

ARTICLE

Approches haut débit pour l'étude des séquences *cis*-régulatrices

Alexandre P. España^{1,2}, David Santiago-Algarra^{1,2}, Lydie Pradel^{1,2}, et Salvatore Spicuglia^{1,2,*}

¹ Aix-Marseille Université, INSERM, TAGC, UMR 1090, 13288 Marseille, France

² Équipe Labellisée Ligue Contre le Cancer, Laboratoire TAGC, INSERM U1090, Aix-Marseille Université, Parc Scientifique de Luminy, 163 avenue de Luminy, 13288 Marseille Cedex 09, France

Reçu le 16 mai 2018

Résumé—La régulation de la transcription des gènes chez les eucaryotes supérieurs implique l'action d'éléments régulateurs proximaux (promoteurs) ou distaux (amplificateurs ou *enhancers*) du site d'initiation de la transcription (*Transcription Start Site*, TSS). Il est aujourd'hui bien connu que les éléments *enhancers* jouent un rôle essentiel dans le développement et la différenciation cellulaire. Les altérations génétiques de ces éléments sont une cause majeure de pathologies humaines. De nombreuses stratégies ont été développées pour identifier et caractériser les *enhancers*. Ici, nous discutons des progrès récents pour évaluer de façon systématique l'activité des *enhancers* allant des approches haut débit de type « gène rapporteur » aux récentes technologies basées sur le système CRISPR/Cas9. Nous soulignons comment ces approches contribuent à une meilleure compréhension de la fonction des *enhancers* en conduisant à la découverte de nouveaux types de séquences de régulation et comment l'altération des *enhancers* peut affecter la régulation transcriptionnelle.

Mots clés : amplificateur, région *cis*-régulatrice, CRISPR-Cas9, génomique fonctionnelle, édition génomique

Abstract—**High-throughput approaches to study *cis*-regulating elements.** Gene expression in higher eukaryotes is regulated through the involvement of transcription start site (TSS)-proximal (promoters) and -distal (*enhancers*) regulatory elements. Enhancer elements play an essential role during development and cell differentiation, while genetic alterations in these elements are a major cause of human disease. Here, we discuss recent advances in high-throughput approaches to identify and characterize enhancer elements, from the well-established massively parallel reporter assays to the recent clustered regularly interspaced short palindromic repeats (CRISPR)/Cas9-based technologies. We discuss how these approaches contribute toward a better understanding of enhancer function in normal and pathological conditions.

Keywords: *enhancers*, CRISPR-Cas9, *cis*-regulatory elements, functional genomics, genome edition

Introduction

L'expression des gènes est rigoureusement contrôlée par une combinaison de promoteurs et de régions régulatrices distales connues sous le nom d'*enhancers* (Kim & Shiekhattar, 2015; Vernimmen & Bickmore, 2015). Parallèlement avec la prise de conscience croissante du rôle important des *enhancers* dans le développement normal et pathologique, il s'est développé un fort intérêt scientifique pour identifier et caractériser ces éléments. Or, cette tâche est difficile car un *enhancer* n'est pas nécessairement placé directement à côté du gène qu'il régule. Les *enhancers* potentiels peuvent être identifiés sur l'ensemble du génome à partir des régions ouvertes de la chromatine capturées à partir d'expériences de DNase,

ATAC-seq ou encore à partir de signatures chromatinien-nes (H3K4me1, H3K27ac) pour la cartographie des *enhancers* potentiellement actifs (Chatterjee & Ahituv, 2017). Bien qu'informatives, ces approches ne fournissent pas de preuves directes de la fonction *enhancer* d'une région. Elles ne permettent pas non plus d'identifier précisément les séquences responsables de l'activité *enhancer*. Par conséquent, il est crucial de vérifier la fonction *enhancer* de ces régions génomiques dans les cellules.

Ces dernières années, des techniques de pointe variées intégrant du séquençage haut débit dans des tests de type « gène rapporteur » ont permis de mesurer de manière directe et quantitative l'activité *enhancer* de plusieurs milliers de régions génomiques. Plus récemment, l'avènement des approches liées au CRISPR (*Clustered Regularly Interspaced Short Palindromic Repeats*) permet d'établir

