



**UNIL** | Université de Lausanne

Unicentre

CH-1015 Lausanne

<http://serval.unil.ch>

---

*Year : 2016*

## Genomic rearrangements and diseases

LOVIGLIO Maria Nicla

LOVIGLIO Maria Nicla, 2016, Genomic rearrangements and diseases

Originally published at : Thesis, University of Lausanne

Posted at the University of Lausanne Open Archive <http://serval.unil.ch>

Document URN : urn:nbn:ch:serval-BIB\_EE2C2B49339E1

### **Droits d'auteur**

L'Université de Lausanne attire expressément l'attention des utilisateurs sur le fait que tous les documents publiés dans l'Archive SERVAL sont protégés par le droit d'auteur, conformément à la loi fédérale sur le droit d'auteur et les droits voisins (LDA). A ce titre, il est indispensable d'obtenir le consentement préalable de l'auteur et/ou de l'éditeur avant toute utilisation d'une oeuvre ou d'une partie d'une oeuvre ne relevant pas d'une utilisation à des fins personnelles au sens de la LDA (art. 19, al. 1 lettre a). A défaut, tout contrevenant s'expose aux sanctions prévues par cette loi. Nous déclinons toute responsabilité en la matière.

### **Copyright**

The University of Lausanne expressly draws the attention of users to the fact that all documents published in the SERVAL Archive are protected by copyright in accordance with federal law on copyright and similar rights (LDA). Accordingly it is indispensable to obtain prior consent from the author and/or publisher before any use of a work or part of a work for purposes other than personal use within the meaning of LDA (art. 19, para. 1 letter a). Failure to do so will expose offenders to the sanctions laid down by this law. We accept no liability in this respect.



**UNIL** | Université de Lausanne

Faculté de biologie  
et de médecine

**Center for Integrative Genomics (CIG)**

**Genomic rearrangements and diseases**

**Thèse de doctorat ès sciences de la vie (PhD)**

présentée à la

Faculté de biologie et de médecine  
de l'Université de Lausanne

Par

**Maria Nicla LOVIGLIO**

Master en Biotechnologie Médicale de l'Université de Bari

**Jury**

Prof. Jérôme Goudet, Président  
Prof. Alexandre Reymond, Directeur de thèse  
Prof. Brunhilde Wirth, expert  
Prof. Bart Deplancke, expert  
Prof. Ioannis Xenarios, expert  
Prof. Vincent Dion, expert

Lausanne 2016



UNIL | Université de Lausanne

Faculté de biologie  
et de médecine

Ecole Doctorale

Doctorat ès sciences de la vie

# Imprimatur

Vu le rapport présenté par le jury d'examen, composé de

<i>Président · e</i>	Monsieur Prof. Jérôme <b>Goudet</b>
<i>Directeur · rice de thèse</i>	Monsieur Prof. Alexandre <b>Reymond</b>
<i>Experts · es</i>	Madame Prof. Brunhilde <b>Wirth</b>
	Monsieur Prof. Bart <b>Deplancke</b>
	Monsieur Prof. Ioannis <b>Xenarios</b>
	Monsieur Prof. Vincent <b>Dion</b>

le Conseil de Faculté autorise l'impression de la thèse de

**Madame Maria Nicla Loviglio**

Master en Science de l' Université de Bari, Italie

intitulée

**Genomic rearrangements and diseases**

Lausanne, le 20 juin 2016

pour le Doyen  
de la Faculté de biologie et de médecine

Prof. Jérôme Goudet

## Table of Contents

<b>ACKNOWLEDGEMENTS</b> .....	<b>4</b>
<b>ABSTRACT (ENGLISH)</b> .....	<b>5</b>
<b>ABSTRACT (FRENCH)</b> .....	<b>6</b>
<b>ABSTRACT (FRENCH, FOR THE GENERAL PUBLIC)</b> .....	<b>7</b>
<b>INTRODUCTION</b> .....	<b>8</b>
<b>CNVs AND COMPLEXITY OF GENETIC DISORDERS</b> .....	<b>8</b>
<b>CNVs AND INTELLECTUAL DISABILITY SYNDROMES</b> .....	<b>11</b>
16P11.2 COPY NUMBER VARIANTS .....	16
17P11.2 COPY NUMBER VARIANTS .....	22
<b>NOVEL APPROACHES TO STUDY CNVs</b> .....	<b>24</b>
<b>REFERENCES</b> .....	<b>30</b>
<b>CHAPTER 1: CHROMOSOMAL CONTACTS CONNECT LOCI ASSOCIATED WITH AUTISM, BMI AND HEAD CIRCUMFERENCE PHENOTYPES</b> .....	<b>47</b>
<b>SUMMARY OF THE CONTRIBUTION</b> .....	<b>47</b>
<b>ABSTRACT</b> .....	<b>50</b>
<b>INTRODUCTION</b> .....	<b>51</b>
<b>MATERIALS AND METHODS</b> .....	<b>52</b>
RECRUITMENT AND PHENOTYPING OF PATIENTS .....	52
LYMPHOBLASTOID CELL LINES AND TRANSCRIPTOME PROFILING.....	54
QUANTITATIVE RT-PCR .....	55
VIEWPOINT SELECTION.....	56
4C-SEQ.....	57
4C-SEQ DATA ANALYSIS.....	58
Hi-C DATA .....	59
ENRICHMENT ANALYSES.....	59
FLUORESCENCE IN SITU HYBRIDIZATION (FISH) .....	61
CHIP-SEQ AND RNA-SEQ MOLECULAR ASSOCIATIONS.....	62
<b>RESULTS</b> .....	<b>62</b>
DISTINCT AND NON-OVERLAPPING LOCI AT 16P11.2 ARE ASSOCIATED WITH MIRROR PHENOTYPES ON BMI AND HC AND AUTISM SUSCEPTIBILITY .....	62
CIS-ACTING CHROMATIN LOOPS THAT LINK THE 16P11.2 BP4-BP5 AND BP2-BP3 GENOMIC INTERVALS ARE PERTURBED IN BP4-BP5 CNV CARRIERS .....	64
PERTURBATIONS OF THE CHROMATIN INTERACTIONS' LANDSCAPE AT 16P11.2 ARE ASSOCIATED WITH GENE EXPRESSION MODIFICATION .....	69
GENOMIC REGIONS CONTACTED BY 16P11.2 VIEWPOINTS ARE ASSOCIATED WITH AUTISM, BMI AND HC PHENOTYPES AND ENRICHED IN CILIARY GENES .....	71
PHENOTYPES ASSOCIATED WITH 2P15-16.1 CNVs.....	74
<b>DISCUSSION</b> .....	<b>76</b>

<b>SUPPLEMENTARY FIGURES</b> .....	<b>79</b>
<b>SUPPLEMENTARY TABLES</b> .....	<b>88</b>
<b>SUPPLEMENTAL DATA</b> .....	<b>109</b>
16P11.2 CONSORTIUM MEMBERS.....	109
2P15 CONSORTIUM MEMBERS .....	110
ACKNOWLEDGMENTS.....	110
AUTHORS CONTRIBUTION.....	111
COMPETING INTEREST .....	111
ACCESSION NUMBERS SECTION .....	111
THE WEB RESOURCES SECTION .....	112
<b>REFERENCES</b> .....	<b>113</b>

**CHAPTER 2: A POTENTIAL ROLE FOR THE LINKER FOR ACTIVATION OF T-CELLS (LAT) IN THE NEUROANATOMICAL PHENOTYPE OF 16P11.2 BP2-BP3 CNVS** ..... **126**

<b>SUMMARY OF THE CONTRIBUTION</b> .....	<b>126</b>
<b>ABSTRACT</b> .....	<b>127</b>
<b>INTRODUCTION</b> .....	<b>128</b>
<b>MATERIALS AND METHODS</b> .....	<b>130</b>
IN VIVO ANALYSIS OF GENE EXPRESSION AND EMBRYO MANIPULATIONS .....	130
ZEBRAFISH WHOLE-MOUNT IMMUNOSTAINING .....	130
<b>RESULTS</b> .....	<b>131</b>
<i>IN VIVO</i> TESTING OF 16P11.2 220 KB BP2-BP3 CNV GENES .....	131
GENETIC INTERACTION BETWEEN <i>LAT</i> AND THE 16P11.2 600KB BP4-BP5 CNV DRIVER <i>KCTD13</i> AND MODIFIERS <i>MVP/MAPK3</i> .....	132
<i>LAT</i> INDUCES COMPARABLE NEURON MORPHOLOGY AND MATURATION DEFECTS AS ITS TCR SIGNAL TRANSDUCTION PARTNERS <i>CD247 (CD3z)</i> AND <i>ZAP70</i> IN THE ZEBRAFISH EMBRYOS .....	136
<b>DISCUSSION</b> .....	<b>138</b>
<b>SUPPLEMENTARY FIGURES</b> .....	<b>142</b>
<b>REFERENCES</b> .....	<b>143</b>

**CHAPTER 3: IDENTIFICATION OF A *RAI1*-ASSOCIATED DISEASE NETWORK THROUGH INTEGRATION OF EXOME SEQUENCING, TRANSCRIPTOMICS AND 3D GENOMICS** ..... **148**

<b>SUMMARY OF THE CONTRIBUTION</b> .....	<b>148</b>
<b>ABSTRACT</b> .....	<b>150</b>
<b>INTRODUCTION</b> .....	<b>151</b>
<b>MATERIALS AND METHODS</b> .....	<b>152</b>
ENROLLMENT .....	152
DETAILED SMS PATIENTS' PHENOTYPES .....	152
ACGH .....	152
EXOME SEQUENCING.....	153
MODELING.....	153
LITERATURE MINING .....	154
IDENTIFICATION OF <i>RAI1</i> INTERACTING PROTEINS.....	154
EMBRYO COLLECTION AND RNA EXTRACTION.....	155
MICROARRAY PROCESSING AND ANALYSIS.....	155

4C-SEQ.....	157
4C-SEQ DATA ANALYSIS.....	157
ENRICHMENT ANALYSES.....	158
HI-C DATA.....	159
<b>RESULTS.....</b>	<b>159</b>
CLINICAL AND MOLECULAR FINDINGS.....	159
VARIANT ANALYSIS AND MODELING.....	161
THE IDENTIFIED RARE VARIANTS AFFECT RAI1-ASSOCIATED GENES.....	164
<b>DISCUSSION.....</b>	<b>170</b>
<b>SUPPLEMENTARY FIGURES.....</b>	<b>174</b>
<b>SUPPLEMENTARY TABLES.....</b>	<b>179</b>
<b>SUPPLEMENTARY TEXT.....</b>	<b>189</b>
DETAILED SMS PATIENTS' PHENOTYPES.....	189
ACKNOWLEDGMENTS.....	194
COMPETING INTEREST.....	194
<b>REFERENCES.....</b>	<b>196</b>
<b><u>THESIS DISCUSSION AND FUTURE PLANS.....</u></b>	<b><u>203</u></b>
<b><u>THESIS DISCUSSION'S REFERENCES.....</u></b>	<b><u>209</u></b>

## Acknowledgements

The thesis is written and ready to be defended, and this amazing five years journey in Lausanne is coming to an end.

I wish to thank the Committee members who will evaluate my work and, in particular, my Thesis Director, Prof. Alexandre Reymond, for supporting me during these years, for the trust and the encouragement I could always find when I knocked at his door. I am very grateful for the amazing opportunities I had in the lab to grow and improve myself, and for the possibility to spend a period of my PhD at Duke University. There I found the best supervisors I could aim to, Prof. Nicholas Katsanis and Prof. Christelle Golzio, who was an inspiring scientific guide and a good friend. I would like to acknowledge all the collaborators, especially Marion Leleu, and all the people and patients who participated in the projects I was involved.

I would like to thank the wonderful people that are part of the Reymond's group, in particular Giuliana, Catia, Elena, Mariana, Norine, Jacqueline, Anne-Maud and Andreas, who helped me, sustained my efforts, stimulated scientific (and less scientific, but still brilliant!) discussions inside and outside the lab, and filled this experience with joy and laughing! I wish to thank the Italian -mostly Apulian, mostly biotech- crew in Lausanne: Francesco, Floriana, Daniele, Giulia, Giuliana, Maurizio and Elena! Aurélien, Min, Anne-Maud and Andreas (aka the "Winter school dinner" friends), I am so happy to have you in my life! Last, but not least, Iris. You have been the best friend I could find, maybe a bit "special" sometimes, but for sure one of the most constant presences in the last few years, and one of the smartest, funniest and most brilliant people I have ever met. ☺

Vorrei inoltre ringraziare la mia Famiglia, mamma, papà e Giosy in particolar modo, che da sempre mi supportano e mi sostengono. Siete stati la mia forza quando, più di quattro anni fa, ho lasciato la mia terra e le mie certezze e sono arrivata piccola e impaurita a Losanna. È passato tanto tempo da allora, e quella scelta mi ha resa più forte, più indipendente e aperta al mondo. Se sono qui oggi, e se ci sono arrivata con tanta grinta ed entusiasmo, lo devo tanto anche all'amore con cui mi avete nutrita e incoraggiata. Grazie agli Amici di sempre, al mitico Campetto che è sempre parte importantissima della mia vita, a Stefania, Enica, Anna, Florinda, Mariella, Nicla, agli amici vicini con il cuore, perché non c'è distanza che possa cambiare questa cosa. Concludo con te, Michele, che sei stato la mia roccia, la certezza più bella di questo percorso, sempre qui, con me e per me, ad abbattere le mie mille paure, l'incertezza nelle difficoltà, a festeggiare i piccoli successi, a tenermi per mano, a sfidare con pazienza e dolcezza la distanza che voleva separarci. La tua presenza è stata forte, costante, essenziale. Tu sai di cosa io sia capace, anche quando io stessa vacillo, e credi fortemente in me. Per questo ti ringrazio, per le cose che abbiamo compiuto insieme, e per le sfide che ci attendono, perché insieme possiamo affrontare ogni cosa! ☺

Ad maiora!

## Abstract (English)

### “Genomic rearrangements and diseases”

Copy number variations (CNVs) are major contributors of genomic imbalances disorders. On the short arm of chromosome 16, CNVs of the distal 220 kb BP2-BP3 region show mirror effect on BMI and head size, and association with autism and schizophrenia, as previously reported for the proximal 600 kb BP4-BP5 deletion and duplication. These two CNVs-prone regions at 16p11.2 are also reciprocally engaged in complex chromatin looping, successfully confirmed by 4C-seq, FISH, Hi-C and concomitant expression changes, and are chromatin interactors of other loci linked to autism and/or mirror phenotypes of BMI and head circumference, for example the 2p15 cytoband. Zebrafish modeling of the BP2-BP3 duplication revealed that the overexpression of the linker for activation of T cells (*LAT*) induces a reduction in dividing cells in the brain and number of post-mitotic neurons in the anterior forebrain, and of intertectal axonal tracts, resulting in microcephaly, and suggested this gene as major contributor in the BP2-BP3 CNVs neurodevelopmental phenotypes. *KCTD13*, *MVP*, and *MAPK3*, three genes mapping within the BP4-BP5 locus and major driver and modifiers, respectively, of the head circumference phenotype linked to that region, and *LAT* act in additive manner to increase the severity of the microcephaly phenotype, supporting the presence of genetic interaction, in addition to proximity in 3D nuclear space, between these two loci.

Smith-Magenis syndrome (SMS) is a developmental disability/multiple congenital anomaly disorder resulting from deletion at 17p11.2 that includes the *RAI1* gene or a nucleotide variant in that gene. We investigated a cohort of 15 individuals with a clinical suspicion of SMS, who showed negative deletion and mutational analysis in *RAI1*. Potentially deleterious variants were identified in eight of these subjects using WES in *KMT2D*, *ZEB2*, *MAP2K2*, *GLDC*, *CASK*, *MECP2*, *KDM5C* and *POGZ*. Analyses of coexpression, biomedical text mining, transcriptome profiling of *Rai1*<sup>-/-</sup> mice and chromosome contacts suggest that these genes and *RAI1* are part of the same disease network.

Our 4C-seq results from 16p11.2 and 17p11.2 studies indicate that chromosomal contacts' maps can be exploited to uncover functionally and clinically related genes. These findings also encourage the integration of the results obtained from various genomic approaches to unravel complex disorders and CNVs.



## Abstract (French)

### « Structure du génôme et pathologies »

La variation du nombre de copies (en anglais, Copy Number Variation, CNV) est un des contributeurs principaux à la pathogenèse des syndromes génétiques rares, mais aussi des maladies multifactorielles fréquentes. Sur le bras court du chromosome 16, les CNVs de la région distale BP2-BP3 de longueur 220 kb conduisent à un effet miroir entre sous-poids et obésité sévère, et micro- et macrocéphalies, et ils sont aussi associés avec l'autisme et la schizophrénie. Des phénotypes similaires ont été observés précédemment sur la même bande chromosomique (16p11.2) pour des délétions et duplications proximales dans la région BP4-BP5 (600 kb). Ces régions BP2-BP3 et BP4-BP5 présentent des contacts chromatiniens réciproques, confirmés avec succès par différentes techniques (4C-seq, FISH, co-régulation dans l'expression des gènes, et des données Hi-C). Elles décrivent aussi des interactions au niveau de la chromatine avec d'autres loci liés à l'autisme et/ou aux phénotypes miroir de l'IMC (Indice de Masse Corporelle) et de la circonférence de la tête, par exemple avec la bande chromosomique 2p15. La modélisation de la duplication de la région BP2-BP3 dans le poisson-zèbre a révélé que la surexpression du gène *LAT* (en anglais, Linker for Activation of T-cells) diminue la prolifération des cellules dans le cerveau et des neurones post-mitotiques dans le cerveau antérieur et le nombre des axones entre les tecta optiques, au début du développement embryonnaire. Dans les stades suivants du développement, nous observons une microcéphalie des poissons. Tous ces éléments indiquent que ce gène est le contributeur essentiel des phénotypes neuro-développementaux des CNVs de la région BP2-BP3. *KCTD13*, *MVP* et *MAPK3* sont situés dans la région BP4-BP5 et, respectivement, un gène principal et deux gènes modificateurs des anomalies de la taille de la tête liée à cette région. Ces trois gènes et *LAT* augmentent ensemble de manière additive la gravité de la microcéphalie, en soutenant la présence d'une interaction, pas seulement dans l'espace 3D nucléaire, mais aussi génétique entre les deux loci.

Le syndrome de Smith-Magenis (SMS) se caractérise par un retard mental, des dysmorphies, des troubles du comportement et du sommeil très sévères, dues à une microdélétion dans la bande 17p11.2 du chromosome 17, qui comprend le gène *RAI1* ou une mutation de ce gène. Nous avons étudié une cohorte de quinze personnes avec un diagnostic de SMS, mais n'ayant pas de délétion ou mutation du gène *RAI1*. Par le séquençage de l'exome, des mutations potentiellement délétères ont été identifiées chez huit de ces sujets dans les gènes *KMT2D*, *ZEB2*, *MAP2K2*, *GLDC*, *CASK*, *MECP2*, *KDM5C* et *POGZ*. Les analyses de la co-expression des gènes, des données de text mining, du profilage du transcriptome des souris *Rai1*<sup>-/-</sup> et des contacts chromatiniens font penser que ces gènes et *RAI1* font partie du même « disease network ».

Les résultats de 4C-seq obtenus par les études des bandes 16p11.2 et 17p11.2 indiquent que les contacts chromosomiques peuvent être exploitées pour découvrir des gènes liés d'un point de vue fonctionnel et clinique. Ces résultats encouragent également l'intégration des données obtenues à partir de différentes approches génomiques pour démêler des troubles complexes et les larges CNVs.

## Abstract (French, for the general public)

### «Structure du génôme et pathologies»

Tout individu dispose normalement de 23 paires de chromosomes, chaque parent donnant une copie de ses gènes à ses descendants. Occasionnellement, de courts segments d'ADN peuvent être perdus pendant ce processus - on parle alors de « délétion » - ou être présents en trois exemplaires - on parle dans ce cas de « duplication ». Ces différences en nombre de copies sur des segments d'ADN, sont désignées par l'acronyme « CNV » (pour « Copy Number Variants ») et leurs manifestations cliniques peuvent être très diverses.

Deux segments situés sur le bras court du chromosome 16 sont réputés pour particulièrement sujets aux délétions et duplications. Ces variations génétiques sont des processus opposés, une opposition que l'on retrouve au niveau du phénotype, autrement dit l'aspect extérieur d'une personne. Nos résultats ont montré l'association de ces CNVs avec un effet miroir entre sous-poids et obésité sévère, micro- et macrocéphalies, mais aussi avec l'autisme et la schizophrénie. Ces deux régions, qui sont éloignées sur la molécule linéaire de l'ADN, se rapprochent et interagissent dans l'espace tridimensionnel du noyau de la cellule. Ces interactions sont probablement importantes dans le contexte de la chromatine (i.e. la forme sous laquelle se présente l'ADN dans le noyau) pour des raisons fonctionnelles. Nous avons aussi travaillé avec le poisson-zèbre pour développer un modèle qui nous permet de tester les effets des duplications sur l'expression phénotypique. Cette approche nous a permis d'identifier le gène dont les niveaux d'expression influencent la taille du cerveau.

Durant ma thèse, j'ai également travaillé sur les erreurs de diagnostics. Le syndrome de Smith-Magenis (SMS) se caractérise par un retard mental, des dysmorphies, des troubles du comportement et du sommeil. Normalement, il est causé par une délétion du gène RAI1 sur le chromosome 17. Nous avons identifié des patients souffrant de SMS, mais ayant des mutations dans d'autres gènes. Le fait que ces gènes aient une fonction dans des voies métaboliques similaires à celle de RAI1 suggèrent que ces résultats sont probants. En résumé, ces gènes sont des pièces du même puzzle et donc le syndrome peut être provoqué par le manque de n'importe quelle pièce.

## Introduction

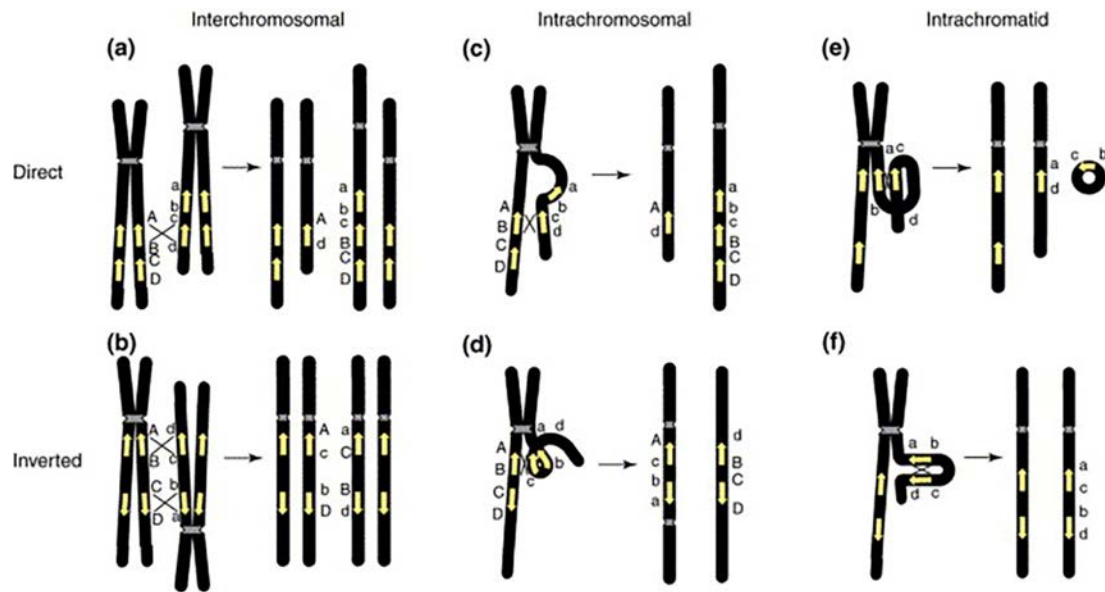
### CNVs and complexity of genetic disorders

With the advent of new technologies for analyzing the genome, our view of its dynamic structure has changed. The human genome contains a diverse array of genomic variants, including single nucleotide polymorphisms (SNPs), length polymorphisms of microsatellite sequences, and several types of structural variations (SVs)<sup>1</sup>. SVs comprise dosage-altering variations, such as insertions and deletions, and dosage-invariant rearrangements, such as inversions and translocations. In recent years, the discovery of copy number variations (CNVs) have been accelerated by the introduction of single nucleotide polymorphism (SNP) genotyping arrays and array comparative genomic hybridization (array-CGH) in both research and diagnostic settings, and even further by the recent advances in exome and genome sequencing analysis<sup>2,3</sup>, revealing such events to be extremely abundant in genomes<sup>4</sup>. The traditional definition of CNVs is of DNA segments which are greater than 1 kb in length and which are present at a variable copy number (gain and losses of DNA) across individuals, while smaller (<1 kb) insertions or deletions are referred to as indels<sup>5</sup>. Their size range can vary from kilobases to several megabase<sup>6</sup>, which makes them not identifiable by conventional chromosomal banding. Several studies have revealed important information regarding the size and frequency of CNVs in normal general populations<sup>7-9</sup>. Approximately 65–80% of normal individuals carry a CNV less than 100 kb in size, 5–10% carry a CNV less than 500 kb in size and 1% of individuals carry a CNV at least 1Mb in size. The current version of the Database of Genomic Variants (DGV) includes about 7 millions entries and overlaps 94.47% of transcripts<sup>6, 10, 11</sup> (<http://dgv.tcag.ca/dgv/app/home>). A large proportion of CNVs is polymorphic (copy number polymorphism, CNPs) in the population and considered as a potential source of inter-individual phenotypic variance<sup>12-14</sup>. CNPs can encompass genes, especially those of the immune and environmental response pathways, which suggests that CNPs likely play an important role in local adaptive selection in human populations with large differences in frequency between populations<sup>15</sup>. The same molecular mechanisms that cause rare pathogenic CNVs are likely to also cause CNPs.

An association to distinctive syndromes (e.g. 15q24 deletion syndrome<sup>16</sup> and 17q21.31 deletion syndrome<sup>17, 18</sup>) was instead shown for several other CNVs. Large CNVs, especially *de novo*, have been predicted to account for a significant component of variation in rare and complex genomic imbalance disorders and birth defects (e.g. craniofacial, cardiac, respiratory, renal)<sup>4</sup>. Locus-specific mutation rates for genomic rearrangements are 1'000-10'000-fold higher than for point mutations<sup>19</sup>, with similar frequency as genomic disorders, although there is some variability across different world populations.

Stankiewicz and Lupski defined duplicated genomic DNA segments ranging from 10-400kb and sharing  $\geq 95\%$ -97% sequence identity<sup>20</sup> as region-specific low copy repeats (LCRs), also termed segmental duplications (SDs)<sup>5, 21</sup>. In the human genome they are characterized by a complex structure, originated from serial segmental duplications over the past 25-40 million years during primate speciation. Particular structural features in genomic repeats, such as repeat length, degree of homology and distance between LCRs, can facilitate the formation of constitutional (i.e. inherited, both recurrent and non-recurrent) and somatic genomic rearrangements and can lead to local genomic instability<sup>22, 23</sup>.

The most frequent mechanism for common-sized recurrent rearrangements is non-allelic homologous recombination (NAHR) between misaligned directly oriented LCRs. The inter-chromatid NAHR events between directly oriented LCRs/SDs result in deletions and duplications. The intra-chromatid NAHR events between directly oriented LCRs/SDs can generate deletions and ring-shaped DNA segments that will be lost in subsequent cell divisions. The NAHR event between reversely oriented LCRs/SDs can cause inversion, a copy-neutral structural variation<sup>5, 24</sup> (**Figure 1**). Complex structure consisting of both direct and inverted LCRs can serve as NAHR substrates leading to, respectively, genomic deletions/duplications and inversions.



**Figure 1. Schematic representation of low-copy repeat/nonallelic homologous recombination (LCR/NAHR)-based mechanisms for genomic rearrangements.** Chromosomes are shown in black, with the centromere depicted by hashed lines. Yellow arrows depict LCRs. The figure depicts LCRs arranged horizontally according to orientation and structure (direct, inverted, complex). The chromosome rearrangements and predicted products of recombination are listed vertically by mechanisms (interchromosomal; intrachromosomal; and intrachromatid). Interchromosomal misalignment leads to deletion/duplication (directly oriented LCRs) **(A)** and inversion (inverted repeats) **(B)**. Intrachromatid loop of inverted repeats results in inversion **(F)**. Interchromatid mispairing of direct repeats results in deletion/duplication **(C)**. Intrachromatid misalignment of directed repeats **(E)** results in deletion and an acentric fragment. Inv dup chromosomes can be the consequence of intrachromosomal **(D)** unequal exchange between inverted LCRs. Modified from<sup>21</sup>.

Non-homologous end joining (NHEJ) is a different recombination-based mechanism participating to the formation of non-recurrent genomic rearrangements. It participates to the repair of DNA double strand breaks by direct rejoining of the DNA ends, a process that can lead to nucleotide gain or loss. It doesn't rely on the presence of low-copy repeat, although particular features of genome architecture can promote it<sup>25</sup>. In addition to these processes, complex non-recurrent rearrangements can be also originated by DNA replication errors, for example by fork stalling and template switching (FoSTeS) and/or microhomology-mediated break induced replication (MMBIR). When the

DNA replication fork stalls, the lagging strand can anneal to another replication fork in physical proximity, by means of microhomology at the 3' end, “priming” or reinitiating DNA synthesis<sup>26, 27</sup>. The FoSTeS mechanism has been further generalized in a replicative template-switch model based on experimental observations from multiple model organisms, the MMBIR model<sup>27</sup>.

The phenotypic effects of CNVs are highly variable and depend on the impact of the genomic rearrangement on regulatory sequences and genes within/nearby the locus of interest, which can be affected in different ways, comprising gene dosage, gene disruption and position effect. These mechanisms have been documented also for apparently balanced translocations and even exert their effects when the breakpoints map as far as ~1.5 Mb away either upstream or downstream from the causative gene<sup>5, 28, 29</sup>. In case of a deletion, the rearrangement are generally produced by gene imbalance but can also be by unmasking recessive mutations or functional polymorphisms on the remaining allele<sup>30</sup>. Furthermore, the deletion of regulatory elements or the disruption of a coding sequence can potentially perturb transvection, i.e. the communication between alleles on homologous chromosomes<sup>31</sup>.

### **CNVs and intellectual disability syndromes**

The term “neurodevelopmental disorders” (NDs) groups a large subset of clinical entities encompassing a spectrum of neurological and psychiatric manifestations which arise upon disruption of brain development, from conception to early adulthood, including autism spectrum disorders (ASD), intellectual disability (ID), communication disorders, attention deficit and hyperactivity disorder (ADHD), specific learning disorders and motor disorders<sup>32</sup>. Schizophrenia (SCZ) has also been suggested to result from neurodevelopmental abnormalities, but it usually manifests only at the adult stage<sup>33</sup>. NDs are usually characterized by a large overlap of symptoms varying from cognitive impairment, behavioral abnormalities, psychotic symptoms, sensory impairment, seizures, neuromotor dysfunction and/or speech and language difficulties<sup>34</sup>. The last 50 years have

been incredibly important in the dissection of the causes of neuropsychiatric disorders, investigation of their etiology and development of new diagnostic methods<sup>35</sup>. Originally, cytogenetic analysis allowed the identification of chromosomal abnormalities in the late 50s<sup>36</sup>, whereas 20 years later the techniques for chromosome banding were discovered, allowing the detection of a substantial number of microdeletion, microduplications and translocation syndromes<sup>37-42</sup>. More recently, FISH (fluorescence in situ hybridization) allowed the detection of the gain/loss of 1-5Mb genomic segments involved in new recurrent patterns of ID and congenital anomalies<sup>43</sup>, but the advent of comparative genome hybridization (CGH) was the true catalyst for a series of significant discoveries in clinical cytogenetics<sup>44-46</sup>. This leap forward was followed by next generation sequencing (NGS) using paired-end methods, where additional small variants, down to tens of bases, could be also detected<sup>47, 48</sup>. In a subset of NDs the classical Mendelian disease model where one gene explains a given trait does not provide an adequate fit. In fact, most of NDs are polygenic or multifactorial and tend to cluster in families because of genetic and environmental factors' influence<sup>49</sup>.

Rare CNVs are involved in the pathogenesis of neurodevelopmental and neurocognitive disorders, such as ID, SCZ and ASD. ID is defined by significant limitations in intellectual functioning and adaptive behavior, which starts early in life, often before the age of 18, and is diagnosed when individuals score less than 70 in the IQ test<sup>50</sup>. Several environmental factors, such as teratogenic agents (for example, maternal alcohol abuse during pregnancy), infections, perinatal hypoxia/birth complications and extreme malnutrition, can be implicated in the onset of ID, as well as genetic factors. Severe ID is mostly sporadic, and can be due to the effect of dominantly acting rare *de novo* mutations or rare CNVs<sup>51-57</sup>. Its severity strongly correlates with number of genes affected by the copy number variants<sup>58</sup>. Milder forms can, instead, occur in families, with more common and complex pattern of inheritance. Chromosomal abnormalities account for up to 15% of case<sup>59, 60</sup>, while *de novo* CNVs arise in ~10% of patients<sup>43, 61-63</sup>. Overall, over 700 genes have currently been reported for autosomal-dominant, autosomal-recessive and X-linked (more than 100 genes,

collectively explaining up to 10% of IDs in males<sup>64</sup>) forms of ID<sup>50</sup>. The validation of a candidate ID gene can take advantage of several approaches, including the replication in unrelated individuals, the assessment of the frequency of mutations in cases versus controls, the enrichment of gene mutations within a patients cohort<sup>65</sup>, as well as modelization with patient-derived cells, animal models. The comparison of CNV morbidity maps with lists of candidate ID genes can also be very powerful for validation purposes<sup>66</sup>.

Many studies have reported the high co-morbidity shared between ID and other cognitive impairments, like ASD, SCZ and epilepsy<sup>67</sup>, suggesting common molecular pathways impacted by these neurodevelopmental disorders<sup>68,69</sup>. SCZ is a complex and highly heritable disorder, characterized by several signs clusters including positive symptoms, negative symptoms, cognitive impairments, and affective dysregulation, and resulting in substantial disability and 10% risk of suicide. It can develop at any age, but in particular between the ages of 16 and 30, but many of its causes stem back to prenatal development<sup>70</sup>. An increased burden of rare mutations has been reported in SCZ, such as large CNVs<sup>71-73</sup> and single-nucleotide variants (SNVs)<sup>74</sup>, which often occur as *de novo* mutations<sup>75-77</sup>. In general, CNVs tend to increase disease risk to a much greater extent than individual SNPs. Consistent with this hypothesis, no individual SNV has been robustly linked to SCZ<sup>74</sup> yet, while 11 large CNVs are strongly associated<sup>72</sup>. About 2.5% of patients versus 0.9% of unaffected controls carry a CNV, supporting copy number variants as a risk factor for SCZ<sup>72, 78</sup>. ASD is a neurodevelopmental disorder defined by a significant impairment in reciprocal social interaction, communication deficits and repetitive and restricted patterns of behavior and interests. The affected individuals can show very variable severity and atypical development, which arise before the age of 3 in most of the cases. According to estimates from U.S. Centers for Disease Control and Prevention (CDC)<sup>79</sup>, ASD affects 1 in every 68 children and it is one of the most prevalent childhood disorders. The Diagnostic and Statistical Manual of Mental Disorders (DSM-5) defines the core symptoms of autism spectrum disorder (ASD) with two categories of behavior: (i) deficit in social communication; and (ii) stereotyped behavior. Autistic disorder or classic autism, Asperger's

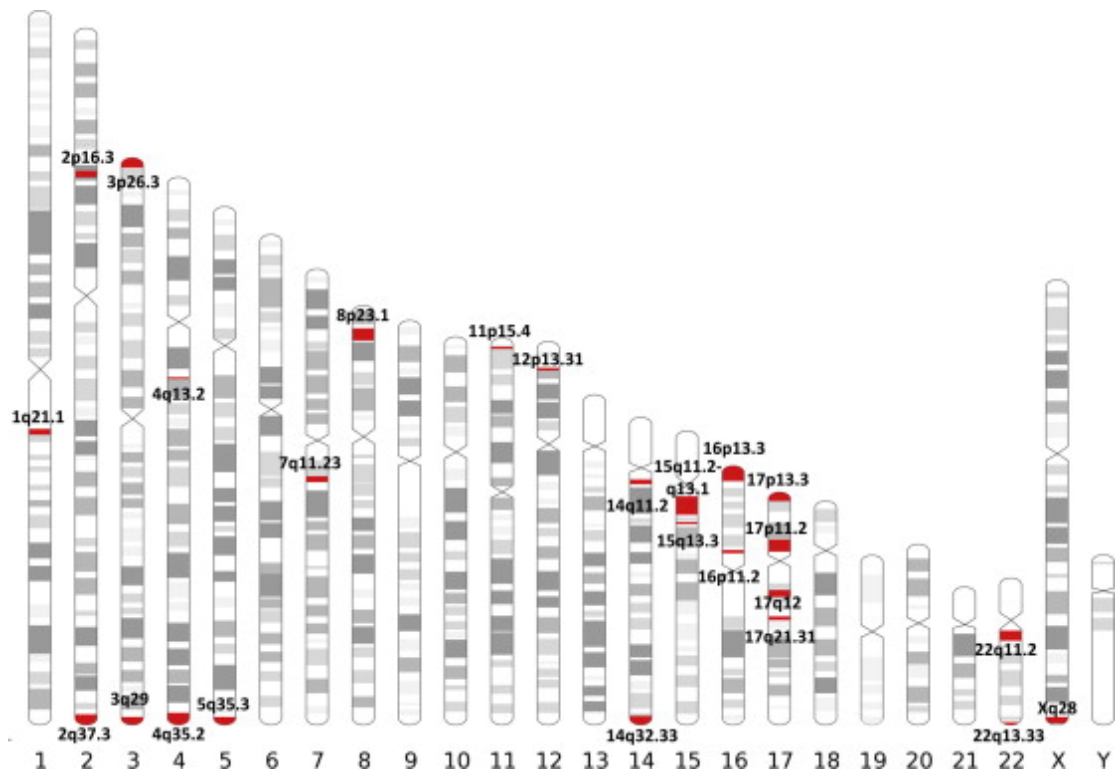


syndrome and pervasive developmental disorder – not otherwise specified (PDD-NOS), which all fit in the autism spectrum, differ with regard to symptom severity and early development of language, cognitive and social behavior<sup>80</sup>. ASD has strong genetic bases and is highly heritable, although it is genetically complex and its underlying genetic architecture is not completely understood. Two contrasting models have tried to explain the complexity of this disease, the common variant common disease (CVCD) and rare variant common disease (RVCD) model. The first one proposes that the genetic risk in an individual is due to the combination of many high frequency (minor allele frequency >1%) / low risk variants (odds ratio <1.5), while the second hypothesis favors the idea of rare mutations conferring high risk<sup>81, 82</sup>. In 2014, Gaugler et colleagues showed that, although ASD individual risk-associated genes have been mostly identified from rare variation, especially *de novo* mutations<sup>83-92</sup>, its heritability is ~52.4%, due to common variation<sup>93</sup>. Both contributions are, therefore, likely to be important. CNVs have a very critical role in the etiology of ASD. Initially, Sebat and colleagues compared CNVs between autistic patients and unaffected parents/siblings using aCGH, allowing efficient detection of *de novo* CNVs in 10% of ASD cases from simplex families, 3% in multiplex families, but only 1% in the general population. Later studies confirmed high *de novo* CNVs rates from 5.8% and 8.4% in ASD<sup>88, 94, 95</sup>, with a few major examples of syndromic CNV listed in **Table 1** and most frequent chromosomal abnormalities shown in **Figure 2**<sup>81, 88, 96, 97</sup>.

CNV locus	Frequency in ASD <sup>88, 96, 97</sup> (n = 2120)	Frequency in developmental disorders <sup>98</sup> (n = 15767)	Frequency in control <sup>88, 96-98</sup> (n = 8329)	Association to ASD	Related disorders <sup>81, 88, 96, 97</sup>
1q21	0.20%	0.50%	<0.1%	dup	Schizophrenia (deletion)
5p15.2	0.10%	-	0%	del	Novel region found for ASD
7q11.23	0.20%	0.40%	0%	dup	Williams-Beuren syndrome (deletion)
15q11-13	0.10%	0.30%	0%	del, dup	Prader-Willi syndrome (paternal deletion), Angelman syndrome (maternal deletion), ASD (maternal duplication)
15q13	0.20%	0.40%	<0.1%	dup	Schizophrenia (deletion), bipolar disorder (duplication)
<b>16p11.2</b>	0.80%	0.50%	<0.1%	del, dup	Schizophrenia (deletion), bipolar disorder, OCD (duplication)
<b>17p11.2</b>	0.20%	0.20%	0%	dup	Smith-Magenis syndrome (deletion)
22q11.2	0.10%	0.90%	0%	del, dup	DiGeorge syndrome, schizophrenia (deletion)

ASD, autism spectrum disorder; CNV, copy-number variation; OCD, obsessive-compulsive disorder.

**Table 1. Major CNV with strong association with ASD and developmental disorders. Modified from<sup>80</sup>.**

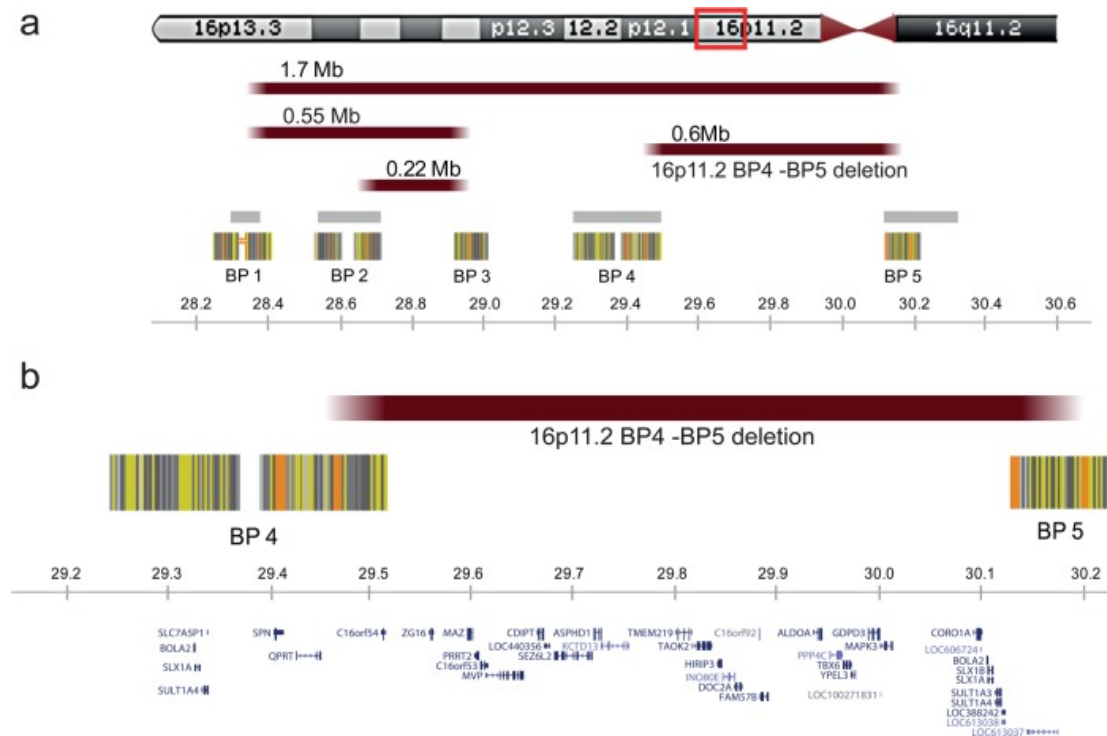


**Figure 2. Karyogram for most frequent chromosomal abnormalities associated with ASD.** 25 loci are shown in the karyogram, represented as red bands. Notably, many loci are located in telomere proximal regions. Modified from<sup>99</sup>.

### 16p11.2 copy number variants

The last 15 million years of hominoid evolution have seen a rapid integration of segmental duplications or low-copy repeats in two of the most proximal bands of the short arm of human chromosome 16 (chromosomal bands 16p11.2 and 16p12.1)<sup>100, 101</sup>, predisposing several loci to a higher frequency of pathogenic rearrangements through Non-Allelic Homologous Recombination during meiosis<sup>102</sup>. Consistently, different recurrent rearrangements have been described within this interval: i) the 16p11.2-p12.1 ~8 Mb (Megabase) deletion that covers the entire region (chr16: 21.3-29.4 Mb)<sup>98, 103-106</sup>; ii) the 16p12.1 ~520 kb deletions and duplications encompassing *EEF2K* and *CDR2*<sup>58, 98</sup>; iii) the 16p11.2 distal ~220 kb BP2-BP3 deletions and duplications; iv) the 16p11.2

proximal ~600 kb BP4-BP5 deletions and duplications, as well as v) 550 kb BP1-BP3 and vi) 1.7 Mb BP1- BP5 rearrangements (iii-vi in **Figure 3**). The ~600 kb deletion/duplication region at 16p11.2 (29.5 to 30.1Mb) is a typical example of such novel genomic imbalances. This region is flanked by 147 kb segmental duplication with 99.5% sequence identity<sup>107</sup>.



**Figure 3. The 16p11.2 rearrangements. Five highly homologous blocks of low copy repeats (LCRs) acting as substrates for non-allelic homologous recombination have been defined as mediators of recurrent and clinically relevant imbalances within the 16p11.2 chromosomal band. These ‘recombination hotspots’ from telomere to centromere are renamed as breakpoints BP1 to BP5. (A) 16p11.2 BP1-BP5, BP1-BP3, BP2-BP3, BP4-BP5 rearrangements are schematically pinpointed with brown bars while grey bars and striated blocks indicate intervals of recurrent polymorphisms reported in the Database of Genomic Variants (<http://projects.tcag.ca/variation>) and common sequence stretches, respectively. (B) Genes encompassed by the genomic region between BP4 and BP5 are shown. All genomic positions are given according to the human genome build hg18/NCBI 36. Modified from <sup>107</sup>.**

The deletion of this interval, which has a prevalence of ~0.05% in the general population, has recently been implicated as one of the most significant genetic

risk factors for ASD in multiple cohorts, among a variety of other childhood-onset neurodevelopmental and psychiatric conditions, such as mental retardation, seizure, language delay, and schizophrenia<sup>108</sup>. ASD is diagnosed in approximately 18% of deletion carriers and global cognition is affected, with a shift of approximately 2 SDs in IQ without altering the variance<sup>107-110</sup>. Pooling the data obtained from a general population cohort with those from the obesity cohorts in an overall case-control association analysis, significant evidence for association of the 16p11.2 deletions with obesity and morbid obesity (43-fold increased risk), often accompanied by hyperphagia, was shown<sup>111, 112</sup>, and significant increase in head circumference was linked to the deletion of this region as well<sup>108, 113</sup>. The corresponding reciprocal duplication, recurring with ~0.04% prevalence in the general population similar to what observed for the deletion, was linked with schizophrenia (SCZ), underweight (8-fold increased risk) and microcephaly. Furthermore, the 600 kb duplication confers risk for typical and atypical Rolandic epilepsy, which is the most frequent form of focal epilepsy in childhood<sup>114</sup>. Other physiological abnormalities associated with 16p11.2 copy number variants include motor hypotonia, seizures, feeding difficulty, immune deficiency, syringomyelia, hearing loss, and cardiac defects<sup>88, 108, 112-123</sup>.

The reciprocal impact of the above-mentioned 16p11.2 copy number variants suggests that severe obesity and being underweight could have mirror etiologies, possibly through contrasting effects on energy balance<sup>108</sup>. Of note, some of these mirroring comorbidities might be causally linked, as macrocephaly was associated with ASD and microcephaly with SCZ, driving the hypothesis that these syndromic conditions may represent mirror states associated with reciprocal changes in copy number at this locus<sup>108, 113</sup>. Reciprocal CNVs at 16p11.2 BP4-BP5 were recently associated with global brain metrics, using a combination of molecular, neuroimaging and clinical approaches. Specifically the copy number negatively correlates to the gray matter volume and white matter tissue properties in cortico-subcortical regions implicated in reward, language and social cognition, resembling the well-established structural abnormalities in ASD and SCZ<sup>110</sup>. The largest effect was detected in the thalamus, with trends in cerebellum and hippocampus<sup>124</sup>. Through a comprehensive phenotypic

characterization of the 16p11.2 BP4-BP5 duplication and deletion ascertained in US and European cohorts it was possible to assess that the frequency of ASD, consistent with previous studies, is similar in deletion and duplication probands, suggesting that both CNVs equally predispose to autism. The duplication is linked with a form of low-functioning ASD and epilepsy compared to the normal cognition range observed in the deletion carriers, which could be due to the presence of additional factors with a negative effect on IQ. A large variance and multimodal distribution in FSIQ are important features of the duplication, while the deletion group shows a consistent effect of the rearrangement on FSIQ across carrier groups and a normal distribution consistent with what is observed in the general population. Duplication and deletion carriers ascertained for neurodevelopmental disorders were also characterized by an overrepresentation of males and lower IQ in female participants, similar to ASD cohorts<sup>125</sup>. Concerning the obesity/underweight mirror phenotype, young deletion carrier present altered satiety responsiveness, prior to any diagnosis of obesity, which results in an increased responsiveness to food, emotional overeating, and is later responsible of an increase in the BMI z-score<sup>126</sup>. Although it is very clear that the 16p11.2 CNVs confer a high risk for neurodevelopmental disease, other factors, and in particular the genetic background, are likely playing a role in the determination of the phenotypic outcome. Duyzend and colleagues have reported a maternal parent-of-origin bias for *de novo* 16p11.2 deletions, possibly due to the presence of a strong hotspot of female recombination within the critical region, and the same bias for secondary CNVs<sup>127</sup>.

No data at present point towards particular candidate genes for obesity within the more proximal 600 kb BP4-BP5 recurrent rearrangements. The locus contains 28 “unique” genes (*SPN*, *QPRT*, *C16orf54*, *ZG16*, *KIF22*, *MAZ*, *PRRT2*, *PAGR1* (aka *C16orf53*), *MVP*, *CDIPT*, *CDIPT-AS*, *SEZ6L2*, *ASPHD1*, *KCTD13*, *TMEM219*, *TAOK2*, *HIRIP3*, *INO80E*, *DOC2A*, *C16orf92*, *FAM57B*, *ALDOA*, *PPP4C*, *TBX6*, *YPEL3*, *GDPD3*, *MAPK3*, *CORO1A*) and multiple copies of *BOLA2/2B*, *SLX1A/1B*, *SULT1A3/4*, *SMG1P* and *NP1P*, the majority of which are expressed in the brain and thus of potential functional relevance for normal neurodevelopment<sup>108</sup>. A three-generation family with ASD was reported with a

co-segregating deletion of ~118 kb, potentially refining the critical region for the autism, with five genes residing within this region, *MVP*, *CDIPT1*, *SEZ6L2*, *ASPHD1* and *KCTD13*<sup>128</sup>. Furthermore, an atypical 136 kb duplication encompassing the *QPRT* and *SPN* genes was found in a female with anorexia nervosa<sup>108</sup>. In vivo modeling of the 16p11.2 rearrangements in zebrafish embryos described *KCTD13* as the major driver of the mirrored head circumference phenotype. Furthermore, it was shown that overexpression of *KCTD13* together with *MVP* or *MAPK3* leads to more profound changes in mean head size compared to overexpression of *KCTD13* alone, suggesting that the epistatic effect of these three genes is involved in the microcephaly/macrocephaly phenotype of 600 kb rearrangement carriers<sup>129</sup>. The participation of *KCTD13* to the Cul3-RhoA pathway in particular brain region, i.e. the layer 4 of the inner cortical plate, was suggested to be important for controlling brain size. The late mid-fetal period of cortical development is a crucial stage for the establishment of the 16p11.2 proteins connectivity with their partners; its dysregulation at may be a potential determinant of 16p11.2 CNV deletion and duplication phenotypes<sup>130</sup>. The T-box 6 (*TBX6*) gene, a transcription factor involved in regulation of early developmental processes, was initially proposed to play a role in the congenital anomalies among a series of patients<sup>113</sup> and in vertebral malformation in mice<sup>131, 132</sup>. Recently, this hypothesis has been confirmed by Wu and colleagues' finding that *TBX6* null mutations and noncoding variants therefore contribute substantively to the complex trait of sporadic congenital scoliosis<sup>133</sup>.

Data from three independently generated mouse models of 16p11.2 deletion syndrome were published in the last years, by the Mills group at Cold Spring Harbor Laboratory<sup>134</sup>, by the Dolmetsch group at Stanford University<sup>135</sup>, and by the Herault group at IGBMC, which also provided the first model of duplication<sup>136</sup>. All deletion lines showed low body weight and impaired adipogenesis, perinatal mortality, increased spontaneous locomotor activity in a novel home cage environment and hyperactivity, repetitive behaviors and sporadic motor stereotypies. The deficits in pairwise discrimination and reversal learning in the touchscreen task, well-replicated novel object recognition

deficits, an object location memory deficit, and a preference for social novelty deficit recapitulated the cognitive impairments displayed by the 16p11.2 deletion carriers, with differences in contextual learning and memory between the lines due to genetic background, exact deletion size, environmental factors, and differing testing procedures<sup>134, 135, 137</sup>. Neuroanatomical defects included reduced cortical thickness, due to possibly altered cortical neurogenesis, with a reduced numbers of Pax6+ progenitor cells and Satb2+ callosal projection neurons in both the Dolmetsch and Mills 16p11.2 +/- mice<sup>134, 135, 137</sup>. Interestingly, the cortical thickness is decreased in human carriers<sup>110</sup> and mice harboring the heterozygous deletion, although the syntenic human carriers show a consistent increase in overall brain volume and weight, which differs to mouse models of the deletion exhibiting smaller brain volumes<sup>124</sup>.

Intriguingly, 16p11.2 +/- mice exhibit anxiety-like behavior and memory deficits similar to phenotypes observed in the ERK1 and ERK2 null mice and their the cortex shows strikingly similar perturbations in cortical cytoarchitecture to those we previously observed in mice with ERK deletions<sup>138</sup>. Together with the observation of altered levels of ERK signaling effectors cyclin D1 and p27 (Kip1) during mid-neurogenesis, these results support the hypothesis that the ERK dysregulation might be contributing to the developmental anomalies and behavioral deficits observed in the 16p11.2 deletion carriers<sup>139</sup>. Concerning the effects on metabolism, the duplication model allowed a better understanding of the opposite energy imbalances occurring in Del/+ and Dup/+ models when challenged with high fat and high sugar diet<sup>136</sup>. The analysis of the transcriptome revealed a marked impact of the 16p11.2 gene expression perturbations on several pathways with neurocognitive and metabolic phenotypes. Interestingly, differently from the behavioral, activity and memory impairments, which perfectly reproduced the human phenotype of 16p11.2 600kb rearrangements, the mirrored metabolic defects were replicated in mice with an opposite direction effect, i.e. adult Del/+ mice were lean in comparison to the human obese phenotype and the Dup/+ mice were overweight in comparison to the human underweight phenotype<sup>136</sup>.



Less frequently, another locus, distal to the above-mentioned CNV in the 16p11.2 and spanning 220 kb (28.7 Mb to 28.9 Mb) (**Figure 3**), can also be affected by recurrent microdeletions and microduplications. The deletion of this region, which is centered around the *SH2B1* gene, previously associated by GWAS with BMI, serum leptin and body fat<sup>140-142</sup>, was reported in individuals with developmental delay and obesity<sup>143, 144</sup>, and also significantly linked with SCZ in a recent large-scale association study<sup>145</sup>. Detailed data about the phenotypes associated with the reciprocal duplication were lacking so far. Results obtained in our lab confirmed the above-mentioned association of the 220 kb deletion with increased body mass index (BMI) as well as with developmental delay/intellectual disability (DD/ID), but also revealed a new link with increased head circumference and ASD. The reciprocal duplication was associated with the mirrored phenotypes of decreased BMI and head circumference, and also linked to DD/ID, ASD and epilepsy [see **Chapter 1** of the Results section]. The observation that similar phenotypic manifestations can be derived from distinct genomic imbalances challenges our understanding of possible co-regulation of distant loci. Furthermore, the delineation of the dosage-sensitive gene(s) underlying the observed phenotypes is hampered by the recurrent rearrangement-specific breakpoints of the distal and proximal 16p11.2 deletions/duplications, preventing identification of the disease-associated critical region. However, there is mounting evidence supporting a role for haploinsufficiency of *SH2B1* in the obesity phenotype of patients with the 220 kb BP2-BP3 deletion: it encodes an adaptor protein involved in the leptin and insulin signaling; it is a likely causal obesity gene from the GWAS studies; mice lacking *Sh2B1* are characterized by obesity and severe insulin resistance<sup>142, 146</sup>.

### 17p11.2 copy number variants

Deletion and duplication of a 3.7Mb region in 17p11.2 result in two reciprocal syndromes, Smith–Magenis (SMS; MIM182290)<sup>147</sup> and Potocki–Lupski syndrome (MIM610883)<sup>148</sup>. SMS is a complex and rare neurobehavioral disorder with an

estimated prevalence of 1:15,000 to 1:25,000 live births. Characteristic associated features include craniofacial features, DD, metabolic problems and obesity, skeletal anomalies, moderate to profound ID, self-injurious and stereotypic behaviors<sup>147, 149</sup>. Sleep disturbances are some of the most peculiar and penetrant observations in SMS patients and comprise difficulties in falling asleep at night, reduced or absent rapid eye movement (REM) sleep, early waking, frequent night-time arousals for feeding, and daytime prolonged napping<sup>150-152</sup>. Several studies have indicated as the underlying cause of the sleep perturbation an inversion in the rhythm of melatonin secretion from the pineal gland, which is very high in the daytime instead than during nighttime, although this is not present in 100% of the patients<sup>152, 153</sup>. ASD features are also displayed, with restricted interest and obsessional thinking<sup>154</sup>. Repetitive behaviors include body squeeze, licking and flipping, self-hugging, seeking constant attention, rock, spin or twirl their body and grind their teeth<sup>155-158</sup>. Furthermore, self-injurious, destructive and aggressive behavior is present in 70-97% of patients and more prevalent in SMS than in other ID with different etiologies<sup>159</sup>. Its manifestations encompass polyembolokoilomania (insertion of foreign objects into bodily orifices), onychotillomania (pulling finger and toe nail out) and other unusual behaviors like poking others' eyes, forceful hugging and punching fists through walls and windows.

As previously mentioned, molecular cytogenetic analysis of SMS patients classically associated the syndrome with a deletion within cytoband 17p11.2 spanning ~3.8Mb and encompassing 34 genes, including the *RAI1* gene (90% of patients). About 7-10% of SMS individuals carry a nucleotide variant in *RAI1*, including insertions or deletions within its coding region that result in frameshifts, as well as missense and nonsense mutations<sup>147, 160-163</sup>. *RAI1* was identified as a target of retinoic acid induction in P19 mouse embryonic carcinoma cells<sup>164</sup>. Its subcellular localization is in the nucleus, given its role of transcriptional regulator, with a PHD (plant homeodomain) motif; it is expressed in migrating neural crest cells, early developing nervous system and, at lower levels, in adult brain<sup>160, 165</sup>.

The abnormal functioning of *RAI1* can possibly explain the circadian disturbances in SMS, as recent studies have shown that *RAI1* is a crucial player in the transcription of circadian locomotor output cycles kaput (CLOCK), a key component of the mammalian circadian oscillator that transcriptionally regulates many critical circadian genes, both in the hypothalamus, where the suprachiasmatic nucleus (SCN), responsible for controlling the central circadian rhythm, resides, and also within the peripheral circadian oscillators, i.e. liver, heart and kidney<sup>166</sup>.

SMS-like individuals were found to recurrently harbor deletions of the 2q37.3 or 2q23.1 cytobands encompassing *HDAC4* and *MBD5*, respectively<sup>167-169</sup>. Similarly, *PITX3* was proposed to be responsible for the SMS-like neurobehavioral abnormalities observed in a patient<sup>170</sup>. These findings suggest genetic heterogeneity for SMS syndrome, which is further investigated in this thesis [see **Chapter 3** of the Results section].

### Novel approaches to study CNVs

The study of large and complex copy number variants involved in neurodevelopmental disorders can benefit from different kinds of approaches. The deep characterization (including genetic and environmental aspects) of: (A) large groups of patients with a given CNV and (B) large cohorts of control individuals carrying the same CNV, can provide some insights about the events/additional variants/risk factors that can be necessary to push the neuropsychiatric phenotype beyond threshold of disease. The analysis of the transcriptome (either using lymphoblastoid cell lines, fibroblast or postmortem brain samples) is necessary to address the impact of CNVs in dosage-sensitive genes and consequent affected pathways, and to infer the potential contribution in perturbing brain development. Functional studies using animal models and/or, more recently, human induced pluripotent stem cells (hiPSCs) lines provide a deeper understanding of pathogenic mechanisms and are useful for the development of new therapeutic strategies. As the Encyclopedia of DNA

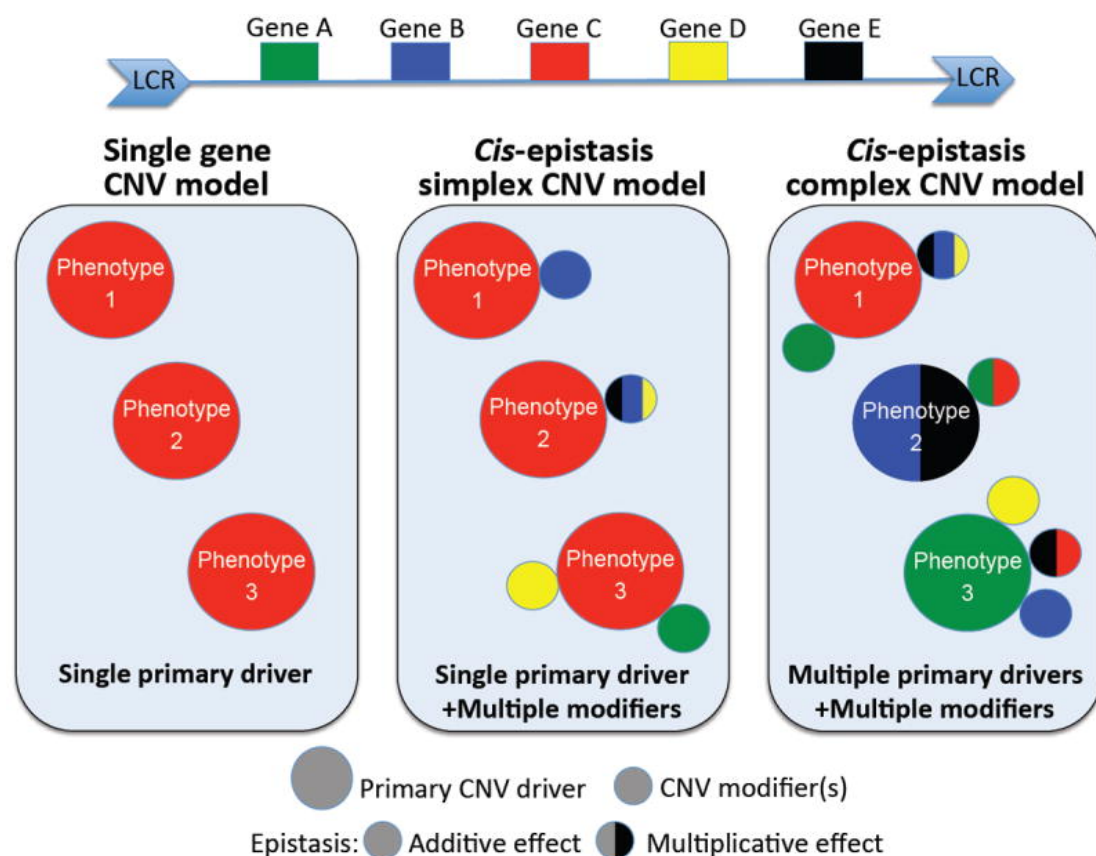
Elements (ENCODE) project moves forward, the investigation of the regulatory segments of the genome, including also 'non-coding' elements, together with the study of its 3D structure will provide additional information to be integrated in these studies, in order to assess their roles in human biology and disease.

One of the biggest challenges in the study of a copy number variant is that the affected genomic fragment in most cases encompasses several genes, exacerbating the problem of assigning causality to one or more specific transcripts within a CNV. Different models summarize the contribution of genes in a rare CNV to a certain phenotype(s)<sup>4,171</sup>. In the single gene model, one major driver gene can be associated to the phenotype (**Figure 4**); an example of this model is represented by *RAI1*'s contribution to SMS. The simplex and complex cis-epistatic model posit that one gene or numerous genes, respectively, are necessary and sufficient to determine phenotype, in epistasis with other genes within the CNV; in the complex model some of the involved genes drive specific endophenotypes while others trigger complex additive and/or multiplicative effects (**Figure 4**)<sup>4</sup>. In the contiguous gene model, each or several genes within the CNV contribute independently to the multiple unrelated features that can be displayed by the individuals carrying the CNV. In addition, the parent-of-origin model postulates a different effect on the phenotype, depending on whether the disruption occurs in the gene derived from maternal or parental chromosome<sup>171</sup>.

Animal models of human diseases are important for validation of pathogenicity, and help to uncover the candidate gene(s)/genomic regions for a certain phenotype, investigate the underlying biological and physiological mechanism and explore potential treatments. Among the other well-established models used for the functional evaluation of rare CNV disorders, including mouse and fruit fly, zebrafish has become increasingly attractive to study embryogenesis and organ development, since about 70% of the annotated zebrafish genes have orthologs in humans, and because of its short generation time, the ability to assess large populations, and the transparency of the embryos, which makes their observation and manipulation extremely easy<sup>172, 173</sup>. In the last few years, this model has become extremely popular for the study of copy number variants and

the functional assessment of the encompassed genes. A good example of successful prioritization of candidate genes is represented by Blaker-Lee's and Golzio's studies of the 16p11.2 600 kb BP4-BP5 rearrangements<sup>129, 174</sup>, which allowed, respectively, the identification of *ALDOA* and *KIF22* as deletion dosage sensors impacting brain and body morphology, and of *KCTD13*, in epistasis with *MAPK3* and *MVP*, as major driver of the head circumference phenotype.

**Chapter 2** of the Result section will focus on the functional characterization of the 16p11.2 220 kb BP2-BP3 locus and encompassed genes by using zebrafish as animal model.



**Figure 4. Some models to explain causality of disease owing to rare CNVs.** The figure represents a genomic segment flanked by LCRs (blue arrows) and encompassing five genes (Gene A-E). According to the “single gene CNV model” a single gene within the CNV is the primary driver of the observed phenotype, explaining 100% of its expressivity and penetrance; the “cis-epistasis” simplex and complex models, instead, allow the presence of one or more major driver genes contributing to the phenotype(s), whose expressivity and penetrance is modulated by one or several modifiers. Modified from <sup>4</sup>

Recent studies have shown that a big portion of the genome is, in one way or another, involved in gene regulation<sup>175</sup>. Transcription of co-regulated genes occurs in a context of proximal and distal chromatin looping interactions, which represent the basic organizing principle of the nuclear architecture and gene regulation<sup>176-178</sup>. On a large scale, the organization of chromosomes is not random, but influenced by gene density and transcriptional status<sup>179-181</sup>. It has also been revealed that, even if the interphase nuclei the chromosomes tend to occupy their own distinct special territories<sup>182</sup>, intermingling occurs at those regions where the chromosomes are in contact, creating the opportunity for likely functioning interchromosomal interactions<sup>183</sup>. Many studies have recently focused on the looping mechanisms of long-range regulatory elements, as the shape of the genome is described as a network of interactions between genomic elements, which can play an important part in the coordination of transcription and other DNA-metabolic processes<sup>184-186</sup>. The hypothesis is that the looping of the chromatin fibers can bring genes in close physical proximity to distal regulatory elements (DREs)<sup>184, 187</sup>. Therefore, looping is not only due to well-established chromosomal folding processes, but it requires protein-mediated contacts between regulatory elements and gene promoters. There is increasing evidence that this spatial connectivity between DREs and corresponding promoters can work over long distances<sup>187</sup> and even on different chromosomes<sup>188</sup>, coinciding with alterations in gene expression<sup>188-191</sup>. For example, an important limb enhancer (ZRS) of Sonic Hedgehog (*SHH*) is located more than 1Mb away from its transcription start site<sup>192, 193</sup>, and, more generally, only 7% of distal elements contact their closest promoter<sup>194</sup>. However, looping seems to be necessary, but not sufficient to drive expression, since most of DNA contacts are established long before gene expression occurs, mostly with a non transcription-dependent specificity and also in absence of regulatory elements<sup>195, 196</sup>. A better understanding of these processes is crucial to determine how gene expression is regulated and achieved, and whether structural perturbations, like copy number

variations, can produce a phenotype also because of their impact on these mechanisms.

