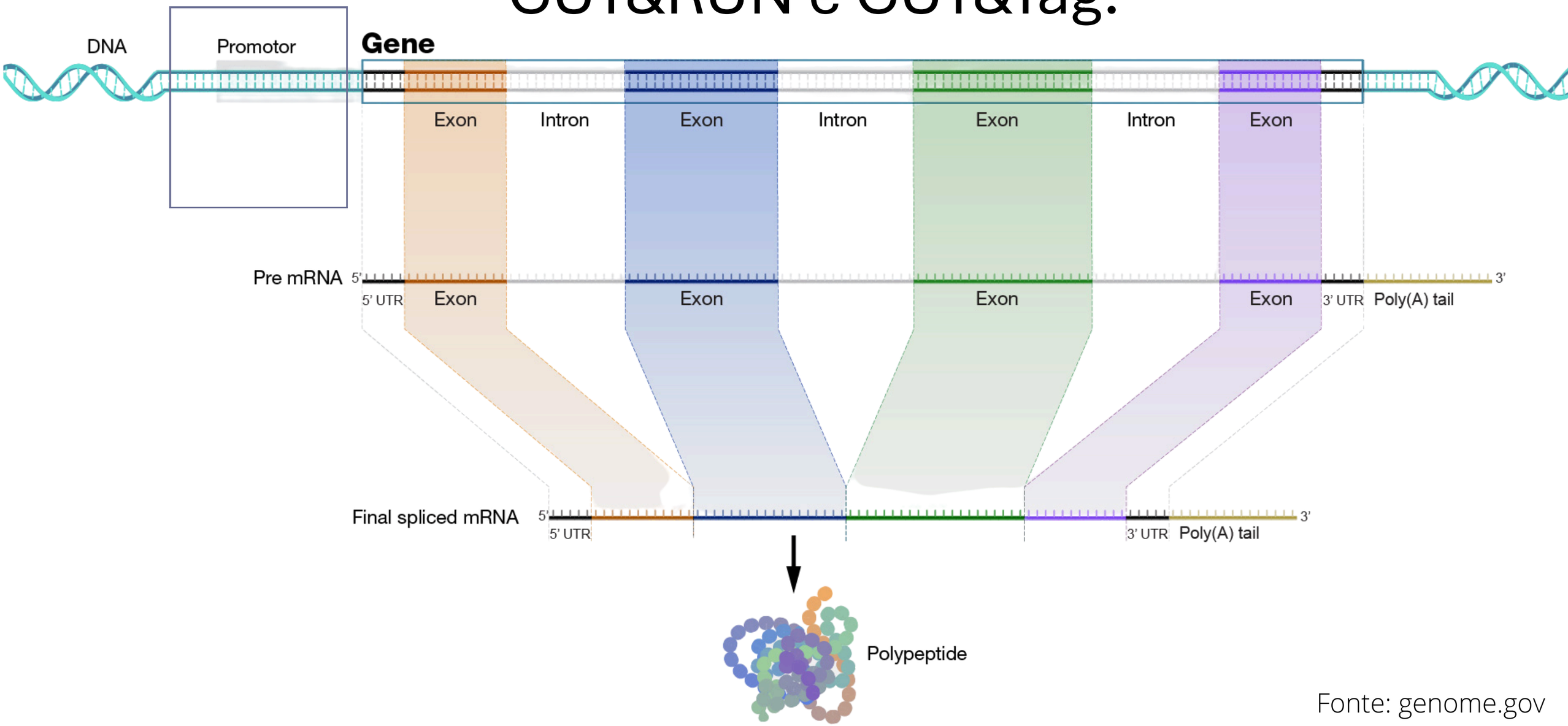
Fonte: genome.gov

Epigenética: WGBS, EM-seq, metilação direta e sequenciamento da 5-base.