*Auteur correspondant : salvatore.spicuglia@inserm.fr

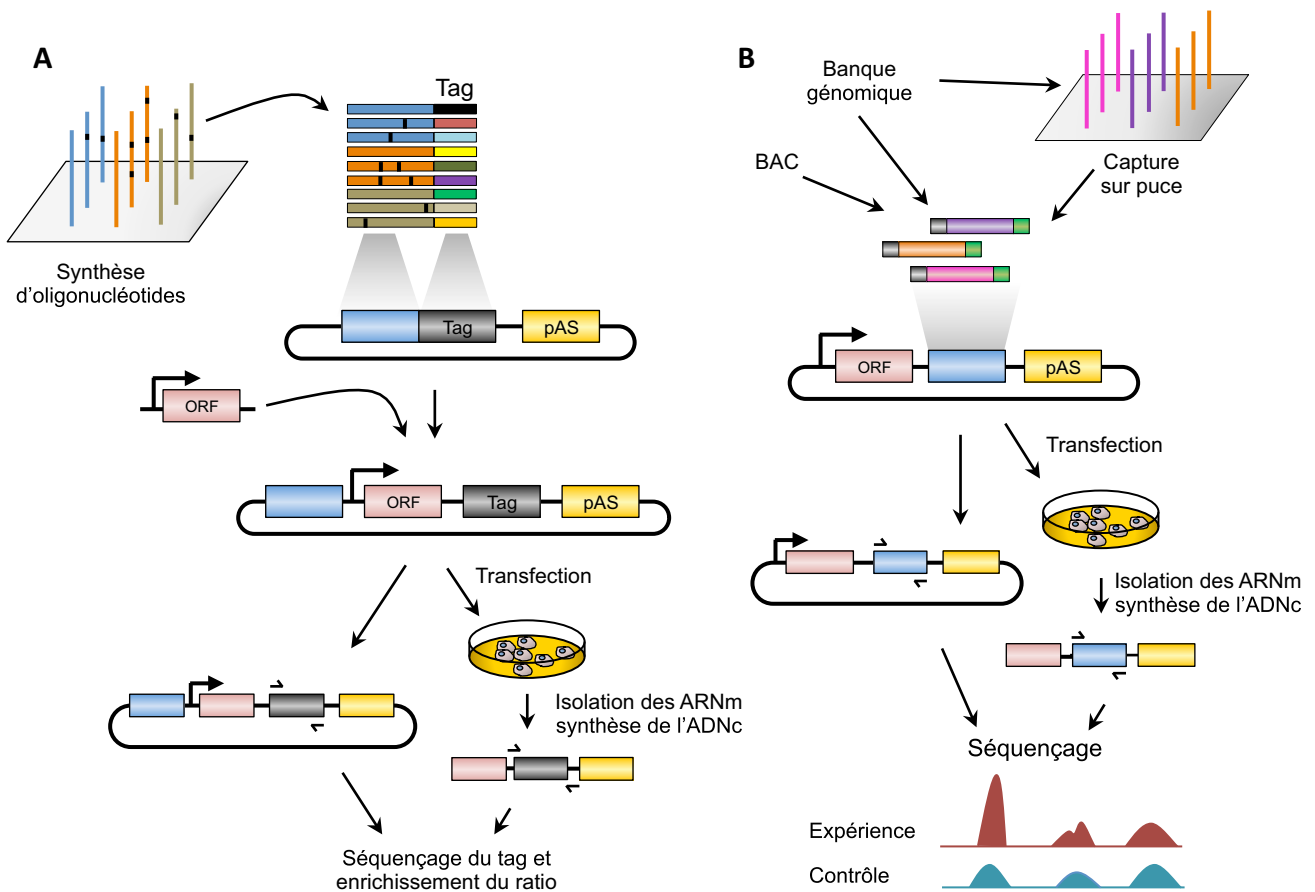


Figure 1. Principe des tests haut débit de l'activité enhancer. (A) Aperçu de la méthode MPRA (*Massively Parallel Reporter Assay*). Les séquences testées (variants, sauvages, etc.) sont généralement synthétisées *in silico* par synthèse massive d'oligonucléotides avec des codes-barres uniques (Tag) puis clonées dans le plasmide. Les tags peuvent être synthétisés en même temps que les séquences ou ajoutés après synthèse par amplification PCR (*Polymerase Chain Reaction*). Un promoteur basal et un ORF (*Open Reading Frame*) de gène rapporteur sont insérés entre l'élément testé et les codes-barres. La banque de gènes rapporteurs est ensuite transfectée dans des cellules en culture. Par la suite, l'ARNm est isolé et l'ADNc est synthétisé. Les codes-barres sont séquençés avant (banque plasmidique pour normalisation) et après transfection. La différence dans l'enrichissement de chaque code-barre est proportionnelle à l'activité *enhancer* de la séquence testée. Dans le cas de l'ajout de codes-barres après la synthèse, une étape de séquençage supplémentaire est nécessaire à la première étape de clonage. (B) Aperçu du séquençage de l'auto-transcription des régions régulatrices actives (STARR-Seq). Une banque de chromosomes bactériens artificiels (BAC) ou génomiques est clonée dans le plasmide rapporteur en aval de l'ORF et en amont du site de polyadénylation (pAS). Par ailleurs, les régions d'intérêt peuvent être enrichies par une approche de capture. La banque de gènes rapporteurs est transfectée dans des cellules en culture. Par la suite, l'ARNm est isolé et l'ADNc est synthétisé. Les régions clonées sont séquençées à partir de la banque de plasmides (contrôle) et de l'ADNc. Les différences d'enrichissement par rapport au contrôle sont proportionnelles à l'activité de l'*enhancer*. Dans les deux cas, l'effet de l'*enhancer* sur le promoteur basal est indiqué par une flèche.

à grande échelle la pertinence de la fonction *enhancer* dans le contexte endogène. Cette revue résume les essais mis au point pour tester l'activité des *enhancers* à l'échelle du génome, et pointe leurs limites, ainsi que les principales découvertes recueillies à l'aide des techniques présentées.

Principe des tests de gènes rapporteurs à haut débit

Les tests «gène rapporteur» épisomaux ont été largement utilisés pour caractériser les régions régulatrices potentielles. Plusieurs stratégies à haut débit ont été mises au point, permettant une analyse simultanée de centaines

de milliers de plasmides rapporteurs. Elles ont fait l'objet de plusieurs revues exhaustives (par ex. Dailey, 2015; Inoue & Ahituv, 2015; White, 2015). Ces méthodes conçues pour tester l'activité des *enhancers* ou des promoteurs peuvent être qualitatives (généralement grâce au tri cellulaire) ou quantitatives (basées sur les analyses de RNA-seq). Nous nous intéresserons ici aux méthodes quantitatives récentes visant à caractériser les *enhancers*. Deux approches en particulier ont été largement utilisées au cours de ces dernières années (Figure 1, Tableau 1) : l'analyse à haut débit en parallèle de gènes rapporteurs *Massively Parallel Reporter Assays* (MPRA) et le séquençage de régions régulatrices auto-transcrites (STARR-seq).

Tableau 1. Exemples de tests fonctionnels *enhancers* à grande échelle chez les mammifères.

Technique originelle	Nom spécifique	Origine de l'ADN	Application	Traits spécifiques	Nb. de régions ¹	Taille (pb ²)	Lignée cellulaire	Promoteur	Espèce	Ref.
MPRA		Synthétique	Caractériser des <i>enhancers</i> putatifs et leurs régions actives	Centré sur les sites de FT	2104	145	K562, HepG2	SV40	Humaine	(Kheradpour <i>et al.</i> , 2013)
MPRA	CRE-seq	Synthétique	Caractériser les régions génomiques avec une fonction <i>enhancer</i> prédite	Mutation des sites de fixation des FT	2100	130	K562	Hsp68	Humaine	(Kwasnieski <i>et al.</i> , 2014)
FIREWACH		Génomique	Identification d' <i>enhancers</i> à partir de régions accessibles aux enzymes de restriction	Isolation des régions GFP-positives	84 240	154	ESC	Minimal FpG5	Souris	(Murtha <i>et al.</i> , 2014)
STARR-seq	CapSTARR-seq	Génomique	Identification des <i>enhancers</i> par DHS	Capture des DHS	7542	330-430	P5424, 3T3	SCP1	Souris	(Vanhille <i>et al.</i> , 2015)
STARR-seq		Génomique	Caractériser des variants génétiques dans des séquences régulatrices de 95 patients	Capture des DHS	104	402	HepG2	SCP1	Humaine	(Vockley <i>et al.</i> , 2015)
MPRA	CRE-seq	Génomique	Identification de séquences régulatrices	Capture des DHS	4000	464	Retina	Minimal Rho	Souris	(Shen <i>et al.</i> , 2016)
MPRA	Sharp-MPRA	Synthétique	Caractériser les nucléotides comme activateurs ou répresseurs dans des séquences régulatrices	Synthèse d'oligonucléotides chevauchant (résolution de 30 ou 5 pb)	15 720	145	HepG2, K562	Minimal TATA, SV40	Humaine	(Ernst <i>et al.</i> , 2016)
MPRA		Synthétique	Identification des variants régulateurs associés aux eQTL	Centré sur les variants eQTL	3642	150	Lymphoblastoid, HepG2	Minimal TATA	Humaine	(Tewhey <i>et al.</i> , 2016)

Tableau 1. (suite).

Technique originelle	Nom spécifique	Origine de l'ADN	Application	Traits spécifiques	Nb. de régions ¹	Taille (pb ²)	Lignée cellulaire	Promoteur	Espèce	Ref.
MPRA		Synthétique	Identification des variants régulateurs associés aux traits des globules rouges	3 fenêtres couissantes par rapport aux variants GWAS	2756	145	K562, K562 +GATA1	Minimal TATA	Humaine	(Ulirsch <i>et al.</i> , 2016)
STARR-seq	CapSTARR-seq	Génomique	Identification des régions promotrices avec activité <i>enhancer</i>	Capture de -200 à +50 pnb par rapport aux TSS des gènes codants	20 719	250	K562, HeLa	SCP1	Humaine	(Dao <i>et al.</i> , 2017)
MPRA	LentiMPRA	Synthétique	Identification des <i>enhancers</i> par intégration chromosomique	Centré sur le signal de plusieurs pics de ChIP-seq	2236	171	HepG2	pGL4.23 promoteur	Humaine	(Inoue <i>et al.</i> , 2017)
CRISPR-Cas9		Synthétique	Identification d' <i>enhancers</i> endogènes fixés par p53 et ER α	Ciblé sur les sites de fixation de p53 dans des <i>enhancers</i> putatifs	685	I	BJ-RASg12v	I	Humaine	(Korkmaz <i>et al.</i> , 2016)

¹ Nombre de séquences d'ADN ciblées, pas nécessairement le nombre unique de fragments qui ont été testés.