Microscope-based approaches do not allow a comprehensive analysis of the organization of the chromosomes at kilobase resolution, but this limitation was passed by the development of the Chromosome Conformation Capture (3C) based technologies<sup>197, 198</sup>. The rationale behind these approaches is that the 3D arrangement of a genomic region, chromosome or whole genome can be inferred from the determination of a sufficient number of pairwise interaction frequencies<sup>199</sup>. In the 3C, cells are fixed with formaldehyde, in order to cross-link proteins to other proteins and to DNA segments that are close in the nuclear space. The cross-linked chromatin is then digested with an excess of restriction enzyme and ligated under dilute conditions that favor junctions between cross-linked DNA fragments, before reversal of the crosslinking and DNA purification. The detection of the interactions relies on the use of locus-specific PCR primers, for example with a forward anchor primer in a region of interest, such as an enhancer, and a reverse primer located at any position that it is desirable to interrogate<sup>197</sup>. This was originally done in studies investigating the  $\beta$ -globin locus, where the upstream LCR was shown to physically interact with the active globin genes, thereby looping out the intervening 30–50 kb of chromatin fiber<sup>186</sup>. The 4C (also known as “circular 3C” or “3C-on-chip”) uses instead inverse PCR to generate genome-wide interaction profile for a single locus<sup>200, 201</sup>. 5C and Hi-C methods are not anchored on a single locus as the first two; they generate matrices of interaction frequencies that can be visualized as two-dimensional heat maps with genomic positions along the two axes, providing “many-to-many” or “all-to-all”, respectively, genome-wide interaction maps<sup>179, 194, 202</sup>. Hi-C interaction maps are derived by using a population of cells and reflect average configuration of regulatory landscapes, yet in agreement with the observations made by single-cell Hi-C<sup>203</sup>. Nevertheless, predictive models’ observations suggest that these data represent the result of a huge number of transient contacts, which are extremely variable across cells<sup>204</sup>. Despite some limitations, Hi-C data provided a great amount of information to describe the general principles of spatial proximity, the presence of chromosome territories and the existence of suborders of chromosome organization at megabase scale,

characterized by a higher frequency of interactions with themselves than with the rest of the genome, known as Topological Associated Domains (TADs)<sup>205-207</sup>. The TADs are considered the fundamental structural unit of the genome; in these domains, which are conserved across species, cell type and tissues, gene co-regulation is observed<sup>206</sup> and several genomic features are highly correlated, including chromatin marks, DNA replication and lamina-associated domains (LADs)<sup>205, 207, 208</sup>.

Structural rearrangements of the genome have the potential to disrupt the TADs organization, for example promoting the shifting of regulatory elements between domains or modifying the position of boundaries, resulting in ectopic enhancer-promoter interactions, gene misexpression and disease<sup>209, 210</sup>. This needs to be taken into consideration when investigating copy number variants and assessing their contribution to the disease's onset.

In the light of growing evidence of the relevance of these chromatin substructures and regulatory looping, **Chapter 1** and **3** of the Results section will investigate the potential role of chromatin interactions in linking loci that, when rearranged, associate with similar phenotype and/or in positioning in close physical proximity genes in the same pathway/displaying similar functions.



## References

1. Xi R, Kim TM, Park PJ. Detecting structural variations in the human genome using next generation sequencing. *Brief Funct Genomics* 2010; **9**(5-6): 405-415.
2. Genomes Project C, Abecasis GR, Altshuler D, Auton A, Brooks LD, Durbin RM *et al.* A map of human genome variation from population-scale sequencing. *Nature* 2010; **467**(7319): 1061-1073.
3. Krumm N, Sudmant PH, Ko A, O'Roak BJ, Malig M, Coe BP *et al.* Copy number variation detection and genotyping from exome sequence data. *Genome Res* 2012; **22**(8): 1525-1532.
4. Golzio C, Katsanis N. Genetic architecture of reciprocal CNVs. *Curr Opin Genet Dev* 2013; **23**(3): 240-248.
5. Stankiewicz P, Lupski JR. Structural variation in the human genome and its role in disease. *Annu Rev Med* 2010; **61**: 437-455.
6. Redon R, Ishikawa S, Fitch KR, Feuk L, Perry GH, Andrews TD *et al.* Global variation in copy number in the human genome. *Nature* 2006; **444**(7118): 444-454.
7. Itsara A, Cooper GM, Baker C, Girirajan S, Li J, Absher D *et al.* Population analysis of large copy number variants and hotspots of human genetic disease. *American journal of human genetics* 2009; **84**(2): 148-161.
8. Shaikh TH, Gai X, Perin JC, Glessner JT, Xie H, Murphy K *et al.* High-resolution mapping and analysis of copy number variations in the human genome: a data resource for clinical and research applications. *Genome Res* 2009; **19**(9): 1682-1690.
9. Genomes Project C, Abecasis GR, Auton A, Brooks LD, DePristo MA, Durbin RM *et al.* An integrated map of genetic variation from 1,092 human genomes. *Nature* 2012; **491**(7422): 56-65.
10. Conrad DF, Pinto D, Redon R, Feuk L, Gokcumen O, Zhang Y *et al.* Origins and functional impact of copy number variation in the human genome. *Nature* 2010; **464**(7289): 704-712.
11. MacDonald JR, Ziman R, Yuen RK, Feuk L, Scherer SW. The Database of Genomic Variants: a curated collection of structural variation in the human genome. *Nucleic Acids Res* 2014; **42**(Database issue): D986-992.
12. Iafrate AJ, Feuk L, Rivera MN, Listewnik ML, Donahoe PK, Qi Y *et al.* Detection of large-scale variation in the human genome. *Nat Genet* 2004; **36**(9): 949-951.
13. Sebat J, Lakshmi B, Troge J, Alexander J, Young J, Lundin P *et al.* Large-scale copy number polymorphism in the human genome. *Science* 2004; **305**(5683): 525-528.

14. Zhang F, Gu W, Hurles ME, Lupski JR. Copy number variation in human health, disease, and evolution. *Annual review of genomics and human genetics* 2009; **10**: 451-481.
15. Mikhail FM. Copy number variations and human genetic disease. *Current opinion in pediatrics* 2014; **26**(6): 646-652.
16. Sharp AJ, Selzer RR, Veltman JA, Gimelli S, Gimelli G, Striano P *et al.* Characterization of a recurrent 15q24 microdeletion syndrome. *Hum Mol Genet* 2007; **16**(5): 567-572.
17. Koolen DA, Vissers LE, Pfundt R, de Leeuw N, Knight SJ, Regan R *et al.* A new chromosome 17q21.31 microdeletion syndrome associated with a common inversion polymorphism. *Nat Genet* 2006; **38**(9): 999-1001.
18. Shaw-Smith C, Pittman AM, Willatt L, Martin H, Rickman L, Gribble S *et al.* Microdeletion encompassing MAPT at chromosome 17q21.3 is associated with developmental delay and learning disability. *Nat Genet* 2006; **38**(9): 1032-1037.
19. Lupski JR. Genomic rearrangements and sporadic disease. *Nat Genet* 2007; **39**(7 Suppl): S43-47.
20. Eichler EE, Clark RA, She XW. An assessment of the sequence gaps: Unfinished business in a finished human genome. *Nature Reviews Genetics* 2004; **5**(5): 345-354.
21. Stankiewicz P, Lupski JR. Genome architecture, rearrangements and genomic disorders. *Trends Genet* 2002; **18**(2): 74-82.
22. Gu W, Zhang F, Lupski JR. Mechanisms for human genomic rearrangements. *PathoGenetics* 2008; **1**(1): 4.
23. Stankiewicz P, Inoue K, Bi W, Walz K, Park SS, Kurotaki N *et al.* Genomic disorders: genome architecture results in susceptibility to DNA rearrangements causing common human traits. *Cold Spring Harbor symposia on quantitative biology* 2003; **68**: 445-454.
24. Chen L, Zhou W, Zhang L, Zhang F. Genome architecture and its roles in human copy number variation. *Genomics Inform* 2014; **12**(4): 136-144.
25. Weterings E, van Gent DC. The mechanism of non-homologous end-joining: a synopsis of synapsis. *DNA repair* 2004; **3**(11): 1425-1435.
26. Lee JA, Carvalho CM, Lupski JR. A DNA replication mechanism for generating nonrecurrent rearrangements associated with genomic disorders. *Cell* 2007; **131**(7): 1235-1247.
27. Zhang F, Khajavi M, Connolly AM, Towne CF, Batish SD, Lupski JR. The DNA replication FoSTeS/MMBIR mechanism can generate genomic, genic and exonic complex rearrangements in humans. *Nat Genet* 2009; **41**(7): 849-853.

28. Lupski JR. Genomic disorders ten years on. *Genome Med* 2009; **1**(4): 42.
29. Lupski JR. Genomic disorders: structural features of the genome can lead to DNA rearrangements and human disease traits. *Trends Genet* 1998; **14**(10): 417-422.
30. Kurotaki N, Stankiewicz P, Wakui K, Niikawa N, Lupski JR. Sotos syndrome common deletion is mediated by directly oriented subunits within inverted Sos-REP low-copy repeats. *Hum Mol Genet* 2005; **14**(4): 535-542.
31. Lupski JR, Stankiewicz P. Genomic disorders: molecular mechanisms for rearrangements and conveyed phenotypes. *PLoS genetics* 2005; **1**(6): e49.
32. Torres F, Barbosa M, Maciel P. Recurrent copy number variations as risk factors for neurodevelopmental disorders: critical overview and analysis of clinical implications. *J Med Genet* 2016; **53**(2): 73-90.
33. Lewis DA, Levitt P. Schizophrenia as a disorder of neurodevelopment. *Annual review of neuroscience* 2002; **25**: 409-432.
34. Lee JA, Lupski JR. Genomic rearrangements and gene copy-number alterations as a cause of nervous system disorders. *Neuron* 2006; **52**(1): 103-121.
35. Grayton HM, Fernandes C, Rujescu D, Collier DA. Copy number variations in neurodevelopmental disorders. *Progress in neurobiology* 2012; **99**(1): 81-91.
36. Ford CE, Hamerton JL. The chromosomes of man. *Nature* 1956; **178**(4541): 1020-1023.
37. Caspersson T, Zech L, Johansson C. Differential binding of alkylating fluorochromes in human chromosomes. *Experimental cell research* 1970; **60**(3): 315-319.
38. Drets ME, Shaw MW. Specific banding patterns of human chromosomes. *Proceedings of the National Academy of Sciences of the United States of America* 1971; **68**(9): 2073-2077.
39. Patil SR, Lubs HA. Non-random association of human acrocentric chromosomes. *Humangenetik* 1971; **13**(2): 157-159.
40. Schnedl W. Analysis of the human karyotype using a reassociation technique. *Chromosoma* 1971; **34**(4): 448-454.
41. Seabright M. A rapid banding technique for human chromosomes. *Lancet* 1971; **2**(7731): 971-972.
42. Sumner AT, Evans HJ, Buckland RA. New technique for distinguishing between human chromosomes. *Nature: New biology* 1971; **232**(27): 31-32.

43. Vissers LE, de Vries BB, Osoegawa K, Janssen IM, Feuth T, Choy CO *et al.* Array-based comparative genomic hybridization for the genomewide detection of submicroscopic chromosomal abnormalities. *American journal of human genetics* 2003; **73**(6): 1261-1270.
44. Cook EH, Jr., Scherer SW. Copy-number variations associated with neuropsychiatric conditions. *Nature* 2008; **455**(7215): 919-923.
45. Slavotinek AM. Novel microdeletion syndromes detected by chromosome microarrays. *Human genetics* 2008; **124**(1): 1-17.
46. Ramocki MB, Zoghbi HY. Failure of neuronal homeostasis results in common neuropsychiatric phenotypes. *Nature* 2008; **455**(7215): 912-918.
47. Mills RE, Walter K, Stewart C, Handsaker RE, Chen K, Alkan C *et al.* Mapping copy number variation by population-scale genome sequencing. *Nature* 2011; **470**(7332): 59-65.
48. Nord AS, Lee M, King MC, Walsh T. Accurate and exact CNV identification from targeted high-throughput sequence data. *BMC genomics* 2011; **12**: 184.
49. Coe BP, Girirajan S, Eichler EE. The genetic variability and commonality of neurodevelopmental disease. *American journal of medical genetics Part C, Seminars in medical genetics* 2012; **160C**(2): 118-129.
50. Vissers LE, Gilissen C, Veltman JA. Genetic studies in intellectual disability and related disorders. *Nature reviews Genetics* 2016; **17**(1): 9-18.
51. de Ligt J, Willemsen MH, van Bon BW, Kleefstra T, Yntema HG, Kroes T *et al.* Diagnostic exome sequencing in persons with severe intellectual disability. *The New England journal of medicine* 2012; **367**(20): 1921-1929.
52. Deciphering Developmental Disorders S. Large-scale discovery of novel genetic causes of developmental disorders. *Nature* 2015; **519**(7542): 223-228.
53. Gilissen C, Hehir-Kwa JY, Thung DT, van de Vorst M, van Bon BW, Willemsen MH *et al.* Genome sequencing identifies major causes of severe intellectual disability. *Nature* 2014; **511**(7509): 344-347.
54. Hamdan FF, Gauthier J, Araki Y, Lin DT, Yoshizawa Y, Higashi K *et al.* Excess of de novo deleterious mutations in genes associated with glutamatergic systems in nonsyndromic intellectual disability. *American journal of human genetics* 2011; **88**(3): 306-316.
55. Rauch A, Wieczorek D, Graf E, Wieland T, Ende S, Schwarzmayr T *et al.* Range of genetic mutations associated with severe non-syndromic sporadic intellectual disability: an exome sequencing study. *Lancet* 2012; **380**(9854): 1674-1682.
56. Vissers LE, de Ligt J, Gilissen C, Janssen I, Steehouwer M, de Vries P *et al.* A de novo paradigm for mental retardation. *Nat Genet* 2010; **42**(12): 1109-1112.

57. Yang Y, Muzny DM, Reid JG, Bainbridge MN, Willis A, Ward PA *et al.* Clinical whole-exome sequencing for the diagnosis of mendelian disorders. *The New England journal of medicine* 2013; **369**(16): 1502-1511.
58. Girirajan S, Rosenfeld JA, Cooper GM, Antonacci F, Siswara P, Itsara A *et al.* A recurrent 16p12.1 microdeletion supports a two-hit model for severe developmental delay. *Nat Genet* 2010; **42**(3): 203-209.
59. Michelson DJ, Shevell MI, Sherr EH, Moeschler JB, Gropman AL, Ashwal S. Evidence report: Genetic and metabolic testing on children with global developmental delay: report of the Quality Standards Subcommittee of the American Academy of Neurology and the Practice Committee of the Child Neurology Society. *Neurology* 2011; **77**(17): 1629-1635.
60. van Karnebeek CD, Jansweijer MC, Leenders AG, Offringa M, Hennekam RC. Diagnostic investigations in individuals with mental retardation: a systematic literature review of their usefulness. *European journal of human genetics : EJHG* 2005; **13**(1): 6-25.
61. de Vries BB, Pfundt R, Leisink M, Koolen DA, Vissers LE, Janssen IM *et al.* Diagnostic genome profiling in mental retardation. *American journal of human genetics* 2005; **77**(4): 606-616.
62. Shaw-Smith C, Redon R, Rickman L, Rio M, Willatt L, Fiegler H *et al.* Microarray based comparative genomic hybridisation (array-CGH) detects submicroscopic chromosomal deletions and duplications in patients with learning disability/mental retardation and dysmorphic features. *J Med Genet* 2004; **41**(4): 241-248.
63. Wagenstaller J, Spranger S, Lorenz-Depiereux B, Kazmierczak B, Nathrath M, Wahl D *et al.* Copy-number variations measured by single-nucleotide-polymorphism oligonucleotide arrays in patients with mental retardation. *American journal of human genetics* 2007; **81**(4): 768-779.
64. Lubs HA, Stevenson RE, Schwartz CE. Fragile X and X-linked intellectual disability: four decades of discovery. *American journal of human genetics* 2012; **90**(4): 579-590.
65. Samocha KE, Robinson EB, Sanders SJ, Stevens C, Sabo A, McGrath LM *et al.* A framework for the interpretation of de novo mutation in human disease. *Nat Genet* 2014; **46**(9): 944-950.
66. Coe BP, Witherspoon K, Rosenfeld JA, van Bon BW, Vulto-van Silfhout AT, Bosco P *et al.* Refining analyses of copy number variation identifies specific genes associated with developmental delay. *Nat Genet* 2014; **46**(10): 1063-1071.
67. van Bokhoven H. Genetic and epigenetic networks in intellectual disabilities. *Annual review of genetics* 2011; **45**: 81-104.
68. Cristino AS, Williams SM, Hawi Z, An JY, Bellgrove MA, Schwartz CE *et al.* Neurodevelopmental and neuropsychiatric disorders represent an interconnected molecular system. *Molecular psychiatry* 2014; **19**(3): 294-301.

69. Hormozdiari F, Penn O, Borenstein E, Eichler EE. The discovery of integrated gene networks for autism and related disorders. *Genome Res* 2015; **25**(1): 142-154.
70. Cardno AG, Gottesman, II. Twin studies of schizophrenia: from bow-and-arrow concordances to star wars Mx and functional genomics. *American journal of medical genetics* 2000; **97**(1): 12-17.
71. International Schizophrenia C. Rare chromosomal deletions and duplications increase risk of schizophrenia. *Nature* 2008; **455**(7210): 237-241.
72. Rees E, Walters JT, Georgieva L, Isles AR, Chambert KD, Richards AL *et al.* Analysis of copy number variations at 15 schizophrenia-associated loci. *The British journal of psychiatry : the journal of mental science* 2014; **204**(2): 108-114.
73. Walsh T, McClellan JM, McCarthy SE, Addington AM, Pierce SB, Cooper GM *et al.* Rare structural variants disrupt multiple genes in neurodevelopmental pathways in schizophrenia. *Science* 2008; **320**(5875): 539-543.
74. Purcell SM, Moran JL, Fromer M, Ruderfer D, Solovieff N, Roussos P *et al.* A polygenic burden of rare disruptive mutations in schizophrenia. *Nature* 2014; **506**(7487): 185-190.
75. Kirov G, Pocklington AJ, Holmans P, Ivanov D, Ikeda M, Ruderfer D *et al.* De novo CNV analysis implicates specific abnormalities of postsynaptic signalling complexes in the pathogenesis of schizophrenia. *Molecular psychiatry* 2012; **17**(2): 142-153.
76. Malhotra D, McCarthy S, Michaelson JJ, Vacic V, Burdick KE, Yoon S *et al.* High frequencies of de novo CNVs in bipolar disorder and schizophrenia. *Neuron* 2011; **72**(6): 951-963.
77. Xu B, Roos JL, Levy S, van Rensburg EJ, Gogos JA, Karayiorgou M. Strong association of de novo copy number mutations with sporadic schizophrenia. *Nat Genet* 2008; **40**(7): 880-885.
78. Pocklington AJ, Rees E, Walters JT, Han J, Kavanagh DH, Chambert KD *et al.* Novel Findings from CNVs Implicate Inhibitory and Excitatory Signaling Complexes in Schizophrenia. *Neuron* 2015; **86**(5): 1203-1214.
79. Autism, Developmental Disabilities Monitoring Network Surveillance Year Principal I, Centers for Disease C, Prevention. Prevalence of autism spectrum disorders--Autism and Developmental Disabilities Monitoring Network, 14 sites, United States, 2008. *Morbidity and mortality weekly report Surveillance summaries* 2012; **61**(3): 1-19.
80. Shishido E, Aleksic B, Ozaki N. Copy-number variation in the pathogenesis of autism spectrum disorder. *Psychiatry and clinical neurosciences* 2014; **68**(2): 85-95.
81. Malhotra D, Sebat J. CNVs: harbingers of a rare variant revolution in psychiatric genetics. *Cell* 2012; **148**(6): 1223-1241.

82. Bodmer W, Bonilla C. Common and rare variants in multifactorial susceptibility to common diseases. *Nat Genet* 2008; **40**(6): 695-701.
83. Iossifov I, O'Roak BJ, Sanders SJ, Ronemus M, Krumm N, Levy D *et al.* The contribution of de novo coding mutations to autism spectrum disorder. *Nature* 2014; **515**(7526): 216-221.
84. Iossifov I, Ronemus M, Levy D, Wang Z, Hakker I, Rosenbaum J *et al.* De novo gene disruptions in children on the autistic spectrum. *Neuron* 2012; **74**(2): 285-299.
85. De Rubeis S, He X, Goldberg AP, Poultney CS, Samocha K, Cicek AE *et al.* Synaptic, transcriptional and chromatin genes disrupted in autism. *Nature* 2014; **515**(7526): 209-215.
86. Neale BM, Kou Y, Liu L, Ma'ayan A, Samocha KE, Sabo A *et al.* Patterns and rates of exonic de novo mutations in autism spectrum disorders. *Nature* 2012; **485**(7397): 242-245.
87. O'Roak BJ, Vives L, Girirajan S, Karakoc E, Krumm N, Coe BP *et al.* Sporadic autism exomes reveal a highly interconnected protein network of de novo mutations. *Nature* 2012; **485**(7397): 246-250.
88. Sanders SJ, Ercan-Sencicek AG, Hus V, Luo R, Murtha MT, Moreno-De-Luca D *et al.* Multiple recurrent de novo CNVs, including duplications of the 7q11.23 Williams syndrome region, are strongly associated with autism. *Neuron* 2011; **70**(5): 863-885.
89. Sanders SJ, Murtha MT, Gupta AR, Murdoch JD, Raubeson MJ, Willsey AJ *et al.* De novo mutations revealed by whole-exome sequencing are strongly associated with autism. *Nature* 2012; **485**(7397): 237-241.
90. Sebat J, Lakshmi B, Malhotra D, Troge J, Lese-Martin C, Walsh T *et al.* Strong association of de novo copy number mutations with autism. *Science* 2007; **316**(5823): 445-449.
91. Glessner JT, Wang K, Cai G, Korvatska O, Kim CE, Wood S *et al.* Autism genome-wide copy number variation reveals ubiquitin and neuronal genes. *Nature* 2009; **459**(7246): 569-573.
92. Pinto D, Delaby E, Merico D, Barbosa M, Merikangas A, Klei L *et al.* Convergence of genes and cellular pathways dysregulated in autism spectrum disorders. *American journal of human genetics* 2014; **94**(5): 677-694.
93. Gaugler T, Klei L, Sanders SJ, Bodea CA, Goldberg AP, Lee AB *et al.* Most genetic risk for autism resides with common variation. *Nat Genet* 2014; **46**(8): 881-885.
94. Marshall CR, Noor A, Vincent JB, Lionel AC, Feuk L, Skaug J *et al.* Structural variation of chromosomes in autism spectrum disorder. *American journal of human genetics* 2008; **82**(2): 477-488.

95. Levy D, Ronemus M, Yamrom B, Lee YH, Leotta A, Kendall J *et al.* Rare de novo and transmitted copy-number variation in autistic spectrum disorders. *Neuron* 2011; **70**(5): 886-897.
96. Devlin B, Scherer SW. Genetic architecture in autism spectrum disorder. *Curr Opin Genet Dev* 2012; **22**(3): 229-237.
97. Pinto D, Pagnamenta AT, Klei L, Anney R, Merico D, Regan R *et al.* Functional impact of global rare copy number variation in autism spectrum disorders. *Nature* 2010; **466**(7304): 368-372.
98. Cooper GM, Coe BP, Girirajan S, Rosenfeld JA, Vu TH, Baker C *et al.* A copy number variation morbidity map of developmental delay. *Nature genetics* 2011; **43**(9): 838-846.
99. Liu X, Takumi T. Genomic and genetic aspects of autism spectrum disorder. *Biochemical and biophysical research communications* 2014; **452**(2): 244-253.
100. Johnson ME, Viggiano L, Bailey JA, Abdul-Rauf M, Goodwin G, Rocchi M *et al.* Positive selection of a gene family during the emergence of humans and African apes. *Nature* 2001; **413**(6855): 514-519.
101. Antonacci F, Kidd JM, Marques-Bonet T, Teague B, Ventura M, Girirajan S *et al.* A large and complex structural polymorphism at 16p12.1 underlies microdeletion disease risk. *Nature genetics* 2010; **42**(9): 745-750.
102. Hastings PJ, Lupski JR, Rosenberg SM, Ira G. Mechanisms of change in gene copy number. *Nature reviews Genetics* 2009; **10**(8): 551-564.
103. Ballif BC, Hornor SA, Jenkins E, Madan-Khetarpal S, Surti U, Jackson KE *et al.* Discovery of a previously unrecognized microdeletion syndrome of 16p11.2-p12.2. *Nat Genet* 2007; **39**(9): 1071-1073.
104. Battaglia A, Novelli A, Bernardini L, Iglizzi R, Parrini B. Further characterization of the new microdeletion syndrome of 16p11.2-p12.2. *American journal of medical genetics Part A* 2009; **149A**(6): 1200-1204.
105. Hempel M, Rivera Brugues N, Wagenstaller J, Lederer G, Weitensteiner A, Seidel H *et al.* Microdeletion syndrome 16p11.2-p12.2: clinical and molecular characterization. *American journal of medical genetics Part A* 2009; **149A**(10): 2106-2112.
106. Hernando C, Plaja A, Rigola MA, Perez MM, Vendrell T, Egocue J *et al.* Comparative genomic hybridisation shows a partial de novo deletion 16p11.2 in a neonate with multiple congenital malformations. *J Med Genet* 2002; **39**(5): E24.
107. Zufferey F, Sherr EH, Beckmann ND, Hanson E, Maillard AM, Hippolyte L *et al.* A 600 kb deletion syndrome at 16p11.2 leads to energy imbalance and neuropsychiatric disorders. *J Med Genet* 2012; **49**(10): 660-668.



108. Jacquemont S, Reymond A, Zufferey F, Harewood L, Walters RG, Kutalik Z *et al.* Mirror extreme BMI phenotypes associated with gene dosage at the chromosome 16p11.2 locus. *Nature* 2011; **478**(7367): 97-102.
109. Hanson E, Bernier R, Porche K, Jackson FI, Goin-Kochel RP, Snyder LG *et al.* The cognitive and behavioral phenotype of the 16p11.2 deletion in a clinically ascertained population. *Biological psychiatry* 2015; **77**(9): 785-793.
110. Maillard AM, Ruef A, Pizzagalli F, Migliavacca E, Hippolyte L, Adaszewski S *et al.* The 16p11.2 locus modulates brain structures common to autism, schizophrenia and obesity. *Molecular psychiatry* 2015; **20**(1): 140-147.
111. Benninger MS, Appelbaum PC, Denny JC, Osguthorpe DJ, Stankiewicz JA. Maxillary sinus puncture and culture in the diagnosis of acute rhinosinusitis: the case for pursuing alternative culture methods. *Otolaryngol Head Neck Surg* 2002; **127**(1): 7-12.
112. Walters RG, Jacquemont S, Valsesia A, de Smith AJ, Martinet D, Andersson J *et al.* A new highly penetrant form of obesity due to deletions on chromosome 16p11.2. *Nature* 2010; **463**(7281): 671-675.
113. Shinawi M, Liu P, Kang SH, Shen J, Belmont JW, Scott DA *et al.* Recurrent reciprocal 16p11.2 rearrangements associated with global developmental delay, behavioural problems, dysmorphism, epilepsy, and abnormal head size. *J Med Genet* 2010; **47**(5): 332-341.
114. Reinthaler EM, Lal D, Lebon S, Hildebrand MS, Dahl HH, Regan BM *et al.* 16p11.2 600 kb Duplications confer risk for typical and atypical Rolandic epilepsy. *Hum Mol Genet* 2014; **23**(22): 6069-6080.
115. Bassuk AG, Geraghty E, Wu S, Mullen SA, Berkovic SF, Scheffer IE *et al.* Deletions of 16p11.2 and 19p13.2 in a family with intellectual disability and generalized epilepsy. *American journal of medical genetics Part A* 2013; **161A**(7): 1722-1725.
116. Bijlsma EK, Gijsbers AC, Schuurs-Hoeijmakers JH, van Haeringen A, Fransen van de Putte DE, Anderlid BM *et al.* Extending the phenotype of recurrent rearrangements of 16p11.2: deletions in mentally retarded patients without autism and in normal individuals. *European journal of medical genetics* 2009; **52**(2-3): 77-87.
117. Ciuladaite Z, Kasnauskiene J, Cimbalistiene L, Preiksaitiene E, Patsalis PC, Kucinskas V. Mental retardation and autism associated with recurrent 16p11.2 microdeletion: incomplete penetrance and variable expressivity. *Journal of applied genetics* 2011; **52**(4): 443-449.
118. Hanson E, Nasir RH, Fong A, Lian A, Hundley R, Shen Y *et al.* Cognitive and behavioral characterization of 16p11.2 deletion syndrome. *Journal of developmental and behavioral pediatrics : JDBP* 2010; **31**(8): 649-657.
119. Puvabanditsin S, Nagar MS, Joshi M, Lambert G, Garrow E, Brandsma E. Microdeletion of 16p11.2 associated with endocardial fibroelastosis. *American journal of medical genetics Part A* 2010; **152A**(9): 2383-2386.

120. Schaaf CP, Goin-Kochel RP, Nowell KP, Hunter JV, Aleck KA, Cox S *et al.* Expanding the clinical spectrum of the 16p11.2 chromosomal rearrangements: three patients with syringomyelia. *European journal of human genetics : EJHG* 2011; **19**(2): 152-156.
121. Shioh LR, Paris K, Akana MC, Cyster JG, Sorensen RU, Puck JM. Severe combined immunodeficiency (SCID) and attention deficit hyperactivity disorder (ADHD) associated with a Coronin-1A mutation and a chromosome 16p11.2 deletion. *Clinical immunology* 2009; **131**(1): 24-30.
122. Steinberg S, de Jong S, Mattheisen M, Costas J, Demontis D, Jamain S *et al.* Common variant at 16p11.2 conferring risk of psychosis. *Molecular psychiatry* 2014; **19**(1): 108-114.
123. Rosenberg C, Freitas EL, Uehara DT, Auricchio MT, Costa SS, Oiticica J *et al.* Short Report - Clinical Genetics Genomic copy number alterations in non-syndromic hearing loss. *Clinical genetics* 2015.
124. Qureshi AY, Mueller S, Snyder AZ, Mukherjee P, Berman JI, Roberts TP *et al.* Opposing brain differences in 16p11.2 deletion and duplication carriers. *The Journal of neuroscience : the official journal of the Society for Neuroscience* 2014; **34**(34): 11199-11211.
125. D'Angelo D, Lebon S, Chen Q, Martin-Brevet S, Snyder LG, Hippolyte L *et al.* Defining the Effect of the 16p11.2 Duplication on Cognition, Behavior, and Medical Comorbidities. *JAMA Psychiatry* 2016; **73**(1): 20-30.
126. Maillard AM, Hippolyte L, Rodriguez-Herrerros B, Chawner SJ, Dremmel D, Aguera Z *et al.* 16p11.2 Locus modulates response to satiety before the onset of obesity. *International journal of obesity* 2015.
127. Duyzend MH, Nettle X, Coe BP, Baker C, Nickerson DA, Bernier R *et al.* Maternal Modifiers and Parent-of-Origin Bias of the Autism-Associated 16p11.2 CNV. *American journal of human genetics* 2016; **98**(1): 45-57.
128. Crepel A, Steyaert J, De la Marche W, De Wolf V, Fryns JP, Noens I *et al.* Narrowing the critical deletion region for autism spectrum disorders on 16p11.2. *Am J Med Genet B Neuropsychiatr Genet* 2011; **156**(2): 243-245.
129. Golzio C, Willer J, Talkowski ME, Oh EC, Taniguchi Y, Jacquemont S *et al.* KCTD13 is a major driver of mirrored neuroanatomical phenotypes of the 16p11.2 copy number variant. *Nature* 2012; **485**(7398): 363-367.
130. Lin GN, Corominas R, Lemmens I, Yang X, Tavernier J, Hill DE *et al.* Spatiotemporal 16p11.2 protein network implicates cortical late mid-fetal brain development and KCTD13-Cul3-RhoA pathway in psychiatric diseases. *Neuron* 2015; **85**(4): 742-754.
131. Sparrow DB, Chapman G, Smith AJ, Mattar MZ, Major JA, O'Reilly VC *et al.* A mechanism for gene-environment interaction in the etiology of congenital scoliosis. *Cell* 2012; **149**(2): 295-306.

132. White PH, Farkas DR, McFadden EE, Chapman DL. Defective somite patterning in mouse embryos with reduced levels of Tbx6. *Development* 2003; **130**(8): 1681-1690.
133. Wu N, Ming X, Xiao J, Wu Z, Chen X, Shinawi M *et al.* TBX6 null variants and a common hypomorphic allele in congenital scoliosis. *The New England journal of medicine* 2015; **372**(4): 341-350.
134. Horev G, Ellegood J, Lerch JP, Son YE, Muthuswamy L, Vogel H *et al.* Dosage-dependent phenotypes in models of 16p11.2 lesions found in autism. *Proceedings of the National Academy of Sciences of the United States of America* 2011; **108**(41): 17076-17081.
135. Portmann T, Yang M, Mao R, Panagiotakos G, Ellegood J, Dolen G *et al.* Behavioral abnormalities and circuit defects in the basal ganglia of a mouse model of 16p11.2 deletion syndrome. *Cell reports* 2014; **7**(4): 1077-1092.
136. Arbogast T, Ouagazzal AM, Chevalier C, Kopanitsa M, Afinowi N, Migliavacca E *et al.* Reciprocal Effects on Neurocognitive and Metabolic Phenotypes in Mouse Models of 16p11.2 Deletion and Duplication Syndromes. *PLoS genetics* 2016; **12**(2): e1005709.
137. Yang M, Mahrt EJ, Lewis F, Foley G, Portmann T, Dolmetsch RE *et al.* 16p11.2 Deletion Syndrome Mice Display Sensory and Ultrasonic Vocalization Deficits During Social Interactions. *Autism research : official journal of the International Society for Autism Research* 2015; **8**(5): 507-521.
138. Pucilowska J, Puzerey PA, Karlo JC, Galan RF, Landreth GE. Disrupted ERK signaling during cortical development leads to abnormal progenitor proliferation, neuronal and network excitability and behavior, modeling human neuro-cardio-facial-cutaneous and related syndromes. *The Journal of neuroscience : the official journal of the Society for Neuroscience* 2012; **32**(25): 8663-8677.
139. Pucilowska J, Vithayathil J, Tavares EJ, Kelly C, Karlo JC, Landreth GE. The 16p11.2 deletion mouse model of autism exhibits altered cortical progenitor proliferation and brain cytoarchitecture linked to the ERK MAPK pathway. *The Journal of neuroscience : the official journal of the Society for Neuroscience* 2015; **35**(7): 3190-3200.
140. Thorleifsson G, Walters GB, Gudbjartsson DF, Steinthorsdottir V, Sulem P, Helgadóttir A *et al.* Genome-wide association yields new sequence variants at seven loci that associate with measures of obesity. *Nat Genet* 2009; **41**(1): 18-24.
141. Willer CJ, Speliotes EK, Loos RJ, Li S, Lindgren CM, Heid IM *et al.* Six new loci associated with body mass index highlight a neuronal influence on body weight regulation. *Nat Genet* 2009; **41**(1): 25-34.
142. Speliotes EK, Willer CJ, Berndt SI, Monda KL, Thorleifsson G, Jackson AU *et al.* Association analyses of 249,796 individuals reveal 18 new loci associated with body mass index. *Nat Genet* 2010; **42**(11): 937-948.

143. Bochukova EG, Huang N, Keogh J, Henning E, Purmann C, Blaszczyk K *et al.* Large, rare chromosomal deletions associated with severe early-onset obesity. *Nature* 2010; **463**(7281): 666-670.
144. Bachmann-Gagescu R, Mefford HC, Cowan C, Glew GM, Hing AV, Wallace S *et al.* Recurrent 200-kb deletions of 16p11.2 that include the SH2B1 gene are associated with developmental delay and obesity. *Genetics in medicine : official journal of the American College of Medical Genetics* 2010; **12**(10): 641-647.
145. Guha S, Rees E, Darvasi A, Ivanov D, Ikeda M, Bergen SE *et al.* Implication of a rare deletion at distal 16p11.2 in schizophrenia. *JAMA psychiatry* 2013; **70**(3): 253-260.
146. Ren D, Li M, Duan C, Rui L. Identification of SH2-B as a key regulator of leptin sensitivity, energy balance, and body weight in mice. *Cell Metab* 2005; **2**(2): 95-104.
147. Smith ACM, Boyd KE, Elsea SH, Finucane BM, Haas-Givler B, Gropman A *et al.* Smith-Magenis Syndrome. In: Pagon RA, Adam MP, Ardinger HH, Wallace SE, Amemiya A, Bean LJH *et al.* (eds). *GeneReviews(R)*: Seattle (WA), 1993.
148. Potocki L, Chen KS, Park SS, Osterholm DE, Withers MA, Kimonis V *et al.* Molecular mechanism for duplication 17p11.2- the homologous recombination reciprocal of the Smith-Magenis microdeletion. *Nat Genet* 2000; **24**(1): 84-87.
149. Elsea SH, Girirajan S. Smith-Magenis syndrome. *European journal of human genetics : EJHG* 2008; **16**(4): 412-421.
150. Smith AC, Dykens E, Greenberg F. Sleep disturbance in Smith-Magenis syndrome (del 17 p11.2). *American journal of medical genetics* 1998; **81**(2): 186-191.
151. Potocki L, Glaze D, Tan DX, Park SS, Kashork CD, Shaffer LG *et al.* Circadian rhythm abnormalities of melatonin in Smith-Magenis syndrome. *J Med Genet* 2000; **37**(6): 428-433.
152. De Leersnyder H, De Blois MC, Claustrat B, Romana S, Albrecht U, Von Kleist-Retzow JC *et al.* Inversion of the circadian rhythm of melatonin in the Smith-Magenis syndrome. *The Journal of pediatrics* 2001; **139**(1): 111-116.
153. De Leersnyder H, de Blois MC, Bresson JL, Sidi D, Claustrat B, Munnich A. [Inversion of the circadian melatonin rhythm in Smith-Magenis syndrome]. *Revue neurologique* 2003; **159**(11 Suppl): 6S21-26.
154. Laje G, Morse R, Richter W, Ball J, Pao M, Smith AC. Autism spectrum features in Smith-Magenis syndrome. *American journal of medical genetics Part C, Seminars in medical genetics* 2010; **154C**(4): 456-462.
155. Moss J, Oliver C, Arron K, Burbidge C, Berg K. The prevalence and phenomenology of repetitive behavior in genetic syndromes. *Journal of autism and developmental disorders* 2009; **39**(4): 572-588.

156. Dykens EM, Finucane BM, Gayley C. Brief report: cognitive and behavioral profiles in persons with Smith-Magenis syndrome. *Journal of autism and developmental disorders* 1997; **27**(2): 203-211.
157. Dykens EM, Smith AC. Distinctiveness and correlates of maladaptive behaviour in children and adolescents with Smith-Magenis syndrome. *Journal of intellectual disability research : JIDR* 1998; **42 ( Pt 6)**: 481-489.
158. Finucane BM, Konar D, Haas-Givler B, Kurtz MB, Scott CI, Jr. The spasmodic upper-body squeeze: a characteristic behavior in Smith-Magenis syndrome. *Developmental medicine and child neurology* 1994; **36**(1): 78-83.
159. Sloneem J, Oliver C, Udwin O, Woodcock KA. Prevalence, phenomenology, aetiology and predictors of challenging behaviour in Smith-Magenis syndrome. *Journal of intellectual disability research : JIDR* 2011; **55**(2): 138-151.
160. Bi W, Saifi GM, Shaw CJ, Walz K, Fonseca P, Wilson M *et al*. Mutations of RAI1, a PHD-containing protein, in nondeletion patients with Smith-Magenis syndrome. *Human genetics* 2004; **115**(6): 515-524.
161. Girirajan S, Elsas LJ, 2nd, Devriendt K, Elsea SH. RAI1 variations in Smith-Magenis syndrome patients without 17p11.2 deletions. *J Med Genet* 2005; **42**(11): 820-828.
162. Slager RE, Newton TL, Vlangos CN, Finucane B, Elsea SH. Mutations in RAI1 associated with Smith-Magenis syndrome. *Nat Genet* 2003; **33**(4): 466-468.
163. Vieira GH, Rodriguez JD, Carmona-Mora P, Cao L, Gamba BF, Carvalho DR *et al*. Detection of classical 17p11.2 deletions, an atypical deletion and RAI1 alterations in patients with features suggestive of Smith-Magenis syndrome. *European journal of human genetics : EJHG* 2012; **20**(2): 148-154.
164. Imai Y, Suzuki Y, Matsui T, Tohyama M, Wanaka A, Takagi T. Cloning of a retinoic acid-induced gene, GT1, in the embryonal carcinoma cell line P19: neuron-specific expression in the mouse brain. *Brain research Molecular brain research* 1995; **31**(1-2): 1-9.
165. Toulouse A, Rochefort D, Roussel J, Joobert R, Rouleau GA. Molecular cloning and characterization of human RAI1, a gene associated with schizophrenia. *Genomics* 2003; **82**(2): 162-171.
166. Williams SR, Zies D, Mullegama SV, Grotewiel MS, Elsea SH. Smith-Magenis syndrome results in disruption of CLOCK gene transcription and reveals an integral role for RAI1 in the maintenance of circadian rhythmicity. *American journal of human genetics* 2012; **90**(6): 941-949.
167. Mullegama SV, Pugliesi L, Burns B, Shah Z, Tahir R, Gu Y *et al*. MBD5 haploinsufficiency is associated with sleep disturbance and disrupts circadian pathways common to Smith-Magenis and fragile X syndromes. *European journal of human genetics : EJHG* 2015; **23**(6): 781-789.

168. Williams SR, Aldred MA, Der Kaloustian VM, Halal F, Gowans G, McLeod DR *et al.* Haploinsufficiency of HDAC4 causes brachydactyly mental retardation syndrome, with brachydactyly type E, developmental delays, and behavioral problems. *American journal of human genetics* 2010; **87**(2): 219-228.
169. Williams SR, Girirajan S, Tegay D, Nowak N, Hatchwell E, Elsea SH. Array comparative genomic hybridisation of 52 subjects with a Smith-Magenis-like phenotype: identification of dosage sensitive loci also associated with schizophrenia, autism, and developmental delay. *J Med Genet* 2010; **47**(4): 223-229.
170. Derwinska K, Mierzewska H, Goszczanska A, Szczepanik E, Xia Z, Kusmierska K *et al.* Clinical improvement of the aggressive neurobehavioral phenotype in a patient with a deletion of PITX3 and the absence of L-DOPA in the cerebrospinal fluid. *Am J Med Genet B Neuropsychiatr Genet* 2012; **159B**(2): 236-242.
171. Iyer J, Girirajan S. Gene discovery and functional assessment of rare copy-number variants in neurodevelopmental disorders. *Brief Funct Genomics* 2015; **14**(5): 315-328.
172. Streisinger G, Walker C, Dower N, Knauber D, Singer F. Production of clones of homozygous diploid zebra fish (*Brachydanio rerio*). *Nature* 1981; **291**(5813): 293-296.
173. Davis EE, Frangakis S, Katsanis N. Interpreting human genetic variation with in vivo zebrafish assays. *Biochimica et biophysica acta* 2014; **1842**(10): 1960-1970.
174. Blaker-Lee A, Gupta S, McCammon JM, De Rienzo G, Sive H. Zebrafish homologs of genes within 16p11.2, a genomic region associated with brain disorders, are active during brain development, and include two deletion dosage sensor genes. *Disease models & mechanisms* 2012; **5**(6): 834-851.
175. Consortium EP. An integrated encyclopedia of DNA elements in the human genome. *Nature* 2012; **489**(7414): 57-74.
176. Miele A, Dekker J. Long-range chromosomal interactions and gene regulation. *Mol Biosyst* 2008; **4**(11): 1046-1057.
177. Palstra RJ, Simonis M, Klous P, Brasslet E, Eijkelkamp B, de Laat W. Maintenance of long-range DNA interactions after inhibition of ongoing RNA polymerase II transcription. *PLoS One* 2008; **3**(2): e1661.
178. Tan-Wong SM, Zaugg JB, Camblong J, Xu Z, Zhang DW, Mischo HE *et al.* Gene loops enhance transcriptional directionality. *Science* 2012; **338**(6107): 671-675.
179. Lieberman-Aiden E, van Berkum NL, Williams L, Imakaev M, Ragoczy T, Telling A *et al.* Comprehensive mapping of long-range interactions reveals folding principles of the human genome. *Science* 2009; **326**(5950): 289-293.
180. Bickmore WA, van Steensel B. Genome architecture: domain organization of interphase chromosomes. *Cell* 2013; **152**(6): 1270-1284.

181. Gibcus JH, Dekker J. The hierarchy of the 3D genome. *Mol Cell* 2013; **49**(5): 773-782.
182. Ciribilli Y, Andreotti V, Menendez D, Langen JS, Schoenfelder G, Resnick MA *et al*. The coordinated p53 and estrogen receptor cis-regulation at an FLT1 promoter SNP is specific to genotoxic stress and estrogenic compound. *PLoS One* 2010; **5**(4): e10236.
183. Branco MR, Pombo A. Intermingling of chromosome territories in interphase suggests role in translocations and transcription-dependent associations. *PLoS Biol* 2006; **4**(5): e138.
184. Sexton T, Bantignies F, Cavalli G. Genomic interactions: chromatin loops and gene meeting points in transcriptional regulation. *Semin Cell Dev Biol* 2009; **20**(7): 849-855.
185. Carter D, Chakalova L, Osborne CS, Dai YF, Fraser P. Long-range chromatin regulatory interactions in vivo. *Nat Genet* 2002; **32**(4): 623-626.
186. Tolhuis B, Palstra RJ, Splinter E, Grosveld F, de Laat W. Looping and interaction between hypersensitive sites in the active beta-globin locus. *Mol Cell* 2002; **10**(6): 1453-1465.
187. Vorhees CV, Graham DL, Braun AA, Schaefer TL, Skelton MR, Richtand NM *et al*. Prenatal immune challenge in rats: altered responses to dopaminergic and glutamatergic agents, prepulse inhibition of acoustic startle, and reduced route-based learning as a function of maternal body weight gain after prenatal exposure to poly IC. *Synapse* 2012; **66**(8): 725-737.
188. Spilianakis CG, Lalioti MD, Town T, Lee GR, Flavell RA. Interchromosomal associations between alternatively expressed loci. *Nature* 2005; **435**(7042): 637-645.
189. Apostolou E, Thanos D. Linking differential chromatin loops to transcriptional decisions. *Mol Cell* 2008; **29**(2): 154-156.
190. Fullwood MJ, Ruan Y. CHIP-based methods for the identification of long-range chromatin interactions. *J Cell Biochem* 2009; **107**(1): 30-39.
191. Schoenfelder S, Sexton T, Chakalova L, Cope NF, Horton A, Andrews S *et al*. Preferential associations between co-regulated genes reveal a transcriptional interactome in erythroid cells. *Nat Genet* 2010; **42**(1): 53-61.
192. Lettice LA, Heaney SJ, Purdie LA, Li L, de Beer P, Oostra BA *et al*. A long-range Shh enhancer regulates expression in the developing limb and fin and is associated with preaxial polydactyly. *Hum Mol Genet* 2003; **12**(14): 1725-1735.
193. Sagai T, Hosoya M, Mizushina Y, Tamura M, Shiroishi T. Elimination of a long-range cis-regulatory module causes complete loss of limb-specific Shh expression and truncation of the mouse limb. *Development* 2005; **132**(4): 797-803.

194. Sanyal A, Lajoie BR, Jain G, Dekker J. The long-range interaction landscape of gene promoters. *Nature* 2012; **489**(7414): 109-113.
195. Amano T, Sagai T, Tanabe H, Mizushima Y, Nakazawa H, Shiroishi T. Chromosomal dynamics at the Shh locus: limb bud-specific differential regulation of competence and active transcription. *Developmental cell* 2009; **16**(1): 47-57.
196. Ghavi-Helm Y, Klein FA, Pakozdi T, Ciglar L, Noordermeer D, Huber W *et al.* Enhancer loops appear stable during development and are associated with paused polymerase. *Nature* 2014; **512**(7512): 96-100.
197. Dekker J, Rippe K, Dekker M, Kleckner N. Capturing chromosome conformation. *Science* 2002; **295**(5558): 1306-1311.
198. van Steensel B, Dekker J. Genomics tools for unraveling chromosome architecture. *Nat Biotechnol* 2010; **28**(10): 1089-1095.
199. Dekker J, Marti-Renom MA, Mirny LA. Exploring the three-dimensional organization of genomes: interpreting chromatin interaction data. *Nature reviews Genetics* 2013; **14**(6): 390-403.
200. Simonis M, Klous P, Splinter E, Moshkin Y, Willemsen R, de Wit E *et al.* Nuclear organization of active and inactive chromatin domains uncovered by chromosome conformation capture-on-chip (4C). *Nat Genet* 2006; **38**(11): 1348-1354.
201. Zhao Z, Tavoosidana G, Sjolinder M, Gondor A, Mariano P, Wang S *et al.* Circular chromosome conformation capture (4C) uncovers extensive networks of epigenetically regulated intra- and interchromosomal interactions. *Nat Genet* 2006; **38**(11): 1341-1347.
202. Rao SS, Huntley MH, Durand NC, Stamenova EK, Bochkov ID, Robinson JT *et al.* A 3D map of the human genome at kilobase resolution reveals principles of chromatin looping. *Cell* 2014; **159**(7): 1665-1680.
203. Nagano T, Lubling Y, Stevens TJ, Schoenfelder S, Yaffe E, Dean W *et al.* Single-cell Hi-C reveals cell-to-cell variability in chromosome structure. *Nature* 2013; **502**(7469): 59-64.
204. Giorgetti L, Galupa R, Nora EP, Piolot T, Lam F, Dekker J *et al.* Predictive polymer modeling reveals coupled fluctuations in chromosome conformation and transcription. *Cell* 2014; **157**(4): 950-963.
205. Dixon JR, Selvaraj S, Yue F, Kim A, Li Y, Shen Y *et al.* Topological domains in mammalian genomes identified by analysis of chromatin interactions. *Nature* 2012; **485**(7398): 376-380.
206. Nora EP, Lajoie BR, Schulz EG, Giorgetti L, Okamoto I, Servant N *et al.* Spatial partitioning of the regulatory landscape of the X-inactivation centre. *Nature* 2012; **485**(7398): 381-385.



207. Sexton T, Yaffe E, Kenigsberg E, Bantignies F, Leblanc B, Hoichman M *et al.* Three-dimensional folding and functional organization principles of the Drosophila genome. *Cell* 2012; **148**(3): 458-472.
208. Pope BD, Ryba T, Dileep V, Yue F, Wu W, Denas O *et al.* Topologically associating domains are stable units of replication-timing regulation. *Nature* 2014; **515**(7527): 402-405.
209. Lupianez DG, Spielmann M, Mundlos S. Breaking TADs: How Alterations of Chromatin Domains Result in Disease. *Trends Genet* 2016.
210. Lupianez DG, Kraft K, Heinrich V, Krawitz P, Brancati F, Klopocki E *et al.* Disruptions of topological chromatin domains cause pathogenic rewiring of gene-enhancer interactions. *Cell* 2015; **161**(5): 1012-1025.

## Chapter 1: Chromosomal contacts connect loci associated with autism, BMI and head circumference phenotypes

### Summary of the contribution

The observations described in this chapter refer to the article “Chromosomal contacts connect loci associated with autism, BMI and head circumference phenotypes”, currently under review (second round of revision) in the journal “Molecular Psychiatry”.

As co-first author of this paper, I contributed to a substantial subset of the analysis and experiments presented in the following sections. Specifically, I designed and prepared the 4C-seq libraries, whose sequencing results were then analyzed by Marion Leleu (co-first author), conducted the statistical analysis on the 4C-seq dataset, GO, network analysis, and comparison with 16p11.2 transcriptome dataset. I also participated in the analysis of the anthropometric data collected for the 16p11.2 BP2-BP3 and 2p15 patients’ cohorts.

The results are presented in five main figures (named **Figure 1-5**), eleven supplementary figures (**Figure S1-11**) and thirty-seven supplementary tables (**Table S1-37**; **Tables S1-2, S6-27, S31, S35-37** are available upon request; the remaining tables are included in the text).

## **Chromosomal contacts connect loci associated with autism, BMI and head circumference phenotypes**

Maria Nicla Loviglio<sup>1,27</sup>, Marion Leleu<sup>2,3,27</sup>, Katrin Männik<sup>1,4</sup>, Marzia Passeggeri<sup>5</sup>, Giuliana Giannuzzi<sup>1</sup>, Ilse van der Werf<sup>1</sup>, Sebastian M. Waszak<sup>2,3,6</sup>, Marianna Zazhytska<sup>1</sup>, Inês Roberts-Caldeira<sup>5</sup>, Nele Gheldof<sup>1</sup>, Eugenia Migliavacca<sup>1,2</sup>, Ali Abdullah Alfaiz<sup>1,2</sup>, Loyse Hippolyte<sup>5</sup>, Anne M. Maillard<sup>5</sup>, 16p11.2 Consortium<sup>7</sup>, 2p15 Consortium<sup>8</sup>, Anke Van Dijck<sup>9</sup>, R. Frank Kooy<sup>9</sup>, Damien Sanlaville<sup>10</sup>, Jill A. Rosenfeld<sup>11,12</sup>, Lisa G. Shaffer<sup>13</sup>, Joris Andrieux<sup>14</sup>, Christian Marshall<sup>15</sup>, Stephen W. Scherer<sup>16,17</sup>, Yiping Shen<sup>18,19,20</sup>, James F. Gusella<sup>21,22</sup>, Unnur Thorsteinsdottir<sup>23</sup>, Gudmar Thorleifsson<sup>23</sup>, Emmanouil T. Dermitzakis<sup>2,24</sup>, Bart Deplancke<sup>2,3</sup>, Jacques S Beckmann<sup>2,5,25</sup>, Jacques Rougemont<sup>2,3</sup>, Sébastien Jacquemont<sup>5,26,28</sup>, Alexandre Reymond<sup>1,28,\*</sup>

<sup>1</sup>Center for Integrative Genomics, University of Lausanne, CH-1015 Lausanne, Switzerland

<sup>2</sup>Swiss Institute of Bioinformatics (SIB), CH-1015 Lausanne, Switzerland

<sup>3</sup>School of Life Sciences, EPFL (Ecole Polytechnique Fédérale de Lausanne), CH-1015 Lausanne, Switzerland

<sup>4</sup>Estonian Genome Center, University of Tartu, 51010 Tartu, Estonia

<sup>5</sup>Service of Medical Genetics, Lausanne University Hospital (CHUV), CH-1011 Lausanne, Switzerland

<sup>6</sup>Genome Biology Unit, European Molecular Biology Laboratory (EMBL), 69117 Heidelberg, Germany

<sup>7</sup>The list of the 16p11.2 Consortium members is specified at the end of the text, while their affiliations are detailed in Supplemental Data.

<sup>8</sup> The list of the 2p15 Consortium members is specified at the end of the text, while their affiliations are detailed in Supplemental Data.

<sup>9</sup>Department of Medical Genetics, University of Antwerp, 2610 Antwerp, Belgium

<sup>10</sup>Service de génétique, Hôpital Femme Mère Enfant, Centre de Biologie et Pathologie Est, Hospices Civils de Lyon, CRNL, CNRS UMR5292, INSERM U1028, Université Claude Bernard Lyon I, 69677 Bron, France.

<sup>11</sup>Signature Genomic Laboratories, PerkinElmer, Inc., Spokane, WA 99207

<sup>12</sup>Department of Molecular & Human Genetics, Baylor College of Medicine, Houston, TX 77030

<sup>13</sup>Paw Print Genetics, Genetic Veterinary Sciences, Inc., Spokane, WA 99202

<sup>14</sup>Institut de Génétique Médicale, CHRU de Lille - Hôpital Jeanne de Flandre, 59037 Lille, France

<sup>15</sup>Genome Diagnostics, Department of Paediatric Laboratory Medicine, The Hospital for Sick Children, Toronto, ON M5G 1X8

<sup>16</sup>The Centre for Applied Genomics, The Hospital for Sick Children, Toronto, ON M5G 1X8

<sup>17</sup>McLaughlin Centre and Department of Molecular Genetics, University of Toronto, ON M5G 1X8

<sup>18</sup>Department of Pathology, Harvard Medical School, Boston MA 02114

<sup>19</sup>Department of Laboratory Medicine, Boston Children's Hospital, Boston MA 02115

<sup>20</sup>Shanghai Jiaotong University, School of Medicine, Shanghai, 200127 China

<sup>21</sup>Center for Human Genetic Research, Massachusetts General Hospital, Boston, MA 02114

<sup>22</sup>Departments of Genetics and Neurology, Harvard Medical School, Boston, MA 02114-2696

<sup>23</sup>deCODE Genetics, Amgen, IS-101 Reykjavik, Iceland

<sup>24</sup>Department of Genetic Medicine and Development, University of Geneva Medical School, CH-1211 Geneva, Switzerland

<sup>25</sup>Department of Medical Genetics, University of Lausanne, CH-1011 Lausanne, Switzerland

<sup>26</sup>Current address: CHU Sainte Justine, University of Montreal, Montreal, QC QC H3T 1C4

<sup>27</sup>Co-first authors

<sup>28</sup>Co-senior authors

\*corresponding author

**Correspondence should be addressed to:**

Alexandre Reymond, alexandre.reymond@unil.ch

Center for Integrative Genomics, University of Lausanne, Genopode building, CH-1015 Lausanne, Switzerland

+41 21 692 3960 (ph); +41 21 692 3965 (fx).

**Running Title:** Chromosomal contacts of 16p11.2 CNVs

**Keywords:** 16p11.2; autism; obesity; head size; chromosome conformation capture; transcriptome; 2p15; genome architecture

## Abstract

Copy number variants (CNVs) are major contributors to genomic imbalances disorders. Phenotyping of 137 unrelated deletion and reciprocal duplication carriers of the distal 16p11.2 220kb BP2-BP3 interval showed that these rearrangements are associated with autism spectrum disorders (ASD) and mirror phenotypes of obesity/underweight and macro-/microcephaly. Such phenotypes were previously associated with rearrangements of the non-overlapping proximal 16p11.2 600kb BP4-BP5 interval. These two CNVs-prone regions at 16p11.2 are reciprocally engaged in complex chromatin looping, as successfully confirmed by 4C-seq, FISH and Hi-C, as well as coordinated expression and regulation of encompassed genes. We observed that genes differentially expressed in 16p11.2 BP4-BP5 CNV carriers are concomitantly modified in their chromatin interactions, suggesting that disruption of chromatin interplays could participate in the observed phenotypes.

We also identified cis- and trans-acting chromatin contacts to other genomic regions previously associated with analogous phenotypes. For example, we uncovered that individuals with reciprocal rearrangements of the trans-contacted 2p15 locus similarly display mirror phenotypes on head circumference and weight. Our results indicate that chromosomal contacts' maps could uncover functionally and clinically related genes.

## Introduction

Long-range chromatin contacts that bring genes and regulatory sequences in close proximity are necessary for co-transcription of biologically-related and developmentally co-regulated genes<sup>1, 2</sup>. Correspondingly, genomic structural changes were associated with disruption of the organization of chromatin compartments by shifting regulatory elements between domains and/or modifying domain boundaries, which resulted in ectopic interactions, gene misexpression and disease<sup>3, 4</sup>. In the last 15 million years the 16p11.2–12.2 region rapidly integrated segmental duplications contributing to profound modifications of these chromosomal bands in hominoids<sup>5, 6</sup>. It allowed the emergence of new transcripts<sup>7</sup> and placed the whole region at risk for various recurrent rearrangements<sup>8-10</sup> through non-allelic homologous recombination<sup>11</sup>(**Figure 1**). These rearrangements include a recurrent interstitial deletion of ~600kb defined by 16p11.2 breakpoints 4-5 (BP4-BP5; OMIM#611913), which encompasses 28 “unique” genes and four genes with multiple copies<sup>12</sup>(**Figure 1**). With a population prevalence of ~0.05% this variant is one of the most frequent known etiologies of autism spectrum disorder (ASD)<sup>9, 13-15</sup>. It impacts adaptive behavior and language skills and predisposes to a highly penetrant form of obesity and macrocephaly<sup>15-19</sup>. A mirror phenotype is observed in carriers of the reciprocal duplication (OMIM#614671), who present a high risk of schizophrenia (SCZ), Rolandic epilepsy, being underweight and microcephalic<sup>18-22</sup>. Case series have reported variable expressivity; systematic phenotyping showed that deletion and duplication lead to an average IQ decrease of 26 and 16 points in proband compared to non-carrier family members<sup>15, 23</sup>. Correspondingly, the phenotypes of carriers identified in unselected populations are reminiscent of those described for carriers of 16p11.2 rearrangements ascertained in clinical cohorts. <sup>24</sup>. Deletions and duplications show a mirroring impact on brain volume and specific cortico-striatal structures implicated in reward, language and social cognition<sup>25</sup>. Changes in copy numbers of this interval are associated with significant modifications of the mRNA levels of ciliopathy and ASD-associated genes in humans and mice<sup>12, 26</sup>.

Correspondingly, mouse models engineered to have three copies of the 7qF3 orthologous region showed reduced cilia length in the CA1 hippocampal region, whereas modulation of the expression of ciliopathy-associated genes rescued phenotypes induced by *KCTD13* (MIM#608947) under- and overexpression<sup>12</sup>, one of the key drivers of the 16p11.2 600kb BP4-BP5 CNV genomic-interval associated traits<sup>27</sup>. Distal to BP4-BP5, the deletion of 16p11.2 220kb BP2-BP3 interval was similarly associated with obesity, developmental delay (DD), intellectual disability (ID) and SCZ<sup>16, 28-35</sup>. However, detailed data about the phenotypes associated with the reciprocal duplication are still lacking.

We hypothesized that copy number modification of the 16p11.2 600kb BP4-BP5 interval alters the three dimensional positioning of these genes resulting in expression alterations of pathways involved in its phenotypic manifestation. We used chromatin conformation capture to explore the chromosome-wide effects of the 16p11.2 600kb BP4-BP5 structural rearrangements on chromatin structure and assessed how these underlay the associated phenotypes. This region engages in multi-gene complex structures that are disrupted when its copy number changes. The implicated genes are known to be linked to 16p11.2-associated phenotypes, such as primary cilium alteration, energy imbalance, head circumference (HC) and ASD. We also demonstrate that our approach could be used to identify additional loci, whose copy number changes are associated with strikingly similar phenotypic manifestations.

## **Materials and Methods**

### **Recruitment and phenotyping of patients**

The institutional review board of the University of Lausanne, Switzerland approved this study. Participants were enrolled in the study after signing an informed consent form and being clinically assessed by their respective physicians. For the data collected through questionnaires, information was gathered retrospectively and anonymously by physicians who had ordered chromosomal microarray analyses performed for clinical purposes only.

Consequently, research-based informed consent was not required by the institutional review board of the University of Lausanne, which granted an exemption for this part of the data collection. Overall cognitive functioning was assessed as published<sup>15</sup>.

To better assess the phenotypic features associated with the 16p11.2 220kb BP2-BP3 rearrangements, we recruited and phenotyped 110 and 57 carriers of the 220kb BP2-BP3 deletion (OMIM#613444) and duplication from 88 and 49 families, respectively (**Table S1**). Whereas these structural variants were previously reported to be among the CNVs most frequently harboring a possibly deleterious second genetic lesion (29% and 13% of the time, respectively)<sup>32</sup>, we do not confirm such propensity. Indeed, second-site structural variants were identified in “only” 7% (6/88) and 4% (2/49) of the enrolled BP2-BP3 deletion and duplication carrier probands, respectively. Deleterious CNVs were defined as: i) known recurrent genomic disorder, ii) CNV encompassing published critical genomic region or disrupting a gene that is a known etiology of neurodevelopmental disorders or iii) >500kb CNV with AF<0.001. We compared available data on weight, height, BMI (Body Mass Index) and HC for 77 and 39 unrelated deletion and duplication carriers, respectively (including published cases). The mean age of this group of patients was 16 years (range 0.42-78 years, with 34 cases older than 18 years). The prevalence of the 16p11.2 deletion and duplication were inferred from six European population-based GWAS cohorts, sets of chromosomal microarray-genotyped control individuals and clinical cohorts<sup>9, 30, 33, 35-39</sup>. CNV analyses were carried out as described in<sup>18</sup>.

We similarly enrolled 26 and 9 unrelated carriers of 2p15 deletions and duplications, including 12 deletion cases from the literature<sup>40-49</sup>. The Signature Genomics cases were recently described in<sup>50</sup>. Patients were identified through routine etiological work-ups of patients ascertained for developmental delay/intellectual disability in cytogenetic centers. The coordinates of the



rearrangements' breakpoints (**Table S2**) were recognized by different chromosomal microarray platforms.

### Lymphoblastoid cell lines and transcriptome profiling

We had previously established by EBV transformation LCLs from 16p11.2 BP4-BP5 patients, as well as controls. The LCL transcriptome of 50 deletion and 31 duplication carriers, as well as 17 control individuals was previously profiled with Affymetrix GeneChips Human Genome U133+ PM 24 array plates. The results are deposited in the NCBI Gene Expression Omnibus under accession number GSE57802. The Robust Multi-array Average (RMA) approach was used for the creation and normalization of the summarized probe set signals. We applied a non-specific filter to discard probe sets with low variability and low signal, i.e. detectable expression levels. Specifically, probe sets with both i) signal SD > median of signal SD of all probe sets, and ii) larger signal > median of larger signal of all probe sets, were retained as described in <sup>12</sup>. This selection yielded a total of 23,602 probe sets. To reduce a potential bias toward genes with multiple probe sets, for the modular analysis, only one probe set with the highest variance per gene was kept, for a total of 15,112 probe sets. Using a dosage effect model and moderated t-statistics, we identified 1,188 and 2,209 significantly differentially expressed genes (FDR ≤ 1% and 5%, respectively; uniquely mapping probes)<sup>12</sup>. We used Geneprof to access data pertaining to gene expression and co-regulation.

We are well aware of the limitations of the study of LCLs, for instance for genes whose expression specificity resides in other cell lineages. These experiments are nevertheless worth pursuing simply because i) the primary human target tissues remain often beyond reach; ii) we cannot exclude a broad to ubiquitous expression pattern and chromatin contacts for the genes involved in these disease processes; and iii) the pattern of expressions in peripheral tissue may be used as a biomarker in translational project. Similar limitations apply to the use

of embryonic stem cells-derived material, while animal tissues have a different set of shortcomings.

### Quantitative RT-PCR

For qPCR, 100 ng of high-quality total RNA was converted to cDNA using Superscript VILO (Invitrogen, Carlsbad, CA, USA) according to the manufacturer's protocol. Primers were designed using PrimerExpress 2.0 software (Applied Biosystems, Foster City, CA, USA), with default parameters except for the primer- and minimal amplicon lengths, which were set at 17-26 bp and 60 bp respectively. The amplification factor of each primer pair was tested using a cDNA dilution series and only assays with amplification factors between 1.75 and 2.00 were retained. A representative set of samples was tested for genomic contamination. qPCR experiments were performed in triplicate using SYBR-Green (Roche, Basel, Switzerland) as reporter. The reaction mixtures were prepared in 384-well plates using a Freedom Evo robot (Tecan, Männedorf, Switzerland) and run in an ABI 7900HT sequence detection system (Applied Biosystems, Foster City, CA, USA) using the following conditions: 50°C for 2 minutes, 95°C for 10 minutes, followed by 45 cycles of 95°C for 15 seconds and then 60°C for 1 minute, after which dissociation curves were established. Applicable normalization genes were included in each experiment to enable compensation for fluctuations in expression levels between experiments. Using SDS v2.4 software (Applied Biosystems, Foster City, CA, USA) the threshold and baseline values were adjusted when necessary to obtain raw cycle threshold (Ct) values. The Ct values were further analysed using qBase plus software (Biogazelle, Zwijnaarde, Belgium), which calculates relative expression values per sample per tested gene upon designation of the normalization genes and corrects for the amplification efficiency of the performed assay. We assessed by qPCR the RNA levels of seven DE genes belonging to the ciliopathy or PTEN pathway (*BBS4* (MIM#600374), *BBS7* (MIM#607590), *BBS10* (MIM#610148), *XPOT* (MIM#603180), *NUP58* (MIM#607615), *PTPN11* (MIM#176876) and *SMAD2* (MIM#601366)), and five others that map either to the BP4-BP5 (*ALDOA* (MIM#103850), *KCTD13*, *MAPK3* (MIM#601795) and *MVP* (MIM#605088)) or

the BP2-BP3 interval (*SH2B1* (MIM#608937)) in LCLs from eight carriers of the 220kb BP2-BP3 deletion, eight carriers of the 600kb BP4-BP5 deletion and 10 control individuals. In particular we identified a significant diminution of the hemizygote gene *SH2B1* but not of the neighboring normal-copy *KCTD13*, *MVP* and *MAPK3* in BP2-BP3 deletion carriers.

