

Cis-ruptions d'éléments génomiques hautement conservés non codants à distance du gène *SOX9* dans la séquence de Pierre Robin

Sabina Benko, Christopher T. Gordon, Jeanne Amiel et Stanislas Lyonnet

Département de Génétique, Université Paris Descartes et Inserm U-781, Hôpital Necker-Enfants Malades AP-HP, 149 rue de Sèvres, 75743 Paris Cedex 15, France

Auteur correspondant : Stanislas Lyonnet, stanislas.lyonnet@inserm.fr

Reçu le 10 février 2011

Résumé – Les gènes majeurs du développement présentant des patrons d'expression complexes sont souvent entourés d'un désert génique particulièrement riche en régions non codantes hautement conservées au cours de l'évolution. L'expression de ces gènes dans de nombreux territoires au cours du développement nécessite une régulation coordonnée dans le temps et l'espace, orchestrée par certaines de ces régions conservées non codantes agissant comme des éléments régulateurs de transcription. *SOX9* est un gène essentiel à de nombreux processus développementaux, comme la chondrogénèse, le développement des cellules de crêtes neurales et le développement testiculaire. En accord avec ces principaux territoires d'expression, l'haplo-insuffisance de *SOX9* liée à des altérations de sa séquence codante conduit à un syndrome polymalformatif – la dysplasie campomélique – dont les signes majeurs sont une atteinte osseuse, une séquence de Pierre Robin (SPR) et une anomalie de différenciation sexuelle (*Disorder of Sex Development*, DSD). *SOX9* est localisé dans un désert génique de ~2,5 Mb extrêmement riche en séquences conservées. Nous avons utilisé le locus *SOX9* et la dysplasie campomélique comme modèle d'étude pour montrer qu'un ou plusieurs endophénotypes d'un syndrome complexe peuvent résulter d'une dérégulation tissu-spécifique de la transcription d'un gène majeur du développement. Notre travail a porté attention à l'un de ces endophénotypes, la SPR, caractérisée par la triade micro-et/ou rétrognathie, glossoptose et fente palatine). Nous détaillons ici l'identification des altérations (translocations, délétions, mutations ponctuelles) des régions non codantes localisées à très grande distance (plus de 1,2 Mb) en amont et en aval de *SOX9* dans des cas de SPR isolée familiale ou sporadique. Nos études *in vitro* et *in vivo* ont indiqué que ces altérations concernent des éléments conservés au cours de l'évolution et ayant une fonction régulatrice d'expression tissu-spécifique, notamment dans le mésenchyme mandibulaire. Ces données permettent de proposer un modèle moléculaire expliquant une présentation clinique d'un syndrome restreint à un seul des endophénotypes qui le composent par invalidation ciblée tissu-spécifique de l'expression d'un gène majeur du développement.

Mots clés : Éléments régulateurs / régulation transcriptionnelle tissu-spécifique / endophénotype / séquence de Pierre Robin / dysplasie campomélique

Abstract – *Cis*-ruptions of highly conserved non-coding genomic elements distant from the *SOX9* gene in the Pierre Robin sequence.

Major developmental genes, exhibiting complex expression patterns, are often embedded within a genic desert particularly rich in regions, which though non-coding are highly conserved. The developmental expression of these genes in many areas requires coordinated regulation in time and space, which is orchestrated by some of these conserved non-coding regions, acting as transcriptional regulators. *SOX9* is an

essential gene for many developmental processes, such as chondrogenesis, migration and differentiation of neural crest cells and testis development. In agreement with these major expression areas, *SOX9* haploinsufficiency, linked to alterations in coding sequence, leads to a polymorphic malformation syndrome – campomelic dysplasia – whose major symptoms are a bone anomaly, a Pierre Robin sequence, and a sexual differentiation anomaly (Disorder of Sex Development, DSD). *SOX9* is located in a ~2.5 Mb gene desert extremely rich in conserved sequences. We have used the *SOX9* locus and campomelic dysplasia as a model to show that one or several endophenotypes within a complex syndrome may arise from a tissue-specific deregulation of a major developmental gene transcription. Our work has focused on one of these endophenotypes, SPR, characterized by the triad micro- and/or retrognathia, glossoptosis and cleft palate. Here we report in detail how we identified alterations (translocations, deletions, point mutations) in non-coding regions, located far away (more than 1.2 Mb) upstream and downstream of *SOX9*, in clustered or sporadic SPR cases.

Key words: Regulation elements / tissue-specific regulating transcription / endophenotype / Pierre Robin's sequence / campomelic dysplasia

Abréviations :

CCN, Cellules de la crête neurale
 ChIP, *Chromatin Immuno-Precipitation*
 CGH, *Comparative Genomic Hybridization*
 CNV, *Copy Number Variation*
 DC, Dysplasie Campomélique
 DSD, *Disorder of Sex Development*
 HCNE, *Highly Conserved Non-coding Element*
 MEPA, *Mouse Embryonic Pharyngeal Arch*
 ou Arc Pharyngé Embryonnaire Murin
 QTL, *Quantitative Trait Loci*
 SNP, *Single Nucleotide Polymorphism*
SOX9, *SRY-related HMG-box gene 9*
 SPR, Séquence de Pierre Robin

Éléments régulateurs de transcription en cis

Les gènes du développement des organismes complexes sont accompagnés d'une multitude de séquences d'ADN non codantes régulatrices, d'une importance capitale dans le contrôle de leur expression tissu- et stade-spécifique au cours du développement ; il s'agit d'éléments régulateurs en *cis* qui peuvent avoir une fonction d'éléments activateurs (*enhancer*) ou inhibiteurs de transcription (*silencer*). Un même élément régulateur peut pourtant jouer un rôle d'activateur ou d'inhibiteur de transcription en fonction du contexte cellulaire et des différents facteurs de transcription présents (Uhlenhaut *et al.*, 2009). Ces éléments régulateurs sont des séquences nucléotidiques relativement courtes, généralement d'une taille de 100–500 pb, et peuvent être localisés dans les introns ou bien à une distance plus ou moins grande, en 5' et/ou en 3' du gène cible qu'ils régulent. En fait, ils sont souvent dispersés sur de larges séquences génomiques

pouvant s'étendre à plus de 1 Mb de leur promoteur cible (Lettice *et al.*, 2002 ; Sagai *et al.*, 2005). Ils comprennent souvent de multiples séquences spécifiques de fixation de facteurs de transcription.

Un élément activateur régule l'expression génique de façon tissu- et stade-spécifique dans une partie plus ou moins étendue du territoire d'expression du gène cible. Un exemple est celui des éléments activateurs dénommés E1–E7, localisés dans le domaine de régulation du gène *SOX9*, qui sont capables de réguler l'expression d'un gène rapporteur chez l'embryon de souris dans une fraction restreinte du patron d'expression du gène *SOX9* : E1 dirige l'expression dans le noeud embryonnaire, la notochorde, l'intestin et le pancréas ; E3 dirige l'expression dans les cellules de crête neurale du premier arc branchial et dans l'oreille interne ; E7 dirige l'expression dans le télencéphale et le mésencéphale (Bagheri-Fam *et al.*, 2006). Au-delà de l'étude des éléments activateurs sélectionnés par des essais de gène rapporteur, la tissu-spécificité des éléments activateurs est aussi suggérée par des études comparatives de sites hypersensibles à la DNaseI et par des patrons de modification des histones caractéristiques de l'état de ces éléments dans plusieurs types cellulaires (Xi *et al.*, 2007 ; Heintzman *et al.*, 2009).

Une étude récente a aussi montré que les éléments activateurs de transcription des vertébrés sont fortement modulaires, autonomes et additifs dans leur effets (Visel *et al.*, 2009a). Dans cette étude, Visel *et al.* ont comparé l'activité *in vivo* d'éléments combinés par rapport à des éléments isolés. Différentes combinaisons de deux ou quatre éléments activateurs sur un total de six ont été testées par essai de gène rapporteur chez l'embryon de souris. Les éléments activateurs testés appartenaient aux domaines de régulation de différents gènes fonctionnellement indépendants,

localisés sur différents chromosomes ou espacés de plusieurs mégabases. Les éléments activateurs conservaient une spécificité d'expression spatiale et temporelle quelle que soit la combinaison testée. L'expression du gène rapporteur représentait ainsi la superposition de patrons individuels, n'interférant pas entre eux. Ces résultats montrent l'indépendance fonctionnelle et l'additivité spatio-temporelle des éléments activateurs chez les vertébrés, et suggèrent que la somme de leurs actions individuelles tissu- et stade-spécifiques récapitule le patron d'expression complet d'un gène ainsi régulé. De plus, ces résultats vont dans le sens de l'indépendance fonctionnelle relative entre éléments régulateurs et de l'absence d'interférence entre eux. Ce modèle offre une plasticité maximale dans la mise en place de patrons d'expression complexes au cours de l'évolution.

Les éléments régulateurs activateurs de transcription représentent un groupe important des éléments non codants fonctionnels. Ainsi, les études du génome entier sur deux différentes lignées cellulaires ont identifié environ 55 000 séquences portant une signature chromatinienne (monométhylation H3K4) indicatrice de probables éléments activateurs (Heintzman *et al.*, 2009). Dans la majorité des cas, la signature est spécifique d'un type cellulaire. Par extrapolation, prenant en compte la variabilité des types cellulaires chez les mammifères, on estime que le génome humain comporte de l'ordre de 10^5 à 10^6 éléments activateurs avec un espacement moyen de 3000 à 30 000 pb (Heintzman *et al.*, 2009). Le rôle de ces éléments est déjà établi dans le développement (Stathopoulos & Levine, 2005), la diversité phénotypique (Levine & Tjian, 2003) et l'origine de maladies génétiques humaines (Lettice *et al.*, 2003; Emison *et al.*, 2005; Kleinjan & van Heyningen, 2005).

