

Université de Montréal

**Évaluation du caryotype moléculaire en tant qu'outil diagnostique
chez les enfants avec déficience intellectuelle et/ou malformations
congénitales**

par
Guylaine D'Amours

Programme de Sciences biomédicales
Faculté de Médecine

Mémoire présenté à la Faculté des études supérieures et postdoctorales
en vue de l'obtention du grade de
Maître ès science
en Sciences biomédicales

Mai 2013

© Guylaine D'Amours, 2013

Université de Montréal
Faculté des études supérieures et postdoctorales

Ce mémoire intitulé :

Évaluation du caryotype moléculaire en tant qu'outil diagnostique chez les enfants avec
déficience intellectuelle et/ou malformations congénitales

Présenté par :
Guylaine D'Amours

a été évalué par un jury composé des personnes suivantes :

Dre Josée Hébert, président-rapporteur
Dre Emmanuelle Lemyre, directrice de recherche
Dr Jacques L. Michaud, co-directeur
Dre Alessandra M. Duncan, membre du jury

RÉSUMÉ

Le caryotype moléculaire permet d'identifier un CNV chez 10-14% des individus atteints de déficience intellectuelle et/ou de malformations congénitales. C'est pourquoi il s'agit maintenant de l'analyse de première intention chez ces patients. Toutefois, le rendement diagnostique n'est pas aussi bien défini en contexte prénatal et l'identification de CNVs de signification clinique incertaine y est particulièrement problématique à cause du risque d'interruption de grossesse. Nous avons donc testé 49 fœtus avec malformations majeures et un caryotype conventionnel normal avec une micropuce CGH pangénomique, et obtenu un diagnostic dans 8,2% des cas. Par ailleurs, des micropuces à très haute résolution combinant le caryotype moléculaire et le génotypage de SNPs ont récemment été introduites sur le marché. En plus d'identifier les CNVs, ces plateformes détectent les LOHs, qui peuvent indiquer la présence d'une mutation homozygote ou de disomie uniparentale. Ces anomalies pouvant être associées à la déficience intellectuelle ou à des malformations, leur détection est particulièrement intéressante pour les patients dont le phénotype reste inexpliqué. Cependant, le rendement diagnostique de ces plateformes n'est pas confirmé, et l'utilité clinique réelle des LOHs n'est toujours pas établie. Nous avons donc testé 21 enfants atteints de déficience intellectuelle pour qui les méthodes standards d'analyse génétique n'avaient pas résulté en un diagnostic, et avons pu faire passer le rendement diagnostique de 14,3% à 28,6% grâce à l'information fournie par les LOHs. Cette étude démontre l'utilité clinique d'une micropuce CGH pangénomique chez des fœtus avec malformations, de même que celle d'une micropuce SNP chez des enfants avec déficience intellectuelle.

Mots-clés : Consanguinité, Déficience intellectuelle, Diagnostic prénatal, Disomie uniparentale, Hybridation génomique comparative (CGH), Malformations congénitales, Micropuce, Pertes d'hétérozygotie (LOH), Polymorphisme de nucléotide simple (SNP), Variation de nombre de copies de segment d'ADN (CNV).

ABSTRACT

Molecular karyotyping identifies a CNV in 10-14% of individuals affected with intellectual disability and/or congenital abnormalities. Therefore, it is now the first-tier analysis for these patients. However, the diagnostic yield is not as clear in the prenatal context, and the risk of pregnancy termination makes the detection of variants of uncertain clinical significance particularly problematic. We tested 49 fetuses with major malformations and a normal karyotype, using a pangenomic CGH array, and obtained a diagnosis in 8.2% of cases. Furthermore, high-resolution microarrays combining molecular karyotyping and SNP genotyping were recently introduced on the market. In addition to identifying CNVs, these platforms detect LOHs, which can indicate the presence of a homozygous mutation or of uniparental disomy. Since these abnormalities can be associated with intellectual disability or congenital abnormalities, their detection is of particular interest for patients whose phenotype remains unexplained. However, the diagnostic yield obtained with these platforms is not confirmed, and the real clinical value of LOH detection is not yet established. We tested 21 children affected with intellectual disability for whom standard genetic analyses failed to provide a diagnosis, and were able to increase the diagnostic yield from 14.3% to 28.6% as a result of the information provided by LOHs. This study shows the clinical usefulness of pangenomic CGH arrays in fetuses with malformation(s), as well as that of SNP arrays in children with intellectual disability.

Keywords : Comparative genomic hybridization (CGH), Congenital abnormalities, Consanguinity, DNA copy number variation (CNV), Intellectual disability, Loss of heterozygosity (LOH), Microarray analysis, Prenatal diagnosis, Single nucleotide polymorphism (SNP), Uniparental disomy.