### Viewpoint selection

We used an adaptation of the 4C method<sup>51-53</sup>, the high-resolution Chromosome Conformation Capture Sequencing technology (4C-seq)<sup>54</sup> to identify chromosomal regions that physically associate with the promoters of *MVP*, *KCTD13*, *ALDOA*, *TBX6* (MIM#602427) and *MAPK3*, five of the 28 “unique” genes of the BP4-BP5 interval selected according to their potential role in the described phenotype. Reduction by ~50% of the RNA levels of the ortholog of *ALDOA* (Aldolase A) was associated with a change in brain morphology in zebrafish, suggesting that this gene is dosage sensitive<sup>55</sup>. In human, recessive *ALDOA* deficiency is associated with glycogen storage disease XII (OMIM#611881)<sup>56</sup>. Morpholino-driven reduction of the expression level of the *KCTD13* (Potassium Channel Tetramerisation Domain containing protein 13) ortholog resulted in macrocephaly in zebrafish, while its depletion in the brain of mouse embryos resulted in an increase of proliferating cells. The mirroring microcephaly was seen upon overexpression of human *KCTD13* cDNA in zebrafish embryos’ heads, a phenotype further amplified upon concomitant overexpression of either *MAPK3* (Mitogen-Activated Protein Kinase 3) or *MVP* (Major Vault Protein)<sup>27</sup>. *TBX6* (T-Box Transcription Factor 6) is a candidate gene for the vertebral malformations observed in some deletion carriers since i) mice homozygous for a *Tbx6* mutation showed rib and vertebral body anomalies<sup>57</sup>; ii) *TBX6* polymorphisms were associated with congenital scoliosis in the Han population<sup>58</sup>; iii) a stoploss variant in *TBX6* segregates with congenital spinal defects in a three-generation family<sup>59</sup> (OMIM#122600); and iv) carriers of 16p11.2 600kb BP4-BP5 deletions and a common hypomorphic *TBX6* allele suggest a compound inheritance in congenital scoliosis<sup>60</sup>. *TBX6* was selected as a viewpoint even though this gene is not expressed in LCLs (or only at extremely

low level), as studies have shown that the contacted domains are stable across cell lines and tissues regardless of expression status<sup>3</sup>. Within the group of genes chromatin-contacted by the above viewpoints we selected two more viewpoints within the 16p11.2 220kb BP2-BP3 region (**Figure 1**), *i.e.* the promoters of *SH2B1* and *LAT*. The *SH2B1* gene was suggested to be a crucial candidate for the obesity phenotype associated with this genomic interval<sup>16,28</sup> as it encodes a Src homology adaptor protein involved in leptin and insulin signaling<sup>61,62</sup>. Common variants in this locus were repeatedly associated with BMI, serum leptin and body fat in genome-wide association studies (GWAS)<sup>63-66</sup>, while rare dominant mutations were reported to cause obesity, social isolation, aggressive behavior and speech and language delay<sup>67</sup>. In a recent large-scale association study, the deletion was also significantly linked with SCZ<sup>35</sup>. The *LAT* (linker for activation of T cells) adaptor molecule participates in AKT activation and plays an important role in the regulation of lymphocyte maturation, activation and differentiation<sup>68,69</sup>. Its inactivation could be circumvented by Ras/MAPK constitutive activation<sup>70</sup>.

#### 4C-seq

4C libraries were prepared from LCLs of two control individuals and two carriers each of the 16p11.2 BP4-BP5 deletion and duplication, sex- and age-matched (**Table S3**). Briefly, LCLs were grown at 37°C.  $5 \times 10^7$  exponentially growing cells were harvested and crosslinked with 1% formaldehyde, lysed and cut with DpnII, a 4-cutter restriction enzyme that allows higher resolution<sup>53</sup>. After ligation and reversal of the crosslinks, the DNA was purified to obtain the 3C library. This 3C library was further digested with NlaIII and circularized to obtain a 4C library. The inverse PCR primers to amplify 4C-seq templates were designed to contain Illumina adaptor tails, sample barcodes and viewpoint-specific sequences. Viewpoints were selected at the closest suitable DpnII fragment relative to the transcriptional start sites of the targeted genes. The sequence of the 4C-seq primers is reported in **Table S4**. For all viewpoints, we amplified at least 1.6 µg of 4C template (using about 100 ng of 4C template per inverse PCR reaction, for a total number of 16 PCRs). We multiplexed the 4C-seq templates in equimolar ratios and analyzed them on a 100-bp single-end

Illumina HiSeq flow cell. The numbers of raw, excluded, and mapped reads for each viewpoint and LCL sample are detailed in **Table S5**.

#### 4C-seq data analysis

4C-seq data were analyzed as described in<sup>53, 54</sup> through the 4C-seq pipeline available at <http://htsstation.epfl.ch/>)<sup>71</sup> and visualized with gFeatBrowser. Briefly, the multiplexed samples were separated, undigested and self-ligated reads removed. Remaining reads were aligned and translated to a virtual library of DpnII fragments. Read counts were then normalized to the total number of reads and replicates combined by averaging the resulting signal densities (**Figure S1**). The local correlation between the profiles of the two samples per viewpoint was calculated ( $0.46 \leq r^2 \leq 0.74$  for controls,  $0.29 \leq r^2 \leq 0.67$  for deletions and  $0.22 \leq r^2 \leq 0.61$  for the duplications). The combined profiles were then smoothed with a window size of 29 fragments. The region directly surrounding the viewpoint is usually highly enriched and can show considerable experimental variation, thereby influencing overall fragment count. To minimize these effects, the viewpoint itself and the directly neighboring 'undigested' fragment were excluded during the procedure. In addition to this filtering, we modeled the data to apply a profile correction similar to the one described in<sup>72</sup> using a fit with a slope -1 in a log-log scale<sup>73</sup>. Significantly interacting regions were detected by applying a domainogram analysis as described<sup>74</sup>. We selected BRICKS (Blocks of Regulators In Chromosomal Kontext) with a p-value threshold smaller than 0.01 for both "cis" and "trans" interactions. To determine differentially interacting regions between the 16p11.2 600kb BP4-BP5 deletion (Del), duplication (Dup) and control (Ctrl), we considered all non-null BRICKS found by a domainogram analysis<sup>74</sup> in either condition and quantified both signals in each BRICK. The resulting table was scaled to the sample with the largest IQR and the difference of signals was compared to random in order to associate a p-value (FDR) to each BRICK. Finally, only BRICKS with a p-value < 0.01 were considered.

All the viewpoints mapping on the BP4-BP5 interval, except *KCTD13*, contact the 146 and 147kb long Low Copy Repeats (LCR) that flank the 16p11 600kb BP4-BP5 rearrangements. To unravel whether the signal was reflecting the interaction with the centromeric, the telomeric or both LCRs and given the high similarity (99.5% identity) of the two blocks, we separately treated the reads mapping within these regions (chr16:29460515-29606852 and chr16:30199854-30346868 according to GRCh 37/*hg19* assembly, February 2009) using different and more stringent criteria, *i.e.* no mismatch and unique site mapping. All values were normalized to the total number of reads mapping to the 2 regions (per thousands of reads). We observed a higher proportion of contacts occurring with the centromeric segmental duplication compared to the telomeric one for *MAPK3* and *TBX6*, while the trend was reversed for *MVP*, in agreement with their proximity to the centromeric and telomeric LCR blocks, respectively. No conclusive results were obtained for *ALDOA*.

### Hi-C data

Hi-C matrices from <sup>75</sup> were prepared by first applying a KR normalization to the 5kb and 100kb resolution observed matrices, and then by dividing each normalized score by the expected one extracted from the KR expected file (as described in section II.c of the Extended Experimental Procedures of reference <sup>75</sup>). KR Expected values less than 1 were set to 1 to avoid long-distance interaction biases.

### Enrichment analyses

Gene annotation was obtained through BioScript (<http://gdv.epfl.ch/bs>). Protein interaction networks for the genes selected by BRICKS calling and from the list of interacting regions affected by the rearrangements were determined using STRING (Search Tool for the Retrieval of Interacting Genes/Proteins) v9.1 (<http://string-db.org/>)<sup>76</sup>. We exploited Bioscript for Gene Ontology analysis (topGO) (<http://gdv.epfl.ch/bs>), DAVID GO and KEGG, OMIM Disease and KEGG coupled with Enrichr to assess if the chromatin-contacted genes were enriched

in specific pathways and genes associated with Mendelian diseases<sup>77-80</sup> (<http://www.omim.org/downloads>). The OMIM gene-set library was obtained directly from the NCBI's OMIM Morbid Map<sup>81</sup>. We exploited the SFARI Gene lists and scores (<https://sfari.org/>; March 2014 release), the union of the genes cataloged in<sup>82-84</sup> and the GWAS hits for BMI<sup>85</sup> to assess enrichment for ASD, SCZ and BMI genes, respectively. We also used the *de novo* "high confidence" ASD targets (selected with FDR<0.1 in <sup>86</sup> and likely gene disrupting recurrent mutations target genes in <sup>87</sup>) to assess enrichment of ASD-associated genes. Ciliary genes enrichment was computed merging the SYSCILIA gold standard (SCGS) and potential ciliary gene lists (genes with no additional evidence for ciliary function were excluded)<sup>88</sup>. We used Enrichr Chromosome Location tool and BRICKS count in different window sizes (5Mb, 1Mb and 500kb) to determine whether any cytogenetic band other than 16p11.2 was enriched for BRICKS. Other than 16p11.2, we identified enrichments at 16p12, 16p13, 16q13, 1p36, 11q13, 16q22, 7q31, 15q15 and 1q32 cytobands. As a large proportion of the 16p11.2 BRICKS map to segmental duplications that are highly similar to the LCR flanking the 16p11 600kb BP4-BP5 rearrangements we conservatively did not consider this region.

All seven tested viewpoints showed enrichment for contacts with loci that encode proteins that interact together (all  $p < 0.01$ ). A single process, focal adhesion assembly, was shared between the BP2-BP3 and the BP4-BP5 groups of viewpoints (GO:0048041,  $p = 6.02e^{-03}$  for the BP2-BP3 group and  $P = 1.12e^{-03}$  for the BP4-BP5). Focal adhesion links the internal actin cytoskeleton to the extracellular matrix; it is used by cells to explore their environment, and depends strongly on microtubule dynamics<sup>89</sup> in coordination with the primary cilium<sup>90, 91</sup>. Genes participating in focal adhesion (*GSK3B* (MIM#605004); *PAK7* (MIM#608038)), axon guidance (*ROBO1* (MIM#602430); *EPHA6* (MIM#600066); *PAK7*; *GSK3B*) and Golgi apparatus-related processes (*SND1* (MIM#602181); *FRMD5* (MIM#616309)) are among the 24 genes trans-contacted by both the BP4-BP5 and BP2-BP3 viewpoints. The gene *MPZL1* (MIM#604376), contacted

by both *KCTD13* and *SH2B1*, is associated with SCZ<sup>92</sup>. It is a downstream target of *PTPN11*<sup>93</sup>.

### Fluorescence in situ hybridization (FISH)

Interphase nuclei were prepared from a LCL of a control individual. FISH experiments were performed using fosmid clones (300 ng) directly labeled by nick-translation with Cy3-dUTP and fluorescein-dUTP as previously described<sup>94</sup> with minor modifications. Fosmid and BAC clones (G248P86150B3 for the *ALDOA* locus, G248P800063B6 for the *SH2B1* locus; G248P86115A10 for the *KIAA0556* locus; RP11-301D18 for the *KCTD13* and *MVP* loci; RP11-383D9 for *PTEN* (MIM#601728); RP11-477N2 for *USP34* (MIM#615295)/*XPO1* (MIM#602559) and RP11-43E18 for *MARK4* (MIM#606495)) were obtained from the CHORI BACPAC Resources Center (<https://bacpac.chori.org/>). We picked the *MARK4*-encompassing BAC as “control BAC” because it maps to a gene-rich region on chromosome 19, a centrally-positioned chromosome within the nucleus. Hybridization was performed at 37°C in 2xSSC, 50% (v/v) formamide, 10% (w/v) dextran sulfate, 3 µg C0t-1 DNA, and 3 µg sonicated salmon sperm DNA, in a volume of 10 µl. Post-hybridization washing was at 60°C in 0.1xSSC, three times. Nuclei were DAPI-stained and digital images were obtained using a Zeiss Imager A1 fluorescence microscope. We considered 50-60 cells per experiment (i.e. at least 100 distances) and co-localization was defined if the distance between signals was  $\leq 0.3\mu\text{m}$ . The contact between *SH2B1* and *ALDOA*, versus the control *KIAA0556*, was estimated by calculating the distance between BAC probes (median *SH2B1-ALDOA* and *-KIAA0556* distances = 0.43 and 1.24 µm, respectively; Wilcoxon rank sum test,  $P=1.45e^{-17}$ ). The contact of *MVP* and *KCTD13* with, respectively, *PTEN* and *USP34/XPO1*, compared to the control *MARK4*, was estimated as percentage of co-localization (25% and 14% co-localization versus 2% with the control locus; Fisher’s test enrichment:  $P=6.9e^{-05}$  and  $P=0.01$ , respectively; median *MVP/KCTD13-USP34/XPO1* distances = 1.76, *MVP/KCTD13-PTEN* = 2.61 and *MVP/KCTD13-MARK4* = 4.96 µm; Wilcoxon rank sum test,  $P=5.4e^{-10}$  and  $P=9.3e^{-05}$ , respectively).



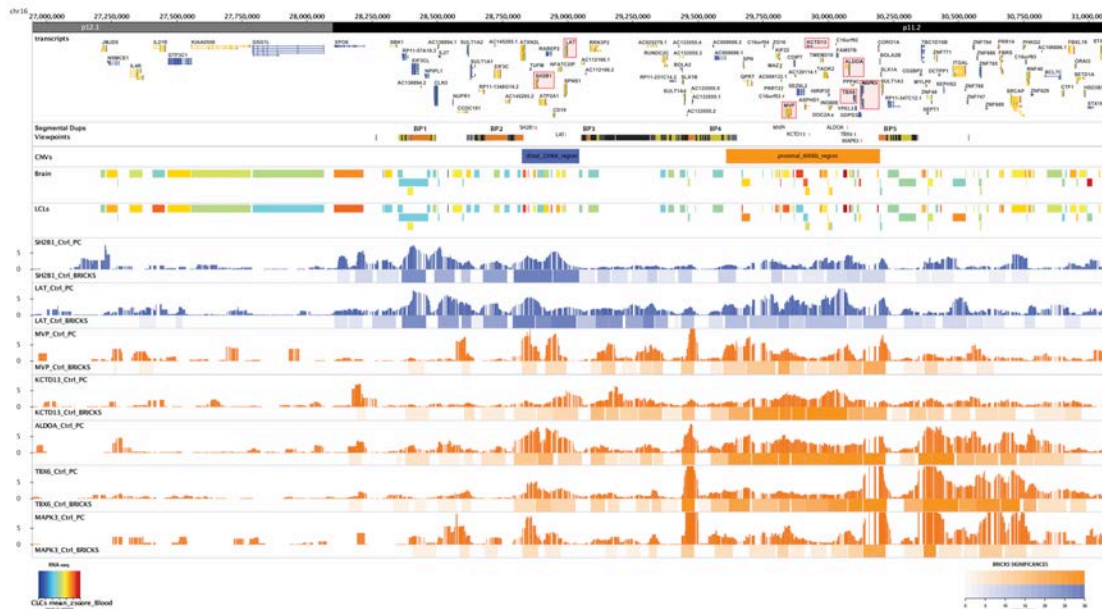
### ChIP-seq and RNA-seq molecular associations

Detailed experimental procedures and results are described in <sup>95</sup>. Briefly, ChIP-seq (chromatin immuno-precipitation coupled with sequencing) and mRNA-seq data were produced from LCLs of 54 individuals of European origin from the 1000 Genomes Project <sup>96</sup>. ChIP-seq with antibodies recognizing H3K4me1, H3K4me3, H3K27ac, PU.1 and RNA Pol2 binding, as well as mRNA-seq gene expression profiling, were carried out from a single growth of LCLs as previously described <sup>97</sup>. Genotypes were obtained from the GEUVADIS consortium <sup>98</sup>. To map associations between pairs of ChIP-seq and/or RNA-seq peaks, we retained 47 individuals after data quality control and proceeded as follows for each of the 15 possible unordered pairs of distinct molecular phenotypes ( $A_1, A_2$ ). First, we measured inter-individual Pearson correlation between every possible pair of normalized quantifications at peaks ( $p_1, p_2$ ) within the 16p11.2 interval (28.1-34.6 Mb) such that  $p_1$  and  $p_2$  belong to  $A_1$  and  $A_2$ , respectively. Note that the distances here were measured between the respective peak centers, excepted for mRNA for which we used the transcription start site. Then, we assessed to what extent the correlations significantly differed from zero by calculating P-values using the R function *cor.test* and corrected them for multiple-testing by using the Benjamini & Hochberg procedure as implemented in the R function *p.adjust* (FDR 5% and 10%).

## Results

### Distinct and non-overlapping loci at 16p11.2 are associated with mirror phenotypes on BMI and HC and autism susceptibility

To comprehensively assess phenotypic features associated with the distal 16p11.2 220kb BP2-BP3 CNVs (**Figure 1**), we collected de-identified data on 137 unrelated carriers (88 deletions and 49 duplications; **Table S1**) and compared BMI and HC to gender-, age- and geographical location-matched reference



**Figure 1: The 16p11.2 region and its 4C interactions profile (Panels from top to bottom)**

**Transcripts:** The transcripts mapping within the human chromosome 16 GRCh37/hg19 27-31Mb region are indicated. The 4C-targeted *SH2B1*, *LAT*, *MVP*, *KCTD13*, *ALDOA*, *TBX6* and *MAPK3* genes are highlighted in red.

**Segmental duplications/viewpoints:** The duplicated regions containing the low-copy repeats (LCR) that flank these rearrangements telomerically and centromerically are shown, whereas the position of the restrictions fragments used as viewpoints are marked with red ticks.

**CNVs:** The position of the 600kb BP4-BP5 (orange) and 220kb BP2-BP3 intervals (blue) are depicted.

**Brain/LCLs:** The mean z-score for transcript expression per group (Brain or LCLs) from GTEx are displayed. The corresponding RNA-seq heatmap color legend is showed at the bottom left corner.

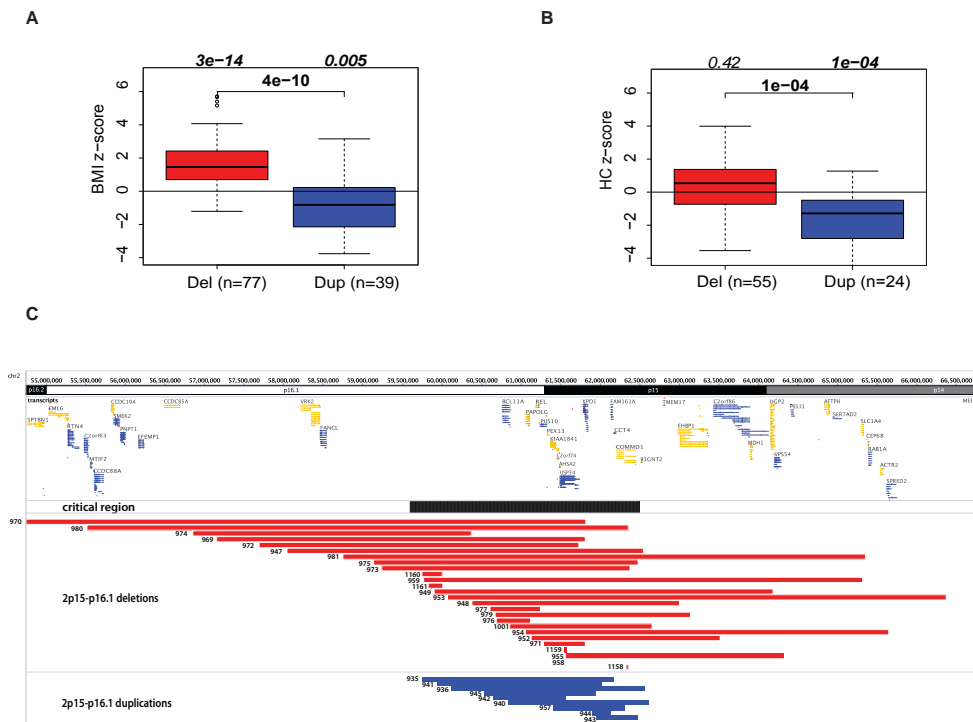
**PC/BRICKs:** Smoothed and profile-corrected 4C signal (upper part of each panel) and BRICKs (lower part) identified for each of the seven 4C viewpoints within the 16p11.2 cytoband, i.e. from top to bottom *SH2B1*, *LAT*, *MVP*, *KCTD13*, *ALDOA*, *TBX6* and *MAPK3*. The corresponding BRICKs significance heatmap color legend is showed at the bottom right corner.

population as described<sup>18</sup>(**Figure 2A-B**). The BMI mean Z-score of deletion carriers deviated significantly from that of the general population (t-test,  $P=3.1e^{-14}$ ), replicating the earlier described association of the deletion with obesity<sup>16, 28</sup>. We observe a trend towards increased HC in deletion carriers. The duplication carriers showed a mirroring decrease of BMI and HC values when compared to those of the control population (t-test,  $P=0.005$  and  $1.1e^{-4}$ , respectively). We also observe an increase in ASD prevalence in both deletion (23/88; 26%) and duplication (11/49; 22%) carriers compared to the general population (5,338/363,749; 1.5%)<sup>99</sup> (Fisher's enrichment test:  $OR=23.7$ ,  $P=2.5e^{-22}$ ;  $OR=19.4$ ,  $P=1.2e^{-10}$ ) in agreement with published results<sup>29-35</sup>. Thus, genomic rearrangements at 600kb BP4-BP5 and 220kb BP2-BP3, two loci 650kb apart, present similar clinical patterns: large effect sizes on BMI and HC, as well as association with ASD.

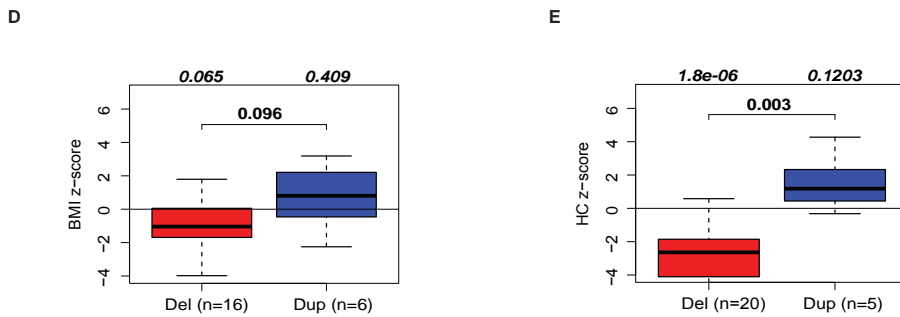
#### Cis-acting chromatin loops that link the 16p11.2 BP4-BP5 and BP2-BP3 genomic intervals are perturbed in BP4-BP5 CNV carriers

We posited that the remarkable overlap of phenotypic features associated with the BP2-BP3 and BP4-BP5 CNVs might derive from the rearrangement-mediated disruption of the 3D chromatin structure within the 16p11.2 cytoband. To challenge this hypothesis, we assessed the pattern of chromosomal interactions of selected "viewpoints" from both loci in two lymphoblastoid cell lines (LCLs) derived from control individuals using an adapted version of the 4C method (4C-seq: Circularized Chromosome Conformation Capture combined with multiplexed high-throughput sequencing)<sup>51, 52, 54, 100</sup> (**Methods and Table S3**). Despite the limitations of the study of LCLs (**Methods**), these experiments are worth pursuing as studies have shown that chromatin contacts are stable across cell lines and tissues regardless of contacted-gene expression status<sup>3</sup> and that LCL transcriptome profiles can be recapitulated in other tissues and species<sup>12</sup>. Specifically, our previous analyses of LCL transcriptomes showed that genes whose expression correlated with dosage of the 16p11.2 locus are significantly enriched in genes associated with ASD and ciliopathies both in human LCLs and mouse cortex<sup>12</sup>. In particular, we identified chromosomal regions that physically

16p11.2 220 kb CNVs



2p15-16.1 CNVs



**Figure 2: Phenotypic characterization of carriers of 16p11.2 BP2-BP3 and 2p15 rearrangements**

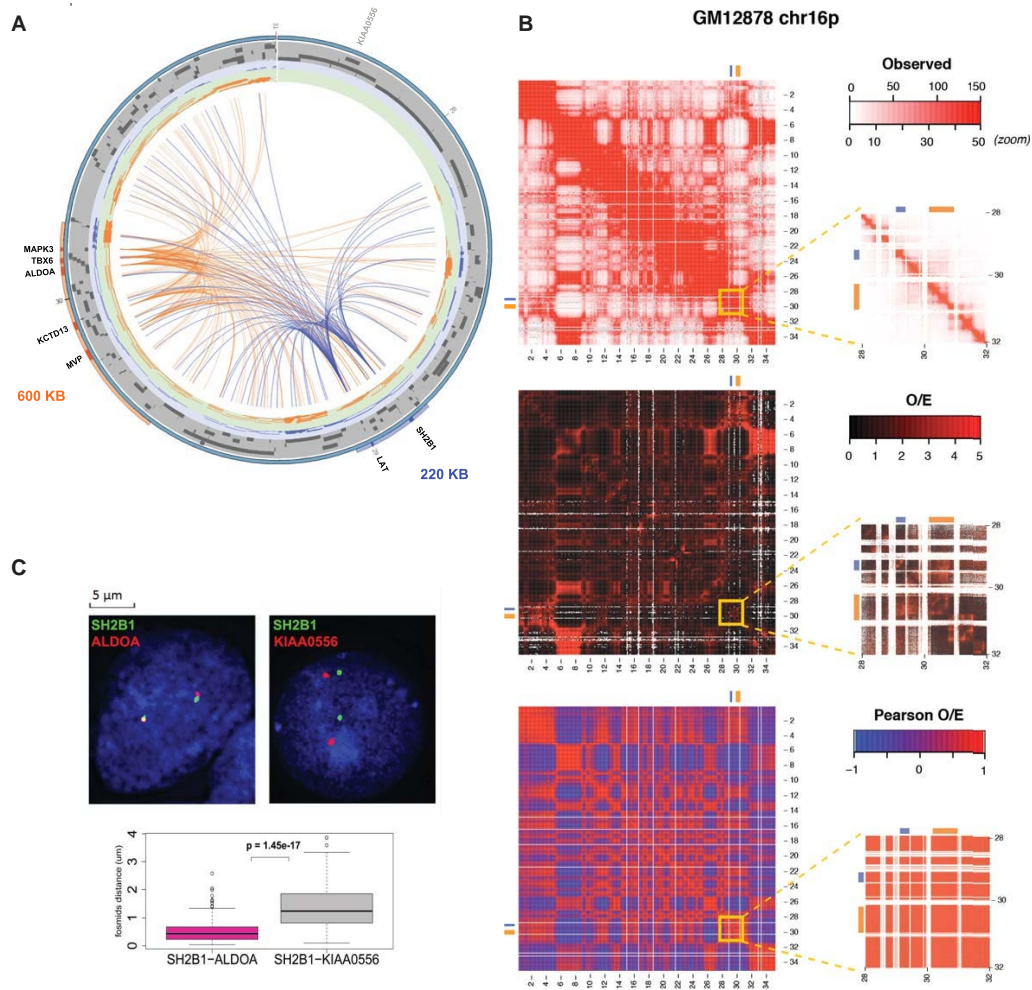
Distribution of Z-score values of BMI **(A)** and head circumference **(B)** in unrelated carriers of the 16p11.2 220kb BP2-BP3 deletion (red) and duplication

(blue) taking into account the normal effect of age and gender observed in the general population as described in <sup>18</sup>. The general population has a mean of zero. **(C)** Comparison of the genomic breakpoints of 2p15 deletions (red bars) and duplications (blue bars) in 26 and 9 unrelated carriers, respectively. The breakpoints' coordinates are detailed in **Table S2**. The genes mapping within the interval and cytobands' positions are shown above, while the extent of the critical region is indicated by a black bar. Distribution of Z-score values of BMI **(D)** and head circumference **(E)** in carriers of the 2p15 deletion (red) and duplication (blue).

associate with the promoters of *MVP*, *KCTD13*, *ALDOA*, *TBX6* and *MAPK3*, five genes mapping to the BP4-BP5 interval, and *SH2B1* and *LAT*, two genes mapping to the BP2-BP3 one. These were investigated based on their potential role in the phenotype (**Figure S2; Methods**)<sup>27, 55, 57-59</sup>.

Genome-wide we identified an average of 265 BRICKs (Blocks of Regulators In Chromosomal Kontext; FDR≤1%), *i.e.* 3-dimensionally interacting genomic fragments, for the seven viewpoints (range: 168-442; **Table S6-S12**). In particular, we observed complex chromatin looping between genes located in the proximal BP4-BP5 and those mapping both to the distal BP2-BP3 region and the equidistant downstream region rich in Zn-finger genes (**Figure 1, 3A**). For instance each of the nine genes of the BP2-BP3 interval (*ATXN2L* (MIM#607931), *TUFM* (MIM#602389), *SH2B1*, *ATP2A1* (MIM#108730), *RABEP2* (MIM#611869), *CD19* (MIM#107265), *NFATC2IP* (MIM#614525), *SPNS1* (MIM#612583) and *LAT*) is contacted by at least one of the five assessed viewpoints in the BP4-BP5 interval (**Figure 1, 3A**). We reciprocally validated these chromatin interactions using the promoters of *SH2B1* and *LAT* as viewpoints (e.g. the chromatin loops of *MVP*, *KCTD13*, *ALDOA*, *TBX6* and *MAPK3* with *SH2B1* are all recapitulated using *SH2B1* as viewpoint; **Figure 3A; Table S11-S12**). The preferential contacted domain of the BP2-BP3 viewpoints extends proximally to the BP4-BP5 and Zn-finger gene-rich regions (**Figure 3A-B, S3**). Inversely, significantly less interactions are called in the gene-rich and equidistant distal region (t-test

$p=0.011$ , **Figure S3**), suggesting that these interactions do not merely reflect the spatial clustering of gene-dense regions.



**Figure 3: Chromatin interactions between the 16p11.2 600kb BP4-BP5 and 220kb BP2-BP3 genomic intervals.**

**(A)** Circos plot representation of the chromatin loops identified in the human chromosome 16 27.5 to 31.0Mb window. The 220kb BP2-BP3 and 600kb BP4-BP5 intervals are depicted by blue and orange bars on the peripheral circle, respectively. Darker sections indicated the positions of the viewpoints. Central blue and orange lines indicate the chromatin interactions corresponding to BP2-BP3 and BP4-BP5 viewpoints, respectively. Note the quasi absence of loops between the BP2-BP3 viewpoints (*LAT* and *SH2B1*) and the 27.5-28.4Mb region. The mapping position of the *KIAA0556* gene, used as control locus in FISH experiments, is indicated.

**(B)** High resolution Hi-C chromosome conformation capture results obtained in reference<sup>75</sup> with the GM12878 LCL within the chromosome 16 0-34Mb window (left panels) and zoom in within the 28-31 Mb region encompassing the two CNVs (5kb resolution; right panels). The positions of the 220kb BP2-BP3 and 600kb BP4-BP5 intervals are shown by blue and orange bars, respectively. Observed (top panel), observed/expected (central panel) and Pearson correlation results are presented (bottom panel).

**(C)** Fluorescence *in situ* hybridization experiments show colocalization of *SH2B1* foci (green) that map to the 220kb BP2-BP3 interval with *ALDOA* foci (red) that map to the 600kb BP4-BP5 genomic interval (left panel) but not with the equidistant *KIAA0556* (red) foci (central panel). The distribution of interphase nuclei distances between the *SH2B1* and *ALDOA* (deep pink) and *SH2B1* and *KIAA0556* foci (grey) are shown in the lower panel. The mapping positions of *ALDOA*, *SH2B1* and *KIAA0556* are indicated in **(A)**.

We confirmed the genomic interaction between the 600kb BP4-BP5 and 220kb BP2-BP3 intervals using fluorescence *in situ* hybridization (FISH). This independent method showed that the BP2-BP3-mapping *SH2B1* locus was significantly closer to the BP4-BP5-encompassed *ALDOA* locus than to a control region, the *KIAA0556* locus, situated equidistantly on its telomeric side (median *SH2B1-ALDOA* and *SH2B1-KIAA0556* distances=0.43 and 1.24  $\mu\text{m}$ , respectively; Wilcoxon rank sum test,  $P=1.45e^{-17}$ )(**Figure 3C, Figure S4, Methods**). We also examined published Hi-C (genome-wide conformation capture) and high-resolution Hi-C data from LCLs. Although they cannot confirm our chromatin connections given their limited resolution, they support a preferential 3-dimensional proximity of these two regions<sup>73,75</sup> (**Figure 3B**). Concordant results were found in both human IMR90 fibroblasts and ES cells<sup>26,101</sup> and mouse cortex and ES cells suggesting the conservation of the structure of this topological

associated domain (TAD) across species and tissues<sup>101</sup>, as recently shown for the TAD spanning the *WNT6/IHH/EPHA4/PAX3* (MIM#604663; #600726; #602188; #606597) locus<sup>3</sup>.

As chromatin interactions were determined in normal diploid context, we next assessed the effect of BP4-BP5 CNVs on these chromatin loops. We identified genomic fragments that interact with the same seven viewpoints in LCLs of two BP4-BP5 deletion- and two reciprocal duplication-patients (**Methods**). We observed a genome-wide decrease in the number of BRICKS per viewpoint ranging from 27 to 84%, suggesting that both rearrangements triggered dramatic reorganizations. Consistent with this hypothesis, the *SH2B1* viewpoint, whose copy number is not affected by the proximal BP4-BP5 CNV, shows a 36% reduction in the amount of interacting regions (all BRICKS listed in **Table S13-S26**). We compared the 4C-seq results from control individuals and the four patients and identified, across all conditions and considering all viewpoints, 1193 genes with significantly modified chromosomal contacts (FDR<1%, **Methods; Table S27**). These results support the idea that large structural rearrangements perturb the 3D genomic structure by modifying both cis- and trans-contacts.

#### [Perturbations of the chromatin interactions' landscape at 16p11.2 are associated with gene expression modification](#)

Our results show that the gene-rich BP2-BP3 and BP4-BP5 16p11.2 intervals, whose CNVs are linked to overlapping phenotypes, are reciprocally engaged in complex chromatin looping as determined by 4C, FISH and Hi-C. The recent discovery of multigene complexes where chromosomal loops orchestrate co-transcription of interacting genes<sup>2, 102</sup> is suggestive of functional implications for the chromosomal contacts between the BP2-BP3 and BP4-BP5 intervals.



To assess this possibility, we first used our recent association analyses of population-wide transcription factor DNA binding (PU.1 and RPB2 - the second largest subunit of RNA polymerase II), histone modification enrichment patterns (H3K4me1, H3K4me3, and H3K27ac) and gene expression measured by ChIP-seq and RNA-seq in LCLs derived from 47 European unrelated individuals whose genomes were sequenced in the frame of the 1000 Genomes Project<sup>95</sup>. We measured the extent of quantitative coordination of natural inter-individual variation between pairs of these six molecular phenotypes at putative regulatory regions mapping within cytoband 16p11.2 and identified coordinated behavior in terms of mapping enrichment. For example, we found association between active regulatory regions mapping within the BP4-BP5 interval and expression of BP2-BP3 genes (**Figure S5**)<sup>95</sup>, consistent with the notion that some of the chromatin loops uncovered between these two intervals might bring together regulatory elements and genes.

Secondly, we examined if genes involved in primary cilium function and related pathways, that are modified in BP4-BP5 deletion patients' cells<sup>12</sup>, are also changed in BP2-BP3 deletion carriers (**Methods**). We found that the ciliary genes *BBS4*, *BBS7*, *SMAD2*, *XPOT* and *NUP58* are correspondingly modified in LCLs derived from both BP4-BP5 and BP2-BP3 deletion carriers (**Figure S6**). The interplay between the 600kb BP4-BP5 and the 220kb BP2-BP3 interval is further substantiated: (i) by published data showing perturbed expression of genes mapping within BP2-BP3 distal interval (i.e. *LAT*, *SPNS1* and *ATP2A1*) in cells derived from BP4-BP5 patients<sup>26</sup>; as well as (ii) by the observation that, within the top-10 genes correlated with *SH2B1* expression according to GeneProf ([www.geneprof.org](http://www.geneprof.org)), three (*ZNF500*; *CDAN1* (MIM#607465); *LRRC14*) are contacted by viewpoints located in the BP4-BP5 region and two (*PIGO* (MIM#614730); *TUBGCP6* (MIM#610053)) are differentially expressed in BP4-BP5 CNV patients.

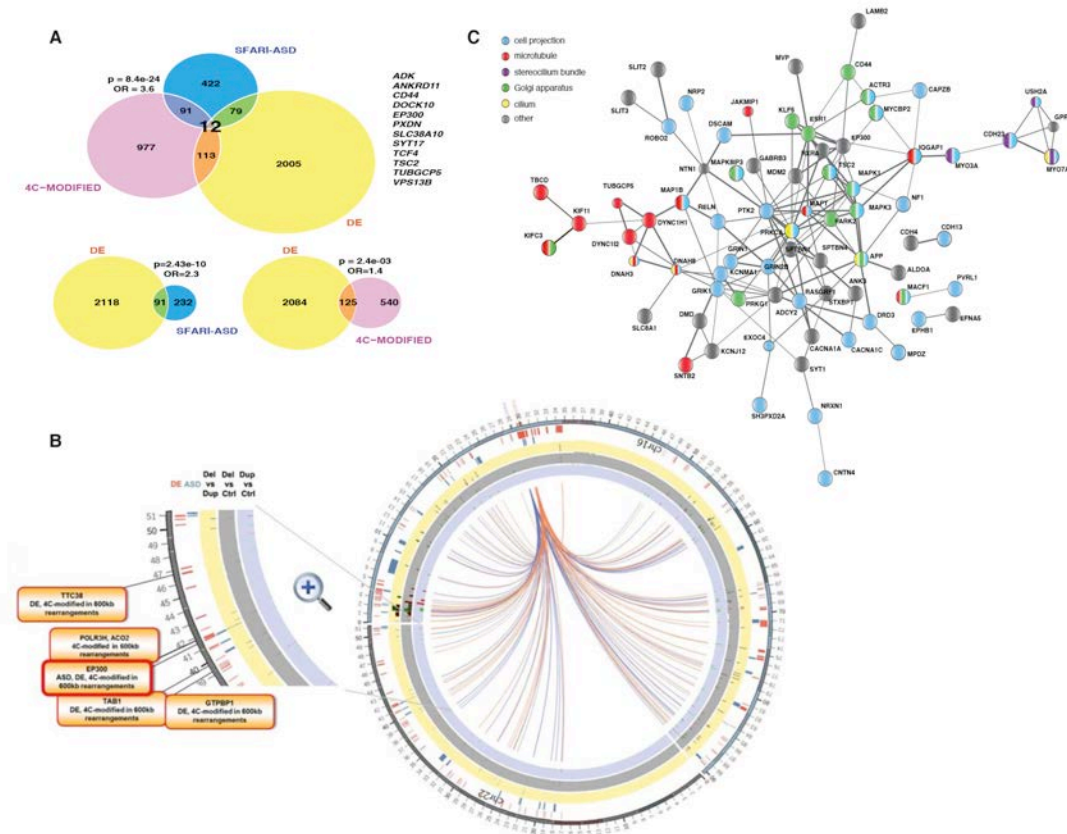
We then assessed whether the structural chromatin changes identified in the 600kb BP4-BP5 CNVs carriers (i.e. 1193 genes with significantly modified chromosomal contacts, FDR<1%) are paralleled by transcriptome modifications identified in <sup>12</sup> (2,209 significantly differentially expressed genes, FDR ≤ 5%; uniquely mapping probes)). We find a significant overlap between the two gene lists as 125 genes with modified chromatin loops are concomitantly differentially expressed in 16p11.2 600kb BP4-BP5 CNV carriers' cells (125/665 4C-modified genes with detectable expression (see Methods); Fisher's enrichment test: OR=1.4,  $P=0.002$ ; **Table S28-S30; Figure 4A**). There is no correlation between quality of the change in 4C contact (increased or decreased) and the sense of the perturbation in gene expression, i.e. a loss of contact does not necessarily imply a decrease in expression and vice versa (**Figure S7, Table S31**).

These results show that the BP2-BP3 and BP4-BP5 loci are linked by chromatin loops, coordinated molecular phenotypes and co-regulation of genes.

#### Genomic regions contacted by 16p11.2 viewpoints are associated with autism, BMI and HC phenotypes and enriched in ciliary genes

Consistent with the notion that chromatin contacts connect biologically-related genes, BRICKs genes are enriched for genes that encode proteins that interact ( $p<0.01$  for all viewpoints; **Table S32; Methods**). Processes overrepresented within BRICKs genes are listed in **Table S33**. They are also enriched for genes listed in SFARI Gene (<https://sfari.org/>), an annotated list of candidate genes for ASD (union of the SFARI Syndromic, High Confidence and Strong Candidate Gene categories (Categories S+1-2), 13/76, OR=2.9,  $P=0.0014$ ) and for ASD-associated genes identified by whole exome studies (8/50 of the "high confidence *de novo*" ASD-associated genes<sup>86, 87</sup>; OR=2.7,  $P=0.016$ ; **Table S34**). The BP4-BP5 and BP2-BP3 viewpoints contacted genes include, *GRID1* (MIM#610659) and *PTEN* at 10q23.2-q23.31, *USP34/XPO1* at 2p15, but also genes linked to HC phenotypes like *CHD1L* (MIM#613039) at 1q21.1<sup>103-112</sup> and *EP300* (MIM#602700) at 22q13<sup>113</sup> (**Figure 5A, Figure S8, Table S35**). To validate these interactions, we

verified a subset by FISH (e.g. *MVP-PTEN* and *KCTD13-USP34/XPO1*; **Figure 5B-E, Methods**). The comparison of distributions of Hi-C scores in selected vs non-selected BRICKS for each of our 4C viewpoints (**Methods**) further demonstrates the reproducibility of the 4C results (**Figure S9**).



**Figure 4: Extensive overlap between differentially expressed genes and loci that show modified chromatin interactions.**

**(A)** Top panel: weighted Venn diagram showing the overlap between the 2209 genes that are differentially expressed in 16p11.2 rearrangement carriers (DE, yellow disk;  $FDR \leq 5\%$  <sup>12</sup>), the 1193 genes that show modified chromatin interactions in 16p11.2 rearrangement carriers (4C-modified, purple disk; only 665 with detectable expression are considered for the DE enrichment; see **Table S31**) and the 604 genes listed in SFARI Gene (<https://sfari.org/>; 323 expressed), an annotated list of candidate genes for ASD (ASD; blue disk). The numbers of common genes are indicated and the twelve 4C-modified, DE and ASD-SFARI genes are specified on the right. Bottom panels: weighted Venn diagrams showing the overlap between the DE genes and the LCLs-expressed ASD and 4C-

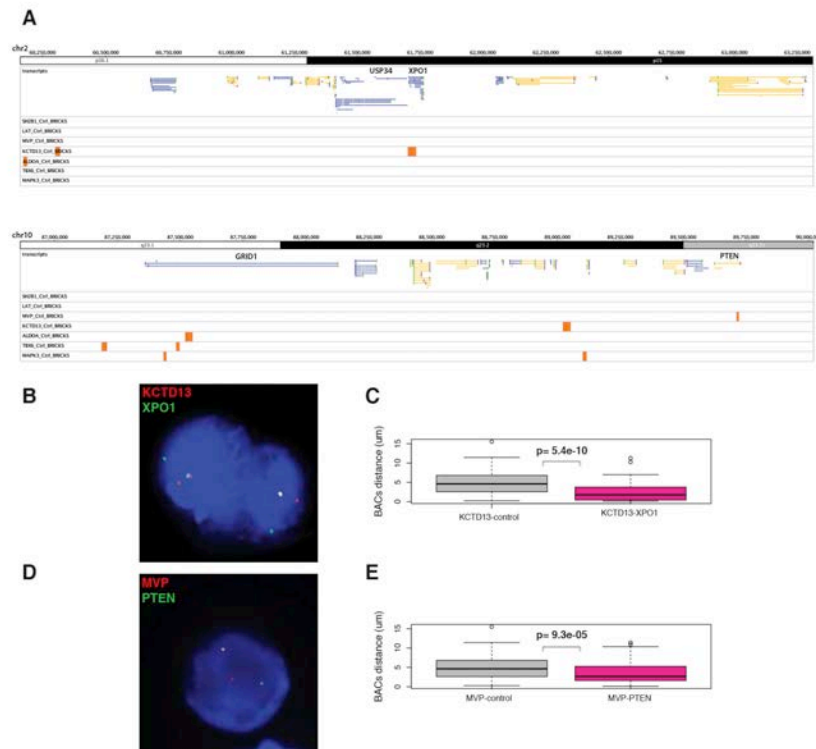
modified genes (lower left and right, respectively). **(B)** Circos plot representation of the chromatin loops identified in human chromosomes 16 and 22 (right hand panel). The 220kb BP2-BP3 and 600kb BP4-BP5 intervals are depicted by blue and orange bars on the peripheral circle, respectively. Central blue and orange lines indicate the CNVs-modified chromatin interactions corresponding to BP2-BP3 and BP4-BP5 viewpoints, respectively. Ticks on the three internal rings indicate BRICKS with significantly modified interactions between 16p11.2 600kb BP4-BP5 duplication and control samples (light blue ring), between 16p11.2 600kb BP4-BP5 deletion and control samples (dark grey), and between 16p11.2 600kb BP4-BP5 deletion and duplication samples (yellow). Blue and red ticks on the most external rings denote genes differentially expressed in 16p11.2 patients (DE) and SFARI-ASD-associated genes (ASD), respectively. A zoomed-in view with examples of genes with modified chromatin interactions mapping within the 22q13 cytoband is presented in the left hand panel. **(C)** The 1193 genes that show modified chromatin interaction in 16p11.2 cells with 16p11.2 600kb BP4-BP5 rearrangements encode proteins that interact. The confidence view interaction network of the encoded proteins corresponding to the enriched GO terms (GO:0030030 cell projection organization, GO:0042995 cell projection, GO:0030173 integral to Golgi membrane, GO:0000138 Golgi trans cisterna, GO:0034504 protein localization to nucleus and GO:0005874 microtubule) is visualized with STRING. Proteins belonging to cell projection (blue), microtubule (red), Golgi apparatus (green), stereocilium bundle (purple) and cilium (yellow) process/cell components are highlighted by colored beads. Disconnected nodes are not shown.

Reminiscent of our transcriptome findings<sup>12</sup>, enrichment analysis of the 1193 genes with modified chromosomal contacts showed significant over-representation of ciliary genes<sup>88</sup> (40/493, OR=1.47,  $P=0.030$ , **Table S29-30**), OMIM terms associated with dysfunction of ciliary structures (**Table S28**) and candidate genes for ASD (**Table S29-S30**). We showed previously<sup>12</sup> that differentially expressed genes were similarly enriched for SFARI-ASD-associated genes (91/323 with detectable expression; OR=2.3,  $P=2.43e^{-10}$ ); **Figure 4A**). Notably five (*TCF4* (MIM#602272), *EP300*, *ADK* (MIM#102750), *TUBGCP5* (MIM#608147), *VPS13B* (MIM#607817)) of the twelve genes that are concurrently SFARI-ASD, differentially expressed and modified in 4C contacts (OR=5.01,  $P=1.5e^{-05}$ ) were previously associated with head circumference changes<sup>114-117</sup> (**Figure 4A-B; Table S29-S30**).

### Phenotypes associated with 2p15-16.1 CNVs

Our results suggest that chromatin interactions captured by the BP4-BP5 and BP2-BP3 viewpoints and their perturbations can be exploited to identify additional genes/loci, which when genetically perturbed, are associated with similar pathways, diseases and phenotypes. We challenged this hypothesis by assessing whether the phenotypic features of 2p15-p16.1 deletion and duplication carriers overlap with those of carriers of the 600kb BP4-BP5 and 220kb BP2-BP3 rearrangements. Haploinsufficiency of the chromatin-contacted *USP34/XPO1* was suggested to be responsible for the 2p15-p16.1 deletion syndrome (MIM#612513) phenotypes that included ID, ASD, microcephaly, dysmorphic facial features and a variety of congenital organ defects<sup>111, 112</sup>.

We collected clinical data on 26 and 9 unrelated 2p15-p16.1 deletion and duplication carriers, respectively (**Figure 2C; Table S2**). Comparison of data on BMI and HC of both variants pinpoints mirror phenotypes for these two traits (**Figure 2D-E**). Whereas we do not formally demonstrate a mechanistic and functional link between the 600kb BP4-BP5 and the 2p15 interval, it should be noted that the *KCTD13-USP34/XPO1* interactions are present in controls LCLs, but neither in 16p11.2 deletion nor duplication LCLs (see below). Furthermore, five (*C2orf74*, *COMMD1* (MIM#607238), *FAM161A* (MIM#613596), *PEX13* (MIM#601789), *PUS10* (MIM#612787)) and 11 (*AHSA2*, *BCL11A* (MIM#606557), *PAPOLG*, *REL* (MIM#164910), *USP34*, *XPO1*) of the 13 genes mapping within the 2p15-16.1 syndrome minimal overlapping interval<sup>111, 112</sup> show perturbed expression levels in 16p11.2 CNV patients with cutoffs of 5 and 15% FDR, respectively. Thus the cis- and trans-chromatin contacts we uncovered bridge genomic regions, whose rearrangements are associated with ASD and mirror phenotypes on BMI and HC.



**Figure 5: Examples of 16p11.2 viewpoints chromatin-contacted regions**  
**(A)** Examples of regions (BRICKS) interacting with 16p11.2 viewpoints showing some of the interacting genes, *i.e.* *GRID1*, *PTEN*, and *USP34/XPO1*. Other examples (*CHD1L* and *EP300*) are shown in **Figure S8**). Fluorescence *in situ* hybridization experiments show colocalization of the 600kb BP4-BP5 interval-encompassed *KCTD13* (red) and 2p15-mapping *XPO1* foci (green) **(B)** and the 600kb BP4-BP5 interval-encompassed *MVP* (red) and 10q23.31-mapping *PTEN* foci (green) **(D)**. The distribution of interphase nuclei distances between *KCTD13* and *XPO1* **(C)** and between *MVP* and *PTEN* **(E)** foci are compared with to those between *KCTD13/MVP* and *MARK4* (control) foci (25% and 14% co-localization versus 2% with the control locus; Fisher's test enrichment:  $P=6.9e^{-05}$  and  $P=0.01$ , respectively; median *MVP/KCTD13-USP34/XPO1* distances = 1.76, *MVP/KCTD13-PTEN* = 2.61 and *MVP/KCTD13-MARK4* = 4.96  $\mu\text{m}$ ; Wilcoxon rank sum test,  $P=5.4e^{-10}$  and  $P=9.3e^{-05}$ , respectively).

## Discussion

The 16p11.2 600kb BP4-BP5 rearrangements allow investigation of molecular mechanisms underlying the co-morbidity triad of neurodevelopmental disorders, energy imbalance and HC alterations, all associated with changes in gene dosage. To identify pathways that are perturbed when the dosage of this region is modified we cataloged the chromosomal contacts of genes mapping within this genomic interval. Using chromosome conformation capture we uncovered a network of chromatin loops with genes previously associated with ASD and HC demonstrating the pertinence of this approach. We show, for example, that the 16p11.2 phenotype drivers *MVP* and *MAPK3* promoters have long-range chromatin interactions with *PTEN* and *CHD1L*, respectively. The MVP protein regulates the intracellular localization of PTEN<sup>118</sup>, a dual-specificity phosphatase that antagonizes PI3K/AKT and Ras/MAPK signaling pathways. Both *PTEN* germline mutations in humans and targeted inactivation in mice are associated with macrocephaly/ASD syndrome (MIM#605309)<sup>103-106, 119</sup>. Congruently, germline mutations in the Ras/MAPK pathway cause a group of syndromes frequently regrouped under the term RASopathies, recently shown to affect social interactions<sup>120, 121</sup>. We corroboratingly revealed that expression of *PTEN* pathway members is sensitive to gene dosage at the 16p11.2 locus<sup>12</sup>. *CHD1L* was suggested to be a major driver of the phenotypes associated with 1q21.1 rearrangements (OMIM#612474; #612475)<sup>107, 108</sup>. Analogous to 16p11.2, deletions and duplications of this interval are linked to micro- and macrocephaly, respectively<sup>109</sup>.

These results suggest that chromatin interactions, even when tested in peripheral tissues (such as LCLs) not considered to play a central role in the resulting neurodevelopmental phenotype, could reveal genes or pathways, which are co-regulated and associated with similar phenotypes. Several studies have shown that the contacted domains can be highly stable across species and cell lines (even when the contacted genes are not expressed)<sup>3, 101</sup>, supporting the notion that patient-derived samples can provide direct insight into regulatory abnormalities and that LCLs still contain valuable information for the study of

the patients phenotype<sup>12</sup>. Consistent with this hypothesis, we established that dosage perturbation of two chromatin-contacted loci, the *cis*-contacted 16p11.2 220kb BP2-BP3 and the *trans*-interacting 2p15 intervals, are associated with mirror phenotypes on BMI and HC. Similar regulatory cores engaged in multiple physical interactions were recently described, e.g. the 8q24 oncogenetic locus<sup>122</sup>.

The physiological relevance of the underlying chromatin architecture is further exemplified by the extensive overlap between differentially expressed genes and the loci that show modified loops upon dosage changes of the 16p11.2 600kb BP4-BP5 interval. Together with the observed enrichments in various pathology-relevant gene ontology terms and pathways, this sheds light on a possible “chromatin hub” role of the 16p11.2 locus in the observed phenotypes. The twelve genes that are concurrently SFARI-ASD, differentially expressed and modified in 4C contacts include i) *VPS13B*, whose mutations cause Cohen syndrome (OMIM#216550), an autosomal recessive disorder characterized by ID, microcephaly, retinal dystrophy and truncal obesity<sup>123</sup>; ii) *TCF4*, whose haploinsufficiency is associated with Pitt-Hopkins syndrome (OMIM#610954) typified among other traits by ID, recurrent seizures and microcephaly (Of note, the expression level of this transcription factor of the Wnt/ $\beta$ -catenin signaling cascade was shown to be altered in the cortex of mice models engineered to carry one or three copies of the 16p11.2 orthologous region<sup>26</sup>); iii) the *TSC2* (MIM#191092) tuberous sclerosis (OMIM#613254) gene, which encodes an inhibitor of mTORC1 signaling that limits cell growth and was linked to ciliary dysfunction<sup>124</sup>; and iv) *EP300*, which is mutated in a form of Rubinstein-Taybi syndrome 2 (OMIM#613684) associated with a more severe microcephaly<sup>114, 125</sup>.

Whereas the functions and natures of the detected contacts remains to be elucidated, recent reports show that transcription of co-regulated genes occurs in the context of spatial proximity, which is lost in knockout studies<sup>2, 102</sup>. While we cannot exclude that the observed proximal positions of genes is brought about by proteins that do not directly interact with any of them <sup>126</sup> or occur

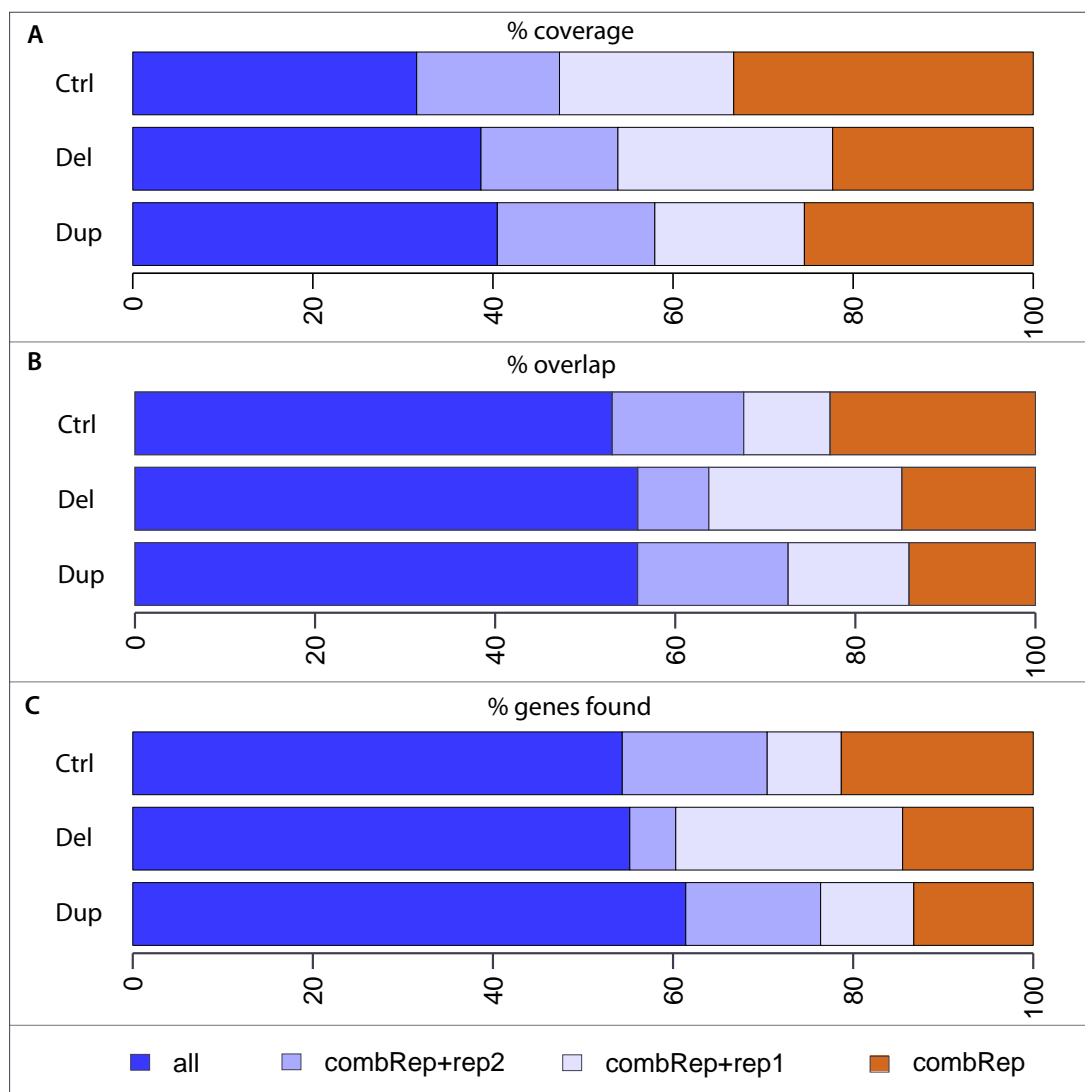


through their use of identical transcription factory for example <sup>127, 128</sup>, we *de facto* witness that expression of multiple genes of converging pathways are modified and that the chromatin-contacted genes are encoding proteins of overlapping interactomes (**Figure 4C, S10-S11**). For example, *MVP* loops with genes implicated in maintenance of cell polarity ( $P=2.75e^{-03}$ ) (**Table S33**).

Chromatin spatial organization is conserved, to some extent at least, through evolution<sup>129</sup>. The distal and proximal 16p11.2 regions are physically interacting in both human and mouse cells<sup>73, 101</sup>. This chromatin crosstalk is preserved despite modifications in the human lineage of the orientation of both the BP2-BP3 and the BP4-BP5 regions, as well as doubling of the size of the intervening region. While we cannot rule out that similar studies on the clinically relevant tissues might uncover additional important partners, these studies demonstrate that maintenance of chromatin crosstalk across tissues (from fibroblasts to cortical neurons) and in different lineages lends credence to the use of LCLs and animal models as proxies to study chromatin properties of the human central nervous system, the more likely tissue determining the phenotypes associated with 16p11.2 600kb BP4-BP5 and BP2-BP3 CNVs.

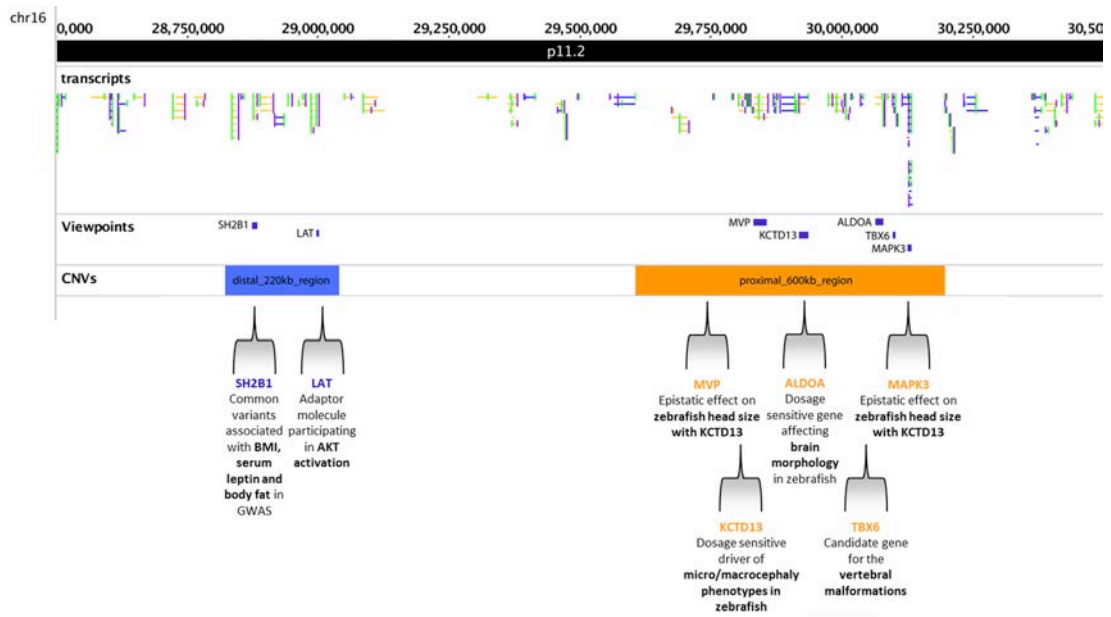
The identified cis- and trans-chromatin contacts bridge loci whose rearrangements result in mirror phenotypes on BMI and HC, as well as involve known ASD candidate genes. While investigations of the 3D genomic structures of additional regions are warranted, the results present here support the idea that the elucidation of chromatin contacts can be proposed as a new and effective tool to unravel genes participating in similar pathways or disease mechanisms, and identify loci associated with overlapping phenotypic manifestations. Our study also suggests that modifications of chromatin interplays play a crucial role in the observed phenotypes.

## Supplementary figures



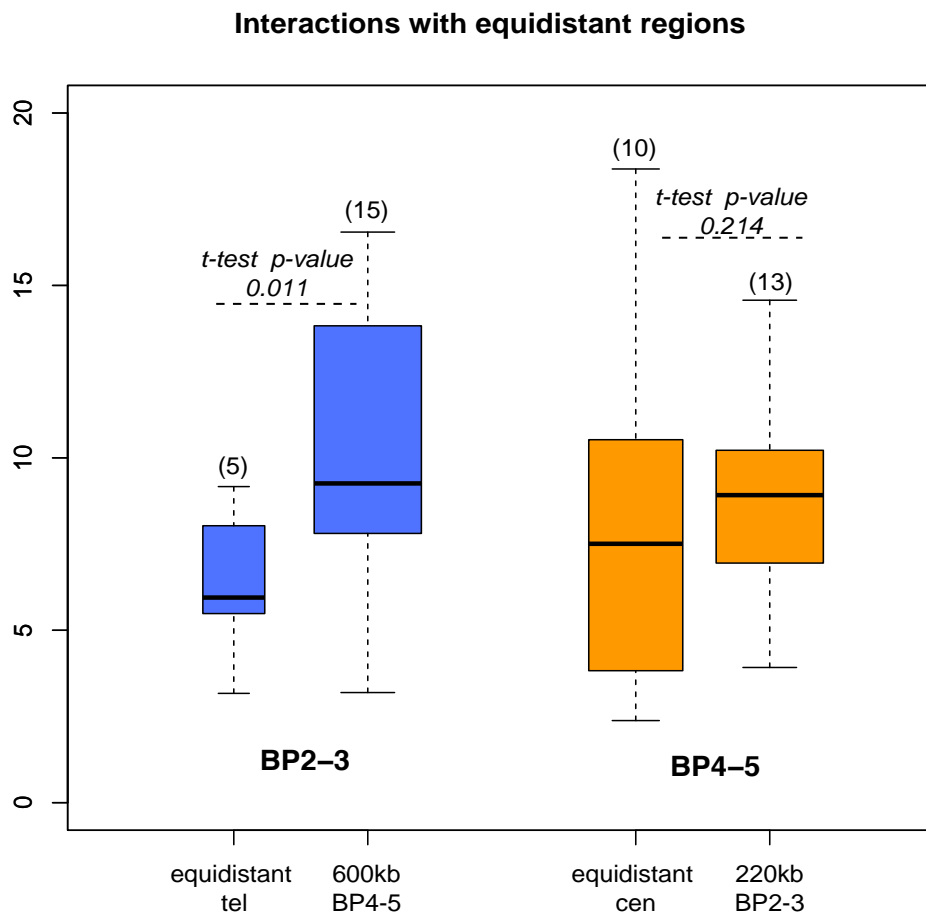
### Figure S1: Comparison of 4C-seq replicates

BRICKS obtained after averaging replicates (combRep) are compared to BRICKS found in each replicate separately (rep1 and rep2) according to **(A)** percentage of the total length of combRep BRICKS also included in the intersection of rep1 and rep2 BRICKS, in only one of the replicates or in none; **(B)** percentage of the combRep BRICKS that overlap with the intersection of rep1 and rep2 BRICKS, with only one of the replicates or with neither; and **(C)** percentage of genes found within combRep BRICKS that are also found in the intersection of rep1 and rep2 BRICKS, in only one of the replicates or in none.