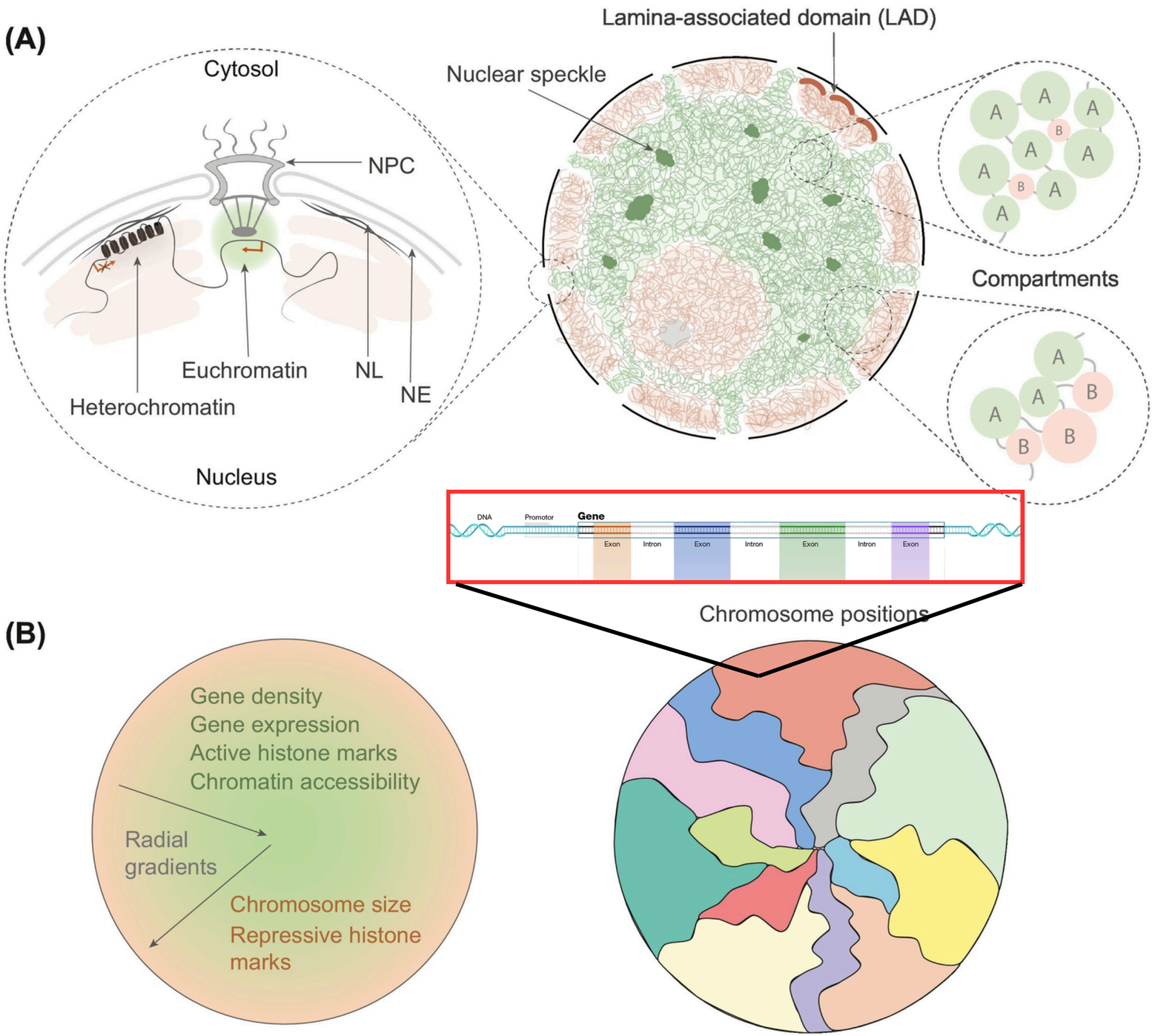
Fonte: genome.gov

Regulação gênica: ATAC-seq, ChIP-seq, CUT&RUN e CUT&Tag.



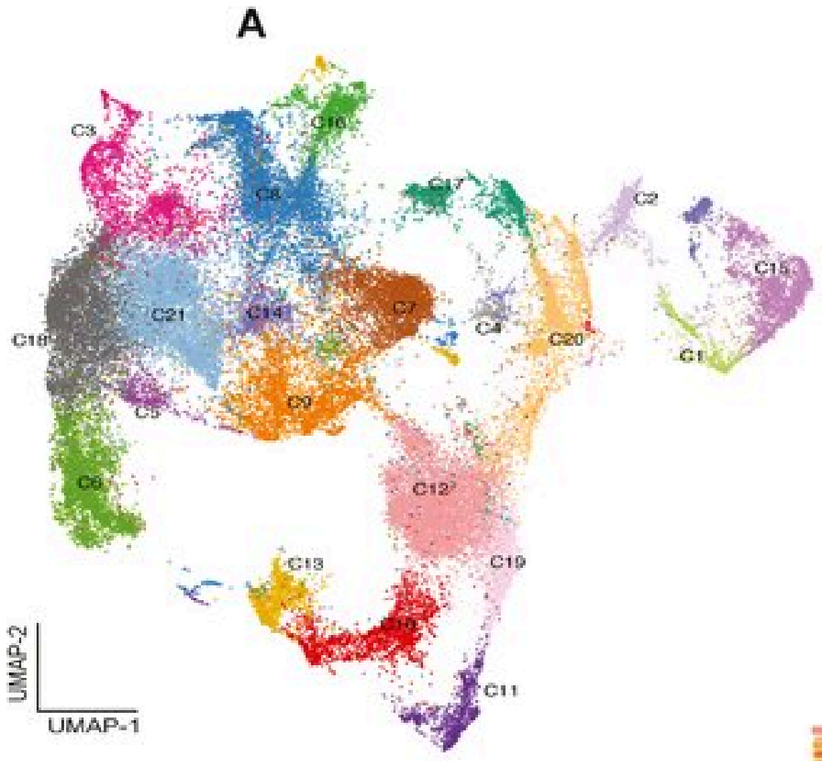
Fonte: genome.gov

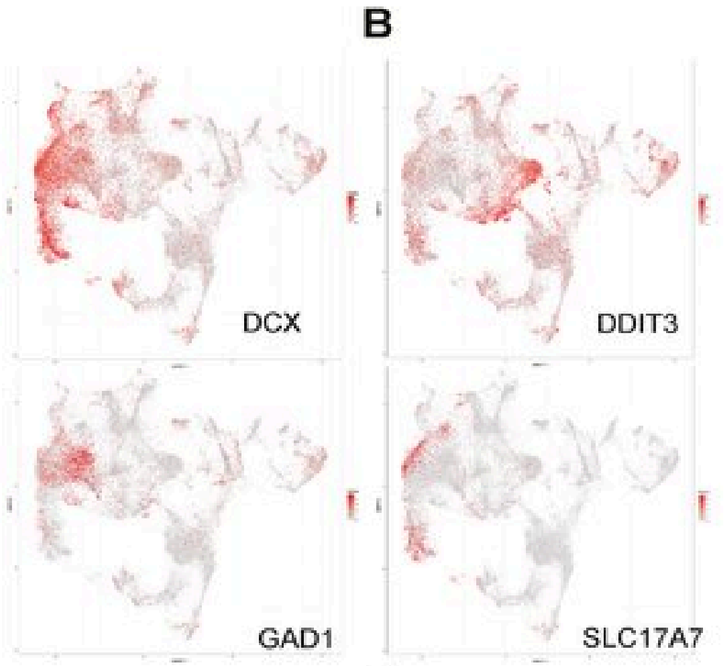
Organização do genoma: Hi-C, Micro-C, HiChIP e Pore-C.

