

# LU3IN013:

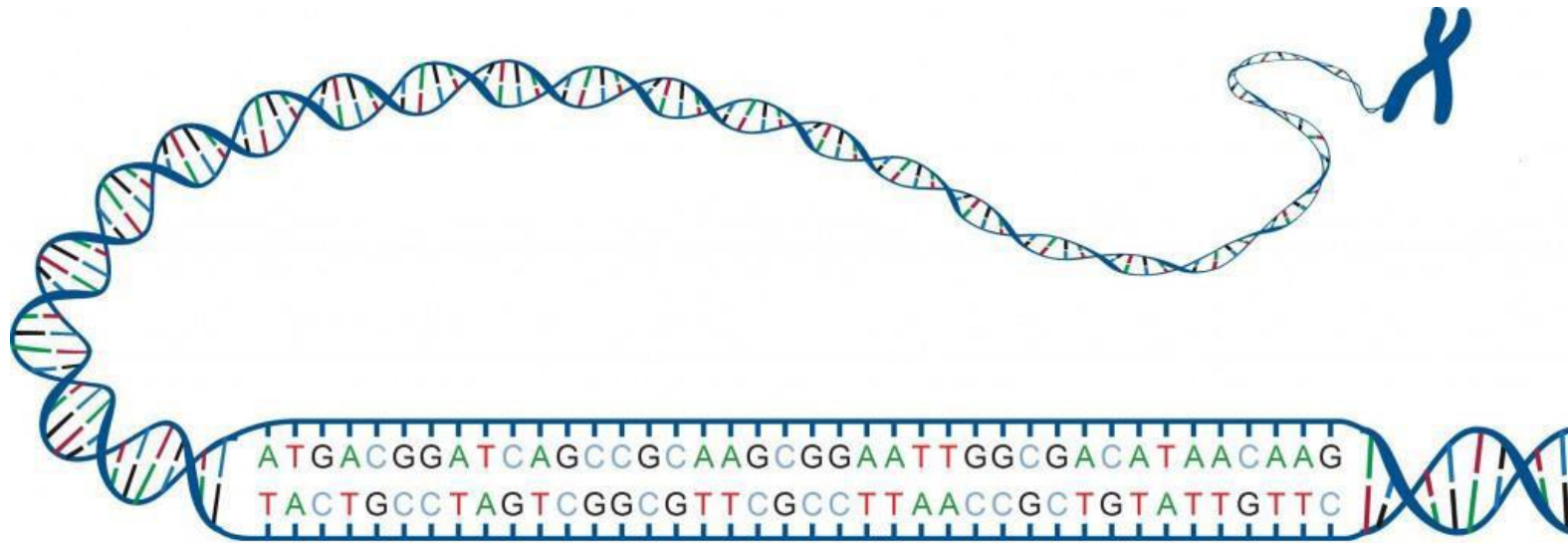
## Initiation à la recherche

### Introduction au projet

**Recherche de motifs dans les génomes:  
les sites fixation à l'ADN des facteurs de transcription**

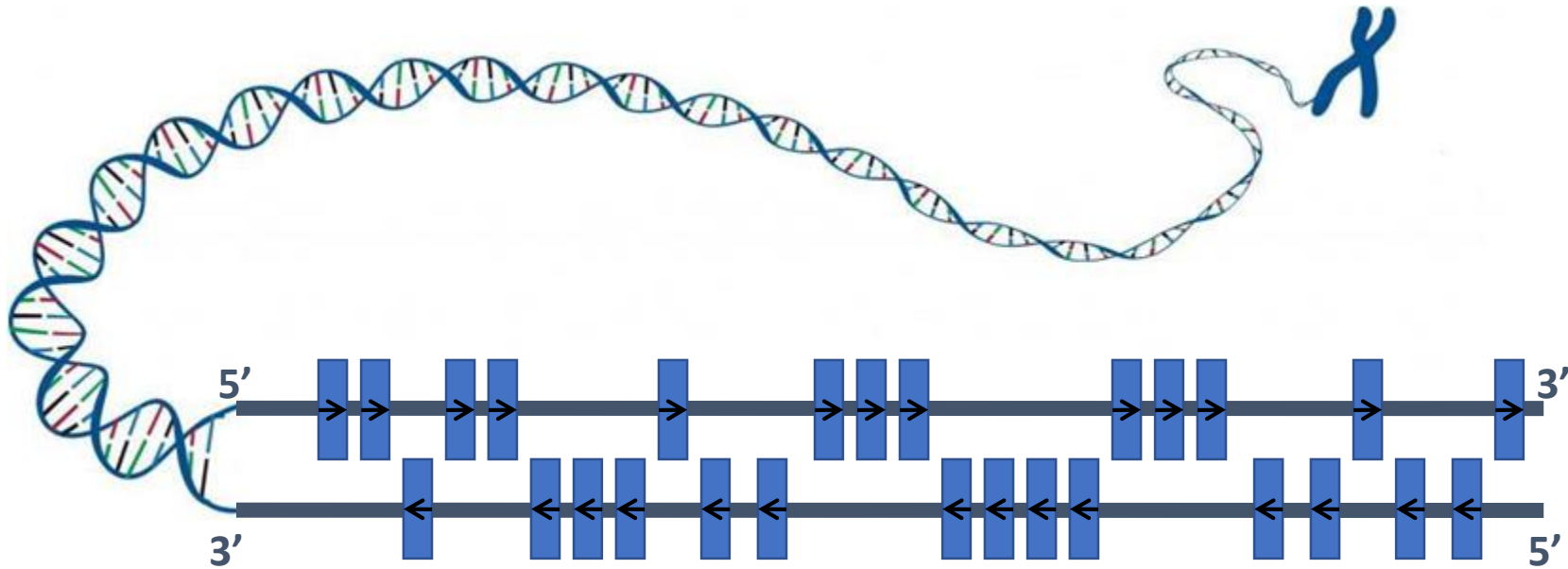
# Organisation et activation de l'information des génomes