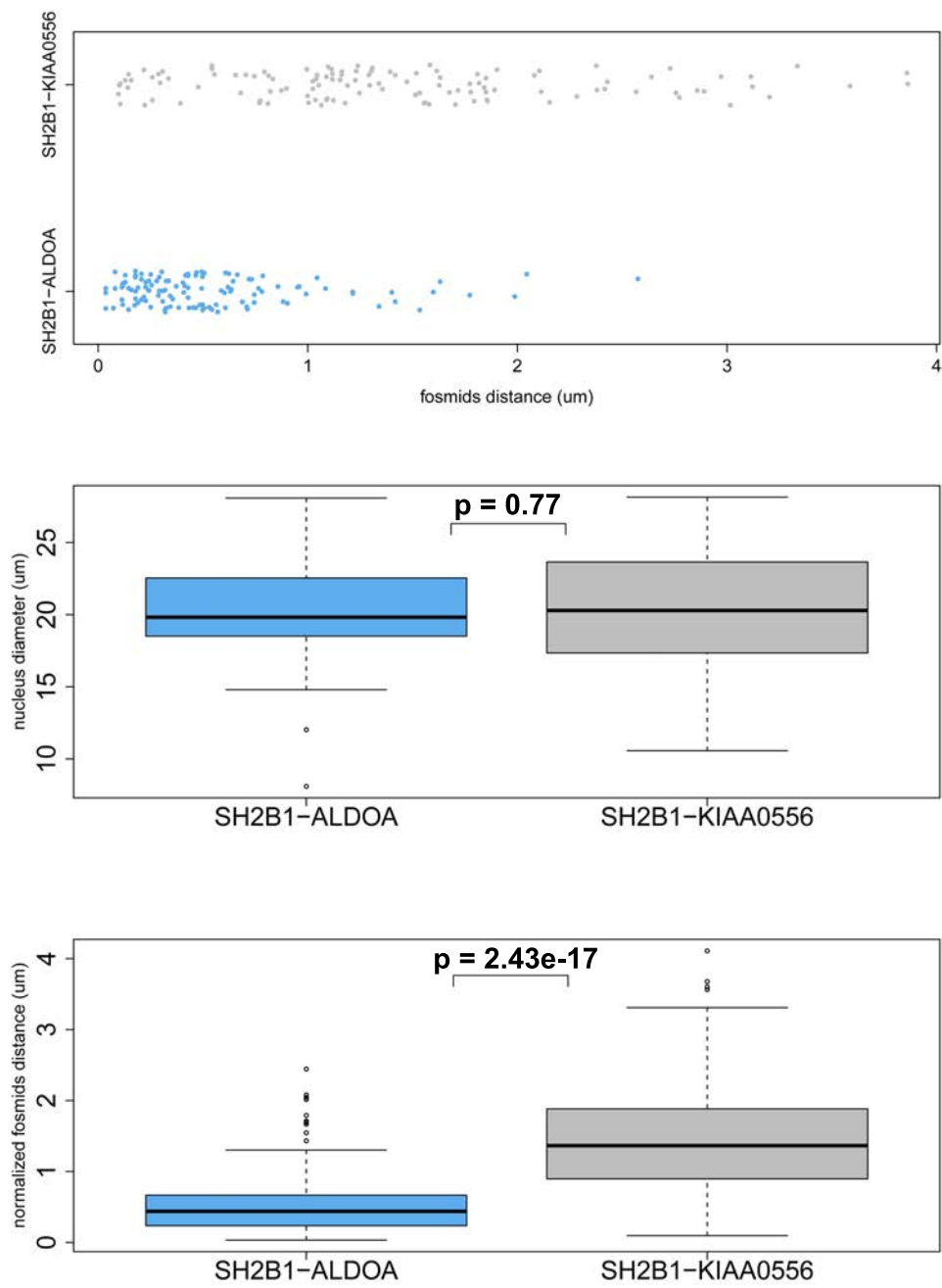
### Figure S2: Selection of 4C viewpoints

Function and mapping position of the genes selected as viewpoints for 4Cseq analysis.



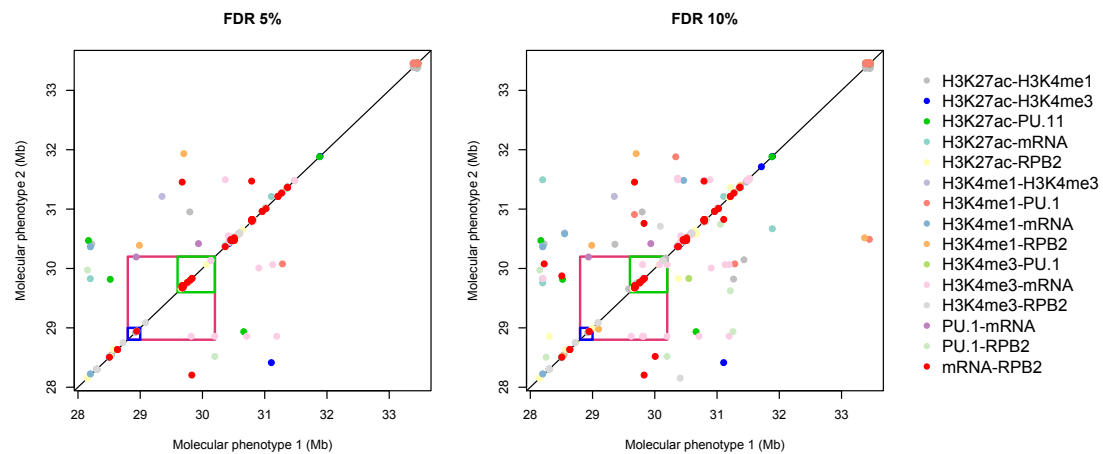
**Figure S3: The 220kb BP2-3 viewpoints display preferential interactions with the BP4-BP5 region compared with the telomeric equidistant region**

Significances of BRICKS identified with viewpoints mapping within the 220kb BP2-BP3 region and mapping within the 600kb BP4-BP5 interval (right blue boxplot) compared with those in the equidistant, same lengths and telomeric region (left blue boxplot). Significances of BRICKS identified with viewpoints mapping within the 600kb BP4-BP5 region and mapping within the 220kb BP2-BP3 interval (right orange boxplot) compared with those in the equidistant, same lengths and centromeric region (left orange boxplot). The number of BRICKs is indicated in brackets and p-values of a one-sided t-test (alternative "less") are reported above each comparison.



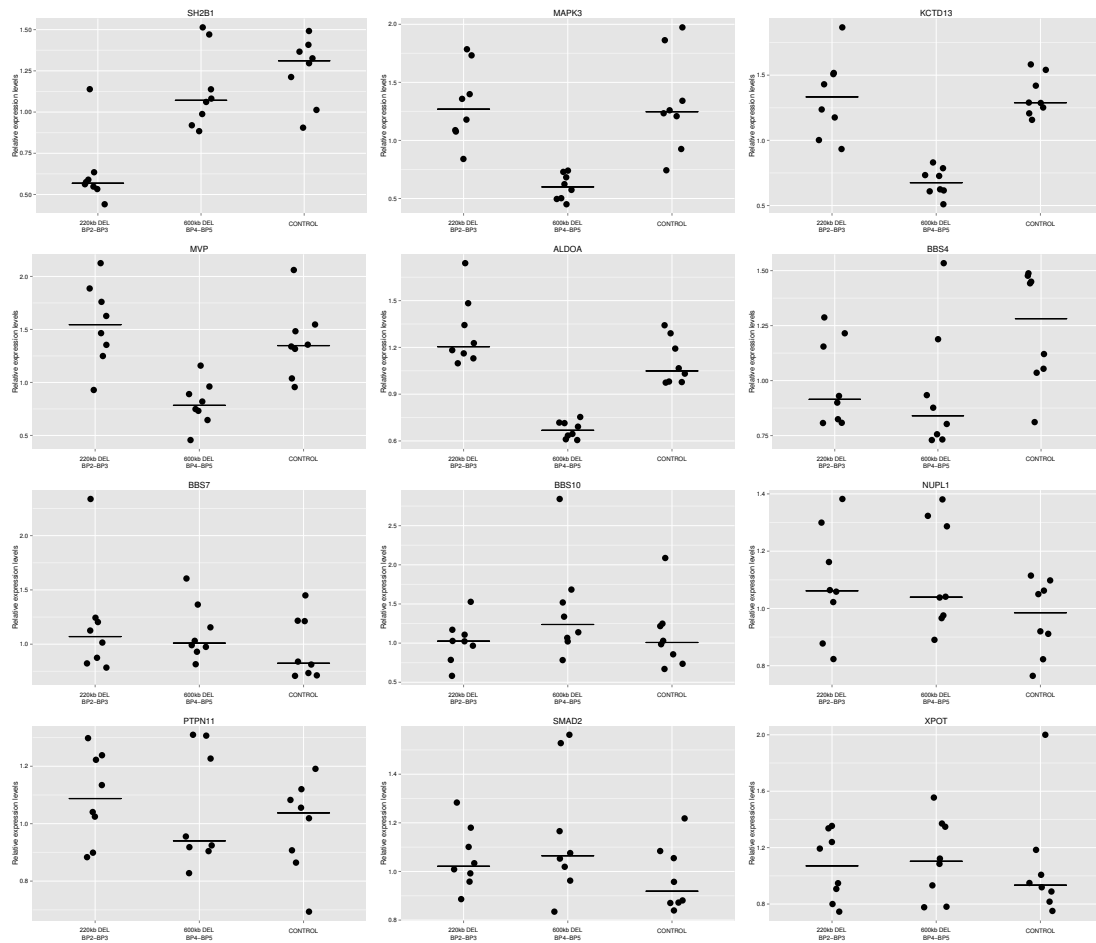
### Figure S4: FISH shows co-localization of the 220kb and 600kb intervals

Fluorescence in situ hybridization experiments show co-localization of SH2B1 that maps to the 220 kb interval with ALDOA that map to the 600kb interval but not with the equidistant KIAA0556 locus. **(A)** Distribution of interphase nuclei foci distances SH2B1-ALDOA (light blue) and SH2B1-KIAA0556 (grey)(n=120 for each experiment). **(B)** Distribution of the measured nucleus diameters in the SH2B1-ALDOA (light blue) and SH2B1-KIAA0556 experiments (grey)(n=60 nuclei). **(C)** Distribution of interphase nuclei foci distances SH2B1-ALDOA (light blue) and SH2B1-KIAA0556 (grey) normalized by the nucleus diameter (n=120 for each experiment).



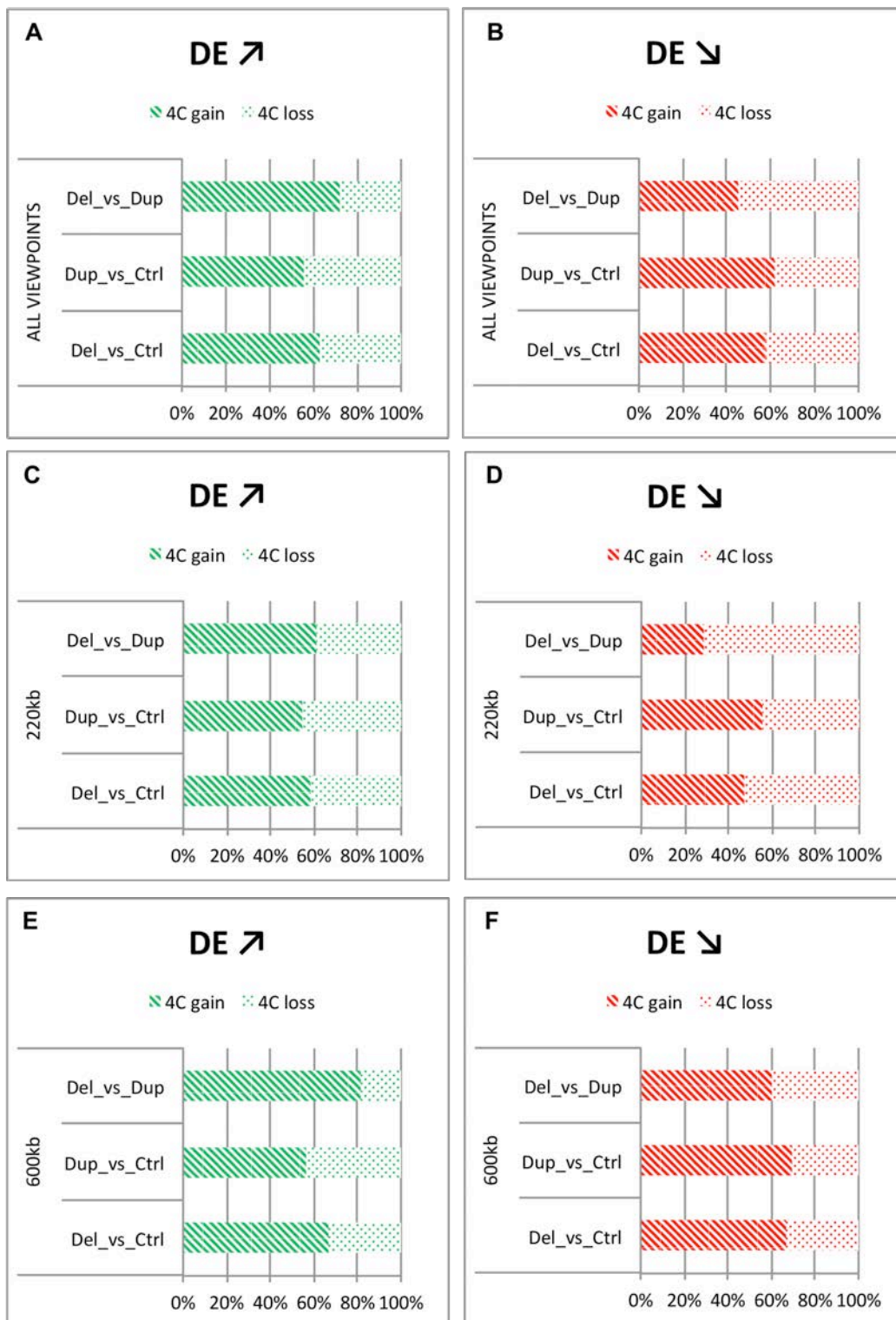
### Figure S5: Associations between molecular phenotypes within the 16p11.2 cytoband

Pairwise molecular associations within the 16p11.2 region (28.1-34.6 Mb) between H3K4me1, H3K4me3, H3K27ac, PU.1 and RPB2 binding and mRNA levels in LCLs of 47 unrelated individuals at 5% (left panel) and 10% FDR (right panel). The positions of the 220kb BP2-BP3 and 600kb BP4-BP5 intervals are delimited by blue and green rectangles, respectively. Associations that support the 4C observed loopings between the 16p11.2 600kb BP4-BP5 and BP2-BP3 intervals are enclosed in the red rectangle. The nature of the molecular phenotypes participating in the associations is indicated on the right.

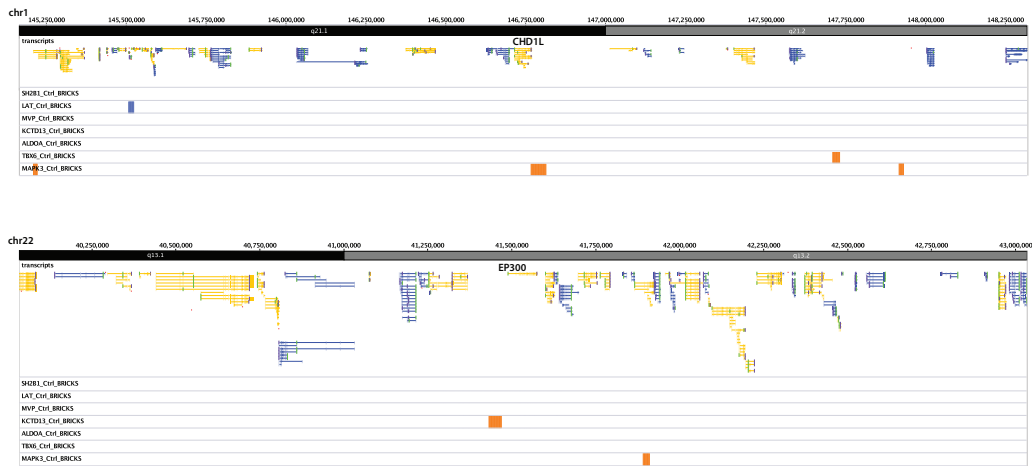


**Figure S6: Ciliopathy gene expression levels are modified in cells of both 16p11.2 600kb BP4-BP5 and 16p11.2 BP2-BP3 deletion carriers.**

RNA levels of genes involved in primary cilium function and related pathways are modified in both 600kb BP4-BP5 and 220kb BP2-BP3 deletion patients' cells. Relative expression levels measured by quantitative PCR of BBS4, BBS7, BBS10, PTPN11, SMAD2, XPOT, NUP58 (a.k.a NUPL1), as well as 16p11.2 imbalanced interval genes ALDOA, KCTD13, MAPK3, MVP and SH2B1 in LCLs of eight unrelated carriers of the 220kb BP2-BP3 deletion, eight unrelated carriers of the 600kb BP4-BP5 deletion and eight unrelated age- and sex-matched control individuals (CONTROL). Note the significant diminution of the hemizygote gene SH2B1, but not of the neighboring normal-copy ALDOA, KCTD13, MVP and MAPK3 in 220kb deletion carriers and the reciprocal diminutions of the hemizygote ALDOA, KCTD13, MVP and MAPK3 but not of the neighboring normal-copy SH2B1 in 600 kb deletion carriers. The bar indicates the median.



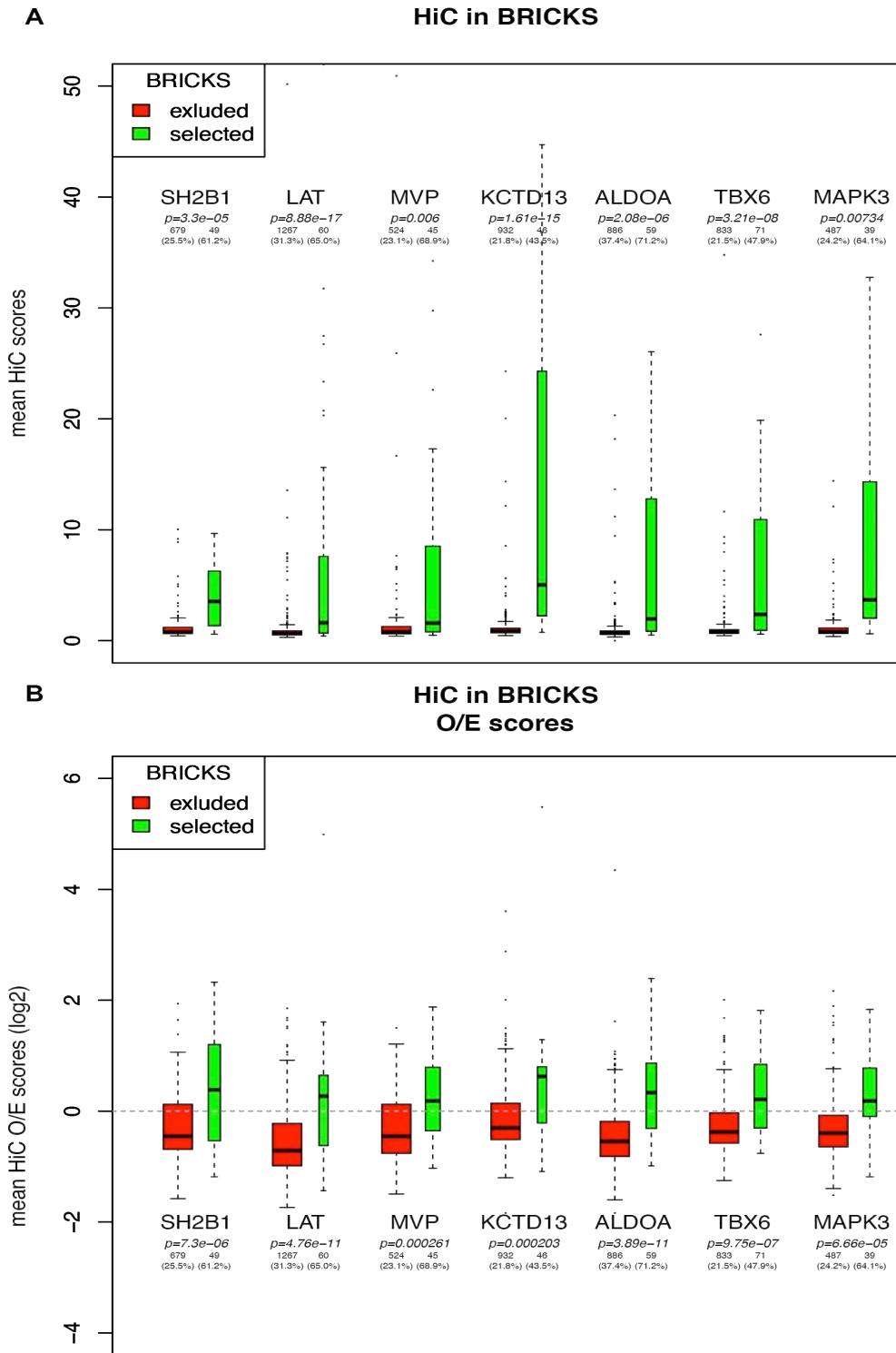
**Figure S7 : 4C contacts perturbation and gene differential expression.** Distribution of gain / loss of contacts for up- and down-regulated genes for each comparison: Deletion (Del) vs. Duplication (Dup), Dup vs. Control (Ctrl) and Del vs. Ctrl. Panels (A) and (B) show up-/down-regulated genes with altered BRICKS found in all viewpoints, panels (C) and (D) in the 220kb viewpoints only, and panels (E) and (F) in the 600kb viewpoints only.



**Figure S8: Chromatin-interacting regions.**

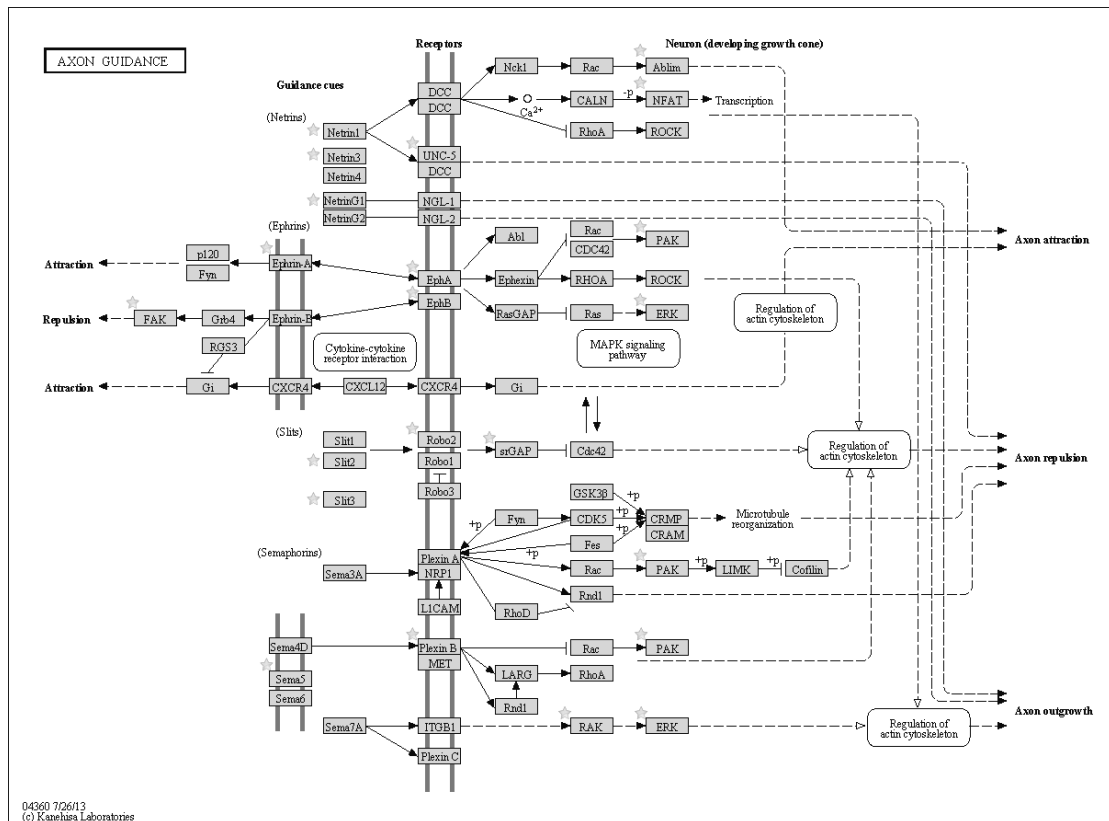
Examples of regions (BRICKS) interacting with 16p11.2 viewpoints showing some of the interacting genes, i.e. CHD1L (A) and EP300 (B). Other examples are shown in Figure 4A.





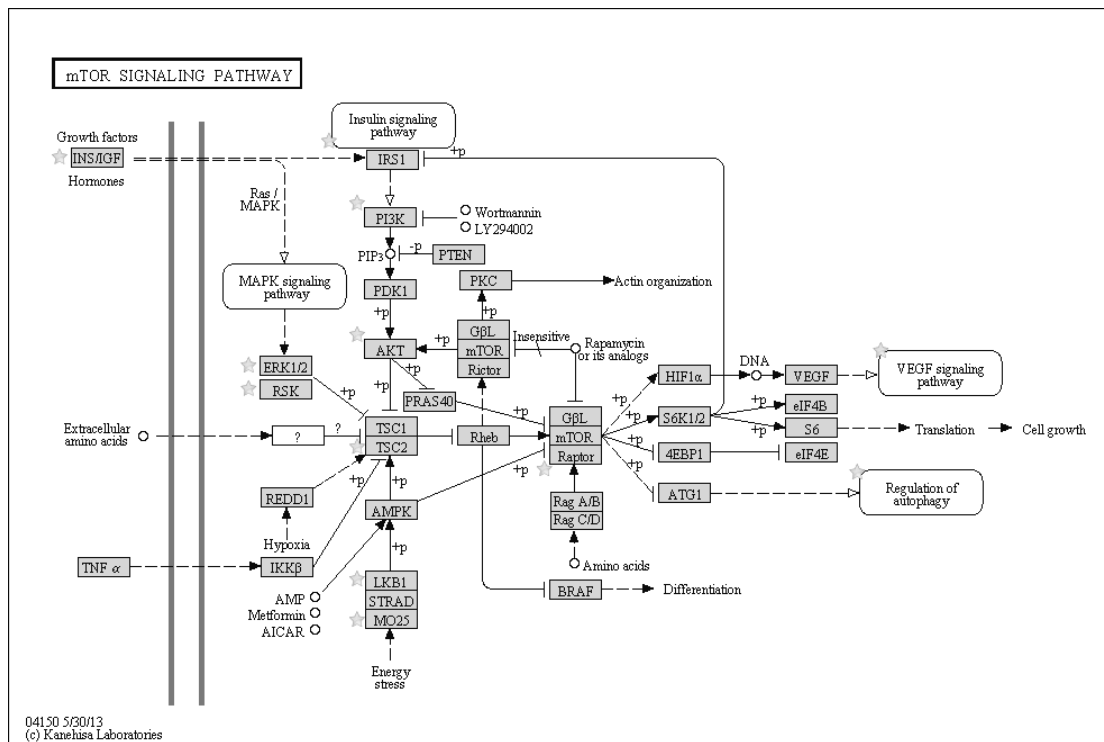
**Figure S9: Distribution of Hi-C scores in selected versus non-selected BRICKS**

Virtual 4C-seq tracks were generated for each viewpoint from the GM12878 Hi-C results of reference 70 (5kb resolution) by extracting the Hi-C vectors from the KR normalized observed (top panel) and observed/expected matrices (bottom panel). BRICKS found with each viewpoint were quantified by the mean Hi-C signals. The p-values of two-sided t-tests are reported for each comparison, together with the number of Hi-C bins and the % of non-NA bins.



**Figure S10: KEGG pathway with modified chromatin interactions**

The 1193 genes that show modified chromatin interaction in cells of 16p11.2 600kb rearrangement carriers are enriched in members of the hsa04360 “axon guidance” KEGG pathway. The genes with modified chromatin interactions are marked with stars.



**Figure S11: KEGG pathway with modified chromatin interactions**

The 1193 genes that show modified chromatin interaction in cells of 16p11.2 600kb rearrangement carriers are enriched in members of the hsa04150 “mTOR signaling” KEGG pathway. The genes with modified chromatin interactions are marked with stars.

### Supplementary tables

**Table S1: Catalog of the 16p11.2 220kb BP2-BP3 deletion and duplication carriers used in this study.**

**Table S2: Catalog of 2p15-p16.1 deletion and duplication carriers used in this study.**

**Table S6: Selected BRICKs (Blocks of Regulators In Chromosomal Kontext; FDR≤1%) using MVP as a viewpoint in controls combined replicates.**

**Table S7: Selected BRICKs (Blocks of Regulators In Chromosomal Kontext; FDR≤1%) using KCTD13 as a viewpoint in controls combined replicates.**

**Table S8: Selected BRICKs (Blocks of Regulators In Chromosomal Kontext; FDR≤1%) using ALDOA as a viewpoint in controls combined replicates.**

**Table S9: Selected BRICKs (Blocks of Regulators In Chromosomal Kontext; FDR≤1%) using TBX6 as a viewpoint in controls combined replicates.**

**Table S10: Selected BRICKs (Blocks of Regulators In Chromosomal Kontext; FDR≤1%) using MAPK3 as a viewpoint in controls combined replicates.**

**Table S11: Selected BRICKs (Blocks of Regulators In Chromosomal Kontext; FDR≤1%) using SH2B1 as a viewpoint in controls combined replicates.**

**Table S12: Selected BRICKs (Blocks of Regulators In Chromosomal Kontext; FDR≤1%) using LAT as a viewpoint in controls combined replicates.**

**Table S13: Selected BRICKs (Blocks of Regulators In Chromosomal Kontext; FDR≤1%) using SH2B1 as a viewpoint in 16p11.2 600kb BP4-BP5 deletion combined replicates.**

**Table S14: Selected BRICKs (Blocks of Regulators In Chromosomal Kontext; FDR≤1%) using LAT as a viewpoint in 16p11.2 600kb BP4-BP5 deletion combined replicates.**

**Table S15: Selected BRICKs (Blocks of Regulators In Chromosomal Kontext; FDR≤1%) using MVP as a viewpoint in 16p11.2 600kb BP4-BP5 deletion combined replicates.**

**Table S16: Selected BRICKs (Blocks of Regulators In Chromosomal Kontext; FDR≤1%) using KCTD13 as a viewpoint in 16p11.2 600kb BP4-BP5 deletion combined**

**Table S17: Selected BRICKs (Blocks of Regulators In Chromosomal Kontext; FDR≤1%) using ALDOA as a viewpoint in 16p11.2 600kb BP4-BP5 deletion combined replicates.**

**Table S18: Selected BRICKs (Blocks of Regulators In Chromosomal Kontext; FDR≤1%) using TBX6 as a viewpoint in 16p11.2 600kb BP4-BP5 deletion combined replicates.**

**Table S19: Selected BRICKs (Blocks of Regulators In Chromosomal Kontext; FDR≤1%) using MAPK3 as a viewpoint in 16p11.2 600kb BP4-BP5 deletion combined replicates.**

**Table S20: Selected BRICKs (Blocks of Regulators In Chromosomal Kontext; FDR≤1%) using SH2B1 as a viewpoint in 16p11.2 600kb BP4-BP5 duplication combined replicates.**

**Table S21: Selected BRICKs (Blocks of Regulators In Chromosomal Kontext; FDR≤1%) using LAT as a viewpoint in 16p11.2 600kb BP4-BP5 duplication combined replicates.**

**Table S22: Selected BRICKs (Blocks of Regulators In Chromosomal Kontext; FDR≤1%) using MVP as a viewpoint in 16p11.2 600kb BP4-BP5 duplication combined replicates.**

**Table S23: Selected BRICKs (Blocks of Regulators In Chromosomal Kontext; FDR≤1%) using KCTD13 as a viewpoint in 16p11.2 600kb BP4-BP5 duplication combined.**

**Table S24: Selected BRICKs (Blocks of Regulators In Chromosomal Kontext; FDR≤1%) using ALDOA as a viewpoint in 16p11.2 600kb BP4-BP5 duplication combined.**

**Table S25: Selected BRICKS (Blocks of Regulators In Chromosomal Kontext; FDR≤1%) using TBX6 as a viewpoint in 16p11.2 600kb BP4-BP5 duplication combined.**

**Table S26: Selected BRICKS (Blocks of Regulators In Chromosomal Kontext; FDR≤1%) using MAPK3 as a viewpoint in 16p11.2 600kb BP4-BP5 duplication combined.**

**Table S27: Significantly 4C-modified chromosomal contacts (FDR<1%) in 16p11.2 600kb BP4-BP5 deletion and duplication samples compared to controls, and deletions.**

**Table S31: List of 4C-modified gene, with direction of the change (gain or loss of contact), SFARI-ASD-association, and differential expression data in 16p11.2 600kb del/dup carriers LCLs (DE, with log fold change).**

**Table S35: BRICKS density (number of BRICKS counted) per viewpoint and group (220kb, 600kb and all viewpoints) at 5Mb (Table S35.1; at least 8 BRICKS called per window, all viewpoints considered), 1Mb (Table S35.2; at least 4 BRICKS) and 500kb (Table S35.3; at least 3 BRICKS) sized windows; enrichment of BRICKS genes at chromosomal cytobands calculated by the Chromosome Location tool in Enrichr (<http://amp.pharm.mssm.edu/Enrichr/>) per group (Table S35.4).**

**Table S36: 16p11.2 Consortium Members.**

**Table S37: 2p15 Consortium Members.**

Tables available upon request.

**Table S3: Catalog of lymphoblastoid cell lines used in this study.**

Sample code	Origin	Gender	Age	Copy Number of 16p11.2 600 kb BP4-BP5 CNV
GRAEM-control1	CIG, Lausanne	F	31	2
SOMAG-control2	CIG, Lausanne	F	28	2
ATNAT-deletion1	Lille	F	39	1
VUCAT-deletion2	Lausanne	F	36	1
TABCA-duplication1	Nantes	F	36	3
BONAD-duplication2	Caen	F	28	3

**Table S4: 4C primers sequence. The Illumina adapter tails are highlighted in red (forward primers) and blue (reverse primers).**

Viewpoint	Primer name	Sequence	Primer name	Sequence
SH2B1	SH2B1_4C_seq_F	AATGATACGGCGACC ACCGAACACTCTTTC CCTACACGAGCTCT TCCGATCTGGAGGGG AAGAGTGGTCTTT	SH2B1_4C_seq_R	CAAGCAGAAGACGGC ATACGACCAACAAAA AGTGAGCGACA
		AATGATACGGCGACC ACCGAACACTCTTTC CCTACACGAGCTCT TCCGATCTCTGACCG GAGTCCTGGGTGT		CAAGCAGAAGACGGC ATACGACCTCACAAC CAAGGTCCCTA
LAT	LAT_4C_seq_F	AATGATACGGCGACC ACCGAACACTCTTTC CCTACACGAGCTCT TCCGATCTCTGACCG GAGTCCTGGGTGT	LAT_4C_seq_R	CAAGCAGAAGACGGC ATACGACCTCACAAC CAAGGTCCCTA
		AATGATACGGCGACC ACCGAACACTCTTTC CCTACACGAGCTCT TCCGATCTAGCTGGC TCCAAGGTAGAAA		CAAGCAGAAGACGGC ATACGACTGCCGAGG GAAGAGACTAC
KCTD13	KCTD13_4C_seq_F	AATGATACGGCGACC ACCGAACACTCTTTC CCTACACGAGCTCT TCCGATCTGCCTGAG TGTCTCACATAGC	KCTD13_4C_seq_R	CAAGCAGAAGACGGC ATACGAGCCCTAGCC AAAATCCAGAC
		AATGATACGGCGACC ACCGAACACTCTTTC CCTACACGAGCTCT TCCGATCTAGCCTCA ACTGTCTCTGCTTC		CAAGCAGAAGACGGC ATACGAGGCAGTAGA CAGAGAAAGCACA
ALDOA	ALDOA_4C_seq_F	AATGATACGGCGACC ACCGAACACTCTTTC CCTACACGAGCTCT TCCGATCTAGCCTCA ACTGTCTCTGCTTC	ALDOA_4C_seq_R	CAAGCAGAAGACGGC ATACGAGGCAGTAGA CAGAGAAAGCACA
		AATGATACGGCGACC ACCGAACACTCTTTC CCTACACGAGCTCT TCCGATCTACGCTGC AGATGAGCAGAC		CAAGCAGAAGACGGC ATACGACGAACCTTG GAGGCTTGG
MAPK3	MAPK3_4C_seq_F	AATGATACGGCGACC ACCGAACACTCTTTC CCTACACGAGCTCT TCCGATCTGGTGGGG TTTGAATGAGATG	MAPK3_4C_seq_R	CAAGCAGAAGACGGC ATACGATGACTCAGG AGCACCTACA

**Table S5: Numbers of raw, excluded, mappable and mapped 4C reads for each viewpoint and sample.**

Viewpoint	Condition	RawReads	ExcludedReads *reads that have been filtered, due to the presence of Undigested, Self-ligated and/or bait sequence	MappableReads *reads considered for the rest of the analysis (Bowtie)	MappedReads
SH2B1	control1	2580767	332658	2248109	1695361
	control2	12314376	560126	11754250	7730314
	deletion1	3627369	544009	3083360	1959941
	deletion2	2248115	14608	2233507	1365868
	duplication1	3507290	36124	3471166	2072844
	duplication2	2635340	31301	2604039	1656997
LAT	control1	29272291	5628091	23644200	14614299
	control2	22707491	5935628	16771863	7548802
	deletion1	19565978	7257205	12308773	7360169
	deletion2	38619273	2871782	35747491	19061908
	duplication1	24112972	2773773	21339199	11404819

MVP	duplication2	21233243	1910256	19322987	7743731
	control1	8773647	2248142	6525505	5590973
	control2	30035542	7073964	22961578	17091321
	deletion1	7711725	1335371	6376354	3801555
	deletion2	1035840	101384	934456	596077
	duplication1	10041056	385709	9655347	5240247
	duplication2	2556332	609469	1946863	1452251
	control1	13542833	6729673	6813160	4942638
KCTD13	control2	43023192	21855871	21167321	10990790
	deletion1	31000280	5393490	25606790	4564507
	deletion2	8211454	3169881	5041573	2594193
	duplication1	14536379	3971770	10564609	2190634
	duplication2	8665086	5135489	3529597	1957162
	control1	14343228	4659257	9683971	6829243
	control2	25798974	3287975	22510999	1442397
	deletion1	36683622	500699	36182923	409218
ALDOA	deletion2	8940937	683053	8257884	1492805
	duplication1	16460525	136012	16324513	391857
	duplication2	11888366	2753366	9135000	2769530
	control1	5611067	1231749	4379318	3513837
	control2	28364800	6537550	21827250	14242024
	deletion1	5156129	738117	4418012	2231652
	deletion2	1713771	248208	1465563	745713
	duplication1	4994049	256241	4737808	2531145
TBX6	duplication2	4260311	725640	3534671	2320175
	control1	15617754	7247480	8370274	6564936
	control2	39606162	21480326	18125836	7936212
	deletion1	17460008	8616655	8843353	2996322
	deletion2	7152784	719585	6433199	943065
	duplication1	4491814	1561047	2930767	931515
	duplication2	7013772	3407502	3606270	2135199
	MAPK3				

**Table S28: Top 3 OMIM Disease terms and KEGG Pathways found for the 4C-modified genes in Enrichr (<http://amp.pharm.mssm.edu/Enrichr/>)(Table S28.1); top 10 biological processes (BP), cell component (CC), top 3 KEGG Pathways and top 4 UP-TISSUE annotations for the 4C-modified genes in DAVID (<http://david.abcc.ncifcrf.gov/>) (Table S28.2); top 5 biological processes (BP) and cell component (CC) per group (all 4C-modified genes, 4C-modified contacts with 220kb or 600kb viewpoints only (Table S28.3); all DE 4C-modified genes, DE 4C-modified contacts with 220kb or 600kb viewpoints only (Table S28.4)) in TopGO(<http://gdv.epfl.ch/bs>).**

**Table S28.1**

OMIM Disease						
Term	Overlap	P-value	Adjusted P-value	Z-score	Combined Score	Genes
Usher syndrome	4/11	0.004926844	0.285756974	-1.244750646	1.559191551	MYO7A;GPR98;USH2A;CDH23
Cone-rod dystrophy	4/15	0.012026801	0.348777243	-1.260895613	1.328128877	RPGR;RPGRIP1;RIMS1;PITPNM3
Retinitis pigmentosa	6/51	0.057647588	0.499509822	-0.862502211	0.598686949	RP1; IMPDH1;USH2A;RPGR;MFRP; EYS

KEGG Pathways						
Term	Overlap	P-value	Adjusted P-value	Z-score	Combined Score	Genes
HSA04360 AXON GUIDANCE	20/128	0.000153962	0.021708609	-1.844618407	7.064974017	UNC5D;ROBO2;PLXNB1;PLXNB2;MAPK1;EFNA5;EPHB1;SEMA5B;PTK2;NTN1;EPHA3;EPHA6;PAK7;SLIT2;SLIT3;SRGAP3;ABLIM3;NFATC2;NFATC3;NTNG1
HSA04010 MAPK SIGNALING PATHWAY	26/257	0.006576726	0.230113552	-2.048765557	3.01001027	MAP3K7;MAP2K5;MAPK1;MAPK8;MAP2K4;MAPT;CACNG2;FLNB;MAPK11;RPS6KA1;RPS6KA2;FGF12;FGF23;STK3;RASGRF1;ATF2;CACNA2D3;CACNA2D2;CACNA1E;CACNA1C;CACNA1A;AKT3;PRKCA;NFATC2;NF1;MAPK8IP3
HSA04150 MTOR SIGNALING PATHWAY	9/48	0.003666952	0.230113552	-1.996200463	2.932782559	CAB39;AKT3;MAPK1;STK11;PIK3R1;IGF1;RPS6KA1;RPS6KA2;TSC2

**Table S28.2**

All 4C-modified										
BP										
Category	Term	Count	%	PValue	List Total	Pop Hits	Pop Total	Fold Enrich	Bonferroni	Benjamini
Goterm_Bp_Fat	GO:0048667~ cell morphogenesis involved in neuron differentiation	36	3.20	3.35E-08	818	209	13528	2.85	9.90E-05	9.90E-05
Goterm_Bp_Fat	GO:0000904~ cell morphogenesis	39	3.46	6.44E-08	818	244	13528	2.64	1.91E-04	9.53E-05



Goterm_Bp_Fat	involved in differentiation GO:0007155~	79	7.02	8.72E-08	818	700	13528	1.87	2.58E-04	8.60E-05
Goterm_Bp_Fat	cell adhesion GO:0022610~	79	7.02	9.06E-08	818	701	13528	1.86	2.68E-04	6.70E-05
Goterm_Bp_Fat	biological adhesion GO:0048812~	35	3.11	1.74E-07	818	213	13528	2.72	5.14E-04	1.03E-04
Goterm_Bp_Fat	neuron projection morphogenesis GO:0048666~	46	4.09	4.90E-07	818	339	13528	2.24	0.0014486 09	2.42E-04
Goterm_Bp_Fat	neuron development GO:0007409~	32	2.84	5.26E-07	818	193	13528	2.74	0.0015556 23	2.22E-04
Goterm_Bp_Fat	axonogenesis GO:0031175~	38	3.37	6.46E-07	818	256	13528	2.45	0.0019094 61	2.39E-04
Goterm_Bp_Fat	neuron projection development GO:0030030~	48	4.26	8.48E-07	818	368	13528	2.16	0.0025070 72	2.79E-04
Goterm_Bp_Fat	cell projection organization GO:0048858~	36	3.20	1.72E-06	818	245	13528	2.43	0.0050733 57	5.08E-04
Goterm_Bp_Fat	cell projection morphogenesis									

Cc		Category	Term	Count	%	PValue	List Total	Pop Hits	Pop Total	Fold Enrich	Bonferroni	Benjamini
Goterm_Cc_Fat	GO:0043005~	neuron projection	46	4.09	1.77E-06	803	342	12782	2.14	8.61E-04	8.61E-04	
Goterm_Cc_Fat	GO:0042995~	cell projection	72	6.39	3.06E-05	803	697	12782	1.64	0.0147993 07	7.43E-03	
Goterm_Cc_Fat	GO:0030054~	cell junction	55	4.88	1.56E-04	803	518	12782	1.69	0.0733443 63	0.0250714 45	

KEGG Pathways		Category	Term	Count	%	PValue	List Total	Pop Hits	Pop Total	Fold Enrich	Bonferroni	Benjamini
Kegg_Pathway	hsa04360:	Axon guidance	21	1.87	4.97E-05	299	129	5085	2.77	0.0074767 01	0.0074767 01	
Kegg_Pathway	hsa05414:	Dilated cardio- myopathy	15	1.33	8.20E-04	299	92	5085	2.77	0.1165181 12	0.0600628 28	
Kegg_Pathway	hsa04150:	mTOR signaling pathway	10	0.89	2.85E-03	299	52	5085	3.27	0.3497013 27	0.1336282 37	

Up-Tissue		Category	Term	Count	%	PValue	List Total	Pop Hits	Pop Total	Fold Enrich	Bonferroni	Benjamini
Up_Tissue	Brain	600	53.29	1.32E-20	1063	7789	18201	1.32	4.02E-18	4.02E-18		
Up_Tissue	Fetal brain	90	7.99	4.07E-10	1063	770	18201	2.00	1.24E-07	6.21E-08		
Up_Tissue	Epithelium	212	18.83	6.66E-08	1063	2567	18201	1.41	2.03E-05	6.77E-06		
Up_Tissue	Amygdala	55	4.88	1.25E-03	1063	605	18201	1.56	3.17E-01	9.10E-02		

**Table S28.3**

All 4C-modified		BP		GO.ID	Term	Annotated	Significant	Expected	Classic Fisher	Elim Fisher
GO:0006468	protein phosphorylation	1164	105	69.17	8.70E-06	0.00011				
GO:0007202	activation of phospholipase C activity	58	12	3.45	0.00013	0.00013				
GO:0071322	cellular response to carbohydrate stimulus	32	10	1.9	1.00E-05	0.00017				

GO:0045773	<b>positive regulation of axon extension</b>	22	7	1.31	0.0002	0.0002
GO:0043113	receptor clustering	16	6	0.95	0.00021	0.00021

CC						
GO.ID	Term	Annotated	Significant	Expected	Classic Fisher	Elim Fisher
GO:0005886	plasma membrane	4260	303	246.29	1.60E-05	0.00015
GO:0045202	<b>synapse</b>	462	50	26.71	1.40E-05	0.00015
GO:0030425	<b>dendrite</b>	326	40	18.85	5.60E-06	0.00037
GO:0005578	proteinaceous extracellular matrix	370	38	21.39	0.00043	0.00043
GO:0043025	<b>neuronal cell body</b>	294	31	17	0.00091	0.00091

220 kb						
BP						
GO.ID	Term	Annotated	Significant	Expected	Classic Fisher	Elim Fisher
GO:0007155	cell adhesion	892	51	27.11	1.00E-05	0.00019
GO:0007411	<b>axon guidance</b>	357	26	10.85	3.70E-05	0.00021
GO:0006112	energy reserve metabolic process	163	14	4.95	0.00046	0.00046
GO:0007268	<b>synaptic transmission</b>	613	44	18.63	1.10E-07	0.00049
GO:0060442	branching involved in prostate gland morphogenesis	15	4	0.46	0.00088	0.00088

CC						
GO.ID	Term	Annotated	Significant	Expected	Classic Fisher	Elim Fisher
GO:0045202	<b>synapse</b>	462	31	13.73	2.20E-05	0.0002
GO:0005874	<b>microtubule</b>	319	22	9.48	0.00023	0.00023
GO:0005578	proteinaceous extracellular matrix	370	24	10.99	0.00031	0.00031
GO:0044463	<b>cell projection part</b>	560	40	16.64	2.90E-07	0.00031
GO:0005913	cell-cell adhesion junction	46	7	1.37	0.00038	0.00038

600 kb						
BP						
GO.ID	Term	Annotated	Significant	Expected	Classic Fisher	Elim Fisher
GO:0001952	regulation of cell-matrix adhesion	44	7	1.42	0.00048	0.00048
GO:0045773	<b>positive regulation of axon extension</b>	22	5	0.71	0.00058	0.00058
GO:0007507	heart development	341	23	11.03	0.00075	0.00075
GO:0010529	negative regulation of transposition	7	3	0.23	0.00107	0.00107
GO:0006688	glycosphingolipid biosynthetic process	15	4	0.49	0.00111	0.00111

CC						
GO.ID	Term	Annotated	Significant	Expected	Classic Fisher	Elim Fisher
GO:0005886	plasma membrane	4260	165	133.47	0.00104	0.001
GO:0002102	podosome	9	3	0.28	0.00223	0.0022
GO:0043005	<b>neuron projection</b>	583	36	18.27	8.80E-05	0.0027
GO:0016021	integral to membrane	5530	208	173.26	0.00079	0.0028
GO:0030425	<b>dendrite</b>	326	19	10.21	0.00728	0.0073

**Table  
S28.3**

All 4C-modified								
BP								
GO.ID	Term	Annotated	Significant	Expected	Classic Fisher	Elim Fisher		
GO:0006468	protein phosphorylation	1164	105	69.17	8.70E-06	0.00011		
GO:0007202	activation of phospholipase C activity	58	12	3.45	0.00013	0.00013		
GO:0071322	cellular response to carbohydrate stimulus	32	10	1.9	1.00E-05	0.00017		
GO:0045773	<b>positive regulation of axon extension</b>	22	7	1.31	0.0002	0.0002		
GO:0043113	receptor clustering	16	6	0.95	0.00021	0.00021		

CC								
GO.ID	Term	Annotated	Significant	Expected	Classic Fisher	Elim Fisher		
GO:0005886	plasma membrane	4260	303	246.29	1.60E-05	0.00015		
GO:0045202	<b>synapse</b>	462	50	26.71	1.40E-05	0.00015		
GO:0030425	<b>dendrite</b>	326	40	18.85	5.60E-06	0.00037		
GO:0005578	proteinaceous extracellular matrix	370	38	21.39	0.00043	0.00043		
GO:0043025	<b>neuronal cell body</b>	294	31	17	0.00091	0.00091		

220 kb								
BP								
GO.ID	Term	Annotated	Significant	Expected	Classic Fisher	Elim Fisher		
GO:0007155	cell adhesion	892	51	27.11	1.00E-05	0.00019		
GO:0007411	<b>axon guidance</b>	357	26	10.85	3.70E-05	0.00021		
GO:0006112	energy reserve metabolic process	163	14	4.95	0.00046	0.00046		
GO:0007268	<b>synaptic transmission</b>	613	44	18.63	1.10E-07	0.00049		
GO:0060442	branching involved in prostate gland morphogenesis	15	4	0.46	0.00088	0.00088		

CC								
GO.ID	Term	Annotated	Significant	Expected	Classic Fisher	Elim Fisher		
GO:0045202	<b>synapse</b>	462	31	13.73	2.20E-05	0.0002		
GO:0005874	<b>microtubule</b>	319	22	9.48	0.00023	0.00023		
GO:0005578	proteinaceous extracellular matrix	370	24	10.99	0.00031	0.00031		
GO:0044463	<b>cell projection part</b>	560	40	16.64	2.90E-07	0.00031		
GO:0005913	cell-cell adherens junction	46	7	1.37	0.00038	0.00038		

600 kb								
BP								
GO.ID	Term	Annotated	Significant	Expected	Classic Fisher	Elim Fisher		
GO:0001952	regulation of cell-matrix adhesion	44	7	1.42	0.00048	0.00048		
GO:0045773	<b>positive regulation of axon extension</b>	22	5	0.71	0.00058	0.00058		
GO:0007507	heart development	341	23	11.03	0.00075	0.00075		
GO:0010529	negative regulation of transposition	7	3	0.23	0.00107	0.00107		
GO:0006688	glycosphingolipid biosynthetic process	15	4	0.49	0.00111	0.00111		

CC								
GO.ID	Term	Annotated	Significant	Expected	Classic Fisher	Elim Fisher		

GO:0005886	plasma membrane	4260	165	133.47	0.00104	0.001
GO:0002102	podosome	9	3	0.28	0.00223	0.0022
GO:0043005	<b>neuron projection</b>	583	36	18.27	8.80E-05	0.0027
GO:0016021	integral to membrane	5530	208	173.26	0.00079	0.0028
GO:0030425	<b>dendrite</b>	326	19	10.21	0.00728	0.0073

**Table  
S28.4**

BP						
GO.ID	Term	Annotated	Significant	Expected	Classic Fisher	Elim Fisher
GO:0010604	positive regulation of macromolecule metabolic process	1644	22	11.24	0.0015	0.0015
GO:0043518	negative regulation of DNA damage response, signal transduction by p53 class mediator	10	2	0.07	0.002	0.002
GO:0045070	positive regulation of viral genome replication	10	2	0.07	0.002	0.002
GO:0050650	chondroitin sulfate proteoglycan biosynthetic process	11	2	0.08	0.0024	0.0024
GO:0034504	<b>protein localization to nucleus</b>	228	6	1.56	0.0048	0.0048

CC						
GO.ID	Term	Annotated	Significant	Expected	Classic Fisher	Elim Fisher
GO:0044424	intracellular part	12113	91	76.88	0.0015	0.0039
GO:0005730	nucleolus	527	9	3.34	0.0064	0.0064
GO:0043231	intracellular membrane-bounded organelle	9403	74	59.68	0.0037	0.0124
GO:0033596	<b>TSC1-TSC2 complex</b>	2	1	0.01	0.0127	0.0127
GO:0043227	membrane-bounded organelle	9413	74	59.74	0.0038	0.0128

220 kb						
BP						
GO.ID	Term	Annotated	Significant	Expected	Classic Fisher	Elim Fisher
GO:0043518	negative regulation of DNA damage response, signal transduction by p53 class mediator	10	2	0.03	0.0005	0.0005
GO:0070647	protein modification by small protein conjugation or removal	510	7	1.73	0.0016	0.0016
GO:0002316	follicular B cell differentiation	1	1	0	0.0034	0.0034
GO:0006843	mitochondrial citrate transport	1	1	0	0.0034	0.0034
GO:0035280	miRNA loading onto RISC involved in gene silencing by miRNA	1	1	0	0.0034	0.0034

CC						
GO.ID	Term	Annotated	Significant	Expected	Classic Fisher	Elim Fisher
GO:0043229	intracellular organelle	10398	42	32.4	0.0045	0.0075
GO:0030173	<b>integral to Golgi membrane</b>	45	2	0.14	0.0087	0.0087
GO:0000138	<b>Golgi trans cisterna</b>	4	1	0.01	0.0124	0.0124
GO:0032437	<b>cuticular plate</b>	4	1	0.01	0.0124	0.0124
GO:0005634	nucleus	5740	26	17.88	0.0154	0.0154

600 kb						
BP						
GO.ID	Term	Annotated	Significant	Expected	Classic Fisher	Elim Fisher
GO:0043491	protein kinase B signaling cascade	66	3	0.26	0.0022	0.0022

GO:0009887	organ morphogenesis	753	9	2.94	0.0024	0.0024
GO:0010604	positive regulation of macromolecule metabolic process	1644	14	6.42	0.0037	0.0037
GO:0000073	<b>spindle pole body separation</b>	1	1	0	0.0039	0.0039
GO:0006175	dATP biosynthetic process	1	1	0	0.0039	0.0039

CC						
GO.ID	Term	Annotated	Significant	Expected	Classic Fisher	Elim Fisher
GO:0044451	nucleoplasm part	793	8	2.84	0.007	0.007
GO:0033596	<b>TSC1-TSC2 complex</b>	2	1	0.01	0.0071	0.0071
GO:0008287	protein serine/threonine phosphatase complex	51	2	0.18	0.0143	0.0143
GO:0001940	male pronucleus	5	1	0.02	0.0178	0.0178
GO:0005797	<b>Golgi medial cisterna</b>	5	1	0.02	0.0178	0.0178

**Table S29: List of 4C-modified genes which are cilia-related, SFARI-ASD-associated, differentially expressed (DE), and both ASD-associated and DE.**

SFARI-ASD 4C-modified genes		SFARI-ASD (Categories S,1,2) 4C-modified genes		DE 4C-modified genes	
Ensembl ID	Gene	Ensembl ID	Gene	Ensembl ID	Gene
ENSG00000026508	CD44	ENSG00000167522	ANKRD11	ENSG00000004961	HCCS
ENSG00000036257	CUL3	ENSG00000151067	CACNA1C	ENSG00000005194	CIAPIN1
ENSG00000040731	CDH10	ENSG00000174469	CNTNAP2	ENSG00000012983	MAP4K5
ENSG00000041515	MYO16	ENSG00000198947	DMD	ENSG00000020577	SAMD4A
ENSG00000071242	RPS6KA2	ENSG00000181090	EHMT1	ENSG00000026508	CD44
ENSG00000075884	ARHGAP15	ENSG00000102081	FMR1	ENSG00000048471	SNX29
ENSG00000078328	RBFOX1	ENSG00000177807	KCNJ10	ENSG00000054282	SDCCAG8
ENSG00000080224	EPHA6	ENSG00000196712	NF1	ENSG00000058056	USP13
ENSG00000081237	PTPRC	ENSG00000162631	NTNG1	ENSG00000063438	AHRR
ENSG00000091831	ESR1	ENSG00000103197	TSC2	ENSG00000065000	AP3D1
ENSG00000100030	MAPK1	ENSG00000132549	VPS13B	ENSG00000072736	NFATC3
ENSG00000100150	DEPDC5	ENSG00000145362	ANK2	ENSG00000074695	LMAN1
ENSG00000100393	EP300	ENSG00000150086	GRIN2B	ENSG00000075234	TTC38
ENSG00000102081	FMR1	ENSG00000157445	CACNA2D3	ENSG00000080345	RIF1
ENSG00000103197	TSC2	ENSG00000144619	CNTN4	ENSG00000090863	GLG1
ENSG00000103528	SYT17	ENSG00000169862	CTNND2	ENSG00000092201	SUPT16H
ENSG00000109906	ZBTB16	ENSG00000036257	CUL3	ENSG00000100075	SLC25A1
ENSG00000110076	NRXN2	ENSG00000171587	DSCAM	ENSG00000100226	GTPBP1
ENSG00000112232	KHDRBS2	ENSG00000166206	GABRB3	ENSG00000100324	TAB1
ENSG00000112679	DUSP22	ENSG00000155974	GRIP1	ENSG00000100393	EP300
ENSG00000118257	NRP2	ENSG00000179915	NRXN1	ENSG00000100664	EIF5
ENSG00000119715	ESRRB	ENSG00000189056	RELN	ENSG00000101347	SAMHD1
ENSG00000122584	NXPH1			ENSG00000102554	KLF5
ENSG00000125780	TGM3			ENSG00000102931	ARL2BP

**De Rubeis et al., lossifov et al. ASD 4C-modified genes**

ENSG00000128512	DOCK4			ENSG00000103018	CYB5B
ENSG00000128573	FOXP2			ENSG00000103197	TSC2
ENSG00000130226	DPP6			ENSG00000103528	SYT17
ENSG00000130508	PXDN			ENSG00000103550	C16orf88
ENSG00000132549	VPS13B			ENSG00000103657	HERC1
ENSG00000134780	DAGLA			ENSG00000104213	PDGFRL
ENSG00000135127	CCDC64			ENSG00000104731	KLHDC4
ENSG00000135905	DOCK10			ENSG00000109189	USP46
ENSG00000136854	STXBP1			ENSG00000110497	AMBRA1
ENSG00000139734	DIAPH3			ENSG00000111615	KRR1
ENSG00000139915	MDGA2			ENSG00000115306	SPTBN1
ENSG00000140488	CELF6			ENSG00000115677	HDLBP
ENSG00000140538	NTRK3			ENSG00000115935	WIPF1
ENSG00000142192	APP			ENSG00000115966	ATF2
ENSG00000144278	GALNT13			ENSG00000120659	TNFSF11
ENSG00000144619	CNTN4			ENSG00000120686	UFM1
ENSG00000145362	ANK2			ENSG00000121281	ADCY7
ENSG00000146469	VIP			ENSG00000121964	GTDC1
ENSG00000146555	SDK1			ENSG00000123091	RNF11
ENSG00000147010	SH3KBP1			ENSG00000124067	SLC12A4
ENSG00000148219	ASTN2			ENSG00000126264	HCST
ENSG00000149972	CNTN5			ENSG00000129292	PHF20L1
ENSG00000150086	GRIN2B			ENSG00000130508	PXDN
ENSG00000151067	CACNA1C			ENSG00000130517	PGPEP1
ENSG00000151150	ANK3			ENSG00000130723	BAT2L1
ENSG00000151577	DRD3			ENSG00000131153	GINS2
ENSG00000152495	CAMK4			ENSG00000132549	VPS13B
ENSG00000153575	TUBGCP5			ENSG00000133997	MED6
ENSG00000154678	PDE1C			ENSG00000134851	TMEM165
ENSG00000155052	CNTNAP5			ENSG00000135250	SRPK2
ENSG00000155966	AFF2			ENSG00000135679	MDM2
ENSG00000155974	GRIP1			ENSG00000135905	DOCK10
ENSG00000156110	ADK			ENSG00000136273	HUS1
ENSG00000156113	KCNMA1			ENSG00000137764	MAP2K5
ENSG00000157423	HYDIN			ENSG00000138185	ENTPD1
ENSG00000157637	SLC38A10			ENSG00000139546	TARBP2
ENSG00000158321	AUTS2			ENSG00000139574	NPFF
ENSG00000162599	NFIA			ENSG00000140332	TLE3
ENSG00000162631	NTNG1			ENSG00000140993	TIGD7
ENSG00000165973	NELL1			ENSG00000141034	C17orf39
ENSG00000166206	GABRB3			ENSG00000144339	TMEFF2
ENSG00000166501	PRKCB			ENSG00000144451	SPAG16
ENSG00000166736	HTR3A			ENSG00000144567	FAM134A
ENSG00000167522	ANKRD11			ENSG00000144857	BOC
ENSG00000169306	IL1RAPL1			ENSG00000145819	ARHGAP26
ENSG00000171444	MCC			ENSG00000146592	CREB5
ENSG00000172264	MACROD2			ENSG00000147050	KDM6A
ENSG00000173406	DAB1			ENSG00000149970	CNKSR2
ENSG00000174469	CNTNAP2			ENSG00000152056	AP1S3

Ensembl ID	Gene
ENSG00000145362	ANK2
ENSG00000167522	ANKRD11
ENSG00000157445	CACNA2D3
ENSG00000036257	CUL3
ENSG00000171587	DSCAM
ENSG00000166206	GABRB3
ENSG00000150086	GRIN2B
ENSG00000189056	RELN
ENSG00000079841	RIMS1
ENSG00000038382	TRIO
ENSG00000163625	WDFY3

Ciliary 4C-modified genes	
Ensembl ID	Gene
ENSG00000138031	ADCY3
ENSG00000151150	ANK3
ENSG00000197826	C4orf22
ENSG00000213204	C6orf165
ENSG00000077549	CAPZB
ENSG00000103021	CCDC113
ENSG00000096401	CDC5L
ENSG00000107736	CDH23
ENSG00000197102	DYNC1H1
ENSG00000131558	EXOC4
ENSG00000164199	GPR98
ENSG00000157423	HYDIN
ENSG00000137474	MYO7A
ENSG00000114904	NEK4
ENSG00000101004	NINL
ENSG00000179915	NRXN1
ENSG00000102900	NUP93
ENSG00000145730	PAM
ENSG00000185345	PARK2
ENSG00000158683	PKD1L1
ENSG00000170927	PKHD1
ENSG00000158528	PPP1R9A
ENSG00000112210	RAB23
ENSG00000104237	RP1
ENSG00000156313	RPGR
ENSG00000092200	RPGRIIP1
ENSG00000103494	RPGRIIP1L
ENSG00000054282	SDCCAG8
ENSG00000168385	SEPT2
ENSG00000144451	SPAG16
ENSG00000152582	SPEF2
ENSG00000211455	STK38L
ENSG00000107882	SUFU
ENSG00000196628	TCF4

ENSG00000175497	DPP10	ENSG00000100815	TRIP11	ENSG00000153575	TUBGCP5
ENSG00000176771	NCKAP5	ENSG00000128881	TTBK2	ENSG00000153944	MSI2
ENSG00000177807	KCNJ10	ENSG00000153575	TUBGCP5	ENSG00000154001	PPP2R5E
ENSG00000179915	NRXN1	ENSG00000168038	ULK4	ENSG00000154122	ANKH
ENSG00000181090	EHMT1	ENSG00000042781	USH2A	ENSG00000154265	ABCA5
ENSG00000183098	GPC6	ENSG00000158023	WDR66	ENSG00000155100	OTUD6B
ENSG00000183117	CSMD1			ENSG00000156030	C14orf43
ENSG00000183230	CTNNA3			ENSG00000156110	ADK
ENSG00000184156	KCNQ3			ENSG00000156976	EIF4A2
ENSG00000184226	PCDH9			ENSG00000157637	SLC38A10
ENSG00000184254	ALDH1A3			ENSG00000157800	SLC37A3
ENSG00000184304	PRKD1			ENSG00000159140	SON
ENSG00000184347	SLIT3			ENSG00000159579	RSPRY1
ENSG00000185008	ROBO2			ENSG00000162065	TBC1D24
ENSG00000185345	PARK2			ENSG00000162613	FUBP1
ENSG00000186094	AGBL4			ENSG00000164330	EBF1
ENSG00000187957	DNER			ENSG00000164880	INTS1
ENSG00000188641	DPYD			ENSG00000164970	C9orf25
ENSG00000189056	RELN			ENSG00000165572	KBTBD6
ENSG00000189283	FHIT			ENSG00000165689	SDCCAG3
ENSG00000196628	TCF4			ENSG00000167522	ANKRD11
ENSG00000196712	NF1			ENSG00000167978	SRRM2
ENSG00000196876	SCN8A			ENSG00000168386	FILIP1L
ENSG00000197157	SND1			ENSG00000168675	C18orf1
ENSG00000197381	ADARB1			ENSG00000168763	CNNM3
ENSG00000198947	DMD			ENSG00000169398	PTK2
ENSG00000221866	PLXNA4			ENSG00000169826	CSGALNACT2
ENSG00000249108	CLTCL1			ENSG00000171310	CHST11
ENSG00000152910	CNTNAP4			ENSG00000173442	EHBP1L1
ENSG00000176884	GRIN1			ENSG00000173451	THAP2
				ENSG00000173482	PTPRM
				ENSG00000174197	MGA
				ENSG00000179364	PACS2
				ENSG00000182670	TTC3
				ENSG00000183513	C2orf64
				ENSG00000184575	XPOT
				ENSG00000184743	ATL3
				ENSG00000185722	ANKFY1
				ENSG00000186566	GPATCH8
				ENSG00000187742	SECISBP2
				ENSG00000189079	ARID2
				ENSG00000196284	SUPT3H
				ENSG00000196628	TCF4
				ENSG00000197102	DYNC1H1
				ENSG00000197782	ZNF780A
				ENSG00000197943	PLCG2
				ENSG00000198964	SGMS1
				ENSG00000204569	PPP1R10
				ENSG00000240771	ARHGEF25

ENSG00000244509	APOBEC3C
ENSG00000254726	MEX3A
ENSG0000003509	C2orf56

SFARI-ASD DE 4C-modified genes							
Ensembl ID	Gene	Head size	Ophthalmological	Epilepsy	Urogenital	Psychiatric	BMI/growth
ENSG00000135905	DOCK10						
ENSG00000132549	VPS13B	microcephaly, syndromic	retinopathy/visual loss				obesity
ENSG00000196628	TCF4	microcephaly, syndromic	corneal dystrophy	epileptic encephalopathy		schizophrenia	
ENSG00000103528	SYT17						
ENSG00000157637	SLC38A10						
ENSG00000130508	PXDN		ocular anomalies		Hydrocele		
ENSG00000026508	CD44						
ENSG00000153575	TUBGCP5	microcephaly, syndromic					
ENSG00000103197	TSC2			seizures/epilepsy	polycystic kidney disease		
ENSG00000100393	EP300	microcephaly, syndromic					growth abnormalities skeletal malformations
ENSG00000156110	ADK	macrocephaly, syndromic					
ENSG00000167522	ANKRD11			EEG abnormalities	cryptorchidism		growth abnormalities skeletal malformations

**Table S30: Enrichment of ASD associated genes in the 4C-modified genes (Table S30.1), of "high confidence" ASD target genes identified in whole exome sequencing studies (De Rubeis et al., 2014; Iossifov et al., 2014)(TableS30.2), of ciliary genes (van Dam, 2013)(Table S30.3), and of differentially expressed (DE; FDR5%) genes in the 4C-modified genes (Table S30.4), and of ASD genes in the DE and 4C-modified genes (Table S30.5)(Fisher's exact test)**

**Table S30.1**

Phenotype	Listed genes	Unique 4C-modified genes (all viewpoints)	4C-modified ASD genes (all viewpoints)	p-value	OR	5% CI	95% CI	Reference list
Autism spectrum disorder (ASD)	604	1193	103	8.41E-24	3.6	2.9	4.5	SFARI Autism Research (March 2014)
Autism spectrum disorder (ASD)	76	1193	22	1.48E-10	6.8	3.9	11.4	SFARI-Categories S, 1, 2



**Table S30.2**

Phenotype	Listed genes	Unique 4C-modified genes (all viewpoints)	Interacting genes - all viewpoint	p-value	OR	5% CI	95% CI	Reference list
Autism spectrum disorder (ASD)	50		11	9.93E-05	4.7	2.2	9.3	De Rubeis, 2014; Iossifov, 2014
Autism spectrum disorder (ASD)	33	1193	7	2.30E-03	4.4	1.6	10.5	De Rubeis, 2014
Autism spectrum disorder (ASD)	27		6	3.67E-03	4.7	1.6	12.1	Iossifov, 2014

