

Université de Montréal

**Anomalies du tube neural: Mieux comprendre les causes génétiques de
cette pathologie complexe**

par

Philippe Lemay

Département de biochimie et médecine moléculaire

Faculté de Médecine

Thèse présentée à la Faculté des études supérieures

en vue de l'obtention du grade de PhD

en Biochimie, option génomique humaine

Août, 2017

© Philippe Lemay, 2017

Université de Montréal Faculté de Médecine

Cette thèse intitulée:

Anomalies du tube neural: Mieux comprendre les causes génétiques de cette pathologie complexe

Présenté par : Philippe Lemay

est sous-évaluation par un jury composé des personnes suivantes :

Jacques Drouin – Président-rapporteur

Zoha Kibar – Directeur

Marie-Pierre Dubé – Membre du jury

Simon Girard – Examineur externe

Dindial Ramotar – Représentant du doyen

Résumé

Les anomalies du tube neural (ATN) sont des malformations congénitales affectant 1:1000 naissances et causées par une fermeture incomplète du tube neural résultant en des malformations du cerveau et de la moelle épinière. Les présentations les plus courantes de ces malformations sont l'Anencéphalie et le Spina Bifida qui résultent d'échec de fermeture dans les régions crânielles et spinales respectivement. L'étiologie de la maladie est complexe et inclut des facteurs génétiques et environnementaux. À ce jour, près de ~250 gènes causatifs ont été identifiés chez le modèle murin permettant entre autres de conclure à un rôle important de la voie PCP dans l'étiologie des ATN. Des études de liaison et l'étude des gènes de la voie de l'acide folique ont aussi eu un certain succès, mais une majorité de l'étiologie génétique de la pathologie humaine demeure inconnue suggérant que de nouvelles approches sont nécessaires. Le séquençage de l'exome entier combiné aux méthodes d'analyse de variants *de novo* et de reséquençage de masse offre une option d'investigation puissante permettant d'éclaircir l'étiologie génétique des ATN.

L'investigation de variants *de novo* dans 43 trios a permis d'identifier 42 variants codants dont 6 entraînent une version tronquée de la protéine. De ceux-ci, 2 variants tronquants (p.Tyr392X et p.Leu948fs) touchaient le gène candidat *SHROOM3* ayant précédemment été associé aux ATN chez la souris. L'enrichissement statistique de variants tronquants dans ce gène suggère une association à la pathologie humaine. De plus, le nombre élevé de variants tronquants identifiés suggère que l'apparition de variants *de novo* est un des mécanismes permettant de maintenir la prévalence élevée de cette maladie létale dans la population.

L'investigation des variants transmis à l'aide d'une approche des gènes candidats chez 6 familles affectées par des ATN a permis l'identification d'un variant homozygote tronquant p.Asp16Aspfs*10 sur le gène *GRHL3* chez deux sœurs affectées par un ATN. *GRHL3* étant un gène candidat ayant été précédemment associé aux ATN chez la souris, un reséquençage massif de ce gène dans 192 patients sporadiques et une réanalyse des 43 trios précédents ont permis d'identifier 7 variants avec un haut potentiel pathogénique. L'analyse statistique de ces variants a permis de conclure un enrichissement significatif suggérant une association de *GRHL3* aux ATN humaines. Des études fonctionnelles subséquentes a permis de conclure que deux des variants entraînaient une perte de fonctions (p.Arg391Cys et p.Ala529Val), alors qu'un autre affectait

l'épissage de l'ARN (p.Ala318Glyfs*26). Ces résultats supportent fortement l'implication du gène *GRHL3* dans les ATN humains.

L'étiologie génétique des anomalies du tube neural demeure largement inconnue. Ces nouveaux gènes ouvrent la porte sur de nouveaux processus pathogéniques potentiels de la maladie. Ces résultats permettront de mieux comprendre la maladie et d'offrir une meilleure prise en charge clinique des familles affectées.

Mots clefs : Anomalies du tube neural, *GRHL3*, *SHROOM3*, Séquençage de l'exome entier, Génétique complexe, Mutations *de novo*

Abstract

Neural tube defects (NTD) are congenital malformations affecting 1 :1000 births and are caused by incomplete neural tube closure resulting in spinal cord and brain malformations. The most frequent presentations include Anencephaly and Spina Bifida which result from failure of closure of the cranial and spinal regions respectively. The etiology of the disease is complex and include genetic and environmental factors. To this day, ~250 causative genes have been identified in the mouse model and demonstrated the role of the planar cell polarity pathway in the NTD' etiology. Linkage studies and candidate gene studies of folic acid genes had limited success and a large portion of the genetic etiology of the human pathology remain unknown suggesting the need for novel approaches. Whole exome sequencing combined with *de novo* variant and massive resequencing methods offer a powerful tool that may lead to a better understanding of the genetic etiology of NTD.

Investigation of *de novo* variants in 43 trios identified 42 coding variants including 6 truncating ones. Of these, 2 truncating variants (p.Tyr392X et p.Leu948fs) were found in the *SHROOM3* gene which had previously been associated to NTD in mouse. Statistical enrichment of truncating variants in this gene suggest an association to the human pathology. The high number of truncating variants identified in this cohort also suggest that *de novo* variants are one of the mechanisms that maintain the high prevalence of this lethal disease in the population.

Investigation of transmitted variants using a candidate gene approach in 6 NTD affected family identified a homozygous truncating variant p.Asp16Aspfs*10 in the *GRHL3* gene in two NTD affected sisters. *GRHL3* is a strong candidate gene previously associated to NTD in mouse. Massive parallel resequencing of 192 sporadic patients and the analysis of the 43 previous trios identified 7 variants with a high pathogenic potential. Statistical analysis of these variants demonstrated a significant enrichment suggesting an association of *GRHL3* with human NTD. Functional studies supported these findings through the identification of two loss of function variants (p.Arg391Cys et p.Ala529Val) and a variant affecting RNA splicing (p.Ala318Glyfs*26).

The genetic etiology of NTD remain largely unknown. These new genes are essential to identification of new potential pathogenic processes of this disease. These results will help better understand the disease and offer better genetic counseling for the affected families.

Key words : Neural tube defects, *GRHL3*, *SHROOM3*, Whole exome sequencing, Complex genetics, *De novo* mutations

Table des matières

| | |
|---|----|
| 1. Introduction..... | 1 |
| 1.1 Formation du tube neural | 1 |
| 1.1.1 Gastrulation..... | 1 |
| 1.1.2 Neurulation primaire | 2 |
| 1.1.2.1 Induction neurale | 2 |
| 1.1.2.2 Neurulation..... | 3 |
| 1.1.2.2.1 Extension convergence..... | 3 |
| 1.1.2.2.2 Constriction apicale | 7 |
| 1.1.2.3 Fermeture du tube neural | 8 |
| 1.1.3 Neurulation secondaire | 12 |
| 1.2 Anomalies du tube neural | 13 |
| 1.2.1 Anomalies du tube neural ouvertes | 14 |
| 1.2.1.1 Les ATN ouvertes crânielles..... | 14 |
| 1.2.1.2 Les ATN ouvertes spinales | 15 |
| 1.2.2 Anomalies du tube neural fermées | 16 |
| 1.2.2.1 Les ATN fermées avec masse sous-cutanée | 16 |
| 1.2.2.2 Les ATN fermées sans masse..... | 17 |
| 1.2.2.3 Les défauts de formation de la notochorde | 18 |
| 1.2.3 Facteurs environnementaux..... | 19 |
| 1.2.4 Facteurs génétiques | 21 |
| 1.2.4.1 La voie des folates | 22 |
| 1.2.4.2 Étude des gènes candidats chez la souris..... | 24 |
| 1.2.4.2.1 La voie de la polarité cellulaire planaire..... | 24 |
| 1.2.4.2.2 Survol des associations identifiées dans le modèle murin..... | 26 |
| 1.2.5 Santé publique..... | 27 |
| 1.3 Le gène <i>SHROOM3</i> | 28 |
| 1.3.1 Association aux ATN et à d'autres pathologies | 28 |
| 1.3.2 Localisation et expression | 29 |
| 1.3.3 Implication dans la constriction apicale et dans l'asymétrie..... | 29 |
| 1.3.4 Membres de la famille Shroom | 31 |
| 1.4 Le gène <i>GRHL3</i> | 32 |
| 1.4.1 Association aux ATN et à d'autres pathologies | 32 |

| | |
|---|-----|
| 1.4.2 Localisation et expression | 34 |
| 1.4.3 Implication dans l'épiderme et régulation complexe..... | 34 |
| 1.4.4 Membres de la famille GRHL..... | 37 |
| 1.5 Études génétiques des maladies humaines | 38 |
| 1.5.1 Génétique traditionnelle | 40 |
| 1.5.1.1 Étude de gènes candidats..... | 40 |
| 1.5.1.2 Études de liaison et d'association | 40 |
| 1.5.2 Génétique de nouvelle génération..... | 41 |
| 1.5.2.1 Études de séquençage de l'exome et du génome..... | 41 |
| 1.5.2.1 Étude de reséquençage dirigé..... | 45 |
| 2. Projet de recherche | 46 |
| 2.1 Problématique | 46 |
| 2.2 Hypothèse et but..... | 46 |
| 2.3 Pertinence et retombées..... | 47 |
| 3. Contributions..... | 48 |
| Contributions aux anomalies du tube neural..... | 48 |
| Contributions supplémentaires..... | 48 |
| 4. Loss of function de novo mutations play an important role in severe human neural tube defects..... | 50 |
| 5. Rare deleterious variants in <i>GRHL3</i> are associated with human spina bifida..... | 76 |
| 6. Discussion | 101 |
| 6.1 Critiques et limites des résultats | 101 |
| 6.1.1 Détection de mutations mosaïques | 101 |
| 6.1.2 Limites de l'essai luciférase | 101 |
| 6.1.3 Analyses des bases de données publiques..... | 103 |
| 6.2 Analyse des gènes candidats <i>SHROOM3</i> et <i>GRHL3</i> | 104 |
| 6.2.1 <i>SHROOM3</i> | 104 |
| 6.2.1.1 Mécanismes potentiels impliquant <i>SHROOM3</i> dans différentes pathologies.... | 104 |
| 6.2.1.2 Les données publiques informent les hypothèses sur le mode d'action des variants..... | 107 |
| 6.2.2 <i>GRHL3</i> | 108 |
| 6.2.2.1 <i>GRHL3</i> est impliqué dans le développement oro-facial et dans la neurulation.. | 109 |
| 6.2.2.2 Des indices d'une régulation complexe basée sur les isoformes..... | 111 |
| 6.2.2.3 <i>GRHL3</i> dans la neurulation..... | 112 |

| | |
|---|-----|
| 6.3 Évolution de la recherche sur les ATN..... | 114 |
| 6.3.1 Les variants <i>de novo</i> dans les méta-analyses, l'identification de gènes candidats et leur association à l'acide folique | 114 |
| 6.3.2 Les familles multiplexes, les variants mendéliens et hautement délétères..... | 116 |
| 6.3.3 Une étiologie polygénique pour une maladie rare | 117 |
| 6.4 Impact clinique des découvertes..... | 118 |
| 6.4.1 Profil risque-bénéfice | 119 |
| 6.4.2 Analyse des coûts | 119 |
| 6.4.3 Retour aux familles et validations dirigées..... | 120 |
| 7. Plans futurs..... | 122 |
| 8. Conclusion | 124 |
| 9. Références | 125 |

Liste des tableaux

| | |
|---|----|
| Manuscript 1, Table 1: <i>De novo</i> mutations identified in NTD affected trios and confirmed by Sanger sequencing..... | 65 |
| Manuscript 1, Table S1: Description of the trios affected with neural tube defects included in this study..... | 70 |
| Manuscript 2, Table 1: Highly deleterious variants identified in GRHL3 in spina bifida..... | 95 |

Liste des figures

| | |
|--|----|
| Figure 1 : Représentation graphique du processus de formation du tube neural de la gastrulation jusqu'à la fermeture complète. La gastrulation permet la formation de la notochorde. L'induction neurale permet la formation du neuroectoderme et des plis neuraux. Le neuroectoderme se déforme et se soulève à l'aide des charnières dorsolatérales (CDL) et médiane (CM). Finalement, les extrémités des plis neuraux se fusionnent pour compléter la fermeture du tube neural. Le tube neural complété est recouvert d'épiderme et les cellules de la crête neurale sont formées..... | 4 |
| Figure 2 : Représentation graphique de l'effet de l'extension convergence sur la plaque neurale...5 | 5 |
| Figure 3 : Représentation des différentes protéines et interactions impliquées dans la voie PCP...6 | 6 |
| Figure 4 : Représentation graphique de l'effet de la constriction apicale sur la plaque neurale.....7 | 7 |
| Figure 5 : Représentation graphique des 3 principaux points de fermeture chez la souris et descriptions des différents modes de fermeture le long de l'axe antéro-postérieur.....10 | 10 |
| Figure 6 : Représentation graphique de la création des protrusions cellulaires lors de la fusion du tube neural.....11 | 11 |
| Figure 7 : Représentation de la neurulation secondaire démontrant la canalisation de l'éminence caudale et son rattachement au tube neural13 | 13 |
| Figure 8 : Représentation graphique des domaines fonctionnels du gène <i>SHROOM3</i> humain et de leur fonction. L'encadré transparent représente le domaine prédit EVH1.....30 | 30 |
| Figure 9 : Représentation graphique de la localisation et des remaniements nécessaires aux changements conformationnels impliqués dans la CA. Dans la forme relaxée, l'actine maintient la forme colonnaire des cellules. <i>SHROOM3</i> est ensuite recruté aux jonctions apicales et lie <i>DVL2</i> en entraînant un remaniement de l'actine pour former un anneau contractile. Finalement, l'activation des <i>ROCK</i> et de <i>MCL</i> entraîne l'activation de la myosine induisant la contraction.....31 | 31 |
| Figure 10 : Représentation graphique des 4 isoformes de <i>GRHL3</i> et description de la fonction des différents domaines protéiques.....33 | 33 |

Figure 11 : Représentation graphique des principaux modèles impliquant GRHL3 dans les anomalies du tube neural. La colonne de gauche représente le processus normal et la colonne de droite représente le processus perturbé par une insulte majeure à *GRHL3*.....36

Figure 12: Représentation graphique de la génétique humaine traditionnelle (gauche) ou de nouvelle génération (droite).....39

Figure 13 : Représentation graphique des méthodes de capture du séquençage parallèle massif avec à gauche le séquençage de l'exome, au centre le séquençage du génome et à droite le reséquençage dirigé (MIP).....43

Manuscript 1, Figure 1: *De novo* mutations identified in *SHROOM3* and in *PAX3* in trios affected with NTDs (A) Protein schematic representation of the long and predicted short isoforms of *SHROOM3* indicating the position of the two DNMs identified in NTD probands. (B) Chromatograms of the 3 individuals of family PrZRV and Pr394 (from top to bottom: proband, mother, father) surrounding the *SHROOM3* c.1176C>G nonsense variant and the c.2843_2844insG frameshift variant. (C) Protein schematic representation of the long and predicted short isoforms of *PAX3* indicating the position of the DNM identified in NTD probands. (D) Chromatograms of the 3 individuals of family Pr134 (from top to bottom: proband, mother, father) surrounding the *PAX3* c.218C>A nonsense variant.....64

Manuscript 1, Supplementary Figure S1: LoF *de novo* mutations identified in *DDX3X*, *MFAP1* and *WBSCR28* in 3 trios affected with NTDs. Chromatograms of the 3 trios (from top to bottom: proband, mother, father) showing the c.620dupA *DDX3X* frameshift variant (panel A), the c.69del7 *MFAP1* frameshift variant (panel B) and the c.380G>A *WBSCR28* nonsense variant (panel C).....68

Manuscript 1, Supplementary Figure S2: *De novo* mutations identified in orthologues of genes associated with NTDs in mouse models. Chromatograms of the 2 trios (from top to bottom: proband, mother, father) showing the c.1171C>T *GRHL3* variant and (panel A) and the c.4475G>A *PTPRS* variant (panel C). Multi-species alignment of the amino acids surrounding the p.R391C variant in *GRHL3* (panel B) and the p.R1492Q variant in *PTPRS* (panel D).....69

Manuscript 2, Figure 1: Pedigree of the family segregating c.48dupC (p.Asp16Aspfs*10) variant. Parents were heterozygous for the variant, while both children affected with spina bifida were homozygous. Sequence chromatograms are shown next to each family member.....91

Manuscript 2, Figure 2: Deleterious rare variants identified in GRHL3 in spina bifida. (A) A schematic representation of the GRHL3 gene and protein showing the position of the deleterious variants identified in GRHL3 in spina bifida. The different isoforms only affect the first and last exons which in turn modify the trans-activation and the dimerization domains found in N and C-terminus of the GRHL3 protein. Isoform 4 (NM_001195010.1) is the shortest with an alternative start codon in exon 2 and excludes the p.Asp16Aspfs*10 variant. Isoform 3 (NM_198174.2) is the longest using the alternative last exon 16'. Isoform 2 (NM_198173.2) shares exons 1-15 of isoform 3 and the last exon 16 of isoform 4. Isoform 1 (NM_021180.3) has an alternative first exon 1' and shares the exons 2-16 of isoforms 2. TA, trans-activation domain. (B) Western blot analysis of HEK293 cells transfected with either isoform 2 or 3 of GRHL3. (C) Western blot analysis of HEK293 transfected with either wild type or mutant GRHL3 isoform 2. (D) Luciferase assay of the isoforms 2 and 3 of GRHL3. Ratios of the luciferase to the renilla are reported. Isoform 3 has no reporter activity in this assay while isoform 2 results in a 17-fold increase in reporter activity. (E) Luciferase assay of the 6 GRHL3 variants identified in spina bifida and cloned in GRHL3 isoform 2. Ratios of the luciferase to the renilla are reported. The variants, p.Arg391Cys and p.Ala529Val, cause reduced reporter activity.....92

Manuscript 2, Figure 3: The c.1047+2T>C (p.Ala318Glyfs*26) variant in GRHL3 in spina bifida caused skipping of exon 8. (A) Exon 8 of GRHL3 flanked by 179bp of intron 7 and 197 bp of intron 8 were cloned in the exon trapping vector pSPL3. The c.1047+2T>C variant was introduced by mutagenesis. β -Globin primers were used to amplify the final fragment. (B) RT-PCR amplification using β -Globin primers shows a difference between the size of the WT and c.1047+2T>C products. Chromatograms of the RT-PCR products demonstrate skipping of exon 8.....94

Manuscript 2, Figure S1: Alignments of the *GRHL3* missense variants on other species. Alignments were done using human (NM_198173.2), mouse (NM_001013756.1), chicken (XM_004947906.2), xenopus (NM_001005642.1) and zebrafish (XM_009293109.2).....96

Manuscript 2, Figure S2: Graphical representation showing previously identified variants associated to Van Der Woude syndrome and nonsyndromic cleft palate on top and spina bifida variants on the bottom. A clustering of highly deleterious variants in the trans-activation domain of the GRHL3 protein is observed in spina bifida while no variant had previously been identified in this region in the Van Der Woude syndrome or in nonsyndromic cleft palate.....97

Liste des abréviations

| | |
|---------|---|
| 5meTHF | 5-methyl-tetrahydrofolate |
| ADN | Acide désoxyribonucléique |
| AKT | AKT serine/threonine kinase |
| ANKRD | Ankyrin repeat domain 6 |
| ARX | Aristaless related homeobox |
| ARN | Acide ribonucléique |
| ATN | Anomalie du tube neural |
| BMP | Bone morphogenic proteins |
| CA | Constriction apicale |
| CDC | Cell division cycle |
| CDH | Cadhérine |
| CDL | Charnières dorsolatérales |
| CELSR | Cadherin EGF LAG seven-pass G-type receptor |
| CITED | Cbp/P300 interacting transactivator with Glu/Asp rich carboxy-terminal domain |
| CM | Charnière médiane |
| CNV | Variant par nombre de copies |
| DAAM | Dishevelled associated activator morphogenesis |
| DDX | DEAD-box helicase |
| DHFR | Dihydrofolate réductase |
| DIA | Diaphanous |
| DGO | Diego |
| DKK | Dickkopf WNT signaling pathway inhibitor |
| DPI | Diagnostic préimplantatoire |
| DSH/DVL | Dishevelled |

| | |
|--------|---|
| EC | Extension convergence |
| EGFR | Epidermal growth factor receptor |
| ENU | N-ethyl-N-nitrosourea |
| EPH | Éphrine |
| ERK | Extracellular signal-regulated kinases |
| FGF | Fibroblast growth factors |
| FIV | Fécondation in-vitro |
| FLP | fente labio-palatine |
| FMI | Flamingo |
| FOLR | Récepteurs du folates 1 et 2 |
| FOX | Forkhead box |
| FZ/FZD | Frizzled |
| FUZ | Fuzzy |
| GPCR | G-protein coupled receptor |
| GRHL | Grainyhead like |
| HNF | Hepatocyte nuclear factor |
| IP | Inositol phosphate |
| IRF | Interferon regulatory factor |
| JNK | C-jun |
| LGL | Lethal giant larvae |
| LRP | LDL Receptor Related Protein |
| MAPK | MAP kinases |
| MCAM | Melanoma cell adhesion molecule |
| MCL | Myosines régulatrices des chaînes légères |
| MFAP | Microfibrillar associated protein |

| | |
|--------|---|
| MIP | Sondes d'inversion moléculaire |
| MTHFD | Methylene-tetrahydrofolate deshydrogenase, cyclohydrolase, formyltetrahydrofolate |
| MTHFR | Methyl-tetrahydrofolate reductase |
| MTR | 5-methyltetrahydrofolate-homocysteine methyl transferase |
| MTRR | 5-methyltetrahydrofolate-homocysteine methyl reductase |
| NKX2 | NK2 homeobox |
| NO | Acide nitrique |
| NP63 | Tumor protein P63 |
| PAR | Protease-activated receptor |
| PAX | Paired box |
| PCP | Polarité cellulaire planaire |
| PIP | Phosphatidylinositol phosphate |
| PITX | Paired Like Homeodomain |
| PK | Prickle |
| PTK | Protein tyrosine kinase |
| PTPRS | Protein tyrosine phosphatase receptor type S |
| RAC | Ras-related C3 botulinum toxin substrate |
| RHO | Ras homolog family member |
| RHOGEF | PDZ-Rho guanine exchange factor |
| ROCK | Rho-associated kinase |
| SAH | S-adenosylhomocysteine |
| SAM | S-adenosylmethionine |
| SCRB | Scribble |
| SDC | Syndecan |
| SHH | Sonic hedgehog |

| | |
|--------|---|
| SHROOM | Shroom Family Member |
| SHMT | Serine hydroxymethyltransferase |
| SLC19A | Solute carrier family 19 member |
| SLC46A | Solute carrier family 46 member |
| SMAD | SMAD family member |
| SNAI | Snail family transcriptional repressor |
| SNP | Single nucleotide polymorphism (Polymorphismes d'un nucléotide) |
| STBM | Strabismus |
| SVW | Syndrome de Van der Woude |
| TEM | Transition épithélio-mésenchymateuse |
| TGASE | Transglutaminase 1 |
| TGF | Transforming growth factor |
| TME | Transition mésenchymo-épithéliale |
| TYMS | Thymidylate synthétase |
| TWIST | Twist family BHLH transcription factor |
| UTR | Région non-translaté |
| VANGL | <i>Van Gogh-like protein</i> |
| VASP | Vasodilator-stimulated phosphoprotein |
| WBSCR | Williams-Beuren syndrome chromosome region |
| WNT | Wingless-type |
| ZIC | Zic family member |

Remerciements

Les travaux et les résultats présentés dans ce document n'auraient pas été possible sans l'apport considérable qu'un large groupe de personnes y ont fait. Je souhaite tout d'abord remercier ma superviseuse de thèse Zoha Kibar, une contributrice clef de ces travaux. Sous ta supervision et à travers nos conversations, j'ai appris tout ce qui était nécessaire pour avancer dans le domaine scientifique. Ta confiance en moi m'a permis d'aller acquérir les outils et les compétences qui m'ont permis d'avancer dans le projet et dans la vie. Je remercie aussi les membres du laboratoire qui à travers nos conversations ont su me transmettre leur passion et les connaissances qui furent nécessaires à l'écriture de ce document.

Je souhaite aussi remercier mes collaborateurs de l'équipe de Valeria Capra qui ont permis par leurs idées ou leur travail acharné de réaliser les travaux présentés ici. Je souhaite spécialement remercier le groupe de bio-informatique du RMGA. Alexandre Dionne-Laporte, Dan Spiegelman, Édouard Henrion, Ousmane Diallo vous avez su me transmettre votre passion pour l'analyse de données et me donner une perspective unique sur mes travaux. Vous avez tous et chacun été une influence clef dans l'avancement de mes travaux et dans mon développement personnel.

Je souhaite aussi remercier mes amis et ma famille pour le support qu'ils m'ont offert tout au cours de mes études, votre support est la clef de ma réussite. Un merci spécial à Mélanie Ratelle qui a aidé chaque jour à me garder motivé durant l'écriture de ce document.

Finalement, je remercie le FRQ-S et les IRSC pour avoir subventionné mes travaux et m'avoir permis d'accomplir ma vision de ce projet. Bien entendu, ces travaux seraient impossibles sans les patients, ainsi que leur famille que je remercie chaudement.

1. Introduction

1.1 Formation du tube neural

La formation du système nerveux chez l'humain commence dès la 3^{ème} semaine du développement embryonnaire¹ par la formation d'une structure circulaire nommée tube neural. Cette structure se forme à travers plusieurs étapes incluant la gastrulation, la neurulation primaire et la neurulation secondaire^{2 3}. L'aspect séquentiel de ces étapes varie d'une espèce à l'autre, mais s'établit à travers des processus biologiques communs¹. Une meilleure compréhension de ces derniers est nécessaire à l'étude des malformations pathologiques de cette structure.

1.1.1 Gastrulation

La gastrulation est le processus embryonnaire transformant le blastocyste mono-laminaire à une structure tri-laminaire connue sous le nom de gastrula. Au jour 8 post-fertilisation chez l'humain, la masse cellulaire interne du blastocyste se sépare en un feuillet bi-laminaire composé de l'épiblaste et de l'hypoblaste⁴. L'hypoblaste devient l'endoderme extra-embryonnaire qui donnera naissance au sac vitellin⁵. Au 15^{ème} jour post-fertilisation, des cellules provenant de ce qui deviendra le pôle postérieur s'invaginent de manière à créer la ligne primitive. Cette invagination de la ligne médiane caudale du disque bi-laminaire s'allonge ensuite vers la région antérieure du nœud primitif¹. Ce nœud primitif agit de manière à exprimer certains gènes nécessaires à la direction de la migration des cellules et à leur différenciation⁴. Les cellules subissent aussi une transition épithélio-mésenchymateuse résultant en une perte de E-Cadhérine⁶. Cette perte permet une migration des cellules à travers la ligne primitive. Une première vague de cellules s'infiltrent et viennent déplacer l'hypoblaste pour devenir l'endoderme définitif. Suivant cette vague, d'autres cellules s'infiltrent et se différencient en mésoderme de manière à occuper l'espace entre l'épiblaste et la couche nouvellement formée d'endoderme. À travers cette migration, le nœud primitif génère le mésoderme axial qui entraîne la formation de la plaque notochordale qui s'intègre initialement à l'endoderme puis se sépare ensuite pour former le notochorde⁷. Suite à cette migration, l'épiblaste restant se transforme en ectoderme dont les principales structures dérivées seront l'ectoderme neural et l'ectoderme de surface⁴.

1.1.2 Neurulation primaire

La neurulation est le processus principal entraînant la formation du tube neural. Elle se produit en deux phases : La neurulation primaire et la neurulation secondaire. La première de ces deux phases résulte en la formation du tube neural rostral, alors que la neurulation secondaire s'effectue plus tardivement et entraîne la formation du tube neural caudal au neuropore postérieur.

1.1.2.1 Induction neurale

Chez les mammifères, le processus de neurulation primaire commence par la transformation de l'ectoderme en cellules colonnaires de la plaque neurale ou neuroectoderme (Figure 1)¹. Ce processus connu sous le nom d'induction neurale s'initie durant la gastrulation à travers la répression des *bone morphogenic proteins* (BMP), entraînant la formation du tissu neural antérieur. Ces dernières sont responsables de la différenciation de l'ectoderme en épiderme et inhibent la différenciation neuronale. Ce processus est indépendant du notochorde/mésoderme chez le *Xenopus*^{8 9}. Afin d'initier l'induction neurale, des facteurs tels que *noggin*, *follistatin* et *chordin* sont produits par le nœud primitif et inhibent les BMP par interaction directe. Les éléments impliqués dans ce modèle chez les mammifères demeurent néanmoins incertains. La difficulté d'investiguer ce processus dans des animaux vivipares a dirigé les études vers le *Xenopus* et le poisson zèbre. Ces études chez le poisson zèbre suggèrent que les fibroblast growth factors (*fgf*) pourraient être impliqués¹⁰. Cette étude suggère que les *fgf* agissent à travers la voie des MAP kinases (MAPK) comme inhibiteurs des SMAD family member 1/5 (*smad1/5*) qui sont des activateurs du facteur de transcription Tumor Protein P63 (*np63*) qui inhibe l'expression de gènes entraînant un destin neural¹⁰. Cette théorie fut subséquemment approfondie chez le *Xenopus* où les *fgf* ont une fois de plus été associés à l'inhibition directe de *smad1*. Ils semblent aussi être associés à une inhibition de la transcription des BMP et à l'activation de la transcription des inhibiteurs de BMP. Finalement, une activation directe des gènes induisant un destin neural a aussi été reportée¹¹. Il semblerait que ce type d'activité s'effectue de manière dose dépendante et qu'une faible dose de *fgf* est nécessaire, alors qu'une forte dose entraînerait une transition de l'ectoderme en mésoderme^{11 12}. Des études précédentes ont aussi démontré que la perte de fonction de *fgf* n'est pas suffisante pour inhiber l'induction neurale suggérant que d'autres facteurs agissent de pair dans cette régulation¹³. La voie de signalisation wingless-type (WNT) est aussi associée à l'induction neurale, mais son rôle est

controversé. Une étude a démontré qu'une surexpression de WNT permet d'inhiber les BMP et d'induire un destin neural¹⁴. Cependant, des études plus récentes démontrent qu'au contraire les WNT inhibent l'effet des *fgf* et qu'ils doivent être d'abord réprimés pour permettre l'induction neurale^{15 16}.

1.1.2.2 Neurulation

La plaque neurale, créée par l'induction neurale, est le siège des transformations mécaniques nécessaires à la formation du tube neural. Ces transformations entraînent la formation des plis neuraux qui permettent à la plaque neurale de se déformer (Figure 1). Suite à la formation du neuroectoderme, les plis neuraux se forment à la jonction avec l'ectoderme. Ces derniers se soulèvent dorsalement et suite à la formation de charnières se réunissent et fusionnent de manière à donner le tube neural. Les deux processus morphogéniques impliqués sont l'extension convergence et la constriction apicale.

1.1.2.2.1 Extension convergence

L'extension convergence (EC) entraîne une migration des cellules vers la ligne médiane et une extension dans l'axe antéro-postérieur (Figure 2)¹⁷. Deux mécanismes ont été proposés pour expliquer cette migration: Le glissement cellulaire¹⁸ ou le remodelage des jonctions¹⁹. L'idée commune à ces deux approches est qu'un explant de plaque neurale de *Xenopus* réussit à se polariser et à procéder à une EC indépendamment de son environnement²⁰. Ceci suggère que la régulation et les forces motrices sont internes à la plaque neurale. Le premier modèle suggère que les cellules forment des protrusions qui se lient aux cellules adjacentes et qui serviront à induire une tension permettant la migration^{21 22 23 24}. Le deuxième modèle suggère plutôt que les protrusions sont secondaires à un remodellement des jonctions apicales et que la myosine activée de ces jonctions entraînerait la migration^{24 25 26}. Le glissement cellulaire a été observé dans les cellules mésenchymateuses profondes de la plaque neurale des *Xenopus* et dans la surface baso-latérale de l'épithélium neural des souris. Le remodelage des jonctions a été observé dans l'épithélium de la drosophile, du poulet et de la souris. Ces modèles ne sont pas nécessairement exclusifs, mais des études plus approfondies sont nécessaires afin d'éclaircir les bases mécanistiques de l'EC.

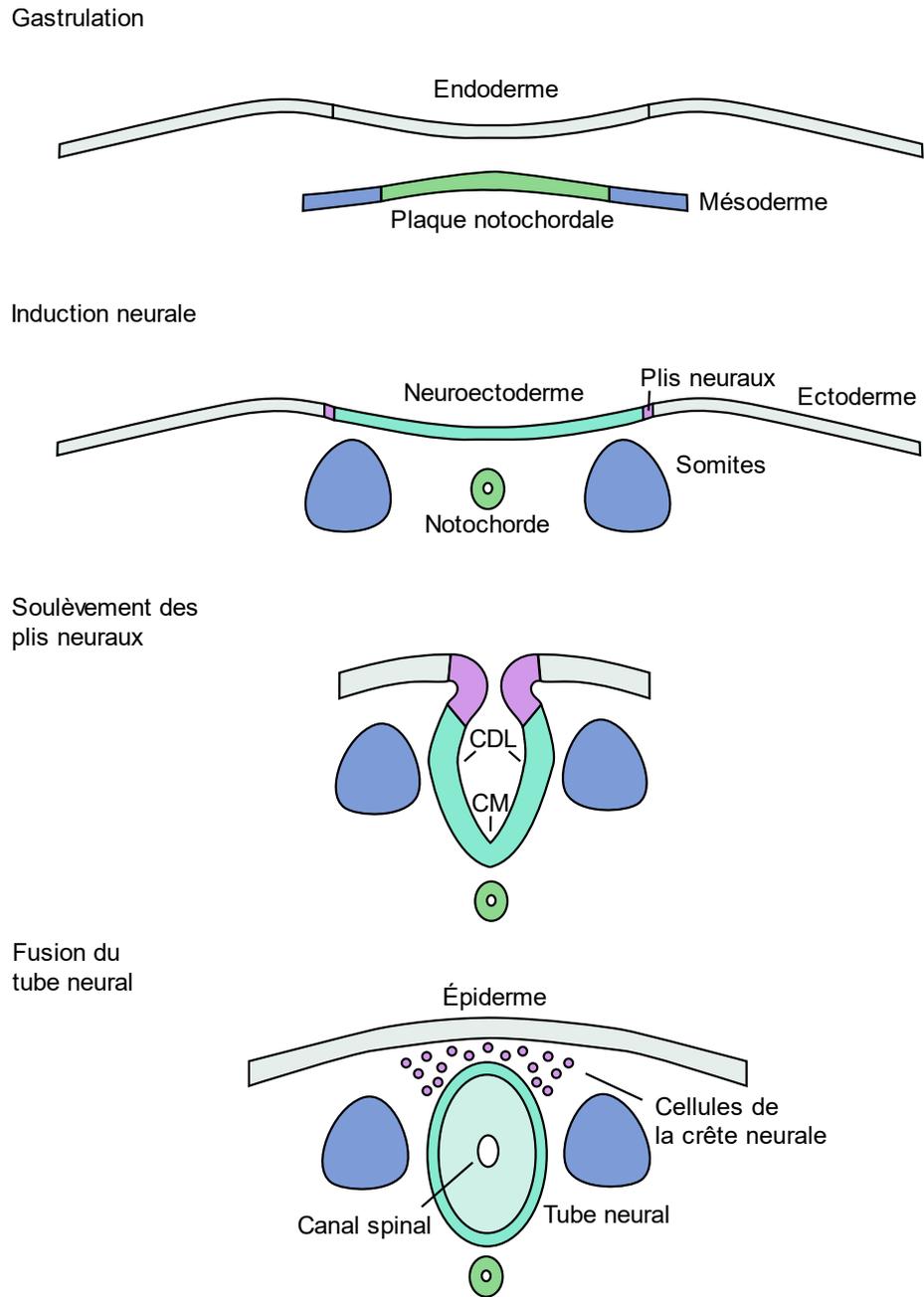


Figure 1 : Représentation graphique du processus de formation du tube neural de la gastrulation jusqu'à la fermeture complète. La gastrulation permet la formation de la notochorde. L'induction neurale permet la formation du neuroectoderme et des plis neuraux. Le neuroectoderme se déforme et se soulève à l'aide des charnières dorsolatérales (CDL) et médiane (CM). Finalement, les extrémités des plis neuraux se fusionnent pour compléter la fermeture du tube neural. Le tube neural complété est recouvert d'épiderme et les cellules de la crête neurale sont formées.

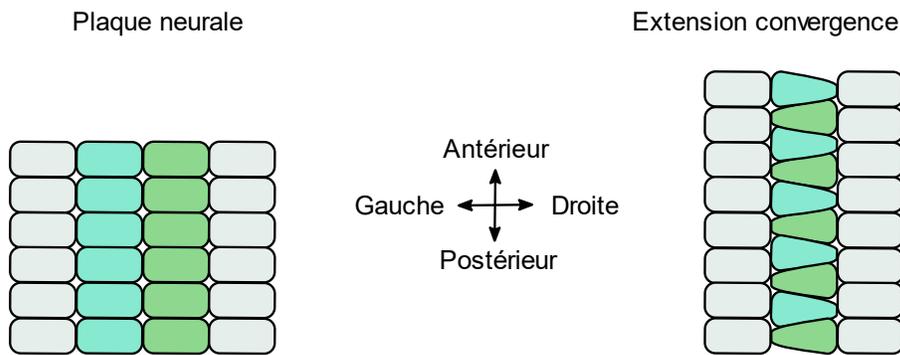


Figure 2 : Représentation graphique de l'effet de l'extension convergence sur la plaque neurale

La régulation de l'EC est principalement associée à la voie de la polarité cellulaire planaire (PCP) (Figure 3). Cette voie, aussi connue sous le nom de la voie Wnt non canonique frizzled/dishevelled (*fz/dsh*), a originellement été étudiée dans la drosophile²⁷. Son implication dans la formation des trichomes, du thorax et des ommatidies a permis de l'associer à la polarité des cellules dans le plan d'un épithélium^{28 29 30}. *fz* voyage le long de l'axe proximo-distal sur des microtubules apicaux pour enrichir les interfaces cellulaires distales^{31 32}. *fz* recrute *dsh* le concentrant à la membrane distale et le retirant du cytoplasme et de la membrane apicale³³. Ce complexe coordonne la localisation de strabismus (*stbm*) aux jonctions adhérentes proximales à l'intérieur des cellules avoisinantes. Cette protéine recrute prickle (*pk*)³⁴ qui entraîne l'exclusion de *dsh* de la membrane proximale. *dsh* interagit avec diego (*dgo*) et l'entraîne à la membrane distale. *dgo* exclut *pk* de la membrane distale et le complexe *dgo/dsh* maintient la localisation de flamingo (*fmi*) au niveau distal³⁵. Cette exclusion mutuelle de groupe protéiques participe ainsi à la création d'une polarité proximo-distale. Cette polarité permet d'activer différenciellement certains gènes par l'intermédiaire de la protéine *dsh* qui active les petites GTPase Ras Homolog Family Member (*Rho*) et Ras-Related C3 Botulinum Toxin Substrate (*Rac*)³⁶. L'activation de ces protéines se fait à l'aide de *Dishevelled associated activator morphogenesis 1 (daam1)* et permet l'activation de C-Jun (*jnk*). Chez l'humain, RHO est activé par PDZ-Rho guanine exchange factor (RHOGEF2) dont la transcription est activée par grainyhead like 3 (GRHL3). Rho permet ensuite l'activation de Rho-associated kinase (*rock*) et de la myosine. Tous ces effecteurs convergent finalement afin d'assurer la modulation de l'actine et des microtubules permettant une modification de la forme des cellules³⁶. Quoiqu'initialement étudié dans la drosophile, cette voie de signalisation a aussi été démontrée comme essentielle à l'EC chez plusieurs autres modèles^{37 38 39}. L'étude chez la drosophile demeure le portrait le plus complet des interactions protéiques

de la voie PCP et il est assumé que la voie chez les vertébrés ressemble à cette dernière. Cependant, chez la souris des familles multigéniques remplacent des gènes uniques chez la drosophile²⁷. De plus, d'autres gènes tel que Scribble (*SCRB*)⁴⁰, Protein tyrosine kinase 7 (*PTK7*)⁴¹ ou même d'autres WNT sont spécifiques aux vertébrés^{42 43}.

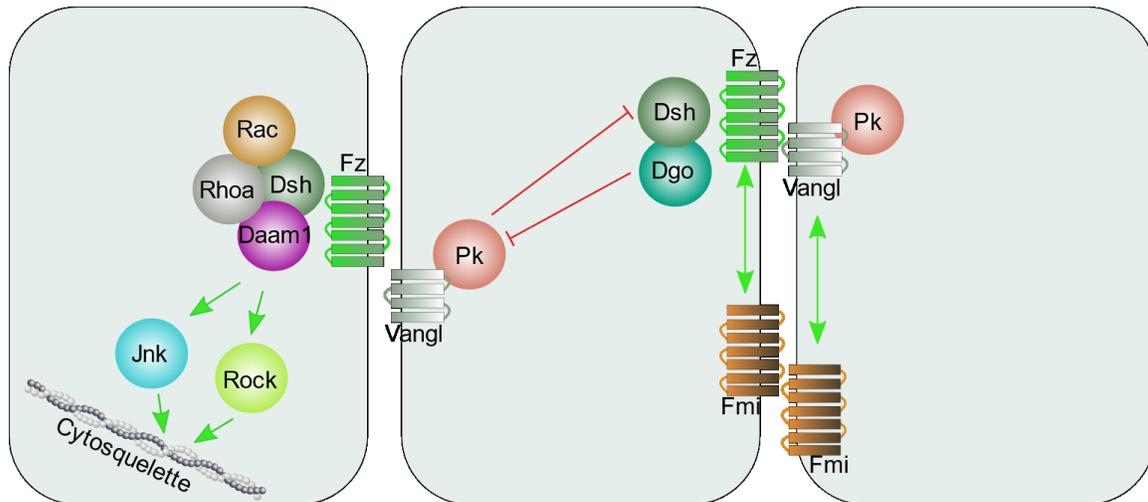


Figure 3: Représentation des différentes protéines et interactions impliquées dans la voie PCP.

Cette voie est aussi intrinsèquement liée à la voie WNT canonique à travers l'interrupteur moléculaire *dsh*. La phosphorylation de ce dernier permet son recrutement à la membrane et la stabilisation de la β -caténine cytosolique. Le tout s'effectue par l'inhibition du complexe APC-Axine-GSK3 β à travers la dégradation de l'axine et l'inhibition de GSK3 β . La β -caténine accumulée dans le cytoplasme transloque alors au noyau où elle active ses cibles géniques à l'aide des facteurs de transcriptions LEF/TCF. L'absence de phosphorylation de *dsh* permet d'assurer la dégradation de la β -caténine à travers le complexe APC-Axine-GSK3 β qui ubiquitine la protéine et la dirige vers la dégradation par le protéasome. Le maintien de cette inhibition est nécessaire à l'activation de PCP démontrant l'effet complexe de la régulation par la protéine *dsh*^{44 45}.

1.1.2.2.2 Constriction apicale

La constriction apicale permet d'induire une déformation de la plaque neurale nécessaire à la formation des charnières qui entraîneront les plis neuraux (Figure 1). La constriction apicale (CA) permet une réduction de la surface apicale des cellules à l'aide de forces mécaniques intrinsèques au feuillet cellulaire (Figure 4)⁴⁶. Cette force provient de la myosine II qui induit une tension au niveau des filaments d'actine⁴⁷. Les molécules de myosine sont un hexamère formé de deux chaînes lourdes, deux chaînes légères essentielles et deux chaînes légères régulatrices. Les chaînes lourdes couplent l'hydrolyse de l'ATP à l'association au filament d'actine, puis à son changement de conformation suivi d'une dissociation de l'actine⁴⁷. La tension générée par la myosine vient de son association à la F-actine qui permet de générer une contraction par la formation d'un complexe sarcomérique⁴⁸ ou à travers un assemblage désordonné⁴⁹. Le complexe sarcomérique semble générer une tension en périphérie de la surface apicale des cellules et ressemble à une ceinture mise sous tension. La deuxième source de tension provient d'un réseau d'actine/myosine organisé sous la membrane plasmique apicale des cellules. Ce réseau se nomme le cortex d'actine⁵⁰. Cette structure de 2 dimensions permet d'induire un flux cortical permettant d'enrichir la membrane apicale en actine permettant ainsi un potentiel maintien du gradient de tension^{51 52}. Cette contraction s'effectue sous forme de pulse dont la régulation complète demeure incertaine^{53 54}. Il est cependant clair que la présence d'APKC permettrait d'inhiber cette contraction^{53 55}.

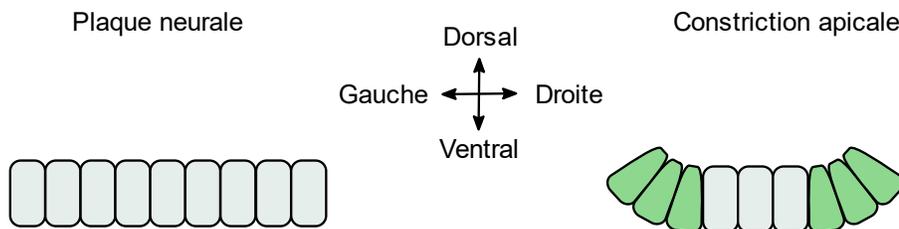


Figure 4 : Représentation graphique de l'effet de la constriction apicale sur la plaque neurale

La régulation de ce processus est fortement associée à la présence de la petite GTPase RHOA. Elle active la protéine ROCK qui est capable d'inhiber les phosphatases de myosines en phosphorylant directement les myosines⁵⁶. RHOA peut aussi activer les formines associées à diaphanous (Dia) qui facilitent l'assemblage des filaments d'actines non-associés⁵⁷. Chez la drosophile, l'activation de Rhoa est initiée par les facteurs de transcription twist family BHLH transcription factor 1 (Twist) et snail family transcriptional repressor 1 (Snai) qui activent coopérativement la Rhogef2 et la recrute à la surface apicale où elle activera Rhoa⁵⁸. Snail semble

impliqué dans l'induction de la contraction, alors que Twist a une fonction dans la stabilisation de la cellule contractée⁵⁹. Alors que Rhoa et Rock ségrègent au centre du cortex apical, Dia semble être plus également distribué sur le cortex. Rhoa et Dia se retrouvent aussi aux jonctions adhérentes, mais Rock semble en être exclue⁵⁹. Cette polarisation radiale semble primordiale pour assurer le maintien de ces jonctions sans lesquelles il est impossible de générer des changements conformationnels⁶⁰. Ce processus semble suffisant pour la CA chez la drosophile, cependant les vertébrés nécessitent aussi une ceinture d'actine. Ce processus, semble fortement associé à la protéine Shroom Family Member 3 (Shroom3). Ce gène est localisé à la surface apicale des cellules et lie la F-actine⁶¹. Il est capable de recruter Rock aux jonctions adhérentes, un événement responsable de la formation de câbles d'actine-myosine pouvant prendre la forme d'un sarcomère⁶². La voie PCP permet d'assurer la polarisation planaire de ce gène. Elle contrôle l'activation de Rock dans les jonctions adhérentes orientées dans l'axe médio-latéral et permet ainsi d'induire la constriction qui permettra la formation des plis neuraux nécessaires à la fermeture du tube neural²⁵. La régulation de *SHROOM3* et son implication dans les ATN sera plus longuement discutée dans la section 1.3.

1.1.2.3 Fermeture du tube neural

Les changements morphologiques induits par l'extension convergence et la CA permet la formation des CDL et CM. Ces dernières agissent à titre de pivots et permettent un soulèvement des régions distales de la plaque neurale (Figure 1)¹. La formation de ces charnières est séquentielle, et varie le long de l'axe antéro-postérieur. La concentration du morphogène sonic hedgehog (Shh) est responsable de cette variation. L'expression de ce gène chez le poulet est régulée par forkhead box A2 (Foxa2) qui induit aristaless related homeobox (Arx) qui à son tour régule positivement l'expression de Shh qui dans une boucle de régulation négative induit NK2 homeobox (Nkx2.2), un inhibiteur de Foxa2⁶³. Chez la souris, l'expression de Shh dans les papilles gustatives est réprimée par le facteur de transcription Grhl3, un facteur important dans la formation du tube neural⁶⁴. La co-localisation de ces protéines suggère que cette régulation pourrait être transposée au tube neural. La régulation de Shh induit une expression différente à travers l'axe antéro-postérieur du tube neural. Au niveau crânial, Shh est hautement exprimé et entraîne la formation de la CM en absence de CDL (mode de fermeture 1)^{65 66}. À haute concentration, Shh inhibe Noggin un inhibiteur de Bmp2. Bmp2 produit par l'ectoderme de surface peut alors inhiber la formation des CDL⁶⁷. La diminution de la concentration de Shh dans

la région spinale permet de lever l'inhibition de Noggin et de maintenir la répression de Bmp2. Dans cette condition, les CDL en plus de la CM sont formées (mode de fermeture 2). Le mode de fermeture 3 consiste à une absence de la CM en présence de CDL. Il est uniquement présent dans le bas de la région spinale (Figure 5)⁶⁸. La formation des CDL dépend principalement de la signalisation BMP⁶⁹. Cette voie de signalisation permet la phosphorylation des Smad1,5,8 et leur recrutement aux jonctions serrées en favorisant leur liaison au complexe protease-activated receptor (PAR). La signalisation BMP diminue la phosphorylation des Smad2,3 et entraîne leurs exclusions des jonctions serrées en augmentant leurs liaisons à lethal giant larvae (LGL). Cette activité est contrebalancée par transforming growth factor beta (TGFb) qui a une fonction contraire⁶⁹. La présence des Smad1,5,8 de manière apicale permet de stabiliser l'épithélium et réduit la réorganisation cellulaire. L'inhibition des BMP est médiée par Noggin qui lui dépend de zic family member 2 (*Zic2*) pour son expression spatiale⁶⁶. Les cellules soumises à cette inhibition obtiennent une flexibilité accrue laissant place aux changements structuraux nécessaires à la formation des CDL⁶⁹.

Les plis neuraux tournent ensuite autour de leurs charnières et convergent vers la ligne médiane dorsale. Ceci entraîne une apposition des plis neuraux et leur fusion. Les mécanismes régulant cette fusion demeurent encore largement inconnus cependant, l'apparence de protrusions cellulaires a été détectée (Figure 6)⁷⁰. Deux des facteurs importants dans ce processus de fusion sont les récepteurs éphrines (EPH) et les récepteurs PAR⁷¹. Chez la souris, *Ephrin-A5* et *EphA7* ont tous deux été identifiés sur le bord des plis neuraux crâniens et leur interaction serait nécessaire à l'adhésion cellulaire⁷¹. Ces récepteurs sont habituellement répulsifs, mais certains de leurs isoformes semblent modifier cette fonction. Le knockout de certains isoformes de *Ephrin-A5* entraîne un échec de fermeture suggérant un rôle positif dans la fusion⁷¹. Les récepteurs PAR sont des G-protein coupled receptor (GPCR). *Par1* et *Par2* agissent à travers la sous-unité $G\alpha_{i/o/z}$ dont le knockout peut aussi causer des anomalies du tube neural (ATN). Cette malformation est causée par un défaut de fermeture du tube neural. Les PAR sont exprimés dans l'ectoderme couvrant le neuroépithélium, durant l'apposition des plis neuraux. Leur localisation et leur

association aux ATN en font un excellent candidat dans la fusion des pli neuraux, mais le mode d'action demeure inconnu⁷².