**1 chromosome= 1 molécule d'ADN = 2 brins d'ADN avec des séquences complémentaires**

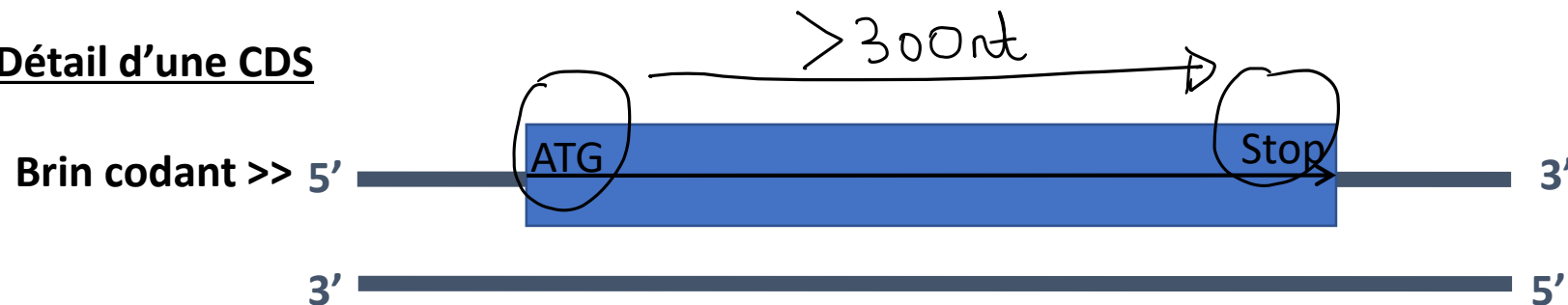


# Organisation et activation de l'information des génomes

**1 chromosome= 1 molécule d'ADN = 2 brins d'ADN avec des séquences complémentaires  
=> des centaines/milliers de séquences codant des protéines (CDS)**

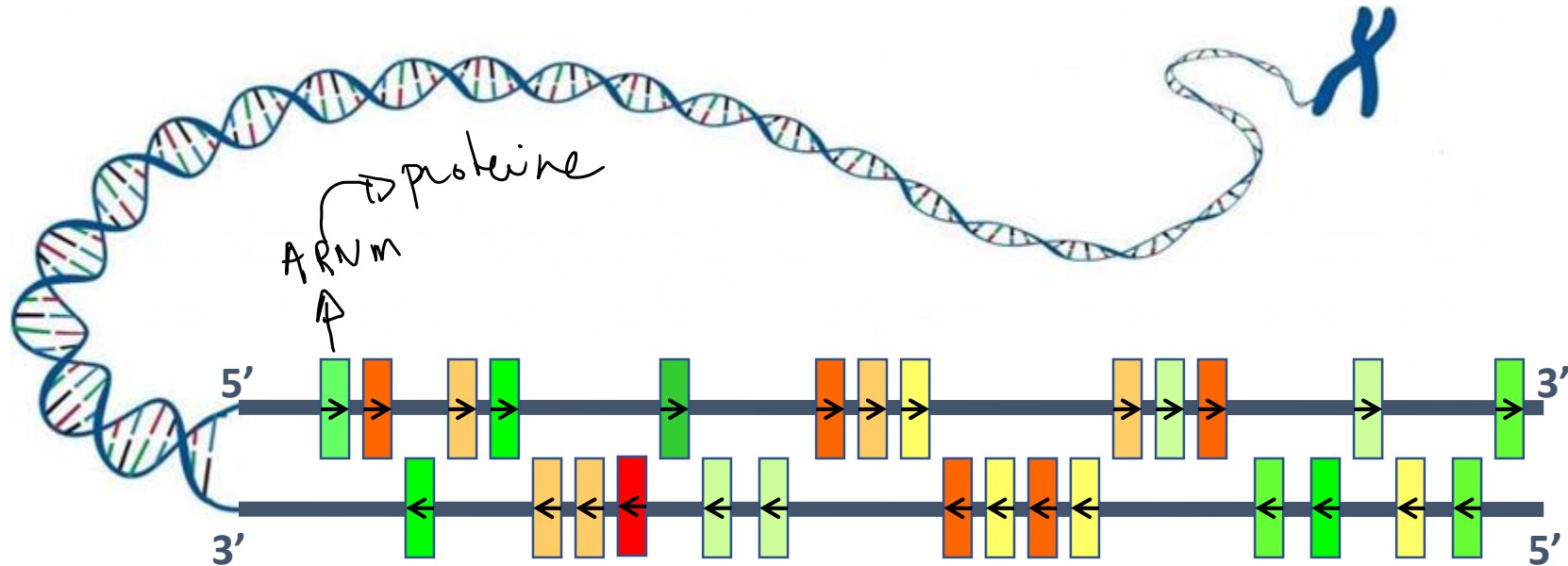


## Détail d'une CDS

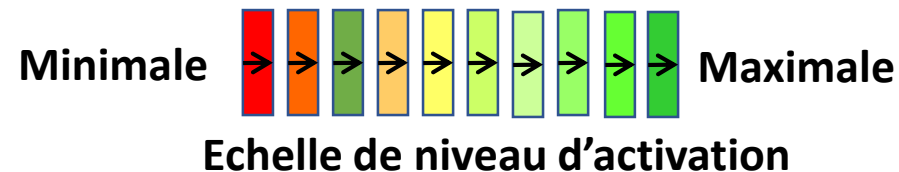


# Organisation et activation de l'information des génomes

**1 chromosome = 1 molécule d'ADN = 2 brins d'ADN avec des séquences complémentaires**  
**⇒ des centaines/milliers de séquences codant des protéines (CDS)**