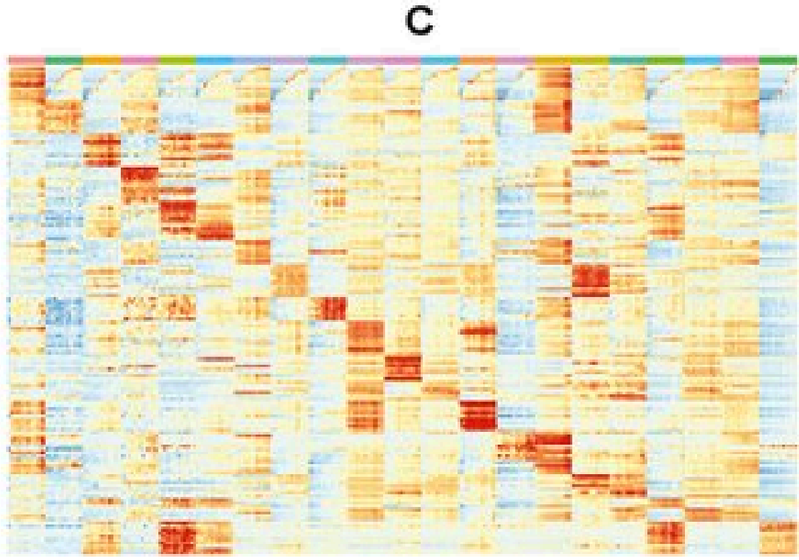


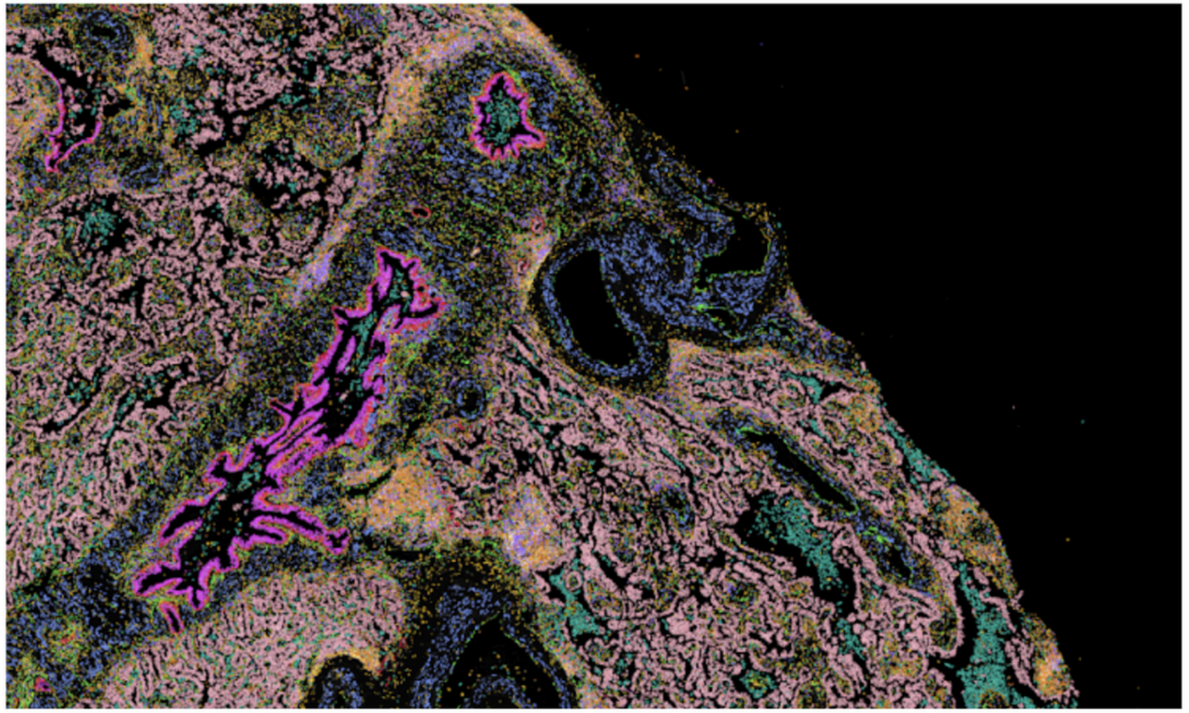
Trends in Cell Biology

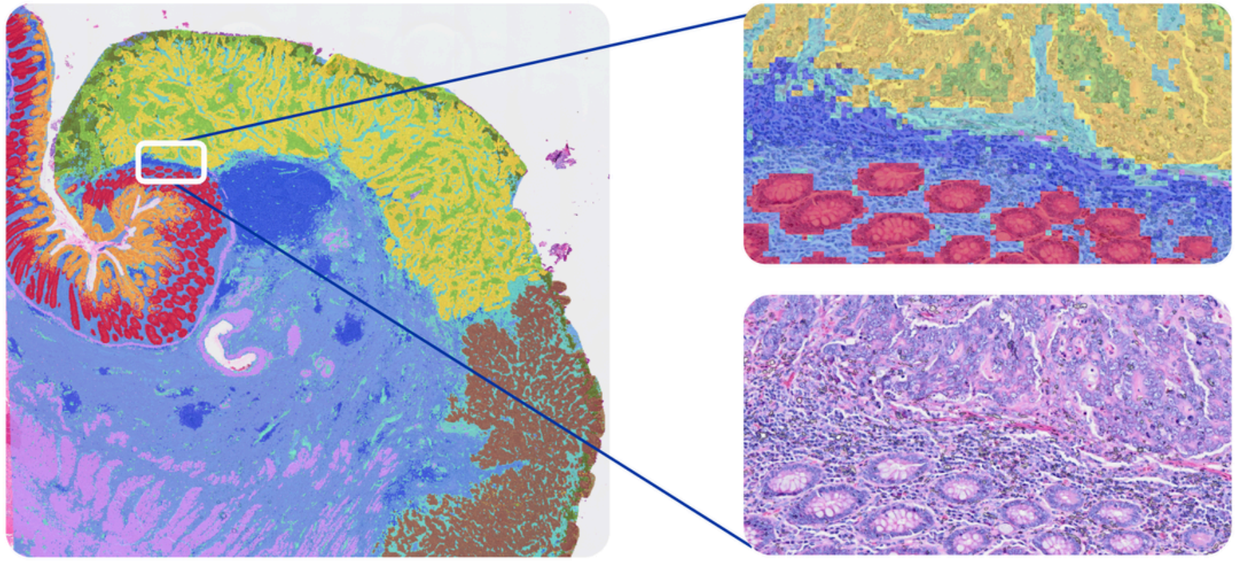
- **Heterogeneidade celular:** scRNA-seq, snRNA-seq, scATAC-seq, CITE-seq, Multiome and Perturb-seq.