**Table S30.3**

Ciliary genes	Unique 4C-modified genes - all viewpoints	4C-modified ciliary genes - all viewpoints	p-value	OR	5% CI	95% CI	Reference list
493	1193	40	3.02E-02	1.5	1.0	2.0	van Dam, 2013

**Table S30.4**

Unique 4C-modified genes - all viewpoints	Unique 5%FDR DE genes	4C-modified and DE genes - all viewpoints	p-value	OR	5% CI	95% CI	DE reference list	Reference list
1193 (665 expressed in array)	2209	125	2.40E-03	1.4	1.1	1.7	Migliavacca et al., 2015	SFARI Autism Research (March 2014)

**Table S30.5**

Phenotype	Listed genes	Expressed in array	4C-modified and DE genes - all viewpoints	4C-modified ASD genes - all viewpoints	p-value	OR	5% CI	95% CI
Autism spectrum disorder (ASD)	604	323	125	12	1.51E-05	5.0	2.5	9.2

**Table S32: Enrichment of protein-protein interactions of BRICKs encompassed genes per viewpoint (computed using STRING (<http://string-db.org/>; default parameters, medium confidence score: 0.4)**

	SH2B1	LAT	MVP	KCTD3	ALDOA	TBX6	MAPK3
<b>P-value</b>	2.64E-05	5.57E-05	3.48E-04	2.93E-03	1.22E-04	3.96E-03	6.05E-04
<b>Interactions observed</b>	47	237	155	54	88	117	101
<b>Interactions expected</b>	2.42E+01	1.82E+02	1.16E+02	3.59E+01	5.76E+01	9.02E+01	7.15E+01
<b>Proteins</b>	187	444	323	205	273	279	243

**Table S33: Top 15 biological processes (BP) and cell component (CC) per group of viewpoints (Table S33.1) and per viewpoint (Table S33.2) according to TopGO (<http://gdv.epfl.ch/bs>)**

**Table S33.1**

220 kb		BP					
GO.ID	Term	Annotated	Significant	Expected	classicFisher	elimFisher	
GO:0050427	3'-phosphoadenosine 5'-phosphosulfate metabolic process	15	4	0.41	0.00061	0.00061	
<b>GO:0071109</b>	<b>superior temporal gyrus development</b>	<b>2</b>	<b>2</b>	<b>0.06</b>	<b>0.00076</b>	<b>0.00076</b>	
GO:0051923	sulfation	16	4	0.44	0.00079	0.00079	
GO:0006584	catecholamine metabolic process	43	6	1.18	0.00108	0.00108	
GO:0016559	peroxisome fission	3	2	0.08	0.00223	0.00223	
GO:0010792	DNA double-strand break processing involved in repair via single-strand annealing	4	2	0.11	0.00437	0.00437	
GO:0042373	vitamin K metabolic process	4	2	0.11	0.00437	0.00437	
GO:0006733	oxidoreduction coenzyme metabolic process	57	6	1.57	0.00466	0.00466	
GO:0009812	flavonoid metabolic process	13	3	0.36	0.00482	0.00482	
<b>GO:0048041</b>	<b>focal adhesion assembly</b>	<b>27</b>	<b>4</b>	<b>0.74</b>	<b>0.00602</b>	<b>0.00602</b>	
<b>GO:2000344</b>	<b>positive regulation of acrosome reaction</b>	<b>5</b>	<b>2</b>	<b>0.14</b>	<b>0.00716</b>	<b>0.00716</b>	

GO:0007596	blood coagulation skeletal muscle tissue	507	24	13.96	0.00718	0.00718
GO:0007519	development	149	10	4.1	0.00825	0.00825
GO:0046700	heterocycle catabolic process	872	38	24.01	0.00345	0.00873
<b>GO:0016079</b>	<b>synaptic vesicle exocytosis</b>	<b>30</b>	<b>4</b>	<b>0.83</b>	<b>0.00881</b>	<b>0.00881</b>
<b>CC</b>						
GO.ID	Term	Annotated	Significant	Expected	classicFisher	elimFisher
<b>GO:0034704</b>	<b>calcium channel complex</b>	<b>31</b>	<b>5</b>	<b>0.83</b>	<b>0.0013</b>	<b>0.0013</b>
<b>GO:0031463</b>	<b>Cul3-RING ubiquitin ligase complex</b>	<b>10</b>	<b>3</b>	<b>0.27</b>	<b>0.002</b>	<b>0.002</b>
GO:0033557	Slx1-Slx4 complex	3	2	0.08	0.0021	0.0021
GO:0005947	mitochondrial alpha-ketoglutarate dehydrogenase complex	4	2	0.11	0.0041	0.0041
GO:0001772	immunological synapse	15	3	0.4	0.0068	0.0068
GO:0005923	tight junction	86	7	2.3	0.0082	0.0082
GO:0005737	cytoplasm	8838	263	236.62	0.0073	0.0092
GO:0031082	BLOC complex	7	2	0.19	0.0137	0.0137
GO:0031083	BLOC-1 complex	7	2	0.19	0.0137	0.0137
GO:0044449	contractile fiber part	128	8	3.43	0.0219	0.0219
<b>GO:0044456</b>	<b>synapse part</b>	<b>342</b>	<b>16</b>	<b>9.16</b>	<b>0.0224</b>	<b>0.0224</b>
<b>GO:0005891</b>	<b>voltage-gated calcium channel complex</b>	<b>23</b>	<b>3</b>	<b>0.62</b>	<b>0.0227</b>	<b>0.0227</b>
<b>GO:0005874</b>	<b>microtubule</b>	<b>319</b>	<b>15</b>	<b>8.54</b>	<b>0.0254</b>	<b>0.0254</b>
GO:0002095	caveolar macromolecular signaling complex	1	1	0.03	0.0268	0.0268
GO:0008352	katanin complex	1	1	0.03	0.0268	0.0268
<b>600 kb</b>						
<b>BP</b>						
GO.ID	Term	Annotated	Significant	Expected	classicFisher	elimFisher
<b>GO:0045747</b>	<b>positive regulation of Notch signaling pathway</b>	<b>8</b>	<b>4</b>	<b>0.36</b>	<b>0.00026</b>	<b>0.00026</b>
<b>GO:0046854</b>	<b>phosphatidylinositol phosphorylation</b>	<b>15</b>	<b>5</b>	<b>0.68</b>	<b>0.00039</b>	<b>0.00039</b>
GO:0050427	3'-phosphoadenosine 5'-phosphosulfate metabolic process	15	5	0.68	0.00039	0.00039
<b>GO:0007411</b>	<b>axon guidance</b>	<b>357</b>	<b>31</b>	<b>16.22</b>	<b>0.00043</b>	<b>0.00043</b>
GO:0043149	stress fiber assembly	32	7	1.45	0.00048	0.00048
<b>GO:0048041</b>	<b>focal adhesion assembly</b>	<b>27</b>	<b>6</b>	<b>1.23</b>	<b>0.00112</b>	<b>0.00112</b>
GO:0007169	transmembrane receptor protein tyrosine kinase signaling pathway	585	43	26.58	0.0013	0.0013
GO:0048096	chromatin-mediated maintenance of transcription	6	3	0.27	0.00168	0.00168
GO:0016074	snoRNA metabolic process	2	2	0.09	0.00206	0.00206
GO:0016573	histone acetylation	101	12	4.59	0.00206	0.00206
<b>GO:0000132</b>	<b>establishment of mitotic spindle orientation</b>	<b>13</b>	<b>4</b>	<b>0.59</b>	<b>0.00217</b>	<b>0.00217</b>
GO:0007520	myoblast fusion	14	4	0.64	0.00293	0.00293
GO:0050919	negative chemotaxis	14	4	0.64	0.00293	0.00293
GO:0033554	cellular response to stress	1031	65	46.84	0.00437	0.00437
GO:0000738	DNA catabolic process, exonucleolytic	8	3	0.36	0.0044	0.0044
<b>CC</b>						
GO.ID	Term	Annotated	Significant	Expected	classicFisher	elimFisher
<b>GO:0008305</b>	<b>integrin complex</b>	<b>29</b>	<b>6</b>	<b>1.3</b>	<b>0.00155</b>	<b>0.0016</b>

GO:0005887	integral to plasma membrane	1281	81	57.36	0.00096	0.005
GO:0033557	Slx1-Slx4 complex	3	2	0.13	0.00583	0.0058
GO:0005829	cytosol mitochondrial respiratory chain complex IV	2279	126	102.04	0.0064	0.0064
GO:0005751	respiratory chain complex IV	4	2	0.18	0.01131	0.0113
GO:0045277	respiratory chain complex IV	4	2	0.18	0.01131	0.0113
<b>GO:0030425</b>	<b>dendrite</b>	<b>326</b>	<b>24</b>	<b>14.6</b>	<b>0.01194</b>	<b>0.0119</b>
GO:0005912	adherens junction phosphatidylinositol 3-kinase complex	178	15	7.97	0.01409	0.0141
<b>GO:0005942</b>	<b>synapse part</b>	<b>13</b>	<b>3</b>	<b>0.58</b>	<b>0.01827</b>	<b>0.0183</b>
GO:0005654	nucleoplasm	1505	84	67.39	0.0201	0.0201
<b>GO:0044456</b>	<b>DNA-directed RNA polymerase II, core complex</b>	<b>342</b>	<b>24</b>	<b>15.31</b>	<b>0.02022</b>	<b>0.0202</b>
GO:0005665	BAF-type complex	14	3	0.63	0.02249	0.0225
GO:0090544	BAF-type complex	14	3	0.63	0.02249	0.0225
<b>GO:0030496</b>	<b>midbody beta-catenin destruction complex</b>	<b>65</b>	<b>7</b>	<b>2.91</b>	<b>0.02575</b>	<b>0.0257</b>
GO:0030877	complex	6	2	0.27	0.02663	0.0266

**Table S33.2**

SH2B1							
BP							
GO.ID	Term	Annotated	Significant	Expected	classicFisher	elimFisher	
GO:0009812	flavonoid metabolic process	13	3	0.12	0.00021	0.00021	
GO:0010792	DNA double-strand break processing involved in repair via single-strand annealing	4	2	0.04	0.00051	0.00051	
GO:0006729	tetrahydrobiopterin biosynthetic process	6	2	0.06	0.00126	0.00126	
GO:0019369	arachidonic acid metabolic process	12	2	0.11	0.00534	0.00534	
GO:0009060	aerobic respiration	39	3	0.36	0.00565	0.00565	
CC							
GO.ID	Term	Annotated	Significant	Expected	classicFisher	elimFisher	
GO:0033557	Slx1-Slx4 complex	3	2	0.03	0.00024	0.00024	
GO:0001772	immunological synapse	15	3	0.14	0.00031	0.00031	
GO:0005852	eukaryotic translation initiation factor 3 complex	15	2	0.14	0.00792	0.00792	
GO:0031673	H zone	1	1	0.01	0.00906	0.00906	
GO:0045335	phagocytic vesicle	18	2	0.16	0.01134	0.01134	
LAT							
BP							
GO.ID	Term	Annotated	Significant	Expected	classicFisher	elimFisher	
GO:0050427	3'-phosphoadenosine 5'-phosphosulfate metabolic process	15	4	0.34	0.00028	0.00028	
GO:0051923	sulfation	16	4	0.36	0.00036	0.00036	
<b>GO:0071109</b>	<b>superior temporal gyrus development</b>	<b>2</b>	<b>2</b>	<b>0.04</b>	<b>0.0005</b>	<b>0.0005</b>	
GO:0016559	peroxisome fission	3	2	0.07	0.00148	0.00148	
GO:0007519	skeletal muscle tissue development	149	10	3.34	0.00194	0.00194	
CC							
GO.ID	Term	Annotated	Significant	Expected	classicFisher	elimFisher	
GO:0034704	calcium channel complex	31	5	0.68	0.00053	0.00053	

GO:0031463	Cul3-RING ubiquitin ligase complex	10	3	0.22	0.00113	0.00113
GO:0033557	Slx1-Slx4 complex	3	2	0.07	0.00143	0.00143
GO:0005947	mitochondrial alpha-ketoglutarate dehydrogenase complex	4	2	0.09	0.00281	0.00281
GO:0001772	immunological synapse	15	3	0.33	0.00394	0.00394
<b>MVP</b>						
<b>BP</b>						
GO.ID	Term	Annotated	Significant	Expected	classicFisher	elimFisher
GO:0006584	catecholamine metabolic process	43	5	0.67	0.00054	0.00054
GO:0009812	flavonoid metabolic process	13	3	0.2	0.00097	0.00097
GO:0010792	DNA double-strand break processing involved in repair via single-strand annealing	4	2	0.06	0.00144	0.00144
GO:0042098	T cell proliferation	110	7	1.73	0.00173	0.00173
<b>GO:0007163</b>	<b>establishment or maintenance of cell polarity</b>	<b>89</b>	<b>6</b>	<b>1.4</b>	<b>0.00275</b>	<b>0.00275</b>
<b>CC</b>						
GO.ID	Term	Annotated	Significant	Expected	classicFisher	elimFisher
GO:0033557	Slx1-Slx4 complex	3	2	0.05	0.00077	0.00077
GO:0001772	immunological synapse	15	3	0.24	0.00163	0.00163
GO:0001891	phagocytic cup	23	3	0.37	0.00576	0.00576
GO:0030141	secretory granule	243	10	3.91	0.00616	0.00616
<b>GO:0030672</b>	<b>synaptic vesicle membrane</b>	<b>47</b>	<b>4</b>	<b>0.76</b>	<b>0.00681</b>	<b>0.00681</b>
<b>KCTD13</b>						
<b>BP</b>						
GO.ID	Term	Annotated	Significant	Expected	classicFisher	elimFisher
GO:0009812	flavonoid metabolic process	13	3	0.13	0.00025	0.00025
GO:0010792	DNA double-strand break processing involved in repair via single-strand annealing	4	2	0.04	0.00058	0.00058
GO:0006584	catecholamine metabolic process	43	4	0.43	0.00084	0.00084
<b>GO:0045945</b>	<b>positive regulation of transcription from RNA polymerase III promoter</b>	<b>6</b>	<b>2</b>	<b>0.06</b>	<b>0.00142</b>	<b>0.00142</b>
GO:0002862	negative regulation of inflammatory response to antigenic stimulus	7	2	0.07	0.00198	0.00198
<b>CC</b>						
GO.ID	Term	Annotated	Significant	Expected	classicFisher	elimFisher
GO:0033557	Slx1-Slx4 complex	3	2	0.03	0.00027	0.00027
GO:0001772	immunological synapse	15	3	0.14	0.00035	0.00035
GO:0070603	SWI/SNF-type complex	58	4	0.55	0.0022	0.0022
GO:0005833	hemoglobin complex	13	2	0.12	0.00648	0.00648
GO:0005665	DNA-directed RNA polymerase II, core complex	14	2	0.13	0.00752	0.00752
<b>ALDOA</b>						
<b>BP</b>						
GO.ID	Term	Annotated	Significant	Expected	classicFisher	elimFisher
GO:0016074	snoRNA metabolic process	2	2	0.03	0.00017	0.00017
GO:0006584	catecholamine metabolic process	43	5	0.57	0.00025	0.00025
GO:0051561	elevation of mitochondrial calcium ion concentration	3	2	0.04	0.00052	0.00052

GO:0009812	flavonoid metabolic process	13	3	0.17	0.00059	0.00059
<b>GO:0046854</b>	<b>phosphatidylinositol phosphorylation</b>	<b>15</b>	<b>3</b>	<b>0.2</b>	<b>0.00092</b>	<b>0.00092</b>
<b>CC</b>						
GO.ID	Term	Annotated	Significant	Expected	classicFisher	elimFisher
GO:0033557	Slx1-Slx4 complex	3	2	0.04	0.00049	0.00049
GO:0001772	immunological synapse	15	3	0.19	0.00086	0.00086
GO:0008385	IkappaB kinase complex	9	2	0.12	0.00564	0.00564
GO:0005887	integral to plasma membrane	1281	27	16.56	0.00817	0.00817
GO:0016604	nuclear body	272	9	3.52	0.00906	0.00906
<b>TBX6</b>						
<b>BP</b>						
GO.ID	Term	Annotated	Significant	Expected	classicFisher	elimFisher
GO:0009812	flavonoid metabolic process	13	3	0.19	0.00076	0.00076
GO:0006450	regulation of translational fidelity	4	2	0.06	0.00122	0.00122
GO:0010792	DNA double-strand break processing involved in repair via single-strand annealing	4	2	0.06	0.00122	0.00122
GO:0042373	vitamin K metabolic process	4	2	0.06	0.00122	0.00122
GO:0007596	blood coagulation	507	20	7.33	4.80E-05	0.00132
<b>CC</b>						
GO.ID	Term	Annotated	Significant	Expected	classicFisher	elimFisher
GO:0033557	Slx1-Slx4 complex	3	2	0.04	0.00058	0.00058
GO:0008305	integrin complex	29	4	0.4	0.00067	0.00067
GO:0001772	immunological synapse	15	3	0.21	0.00108	0.00108
GO:0005751	mitochondrial respiratory chain complex IV	4	2	0.06	0.00114	0.00114
GO:0031674	I band	68	5	0.95	0.0026	0.0026
<b>MAPK3</b>						
<b>BP</b>						
GO.ID	Term	Annotated	Significant	Expected	classicFisher	elimFisher
<b>GO:0010226</b>	<b>response to lithium ion</b>	<b>28</b>	<b>4</b>	<b>0.34</b>	<b>0.00033</b>	<b>0.00033</b>
GO:0009812	flavonoid metabolic process	13	3	0.16	0.00045	0.00045
<b>GO:0043149</b>	<b>stress fiber assembly</b>	<b>32</b>	<b>4</b>	<b>0.39</b>	<b>0.00056</b>	<b>0.00056</b>
GO:0010792	DNA double-strand break processing involved in repair via single-strand annealing	4	2	0.05	0.00085	0.00085
GO:0006584	catecholamine metabolic process	43	4	0.52	0.00174	0.00174
<b>CC</b>						
GO.ID	Term	Annotated	Significant	Expected	classicFisher	elimFisher
GO:0033557	Slx1-Slx4 complex	3	2	0.04	0.00041	0.00041
GO:0005634	nucleus	5740	91	67.23	0.00032	0.00059
GO:0008305	integrin complex	29	3	0.34	0.00462	0.00462
GO:0030131	clathrin adaptor complex	32	3	0.37	0.00612	0.00612
GO:0005665	DNA-directed RNA polymerase II, core complex	14	2	0.16	0.01132	0.01132

**Table S34: Enrichment of SFARI-ASD, ID/DD, Schizophrenia and BMI in BRICKs genes (Fisher's exact test). Table S34.4: Enrichment of "high confidence" ASD target genes in SFARI (Categories S, 1, 2; [https://gene.sfari.org/autdb/GS\\_Home.do](https://gene.sfari.org/autdb/GS_Home.do)) and identified in whole exome sequencing studies (De Rubeis et al., 2014; Iossifov et al., 2014) in BRICKs genes (Fisher's exact test).**

**Table S34.2**

Phenotype	Listed genes	BRICKs genes all viewpoints	Interacting genes all viewpoint	p-value	OR	5% CI	95 % CI	Reference list
Autism spectrum disorder (ASD)	604		78	1.65E-08	2.1	1.7	2.8	SFARI Autism Research (March 2014)
ID/DD	203	1382 (all)	17	3.20E-01	1.3	0.7	2.1	Reference-based
Schizophrenia	214		17	4.08E-01	1.2	0.7	2.0	Girard, 2011; Xu, 2011; Xu, 2012
BMI	541		30	3.36E-01	0.8	0.5	1.2	Vimalaswaran, 2012

**Table S34.2**

Phenotype	Listed genes	BRICKs genes per groups of viewpoints	Interacting genes per group	p-value	OR	5% CI	95 % CI	Reference list
Autism spectrum disorder (ASD)	604	588(220 kb)	32	6.69E-04	2.0	1.3	2.9	SFARI Autism Research (March 2014)
		978 (600 kb)	54	5.71E-06	2.0	1.5	2.7	

**Table S34.3**

Phenotype	Listed genes	BRICKs genes per viewpoint	Interacting genes per viewpoint	p-value	OR	5% CI	95 % CI	Reference list
Autism spectrum disorder (ASD)	604	213 (SH2B1)	11	6.07E-02	1.8	0.9	3.4	SFARI Autism Research (March 2014)
		484 (LAT)	25	5.58E-03	1.9	1.2	2.8	
		347 (MVP)	23	2.56E-04	2.4	1.5	3.7	
		232 (KCTD13)	9	3.26E-01	1.4	0.6	2.6	
		299 (ALDOA)	14	7.91E-02	1.7	0.9	2.9	
		303 (TBX6)	17	8.92E-03	2.0	1.2	3.3	
		259 (MAPK3)	11	1.90E-01	1.5	0.7	2.7	

**Table  
S34.4**

Phenotype	Listed genes	BRICKS genes all viewpoints	Interacting genes all viewpoint	p-value	OR	5% CI	95 % CI	Reference list
Autism spectrum disorder (ASD)	76	1382 (all)	13	1.40E-03	2.9	1.5	5.4	SFARI-Categories S, 1, 2
Autism spectrum disorder (ASD)	50		8	1.63E-02	2.7	1.1	5.8	De Rubeis, 2014; Iossifov, 2014
Autism spectrum disorder (ASD)	33		7	5.17E-03	3.8	1.4	9.0	De Rubeis, 2014
Autism spectrum disorder (ASD)	27		4	1.00E-01	2.5	0.6	7.2	Iossifov, 2014

## Supplemental Data

### 16p11.2 Consortium members

Maria Nicla Loviglio, Katrin Männik, Ilse van der Werf, Giuliana Giannuzzi, Marianna Zazhytska, Nele Gheldof, Eugenia Migliavacca, Ali A. Alfaiz, Inês Roberts-Caldeira, Loyse Hippolyte, Anne M. Maillard, Alessandra Ferrarini, Florence Niel Butsch, Bernard Conrad, Marie-Claude Addor, Marco Belfiore, Katharina Roetzer, Anke Van Dijck, Bettina Blaumeiser, Frank Kooy, Filip Roelens, Annelies Dheedene, Barbara Delle Chiaie, Björn Menten, Ann Oostra, Jean-Hubert Caberg, Melissa Carter, Barbara Kellam, Dimitri J Stavropoulos, Christian Marshall, Stephen W Scherer, Rosanna Weksberg, Cheryl Cytrynbaum, Anne Bassett, Chelsea Lowther, Jane Gillis, Sara MacKay, Iben Bache, Lilian B. Ousager, Maja Patricia Smerdel, Jesper Graakjaer, Susanne Kjaergaard, Andres Metspalu, Michele Mathieu, Dominique Bonneau, Agnes Guichet, Philippe Parent, Claude Férec, Marion Gerard, Ghislaine Plessis, James Lespinasse, Alice Masurel, Nathalie Marle, Laurence Faivre, Patrick Callier, Valerie Layet, Nathalie Le Meur, Céline Le Goff, Bénédicte Duban-Bedu, Sylvie Sukno, Odile Boute, Joris Andrieux, Patricia Blanchet, David Geneviève, Jacques Puechberty, Anouck Schneider, Bruno Leheup, Philippe Jonveaux, Sandra Mercier, Albert David, Cédric Le Caignec, Loïc de Pontual, Eva Pipiras, Aurelia Jacquette, Boris Keren, Brigitte Gilbert-Dussardier, Frederic Bilan, Alice Goldenberg, Pascal Chambon, Annick Toutain, Marianne Till, Damien Sanlaville, Barbara Leube, Brigitte Royer-Pokora,



Hans Jürgen Grabe, Carsten Oliver Schmidt, Claudia Schurmann, Georg Homuth, Gudmar Thorleifsson, Unnur Thorsteinsdottir, Laura Bernardini, Antonio Novelli, Lucia Micale, Giuseppe Merla, Marcella Zollino, Francesca Mari, Caterina Lo Rizzo, Alessandra Renieri, Margherita Silengo, Anneke T. Vulto-van Silfhout, Meyke Schouten, Rolph Pfundt, Nicole de Leeuw, Fleur Vansenne, Saskia M. Maas, Daniela Q.C.M. Barge-Schaapveld, Alida C. Knegt, Barbro Stadheim, Olaug Rodningen, Gunnar Houge, Sue Price, Lara Hawkes, Carolyn Campbell, Usha Kini, Julie Vogt, Robin Walters, Alexandra Blakemore, James F. Gusella, Yiping Shen, Daryl Scott, Carlos A Bacino, Karen Tsuchiya, Roger Ladda, Susan Sell, Alexander Asamoah, Aline I. Hamati, Jill A. Rosenfeld, Lisa G. Shaffer, Elyse Mitchell, Jennelle C. Hodge, Jacques S Beckmann, Sébastien Jacquemont, Alexandre Reymond. The affiliations and email addresses of the members of the 16p11.2 Consortium are listed in **Supplementary Table S36**.

#### 2p15 Consortium members

Alexandre Reymond, Lisa J. Ewans, David Mowat, Jan Walker, David J. Amor, Hilde Van Esch, Patricia Leroy, Jean-Hubert Caberg, John-Steven Bamforth, Deepti Babu, Marianne Till, Damien Sanlaville, David Geneviève, Jacques Puechberty, Bertrand Isidor, Nataliya DiDonato, Karl Hackmann, Marzia Passeggeri, Arie van Haeringen, Jill A. Rosenfeld, Lisa G. Shaffer, Rosemarie Smith, Sara Ellingwood, Darren M. Farber, Vinay Puri, Neda Zadeh, David D. Weaver, Mandy Miller, Timothy Wilks, Carolina J. Jorgez, DeeDee Lafayette, Sébastien Jacquemont. The affiliations and email addresses of the members of the 2p15 Consortium are listed in **Supplementary Table S37**.

#### Acknowledgments

We thank the members of the Lausanne Genomic Technologies Facility and Jacqueline Chrast for technical help and Catia Attanasio for comments on the manuscript. NG is a grantee of the Marie Heim Vögtlin and the Pro-Women programs of the SNSF and the Faculty of Biology and Medicine, University of Lausanne, respectively. KM was awarded a fellowship from the Swiss Scientific

Exchange NMS Program. IvdW was recipient of scholarships from the Nora Baart Foundation and Erasmus European Union Lifelong Learning Programme. AAA was awarded a scholarship from the Saudi Arabian National Guard Health Affairs. SMW was supported by SNSF (Swiss National Science Foundation)(P2ELP3\_155365) and EMBO (ALTF 755-2014) fellowships. SJ is recipient of a SNSF Bursary Professor fellowship. This work is supported by grants from the Simons Foundation (SFARI274424 to AR), the Swiss National Science Foundation 31003A\_160203 (AR), SNSF Sinergia grants CRSII33-133044 (AR) and CRSI33-130326 (ETD, BD, AR), the Leenaards Foundation Prize (SJ, AR) and the SystemsX.ch initiative grant 3826 (ETD and AR). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

#### Authors contribution

MNL, NG and GG performed the 4C and FISH experiments. ML, MNL, EM, AAA and JR conducted the statistical analyses. KM, MNL and IvdW established cell lines. IvdW, KM and MZ performed the qPCR experiments. SMW, ETD and BD run the molecular associations. IRC, KM, IvdW, MP, LH, AMM, RFK, DS, JAR, LGS, JA, CM, SWS, YS, JFG, UT, GT, JSB, SJ and all other members of the 16p11.2 and 2p15 consortia phenotyped and genotyped individuals. MNL and AR wrote the manuscript with contributions from JSB and SJ. AR designed the study and obtained the necessary funding. All authors commented on and approved the manuscript.

#### Competing interest

The authors declare no competing interests.

#### Accession Numbers section

GEO Series accession number: GSE57802

### The Web Resources section

4C-seq pipeline: <http://htsstation.epfl.ch/>

BioScript: <http://gdv.epfl.ch/bs>

GeneProf: <http://www.geneprof.org>

GTEx: <http://www.gtexportal.org/home/>

gFeatBrowser: <http://www.gfeatbrowser.com>

Online Mendelian Inheritance in Man: <http://www.omim.org>

SFARI: <https://sfari.org/>

STRING: <http://string-db.org/>

## References

1. de Laat W, Duboule D. Topology of mammalian developmental enhancers and their regulatory landscapes. *Nature* 2013; **502**(7472): 499-506.
2. Fanucchi S, Shibayama Y, Burd S, Weinberg MS, Mhlanga MM. Chromosomal contact permits transcription between coregulated genes. *Cell* 2013; **155**(3): 606-620.
3. Lupianez DG, Kraft K, Heinrich V, Krawitz P, Brancati F, Klopocki E *et al.* Disruptions of topological chromatin domains cause pathogenic rewiring of gene-enhancer interactions. *Cell* 2015; **161**(5): 1012-1025.
4. Lupianez DG, Spielmann M, Mundlos S. Breaking TADs: How Alterations of Chromatin Domains Result in Disease. *Trends in genetics : TIG* 2016.
5. Johnson ME, Viggiano L, Bailey JA, Abdul-Rauf M, Goodwin G, Rocchi M *et al.* Positive selection of a gene family during the emergence of humans and African apes. *Nature* 2001; **413**(6855): 514-519.
6. Antonacci F, Kidd JM, Marques-Bonet T, Teague B, Ventura M, Girirajan S *et al.* A large and complex structural polymorphism at 16p12.1 underlies microdeletion disease risk. *Nature genetics* 2010; **42**(9): 745-750.
7. Giannuzzi G, Migliavacca E, Reymond A. Novel H3K4me3 marks are enriched at human- and chimpanzee-specific cytogenetic structures. *Genome research* 2014; **24**(9): 1455-1468.
8. Ballif BC, Hornor SA, Jenkins E, Madan-Khetarpal S, Surti U, Jackson KE *et al.* Discovery of a previously unrecognized microdeletion syndrome of 16p11.2-p12.2. *Nature genetics* 2007; **39**(9): 1071-1073.
9. Weiss LA, Shen Y, Korn JM, Arking DE, Miller DT, Fossdal R *et al.* Association between microdeletion and microduplication at 16p11.2 and autism. *The New England journal of medicine* 2008; **358**(7): 667-675.
10. Girirajan S, Rosenfeld JA, Cooper GM, Antonacci F, Siswara P, Itsara A *et al.* A recurrent 16p12.1 microdeletion supports a two-hit model for severe developmental delay. *Nature genetics* 2010; **42**(3): 203-209.

11. Hastings PJ, Lupski JR, Rosenberg SM, Ira G. Mechanisms of change in gene copy number. *Nature reviews Genetics* 2009; **10**(8): 551-564.
12. Migliavacca E, Golzio C, Mannik K, Blumenthal I, Oh EC, Harewood L *et al.* A Potential Contributory Role for Ciliary Dysfunction in the 16p11.2 600 kb BP4-BP5 Pathology. *Am J Hum Genet* 2015; **96**(5): 784-796.
13. Glessner JT, Wang K, Cai G, Korvatska O, Kim CE, Wood S *et al.* Autism genome-wide copy number variation reveals ubiquitin and neuronal genes. *Nature* 2009; **459**(7246): 569-573.
14. Marshall CR, Noor A, Vincent JB, Lionel AC, Feuk L, Skaug J *et al.* Structural variation of chromosomes in autism spectrum disorder. *Am J Hum Genet* 2008; **82**(2): 477-488.
15. Zufferey F, Sherr EH, Beckmann ND, Hanson E, Maillard AM, Hippolyte L *et al.* A 600 kb deletion syndrome at 16p11.2 leads to energy imbalance and neuropsychiatric disorders. *Journal of medical genetics* 2012; **49**(10): 660-668.
16. Bochukova EG, Huang N, Keogh J, Henning E, Purmann C, Blaszczyk K *et al.* Large, rare chromosomal deletions associated with severe early-onset obesity. *Nature* 2010; **463**(7281): 666-670.
17. Walters RG, Jacquemont S, Valsesia A, de Smith AJ, Martinet D, Andersson J *et al.* A new highly penetrant form of obesity due to deletions on chromosome 16p11.2. *Nature* 2010; **463**(7281): 671-675.
18. Jacquemont S, Reymond A, Zufferey F, Harewood L, Walters RG, Kutalik Z *et al.* Mirror extreme BMI phenotypes associated with gene dosage at the chromosome 16p11.2 locus. *Nature* 2011; **478**(7367): 97-102.
19. Shinawi M, Liu P, Kang SH, Shen J, Belmont JW, Scott DA *et al.* Recurrent reciprocal 16p11.2 rearrangements associated with global developmental delay, behavioural problems, dysmorphism, epilepsy, and abnormal head size. *Journal of medical genetics* 2010; **47**(5): 332-341.
20. Reinthaler EM, Lal D, Lebon S, Hildebrand MS, Dahl HH, Regan BM *et al.* 16p11.2 600 kb Duplications confer risk for typical and atypical Rolandic epilepsy. *Human molecular genetics* 2014.
21. Dimassi S, Labalme A, Lesca G, Rudolf G, Bruneau N, Hirsch E *et al.* A subset of genomic alterations detected in rolandic epilepsies contains candidate or known epilepsy genes including GRIN2A and PRRT2. *Epilepsia* 2014; **55**(2): 370-378.

22. McCarthy SE, Makarov V, Kirov G, Addington AM, McClellan J, Yoon S *et al.* Microduplications of 16p11.2 are associated with schizophrenia. *Nature genetics* 2009; **41**(11): 1223-1227.
23. D'Angelo D, Lebon S, Chen Q, Martin-Brevet S, Snyder LG, Hippolyte L *et al.* Defining the Effect of the 16p11.2 Duplication on Cognition, Behavior, and Medical Comorbidities. *JAMA Psychiatry* 2016; **73**(1): 20-30.
24. Mannik K, Magi R, Mace A, Cole B, Guyatt AL, Shihab HA *et al.* Copy number variations and cognitive phenotypes in unselected populations. *JAMA : the journal of the American Medical Association* 2015; **313**(20): 2044-2054.
25. Maillard AM, Ruef A, Pizzagalli F, Migliavacca E, Hippolyte L, Adaszewski S *et al.* The 16p11.2 locus modulates brain structures common to autism, schizophrenia and obesity. *Molecular psychiatry* 2015; **20**(1): 140-147.
26. Blumenthal I, Ragavendran A, Erdin S, Klei L, Sugathan A, Guide JR *et al.* Transcriptional consequences of 16p11.2 deletion and duplication in mouse cortex and multiplex autism families. *Am J Hum Genet* 2014; **94**(6): 870-883.
27. Golzio C, Willer J, Talkowski ME, Oh EC, Taniguchi Y, Jacquemont S *et al.* KCTD13 is a major driver of mirrored neuroanatomical phenotypes of the 16p11.2 copy number variant. *Nature* 2012; **485**(7398): 363-367.
28. Walters RG, Coin LJ, Ruukonen A, de Smith AJ, El-Sayed Moustafa JS, Jacquemont S *et al.* Rare genomic structural variants in complex disease: lessons from the replication of associations with obesity. *PloS one* 2013; **8**(3): e58048.
29. Barge-Schaapveld DQ, Maas SM, Polstra A, Knegt LC, Hennekam RC. The atypical 16p11.2 deletion: a not so atypical microdeletion syndrome? *American journal of medical genetics Part A* 2011; **155A**(5): 1066-1072.
30. Bachmann-Gagescu R, Mefford HC, Cowan C, Glew GM, Hing AV, Wallace S *et al.* Recurrent 200-kb deletions of 16p11.2 that include the SH2B1 gene are associated with developmental delay and obesity. *Genetics in medicine : official journal of the American College of Medical Genetics* 2010; **12**(10): 641-647.
31. Sampson MG, Coughlin CR, 2nd, Kaplan P, Conlin LK, Meyers KE, Zackai EH *et al.* Evidence for a recurrent microdeletion at chromosome 16p11.2 associated with congenital anomalies of the kidney and urinary tract (CAKUT) and Hirschsprung disease. *American journal of medical genetics Part A* 2010; **152A**(10): 2618-2622.

32. Girirajan S, Rosenfeld JA, Coe BP, Parikh S, Friedman N, Goldstein A *et al.* Phenotypic heterogeneity of genomic disorders and rare copy-number variants. *The New England journal of medicine* 2012; **367**(14): 1321-1331.
33. Cooper GM, Coe BP, Girirajan S, Rosenfeld JA, Vu TH, Baker C *et al.* A copy number variation morbidity map of developmental delay. *Nature genetics* 2011; **43**(9): 838-846.
34. Bijlsma EK, Gijsbers AC, Schuurs-Hoeijmakers JH, van Haeringen A, Fransen van de Putte DE, Anderlid BM *et al.* Extending the phenotype of recurrent rearrangements of 16p11.2: deletions in mentally retarded patients without autism and in normal individuals. *European journal of medical genetics* 2009; **52**(2-3): 77-87.
35. Guha S, Rees E, Darvasi A, Ivanov D, Ikeda M, Bergen SE *et al.* Implication of a rare deletion at distal 16p11.2 in schizophrenia. *JAMA Psychiatry* 2013; **70**(3): 253-260.
36. Sabatti C, Service SK, Hartikainen AL, Pouta A, Ripatti S, Brodsky J *et al.* Genome-wide association analysis of metabolic traits in a birth cohort from a founder population. *Nature genetics* 2009; **41**(1): 35-46.
37. Nelis M, Esko T, Magi R, Zimprich F, Zimprich A, Toncheva D *et al.* Genetic structure of Europeans: a view from the North-East. *PLoS one* 2009; **4**(5): e5472.
38. John U, Greiner B, Hensel E, Ludemann J, Piek M, Sauer S *et al.* Study of Health In Pomerania (SHIP): a health examination survey in an east German region: objectives and design. *Sozial- und Präventivmedizin* 2001; **46**(3): 186-194.
39. Firmann M, Mayor V, Vidal PM, Bochud M, Pecoud A, Hayoz D *et al.* The CoLaus study: a population-based study to investigate the epidemiology and genetic determinants of cardiovascular risk factors and metabolic syndrome. *BMC cardiovascular disorders* 2008; **8**: 6.
40. Chabchoub E, Willekens D, Vermeesch JR, Fryns JP. Holoprosencephaly and ZIC2 microdeletions: novel clinical and epidemiological specificities delineated. *Clinical genetics* 2012; **81**(6): 584-589.
41. de Leeuw N, Pfundt R, Koolen DA, Neefs I, Scheltinga I, Mieloo H *et al.* A newly recognised microdeletion syndrome involving 2p15p16.1: narrowing down the critical region by adding another patient detected by genome wide tiling path array comparative genomic hybridisation analysis. *Journal of medical genetics* 2008; **45**(2): 122-124.

42. Felix TM, Petrin AL, Sanseverino MT, Murray JC. Further characterization of microdeletion syndrome involving 2p15-p16.1. *American journal of medical genetics Part A* 2010; **152A**(10): 2604-2608.
43. Florisson JM, Mathijssen IM, Dumeé B, Hoogeboom JA, Poddighe PJ, Oostra BA *et al.* Complex craniosynostosis is associated with the 2p15p16.1 microdeletion syndrome. *American journal of medical genetics Part A* 2013; **161A**(2): 244-253.
44. Hancarova M, Vejvalkova S, Trkova M, Drabova J, Dleskova A, Vlckova M *et al.* Identification of a patient with intellectual disability and de novo 3.7 Mb deletion supports the existence of a novel microdeletion syndrome in 2p14-p15. *Gene* 2013; **516**(1): 158-161.
45. Huchtagowder V, Liu TC, Paciorkowski AR, Thio LL, Keller MS, Anderson CD *et al.* Chromosome 2p15p16.1 microdeletion syndrome: 2.5 Mb deletion in a patient with renal anomalies, intractable seizures and a choledochal cyst. *European journal of medical genetics* 2012; **55**(8-9): 485-489.
46. Liang JS, Shimojima K, Ohno K, Sugiura C, Une Y, Ohno K *et al.* A newly recognised microdeletion syndrome of 2p15-16.1 manifesting moderate developmental delay, autistic behaviour, short stature, microcephaly, and dysmorphic features: a new patient with 3.2 Mb deletion. *Journal of medical genetics* 2009; **46**(9): 645-647.
47. Piccione M, Piro E, Serraino F, Cavani S, Ciccone R, Malacarne M *et al.* Interstitial deletion of chromosome 2p15-16.1: report of two patients and critical review of current genotype-phenotype correlation. *European journal of medical genetics* 2012; **55**(4): 238-244.
48. Prontera P, Bernardini L, Stangoni G, Capalbo A, Rogaia D, Romani R *et al.* Deletion 2p15-16.1 syndrome: case report and review. *American journal of medical genetics Part A* 2011; **155A**(10): 2473-2478.
49. Rajcan-Separovic E, Harvard C, Liu X, McGillivray B, Hall JG, Qiao Y *et al.* Clinical and molecular cytogenetic characterisation of a newly recognised microdeletion syndrome involving 2p15-16.1. *Journal of medical genetics* 2007; **44**(4): 269-276.
50. Jorgez CJ, Rosenfeld JA, Wilken NR, Vangapandu HV, Sahin A, Pham D *et al.* Genitourinary defects associated with genomic deletions in 2p15 encompassing OTX1. *PloS one* 2014; **9**(9): e107028.
51. Simonis M, Klous P, Splinter E, Moshkin Y, Willemsen R, de Wit E *et al.* Nuclear organization of active and inactive chromatin domains uncovered by chromosome conformation capture-on-chip (4C). *Nature genetics* 2006; **38**(11): 1348-1354.



52. Simonis M, Kooren J, de Laat W. An evaluation of 3C-based methods to capture DNA interactions. *Nature methods* 2007; **4**(11): 895-901.
53. Noordermeer D, Leleu M, Splinter E, Rougemont J, De Laat W, Duboule D. The dynamic architecture of Hox gene clusters. *Science* 2011; **334**(6053): 222-225.
54. Gheldof N, Leleu M, Noordermeer D, Rougemont J, Reymond A. Detecting long-range chromatin interactions using the chromosome conformation capture sequencing (4C-seq) method. *Methods Mol Biol* 2012; **786**: 211-225.
55. Blaker-Lee A, Gupta S, McCammon JM, De Rienzo G, Sive H. Zebrafish homologs of genes within 16p11.2, a genomic region associated with brain disorders, are active during brain development, and include two deletion dosage sensor genes. *Disease models & mechanisms* 2012; **5**(6): 834-851.
56. Kishi H, Mukai T, Hirono A, Fujii H, Miwa S, Hori K. Human aldolase A deficiency associated with a hemolytic anemia: thermolabile aldolase due to a single base mutation. *Proceedings of the National Academy of Sciences of the United States of America* 1987; **84**(23): 8623-8627.
57. Watabe-Rudolph M, Schlautmann N, Papaioannou VE, Gossler A. The mouse rib-vertebrae mutation is a hypomorphic Tbx6 allele. *Mechanisms of development* 2002; **119**(2): 251-256.
58. Fei Q, Wu Z, Wang H, Zhou X, Wang N, Ding Y *et al.* The association analysis of TBX6 polymorphism with susceptibility to congenital scoliosis in a Chinese Han population. *Spine* 2010; **35**(9): 983-988.
59. Sparrow DB, McInerney-Leo A, Gucev ZS, Gardiner B, Marshall M, Leo PJ *et al.* Autosomal dominant spondylocostal dysostosis is caused by mutation in TBX6. *Human molecular genetics* 2013.
60. Wu N, Ming X, Xiao J, Wu Z, Chen X, Shinawi M *et al.* TBX6 null variants and a common hypomorphic allele in congenital scoliosis. *The New England journal of medicine* 2015; **372**(4): 341-350.
61. Duan C, Li M, Rui L. SH2-B promotes insulin receptor substrate 1 (IRS1)- and IRS2-mediated activation of the phosphatidylinositol 3-kinase pathway in response to leptin. *The Journal of biological chemistry* 2004; **279**(42): 43684-43691.

62. Ren D, Zhou Y, Morris D, Li M, Li Z, Rui L. Neuronal SH2B1 is essential for controlling energy and glucose homeostasis. *The Journal of clinical investigation* 2007; **117**(2): 397-406.
63. Willer CJ, Speliotes EK, Loos RJ, Li S, Lindgren CM, Heid IM *et al.* Six new loci associated with body mass index highlight a neuronal influence on body weight regulation. *Nature genetics* 2009; **41**(1): 25-34.
64. Thorleifsson G, Walters GB, Gudbjartsson DF, Steinthorsdottir V, Sulem P, Helgadóttir A *et al.* Genome-wide association yields new sequence variants at seven loci that associate with measures of obesity. *Nature genetics* 2009; **41**(1): 18-24.
65. Speliotes EK, Willer CJ, Berndt SI, Monda KL, Thorleifsson G, Jackson AU *et al.* Association analyses of 249,796 individuals reveal 18 new loci associated with body mass index. *Nature genetics* 2010; **42**(11): 937-948.
66. Jamshidi Y, Snieder H, Ge D, Spector TD, O'Dell SD. The SH2B gene is associated with serum leptin and body fat in normal female twins. *Obesity (Silver Spring)* 2007; **15**(1): 5-9.
67. Doche ME, Bochukova EG, Su HW, Pearce LR, Keogh JM, Henning E *et al.* Human SH2B1 mutations are associated with maladaptive behaviors and obesity. *The Journal of clinical investigation* 2012; **122**(12): 4732-4736.
68. Fuller DM, Zhu M, Ou-Yang CW, Sullivan SA, Zhang W. A tale of two TRAPs: LAT and LAB in the regulation of lymphocyte development, activation, and autoimmunity. *Immunologic research* 2011; **49**(1-3): 97-108.
69. Shim EK, Jung SH, Lee JR. Role of two adaptor molecules SLP-76 and LAT in the PI3K signaling pathway in activated T cells. *Journal of immunology* 2011; **186**(5): 2926-2935.
70. Fuller DM, Zhu M, Koonpaew S, Nelson MI, Zhang W. The importance of the Erk pathway in the development of linker for activation of T cells-mediated autoimmunity. *Journal of immunology* 2012; **189**(8): 4005-4013.
71. David FP, Delafontaine J, Carat S, Ross FJ, Lefebvre G, Jarosz Y *et al.* HTSstation: a web application and open-access libraries for high-throughput sequencing data analysis. *PloS one* 2014; **9**(1): e85879.
72. Tolhuis B, Blom M, Kerkhoven RM, Pagie L, Teunissen H, Nieuwland M *et al.* Interactions among Polycomb domains are guided by chromosome architecture. *PLoS genetics* 2011; **7**(3): e1001343.

73. Lieberman-Aiden E, van Berkum NL, Williams L, Imakaev M, Ragoczy T, Telling A *et al.* Comprehensive mapping of long-range interactions reveals folding principles of the human genome. *Science* 2009; **326**(5950): 289-293.
74. de Wit E, Braunschweig U, Greil F, Bussemaker HJ, van Steensel B. Global chromatin domain organization of the Drosophila genome. *PLoS genetics* 2008; **4**(3): e1000045.
75. Rao SS, Huntley MH, Durand NC, Stamenova EK, Bochkov ID, Robinson JT *et al.* A 3D Map of the Human Genome at Kilobase Resolution Reveals Principles of Chromatin Looping. *Cell* 2014; **159**(7): 1665-1680.
76. Franceschini A, Szklarczyk D, Frankild S, Kuhn M, Simonovic M, Roth A *et al.* STRING v9.1: protein-protein interaction networks, with increased coverage and integration. *Nucleic acids research* 2013; **41**(Database issue): D808-815.
77. Alexa A, Rahnenfuhrer J. topGO: Enrichment analysis for Gene Ontology. In: BioConductor (ed).2010.
78. Huang da W, Sherman BT, Lempicki RA. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nature protocols* 2009; **4**(1): 44-57.
79. Reimand J, Arak T, Vilo J. g:Profiler--a web server for functional interpretation of gene lists (2011 update). *Nucleic acids research* 2011; **39**(Web Server issue): W307-315.
80. Chen EY, Tan CM, Kou Y, Duan Q, Wang Z, Meirelles GV *et al.* Enrichr: interactive and collaborative HTML5 gene list enrichment analysis tool. *BMC bioinformatics* 2013; **14**: 128.
81. Hamosh A, Scott AF, Amberger J, Valle D, McKusick VA. Online Mendelian Inheritance in Man (OMIM). *Human mutation* 2000; **15**(1): 57-61.
82. Girard SL, Gauthier J, Noreau A, Xiong L, Zhou S, Jouan L *et al.* Increased exonic de novo mutation rate in individuals with schizophrenia. *Nature genetics* 2011; **43**(9): 860-863.
83. Xu B, Ionita-Laza I, Roos JL, Boone B, Woodrick S, Sun Y *et al.* De novo gene mutations highlight patterns of genetic and neural complexity in schizophrenia. *Nature genetics* 2012; **44**(12): 1365-1369.

84. Xu B, Roos JL, Dexheimer P, Boone B, Plummer B, Levy S *et al.* Exome sequencing supports a de novo mutational paradigm for schizophrenia. *Nature genetics* 2011; **43**(9): 864-868.
85. Vimalaswaran KS, Tachmazidou I, Zhao JH, Hirschhorn JN, Dudbridge F, Loos RJ. Candidate genes for obesity-susceptibility show enriched association within a large genome-wide association study for BMI. *Human molecular genetics* 2012; **21**(20): 4537-4542.
86. De Rubeis S, He X, Goldberg AP, Poultney CS, Samocha K, Ercument Cicek A *et al.* Synaptic, transcriptional and chromatin genes disrupted in autism. *Nature* 2014; **515**(7526): 209-215.
87. Iossifov I, O'Roak BJ, Sanders SJ, Ronemus M, Krumm N, Levy D *et al.* The contribution of de novo coding mutations to autism spectrum disorder. *Nature* 2014; **515**(7526): 216-221.
88. van Dam TJ, Wheway G, Slaats GG, Group SS, Huynen MA, Giles RH. The SYSCILIA gold standard (SCGSv1) of known ciliary components and its applications within a systems biology consortium. *Cilia* 2013; **2**(1): 7.
89. Stehbens S, Wittmann T. Targeting and transport: how microtubules control focal adhesion dynamics. *The Journal of cell biology* 2012; **198**(4): 481-489.
90. Bukoreshtliev NV, Haase K, Pelling AE. Mechanical cues in cellular signalling and communication. *Cell and tissue research* 2013; **352**(1): 77-94.
91. Seeger-Nukpezah T, Golemis EA. The extracellular matrix and ciliary signaling. *Current opinion in cell biology* 2012; **24**(5): 652-661.
92. He G, Liu X, Qin W, Chen Q, Wang X, Yang Y *et al.* MPZL1/PZR, a novel candidate predisposing schizophrenia in Han Chinese. *Molecular psychiatry* 2006; **11**(8): 748-751.
93. Eminaga S, Bennett AM. Noonan syndrome-associated SHP-2/Ptpn11 mutants enhance SIRPalpha and PZR tyrosyl phosphorylation and promote adhesion-mediated ERK activation. *The Journal of biological chemistry* 2008; **283**(22): 15328-15338.
94. Lichter P, Tang CJ, Call K, Hermanson G, Evans GA, Housman D *et al.* High-resolution mapping of human chromosome 11 by in situ hybridization with cosmid clones. *Science* 1990; **247**(4938): 64-69.

95. Waszak SM, Delaneau O, Gschwind AR, Kilpinen H, Raghav SK, Witwicki RM *et al.* Population Variation and Genetic Control of Modular Chromatin Architecture in Humans. *Cell* 2015; **162**(5): 1039-1050.
96. Genomes Project C, Abecasis GR, Altshuler D, Auton A, Brooks LD, Durbin RM *et al.* A map of human genome variation from population-scale sequencing. *Nature* 2010; **467**(7319): 1061-1073.
97. Kilpinen H, Waszak SM, Gschwind AR, Raghav SK, Witwicki RM, Orioli A *et al.* Coordinated effects of sequence variation on DNA binding, chromatin structure, and transcription. *Science* 2013; **342**(6159): 744-747.
98. Lappalainen T, Sammeth M, Friedlander MR, t Hoen PA, Monlong J, Rivas MA *et al.* Transcriptome and genome sequencing uncovers functional variation in humans. *Nature* 2013; **501**(7468): 506-511.
99. CDC. Prevalence of Autism Spectrum Disorder Among Children Aged 8 Years — Autism and Developmental Disabilities Monitoring Network, 11 Sites, United States, 2010. *MMWR Surveill Summ*, vol. 632014, pp 1-21.
100. Gheldof N, Witwicki RM, Migliavacca E, Leleu M, Didelot G, Harewood L *et al.* Structural variation-associated expression changes are paralleled by chromatin architecture modifications. *PloS one* 2013; **8**(11): e79973.
101. Dixon JR, Selvaraj S, Yue F, Kim A, Li Y, Shen Y *et al.* Topological domains in mammalian genomes identified by analysis of chromatin interactions. *Nature* 2012; **485**(7398): 376-380.
102. Zhang Y, Wong CH, Birnbaum RY, Li G, Favaro R, Ngan CY *et al.* Chromatin connectivity maps reveal dynamic promoter-enhancer long-range associations. *Nature* 2013; **504**(7479): 306-310.
103. Butler MG, Dasouki MJ, Zhou XP, Talebizadeh Z, Brown M, Takahashi TN *et al.* Subset of individuals with autism spectrum disorders and extreme macrocephaly associated with germline PTEN tumour suppressor gene mutations. *Journal of medical genetics* 2005; **42**(4): 318-321.
104. Herman GE, Butter E, Enrile B, Pastore M, Prior TW, Sommer A. Increasing knowledge of PTEN germline mutations: Two additional patients with autism and macrocephaly. *American journal of medical genetics Part A* 2007; **143**(6): 589-593.

105. O'Roak BJ, Vives L, Fu W, Egertson JD, Stanaway IB, Phelps IG *et al.* Multiplex targeted sequencing identifies recurrently mutated genes in autism spectrum disorders. *Science* 2012; **338**(6114): 1619-1622.
106. Pal A, Barber TM, Van de Bunt M, Rudge SA, Zhang Q, Lachlan KL *et al.* PTEN mutations as a cause of constitutive insulin sensitivity and obesity. *The New England journal of medicine* 2012; **367**(11): 1002-1011.
107. Girirajan S, Dennis MY, Baker C, Malig M, Coe BP, Campbell CD *et al.* Refinement and discovery of new hotspots of copy-number variation associated with autism spectrum disorder. *Am J Hum Genet* 2013; **92**(2): 221-237.
108. Harvard C, Strong E, Mercier E, Colnaghi R, Alcantara D, Chow E *et al.* Understanding the impact of 1q21.1 copy number variant. *Orphanet journal of rare diseases* 2011; **6**: 54.
109. Brunetti-Pierri N, Berg JS, Scaglia F, Belmont J, Bacino CA, Sahoo T *et al.* Recurrent reciprocal 1q21.1 deletions and duplications associated with microcephaly or macrocephaly and developmental and behavioral abnormalities. *Nature genetics* 2008; **40**(12): 1466-1471.
110. Griswold AJ, Ma D, Cukier HN, Nations LD, Schmidt MA, Chung RH *et al.* Evaluation of copy number variations reveals novel candidate genes in autism spectrum disorder-associated pathways. *Human molecular genetics* 2012; **21**(15): 3513-3523.
111. Liu X, Malenfant P, Reesor C, Lee A, Hudson ML, Harvard C *et al.* 2p15-p16.1 microdeletion syndrome: molecular characterization and association of the OTX1 and XPO1 genes with autism spectrum disorders. *European journal of human genetics : EJHG* 2011; **19**(12): 1264-1270.
112. Fannemel M, Baroy T, Holmgren A, Rodningen OK, Haugsand TM, Hansen B *et al.* Haploinsufficiency of XPO1 and USP34 by a de novo 230 kb deletion in 2p15, in a patient with mild intellectual disability and cranio-facial dysmorphisms. *European journal of medical genetics* 2014.
113. Vaags AK, Lionel AC, Sato D, Goodenberger M, Stein QP, Curran S *et al.* Rare deletions at the neurexin 3 locus in autism spectrum disorder. *Am J Hum Genet* 2012; **90**(1): 133-141.
114. Seltzer LE, Paciorkowski AR. Genetic disorders associated with postnatal microcephaly. *American journal of medical genetics Part C, Seminars in medical genetics* 2014; **166**(2): 140-155.

115. Bjursell MK, Blom HJ, Cayuela JA, Engvall ML, Lesko N, Balasubramaniam S *et al.* Adenosine kinase deficiency disrupts the methionine cycle and causes hypermethioninemia, encephalopathy, and abnormal liver function. *Am J Hum Genet* 2011; **89**(4): 507-515.
116. Doornbos M, Sikkema-Raddatz B, Ruijvenkamp CA, Dijkhuizen T, Bijlsma EK, Gijsbers AC *et al.* Nine patients with a microdeletion 15q11.2 between breakpoints 1 and 2 of the Prader-Willi critical region, possibly associated with behavioural disturbances. *European journal of medical genetics* 2009; **52**(2-3): 108-115.
117. Kolehmainen J, Wilkinson R, Lehesjoki AE, Chandler K, Kivitie-Kallio S, Clayton-Smith J *et al.* Delineation of Cohen syndrome following a large-scale genotype-phenotype screen. *Am J Hum Genet* 2004; **75**(1): 122-127.
118. Chung JH, Eng C. Nuclear-cytoplasmic partitioning of phosphatase and tensin homologue deleted on chromosome 10 (PTEN) differentially regulates the cell cycle and apoptosis. *Cancer research* 2005; **65**(18): 8096-8100.
119. Kwon CH, Luikart BW, Powell CM, Zhou J, Matheny SA, Zhang W *et al.* Pten regulates neuronal arborization and social interaction in mice. *Neuron* 2006; **50**(3): 377-388.
120. Adviento B, Corbin IL, Widjaja F, Desachy G, Enrique N, Rosser T *et al.* Autism traits in the RASopathies. *Journal of medical genetics* 2014; **51**(1): 10-20.
121. Alfieri P, Piccini G, Caciolo C, Perrino F, Gambardella ML, Mallardi M *et al.* Behavioral profile in RASopathies. *American journal of medical genetics Part A* 2014; **164A**(4): 934-942.
122. Du M, Yuan T, Schilter KF, Dittmar RL, Mackinnon A, Huang X *et al.* Prostate cancer risk locus at 8q24 as a regulatory hub by physical interactions with multiple genomic loci across the genome. *Human molecular genetics* 2014.
123. Balikova I, Lehesjoki AE, de Ravel TJ, Thienpont B, Chandler KE, Clayton-Smith J *et al.* Deletions in the VPS13B (COH1) gene as a cause of Cohen syndrome. *Human mutation* 2009; **30**(9): E845-854.
124. Mahjoub MR, Stearns T. Supernumerary centrosomes nucleate extra cilia and compromise primary cilium signaling. *Current biology : CB* 2012; **22**(17): 1628-1634.
125. Bartsch O, Labonte J, Albrecht B, Wieczorek D, Lechno S, Zechner U *et al.* Two patients with EP300 mutations and facial dysmorphism different from the classic Rubinstein-Taybi syndrome. *American journal of medical genetics Part A* 2010; **152A**(1): 181-184.

126. Gavrilov AA, Gushchanskaya ES, Strelkova O, Zhironkina O, Kireev, II, Iarovaia OV *et al.* Disclosure of a structural milieu for the proximity ligation reveals the elusive nature of an active chromatin hub. *Nucleic acids research* 2013; **41**(6): 3563-3575.
127. Edelman LB, Fraser P. Transcription factories: genetic programming in three dimensions. *Current opinion in genetics & development* 2012; **22**(2): 110-114.
128. Park SK, Xiang Y, Feng X, Garrard WT. Pronounced cohabitation of active immunoglobulin genes from three different chromosomes in transcription factories during maximal antibody synthesis. *Genes & development* 2014; **28**(11): 1159-1164.
129. Mitchell AC, Bharadwaj R, Whittle C, Krueger W, Mirnics K, Hurd Y *et al.* The genome in three dimensions: a new frontier in human brain research. *Biological psychiatry* 2014; **75**(12): 961-969.



## Chapter 2: A potential role for the linker for activation of T-cells (LAT) in the neuroanatomical phenotype of 16p11.2 BP2-BP3 CNVs

### Summary of the contribution

The present chapter presents the unpublished results obtained at Center for Human Disease Modeling (CHDM) under the supervision of Prof. Nicholas Katsanis and Prof. Christelle Golzio. I conducted all the experiments and analysis described in the following paragraphs.

The results are summarized in three main figures (named **Figure 1-3**) and two supplementary figures (**Figure S1-2**).

## **A potential role for the linker for activation of T-cells (LAT) in the neuroanatomical phenotype of 16p11.2 BP2-BP3 CNVs**

### **Abstract**

Copy number variants (CNVs) of the distal 16p11.2 220 kb BP2-BP3 region show mirror effect on BMI and head size, and association with autism spectrum disorders (ASD) and schizophrenia, as reported previously for the proximal 16p11.2 600 kb BP4-BP5 deletion and duplication. These two CNV-prone regions are also engaged reciprocally in complex chromatin looping confirmed by 4C, FISH, Hi-C and concomitant expression changes. We assessed the 220 kb BP2-BP3 duplication by overexpressing each of the nine encompassed human transcripts, *CD19*, *NFATC2IP*, *ATXN2L*, *TUFM*, *ATP2A1*, *RABEP2*, *SPNS1*, *LAT* and *SH2B1*, in zebrafish embryos. Overexpression of the linker for activation of T cells (*LAT*) induced a reduction of proliferating cells in the brain, a reduced number of post-mitotic neurons in the anterior forebrain at 2 days-post-proliferation (dpf), and of intertectal axonal tracts at 3dpf, resulting in microcephaly at later stages. We found similar effects upon overexpression of *CD247* and *ZAP70*, two genes encoding members of the immune system signaling pathway mediated by *LAT*. Co-injections experiments showed that *KCTD13*, *MVP*, and *MAPK3* (major driver and modifiers of the 600 kb BP4-BP5 syndromes, respectively) showed that these genes interacted with *LAT* in an additive manner, and co-injected embryos exhibited an increased severity of the microcephaly compared to single-injected embryos, suggesting the presence of genetic interaction. Taken together, our results suggest that *LAT*, besides its well-recognized function in T-cells development, is a major contributor of the 16p11.2 (BP2-BP3) 220 kb CNV-associated neurodevelopmental phenotypes.

## Introduction

A major challenge in the interpretation of CNV spanning over large regions and encompassing several genes is the identification of the critical locus whose dosage sensitivity drives the phenotype. Deletions and duplications at the 16p11.2 600 kb BP4-BP5 regions are among the most frequent causes of neurodevelopmental and neuropsychiatric disorders<sup>1-4</sup> and are associated with mirror phenotypes on body mass index (BMI), head circumference (HC), and brain volume<sup>5, 6</sup>. Distal to this region, the 16p11.2 220 kb BP2-BP3 locus was initially described for the enrichment of its deletion in patients with early-onset obesity compared to unscreened population controls<sup>7, 8</sup>. This deletion was also implicated in developmental delay, intellectual disability, ASD and defined as well-replicated schizophrenia-associated CNV locus<sup>2, 9, 10</sup>. However, the clinical significance of the reciprocal duplication of the same interval was never fully assessed.

As a preliminary study, we recently assessed the phenotypic features associated with the 16p11.2 220 kb BP2-BP3 deletion and reciprocal duplication, by collecting clinical data on 137 unrelated carriers (88 deletions and 49 duplications). We compared their BMI and HC to gender-, age- and geographical location-matched reference populations (Loviglio et al., under review; see **Chapter 1**). The BMI mean Z-score of deletion carriers differed significantly from that of the general population (t-test,  $P=3.1e^{-14}$ ), confirming the aforementioned association of the deletion with obesity<sup>8, 11</sup>. Similarly, we observed a trend towards an increased HC in deletions. The duplication carriers showed a mirroring decrease of BMI and HC values when compared to those of the control population (t-test,  $P=0.005$  and  $1.1e^{-4}$ , respectively). Our results indicate that the 16p11.2 220 kb deletion and reciprocal duplication oppositely affect growth parameters, with an early-onset in childhood, and a strong gender-bias toward male carriers (unpublished data). We also observed an increase in ASD prevalence in both deletion (23/88; 26%) and duplication (11/49; 22%) carriers compared to the general population (5,338/363,749)<sup>12</sup>(Fisher's enrichment test: OR=23.7,  $P=2.5e^{-22}$ ; OR=19.4,  $P=1.2e^{-10}$ ). Thus, genomic rearrangements at both 600 kb BP4-BP5 and 220 kb BP2-BP3, two loci 650 kb apart present similar

clinical patterns: large effect sizes on BMI and HC, as well as association with ASD. These rearrangements represent a unique opportunity to study the etiology of neuropsychiatric and energy balance disorders and their underlying pathophysiological pathways.

The 220 kb interval encompasses nine genes: *ATXN2L*, *TUFM*, *SH2B1*, *ATP2A1*, *RABEP2*, *CD19*, *NFATC2IP*, *SPNS1* and *LAT*. *SH2B1* was suggested to be a crucial candidate for the obesity phenotype as it encodes a Src homology adaptor protein involved in leptin and insulin signaling<sup>13, 14</sup>. Common variants in this locus were repeatedly associated with BMI, serum leptin and body fat in genome-wide association studies (GWAS)<sup>15-18</sup>, whereas rare dominant mutations were reported to cause obesity, social isolation, aggressive behavior and speech and language delay<sup>19</sup>. It is however unclear whether the complex phenotype associated with the 16p11.2 220 kb rearrangements are resulting from imbalance of one or several of the nine gene(s) mapping to this genomic interval.

The zebrafish embryo is emerging as a powerful *in vivo* model in translational medicine and notably in neuroscience and neurogenetics due to the high evolutionary conservation of key genes and pathways in this animal model<sup>20</sup>. HC defects observed at birth or during infancy is a key diagnostic feature for ASD and other major neurological disorders<sup>21</sup>. In such instance, the measurement of the head size of zebrafish embryo is a powerful proxy to identify causal genes for micro- and macrocephaly, often associated with disease in humans<sup>22</sup>, as well as a valuable alternative to the generation of mouse models, where systematic engineering of multiple genes is more expensive and time-consuming even in the burgeoning CRISPR-CAS9 era. Golzio et al. developed a surrogate measurement for head size at 4.25–4.5 d.p.f. using objective measurements, with the distance across the convex tips of the eye cups recorded blindly in 50 embryos per injection<sup>23</sup>.

Given the strong association between the 16p11.2 220 kb BP2-BP3 CNVs and head size defects and ASD and the above-described successes, we used zebrafish embryos as an *in vivo* model to identify the causal gene(s) for the microcephaly phenotype.

## Materials and Methods

### In vivo analysis of gene expression and embryo manipulations

For overexpression experiments, the human wild-type mRNAs (*CD19* - NM\_001178098, *NFATC2IP* - NM\_032815, *ATXN2L* - NM\_007245, *TUFM* - NM\_003321, *ATP2A1* - NM\_004320, *RABEP2* - NM\_024816, *SPNS1* - NM\_032038, *LAT* - NM\_001014987, *SH2B1* - NM\_001145795, *CD247* isoform1 - NM\_000734, *CD247* isoform2 - NM\_198053, *ZAP70* isoform1 - NM\_001079, *ZAP70* isoform2 - NM\_207519) were cloned into the pCS2 vector and transcribed *in vitro* using the SP6 Message Machine kit (Ambion). We injected 1 nl of diluted RNA (50, 100 or 150 ng) into wild-type zebrafish embryos at the 1- to 2-cell stage. Injected embryos were scored at 4.25 d.p.f. for the head size and classified into two groups, normal and mutant, on the basis of the relative head size compared with age-matched controls from the same clutch. All the experiments were repeated three times and a *t*-test was performed to determine the significance of the morphant phenotype.

### Zebrafish whole-mount immunostaining

Embryos were fixed in 4% PFA overnight and stored in 100% methanol at -20 °C. For acetylated tubulin staining, embryos were fixed in Dent's fixative (80% methanol, 20% dimethylsulphoxide (DMSO)) overnight at 4 °C. The embryos were permeabilized with proteinase K, then postfixed with 4% PFA, washed in PBSTX (PBS+0.5%, Triton X-100). After rehydration in PBS, PFA-fixed embryos were washed in IF buffer (0.1% Tween-20, 1% BSA in PBS 1×) for 10 min at room temperature. The embryos were incubated in the blocking buffer (10% FBS, 1% BSA in PBS 1×) for 1 h at room temperature. After two washes in IF Buffer for 10 min each, embryos were incubated in the first antibody solution, 1:750 anti-histone H3 (ser10)-R, (sc-8656-R, Santa Cruz), 1:1000 anti-HuC/D (A21271, Invitrogen), 1:1000 anti-acetylated tubulin (T7451, Sigma-Aldrich), in blocking solution, overnight at 4 °C. After two washes in IF Buffer for 10 min each, embryos were incubated in the secondary antibody solution, 1:1000 Alexa

Fluor donkey anti-rabbit IgG and Alexa Fluor goat anti-mouse IgG (A21207, A11001, Invitrogen), in blocking solution, for 1 h at room temperature. For the anti-H3 protocol, staining levels were tested by counting positive cells in defined regions of the head using the ImageJ software.