² pb : paire de base ; capSTARR-seq : *capture-based self-transcribing active regulatory region sequencing* ; CHIP-seq : *chromatin immunoprecipitation sequencing* ; CRE : *cis-regulatory elements* ; CRISPR : *clustered regularly interspaced short palindromic repeats* ; DHS : *DNase I hypersensitive sites* ; eQTL : *expression quantitative trait loci* ; ER α : *estrogen receptor alpha* ; ESC : *embryonic stem cell* ; FIREWACH : *functional identification of regulatory elements within accessible chromatin* ; GFP : *green fluorescent protein* ; GWAS : *genome-wide association study* ; Hsp68 : *heat shock promoter 68* ; lentiMPRA : *lentiviral massively parallel reporter assay* ; MPRA : *massively parallel reporter sequencing* ; N/A : non applicable ; Sharpr : *systematic high-resolution activation and repression profiling with reporter-tiling* ; STARR-seq : *self-transcribing active regulatory region sequencing* ; SCP1 : *super core promoter 1* ; SV40 : *simian virus 40* ; FT : Facteur de transcription ; TSS : *transcription start site*.

La méthode MPRA consiste en la génération d'une banque de plasmides rapporteurs contenant les séquences d'ADN d'intérêt généralement clonées en amont d'un promoteur basal. Les régions d'intérêt sont préalablement produites grâce à la technologie de synthèse massive d'ADN sur micropuce et amplifiées par PCR à l'aide d'oligonucléotides contenant des séquences tags ou codes-barres uniques (placées dans les 3' UTR du gène rapporteur). Pour augmenter la sensibilité et la reproductibilité, plusieurs codes-barres peuvent être ajoutés à n'importe quelle séquence. La banque de rapporteurs est ensuite transfectée dans les lignées cellulaires pertinentes et le séquençage ARN des codes barre est réalisé, fournissant ainsi une lecture quantitative de l'activité régulatrice des régions testées (Figure 1A).

Le MPRA a été utilisé pour étudier un certain nombre de questions biologiques. À l'origine, il a été conçu pour examiner les composants fonctionnels d'*enhancers* précédemment identifiés à la résolution d'un seul nucléotide (Melnikov *et al.*, 2012; Patwardhan *et al.*, 2012). Par la suite, une approche similaire (appelée CRE-seq) a été utilisée pour tester fonctionnellement environ 2000 régions prédites par ENCODE comme étant des *enhancers*, des *enhancers* faibles ou des éléments réprimés (Kwasnieski *et al.*, 2014) et des *enhancers* synthétiques afin de modéliser des règles grammaticales des séquences régulatrices (Smith *et al.*, 2013; Nguyen *et al.*, 2016). Le MPRA peut également être utilisé pour évaluer systématiquement la pertinence des motifs de régulation prédits à l'intérieur d'*enhancers*. Kheradpour *et al.* ont en effet testé environ 2000 *enhancers* putatifs ainsi que des *enhancers* synthétiques contenant des variants de motifs ciblés par certains facteurs de transcription (FT) (Kheradpour *et al.*, 2013). Par la suite, le même laboratoire a mis au point une approche MPRA à haute résolution (également appelée Sharpr-MPRA) qui a permis de cartographier à l'échelle du génome les nucléotides activateurs et répresseurs dans les régions régulatrices (Ernst *et al.*, 2016). Grâce à la synthèse de constructions MPRA chevauchantes à large couverture, ces auteurs sont parvenus à déduire les effets régulateurs de nucléotides ayant des propriétés soit activatrices soit répressives (Ernst *et al.*, 2016; Nguyen *et al.*, 2016). Enfin, le MPRA peut être utilisé afin de tester l'impact de SNP (*Single-Nucleotide Polymorphisms*) et identifier les variants régulateurs fonctionnels associés à des caractères ou des maladies humaines. Deux études récentes du *Broad Institute* ont fourni la preuve de principe de telles approches. Tewhey *et al.* ont utilisé une version améliorée du MPRA pour analyser des milliers d'eQTL (*expression Quantitative Trait Loci*) et identifier les allèles qui ont un impact sur l'expression génique dans des lignées cellulaires lymphoblastiques (Tewhey *et al.*, 2016). Ulirsch *et al.* ont utilisé le MPRA pour tester 2756 variants associés à 75 loci d'association pangénomique (GWAS) impliqués dans les caractères érythrocytaires (Ulirsch *et al.*, 2016). Dans les deux cas, la pertinence des résultats du MPRA a été confirmée par l'ingénierie génétique CRISPR. Il est intéressant de noter que certains de ces variants régulateurs identifiés ne se

situait pas dans des motifs connus, suggérant ainsi qu'ils influencent la structure de l'ADN plutôt que la liaison des FT (Ulirsch *et al.*, 2016).

Alexander Stark et ses collègues ont introduit une nouvelle méthode innovante appelée STARR-seq (Arnold *et al.*, 2013) (pour revue, voir Muerdter *et al.*, 2015). Le STARR-seq a pour but d'identifier et de quantifier les *enhancers* transcriptionnels directement en se basant sur leur activité dans le génome entier (Figure 1B). En bref, un grand nombre de fragments d'ADN de provenances arbitraires sont clonés en aval dans le 3' UTR d'un gène rapporteur (GFP). Une fois dans un contexte cellulaire, les *enhancers* actifs activeront le promoteur en amont et se transcriront eux-mêmes, générant des transcrits rapporteurs parmi les ARN cellulaires. Ainsi, chaque transcrit contient le gène rapporteur et son « code-barre ». Ces transcrits peuvent être isolés séparément par PCR ciblée et éventuellement détectés par séquençage haut débit. De cette façon, l'activité de millions d'*enhancers* putatifs peut être mesurée simultanément, sans être affectée par l'emplacement des séquences candidates et leur orientation. Le principal avantage par rapport au MPRA classique est que la séquence testée elle-même est utilisée comme un « code-barre », simplifiant considérablement l'ensemble de la procédure de quantification de l'activité des *enhancers*. Le laboratoire d'Alexander Stark a utilisé les approches STARR-seq pour poser plusieurs questions mécanistiques de base sur la biologie des *enhancers* chez la drosophile et notamment (i) l'identification et la caractérisation des *enhancers* spécifiques d'un type cellulaire (Arnold *et al.*, 2013; Yáñez-Cuna *et al.*, 2014) et hormonosensibles (Shlyueva *et al.*, 2014), (ii) l'impact de la variation de séquence *cis*-régulatrice sur l'activité et l'évolution des *enhancers* (Arnold *et al.*, 2014) et (iii) la dissection des bases de spécificité des *enhancers* sur différents types de promoteurs (de gènes développementaux *versus* gènes constitutifs) (Zabidi *et al.*, 2015).