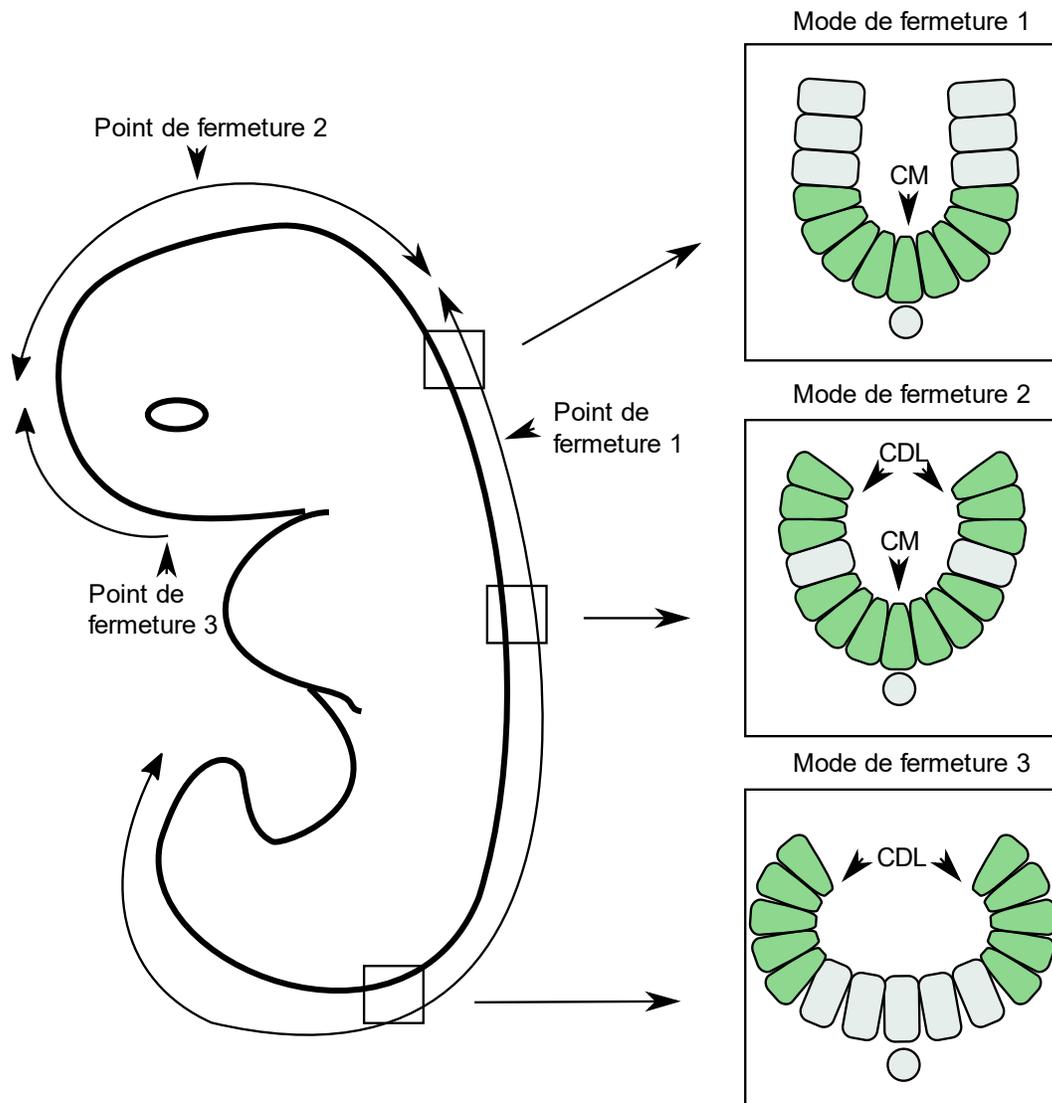


Figure 5 : Représentation graphique des 3 principaux points de fermeture chez la souris et descriptions des différents modes de fermeture le long de l'axe antéro-postérieur.

Une étude récente a aussi démontré l'importance du maintien des divers types cellulaires lors de la fermeture du tube neural⁷³. Trois types cellulaires se retrouvent au site de fusion : l'ectoderme de surface, le neuroectoderme et des cellules non-différenciées. Un des principaux régulateurs du maintien de ces types est le gène *Grhl3*. Ce gène est un facteur de transcription exprimé à la frontière entre l'ectoderme de surface et le neuroectoderme. Il entraîne la différenciation des cellules non-différenciées en ectoderme de surface lors de la fusion des plis neuraux. En absence de *Grhl3*, ces cellules sont transformées en neuroectoderme, inhibant une

fermeture adéquate. Cette différenciation est causée par la voie de WNT canonique, une des cibles de *Grhl3*. L'expression de cette voie est en parallèle inhibée par dickkopf WNT signaling pathway inhibitor 1 (*Dkk1*) qui agit en tant que régulateur temporel, permettant de maintenir les cellules dans leur forme non-différenciées, jusqu'à l'étape de fusion des plis neuraux. *Dkk1* permet aussi la différenciation en neuroectoderme à travers l'inhibition de la voie WNT canonique⁷³. *GRHL2*, un membre de la même famille génique, a été impliqué dans l'expression de gènes d'adhésion importants à la fusion du tube neural⁷⁴. Un de ces gènes est la Cadhérine 1 (CDH1) qui est une molécule d'adhésion fortement exprimée dans les plis neuraux⁷⁴. Cette fusion complète ainsi la neurulation primaire et se termine autour de la 6^{ème} semaine embryonnaire chez l'humain.

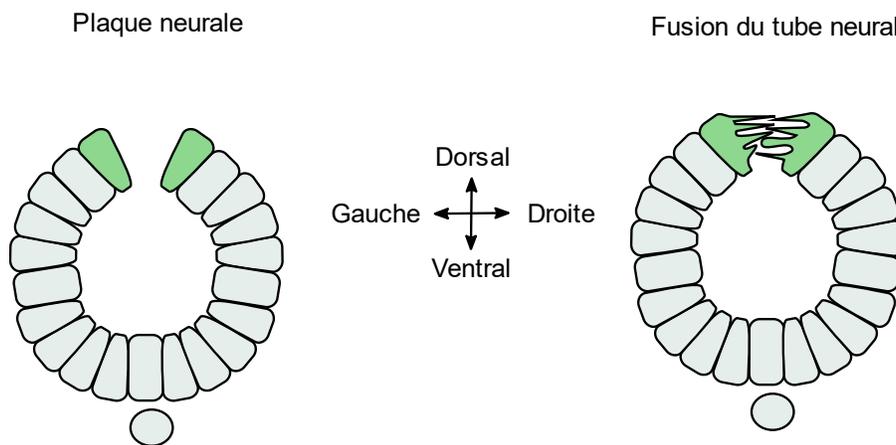


Figure 6 : Représentation graphique de la création des protrusions cellulaires lors de la fusion du tube neural

Physiologiquement, cette fermeture se fait sous la forme d'une fermeture éclair. Chez la souris, cette fermeture est initiée à E8.5 au niveau de la jonction cervicale et du rhombencéphale au point de fermeture 1. La fermeture 1 s'étend ensuite de manière bidirectionnelle vers la région cervicale et caudale de la souris. Le point de fermeture 2 s'initie au niveau de la jonction du prosencéphale et du mésencéphale et s'étend bi-directionnellement. Le point de fermeture 3 s'initie à l'extrémité rostrale du prosencéphale et se propage uniquement vers la région caudale. La position du point de fermeture 2 est flexible et peut être plus rostrale ou caudale. Les points de fermeture 2 et 3 complètent la neurulation crânienne, alors que le point de fermeture 1 s'étend jusqu'à la fermeture du neuropore et complète la neurulation primaire (Figure 5). Chez l'humain, le nombre de points de fermeture demeure incertain variant de 2 à 5^{75 76}. Les deux points principaux sont les mêmes que les points de fermeture 1 et 3 chez la souris. En plus du

point de fermeture 2, il est suggéré que des fermetures optionnelles peuvent s'initier entre le point 1 et 2 (point de fermeture 4) et de manière rostrale au point 1 entraînant la fermeture du neuropore (point de fermeture 5)⁷⁶. La présence de ces points est principalement supportée par la description des patients avec des ATN multiples⁷⁷.

1.1.3 Neurulation secondaire

Suite à la fermeture du neuropore postérieur, la neurulation primaire se termine⁷⁸. La neurulation secondaire débute ensuite par l'agrégation de cellules mésenchymateuses pluripotentes dans un cylindre qui se transforme en cellules épithéliales polarisées à travers la transition mésenchymo-épithéliale (TME)^{79 80}. Ce processus réversible permet à des cellules épithéliales, polarisées et riches en jonctions de devenir des cellules mésenchymateuses, non-polarisées et capable de migrer⁸¹. Ce processus est impliqué dans la réparation des blessures^{82 83}, dans la transformation oncogénique^{84 85}, dans la formation des organes^{86 87} et dans la formation des cellules de la crête neural⁸⁸. Le processus semble être dirigé par les BMP et leur inhibiteur Noggin⁸⁸. Nodal et Vg1 sont nécessaires à la modulation du signal WNT^{89 90} et Fgfr coopère avec Wnt3 pour réguler le signal⁹¹. Les facteurs de transcription Snail, Eomesodermin et Mesoderm posterior BHLH transcription factor^{92 93 94} sont aussi nécessaires à ce processus. Suite à son activation par les Claudines et les Occludines, Snail est même reconnu pour réprimer la *Cdh1*, un composant principal des cellules épithéliales⁹⁵. Snai2 permet aussi de maintenir la répression de Grhl3 durant la transition épithélio-mésenchymateuse (TEM) afin d'assurer que le retour à l'état épithélial ne se fasse pas trop rapidement⁹⁶. En effet, les gènes Grhl2 et Grhl3 semblent fortement impliqués dans le processus de TME. À travers deux types cellulaires distincts, ces gènes ont été associés à l'expression de la *Cdh1* impliquée dans un destin épithélial^{97 98}. Une ablation de ces gènes inhibe complètement la TME démontrant leur rôle clef dans le processus. Grhl2 et Grhl3 agissent tous deux sur les mêmes régions amplificatrices, mais ne démontrent aucune redondance dans différents types cellulaires^{97 98}. Ceci semble indiquer que malgré une fonction cellulaire similaire, ces deux gènes acquièrent une spécificité en fonction du type cellulaire. Il a été suggéré que cette spécificité est conférée par des co-facteurs. Dans le cadre de Grhl3, hepatocyte nuclear factor 4 alpha (Hnf4a) a été identifié, mais ils demeurent inconnus chez Grhl2⁹⁷. Les protéines GRHL sont aussi impliquées dans l'expression de Claudine 4, une protéine réprimant *Cdh1*, démontrant une potentielle boucle de régulation entre la TEM et la TME⁹⁸.

Le cylindre ainsi formé par ce processus se nomme masse cellulaire caudale chez l'humain⁹⁹ et bourgeon caudal chez les souris¹⁰⁰ et donne naissance au notochorde caudal, aux somites et au tube neural caudal. La masse cellulaire caudale subit une agrégation et une condensation formant un cordon médullaire solide. Le cordon subit ensuite une canalisation qui est représentée par l'apparition d'une lumière à l'intérieur du cordon¹⁰¹. Le cordon subit finalement une apoptose programmée entraînant sa régression et la formation du conus médullaire qui représente la fin conique du tube neural¹⁰¹. L'éminence caudale est ainsi formée et va ensuite fusionner avec le tube neural principal, processus qui entraîne des ATN lorsque troublé (Figure 7). Malgré la présence de voies moléculaires partagées avec la neurulation primaire, la neurulation secondaire semble avoir un processus partiellement distinct qui demeure néanmoins largement incompris⁸⁰.

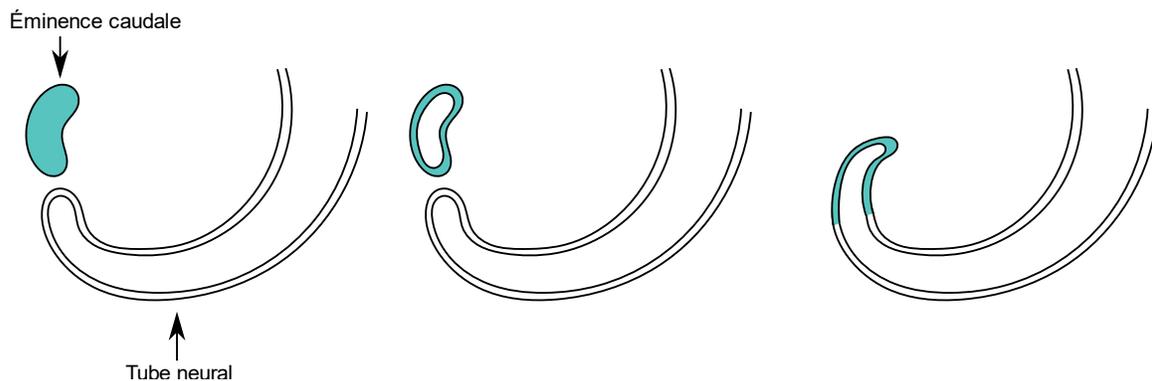


Figure 7: Représentation de la neurulation secondaire démontrant la canalisation de l'éminence caudale et son rattachement au tube neural

1.2 Anomalies du tube neural

Les ATN sont un groupe de malformations congénitales résultant d'un échec partiel ou complet de fermeture du tube neural durant l'embryogénèse. La prévalence de ces malformations est de 1-2 par 1000 naissances globalement dans les pays sans supplémentation en acide folique. Avant cette supplémentation, la prévalence canadienne était de 1.58 par 1000 naissances, mais elle est aujourd'hui de 0.86 par 1000 naissances¹⁰². L'emplacement du défaut sur l'axe rostro-caudal entraîne des présentations physiologiques diverses. La principale classification de ces malformations est faite en fonction de la présence de peau par-dessus le défaut. Si le défaut est exposé à l'environnement, il est défini comme ouvert, alors qu'en présence de peau par-dessus, il est défini comme fermé¹⁰³.

1.2.1 Anomalies du tube neural ouvertes

Les ATN ouvertes sont la forme la plus fréquente des ATN. Le myéломéningocèle (spina bifida) et l'anencéphalie sont les deux formes les plus fréquentes décrivant près de 90% des ATN ouvertes^{104 105}.

1.2.1.1 Les ATN ouvertes crânielles

L'**anencéphalie** est causée par un défaut de fermeture au point 2 ou 3 durant la troisième à la quatrième semaine embryonnaire et compte pour 40% des ATN ouvertes^{105 106}. Un défaut au point 2 entraîne une exencéphalie causée par un sous-développement de la voûte crânienne. La dégradation de l'encéphale par le liquide amniotique entraîne une anencéphalie¹⁰⁵. Un défaut au point 3 de fermeture va plutôt entraîner un sous-développement de l'encéphale dans la région du prosencéphale et est souvent associé à une malformation des os de la boîte crânienne¹⁰³. Il est possible d'observer aussi une forte association à une fente labio-palatine¹⁰⁷. Cette malformation est létale dès le stade embryonnaire ou tôt après la naissance¹⁰⁵. Il est aussi possible d'observer une proportion plus élevée de filles avec l'anencéphalie à un ratio de 3 :1¹⁰⁵. Étant donné l'absence de viabilité, aucune prise en charge clinique de la maladie n'est effectuée.

Le **céphalocèle** représente 7% des ATN ouvertes et se présente comme un engagement du cerveau et/ou des méninges à travers un défaut dans le crâne^{105 108 109}. Le terme englobant de céphalocèle peut être précisé en encéphalocèle et méningocèle crânien en fonction de l'engagement observé^{108 109}. Les causes demeurent inconnues, mais il est probable qu'une malformation cause une faiblesse au niveau du crâne, qui suite à l'accumulation de liquide céphalo-rachidien, se fissure entraînant des structures cérébrales¹⁰⁹. Les encéphalocèles peuvent être antérieures, occipitales ou pariétales. L'association de cette malformation à l'hydrocéphale pourrait être causée par une obstruction du flux de liquide céphalo-rachidien par du tissu cérébral¹⁰⁷. La prise en charge clinique de cette malformation commence par l'identification du défaut. À l'aide des techniques d'imagerie et d'injection de colorant, il est possible d'identifier les 2 principaux éléments : Le type de défaut crânien et la composition des structures cérébrales entraînées dans le défaut. Ce dernier est ensuite ouvert et les sections non-neurales sont extraites, alors que les sections neurales sont réintroduites dans le crâne. Le défaut dans le crâne est finalement fermé à l'aide d'une greffe osseuse. Cette correction est considérée comme définitive lorsque bien effectuée. Les corrections esthétiques subséquentes dépendent fortement de l'emplacement du défaut initial et vont parfois être effectuées à travers toute la croissance du

patient¹¹⁰. Après le traitement, des retards développementaux visuels et auditifs, de l'hypothyroïdie et des retards de croissance ont été reportés. Cependant, un développement cognitif normal, ainsi qu'une espérance de vie proche de la normale sont possibles¹¹¹.

Le **craniorachischisis** est la plus rare des ATN ouvertes avec seulement 3% des cas. Il se présente comme une anencéphalie continue à un myeloméningocèle qui peut s'étendre sur toute la longueur de l'embryon¹⁰⁵. Souvent considérée comme la plus sévère forme d'ATN, il résulte en une mort prénatale de l'embryon ou une fausse couche¹¹². Cette ATN dérive d'un échec de fermeture du point de fermeture 1¹⁰³. Une prévalence 7x plus élevée de la maladie a été identifiée au nord de la Chine, mais il est impossible de confirmer si la cause est génétique ou environnementale¹¹³. Considérant l'aspect complexe de la maladie, il s'agit probablement d'une combinaison de ces deux facteurs. Étant donné l'absence de viabilité, aucune prise en charge clinique de la maladie n'est effectuée.

1.2.1.2 Les ATN ouvertes spinales

Le **myelocèle** est la version la plus rare des ATN ouvertes spinales¹¹⁴. La placode neurale, un épaissement de l'ectoderme neural, se retrouve découverte et ne dépasse pas du niveau de la peau¹⁰⁴. Cette malformation est souvent viable, mais les ulcérations et les infections sont un large risque sur les enfants non-opérés¹⁰⁴. La cause est un échec de neurulation entraînant un maintien de l'adhésion entre l'ectoderme cutané et neural résultant en un défaut de fermeture de la peau. Étant donné que le mésenchyme ne migre pas derrière le tube neural, les structures dorsales se forment latéralement¹⁰⁴. Le **myeloméningocèle** représente 98% des ATN ouvertes spinales. Il a une pathogénèse similaire au myelocèle, mais est combiné avec une expansion de l'espace sous-arachnoïdien. La placode neurale est poussée vers l'extérieur créant un renflement dorsal¹¹⁴. Ces malformations sont dues à un échec de fermeture du point 1 et apparaissent principalement dans la région lombaire¹⁰³. Étant donné l'aspect ouvert de la malformation, la priorité de la prise en charge est de diminuer les risques d'infection. Pour ce faire, un recouvrement en silicone est appliqué jusqu'à l'opération qui consiste à réinsérer les méninges dans la moelle épinière et de fermer le défaut. Cette opération est effectuée chez près de 95% des patients aux États-Unis¹¹⁵. Suite à l'opération, une prévalence élevée d'hydrocéphalie, d'incontinence, de retards du développement verbal et de paralysie demeurent nécessitant des soins constants¹¹⁶.

L'**hémimyeloméningocèle** et l'**hémimyelocèle** sont des versions syndromiques des ATN ouvertes spinales incluant une diastématomyélie (séparation de la moelle épinière) dont l'une des moitiés échoue à neuruler¹¹⁷. La prise en charge clinique requiert une étape supplémentaire aux myeloméningocèles qui consiste à l'ablation du septum médian parfois présent et à relier les manchons durs. Les problèmes post-opératoires sont similaires aux myeloméningocèles¹¹⁸.

1.2.2 Anomalies du tube neural fermées

Les ATN fermées sont plus hétérogènes que les ouvertes. Aucune forme ne prédomine aussi fortement. Près de 50% des ATN fermées possèdent une marque cutanée à la naissance permettant de diagnostiquer le défaut. Ainsi, l'utilisation de l'appellation *occulta* est controversée. Ces malformations sont divisées en fonction de la présence ou non de masses sous-cutanées¹⁰⁴. Les défauts de la formation de la notochorde sont eux aussi considérés comme fermés et seront discutés dans une troisième catégorie.

1.2.2.1 Les ATN fermées avec masse sous-cutanée

Les **lipomyelocèles** et les **lipomyeloméningocèles** sont caractérisés par une masse lipidique (lipome) sous-cutanée positionnée en haut du sillon inter-fessier et qui s'étend asymétriquement¹¹⁹. La masse visible à l'œil nu contient un mélange d'adipocytes matures associés à des muscles striés, du cartilage, des os et des cellules nerveuses¹²⁰. Cette malformation causée par un échec de la neurulation primaire provient d'une disjonction prématurée de l'ectoderme cutané et du neuroectoderme ouvrant la voie à l'infiltration de cellules mésenchymateuses dans le tube neural¹¹⁹. Ces cellules se développent ensuite en tissu adipeux entraînant la formation de la masse lipidique. Dans le lipomyelocèle, l'interface entre le lipome et la placode se trouve à l'intérieur du canal spinal et le lipome se continue jusqu'au gras sous-cutané à l'aide d'une fissure postérieure. Le lipomyeloméningocèle est causé par une expansion de l'espace sous-arachnoïdien qui entraîne une dilatation du canal spinal. Dans la plupart des cas, cette expansion force une rotation de la placode et un déplacement des méninges¹¹⁴. Étant donné l'aspect fermé, les patients sont identifiés à l'âge adulte. Lorsque l'on considère l'aspect souvent asymptomatique du défaut, l'opération chirurgicale complexe nécessaire au traitement est considérée risquée. L'opération consiste à une ouverture de la peau au-dessus du défaut et d'une ablation du lipome pouvant nécessiter l'utilisation du laser. L'ablation doit avoir lieu le plus proche de la placode neurale sans l'endommager. La suite de l'opération converge avec les myeloméningocèles¹²¹.

Le **méningocèle** est caractérisé par un engagement de poches de liquide céphalo-rachidien entourées de dure-mère et d'arachnoïdes à travers une fissure postérieure¹²². Ce défaut est principalement lombaire ou sacré, mais peut aussi être occipital ou cervical¹²². L'origine demeure inconnue, mais il est probable qu'un renflement des méninges soit créé par une pulsation continue du liquide céphalo-rachidien¹²². Comparativement au lipomyelocèle, cette malformation est diagnostiquée tôt étant donné la protrusion visible. L'opération s'apparente aux myeloméningocèles combinée à un drainage du liquide céphalo-rachidien accumulé dans la poche¹²¹. Une large proportion des symptômes causés par le méningocèle peut être renversée étant donné l'atteinte minime aux méninges¹²³.

Le **myelocystocèle** est caractérisé par l'engagement d'un syrinx terminal dans un méningocèle postérieur. Le syrinx communique avec le canal épendymaire, alors que le méningocèle communique avec l'espace sous-arachnoïdien. La cause pourrait être une incapacité de sortie du liquide céphalo-rachidien lors de l'embryogénèse entraînant la formation du kyste¹²⁴¹²⁵. Le traitement passe par une chirurgie similaire aux méningocèles combinée à une ablation du kyste¹²⁶.

Les **lipomes intramédullaires** sont physiologiquement identiques aux lipomes affectant la dure-mère, mais demeurent à l'intérieur du sac thécal. Un large lipome peut entraîner un déplacement ou une compression de la moelle épinière¹²⁷. Les **lipomes filaires** quant à eux, proviennent d'un échec de la neurulation secondaire entraînant un épaissement lipidique du filum terminale. Cette malformation est causée par une canalisation incomplète et la persistance de cellules lipidiques¹²⁸. Les symptômes tel que les faiblesses musculaires et les atteintes neuronales ne peuvent pas être réglés par la chirurgie, cependant la décompression diminue la progression des symptômes par ablation de la masse¹²⁹.

1.2.2.2 Les ATN fermées sans masse

Les **filum terminale étroits** ont une présentation physiologique qui consiste à un filum terminale hypertrophique qui produit une fissuration et un blocage de l'ascension du cône médullaire¹²⁷. Cette malformation est causée par une différenciation anormale durant la neurulation secondaire entraînant la formation d'un filum plus épais¹³⁰. L'évolution des symptômes tel que la douleur, des problèmes d'incontinences et des faiblesses des membres inférieurs peuvent être stabilisés ou améliorés par un sectionnement du filum terminale¹²¹.

Les **sinus dermiques** se présentent comme une fistule connectant la surface de la peau au système nerveux central et ses méninges¹³¹. Ils sont causés par une disjonction incomplète du neuroectoderme et l'ectoderme cutanée durant la neurulation et se retrouvent à n'importe quel niveau de la ligne médiane¹³². Les principaux risques associés à cette malformation sont des infections et des abcès qui peuvent être causés par une entrée de bactéries dans le système nerveux central. La vitesse de traitement de cette malformation est clef étant donné que les individus dont le sinus est fermé chirurgicalement avant la première infection ont une rémission complète¹²¹.

Les **ventricules terminaux persistants** sont une cavité dans le cône médullaire qui est principalement diagnostiquée de manière post-mortem¹³³. Ce défaut est généralement asymptomatique et provient d'une régression incomplète du ventricule terminal durant la neurulation secondaire¹³⁴. Étant donné le caractère asymptomatique de cette malformation, la chirurgie est souvent évitée. Cependant, dans les cas symptomatiques, la douleur est souvent présente et une ablation du ventricule permet de guérir les symptômes 89% du temps¹³⁵.

1.2.2.3 Les défauts de formation de la notochorde

Les **diastématomyélie** se présentent comme une séparation de la moelle épinière en 2 segments. Cette séparation peut être vue sur un spectre continu variant d'une bifurcation partielle jusqu'à complète. La cause est un échec d'intégration à la ligne médiane résultant en 2 processus notochordaux parallèles. Le développement subséquent entraîne l'enveloppement ou non dans des sacs durs distincts¹³⁶. Cette pathologie cause des hypertrichoses, des atteintes neurologiques et osseuses. Ces symptômes (principalement les atteintes neurologiques) s'améliorent dans 33% des cas suite à une opération consistant à enlever les protrusions osseuses et à réunir la moelle épinière¹³⁷.

Les **agénésies caudales (syndrome de régression caudal)** peuvent être partielles ou complètes et se présentent à travers une agénésie de la colonne vertébrale, une imperforation anale, des anomalies génitales, une hypoplasie pulmonaire et des anomalies des membres inférieurs¹³⁸. Dépendant de la forme du cône médullaire, les agénésies caudales sont caractérisées différemment. Un cône haut et abrupte signifie que le bourgeon de la queue, ainsi que la notochorde sont affectés, alors qu'un cône bas et attaché signifie que seul le bourgeon est affecté. Leurs causes demeurent incertaines, mais sont caractéristiques de troubles de

neurulation et de gastrulation¹³⁹. Étant un spectre de malformation, aucun traitement clair n'existe. Les symptômes seront traités chirurgicalement l'un après l'autre¹²¹.

Les **dysgénésies spinales segmentaires** sont une agénésie ou dysgénésie de la colonne vertébrale lombaire, des anomalies segmentaires de la moelle épinière, une paraplégie congénitale et des déformations congénitales des membres inférieurs¹⁴⁰. Elles sont associées à un problème de gastrulation affectant un segment intermédiaire de la notochorde¹⁴⁰. Le résultat est une colonne vertébrale et une moelle épinière segmentées et un canal spinal rétréci ou même interrompu entre ces deux segments¹⁴⁰. Tout comme pour les agénésies caudales, il est impossible de définir un traitement clair à cette pathologie étant dépendante des symptômes observés.

1.2.3 Facteurs environnementaux

Le plus important facteur environnemental impliqué dans le développement des ATN est la prise d'acide folique (Vitamine B₉, folate)^{141 142}. Prise de manière péri-conceptionnelle, cette vitamine permet de réduire la prévalence d'ATN de 50 à 70% et varie en fonction des différentes régions¹⁰² (voir section 1.2.5 pour historique de cette association). L'accès est assuré par la distribution de médicaments ou par une supplémentation de l'alimentation tel qu'observé au Canada (discuté dans la section 1.2.5)¹⁰². L'acide folique est une vitamine hydrosoluble impliquée dans le cycle des folates. Ce processus biologique génère les substrats nécessaires à la formation des purines, des thymidines, de la cystéine, des protéines et à la méthylation¹⁴³. À travers ce cycle, l'acide folique est transformé en 5-méthyl-tétrahydrofolate. Sous cette forme active, il devient un substrat de la transformation de l'homocystéine en méthionine et un donneur de carbones pour les purines et les thymidines. La méthionine est ensuite impliquée dans la formation des protéines et de S-adenosylmethionine (SAM), un donneur de méthyl nécessaire à la méthylation¹⁰³. Le cycle de l'acide folique est présenté en détails dans la section 1.2.4.1.

Le rôle de l'acide folique dans le développement des ATN demeure inconnu, mais la formation des thymidines et la méthylation sont les principaux processus suspectés. Lors de la formation du tube neural, une grande quantité de nucléotides est nécessaire à la réplication rapide de l'acide désoxyribonucléique (ADN) et un manque de nucléotides pourrait causer un ralentissement du développement du tissu¹⁴⁴. Cette hypothèse est supportée par la présence d'une déficience en thymidine chez le modèle d'ATN murin *Spotch* possédant une mutation homozygote dans le gène *paired box 3 (Pax3)*. Le fait que le phénotype puisse être rétabli par une

supplémentation en acide folique ou en thymidine démontre la spécificité du phénotype¹⁴⁴. Une souris knockout pour serine hydroxymethyltransferase 1 (*Shmt1*), un gène impliqué dans le cycle de l'acide folique entre la formation des thymidines et la reméthylation de l'homocystéine, entraîne des ATN combinées à une faible synthèse des thymidines en présence d'une consommation insuffisante d'acide folique¹⁴⁵. Cependant, cette théorie a aussi ses failles tel qu'observé chez la souris mutante pour Cbp/P300 interacting transactivator with Glu/Asp rich carboxy-terminal domain 2 (*Cited2*) qui développe une ATN mais qui n'a aucune différence de synthèse de thymidines lorsque comparée aux contrôles¹⁴⁶. Le deuxième mécanisme suggéré est à travers la méthylation associée au cycle du folate. La première preuve d'association provient de l'augmentation de la prévalence des ATN lors d'une disruption chimique de la méthylation^{147 148}. Ceci est supplémenté par le fait que chez deux modèles d'ATN murins insensibles à l'acide folique, une forte dose de méthionine a permis de réduire la prévalence des ATN^{149 150}. Les ATN ont aussi été associées à l'hypométhylation et à un gène impliqué uniquement dans la méthylation^{151 152}. La méthylation des protéines a aussi été suggérée comme importante pour la différenciation du tissu neural et la distribution de l'actine et de la tubuline durant la fermeture du tube neural^{153 154}. Cependant, un knockout de Methyl-tetrahydrofolate reductase (*Mthfr*) chez la souris entraînant une diminution de la méthylation n'a pas mené à des ATN limitant cette association¹⁵⁵.

Malgré l'effet marqué de l'acide folique, sa carence n'a jamais été directement associée aux ATN¹⁵⁶. Les carences en inositol quant à eux ont été associées à une prévalence accrue d'ATN¹⁵⁷. L'inositol est un sucre alcool impliqué comme messenger secondaire dans les cellules eucaryotes principalement sous forme d'inositol phosphate (IP). Il est aussi un composant important des lipides structuraux phosphatidylinositol phosphate (PIP)¹⁵⁸. Il a été démontré qu'une supplémentation en inositol durant la grossesse de souris *curly tail* (modèle de transmission oligogénique complexe associé au gène *Grhl3*) résistante à l'acide folique permettait de diminuer la prévalence d'ATN^{159 160}. Le mécanisme d'action proposé est une correction de la sous-prolifération des tissus ventraux sous-tendant la fermeture du tube neural^{161 162}. Ce mécanisme est controversé étant donné un changement de paradigme récent dans la perception du rôle de *GRHL3* dans les ATN, l'associant plus fortement à la différenciation des cellules de la plaque neurale et à la TEM^{73 163}. De plus, l'absence d'effets de l'inositol sur une souris *Grhl3*^{-/-} supporte un mécanisme indépendant de ce gène¹⁶⁴. Chez l'humain, une étude chez 12 mères ayant précédemment eu un grossesse avec un ATN a démontré une réduction de la récurrence attendue¹⁶⁵. Une concentration sanguine d'inositol 5% plus basse que la moyenne a été

démontrée chez 63 femmes avec des grossesses affectées par une ATN¹⁶⁶. Le lien biochimique de l'inositol et des ATN est associé à la production de diacylglycérol par l'action de la phospholipase C sur les phosphatidylinositols entraînant l'activation des isoformes $\beta 1$, de la protéine kinase C et une progression du cycle cellulaire¹⁶⁷.

De faibles associations avec la prise péri-conceptionnelle d'autres nutriments ont aussi été reportées. Il a été suggéré que l'homocystéine, la vitamine B12, la choline et la bétaine pourraient toutes être impliquées dans une réduction du risque de développer des ATN^{168 169 170}. Cependant, plusieurs articles n'ont trouvé aucune association entre ces facteurs ne permettant pas de conclure sur leur importance. Plusieurs facteurs ont aussi été reconnus pour augmenter le risque de développer des ATN. L'acide valproïque est reconnu pour son effet tératogène entraînant une augmentation de 10x des risques d'ATN¹⁷¹. Le mécanisme demeure incertain mais est potentiellement lié à l'hyperacétylation de protéines entraînant des problèmes de stabilité¹⁷¹. L'hyperglycémie maternelle est aussi associée à une prévalence accrue d'ATN à travers une mort cellulaire accrue dans le neuroépithélium¹⁷². L'obésité maternelle et l'hyperthermie ont aussi été observées comme facteurs de risque, mais les mécanismes demeurent inconnus^{173 174}.

1.2.4 Facteurs génétiques

L'implication de facteurs environnementaux n'est qu'un aspect de l'étiologie complexe des ATN. Plusieurs articles démontrent une forte implication de facteurs génétiques dans le développement de la maladie et estime une héritabilité de 60%¹⁰³. L'héritabilité a été calculée à partir de la prévalence moyenne et du risque de récurrence dans les fratries à l'aide de la méthode de Falconer¹⁷⁵. Ces valeurs ont été obtenues par l'étude de 59 561 naissances dans un hôpital de Salt Lake City¹⁷⁶. La récurrence observée chez des fratries varie de 2 à 5% d'ATN, ce qui représente une augmentation de 50x par rapport à la population¹⁷⁷. L'explication de facteurs environnementaux a été rejetée en démontrant qu'une augmentation aussi forte de la prévalence nécessiterait un tératogène aussi puissant que la thalidomide, un événement peu probable¹⁷⁸. De plus, une analyse de 363 patients affectés a démontré que 8.5% des individus affectés proviennent de familles avec des antécédents d'ATN¹⁷⁹. Cependant, les paires affectées sont au deuxième ou troisième degré de séparation suggérant une transmission complexe et oligogénique¹⁷⁹. Des études de jumeaux critiquées pour leurs petits échantillons ont aussi démontré une concordance variant entre 3.7% et 18%¹⁷⁷. La ségrégation de la maladie avec des syndromes génétiques tel que Meckel et Mohr supporte aussi cette hypothèse¹⁷⁷. Une forte

association à l'aneuploïdie a aussi été établie avec 5 à 17% des cas d'ATN affectés et une fréquence élevée de trisomie 13 et 18^{180 181}.

La transmission complexe et le potentiel léthal de la pathologie ont nui à l'identification de larges familles affectées limitant ainsi les études génétiques traditionnelles. Cependant, deux études de liaison ont tout de même permis d'identifier des régions d'intérêts sur les gènes 2, 7, 10^{182 183}. Ces résultats largement complexifiés par l'hétérogénéité génétique et l'aspect oligogénique de la maladie n'ont à ce jour mené à l'identification d'aucun gène causatif¹⁰³. Les études ayant mené à l'identification de gènes impliqués dans l'étiologie se déclinent plutôt en trois segments : L'étude de la voie des folates l'étude et les études de gènes candidats chez la souris (incluant la voie de la polarité cellulaire planaire).

1.2.4.1 La voie des folates

Étant donné la grande importance de l'acide folique dans la prévention des ATN, la voie des folates a été fortement étudiée dans l'étiologie de ces maladies. L'acide folique est principalement absorbé par l'intestin grêle à l'aide des récepteurs solute carrier family 19 member 1/2, solute carrier family 46 member 1 (SLC19A1, SLC19A2 et SLC46A1)¹⁸⁴. Par l'enzyme dihydrofolate réductase (DHFR), il est transformé en tétrahydrofolate qui servira à la synthèse des purines et à la formation de 5-10-methyl-tétrahydrofolate par l'enzyme SHMT. Ce nouveau produit sera quant à lui la base de la synthèse des thymidines par l'enzyme thymidylate synthétase (TYMS) et sera transformé par les enzymes méthylène-tétrahydrofolate deshydrogenase, cyclohydrolase, formyltétrahydrofolate (MTHFD1) et méthylènetétrahydrofolate réductase (MTHFR) en 5-methyl-tétrahydrofolate (*5meTHF*), la principale source d'acide folique circulant dans le volume sanguin. Le *5meTHF* est ensuite absorbé par les récepteurs du folates 1 et 2 (FOLR1/FOLR2). À l'intérieur de la cellule, il est utilisé par l'enzyme 5-methyltétrahydrofolate-homocystéine methyl transferase (MTR) et sa réductase (MTRR) afin de remétyler l'homocystéine en méthionine et retourne sous la forme de tétrahydrofolate initial. La méthionine est ensuite transformée en SAM, un donneur de méthyl impliqué dans la majorité des procédés de méthylation. Le S-adenosylhomocystéine (SAH) résultant est retransformé en homocystéine qui attendra sa reméthylation¹⁰³.

Plusieurs variants ont été identifiées dans ce processus cyclique complexe, mais le principal demeure c.677C>T du gène *MTHFR*¹⁸⁵. Ce variant réduit de 65% l'activité de *MTHFR* et

possède un rapport des chances de 1.7 de développer une ATN comparé aux non-porteurs^{186 187}. Cependant, certaines populations ne démontrent aucune association entre les ATN et ce variant suggérant une pénétrance incomplète¹⁸⁸. Le variant c.1298A>C réduit lui aussi l'activité de *MTHFR*, mais une seule étude a démontré une association aux ATN malgré une absence de changement dans les niveaux d'homocystéine^{189 190}. Au niveau des récepteurs, aucune association aux ATN n'a été faite, mais une ablation de l'orthologue du récepteur *FOLR1* chez la souris résulte en une mort embryonnaire¹⁹¹. Le variant c.80A>G sur le gène *SLC19A1* a été associé à une prévalence accrue d'ATN qui semble interagir génétiquement avec les variants de *MTHFR*¹⁹². Un variant substituant une arginine à une glycine (c.1958G>A) sur le gène *MTHFD1* a été associé aux ATN par un risque accru de transmission maternel, mais n'a démontré aucune association chez les patients^{193 194}. Une répétition de 28 pb dans la région amplificatrice (Enhancer) ainsi qu'une délétion de 6 pb dans la région non-transcrite 3' chez le gène *TYMS* a aussi été associé aux ATN¹⁹⁵. Des études ont aussi démontré l'association de ces mutations à une quantité accrue d'homocystéine plasmatique¹⁹⁶, mais d'autres études ont été incapables de répéter ces observations^{197 198}. Une délétion de 19 pb dans l'intron de *DHFR* a aussi été associée aux ATN chez les mères d'enfants affectés¹⁹⁹. Pour le gène *SHMT*, le variant c.1420C>T a été défini comme un facteur protecteur des ATN²⁰⁰, mais ces résultats n'ont pas été répliqués²⁰¹. Les gènes de reméthylation de l'homocystéine ont aussi été associés aux ATN. Le variant c.2756A>G du gène *MTR* a été hautement étudié et a été associé à une prévalence accrue d'ATN^{202 203}, à une absence d'association^{204 205} et même à un effet protecteur²⁰⁶. Le gène *MTRR* a quant à lui été associé aux ATN à travers une méta-analyse de plusieurs études du variant c.66A>G²⁰⁷. Le rapport des chances de 2.1 de développer une ATN chez les porteurs versus les non-porteurs démontre un effet marqué de ce polymorphisme d'un nucléotide (SNP) et en fait un des candidats les plus forts de la voie des folates. L'implication de cette voie est aussi suggérée au niveau immunitaire où près de 75% des femmes ayant eu une grossesse affectée par les ATN produisent des anticorps auto-immuns bloquant les récepteurs folates²⁰⁸. Ces observations sont cependant contestées par une étude n'ayant trouvée aucun lien entre la présence de ces anticorps chez la mère et les ATN²⁰⁹.

Ce grand nombre de variants identifiés dans la voie des folates soutient son importance dans l'étiologie génétique de la maladie. Cependant, la pénétrance incomplète de la plupart d'entre eux suggère qu'un effet combinatoire entre ces mutations est peut-être nécessaire à l'expression d'un phénotype d'ATN. La présence d'effets parfois contradictoires entre différentes études supporte un rôle majeur de l'environnement et de l'ethnicité dans l'expressivité de ces

variants. Des études répliquant ces résultats sont nécessaires afin de mieux comprendre l'effet de ces mutations.

1.2.4.2 Étude des gènes candidats chez la souris

Plus de 250 gènes ont été identifiés comme causatifs d'ATN chez des souris modèles^{210,211}. Ces études offrent des avantages majeurs sur les études humaines précédentes. Elles permettent d'étudier en détail le phénotype et l'implication du gène dans ce dernier. De plus, la similarité du phénotype murin et humain en fait un excellent modèle biologique. La large majorité de ces modèles ont été construits par l'ablation de gènes, par des pièges géniques et par des mutations aléatoires à l'aide de N-ethyl-N-nitrosourea (ENU)¹⁰³. Ceci a permis de cribler une large proportion de gènes résultant en un portrait très complet de la pathologie chez la souris.

1.2.4.2.1 La voie de la polarité cellulaire planaire

La voie de signalisation de la polarité cellulaire planaire (PCP) a elle aussi été associée aux ATN. C'est l'identification d'un variant sur le gène Van Gogh-like protein 2 (*Vangl2*) chez le modèle murin LP qui a pour la première fois confirmé une association de cette voie avec les ATN. L'identification subséquente de variants potentiellement pathogéniques chez *VANGL1* dans des familles affectées par les ATN a permis de renforcer cette association^{212 213}. Étant l'orthologue du gène *Stbm* chez la drosophile, *VANGL1* est impliqué dans la localisation de *PRICKLE1* et permet l'exclusion de DVL²¹⁴. Cette liaison aux trois isoformes de *Dvl* murin est compromise par le variant p.Val239Ile identifié par le séquençage du gène dans une cohorte de 171 affectés²¹³. L'effet pathogénique des variants p.Val239Ile et p.Met328Thr a aussi été démontré à l'aide d'une étude de leurs fonctions chez le poisson zèbre²¹⁵. Des variants associés aux ATN fermées ont été identifiés dans des régions conservées du gène *VANGL2* et subséquemment validés^{216 217}. Cependant, l'implication de ce gène est fortement supporté par les études chez la souris^{212 218}.

Afin d'approfondir l'association de la voie PCP et des ATN, d'autres membres de la voie ont été investigués. Le gène *FZD* possède 10 isoformes et quoi qu'une redondance est observée entre les gènes, une ablation de *FZD3* et *FZD6* cause un craniorachischisis chez 100% des souris homozygotes^{219 220}. L'étude de ces gènes chez l'humain a identifié 5 variants intéressants dans *FZD6*, incluant l'insertion *de novo* p.Cys615X, mais n'a à ce jour rien identifié dans le gène *FZD3*. Aucune validation fonctionnelle n'a été effectuée. *DVL1*, *DVL2*, *DVL3* ont été principalement associés aux ATN par l'étude de la souris *Dvl2*^{-/-} causant un spina bifida²²¹. L'ablation de *Dvl1* et *Dvl3* n'a pas été associée aux ATN, mais la combinaison de *Dvl1*^{-/-} et *Dvl2*^{-/-} résulte en un

craniorachischis démontrant une interaction génétique²²¹. Le séquençage de ces gènes chez une cohorte de 453 patients a permis d'identifier six variants intéressants sur *DVL1*, cinq sur *DVL2* et trois sur *DVL3*^{214 222}. Tous ces variants identifiés chez des patients atteints d'ATN fermées sont faux-sens, sauf l'insertion p.Glu620X identifiée sur *DVL2*. Aucune validation fonctionnelle de ces variants n'a été effectuée. Le gène *Fuzzy (FUZ)* impliqué dans le maintien de la position cellulaire de DVL a aussi été associé aux ATN par une ablation chez la grenouille résultant en un craniorachischis²²³ et chez la souris résultant en une exencéphalie²²⁴. Le séquençage de ce gène chez 234 patients a permis d'identifier trois variants chez des patients avec des ATN fermées. Deux d'entre eux affectent la formation des cils et le mouvement cellulaire supportant leur effet pathogénique et l'association de ce gène aux ATN humaines²²⁵. Pour *PRICKLE1*, une cohorte de 364 patients a été séquencée et neuf mutations intéressantes ont été identifiées²²⁶. De ces variants, p.Arg682Cys a démontré un effet compatible avec un dominant négatif dans une validation chez le poisson zèbre²²⁶. L'orthologue de *fmi* connu sous le nom de cadherin EGF LAG seven-pass G-type receptor (*CELSR1, CELSR2, CELSR3*) a été associé aux ATN par une souris *Celsr*^{-/-} causant un craniorachischis²²⁷. Le séquençage de ce gène dans 473 patients a identifié 27 variants intéressants dont la délétion c.8887_8898del12 et le non-sens p.Gln834X²²⁸. Des validations fonctionnelles de ces variants sont nécessaires afin de conclure sur leur implication.

Les interrupteurs moléculaires modulant la transition entre la voie canonique et non-canonique de WNT ont aussi été associés aux ATN. Le gène *PTK7* a été associé aux ATN à travers une souris *Ptk7*^{-/-} résultant en un craniorachischis²²⁹. Ce membre de la voie PCP est unique aux mammifères et semble agir comme un interrupteur moléculaire entre les voies^{230 231}. Le séquençage de ce gène sur 473 patients a permis d'identifier six variants intéressants desquels un variant agissait comme hypermorphe dans le poisson zèbre²³². L'orthologue de *dgo* connu sous le nom de ankyrin repeat domain 6 (*ANKRD6*) a été associé aux ATN par l'identification de 6 variants dans une cohorte de 473 patients dont p.Pro548Leu, p.Arg632His affectent sa fonction d'interrupteur entre les deux voies^{233 234}. Finalement, LDL Receptor Related Protein 6 (*LRP6*) a aussi été associé aux ATN à travers le variant p.Ile681Arg identifié chez la souris qui interagit génétiquement avec *Vangl2* causant une prévalence de spina bifida de 36%²³⁵. Les validations fonctionnelles ont démontré une perte de la fonction d'interrupteur moléculaire²³⁵. Chez l'humain, le séquençage de 285 patients a permis d'identifier trois variants desquels p.Val1386Leu a été associé à une activité canonique réduite et une activité non-canonique accrue²³⁵. Une étude subséquente sur 192 patients a identifié 3 variants (p.Ala3Val, p.Tyr544Cys, et p.Arg1574Leu)

entraînant une augmentation de l'activité WNT non-canonique. L'effet de ces variants sur l'activité WNT canonique est cependant variable avec une diminution pour p.Tyr544Cys, une augmentation pour p.Arg1574Leu et un effet neutre pour p.Ala3Val. Ces effets divers sur la voie canonique suggère que les variants fonctionnels peuvent causer une prédisposition ou être la cause des ATN en fonction de leur positionnement sur la protéine²³⁶.

La forte association de la voie PCP et du craniorachischisis supporte une importance de cette voie dans le point de fermeture 1 chez la souris. Cependant, les variants identifiés chez l'humain proviennent d'ATN diverses et une faible proportion co-ségrègent avec des variants dans la voie PCP. Ainsi, il semble que la voie PCP n'explique qu'une faible proportion des patients atteints d'ATN, mais démontre néanmoins un rôle clef dans le développement de la pathologie.

1.2.4.2.2 Survol des associations identifiées dans le modèle murin

Les études du modèle murin le potentiel d'investiguer l'interaction gène-environnement dans le cadre de l'effet protecteur de l'acide folique et de l'inositol. Ainsi, elles ont déterminé que les souris avec un ablation de *Cart1*²³⁷, *Cited2*¹⁴⁶, *Folr1*²³⁸, *Gcn5*²³⁹ ou un variant pathologique sur *Lrp6*²⁴⁰, *Pax3*²⁴¹ avaient une pénétrance d'ATN diminuée en présence d'acide folique. D'autres nutriments ont aussi été associés à une réduction de la prévalence d'ATN tel que la méthionine chez la souris *Axd*^{-/-242}, l'inositol¹⁶⁰ chez la souris curly tail souvent associée à *Grhl3* et le Purina 5001 chez la ligné de souris SELH/Bc²⁴³. Il est cependant intéressant de noter que malgré l'association de curly tail, une ablation directe du gène *Grhl3* n'a démontré aucune sensibilité à l'inositol¹⁶⁴.

À travers la large quantité de gènes causatifs, certains processus biologiques ont été associés aux ATN. Les protéines du cytosquelette ont été associées à travers les gènes *Abl1*, *Abl2*²⁴⁴, *Cofilin*²⁴⁵, *Grlf1*²⁴⁶, *Mapk8*, *Mapk9*²⁴⁷, *Marcks*²⁴⁸, *Mena*, *Profilin*²⁴⁹, *Vasp*²⁵⁰, *Mlp*²⁵¹, *Palladin*²⁵², *Shrm*²⁵³, *Vcl*²⁵⁴. Le cycle cellulaire et la neurogenèse sont aussi importants dans la pathogénèse des ATN tel que démontré par *Brd2*²⁵⁵, *Hes1*²⁵⁶, *Hes3*²⁵⁷, *Jmj*²⁵⁸, *Nf1*²⁵⁹, *Notch3*²⁶⁰, *Numb*²⁶¹, *Nup50*²⁶², *Pax3*²⁶³, *Pax7*²⁶⁴, *RBP-Jk*²⁶⁵, *Ptprs*²⁶⁶. La mort cellulaire joue un rôle clair à travers les gènes *Ap2*²⁶⁷, *Apaf1*²⁶⁸, *Bcl10*²⁶⁹, *Brca1*²⁷⁰, *Gadd45*²⁷¹, *Casp3*²⁷², *Casp9*²⁷³, *Chuk*, *Ikbkb*²⁷⁴, *Mekk4*²⁷⁵, *Tcof1*²⁷⁶, *Trx2*²⁷⁷. Il est aussi clair que la transcription génétique a un rôle très important dans la pathologie des ATN avec près d'une quinzaine de co-activateurs (ex : *Cecr2*, *Smarcc1*, *Xnp*) et une vingtaine de facteurs de transcription (ex : *Grhl3*, *Zic2*)¹⁸⁵ causant des ATN suivant leur ablation chez la souris.

La souris représente un outil puissant pour les études mécanistiques de la pathogénèse des ATN. En plus d'identifier des gènes clefs de la pathologie, la souris permet d'investiguer des processus pathologiques des ATN. Ainsi, le rôle des micro acides ribonucléiques (microARN) dans l'étiologie des ATN²⁷⁸ et le rôle de certaines voies de signalisation dans le développement de différents points de fermetures ont été étudiés¹⁸⁵. L'importance de ce modèle dans les validations fonctionnelles permet parfois même de soulever des questions en absence de phénotype. Ainsi, l'absence d'ATN chez la souris *Mthfr*^{-/-} soulève des doutes sur l'importance du gène dans la pathologie humaine¹⁵⁵.

1.2.5 Santé publique

La santé publique a pour but de minimiser les coûts et de maximiser la qualité de vie des habitants d'un pays. Étant donné le fort impact social et personnel des ATN, elles sont un excellent candidat pour ce type de politique. La Chine a été le premier pays à suggérer une prise péri-conceptionnelle de 400ug/jour d'acide folique entraînant une réduction de la prévalence d'ATN de 80%²⁷⁹. Cette étude a confirmé le rôle de l'acide folique dans la prévention des ATN et en a fait une excellente option pour des politiques de santé publique. C'est ainsi que le Canada et les États-Unis ont décidé de mettre en place une politique similaire. Cependant, afin d'augmenter le niveau de protection, ils ont implémenté une politique de fortification des aliments communs. Aujourd'hui, le niveau de fortification au Canada est de 0.15mg par 100g de farine (0.14mg aux États-Unis) et de 0.27mg par 100g de pâtes (conforme aux États-Unis). Les États-Unis fortifient aussi le riz à un taux de 0.308mg par 100g. Les céréales sont fortifiées à un taux de 400ug aux États-Unis et de 60ug au Canada. Les doses décrites visent à ne pas dépasser la dose quotidienne maximale de 1000ug chez un adulte¹⁰².

Au Canada, cette supplémentation a été accompagnée d'une réduction totale des ATN de 46% en 9 ans. Cette réduction touche principalement le spina bifida avec une réduction de 53%. Cette réduction diminue à 38% pour les anencéphalies et 31% pour les encéphalocèles¹⁰². En absence d'une augmentation de consommation de suppléments, la concentration moyenne de folate sanguin a augmenté de 527nM à 741nM^{280 281}.