TABLE DES MATIÈRES

RÉSUMÉ	iii
ABSTRACT	iv
LISTE DES SIGLES ET ABRÉVIATIONS	viii
LISTE DES TABLEAUX	x
LISTE DES FIGURES	xi
REMERCIEMENTS	xiii
1 INTRODUCTION	1
1.1 GÉNÉRALITÉS.....	1
1.2 VARIANTS DU NOMBRE DE COPIES.....	3
1.2.1 Définition	3
1.2.2 Mécanismes de formation	3
1.2.3 Mécanismes pathogéniques	7
1.2.4 Signification clinique	9
1.2.4.1 Éléments compliquant l'interprétation des résultats.....	12
1.2.4.2 Outils informatiques nécessaires	14
1.3 PERTES D'HÉTÉROZYGOTIE.....	17
1.3.1 Définition	17
1.3.2 Ce que révèle la présence de pertes d'hétérozygotie.....	17
1.3.2.1 Endogamie & consanguinité.....	18
1.3.2.2 Disomie uniparentale.....	20
1.3.3 Mécanismes pathogéniques associés aux pertes d'hétérozygotie.....	24
1.3.3.1 Transmission de mutations homozygotes.....	24
1.3.3.2 Gènes soumis à l'empreinte parentale.....	25
1.4 MICROPUCES EN CYTOGÉNÉTIQUE CLINIQUE.....	27
1.4.1 Hybridation génomique comparative (CGH)	27
1.4.1.1 Principe et évolution de la technique de CGH.....	27
1.4.1.2 Avantages et limites de la CGH.....	30
1.4.2 Génotypage SNP sur micropuce	32
1.4.2.1 Principe du génotypage SNP sur micropuce	32
1.4.2.2 Avantages et limites des micropuces SNP.....	33
1.4.3 Le caryotype moléculaire aujourd'hui.....	37
1.4.3.1 Combinaison CGH / génotypage SNP.....	37
1.4.3.2 Utilisation actuelle des micropuces en clinique.....	37
2 HYPOTHÈSES ET OBJECTIFS	39
2.1 HYPOTHÈSES.....	39
2.1.1 CGH pangénomique en prénatal	39
2.1.2 Combinaison CGH / génotypage SNP en postnatal	40
2.2 OBJECTIFS.....	41
2.2.1 CGH pangénomique en prénatal	41

2.2.2	<i>Combinaison CGH / géotypage SNP en postnatal</i>	41
3	WHOLE-GENOME ARRAY CGH IDENTIFIES PATHOGENIC COPY NUMBER VARIATIONS IN FETUSES WITH MAJOR MALFORMATIONS AND A NORMAL KARYOTYPE	42
3.1	ABSTRACT	43
3.2	INTRODUCTION	44
3.3	MATERIALS AND METHODS	46
3.3.1	<i>Patients and samples</i>	46
3.3.2	<i>Array CGH analysis and interpretation</i>	46
3.3.3	<i>Validation of CNVs</i>	48
3.4	RESULTS	50
3.4.1	<i>Clinically significant results</i>	53
3.4.2	<i>Results of unclear clinical significance</i>	55
3.4.3	<i>Novel benign CNVs</i>	57
3.5	DISCUSSION	58
3.5.1	<i>Diagnostic yield</i>	60
3.5.2	<i>Size of detected chromosomal aberrations</i>	61
3.5.3	<i>Interpretation challenges</i>	61
3.5.4	<i>Implementation in the prenatal setting</i>	62
3.6	ACKNOWLEDGEMENTS	64
3.7	REFERENCES	65
3.8	SUPPORTING INFORMATION	71
3.8.1	<i>Methods</i>	71
3.8.1.1	<i>Samples preparation</i>	74
3.8.1.2	<i>Targeted FISH</i>	74
3.8.1.3	<i>Arrays</i>	74
3.8.1.4	<i>Web database links</i>	75
3.8.1.5	<i>Statistical analysis</i>	75
3.8.2	<i>Results & discussion</i>	76
3.8.2.1	<i>Clinically significant results: additional clinical & interpretation details</i>	76
3.8.2.2	<i>Results of unclear clinical significance: additional clinical & interpretation details</i>	78
3.8.2.3	<i>Benign CNVs: clinical & interpretation details</i>	83
3.8.3	<i>References</i>	86
4	SNP ARRAYS: COMPARING DIAGNOSTIC YIELDS FOR FOUR PLATFORMS IN CHILDREN WITH DEVELOPMENTAL DELAY	94
4.1	ABSTRACT	95
4.2	INTRODUCTION	96
4.3	MATERIAL & METHODS	98
4.3.1	<i>Patients</i>	98
4.3.2	<i>Samples</i>	101
4.3.3	<i>Array genomic hybridization</i>	101
4.3.4	<i>Visualization software settings</i>	101
4.3.5	<i>CNV interpretation</i>	104
4.3.6	<i>LOH interpretation</i>	104

