



**HAL**  
open science

# Hétérogénéité phénotypique dans les populations d'origine clonale: Origine et conséquences

Guillaume Corre

► **To cite this version:**

Guillaume Corre. Hétérogénéité phénotypique dans les populations d'origine clonale: Origine et conséquences. Biologie cellulaire. Ecole Pratique des Hautes Etudes, 2012. Français. NNT: . tel-01116789

**HAL Id: tel-01116789**

**<https://hal.science/tel-01116789v1>**

Submitted on 14 Feb 2015

**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.

---

*École Pratique des Hautes Études*

*École Doctorale ED472*

*Mention Systèmes Intégrés, Environnement et Biodiversité*

---

## **Thèse de doctorat**

Discipline

*Biologie cellulaire, moléculaire et sciences de la santé*

---

# ***Hétérogénéité phénotypique dans les populations d'origine clonale : Origine et conséquences***

---

Présentée par

**Guillaume CORRE**

Soutenue le 27 juin 2012 devant le jury composé de :

<b>Monsieur de Professeur Bernard Mignotte</b>	<b>Président</b>
<b>Monsieur le Docteur Hugues Berry</b>	<b>Rapporteur</b>
<b>Monsieur le Docteur Bernard Vandebunder</b>	<b>Rapporteur</b>
<b>Monsieur le Docteur Olivier Gandrillon</b>	<b>Examineur</b>
<b>Monsieur le Professeur Andras Paldi</b>	<b>Directeur de Thèse</b>

Laboratoire d'Immunologie Moléculaire et Biothérapies Innovantes

INSERM U951 - EPHE - UEVE - GENETHON

1 bis rue de l'Internationale

91000 EVRY

## Résumé

---

Le paradigme actuel de la biologie soutient que des cellules possédant le même programme génétique doivent se comporter de façon similaire lorsqu'elles évoluent dans le même environnement. L'origine clonale de cellules est alors une garantie de grande homogénéité pour de nombreuses applications. Néanmoins, des observations montrent que malgré leur identité génétique, certaines cellules arborent des caractères différents dans le même environnement. L'objectif principal de ce travail est d'étudier la dynamique des fluctuations qui conduit à la diversification de ces cellules.

Dans un premier temps, nous avons étudié la dynamique qui permet d'obtenir une population clonale hétérogène, à partir d'une cellule. En isolant des cellules et en analysant leurs descendants pendant plusieurs semaines, nous avons montré que ce processus est relativement long. Il semble subir un effet mémoire à l'échelle cellulaire qui ralentit la diversification phénotypique. Néanmoins, après plusieurs semaines, un retour à l'équilibre s'opère et la variabilité des populations obtenues atteint celle de la population d'origine.

Par la suite, nous avons quantifié les changements d'expression génique à l'échelle cellulaire. Cette seconde étape, basée sur la vidéo-microscopie couplée à l'analyse d'images et au traitement du signal, a permis d'estimer la dynamique qui aboutit à l'hétérogénéité observée dans la première partie. Nous avons aussi mis en évidence qu'une proximité spatiale entre deux constructions transgéniques pouvait entraîner une compétition entre-elles. Enfin, nous avons identifié une limite à l'utilisation de certains gènes rapporteurs dans les études dynamiques de l'expression génique.

**Mots-clés :** expression génique, stochasticité, population clonale, cellules eucaryotes, phénotype, variabilité.



## Abstract

---

According to the actual paradigm, isogenic cells in the same environment must display the same behavior as they share the same genetic program. In consequences, the clonal origin of cells is considered as the best guaranty of their homogeneity. Nevertheless, observing differences between isogenic cells, grown in the same environment is not a rare event as observed in different studies. The main objective of this work is to understand how those differences between genetically identical cells appear and to quantify the dynamics of these changes.

The first part is dedicated to the study of long term variations in cells' phenotype. Starting from single cells, we monitored the emergence of cell-to-cell variations in the resulting populations by weekly analysis during several weeks. We show that the establishment of cell-to-cell heterogeneity in our system is a long lasting process, suffering a memory effect which slowed the diversification of phenotype. Nevertheless, several weeks after sub-cloning individual cells, a relaxation process is observed in the resulting populations and their variability reached the same level as in the original population.

In the second part, we quantified rapid changes in gene expression at the level of single cells. Based on video-microscopy experiments coupled with images analysis and signal processing, we highlight different single-cell dynamics that could explain the emergence of cell-to-cell differences in isogenic cells populations. We also show that a competition effect appears between similar transgenic constructs when they are close to each other. Finally, we highlight the limitations of the use of long-life proteins for the study of gene expression dynamics.

**Keywords:** gene expression, stochasticity, clonal populations, eukaryotic cells, phenotype, variability.

## Remerciements

---

Je tiens tout d'abord à remercier les membres du jury de m'avoir fait l'honneur de juger mon travail.

J'exprime toute ma gratitude au Pr Andras Paldi qui m'a accueilli dans son équipe de recherche dès la fin de mon stage de master 2. Je le remercie d'avoir partagé sa vision éclairée et novatrice sur les questions biologiques qui nous ont préoccupées pendant ces trois années, ainsi que sur tous les autres sujets qui m'ont permis d'évoluer. Merci pour son aide au quotidien, pour la liberté d'entreprendre qu'il m'a confiée, sa motivation et son enthousiasme dans les moments plus difficiles. Je le remercie enfin pour son aide à la rédaction de ce manuscrit.

Je souhaite particulièrement remercier le Dr Thi My Ahn Neildez qui m'a encadré pendant mon stage de master 2 et qui m'a motivé à poursuivre le travail entamé en ce temps, en m'aidant à la recherche de financement pour cette thèse. Merci pour sa sympathie et sa bonne humeur au quotidien.

J'adresse également mes remerciements aux autres membres de l'équipe anciennement ICC et actuelle IMBI. Merci au Dr Anne Galy pour son soutien et à Maguy Sylvestre pour son aide aux questions administratives.

Je remercie toutes les personnes ayant participé de près ou de loin à ce projet parmi lesquelles Peggy, Perrine et Sophie. Un remerciement particulier au Dr Daniel Stockholm pour ses compétences en imagerie, en analyse d'images et en programmation, ainsi que pour son enthousiasme contagieux et son esprit passionné.

Je remercie les « stocka-boys » de Lyon et Paris dont Olivier Gandrillon, José Vinuelas, Gaël Kaneko, Jean-Jacques Kupiec et l'équipe modélisation de Guillaume Beslon, pour leur

aide matérielle et technique, ainsi que pour les discussions sur la thématique et les résultats.

Merci aux différents services de Généthon qui m'ont aidé pendant cette période, le service imagerie/cytométrie en premier lieu ainsi que les services séquençage, histologie et analyses moléculaires.

Je souhaite remercier les locataires successifs du 3N7-27 qui m'ont forcé à parler un peu plus : Céline, Justine, Jérémie n°2, Wissem, Catherine, Aurore et Jérôme. Merci de m'avoir supporté moi, mes silences interminables, ma musique et mes sifflotements ! Merci à Yoshiaki pour son aide technique et nos discussions en anglais qui ont élevé mon niveau oral.

Je pense aussi évidemment à ceux que j'ai eu l'occasion de côtoyer au quotidien : Fédor, David, Smahane, Julie, Pierre, Marianne, Jérémie n°1, Robin et Thibaut ainsi que les autres cruciverbistes amateurs de café du midi.

Je ne saurai oublier ma famille, ma belle famille et mes amis qui m'ont soutenu tout au long de cette expérience, malgré la distance. Merci aux amis du pays (Mc Ouille, Antho, Noémie, Rose, Natcha et cie, bibi, le gâteau breton, le kig ah farz, le pâté Hénaff et les traou mad) et aux locaux (Véro, Méli, Soso, Cyril et Clément).

Je remercie ma sœur pour l'entretien et l'approvisionnement de l'appartement pendant ses (très) fréquentes escales ainsi que les autres visiteurs qui ont peuplé les lieux, même brièvement. Je remercie plus encore celle qui me soutient et me supporte depuis presque 5 ans et sans qui la vie serait bien plus monotone : coin coin Gaëlle !

Je remercie Généthon et l'AFM pour l'accueil dans des conditions de travail privilégiées, ainsi que le Fond AXA pour la recherche qui a financé ce travail.

Que chacun trouve dans ce travail l'expression de ma sincère gratitude (notamment ceux que j'ai oublié de citer ☺).

# Table des matières

---

<b>RESUME</b> .....	<b>III</b>
<b>ABSTRACT</b> .....	<b>IV</b>
<b>REMERCIEMENTS</b> .....	<b>V</b>
<b>TABLE DES MATIERES</b> .....	<b>VII</b>
<b>LISTE DES ABREVIATIONS</b> .....	<b>IX</b>
<b>LISTE DES FIGURES</b> .....	<b>XI</b>
<b>LISTE DES TABLEAUX</b> .....	<b>XV</b>
<b>LISTE DES ANNEXES</b> .....	<b>XV</b>
<b>PREAMBULE</b> .....	<b>1</b>
<b>INTRODUCTION</b> .....	<b>1</b>
I.  VARIABILITE NON-GENETIQUE DU PHENOTYPE .....	2
I.1. <i>Du génotype au phénotype</i> .....	2
I.2. <i>L'origine clonale</i> .....	3
I.3. <i>Une préoccupation récente pour un phénomène « ancien »</i> .....	4
I.4. <i>La variabilité biologique : la plus constante des variables ?</i> .....	5
I.5. <i>La variabilité : un phénomène ignoré</i> .....	8
I.6. <i>De nouveaux outils pour la quantification de l'hétérogénéité</i> .....	11
II.  ORIGINE DE L'HETEROGENEITE INTER-CELLULAIRE .....	17
II.1. <i>L'expression génique est un processus stochastique</i> .....	17
II.2. <i>La régulation du bruit</i> .....	32
II.3. <i>L'organisation cellulaire et le bruit structural</i> .....	37
III.  CONSEQUENCES DE LA STOCHASTICITE DE L'EXPRESSION GENIQUE .....	47
III.1. <i>Lysogénie et cycle lytique chez le phage lambda</i> .....	48
III.2. <i>Virulence et latence chez le VIH-1</i> .....	49
III.3. <i>Entrée en compétence chez Bacillus subtilis</i> .....	50
III.4. <i>Différenciation cellulaire chez la drosophile</i> .....	51
III.5. <i>Différenciation lors de l'hématopoïèse</i> .....	52
<b>OBJECTIF</b> .....	<b>55</b>
<b>MATERIEL ET METHODES</b> .....	<b>59</b>
I.  MATERIEL BIOLOGIQUE .....	60
I.1. <i>Lignée cellulaire HER 911</i> .....	60
I.2. <i>Modèle bi-fluorescent eCFP/eYFP</i> .....	60
I.3. <i>Constructions plasmidiques et propriété spectrales protéiques</i> .....	60
I.4. <i>Oligonucléotides</i> .....	65
II.  METHODES GENERALES .....	67

II.1.	<i>Culture cellulaire</i> .....	67
II.2.	<i>Biologie moléculaire</i> .....	70
II.3.	<i>Techniques d'imagerie</i> .....	78
II.4.	<i>Cytométrie en flux et tri cellulaire</i> .....	81
II.5.	<i>Bio-informatique</i> .....	83
III.	PROTOCOLES SPECIFIQUES.....	95
III.1.	<i>Détermination de la durée de vie des protéines eCFP et eYFP</i> .....	95
III.2.	<i>Détermination de la durée de vie des ARNm</i> .....	95
<b>RESULTATS</b> .....		<b>97</b>
I.	ÉTUDE DES FLUCTUATIONS PHENOTYPIQUES A LONG TERME. ....	98
I.1.	<i>Présentation et caractérisation du modèle cellulaire</i> .....	98
I.2.	<i>Émergence de la variabilité phénotypique à long terme</i> .....	106
I.3.	<i>Analyse de la dynamique de l'expression génique</i> .....	121
I.4.	<i>Fluctuations de l'expression génique à l'échelle d'un cycle cellulaire</i> .....	128
I.5.	<i>Photo-blanchiment de cellules entières</i> .....	130
II.	AMELIORATION DU MODELE POUR L'ETUDE DES FLUCTUATIONS RAPIDES.....	136
II.1.	<i>Construction du modèle cellulaire avec protéines déstabilisées</i> .....	136
II.2.	<i>Caractérisation des clones indépendants et dépendants</i> .....	139
II.3.	<i>Analyse des fluctuations par imagerie en temps-réel</i> .....	147
<b>DISCUSSION ET PERSPECTIVES</b> .....		<b>159</b>
I.	CHOIX DU MODELE CELLULAIRE .....	160
II.	MACRO-HETEROGENEITE ET DYNAMIQUE CHROMATINIENNE .....	164
III.	ANALYSE DYNAMIQUE DE LA VARIABILITE .....	166
IV.	ÉTUDE DES CHANGEMENTS D'EXPRESSION RAPIDES .....	170
V.	DESTABILISATION DES PROTEINES ET FLUCTUATIONS A COURT TERME .....	174
VI.	DES OBSERVATIONS DIFFERENTES ENTRE LES DEUX MODELES .....	178
<b>CONCLUSION</b> .....		<b>180</b>
<b>REFERENCES BIBLIOGRAPHIQUES</b> .....		<b>181</b>
<b>ANNEXES</b> .....		<b>205</b>
<b>PUBLICATIONS</b> .....		<b>215</b>

## Liste des abréviations

---

<b>ADN(c)</b>	Acide Déoxyribo Nucléique (complémentaire)
<b>3'UTR</b>	3' Untranscribe Terminal Region
<b>ACF</b>	Auto Correlation Function
<b>ActD</b>	Actinomycine D
<b>ARE</b>	AU-Rich Element
<b>aTc</b>	anhydro Tetracyclin
<b>BLAST</b>	Basic Local Alignment Search Tool
<b>CaCl<sub>2</sub></b>	Chlorure de Calcium
<b>CCD</b>	Couple Charge Device
<b>CCF</b>	Cross Correlation Function
<b>CGG</b>	Commission de Génie Génétique
<b>ChIP</b>	Chromatin Immuno Precipitation
<b>CHX</b>	CycloHeximide
<b>CMF</b>	CytoMétrie en Flux
<b>CMV</b>	CytoMegaloVirus
<b>Cov</b>	Covariance
<b>CSV</b>	Comma Separated Values
<b>CV</b>	Coefficient de Variation
<b>DAPI</b>	4',6-diAmidino-2-PhenylIndole
<b>DEPC</b>	DiEthylPyroCarbonate
<b>DMEM</b>	Dubbelco Modified Eagle Medium
<b>DMSO</b>	DiMethylSulfOxyde
<b>DNMT</b>	DNA Methyl Transferase
<b>dNTP</b>	deoxyNucléotides TriPhosphate
<b>DRB</b>	5,6-Dichlorobenzimidazole 1-β-D-ribofuranoside
<b>EDTA</b>	Ethylene Diamine Tetraacetic acid
<b>ELISA</b>	Enzyme Linked Immuno Sorbent Assay
<b>ETGC</b>	Etablissement de Thérapie Génique et Cellulaire
<b>FACS</b>	Fluorescence Activated Cell Sorting
<b>FCS</b>	Flow Cytometry Standard
<b>FDA</b>	Food and Drug Administration
<b>FISH</b>	Fluorescent InSitu Hybridization
<b>FITC</b>	Fluorescein isoThioCyanate
<b>FLIP</b>	Fluorescence Loss in Photobleaching
<b>FRAP</b>	Fluorescence Recovering After Photobleaching
<b>FRET</b>	Fluorescence Resonance by Energy Transfer
<b>GAPDH</b>	GlycerAldehyde 3-Phosphate DeHydrogenase
<b>GFP, CFP,YFP,RFP</b>	Green, Cyan, Yellow, Red Fluorescent Protein
<b>HAT</b>	Histone Acetyl Transférase
<b>HBS</b>	HEPES Buffer Saline
<b>HDAC</b>	Histone DeAcétylase
<b>HEK</b>	Human Embryonic Kidney
<b>HER</b>	Human Embryonic Retinoblastome
<b>HIF</b>	Hypoxia Induced Factor

<b>HMT</b>	Histone Methyl Transferase
<b>IPTG</b>	IsoPropyl $\beta$ -D-1-ThioGalactopyranoside
<b>IRES</b>	Internal Ribosome Entry Site
<b>LB</b>	Luria Bertini
<b>LSM</b>	Laser Scanning Microscope
<b>LTR</b>	Long Terminal Repeat
<b>M</b>	Molar concentration
<b>mARN, siRNA, shRNA</b>	messenger, small interference, shot hairpin RiboNucleic Acide
<b>MBP</b>	Methyl Bindind Protein
<b>MDR</b>	Multi Drug Resistance
<b>MFI</b>	Mean or Median Fluorescence Intensity
<b>NCBI</b>	National Center for Biotechnology Information
<b>nm</b>	nano meter
<b>NV</b>	Normalized Variance
<b>OGM</b>	Organisme Génétiquement Modifié
<b>PBS</b>	Phosphate Buffer Saline
<b>PCR, qPCR, qRTPCR, qRT<sup>2</sup>PCR</b>	(Quantitative)(Real Time) (quantitative Real Time Reverser Transcription) Polymerase Chain Reaction
<b>PEST</b>	proline (P), glutamic acid (E), serine (S), and threonine (T) rich sequence
<b>pH</b>	potentiel Hydrogène
<b>PI</b>	Propidium Iodide
<b>PS</b>	Penicilline/Streptomycine
<b>rAdV</b>	recombinant AdenoVirus
<b>RBS</b>	Ribosome Binding Site
<b>ROI</b>	Region Of Interest
<b>SAM</b>	S-Adenosyl methionine
<b>SQL</b>	Structured Query Language
<b>SVF</b>	Sérum de Veau Foetal
<b>TAE</b>	TRIS Acetate EDTA
<b>TE</b>	TRIS EDTA
<b>TFIID</b>	Transcription Factor II
<b>TRITC</b>	Tetra methyl Rhodamine Iso Thio Cyanate
<b>TSA</b>	Tricostatine A
<b>UV</b>	Ultra Violets
<b>Var</b>	Variance
<b>VIH</b>	Virus de l'Immunodéficience Humaine
<b>X-GAL</b>	5-bromo-4-chloro-indolyl- $\beta$ -D-galactopyranoside

## Liste des figures

---

Figure 1 Relation entre génotype, environnement et phénotype.....	3
Figure 2 Expansion cellulaire clonale dans le système immunitaire.....	4
Figure 3 Bibliométrie sur la thématique de l'hétérogénéité inter-cellules.....	5
Figure 4 Effet d'homogénéisation des techniques d'analyse sur population.....	10
Figure 5 Expression génique graduelle ou binaire. ....	11
Figure 6 Constructions plasmidiques utilisées dans le système "repressilator". ....	19
Figure 7 Profil d'expression de la GFP dans le système « repressilator ». ....	20
Figure 8 Traduction et variabilité de l'expression génique chez <i>B.Subtilis</i> . ....	23
Figure 9 Transcription, traduction et variabilité de l'expression génique ( <i>S. cerevisiae</i> ). ..	24
Figure 10 Analyse de la dynamique d'expression par suivi de molécule unique.....	26
Figure 11 Quantification du nombre de transcrits cellulaires par FISH.....	26
Figure 12 Détection des transcrits in vivo par des sondes MS2.....	27
Figure 13 Dynamique de la transcription et de la traduction avec les sondes MS2. ....	28
Figure 14 Transcription en temps réel sur cellule isolée. ....	28
Figure 15 Chromosome bactérien utilisé pour le modèle bifluorescent. ....	30
Figure 16 Identification de la source de fluctuation de l'expression génique.....	31
Figure 17 Visualisation du bruit intrinsèque dans un modèle bactérien. ....	31
Figure 18 Exemples de réseaux géniques synthétiques. ....	33
Figure 19 Propagation du bruit dans un réseau génique.....	34
Figure 20 Conséquences de la longueur des cascades de régulation. ....	34
Figure 21 Longueur des cascades de régulation et délai de réponse.....	35
Figure 22 Auto-régulation négative et réduction du bruit. ....	36
Figure 23 Organisation cellulaire chez les eucaryotes et les procaryotes. ....	38
Figure 24 Organisation du génome eucaryote.....	39
Figure 25 Modifications épigénétiques au niveau des histones.....	40
Figure 26 Étude de la dynamique protéique par FRAP. ....	41
Figure 27 Coopérativité des modifications épigénétiques.....	43
Figure 28 Équilibre dynamique de la chromatine.....	45



Figure 29 L'expression génique est un processus multi-étape. ....	46
Figure 30 Hétérogénéité phénotypique et sélection clonale .....	48
Figure 31 Rétro-action négative, fluctuations et cycle infectieux du phage lambda. ....	49
Figure 32 Boucle de rétroaction positive chez le VIH. ....	50
Figure 33 Rétrocontrôle, initiation et maintenance de la compétence chez B.Subtilis. ....	51
Figure 34 Stochasticité de l'expression génique et devenir cellulaire chez la drosophile...	51
Figure 35 Variabilité phénotypique et différenciation cellulaire hématopoïétique.....	53
Figure 36 Rétroaction et différenciation des cellules souches hématopoïétiques. ....	53
Figure 37 Illustration de l'hétérogénéité phénotypique dans notre modèle. ....	55
Figure 38 Différentes dynamiques cellulaires peuvent expliquer l'hétérogénéité. ....	56
Figure 39 Carte des plasmides eCFP et eYFP sauvages. ....	61
Figure 40 Spectres d'absorption et d'émission des protéines eCFP et eYFP. ....	63
Figure 41 Plasmides codants eCFP et eYFP destabilisées. ....	64
Figure 42 Construction mixte eCFP/eYFP destabilisées et plasmide Transposase. ....	65
Figure 43 Régions amplifiées pour l'étude de la méthylation du promoteur CMV. ....	74
Figure 44 Estimation du site d'ouverture des plasmides. ....	76
Figure 45 Principe de la Splinkerette PCR. ....	76
Figure 46 Colorisation des images RGB. ....	78
Figure 47 Principe du photo-blanchiment de cellule entière. ....	80
Figure 48 Procédure d'analyse en cytométrie en flux. ....	82
Figure 49 Étapes de l'analyse d'images par CellProfiler.....	85
Figure 50 Différences d'illumination mesurées sur un champ vide. ....	86
Figure 51 Segmentation des cellules. ....	87
Figure 52 Montage d'un suivi automatique de cellules.....	88
Figure 53 Trois signaux artificiels et leur auto-corrélation respective.....	92
Figure 54 Exemple de calcul d'inter-corrélation. ....	94
Figure 55 Expression d'eCFP et eYFP dans les clones parentaux.....	98
Figure 56 Vérification de la présence des transgènes par PCR.....	101
Figure 57 Corrélations entre MFI et quantité d'ARNm .....	102
Figure 58 Cartographie des sites d'intégration dans le clone parental C3. ....	103
Figure 59 Expression des gènes autour du site d'intégration d'eCFP et d'eYFP.....	105

Figure 60 Analyse de la méthylation des îlots CpG. ....	106
Figure 61 Exemple de distribution de la fluorescence eYFP d'un clone parental. ....	107
Figure 62 Contrôle de l'expression des transgènes eCFP et eYFP dans 2 sous-clones. ....	108
Figure 63 Évolution de la fluorescence moyenne eYFP dans les sous-clones ....	111
Figure 64 Niveau moyen de fluorescence de chaque sous-clone (J45 à 55). ....	112
Figure 65 Évolution de la variance dans les 3 types de sous-clones. ....	113
Figure 66 Évolution du coefficient de variation dans les 3 types de sous-clones. ....	113
Figure 67 Relation entre moyenne de fluorescence et coefficient de variation. ....	114
Figure 68 Augmentation de la proportion de cellules positives. ....	115
Figure 69 Diminution de la proportion de cellules positives. ....	116
Figure 70 Évolution de la proportion de cellules eYFP positives après sous-clonage. ....	117
Figure 71 Estimation du moment d'apparition des premières cellules eYFP positives. ....	118
Figure 72 Évolution du pourcentage de cellules eCFP positives. ....	119
Figure 73 Analyse de l'expression des transgènes par vidéo-microscopie. ....	122
Figure 74 Montage d'images de vidéo-microscopie. ....	123
Figure 75 Dynamique de l'expression des transgènes 3 cellules et leur descendance. ....	124
Figure 76 Variabilité du cycle cellulaire et asymétrie de la division. ....	125
Figure 77 Observation de la diversification du phénotype. ....	126
Figure 78 Inhibition de l'expression des transgènes par la densité cellulaire. ....	127
Figure 79 Schéma de montage du dispositif expérimental. ....	128
Figure 80 Détails de la puce à cellules CYTOO®. ....	129
Figure 81 Évolution de l'expression des transgènes eCFP et eYFP. ....	129
Figure 82 Photo-blanchiment de cellules entières. ....	132
Figure 83 Retour de fluorescence après inhibition de la transcription ou traduction. ....	133
Figure 84 Retour de fluorescence sur cellules non traitées. ....	134
Figure 85 Retour de fluorescence de cellules traitées à la TSA. ....	135
Figure 86 Contrôle visuel de la transfection à J+1. ....	137
Figure 87 Contrôle de la transfection par cytométrie en flux. ....	137
Figure 88 Comparaison des constructions stables / déstabilisées. ....	140
Figure 89 Cinétique d'inhibition de la traduction. ....	141
Figure 90 Estimation de la demi-vie des protéines fluorescentes. ....	142

Figure 91 Estimation de la demi-vie des ARNm eCFP et eYFP. ....	143
Figure 92 Clones dépendants et indépendants exprimant les protéines déstabilisés. ....	144
Figure 93 Corrélation de l'expression des deux transgènes par catégorie de clones. ....	145
Figure 94 Analyse en cytométrie de deux clones exprimant les transgènes déstabilisés..	145
Figure 95 Niveau moyen d'expression de chaque transgène suivant la catégorie. ....	146
Figure 96 Corrélation de l'expression moyenne des deux transgènes. ....	146
Figure 97 Variabilité de l'expression des transgènes suivant la catégorie. ....	147
Figure 98 Vidéo-microscopie de cellules exprimant les protéines déstabilisées. ....	148
Figure 99 Exemples de dynamiques d'expression des transgènes. ....	150
Figure 100 Lissage des signaux bruts et normalisation. ....	152
Figure 101 Exemples d'inter-corrélation calculée sur des signaux variés. ....	153
Figure 102 Pourcentage de cellules ayant une expression non-corrélées des transgènes.	154
Figure 103 Exemples d'auto-corrélations sur différents type de signaux. ....	155
Figure 104 distributions des demi-vies d'auto-corrélation. ....	156
Figure 105 Corrélation des demi-vies d'auto-corrélation entre transgènes. ....	157

---

## Liste des tableaux

---

Tableau 1	Systèmes d'étude de l'expression génique à l'échelle unicellulaire.....	16
Tableau 2	Modifications épigénétiques et turnover des histones.....	44
Tableau 3	Liste des oligonucléotides.....	66
Tableau 4	Bibliothèques additionnelles utilisées sous R. ....	89
Tableau 5	Amorces de screening pour la présence des transgènes.....	100
Tableau 6	Quantification de l'expression des transgènes.....	102
Tableau 7	Bilan des sous-clones obtenus après tri cellulaire. ....	109
Tableau 8	Bilan de l'évolution du phénotype des sous-clones. ....	120
Tableau 9	Conditions expérimentales pour la mise au point du FRAP.....	131
Tableau 10	Nombre de clones "Dépendants" obtenus.....	138
Tableau 11	nombre de clones « indépendants » obtenus. ....	139
Tableau 12	Bilan des clones analysés par imagerie temps-réel.....	149

## Liste des annexes

---

Annexe 1	Analyse par cytométrie en flux des clones simple couleur eYFP.....	206
Annexe 2	Analyse par cytométrie en flux des clones simple couleur eCFP.....	207
Annexe 3	Analyse par cytométrie en flux des clones indépendants eCFP4 + YFP.....	208
Annexe 4	Analyse par cytométrie en flux des clones indépendants eCFP2A + eYFP.....	209
Annexe 5	Analyse par cytométrie en flux des clones dépendants eCFP/eYFP. ....	210
Annexe 6	Analyse par cytométrie en flux des clones dépendants eCFP/eYFP (suite).....	211
Annexe 7	Dynamique d'expression des transgènes eCFP et eYFP dans des cellules isolées d'un clone « Dépendant ».....	212
Annexe 8	Dynamique d'expression des transgènes eCFP et eYFP dans des cellules isolées d'un clone « Indépendant ».....	213



## Préambule

---

Il est communément accepté que le devenir d'une cellule et par là, la constitution d'un organisme, est déterminé par la séquence de son ADN. Cette réflexion déterministe constitue le paradigme actuel dominant de la biologie. Elle est intimement liée à l'essor de la biologie moléculaire qui a aboutie au dogme central ADN-ARN-Protéine (Crick, 1958, 1970). En conséquence, qu'il s'agisse de cellules procaryotes ou de cellules de mammifères supérieurs, l'origine clonale de cellules est aujourd'hui considérée comme la meilleure garantie de leur homogénéité et de leur stabilité phénotypique.

Malgré cette promesse, il n'est pas rare de voir apparaître des différences entre les cellules d'une culture *in vitro*, bien que la population cellulaire soit isogénique, et cultivée dans un environnement homogène. Le développement et la différenciation cellulaire sont le reflet par excellence de cette diversité non-génétique, les cellules de chaque organisme résultant de l'expansion d'une seule cellule originelle.

Cette observation soulève plusieurs questions, d'ordre fondamental mais aussi d'ordre pratique. Comment expliquer cette variabilité « non-génétique » et l'émergence de phénotypes cellulaires atypiques quand on pense avoir maîtrisé toutes les sources de variations externes ? Quelles conséquences cette variabilité a-t-elle en pratique, notamment dans les applications biotechnologiques basées sur des systèmes cellulaires ? Plus important encore, si cette variabilité existe *in-vitro*, est-elle présente *in vivo* ? Est-elle impliquée dans les processus biologiques ou est-ce délétère pour la cellule ou l'organisme ? Quel que soit le rôle de la variabilité, est-il possible de la canaliser ou de trouver des conditions l'exacerbant ?



## **Introduction**

---



## I. Variabilité non-génétique du phénotype

---

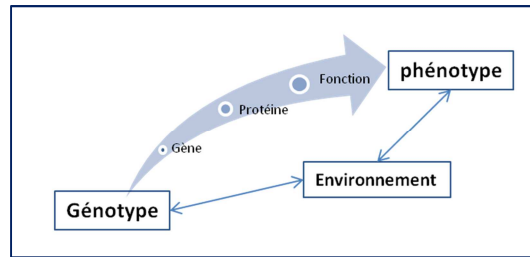
### I.1. Du génotype au phénotype

---

Il convient d'abord de définir ce que nous entendons par « phénotype ». Étymologiquement, ce terme provient des mots grecs « phenon » (paraître) et « typus » (empreinte). Il fait référence, dans sa définition la plus large, aux différents caractères observables d'un individu. Suivant l'échelle à laquelle on étudie cet individu, on définit le phénotype de façon plus restrictive en observant un sous-ensemble de caractères, voire uniquement le caractère d'intérêt. Au milieu du XIXe siècle, Mendel caractérisait les petits pois qu'il étudiait en utilisant leur couleur et leur rugosité.

Au niveau cellulaire, on définit le phénotype en fonction de critères morphologiques (taille, forme, ramifications, adhérence...) ou fonctionnels (contraction, sécrétion...) de la cellule. Au niveau moléculaire, on peut uniquement s'intéresser à l'expression d'une protéine par la cellule. Par exemple, le phénotype enzymologique de bactéries permet aisément d'identifier la souche, en fonction des diverses activités biochimiques identifiées. Il en est de même avec les marqueurs membranaires pour les cellules du système immunitaire (CD34+, CD4+, CD8+...) qui permettent de définir des sous-populations de lymphocyte par exemple.

Le phénotype correspond donc à ce que nous pouvons observer chez un individu, qu'il soit uni- ou multicellulaire. Il est classiquement décrit comme étant la résultante de l'expression de son génotype (un ou plusieurs gènes) à laquelle il faut ajouter les effets d'un environnement souvent complexe, qui peut être physique mais aussi chimique ou biochimique (Figure 1).



*Figure 1 Relation entre génotype, environnement et phénotype.*

En considérant cette définition simple, comment expliquer que des cellules génétiquement identiques puissent arborer un phénotype variable dans un environnement constant ?

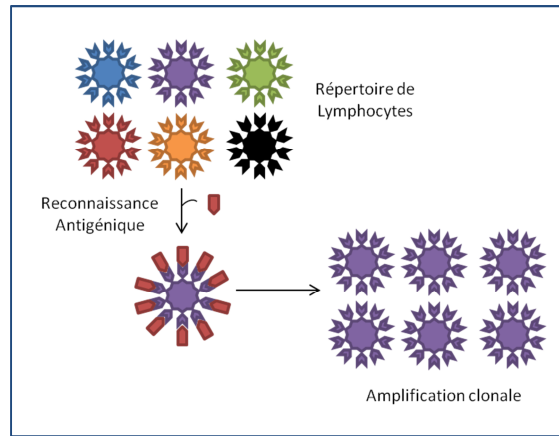
## I.2. L'origine clonale

---

Dans ce travail, nous nous sommes intéressés à la biologie de populations d'origine clonale afin de s'affranchir de différences génétiques. Un clone est alors défini comme l'ensemble des individus partageant le même patrimoine génétique. On parle alors d'individus isogéniques. Dans le langage commun, l'amalgame est souvent fait entre l'individu et l'ensemble des individus génétiquement identiques lorsque l'on parle de clone.

Suivant l'échelle à laquelle on se place, le clone peut donc être un organisme pluricellulaire (vrais jumeaux, plantes bouturées, etc.) ou des cellules isolées (colonie bactérienne ou cellules en culture).

Cette notion de clone est particulièrement usitée en immunologie, notamment dans le cadre de l'étude de l'expansion des lymphocytes (Figure 2). En réponse à la présentation antigénique par une cellule spécialisée, des lymphocytes naïfs vont se spécialiser pour cibler spécifiquement l'antigène. Ils vont alors se multiplier pour constituer un clone de cellules aptes à reconnaître et à détruire l'antigène. On parle d'expansion clonale.



**Figure 2 Expansion cellulaire clonale dans le système immunitaire.**

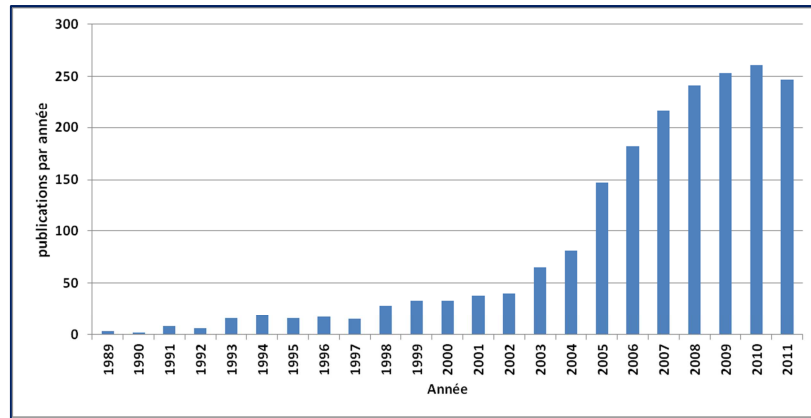
*En réponse à une stimulation antigénique (en rouge), une expansion clonale a lieu pour constituer un pool de cellules effectrices dirigées contre l'antigène.*

Dans le cadre de ce projet, nous parlerons d'un clone comme une population de cellules issue de l'amplification d'une cellule parentale par divisions successives, et qui ont été cultivées dans des conditions de prolifération adaptées. Nous discuterons par la suite de ce que l'on entend par « conditions adaptées ». En effet, il existe malgré tous les efforts des industriels, une différence non négligeable entre le contexte biologique in-vivo et les conditions de culture.

Le clonage prend plusieurs sens en biologie, du clonage moléculaire jusqu'au transfert de noyau chez les mammifères (Campbell et al., 1996). Nous parlerons ici de clonage pour définir l'isolement d'une cellule présentant un intérêt pour l'étude, et à son amplification par divisions successives. Le sous-clonage désignera le fait d'isoler à son tour une cellule particulière dans une population d'origine clonale. On parlera alors de sous-clone ou de sous-population clonale pour désigner cet individu et sa descendance.

### I.3. Une préoccupation récente pour un phénomène « ancien »

L'intérêt pour l'hétérogénéité intercellulaire est de plus en plus important comme en témoigne la littérature scientifique de ces dernières années (Figure 3). Une simple recherche par mots-clés montre l'évolution spectaculaire de cette thématique depuis les années 1990, dont une explosion à partir de 2002. Cette date marque la publication d'un des articles pionniers de la discipline par (Elowitz et al., 2002)



**Figure 3 Bibliométrie sur la thématique de l'hétérogénéité inter-cellules.**  
*Toutes revues confondues (Web of Knowledge, déc.2011).*

Plus qu'un effet de mode, il s'agit à l'heure actuelle de la redécouverte d'un phénomène longtemps ignoré pour des raisons que nous discuterons et qui semble aujourd'hui prendre une importance majeure en biologie. De façon intéressante, l'intérêt pour ce domaine est transdisciplinaire, attirant autant les physiciens, les mathématiciens modélisateurs que les biologistes. Il est donc nécessaire de franchir la barrière des disciplines, aussi difficile soit-il pour des biologistes de se plonger dans les équations. Nous verrons par la suite que cette interdisciplinarité est néanmoins essentielle pour aborder certains aspects de cette thématique. L'utilisation de modèles mathématiques ou les simulations informatiques sont des aspects mal maîtrisés par la plupart des biologistes de formation.

L'hétérogénéité phénotypique est observée dans de nombreux types cellulaires, de la bactérie aux cellules eucaryotes supérieures. Cultivées dans un environnement que nous qualifierons d'adapté et homogène, des cellules isogéniques peuvent se comporter différemment en réponse à un stimulus. Dans la partie à suivre, nous détaillerons quelques-unes des observations historiques ayant participé à l'essor actuel de cette thématique.

#### I.4. La variabilité biologique : la plus constante des variables ?

Le concept de variabilité inter-cellules émerge bien avant la biologie moléculaire des années 1980, principalement par la manipulation de micro-organismes bactériens tels

qu'*E. coli*, qui est alors l'organisme le plus étudié. A cette époque, la majeure partie des mises en évidence de l'hétérogénéité entre cellules est uniquement le fait d'observations.

Une des premières caractérisations revient à Max Delbrück, un biophysicien, qui observe une grande variabilité intercellulaire lorsqu'il quantifie le nombre de phages produits par des bactéries infectées (Delbrück, 1940). Il propose pour expliquer cette différence que dans chaque cellule, une incertitude dans les réactions biochimiques et les interactions moléculaires peut affecter la production virale. Il sous-entend aussi que des facteurs prédéterminants comme la taille des cellules peuvent entraîner cette variabilité, sans en faire la démonstration.

En 1944, Bigger et al. montrent que dans une population isogénique de staphylocoques traitée par un antibiotique, la pénicilline, une sous-population de bactéries survit au traitement. Il les appelle les « persistants ». La majorité des individus de la population en croissance meurt suite au traitement alors que d'autres survivent, dans une phase de croissance lente ou de dormance. Après arrêt du traitement, ces cellules persistantes se divisent à nouveau et reconstituent la population. Elles demeurent néanmoins sensibles à l'antibiotique, ce qui les rend différentes de mutants génétiques ayant acquis une résistance (Bigger, 1944). Ce phénomène de persistance a depuis été observé dans d'autres espèces de bactéries, notamment *E. coli* (Balaban et al., 2004).

Un comportement analogue existe chez la bactérie *Bacillus subtilis* où l'on observe un état dit de « compétence » dans certaines cellules d'une population isogénique (Nester and Stocker, 1963; Cahn and Fox, 1968; Hadden and Nester, 1968). La compétence est un état dans lequel la bactérie peut capter de l'ADN étranger présent dans son environnement, lui permettant d'acquérir des propriétés nouvelles dans des conditions de stress (résistance aux antibiotiques, activité enzymatique métabolique...). C'est un état temporaire et réversible, et de façon intéressante, seule une partie des cellules d'une colonie (environ 10%) peut devenir compétente à un instant donné. Ces deux exemples semblent indiquer que la variabilité non-génétique joue un rôle dans la biodiversité de ces organismes et pourrait constituer une stratégie de survie répandue.

En 1957, il a été montré que la production cellulaire de b-galactosidase dans des bactéries individuelles était hautement variable et aléatoire. Lorsque l'induction est de faible intensité, elle augmente la proportion de cellules exprimant l'enzyme plutôt que le taux d'expression de chaque cellule de façon équivalente (Novick and Weiner, 1957). Pourtant soumises au même stimulus, les cellules génétiquement identiques se comportent différemment. Les auteurs montrent aussi que chez ces bactéries, la décision de passer du métabolisme du glucose au métabolisme du lactose lorsque la dose de lactose est intermédiaire, pourrait être liée à des fluctuations aléatoires dans le niveau de « facteurs régulateurs ». Rappelons que la découverte du fonctionnement de l'opéron lactose date du début des années 1960 (Jacob and Monod, 1961). Le mécanisme sous-jacent a été explicité dans une étude récente (Choi et al., 2010).

Dans les années 1990, plusieurs études ont confirmé l'observation d'induction binaire des gènes, avec différents modèles. L'expression d'un gène rapporteur sous le contrôle d'un promoteur inductible est étudiée par Ko et al. Sous différentes doses d'inducteur (glucocorticoïde), ils notent que la variabilité d'expression du transgène entre cellules est très importante (Ko et al., 1990). Les mêmes résultats ont été observés par Ross et al. dans un modèle de macrophages murins exprimant la B-Galactosidase sous le contrôle du promoteur LTR de HIV1. Dans une population clonale, la fréquence de cellules exprimant le transgène est comprise entre 1/10000 et 1/1000 et peut atteindre 1/100 en présence de la protéine virale TAT, un trans-activateur du promoteur LTR (Ross et al., 1994).

En étudiant le chimiotactisme de bactéries individuelles stimulées par un attractant, on observe que le temps de réponse pour la modification de la nage varie énormément d'un individu à l'autre, malgré leur identité génétique (Spudich and Koshland, 1976). A cette époque déjà, l'hypothèse avancée pour expliquer cette différence est la distribution discrète du faible nombre de moteurs moléculaires présents par cellule. Une distribution de Poisson du nombre de ces molécules aurait alors pour conséquence une variabilité de cellule à cellule. Compte-tenu des moyens disponibles à cette époque, cette hypothèse n'a pu être confirmée. Les auteurs de cette étude parlent alors, probablement pour la première

fois, « d'individualité non-génétique » pour qualifier cette variabilité interindividuelle. Ils montrent aussi que le taux de changement de nage en réponse à la molécule attractive semble hérité. Il existe une sorte de « mémoire » du style de nage dans la descendance d'une cellule. Un déséquilibre dans la répartition des composants cellulaires lors de la division est alors proposé pour expliquer la variabilité, et non pas la taille des cellules ou la désynchronisation du cycle cellulaire.

Berg et Purcell étudient la précision avec laquelle des bactéries peuvent estimer la quantité d'attractants dans son environnement. Cette précision semble limitée par la diffusion du ligand à la surface des cellules impliquant une fluctuation de la quantité des récepteurs à ce ligand. L'environnement joue ici un rôle prépondérant dans la variabilité entre cellules (Berg and Purcell, 1977).

Les premiers travaux sur des cellules eucaryotes datent des années 1960 mais sont moins aboutis compte-tenu de la complexité de ces cellules par rapport aux bactéries. Peu d'eucaryotes supérieurs sont utilisés comme modèles, les recherches se focalisant principalement sur la levure. Néanmoins, une large variabilité de la taille des cellules, de leur indice de prolifération ou de leur mobilité est observée. Elle est mise en relation avec l'environnement des cellules, notamment leur densité. Malgré le grand nombre d'exemples biologiques où les fluctuations stochastiques sont observées, leur rôle biologique potentiel n'est pas pris en compte par la plupart des biologistes.

### I.5. La variabilité : un phénomène ignoré.

Pour de multiples raisons, la notion d'hétérogénéité ou de variabilité phénotypique a été laissée de côté. Nous pouvons néanmoins pointer deux d'entre-elles : en premier lieu, la conception que l'on se faisait du fonctionnement cellulaire à cette époque et, dans un second temps la méthodologie adoptée. Du fait de ces deux limitations, conceptuelle et technique, la variabilité est considérée aujourd'hui encore comme un phénomène marginal, voire comme un simple bruit de fond lié à l'expérimentation.

### *1.5.a. Un paradigme fort*

---

La vision déterministe du fonctionnement cellulaire, que beaucoup de biologistes ont encore à l'heure actuelle et telle qu'on l'enseigne dans les universités, est intimement liée à l'essor de la biologie moléculaire dans les années 1980. Les techniques révolutionnaires qui ont été développées ont permis d'analyser les gènes un à un, de décortiquer leur fonctionnement et leur régulation individuelle. Néanmoins, cette approche mécanistique a nécessité l'adoption d'un réductionnisme très fort afin d'expliquer le fonctionnement des organismes mais aussi l'origine de certaines maladies génétiques à partir de gènes individuels.

Ce réductionnisme s'est imposé dans le raisonnement collectif avec le célèbre dogme de la biologie moléculaire : ADN-ARN-Protéine. Dans ce modèle déterministe de la génétique, l'ADN joue le rôle central. Les réactions sont caractérisées par des relations linéaires unidirectionnelles et chaque cause a une conséquence bien déterminée. Ainsi, le phénotype est directement déterminé par le génotype (Figure 1). Dans ce cadre théorique, les cellules sont stables et nécessitent un facteur externe pour évoluer. Un facteur donné induit une réaction spécifique et reproductible dans les mêmes conditions. Il s'agit typiquement de la vision que nous avons de la différenciation cellulaire et ce concept d'induction est à la base de l'embryologie et de la biologie du développement.

Avec une telle vision mécanistique, il est intuitivement difficile de concevoir que des processus aléatoires, quelle que soit leur origine, puissent avoir un rôle dans des processus hautement déterministes, reproductibles et robustes. La cellule est une horloge de haute précision qui ne semble pouvoir tolérer la moindre incertitude.

### *1.5.b. Une approche méthodologique inadaptée*

---

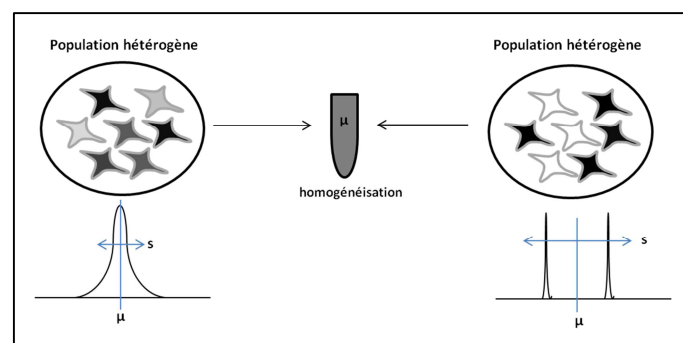
Comme nous l'avons évoqué précédemment, le paradigme de la biologie cellulaire et de la biologie du développement est en quelque sorte l'héritage d'un mode de pensée adopté par



la biologie moléculaire. D'un point de vue méthodologique, la plupart des expériences est réalisée sur des échantillons contenant un grand nombre d'individus, par commodité, mais aussi pour répondre aux limites de sensibilité des techniques.

Par exemple, les techniques d'extraction d'ADN, d'ARN ou de protéines, possèdent un rendement limité qui impose de travailler avec quelques milliers de cellules pour certaines d'entre-elles. Aujourd'hui encore, cette limite existe pour certains kits commerciaux. Or, cette étape est essentielle pour obtenir le matériel biologique purifié qui sera analysé par la suite. Malgré la sensibilité de certaines techniques, cette étape d'extraction/purification est une limite à l'analyse des individus isolés. La PCR, technique de biologie moléculaire de référence qui a révolutionné notre façon de manipuler l'ADN, en est l'exemple type. Malgré sa haute sensibilité, il est nécessaire de travailler avec une quantité d'ADN de quelques nanogrammes. Rappelons que le contenu en ADN d'une cellule ne dépasse pas 6pg et qu'il est donc nécessaire d'accumuler quelques milliers de cellules pour atteindre cette exigence.

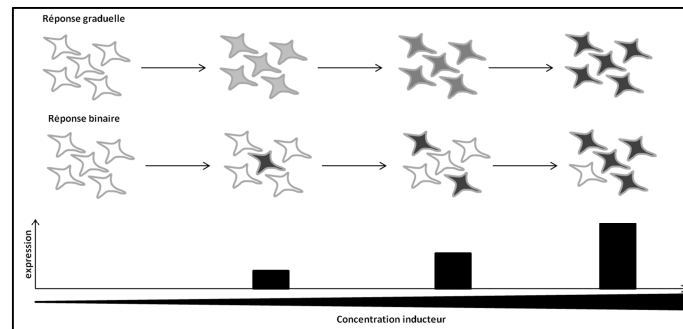
Ce genre de techniques (PCR et dérivés, Western Blot, ELISA...) est souvent suffisant pour montrer qu'un traitement a un effet global se traduisant à l'échelle de la population. Néanmoins, leur inconvénient est que le résultat obtenu est une moyenne du comportement de la population de cellules pour la variable d'intérêt (Figure 4). Aucune information sur une sous-population ou des individus atypiques ne pourra être obtenue (Altschuler and Wu, 2010).



**Figure 4 Effet d'homogénéisation des techniques d'analyse sur population.**

*L'information sur la variabilité dans la population est perdue si les individus sont mélangés. A gauche, la moyenne reflète globalement la distribution du marqueur dans la population, ce qui n'est pas le cas dans l'exemple de droite. ( $\mu$  = moyenne,  $s$  = variance).*

De façon automatique et puisque les cellules sont isogéniques, cultivées dans un environnement homogène, on transpose à chaque individu un effet global mesuré sur cette population. On peut citer l'exemple de l'induction d'un gène en réponse à un gradient d'un signal inducteur. En quantifiant le produit du gène, dans des populations traitées par différentes doses d'inducteurs, on observe une réponse dose dépendante. Plus la concentration de l'inducteur est importante plus le gène est exprimé. Logiquement, un modèle de réponse graduelle est adopté pour caractériser la réponse de chaque cellule. Or, une autre explication aussi convaincante peut être avancée : les cellules ne répondent pas de façon graduelle mais de façon binaire/stochastique (Figure 5). Chaque cellule possède une certaine probabilité de répondre, selon un modèle tout-ou-rien en imaginant un seuil de réponse.



**Figure 5 Expression génique graduelle ou binaire.**

*Sous différentes doses d'inducteur, un même niveau d'expression observé peut refléter différentes dynamiques des individus de la population. Dans le premier cas, la moyenne observée correspond à celle de chaque cellule. Dans le second cas, la proportion de cellules positives augmente mais la moyenne ne reflète pas le comportement individuel.*

L'utilisation massive de ce genre de techniques a masqué pendant longtemps des phénomènes que nous découvrons aujourd'hui. L'émergence de nouvelles techniques basées sur l'analyse de cellules uniques, mais aussi l'amélioration de la sensibilité des techniques mentionnées précédemment, ont abouti à un changement d'échelle d'analyse.

## I.6. De nouveaux outils pour la quantification de l'hétérogénéité

Comme nous l'avons vu précédemment, les analyses sur population entière ne permettent pas de surligner la variabilité entre les individus et réduisent leur comportement à celui de la population entière. Il est donc essentiel de travailler à l'échelle individuelle pour mettre

en évidence des comportements cellulaires uniques. Ces dernières années ont vu l'amélioration considérable de l'instrumentation en terme de sensibilité, de rapidité et de robustesse mais aussi la commercialisation d'outils adaptés à la détection et à la quantification de processus biologiques dynamiques et de faible intensité. Ces outils ont permis de mieux détecter la variabilité entre cellules mais aussi de passer de son observation à sa quantification (Arriaga, 2009).

### *1.6.a. L'instrumentation au service de la variabilité*

---

Il est nécessaire pour étudier la variabilité d'avoir accès à tous les individus et donc d'utiliser des techniques permettant une analyse à l'échelle individuelle. Parmi les techniques d'analyse sur cellule unique, l'imagerie et la cytométrie en flux sont les plus utilisées.

#### *◆ Microscopie*

---

La microscopie est la technique la plus à même d'offrir une résolution à l'échelle de la cellule unique. Elle permet de visualiser directement le phénomène d'intérêt pour autant que l'on dispose d'un bon moyen de le révéler (protéine fluorescente, sonde nucléique,...). Bien qu'ancienne, cette technologie a profité de nombreuses avancées et la plupart des microscopes sont aujourd'hui équipés de sources de fluorescence qui permettent de détecter une grande diversité de fluorophores. Ces molécules fluorescentes conjuguées à des anticorps, à des acides nucléiques ou à des protéines, permettent de détecter une infinité de structures cellulaires ou tissulaires, de protéines spécifiques ou d'activités biochimiques.

Elle permet bien sûr de contrôler visuellement les échantillons (observation statique) mais elle autorise aussi la réalisation d'expériences en temps-réel grâce à l'intégration du système d'imagerie dans un environnement contrôlé. Cette amélioration permet donc d'analyser une cellule et sa descendance au cours du temps, de suivre des interactions cellule/cellule ou cellule/environnement et de quantifier de façon dynamique un

phénomène d'intérêt. L'exemple type est le suivi de la mobilité intracellulaire d'une protéine fluorescente par la technique de FRAP. L'inconvénient majeur est la complexité expérimentale de ce type d'expériences et le débit d'analyse qui est limité à quelques cellules.

#### ◆ *Cytométrie en flux*

---

Technique plus récente, la cytométrie en flux (CMF) est le plus souvent utilisée pour analyser à grande vitesse des milliers de cellules. On l'utilise beaucoup en immunologie/hématologie pour identifier des sous-populations cellulaires parfois rares dans des échantillons complexes. De même, c'est la technique de référence pour étudier le cycle cellulaire. Le principe de la CMF réside dans l'association d'un système fluidique permettant de faire circuler les cellules de l'échantillon et de les focaliser avant leur passage dans une chambre optique. Les cellules passent alors une à une devant une série de lasers de longueurs d'ondes différentes afin d'exciter les fluorophores. Ceux-ci sont souvent couplés à des anticorps ciblant une protéine membranaire d'intérêt, mais on peut bien sûr détecter directement des protéines fluorescentes intra-cellulaires ou membranaires. Une partie électronique, composée de photomultiplicateurs, d'amplificateurs et de convertisseurs analogiques, permet de collecter et de traiter les signaux. Certains appareils ont la capacité d'isoler physiquement des cellules d'intérêt en fonction d'un critère de morphologie ou de fluorescence (FACS). Ils permettent le clonage d'individus à partir d'une population mixte.

La CMF a rarement été utilisée pour mesurer des variations à l'intérieur d'une population homogène. Ceci est assez surprenant compte-tenu des différences de fluorescence que l'on observe dans la plupart des articles scientifiques utilisant cette technique. En effet, le marquage fluorescent de certaines molécules membranaires se répartit parfois sur 2 ou 3 log d'intensité sans être commenté. La présence de cent fois plus de molécules à la surface de certaines cellules est-elle sans conséquence ? Cette observation montre que l'intérêt pour la variabilité est encore quelque chose de récent alors qu'il y a tant d'exemples à disposition pour la mettre en évidence. L'avantage de cette technique réside donc dans son

extrême sensibilité ainsi que dans sa rapidité. Elle peut être facilement standardisée grâce à des billes fluorescentes pour limiter la variabilité instrumentale.

Récemment, des appareils apparus sur le marché proposent de coupler l'imagerie et la cytométrie afin de combiner les avantages de chacune.

#### ◆ *PCR sur cellules uniques*

---

Nous avons mentionné plus tôt la technique de PCR, technique de base de la biologie moléculaire, en soulignant ses limites pour l'étude de la variabilité. Cette technique a aussi bénéficié d'avancées majeures qui ont rendu possible l'analyse d'un plus faible nombre de cellules. Il est actuellement possible d'analyser des cellules uniques de façon quantitative par qRT-PCR ou qRT<sup>2</sup>PCR, mais la précision de la technique est assez limitée. Il est difficile de faire la part des choses entre les variabilités biologiques et instrumentales, même si certains articles affirment pouvoir analyser des cellules uniques (Diercks et al., 2009).

Les prochaines années verront très probablement l'amélioration de la sensibilité de cette technique qui sera alors un outil de choix. En effet, c'est une technique rapide qui, couplée avec un trieur de cellules par exemple, permettra d'analyser facilement des centaines de cellules isolées en parallèle.

#### ◆ *Dispositifs micro-fluidiques*

---

Des systèmes dit micro-fluidiques ont été développés ces dernières années (Zhang et al., 2007). Ils ont l'avantage d'être de taille réduite et ainsi de minimiser les volumes, notamment de réactifs. Leur taille permet aussi une régulation extrêmement précise de l'environnement et des flux internes. Ces dispositifs modelables trouvent aujourd'hui une application en microscopie pour le suivi de cellules isolées. Ils permettent de piéger les cellules, de perfuser des milieux de cultures différents ou d'ajouter des drogues.

En biologie moléculaire, ces dispositifs sont utilisés pour la réalisation de PCR sur cellules isolées (Zhang et al., 2006; Zhang and Ozdemir, 2009). Les volumes sont extrêmement

réduits (quelques pl) et les changements de température plus rapides, ce qui réduit la durée de l'expérience. Toutes les étapes sont réunies dans le même dispositif pour limiter les pertes de matériel. Certains dispositifs commerciaux proposent jusqu'à 96 réactions différentes pour 96 échantillons uniques, sur le même dispositif (fluidigm, Advalytix.). Le système permet ainsi de coupler toutes les réactions mais aussi d'en détecter le résultat.

### *1.6.b. Outils biologiques*

---

La grande majorité des techniques décrites ci-dessus fait appel à la détection de fluorescence et il existe aujourd'hui une grande diversité de molécules, purement chimiques ou de nature protéique, ayant cette propriété. On peut aujourd'hui, sur des critères tant fonctionnels que structuraux, détecter des propriétés et activités cellulaires très diverses. Les anciennes techniques de biologie moléculaire telles que la PCR ou l'hybridation sur membrane (western & northern blot) ont aussi profité de ces molécules fluorescentes. Elles sont devenues plus sensibles et quantitatives, permettant de descendre à l'échelle de la cellule unique, notamment pour la PCR.

Depuis la découverte de la GFP (Shimomura et al., 1962), nous assistons à une explosion du nombre de protéines fluorescentes disponibles (Rizzo et al., 2009). Isolée et purifiée à partir d'une méduse (*Aequorea victoria*), cette protéine a révolutionné la biologie, si bien qu'un prix Nobel de chimie a récompensé les auteurs de sa découverte et de sa caractérisation biochimique en 2008. Sa version sauvage a été peu utilisée comme outil en biologie. Elle a l'inconvénient d'être excitée à une longueur d'onde proche des UV qui est délétère pour la cellule. Dès lors, de nombreux mutants ont été obtenus par mutagenèse dirigée ou aléatoire, notamment contre le chromophore (Heim et al., 1995). Les mutations sélectionnées ont eu pour conséquences l'augmentation de l'intensité de la fluorescence et une meilleure photo-stabilité. Ces nouvelles variantes sont de mieux en mieux optimisées pour être compatibles avec les filtres standards en imagerie, notamment le FITC. D'autres mutations ont entraîné des changements dans les spectres d'excitation et d'émission, générant des couleurs différentes, une sensibilité au pH, des protéines photo-activables ou encore une maturation plus rapide à 37°C. Chaque mutant possède des caractéristiques

spécifiques qui font des protéines fluorescentes un outil de choix pour de nombreuses situations expérimentales.

Les différentes avancées technologiques mentionnées ci-dessus ont clairement participé à l'intérêt actuel pour la variabilité inter cellulaire (Tableau 1). Naturelles ou issues du génie génétique, la fusion de ces protéines fluorescentes avec une protéine d'intérêt permet de visualiser directement en temps réel et *in vivo* cette protéine, d'étudier sa synthèse, sa dégradation, sa localisation ou encore ses interactions avec d'autres protéines (FRET). On peut suivre la dynamique membranaire, nucléaire et chromatinienne mais aussi celle de l'expression génique à l'échelle cellulaire.

Reference	Fluorescent reporters used	Analytical method	Organism
Atkinson <i>et al</i> (2003)	CFP	Fluorescence microscopy	<i>E. coli</i>
Austin <i>et al</i> (2006)	GFP <sub>pasv</sub> -destabilized GFP variant with half-life around 110 min, GFP <sub>paav</sub> -destabilized GFP variant (Andersen 98) with half-life around 60 min	Time-lapse microscopy	<i>E. coli</i>
Bar-Even <i>et al</i> (2006)	GFP fusion proteins (from Invitrogen's yeast GFP clone collection) containing the coding sequence of <i>Aequorea victoria</i> GFP (S65T). CFP and YFP fusions used for dual-reporter assay	Flow cytometry, fluorescence microscopy	<i>S. cerevisiae</i>
Bean <i>et al</i> (2006)	Yeast- and FACS-optimized, destabilized GFP variant with half-life around 30 min (Mateus and Avery, 2000)	Time-lapse microscopy	<i>S. cerevisiae</i>
Becskei and Serrano (2000)	EGFP	Fluorescence microscopy	<i>E. coli</i>
Becskei <i>et al</i> (2001)	yEGFP (yeast-enhanced green fluorescent protein)	Fluorescence microscopy	<i>S. cerevisiae</i>
Colman-Lerner <i>et al</i> (2005)	CFP, YFP	Time-lapse microscopy	<i>S. cerevisiae</i>
Elowitz and Leibler (2000)	GFP <sub>paav</sub> -destabilized GFP variant (Andersen 98) with half-life around 90 min	Time-lapse microscopy	<i>E. coli</i>
Elowitz <i>et al</i> (2002)	CFP, YFP (wild-type codons developed by the University of Washington Yeast Resource Center)	Time-lapse microscopy	<i>E. coli</i>
Fung <i>et al</i> (2005)	gfpmut3.1AAV-destabilized GFP variant	Time-lapse microscopy	<i>E. coli</i>
Gardner <i>et al</i> (2000)	gfpmut3 (FACS-optimized GFP mutant; see Cormack <i>et al</i> , 1996)	Flow cytometry	<i>E. coli</i>
Geva-Zatorsky <i>et al</i> (2006)	ECFP from pECFP-C1 (Clontech), EYFP from pEYFP-1 (Clontech)	Time-lapse microscopy	MCF-7 (human breast cancer)
Guido <i>et al</i> (2006)	gfpmut3 (FACS-optimized GFP mutant; see Cormack <i>et al</i> , 1996)	Flow cytometry	<i>E. coli</i>
Lahav <i>et al</i> (2004)	ECFP from pECFP-C1 (Clontech), EYFP from pEYFP-1 (Clontech)	Time-lapse microscopy	H1299 & MCF-7 (human lung and breast cancer)
Nelson <i>et al</i> (2004)	DsRed and EGFP from pEGFP-N1 and pDsRed-N1 (Clontech)	Time-lapse microscopy	HeLa and SK-N-AS
Ozbudak <i>et al</i> (2002)	gfpmut2 (FACS-optimized GFP mutant; see Cormack <i>et al</i> , 1996)	Flow cytometry	<i>B. subtilis</i>
Pedraza and van Oudenaarden (2005)	CFP, YFP, RFP from pECFP, pEYFP, and pDsRed-Express (Clontech)	Fluorescence microscopy	<i>E. coli</i>
Raser and O'Shea (2004)	yECFP (yeast-optimized enhanced cyan fluorescent protein) and yVYFP (yeast-optimized Venus yellow fluorescent protein)	Fluorescence microscopy	<i>S. cerevisiae</i>
Rosenfeld <i>et al</i> (2002)	gfpmut3 (FACS-optimized GFP mutant; Cormack <i>et al</i> , 1996)	Fluorimeter	<i>E. coli</i>
Rosenfeld <i>et al</i> (2005)	CFP, YFP (yfp gene from pDHS plasmid, University of Washington Yeast Resource Center)	Time-lapse microscopy	<i>E. coli</i>
Sigal <i>et al</i> (2006)	YFP	Time-lapse microscopy	H1299 (human lung cancer)
Suel <i>et al</i> (2006)	CFP, YFP	Time-lapse microscopy	<i>B. subtilis</i>
Volfsen <i>et al</i> (2005)	yEGFP (yeast-enhanced green fluorescent protein)	Flow cytometry	<i>S. cerevisiae</i>
Weinberger <i>et al</i> (2005)	Enhanced GFP (Clontech, Palo Alto, CA)	Flow cytometry	Jurkat cells

**Tableau 1 Systèmes d'étude de l'expression génique à l'échelle unicellulaire.**

Toutes ces études font appel à des techniques sur cellules unique et à des protéines fluorescentes.

Tiré de (Longo and Hasty, 2006)

L'utilisation de plus en plus fréquente de ces techniques a généré une somme d'observations non-négligeables mettant en avant le rôle prépondérant et irréfutable de la variabilité en biologie, notamment chez les eucaryotes dont la complexité semble percée un peu plus chaque jour. Après avoir décrit quelques exemples de variabilité en biologie, la seconde partie est consacrée à la mise en évidence et à la recherche de l'origine de la variabilité inter-cellules.

## II. Origine de l'hétérogénéité inter-cellulaire

---

L'hétérogénéité entre cellules isogéniques apparaît spontanément dans un environnement homogène ( Variabilité non-génétique du phénotype). Ces cellules peuvent exprimer des protéines données à des niveaux différents malgré leur origine clonale. Il a donc été proposé l'existence d'une certaine part aléatoire dans le processus d'expression génique aboutissant à une telle variabilité.

### II.1. L'expression génique est un processus stochastique

---

La nature aléatoire de l'expression génique et du fonctionnement cellulaire n'a été théorisée que dans les années 1980 (Kupiec, 1983). Il aura fallu attendre les années 2000 pour qu'elle soit reconnue par la communauté scientifique et commence à faire l'objet d'études spécifiques.

En 1994, Ross et al. étudient la production de l'enzyme bêta-galactosidase sous le contrôle du promoteur LTR de HIV-1 et observent, comme Ko et al., l'apparition sporadique de cellules exprimant le transgène (Ko et al., 1990; Ross et al., 1994). Dans cet article, les auteurs vont plus loin et postulent que la transcription chez les eucaryotes se déroule de façon pulsative, espacée de périodes de longue inactivité. La régulation transcriptionnelle est alors proposée comme reposant sur une modification de la probabilité de transcription plutôt que sur la modification du taux de transcription.

En utilisant des modèles informatiques et des algorithmes de modélisation stochastique des réactions biochimiques (dérivés de (Gillespie, 1977)), McAdams et Arkin postulent comme Ross et al., que la synthèse protéique est un processus sporadique se produisant par évènements rares (McAdams and Arkin, 1997). Ils prédisent que ce comportement stochastique sera renforcé si les évènements de transcription et le nombre de protéines par transcrit sont des évènements rares (ie. Si l'efficacité de traduction est faible). Ces conclusions les amènent à prédire que la nature stochastique de l'expression génique et de

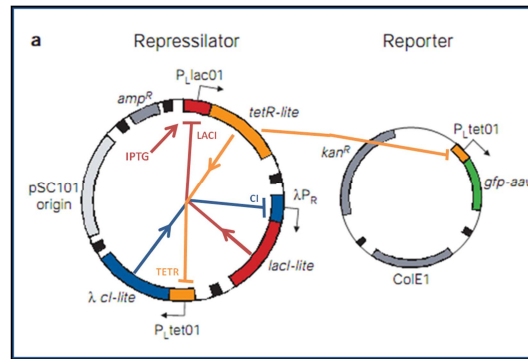


sa régulation sont des sources majeures de la variabilité des événements cellulaires observés.