L'impact économique de cette politique est majeur. Un enfant atteint d'un myeloméningocèle coûtait 70 000\$ par année en 2001 en plus de près de 5 chirurgies par année durant les 5 premières années de vie aux États-Unis^{177 282}. Pour les 20 premières années de vie

d'un patient, ce coût s'élève à presque 1.4 millions dans le système privé des États-Unis. Lorsque les ATN sont étudiées en groupe, ce coût est légèrement plus bas à presque 620 000\$ dans un contexte public Européen²⁸³. Considérant une réduction de prévalence des ATN de 1.58 :1000 à 0.86 :1000 et une population canadienne de 35 millions d'habitants, il est possible de calculer que le Canada a économisé jusqu'à 15 milliards par années¹⁰². Quoique ce chiffre puisse être plus petit étant donné les différences du système de santé canadien et l'évolution des techniques médicales, il offre une perspective unique sur l'importance de diminuer l'incidence de cette maladie et l'importance de son étude. L'implémentation de tests génétiques efficaces ou de nouvelles politiques de fortification en nutriments protégeant contre les ATN méritent une considération importante.

1.3 Le gène *SHROOM3*

1.3.1 Association aux ATN et à d'autres pathologies

Shroom3 a été caractérisé chez une souris affectée par les ATN et provenant d'une série de pièges géniques aléatoires. Cette souris meurt de manière prénatale et le phénotype suit une transmission homozygote. À E10.25, une ouverture du tube neural s'étendant rostralement à partir de la vésicule otic est observé chez 100% des embryons et engendre une exencéphalie (version murine de l'anencéphalie). Certaines souris ont aussi un échec de fermeture caudale engendrant un spina bifida et un échec de fermeture ventrale engendrant un gastroschisis. Le séquençage du piège génique a permis d'identifier deux isoformes de *Shroom3* : *ShrmS* (sans domaine PDZ) et *ShrmL* (pleine longueur). L'initiation de l'isoforme *ShrmS* est en aval du piège génique permettant de conclure que l'isoforme court est insuffisant pour compenser la perte de l'isoforme long²⁵³. Cette souris a aussi été étudiée dans le cadre des maladies chroniques des reins. Un affaissement des glomérules rénaux causé par un problème dans les podocytes est observé²⁸⁴. Il est intéressant de noter que ce phénotype affecte uniquement les souris homozygotes ne permettant pas d'étudier l'évolution du phénotype chez l'adulte.

Depuis ces études, *SHROOM3* a été impliqué dans différentes pathologies et processus biologiques. Par association du génome entier, il a été impliqué dans les fentes labio-palatines (FLP)²⁸⁵. De la même manière, il a été impliqué dans la microsomie cranio-faciale²⁸⁶. Son implication dans la fermeture du tube neural crânien en fait un candidat intéressant pour ces malformations crâniennes. *SHROOM3* a aussi été associé aux maladies rénales^{287 288 289}. Une méta-analyse d'associations du génome entier a permis aussi de démontrer un lien entre *SHROOM3*

et le développement des érythrocytes²⁹⁰. Ceci est surprenant considérant l'absence d'expression de ce gène dans la moelle osseuse. Il a aussi été impliqué dans le développement des neurones²⁹¹²⁹² et dans l'hétérotaxie²⁹³.

1.3.2 Localisation et expression

SHROOM3 est une protéine principalement connue pour son rôle dans la CA, mais il est aussi important dans l'épaississement des tissus²⁹⁴. Chez le *Xenopus*, *Shroom3* apparaît durant la transition blastulaire. Il a été confirmé que *Shroom3* est exprimé du côté dorsal de l'ectoderme et du mésoderme durant la gastrulation. Lors de la neurulation, son expression est restreinte à la partie superficielle de la plaque neurale et concentrée au niveau antérieur de l'embryon. Suite à la fermeture du tube neural, *Shroom3* est exprimé dans la placode otic nécessaire au développement de l'oreille et dans la placode olfactive. Comme *Shroom1*, il est aussi exprimé dans le cœur, les reins, le mésencéphale, le prosencéphale et dans le bourgeon de la queue chez la grenouille²⁹⁵. Chez la souris à E8.5, l'expression semble plutôt restreinte au neuroépithélium cranial, le tube neural, le mésoderme paraxial et l'intestin. Elle migre vers le tube neural, le prosencéphale, les somites, le cœur et l'intestin à E10.5. Par E14.5, cette expression se limite aux muscles squelettiques, les régions distales des doigts et le prosencéphale²⁵³. Cette étude manque la granularité du *xenopus*, mais confirme la conservation de l'expression à travers l'évolution. L'expression de ce gène a aussi été démontrée dans les placodes de la lentille²⁹⁶.

1.3.3 Implication dans la constriction apicale et dans l'asymétrie

SHROOM3 est une protéine d'ancrage avec un domaine PDZ, un domaine suggéré EVH1, un domaine ASD1 et un domaine ASD2. Deux isoformes sont connus chez la souris, *ShrmL* et *ShrmS*, mais un seul isoforme est présent chez l'humain (Figure 8)²⁵³. Les domaines PDZ et ASD1 sont impliqués dans la position cellulaire du gène, alors que ASD2 active la CA. Cette activation est médiée par les protéines *Rock1* et *Rock2* recrutées aux jonctions apicales⁶². Les *ROCK* sont des Ser/Thr kinases qui lient *RHOA*, préalablement activé par *RHOGEF2*, et phosphorylent les myosines régulatrices des chaînes légères (MCL)²⁹⁷. À travers des repositionnements d'actine et de gamma-tubuline attribués au domaine ASD1 de *Shroom3*, un anneau d'actine-myosine est créé à la surface apicale des cellules^{253 298 299}. Les MCL activent la myosine2 qui induit une contraction de cet anneau et du réseau apical d'actine permettant la réduction de la surface apicale (Figure 9)³⁰⁰. La localisation est régulée par les domaines PDZ et ASD1 qui assurent la liaison à la F-actine²⁹⁸. Sa localisation planaire dans la plaque neurale semble quant à elle associée à la voie

PCP. *Shroom3* interagit génétiquement avec les composants de la voie PCP *Vangl2* et *Wnt5a*, alors que *Dvl2* dirige le recrutement de *Shroom3* associé à *Rock*. Cette interaction se fait à travers le domaine DEP de *Dvl2* et une région non caractérisée de *Shroom3* se situant entre les acides aminés 286 et 776. Cette interaction permet alors d'enrichir des interfaces cellulaires spécifiques orientées dans l'axe médio-latéral. En absence de *Dvl2*, *Shroom3* peut tout de même aller aux jonctions, mais le positionnement médio-latéral est perdu suggérant que *Shroom3* est en aval de la voie PCP³⁰¹. Des poissons zèbres traités par un morpholino contre *Vangl2* échouent à la CA supportant le rôle de la voie PCP dans ce processus³⁰².

Dans le cadre des tissus non-neuraux, Pax6 est capable d'induire l'expression de *Shroom3* dans les placodes de lentilles murines créant ainsi un lien entre la machinerie de morphogénèse

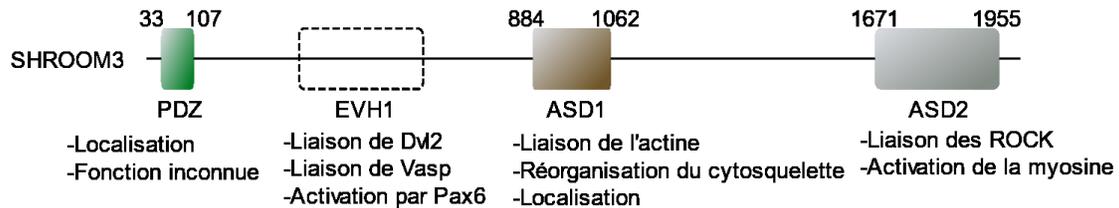


Figure 8 : Représentation graphique des domaines fonctionnels du gène *SHROOM3* humain et de leur fonction. L'encadré transparent représente le domaine prédit EVH1.

de ces cellules et le gène impliqué dans leur épaissement²⁹⁶. Cette régulation s'effectue à travers le domaine EVH1 de *Shroom3* qui permet aussi d'assurer la localisation de vasodilator-stimulated phosphoprotein (Vasp) une autre protéine impliquée dans la CA²⁹⁶. Il a aussi été démontré que le facteur paired like homeodomain 1 (*Pitx1*) est nécessaire à l'expression de *Shroom3* dans l'intestin de *Xenopus* à travers une séquence promotrice en amont du gène³⁰³. Finalement, cell division cycle 42 (*Cdc42*) est impliqué dans la polarité cellulaire de l'intestin et dans la formation des organes. Il semble être nécessaire au maintien de la polarité de *Shroom3*³⁰⁴.

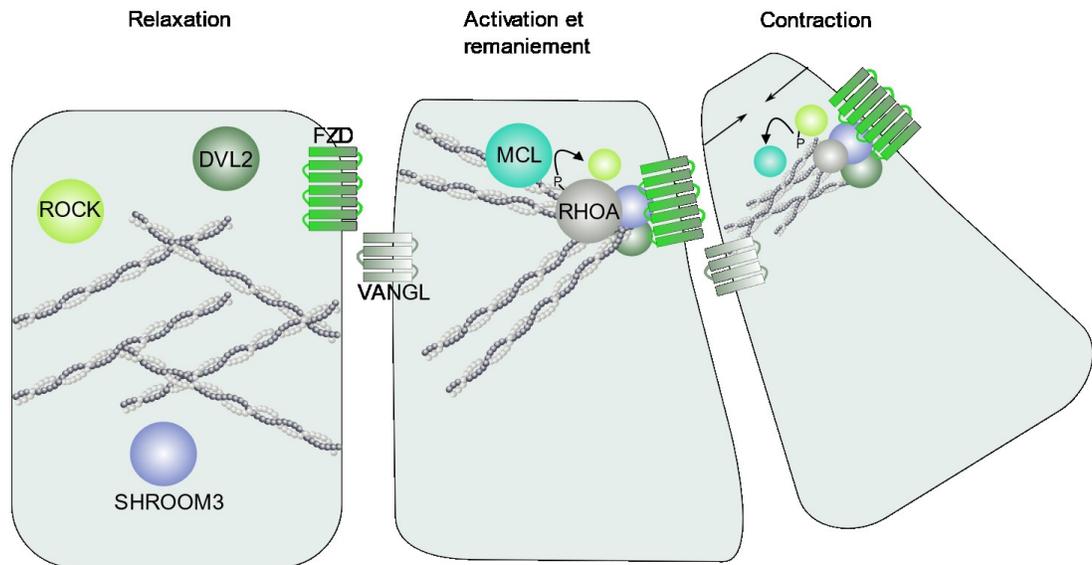


Figure 9 : Représentation graphique de la localisation et des remaniements nécessaires aux changements conformationnels impliqués dans la CA. Dans la forme relaxée, l'actine maintient la forme colonnaire des cellules. SHROOM3 est ensuite recruté aux jonctions apicales et lie DVL2 en entraînant un remaniement de l'actine pour former un anneau contractile. Finalement, l'activation des ROCK et de MCL entraîne l'activation de la myosine induisant la contraction.

1.3.4 Membres de la famille Shroom

La famille des *Shroom* est composée de 4 membres. Ces gènes sont connectés par la présence de domaines ASD spécifiques à cette famille²⁹⁸. Ces gènes ont été connus sous le nom de *Apx* (*Shroom1*), *Apxl* (*Shroom2*), *Shrm* (*Shroom3*) et *kiaa1202* (*Shroom4*)²⁹⁸. Excepté pour *Shroom3*, les fonctions des protéines de cette famille demeurent encore largement inconnues. Cependant, chacune de ces protéines ont démontré une capacité à interagir avec l'actine. De plus, *Shroom2* et *Shroom4* ont la capacité intrinsèque d'induire la CA, mais leur localisation les empêchent²⁹⁸. En effet, *Shroom2* lie principalement l'actine corticale, alors que *Shroom4* lie des populations d'actines différentes de *Shroom2* et *Shroom3*²⁹⁸. Il n'y a aucune étude sur le type d'actine lié par *Shroom1*. Cependant, il lie la protéine de surface Melanoma cell adhesion molecule (MCAM) et régule les microtubules à travers la gamma-tubuline²⁹⁹.

Une étude de l'expression de ces gènes dans le xenopus a démontré plusieurs spécificités. Durant la neurulation, *Shroom1* et *Shroom2* sont exprimés dans la partie profonde de la plaque

neurale, alors que Shroom4 est restreint aux somites. Shroom1 et Shroom2 sont exprimés dans l'ARN maternelle chez le xenopus, alors que Shroom3 apparaît durant la transition blastulaire et que Shroom4 apparaît encore plus tard. Par la suite, Shroom1 est exprimé dans le notochorde, le mésoderme paraxiale et les somites. Il se retrouve aussi dans la placode otic et les conduits pronéphriques. Shroom2 est principalement exprimé de manière antérieure durant la neurulation. Il est ensuite exprimé dans le prosencéphale, le mésencéphale, les yeux, la placode otic, les reins, le notochorde et l'épithélium menant aux pigments oculaires²⁹⁵. L'étude de la fonction de ce gène suggère une implication dans la contractilité cellulaire nécessaire à la migration et indépendance de la CA³⁰⁵. Les problèmes de pigmentation de l'œil associés à ce gène peuvent alors s'expliquer par un échec de migration dans l'épithélium pré-oculaire³⁰⁶. Shroom4 est exprimé dans le mésoderme ventral, les yeux, les conduits pronéphriques, la vésicule otic et la placode olfactive. Cependant, son seuil d'expression principal est les somites de la région postérieure au cours du développement²⁹⁸. Ce gène a été associé au retard mental lié à l'X, mais le processus pathologique demeure inconnu³⁰⁷. Les spécificités d'expression des Shroom au niveau des divers tissus et même au niveau des différentes actines démontrent un mécanisme non redondant entre les membres de cette famille. Cependant, l'importance de ces protéines au niveau du remaniement cellulaire de l'actine combiné à leur forte expression au niveau de la plaque neurale en font des candidats intéressants qui devraient être étudiés dans le cadre des ATN.

1.4 Le gène *GRHL3*

1.4.1 Association aux ATN et à d'autres pathologies

GRHL3 a été caractérisé suite à l'identification de *GRHL1* et *GRHL2* par l'analyse de la «high throughput genomic sequence database». Il s'agit d'un facteur de transcription impliqué dans la régulation d'une large quantité de gènes. Sa similarité au gène *grainyhead* identifié chez la drosophile lui a valu le nom *GRHL*³⁰⁸. Les domaines protéiques de trans-activation, de liaison de l'ADN et de dimérisation rappellent ceux de *GRHL1* et *GRHL2* (Figure 10). Dans cet article, les auteurs identifient 3 isoformes de *GRHL3* et démontrent une différence au niveau de leur patron d'expression chez l'humain. Les trois isoformes sont présent dans le pancréas et le cerveau. L'isoforme 1 est exprimé dans la prostate, le placenta, le colon, le rein et les ganglions lymphatiques. L'isoforme 2 est exprimé dans les amygdales, le thymus, le placenta et les reins³⁰⁸. La création de la souris *Grhl3*^{-/-} a permis de lier *Grhl3* aux ATN. Un spina bifida est observé chez

100% des embryons homozygotes avec 2% d'exencéphalie. La charnière médiane se forme normalement, mais le phénotype semble causé par une absence de soulèvement ou de rapprochement des plis neuraux¹⁶⁴. La souris hétérozygote possède une cavité crâniale significativement réduite, mais aucune ATN³⁰⁹. Une interaction génétique entre cette souris et la souris Curly Tail, un modèle de transmission complexe des ATN associé à la région contenant *Grhl3*, permet de supporter l'importance de ce gène dans la pathologie. De plus, une surexpression de ce gène permet de rétablir le développement de cette souris¹⁶². Le développement des ATN dans ces souris a été associé à un sous-développement de la région ventrale entraînant des tensions physiologiques contraignant la fermeture du tube neural. Ce mécanisme est controversé et souvent ignoré au profit de ses rôles dans la voie PCP, la fermeture du tube neural et la régulation de Shh discutés dans les sections 1.1.2 et 1.1.4¹⁶⁴. Le rôle réel de *Grhl3* dans les ATN demeure encore incertain, mais un variant sur le gène *LaminB1* augmente significativement la prévalence d'ATN dans la souris Curly Tail. Cela semble associé à son action sur la capacité proliférative et la morphologie nucléaire³¹⁰. Une meilleure étude de cette interaction permettrait potentiellement d'éclaircir le mécanisme de *Grhl3* dans la pathologie.

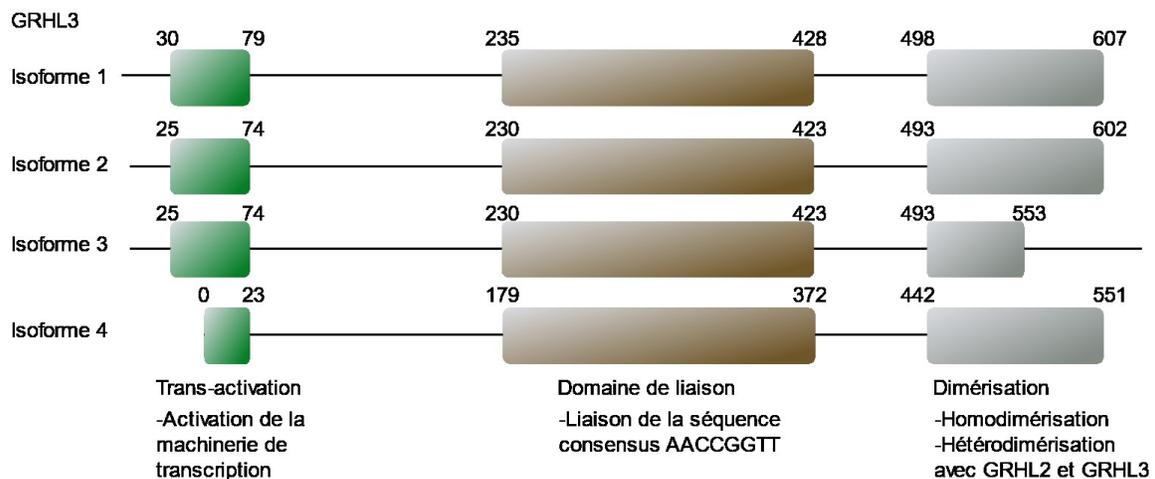


Figure 10 : Représentation graphique des 4 isoformes de GRHL3 et description de la fonction des différents domaines protéiques.

Grhl3 est aussi associé à la régulation de l'intégrité épidermale. Chez la souris *Grhl3*^{-/-}, une diminution des lamelles lipidiques et de l'enveloppe cornifiée sont observées entraînant une perte d'imperméabilité de l'épiderme³¹¹. La migration cellulaire nécessaire à la réparation des blessures cutanées est aussi compromise. *Grhl3* est aussi reconnu pour activer l'expression du

gène *uroplakinII* impliqué dans la différenciation urothéliale³¹². Chez l'humain, *GRHL3* est impliqué dans les fentes labiales, les fentes palatines, dans le syndrome de Van Der Woude 2 (SVW2) et dans la formation des papilles gustatives^{64 313 314}. Au niveau cognitif, le gène est important dans le développement des capacités locomotrices et d'inhibition des souris. Le processus biologique justifiant ce lien demeure encore largement inconnu³¹⁵.

1.4.2 Localisation et expression

L'expression de ce gène a été étudiée chez la souris. Il est fortement exprimé dans l'ectoderme de surface à partir de E8.5 jusqu'à E17.5. Il est ensuite fortement exprimé dans le cœur de E10.5 et cette expression diminue rapidement jusqu'à être nulle à E17.5. Il est faiblement exprimé dans les poumons à E15.5 et dans les reins à E17.5. Il s'exprime de manière croissante dans l'épithélium oral entre E10.5 à E17.5. Il est aussi exprimé de manière stable dans l'œsophage et le gros intestin entre E10.5 et E17.5. Il s'exprime fortement dans l'estomac entre E15.5 et E17.5, alors qu'il est complètement absent de l'intestin grêle. Finalement, il s'exprime de manière croissante dans la vessie entre E10.5 et E17.5³¹⁶.

1.4.3 Implication dans l'épiderme et régulation complexe

Le rôle de ce gène dans les ATN est principalement attribué à trois fonctions principales : Son implication dans la voie PCP, la fermeture du tube neural et la régulation de Shh. Les mécanismes d'actions de ces processus ayant été traités dans les sections 1.1.2 et 1.1.3 (Figure 11), cette section s'attardera plutôt aux autres aspects de la fonction et la régulation de *GRHL3*.

Grhl3 est particulièrement connu pour son implication dans le développement de l'épiderme. Cette association a initialement été attribuée au gène transglutaminase 1 (*Tgase1*) impliqué dans la liaison des protéines et des lipides dans l'épiderme³¹⁷. Cependant, une étude subséquente a démontré qu'un réseau beaucoup plus complexe de gènes est dérégulé en absence de *Grhl3*. Au total, 231 gènes sont dérégulés chez les souris *Grhl3*^{-/-}. De ceux-ci, 51 gènes sont fortement associés à des familles ou des voies de signalisation importantes dans la formation de l'épiderme. Les principaux regroupements étaient les protéines structurales (21 gènes), les molécules d'adhésion (9 gènes), les enzymes de modification des protéines et lipides (2 gènes incluant *Tgase1*), les inhibiteurs des protéases et des sérines (2 gènes), les enzymes et transporteurs impliqués dans le métabolisme des lipides (15 gènes). L'étude démontrait aussi qu'en absence du gène *Lmo4*, qui n'est pas régulé par *Grhl3*, le développement de l'épiderme est

encore plus compromis suggérant un lien entre ce gène et *Grhl3*³¹⁸. Ce processus est aussi impliqué dans la migration et la fusion nécessaire à la formation des paupières à travers l'activation de la voie epidermal growth factor receptor/extracellular signal-regulated kinases (EGFR/ERK) activée par Tgfa^{319 320}. Interferon regulatory factor 6 (*Irf6*) semble aussi jouer un rôle dans la régulation de l'expression de *Grhl3* dans le périoderme et l'épithélium oral des embryons de souris³²¹. C'est un facteur de transcription impliqué dans la différenciation de la peau, de l'épithélium mammaire et oral. Il a aussi été associé à l'expression de l'endothéline, un facteur primordial au développement cranio-facial chez le poisson zèbre³²².

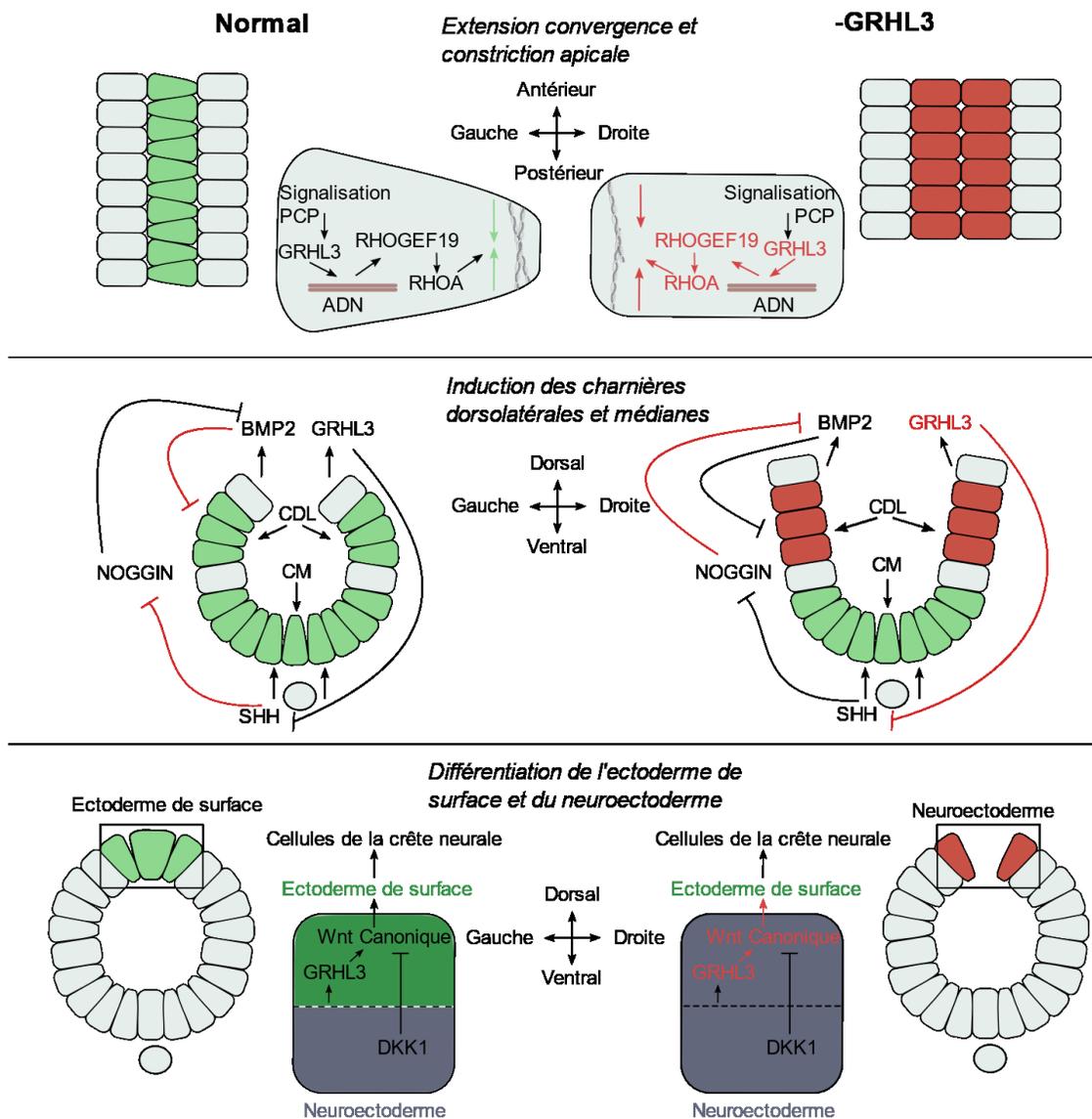


Figure 11 : Représentation graphique des principaux modèles impliquant GRHL3 dans les anomalies du tube neural. La colonne de gauche représente le processus normal et la colonne de droite représente le processus perturbé par une insulte majeure à GRHL3.

En plus de patrons d'expression différents, les isoformes de GRHL3 démontrent des interactions complexes qui semblent assurer une régulation spécifique de ce gène. Ces isoformes ont uniquement été identifiés chez l'humain et seul l'isoforme 2 est reporté chez la souris. Un total de 4 isoformes sont à ce jour identifiés, mais uniquement l'isoforme 1, 2 et 4 sont discutés dans la littérature. L'isoforme 3 possède un codon 3' alternatif qui modifie le domaine de

dimérisation et il partage le 5' de l'isoforme 2 (Figure 10). L'étude des 3 autres isoformes a permis de conclure que l'isoforme 1 active la migration et inhibe l'apoptose, alors que l'isoforme 4 réduit la migration et ne modifie pas l'apoptose. Une différence dans l'expression de ces isoformes peut être induite par l'expression d'acide nitrique (NO), un inducteur de migration cellulaire et un inhibiteur de l'apoptose, qui entraîne une surexpression de l'isoforme 1 et une diminution de l'isoforme 4³²³. Le NO semble donc agir à la manière d'une boucle d'activation étant donné que *GRHL3* est aussi impliqué dans sa phosphorylation à travers l'expression d'AKT serine/threonine kinase 1 (*AKT1*)³²⁴. Cette phosphorylation est nécessaire à l'activité anti-apoptotique de NO et est augmentée en présence de l'isoforme 1 et réduite en présence de l'isoforme 4. Ces deux isoformes peuvent tout deux activer l'expression génique, cependant seulement 28 gènes sont communs à cette activation entre ces deux isoformes, expliquant la large divergence dans leur fonction. L'isoforme 4 agit de manière pathogénique lorsque surexprimé dans le poisson zèbre suggérant qu'il peut potentiellement agir en tant qu'inhibiteur compétitif des autres isoformes³²³. La voie PCP ajoute une régulation supplémentaire tel que démontré par des études d'interactions génétiques entre des variants murins de *Grhl3*, *Vangl2*, *Scrb1* et *Celsr1*³²⁵. L'investigation de cette interaction suggère que *Grhl3* exprime *Rhogef19* qui à son tour activera l'effecteur de la voie PCP *Rhoa*. Cette interaction pourrait expliquer comment la régulation spatio-temporelle précise de *Grhl3* est effectuée³²⁵. Au niveau post-transcriptionnel, une régulation supplémentaire est attribuée au micro ARN miR21. Il est impliqué dans la répression de *Grhl3* à travers la signalisation leptine et les hormones de la thyroïde³²⁶. Cette répression de *Grhl3* induit aussi une diminution de la phosphorylation de NO³²⁷.

1.4.4 Membres de la famille GRHL

Cette famille de facteurs de transcription agit dans un large éventail de domaines. *Grhl1* est nécessaire à la fonction normale de l'épiderme chez la souris. Une ablation entraîne une dérégulation sévère de la différenciation terminale des kératinocytes rappelant la kératodermie palmoplantaire^{328 329}. À travers un variant ENU qui entraîne la perte de fonction du gène, *Grhl2* a été impliqué dans la formation du tube neural chez la souris⁷⁴. Ce gène est aussi impliqué dans la morphogénèse épithéliale et dans la formation de la barrière épithéliale du foie³³⁰. Dans le rein, il est responsable de la formation de la barrière épithéliale et l'expansion de la lumière³³¹. Il est aussi impliqué dans la morphologie de l'épithélium des poumons et dans la morphogénèse des trophoblastes du placenta^{332 333}. Finalement, comme *Grhl3*, il est impliqué dans la formation de

l'épithélium³³⁴. *Grhl2* et *Grhl3* sont partiellement redondants dans ce tissu étant donné que l'expression de *Grhl2* dans une souris *Grhl3*^{-/-} rétablit le défaut de réparation des blessures, mais n'affecte pas le défaut de l'épiderme. Des différences entre les spectres de régulation de ces gènes, tel que *Tgase1* exprimé par *Grhl3* et insensible à *Grhl2*, justifie cette redondance incomplète³³⁵. Il est aussi intéressant de noter que *Grhl3* et *Grhl2* induisent tous deux l'expression de *Cdh1* à travers le même amplificateur de transcription, mais certains types cellulaires ne sont sensibles qu'à une de ces protéines⁹⁷.

L'expression de ces gènes corrèle fortement avec les rôles identifiés. *Grhl1* est exprimé dans l'ectoderme de surface à partir de E17.5, alors que *Grhl2* possède un patron d'expression similaire à *Grhl3*. *Grhl1* et 2 sont tous deux fortement exprimés dans les glandes sous-linguales à E17.5. *Grhl1* est très fortement exprimé dans le cœur dès E8.5. *Grhl1* et 2 sont exprimés dans les poumons et les reins dès E15.5. Les deux gènes sont fortement exprimés dans l'épithélium olfactif à partir de E10.5 pour *Grhl2* et E12.5 pour *Grhl1*. Ceci corrèle aussi avec l'épithélium oral excepté pour *Grhl1* qui commence à être exprimé seulement à E15.5. *Grhl2* est exprimé dans l'œsophage dès E8.5 et *Grhl1* dès E15.5. Ils sont tous deux exprimés dans l'estomac dès E15.5, mais *Grhl1* est beaucoup plus fortement exprimé. *Grhl2* est le seul membre de la famille à être exprimé dans l'intestin grêle et ce dès E10.5. Il est aussi exprimé dans le gros intestin et la vessie durant la même période. *Grhl1* est seulement exprimé dans la vessie à partir de E15.5. Ces différents patrons d'expression expliquent partiellement pourquoi malgré des fonctions redondantes, tel que la liaison de la région amplificatrice de *Cdh1*, un gène ne peut pas rétablir le phénotype causé par l'ablation de l'autre dans différents types cellulaires^{97 98}.

1.5 Études génétiques des maladies humaines

Les études génétiques ont grandement changé au cours des dernières années. L'accès au séquençage parallèle massif a grandement accéléré le processus et réduit les coûts associés. Plusieurs méthodes ont évolué et sont encore utilisées. Cette section fera un survol de ces méthodes ainsi que leurs forces et faiblesses (Figure 12).

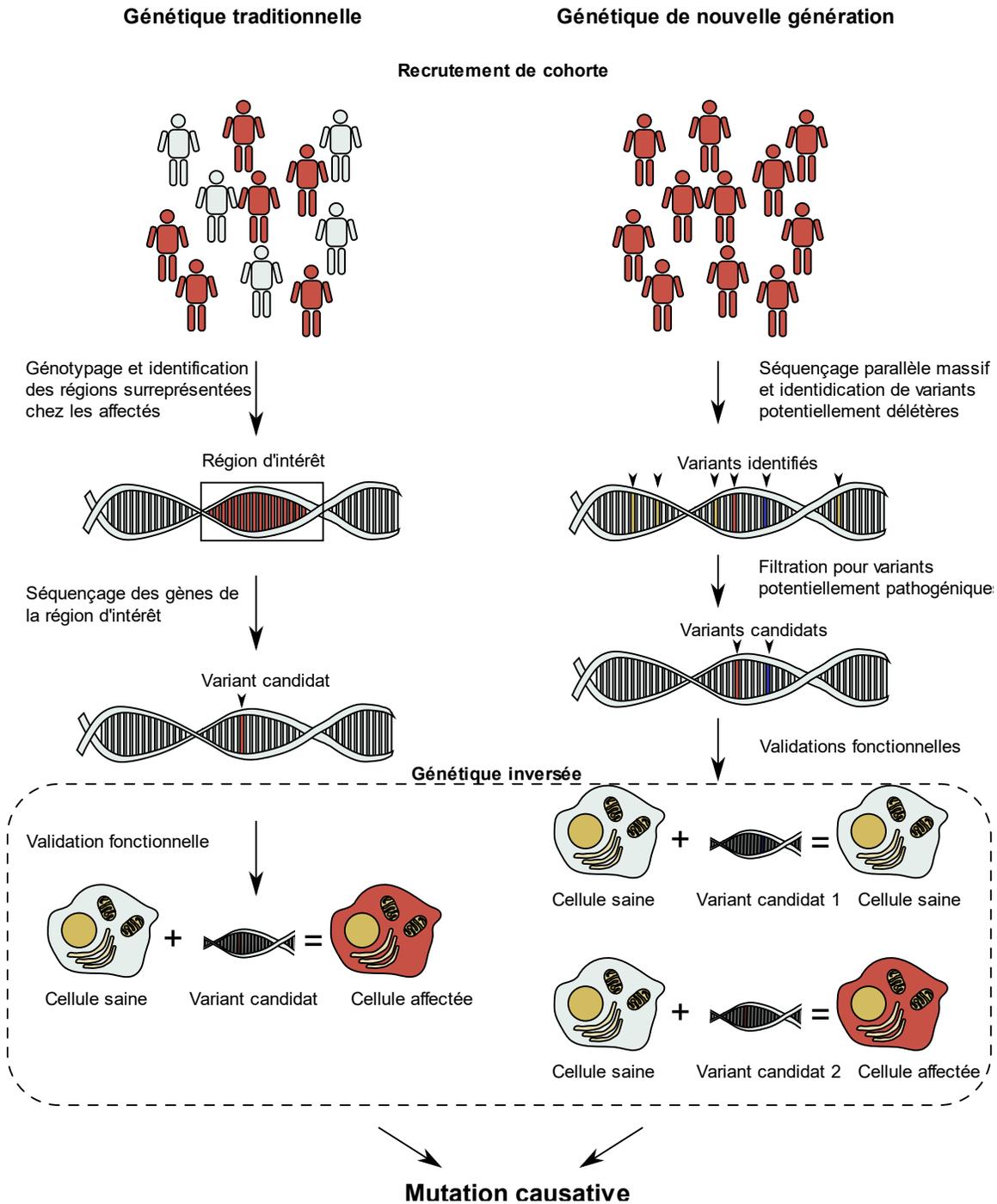


Figure 12: Représentation graphique de la génétique humaine traditionnelle (gauche) ou de nouvelle génération (droite).

1.5.1 Génétique traditionnelle

1.5.1.1 Étude de gènes candidats

Le mode d'investigation traditionnel consiste à l'étude de gènes spécifiques. Pour ce faire deux méthodes sont utilisées : La génétique directe et la génétique inversée. **La génétique directe** consiste à identifier la cause génétique d'un phénotype identifié dans un modèle³³⁶. La génération de modèles s'effectue principalement à travers des variants spontanés, la mutagenèse et les pièges géniques^{337 338 339}. Une sélection basée sur le phénotype observé est suivie par du séquençage visant l'identification du gène causatif³⁴⁰. L'identification directe de la mutation permet de créer un lien de causalité clair entre un gène et la pathologie, mais est toujours limitée par l'aléa des mutations et des emplacements d'insertion des diverses méthodes. Les orthologues des gènes identifiés chez les animaux modèles représentent des bons candidats pour la maladie humaine. **La génétique inversée** consiste à créer un variant connu dans un gène et de décrire le phénotype (Figure 12). Une fois de plus, le modèle est clef dans la description du phénotype. Cependant, cette approche favorise l'étude du gène, alors des modèles plus distants de l'être humain peuvent être utilisés pour son étude³⁴¹. Malgré l'aspect direct de cette méthode, elle nécessite d'inférer un lien entre le phénotype et la maladie. En effet, chez des organismes plus simples tel que la cellule et le ver, le phénotype observé est parfois difficile à corréler avec une pathologie. Ainsi, ces deux approches ont leurs forces et faiblesses quant à l'avancement de la génétique, cependant la dépendance de ces dernières sur des modèles diminue leurs applicabilités et supporte l'utilisation de nouvelles méthodes.

1.5.1.2 Études de liaison et d'association

L'utilisation d'un modèle permet d'identifier facilement des régions candidates grâce à une homogénéité génétique élevée et un contrôle sur les croisements permettant de réduire la région d'investigation. Cependant, les conclusions ne sont pas aisément applicables à la génétique parfois complexe de l'humain. Ainsi, des technologies permettant de sonder la totalité du génome humain chez de larges cohortes ont été développées. Les études utilisant cette technologie suivent un protocole qui consiste à identifier une pathologie, recruter une cohorte, séquencer la cohorte, identifier plusieurs gènes candidats et valider fonctionnellement ces gènes (Figure 12).

Le **génotypage de SNP** offre cette possibilité. Le séquençage de variations génétiques communes est effectué depuis longtemps et a servi entre autres à positionner les régions causatives dans les approches de génétique directe. Cependant, les connaissances acquises par le

séquençage du génome, ainsi qu'une réduction des coûts a permis d'ouvrir la voie au séquençage parallèle^{342 343}. La logique de cette méthode est qu'un SNP près d'un variant causatif suivra la même transmission et permettra d'identifier la région d'intérêt³⁴⁴. Deux types d'approches statistiques ont fait appel à cette technologie : Les études de liaison et les études d'association. Ces deux études visent à identifier des régions génétiques co-ségrégant avec la pathologie ou le trait à l'étude. Les études de liaison testent si une région suit statistiquement la transmission d'un trait dans une ou plusieurs familles. Les études d'association utilisent de larges cohortes d'individus non-affiliés afin de trouver des régions communes associées à un trait. Ces techniques ont l'avantage d'être peu dispendieuses permettant d'investiguer des cohortes beaucoup plus larges. Étant donné la logique de ces approches, l'identification d'une large région complique l'identification d'un variant causatif limitant les conclusions³⁴⁵.

1.5.2 Génétique de nouvelle génération

1.5.2.1 Études de séquençage de l'exome et du génome

Les limites de la technologies de génotypage de SNP ont mené au développement en 2007 du **séquençage de l'exome**³⁴⁶. Cette méthode consiste à utiliser des sondes spécifiques aux régions codantes du génome (les exons) suivi par un séquençage à l'aides d'amorces reconnaissant les régions ciblées par ces sondes (Figure 13)³⁴⁷. Les séquences ainsi recueillies sont filtrées pour une qualité minimale et pour des duplicatas causés par l'amplification. Elles sont ensuite alignées sur un génome de référence et les séquences non-alignées sont ensuite réalignées avec des critères moins stricts afin d'identifier les insertions et les délétions. Les divergences entre les séquences et le génome de référence sont ensuite identifiées et la position de ces variants est annotée pour comprendre le contexte génique³⁴⁸. Contrairement au génotypage de SNP, cette technique permet d'identifier directement les variants potentiellement causatifs. Cette technique est cependant beaucoup plus sensible à la qualité de l'ADN et le nombre de séquences obtenues. Plusieurs variants sont insérés durant le processus d'amplification et ils doivent être différenciés des variants réels. Ainsi, des critères de qualité élevés sont nécessaires afin de réduire les faux positifs³⁴⁹. À ce jour, 85% des mutations causatives ont été identifiées dans les régions codantes justifiant l'utilisation de cette technique³⁵⁰. Ceci est causé par une combinaison de deux facteurs : Le potentiel délétère et l'aptitude d'investigation. Des projets comme Encode démontrent qu'une large fraction du génome non-codant possède une activité régulatrice, cependant les chances d'avoir un effet délétère est plus élevé dans la

région codante. De plus, les validations fonctionnelles de variants dans ces régions sont complexes diminuant l'aptitude à démontrer une association³⁵¹. Cette technologie permet aujourd'hui d'identifier des variants *de novo* et des variants rares sans biais d'investigation. Ceci facilite grandement l'identification de nouveaux gènes associés à des maladies. Cependant, étant donné le grand nombre de variants identifiés, il est nécessaire de trouver des méthodes de priorisation appropriées.

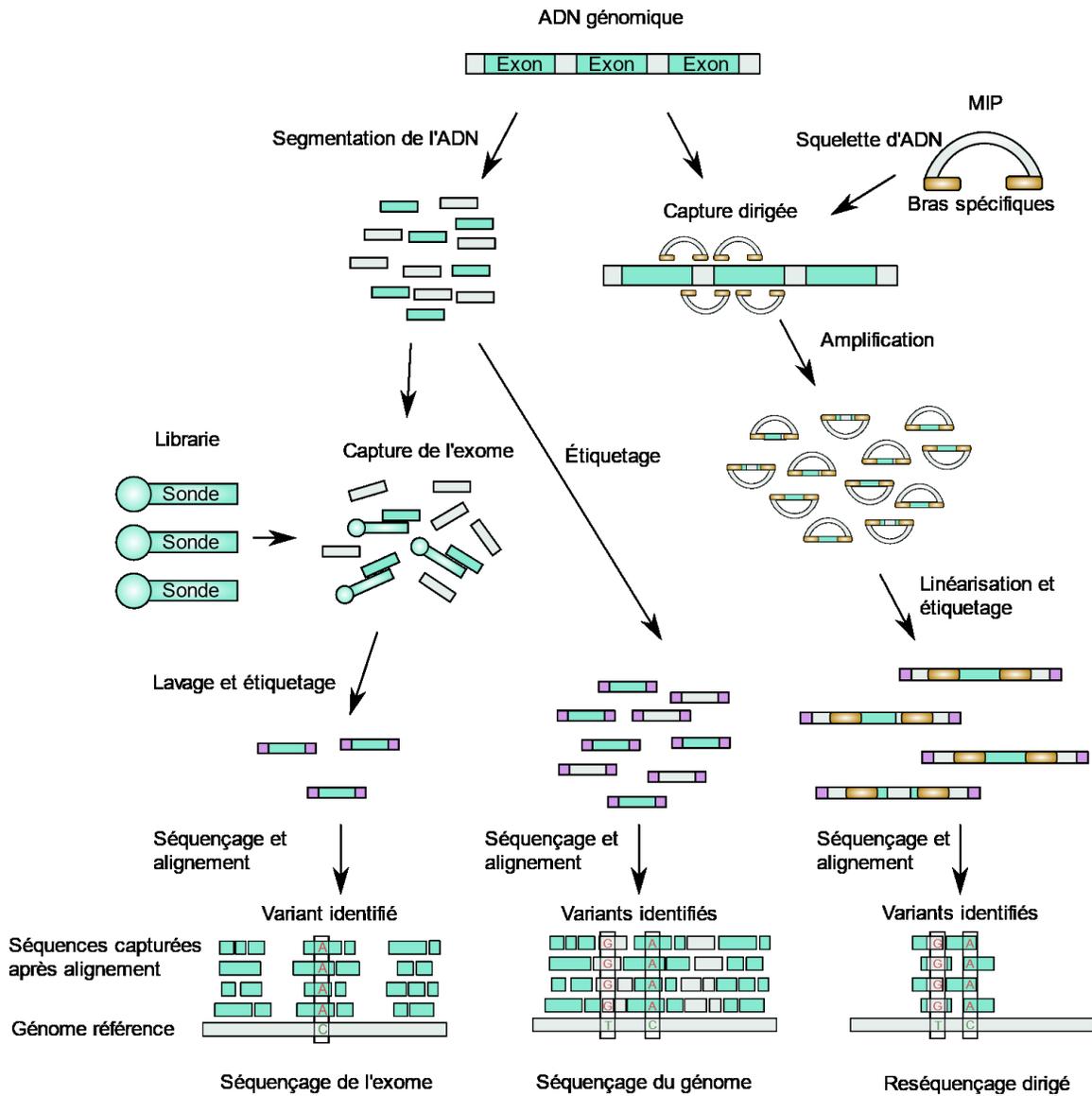


Figure 13 : Représentation graphique des méthodes de capture du séquençage parallèle massif avec à gauche le séquençage de l'exome, au centre le séquençage du génome et à droite le reséquençage dirigé (MIP).

La première méthode pour identifier des variants causatifs est de filtrer pour des variants hautement délétères. Pour définir l'impact potentiel d'une mutation, plusieurs programmes de prédiction existent. Des programmes comme Polyphen³⁵², SIFT³⁵³ et CADD³⁵⁴ utilisent des informations tel que la conservation d'un nucléotide à travers les espèces, l'effet potentiel de la modification ou la fréquence de mutation d'un trinucleotide. Ces méthodes permettent ainsi de

déterminer le risque que ce variant ait un effet pathogénique et ainsi de prioriser les variants intéressants. Un fardeau de variants est aussi une méthode de priorisation. Cette dernière consiste à déterminer le nombre attendu de variant dans un gène. Pour ce faire, des bases de données tel que ExAC³⁵⁵ et EVS³⁵⁶, des prédictions gènes spécifiques³⁵⁷ ou des approximations par génomes³⁵⁸ peuvent être utilisées. En comparant le nombre de variants observés dans la cohorte affectée à celui attendu dans un gène, il est possible d'identifier des gènes candidats. En assumant que la pathologie est la principale différence entre la cohorte affectée et contrôle, il est possible de conclure à une association avec le gène enrichi en variant^{358 359}. Les variants *de novo* sont aussi d'excellents candidats^{360 361 362}. Ces variants apparaissent dans les gamètes parentaux, sont uniquement présents chez l'enfant et ne sont pas soumis à une sélection naturelle. Pour cette raison, ils ont le potentiel d'être plus délétères que des variants transmis qui ne peuvent pas être individuellement létaux. Ces variants sont aussi d'excellents candidats dans le cadre de malformations létales et peuvent agir comme un renouvellement des variants perdus à travers les individus décédés³⁶³.

Le séquençage de l'exome est utilisé principalement pour son rapport efficacité-prix. La capture d'une section réduite du génome permet de maximiser la couverture des régions codantes et de multiplexer plusieurs individus tout en maintenant un niveau de couverture acceptable. Il est choisi au-dessus du génotypage de SNP grâce à sa capacité d'identifier directement les variants³⁴⁷. Un autre avantage est sa capacité à capturer l'ADN mitochondriale sans aucune sonde spécifique. Ce processus semble être associé à une surreprésentation de l'ADN mitochondriale permettant des associations non-spécifiques de ces séquences aux sondes de capture³⁶⁴. Les principales faiblesses de cette méthode sont l'inégalité de la capture et l'absence de données non-codantes. En effet, l'utilisation de sondes crée un large biais au niveau de la capture compliquant l'identification de variants par nombre de copies (CNV). Ces variants sont représentés par une séquence répétée sur le même allèle d'ADN et ont un potentiel pathogénique élevé³⁶⁵. L'absence de données non-codantes limite aussi l'identification de variants causatifs chez certaines familles. Cependant, une faible proportion des études génétiques a identifié des variants causatifs dans la région non-codante étant donné la complexité de l'analyse et de la validation de ces variants. De plus, les régions d'épissage, les 5' et les 3' non-translaté (UTR), très importantes dans l'analyse de la région codante, sont bien couvertes dans les nouvelles trousse de capture³⁶⁶. Ainsi, cette faiblesse est largement minimisée.

Afin de compenser pour ces faiblesses, la technique par excellence est le **séquençage du génome**. Cette technique n'utilise pas de sonde, mais capture la totalité de l'ADN génomique et la séquence (Figure 13). Le niveau de couverture est généralement plus bas, mais beaucoup plus uniforme permettant une meilleure analyse des CNV. Il permet aussi l'identification de variants non-codants. Le principal obstacle est le coût minimisant la grosseur des études potentielles tout en apportant un gain qui demeure encore difficilement quantifiable³⁶⁷.

1.5.2.1 *Étude de reséquençage dirigé*

Le **reséquençage dirigé** est utilisé pour de courtes régions d'intérêt nécessitant une haute couverture (Figure 13). En utilisant le séquençage parallèle dans une petite région, une large quantité de patients sont séquencés avec une forte couverture et un coût minimal. Pour ce faire, il est nécessaire de créer une librairie de capture sur mesure. Plusieurs compagnies offrent ce service^{368 369}, mais une méthode récente facilite ce processus. Les sondes d'inversion moléculaire (MIP) sont des courtes séquences d'ADN composées de deux bras spécifiques à la région à capturer et liées par un squelette d'ADN³⁷⁰. Suite à l'hybridation à la séquence capturée, le squelette d'ADN est ouvert et marqué à l'aide d'adaptateurs spécifiques à chaque sonde. Cette capture est séquencée à l'aide du séquençage parallèle massif. Le principal obstacle à cette approche est la génération de sondes de qualité³⁷⁰. Cependant, des programmes sont maintenant disponibles afin de faciliter cette étape diminuant les coûts associés au reséquençage dirigé.

Tout comme le séquençage de l'exome et du génome, cette étude ne permet pas d'adéquatement identifier les mutations affectant l'expression ou l'épigénétique. Ces dernières sont plutôt faites à l'aide de techniques telles que les micropuces, le séquençage de l'ARN ou le séquençage bi-sulfates. Elles sont une source d'information importante et complémentaires aux études génétiques de nouvelle génération.

2. Projet de recherche

2.1 Problématique

Malgré la supplémentation en acide folique, près de 1 naissance sur 2000 sont affectées par les ATN au Canada. Ces familles affectées et souvent insensibles à l'acide folique sont alors sans recours étant donné les connaissances anémiques sur l'étiologie génétique de la maladie et sur d'autres modes de prévention. L'absence d'information nuit à la prise de décision éclairée des familles vis-à-vis d'une deuxième grossesse. Une mauvaise compréhension des risques peut augmenter la détresse psychologique des familles en plus de leur imposer un fardeau fiscal imprévu, ainsi qu'à l'état³⁷¹.

L'identité et le nombre de gènes impliqués dans le développement des ATN demeurent encore largement inconnus. Des approches basées sur une priorisation des candidats de la voie PCP et de l'acide folique ont identifié des variants intéressants, mais ont stagné à l'investigation des mêmes gènes¹⁰³. De plus, le conseil génétique est limité par la faible pénétrance des variants identifiés et le manque de connaissances sur la transmission de la maladie³⁷².

2.2 Hypothèse et but

Notre hypothèse est que la plupart des variants causatifs se retrouveront dans la région codante du génome et que l'utilisation de techniques de séquençage parallèle massif combiné à des approches statistiques et gènes candidats permettra d'identifier de nouveaux gènes impliqués dans l'étiologie génétique complexe des ATN humaines.

Notre but ultime vise à identifier de nouveaux variants chez des gènes connus ou de nouveaux gènes impliqués dans les ATN à travers l'utilisation du séquençage parallèle massif de l'exome entier. Plus spécifiquement 2 objectifs principaux sont définis :

- (1) Investiguer la contribution des variants *de novo* à l'étiologie des ATN dans une cohorte de 43 trios
- (2) Identifier des variants transmis hautement délétères dans une cohorte de 241 familles, trios et cas sporadiques et en faire la validation fonctionnelle

2.3 Pertinence et retombées

L'identification de nouveaux gènes impliqués dans les ATN permettra de mieux comprendre les mécanismes pathogéniques de la maladie. Ces gènes permettront aussi, le développement de nouvelles stratégies préventives. Finalement, une meilleure compréhension des différents modes de transmission de cette pathologie permettra d'éclairer le conseil génétique pour les familles affectées.