Le STARR-seq a été appliqué aux cellules humaines en utilisant les BAC (*Bacterial Artificial Chromosomes*) (Arnold *et al.*, 2013). Cependant, avec la complexité et la taille des génomes de mammifères, cette technique n'est pas facile à mettre en œuvre. La préparation de banques représentatives est un défi et requiert une très grande profondeur de séquençage. Pour éviter ce problème, nous avons développé une approche basée sur la capture de séquences (appelée CapSTARR-seq) pour évaluer un sous-ensemble de Sites Hypersensibles à la DNase (DHS) dans des thymocytes en développement chez la souris (Vanhille *et al.*, 2015). Les régions d'intérêt sont capturées à l'aide de puces à ADN synthétisées à façon et clonées dans le vecteur STARR-seq, ce qui permet une quantification précise et économiquement rentable de l'activité des *enhancers* chez les mammifères. Des approches similaires ont été publiées par d'autres laboratoires, y compris la capture de variants génomiques naturels (Vockley *et al.*, 2015) et de séquences DHS du système nerveux central en utilisant une approche de capture avec des oligo-appâts (Shen *et al.*, 2016). Alternativement, il serait possible de cloner directement les régions de chromatine ouvertes

(Murtha *et al.*, 2014; Barakat *et al.*, 2017; Wang *et al.*, 2017). En effet, récemment, deux études indépendantes ont réussi à réaliser des expériences de STARR-seq sur le génome humain entier (Liu *et al.*, 2017; Muerdter *et al.*, 2018).

Mise en garde concernant les essais de gènes rapporteurs à haut débit

Les sources d'ADN utilisées sont un problème potentiel pour les tests de gènes rapporteurs à haut débit. La plupart des approches MPRA ont fait appel à de la synthèse massive d'oligonucléotides (Figure 1A), ce qui permet de définir avec précision les régions testées et de modifier à façon les séquences sous-jacentes. Toutefois, cette approche comporte actuellement deux limites. D'une part, la taille du fragment testé est limitée à ~200 bp (en incluant les adaptateurs), ce qui peut empêcher de tester des régions régulatrices complètes. De plus, le nombre d'oligonucléotides qui peut être synthétisé est limité (actuellement 100 000 au maximum). Ces contraintes devraient être surmontées dans un proche avenir grâce à l'amélioration des technologies de synthèse des oligonucléotides. Un autre biais concerne le fait que la transfection des plasmides dans les cellules de mammifères induit l'activation d'une réponse interféron de type I pouvant fausser l'interprétation des résultats (Muerdter *et al.*, 2018).

Dans l'approche STARR-seq, les fragments d'ADN sont clonés dans la région transcrite (Figure 1B). Leurs séquences fournissent alors directement les informations sur l'activité de l'*enhancer*. Cependant, cela introduit également une source d'artefact potentiel car certaines séquences peuvent influencer la stabilité du transcrit plutôt que l'activation de la transcription. Ce biais éventuel pourrait être évité en comparant les résultats des régions testées dans les deux orientations permettant ainsi d'éliminer l'effet stabilisant spécifique d'un brin.

Une préoccupation générale au sujet des tests rapporteurs épisomaux est qu'ils peuvent ne pas refléter précisément la fonction des éléments *enhancers* dans leur contexte endogène. Pour éviter partiellement cet inconvénient, des MPRA lentiviraux (Murtha *et al.*, 2014; Inoue *et al.*, 2017) ou associés à de la chromatine d'adénovirus (Nguyen *et al.*, 2016; Shen *et al.*, 2016) ont été réalisés. Ces méthodes permettent d'effectuer des tests rapporteurs dans des cellules ou des tissus difficiles à transfecter. Bien sûr, un argument également valable est que les tests de gènes rapporteurs épisomaux permettent l'étude non biaisée de la fonction d'*enhancer* indépendamment de l'influence de la conformation de la chromatine ou du contexte génomique. Il est intéressant de noter qu'une étude récente a réalisé une comparaison systématique de l'activité *enhancer* épisomale *versus* chromosomique en utilisant des versions intégratives et non intégratives d'un test rapporteur lentiviral (Inoue *et al.*, 2017). Bien que le test gène rapporteur chromosomal soit plus prévisible par les modèles épigénomiques et

séquentiels, les deux tests étaient relativement bien corrélés. Une autre approche alternative est l'introduction de gènes rapporteurs dans le génome à l'aide de systèmes de transposition (Ruf *et al.*, 2011; Akhtar-Zaidi *et al.*, 2012; Symmons *et al.*, 2014)). Bien que ces approches n'évaluent pas directement l'activité *enhancer*, elles permettent de tester l'influence du contexte régulateur de loci endogènes.

Approches fondées sur le CRISPR pour l'évaluation de la fonction *enhancer*

Comme nous l'avons déjà mentionné, une des limites potentielles des essais de gènes rapporteurs est le fait que les *enhancers* candidats soient étudiés en dehors de leur contexte génomique endogène, lequel est probablement nécessaire pour leur fonction *in vivo*. L'avènement des technologies CRISPR permet maintenant de contourner cette restriction (Lo & Qi, 2017; Montalbano *et al.*, 2017). Plusieurs études ont effectué une dissection systématique d'*enhancers* individuels à l'aide de mutagenèses induites par TALEN (Vierstra *et al.*, 2015) ou CRISPR (Canver *et al.*, 2015). Dans le cadre de ces études, une banque d'ARN guide unique ou TALEN, qui couvre des *enhancers* sélectionnés, a été conçue pour effectuer des analyses saturantes de mutagenèse *in situ*, en mettant en évidence des séquences ayant un impact positif ou négatif sur la fonction *enhancer*. L'approche du CRISPR a par la suite été élargie pour évaluer la fonction *enhancer* dans de grandes régions génomiques entourant des loci d'intérêt (Diao *et al.*, 2016, 2017; Rajagopal *et al.*, 2016; Sanjana *et al.*, 2016) ou pour identifier des éléments *enhancers* impliqués dans des voies de régulation génique spécifiques (Korkmaz *et al.*, 2016) (Tableau 1).

La mutagenèse médiée par le CRISPR est limitée par le fait qu'une densité élevée d'ARN guides est nécessaire pour saturer tous les éléments régulateurs possibles et que des mutations spécifiques sont difficiles à mettre en œuvre à l'échelle du génome. Dans certains cas, il peut aussi y avoir un biais en ce qui concerne les régions qui peuvent être ciblées par les ARN guides ou des limitations dues à l'efficacité de la transfection dans certains types de cellules. Alternativement, la Cas9 avec l'activité nucléase désactivée (dCas9) peut être fusionnée à des domaines activateurs ou répresseurs pour modifier précisément l'expression génique des promoteurs et des éléments régulateurs distaux (Hsu *et al.*, 2014). Sur la base de cette propriété, les domaines répresseurs et activateurs fusionnés avec la dCas9 et combinés à un pool d'ARN guides ont été utilisés pour générer du CRISPR d'activation ou d'inactivation (CRISPRi) en ciblant des sites DHS associés à un gène d'intérêt (Klann *et al.*, 2017) ou l'ensemble d'un locus (Fulco *et al.*, 2016). Dans la première étude, un gène rapporteur, introduit à la place des gènes cibles, a été utilisé pour évaluer l'activité des *enhancers*. Dans la deuxième étude, les critères de sélection se fondaient sur l'avantage ou le désavantage de croissance que procurait le changement d'expression du gène associé

à l'*enhancer*, fournissant ainsi une preuve de concept pour l'identification à l'échelle du génome des régions *enhancers* fonctionnelles. Une des limites actuelles de ces approches est que la stratégie d'étude est fondée sur des caractéristiques phénotypiques (comme la capacité de croissance cellulaire, les marqueurs du développement, etc...) au lieu d'évaluer directement les niveaux d'expression des gènes régulés. Pour surmonter cette limitation, une nouvelle et puissante méthode a combiné le CRISPRi et le RNA-seq sur cellule unique (Xie *et al.*, 2017) permettant ainsi de relier directement la fonction *enhancer* et l'effet combinatoire de plusieurs variants de chaque *enhancer* avec le ou les gènes cibles. Bien que ces approches aient été utilisées jusqu'à présent pour étudier des zones génomiques restreintes, elles seront vraisemblablement mises en œuvre dans l'avenir pour une véritable dissection des éléments *enhancers* à l'échelle du génome.