Identification des éléments régulateurs en *cis*

À ce jour nous ne connaissons ni les « codes » ni la « grammaire » qui sous-tendent la fonctionnalité des éléments régulateurs en *cis*. Il est donc toujours difficile de discerner les séquences non codantes non fonctionnelles et celles qui sont impliquées dans la régulation transcriptionnelle.

Pourtant, l'analyse génomique comparative des génomes humain et murin a démontré qu'au cours de l'évolution 5 % du génome humain était sous pression de sélection. Parmi ces 5 %, seul ~1 % code des protéines (Waterston *et al.*, 2002) et environ 1 % code les ARNs de structure ou régulateurs (Mattick & Makunin, 2006; Pheasant & Mattick, 2007). Les ~3 % restants, en séquences conservées entre Homme et souris, sont donc des séquences non codantes. Donc, de même que les séquences codantes pour les protéines, les séquences conservées non codantes ont été soumises

au cours de l'évolution à une pression de sélection négative. Cela suggère fortement que ces séquences ont un rôle fonctionnel (Waterston *et al.*, 2002; Nobrega *et al.*, 2003; Pennacchio *et al.*, 2006). De plus, la distribution des éléments non codants conservés n'est pas aléatoire; leur concentration est 4,5 fois plus élevée dans les déserts géniques entourant des gènes du développement (Nobrega *et al.*, 2003; Siepel *et al.*, 2005; Woolfe *et al.*, 2005), suggérant qu'ils pourraient avoir un rôle dans le développement des organismes complexes.

Plusieurs fonctions ont été proposées pour les éléments non codants conservés au cours de l'évolution. Ils pourraient agir comme des éléments régulateurs activateurs ou inhibiteurs de transcription, avoir un rôle structural, être impliqués dans le remodelage de la chromatine, avoir un lien avec la transcription pervasive du génome, réguler l'épissage alternatif, etc. (Dermitzakis *et al.*, 2005). À ce jour, la plupart des études expérimentales sur la fonction des séquences non codantes conservées se sont pourtant focalisées sur leur potentiel de régulation de l'expression génique en *cis* par activation de la transcription. Pour une fraction des éléments non codants, il a été démontré *in vivo* chez l'embryon de poisson-zèbre ou de souris, qu'ils fonctionnent bien comme des éléments activateurs de transcription ayant un rôle dans la régulation de l'expression stade- et tissu-spécifique des gènes du développement (Nobrega *et al.*, 2003; Woolfe *et al.*, 2005; Pennacchio *et al.*, 2006). Nous pouvons ainsi considérer ces régions comme candidates au criblage par séquençage à haut débit dans des anomalies du développement chez l'Homme.

D'autre part, avec une meilleure connaissance des signatures chromatinienne de diverses classes d'éléments régulateurs et l'introduction de technologies permettant des études du génome entier, les études expérimentales de cartographie des éléments régulateurs par immunoprécipitation de chromatine (ChIP) se montrent de plus en plus efficaces. Par exemple, une expérience de ChIP-seq contre le cofacteur p300 permet d'identifier les éléments régulateurs activateurs d'expression tissu-spécifique avec une efficacité de près de 80 % dans chaque tissu analysé chez l'embryon de souris : le diencephale, le prosencéphale, les bourgeons de membres et le coeur murin au stade E11.5 du développement (Visel *et al.*, 2009b; Blow *et al.*, 2010). De manière intéressante, l'une de ces études menée sur le coeur embryonnaire murin a révélé que la majorité de ces éléments ne montre pas une conservation extrême. De plus, environ 30 % des activateurs de transcription ne sont pas conservés au cours de l'évolution, alors que des essais de gène rapporteur *in vivo* confirment leur activité régulatrice spécifique d'un tissu. Ces résultats sont en accord avec les études du projet ENCODE dont les expériences

de ChIP, visant à identifier différentes modifications de la chromatine ou localiser des sites de fixation de différents facteurs de transcription, ont révélé que la séquence d'une grande partie des éléments détectés est faiblement ou voire non conservée au cours de l'évolution (Consortium, 2007). Cela suggère qu'une partie des éléments régulateurs ne peut pas être identifiée sur la base de la conservation de séquence, ou que les séquences conservées au sein de ces éléments sont trop petites et vont échapper à la détection par les algorithmes couramment utilisés et ce d'autant plus que les critères de sélection appliqués sont stringents (Elgar & Vavouri, 2008; Heintzman & Ren, 2009).

Identification de(s) gène(s) cibles d'un élément régulateur altéré dans une maladie génétique

Identifier le (ou les) gène(s) cible(s) d'un élément régulateur en *cis* n'est pas évident et peut exiger des expériences laborieuses. En effet, nous avons vu que les éléments régulateurs en *cis* peuvent être localisés à très grande distance par rapport au(x) promoteur(s) qu'ils contrôlent. De plus, le gène et ses éléments régulateurs ne sont pas nécessairement organisés comme une unité continue; ils peuvent dépasser les « frontières » représentées par des gènes avoisinants (Lettice *et al.*, 2002, 2003; Kleinjan & Lettice, 2008), voire être localisés sur d'autres chromosomes (Lomvardas *et al.*, 2006; Fuss *et al.*, 2007). Enfin, un élément régulateur peut moduler l'expression de plusieurs gènes (Spitz & Duboule, 2008). Ces arguments remettent en question la définition présentant un gène comme une unité d'hérédité et suggèrent le concept nouveau de module fonctionnel d'expression, englobant la région codante transcrite, les éléments régulateurs en *cis* de sa transcription et la structure chromatinienne locale (Dillon & Sabbattini, 2000; Dillon, 2003; de Laat & Grosveld, 2003). La clef de l'autonomie fonctionnelle d'une telle unité par rapport aux gènes adjacents réside dans la spécificité de l'interaction entre les éléments régulateurs et leur(s) promoteur(s) cible(s).

Les arguments suivants peuvent être utilisés pour évoquer (indirectement) le gène (ou les gènes) cible(s) d'un élément régulateur distant :

- le patron d'expression du gène cible candidat ;
- le phénotype de la maladie causée par des mutations de la séquence codante du gène cible identique ou incluant le phénotype étudié comme endophénotype ;
- la démonstration, par essai de gène rapporteur, que l'élément concerné par les altérations de séquences non codantes identifiées chez les patients active l'expression dans le tissu touché au cours du développement des cas étudiés et compris dans le

territoire complet d'expression du gène cible candidat ;

- l'identité des facteurs de transcription interagissant avec l'élément régulateur touché, leur rapport aux voies moléculaires et aux événements du développement prédits comme dérégulés dans la maladie étudiée et leur relation avec le gène cible candidat ;
- la démonstration que l'élément régulateur et le gène cible font partie du même bloc de synténie, conservé au cours de l'évolution ;
- l'organisation spatiale des fibres de chromatine dans le noyau en interphase qui varie en fonction du niveau d'expression du gène cible (pouvant être étudié par FISH interphasique) ;
- les interactions physiques directes entre l'élément régulateur et le promoteur du gène cible candidat identifiées par méthodes de capture de conformation de la chromatine (3C et ses dérivés) (Crutchley *et al.*, 2010).

Mais la complexité de ces nouveaux mécanismes de pathologie moléculaire va bien plus loin. Ainsi, les mutations d'éléments régulateurs d'un même gène peuvent entraîner des phénotypes différents. C'est le cas au locus *Shh* où la dérégulation de l'expression de *Shh* par des réarrangements proches du gène *Shh* cause une holoprosencéphalie, alors que la perte d'éléments régulateurs plus distaux et régulant l'expression dans les bourgeons de membre conduit à des anomalies du développement des membres (Kleinjan & Lettice, 2008). De plus, un élément régulateur peut acquérir par mutation une nouvelle fonction causant un phénotype nouveau non inclus dans le spectre des anomalies par mutations de la séquence codante du gène cible. Par exemple, les mutations ponctuelles de l'élément régulateur de l'expression de *Shh* dans les bourgeons de membre induisent une polydactylie préaxiale chez la souris et chez l'Homme (Lettice *et al.*, 2002, 2003). Ce phénotype est donc différent de celui qui est provoqué par les délétions de ce même élément chez la souris, qui présente alors des membres sévèrement raccourcis, et n'appartient pas au spectre des anomalies associées aux mutations de la séquence codante de *Shh* (Sagai *et al.*, 2005; Kleinjan & Lettice, 2008). Un dernier facteur de complexité tient au fait qu'un élément régulateur peut contrôler la transcription de plusieurs gènes (Spitz *et al.*, 2003).

La preuve biologique la plus formelle qu'un élément régulateur est impliqué dans le contrôle d'un gène donné au cours du développement serait l'invalidation de cet élément chez un modèle animal par mutation ponctuelle ou délétion. La cible *in vivo* a ainsi été démontrée par exemple pour des éléments régulateurs aux loci de *Emx2* (Suda *et al.*, 2010) et *Shh* (Sagai *et al.*, 2005), par analyse de l'expression génique dans le tissu d'intérêt chez les souris transgéniques.