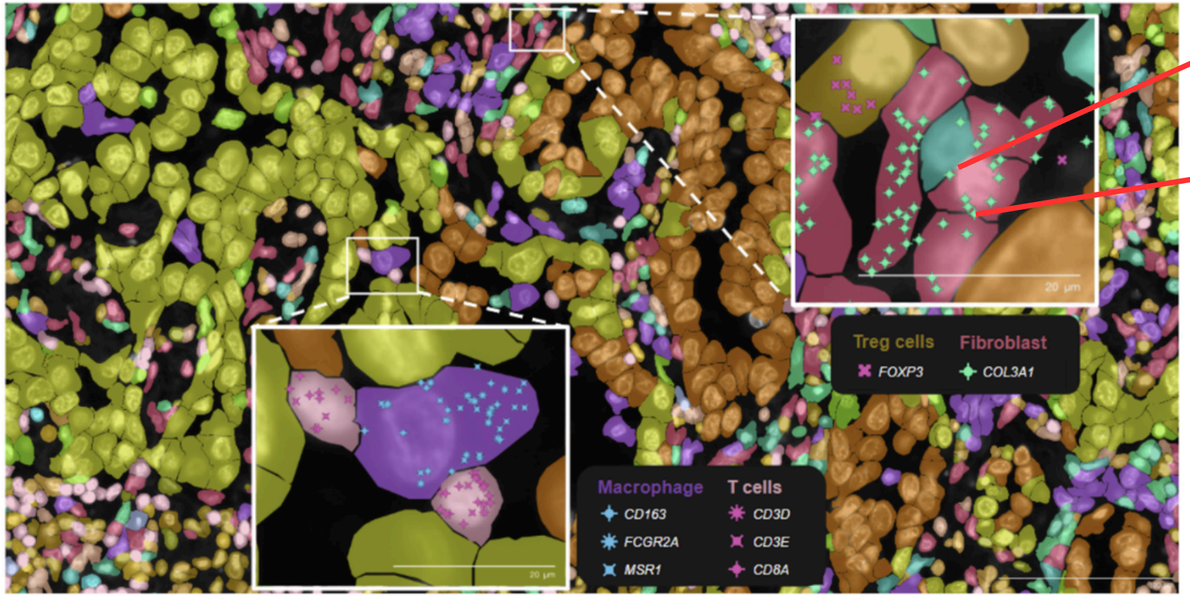


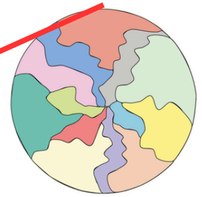












Biologia espacial: Visium, Xenium, CosMx, MERSCOPE, GeoMx, StrataMap, Atera e AVITI24.

Genoma microbiano: 16S, ITS, metagenômica por shotgun, metatranscriptômica e captura patogênica.



Como se dá o fluxo de preparação da amostra antes do sequenciamento?

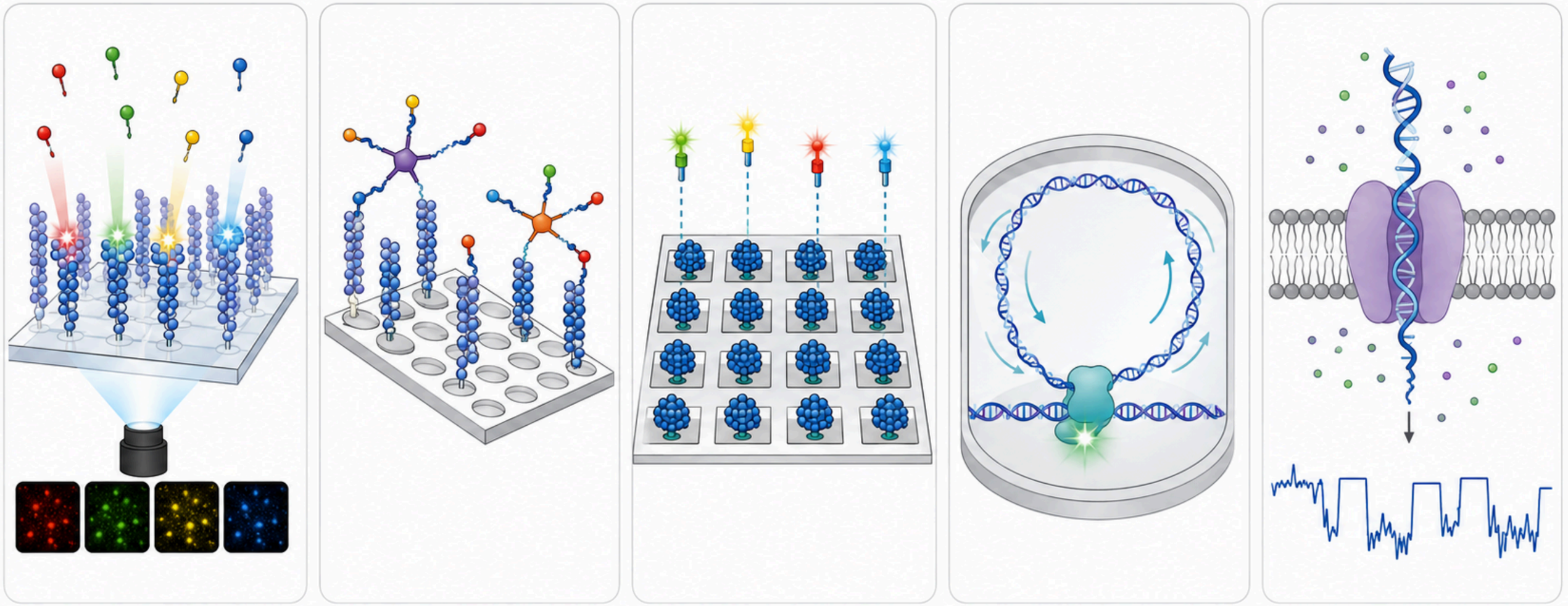
Extração de RNA / DNA ⟶ Preparação da biblioteca ⟶ Sequenciamento massivo