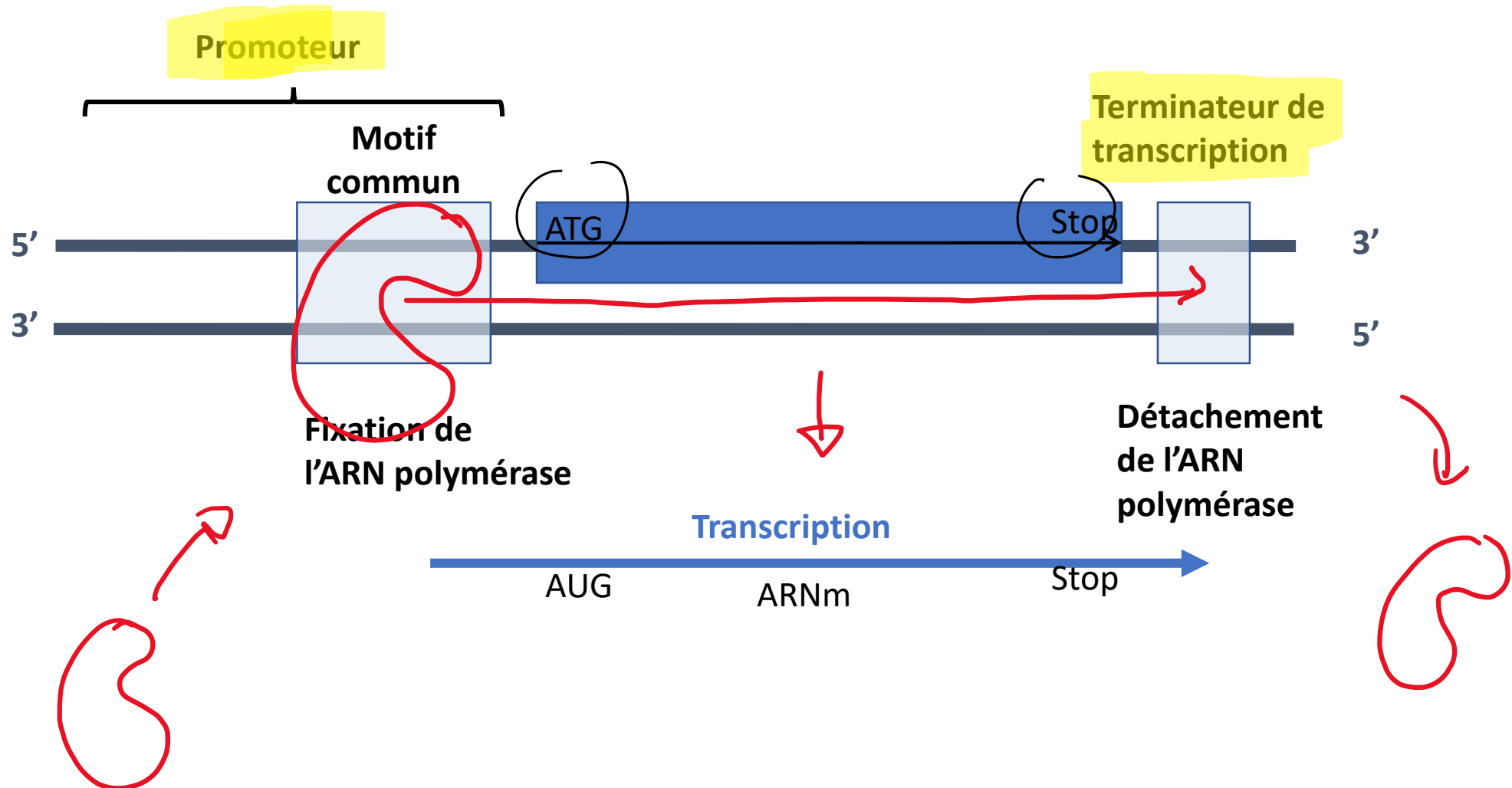


**⇒ Une activation variable des CDS en fonction des conditions cellulaires et environnementales**



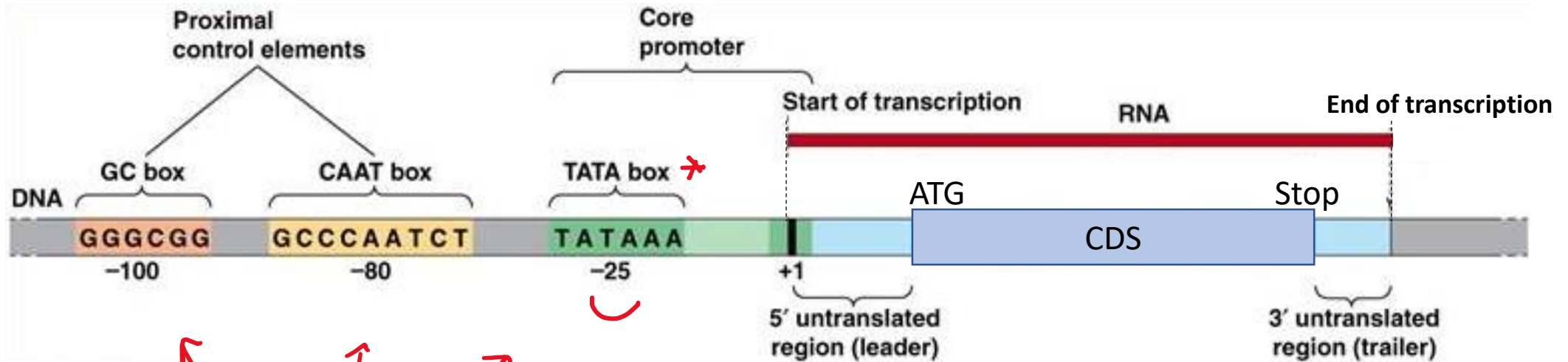
# Le promoteur des CDS: le site d'activation de la transcription des CDS