3. Contributions

Contributions aux anomalies du tube neural

Les sections suivantes de cet ouvrage incluent deux articles qui contribuent aux connaissances dans le domaine de la génétique des anomalies du tube neural. Ces articles sont « Loss of function de novo mutations play an important role in severe human neural tube defects », originalement publié dans « Journal of Medical Genetics » en date du 24 Mars 2015³⁷³ et « Rare deleterious variants in GRHL3 are associated with human spina bifida », originalement publié dans « Human Mutation » en date du 24 Mars 2017³⁷⁴. Un troisième article identifiant des approches d'investigation de mutations avec de plus faibles potentiels pathogéniques dans les anomalies du tube neural est aussi en préparation. Je suis responsable de l'écriture de la planification/implémentation des analyses et du développement des hypothèses. L'article devrait être prêt pour soumission au cours de l'hiver 2018.

Contributions supplémentaires

J'ai aussi été impliqué dans plusieurs autres articles non-reliés aux anomalies du tube neural ou ayant nécessité une contribution plus faible. J'ai été impliqué dans la publication d'un article sur la malformation Chiari-Like chez le chien Griffon Bruxellois³⁷⁵. Un second article sur la Syringomyélie chez le chien Cavalier King Charles Spaniel est présentement en révision. Je suis premier auteur sur ces deux articles ayant été impliqué dans le développement des hypothèses, l'analyse des résultats et dans une large partie de l'écriture du manuscrit. Je suis aussi premier auteur sur un article caractérisant un nouveau modèle murin de l'Anémie de Blackfan Diamond qui est présentement en préparation et qui sera probablement soumis au cours du printemps 2018. J'étais responsable d'identifier le gène causatif impliqué dans ce nouveau modèle ainsi que de développer un programme d'analyse des données génétiques provenant de la souris. J'ai aussi participé à la caractérisation du variant et à l'écriture de l'article.

J'ai aussi contribué à l'écriture et au développement de plusieurs projets de recherche à un niveau plus faible. Ainsi, j'ai aidé à développer les analyses génétiques dans d'une cohorte affectée par l'Amyloïdose cérébrale³⁷⁶ et pour une cohorte affectée par l'encéphalopathie épileptique³⁷⁷. Finalement, j'ai été impliqué dans des analyses ayant mené à l'identification d'une mutation dans la Syringomyélie chez le Cavalier King Charles Spaniel. Cet article diffère de celui

décrit plus tôt par les méthodes utilisées et les résultats complémentaires présentés. J'ai aussi participé à une partie de l'écriture de cet article qui devrait être soumis sous peu.

4. Loss of function de novo mutations play an important role in severe human neural tube defects

Philippe Lemay¹, Marie-Claude Guyot¹, Élizabéth Tremblay¹, Alexandre Dionne-Laporte², Dan Spiegelman², Édouard Henrion², Ousmane Diallo², Patrizia De Marco³, Elisa Merello³, Christine Massicotte¹, Valérie Désilets¹, Jacques L. Michaud¹, Guy A. Rouleau², Valeria Capra³, Zoha Kibar^{1,*}

¹ CHU Ste-Justine Research Center, Université de Montréal, Montréal, Québec H3T 1C5 Canada ;

² Montreal Neurological Institute, McGill University, Montréal, Québec H3A 2B4,Canada ³ Istituto Giannina Gaslini, 16148, Genoa, Italy

*Corresponding Author at University of Montréal, Department of Neurosciences, CHU Sainte Justine Research Center, 3175 Cote-Ste-Catherine, Room 5999C, Montreal, QC H3T 1C5 Canada; Tel: (514) 345 4931 X 3984; Fax: (514) 345 4801; E-mail : zoha.kibar@ recherche-ste-justine.qc.ca

Contributions des co-auteurs

Philippe Lemay:

- Participation au développement de l'hypothèse
- Développement de filtres et de programmes informatiques permettant l'analyse des données finales du séquençage de l'exome
- Validation des variants identifiés
- Analyses statistiques
- Conception des figures et rédaction du manuscrit

Marie-Claude Guyot:

- Validation des variants identifiés
- Support technique

Élizabeth Tremblay:

- Validation des variants identifiés

Alexandre Dionne-Laporte, Dan Spiegelman, Édouard Henrion, Ousmane Diallo:

- Transformation des données brutes de séquençage de l'exome en format d'analyse
- Support informatique

Patrizia De Marco, Elisa Merello, Christine Massicotte, Valérie Désilets:

- Recrutement de la cohorte de 43 trios

Jacques L. Michaud, Guy A. Rouleau, Valeria Capra:

- Support financier
- Support technique et participation au développement de l'hypothèse
- Aide au recrutement

Zoha Kibar:

- Conception et supervision du projet
- Soutien scientifique, technique et matériel pour la réalisation du projet
- Rédaction du manuscrit

Abstract

Background. Neural tube defects (NTDs) are very common and severe birth defects that are caused by failure of neural tube closure and that have a complex etiology. Anencephaly and spina bifida are severe NTDs that affect reproductive fitness and suggest a role for *de novo* mutations (DNMs) in their etiology. **Methods.** We used whole exome sequencing in 43 sporadic cases affected with myelomeningocele or anencephaly and their unaffected parents to identify DNMs in their exomes. **Results.** We identified 42 coding DNMs in 25 cases, of which six were Loss of Function (LoF) showing a higher rate of LoF DNM in our cohort compared to control cohorts. Notably, we identified two protein truncating DNMs in two independent cases in *SHROOM3*, previously associated with NTDs only in animal models. We have demonstrated a significant enrichment of LoF DNMs in this gene in NTDs when compared to the gene specific DNM rate and to the DNM rate estimated from control cohorts. We also identified one nonsense DNM in *PAX3* and 2 potentially causative missense DNMs in *GRHL3* and *PTPRS*. **Conclusions.** Our study demonstrates an important role of LoF DNMs in the development of NTDs and strongly implicates *SHROOM3* in its etiology.

Introduction

Neural tube defects (NTDs) are a group of congenital malformations affecting 1 to 2 individuals per 1000 births.¹ They are caused by an incomplete closure of the neural tube during embryogenesis.¹ The most frequent forms of NTDs are anencephaly and myelomeningocele (or spina bifida), which are caused by a closure defect in the brain and the spinal cord region respectively.¹ Children affected with anencephaly die early in development or soon after birth, while children affected with myelomeningocele survive but are affected by developmental physical defects with varying degrees of severity. Most cases of NTDs are sporadic and non-syndromic.² Peri-conceptual folic acid intake has been shown to reduce prevalence of NTDs by 50-70%,³ but a large amount of cases remain resistant to this preventive treatment urging the need for identification of other causative factors and development of novel preventive and counseling strategies.

NTDs have a strong genetic component with an estimated heritability of 60%,¹ but so far the genetics of the disease remains largely unknown. Previous linkage studies in NTDs have identified candidate regions on chromosomes 2, 7 and 10 but failed to identify any causative NTD gene.⁴ Few common variants in folic acid-related genes have also been shown to be associated with NTDs in certain populations, but these variants seem to contribute only to a small part of the etiology of the disease.⁵ Previous gene identification studies in NTDs have mainly adopted a candidate gene approach and focused on folate-related genes and on candidate genes from animal studies.^{1,2,5} Animal models have demonstrated an important role of the planar cell polarity pathway in the etiology of NTDs.¹ Subsequent investigation of genes of this pathway, including *VANGL1* and *VANGL2*, in human NTDs has implicated them as risk factors in a small fraction of patients.¹ Generally, candidate gene studies in NTDs have faced limited success in identifying major causative genes predisposing to NTDs suggesting the need for novel approaches. Several

recent studies strongly suggest that *de novo* variants (DNMs) represent a common cause of birth defects and neurodevelopmental diseases.⁶ DNM could provide a mechanism by which early-onset reproductively lethal diseases remain frequent in the population. Therefore, these variants are strong candidates for causing diseases that occur sporadically and that have a reduced reproductive fitness.⁶ Severe forms of NTDs, such as anencephaly and myelomeningocele, fall in this category and hence investigation of DNMs may therefore increase the chance of identifying loss of function (LoF) DNMs implicated in NTDs.

Results and discussion

Forty-three families each composed of one affected child and two unaffected parents with no family history of NTDs were recruited through the Montreal Ste-Justine Hospital Spina Bifida Center, the 3D study of the Integrated Research Network in Perinatology of Quebec and Eastern Ontario and the Istituto Giannina Gaslini in Genoa, Italy. Samples from families were collected with the approval of the Local Ethics Committees and written informed consent was obtained from all participating individuals. Detailed information including folate status, tissue of origin and type of NTDs of this cohort is summarized in Table S1. Briefly, all 43 cases were affected with NTDs including 35 myelomeningocele and 8 anencephaly cases. A total of 21 cases were fetuses and 55.6 % took folate peri-conceptionally. Tissues from fetuses were all obtained following induced abortions. The average maternal and paternal ages were 30.0 ± 4.8 years and 30.7 ± 5.9 years respectively.

Exome data from 43 NTD trios was aligned to the GRCh37 genome reference and called variants were filtered for minimal quality (GQ> 20, total variant reads> 3, total reads> 7) and for absence in parents and public databases (1000 genome at <http://www.1000genomes.org/home> and Exome Variant Server at <http://evs.gs.washington.edu/EVS/>) to identify DNMs. Details of the alignment strategy and SNV annotation are included in Supplementary data. Seventy variants

were identified and validated by Sanger sequencing to minimize false negative results. This approach resulted in the identification of 42 coding mutations in 25 trios including 3 nonsense, 3 frameshift, 30 missense and 6 synonymous (table 1). The average DNM rate per base pair was 1.62×10^{-8} which was consistent with published ratios.^{7,8}

We identified 6 LoF DNMs defined as nonsense, frameshift and splicing variants, in 6 NTD trios resulting in a per base DNMs rate of 0.23×10^{-8} . This rate was higher than two previously published per base LoF DNMs rates, 0.05×10^{-8} (reference 8) and 0.17×10^{-8} (reference 9) but only reached significance ($P = 0.002$) when compared to the lower rate suggesting that larger cohorts of MMC and anencephaly are needed to confirm these initial findings. This increased rate suggests that LoF DNMs are an important part of the pathogenicity of NTDs. Details of all statistical analysis can be found in Supplementary data.

Notably, we identified 2 LoF DNMs in *SHROOM3* (NM_020859.3) in two unrelated families: one nonsense variant c.1176C>G (p.Y392X) and one frameshift variant c.2843_2844insG leading to a premature stop codon (p.L948fs) (Figure 1B). The c.1176C>G variant was detected in the PrZRV proband affected with anencephaly with cranial fossa agenesis and facial dysmorphism (table 1). The mother took folic acid peri-conceptionally (Supplementary table S1). The c.2843_2844insG variant was detected in the Pr394 proband affected with a thoracic myelomeningocele and Chiari type IV malformation (table 1). The mother did not take folic acid peri-conceptionally (Supplementary table S1). None of these mutations were reported in the ExAC database (<http://exac.broadinstitute.org/>). *SHROOM3* is an actin-binding protein that is known to be a key regulator of apical constriction, a process by which cells convert their shape from cuboidal to wedge-like due to a decrease in their apical area.⁹ This function is essential for hinge-point formation and bending of the neural tube during its formation and closure in both

vertebrate embryos.⁹ Recessive mutations in *SHROOM3* were previously associated to heterotaxy which represents a multiple congenital anomaly syndrome resulting from abnormalities of the proper specification of left-right asymmetry during embryonic development.¹⁰ The LoF DNM in this gene described on our study was detected in two NTD patients who showed no sign of heterotaxy, and hence these NTD mutations might act in a haploinsufficient or a dominant negative manner. *Shroom3* exists in 2 isoforms that were demonstrated to have similar functions and expression patterns in the mouse and frog models.^{11 12} In the mouse gene trap mutant, both isoforms are knocked out leading to exencephaly and spina bifida in homozygous embryos at a penetrance of 100% and 23% respectively.¹¹ In our cohort, both truncating DNMs detected in *SHROOM3* seem to affect both isoforms since they map at positions 1176bp and 2843bp respectively 5' or inside the ASD1 domain (Figure 1A). These 2 DNMs clearly remove important functional domains and might confer nonsense mediated RNA decay and hence they are most likely LoF mutations. Furthermore, previous studies have demonstrated the potential of a truncated version of *Shroom3* to act in a dominant negative manner in *Xenopus*¹². This support the potential pathogenicity of these heterozygous mutations which could result in limited apical constriction causing the NTD.

The phenotypic variation between the two probands (myelomeningocele and anencephaly) who carry LoF DNM in *SHROOM3* could be caused by many factors including genetic modifiers and environmental factors. The two variants might also have different pathogenic effects on the protein with different outcomes in the implicated developmental pathway. These factors could act alone or in combination to affect the clinical expressivity of the NTD phenotype.

Considering the estimated LoF DNM rates between 0.05×10^{-8} and 0.17×10^{-8} SNV per base^{7 8} and the length of *SHROOM3* long isoform (NM_020859.3: 5991bp), we expect to find

respectively between 0.00026 and 0.00088 LoF DNM in this gene in our cohort. However, we detected 2 DNMs in our cohort of 43 trios resulting in a frequency of DNM per bp of 3.88×10^{-6} . This represents a significant enrichment of LoF DNMs in *SHROOM3* under the two rates with respective Bonferroni corrected *P* values of 6.77×10^{-4} and 0.008 using a two tailed binomial exact test. A recently published gene specific per trio mutation rate for LoF DNMs was also used to test the significant enrichment of LoF DNMs in *SHROOM3*.¹³ Under the *SHROOM3* gene specific rate of 1.34×10^{-5} , we expected to find 0.00058 LoF DNMs in our cohort which was significantly lower than our observed rate (*P* bonferroni = 0.003). Details of all statistical analyses can be found in Supplementary data.

Also importantly we detected one *de novo* stopgain in the NTD-associated gene *PAX3*, c.218C>A (p.S73X), in the Pr134 case affected with myelomeningocele with type II Arnold-Chiari Malformation, hydrocephalus and Waardenburg syndrome (Figure 1D). The mother did not take folic acid or multivitamins peri-conceptionally (Supplementary table S1). This mutation was not reported in the ExAC database. *PAX3* is a paired box transcription factor that is a key developmental regulator of the neural crest and its derivatives and that plays an important role during neurogenesis and myogenesis. It contains two DNA binding domains, the paired box domain (PAX) and a paired-type homeodomain (HOX) that interact cooperatively for DNA binding.¹⁴ The putative truncated protein caused by the DNM p.S73X lacks all functional domains including both PAX and HOX (Figure 1C) and might confer nonsense mediated RNA decay.

In the Splotch mouse model, homozygous loss-of-function mutations of *Pax3* cause spina bifida and other neural crest abnormalities. In humans, the relative contribution of *PAX3* to the overall burden of NTDs remains unclear.¹⁴ Heterozygous mutations in *PAX3* are known to cause Waardenburg syndrome (WS), an autosomal dominant condition that affects neural crest derived

structures and which is occasionally associated with NTDs. Few individuals with both WS and NTDs were demonstrated to carry heterozygous *PAX3* mutations or deletions.¹⁴ Our report of a new protein truncating SNV in *PAX3* in one NTD patient provides additional evidence for a pathogenic role of this gene in spina bifida.

While *SHROOM3* and *PAX3* are by far the strongest candidate genes identified in this study, other interesting DNMs have been identified in our cohort (table 1). We detected 3 LoF DNMs, p.W127X in *WBSCR28*, p.K23fs in *MFAP1* and p.Q207fs in *DDX3X*, with no previous association to NTDs and that were not reported in the ExAC database (table 1, Supplementary Figure S1). The *WBSCR28* DNM was detected in the PrYPT proband affected with lumbosacral myelomeningocele and type II Arnold-Chiari Malformation. This gene encodes a putative transmembrane protein of unknown function that maps to the region deleted in the Williams-Beuren syndrome (table 1). This syndrome is characterized by a range of phenotypes including mental retardation, dysmorphic facies, heart abnormalities, short stature and infantile hypocalcaemia. *WBSCR28* is not the main candidate gene in this disease.¹⁵ The *MFAP1* DNM was found in the Pr201 proband affected with lumbosacral myelomeningocele. This gene encodes the microfibrillar-associated protein 1 that represents an uncharacterized protein found in some human spliceosomal fractions.¹⁶ The *DDX3X* DNM was found in the Pr389 proband affected with lumbosacral myelomeningocele and hydrocephalus. This gene belongs to the DEAD-box proteins, a large family of ATP-dependent RNA helicases that participate in all aspects of RNA metabolism.¹⁷ While none of these 3 genes represent strong candidates for NTDs based on published data, the presence of LoF DNMs in these genes still suggest them as potentially interesting NTD candidates. Additional genetic studies in larger cohorts and functional studies are needed to validate their role in NTDs.

Two other interesting missense DNM identified in 2 NTD trios implicated genes whose orthologues cause NTD in mice: c.1171C>T (p.R391C) in *GRHL3* (*GRAINYHEAD-LIKE 3*) and c.4475G>A (p.R1492Q) in *PTPRS* (*PROTEIN TYROSINE PHOSPHATASE, RECEPTOR TYPE, S*) (supplementary Figure S2). The *GRHL3* mutation was identified in the Pr548 proband affected with myelomeningocele with type II Arnold-Chiari Malformation and hydromyelia. This mutation was not previously reported in the ExAC database. *Grhl3* is a transcription factor that plays an important role in epidermal integrity and wound healing. Importantly, null alleles at this gene caused mainly severe spina bifida and occasionally exencephaly. *Grhl3* was tightly linked to Curly tail that represents one of the most well-established mouse models for NTDs.¹⁸ The p.R391C maps to the DNA binding domain of the GRHL3 protein. The modified amino acid is highly conserved (supplementary Figure S2) and the mutation is defined as probably damaging by PolyPhen-2 HDIV (table 1). Surprisingly, the same DNM p.R391C was detected in Van der Woude syndrome, the most common syndromic form of cleft lip and palate.¹⁹ While this could point toward a chance finding, we hypothesize that additional genetic and/or environmental factors may modify the phenotypic expression of the same *GRHL3* mutation in different individuals. The p.R1492Q in *PTPRS* was found in the Pr125 proband affected with lumbosacral myelomeningocele with type II Arnold-Chiari Malformation. The mutation was previously reported on one allele in the ExAC database (<http://exac.broadinstitute.org/>) resulting in a frequency of 0.000008368. This gene is a phosphatase involved in regulating cell proliferation, cell adhesion and nervous system maturation and causes exencephaly in mice.²⁰ *PTPRS* fibronectin and Ig-like domains are extracellular and mainly involved in cell interaction, while the catalytic phosphatase domain is intracellular and mainly involved in signal transmission. The p.R1049Q variant was identified in a proband affected with myelomeningocele (table 1) and resides in a conserved region that forms part of the phosphatase domain (supplementary Figure S2). It modifies a conserved arginine to a

glutamine, a nonconservative substitution that removes the positive charge of the amino acid side chain, changes its size and is predicted to be probably damaging by PolyPhen-2 (table 1).

Molecular genetic studies of bigger cohorts and careful phenotyping are still needed to better understand the mechanism of action of potential DNMs detected in this study and to assess their role in the pathogenicity in NTDs. While these DNMs are most likely highly penetrant, it is possible that they act in concert with other events to cause the disease. This is consistent with the two hit model that was initially proposed as a cancer mechanism and later suggested as a potential mechanism in complex diseases.²¹ This model was more recently explored in explaining variable expressivity in severe developmental genomic disorders.²¹ In this model, a secondary insult is necessary to result in a different or more severe clinical manifestation of a complex disease. This insult occurs during development and could be genetic or epigenetic (germ line or somatic) or environmental. The two “hits” could act independently or additively to each other resulting in a phenotype that differs from either hit alone. Alternatively, the two hits could involve gene(s) from the same or similar biochemical pathway and hence could interact in an epistatic manner.

The novel approach of whole exome sequencing has long been due in the genetic investigation of NTDs. This study uses this powerful approach in this complex trait and has successfully identified potential candidate de novo mutations in novel genes in the development of human NTDs. We have demonstrated the presence of LoF variants in 5 genes, have reported more LoF DNMs in our cohort than expected and we have identified 2 of those mutations in orthologues of mouse NTD genes suggesting the involvement of those variants in the etiology of human NTDs. We have also demonstrated the presence of two independent protein truncating variants in *SHROOM3* in 43 trios and presented a statistically significant enrichment of LoF DNMs

in this gene when compared to control and gene specific DNMs rates. Our data strongly suggest that highly penetrant pathogenic variants in this gene may account for a significant part of the genetic etiology of severe forms of NTD. Further studies of this gene in a bigger cohort of sporadic cases may help better assess the significance of these findings.

Acknowledgments

We would like to thank all participants in this study. ZK has a salary award from the “Fonds de Recherche du Québec – Santé”. JLM is a National Scientist of the Fonds de Recherche du Québec-Santé. PL is supported by “Fondation du CHU Ste-Justine” and “Fonds de Recherche du Québec – Santé”. This project was conducted as part of the research programme of the Integrated Research Network in Perinatology of Quebec and Eastern Ontario (IRNPQEO). This work was supported by the Canadian Institutes of Health Research [CRI 88413].

References

- 1 Bassuk AG, Kibar Z. Genetic basis of neural tube defects. *Semin Pediatr Neurol* 2009;16:101–10.
- 2 Copp AJ, Stanier P, Greene ND. Neural tube defects: recent advances, unsolved questions, and controversies. *Lancet Neurol* Published Online First: 18 June 2013. doi:10.1016/S1474-4422(13)70110-8
- 3 Prevention of neural tube defects: results of the Medical Research Council Vitamin Study. MRC Vitamin Study Research Group. *Lancet* 1991;338:131–7.
- 4 Stamm DS, Siegel DG, Mehlretter L, Connelly JJ, Trott A, Ellis N, Zismann V, Stephan DA, George TM, Vekemans M, Ashley-Koch A, Gilbert JR, Gregory SG, Speer MC. Refinement of 2q and 7p loci in a large multiplex NTD family. *Birt Defects Res A Clin Mol Teratol* 2008;82:441–52.
- 5 Zhang T, Lou J, Zhong R, Wu J, Zou L, Sun Y, Lu X, Liu L, Miao X, Xiong G. Genetic variants in the folate pathway and the risk of neural tube defects: a meta-analysis of the published literature. *PLoS One* 2013;8:e59570.
- 6 Veltman JA, Brunner HG. De novo mutations in human genetic disease. *Nat Rev Genet* 2012;13:565–75.
- 7 Sanders SJ, Murtha MT, Gupta AR, Murdoch JD, Raubeson MJ, Willsey AJ, Ercan-Sencicek AG, DiLullo NM, Parikshak NN, Stein JL, Walker MF, Ober GT, Teran NA, Song Y, El-Fishawy P, Murtha RC, Choi M, Overton JD, Bjornson RD, Carriero NJ, Meyer KA, Bilguvar K, Mane SM, Sestan N, Lifton RP, Günel M, Roeder K, Geschwind DH, Devlin B, State MW. De novo mutations revealed by whole-exome sequencing are strongly associated with autism. *Nature* 2012;485:237–41.
- 8 Iossifov I, O’Roak BJ, Sanders SJ, Ronemus M, Krumm N, Levy D, Stessman HA, Witherspoon KT, Vives L, Patterson KE, Smith JD, Paepfer B, Nickerson DA, Dea J, Dong S, Gonzalez LE, Mandell JD, Mane SM, Murtha MT, Sullivan CA, Walker MF, Waqar Z, Wei L, Willsey AJ, Yamrom B, Lee Y, Grabowska E, Dalkic E, Wang Z, Marks S, Andrews P, Leotta A, Kendall J, Hakker I, Rosenbaum J, Ma B, Rodgers L, Troge J, Narzisi G, Yoon S, Schatz MC, Ye K, McCombie WR, Shendure J, Eichler EE, State MW, Wigler M. The contribution of de novo coding mutations to autism spectrum disorder. *Nature* 2014;515:216–21.
- 9 Sawyer JM, Harrell JR, Shemer G, Sullivan-Brown J, Roh-Johnson M, Goldstein B. Apical constriction: a cell shape change that can drive morphogenesis. *Dev Biol* 2010;341:5–19.
- 10 Tariq M, Belmont JW, Lalani S, Smolarek T, Ware SM. SHROOM3 is a novel candidate for heterotaxy identified by whole exome sequencing. *Genome Biol* 2011;12:R91.
- 11 Hildebrand JD, Soriano P. Shroom, a PDZ domain-containing actin-binding protein, is required for neural tube morphogenesis in mice. *Cell* 1999;99:485–97.
- 12 Haigo SL, Hildebrand JD, Harland RM, Wallingford JB. Shroom induces apical constriction and is required for hinge-point formation during neural tube closure. *Curr Biol* 2003;13:2125–37.
- 13 Samocha KE, Robinson EB, Sanders SJ, Stevens C, Sabo A, McGrath LM, Kosmicki JA, Rehnström K, Mallick S, Kirby A, Wall DP, MacArthur DG, Gabriel SB, DePristo M, Purcell SM,

Palotie A, Boerwinkle E, Buxbaum JD, Cook EH, Gibbs RA, Schellenberg GD, Sutcliffe JS, Devlin B, Roeder K, Neale BM, Daly MJ. A framework for the interpretation of de novo mutation in human disease. *Nat Genet* 2014;46:944–50.

14 Greene NDE, Massa V, Copp AJ. Understanding the causes and prevention of neural tube defects: Insights from the splotch mouse model. *Birt Defects Res A Clin Mol Teratol* 2009;85:322–30.

15 Hockenhull EL, Carette MJ, Metcalfe K, Donnai D, Read AP, Tassabehji M. A complete physical contig and partial transcript map of the Williams syndrome critical region. *Genomics* 1999;58:138–45.

16 Hegele A, Kamburov A, Grossmann A, Sourlis C, Wowro S, Weimann M, Will CL, Pena V, Lührmann R, Stelzl U. Dynamic protein-protein interaction wiring of the human spliceosome. *Mol Cell* 2012;45:567–80.

17 Soto-Rifo R, Ohlmann T. The role of the DEAD-box RNA helicase DDX3 in mRNA metabolism. *Wiley Interdiscip Rev RNA* 2013;4:369–85.

18 Gustavsson P, Copp AJ, Greene NDE. Grainyhead genes and mammalian neural tube closure. *Birt Defects Res A Clin Mol Teratol* 2008;82:728–35.

19 Peyrard-Janvid M, Leslie EJ, Kousa YA, Smith TL, Dunnwald M, Magnusson M, Lentz BA, Unneberg P, Fransson I, Koillinen HK, Rautio J, Pegelow M, Karsten A, Basel-Vanagaite L, Gordon W, Andersen B, Svensson T, Murray JC, Cornell RA, Kere J, Schutte BC. Dominant mutations in GRHL3 cause Van der Woude Syndrome and disrupt oral periderm development. *Am J Hum Genet* 2014;94:23–32.

20 Uetani N, Bertozzi K, Chagnon MJ, Hendriks W, Tremblay ML, Bouchard M. Maturation of ureter-bladder connection in mice is controlled by LAR family receptor protein tyrosine phosphatases. *J Clin Invest* 2009;119:924–35.

21 Kumar RA. Two-hit wonder: a novel genetic model to explain variable expressivity in severe pediatric phenotypes. *Clin Genet* 2010;78:517–9.

Figures

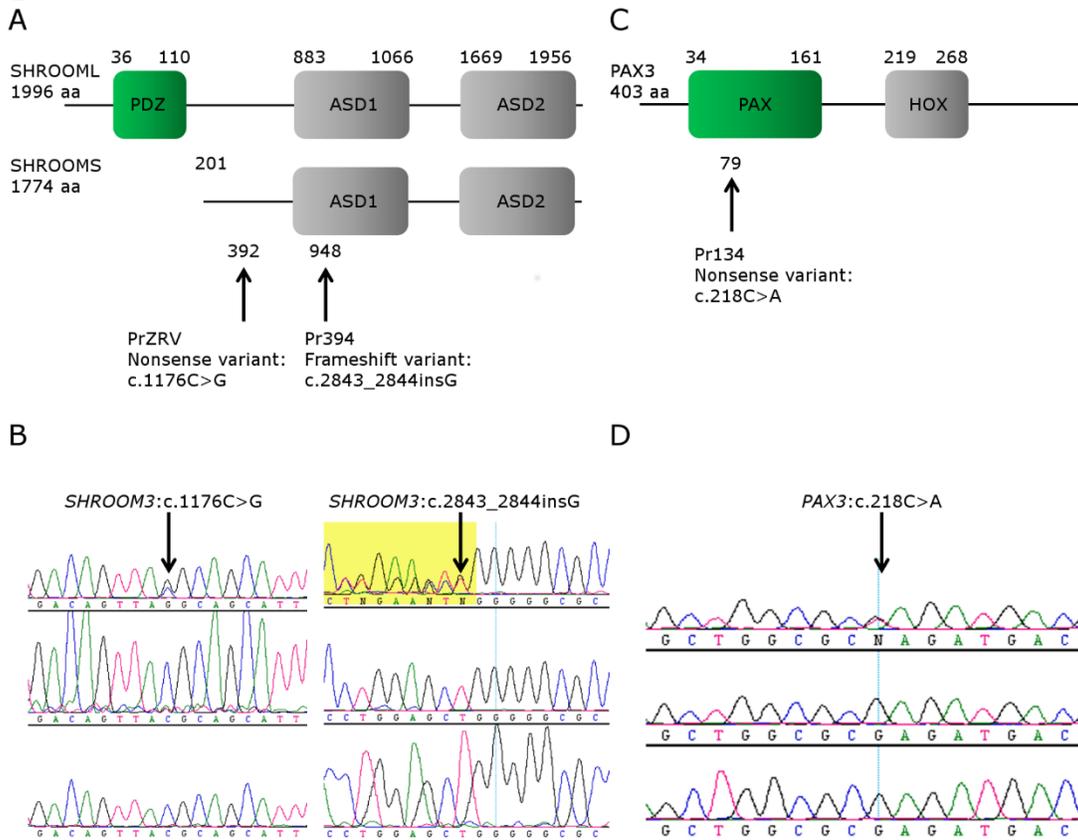


Figure 1: *De novo* mutations identified in *SHROOM3* and in *PAX3* in trios affected with NTDs (A) Protein schematic representation of the long and predicted short isoforms of *SHROOM3* indicating the position of the two DNMs identified in NTD probands. (B) Chromatograms of the 3 individuals of family PrZRV and Pr394 (from top to bottom: proband, mother, father) surrounding the *SHROOM3* c.1176C>G nonsense variant and the c.2843_2844insG frameshift variant. (C) Protein schematic representation of the long and predicted short isoforms of *PAX3* indicating the position of the DNM identified in NTD probands. (D) Chromatograms of the 3 individuals of family Pr134 (from top to bottom: proband, mother, father) surrounding the *PAX3* c.218C>A nonsense variant.

Table 1: *De novo* mutations identified in NTD affected trios and confirmed by Sanger sequencing.

| Loss of function <i>de novo</i> mutations (Nonsense, frameshift, splicing) | | | | | | | |
|--|-----------|-----|-----------|-----------------|-----------------|-------------------|-----------------|
| Family | NTD type* | Chr | Position | Genes | cDNA change | Amino acid change | Polyphen HDIV** |
| Pr394 | MMC | 4 | 77662169 | <i>SHROOM3</i> | c.2843_2844insG | p.L948fs | NA*** |
| PrZRV | Anen | 4 | 77660502 | <i>SHROOM3</i> | c.1176C>G | p.Y392X | NA |
| Pr134 | MMC | 2 | 223161800 | <i>PAX3</i> | c.218C>A | p.S73X | NA |
| Pr201 | MMC | 15 | 44116692 | <i>MFAP1</i> | c.69del7 | p.K23fs | NA |
| Pr389 | MMC | X | 41202544 | <i>DDX3X</i> | c.620dupA | p.Q207fs | NA |
| PrYPT | MMC | 7 | 73279630 | <i>WBSCR28</i> | c.380G>A | p.W127X | NA |
| Missense and synonymous <i>de novo</i> mutations | | | | | | | |
| Pr548 | MMC | 1 | 24668728 | <i>GRHL3</i> | c.1171C>T | p.R391C | 1 |
| Pr548 | MMC | 19 | 39798985 | <i>LRFN1</i> | c.1604C>T | p.T535I | 1 |
| Pr548 | MMC | 9 | 124751686 | <i>TLL11</i> | c.1327A>G | p.N443D | 0.968 |
| Pr125 | MMC | 19 | 5214591 | <i>PTPRS</i> # | c.4475G>A | p.R1049Q | 0.998 |
| Pr125 | MMC | 2 | 27435209 | <i>ATRAID</i> | c.138G>A | p.A46A | NA |
| Pr122 | MMC | 12 | 123341629 | <i>HIP1R</i> | c.1682G>T | p.G561V | 0.019 |
| Pr122 | MMC | 17 | 71232301 | <i>C17orf80</i> | c.1441C>T | p.R481W | 0.099 |
| Pr122 | MMC | X | 53592096 | <i>HUWE1</i> | c.6812G>A | p.S2271N | 0 |
| Pr134 | MMC | 14 | 105179874 | <i>INF2</i> # | c.2971C>T | p.R991W | 1 |
| Pr191 | MMC | 8 | 10464772 | <i>RP1L1</i> # | c.6836C>T | p.P2279L | 0.053 |
| Pr191 | MMC | 20 | 62371335 | <i>SLC2A4RG</i> | c.70C>T | p.R24C | 0.426 |
| Pr20 | MMC | 3 | 38317786 | <i>SLC22A13</i> | c.1246G>A | p.V416M | 0.948 |
| Pr20 | MMC | 16 | 31383022 | <i>ITGAX</i> | c.2077C>G | p.Q693E | 0.001 |
| Pr201 | MMC | 2 | 47703654 | <i>MSH2</i> # | c.1956A>G | p.Q652Q | NA |

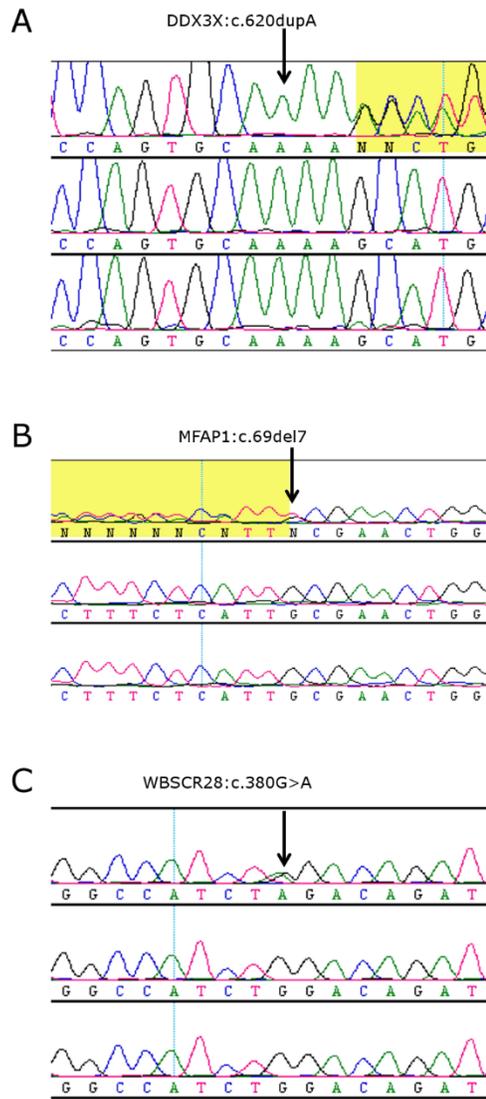
| | | | | | | | |
|-------|------|----|-----------|-----------------------------|-----------|----------|-------|
| Pr202 | MMC | 6 | 30122164 | <i>TRIM10</i> | c.1028A>T | p.D343V | 1 |
| Pr202 | MMC | 7 | 23775208 | <i>STK31</i> | c.535A>C | p.I179L | 0.039 |
| Pr25 | MMC | 9 | 134183554 | <i>PPAPDC3</i> [#] | c.696C>T | p.I232I | NA |
| Pr263 | MMC | 6 | 1390351 | <i>FOXF2</i> | c.169G>A | p.A57T | 0.01 |
| Pr28 | MMC | 8 | 124333387 | <i>ATAD2</i> | c.4160G>A | p.S1387N | 0.001 |
| Pr282 | MMC | 19 | 808439 | <i>PTBP1</i> | c.1233C>A | p.N411K | 0.013 |
| Pr389 | MMC | 7 | 100285176 | <i>GIGYF1</i> | c.325C>T | p.P109S | 0.728 |
| Pr402 | MMC | 17 | 19319353 | <i>RNF112</i> | c.1761C>T | p.A587A | NA |
| Pr402 | MMC | 3 | 49775724 | <i>IP6K1</i> | c.355C>T | p.R119C | 1 |
| Pr402 | MMC | 4 | 48424093 | <i>SLAIN2</i> | c.1745G>C | p.X582S | NA |
| Pr530 | MMC | 3 | 57616163 | <i>DENND6A</i> | c.1605A>C | p.E535D | 0.997 |
| Pr551 | MMC | 9 | 130279261 | <i>FAM129B</i> | c. 848C>T | p.A283V | 0.004 |
| Pr553 | MMC | 19 | 39329153 | <i>HNRNPL</i> | c.1441C>T | p.R481W | 1 |
| Pr554 | Anen | 3 | 142741447 | <i>U2SURP</i> | c.961G>A | p.G321S | 1 |
| Pr554 | Anen | 5 | 93966388 | <i>ANKRD32</i> | c.371T>C | p.F124S | 0.999 |
| Pr67 | MMC | 2 | 44566318 | <i>PREPL</i> | c.937C>G | p.L313V | 1 |
| Pr67 | MMC | 5 | 176314262 | <i>HK3</i> | c.1677G>A | p.V559V | NA |
| Pr67 | MMC | 9 | 33264606 | <i>BAG1</i> | c.67G>A | p.A23T | 0.897 |
| PrKKS | MMC | 15 | 43621819 | <i>LCMT2</i> [#] | c.869T>C | p.I290T | 0.201 |
| PrKKS | MMC | 7 | 11076097 | <i>PHF14</i> | c.1655G>C | p.R552P | 0.998 |
| PrTVB | MMC | 22 | 41558745 | <i>EP300</i> [#] | c.3690A>G | p.Q1230Q | NA |
| PrVWA | MMC | 9 | 120475791 | <i>TLR4</i> | c.1385C>T | p.A462V | 0.006 |

*MMC = myelomeningocele. Anen= anencephaly

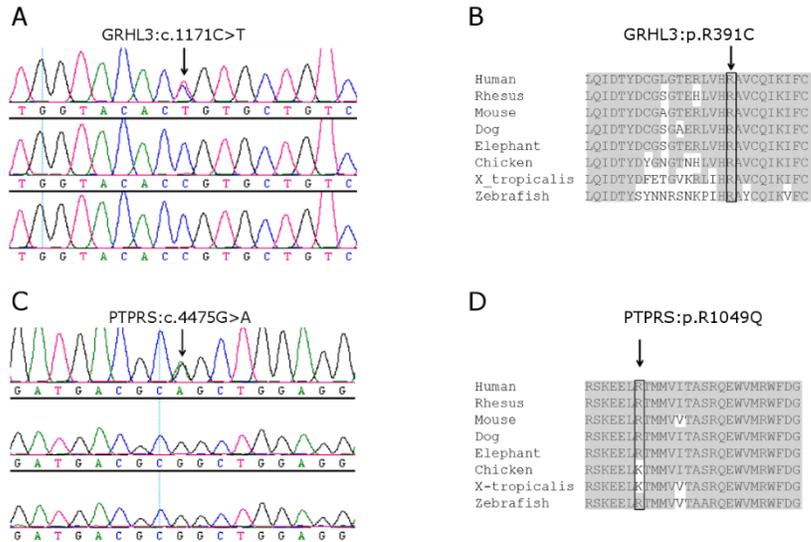
**Probably damaging (Polyphen HDIV \geq 0.957), possibly damaging (0.453 \leq Polyphen HDIV \leq 0.956); benign (Polyphen HDIV \leq 0.452)

*** NA = Non-applicable

= mutation previously reported in the ExAC database (<http://exac.broadinstitute.org/>).
Reported mutations frequencies are: *PTPRS* (frequency: 0.000008368); *INF2* (frequency: 0.000009451); *RP1L1* (frequency: 0.0001411); *MSH2* (frequency: 0.0001730), *PPAPDC3* (frequency: 0.0003366); *LCMT2* (frequency: 0.000008281); *EP300* (frequency: 0.0000082).



Supplementary Figure S1: LoF *de novo* mutations identified in *DDX3X*, *MFAP1* and *WBSR28* in 3 trios affected with NTDs. Chromatograms of the 3 trios (from top to bottom: proband, mother, father) showing the c.620dupA *DDX3X* frameshift variant (panel A), the c.69del7 *MFAP1* frameshift variant (panel B) and the c.380G>A *WBSR28* nonsense variant (panel C).



Supplementary Figure S2: *De novo* mutations identified in orthologues of genes associated with NTDs in mouse models. Chromatograms of the 2 trios (from top to bottom: proband, mother, father) showing the c.1171C>T *GRHL3* variant and (panel A) and the c.4475G>A *PTPRS* variant (panel C). Multi-species alignment of the amino acids surrounding the p.R391C variant in *GRHL3* (panel B) and the p.R1492Q variant in *PTPRS* (panel D).

Table S1: Description of the trios affected with neural tube defects included in this study

| Proband | NTD type | <i>De novo</i> per proband | Peri- conceptional folic acid intake | Multi-vitamins | Type of tissue |
|---|------------------|----------------------------------|---|----------------|----------------|
| Fetuses (average age at induced abortion: 21.7 ± 5.8 weeks) | | | | | |
| Pr101 | myelomeningocele | 0 | Yes | Yes | Liver |
| Pr122 | anencephaly | 3 | No | Unknown | Liver |
| Pr125 | myelomeningocele | 2 | Unknown | Unknown | Liver |
| Pr126 | anencephaly | 0 | Unknown | Unknown | Liver |
| Pr138 | myelomeningocele | 0 | Unknown | Yes | Liver |
| Pr258 | myelomeningocele | 0 | Yes | Yes | Amniocytes |
| Pr260 | myelomeningocele | 0 | Yes | Yes | Blood |
| Pr263 | myelomeningocele | 1 | Yes | Yes | Blood |
| Pr267 | myelomeningocele | 0 | No | No | Blood |
| Pr269 | anencephaly | 0 | Unknown | Unknown | Blood |
| Pr28 | myelomeningocele | 1 | Yes | Yes | Liver |
| Pr282 | anencephaly | 1 | Yes | Yes | Placenta |
| Pr32 | anencephaly | 0 | Unknown | Yes | Umbilical cord |
| Pr41 | anencephaly | 0 | Unknown | Unknown | Amniocytes |
| Pr554 | anencephaly | 2 | Unknown | Unknown | Blood |
| Pr79 | myelomeningocele | 0 | Yes | Yes | Umbilical cord |
| PrGVK | myelomeningocele | 0 | No | Yes | Umbilical cord |
| PrKKS | myelomeningocele | 2 | Yes | No | Umbilical cord |
| PrQKV | myelomeningocele | 0 | No | Yes | Umbilical cord |
| PrTVB | myelomeningocele | 1 | Yes | Yes | Umbilical cord |

| | | | | | |
|---|------------------|---|-----|-----|----------------|
| PrVWA | myelomeningocele | 1 | No | Yes | Umbilical cord |
| PrYPT | myelomeningocele | 1 | Yes | No | Umbilical cord |
| PrZRV | anencephaly | 1 | Yes | Yes | Umbilical cord |
| Patients (average age of proband: 17.5 ±14.8) | | | | | |
| Pr134 | myelomeningocele | 2 | No | No | Blood |
| Pr191 | myelomeningocele | 2 | No | No | Blood |
| Pr20 | myelomeningocele | 2 | No | No | Blood |
| Pr201 | myelomeningocele | 2 | No | No | Blood |
| Pr202 | myelomeningocele | 2 | No | No | Blood |
| Pr215 | myelomeningocele | 0 | No | No | Blood |
| Pr231 | myelomeningocele | 0 | No | No | Blood |
| Pr25 | myelomeningocele | 1 | Yes | No | Blood |
| Pr265 | myelomeningocele | 0 | Yes | No | Blood |
| Pr389 | myelomeningocele | 2 | No | No | Blood |
| Pr394 | myelomeningocele | 1 | No | No | Blood |
| Pr402 | myelomeningocele | 3 | No | No | Blood |
| Pr53 | myelomeningocele | 0 | No | No | Blood |
| Pr530 | myelomeningocele | 1 | Yes | No | Blood |
| Pr548 | myelomeningocele | 3 | Yes | Yes | Blood |
| Pr550 | myelomeningocele | 0 | Yes | No | Blood |
| Pr551 | myelomeningocele | 1 | Yes | No | Blood |
| Pr552 | myelomeningocele | 0 | Yes | No | Blood |
| Pr553 | myelomeningocele | 1 | Yes | No | Blood |
| Pr67 | myelomeningocele | 3 | Yes | No | Blood |

Supplementary Methods:

Ethics statement

All samples were collected with the approval of the Local Ethics Committees: CHU Sainte Justine Hospital (Protocols' numbers: 2598 and 2899) and Istituto Giannina Gaslini, Genoa, Italy (Protocol number: 213/2013). Written informed consent was obtained from all participating individuals.

Cohort

Forty-three families each composed of one affected child and two unaffected parents with no family history of NTDs were recruited through the Montreal Ste-Justine Hospital Spina Bifida Center, the 3D study of the Integrated Research Network in Perinatology of Quebec and Eastern Ontario and the Istituto Giannina Gaslini in Genoa, Italy. Detailed information including folate status, tissue of origin and type of NTDs of this cohort is summarized in Table S1. Further description of the families with discussed findings can be found in Table S2. Briefly, all 43 cases were affected with NTDs including 35 myelomeningocele and 8 anencephaly cases. A total of 21 cases were fetuses and 55.6 % took folate peri-conceptionally. Tissues from fetuses were all obtained following induced abortions. The average maternal and paternal ages were 30.0 ± 4.8 years and 30.7 ± 5.9 years respectively. Saliva or blood samples were obtained from affected patients and their parents. Liver biopsies or umbilical cord tissues were obtained from aborted fetuses.

Alignment and SNV annotation

Whole exome sequencing (WES) on genomic DNA was performed according to the manufacturer's procedures at the McGill University and Genome Quebec Innovation Center (MUGQIC). Capture was done using SureSelectXT Human All Exon V4 or the SureSelect Human All Exon 50Mb. Sequencing was done on Hiseq 2000 as pair-end with 3 samples per lane. Raw data were aligned to the GRCh37 build using Burrows-Wheeler Aligner (BWA)¹. Duplicated reads were removed and the rest were locally realigned followed by variant calling using GATK 2.6.4.^{2 3 4} Variants were annotated using ANNOVAR (<http://www.openbioinformatics.org/annovar/>). Coding regions defined by refseq (33 499 779bp) were covered in average by 132 reads with 90% of target covered 20 times or more. Identified DNMs were annotated with the following bioinformatics tools: CADD score⁵, Polyphen HDIV^{6 7 8}, RVIS rank⁹ and HI score¹⁰.

Statistical Analysis

To allow an accurate comparison of the LoF *de novo* rate in our sample to the control in the published data, the number of LoF *de novo* were divided by the number of coding bases (33 499 779bp) covered at least with 20 or more reads. This per-base LoF *de novo* rate was then compared to the published control rate for two LoF DNM rates of 0.05×10^{-8} and 0.17×10^{-8} .^{11 12} The rate of LoF DNMs in *SHROOM3* was obtained by multiplying the two per-base *de novo* rate with the length of *SHROOM3* long isoform (5991bp). The gene specific per-trio LoF *de novo* rate from Samocha KE et al., Nat genet, 2014¹³ was obtained by combining the rate of splicing, nonsense and frameshift mutations in *SHROOM3*. Significance for differences between those rates and the observed number of mutations was assessed using a two tailed binomial exact test performed using R V3.0.1.¹⁴ Bonferroni correction was defined using the Ensembl gene annotation (http://aug2014.archive.ensembl.org/Homo_sapiens/Info/Annotation#assembly). A total of 20 389 genes has been annotated in the GRCh38 suggesting an acceptance threshold of $0.05 / 20\ 389 = 2.45 \times 10^{-6}$.

Supplementary references:

- 1 Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinforma Oxf Engl* 2009;25:1754–60.
- 2 McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernytsky A, Garimella K, Altshuler D, Gabriel S, Daly M, DePristo MA. The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res* 2010;20:1297–303.
- 3 DePristo MA, Banks E, Poplin R, Garimella KV, Maguire JR, Hartl C, Philippakis AA, del Angel G, Rivas MA, Hanna M, McKenna A, Fennell TJ, Kernytsky AM, Sivachenko AY, Cibulskis K, Gabriel SB, Altshuler D, Daly MJ. A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nat Genet* 2011;43:491–8.
- 4 Van der Auwera GA, Carneiro MO, Hartl C, Poplin R, del Angel G, Levy-Moonshine A, Jordan T, Shakir K, Roazen D, Thibault J, Banks E, Garimella KV, Altshuler D, Gabriel S, DePristo MA. From FastQ Data to High-Confidence Variant Calls: The Genome Analysis Toolkit Best Practices Pipeline. In: *Current Protocols in Bioinformatics*. John Wiley & Sons, Inc. 2002. <http://onlinelibrary.wiley.com/doi/10.1002/0471250953.bi1110s43/abstract> (accessed 23 May2014).
- 5 Kircher M, Witten DM, Jain P, O’Roak BJ, Cooper GM, Shendure J. A general framework for estimating the relative pathogenicity of human genetic variants. *Nat Genet* 2014;46:310–5.
- 6 Adzhubei IA, Schmidt S, Peshkin L, Ramensky VE, Gerasimova A, Bork P, Kondrashov AS, Sunyaev SR. A method and server for predicting damaging missense mutations. *Nat Methods* 2010;7:248–9.
- 7 Liu X, Jian X, Boerwinkle E. dbNSFP: A Lightweight Database of Human Nonsynonymous SNPs and Their Functional Predictions. *Hum Mutat* 2011;32:894–9.
- 8 Liu X, Jian X, Boerwinkle E. dbNSFP v2.0: A Database of Human Non-synonymous SNVs and Their Functional Predictions and Annotations. *Hum Mutat* 2013;34:E2393–402.
- 9 Petrovski S, Wang Q, Heinzen EL, Allen AS, Goldstein DB. Genic intolerance to functional variation and the interpretation of personal genomes. *PLoS Genet* 2013;9:e1003709.
- 10 Huang N, Lee I, Marcotte EM, Hurles ME. Characterising and Predicting Haploinsufficiency in the Human Genome. *PLoS Genet* 2010;6:e1001154.
- 11 Sanders SJ, Murtha MT, Gupta AR, Murdoch JD, Raubeson MJ, Willsey AJ, Ercan-Sencicek AG, DiLullo NM, Parikshak NN, Stein JL, Walker MF, Ober GT, Teran NA, Song Y, El-Fishawy P, Murtha RC, Choi M, Overton JD, Bjornson RD, Carriero NJ, Meyer KA, Bilguvar K, Mane SM, Sestan N, Lifton RP, Günel M, Roeder K, Geschwind DH, Devlin B, State MW. De novo mutations revealed by whole-exome sequencing are strongly associated with autism. *Nature* 2012;485:237–41.
- 12 Iossifov I, O’Roak BJ, Sanders SJ, Ronemus M, Krumm N, Levy D, Stessman HA, Witherspoon KT, Vives L, Patterson KE, Smith JD, Paepfer B, Nickerson DA, Dea J, Dong S, Gonzalez LE, Mandell JD, Mane SM, Murtha MT, Sullivan CA, Walker MF, Waqar Z, Wei L, Willsey

AJ, Yamrom B, Lee Y-H, Grabowska E, Dalkic E, Wang Z, Marks S, Andrews P, Leotta A, Kendall J, Hakker I, Rosenbaum J, Ma B, Rodgers L, Troge J, Narzisi G, Yoon S, Schatz MC, Ye K, McCombie WR, Shendure J, Eichler EE, State MW, Wigler M. The contribution of de novo coding mutations to autism spectrum disorder. *Nature* Published Online First: 29 October 2014.
doi:10.1038/nature13908

13 Samocha KE, Robinson EB, Sanders SJ, Stevens C, Sabo A, McGrath LM, Kosmicki JA, Rehnström K, Mallick S, Kirby A, Wall DP, MacArthur DG, Gabriel SB, DePristo M, Purcell SM, Palotie A, Boerwinkle E, Buxbaum JD, Cook Jr EH, Gibbs RA, Schellenberg GD, Sutcliffe JS, Devlin B, Roeder K, Neale BM, Daly MJ. A framework for the interpretation of de novo mutation in human disease. *Nat Genet* 2014;46:944–50.