On trouve aujourd'hui de nombreux synonymes dans la littérature à cette notion de variabilité, tels que « stochasticité » ou « bruit » de l'expression génique. Tous ces termes rappellent que le niveau d'expression d'un gène donné n'est pas identique dans toutes les cellules d'une population, même d'origine clonale. Elle génère en conséquence une identité cellulaire non-génétique. La notion de « bruit », héritée de l'anglo-saxon « noise », porte une connotation péjorative car elle rappelle le « bruit » qui perturbe les signaux. Comme de nombreux termes en génétique, ce terme est emprunté au langage informatique ou à l'électronique, sans doute un reliquat de l'âge d'or de la biologie moléculaire. En faisant ce parallèle, on perçoit alors ces fluctuations de l'expression génique comme un processus délétère pour la cellule. Celle-ci doit alors mettre en œuvre des processus de compensation pour stabiliser son état et tamponner ces fluctuations.

Durant cette période, on assiste à l'essor d'une autre discipline : la biologie des systèmes ou biologie synthétique. Grâce aux connaissances accumulées, notamment en biologie moléculaire, les recherches évoluent de la caractérisation des gènes et de leur produit vers la compréhension de leur régulation, de leurs interactions les uns avec les autres. De façon intéressante, la plupart des articles théoriques traitant de ces réseaux géniques sont l'œuvre de biophysiciens modélisateurs.

L'un des premiers articles, publié en 2000, utilise un système purement synthétique pour étudier la régulation d'un réseau génique. Le principe est simple en apparence mais va déboucher sur une révolution en biologie (Elowitz and Leibler, 2000). Le système utilisé, dénommé « repressilator », est basé sur l'utilisation de 3 répresseurs transcriptionnels dans le but de construire un système oscillant par répression réciproque des trois gènes (Figure 6). Les trois répresseurs utilisés n'appartiennent pas à l'organisme utilisé, ici la bactérie *E. coli*.

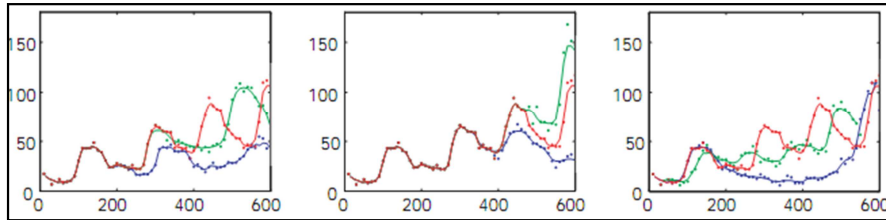


**Figure 6 Constructions plasmidiques utilisées dans le système "repressilator".**

Le premier plasmide contient les trois gènes auto-régulés. Le second plasmide code pour le gène rapporteur GFP. Adapté de (Elowitz and Leibler, 2000).

Le premier répresseur, LacI (en rouge), inhibe la transcription du second répresseur *terR* provenant du transposon Tn10 (en jaune) en bloquant le promoteur  $P_{lac01}$ . TetR inhibe à son tour la transcription du troisième répresseur *cI* du phage lambda (en bleu) en bloquant le promoteur  $P_{tet01}$ . Pour finir, CI inhibe l'expression de *lacI* en bloquant le promoteur  $P_r$ , complétant ainsi le cycle. Le niveau de TetR peut être modulé par ajout d'IPTG, qui inhibe l'action de LacI. TetR possède une seconde activité, celle d'inhiber l'expression de la GFP sur un autre plasmide. Cette protéine fluorescente sera le témoin de l'activité du système. Des bactéries ont été transformées avec ces deux constructions puis ont été clonées pour s'assurer de l'homogénéité génétique des cellules.

Dans ce modèle, le comportement du système dépend de plusieurs facteurs, dont la transcription de chacun des répresseurs, leur traduction et la dégradation des ARNm et des protéines. En fonction des paramètres initiaux, deux états peuvent être imaginés : un état stable en équilibre ou un état perturbé. Malgré la faible complexité apparente du réseau mis en œuvre, on observe une forte variabilité dans la durée et l'amplitude des cycles d'expression de la GFP (Figure 7). Elle se traduit par une désynchronisation et une perte des oscillations de l'expression de la GFP dans une même cellule mais aussi entre cellules sœurs.



**Figure 7 Profil d'expression de la GFP dans le système « repressilator ».**

*3 exemples de fluctuations observés dans des cellules isolées au cours du temps (minutes). La ligne rouge est une cellule de référence. Les autres lignes correspondent à des cellules filles. On observe une perte de périodicité entre cellules filles dans la plupart des cas (Elowitz and Leibler, 2000).*

Ceci indique qu'il existe un « bruit » dans le système à l'échelle de la cellule unique. Ce bruit est lié à la fluctuation de la concentration des diverses molécules impliquées dans la régulation du circuit. Même s'il ne permet d'étudier qu'un seul des composants du réseau, ce modèle a été l'un des premiers systèmes de nature artificielle, développé pour comprendre la logique de réseau et leur propriété. Il a participé à l'intérêt de la communauté scientifique pour cette thématique de recherche en expansion. Puisque l'on pensait avoir tout sous contrôle, ce « bruit » observé a intrigué les concepteurs du réseau qui ont poursuivi leur recherche dans ce domaine (Cox et al., 2010).

A partir de ce moment, énormément d'efforts sont faits pour comprendre les processus moléculaires mis en œuvre à l'échelle du gène pour son expression. On commence alors à expliquer la variabilité entre cellules non plus uniquement par les interactions qui existent entre-elles ou avec leur environnement mais aussi par la perturbation de l'expression génique par des phénomènes aléatoires, sans en connaître précisément l'origine.

Au début des années 2000, on assiste à un gain d'intérêt pour cette notion de stochasticité de l'expression génique et plusieurs équipes de recherche commencent à modéliser de façon informatique, avec une approche stochastique, l'expression des gènes. Dans un premier temps à l'échelle d'un gène unique puis à l'échelle de réseaux de gènes. Les premières études sont plutôt théoriques et réalisées par des laboratoires de physique ou de biologie des systèmes mais les prédictions de ces articles se confirmeront dans les années qui suivront.

### II.1.a. La loi des petits nombres

---

Les systèmes cellulaires (bactérie, levure ou cellules eucaryotes supérieures) sont des structures de petite taille dans lesquelles se déroulent de nombreuses réactions biochimiques nécessaires à leur survie. L'efficacité de chacune de ces réactions biochimiques dépend de nombreux facteurs biologiques et physiques (température, pH, etc.) et en particulier la concentration de substrat.

En réalité, il existe parfois un très faible nombre de partenaires moléculaires dans une cellule. De même, la quantité de cofacteurs peut être limitée, réduisant ainsi l'efficacité de certaines réactions biochimiques. Cet aspect quantitatif est très important puisque qu'il fait des réactions biochimiques un processus hautement probabiliste, gouverné par les lois de la diffusion. La taille du système et le nombre de partenaires sont donc des variables de première importance.

L'expression génique ne fait pas exception. Nous connaissons à présent beaucoup de détails moléculaires à propos des différentes étapes de l'expression génique, tant chez les procaryotes que chez les eucaryotes. Chaque livre de biologie moléculaire regorge de schémas toujours plus complexes décrivant les différents partenaires moléculaires recrutés pour la transcription d'un gène ou la traduction d'un ARNm. Néanmoins il ne s'agit que d'une description qualitative de ces processus. Il est rare de voir évoquer l'aspect quantitatif et dynamique de l'expression génique. Pourtant, des questions simples et essentielles doivent être posées : les partenaires sont-ils présents ? Au bon endroit ? Au bon moment ? Quelle est la stœchiométrie du système, le nombre de partenaires est-il suffisant ? (Guptasarma, 1995).

Il est maintenant connu que le nombre d'ARNm à un instant donné dans une cellule peut être très faible, de l'ordre de 5 à 10 molécules (Velculescu et al., 1997; Holland, 2002; Ghaemmaghami et al., 2003) tout comme le nombre de protéines (Taniguchi et al., 2010). Plus important encore, le nombre de copies d'un gène est limité à 2 pour la plupart d'entre

eux et souvent, un seul allèle est transcriptionnellement actif. Sur la base de cet aspect quantitatif, les réactions biochimiques ont une propriété « bruitée » intrinsèque qui est susceptible d'être une source de variabilité de l'expression génique.

L'expression génique étant un processus multi-étapes, on peut facilement imaginer que chacune des étapes soit une source de variations et que les variations des étapes en amont peuvent se transmettre aux étapes suivantes.

### *II.1.b. Des évènements rares sont sources de variabilité*

---

#### *◆ Efficacité de traduction chez les bactéries*

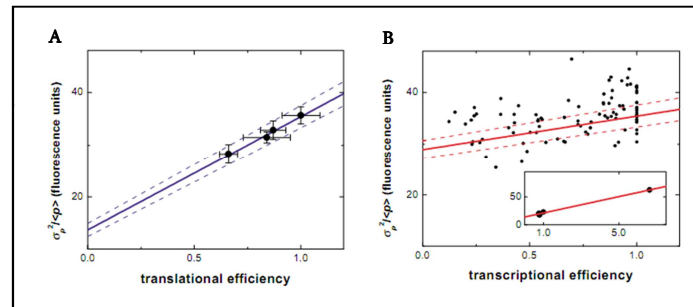
---

En se basant sur l'hypothèse que le faible nombre de molécules était une source de variabilité, Thattai et Van Oudenaarden prédisent par la modélisation que la variabilité de l'expression génique chez la bactérie est fortement liée à l'efficacité de traduction (nombre de protéines synthétisées par molécule d'ARNm) et que des fluctuations intrinsèques en sont à l'origine (Thattai and van Oudenaarden, 2001). Ces fluctuations semblent causées par la structure, la cinétique des réactions et la concentration des substrats des réseaux biochimiques. Ils prédisent qu'une faible quantité d'ARNm et une haute efficacité de traduction se traduiront par une grande variabilité, ce qui a déjà été mentionné (McAdams and Arkin, 1997). De plus, ils discutent de l'implication de ces variations lorsque les protéines soumises à fluctuation font partie d'un réseau dans lequel elles jouent un rôle régulateur, ce qui constitue une part importante des protéines connues (transduction du signal, facteurs de transcription, boucles de rétroaction...).

Le rôle de l'efficacité de la traduction est confirmé par le même laboratoire par la modulation du taux de transcription (par induction du promoteur) et du taux de traduction (mutation du RBS) d'un rapporteur fluorescent chez *Bacillus subtilis* (Ozbudak et al., 2002). Les auteurs décrivent une façon de quantifier le bruit, en se basant sur la moyenne et la variance de la fluorescence de leur population cellulaire. Ils définissent le

bruit comme le rapport de la variance et de la moyenne de fluorescence (on parle aussi de variance normalisée).

Une augmentation de l'induction (augmentation de l'efficacité de transcription) se traduit par une augmentation de la synthèse protéique, mais n'a pas d'influence sur le bruit (Figure 8, B). En revanche, une diminution de l'efficacité de traduction a pour effet de diminuer le bruit de l'expression du gène rapporteur (Figure 8, A).



**Figure 8 Traduction et variabilité de l'expression génique chez *B. Subtilis*.**

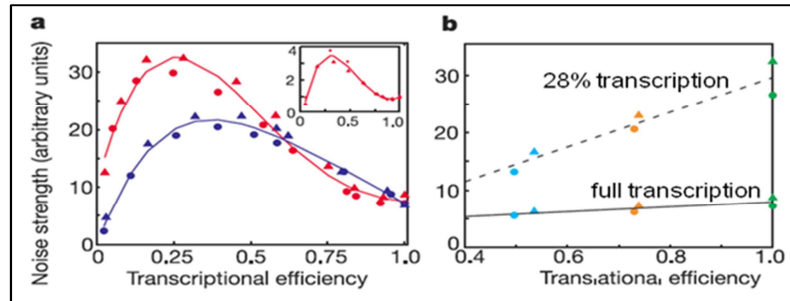
La réduction de l'efficacité de traduction (protéines par mRNA) réduit le bruit (b) alors que la réduction de l'efficacité de transcription (mRNA par burst) n'a presque pas d'effet(c) (Ozbudak et al., 2002).

La quantité d'ARNm et le taux de traduction sont donc les limites majeures à la stabilité de l'expression génique dans ce modèle. Le rôle dominant de l'efficacité de traduction dans les fluctuations de l'expression génique chez *Bacillus subtilis*, permet de supposer qu'il est possible de minimiser le bruit en la modulant. La faible efficacité de traduction serait un moyen de réduire le bruit de l'expression de certains gènes, et serait potentiellement sélectionnée pour cet avantage. L'hypothèse selon laquelle la synthèse protéique est de nature pulsative est ainsi confirmée.

#### ◆ Efficacité de transcription chez les eucaryotes

Dans une étude comparable chez la levure *Saccharomyces cerevisiae*, un organisme eucaryote unicellulaire, la variabilité de l'expression génique a été étudiée en modulant la transcription et la traduction d'un gène rapporteur inducible (Blake et al., 2003). L'augmentation de l'efficacité de la transcription joue ici un rôle majeur dans la génération de bruit (Figure 9). La relation entre le bruit et l'efficacité de transcription n'est pas linéaire. Lorsque l'efficacité de transcription est d'environ 1/4, le bruit atteint un

maximum. Si l'efficacité de traduction diminue dans cette condition, on observe une diminution du bruit, comme chez la bactérie (Figure 9, a). En revanche, lorsque l'efficacité de transcription est maximale, l'augmentation de l'efficacité de traduction a très peu d'effet contrairement à l'étude précédente (Figure 9, b).



**Figure 9 Transcription, traduction et variabilité de l'expression génique (*S. cerevisiae*).**

*La réduction de l'efficacité de transcription (mRNA par burst) augmente le bruit de l'expression génique (A). La réduction de l'efficacité de traduction (protéines par mRNA) réduit le bruit lorsqu'il est à un niveau élevé (B). Les deux étapes semblent importantes dans la régulation du bruit (Blake et al., 2003).*

Contrairement aux procaryotes, l'efficacité de transcription et celle de traduction sont impliquées dans la variabilité de l'expression génique dans ce modèle eucaryote. Néanmoins, la transcription joue un rôle plus important que l'efficacité de traduction. Grâce à un modèle mathématique, les auteurs démontrent la nécessité d'une expression sporadique des ARNm pour que le bruit soit influencé par l'efficacité de transcription. Ils prédisent que la variabilité de l'expression génique chez les eucaryotes est liée à la dynamique et à la stabilité des complexes multi protéiques impliqués dans la régulation de la transcription. Le rôle de la chromatine est évoqué afin d'expliquer la production sporadique d'ARNm. La chromatine aurait pour effet de modifier l'état d'activation du promoteur, autorisant une forte transcription lorsque celui-ci est actif puis en la réprimant en rendant le promoteur inactif. Il y aurait alors de fortes périodes de transcription et de traduction générant des fluctuations de l'abondance de protéines au cours du temps.

En 2006, une étude chez la levure mesure la variabilité de l'expression de plusieurs gènes endogènes (Bar-Even et al., 2006). En réalisant des protéines de fusion entre la GFP et diverses protéines d'intérêt, les auteurs mesurent le bruit en fonction de la quantité moyenne de 43 protéines différentes et dans 11 conditions de culture différentes. La même

année, Newman et al. publient un travail similaire dévoilant que le bruit est inversement corrélé avec l'abondance de protéines soulignant l'effet du bruit intrinsèque lorsque le nombre de protéines est faible (Newman et al., 2006). De façon intéressante, certaines protéines ne respectent pas cette dynamique et sont encore plus variables. Il s'agit de protéines impliquées dans la réponse au stress qui ont une composition particulière de leur promoteur (présence de TATAA box, mentionnée par (Blake et al., 2003; Raser and O'Shea, 2004)) indiquant que la transcription joue un rôle important.

Chez les cellules humaines, la même approche a été utilisée en étudiant 20 protéines de fusion situées dans le noyau (Sigal et al., 2006). Contrairement aux autres approches utilisant la cytométrie en flux, celle-ci utilise l'imagerie pour suivre l'évolution de la fluorescence au cours de 2 à 3 générations cellulaires. L'expression des protéines dans ce modèle est relativement stable dans des cellules individuelles et il faut plusieurs générations pour observer des variations significatives. Les fluctuations ne sont pas aussi importantes que celles observées dans d'autres modèles. Un biais de sélection a pu être introduit lors de la construction des lignées cellulaires exprimant les protéines de fusion en favorisant celles qui avaient un niveau d'expression suffisamment élevé, et donc avec une variabilité plus faible. Il est aussi très probable que la stabilité des protéines (leur taux de dégradation) joue un rôle important dans la variabilité observée. Une durée de vie longue aura pour effet de masquer la nature sporadique de la transcription alors que des protéines à courte durée de vie verront leur abondance corrélée avec le niveau d'ARNm.

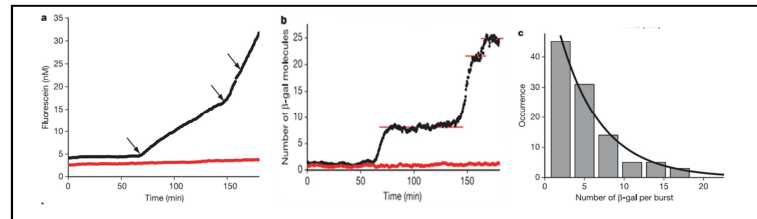
#### ◆ *Dynamique d'expression à l'échelle moléculaire*

---

De récentes études ont montré que le suivi de protéines uniques peut être utilisé pour visualiser la dynamique d'expression d'un gène chez la bactérie (Cai et al., 2006; Yu et al., 2006; Choi et al., 2008). En mesurant la fluorescence d'une protéine ou d'un produit de réaction au cours du temps (Figure 10, a), on peut estimer le nombre absolu d'enzymes présentes dans la cellule ayant participé aux fluctuations du signal observé (Figure 10, b).



A partir de cette estimation on peut calculer le nombre de molécules produites par évènement de synthèse (Figure 10, c).



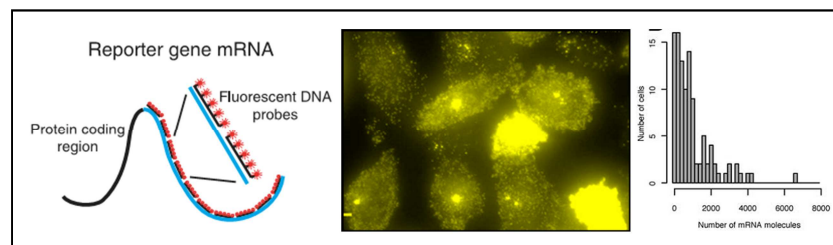
**Figure 10 Analyse de la dynamique d'expression par suivi de molécule unique.**

*A : Évolution de la fluorescence dans une cellule unique au cours du temps. B : Estimation du nombre de molécules dans la cellule. C : distribution du nombre de molécules produites par évènement de synthèse.*

*Tiré de (Cai et al., 2006).*

Cette résolution moléculaire a ainsi permis de confirmer expérimentalement la prédiction selon laquelle la synthèse protéique était sporadique. Néanmoins dans certaines conditions, il n'est pas possible de visualiser directement le produit du gène si la concentration est trop faible. C'est particulièrement le cas chez les eucaryotes qui ont un volume cellulaire important. D'autre part, un faible signal de la GFP se confond aisément avec le bruit de fond. Une autre limite à l'analyse de fluorescence est que celle-ci n'est pas parfaitement proportionnelle à la quantité absolue de protéines.

Pour contourner cette limite, il est possible de quantifier directement les ARNm dans des cellules uniques. Les premières approches par hybridation *in situ* (Figure 11) montrent une grande disparité entre cellules isolées dans la même culture (Raj et al., 2006, 2008).



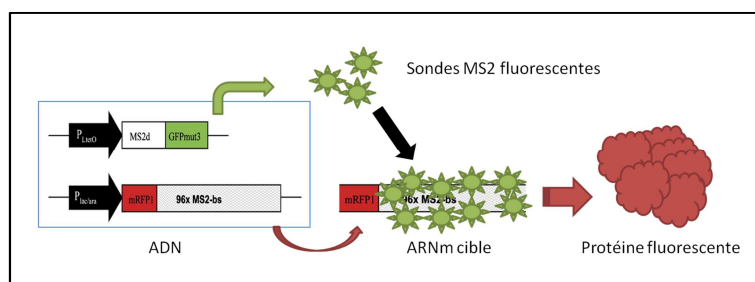
**Figure 11 Quantification du nombre de transcrits cellulaires par FISH.**

*A : L'ARNm possède des séquences répétées artificielles permettant l'hybridation de sondes nucléiques fluorescentes. B : Les ARNm sont visibles par imagerie à fluorescence, leur distribution est hétérogène entre les cellules d'une même population. C : Distribution du nombre d'ARNm par cellule obtenue après analyse d'images.*

*Tiré de (Raj et al., 2006)*

Des zones de forte transcription témoignent de la nature sporadique de la production d'ARNm dans des cellules de mammifères, probablement liée à la dynamique d'activation du gène par des mécanismes chromatinien de régulation. En effet, les auteurs montrent que des gènes situés à proximité du même locus sont soumis à une dynamique d'expression similaire, en termes de pics de synthèse.

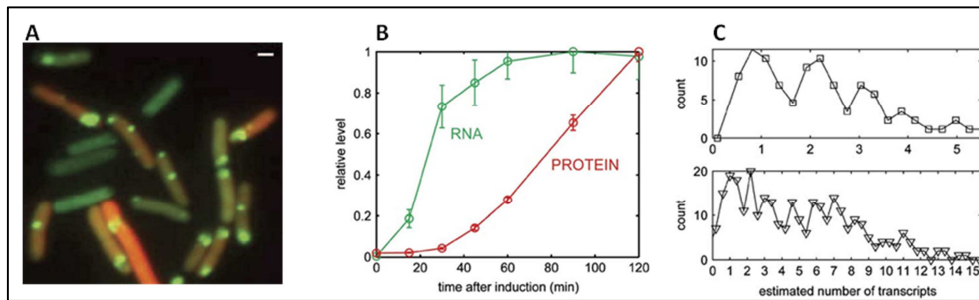
Une seconde approche consiste à quantifier en temps réel la production d'ARNm dans des cellules isolées (Figure 12). Par exemple, Golding et al. utilisent chez la bactérie, des sondes couplées à la GFP appelées MS2-GFP. Cette protéine maturase du phage MS2 se fixe sur des structures secondaires présentes sur l'ARNm d'intérêt (Bertrand et al., 1998; Golding and Cox, 2004; Golding et al., 2005).



**Figure 12 Détection des transcrits in vivo par des sondes MS2.**

*Exprimées dans les cellules cibles en même temps que le transgène d'intérêt, les sondes MS2d-GFP se fixent sur leur cible ARNm au niveau de séquences spécifiques MS2-bs. Dans une cellule, les ARNm et les protéines peuvent être quantifiés simultanément grâce à leur fluorescence respective. Adapté de (Golding et al., 2005).*

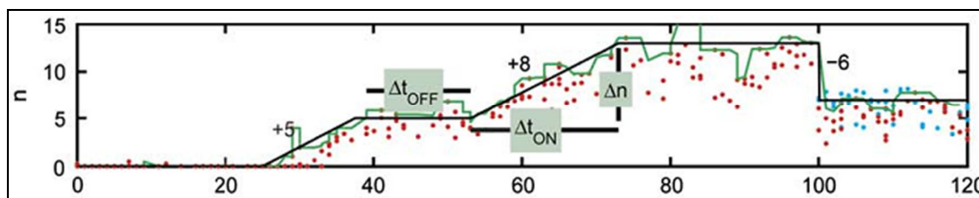
L'expression de la sonde MS2-GFP dans ce système est inductible ce qui permet de contrôler la quantité de sonde dans chaque cellule. Il faut être dans une gamme de concentration suffisamment élevée pour saturer tous les ARNm tout en évitant le bruit de fond trop important. Golding et al. (2005) ont estimé que dans leurs conditions expérimentales, chaque cellule contient environ 10 000 sondes dont 3 à 4% seulement sont fixés à l'ARN cible. L'intérêt de ce système est que le gène rapporteur, dont l'ARNm constitue la cible des sondes MS2, code pour une protéine fluorescente (mRFP). Sur une même image, il est possible de quantifier simultanément le niveau d'ARNm et de protéines, suite à l'induction du promoteur du gène rapporteur (Figure 13 A et B). Grâce à une calibration de la fluorescence, le nombre de transcrits par cellule peut être estimé (Figure 13 C).



**Figure 13 Dynamique de la transcription et de la traduction avec les sondes MS2.**

*A : Bactéries exprimant le transgène MS2-GFP (vert) et le transgène mRFP1 (rouge). B : Quantification des transcrits et des protéines lors de l'induction de l'expression du transgène mRFP1. C : Distribution du nombre de transcrits par cellule, dans deux expériences indépendantes. Tiré de (Golding et al., 2005).*

La dynamique de synthèse des ARNm et de la protéine est différente si bien que la corrélation entre la quantité des deux types de molécules est faible. Cette mauvaise corrélation entre ARNm et protéines a été confirmée dans une autre étude (Taniguchi et al., 2010). Grâce à cette technique, l'évolution en temps réel du nombre de transcrits peut aussi être étudiée par vidéo-microscopie, en analysant des cellules individuelles (Figure 14).



**Figure 14 Transcription en temps réel sur cellule isolée.**

*Le suivi d'une cellule isolée permet de quantifier l'importance des événements ( $\Delta N$ ), les périodes d'activité ( $\Delta t_{ON}$ ) et d'inactivité ( $\Delta t_{OFF}$ ) de la transcription. Tiré de (Golding et al., 2005).*

Les événements de synthèse et de dégradation des ARNm peuvent être décrits de façon très précise. Ces sondes ont ainsi permis une avancée majeure dans la compréhension de la dynamique de l'expression génique. Elles permettent de confirmer la nature sporadique de la transcription chez les bactéries par l'observation directe de l'accumulation et de la dégradation de molécules individuelles. Cette stratégie a aussi été utilisée avec des cellules eucaryotes avec les mêmes conclusions (Yunger et al., 2010).

La propriété intrinsèquement aléatoire des réactions biochimiques liée au faible nombre d'acteurs n'est pas la seule source possible de fluctuation dans l'expression génique.

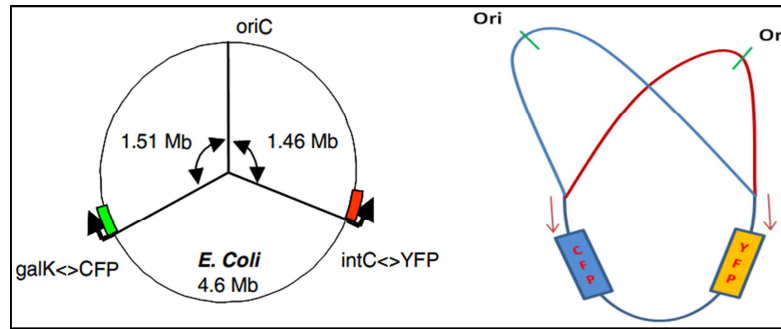
Chaque réaction est potentiellement « bruitée » et participe à un bruit général. Ainsi, certains facteurs qui influencent l'efficacité de transcription (facteurs de transcription, polymérase...) ou celle de la traduction (nombre de ribosomes, disponibilité en acides aminés ...), la progression dans le cycle cellulaire ou des changements environnementaux peuvent être sources de fluctuations. Les fluctuations observées peuvent donc provenir du phénomène d'intérêt et/ou du bruit de fond général lié à la stochasticité de l'ensemble des réactions biochimiques cellulaires.

### *II.1.c. Identification de l'origine des variations*

---

Afin de faire la différence entre ces deux sources de fluctuation, Elowitz et al. proposent un modèle biologique qui permet d'identifier la source des variations de l'expression génique (Elowitz et al., 2002). Il s'agit d'un modèle bactérien mais qui sera le point de départ de nombreux travaux expérimentaux dans ce domaine et chez de nombreux organismes.

Dans cet article, ils introduisent la notion de sources intrinsèque et extrinsèque des fluctuations de l'expression génique. Ces fluctuations seront analysées de façon mathématique dans un second article la même année (Swain et al., 2002). Ces notions ont depuis été approfondies mais nécessitent néanmoins toujours une description simple du système, afin de définir ce qui lui appartient (« intrinsèque ») de ce qui lui est étranger (« extrinsèque ») (Paulsson, 2005; Hilfinger and Paulsson, 2011). Pour mesurer la contribution de chacune des sources de fluctuation, ils intègrent au génome bactérien d'*E. coli*, 2 gènes rapporteurs sous le contrôle de promoteurs identiques (Figure 15). L'un d'eux contrôle l'expression de la eCFP, l'autre celle de la eYFP, deux protéines fluorescentes : le système étudié étant alors les deux gènes rapporteurs.



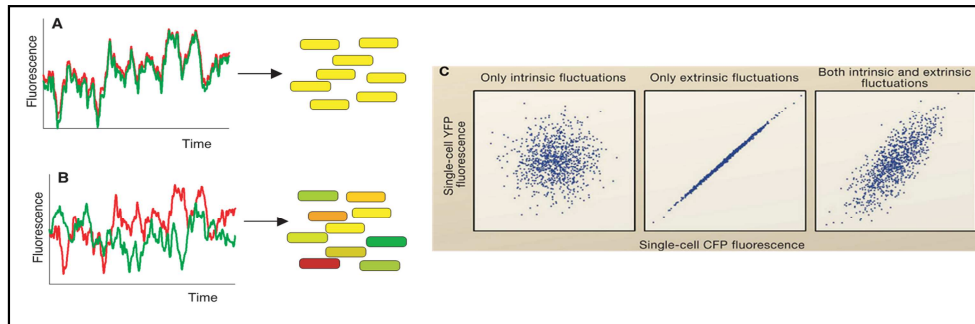
**Figure 15 Chromosome bactérien utilisé pour le modèle bifluorescent.**

*Les deux transgènes possédant le même promoteur sont situés à égale distance de l'origine de réplication.*

*Tiré de (Elowitz et al., 2002)*

En reprenant les définitions précédentes, les fluctuations extrinsèques sont alors liées à la concentration, la position ou plus globalement la disponibilité des éléments régulateurs nécessaires à l'expression des transgènes. Elles sont communes pour chacun des transgènes d'une cellule qui partagent le même environnement global. Ils peuvent varier dans une cellule au cours du temps et entre cellules d'une même population (Figure 16). Les fluctuations extrinsèques expliquent principalement la différence entre cellules dans une population. La position dans le cycle cellulaire participe de façon majeure à cette composante du bruit chez la levure par exemple (Colman-Lerner et al., 2005), tout comme la taille de la cellule et le taux de croissance (Snijder et al., 2009; Tsuru et al., 2009; Zeng et al., 2010). Chez les cellules de mammifères, la densité cellulaire peut être source de variabilité extrinsèque (Stockholm et al., 2007; Neildez-Nguyen et al., 2008).

La source intrinsèque correspond à la nature discrète des événements biochimiques décrits précédemment. Elle est responsable de la différence d'expression entre des transgènes dans une même cellule. La quantification de l'expression des deux transgènes dans des cellules uniques permet alors de déterminer l'origine de la source de fluctuations. Si les deux transgènes évoluent de la même façon, la source est extrinsèque, dans le cas contraire, elle est intrinsèque (Figure 16).



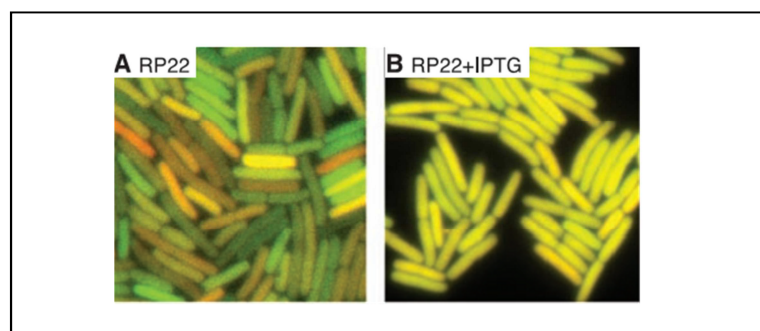
**Figure 16 Identification de la source de fluctuation de l'expression génique.**

*Les fluctuations extrinsèques perturbent les deux transgènes simultanément : leur expression est donc corrélée. Le bruit intrinsèque, en revanche, est spécifique à un des transgènes : l'expression n'est plus corrélée.*

*Tiré de (Raj and van Oudenaarden, 2008)*

Les auteurs ont construit plusieurs souches bactériennes dont les deux transgènes sont sous le contrôle d'un promoteur inductible par le lactose. Ils ont montré clairement que l'expression constitutive d'un gène dans une souche sans répresseur lacI<sup>-</sup> peut être extrêmement uniforme dans certaines conditions. De même, ceci ne dépend pas du promoteur utilisé. Toutes les bactéries expriment les deux transgènes au même niveau.

Dans un deuxième temps, ils montrent dans une souche LacI<sup>+</sup> sans induction, que l'expression n'est plus que de 5% environ et que le bruit intrinsèque augmente d'un facteur 5 (Figure 17, A). L'effet est réversible en ajoutant une quantité saturante d'inducteur (IPTG), ce qui démontre que le bruit intrinsèque est directement lié à l'activité de LacI (Figure 17, B).



**Figure 17 Visualisation du bruit intrinsèque dans un modèle bactérien.**

*Le bruit intrinsèque est réduit par ajout d'une quantité saturante d'inducteur.*

*Tiré de (Elowitz et al., 2002)*

Malgré l'induction par une quantité saturante d'IPTG, des différences persistent entre cellules ce qui indique qu'il existe des différences entre cellules dans l'expression de LacI. Ces travaux montrent ainsi que tout facteur soumis à une source intrinsèque de

fluctuation de son abondance (LacI dans ce cas) peut à son tour être une source extrinsèque de fluctuations pour un autre système.

Raser et O'Shea étudient plus précisément la contribution intrinsèque et extrinsèque de la variabilité de l'expression génique chez la levure en utilisant un modèle proche de celui d'Elowitz et al. décrit précédemment (Elowitz et al., 2002; Raser and O'Shea, 2004). Ils montrent que le bruit intrinsèque est lié à l'induction de la transcription et au taux d'activation du promoteur. Ceci est confirmé en créant des mutants pour lesquels les facteurs de remodelage de la chromatine sont recrutés plus lentement. L'induction et les mutations ont le même effet (augmentation du bruit intrinsèque) ce qui prouve que le remodelage chromatinien autour du promoteur joue un rôle dans la stochasticité de l'expression génique. Dans un deuxième temps, des mutations de la TATA box, qui régule le taux de transcription, provoquent une diminution du bruit intrinsèque de l'expression génique, traduisant le rôle de l'étape de transcription dans le bruit intrinsèque.

L'importance de la structure du promoteur dans la variabilité de l'expression génique a été démontrée dans d'autres études. Chaque promoteur endogène possède une activité propre créant une dynamique d'expression qui lui est spécifique (Suter et al., 2011).

## II.2. La régulation du bruit

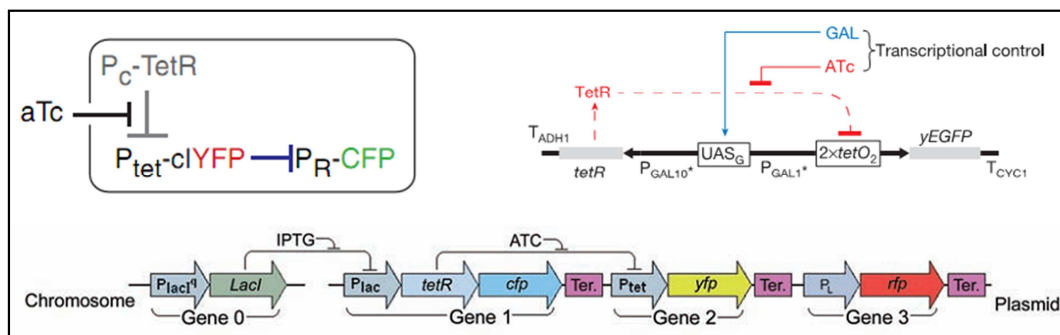
---

L'étude de la stochasticité de l'expression génique est encore récente mais elle a déjà fait l'objet de nombreuses études. De façon intéressante, c'est une thématique qui a regroupé plusieurs champs disciplinaires avec bien entendu des biologistes, mais aussi des physiciens et des mathématiciens. L'apport de ces deux dernières disciplines pour la compréhension de la stochasticité de l'expression est majeur. Elles ont permis notamment de créer des modèles mathématiques et des simulations numériques afin de tenter d'expliquer l'origine, mais aussi les conséquences de la stochasticité de l'expression génique. Des modèles probabilistes de l'expression génique ont vu le jour et sont à même de reproduire les observations expérimentales. Plus important encore, ces modèles sont aussi capables d'expliquer des phénomènes considérés comme déterministes.

La plupart des réactions biochimiques cellulaires font partie de réseaux, les produits de certaines réactions pouvant servir de substrat à d'autres réactions, propageant ainsi le signal primaire tout au long de la cascade (ie. transduction du signal,...).

### II.2.a. Transmission du bruit dans les réseaux géniques

Pour la plupart d'entre eux, les gènes endogènes sont aussi organisés en réseaux. Ils comprennent un certain nombre de partenaires interagissant les uns avec les autres, en suivant des dynamiques de régulation variées. Néanmoins, ces réseaux sont peu étudiés dans leur contexte biologique en raison de leur grande complexité (certains partenaires sont encore inconnus ou la relation entre partenaires n'est pas identifiée). La plupart des recherches sur ces réseaux fait appel à des réseaux synthétiques ou à la modélisation informatique (Figure 18). A partir de ces réseaux synthétiques simples, il est possible d'étudier la dynamique de réseaux plus complexes (Guido et al., 2006).



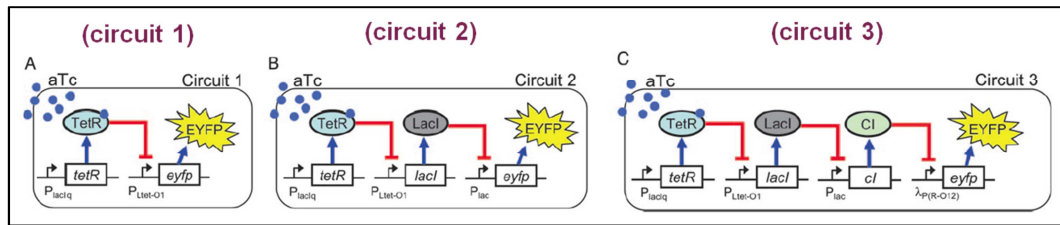
**Figure 18 Exemples de réseaux géniques synthétiques.**

*L'analyse des dynamiques d'expression et de la transmission des fluctuations se fait le plus fréquemment grâce à des protéines fluorescentes.*

Les cascades sont des structures de régulation fréquentes et la mise en évidence de la stochasticité de l'expression génique à l'échelle d'un gène a soulevé la question de la propagation du bruit dans ces cascades de régulation. Les premières études ont montré que le bruit à chaque étape pouvait se transmettre dans une cascade de régulation transcriptionnelle, s'ajoutant au bruit intrinsèque de l'étape suivante (Rosenfeld et al., 2002, 2005; Pedraza and van Oudenaarden, 2005).



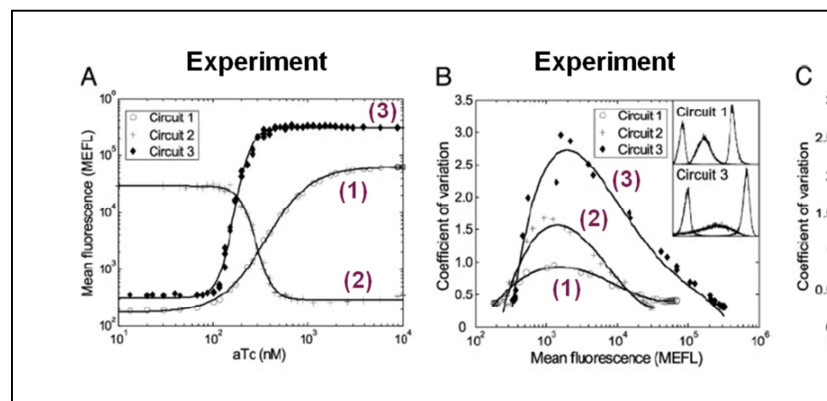
Hooshangi et al. étudient ce phénomène dans un réseau synthétique inductible chez E.Coli, comportant 1, 2 ou 3 étapes de répression transcriptionnelle (Hooshangi et al., 2005; Hooshangi and Weiss, 2006).



**Figure 19 Propagation du bruit dans un réseau génétique.**

Tiré de (Hooshangi et al., 2005)

Ils montrent dans un premier temps qu'une architecture différente du réseau peut modifier sa sensibilité/robustesse (Figure 20, A). Le niveau d'expression n'est pas linéairement lié au niveau d'induction et dépend de la longueur de la cascade.



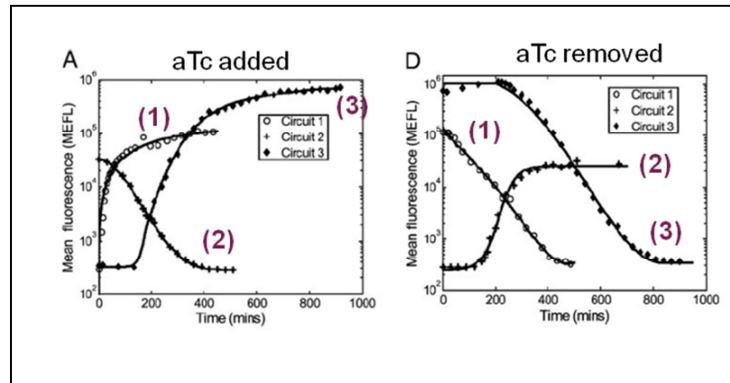
**Figure 20 Conséquences de la longueur des cascades de régulation.**

La longueur de la cascade détermine la dynamique, la sensibilité et la variabilité du réseau génétique.

Tiré de (Hooshangi et al., 2005)

Lorsque la concentration d'inducteur est faible ou élevée, la variabilité dans la population de cellules est limitée. En revanche, lorsque la concentration d'inducteur est intermédiaire et compte-tenu de la courbe de réponse, la sensibilité est maximale si bien que de petites variations vont se traduire par une grande hétérogénéité. De plus, ils montrent que plus la cascade comporte d'étapes, plus le bruit est amplifié (Figure 20, B). Blake et al. montrent dans un système similaire que lorsque le bruit augmente dans cette région de forte sensibilité, on observe une bimodalité de la distribution de la fluorescence dans la

population cellulaire (Blake et al., 2003). La structure du réseau a aussi pour particularité de modifier la dynamique de réponse à l'inducteur (Figure 21).



**Figure 21** Longueur des cascades de régulation et délai de réponse.

Plus la cascade contient de niveaux, plus le délai de réponse est important lorsque l'on ajoute ou retire l'inducteur (aTc) dans les 3 réseaux de la Figure 19.

Tiré de (Hooshangi et al., 2005)

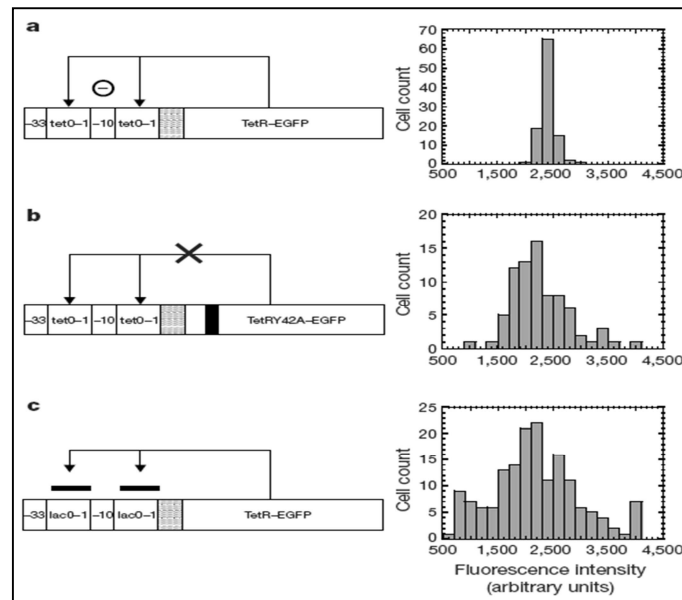
Plus la cascade est longue (ici de 1 à 3 niveaux), plus le délai de réponse est important. Ceci est valable lorsque l'on ajoute l'inducteur (Figure 21, A) mais aussi quand on le retire (Figure 21, B).

### II.2.b. Atténuation ou amplification du bruit par autorégulation

Il existe d'autres dynamiques que les cascades dans les réseaux de gènes. Souvent, l'expression d'un gène peut être régulée par son propre produit d'expression. On parle alors de rétroaction. Elle peut être positive ou négative, si l'expression du gène est augmentée ou réprimée. Ces deux structures couplées à la stochasticité de l'expression génique peuvent avoir des propriétés émergentes très intéressantes (Savageau, 1974a, 1974b; Thattai and van Oudenaarden, 2001).

Dans un contexte de réseau synthétique chez E.Coli, la rétroaction négative s'est révélée utile pour réduire la variabilité de l'expression d'un gène rapporteur (Becskei and Serrano, 2000; Dublanche et al., 2006). Dans ce système, la synthèse de TetR va réprimer sa propre expression grâce à la fixation des molécules TET sur des opérateurs en amont du promoteur (Figure 22, a). Lorsque la quantité de répresseur augmente, son expression

diminue et il se dégrade progressivement pour revenir vers la moyenne. Lorsque la quantité de répresseur diminue sous la moyenne, la répression diminue et l'expression augmente. Cette dynamique est intéressante pour réduire la variabilité mais peut créer en contrepartie des oscillations, à la manière d'un thermostat.



**Figure 22 Auto-régulation négative et réduction du bruit.**

*Tiré de (Becskei and Serrano, 2000)*

Privé de cet effet inhibiteur par mutation de la protéine TET ou par changement des opérateurs (de tetO en lacO, non-reconnu par TET), la rétroaction n'a pas lieu et la variabilité augmente dans la population (Figure 22, b et c). Les histogrammes de distribution de la fluorescence témoignent de la perte de la rétroaction négative et de l'augmentation de la variabilité.

La rétroaction positive provoque un comportement différent, notamment l'amplification du signal. En effet, une expression très faible du gène est suffisante pour induire une auto-activation et un renforcement de son expression. Ce genre de régulation peut donc fonctionner comme une bascule entre deux états distincts générant une bistabilité (Becskei et al., 2001; Isaacs et al., 2003; Ozbudak et al., 2004; Acar et al., 2005). Les conséquences de la stochasticité de l'expression génique peuvent être dramatiques dans le sens où ces fluctuations peuvent faire basculer le système d'un état off à un état on de façon aléatoire. Néanmoins, cette sensibilité peut être atténuée en présence de

coopérativité entre les activateurs (dimérisation ou augmentation du nombre de sites de fixation) qui va créer un seuil pour l'auto-activation (Hasty et al., 2000).

Ainsi, la nature des interactions mutuelles est susceptible de créer des propriétés inattendues, par exemple de multi-stabilité. Le passage d'un état stable à l'autre peut alors être la conséquence de fluctuations à l'échelle moléculaire impliquant dans certains cas une seule molécule et parfois, plusieurs molécules à travers des cascades de régulation (Choi et al., 2010).

### II.3. L'organisation cellulaire et le bruit structural

---

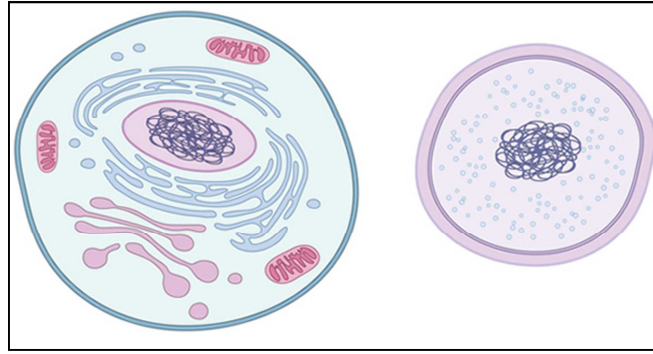
Le bruit moléculaire est une source importante de fluctuations intrinsèques qui est liée au nombre de partenaires mais aussi à leur dynamique, rendant les réactions biochimiques hautement probabilistes. Il existe une autre source de variabilité qui résulte de l'organisation des différentes structures cellulaires et subcellulaires, notamment celle du génome.

#### II.3.a. Organisation cellulaire

---

Contrairement aux procaryotes, les cellules eucaryotes sont des systèmes organisés et compartimentés par des membranes internes qui délimitent des organelles (Figure 23). Ces organelles accomplissent des fonctions particulières qui nécessitent un environnement biochimique interne différent du cytoplasme (réticulum endoplasmique, golgi, mitochondries, lysosomes, noyau,...). Cette différence d'environnement est exploitée pour produire un travail biochimique grâce à un gradient de molécules ou d'ions (mitochondries, lysosomes...).

En contrepartie, cette organisation nécessite de mettre en place des mécanismes de transport, souvent actifs, entre compartiments. Ils permettent de maintenir le gradient, d'importer des substrats ou d'exporter des produits et déchets biochimiques.



**Figure 23 Organisation cellulaire chez les eucaryotes et les procaryotes.**

Ces représentations très simplifiées montrent les différences d'organisation intracellulaire entre eucaryotes (à gauche) et procaryotes (à droite). La présence d'organelles chez les eucaryotes crée des compartiments bien identifiés et spécialisés.

Dans le cadre de l'expression génique chez les eucaryotes, les échanges sont essentiels. Le matériel génétique est confiné dans le noyau où a lieu la transcription alors que la traduction des ARNm se déroule à l'extérieur de ce compartiment. Une étape de transport des ARNm hors du noyau est donc nécessaire et peut ajouter de la variabilité en créant un délai entre la transcription et la traduction. Il en est de même pour le transport des substrats et enzymes nécessaires à la transcription (ARN polymérase, nucléotides, cofacteurs ioniques...).

Il est néanmoins difficile de quantifier précisément l'impact de cette étape sur la variabilité observée et seules les modélisations informatiques prennent en compte l'export des ARNm, la taille de la cellule ou la dynamique du noyau comme sources de variabilité (Coulon et al., 2010).

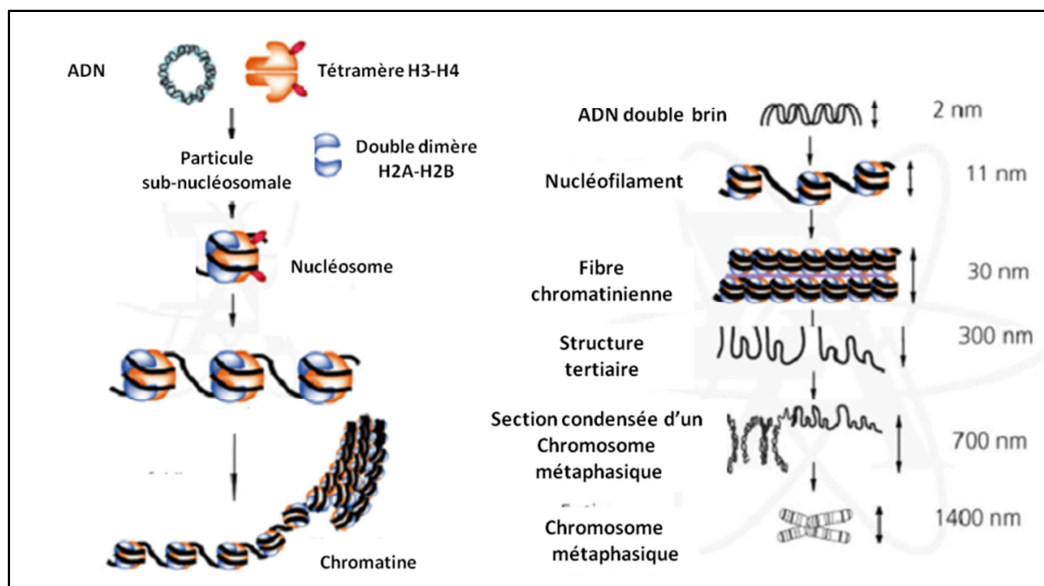
### *II.3.b. Organisation du génome*

---

Chez les eucaryotes, l'ADN est confiné dans le noyau de la cellule. Il est associé à diverses protéines pour former un complexe nucléoprotéique appelé chromatine. Celle-ci joue un rôle structural important en structurant en différents niveaux d'organisation les 2 mètres d'ADN que contient chaque cellule (humaine). Grâce à sa flexibilité, elle permet d'atteindre un facteur de compaction de près de 400 fois pendant l'interphase et jusqu'à 10 000 fois pendant la mitose pour faire tenir cette quantité d'ADN dans un volume nucléaire de moins de  $500\mu\text{m}^3$ .

### ◆ Structure de la chromatine

L'unité de base de la chromatine est le nucléosome, constitué de 4 types d'histones constitutifs (H2A, H2B, H3 et H4), petites protéines basiques, formant un octamère autour duquel s'enroulent environ 160 pb d'ADN (Figure 24). Cette structure est maintenue par des histones de type H1. Il a été montré que d'autres protéines de nature non-histone ainsi que des ARN pouvaient participer à la stabilité de ce complexe nucléoprotéique. Au cours du cycle cellulaire, le niveau de compaction de la chromatine évolue du nucléo-filament jusqu'au chromosome métaphasique hyper condensé (Figure 24, droite), notamment grâce à la présence d'histones H1 (Grigoryev et al., 2009).



**Figure 24 Organisation du génome eucaryote.**

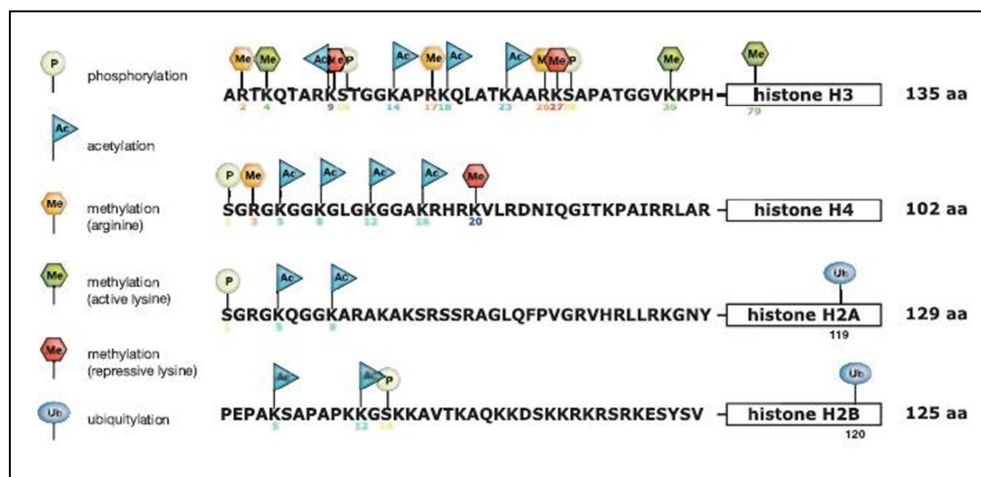
*Le nucléosome, unité de base de la chromatine, est constitué de dimères d'histones. L'ADN s'enroule autour de cette structure pour atteindre un premier niveau de compaction (à gauche). Des niveaux de compaction supérieurs peuvent être atteints, notamment au cours du cycle cellulaire (à droite).*

Les différents niveaux de compaction de la chromatine sont caractérisés par un profil de modifications post-traductionnelles au niveau des histones, mais aussi par des modifications covalentes de l'ADN. Ces changements modifient l'affinité de l'ADN pour les histones ainsi que le recrutement de facteurs de stabilisation et de remodelage de la chromatine, facilitant ainsi le passage d'un état condensé à un état moins condensé.

### ◆ Modifications épigénétiques

Les modifications covalentes de l'ADN et post-traductionnelles des histones sont dites « épigénétiques » par opposition aux mutations génétiques. Elles concernent uniquement les changements qui n'altèrent pas la séquence primaire de l'ADN. La paternité de ce terme revient à Conrad Waddington (1905-1975) qui définit alors l'épigénétique comme l'étude de l'interaction entre l'environnement et les gènes aboutissant au phénotype (1942). De nos jours, ce terme regroupe globalement les phénomènes liés à la régulation de l'expression génique (inactivation du chromosome X, empreinte parentale, miRNA, shRNA...).

On peut caractériser l'état d'activité d'une région chromatinienne par les modifications présentes au niveau des histones et de l'ADN de cette région. Les modifications des histones se produisent principalement sur leur partie N-terminale. Elles consistent en majorité à l'ajout de résidus acétyle, méthyle, phosphate... par des enzymes de type HAT ou HMT (Figure 25).



**Figure 25 Modifications épigénétiques au niveau des histones.**

La queue des histones (Lettres Majuscules) représente environ  $\frac{1}{4}$  de leur taille et accueille la plupart des modifications.  
Tiré de (Allis, 2007)

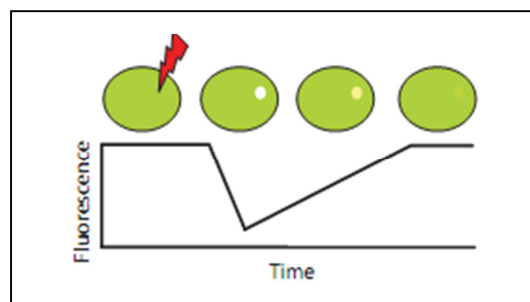
La présence de ces résidus modifie l'affinité entre l'ADN et les histones et en conséquence, le niveau de compaction de cette macro-structure. Le deuxième type de modification épigénétique est la méthylation de l'ADN. Elle se produit sur le carbone 5 de la cytosine

grâce à l'activité d'enzymes DNMT qui catalysent le transfert de groupements méthyles à partir de S-adénosyl-1-méthionine (SAM). Les cytosines ciblées par cette modification appartiennent à des clusters appelés « îlots CpG ». Ces îlots contiennent statistiquement une plus grande proportion de répétitions CG que le reste du génome et sont souvent situés à proximité de régions régulatrices de l'expression d'un gène (Saxonov et al., 2006). La méthylation a pour conséquence l'insertion des cytosines méthylées dans le grand sillon de l'ADN, provoquant ainsi un encombrement stérique et une compaction de la chromatine. La méthylation de l'ADN est donc considérée comme une caractéristique de régions transcriptionnellement inactives.

### II.3.c. Dynamique chromatinienne

Différentes études montrent que malgré son apparente stabilité, la chromatine possède une dynamique interne intense. La technique de référence utilisée pour étudier le temps de résidence de protéines (ou leur diffusion) est le FRAP.

Cette technique consiste à exciter une région de la cellule au moyen d'un laser de longueur d'onde spécifique. L'excitation laser prolongée va réduire la fluorescence de la protéine d'intérêt (couplée chimiquement avec un fluorophore, protéine de fusion...) dans la zone ciblée (Figure 26). En mesurant le temps nécessaire au retour de fluorescence dans cette zone, on peut estimer la dynamique de cette protéine. Plus le temps de résidence est élevé, moins les protéines vont diffuser et reconquérir la zone éteinte.



**Figure 26 Étude de la dynamique protéique par FRAP.**

*Une région de la cellule fluorescente est illuminée intensément pour réduire sa fluorescence. La vitesse du retour de la fluorescence traduit alors la vitesse de diffusion de la molécule fluorescente.*

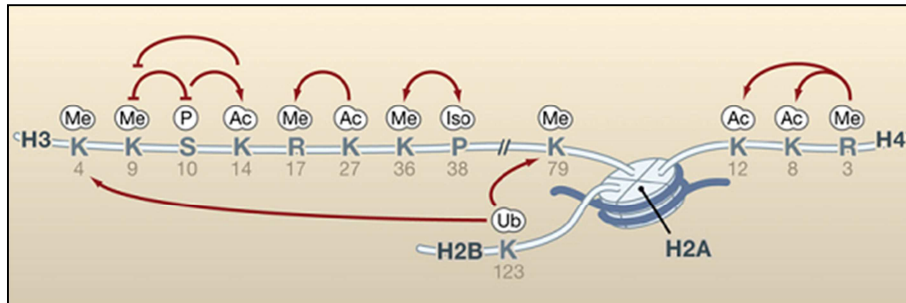


Cette technique a été utilisée pour étudier la dynamique des nombreuses protéines nucléaires. Le temps de résidence des facteurs de transcription, d'enzymes de remodelage de la chromatine ou de récepteur d'hormones (Eostrogen Receptor et Glucocorticoid Receptor) est très court, de l'ordre de quelques secondes seulement (McNally et al., 2000; Phair and Misteli, 2000; Stenoien et al., 2001; Phair et al., 2004). La même technique appliquée aux histones de base (H2A, H2B, H3 ou H4) montre que leur temps de résidence est supérieur à la plupart des autres protéines associées à l'ADN, pouvant atteindre quelques dizaines de minutes à 1 heure (Kimura and Cook, 2001). L'histone H1, qui se fixe à l'ADN entre les nucléosomes, possède une dynamique plus intense que les histones « cœur » (H2A, H2B, H3 et H4), avec un temps de résidence proche d'une minute (Lever et al., 2000; Misteli et al., 2000).

Ces expériences montrent que la chromatine est bien un élément instable et que son étude par des techniques statiques comme le ChIP ne sont que des aperçus temporaires de l'état chromatinien. La répartition des histones et leurs modifications ne peuvent être analysées que de façon statistique compte-tenu de leur nature éphémère. La dynamique de la chromatine est intimement liée à la nature des modifications épigénétiques qu'elle porte mais surtout à leur réversibilité. Toutes les modifications sont catalysées par des enzymes particulières mais il existe dans la cellule des enzymes qui catalysent les réactions inverses de façon quasi-immédiate (HAT / HDAC). La méthylation des histones et de l'ADN était considérée jusqu'à peu comme une modification permanente. L'identification d'enzymes capables de démétyler les histones (Jumonji domain-containing (JmjC) histone demethylases) a montré que ces modifications pouvaient aussi être réversibles (Tsukada et al., 2005).

En ce qui concerne l'ADN, seule la démétylation passive lors de la réplication de l'ADN était connue. Il a été démontré récemment dans des cellules quiescentes, que le niveau de méthylation de l'ADN résulte d'un équilibre entre méthylation et démétylation (active et passive) alors qu'elle était considérée jusqu'alors comme non réversible (Yamagata et al., 2012).

Les modifications épigénétiques sont donc réversibles et ont un turnover important si bien que le statut épigénétique d'une histone donnée est très dynamique. De plus, il existe une coopérativité entre les différentes modifications (Figure 27).



**Figure 27** *Coopérativité des modifications épigénétiques.*

*La modification de certains résidus sur la queue des histones peut favoriser la modification des résidus adjacents.*

Certaines modifications sont utilisées pour le recrutement de protéines adaptatrices (Methyl Binding Proteins, Heterochromatin Protein 1) qui vont à leur tour recruter des enzymes pour renforcer les modifications existantes ou en ajouter de nouvelles. Ces nouvelles modifications vont accroître la stabilité d'une histone ou participer à sa déstabilisation.

Des études menées par spectrométrie de masse sur des histones isotypiquement marquées ont permis de mesurer le turnover des différents histones de base ainsi que des variantes d'histones, en fonction des modifications épigénétiques qu'elles portent (Zee et al., 2010) (Tableau 2).

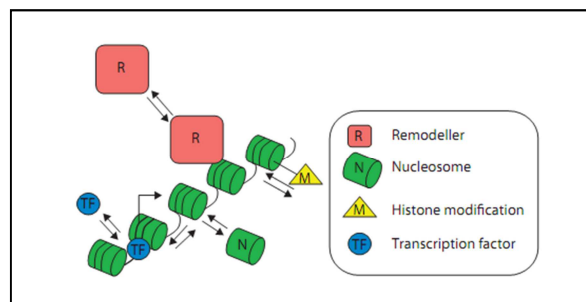
D'une façon générale, seule l'histone H2A présente un turnover plus important que les autres. La différence majeure dans le turnover des autres histones réside dans la nature, la position et le nombre de modifications épigénétiques qu'elles portent (Tableau 2). Plus une histone est méthylée, plus le turnover diminue, indiquant une meilleure stabilité (par exemple H3K9). De même, plus l'histone est acétylée, plus son turnover semble augmenter (H4K5), indiquant une plus grande instabilité.

Modification	Turnover	Rôle activateur / répresseur (référence)
H3K4me1	0.7326	(Barski et al., 2007)
H3K9me1	0.7029	(Barski et al., 2007)
H3K9me2	0.5840	(Barski et al., 2007)
H3K9me3	0.4104	(Barski et al., 2007)
H3K9/K14ac1	1.1816	(Wang et al., 2009)
H3K18/K23ac1	1.2919	(Wang et al., 2009)
H3K18ac1/K23ac1	1.6818	(Wang et al., 2009)
H3K27me1	0.5929	(Barski et al., 2007)
H3K27me2	0.4272	(Barski et al., 2007)
H3K27me3	0.2679	(Barski et al., 2007)
H3K36me1	0.8803	(Barski et al., 2007)
H3K36me2	0.5669	(Rao et al., 2005)
H3K56ac1	3.8155	(Wang et al., 2009)
H3K79me1	0.6671	(Steger et al., 2008)
H3K79me2	0.5661	(Steger et al., 2008)
H4K5/K8/K12/K16ac1	1.1968	(Wang et al., 2009)
H4K5/K8/K12/K16ac2	1.5118	(Wang et al., 2009)
H4K5/K8/K12/K16ac3	1.6048	(Wang et al., 2009)
H4K5/K8/K12/K16ac4	1.5483	(Wang et al., 2009)
H4K20me1	0.5884	(Barski et al., 2007)
H4K20me2	0.2283	(Trojer et al., 2007)
H4K20me3	0.1459	(Schotta et al., 2004)

**Tableau 2 Modifications épigénétiques et turnover des histones**  
 Le turnover est normalisé par rapport à la forme non modifiée de l'histone.  
 Tiré de (Zee et al., 2010).

De façon intéressante, les marques épigénétiques considérées comme activatrices de l'expression génique (en bleu dans le tableau) sont retrouvées sur les histones ayant un turnover important, contrairement aux marques inhibitrices (en rouge). Cette observation permet de répondre à une des questions fondamentales qui se posent quand on étudie l'expression d'un gène. Plus que la nature de son promoteur, la présence de séquences régulatrices ou le nombre de facteurs de transcription, c'est l'accessibilité de ces séquences aux facteurs de transcription qui est primordiale. Il est souvent fait mention de la spécificité de certains facteurs de transcription, mais comment atteindre une séquence spécifique d'ADN lorsqu'elle est inaccessible ?

On peut aisément imaginer que l'instabilité des interactions histone/ADN ainsi que celle des protéines s'y fixant va potentiellement augmenter la probabilité d'expression d'un gène en facilitant temporairement et spontanément l'accès à l'ADN (Li et al., 2005) (Figure 28). Cette hypothèse de décompaction spontanée est soutenue par le fait qu'il existe dans chaque cellule, des épisodes de transcription que l'on qualifie « d'illégitimes ». Ces événements concernent tous les gènes, particulièrement ceux qui ne sont *a priori* pas liés au type cellulaire considéré, d'où leur caractère illégitime (Chelly et al., 1988, 1989; Fonknechten et al., 1992).