Cependant, ce type d'étude comporte un risque important de résultat négatif. En effet, les souris transgéniques homozygotes pour des délétions emportant de grands segments non codants, rapportées par Nobrega *et al.* (2004), ainsi que les souris transgéniques homozygotes pour des délétions de séquences de quelques centaines de bases extrêmement conservées et présentant une activité régulatrice dans les essais de gène rapporteur *in vivo*, décrites par Ahituv *et al.* (2007), sont viables, fertiles et sans phénotype identifiable par rapport aux souris de phénotype sauvage des mêmes portées (Nobrega *et al.*, 2004; Ahituv *et al.*, 2007). Un tel résultat négatif peut s'expliquer par plusieurs phénomènes. Tout d'abord, un même gène du développement peut montrer une différence de sensibilité aux effets de dosage dans différents tissus chez l'Homme et la souris. Pour prendre l'exemple de *SOX9*, le développement mandibulaire paraît aussi sensible aux modifications de l'expression de *SOX9* chez l'Homme que chez la souris (Mori-Akiyama *et al.*, 2003); à l'inverse, le développement testiculaire est touché par l'haploinsuffisance de *SOX9* chez l'Homme alors qu'une réversion sexuelle XY n'est notée que pour une invalidation homozygote de *Sox9* chez la souris (Barrionuevo *et al.*, 2006). De plus, les gènes impliqués dans le développement peuvent être contrôlés par plusieurs éléments régulateurs ayant des fonctions redondantes. Cette redondance fonctionnelle entre éléments régulateurs semble jouer un rôle dans la robustesse phénotypique par rapport à la variabilité génétique et environnementale. Dans des conditions optimales, la perte d'un de ces éléments n'aura donc que des effets subtils ou non détectables sur le phénotype (Visel *et al.*, 2008; Frankel *et al.*, 2010). Enfin, même si sa séquence est hautement conservée au cours de l'évolution, un élément régulateur a pu acquérir une fonction nouvelle dans une espèce donnée en accumulant des variants au cours de l'évolution (Prabhakar *et al.*, 2008). L'invalidation d'un tel élément dans une autre espèce peut alors avoir des conséquences différentes, voire aucune conséquence sur le phénotype.

Maladies génétiques associées à une perturbation de la régulation transcriptionnelle à distance

Les mutations qui touchent l'intégrité des séquences codantes ou le nombre de copies d'un gène sont les causes moléculaires les plus souvent identifiées dans les maladies génétiques monogéniques. Cependant, le promoteur seul est insuffisant pour réguler l'expression spatio-temporelle complexe d'un gène du développement et nécessite un domaine de régulation plus ou moins vaste. Ce domaine comprend des éléments régulateurs en *cis* ayant une activité directe

dans le contrôle de la transcription et des éléments qui contribuent à l'organisation d'une structure appropriée de la chromatine, sans laquelle les premiers ne pourront pas agir. Il n'est donc pas étonnant que les réarrangements ou les mutations affectant les régions régulatrices d'un gène conduisent à une transcription aberrante ou défectueuse, ayant pour conséquence des maladies génétiques. La prévalence de tels effets parmi les maladies génétiques humaines est d'ailleurs probablement sous-estimée puisque les régions régulatrices en *cis* des gènes responsables de maladies génétiques sont encore loin d'être systématiquement analysées lors des études visant à identifier les variants délétères associés à un phénotype donné. Une des raisons est liée au fait que les bornes du domaine de régulation d'expression sont très difficiles à préciser, pouvant se situer à plus de 1 Mb de part et d'autre de l'unité transcriptionnelle du gène cible, au-delà des gènes avoisinants (Pfeifer *et al.*, 1999; Lettice *et al.*, 2002, 2003; Velagaleti *et al.*, 2005), dans les introns des gènes voisins comme au locus *Shh* (Kleinjan & Lettice, 2008), voire être localisées sur d'autres chromosomes (Lomvardas *et al.*, 2006).

Récemment, Kleinjan & Coutinho (2009) ont proposé le terme de *cis*-ruption pour décrire les différents mécanismes des maladies génétiques provoquées par un dysfonctionnement du système de régulation en *cis*, de l'architecture de la chromatine normale, ou du positionnement nucléaire du fragment chromosomique comprenant le gène dont l'expression est affectée. Ces *cis*-ruptions comprennent aussi les situations où le niveau d'expression génique est modifié par altération du contexte chromatinien alors que l'unité transcriptionnelle reste intacte, correspondant à la définition des effets de position (Kleinjan & Lettice, 2008). La liste des *cis*-ruptions associées à des maladies génétiques mono- ou multigéniques est déjà longue (pour revue voir Kleinjan & van Heyningen, 2005; Kleinjan & Lettice, 2008; Kleinjan & Coutinho, 2009).

Altérations de génome non codant à grande distance de *SOX9* dans la séquence de Pierre Robin

SOX9 et dysplasie campomélique

SOX9 (*SRY-related HMG-box gene 9*, OMIM 608160) est un gène majeur du développement codant un facteur de transcription. Il assume des fonctions essentielles au cours des différents stades de la chondrogénèse, de la différenciation des testicules et dans le développement des cellules de la crête neurale (CCN), une population de cellules transitoires et migratoires, qui, au niveau crânien, vont, entre autres, coloniser

les arcs branchiaux et former les éléments du squelette facial (Spokony *et al.*, 2002; Cheung *et al.*, 2005; McKeown *et al.*, 2005; Sakai *et al.*, 2006).

Les mutations des séquences codantes de *SOX9* sont la cause d'un syndrome polymalformatif du nouveau-né : la dysplasie campomélique (DC, OMIM 114290). La DC est un syndrome autosomique dominant causé par l'haploinsuffisance de *SOX9*. En accord avec le rôle de *SOX9*, ce syndrome est caractérisé par : une campomélie osseuse (incurvation des os longs), une hypoplasie des ailes iliaques et du pelvis, un thorax étroit par hypoplasie costale, 11 paires de côtes, une platyspondylie, une scoliose, des anomalies du développement sexuel chez les patients 46,XY (DSD 46,XY, *Disorder of Sex Development*; réversion sexuelle) et une séquence de Pierre Robin (SPR) (Houston *et al.*, 1983; Wagner *et al.*, 1994; Mansour *et al.*, 1995; Foster, 1996; Velagaleti *et al.*, 2005; Leipoldt *et al.*, 2007). La majorité des patients DC meurent dans la période néonatale par insuffisance respiratoire.

De plus, plusieurs cas de DC ont été rapportés, qui résultaient de réarrangements en 5' ou en 3' des séquences codantes de *SOX9* (délétions, translocations et inversions) (Foster *et al.*, 1994; Wagner *et al.*, 1994; Ninomiya *et al.*, 1996; Wirth *et al.*, 1996; Wunderle *et al.*, 1998; Pfeifer *et al.*, 1999; Pop *et al.*, 2004; Hill-Harfe *et al.*, 2005; Velagaleti *et al.*, 2005; Leipoldt *et al.*, 2007; Gordon *et al.*, 2009). Le phénotype des DC par translocation est parfois moins sévère que celui qui est lié à des mutations des régions codantes de *SOX9*, et la sévérité semble décroître avec la distance entre point de cassure de la translocation et séquences codantes (Leipoldt *et al.*, 2007). Ces observations incitent à penser que l'immense désert génique environnant *SOX9* comporte le domaine de régulation de son expression. Il a ainsi été proposé que ces translocations altèrent l'expression de *SOX9* par rupture de la continuité entre les séquences codantes de *SOX9* et les éléments régulateurs de son expression tissu- et stade-spécifique. Ainsi, les phénotypes observés seraient de sévérité variable en fonction du nombre d'éléments régulateurs perdus en *cis* (Pop *et al.*, 2004; Leipoldt *et al.*, 2007). Cette hypothèse est étayée par l'observation rapportée par Bagheri-Fam *et al.* (2006) qui démontre que 7 éléments conservés entre l'Homme et le *Fugu rubripes*, localisés entre -290 kb et +95 kb par rapport à *SOX9*, activent l'expression de *lacZ* de façon tissu-spécifique dans les différents domaines d'expression de *SOX9* (Bagheri-Fam *et al.*, 2006).

La séquence de Pierre Robin

La séquence de Pierre Robin (SPR, OMIM 261800) est définie par la triade : rétro- et/ou micrognathie,

glossoptose et fente palatine postérieure en forme de U (fente de forme « romane ») (Holder-Espinasse *et al.*, 2001; Kahn *et al.*, 2003). La triade de la SPR est souvent associée à des troubles fonctionnels parmi lesquels les plus importants sont les troubles de succion-déglutition et les troubles respiratoires. Une insuffisance vélo-pharyngée, un reflux œsophagien, une dérégulation du rythme cardiaque et une hyperactivité vagale sont aussi fréquemment décrits. La SPR peut être incomplète, associant une combinaison et une expression variables des signes cardinaux.

Une étude rétrospective de 117 cas suivis sur une période de 10 ans à l'Hôpital Necker-Enfants Malades a proposé de classer les patients atteints d'une SPR en trois catégories (Holder-Espinasse *et al.*, 2001) : (i) SPR isolée (SPR_i; 48 % des cas); la SPR est alors le plus souvent sporadique mais on note des cas familiaux avec un mode de transmission probablement autosomique dominant, avec pénétrance incomplète et expression variable; (ii) SPR syndromique (35 % des cas) : la SPR est associée à d'autres signes cliniques qui décrivent un syndrome; (iii) SPR associée à d'autres malformations.

La SPR est donc cliniquement et génétiquement très hétérogène (Cohen, 1999; Houdayer *et al.*, 2001; Melkonimi *et al.*, 2003; Jamshidi *et al.*, 2004; Ounap *et al.*, 2005; Jakobsen *et al.*, 2006).

Hypothèse

Le fait que les éléments régulateurs activateur(s) et inhibiteur(s) de transcription ont souvent une fonction stade- et tissu-spécifique suggère que leurs mutations ou réarrangements peuvent altérer l'expression génique dans certains tissus alors que les autres domaines d'expression resteront intacts. On peut dès lors émettre l'hypothèse qu'il s'agit d'un phénotype plus restreint (endophénotype) que celui qui est associé à des mutations de séquences codantes du même gène (spectre complet d'un syndrome).

Suivant cette proposition, nous pouvons imaginer que les endophénotypes d'un syndrome polymalformatif complexe (ici la DC) puissent se manifester dans leurs formes isolées (ici la SPR) en conséquence d'une dérégulation plus fine, d'un ou plusieurs éléments régulateurs en *cis*, tissu- et stade-spécifiques, localisés à distance du gène et dont les mutations des séquences codantes sont la cause du syndrome en question (figure 1). Ainsi, un syndrome polymalformatif, caractérisé par un certain nombre d'endophénotypes (symptômes), et un endophénotype unique de ce syndrome apparaissant dans sa forme isolée pourraient représenter les deux extrêmes du spectre d'altérations d'expression d'un même gène.