Quelques résultats pertinents

L'un des intérêts des tests *enhancer* à haut débit est la possibilité d'explorer la fonction d'*enhancer* sans notions préconçues, menant ainsi à de nouvelles découvertes imprévues. Une observation commune de plusieurs études est que de nombreuses régions prédites pour avoir une fonction *enhancer* ne présentent pas d'activité rapporteur (Arnold *et al.*, 2013 ; Kheradpour *et al.*, 2013 ; Kwasnieski *et al.*, 2014 ; Vanhille *et al.*, 2015 ; Ernst *et al.*, 2016). Par exemple seulement 26 % des *enhancers* prédits, basés sur les signatures de chromatine dans les cellules K562, présentaient une activité *enhancer* dans les tests rapporteurs effectués sur la même lignée (Kwasnieski *et al.*, 2014). Ceci suggère qu'en plus des modifications d'histones, des spécificités séquentielles additionnelles, comme des sites de liaison aux facteurs de transcription, sont des déterminants essentiels de l'activité *cis*-régulatrice. En effet, la concentration de sites ou de motifs liant des facteurs de transcription est hautement prédictive d'une forte activité *enhancer* (Dogan *et al.*, 2015 ; Vanhille *et al.*, 2015 ; Ernst *et al.*, 2016). Par ailleurs, cela pourrait également indiquer que toutes les séquences requises ne sont pas présentes dans les régions testées ou que les contextes des promoteurs endogènes sont essentiels à l'activité *enhancer*. Enfin, il est également possible que certaines régions de chromatine ouvertes, bien que contribuant à la régulation transcriptionnelle, aient des fonctions indépendantes des *enhancers* (Bevington *et al.*, 2017) ou n'aient pas les fonctions classiques des *enhancers* (Engreitz *et al.*, 2016).

Les résultats les plus surprenants des tests des *enhancers* fonctionnels proviennent peut-être de l'observation selon laquelle beaucoup de promoteurs affichent une activité *enhancer* (Arnold *et al.*, 2013 ; Zabidi *et al.*, 2015 ; Nguyen *et al.*, 2016 ; Dao *et al.*, 2017 ; Medina-Rivera *et al.*, 2018 ; Muerdter *et al.*, 2018). La définition originale des *enhancers* implique la capacité à activer l'expression de gènes à distance, tandis que les promoteurs impliquent la capacité d'induire une expression de gènes au niveau local. Toutefois, cette dichotomie fondamentale

des éléments *cis*-régulateurs a été contestée par la mise en évidence de grandes similitudes entre les promoteurs et les *enhancers*, telles que les caractéristiques de séquences d'ADN, les marques de chromatine, le recrutement de la polymérase II et la transcription bidirectionnelle (Andersson, 2015). Par exemple, la modification de l'histone généralement observée dans les régions de promoteurs, H3K4me3, a également été associée à des *enhancers* (Koch *et al.*, 2011 ; Pekowska *et al.*, 2011 ; Core *et al.*, 2014). L'évaluation de l'activité des *enhancers* par CapSTARR-seq a montré que les *enhancers* distaux du site d'initiation de la transcription (*Transcription Start Site*, TSS) sont effectivement associés à un enrichissement en H3K4me3 dans les *loci* endogènes (Vanhille *et al.*, 2015). Plusieurs études ont également suggéré que certains promoteurs pourraient avoir des fonctions *enhancers* (Medina-Rivera *et al.*, 2018). Plusieurs études indépendantes basées sur les tests gènes rapporteurs à haut débit ont notamment fait état d'une activité *enhancer* étendue à partir des régions proximales aux TSS. En appliquant la technique du STARR-seq, Zabidi *et al.* (2015) ont passé au crible le génome de la drosophile à l'aide de différents promoteurs obtenus à partir de gènes de ménage, de gènes impliqués dans le développement et de gènes spécifiques de type cellulaire. Ils ont découvert que les promoteurs des gènes constitutifs étaient principalement régulés par des promoteurs proximaux à activité *enhancer* tandis que les promoteurs des gènes développementaux et des gènes tissu-spécifiques ont quant à eux besoin d'*enhancers* distaux. Ernst *et al.* ont montré dans les lignées cellulaires humaines que les *enhancers* actifs étaient enrichis en sites DNaseI qui chevauchent les TSS (Ernst *et al.*, 2016). Nguyen *et al.* (2016) ont réalisé une comparaison fonctionnelle d'une série de promoteurs et *enhancers* issus de cellules neuronales de souris en utilisant une approche MPRA intégrative. Il est intéressant de noter que les promoteurs de gènes ont généré une activité *enhancer* similaire aux régions régulatrices distales. Dans une étude récente, nous avons constaté que 2 à 3 % de tous les promoteurs humains présentaient une activité *enhancer* dans une lignée cellulaire donnée (Dao *et al.*, 2017). Comparé aux promoteurs et aux *enhancers* distaux classiques, ces promoteurs-*enhancers* se chevauchaient, présentaient des caractéristiques génomiques et épigénomiques distinctes et étaient associés à des gènes constitutifs et de réponse au stress. Les délétions génomiques par CRISPR ont démontré que plusieurs promoteurs, avec une activité *enhancer* dans leur contexte naturel, sont en effet impliqués dans la *cis*-régulation de l'expression de gènes distaux et fonctionnent en outre comme de véritables *enhancers* (Dao *et al.*, 2017). De plus, des variations génétiques humaines au sein de ce type de promoteur ont été associées à un effet important sur l'expression des gènes distaux. Parallèlement, une autre étude utilisant la manipulation génétique par CRISPR/Cas9 des régions promotrices a fait état d'une *cis*-régulation distale fréquente par des *loci* associés à des promoteurs d'ARNnc et dans une moindre mesure par des *loci* avec des gènes codants (Engreitz *et al.*, 2016). Enfin, deux études récentes