14 R Development Core Team. The R Project for Statistical Computing. <http://www.r-project.org/> (accessed 14 Aug2013).

5. Rare deleterious variants in *GRHL3* are associated with human spina bifida

Philippe Lemay,¹ Patrizia De Marco,² Alexandre Emond,¹ Dan Spiegelman,³ Alexandre Dionne-Laporte,³ Sandra Laurent,³ Elisa Merello,² Andrea Accogli,² Guy A Rouleau,³ Valeria Capra² and Zoha Kibar¹

¹CHU Sainte Justine Research Center and University of Montréal, Montréal, Québec H3T 1C5 Canada, ²U.O. Neurochirurgia, Istituto Giannina Gaslini, Genova, 16147, Italy, ³Montreal Neurological Institute, McGill University, Montréal, Québec H3A 2B4, Canada

* Correspondence to Zoha Kibar, University of Montréal, Department of Neurosciences, CHU Sainte Justine Research Center, 3175 Cote-Sainte-Catherine, Room A711, Montreal, QC H3T 1C5 Canada; Tel: (514) 345 4931 X 3984; Fax: (514) 345 4801; E-mail : zoha.kibar@recherche-ste-justine.qc.ca

Contributions des co-auteurs

Philippe Lemay:

- Participation au développement de l'hypothèse
- Développement de filtres et de programmes informatiques permettant l'analyse des données finales du séquençage de l'exome et des mips
- Validation des variants identifiés
- Validations fonctionnelles par luciférase et mini-gène
- Analyses statistiques
- Conception des figures et rédaction du manuscrit

Patrizia De Marco :

- Caractérisation des phénotypes et analyses des données cliniques
- Participation au recrutement de la cohorte

Alexandre Emond :

- Clonage des gènes utilisés dans les validations par luciférase et mini-gène

Dan Spiegelman, Alexandre Dionne-Laporte :

- Transformation des données brutes de séquençage de l'exome et des mips en format d'analyse
- Support informatique

Sandra Laurent :

- Contrôle de qualité et analyse des données brutes des mips

Elisa Merello, Andrea Accogli :

- Recrutement de la cohorte de 192 patients et des 7 familles multiplexes

Guy A Rouleau, Valeria Capra :

- Aide au recrutement
- Support financier et technique

Zoha Kibar :

- Supervision du projet
- Soutien scientifique, technique et matériel pour la réalisation du projet
- Rédaction du manuscrit

Abstract

Neural tube defects (NTDs), including spina bifida, are among the most common birth defects caused by failure of neural tube closure during development. They have a complex etiology involving largely undetermined environmental and genetic factors. Previous studies in mouse models have implicated the transcription factor *Grhl3* as an important factor in the pathogenesis of spina bifida. In the present study, we conducted a resequencing analysis of *GRHL3* in a cohort of 233 familial and sporadic cases of spina bifida. We identified 2 novel truncating variants: one homozygous frameshift variant, p.Asp16Aspfs*10, in 2 affected siblings and one heterozygous intronic splicing variant, p.Ala318Glyfs*26. We also identified 5 missense variants, one of which was demonstrated to reduce activation of the gene targets in a luciferase reporter assay. With the previously identified p.Arg391Cys variant, 8 variants were found in *GRHL3*. Comparison of the variant rate between our cohort and the ExAC database identified a significant enrichment of deleterious variants in *GRHL3* in the whole gene and the trans-activation region in spina bifida patients. These data provide strong evidence for a role of *GRHL3* as a predisposing factor to spina bifida and will help dissect the complex etiology and pathogenic mechanisms of these malformations.

Key Words

GRHL3, Neural Tube Defects, Spina Bifida, Whole Exome Sequencing, Molecular Inversion Probes

Introduction

Neural tube defects (NTDs, MIM# 182940) are among the most common congenital malformations affecting 1-2:1000 births. They are caused by an incomplete closure of the neural tube during early embryonic development [Bassuk and Kibar 2009]. They are phenotypically diverse depending on the site and severity of the closure defect. The most common clinical presentations are spina bifida (myelomeningocele) and anencephaly which result from failure of neural tube closure in the spinal and cranial regions respectively [Bassuk and Kibar 2009]. Consequences of these malformations range from large developmental physical and psychological disabilities in spina bifida to death during the early embryonic stage or soon after birth in anencephaly. Peri-conceptual folic acid intake has been shown to reduce the prevalence of NTDs by 50-70% [MRC Vitamin Study Research Group 1991]; however, a large number of cases remain folate resistant urging the need for identification of novel causative factors which could lead to the development of novel preventive strategies.

The etiology of NTDs in humans is multifactorial involving environmental and genetic factors with an estimated heritability of 60% [Bassuk and Kibar 2009]. While having a complex genetic etiology in humans, previous studies have identified over ~250 causative genes for NTDs in mice segregating in a Mendelian fashion [Harris and Juriloff 2007; Harris and Juriloff 2010; Greene and Copp 2005]. Molecular genetic studies of orthologues of the mouse NTD genes have had some success in identification of predisposing NTD genes in humans. Such studies were particularly powerful in demonstrating an important role of the planar cell polarity pathway in a fraction of human NTD patients [Juriloff and Harris 2012; Kibar et al. 2007; Copp et al. 2013]. Other candidate gene approaches investigated the role of folate-related genes in the etiology of human NTDs and identified few NTD-associated variants that accounted for a small part of the disease etiology [Zhang et al. 2013]. Whole-genome linkage analysis have shown significant association to three chromosomal regions on chromosome 2, 7 and 10, but failed to identify any causative gene [Stamm et al. 2008; Rampersaud et al. 2005]. These results need to be interpreted with caution as these kinds of studies are most likely complicated by clinical and genetic heterogeneity of NTDs.

Our group has recently conducted a whole exome sequencing study in 43 trios affected with either spina bifida or anencephaly and identified among others, the de novo missense variant, p.Arg391Cys, in *GRHL3* (Grainyhead Like Transcription Factor 3, MIM# 608317), predicted to be deleterious, in one case of spina bifida [Lemay et al. 2015]. *GRHL3* is a transcription factor involved in the epidermal integrity and

wound healing [Gustavsson et al. 2008]. It was also associated to the regulation of E-Cadherin in the epithelial to mesenchymal transition [Alotaibi et al. 2015]. The human GRHL3 consists of three known domains: the trans-activation, the DNA binding and the dimerization domains. It has four isoforms, with three alternative first exons and two alternative last exons which were previously shown to have distinct and even opposing functions [Peyrard-Janvid et al. 2014; Caddy et al. 2010]. While the full spectrum of activity of this gene remains unknown, it was strongly associated to NTDs in the Curly tail mouse model [Gustavsson et al. 2008] and in *Grhl3* knockout mice [Ting et al. 2003]. Curly tail (Ct) is one of the most well-established models for NTDs due to the complex nature of its transmission which mimics the human pathology. Homozygous mutant ct/ct embryos develop lumbosacral spina bifida (10–20%) or exencephaly (1–5%). Linkage studies have mapped the Ct gene to a candidate genomic region containing *Grhl3* but has not identified a pathogenic coding variant. However, the diminished expression of *Grhl3* in mutant neural tissues, the complete rescue of spinal NTDs by reinstatement of *Grhl3* expression in ct/ct embryos, the 100% penetrance of spina bifida in *Grhl3* knockout mice and the genetic interaction of Ct with these knockout mice, provide strong evidence that Ct is a hypomorphic allele of *Grhl3* [Ting et al. 2003; Neumann et al. 1994]. While these data confirm an important role of *Grhl3* in the pathogenesis of NTDs in mouse models, its role in the human NTD pathology remained unclear [Ting et al. 2003]. To date, variants in *GRHL3* in humans have been associated only with a minority (5%) of cases presenting Van Der Woude syndrome 2 (VWS2, MIM# 606713) (the most common syndromic form of cleft lip and palate [Peyrard-Janvid et al. 2014]) and recently with nonsyndromic cleft palate [Mangold et al. 2016; Leslie et al. 2016].

The strong implication of *Grhl3* in mouse NTDs as well as our initial WES findings have prompted us to investigate the role of this gene in a large cohort of human NTDs. We focused our study on spina bifida that is the most common and severe form of spinal NTDs. We hypothesized that de novo as well as inherited variants in *GRHL3* predispose to spina bifida in a fraction of patients. Using new WES data from 6 familial spina bifida cases (2 affected cases per family), our previously published WES data in 35 spina bifida trios and new Molecular Inversion Probe (MIP) sequencing in 192 unrelated sporadic cases, we identified 8 deleterious variants in *GRHL3* in spina bifida. When compared to public databases, this represented a significant enrichment of deleterious variants which disproportionately affected the first 74 amino acid of GRHL3 that mainly includes the trans-activation domain. Notably, a frameshift homozygous variant and a heterozygous splicing variant both leading to premature stop codons were identified in 2 unrelated spina bifida cases. We next used a *GRHL3* specific binding sequence coupled to a luciferase reporter for functional validation of NTD-associated variants in *GRHL3* where we demonstrated

a reduced gene activity for one heterozygous missense variant. Our data provide a first line of evidence of a pathogenic role of *GRHL3* in human spina bifida and help better understand the complex etiology of this group of spinal NTDs.

Subjects, Materials and Methods

Cohort

Three multiplex families each with 2 affected cases and 192 unrelated sporadic cases of spina bifida were recruited through the Istituto Giannina Gaslini in Genoa. All samples were collected with the approval of the Local Ethics Committees: Istituto Giannina Gaslini, Genoa, Italy (Protocol number: 213/2013). Written informed consent was obtained from all participating individuals.

Whole exome sequencing and molecular inversion probe sequencing

Whole exome sequencing (WES) of the genomic DNA of the 6 spina bifida familial cases was performed according to the manufacturer's procedure at the McGill University and Genome Quebec Innovation Center (MUGQIC). Exome Capture was done using SureSelectXT Human All Exon V4. Pair-end sequencing was done on Hiseq 2000 with 3 samples per lane. Raw data was aligned to the GRCh37 genome reference build using the Burrows-Wheeler Aligner (BWA). Duplicate reads were removed, remaining reads were realigned locally and variants were called using GATK 2.6.4. Variants annotation was done using Annovar (<http://www.openbioinformatics.org/annovar/>). This resulted in 96.1% of the 33 499 779 coding base pairs being covered at least 20x. Called variants were filtered for minimal quality (GQ>50, total variant reads>10, total reads>20), a MAF of <1% in the Exome Variant Server at <http://evs.gs.washington.edu/EVS/> and a shared transmission in both affected of a family. Identified variants in *GRHL3* were validated by Sanger sequencing.

Regions of interest were defined as the coding regions of all the isoforms of *GRHL3* (NM_021180.3, NM_198173.2, NM_198174.2, NM_001195010.1). Molecular inversion probes (MIPs) were designed using the scripts of O'Roak et al [O'Roak et al. 2012] available at: http://krishna.gs.washington.edu/mip_pipeline/. The absence of SNPs in the MIPs arms was verified to ensure unbiased capture. A total of 57 MIPs was designed to cover the *GRHL3* gene using the standard capture length of 112bp. The MIPs were then tested for capture level, and poor performers were adjusted in the final mix. The sequencing reactions of the 57MIPs in the 192 unrelated cases were ran on one lane of Hiseq 2000 resulting in an average coverage of at least 500X in 93.4% of the *GRHL3* gene. Identified

variants were validated by Sanger sequencing. The numbering system of the variants at the nucleotide level and the amino acid change are based on *GRHL3* cDNA and protein sequence with the accessions NM_198173.2 and NP_937816.1 respectively. Nucleotide numbering uses +1 as the A of the ATG translation initiation codon in the reference sequence, with the initiation codon as codon 1.

Statistical analysis of the sequencing data

GRHL3 variants identified in the exome sequencing data of the 6 familial spina bifida cases (2 cases per family), the MIPS sequencing of 192 cases and our previously published exome data on 35 spina bifida cases [Lemay et al. 2015] were filtered for a phred CADD score [Kircher et al. 2014] over 20 and frequency in the EXAC public databases under 1%. CADD phred score was obtained through CADD v1.0 which was done through the CADD portal (<http://cadd.gs.washington.edu/>). Familial variants were counted in only one individual as to not bias the enrichment analysis. Control variants were obtained through the ExAC database and filtered in the same manner. The expected frequency for these variants was obtained by dividing the number of control filtered variants by the number of coding and intronic splicing bases in *GRHL3* isoform 2 multiplied by the number of alleles sequenced ($860 \text{ variants} / ((1809 + 30) \text{bp} * 60706 * 2 \text{ alleles}) = 3.851 \times 10^{-6}$). The region containing the trans-activation domain was defined by the first 74 amino acids of the protein. The variant frequency for this region was obtained by adapting the number of variants and the number of base pairs covered ($87 \text{ variants} / ((222 + 8) \text{bp} * 60706 * 2 \text{ alleles}) = 3.116 \times 10^{-6}$). Variant enrichment was assessed using a two-tailed exact Poisson test with R v3.3.1. Identified variants were submitted to the LOVD 3.0 locus-specific database (<http://databases.lovd.nl/shared/genes/GRHL3>).

The *GRHL3* exon 8 trapping experiment

Wild-type *GRHL3* exon 8 flanked by 179bp of intron 7 and 197 bp of wild-type or mutant intron 8 (carrying the intronic variant c.1047+2T>C) was cloned in the exon trapping vector pSPL3 [Nisson et al. 1994] using the SLIC [Li and Elledge 2012] method and the following primers: forward, 5'-GAGCGGCCGCTGCAGGATCCCAGATCAGAGCTTTGGAATCC-3' and reverse, 5'-GTACGGGATCACCAGATATCGGTACAAGTTCTGAGCTCC-3'. The restriction sites used for cloning are underlined and the gene specific sequences are indicated in Italics. This resulted in the cloning of the exon 8 of the wild-type or mutant *GRHL3*. Exon 8 trapping vector was transfected in HEK293 cells with Lipofectamine 2000 (catalog#:11668027) using 5 ug of DNA and 15 ul of lipofectamine in a 10cm dish. RNA extraction was performed using TRIZOL (catalog#: 15596026) with the manufacturer's protocol. RNA was treated with 1 unit of invitrogen DNase (catalog#: 18068015) for an hour and RNA was purified using

phenol-chloroform method. Reverse transcriptase was performed on 1ug of RNA using poly dT oligos, Invitrogen's superscript II (catalog#: 18064014) and Invitrogen's RNaseOUT (catalog#: 10777019) following the manufacturer protocol. Finally, beta-globin primers of the pSL3B vector were used to amplify the expressed construct.

Luciferase Reporter Assay

Isoform 2 and isoform 3 of *GRHL3* were both cloned in the Invitrogen's pcDNA3.1 (-) Myc-His A vector (catalog#: V85520). Isoform 2 was cloned using the SLIC [Li and Elledge 2012] method with the following primers: forward, 5'-ACGGGCCCTCTAGACTCGAGATGTCTGAATGAACTTGATTTACAG-3' and reverse, 5'-TTTTGTTCGGGCCCAAGCTTTTACAGCTCCTTAAGGATGATC-3' while isoform 3 was cloned using the same forward primer and the reverse primer, 5'-TTTTGTTCGGGCCCAAGCTTTCATTCCGAGTTTACAGTGAAG-3'. *GRHL3* variants were introduced in isoform 2 by in site directed mutagenesis.

Functional validation of the *GRHL3* variants was done using a *GRHL3* specific binding sequence coupled to a luciferase reporter construct that was kindly provided by Dr. Robert A. Cornell (University of Iowa) [Leslie et al. 2016]. HEK293 cells were grown to confluence in a 10cm Petri dish and then equally divided in a 24 wells plate. Transfection was done 24 hours after dividing the cells using Lipofectamine 2000 with 500 ng of the wildtype or mutated *GRHL3*, 500 ng of the luciferase reporter plasmid and 100 ng of Renilla pRL-TK plasmid. Cells were collected 24h after transfection and lysed using the promega Dual-Luciferase passive lysis buffer. Ten measurements were done for each of the 3 biological replicates and the average was used for comparison. The luciferase activity was normalised to the Renilla activity for every sample and the ratios are reported. The wildtype *GRHL3* isoform 2 and 3 were both tested using this protocol, but only the isoform 2 was finally used for functional validation of *GRHL3* variants. Differences between isoform 2 and 3 were assessed using a two-tailed unpaired t-test, while differences between variants of the isoform 2 were tested using an ANOVA with Tukey Honest Significant Differences post-hoc test using R v3.3.1 with a threshold of 0.05.

Western blot analysis

Cellular extracts from HEK293 transfected cells were extracted using RIPA and protease inhibitor (Roche's cOmplete). Cellular extracts were centrifuged to remove membrane extracts. Proteins were quantified using the Pierce BCA assay from Thermo Fisher (catalog#: 23225) and 75 ug of each extract was loaded on a 8% bis-acrylamide gel subjected to SDS PAGE. Protein were transferred to a PVDF membrane

from Immobilon (catalog#: IPVH00010). The membrane was blocked using a 5% milk/TBS solution overnight at 4°C. Primary incubation was done using a 5% milk/TBS solution with anti-Myc antibody (Invitrogen 46-0603) for 4 hours at room temperature followed by a 10 minutes' wash in TBST. Secondary incubation was done using a 5% milk/TBS solution with anti-mouse HRP antibody (abcam, Cambridge, MA) followed by three 10 minutes' wash. Finally, the signal was revealed using SuperSignal™ West Femto Maximum Sensitivity Substrate (catalog#: 34095).

Results

Sequencing analyses of *GRHL3* in spina bifida

WES was performed on 6 spina bifida familial cases (3 families with each having 2 affected individuals). Annotated variants in *GRHL3* shared by both affected members of each family were investigated for loss of function (nonsense, frameshift, splicing) variants and variants with a phred CADD score over 20 [Kircher et al. 2014]. This successfully identified a novel homozygous frameshift variant in *GRHL3* shared by the two affected siblings in one of the three families. Both siblings were also affected with scoliosis, and one sibling was also affected with non-syndromic cleft lip and palate. This family had reported consanguinity with both parents being first degree cousins. The identified variant is heterozygous in both unaffected parents. It introduced a C at base pair 48 leading to a c.48dupC (p.Asp16Aspfs*10) protein change and truncation (Table 1, Figure 1 and 2A). This variant completely truncates isoforms 1, 2 and 3 of *GRHL3*, while it does not affect isoform 4 as it maps 5' of the alternative start codon of this isoform.

Following the identification of an inherited frameshift variant in a multiplex spina bifida family, we re-analysed our published WES data on 35 spina bifida trios for the presence of inherited potentially deleterious variants. We detected a heterozygous father-inherited variant c.101C>T (p.Thr34Met) in a 10 year old girl affected with lumbosacral myelomeningocele. This variant was present at a very low frequency in the ExAC database. It is a non-conservative change that affects a conserved threonine residue in the trans-activation domain of *GRHL3* (Table 1, Fig 2A, Supp. Figure S1).

We next extended our resequencing analyses of *GRHL3* to a large cohort of 192 sporadic spina bifida cases. MIP sequencing identified 4 heterozygous missense variants with CADD score >20, c.152A>G (p.Asn51Ser), c.208C>A (p.Pro70Thr), c.862G>C (p.Asp288His), and c.1586C>T (p.Ala529Val), and one heterozygous intronic splicing variant (Table 1, Figure 2A). The intronic c.1047+2T>C (p.Ala318Glyfs*26)

splicing variant was novel and not detected in public databases. It affects the second base pair of intron 8 and was detected in a 27 years old male affected with myelomeningocele, Chiari II malformation and hydrocephalus. To determine whether this variant affects the splicing of intron 8, we cloned exon 8 flanked by 179bp of intron 7 and 197 bp of intron 8 in the exon trapping vector pSPL3 and performed reverse transcribed-PCR on RNA isolated from HEK293 cells transfected with either wild-type or mutant *GRHL3* exon 8 trapping vector (Figure 3A). This resulted in a product of 350bp (77 exonic + 273 artificial exons) for the wildtype and of 273bp for the mutant. Complete absence of the original product in the mutant RNA suggests a complete disruption of the splicing around exon 8 (Figure 3B). Sequencing of the wild-type and mutant RT-PCR products further confirmed that the variant resulted in the skipping of exon 8 which would lead to a premature stop codon p.Ala318Glyfs*26. The resulting protein would have half of its binding domain and its whole dimerization domain truncated (Figure 2A).

The variant p.Pro70Thr was novel and not detected in public databases. It was identified in a 19 years old female affected with myelomeningocele and Chiari II malformation and affects a highly conserved proline residue in the trans-activation domain of GRHL3 (Figure 2, Supp. Figure S1). The 3 other missense variants in *GRHL3* detected by MIP sequencing were reported at very low frequencies in the ExAC database. The variant p.Asn51Ser was detected in a 26 year old female affected with myelomeningocele, Chiari II malformation, hydrocephalus and hydromyelia. It affects a conserved asparagine residue in the trans-activation domain of GRHL3. The variant p.Asp288His was detected in a 44 year old male affected with myelomeningocele and affects a highly conserved residue in the DNA binding domain of GRHL3. The variant p.Ala529Val was detected in a 26 year old female affected with myelomeningocele Chiari II malformation, hydrocephalus and hydrosyringomyelia and affects a perfectly conserved residue across 5 species in the dimerization domain of GRHL3.

In total, our resequencing analysis of *GRHL3* in 6 familial cases, 35 previously published spina bifida trios and 192 spina bifida sporadic cases, identified one homozygous frameshift variant, one heterozygous intronic splicing variant and 6 heterozygous missense variants (one of which was previously reported) which resulted in a total of 8 variants. One variant was detected in a homozygous state resulting in a total of 9 mutated alleles. Interestingly, 5 of the 9 affected alleles mapped to the first 74 amino acids of the N terminal region including the trans-activation domain of the gene (Figure 2A). Statistical enrichment for deleterious alleles in *GRHL3* in spina bifida was assessed using the ExAC database as controls. In the latter, 3486 rare mutated alleles (MAF 1%) were identified in *GRHL3* of which 860 had a phred CADD score over 20. Of these 860 potentially deleterious alleles, 87 mapped to the region including

the trans-activation domain. The number of variants was normalized on the total number of base pairs screened and enrichment of deleterious variants in *GRHL3* in spina bifida was assessed using a two-tailed exact Poisson test. This resulted in a significant enrichment of deleterious variants in the whole gene ($P = 6.922 \times 10^{-3}$) and in the first 74 amino acids containing the trans-activation domain ($P = 2.622 \times 10^{-5}$). The rate of affected alleles in this region was then compared to the rate of variants in the whole gene in our spina bifida cohort. This once again resulted in a significant enrichment ($P = 5.973 \times 10^{-3}$) suggesting that the region affecting the trans-activation domain of *GRHL3* is disproportionately mutated in our cohort as compared to the rest of the gene. In the ExAC database, this trend was reversed and the trans-activation domain had significantly less variants ($P = 0.048$) than the whole *GRHL3* gene further supporting our observation. This enrichment in this specific domain in spina bifida is quite interesting as none of the previously reported pathogenic variants in either VWS2 or non-syndromic cleft lip/palate mapped to this domain.

Functional validation of *GRHL3* variants in spina bifida

We next examined the effect of the 6 potentially deleterious missense variants identified in *GRHL3* in spina bifida on its ability to activate luciferase reporter expression in HEK293 cells. We used a previously described luciferase reporter vector that has four replicates of a consensus *GRHL3* binding motif [Leslie et al. 2016]. We first chose the longest isoform 3 of *GRHL3* for this assay. Surprisingly, isoform 3 failed to activate the reporter and behaved like the control empty expression vector (Figure 2D). We next tested isoform 2 that is identical to isoform 3 except at the dimerization domain. The dimerization domain of isoform 2 is also shared by isoforms 1 and 4, while isoform 3 uses an alternative exon which results in a dimerization domain containing 61 different amino acids. In cells transfected with wild-type *GRHL3* isoform 2, the level of reporter activity was 17-fold higher than in cells transfected with an empty expression plasmid (control-transfected cells) (Figure 2D). Western blot analysis showed that both proteins were expressed and surprisingly isoform 3 that has no reporter activity seemed to have a higher expression level. All 6 missense variants were then inserted in *GRHL3* isoform 2 and tested for their ability to activate the reporter in HEK293 cells. As previously reported, the variant p.Arg391Cys (also detected in one patient of VWS2) had no ability to activate the reporter. One rare variant p.Ala529Val identified in our study in spina bifida had also lost this activity providing strong functional evidence for a pathogenic effect. The other 4 missense variants predicted to be pathogenic behaved like the wild-type in this assay (Figure 2E). All variants had comparable expression levels upon western analysis in transfected HEK293 cells (Figure 2C).

Discussion

Variants in *Grhl3* have long been associated with NTDs in mice. Our study represents the first line of evidence of association of this gene with NTDs in humans. By using a combination of whole exome and MIP sequencing of *GRHL3* in sporadic and familial spina bifida patients, we identified one homozygous frameshift deletion, one heterozygous intron splicing variant and 6 heterozygous missense variants with a CADD score >20 (including a de novo previously reported p.Arg391Cys variant). Three of these variants were novel, p.Asp16Aspfs*10, p.Pro70Thr, p.Ala318Glyfs*26, and were not reported in the ExAC database. The p.Asp16Aspfs*10 variant maps to exon 2 of the gene and truncates most of the protein in 3 of its 4 isoforms. It was detected in a homozygous state in 2 affected siblings where both parents were carriers suggesting a recessive mode of transmission and making it the most deleterious variant reported to date for *GRHL3*. No truncating variant mapping to this region in the 5' end of *GRHL3* has ever been reported in the literature or public databases. The p.Ala318Glyfs*26 resulted from an intronic splice variant that caused skipping of exon 8 in a gene trapping experiment and which leads to truncation of the protein at amino acid 345 and removing almost half of the DNA binding domain and the whole dimerization domain. The latter is important for stabilization of the DNA-protein complex.

Calculations on genic intolerance by ExAC classified *GRHL3* as highly intolerant toward loss-of-function variants with a pLI of 1. This score is justified by the fact that only 1 LoF variant was found, while a predictive algorithm expected to find 20 [Consortium et al. 2016]. This high discrepancy strongly supports the intolerance of the *GRHL3* gene. This gene was also predicted to be intolerant to missense variants using the same method with a Z score of 1.38. In our study, we compared the frequency of deleterious variants in *GRHL3* in our spina bifida cohort to that reported in the ExAC database and we detected a statistical enrichment of deleterious variants in *GRHL3* in spina bifida further supporting a pathogenic role of this gene in human NTDs.

Functional validation of the 6 missense *GRHL3* variants reported in this study using a luciferase reporter assay confirmed the inability of p.Arg391Cys to activate its target and identified p.Ala529Val as a new variant impairing target activation. The 4 other variants did not affect gene function in this reporter assay despite their high CADD score. Additional cell- or animal model- based assays are needed to investigate their predicted pathogenicity. Interestingly, *Grhl3* was demonstrated to be an important regulator of the transcription of RhoGEF19 during epidermal migration and wound repair, in the context of non-canonical WNT planar cell polarity signalling (PCP) pathway [Caddy et al. 2010]. In this same study,

authors proposed that transcriptional regulation of RhoGEF19 will also be important for other Grhl3-dependent morphogenetic events, such as neural tube closure. In the fly, the function of the grainy head transcription factor is essential for the function of the frizzled pathway during planar polarity of wing hair cells, providing additional evidence for a role of this gene in PCP signaling. Variants in PCP signaling have been strongly associated with NTDs in mouse models and human cohorts [Juriloff and Harris 2012; Kibar et al. 2007; Copp et al. 2013]. Another study has demonstrated an important role of Grhl3 in directing canonical Wnt signaling during cell fate specification of the neural plate border, which is crucial for neural tube closure [Kimura-Yoshida et al. 2015]. It was proposed that failure of fate specification of progenitor cells in the neural plate border/neural fold coordinated with neural tube closure may be associated with human NTDs. Testing the potential pathogenic effect of *GRHL3* variants identified in our spina bifida cohort on non-canonical Wnt/PCP as well as Wnt canonical signaling would help better define the role of GRHL3 in normal and abnormal neurulation.

Pathogenic variants in *GRHL3* have been described in VWS2 (the most common syndromic form of cleft lip and palate) and in nonsyndromic cleft palate [Peyrard-Janvid et al. 2014; Caddy et al. 2010]. In mouse, loss of Grhl3 results in spina bifida and in abnormal oral periderm leading to a cleft palate in 17% of embryos confirming a multifaceted role of this gene in both neurulation and craniofacial development. A recent study based on literature review and bioinformatics analyses have identified common regulatory factors in orofacial and neural tube development, including *GRHL3* [Kousa et al. 2016]. We have previously reported that a de novo variant p.Arg391Cys detected in one patient affected with spina bifida was also reported in one case of VWS2. Interestingly, in this study, one of the two siblings carrying the homozygous p.Asp16Aspfs*10 frameshift variant was also affected with cleft lip. Hence, we hypothesize that the phenotypic expression of the same variant in *GRHL3* in different individuals is likely to be modified by the presence of other genetic risk factors, environmental exposures, epigenetic marks, or stochastic events. These factors represent a secondary insult that could affect the same biochemical pathway as *GRHL3* or similar pathways implicated in craniofacial and/or neural tube developmental processes.

Another interesting finding in our study is the significant enrichment of deleterious variants in the region containing the trans-activation domain of GRHL3 as compared to the rest of the gene in our spina bifida cohort. This is in sharp contrast to variant reported in VWS2 or nonsyndromic cleft palate where none mapped to this domain (Supp. Figure S2). While this could be due to chance because of the low number of identified variants in VWS2, we hypothesize that the NTD-associated variants identified in this domain most likely affect the interaction of GRHL3 with co-regulators that specifically modulate the

transcriptional activity of this gene during neural tube formation. This important genotype-phenotype correlation needs to be confirmed in a larger cohort of patients.

Transcriptional properties of *GRHL3* can also be finely tuned by the different roles of the various isoforms as previously reported for other transcription factors including Pax genes [Monsoro-Burq 2015]. Different isoforms of *Grhl3* were previously shown to have different and even opposing functions in cell migration, apoptotic activity and nitric oxide synthase in endothelial cells as well as in their abilities to induce severe malformations in zebrafish embryos [Haendeler et al. 2013]. Microarray analyses revealed different target genes between the two *Grhl3* isoforms, and it was proposed that this could explain the opposite effects of these proteins in endothelial cells. In our study, we demonstrated that while isoform 2 could activate the reporter gene in HEK293 cells, isoform 3 failed to do so. This was not due to lack of expression of isoform 3 since it was proven by western blot that both proteins were expressed. Further investigations through combined transfections of multiple isoforms affected by various variants or deletion constructs are needed to better understand the seemingly complex interactions affecting this gene function. Identification of key target genes of all *GRHL3* isoforms would be crucial for a better understanding of their versatile roles in multiple developmental processes including neurulation and in human diseases.

GRHL3 belongs to a family of 3 highly conserved transcription factors that have numerous roles in developmental processes including neural tube closure, epidermal barrier formation, wound healing, craniofacial development and neural patterning/morphogenesis. These homologues are highly similar in the DNA binding and dimerization domains. *Grhl1* knockout mice do not show any NTDs, are viable and fertile but they display hair loss and symptoms reminiscent of palmoplantar keratoderma [Wilanowski et al. 2008]. *Grhl2* variants, on the other hand, displayed a cleft face, cranial NTDs (exencephaly) and other embryonic closure defects, demonstrating an important role of this gene in fusion of embryonic tissues including neural fold and facial prominences [Pyrgaki et al. 2011]. Extensive studies of mouse knock-outs and knock-ins have demonstrated that *Grhl2* and *Grhl3* provided distinct but cooperative regional signals for rostral and caudal neural tube closure respectively [Rifat et al. 2010]. These data represent *GRHL2* as an excellent candidate for both NTDs and cleft lip and palate disease in humans. Additional molecular and genetic studies of both family members, *GRHL2* and *GRHL3*, in larger cohorts are needed to dissect their role in the pathogenesis of these diseases.

Identification of deleterious variants in *GRHL3* in spina bifida has potential implications for novel preventive strategies. NTDs in *Grhl3* mouse variants (Ct) are resistant to folic acid, whereas they can be

prevented by myo-inositol, suggesting a possible alternative therapy for human folate-resistant NTDs [Greene et al. 2016]. A large NTD consortium is needed for identification of families carrying deleterious variants in *GRHL3* and having a higher risk for NTDs who would offer an ideal cohort for developing alternative inositol-based prevention strategies.

Acknowledgments

We would like to thank all participants in this study. P.L. is awarded by the CHU Sainte Justine Foundation and the Fonds de Recherche du Québec—Santé (FRQS). P.D.M. is funded by Fondazione Gerolamo Gaslini and Trust Aletti. E.M. was funded by Trust Aletti. Z.K. is funded by the Institutes of Health Research and the Réseau de Médecine Génétique Appliquée (FRQS). We are also grateful to the Associazione Spina Bifida Italia (ASBI). The authors have no competing interests.

Figures

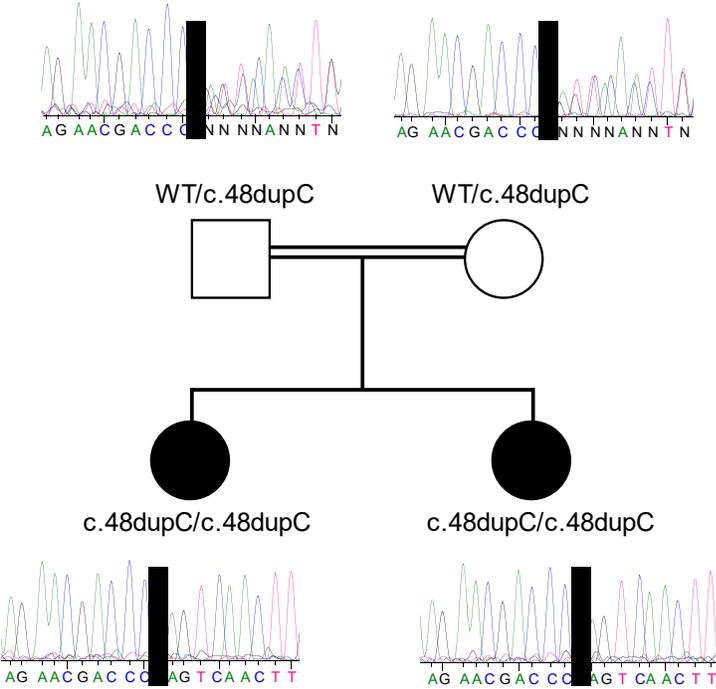


Figure 1: Pedigree of the family segregating c.48dupC (p.Asp16Aspfs*10) variant. Parents were heterozygous for the variant, while both children affected with spina bifida were homozygous. Sequence chromatograms are shown next to each family member.

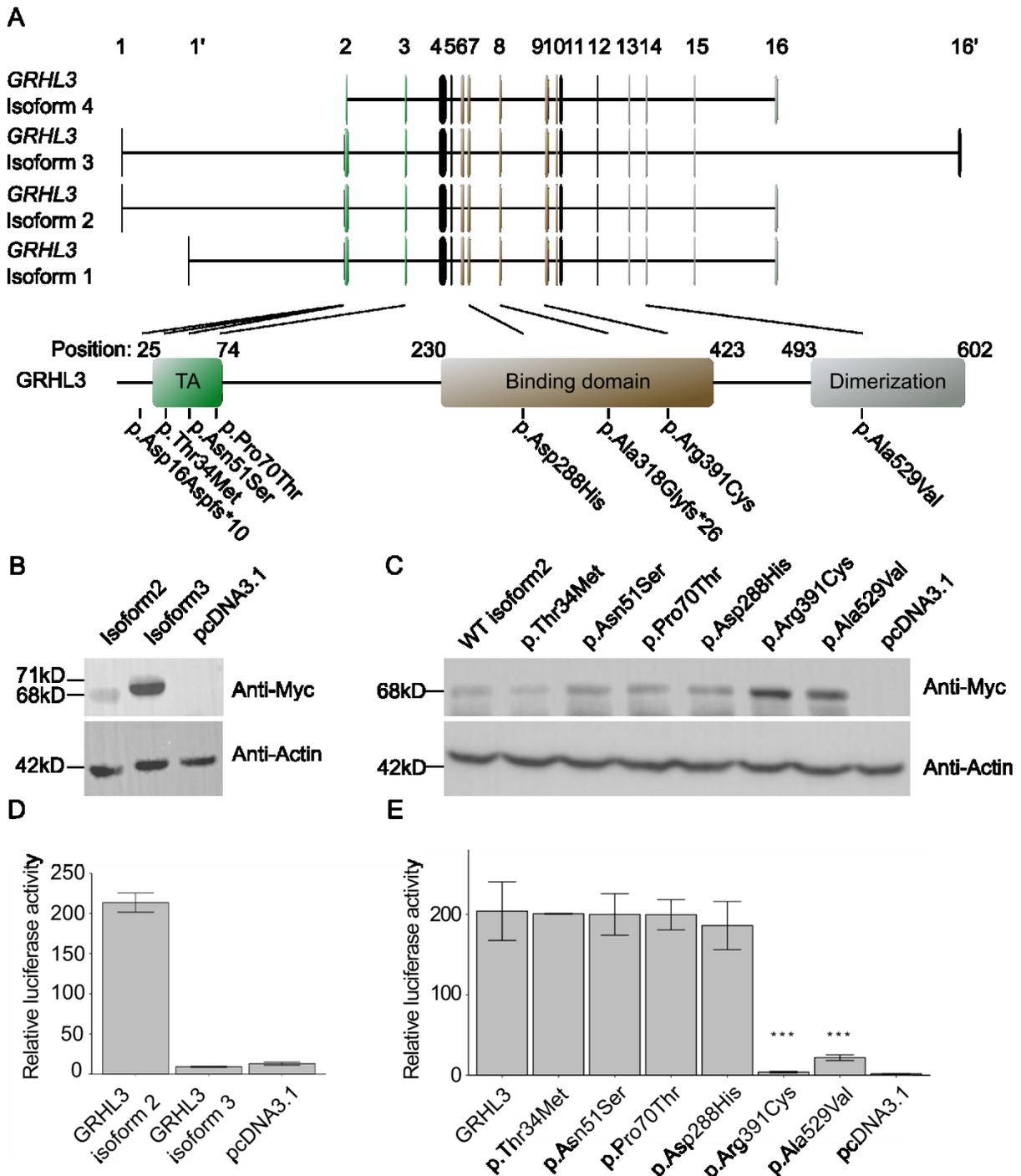


Figure 2: Deleterious rare variants identified in *GRHL3* in spina bifida. (A) A schematic representation of the *GRHL3* gene and protein showing the position of the deleterious variants identified in *GRHL3* in spina bifida. The different isoforms only affect the first and last exons which in turn modify the trans-activation and the dimerization domains found in N and C-terminus of the *GRHL3* protein. Isoform 4

(NM_001195010.1) is the shortest with an alternative start codon in exon 2 and excludes the p.Asp16Aspfs*10 variant. Isoform 3 (NM_198174.2) is the longest using the alternative last exon 16'. Isoform 2 (NM_198173.2) shares exons 1-15 of isoform 3 and the last exon 16 of isoform 4. Isoform 1 (NM_021180.3) has an alternative first exon 1' and shares the exons 2-16 of isoforms 2. TA, trans-activation domain. (B) Western blot analysis of HEK293 cells transfected with either isoform 2 or 3 of *GRHL3*. (C) Western blot analysis of HEK293 transfected with either wild type or mutant *GRHL3* isoform 2. (D) Luciferase assay of the isoforms 2 and 3 of *GRHL3*. Ratios of the luciferase to the renilla are reported. Isoform 3 has no reporter activity in this assay while isoform 2 results in a 17-fold increase in reporter activity. (E) Luciferase assay of the 6 *GRHL3* variants identified in spina bifida and cloned in *GRHL3* isoform 2. Ratios of the luciferase to the renilla are reported. The variants, p.Arg391Cys and p.Ala529Val, cause reduced reporter activity.

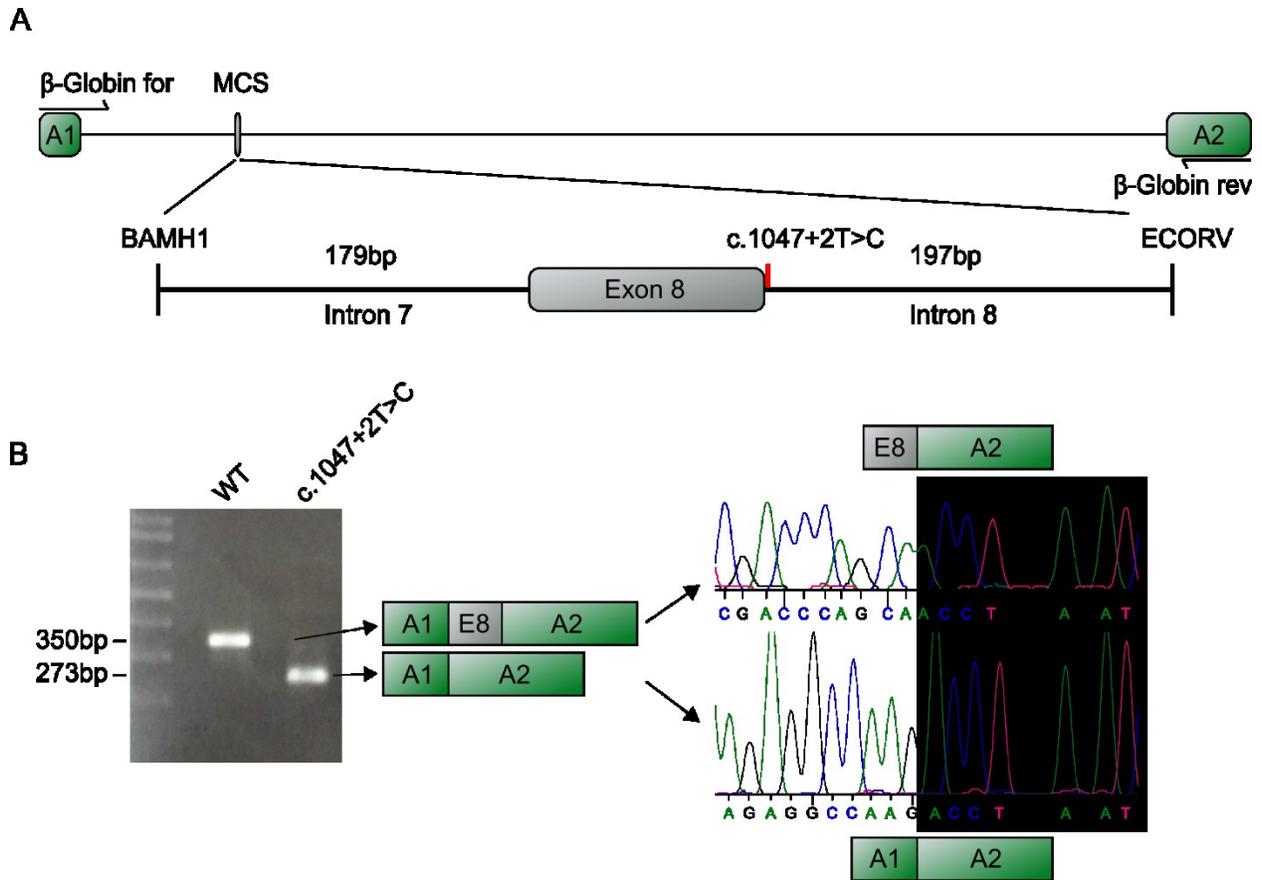


Figure 3: The c.1047+2T>C (p.Ala318Glyfs*26) variant in *GRHL3* in spina bifida caused skipping of exon 8. (A) Exon 8 of *GRHL3* flanked by 179bp of intron 7 and 197 bp of intron 8 were cloned in the exon trapping vector pSPL3. The c.1047+2T>C variant was introduced by mutagenesis. β -Globin primers were used to amplify the final fragment. (B) RT-PCR amplification using β -Globin primers shows a difference between the size of the WT and c.1047+2T>C products. Chromatograms of the RT-PCR products demonstrate skipping of exon 8.

Table 1: Highly deleterious variants identified in *GRHL3* in spina bifida

| Chr | Position | cDNA variant ^a | Amino acid change ^a | Protein domain ^b | ExAC frequency | CADD |
|-----|----------|---------------------------|--------------------------------|-----------------------------|----------------|------|
| 1 | 24657945 | c.48dupC | p.Asp16Aspfs*10 | Pre-TA | 0 | NA |
| 1 | 24657999 | c.101C>T | p.Thr34Met | TA | 6.59E-05 | 20.4 |
| 1 | 24658050 | c.152A>G | p.Asn51Ser | TA | 3.25E-05 | 28.4 |
| 1 | 24661138 | c.208C>A | p.Pro70Thr | TA | 0 | 23.1 |
| 1 | 24664503 | c.862G>C | p.Asp288His | BD | 8.13E-06 | 23.4 |
| 1 | 24666254 | c.1047+2T>C | p.Ala318Glyfs*26 | BD | 0 | 22.2 |
| 1 | 24668728 | c.1171C>T | p.Arg391Cys | BD | 0 | 28.9 |
| 1 | 24674000 | c.1586C>T | p.Ala529Val | DD | 4.07E-05 | 33 |

^aThe numbering systems of the variant at the nucleotide level and the amino acid change are based on *GRHL3* cDNA and protein sequence with the accessions NM_198173.2 and NP_937816.1 respectively. Nucleotide numbering uses +1 as the A of the ATG translation initiation codon in the reference sequence, with the initiation codon as codon 1.

^bAbbreviations: BD, binding domain, DD, dimerization domain; TA, trans-activation.

| | p.Thr34Met | p.Asn51Ser | p.Pro70Thr |
|-----------|---|------------|------------|
| Human | YTSEDEAWKTYLENPLTAATKAMMRVNGDDDSVAALSFLYDYVMGPKKEKRILSSST | | |
| Mouse | YSNEDEAWKTYLENPLTAATKAMMRVNGDEESVAALSFLYDYVMGPKKEKRILSSST | | |
| Chicken | YTNEDEAWKTYLENPLTAATKAMMRVNGDDDSVAALSLLYDYVMVPEKEKRILPSGM | | |
| Xenopus | SPNDDEAWSKYLENPMATAATKAMMRANGDDDGVAALSLLYDYRVPKEKRIITQGS | | |
| Zebrafish | ESYVMDYWTNMDSGNLSKTKPRL---ASDEDLATLNLLYDACKPSKEQKMTS--- | | |

p.Asp288His

| | |
|-----------|-----------------------------------|
| Human | ALSSNKVKSVVMVVFENEKVPVEQLRFWKHWHS |
| Mouse | ALSSSKVKSVVMVVFNDKVPVEQLRFWRHWHS |
| Chicken | HLSSNKVKSVVMIVVFNEKIPTQLKFWKHWS |
| Xenopus | HLTSNKVKSVVMVVFNEKNPEEQLKRWKHWHS |
| Zebrafish | --PCSKVKTVMIMAVFENDKSPMQLKYWNHWA |

p.Arg391Cys

| | |
|-----------|-----------------------------------|
| Human | IDTYDCGLGTERLVHRAVCQIKIFCDKGAERKM |
| Mouse | IDTYDCGAGTERLVHRAVCQIKIFCDKGAERKM |
| Chicken | IDTYDYGNGTNHLVHRAVCQIKIFCDKGAERKM |
| Xenopus | IDTYDFETGVKRLIHRAVCQIKIFCDKGAERKM |
| Zebrafish | IDTYDFSSGNNRLIHRAVCQVKIFCDKGAERKM |

p.Ala529Val

| | |
|-----------|-----------------------------------|
| Human | -VLLYVRRETEEVFDALMLKTPDLKGLRNAISE |
| Mouse | -VLLYVRRETEEVFDALMLKTPDLKGLRNAISE |
| Chicken | -VLLYVRRESEEVFDALMLKTPDLQGLRMAISE |
| Xenopus | -VLLYVRRETEEVFDALMLKTPDLKGLRNAISE |
| Zebrafish | AVLLYVRRESEEVFDALMLNTPNLKGLKEAISE |

Figure S1: Alignments of the *GRHL3* missense variants on other species. Alignments were done using human (NM_198173.2), mouse (NM_001013756.1), chicken (XM_004947906.2), xenopus (NM_001005642.1) and zebrafish (XM_009293109.2).

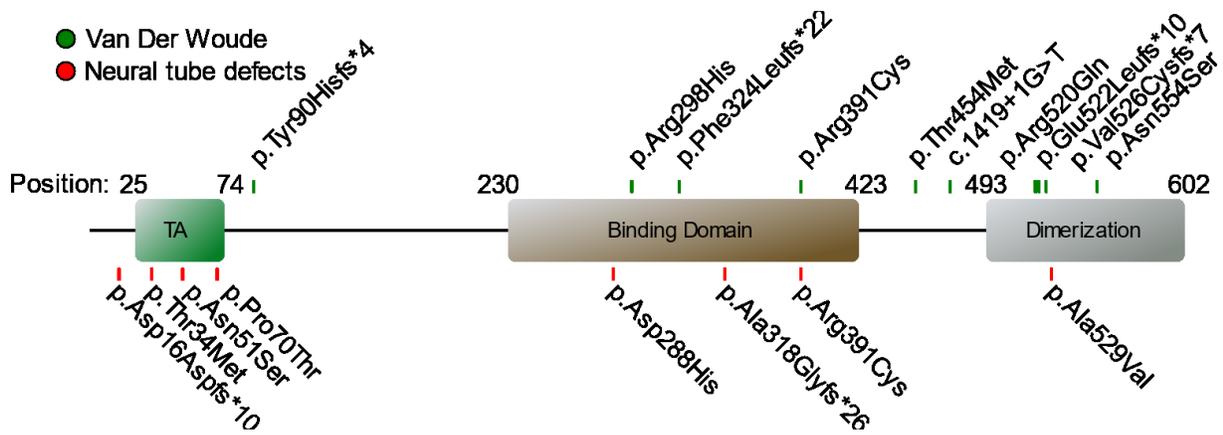


Figure S2: Graphical representation showing previously identified variants associated to Van Der Woude syndrome and nonsyndromic cleft palate on top and spina bifida variants on the bottom. A clustering of highly deleterious variants in the trans-activation domain of the GRHL3 protein is observed in spina bifida while no variant had previously been identified in this region in the Van Der Woude syndrome or in nonsyndromic cleft palate

References

- Alotaibi H, Basilicata MF, Shehwana H, Kosowan T, Schreck I, Braeutigam C, Konu O, Brabletz T, Stemmler MP. 2015. Enhancer cooperativity as a novel mechanism underlying the transcriptional regulation of E-cadherin during mesenchymal to epithelial transition. *Biochim Biophys Acta BBA - Gene Regul Mech* 1849:731–742.
- Bassuk AG, Kibar Z. 2009. Genetic Basis of Neural Tube Defects. *Semin Pediatr Neurol* 16:101–110.
- Caddy J, Wilanowski T, Darido C, Dworkin S, Ting SB, Zhao Q, Rank G, Auden A, Srivastava S, Papenfuss TA, Murdoch JN, Humbert PO, et al. 2010. Epidermal Wound Repair Is Regulated by the Planar Cell Polarity Signaling Pathway. *Dev Cell* 19:138–147.
- Consortium EA, Lek M, Karczewski K, Minikel E, Samocha K, Banks E, Fennell T, O'Donnell-Luria A, Ware J, Hill A, Cummings B, Tukiainen T, et al. 2016. Analysis of protein-coding genetic variation in 60,706 humans. [bioRxiv 030338](https://doi.org/10.1101/030338).
- Copp AJ, Stanier P, Greene ND. 2013. Neural tube defects: recent advances, unsolved questions, and controversies. *Lancet Neurol* 12:799–810.
- Greene NDE, Copp AJ. 2005. Mouse models of neural tube defects: Investigating preventive mechanisms. *Am J Med Genet C Semin Med Genet* 135C:31–41.
- Greene NDE, Leung K-Y, Copp AJ. 2016. Inositol, neural tube closure and the prevention of neural tube defects. *Birt Defects Res A Clin Mol Teratol*. In press
- Gustavsson P, Copp AJ, Greene NDE. 2008. Grainyhead genes and mammalian neural tube closure. *Birt Defects Res A Clin Mol Teratol* 82:728–735.
- Haendeler J, Mlynek A, Büchner N, Lukosz M, Graf M, Guettler C, Jakob S, Farrokh S, Kunze K, Goy C, Guardiola-Serrano F, Schaal H, et al. 2013. Two Isoforms of Sister-of-Mammalian Grainyhead Have Opposing Functions in Endothelial Cells and In Vivo. *Arterioscler Thromb Vasc Biol* 33:1639–1646.
- Harris MJ, Juriloff DM. 2007. Mouse mutants with neural tube closure defects and their role in understanding human neural tube defects. *Birt Defects Res A Clin Mol Teratol* 79:187–210.
- Harris MJ, Juriloff DM. 2010. An update to the list of mouse mutants with neural tube closure defects and advances toward a complete genetic perspective of neural tube closure. *Birt Defects Res A Clin Mol Teratol* 88:653–669.
- Juriloff DM, Harris MJ. 2012. A consideration of the evidence that genetic defects in planar cell polarity contribute to the etiology of human neural tube defects. *Birt Defects Res A Clin Mol Teratol* 94:824–840.