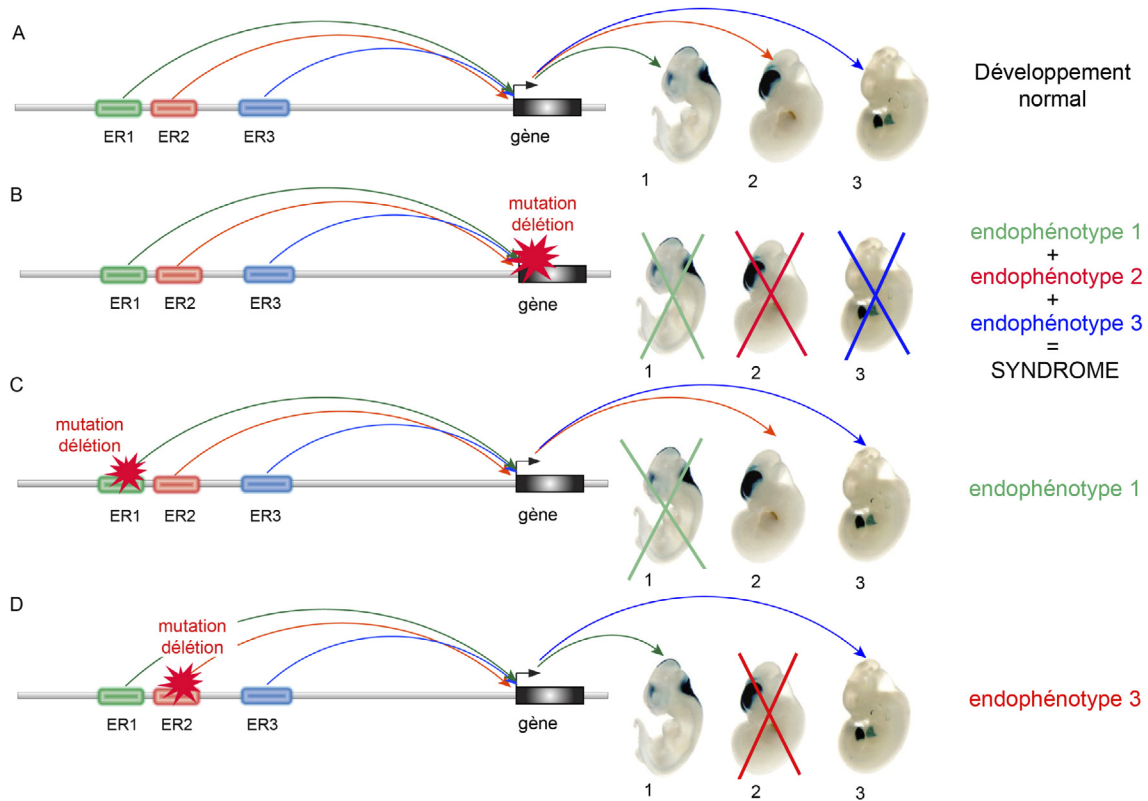


Fig. 1. Le modèle prédit des endophénotypes de syndromes humains polymalformatifs par altérations génomiques d'éléments régulateurs en *cis*. (A) L'expression spatio-temporelle d'un gène majeur du développement nécessite l'activité d'éléments régulateurs stade- et tissu-spécifiques. Ces éléments peuvent être localisés à plus ou moins grande distance par rapport au promoteur du gène cible qu'ils régulent. Dans l'exemple schématisé, trois éléments (ER1, ER2, ER3) activent l'expression du gène cible (représenté par un rectangle noir) de façon tissu-spécifique dans le prosencéphale, le rhombencéphale et les bourgeons de membres. (B) Les mutations de séquences codantes du gène cible causent un syndrome polymalformatif comprenant le spectre complet des trois endophénotypes tissu-spécifiques. (C, D) Les délétions ou les mutations d'un élément régulateur en *cis* entraînent un endophénotype 2 du syndrome polymalformatif par perte d'expression du gène cible dans le(s) tissu(s) correspondant(s) à l'action de l'élément affecté. Par exemple, les altérations de l'élément ER1 entraînent un défaut du développement du rhombencéphale, les altérations concernant l'élément ER3 affectent le développement des bourgeons de membres. Les images des embryons sont reproduites d'après Visel *et al.* (2009a).

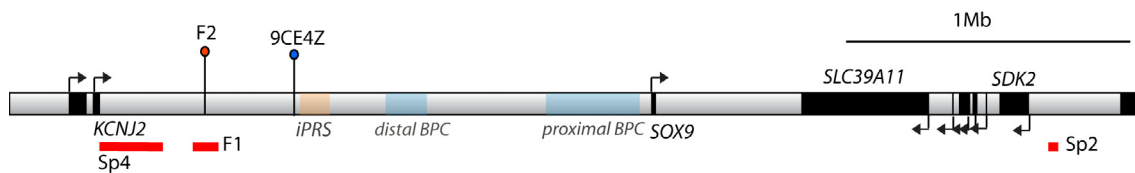


Fig. 2. Altérations génomiques dans le désert génique entourant *SOX9* identifiées chez les patients SPRi. Le locus *SOX9* est représenté par une barre grise. Les gènes sont indiqués en noir avec une flèche indiquant la direction de la transcription. Les régions comprenant les translocations causant la dysplasie campomélique (distal BPC, *distal breakpoint cluster*; proximal BPC, *proximal breakpoint cluster*) sont représentées en bleu. La région regroupant les translocations T1-T3 identifiées chez les SPRi (translocation SPRi) est représentée en orange. Les délétions localisées dans le désert génique centromérique et télomérique à *SOX9*, identifiées par CGH chez les patients SPRi familiaux (F1) et chez le patient sporadique Sp4, sont représentées par des barres rouges. La position du HCNE-F2 muté dans la famille F2 est indiquée par le point rouge. Le deuxième élément testé par essai de gène rapporteur (9CE4Z) est indiqué par le point bleu.

La situation de *SOX9* nous a semblé idéale pour tester cette hypothèse, l'endophénotype de la DC qui a fait objet de nos premiers travaux étant la SPR.

Identification d'un locus de séquence de Pierre Robin isolée

Nous avons débuté ce travail sur la séquence de Pierre Robin isolée en identifiant un locus unique sur le chromosome 17q24.3-25.1 de 5,4 Mb ségrégeant avec la SPRi dans la famille F1 (locus *PRS1*). D'autre part, nous avons étudié trois cas familiaux non apparentés de SPRi (familles T1, T2 et T3) où ségrège une translocation équilibrée dont un des points de cassure est situé sur le chromosome 17q24.3, dans le désert génique proximal au gène *SOX9*. Ces points de cassure sont groupés dans une région d'environ 200 kb (groupe de translocation SPR dans la figure 2), localisée à plus de 1 Mb en position centromérique par rapport à *SOX9*. Cette région est incluse dans le locus qui coségrège avec la SPRi dans la famille F1. De plus, une translocation similaire associée à un cas familial de SPRi a été rapportée par Jakobsen *et al.* (2007).

Prenant en compte l'absence de mutation de séquences codantes des gènes candidats au locus *PRS1* et les cas de translocations localisées dans le désert génique centromérique à *SOX9* dans des familles indépendantes de SPRi, nous avons fait l'hypothèse qu'une altération d'un ou plusieurs éléments régulateurs de l'expression de *SOX9* au cours du développement mandibulaire, et centromérique aux points de cassures des translocations T1–T3, est la cause de la SPRi dans la famille F1. Pour tester cette hypothèse, nous avons entrepris une expérience de gène rapporteur chez la souris, à l'aide d'une construction comprenant le gène rapporteur *lacZ*, un promoteur minimal et un élément conservé (9CE4Z) immédiatement centromérique aux points de cassures T1–T3. L'élément 9CE4Z a été sélectionné sur une conservation de séquence au cours de l'évolution de plus de 75 % entre Homme, singe, chien et souris, sur plus de 350 pb, et correspondant ainsi à un HCNE (*Highly Conserved Non-coding Element*). L'élément 9CE4Z régule effectivement une expression tissu-spécifique dans le mésenchyme mandibulaire.

Par la suite, nous avons entrepris de rechercher des variants du nombre de copies au locus *PRS1* chez des patients atteints de SPRi, par CGH de haute résolution. Cette partie de l'étude a été entreprise sur une série consécutive composée de 12 patients SPRi, les cas sélectionnés pour cette étude étant des SPRi sporadiques ou familiales (la famille F1 était incluse) sans translocation au locus *PRS1*. Nous avons ainsi identifié trois délétions au locus *SOX9*, emportant des séquences hautement conservées au

cours de l'évolution (HCNE) : (i) deux délétions centromériques aux points de cassures T1–T3 (dont une identifiée chez les patients de la famille F1) et (ii) une délétion télomérique à *SOX9*.

La délétion identifiée dans la famille F1 comprend 10 séquences conservées au cours de l'évolution à plus de 75 % entre l'Homme, le singe, le chien et la souris, sur plus de 350 pb (HCNE). Le séquençage direct de ces HCNE chez les patients présentant une SPRi, un caryotype normal et pas de polymorphisme au locus *SOX9*, a permis d'identifier une mutation ponctuelle coségréant avec la malformation dans la famille F2, localisée dans un HCNE que nous avons dénommé HCNE-F2, et absente dans plus de 440 chromosomes témoins.