4.4	RESULTS	106
4.4.1	<i>Detected CNVs: standard thresholds</i>	106
4.4.2	<i>Detected CNVs: optimized thresholds</i>	106
4.4.3	<i>CNV validation</i>	107
4.4.4	<i>Clinical significance of detected de novo CNVs</i>	109
4.4.5	<i>Clinical significance of confirmed de novo CNVs</i>	116
4.4.6	<i>UPD</i>	117
4.4.7	<i>Informative LCSHs</i>	118
4.4.8	<i>Interpretation software</i>	121
4.5	DISCUSSION	122
4.5.1	<i>CNVs</i>	122
4.5.2	<i>LCSHs & UPD</i>	126
4.5.3	<i>Interpretation software</i>	128
4.6	CONCLUSION	130
4.7	ACKNOWLEDGEMENTS.....	131
4.8	REFERENCES	132
4.9	SUPPORTING INFORMATION.....	137
5	SYNTHÈSE ET CONCLUSION	158
5.1	CGH PANGÉNOMIQUE EN PRÉNATAL	159
5.1.1	<i>Rendement diagnostique</i>	159
5.1.2	<i>CNVs de signification clinique incertaine</i>	160
5.2	COMBINAISON CGH / GÉNOTYPAGE SNP EN POSTNATAL.....	161
5.2.1	<i>Rendement diagnostique</i>	161
5.2.2	<i>CNVs bénins & de signification clinique incertaine</i>	162
5.2.3	<i>Logiciels de visualisation</i>	163
5.2.4	<i>Défis & charge de travail associés aux micropuces SNP</i>	164
5.3	LIMITES ET AMÉLIORATIONS	166
5.4	PERSPECTIVES	170
5.4.1	<i>Implantation en clinique</i>	170
5.4.2	<i>Interprétation des résultats : les défis</i>	171
5.4.3	<i>Séquençage à très haut débit</i>	174
5.4.4	<i>Stratégie d'analyse : ciblée ou non?</i>	178
5.5	CONCLUSION	180
	BIBLIOGRAPHIE	182
	ANNEXE	xiv
	ANNEXE 1 : APPORT DES AUTEURS AUX ARTICLES.....	xiv

LISTE DES SIGLES ET ABRÉVIATIONS

2.7M	<i>Affymetrix Cytogenetics Whole-Genome 2.7M Array</i>
aCGH	<i>Array-based comparative genomic hybridization</i>
ADN	<i>Acide désoxyribonucléique</i>
AF	<i>Amniotic fluid</i>
ARN	<i>Acide ribonucléique</i>
ARNnc	<i>ARN non-codant</i>
AS	<i>Angelman syndrome</i>
BAC	<i>Bacterial artificial chromosome</i>
BAF	<i>B Allele Frequency</i>
BIR	<i>Break-induced replication</i>
CB	<i>Cord blood</i>
CGH	<i>Comparative genomic hybridization</i>
CGX-12	<i>NimbleGen CGX-12 Cytogenetics Array</i>
ChAS	<i>Chromosome Analysis Suite</i>
CN-LOH	<i>Copy neutral loss of heterozygosity</i>
CNV	<i>Copy number variant/variation</i>
CVS	<i>Chorionic villus sampling</i>
CytoSNP	<i>Illumina HumanCytoSNP-12 DNA Analysis BeadChip</i>
DD	<i>Developmental delay</i>
DECIPHER	<i>Database of Chromosomal Imbalance and Phenotype in Humans Using Ensembl Resources</i>
DGV	<i>Database of Genomic Variants</i>
DG	<i>DiGeorge</i>
DNA	<i>Deoxyribonucleic acid</i>
DUP	<i>Disomie uniparentale</i>
DWS	<i>Dandy-Walker syndrome</i>
FFT	<i>Frozen fetal tissue</i>
FISH	<i>Fluorescence in situ hybridization</i>
FoSTeS	<i>Fork stalling and template switching</i>
GCAD	<i>Genoglyphix Chromosome Aberration Database</i>
GTG	<i>G-banding with trypsin and giemsa</i>
GWAS	<i>Genome-wide association study</i>
ID	<i>Intellectual disability</i>
IUFD	<i>Intrauterine fetal death</i>

IUGR	<i>Intrauterine growth retardation</i>
kb	<i>Kilobase</i>
LCR	<i>Low-copy repeats</i>
LCSH	<i>Long contiguous stretches of homozygosity</i>
LOH	<i>Loss of heterozygosity</i>
MAF	<i>Minor allele frequency</i>
Mb	<i>Mégabase</i>
MCA	<i>Multiple congenital anomalies</i>
MIM	<i>Mendelian Inheritance in Man</i>
MLPA	<i>Multiplex ligation probe amplification</i>
MMBIR	<i>Microhomology-mediated break-induced replication</i>
NAHR	<i>Nonallelic homologous recombination</i>
NCBI	<i>National Center for Biotechnology Information</i>
NHEJ	<i>Nonhomologous end joining</i>
OMIM	<i>Online Mendelian Inheritance in Man</i>
Omni1	<i>Illumina HumanOmni1-Quad BeadChip</i>
PAC	<i>P1-derived artificial chromosome</i>
PCR	<i>Polymerase chain reaction</i>
PWS	<i>Prader-Willi syndrome</i>
QMPSF	<i>Quantitative multiplex PCR of short fragments</i>
ROH	<i>Runs/regions of homozygosity</i>
SHFM3	<i>Split-hand/foot malformation 3</i>
snoRNA	<i>Small nucleolar RNA</i>
SNP	<i>Single nucleotide polymorphism</i>
SNP 6.0	<i>Affymetrix Genome-Wide Human SNP Array 6.0</i>
SUA	<i>Single umbilical artery</i>
TOP	<i>Termination of pregnancy</i>
UCSC	<i>University of California Santa Cruz</i>
UPD	<i>Uniparental disomy</i>
US	<i>Ultrasound</i>
VCF	<i>Velocardiofacial</i>
VOUS	<i>Variant of uncertain clinical significance</i>