## Détail d'une CDS avec son promoteur



## Détail de la structure globale de la régions promotrice d'un gène eucaryote

Promoteur



3 motifs

Reconnaissance efficace de l'ARN polymérase

## Exemple de recherche de motifs:

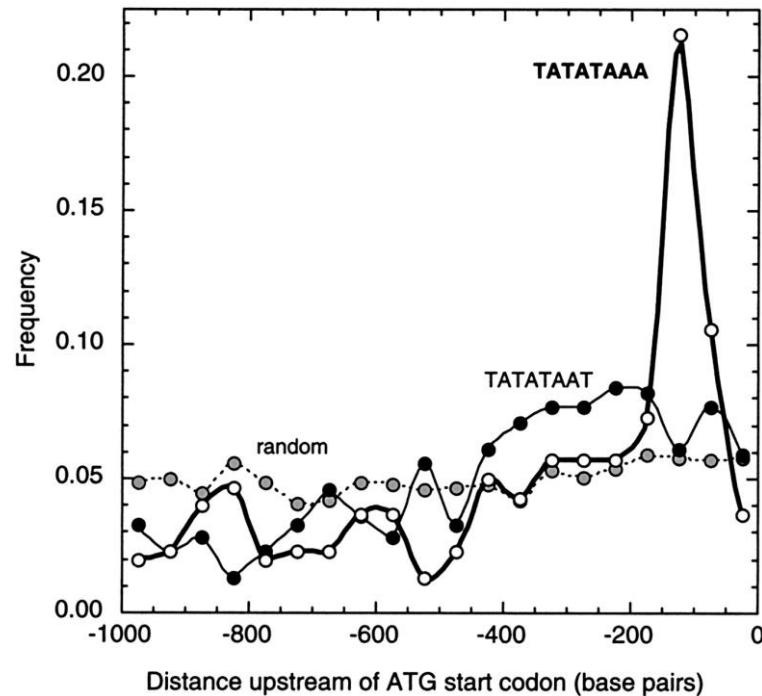
### Recherche de la TATA box dans le génome de la levure *S. cerevisiae*

→ motif le + fréquent

**Consensus du motif TATA box :** TATA(A/T)A(A/T)(A/G) → IUPAC nucleotide code: TATAWAWR

**Position de la TATA box:**

20 à 40 bases avant le site initiation de la transcription, ie 50 à 200 base avant l'ATG initiateur

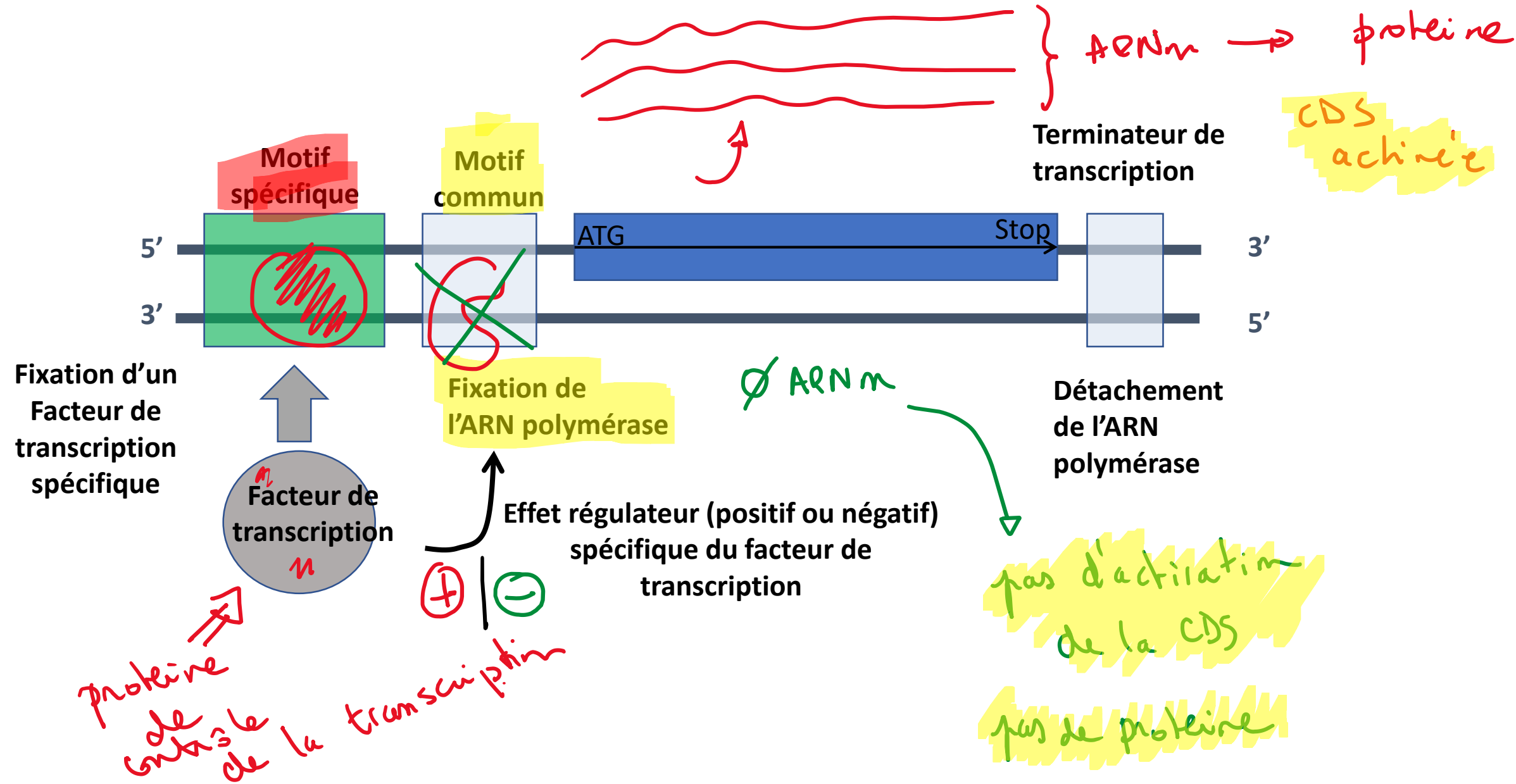


→ Utilisation d'un outil en ligne de recherche de motif:

[http://rsat.sb-roscoff.fr/genome-scale-dna-pattern\\_form.cgi](http://rsat.sb-roscoff.fr/genome-scale-dna-pattern_form.cgi)

1641 occurrence sur 6604 CDS: ~ 25% des CDS

Le promoteur des CDS:  
une région régulatrice de l'activation génique





## Exemple de motifs régulateur: Le motif de recrutement du facteur de transcription Gal4

Details for **GAL4** from [Macisaac et. al](#)

TF

Logo

GAL4

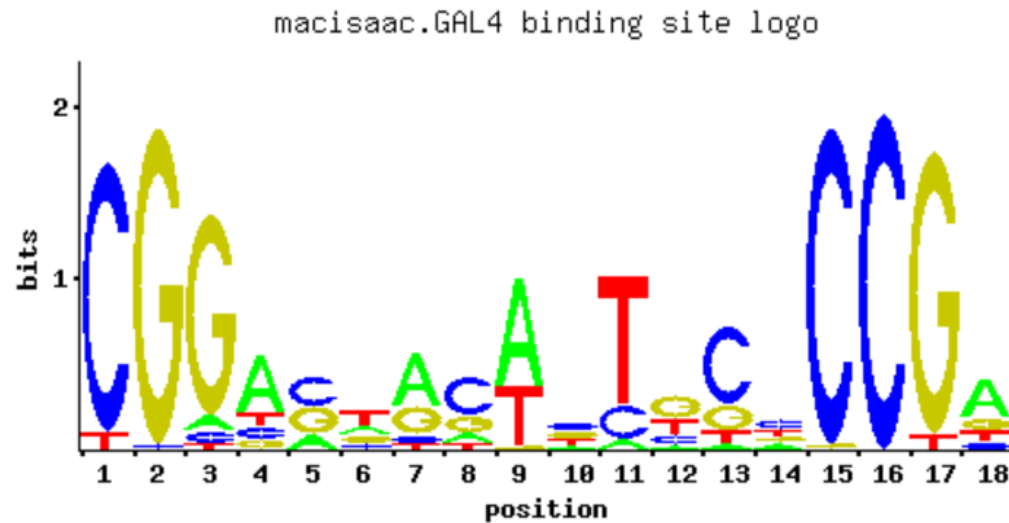
Matrices:

PFM

PCM

PWM

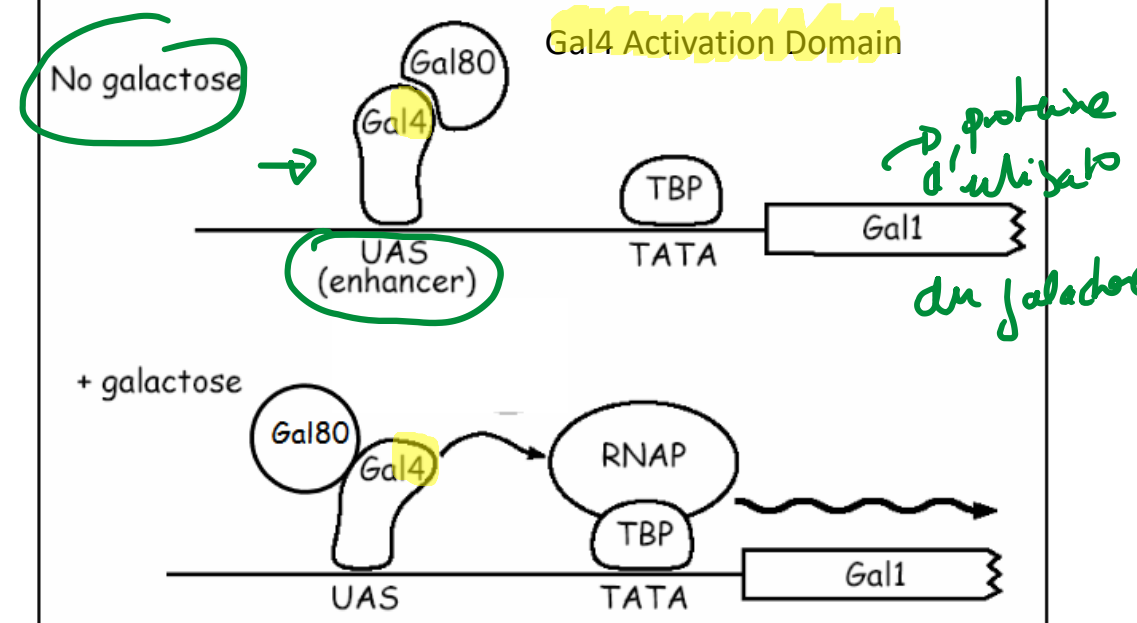
Data Source:  
[Macisaac et. al](#)



Position Count Matrix for Macisaac et. al GAL4:

A		0	0	5	64	18	19	57	14	64	11	5	6	3	18	0	0	0	57
C		94	1	4	12	42	14	7	56	0	36	19	18	67	42	98	99	0	9
G		0	98	88	10	37	18	29	21	1	31	0	47	21	18	1	0	95	17
T		5	0	1	13	1	47	5	7	34	20	74	28	9	20	0	0	4	15