Analyse fonctionnelle du HCNE-F2

Un essai de gène rapporteur chez l'embryon de souris a démontré que le HCNE-F2 active l'expression tissu-spécifique dans le mésenchyme mandibulaire. De plus, des vecteurs d'expression contenant le HCNE-F2 sauvage ou muté, le promoteur *SOX9* minimal et le gène rapporteur *lac+*, ont été transfectés dans différents systèmes cellulaires : lignées cellulaires dérivées du mésenchyme mandibulaire (MEPA mandibulaires) et maxillaire (MEPA maxillaires) de l'embryon murin en culture et la lignée HEK293T et SKNBE(2)c. Ainsi nous avons démontré que la fonction activatrice du HCNE-F2 spécifique des cellules MEPA mandibulaires est abolie par la mutation ponctuelle identifiée dans la famille F2.

L'analyse *in silico* du HCNE-F2 suggérait que la mutation « F2 » touche le site de fixation de MSX1, un répresseur de transcription, fortement exprimé dans le mésenchyme mandibulaire au cours du développement crânio-facial (Cai *et al.*, 2005). La souris homozygote mutante *Msx1*^{-/-} présente en effet des défauts du squelette facial, comprenant une fente palatine et une micrognathie rappelant la SPR (Satokata & Maas, 1994). Nous avons confirmé la liaison de *Msx1* au HCNE-F2 par immunoprécipitation de chromatine (ChIP) sur cellules MEPA mandibulaires. Par expérience de retardement sur gel (EMSA), nous avons aussi démontré que MSX1 se lie avec une affinité supérieure à la séquence du HCNE-F2 muté comparée à sa séquence sauvage. Cela suggère que la mutation « F2 » conduit à une répression de l'expression de *SOX9* par MSX1, dans le mésenchyme mandibulaire.

Signature épigénétique au locus *PRS1*

L'étude de signature chromatinienne a été réalisée par ChIP parallèlement sur deux types cellulaires : les

cellules MEPA mandibulaires et les cellules MEPA maxillaires. Deux éléments ont été analysés : le HCNE-F2 et le HCNE délété dans le cas Sp2 (HCNE-SP2). Nos résultats montrent que, dans les cellules du mésenchyme mandibulaire, les deux éléments présentent un fort enrichissement de liaison du cofacteur p300 et de H3K4me, caractérisant les éléments activateurs. Cette signature était absente dans les cellules maxillaires. D'autre part, spécifiquement dans les cellules maxillaires, ces éléments portaient la signature de la chromatine répressive (H3K27me3). Ces résultats suggèrent encore une fois que HCNE-F2 et HCNE-Sp2 sont des éléments régulateurs spécifiquement actifs dans le mésenchyme mandibulaire, et dont l'altération restreint les effets phénotypiques à la mandibule en développement.

La séquence de Pierre Robin isolée, conséquence d'une perte d'expression tissu-spécifique de *SOX9* dans le mésenchyme mandibulaire au cours de l'embryogenèse

L'étude des cas familiaux de SPR isolée F1, T1, T2 et T3 a permis de cartographier un locus *PRS1* sur le chromosome 17q24.3. L'analyse de ce locus par CGH et par séquençage direct chez des patients atteints de SPRi sporadique ou familiale a aussi permis d'identifier des altérations de séquences non codantes localisées dans le désert génique environnant *SOX9*, un gène majeur du développement mandibulaire, chez 4 des 12 cas étudiés. Toutes ces altérations impliquent des éléments conservés au cours de l'évolution.

Pour ces patients, nous mettons en avant l'hypothèse physiopathologique fondée sur une dérégulation stade- et tissu-spécifique de l'expression de *SOX9* dans le mésenchyme mandibulaire par altération d'éléments régulateurs en *cis* au locus *PRS1*, qu'il s'agisse de translocations, délétions ou mutations ponctuelles. Les arguments désignant *SOX9* comme meilleure cible des éléments régulateurs en question sont les suivants :

- *SOX9* est fortement exprimé dans le mésenchyme mandibulaire au cours du développement chez la souris comme chez l'Homme (Cai *et al.*, 2005).
- *SOX9* est impliqué dans le développement des cellules de la crête neurale et du cartilage de Meckel (Mori-Akiyama *et al.*, 2003).

Une perte de fonction par haploinsuffisance ou mutation hétérozygote de la séquence codante de *SOX9* est responsable de la DC, un syndrome polymalformatif sévère souvent associé à la SPR. La dysplasie campomélique acampomélique,

causée par des translocations ou des délétions localisées dans le désert génique entourant *SOX9*, est moins sévère que la DC, et également associée à la SPR. Ces cas indiquent que la SPR résulte d'une haploinsuffisance de *SOX9* (Pop *et al.*, 2004; Lecointre *et al.*, 2009; Leipoldt *et al.*, 2007).

- *SOX9* active la transcription des gènes codant les collagènes *Col2a1*, *Col11a1* et *Col11a2* dans les chondrocytes (Lefebvre *et al.*, 1997; Bridgewater *et al.*, 1998; Sekiya *et al.*, 2000; Mikkelsen *et al.*, 2007). Des mutations hétérozygotes de la séquence codante de chacun de ces trois gènes sont impliquées dans le syndrome de Stickler dont la SPR est un signe clinique classique (Holder-Espinasse *et al.*, 2001).

Les souris mutantes, chez lesquelles *SOX9* est invalidé dans les cellules de la crête neurale, présentent une mandibule réduite et une fente palatine, phénotype rappelant la SPR (Mori-Akiyama *et al.*, 2003).

Les éléments HCNE-F2 et 9CE4Z affectés par les altérations au locus *PRS1*, étudiés par essai de gène rapporteur, activent l'expression du gène rapporteur chez l'embryon de souris de manière tissu-spécifique dans une partie du territoire d'expression de *SOX9* chez l'embryon humain, incluant le mésenchyme mandibulaire.

Cependant, ces arguments ne permettent pas d'exclure l'hypothèse selon laquelle l'expression d'autres gènes est modifiée par les altérations identifiées au locus *PRS1*. En effet, le gène *KCNJ2*, immédiatement proximal à *SOX9*, est aussi une cible potentielle des éléments régulateurs du locus *PRS1* puisque des mutations hétérozygotes de *KCNJ2* entraînent la paralysie périodique cardio-dysrythmique d'Andersen (OMIM 170390), parfois associée à une micrognathie et une fente palatine. Pourtant, un effet dominant négatif sur la fonction de *KCNJ2* est rapporté pour toutes les mutations de séquence codante de *KCNJ2* identifiées à ce jour dans le syndrome d'Andersen. Ce modèle s'accorde mal avec l'hypothèse d'une perte de fonction par délétions ou mutations hétérozygotes dans un élément régulant son expression.

Nous proposons donc que les mutations SPRi/*SOX9*, portant sur des éléments hautement conservés non codants, ont pour effet une réduction de la transcription de *SOX9* stade- et tissu-spécifique dans les cellules du mésenchyme mandibulaire, rappelant les modèles d'inactivation conditionnelle chez la souris. Compte tenu du rôle de *SOX9* dans le développement du cartilage, ainsi que de son rôle d'activateur de la transcription de *COL2A1*, *COL11A1* et *COL11A2* (dont les mutations causent le syndrome de Stickler, souvent associé à la SPR), nous proposons que la perte d'activité des éléments régulateurs d'expression de *SOX9* dans le mésenchyme mandibulaire conduit à une réduction de

la transcription de *COL2A1*, *COL11A1* et *COL11A2* dans ce tissu. De cette cascade résulterait un défaut du développement du cartilage de Meckel et/ou du cartilage condylien, puis une hypoplasie mandibulaire avec micro- et/ou rétrognathie, aboutissant à la SPR.

Conclusion

Maladies génétiques par altération des séquences non codantes à grande distance du gène du développement cible

La majorité des maladies mendéliennes dont les bases moléculaires sont connues correspondent à des mutations de séquences codantes (Valle, 2004). Cependant, les gènes majeurs du développement présentent le plus souvent des patrons d'expression complexes dans le temps et dans l'espace. Leurs séquences codantes seules ne permettent pas une expression génique quantitativement et spatio-temporellement correcte. Cela implique l'existence de domaines complexes de régulation fonctionnelle associés à ces gènes.

Il semble que la régulation de l'expression des gènes majeurs du développement dépende avant tout de la régulation du taux de transcription sous l'influence de deux facteurs intimement liés : (1) les modifications de la chromatine au locus du gène en question et (2) l'activité de facteurs de transcription agissant en *trans* sur les séquences régulatrices en *cis* localisées à plus ou moins grande distance par rapport au site d'initiation de la transcription des gènes cibles. Les altérations du contexte chromatinien, comme cela peut être le cas avec les remaniements génomiques conduisant à un effet de position, ou les dysfonctionnements des mécanismes de régulation par des éléments régulateurs en *cis* pourraient modifier l'expression d'un gène régulé alors que l'unité transcriptionnelle demeure intacte. Ces modifications de la régulation d'un gène du développement peuvent être responsables de malformations congénitales (Kleinjan & van Heyningen, 2005; Kleinjan & Lettice, 2008; Kleinjan & Coutinho, 2009). La ségrégation des altérations de séquences non codantes avec le phénotype dans les cas familiaux, et/ou sa survenue *de novo* dans les cas sporadiques, et l'absence de ces altérations chez des témoins et dans les bases de données de SNP et CNV sont des arguments pour retenir leur caractère causal.

Dérégulation tissu-spécifique de l'expression génique dans des endophénotypes de syndromes complexes : un modèle de mécanisme des malformations communes

Les éléments régulateurs en *cis* ont souvent une fonction stade- et tissu-spécifique. Cela suggère que leurs

mutations ou réarrangements peuvent altérer l'expression d'un gène cible dans certains tissus alors que les autres domaines d'expression demeurent intacts. Nous avons démontré qu'un endophénotype de la dysplasie campomélique – la séquence de Pierre Robin – peut survenir de façon isolée à la suite d'altérations de la séquence nucléotidique localisées dans des régions non codantes à grande distance du gène cible, *SOX9*. Ces altérations (mutations ponctuelles, délétions, duplications ou translocations), qui entraîneraient une dérégulation tissu-spécifique de l'expression de *SOX9*, concernent un ou plusieurs éléments régulateurs tissu-spécifiques conservés au cours de l'évolution.