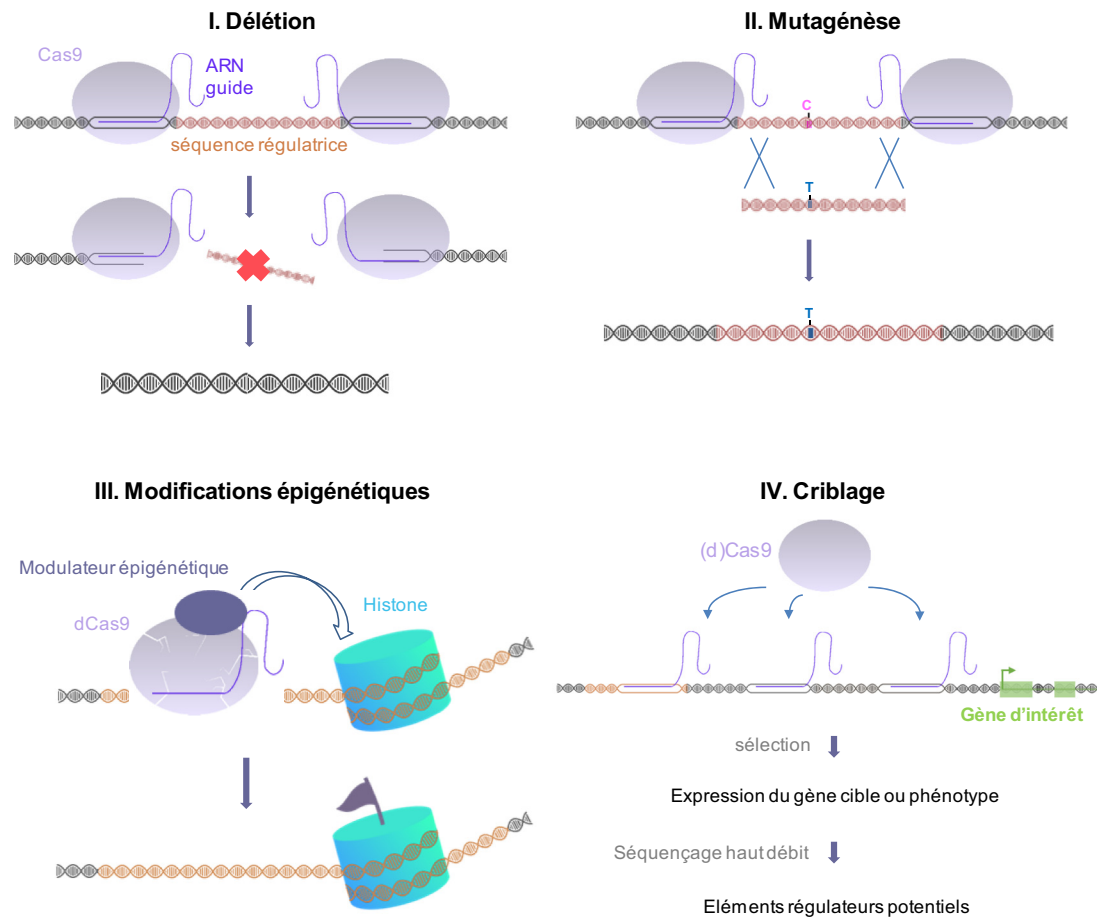


Figure 2. Étude d'éléments *cis*-régulateurs par CRISPR/Cas9. La technologie CRISPR/Cas9 permet d'éditer l'ADN génomique de manière ciblée grâce à un petit fragment d'ARN guide. L'endonuclease Cas9 reconnaît le complexe ARN guide/ADN et coupe le double brin d'ADN, déclenchant ainsi dans la cellule les systèmes de réparation de l'ADN. L'utilisation d'une Cas9 ayant perdu son activité d'endonuclease (dCas9) permet de cibler l'ADN sans le couper. Ainsi, la technologie CRISPR peut être utilisée pour étudier les séquences régulatrices dans leur contexte naturel par délétion d'une séquence régulatrice *via* l'utilisation de deux ARN guides et réparation de l'ADN par jonction d'extrémités non homologues (I) ou par remplacement d'une séquence par sa copie mutée ajoutée grâce au système de réparation par recombinaison homologues (II), modifications épigénétiques de l'ADN grâce à un modulateur épigénétique (activateur ou répresseur) couplé à la dCas9 (III), identification d'éléments régulateurs potentiels d'un gène d'intérêt en passant au crible les régions génomiques voisines par CRISPR/(d)Cas9 (IV).

qui ont effectué des cribles d'éléments *cis*-régulateurs par des approches CRISPR/Cas9 ont montré que l'expression de certains gènes est contrôlée par des promoteurs de gènes distaux (Rajagopal *et al.*, 2016 ; Diao *et al.*, 2017). Dans l'ensemble, ces résultats ouvrent une possibilité intéressante : les variants associés à des caractères ou des maladies situés dans un sous-ensemble de promoteurs pourraient directement impacter l'expression de gènes distaux (Figure 2).

Conclusions

La mise en œuvre de tests de gènes rapporteurs à haut débit et de criblages basés sur le CRISPR permet la validation expérimentale de l'activité des *enhancers* dans différents types de cellules et de contextes cellulaires. Ces essais sont maintenant suffisamment robustes et sensibles

pour être largement utilisés par les chercheurs qui s'intéressent à la régulation des gènes. Ces approches ont également donné lieu à des découvertes imprévues, comme le rôle de certains promoteurs en tant que régulateurs de type *enhancer*. L'une des principales limites de ces approches réside dans le fait qu'elles ne fournissent pas d'informations directes sur la découverte des gènes ciblés par les *enhancers*. Par conséquent, la combinaison de tests d'*enhancers* avec les méthodologies 3C récemment développées, telles que le 4C-seq, Hi-C ou capture Hi-C (Sati & Cavalli, 2017), devrait considérablement faciliter l'assignation des *enhancers* découverts à leurs gènes cibles présumés. Enfin, avec la baisse attendue du coût du séquençage et de la synthèse massive d'oligonucléotides, il sera possible de tester systématiquement l'impact des variants dans les régions régulatrices dans différents contextes pathologiques et développementaux.

Remerciements. Le travail au sein du laboratoire TAGC est soutenu par des financements récurrents de la part de l'Inserm et d'Aix-Marseille Université (AMU). Le projet menant à cette publication a reçu des subventions de l'ARC (PJA 20151203149), du Plan Cancer (P036496), de l'Initiative d'Excellence AMU (A*Midex), du programme français « Investissements d'Avenir » et de la Ligue Contre le Cancer (Équipe Labellisée). A.E. et D.S.G. perçoivent respectivement une bourse du Ministère de l'Enseignement Supérieur et de la Recherche d'une part, et du Conacyt (Mexique), d'autre part.