### Activation of GAL transcription



→ Utilisation d'un outil en ligne de recherche de motif:

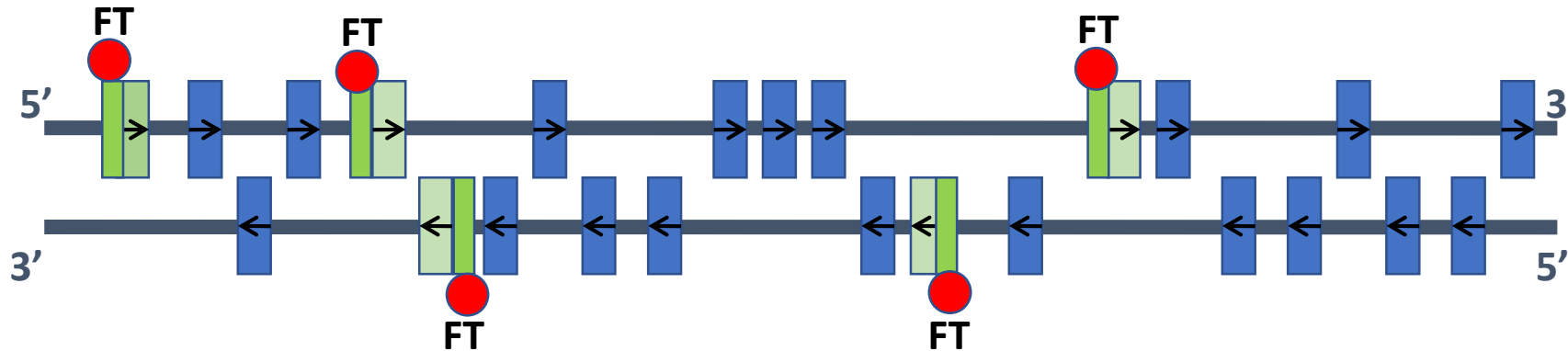
<http://rsat.sb-roscoff.fr/genome-scale-dna-pattern.cgi>

Motif code IUPAC: CGGNNNNNNNNNNNNCCG

189 occurrence sur 6604 CDS (2%) dont 13 dans des gènes du métabolisme du galactose

## Répartition des motifs régulateurs dans les génome et co-régulation des gènes impliqués dans un même processus

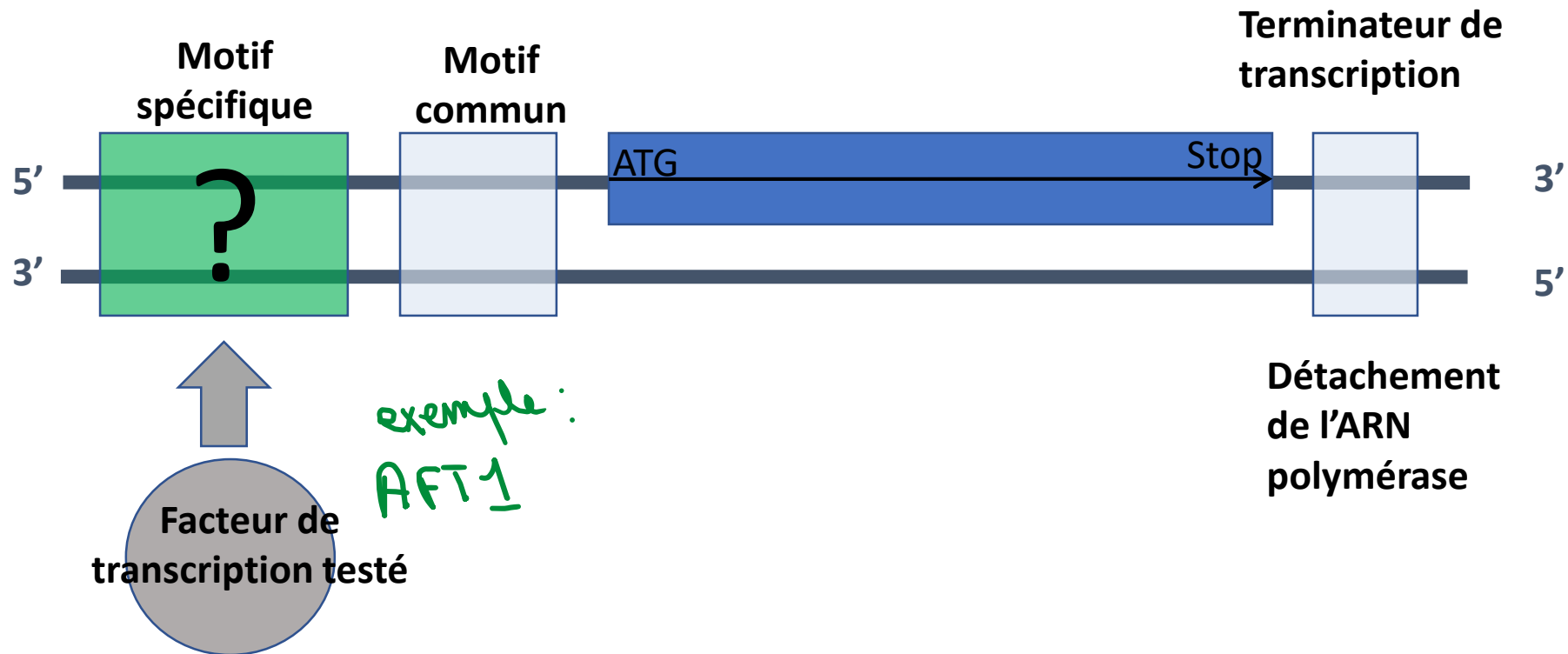
FT: Facteur de transcription



Projet: comment découvrir les motifs régulateurs de FT dans un génome?