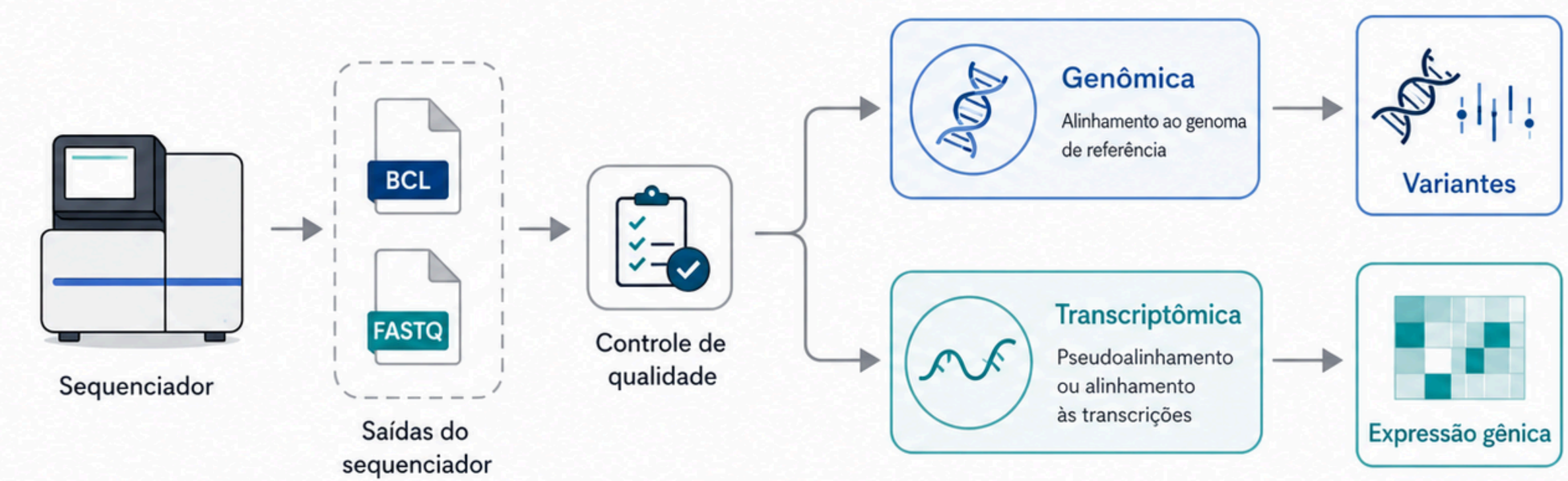


Tecnologias para sequenciamento: Illumina SBS, Element avidity, MGI DNBSEQ, PacBio HiFi e Oxford Nanopore.

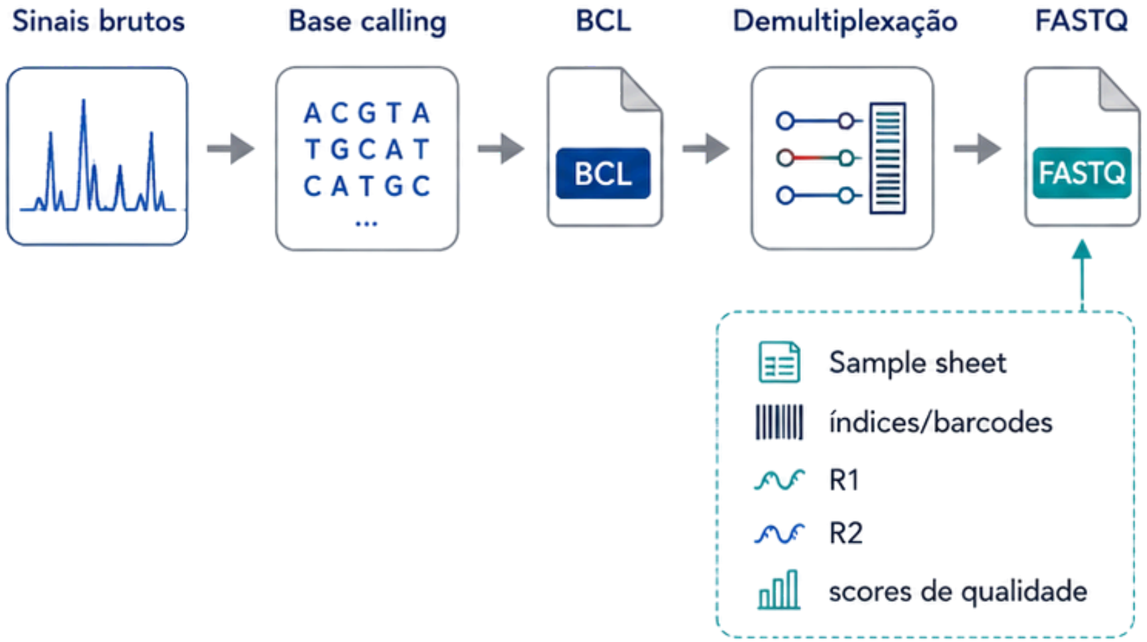


Intervalo para conhecer o sequenciador na Cegenbio

Genoma e transcriptoma



O que sai do sequenciador?