Ainsi, le modèle du gène *SOX9* et du syndrome de la dysplasie campomélique nous a permis de montrer qu'un syndrome polymalformatif caractérisé par plusieurs symptômes et un endophénotype isolé de ce même syndrome représentent deux extrêmes du spectre des altérations d'expression d'un même gène, la somme des endophénotypes possibles récapitulant le phénotype du syndrome par mutation perte-de-fonction de séquence codante du gène cible (figure 1).

De façon intéressante, cela ouvre de nouvelles voies de dissection moléculaire des malformations congénitales fréquentes comme les fentes labio-palatines, les défauts de fermeture du tube neural ou les malformations cardiaques. De plus, comme il a été démontré au locus *RET* (de Pontual *et al.*, 2006) impliqué dans la maladie de Hirschsprung isolée, les mutations touchant un élément régulateur d'un gène impliqué dans une maladie ne sont pas toujours de pénétrance complète; elles peuvent rejoindre la définition d'un polymorphisme et être un facteur de prédisposition dans un modèle multigénique. De grandes séries de patients présentant une malformation congénitale fréquente isolée pourraient être étudiées pour rechercher des altérations d'éléments régulateurs d'un gène du développement dont les mutations de séquence codante entraînent la malformation étudiée mais dans un contexte polymalformatif.

Implication pour des maladies génétiques communes

Il est aujourd'hui indiscutable que la régulation transcriptionnelle à grande distance joue un rôle critique dans de nombreux processus du développement et qu'une dérégulation peut entraîner des maladies rares ou fréquentes. Cependant, même si les éléments régulateurs tissu-spécifiques semblent être plus nombreux autour des gènes du développement nécessitant des patrons d'expression spatio-temporels complexes, la régulation à grande distance ne se borne pas aux gènes du développement. Les effets des variations concernant les éléments régulateurs de ces gènes

semblent jouer un rôle important aux loci à effet quantitatif (QTL- *Quantitative Trait Loci*). Dans ce contexte, nous pouvons citer l'exemple des SNPs dans des régions régulatrices distales du gène de la lactase permettant un maintien de l'expression de la lactase au cours de la vie adulte (Troelsen *et al.*, 2003; Lewinsky *et al.*, 2005; Tishkoff *et al.*, 2007). Les effets quantitatifs causés par des SNPs régulateurs affectant le niveau d'expression génique pourraient de manière similaire être associés à de nombreuses maladies fréquentes comprenant une composante génétique, comme le diabète, les maladies cardiovasculaires, l'hypertension ou l'obésité. Ainsi, dans un autre travail récent, l'invalidation d'un segment génomique non codant, comprenant des SNPs associés aux maladies coronariennes chez l'homme, a eu pour conséquences des défauts des cellules vasculaires et une altération de l'expression cardiaque des gènes avoisinants (Visel *et al.*, 2010).

Les conséquences cliniques des altérations de régulation spatio-temporelle des gènes commencent donc à émerger, notamment du fait de l'amélioration des méthodes d'investigation génomiques à haut débit, et pourraient peut-être se révéler aussi fréquentes que les mutations des séquences codantes.

Références

- Ahituv N., Zhu Y., Visel A., Holt A., Afzal V., Pennacchio L.A., Rubin E.M., Deletion of ultraconserved elements yields viable mice. *PLoS Biol*, 2007, 5, e234.
- Bagheri-Fam S., Barrionuevo F., Dohrmann U., Gunther T., Schule R., Kemler R., Mallo M., Kanzler B., Scherer G., Long-range upstream and downstream enhancers control distinct subsets of the complex spatio-temporal *Sox9* expression pattern. *Dev Biol*, 2006, 291, 382–397.
- Barrionuevo F., Bagheri-Fam S., Klattig J., Kist R., Taketo M.M., Englert C., Scherer G., Homozygous inactivation of *Sox9* causes complete XY sex reversal in mice. *Biol Reprod*, 2006, 74, 195–201.
- Blow M.J., McCulley D.J., Li Z., Zhang T., Akiyama J.A., Holt A., Plajzer-Frick I., Shoukry M., Wright C., Chen F., Afzal V., Bristow J., Ren B., Black B.L., Rubin E.M., Visel A., Pennacchio L.A., ChIP-Seq identification of weakly conserved heart enhancers. *Nat Genet*, 2010, 42, 806–810.
- Bridgewater L.C., Lefebvre V., de Crombrughe B., Chondrocyte-specific enhancer elements in the *Col11a2* gene resemble the *Col2a1* tissue-specific enhancer. *J Biol Chem*, 1998, 273, 14998–15006.
- Cai J., Ash D., Kotch L.E., Jabs E.W., Attie-Bitach T., Auge J., Mattei G., Etchevers H., Vekemans M., Korshunova Y., Tidwell R., Messina D.N., Winston J.B., Lovett M., Gene expression in pharyngeal arch 1 during human embryonic development. *Hum Mol Genet*, 2005, 14, 903–912.
- Cheung M., Chaboissier M.C., Mynett A., Hirst E., Schedl A., Briscoe J., The transcriptional control of trunk neural crest induction, survival, and delamination. *Dev Cell*, 2005, 8, 179–192.
- Cohen M.M., Jr., Robin sequences and complexes : causal heterogeneity and pathogenetic/phenotypic variability. *Am J Med Genet*, 1999, 84, 311–315.
- Consortium T.E.P., Identification and analysis of functional elements in 1% of the human genome by the ENCODE pilot project. *Nature*, 2007, 447, 799–816.
- Crutchley J.L., Wang X.Q., Ferraiuolo M.A., Dostie J., Chromatin conformation signatures: ideal human disease biomarkers? *Biomark Med*, 2010, 4, 611–629.
- de Laat W., Grosveld F., Spatial organization of gene expression: the active chromatin hub. *Chromosome Res*, 2003, 11, 447–459.
- de Pontual L., Pelet A., Trochet D., Jaubert F., Espinosa-Parrilla Y., Munnich A., Brunet J.F., Goriadis C., Feingold J., Lyonnet S., Amiel J., Mutations of the *RET* gene in isolated and syndromic Hirschsprung's disease in human disclose major and modifier alleles at a single locus. *J Med Genet*, 2006, 43, 419–423.
- Dermitzakis E.T., Raymond A., Antonarakis S.E., Conserved non-genic sequences – an unexpected feature of mammalian genomes. *Nat Rev Genet*, 2005, 6, 151–157.
- Dillon N., Gene autonomy: positions, please. *Nature*, 2003, 425, 457.
- Dillon N., Sabbattini P., Functional gene expression domains: defining the functional unit of eukaryotic gene regulation. *Bioessays*, 2000, 22, 657–665.
- Elgar G., Vavouri T., Tuning in to the signals: noncoding sequence conservation in vertebrate genomes. *Trends Genet*, 2008, 24, 344–352.
- Emison E.S., McCallion A.S., Kashuk C.S., Bush R.T., Grice E., Lin S., Portnoy M.E., Cutler D.J., Green E.D., Chakravarti A., A common sex-dependent mutation in a *RET* enhancer underlies Hirschsprung disease risk. *Nature*, 2005, 434, 857–863.
- Foster J.W., Mutations in *SOX9* cause both autosomal sex reversal and campomelic dysplasia. *Acta Paediatr Jpn*, 1996, 38, 405–411.
- Foster J.W., Dominguez-Steglich M.A., Guioli S., Kwok C., Weller P.A., Stevanovic M., Weissenbach J., Mansour S., Young I.D., Goodfellow P.N., Brook J.D., Schafer A.J., Campomelic dysplasia and autosomal sex reversal caused by mutations in an *SRY*-related gene. *Nature*, 1994, 372, 525–530.
- Frankel N., Davis G.K., Vargas D., Wang S., Payre F., Stern D.L., Phenotypic robustness conferred by apparently redundant transcriptional enhancers. *Nature*, 2010, 466, 490–493.
- Fuss S.H., Omura M., Mombaerts P., Local and cis effects of the H element on expression of odorant receptor genes in mouse. *Cell*, 2007, 130, 373–384.
- Gordon C.T., Tan T.Y., Benko S., Fitzpatrick D., Lyonnet S., Farlie P.G., Long-range regulation at the *SOX9* locus in development and disease. *J Med Genet*, 2009, 46, 649–656.