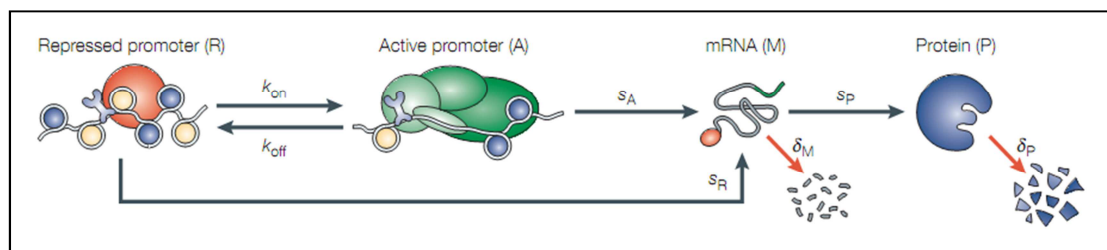
**Figure 28 Équilibre dynamique de la chromatine.**  
Tiré de (Deal and Henikoff, 2010)

L'équilibre entre les différentes marques épigénétiques va donc conditionner l'état d'activité transcriptionnelle au locus concerné. En fonction de sa localisation sur le génome, une séquence d'ADN peut donc se trouver dans une région dynamique ou dans une région stable. La dynamique de la chromatine joue par conséquent un rôle important dans la régulation de l'expression génique mais aussi dans tous les processus liés à l'ADN, comme sa réplication et sa réparation.

#### *II.3.d. La cellule stochastique.*

Grâce à l'accumulation de données sur l'aspect quantitatif et spatial des réactions biochimiques, notre perception du fonctionnement cellulaire évolue. La linéarité et le déterminisme des processus biologiques laissent place à une vision dynamique régie par les probabilités et le hasard. Nous avons désormais un aperçu différent de l'expression génique. Ce processus, essentiel au fonctionnement cellulaire et à leur survie, a évolué depuis la proclamation du dogme de la biologie moléculaire.

De nombreux aspects quantitatifs, qualitatifs, spatiaux et temporels de ce processus ont été révélés. Nous savons maintenant que l'expression d'un gène est conditionnée par son accessibilité à la machinerie de transcription, elle-même dépendante de réactions complexes permettant le remodelage de la chromatine. Elle est aussi régulée par la mobilité des différents partenaires du complexe de transcription, qu'il s'agisse de simple diffusion passive ou de transport actif entre compartiments cellulaires et subcellulaires (Smolen et al., 1999). La probabilité d'assemblage d'un complexe de transcription est alors influencée par la diffusion des partenaires, leur disponibilité et leur nombre (Greive and von Hippel, 2005). L'expression génique étant un processus multi-étapes, on imagine que chacune d'elles est soumise à ces contraintes spatiales et temporelles (Figure 29). Ainsi, l'absence d'un des facteurs, au bon moment et au bon endroit, est suffisante pour perturber tout le processus.



**Figure 29 L'expression génique est un processus multi-étape.**  
tiré de (Kaern et al., 2005)

Il a d'ailleurs été montré que l'assemblage de la machinerie de transcription était un processus très inefficace, pouvant durer plusieurs minutes. De même, l'élongation n'est pas continue mais parsemée de pauses de 10 à 20 secondes (Herbert et al., 2006). Ces pauses peuvent néanmoins avoir un rôle régulateur (Landick, 2006; Rajala et al., 2010) dans l'expression génique mais pourraient contribuer de façon significative à sa variabilité (Voliotis et al., 2008; Dobrzyński and Bruggeman, 2009), au même titre que l'initiation de la transcription et de la traduction (Bai et al., 2006).

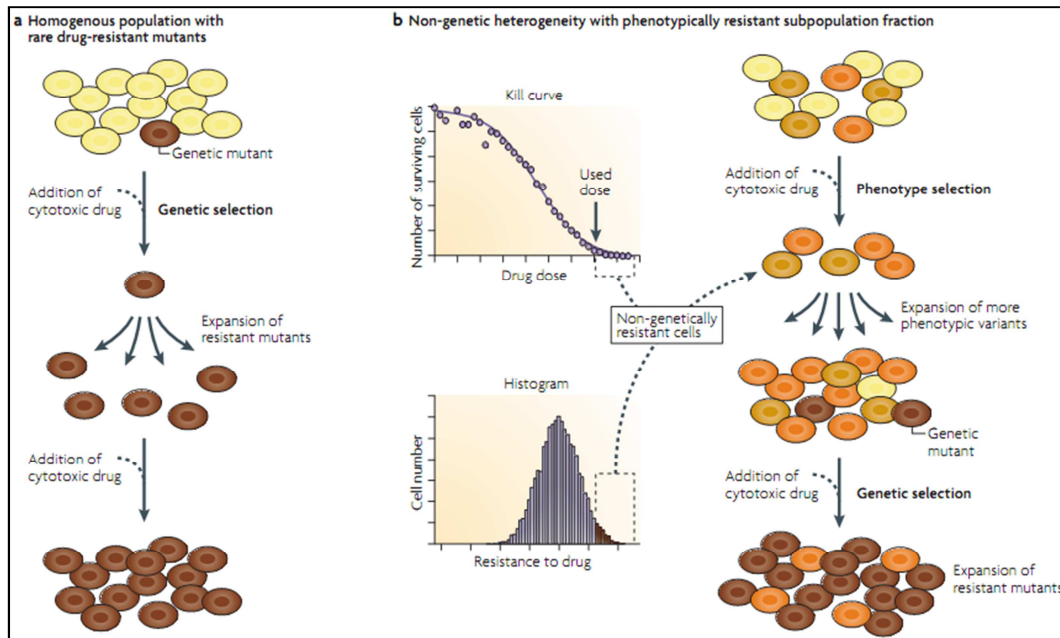
### III. Conséquences de la stochasticité de l'expression génique

---

Jusqu'alors considérée comme un bruit de fond, la variabilité semble à la vue des dernières avancées, jouer un rôle biologique important. Les études de Bar-Even et al. et Newman et al. avec la levure ont montré que certains gènes semblent optimisés pour fonctionner avec un certain niveau du bruit tandis que d'autres sont hautement régulés (Bar-Even et al., 2006; Newman et al., 2006). La plupart des gènes dont l'expression est considérée comme stable fait partie d'un groupe de gènes essentiels au fonctionnement cellulaire alors que les gènes décrits comme plus variables que la moyenne sont impliqués dans des mécanismes de réponse au stress. De même, les gènes exprimés de façon constitutive semblent moins variables que ceux qui sont transcriptionnellement régulés (Zenklusen et al., 2008). Considérée comme nuisible pour la cellule, la variabilité phénotypique peut fournir à la cellule, dans certains contextes, la flexibilité nécessaire pour s'adapter à un environnement changeant. Certains chercheurs avancent donc l'idée selon laquelle la propriété « variable » de certains gènes a été sélectionnée au cours de l'évolution pour répondre à des contraintes environnementales particulières (Stern et al., 2007; Silander et al., 2012).

Il a été démontré par modélisation qu'une population dont le phénotype est variable possède un taux de croissance significativement supérieur à celui d'une population homogène (Thattai and van Oudenaarden, 2004). La variabilité à l'échelle de la cellule individuelle augmente la probabilité de survivre dans un environnement fluctuant en proposant à l'échelle de la population, différentes « solutions » biologiques potentielles. Cette stratégie de survie est particulièrement importante pour des microorganismes comme les bactéries. Nous avons vu précédemment l'exemple des bactéries entrant en compétence en cas de stress ou les « persistants » qui survivent suite au traitement antibiotique (Bigger, 1944; Balaban et al., 2004). Néanmoins, d'un point de vue médical, cette variabilité peut être une limite à l'utilisation d'antibiotiques compte-tenu des individus phénotypiquement sélectionnés. La même interrogation peut être formulée vis-à-vis des cellules cancéreuses qui sont extrêmement hétérogènes (génétiquement et phénotypiquement) (Figure 30). La variation de l'expression d'une molécule quelconque

(canal MDR par exemple) peut générer aléatoirement dans la population de cellules cancéreuses, des cellules plus résistantes aux chimiothérapies et limiter l'action de ces composés chimiques (Niepel et al., 2009; Sharma et al., 2010).



**Figure 30 Hétérogénéité phénotypique et sélection clonale**

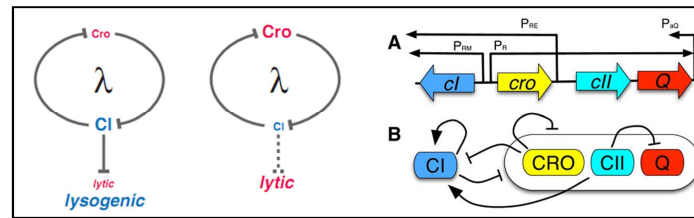
Tiré de (Brock et al., 2009)

Contrairement aux mutations génétiques, les phénotypes obtenus par l'hétérogénéisation de la population puis sélectionnés sont réversibles. La variabilité de l'expression génique peut aussi être impliquée dans des processus de décision à l'échelle de la cellule individuelle. Il existe des réseaux géniques, notamment les boucles de rétrocontrôle, qui peuvent faire émerger de la bistabilité. L'équilibre entre les deux états peut être perturbé par des fluctuations aléatoires dans l'abondance de l'un des partenaires moléculaires. Il existe de nombreux exemples notamment en microbiologie et en virologie.

### III.1. Lysogénie et cycle lytique chez le phage lambda

L'exemple le plus connu est celui du cycle lytique du phage lambda (Lieb, 1953; Arkin et al., 1998; Tian and Burrage, 2004; St-Pierre and Endy, 2008; Joh and Weitz, 2011). Dans une bactérie infectée par ce phage, deux états sont possibles : la lyse et la lysogénie (

Figure 31). La première correspond à une production active de nouveaux virus avec pour conséquence la lyse de la cellule infectée et la seconde à un état de latence du provirus avec maintien de l'intégrité de la cellule jusqu'à l'activation du cycle lytique.



**Figure 31** *Rétro-action négative, fluctuations et cycle infectieux du phage lambda.*  
Tiré de (Joh and Weitz, 2011)

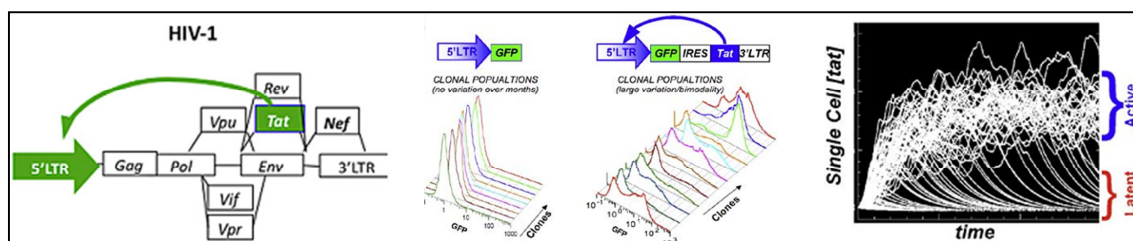
L'équilibre entre lyse et lysogénie repose sur l'interaction de partenaires dont CRO et CI sous forme d'une inhibition mutuelle. Ces deux partenaires s'auto-inhibent générant une bistabilité, qui se traduit soit par l'entrée en phase lytique soit par la lysogénie. Des fluctuations dans l'abondance de l'un ou l'autre des partenaires pouvant modifier le devenir du système.

### III.2. Virulence et latence chez le VIH-1

En virologie, on retrouve cet équilibre entre phase de latence et de virulence chez le virus VIH. Le promoteur LTR de HIV-1 est une source de variabilité, principalement en conséquence de la dynamique sporadique de la transcription à partir de ce promoteur (2 à 10 ARNm par période d'activité) (Weinberger et al., 2005, 2008a; Singh and Weinberger, 2009; Singh et al., 2010; Razooky and Weinberger, 2011). En conséquence, l'abondance des produits des gènes sous le contrôle de ce promoteur est variable. Ceci est particulièrement important pour la protéine TAT qui est un trans-activateur du promoteur LTR (Figure 32, gauche). Elle exerce un rétrocontrôle positif et renforce l'activité du promoteur LTR. De cette structure de réseau peut émerger de la bistabilité, comme décrit précédemment. Cette dynamique a été mise en évidence dans un modèle utilisant un gène rapporteur (GFP) sous le contrôle du promoteur LTR. En absence du gène Tat, la distribution de la fluorescence dans des populations isogéniques est monomodale (Figure 32, gauche). Il n'y a pas de rétrocontrôle et la variance est liée à la variabilité intrinsèque



de l'expression génique. En présence de Tat, on note l'apparition d'une bi-modalité dans la distribution de la fluorescence de la GFP dans des populations isogéniques. L'importance de chaque mode est différente entre les clones et semble dépendre du site d'intégration du transgène (Figure 32, centre).



**Figure 32** Boucle de rétroaction positive chez le VIH.

La protéine TAT est un trans-activateur du LTR. Une variabilité de son expression peut générer un état bistable qui conditionne l'évolution de l'infection entre phase de latence et de virulence.

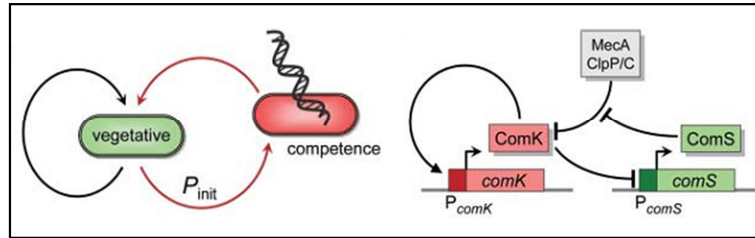
Tiré de (Singh and Weinberger, 2009)

A l'échelle cellulaire, lorsque la concentration de TAT dans la cellule infectée est faible, le virus reste sous forme latente car les autres gènes viraux dépendants du LTR ne sont pas exprimés. A forte concentration, le virus devient virulent et répliatif (Figure 32, droite) car tous les gènes nécessaires sont exprimés.

Le système est ainsi très sensible lorsque la concentration intracellulaire de TAT est intermédiaire, de petites fluctuations étant suffisantes pour faire basculer l'équilibre vers la latence ou la virulence.

### III.3. Entrée en compétence chez *Bacillus subtilis*

La compétence chez la bactérie *Bacillus subtilis*, est une stratégie dont nous avons vu précédemment l'intérêt biologique pour la survie. L'induction de la compétence repose sur l'expression d'une centaine de gènes dont ComK (Maamar and Dubnau, 2005; Maamar et al., 2007; Suel et al., 2007). La protéine ComK exerce un rétrocontrôle positif sur son promoteur ce qui induit comme précédemment, une bistabilité.



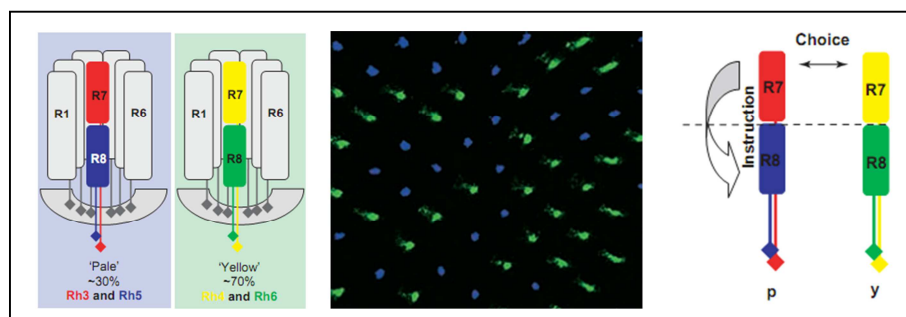
**Figure 33** *Rétrocontrôle, initiation et maintenance de la compétence chez B.Subtilis.*  
L'entrée en compétence dépend de la rétroaction de ComK sur sa propre expression.

Il existe de nombreux autres exemples biologiques d'équilibre bistable où la stochasticité de l'expression génique semble impliquée pour provoquer le passage d'un état à un autre (pour review : (Dubnau and Losick, 2006; Losick and Desplan, 2008; Fraser and Kaern, 2009).

Il a aussi été proposé que la stochasticité de l'expression génique puisse être impliquée dans des mécanismes que l'on pensait hautement régulés tels que la différenciation cellulaire, l'auto renouvellement, le cycle cellulaire, la prolifération cellulaire, l'apoptose mais aussi à l'échelle de l'organisme dans le développement (Paldi, 2003; Spencer et al., 2009; Stockholm et al., 2010; Skommer et al., 2011).

#### III.4. Différenciation cellulaire chez la drosophile

L'exemple le plus connu de différenciation stochastique est celui des yeux de drosophile dans lesquels on trouve une distribution aléatoire des photo-récepteurs. Chaque œil est composé d'environ 800 ommatidies qui comptent 8 photo-récepteurs chacune (R1 à R8). R7 et R8 sont entourés des 6 autres photo-récepteurs, R7 étant situé au-dessus de R8.



**Figure 34** *Stochasticité de l'expression génique et devenir cellulaire chez la drosophile.*

On distingue deux types de récepteurs (pale et yellow) selon la rhodopsine exprimée par le récepteur R7. Le choix entre Rh3 et Rh4 semble aléatoire. La rhodopsine exprimée par le récepteur R8 sous-jacent semble quand à elle dépendre de celle exprimée par R7.

Ces photo-récepteurs expriment des rhodopsines de différents types permettant une perception de différentes longueurs d'onde. R7 exprime Rh3 ou Rh4 et R8 exprime Rh5 ou Rh6. On distingue donc deux types de photo-récepteurs, des « pâles » et des « jaunes » (Figure 34, gauche et centre).

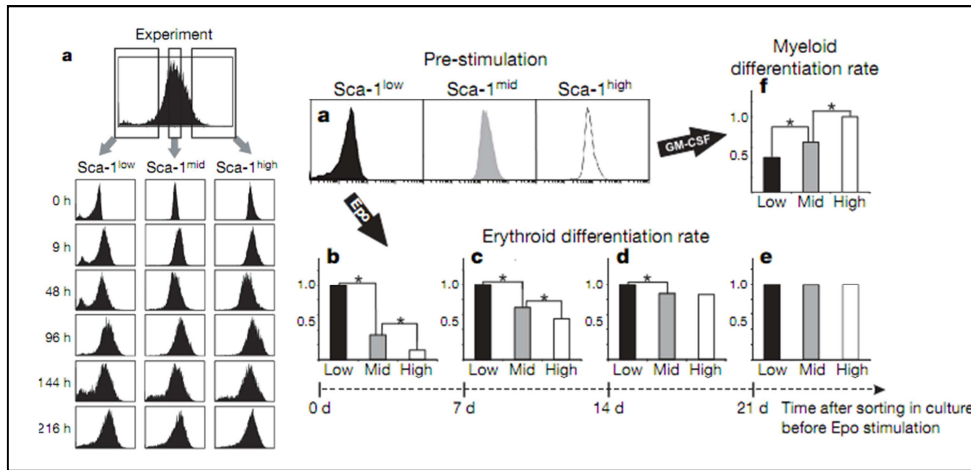
Il semble que le choix entre Rh3 et Rh4 dans le récepteur R7 soit stochastique. En revanche, l'expression de Rh5 ou Rh6 dans le récepteur R8 est couplé avec celle du récepteur R7 : Rh5 est toujours exprimé avec Rh3 et Rh6 avec Rh4 (Figure 34, droite). De façon surprenante, la proportion de récepteurs « pâles » ou « jaunes » est biaisée (30%/70%) mais est une constante entre plusieurs espèces (Wernet et al., 2006).

### III.5. Différenciation lors de l'hématopoïèse

---

L'hématopoïèse est un processus de différenciation bien connu en biologie qui génère différents types de cellules effectrices à partir d'une cellule souche hématopoïétique. Il a été proposé que la différenciation au cours de ce processus résulte d'évènements stochastiques contrairement à la vision linéaire classique de ce processus (Enver and Greaves, 1998; Enver et al., 1998; Hume, 2000).

En 2008, Chang et al. étudient la variabilité du marqueur Sca-1 (Stem Cell Antigen 1) dans des cellules souches hématopoïétiques de souris (Chang et al., 2008). En triant des populations de cellules exprimant différents niveaux de Sca-1, ils notent que des populations de cellules triées sur leur niveau d'expression de Sca-1 (faible, intermédiaire ou élevé) sont capables de reconstituer la population initiale. Les extrêmes sont en quelque sorte des états instables qui tendent à revenir vers la moyenne avec une certaine inertie (Figure 35, gauche). Néanmoins, ces états extrêmes semblent préparer la cellule à la différenciation (« priming »). En effet, ils montrent que le niveau d'expression de ce marqueur corrèle bien avec la probabilité de différenciation dans la lignée myéloïde ou érythrocytaire (Figure 35, droite).

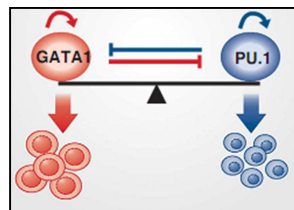


**Figure 35 Variabilité phénotypique et différenciation cellulaire hématopoïétique.**

Une population triée sur un niveau d'expression est capable de régénérer la population d'origine. Le niveau d'expression de Sca-1 est pronostique du potentiel de différenciation des cellules dans la lignée lymphoïde ou myéloïde.

Tiré de (Chang et al., 2008).

Le choix entre la lignée érythrocytaire et myéloïde ne repose pas uniquement sur l'expression de Sca-1 mais sur un équilibre entre deux autres facteurs, GATA1 et Pu.1 (GATA2).



**Figure 36 Rétroaction et différenciation des cellules souches hématopoïétiques.**

La différenciation dans la lignée lymphoïde ou myéloïde repose sur un équilibre de répression réciproque entre GATA1 et Pu.1.

Le premier favorise la lignée érythrocytaire et est plus abondant dans les cellules exprimant faiblement Sca-1, le second favorise la lignée myéloïde et est exprimé de façon accrue dans les cellules exprimant Sca-1 de façon abondante.

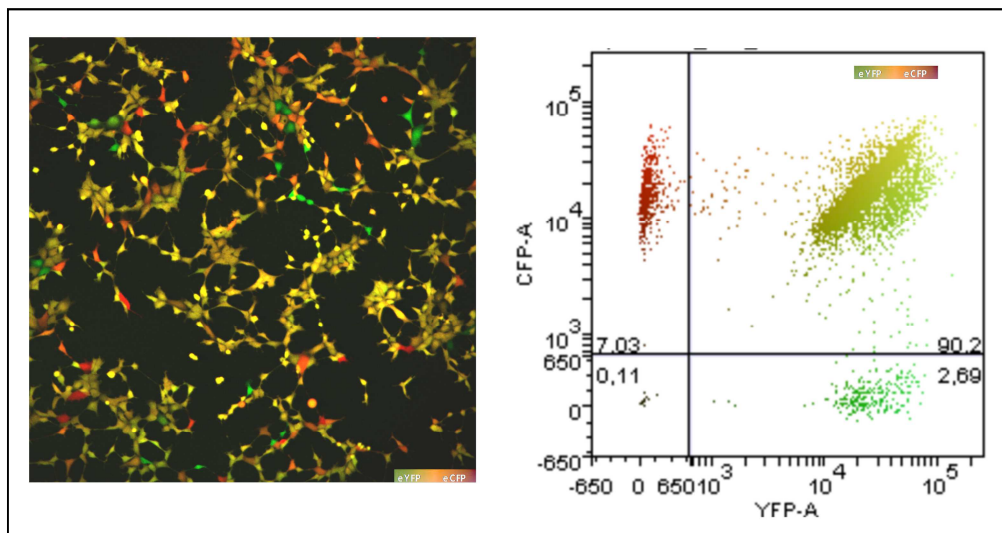
D'un côté, nous imaginons aisément le rôle positif d'une telle variabilité pour l'adaptation à l'environnement ou à des mutations génétiques. D'un autre côté, un paradoxe apparaît entre la présence indéniable du bruit dans certains circuits géniques et la robustesse nécessaire à certains processus biologiques (Johnston and Desplan, 2010). Cette dernière semble absolument nécessaire pour le développement d'organismes multicellulaires qui dépend de gradients et dont l'aspect quantitatif est absolument primordial. Or ces

organismes maintiennent un plan de développement malgré les différentes sources de variabilité, fait particulièrement intéressant. Quelques années auparavant, une nouvelle théorie dite de l'onto-phylogenèse a été formulée. Cette théorie, centrée sur l'individu, minimise grandement l'importance du programme génétique et de la spécificité des interactions pour laisser place à la variabilité dans le développement (Kupiec, 1983, 1996, 1997). Nous ne détaillerons pas cette théorie dans ce travail mais gardons à l'esprit qu'elle fait partie du changement de paradigme que nous observons aujourd'hui en biologie cellulaire et moléculaire.

## Objectif

---

L'objectif de ce travail est d'évaluer comment apparaît l'hétérogénéité phénotypique dans une population clonale et d'en extraire des hypothèses quant à son rôle biologique. En 2007, le laboratoire a développé un modèle cellulaire dérivé de ceux décrits en introduction pour étudier la stochasticité de l'expression génique (Neildez-Nguyen et al., 2008). Ce modèle cellulaire eucaryote, basé sur l'utilisation de deux gènes rapporteurs quasiment identiques, a clairement mis en évidence l'hétérogénéité phénotypique pouvant apparaître dans une population clonale cultivée dans des conditions homogènes (Figure 37). Dans les premiers temps suivant l'établissement de clones, les populations isogéniques obtenues présentaient une grande diversité de niveau d'expression des deux transgènes à tel point que certaines cellules n'exprimaient plus aucun des transgènes. Certaines en exprimaient un seul, d'autres les deux mais à des niveaux très différents, l'étendue pouvant atteindre 3 log de fluorescence.

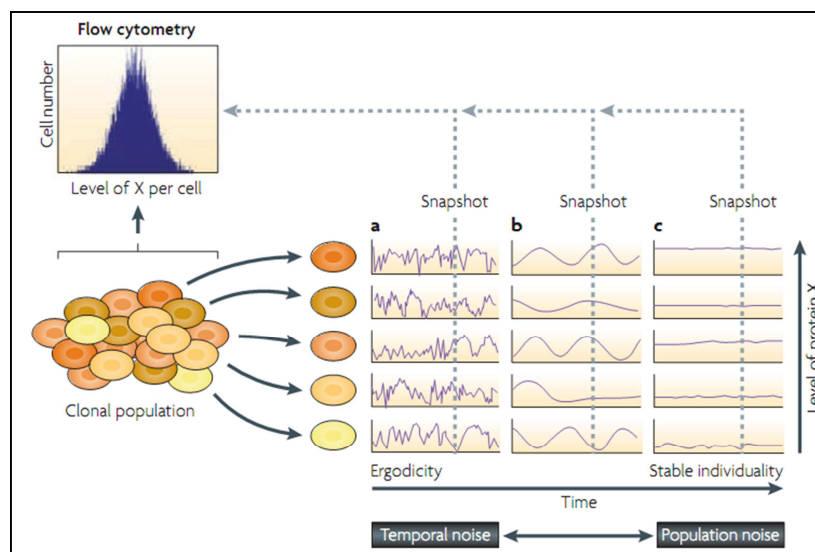


**Figure 37 Illustration de l'hétérogénéité phénotypique dans notre modèle.**

*La majorité des cellules exprime les deux transgènes mais des sous-populations n'exprimant pas un/les transgène(s) sont visibles. Microscopie confocale 10x et cytométrie en flux.*

Les observations effectuées lors de cette étude sont le point de départ de ce travail. La diversité dans les populations clonales est maintenant bien établie. En revanche, la dynamique qui permet d'y aboutir n'est que peu étudiée. Les études menées dans ce

domaine se cantonnent généralement à l'analyse de distributions obtenues par cytométrie en flux, en faisant l'hypothèse forte d'ergodicité dans le système (Huang, 2009). Les fluctuations sont alors considérées comme suffisamment rapides pour que le système explore les différents états possibles en un laps de temps court. La moyenne temporelle d'un individu est alors équivalente à la moyenne de la population (Figure 38). Les fluctuations observées chez les eucaryotes sont de nature plutôt lente et il est dangereux de considérer la variabilité du profil obtenu par cytométrie en flux sur une population comme l'unique reflet de la stochasticité de l'expression génique (Huang, 2009).



**Figure 38** *Différentes dynamiques cellulaires peuvent expliquer l'hétérogénéité. L'hétérogénéité peut être quantifiée par cytométrie en flux ou par analyse temporelle de cellules isolées. Tiré de (Brock et al., 2009)*

Nous avons donc profité du modèle cellulaire développé au laboratoire et déjà caractérisé sur de nombreux points, pour étudier l'émergence de l'hétérogénéité phénotypique. L'objectif est de comprendre comment s'établit la diversité au sein d'une population clonale à travers ce modèle.

Le phénotype étudié sera l'expression de chaque transgène via la fluorescence de leur produit, dans des cellules individuelles. Plus que le phénotype à proprement parler, ce modèle va nous permettre d'analyser l'expression génique et ses fluctuations à l'origine de la variabilité du phénotype. Nous examinerons donc à quelle vitesse les différentes sous-populations apparaissent, si leur phénotype est stable ou dans quelles mesures il est réversible.

Deux approches ont été privilégiées pour couvrir différentes résolutions temporelles. Une approche à long terme, à l'échelle de plusieurs générations de cellules, permettra d'étudier globalement la génération d'une population clonale. Une seconde approche, à l'échelle d'un cycle cellulaire, permettra d'analyser des individus isolés d'une population. Ces deux approches reposent sur l'analyse de cellules isolées.





## Matériel et méthodes

---

## I. Matériel biologique

---

### I.1. Lignée cellulaire HER 911

---

Les modèles cellulaires utilisés dans tous ces travaux ont été obtenus à partir de la lignée HER 911 provenant d'un Rétinoblastome Embryonnaire Humain (HER) (Fallaux et al., 1996). Disponible à la banque de cellules de l'ETGC au laboratoire, elle a été développée à l'origine pour remplacer la lignée 293, dérivée de cellules HEK, pour la production d'adénovirus recombinants (rAdV). Les cellules HER 911 ont des propriétés proches des cellules 293 : elles sont facilement transfectables, se cultivent facilement dans un milieu complet simple et ont la même fréquence de recombinaison homologue.

### I.2. Modèle bi-fluorescent eCFP/eYFP

---

Comme décrit en introduction, l'utilisation de deux gènes rapporteurs est un outil de choix pour identifier la source des fluctuations à l'échelle de la cellule unique (Intrinsèque vs Extrinsèque). Un premier modèle a été produit en 2007 dans le laboratoire, en transfectant successivement la lignée HER 911 avec deux constructions plasmidiques (Neildez-Nguyen et al., 2008). Une première transfection a permis d'intégrer le premier gène rapporteur, l'eCFP. Après sélection des cellules fluorescentes grâce à un trieur de cellules, le deuxième transgène, l'eYFP, a été introduit par la même technique. Des cellules exprimant les deux constructions ont par la suite été clonées et amplifiées pour générer des clones. Une analyse par PCR quantitative a permis de sélectionner les clones ne possédant qu'une seule copie de chaque transgène. Ces clones seront nommés « clones parentaux » par la suite et constituent le matériel biologique de la première partie de ce travail.

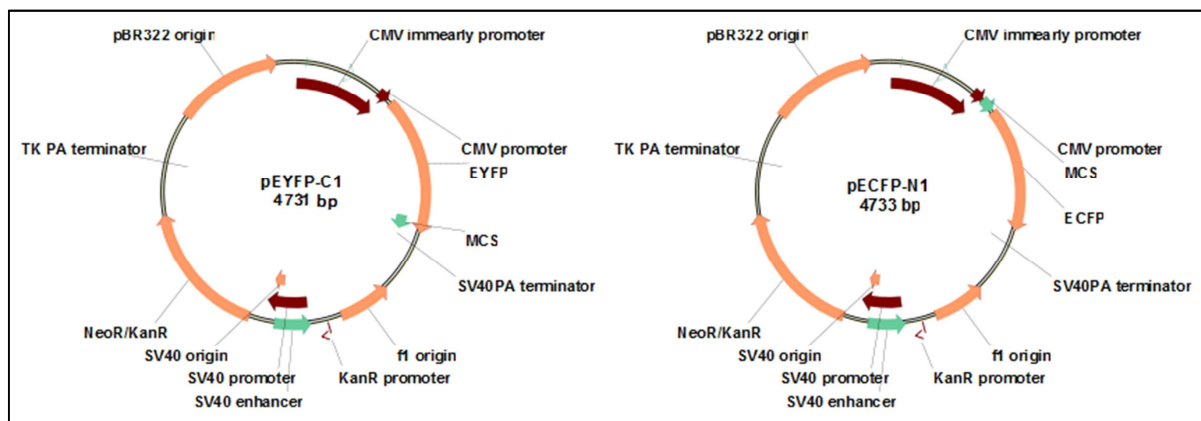
### I.3. Constructions plasmidiques et propriété spectrales protéiques

---

Plusieurs constructions plasmidiques ont été utilisées afin de construire nos modèles de cellules fluorescentes. Leurs propriétés sont résumées ci-dessous.

### 1.3.a. Plasmides commerciaux non-modifiés

peCFP-N1 et peYFP-C1 sont deux plasmides commerciaux (ClonTech, cat.#6005-1 et cat.#6900-1) qui codent respectivement pour des variants jaune (eYFP) et bleu (eCFP) de la GFP (protéine fluorescente verte) (Figure 39). Les deux transgènes sont sous le contrôle du promoteur viral CMV et seules 16 paires de bases sont différentes entre les deux séquences codantes. Les plasmides contiennent également une origine de réplication SV40 et un gène de résistance à la néomycine pour une sélection des cellules transfectées. Un gène de résistance à la kanamycine sous le contrôle d'un promoteur bactérien permet de faire de la sélection chez *E. coli*, pour la production des plasmides. Les deux plasmides contiennent aussi une origine de réplication provenant du plasmide pUC19 pour la réplication chez *E. coli* et une origine f1 pour la production d'ADN simple brin.



**Figure 39** Carte des plasmides eCFP et eYFP sauvages.

*Ces plasmides ont été utilisés pour la construction d'un modèle bi-fluorescent par transfections successives.*

EYFP et eCFP contiennent plus de 190 mutations silencieuses qui créent un cadre ouvert de lecture, composé principalement de codons optimisés pour l'humain. De plus, la séquence en amont de l'eYFP et de l'eCFP a été convertie en une séquence consensus d'initiation de transcription de Kovac. Ces modifications augmentent l'efficacité de traduction de l'ARNm et en conséquence l'expression de la eYFP et eCFP dans les cellules de mammifères.

Nous considérons pour la suite ces constructions et les cellules transfectées par ces constructions comme le modèle « **stable** ».

◆ *Plasmide peYFP-C1*

---

La séquence codante du gène rapporteur eYFP contient 4 substitutions d'acides aminés par rapport à la GFP sauvage : Ser-65 par Gly; Val-68 par Leu; Ser-72 par Ala; et Thr-203 par Tyr. Comme il s'agit d'un plasmide de clonage, un site de multi-clonage (MCS) se trouve entre la séquence codante de l'eYFP et le codon stop. Il est ainsi possible de créer une protéine de fusion marquée en N-term.

◆ *Plasmide peCFP-N1*

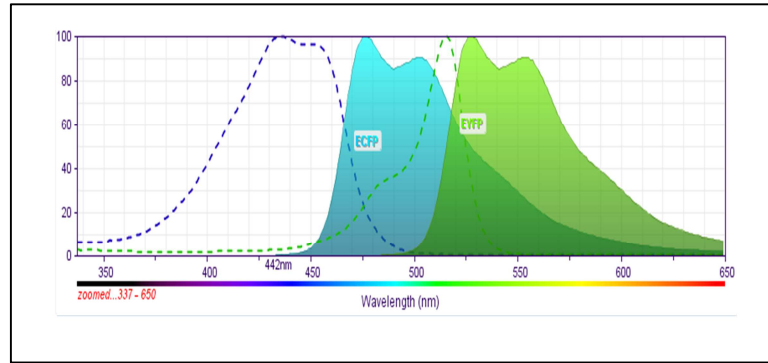
---

La séquence codante d'eCFP contient 6 substitutions d'acides aminés par rapport à la GFP. La substitution de la Tyr-66 par Trp apporte le changement majeur et donne à cette protéine des propriétés semblables aux autres variants cyan. Les 5 autres substitutions (Phe-64 par Leu; Ser-65 par Thr; Asn-146 par Ile; Met-153 par Thr; et Val-163 par Ala) augmentent la brillance et la solubilité de la protéine en améliorant son repliement tridimensionnel. Contrairement à l'eYFP, le site de multi-clonage se trouve entre le promoteur précoce CMV et le début de la séquence codant pour l'eCFP. Cette configuration permet de produire une protéine de fusion marquée en C-term par l'eCFP.

◆ *Propriétés spectrales*

---

Malgré leurs séquences très similaires, les protéines eCFP et eYFP ont des caractéristiques de fluorescence relativement distinctes (Figure 40). Ce couple de protéines est souvent utilisé pour des analyses de FRET compte-tenu du recouvrement partiel des spectres d'émission et d'absorption d'eCFP et eYFP respectivement.



**Figure 40 Spectres d'absorption et d'émission des protéines eCFP et eYFP.**

Leurs pics d'excitation se situent respectivement à 433-453nm et 513nm. Le maximum d'émission est atteint entre 475 et 500nm pour l'eCFP et entre 525 et 560nm pour l'eYFP.

Le spectre d'excitation de ces deux protéines est très différent ce qui permet de les exciter spécifiquement avec une source lumineuse adaptée. De même, des filtres d'émission adaptés permettent d'utiliser ces deux gènes rapporteurs simultanément, en cytométrie en flux et en imagerie. La configuration utilisée sera précisée à chaque expérience en fonction du matériel utilisé.

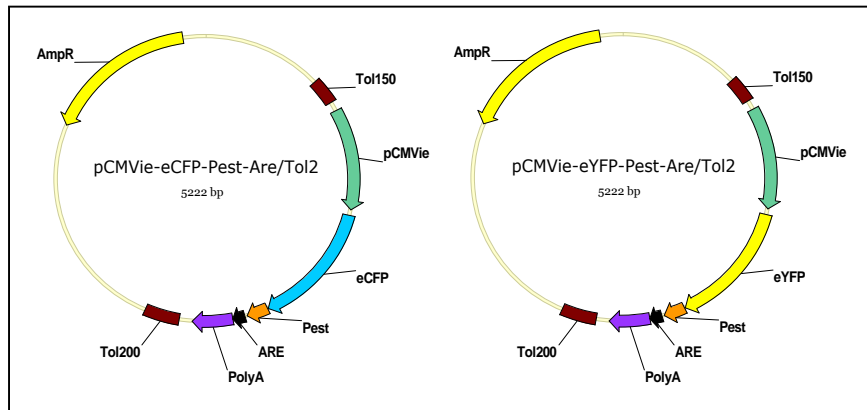
### *1.3.b. Plasmides modifiés*

A partir des deux constructions décrites précédemment, peCFP-N1 et peYFP-C1, deux nouvelles constructions ont été dérivées (Figure 41). La construction de ces deux nouveaux plasmides a été réalisée par nos collaborateurs de Lyon (Gandrillon O., Vinuela J). L'objectif est de réduire la durée de vie des protéines fluorescentes, ceci afin de suivre avec une meilleure résolution temporelle les changements d'expression génique.

La séquence codante de l'eYFP et de l'eCFP a été modifiée en ajoutant deux types de séquences pour déstabiliser d'une part l'ARNm, et d'autre part la protéine (Figure 41) :

- Une séquence dite « ARE », présente dans la partie 3'UTR des ARNm dont la durée de vie est courte (Chen and Shyu, 1995; Fialcowitz et al., 2005).
- une séquence dite « PEST », très répandue dans les protéines ayant une courte durée de vie et qui fonctionne comme un signal d'adressage au protéasome (Rechsteiner, 1988, 1990; Rechsteiner and Rogers, 1996).

De part et d'autre de la séquence codante, deux sites de reconnaissance de la transposase TOL2 (Tol150 et Tol200) ont été ajoutés. Ils facilitent l'intégration de notre transgène dans le génome cible lorsque les constructions sont co-transfectées avec un plasmide codant la transposase TOL2 (Koga and Hori, 2001; Kawakami, 2007).



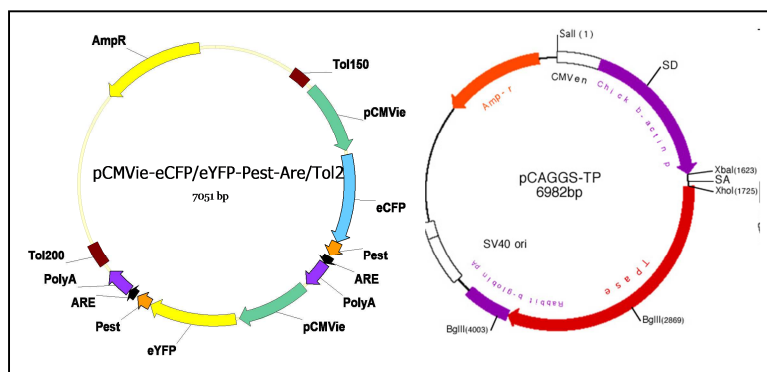
**Figure 41 Plasmides codants eCFP et eYFP déstabilisés.**

*La séquence ARE permet la déstabilisation de l'ARNm et la séquence PEST, celle de la protéine.*

L'ajout des deux types de séquences (ARE et PEST) à la séquence codante ne semble pas modifier les propriétés spectrales des deux protéines. La fluorescence est détectable avec les mêmes outils (lasers, lampes et filtres à fluorescence) que pour les constructions commerciales.

Les deux nouvelles constructions obtenues sont appelées *pCMVie-eCFP-PEST-ARE/Tol2* et *pCMVie-eYFP-PEST-ARE/Tol2*. Pour des raisons de commodité, nous les appellerons par la suite peCFPdes et peYFPdes, « des » pour déstabilisés.

Une troisième construction intégrant les deux séquences codantes modifiées a été produite. Dans cette construction, les séquences codantes sont situées à la suite l'une de l'autre (Figure 42, gauche) : l'eCFPdes précède l'eYFPdes et chacune des séquences codantes est exprimée de façon indépendante grâce à un promoteur dédié (CMV dans les deux cas). Sa dénomination pour le reste de l'étude sera peCFPdes/eYFPdes.



**Figure 42 Construction mixte eCFP/eYFP destabilisées et plasmide Transposase.**

La construction mixte contient les séquences codantes de l'eCFP et de l'eYFP modifiées pour destabiliser les ARNm et les protéines. La construction de droite code pour l'enzyme transposase.

Un plasmide auxiliaire commercial a par ailleurs été utilisé pour apporter la transposase dans les cellules cibles (Figure 42, droite). Ce plasmide est co-transfecté avec les plasmides codant les protéines fluorescentes. La transposase exprimée à partir de cette construction reconnaît les séquences TOL présentes de part et d'autre du transgène et permet de l'intégrer au génome hôte (Koga, 2004). Elle va ainsi exciser la séquence codante des plasmides peCFPdes, peYFPdes et peCFPdes/eYFPdes et faciliter leur intégration dans le génome de la cellule.

#### I.4. Oligonucléotides

Les amorces et sondes utilisées pour les PCR ont toutes été synthétisées par la société Eurogentec (Tableau 3). La solution stock à 200  $\mu$ M est préparée dans de l'eau ultra pure (DNase et RNase Free) puis stockée à  $-20^{\circ}\text{C}$ . Les solutions de travail sont diluées à 10  $\mu$ M dans de l'eau ultra pure et sont aussi conservées à  $-20^{\circ}\text{C}$ .

Dénomination	Séquence (5'-3')
YFP. fw	CTCGTGACCACCTTCGGCT
YFP. rv	TCCTGGACGTAGCCTTCGG
CFP. fw	CGTGACCACCCTGACCTGG
CFP. rv	TCCTGGACGTAGCCTTCGG
hTFIID.fw	GAGAGCCACGAACCACGG
hTFIID.rv	ACATCACAGCTCCCCACCAT
CFP / YFP.pr	FAM -CGACCACATGAA GCAG
Splink 1	CGAAGAGTAACCGTTGCTAGGAGAGACC



<b>Splink 2</b>	GTGGCTGAATGAGACTGGTGTGCGAC
<b>Splink - long</b>	GAAGAGTAACCGTTGCTAGGAGAGACC- GTGGCTGAATGAGACTGGTGTGCGACACTAGTGG
<b>Splink - short</b>	TACCACTAGTGTGCGACACCAGTCTCTAATTTTTTTTTTCAAAAAA
<b>SP6</b>	ATTTAggTgACACTATAg
<b>T7</b>	TAATACgACTCACTATAggg
<b>huZNF143.fw</b>	ACTCTGTTGCTATGGTTACTGC
<b>huZNF143.rv</b>	AGGTCTGGTTTCTGTGGTTAC
<b>huSWAP70.fw</b>	TCCTTTCCATAACCTGTGC
<b>huSWAP70.rv</b>	CACAGAGGGTCCAACACATC
<b>huRFX3.fw</b>	TCGAGCACTTCCAGAGTTTG
<b>huRFX3.rv</b>	TGGAGTAGAGGGAGAATAGCG
<b>huSLC1A1.fw</b>	GTCCTGACTGGGCTTGCAA
<b>huSLC1A1.rv</b>	CAACGGGTAACACGAATCGA
<b>hTFIID.pr</b>	TGTGCACAGGAGCCAAGAGTGAAGA
<b>CMV-GFP.Bi.Fw</b>	TTTGGTAGTATATTAATGGG
<b>CMV-GFP.Bi.Rv</b>	ATAAACTTCAAAATCAACTT
<b>CMV-GFP.Bi.Fo2</b>	CGGGGATTTTTAAGTTTTTAT

Tableau 3 Liste des oligonucléotides

## II. Méthodes Générales

---

### II.1. Culture cellulaire

---

#### II.1.a. Entretien des cellules

---

Les cellules HER911 sont classées parmi les OGM de classe 2 (agrément CGG) et sont donc manipulées dans un laboratoire de type L2, selon les recommandations de bonne pratique de laboratoire. Ce sont des cellules adhérentes, cultivées dans des flasques de facilitant leur adhérence (flask de surface variable, Beckton Dickinson). Elles mesurent environ 30µm de diamètre lorsqu'elles adoptent une forme étoilée. Selon les expériences, une surface différente sera utilisée, notamment en verre pour la microscopie. De plus, selon les expériences, un revêtement particulier comme du collagène ou de la fibronectine a été utilisé. Cette particularité sera détaillée dans les expériences concernées.

Le temps de doublement de ces cellules est d'environ 22 heures en conditions standards mais varie de 16 à plus de 35 heures. Les cellules sont cultivées en routine dans du milieu DMEM (GIBCO) additionné de 10% de SVF (GIBCO) et d'antibiotiques (Pénicilline 100U/ml et Streptomycine 100µg/ml, GIBCO), filtré à 0.22µm. Le SVF a préalablement été chauffé à 56°C pendant 20 minutes afin de dénaturer les protéines du complément.

Les cellules sont incubées en atmosphère humide, à 37°C et 5% de CO<sub>2</sub>.

#### II.1.b. Passage des cellules à confluence

---

Lorsque la confluence atteint 80%, les cellules sont détachées et diluées dans une nouvelle boîte de culture. La densité cellulaire étant un facteur-clé du microenvironnement cellulaire, nous veillerons à ne pas dépasser cette limite.

Les cellules sont lavées avec du PBS (GIBCO) puis détachées par ajout de trypsine/EDTA 0.05% (GIBCO) pendant 5 minutes à 37°C. L'action de cette enzyme est bloquée par ajout de 2 volumes de milieu de culture complet (contenant du SVF) pour un volume de trypsine. La suspension est homogénéisée par aspirations successives et, selon les besoins,

tamisée pour retirer les amas de cellules restants. Ceci est particulièrement important pour la cytométrie en flux, le tri cellulaire et la microscopie sur cellules vivantes que nous décrirons plus loin. Selon les besoins, les cellules sont dénombrées sur des lames de numération. Après centrifugation à 800g pendant 5 minutes, le surnageant est remplacé par 1ml de milieu complet et les cellulesensemencées. En routine, les cellules sont typiquement diluées au 1/10<sup>e</sup> mais pour des expériences particulières, la dilution a été adaptée aux besoins.

### *II.1.c. Congélation et décongélation de cellules*

---

Les cellules sont conservées par congélation. A long terme, elles sont maintenues dans de l'azote liquide à -196°C et à court terme, une conservation à -80°C est privilégiée. Brièvement, les cellules à congeler sont cultivées dans une flasque de 75cm<sup>2</sup>. A 70-80% de confluence, elles sont détachées de leur support de culture. Cette densité correspond à environ 8 millions de cellules. Après centrifugation, le culot est repris dans du milieu complet pour obtenir une concentration de 2 million de cellules par ml. Un millilitre de la suspension cellulaire obtenue est mélangé volume à volume avec un milieu de congélation constitué à 80% de SVF et à 20% de DMSO (SIGMA). Les 2 ml ainsi obtenus sont répartis dans deux ampoules de congélation (10% DMSO final et 1 million de cellules par ampoule). Les ampoules sont placées dans une boîte de congélation contenant de l'isopropanol puis entreposée à -80°C où leur température va décroître régulièrement de 1°C/minute environ. Après 6 heures, les ampoules restent stockées à -80°C ou sont transférées à -196°C dans une cuve d'azote liquide.

Pour décongeler ces cellules, les ampoules sont réchauffées rapidement dans un bain-marie à 37°C. Le contenu est transféré dans un tube contenant 9 ml de milieu complet préchauffé à 37°C puis centrifugé à 800g pendant 5 minutes pour retirer le DMSO. Le culot est repris doucement dans du milieu complet puis transféré dans une flasque de culture annotée de 75cm<sup>2</sup>. Après attachement des cellules, la confluence est d'environ 10%. Le milieu de culture est renouvelé le lendemain pour éliminer toute trace de DMSO et les cellules mortes.

### II.1.d. Transfection de cellules HER 911

---

Il existe différentes méthodes non virales pour introduire un ADN étranger dans une cellule parmi lesquelles les liposomes (micelles lipidiques complexées à l'ADN), les lipoplexes (complexe d'ADN et de molécules polycationiques), l'électroporation ou celle que nous avons utilisée, la transfection. Plusieurs solutions commerciales dérivées du protocole original utilisant un précipité de phosphate de calcium sont disponibles. Nous avons utilisé cette technique historique sur nos cellules qui sont très facilement transfectables. L'objectif étant d'introduire de façon stable les constructions plasmidiques dans les cellules.

Le jour précédent la transfection, des cellules HER911 natives sontensemencées dans une plaque de 6 puits à raison de  $5 \times 10^5$  cellules par puits. Le lendemain, le milieu de culture est renouvelé, puis les cellules HER 911 natives sont co-transfectées avec le plasmide Tol2/transposase et au choix :

- Le plasmide peYFPdes
- Le plasmide peCFPdes
- Le plasmide peCFPdes/eYFPdes

Le milieu de transfection est préparé stérilement de la façon suivante :  $36 \mu\text{l}$  de tampon HBS1x sont mélangé à  $5 \mu\text{g}$  de plasmide. Après agitation,  $5.1 \mu\text{l}$  de  $\text{CaCl}_2$  2.5M sont ajoutés. Le mélange est incubé à température ambiante pendant 20 minutes. Après ce délai, la totalité du volume est déposé sur les cellules en culture au goutte à goutte. Douze heures plus tard, le milieu de culture est changé et la transfection contrôlée au microscope à fluorescence, avec les filtres adaptés. Les cellules d'un puits contrôle sont analysées par cytométrie en flux pour déterminer le pourcentage de cellules transfectées. A ce stade, beaucoup de cellules sont transfectées de façon transitoire, sans intégration au génome. Les cellules sont entretenues classiquement pendant 1 semaine sans sélection. Après ce délai, la majorité de l'expression transitoire a disparu. Seules les intégrations stables restent visibles. Les cellules sont alors clonées grâce à un trieur de cellules.

## II.2. Biologie moléculaire

---

### II.2.a. Électrophorèse

---

L'électrophorèse a été utilisée afin de vérifier la qualité des acides nucléiques extraits, la digestion par les enzymes de restrictions ou la taille de produits de PCR. Les acides nucléiques sont résolus sur gel d'agarose (Invitrogen) à 0.8% m/v (ADN génomique), 1% m/v (ARN et ADN génomique digéré) ou 2% m/v (Produits de PCR) dissous dans du tampon TAE 1x (EUROMEDEX). Un agent intercalant, le SyBrSafe (1x, Invitrogen), a été utilisé pour visualiser les acides nucléiques dans le gel par émission de fluorescence sous exposition aux UV.

Classiquement, 10µl d'ADN sont déposés sur gel avec 2µl de tampon de charge 6x (Fermentas). Pour l'ARN, 2µl sont mélangés à 2µl de formamide 99%, chauffés 10 minutes à 50°C pour les dénaturer puis déposés avec 1µl de tampon de charge 5x RNAses free (Gel pilot loading dye ;Qiagen). Pour chaque électrophorèse, 5 µl (0.5µg) de marqueur de taille (GeneRuler Express DNA Ladder ; Fermentas) sont déposés aux extrémités du gel. La migration s'effectue entre 3 et 6 volts/cm, dans une cuve contenant du TAE 1x.

### II.2.b. Extraction d'ADN génomique

---

L'extraction d'ADN a été réalisée avec le kit « Genomic DNA » (Qiagen) suivant les instructions du fabricant. Brièvement, les membranes plasmiques des cellules (entre 5 et 10 millions) conservées sous forme de culots secs à -80°C sont lysées par un détergent. Après centrifugation, les noyaux sont à leur tour lysés. Le lysat est dans un premier temps incubé avec de la RNaseA (15minutes, 37°C, Qiagen) pour éliminer les ARN puis incubé avec de la protéinase K (1 heure, 50°C, Qiagen) afin de digérer les protéines nucléaires.

L'ADN est ensuite isolé et purifié sur une colonne de silice par chromatographie échangeuse d'anions grâce à différents tampons de pH et force ionique variables. L'ADN est élué dans du tampon TE puis précipité par de l'isopropanol. Après centrifugation, le culot d'ADN est lavé par de l'éthanol 70% puis solubilisé dans du tampon TE. La qualité de

l'ADN extrait et sa concentration sont estimés par spectrophotométrie (NanoDrop). La lecture de l'absorbance à 260nm (Acides nucléiques) et 280nm (protéines) nous renseignent sur la concentration de l'échantillon et sa pureté (ratio 260/280).

### *II.2.c. Extraction d'ARN totaux*

---

Les ARN totaux ont été extraits avec le kit « RNeasy Plus mini » (Qiagen) suivant les instructions du fabricant à partir de  $5 \times 10^5$  cellules. Cette étape est réalisée dans une pièce dédiée à la manipulation d'ARN, où aucune trace de RNase ne doit apparaître. Le principe est le même que pour l'extraction d'ADN décrite ci-dessus. Après lyse des cellules, l'ADN génomique est séparé des ARN grâce à une colonne de silice par exclusion de taille et affinité. Les ARN sont purifiés sur une nouvelle colonne afin d'éliminer les protéines résiduelles puis élués dans de l'eau UltraPure traitée au DEPC (Ambion), RNase et DNase free. Comme précédemment, la concentration est estimée par lecture de l'absorbance à 260nm. La qualité des ARN est quant à elle vérifiée par électrophorèse sur gel d'agarose 1% après leur dénaturation par du formamide (Sigma).

### *II.2.d. Purification des ARN totaux*

---

Les ARN totaux sont purifiés par traitement à la DNase I afin de retirer toute contamination par de l'ADN génomique (Turbot DNA-free Kit, Ambion). 10 µg d'ARN totaux sont mélangés à 5µl de tampon 10x, 1µl de DNase I (2U/µl), 1.25µl de RNAsin (40U/µl) et qsp 50µl d'eau RNase free. Après 30 minutes à 37°C, 5µl de résine d'inactivation sont ajoutés. Après 2 minutes, le tube est centrifugé à 10 000g pendant 5 minutes et la phase liquide prélevée. Un nouveau dosage est effectué après cette étape de purification et les ARN conservés à -80°C dans des tubes RNase Free ou directement retro-transcrits en ADNc.

### *II.2.e. Rétro-transcription des ARNm*

---

La synthèse de l'ADNc est réalisée avec le kit « Superscript First strand synthesis System » (Invitrogen) suivant les instructions du fabricant dans une pièce dédiée à ce type de

manipulations. Brièvement, 1µg d'ARN purifié est incubé avec des amorces aléatoires fournies dans le kit. Une enzyme reverse transcriptase (ADN polymérase ARN dépendante) effectue la synthèse d'un ADN complémentaire à la matrice ARN à partir de ces amorces aléatoires. Les cDNA obtenus sont plus stables que les ARN et insensibles aux RNAses. Ils sont conservés à -80°C.

### *II.2.f. qPCR*

---

La PCR quantitative en temps réel est utilisée pour estimer le nombre de copies des gènes rapporteurs intégrés dans le génome cellulaire et pour tester la spécificité des amorces.

#### *◆ Quantification*

---

La chimie Taqman est utilisée pour la quantification. La sonde d'hydrolyse contient un fluorophore rapporteur situé en 5' et un « quencher » de fluorescence spécifique en 3' (Se référer à la description des oligonucléotides). Une amplification est réalisée pour les transgènes eYFP, eCFP et albumine. L'albumine étant présente en deux exemplaires dans chaque cellule, cela nous permet de standardiser les valeurs de Ct obtenues pour chaque couple d'amorce utilisé.

Une gamme étalon est réalisée pour chaque gène, allant de  $10^3$  à  $10^7$  copies grâce aux différents plasmides (ALB et GFP : plasm ID 769, eYFP : plasm ID 140, eCFP plam ID 141). Chaque point est réalisé en double. Pour une réaction, le mélange réactionnel suivant est utilisé : 1,75µl d'eau, 12,5µl de mix de qPCR 2x (Absolute QPCR ROX mix ; ABgene), 0,25µl de chaque amorce et sonde 10µM. 15µl sont déposés dans chaque puits d'une plaque 96 puits. Pour chaque échantillon, 10µl d'ADN à 10 ng/µl sont ajoutés.

La réaction s'effectue sur un automate Applied Biosystems 7900HT Fast Real-Time PCR System dont la programmation est la suivante : - 95°C, 15 minutes - 40 cycles : 95°C/15sec puis 60°C/1min. L'analyse des données est effectuée avec le logiciel « Sequence Detection Systems v2.2.2 ».

### ◆ *Évaluation des amorces*

---

La PCR en temps réel est utilisée pour évaluer la spécificité des différents couples d'amorces utilisés à travers ce travail. Pour cela, une qPCR est réalisée avec les couples d'amorces à tester sur un ADN extrait de cellules 911 et un mix commercial (MESA BLUE qPCR MasterMix Plus for SYBR, Eurogentec). L'amplification se fait à différentes température grâce à un thermo-cycleur à gradient (Chromo4, Biorad). Une courbe de fusion entre 65 et 90°C à la fin des cycles d'amplification permet de contrôler la présence d'un seul amplicon. Une électrophorèse suite à l'amplification permet de contrôler la taille des amplicons à la taille théorique estimée lors de la conception des amorces.

#### *II.2.g. Analyse de la méthylation par la méthode du bisulfite*

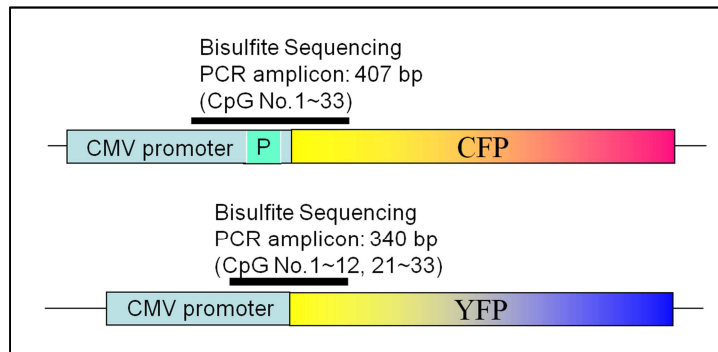
---

L'analyse de la méthylation des promoteurs CMV a été réalisée par la technique du bisulfite (EpiTect Bisulfite, Qiagen). Cette technique consiste à convertir chimiquement les cytosines non-méthylées en uracile et à séquencer les différents CpG. Les cytosines méthylées seront converties en Thymine.

Le mélange réactionnel contient : 2µg d'ADN génomique, 85µl de réactif « bisulfite mix », 15 µl de tampon « DNA protect » et de l'eau (qsq 140µl). La réaction se déroule ainsi : 99°C-5', 60°C-25', 99°C-5', 60°C-85', 99°C-5' puis 60°C-175'. L'ADN converti est ensuite purifié sur colonne, selon les indications du fabricant.

Cette étape est suivie d'une amplification par PCR avec des amorces spécifiques. Une première PCR est réalisée avec un couple d'amorces qui s'hybride entre le promoteur CMV et la séquence codante d'eCFP ou d'eYFP (CMV-GFP BiFw et Rv, Tableau 3, Figure 43). Le mix de PCR contient : 2µl tampon 10x (Qiagen), 0,4µl 10mM dNTPs mix, 0,2µl Taq polymérase (TaqGold, Applied), 0,2 µl mélange Amorces10µM, 100ng d'ADN et de l'eau (qsp 20µl). Les conditions de température sont les suivantes : 94°C-3' puis 20 cycles 94°C-1', 55°C-45", 72°C-1'. Cette PCR est suivie d'une PCR emboîtée réalisée avec les amorces CMV-GFP Bi Fo2 et CMV-GFP Bi Rv (Tableau 3). Le mélange réactionnel et les cycles de température sont identiques à la première PCR.





**Figure 43 Régions amplifiées pour l'étude de la méthylation du promoteur CMV.**

*La taille de l'amplicon est différente entre les transgènes eCFP et eYFP. 33 et 24 CpG sont respectivement analysés par amplicon.*

Après électrophorèse sur gel d'agarose, les bandes correspondantes à l'amplification de l'eCFP et de l'eYFP sont extraites et purifiées (QIAquick Spin, Qiagen). Les produits de PCR purifiés sont clonés dans un vecteur pGEM-easy (Promega) et transformés dans des bactéries JM109 (promega) suivant les instructions du fabricant.

Les bactéries sontensemencées sur une boîte de Pétri contenant du milieu solide LB-agar-Ampicilline supplémenté en X-gal (40µl @ 20mg/ml) et IPTG (40µl @ 20mg/ml). Le lendemain, quelques colonies blanches sont repiquées et amplifiées en milieu LB – ampicilline liquide, et une extraction de plasmides est réalisée pour chaque échantillon (QIAprep Miniprep, QIAGEN). L'insert des plasmides extraits et purifiés est séquencé grâce aux amorces SP6 et T7.

#### *II.2.h. Identification des sites d'intégration par Splinkerette PCR*

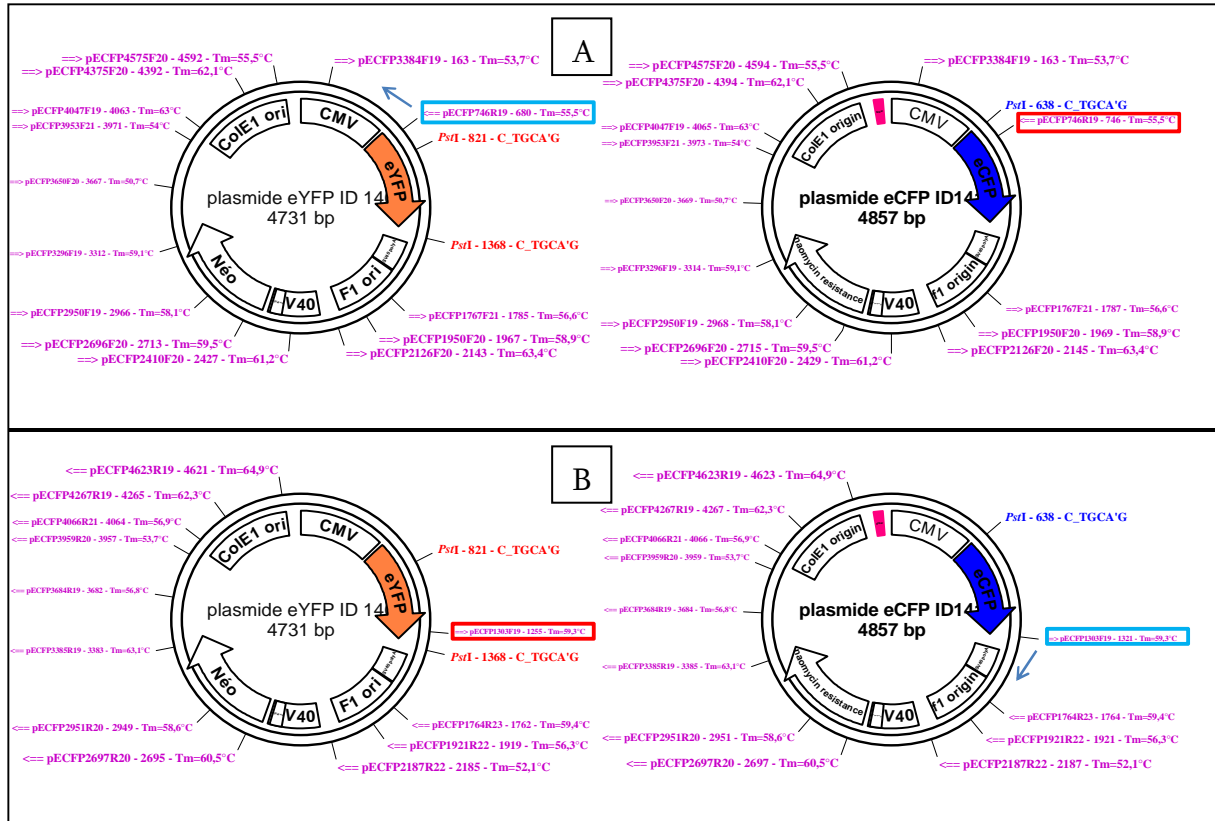
Le site d'intégration de chaque transgène est identifié par une technique dérivée de la PCR, la splinkerette PCR (Uren et al., 2009). Cette technique consiste à amplifier un fragment d'ADN contenant la jonction entre l'ADN génomique et l'ADN plasmidique puis à séquencer ce fragment.

Deux approches ont été utilisées en fonction du modèle : dans le modèle établi avec les plasmides commerciaux, l'identification a été manuelle. Dans le modèle fabriqué avec les plasmides codant les protéines déstabilisées, une technique à haut débit a été utilisée,

suivant le protocole proposé par Uren et al. (2009). Cette seconde approche a été développée par nos collaborateurs de Lyon (Olivier Gandrillon et José Vinuelas).

Dans la première approche, la difficulté est que les plasmides commerciaux n'ont pas été conçus pour ce genre d'identification. Ils peuvent s'être ouverts à n'importe quel endroit pour se linéariser lors de leur intégration. Nous savons néanmoins que le site d'ouverture ne se trouve pas dans la séquence codante pour les gènes rapporteurs eCFP et eYFP, puisque ceux-ci sont exprimés dans les clones cellulaires. Une autre difficulté est la quasi-homologie de séquence entre les plasmides peCFP-N1 et peYFP-C1 qui rend la discrimination entre les deux difficile.