LISTE DES TABLEAUX

Tableau 1-1 : Corrélation entre le pourcentage de LCSH et le degré de parenté des parents. ____	19
Table 3-1 : Summary of the aCGH results obtained with each type of array. _____	47
Table 3-2 : Details for fetuses with an aCGH result of defined or unclear clinical significance. _____	51
Table 3-3 : Summary of previous and present prenatal studies investigating fetuses with an abnormal ultrasound. _____	59
Table 3-S1 : Cohort description. _____	72
Table 3-S2 : List of fetuses tested with FISH and probes used. _____	74
Table 4-1 : Phenotype information of patients included in the study. _____	99
Table 4-2 : ChAS and KaryoStudio thresholds used for CNV and LCSH visualization. _____	103
Table 4-3 : Number of confirmed CNVs and confirmation rate for each tested array. _____	108
Table 4-4 : Clinical significance of detected de novo CNVs. _____	110
Table 4-5 : List of de novo CNVs confirmed on at least two arrays. _____	113
Table 4-6 : Informative LCSHs larger than 5 Mb. _____	120
Table 4-S1 : Number of CNVs detected by each tested array using standard thresholds. _____	137
Table 4-S2 : Number of CNVs detected by each tested array using optimized thresholds. _____	138
Table 4-S3 : List of CNVs confirmed on at least two arrays. _____	139
Table 4-S4 : List of LCSHs larger than 5 Mb. _____	145
Table 4-S5 : Percentage of homozygosity calculated from copy-neutral LOHs. _____	148
Table 4-S6 : Software features and their impact on result interpretation. _____	149

LISTE DES FIGURES

Figure 1-1 : Principaux mécanismes de formation des CNVs. _____	6
Figure 1-2 : Idéogramme des chromosomes humains avec des loci génomiques responsables de déficience intellectuelle syndromique et non syndromique. _____	11
Figure 1-3 : Disomie uniparentale résultant d'erreurs méiotiques. _____	22
Figure 1-4 : Disomie uniparentale résultant d'erreurs mitotiques. _____	23
Figure 1-5 : Résumé de la carte génomique de la région chromosomique 15q11.2-q13. _____	26
Figure 1-6 : Principe de la technique de CGH. _____	28
Figure 1-7 : Principe du génotypage SNP sur micropuce. _____	34
Figure 1-8 : Patrons de fréquence de l'allèle B (BAF : <i>B Allele Frequency</i>) et log ratios détectés par les micropuces SNP en cas de polyploïdie et de mosaïcisme. _____	35
Figure 3-1 : Interpretation algorithm used to establish clinical significance of the CNVs detected by aCGH in our cohort. _____	49
Figure 3-2 : Array CGH result for fetus 20. _____	54
Figure 4-1 : Number of <i>de novo</i> CNVs detected by each array for all 21 patients, categorized according to clinical significance. _____	111
Figure 4-2 : Number of <i>de novo</i> CNVs detected by each array, as a function of minimum size. _	115
Figure 4-S1 : VOUS detected in patient 3.2, as visualized in each software. _____	151
Figure 4-S2 : Pathogenic CNVs detected in patient 14.5, as visualized in each software. _____	152
Figure 4-S3 : Pathogenic CNV detected in patient 14.5, as visualized in each software. _____	153
Figure 4-S4 : Pathogenic CNV detected in patient 44.15, as visualized in each software. _____	154
Figure 4-S5 : Informative LCSH detected in patient 1.1, as visualized in each software. _____	155
Figure 4-S6 : Informative LCSHs detected in patients 10.4 and 11.4, as visualized in each software. _____	156
Figure 4-S7 : Informative LCSH detected in patient 26.9, as visualized in each software. _____	157
Figure 5-1 : Principes de détection de remaniements chromosomiques par séquençage à très haut débit. _____	177

À Papa

REMERCIEMENTS

Je tiens à remercier mes directeurs de recherche, Dre Emmanuelle Lemyre et Dr Jacques L. Michaud, qui m'ont tous deux appuyée à chacune des étapes de ce projet. Vos commentaires et critiques ont enrichi mon apprentissage et je me sens privilégiée d'avoir pu travailler à vos côtés. Dre Lemyre, merci pour votre guidance, votre disponibilité, et surtout de m'avoir fait confiance pour relever ce défi. Dr Michaud, ce fut un réel plaisir de pouvoir profiter de votre expertise et de vos précieuses recommandations.

Aux techniciennes du laboratoire de cytogénétique, un immense merci pour votre appui et pour m'avoir permis « d'envahir » votre espace de travail sans jamais me faire sentir que j'étais de trop, au contraire! Vos encouragements et votre enthousiasme ont rendu mes journées — et soirées! — au laboratoire des plus stimulantes. Merci également à Géraldine, qui m'a aidée à résoudre de multiples pépins techniques et surtout, m'a souvent rassurée grâce à ses sages conseils.