- Kibar Z, Torban E, McDearmid JR, Reynolds A, Berghout J, Mathieu M, Kirillova I, De Marco P, Merello E, Hayes JM, Wallingford JB, Drapeau P, et al. 2007. Mutations in VANGL1 Associated with Neural-Tube Defects. *N Engl J Med* 356:1432–1437.
- Kimura-Yoshida C, Mochida K, Ellwanger K, Niehrs C, Matsuo I. 2015. Fate Specification of Neural Plate Border by Canonical Wnt Signaling and Grhl3 is Crucial for Neural Tube Closure. *EBioMedicine* 2:513–527.
- Kircher M, Witten DM, Jain P, O’Roak BJ, Cooper GM, Shendure J. 2014. A general framework for estimating the relative pathogenicity of human genetic variants. *Nat Genet* 46:310–315.
- Kousa YA, Mansour TA, Seada H, Matoo S, Schutte BC. 2016. Shared molecular networks in orofacial and neural tube development. *Birt Defects Res A Clin Mol Teratol*. In press
- Lemay P, Guyot M-C, Tremblay É, Dionne-Laporte A, Spiegelman D, Henrion É, Diallo O, De Marco P, Merello E, Massicotte C, Désilets V, Michaud JL, et al. 2015. Loss-of-function de novo mutations play an important role in severe human neural tube defects. *J Med Genet* 52:493–497.
- Leslie EJ, Liu H, Carlson JC, Shaffer JR, Feingold E, Wehby G, Laurie CA, Jain D, Laurie CC, Doheny KF, McHenry T, Resick J, et al. 2016. A Genome-wide Association Study of Nonsyndromic Cleft Palate Identifies an Etiologic Missense Variant in GRHL3. *Am J Hum Genet* 98:744–754.
- Li MZ, Elledge SJ. 2012. SLIC: a method for sequence- and ligation-independent cloning. *Methods Mol Biol Clifton NJ* 852:51–59.
- Mangold E, Böhmer AC, Ishorst N, Hoebel A-K, Gültepe P, Schuenke H, Klamt J, Hofmann A, Gözl L, Raff R, Tessmann P, Nowak S, et al. 2016. Sequencing the GRHL3 Coding Region Reveals Rare Truncating Mutations and a Common Susceptibility Variant for Nonsyndromic Cleft Palate. *Am J Hum Genet* 98:755–762.
- Monsoro-Burq AH. 2015. PAX transcription factors in neural crest development. *Semin Cell Dev Biol* 44:87–96.
- MRC Vitamin Study Research Group. 1991. Prevention of neural tube defects: Results of the Medical Research Council Vitamin Study. *The Lancet* 338:131–137.
- Neumann PE, Frankel WN, Letts VA, Coffin JM, Copp AJ, Bernfield M. 1994. Multifactorial inheritance of neural tube defects: localization of the major gene and recognition of modifiers in ct mutant mice. *Nat Genet* 6:357–362.
- Nisson PE, Ally A, Watkins PC. 1994. Protocols for trapping internal and 3’-terminal exons. *Genome Res* 4:S24–S39.
- O’Roak BJ, Vives L, Fu W, Egertson JD, Stanaway IB, Phelps IG, Carvill G, Kumar A, Lee C, Ankenman K, Munson J, Hiatt JB, et al. 2012. Multiplex targeted sequencing identifies recurrently mutated genes in autism spectrum disorders. *Science* 338:1619–1622.

Peyrard-Janvid M, Leslie EJ, Kousa YA, Smith TL, Dunnwald M, Magnusson M, Lentz BA, Unneberg P, Fransson I, Koillinen HK, Rautio J, Pegelow M, et al. 2014. Dominant Mutations in GRHL3 Cause Van der Woude Syndrome and Disrupt Oral Periderm Development. *Am J Hum Genet* 94:23–32.

Pyrgaki C, Liu A, Niswander L. 2011. Grainyhead-like 2 regulates neural tube closure and adhesion molecule expression during neural fold fusion. *Dev Biol* 353:38–49.

Rampersaud E, Bassuk AG, Enterline DS, George TM, Siegel DG, Melvin EC, Aben J, Allen J, Aylsworth A, Brei T, Bodurtha J, Buran C, et al. 2005. Whole genomewide linkage screen for neural tube defects reveals regions of interest on chromosomes 7 and 10. *J Med Genet* 42:940–946.

Rifat Y, Parekh V, Wilanowski T, Hislop NR, Auden A, Ting SB, Cunningham JM, Jane SM. 2010. Regional neural tube closure defined by the Grainy head-like transcription factors. *Dev Biol* 345:237–245.

Stamm DS, Siegel DG, Mehlretter L, Connelly JJ, Trott A, Ellis N, Zismann V, Stephan DA, George TM, Vekemans M, Ashley-Koch A, Gilbert JR, et al. 2008. Refinement of 2q and 7p loci in a large multiplex NTD family. *Birt Defects Res A Clin Mol Teratol* 82:441–452.

Ting SB, Wilanowski T, Auden A, Hall M, Voss AK, Thomas T, Parekh V, Cunningham JM, Jane SM. 2003. Inositol- and folate-resistant neural tube defects in mice lacking the epithelial-specific factor Grhl-3. *Nat Med* 9:1513–1519.

Wilanowski T, Caddy J, Ting SB, Hislop NR, Cerruti L, Auden A, Zhao L-L, Asquith S, Ellis S, Sinclair R, Cunningham JM, Jane SM. 2008. Perturbed desmosomal cadherin expression in grainy head-like 1-null mice. *EMBO J* 27:886–897.

Zhang T, Lou J, Zhong R, Wu J, Zou L, Sun Y, Lu X, Liu L, Miao X, Xiong G. 2013. Genetic Variants in the Folate Pathway and the Risk of Neural Tube Defects: A Meta-Analysis of the Published Literature. *PLOS ONE* 8:e59570.

6. Discussion

6.1 Critiques et limites des résultats

Les résultats présentés dans les sections 4 et 5 démontrent une évolution certaines des connaissances associées à la génétique des ATN. La limite de certaines des techniques et des conclusions de ces résultats mérite néanmoins d'être discuté. La section suivante s'attarde sur quelques éléments qui méritent des clarifications supplémentaires.

6.1.1 Détection de mutations mosaïques

Un élément permettant de mieux comprendre l'impact de certaines des mutations identifiées est le mosaïsme génétique. En effet, des mutations présentes dans un nombre limité de tissus permettent souvent de conclure à un effet pathologique moins élevé et doit donc être exploré. Cet élément est malheureusement absent des analyses présentées à la section 4 et sera analysé ici. Des barèmes d'investigations établis sur le séquençage du génome entier permet d'identifier de potentielles mutations mosaïques à partir des données de séquençage de l'exome. Ces barèmes suggèrent qu'une mutation avec une couverture minimale de 80X et un ratio allélique inférieur à 32.8% est probablement mosaïque³⁶³. Six des mutations présentées dans la section 4 passent ce barème. Il s'agit des mutations c.1604C>T sur *LRFN1*, c.1441C>T sur *C17orf80*, c.535A>C sur *STK31*, c.355C>T sur *IP6K1*, c.1605A>C sur *DENND6A* et c.961G>A sur *U2SURP*. Afin de conclure définitivement au mosaïsme, il serait nécessaire de séquencer ces mutations dans plusieurs tissus ce qui sera difficile étant donné que plusieurs des patients affectés dans cette cohorte sont décédés. Un mosaïsme démontré permettrait d'informer de potentielles études subséquentes sur la pathogénicité de ces mutations. Cependant, il est important de noter que le séquençage de l'exome est reconnu pour avoir un biais de capture plus élevé que le séquençage du génome³⁷⁸. Ainsi, les critères ayant été établi sur le séquençage du génome ne sont peut-être pas appropriés chez le séquençage de l'exome supportant la nécessité pour une étude plus approfondie de ces mutations. Les principales mutations candidates (*SHROOM3*, *PAX3*, *GRHL3*) suivent un ratio allélique suggérant une transmission germinale ce qui limite l'intérêt pour une validation supplémentaire du mosaïsme dans cette cohorte.

6.1.2 Limites de l'essai luciférase

L'essai luciférase utilisé dans la section 5 permet de démontrer l'effet de nouvelles mutations sur la capacité de *GRHL3* de lier et d'activer ses cibles, mais quelques éléments supplémentaires doivent être discutés.

Premièrement, une meilleure analyse de la séquence consensus permettant la liaison de *GRHL3* à l'ADN est nécessaire. En effet, la séquence consensus 5'- AACCGGTT- 3' est reconnue pour permettre la liaison de *GRHL3* à l'ADN³¹⁷. Cette séquence consensus est aussi reconnue comme le principal site de liaison des autres membres de la famille des *GRHL*³³⁵. L'essai luciférase utilisé dans la section 5 contient 4 répétitions de cette séquence consensus suivi du gène *FLuc*. L'essai a initialement été testé sur l'isoforme 4 qui a démontré une capacité à l'activer spécifiquement. Cet isoforme possède la même séquence 3' que l'isoforme 2 décrit dans la section 5, mais diffère en 5'. L'isoforme 2 a lui aussi démontré son habilité à activer spécifiquement l'essai luciférase tel que démontré dans la section 5. L'essai demeure pour l'instant non testé sur l'isoforme 1 et n'a pas démontré d'association à l'isoforme 3. Cependant, étant donné le haut niveau de conservation de la séquence consensus à travers toute la famille *GRHL*, il est possible de suggérer que ces isoformes reconnaîtront aussi la séquence. Ceci permettrait donc d'utiliser l'essai luciférase dans l'étude des différents isoformes de *GRHL3*.

L'essai luciférase est cependant limité au niveau des conclusions fonctionnelles. En effet, les conclusions offertes par l'essai luciférase se limitent à analyser la capacité d'interaction du gène à la séquence consensus et son habilité à recruter la machinerie de transcription. Malgré le fait que cette différenciation n'a pas été faite dans le contexte de la section 5, il aurait été possible d'analyser la fonction affectée par la mutation. Pour ce faire, il faudrait co-transfecter en quantité égale un *GRHL3* natif et un *GRHL3* affecté par une mutation réduisant son activité. En présence d'une mutation affectant la liaison, l'activité luciférase sera égale au contrôle positif composé d'une double transfection de *GRHL3* natif et d'un plasmide contrôle. Cependant, si c'est la fonction de recrutement qui est affectée, le *GRHL3* affecté agira comme un inhibiteur compétitif de la liaison de *GRHL3* natif diminuant de moitié l'activité. Considérant les études de calibration faites à l'interne sur cet essai, la sensibilité de l'essai permettrait de détecter une telle diminution. L'essai est principalement limité quant aux conclusions sur l'impact cellulaire des mutations. En effet, une diminution d'activité tel qu'observée avec la mutation p.Ala529Val est peu informative sur les conséquences biologiques et son implication dans les ATN. Des essais plus spécifiques aux principaux processus impliqués dans les ATN (figure 11) sont nécessaires. Un essai visant à déterminer la capacité d'induire la TEM et la TME par transfection d'une version native ou mutée de *GRHL3* a déjà été développé⁹⁷. Des études *in cellulo* analysant les changements morphologiques du cytosquelette d'actine par immunofluorescence suite à la transfection de *GRHL3* natif ou muté permettrait de mieux comprendre le processus de *GRHL3* impliqué dans les ATN. Ces études permettraient aussi d'analyser l'interaction complexe entre les isoformes. Pour ces études, des cellules spécifiques réactives à *GRHL3* et pouvant générer une constriction apicale seraient nécessaires. La plupart

de ces cellules se retrouvent sous forme de lignée primaire et une certaine expertise dans la culture et la transfection de ces cellules est nécessaire. Bien entendu, l'introduction des mutations identifiées dans un modèle murin serait la plus forte preuve d'association aux ATN, mais cette technique est longue et coûteuse. Un bon compromis entre les techniques *in cellulo* et le modèle murin serait l'utilisation du modèle de *Xenopus* qui est reconnu pour développer des ATN et chez lequel *GRHL3* est aussi impliqué.

6.1.3 Analyses des bases de données publiques

L'utilisation de bases de données publiques est un outil puissant pour l'identification d'enrichissements de mutations et pour l'évaluation de la pathogénicité. Cependant, ces bases de données peuvent entraîner plusieurs biais d'analyse qui seront discutés. Dans les résultats de la section 4 et 5, les bases de données EVS et ExAC sont utilisées, mais seront discutées comme un tout dans cette section étant donné l'inclusion d'une large partie des données d'EVS dans ExAC.

Les bases de données publiques incluent une large quantité de patients (6500 pour EVS, 60706 pour ExAC) qui proviennent d'études diverses ce qui augmente le risque d'inclusion de patients affectés par un ATN fermé. Considérant que près de 50% des ATN fermés ne possèdent pas de signes visibles et qu'une large partie demeurent asymptomatiques, il est très probable qu'une fraction des patients sont affectés par cette pathologie. Ce problème est accentué par le fait que plusieurs patients inclus dans ces bases de données n'ont pas consenti à partager leurs données cliniques, limitant ainsi la capacité à les identifier. La similarité biologique entre les ATN fermés et ouverts suggèrent une étiologie commune suggérant que certains variants de ces bases de données pourraient être pathogéniques. Ceci a comme impact de limiter l'habilité d'évaluer la pathogénicité d'un variant en fonction des bases de données publiques. Cet impact est néanmoins minimal étant donné qu'un calcul rapide utilisant la prévalence moyenne des ATN de 1-2 :1000 naissance suggère qu'au maximum 60 à 120 patients potentiels seraient présents dans ExAC. Ces mutations seraient ensuite réparties entre plusieurs gènes responsables des ATN réduisant encore leur risque d'affecter notre analyse. Un autre risque associé à ces bases de données est la présence de faux positifs. Étant donné l'utilisation de différentes technologies de séquençage et des filtres de qualité différents d'une étude à l'autre, les mutations incluses dans les bases de données ne sont pas uniformes. De plus, l'absence de validation par Sanger augmente le risque de faux positifs associé à ces variants. Les disparités au niveau des filtres et des technologies utilisés limitent la capacité à identifier de potentielles mutations mosaïques. Ainsi plusieurs mutations présentes dans ces bases de données sont potentiellement des artefacts ou mosaïques ce qui affecte notre habilité à évaluer la pathogénicité des variants identifiés dans notre cohorte. Malgré ces limites, ces bases de données demeurent le meilleur

outil de comparaison et d'analyse permettant de mieux comprendre la complexité génétique d'une maladie. Monétairement, ces bases de données permettent aussi d'attribuer les fonds de recherche pour une plus large cohorte d'individus affectés au lieu de séquencer des contrôles. Finalement, la présence de faux positifs augmente artificiellement le seuil statistique pour identifier des enrichissements de mutations dans certains gènes. Cependant, ce seuil plus élevé n'affecte pas les résultats présentés à la section 4 et 5 étant donné que les critères de filtrations et de validation de nos variants sont plus restrictifs que ceux de ces bases de données.

6.2 Analyse des gènes candidats *SHROOM3* et *GRHL3*

Tel que discuté dans les sections 4 et 5 deux principaux candidats, les gènes *SHROOM3* et *GRHL3* ont été statistiquement associés aux ATN. À travers leur association aux ATN chez la souris, ces gènes étaient d'excellent candidats, mais leurs associations à d'autres pathologies suggère une complexité dans leur génétique. Dans cette section, nous discuterons des rôles de ces gènes dans leurs différentes pathologies et analyseront plus en détails les mécanismes à travers lesquels ils peuvent entraîner ces différents phénotypes.

6.2.1 SHROOM3

SHROOM3 est un excellent candidat dans le cadre des ATN et nos résultats démontrent pour la première fois une association à la pathologie humaine. Afin d'identifier ces mutations des critères de filtration très permissif ont été appliqués de manière à minimiser les faux négatifs. Ceci entraîna la validation de seulement 60% des 70 mutations initialement identifiées. Tel qu'observé dans la littérature, ce taux de validation est particulièrement bas³⁵⁸, mais en considérant la permissivité des critères de filtration, il est acceptable. L'identification de deux variants *de novo* tronquant la protéine est un phénomène rare et le fait qu'il soit observé dans une aussi petite cohorte est encore plus surprenant. Cependant, l'association de ce gène à diverses pathologies et l'existence de variants tronquant chez des contrôles suggèrent un mécanisme pathologique complexe.

6.2.1.1 Mécanismes potentiels impliquant *SHROOM3* dans différentes pathologies

L'implication de *SHROOM3* dans les FLP, la microsomie et l'hétérotaxie demeure faiblement supporté par la littérature^{285 286 293}. Néanmoins, la validation de ces associations permettrait d'approfondir les connaissances sur le fonctionnement de cette protéine. Cette section s'attardera à critiquer ces associations et déterminer les conclusions qu'elles généreraient.

L'association avec les FLP dérive d'une étude d'association du génome entier ayant identifié un SNP approchant le seuil significatif dans le gène *SHROOM3*. L'absence de réplification de cette association et le devis d'étude ne permettent de conclure à un lien direct. Ainsi, cette association est statistiquement faible, mais biologiquement forte. En effet, la co-ségrégation des malformations cranio-faciales et des ATN, ainsi que leur lien commun à l'acide folique suggèrent une étiologie commune^{379 380}. Plusieurs mécanismes rappellent aussi la fermeture du tube neural. La fusion palatale est régulée par les éphrines, les gènes *GRHL* et leur régulateur *IRF6*. Ils impliquent aussi la TEM et la signalisation WNT. *SHROOM3* pourrait être impliqué dans les FLP à travers la signalisation WNT et le gène *CDC42*. *CDC42* est un régulateur du cycle cellulaire qui est activé par *WNT5A*³⁸¹ et contribue avec PCP pour assurer la polarité cellulaire³⁸². Une ablation de *CDC42* dans les cellules de la crête neurale a précédemment été associée à des FLP³⁸³. Cette protéine régule la localisation cellulaire de *SHROOM3*³⁰⁴. Cette localisation est nécessaire à un remaniement du cytosquelette impliqué dans la fermeture des FLP. Ainsi, une insulte aux cellules de la crête neurale ou au remaniement du cytosquelette pourrait relier *SHROOM3* aux FLP.

La microsomie cranio-faciale, définie par un sous-développement des structures cranio-faciales, a été associée à une région incluant *SHROOM3* par une étude d'association du génome entier²⁸⁶. Un séquençage du génome chez une cohorte de 21 patients a aussi identifié un variant faux-sens hautement délétère²⁸⁶. Cette pathologie est associée aux cellules de la crête neurale et entraîne un large spectre d'anomalies cranio-faciales suggérant que *SHROOM3* pourrait agir à travers un mécanisme similaire aux FLP²⁸⁶. De plus, les souris *Shroom3*^{-/-} montrent des anomalies cranio-faciales avec une pénétrance élevée renforçant le lien avec la microsomie.

L'hétérotaxie est une maladie causant une divergence du positionnement des organes lors du développement³⁸⁴. *SHROOM3* y est associé à travers cinq variants faux-sens dans une cohorte de 97 patients. Cependant, l'absence de statistiques minimise l'impact de ces résultats qui ne sont pas significativement différents de la fréquence de variation attendue pour ce gène. Cette maladie est associée à une perte de l'asymétrie gauche-droite et les auteurs suggèrent que *SHROOM3* y serait lié à travers son lien à *PITX2*²⁹³. Ceci est supporté par l'expression de *SHROOM3* dans plusieurs organes et par son rôle dans la régulation de *Pitx1*, un facteur affectant l'asymétrie gauche-droite à travers une modulation de *Pitx2*³⁸⁵. De plus, l'observation du phénotype d'ATN chez la souris *Shroom3*^{-/-} inclut une déformation du tube neural dans l'axe gauche-droite. Cependant, aucune hétérotaxie n'est reportée chez ces souris²⁵³.

Les analyses présentées suggèrent l'implication de *SHROOM3* dans plusieurs malformations non-syndromiques. Cependant, une analyse critique des résultats démontre des liens faibles. L'association aux ATN démontrée dans la section 4 offre le plus fort lien de causalité entre ce gène et une pathologie. Cependant, le lien biologique fort de ce gène avec ces pathologies supporte une variabilité du phénotype. Deux principales hypothèses peuvent expliquer cette variation : Le type de variant et la présence de variants modificateurs. Dans le cadre des ATN, les variants p.Y392X et p.L948fs sont tronquants et peuvent entraîner une perte de fonction ou un dominant négatif tel que démontré chez le *Xenopus*³⁸⁶. Dans le cas de l'hétérotaxie et de la microsomie, les variants sont faux-sens et spécifiques à l'isoforme long (*Shrml*) de *SHROOM3*. Les variants faux-sens peuvent entraîner un gain de fonction engendrant un phénotype différent des variants tronquants. De plus, un seul des isoformes de la souris est touché justifiant un phénotype différent. Cependant, à ce jour, un seul isoforme de *SHROOM3* a été identifié chez l'humain limitant cette possibilité. Ces variants faux-sens se retrouvent aussi dans le domaine PDZ dont la fonction demeure inconnue. Il est possible que ce domaine régule une interaction différente qui impliquerait le gène *SHROOM3* dans ces pathologies. Aucun variant n'a encore été identifié pour la FLP, mais leur emplacement pourrait aussi entraîner une variation dans le phénotype. Cependant, la co-ségrégation des phénotypes de malformations oro-faciales et d'ATN chez la souris *Shroom3*^{-/-} suggère que les variants tronquants agissent dans le contexte des deux pathologies. Les variants modificateurs pourraient alors être en cause³⁸⁷. En effet, la présence de patients et de modèles avec une combinaison de ces malformations^{253 388 389} suggère une modulation du phénotype par d'autres variants ou événements.

Au niveau de l'étiologie de ces différentes pathologies, le gène *SHROOM3* peut être impliqué à travers deux hypothèses : l'implication dans un processus commun ou dans différents processus. Pour le processus commun, l'hypothèse serait que *SHROOM3* modifie un processus clef du développement qui en aval modifiera d'autres processus entraînant l'apparition de différents phénotypes. Le développement humain suggérerait que le premier processus affecté serait la formation du tube neural. À travers la fonction de *SHROOM3* dans la constriction apicale (CA), le tube neural ne fusionne pas et les cellules de la crête neurale, importantes dans le contexte de la FLP et de la macrosomie, ne sont pas générées. Ces cellules sont aussi impliquées dans la formation de plusieurs organes offrant un lien à l'hétérotaxie. Cependant, la pénétrance incomplète des malformations oro-faciales chez la souris et l'absence de ces phénotypes chez les patients identifiés dans la section 4 supportent une modulation plus indépendante. Il est possible que des processus avec une régulation distincte de la CA soient responsables de ces pathologies. En effet, les cellules de la crête neurale sont partiellement générées avant la fermeture du tube neural, suggérant que la CA n'est pas la cause des malformations oro-faciales⁷³. L'implication de

Shroom3 dans les remaniements du cytosquelette en fait un candidat intéressant dans le contexte de la TEM, un processus clef dans la délamination des cellules de la crête neurale, dans la formation de l'épithélium oral³⁹⁰ et dans la fusion des FLP³⁹¹. Ceci expliquerait l'implication de *SHROOM3* dans la FLP et la microsomie. Une voie indépendante permettrait aussi de mieux justifier le lien avec l'hétérotaxie et *PITX2* dont l'importance dans le tube neural demeure mitigée.

L'implication d'un gène dans plusieurs pathologies suggère une régulation complexe et/ou une implication dans un large éventail de fonctions biologiques. L'investigation de ces associations suggèrent une régulation complexe de *SHROOM3*. Son expression dans plusieurs tissus et son implication dans la réorganisation du cytosquelette en fait un excellent candidat dans ces pathologies. Malgré le lien fort de *SHROOM3* aux ATN identifié dans la section 4, des études génétiques supplémentaires sont nécessaires afin d'associer ce gène aux autres pathologies. Une connaissance approfondie des phénotypes associés à ce gène aidera le développement d'études sur son mécanisme d'action.

6.2.1.2 Les données publiques informent les hypothèses sur le mode d'action des variants

L'utilisation de données publiques permet aussi d'approfondir les conclusions quant à la pathogénicité des variants identifiés sur *SHROOM3*. Dans la section 4, cette analyse est effectuée avec la base de données EVS³⁵⁶, mais de meilleures données sont maintenant disponibles. La base de données ExAC³⁵⁵, provenant du séquençage de l'exome de 67 000 individus, est la plus large source de données génétiques. L'analyse du gène *SHROOM3* identifie 13 variants tronquants provenant de 15 individus. Cinq de ces variants précèdent p.Y392X, suivi de 2 entre p.Y392X et p.L948fs et 6 après p.L948fs. Les études *de novo* permettent d'identifier des variants délétères et causatifs, mais les variants tronquants chez des individus sains suggèrent autrement. En effet, si p.Y392X et p.L948fs créent des fragments dominants, des variants entre ces derniers auraient une haute probabilité de créer des fragments dominants. Ceci corrèle avec le modèle murin où les souris *Shroom3*^{+/-} n'ont pas de phénotype²⁵³. Ceci va à l'encontre de l'hypothèse proposée à la section 4 qui suggère que ces deux variants agissent en tant que dominant négatifs³⁸⁶. Cette discordance s'explique par le fait que l'hypothèse provient de l'analyse des données EVS qui ne répertorient aucun variant entre p.Y392X et p.L948fs. Ainsi, quoi qu'il soit possible que l'un de ces variants agissent à travers ce mécanisme, il est peu probable que les deux aient cet effet. Il faut cependant prendre en considération le risque de faux positifs de ces bases de données. En effet, tel que décrit dans la section 6.1.3, plusieurs mutations provenant de ces bases sont des faux positifs, sont mosaïques ou tout simplement associés à un individu avec une ATN fermée. Seulement 2 mutations ont été identifiées entre p.Y392X et p.L948fs et si ces dernières sont des faux positifs, il est possible que l'hypothèse du dominant

négligé soit possible. Cependant, ces faux positifs sont relativement rares suggérant que d'autres hypothèses pourraient être plus appropriées. Dans la section 4, l'hypothèse de la double insulte est aussi soulevée³⁹². Elle suggère qu'un variant hautement pathogénique va créer un phénotype qui est ensuite modulé par un deuxième événement qui peut être d'origine génétique, épigénétique ou environnemental³⁹². Cette hypothèse justifie pourquoi deux variants tronquants entraînent des phénotypes différents (anencéphalie et myeloméningocèle). Selon celle-ci, les 15 individus contrôles avec des variants tronquants dans *SHROOM3* n'auraient pas été soumis à la deuxième insulte nécessaire au développement du phénotype. Ce mécanisme peut aussi soutenir la modulation du phénotype de *SHROOM3* dans les ATN, les FLP, la microsomie et l'hétérotaxie. Cependant, le mécanisme tel que suggéré par Kumar et al, 2010³⁹² suggère que chaque insulte est associée à un phénotype et que ce mécanisme est plus rare chez les individus affectés par un variant *de novo*. La version suggérée ici supporte plutôt un mécanisme synergétique entre les variants avec une insulte principale transformée en pathologie à travers plusieurs insultes secondaires³⁹³. Ceci suggère que les individus affectés ont un contexte génétique qui prédisposent au développement des ATN.

En conclusion, le mécanisme pathologique à travers lequel les variants de *SHROOM3* induisent des ATN nécessite d'être approfondi. L'étude des données publiques permet de conclure que la position des mutations n'est pas un facteur majeur dans l'apparition du phénotype d'ATN et que la double insulte est un mécanisme pathogénique plus probable. Ceci corrèle avec la discussion de la section 6.2.1.1 qui suggère que la position des variants dans le gène n'était pas suffisant pour justifier la modulation du phénotype. Les études de *SHROOM3* dans le cadre des ATN devraient donc porter une attention particulière aux méthodes génétiques, épigénétiques et environnementales affectant la fonction de ce gène.

6.2.2 GRHL3

Par son association aux ATN chez la souris, *GRHL3* est un excellent gène candidat. L'identification d'un variant homozygote tronquant chez deux patientes affectées par un myeloméningocèle renforce fortement cette candidature dans la pathologie humaine. L'enrichissement significatif de variants hautement délétères, combiné avec l'identification d'un deuxième variant tronquant et la démonstration d'une perte de fonction chez un variant faux-sens démontre l'importance de ce gène. L'association de ce gène au SVW2 et aux FLP suggère un large spectre de fonction. La présence d'un isoforme non-actif tel que décrit dans la section 5 suggère aussi une régulation complexe de ce gène qui nécessite une investigation plus approfondie. Finalement, l'implication de ce gène à plusieurs étapes de la neurulation

en fait un candidat très intéressant, mais suggère potentiellement un mécanisme pathologique complexe pour les variants identifiés.

6.2.2.1 *GRHL3* est impliqué dans le développement oro-facial et dans la neurulation

Contrairement à l'association entre *SHROOM3* et FLP, les démonstrations liant *GRHL3* et FLP sont particulièrement fortes. Des enrichissements significatifs de variants suivis de validations fonctionnelles démontrant une diminution de l'activité du gène *GRHL3* sont observés chez des patients affectés par le SVW2 et des FLP^{313 314}. Son lien aux ATN est lui aussi fortement démontré par l'étude de la souris *Grhl3*^{-/-} et par son rôle dans la formation de l'ectoderme de surface^{73 309}. Cependant, ce sont les résultats présentés à la section 5 qui l'impliquent dans la pathologie humaine³⁷⁴. Ces observations combinées à une étude récente semble pointer vers un lien entre ces deux pathologies³⁹⁴. Les résultats obtenus supportent ce lien à travers l'identification du variant *de novo* p.Arg391Cys. Tel que discuté, ce variant entraîne une perte de fonction de *GRHL3* dans le SVW2 et dans les ATN en faisant un excellent candidat pour le mécanisme de la double insulte discuté dans la section 6.2.1.2³⁹². En effet, il est probable que ce variant entraîne un phénotype et que la présence d'un deuxième événement soit nécessaire afin de moduler ce phénotype vers le SVW2 ou les ATN. Ceci est supporté par une des patientes avec p.Asp16Aspfs*10, le variant homozygote tronquant ségrégant chez deux sœurs décrit dans la section 5. En effet, l'une des sœurs est aussi affecté par une FLP, alors qu'aucun autre membre de la famille incluant la sœur homozygote et les parents hétérozygotes n'ont ce phénotype. L'association de ce variant aux ATN peut être justifié par le fait que la majorité des variants associés au SVW2 agissent en dominants négatifs, mais le variant homozygote agit comme un ablation complète du gène³¹³. Cependant, le type et la position du variant n'expliquent pas la divergence entre les deux sœurs et ne justifient pas la modulation du phénotype pour p.Arg391Cys supportant l'hypothèse qu'un événement supplémentaire (potentiellement *de novo*) agirait sur la modulation du phénotype. La nature de cet événement et la méthode par laquelle elle module le phénotype est quant à elle incertaine, mais sous-tend l'idée qu'un élément modulateur entre SVW2 et ATN est possible.

La présence du variant p.Arg391Cys dans les deux pathologies suggère que les processus pathologiques se superposent partiellement et qu'une deuxième insulte est nécessaire à la modulation du phénotype. Ainsi, une voie de signalisation commune à ces deux pathologies serait une approche intéressante pour l'identification de la deuxième insulte. Les trois processus candidats décrits dans la figure 11 sont potentiellement impliqués. Tout d'abord, la voie PCP est impliquée dans le développement oro-facial par les cellules de la crête neurale³⁸³ et la fermeture du palais³⁹⁵. Les cellules de la crête neurale

sont créées durant la fermeture du tube neurale et un échec de formation de ces cellules entraîne une ATN^{396 397}. La migration générée par la TEM permet ensuite de créer une partie du cartilage et des os impliqués dans la morphologie faciale³⁸³. Grhl3 active Rhoa qui en collaboration avec syndecan 4 (Sdc4) inhibe Rac. Ceci permet d'inhiber les protrusions cellulaires et de moduler la migration ces cellules de la crête neurale. Cependant, seul les ATN crâniens dépendent des cellules de la crête neurale et même ces évidences semblent circonstancielles³⁹⁸. La voie PCP a aussi été suggérée comme étant impliquée dans la formation du palais par la souris *Fzd2*^{-/-}, un membre de la voie PCP³⁹⁹. Cette dernière développe une FLP, mais ce phénotype ne se retrouve pas chez d'autres membres de la voie limitant la candidature de ce processus comme lien des ATN et des malformations oro-faciales²¹². Le deuxième processus est associé à Shh (Figure 11). Ce dernier est impliqué dans la formation des CDL et est régulé par *GRHL3*. Il est aussi impliqué dans les malformations oro-faciales à travers la formation des cellules de la crête neurale. Au niveau de la fusion du palais, la croissance du mésenchyme oral est assurée par l'expression de Shh⁴⁰⁰. Une dérégulation de l'expression de Shh en absence de Grhl3 débalancerait les ratios nécessaires à la croissance normale du palais induisant une malformation. Finalement, les cellules de la crête neurale dérivent du maintien de l'interaction entre l'ectoderme de surface et le neuroectoderme lors de la fermeture du tube neural. Une perte de cette interaction causée par un échec de fermeture ou une perte de la lignée d'ectoderme de surface inhiberait la formation des cellules de la crête neurale et affecterait le développement oro-facial⁷³. Les cellules de la crête neurale sont générées en absence de fusion du tube neural, permettant d'obtenir un phénotype sans l'autre.

Il est aussi possible que des voies de signalisation différentes sont impliquées dans les différents phénotypes et que la deuxième insulte agirait à l'intérieur de ces voies. Il est maintenant accepté que les cibles activées par *GRHL3* semblent dépendantes de co-facteurs. En effet, en fonction du type cellulaire, *GRHL3* est incapable d'activer certaines de ses cibles suggérant que des facteurs intrinsèques aux types cellulaires sont nécessaires à moduler sa spécificité^{97 401}. Cette hypothèse permettrait d'expliquer comment les malformations oro-faciales et les ATN apparaissent de manières indépendantes en présence de variants sur le même gène. Il est aussi possible que certaines de ces deuxième insultes soient exprimées par *GRHL3* dans plusieurs tissus, permettant d'expliquer les patients combinant les deux phénotypes tel qu'observé dans la famille arborant le variant p.Asp16Aspfs*10.

Tel que discuté dans la section 5, la variabilité phénotypique pourrait aussi être partiellement justifiée par le domaine affecté par les variants. En effet, le fort enrichissement de variants délétères dans le domaine de trans-activation du gène *GRHL3* semble spécifique à notre cohorte d'ATN et absente des

bases de données publiques et des variants identifiés dans les malformations oro-faciales. Ce domaine est impliqué dans le recrutement de la machinerie de transcription, donc une mutation fonctionnelle n'affecterait pas la liaison de la protéine à ses cibles. Incapable d'activer la transcription, ce GRHL3 muté resterait lié indéfiniment à ses cibles inhibant la liaison de protéines saines. Ce rôle en tant qu'inhibiteur compétitif en ferait un variant plus délétère qu'une perte de fonction et il pourrait potentiellement induire une ATN en absence de deuxième insulte. Il serait possible de valider cette hypothèse à l'aide de la méthode de luciférase décrite à la section 5. En double transfectant les variants du domaine de trans-activation et une version sauvage de *GRHL3*, il serait possible de tester le potentiel d'inhibition compétitive de ces variants. Des expériences de CHIP-seq utilisant la même double transfection pourraient aussi ajouter à ces résultats en investiguant cette interaction dans un modèle biologique. Une présence diminuée du *GRHL3* sauvage à ses sites de liaison en présence d'un *GRHL3* arborant des variants dans le domaine de trans-activation permettrait de conclure à un effet d'inhibiteur compétitif.

En ce moment, il est impossible de définir si la modulation du phénotype entre les malformations oro-faciales et les ATN est associée à une voie commune, à plusieurs processus différents ou à l'emplacement des variants. Une meilleure compréhension du mécanisme de modulation du phénotype est primordiale pour mieux comprendre la fonction de ce gène dans le contexte de ces pathologies et nous informer sur les processus biologiques importants dans l'étiologie des ATN.

6.2.2.2 Des indices d'une régulation complexe basée sur les isoformes

Des articles précédents ont démontré une discordance entre les isoformes 1 et 4 de *GRHL3* et une concordance entre les isoformes 1 et 2 dans le contexte de l'apoptose et la migration^{324 402}. In vitro, les 3 isoformes activent la transcription d'un gène rapporteur précédé de séquences de liaison de *GRHL3*. L'isoforme 3 n'a pas été inclus dans ces études. Les résultats obtenus à la section 5 sont le premier rapport sur la fonction de cet isoforme. Le test rapporteur utilisé pour l'isoforme 3, diffère peu de ceux utilisés pour les isoformes 1,2 et 4. Il possède 4 régions de liaison de *GRHL3* au lieu de 2 et se trouve dans un vecteur différent. De plus, en comparant les résultats présentés pour l'isoforme 2 dans la section 5, il est possible de confirmer une activation similaire à celle de l'isoforme 4 rapportée par un autre groupe³¹⁴. Ceci supporte l'idée que les fonctions différentes entre les isoformes sont causées par l'expression de gènes différents et pas une variabilité dans les séquences liées. L'isoforme 3 échappe à cette logique. Tel que précédemment discuté, la principale différence de cet isoforme est la présence d'une section 3' qui modifie le site de dimérisation de la protéine. Ce site de dimérisation est fortement associé à la spécificité de liaison des protéines de la famille *GRHL* à leurs sites de liaisons. Les protéines de la famille peuvent

créer des hétérodimères entre eux ou des homodimères³⁰⁸. Ces interactions ont été suggérées comme une des sources potentielles de l'expression différentielle de certains gènes⁴⁰³. Cependant la découverte du facteur *HNF4a* suggère que d'autres facteurs pourraient être impliqués dans cette régulation⁹⁷. L'absence d'activité de l'isoforme 3 est donc probablement liée à une perte d'interaction avec des co-facteurs nécessaires à sa fonction. Afin de comprendre si cette interaction est nécessaire au recrutement de la machinerie de transcription ou à la liaison de la séquence consensus, des études co-immunoprécipitation ou de CHIP-Seq seraient nécessaires. La co-immunoprécipitation des protéines de la machinerie de transcription et de l'isoforme 3 permettrait d'éclaircir sa capacité de recrutement, alors que le CHIP-Seq de l'isoforme 3 permettrait d'éclaircir son potentiel de liaison de sa séquence consensus. Il est aussi intéressant de noter que l'isoforme est plus fortement exprimé par immunobuvardage de type western suggérant une stabilité accrue. Cette stabilité accrue pourrait permettre une inhibition compétitive efficace de la liaison des autres isoformes à leurs cibles ou être un effet collatéral de son manque d'activité dans ce type cellulaire.

La section 5 identifie le premier isoforme de *GRHL3* incapable d'activer un gène rapporteur précédé par sa séquence de liaison. Cette nouvelle information démontre une régulation interne du gène beaucoup plus complexe qu'initialement démontrée avec des variations au niveau de la spécificité des cibles (isoforme 4) et au niveau de la capacité de les activer (isoforme 3). Une meilleure compréhension des interrelations entre les isoformes de *GRHL3* permettrait de mieux comprendre la fonction de ce gène. Tel que suggéré dans la section 5, une étude des variants identifiés chez les patients ATN dans les 4 différents isoformes permettra d'éclairer le mécanisme reliant ce gène à la pathologie.

6.2.2.3 *GRHL3* dans la neurulation

L'implication de *GRHL3* dans la neurulation a été plusieurs fois démontrée, mais son rôle demeure incertain. Dans la section 5, les hypothèses de liens avec la voie PCP ou de la spécification de la plaque neurale sont discutées. Cette section fera un court rappel de cette discussion en plus de discuter d'une troisième hypothèse associée à *SHH*.

Grhl3 est responsable de l'expression de *Rhogef2*, une protéine nécessaire à l'activation de *Rhoa* (Figure 11). Cette information combinée au lien entre *GRHL3* et plusieurs membres de la voie PCP suggère une implication de ce gène dans la régulation de l'effecteur *Rhoa* dans le contexte de la voie PCP. Cependant, une ablation de *Rhoa* chez la souris résulte en un phénotype différent de celui identifié dans la souris *Grhl3*^{-/-}, suggérant que d'autres aspects de *Grhl3* sont impliqués dans l'apparition des ATN. En effet, l'ablation de *Rhoa* entraîne l'apparition d'une exencéphalie à partir de E15.5 suite à une expansion

des progéniteurs neuraux⁴⁰⁴. La présence d'un épaissement de la plaque neurale, typique des défauts dans PCP, demeure incertain^{164 325 405}. Cette voie est un excellent candidat pour expliquer l'implication de *GRHL3* dans les ATN, cependant la différence phénotypique avec *Grhl3*^{-/-} suggère que d'autres facteurs sont impliqués.

Le maintien de l'ectoderme de surface et du neuroectoderme pourrait aussi expliquer son implication dans les ATN (Figure 11). Ces tissus sont nécessaires à la fusion du tube neural et important dans la formation des cellules de la crête neurale. Ces cellules sont ensuite impliquées dans le développement oro-facial. L'échec de rapprochement des pli neuraux, pourrait être causé par l'absence de protrusions nécessaires à générer la force motrice permettant la fermeture. Cependant, cela n'explique pas l'absence de CDL observé chez les souris *Grhl3*^{-/-}¹⁶⁴.

La dérégulation de *Shh* est aussi un mécanisme possible. *Shh* est impliqué la formation de la CM et inhibe les CDL (Figure 11). Une perte de *Grhl3* entraîne une augmentation de *Shh*⁶⁴. Ainsi, une perte de fonction de *Grhl3* peut induire une surexpression de *Shh* qui inhibe la formation des CDL affectant principalement les modes de fermeture 2 et 3. Ces modes associés à la section caudale du tube neural justifient le phénotype de spina bifida chez la souris *Grhl3*^{-/-}. La présence d'une CM et l'absence de CDL s'approche aussi plus du phénotype des souris *Grhl3*^{-/-}. Il faut aussi noter que *Shh* inhibe la formation des CDL seulement à haute concentration, alors qu'une petite concentration est nécessaire à la formation de la CM. Ce système est donc sensible aux petites modifications de concentration. Sachant qu'une surexpression de *Shh* et de divers facteurs de sa signalisation ont été associés aux ATN, il s'agit d'un candidat intéressant⁴⁰⁶. Cependant, l'étude d'une souris mutante *Grhl3*^{-/-} a démontré une quantité d'ARN *Bmp2* et *Noggin* (principaux effecteurs connus de *SHH* dans l'induction des CDL) similaire aux souris saines. Ceci amène les auteurs à suggérer que le défaut de CDL serait généré par d'autres effecteurs que *Noggin* et *BMP2*⁴⁰⁷. Les données de cette étude in situ, sont cependant insuffisantes pour soutenir ces conclusions. La réalité des morphogènes comme *Noggin* et *BMP2* est qu'une faible diminution de leur concentration est suffisante pour modifier leur impact cellulaire⁴⁰⁸. Ainsi, l'étude qualitative utilisée par les auteurs serait incapable d'observer un changement subtil du niveau d'expression. De plus, la fonction principale de *Noggin* est de moduler l'internalisation de BMP en interagissant avec ses récepteurs⁴⁰⁹. Le maintien de l'expression de *Bmp2* dans les souris mutantes ne permet donc pas de conclure à un maintien de sa fonction qui pourrait être modulée par d'autres de ses modulateurs tel que *Chordin* et *Gremlin*. Des études quantitatives sur l'expression de *Noggin*, ainsi que des études sur l'internalisation de *Bmp2* dans

le tube neural murin sont nécessaire afin de clarifier ce mécanisme. Ces résultats suggèrent néanmoins que la voie *Shh-Noggin-Bmp2* n'explique pas la totalité du phénotype.

Chacune de ces hypothèses aurait le potentiel de générer une ATN. Cependant, ces processus ne sont pas exclusifs. Il est possible qu'un échec de plusieurs de ces processus soit nécessaire afin de générer une ATN. Par exemple, WNT pourrait assurer l'expression des mutants *GRHL3* à des endroits spécifiques de la plaque neurale. *SHH* affecterait la formation des CDL éloignant les plis neuraux et cet éloignement contribuerait avec une diminution des protrusions cellulaires à un échec de fusion.

6.3 Évolution de la recherche sur les ATN

À ce jour, l'étude des ATN humains investigate 2 grands axes: Les études de gènes candidats de la souris²³⁵ et les études d'association^{182 183 226 228}. Ces techniques ont participé à la croissance des connaissances sur les ATN, mais sont limitées dans leur spectre d'investigation. L'étude de gènes candidats est souvent limité par l'identification de variants forts chez l'humain qui permettrait de conclure à l'implication de ce gène. Les études d'association se frappent à un mur étant donné la complexité de la maladie et la faible pénétrance des variants. L'étude des voies de signalisation et des processus biologiques candidats, tel que PCP et la voie de l'acide folique, est limitée dans l'identification de nouveaux candidats. Les résultats présentés démontrent l'implication de plusieurs modes de transmission dans cette pathologie ouvrant la porte à de nouveaux devis d'étude qui permettraient d'accélérer la compréhension de l'étiologie génétique de cette pathologie.

6.3.1 Les variants *de novo* dans les méta-analyses, l'identification de gènes candidats et leur association à l'acide folique

En plus d'identifier une forte association au gène *SHROOM3*, la section 4 démontre pour la première fois l'implication de variants *de novo* dans l'étiologie de la maladie. Tel que discuté, l'enrichissement statistique de variants hautement délétères est seulement significatif lorsque comparé à la fréquence de variation la plus faible, mais demeure plus élevé que la fréquence la plus conservative. L'analyse pan-génomique de cette section qui compare la fréquence globale de mutations dans notre cohorte aux fréquences références est fait avec un test binomial et sans Bonferroni étant donné que l'unité du test est le génome entier. La correction pour les tests multiples est uniquement appliquée aux tests qui utilisent le gène comme unité de test (ex : test d'enrichissement de mutations dans *SHROOM3*). Le choix du test binomial est quant à lui attribué au fait que plusieurs articles supportent son utilisation dans ce contexte^{358 410 411}. Cependant, il est aussi intéressant de réaliser qu'en présence de beaucoup d'événements (tel que dans le contexte de la section 4), le test de poisson et le test binomial

convergent. En effet, la valeur P est significative et identique pour ces deux tests avec les données de la section 4. Les résultats ainsi obtenus supportent l'implication des variants *de novo* dans les ATN, mais des études subséquentes sont nécessaires pour le confirmer. L'implication des variants *de novo* ouvre beaucoup de voies intéressantes pour l'étude des ATN.

Premièrement, elle facilite les projets coopératifs de méta-analyse. Ces projets sont une des principales sources d'information génétique dans les maladies neurodéveloppementales tel que la déficience intellectuelle et l'autisme^{412 413}. Les variants *de novo* facilitent ce type d'analyse, car ils sont en général plus faciles à identifier avec confiance. En effet dans le contexte de familles, les variants ont souvent une pénétrance incomplète compliquant leur identification. De plus, l'utilisation d'un mode de filtration basé sur la ségrégation entraîne l'identification de plusieurs variants candidats, créant une banque de variants contaminées par des faux positifs. Les variants *de novo* sont plus rares. Une réduction des variants identifiés diminue le nombre de faux positifs pouvant biaiser une méta-analyse. Parce qu'ils n'ont pas besoin d'être transmis, ces variants ont aussi un potentiel délétère plus élevé, facilitant leur identification et l'élimination de faux positifs. Finalement, la petite taille de ces familles (3 individus), permet aux différents laboratoires de séquencer de plus larges cohortes accélérant la croissance des bases de mutations. Ces bases de données peuvent être ensuite utilisées par les chercheurs afin d'augmenter le pouvoir statistique de leurs analyses et identifier de nouveaux gènes impliqués dans le développement de la pathologie. Cette première démonstration de l'implication des variants *de novo* dans les ATN ouvre la voie à ces analyses qui permettront une croissance rapide de la compréhension de la pathologie. De plus, elle permet de cibler un des mécanismes à travers lequel une pathologie aussi létale que les ATN peut demeurer aussi prévalente dans la société.

L'étude de variants *de novo* a aussi permis d'identifier plusieurs autres candidats intéressants pour les ATN. Tel que discuté dans la section 4, 6 candidats ressortent de l'analyse des trios en dehors de *SHROOM3* et *GRHL3*. Ces candidats sont *PAX3*, microfibrillar associated protein 1 (*MFAP1*), DEAD-box helicase 3 (*DDX3X*), Williams-Beuren syndrome chromosome region 28 (*WBSCR28*) et protein tyrosine phosphatase receptor type S (*PTPRS*). De ces variants, p.S73X sur *PAX3* est le principal candidat. La présence d'ATN chez la souris *Pax3*^{-/-}, la nature haplo-insuffisante de *PAX3* et la délétion de l'allèle protéique suggère un effet pathogénique. Ce gène avait à ce jour seulement été associé au syndrome de Waardenburg, ce qui permet de créer un lien potentiel entre cette pathologie et les ATN. Malgré la candidature plus faible des autres variants, l'association de *PTPRS* aux ATN et les variants hautement délétères chez les autres gènes en font d'excellents candidats. Même en absence d'associations

statistiques, ces gènes contribuent à améliorer les connaissances génétiques sur les ATN en confirmant d'excellents gènes candidats (ex : PAX3) et en identifiant de nouveaux candidats.

De par leur nature, les variants *de novo* ont un potentiel délétère plus élevé. Dans le contexte d'une maladie létale comme les ATN, le potentiel de transmission de variants causatifs est fortement diminué. Les variants *de novo* ne sont pas soumis à cette limitation étant donné leur absence de transmission. Ils ont donc le potentiel d'être hautement pénétrant. Ceci permet d'associer aux ATN de nouveaux gènes intolérants aux variants ou des variants causatifs trop délétères pour être transmis. L'identification de ces facteurs permettra une meilleure compréhension de la maladie.

Finalement, le rôle de l'acide folique dans le développement des ATN demeure encore largement inconnu tel que décrit à la section 1.2.3. L'identification d'un lien entre le risque de développer des mutations *de novo* et la prise péri-conceptionnelle d'acide folique par les parents auraient permis d'éclaircir ce lien. Cependant, malgré la présence d'information sur la prise d'acide folique chez les trios étudiés, aucune association entre la présence de mutations *de novo* et la prise d'acide folique n'a été identifiée. Ceci pourrait être attribué à la taille insuffisante de la cohorte ne permettant pas la puissance statistique nécessaire à la détection d'une telle association. De plus, de par la nature de l'hypothèse nulle, il nous est aussi impossible de conclure à l'absence de lien entre ces facteurs. Néanmoins, nos résultats s'alignent avec les études murines qui démontrent que plusieurs mutations causatives d'ATN sont insensibles à la protection par l'acide folique^{159 160}. En effet, si l'une des 13 mutations identifiées chez des enfants dont les parents ont pris de l'acide folique s'avère causative, cela supporterait la présence de cas génétiques humains insensibles à l'effet protecteur de l'acide folique.

6.3.2 Les familles multiplexes, les variants mendéliens et hautement délétères

Ces résultats identifient aussi pour la première fois un potentiel variant mendélien. Le variant homozygote p.Asp16Aspfs*10 sur le gène *GRHL3*, présent chez deux sœurs d'une famille consanguine affectées par des myeloméningocèles, a toutes les caractéristiques d'un variant mendélien. À ce jour, les études génétiques visent plutôt de larges cohortes de cas sporadiques qui sont plus faciles à recruter, mais qui ont un potentiel limité dans l'identification de variants transmis. Les résultats présentés ici démontrent que le recrutement d'une famille consanguine multiplexe avec un phénotype concordant entre les affectés peut être la clef permettant l'identification de nouveaux variants candidats. Cependant, l'étiologie complexe de cette maladie est parfois mal interprétée par les cliniciens et les chercheurs diminuant la volonté et l'intérêt de recruter ce type de familles. En réalité, la complexité génétique des ATN suggère un mode de transmission principalement oligogénique, mais elle n'exclut pas qu'une faible

proportion des cas sont causés par des variants mendéliens tel qu'observé dans le retard mental non-syndromique⁴¹⁴ et la sclérose latérale amyotrophique⁴¹⁵. Ces cas n'excluent pas non plus la possibilité d'avoir des variants modificateurs qui modulent l'expression physiologique du phénotype. Étant donné la rareté des cas consanguins, un effort global est nécessaire pour le recrutement. Un plus large échantillon de familles consanguines est nécessaire afin de conclure au rôle des variants mendéliens dans l'étiologie des ATN, mais ces résultats ouvrent la porte à un changement de perception quant à l'étiologie complexe des ATN qui est nécessaire à un recrutement accru de ces familles.