O que significam

BCL
base calls por ciclo

FASTQ
sequência + qualidade por leitura

Plataformas principais

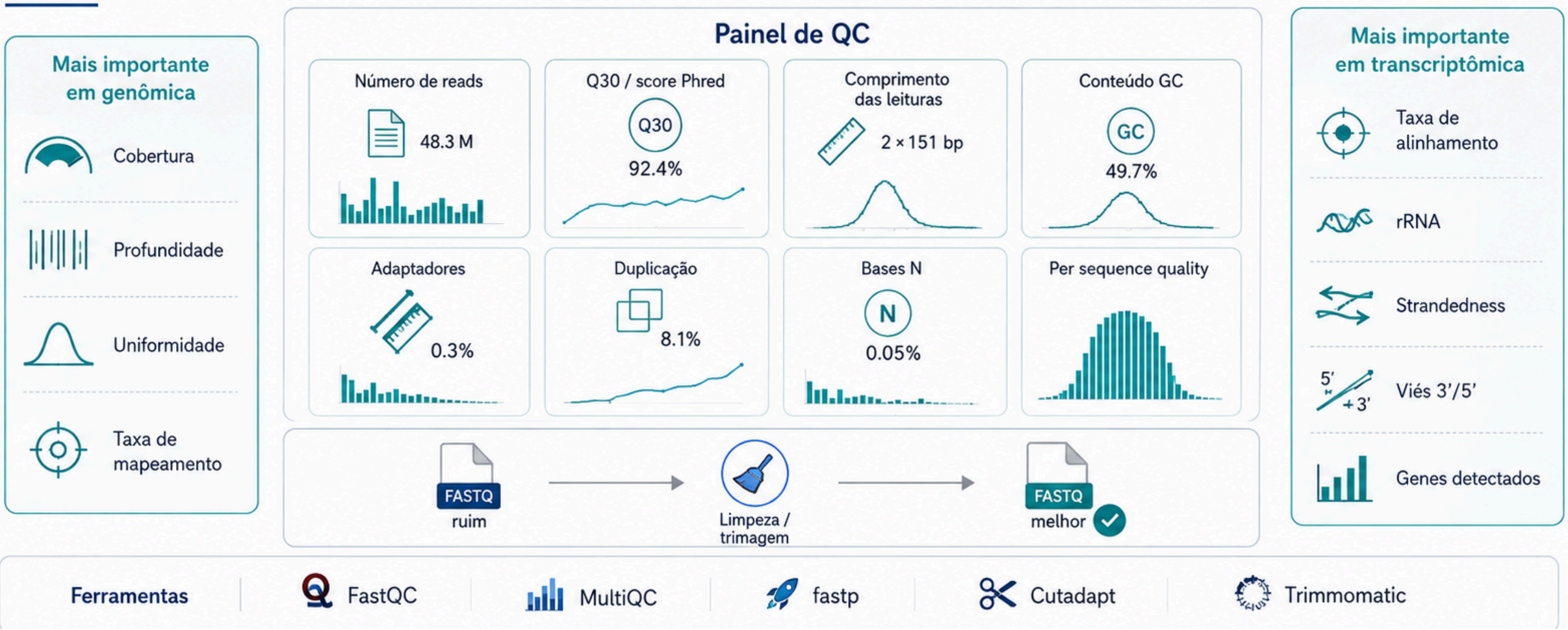
Illumina
leituras curtas, BCL e FASTQ

Oxford Nanopore
leituras longas, FASTQ

PacBio HiFi
leituras longas de alta acurácia

Ion Torrent
semicondutor

Controle de qualidade



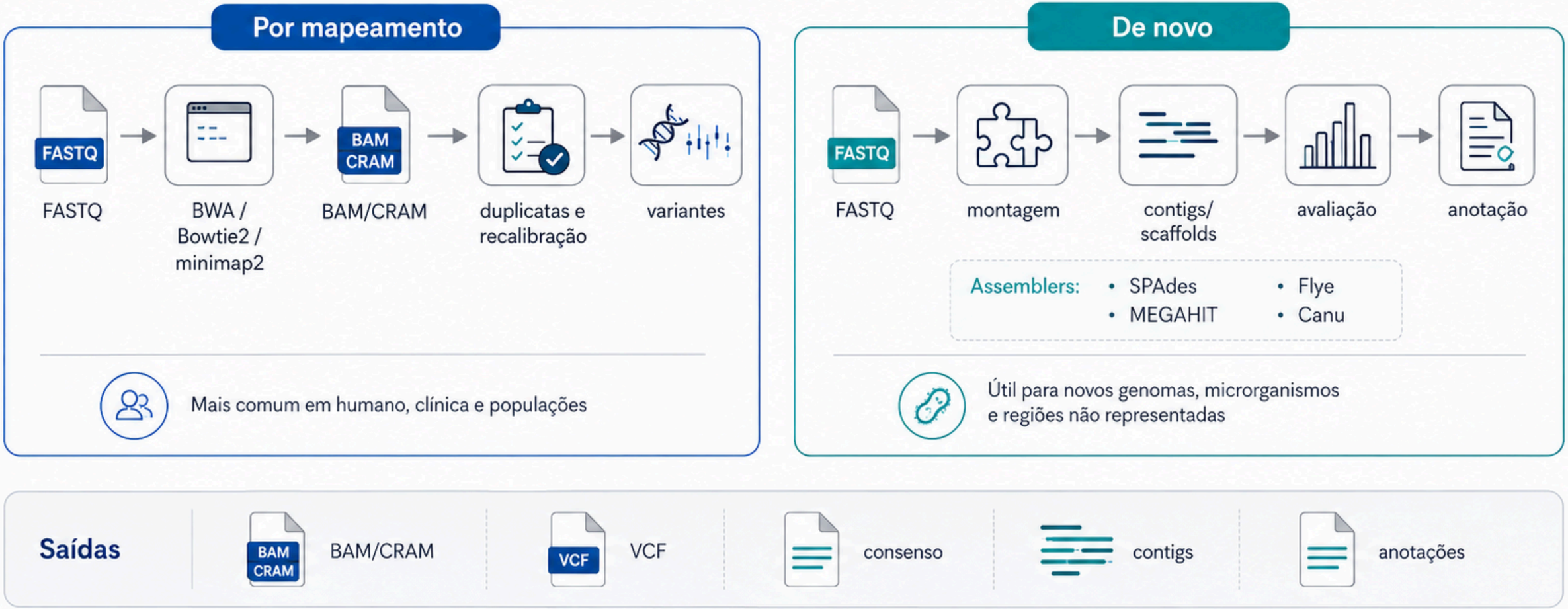
Genômica

1. DNA / biblioteca
2. FASTQ
3. QC
4. Alinhamento ao genoma de referência ou de *novo*
5. BAM/CRAM ou contigs
6. Chamada de variantes
7. Anotação
8. Pergunta biológica → Variantes