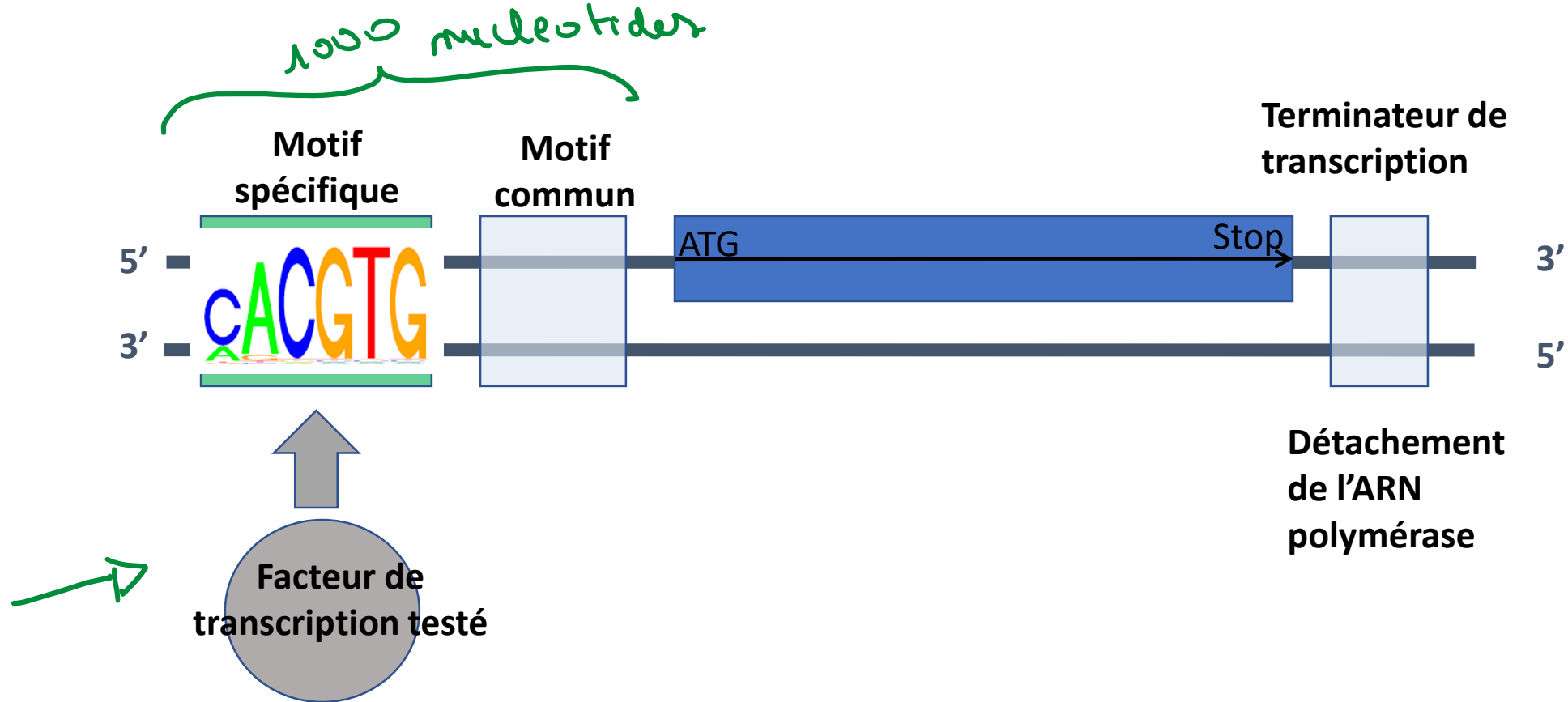
Objectif du projet:  
Identifier le motif fixé par un facteur de transcription d'intérêt