## Results

### *In vivo* testing of 16p11.2 220 kb BP2-BP3 CNV genes

We first assessed if the nine genes encompassed within the 16p11.2 220kb BP2-BP3 region, i.e. *CD19*, *NFATC2IP*, *ATXN2L*, *TUFM*, *ATP2A1*, *RABEP2*, *SPNS1*, *LAT* and *SH2B1*, had orthologous genes in the zebrafish genome by performing reciprocal BLAST (basic local alignment search tool). Seven genes (*NFATC2IP*, *ATXN2L*, *TUFM*, *SH2B1*, *ATP2A1*, *RABEP2* and *SPNS1*) have orthologs in zebrafish, all mapping in a syntenic locus on chromosome 3 in single copy, with the exception of *atp2a1*, which has a paralog, *atp2a1l* (89% identical), on chromosome 12. The remaining two genes, *CD19* and *LAT* appear to be absent from the zebrafish genome, even if a sequence similar to *LAT* (27% identical at the amino acid level) is present in the syntenic chromosome 3 region between the genes *atxn2l* and *spns1*. These two genes are involved in the immune response of B- and T-cells, respectively. *CD19* acts as a co-receptor to amplify signals from the pre-B-cell antigen receptor (BCR) thus modulating B-cell fate decisions at multiple stages of development<sup>24</sup>. *LAT* is a membrane-associated adaptor protein which, upon stimulation of the T-cell receptor (TCR), selectively induces pathways critical for T-cell activation and adaptive immune response<sup>25, 26</sup>.

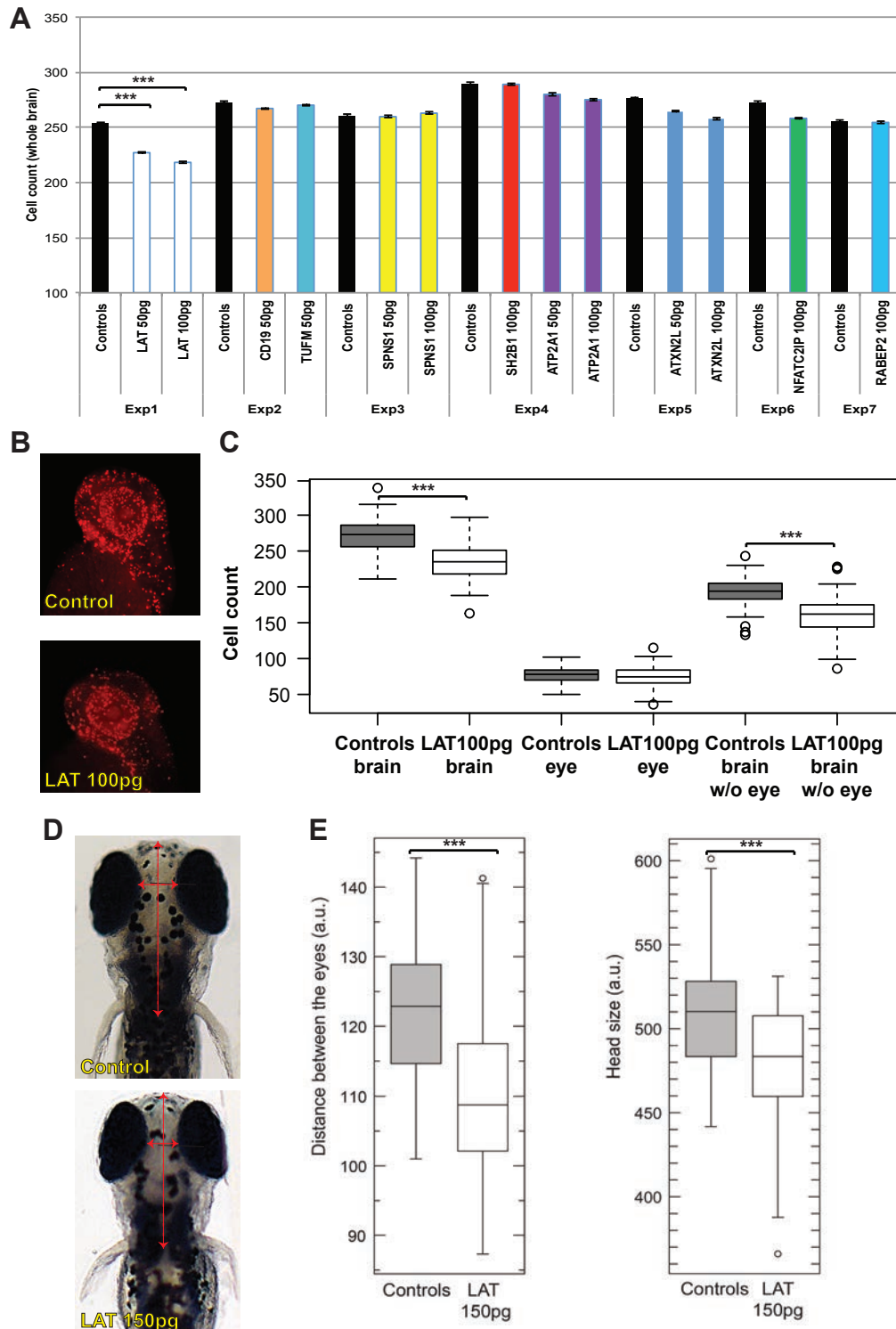
Similarly to the classic *Drosophila* misexpression experiments, we generated capped messenger RNA for these human genes and modeled the duplication by overexpressing each individual human transcript in zebrafish embryos. To achieve substantial overexpression above the baseline of any single transcript, we typically used an injection amount, corresponding to 0.25–0.5% of the total polyA mRNA found in a zebrafish embryo<sup>27</sup>. Previous studies confirmed the persistence of the injected human mRNAs in zebrafish up to 4.5 dpf<sup>23</sup>. To test the

effect of the overexpression of each transcript, we injected separately 50 or 100 pg of RNA encoding each of the nine candidate genes into the zebrafish yolk at the single- or two-cell stage. We did not observe toxicity, lethality or gross morphological defects upon injection. Next, we determined the level of cell proliferation in the zebrafish head by immunostaining with an anti-phospho-histone H3 antibody, a M-phase marker (**Figure 1A**). Injection of the linker for activation of T-cells (*LAT*) messenger resulted in a decreased number of proliferating head cells at 2 dpf (two-tailed t-test  $P=2.6E^{-17}$ ) (**Figure 1B-C**); in contrast, the proliferating cells count of the embryos injected with the eight other transcripts was not significantly different from that of control embryos (**Figure 1A**). Furthermore, the observed effect was dosage-dependent, since increasing amounts of *LAT* mRNA (50 and 100pg) resulted in an increased reduction of cell proliferation (**Figure 1A**). This decrease was specific to the brain (two-tailed t-test  $P=5.03E^{-16}$ ) as no effect was detected in the eye ( $P=0.12$ ) (**Figure 1B-C**). Such zebrafish phenotype is often associated with microcephaly at later developmental stages (e.g. *KCTD13*, *AUTS2* and *BTG2*)<sup>23, 28, 29</sup>. To further assess this possibility, we measured the head size at 4.25–4.5dpf in control and *LAT* mRNA-injected embryos. We observed that increasing amounts of *LAT* mRNA injected into zebrafish embryos (100 and 150pg) yielded a significant increase in the percentage of embryos with microcephaly at this stage (**Figure S1, Figure 1D-E**) (measurement of the distance between the eyes in *LAT*-150pg RNA-injected embryos compared to controls, two-tailed t-test  $P=2.96E^{-8}$ ; measurement of the head size, two-tailed t-test  $P=3.14E^{-5}$ ).

#### Genetic interaction between *LAT* and the 16p11.2 600kb BP4-BP5 CNV driver *KCTD13* and modifiers *MVP/MAPK3*

Manipulation of *LAT* expression leads to head size differences mimicking the HC phenotype observed in the carriers of the 16p11.2 220 kb BP2-BP3 rearrangements. To assess the possible additional contribution of the other genes mapping within this genomic interval, we re-injected *LAT* with each of the remaining transcripts encompassed by the BP2-BP3 region in pairwise combination and determined the number of proliferating cells in the brain by

anti-phospho H3 immunostaining at 2 dpf. We did not observe any significant change in the expressivity (percentile changes in mean count of stained cells) of the phenotype driven by *LAT* injected alone (**Figure 2A**), suggesting that *LAT* is the major and lone phenotypic driver of the BP2-BP3 CNV and that none of the other genes from the BP2-BP3 act as modifiers for the head size phenotype.



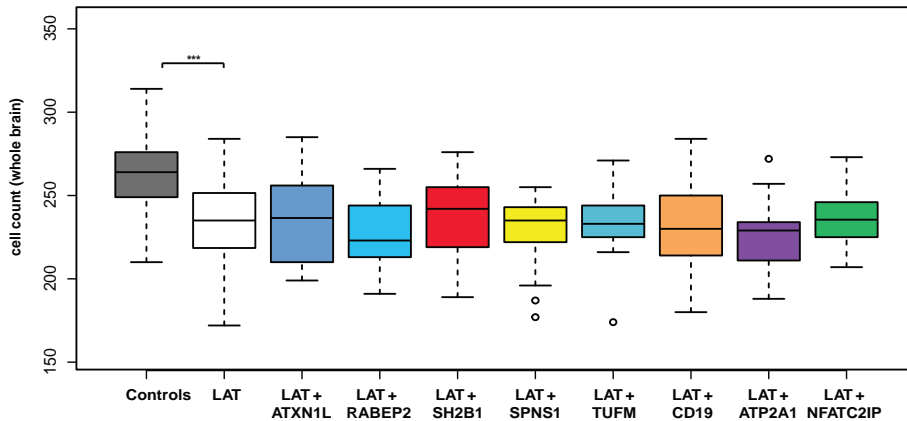


**Figure 1: LAT overexpression leads to brain-specific reduction in cell proliferation and decreased head size.** (A) Bar graph of phospho-histone H3 staining. Quantification of proliferating cells in the brain of controls and RNA-injected embryos at 2dpf using ImageJ software (<http://imagej.nih.gov/ij/>). All transcripts were tested with at least one dosage (50pg or 100pg or, whenever possible, both). Average N=20 for each subgroup. (B-C) Phospho-histone H3 staining for proliferating cells and boxplot representation of its ImageJ quantification (in the whole brain, eye and brain excluding the eye area) of control and LAT-overexpressing embryos at 2dpf. Average N=80 for each subgroup. (D-E) Dorsal views of representative control and LAT 150pg-overexpressing embryos (from top to bottom) and boxplot representation of the measurements (expressed in arbitrary units, a.u.) of the distance between the eye-cups and head size at 4.25dpf. Red arrows in (D) highlight the measured distances. Average N=60 per injection. Significance calculated by two-tailed *t*-test comparisons between control and mRNA-injected embryos (\* if  $p \leq 0.05$ , \*\* if  $p \leq 0.01$ , \*\*\* if  $p \leq 0.001$ ).

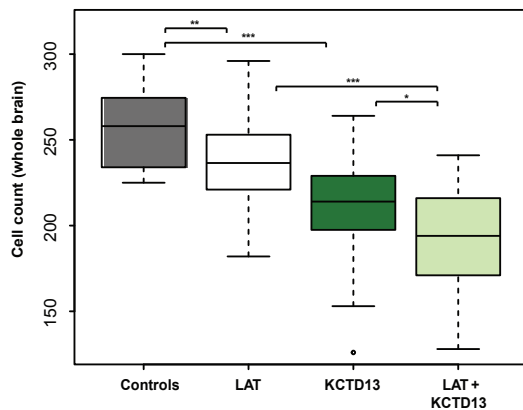
Less than 1Mb apart and proximal to the 220 kb locus, the 16p11.2 600 kb BP4-BP5 CNVs exhibits phenotypes similar to those found in BP2-BP3 CNV carriers, i.e. mirror effect on BMI and head size, and association with ASD and schizophrenia<sup>1-6</sup>. The two CNVs have also been shown to interact preferentially at the chromatin level, by 4C, FISH, Hi-C and concomitant expression changes<sup>30</sup>(Loviglio et al., under review, **Chapter 1**). To investigate whether the two regions interact also genetically, with the candidate drivers of the HC phenotype of the two regions (*LAT* for the BP2-BP3 CNVs and *KCTD13*, candidate driver and *MAPK3* and *MVP*, modifier genes for BP4-BP5 CNVs), we co-injected *LAT* with mRNAs encoding *KCTD13*, *MAPK3* or *MVP* and evaluated the number of proliferating cells with phospho-H3 staining. Single injections of *LAT* and *KCTD13* led to a decreased head size ranging from 10% to 18% respectively. In combination, the severity of the phenotype increased to 25% when *LAT* was co-injected with *KCTD13*, 20% with *MAPK3* and 22% with *MVP* (**Figure 2B-C**). Of note, the two modifiers of the BP4-BP5 CNV, *MAPK3* and *MVP* do not drive any significant reduction in cell proliferation when overexpressed alone as shown previously<sup>23</sup>. Taken together, these results suggest that *LAT* and *KCTD13* interact genetically and are influenced by the same modifier genes. They also indicate

that rearrangements that encompass both CNVs at 16p11.2, i.e. from BP2 to BP5, would produce a more detrimental effect on head size.

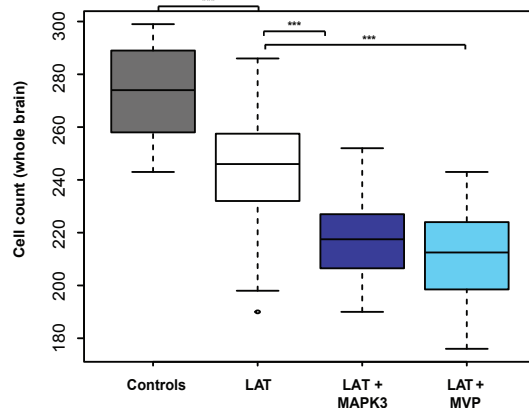
**A**



**B**



**C**



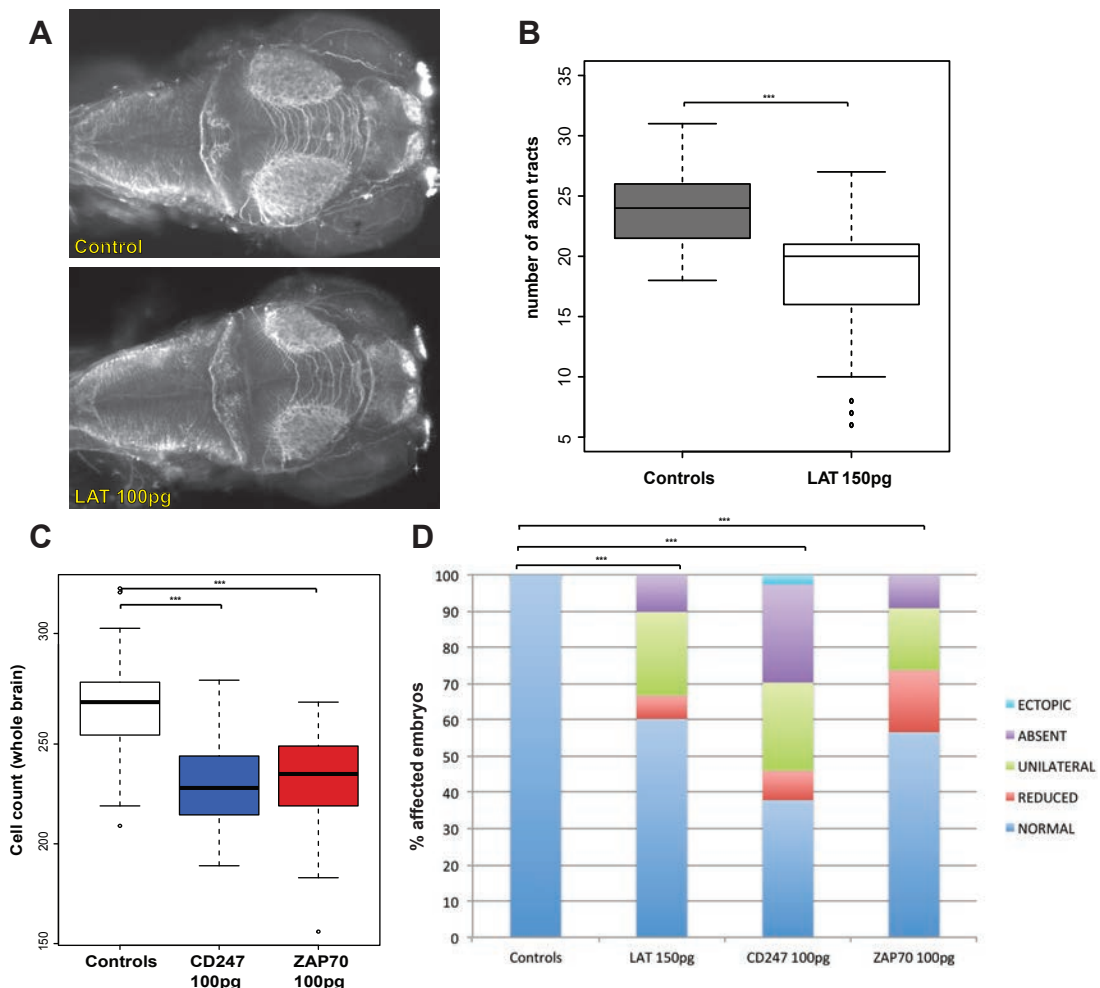
**Figure 2. Pairwise injections of *LAT* mRNA and the driver or modifier transcripts encoded by the 16p11.2 600 kb CNV genes negatively impact cell proliferation in the brain.** Bar graph of phospho-histone H3 staining quantification of proliferating cells in the zebrafish brain of 2dpf control embryos, and embryos injected with *LAT* and pairwise injections of *LAT* with each of the 8 remaining transcripts in the 220 kb region (**A**), *LAT*, *KCTD13* e pairwise injection of the two (**B**), *LAT*, and pairwise injection with *KCTD13*'s modifiers *MVP*/*MAPK3* (**C**) using ImageJ software (<http://imagej.nih.gov/ij/>). All transcripts were tested at a dosage of 100pg. Average N=40 for each subgroup. Significance calculated by two-tailed *t*-test comparisons between control and mRNA-injected embryos (\* if  $p \leq 0.05$ , \*\* if  $p \leq 0.01$ , \*\*\* if  $p \leq 0.001$ ).

We further characterized the neuroanatomical phenotypes induced by the overexpression of *LAT* by looking at possible defects in neuron morphology and/or neurogenesis that could contribute to the onset of the microcephalic phenotype. Acetylated tubulin staining on 3dpf embryos injected with 150pg of *LAT* mRNA revealed a significant reduction in the number of axon tracts projecting from the optic tecta compared to controls (two-tailed t-test  $P=4.4E^{-14}$ ) (**Figure 3A-B**). We hypothesized that the observed phenotypes could be driven also by changes in the number of cells in the developing brain. To test this possibility, we counted HuC/D-positive cells (a marker for post-mitotic neurons). As shown in **Figure 3D**, we observed significant differences in the forebrain, showing an over-representation of unilateral, reduced and absent HuC/D protein levels in the embryos injected with *LAT* 150pg compared to control ones (Fisher's exact test  $P=1.3E^{-14}$ ). These results suggest that *LAT* might likely cause the onset of microcephaly by affecting early neurogenesis and cell projection's organization in the zebrafish brain.

#### *LAT* induces comparable neuron morphology and maturation defects as its TCR signal transduction partners *CD247* (*CD3 $\zeta$* ) and *ZAP70* in the zebrafish embryos

Recently, studies have addressed whether immune molecules have critical functions in the developing brain. For instance, a previously uncharacterized role in regulating early neuronal morphogenesis was assigned to *CD247* (a.k.a. (also known as) *CD3 $\zeta$* ) and *ZAP70*<sup>31-33</sup>. Both genes encode for proteins belonging to the T-cells receptor's signaling module, similarly to *LAT*. We observed that the overexpression of the *CD247* and *ZAP70* in zebrafish embryos recapitulates the phenotype observed upon *LAT* overexpression, i.e. the decreased cell proliferation in the brain compared to control embryos (**Figure 3C**). Of note, the neuronal defect is driven by *ZAP70* isoform 1 but not isoform 2, that differs from the canonical one, at the protein level, for missing amino acids 1-307, whereas both *CD247* isoforms produce a significant decrease in PH3-stained cells (see **Methods** and **Figure S2**). Conversely, HuC/D level assessment in the forebrain of embryos injected with *CD247* isoform 1 and *ZAP70* isoform 1 reproduced the

same reduction observed upon injection of *LAT*. Furthermore, we observed a trend of increased severity of *CD247* compared to the other two transcripts, both in the HuC/D and PH3-staining experiments (**Figure 3D**). Taken together, these studies on *LAT* partners confirmed that perturbation of the signaling pathway mediated by *LAT* is necessary for the proliferation and differentiation of neurons in the developing brain. The manipulation of the expression of the three complex partners *LAT*, *ZAP70*, and *CD247* in zebrafish confirms the function of this immune T-cell complex in neurogenesis, a function that was hitherto unknown.



**Figure 3. The overexpression of *LAT* and its signalosome partners *CD247* and *ZAP70* affects neuron morphology and maturation.** Dorsal views (**A**) of control and *LAT* mRNA-injected embryos at 2dpf stained with anti-acetylated

tubulin (AcTub). **(B)** Boxplots of inter-tecta axonal tracts' count after acetylated Tubulin staining of 3dpf control embryos and embryos injected with 150pg of *LAT*. Average N=80 for each subgroup. **(C)** Boxplots of phospho-histone H3 staining quantification of proliferating cells in the zebrafish brain of 2dpf control embryos, and embryos injected with *CD247* isoform1 and *ZAP70* isoform1. Average N=60 for each group. **(D)** Percentage of embryos with normal bilateral HuC/D protein levels (blue) or unilateral HuC/D (green), ectopic (light blue), and absent (purple)/reduced protein levels (red) in the anterior forebrain in embryo batches injected with *LAT* 150pg, *CD247* 100pg and *ZAP70* 100pg mRNAs. HuC/D levels in the anterior forebrain of the embryo injected with all three mRNAs are considerably decreased compared to those of the control embryo. Significance was estimated by two-tailed *t*-test comparisons between control and mRNA-injected embryos (panels **B** and **C**); enrichment of "abnormal" pattern of HuC/D staining in the injected embryos versus controls was calculated by fisher's exact test **(D)**(\* if  $p \leq 0.05$ , \*\* if  $p \leq 0.01$ , \*\*\* if  $p \leq 0.001$ ).

## Discussion

Growing evidence points to the intimate connection existing between the immune and nervous systems. A reciprocal influence has been documented, with different metabolites provided by the brain, as well as direct stimulation of lymphoid organs, and immune system influencing brain's function through multiple mechanisms<sup>34</sup>. Disrupting the fine regulation between circulating immune cells, macrophages, microglia and neurons or the imbalance in immune molecules can cause a pro-inflammatory skew and produce changes in neuronal function<sup>35, 36</sup>. This effect is not limited to the fetal period since neurodevelopment continues in infancy<sup>37</sup>. In particular, susceptibility to ASD can be substantially influenced through perturbation of these mechanisms. ASD regroups diseases with different etiologies, one of which is the presence of environmental risk factors during pregnancy, followed by immunoneuroendocrine response from the mother (a.k.a MIA (maternal immune activation)) to the developing embryo/fetus<sup>38</sup>. MIA can also be responsible of crude anatomical changes (e.g. enlarged ventricles and small hippocampal volumes) affecting regions that are critical for cognition<sup>39</sup>. T cells have been recognized to play a beneficial role in cognition and behavior<sup>40, 41</sup>. Severe combined immunodeficiency (SCID) seen in *Rag1*<sup>-/-</sup> and *Rag2*<sup>-/-</sup> mice (lacking T

and B cells), nude mice (lacking mature T cells) and OTII mice (whose T cells fail at recognizing self-antigens) impact learning and memory processes<sup>41-45</sup>. Although it is well known that the communication between the immune cells and the central nervous system (CNS) occurs within the meningeal compartment, the recent discovery of CNS lymphatic vessels strongly supports the possibility that immune system dysfunction might be associated with brain disorders<sup>46</sup>. Moreover, several immune-related molecules have been shown to have pleiotropic functions in the brain that can directly influence synaptic function<sup>47, 48</sup>.

In this report, we present the *in vivo* functional characterization of the nine genes mapping at 16p11.2 220kb BP2-BP3 locus, whose CNVs are linked with mirror effect on BMI and HC and associated with ASD and schizophrenia (Loviglio et al., under review, **Chapter 1**). The assessment of the functional effects of overexpression of each genes showed that the linker for activation of T cells, *LAT*, is a candidate gene driver for the head size phenotypes seen in carriers of the 220kb rearrangements. An increase of *LAT* dosage reduces the number of dividing cells in the brain and mature post-mitotic neurons in the anterior forebrain, resulting at later stages in microcephaly. Such reduction of the head size of the zebrafish larvae recapitulates the phenotype presented by the human carriers of the duplication. Overexpression of *LAT* affects the count of axonal projections between the optic tecta, a result that is particularly intriguing in the light of recent findings about two genes belonging to the same pathway, *CD247* (*CD3ζ*) and *ZAP70*. These genes are a prototypical example of the increasing number of immunoreceptors found to have a role not only in immunity-related processes, but also in brain development, synaptogenesis and/or behavior. The transmembrane adaptor signaling protein *CD3ζ*, initially characterized only in T lymphocytes and natural killer (NK), similarly to *LAT*, is part of the CD3 complex, the signaling module of the T-cell receptor responsible for antigenic recognition<sup>49</sup>. This complex participates to the recruitment and activation of downstream effector molecules, including the kinase *ZAP70*, which in turns phosphorylates tyrosine residues present in the intracytoplasmic segments of *LAT* and *CD6*, both representing signal-amplification and diversification

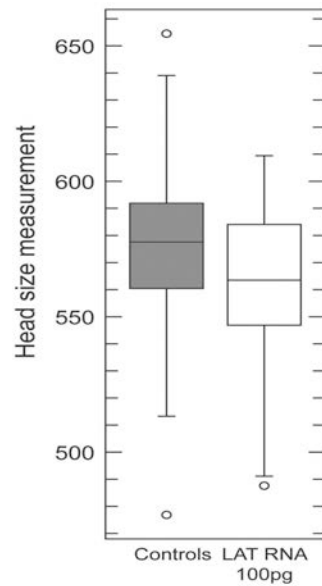
modules<sup>50</sup>. However, *CD247* contribution to dendritic arborization, synaptic plasticity and early neuronal morphogenesis through *ZAP70/Syk* kinase activation by ephrineA-activated pathway was highlighted in recent publications<sup>31, 33, 51, 52</sup>, confirming the hypothesis of a novel role for both genes in the nervous system. Their contribution to dendritogenesis and neuritogenesis, assessed in mice and cultured cells, provides a functional hypothesis about the impact of *LAT* overexpression on cell projections in zebrafish. *CD247*<sup>-/-</sup> mice show brain morphological defects (<http://www.informatics.jax.org/marker/MGI:88334>), in particular reduced glutamatergic synaptic activity in the retina and synaptic plasticity in the hippocampus<sup>53</sup>, and exhibited impaired learning and memory, with a T-cell-independent mechanism<sup>52</sup>. In zebrafish, *ZAP70* is expressed ubiquitously in early development, and highly and specifically in the head from the 16-somites stage, with low expression level throughout the rest of the embryo<sup>54</sup>. Given the absence of a defined *LAT* ortholog in zebrafish, that prevented us to suppress the gene expression by antisense morpholino or by CRISPR/Cas9, we took advantage of the presence of orthologs of the two partners of *LAT*, *CD247* and *ZAP70*, to assess whether their overexpression could recapitulate the phenotypes induced by *LAT* in the zebrafish embryo. Similar to *LAT*, we observed a negative effect on cell proliferation and neuronal maturation, and *CD247* overexpression exhibited the most severe phenotype possibly due to the fact that *CD247* is the most upstream component in the pathway mediated by *LAT*.

Further functional analyses are required to elucidate whether the phenotype we observed is due to a T-cell dependent mechanism or to a novel function of *LAT* in the CNS, a compartment where this gene has never been studied, since its expression was thought to be exclusive to T-cells, mast cells, NK cells and megakaryocytes<sup>55</sup>. However, the GenePaint expression database (<http://www.genepaint.org>) reports that *LAT* is also expressed in the upper layer of the cortical plate at E14.5, a region known to be implicated in neuronal maturation<sup>56</sup>, although at substantially lower level compared to thymus. Low expression levels in brain structures, in particular the cerebellum, are also reported in human in GTEX (<http://gtexportal.org/home/>)<sup>57</sup> and

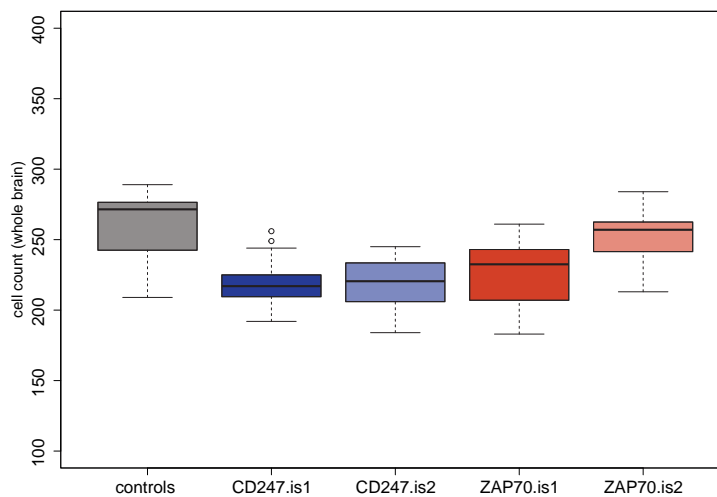
GeneInvestigator (<https://geneinvestigator.com/gv/>). Furthermore, the phenotypes “neuronal loss in central nervous system” and “cognitive impairment” are the top two phenotypes inferred for *LAT* in Monarch (<http://monarchinitiative.org/>)<sup>58</sup>, suggesting that we cannot exclude a T-cell independent mechanism in the observed phenotype. We therefore propose a new role for linker for activation of T-cells (*LAT*), besides its well-known function in T-cells development, as an important player in the 16p11.2 (BP2-BP3) 220kb CNVs neurodevelopmental phenotypes. Further investigations will be needed to clarify the role of *LAT* in connection with the main driver and modifiers of the head size phenotype for the 600kb BP4-BP5 rearrangements at 16p11.2, possibly implying the existence of genetic interaction (see **Thesis discussion**). The “co-localization” at the same cytoband of a cluster of genes participating to convergent pathways relevant to brain phenotypes suggests a new way to look at the 16p11.2 220kb and 600kb CNV regions, as intimately “connected loci” rather than completely independent entities.



## Supplementary Figures



**Figure S1. Measurement (expressed in arbitrary units, a.u.) of the head size of control uninjected embryos and LAT 100pg-overexpressing embryos at 5dpf. Average N=60 per injection. One-tailed t-test P= 0.048**



**Figure S2. Boxplots of phospho-histone H3 staining quantification of proliferating cells in the zebrafish brain of 2dpf control embryos, and embryos injected with *CD247* isoform1 and 2, and *ZAP70* isoform1 and 2.**

## References

1. Zufferey F, Sherr EH, Beckmann ND, Hanson E, Maillard AM, Hippolyte L *et al.* A 600 kb deletion syndrome at 16p11.2 leads to energy imbalance and neuropsychiatric disorders. *J Med Genet* 2012; **49**(10): 660-668.
2. Cooper GM, Coe BP, Girirajan S, Rosenfeld JA, Vu TH, Baker C *et al.* A copy number variation morbidity map of developmental delay. *Nature genetics* 2011; **43**(9): 838-846.
3. McCarthy SE, Makarov V, Kirov G, Addington AM, McClellan J, Yoon S *et al.* Microduplications of 16p11.2 are associated with schizophrenia. *Nat Genet* 2009; **41**(11): 1223-1227.
4. Green EK, Rees E, Walters JT, Smith KG, Forty L, Grozeva D *et al.* Copy number variation in bipolar disorder. *Molecular psychiatry* 2016; **21**(1): 89-93.
5. Jacquemont S, Reymond A, Zufferey F, Harewood L, Walters RG, Kutalik Z *et al.* Mirror extreme BMI phenotypes associated with gene dosage at the chromosome 16p11.2 locus. *Nature* 2011; **478**(7367): 97-102.
6. Maillard AM, Ruef A, Pizzagalli F, Migliavacca E, Hippolyte L, Adaszewski S *et al.* The 16p11.2 locus modulates brain structures common to autism, schizophrenia and obesity. *Molecular psychiatry* 2015; **20**(1): 140-147.
7. Bachmann-Gagescu R, Mefford HC, Cowan C, Glew GM, Hing AV, Wallace S *et al.* Recurrent 200-kb deletions of 16p11.2 that include the SH2B1 gene are associated with developmental delay and obesity. *Genetics in medicine : official journal of the American College of Medical Genetics* 2010; **12**(10): 641-647.
8. Bochukova EG, Huang N, Keogh J, Henning E, Purmann C, Blaszczyk K *et al.* Large, rare chromosomal deletions associated with severe early-onset obesity. *Nature* 2010; **463**(7281): 666-670.
9. Barge-Schaapveld DQ, Maas SM, Polstra A, Knegt LC, Hennekam RC. The atypical 16p11.2 deletion: a not so atypical microdeletion syndrome? *American journal of medical genetics Part A* 2011; **155A**(5): 1066-1072.
10. Guha S, Rees E, Darvasi A, Ivanov D, Ikeda M, Bergen SE *et al.* Implication of a rare deletion at distal 16p11.2 in schizophrenia. *JAMA psychiatry* 2013; **70**(3): 253-260.
11. Walters RG, Coin LJ, Ruukonen A, de Smith AJ, El-Sayed Moustafa JS, Jacquemont S *et al.* Rare genomic structural variants in complex disease: lessons from the replication of associations with obesity. *PloS one* 2013; **8**(3): e58048.
12. CDC. Prevalence of Autism Spectrum Disorder Among Children Aged 8 Years — Autism and Developmental Disabilities Monitoring Network, 11 Sites, United States, 2010. *MMWR Surveill Summ*, vol. 632014, pp 1-21.

13. Duan C, Li M, Rui L. SH2-B promotes insulin receptor substrate 1 (IRS1)- and IRS2-mediated activation of the phosphatidylinositol 3-kinase pathway in response to leptin. *The Journal of biological chemistry* 2004; **279**(42): 43684-43691.
14. Ren D, Zhou Y, Morris D, Li M, Li Z, Rui L. Neuronal SH2B1 is essential for controlling energy and glucose homeostasis. *The Journal of clinical investigation* 2007; **117**(2): 397-406.
15. Willer CJ, Speliotes EK, Loos RJ, Li S, Lindgren CM, Heid IM *et al.* Six new loci associated with body mass index highlight a neuronal influence on body weight regulation. *Nat Genet* 2009; **41**(1): 25-34.
16. Thorleifsson G, Walters GB, Gudbjartsson DF, Steinthorsdottir V, Sulem P, Helgadottir A *et al.* Genome-wide association yields new sequence variants at seven loci that associate with measures of obesity. *Nat Genet* 2009; **41**(1): 18-24.
17. Speliotes EK, Willer CJ, Berndt SI, Monda KL, Thorleifsson G, Jackson AU *et al.* Association analyses of 249,796 individuals reveal 18 new loci associated with body mass index. *Nat Genet* 2010; **42**(11): 937-948.
18. Jamshidi Y, Snieder H, Ge D, Spector TD, O'Dell SD. The SH2B gene is associated with serum leptin and body fat in normal female twins. *Obesity (Silver Spring)* 2007; **15**(1): 5-9.
19. Doche ME, Bochukova EG, Su HW, Pearce LR, Keogh JM, Henning E *et al.* Human SH2B1 mutations are associated with maladaptive behaviors and obesity. *The Journal of clinical investigation* 2012; **122**(12): 4732-4736.
20. Stewart AM, Braubach O, Spitsbergen J, Gerlai R, Kalueff AV. Zebrafish models for translational neuroscience research: from tank to bedside. *Trends Neurosci* 2014; **37**(5): 264-278.
21. Courchesne E, Mouton PR, Calhoun ME, Semendeferi K, Ahrens-Barbeau C, Hallet MJ *et al.* Neuron number and size in prefrontal cortex of children with autism. *JAMA* 2011; **306**(18): 2001-2010.
22. Ceol CJ, Houvras Y, Jane-Valbuena J, Bilodeau S, Orlando DA, Battisti V *et al.* The histone methyltransferase SETDB1 is recurrently amplified in melanoma and accelerates its onset. *Nature* 2011; **471**(7339): 513-517.
23. Golzio C, Willer J, Talkowski ME, Oh EC, Taniguchi Y, Jacquemont S *et al.* KCTD13 is a major driver of mirrored neuroanatomical phenotypes of the 16p11.2 copy number variant. *Nature* 2012; **485**(7398): 363-367.
24. Del Nagro CJ, Otero DC, Anzelon AN, Omori SA, Kolla RV, Rickert RC. CD19 function in central and peripheral B-cell development. *Immunologic research* 2005; **31**(2): 119-131.

25. Fuller DM, Zhang W. Regulation of lymphocyte development and activation by the LAT family of adapter proteins. *Immunological reviews* 2009; **232**(1): 72-83.
26. Bartelt RR, Houtman JC. The adaptor protein LAT serves as an integration node for signaling pathways that drive T cell activation. *Wiley interdisciplinary reviews Systems biology and medicine* 2013; **5**(1): 101-110.
27. Takeda H, Matsuzaki T, Oki T, Miyagawa T, Amanuma H. A novel POU domain gene, zebrafish pou2: expression and roles of two alternatively spliced twin products in early development. *Genes & development* 1994; **8**(1): 45-59.
28. Beunders G, Voorhoeve E, Golzio C, Pardo LM, Rosenfeld JA, Talkowski ME *et al.* Exonic deletions in APTS2 cause a syndromic form of intellectual disability and suggest a critical role for the C terminus. *American journal of human genetics* 2013; **92**(2): 210-220.
29. Jordan DM, Frangakis SG, Golzio C, Cassa CA, Kurtzberg J, Task Force for Neonatal G *et al.* Identification of cis-suppression of human disease mutations by comparative genomics. *Nature* 2015; **524**(7564): 225-229.
30. Blumenthal I, Ragavendran A, Erdin S, Klei L, Sugathan A, Guide JR *et al.* Transcriptional consequences of 16p11.2 deletion and duplication in mouse cortex and multiplex autism families. *American journal of human genetics* 2014; **94**(6): 870-883.
31. Baudouin SJ, Angibaud J, Loussouarn G, Bonnamain V, Matsuura A, Kinebuchi M *et al.* The signaling adaptor protein CD3zeta is a negative regulator of dendrite development in young neurons. *Molecular biology of the cell* 2008; **19**(6): 2444-2456.
32. Angibaud J, Baudouin SJ, Louveau A, Nerriere-Daguin V, Bonnamain V, Csaba Z *et al.* Ectopic expression of the immune adaptor protein CD3zeta in neural stem/progenitor cells disrupts cell-fate specification. *Journal of molecular neuroscience : MN* 2012; **46**(2): 431-441.
33. Hatterer E, Benon A, Chounlamountri N, Watrin C, Angibaud J, Jouanneau E *et al.* Syk kinase is phosphorylated in specific areas of the developing nervous system. *Neuroscience research* 2011; **70**(2): 172-182.
34. Filiano AJ, Gadani SP, Kipnis J. Interactions of innate and adaptive immunity in brain development and function. *Brain research* 2015; **1617**: 18-27.
35. Zhan Y, Paolicelli RC, Sforazzini F, Weinhard L, Bolasco G, Pagani F *et al.* Deficient neuron-microglia signaling results in impaired functional brain connectivity and social behavior. *Nature neuroscience* 2014; **17**(3): 400-406.
36. Hsiao EY, Patterson PH. Placental regulation of maternal-fetal interactions and brain development. *Developmental neurobiology* 2012; **72**(10): 1317-1326.

37. Lebel C, Beaulieu C. Longitudinal development of human brain wiring continues from childhood into adulthood. *The Journal of neuroscience : the official journal of the Society for Neuroscience* 2011; **31**(30): 10937-10947.
38. Gottfried C, Bambini-Junior V, Francis F, Riesgo R, Savino W. The Impact of Neuroimmune Alterations in Autism Spectrum Disorder. *Frontiers in psychiatry* 2015; **6**: 121.
39. Li Q, Cheung C, Wei R, Hui ES, Feldon J, Meyer U *et al*. Prenatal immune challenge is an environmental risk factor for brain and behavior change relevant to schizophrenia: evidence from MRI in a mouse model. *PloS one* 2009; **4**(7): e6354.
40. Wolf SA, Steiner B, Akpınarlı A, Kammertoens T, Nassenstein C, Braun A *et al*. CD4-positive T lymphocytes provide a neuroimmunological link in the control of adult hippocampal neurogenesis. *Journal of immunology* 2009; **182**(7): 3979-3984.
41. Kipnis J, Cohen H, Cardon M, Ziv Y, Schwartz M. T cell deficiency leads to cognitive dysfunction: implications for therapeutic vaccination for schizophrenia and other psychiatric conditions. *Proceedings of the National Academy of Sciences of the United States of America* 2004; **101**(21): 8180-8185.
42. Ziv Y, Ron N, Butovsky O, Landa G, Sudai E, Greenberg N *et al*. Immune cells contribute to the maintenance of neurogenesis and spatial learning abilities in adulthood. *Nature neuroscience* 2006; **9**(2): 268-275.
43. Radjavi A, Smirnov I, Kipnis J. Brain antigen-reactive CD4+ T cells are sufficient to support learning behavior in mice with limited T cell repertoire. *Brain, behavior, and immunity* 2014; **35**: 58-63.
44. Brynskikh A, Warren T, Zhu J, Kipnis J. Adaptive immunity affects learning behavior in mice. *Brain, behavior, and immunity* 2008; **22**(6): 861-869.
45. Derecki NC, Cardani AN, Yang CH, Quinlan KM, Crihfield A, Lynch KR *et al*. Regulation of learning and memory by meningeal immunity: a key role for IL-4. *The Journal of experimental medicine* 2010; **207**(5): 1067-1080.
46. Louveau A, Smirnov I, Keyes TJ, Eccles JD, Rouhani SJ, Peske JD *et al*. Structural and functional features of central nervous system lymphatic vessels. *Nature* 2015; **523**(7560): 337-341.
47. Boulanger LM. Immune proteins in brain development and synaptic plasticity. *Neuron* 2009; **64**(1): 93-109.
48. Garay PA, McAllister AK. Novel roles for immune molecules in neural development: implications for neurodevelopmental disorders. *Frontiers in synaptic neuroscience* 2010; **2**: 136.
49. Samelson LE, Harford JB, Klausner RD. Identification of the components of the murine T cell antigen receptor complex. *Cell* 1985; **43**(1): 223-231.

50. Malissen B, Gregoire C, Malissen M, Roncagalli R. Integrative biology of T cell activation. *Nature immunology* 2014; **15**(9): 790-797.
51. Angibaud J, Louveau A, Baudouin SJ, Nerriere-Daguin V, Evain S, Bonnamain V *et al.* The immune molecule CD3zeta and its downstream effectors ZAP-70/Syk mediate ephrin signaling in neurons to regulate early neuritogenesis. *Journal of neurochemistry* 2011; **119**(4): 708-722.
52. Louveau A, Angibaud J, Haspot F, Opazo MC, Thinard R, Thepenier V *et al.* Impaired spatial memory in mice lacking CD3zeta is associated with altered NMDA and AMPA receptors signaling independent of T-cell deficiency. *The Journal of neuroscience : the official journal of the Society for Neuroscience* 2013; **33**(47): 18672-18685.
53. Xu HP, Chen H, Ding Q, Xie ZH, Chen L, Diao L *et al.* The immune protein CD3zeta is required for normal development of neural circuits in the retina. *Neuron* 2010; **65**(4): 503-515.
54. Christie TL, Carter A, Rollins EL, Childs SJ. Syk and Zap-70 function redundantly to promote angioblast migration. *Developmental biology* 2010; **340**(1): 22-29.
55. Zhang W, Sommers CL, Burshtyn DN, Stebbins CC, DeJarnette JB, Tribble RP *et al.* Essential role of LAT in T cell development. *Immunity* 1999; **10**(3): 323-332.
56. Meyer G. Genetic control of neuronal migrations in human cortical development. *Advances in anatomy, embryology, and cell biology* 2007; **189**: 1 p preceding 1, 1-111.
57. Consortium GT. Human genomics. The Genotype-Tissue Expression (GTEx) pilot analysis: multitissue gene regulation in humans. *Science* 2015; **348**(6235): 648-660.
58. Haendel MA, Vasilevsky N, Brush M, Hochheiser HS, Jacobsen J, Oellrich A *et al.* Disease insights through cross-species phenotype comparisons. *Mammalian genome : official journal of the International Mammalian Genome Society* 2015; **26**(9-10): 548-555.

## Chapter 3: Identification of a *RAI1*-associated disease network through integration of exome sequencing, transcriptomics and 3D genomics

### Summary of the contribution

The observations described in this chapter refer to the article “Identification of a *RAI1*-associated disease network through integration of exome sequencing, transcriptomics and 3D genomics”, currently submitted to the journal “Genome Medicine”.

As co-first author of this paper, I contributed to a substantial subset of the analysis and experiments presented in the following sections. Specifically, I participated to the variant pathogenicity assessment; I designed and prepared the 4C-seq libraries, conducted the statistical analysis on the 4C-seq dataset, GO and network analysis and comparison with *Rai1*<sup>-/-</sup> mice transcriptome data.

The results are presented in four main figures (named **Figure 1-4**), six supplementary figures (**Figure S1-6**) and fourteen supplementary tables (**Table S1-14**; **Tables S1-10** are available upon request, **Tables S11-14** are included in the text). In addition, a supplementary text section is present.

## **Identification of a *RAI1*-associated disease network through integration of exome sequencing, transcriptomics and 3D genomics**

Maria Nicla Loviglio<sup>1\*</sup>, Christine R. Beck<sup>2\*</sup>, Janson J. White<sup>2</sup>, Marion Leleu<sup>3,4</sup>, Tamar Harel<sup>2</sup>, Nicolas Guex<sup>4</sup>, Anne Niknejad<sup>4</sup>, Weimin Bi<sup>2</sup>, Edward S. Chen<sup>2</sup>, Isaac Crespo<sup>4</sup>, Jiong Yan<sup>2,8</sup>, Wu-Lin Charng<sup>2</sup>, Shen Gu<sup>2</sup>, Ping Fang<sup>2,9</sup>, Zeynep Coban-Akdemir<sup>2</sup>, Chad A. Shaw<sup>2</sup>, Shalini N. Jhangiani<sup>5</sup>, Donna M. Muzny<sup>5</sup>, Richard A. Gibbs<sup>2,5</sup>, Jacques Rougemont<sup>3,4</sup>, Ioannis Xenarios<sup>1,4</sup>, James R. Lupski<sup>2,5,6,7</sup>, Alexandre Reymond<sup>1</sup>

<sup>1</sup>Center for Integrative Genomics, University of Lausanne, CH-1015 Lausanne, Switzerland

<sup>2</sup>Department of Molecular and Human Genetics, Baylor College of Medicine, Houston, TX 77030, USA

<sup>3</sup>School of Life Sciences, EPFL (Ecole Polytechnique Fédérale de Lausanne), CH-1015 Lausanne, Switzerland:

<sup>4</sup>Swiss Institute of Bioinformatics (SIB), CH-1015 Lausanne, Switzerland

<sup>5</sup>Human Genome Sequencing Center, Baylor College of Medicine, Houston, TX 77030, USA

<sup>6</sup>Department of Pediatrics, Baylor College of Medicine, Houston, TX 77030, USA

<sup>7</sup>Texas Children's Hospital, Houston, TX 77030, USA

Present address:

<sup>8</sup>Department of Pathology and Laboratory Medicine, University of Saskatchewan, Saskatoon, SK S7N 0W8, Canada

<sup>9</sup>WuXiNextCODE, 101 Main Street, Cambridge, MA 02142, USA

\*These two authors contributed equally to this work

Correspondence to

alexandre.reymond@unil.ch

or

jlupski@bcm.edu

**Running Title:** Smith-Magenis syndrome and rare variants

**Keywords:** diagnostic, intellectual disability, chromatin conformation, text mining, disease network



## Abstract

Smith-Magenis syndrome (SMS) is a developmental disability/multiple congenital anomaly disorder resulting from haploinsufficiency of *RAI1*. It is characterized by distinctive facial features, brachydactyly, sleep disturbances and stereotypic behaviors. We investigated a cohort of 15 individuals with a clinical suspicion of SMS, who showed neither deletion in the SMS critical region nor damaging variants in *RAI1*. Potentially deleterious variants were identified in eight of these subjects using whole-exome sequencing. These changes affect *KMT2D*, *ZEB2*, *MAP2K2*, *GLDC*, *CASK*, *MECP2*, *KDM5C* and *POGZ*, known to be associated with Kabuki syndrome 1, Mowat-Wilson syndrome, cardiofaciocutaneous syndrome, glycine encephalopathy, mental retardation and microcephaly with pontine and cerebellar hypoplasia, X-linked mental retardation 13, X-linked mental retardation Claes-Jensen type and White-Sutton syndrome, respectively. Analyses of coexpression and biomedical text mining suggest that these pathologies and SMS are part of the same disease network. Further support for this hypothesis was obtained from transcriptome profiling that showed that the expression levels of both *Zeb2* and *Map2k2* are perturbed in *Rai1*<sup>-/-</sup> mice. As an orthogonal approach to potentially contributory disease gene variants, we used chromatin conformation capture to reveal chromatin contacts between *RAI1* and the loci flanking *ZEB2* and *GLDC*, as well as between *RAI1* and human orthologs of the genes that show perturbed expression in our *Rai1*<sup>-/-</sup> mouse model. These holistic studies of *RAI1* and its interactions allow insights into SMS and other disorders associated with intellectual disability and behavioral abnormalities. Our findings support a pan-genomic approach to the molecular diagnosis of a distinctive disorder.

## Introduction

Smith-Magenis syndrome (SMS; MIM #182290) is a rare genomic disorder with a prevalence of 1 in 15,000. It is associated with specific craniofacial dysmorphism, developmental delay (DD), moderate to profound intellectual disability (ID) and self-injurious and stereotypic behaviors<sup>1, 2</sup>. SMS individuals show sleep disturbance with frequent daytime napping and nighttime awakenings. They display restricted interest and obsessional thinking and social responsiveness scale scores consistent with autism spectrum disorder (ASD)<sup>3</sup>. They repetitively mouth objects, rock, spin or twirl their body and grind their teeth<sup>4</sup>. This distinctive profile is complemented by specific lick and flip and self hug behaviors, as well as attachment to people<sup>5-7</sup>. Challenging behaviors such as self-injuries, physical aggression and destructive behavior are significantly more prevalent in SMS than in ID with mixed etiologies<sup>8</sup>. Self-injuries are present in 70-97% of subjects and include polyembolokoilamania (insertion of foreign objects into bodily orifices) and onychotillomania (pulling finger and toe nails out). Unusual behaviors can comprise poking others' eyes, forceful hugging and punching fists through walls and windows.

Whereas SMS is classically associated with a deletion within cytogenetic G-band 17p11.2 that includes the *RAI1* gene (about 90% of subjects) or a nucleotide variant in that gene (about 5%)<sup>1, 9-12</sup>, some reports suggested genetic heterogeneity as SMS-like individuals were found to recurrently harbor deletions of the 2q37.3 or 2q23.1 cytobands encompassing *HDAC4* and *MBD5*, respectively<sup>13-15</sup>. Similarly, *PITX3* was proposed to be responsible for the SMS-like neurobehavioral abnormalities observed in an individual<sup>16</sup>.

Here we use recent advances in genome sequencing technologies to further assess the genetic heterogeneity of SMS and the possible clinical overlap of this syndrome with other intellectual disability and cognitive dysfunction disorders, as some of the seemingly characteristic phenotypic features are non-discriminating amongst ID syndromes. We also evaluate the pertinence of

network interactions and provide experimental data in support of potential molecular diagnoses.

## Materials and Methods

### Enrollment

The institutional review board of the Baylor College of Medicine (BCM) approved this study. Participants were enrolled in the study after written informed consent was obtained from parents or legal guardians. Each of the 149 patients was clinically assessed by their respective physicians.

BCM patients were diagnosed as potentially affected by SMS through clinical assessment. Briefly, they all presented intellectual disability and/or developmental delay, and typically had sleep disturbances, stereotypies, or other endophenotypes common to SMS (e.g. distinctive facial features, tantrums, self-injurious behaviors, hoarse voice). The clinical presentation of SMS is heterogeneous; therefore, the indication of SMS by a clinician can be either premature in the case of a young infant, or possibly a misdiagnosis in an individual with behavioral issues and ID.

### Detailed SMS patients' phenotypes

The detailed phenotype descriptions of 13 of the 15 patients without *RAI1* genetic alteration are described in **Supplementary text and Supplementary Table S1**. The remaining 2 individuals had no clinical data available.

### aCGH

Targeted chromosome 17p aCGH analyses were carried out on each proband as previously reported<sup>17</sup>. Additional genome-wide aCGH was conducted on each person using Baylor Miraca Genetics Laboratory design version 10.1, an Agilent 180K oligo array. All array data were analyzed as previously described<sup>18</sup>.

## Exome sequencing

To uncover genetic variants associated with the abnormalities shown by the 15 patients without *RAI1* genetic alteration, we performed whole exome sequencing of DNA extracted from blood of the proband and both their parents whenever possible (8 trios) at the BCM Human Genome Sequencing Center via the Baylor-Hopkins Center for Mendelian Genetics. Exomes were captured and sequenced on an Illumina HiSeq platform using previously described methods<sup>19</sup>. Variants were filtered based on inheritance patterns including autosomal recessive, X-linked and *de novo*/autosomal dominant. Variants with MAF<0.05 in control cohorts (ARIC, 1000 genomes, the exome variant server and our internal BCM database of >5000 exomes) and predicted to be deleterious by SIFT10 and/or PolyPhen were prioritized. Sanger sequencing confirmed putatively causative variants and their familial segregation.

## Modeling

The primary sequence of each candidate protein was loaded in Swiss-PdbViewer aligned onto suitable modeling templates retrieved from SWISS-MODEL and superposed in 3D-space using Swiss-PdbViewer<sup>20, 21</sup>. Each variant was modeled in the context of the overall 3D-structure to evaluate its potential impact with respect to protein folding, as well as to position of known disease-associated variants. We also assessed if missense variants perturbing the protein function clustered in 3D around key regions of the protein<sup>22</sup>.

The ZEB2 Zinc finger residues 995-1078 were modeled using the pdb entry 1mey as template<sup>23</sup>. MAP2K2 was modeled using both MAP2K2 (pdb entry 1s9i 3.2A resolution<sup>24</sup> and MAP2K1 (pdb entry 3eqi, 1.9A resolution) structures<sup>25</sup>. The GLDC residues were aligned on the *Synechocystis sp.* glycine decarboxylase model PCC 6803 (pdb entry 4LHD)<sup>26</sup>. To model the CASK variants, two partial CASK crystal structures (pdb entries 1kwa, chain A<sup>27</sup> and 1kgd, chain A [<http://www.ncbi.nlm.nih.gov/pubmed/11729206?dopt=Abstract>]) covering

residues 487-572 and 739-914, respectively, were superposed on the crystal structure of PALS1/Crb (pdb entry 4wsi<sup>28</sup>) that present 35% identity with CASK.

### Literature mining

Because literature resources do not use entity name in a consistent way, we first checked each gene identifier by using UniProtKB (<http://www.uniprot.org>) or HUGO Gene Nomenclature Committee (HGNC) database (<http://www.genenames.org>) in order to retrieve the recommended/approved name, short name(s), alternative and synonymous name(s) if any for each targeted gene, as well as the name(s) of the encoded protein. These were used as singleton and/or pairwise strings to extract information from various literature resources: PubMed (<http://www.ncbi.nlm.nih.gov/pubmed>), Google Scholar, iHOP (<http://www.ihop-net.org/UniPub/iHOP/>) and EVEX (<http://evexdb.org/Gene>), to cite here the original source of reference for this project. The obtained results were curated and the reported relationships were visualized using Cytoscape (3.2.1; <http://www.cytoscape.org/>). The connectivity was assessed using the Knet-function, which is based on the adaptation of spatial statistics concepts to network analysis proposed in <sup>29</sup>. The statistical significance of the obtained Knet-function value was calculated with respect to a population of permuted networks ( $n=10^6$ ) derived from the original prior knowledge network. It is worth noting here that the connectivity is not only based on direct but also on indirect connections through shortest paths.

### Identification of RAI1 Interacting Proteins

We identified ZBTB17/MIZ1 and BRD2 as likely interactors for RAI1 with a yeast two-hybrid assay. Briefly, two fragments of the carboxyl-terminus of mouse *Rai1* (a.a 1246-1841 and a.a. 1246-1890) were cloned into pCWX200 as baits. Around 10 million independent cDNA library clones (10x library coverage) were screened for protein-protein interaction with both baits. From this analysis, we identified ZBTB17/MIZ1, BRD2, and SOGA3 as reasonable candidates (at least two clones, supported by both baits) for RAI1 interaction candidates. These

interactions were further assessed using co-immunoprecipitation (co-IP) analysis in HEK293 cells. Full-length *Rai1* was cloned in pCMV-3xFLAG vector while the three candidates were cloned into pCMV-HA vectors to confirm the yeast two-hybrid results. Lysate from the co-transfected HEK293 cells (RAI1 and one of the candidates) was purified with EZview FLAG-M2 beads (Sigma) and analyzed with rat anti-HA (Abcam) on Western Blot. The interaction between RAI1 and ZBTB17/MIZ1 was confirmed by co-IP, however *BRD2* did not express well enough on Western blot, and *SOGA3* was too sticky to conduct co-IP with, as it bound to the beads in the absence of FLAG-RAI1.

### Embryo collection and RNA extraction

Mice were housed in standard specific pathogen free conditions. All animal studies were conducted under protocols approved by the Baylor Institutional Animal Care and Use Committee and followed NIH guidelines. Timed matings between *Rai1* heterozygous females and males in F2 generation in the C57BL/6 *Tyrc-Brd* and 129SvEv mixed genetic background were implemented to generate *Rai1*<sup>-/-</sup> embryos. To harvest embryos, pregnant females were sacrificed by cervical dislocation and the embryos were dissected from the uterus in ice-cold PBS solution. Similar sized embryos at 10.5 dpc were collected in 1.5 ml eppendorf tubes, frozen immediately in liquid nitrogen, and stored in -80°C. Portions of the yolk sac were saved for genotyping as described previously<sup>30</sup>. For RNA extraction, the whole embryos were homogenized in Trizol and RNA was extracted according to the manufacturer's instruction (Invitrogen) followed by purification on columns using an RNeasy mini kit (Qiagen Sciences, Germantown, MD). The RNA integrity, concentration, and overall quality were tested with an Agilent Bioanalyzer 2100 and a NanoDrop ND-1000 spectrophotometer.

### Microarray processing and analysis

Five to ten micrograms of total RNA from each individual embryo of three *Rai1*<sup>-/-</sup> at 10.5 dpc and three wild type controls were used to produce cRNA target microarray transcriptome analyses. Day 10.5 dpc embryos were chosen because

*Rai1* functions during this stage as indicated by its strong expression and embryonic lethality of *Rai1*<sup>-/-</sup> embryos from 7.5 to 18.5 dpc (Bi et al 2005 HMG). In addition, the size of the *Rai1*<sup>-/-</sup> embryos at 10.5 dpc is comparable to that of their wild type littermates whereas the few surviving *Rai1*<sup>-/-</sup> mice at birth are significantly smaller than the wild type<sup>30</sup>. The integrity and quality of the extracted RNAs were assessed on a 2100 Bioanalyzer (Agilent, Santa Clara, CA). The target was generated using a reverse transcription reaction to produce cDNA (SuperScript Choice System, Gibco), which was subsequently subjected to *in vitro* transcription with biotinylated cytidine-5'-triphosphate and uridine-5'-triphosphate using ENZO BioArray High Yield RNA Transcript Labeling kit to produce biotinylated cRNA. The target was then fragmented and hybridized to Affymetrix Mouse Genome 430 2.0 Array GeneChips (Affymetrix, Santa Clara, CA) in duplicates using an Affymetrix GeneChip Fluidics Station 400. The arrays were stained with phycoerythrin-coupled avidin and scanned using a GeneArray Scanner 3000. The resultant output was analyzed using Affymetrix Microarray Suite software and examined for excessive background or evidence for RNA degradation. The chips were assessed by scaling factor, average background, percent of probe sets that are present, number of probes present and the 3'-end to 5'-end probe intensity ratio for housekeeping probe sets (*b-actin* and *GAPDH*), as well as the number of probes present for the "spiked in" probe sets (*BioB*, *BioC*, *BioD*, and *Crex*). All the chips were of good quality, which is further supported by the observations that they have similar RNA degradation patterns and the chips were well replicated within the same genotype group as shown by scatter plot analyses. The criteria for genes differentially expressed are that the log ratio of the normalized expression values in the *Rai1* deficient embryos versus the controls is > 0.5 and the *P* value < 0.05, which empirically gives a very low false detection rate. The probesets with very low expression values were filtered out. We analyzed the chromosomal position of all the regulated genes using the chromosomal coordinates within recent genome assemblies of the mouse (mm5). The array data were analyzed using the GC-RMA program to estimate the expression measures from the probe level data<sup>31</sup>. The program corrects the background, normalizes the raw perfect match data using the quantile normalization method, and summarizes the probe values to probe set

values (expression values, one per probe set per chip), in  $\log_2$  scale. The fold change for each probe is the log ratio of average expression value in the mutant samples divided by that in the wild type controls. The fold change is considered to be significant if  $P \leq 0.05$ .

#### 4C-seq

4C libraries were prepared from LCLs of two age-matched female control individuals. Briefly, LCLs were grown at 37°C.  $5 \times 10^7$  exponentially growing cells were harvested and crosslinked with 1% formaldehyde, lysed and cut with DpnII, a 4-cutter restriction enzyme that allows higher resolution<sup>32</sup>. After ligation and reversal of the crosslinks, the DNA was purified to obtain the 3C library. This 3C library was further digested with NlaIII and circularized to obtain a 4C library. The inverse PCR primers to amplify 4C-seq templates were designed to contain Illumina adaptor tails, sample barcodes and viewpoint-specific sequences. The selected viewpoint maps within the 5' portion of the first intron of the *RAI1* gene (700bp from the donor site of exon 1), a region enriched in DNaseI hypersensitive and transcription factor binding sites<sup>33</sup>. It corresponds to the closest suitable DpnII fragment relative to the transcriptional start sites of the targeted gene. The sequence of the 4C-seq primers is reported in **Supplementary Table S13**. We amplified at least 1.6  $\mu\text{g}$  of 4C template (using about 100 ng of 4C template per inverse PCR reaction, for a total number of 16 PCRs). We multiplexed the two 4C-seq templates in equimolar ratios and analyzed them on a 100-bp single-end Illumina HiSeq flow cell. The numbers of raw, excluded, and mapped reads for each LCL sample are detailed in **Supplementary Table S14**.

#### 4C-seq data analysis

4C-seq data were analyzed as previously described<sup>32, 34, 35</sup> through the 4C-seq pipeline available at <http://htsstation.epfl.ch/><sup>36</sup> and visualized with gFeatBrowser (<http://www.gfeatbrowser.com>). Briefly, the multiplexed samples were separated, undigested and self-ligated reads removed. Remaining reads



were aligned and translated to a virtual library of DpnII fragments. Read counts were then normalized to the total number of reads and replicates combined by averaging the resulting signal densities (**Supplementary Figure S5**). The local correlation between the profiles of the two samples per viewpoint was calculated (Spearman correlation: 0.83). The combined profiles were then smoothed with a window size of 29 fragments. The region directly surrounding the viewpoint is usually highly enriched and can show considerable experimental variation, thereby influencing overall fragment count. To minimize these effects, the viewpoint itself and the directly neighboring ‘undigested’ fragment were excluded during the procedure. In addition to this filtering, we modeled the data to apply a profile correction similar to the one described in <sup>37</sup> using a fit with a slope -1 in a log-log scale<sup>38</sup>. Significantly interacting regions were detected by applying a domainogram analysis as described<sup>39</sup>. We selected BRICKS (Blocks of Regulators In Chromosomal Kontext) with a p-value threshold smaller than 0.01 for both “cis” and “trans” interactions, and annotated the BRICKs overlapping genes as well as the closest upstream and the closest downstream genes, in a window of +/-500kb.

### Enrichment analyses

Gene annotation was obtained through BioScript (<http://gdv.epfl.ch/bs>). Protein interaction networks for BRICKs genes were determined using STRING (Search Tool for the Retrieval of Interacting Genes/Proteins) v9.1 (<http://string-db.org/>)<sup>40</sup>. We exploited GO with Enrichr (<http://amp.pharm.mssm.edu/Enrichr/>) to assess if the chromatin-contacted genes were enriched in specific pathways and genes associated with Mendelian diseases and GIANT (<http://giant.princeton.edu/>) and Genemania (<http://www.genemania.org/>) to test tissue-specific functional interactions and produce association networks, respectively<sup>41-44</sup>. We used Enrichr Chromosome Location tool and BRICKS count in different window sizes (5Mb, 1Mb and 500kb) to determine whether any cytogenetic band other than 17p11.2 was enriched for BRICKS. Other than 17p11.2, we identified significant enrichments at cytobands 17p12, 17p13 and 2q22, where the gene *ZEB2* is located.

## Hi-C data

Hi-C matrices from Rao *et al.*<sup>45</sup> were prepared by first applying a KR normalization to the 5kb and 100kb resolution observed matrices, and then by dividing each normalized score by the expected one extracted from the KR expected file (as described previously in section II.c of the Extended Experimental Procedures of reference<sup>45</sup>). KR Expected values less than 1 were set to 1 to avoid long-distance interaction biases. HiC matrices from Dixon *et al.*<sup>46</sup> were generated from the normalized datasets at a 40kb resolution and transformed to a 400kb resolution by summing the contacts observed in 10x10 sub-matrices. Expected vectors represent the mean number of contacts observed at a given distance and were used to calculate the observed/expected matrices.

## Results

### Clinical and Molecular findings

Through physicians from a large network of medical genetics centers we enrolled a cohort of 149 individuals presenting with a constellation of SMS features. High-density 17p11.2 array Comparative Genomic Hybridization (aCGH) and Sanger sequencing of *RAI1* showed that 134 out of 149 subjects presented genetic or genomic alteration of the *RAI1* gene<sup>9, 11, 17, 47-50</sup>, 96/134 (72%) individuals carried the classic recurrent 3.7 Mb SMS deletion, 10 (7.5%) contained an Uncommon Recurrent 1 (UR1) or UR2 rearrangement, 24 (18%) a non-recurrent *RAI1* deletion and 4 (3%) had a *de novo* variant in *RAI1*<sup>9, 11, 47, 50, 51</sup> (**Supplementary Table S1**). Whereas these proportions are similar to published results<sup>12, 52</sup>, it is likely that some clinicians did not refer individuals with SMS features who were negative for SMS molecular diagnosis (via aCGH or Fluorescence *In Situ* Hybridization, FISH) or who were positive for another potentially causative CNV, for example 1p36 deletion syndrome<sup>15, 53, 54</sup> that shares multiple similarities with SMS. Indeed, many individuals were molecularly diagnosed prior to sample submission. Consistent with this

hypothesis, a study identified mutations affecting *RAI1* in only 30% of SMS subjects<sup>12</sup>.