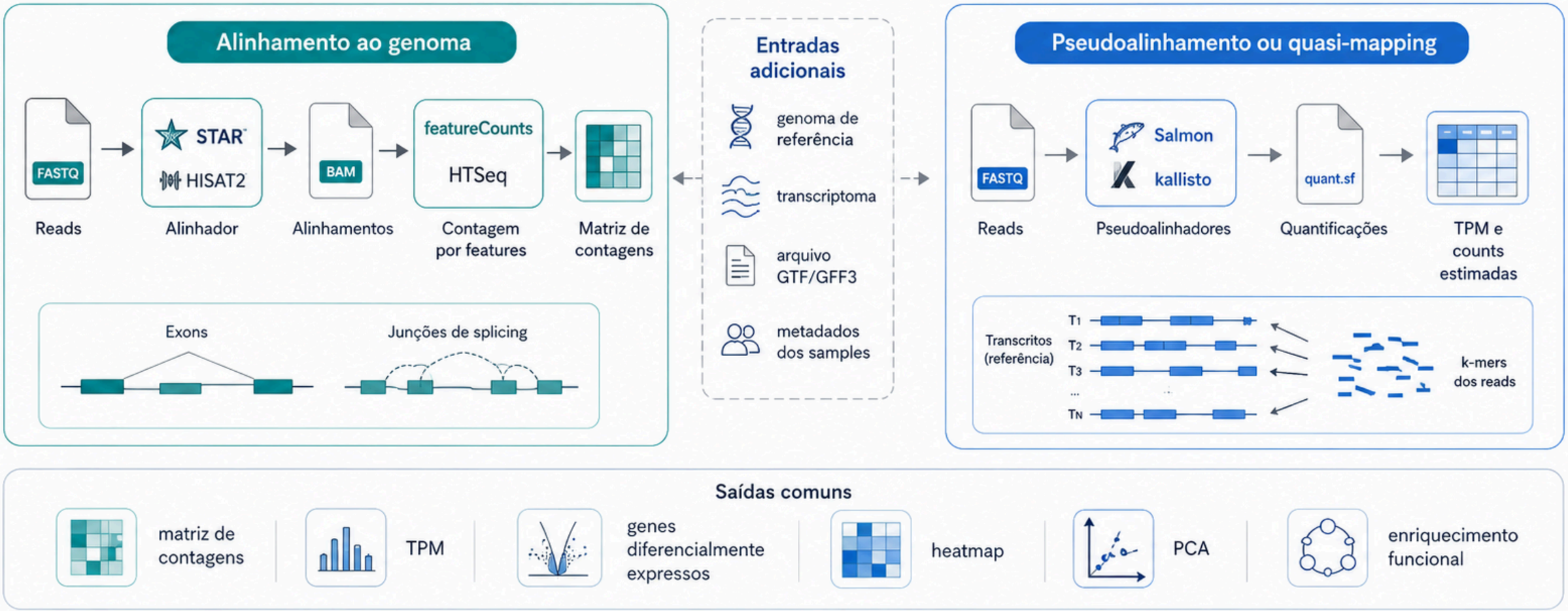
Transcriptômica

1. RNA / biblioteca
2. FASTQ
3. QC
4. Alinhamento *spliced* ao genoma ou pseudoalinhamento ao transcriptoma
5. Contagens / TPM
6. Normalização
7. Expressão diferencial
8. Interpretação funcional → Heatmap / PCA