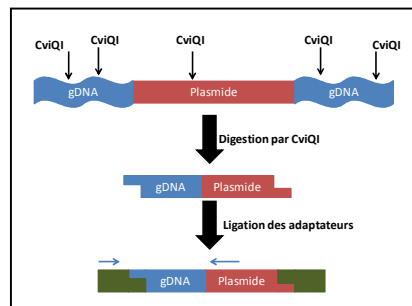
Dans un premier temps, une PCR est réalisée après digestion de l'ADN génomique par l'enzyme PstI. Cette enzyme possède deux sites de reconnaissance sur le plasmide eYFP et un seul sur le plasmide eCFP. Nous avons utilisé deux amorces qui permettent de faire la distinction entre les deux plasmides (Figure 44). Grâce à la digestion, l'amorce 746R19 ne peut amplifier que le plasmide eYFP car le site de restriction sur le plasmide eCFP se trouve devant l'amorce (Figure 44, A, encadré). L'amorce 1303F19 est utilisée avec la même stratégie pour amplifier le plasmide eCFP (Figure 44, B, encadré). La PCR est réalisée à partir de chacune des amorces, couplée à diverses autres amorces disposées autour de chacun des plasmides. L'absence d'amplification permet alors d'estimer approximativement le site d'ouverture du plasmide entre deux amorces.



**Figure 44** Estimation du site d'ouverture des plasmides.

Les plasmides eCFP et eYFP sont digérés différemment par l'enzyme *PstI* (gauche vs droite). Deux amorces (encadrées) peuvent alors amplifier spécifiquement chacun d'eux (A vs B).

Après avoir estimé le site d'ouverture des plasmides, une seconde approche est utilisée pour identifier exactement la position du site d'ouverture et du site d'intégration génomique des plasmides. La stratégie consiste à digérer l'ADN génomique contenant le transgène avec une endonucléase qui, idéalement, ne possède pas de site de coupure sur le plasmide mais coupe fréquemment dans l'ADN génomique. Des adaptateurs seront ajoutés de part et d'autre du fragment digéré pour servir de site d'hybridation pour une des amorces de PCR (Figure 45).



**Figure 45** Principe de la Splinkerette PCR.

L'ADN génomique contenant les plasmides intégrés est digéré par un endonucléase. Des adaptateurs compatibles sont alors ajoutés aux extrémités. Une PCR est réalisée pour amplifier la jonction plasmide / ADNg.

L'enzyme choisie est CviQI (ou Csp6I, Fermentas) dont la séquence cible est composée de seulement 4 bases (G|TA\_C) et coupe l'ADN statistiquement toutes les 256 bases. Suite à la digestion, des adaptateurs oligonucléiques de synthèse sont ajoutés par ligation (T4 DNA ligase, Invitrogen). Ces adaptateurs sont formés par hybridation des oligonucléotides Splink long et Splink short (Tableau 3), et sont compatibles avec l'extrémité générée par la digestion. Ils contiennent le site d'hybridation pour les amorces anti-sens splink1 et splink2.

Le site d'ouverture n'étant pas connu, nous avons dessiné des amorces sens compatibles avec Splink 1 et Splink 2 (Tableau 3) régulièrement disposées sur les plasmides. Les amorces peuvent s'hybrider aussi bien sur peCFP-N1 que sur peYFP-C1.

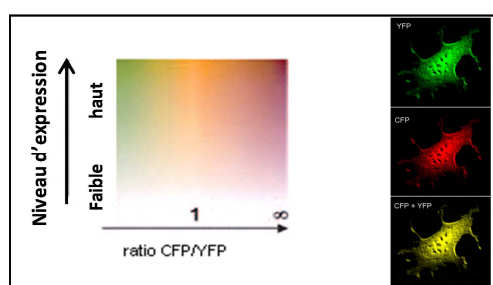
Une amplification est réalisée entre l'amorce Splink1 et les différentes amorces s'hybridant sur les plasmides. Après contrôle par électrophorèse, une seconde PCR est réalisée avec l'amorce splink2 et les amorces sens qui ont fonctionné. Après électrophorèse, les puits présentant des bandes différentes du contrôle négatif sont purifiés et les fragments clonés dans un vecteur pGEM-easy (Promega). Les plasmides recombinants sont utilisés pour transformer des bactéries compétentes (JM109, Promega) afin d'amplifier chaque insert. Les bactéries sontensemencées sur une boîte de Pétri contenant du milieu solide LB-agar-Ampicilline supplémenté en X-gal (40µl @ 20mg/ml) et IPTG (40µl @ 20mg/ml). Le lendemain, une vingtaine de colonies blanches sont repiquées et amplifiées en milieu LB – ampicilline liquide. Une extraction de plasmides est réalisée pour chaque échantillon (QIAprep Miniprep, QIAGEN) et l'insert des plasmides est séquencé grâce aux amorces SP6 et T7.

Après avoir identifié le site d'intégration et le site d'ouverture des plasmides, nous avons réalisé une amplification par PCR avec une amorce située au niveau du site d'ouverture sur la jonction ADN/plasmide. Une seconde amorce située dans la région codante de l'eCFP et l'eYFP est utilisée. Afin de faire la différence entre eCFP et eYFP, l'ADN génomique du clone est digéré par l'enzyme PstI comme précédemment.

## II.3. Techniques d'imagerie

### II.3.a. Colorisation des images

Les différentes images acquises par les différents microscopes ont été retraitées et parfois colorisées pour améliorer la présentation. C'est notamment le cas lors de l'analyse de cellules exprimant simultanément les transgènes eCFP et eYFP. Pour l'ensemble des images en couleur (RGB) présentées dans ce travail, les couleurs rouge et verte correspondent respectivement aux transgènes eCFP et eYFP (Figure 46).



**Figure 46 Colorisation des images RGB.**

*Le canal rouge est attribué à la fluorescence eCFP et le vert à la eYFP. L'expression des deux transgènes résulte en une couleur jaune de la cellule.*

Par synthèse additive des couleurs, l'expression des deux transgènes dans une cellule se traduira par une couleur jaune-orangée de celle-ci.

### II.3.b. Microscopie à fluorescence

L'efficacité de transfection et la fluorescence des cellules de nos modèles ont été contrôlées avec un microscope inversé Axiovert135 (ZEISS) équipé d'une caméra CCD (Hamamatsu) et d'une lampe à mercure HBO100W. Différents blocs de filtres dichroïques permettent de sélectionner les longueurs d'ondes souhaitées en fonction du fluorochrome. Les images sont acquises avec le logiciel WASABI et retraitées avec le logiciel ImageJ.

### II.3.c. Microscopie temps réel

---

#### ◆ *Biostation(NIKON)*

---

L'appareil majoritairement utilisé est la Biostation IM-Q (NIKON). C'est un appareil d'imagerie compact, thermo-régulé et couplé à un mélangeur de CO<sub>2</sub> qui permet de reproduire les conditions d'un incubateur. Ce microscope inversé permet d'observer à long terme des cellules vivantes et de réaliser des expériences en continu. Il est équipé d'un objectif 20x couplé à deux lentilles qui permettent d'accroître le grossissement à 40x et 80x. La mobilité de l'objectif est un atout car elle permet de scanner presque entièrement un échantillon sans le bouger. La présence d'un bloc de filtres optiques en association avec une lampe à mercure permet de collecter des signaux fluorescents. Nous utilisons ici cet appareil pour suivre l'évolution de la fluorescence dans nos modèles de cellules bifluorescentes à travers plusieurs divisions cellulaires et dans des conditions diverses.

Les cellules à analyser sont décollées de leur support de culture comme décrit précédemment et tamisées. La concentration cellulaire de la suspension est déterminée puis 1500 cellules sontensemencées dans une boîte de culture IBIDI HI-Q4 (IBIDI) spécialement conçue pour cet appareil. Elles sont divisées en 4 compartiments indépendants, permettant de faire 4 conditions expérimentales ou échantillons simultanément. Le fond de cette boîte est de qualité optique pour une résolution optimale. Après attachement des cellules, la boîte est placée dans l'appareil pendant 1 heure jusqu'à stabilisation thermique des différents éléments évitant ainsi toute fluctuation pendant l'acquisition. Pour chaque compartiment, 20 à 25 champs à analyser sont répertoriés soit environ 100 champs par expérience. On voit nettement l'avantage d'un tel objectif mobile, qui ne nous restreint pas à un seul champ. Le délai entre deux prises de vue est de 15 à 20 minutes en moyenne lorsque 100 champs sont analysés et 3 canaux utilisés.

#### ◆ *Station confocale thermo-régulée (Zeiss)*

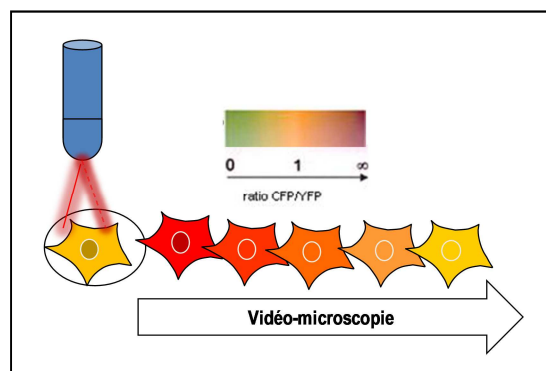
---

Nous avons utilisé un microscope inversé Zeiss Axiovert 100M équipé d'une tête confocale LSM510. Cet appareil est confiné dans une enceinte thermo-régulée et une chambre

spéciale permet de maintenir 5% de CO<sub>2</sub> autour de la culture cellulaire. Il est équipé de plusieurs objectifs secs et à immersion (10x à 63x), de 3 lasers (5 raies utilisables : 457nm, 488nm, 514nm, 543nm et 633nm), de 2 détecteurs et d'un détecteur spectral (2+8 détections simultanées). Contrairement à l'équipement précédent, ni les objectifs ni la platine ne sont mobiles. Les cellules sont préparées de la même façon mais sont incubées dans une boîte de culture de 35mm de diamètre MatTek à fond en verre. 4000 cellules sontensemencées dans chaque boîte.

### II.3.d. Photo blanchiment de cellules entières

La technique de photo blanchiment sur cellules entières a été utilisée pour mesurer la cinétique de synthèse d'une protéine fluorescente dans notre modèle bifluorescent. C'est une technique dérivée du FRAP qui permet de mesurer des coefficients de diffusion en étudiant la vitesse de recouvrement de fluorescence dans une zone de la cellule où l'on a volontairement éteint cette fluorescence par excitation laser (Figure 47). Ici, nous avons décidé d'éteindre la fluorescence de l'eYFP et de conserver l'eCFP comme référence. Ce choix résulte de la configuration du microscope utilisé, notamment d'une puissance de laser insuffisante pour réaliser cette expérience sur l'eCFP.



**Figure 47 Principe du photo-blanchiment de cellule entière.**

*La cellule exprimant les deux transgènes (jaune) est illuminée par un laser à 514nm jusqu'à extinction de la fluorescence eYFP. La synthèse de nouvelles protéines fluorescentes est analysée par vidéo-microscopie.*

Le microscope utilisé est le confocale Zeiss décrit précédemment. Pour chaque champ, une à quatre cellules ont été analysées et une à deux cellules non perturbées afin de servir de référence. Dans un premier temps, une ROI est déterminée autour des cellules à

illuminer. Une première image est sauvegardée avant l'extinction de fluorescence. Ensuite, les cellules repérées sont excitées par le laser 514nm à pleine puissance pendant 20 pulses. La fluorescence est diminuée de 80% environ, de façon reproductible. Suite à cette excitation laser, une image est capturée toutes les 10 minutes pendant 5 à 10 heures.

## II.4. Cytométrie en flux et tri cellulaire

---

### II.4.a. Analyse de la distribution de fluorescence

---

La cytométrie en flux est une technique alliant fluide, optique et électronique permettant d'analyser rapidement un grand nombre d'évènements avec une grande sensibilité. Elle est basée sur la détection de lumière diffusée ou de signaux fluorescents après excitation laser de particules (ie. virus, bactéries, cellules...) individuelles entrainées par un flux sous pression. Nous utiliserons cette technique principalement dans le but de mesurer la fluorescence des cellules de notre modèle bifluorescent et pour générer des clones.

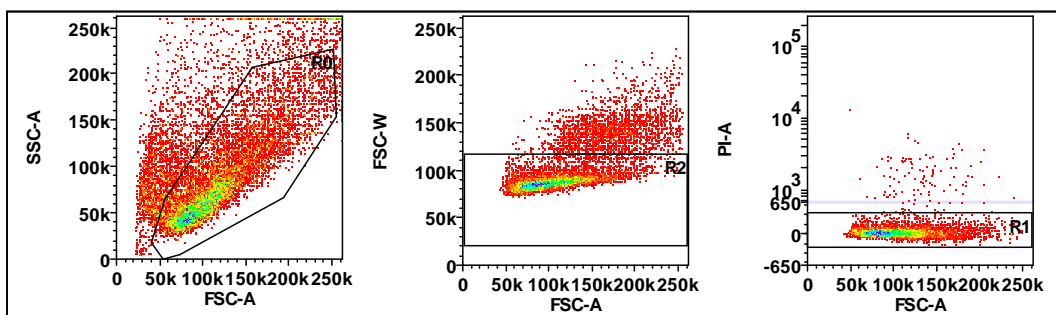
Les cellules à analyser sont détachées de leur support comme décrit précédemment. La suspension cellulaire est tamisée sur un maillage de 30µm afin de retenir les amas de cellules qui n'ont pas été dissociés par la trypsine/EDTA. Typiquement, la taille d'une cellule HER 911 isolée en suspension est de 10-15µm de diamètre. Après centrifugation, les cellules sont reprises dans du PBS1x contenant 2% de SVF et 2µg/ml d'iodure de propidium, un marqueur de viabilité qui se fixe à l'ADN. Seules les cellules mortes sont perméables à cette molécule et sont donc fluorescentes. Les cellules sont conservées à 4°C en attendant leur analyse. L'appareil utilisé est un LSR II (Beckton Dickinson) équipé de 3 lasers (405,488 et 633nm) et pouvant collecter 11 paramètres simultanément en plus de la taille et la structure. Une maintenance hebdomadaire est réalisée sur cet appareil garantissant un fonctionnement optimal.

Pour notre modèle, l'eCFP et l'eYFP sont excités respectivement à 405nm et 488nm. La fluorescence émise est collectée en utilisant un filtre commun 505LP puis un filtre 525/50nm pour l'eCFP et un filtre 530/30nm pour l'eYFP. Afin de standardiser les mesures



d'un jour à l'autre, de billes de calibration sont utilisées lors de chaque analyse. Ces billes sont recouvertes de composés fluorescents ayant les mêmes propriétés spectrales que nos protéines fluorescentes. Leur fluorescence est stable dans le temps et certifiée comparable d'un lot à l'autre par le fabricant. Elles permettent de vérifier que la puissance des lasers notamment ne varie pas. Les cellules mortes ou en voie d'apoptose, marquées par le PI, émettent une fluorescence rouge lorsqu'elles sont excitées par le laser 488nm. Cette fluorescence est collectée à travers un filtre 600LP puis 610/20nm.

Pour chaque échantillon, des régions d'intérêt sont utilisées en fonction des différents paramètres mesurés afin de sélectionner la population qui nous intéresse (Figure 48). Dans un premier temps, les débris sont retirés grâce aux informations de taille et de structure. Les cellules mortes sont à leur tour retirées grâce à la fluorescence du PI. En général, 10 000 cellules PI négatives (vivantes) sont analysées pour chaque échantillon.



**Figure 48 Procédure d'analyse en cytométrie en flux.**

*De gauche à droite : les objets de forme correcte sont sélectionnés (R0), puis les objets uniques (R2) et enfin les objets n'étant pas marqués par le PI (R1). Le code couleur est un gradient de densité.*

Les données sont exportées en format propriétaire pour un retraitement éventuel sur le logiciel de l'appareil (FACS Diva, BD) ou en format FCS 3.0 reconnu par de nombreux logiciels. Nous utiliserons par la suite le logiciel WEASEL 3.0<sup>1</sup> (Walter and Elisa Hall Institute, Melbourne) pour les représentations graphiques, l'analyse en batch et l'extraction automatique de statistiques. Les distributions de fluorescence n'étant pas normales, la moyenne arithmétique de fluorescence n'a pas été utilisée, au profit de la médiane ou de la moyenne géométrique.

<sup>1</sup> <http://www.wehi.edu.au/>

### *II.4.b. Clonage et sous-clonage cellulaire*

---

Dans un second temps, nous avons cloné des cellules transfectées avec les constructions plasmidiques pour générer des populations isogéniques de notre modèle. Cette étape a été réalisée avec un trieur de cellules MoFlo (Cytomation) possédant 3 lasers (445nm, 488nm et 633nm) et pouvant analyser 11 signaux fluorescents simultanément. Le fonctionnement de cet appareil est très similaire à celui du cytomètre en flux à la différence que l'on peut récupérer les cellules après leur analyse. En fonction du critère d'intérêt, il est possible de réensemencer les cellules dans une plaque 96 puits, à raison d'une cellule par puits et de façon stérile. Les protéines eCFP et eYFP sont excitées par le laser 445nm et 488nm respectivement et leur fluorescence collectée avec un filtre 485/25nm et 575/25nm.

Les cellules triées sont collectées dans une plaque 96 puits contenant 100µl de milieu DMEM complet et cultivées comme décrit précédemment. Le lendemain, 100µl de milieu de culture sont ajoutés et la présence de cellules dans chaque puits vérifiée par microscopie. Après une semaine, les cellules de chaque puits sont transférées dans une plaque plus grande (6 puits de 10cm<sup>2</sup>) puis à 80% de confluence dans une flasque de T75. Lorsque la confluence atteint de nouveau 80%, les cellules sont congelées en partie, en suivant le protocole décrit précédemment et une autre partie est gardée en culture pour différentes expériences.

## *II.5. Bio-informatique*

---

### *II.5.a. Conception d'amorces de PCR*

---

Les amorces utilisées pour la splinkerette PCR ont été choisies et optimisées à l'aide du logiciel oligo7, de la séquence des plasmides eCFP et eYFP. Elles sont compatibles avec les amorces SPL1 et SPL2. L'outil Primer BLAST du site NCBI a été utilisé pour vérifier l'absence d'hybridation non spécifique dans le génome humain. Les amorces de qRT-PCR ont été dessinées à partir des séquences de cDNA.

### *II.5.b. Analyse de séquences*

---

Les séquences obtenues après le séquençage des inserts sont analysées avec le logiciel BioEdit V7.0.5.1. (Hall, 1999) Dans un premier temps, les chromatogrammes sont analysés visuellement afin de juger de leur qualité. Les parties de mauvaise qualité, généralement au début et à la fin du tracé sont éliminées. Après cet examen visuel, les séquences sont alignées avec les amorces de séquençage et avec les autres séquences connues. Les séquences nettoyées sont comparées à la base de données NCBI par BLAST afin de d'identifier de l'ADN génomique humain. En identifiant les séquences plasmidiques et génomiques, nous pouvons caractériser le site d'intégration dans le génome mais aussi la position d'ouverture du plasmide lors de l'intégration.

### *II.5.c. Analyse d'images*

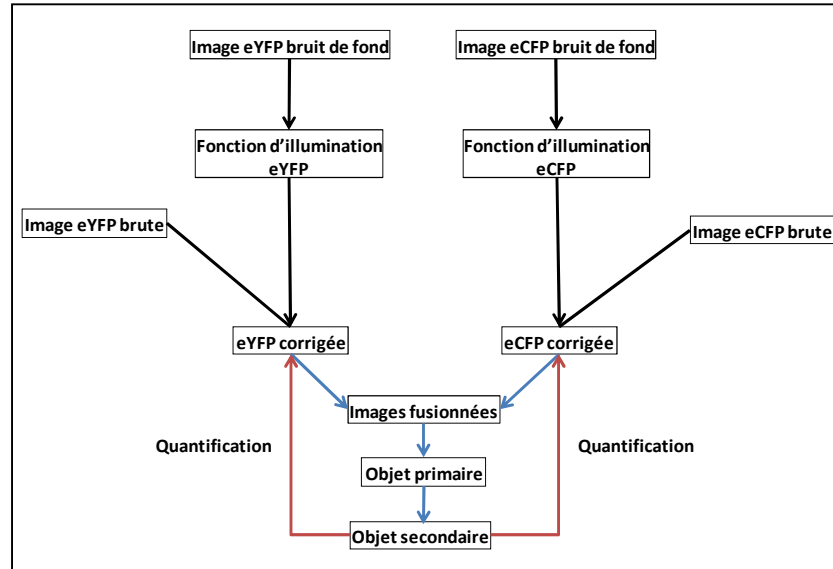
---

Les images issues des expériences de vidéo-microscopie et de photo-blanchiment sont analysées automatiquement à l'aide du logiciel CellProfiler<sup>1</sup> (Carpenter et al., 2006; Jones et al., 2008, 2009). C'est un logiciel open-source développé par le BROAD INSTITUTE en collaboration avec de nombreux partenaires universitaires et industriels, notamment pharmaceutiques. Ils utilisent de plus en plus ce genre d'analyse d'images au cours de screening cellulaires à haut débit de molécules potentiellement pharmaco-actives. Le logiciel utilise une programmation et des algorithmes développés sous MatLab ce qui le rend très puissant et accessible. De plus, sa licence open-source regroupe, à l'instar du logiciel ImageJ, une communauté de développeurs qui proposent des solutions en ligne pour de nombreuses problématiques.

L'objectif principal de cette analyse est de segmenter les cellules pour en réaliser un suivi automatique sur toute la durée de l'expérience. Le second objectif sera de mesurer différents paramètres sur l'image entière ou pour chaque objet identifié (Figure 49).

---

<sup>1</sup> <http://www.cellprofiler.org/>

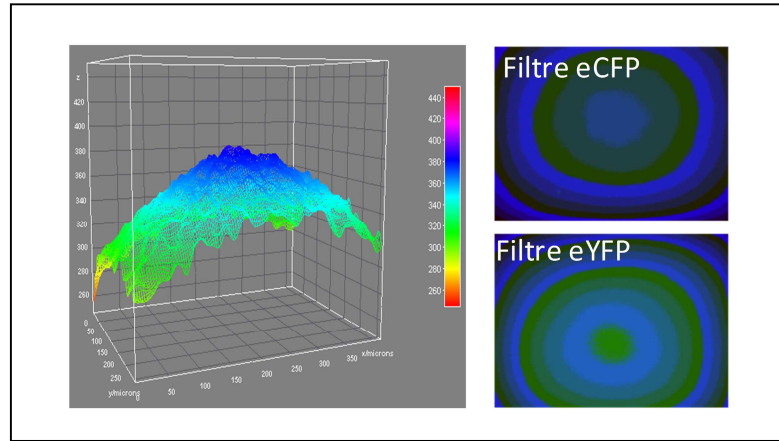


**Figure 49** Étapes de l'analyse d'images par CellProfiler.

Les images brutes sont normalisées grâce à une fonction d'illumination calculée sur le bruit de fond. Les images corrigées sont utilisées pour la détection des objets (cellules) et la quantification de la fluorescence.

Brièvement, voici la démarche globale adoptée :

1. Les images correspondantes aux canaux eCFP et eYFP sont ouvertes séparément en format TIFF 8 bits.
2. Le bruit de fond mesuré sur une image acquise sur un champ vide est retiré des images brutes par soustraction.
3. Une fonction d'auto-illumination est calculée sur une image provenant du même champ vide à chaque point de temps, pour eCFP et eYFP. Celle-ci permet de corriger des différences d'illumination pouvant apparaître sur une image, notamment dans les coins de celle-ci (vignettage)(Figure 50).



**Figure 50 Différences d'illumination mesurées sur un champ vide.**

*L'intensité de fluorescence est plus importante au centre de l'image (gauche). De plus, cette hétérogénéité diffère selon le filtre utilisé (droite). Images colorées par gradient d'intensité.*

4. Cette image de normalisation (dont l'intensité des pixels est comprise entre 0 et 1) est appliquée aux images sans bruit de fond correspondantes par simple division. La variabilité spatiale et temporelle est ainsi éliminée, sur la base du bruit de fond (Schwarzfischer et al., 2011).
5. Les deux images corrigées (eCFP et eYFP) sont associées en fonction de l'intensité des pixels :
  - a. Par addition si l'intensité des deux canaux est faible
  - b. Par valeur moyen ou maximale si l'intensité entre les deux canaux est différente.

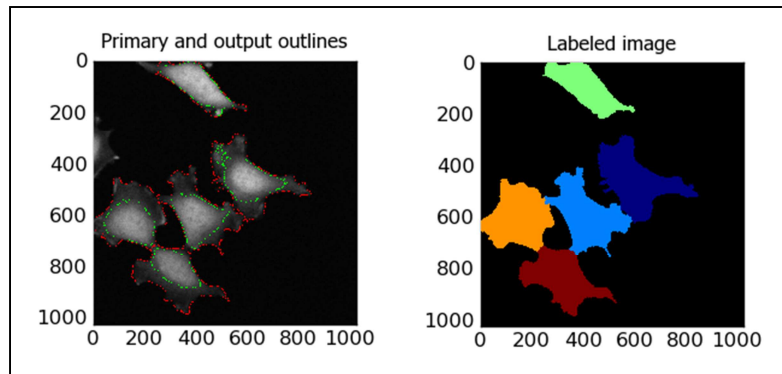
Ceci permet d'obtenir un signal suffisamment fort mais ne saturant pas, afin de conserver un ratio signal/bruit de fond acceptable. Le contour des cellules est alors suffisamment intense pour être entièrement détecté. L'économie de cette étape induit un risque de biais dans la détection du contour des cellules et par la suite sur le calcul de la taille et des moyennes de fluorescence.

La détection des objets se fait en deux étapes, à partir des images fusionnées (Figure 51) :

1. Les objets dits « primaires » sont détectés en appliquant un seuil d'intensité élevé (algorithme Otsu avec une approche adaptative). Ainsi, seules les parties intenses de chaque cellule seront identifiées, en général la surface au dessus du noyau qui

est en relief. Un filtre de taille est appliqué pour éliminer les objets trop petits ou trop grands, ainsi que les objets touchant le bord de l'image.

2. Les objets « secondaires » sont identifiés par expansion des objets primaires jusqu'à la limite du bruit de fond par un algorithme de ligne de partage des eaux (Figure 51).



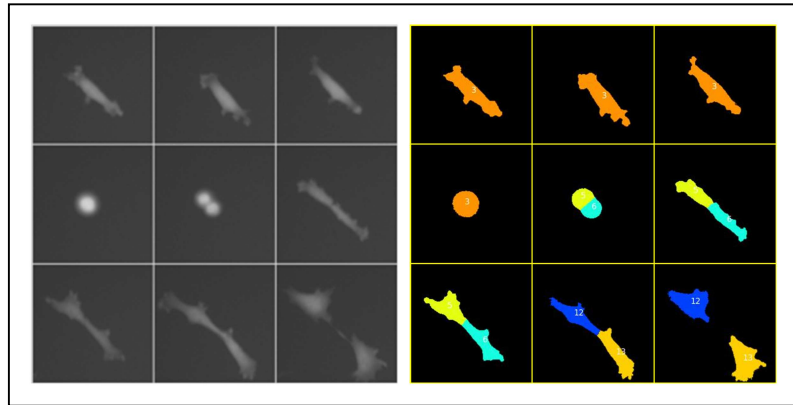
**Figure 51 Segmentation des cellules.**

*Les objets primaires (contours verts) sont identifiés grâce à un seuil d'intensité élevé. Les objets secondaires (contours rouges) sont identifiés par expansion des objets primaires (Gauche). On obtient ainsi un masque des objets identifiés (Droite).*

L'intensité des pixels dans chaque objet est mesurée sur chacune des images corrigées (eCFP et eYFP). A partir de ces mesures, plusieurs paramètres sont calculés :

1. La fluorescence totale, la moyenne, la médiane, le minimum, le maximum et la variance de fluorescence par cellule.
2. La taille de la cellule (surface, périmètre)
3. La forme de la cellule, son orientation...
4. Sa position, la distance qui la sépare des autres cellules...

Enfin, le logiciel attribue un numéro à chaque cellule et conserve ce même numéro d'image en image. Entre deux images, le même numéro est attribué aux objets qui partagent le maximum de superposition. Lorsque la cellule se divise, l'algorithme de segmentation attribue un nouveau numéro à chacune des cellules sœurs (Figure 52).



**Figure 52 Montage d'un suivi automatique de cellules.**

*Gauche* : images brutes d'une vidéo (de gauche à droite et de haut en bas). *Droite* : Segmentation et tracking. Chaque cellule garde son numéro jusqu'à la division.

Les résultats sont compilés sous forme de tableau dans un fichier au format CSV et une base de donnée SQL, contenant en moyenne une trentaine de variables par cellule.

#### II.5.d. Analyses statistiques et descriptives

Les données issues de l'analyse d'images par le logiciel CellProfiler sont analysées à l'aide du logiciel R (Team, 2011)<sup>1</sup>. C'est un logiciel open-source sous licence GNU qui fournit une large gamme d'outils statistiques et graphiques pour l'analyse de données (modélisation, test statistiques, analyse de séries temporelles, classification, clustering...). Il est possible d'écrire des scripts qui seront exécutés de façon automatique, ce qui constitue un avantage pour l'analyse de données redondantes comme celles obtenues dans ce travail.

#### ◆ Bibliothèques spécialisée pour R

Plusieurs bibliothèques spécialisées ont été utilisées afin d'obtenir de nouvelles fonctions.

Bibliothèque	Utilisation	Référence
<b>ggplot2</b>	Graphiques	(Wickham, 2009)
<b>msProcess</b>	Détection pics et vallées	(Gong et al., 2011)
<b>Ade4/Ade4TkGUI</b>	Analyse multivariée	(Thioulouse et al., 1997; Thioulouse and Dray,

<sup>1</sup> <http://www.r-project.org/>

		2007)
<b>MADE4</b>	Analyse multivariée	(Culhane et al.)
<b>Vegan</b>	Clustering	(Oksanen et al., 2011)
<b>PRADA</b>	Analyse cytométrie en flux	(Hahne et al., 2008)
<b>FlowCore</b>	Analyse cytométrie en flux	(Ellis et al., 2010)
<b>png</b>	Importation images	(Urbanek, 2011)
<b>FactoMinR</b>	clustering	(Husson et al., 2011)
<b>Moments</b>	Maxima et minima	(Komsta and Novomestky, 2011)

*Tableau 4 Bibliothèques additionnelles utilisées sous R.*

De plus, des fonctions et scripts personnels ont été écrits pour l'automatisation des analyses de résultats, notamment ceux de cytométrie en flux et de vidéo-microscopie.

#### ◆ *Analyses statistiques*

---

Les analyses statistiques classiques ont été réalisées avec les fonctions de base du logiciel (bibliothèque « Stats »). Les comparaisons de moyennes sont réalisées avec un test T de Student ou un test de Mann-Whitney-Wilcoxon selon la normalité et l'homosédasticité des données. Ces deux propriétés sont testées respectivement par un test de Shapiro et un test de Fisher-Snedecor ou d'ansari-Bradley. Les comparaisons de distributions sont réalisées avec un test de Kolmogorov-Smirnov et les tests de corrélation avec un test du coefficient de corrélation linéaire de Pearson ou de Spearman, selon la situation.

#### ◆ *Analyse de séries temporelles*

---

Une série temporelle consiste classiquement en une suite de valeurs mesurées à intervalle de temps régulier permettant de suivre l'évolution d'un paramètre au cours du temps. Des outils mathématiques permettent de comprendre son évolution et aussi de formuler des modèles susceptibles de prévoir son comportement futur. Ainsi, l'analyse de séries temporelles est utilisée dans de nombreux domaines tels que l'économie, la météorologie, l'électronique ou en médecine.



L'objectif est d'extraire de ces séries de données une tendance générale et d'identifier une éventuelle périodicité. Deux approches différentes mais non exclusives peuvent être utilisées : l'analyse temporelle et l'analyse spectrale (Shumway and Stoffer, 2006). Nous avons utilisé uniquement la première approche. Elle suppose que la corrélation entre chaque point de temps peut être mieux expliquée en termes de dépendance vis-à-vis des points de temps précédents. Il s'agit de prévoir l'évolution du système en fonction de son état actuel et passé. Les principaux outils mathématiques utilisés pour la mesure de cette dépendance sont l'auto-corrélation et l'inter-corrélation.

• *Variance, covariance et corrélation*

La variance est une mesure permettant de caractériser la dispersion d'une distribution autour de sa moyenne. Elle se définit comme *la moyenne des carrés des écarts à la moyenne* et se note :

$$var(x) = \sigma^2 x = \frac{\sum_{i=1}^n (x_i - \mu)^2}{n}$$

Avec  $\mu$  la moyenne et  $n$  le nombre de valeurs  $x$ .

La covariance évalue les sens de variation de deux variables et permet ainsi de quantifier la dépendance de ces deux variables. Elle mesure la force qui relie les deux variables :

$$Cov(x, y) = \frac{\sum_{i=1}^n [(X_i - \mu_x)(Y_i - \mu_y)]}{n}$$

Avec  $\mu_x$  et  $\mu_y$  les moyennes respectives des variables  $x$  et  $y$  et  $n$  le nombre de valeurs.

L'inconvénient de la covariance est que la valeur dépend des unités dans lesquelles sont exprimées les variables  $X$  et  $Y$ . C'est la raison pour laquelle on préfère le coefficient de corrélation dont l'étendue est  $(-1 : 1)$  :

$$\rho = \frac{cov(X, Y)}{\sqrt{var(X) \cdot var(Y)}}$$

Autre inconvénient de ce genre de mesure, c'est que la dépendance mesurée est forcément linéaire. Un coefficient de corrélation de 0 ne signifie pas qu'il n'existe aucune dépendance entre  $X$  et  $Y$  mais seulement qu'elle n'est pas linéaire.

• *Auto-covariance et auto-corrélation*

Par définition, l'auto-covariance se définit comme la covariance d'un signal  $x$  par rapport à lui-même, plus exactement une version décalée dans le temps de lui-même :

$$\gamma x(s, t) = cov(x_s, x_t) = E[(x_s - \mu_s)(x_t - \mu_t)]$$

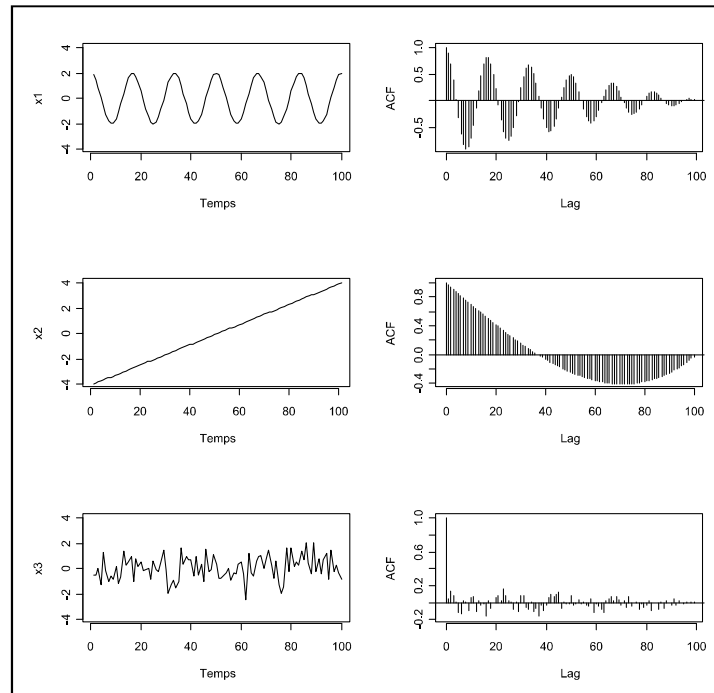
Où  $x_s$  et  $x_t$  sont deux points de temps, et  $\mu$  la moyenne temporelle. Si  $s = t$ , alors  $\gamma x(s, t) = var(x_t)$ .

L'auto-corrélation est souvent confondue avec l'auto-covariance mais correspond à cette dernière divisée par la variance et permet de s'affranchir du problème lié aux unités :

$$\rho(t, t - k) = \frac{\text{auto covariance}}{\text{variance}} = \frac{\gamma x(s, t)}{\sqrt{(var(xs))(var(xt))}}$$

L'ACF est une fonction dont le résultat est dans l'intervalle  $[-1,1]$  avec 1 indiquant une corrélation parfaite, c'est-à-dire que les signaux se recouvrent exactement quand le temps est décalé de  $k$ . Une corrélation de -1 indique en revanche une anti-corrélation parfaite. Dans R, on utilise la fonction  $ACF(x, \text{max.lag}=n)$  sur la série de données préalablement convertie en série temporelle par la fonction  $ts(x)$ .

La Figure 53 montre 3 exemples d'auto-corrélation. La colonne de gauche représente 3 séries temporelles et la colonne de droite la fonction d'auto-corrélation appliquée à ces séries, avec un décalage de 0 à 100 intervalles de temps (noté « Lag »).



**Figure 53** Trois signaux artificiels et leur auto-corrélation respective.

*La périodicité du signal sinusoïdal se retrouve dans son ACF. Plus le signal est constant, plus la décroissance de la fonction d'auto-corrélation est lente.*

Les trois exemples représentent de haut en bas : un signal sinusoïdal, un signal linéaire croissant et un signal aléatoire. Dans les trois cas, l'auto-corrélation au point 0 vaut 1 car on compare les signaux à eux-mêmes, sans décalage.

Dans le cas du signal aléatoire, par définition il n'est pas possible de prédire les valeurs suivantes. Le décalage d'une seule unité de temps suffit à perdre toute corrélation : l'ACF reste proche de 0.

L'observation est différente dans le cas du signal linéaire constant. Les valeurs sont prédictibles d'un point à l'autre et la corrélation se maintient malgré le décalage.

La série sinusoïdale est particulière puisque lorsque le décalage vaut  $\frac{1}{2}$  période, la corrélation est négative. Elle redevient positive après une autre  $\frac{1}{2}$  période. On retrouve ainsi la propriété de périodicité dans l'ACF.

Nous utiliserons dans ce travail la vitesse de décroissance de la fonction d'auto-corrélation pour caractériser les signaux de fluorescence obtenus. Nous estimerons le décalage nécessaire à l'obtention d'une auto-corrélation de 0.5, que nous appellerons « demi-vie de l'auto-corrélation » (Weinberger et al., 2008b).

• *Covariance et corrélation croisées*

Les fonctions de covariance croisée et de corrélation croisée (CCF) mesurent la corrélation entre deux signaux différents décalés dans le temps. En reprenant les définitions de l'auto-covariance et de l'auto-corrélation, on obtient les formules suivantes :

Covariance croisée :

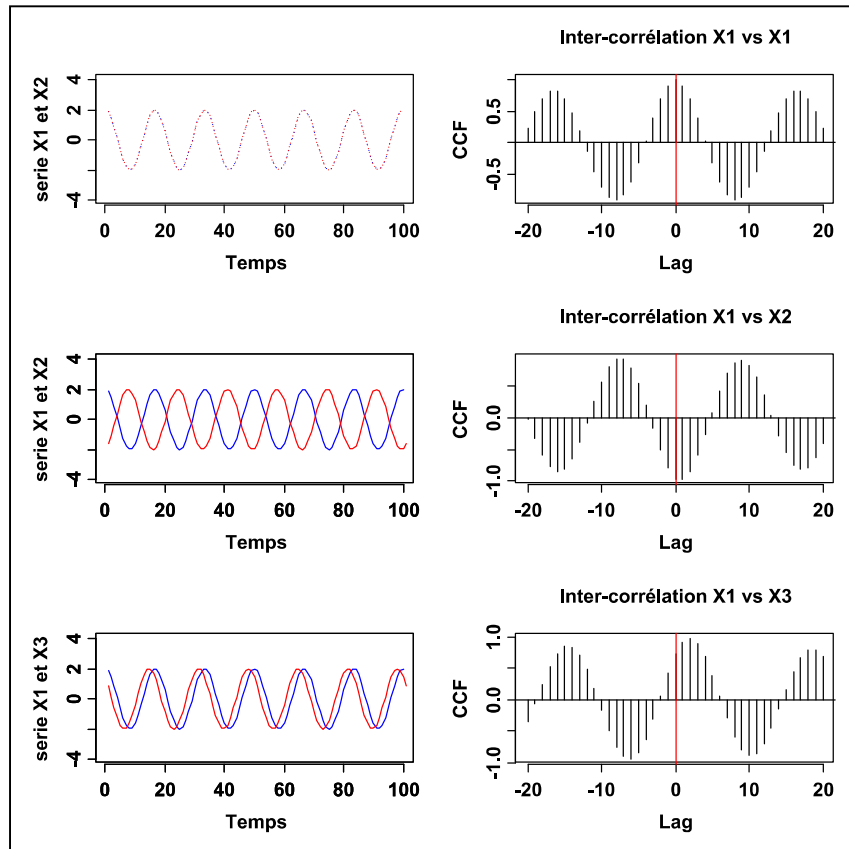
$$\gamma_{xy}(s, t) = cov(x_s, y_t) = E[(x_s - \mu_{x_s})(y_t - \mu_{y_t})]$$

Corrélation croisée :

$$\rho_{xy}(t, t - k) = \frac{\text{covariance croisée}}{\text{variance}} = \frac{\gamma_{xy}(s, t)}{\sqrt{(\text{var}(x_{s,s}))(\text{var}(y_{t,t}))}}$$

Comme l'auto-corrélation, les valeurs vont de -1 à 1, correspondant respectivement à une anti-corrélation parfaite et à une corrélation parfaite. L'inter-corrélation se calcule avec la fonction  $CCF(x, y, \text{max.lag}=n)$  sous R.

La Figure 54 décrit 3 exemples de superposition de signaux sinusoïdaux : une superposition parfaite, une anti-corrélée et une légèrement déphasée. La colonne de droite correspond à la fonction de corrélation-croisée entre chacun des signaux à des décalages allant de -20 à 20 intervalles de temps. La ligne rouge sur le graphique de corrélation croisée représente la corrélation lorsqu'il n'y a pas de décalage, c'est-à-dire tel que représenté sur le graphique de gauche correspondant.



**Figure 54 Exemple de calcul d'inter-corrélation.**

*Lorsque les signaux sont parfaitement superposés, la corrélation vaut 1. La figure d'inter-corrélation prédit qu'un décalage de 9 unités de temps à gauche ou à droite, va aboutir à une corrélation de -1. En effectuant ce décalage (1/2 période), les deux signaux sont parfaitement anti-corrélés.*

On note alors que la corrélation pour le premier cas est parfaite ( $=1$ ), les deux signaux étant superposés. L'inter-corrélation prédit néanmoins qu'un décalage de 9 unités de temps aboutira à une inter-corrélation de  $-1$ . Ce décalage est représenté par le second cas. les signaux sont parfaitement anti-corrélés ce qui se traduit par une corrélation de  $-1$ . Comme il s'agit ici de signaux sinusoïdaux, il suffit de se décaler d'une demi-période pour retrouver une corrélation parfaite.

Le troisième cas est intermédiaire, les signaux rouge et bleu sont légèrement déphasés ce qui se traduit par une corrélation inférieure à 1.

Dans ce travail, nous avons utilisé l'ACF et la CCF (fonctions  $acf(x)$  et  $ccf(x,y)$  dans R) afin d'analyser les mesures de fluorescence au cours du temps, obtenues par vidéo-microscopie sur notre modèle de cellules bi-fluorescentes. Nous analyserons l'évolution de l'eCFP et de l'eYFP séparément grâce à une ACF sur chacun des signaux puis l'un par rapport à l'autre grâce à la CCF.

### III. Protocoles Spécifiques

---

#### III.1. Détermination de la durée de vie des protéines eCFP et eYFP

---

Des cellules exprimant les deux gènes rapporteurs eCFP et eYFP sontensemencées à une densité de  $5 \times 10^4$  cellules par  $\text{cm}^2$  dans une plaque 6 puits. Le lendemain, le milieu de culture est changé. Les cellules sont ensuite traitées pendant 1 à 5 heures avec du cycloheximide (CHX) à une concentration finale de  $50 \mu\text{g/ml}$ . Cette drogue bloque l'étape de traduction pendant la synthèse protéique. Il n'y a donc plus de néo-synthèse de eCFP ni eYFP pendant la durée du traitement ce qui nous permet de suivre la dégradation de chacune des protéines. La fluorescence moyenne par cellule est analysée par cytométrie en flux comme décrit précédemment au bout de 1, 2, 3, 4 et 5 heures de traitement. Un contrôle avec adjuvant seul (DMSO) sert de référence.

Un modèle de régression exponentielle est appliqué aux données obtenues pour déterminer la demi-vie de chaque protéine. Celle-ci correspond à la durée nécessaire pour que la fluorescence diminue de 50% par rapport au contrôle adjuvant.

#### III.2. Détermination de la durée de vie des ARNm

---

Des cellules exprimant les deux gènes rapporteurs eCFP et eYFP sontensemencées à une densité de  $5 \times 10^4$  cellules par  $\text{cm}^2$  dans une plaque 6 puits. Le lendemain, le milieu de culture est changé. Les cellules sont ensuite traitées pendant 30 minutes à 4 heures avec de l'actinomycine D (ActD) à une concentration finale de  $5 \mu\text{g/ml}$ . Cette drogue bloque l'étape de transcription pendant la synthèse des ARNm. Il n'y a donc plus de néo-synthèse d'ARNm pendant le temps du traitement ce qui nous permet de suivre la dégradation de chacun des transcrits. Les cellules sont détachées classiquement, comptées, et les ARN totaux extraits puis reverse-transcrits comme décrit précédemment. La quantité de transcrits est estimée pour chaque temps de traitement par qRTPCR en utilisant les

amorces YFP.fw, YFP.Rv, CFP.fw, CFP.Rv et la sonde CFP/YFP.Pr. Un contrôle avec adjuvant seul (éthanol 100%) sert de référence.

Le nombre de transcrits est normalisé par le nombre de cellules dans chaque condition afin de tenir compte de la mortalité cellulaire. Nous n'avons pas utilisé de gène de ménage pour standardiser la quantité d'ARN en fonction des conditions. L'Actinomycine-D possède une action globale sur la cellule, y compris sur la transcription des gènes de ménage classiquement utilisés, tels que l'actine, la GAPDH ou l'albumine. Un modèle de régression exponentielle est appliqué aux données obtenues pour déterminer la demi-vie de chaque transcrit. Celle-ci correspond à la durée nécessaire pour que la quantité de transcrits diminue de 50% par rapport au contrôle adjuvant.

# Résultats

---



## I. Étude des fluctuations phénotypiques à long terme.

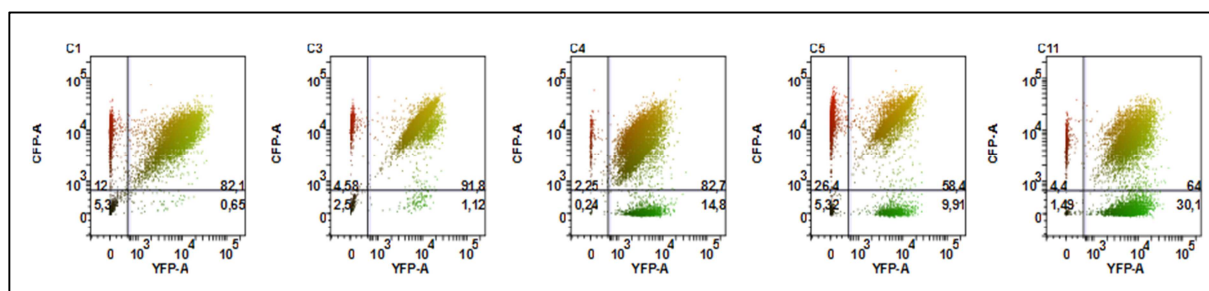
Dans cette première partie, nous avons étudié l'apparition de l'hétérogénéité phénotypique en analysant la dynamique qui permet de passer d'un phénotype unique bien caractérisé à une population mixte arborant une diversité de phénotypes. Nous avons fait appel à un modèle cellulaire proche de celui décrit en introduction (Page 29).

### I.1. Présentation et caractérisation du modèle cellulaire

#### I.1.a. Description du modèle et observations antérieures

Ce modèle, conçu en 2007 afin d'étudier l'origine de l'hétérogénéité phénotypique, a fait l'objet d'une publication (Neildez-Nguyen et al., 2008). Il s'agit de cellules HER 911 transfectées successivement par deux constructions plasmidiques commerciales. Des cellules exprimant les deux transgènes ont été clonées et le nombre de copies de chacun des transgènes estimé par PCR quantitative. Seules les populations clonales ne possédant qu'une seule copie de chacun des transgènes par cellule ont été conservées et analysées.

Lors de cette première étude, 16 clones avaient été conservés. Nous les appellerons par la suite les clones « parentaux ». L'observation majeure est que, dès les premiers jours après le clonage de cellules pour établir les clones parentaux, la population isogénique alors obtenue était déjà hétérogène en termes d'expression de chacun des deux transgènes (Figure 55).



**Figure 55 Expression d'eCFP et eYFP dans les clones parentaux.**

Analyse par cytométrie en flux des clones C1, C3, C4, C5 et C11 exprimant les protéines non-déstabilisées.

En effet, des cellules exprimant eCFP et eYFP à différents niveaux étaient visibles dans la population, ainsi que certaines cellules exprimant uniquement un seul des transgènes. Dans une plus faible proportion, certaines cellules n'expriment aucun des transgènes.

Lors de cette première étude, les auteurs ont montré que dans ce modèle cellulaire, il n'y avait pas de corrélation entre le niveau moyen d'expression des deux transgènes entre les différents clones parentaux obtenus. L'hypothèse d'un effet de position est alors la plus probable. Les transgènes étant identiques (même promoteur et séquence codante très proche), on s'attendait intuitivement à une expression équivalente si uniquement des mécanismes spécifiques au promoteur étaient mis en jeu. La seule différence entre les clones parentaux étant la position des transgènes, le rôle prédominant du contexte chromatinien semble indéniable dans la régulation du niveau d'expression des transgènes.

Dans un second temps, la variabilité au sein de chaque clone parental a été évaluée. Dans chacun d'eux, la différence d'expression des deux transgènes entre les cellules est importante et va à l'encontre de ce qui est espéré pour une population clonale. En effet, malgré l'identité de leur génome et de leur condition de culture (micro environnement), les cellules expriment à différents niveaux chacun des transgènes. Le niveau d'expression varie non seulement significativement au sein de chaque population parentale, mais aussi entre chacune des populations. Chaque clone parental est donc caractérisé par le niveau d'expression moyen des deux transgènes des cellules qui le compose mais aussi par la variabilité d'expression entre cellules. Cette observation suggère que le site d'intégration est impliqué dans la régulation du niveau d'expression mais participe aussi à sa variabilité. Les auteurs ont identifié la source de variation comme étant majoritairement liée à la position du transgène. La part extrinsèque de la variabilité observée est la même d'une population à l'autre indiquant une certaine homogénéité des conditions générales de culture. Inversement, la part intrinsèque diffère d'un clone à l'autre, indiquant alors que la source de variation n'est pas liée aux transgènes (i.e. le promoteur), mais à un paramètre qui diffère entre les deux. La seule différence dans ce modèle est le site d'intégration des transgènes dans le génome hôte et cette variabilité intrinsèque reflète probablement l'état

chromatinien spécifique de chacun des sites d'intégration ou plus largement l'organisation spatiale nucléaire à ce locus.

Enfin, les auteurs ont montré que des cellules exprimant aucun des transgènes sont présentes et peuvent les ré-exprimer spontanément de façon synchronisée ou indépendante. La décongélation de cellules négatives a permis de faire exprimer les deux transgènes dans plusieurs cas. Des conditions de culture stressantes, telle qu'une haute densité, sont susceptibles d'augmenter la variabilité de chaque population clonale.

### *1.1.b. Caractérisation du modèle et contrôles préliminaires*

Avant de poursuivre le travail avec les clones parentaux présentés dans la partie précédente, nous avons procédé à des contrôles préliminaires afin de mieux caractériser ces clones.

#### *◆ Vérification de la présence des transgènes*

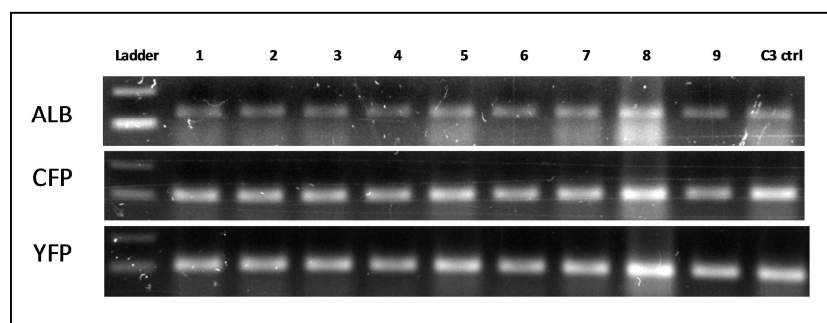
Les populations clonales parentales issues de ce modèle sont très hétérogènes en termes d'expression des deux transgènes présents dans leur génome. Certaines expriment les transgènes à différents niveaux pendant que d'autres ne les expriment pas. Il est alors possible que ceux-ci aient été éliminés du génome. Afin de vérifier cette hypothèse, une analyse par PCR a été réalisée sur plusieurs sous-populations clonales n'exprimant aucun des transgènes à l'aide d'amorces spécifiques de chacun des transgènes eCFP et eYFP (Tableau 5, Figure 56).

Cible génomique	Amorces
Albumine	Alb.F et Alb.R
eCFP	CFP.F et CFP.R
eYFP	YFP.F et CFP.R

**Tableau 5 Amorces de screening pour la présence des transgènes.**

L'albumine est utilisée comme contrôle positif de PCR pour chaque sous-clone testé. Le clone parental C3, double fluorescent, est lui utilisé comme témoin positif (puits 10).

L'électrophorèse sur gel d'agarose révèle que pour tous les sous-clones négatifs analysés une bande spécifique à eCFP et à eYFP est présente.



**Figure 56 Vérification de la présence des transgènes par PCR.**

*L'amplification a été réalisée sur plusieurs clones n'exprimant pas les transgènes (numéro 1 à 9) et sur le clone parental C3. Marqueur de taille : 100 et 200 bp.*

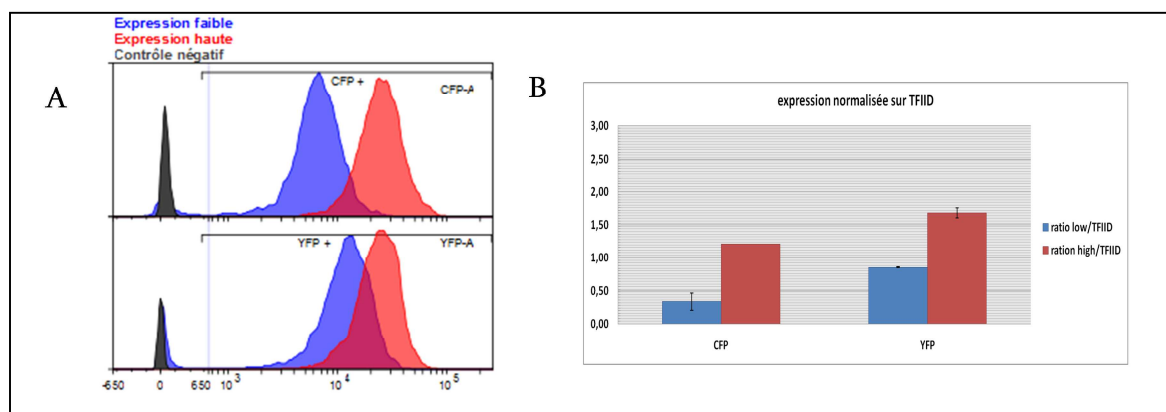
Cette analyse PCR confirme que tous les sous-clones analysés possèdent les deux transgènes. L'absence d'expression est donc probablement liée à une répression active de l'expression du/des transgène(s). Ce type d'analyse a été réalisé au cours de l'expérience chaque fois qu'il a été jugé utile de procéder à cette vérification.

#### ◆ *Régulation transcriptionnelle de l'expression des transgènes*

Nous avons constaté que dans une population cellulaire clonale, des disparités phénotypiques apparaissent entre les individus qui la composent. Dans l'objectif de vérifier que le niveau d'expression est corrélé au niveau de transcription, une analyse qRT-PCR est réalisée sur deux populations triées en fonction du niveau d'expression des deux transgènes eCFP et eYFP.

Une population exprimant fortement les deux transgènes et une seconde les exprimant environ dix fois moins ont été générées par tri-cellulaire et entretenues en culture classique jusqu'à obtention d'un nombre suffisant de cellules. Une analyse par cytométrie avant l'extraction d'ARN a été réalisée pour confirmer les phénotypes (Figure 57, A). Un standard interne, le TFIID, a été utilisé pour la normalisation de l'expression. La quantification est réalisée avec une gamme de dilutions de plasmides pour chacun des

transgènes et le contrôle. Le ratio transgène/TFIID est alors l'expression normalisée des transgènes entre les deux conditions (Figure 57, B).



**Figure 57** Corrélation entre MFI et quantité d'ARNm

Le niveau de fluorescence (A) et l'expression (B) ont été quantifiés dans une population de cellules exprimant fortement (en rouge) ou faiblement (en bleu) chacun des transgènes.

Le niveau d'ARNm des deux transgènes, dans les deux catégories, corrèle bien avec le niveau de fluorescence dans les deux populations testées (Tableau 6).

	Expression eCFP			Expression eYFP		
	Faible	Haute	ratio	Faible	Haute	ratio
FCM (fluorescence médiane)	6930	26776	<b>3.86</b>	12098	24926	<b>2.06</b>
qRTPCR (expression normalisée)	0.34	1.21	<b>3.55</b>	0.85	1.68	<b>1.97</b>

**Tableau 6** Quantification de l'expression des transgènes.

La différence observée, entre les deux phénotypes et pour les deux transgènes, est comparable que l'on regarde le niveau des ARNm et de la fluorescence des protéines.

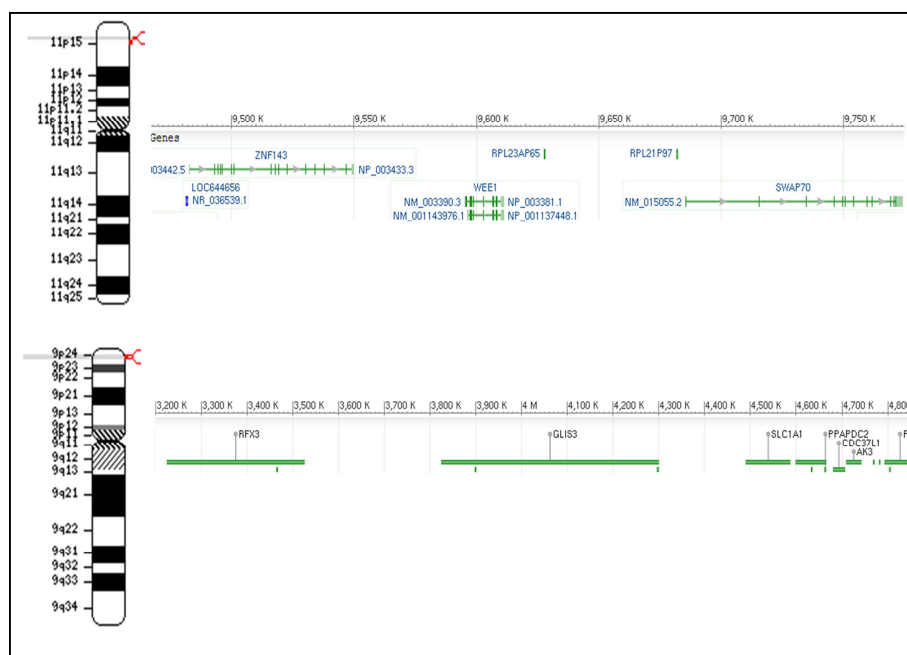
On peut donc conclure que la régulation du niveau d'expression de chacun des transgènes est majoritairement liée à la transcription. De même, la variabilité d'expression observée entre les cellules positives (exprimant faiblement ou non les transgènes) est probablement le résultat de la stochasticité moléculaire à l'étape de transcription (micro-hétérogénéité).

#### ◆ Identification du site d'intégration des deux transgènes

Dans ce modèle cellulaire, chacun des transgènes est présent en une seule copie dans le génome. La position d'un gène dans le génome joue un rôle important pour son

expression. Il est fréquent d'observer la répression de l'expression d'un gène s'il se trouve dans une région chromatinienne constituée d'hétérochromatine qui par sa condensation isole l'ADN des différents facteurs de transcription. Le site d'intégration des deux transgènes eCFP et eYFP a été identifié dans deux clones parentaux utilisés pour la suite des expériences. Le site d'ouverture de chacun des plasmides est différent. Par exemple, dans le clone C3, eYFP est ouvert en position 1828 et eCFP en position 4375.

La jonction entre l'ADN génomique et l'ADN plasmidique des transgènes a été amplifiée par splinkerette PCR. Les produits d'amplification ont été clonés par transformation de bactéries et un séquençage a permis d'identifier le point d'ouverture de chaque plasmide lors de l'intégration et le site d'intégration dans le génome hôte (Figure 58).



**Figure 58** Cartographie des sites d'intégration dans le clone parental C3.

Le transgène eCFP est intégré dans un intron du gène *WEE1* sur le chromosome 11. eYFP est intégré dans un intron du gène *GLIS3* sur le chromosome 9.

Dans le cas du clone C3, le transgène eCFP est localisé sur le chromosome 11 alors que le transgène eYFP se trouve sur le chromosome 9. Dans les deux cas, le transgène est intégré dans l'intron d'un gène, *WEE1* et *GLIS3* pour eCFP et eYFP respectivement.

*WEE1* est un gène codant une protéine nucléaire appartenant à la famille des Serine/Tyrosine kinase. Ce gène est entouré de chaque côté par les gènes *ZNF143* et *SWAP70*, situés de part et d'autre, respectivement à 45 et 74 kb.

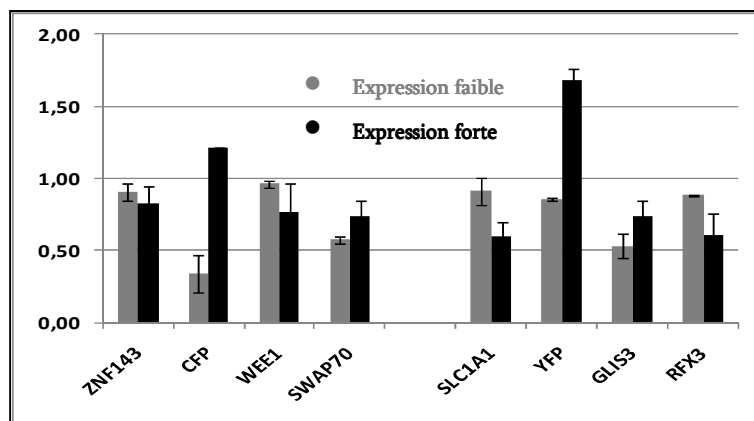
*GLIS3* code une protéine nucléaire appartenant à la famille des doigts de zinc « GLIS-like ». Elle a une fonction ambivalente d'activateur mais aussi de répresseur de la transcription. Le gène est entouré d'un bord par *SLC1A1* (190kb) et de l'autre par le gène *RFX3* (300kb).

#### ◆ *Expression des gènes partageant le locus des transgènes*

---

Le site d'intégration des deux transgènes a été identifié dans le clone C3. Afin de déterminer si le contexte génomique joue un rôle important dans les différences d'expression génique que nous observons, une analyse de l'expression des gènes adjacents à nos transgènes dans le génome a été réalisée. En effet, la position dans le génome est très importante compte-tenu de la dynamique de la chromatine. Certaines régions sont plus ou moins susceptibles d'être sous forme d'euchromatine (constitutive ou facultative) ou inversement d'hétérochromatine. De plus, la réversibilité de cet état chromatinien peut influencer l'expression des gènes qui s'y trouvent, en termes de niveau d'expression et de variabilité.

Les gènes interrompus par l'intégration des transgènes ainsi que les gènes les plus proches ont été analysés par qRTPCR dans une population exprimant fortement les transgènes et une seconde les exprimant plus faiblement (Figure 59). La différence d'expression entre les deux populations est la même que dans la Figure 57. Les gènes étudiés sont : *ZNF143*, *WEE1*, *SWAP70*, *GLIS3*, *SLC1A1* et *RFX3*. Les amorces utilisées sont détaillées dans le Tableau 3. Comme précédemment, nous confirmons que le niveau d'ARNm de chacun des transgènes (*eCFP* et *eYFP*) corrèle avec le niveau de protéines (i.e. la fluorescence).



**Figure 59** Expression des gènes autour du site d'intégration d'eCFP et d'eYFP.

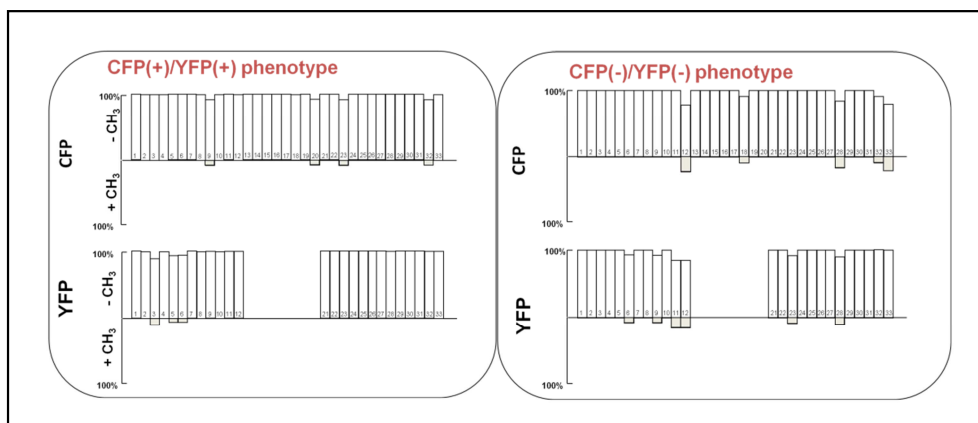
Le niveau d'expression est mesuré dans des cellules exprimant fortement ou faiblement les transgènes.

Les gènes interrompus par les transgènes (*GLIS3* et *WEE1*) sont toujours exprimés sans différence significative entre les deux populations considérées. Les gènes adjacents à *WEE1* / *eCFP* ne semblent pas affectés non plus et leur expression est identique entre les deux populations. En revanche, *SLC1A1* et *RFX3* semblent sous-exprimés dans la population High. Pour chaque gène, au moins deux allèles sont présentes. Dans ce modèle, une seule intégration a eu lieu pour chacun des transgènes, interrompant un allèle des gènes mentionnés plus tôt. La limite majeure est que nous ne savons pas si l'expression observée est liée à l'allèle du gène ciblé, qui continue à être actif malgré l'intégration, ou au second allèle du gène.

#### ◆ Analyse de la méthylation du promoteur CMV

La méthylation du promoteur CMV a été réalisée sur deux populations de cellules : la première exprimant les deux transgènes et sur une seconde ne les exprimant pas. Suite à la conversion de l'ADN génomique de ces deux populations par le bisulfite, l'amplification par PCR et le clonage des produits de PCR, 26 colonies bactériennes blanches ont été étudiées dans les deux conditions. Pour chacune d'elles, 33 îlots CpG ont été étudiés pour le transgène eCFP et 21 pour le transgène eYFP (1-12 et 24-33) (Figure 60). La différence est liée à la présence du site de multi-clonage entre le promoteur CMV et la séquence codante dans le plasmide eCFP.