Références

- Akhtar-Zaidi, B., Cowper-Sal-lari, R., Corradin, O., Saiakhova, A., Bartels, C.F., Balasubramanian, D., Myeroff, L., Lutterbaugh, J., Jarrar, A., Kalady, M.F., Willis J., Moore J.H., Tesar P.J., Laframboise T., Markowitz S., Lupini M., Scacheri P.C. (2012). Epigenomic enhancer profiling defines a signature of colon cancer. *Science*, 336, 736-739.
- Andersson, R. (2015). Promoter or enhancer, what's the difference? Deconstruction of established distinctions and presentation of a unifying model. *BioEssays*, 37, 314-323.
- Arnold, C.D., Gerlach, D., Stelzer, C., Boryn, M., Rath, M., Stark, A. (2013). Genome-wide quantitative enhancer activity maps identified by STARR-seq. *Science*, 339, 1074-1077.
- Arnold, C.D., Gerlach, D., Spies, D., Matts, J.A., Sytnikova, Y. A., Pagani, M., Lau, N.C., Stark, A. (2014). Quantitative genome-wide enhancer activity maps for five Drosophila species show functional enhancer conservation and turnover during *cis*-regulatory evolution. *Nat Genet*, 46, 685-692.
- Barakat, T.S., Halbritter, F., Zhang, M., Rendeiro, A.F., Bock, C., Chambers, I. (2017). Functional dissection of the enhancer repertoire in human embryonic stem cells. *BioRxiv*, 146696.
- Bevington, S.L., Cauchy, P., Cockerill, P.N. (2017). Chromatin priming elements establish immunological memory in T cells without activating transcription. *BioEssays*, 39(2), 1600184.
- Canver, M.C., Smith, E.C., Sher, F., Pinello, L., Sanjana, N.E., Shalem, O., Chen, D.D., Schupp, P.G., Vinjamur, D.S., Garcia, S.P., Luc S., Kurita R., Nakamura Y., Fujiwara Y., Maeda T., Yuan G.C., Zhang F., Orkin S.H., Bauer D.E. (2015). BCL11A enhancer dissection by Cas9-mediated *in situ* saturating mutagenesis. *Nature*, 527, 192-197.
- Chatterjee, S., Ahituv, N. (2017). Gene regulatory elements, major drivers of human disease. *Annu Rev Genomics Hum Genet*, 18, 45-63.
- Core, L.J., Martins, A.L., Danko, C.G., Waters, C., Siepel, A., Lis, J.T. (2014). Analysis of nascent RNA identifies a unified architecture of initiation regions at mammalian promoters and enhancers. *Nat Genet*, 46, 1311-1320.
- Dailey, L. (2015). High throughput technologies for the functional discovery of mammalian enhancers: New approaches for understanding transcriptional regulatory network dynamics. *Genomics*, 106, 151-158.
- Dao, L.T.M., Galindo-Albarrán, A.O., Castro-Mondragon, J.A., Andrieu-Soler, C., Medina-Rivera, A., Souaid, C., Charbonnier, G., Griffon, A., Vanhille, L., Stephen, T., Alomairi J., Martin D., Torres M., Fernandez N., Soler E., van Helden J., Puthier D., Spicuglia S. (2017). Genome-wide characterization of mammalian promoters with distal enhancer functions. *Nat Genet*, 49, 1073-1081
- Diao, Y., Li, B., Meng, Z., Jung, I., Lee, A.Y., Dixon, J., Maliskova, L., Guan, K., Shen, Y., Ren, B. (2016). A new class of temporarily phenotypic enhancers identified by CRISPR/Cas9-mediated genetic screening. *Genome Res*, 26, 397-405.
- Diao, Y., Fang, R., Li, B., Meng, Z., Yu, J., Qiu, Y., Lin, K.C., Huang, H., Liu, T., Marina, R.J., Jung I., Shen Y., Guan K.L., Ren B. (2017). A tiling-deletion-based genetic screen for *cis*-regulatory element identification in mammalian cells. *Nat Methods*, 14, 629.
- Dogan, N., Wu, W., Morrissey, C.S., Chen, K.-B., Stonestrom, A., Long, M., Keller, C.A., Cheng, Y., Jain, D., Visel, A., Pennacchio L.A., Weiss M.J., Blobel G.A., Hardison R.C. (2015). Occupancy by key transcription factors is a more accurate predictor of enhancer activity than histone modifications or chromatin accessibility. *Epigenetics Chromatin*, 8.
- Engreitz, J.M., Haines, J.E., Perez, E.M., Munson, G., Chen, J., Kane, M., McDonel, P.E., Guttman, M., Lander, E.S. (2016). Local regulation of gene expression by lncRNA promoters, transcription and splicing. *Nature*, 539, 452.
- Ernst, J., Melnikov, A., Zhang, X., Wang, L., Rogov, P., Mikkelsen, T.S., Kellis, M. (2016). Genome-scale high-resolution mapping of activating and repressive nucleotides in regulatory regions. *Nat Biotechnol*, 34, 1180-1190.
- Fulco, C.P., Munschauer, M., Anyoha, R., Munson, G., Grossman, S.R., Perez, E.M., Kane, M., Cleary, B., Lander, E.S., Engreitz, J.M. (2016). Systematic mapping of functional enhancer-promoter connections with CRISPR interference. *Science*, 354, 769-773.
- Hsu, P.D., Lander, E.S., Zhang, F. (2014). Development and applications of CRISPR-Cas9 for genome engineering. *Cell*, 157, 1262-1278.
- Inoue, F., Ahituv, N. (2015). Decoding enhancers using massively parallel reporter assays. *Genomics*, 106, 159-164.
- Inoue, F., Kircher, M., Martin, B., Cooper, G.M., Witten, D.M., McManus, M.T., Ahituv, N., Shendure, J. (2017). A systematic comparison reveals substantial differences in chromosomal *versus* episomal encoding of enhancer activity. *Genome Res*, 27, 38-52.
- Kheradpour, P., Ernst, J., Melnikov, A., Rogov, P., Wang, L., Zhang, X., Alston, J., Mikkelsen, T.S., Kellis, M. (2013). Systematic dissection of regulatory motifs in 2000 predicted human enhancers using a massively parallel reporter assay. *Genome Res*, 23, 800-811.
- Kim, T.-K., Shiekhattar, R. (2015). Architectural and functional commonalities between enhancers and promoters. *Cell*, 162, 948-959.
- Klann, T.S., Black, J.B., Chellappan, M., Safi, A., Song, L., Hilton, I.B., Crawford, G.E., Reddy, T.E., Gersbach, C.A. (2017). CRISPR-Cas9 epigenome editing enables high-throughput screening for functional regulatory elements in the human genome. *Nat Biotechnol*, 35, 561-568.
- Koch, F., Fenouil, R., Gut, M., Cauchy, P., Albert, T.K., Zacarias-Cabeza, J., Spicuglia, S., de la Chapelle, A.L., Heidemann, M., Hintermair, C., Eick D., Gut I., Ferrier P., Andrau J.C. (2011). Transcription initiation platforms and GTF recruitment at tissue-specific enhancers and promoters. *Nat Struct Mol Biol*, 18, 956.
- Korkmaz, G., Lopes, R., Ugalde, A.P., Nevedomskaya, E., Han, R., Myacheva, K., Zwart, W., Elkon, R., Agami, R. (2016). Functional genetic screens for enhancer elements in the human genome using CRISPR-Cas9. *Nat Biotechnol*, 34, 192.
- Kwasnieski, J.C., Fiore, C., Chaudhari, H.G., Cohen, B.A. (2014). High-throughput functional testing of ENCODE segmentation predictions. *Genome Res*, 24, 1595-1602.
- Liu, Y., Yu, S., Dhiman, V.K., Brunetti, T., Eckart, H., White, K.P. (2017). Functional assessment of human enhancer activities using whole-genome STARR-sequencing. *Genome Biol*, 18, 219.
- Lo, A., Qi, L. (2017). Genetic and epigenetic control of gene expression by CRISPR-Cas systems. *F1000 Research*, 6.