Je ne peux passer sous silence l'aide appréciable apportée par les personnes suivantes : Frédérique Tihy, Raouf Fetni, Dre Sonia Nizard et Mark E. Samuels. Les discussions que nous avons eues et vos mots d'encouragement m'ont permis de mieux naviguer à travers les divers écueils propres aux études supérieures, et je vous en suis sincèrement reconnaissante.

Je tiens aussi à remercier les familles qui ont accepté de participer aux projets de recherche, de même que Christine Massicotte pour sa collaboration et son impressionnante efficacité au recrutement. Merci également aux organismes m'ayant apporté un soutien financier : le Conseil de recherche en sciences naturelles et en génie du Canada, le Fonds de la recherche en santé du Québec, la Fondation du CHU Sainte-Justine et la Fondation des Étoiles.

Finalement, je souhaite remercier ma famille. Votre appui inconditionnel n'a pas de prix, et il est la source à laquelle j'ai pu puiser l'assurance et la motivation dont j'avais besoin pour atteindre mes objectifs, malgré les difficultés et les obstacles.

1 INTRODUCTION

1.1 Généralités

La biologie moléculaire permet l'analyse du génome nucléotide par nucléotide, et ses méthodes sont utilisées pour étudier tant les polymorphismes nucléotidiques (SNP : *single nucleotide polymorphism*) que les mutations ponctuelles pathogéniques. À l'autre spectre du génome, l'analyse cytogénétique détecte les aberrations chromosomiques telles que les trisomies, les translocations, les délétions et les duplications. L'analyse étant effectuée en observant les chromosomes au microscope, les anomalies détectées sont donc d'une taille minimale de 5 à 10 mégabases (Mb), ce qui correspond à la résolution obtenue dans des conditions idéales. Du fait de leur grande taille, ces anomalies touchent une portion importante du génome, et ont ainsi presque toujours un impact pathologique. Quelques exceptions sont toutefois bien connues, telles que les chromosomes marqueurs, qui souvent ne touchent que l'hétérochromatine péricentromérique, pauvre en gènes, ou les satellites des chromosomes acrocentriques, qui touchent les gènes répétitifs transcrits en ARNs ribosomiaux.

L'analyse cytogénétique est utilisée dans un contexte clinique depuis les années 60. À l'époque, on se contentait de déterminer le nombre de chromosomes, ce qui permettait tout de même d'établir un diagnostic chez bien des patients, atteints par exemple d'une trisomie, ou du syndrome de Turner. Puis le marquage en bandes G a permis une analyse plus précise des patrons de bandes spécifiques à chaque chromosome, ce qui a mené à l'identification d'anomalies touchant seulement une partie d'un chromosome, comme les translocations, les délétions, et les duplications. L'amélioration des techniques de marquage, de même que la mise au point de techniques moléculaires telles que l'hybridation *in situ* en fluorescence (FISH : *fluorescence in situ hybridization*), ont mené à l'identification d'anomalies chromosomiques de plus en plus petites, tant dans un contexte prénatal, qu'en oncologie et en cytogénétique constitutionnelle. L'utilisation des diverses analyses cytogénétiques dans la population prénatale et chez les patients atteints de déficience intellectuelle est détaillée dans l'introduction de chacun des articles de ce mémoire (sections 3.2 et 4.2).

Les avancées technologiques des dernières années, principalement l'hybridation génomique comparative sur micropuce (aCGH : *array-based comparative genomic hybridization*) ont permis de faire l'analyse cytogénétique du génome en entier en une seule expérience, et ce à une résolution qui ne cesse d'augmenter. Ces prouesses techniques ont toutefois été accompagnées de découvertes qui sont venues brouiller la compréhension que l'on croyait avoir du génome humain. Il s'avère que l'architecture de ce dernier est beaucoup plus complexe que ce que l'on pouvait imaginer il y a quinze ans. Bien que nous supposions la présence d'anomalies chromosomiques de taille intermédiaire — donc indétectables au microscope — l'expérience de la cytogénétique pouvait laisser penser qu'une aberration chromosomique impliquant des gènes uniques résultait pratiquement toujours en un phénotype anormal. Or, les séquences répétées s'avèrent beaucoup plus nombreuses que ce que l'on imaginait, et ne touchent pas seulement l'ADN dit « poubelle », mais également des régions plus ou moins riches en gènes uniques. C'est principalement cette notion qui nous rend prudents face à l'implantation dans les laboratoires cliniques de plateformes d'aCGH ayant une résolution toujours plus élevée. S'ajoutent à cette limitation, la performance technique des plateformes, qui semble inégale d'un fabricant à l'autre, de même que les « extras » disponibles sur certaines micropuces, qui théoriquement peuvent présenter un certain avantage.