The remaining 15 individuals (10%) showed no discernable perturbation of the *RAI1* gene. The 13 with available clinical data presented the following classical SMS features: ID (12/13), DD (13/13), sleep disturbances (8/13) and/or self-injurious behavior (10/13) in particular onychotillomania (7/13) (**Supplementary Text, Supplementary Table S1**). To identify the underlying cause of the phenotypes of these 15 individuals, the probands and their parents when available (8 cases) were subjected to high-resolution genome-wide aCGH and whole exome sequencing. We identified potentially causative variants in nine individuals (**Supplementary Table S1**). These group into five categories: (i) a 47, XYY karyotype (subject BAB2492); (ii) *de novo* variants in *ZEB2* (BAB2386), *CASK* (BAB2540) and *KMT2D* (BAB2319) (iii) compound heterozygote variants in *GLDC* (BAB4947); (iv) a *MECP2* variant in a female with random X-inactivation (BAB2552) inherited from the individual's mother, who presented with a skewed X-inactivation pattern (away from this allele) in her blood (**Supplementary Figure S1**); and (v) variants in *POGZ* (BAB2330, variant not maternally inherited), *MAP2K2* (BAB2474) and the X-linked *KDM5C* (BAB2293), the origins of which could not be assessed. We confirmed by Sanger sequencing the segregation of sequence variants in available family members.

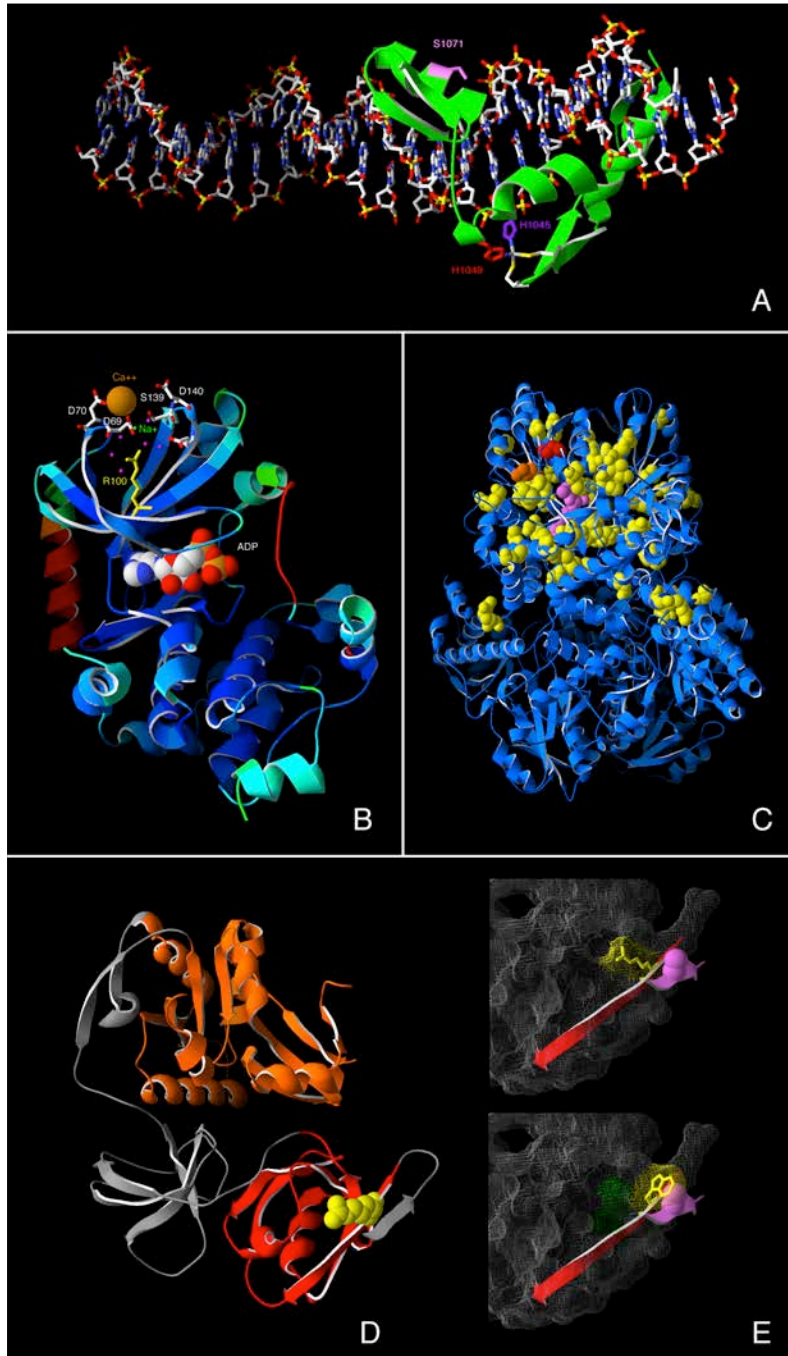
Individual BAB2330 and four other carriers of heterozygous truncating variants in *POGZ* allowed the recent description of a new syndromic form of intellectual disability<sup>55</sup>. We compared the phenotype of the remaining subjects (**Supplementary Text, Supplementary Table S1**) with those associated with the identified molecular diagnoses, including 47,XYY, Mowat-Wilson syndrome (MOWS; OMIM#235730), mental retardation and microcephaly with pontine and cerebellar hypoplasia (MICPCH; OMIM#300749), Kabuki syndrome-1 (KABUK1; OMIM#147920), glycine encephalopathy (GCE; OMIM#605899), X-linked syndromic mental retardation 13 (MRXS13; OMIM#300055), cardiofaciocutaneous syndrome (CFC4; OMIM #615280) and X-linked syndromic mental retardation Claes-Jensen type (MRXSC; OMIM#300534) (**Supplementary**

**Table S2).** While we observed some distinct clinical features (e.g. macrocephaly and seizures in the carriers of variants in *KDM5C* and *GLDC*, respectively (**Supplementary text**)) a sufficient group of features are specific to SMS<sup>56</sup> to hypothesize that in some cases the molecular diagnosis hinted at potential underlying genetic heterogeneity for SMS rather than misdiagnoses of other syndromes and that some of the 47,XYY, MOWS, MICPCH, KABUK1, GCE, MRXS13, CFC4 and MRXSC syndromes have a greater clinical phenotypic variability than anticipated. This prompted investigation of the presumptive effect of the variants on the encoded proteins and molecular perturbations that may underlay the observed phenotypic manifestations.

### Variant analysis and Modeling

The variants identified in *KMT2D* (p.E3418X) and *MECP2* (p.P389fsX) are predicted to be loss-of-function alleles, which are likely pathogenic alleles as *KMT2D* and *MECP2* are «extremely intolerant» and «intolerant» to loss-of-function variation according to the Exome Aggregation Consortium database version 0.3 (<http://exac.broadinstitute.org>) [10.2015] (pLI=1.0 and 0.7, respectively) and as analogous loss-of-function variants in *KMT2D* and *MECP2* were identified in KABUK1<sup>57</sup> and MRXS13<sup>58</sup> individuals, respectively (**Supplementary Table S3-S4**). When possible we used X-ray structures and/or cryo-EM modeling to obtain a three-dimensional representation of the remaining encoded proteins and compared the variants we identified with those previously reported in MOWS, MICPCH, GCE, CFC4 and MRXSC subjects (**Supplementary Table S5-S9**). By and large these models suggest that the variants identified in the current study are detrimental to the encoded proteins: i) the ZEB2 p.H1049P variant substitutes a residue that participates in the coordination of the Zn<sup>++</sup> atom of one of the Zinc fingers, similar to the variant p.H1045R identified in a MOWS subject (**Figure 1A; Supplementary Table S9**); ii) the MAP2K2 p.D69del variant removes one of the two aspartic acid residues involved in the binding of a Ca<sup>++</sup> ion in the conserved GELKDD loop (**Figure 1B**); iii) the *GLDC* p.L726Q and p.P647L variants likely affect the packing of the encoded protein in the neighborhood of the catalytic lysine K754 residue similar to the 61 missense

variants identified in GCE subjects (**Figure 1C, Supplementary Table S6**); and iv) the CASK p.R489W variant places a bulky tryptophan sidechain that cannot be accommodated in the structure without changing the molecular surface (**Figure 1D**). The possible impact of the KDM5C p.K1023R variant on this conserved position (**Supplementary Figure S2**) could not be evaluated as no template is available for this region.



**Figure 1: Modeling of variants identified in SMS patients**

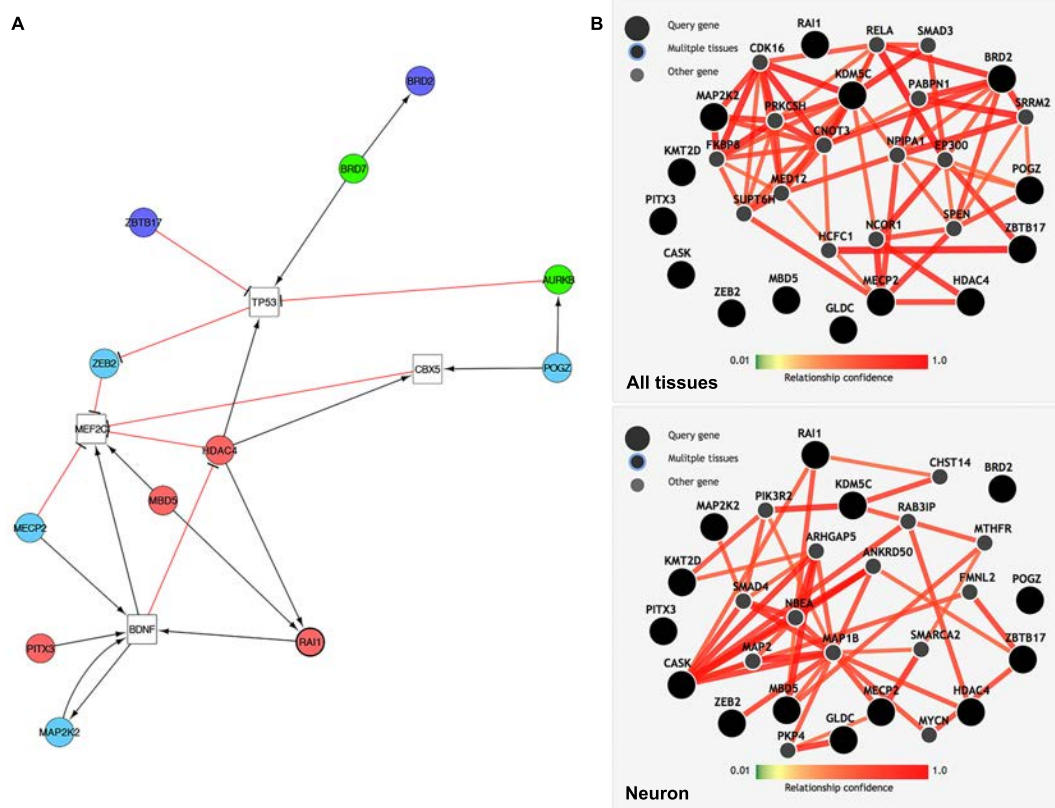
**(A)** ZEB2: Position of the H1049P variant identified in patient BAB2386 with respect to the Zinc finger residues 995-1078 (green ribbon). Residue H1049 (red) together with residue H1045 (purple), which was reported mutated in a Mowat-Wilson syndrome patient (**Supplementary Table S10**), is essential for the coordination of the Zn<sup>++</sup> ion of the Zinc finger. The S1071 residue located within an alpha-helix that directly contacts the DNA major groove (pink) was similarly found mutated in another MOWS patient (**Supplementary Table S10**). A proline at this position will likely disrupt the helix through steric hindrances, thus impacting the DNA binding ability of the Zinc finger. **(B)** MAP2K2: Position of the D69del variant identified in patient BAB2474 with respect to the overall MAP2K2/MAP2K1 structures. Since (i) MAP2K1 has a higher resolution than MAP2K2; (ii) is readily superposable onto MAP2K2; and (iii) residues close to D69del in 3D space are strictly conserved, its model has been used to highlight the context of the variant. The ribbon representation is colored by backbone root-mean-square deviation to MAP2K1 (with a dark blue, blue, cyan, green gradient representing region with rmsd < 0.5, < 1.0, < 1.5, < 2.0 respectively). Regions not resolved in MAP2K2 are shown in red and ADP is represented as a space-filling model. The D69 and D70 residues are located at the top of a conserved GELKDD loop facing the conserved residues S139 and D140. These residues are involved in the binding of a Ca<sup>++</sup> ion (orange). D69 is also binding a Na<sup>+</sup> ion (green), itself coordinated by a network of water molecules (purple) held in place by the conserved R100 (yellow) and E138 (not highlighted). A deletion of D69 will alter the local loop geometry and disrupt the ion coordination ability of this conserved region. **(C)** GLDC: Positions of the compound heterozygote variants P647L (red space-filling model) and L726Q (orange space-filling model) found in patient BAB4947 relative to the 61 missense variants conferring glycine encephalopathy (GCE) reported in the literature, i.e. R59T, L82W, S132W, Y138F, N150T, Y161C, G171A, T187K, A202V, R212K, C225R, T269M, A283P, D295Y, L296R, A313P, P329T, H371D, I372F, R373W, A389V, R410K, I440N, R461Q, P509A, R515S, T532R, N533Y, R536E, M552V, S564I, A569T, E597K, V611G, C644F, H651R, A694P, P700A, S701F, V705M, G728R, N732K, R739H, H753P, H760Q, G761R, G762R, P765S, G766C, P769L, G771R, H775R, R790W, A802V, T830M, M840K, A841P, V905G, P949L, S957P and R966G (yellow space-filling models). Although disease-conferring variants are scattered throughout the peptide chain, once analyzed in their 3D context they cluster in a specific structural region, which in its center bears the catalytic lysine K754 and pyridoxal phosphate (pink space-filling model). All GCE-conferring variants result in bulkier residues or elicit a change of charge (**Supplementary Table S7**). The variants identified in the SMS patient cannot, likewise, be accommodated without forcing a local structural rearrangement; specifically P647L clashes with I676, whereas L726Q collides with P740 and Q734. **(D)** CASK: Two partial CASK crystal structures covering residues 487-572 (red) and 739-914 (orange) were superposed onto the crystal structure of PALS1/Crb to model their relative orientation and highlight the position of the R489 residue (yellow space-filling model) found mutated in patient BAB2540 in context of the overall C-terminal domain of the protein (left panel). The R489 residue (yellow space-filling model) adopts an extended conformation perpendicular to the beta sheet (red; top right panel). Whereas in some species a Lysine residue is found at this position as it can adopt a similar conformation and be positively charged, the

mutant R489W identified in SMS patient BAB2540 cannot be accommodated in this structural context. A bulkier tryptophan residue in this position will have to point outward in a direction parallel to the beta-sheet (bottom right panel). Since it is located at the end of a domain, the protrusion of an unexpected sidechain in this region could prevent the formation of an extended beta-sheet, for example with the N-terminal domain of CASK or another protein, because of collision with its sidechains. The position of a beta-strand present in the crystal structure of PALS1/Crb (pink arrow) is shown to illustrate this concept with R489 (top right panel) and W489 (bottom right panel). The R489W variant will affect the molecular surface of the domain, revealing a recessed hydrophobic patch formed by the hydrophobic residues V491 (dark green) and I567 (light green), which possibly will affect interaction with other molecules.

### The identified rare variants affect Rai1-associated genes

We next proceeded to test the hypothesis that genes mutated in individuals with SMS-like features were associated with *RAI1*. To challenge this assumption we first assessed if *HDAC4*, *MBD5* and *PITX3*, three genes previously reported to be associated with SMS phenotypes<sup>13, 14, 16</sup>, *BRD2* and *ZBTB17* (a.k.a. *MIZ1*), two genes encoding high-confidence *RAI1* interactors we identified by two-hybrid assay (**Methods**) and *ZEB2*, *CASK*, *KMT2D*, *GLDC*, *MECP2*, *MAP2K2*, *POGZ* and *KDM5C*, the eight genes identified here, were part of a *RAI1* functional network. Manual curation of the literature revealed single or double edges functional relationships between 10 of these 14 genes. This network includes *ZEB2*, *MECP2* and *MAP2K2* (**Figure 2A**) indicating that may have uncovered a “disease network” as previously described<sup>59</sup>. The significance of the observed connectivity ( $P=0.00328$ ) was assessed adapting spatial statistics concepts to network analysis<sup>29</sup> (**Methods**). Second, we used the GIANT database (Genome-scale Integrated Analysis of gene Networks in Tissues<sup>60</sup>) to assess whether these fourteen genes form a functional network and eventually capture tissue-specific functional interactions. When considering GIANT data from neurons, *CASK* functions as a provincial hub with nine edges and 11 genes (*CASK*, *GLDC*, *HDAC4*, *KDM5C*, *KMT2D*, *MAP2K2*, *MBD5*, *MECP2*, *RAI1*, *ZBTB17*, *ZEB2*) of the 14 assessed are connected, further supporting the notion of “disease network” as in particular *HDAC4* and *MBD5*, two of the three genes previously associated with SMS-like phenotypes are included<sup>13-15</sup>. 8 out of 14 genes (*BRD2*, *HDAC4*, *KDM5C*, *MAP2K2*, *MECP2*, *POGZ*, *RAI1* and *ZBTB17*) including in particular the two genes,

*BRD2* and *ZBTB17*, encoding high-confidence *RAI1* interactors are functionally linked in the “all tissue” network (**Figure 2**). Furthermore, when *RAI1* is used as single query gene, *RAI1* and *CASK* are directly linked in the resulting gene network but, again, specifically in neurons (**Supplementary Figure S3**). These results and the data extracted from the literature suggest that at least some of the eight genes with variants potentially causing SMS-like phenotypes could possibly be causative as they are functionally associated with *RAI1*.



**Figure 2: *RAI1* Molecular interactions**

**(A) Literature-defined molecular interactions.** Genes interactions network obtained combining literature text-mining resources (i.e. Pubmed, Google Scholar, iHOP and EVEX) and visualized using Cytoscape. Nodes are colored depending on the role of the gene: genes associated with SMS phenotypes in red (*RAI1* is highlighted with a thicker outline); *RAI1* interactors identified in yeast two-hybrid screens in dark blue; genes with rare possibly damaging variants in SMS patients identified in this report in light blue; ‘extra nodes’ required to make connection are shown as white squares, while ‘extra nodes’ found in the 4C BRICKs of the *RAI1* viewpoint are depicted in green.



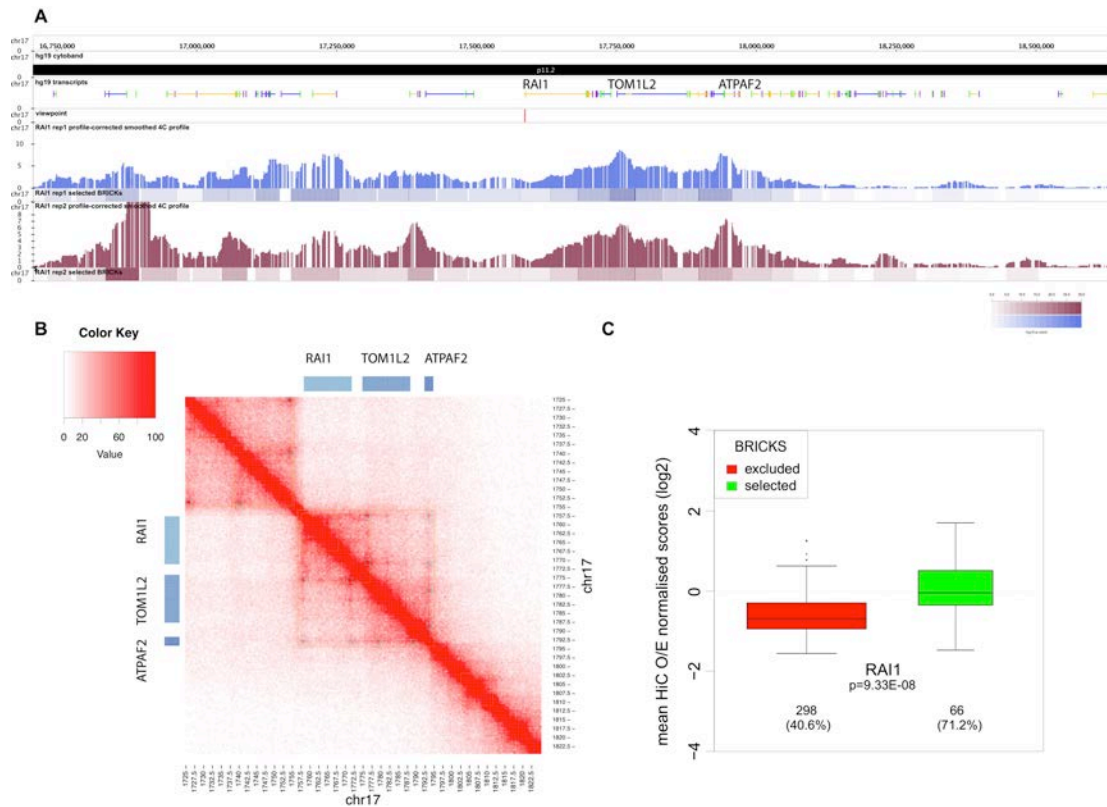
## **(B) Co-expression-based molecular interactions**

Tissue-specific functional interaction network built using GIANT (Genome-scale Integrated Analysis of gene Networks in Tissues, <http://giant.princeton.edu/>) in “all tissues” (top panel) versus “neuron” (bottom panel), with minimum relationship confidence = 0.5 and maximum number of genes = 15. *RAI1*, *HDAC4*, *PITX3*, *MBD5*, *BRD2*, *MIZ1/ZBTB17*, *CASK*, *KMT2D*, *MECP2*, *GLDC*, *KDM5C*, *MAP2K2*, *POGZ* and *ZEB2* were used as queries.

To gain further insights about the genes regulated by *Rai1* during mouse embryonic development, we performed microarray analysis on total RNA prepared from three 10.5 dpc *Rai1*<sup>-/-</sup> embryos and from three of their wild type littermates. The two *Rai1* transcripts present on the array are significantly down-regulated in *Rai1*<sup>-/-</sup> embryos compared to wild type littermates (e.g. the AK013909 transcript with a fold change of 6.2 shows the largest down-regulation amongst the 45,037 assessed probesets). In fact the expression values for both transcripts are within background levels in the *Rai1*<sup>-/-</sup> embryos indicating that both transcripts are not expressed in the *Rai1*<sup>-/-</sup> mutants and further corroborating the contention that the engineered *Rai1* mutant allele is a complete null allele<sup>30</sup>. In total, 142 and 157 probe sets showed an over 2-fold increase or decrease, respectively (**Supplementary Table S10; Methods**) in the mutant mice. Consistent with the hypothesis that genes potentially causative of the SMS-like phenotypes are functionally associated to or transcriptionally regulated by *RAI1*, the expression levels of both *Zeb2* (ENSMUSG00000026872) and *Map2k2* (ENSMUSG00000035027) were perturbed in *Rai1*<sup>-/-</sup> mice (**Supplementary Table S10**). We then assessed the chromosomal position of the dysregulated genes. The enrichment score using a Pearson Chi-Square goodness of fit statistic indicated that they showed a biased chromosome distribution with 22% of the genes down- and 26% of the genes up-regulated in the *Rai1* mutants mapping to mouse chromosome 11 (MMU11) where the *Rai1* gene resides. Less than 5% of the differentially expressed genes are located on any chromosome other than MMU11. This enrichment on MMU11 for down- and up-regulated genes in *Rai1*<sup>-/-</sup> embryos is reminiscent of our previous finding that the engineered MMU11 deletion and reciprocal duplication that mimic SMS

and Potocki-Lupski syndrome rearrangements were associated with a MMU11-wide transcriptome perturbation in the five assessed adult male tissues<sup>61</sup>.

Chromatin architecture can similarly be exploited to identify genes that belong to the same pathway. Long-range chromatin contacts, which bring genes in close proximity to regulatory sequences, have been shown to be necessary for co-transcription of biologically-related and developmentally co-regulated genes<sup>62, 63</sup>.



**Figure 3: 4C interactions profile of *RAI1* and comparison with Hi-C interactions profiles locally and globally (A) (Panels from top to bottom)**

**Transcripts:** The structure of the transcripts mapping within human 17p11.2 cytoband from approximately 16.5Mb to 18.5Mb are indicated, in particular those of the *RAI1*, *TOM1L2* and *ATPAF2* genes.

**Viewpoint:** The red tick shows the mapping position of the *RAI1* viewpoint used in the 4C experiments.

**PC/BRICKS:** Smoothed and profile-corrected 4C signal (upper part of each panel) and BRICKs (lower part) identified for each replicate (blue and burgundy). The corresponding BRICKs significance heatmap color legend is showed in the bottom right corner.

**(B)** High resolution Hi-C chromosome conformation capture results obtained with the GM12878 LCL within the chromosome 17 17.25-18.22Mb window (5kb resolution). Yellow and light blue squares highlight the contact domains and peaks identified in <sup>45</sup>, respectively. The position of the *RAI1*, *TOM1L2* and *ATPAF2* genes is indicated.

**(C)** Distribution of Hi-C scores in selected (FDR1%) vs non-selected BRICKs (FDR10%). Virtual 4C-seq tracks were generated for the *RAI1* viewpoint from the GM12878 Hi-C results published in ENREF\_39<sup>45</sup> (5kb resolution) by extracting the Hi-C vectors from the KR normalized observed/expected matrices. BRICKs found with the viewpoint were quantified by the mean Hi-C signals. The p-value of two-sided t-test is reported for the comparison, together with the number of Hi-C bins and the % of non-NA bins.

We recently showed the pertinence of this approach by documenting that genes associated with ASD and head circumference phenotypes were linked by chromatin loops<sup>34</sup>. As a third approach to assess if *RAI1* is biologically related to the eight genes identified in the SMS-like individuals, we used an adapted version of the 4C method (4C-seq: Circularized Chromosome Conformation Capture combined with multiplexed high-throughput sequencing)<sup>34, 35, 64-66</sup> to identify chromosomal regions that physically associate with the *RAI1* “viewpoint”. We independently analyzed the local pattern of chromosomal interactions in LCLs of two control individuals (**Supplementary Figure S4-S5, Methods**). Genome-wide we detected 153 significant BRICKs (Blocks of Regulators In Chromosomal Kontext; FDR ≤ 1%), *i.e.* 3-dimensionally interacting genomic fragments (**Methods; Supplementary Table S11**), encompassing 147 genes. Within the 66 (43%) intrachromosomal BRICKs we identified, in particular, two genomic intervals that flank the *RAI1* viewpoint (**Figure 3A-B**) and which are *de facto* positive controls, as they were previously reported to interact with *RAI1* in high resolution Hi-C (genome-wide conformation capture)

from LCLs<sup>38,45</sup>. Although trans-DNA contacts from Hi-C datasets are only reliable when determined over genomic windows larger than single-restriction fragment<sup>67</sup>, we can report consistency between Hi-C results<sup>38</sup> and the inter- and intrachromosomal contacts we identified in this report (**Figure 3C, Methods**).

The genes mapping within the *RAI1*-chromatin contacted genomic loci (BRICKs genes) are enriched for genes that encode proteins that interact together (82 observed interactions versus 35 expected;  $P=6.41e^{-12}$ ). BRICKs genes are also enriched for the GO term “detection of light stimulus involved in sensory perception” in Enrichr ( $P=5.45e^{-3}$ ) (**Methods; Supplementary Table S12**). Similarly, Enrichr showed that chromosome contacts were enriched in inter- and intrachromosomal cytobands (17p11, adjusted  $P < 1e^{-09}$ ; 17p12, adjusted  $P =9.7e^{-09}$ ; 17p13, adjusted  $P =1.8e^{-03}$ ; and 2q22 adjusted  $P =4.95e^{-02}$ ). *ZEB2*, one of the eight genes found mutated in the SMS subjects, maps to the latter 2q22.3 region and is flanked by BRICKs. To further assess possible functional relationships between *RAI1* and chromatin-contacted genes, we retrieved the list of 322 genes flanking the BRICKs (BRICKs flanking genes, i.e. the closest genes to be found upstream and downstream of a BRICK within a 500kb window). The 4C assays in particular identified interchromosomal contacts with restriction fragments mapping 200kb away from the *ZEB2* and *GLDC* gene loci. We then compared the lists of BRICKs genes and BRICKs flanking genes with the list of genes whose expression levels were perturbed in *Rai1*<sup>-/-</sup> mouse embryos. Although our analysis is restricted by a small sample size, we found a consistent trend of overrepresentation (Fisher’s enrichment test,  $P=0.22$ , OR=1.5, and  $P=0.2$ , OR=1.4) with 10 and 18 chromatin-contacted BRICKs genes and BRICKs flanking genes, respectively, differentially expressed in the mouse knockdown model. Interestingly, 6 out of 10 of these BRICKs genes mapping at cytobands 17p13, 17p11 (2 genes), 17q21 and 17q23 (2) have mouse orthologs that map on mouse chromosome MMU11, thus possibly explaining the enrichment of MMU11-mapping genes within genes differentially expressed in *Rai1*<sup>-/-</sup> mouse embryos (**Supplementary Figure S6**).

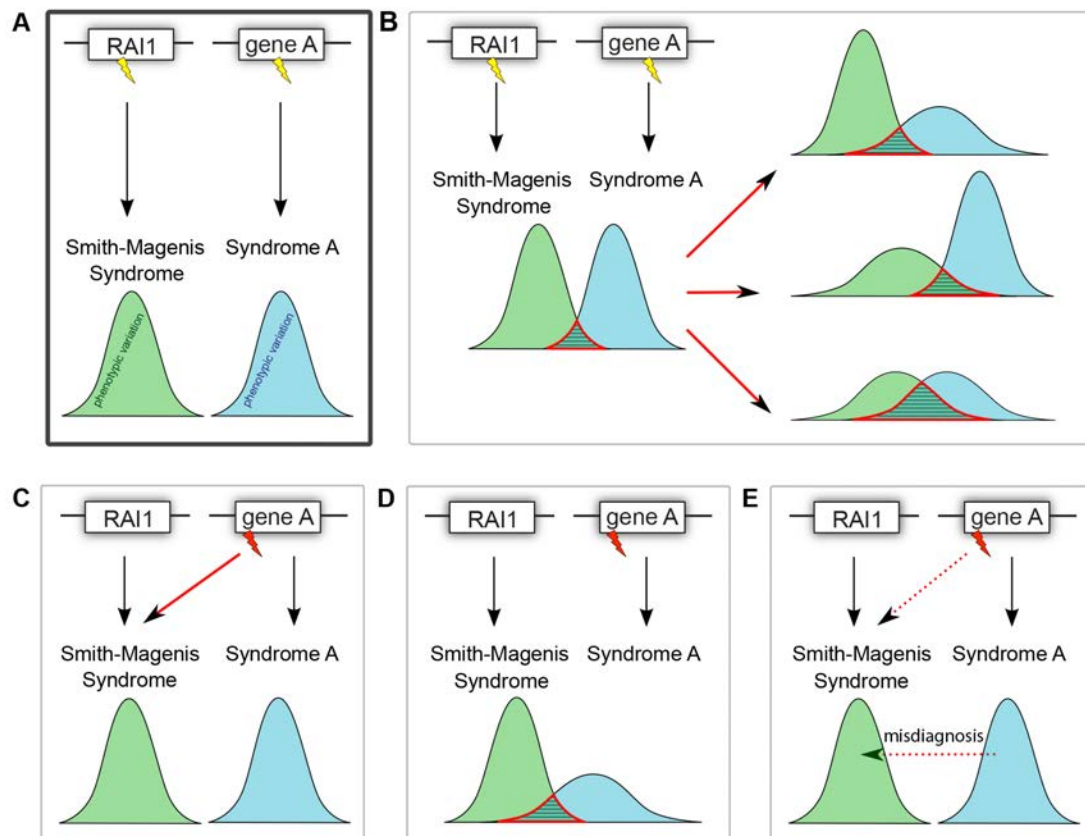
## Discussion

Within a cohort of 149 individuals presenting clinical features of SMS we identified 90% (134/149) of individuals with either a heterozygous deletion of *RAI1* or a predicted deleterious variant of the *RAI1* gene. We used recent advances in genome sequencing technologies to possibly identify genetic alteration(s) associated with SMS in the remaining subjects. These strategies were successfully applied to discover loci associated with ID<sup>68</sup>. They revealed a large genic overlap between ID and ASD, schizophrenia and epileptic encephalopathy<sup>69</sup> suggesting that some developmental disorders have highly variable clinical presentations. They similarly uncovered limitations to the phenotype-driven strategy and conventional clinical paradigm of identifying individuals with very similar presentations as they revealed an unsuspected phenotypic variance of known disorders<sup>70,71</sup>.

The diagnosis of SMS has primarily relied on clinical suspicion and consideration in a differential diagnosis followed by laboratory studies and confirmatory molecular findings. Since the subjects studied in this cohort were ascertained by experienced clinicians, an aptitude supported by the low number of individuals without a *RAI1* molecular diagnosis, we exploited the remaining 15 individuals to assess the possible heterogeneity of this syndrome (**Figure 4**). We identified potentially causal genetic alterations in nine individuals. They comprise variants in the *ZEB2*, *CASK*, *KMT2D*, *GLDC*, *MECP2*, *MAP2K2*, *KDM5C* and *POGZ* genes, which are associated with MOWS, MICPCH, KABUK1, GCE, MRXS13, CFC4, MRXSC and a new ID syndrome<sup>55</sup>, respectively, as well as a 47, XYY karyotype.

It is important to the medical community to identify phenotypic overlap between diseases, which suggest common causes and alterations of the same pathways, as this knowledge could be exploited therapeutically. In this report we identify previously unappreciated relationships between SMS and its major driver *RAI1* and other diseases that include MOWS, MICPCH, KABUK1, GCE, MRXS13, CFC4 and MRXSC. Literature mining, co-expression data, transcriptome profiling of *Rai1*<sup>-/-</sup> animal models and chromosomal contacts support the existence of a

comprehensive “biological module”<sup>72</sup> or “disease network”<sup>59</sup> underlying these diseases.



**Figure 4: Genetic heterogeneity, phenotypic variance and misdiagnosis**  
**(A)** *RAI1* and gene A are associated with Smith-Magenis syndrome (SMS) and syndrome A, respectively and variants (yellow “thunders”) in those genes cause the green and blue phenotypes, respectively. **(B)** The phenotypic spectra of these diseases could be more variable than anticipated and result in overlapping features. Such overlap could be due to a broader phenotypic variability of syndrome A, of SMS or of both syndromes (right panels). The rare variants in *Rai1*-associated genes identified in individuals with SMS-like features and reported here (red “thunders”) could be explained by a combination of genetic heterogeneity of SMS and allelic heterogeneity of gene A **(C)**, an increased variance of syndrome A **(D)** or a misdiagnosis of SMS **(E)**.

Are we exploring genetic heterogeneity in SMS or misdiagnosis of these syndromes?—Although none of the 15 subjects described in this study have traditional molecular diagnoses involving *RAI1* haploinsufficiency and thus

should formally be considered misdiagnoses, many have phenotypes with considerable overlap with SMS (**Figure 4, Supplementary Table S1, Supplementary text**). BAB4947 presented facial dysmorphisms, SMS-like behavioral disturbances that include sleep problems, polyembolokoilamania, onychotillomania, brachycephaly and brachydactyly, as well as known *GLDC*-variants associated features such as seizures. His clinical diagnosis could possibly be confounded by the likely presence of two molecular diagnoses: compound heterozygous variants in *GLDC* and an inherited frameshift variant in *TCOF1*, a gene associated with Treacher Collins syndrome-1 (OMIM #154500) and possibly accounting for the down-slanting eyes, everted lateral eyelids and malar hypoplasia. The clinical scenario is similar with cases BAB2474 and BAB2540, who did not show CFC4- (e.g. ectodermal anomalies, craniofacial features) and MICPCH-distinctive features (e.g. microcephaly and pontocerebellar hypoplasia). Likewise individual BA2492 has a constellation of symptoms (sleep disturbance, DD, cognitive impairment, brachydactyly) compatible with only the most severe 47,XYY sex chromosome aneuploidy cases<sup>73</sup>. Consistent with the hypothesis of expanded phenotypes, the phenotypic variability of White-Sutton syndrome associated with variants in *POGZ* keeps extending with clinical features including ASD, DD, ID, schizophrenia and microcephaly<sup>55, 68, 74-80</sup>. We can also not formally rule out that we have not yet determined the true genetic cause(s) of the phenotypic spectrum of these subjects or they occur in presence of more complex, blended phenotypes as exemplified by BAB4947 above.

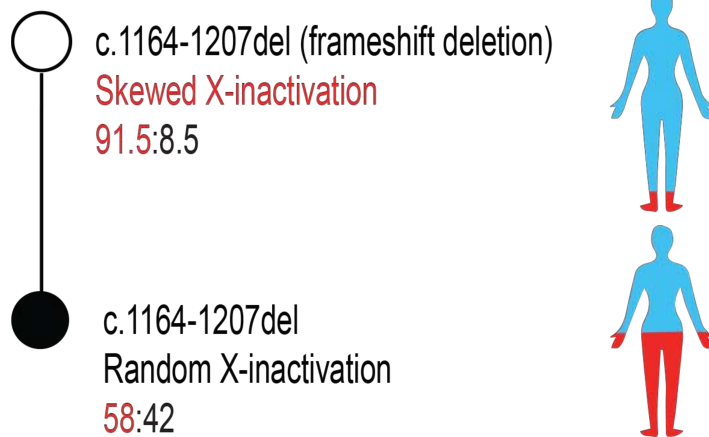
The presented results support the notion that at least some of the identified variants in candidate SMS contributory genes *CASK*, *GLDC*, *KDM5C*, *KMT2D*, *MAP2K2*, *MECP2*, *POGZ* and *ZEB2* are causative of the observed phenotypes and thus that modification of the function of these genes is associated with a greater phenotypic variability than previously expected (**Figure 4**). Conversely, one and two carriers of damaging *RAI1* variants were identified within a total of 6,381 ASD<sup>76, 81</sup> and 2,426 ID<sup>68, 75, 82-85</sup> individuals, respectively. Whereas the phenotype of one of the ID subjects was retrospectively found to be consistent with SMS<sup>85</sup>, we lack detailed phenotypic information regarding the other two cases. If we

assume that these two individuals do not present with typical SMS features that would have excluded them from these cohorts, it suggests that the phenotype of carriers of *RAI1* deleterious variants is similarly more variable than anticipated.

Structural variations, especially large rearrangements involving several genes, shape tissue transcriptomes and impact the expression of genes mapping to their flanks<sup>86,61</sup>. We show that the homozygous deletion of *Rai1* in mouse embryos<sup>30</sup> influences the expression of several genes and in particular MMU11 genes. The *RAI1* viewpoint contacts the orthologous genes at the chromatin level. As some of these genes contribute to phenotypes associated with *RAI1* variation (e.g. *KRT17* with “hoarse voice” (HP:0001609), *B9D1* with “low-set, posteriorly rotated ears” (HP:0000368), “hypertelorism” (HP:0000316) and “microcornea” (HP:0000482)), they could be involved in *RAI1* pathways. It is thus possible that the disruption of the orthologous locus in the *Rai1*<sup>-/-</sup> mice perturbs chromatin loops and affects expression levels of *RAI1*-contacted/functionally associated genes. Our results strongly support a disease network associated with *RAI1*, and illustrate the utility of a comprehensive multi faceted diagnostic approach even in the presence of a distinctive disorder.



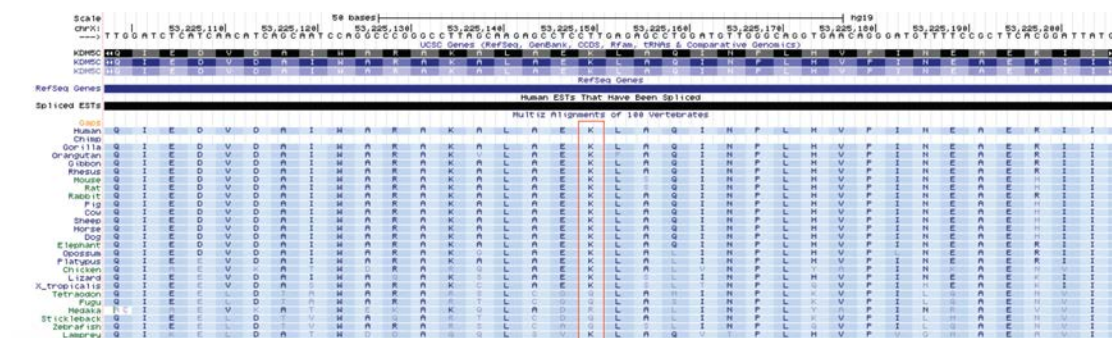
## Supplementary Figures



Maternally inherited *MECP2* frameshift deletion. X-inactivation studies revealed highly skewed inactivation in the mother but not in the affected daughter.

## Supplementary Figure S1: X-inactivation in patient BAB2552 and her mother

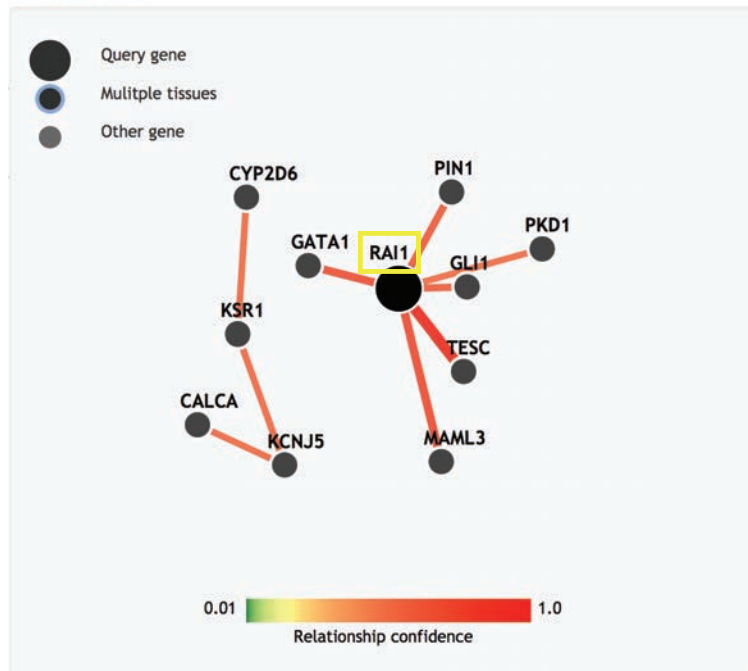
X-inactivation analyses of patient BAB2552 (top) and her mother (bottom).



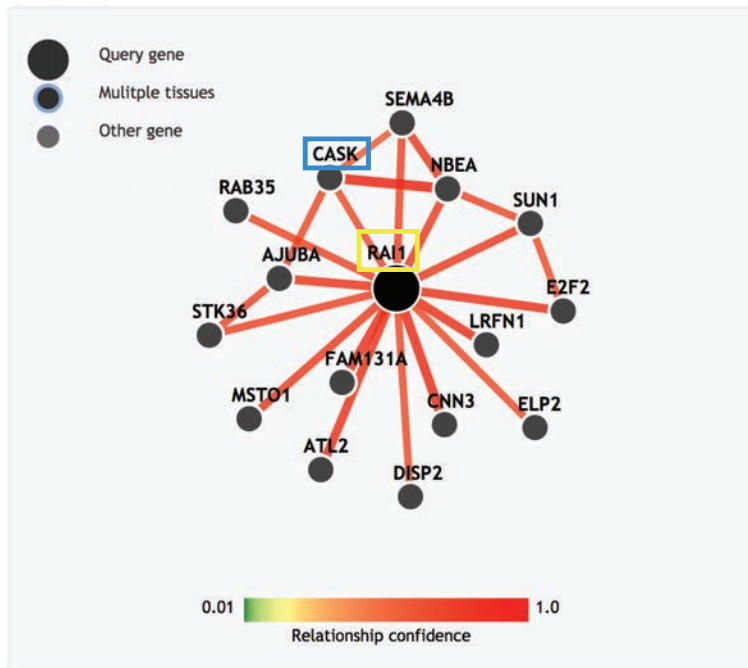
## Supplementary Figure S2: Conservation of the K1023 residue of KDM5C

PhyloP and Multiz Alignment UCSC tools showing the conservation of the K1023 residue of KDM5C (position highlighted by a red rectangle) across vertebrate species (human, chimp, gorilla, orangutan, gibbon, rhesus, mouse, rat, rabbit, pig, cow, sheep, horse, dog, elephant, opossum, platypus, chicken, lizard, xenopus tropicalis, tetraodon, fugu, medaka, stickleback, zebrafish and lamprey).

### all tissues



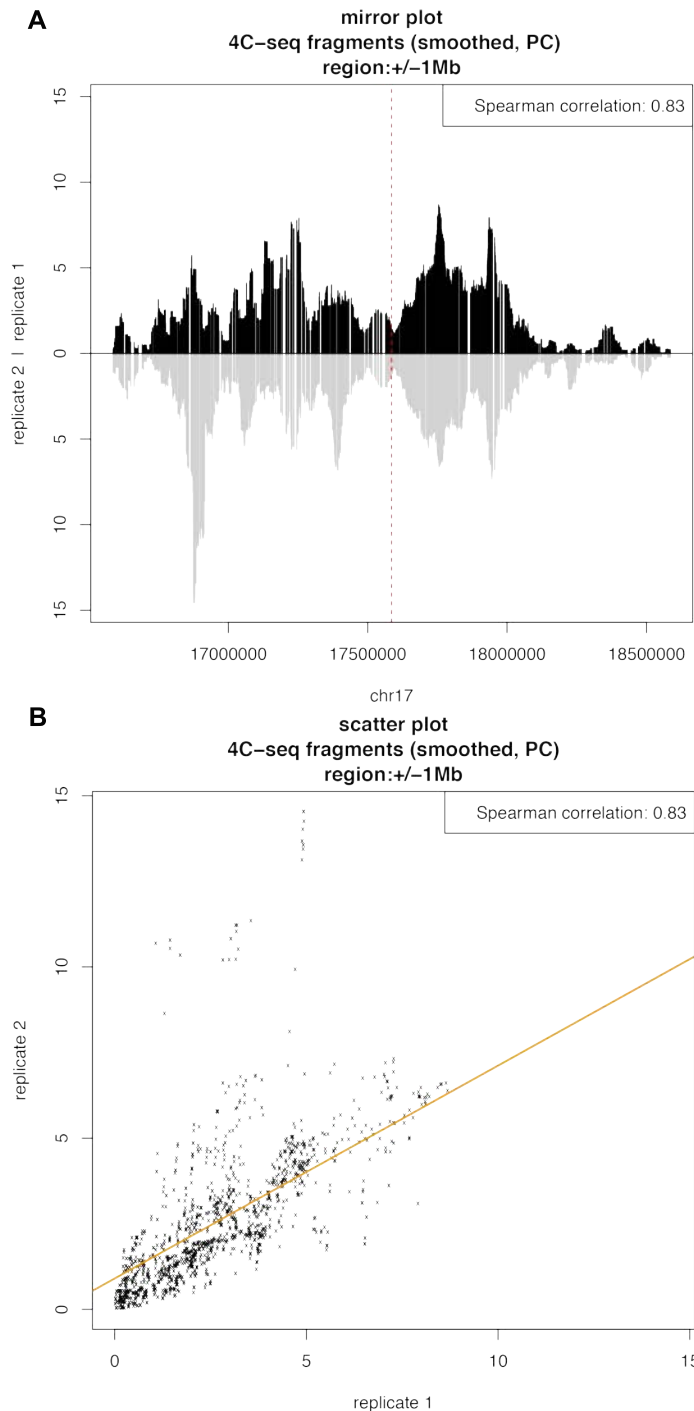
### neuron



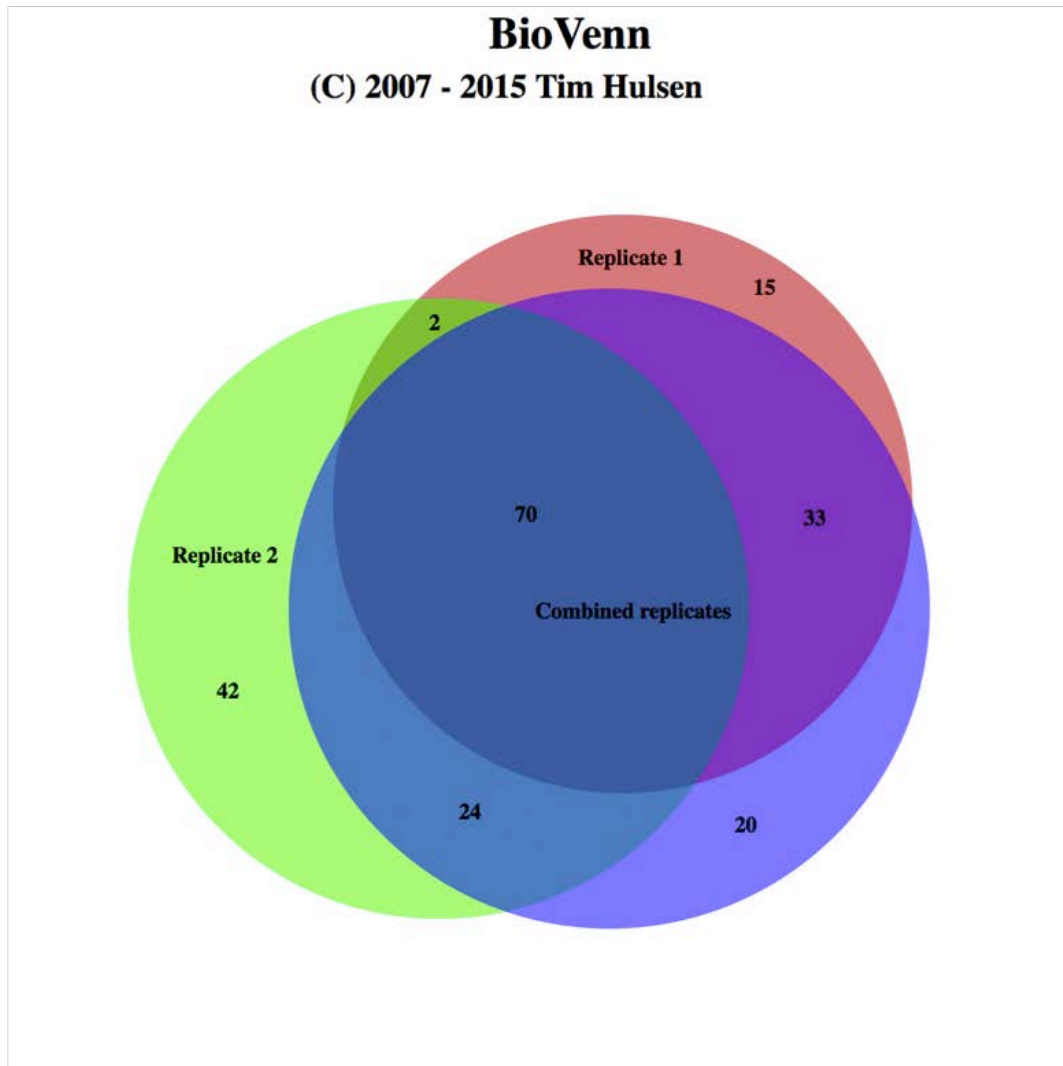
### Supplementary Figure S3: Co-expression-based molecular interactions

Tissue-specific functional interaction network built using GIANT (Genome-scale Integrated Analysis of gene Networks in Tissues, <http://giant.princeton.edu/>) in “all tissues” (top panel) versus “neuron” (bottom panel), with minimum

relationship confidence = 0.5 and maximum number of genes = 15. *RAI1* was used as query (relationship confidence: 0.58; 77% due to co-expression).

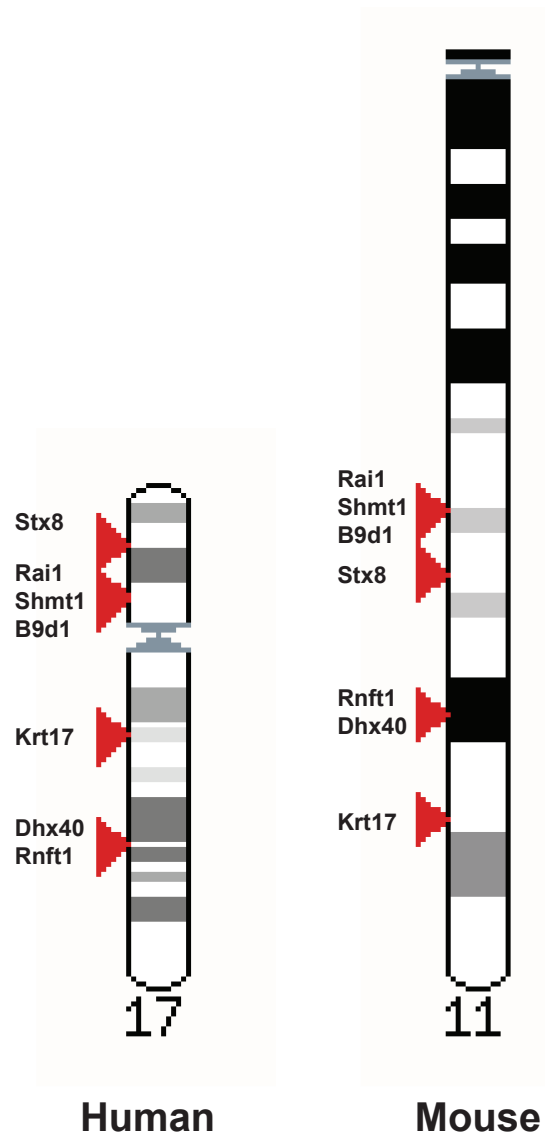


**Supplementary Figure S4: Comparison of 4C-seq replicates**  
Mirror **(A)** and Scatter plot **(B)** of 4C-seq profile of replicate 1 (top portion) and 2 (bottom portion) in the region defined as viewpoint +/- 1Mb with Spearman's rank correlation calculated between the two replicates.



**Supplementary Figure S5: Comparison of 4C-seq replicates**

Venn diagram showing the overlap between the BRICKS obtained after averaging replicates (combRep) compared to BRICKS found in each replicate separately (rep1 and rep2).



**Supplementary Figure S6:**

Mapping of the HSA17 BRICKs genes and their mouse MMU11 orthologs differentially expressed in 10.5 dpc *Rai1*<sup>-/-</sup> mouse embryos.

## Supplementary Tables

**Supplementary Table S1: Clinical phenotype of SMS patients without RAI1 alteration compared with classical SMS phenotype.**

**Supplementary Table S2: Comparison of typical SMS, MOWS, MICPCH, KABUK1, GCE, MRXS13, CFC4 and MRXSC phenotypes.**

**Supplementary Table S3: pathogenic variants identified in KMT2D in this report (bold, patient BAB2319) and in literature (references below).**

**Supplementary Table S4: pathogenic variants identified in MECP2 in this report (bold, patient BAB2552) and in literature (references below).**

**Supplementary Table S5: pathogenic variants identified in CASK in this report (bold, patient BAB2540) and in literature (references below).**

**Supplementary Table S6: pathogenic variants identified in GLDC in this report (bold, patient BAB4947) and in literature (references below).**

**Supplementary Table S7: pathogenic variants identified in KDM5C in this report (bold, patient BAB2293) and in literature (references below).**

**Supplementary Table S8: pathogenic variants identified in MAP2K2 in this report (bold, patient BAB2474) and in literature (references below).**

**Supplementary Table S9: pathogenic variants identified in ZEB2 in this report (bold, patient BAB2386) and in literature (references below).**

**Supplementary Table S10: genes (Affymetrix probesets) with modified expression levels in RAI1<sup>-/-</sup> embryos.**

Tables available upon request

**Supplementary Table S11: List of BRICKs encompassed genes and BRICKs flanking genes, including the closest genes upstream and downstream of a BRICK within a 500kb window.**

BRICKs genes		BRICKs flanking genes	
Ensembl ID	Gene Symbol	Ensembl ID	Gene Symbol
ENSG00000162415	ZSWIM5	ENSG00000162415	ZSWIM5
ENSG00000066294	CD84	ENSG00000184005	ST6GALNAC3
ENSG00000184005	ST6GALNAC3	ENSG00000162631	NTNG1
ENSG00000178104	PDE4DIP	ENSG00000178104	PDE4DIP
ENSG00000163386	NBPF10	ENSG00000255168	RP11-458D21.5
ENSG00000162631	NTNG1	ENSG00000163386	NBPF10
ENSG00000107643	MAPK8	ENSG00000066294	CD84
ENSG00000107957	SH3PXD2A	ENSG00000107643	MAPK8
ENSG00000156026	CCDC109A	ENSG00000156026	CCDC109A
ENSG00000110324	IL10RA	ENSG00000138315	OIT3
ENSG00000133812	SBF2	ENSG00000107957	SH3PXD2A
ENSG00000130592	LSP1	ENSG00000130592	LSP1
ENSG00000157837	AC069214.1	ENSG00000133812	SBF2
ENSG00000099814	KIAA0284	ENSG00000110324	IL10RA
ENSG00000042088	TDP1	ENSG00000157837	AC069214.1
ENSG00000100592	DAAM1	ENSG00000100592	DAAM1
ENSG00000157890	MEGF11	ENSG00000042088	TDP1
ENSG00000206193	AC124312.1	ENSG00000099814	KIAA0284
ENSG00000182175	RGMA	ENSG00000206193	AC124312.1
ENSG00000150667	FSIP1	ENSG00000150667	FSIP1
ENSG00000233356	AC133919.1	ENSG00000157890	MEGF11
ENSG00000166164	BRD7	ENSG00000182175	RGMA
ENSG00000214627	AC142381.2	ENSG00000214627	AC142381.2
ENSG00000125124	BBS2	ENSG00000121281	ADCY7
ENSG00000170160	CCDC144A	ENSG00000166164	BRD7
ENSG00000131885	KRT17P1	ENSG00000125124	BBS2
ENSG00000240505	TNFRSF13B	ENSG00000233356	AC133919.1
ENSG00000133030	MPRIP	ENSG00000083457	ITGAE
ENSG00000214899	C17orf84	ENSG00000167840	ZNF232
ENSG00000179598	PLD6	ENSG00000129204	USP6
ENSG00000141030	COPS3	ENSG00000180626	ZNF594
ENSG00000205309	NT5M	ENSG00000179029	TMEM107
ENSG00000141026	MED9	ENSG00000178999	AURKB
ENSG00000133027	PEMT	ENSG00000170310	STX8
ENSG00000108557	RAI1	ENSG00000007237	GAS7
ENSG00000072310	SREBF1	ENSG00000006695	COX10
ENSG00000203520	AC122129.2	ENSG00000221926	TRIM16
ENSG00000171962	LRRC48	ENSG00000255104	AC005324.8-001
ENSG00000141034	C17orf39	ENSG00000214946	TBC1D26

ENSG00000091536	MYO15A	ENSG00000170315	UBB
ENSG00000108591	DRG2	ENSG00000187688	TRPV2
ENSG00000091542	ALKBH5	ENSG00000181350	C17orf76
ENSG00000214860	EVPLL	ENSG00000170160	CCDC144A
ENSG00000171916	LGALS9C	ENSG00000131885	KRT17P1
ENSG00000250766	GRAPL	ENSG00000240505	TNFRSF13B
ENSG00000108641	B9D1	ENSG00000230969	AC104024.2
ENSG00000205217	FAM106B	ENSG00000133030	MPRIP
ENSG00000170298	LGALS9B	ENSG00000214899	C17orf84
ENSG00000214819	CDRT15L2	ENSG00000179598	PLD6
ENSG00000141127	PRPSAP2	ENSG00000154803	FLCN
ENSG00000170315	UBB	ENSG00000141030	COP3
ENSG00000141068	KSR1	ENSG00000205309	NT5M
ENSG00000170310	STX8	ENSG00000141026	MED9
ENSG00000083290	ULK2	ENSG00000108551	RASD1
ENSG00000205160	AC005726.4	ENSG00000133027	PEMT
ENSG00000231421	AC006050.2	ENSG00000108557	RAI1
ENSG00000186847	KRT14	ENSG00000072310	SREBF1
ENSG00000141543	EIF4A3	ENSG00000175662	TOM1L2
ENSG00000214946	TBC1D26	ENSG00000203520	AC122129.2
ENSG00000177731	FLII	ENSG00000171962	LRRC48
ENSG00000167434	CA4	ENSG00000171953	ATPAF2
ENSG00000255104	AC005324.8-001	ENSG00000141034	C17orf39
ENSG00000109016	DHRS7B	ENSG00000108591	DRG2
ENSG00000180626	ZNF594	ENSG00000091536	MYO15A
ENSG00000167840	ZNF232	ENSG00000091542	ALKBH5
ENSG00000007237	GAS7	ENSG00000131899	LLGL1
ENSG00000108406	DHX40	ENSG00000177731	FLII
ENSG00000006695	COX10	ENSG00000177427	SMCR7
ENSG00000161649	CD300LG	ENSG00000177302	TOP3A
ENSG00000083457	ITGAE	ENSG00000176974	SHMT1
ENSG00000109065	NAT9	ENSG00000214860	EVPLL
ENSG00000189050	RNFT1	ENSG00000186831	AL353997.5
ENSG00000179029	TMEM107	ENSG00000171916	LGALS9C
ENSG00000101605	MYOM1	ENSG00000141127	PRPSAP2
ENSG00000226924	AC139100.2	ENSG00000154016	GRAP
ENSG00000060069	CTDP1	ENSG00000228157	AC007952.5
ENSG00000182315	AC010606.1	ENSG00000250766	GRAPL
ENSG00000213988	ZNF90	ENSG00000230493	AC106017.3
ENSG00000105135	ILVBL	ENSG00000108641	B9D1
ENSG00000163046	ANKRD30BL	ENSG00000166482	MFAP4
ENSG00000221823	PPP3R1	ENSG00000083290	ULK2
ENSG00000121964	GTDC1	ENSG00000154898	CCDC144C
ENSG00000182389	CACNB4	ENSG00000205217	FAM106B
ENSG00000135931	ARMC9	ENSG00000170298	LGALS9B
ENSG00000088766	CRLS1	ENSG00000214819	CDRT15L2



ENSG00000160216	AGPAT3	ENSG00000124422	USP22
ENSG00000176597	B3GNT5	ENSG00000109016	DHRS7B
ENSG00000236438	FAM157A	ENSG00000141068	KSR1
ENSG00000114841	DNAH1	ENSG00000167524	AC005726.6
ENSG00000174748	RPL15	ENSG00000205160	AC005726.4
ENSG00000205097	FRG2	ENSG00000231421	AC006050.2
ENSG00000145734	BDP1	ENSG00000186847	KRT14
ENSG00000113504	SLC12A7	ENSG00000128422	KRT17
ENSG00000249160	RP11-1C1.5	ENSG00000161647	MPP3
ENSG00000112218	GPR63	ENSG00000161649	CD300LG
ENSG00000145990	GFOD1	ENSG00000108406	DHX40
ENSG00000214113	LYRM4	ENSG00000189050	RNFT1
ENSG00000170786	SDR16C5	ENSG00000167434	CA4
ENSG00000182197	EXT1	ENSG00000109062	SLC9A3R1
ENSG00000177257	DEFB4B	ENSG00000109065	NAT9
ENSG00000171711	DEFB4A	ENSG00000141543	EIF4A3
ENSG00000164830	OXR1	ENSG00000101605	MYOM1
ENSG00000197859	ADAMTSL2	ENSG00000060069	CTDP1
ENSG00000185963	BICD2	ENSG00000101544	ADNP2
ENSG00000136868	SLC31A1	ENSG00000226924	AC139100.2
ENSG00000130635	COL5A1	ENSG00000230522	MBD3L2
ENSG00000107077	KDM4C	ENSG00000182315	AC010606.1
ENSG00000136895	GARNL3	ENSG00000105135	ILVBL
ENSG00000181090	EHMT1	ENSG00000213988	ZNF90
ENSG00000169306	IL1RAPL1	ENSG00000221823	PPP3R1
ENSG00000255168	RP11-458D21.5	ENSG00000163046	ANKRD30BL
ENSG00000138315	OIT3	ENSG00000121964	GTDC1
ENSG00000121281	ADCY7	ENSG00000182389	CACNB4
ENSG00000230969	AC104024.2	ENSG00000135931	ARMC9
ENSG00000154803	FLCN	ENSG00000088766	CRLS1
ENSG00000108551	RASD1	ENSG00000160216	AGPAT3
ENSG00000175662	TOM1L2	ENSG00000197885	NKIRAS1
ENSG00000171953	ATPAF2	ENSG00000174748	RPL15
ENSG00000177302	TOP3A	ENSG00000114841	DNAH1
ENSG00000186831	AL353997.5	ENSG00000053524	MCF2L2
ENSG00000154016	GRAP	ENSG00000176597	B3GNT5
ENSG00000230493	AC106017.3	ENSG00000236438	FAM157A
ENSG00000166482	MFAP4	ENSG00000205097	FRG2
ENSG00000154898	CCDC144C	ENSG00000113504	SLC12A7
ENSG00000181350	C17orf76	ENSG00000236454	AC116351.3
ENSG00000167524	AC005726.6	ENSG00000249160	RP11-1C1.5
ENSG00000128422	KRT17	ENSG00000145734	BDP1
ENSG00000177427	SMCR7	ENSG00000214113	LYRM4
ENSG00000221926	TRIM16	ENSG00000145990	GFOD1
ENSG00000124422	USP22	ENSG00000112218	GPR63
ENSG00000129204	USP6	ENSG00000177257	DEFB4B

ENSG00000161647	MPP3
ENSG00000109062	SLC9A3R1
ENSG00000178999	AURKB
ENSG00000101544	ADNP2
ENSG00000230522	MBD3L2
ENSG00000053524	MCF2L2
ENSG00000197885	NKIRAS1
ENSG00000236454	AC116351.3
ENSG00000177243	DEFB103B
ENSG00000176797	DEFB103A
ENSG00000187616	TMEM8C
ENSG00000119321	FKBP15
ENSG00000131899	LLGL1
ENSG00000176974	SHMT1
ENSG00000228157	AC007952.5
ENSG00000187688	TRPV2

ENSG00000177243	DEFB103B
ENSG00000176797	DEFB103A
ENSG00000171711	DEFB4A
ENSG00000170786	SDR16C5
ENSG00000164830	OXR1
ENSG00000182197	EXT1
ENSG00000107077	KDM4C
ENSG00000185963	BICD2
ENSG00000119321	FKBP15
ENSG00000136868	SLC31A1
ENSG00000136895	GARNL3
ENSG00000187616	TMEM8C
ENSG00000197859	ADAMTSL2
ENSG00000130635	COL5A1
ENSG00000181090	EHMT1
ENSG00000169306	IL1RAPL1
ENSG00000101577	LPIN2
ENSG00000102891	MT4
ENSG00000103769	RAB11A
ENSG00000105137	SYDE1
ENSG00000107745	MICU1
ENSG00000107960	OBFC1
ENSG00000108443	RPS6KB1
ENSG00000108599	AKAP10
ENSG00000108602	ALDH3A1
ENSG00000109047	RCVRN
ENSG00000109066	TMEM104
ENSG00000109107	ALDOC
ENSG00000109536	FRG1
ENSG00000112214	FHL5
ENSG00000115145	STAM2
ENSG00000115946	PNO1
ENSG00000119865	CNRIP1
ENSG00000121274	PAPD5
ENSG00000122497	NBPF14
ENSG00000122971	ACADS
ENSG00000123545	NDUFAF4
ENSG00000124523	SIRT5
ENSG00000125872	LRRN4
ENSG00000126088	UROD
ENSG00000126856	PRDM7
ENSG00000127080	IPPK
ENSG00000128482	RNF112
ENSG00000128487	SPECC1
ENSG00000128805	ARHGAP22
ENSG00000130544	ZNF557

ENSG00000130598	TNNI2
ENSG00000131844	MCCC2
ENSG00000132581	SDF2
ENSG00000135914	HTR2B
ENSG00000136856	SLC2A8
ENSG00000136859	ANGPTL2
ENSG00000136867	SLC31A2
ENSG00000137648	TMPRSS4
ENSG00000137747	TMPRSS13
ENSG00000137801	THBS1
ENSG00000138308	PLA2G12B
ENSG00000140807	NKD1
ENSG00000141034	GID4
ENSG00000141040	ZNF287
ENSG00000141367	CLTC
ENSG00000141527	CARD14
ENSG00000145506	NKD2
ENSG00000145979	TBC1D7
ENSG00000145982	FARS2
ENSG00000148408	CACNA1B
ENSG00000148828	LOC100288255
ENSG00000148926	ADM
ENSG00000150991	UBC
ENSG00000152315	KCNK13
ENSG00000154025	SLC5A10
ENSG00000154898	CCDC144CP
ENSG00000156026	MCU
ENSG00000156966	B3GNT7
ENSG00000157837	SPPL3
ENSG00000160218	TRAPPC10
ENSG00000160326	SLC2A6
ENSG00000162739	SLAMF6
ENSG00000162980	ARL5A
ENSG00000163930	BAP1
ENSG00000164236	ANKRD33B
ENSG00000164871	SPAG11B
ENSG00000166073	GPR176
ENSG00000166484	MAPK7
ENSG00000166582	CENPV
ENSG00000166596	WDR16
ENSG00000166938	DIS3L
ENSG00000167524	SGK494
ENSG00000168614	NBPF9
ENSG00000168961	LGALS9
ENSG00000169040	PMCHL2
ENSG00000170142	UBE2E1

ENSG00000170298	LGLS9B
ENSG00000170324	FRMPD2
ENSG00000170791	CHCHD7
ENSG00000171298	GAA
ENSG00000171403	KRT9
ENSG00000171928	TVP23B
ENSG00000172794	RAB37
ENSG00000174738	NR1D2
ENSG00000175061	C17orf76-AS1
ENSG00000175061	SNORD65
ENSG00000176386	CDC26
ENSG00000176994	SMCR8
ENSG00000177602	GSG2
ENSG00000178184	PARD6G
ENSG00000178307	TMEM11
ENSG00000178342	KCNG2
ENSG00000178445	GLDC
ENSG00000178977	LINC00324
ENSG00000180787	ZFP3
ENSG00000181350	FAM211A
ENSG00000181619	GPR135
ENSG00000182315	MBD3L3
ENSG00000184635	ZNF93
ENSG00000185947	ZNF267
ENSG00000186603	HPDL
ENSG00000186825	C2orf27B
ENSG00000186832	KRT16
ENSG00000187607	ZNF286A
ENSG00000187984	ANKRD19P
ENSG00000188522	FAM83G
ENSG00000188933	USP32P1
ENSG00000196544	C17orf59
ENSG00000196990	FAM163B
ENSG00000197124	ZNF682
ENSG00000197223	C1D
ENSG00000197566	ZNF624
ENSG00000198890	PRMT6
ENSG00000203832	NBPF20
ENSG00000205718	MBD3L4
ENSG00000206609	SNORD116-11
ENSG00000206656	SNORD116-17
ENSG00000206688	SNORD116-18
ENSG00000207137	SNORD116-13
ENSG00000207174	SNORD116-15
ENSG00000207197	SNORD116-12
ENSG00000207263	SNORD116-16

ENSG00000207460	SNORD116-19
ENSG00000207693	MIR602
ENSG00000207839	MIR33B
ENSG00000207926	MIR135A1
ENSG00000213240	NOTCH2NL
ENSG00000214514	KRT42P
ENSG00000214822	KRT16P3
ENSG00000214856	KRT16P1
ENSG00000215356	ZNF705B
ENSG00000215372	ZNF705G
ENSG00000215458	LOC284837
ENSG00000215861	LOC653513
ENSG00000219607	PPP1R3G
ENSG00000220205	VAMP2
ENSG00000221288	MIR663B
ENSG00000221540	MIR1180
ENSG00000223380	SEC22B
ENSG00000223510	CDRT15
ENSG00000224074	LINC00691
ENSG00000226435	ANKRD18DP
ENSG00000226521	LOC100287072
ENSG00000226746	SMCR5
ENSG00000227255	CDRT15P2
ENSG00000227300	KRT16P2
ENSG00000076382	SPAG5
ENSG00000231256	C17orf105
ENSG00000231528	FAM225A
ENSG00000233327	USP32P2
ENSG00000072134	EPN2
ENSG00000169554	ZEB2
ENSG00000240024	LINC00888
ENSG00000243667	WDR92
ENSG00000250462	LRRC37BP1
ENSG00000252577	SCARNA20
ENSG00000255168	RP11-458D21.5
ENSG00000256229	ZNF486
ENSG00000256771	ZNF253
ENSG00000258472	RP11-192H23.4
ENSG00000258701	LINC00638
ENSG00000261126	RBFADN
ENSG00000261879	LOC100130950
ENSG00000263512	MIR4311
ENSG00000263641	MIR4777
ENSG00000263834	MIR4635
ENSG00000264173	MIR3175
ENSG00000264493	MIR4298

ENSG00000264624	MIR3615
ENSG00000264744	MIR3689C
ENSG00000264943	SH3GL1P2
ENSG00000265083	MIR3691
ENSG00000266433	TBC1D3P5
ENSG00000266719	MIR4676
ENSG00000266733	TBC1D29

**Supplementary Table S12: BRICKs genes enrichments. Table S12.1. STRING protein-protein interactions. Table S12.2. Enrichr GO Biological process. Table S12.3. Enrichr Chromosomal location**

**Table S12.1**

P-value	Interactions observed	Interactions expected	Proteins in the network
6.41E-12	82	3.50E+01	131

**Table S12.2**

Term	Overlap	P-value	Adjusted P-value	Z-score	Combined Score	Genes
negative regulation of multicellular organism growth (GO:0040015)	2/13	0.003787963	0.510487273	-2.72006165	1.828941089	BBS2;RAI1
fibril organization (GO:0097435)	2/14	0.004311708	0.510487273	-2.71061243	1.822587532	MFAP4;COL5A1
detection of light stimulus involved in sensory perception (GO:0050962)	2/16	0.005453376	0.510487273	-2.70152996	1.816480574	SDR16C5;CACNB4
detection of light stimulus involved in visual perception (GO:0050908)	2/16	0.005453376	0.510487273	-2.69757626	1.813822147	SDR16C5;CACNB4
negative regulation of protein kinase B signaling (GO:0051898)	2/28	0.014776359	0.510487273	-2.676383536	1.799572381	SLC9A3R1;FLCN

**Table S12.3**

Term	Overlap	P-value	Adjusted P-value	Z-score	Combined Score	Genes
chr17p11	29/193	0	0	-2.5986888	217.9704849	C17ORF76;TNFRSF13B;FLII; SHMT1;FLCN;RAI1;ALKBH5; RASD1;PRPSAP2;NT5M;ATP AF2;SREBF1;LRRC48;KSR1; USP22;TRPV2;MED9;C17OR F84;FAM106B;MFAP4;COPS 3;TOM1L2;MYO15A;ULK2;P EMT;C17ORF39;SMCR7;TRI M16;LLGL1
chr17p12	7/46	3.8047E-10	9.70187E-09	-3.0372371	56.03990076	ZNF232;UBB;DHRS7B;STX8; TOP3A;GRAP;COX10
chr17p13	7/328	0.000104393	0.001774689	-2.1966188	13.91366902	ZNF594;USP6;TMEM107;D RG2;ITGAE;AURKB;GAS7
chr2q22	2/27	0.003884708	0.049530021	-2.6388441	7.930191849	CACNB4;GTDC1

**Supplementary Table S13: Catalog of lymphoblastoid cell lines and 4C primers used in this study.**

**Table S13.1**

Sample code	Origin	Gender	Age	Copy Number of 17p11.2 RAI1 locus
GRAEM-control1	CIG, Lausanne	F	31	2
SOMAG-control2	CIG, Lausanne	F	28	2

**Table S13.2**

Viewpoint	Primer name	Sequence	Primer name	Sequence
RAI1	RAI1_4CadaptAAC_F	AATGATACGGCGACCACCGAACA CTCTTCCCTACACGACGCTCTTCC GATCTAACC GATTTCAGGGACCC ACT	RAI1_4Cadapt_R	CAAGCAGAAGACGGCATAAC GAGTCTGCCACCCCTTGTC
RAI1	RAI1_4CadaptGTG_F	AATGATACGGCGACCACCGAACA CTCTTCCCTACACGACGCTCTTCC GATCTGTG GATTTCAGGGACCC ACT	RAI1_4Cadapt_R	CAAGCAGAAGACGGCATAAC GAGTCTGCCACCCCTTGTC

footnote: the Illumina adapter tails are highlighted in red (forward primers) and blue (reverse primers), while the barcode is highlighted in green.