- Medina-Rivera, A., Santiago-Algarra, D., Puthier, D., Spicuglia, S. (2018). Widespread enhancer activity from core promoters. *Trends Biochem Sci*, 43, 452-468.
- Melnikov, A., Murugan, A., Zhang, X., Tesileanu, T., Wang, L., Rogov, P., Feizi, S., Gnirke, A., Callan, C.G., Kinney, J.B., Kellis M., Lander E.S., Mikkelsen T.S. (2012). Rapid dissection and model-based optimization of inducible enhancers in human cells using a massively parallel reporter assay. *Nat Biotechnol*, 30, 271-277.
- Montalbano, A., Canver, M.C., Sanjana, N.E. (2017). High-throughput approaches to pinpoint function within the noncoding genome. *Mol Cell*, 68, 44-59.
- Muerdter, F., Boryn, M., Arnold, C.D. (2015). STARR-seq – principles and applications. *Genomics*, 106, 145-150.
- Muerdter, F., Boryn, M., Woodfin, A.R., Neumayr, C., Rath, M., Zabidi, M.A., Pagani, M., Hablerle, V., Kazmar, T., Catarino, R.R., Schernhuber K., Arnold C.D., Stark A. (2018). Resolving systematic errors in widely used enhancer activity assays in human cells. *Nat Methods*, 15, 141-149.
- Murtha, M., Tokcaer-Keskin, Z., Tang, Z., Strino, F., Chen, X., Wang, Y., Xi, X., Basilico, C., Brown, S., Bonneau, R., Kluger Y., Dailey L. (2014). FIREWACH: High-throughput functional detection of transcriptional regulatory modules in mammalian cells. *Nat Methods*, 11, 559-565.
- Nguyen, T.A., Jones, R.D., Snavely, A.R., Pfenning, A.R., Kirchner, R., Hemberg, M., Gray, J.M. (2016). High-throughput functional comparison of promoter and enhancer activities. *Genome Res*, 26, 1023-1033.
- Patwardhan, R.P., Hiatt, J.B., Witten, D.M., Kim, M.J., Smith, R.P., May, D., Lee, C., Andrie, J.M., Lee, S.-I., Cooper, G.M., Ahituv N., Pennacchio L.A., Shendure J. (2012). Massively parallel functional dissection of mammalian enhancers *in vivo*. *Nat Biotechnol*, 30, 265-270.
- Pekowska, A., Benoukraf, T., Zacarias-Cabeza, J., Belhocine, M., Koch, F., Holota, H., Imbert, J., Andrau, J.-C., Ferrier, P., Spicuglia, S. (2011). H3K4 tri-methylation provides an epigenetic signature of active enhancers. *EMBO J*, 30, 4198-4210.
- Rajagopal, N., Srinivasan, S., Kooshesh, K., Guo, Y., Edwards, M.D., Banerjee, B., Syed, T., Emons, B.J., Gifford, D.K., Sherwood, R.I. (2016). High-throughput mapping of regulatory DNA. *Nat Biotechnol*, 34, 167-174.
- Ruf, S., Symmons, O., Uslu, V.V., Dolle, D., Hot, C., Ettwiller, L., Spitz, F. (2011). Large-scale analysis of the regulatory architecture of the mouse genome with a transposon-associated sensor. *Nat Genet*, 43, 379.
- Sanjana, N.E., Wright, J., Zheng, K., Shalem, O., Fontanillas, P., Joung, J., Cheng, C., Regev, A., Zhang, F. (2016). High-resolution interrogation of functional elements in the non-coding genome. *Science*, 353, 1545-1549.
- Sati, S., Cavalli, G. (2017). Chromosome conformation capture technologies and their impact in understanding genome function. *Chromosoma*, 126, 33-44.
- Shen, S.Q., Myers, C.A., Hughes, A.E.O., Byrne, L.C., Flannery, J.G., Corbo, J.C. (2016). Massively parallel *cis*-regulatory analysis in the mammalian central nervous system. *Genome Res*, 26, 238-255.
- Shlyueva, D., Stelzer, C., Gerlach, D., Yáñez-Cuna, J.O., Rath, M., Boryn, M., Arnold, C.D., Stark, A. (2014). Hormone-responsive enhancer-activity maps reveal predictive motifs, indirect repression, and targeting of closed chromatin. *Mol Cell*, 54, 180-192.
- Smith, R.P., Taher, L., Patwardhan, R.P., Kim, M.J., Inoue, F., Shendure, J., Ovcharenko, I., Ahituv, N. (2013). Massively parallel decoding of mammalian regulatory sequences supports a flexible organizational model. *Nat Genet*, 45, 1021-1028.
- Symmons, O., Uslu, V.V., Tsujimura, T., Ruf, S., Nassari, S., Schwarzer, W., Ettwiller, L., Spitz, F. (2014). Functional and topological characteristics of mammalian regulatory domains. *Genome Res*, 24, 390-400.
- Tewhey, R., Kotliar, D., Park, D.S., Liu, B., Winnicki, S., Reilly, S.K., Andersen, K.G., Mikkelsen, T.S., Lander, E.S., Schaffner, S.F., Sabeti P.C. (2016). Direct identification of hundreds of expression-modulating variants using a multiplexed reporter assay. *Cell*, 165, 1519-1529.
- Ulirsch, J.C., Nandakumar, S.K., Wang, L., Giani, F.C., Zhang, X., Rogov, P., Melnikov, A., McDonel, P., Do, R., Mikkelsen, T.S., Sankaran V.G. (2016). Systematic functional dissection of common genetic variation affecting red blood cell traits. *Cell*, 165, 1530-1545.
- Vanhille, L., Griffon, A., Maqbool, M.A., Zacarias-Cabeza, J., Dao, L.T.M., Fernandez, N., Ballester, B., Andrau, J.C., Spicuglia, S. (2015). High-throughput and quantitative assessment of enhancer activity in mammals by CapStarr-seq. *Nat Commun*, 6, 6905.
- Vernimmen, D., Bickmore, W.A. (2015). The hierarchy of transcriptional activation: From enhancer to promoter. *Trends Genet TIG*, 31, 696-708.
- Vierstra, J., Reik, A., Chang, K.-H., Stehling-Sun, S., Zhou, Y.-Y., Hinkley, S.J., Paschon, D.E., Zhang, L., Psatha, N., Bendana, Y.R., O'Neil C.M., Song A.H., Mich A.K., Liu P.Q., Lee G., Bauer D.E., Holmes M.C., Orkin S.H., Papayannopoulou T., Stamatoyannopoulos G., Rebar E.J., Gregory P.D., Urnov F.D., Stamatoyannopoulos J.A. (2015). Functional footprinting of regulatory DNA. *Nat Methods*, 12, 927-930.
- Vockley, C.M., Guo, C., Majoros, W.H., Nodzenski, M., Scholtens, D.M., Hayes, M.G., Lowe, W.L., Reddy, T.E. (2015). Massively parallel quantification of the regulatory effects of noncoding genetic variation in a human cohort. *Genome Res*, 25, 1206-1214.
- Wang, X., He, L., Goggin, S., Saadat, A., Wang, L., Claussnitzer, M., Kellis, M. (2017). High-resolution genome-wide functional dissection of transcriptional regulatory regions in human. *BioRxiv*, 193136.
- White, M.A. (2015). Understanding how *cis*-regulatory function is encoded in DNA sequence using massively parallel reporter assays and designed sequences. *Genomics*, 106, 165-170.
- Xie, S., Duan, J., Li, B., Zhou, P., Hon, G.C. (2017). Multiplexed engineering and analysis of combinatorial enhancer activity in single cells. *Mol Cell*, 66, 285-299.e5.
- Yáñez-Cuna, J.O., Arnold, C.D., Stampfel, G., Boryn, M., Gerlach, D., Rath, M., Stark, A. (2014). Dissection of thousands of cell type-specific enhancers identifies dinucleotide repeat motifs as general enhancer features. *Genome Res*, 24, 1147-1156.
- Zabidi, M.A., Arnold, C.D., Schernhuber, K., Pagani, M., Rath, M., Frank, O., Stark, A. (2015). Enhancer-core-promoter specificity separates developmental and housekeeping gene regulation. *Nature*, 518, 556-559.