1.2 Variants du nombre de copies

1.2.1 Définition

Plusieurs microremaniements chromosomiques, c'est-à-dire des anomalies chromosomiques non détectables au microscope, ont été identifiés et étudiés bien avant l'avènement de la CGH. Les syndromes de Miller-Dieker (délétion 17p13.3) ou de DiGeorge (délétion 22q11.2), par exemple, sont bien caractérisés et peuvent être diagnostiqués par FISH depuis de nombreuses années. Toutefois, c'est l'analyse pangénomique par CGH qui a mené à l'utilisation du terme « variant du nombre de copies » (CNV : *copy number variant* ou *copy number variation*) pour parler des variations de structure (délétions ou duplications) indétectables au microscope. Initialement, un CNV se définissait comme étant un segment génomique d'au moins 1 kilobase (kb) dont le nombre de copies dévie de celui d'un génome de référence ¹. Ce seuil avait été établi arbitrairement, et grâce à l'amélioration des techniques de détection des CNVs, on parle maintenant de CNV dès que le nombre de copies d'une région varie par rapport à la « normale », peu importe la taille du segment en cause ².

1.2.2 Mécanismes de formation

On distingue deux types de CNVs : les CNVs récurrents, qui sont retrouvés chez plusieurs individus et ont des points de cassure similaires, et les CNVs non-récurrents, qui, bien qu'ayant des régions génomiques pouvant se chevaucher, ont des points de cassures différents d'une personne à l'autre. Plusieurs mécanismes sont à l'origine de la formation des CNVs, les trois principaux (Figure 1-1) étant : la recombinaison homologue non allélique (NAHR : *nonallelic homologous recombination*), la réunion d'extrémités non homologues (NHEJ : *nonhomologous end joining*) et le changement de matrice suivant un blocage de la fourche de réplication (FoSTeS : *fork stalling and template switching*).

La NAHR (Figure 1-1 A) dépend de la présence de séquences répétées : les duplications segmentaires (de l'anglais *segmental duplications*, aussi appelées LCRs : *low-copy repeats*). Les duplications segmentaires sont des séquences non-identiques ayant une grande homologie (> 90%) entre elles, qui sont présentes en plusieurs copies, et réparties à

travers le génome ³. La présence de duplications segmentaires rapprochées peut mener au désalignement des chromosomes ou des chromatides durant la mitose ou la méiose, causant une recombinaison inégale qui résulte en une délétion, une duplication, ou une inversion (si les séquences répétées ont une orientation opposée). Puisque les duplications segmentaires sont répandues dans la population, c'est ce mécanisme qui serait responsable de la majorité des microremaniements récurrents, dont les points de cassures se trouvent dans ces séquences répétées ⁴. Par exemple, les délétions associées au syndrome de Di George résultent de ce mécanisme de formation. La plus fréquente est une délétion de 3 Mb, mais une délétion imbriquée de 1.5 Mb, dont les points de cassure correspondent à la position des LCRs dans cette région, est également récurrente, bien que moins fréquente ⁵.

La NHEJ (Figure 1-1 B) est avant tout un mécanisme permettant la réparation des cassures d'ADN double brin, qu'elles soient physiologiques (la recombinaison V(D)J et la commutation de classe des immunoglobulines) ou pathologiques (cassures causées par les radiations ionisantes, les radicaux libres ou autres) ⁶. Elle s'effectue en quatre étapes : 1) la détection des cassures double brin, 2) l'établissement d'un « pont » moléculaire entre les deux extrémités à réparer, 3) la modification des extrémités pour les rendre « collantes », 4) la ligation des extrémités ⁷. C'est la troisième étape qui laisse une « cicatrice » de réparation, soit une délétion ou une insertion de nucléotides. Ce type de réparation mène le plus souvent à des CNVs non-récurrents, puisqu'elle intervient pour réparer des cassures double brin survenant dans n'importe quelle région du génome, et ne requiert pas d'homologie de séquence pour avoir lieu ⁸. Il semblerait toutefois que l'architecture du génome puisse stimuler ce type de réparation, puisque les points de cassure des CNVs ainsi formés sont souvent à proximité ou dans des séquences répétées. La raison de cet état de fait reste toutefois à démontrer.

Le FoSTeS (Figure 1-1 C), qui implique une erreur de réplication, est également responsable de CNVs non-récurrents mais dont la structure est plus complexe que les CNVs issus de la NHEJ. Ce mécanisme serait une variante de la réplication induite par une cassure (BIR : *break-induced replication*), et est aussi appelée la BIR médiée par microhomologie (MMBIR : *microhomology-mediated break-induced replication*) ⁹. La BIR se produit lorsque, en cours de synthèse, la fourche de réplication rencontre une coupure (un *nick* : un nucléotide manquant) dans le brin matrice, ce qui entraîne « l'effondrement » de la fourche de

réplication. Afin de reprendre la synthèse, l'extrémité 5' du double brin contenant le *nick* est alors réséquée, ce qui donne un bout 3' saillant pouvant envahir le double brin homologue afin de l'utiliser comme brin matrice⁹. Le FoSTeS/MMBIR a lieu quand l'extrémité 3' envahit le double brin à une position qui présente une microhomologie (au lieu de la position correspondant à la séquence homologue). Ceci mène à une délétion ou une duplication, dépendamment de la position de la nouvelle fourche de réplication par rapport à la fourche initiale (en aval ou en amont).