Les résultats présentés, démontrent aussi des méthodes d'investigation de ces familles en absence de variants mendéliens. L'identification de variants candidats à travers la large quantité de données générées par le séquençage parallèle massif est compliquée. Le développement de méthodes permettant de prioriser des variants candidats est nécessaire. Les résultats présentés dans la section 5 démontrent l'efficacité de diverses méthodes de priorisation dans le cadre des ATN. En effet, l'utilisation de programmes de prédiction, tel que CADD, est très utile afin de filtrer des variants potentiels. Cette méthode a permis d'identifier un enrichissement statistique de variants dans le domaine de trans-activation de *GRHL3*, en plus d'un nouveau variant faux-sens p.A1a529Val entraînant une perte de fonction de GRHL3. CADD combine plusieurs programmes de prédiction en un seul score. Ce score est ensuite comparé aux autres variants potentiels du génome afin de créer un ordre de pathogénicité de toutes les variations potentielles. Cette comparaison permet de prioriser plus aisément les candidats pour les validations fonctionnelles³⁵⁴. Quatre des huit variants identifiés avec un score supérieur à 20 ont un impact fonctionnel démontrant le fort potentiel prédictif de ce programme. Le fardeau mutationnel a aussi démontré son efficacité dans le contexte de ces études. Il remplit le rôle d'associer un gène avec les ATN et a fait ses preuves dans le contexte de SHROOM3 et de GRHL3.

Il est parfois intimidant d'utiliser le séquençage parallèle massif étant donné la quantité de connaissances nécessaire à son application efficace. Près de 10 ans après l'apparition de cette technique, ces résultats sont les premiers publiés avec cette méthode dans les ATN. Pour cette raison, il est primordial d'abaisser la courbe d'apprentissage en identifiant et en démontrant l'efficacité de différentes méthodes pour l'étude des ATN. Ceci permettra d'ouvrir la voie à de nouveaux laboratoires et aider au développement des connaissances sur la génétique de cette pathologie.

6.3.3 Une étiologie polygénique pour une maladie rare

Un large nombre de gènes suggère en général une prévalence élevée, mais tel qu'observé ici et à travers les études murines, les ATN ont une étiologie hautement polygénique malgré une faible

prévalence. La première hypothèse est que les gènes impliqués dans les ATN sont en moyenne plus conservés diminuant ainsi la prévalence de la maladie due à une rareté des mutations sur ces gènes. Cependant, un test de cette hypothèse utilisant la fréquence d'apparition de mutations *de novo* tronquantes³⁵⁷ sur une liste de gènes précédemment associée à la pathologie murine^{210 211} ne démontre aucune différence significative entre le risque moyen d'apparition de ces mutations dans ces gènes lorsque comparé au risque global moyen (Gènes ATN = 6.768×10^{-6} , moyenne globale = 4.691×10^{-6}). Une deuxième hypothèse supporte la présence du processus de compensation permettant la formation du tube neural même en présence d'une insulte sur un gène clef du développement embryonnaire. Un tel réseau est évolutivement probable considérant le rôle critique de la formation du tube neural dans la survie de l'embryon. La voie PCP, une des principales voies candidates impliquées dans les ATN, supporte cette hypothèse. En effet, l'étude des modèles de cette voie supporte la nécessité de mutations homozygotes ou de mutations hétérozygotes sur deux gènes de la voie afin de développer un phénotype^{212 219 221}. Ceci supporte une certaine résilience de la voie à une expression plus faible de ces gènes. Une accumulation de mutations faiblement pénétrantes, mais affectant néanmoins la capacité de compensation d'une voie nécessaire à la formation du tube neural pourrait ainsi expliquer la prévalence plus faible de cette pathologie.

6.4 Impact clinique des découvertes

La science moderne a entraîné une augmentation rapide des connaissances génétiques créant une révolution au niveau du diagnostic⁴¹⁶. Il est accepté qu'un niveau d'évidences élevé au niveau scientifique et clinique est nécessaire à l'implémentation d'un test diagnostic⁴¹⁷. Cependant, la croissance rapide des connaissances génétiques a rapidement dépassé la capacité d'effectuer des validations appropriées nécessitant une révision des critères⁴¹⁸. Pour cette raison, de nouveaux guides ont commencé à faire surface pour aider les cliniciens à prendre des décisions quant à l'implémentation de ces tests⁴¹⁹⁴²⁰. L'un d'entre eux suggère l'établissement de tiers d'utilisation. Ces tiers sont basés sur le niveau d'évidence, le profil risque-bénéfice, l'impact potentiel sur la santé et l'analyse des coûts. Malheureusement, considérant qu'un niveau d'évidence suffisant nécessite des répliques et des tests d'implémentation clinique, il est impossible d'atteindre ce critère minimal avec les résultats de la section 4 et 5. Une discussion sur le profil risque-bénéfice et l'analyse des coûts permet néanmoins de mieux comprendre le potentiel d'implémentation de différentes approches sur le court et moyen termes.

6.4.1 Profil risque-bénéfice

Deux différents niveaux d'implémentation peuvent être imaginés dans le contexte clinique : Le rapport des données aux familles investiguées ou l'introduction d'un test diagnostique pour des familles ayant des antécédents d'ATN. À ce jour, le conseil génétique consiste à clarifier le risque de récurrence chez un deuxième enfant afin de participer au processus décisionnel des familles. Le diagnostic préimplantatoire (DPI) et la fécondation in-vitro (FIV) sont impossibles étant donné le manque d'évidence pour des candidats génétiques forts. Ce support minimal offert à ces familles suggère donc qu'en présence d'évidences plus fortes, le profil risque-bénéfice s'améliorerait rapidement. Ce profil est le plus positif dans le contexte du rapport des données aux familles investiguées. Les candidats les plus forts identifiés dans la section 4 et 5 (*PAX3*, *SHROOM3*, *GRHL3*) ont un haut potentiel pathogénique et permettrait d'aider au processus décisionnel des parents. Le niveau d'évidence pour les mutations transmises demeure trop faible pour envoyer une famille vers le DPI et FIV, mais la famille atteinte de p.Asp16Aspfs*10 aurait néanmoins beaucoup à gagner à être rencontrée. Les évidences génétiques supportant une transmission homozygote dans cette famille sont extrêmement fortes et mériteraient d'être revues par un clinicien qui pourrait décider de la meilleure marche à suivre pour cette famille. L'implémentation d'un test diagnostique chez les familles avec antécédents semble quant à lui avoir un profil beaucoup moins positif. Une trop faible quantité de gènes sont connus et même ceux identifiés dans les sections 4 et 5 nécessitent des validations supplémentaires, n'offrant quasiment aucune information diagnostique pour ces familles. Le test deviendrait alors un fardeau supplémentaire sur ces familles sans bénéfice clair.

6.4.2 Analyse des coûts

Une analyse applicable dans le système de santé public canadien, suggère qu'en attribuant une partie de l'argent à l'implémentation clinique de ces résultats, on réduit la capacité d'investir dans d'autres aspects du système de santé. Cette analyse des coûts dépasse le spectre de cette thèse. Cependant, quelques pistes d'investigations seront discutées dans ce contexte.

Dans le contexte du retour des résultats aux familles, le coût des tests et de l'analyse des résultats a déjà été payé et les coûts supplémentaires associés au conseil génétique sont minimes supportant ainsi les conclusions effectuées dans la section 6.4.1. C'est surtout au niveau des nouveaux tests que cette analyse se complique. Dans le contexte des variants *de novo*, peu de bénéfices monétaires sont possibles étant donné le risque déjà bas d'une deuxième grossesse affectée. Quoiqu'il soit parfois acceptable d'investir pour pouvoir offrir un diagnostic clair au patient, dans un contexte d'évidences insuffisantes, ce critère n'est pas suffisant⁴¹⁹. Ainsi, la question se positionne plutôt dans le contexte des variants transmis

pour lesquels il serait théoriquement possible de diminuer le risque d'une deuxième grossesse affectée. Le séquençage de l'exome est souvent peu coûteux lorsque comparé à d'autres tests génétiques, cependant le contexte actuel n'effectue aucun test entraînant un nouveau coût net. Ces coûts doivent ensuite être mis en balance avec les coûts associés aux ATN. Le problème est la connaissance de très peu de gènes impliqués et le risque relativement faible d'une deuxième grossesse affectée. En effet, même en priorisant les parents ayant eu une première grossesse affectée, seulement 50% ont un risque de récurrence. De plus, en fonction de résultats de la section 5, 5 cas sur 233 seraient résolus par l'analyse de *GRHL3*¹⁷⁷. Ceci entraînerait un coût majeur, pour trouver un très faible taux de mutations pathogéniques. Le manque de répliquons et de connaissances sur le type de mutations entraînant des pathologies est aussi un problème étant donné que plusieurs des mutations qui seraient identifiées ne pourraient être évaluées correctement. Cette balance de coûts hautement négative suggère qu'une approche de séquençage systématique nécessitera non seulement des validations supplémentaires, mais une connaissance beaucoup plus approfondie pour être implémentée cliniquement.

Suite à des validations de ces gènes candidats, il demeurera néanmoins nécessaire de minimiser les coûts étant donné la faible habileté à prévoir les risques sur un deuxième enfant. La présente discussion démontre clairement l'implication des gènes des ATN dans plusieurs pathologies. Cette implication varie d'un gène à l'autre suggérant qu'un phénotypage adéquat d'un cas syndromique permettrait d'informer le clinicien sur le gène impliqué dans cette famille. Par exemple, en présence du cas syndromique d'ATN/FLP observé dans la famille arborant le variant p.Asp16Aspfs*10, il aurait été possible de prioriser le séquençage de *GRHL3*. Afin de démontrer ces liens, des études plus larges sur des cas syndromiques d'ATN sont nécessaires afin de renforcer ces corrélations. Ceci permettrait de restreindre le séquençage aux parents dont la première grossesse présentait un phénotype syndromique diminuant drastiquement les coûts. Considérant les économies majeures pour chaque grossesse affectée prévenue, ceci permettrait de changer la balance de coûts et de la rendre potentiellement positive.

6.4.3 Retour aux familles et validations dirigées

En conclusion, l'implémentation clinique de tests génétiques pour les ATN est encore très loin. Un effort de toute la communauté est nécessaire afin de démontrer l'impact clinique des gènes identifiés en plus de continuer l'identification de nouveaux candidats. Cependant, suite à l'analyse précédente, plusieurs choses devraient être considérées dans une perspective clinique. De revenir aux patients avec les résultats obtenus dans le cadre des variants *de novo* et du variant p.Asp16Aspfs*10 permettrait d'aider la prise de décision de ces patients et de diminuer le stress subi par l'absence de diagnostic.

Deuxièmement, en approfondissant les connaissances entre les différents phénotypes et les gènes impliqués dans les ATN, il serait possible d'implémenter des tests génétiques pour la partie syndromique de la population à coût plus faible. Ceci permettrait aussi de confirmer ou infirmer l'hypothèse du lien entre les pathologies et participerait à mieux comprendre les ATN.

7. Plans futurs

L'étiologie des ATN demeure encore largement incomprise. Les variants identifiés à ce jour acompte pour une faible proportion des cas. Pour cette raison, des études subséquentes sur des plus larges cohortes ou le reséquençage d'autres gènes candidats dans la cohorte identifiée ici sont nécessaires afin d'approfondir les connaissances associées à cette maladie. L'identification de techniques permettant l'analyse de familles multiplexes sans candidats mendéliens clair est aussi en route. Des résultats préliminaires démontrent que l'utilisation d'ontologies identifie des fardeaux de variations dans certains processus biologiques pertinent aux ATN. Une validation des résultats de cette technique permettrait de conférer un pouvoir d'investigation accrue chez les familles multiplexes. Le séquençage de l'exome est aussi limité dans son spectre d'investigation. En utilisant le séquençage du génome, il serait possible d'identifier de potentiels variants introniques et intergéniques. Certains variants introniques ont le potentiel d'expliquer une large partie de l'étiologie génétique d'une maladie complexe tel qu'observé avec la région C9ORF72 dans la sclérose latérale amyotrophique⁴²¹. Les études de méthylation et d'expression à l'aide de séquençage bi-sulfate ou de RNA-Seq offrent aussi des perspectives différentes sur la pathologie et permettrait de mieux comprendre les causes génétiques des ATN. Plusieurs investigations en lien avec les résultats présentés ici sont aussi à considérer.

La fonction de *Shroom3* dans la souris a été hautement caractérisée et son implication dans la CA n'est plus à démontrer. Cependant, la présence d'autres variants intermédiaires dans les bases de données publiques suggère que les variants identifiés sont modulés par d'autres facteurs entraînant la pathologie. Un bon endroit pour commencer l'investigation de ces autres facteurs serait dans les gènes impliqués dans la CA. Il est probable qu'un effet combiné de plusieurs variants à l'intérieur du même processus aurait un potentiel pathogénique. L'investigation des voies communes entre les ATN, les FLP, la microsomie et l'hétérotaxie serait aussi informative. Tel que discuté dans la section 6.4.2, la capacité d'identifier une variabilité phénotypique permet d'offrir un outil diagnostique qui facilite l'introduction de tests génétiques cliniques. Ainsi une étude approfondie de la co-ségrégation de ces pathologies dans des familles est nécessaire. Cette dernière informerait aussi sur la fonction du gène *SHROOM3* offrant une meilleure compréhension de son mécanisme biologique et de l'implication des variants identifiés dans les ATN.

Pour *GRHL3*, les données suggèrent plusieurs nouvelles pistes d'investigation. Premièrement, l'interaction complexe entre les isoformes doit d'être investiguée. Les variants identifiés seraient introduits dans les 4 isoformes et soumis à des tests d'activation suivi de tests compétitifs. Cela

permettrait de conclure si la liaison ou l'activation sont préférentielles à certain d'entre eux. L'identification des cibles spécifiques à ces isoformes à l'aide de Chip-Seq ou de séquençage d'ARN permettrait aussi d'aider à décrypter cette régulation complexe. Ces expériences offriraient une meilleure compréhension de la régulation d'un excellent candidat pour les ATN. Étant un régulateur de l'expression génique, il est possible que ces tests identifient des candidats intéressants pour la pathologie. De plus, l'investigation des variants dans les différents isoformes permettrait de valider l'impact fonctionnel des variants du domaine de trans-activation et de conclure sur l'implication de la position des variants dans le développement des ATN.

Une meilleure compréhension du lien biologique entre ce gène et les ATN est aussi nécessaire. Dans la section 6.1.2, il est discuté que le rôle de *Grhl3* dans les ATN demeure encore inconnu et l'hypothèse d'un lien avec *Shh* est soulevé. Ceci pourrait être testé à l'aide d'une ablation de *grhl3* par morpholinos chez le xenopus, suivi d'un rétablissement du phénotype à l'aide de la surexpression d'un membre de la voie des *shh*. Une étude plus approfondie du lien entre *GRHL3* et *SHH* à travers des études de PCR quantitatif permettrait aussi de valider dans un modèle cellulaire si les variants identifiés ou si une diminution globale de *GRHL3* modifie l'expression ou la fonction de *SHH*. L'utilisation de in-situ à différents niveaux de l'axe antéro-postérieur de la souris *Grhl3*^{-/-} permettrait aussi de confirmer l'hypothèse selon laquelle ce défaut affecterait préférentiellement la région caudale.

Finalement, l'identification des principaux éléments modulant les cibles de *GRHL3* permettrait de mieux en comprendre la fonction. Le fait que l'isoforme 3 n'a aucune capacité d'activer ses cibles suggère que le domaine de dimérisation est important à la spécificité du gène. Cette observation est supportée par un domaine et une séquence de liaison communs à tous les isoformes. L'identification des éléments modulant cette spécificité permettrait de mieux comprendre la variabilité phénotypique entre SVW2/FLP et les ATN. L'implication de *GRHL2* dans les ATN et le rapport d'une dimérisation entre *GRHL3* et *GRHL2* en fait un candidat intéressant³⁰⁸. En testant l'impact des divers variants identifiés sur la dimérisation de *GRHL3* permettrait potentiellement d'identifier un mécanisme important à la variabilité phénotypique.

8. Conclusion

Les résultats présentés sont une preuve de l'utilité du séquençage parallèle massif dans l'étude des ATN. Ces études démontrent qu'en présence de différentes formes familiales, il est important de modifier les approches utilisées. Le fait de démontrer que les variants *de novo* ont un rôle important à jouer dans l'étiologie des ATN permet d'offrir un mécanisme potentiel pour le maintien de la prévalence de cette maladie létale. Ils sont la première étape vers des méta-analyses avec un pouvoir de détection accru et permettant d'identifier des associations statistiques de nouveaux gènes. Cette approche est l'une des meilleures méthodes pour identifier de nouveaux candidats intéressants dans le cadre de cette pathologie. Les résultats démontrent aussi que l'approche gène candidat est une méthode puissante qui est amplifiée par le pouvoir du séquençage parallèle massif. En effet, à l'aide de méthodes comme les MIPS, il est possible de séquencer un gène sur une très large cohorte pour une fraction du prix. Cette méthode permet de bâtir sur la base de gènes candidats identifiés chez les souris ou, comme dans le cadre de cet ouvrage, sur l'identification de variants intéressants chez l'humain.

Ces résultats démontrent aussi pour la première fois l'association de deux excellents candidats dans la pathologie humaine. Ceci permet de briser les mythes associant ces gènes à d'autres pathologies chez l'humain. Tel que discuté, ces résultats sont aussi intéressants dans un contexte clinique et potentiellement applicables à certains niveaux du conseil génétique des familles ayant eu une grossesse affectée par les ATN. Les résultats démontrent aussi un rôle très intéressant des isoformes de *GRHL3* dans sa régulation. Ils suggèrent aussi que le mode de transmission oligogénique est la norme, mais que des variants avec des potentiels pathogéniques très élevés pourraient potentiellement agir comme des variants mendéliens, ce qui à ce jour n'avait jamais été observé dans la pathologie humaine.

9. Références

1. Gilbert, S. *Developmental Biology*. (Sinauer Associates, 2000).
2. Copp, A. J. & Greene, N. D. Genetics and development of neural tube defects. *J. Pathol.* **220**, 217–230 (2010).
3. Tam, P. P., Loebel, D. A. & Tanaka, S. S. Building the mouse gastrula: signals, asymmetry and lineages. *Curr. Opin. Genet. Dev.* **16**, 419–425 (2006).
4. Beddington, R. S. P. & Robertson, E. J. Axis Development and Early Asymmetry in Mammals. *Cell* **96**, 195–209 (1999).
5. Palis, J. & Yoder, M. C. Yolk-sac hematopoiesis: The first blood cells of mouse and man. *Exp. Hematol.* **29**, 927–936 (2001).
6. Burdsal, C. A., Damsky, C. H. & Pedersen, R. A. The role of E-cadherin and integrins in mesoderm differentiation and migration at the mammalian primitive streak. *Development* **118**, 829–844 (1993).
7. Sulik, K. *et al.* Morphogenesis of the murine node and notochordal plate. *Dev. Dyn.* **201**, 260–278 (1994).
8. Grunz, H. & Tacke, L. Neural differentiation of *Xenopus laevis* ectoderm takes place after disaggregation and delayed reaggregation without inducer. *Cell Differ. Dev. Off. J. Int. Soc. Dev. Biol.* **28**, 211–217 (1989).
9. Godsave, S. F. & Slack, J. M. Single cell analysis of mesoderm formation in the *Xenopus* embryo. *Dev. Camb. Engl.* **111**, 523–530 (1991).
10. Rentzsch, F., Bakkers, J., Kramer, C. & Hammerschmidt, M. Fgf signaling induces posterior neuroectoderm independently of Bmp signaling inhibition. *Dev. Dyn. Off. Publ. Am. Assoc. Anat.* **231**, 750–757 (2004).
11. Delaune, E., Lemaire, P. & Kodjabachian, L. Neural induction in *Xenopus* requires early FGF signalling in addition to BMP inhibition. *Development* **132**, 299–310 (2005).
12. Isaacs, H. V., Pownall, M. E. & Slack, J. M. eFGF regulates Xbra expression during *Xenopus* gastrulation. *EMBO J.* **13**, 4469–4481 (1994).
13. Smukler, S. R., Runciman, S. B., Xu, S. & van der Kooy, D. Embryonic stem cells assume a primitive neural stem cell fate in the absence of extrinsic influences. *J. Cell Biol.* **172**, 79–90 (2006).
14. Baker, J. C., Beddington, R. S. & Harland, R. M. Wnt signaling in *Xenopus* embryos inhibits bmp4 expression and activates neural development. *Genes Dev.* **13**, 3149–3159 (1999).
15. Wilson, S. I. *et al.* The status of Wnt signalling regulates neural and epidermal fates in the chick embryo. *Nature* **411**, 325–330 (2001).
16. Heeg-Truesdell, E. & LaBonne, C. Neural induction in *Xenopus* requires inhibition of Wnt-beta-catenin signaling. *Dev. Biol.* **298**, 71–86 (2006).
17. Tada, M. & Heisenberg, C.-P. Convergent extension: using collective cell migration and cell intercalation to shape embryos. *Development* **139**, 3897–3904 (2012).
18. Keller, R. *et al.* Mechanisms of convergence and extension by cell intercalation. *Philos. Trans. R. Soc. B Biol. Sci.* **355**, 897–922 (2000).
19. Sullivan-Brown, J. & Goldstein, B. Neural Tube Closure: The Curious Case of Shrinking Junctions. *Curr. Biol.* **22**, R574–R576 (2012).
20. Ninomiya, H., Elinson, R. P. & Winklbauer, R. Antero-posterior tissue polarity links mesoderm convergent extension to axial patterning. *Nature* **430**, 364–367 (2004).
21. Keller, R., Shih, J. & Sater, A. The cellular basis of the convergence and extension of the *Xenopus* neural plate. *Dev. Dyn. Off. Publ. Am. Assoc. Anat.* **193**, 199–217 (1992).
22. Elul, T., Koehl, M. A. & Keller, R. Cellular mechanism underlying neural convergent extension in *Xenopus laevis* embryos. *Dev. Biol.* **191**, 243–258 (1997).

23. Ezin, A. M., Skoglund, P. & Keller, R. The midline (notochord and notoplate) patterns the cell motility underlying convergence and extension of the *Xenopus* neural plate. *Dev. Biol.* **256**, 100–114 (2003).
24. Williams, M., Yen, W., Lu, X. & Sutherland, A. Distinct apical and basolateral mechanisms drive planar cell polarity-dependent convergent extension of the mouse neural plate. *Dev. Cell* **29**, 34–46 (2014).
25. Nishimura, T., Honda, H. & Takeichi, M. Planar cell polarity links axes of spatial dynamics in neural-tube closure. *Cell* **149**, 1084–1097 (2012).
26. Bertet, C., Sulak, L. & Lecuit, T. Myosin-dependent junction remodelling controls planar cell intercalation and axis elongation. *Nature* **429**, 667–671 (2004).
27. Gray, R. S., Roszko, I. & Solnica-Krezel, L. Planar Cell Polarity: Coordinating Morphogenetic Cell Behaviors with Embryonic Polarity. *Dev. Cell* **21**, 120–133 (2011).
28. Adler, P. N. Planar Signaling and Morphogenesis in *Drosophila*. *Dev. Cell* **2**, 525–535 (2002).
29. Lawrence, P. A., Casal, J. & Struhl, G. Cell interactions and planar polarity in the abdominal epidermis of *Drosophila*. *Development* **131**, 4651–4664 (2004).
30. Strutt, D. Frizzled signalling and cell polarisation in *Drosophila* and vertebrates. *Development* **130**, 4501–4513 (2003).
31. Shimada, Y., Yonemura, S., Ohkura, H., Strutt, D. & Uemura, T. Polarized Transport of Frizzled along the Planar Microtubule Arrays in *Drosophila* Wing Epithelium. *Dev. Cell* **10**, 209–222 (2006).
32. Strutt, H. & Strutt, D. Differential Stability of Flamingo Protein Complexes Underlies the Establishment of Planar Polarity. *Curr. Biol.* **18**, 1555–1564 (2008).
33. Axelrod, J. D. Unipolar membrane association of Dishevelled mediates Frizzled planar cell polarity signaling. *Genes Dev.* **15**, 1182–1187 (2001).
34. Bastock, R., Strutt, H. & Strutt, D. Strabismus is asymmetrically localised and binds to Prickle and Dishevelled during *Drosophila* planar polarity patterning. *Development* **130**, 3007–3014 (2003).
35. Usui, T. *et al.* Flamingo, a Seven-Pass Transmembrane Cadherin, Regulates Planar Cell Polarity under the Control of Frizzled. *Cell* **98**, 585–595 (1999).
36. Habas, R., Dawid, I. B. & He, X. Coactivation of Rac and Rho by Wnt/Frizzled signaling is required for vertebrate gastrulation. *Genes Dev.* **17**, 295–309 (2003).
37. Wallingford, J. B. *et al.* Dishevelled controls cell polarity during *Xenopus* gastrulation. *Nature* **405**, 81–85 (2000).
38. Topczewski, J. *et al.* The Zebrafish Glypican Knypek Controls Cell Polarity during Gastrulation Movements of Convergent Extension. *Dev. Cell* **1**, 251–264 (2001).
39. Kibar, Z. *et al.* Mutations in VANGL1 Associated with Neural-Tube Defects. *N. Engl. J. Med.* **356**, 1432–1437 (2007).
40. Murdoch, J. N. *et al.* Disruption of scribble (*Scrb1*) causes severe neural tube defects in the circletail mouse. *Hum. Mol. Genet.* **12**, 87–98 (2003).
41. Paudyal, A. *et al.* The novel mouse mutant, *chuzhoi*, has disruption of Ptk7 protein and exhibits defects in neural tube, heart and lung development and abnormal planar cell polarity in the ear. *BMC Dev. Biol.* **10**, 87 (2010).
42. Heisenberg, C. P. *et al.* Silberblick/Wnt11 mediates convergent extension movements during zebrafish gastrulation. *Nature* **405**, 76–81 (2000).
43. Kilian, B. *et al.* The role of Ppt/Wnt5 in regulating cell shape and movement during zebrafish gastrulation. *Mech. Dev.* **120**, 467–476 (2003).
44. MacDonald, B. T., Tamai, K. & He, X. Wnt/ β -catenin signaling: components, mechanisms, and diseases. *Dev. Cell* **17**, 9–26 (2009).
45. Sonderegger, S., Pollheimer, J. & Knöfler, M. Wnt Signalling in Implantation, Decidualisation and Placental Differentiation – Review. *Placenta* **31**, 839–847 (2010).

46. Sawyer, J. M. *et al.* Apical constriction: A cell shape change that can drive morphogenesis. *Dev. Biol.* **341**, 5–19 (2010).
47. Lecuit, T., Lenne, P.-F. & Munro, E. Force Generation, Transmission, and Integration during Cell and Tissue Morphogenesis. *Annu. Rev. Cell Dev. Biol.* **27**, 157–184 (2011).
48. Ebrahim, S. *et al.* NMII Forms a Contractile Transcellular Sarcomeric Network to Regulate Apical Cell Junctions and Tissue Geometry. *Curr. Biol.* **23**, 731–736 (2013).
49. Wu, S. K. *et al.* Cortical F-actin stabilization generates apical–lateral patterns of junctional contractility that integrate cells into epithelia. *Nat. Cell Biol.* **16**, 167–178 (2014).
50. Salbreux, G., Charras, G. & Paluch, E. Actin cortex mechanics and cellular morphogenesis. *Trends Cell Biol.* **22**, 536–545 (2012).
51. Mayer, M., Depken, M., Bois, J. S., Jülicher, F. & Grill, S. W. Anisotropies in cortical tension reveal the physical basis of polarizing cortical flows. *Nature* **467**, 617–621 (2010).
52. Goehring, N. W. & Grill, S. W. Cell polarity: mechanochemical patterning. *Trends Cell Biol.* **23**, 72–80 (2013).
53. David, D. J. V., Tishkina, A. & Harris, T. J. C. The PAR complex regulates pulsed actomyosin contractions during amnioserosa apical constriction in *Drosophila*. *Development* **137**, 1645–1655 (2010).
54. Azevedo, D. *et al.* DRhoGEF2 Regulates Cellular Tension and Cell Pulsations in the Amnioserosa during *Drosophila* Dorsal Closure. *PLOS ONE* **6**, e23964 (2011).
55. David, D. J. V., Wang, Q., Feng, J. J. & Harris, T. J. C. Bazooka inhibits aPKC to limit antagonism of actomyosin networks during amnioserosa apical constriction. *Development* **140**, 4719–4729 (2013).
56. Bresnick, A. R. Molecular mechanisms of nonmuscle myosin-II regulation. *Curr. Opin. Cell Biol.* **11**, 26–33 (1999).
57. Goode, B. L. & Eck, M. J. Mechanism and Function of Formins in the Control of Actin Assembly. *Annu. Rev. Biochem.* **76**, 593–627 (2007).
58. Kölsch, V., Seher, T., Fernandez-Ballester, G. J., Serrano, L. & Leptin, M. Control of *Drosophila* Gastrulation by Apical Localization of Adherens Junctions and RhoGEF2. *Science* **315**, 384–386 (2007).
59. Mason, F. M., Tworoger, M. & Martin, A. C. Apical domain polarization localizes actin–myosin activity to drive ratchet-like apical constriction. *Nat. Cell Biol.* **15**, 926–936 (2013).
60. Gorfinkiel, N. & Arias, A. M. Requirements for adherens junction components in the interaction between epithelial tissues during dorsal closure in *Drosophila*. *J. Cell Sci.* **120**, 3289–3298 (2007).
61. Hildebrand, J. D. Shroom regulates epithelial cell shape via the apical positioning of an actomyosin network. *J. Cell Sci.* **118**, 5191–5203 (2005).
62. Nishimura, T. & Takeichi, M. Shroom3-mediated recruitment of Rho kinases to the apical cell junctions regulates epithelial and neuroepithelial planar remodeling. *Development* **135**, 1493–1502 (2008).
63. Cho, G., Lim, Y., Cho, I.-T., Simonet, J. C. & Golden, J. A. Arx together with FoxA2, regulates Shh floor plate expression. *Dev. Biol.* **393**, 137–148 (2014).
64. Adhikari, N. *et al.* Grhl3 modulates epithelial structure formation of the circumvallate papilla during mouse development. *Histochem. Cell Biol.* **147**, 5–16 (2017).
65. Ybot-Gonzalez, P., Cogram, P., Gerrelli, D. & Copp, A. J. Sonic hedgehog and the molecular regulation of mouse neural tube closure. *Development* **129**, 2507–2517 (2002).
66. Ybot-Gonzalez, P. *et al.* Neural plate morphogenesis during mouse neurulation is regulated by antagonism of Bmp signalling. *Development* **134**, 3203–3211 (2007).
67. Jacobson, A. G. & Moury, J. D. Tissue Boundaries and Cell Behavior during Neurulation. *Dev. Biol.* **171**, 98–110 (1995).

68. Davidson, B. P., Kinder, S. J., Steiner, K., Schoenwolf, G. C. & Tam, P. P. L. Impact of Node Ablation on the Morphogenesis of the Body Axis and the Lateral Asymmetry of the Mouse Embryo during Early Organogenesis. *Dev. Biol.* **211**, 11–26 (1999).
69. Amarnath, S. & Agarwala, S. Cell-cycle-dependent TGF β -BMP antagonism regulates neural tube closure by modulating tight junctions. *J. Cell Sci.* (2016). doi:10.1242/jcs.179192
70. Geelen, J. A. & Langman, J. Ultrastructural observations on closure of the neural tube in the mouse. *Anat. Embryol. (Berl.)* **156**, 73–88 (1979).
71. Holmberg, J., Clarke, D. L. & Frisén, J. Regulation of repulsion versus adhesion by different splice forms of an Eph receptor. *Nature* **408**, 203–206 (2000).
72. Copp, A. J. & Greene, N. D. E. Defining a PARTICULAR Pathway of Neural Tube Closure. *Dev. Cell* **18**, 1–2 (2010).
73. Kimura-Yoshida, C., Mochida, K., Ellwanger, K., Niehrs, C. & Matsuo, I. Fate Specification of Neural Plate Border by Canonical Wnt Signaling and Grhl3 is Crucial for Neural Tube Closure. *EBioMedicine* **2**, 513–527 (2015).
74. Pyrgaki, C., Liu, A. & Niswander, L. Grainyhead-like 2 regulates neural tube closure and adhesion molecule expression during neural fold fusion. *Dev. Biol.* **353**, 38–49 (2011).
75. O’Rahilly, R. & Müller, F. The two sites of fusion of the neural folds and the two neuropores in the human embryo. *Teratology* **65**, 162–170 (2002).
76. Van Allen, M. I. *et al.* Evidence for multi-site closure of the neural tube in humans. *Am. J. Med. Genet.* **47**, 723–743 (1993).
77. Mahalik, S. K., Vaze, D., Lyngdoh, T. S., Tewari, M. K. & Narasimhan, K. L. Embryogenesis of triple neural tube defects: sonic hedgehog—a key? *J. Pediatr. Surg.* **46**, e5-8 (2011).
78. Colas, J. F. & Schoenwolf, G. C. Towards a cellular and molecular understanding of neurulation. *Dev. Dyn. Off. Publ. Am. Assoc. Anat.* **221**, 117–145 (2001).
79. Hughes, A. F. & Freeman, R. B. Comparative remarks on the development of the tail cord among higher vertebrates. *J. Embryol. Exp. Morphol.* **32**, 355–363 (1974).
80. Lowery, L. A. & Sive, H. Strategies of vertebrate neurulation and a re-evaluation of teleost neural tube formation. *Mech. Dev.* **121**, 1189–1197 (2004).
81. Kalluri, R. & Weinberg, R. A. The basics of epithelial-mesenchymal transition. *J. Clin. Invest.* **119**, 1420–1428 (2009).
82. Zeisberg, E. M. *et al.* Endothelial-to-mesenchymal transition contributes to cardiac fibrosis. *Nat. Med.* **13**, 952–961 (2007).
83. Zeisberg, M. *et al.* Fibroblasts derive from hepatocytes in liver fibrosis via epithelial to mesenchymal transition. *J. Biol. Chem.* **282**, 23337–23347 (2007).
84. Hanahan, D. & Weinberg, R. A. The hallmarks of cancer. *Cell* **100**, 57–70 (2000).
85. Thiery, J. P. Epithelial-mesenchymal transitions in tumour progression. *Nat. Rev. Cancer* **2**, 442–454 (2002).
86. Hay, E. D. Role of cell-matrix contacts in cell migration and epithelial-mesenchymal transformation. *Cell Differ. Dev. Off. J. Int. Soc. Dev. Biol.* **32**, 367–375 (1990).
87. Hay, E. D. The mesenchymal cell, its role in the embryo, and the remarkable signaling mechanisms that create it. *Dev. Dyn. Off. Publ. Am. Assoc. Anat.* **233**, 706–720 (2005).
88. Burstyn-Cohen, T., Stanleigh, J., Sela-Donenfeld, D. & Kalcheim, C. Canonical Wnt activity regulates trunk neural crest delamination linking BMP/noggin signaling with G1/S transition. *Dev. Camb. Engl.* **131**, 5327–5339 (2004).
89. Chea, H. K., Wright, C. V. & Swalla, B. J. Nodal signaling and the evolution of deuterostome gastrulation. *Dev. Dyn. Off. Publ. Am. Assoc. Anat.* **234**, 269–278 (2005).
90. Shah, S. B. *et al.* Misexpression of chick Vg1 in the marginal zone induces primitive streak formation. *Dev. Camb. Engl.* **124**, 5127–5138 (1997).

91. Ciruna, B. & Rossant, J. FGF signaling regulates mesoderm cell fate specification and morphogenetic movement at the primitive streak. *Dev. Cell* **1**, 37–49 (2001).
92. Lindsley, R. C. *et al.* Mesp1 coordinately regulates cardiovascular fate restriction and epithelial-mesenchymal transition in differentiating ESCs. *Cell Stem Cell* **3**, 55–68 (2008).
93. Arnold, S. J., Hofmann, U. K., Bikoff, E. K. & Robertson, E. J. Pivotal roles for eomesodermin during axis formation, epithelium-to-mesenchyme transition and endoderm specification in the mouse. *Dev. Camb. Engl.* **135**, 501–511 (2008).
94. Nieto, M. A. The snail superfamily of zinc-finger transcription factors. *Nat. Rev. Mol. Cell Biol.* **3**, 155–166 (2002).
95. Cavatorta, A. L., Giri, A. A., Banks, L. & Gardiol, D. Cloning and functional analysis of the promoter region of the human Disc large gene. *Gene* **424**, 87–95 (2008).
96. Mistry, D. S., Chen, Y., Wang, Y., Zhang, K. & Sen, G. L. SNAI2 Controls the Undifferentiated State of Human Epidermal Progenitor Cells. *STEM CELLS* **32**, 3209–3218 (2014).
97. Alotaibi, H. *et al.* Enhancer cooperativity as a novel mechanism underlying the transcriptional regulation of E-cadherin during mesenchymal to epithelial transition. *Biochim. Biophys. Acta BBA - Gene Regul. Mech.* **1849**, 731–742 (2015).
98. Werth, M. *et al.* The transcription factor grainyhead-like 2 regulates the molecular composition of the epithelial apical junctional complex. *Development* **137**, 3835–3845 (2010).
99. Saitu, H., Yamada, S., Uwabe, C., Ishibashi, M. & Shiota, K. Development of the posterior neural tube in human embryos. *Anat. Embryol. (Berl.)* **209**, 107–117 (2004).
100. Kostović-Knezević, L., Gajović, S. & Svajger, A. Morphogenetic features in the tail region of the rat embryo. *Int. J. Dev. Biol.* **35**, 191–195 (1991).
101. Pang, D., Zovickian, J. & Moes, G. S. Retained medullary cord in humans: late arrest of secondary neurulation. *Neurosurgery* **68**, 1500–1519; discussion 1519 (2011).
102. De Wals, P. *et al.* Reduction in Neural-Tube Defects after Folic Acid Fortification in Canada. *N. Engl. J. Med.* **357**, 135–142 (2007).
103. Bassuk, A. G. & Kibar, Z. Genetic Basis of Neural Tube Defects. *Semin. Pediatr. Neurol.* **16**, 101–110 (2009).
104. Rossi, A. *et al.* Imaging in spine and spinal cord malformations. *Eur. J. Radiol.* **50**, 177–200 (2004).
105. Copp, A. J., Stanier, P. & Greene, N. D. Neural tube defects: recent advances, unsolved questions, and controversies. *Lancet Neurol.* **12**, 799–810 (2013).
106. Cook, R. J., Erdman, J. N., Hevia, M. & Dickens, B. M. Prenatal management of anencephaly. *Int. J. Gynecol. Obstet.* **102**, 304–308 (2008).
107. Cohen, M. M. Malformations of the craniofacial region: Evolutionary, embryonic, genetic, and clinical perspectives. *Am. J. Med. Genet.* **115**, 245–268 (2002).
108. Cohen, M. M. & Lemire, R. J. Syndromes with cephaloceles. *Teratology* **25**, 161–172 (1982).
109. Stevenson, R. E., Solomon, B. D. & Everman, D. B. *Human Malformations and Related Anomalies.* (Oxford University Press, 2015).
110. Henry, C., MacKay, D. & Dias, M. Encephalocele (Cephalocele; Encephalocystocele; Encephalomyelocele; Hydroencephalocele; Hydromeningocele, cranial; Meningocele; Meningoencephalocele) | The Journal of Family Practice. Available at: <http://www.mdedge.com/jfponline/dsm/13810/dermatology/encephalocele-cephalocele-encephalocystocele-encephalomyelocele>. (Accessed: 3rd April 2017)
111. Lesavoy, M. A., Nguyen, D. T., Yospur, G. & Dickinson, B. P. Nasopharyngeal encephalocele: report of transcranial and transpalatal repair with a 25-year follow-up. *J. Craniofac. Surg.* **20**, 2251–2256 (2009).

112. Coskun, A., Kiran, G. & Ozdemir, O. Craniorachischisis Totalis: A Case Report and Review of the Literature. *Fetal Diagn. Ther.* **25**, 21–25 (2009).
113. Moore, C. A. *et al.* Elevated rates of severe neural tube defects in a high-prevalence area in northern China. *Am. J. Med. Genet.* **73**, 113–118 (1997).
114. Rufener, S. L., Ibrahim, M., Raybaud, C. A. & Parmar, H. A. Congenital Spine and Spinal Cord Malformations—Pictorial Review. *Am. J. Roentgenol.* **194**, S26–S37 (2010).
115. Spina Bifida Treatment & Management: Approach Considerations, Bladder Management, Bowel Management. (2017).
116. McLone, D. G. Care of the neonate with a myelomeningocele. *Neurosurg. Clin. N. Am.* **9**, 111–120 (1998).
117. Parmar, H., Patkar, D., Shah, J. & Maheshwari, M. Diastematomyelia with terminal lipomyelocystocele arising from one hemicord: case report. *Clin. Imaging* **27**, 41–43 (2003).
118. Kim, S. K. *et al.* Diastematomyelia--clinical manifestation and treatment outcome. *J. Korean Med. Sci.* **9**, 135–144 (1994).
119. Naidich, T. P., McLone, D. G. & Mutluer, S. A new understanding of dorsal dysraphism with lipoma (lipomyeloschisis): radiologic evaluation and surgical correction. *AJR Am. J. Roentgenol.* **140**, 1065–1078 (1983).
120. Pierre-Kahn, A. *et al.* Congenital lumbosacral lipomas. *Childs Nerv. Syst. ChNS Off. J. Int. Soc. Pediatr. Neurosurg.* **13**, 298–334; discussion 335 (1997).
121. Castillo, M. *Neuroradiology*. (Lippincott Williams & Wilkins, 2002).
122. Lee, K. S., Gower, D. J., McWhorter, J. M. & Albertson, D. A. The role of MR imaging in the diagnosis and treatment of anterior sacral meningocele. Report of two cases. *J. Neurosurg.* **69**, 628–631 (1988).
123. Chambers, W. R. Technic for the early operation of myelocele and meningomyelocele. *Am. J. Surg.* **80**, 386–393 (1950).
124. Byrd, S. E., Harvey, C. & Darling, C. F. MR of terminal myelocystoceles. *Eur. J. Radiol.* **20**, 215–220 (1995).
125. McLone, D. G. & Naidich, T. P. Terminal myelocystocele. *Neurosurgery* **16**, 36–43 (1985).
126. Pang, D., Zovickian, J., Lee, J. Y., Moes, G. S. & Wang, K.-C. Terminal myelocystocele: surgical observations and theory of embryogenesis. *Neurosurgery* **70**, 1383-1404; discussion 1404-1405 (2012).
127. Tortori-Donati, P., Rossi, A. & Cama, A. Spinal dysraphism: a review of neuroradiological features with embryological correlations and proposal for a new classification. *Neuroradiology* **42**, 471–491 (2000).
128. Brown, E., Matthes, J. C., Bazan, C. & Jinkins, J. R. Prevalence of incidental intraspinal lipoma of the lumbosacral spine as determined by MRI. *Spine* **19**, 833–836 (1994).
129. Kim, C. H. *et al.* Spinal intramedullary lipoma: report of three cases. *Spinal Cord* **41**, 310–315 (2003).
130. Yundt, K. D., Park, T. S. & Kaufman, B. A. Normal diameter of filum terminale in children: in vivo measurement. *Pediatr. Neurosurg.* **27**, 257–259 (1997).
131. Scotti, G., Harwood-Nash, D. C. & Hoffman, H. J. Congenital thoracic dermal sinus: diagnosis by computer assisted metrizamide myelography. *J. Comput. Assist. Tomogr.* **4**, 675–677 (1980).
132. Singh, I., Rohilla, S., Kumar, P. & Sharma, S. Spinal dorsal dermal sinus tract: An experience of 21 cases. *Surg. Neurol. Int.* **6**, S429–S434 (2015).
133. Coleman, L. T., Zimmerman, R. A. & Rorke, L. B. Ventriculus terminalis of the conus medullaris: MR findings in children. *AJNR Am. J. Neuroradiol.* **16**, 1421–1426 (1995).
134. Sigal, R. *et al.* Ventriculus terminalis of the conus medullaris: MR imaging in four patients with congenital dilatation. *AJNR Am. J. Neuroradiol.* **12**, 733–737 (1991).

135. Suh, S. H., Chung, T.-S., Lee, S.-K., Cho, Y.-E. & Kim, K. S. Ventriculus Terminalis in Adults: Unusual Magnetic Resonance Imaging Features and Review of the Literature. *Korean J. Radiol.* **13**, 557–563 (2012).
136. Pang, D. Split cord malformation: Part II: Clinical syndrome. *Neurosurgery* **31**, 481–500 (1992).
137. Miller, A., Guille, J. T. & Bowen, J. R. Evaluation and treatment of diastematomyelia. *J. Bone Joint Surg. Am.* **75**, 1308–1317 (1993).
138. Duhamel, B. From the Mermaid to Anal Imperforation: The Syndrome of Caudal Regression. *Arch. Dis. Child.* **36**, 152–155 (1961).
139. Nievelestein, R. A., Valk, J., Smit, L. M. & Vermeij-Keers, C. MR of the caudal regression syndrome: embryologic implications. *AJNR Am. J. Neuroradiol.* **15**, 1021–1029 (1994).
140. Tortori-Donati, P. *et al.* Segmental spinal dysgenesis: neuroradiologic findings with clinical and embryologic correlation. *AJNR Am. J. Neuroradiol.* **20**, 445–456 (1999).
141. Smithells, R. W., Sheppard, S. & Schorah, C. J. Vitamin deficiencies and neural tube defects. *Arch. Dis. Child.* **51**, 944–950 (1976).
142. Czeizel, A. E. & Dudás, I. Prevention of the first occurrence of neural-tube defects by periconceptional vitamin supplementation. *N. Engl. J. Med.* **327**, 1832–1835 (1992).
143. Blom, H. J., Shaw, G. M., den Heijer, M. & Finnell, R. H. Neural tube defects and folate: case far from closed. *Nat. Rev. Neurosci.* **7**, 724–731 (2006).
144. Barber, R. C., Lammer, E. J., Shaw, G. M., Greer, K. A. & Finnell, R. H. The Role of Folate Transport and Metabolism in Neural Tube Defect Risk. *Mol. Genet. Metab.* **66**, 1–9 (1999).
145. Beaudin, A. E. *et al.* Shmt1 and de novo thymidylate biosynthesis underlie folate-responsive neural tube defects in mice. *Am. J. Clin. Nutr.* **93**, 789–798 (2011).
146. Barbera, J. P. M. *et al.* Folic acid prevents exencephaly in Cited2 deficient mice. *Hum. Mol. Genet.* **11**, 283–293 (2002).
147. Dunlevy, L. P. E. *et al.* Integrity of the methylation cycle is essential for mammalian neural tube closure. *Birt. Defects Res. A. Clin. Mol. Teratol.* **76**, 544–552 (2006).
148. Dunlevy, L. P. E., Burren, K. A., Chitty, L. S., Copp, A. J. & Greene, N. D. E. Excess methionine suppresses the methylation cycle and inhibits neural tube closure in mouse embryos. *FEBS Lett.* **580**, 2803–2807 (2006).
149. Essien, F. B. & Wannberg, S. L. Methionine but not folinic acid or vitamin B-12 alters the frequency of neural tube defects in Axd mutant mice. *J. Nutr.* **123**, 27–34 (1993).
150. Narisawa, A. *et al.* Mutations in genes encoding the glycine cleavage system predispose to neural tube defects in mice and humans. *Hum. Mol. Genet.* **21**, 1496–1503 (2012).
151. Chang, H. *et al.* Tissue-specific distribution of aberrant DNA methylation associated with maternal low-folate status in human neural tube defects. *J. Nutr. Biochem.* **22**, 1172–1177 (2011).
152. Okano, M., Bell, D. W., Haber, D. A. & Li, E. DNA Methyltransferases Dnmt3a and Dnmt3b Are Essential for De Novo Methylation and Mammalian Development. *Cell* **99**, 247–257 (1999).
153. Bjorklund, N. K. & Gordon, R. A hypothesis linking low folate intake to neural tube defects due to failure of post-translation methylations of the cytoskeleton. *Int. J. Dev. Biol.* **50**, 135–141 (2003).
154. Moephuli, S. R., Klein, N. W., Baldwin, M. T. & Krider, H. M. Effects of methionine on the cytoplasmic distribution of actin and tubulin during neural tube closure in rat embryos. *Proc. Natl. Acad. Sci. U. S. A.* **94**, 543–548 (1997).
155. Chen, Z. *et al.* Mice deficient in methylenetetrahydrofolate reductase exhibit hyperhomocysteinemia and decreased methylation capacity, with neuropathology and aortic lipid deposition. *Hum. Mol. Genet.* **10**, 433–444 (2001).
156. Burren, K. A. *et al.* Gene-environment interactions in the causation of neural tube defects: folate deficiency increases susceptibility conferred by loss of Pax3 function. *Hum. Mol. Genet.* **17**, 3675–3685 (2008).

157. Cockroft, D. L. Changes with gestational age in the nutritional requirements of postimplantation rat embryos in culture. *Teratology* **38**, 281–290 (1988).
158. Levine, J. Controlled trials of inositol in psychiatry. *Eur. Neuropsychopharmacol. J. Eur. Coll. Neuropsychopharmacol.* **7**, 147–155 (1997).
159. Greene, N. D. E. & Copp, A. J. Inositol prevents folate-resistant neural tube defects in the mouse. *Nat. Med.* **3**, 60–66 (1997).
160. Cogram, P. *et al.* D-chiro-inositol is more effective than myo-inositol in preventing folate-resistant mouse neural tube defects. *Hum. Reprod. Oxf. Engl.* **17**, 2451–2458 (2002).
161. Copp, A. J., Brook, F. A. & Roberts, H. J. A cell-type-specific abnormality of cell proliferation in mutant (curly tail) mouse embryos developing spinal neural tube defects. *Dev. Camb. Engl.* **104**, 285–295 (1988).
162. Gustavsson, P. *et al.* Increased expression of Grainyhead-like-3 rescues spina bifida in a folate-resistant mouse model. *Hum. Mol. Genet.* **16**, 2640–2646 (2007).
163. Bhandari, A. *et al.* The Grainyhead transcription factor Grhl3/Get1, suppresses miR-21 expression and tumorigenesis in skin: Modulation of the miR-21 target MSH2 by RNA-binding protein DND1. *Oncogene* **32**, 1497–1507 (2013).
164. Ting, S. B. *et al.* Inositol- and folate-resistant neural tube defects in mice lacking the epithelial-specific factor Grhl-3. *Nat. Med.* **9**, 1513–1519 (2003).
165. Cavalli, P., Tonni, G., Grosso, E. & Poggiani, C. Effects of inositol supplementation in a cohort of mothers at risk of producing an NTD pregnancy. *Birt. Defects Res. A. Clin. Mol. Teratol.* **91**, 962–965 (2011).
166. Groenen, P. M. *et al.* Maternal myo-inositol, glucose, and zinc status is associated with the risk of offspring with spina bifida. *Am. J. Obstet. Gynecol.* **189**, 1713–1719 (2003).
167. Cogram, P., Hynes, A., Dunlevy, L. P. E., Greene, N. D. E. & Copp, A. J. Specific isoforms of protein kinase C are essential for prevention of folate-resistant neural tube defects by inositol. *Hum. Mol. Genet.* **13**, 7–14 (2004).
168. Afman, L. a., Blom, H. j., Van Der Put, N. m. j. & Van Straaten, H. w. m. Homocysteine interference in neurulation: A chick embryo model. *Birt. Defects Res. A. Clin. Mol. Teratol.* **67**, 421–428 (2003).
169. Carmichael, S. L., Yang, W. & Shaw, G. M. Periconceptional nutrient intakes and risks of neural tube defects in California. *Birt. Defects Res. A. Clin. Mol. Teratol.* **88**, 670–678 (2010).
170. Shaw, G. M., Carmichael, S. L., Yang, W., Selvin, S. & Schaffer, D. M. Periconceptional Dietary Intake of Choline and Betaine and Neural Tube Defects in Offspring. *Am. J. Epidemiol.* **160**, 102–109 (2004).
171. Menegola, E. *et al.* Inhibition of histone deacetylase activity on specific embryonic tissues as a new mechanism for teratogenicity. *Birth Defects Res. B. Dev. Reprod. Toxicol.* **74**, 392–398 (2005).
172. Chappell, J. H., Wang, X. D. & Loeken, M. R. Diabetes and apoptosis: neural crest cells and neural tube. *Apoptosis Int. J. Program. Cell Death* **14**, 1472–1483 (2009).
173. Rasmussen, S. A., Chu, S. Y., Kim, S. Y., Schmid, C. H. & Lau, J. Maternal obesity and risk of neural tube defects: a metaanalysis. *Am. J. Obstet. Gynecol.* **198**, 611–619 (2008).
174. Moretti, M. E., Bar-Oz, B., Fried, S. & Koren, G. Maternal hyperthermia and the risk for neural tube defects in offspring: systematic review and meta-analysis. *Epidemiol. Camb. Mass* **16**, 216–219 (2005).
175. Falconer, D. S. The inheritance of liability to certain diseases, estimated from the incidence among relatives. *Ann. Hum. Genet.* **29**, 51–76 (1965).
176. Woolf, C. M. A genetic study of spina bifida cystica in Utah. *Soc. Biol.* **22**, 216–220 (1975).
177. Dtrait, E. R. *et al.* Human Neural Tube Defects: Developmental Biology, Epidemiology, and Genetics. *Neurotoxicol. Teratol.* **27**, 515–524 (2005).