**Figure 60 Analyse de la méthylation des îlots CpG.**

*33 îlots ont été analysés pour eCFP et 21 pour eYFP (barres verticales). Les îlots CpG des deux transgènes, qu'il soit exprimé ou non, sont complètement déméthylés dans les 26 colonies analysées.*

Aucune différence n'est observée entre les cellules qui expriment ou non les transgènes. Dans les deux conditions et pour les deux transgènes, l'ensemble des îlots CpG est entièrement dé-méthylé. D'après ce que nous avons vu en introduction, l'hypo-méthylation de l'ADN est propice à l'expression du gène.

Cette étape de caractérisation nous a permis de nous assurer que dans un premier temps, l'absence d'expression n'est pas la conséquence d'une absence de transgène. Nous avons aussi montré que le niveau de fluorescence corrèle avec le niveau d'ARNm et donc que la régulation de l'expression est majoritairement transcriptionnelle. Enfin, l'analyse des sites d'intégration et de l'expression des gènes adjacents au site d'intégration montre que le contexte chromatinien est propice à l'expression.

## I.2. Émergence de la variabilité phénotypique à long terme

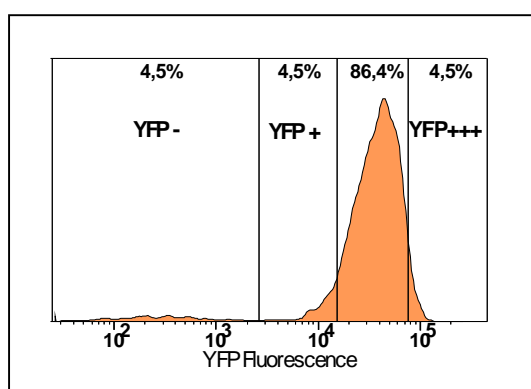
Dans le modèle cellulaire, chaque clone parental est hétérogène : certaines cellules expriment ou non les transgènes et quand elles les expriment, plusieurs niveaux d'expression sont visibles. Dans chaque clone parental, la proportion de cellules exprimant ou non les transgènes semble stable. Mais cette apparente stabilité est paradoxale pour expliquer l'apparition de l'hétérogénéité.

Dans le but d'étudier l'émergence de la variabilité au sein d'une population clonale, nous avons sous-cloné des cellules de clones parentaux. Cette étape permet de démarrer l'expérience avec des cellules dont le niveau d'expression des deux transgènes est connu. Par la suite, nous avons mesuré à différents temps, l'expression des transgènes dans les populations clonales issues de ces cellules isolées, grâce à la fluorescence des protéines synthétisées.

### *1.2.a. Établissement de sous-clones phénotypiquement caractérisés*

Parmi les différents clones parentaux, nous avons sélectionné trois d'entre eux (C1, C3 et C4) qui présentent la plus forte hétérogénéité phénotypique lors de leur analyse en cytométrie en flux. Rappelons que chacun d'eux possède une seule copie des deux transgènes à des positions différentes dans le génome. Les cellules provenant de ces différents clones ont été préparées afin de les individualiser et ont été isolées en fonction de leur phénotype grâce à un trieur de cellules (Moflo cytometry). Pour des raisons techniques, les cellules ont été triées uniquement en fonction de l'expression du transgène eYFP. L'objectif étant d'étudier l'apparition des cellules positives et négatives, nous avons isolé ces deux catégories de cellules.

Parmi les cellules positives, nous avons fait la distinction entre les cellules exprimant fortement (YFP<sup>+++</sup>) et faiblement (YFP<sup>+</sup>) l'eYFP. Elles ont été triées à partir des extrémités (~5%) de la distribution de la fluorescence eYFP de la population parentale (Figure 61).

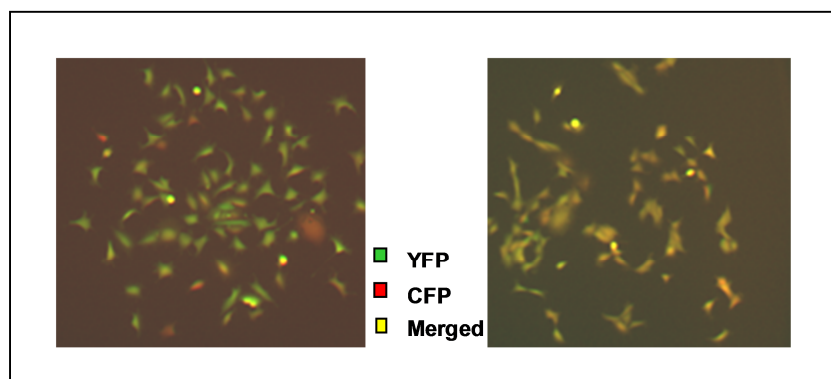


**Figure 61 Exemple de distribution de la fluorescence eYFP d'un clone parental.**

*En considérant un seul transgène, le clone parental est composé de 2 sous-populations, une positive et une négative. La population positive peut être subdivisée en 3 : les 2 extrémités de la distribution et le centre. Ces populations seront appelées YFP-, YFP+ et YFP+++.*

D'une part, les cellules négatives et positives vont permettre d'étudier ce que l'on appelle la macro-hétérogénéité. Elle correspond à la présence de catégories discrètes de phénotypes ou d'états (Huang, 2009). D'autre part, la distinction entre cellules exprimant fortement ou faiblement l'eYFP va permettre d'étudier la micro hétérogénéité dans les populations, qui correspond aux variations dans une population apparemment homogène (monomodale).

Pour chaque clone parental, 1 plaque 96 puits a été réalisée, avec 32 puits par condition (YFP-, YFP+ et YFP+++). Ces cellules isolées ainsi que leur descendance seront appelées « sous-clones » pour ne pas les confondre avec les clones parentaux. Après 5 à 6 jours de culture, la présence de cellules dans chaque puits a été contrôlée par microscopie (Figure 62). L'expression de l'eCFP et/ou de l'eYFP est contrôlée avec les filtres adaptés. Cette vérification optique est essentielle car une analyse par cytométrie est impossible à ce stade, compte-tenu du faible nombre de cellules.



**Figure 62** Contrôle de l'expression des transgènes eCFP et eYFP dans 2 sous-clones.

*Microscopie à fluorescence 10x, J+6. Une soixantaine de cellules sont observables à ce stade et présentent déjà une certaine diversité de niveaux d'expression des transgènes.*

Le rendement du clonage lors de cette expérience a été relativement faible puisque plusieurs cellules isolées n'ont pas proliféré après leur ensemencement. La présence de cellules est détectée dans seulement la moitié des puits après 5-6 jours de culture, les autres puits ne contenant pas de cellules, ou des cellules visiblement mortes. Néanmoins, 44 sous-clones ont été obtenus à l'issue du tri cellulaire (Tableau 7).

Phénotype	Nombre de sous-clones	eCFP +	eYFP +
eYFP-	14	13	0
eYFP+	18	17	18
eYFP+++	12	9	12

**Tableau 7 Bilan des sous-clones obtenus après tri cellulaire.**

*L'expression de l'eYFP est compatible avec le critère de sélection eYFP positif ou eYFP négatif.*

Tous les sous-clones YFP+ et YFP+++ expriment bien le transgène eYFP. La majorité d'entre eux exprime également le transgène eCFP. La population parentale étant constituée de près de 80% de cellules exprimant les deux transgènes, il n'est donc pas étonnant de trier les deux phénotypes simultanément (Figure 55). De la même façon, aucun des sous-clones eYFP- n'exprime le transgène eYFP à ce point de temps. En revanche, ils expriment quasiment tous l'eCFP. Ceci n'est pas étonnant car 95 à 99% des cellules eYFP négatif sont eCFP positif dans la population parentale.

Parmi les sous-clones obtenus, un seul n'exprime aucun transgène et 4 n'expriment pas l'eCFP. L'eCFP n'étant pas un critère lors du tri cellulaire, le niveau de fluorescence lors du tri n'est pas connu. Néanmoins, l'observation microscopique nous renseigne sur l'expression ou non de ce transgène.

Les cellules arrivées à 80% de confluence dans le puits de clonage sont transférées dans un des puits d'une plaque 6 puits et entretenues en routine dans ce support pour faciliter la manipulation de tous les sous-clones simultanément. Une attention particulière est portée lors du passage des cellules afin de ne pas avoir de contamination croisée entre les puits d'une même plaque. Les sous-clones de chaque phénotype issus du même clone parental sont cultivés dans des plaques différentes. Cette précaution permet d'éviter d'éventuelles contaminations de sous-clones de phénotypes différents. De même, la densité cellulaire a été surveillée afin de ne pas atteindre une confluence supérieure à 80% de la surface de culture.

Une vingtaine de jours est nécessaire pour obtenir 80% de confluence dans la plaque 6 puits à partir du jour de clonage. C'est un délai non négligeable que nous discuterons plus

tard. Chacun des sous-clones obtenus est cultivé de façon standard et analysé par cytométrie en flux pendant 2 mois, à raison d'une analyse par semaine en moyenne. Chaque jour, des billes fluorescentes sont utilisées pour calibrer l'appareil et garantir les mêmes conditions d'acquisition. Lorsque les cellules arrivent à 80% de confluence dans le support de culture, elles sont détachées et préparées pour l'analyse en cytométrie. Un dixième des cellules est réensemencé dans un puits de plaque 6 puits ( $10^5$  cellules) et le reste analysé (environ  $9 \times 10^5$  cellules).

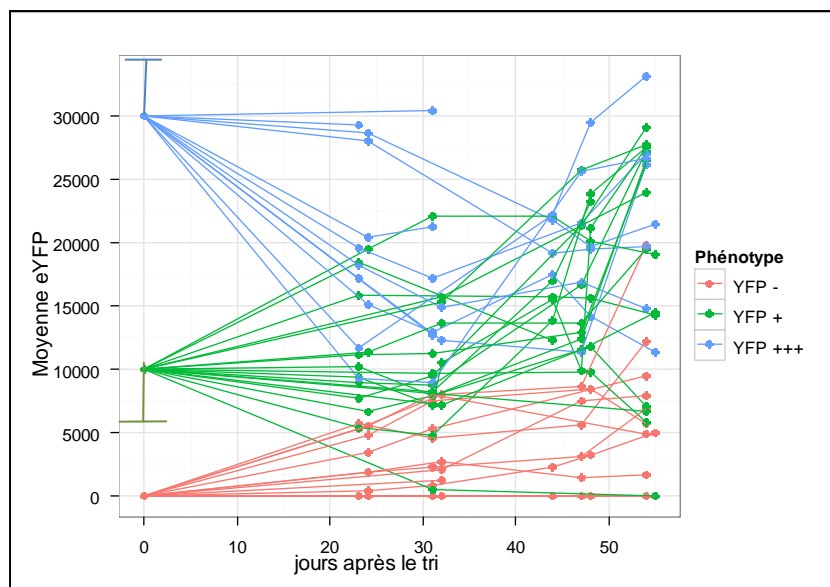
### *1.2.b. Évolution du niveau de fluorescence moyen des sous-clones*

---

Plusieurs indicateurs peuvent être pris en considération pour caractériser la diversité dans chaque sous-population clonale. Dans un premier temps, nous avons suivi la fluorescence moyenne de l'expression du transgène eYFP dans chaque sous-population, en fonction du temps. La moyenne permet d'observer l'évolution de la population entière, malgré les limites évoquées dans l'introduction. Cet indicateur ne met pas en évidence les comportements atypiques et réduit le comportement de chaque individu à celui de la population. Néanmoins, dans une première approche, c'est l'indicateur le plus simple à calculer et qui donne une vision d'ensemble.

Afin d'étudier les fluctuations à long terme de l'expression des deux transgènes, le niveau moyen de la fluorescence d'eYFP (MFI) de chaque sous-clone est calculé à chaque point de temps (Figure 63). Le niveau initial de fluorescence a été estimé en analysant les clones parentaux sur l'appareil d'analyse, le jour du tri, effectué sur un appareil différent. Les clones « YFP +++ » ont un niveau moyen supérieur ou égal à 30 000, les sous-clones « YFP + » un niveau compris entre 10 000 et 5000. Bien entendu les sous-clones « YFP - » ont un niveau de fluorescence équivalent au bruit de fond.

Ces trois catégories sont représentées respectivement par des lignes bleues, vertes et rouges sur la figure suivante.



**Figure 63 Évolution de la fluorescence moyenne eYFP dans les sous-clones**

*La MFI des trois types de sous-clones tend vers une valeur intermédiaire. Certains sous-clones positifs (en vert et bleu) semblent cesser d'exprimer le transgène. La majorité des sous-clones YFP- ont une MFI qui augmente au cours du temps.*

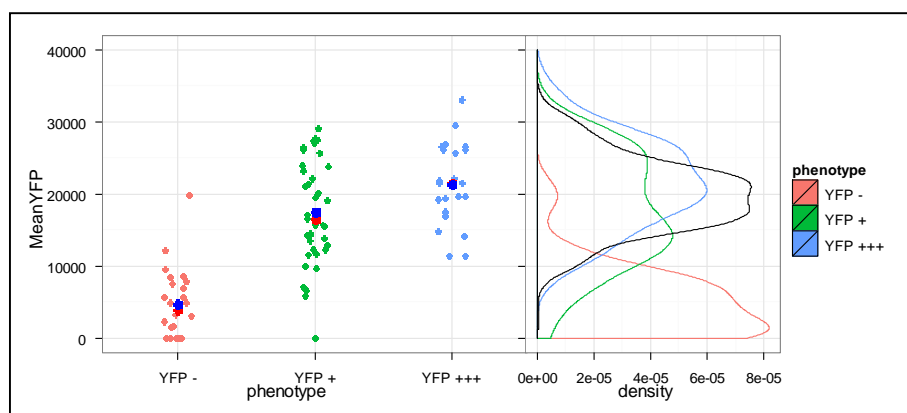
L'observation majeure dans cette expérience est l'impression de convergence qui ressort de ces tracés, quelque soit l'origine des cellules.

On remarque clairement que la majorité des sous-clones dérivés de cellules négatives (rouge) présente une augmentation de la MFI. Elle traduit une expression du transgène qui était réprimé lors du tri. Sur les 14 clones négatifs isolés lors du tri, 10 expriment l'eYFP au bout des deux mois de l'expérience. Le niveau moyen de fluorescence des ces sous-clones reste toutefois inférieur à celui des sous-clones positif.

Les deux catégories de sous-clones positifs ont tendance à converger vers une moyenne de fluorescence intermédiaire, correspondant à la moyenne de fluorescence de la population parentale dont ils sont issus. Les sous-clones « YFP+ » ont une MFI qui augmente ou reste stable au cours du temps tandis que l'expression du transgène eYFP dans les sous-clones « YFP+++ » diminue. On note également un sous-clone « YFP+ » dont le niveau de fluorescence décroît, pour ne plus être détectable à la fin de l'expérience.

Après un délai de 25 jours, soit environ 20 divisions cellulaires, la MFI de chaque type de sous-clones reste encore comparable au niveau d'expression initial. On peut ainsi reconnaître le phénotype d'origine de la majorité des sous-clones. Néanmoins, après 55 jours d'expansion, il est difficile de faire la différence entre des cellules dont le phénotype

d'origine était « YFP+ » (vert) ou « YFP+++ » (bleu) (Figure 64). Le niveau moyen d'expression de chaque sous-clone est très hétérogène et on observe un mélange de tous les sous-clones, quelque soit leur origine phénotypique lors du tri.



**Figure 64 Niveau moyen de fluorescence de chaque sous-clone (J45 à 55).**

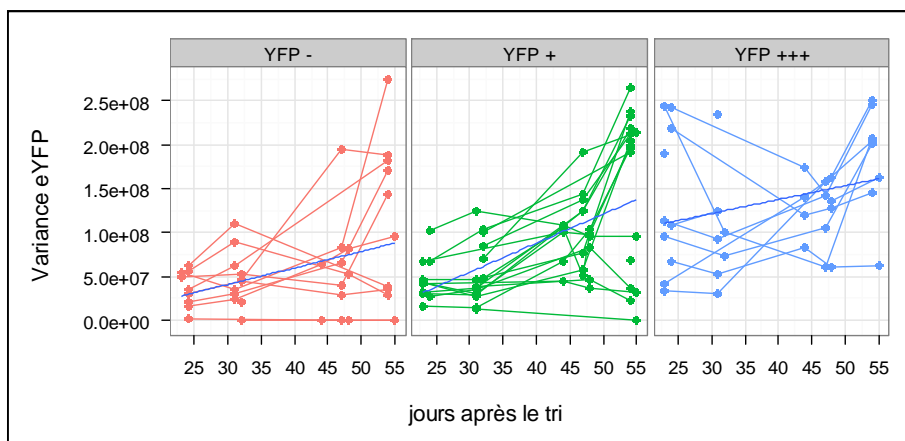
*A la fin de l'expérience, il est difficile de faire la différence entre les clones YFP+ et YFP+++.*

*Point rouge : médiane, point bleu : moyenne. Courbe noire : distribution de la fluorescence dans la population parentale.*

La distribution de la moyenne de fluorescence eYFP des clones « YFP+ » et « YFP+++ » se rapproche fortement de celle de la population d'origine (courbe noire).

### *1.2.c. Évolution de la variabilité des sous-clones*

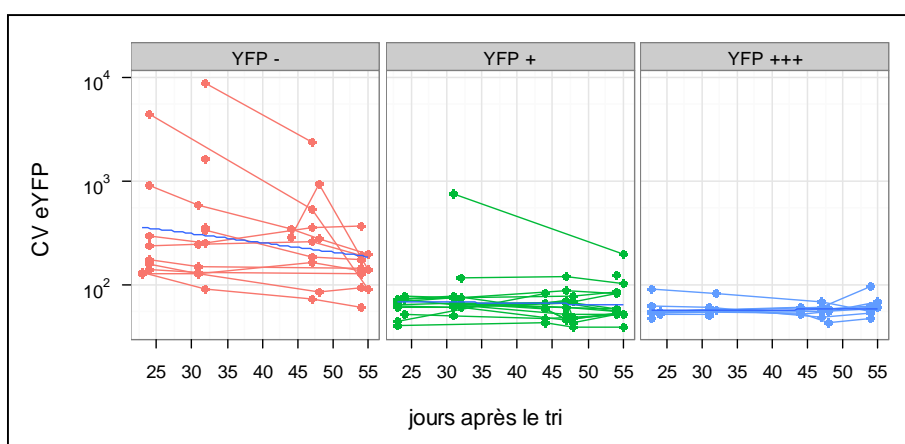
Le suivi de la moyenne de fluorescence de la population nous informe uniquement sur l'évolution globale de celle-ci sans nous donner d'aperçu précis des individus qui la composent. Ce critère n'est donc pas le plus adapté à la description de l'hétérogénéité qui s'installe dans une population. La variance permet d'évaluer dans quelle proportion la fluorescence de chaque individu s'éloigne de la moyenne de la population à laquelle il appartient. Néanmoins, comparer des variances n'a d'intérêt que si les moyennes sont identiques. Dans le cas contraire, la variance n'aura pas le même impact. De plus, en général la variance augmente avec la moyenne et il est difficile de juger de son importance (Figure 65).



**Figure 65 Évolution de la variance dans les 3 types de sous-clones.**

*La variance augmente de façon similaire dans les clones YFP- et YFP + au cours du temps. L'augmentation est plus modérée dans les clones YFP +++.*

Dans cette expérience, le niveau moyen de fluorescence des différents sous-clones varie beaucoup et une comparaison de variance n'aura pas d'intérêt majeur. Pour remédier à cette limite, on préfère travailler avec le coefficient de variation (CV) de la fluorescence ou la variance normalisée (NV). Ces deux indices sont normalisés par la moyenne et peuvent donc être utilisés pour la comparaison de variabilité de conditions différentes (Figure 66).



**Figure 66 Évolution du coefficient de variation dans les 3 types de sous-clones.**

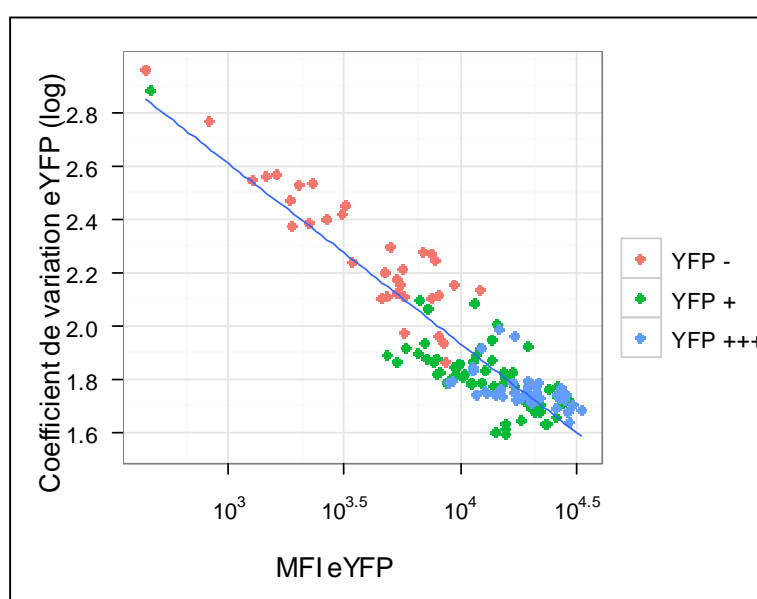
*Le CV est très élevé dans les sous-clones YFP – compte-tenu de leur fluorescence faible. Il diminue au cours du temps puisque le niveau de fluorescence de certains d'eux augmente. Les sous-clones positifs sont beaucoup moins variables, notamment les YFP +++.*

Nous remarquons que le niveau de variation n'est pas le même entre les catégories de sous-clones. Le niveau de fluorescence dans les sous-clones positifs (en vert et bleu) semble moins variable que dans les négatifs (rouge). Dans les sous-clones YFP+, la



moyenne augmente au cours du temps ainsi que la variance. Le coefficient de variation est donc stable. Dans les sous-clones YFP<sup>+++</sup>, la moyenne décroît alors que la variance reste quasiment identique. Le CV est donc stable. Cette stabilité indique que le processus de retour à l'équilibre semble suivre une dynamique commune.

Dans les sous-clones négatifs, le faible niveau de fluorescence donne plus d'impact à la variance et le CV est donc plus élevé. A variance égale, le CV est jusqu'à 10 fois plus importante dans les sous-clones « YFP - ». Au cours du temps, le CV diminue dans certains sous-clones en raison de la forte augmentation de la moyenne (Figure 67).



**Figure 67 Relation entre moyenne de fluorescence et coefficient de variation.**

*Il existe une relation linéaire entre le niveau moyen d'expression par clone et le CV de la population (log/log).  $r^2=0.83$  et droite :  $y = -0.67x + 4.6$*

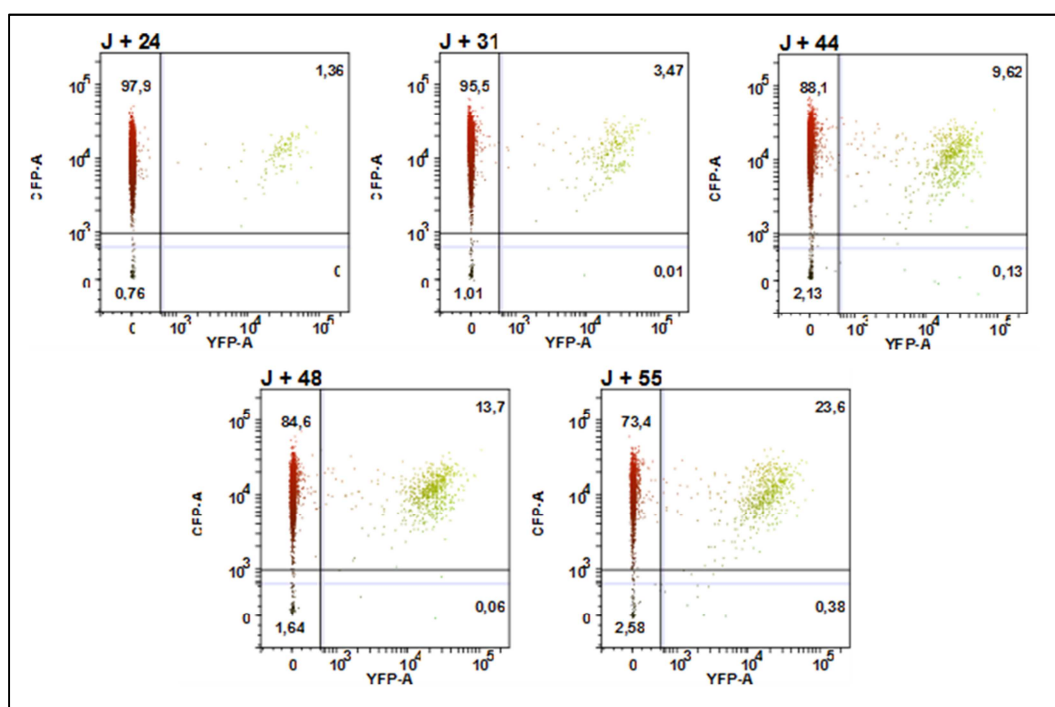
Cette relation linéaire entre CV et MFI est une observation classique dans ce genre d'études qui démontre la variabilité est intimement liée au nombre de molécules.

### *1.2.d. Évolution de la proportion de cellules exprimant le transgène*

Nous avons vu que la fluorescence moyenne dans chaque catégorie de sous-clones (YFP-, + et +++) varie au cours du temps. Les YFP+ et YFP+++ tendent pour la majorité vers un niveau d'expression moyen alors que certains clones arrêtent d'exprimer le transgène. Au contraire, les sous-clones négatifs présentent une augmentation de leur fluorescence

moyenne. Ils peuvent ré-exprimer le transgène spontanément pour atteindre un niveau d'expression légèrement inférieur aux sous-clones positifs.

Que l'on regarde l'extinction d'un transgène ou sa réactivation, on note qu'elles sont liées à l'augmentation et à la diminution du nombre de cellules positives et non pas à une augmentation/diminution globale de la fluorescence dans chaque cellule (Figure 68 et Figure 69). Il s'agit d'un processus que l'on peut qualifier de « tout ou rien », en opposition à un processus graduel qui montrerait un comportement commun de toutes les cellules de la population. Nous revenons ainsi au problème évoqué plus haut par l'utilisation de la moyenne.

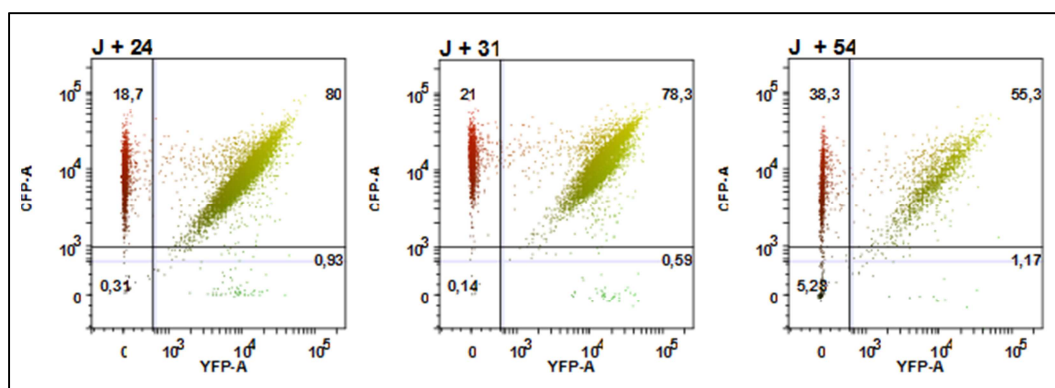


**Figure 68 Augmentation de la proportion de cellules positives.**

*Le sous-clone C3E7 est de phénotype initial « eYFP- » et CFP+. Le pourcentage de cellules exprimant le transgène eYFP augmente ainsi que celui des cellules double négatives.*

On observe effectivement dans ce premier cas une augmentation progressive du nombre de cellules dans le cadran supérieur droit (cellules double positives) au détriment de la proportion de cellules qui étaient uniquement eCFP positives (cadrant haut gauche). L'augmentation du nombre de cellules exprimant le transgène eYFP va en conséquence faire augmenter la fluorescence moyenne de la population entière mais aussi sa variabilité puisque l'écart entre les cellules négatives et positives est très important.

Le deuxième cas (Figure 69) correspond au phénomène inverse, une chute du pourcentage de cellules double positives en faveur de cellules exprimant uniquement l'eCFP.

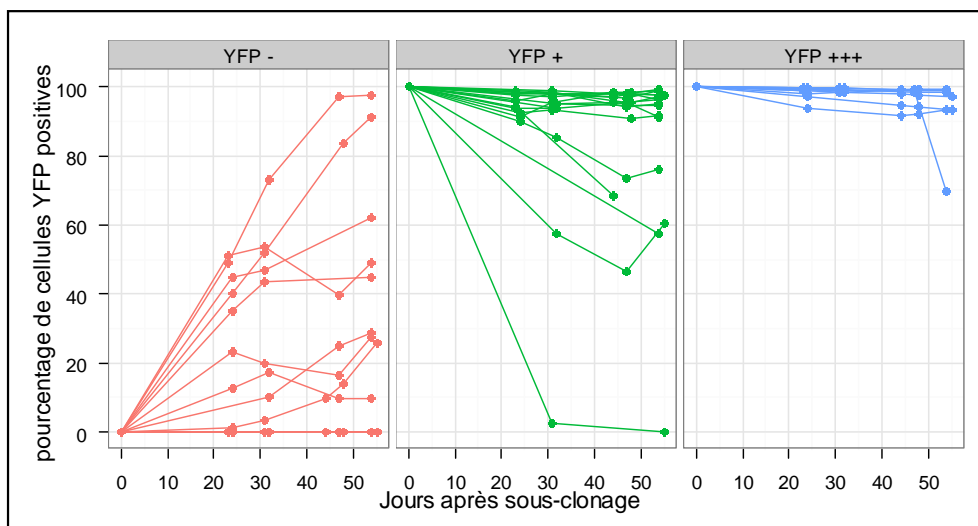


**Figure 69 Diminution de la proportion de cellules positives.**

*Le sous-clone C4H3est de phénotype initial «YFP + » et CFP +. La proportion de cellules exprimant les deux transgènes diminue au profit des autres sous-populations.*

Comme dans l'exemple précédent, la modification des proportions de cellules exprimant ou non le transgène eYFP va modifier la fluorescence moyenne de la population. De façon intéressante, on observe dans cet exemple une extinction préférentielle d'un seul transgène.

Dans chaque sous-clone s'établissent donc des sous-populations au fur et à mesure du temps. Nous avons préalablement vérifié que les cellules exprimant les transgènes n'étaient pas avantagées ou désavantagées en termes de prolifération. Un avantage sélectif aurait biaisé les observations. Les protéines fluorescentes n'ayant aucune activité biologique, le risque majeur est que la présence des deux transgènes soit une charge pour les cellules les exprimant, réduisant ainsi leur prolifération aucune différence significative n'a été observée. Il semble de ce fait intéressant d'étudier ces sous-populations plutôt que la population entière. La Figure 70 résume l'évolution du pourcentage de cellules exprimant l'eYFP dans les sous-clones issus de chacune des catégories YFP-, YFP+ et YFP+++.



**Figure 70** Évolution de la proportion de cellules eYFP positives après sous-clonage.

La majorité des sous-clones YFP- ont un pourcentage de cellules positives qui augmente. Les sous-clones positifs sont perdent quelques cellules positives (5 à 10%) mais restent relativement stables. Les sous-clones YFP+ semblent perdre l'expression du transgène eYFP plus fréquemment

Nous pouvons remarquer que la majorité des sous-clones « YFP - » (8 sur 14) présente une augmentation de plus de 20% du nombre de cellules exprimant le transgène eYFP. Les autres sous-clones restent négatifs même après 2 mois de culture. La présence des transgènes dans ces clones a été confirmée par PCR.

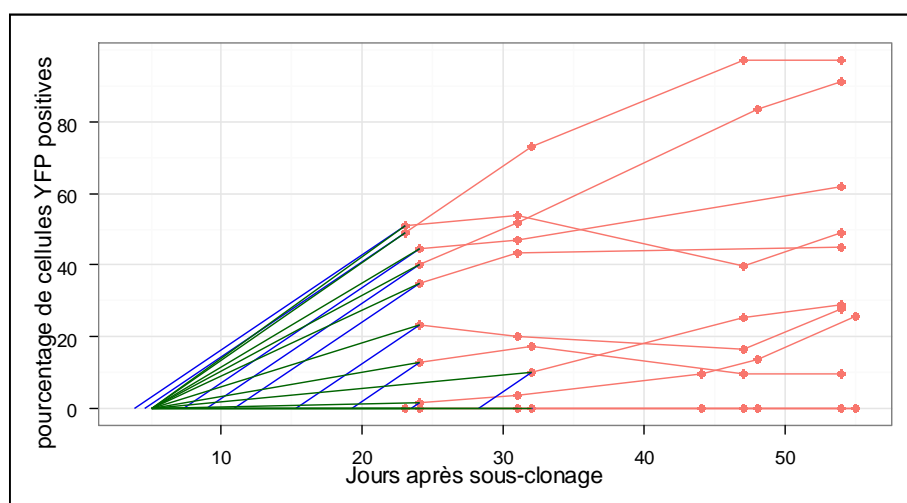
Dans les sous-clones YFP+ et YFP+++, 100% des cellules expriment le transgène eYFP lors du tri puisque c'est le critère de clonage. On remarque que le pourcentage de cellules positives a tendance à décroître dans ces deux catégories. Néanmoins la décroissance est faible pour la majorité des sous-clones (entre 5 et 10%). La catégorie « YFP + » semble plus sensible à l'extinction puisque 5 sous-clones sur 18 présentent une diminution significative du nombre de cellules exprimant le transgène eYFP (-20%). Dans la catégorie « YFP +++ », 1 seul sous-clone sur 12 se comporte de cette façon.

Cette différence de dynamique entre l'apparition des cellules négatives et positives nous laisse supposer qu'il existe un déséquilibre entre la probabilité d'extinction d'un transgène et son activation. Ainsi, comme le niveau d'expression moyen du transgène eYFP, le pourcentage de cellules positives dans les sous-clones semblent converger vers un pourcentage proche de celui de la population d'origine. Les événements de répression

doivent être soit de courte durée soit très rares. Ils peuvent visiblement être stabilisés puisque certains sous-clones négatifs n'expriment jamais le transgène eYFP.

Dans chaque sous-clone, il est possible d'estimer le pourcentage de cellules qui sont devenues positives ou négatives par intervalle de temps. L'augmentation maximale du nombre de cellules positives est de +2.6 % par jour, la valeur minimal de +0.06 %. Ce calcul ne tient pas compte des divisions cellulaires. La pente maximale de décroissance du nombre de cellules positives est de 4.2, le minimum de 0.07.

Grâce aux pourcentages d'augmentation ou de diminution du nombre de cellules positives calculés, une régression à partir du premier point de temps (J23) peut être effectuée pour déterminer la période à laquelle les premières cellules positives sont apparues (Figure 71). Les segments bleus sont l'extrapolation faite avec la pente maximale (2.6%) tandis que les segments verts sont une extrapolation à partir du point 5 jours. A ce temps là, aucune cellule positive n'est observée parmi les négatives au microscope.



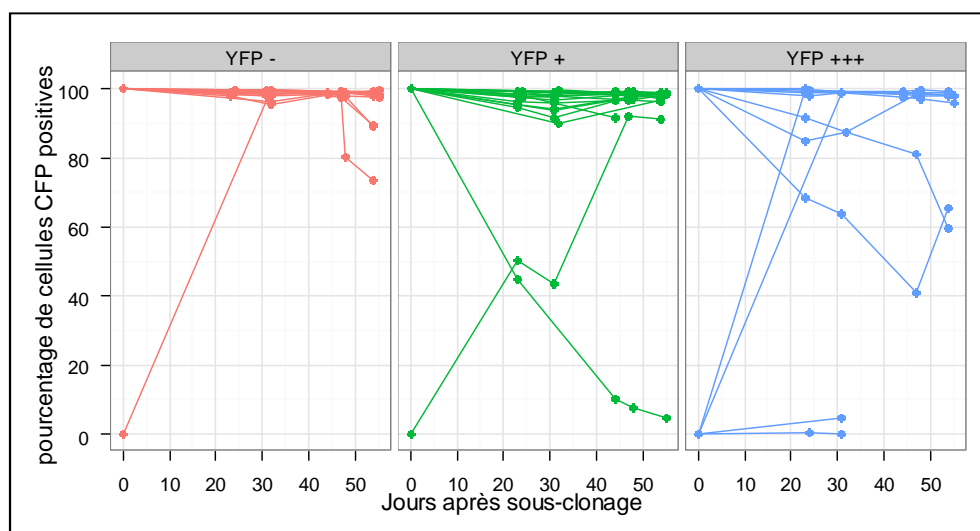
**Figure 71 Estimation du moment d'apparition des premières cellules eYFP positives.**

*La régression montre que dans la majorité des sous-clones, les premières cellules seraient apparues dans les 15 premiers jours après le tri.*

Si on admet que le taux d'apparition des cellules positives est constant depuis le début de l'expérience, la régression nous indique que dans la plupart des sous-clones négatifs, les premières cellules positives sont apparues dans les 10 premiers jours après le clonage. Ce taux d'apparition de cellules positives est en réalité un équilibre entre celles qui s'allument et celles qui s'éteignent. Nous atteignons une des limites de la cytométrie, à savoir

l'impossibilité de suivre une cellule isolée. Nous ne savons donc pas à la vue de ces résultats si les cellules exprimant le transgène dans les premiers temps ont conservé leur expression ou ont de nouveau cessé de l'exprimer. La pente obtenue nous permet aussi de calculer une fréquence de changement phénotypique. La probabilité de changer de phénotype chaque jour est alors comprise entre 0.026 et 0.0007 ce qui est très faible à l'échelle d'une cellule mais suffisant à l'échelle d'une population de milliers d'individus pour que cela soit visible.

Rappelons que pour des raisons techniques, nous avons trié les cellules uniquement sur le critère d'expression de l'eYFP. Néanmoins l'appareil d'analyse étant différent de celui utilisé pour le tri cellulaire, nous avons analysé l'expression de l'eCFP pour chaque sous-clone. Nous avons étudié l'évolution du pourcentage de cellules eCFP positives dans les trois catégories YFP-, + et +++ (Figure 72).



**Figure 72 Évolution du pourcentage de cellules eCFP positives.**

*Dans les trois catégories de sous-clones eYFP, la plupart d'entre eux sont eCFP+. Certains cessent d'exprimer ce transgène. A l'opposé, certains sous-clones sont eCFP- et deviennent positives.*

Par chance, certains des sous-clones n'expriment pas l'eCFP. On note que 5 des 6 sous-clones dont l'expression d'eCFP est réprimée, expriment ce transgène après quelques jours de culture. Un seul sous-clone reste eCFP négatif. Sur les 44 sous-clones établis, 39 étaient eCFP positifs. Après 2 mois de culture, 3 d'entre eux ont clairement un nombre de cellules eCFP positives qui décroît.

Le tableau ci-dessous reprend le nombre de clones obtenus ainsi que leur phénotype observé au microscope après tri et au bout de deux mois de culture.

Phénotype	Nombre de sous-clones	Tri +5 jours		Tri + 55 jours	
		eCFP+	eYFP+	eCFP+	eYFP +
<b>YFP -</b>	14	13 (93%)	0 (0%)	14 (100%)	10 (71%)
<b>YFP +</b>	18	17 (94%)	18 (100%)	18 (100%)	17 (94%)
<b>YFP+++</b>	12	9 (75%)	12 (100%)	11 (92%)	12(100%)

*Tableau 8 Bilan de l'évolution du phénotype des sous-clones.*

Sur les 14 clones eYFP- obtenus après le tri cellulaire, 10 sont devenus eYFP positifs, 1 est devenu eCFP positif et 4 n'expriment toujours pas eYFP. Parmi les clones YFP+, 1 seul est a cessé d'exprimer l'eYFP alors qu'un autre, eCFP négatif, exprime l'eCFP à la fin de l'expérience. Sur les 3 clones YFP+++ n'exprimant pas l'eCFP lors du tri cellulaire, tous expriment l'eCFP après 2 mois de culture et aucun n'a cessé d'exprimer l'eYFP.

Cette analyse à long terme de l'évolution de sous-clones phénotypiquement caractérisés a permis de déterminer que deux phénomènes semblent œuvrer dans l'émergence de l'hétérogénéité dans ce modèle. Des changements rares, liés à l'extinction et activation des transgènes génèrent une macro-hétérogénéité caractérisée par la présence de sous-populations dans chaque clone parental. La proportion de cellules exprimant ou non les transgènes évolue lentement pour atteindre un équilibre et est responsable de la fluorescence moyenne de chaque sous-clone ainsi que de la variabilité.

L'extinction des transgènes est un phénomène rare comme en témoigne le faible pourcentage de cellules négatives apparaissant dans l'expérience de sous-clonage de cellules positives. Ce phénomène est probablement lié à la dynamique de la chromatine au site d'intégration du transgène.

La résolution temporelle dans cette première partie n'a pas permis d'étudier les changements plus rapides probablement responsables de la micro hétérogénéité. Ainsi,

pour étudier ce phénomène, il est nécessaire d'améliorer la résolution temporelle. Nous avons choisi d'étudier les fluctuations par imagerie temps réel.

### I.3. Analyse de la dynamique de l'expression génique

---

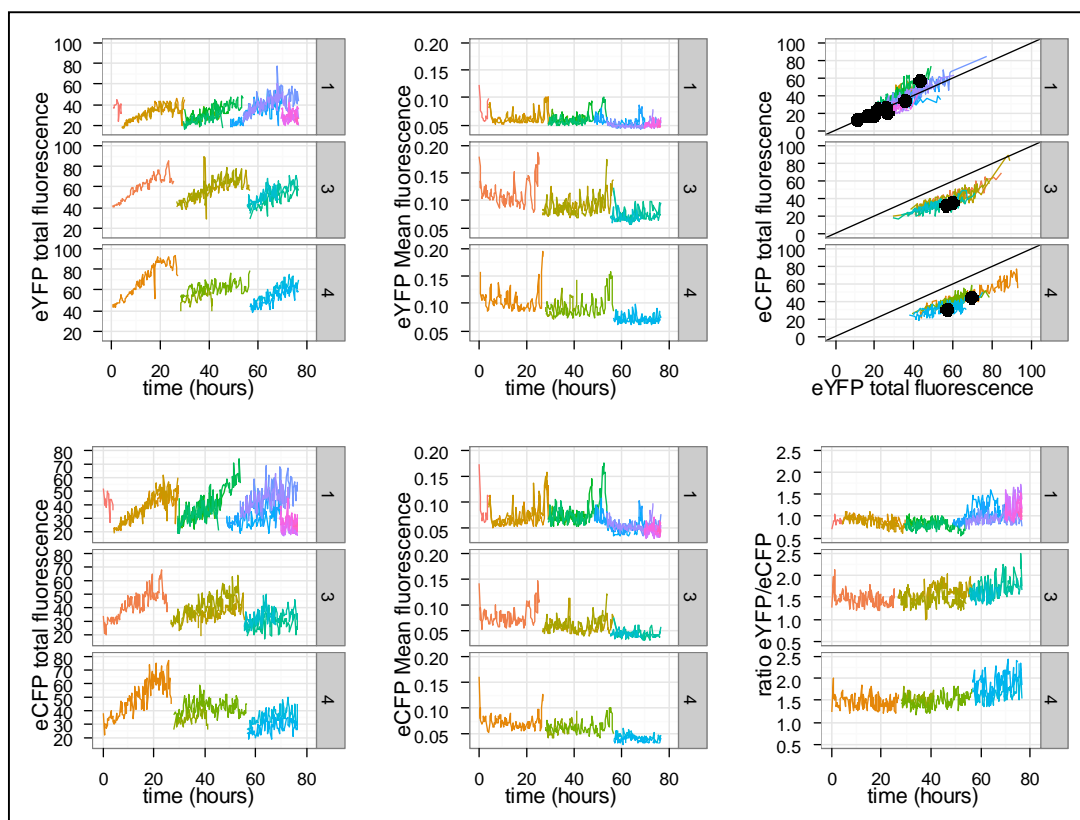
Nous avons observé dans la partie précédente que l'hétérogénéité apparaît rapidement après le clonage d'une cellule. En 25 jours, un début de diversification apparaît. La limite majeure de la stratégie précédente est le délai nécessaire à l'accumulation des cellules avant la première analyse en cytométrie en flux. Entre le jour du clonage et le 25<sup>e</sup> jour d'expansion, nous n'avons aucune information concernant l'évolution du phénotype. Il est donc nécessaire de changer de stratégie.

En utilisant le même modèle cellulaire, nous avons suivi des cellules en culture par microscopie. Il nous est alors possible de suivre l'expression de chacun des transgènes d'une cellule donnée ainsi que de sa descendance au cours du temps. Nous avons une vision à l'échelle des individus et non plus de la population entière. Cet échantillonnage peut néanmoins présenter quelques limites que nous discuterons plus tard.

Sans effectuer de sous-clonage, un millier de cellules issues de chacun des clones parentaux (C1, C3 et C4) ont étéensemencées dans une boîte de culture adaptée à la microscopie. Pendant 1 semaine en moyenne, ce qui correspond à 5-6 divisions cellulaires, les cellules présentes sur le champ ont été imagées à raison d'une image toutes les 20 minutes avec un objectif 10x. Le fait de privilégier une faible densité au début de l'expérience permet d'éviter d'atteindre la densité maximale trop rapidement mais aussi de réaliser un sous-clonage artificiel puisque chaque cellule et sa descendance sont bien isolées. A l'issue de l'expérience, les images sont analysées avec le logiciel CellProfiler afin de mesurer l'intensité de chaque fluorophore et de retracer la généalogie de chaque cellule. Certaines cellules ont été éliminées de l'analyse, d'une part si elles sont sorties du champ du microscope et d'autre part s'il est impossible de réaliser la segmentation correctement. Nous avons analysé l'expression des deux transgènes dans une douzaine de cellules et leur descendance à travers plusieurs cycles cellulaires (Figure 73 et Figure 75).



Les graphiques à gauche et au centre montrent respectivement l'évolution des fluorescences totales et moyenne d'eYFP (en haut) et d'eCFP (en bas) dans 3 cellules et leurs descendances. Les graphiques de droite montrent la corrélation et le ratio entre les deux transgènes.



**Figure 73 Analyse de l'expression des transgènes par vidéo-microscopie.**

3 cellules et leurs descendances sont représentées, chaque couleur du tracé représentant une génération. Les deux graphiques de gauche représentent l'expression de eYFP et eCFP dans les 3 lignées. Les deux graphiques du centre représentent l'évolution du niveau moyen des deux transgènes par cellules. Les graphiques de droite sont la corrélation et le ratio entre les deux transgènes.

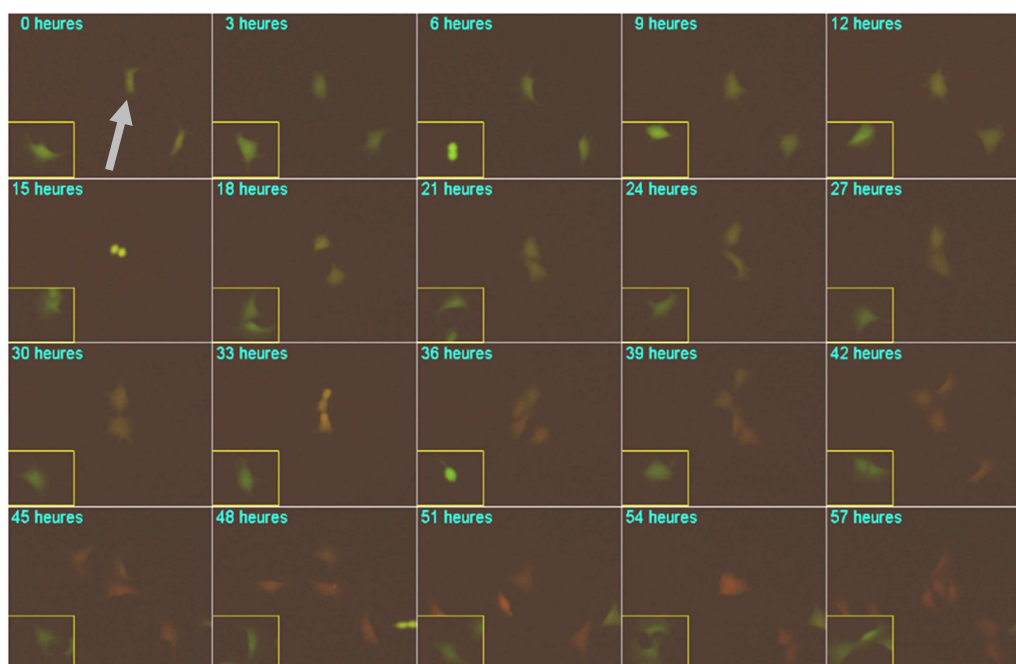
Pour les cellules et leur descendance (différentes couleurs), la fluorescence d'eYFP et d'eCFP varie au cours de chaque cycle cellulaire (graphiques de gauche). La fluorescence totale augmente au cours du cycle parallèlement à la taille de la cellule et en conséquence, la moyenne de fluorescence (i.e. la concentration cellulaire) reste stable (graphiques au centre).

Au moment de la division cellulaire, la cellule se détache de son support et s'arrondit, entraînant une diminution de son volume et une concentration de la fluorescence. La moyenne de fluorescence augmente donc brusquement puis décroît lorsque la cytokinèse

est terminée. La quantité de protéine fluorescente de la cellule mère est distribuée dans les deux nouvelles cellules et la fluorescence totale est donc divisée par deux. Cette dynamique est reproduite au cours des divisions cellulaires suivantes. Après 3 divisions, une certaine variabilité du niveau de fluorescence entre cellules sœurs apparaît. Néanmoins, sur la durée de l'expérience, le niveau moyen de fluorescence de chaque lignée semble stable autour d'une moyenne.

On observe une bonne corrélation entre les deux transgènes dans ces trois lignées comme le montre le graphique supérieur droit. L'expression d'eYFP et eCFP est équivalente dans la lignée 1 alors que dans les lignées 2 et 3, eYFP semble plus exprimée qu'eCFP. Ceci est confirmé sur le graphique inférieur droit où le ratio eYFP/eCFP est supérieur pour les lignées 2 et 3. Le ratio reste constant au cours des divisions successives indiquant une expression corrélée des deux transgènes.

Dans certains cas, il arrive que l'on observe un comportement différent des deux transgènes, dans l'exemple ci-dessous une diminution de la fluorescence d'un des transgènes au cours des divisions cellulaires (Figure 74).

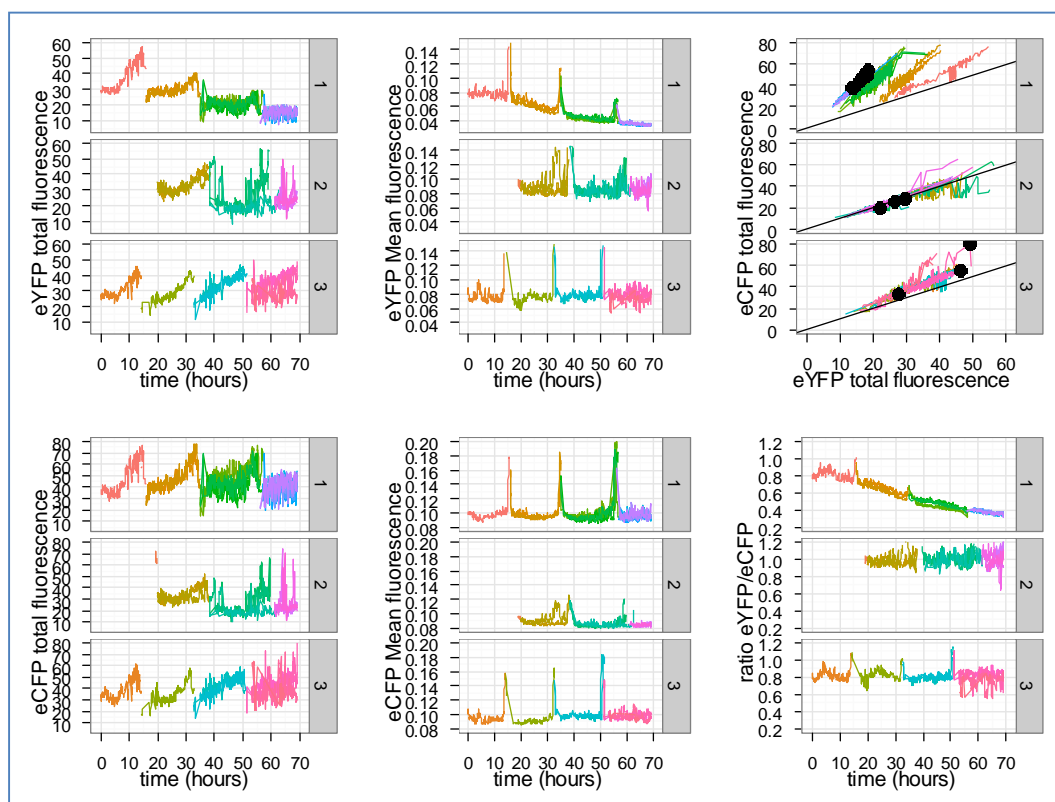


**Figure 74 Montage d'images de vidéo-microscopie.**

*Images RGB eCFP/eYFP (vert/rouge) montrant l'évolution de l'expression des deux transgènes dans une cellule isolée et sa descendance (flèche grise). En incrustation, une cellule contrôle du même champ de microscope. Confocale Zeiss, 10x.*

La cellule marquée par la flèche semble devenir de plus en plus rouge au cours des divisions cellulaires. Une cellule contrôle observée sur le même champ reste stable.

Sur la Figure 75, on observe que suite à la première division cellulaire, la fluorescence eYFP ne remonte que de 50% dans la lignée 1.



**Figure 75** Dynamique de l'expression des transgènes 3 cellules et leur descendance.

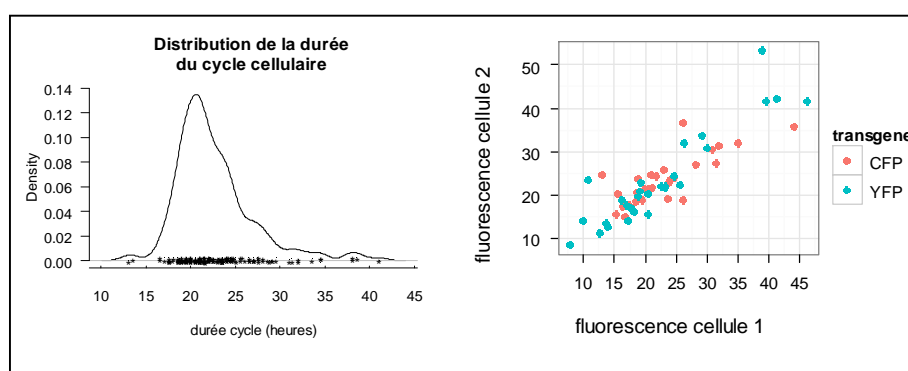
3 cellules et leurs descendance sont représentées, chaque couleur du tracé représentant une génération. Les deux graphiques de gauche représentent l'expression de eYFP et eCFP dans les 3 lignées. Les deux graphiques du centre représentent l'évolution du niveau moyen des deux transgènes par cellules. Les graphiques de droite sont la corrélation et le ratio entre les deux transgènes.

Il semble donc qu'un arrêt de transcription ait eu lieu, probablement dans la cellule-mère. Il est peu probable que l'arrêt d'expression soit apparu dans deux cellules-sœurs simultanément. Même si on ne peut situer précisément le moment, la fluorescence du transgène eYFP continue lentement à décroître, aidée par la dilution liée aux divisions cellulaires. La fluorescence totale par cellule diminue, ainsi que la fluorescence moyenne. En parallèle, la fluorescence liée au second transgène, eCFP, reste constante au cours des cycles cellulaires. Cet arrêt de synthèse de l'eYFP est particulièrement visible sur le graphique inférieur droit, où le ratio eYFP/eCFP décroît d'un facteur 2. Deux autres lignées

analysées sur le même champ de microscope se comportent de façon classique. Nous pouvons donc exclure tout artéfact technique.

A l'issue de cette expérience, nous avons pu constater que le niveau moyen de fluorescence d'une cellule et de sa descendance était relativement stable au bout de 3 à 4 divisions cellulaires. Néanmoins, il arrive que dans certaines cellules un transgène cesse d'être exprimé. Nous n'en avons observé que rarement dans ce genre d'expériences de tracking cellulaire, du fait de la rareté de l'évènement mais aussi du faible nombre de cellules analysées.

Dans la majorité des cas, les seules fluctuations significatives sont celles liées au cycle cellulaire, avec une variation d'un facteur 2 du niveau d'expression des deux transgènes. Une autre source pourrait être le partage inéquitable des composants cellulaire lors de la division. Grâce au tracking de cellules, nous montrons ici que la corrélation des fluorescences entre cellules sœurs est bonne ( $> 0.9$  pour les deux transgènes). Ceci indique donc une faible variabilité liée à une asymétrie lors de la division cellulaire. En revanche, nous notons une variabilité dans la durée du cycle cellulaire entre cellules qui pourrait être une autre source de fluctuations extrinsèques. De plus, la durée du cycle cellulaire est de moins en moins bien corrélée entre générations (0.85 entre cellules sœurs, 0.6 entre cellule mère et fille, 0.2 entre une cellule et sa « grand-mère »).



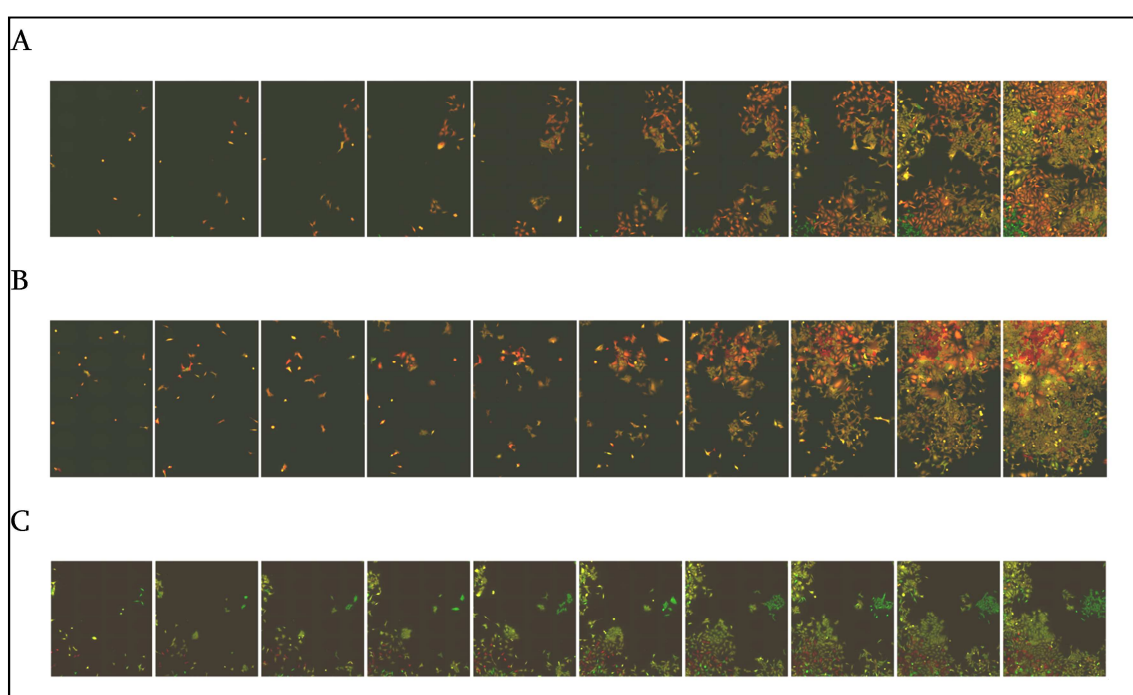
**Figure 76 Variabilité du cycle cellulaire et asymétrie de la division.**

*La distribution de la longueur du cycle cellulaire dans les cellules analysées est asymétrique (gauche). Le partage de la fluorescence entre cellules filles semble équitable (droite).*

La complexité de l'expérience entre la mise en œuvre et l'analyse des images rend l'analyse systématique difficile avec le matériel utilisé. Nous avons vu plus haut que le fait de

démarrer l'expérience avec un faible nombre de cellules était nécessaire pour éviter d'atteindre la confluence maximale trop tôt mais aussi pour faciliter le suivi de cellules. La contrepartie est que peu de cellules ont pu être analysées de façon approfondie.

D'autres expériences de vidéo-microscopie ont donc été réalisées en augmentant le nombre de cellules initiales et en laissant l'expérience se dérouler jusqu'à confluence maximale. La Figure 77 montre trois exemples d'expériences de vidéo microscopie sur de longues périodes (4 à 5 jours). L'expérience commence avec 10 à 20 cellules sur l'ensemble de la mosaïque (30 champs de microscope) et se termine lorsque la densité atteint 80%.



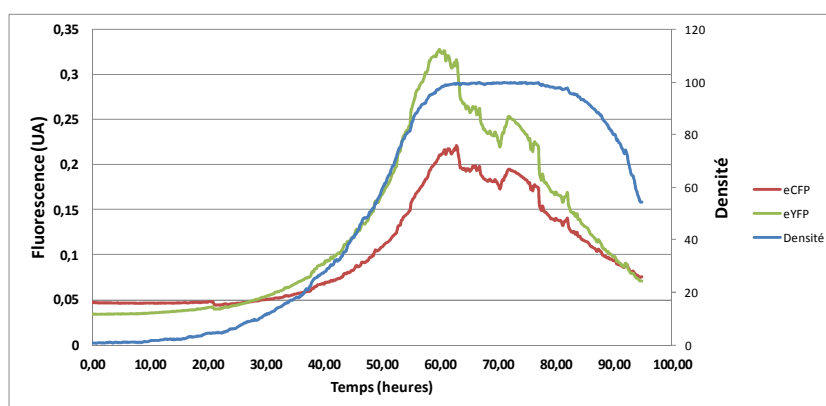
**Figure 77 Observation de la diversification du phénotype.**

*Mosaïque de 30 champs reconstituée, Biostation Nikon 20x, eCFP (vert) et eYFP (rouge), intervalle de 10 heures entre chaque images.*

En observant les cellules dès leur ensemencement, on s'affranchit du délai de 25 jours nécessaire à l'obtention du nombre de cellules suffisant pour une analyse en cytométrie. L'analyse des variations d'expression des transgènes n'a pas été soumise à la même rigueur que précédemment, à l'échelle de la cellule unique. Nous avons analysé visuellement les changements de phénotype.

La vidéo-microscopie à long terme nous a permis de confirmer les observations faites lors du sous-clonage. La première observation est que certaines cellules maintiennent leur phénotype de façon stable au cours du temps. Ceci est particulièrement visible pour les cellules n'exprimant qu'un seul des deux transgènes. En se divisant successivement, ces cellules isolées au début de l'expérience forment des colonies qui sont visibles à la fin de l'expérience. D'autres cellules de la population ont un phénotype qui évolue au cours du temps pour ne plus être comparable à celui du début de l'expérience.

Sur la période de 4 jours de cette expérience, chacun des phénotypes observés dans la population parentale apparaît à partir de cellules isolées. On observe des cellules exprimant les deux transgènes à différents niveaux ou un seul des transgènes. Certaines cellules négatives apparaissent aussi. Il est aussi intéressant de noter qu'à densité maximale, l'expression des transgènes est complètement abolie (Figure 78).



**Figure 78 Inhibition de l'expression des transgènes par la densité cellulaire.**

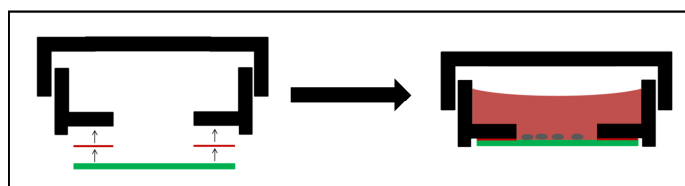
*La fluorescence de chaque transgène est mesurée sur l'image entière. Lorsque la densité atteint 100%, l'expression de chaque transgène dans la population diminue.*

Lorsque la densité augmente par le fait des divisions cellulaires, la fluorescence totale augmente en parallèle. Lorsque la densité atteint son maximum, on observe très rapidement (après moins de 2 heures) une décroissance de la fluorescence des deux transgènes dans la population. Cette décroissance n'est pas liée à la mort cellulaire qui n'intervient que 24h plus tard.

#### I.4. Fluctuations de l'expression génique à l'échelle d'un cycle cellulaire

Toujours en utilisant ce modèle de cellules bifluorescentes eCFP et eYFP, nous avons cherché à isoler les fluctuations de l'expression des deux transgènes à l'échelle d'un cycle cellulaire et non plus à travers plusieurs divisions cellulaires. Cette échelle de temps permet de s'affranchir des fluctuations liées à la division cellulaire et au partage asymétrique des protéines dans les cellules-filles.

Les cellules sont cultivées sur des lames de culture spéciales, recouvertes à la fibronectine dessinant des motifs géométriques. Fabriquées par la compagnie CYTOO®, ces lames permettent d'isoler les cellules les une des autres et de limiter leur mobilité au cours de l'analyse. Nous avons testé une version de lame de démonstration sur laquelle sont imprimés des motifs de taille et forme différentes susceptibles de s'adapter à plusieurs types cellulaires. La lamelle sont fixées sous une boîte de culture stérile de 35mm de diamètre ayant une ouverture de 14mm dans le fond (MatTek, cat#P35-14c, Figure 79). La lamelle est fixée par le dessous avec de la graisse de silicone stérile, non auto-fluorescente et non toxique (Diméthyle Silicone, DOW CORNING 111 valve lubricant and sealant, Agrément FDA 21 CFR 175.300).