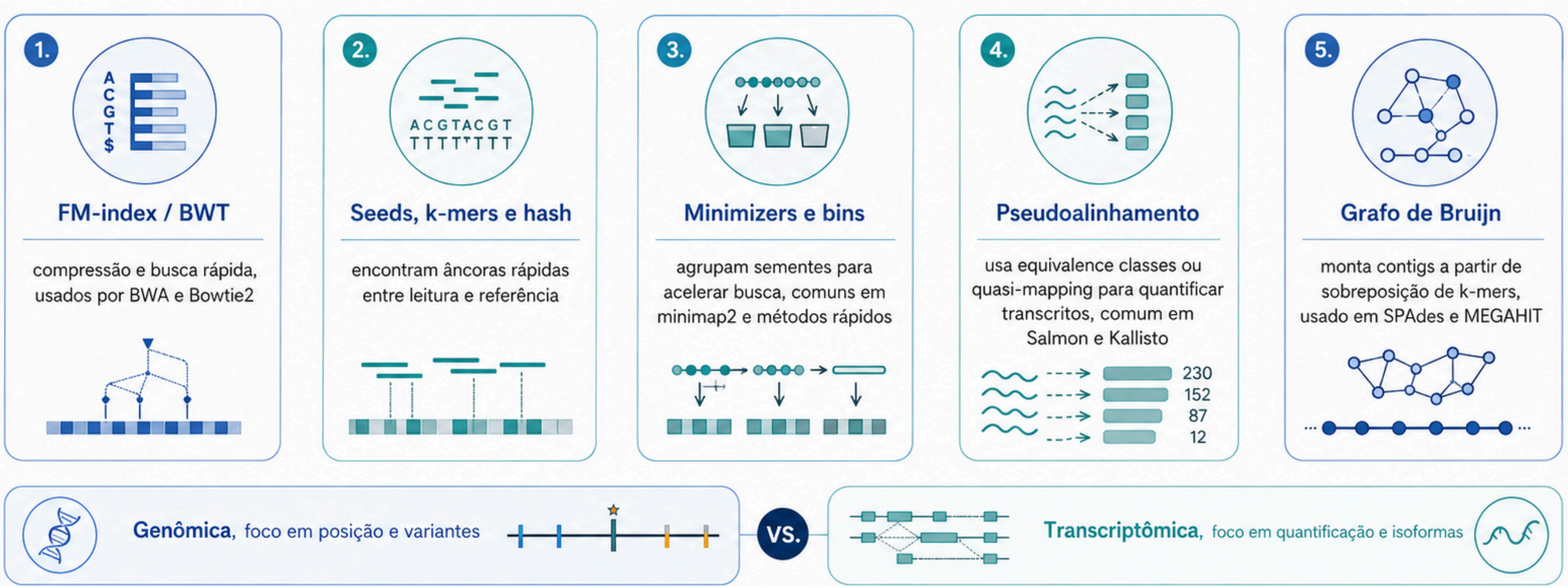
Genômica - análise secundária



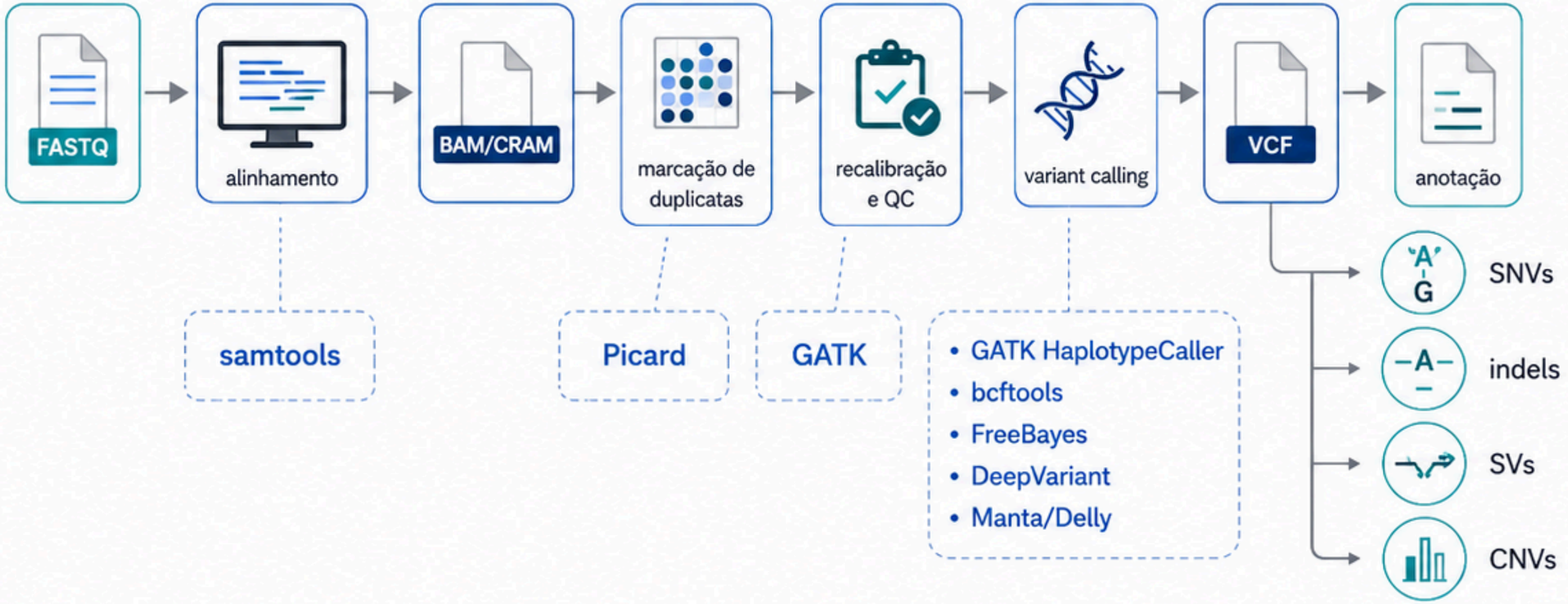
Transcriptômica - análise secundária



Principais algoritmos de alinhamento



Chamada de variantes



Referências e recursos para hg38 / GRCh38

- Genoma de referência GRCh38
- GENCODE / Ensembl
- UCSC Genome Browser
- dbSNP
- ClinVar
- gnomAD
- 1000 Genomes
- Mills indels e outros known sites

Exercício em R - Genômica

Instalar o bioconductor antes

Entradas



FASTQ → BAM/CRAM → sample.vcf.gz +.tbi

Perguntas

- Quantas variantes?
- Qual o QUAL?
- Quais cromossomos?
- Qual profundidade?

Código em R

```
library(VariantAnnotation)

vcf <- readVcf('sample.vcf.gz', 'hg38')

q <- qual(vcf)

keep <- q > 30

sum(keep)

head(rowRanges(vcf[keep]))
```

Saídas

número de variantes

1,234,567

histograma de QUAL

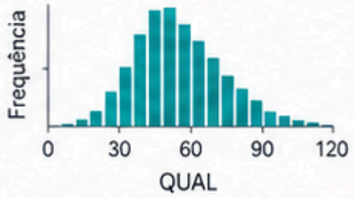


tabela CHROM POS REF ALT

CHROM	POS	REF	ALT
1	10177	A	G
1	10583	C	T
2	20001	G	A
X	15023	T	C
…	…	…	…

variantes filtradas



QUAL > 30

456,789 variantes

Use o seguinte prompt em sua IA de interesse: Encontre um vcf para o fenótipo XYZ abaixo de 100MB em um repositório de sequências, pode ser EGA, por exemplo.

Exercício em R - Transcriptômica

Entradas





sample sheet

Perguntas

- Quantos genes foram detectados?
- Como os grupos se separam?
- Quais genes mudam?

Código em R

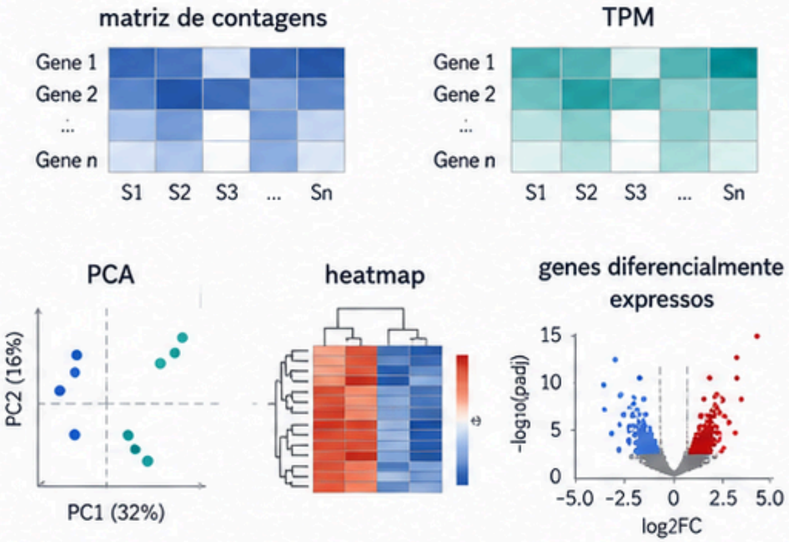
```
library(tximport)
library(DESeq2)

files <- c('S1/quant.sf','S2/quant.sf')
txi <- tximport(files, type='salmon',
                txOut=TRUE)

dds <- DESeqDataSetFromTximport(txi,
       colData=samples, design=~grupo)

dds <- DESeq(dds)
res <- results(dds)
```

Saídas



Diferenças entre os dois fluxos	Entradas	Saídas
Genômica	Entradas: FASTQ, BAM, VCF	Saídas: variantes, anotação, priorização
Transcriptômica	Entradas: FASTQ, quant.sf ou counts	Saídas: contagens, TPM, DEGs, vias