- Heintzman N.D., Ren B., Finding distal regulatory elements in the human genome. *Curr Opin Genet Dev*, 2009, 19, 541–549.
- Heintzman N.D., Hon G.C., Hawkins R.D., Kheradpour P., Stark A., Harp L.F., Ye Z., Lee L.K., Stuart R.K., Ching C.W., Ching K.A., Antosiewicz-Bourget J.E., Liu H., Zhang X., Green R.D., Lobanenkov V.V., Stewart R., Thomson J.A., Crawford G.E., Kellis M., Ren B., Histone modifications at human enhancers reflect global cell-type-specific gene expression. *Nature*, 2009, 459, 108–112.
- Hill-Harfe K.L., Kaplan L., Stalker H.J., Zori R.T., Pop R., Scherer G., Wallace M.R., Fine mapping of chromosome 17 translocation breakpoints $>$ or $=$ 900 Kb upstream of *SOX9* in acampomelic campomelic dysplasia and a mild, familial skeletal dysplasia. *Am J Hum Genet*, 2005, 76, 663–671.
- Holder-Espinasse M., Abadie V., Cormier-Daire V., Beyler C., Manach Y., Munnich A., Lyonnet S., Couly G., Amiel J., Pierre Robin sequence: a series of 117 consecutive cases. *J Pediatr*, 2001, 139, 588–590.
- Houdayer C., Portnoi M.F., Vialard F., Soupre V., Crumiere C., Taillemite J.L., Couderc R., Vazquez M.P., Bahuau M., Pierre Robin sequence and interstitial deletion 2q32.3-q33.2. *Am J Med Genet*, 2001, 102, 219–226.
- Houston C.S., Opitz J.M., Spranger J.W., Macpherson R.I., Reed M.H., Gilbert E.F., Herrmann J., Schinzel A., The campomelic syndrome: review, report of 17 cases, and follow-up on the currently 17-year-old boy first reported by Maroteaux *et al.* in 1971. *Am J Med Genet*, 1983, 15, 3–28.
- Jakobsen L.P., Knudsen M.A., Lespinasse J., Garcia Ayuso, C., Ramos C., Fryns J.P., Bugge M., Tommerup N., The genetic basis of the Pierre Robin Sequence. *Cleft Palate Craniofac J*, 2006, 43, 155–159.
- Jakobsen L.P., Ullmann R., Christensen S.B., Jensen K.E., Molsted K., Henriksen K.F., Hansen C., Knudsen M.A., Larsen L.A., Tommerup N., Tümer Z., Pierre Robin sequence may be caused by dysregulation of *SOX9* and *KCNJ2*. *J Med Genet*, 2007, 44, 381–386.
- Jamshidi N., Macciocca I., Dargaville P.A., Thomas P., Kilpatrick N., McKinlay Gardner R.J., Farlie P.G., Isolated Robin sequence associated with a balanced t(2;17) chromosomal translocation. *J Med Genet*, 2004, 41, e1.
- Kahn J.L., Bourjat P., Barrière P., Imaging of mandibular malformations and deformities. *J Radiol*, 2003, 84, 975–981.
- Kleinjan D.A., van Heyningen V., Long-range control of gene expression: emerging mechanisms and disruption in disease. *Am J Hum Genet*, 2005, 76, 8–32.
- Kleinjan D.A., Lettice L.A., Long-range gene control and genetic disease. *Adv Genet*, 2008, 61, 339–388.
- Kleinjan D.J., Coutinho P., *Cis*-ruption mechanisms: disruption of *cis*-regulatory control as a cause of human genetic disease. *Brief Funct Genomic Proteomic*, 2009, 8, 317–332.
- Lecointre C., Pichon O., Hamel A., Heloury Y., Michel-Calemard L., Morel Y., David A., Le Caignec C., Familial acampomelic form of campomelic dysplasia caused by a 960 kb deletion upstream of *SOX9*. *Am J Med Genet A*, 2009, 149A, 1183–1189.
- Lefebvre V., Huang W., Harley V.R., Goodfellow P.N., de Crombrughe B., *SOX9* is a potent activator of the chondrocyte-specific enhancer of the pro alpha1(II) collagen gene. *Mol Cell Biol*, 1997, 17, 2336–2346.
- Leipoldt M., Erdel M., Bien-Willner G.A., Smyk M., Theurl M., Yatsenko S.A., Lupski J.R., Lane A.H., Shanske A.L., Stankiewicz P., Scherer G., Two novel translocation breakpoints upstream of *SOX9* define borders of the proximal and distal breakpoint cluster region in campomelic dysplasia. *Clin Genet*, 2007, 71, 67–75.
- Lettice L.A., Horikoshi T., Heaney S.J., van Baren M.J., van der Linde H.C., Breedveld G.J., Joosse M., Akarsu N., Oostra B.A., Endo N., Shibata M., Suzuki M., Takahashi E., Shinka T., Nakahori Y., Ayusawa D., Nakabayashi K., Scherer S.W., Heutink P., Hill R.E., Noji S., Disruption of a long-range *cis*-acting regulator for *Shh* causes preaxial polydactyly. *Proc Natl Acad Sci USA*, 2002, 99, 7548–7553.
- Lettice L.A., Heaney S.J., Purdie L.A., Li L., de Beer P., Oostra B.A., Goode D., Elgar G., Hill R.E., de Graaff E., A long-range *Shh* enhancer regulates expression in the developing limb and fin and is associated with preaxial polydactyly. *Hum Mol Genet*, 2003, 12, 1725–1735.
- Levine M., Tjian R., Transcription regulation and animal diversity. *Nature*, 2003, 424, 147–151.
- Lewinsky R.H., Jensen T.G., Moller J., Stensballe A., Olsen J., Troelsen J.T., T-13910 DNA variant associated with lactase persistence interacts with Oct-1 and stimulates lactase promoter activity *in vitro*. *Hum Mol Genet*, 2005, 14, 3945–3953.
- Lomvardas S., Barnea G., Pisapia D.J., Mendelsohn M., Kirkland J., Axel R., Interchromosomal interactions and olfactory receptor choice. *Cell*, 2006, 126, 403–413.
- Mansour S., Hall C.M., Pembrey M.E., Young I.D. A clinical and genetic study of campomelic dysplasia. *J Med Genet*, 1995, 32, 415–420.
- Mattick J.S., Makunin I.V., Non-coding RNA. *Hum Mol Genet*, 2006, 15 Spec No 1, R17–29.
- McKeown S.J., Lee V.M., Bronner-Fraser M., Newgreen D.F., Farlie, P.G., *Sox10* overexpression induces neural crest-like cells from all dorsoventral levels of the neural tube but inhibits differentiation. *Dev Dyn*, 2005, 233, 430–444.
- Melkonien M., Koillinen H., Mannikko M., Warman M.L., Pihlajamaa T., Kaariainen H., Rautio J., Hukki, J., Stofko J.A., Cisneros G.J., Krakow D., Cohn D.H., Kere J., Ala-Kokko L., Collagen XI sequence variations in nonsyndromic cleft palate, Robin sequence and micrognathia. *Eur J Hum Genet*, 2003, 11, 265–270.
- Mikkelsen T.S., Ku M., Jaffe D.B., Issac B., Lieberman E., Giannoukos G., Alvarez P., Brockman W., Kim T.K., Koche R.P., Lee W., Mendenhall E., O'Donovan A., Presser A., Russ C., Xie X., Meissner A., Wernig M., Jaenisch R., Nusbaum C., Lander E.S., Bernstein

- B.E., Genome-wide maps of chromatin state in pluripotent and lineage-committed cells. *Nature*, 2007, 448, 553–560.
- Mori-Akiyama Y., Akiyama H., Rowitch D.H., de Crombrugge B., *Sox9* is required for determination of the chondrogenic cell lineage in the cranial neural crest. *Proc Natl Acad Sci USA*, 2003, 100, 9360–9365.
- Ninomiya S., Isomura M., Narahara K., Seino Y., Nakamura Y., Isolation of a testis-specific cDNA on chromosome 17q from a region adjacent to the breakpoint of t(12;17) observed in a patient with acampomelic campomelic dysplasia and sex reversal. *Hum Mol Genet*, 1996, 5, 69–72.
- Nobrega M.A., Ovcharenko I., Afzal V., Rubin E.M., Scanning human gene deserts for long-range enhancers. *Science*, 2003, 302, 413.
- Nobrega M.A., Zhu Y., Plajzer-Frick I., Afzal V., Rubin E.M., Megabase deletions of gene deserts result in viable mice. *Nature*, 2004, 431, 988–993.
- Ounap K., Ilus T., Laidre P., Uiho O., Tammur P., Bartsch O., A new case of 2q duplication supports either a locus for orofacial clefting between markers D2S1897 and D2S2023 or a locus for cleft palate only on chromosome 2q13-q21. *Am J Med Genet A*, 2005, 137A, 323–327.
- Pennacchio L.A., Ahituv N., Moses A.M., Prabhakar S., Nobrega M.A., Shoukry M., Minovitsky S., Dubchak I., Holt A., Lewis K.D., Plajzer-Frick I., Akiyama J., De Val S., Afzal V., Black B.L., Couronne O., Eisen M.B., Visel A., Rubin E.M., *In vivo* enhancer analysis of human conserved non-coding sequences. *Nature*, 2006, 444, 499–502.
- Pfeifer D., Kist R., Dewar K., Devon K., Lander E.S., Birren B., Korniszewski L., Back E., Scherer G., Campomelic dysplasia translocation breakpoints are scattered over 1 Mb proximal to *SOX9*: evidence for an extended control region. *Am J Hum Genet*, 1999, 65, 111–124.
- Pheasant M., Mattick J.S., Raising the estimate of functional human sequences. *Genome Res* 17, 2007, 1245–1253.
- Pop R., Conz C., Lindenberg K.S., Blesson S., Schmalenberger B., Briault S., Pfeifer D., Scherer G., Screening of the 1 Mb *SOX9* 5' control region by array CGH identifies a large deletion in a case of campomelic dysplasia with XY sex reversal. *J Med Genet*, 2004, 41, e47.
- Prabhakar S., Visel A., Akiyama J.A., Shoukry M., Lewis K.D., Holt A., Plajzer-Frick I., Morrison H., Fitzpatrick D.R., Afzal V., Holt A., Plajzer-Frick I., Morrison H., Fitzpatrick D.R., Afzal V., Pennacchio L.A., Rubin E.M., Noonan J.P., Human-specific gain of function in a developmental enhancer. *Science*, 2008, 321, 1346–1350.
- Sagai T., Hosoya M., Mizushima Y., Tamura M., Shiroishi T., Elimination of a long-range cis-regulatory module causes complete loss of limb-specific *Shh* expression and truncation of the mouse limb. *Development*, 2005, 132, 797–803.
- Sakai D., Suzuki T., Osumi N., Wakamatsu Y., Cooperative action of *Sox9*, *Snail2* and PKA signaling in early neural crest development. *Development*, 2006, 133, 1323–1333.
- Satokata I., Maas R., *Msx1* deficient mice exhibit cleft palate and abnormalities of craniofacial and tooth development. *Nat Genet*, 1994, 6, 348–356.
- Sekiya I., Tsuji K., Koopman P., Watanabe H., Yamada Y., Shinomiya K., Nifuji A., Noda M., *SOX9* enhances aggrecan gene promoter/enhancer activity and is up-regulated by retinoic acid in a cartilage-derived cell line, TC6. *J Biol Chem*, 2000, 275, 10738–10744.
- Siepel A., Bejerano G., Pedersen J.S., Hinrichs A.S., Hou M., Rosenbloom K., Clawson H., Spieth J., Hillier L.W., Richards S., Weinstock G.M., Wilson R.K., Gibbs R.A., Kent W.J., Miller W., Haussler D., Evolutionarily conserved elements in vertebrate, insect, worm, and yeast genomes. *Genome Res*, 2005, 15, 1034–1050.
- Spitz F., Duboule D., Global control regions and regulatory landscapes in vertebrate development and evolution. *Adv Genet*, 2008, 61, 175–205.
- Spitz F., Gonzalez F., Duboule D., A global control region defines a chromosomal regulatory landscape containing the *HoxD* cluster. *Cell*, 2003, 113, 405–417.
- Spokony R.F., Aoki Y., Saint-Germain N., Magner-Fink E., Saint-Jeannet J.P., The transcription factor *SOX9* is required for cranial neural crest development in *Xenopus*. *Development*, 2002, 129, 421–432.
- Stathopoulos A., Levine M., Genomic regulatory networks and animal development. *Dev Cell*, 2005, 9, 449–462.
- Suda Y., Kokura K., Kimura J., Kajikawa E., Inoue F., Aizawa S., The same enhancer regulates the earliest *Emx2* expression in caudal forebrain primordium, subsequent expression in dorsal telencephalon and later expression in the cortical ventricular zone. *Development*, 2010, 137, 2939–2949.
- Tishkoff S.A., Reed F.A., Ranciaro A., Voight B.F., Babbitt C.C., Silverman J.S., Powell K., Mortensen H.M., Hirbo J.B., Osman M., Ibrahim M., Omar S.A., Lema G., Nyambo T.B., Ghorji J., Bumpstead S., Pritchard J.K., Wray G.A., Deloukas P., Convergent adaptation of human lactase persistence in Africa and Europe. *Nat Genet*, 2007, 39, 31–40.
- Troelsen J.T., Olsen J., Moller J., Sjostrom H., An upstream polymorphism associated with lactase persistence has increased enhancer activity. *Gastroenterology*, 2003, 125, 1686–1694.
- Uhlenhaut N.H., Jakob S., Anlag K., Eisenberger T., Sekido R., Kress J., Treier A.C., Klugmann C., Klasen C., Holter N.I., Riethmacher D., Schütz G., Cooney A.J., Lovell-Badge R., Treier M., Somatic sex reprogramming of adult ovaries to testes by *FOXL2* ablation. *Cell*, 2009, 139, 1130–1142.
- Valle D., Genetics, individuality, and medicine in the 21st century. *Am J Hum Genet*, 2004, 74, 374–381.
- Velagaleti G.V., Bien-Willner, G.A., Northup J.K., Lockhart L.H., Hawkins J.C., Jalal S.M., Withers M., Lupski J.R., Stankiewicz P., Position effects due to chromosome breakpoints that map approximately 900 Kb upstream and approximately 1.3 Mb