**Figure 79 Schéma de montage du dispositif expérimental.**

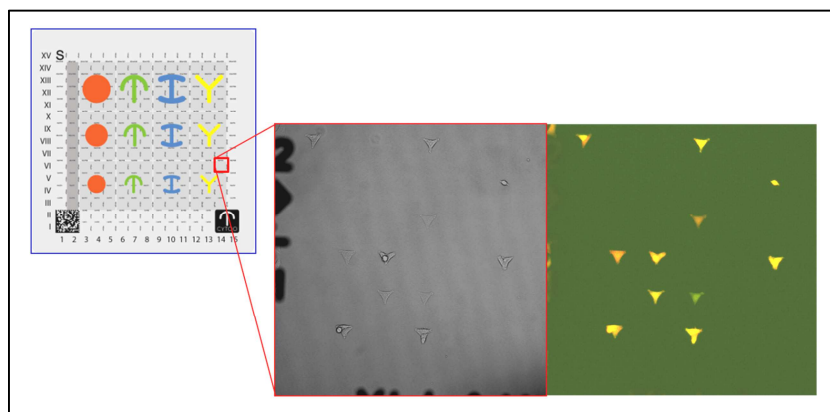
*A gauche : La lamelle Cytoo® (en vert) est collée à la boîte de culture (en noir) par de la graisse de silicone (en rouge).*

*A droite : Les cellules sontensemencées dans le dispositif étanche.*

La stérilité a été vérifiée par ensemencement d'une petite quantité de graisse dans du milieu LB et incubation pendant 24h à 37°C sous agitation. Aucune pousse bactérienne n'est notée. La toxicité a été testée en cultivant des cellules dans une boîte de pétri contenant de la graisse. Les cellules poussent au contact de la graisse et aucune contamination n'est notée. Au microscope, les cellules expriment les deux transgènes normalement et la graisse ne présente pas de fluorescence aux longueurs d'onde de la



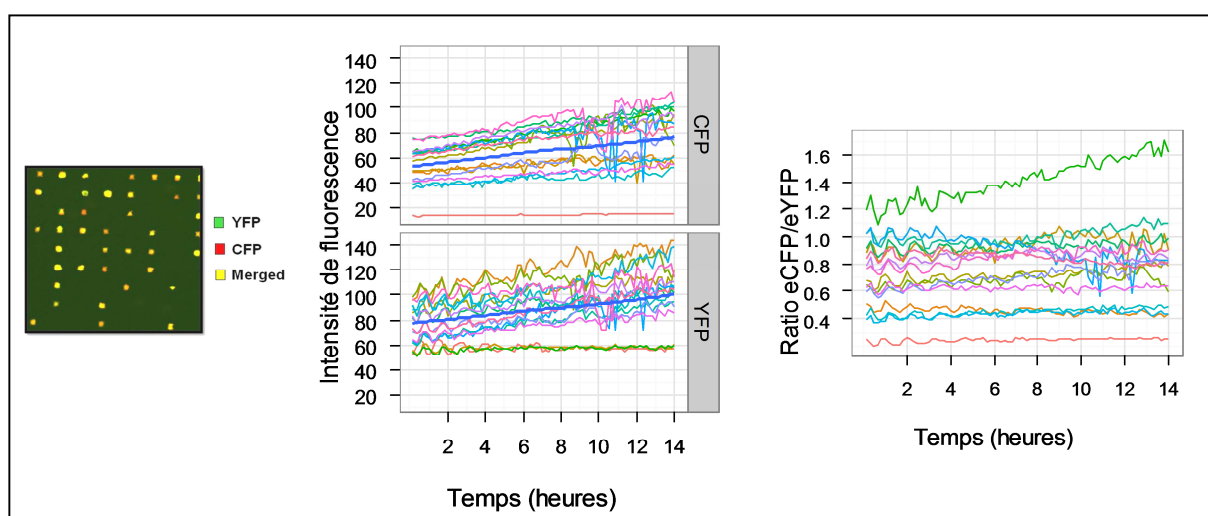
FITC, TRITC, rhodamine, DAPI, eCFP, eYFP, TexasRed et GFP. Les cellules à ensemen- cer sont préparées classiquement et 60 000 d'entre-elles sont déposées sur la lamelle. Après 2 heures d'incubation, un lavage est effectué avec du milieu complet pour éliminer les cellules non adhérentes. Après 2 heures, les cellules adoptent la forme du motif de fibronectine (Figure 80).



**Figure 80** Détails de la puce à cellules CYTOO®.

*A gauche : Schéma de la disposition, de la taille et de la forme des motifs de fibronectine. A droite : observation de la puce au microscope confocale (10x, transmission et fluorescence). Les cellules sont réparties de façon ordonnée sur la puce et adoptent la forme des motifs imprimés.*

La boîte est installée dans la chambre d'acquisition du microscope confocale et la fluorescence d'eCFP et d'eYFP est mesurée toutes les 10 minutes pendant 12 heures (Figure 81).



**Figure 81** Évolution de l'expression des transgènes eCFP et eYFP.

*La fluorescence des deux transgènes est quantifiée dans chaque cellule de la puce au cours du temps. On observe une tendance générale linéaire et croissante pour l'ensemble des cellules. Le rapport entre les deux fluorescences montre qu'il y a peu de changements.*



Les cellules choisies ne se divisent pas pendant la période d'enregistrement. La tendance globale pour les deux transgènes est une augmentation linéaire de la fluorescence (Figure 81). La fluorescence totale de la majorité des cellules augmente au cours du temps. Cette observation est identique à celle de l'analyse précédente (Figure 73 et Figure 74). Elle correspondrait alors à la progression dans le cycle cellulaire.

Le ratio de chaque cellule est relativement stable au cours du temps, fluctuant faiblement autour de sa moyenne. Ceci indique donc que les deux transgènes ont une dynamique similaire. Cette expérience confirme le caractère rare des changements phénotypiques tel qu'observé dans le paragraphe précédent. Ces événements rares génèrent de la variabilité à long terme et les phénotypes obtenus sont relativement stables au cours du temps. (Macro hétérogénéité).

Néanmoins, une des limites observées semble être l'impossibilité de mesurer les fluctuations à long terme avec notre système, ceci étant lié à la longue durée de vie de nos protéines fluorescentes. Cette demi-vie excessive de plus de 24h a pour effet d'amortir les fluctuations rapides. L'arrêt de synthèse protéique ne sera visible qu'après une longue période (au moins 5 ou 6 heures). De plus, pendant cette période, la synthèse peut redémarrer et s'arrêter plusieurs fois. Une telle stabilité n'est pas gênante pour une étude à long terme comme celle réalisée en cytométrie en flux mais limitante pour le cours terme. Il est donc nécessaire de modifier le modèle cellulaire si on veut améliorer la résolution temporelle de notre analyse. Les gènes rapporteurs ont été modifiés pour obtenir une demi-vie des protéines réduite en déstabilisant dans un premier temps l'ARN messager, et dans un second temps la protéine mature.

### I.5. Photo-blanchiment de cellules entières

---

La technique de référence consiste à éteindre la fluorescence dans une région particulière de la cellule et à suivre par la suite le retour de fluorescence dans cette zone, conséquence de la diffusion de molécules à l'intérieur de la cellule. Plusieurs techniques dérivant de ce

principe ont par la suite été développées comme le FLIP ou le FRAP inverse. Ici, nous procédons à une extinction de la fluorescence dans toute la cellule et mesurons le retour de fluorescence. Puisqu'il n'y a presque plus de protéines fluorescentes dans la cellule, le retour de fluorescence que nous observons est la conséquence d'une synthèse de nouvelles protéines ou d'une reformation de protéines dénaturées.

Nous avons ici choisi de mesurer la dynamique de l'expression de l'eYFP et de garder eCFP comme contrôle. Les propriétés spectrales et les caractéristiques de l'équipement rendent plus aisée l'étude de eYFP qui a été éteinte par excitation laser à 514nm.

### *1.5.a. Mise au point des conditions expérimentales*

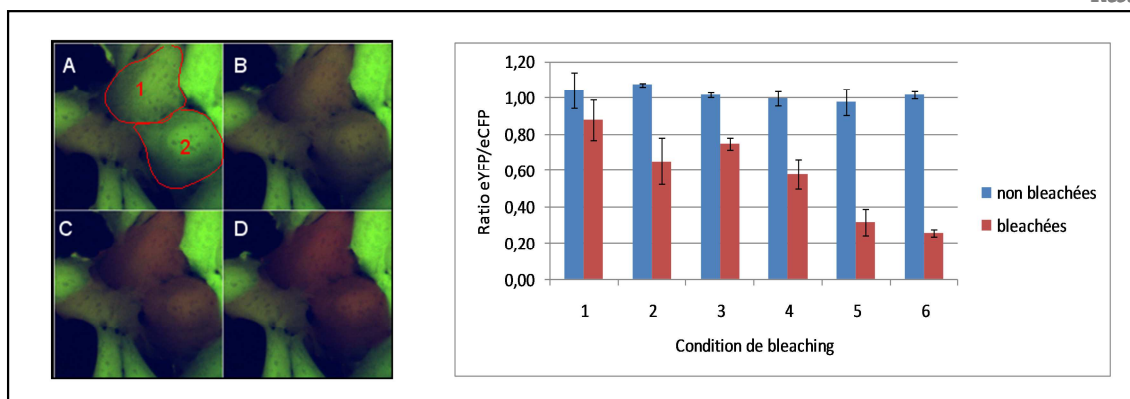
Classiquement pour une analyse FRAP, on tente d'obtenir une diminution de 80% de la fluorescence d'origine. Nous avons donc mis au point les conditions de bleaching afin de répondre à cet objectif. Des cellules exprimant les deux transgènes ont été cultivées dans une boîte de culture spécialement conçue pour l'imagerie jusqu'à leur adhésion puis cette boîte a été incubée sous le microscope thermo-régulé avec 5% de CO<sub>2</sub>.

Plusieurs variables susceptibles de modifier l'extinction de fluorescence ont été testées : la puissance du laser (50 ou 100%), la durée de l'illumination laser (20 à 100 pulses) ainsi que le facteur de grossissement (Tableau 9).

Condition #	1	2	3	4	5	6
Puissance laser (%)	50	100	50	100	100	100
Nombre de pulses	5x20	5x20	100	100	3x20	4x20
zoom	1	1	1	1	2	2

**Tableau 9 Conditions expérimentales pour la mise au point du FRAP**

Pour chaque condition, entre 2 et 5 cellules sont illuminées et, en parallèle, entre 2 et 4 cellules non perturbées sont analysées en guise de contrôle. Une image est acquise avant l'expérience et une image à la fin de chaque illumination laser. Les images sont analysées avec le logiciel CellProfiler (Figure 82).



**Figure 82 Photo-blanchiment de cellules entières.**

*A gauche* : images composites du canal vert (eYFP) et du rouge (eCFP) A : Sélection de la ROI avant photo-blanchiment. B-D : 3 excitations laser successives. *A droite* : Fluorescence eYFP normalisée par eCFP en fonction des conditions.

La Figure 82 (à droite) résume l'extinction relative moyenne du signal eYFP obtenue avec les différentes conditions de photo-blanchiment (Tableau 9) pour des cellules illuminées par le laser et des cellules de contrôle. Pour les 6 conditions testées, aucune modification de la fluorescence de eCFP ou eYFP n'est observée dans les cellules non-excitées. De même, la fluorescence de la eCFP dans les cellules traitées ou non ne change pas significativement. Cette observation confirme que la fluorescence d'eCFP n'est pas modifiée par l'excitation laser à 514nm.

La fluorescence d'eYFP standardisée par celle d'eCFP montre que les conditions 5 et 6 sont les plus optimales pour atteindre le maximum d'extinction de l'eYFP. Le ratio entre la eYFP et eCFP atteint 0.25 pour la condition 6, ce qui équivaut à une chute de 75% de la fluorescence d'eYFP. Ce pourcentage est très reproductible entre cellules et ne dépend pas du niveau de fluorescence initial de la cellule.

A l'issue de l'excitation laser, certaines cellules sont conservées dans les conditions standards de culture afin de vérifier leur viabilité. Après 12h, la majorité des cellules illuminées par le laser est toujours en vie et continue de proliférer normalement. La phototoxicité liée à l'excitation laser ne semble donc pas limitante pour cette expérience. Une mortalité cellulaire systématique est obtenue à partir de 250 pulses d'illumination laser à une puissance de 100%.

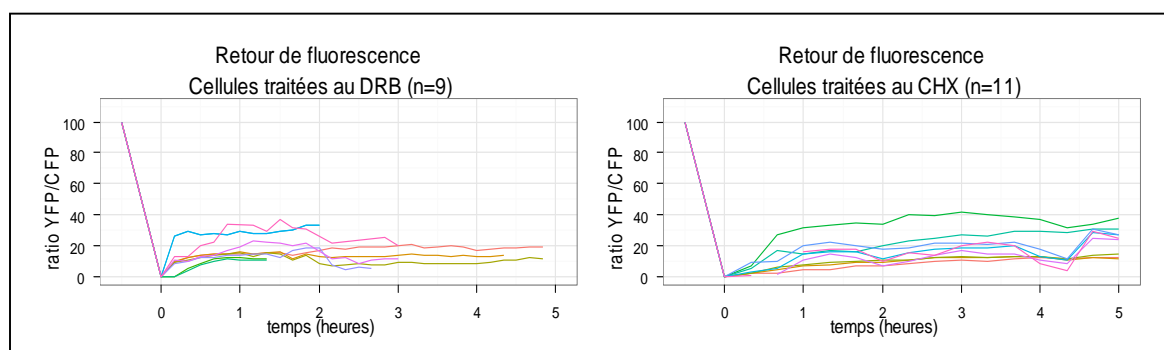
La condition 6 a donc été privilégiée pour la suite des expériences de photo-blanchiment et le protocole suivant a été adopté :

1. Une première image est réalisée avant le photo blanchiment, au zoom 1.
2. Les différentes régions d'intérêt (ROI) sont dessinées sur les cellules d'intérêt, au zoom 2.
3. L'excitation laser est réalisée à pleine puissance pendant 4 séries de 20 pulses, espacées de 30 secondes, au zoom 2.
4. Au zoom1, une série d'images est réalisée pendant 4 à 24h, toutes les 10-15 minutes.

### *1.5.b. Vérification de la dénaturation des protéines illuminées*

Il arrive dans ce genre d'expérience que la protéine d'intérêt ait une stabilité structurale suffisamment forte pour ne pas être dénaturée irréversiblement. Le retour de fluorescence qui est observé est alors la combinaison de la diffusion de protéines natives et de protéines destructurées qui retrouvent leur conformation d'origine spontanément.

Afin de vérifier que ce phénomène est absent dans notre système expérimental, les cellules sont traitées avec un inhibiteur de la traduction : le cycloheximide (CHX), et un inhibiteur de la transcription : le DRB, avant de subir l'illumination laser (Figure 83). Si la dénaturation est irréversible, les cellules illuminées sur toute leur surface ne devraient pas retrouver de fluorescence puisqu'il n'y a plus de synthèse protéique. Le retour de fluorescence est observé pendant 4 à 12 heures après l'illumination laser.



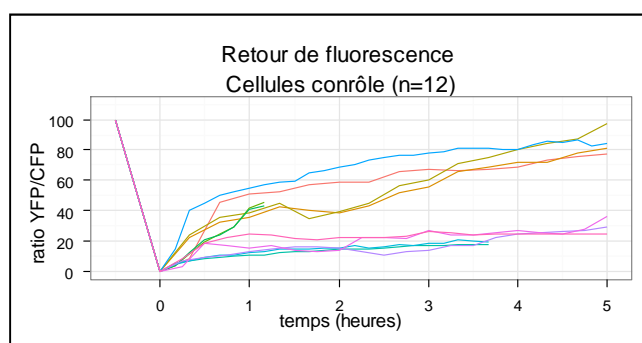
**Figure 83** Retour de fluorescence après inhibition de la transcription ou traduction.  
L'inhibition de la transcription et de la traduction empêche le retour de la fluorescence.

Les cellules traitées avec le DRB ou le CHX, bloquant respectivement la transcription et la traduction, présentent un retour de fluorescence inférieur à 15% et cela même au bout de 8 heures, durée maximale effectuée. Malgré le traitement, les cellules sont toujours vivantes pendant l'expérience mais présentent néanmoins des signes de mortalité après plus de 10 heures de traitement (cellules arrondies, détachement du support).

Ces deux expériences montrent qu'il n'y a que très peu de re-conformation spontanée des protéines et que l'extinction de fluorescence est irréversible dans notre modèle. En conséquence, un retour de fluorescence dans les cellules non traitées sera principalement lié à une synthèse de nouvelles protéines eYFP.

### *1.5.c. Analyse de la cinétique de la synthèse protéique*

La même expérience sur les cellules non-traitées révèle que le profil de retour de la fluorescence eYFP est très hétérogène (Figure 84). Certaines cellules récupèrent jusqu'à 90% de leur fluorescence initiale tandis que d'autres cellules ne récupèrent aucune fluorescence, ou tout au plus autant que les contrôles effectués précédemment (Figure 83). Ces cellules sont pourtant toujours vivantes.



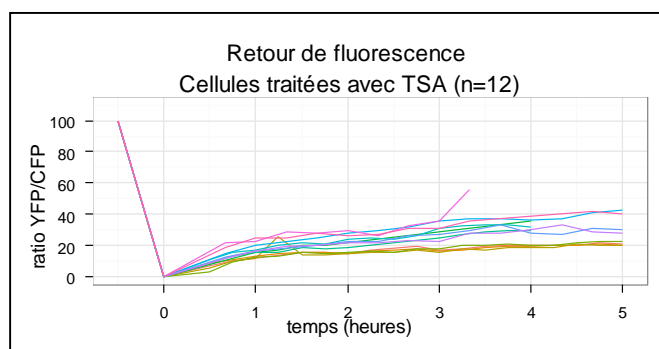
**Figure 84 Retour de fluorescence sur cellules non traitées.**

*Suite à l'illumination laser, la réduction du niveau de fluorescence de cellules isolées et suivie par une récupération hétérogène de la fluorescence. Certaines cellules ne retrouvent pas leur niveau initial de fluorescence après 5 heures de cultures.*

Cette observation est à mettre en relation avec l'extinction de la fluorescence dans certaines cellules lors du time lapse. Elle semble indiquer qu'à certains moments, le

contexte chromatinien à ce locus ne permet pas l'expression du transgène. Néanmoins, sans extinction de la fluorescence par l'illumination laser, ces cellules n'exprimant pas le transgène à cet instant donné auraient malgré tout été fluorescentes, en raison de la durée de vie des protéines extrêmement longue.

Afin de vérifier cette hypothèse, des cellules préparées dans les mêmes conditions ont été traitées avec de la TSA pendant 2 heures, une drogue connue pour ses effets sur la compaction de la chromatine (Figure 85). En inhibant certaines histone-déacétylases, elle provoque l'accumulation de résidus acétyle sur les histones. En conséquence, un état décondensé de la chromatine est privilégié, (euchromatine) favorisant l'accès à l'ADN pour les facteurs de transcription et la machinerie de transcription.



**Figure 85 Retour de fluorescence de cellules traitées à la TSA.**  
L'ensemble des cellules ne retrouve pas leur niveau de fluorescence initial.

L'ensemble des cellules traitées par la TSA ne retrouve pas un niveau de fluorescence comparable à la condition sans traitement. Toutes les cellules traitées recouvrent au moins 20% de fluorescence, 4 heures après l'extinction par illumination laser ce qui n'est pas mieux que les contrôle avec inhibiteur. Des expériences ultérieures nous ont permis de démontrer que l'action de la TSA sur l'expression des transgènes eCFP et eYFP dans notre modèle était visible après 48 heures de traitement. La durée de traitement dans cette expérience de photo-blanchiment n'est probablement pas suffisante pour observer une amélioration du retour mais n'explique pas pourquoi aucune cellule ne retrouve un niveau de fluorescence comparable à celui d'origine.

## II. Amélioration du modèle pour l'étude des fluctuations rapides

---

Après une première tentative de mesure des fluctuations à court terme dans le premier modèle, nous ne sommes pas parvenus à mettre en évidence de tels changements. Cette résolution temporelle courte (quelques dizaines de minutes) implique que l'on soit capable de mettre en évidence des changements subtils. Nous devons donc disposer d'outils sensibles. La limite majeure pour ce genre d'analyse est la durée de vie de la protéine fluorescente qui sert de rapporteur. Une durée de vie trop longue va amortir les fluctuations. Nous avons donc construit un modèle cellulaire plus adapté à l'analyse de changements de niveau d'expression à court terme. Les nouvelles constructions utilisées pour construire ce modèle codent des protéines déstabilisées à courte durée de vie, d'environ 6 heures.

### II.1. Construction du modèle cellulaire avec protéines déstabilisées

---

#### II.1.a. Clones « dépendants » exprimant les protéines déstabilisées

---

##### ◆ *Transfection*

---

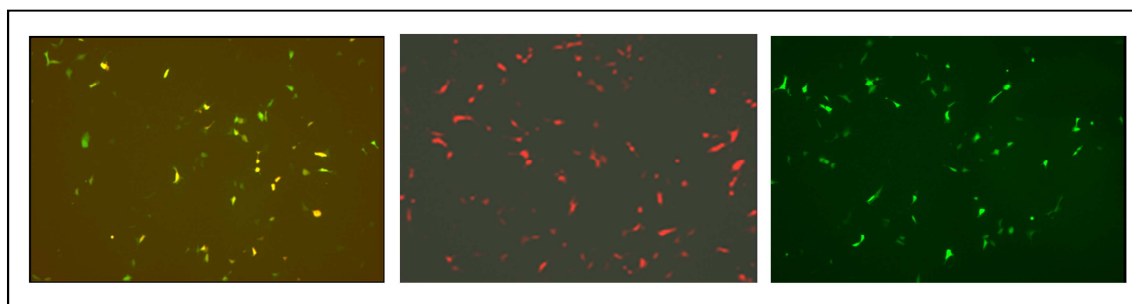
Afin de construire un nouveau modèle de cellules bifluorescentes dont les rapporteurs ont été déstabilisés, 3 pools de cellules natives HER911 ont été transfectés respectivement par 3 constructions plasmidiques :

- pCFPdesTOL2
- pYFPdesTOL2
- pCFPdes/YFPdesTOL2

La séquence codante pour l'eCFP et l'eYFP ont été modifiées pour obtenir une fluorescence plus faible par rapport aux rapporteurs d'origine. La durée de vie des ARNm a été réduite en ajoutant une séquence ARE et celle de la protéine en ajoutant une séquence PEST.

Chacune des constructions est co-transfectée avec un plasmide codant pour une transposase (pTranspTOL2) qui va faciliter l'intégration de chacun des transgènes au

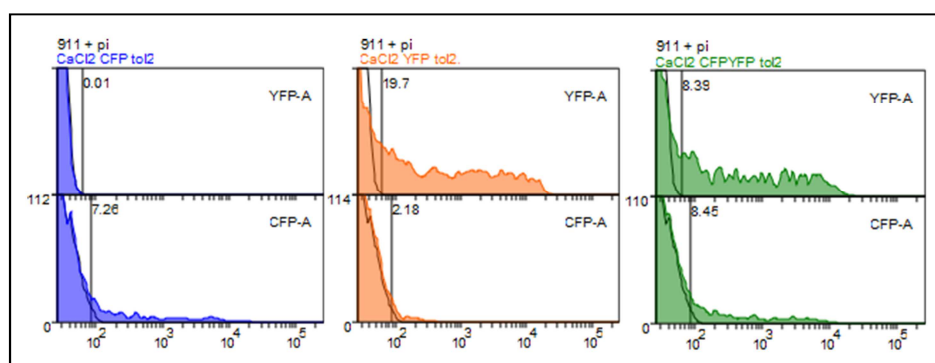
génomique de la cellule hôte. La cotransfection est réalisée avec un ratio 5/1 de copies de plasmides gène rapporteur/plasmide transposase par la technique de précipitation au phosphate de calcium. Au lendemain de la transfection, les puits de culture sont observés au microscope (Figure 86). Pour les 3 conditions, environ 40-50% de transfection est atteint.



**Figure 86 Contrôle visuel de la transfection à J+1.**

*Microscopie 10x. De gauche à droite : pCFPdes/YFPdesTOL2, pCFPdesTOL2 et pYFPdesTOL2.*

Les cellules sont entretenues de façon standard et une semaine plus tard, les populations transfectées sont analysées par cytométrie en flux (Figure 87). Après ce délai, plusieurs divisions ont eu lieu et les plasmides non-intégrés ont été éliminés. Les cellules fluorescentes restantes ont théoriquement intégré les plasmides de façon stable.



**Figure 87 Contrôle de la transfection par cytométrie en flux.**

*Le pourcentage de cellules stablement transfecté est de moins de 20% dans les trois conditions après une semaine de culture.*

Une semaine après la transfection, le pourcentage de cellules exprimant le ou les transgènes suivant les conditions est passé de 40-50% à 7% pour les cellules transfectées avec la construction pCFPdesTOL2, 20% pour les cellules transfectées avec la construction pYFPdesTOL2 et 8% pour les cellules transfectées avec la construction double.



### ◆ Clonage cellulaire

---

Afin de créer des clones, des cellules transfectées avec un seul plasmide (eCFPdes ou eYFPdes) et exprimant le transgène sont isolées grâce à un trieur de cellules (Tableau 10). De même, des cellules transfectées avec la construction double et exprimant les deux transgènes sont isolées.

Transfection	Nombre de clones
HER911 + pCFPdesTOL2+pTranspTOL2	11
HER911 + pYFPdesTOL2+pTranspTOL2	9
HER911 + pCFPdes/YFPdesTOL2+pTranspTOL2	18

*Tableau 10 Nombre de clones "Dépendants" obtenus.*

Pour la suite des expériences, les clones obtenus par transfection avec la construction double seront appelés «*dépendants*», l'eCFPdes et l'eYFPdes se trouvant sur la même construction. Leur abréviation sera «*CFPdes/YFPdes*».

#### *II.1.b. Clones « indépendants » exprimant les protéines déstabilisées*

---

##### ◆ Transfection

---

L'objectif est de produire des clones possédant les deux transgènes déstabilisés mais à des loci différents. Contrairement aux clones décrits ci-dessus dits «*dépendants*», ces clones seront appelés «*indépendants*» et notés par la suite «*CFPdes + YFPdes*».

Pour construire ces clones, deux clones exprimant uniquement CFPdes, produits précédemment, ont été transfectés simultanément avec les plasmides pYFPdesTOL2 et pTranspTOL2. Il s'agit des clones CFP2A et CFP4. Ces deux clones ont été choisis aléatoirement sans *a priori* sur leur niveau de fluorescence, ni sur leur variabilité d'expression. Comme précédemment, l'efficacité de transfection atteint 50% pour les deux expériences.

## ◆ Clonage cellulaire

---

Une semaine plus tard, des clones ont été établis par la technique décrite précédemment. En parallèle, des populations de 50 cellules ont été générées.

Transfection	Nombre de clones	Nombre de populations
CFP2A + pYFPdesTOL2+pTranspTOL2	25	3x50 cellules
CFP4 + pYFPdesTOL2+pTranspTOL2	23	6x50 cellules + 1x 900 cellules

*Tableau 11 nombre de clones « indépendants » obtenus.*

En théorie, chaque clone indépendant obtenu à partir du clone simple CFP2A ou CFP4 possède le même site d'intégration du transgène CFPdes. L'utilisation de la transposase lors de la seconde transfection apportant l'YFPdes a une très faible probabilité d'interagir avec la séquence déjà intégrée. Ce point sera vérifié lors de l'identification des sites d'intégration de chaque transgène. L'ensemble des clones obtenus ont été caractérisé par cytométrie en flux afin de déterminer leur niveau moyen d'expression des deux transgènes et la variabilité de chacun d'eux.

### II.2. Caractérisation des clones indépendants et dépendants

---

Afin de vérifier l'effet des modifications apportées sur les plasmides, nous avons évalué la diminution de fluorescence obtenue, la demi-vie des protéines et celles des ARNm.

#### II.2.a. Évaluation de l'efficacité de la déstabilisation

---

L'objectif de la déstabilisation des transgènes étant de réduire la demi-vie des protéines eCFP et eYFP, nous avons comparé le niveau moyen d'expression entre les clones possédant les constructions stables et les clones possédant les constructions déstabilisées (Figure 88). Les deux types de clones ont été analysés avec les mêmes réglages.

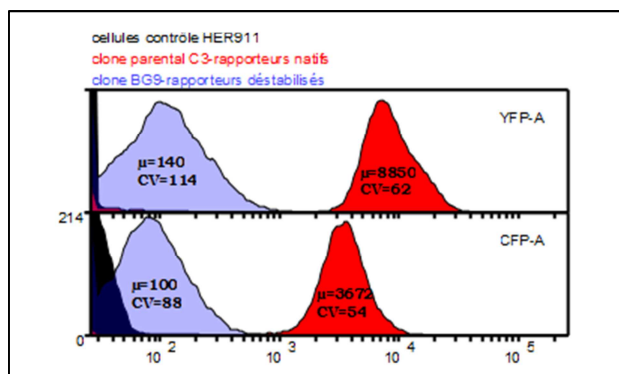


Figure 88 Comparaison des constructions stables / déstabilisées.

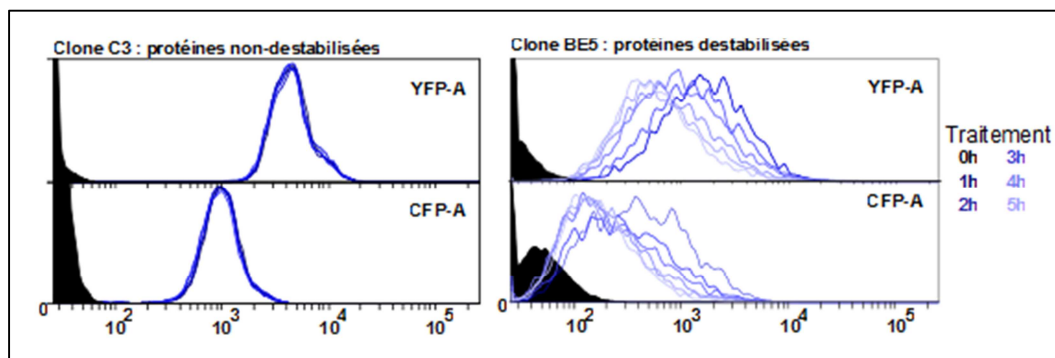
Compte-tenu de la différence observée, la déstabilisation a été très efficace. En conséquence, les réglages du cytomètre ont été adaptés pour que le niveau moyen de fluorescence des clones déstabilisés soit au centre de la plage dynamique (MFI entre  $10^3$  et  $10^4$  UA).

### II.2.b. Estimation de la demi-vie des protéines stables et déstabilisées

Pour mieux caractériser l'efficacité de la déstabilisation des constructions, la durée de vie des protéines fluorescentes a été déterminée. Les cellules de clones exprimant les protéines stables ou déstabilisées sont analysées par cytométrie en flux, après traitement par un inhibiteur de la traduction, le cycloheximide. Extrait de la bactérie *Streptomyces griseus*, cette molécule interfère dans la translocation des ARN de transfert dans le ribosome et bloque ainsi l'élongation du polypeptide en cours de synthèse.

Classiquement, la quantification par western blot après plusieurs temps de traitement permet de suivre la décroissance de la quantité de protéines au cours du temps et d'en extraire la demi-vie. Pour une protéine fluorescente comme dans notre cas, une façon plus rapide et moins contraignante est de quantifier la fluorescence de cette protéine par cytométrie en flux, après différents temps de traitement. On s'affranchit ainsi des étapes fastidieuses d'extraction de protéines, d'électrophorèse, de transfert et d'hybridation sur membrane.

L'indice calculé pour caractériser la durée de vie d'une protéine sera sa demi-vie. Elle correspond au temps pour lequel 50% de la quantité initiale a été dégradé en présence de l'inhibiteur de la traduction. La demi-vie des protéines stables et déstabilisées a donc été estimée par cytométrie en flux dans 2 clones : le clone BE5 exprime les protéines déstabilisées et le clone C3 les protéines stables. Le traitement au cycloheximide (50µg/ml) est réalisé pendant 1 à 5 heures (Figure 89).

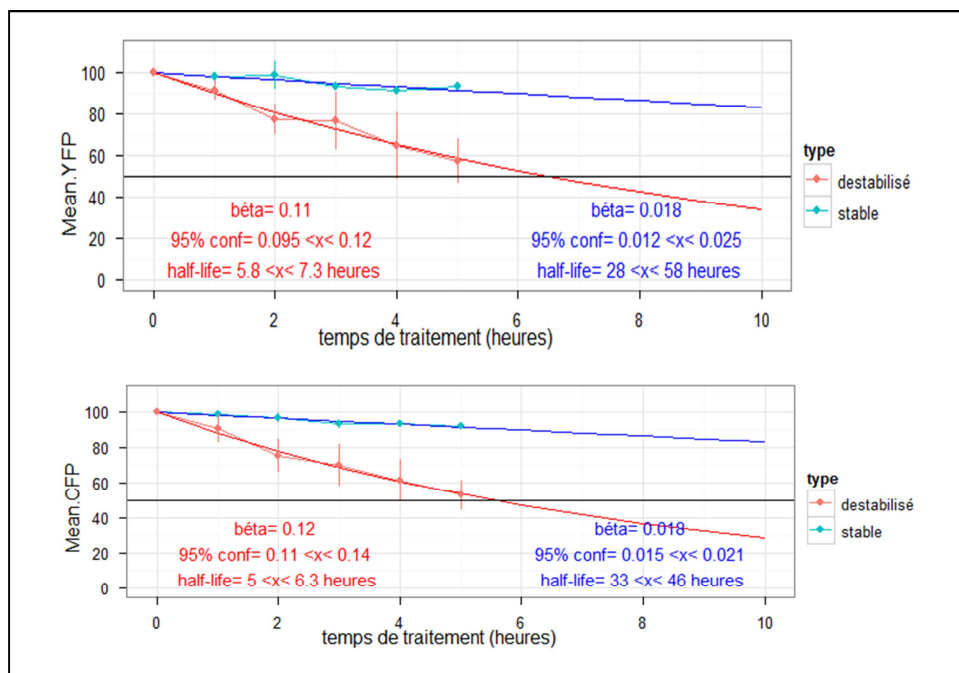


**Figure 89** Cinétique d'inhibition de la traduction.

A gauche : clone C3 parental exprimant les protéines stables. A droite : Clone BE5 C/Y exprimant les protéines déstabilisées.

La distribution des fluorescences dans chacun des clones pendant le traitement montre clairement que le clone BE5 est sensible à l'action du CHX contrairement au clone C3. Dans le premier cas, la distribution se rapproche progressivement du contrôle négatif au cours du temps alors qu'aucun changement significatif n'est observé dans le second cas.

La fluorescence moyenne d'eCFP et d'eYFP de la population traitée à différents temps est reportée sur la Figure 90. A chaque point de temps, elle est normalisée par rapport à la condition sans traitement (100% de fluorescence). Dans le cas des constructions codant les protéines déstabilisées (Lignes rouges), une décroissance rapide apparaît à partir de la première heure de traitement. Cette décroissance est très similaire entre les deux protéines déstabilisées.



**Figure 90 Estimation de la demi-vie des protéines fluorescentes.**

Les lignes bleues correspondent aux protéines non-déstabilisées, les lignes rouges aux protéines déstabilisées. *En haut* : Évolution de la moyenne d'eYFP. *En bas* : Évolution de la moyenne d'eCFP. Bêta correspond au coefficient de décroissance de la régression, il est donné avec son intervalle de confiance à 95%. La demi-vie est égale à  $\ln(2)/\beta$ .

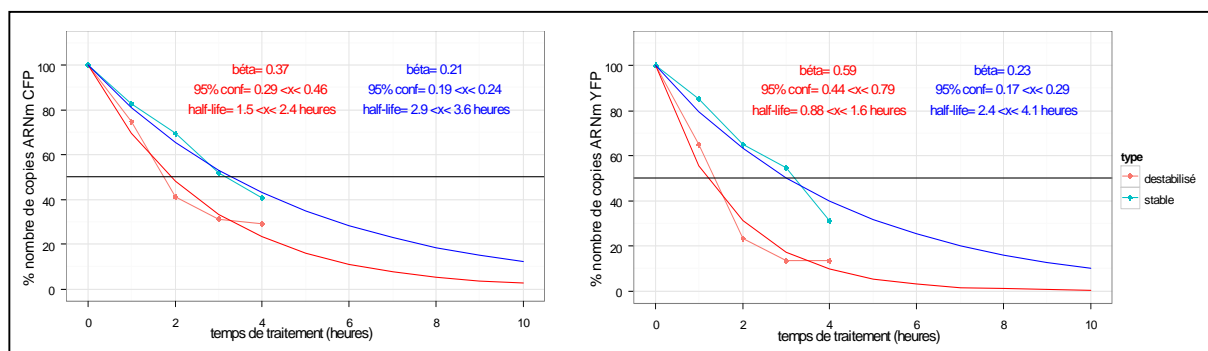
Une régression non-linéaire (de type  $y=100.\exp(-\beta.t)$ ) sur chacune des MFI au cours du temps permet d'estimer le coefficient de décroissance (bêta) ainsi que la demi-vie de chaque type de protéines ( $T_{1/2}=\ln(2)/\beta$ ). La demi-vie des protéines stables eCFP et eYFP est estimée 39h30 +/-6h et 43h +/-8h respectivement. La littérature fait état d'une durée de vie supérieure à 26 heures pour la protéine GFP classique ce qui correspond à l'estimation faite ici. Un temps de traitement supérieur permettrait d'améliorer la prédiction, mais augmente significativement la mortalité cellulaire qui ne permet plus de quantifier la fluorescence.

La décroissance de la fluorescence des protéines déstabilisées a été déterminée de la même façon. Leur demi-vie est estimée à 5h40 +/-40min et 6h30 +/- 40min pour eCFP et eYFP respectivement. La déstabilisation des constructions a donc pour conséquence la réduction de la demi-vie des protéines fluorescentes d'un facteur 6. Ainsi, un arrêt de synthèse protéique d'une heure se traduira par une chute de 8% du niveau de fluorescence contre seulement 1.7% avec les protéines stables ce qui devrait améliorer grandement la résolution temporelle des analyses par vidéo-microscopie.

### II.2.c. Estimation de la demi-vie des ARNm déstabilisés

La déstabilisation des constructions plasmidiques a été réalisée en introduisant un signal de dégradation pour la protéine mais aussi un signal de dégradation des ARNm. Nous avons donc estimé la demi-vie des ARNm déstabilisés. Elle a été déterminée par qRT-PCR après traitement des cellules exprimant les transgènes par un inhibiteur de la transcription, l'actinomycine-D, pendant 1 à 4 heures. Après ce délai de traitement, les cellules sont comptées, les ARN totaux sont extraits, rétro-transcrits en cDNA, puis quantifiés par qRT-PCR.

La quantité d'ARN totaux par condition est standardisée par la quantité de cellules afin de s'affranchir d'une différence de quantité de cellules entre les conditions. Cette différence introduirait un biais dans la décroissance observée de la quantité de mRNA. La quantité théorique d'ARNm est aussi corrigée vis-à-vis de la quantité utilisée pour la retro-transcription (1 $\mu$ g). Cette quantité ne représente pas la même proportion de la quantité totale dans chaque condition et doit être prise en compte. La Figure 91 représente la décroissance normalisée des ARN messagers de chacun des transgènes, après plusieurs du temps de traitement par de l'actinomycine-D.



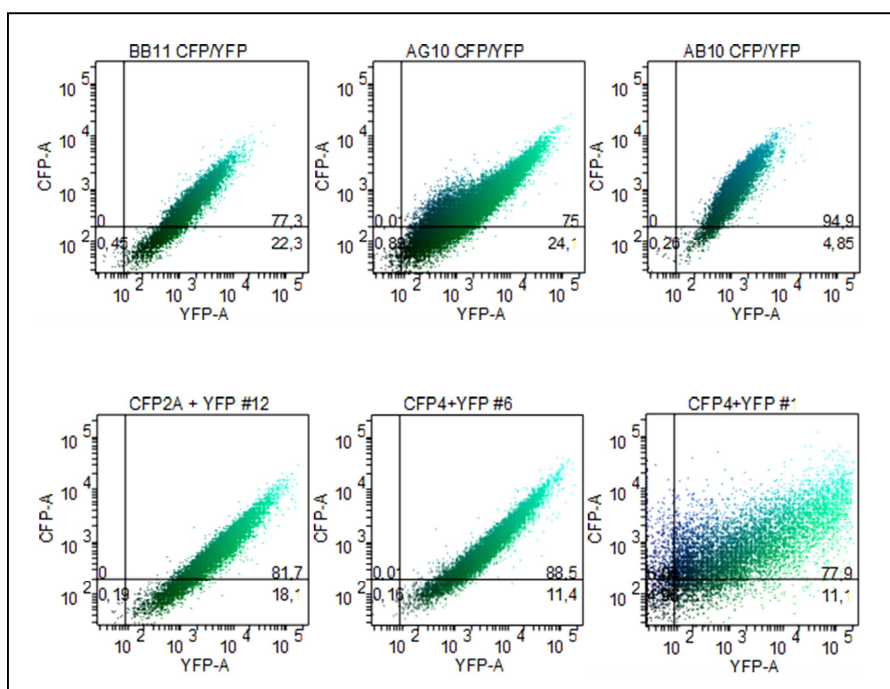
**Figure 91 Estimation de la demi-vie des ARNm eCFP et eYFP.**

Les courbes bleues correspondent aux ARNm déstabilisés et les rouges aux ARNm natifs.

La demi-vie de chacun des ARNm est estimée à 1h30 pour YFPdes et 2h heure pour CFPdes, soit près de deux fois plus faible que celle des ARNm de la version non-déstabilisée (eCFP : 3 heures et eYFP : 3 heures).

## II.2.d. Caractérisation des clones par cytométrie en flux

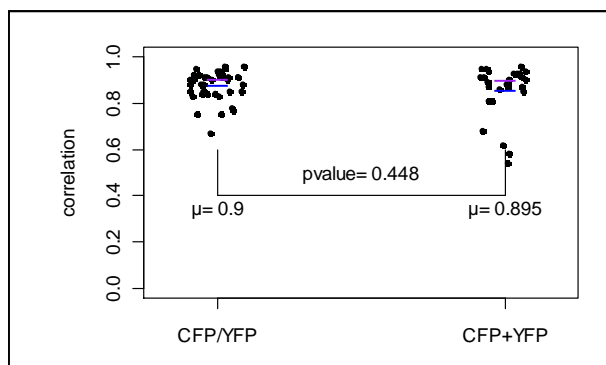
Après amplification, les clones dépendants et indépendants sont analysés par cytométrie en flux afin de caractériser leur niveau moyen de fluorescence de chaque transgène et leur variabilité. Trois clones « dépendants » et « indépendants » représentatifs sont présentés sur la Figure 92. Le reste des clones analysés est présenté dans les annexes 1 à 6.



**Figure 92 Clones dépendants et indépendants exprimant les protéines déstabilisés.**

Les trois premières vignettes sont des clones dépendants (*eCFP/eYFP*) et les trois dernières des clones indépendants (*CFP + YFP*). Le gradient de couleur symbolise le niveau d'expression de chaque transgène et les quadrants les sous-populations exprimant aucun, un seul ou les deux transgènes.

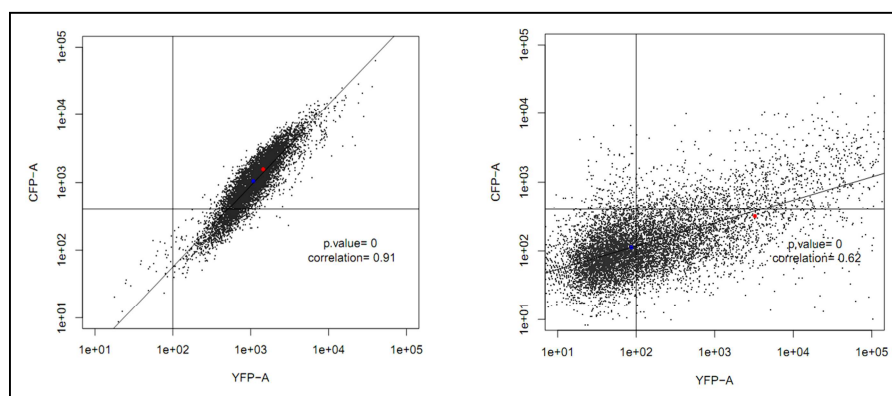
On observe sur ces 6 exemples représentatifs, des niveaux d'expressions des deux transgènes plus ou moins bien corrélés entre les cellules comme le montre la diagonale obtenue sur chaque bi-plot. Chaque sous-clone a un profil d'expression différent. Très peu de cellules expriment un seul des transgènes. Les variations non corrélées sont plutôt visibles lorsque la fluorescence est faible. On voit alors des cellules s'éloigner de la diagonale établie par le reste de la population. La corrélation entre le niveau de fluorescence de l'eCFPdes et de l'eYFPdes est importante (0.9 et 0.89) pour les deux catégories de clones (Figure 93).



**Figure 93** Corrélation de l'expression des deux transgènes par catégorie de clones.

L'expression des deux transgènes est bien corrélée dans les clones dépendants et indépendants. Ligne rouge : médiane, ligne bleue : moyenne.

La différence de corrélation entre les clones dépendants (C/Y) et indépendants (C+Y) n'est pas significative ( $p\text{-value} = 0.45$ ). Seuls certains clones présentent une corrélation plus faible, autour de 0.5 qui se traduit par une plus grande dispersion des cellules (Figure 94, à droite). Dans les autres clones, le niveau de fluorescence des deux transgènes par cellule est très bien corrélé.

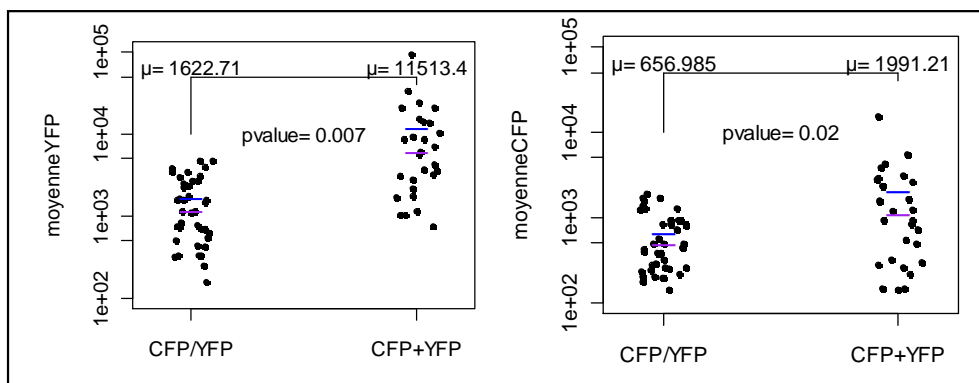


**Figure 94** Analyse en cytométrie de deux clones exprimant les transgènes déstabilisés.

Le profil d'expression des deux transgènes est différent dans les deux clones. Le premier, plus homogène, présente une bonne corrélation des deux transgènes (0.91), le second plus variable, une corrélation plus faible (0.62). Point rouge : médiane, point bleu : moyenne.

Si l'on étudie l'expression des deux transgènes dans les différents types de clones établis, on note que le niveau moyen d'expression des deux transgènes est plus élevé dans les clones indépendants (Figure 95). Cette observation est plus visible pour le transgène eYFP, qui est surexprimé d'un facteur 7 (MFI eYFP = 1622 vs 11513,  $p\text{-value} < 0.01$  et MFI eCFP = 657 vs 1991,  $p\text{-value} < 0.05$ ).

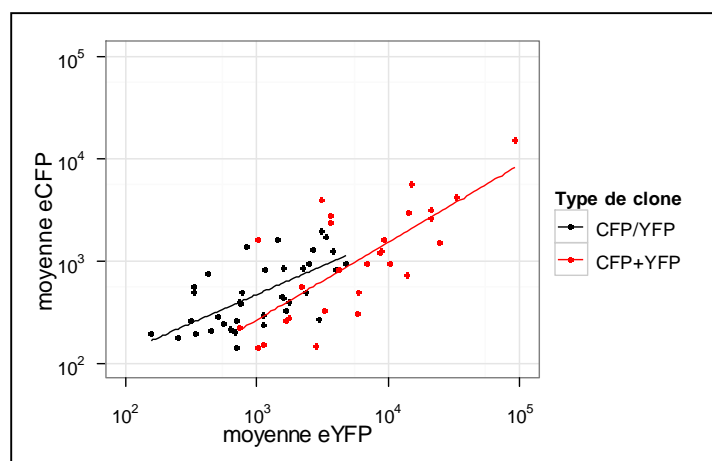




**Figure 95 Niveau moyen d'expression de chaque transgène suivant la catégorie.**

*A gauche : expression de eYFP. A droite : expression d'eCFP. Le niveau d'expression des deux transgènes est significativement supérieur dans les clones indépendants (CFP+YFP). Ligne rouge : médiane, ligne bleu : moyenne.*

Le niveau d'expression plus élevé des deux transgènes dans les clones indépendants s'accompagne d'une meilleure corrélation entre les niveaux moyens d'expression des transgènes dans ces clones (Figure 96). La corrélation atteint 0.62 pour les clones dépendants (en noir) et 0.91 pour les clones indépendants (en rouge).



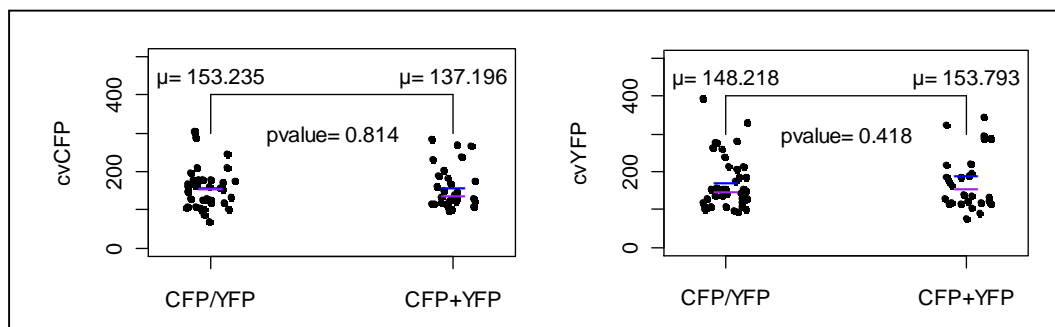
**Figure 96 Corrélation de l'expression moyenne des deux transgènes.**

*La corrélation entre les niveaux d'expression des deux transgènes est plus importante dans la catégorie « indépendant ».*

Paradoxalement, on s'attendait à ce que le niveau moyen d'expression des deux transgènes des clones dépendants soit mieux corrélé. En effet, les deux transgènes étant au même locus, le contexte chromatinien est très probablement plus homogène que dans les clones indépendants.

On remarque également sur les graphiques de cytométrie, que le profil d'expression de l'eYFPdes et de l'eCFPdes dans les deux catégories de clones, est relativement variable,

s'étalant sur plus d'un log (Figure 92, Figure 94 et annexes 1 à 6). La variabilité estimée par le coefficient de variation de chaque transgène permet de se rendre compte que les deux types de clones ont la même variabilité (Figure 97). Cette observation est vraie pour chacun des transgènes (145 vs 152% pour eYFP et 132 vs 134% pour eCFP). Rappelons que le CV est un indice normalisé.



**Figure 97 Variabilité de l'expression des transgènes suivant la catégorie.**

*A gauche : CV du transgène eCFP dans les clones « indépendants » et « dépendants ». A droite : CV du transgène eYFP. La variabilité intra-clone n'est pas différente entre les deux groupes de clones, quelque soit le transgène. Ligne rouge : médiane, ligne bleue : moyenne.*

A ce stade, plusieurs hypothèses peuvent être formulées quant à cette différence d'expression dans les deux catégories de clones. La première est qu'un biais de sélection a eu lieu lors du choix des deux clones exprimant eCFP déstabilisé pour la construction des clones indépendants. La deuxième hypothèse est que dans les clones dépendants, une compétition locale s'est installée au niveau des deux promoteurs pour la machinerie de transcription. Si cette compétition existe, il en résultera une plus grande variabilité dans l'expression des deux transgènes dans ces clones ainsi qu'une moins bonne corrélation entre les deux transgènes dans chaque cellule.

### II.3. Analyse des fluctuations par imagerie en temps-réel

Pour mesurer les variations de niveau d'expression dans ce modèle exprimant les protéines déstabilisées, le protocole décrit en première partie a été suivi. Un arrêt de synthèse des protéines fluorescentes déstabilisées pendant 30 minutes se traduira par une diminution de fluorescence d'environ 6% et, pendant 1 heure, de 11%, ce qui est tout à fait quantifiable.

### II.3.a. Acquisition des images

Les cellules ont été cultivées dans des boîtes de culture spécialement conçues pour l'appareil d'imagerie (Biostation IMQ, Nikon). Contrairement au modèle de cellules exprimant les protéines stables, on remarque au premier coup d'œil que la déstabilisation des protéines permet de mettre en évidence des changements d'expression des transgènes (Figure 98). Ces changements sont visibles à l'échelle de quelques heures et peuvent intervenir plusieurs fois dans l'intervalle d'un cycle cellulaire.



**Figure 98 Vidéo-microscopie de cellules exprimant les protéines déstabilisées.**

*Ce clone dépendant exprime les protéines eCFP et eYFP déstabilisées. On remarque que l'expression des deux transgènes est dynamique par le changement fréquent de couleur des cellules (eYFP en vert et eCFP en rouge). Le délai entre deux images est de 2 heures.*

Dans cet exemple, l'hétérogénéité s'établit très rapidement. Après 2 divisions cellulaires, les 4 cellules sœurs expriment les transgènes à différents niveaux.

Pour les deux catégories de clones, dépendants et indépendants, plusieurs d'entre eux ont été analysés par imagerie temps-réel (Tableau 12). La durée d'analyse s'est limitée à un maximum d'un cycle cellulaire. Les cellules dont le tracking dure moins de 5 heures (15 points de temps) ont été éliminées. Pour chacun des clones, une vingtaine et jusqu'à une centaine de cellules ont été analysées.

Clone	AD2	AB2	AB7	AD6	AE5	BG9		CFP2A-1	CFP2A-2	CFP2A-3	CFP4-1	CFP4-2	CFP4-6	CFP4-3
Type	CFP/YFP	CFP/YFP	CFP/YFP	CFP/YFP	CFP/YFP	CFP/YFP		CFP+YFP	CFP+YFP	CFP+YFP	CFP+YFP	CFP+YFP	CFP+YFP	CFP+YFP
Nombre de cellules	40	110	110	23	17	30		19	28	34	141	29	23	104

**Tableau 12 Bilan des clones analysés par imagerie temps-réel**

La différence de nombre de cellules analysées s'explique principalement par la difficulté à détecter la fluorescence dans certains clones.

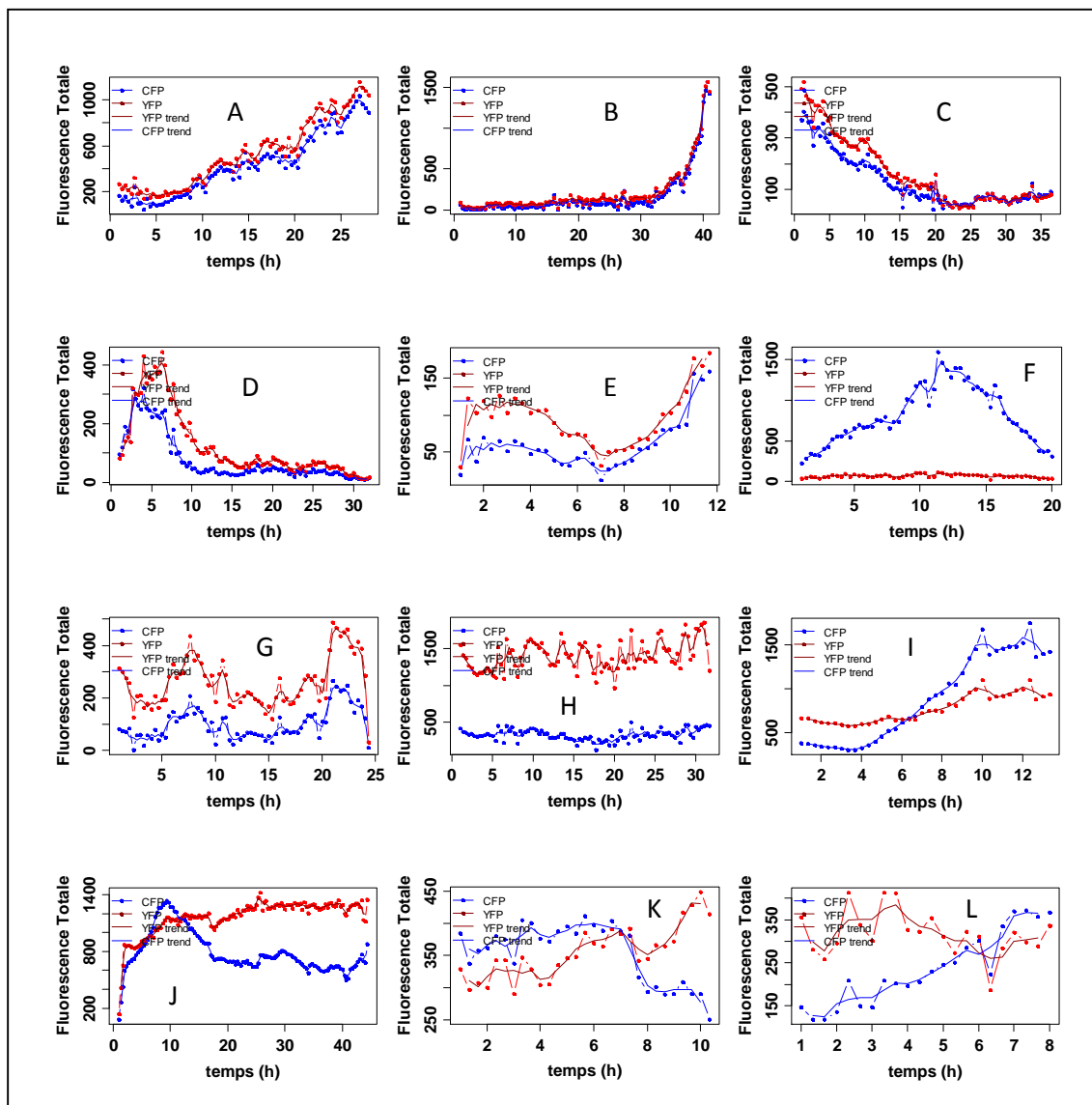
### *II.3.b. Exploitation des données brutes*

Les images de plusieurs cellules ont donc été analysées avec le logiciel CellProfiler® selon le mode opératoire décrit en matériel et méthodes. Les données brutes ont été vérifiées et corrigées si nécessaire (problème de tracking ou de segmentation), puis analysées avec le logiciel R.

On observe sur les données brutes issues des deux catégories de clones, différentes dynamique d'expression des transgènes (Figure 99 , Annexe 7 et Annexe 8). Cette dynamique est parfois identique entre les deux transgènes mais peut aussi différer sur une partie ou la sur la totalité du tracé. Le comportement majoritairement observé est une augmentation de la fluorescence au cours du temps, qui traduit une accumulation de protéines au cours du cycle cellulaire. L'augmentation peut être très progressive (apparemment linéaire, Figure 99, A) mais parfois elle se produit en une courte période de temps (exponentielle, Figure 99, B). Elle peut être continue ou marquée de périodes pendant lesquelles la fluorescence diminue.

On remarque aussi des cellules pour lesquelles la fluorescence d'un ou des deux transgènes décroît au cours du temps (Figure 99, C, D, E, F, K). La décroissance peut être continue ou marquée d'évènements de réexpression qui ralentissent la diminution de fluorescence. La décroissance observée dans la plupart des cas est compatible avec un arrêt total d'expression. En effet, on observe une réduction de fluorescence de 50% en 6 heures qui correspond à la demi-vie des protéines fluorescentes déstabilisées. La déstabilisation est

clairement une amélioration non-négligeable qui permet maintenant d'observer les arrêts de synthèse protéique.



**Figure 99 Exemples de dynamiques d'expression des transgènes.**

*La dynamique est très variable entre cellules, appartenant parfois au même clone, dans l'intervalle d'un cycle cellulaire. Les lignes pleines correspondent au signal brute et les pointillées au lissage par une moyenne mobile. Les lignes bleues et rouges correspondent aux transgènes eCFP et eYFP. Ces exemples appartiennent aux deux catégories de clones.*

On note aussi dans ces quelques exemples, certaines cellules dont le niveau de fluorescence semble évoluer de façon périodique. Il oscille dans l'intervalle de temps de l'expérience (Figure 99, D, E, F). La fréquence est souvent faible mais peut se révéler extrêmement élevée dans certains cas (Figure 99, G et H). Pour ces deux cellules, G et H, le niveau de fluorescence est constant autour d'une moyenne au cours du cycle cellulaire. Néanmoins, il existe une différence entre les niveaux d'expression de deux transgènes dans

la même cellule. La différence d'expression ne semble pas biaisée. On aurait pu supposer que la présence du transgène eCFP devant le transgène eYFP sur les constructions plasmidiques eYFP/eCFPdes puisse favoriser son expression. Cela ne semble pas être le cas, chacun des transgènes pouvant être exprimé plus fortement que le second, et cela dans les cellules d'une même clone.

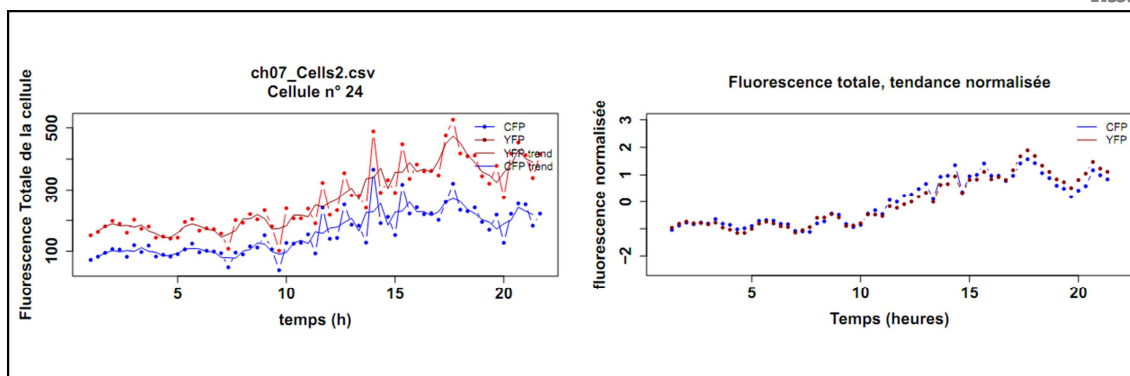
Dans certaines cellules enfin, les deux transgènes ne se comportent pas de la même façon (Figure 99, F, I, J, K, L). La différence entre les transgènes s'observe parfois tout au long de l'expérience mais peut aussi apparaître au cours de l'expérience (Figure 99, J et K).

Nous observons ainsi des différences de tendance globale des signaux (augmentation, décroissance ou stabilité ou périodicité). De plus, il existe des différences entre signaux dans une même cellule. Ces différents types de signaux ont été analysés avec des outils d'analyse de séries temporelles, tels que l'auto-corrélation et l'inter-corrélation. Le premier objectif est de décrire comment se comportent les deux transgènes l'un vis-à-vis de l'autre. Nous avons cherché des dynamiques qui seraient caractéristiques des clones dépendants ou indépendants. Dans un second temps, nous avons analysé séparément chacun des transgènes afin de capturer leur dynamique temporelle et notamment leur périodicité qui traduit les évènements d'extinction et d'activation.

### *II.3.c. Lissage et normalisation des signaux*

---

Les tracés bruts sont lissés pour réduire les fluctuations techniques. Le signal de fluorescence des deux transgènes est décomposé en une tendance grâce à la fonction « Decompose » de la bibliothèque « stat » (Figure 100, gauche, ligne pointillée). La tendance consiste en une moyenne mobile appliquée sur chaque signal avec une fenêtre étroite et symétrique.



**Figure 100 Lissage des signaux bruts et normalisation.**

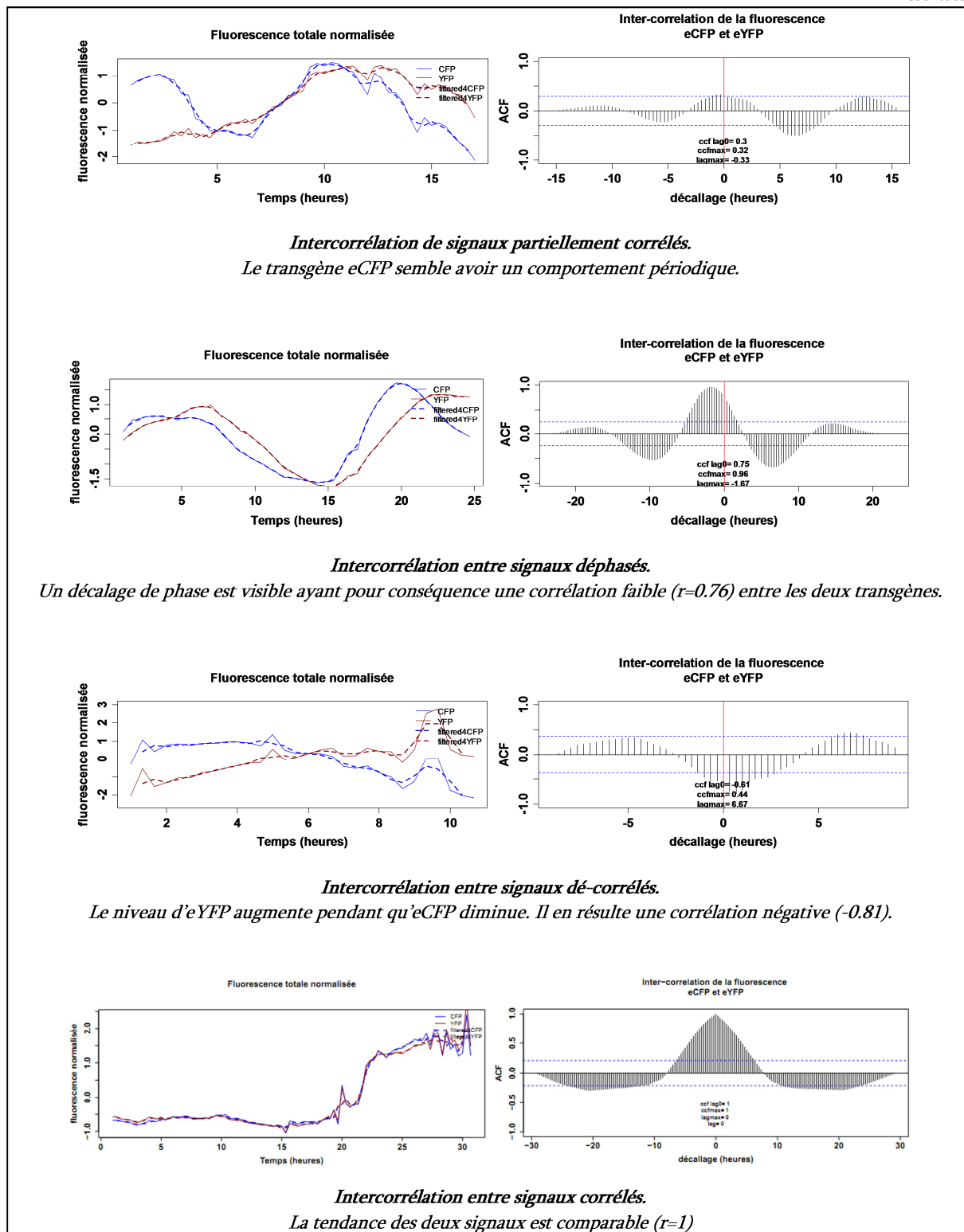
*Le lissage est effectué grâce à une moyenne mobile et la normalisation par centrage et réduction.*

Dans un second temps, la représentation est normalisée (centrées et réduites,  $\mu=0$  et  $SD=1$ ) afin de mieux mettre en évidence les différences de dynamique entre signaux d'intensités différentes (Figure 100, droite).

#### *II.3.d. Comparaison de la dynamique des transgènes*

La corrélation entre les deux signaux au cours du temps est calculée grâce à l'inter-corrélation. Elle permet de se rendre compte de la synchronisation de l'expression des deux gènes rapporteurs. Plusieurs exemples d'inter-corrélations sont illustrés sur la Figure 101.

Les deux premiers signaux témoignent du caractère périodique de l'expression des deux transgènes. Dans le premier, les deux transgènes sont partiellement corrélés alors que dans le second, ils sont légèrement déphasés. Il en résulte une inter-corrélation faible. Le troisième exemple représente deux signaux dont la tendance est totalement inversée. L'inter-corrélation qui en résulte est négative. Le quatrième exemple correspond à une dynamique identique de l'expression des transgènes avec pour conséquence une corrélation parfaite (=1).