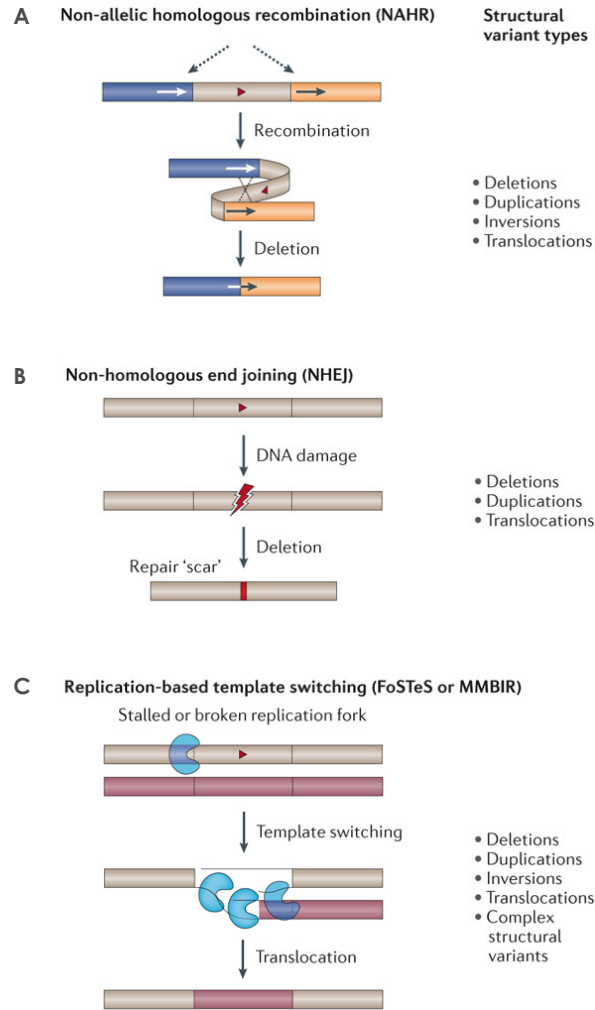


Figure 1-1 : Principaux mécanismes de formation des CNVs.

A | NAHR. Des duplications segmentaires (flèches dans les segments bleu et orange) présentes sur un même chromosome, ou sur des chromosomes différents (non illustré) peuvent se recombiner, résultant par exemple en une délétion du segment entre les duplications segmentaires (segment brun), ou encore en une duplication, une inversion ou une translocation (non illustré). B | NHEJ. La réparation de cassures d'ADN double-brin (éclair rouge) sur une chromatide (segment brun) est accompagnée de la délétion ou de l'addition de plusieurs nucléotides (la cicatrice de réparation : *repair scar*, bande rouge), ce qui peut mener à une délétion (du segment brun avec un triangle rouge) ou une duplication (non illustré). C | FoSTeS/MMBIR. Lors de la réplication de l'ADN, la fourche de réplication du double brin d'ADN (segment brun) peut se bloquer et la machinerie de réplication (en bleu) se transférer sur une séquence de grande homologie à proximité (segment rose), qu'elle soit sur le même chromosome ou non, et utiliser celle-ci comme brin matrice, ce qui mène à une translocation ou insertion (illustré) ou encore à une délétion, une duplication, une inversion ou des variants de structure plus complexes (non illustré). *Adaptée de Weischenfeldt 2013*¹⁰.

1.2.3 Mécanismes pathogéniques

Pour qu'un CNV cause un phénotype anormal, il doit nécessairement altérer l'expression normale d'un ou plusieurs gène(s). Cette altération peut se faire de plusieurs façons^{1,11}, que l'on pourrait classer en quatre grandes catégories : 1) dosage du gène, 2) altération de la structure génique, 3) effet de position et altération des éléments régulateurs, 4) dévoilement d'une mutation hétérozygote.

Certains gènes nécessitent simplement d'être exprimés, peu importe le niveau, pour produire un phénotype normal. Ce n'est pas le cas des gènes touchés par le premier mécanisme pathogénique. On dit de ces gènes qu'ils sont sensibles au dosage, c'est-à-dire qu'une déviation dans leur niveau d'expression entraîne un phénotype anormal. Un gène affecté par une délétion est dit haploinsuffisant lorsque l'expression de la seconde copie du gène n'est pas suffisante pour avoir le phénotype normal. Un gène affecté par une duplication, quant à lui, est considéré comme étant soumis à l'effet de dose, ce qui veut dire que c'est la surexpression de celui-ci qui est responsable du phénotype anormal. Pour les gènes sensibles au dosage, le CNV affecte donc directement le nombre de copies du gène, et c'est ce qui cause le phénotype anormal.

Le second mécanisme implique la structure elle-même du gène. En effet, une délétion ou une duplication — qu'elle soit en tandem, dispersée, ou encore inversée — peut interrompre la séquence du gène et abolir l'expression de la protéine correspondante ou mener à la production d'une protéine tronquée ou anormale, et donc à une perte de fonction. Il en résulte un phénotype anormal si ce gène est haploinsuffisant. Un phénotype anormal peut également être causé par la réunion de deux gènes incomplets — interrompus par le point de cassure d'une délétion ou d'une duplication — résultant en la production d'une protéine de fusion anormale, et donc à un gain de fonction.