Détail d'une CDS avec son promoteur

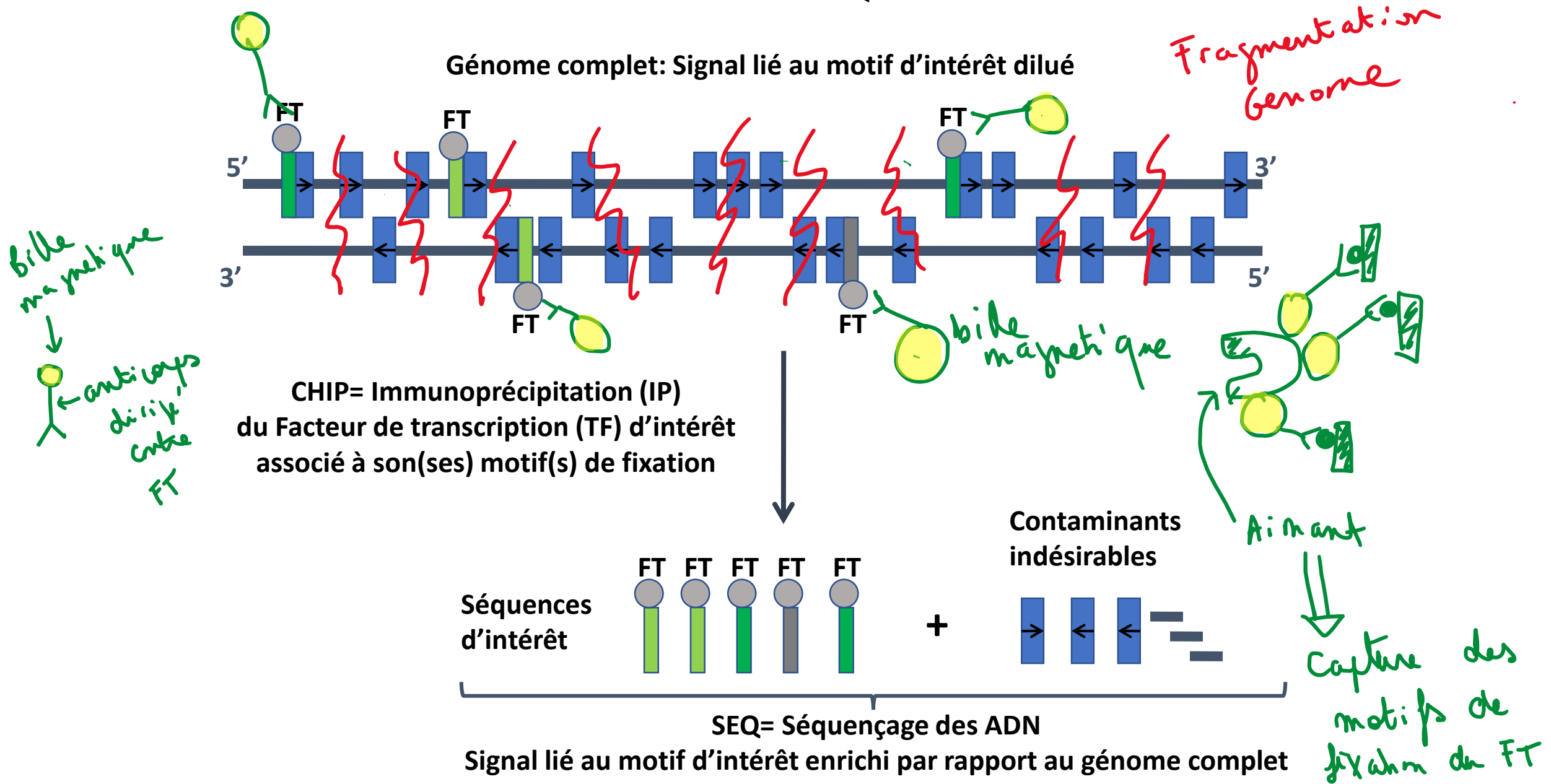


Objectif du projet:  
Identifier le motif fixé par un facteur de transcription d'intérêt

Détail d'une CDS avec son promoteur



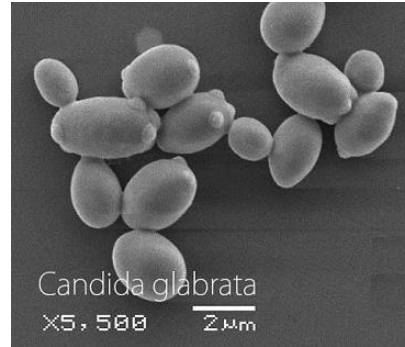
# Méthode d'identification des motifs des facteurs de transcription: le CHIP-SEQ



# Les données expérimentales du projet LU3I019



Organisme étudié: la levure *Candida glabrata*

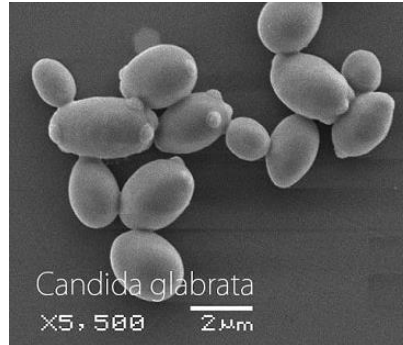


Feature Type	Chromosome													
	ChrA	ChrB	ChrC	ChrD	ChrE	ChrF	ChrG	ChrH	ChrI	ChrJ	ChrK	ChrL	ChrM	Nuclear genome
Total ORFs	205	220	235	288	282	387	440	466	470	523	564	582	621	5283
Verified ORFs	5	10	16	12	19	20	18	23	21	18	30	21	30	243
Uncharacterized ORFs	200	210	219	276	263	367	422	443	449	505	534	561	591	5040
tRNA	9	9	7	18	9	15	18	14	26	17	22	25	18	207
ncRNA	2	2	2	2	1	3	6	4	4	4	14	7	16	67
Pseudogenes	2	3	1	2	1	1	1	1	1	1	0	1	2	17
Centromere	1	1	1	1	1	1	1	0	0	1	1	1	1	11
rRNA	0	0	0	0	0	0	0	0	0	0	0	4	0	4
Long_terminal_repeat	0	0	0	1	0	0	3	0	0	0	0	0	0	4
Repeat_region	0	0	0	1	0	0	1	0	0	0	0	0	0	2
Total	219	235	246	313	294	407	470	485	501	546	601	620	658	5595
Chromosome length (bp)	491,328	502,101	558,804	651,701	687,738	927,101	992,211	1,050,361	1,100,349	1,195,129	1,302,831	1,455,689	1,402,898	12,318,241

Nb de CDS

# Les données expérimentales du projet LU3I019

Organisme étudié: la levure *Candida glabrata*



## Fichier génomique

[C\\_glabrata\\_1000bp\\_only.fasta](#)

Pour chaque CDS, les 1000 nucléotides situés en amont (contenant le promoteur du gène)

```
> CDS1 - - - - - 1000 nt  
> CDS2 - - - - - 1000 nt  
:  
> CDS500 - - - - - 1000 nt
```

## Expériences de CHIP-SEQ pour 8 facteurs de transcription

Aft1 (1 expérience)  
Amt1 (1 expérience)  
Hap4 (2 expériences)  
Hap5 (1 expérience)  
Mac1 (2 expériences)  
Pdr1 (2 expériences)  
Sef1 (1 expérience)  
Zap1 (2 expériences)



## 12 Fichiers CHIP-SEQ

[Sequence by Peaks X.fasta](#)

Séquences enrichies (Peaks) après l'expérience de CHIP-SEQ

```
> 1 21  
> 2 21  
> 3  
> 4
```

Resistante aux stress

# Facteur de transcription contrôlant positivement l'ARN polymérase