**Supplementary Table S14: Numbers of raw, excluded, mappable and mapped 4C reads for each viewpoint and sample.**

Viewpoint	Replicate	RawReads	ExcludedReads *reads that have been filtered, due to the presence of Undigested,Self-ligated and/or bait sequence	MappableReads *reads considered for the rest of the analysis (Bowtie)	MappedReads
RAI1	Replicate1	21'867'968	6'501'347	15'366'621	7'775'204
RAI1	Replicate2	43'592'203	11'281'662	32'310'541	12'163'326

**Supplementary Text**

**Detailed SMS patients' phenotypes**

BAB1604 was a 3-years-2-months old female with global developmental delay, behavioral difficulties, self-injurious behavior, and stereotypies including hand flapping and self-hugging. Sleep cycle was reportedly normal. The patient was reported to have always been hypotonic, and had a femur fracture at 11 months with an otherwise unremarkable skeletal survey. Family history was noncontributory. Growth parameters showed a head circumference at about 5th percentile despite height and weight around the 75th percentile. Features were somewhat coarse with deep-set eyes and synophrys, a small philtrum, thick lips with a cupid's bow, wide spaced teeth, a prominent chin and borderline low set ears. She had hypotonia, joint hypermobility, bilateral muscle wasting of her distal thighs and calves, and absent deep tendon reflexes (DTRs) in the lower extremities. Toenails were hypoplastic and she had bilateral 5th toe clinodactyly. Previous work-up included urine organic acids, plasma amino acids, oligosaccharides, mucopolysaccharides, very long chain fatty acids, ceruloplasmin, liver function tests, B12, folate, lactate, TSH, lead level, creatine kinase, high resolution chromosomes, and FISH for deletion in the Prader-Willi region - all normal. Ophthalmology exam was normal. Brain MRI demonstrated an asymmetric cranial vault but was otherwise normal. EMG demonstrated the



possibility of an axonal neuropathy consistent with Charcot-Marie-Tooth disease.

BAB1952 was a male evaluated at 13-years for global developmental delay, behavioral difficulties including aggressive outbursts and self-injurious behavior, nail biting, and sleep disturbances. He was reported to have had feeding difficulties as an infant, hypotonia, and a severe articulation disorder. MRI showed a mild increase in the CSF spaces around the cerebral hemispheres. Renal ultrasound demonstrated bilateral hydronephrosis with hydroureters, and bilateral ureteropelvic junction obstruction status post surgical correction.

BAB2134 was an 11-year old male with developmental delay, aggressive behavior and bilateral sensorineural hearing loss. He was born at 38 wks of gestation, and weighed 6 lbs 8 oz (appropriate for gestational age, AGA). The patient had slightly delayed motor development, with walking at 18 months. Speech and language development were also delayed, with first words spoken at 4 years. He was reported to have an abnormal sleep pattern, self-injurious behavior including pulling out his toenails, and aggression towards others. His head circumference was at the 40th percentile, height was 50th percentile, and weight was 90<sup>th</sup> percentile. Physical exam revealed frontal bossing, chubby cheeks, micrognathia, upslanting palpebral fissures, a flat bridge nose with scooped out root, and brachydactyly and clinodactyly of the 5th finger bilaterally.

BAB2293 was a male evaluated at 2-years-3-months for global developmental delay and behavioral difficulties. He was born by emergency C-section for failure to progress secondary to deflexed brow presentation. Birth weight was AGA; head circumference at birth was just above the 90<sup>th</sup> percentile. He had delayed motor development, with commando crawling at 18 mo, and was not yet walking when evaluated. Speech and language development were delayed, with a single word "mum" at 27 months. He exhibited aggressive behaviors, hand flapping, and self-injurious behaviors including biting his hands and toes, head banging, and pulling out his own hair. He had an abnormal sleep pattern and chronic constipation. Head CT was unremarkable aside from deformational

plagiocephaly and brachycephaly. Work-up included chromosome analysis, Fragile X testing, thyroid function studies and a urine metabolic screen, all of which resulted normal. Whole exome sequencing showed a variant in *KDM5C* that is associated with X-linked syndromic mental retardation Claes-Jensen type (OMIM#300534).

BAB2321 was a 12-year old male with developmental delay, ADHD, obsessive-compulsive disorder, and aggressive behavior. Height, weight and head circumference were symmetric, at the 5<sup>th</sup>-10<sup>th</sup> percentiles. The patient had an overall coarse appearance, and dysmorphic features included a narrow head, keel-shaped forehead, low anterior hair line, deep-set eyes, slight periorbital fullness, epicanthal folds, a wide mouth with everted tented upper lip, wide spaced teeth with malocclusion, macroglossia, slightly high-arched palate, marked prognathism, and prominent ears. Hands and feet were short with stubby digits, and prominent fingertip pads. Skin involvement included areas of psoriasis. Previous work-up including chromosome analysis, telomeres, Fragile X testing, and enzyme assays for lysosomal storage disorders resulted normal.

BAB2330 was a 5-years-8-months old female evaluated for acquired microcephaly, speech and language delay, and behavioral abnormalities including self-injurious behaviors. She had a disturbed sleep cycle that responded to melatonin. During infancy, seizure-like episodes were witnessed but EEG and brain imaging were normal. The patient was born at full term after an uncomplicated pregnancy. Birth weight was 6 lbs 10 oz (AGA). Weight and length during infancy were around the 5<sup>th</sup> percentile, but weight had increased to 97<sup>th</sup> percentile by age 5-years-8-months. Length at this time was 25<sup>th</sup> percentile. Head circumference, which had been microcephalic during toddler years (-4SD), was at 5<sup>th</sup>-10<sup>th</sup> percentile at 5.5 years. Physical exam showed prominent forehead, narrow eyes with mildly upslanting palpebral fissures, small nose, and somewhat thin upper lip. Brain MRI and EEG were normal, as were ophthalmology evaluation, chromosome analysis, FISH for 17p11.2 and 22q11.2 deletions, lead levels, thyroid function studies, plasma amino acids and

urine organic acids. Whole exome sequencing showed a *de novo* variant in *POGZ*<sup>55</sup>.

BAB2451 was a 10-year old female with developmental delay, intellectual disability, seizures, self-injurious behavior, obesity, onychotillomania, and narcolepsy. Physical examination showed macrocephaly and a down-turned mouth; otherwise, the patient was non-dysmorphic. MRI showed mild global volume loss.

BAB2474 was a 5-year old female with developmental delay, feeding problems and microcephaly. She was born at 33 wks gestation by C-section due to premature onset of labor, acute development of oligohydramnios and breech presentation. Birth weight was 4 lbs 1 oz (50<sup>th</sup> percentile), and Apgars were 9 and 9 at 1 and 5 minutes. The patient had difficulty feeding and failure to thrive, necessitating G-tube feeds for the first four years of life. She had motor delay, with walking at 3.5 years; and speech and language delay. She had seizure-like activity but a normal EEG and MRI. Ophthalmology diagnosed ocular motor apraxia, and she needed eyeglasses since age two years. She had bilateral externally rotated legs and flexible valgus deformity in the feet. She had new onset of hypertension at 5 yrs, with normal renal parenchyma per ultrasound; however, renal arteries were not visualized and follow up imaging was recommended. Previous work up included chromosome analysis, methylation studies for Angelman syndrome, and FISH for 17p11.2, 22q11.2, 15q11, 15q13 and 4p deletions - all negative. Whole exome sequencing showed a variant in *MAP2K2* that is associated with cardiofaciocutaneous syndrome (OMIM #615280).

BAB2492 was a 3-year old male with developmental delay, intellectual disability, hypotonia, seizure-like activity but normal EEG, feeding difficulties, failure to thrive and brachydactyly. Chromosome analysis showed 47,XYY.

BAB2540 was born after a normal pregnancy, at a birth weight of 4350 grams. She walked at 18 months. Speech development was severely delayed. She

exhibited aggressive behavior, self-injury, placing objects in orifices (nose and ears), and hand biting. She had amblyopia and normal hearing. Coarse facial features were apparent, and she had large hands. MRI was reportedly normal, as were chromosome analysis and FISH for 17p11.2 deletion. Whole exome sequencing showed a *de novo* variant in *CASK*, which is associated with mental retardation and microcephaly with pontine and cerebellar hypoplasia (OMIM #300749).

BAB2552 was a female evaluated at 5 years for developmental delay, a very short attention span and behavioral problems. She had an abnormal sleep pattern and self-injurious behaviors including self-biting. Physical exam showed deep-set eyes, and overall craniofacial features were felt to be consistent with Smith-Magenis syndrome (no details available). Whole exome sequencing showed a variant in *MECP2*, which is associated with MECP2 X-linked syndromic mental retardation 13 (OMIM #300055).

BAB2559 was an 8-year-old male evaluated for developmental delay and behavioral difficulties. He exhibited speech delay, aggressive behavior, ADHD, bipolar disorder, self-injurious behavior, and sleep disturbances. He was reportedly hypotonic during infancy, and had difficulty running at 8 years. Physical exam revealed poor dental alignment, obesity specifically in the lower abdomen and pubic area, and absent DTRs in the lower extremities. MRI showed mild widening of CSF spaces around the cerebral and cerebellar hemispheres and mild thinning of the corpus callosum. Previous evaluation was negative, including FISH for 22q11.2 deletion and telomere FISH.

BAB4947 was a male patient born to consanguineous parents of Turkish descent, with intellectual disability, seizures, behavioral disturbances, aggressive behavior, polyembolokoilomania, and onychotillomania. Physical appearance was striking for a Kabuki-like face. He had brachycephaly, arched eyebrows, long palpebral fissures with lateral eversion of the eyelids, a thick and everted upper lip, high arched palate, ear lobe hyperplasia, bilateral preauricular pits, bilateral single palmar creases, and brachydactyly. Chromosome analysis and FISH for

17p11.2 resulted normal. Whole exome sequencing showed compound heterozygote variants in *GLDC* and an inherited frameshift variant in *TCOF1*, which are associated with glycine encephalopathy (OMIM #605899) and Treacher Collins syndrome-1 (OMIM #154500), respectively.

### Acknowledgments

We thank the patients and families for their contribution to this study. We are grateful to the members of the Lausanne Genomic Technologies Facility for technical help. MNL was awarded an EMBO fellowship (ASTF 153-2015). CRB is an HHMI fellow of the Damon Runyon Cancer Research Foundation (DRG 2155-13). TH and WLC are supported by the NIH T32 GM07526 Medical Genetics Research Fellowship Program and the CPRIT RP140102 training Program, respectively. This work was supported by grants from the Swiss National Science Foundation (31003A\_160203) and the Simons Foundation (SFARI274424) to AR; the US National Human Genome Research Institute (NHGRI)/National Heart Lung and Blood Institute (NHLBI) grant number HG006542 to the Baylor-Hopkins Center for Mendelian Genomics; the Smith-Magenis Syndrome Research Foundation (SMSRF); and the National Institute of Neurological Disorders and Stroke (NINDS) NS058529 to JRL. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

### Competing interest

JRL has stock ownership in 23andMe, is a paid consultant for Regeneron Pharmaceuticals, has stock options in Lasergen, Inc., is a member of the Scientific Advisory Board of Baylor Miraca Genetics Laboratories, and is a co-inventor on multiple United States and European patents related to molecular diagnostics for inherited neuropathies, eye diseases and bacterial genomic fingerprinting.

Baylor College of Medicine (BCM) and Miraca Holdings Inc. have formed a joint venture with shared ownership and governance of the Baylor Miraca Genetics Laboratories (BMGL), which performs clinical exome sequencing. The Department of Molecular and Human Genetics at Baylor College of Medicine derives revenue from the chromosomal microarray analysis (CMA) and clinical exome sequencing offered in the Baylor Miraca Genetics Laboratory (BMGL; <http://www.bmgl.com/BMGL/Default.aspx> webcite). The remaining authors declare that they have no competing interests.

## References

1. Smith ACM, Boyd KE, Elsea SH, Finucane BM, Haas-Givler B, Gropman A *et al.* Smith-Magenis Syndrome. In: Pagon RA, Adam MP, Ardinger HH, Wallace SE, Amemiya A, Bean L JH *et al.* (eds). *GeneReviews(R)*: Seattle (WA), 1993.
2. Elsea SH, Girirajan S. Smith-Magenis syndrome. *European journal of human genetics* : *EJHG* 2008; **16**(4): 412-421.
3. Laje G, Morse R, Richter W, Ball J, Pao M, Smith AC. Autism spectrum features in Smith-Magenis syndrome. *American journal of medical genetics Part C, Seminars in medical genetics* 2010; **154C**(4): 456-462.
4. Moss J, Oliver C, Arron K, Burbidge C, Berg K. The prevalence and phenomenology of repetitive behavior in genetic syndromes. *Journal of autism and developmental disorders* 2009; **39**(4): 572-588.
5. Dykens EM, Finucane BM, Gayley C. Brief report: cognitive and behavioral profiles in persons with Smith-Magenis syndrome. *Journal of autism and developmental disorders* 1997; **27**(2): 203-211.
6. Dykens EM, Smith AC. Distinctiveness and correlates of maladaptive behaviour in children and adolescents with Smith-Magenis syndrome. *Journal of intellectual disability research* : *JIDR* 1998; **42 ( Pt 6)**: 481-489.
7. Finucane BM, Konar D, Haas-Givler B, Kurtz MB, Scott CI, Jr. The spasmodic upper-body squeeze: a characteristic behavior in Smith-Magenis syndrome. *Developmental medicine and child neurology* 1994; **36**(1): 78-83.
8. Sloneem J, Oliver C, Udwin O, Woodcock KA. Prevalence, phenomenology, aetiology and predictors of challenging behaviour in Smith-Magenis syndrome. *Journal of intellectual disability research* : *JIDR* 2011; **55**(2): 138-151.
9. Bi W, Saifi GM, Shaw CJ, Walz K, Fonseca P, Wilson M *et al.* Mutations of RAI1, a PHD-containing protein, in nondeletion patients with Smith-Magenis syndrome. *Human genetics* 2004; **115**(6): 515-524.
10. Girirajan S, Elsas LJ, 2nd, Devriendt K, Elsea SH. RAI1 variations in Smith-Magenis syndrome patients without 17p11.2 deletions. *J Med Genet* 2005; **42**(11): 820-828.
11. Slager RE, Newton TL, Vlangos CN, Finucane B, Elsea SH. Mutations in RAI1 associated with Smith-Magenis syndrome. *Nat Genet* 2003; **33**(4): 466-468.
12. Vieira GH, Rodriguez JD, Carmona-Mora P, Cao L, Gamba BF, Carvalho DR *et al.* Detection of classical 17p11.2 deletions, an atypical deletion and RAI1 alterations in patients with features suggestive of Smith-Magenis syndrome. *European journal of human genetics* : *EJHG* 2012; **20**(2): 148-154.

13. Mullegama SV, Pugliesi L, Burns B, Shah Z, Tahir R, Gu Y *et al.* MBD5 haploinsufficiency is associated with sleep disturbance and disrupts circadian pathways common to Smith-Magenis and fragile X syndromes. *European journal of human genetics : EJHG* 2015; **23**(6): 781-789.
14. Williams SR, Aldred MA, Der Kaloustian VM, Halal F, Gowans G, McLeod DR *et al.* Haploinsufficiency of HDAC4 causes brachydactyly mental retardation syndrome, with brachydactyly type E, developmental delays, and behavioral problems. *American journal of human genetics* 2010; **87**(2): 219-228.
15. Williams SR, Girirajan S, Tegay D, Nowak N, Hatchwell E, Elsea SH. Array comparative genomic hybridisation of 52 subjects with a Smith-Magenis-like phenotype: identification of dosage sensitive loci also associated with schizophrenia, autism, and developmental delay. *J Med Genet* 2010; **47**(4): 223-229.
16. Derwinska K, Mierzevska H, Goszczanska A, Szczepanik E, Xia Z, Kusmierska K *et al.* Clinical improvement of the aggressive neurobehavioral phenotype in a patient with a deletion of PITX3 and the absence of L-DOPA in the cerebrospinal fluid. *Am J Med Genet B Neuropsychiatr Genet* 2012; **159B**(2): 236-242.
17. Liu P, Lacia M, Zhang F, Withers M, Hastings PJ, Lupski JR. Frequency of nonallelic homologous recombination is correlated with length of homology: evidence that ectopic synapsis precedes ectopic crossing-over. *Am J Hum Genet* 2011; **89**(4): 580-588.
18. Wiszniewska J, Bi W, Shaw C, Stankiewicz P, Kang SH, Pursley AN *et al.* Combined array CGH plus SNP genome analyses in a single assay for optimized clinical testing. *European journal of human genetics : EJHG* 2014; **22**(1): 79-87.
19. Yuan B, Pehlivan D, Karaca E, Patel N, Charng WL, Gambin T *et al.* Global transcriptional disturbances underlie Cornelia de Lange syndrome and related phenotypes. *The Journal of clinical investigation* 2015; **125**(2): 636-651.
20. Guex N, Peitsch MC. SWISS-MODEL and the Swiss-PdbViewer: an environment for comparative protein modeling. *Electrophoresis* 1997; **18**(15): 2714-2723.
21. Schwede T, Kopp J, Guex N, Peitsch MC. SWISS-MODEL: An automated protein homology-modeling server. *Nucleic acids research* 2003; **31**(13): 3381-3385.
22. Kamburov A, Lawrence MS, Polak P, Leshchiner I, Lage K, Golub TR *et al.* Comprehensive assessment of cancer missense mutation clustering in protein structures. *Proceedings of the National Academy of Sciences of the United States of America* 2015.
23. Kim CA, Berg JM. A 2.2 Å resolution crystal structure of a designed zinc finger protein bound to DNA. *Nature structural biology* 1996; **3**(11): 940-945.
24. Ohren JF, Chen H, Pavlovsky A, Whitehead C, Zhang E, Kuffa P *et al.* Structures of human MAP kinase kinase 1 (MEK1) and MEK2 describe novel noncompetitive kinase inhibition. *Nature structural & molecular biology* 2004; **11**(12): 1192-1197.



25. Fischmann TO, Smith CK, Mayhood TW, Myers JE, Reichert P, Mannarino A *et al.* Crystal structures of MEK1 binary and ternary complexes with nucleotides and inhibitors. *Biochemistry* 2009; **48**(12): 2661-2674.
26. Hasse D, Andersson E, Carlsson G, Maslobov A, Hagemann M, Bauwe H *et al.* Structure of the homodimeric glycine decarboxylase P-protein from *Synechocystis* sp. PCC 6803 suggests a mechanism for redox regulation. *The Journal of biological chemistry* 2013; **288**(49): 35333-35345.
27. Daniels DL, Cohen AR, Anderson JM, Brunger AT. Crystal structure of the hCASK PDZ domain reveals the structural basis of class II PDZ domain target recognition. *Nature structural biology* 1998; **5**(4): 317-325.
28. Li Y, Spangenberg O, Paarmann I, Konrad M, Lavie A. Structural basis for nucleotide-dependent regulation of membrane-associated guanylate kinase-like domains. *The Journal of biological chemistry* 2002; **277**(6): 4159-4165.
29. Cornish AJ, Markowetz F. SANTA: quantifying the functional content of molecular networks. *PLoS computational biology* 2014; **10**(9): e1003808.
30. Bi W, Ohyama T, Nakamura H, Yan J, Visvanathan J, Justice MJ *et al.* Inactivation of Rai1 in mice recapitulates phenotypes observed in chromosome engineered mouse models for Smith-Magenis syndrome. *Hum Mol Genet* 2005; **14**(8): 983-995.
31. Irizarry RA, Hobbs B, Collin F, Beazer-Barclay YD, Antonellis KJ, Scherf U *et al.* Exploration, normalization, and summaries of high density oligonucleotide array probe level data. *Biostatistics* 2003; **4**(2): 249-264.
32. Noordermeer D, Leleu M, Splinter E, Rougemont J, De Laat W, Duboule D. The dynamic architecture of Hox gene clusters. *Science* 2011; **334**(6053): 222-225.
33. Consortium EP. An integrated encyclopedia of DNA elements in the human genome. *Nature* 2012; **489**(7414): 57-74.
34. Loviglio MN, Leleu M, Caldeira Da Silva MI, Mannik K, Passeggeri M, van der Werf I *et al.* Chromosomal contacts connect loci associated with autism, BMI and head circumference phenotypes *submitted* 2014.
35. Gheldof N, Leleu M, Noordermeer D, Rougemont J, Reymond A. Detecting long-range chromatin interactions using the chromosome conformation capture sequencing (4C-seq) method. *Methods Mol Biol* 2012; **786**: 211-225.
36. David FP, Delafontaine J, Carat S, Ross FJ, Lefebvre G, Jarosz Y *et al.* HTSstation: a web application and open-access libraries for high-throughput sequencing data analysis. *PloS one* 2014; **9**(1): e85879.

37. Tolhuis B, Blom M, Kerkhoven RM, Pagie L, Teunissen H, Nieuwland M *et al.* Interactions among Polycomb domains are guided by chromosome architecture. *PLoS genetics* 2011; **7**(3): e1001343.
38. Lieberman-Aiden E, van Berkum NL, Williams L, Imakaev M, Ragozy T, Telling A *et al.* Comprehensive mapping of long-range interactions reveals folding principles of the human genome. *Science* 2009; **326**(5950): 289-293.
39. de Wit E, Braunschweig U, Greil F, Bussemaker HJ, van Steensel B. Global chromatin domain organization of the Drosophila genome. *PLoS genetics* 2008; **4**(3): e1000045.
40. Franceschini A, Szklarczyk D, Frankild S, Kuhn M, Simonovic M, Roth A *et al.* STRING v9.1: protein-protein interaction networks, with increased coverage and integration. *Nucleic acids research* 2013; **41**(Database issue): D808-815.
41. Alexa A, Rahnenfuhrer J. topGO: Enrichment analysis for Gene Ontology. In: BioConductor (ed).2010.
42. Huang da W, Sherman BT, Lempicki RA. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nature protocols* 2009; **4**(1): 44-57.
43. Reimand J, Arak T, Vilo J. g:Profiler--a web server for functional interpretation of gene lists (2011 update). *Nucleic acids research* 2011; **39**(Web Server issue): W307-315.
44. Chen EY, Tan CM, Kou Y, Duan Q, Wang Z, Meirelles GV *et al.* Enrichr: interactive and collaborative HTML5 gene list enrichment analysis tool. *BMC Bioinformatics* 2013; **14**: 128.
45. Rao SS, Huntley MH, Durand NC, Stamenova EK, Bochkov ID, Robinson JT *et al.* A 3D map of the human genome at kilobase resolution reveals principles of chromatin looping. *Cell* 2014; **159**(7): 1665-1680.
46. Dixon JR, Selvaraj S, Yue F, Kim A, Li Y, Shen Y *et al.* Topological domains in mammalian genomes identified by analysis of chromatin interactions. *Nature* 2012; **485**(7398): 376-380.
47. Greenberg F, Guzzetta V, Montes de Oca-Luna R, Magenis RE, Smith AC, Richter SF *et al.* Molecular analysis of the Smith-Magenis syndrome: a possible contiguous-gene syndrome associated with del(17)(p11.2). *Am J Hum Genet* 1991; **49**(6): 1207-1218.
48. Juyal RC, Figuera LE, Hauge X, Elsea SH, Lupski JR, Greenberg F *et al.* Molecular analyses of 17p11.2 deletions in 62 Smith-Magenis syndrome patients. *Am J Hum Genet* 1996; **58**(5): 998-1007.
49. Moncla A, Piras L, Arbex OF, Muscatelli F, Mattei MG, Mattei JF *et al.* Physical mapping of microdeletions of the chromosome 17 short arm associated with Smith-Magenis syndrome. *Human genetics* 1993; **90**(6): 657-660.

50. Shaw CJ, Withers MA, Lupski JR. Uncommon deletions of the Smith-Magenis syndrome region can be recurrent when alternate low-copy repeats act as homologous recombination substrates. *Am J Hum Genet* 2004; **75**(1): 75-81.
51. Chen KS, Manian P, Koeth T, Potocki L, Zhao Q, Chinault AC *et al.* Homologous recombination of a flanking repeat gene cluster is a mechanism for a common contiguous gene deletion syndrome. *Nature genetics* 1997; **17**(2): 154-163.
52. Vilboux T, Ciccone C, Blancato JK, Cox GF, Deshpande C, Introne WJ *et al.* Molecular analysis of the Retinoic Acid Induced 1 gene (RAI1) in patients with suspected Smith-Magenis syndrome without the 17p11.2 deletion. *PLoS one* 2011; **6**(8): e22861.
53. Battaglia A. Commentary: Recognizing syndromes with overlapping features: How difficult is it? Considerations generated by the article on differential diagnosis of Smith-Magenis syndrome by Vieira and colleagues. *American journal of medical genetics Part A* 2011; **155A**(5): 986-987.
54. Vieira GH, Rodriguez JD, Boy R, de Paiva IS, DuPont BR, Moretti-Ferreira D *et al.* Differential diagnosis of Smith-Magenis syndrome: 1p36 deletion syndrome. *American journal of medical genetics Part A* 2011; **155A**(5): 988-992.
55. White J, Beck CR, Harel T, Posey JE, Jhangiani SN, Tang S *et al.* POGZ truncating alleles cause syndromic intellectual disability. *Genome medicine* 2016; **8**(1): 3.
56. Edelman EA, Girirajan S, Finucane B, Patel PI, Lupski JR, Smith AC *et al.* Gender, genotype, and phenotype differences in Smith-Magenis syndrome: a meta-analysis of 105 cases. *Clinical genetics* 2007; **71**(6): 540-550.
57. Micale L, Augello B, Fusco C, Selicorni A, Loviglio MN, Silengo MC *et al.* Mutation spectrum of MLL2 in a cohort of Kabuki syndrome patients. *Orphanet journal of rare diseases* 2011; **6**: 38.
58. Gomot M, Gendrot C, Verloes A, Raynaud M, David A, Yntema HG *et al.* MECP2 gene mutations in non-syndromic X-linked mental retardation: phenotype-genotype correlation. *American journal of medical genetics Part A* 2003; **123A**(2): 129-139.
59. Menche J, Sharma A, Kitsak M, Ghiassian SD, Vidal M, Loscalzo J *et al.* Disease networks. Uncovering disease-disease relationships through the incomplete interactome. *Science* 2015; **347**(6224): 1257601.
60. Greene CS, Krishnan A, Wong AK, Ricciotti E, Zelaya RA, Himmelstein DS *et al.* Understanding multicellular function and disease with human tissue-specific networks. *Nature genetics* 2015; **47**(6): 569-576.
61. Ricard G, Molina J, Chrast J, Gu W, Gheldof N, Pradervand S *et al.* Phenotypic consequences of copy number variation: insights from Smith-Magenis and Potocki-Lupski syndrome mouse models. *PLoS Biol* 2010; **8**(11): e1000543.

62. de Laat W, Duboule D. Topology of mammalian developmental enhancers and their regulatory landscapes. *Nature* 2013; **502**(7472): 499-506.
63. Fanucchi S, Shibayama Y, Burd S, Weinberg MS, Mhlanga MM. Chromosomal contact permits transcription between coregulated genes. *Cell* 2013; **155**(3): 606-620.
64. Simonis M, Klous P, Splinter E, Moshkin Y, Willemsen R, de Wit E *et al.* Nuclear organization of active and inactive chromatin domains uncovered by chromosome conformation capture-on-chip (4C). *Nat Genet* 2006; **38**(11): 1348-1354.
65. Simonis M, Kooren J, de Laat W. An evaluation of 3C-based methods to capture DNA interactions. *Nature methods* 2007; **4**(11): 895-901.
66. Gheldof N, Witwicki RM, Migliavacca E, Leleu M, Didelot G, Harewood L *et al.* Structural variation-associated expression changes are paralleled by chromatin architecture modifications. *PloS one* 2013; **8**(11): e79973.
67. de Wit E, de Laat W. A decade of 3C technologies: insights into nuclear organization. *Genes Dev* 2012; **26**(1): 11-24.
68. Deciphering Developmental Disorders S. Large-scale discovery of novel genetic causes of developmental disorders. *Nature* 2015; **519**(7542): 223-228.
69. Vissers LE, Gilissen C, Veltman JA. Genetic studies in intellectual disability and related disorders. *Nature reviews Genetics* 2016; **17**(1): 9-18.
70. D'Angelo D, Lebon S, Chen Q, Martin-Brevet S, Snyder LG, Hippolyte L *et al.* Defining the Effect of the 16p11.2 Duplication on Cognition, Behavior, and Medical Comorbidities. *JAMA Psychiatry* 2016; **73**(1): 20-30.
71. Mannik K, Magi R, Mace A, Cole B, Guyatt AL, Shihab HA *et al.* Copy number variations and cognitive phenotypes in unselected populations. *JAMA : the journal of the American Medical Association* 2015; **313**(20): 2044-2054.
72. Kochinke K, Zweier C, Nijhof B, Fenckova M, Cizek P, Honti F *et al.* Systematic Phenomics Analysis Deconvolutes Genes Mutated in Intellectual Disability into Biologically Coherent Modules. *Am J Hum Genet* 2016; **98**(1): 149-164.
73. Bardsley MZ, Kowal K, Levy C, Gosek A, Ayari N, Tartaglia N *et al.* 47,XXX syndrome: clinical phenotype and timing of ascertainment. *The Journal of pediatrics* 2013; **163**(4): 1085-1094.
74. Fromer M, Pocklington AJ, Kavanagh DH, Williams HJ, Dwyer S, Gormley P *et al.* De novo mutations in schizophrenia implicate synaptic networks. *Nature* 2014; **506**(7487): 179-184.

75. Gilissen C, Hehir-Kwa JY, Thung DT, van de Vorst M, van Bon BW, Willemsen MH *et al.* Genome sequencing identifies major causes of severe intellectual disability. *Nature* 2014; **511**(7509): 344-347.
76. Iossifov I, O'Roak BJ, Sanders SJ, Ronemus M, Krumm N, Levy D *et al.* The contribution of de novo coding mutations to autism spectrum disorder. *Nature* 2014; **515**(7526): 216-221.
77. Iossifov I, Ronemus M, Levy D, Wang Z, Hakker I, Rosenbaum J *et al.* De novo gene disruptions in children on the autistic spectrum. *Neuron* 2012; **74**(2): 285-299.
78. Isidor B, Kury S, Rosenfeld JA, Besnard T, Schmitt S, Joss S *et al.* De Novo Truncating Mutations in the kinetochore-microtubules attachment gene CHAMP1 Cause Syndromic Intellectual Disability. *Human mutation* 2016.
79. Neale BM, Kou Y, Liu L, Ma'ayan A, Samocha KE, Sabo A *et al.* Patterns and rates of exonic de novo mutations in autism spectrum disorders. *Nature* 2012; **485**(7397): 242-245.
80. Tan B, Zou Y, Zhang Y, Zhang R, Ou J, Shen Y *et al.* A novel de novo POGZ mutation in a patient with intellectual disability. *Journal of human genetics* 2016.
81. De Rubeis S, He X, Goldberg AP, Poultney CS, Samocha K, Ercument Cicek A *et al.* Synaptic, transcriptional and chromatin genes disrupted in autism. *Nature* 2014; **515**(7526): 209-215.
82. de Ligt J, Willemsen MH, van Bon BW, Kleefstra T, Yntema HG, Kroes T *et al.* Diagnostic exome sequencing in persons with severe intellectual disability. *The New England journal of medicine* 2012; **367**(20): 1921-1929.
83. Grozeva D, Carss K, Spasic-Boskovic O, Tejada MI, Gecz J, Shaw M *et al.* Targeted Next-Generation Sequencing Analysis of 1,000 Individuals with Intellectual Disability. *Human mutation* 2015; **36**(12): 1197-1204.
84. Rauch A, Wieczorek D, Graf E, Wieland T, Ende S, Schwarzmayr T *et al.* Range of genetic mutations associated with severe non-syndromic sporadic intellectual disability: an exome sequencing study. *Lancet* 2012; **380**(9854): 1674-1682.
85. Redin C, Gerard B, Lauer J, Herenger Y, Muller J, Quartier A *et al.* Efficient strategy for the molecular diagnosis of intellectual disability using targeted high-throughput sequencing. *Journal of medical genetics* 2014; **51**(11): 724-736.
86. Henrichsen CN, Vinckenbosch N, Zollner S, Chaignat E, Pradervand S, Schutz F *et al.* Segmental copy number variation shapes tissue transcriptomes. *Nature genetics* 2009; **41**(4): 424-429.

## Thesis discussion and future plans

Copy number variants (CNVs) represent a frequent kind of lesion in human genetic disorders, typically affecting multiple genes, regulatory elements, and genomic structures at the same time. The difficulty of investigating these complex loci, often spanning over hundreds of Kilobases or even several Megabases, is due to the necessity of taking into account all the potential contributing mechanisms, assign causality to certain driver(s) and identify complex additive and/or multiplicative relationships among genes<sup>1</sup>. Another layer of complexity is added by the 3D spatial organization of the locus of interest, often overlooked prior to the advent of 3C techniques, and that can be profoundly altered in case of large rearrangements, triggering a cascade of perturbations in the regulatory landscape of the region<sup>2</sup> and impacting the phenotype of the carriers.

The three chapters of this thesis approached the problem of interpreting copy number variants in different, and sometimes complementary, ways. **Chapters 1** and **2** reported the investigations concerning a very well-studied cytoband on the short arm of chromosome 16, the 16p11.2, and in particular two non-overlapping rearrangement-prone loci mapping less than 1 Mb apart, the 16p11.2 distal 220kb BP2-BP3 and the 600kb proximal BP4-BP5 intervals. The interest in these CNVs stems from the observation of their impact on mirror phenotypes of body mass index (BMI) and head circumference (HC), and their association with neurodevelopmental disorders, in particular autism spectrum disorder (ASD)<sup>3-20</sup>. We asked ourselves whether this remarkable similarity in the phenotypic manifestations could originate from a cross-talk between the two loci at the chromatin level (**Chapter 1**), or if the “interaction” could occur at the genetic level, between the identified driver(s) and modifier genes of the observed phenotype in the zebrafish *in vivo* model (**Chapter 2**), or whether both non-mutually exclusive scenarios could take place. **Chapter 3** addressed a different question in a different locus, i.e. the 17p11.2, whose deletion, together with the deletion and point mutations of the encompassed gene *RAI1*, is implicated in the neurobehavioral manifestations of the Smith-Magenis

syndrome<sup>21-30</sup>. We explored an interesting case of genetic and, potentially, also phenotypic heterogeneity, that required a comprehensive approach spanning from protein modeling to network analysis through 3D spatial genomics to identify additional players in the *RAI1*- disease network.

The results reported in **Chapters 1** and **2** shed some light on the interpretation of the phenotypic similarity between the 600kb BP4-BP5 and 220kb BP2-BP3 at 16p11.2, but leave many open questions and challenges for future investigations. The 4C-seq analyses, together with validation from FISH and supporting results in Hi-C published datasets<sup>31,32</sup>, suggests that these two regions share the same chromatin compartment or “megadomain”<sup>32</sup>, meaning that loci in this interval exhibit the same genome-wide contact pattern. This open chromatin compartment spans approximately 4Mb, from ~28Mb to ~32Mb, encompassing both intervals and, more proximally, a Zinc finger genes-rich region. The spatial proximity of these two loci is paralleled by other convergent observations, for example the coordinated behavior in terms of gene expression, enrichment of specific molecular marks, and the sharing of a set of downstream targets, but also the concurrent impact of the change of copy number of one interval on chromatin organization and transcript levels<sup>33,34</sup> of the other.

Our present hypothesis is that the spatial proximity reflects “proximity in the function” of these loci, supported by fact that the genomic regions connected by our selected set of viewpoints revealed enrichment in relevant pathways and ASD-associated genes, similarly to what reported by Cai and colleagues about 8q24 prostate cancer risk locus, where 4C-seq was able to uncover contacts between key genes and pathways involved in prostate cancer onset and progression<sup>35</sup>. Therefore, we can exploit 4C-seq contacts to retrieve a list of potentially biologically significant regions that are close in the 3D space and are possibly implicated in the regulation of the same pathway(s) and contributing to the same phenotypical outcome(s) when perturbed/rearranged/mutated.

Further investigations in this sense are required, and are currently in progress, to evaluate whether the contacts found in human lymphoblastoid cell lines could be recapitulated in a more relevant tissue (i.e. brain cortex compared to liver) and embryonic developmental stage(s), taking advantage of the published mouse

model of the 600kb BP4-BP5 deletion and duplication produced by Yann Hérault laboratory at IGBMC, Strasbourg<sup>36</sup>. At the same, we propose to use an enhancement of Hi-C, the capture Hi-C (cHi-C)<sup>37</sup>, using the target sequence enrichment of approximately 3Mb, encompassing both the 220kb BP2-BP3 and the 600kb BP4-BP5 loci, to capture the interactions profile of the entire region of interest and not to limit the analysis to selected “viewpoints” fragments.

The other hypothesis we had initially made on the possible “cross-talk” scenarios between BP2-BP3 and BP4-BP5 was their genetic interaction, involving the additive and/or synergistic activity of the genes encompassed by the two intervals (**Chapter 2**). First, we identified a candidate gene for the head circumference phenotype of the 220kb BP2-BP3 CNV syndrome, testing each encompassed gene *in vivo* and taking advantage of the zebrafish animal model, successfully used for similar studies in several recent publications<sup>38-40</sup>. Our data interestingly pointed at the linker for activation of T-cells (*LAT*) as the only gene in the region whose overexpression was able to recapitulate the microcephaly features associated with the BP2-BP3 duplication. This result was extremely surprising, given the well-established role of this molecule in T-cell receptor transduction and, to the best of our knowledge, its exclusive implication in immune system-related disorders, i.e. autoimmunity and lymphoproliferative disease<sup>41, 42</sup>. *LAT* impacts cell proliferation, neuronal maturation and axonal projection organization in the zebrafish head, similarly to what previously revealed for the head size phenotype’s driver gene of the BP4-BP5 CNVs, the Potassium Channel Tetramerization Domain Containing 13 (*KCTD13*)<sup>39</sup>. We hypothesized and confirmed that the reduction in cell proliferation, observed upon overexpression of both genes individually, exhibits an increased severity when the fish are exposed to the pairwise ‘cocktail’ combination of the two transcripts, suggesting an “additive” rather than “epistatic” effect for the interplay between these two genes. However, Golzio and colleagues had shown that the expressivity of the reciprocal phenotype induced by *KCTD13* overexpression and suppression could be modulated by other two genes mapping at BP4-P5, the Mitogen-Activated Protein Kinase 3 (*MAPK3*) and Major Vault Protein (*MVP*). The same two modifier genes were also able to enhance the



severity of the phenotype triggered by *LAT*'s increased dosage. Remarkably, these transcripts do not significantly affect in any direction the count of proliferating cells in the brain when overexpressed or suppressed individually<sup>39</sup>. Therefore, *MAPK3* and *MVP* act in “cis-epistasis” with *KCTD13* at the same CNV locus, the BP4-BP5 interval at 16p11.2, and in “trans-epistasis” with *LAT*, which is located 1Mb away at BP2-BP3. *KCTD13* was recently proposed to affect important neuronal pathways through two interconnected modules<sup>43</sup> with different functions: one involved in DNA replication, synthesis, and repair, primarily observed in the prefrontal and motor-sensory cortex during late mid-fetal development; and the other involved in the formation of E3 ubiquitin ligase complexes, which is primarily observed in the parietal, temporal, and occipital cortex during late mid-fetal development. The latter functionally-related group of interacting proteins consists of *KCTD13* and *Cul3-TNFAIP1-KCTD10*, and regulates the ubiquitination and degradation of the small GTPase *RhoA*, which, in turn, is a major regulator of the actin cytoskeleton and cell migration and plays an important role in neuronal development, including neurite outgrowth, axon pathfinding, neuronal migration, dendritic spine formation, and maintenance<sup>43</sup>. T cell receptor (TCR) activation results in the tyrosine phosphorylation of *LAT*, and direct interaction with several proteins, which mediates the indirect interaction of *LAT* with *SLP-76* and *Vav1*<sup>44</sup>. It is tempting to speculate that, since *Vav* proteins are known to switch to active state *Rho-GTP* family members, in particular *RhoA*, *Rac1* and *Cdc42*<sup>45</sup>, *LAT* and *KCTD13* might act via a similar mechanism by influencing *Rho* proteins levels and, consequently, *Rho*-mediated processes of cytoskeletal rearrangement, which could, in turn, regulate cellular processes that influence head and body size during development. Similarly, the signaling pathways activated by the stimulation of the TCR and regulated by *LAT* result in a rise in intracellular calcium, the induction of a number of transcription factors and the activation of mitogen-activated protein kinase (MAPK) cascades<sup>44</sup>. *LAT* and *KCTD13*'s modulator *MAPK3* (also known as *ERK1*) plays an essential role in this cascade<sup>46</sup>. *MVP* is postulated to act as an ERK scaffold that modulates MAPK/ERK signaling<sup>47</sup> and evidence is accumulating that it might be involved in the regulation of this and other important cell signaling pathways, including the PI3K/Akt, also activated by TCR stimulation via *LAT*, based on the

finding that MVP binds several phosphatases and kinases including PTEN<sup>48</sup>, SHP-2<sup>49</sup>, as well as Erk itself. In support of the possible interplay between these three BP4-BP5 mapping loci and *LAT*, Blumenthal and colleagues also showed that copy number variants of the BP4-BP5 interval results in *LAT*'s differential expression in human lymphoblastoid cell lines<sup>33</sup>.

Next step will be to assess whether the activity on brain size revealed in zebrafish for *LAT* can be recapitulated in other organisms and to evaluate if the neurogenic defect is consistent across species. We are currently testing the *LAT*-deficient mice model published by Zhang's lab in 1999<sup>50, 51</sup>, where potentially existing morphological perturbation in the brain might have been missed by previous investigations as only immune system-related phenotypes were plausibly evaluated and reported. This will also help to elucidate whether the phenotype is driven by a T-cells dependent mechanism, since T-cells deficiencies have been suggested to affect brain function and cognitive processes<sup>52</sup>, or not, as reported in *CD247* and *ZAP-70*, for which a novel role was recognized in the CNS<sup>53</sup>. In parallel, we will determine whether the patients carrying the 220kb BP2-BP3 rearrangement in our cohort show any immune-related anomaly, especially signs of dysfunctional adaptive cellular immune function through T-cells activation profiling<sup>54</sup>.

In **Chapter 3** we presented the results of the characterization of a cohort of patients diagnosed with Smith-Magenis syndrome. While the vast majority of them displayed the common heterozygous deletion of *RAI1* or a predicted deleterious variant of the *RAI1* gene, a small subset of individuals was further investigated by whole exome sequencing (WES), leading to the identification of potentially causative variants in genes associated with MOWS, MICPCH, KABUK1, GCE, MRXS13, CFC4, MRXSC and a new ID syndrome<sup>55</sup>, respectively, as well as a 47, XYY karyotype. The challenge of this project, which was approached with a combination of literature mining, co-expression data and chromatin interactions networks, transcriptome profiling of *Rai1*<sup>-/-</sup> animal models and variant modeling through X-ray structures and/or cryo-EM modeling, whenever possible, was to determine the contribution of genetic and phenotypic variability, using a new conceptual framework to describe diseases in their

complexity, for example in view of the pleiotropic relationship among different human diseases and consequent comorbidities. This framework needs to take into account the high multifactorial nature of the causes of disease pathogenesis<sup>56, 57</sup>, including genetic variation, epigenetic modifications, chromatin structure perturbations<sup>2, 58</sup> and several types of genome-environment perturbations. Our results support the existence of a common “biological module”<sup>59</sup> or “disease network”<sup>60</sup> linking the identified genes with *RAI1*, which is in agreement with the idea that genes implicated in the same or related diseases often cluster in the protein-protein interactions (PPI) network or form connected modules within networks<sup>61-63</sup>. Furthermore, recent publications reported cases of mutations occurring at different interaction interfaces in the same protein helping to elucidate how a single gene could be involved in multiple disorders (i.e. pleiotropy) or in disorders with multiple distinct modes of inheritance<sup>64, 65</sup>. It is clear that also the identification of modifier genes, additionally to the main phenotype driver, as in the case of the genes *MAPK3* and *MVP* presented in **Chapter 2**, can shed light on the disease networks, especially given the increased awareness that there are very few, if any, truly single-gene diseases<sup>57</sup>.

In the next future, we can hypothesize that the driving force for the discovery and the deepening of disease networks will be the enhanced ability for high throughput analysis of genomes and molecular phenotype analysis, with a major need for more sophisticated technologies to be used in this effort. The biggest challenge will be the integration of the amount of information derived from, for example, regulatory, protein-protein interactions, proteomic and metabolic networks. These networks operate in very different time-scales, also across different species, and we’re still lacking good bioinformatics and computational tools that would be useful for this integration. The challenge of characterizing and understanding ‘emergent properties’ of these systems will be at the heart of future biological research. This will aid the search of disease causing genes and mutations, and thus the unraveling of complex cases of genetic and phenotypic heterogeneity.

## Thesis discussion's references

1. Golzio C, Katsanis N. Genetic architecture of reciprocal CNVs. *Curr Opin Genet Dev* 2013; **23**(3): 240-248.
2. Lupianez DG, Spielmann M, Mundlos S. Breaking TADs: How Alterations of Chromatin Domains Result in Disease. *Trends Genet* 2016.
3. Bachmann-Gagescu R, Mefford HC, Cowan C, Glew GM, Hing AV, Wallace S *et al.* Recurrent 200-kb deletions of 16p11.2 that include the SH2B1 gene are associated with developmental delay and obesity. *Genetics in medicine : official journal of the American College of Medical Genetics* 2010; **12**(10): 641-647.
4. Ballif BC, Hornor SA, Jenkins E, Madan-Khetarpal S, Surti U, Jackson KE *et al.* Discovery of a previously unrecognized microdeletion syndrome of 16p11.2-p12.2. *Nat Genet* 2007; **39**(9): 1071-1073.
5. Bassuk AG, Geraghty E, Wu S, Mullen SA, Berkovic SF, Scheffer IE *et al.* Deletions of 16p11.2 and 19p13.2 in a family with intellectual disability and generalized epilepsy. *American journal of medical genetics Part A* 2013; **161A**(7): 1722-1725.
6. Bijlsma EK, Gijsbers AC, Schuurs-Hoeijmakers JH, van Haeringen A, Fransen van de Putte DE, Anderlid BM *et al.* Extending the phenotype of recurrent rearrangements of 16p11.2: deletions in mentally retarded patients without autism and in normal individuals. *European journal of medical genetics* 2009; **52**(2-3): 77-87.
7. Crepel A, Steyaert J, De la Marche W, De Wolf V, Fryns JP, Noens I *et al.* Narrowing the critical deletion region for autism spectrum disorders on 16p11.2. *Am J Med Genet B Neuropsychiatr Genet* 2011; **156**(2): 243-245.
8. D'Angelo D, Lebon S, Chen Q, Martin-Brevet S, Snyder LG, Hippolyte L *et al.* Defining the Effect of the 16p11.2 Duplication on Cognition, Behavior, and Medical Comorbidities. *JAMA Psychiatry* 2016; **73**(1): 20-30.
9. Girirajan S, Rosenfeld JA, Cooper GM, Antonacci F, Siswara P, Itsara A *et al.* A recurrent 16p12.1 microdeletion supports a two-hit model for severe developmental delay. *Nat Genet* 2010; **42**(3): 203-209.
10. Grayton HM, Fernandes C, Rujescu D, Collier DA. Copy number variations in neurodevelopmental disorders. *Progress in neurobiology* 2012; **99**(1): 81-91.
11. Guha S, Rees E, Darvasi A, Ivanov D, Ikeda M, Bergen SE *et al.* Implication of a rare deletion at distal 16p11.2 in schizophrenia. *JAMA psychiatry* 2013; **70**(3): 253-260.
12. Hanson E, Bernier R, Porche K, Jackson FI, Goin-Kochel RP, Snyder LG *et al.* The cognitive and behavioral phenotype of the 16p11.2 deletion in a clinically ascertained population. *Biological psychiatry* 2015; **77**(9): 785-793.

13. Hanson E, Nasir RH, Fong A, Lian A, Hundley R, Shen Y *et al.* Cognitive and behavioral characterization of 16p11.2 deletion syndrome. *Journal of developmental and behavioral pediatrics : JDBP* 2010; **31**(8): 649-657.
14. Hempel M, Rivera Bragues N, Wagenstaller J, Lederer G, Weitensteiner A, Seidel H *et al.* Microdeletion syndrome 16p11.2-p12.2: clinical and molecular characterization. *American journal of medical genetics Part A* 2009; **149A**(10): 2106-2112.
15. Horev G, Ellegood J, Lerch JP, Son YE, Muthuswamy L, Vogel H *et al.* Dosage-dependent phenotypes in models of 16p11.2 lesions found in autism. *Proceedings of the National Academy of Sciences of the United States of America* 2011; **108**(41): 17076-17081.
16. Jacquemont S, Reymond A, Zufferey F, Harewood L, Walters RG, Kutalik Z *et al.* Mirror extreme BMI phenotypes associated with gene dosage at the chromosome 16p11.2 locus. *Nature* 2011; **478**(7367): 97-102.
17. Maillard AM, Hippolyte L, Rodriguez-Herreros B, Chawner SJ, Dremmel D, Aguera Z *et al.* 16p11.2 Locus modulates response to satiety before the onset of obesity. *International journal of obesity* 2015.
18. Maillard AM, Ruef A, Pizzagalli F, Migliavacca E, Hippolyte L, Adaszewski S *et al.* The 16p11.2 locus modulates brain structures common to autism, schizophrenia and obesity. *Molecular psychiatry* 2015; **20**(1): 140-147.
19. Marshall CR, Noor A, Vincent JB, Lionel AC, Feuk L, Skaug J *et al.* Structural variation of chromosomes in autism spectrum disorder. *American journal of human genetics* 2008; **82**(2): 477-488.
20. McCarthy SE, Makarov V, Kirov G, Addington AM, McClellan J, Yoon S *et al.* Microduplications of 16p11.2 are associated with schizophrenia. *Nat Genet* 2009; **41**(11): 1223-1227.
21. Bi W, Ohyama T, Nakamura H, Yan J, Visvanathan J, Justice MJ *et al.* Inactivation of Rai1 in mice recapitulates phenotypes observed in chromosome engineered mouse models for Smith-Magenis syndrome. *Hum Mol Genet* 2005; **14**(8): 983-995.
22. Bi W, Saifi GM, Shaw CJ, Walz K, Fonseca P, Wilson M *et al.* Mutations of RAI1, a PHD-containing protein, in nondeletion patients with Smith-Magenis syndrome. *Human genetics* 2004; **115**(6): 515-524.
23. Boone PM, Reiter RJ, Glaze DG, Tan DX, Lupski JR, Potocki L. Abnormal circadian rhythm of melatonin in Smith-Magenis syndrome patients with RAI1 point mutations. *American journal of medical genetics Part A* 2011; **155A**(8): 2024-2027.
24. Carmona-Mora P, Encina CA, Canales CP, Cao L, Molina J, Kairath P *et al.* Functional and cellular characterization of human Retinoic Acid Induced 1 (RAI1) mutations associated with Smith-Magenis Syndrome. *BMC molecular biology* 2010; **11**: 63.

25. De Leersnyder H, De Blois MC, Claustrat B, Romana S, Albrecht U, Von Kleist-Retzow JC *et al.* Inversion of the circadian rhythm of melatonin in the Smith-Magenis syndrome. *The Journal of pediatrics* 2001; **139**(1): 111-116.
26. Dubourg C, Bonnet-Brilhault F, Toutain A, Mignot C, Jacquette A, Dieux A *et al.* Identification of Nine New RAI1-Truncating Mutations in Smith-Magenis Syndrome Patients without 17p11.2 Deletions. *Molecular syndromology* 2014; **5**(2): 57-64.
27. Dykens EM, Finucane BM, Gayley C. Brief report: cognitive and behavioral profiles in persons with Smith-Magenis syndrome. *Journal of autism and developmental disorders* 1997; **27**(2): 203-211.
28. Dykens EM, Smith AC. Distinctiveness and correlates of maladaptive behaviour in children and adolescents with Smith-Magenis syndrome. *Journal of intellectual disability research : JIDR* 1998; **42 ( Pt 6)**: 481-489.
29. Elsea SH, Girirajan S. Smith-Magenis syndrome. *European journal of human genetics : EJHG* 2008; **16**(4): 412-421.
30. Girirajan S, Elsas LJ, 2nd, Devriendt K, Elsea SH. RAI1 variations in Smith-Magenis syndrome patients without 17p11.2 deletions. *J Med Genet* 2005; **42**(11): 820-828.
31. Dixon JR, Selvaraj S, Yue F, Kim A, Li Y, Shen Y *et al.* Topological domains in mammalian genomes identified by analysis of chromatin interactions. *Nature* 2012; **485**(7398): 376-380.
32. Rao SS, Huntley MH, Durand NC, Stamenova EK, Bochkov ID, Robinson JT *et al.* A 3D map of the human genome at kilobase resolution reveals principles of chromatin looping. *Cell* 2014; **159**(7): 1665-1680.
33. Blumenthal I, Ragavendran A, Erdin S, Klei L, Sugathan A, Guide JR *et al.* Transcriptional consequences of 16p11.2 deletion and duplication in mouse cortex and multiplex autism families. *American journal of human genetics* 2014; **94**(6): 870-883.
34. Migliavacca E, Golzio C, Mannik K, Blumenthal I, Oh EC, Harewood L *et al.* A Potential Contributory Role for Ciliary Dysfunction in the 16p11.2 600 kb BP4-BP5 Pathology. *American journal of human genetics* 2015; **96**(5): 784-796.
35. Cai M, Kim S, Wang K, Farnham PJ, Coetzee GA, Lu W. 4C-seq revealed long-range interactions of a functional enhancer at the 8q24 prostate cancer risk locus. *Scientific reports* 2016; **6**: 22462.
36. Arbogast T, Ouagazzal AM, Chevalier C, Kopanitsa M, Afinowi N, Migliavacca E *et al.* Reciprocal Effects on Neurocognitive and Metabolic Phenotypes in Mouse Models of 16p11.2 Deletion and Duplication Syndromes. *PLoS genetics* 2016; **12**(2): e1005709.
37. Jager R, Migliorini G, Henrion M, Kandaswamy R, Speedy HE, Heindl A *et al.* Capture Hi-C identifies the chromatin interactome of colorectal cancer risk loci. *Nature communications* 2015; **6**: 6178.

38. Beunders G, Voorhoeve E, Golzio C, Pardo LM, Rosenfeld JA, Talkowski ME *et al.* Exonic deletions in *AUTS2* cause a syndromic form of intellectual disability and suggest a critical role for the C terminus. *American journal of human genetics* 2013; **92**(2): 210-220.
39. Golzio C, Willer J, Talkowski ME, Oh EC, Taniguchi Y, Jacquemont S *et al.* KCTD13 is a major driver of mirrored neuroanatomical phenotypes of the 16p11.2 copy number variant. *Nature* 2012; **485**(7398): 363-367.
40. Jordan DM, Frangakis SG, Golzio C, Cassa CA, Kurtzberg J, Task Force for Neonatal G *et al.* Identification of cis-suppression of human disease mutations by comparative genomics. *Nature* 2015; **524**(7564): 225-229.
41. O'Brien SA, Zhu M, Zhang W. The Importance of IL-6 in the Development of LAT-Mediated Autoimmunity. *Journal of immunology* 2015; **195**(2): 695-705.
42. Kortum RL, Rouquette-Jazdanian AK, Miyaji M, Merrill RK, Markegard E, Pinski JM *et al.* A phospholipase C-gamma1-independent, RasGRP1-ERK-dependent pathway drives lymphoproliferative disease in linker for activation of T cells-Y136F mutant mice. *Journal of immunology* 2013; **190**(1): 147-158.
43. Lin GN, Corominas R, Lemmens I, Yang X, Tavernier J, Hill DE *et al.* Spatiotemporal 16p11.2 protein network implicates cortical late mid-fetal brain development and KCTD13-Cul3-RhoA pathway in psychiatric diseases. *Neuron* 2015; **85**(4): 742-754.
44. Malissen B, Gregoire C, Malissen M, Roncagalli R. Integrative biology of T cell activation. *Nature immunology* 2014; **15**(9): 790-797.
45. Delaguillaumie A, Lagaudriere-Gesbert C, Popoff MR, Conjeaud H. Rho GTPases link cytoskeletal rearrangements and activation processes induced via the tetraspanin CD82 in T lymphocytes. *Journal of cell science* 2002; **115**(Pt 2): 433-443.
46. Roskoski R, Jr. ERK1/2 MAP kinases: structure, function, and regulation. *Pharmacological research* 2012; **66**(2): 105-143.
47. Liang P, Wan Y, Yan Y, Wang Y, Luo N, Deng Y *et al.* MVP interacts with YPEL4 and inhibits YPEL4-mediated activities of the ERK signal pathway. *Biochemistry and cell biology = Biochimie et biologie cellulaire* 2010; **88**(3): 445-450.
48. Yu Z, Fotouhi-Ardakani N, Wu L, Maoui M, Wang S, Banville D *et al.* PTEN associates with the vault particles in HeLa cells. *The Journal of biological chemistry* 2002; **277**(43): 40247-40252.
49. Kolli S, Zito CI, Mossink MH, Wiemer EA, Bennett AM. The major vault protein is a novel substrate for the tyrosine phosphatase SHP-2 and scaffold protein in epidermal growth factor signaling. *The Journal of biological chemistry* 2004; **279**(28): 29374-29385.

50. Shen S, Chuck MI, Zhu M, Fuller DM, Yang CW, Zhang W. The importance of LAT in the activation, homeostasis, and regulatory function of T cells. *The Journal of biological chemistry* 2010; **285**(46): 35393-35405.
51. Zhang W, Sommers CL, Burshtyn DN, Stebbins CC, DeJarnette JB, Tribble RP *et al.* Essential role of LAT in T cell development. *Immunity* 1999; **10**(3): 323-332.
52. Filiano AJ, Gadani SP, Kipnis J. Interactions of innate and adaptive immunity in brain development and function. *Brain research* 2015; **1617**: 18-27.
53. Angibaud J, Louveau A, Baudouin SJ, Nerriere-Daguin V, Evain S, Bonnamain V *et al.* The immune molecule CD3zeta and its downstream effectors ZAP-70/Syk mediate ephrin signaling in neurons to regulate early neurogenesis. *Journal of neurochemistry* 2011; **119**(4): 708-722.
54. Ashwood P, Krakowiak P, Hertz-Picciotto I, Hansen R, Pessah IN, Van de Water J. Altered T cell responses in children with autism. *Brain, behavior, and immunity* 2011; **25**(5): 840-849.
55. White J, Beck CR, Harel T, Posey JE, Jhangiani SN, Tang S *et al.* POGZ truncating alleles cause syndromic intellectual disability. *Genome medicine* 2016; **8**(1): 3.
56. Carter H, Hofree M, Ideker T. Genotype to phenotype via network analysis. *Curr Opin Genet Dev* 2013; **23**(6): 611-621.
57. del Sol A, Balling R, Hood L, Galas D. Diseases as network perturbations. *Current opinion in biotechnology* 2010; **21**(4): 566-571.
58. Lupianez DG, Kraft K, Heinrich V, Krawitz P, Brancati F, Klopocki E *et al.* Disruptions of topological chromatin domains cause pathogenic rewiring of gene-enhancer interactions. *Cell* 2015; **161**(5): 1012-1025.
59. Kochinke K, Zweier C, Nijhof B, Fenckova M, Cizek P, Honti F *et al.* Systematic Phenomics Analysis Deconvolutes Genes Mutated in Intellectual Disability into Biologically Coherent Modules. *Am J Hum Genet* 2016; **98**(1): 149-164.
60. Menche J, Sharma A, Kitsak M, Ghiassian SD, Vidal M, Loscalzo J *et al.* Disease networks. Uncovering disease-disease relationships through the incomplete interactome. *Science* 2015; **347**(6224): 1257601.
61. Costanzo M, Baryshnikova A, Bellay J, Kim Y, Spear ED, Sevier CS *et al.* The genetic landscape of a cell. *Science* 2010; **327**(5964): 425-431.
62. Goh KI, Choi IG. Exploring the human diseasome: the human disease network. *Brief Funct Genomics* 2012; **11**(6): 533-542.



63. Goh KI, Cusick ME, Valle D, Childs B, Vidal M, Barabasi AL. The human disease network. *Proceedings of the National Academy of Sciences of the United States of America* 2007; **104**(21): 8685-8690.
64. Zhong Q, Simonis N, Li QR, Charlotheaux B, Heuze F, Klitgord N *et al.* Edgetic perturbation models of human inherited disorders. *Molecular systems biology* 2009; **5**: 321.
65. Wang X, Wei X, Thijssen B, Das J, Lipkin SM, Yu H. Three-dimensional reconstruction of protein networks provides insight into human genetic disease. *Nat Biotechnol* 2012; **30**(2): 159-164.