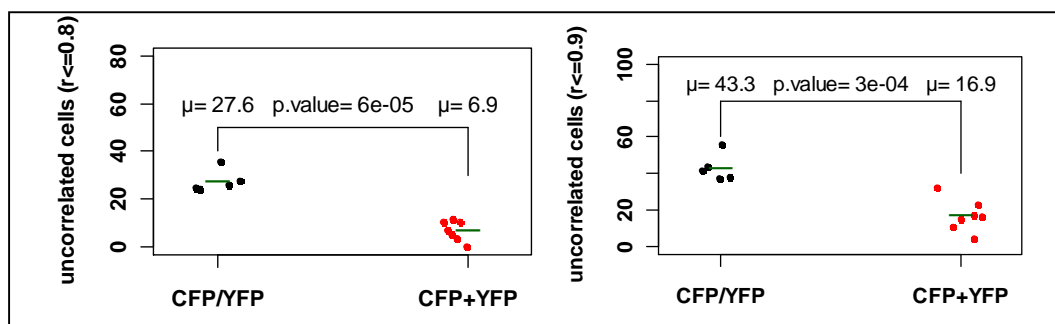
**Figure 101 Exemples d'inter-corrélation calculée sur des signaux variés.**

Les graphiques de gauche représentent l'évolution de la fluorescence des deux transgènes dans des cellules uniques. Ceux de droite sont les inter-corrélations entre les signaux des deux transgènes.

Pour chaque cellule de chaque clone analysé, l'inter-corrélation entre les deux fluorescences est donc mesurée. Deux seuils arbitraires ont été déterminés pour lesquels les cellules sont considérées comme bien corrélées : un premier seuil correspond à une



corrélation de 0.8 et un seuil plus élevé de 0.9. Pour les cellules dont la corrélation est inférieure à ce seuil, l'expression des deux transgènes est considérée comme non synchronisée. Le pourcentage de cellules non-synchronisées par clone est représenté ci-dessous :



**Figure 102 Pourcentage de cellules ayant une expression non-corrélées des transgènes.**

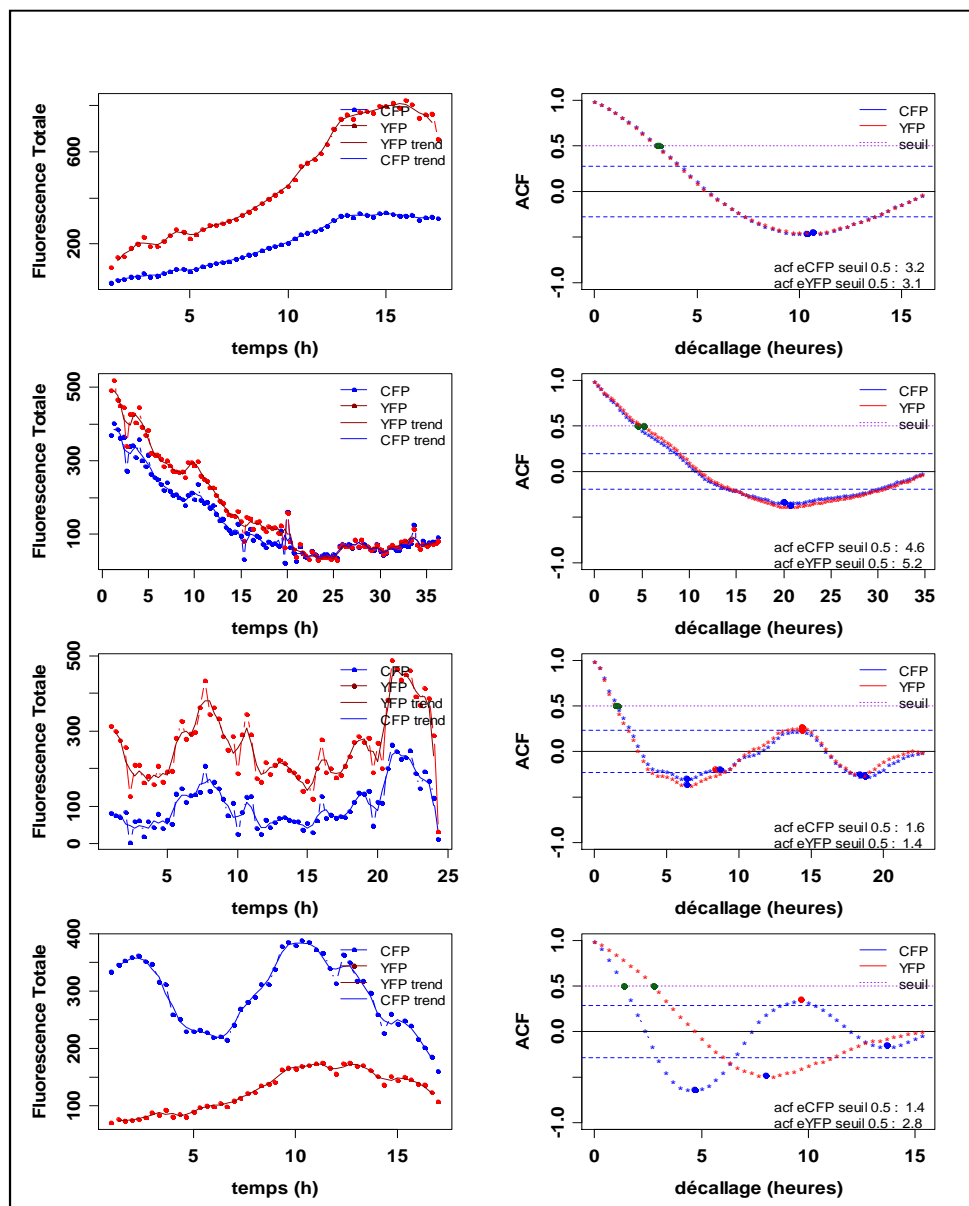
Le pourcentage est calculé pour les clones dépendants (CFP/YFP) et indépendants (CFP+YFP) avec deux seuils. A gauche : pour un seuil de corrélation de 0.8. A droite : pour un seuil de 0.9.

Quel que soit le seuil utilisé, il y a significativement plus de cellules dont l'expression des deux transgènes est désynchronisée dans les clones dépendants CFP/YFP. Parmi les cellules désynchronisées 11% sont anti-corrélées dans la catégorie CFP/YFP et alors qu'aucune n'est visible dans la catégorie CFP+YFP. Malgré leur site d'intégration identique, l'expression des deux transgènes est moins bien corrélée, ce qui est surprenant. Dans la première partie, l'idée d'une compétition locale entre les deux promoteurs pour des facteurs de transcription a été formulée pour expliquer la différence de niveau d'expression des deux transgènes dans les deux catégories de clones. Une telle compétition entraînerait une expression alternée des deux transgènes et ainsi, une dé-corrélation telle que nous la voyons principalement dans les clones dépendants (Annexe 7 et Annexe 8).

### II.3.e. Analyse de la dynamique des transgènes individuels

Afin d'analyser la dynamique individuelle de chacun des transgènes, une auto-corrélation a été calculée sur chaque signal de fluorescence. L'auto-corrélation est, comme son nom l'indique, la corrélation d'un signal par lui-même. Elle permet de détecter des régularités, des profils répétés dans un signal comme un signal périodique perturbé par beaucoup de bruit, ou une fréquence fondamentale d'un signal.

Une auto-corrélation forte indique un phénomène de persistance ou une forte inertie, qui se traduit par une forte dépendance des valeurs présentes vis-à-vis des valeurs passées. A l'inverse, une faible auto-corrélation peut être prédictive de comportements chaotiques, imprévisibles. La Figure 103 illustre différentes auto-corrélations calculées sur la dynamique des transgènes eCFP et eYFP dans 4 cellules.



**Figure 103 Exemples d'auto-corrélations sur différents type de signaux.**

Les graphiques de gauche sont les tracés bruts et ceux de droite les auto-corrélations de chacun des signaux.

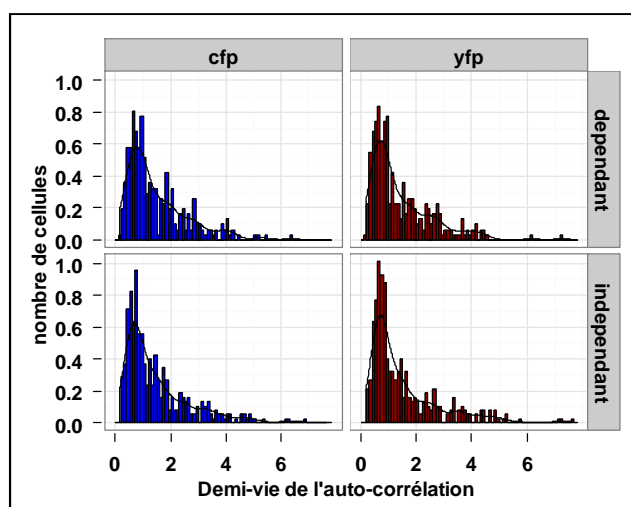
Lorsque les signaux sont constants (croissant ou décroissant) dans le temps (cas 1 et 2), l'auto-corrélation reste élevée plus longtemps par rapport à des signaux périodiques (cas 3 et 4) ou très variables (entre 3 et 5 heures contre 1 à 2 heures).

Lorsque les signaux sont périodiques, l'auto-corrélation possède aussi cette propriété (cas 3 et 4) et il est ainsi possible de déterminer la période des fluctuations. Il arrive que les deux signaux soient différents dans la même cellule et l'auto-corrélation est alors différente (cas4).

Afin de comparer l'auto-corrélation des cellules issues de chaque type de clone (dépendant ou indépendant), il est nécessaire de disposer d'un indice. Nous avons utilisé la « demi-vie » de l'auto-corrélation (ACF) qui correspond au décalage pour lequel l'auto-corrélation prend la valeur de 0,5 (Weinberger et al., 2008b). Ce seuil est représenté par une ligne pointillée violette sur le graphique d'auto-corrélation et l'intersection de chaque signal avec ce seuil est représentée par un point vert sur ce même graphique (Figure 103).

Il est donc possible de comparer l'ACF des deux fluorescences dans une cellule, mais aussi de comparer plusieurs cellules d'un même clone ainsi que les clones entre eux.

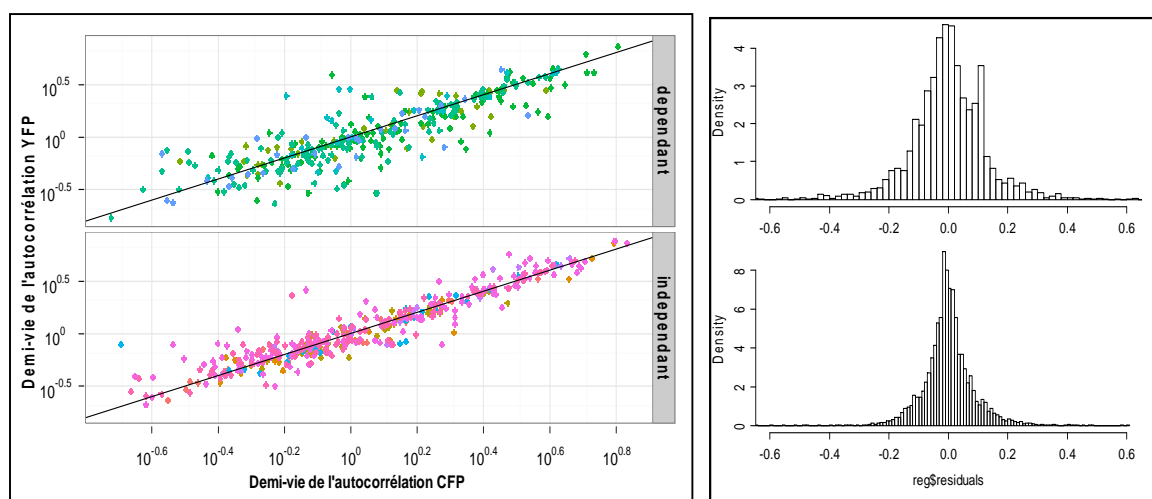
Le temps auquel l'auto-corrélation de chaque transgène atteint le seuil de 0.5 dans chaque cellule est mesuré et la distribution des demi-vies d'auto-corrélation pour les deux transgènes dans les deux catégories de clones est représentée sur la Figure 104.



**Figure 104 distributions des demi-vies d'auto-corrélation.**

*L'auto-corrélation est mesurée pour chaque transgène dans des cellules issues de clones dépendants ou indépendants.*

Il ne semble pas exister de différence moyenne entre la demi-vie de l'auto-corrélation des cellules issues des clones exprimant les deux transgènes à partir du même locus ou de loci différents (Figure 104,  $p > 0.05$ ). Cela signifie donc qu'individuellement, les transgènes se comportent globalement de la même façon. En revanche, lorsque l'on s'intéresse à la corrélation de la demi-vie de l'auto-corrélation de chacun des transgènes dans des cellules individuelles, il semble exister une différence (Figure 105, gauche).



**Figure 105** Corrélation des demi-vies d'auto-corrélation entre transgènes.

*Les histogrammes représentent la distribution des écarts à la régression linéaire représentée sur les graphiques de gauche. La variance est plus élevée dans les clones dépendants.*

Pour les deux types de clones, la majorité des cellules montre une bonne corrélation entre la demi-vie de l'auto-corrélation des deux transgènes. On observe néanmoins des cellules dont la demi-vie de l'auto-corrélation d'un des transgènes est supérieure. La corrélation semble moins bonne dans les cellules exprimant les transgènes à partir du même locus. Une régression linéaire est calculée sur le nuage de points formé par la demi-vie des deux auto-corrélations de chaque cellule (Figure 105, gauche, lignes noires). Elle permet ensuite de calculer la distance qui sépare chaque point et sa valeur projetée sur la droite de régression. Pour les deux types de clones, dépendant ou indépendant, la distribution des écarts à la régression est calculée (Figure 105, droite). On remarque clairement que la variance est plus importante pour les cellules issues de clones dépendants (test des variances d'Ansari,  $F=1.9$ ,  $p < 0.001$ ).

En conclusion de ces analyses par vidéo-microscopie, nous avons montré que la dynamique d'expression des transgènes peut être très variable. Grâce aux protéines déstabilisées, nous avons pu observer nous avons montré l'expression génique n'est pas un processus forcément linéaire et continu. Certaines cellules expriment les transgènes de façon constante alors que d'autre montrent une expression périodique. Des périodes de synthèse protéique sont entrecoupées de périodes d'inactivités pendant lesquelles le niveau de fluorescence diminue. Les analyses des signaux ont montré que les transgènes se comportent individuellement de la même façon dans les deux catégories de clones (démonstré par les analyses d'auto-corrélation). En revanche, dans une même cellule au même instant, leur expression est mieux coordonnées dans les clones indépendants. Il semble donc que la proximité des deux transgènes dans les clones dépendants puisse générer une compétition entre les deux promoteurs.

## Discussion et perspectives

---

Lorsque ce travail a été initié, la littérature dans le domaine de la variabilité de l'expression génique et de l'hétérogénéité intercellulaire était en pleine expansion (Figure 3). Depuis, les observations ont laissé place à des études mécanistiques et moléculaires, ainsi qu'à de nombreux travaux de modélisation de ce phénomène. Les travaux pionniers dans ce domaine émanent de laboratoires de physique et de biophysique, notamment celui de Van Oudenaarden, dans les années 2000. Ils ont réussi à contourner les difficultés techniques inhérentes à l'expérimentation en abordant cette thématique différemment. La majorité de ces travaux sont des modélisations mathématiques et des simulations informatiques, parfois couplés avec des expériences menées sur des organismes unicellulaires procaryotes dans le début des années 2000 puis eucaryotes (notamment la levure).

Ces travaux pionniers ont mis en évidence l'influence du faible nombre de partenaires impliqués dans les réactions biochimiques conduisant à une expression génique de nature stochastique. C'est une propriété importante à prendre en compte pour expliquer les variations d'expression génique mais dont les conséquences ne sont significatives que pour un nombre restreint de situations physiologiques connues à ce jour.

Rares sont les travaux qui ont traité de modèles eucaryotes supérieurs et par conséquent du rôle de l'organisation du génome, en particulier la chromatine, sur la variabilité de l'expression génique. L'organisation du génome et l'épigénétique nous ont donc semblé être une piste intéressante pour expliquer l'origine de la variabilité dans des populations de cellules eucaryotes ayant le même génome. En réalité, plus que la variabilité, c'est sa dynamique qui nous a intéressé dans ce travail.

## I. Choix du modèle cellulaire

---

Le choix du modèle cellulaire s'est porté sur une lignée de cellules eucaryotes, dérivée d'un rétinoblastome embryonnaire humain. Très utilisée au laboratoire, elle semblait adaptée à notre problématique. Afin d'étudier l'impact de la chromatine sur l'expression génique, il était nécessaire de pouvoir différencier la régulation liée au promoteur et aux éléments régulateurs associés, de celle liée à l'environnement chromatinien. Nous avons

donc, dans cet objectif, développé un modèle inspiré de celui d'Elowitz chez la bactérie (Elowitz et al., 2002).

Dans notre modèle, les deux transgènes sont intégrés aléatoirement à différents endroits dans le génome eucaryote. La variabilité intrinsèque définie par (Swain et al., 2002) est la combinaison de la stochasticité liée au faible nombre de molécules, à laquelle s'ajoute l'effet du contexte chromatinien à chaque locus. La stochasticité liée au faible nombre de molécules ne peut pas être directement mesurée dans ces modèles compte-tenu de l'influence de la chromatine qui s'y ajoute. Néanmoins, nous avons estimé qu'elle est comparable entre les différents clones. En effet, la différence de variation non-corrélée dans les clones parentaux mesurée par (Neildez-Nguyen et al., 2008) n'est pas équivalente. Elle peut atteindre un facteur 8, indiquant un rôle prépondérant de la chromatine sur la variabilité intrinsèque. Des différences d'expression dans ce modèle sont donc le reflet d'une dynamique chromatinienne différente à ces loci.

Dans le modèle que nous avons appelé « dépendant » dans la seconde partie de ce travail, les deux transgènes sont portés par la même construction et les promoteurs sont espacés d'une courte distance (Figure 42, page 65). Intégrés au même locus, le contexte chromatinien local des deux transgènes est donc similaire. On peut alors imaginer dans ce cas précis, que la différence d'expression observée entre les deux transgènes est la conséquence de la stochasticité de l'expression génique.

Ces modèles ont été développés dans l'objectif de bien comprendre les mécanismes et leur dynamique avant de proposer un rôle biologique. Ainsi, bien qu'aisément manipulable, la lignée cellulaire utilisée ne permet pas d'étudier de processus biologiques, comme la différenciation cellulaire, pouvant être influencés par l'hétérogénéité cellulaire. Néanmoins, on peut établir des connexions avec des applications biotechnologiques. En effet, certaines applications comme la thérapie génique ou la production de protéines recombinantes ont des similitudes avec notre modèle. Dans ces applications faisant appel au transfert de gène, l'intégration du transgène n'est presque pas maîtrisée, même si certaines études commencent à explorer la possibilité de cibler des loci particuliers, pour



s'affranchir d'une mutagenèse par insertion. Néanmoins, les seules d'entre-elles qui visent à limiter l'effet de position préfèrent se focaliser sur le design de la construction transgénique afin de maîtriser le niveau d'expression, sans se soucier de la variabilité. Il serait donc bénéfique pour ces applications de connaître l'origine des fluctuations observées mais aussi de disposer de moyens pour les limiter. On assiste depuis quelques années à un gain d'intérêt pour des séquences que l'on nomme « insulateur », de part leur capacité à isoler une séquence d'ADN de son environnement génomique immédiat, si elles sont placées de part et d'autre de celle-ci (Yusufzai and Felsenfeld, 2004; Gaszner and Felsenfeld, 2006). Dans cette étude, nous n'avons pas utilisé de telles séquences mais prévoyons leur utilisation pour la suite de ce projet afin de vérifier leur bénéfice sur la variabilité de l'expression génique. Si ces séquences ont la faculté de prévenir la condensation de la chromatine au locus du transgène, l'expression de celui-ci devrait être plus constante, tout au moins débarrassée des fluctuations liées aux changements d'état de la chromatine. Il serait enfin envisageable de vérifier cette hypothèse grâce aux techniques adoptées dans ce travail (imagerie temps réel, cytométrie en flux et traitement du signal).

Dans les applications biotechnologiques citées plus haut, le promoteur utilisé n'est pas toujours endogène. Seule la thérapie génique utilise des promoteurs à régulation physiologique afin de limiter toute toxicité liée à la surexpression du transgène. Pour la production de protéines recombinantes, la problématique n'est pas la même. On utilise souvent des promoteurs forts, d'origine virale, pour produire une grande quantité de protéines. Le produit d'expression du transgène ne joue souvent aucun rôle dans le système-hôte, comme les protéines fluorescentes de notre modèle (hormone de croissance humaine, EPO, hémoglobine, interféron-alpha... produits par des plants de tabac ou des organismes bactériens).

Dans nos modèles, les transgènes sont sous le contrôle de promoteurs constitutifs d'origine virale. Le CMV utilisé ici est considéré comme un promoteur fort compte-tenu du nombre de facteur de transcription pouvant influencer son activité (142 facteurs identifiés avec

TESS<sup>1</sup> (Schug, 2008) avec plus de 1000 sites de fixation putatifs). Ce promoteur n'est *a priori* pas impliqué dans des boucles de rétroactions puisque qu'aucune autre séquence virale, notamment la protéine E2 ne sont présentes. Néanmoins, compte-tenu du nombre de facteurs cellulaires susceptibles de se fixer au promoteur, on ne peut pas affirmer qu'il est totalement indépendant de mécanismes cellulaires. Le fait qu'il s'agisse de promoteurs viraux peut être critiqué et il sera intéressant de tester d'autres types de constructions à l'avenir. Néanmoins, l'intérêt de ce modèle réside dans la présence de ces deux promoteurs identiques que nous comparons. Le fait de comparer l'expression des transgènes dotés de ces promoteurs au sein d'une même cellule, permet de s'affranchir de la nature du promoteur.

Bien entendu, nous avons confirmé que la régulation de l'expression des transgènes se situait à l'étape de la transcription, avant de procéder à la suite des expériences. Une corrélation forte est mesurée entre le niveau d'ARNm et la quantité de protéines (estimée par le niveau moyen de fluorescence) dans diverses populations cellulaires de phénotype sélectionné.

L'autre avantage du modèle est que les gènes rapporteurs codent pour des protéines fluorescentes inactives d'un point de vue biochimique qui n'ont aucun rôle biologique. Elles sont facilement détectables et quantifiables constituant donc un outil puissant pour l'analyse de l'expression génique. Cette indépendance des promoteurs vis-à-vis du reste de la cellule, et la neutralité fonctionnelle des protéines fluorescentes, permettent de mesurer uniquement la dynamique propre à leur expression, sans influence d'une régulation par un réseau complexe de rétroaction, tels que ceux décrits en introduction. Notre modèle est donc un outil idéal pour l'étude de la variabilité de l'expression des transgènes dans le cadre de ces applications ou, devrions-nous plutôt dire dans ce cas, de la stabilité de leur expression.

Cette dualité entre variabilité et stabilité de l'expression génique a été la question principale de cette étude afin de déterminer quelle dynamique était la plus à même

---

<sup>1</sup> Transcription Element Search System at <http://www.cbil.upenn.edu/cgi-bin/tess/tess>

d'expliquer l'hétérogénéité observée dans notre modèle et quelles en étaient les conséquences.

## II. Macro-hétérogénéité et dynamique chromatinienne

---

La variabilité de l'expression génique dans notre modèle a été initialement observée par cytométrie en flux. Cette analyse statique donne un aperçu d'un échantillon de la population à un instant donné. Ces premières observations, confirmées dans cette étude, ont permis de conclure au rôle indéniable joué par le site d'intégration, notamment par la présence de cellules n'exprimant qu'un seul ou aucun des transgènes. Dans ce modèle, les transgènes codent pour des protéines inertes d'un point de vue biochimique et il n'y a pas lieu de penser qu'elles sont impliquées dans des boucles de rétroaction. La bistabilité observée, qui aboutit à ce que l'on a défini en introduction comme la macro-hétérogénéité, est donc envisagée comme la conséquence de la dynamique chromatinienne au locus de chacun des transgènes.

Afin de mieux comprendre ce contexte génomique, les sites d'intégration de deux clones parentaux exprimant les protéines stables ont été identifiés. Nous avons analysé la présence de séquences répétées, la composition en îlots CpG ou les modifications épigénétiques déjà identifiées dans d'autres types cellulaires. Aucune particularité n'a été mise en évidence, si ce n'est que les transgènes sont intégrés dans des régions introniques. Nous avons établi que les gènes dans lesquels se sont intégrés les transgènes sont exprimés normalement. Malheureusement, nous ne savons pas s'il s'agit d'une expression de l'allèle ciblée ou du second allèle du gène. Compte-tenu du faible nombre de sites d'intégration à disposition (deux par clone), il est difficile d'établir une corrélation forte entre le contexte génomique des transgènes et leur dynamique individuelle. Un plus grand nombre de clones parentaux est nécessaire afin de tirer une conclusion significative.

De nombreux mécanismes épigénétiques décrits en introduction sont susceptibles de modifier la dynamique chromatinienne et par voie de conséquence l'expression génique au locus cible. Il a été démontré que le promoteur CMV était susceptible d'être réprimé

par méthylation dans certains types cellulaires et en fonction de l'espèce (Sinclair et al., 1992; Prösch et al., 1996; Radhakrishnan et al., 2008). Chez le rat par exemple, ce processus est dominant et l'expression du transgène diminue au cours du temps après transduction par un adénovirus (Brooks et al., 2004). Chez la souris, ce processus ne semble pas être responsable de l'inactivation du promoteur CMV (Löser et al., 1998). De même, le promoteur CMV a une activité qui dépend fortement du type cellulaire chez les mammifères (Nan et al., 1997; Mehta et al., 2009; Qin et al., 2010). C'est une observation qui a été confirmée dans cette étude en collaboration avec le laboratoire CGMC de Lyon. En utilisant les mêmes plasmides codant les protéines déstabilisées mais dans un modèle d'érythroblastes aviaires (cellules primaires 6C2), l'expression des gènes rapporteurs n'est quasiment pas détectable. Nous avons donc procédé à une analyse de l'état de méthylation des îlots CpG au niveau des promoteurs CMV des deux transgènes. Cette analyse a été réalisée d'une part sur des cellules exprimant les deux transgènes et, d'autre part, sur des cellules ne les exprimant pas après tri cellulaire. Dans les deux cas, les séquences analysées (environ 30 îlots CpG) se sont révélées entièrement déméthylées. Ceci indique que la méthylation locale des îlots CpG du promoteur CMV ne semble pas influencer l'expression de nos transgènes bien que nous n'excluons pas que la méthylation de séquences plus éloignées puisse être responsable de l'inactivation des transgènes dans notre modèle.

La méthylation des CpG est la plus connue mais ne représente qu'un type de modifications épigénétiques susceptibles de modifier l'expression génique. Des analyses par immunoprécipitation de la chromatine (ChIP) se focalisant sur les modifications d'histones les plus courantes comme la méthylation ou l'acétylation (acH3K4, meH3K9) sont envisagées. L'hypothèse de modifications touchant les histones est privilégiée car le traitement de cellules par de la TSA (inhibiteur d'HDAC) a pour effet d'augmenter l'expression génique, indiquant qu'il pourrait exister des modifications, notamment de l'acétylation au niveau des histones (Condreay et al., 1999; Grassi et al., 2003). Cette analyse nous permettra de mieux comprendre le lien entre l'expression ou l'absence d'expression des transgènes dans notre modèle et le contexte chromatinien local. Cette analyse, comme l'étude de la méthylation du promoteur décrite précédemment, nous renseignera sur le contexte chromatinien local. Malheureusement, nous n'avons pas d'informations sur le contexte

génomique plus global qui est aussi susceptible d'avoir un impact sur l'expression des transgènes. Nous avons uniquement vérifié que les gènes les plus proches, situés à plus de 100kb de part et d'autres des transgènes étaient exprimés de façon normale. Le repliement de la chromatine et la présence de régions transcriptionnellement plus actives dans le noyau sont d'autres éléments susceptibles d'influencer l'expression d'un gène, mais ne sont pas identifiables par les techniques employées ici.

Les observations faites sur les clones parentaux semblent indiquer que le site d'intégration des transgènes joue un rôle majeur dans leur niveau d'expression moyen mais aussi sur leur variabilité d'expression. En effet, en fonction des clones parentaux (donc du site d'intégration), le niveau d'expression moyen varie et il n'y a pas de corrélation entre les deux transgènes comme l'ont montré (Neildez-Nguyen et al., 2008)). De plus, le profil d'expression diffère entre les clones. La variabilité intrinsèque et extrinsèque sont différentes ce qui abouti à des formes des profils de cellules positives plus ou moins compactes. La proportion de cellules exprimant le/les transgènes, ou aucun des deux, est également différente entre les clones. Sans en connaitre exactement le mécanisme, le contexte chromatinien dans notre modèle semble donc un acteur majeur de la variabilité observée dans les populations clonales.

### III. Analyse dynamique de la variabilité

---

Nous avons montré que chacun des clones établis présentait des caractéristiques en termes d'expression des deux transgènes (niveau moyen et variance). De façon intéressante, ce profil d'expression n'a pas changé de façon significative depuis 5 ans. Dans chaque clone, la proportion de cellules exprimant les deux transgènes, ou ayant arrêté d'en exprimer, est quasiment identique. Cette observation est contradictoire avec le fait que la population s'est diversifiée à partir d'une seule cellule d'origine lors du clonage. Le système doit donc suivre une dynamique.

Il faut donc envisager deux hypothèses pour expliquer cette stabilité apparente. La première hypothèse soutien que le phénotype de chaque cellule est très stable au cours du

temps mais peut subir des variations brusques. Ainsi, la population demeure elle-même stable au cours du temps et des évènements rares peuvent survenir et faire dériver lentement le phénotype. Dans une seconde hypothèse, la répression et l'activation des transgènes dans une même cellule sont deux processus en équilibre dynamique. La fréquence de transition entre ces deux états reste à définir, mais pourrait dans un cas extrême aboutir à un comportement ergodique de chaque cellule (Figure 38). Cette propriété de stabilisation du phénotype ou la probabilité de changement phénotypique serait alors une propriété liée au site d'intégration. Comme décrite en introduction, la localisation chromosomique est un facteur important dans l'expression d'un gène de part l'environnement génomique proximal mais également par les interactions plus distales qui peuvent exister. De même, des modifications épigénétiques sont susceptibles de modifier la dynamique chromatinienne et, par conséquence, l'expression du gène. Ceci expliquerait alors la différence que nous observons entre les clones parentaux mais aussi entre les cellules d'un même clone parental qui définissent les différentes sous-populations positives et négative. Nous avons donc caractérisé la dynamique d'expression des transgènes afin de déterminer si le profil d'expression de chaque population clonale était un état stationnaire immuable ou un système dynamique pris à un instant donné. En d'autres termes, l'expression des transgènes dans une cellule individuelle est-elle stable au cours du temps ou est-elle soumise à des fluctuations, montrant des propriétés d'ergodicité ? Nous avons pour cela adopté deux approches expérimentales complémentaires, s'appuyant chacune sur un modèle cellulaire adapté.

Dans une première approche, nous avons fait l'hypothèse qu'en isolant une cellule possédant un phénotype donné (i.e. un niveau d'expression du/des transgènes) et en observant le profil de la population obtenue, il était possible de mettre en évidence le caractère variable ou non de l'expression génique. Cette approche a déjà été utilisée dans d'autres études pour montrer la réversibilité d'un phénotype (Chang et al., 2008; Pilbrough et al., 2009; Stockholm et al., 2010). Nous avons simplement reproduit le clonage effectué lors de la construction du modèle en sous-clonant différentes cellules de phénotypes particuliers à partir de populations parentales.

Nous avons montré que le sous-clonage d'une cellule positive présentant un phénotype donné permettait après un temps de culture relativement long (50 jours), de générer une population avec un profil d'expression proche de celui de la population parentale. Ces cellules, sélectionnées sur leur niveau élevé ou faible de fluorescence, tendent vers un niveau d'expression intermédiaire comparable à celui de la population parentale d'origine. Dans ces deux catégories, un faible pourcentage de cellules négatives apparaît traduisant la répression du transgène. De plus, certains sous-clones deviennent totalement négatifs, préférentiellement parmi les cellules exprimant faiblement le transgène lors du clonage. Il peut s'agir de cellules transcriptionnellement inactives, dont la fluorescence est toujours détectable. Après clonage, ces cellules et leur descendance restent transcriptionnellement inactives et perdent peu à peu leur fluorescence jusqu'à devenir négatives. Ce raisonnement est valable uniquement si la durée de vie de protéines fluorescentes est suffisamment longue par rapport au turnover des cycles d'activation et d'inhibition transcriptionnelle. Si la durée de vie est très longue, la fluorescence n'aura pas le temps de diminuer entre l'arrêt de synthèse et une nouvelle période d'activité transcriptionnelle. Il faudrait une période d'inactivité extrêmement longue pour obtenir des cellules négatives. Une analyse par photo-blanchiment de cellules entières a mis en évidence la présence de cellules fluorescentes n'étant pas transcriptionnellement actives. Ceci confirme d'une part que certaines cellules fluorescentes de la population n'expriment pas le transgène (estimé à 50% des cellules) et d'autre part que la durée de vie des protéines peut jouer un rôle important dans l'atténuation des fluctuations ainsi qu'un possible rôle mémoire. On note aussi que la majorité des cellules initialement négatives peuvent réactiver l'expression du transgène, certaines maintenant malgré tout la répression pendant la durée de l'expérience.

Nous avons proposé précédemment deux hypothèses pour expliquer la stabilité apparente du profil de chaque clone parental : un équilibre dynamique entre activation et répression de l'expression des transgènes et une stabilité à long terme, qui revient à attribuer des probabilités extrêmement faibles aux transitions entre état actif et inactif. Cette expérience de sous-clonage montre que la première hypothèse semble la plus probable. Il semble en effet exister à cette échelle de temps, deux mécanismes antagonistes. D'une

part, il existe une mémoire intrinsèque qui ralentie la diversification phénotypique. Cette mémoire semble contrecarrer les fluctuations et fait dériver le phénotype lentement. Il est ainsi toujours possible de deviner le phénotype d'origine d'une population jusqu'à une vingtaine de jours après le sous clonage. Passé ce délai, on assiste à ce que l'on pourrait qualifier de retour à l'équilibre de la population, le niveau moyen d'expression de chaque sous-clone dérivant peu à peu vers celui de sa population parentale. Cette mémoire de la population est similaire à une valeur de consigne qui force le système à s'y rapprocher malgré les perturbations. Cette mémoire pourrait, comme décrit plus haut, être une propriété intrinsèque du site d'intégration conditionnant la probabilité de répression et d'activation du transgène. De façon statistique, après un grand nombre de divisions cellulaires, le phénotype de l'ensemble des individus tendrait vers un équilibre et ceci de façon reproductible. Les fluctuations et la « mémoire » du phénotype à long terme auraient alors la même origine, le site d'intégration du transgène et son contexte épigénétique.

Nos observations permettent aussi de conclure qu'il existe une différence de « potentiel » dans le sens où il semble plus aisé d'activer un transgène lorsqu'il est inactif que de réprimer son expression lorsqu'il est actif. Cette différence de dynamique se retrouve dans tous les clones parentaux et pourrait être liée à la nature des modifications épigénétiques au locus des transgènes et à leur turnover. Même si les modifications évoquées sont réversibles, leur cinétique n'est pas forcément la même. Théoriquement, si les modifications « activatrices » et « répressives » avaient la même probabilité d'occurrence, la proportion de cellules exprimant ou non le/les transgènes devrait être équivalente dans les clones parentaux, ce qui n'est pas le cas. Établies depuis plus de 5 ans, on peut considérer que ces populations ont atteint un état d'équilibre et nous observons que la majorité des cellules (>80%) expriment les deux transgènes. Nous avons montré qu'il n'existe pas de différence dans le profil de méthylation des motifs CpG au niveau des promoteurs CMV et des expériences sont planifiées pour mettre en évidence des modifications au niveau des histones.



## IV. Etude des changements d'expression rapides

---

Dans la première approche, nous avons mis en évidence que des changements phénotypiques avaient lieu et généraient de l'hétérogénéité à l'échelle de la population. Dans une seconde approche, complémentaire au sous-clonage, nous avons souhaité mesurer les fluctuations d'expression des transgènes à plus court terme et à l'échelle de la cellule unique afin de caractériser leur amplitude et leur fréquence.

Nous avons suivi par vidéo-microscopie des cellules isolées exprimant les deux transgènes codant les protéines ayant une longue durée de vie. Dans un premier temps, cette expérience a été réalisée pendant 12 heures puis sans durée fixe, la seule limite étant la densité cellulaire. L'analyse visuelle et quantitative des images ont révélé des changements phénotypiques intervenants avec une occurrence relativement faible.

En regardant des films de vidéo-microscopie, on remarque deux phénomènes distincts. Certaines cellules, comme dans l'expérience de sous-clonage, maintiennent stablement leur phénotype au cours de l'expérience. En prenant les dernières images du film, on peut ainsi repérer des sous-populations qui sont issues de ces cellules, notamment les cellules n'exprimant qu'un seul des transgènes qui sont faciles à suivre. D'autres cellules ont un phénotype qui dérive lentement si bien qu'à la fin de l'expérience on ne peut pas en déterminer le phénotype original.

Il semble donc coexister deux dynamiques opposées dans l'expression des transgènes. Une répression complète et durable qui participe à l'effet mémoire et une certaine variabilité qui a tendance à contrecarrer cet effet mémoire à l'échelle cellulaire, se traduisant à l'échelle de la population clonale par une augmentation de l'hétérogénéité.

Afin d'approfondir cette réflexion, nous avons quantifié l'expression des deux transgènes dans des cellules isolées et leur descendance. Durant moins d'un cycle cellulaire, aucune variation n'est observée sur une trentaine de cellules. A travers plusieurs générations, la majorité des cellules analysées suit la même dynamique : la fluorescence totale de chacun des transgènes augmente progressivement pendant la durée du cycle cellulaire, puis chute

d'un facteur 2 lors de la division. L'augmentation de fluorescence pendant le cycle ne semble pas linéaire mais exponentielle, cette observation restant à confirmer. Seules quelques cellules arrêtent d'exprimer un des transgènes, et cet arrêt de synthèse n'est visible qu'après plusieurs heures de répression. Cette expérience a ainsi soulevé un problème concernant la durée de vie des protéines fluorescente que nous n'avions pas envisagé : plus celle-ci est importante, plus elle va masquer l'activité à court terme du promoteur qui régule son expression. La demi-vie de nos protéines stables est estimée entre 40 et 50 heures, on estime qu'il faut environ 385 heures d'inactivité pour observer une extinction totale. Ceci correspond à 7 cycles cellulaires en prenant en compte la dilution de moitié de la fluorescence à chaque cycle. La fluorescence étant tellement stable, peu de fluctuations ont été observées sur une période de 12 heures.

La limite liée à la durée de vie des protéines fluorescentes utilisées dans ce modèle nous a amené à adopter une approche alternative. Pour mesurer la variabilité d'expression malgré cette limite, nous avons réduit physiquement et irrémédiablement le niveau de fluorescence dans des cellules individuelles, puis mesuré l'augmentation de fluorescence liée à la synthèse de nouvelles protéines.

Nous avons remarqué que la moitié des cellules dont la fluorescence avait été réduite, conservaient un niveau très faible de fluorescence après l'extinction. Ce niveau est comparable avec celui de cellules traitées avec des inhibiteurs de la transcription et de la traduction. Cette observation indique que certaines cellules de la population ne sont pas transcriptionnellement actives mais demeurent fluorescentes. La période d'inactivité transcriptionnelle du transgène semble alors relativement longue, estimée à plus de 8 heures. D'autres cellules, comme escompté, retrouvent un niveau de fluorescence comparable à celui mesuré avant l'extinction, indiquant que le transgène est exprimé au moment du photo-blanchiment.

Cette observation est intéressante et va dans le sens où la durée de vie de la protéine est visiblement une limite à l'analyse de l'expression génique. En effet, la durée de vie des protéines est tellement longue que malgré l'inactivité du transgène, les cellules continuent à émettre de la fluorescence masquant ainsi la nature pulsative de l'expression génique.

Ces observations peuvent aussi expliquer la différence de fréquence de répression de l'expression du transgène observée dans l'expérience précédente de sous-clonage. On peut imaginer que les cellules ayant un faible niveau de fluorescence sont des cellules dont l'expression du transgène a cessé depuis un certains temps et dont le niveau de fluorescence à eu le temps de décroître. Malheureusement, lors des expériences de photo-blanchiment, les paramètres d'observation ont été adaptés à chaque cellule afin que le niveau de fluorescence soit quantifiable (non saturant mais suffisamment fort). On ne peut dans cette expérience établir de corrélation entre le niveau de fluorescence initial et la capacité à synthétiser de nouvelles protéines après le photo-blanchiment.

Dans cette même expérience, des cellules ont été traitées avec une drogue connue pour son activité pro-transcriptionnelle, la TSA. Cette molécule est un inhibiteur d'Histone Déacétylases et permet, par voie de conséquence, l'augmentation de niveau d'acétylation des histones. Cette modification réduit les interactions ADN/histone augmentant la probabilité de transcription d'un gène. Ainsi, le traitement devait permettre aux cellules transcriptionnellement inactives de retrouver leur fluorescence après photo-blanchiment. Curieusement, aucune amélioration du retour de fluorescence n'a été observée suite au traitement. Nous observons même l'effet opposé, aucune cellule ne retrouvant son niveau de fluorescence initial. Le niveau de récupération est équivalent à celui observé sans drogue. Peut-être s'agit-il d'un problème d'échantillonnage, le nombre de cellules analysées étant réduit. Il peut aussi s'agir d'un artefact lié au délai d'action de cette drogue. En effet, nous avons montré par la suite que 48h sont nécessaires pour observer une augmentation du niveau du niveau fluorescence en cytométrie en flux (données non-présentées). En tenant compte de cette observation, nous prévoyons de renouveler ce genre d'expériences en incluant d'autres drogues connues pour leur effet sur la chromatine (5-azacytidine ou butyrate).

Cette expérience a montré que la synthèse protéique est rapide dans ce modèle cellulaire. Après le photo-blanchiment, les cellules qui retrouvent un niveau de fluorescence comparable à celui d'origine le font en 4 à 5 heures, quel que soit leur niveau de fluorescence initial. Cette différence de dynamique entre la synthèse et la dégradation est

probablement une des raisons limitant l'analyse des fluctuations à court terme dans notre modèle. Seules des augmentations importantes de niveaux de fluorescence pourraient être visibles avec ce genre de protéines stables, compte-tenu de leurs demi-vies élevées.

Cette propriété est une limite dans notre système expérimental qui nous amène à réfléchir sur l'impact que la durée de vie des protéines peut avoir dans une situation physiologique. L'augmentation de la durée de vie de certaines protéines serait alors un moyen efficace de réduire le bruit lié à leur expression stochastique. Cette hypothèse a déjà été confortée par des données montrant une relation inverse entre le niveau de bruit et le niveau moyen d'expression de protéines dans divers systèmes biologiques (Bar-Even et al., 2006). En revanche, une durée de vie accrue est un inconvénient dans les mécanismes cellulaires nécessitant une réaction d'adaptation rapide comme la transmission de signaux ou l'induction d'un gène. Le seul moyen d'y remédier serait de pouvoir moduler la dégradation de ces protéines pour que leur concentration s'adapte rapidement à un arrêt de synthèse. Malheureusement, peu de données existent sur cette dégradation ciblée, la majorité des travaux se focalisant sur l'expression génique et la synthèse protéique. Elle est souvent considérée comme constante au cours du temps, se modélisant par une courbe exponentielle décroissante et se caractérisant par sa demi-vie. Puisque la dégradation protéique est comme les différentes étapes de l'expression génique, majoritairement liée à des activités enzymatiques, il est fort probable qu'elle soit soumise aux mêmes genres de fluctuations spatiales, temporelles et quantitatives liées à l'activité de ces enzymes catalytiques (assemblage du protéasome, ubiquitination, phosphorylation, oxydation...).

Il existe quelques exemples de protéines qui peuvent faire évoluer leur stabilité suivant les conditions physiologiques de la cellule. Par exemple, le facteur HIF-1 (Hypoxia Induced factor) est un hétéro dimère obligatoire qui peut se lier à l'ADN au niveau de séquences appelées HER (Hypoxia Response Element), présentes au niveau de gènes comme l'EPO ou le VEGF. La sous-unité  $\alpha$  de cet hétéro dimère est instable en condition de normoxie, sa demi-vie étant limitée à 5 minutes maximum (Jewell et al., 2001). Elle est ainsi synthétisée puis dégradée de façon continue en normoxie par le protéasome, suite à son hydroxylation et à son ubiquitylation par la protéine von Hippel-Lindau. En absence

d'oxygène, un cofacteur des enzymes d'hydroxylation, la sous-unité  $\alpha$  n'est plus dégradée et s'accumule instantanément dans le noyau où elle peut se dimériser avec la sous-unité  $\beta$  (Jewell et al., 2001; Berra et al., 2003; Depping et al., 2008). Il a été démontré, dans cet exemple, que la synthèse de HIF-  $\alpha$  entre normoxie et hypoxie est très similaire, la régulation se faisant majoritairement au niveau de la dégradation de la protéine (Fandrey et al., 2006).

## V. Déstabilisation des protéines et fluctuations à court terme

---

En cytométrie en flux, la plage dynamique importante permet de mettre en évidence des différences de niveaux de fluorescence subtiles entre cellules. En revanche, pour une application en imagerie temps-réel, les protéines stables ne sont pas adaptées et aucune variation de l'expression des transgènes n'est visible. Face à cette limitation technique, nous avons adapté le modèle d'origine pour améliorer la résolution temporelle de notre analyse en temps-réel.

En collaboration avec le laboratoire CGMC à l'université de Lyon, de nouvelles constructions plasmidiques ont été obtenues. Ces nouvelles constructions diffèrent par plusieurs aspects. Le plus important étant la déstabilisation des rapporteurs (au niveau des ARNm et des protéines). Nous avons évalué ces nouvelles constructions et montré que la déstabilisation dans notre lignée cellulaire était efficace, puisque la demi-vie des protéines a été réduite d'un facteur 10 (de 55 heures à 5h30) et le niveau de fluorescence d'un facteur 50 dans 3 clones différents. Une telle demi-vie nous a paru plus adaptée avec notre application d'imagerie temps réel même si paradoxalement, dans certaines conditions, la fluorescence a été difficile à visualiser et à quantifier. C'est notamment le cas juste après la mitose dans certains clones, ce qui entraîne des difficultés lors de l'analyse des images. La détection des cellules étant effectuée à partir de leur fluorescence, nous avons vérifié que la faible fluorescence de certaines cellules n'était pas une limite à leur analyse. Une segmentation manuelle a montré une erreur d'estimation de la taille cellulaire de moins de 3% lorsque la fluorescence est faible. Une amélioration aurait été d'introduire un marquage membranaire lors de l'analyse en temps réel. Néanmoins, l'erreur mesurée

n'étant pas excessive, nous avons pris le parti de ne pas introduire de tels marquages supplémentaires.

Nous avons construit deux types de clones avec les nouvelles constructions à disposition. Le premier est identique au modèle précédent, à l'exception de la stabilité des protéines. Les transgènes sont portés par deux constructions indépendantes et transfectés successivement. Ainsi, les deux transgènes sont intégrés à deux loci différents. Nous avons aussi construit des clones par transfection des transgènes portés par le même plasmide. Dans ce cas là, les deux transgènes sont intégrés au même locus. Pour chaque clone étudié, que les deux transgènes soient au même locus ou non, une analyse du/des sites d'intégration à été débutée. L'analyse est réalisée par nos collaborateurs à Lyon (Gandrillon O. et Vinuelas J.). Des résultats préliminaires nous ont permis de sélectionner les clones ne possédant qu'un seul site d'intégration pour chacun des transgènes pour la suite des expériences.

L'analyse par cytométrie en flux des clones appartenant aux deux catégories montre que le niveau de fluorescence moyen dans les clones dépendants est statistiquement plus faible que celui des clones indépendants. L'hypothèse que nous pouvons avancer à ce stade est qu'une compétition s'installe lorsque les deux promoteurs sont sur la même construction. Cette compétition aurait alors pour conséquence l'activation alternée de chacun des promoteurs réduisant ainsi leur période d'activité et donc la quantité d'ARNm synthétisés.

Après avoir caractérisé ce nouveau modèle, l'analyse par vidéo-microscopie a été effectuée sur plusieurs clones, dépendant ou indépendant. La déstabilisation de la protéine et de l'ARNm semble avoir été efficace car nous avons clairement observé des changements d'expression des transgènes à court termes. En effet, ces changements peuvent intervenir plusieurs fois par cycle cellulaire. A l'échelle de plusieurs jours, ces changements d'expression rapides ont pour conséquence d'accélérer la diversification phénotypique. Contrairement au premier modèle dont les protéines sont stables, nous n'observons pas de cellules phénotypiquement stables indiquant une perte de l'effet mémoire évoqué dans le contexte du premier modèle. Il est alors envisagé de reproduire l'expérience de sous

clonage afin de vérifier si effectivement, des cellules isolées à partir d'une population parentale peuvent reconstituer cette population plus rapidement que dans le premier modèle. Le changement de la durée de vie des protéines fluorescentes nous a permis de mieux distinguer leur dynamique transcriptionnelle même si la résolution temporelle que nous approchons n'est que de 5h. Réduire encore la durée de vie des protéines actuelles serait inutile compte-tenu des difficultés rencontrées pour la détection de fluorescence trop faible dans certains clones. La solution serait peut être de remplacer les protéines fluorescentes actuelles par des protéines plus « brillantes » qui malgré leur faible durée de vie serait toujours détectables.

L'analyse de cellules isolées par vidéo-microscopie nous a néanmoins informés sur la dynamique individuelle des transgènes. Les tracés ont été analysés par deux outils mathématiques, l'auto-corrélation et l'inter-corrélation. Le premier outil montre que la dynamique des deux transgènes est très similaire tout type de clones confondus (indépendant ou dépendant). Il n'y a statistiquement pas plus de périodicité dans l'une ou l'autre des situations. En revanche, si l'on étudie des cellules individuelles, l'auto-corrélation de chacun des transgènes peut varier. De façon statistique, le nombre de cellules dont l'auto-corrélation d'eCFP et d'eYFP est différente est plus important quand les transgènes dans ces cellules, sont au même locus.

Le second outil utilisé, l'inter-corrélation, mesure la corrélation entre deux séries temporelles, ici l'évolution de la fluorescence des deux protéines. Les résultats obtenus par cette analyse montrent qu'il y a statistiquement plus de cellules dont l'expression des deux transgènes n'est pas corrélée dans les clones dits « dépendant », où les deux gènes rapporteurs sont intégrés au même locus.

Ces observations suggèrent donc que dans ce type de clones, la proximité des deux promoteurs (moins de 2kb) puisse engendrer une compétition. Cette compétition pourrait se faire vis-à-vis de facteurs de transcription ou d'autres protéines impliquées dans la transcription. Cette observation est contradictoire avec les observations effectuées dans un système similaire, mais sur la quantité d'ARNm synthétisés à partir d'un même locus, ou

de deux loci différents (Raj et al., 2006). Les auteurs montrent que la transcription est identique entre les deux transgènes lorsqu'ils sont intégrés au même locus. L'objectif prochain est de construire des clones cellulaires, toujours avec les mêmes gènes rapporteurs, mais dont l'espacement entre les promoteurs serait accentué progressivement jusqu'à perdre cette compétition locale.

La vidéo-microscopie a permis d'observer la dynamique d'expression des transgènes au cours du temps, particulièrement durant le cycle cellulaire. Ainsi, on note clairement que la quantité de protéines augmente au cours du cycle cellulaire, pour chuter de moitié lors de la division. Même si le promoteur utilisé ne dépend pas d'une régulation liée au cycle cellulaire, il semble profiter de la dynamique globale d'expression de la cellule pendant le cycle. L'analyse de cellules isolées ainsi que de leur descendance indique que la quantité de protéines héritée (mesuré par le niveau de fluorescence) est similaire entre cellules filles. Ceci est probablement lié au fait que le niveau d'expression des transgènes est très élevé dans ce modèle. Plus la quantité de molécules est importante, moins les petites différences seront importantes. Cette observation est aussi valable dans le second modèle dans lequel les protéines sont pourtant déstabilisées. Il semble donc que l'asymétrie liée à la division soit une source minime de variabilité. Il reste néanmoins que la quantité d'ARNm hérité n'a pas été mesurée dans ce travail et pourrait participer à l'hétérogénéité observée comme proposé par (Darzynkiewicz et al., 1982; Huh and Paulsson, 2011a, 2011b).

La seule observation commune entre les deux modèles sur l'ensemble des expériences d'imagerie temps-réelle, est l'apparition d'une différence de durée de cycle cellulaire (donnée non-présentées). La distribution de la durée des cycles cellulaires, déterminée à partir de l'ensemble des expériences de vidéo-microscopie, est très variable. Elle ne suit pas une loi normale mais ressemble plus à une distribution log-normale, d'une moyenne de 22 heures, et pouvant atteindre dans les cas extrêmes 16h et jusqu'à 40 heures. La durée du cycle cellulaire est bien corrélée entre cellules-sœurs mais aussi entre cellules-sœurs et leur cellule mère. En revanche, la corrélation diminue après une génération. Cette variabilité peut avoir une incidence sur les niveaux de fluorescence observés puisque les



cellules qui se divisent plus tôt que la moyenne n'ont pas le temps de synthétiser autant de protéines que les autres. Ceci entraîne une variabilité de la quantité de protéines entre cellules qui a déjà été observée dans d'autres systèmes (Sigal et al., 2006; Tsuru et al., 2009; Canela-Xandri et al., 2010; Cookson et al., 2010). Dans notre modèle, la contribution de ce phénomène sera uniquement en faveur du bruit extrinsèque, et donc non spécifique à un des transgènes. En revanche, il faut être prudent si l'on utilise un modèle ne possédant qu'une seule fluorescence. La variance de la distribution observée ne sera pas uniquement le résultat de fluctuations intrinsèques, qui pourront être alors surestimées si les cellules ne sont pas synchronisées (Sigal et al., 2006). Il semble donc qu'une grande part de la variabilité extrinsèque observée soit liée au cycle cellulaire. Un facteur 2 correspond à l'augmentation lors du cycle cellulaire, auquel il faut ajouter un facteur indéterminé (entre 2 et 4) lié à la variabilité de la durée du cycle d'une cellule à l'autre dans la population.

## VI. Des observations différentes entre les deux modèles

---

Afin de capturer la variabilité sur une échelle de temps allant de quelques heures jusqu'à plusieurs générations de cellules, nous avons dû adapter le modèle d'origine, tout en gardant ses avantages.

Chacun des clones obtenus avec les constructions modifiées pour déstabiliser les protéines a été analysé par cytométrie en flux. Plusieurs observations majeures ont été effectuées. La première est que le bruit intrinsèque est faible, à l'exception de quelques clones. Peu de cellules négatives ou exprimant un seul des transgènes sont visibles, ce qui était l'observation majeure dans le modèle d'origine avec les protéines non-déstabilisées. Nous avons intuitivement pensé que le bruit intrinsèque dans les clones possédant les transgènes au même locus serait plus faible compte-tenu du contexte chromatinien commun. Cette hypothèse n'a pas été confirmée, le coefficient de variation de chacun des transgènes dans les deux groupes de clones étant identique.

Dans ces nouvelles constructions plasmidiques, les transgènes eCFP, eYFP ou eCFP/eYFP sont entourés par des séquences TOL. Ces séquences sont reconnues par une transposase

qui est co-transfectée avec les plasmides. Elle va exciser le transgène du reste de la construction afin de l'intégrer dans le génome de la cellule hôte, au niveau de séquences TOL homologues.

Ceci a deux conséquences majeures par rapport au premier modèle. Dans un premier temps, il peut y avoir un biais dans le choix du site d'intégration puisqu'il s'agit d'un processus séquence dépendant. Le transgène sera intégré dans le génome où des séquences compatibles sont présentes. Il a été montré que dans des cellules HEK 293, le transposon TOL2 s'intègre de façon préférentielle dans les régions inter-géniques et introniques ainsi que dans les régions riches en CpG (Meir et al., 2011). L'analyse des sites d'intégration en cours devrait nous permettre de déterminer statistiquement si une telle préférence existe dans notre modèle ou si l'intégration est aléatoire. Mis à part la préférence pour le choix du site d'intégration, la seconde différence majeure entre les deux types de constructions plasmidiques réside dans la taille et la composition de la séquence intégrée. Dans le premier modèle, le plasmide complet s'intègre au génome de la cellule-hôte après s'être linéarisé. L'ouverture du plasmide semble aléatoire compte-tenu des sites d'ouverture identifiés par PCR. En conséquence, il persiste dans le premier modèle cellulaire, toute la partie du plasmide contenant des éléments de régulation bactériens utilisés pour la production des plasmides. En revanche, dans le second modèle, compte-tenu de la présence de séquences TOL de part et d'autre du transgène, la transposase va exciser et intégrer uniquement le transgène sans les éléments bactériens du plasmide. La présence ou l'absence de ces séquences pourrait être une origine de la différence observée dans le profil d'expression des deux transgènes entre les deux modèles cellulaires utilisées dans cette étude (Peterson et al., 1987; Ehrhardt et al., 2003; Chen et al., 2004). En participant au contexte chromatinien local, elles pourraient influencer la dynamique de l'expression génique. Malheureusement, nous ne pouvons que supposer l'impact de ces séquences car nous n'avons pas construit de clones ayant le même site d'intégration mais avec les deux constructions plasmidiques pour faire une étude comparative approfondie.

## Conclusion

---

Les résultats obtenus dans ce travail montrent que les fluctuations de l'expression génique sont présentes et se manifestent à plusieurs échelles dans le modèle utilisé. Dans un premier temps, nous avons montré que le site d'intégration d'un transgène est une composante majeure de la régulation de son niveau d'expression et de la régulation de sa variabilité d'expression. Le modèle a permis d'établir que les transgènes sont transcriptionnellement actifs la majeure partie du temps. Ils peuvent néanmoins subir des périodes d'inactivité transitoires ou très prolongées, donnant lieu à l'établissement d'une bistabilité pour chacun d'eux et contribuant à la macro-hétérogénéité de la population. Ainsi, dans une population clonale, des sous-populations se sont établies au cours du temps, jusqu'à atteindre un état stable. Nous avons montré que la durée de vie des protéines utilisées dans ce système joue un rôle majeur sous-estimé, en introduisant une mémoire si elle est supérieure à la durée de vie moyenne d'une cellule. La réduction de la demi-vie des protéines fluorescentes nous a permis d'améliorer la résolution de notre système expérimental pour détecter des fluctuations plus fines, à l'échelle d'un cycle cellulaire. Ces données ont clairement montré que l'état stable des populations clonales était un état dynamique en équilibre, le phénotype de chaque cellule étant réversible. De plus, la dynamique d'expression peut être extrêmement variée entre cellules appartenant au même clone. Des différences d'expression entre deux transgènes dans une cellule peuvent apparaître si la proximité entre les deux constructions est importante. Ce travail, en accord avec les observations publiées depuis de nombreuses années, permet de se rendre compte que la variabilité phénotypique est la plus constante des variables en biologie. Elle semble universelle et constitue encore un défi pour de nombreuses applications. Compte-tenu des progrès techniques et de l'intérêt pour cette thématique, il est certain que ces prochaines années nous donneront de nouveaux exemples du rôle de la variabilité en biologie. L'utilisation de modèles tels que dans celui de ce travail devrait permettre de mieux la cerner et la maîtriser, afin qu'elle ne constitue plus une imperfection inévitable dans certaines applications biotechnologiques.

## Références bibliographiques

---

1. Acar, M., Becskei, A., and van Oudenaarden, A. (2005). Enhancement of cellular memory by reducing stochastic transitions. *Nature* *435*, 228–232.
2. Allis, C.D. (2007). *Epigenetics* (CSHL Press).
3. Altschuler, S.J., and Wu, L.F. (2010). Cellular Heterogeneity: Do Differences Make a Difference? *Cell* *141*, 559–563.
4. Arkin, A., Ross, J., and McAdams, H.H. (1998). Stochastic Kinetic Analysis of Developmental Pathway Bifurcation in Phage  $\lambda$ -Infected *Escherichia coli* Cells. *Genetics* *149*, 1633–1648.
5. Arriaga, E.A. (2009). Determining biological noise via single cell analysis. *Anal Bioanal Chem* *393*, 73–80.
6. Atkinson, M.R., Savageau, M.A., Myers, J.T., and Ninfa, A.J. (2003). Development of genetic circuitry exhibiting toggle switch or oscillatory behavior in *Escherichia coli*. *Cell* *113*, 597–607.
7. Austin, D.W., Allen, M.S., McCollum, J.M., Dar, R.D., Wilgus, J.R., Sayler, G.S., Samatova, N.F., Cox, C.D., and Simpson, M.L. (2006). Gene network shaping of inherent noise spectra. *Nature* *439*, 608–611.
8. Bai, L., Santangelo, T.J., and Wang, M.D. (2006). Single-molecule analysis of RNA polymerase transcription. *Annu Rev Biophys Biomol Struct* *35*, 343–360.
9. Balaban, N.Q., Merrin, J., Chait, R., Kowalik, L., and Leibler, S. (2004). Bacterial persistence as a phenotypic switch. *Science* *305*, 1622–1625.
10. Bar-Even, A., Paulsson, J., Maheshri, N., Carmi, M., O’Shea, E., Pilpel, Y., and Barkai, N. (2006). Noise in protein expression scales with natural protein abundance. *Nat. Genet.* *38*, 636–643.

11. Barski, A., Cuddapah, S., Cui, K., Roh, T.-Y., Schones, D.E., Wang, Z., Wei, G., Chepelev, I., and Zhao, K. (2007). High-resolution profiling of histone methylations in the human genome. *Cell* *129*, 823–837.
12. Bean, J.M., Siggia, E.D., and Cross, F.R. (2006). Coherence and timing of cell cycle start examined at single-cell resolution. *Mol. Cell* *21*, 3–14.
13. Becskei, A., S  raphin, B., and Serrano, L. (2001). Positive feedback in eukaryotic gene networks: cell differentiation by graded to binary response conversion. *Embo J* *20*, 2528–2535.
14. Becskei, A., and Serrano, L. (2000). Engineering stability in gene networks by autoregulation. *Nature* *405*, 590–593.
15. Berg, H.C., and Purcell, E.M. (1977). Physics of chemoreception. *Biophys J* *20*, 193–219.
16. Berra, E., Benizri, E., Ginouv  s, A., Volmat, V., Roux, D., and Pouyssegur, J. (2003). HIF prolyl-hydroxylase 2 is the key oxygen sensor setting low steady-state levels of HIF-1   in normoxia. *Embo J* *22*, 4082–4090.
17. Bertrand, E., Chartrand, P., Schaefer, M., Shenoy, S.M., Singer, R.H., and Long, R.M. (1998). Localization of ASH1 mRNA particles in living yeast. *Mol. Cell* *2*, 437–445.
18. Bigger, J. (1944). Treatment of staphylococcal infections with penicillin by intermittent sterilisation. *The Lancet* *244*, 497–500.
19. Blake, W.J., M, K.A., Cantor, C.R., and Collins, J.J. (2003). Noise in eukaryotic gene expression. *Nature* *422*, 633–637.
20. Brock, A., Chang, H., and Huang, S. (2009). Non-genetic heterogeneity--a mutation-independent driving force for the somatic evolution of tumours. *Nat. Rev. Genet.* *10*, 336–342.

21. Brooks, A.R., Harkins, R.N., Wang, P., Qian, H.S., Liu, P., and Rubanyi, G.M. (2004). Transcriptional silencing is associated with extensive methylation of the CMV promoter following adenoviral gene delivery to muscle. *J Gene Med* *6*, 395–404.
22. Cahn, F.H., and Fox, M.S. (1968). Fractionation of Transformable Bacteria from Competent Cultures of *Bacillus subtilis* on Renografin Gradients. *J Bacteriol* *95*, 867–875.
23. Cai, L., Friedman, N., and Xie, X.S. (2006). Stochastic protein expression in individual cells at the single molecule level. *Nature* *440*, 358–362.
24. Campbell, K.H.S., McWhir, J., Ritchie, W.A., and Wilmut, I. (1996). Sheep cloned by nuclear transfer from a cultured cell line. *Nature* *380*, 64–66.
25. Canela-Xandri, O., Sagués, F., and Buceta, J. (2010). Interplay between Intrinsic Noise and the Stochasticity of the Cell Cycle in Bacterial Colonies. *Biophys J* *98*, 2459–2468.
26. Carpenter, A.E., Jones, T.R., Lamprecht, M.R., Clarke, C., Kang, I.H., Friman, O., Guertin, D.A., Chang, J.H., Lindquist, R.A., Moffat, J., et al. (2006). CellProfiler: image analysis software for identifying and quantifying cell phenotypes. *Genome Biol.* *7*, R100.
27. Chang, H.H., Hemberg, M., Barahona, M., Ingber, D.E., and Huang, S. (2008). Transcriptome-wide noise controls lineage choice in mammalian progenitor cells. *Nature* *453*, 544–547.
28. Chelly, J., Concordet, J.P., Kaplan, J.C., and Kahn, A. (1989). Illegitimate Transcription: Transcription of Any Gene in Any Cell Type. *Proc Natl Acad Sci U S A* *86*, 2617–2621.

29. Chelly, J., Kaplan, J.-C., Maire, P., Gautron, S., and Kahn, A. (1988). Transcription of the dystrophin gene in human muscle and non-muscle tissues. *Proc Natl Acad Sci U S A* *333*, 858–860.
30. Chen, C.Y., and Shyu, A.B. (1995). AU-rich elements: characterization and importance in mRNA degradation. *Trends Biochem. Sci.* *20*, 465–470.
31. Chen, Z.Y., He, C.Y., Meuse, L., and Kay, M.A. (2004). Silencing of episomal transgene expression by plasmid bacterial DNA elements in vivo. *Gene Ther.* *11*, 856–864.
32. Choi, P.J., Cai, L., Frieda, K., and Xie, X.S. (2008). A stochastic single-molecule event triggers phenotype switching of a bacterial cell. *Science* *322*, 442.
33. Choi, P.J., Xie, X.S., and Shakhnovich, E.I. (2010). Stochastic switching in gene networks can occur by a single-molecule event or many molecular steps. *J. Mol. Biol.* *396*, 230–244.
34. Colman-Lerner, A., Gordon, A., Serra, E., Chin, T., Resnekov, O., Endy, D., Pesce, C.G., and Brent, R. (2005). Regulated cell-to-cell variation in a cell-fate decision system. *Nature* *437*, 699–706.
35. Condreay, J.P., Witherspoon, S.M., Clay, W.C., and Kost, T.A. (1999). Transient and stable gene expression in mammalian cells transduced with a recombinant baculovirus vector. *Proc Natl Acad Sci U S A* *96*, 127–132.
36. Cookson, N.A., Cookson, S.W., Tsimring, L.S., and Hasty, J. (2010). Cell cycle-dependent variations in protein concentration. *Nucleic Acids Res.* *38*, 2676–2681.
37. Coulon, A., Gandrillon, O., and Beslon, G. (2010). On the spontaneous stochastic dynamics of a single gene: complexity of the molecular interplay at the promoter. *BMC Syst Biol* *4*, 2.
38. Cox, R.S., 3rd, Dunlop, M.J., and Elowitz, M.B. (2010). A synthetic three-color scaffold for monitoring genetic regulation and noise. *J Biol Eng* *4*, 10.