178. Khoury, M. J., Beaty, T. H. & Liang, K. Y. Can familial aggregation of disease be explained by familial aggregation of environmental risk factors? *Am. J. Epidemiol.* **127**, 674–683 (1988).
179. Partington, M. D. & McLone, D. G. Hereditary factors in the etiology of neural tube defects. Results of a survey. *Pediatr. Neurosurg.* **23**, 311–316 (1995).
180. Hume, R. F. *et al.* Aneuploidy among prenatally detected neural tube defects. *Am. J. Med. Genet.* **61**, 171–173 (1996).
181. Kennedy, D., Chitayat, D., Winsor, E. J., Silver, M. & Toi, A. Prenatally diagnosed neural tube defects: ultrasound, chromosome, and autopsy or postnatal findings in 212 cases. *Am. J. Med. Genet.* **77**, 317–321 (1998).
182. Stamm, D. S. *et al.* High-density single nucleotide polymorphism screen in a large multiplex neural tube defect family refines linkage to loci at 7p21.1–pter and 2q33.1–q35. *Birt. Defects Res. A. Clin. Mol. Teratol.* **76**, 499–505 (2006).
183. Rampersaud, E. *et al.* Whole genomewide linkage screen for neural tube defects reveals regions of interest on chromosomes 7 and 10. *J. Med. Genet.* **42**, 940–946 (2005).
184. Hou, Z. & Matherly, L. H. Biology of the Major Facilitative Folate Transporters SLC19A1 and SLC46A1. *Curr. Top. Membr.* **73**, 175–204 (2014).
185. Greene, N. D. E., Stanier, P. & Copp, A. J. Genetics of human neural tube defects. *Hum. Mol. Genet.* **18**, R113–R129 (2009).
186. Frosst, P. *et al.* A candidate genetic risk factor for vascular disease: a common mutation in methylenetetrahydrofolate reductase. *Nat. Genet.* **10**, 111–113 (1995).
187. Botto, L. D. & Yang, Q. 5,10-Methylenetetrahydrofolate reductase gene variants and congenital anomalies: a HuGE review. *Am. J. Epidemiol.* **151**, 862–877 (2000).
188. Vieira, A. R. *et al.* Studies of reduced folate carrier 1 (RFC1) A80G and 5,10-methylenetetrahydrofolate reductase (MTHFR) C677T polymorphisms with neural tube and orofacial cleft defects. *Am. J. Med. Genet. A.* **135A**, 220–223 (2005).
189. De Marco, P. *et al.* Study of MTHFR and MS polymorphisms as risk factors for NTD in the Italian population. *J. Hum. Genet.* **47**, 319–324 (2002).
190. Parle-McDermott, A. *et al.* Analysis of the MTHFR 1298A-->C and 677C-->T polymorphisms as risk factors for neural tube defects. *J. Hum. Genet.* **48**, 190–193 (2003).
191. Piedrahita, J. A. *et al.* Mice lacking the folic acid-binding protein Folbp1 are defective in early embryonic development. *Nat. Genet.* **23**, 228–232 (1999).
192. De Marco, P. *et al.* Reduced folate carrier polymorphism (80A->G) and neural tube defects. *Eur. J. Hum. Genet.* **11**, 245–252 (2003).
193. Brody, L. C. *et al.* A polymorphism, R653Q, in the trifunctional enzyme methylenetetrahydrofolate dehydrogenase/methenyltetrahydrofolate cyclohydrolase/formyltetrahydrofolate synthetase is a maternal genetic risk factor for neural tube defects: report of the Birth Defects Research Group. *Am. J. Hum. Genet.* **71**, 1207–1215 (2002).
194. Hol, F. A. *et al.* Molecular genetic analysis of the gene encoding the trifunctional enzyme MTHFD (methylenetetrahydrofolate-dehydrogenase, methenyltetrahydrofolate-cyclohydrolase, formyltetrahydrofolate synthetase) in patients with neural tube defects. *Clin. Genet.* **53**, 119–125 (1998).
195. Volcik, K. A. *et al.* Associations between polymorphisms within the thymidylate synthase gene and spina bifida. *Birt. Defects Res. A. Clin. Mol. Teratol.* **67**, 924–928 (2003).
196. Trinh, B. N., Ong, C.-N., Coetzee, G. A., Yu, M. C. & Laird, P. W. Thymidylate synthase: a novel genetic determinant of plasma homocysteine and folate levels. *Hum. Genet.* **111**, 299–302 (2002).
197. Brown, K. S. *et al.* The thymidylate synthase tandem repeat polymorphism is not associated with homocysteine concentrations in healthy young subjects. *Hum. Genet.* **114**, 182–185 (2004).

198. Wilding, C. S. *et al.* Thymidylate synthase repeat polymorphisms and risk of neural tube defects in a population from the northern United Kingdom. *Birt. Defects Res. A. Clin. Mol. Teratol.* **70**, 483–485 (2004).
199. Johnson, W. G. *et al.* New 19 bp deletion polymorphism in intron-1 of dihydrofolate reductase (DHFR): a risk factor for spina bifida acting in mothers during pregnancy? *Am. J. Med. Genet. A.* **124A**, 339–345 (2004).
200. Relton, C. L. *et al.* Gene-gene interaction in folate-related genes and risk of neural tube defects in a UK population. *J. Med. Genet.* **41**, 256–260 (2004).
201. Relton, C. L. *et al.* Low erythrocyte folate status and polymorphic variation in folate-related genes are associated with risk of neural tube defect pregnancy. *Mol. Genet. Metab.* **81**, 273–281 (2004).
202. Guéant-Rodriguez, R. M. *et al.* Transcobalamin and methionine synthase reductase mutated polymorphisms aggravate the risk of neural tube defects in humans. *Neurosci. Lett.* **344**, 189–192 (2003).
203. Doolin, M.-T. *et al.* Maternal genetic effects, exerted by genes involved in homocysteine remethylation, influence the risk of spina bifida. *Am. J. Hum. Genet.* **71**, 1222–1226 (2002).
204. Luccock, M., Daskalakis, I., Briggs, D., Yates, Z. & Levene, M. Altered folate metabolism and disposition in mothers affected by a spina bifida pregnancy: influence of 677c --> t methylenetetrahydrofolate reductase and 2756a --> g methionine synthase genotypes. *Mol. Genet. Metab.* **70**, 27–44 (2000).
205. Luccock, M., Daskalakis, I., Hinkins, M. & Yates, Z. An Examination of Polymorphic Genes and Folate Metabolism in Mothers Affected by a Spina Bifida Pregnancy. *Mol. Genet. Metab.* **73**, 322–332 (2001).
206. Christensen, B. *et al.* Genetic polymorphisms in methylenetetrahydrofolate reductase and methionine synthase, folate levels in red blood cells, and risk of neural tube defects. *Am. J. Med. Genet.* **84**, 151–157 (1999).
207. Linden, I. J. M. van der *et al.* The methionine synthase reductase 66A>G polymorphism is a maternal risk factor for spina bifida. *J. Mol. Med.* **84**, 1047–1054 (2006).
208. Rothenberg, S. P. *et al.* Autoantibodies against Folate Receptors in Women with a Pregnancy Complicated by a Neural-Tube Defect. *N. Engl. J. Med.* **350**, 134–142 (2004).
209. Molloy, A. M. *et al.* Lack of Association between Folate-Receptor Autoantibodies and Neural-Tube Defects. *N. Engl. J. Med.* **361**, 152–160 (2009).
210. Harris, M. J. & Juriloff, D. M. Mouse mutants with neural tube closure defects and their role in understanding human neural tube defects. *Birt. Defects Res. A. Clin. Mol. Teratol.* **79**, 187–210 (2007).
211. Harris, M. J. & Juriloff, D. M. An update to the list of mouse mutants with neural tube closure defects and advances toward a complete genetic perspective of neural tube closure. *Birt. Defects Res. A. Clin. Mol. Teratol.* **88**, 653–669 (2010).
212. Kibar, Z. *et al.* Ltap, a mammalian homolog of Drosophila Strabismus/Van Gogh, is altered in the mouse neural tube mutant Loop-tail. *Nat. Genet.* **28**, 251–255 (2001).
213. Kibar, Z. *et al.* Mutations in VANGL1 Associated with Neural-Tube Defects. *N. Engl. J. Med.* **356**, 1432–1437 (2007).
214. De Marco, P. *et al.* Genetic analysis of disheveled 2 and disheveled 3 in human neural tube defects. *J. Mol. Neurosci. MN* **49**, 582–588 (2013).
215. Reynolds, A. *et al.* VANGL1 rare variants associated with neural tube defects affect convergent extension in zebrafish. *Mech. Dev.* **127**, 385–392 (2010).
216. Kibar, Z. *et al.* Contribution of VANGL2 mutations to isolated neural tube defects. *Clin. Genet.* **80**, 76–82 (2011).

217. Lei, Y.-P. *et al.* VANGL2 mutations in human cranial neural-tube defects. *N. Engl. J. Med.* **362**, 2232–2235 (2010).
218. Guyot, M.-C. *et al.* A novel hypomorphic Looptail allele at the planar cell polarity Vangl2 gene. *Dev. Dyn. Off. Publ. Am. Assoc. Anat.* **240**, 839–849 (2011).
219. Wang, Y., Guo, N. & Nathans, J. The role of Frizzled3 and Frizzled6 in neural tube closure and in the planar polarity of inner-ear sensory hair cells. *J. Neurosci. Off. J. Soc. Neurosci.* **26**, 2147–2156 (2006).
220. De Marco, P. *et al.* FZD6 is a novel gene for human neural tube defects. *Hum. Mutat.* **33**, 384–390 (2012).
221. Hamblet, N. S. *et al.* Dishevelled 2 is essential for cardiac outflow tract development, somite segmentation and neural tube closure. *Dev. Camb. Engl.* **129**, 5827–5838 (2002).
222. Merello, E. *et al.* Rare missense variants in DVL1, one of the human counterparts of the *Drosophila* dishevelled gene, do not confer increased risk for neural tube defects. *Birt. Defects Res. A. Clin. Mol. Teratol.* **97**, 452–455 (2013).
223. Park, T. J., Haigo, S. L. & Wallingford, J. B. Ciliogenesis defects in embryos lacking inturned or fuzzy function are associated with failure of planar cell polarity and Hedgehog signaling. *Nat. Genet.* **38**, 303–311 (2006).
224. Gray, R. S. *et al.* The planar cell polarity effector Fuz is essential for targeted membrane trafficking, ciliogenesis and mouse embryonic development. *Nat. Cell Biol.* **11**, 1225–1232 (2009).
225. Seo, J. H. *et al.* Mutations in the planar cell polarity gene, Fuzzy, are associated with neural tube defects in humans. *Hum. Mol. Genet.* **20**, 4324–4333 (2011).
226. Bosoi, C. M. *et al.* Identification and characterization of novel rare mutations in the planar cell polarity gene PRICKLE1 in human neural tube defects. *Hum. Mutat.* **32**, 1371–1375 (2011).
227. Curtin, J. A. *et al.* Mutation of *Celsr1* disrupts planar polarity of inner ear hair cells and causes severe neural tube defects in the mouse. *Curr. Biol. CB* **13**, 1129–1133 (2003).
228. Allache, R., De Marco, P., Merello, E., Capra, V. & Kibar, Z. Role of the planar cell polarity gene CELSR1 in neural tube defects and caudal agenesis. *Birt. Defects Res. A. Clin. Mol. Teratol.* **94**, 176–181 (2012).
229. Lu, X. *et al.* PTK7/CCK-4 is a novel regulator of planar cell polarity in vertebrates. *Nature* **430**, 93–98 (2004).
230. Peradziryi, H. *et al.* PTK7/Otk interacts with Wnts and inhibits canonical Wnt signalling. *EMBO J.* **30**, 3729–3740 (2011).
231. Hayes, M., Naito, M., Daulat, A., Angers, S. & Ciruna, B. Ptk7 promotes non-canonical Wnt/PCP-mediated morphogenesis and inhibits Wnt/ β -catenin-dependent cell fate decisions during vertebrate development. *Dev. Camb. Engl.* **140**, 1807–1818 (2013).
232. Wang, M. *et al.* Role of the planar cell polarity gene Protein tyrosine kinase 7 in neural tube defects in humans. *Birt. Defects Res. A. Clin. Mol. Teratol.* **103**, 1021–1027 (2015).
233. Allache, R. *et al.* Genetic studies of ANKRD6 as a molecular switch between Wnt signaling pathways in human neural tube defects. *Birt. Defects Res. A. Clin. Mol. Teratol.* **103**, 20–26 (2015).
234. Schwarz-Romond, T. *et al.* The ankyrin repeat protein Diversin recruits Casein kinase Iepsilon to the beta-catenin degradation complex and acts in both canonical Wnt and Wnt/JNK signaling. *Genes Dev.* **16**, 2073–2084 (2002).
235. Allache, R. *et al.* Novel mutations in Lrp6 orthologs in mouse and human neural tube defects affect a highly dosage-sensitive Wnt non-canonical planar cell polarity pathway. *Hum. Mol. Genet.* **23**, 1687–1699 (2014).
236. Lei, Y. *et al.* Rare LRP6 Variants Identified in Spina Bifida Patients. *Hum. Mutat.* **36**, 342–349 (2015).

237. Zhao, Q., Behringer, R. R. & de Crombrughe, B. Prenatal folic acid treatment suppresses acrania and meroanencephaly in mice mutant for the *Cart1* homeobox gene. *Nat. Genet.* **13**, 275–283 (1996).
238. Spiegelstein, O. *et al.* Embryonic development of folate binding protein-1 (*Folbp1*) knockout mice: Effects of the chemical form, dose, and timing of maternal folate supplementation. *Dev. Dyn. Off. Publ. Am. Assoc. Anat.* **231**, 221–231 (2004).
239. Lin, W. *et al.* Proper expression of the *Gcn5* histone acetyltransferase is required for neural tube closure in mouse embryos. *Dev. Dyn. Off. Publ. Am. Assoc. Anat.* **237**, 928–940 (2008).
240. Carter, M., Ulrich, S., Oofuji, Y., Williams, D. A. & Elizabeth Ross, M. Crooked Tail(*Cd*) Models Human Folate-Responsive Neural Tube Defects. *Hum. Mol. Genet.* **8**, 2199–2204 (1999).
241. Fleming, A. & Copp, A. J. Embryonic Folate Metabolism and Mouse Neural Tube Defects. *Science* **280**, 2107–2109 (1998).
242. Essien, F. B. Maternal methionine supplementation promotes the remediation of axial defects in *Axd* mouse neural tube mutants. *Teratology* **45**, 205–212 (1992).
243. Juriloff, D. M. *et al.* Multifactorial genetics of exencephaly in *SELH/Bc* mice. *Teratology* **64**, 189–200 (2001).
244. Koleske, A. J. *et al.* Essential roles for the *Abl* and *Arg* tyrosine kinases in neurulation. *Neuron* **21**, 1259–1272 (1998).
245. Gurniak, C. B., Perlas, E. & Witke, W. The actin depolymerizing factor n-cofilin is essential for neural tube morphogenesis and neural crest cell migration. *Dev. Biol.* **278**, 231–241 (2005).
246. Brouns, M. R. *et al.* The adhesion signaling molecule p190 RhoGAP is required for morphogenetic processes in neural development. *Dev. Camb. Engl.* **127**, 4891–4903 (2000).
247. Sabapathy, K. *et al.* Defective neural tube morphogenesis and altered apoptosis in the absence of both *JNK1* and *JNK2*. *Mech. Dev.* **89**, 115–124 (1999).
248. Stumpo, D. J., Bock, C. B., Tuttle, J. S. & Blackshear, P. J. *MARCKS* deficiency in mice leads to abnormal brain development and perinatal death. *Proc. Natl. Acad. Sci. U. S. A.* **92**, 944–948 (1995).
249. Lanier, L. M. *et al.* *Mena* is required for neurulation and commissure formation. *Neuron* **22**, 313–325 (1999).
250. Menzies, A. S. *et al.* *Mena* and vasodilator-stimulated phosphoprotein are required for multiple actin-dependent processes that shape the vertebrate nervous system. *J. Neurosci. Off. J. Soc. Neurosci.* **24**, 8029–8038 (2004).
251. Wu, M., Chen, D. F., Sasaoka, T. & Tonegawa, S. Neural tube defects and abnormal brain development in *F52*-deficient mice. *Proc. Natl. Acad. Sci. U. S. A.* **93**, 2110–2115 (1996).
252. Luo, H. *et al.* Disruption of palladin results in neural tube closure defects in mice. *Mol. Cell. Neurosci.* **29**, 507–515 (2005).
253. Hildebrand, J. D. & Soriano, P. *Shroom*, a PDZ domain-containing actin-binding protein, is required for neural tube morphogenesis in mice. *Cell* **99**, 485–497 (1999).
254. Xu, W., Baribault, H. & Adamson, E. D. Vinculin knockout results in heart and brain defects during embryonic development. *Dev. Camb. Engl.* **125**, 327–337 (1998).
255. Shang, E., Wang, X., Wen, D., Greenberg, D. A. & Wolgemuth, D. J. Double bromodomain-containing gene *Brd2* is essential for embryonic development in mouse. *Dev. Dyn. Off. Publ. Am. Assoc. Anat.* **238**, 908–917 (2009).
256. Ishibashi, M. *et al.* Targeted disruption of mammalian hairy and Enhancer of split homolog-1 (*HES-1*) leads to up-regulation of neural helix-loop-helix factors, premature neurogenesis, and severe neural tube defects. *Genes Dev.* **9**, 3136–3148 (1995).
257. Hirata, H., Tomita, K., Bessho, Y. & Kageyama, R. *Hes1* and *Hes3* regulate maintenance of the isthmus organizer and development of the mid/hindbrain. *EMBO J.* **20**, 4454–4466 (2001).
258. Takeuchi, T. *et al.* Gene trap capture of a novel mouse gene, *jumonji*, required for neural tube formation. *Genes Dev.* **9**, 1211–1222 (1995).

259. Lakkis, M. M., Golden, J. A., O'Shea, K. S. & Epstein, J. A. Neurofibromin deficiency in mice causes exencephaly and is a modifier for Splotch neural tube defects. *Dev. Biol.* **212**, 80–92 (1999).
260. Lardelli, M., Williams, R., Mitsiadis, T. & Lendahl, U. Expression of the Notch 3 intracellular domain in mouse central nervous system progenitor cells is lethal and leads to disturbed neural tube development. *Mech. Dev.* **59**, 177–190 (1996).
261. Zhong, W. *et al.* Mouse numb is an essential gene involved in cortical neurogenesis. *Proc. Natl. Acad. Sci. U. S. A.* **97**, 6844–6849 (2000).
262. Smitherman, M., Lee, K., Swanger, J., Kapur, R. & Clurman, B. E. Characterization and targeted disruption of murine Nup50, a p27(Kip1)-interacting component of the nuclear pore complex. *Mol. Cell. Biol.* **20**, 5631–5642 (2000).
263. Epstein, D. J., Vekemans, M. & Gros, P. Splotch (Sp2H), a mutation affecting development of the mouse neural tube, shows a deletion within the paired homeodomain of Pax-3. *Cell* **67**, 767–774 (1991).
264. Mansouri, A. & Gruss, P. Pax3 and Pax7 are expressed in commissural neurons and restrict ventral neuronal identity in the spinal cord. *Mech. Dev.* **78**, 171–178 (1998).
265. Oka, C. *et al.* Disruption of the mouse RBP-J kappa gene results in early embryonic death. *Dev. Camb. Engl.* **121**, 3291–3301 (1995).
266. Uetani, N. *et al.* Maturation of ureter-bladder connection in mice is controlled by LAR family receptor protein tyrosine phosphatases. *J. Clin. Invest.* **119**, 924–935 (2009).
267. Schorle, H., Meier, P., Buchert, M., Jaenisch, R. & Mitchell, P. J. Transcription factor AP-2 essential for cranial closure and craniofacial development. *Nature* **381**, 235–238 (1996).
268. Cecconi, F., Alvarez-Bolado, G., Meyer, B. I., Roth, K. A. & Gruss, P. Apaf1 (CED-4 homolog) regulates programmed cell death in mammalian development. *Cell* **94**, 727–737 (1998).
269. Ruland, J. *et al.* Bcl10 is a positive regulator of antigen receptor-induced activation of NF-kappaB and neural tube closure. *Cell* **104**, 33–42 (2001).
270. Gowen, L. C., Johnson, B. L., Latour, A. M., Sulik, K. K. & Koller, B. H. Brca1 deficiency results in early embryonic lethality characterized by neuroepithelial abnormalities. *Nat. Genet.* **12**, 191–194 (1996).
271. Wang, X. *et al.* Genetic interactions between Brca1 and Gadd45a in centrosome duplication, genetic stability, and neural tube closure. *J. Biol. Chem.* **279**, 29606–29614 (2004).
272. Leonard, J. R., Klocke, B. J., D'Sa, C., Flavell, R. A. & Roth, K. A. Strain-dependent neurodevelopmental abnormalities in caspase-3-deficient mice. *J. Neuropathol. Exp. Neurol.* **61**, 673–677 (2002).
273. Kuida, K. *et al.* Reduced apoptosis and cytochrome c-mediated caspase activation in mice lacking caspase 9. *Cell* **94**, 325–337 (1998).
274. Li, Q., Estepa, G., Memet, S., Israel, A. & Verma, I. M. Complete lack of NF-kappaB activity in IKK1 and IKK2 double-deficient mice: additional defect in neurulation. *Genes Dev.* **14**, 1729–1733 (2000).
275. Chi, H., Sarkisian, M. R., Rakic, P. & Flavell, R. A. Loss of mitogen-activated protein kinase kinase 4 (MEKK4) results in enhanced apoptosis and defective neural tube development. *Proc. Natl. Acad. Sci. U. S. A.* **102**, 3846–3851 (2005).
276. Dixon, J., Brakebusch, C., Fässler, R. & Dixon, M. J. Increased levels of apoptosis in the prefusion neural folds underlie the craniofacial disorder, Treacher Collins syndrome. *Hum. Mol. Genet.* **9**, 1473–1480 (2000).
277. Nonn, L., Williams, R. R., Erickson, R. P. & Powis, G. The Absence of Mitochondrial Thioredoxin 2 Causes Massive Apoptosis, Exencephaly, and Early Embryonic Lethality in Homozygous Mice. *Mol. Cell. Biol.* **23**, 916–922 (2003).
278. Hosako, H. *et al.* Gene and microRNA expression in p53-deficient day 8.5 mouse embryos. *Birt. Defects Res. A. Clin. Mol. Teratol.* **85**, 546–555 (2009).

279. Berry, R. J. *et al.* Prevention of Neural-Tube Defects with Folic Acid in China. *N. Engl. J. Med.* **341**, 1485–1490 (1999).
280. Ray, J. G., Vermeulen, M. J., Boss, S. C. & Cole, D. E. C. Increased red cell folate concentrations in women of reproductive age after Canadian folic acid food fortification. *Epidemiol. Camb. Mass* **13**, 238–240 (2002).
281. Morin, P. *et al.* Pregnancy planning and folic acid supplement use: results from a survey in Quebec. *Prev. Med.* **35**, 143–149 (2002).
282. Worley, G., Rosenfeld, L. R. & Lipscomb, J. Financial counseling for families of children with chronic disabilities. *Dev. Med. Child Neurol.* **33**, 679–689 (1991).
283. Yi, Y., Lindemann, M., Colligs, A. & Snowball, C. Economic burden of neural tube defects and impact of prevention with folic acid: a literature review. *Eur. J. Pediatr.* **170**, 1391–1400 (2011).
284. Khalili, H. *et al.* Developmental Origins for Kidney Disease Due to Shroom3 Deficiency. *J. Am. Soc. Nephrol.* **27**, 2965–2973 (2016).
285. Leslie, E. J. *et al.* Genome-wide meta-analyses of nonsyndromic orofacial clefts identify novel associations between FOXE1 and all orofacial clefts, and TP63 and cleft lip with or without cleft palate. *Hum. Genet.* **136**, 275–286 (2017).
286. Zhang, Y.-B. *et al.* Genome-wide association study identifies multiple susceptibility loci for craniofacial microsomia. *Nat. Commun.* **7**, 10605 (2016).
287. Ellis, J. W. *et al.* Validated SNPs for eGFR and their associations with albuminuria. *Hum. Mol. Genet.* **21**, 3293–3298 (2012).
288. Köttgen, A. *et al.* Multiple loci associated with indices of renal function and chronic kidney disease. *Nat. Genet.* **41**, 712–717 (2009).
289. Simino, J., Rao, D. C. & Freedman, B. I. Novel findings and future directions on the genetics of hypertension. *Curr. Opin. Nephrol. Hypertens.* **21**, 500–507 (2012).
290. the CHARGE Consortium Hematology Working Group. Meta-analysis of rare and common exome chip variants identifies S1PR4 and other loci influencing blood cell traits. *Nat. Genet.* **48**, 867–876 (2016).
291. Dickson, H. M., Wilbur, A., Reinke, A. A., Young, M. A. & Vojtek, A. B. Targeted inhibition of the Shroom3-Rho kinase protein-protein interaction circumvents Nogo66 to promote axon outgrowth. *BMC Neurosci.* **16**, 34 (2015).
292. Taylor, J. *et al.* The scaffold protein POSH regulates axon outgrowth. *Mol. Biol. Cell* **19**, 5181–5192 (2008).
293. Tariq, M., Belmont, J. W., Lalani, S., Smolarek, T. & Ware, S. M. SHROOM3 is a novel candidate for heterotaxy identified by whole exome sequencing. *Genome Biol.* **12**, R91 (2011).
294. Schoenwolf, G. C. Cell movements driving neurulation in avian embryos. *Development* **113**, 157–168 (1991).
295. Lee, C., Le, M.-P. & Wallingford, J. B. The shroom family proteins play broad roles in the morphogenesis of thickened epithelial sheets. *Dev. Dyn.* **238**, 1480–1491 (2009).
296. Plageman, T. F. *et al.* Pax6-dependent Shroom3 expression regulates apical constriction during lens placode invagination. *Development* **137**, 405–415 (2010).
297. Riento, K. & Ridley, A. J. Rocks: multifunctional kinases in cell behaviour. *Nat. Rev. Mol. Cell Biol.* **4**, 446–456 (2003).
298. Dietz, M. L., Bernaciak, T. M., Vendetti, F., Kielec, J. M. & Hildebrand, J. D. Differential Actin-dependent Localization Modulates the Evolutionarily Conserved Activity of Shroom Family Proteins. *J. Biol. Chem.* **281**, 20542–20554 (2006).
299. Lee, C., Scherr, H. M. & Wallingford, J. B. Shroom family proteins regulate γ -tubulin distribution and microtubule architecture during epithelial cell shape change. *Development* **134**, 1431–1441 (2007).

300. Matsumura, F. Regulation of myosin II during cytokinesis in higher eukaryotes. *Trends Cell Biol.* **15**, 371–377 (2005).
301. McGreevy, E. M., Vijayraghavan, D., Davidson, L. A. & Hildebrand, J. D. Shroom3 functions downstream of planar cell polarity to regulate myosin II distribution and cellular organization during neural tube closure. *Biol. Open* **4**, 186–196 (2015).
302. Ossipova, O., Chuykin, I., Chu, C.-W. & Sokol, S. Y. Vangl2 cooperates with Rab11 and Myosin V to regulate apical constriction during vertebrate gastrulation. *Development* **142**, 99–107 (2015).
303. Chung, M.-I., Nascone-Yoder, N. M., Grover, S. A., Drysdale, T. A. & Wallingford, J. B. Direct activation of Shroom3 transcription by Pitx proteins drives epithelial morphogenesis in the developing gut. *Development* **137**, 1339–1349 (2010).
304. Loebel, D. A. F. *et al.* Thyroid bud morphogenesis requires CDC42- and SHROOM3-dependent apical constriction. *Biol. Open* **5**, 130–139 (2016).
305. Farber, M. J., Rizaldy, R. & Hildebrand, J. D. Shroom2 regulates contractility to control endothelial morphogenesis. *Mol. Biol. Cell* **22**, 795–805 (2011).
306. Fairbank, P. D. *et al.* Shroom2 (APXL) regulates melanosome biogenesis and localization in the retinal pigment epithelium. *Dev. Camb. Engl.* **133**, 4109–4118 (2006).
307. Piton, A., Redin, C. & Mandel, J.-L. XLID-causing mutations and associated genes challenged in light of data from large-scale human exome sequencing. *Am. J. Hum. Genet.* **93**, 368–383 (2013).
308. Ting, S. B. *et al.* The identification and characterization of human Sister-of-Mammalian Grainyhead (SOM) expands the grainyhead-like family of developmental transcription factors. *Biochem. J.* **370**, 953–962 (2003).
309. Goldie, S. J. *et al.* Mice lacking the conserved transcription factor Grainyhead-like 3 (Grhl3) display increased apposition of the frontal and parietal bones during embryonic development. *BMC Dev. Biol.* **16**, 37 (2016).
310. Castro, S. C. P. D. *et al.* Lamin B1 Polymorphism Influences Morphology of the Nuclear Envelope, Cell Cycle Progression, and Risk of Neural Tube Defects in Mice. *PLOS Genet.* **8**, e1003059 (2012).
311. Ting, S. B. *et al.* The Epidermis of Grhl3-Null Mice Displays Altered Lipid Processing and Cellular Hyperproliferation. *Organogenesis* **2**, 33–35 (2005).
312. Yu, Z., Mannik, J., Soto, A., Lin, K. K. & Andersen, B. The epidermal differentiation-associated Grainyhead gene *Get1/Grhl3* also regulates urothelial differentiation. *EMBO J.* **28**, 1890–1903 (2009).
313. Peyrard-Janvid, M. *et al.* Dominant Mutations in GRHL3 Cause Van der Woude Syndrome and Disrupt Oral Periderm Development. *Am. J. Hum. Genet.* **94**, 23–32 (2014).
314. Leslie, E. J. *et al.* A Genome-wide Association Study of Nonsyndromic Cleft Palate Identifies an Etiologic Missense Variant in GRHL3. *Am. J. Hum. Genet.* **98**, 744–754 (2016).
315. Dworkin, S. *et al.* Grainyhead-like 3 (Grhl3) deficiency in brain leads to altered locomotor activity and decreased anxiety-like behaviors in aged mice. *Dev. Neurobiol.* n/a-n/a (2017). doi:10.1002/dneu.22469
316. Auden, A. *et al.* Spatial and temporal expression of the Grainyhead-like transcription factor family during murine development. *Gene Expr. Patterns* **6**, 964–970 (2006).
317. Ting, S. B. *et al.* A Homolog of *Drosophila* grainy head Is Essential for Epidermal Integrity in Mice. *Science* **308**, 411–413 (2005).
318. Yu, Z. *et al.* The Grainyhead-like epithelial transactivator *Get-1/Grhl3* regulates epidermal terminal differentiation and interacts functionally with LMO4. *Dev. Biol.* **299**, 122–136 (2006).
319. Hislop, N. R. *et al.* Grhl3 and Lmo4 play coordinate roles in epidermal migration. *Dev. Biol.* **321**, 263–272 (2008).
320. Yu, Z. *et al.* Grainyhead-like factor *Get1/Grhl3* regulates formation of the epidermal leading edge during eyelid closure. *Dev. Biol.* **319**, 56–67 (2008).

321. de la Garza, G. *et al.* Interferon Regulatory Factor 6 Promotes Differentiation of the Periderm by Activating Expression of Grainyhead-Like 3. *J. Invest. Dermatol.* **133**, 68–77 (2013).
322. Dworkin, S. *et al.* Grainyhead-like 3 regulation of endothelin-1 in the pharyngeal endoderm is critical for growth and development of the craniofacial skeleton. *Mech. Dev.* **133**, 77–90 (2014).
323. Haendeler, J. *et al.* Two Isoforms of Sister-of-Mammalian Grainyhead Have Opposing Functions in Endothelial Cells and In Vivo. *Arterioscler. Thromb. Vasc. Biol.* **33**, 1639–1646 (2013).
324. Lukosz, M., Mlynek, A., Czypiorski, P., Altschmied, J. & Haendeler, J. The transcription factor Grainyhead like 3 (GRHL3) affects endothelial cell apoptosis and migration in a NO-dependent manner. *Biochem. Biophys. Res. Commun.* **412**, 648–653 (2011).
325. Caddy, J. *et al.* Epidermal Wound Repair Is Regulated by the Planar Cell Polarity Signaling Pathway. *Dev. Cell* **19**, 138–147 (2010).
326. Girolamo, D. D. *et al.* Reciprocal interplay between thyroid hormone and microRNA-21 regulates hedgehog pathway–driven skin tumorigenesis. *J. Clin. Invest.* **126**, 2308–2320 (2016).
327. Pourhoseini, S. *et al.* Upregulation of miR21 and Repression of Grhl3 by Leptin Mediates Sinusoidal Endothelial Injury in Experimental Nonalcoholic Steatohepatitis. *PLOS ONE* **10**, e0116780 (2015).
328. Wilanowski, T. *et al.* Perturbed desmosomal cadherin expression in grainy head-like 1-null mice. *EMBO J.* **27**, 886–897 (2008).
329. Mlacki, M., Darido, C., Jane, S. M. & Wilanowski, T. Loss of Grainy Head-Like 1 Is Associated with Disruption of the Epidermal Barrier and Squamous Cell Carcinoma of the Skin. *PLOS ONE* **9**, e89247 (2014).
330. Senga, K., Mostov, K. E., Mitaka, T., Miyajima, A. & Tanimizu, N. Grainyhead-like 2 regulates epithelial morphogenesis by establishing functional tight junctions through the organization of a molecular network among claudin3, claudin4, and Rab25. *Mol. Biol. Cell* **23**, 2845–2855 (2012).
331. Aue, A. *et al.* A Grainyhead-Like 2/Ovo-Like 2 Pathway Regulates Renal Epithelial Barrier Function and Lumen Expansion. *J. Am. Soc. Nephrol. JASN* **26**, 2704–2715 (2015).
332. Gao, X. *et al.* Evidence for multiple roles for grainyhead-like 2 in the establishment and maintenance of human mucociliary airway epithelium.[corrected]. *Proc. Natl. Acad. Sci. U. S. A.* **110**, 9356–9361 (2013).
333. Walentin, K. *et al.* A Grhl2-dependent gene network controls trophoblast branching morphogenesis. *Dev. Camb. Engl.* **142**, 1125–1136 (2015).
334. Kohn, K. W., Zeeberg, B. M., Reinhold, W. C. & Pommier, Y. Gene expression correlations in human cancer cell lines define molecular interaction networks for epithelial phenotype. *PLoS One* **9**, e99269 (2014).
335. Boglev, Y. *et al.* The unique and cooperative roles of the Grainy head-like transcription factors in epidermal development reflect unexpected target gene specificity. *Dev. Biol.* **349**, 512–522 (2011).
336. Beutler, B. Immunology, phenotype first. Preface. *Curr. Top. Microbiol. Immunol.* **321**, v–viii (2008).
337. Keane, T. M. *et al.* Mouse genomic variation and its effect on phenotypes and gene regulation. *Nature* **477**, 289–294 (2011).
338. Russell, W. L. *et al.* Specific-locus test shows ethylnitrosourea to be the most potent mutagen in the mouse. *Proc. Natl. Acad. Sci. U. S. A.* **76**, 5818–5819 (1979).
339. Austin, C. P. *et al.* The knockout mouse project. *Nat. Genet.* **36**, 921–924 (2004).
340. Russell, W. L. & Kelly, E. M. Specific-locus mutation frequencies in mouse stem-cell spermatogonia at very low radiation dose rates. *Proc. Natl. Acad. Sci. U. S. A.* **79**, 539–541 (1982).
341. Hardy, S., Legagneux, V., Audic, Y. & Paillard, L. Reverse genetics in eukaryotes. *Biol. Cell* **102**, 561–580 (2010).

342. International HapMap Consortium. The International HapMap Project. *Nature* **426**, 789–796 (2003).
343. Wang, D. G. *et al.* Large-scale identification, mapping, and genotyping of single-nucleotide polymorphisms in the human genome. *Science* **280**, 1077–1082 (1998).
344. Witte, J. S. Genome-Wide Association Studies and Beyond. *Annu. Rev. Public Health* **31**, 9–20 (2010).
345. Hodge, S. E. Linkage analysis versus association analysis: distinguishing between two models that explain disease-marker associations. *Am. J. Hum. Genet.* **53**, 367–384 (1993).
346. Albert, T. J. *et al.* Direct selection of human genomic loci by microarray hybridization. *Nat. Methods* **4**, 903–905 (2007).
347. Warr, A. *et al.* Exome Sequencing: Current and Future Perspectives. *G3 GenesGenomesGenetics* **5**, 1543–1550 (2015).
348. Guo, Y., Ding, X., Shen, Y., Lyon, G. J. & Wang, K. SeqMule: automated pipeline for analysis of human exome/genome sequencing data. *Sci. Rep.* **5**, 14283 (2015).
349. Krøigård, A. B., Thomassen, M., Lænkholm, A.-V., Kruse, T. A. & Larsen, M. J. Evaluation of Nine Somatic Variant Callers for Detection of Somatic Mutations in Exome and Targeted Deep Sequencing Data. *PLoS ONE* **11**, (2016).
350. Rabbani, B., Tekin, M. & Mahdieh, N. The promise of whole-exome sequencing in medical genetics. *J. Hum. Genet.* **59**, 5–15 (2014).
351. Qu, H. & Fang, X. A brief review on the Human Encyclopedia of DNA Elements (ENCODE) project. *Genomics Proteomics Bioinformatics* **11**, 135–141 (2013).
352. Adzhubei, I. A. *et al.* A method and server for predicting damaging missense mutations. *Nat. Methods* **7**, 248–249 (2010).
353. Kumar, P., Henikoff, S. & Ng, P. C. Predicting the effects of coding non-synonymous variants on protein function using the SIFT algorithm. *Nat. Protoc.* **4**, 1073–1081 (2009).
354. Kircher, M. *et al.* A general framework for estimating the relative pathogenicity of human genetic variants. *Nat. Genet.* **46**, 310–315 (2014).
355. Consortium, E. A. *et al.* Analysis of protein-coding genetic variation in 60,706 humans. *bioRxiv* 030338 (2016). doi:10.1101/030338
356. Exome Variant Server. Available at: <http://evs.gs.washington.edu/EVS/>. (Accessed: 19th April 2017)
357. Samocha, K. E. *et al.* A framework for the interpretation of de novo mutation in human disease. *Nat. Genet.* **46**, 944–950 (2014).
358. Iossifov, I. *et al.* The contribution of de novo coding mutations to autism spectrum disorder. *Nature* **515**, 216–221 (2014).
359. Wu, M. C. *et al.* Rare-Variant Association Testing for Sequencing Data with the Sequence Kernel Association Test. *Am. J. Hum. Genet.* **89**, 82–93 (2011).
360. Yuen, R. K. *et al.* Genome-wide characteristics of de novo mutations in autism. *Npj Genomic Med.* **1**, 16027 (2016).
361. Fromer, M. *et al.* De novo mutations in schizophrenia implicate synaptic networks. *Nature* **506**, 179–184 (2014).
362. Parker, M. J. *et al.* De novo, heterozygous, loss-of-function mutations in SYNGAP1 cause a syndromic form of intellectual disability. *Am. J. Med. Genet. A.* **167A**, 2231–2237 (2015).
363. Acuna-Hidalgo, R., Veltman, J. A. & Hoischen, A. New insights into the generation and role of de novo mutations in health and disease. *Genome Biol.* **17**, 241 (2016).
364. Griffin, H. R. *et al.* Accurate mitochondrial DNA sequencing using off-target reads provides a single test to identify pathogenic point mutations. *Genet. Med.* **16**, 962–971 (2014).

365. Tan, R. *et al.* An evaluation of copy number variation detection tools from whole-exome sequencing data. *Hum. Mutat.* **35**, 899–907 (2014).
366. Parla, J. S. *et al.* A comparative analysis of exome capture. *Genome Biol.* **12**, R97 (2011).
367. Meienberg, J., Bruggmann, R., Oexle, K. & Matyas, G. Clinical sequencing: is WGS the better WES? *Hum. Genet.* **135**, 359–362 (2016).
368. NimbleGen SeqCap Target Enrichment. Available at: <http://sequencing.roche.com/products/nimblegen-seqcap-target-enrichment.html>. (Accessed: 24th May 2017)
369. Targeted Gene Sequencing | Focused panels targeting key genes. Available at: <https://www.illumina.com/techniques/sequencing/dna-sequencing/targeted-resequencing/targeted-panels.html>. (Accessed: 24th May 2017)
370. O’Roak, B. J. *et al.* Multiplex targeted sequencing identifies recurrently mutated genes in autism spectrum disorders. *Science* **338**, 1619–1622 (2012).
371. Garand, L., Lingler, J. H., Conner, K. O. & Dew, M. A. Diagnostic Labels, Stigma, and Participation in Research Related to Dementia and Mild Cognitive Impairment. *Res. Gerontol. Nurs.* **2**, 112–121 (2009).
372. Joó, J. G. *et al.* Neural tube defects in the sample of genetic counselling. *Prenat. Diagn.* **27**, 912–921 (2007).
373. Lemay, P. *et al.* Loss-of-function de novo mutations play an important role in severe human neural tube defects. *J. Med. Genet.* **52**, 493–497 (2015).
374. Lemay, P. *et al.* Rare deleterious variants in GRHL3 are associated with human spina bifida. *Hum. Mutat.* (2017). doi:10.1002/humu.23214
375. Lemay, P. *et al.* Quantitative Trait Loci (QTL) Study Identifies Novel Genomic Regions Associated to Chiari-Like Malformation in Griffon Bruxellois Dogs. *PLOS ONE* **9**, e89816 (2014).
376. Benedet, A. L. *et al.* Epistasis analysis links immune cascades and cerebral amyloidosis. *J. Neuroinflammation* **12**, 227 (2015).
377. Hamdan, F. F. *et al.* High Rate of Recurrent De Novo Mutations in Developmental and Epileptic Encephalopathies. *Am. J. Hum. Genet.* **101**, 664–685 (2017).
378. D’Aurizio, R. *et al.* Enhanced copy number variants detection from whole-exome sequencing data using EXCAVATOR2. *Nucleic Acids Res.* **44**, e154–e154 (2016).
379. Siegel-Bartlett, J., Levin, A., Teebi, A. S. & Kennedy, S. J. Craniofacial anomalies, cataracts, congenital heart disease, sacral neural tube defects, and growth and developmental retardation in two sisters: a new autosomal recessive MCA/MR syndrome? *J. Med. Genet.* **39**, 145–148 (2002).
380. Finnell, R. H., Greer, K. A., Barber, R. C. & Piedrahita, J. A. Neural tube and craniofacial defects with special emphasis on folate pathway genes. *Crit. Rev. Oral Biol. Med. Off. Publ. Am. Assoc. Oral Biol.* **9**, 38–53 (1998).
381. Prasad, C. P., Chaurasiya, S. K., Axelsson, L. & Andersson, T. WNT-5A triggers Cdc42 activation leading to an ERK1/2 dependent decrease in MMP9 activity and invasive migration of breast cancer cells. *Mol. Oncol.* **7**, 870–883 (2013).
382. Schlessinger, K., McManus, E. J. & Hall, A. Cdc42 and noncanonical Wnt signal transduction pathways cooperate to promote cell polarity. *J. Cell Biol.* **178**, 355–361 (2007).
383. Liu, Y. *et al.* Inactivation of Cdc42 in neural crest cells causes craniofacial and cardiovascular morphogenesis defects. *Dev. Biol.* **383**, 239–252 (2013).
384. PIANO Mortari, E. *et al.* Heterotaxy syndrome with and without spleen: Different infection risk and management. *J. Allergy Clin. Immunol.* (2016). doi:10.1016/j.jaci.2016.10.014
385. Marcil, A., Dumontier, É., Chamberland, M., Camper, S. A. & Drouin, J. Pitx1 and Pitx2 are required for development of hindlimb buds. *Development* **130**, 45–55 (2003).

386. Haigo, S. L., Hildebrand, J. D., Harland, R. M. & Wallingford, J. B. Shroom Induces Apical Constriction and Is Required for Hinge-point Formation during Neural Tube Closure. *Curr. Biol.* **13**, 2125–2137 (2003).
387. Leduc, R. Y. M., Singh, P. & McDermid, H. E. Genetic backgrounds and modifier genes of NTD mouse models: An opportunity for greater understanding of the multifactorial etiology of neural tube defects. *Birth Defects Res.* **109**, 140–152 (2017).
388. Haaning, A. M., Quinn, M. E. & Ware, S. M. Heterotaxy-spectrum heart defects in *Zic3* hypomorphic mice. *Pediatr. Res.* **74**, 494–502 (2013).
389. Zhu, L., Belmont, J. W. & Ware, S. M. Genetics of human heterotaxias. *Eur. J. Hum. Genet.* **14**, 17–25 (2005).
390. Costa, L. C. M. C. *et al.* Expression of epithelial-mesenchymal transition markers at the invasive front of oral squamous cell carcinoma. *J. Appl. Oral Sci.* **23**, 169–178 (2015).
391. Dudas, M., Li, W.-Y., Kim, J., Yang, A. & Kaartinen, V. Palatal fusion – Where do the midline cells go?: A review on cleft palate, a major human birth defect. *Acta Histochem.* **109**, 1–14 (2007).
392. Kumar, R. A. Two-hit wonder: a novel genetic model to explain variable expressivity in severe pediatric phenotypes. *Clin. Genet.* **78**, 517–519 (2010).
393. Nadeau, J. H. Modifier genes in mice and humans. *Nat. Rev. Genet.* **2**, 165–174 (2001).
394. Kousa, Y. A., Mansour, T. A., Seada, H., Matoo, S. & Schutte, B. C. Shared molecular networks in orofacial and neural tube development. *Birt. Defects Res. A. Clin. Mol. Teratol.* (2016). doi:10.1002/bdra.23598
395. Topczewski, J., Dale, R. M. & Sisson, B. E. Planar cell polarity signaling in craniofacial development. *Organogenesis* **7**, 255–259 (2011).
396. Ewart, J. L. *et al.* Heart and neural tube defects in transgenic mice overexpressing the Cx43 gap junction gene. *Dev. Camb. Engl.* **124**, 1281–1292 (1997).
397. Morris-Kay, G. & Tan, S.-S. Mapping cranial neural crest cell migration pathways in mammalian embryos. *Trends Genet.* **3**, 257–261 (1987).
398. Copp, A. J., Greene, N. D. E. & Murdoch, J. N. The genetic basis of mammalian neurulation. *Nat. Rev. Genet.* **4**, 784–793 (2003).
399. Yu, H. *et al.* Frizzled 1 and frizzled 2 genes function in palate, ventricular septum and neural tube closure: general implications for tissue fusion processes. *Dev. Camb. Engl.* **137**, 3707–3717 (2010).
400. Han, J. *et al.* Indirect modulation of Shh signaling by *Dlx5* affects the oral-nasal patterning of palate and rescues cleft palate in *Msx1*-null mice. *Dev. Camb. Engl.* **136**, 4225–4233 (2009).
401. Klein, R. H. *et al.* GRHL3 binding and enhancers rearrange as epidermal keratinocytes transition between functional states. *PLOS Genet.* **13**, e1006745 (2017).
402. Guardiola-Serrano, F. *et al.* Gene trapping identifies a putative tumor suppressor and a new inducer of cell migration. *Biochem. Biophys. Res. Commun.* **376**, 748–752 (2008).
403. Haendeler, J. *et al.* Two isoforms of Sister-Of-Mammalian Grainyhead have opposing functions in endothelial cells and in vivo. *Arterioscler. Thromb. Vasc. Biol.* **33**, 1639–1646 (2013).
404. Katayama, K. *et al.* Loss of RhoA in neural progenitor cells causes the disruption of adherens junctions and hyperproliferation. *Proc. Natl. Acad. Sci. U. S. A.* **108**, 7607–7612 (2011).
405. Greene, N. D., Gerrelli, D., Van Straaten, H. W. & Copp, A. J. Abnormalities of floor plate, notochord and somite differentiation in the loop-tail (Lp) mouse: a model of severe neural tube defects. *Mech. Dev.* **73**, 59–72 (1998).
406. Murdoch, J. N. & Copp, A. J. The relationship between sonic Hedgehog signaling, cilia, and neural tube defects. *Birt. Defects Res. A. Clin. Mol. Teratol.* **88**, 633–652 (2010).
407. Rifat, Y. *et al.* Regional neural tube closure defined by the Grainy head-like transcription factors. *Dev. Biol.* **345**, 237–245 (2010).
408. Gurdon, J. B. & Bourillot, P.-Y. Morphogen gradient interpretation. *Nature* **413**, 797–803 (2001).

409. Alborzinia, H. *et al.* Quantitative kinetics analysis of BMP2 uptake into cells and its modulation by BMP antagonists. *J Cell Sci* **126**, 117–127 (2013).
410. Girard, S. L. *et al.* Increased exonic *de novo* mutation rate in individuals with schizophrenia. *Nat. Genet.* **43**, 860 (2011).
411. Awadalla, P. *et al.* Direct Measure of the De Novo Mutation Rate in Autism and Schizophrenia Cohorts. *Am. J. Hum. Genet.* **87**, 316–324 (2010).
412. Lelieveld, S. H. *et al.* Meta-analysis of 2,104 trios provides support for 10 new genes for intellectual disability. *Nat. Neurosci.* **19**, 1194–1196 (2016).
413. Deciphering Developmental Disorders Study. Prevalence and architecture of *de novo* mutations in developmental disorders. *Nature* **542**, 433–438 (2017).
414. Kaufman, L., Ayub, M. & Vincent, J. B. The genetic basis of non-syndromic intellectual disability: a review. *J. Neurodev. Disord.* **2**, 182–209 (2010).
415. Simpson, C. L. & Al-Chalabi, A. Amyotrophic lateral sclerosis as a complex genetic disease. *Biochim. Biophys. Acta BBA - Mol. Basis Dis.* **1762**, 973–985 (2006).
416. Ojha, R. P. & Thertulien, R. Health care policy issues as a result of the genetic revolution: implications for public health. *Am. J. Public Health* **95**, 385–388 (2005).
417. Wilson, J. M. G., Jungner, G. & Organization, W. H. Principles and practice of screening for disease. (1968).
418. Ronchi, E. Genetic testing: policy issues for the new millennium. *J. Biolaw Bus.* **4**, 34–36 (2001).
419. Khoury, M. J., Coates, R. J. & Evans, J. P. Evidence-based classification of recommendations on use of genomic tests in clinical practice: Dealing with insufficient evidence. *Genet. Med.* **12**, 680–683 (2010).
420. WHO | Revisiting Wilson and Jungner in the genomic age: a review of screening criteria over the past 40 years. *WHO* Available at: <http://www.who.int/bulletin/volumes/86/4/07-050112/en/>. (Accessed: 4th May 2017)
421. Renton, A. E. *et al.* A hexanucleotide repeat expansion in C9ORF72 is the cause of chromosome 9p21-linked ALS-FTD. *Neuron* **72**, 257–268 (2011).