- downstream of *SOX9* in two patients with campomelic dysplasia. *Am J Hum Genet*, 2005, 76, 652–662.
- Visel A., Prabhakar S., Akiyama J.A., Shoukry M., Lewis K.D., Holt A., Plajzer-Frick I., Afzal V., Rubin E.M., Pennacchio L.A., Ultraconservation identifies a small subset of extremely constrained developmental enhancers. *Nat Genet*, 2008, 40, 158–160.
- Visel A., Akiyama J.A., Shoukry M., Afzal V., Rubin E.M., Pennacchio L.A., Functional autonomy of distant-acting human enhancers. *Genomics*, 2009a, 93, 509–513.
- Visel A., Blow M.J., Li Z., Zhang T., Akiyama J.A., Holt A., Plajzer-Frick I., Shoukry M., Wright C., Chen F., Afzal V., Ren B., Rubin E.M., Pennacchio L.A., ChIP-seq accurately predicts tissue-specific activity of enhancers. *Nature*, 2009b, 457, 854–858.
- Visel A., Zhu Y., May D., Afzal V., Gong E., Attanasio C., Blow M.J., Cohen J.C., Rubin E.M., Pennacchio L.A., Targeted deletion of the 9p21 non-coding coronary artery disease risk interval in mice. *Nature*, 2010, 464, 409–412.
- Wagner T., Wirth J., Meyer J., Zabel B., Held M., Zimmer J., Pasantés J., Bricarelli F.D., Keutel J., Hustert E., Wolf U., Tommerup N., Schempp W., Scherer G., Antonarakis S.E., Attwood J., Baertsch R., Bailey J., Barlow K., Beck S., Berry E., Birren B., Bloom T., Bork P., Botcherby M., Bray N., Brent M.R., Brown D.G., Brown S.D., Bult C., Burton J., Butler J., Campbell R.D., Carninci P., Cawley S., Chiaromonte F., Chinwalla A.T., Church D.M., Clamp M., Clee C., Collins F.S., Cook L.L., Copley R.R., Coulson A., Couronne O., Cuff J., Curwen V., Cutts T., Daly M., David R., Davies J., Delehaunty K.D., Deri J., Dermitzakis E.T., Dewey C., Dickens N.J., Diekhans M., Dodge S., Dubchak I., Dunn D.M., Eddy S.R., Elnitski L., Emes R.D., Eswara P., Eyraas E., Felsenfeld A., Fewell G.A., Flicek P., Foley K., Frankel W.N., Fulton L.A., Fulton R.S., Furey T.S., Gage D., Gibbs R.A., Glusman G., Gnerre S., Goldman N., Goodstadt L., Grafham D., Graves T.A., Green E.D., Gregory S., Guigó R., Guyer M., Hardison R.C., Haussler D., Hayashizaki Y., Hillier L.W., Hinrichs A., Hlavina W., Holzer T., Hsu F., Hua A., Hubbard T., Hunt A., Jackson I., Jaffe D.B., Johnson L.S., Jones M., Jones T.A., Joy A., Kamal M., Karlsson E.K., Karolchik D., Kasprzyk A., Kawai J., Keibler E., Kells C., Kent W.J., Kirby A., Kolbe D.L., Korf I., Kucherlapati R.S., Kulbokas E.J., Kulp D., Landers T., Leger J.P., Leonard S., Letunic I., Levine R., Li J., Li M., Lloyd C., Lucas S., Ma B., Maglott D.R., Mardis E.R., Matthews L., Mauceli E., Mayer J.H., McCarthy M., McCombie W.R., McLaren S., McLay K., McPherson J.D., Meldrim J., Meredith B., Mesirov J.P., Miller W., Miner T.L., Mongin E., Montgomery K.T., Morgan M., Mott R., Mullikin J.C., Muzny D.M., Nash W.E., Nelson J.O., Nhan M.N., Nicol R., Ning Z., Nusbaum C., O'Connor M.J., Okazaki Y., Oliver K., Overton-Larty E., Pachter L., Parra G., Pepin K.H., Peterson J., Pevzner P., Plumb R., Pohl C.S., Poliakov A., Ponce T.C., Ponting C.P., Potter S., Quail M., Reymond A., Roe B.A., Roskin K.M., Rubin E.M., Rust A.G., Santos R., Sapojnikov V., Schultz B., Schultz J., Schwartz M.S., Schwartz S., Scott C., Seaman S., Searle S., Sharpe T., Sheridan A., Shownkeen R., Sims S., Singer J.B., Slater G., Smit A., Smith D.R., Spencer B., Stabenau A., Stange-Thomann N., Sugnet C., Suyama M., Tesler G., Thompson J., Torrents D., Trevaskis E., Tromp J., Ucla C., Ureta-Vidal A., Vinson J.P., Von Niederhausern A.C., Wade C.M., Wall M., Weber R.J., Weiss R.B., Wendl M.C., West A.P., Wetterstrand K., Wheeler R., Whelan S., Wierzbowski J., Willey D., Williams S., Wilson R.K., Winter E., Worley K.C., Wyman D., Yang S., Yang S.P., Zdobnov E.M., Zody M.C., Lander E.S., Autosomal sex reversal and campomelic dysplasia are caused by mutations in and around the SRY-related gene *SOX9*. *Cell*, 1994, 79, 1111–1120.
- Waterston R.H., Lindblad-Toh K., Birney E., Rogers J., Abril J.F., Agarwal P., Agarwala R., Ainscough R., Alexandersson M., An P., *et al.*, Initial sequencing and comparative analysis of the mouse genome. *Nature*, 2002, 420, 520–562.
- Wirth J., Wagner T., Meyer J., Pfeiffer R.A., Tietze H.U., Schempp W., Scherer G., Translocation breakpoints in three patients with campomelic dysplasia and autosomal sex reversal map more than 130 kb from *SOX9*. *Hum Genet*, 1996, 97, 186–193.
- Woolfe A., Goodson M., Goode D.K., Snell P., McEwen G.K., Vavouri T., Smith S.F., North P., Callaway H., Kelly K., Walter K., Abnizova I., Gilks W., Edwards Y.J., Cooke J.E., Elgar G., Highly conserved non-coding sequences are associated with vertebrate development. *PLoS Biol*, 2005, 3, e7.
- Wunderle V.M., Critcher R., Hastie N., Goodfellow P.N., Schedl A., Deletion of long-range regulatory elements upstream of *SOX9* causes campomelic dysplasia. *Proc Natl Acad Sci USA*, 1998, 95, 10649–10654.
- Xi H., Shulha H.P., Lin J.M., Vales T.R., Fu Y., Bodine D.M., McKay R.D., Chenoweth J.G., Tesar P.J., Furey T.S., Ren B., Weng Z., Crawford G.E., Identification and characterization of cell type-specific and ubiquitous chromatin regulatory structures in the human genome. *PLoS Genet*, 2007, 3, e136.