39. Crick, F. (1958). On protein synthesis. *Symp. Soc. Exp. Biol.* *12*, 138–163.
40. Crick, F. (1970). Central dogma of molecular biology. *Nature* *227*, 561–563.
41. Culhane, A.C., Thioulouse, J., Perrière, G., and Higgins, D.G. MADE4: an R package for multivariate analysis of gene expression data. *Bioinformatics* *21*, 2789 – 2790.
42. Darzynkiewicz, Z., Crissman, H., Traganos, F., and Steinkamp, J. (1982). Cell heterogeneity during the cell cycle. *J. Cell. Physiol.* *113*, 465–474.
43. Deal, R.B., and Henikoff, S. (2010). Capturing the dynamic epigenome. *Genome Biol.* *11*, 218.
44. Delbrück, M. (1940). The growth of bacteriophage and lysis of the host. *J Gen Physiol* *23*, 643–660.
45. Depping, R., Steinhoff, A., Schindler, S.G., Friedrich, B., Fagerlund, R., Metzen, E., Hartmann, E., and Köhler, M. (2008). Nuclear translocation of hypoxia-inducible factors (HIFs): involvement of the classical importin alpha/beta pathway. *Biochim. Biophys. Acta* *1783*, 394–404.
46. Diercks, A., Kostner, H., and Ozinsky, A. (2009). Resolving Cell Population Heterogeneity: Real-Time PCR for Simultaneous Multiplexed Gene Detection in Multiple Single-Cell Samples. *PLoS ONE* *4*, e6326.
47. Dobrzyński, M., and Bruggeman, F.J. (2009). Elongation dynamics shape bursty transcription and translation. *Proc Natl Acad Sci U S A* *106*, 2583 –2588.
48. Dublanche, Y., Michalodimitrakis, K., Kümmerer, N., Foglierini, M., and Serrano, L. (2006). Noise in transcription negative feedback loops: simulation and experimental analysis. *Mol Syst Biol* *2*, 41.
49. Dubnau, D., and Losick, R. (2006). Bistability in bacteria. *Mol. Microbiol.* *61*, 564–572.

50. Ehrhardt, A., Peng, P.D., Xu, H., Meuse, L., and Kay, M.A. (2003). Optimization of cis-acting elements for gene expression from nonviral vectors in vivo. *Hum. Gene Ther.* *14*, 215–225.
51. Ellis, B., Haaland, P., Hahne, F., Meur, N.L., and Gopalakrishnan, N. (2010). flowCore: flowCore: Basic structures for flow cytometry data.
52. Elowitz, M.B., and Leibler, S. (2000). A synthetic oscillatory network of transcriptional regulators. *Nature* *403*, 335–338.
53. Elowitz, M.B., Levine, A.J., Siggia, E.D., and Swain, P.S. (2002). Stochastic Gene Expression in a Single Cell. *Science* *297*, 1183–1186.
54. Enver, T., and Greaves, M. (1998). Loops, Lineage, and Leukemia. *Cell* *94*, 9–12.
55. Enver, T., Heyworth, C.M., and Dexter, T.M. (1998). Do Stem Cells Play Dice? *Blood* *92*, 348–351.
56. Fallaux, F.J., Kranenburg, O., Cramer, S.J., Houweling, A., Van Ormondt, H., Hoeben, R.C., and Van Der Eb, A.J. (1996). Characterization of 911: a new helper cell line for the titration and propagation of early region 1-deleted adenoviral vectors. *Hum. Gene Ther.* *7*, 215–222.
57. Fandrey, J., Gorr, T.A., and Gassmann, M. (2006). Regulating Cellular Oxygen Sensing by Hydroxylation. *Cardiovasc Res* *71*, 642–651.
58. Fialcowitz, E.J., Brewer, B.Y., Keenan, B.P., and Wilson, G.M. (2005). A Hairpin-Like Structure Within an AU-Rich mRNA-Destabilizing Element Regulates Trans-Factor Binding Selectivity and mRNA Decay Kinetics. *J. Biol. Chem.* *280*, 22406–22417.
59. Fonknechten, N., Chelly, J., Lepercq, J., Kahn, A., Kaplan, J.C., Kitzis, A., and Chomel, J.C. (1992). CFTR illegitimate transcription in lymphoid cells: quantification and applications to the investigation of pathological transcripts. *Hum. Genet.* *88*, 508–512.

60. Fraser, D., and Kaern, M. (2009). A chance at survival: gene expression noise and phenotypic diversification strategies. *Mol. Microbiol.* *71*, 1333–1340.
61. Fung, E., Wong, W.W., Suen, J.K., Bulter, T., Lee, S., and Liao, J.C. (2005). A synthetic gene-metabolic oscillator. *Nature* *435*, 118–122.
62. Gardner, T.S., Cantor, C.R., and Collins, J.J. (2000). Construction of a genetic toggle switch in *Escherichia coli*. *Nature* *403*, 339–342.
63. Gaszner, M., and Felsenfeld, G. (2006). Insulators: exploiting transcriptional and epigenetic mechanisms. *Nat. Rev. Genet.* *7*, 703–713.
64. Geva-Zatorsky, N., Rosenfeld, N., Itzkovitz, S., Milo, R., Sigal, A., Dekel, E., Yarnitzky, T., Liron, Y., Polak, P., Lahav, G., et al. (2006). Oscillations and variability in the p53 system. *Mol Syst Biol* *2*, 2006.0033.
65. Ghaemmaghami, S., Huh, W.-K., Bower, K., Howson, R.W., Belle, A., Dephoure, N., O’Shea, E.K., and Weissman, J.S. (2003). Global analysis of protein expression in yeast. *Nature* *425*, 737–741.
66. Gillespie, D.T. (1977). Exact stochastic simulation of coupled chemical reactions. *J. Phys. Chem.* *81*, 2340–2361.
67. Golding, I., and Cox, E.C. (2004). RNA dynamics in live *Escherichia coli* cells. *Proc Natl Acad Sci U S A* *101*, 11310–11315.
68. Golding, I., Paulsson, J., Zawilski, S.M., and Cox, E.C. (2005). Real-time kinetics of gene activity in individual bacteria. *Cell* *123*, 1025–1036.
69. Gong, L., Constantine, W., and Chen, and Y.A. (2011). msProcess: Protein Mass Spectra Processing.
70. Grassi, G., Maccaroni, P., Meyer, R., Kaiser, H., D’Ambrosio, E., Pascale, E., Grassi, M., Kuhn, A., Di Nardo, P., Kandolf, R., et al. (2003). Inhibitors of DNA methylation and histone deacetylation activate cytomegalovirus promoter-

controlled reporter gene expression in human glioblastoma cell line U87. *Carcinogenesis* 24, 1625–1635.

71. Greive, S.J., and von Hippel, P.H. (2005). Thinking quantitatively about transcriptional regulation. *Nat. Rev. Mol. Cell Biol.* 6, 221–232.
72. Grigoryev, S.A., Arya, G., Correll, S., Woodcock, C.L., and Schlick, T. (2009). Evidence for heteromorphic chromatin fibers from analysis of nucleosome interactions. *Proc Natl Acad Sci U S A* 106, 13317–13322.
73. Guido, N.J., Wang, X., Adalsteinsson, D., McMillen, D., Hasty, J., Cantor, C.R., Elston, T.C., and Collins, J.J. (2006). A bottom-up approach to gene regulation. *Nature* 439, 856–860.
74. Guptasarma, P. (1995). Does replication-induced transcription regulate synthesis of the myriad low copy number proteins of *Escherichia coli*? *Bioessays* 17, 987–997.
75. Hadden, C., and Nester, E.W. (1968). Purification of Competent Cells in the *Bacillus subtilis* Transformation System. *J Bacteriol* 95, 876–885.
76. Hahne, F., Huber, W., Ruschhaupt, M., and Toedling, J. (2008). prada: Data analysis for cell-based functional assays.
77. Hall, T. (1999). BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucleic Acids Symposium Series* 41, 95–98.
78. Hasty, J., Pradines, J., Dolnik, M., and Collins, J.J. (2000). Noise-based switches and amplifiers for gene expression. *Proc Natl Acad Sci U S A* 97, 2075–2080.
79. Heim, R., Cubitt, A.B., and Tsien, R.Y. (1995). Improved green fluorescence. *Nature* 373, 663–664.

80. Herbert, K.M., La Porta, A., Wong, B.J., Mooney, R.A., Neuman, K.C., Landick, R., and Block, S.M. (2006). Sequence-Resolved Detection of Pausing by Single RNA Polymerase Molecules. *Cell* *125*, 1083–1094.
81. Hilfinger, A., and Paulsson, J. (2011). Separating intrinsic from extrinsic fluctuations in dynamic biological systems. *Proc Natl Acad Sci U S A* *108*, 12167–12172.
82. Holland, M.J. (2002). Transcript abundance in yeast varies over six orders of magnitude. *J. Biol. Chem.* *277*, 14363–14366.
83. Hooshangi, S., Thiberge, S., and Weiss, R. (2005). Ultrasensitivity and noise propagation in a synthetic transcriptional cascade. *Proc Natl Acad Sci U S A* *102*, 3581–3586.
84. Hooshangi, S., and Weiss, R. (2006). The effect of negative feedback on noise propagation in transcriptional gene networks. *Chaos* *16*, 026108.
85. Huang, S. (2009). Non-genetic heterogeneity of cells in development: more than just noise. *Development* *136*, 3853–3862.
86. Huh, D., and Paulsson, J. (2011a). Non-genetic heterogeneity from stochastic partitioning at cell division. *Nat. Genet.* *43*, 95–100.
87. Huh, D., and Paulsson, J. (2011b). Random partitioning of molecules at cell division. *Proc Natl Acad Sci U S A* *108*, 15004–15009.
88. Hume, D.A. (2000). Probability in transcriptional regulation and its implications for leukocyte differentiation and inducible gene expression. *Blood* *96*, 2323–2328.
89. Husson, F., Josse, J., Le, S., and Mazet, J. (2011). FactoMineR: Multivariate Exploratory Data Analysis and Data Mining with R.

90. Isaacs, F.J., Hasty, J., Cantor, C.R., and Collins, J.J. (2003). Prediction and measurement of an autoregulatory genetic module. *Proc Natl Acad Sci U S A* *100*, 7714–7719.
91. Jacob, F., and Monod, J. (1961). Genetic regulatory mechanisms in the synthesis of proteins. *J. Mol. Biol.* *3*, 318–356.
92. Jewell, U.R., Kvietikova, I., Scheid, A., Bauer, C., Wenger, R.H., and Gassmann, M. (2001). Induction of HIF-1alpha in response to hypoxia is instantaneous. *Faseb J.* *15*, 1312–1314.
93. Joh, R.I., and Weitz, J.S. (2011). To lyse or not to lyse: transient-mediated stochastic fate determination in cells infected by bacteriophages. *PLoS Comput. Biol.* *7*, e1002006.
94. Johnston, R.J., and Desplan, C. (2010). A Penetrating Look at Stochasticity in Development. *Cell* *140*, 610–612.
95. Jones, T.R., Carpenter, A.E., Lamprecht, M.R., Moffat, J., Silver, S.J., Grenier, J.K., Castoreno, A.B., Eggert, U.S., Root, D.E., Golland, P., et al. (2009). Scoring diverse cellular morphologies in image-based screens with iterative feedback and machine learning. *Proc Natl Acad Sci U S A* *106*, 1826–1831.
96. Jones, T.R., Kang, I.H., Wheeler, D.B., Lindquist, R.A., Papallo, A., Sabatini, D.M., Golland, P., and Carpenter, A.E. (2008). CellProfiler Analyst: data exploration and analysis software for complex image-based screens. *BMC Bioinformatics* *9*, 482.
97. Kaern, M., Elston, T.C., Blake, W.J., and Collins, J.J. (2005). Stochasticity in gene expression: from theories to phenotypes. *Nat Rev Genet* *6*, 451–464.
98. Kawakami, K. (2007). Tol2: a versatile gene transfer vector in vertebrates. *Genome Biol.* *8 Suppl 1*, S7.
99. Kimura, H., and Cook, P.R. (2001). Kinetics of Core Histones in Living Human Cells. *J Cell Biol* *153*, 1341–1354.

- 100.Ko, M.S., Nakauchi, H., and Takahashi, N. (1990). The dose dependence of glucocorticoid-inducible gene expression results from changes in the number of transcriptionally active templates. *Embo J* *9*, 2835–2842.
- 101.Koga, A. (2004). Transposition mechanisms and biotechnology applications of the medaka fish Tol2 transposable element. *Adv. Biophys.* *38*, 161–180.
- 102.Koga, A., and Hori, H. (2001). The Tol2 transposable element of the medaka fish: an active DNA-based element naturally occurring in a vertebrate genome. *Genes Genet. Syst.* *76*, 1–8.
- 103.Komsta, L., and Novomestky, F. (2011). moments: Moments, cumulants, skewness, kurtosis and related tests.
- 104.Kupiec, J.J. (1996). A chance-selection model for cell differentiation. *Cell Death Differentiation* *3*, 385–390.
- 105.Kupiec, J.J. (1997). A Darwinian theory for the origin of cellular differentiation. *Mol. Gen. Genet.* *255*, 201–208.
- 106.Kupiec, J.-J. (1983). A probabilist theory for cell differentiation, embryonic mortality and DNA C-value paradox. *Spec. Sci. Tech* *6*, 471–478.
- 107.Lahav, G., Rosenfeld, N., Sigal, A., Geva-Zatorsky, N., Levine, A.J., Elowitz, M.B., and Alon, U. (2004). Dynamics of the p53-Mdm2 feedback loop in individual cells. *Nat. Genet.* *36*, 147–150.
- 108.Landick, R. (2006). The regulatory roles and mechanism of transcriptional pausing. *Biochem. Soc. Trans.* *34*, 1062–1066.
- 109.Lever, M.A., Th'ng, J.P.H., Sun, X., and Hendzel, M.J. (2000). Rapid exchange of histone H1.1 on chromatin in living human cells. *Nature* *408*, 873–876.
- 110.Li, G., Levitus, M., Bustamante, C., and Widom, J. (2005). Rapid spontaneous accessibility of nucleosomal DNA. *Nat. Struct. Mol. Biol.* *12*, 46–53.

- 111.Lieb, M. (1953). The establishment of lysogenicity in *Escherichia coli*. *J. Bacteriol.* *65*, 642–651.
- 112.Longo, D., and Hasty, J. (2006). Dynamics of single-cell gene expression. *Mol. Syst. Biol.* *2*, 64.
- 113.Löser, P., Jennings, G.S., Strauss, M., and Sandig, V. (1998). Reactivation of the previously silenced cytomegalovirus major immediate-early promoter in the mouse liver: involvement of NFkappaB. *J. Virol.* *72*, 180–190.
- 114.Losick, R., and Desplan, C. (2008). Stochasticity and cell fate. *Science* *320*, 65–68.
- 115.Maamar, H., and Dubnau, D. (2005). Bistability in the *Bacillus subtilis* K-state (competence) system requires a positive feedback loop. *Mol. Microbiol.* *56*, 615–624.
- 116.Maamar, H., Raj, A., and Dubnau, D. (2007). Noise in Gene Expression Determines Cell Fate in *Bacillus subtilis*. *Science* *317*, 526–529.
- 117.McAdams, H.H., and Arkin, A. (1997). Stochastic mechanisms in gene expression. *Proc Natl Acad Sci U S A* *94*, 814–819.
- 118.McNally, J.G., Müller, W.G., Walker, D., Wolford, R., and Hager, G.L. (2000). The glucocorticoid receptor: rapid exchange with regulatory sites in living cells. *Science* *287*, 1262–1265.
- 119.Mehta, A.K., Majumdar, S.S., Alam, P., Gulati, N., and Brahmachari, V. (2009). Epigenetic regulation of cytomegalovirus major immediate-early promoter activity in transgenic mice. *Gene* *428*, 20–24.
- 120.Meir, Y.-J.J., Weirauch, M.T., Yang, H.-S., Chung, P.-C., Yu, R.K., and Wu, S.C.-Y. (2011). Genome-wide target profiling of piggyBac and Tol2 in HEK 293: pros and cons for gene discovery and gene therapy. *BMC Biotechnology* *11*, 28.



121. Misteli, T., Gunjan, A., Hock, R., Bustin, M., and Brown, D.T. (2000). Dynamic binding of histone H1 to chromatin in living cells. *Nature* *408*, 877–881.
122. Nan, X., Campoy, F.J., and Bird, A. (1997). MeCP2 Is a Transcriptional Repressor with Abundant Binding Sites in Genomic Chromatin. *Cell* *88*, 471–481.
123. Neildez-Nguyen, T.M.A., Parisot, A., Vignal, C., Rameau, P., Stockholm, D., Picot, J., Allo, V., Le Bec, C., Laplace, C., and Paldi, A. (2008). Epigenetic gene expression noise and phenotypic diversification of clonal cell populations. *Differentiation* *76*, 33–40.
124. Neildez-Nguyen, T.M.A., Parisot, A., Vignal, C., Rameau, P., Stockholm, D., Picot, J., Allo, V., Le Bec, C., Laplace, C., and Paldi, A. (2008). Epigenetic gene expression noise and phenotypic diversification of clonal cell populations. *Differentiation* *76*, 33–40.
125. Nelson, D.E., Ihekweba, A.E.C., Elliott, M., Johnson, J.R., Gibney, C.A., Foreman, B.E., Nelson, G., See, V., Horton, C.A., Spiller, D.G., et al. (2004). Oscillations in NF-kappaB signaling control the dynamics of gene expression. *Science* *306*, 704–708.
126. Nester, E.W., and Stocker, B.A.D. (1963). Biosynthetic latency in early stages of deoxyribonucleic acid transformation in *Bacillus subtilis*. *J Bacteriol* *86*, 785–796.
127. Newman, J.R.S., Ghaemmighami, S., Ihmels, J., Breslow, D.K., Noble, M., DeRisi, J.L., and Weissman, J.S. (2006). Single-cell proteomic analysis of *S. cerevisiae* reveals the architecture of biological noise. *Nature* *441*, 840–846.
128. Niepel, M., Spencer, S.L., and Sorger, P.K. (2009). Non-genetic cell-to-cell variability and the consequences for pharmacology. *Curr Opin Chem Biol* *13*, 556–561.
129. Novick, A., and Weiner, M. (1957). Enzyme induction as an all-or-none phenomenon. *Proc Natl Acad Sci U S A* *43*, 553–566.

130. Oksanen, J., Blanchet, F.G., Kindt, R., Legendre, P., Minchin, P.R., O'Hara, R.B., Simpson, G.L., Solymos, P., Stevens, M.H.H., and Wagner, H. (2011). *vegan*: Community Ecology Package.
131. Ozbudak, E.M., Thattai, M., Kurtser, I., Grossman, A.D., and van Oudenaarden, A. (2002). Regulation of noise in the expression of a single gene. *Nat. Genet.* *31*, 69–73.
132. Ozbudak, E.M., Thattai, M., Lim, H.N., Shraiman, B.I., and van Oudenaarden, A. (2004). Multistability in the lactose utilization network of *Escherichia coli*. *Nature* *427*, 737–740.
133. Paldi, A. (2003). Stochastic gene expression during cell differentiation: order from disorder? *Cell. Mol. Life Sci.* *60*, 1775–1778.
134. Paulsson, J. (2005). Models of stochastic gene expression. *Phys. Life Rev.* *2*, 157–175.
135. Pedraza, J.M., and van Oudenaarden, A. (2005). Noise propagation in gene networks. *Science* *307*, 1965–1969.
136. Peterson, D.O., Beifuss, K.K., and Morley, K.L. (1987). Context-dependent gene expression: cis-acting negative effects of specific procaryotic plasmid sequences on eucaryotic genes. *Mol Cell Biol* *7*, 1563–1567.
137. Phair, R.D., and Misteli, T. (2000). High mobility of proteins in the mammalian cell nucleus. *Nature* *404*, 604–609.
138. Phair, R.D., Scaffidi, P., Elbi, C., Vecerová, J., Dey, A., Ozato, K., Brown, D.T., Hager, G., Bustin, M., and Misteli, T. (2004). Global nature of dynamic protein-chromatin interactions in vivo: three-dimensional genome scanning and dynamic interaction networks of chromatin proteins. *Mol. Cell. Biol.* *24*, 6393–6402.
139. Pilbrough, W., Munro, T.P., and Gray, P. (2009). Intracloonal protein expression heterogeneity in recombinant CHO cells. *PLoS ONE* *4*, e8432.

140. Prösch, S., Stein, J., Staak, K., Liebenthal, C., Volk, H.D., and Krüger, D.H. (1996). Inactivation of the very strong HCMV immediate early promoter by DNA CpG methylation in vitro. *Biol. Chem. Hoppe-Seyler* *377*, 195–201.
141. Qin, J.Y., Zhang, L., Clift, K.L., Hular, I., Xiang, A.P., Ren, B.-Z., and Lahn, B.T. (2010). Systematic Comparison of Constitutive Promoters and the Doxycycline-Inducible Promoter. *PLoS ONE* *5*, e10611.
142. Radhakrishnan, P., Basma, H., Klinkebiel, D., Christman, J., and Cheng, P.-W. (2008). Cell type-specific activation of the cytomegalovirus promoter by dimethylsulfoxide and 5-aza-2'-deoxycytidine. *Int. J. Biochem. Cell Biol.* *40*, 1944–1955.
143. Raj, A., van den Bogaard, P., Rifkin, S.A., van Oudenaarden, A., and Tyagi, S. (2008). Imaging individual mRNA molecules using multiple singly labeled probes. *Nat Methods* *5*, 877–879.
144. Raj, A., and van Oudenaarden, A. (2008). Nature, Nurture, or Chance: Stochastic Gene Expression and Its Consequences. *Cell* *135*, 216–226.
145. Raj, A., Peskin, C.S., Tranchina, D., Vargas, D.Y., and Tyagi, S. (2006). Stochastic mRNA synthesis in mammalian cells. *PLoS Biol.* *4*, e309.
146. Rajala, T., Häkkinen, A., Healy, S., Yli-Harja, O., and Ribeiro, A.S. (2010). Effects of transcriptional pausing on gene expression dynamics. *PLoS Comput. Biol.* *6*, e1000704.
147. Rao, B., Shibata, Y., Strahl, B.D., and Lieb, J.D. (2005). Dimethylation of Histone H3 at Lysine 36 Demarcates Regulatory and Nonregulatory Chromatin Genome-Wide. *Mol Cell Biol* *25*, 9447–9459.
148. Raser, J.M., and O'Shea, E.K. (2004). Control of Stochasticity in Eukaryotic Gene Expression. *Science* *304*, 1811–1814.

- 149.Razooky, B.S., and Weinberger, L.S. (2011). Mapping the architecture of the HIV-1 Tat circuit: A decision-making circuit that lacks bistability and exploits stochastic noise. *Methods* *53*, 68–77.
- 150.Rechsteiner, M. (1988). Regulation of enzyme levels by proteolysis: the role of pest regions. *Adv. Enzyme Regul.* *27*, 135–151.
- 151.Rechsteiner, M. (1990). PEST sequences are signals for rapid intracellular proteolysis. *Semin. Cell Biol.* *1*, 433–440.
- 152.Rechsteiner, M., and Rogers, S.W. (1996). PEST sequences and regulation by proteolysis. *Trends Biochem. Sci.* *21*, 267–271.
- 153.Rizzo, M.A., Davidson, M.W., and Piston, D.W. (2009). Fluorescent Protein Tracking and Detection: Fluorescent Protein Structure and Color Variants. *Cold Spring Harb Protoc* *2009*, pdb.top63.
- 154.Rosenfeld, N., Elowitz, M.B., and Alon, U. (2002). Negative autoregulation speeds the response times of transcription networks. *J. Mol. Biol.* *323*, 785–793.
- 155.Rosenfeld, N., Young, J.W., Alon, U., Swain, P.S., and Elowitz, M.B. (2005). Gene regulation at the single-cell level. *Science* *307*, 1962–1965.
- 156.Ross, I.L., Browne, C.M., and Hume, D.A. (1994). Transcription of individual genes in eukaryotic cells occurs randomly and infrequently. *Immunol. Cell Biol.* *72*, 177–185.
- 157.Savageau, M.A. (1974a). Comparison of classical and autogenous systems of regulation in inducible operons. *Nature* *252*, 546–549.
- 158.Savageau, M.A. (1974b). Optimal design of feedback control by inhibition. *J. Mol. Evol.* *4*, 139–156.

- 159.Saxonov, S., Berg, P., and Brutlag, D.L. (2006). A genome-wide analysis of CpG dinucleotides in the human genome distinguishes two distinct classes of promoters. *Proc Natl Acad Sci U S A* *103*, 1412–1417.
- 160.Schotta, G., Lachner, M., Sarma, K., Ebert, A., Sengupta, R., Reuter, G., Reinberg, D., and Jenuwein, T. (2004). A silencing pathway to induce H3-K9 and H4-K20 trimethylation at constitutive heterochromatin. *Genes Dev* *18*, 1251–1262.
- 161.Schug, J. (2008). Using TESS to predict transcription factor binding sites in DNA sequence. *Curr Protoc Bioinformatics Chapter 2*, Unit 2.6.
- 162.Schwarzfischer, M., Marr, C., Krumsiek, J., Hoppe, P.S., Schroeder, T., and Theis, F.J. (2011). Efficient fluorescence image normalization for time lapse movies. (Heidelberg),.
- 163.Sharma, S.V., Lee, D.Y., Li, B., Quinlan, M.P., Takahashi, F., Maheswaran, S., McDermott, U., Azizian, N., Zou, L., Fischbach, M.A., et al. (2010). A chromatin-mediated reversible drug tolerant state in cancer cell subpopulations. *Cell* *141*, 69–80.
- 164.Shimomura, O., Johnson, F.H., and Saiga, Y. (1962). Extraction, purification and properties of aequorin, a bioluminescent protein from the luminous hydromedusan, *Aequorea*. *J Cell Comp Physiol* *59*, 223–239.
- 165.Shumway, R.H., and Stoffer, D.S. (2006). *Time Series Analysis and Its Applications: With R Examples* (Springer).
- 166.Sigal, A., Milo, R., Cohen, A., Geva-Zatorsky, N., Klein, Y., Alaluf, I., Swerdlin, N., Perzov, N., Danon, T., Liron, Y., et al. (2006). Dynamic proteomics in individual human cells uncovers widespread cell-cycle dependence of nuclear proteins. *Nat. Genet.* *3*, 525–531.

- 167.Silander, O.K., Nikolic, N., Zaslaver, A., Bren, A., Kikoin, I., Alon, U., and Ackermann, M. (2012). A genome-wide analysis of promoter-mediated phenotypic noise in *Escherichia coli*. *PLoS Genet.* *8*, e1002443.
- 168.Sinclair, J.H., Baillie, J., Bryant, L.A., Taylor-Wiedeman, J.A., and Sissons, J.G. (1992). Repression of human cytomegalovirus major immediate early gene expression in a monocytic cell line. *J. Gen. Virol.* *73 (Pt 2)*, 433–435.
- 169.Singh, A., Razooky, B., Cox, C.D., Simpson, M.L., and Weinberger, L.S. (2010). Transcriptional Bursting from the HIV-1 Promoter Is a Significant Source of Stochastic Noise in HIV-1 Gene Expression. *Biophys J* *98*, L32–L34.
- 170.Singh, A., and Weinberger, L.S. (2009). Noise in viral gene expression as a molecular switch for viral latency. *Curr Opin Microbiol* *12*, 460–466.
- 171.Skommer, J., Raychaudhuri, S., and Wlodkovic, D. (2011). Timing is everything: stochastic origins of cell-to-cell variability in cancer cell death. *Front. Biosci.* *16*, 307–314.
- 172.Smolen, P., Baxter, D.A., and Byrne, J.H. (1999). Effects of macromolecular transport and stochastic fluctuations on dynamics of genetic regulatory systems. *Am. J. Physiol.* *277*, C777–790.
- 173.Snijder, B., Sacher, R., R|[auml]|m|[ouml]|, P., Damm, E.-M., Liberali, P., and Pelkmans, L. (2009). Population context determines cell-to-cell variability in endocytosis and virus infection. *Nature* *461*, 520–523.
- 174.Spencer, S.L., Gaudet, S., Albeck, J.G., Burke, J.M., and Sorger, P.K. (2009). Non-genetic origins of cell-to-cell variability in TRAIL-induced apoptosis. *Nature* *459*, 428–432.
- 175.Spudich, J.L., and Koshland, D.E. (1976). Non-genetic individuality: chance in the single cell. *Nature* *262*, 467–471.

176. Steger, D.J., Lefterova, M.I., Ying, L., Stonestrom, A.J., Schupp, M., Zhuo, D., Vakoc, A.L., Kim, J.-E., Chen, J., Lazar, M.A., et al. (2008). DOT1L/KMT4 recruitment and H3K79 methylation are ubiquitously coupled with gene transcription in mammalian cells. *Mol. Cell. Biol.* **28**, 2825–2839.
177. Stenoien, D.L., Patel, K., Mancini, M.G., Dutertre, M., Smith, C.L., O'Malley, B.W., and Mancini, M.A. (2001). FRAP reveals that mobility of oestrogen receptor-[alpha] is ligand- and proteasome-dependent. *Nat Cell Biol* **3**, 15–23.
178. Stern, S., Dror, T., Stolovicki, E., Brenner, N., and Braun, E. (2007). Genome-wide transcriptional plasticity underlies cellular adaptation to novel challenge. *Mol Syst Biol* **3**, 106.
179. Stockholm, D., Benchaouir, R., Picot, J., Rameau, P., Neildez, T.M.A., Landini, G., Laplace-Builhé, C., and Paldi, A. (2007). The Origin of Phenotypic Heterogeneity in a Clonal Cell Population In Vitro. *PLoS ONE* **2**.
180. Stockholm, D., Edom-Vovard, F., Coutant, S., Sanatine, P., Yamagata, Y., Corre, G., Le Guillou, L., Neildez-Nguyen, T.M.A., and Paldi, A. (2010). Bistable Cell Fate Specification as a Result of Stochastic Fluctuations and Collective Spatial Cell Behaviour. *PLoS ONE* **5**, e14441.
181. St-Pierre, F., and Endy, D. (2008). Determination of cell fate selection during phage lambda infection. *Proc Natl Acad Sci U S A* **105**, 20705–20710.
182. Suel, G.M., Kulkarni, R.P., Dworkin, J., Garcia-Ojalvo, J., and Elowitz, M.B. (2007). Tunability and noise dependence in differentiation dynamics. *Science* **315**, 1716–1719.
183. Suter, D.M., Molina, N., Gatfield, D., Schneider, K., Schibler, U., and Naef, F. (2011). Mammalian Genes Are Transcribed with Widely Different Bursting Kinetics. *Science* **332**, 472–474.

184. Swain, P.S., Elowitz, M.B., and Siggia, E.D. (2002). Intrinsic and extrinsic contributions to stochasticity in gene expression. *Proc Natl Acad Sci U S A* *99*, 12795–12800.
185. Taniguchi, Y., Choi, P.J., Li, G.-W., Chen, H., Babu, M., Hearn, J., Emili, A., and Xie, X.S. (2010). Quantifying E. coli Proteome and Transcriptome with Single-Molecule Sensitivity in Single Cells. *Science* *329*, 533–538.
186. Team, R.D.C. (2011). R: A Language and Environment for Statistical Computing (Vienna, Austria).
187. Thattai, M., and van Oudenaarden, A. (2001). Intrinsic noise in gene regulatory networks. *Proc Natl Acad Sci U S A* *98*, 8614–8619.
188. Thattai, M., and van Oudenaarden, A. (2004). Stochastic gene expression in fluctuating environments. *Genetics* *167*, 523–530.
189. Thioulouse, J., Chessel, D., Dole'dec, S., and Olivier, J.-M. (1997). ADE-4: a multivariate analysis and graphical display software. *Statistics and Computing* *7*, 75–83.
190. Thioulouse, J., and Dray, S. (2007). Interactive Multivariate Data Analysis in R with the ade4 and ade4TkGUI Packages. *Journal of Statistical Software* *22*, 1–14.
191. Tian, T., and Burrage, K. (2004). Bistability and switching in the lysis/lysogeny genetic regulatory network of bacteriophage lambda. *J. Theor. Biol.* *227*, 229–237.
192. Trojer, P., Li, G., Sims, R.J., 3rd, Vaquero, A., Kalakonda, N., Boccuni, P., Lee, D., Erdjument-Bromage, H., Tempst, P., Nimer, S.D., et al. (2007). L3MBTL1, a histone-methylation-dependent chromatin lock. *Cell* *129*, 915–928.
193. Tsukada, Y., Fang, J., Erdjument-Bromage, H., Warren, M.E., Borchers, C.H., Tempst, P., and Zhang, Y. (2005). Histone demethylation by a family of JmjC domain-containing proteins. *Nature* *439*, 811–816.



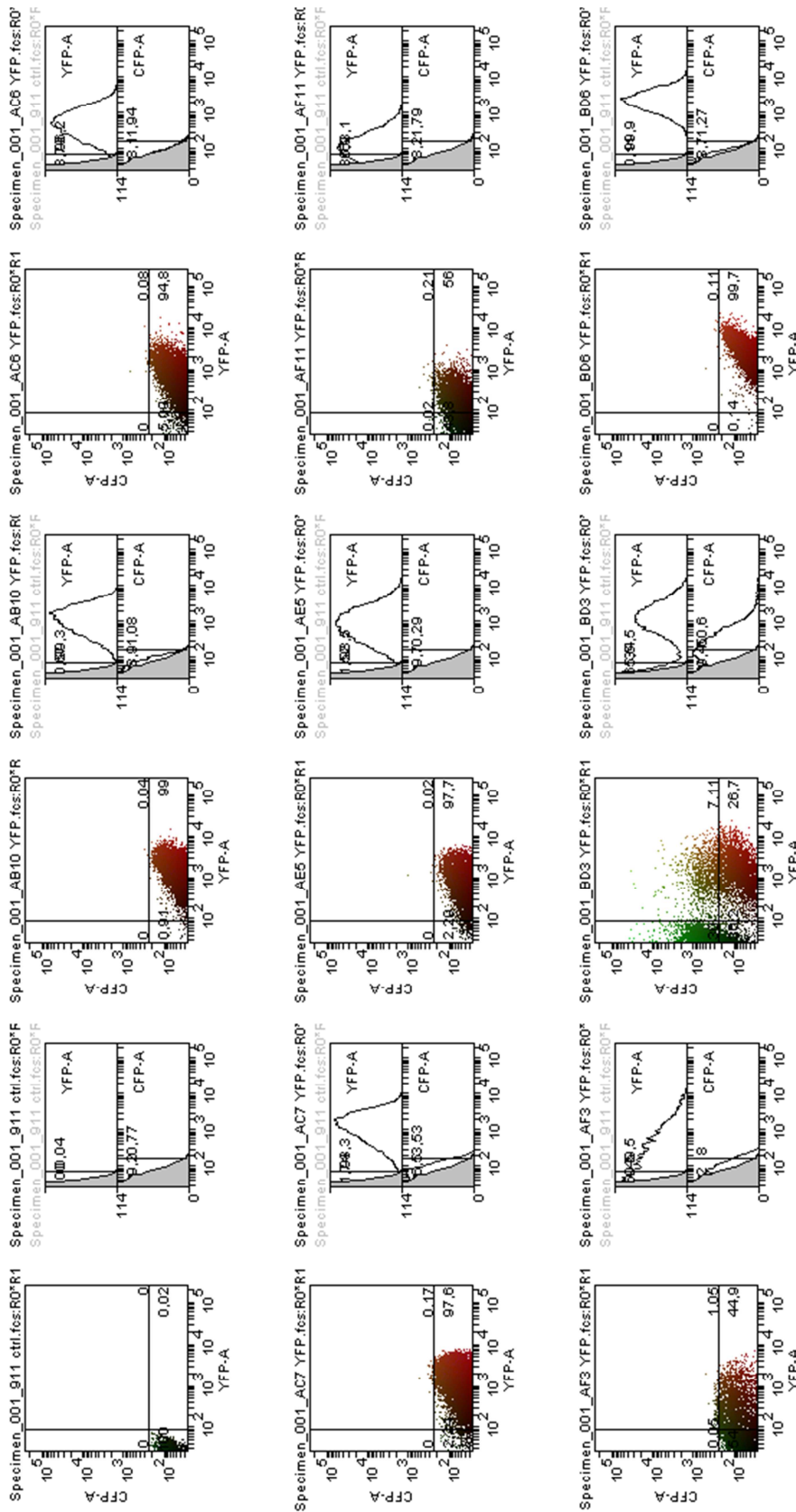
194. Tsuru, S., Ichinose, J., Kashiwagi, A., Ying, B.-W., Kaneko, K., and Yomo, T. (2009). Noisy cell growth rate leads to fluctuating protein concentration in bacteria. *Phys Biol* *6*, 036015.
195. Urbanek, S. (2011). png: Read and write PNG images.
196. Uren, A.G., Mikkers, H., Kool, J., van der Weyden, L., Lund, A.H., Wilson, C.H., Rance, R., Jonkers, J., van Lohuizen, M., Berns, A., et al. (2009). A high-throughput splinkerette-PCR method for the isolation and sequencing of retroviral insertion sites. *Nat Protoc* *4*, 789–798.
197. Velculescu, V.E., Zhang, L., Zhou, W., Vogelstein, J., Basrai, M.A., Bassett Jr., D.E., Hieter, P., Vogelstein, B., and Kinzler, K.W. (1997). Characterization of the Yeast Transcriptome. *Cell* *88*, 243–251.
198. Volfson, D., Marciniak, J., Blake, W.J., Ostroff, N., Tsimring, L.S., and Hasty, J. (2006). Origins of extrinsic variability in eukaryotic gene expression. *Nature* *439*, 861–864.
199. Voliotis, M., Cohen, N., Molina-París, C., and Liverpool, T.B. (2008). Fluctuations, Pauses, and Backtracking in DNA Transcription. *Biophys J* *94*, 334–348.
200. Wang, Z., Zang, C., Cui, K., Schones, D.E., Barski, A., Peng, W., and Zhao, K. (2009). Genome-wide mapping of HATs and HDACs reveals distinct functions in active and inactive genes. *Cell* *138*, 1019–1031.
201. Weinberger, L.S., Burnett, J.C., Toettcher, J.E., Arkin, A.P., and Schaffer, D.V. (2005). Stochastic gene expression in a lentiviral positive-feedback loop: HIV-1 Tat fluctuations drive phenotypic diversity. *Cell* *122*, 169–182.
202. Weinberger, L.S., Dar, R.D., and Simpson, M.L. (2008a). Transient-mediated fate determination in a transcriptional circuit of HIV. *Nat Genet* *40*, 466–470.
203. Weinberger, L.S., Dar, R.D., and Simpson, M.L. (2008b). Transient-mediated fate determination in a transcriptional circuit of HIV. *Nat Genet* *40*, 466–470.

204. Wernet, M.F., Mazzoni, E.O., Celik, A., Duncan, D.M., Duncan, I., and Desplan, C. (2006). Stochastic spineless expression creates the retinal mosaic for colour vision. *Nature* *440*, 174–180.
205. Wickham, H. (2009). *ggplot2: Elegant Graphics for Data Analysis* (Springer).
206. Yamagata, Y., Szabó, P., Szüts, D., Bacquet, C., Arányi, T., and Páldi, A. (2012). Rapid turnover of DNA methylation in human cells. *Epigenetics* *7*, 0–1.
207. Yu, J., Xiao, J., Ren, X., Lao, K., and Xie, X.S. (2006). Probing gene expression in live cells, one protein molecule at a time. *Science* *311*, 1600–1603.
208. Yunger, S., Rosenfeld, L., Garini, Y., and Shav-Tal, Y. (2010). Single-allele analysis of transcription kinetics in living mammalian cells. *Nat Meth* *7*, 631–633.
209. Yusufzai, T.M., and Felsenfeld, G. (2004). The 5'-HS4 chicken  $\beta$ -globin insulator is a CTCF-dependent nuclear matrix-associated element. *Proc Natl Acad Sci U S A* *101*, 8620–8624.
210. Zee, B.M., Levin, R.S., Dimaggio, P.A., and Garcia, B.A. (2010). Global turnover of histone post-translational modifications and variants in human cells. *Epigenetics Chromatin* *3*, 22.
211. Zeng, L., Skinner, S.O., Zong, C., Sippy, J., Feiss, M., and Golding, I. (2010). Decision Making at a Subcellular Level Determines the Outcome of Bacteriophage Infection. *Cell* *141*, 682–691.
212. Zenklusen, D., Larson, D.R., and Singer, R.H. (2008). Single-RNA counting reveals alternative modes of gene expression in yeast. *Nat Struct Mol Biol* *15*, 1263–1271.
213. Zhang, C., Xing, D., and Li, Y. (2007). Micropumps, microvalves, and micromixers within PCR microfluidic chips: Advances and trends. *Biotechnol. Adv.* *25*, 483–514.

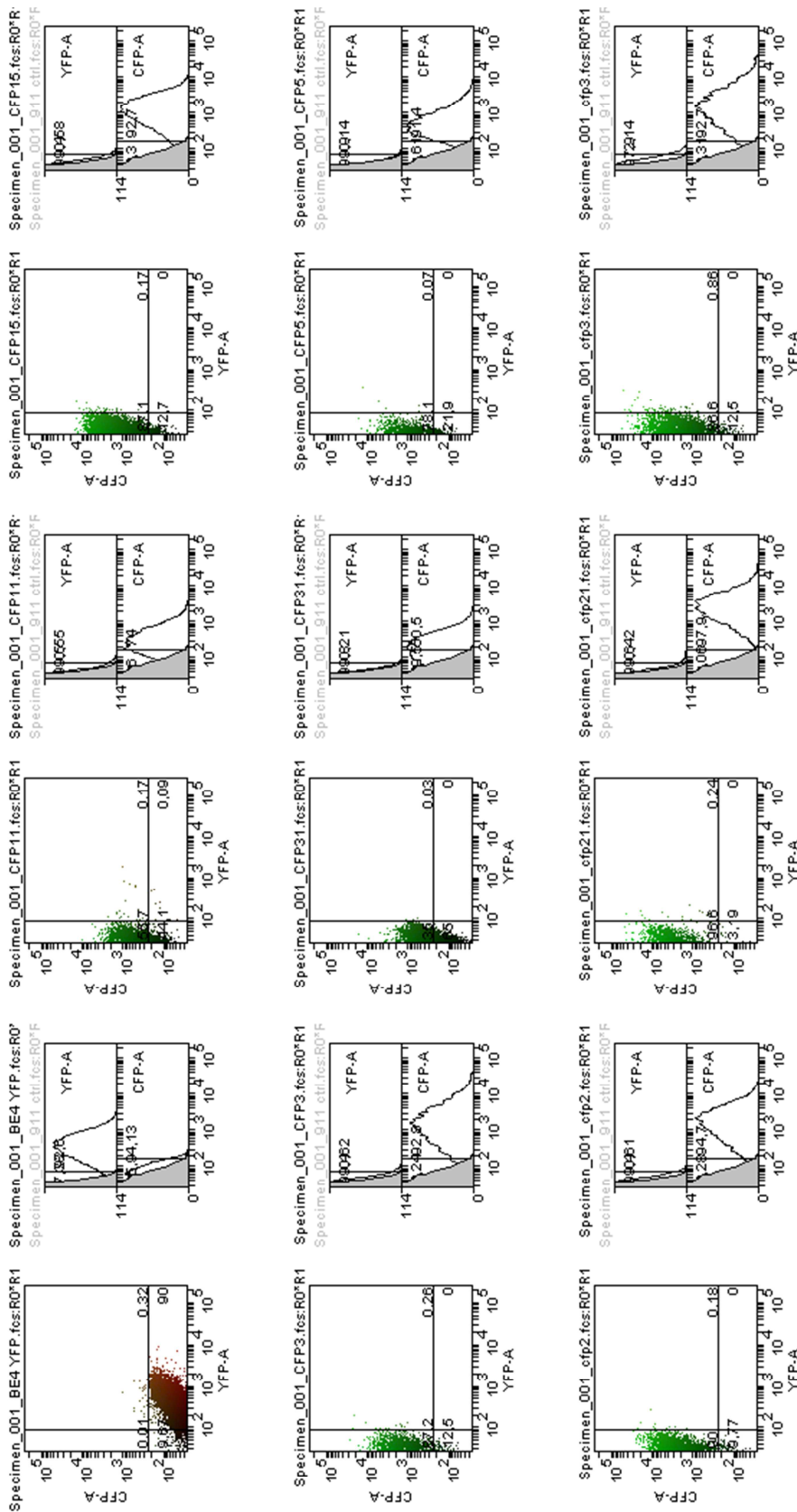
214. Zhang, C., Xu, J., Ma, W., and Zheng, W. (2006). PCR microfluidic devices for DNA amplification. *Biotechnol. Adv.* *24*, 243–284.
215. Zhang, Y., and Ozdemir, P. (2009). Microfluidic DNA amplification--a review. *Anal. Chim. Acta* *638*, 115–125.

## **Annexes**

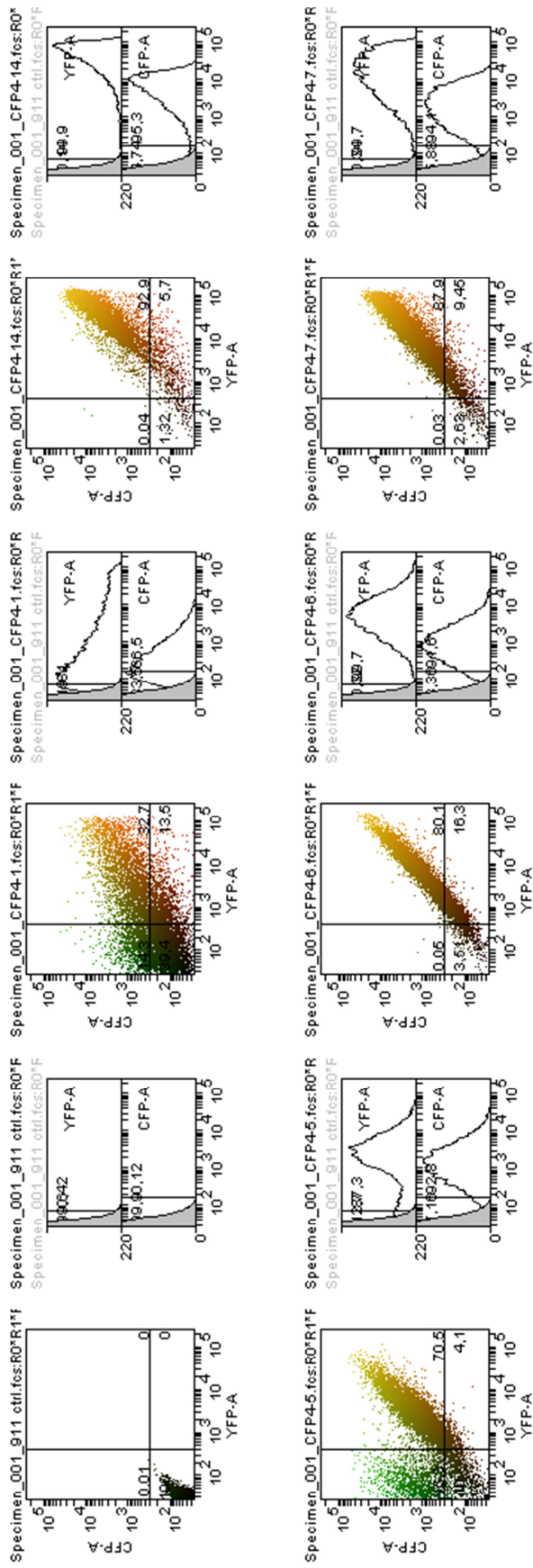
---



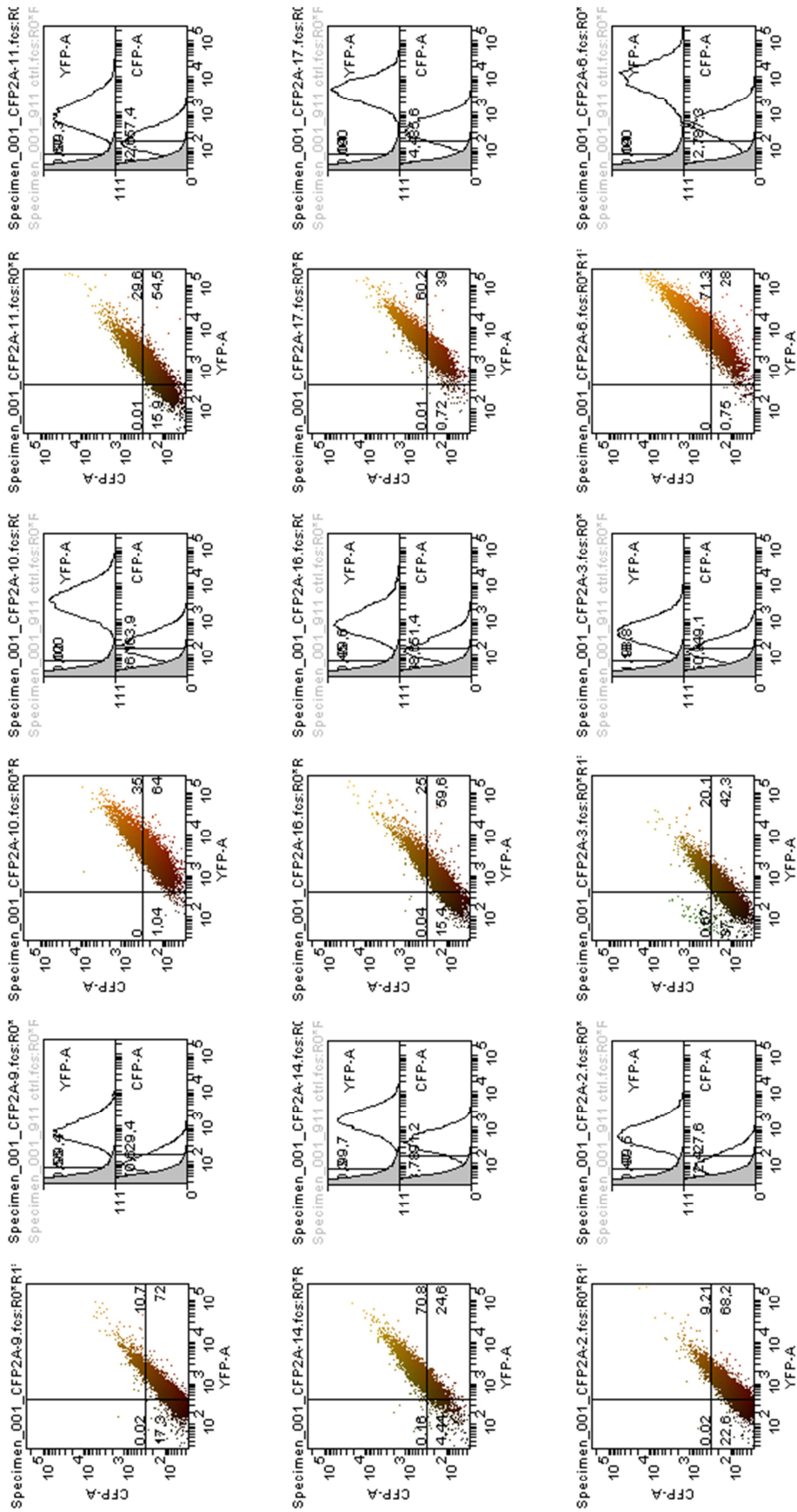
**Annexe 1 Analyse par cytométrie en flux des clones simple couleur eYFP.**



**Annexe 2 Analyse par cytométrie en flux des clones simple couleur eCFP.**

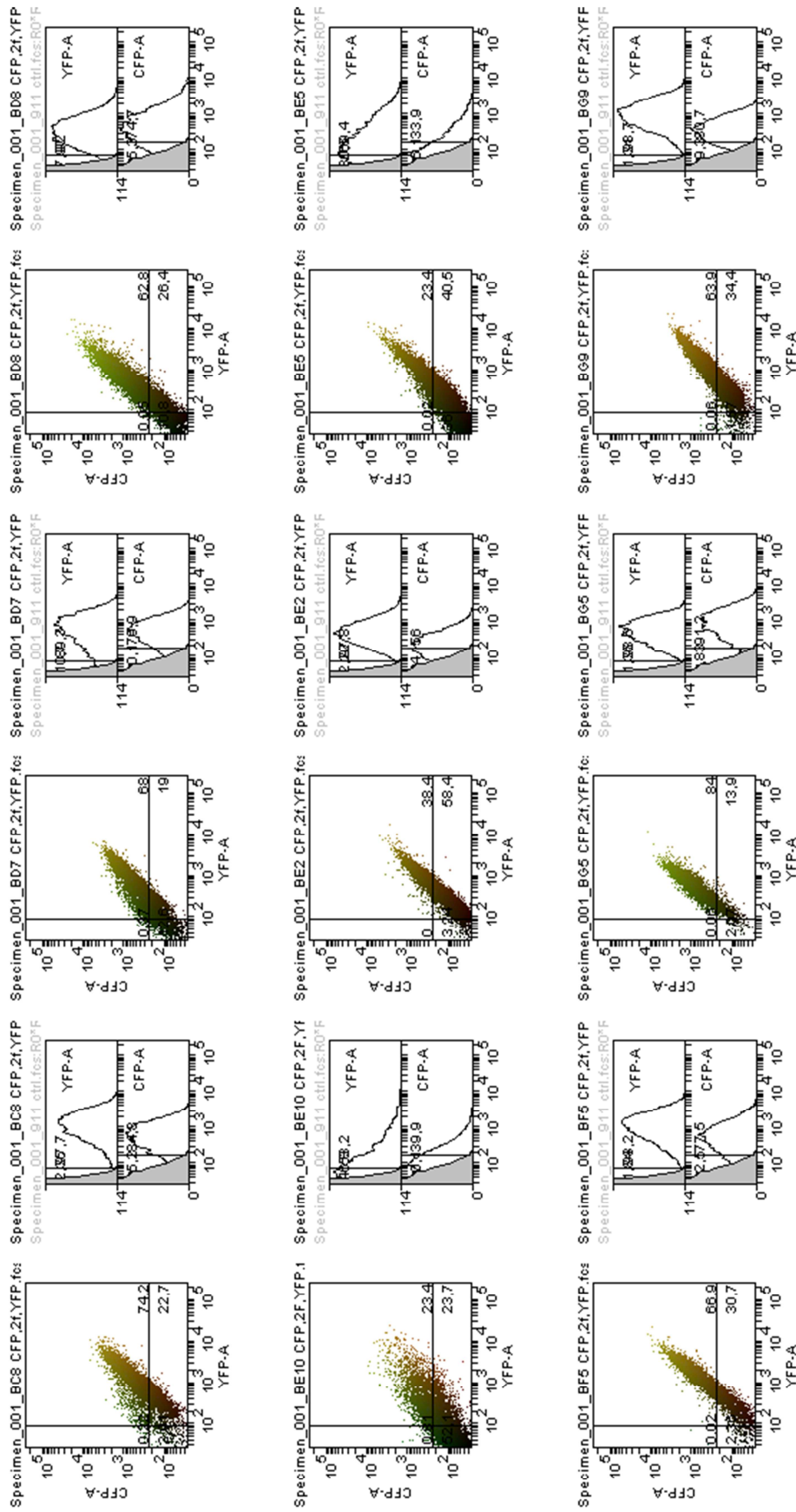


**Annexe 3 Analyse par cytométrie en flux des clones indépendants eCFP4 + YFP**

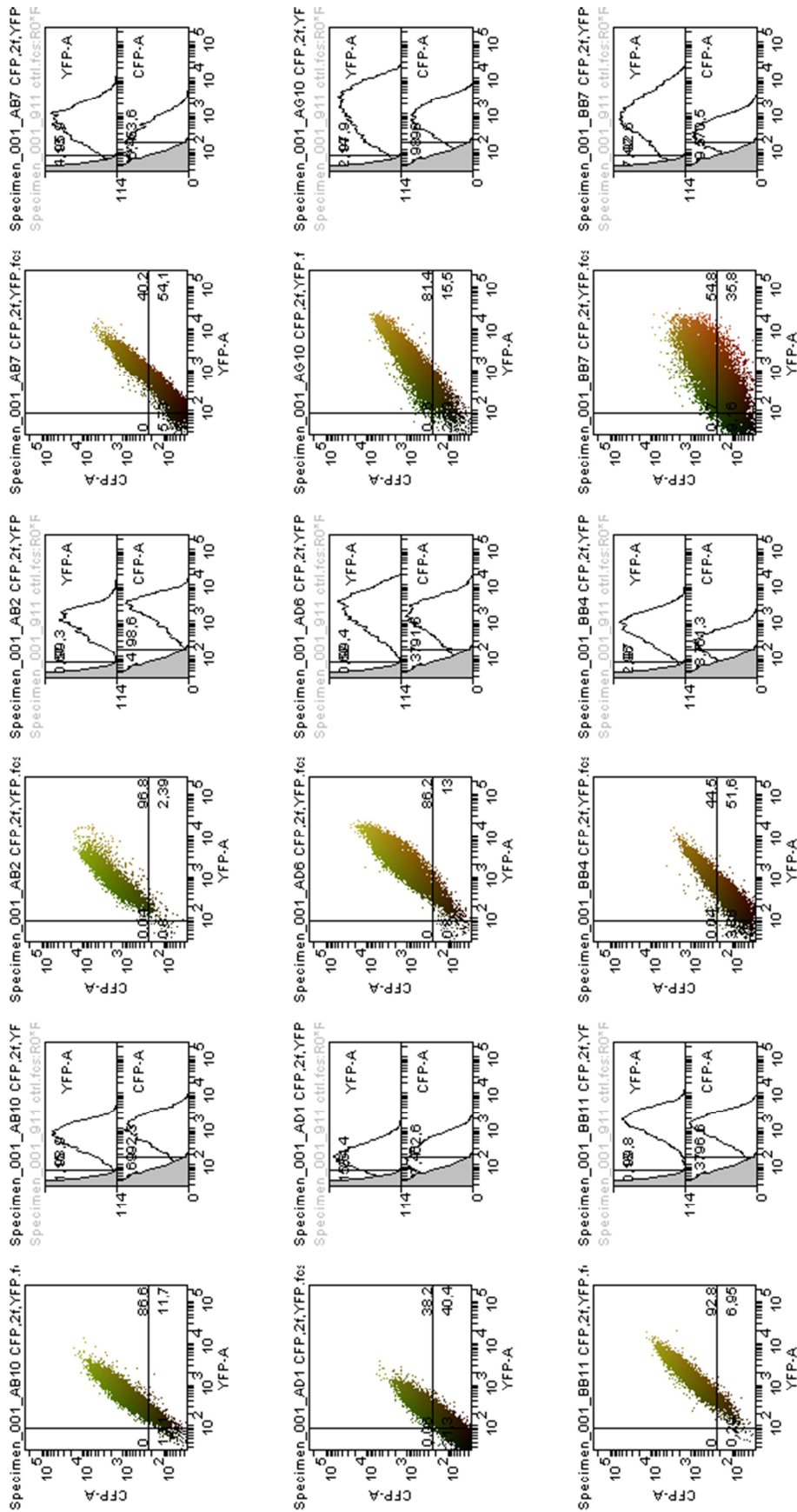


**Annexe 4 Analyse par cytométrie en flux des clones indépendants eCFP2A + eYFP.**

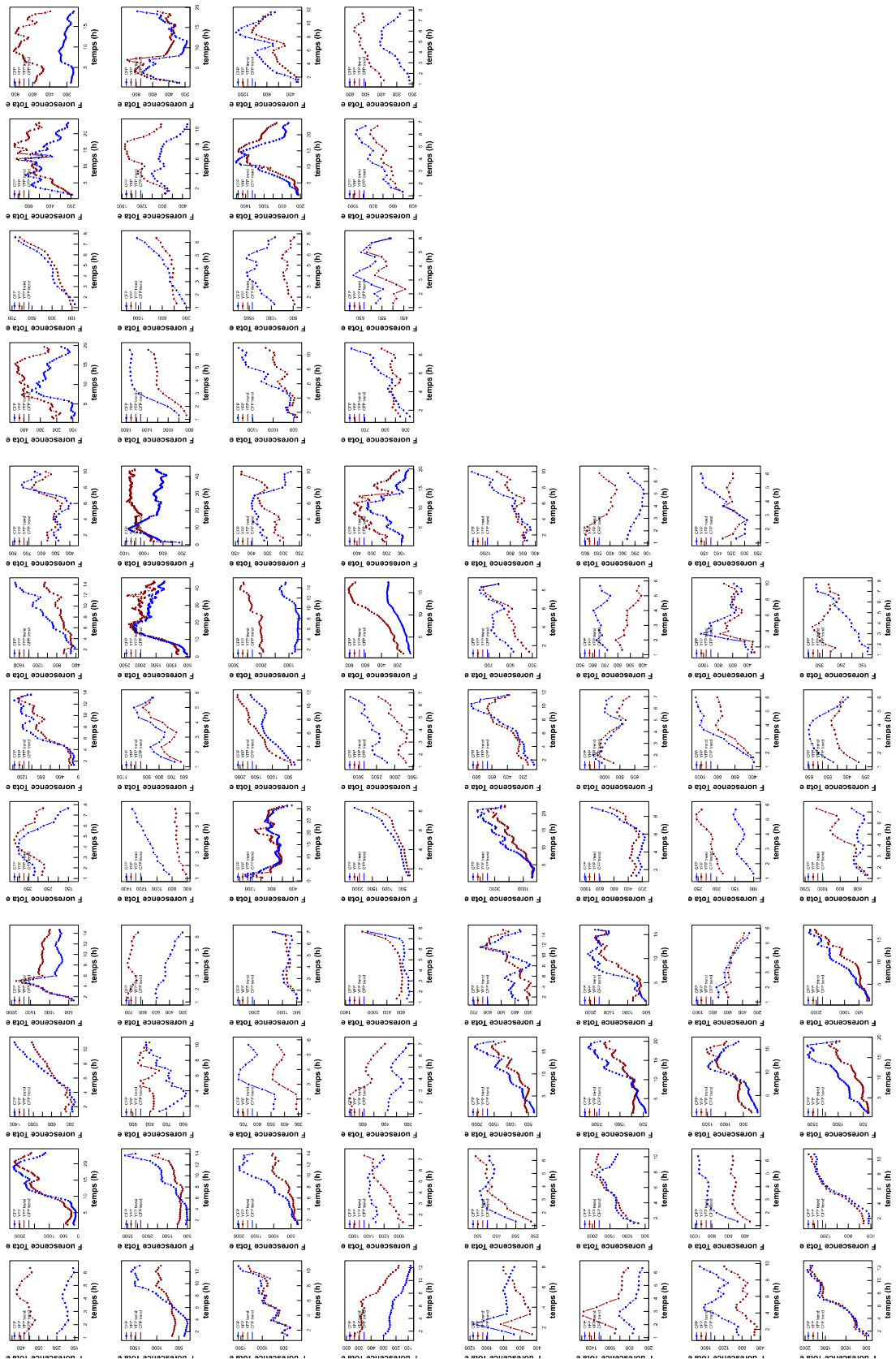




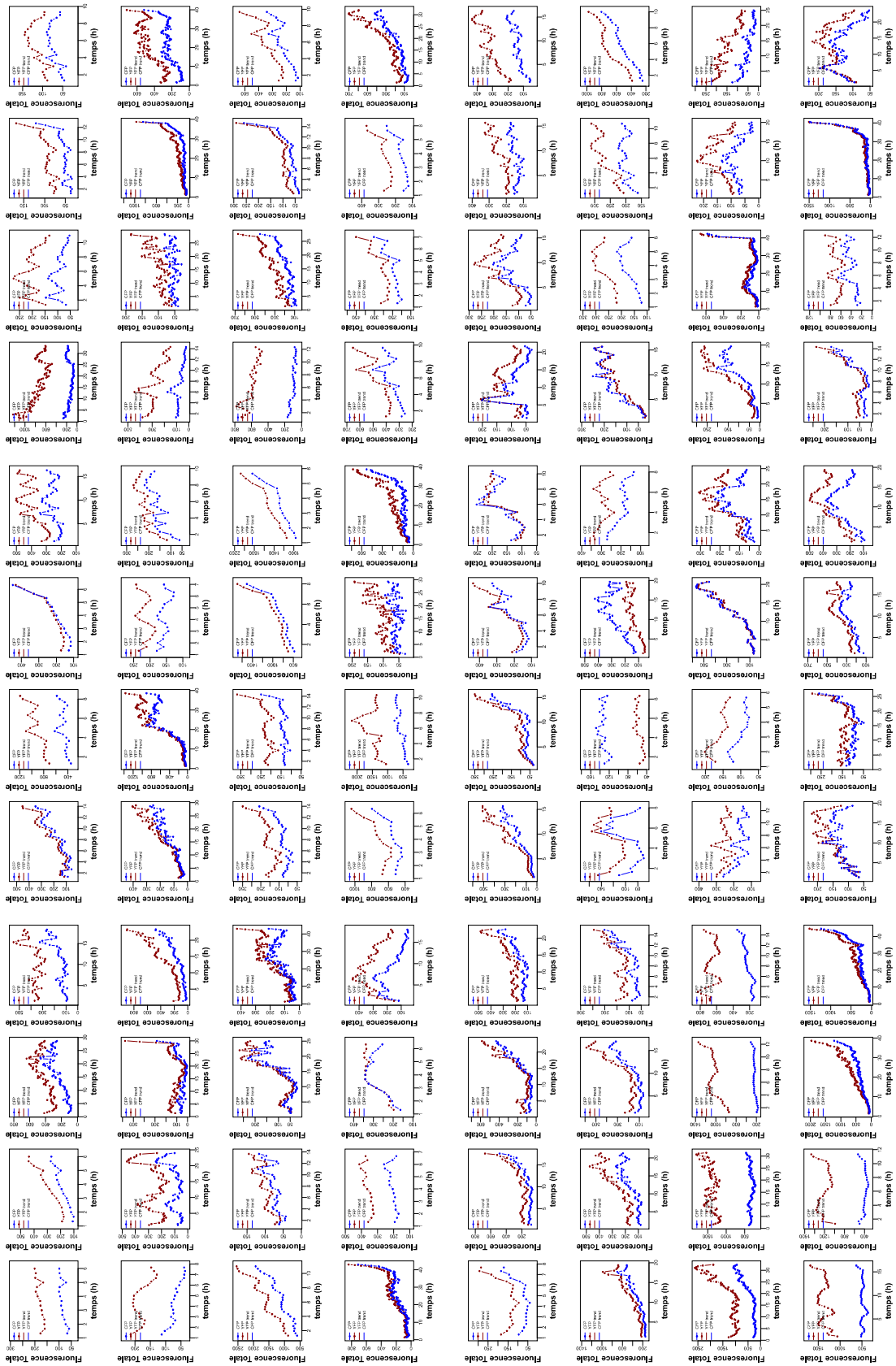
**Annexe 5 Analyse par cytométrie en flux des clones dépendants eCFP/eYFP.**



**Annexe 6 Analyse par cytométrie en flux des clones dépendants eCFP/e YFP (suite).**



**Annexe 7** *Dynamique d'expression des transgènes eCFP et eYFP dans des cellules isolées d'un clone « Dépendant ».*



**Annexe 8** *Dynamique d'expression des transgènes eCFP et eYFP dans des cellules isolées d'un clone*  
 « Indépendant ».



## **Publications**

---