L'altération directe d'un gène n'est pas nécessaire pour que son expression soit affectée. En effet, celle-ci est sous le contrôle d'éléments régulateurs : promoteurs, amplificateurs (*enhancers*), silenceurs, insulateurs et ARNs non-codants (ARNnc). Un promoteur est une séquence d'ADN située directement en amont du gène qu'elle régule, et est nécessaire pour amorcer la transcription du gène. Il agit un peu à la manière d'un interrupteur

qui est allumé ou éteint en fonction du temps (étape du développement) ou de l'espace (le tissu). Il comprend la boîte TATA, la boîte CAT et un segment riche en GC. Les amplificateurs et silenceurs sont des séquences consensus d'ADN situées en aval ou en amont du ou des gène(s) qu'ils régulent. C'est leur liaison par des facteurs de transcription (activateurs et répresseurs), ou encore leur transcription en ARN non-codants (ARNnc) qui influence le taux de transcription du gène. Contrairement aux promoteurs, ces séquences peuvent être très loin du gène régulé et peuvent donc agir à distance. Les insulateurs sont des séquences isolant un gène des promoteurs ou autres séquences régulatrices environnantes. Ils représentent donc une sorte de « frontière » qui délimite le territoire d'influence des autres éléments régulateurs. Les ARNnc comprennent les petits ARNnc (microARNs, petits ARNs interférents endogènes, et ARNs liant piwi)¹² et les longs ARNnc intergéniques (dont le plus connu est sans doute *XIST*, qui contrôle l'inactivation du X)¹³. Ces ARNnc agissent souvent sur les gènes à proximité soit en se liant directement à l'ADN ou encore aux ARNs messager, et exercent une influence directe sur l'expression génique en bloquant ou en activant la transcription ou la traduction. Une modification de la position d'un gène dans le génome, ou de celle d'un élément régulateur peut donc avoir autant d'effet qu'une délétion ou une duplication du gène lui-même. On parle alors d'effet de position. Par exemple, une duplication n'est pas nécessairement en tandem, mais peut être dispersée n'importe où dans le génome. Si ce déplacement a pour effet de placer le gène devant un promoteur toujours actif, alors que son expression est habituellement régulée de façon stricte, la surexpression résultante pourra causer un phénotype anormal. De même, le déplacement d'un élément régulateur à proximité d'un gène (suite à une délétion ou une duplication), ou la délétion d'un insulateur, pourront tous deux avoir un effet important sur l'expression de celui-ci. Enfin, le déplacement, la duplication ou la délétion de gènes codant pour des ARNnc peut avoir pour effet de modifier le taux de régulation des gènes sous leur contrôle.

Le dernier mécanisme pathogénique représente en réalité une deuxième mutation qui en démasque une première. En effet, une délétion, transmise par un parent, peut révéler une mutation ponctuelle présente sur le premier allèle (hérité de l'autre parent). Ce mécanisme pourra mener à l'expression d'un phénotype récessif.

Les mécanismes évoqués précédemment constituent une simplification du potentiel pathogénique des CNVs. Des mécanismes plus complexes sont impliqués dans nombre de pathologies, comme les désordres génomiques causés par de larges CNVs affectant plusieurs gènes contigus, ou les désordres résultant de l'effet cumulatif de multiples CNVs. D'autres pathologies impliquant des gènes modificateurs peuvent être causés par une mutation ou un CNV touchant des gènes situés à des loci distants. Il est probable que plusieurs affections à expression variable ou à pénétrance incomplète, telles que les maladies neuropsychiatriques, entrent dans cette dernière catégorie.

1.2.4 Signification clinique

La littérature scientifique regorge d'exemples de CNVs pathogéniques. Avant même la mise au point des méthodes pangénomiques, la technique de FISH avait identifié un très grand nombre de microremaniements responsables de différentes pathologies, principalement des syndromes cliniquement reconnaissables combinant soit une déficience intellectuelle ou un retard de développement, des dysmorphies et/ou des malformations congénitales ^{14,15}. L'utilisation des micropuces (ou caryotype moléculaire) pour détecter les CNVs a mené à une explosion du nombre de corrélations génotype-phénotype rapportées chez ces patients, tant dans les journaux scientifiques que dans les bases de données recensant les CNVs pathogéniques, si bien que la carte des régions chromosomiques touchées de façon récurrente s'est passablement enrichie (Figure 1-2). C'est ce haut rendement diagnostique qui a répandu l'utilisation du caryotype moléculaire chez les patients atteints de déficience intellectuelle et/ou de malformations congénitales et de plus en plus en contexte prénatal également (voir sections 3.2 et 4.2). La déficience intellectuelle et les malformations congénitales ayant des causes très variées, il peut être difficile d'orienter le diagnostic à partir des informations cliniques. En ce sens, l'apport du caryotype moléculaire est indéniable puisqu'il permet d'interroger le génome en entier en une seule analyse, ce qui est particulièrement utile chez les patients pour qui le processus diagnostique est dans un cul-de-sac alors qu'une origine génétique est suspectée. Les CNVs ainsi détectés peuvent ensuite être comparés avec les connaissances scientifiques les plus à jour pour la région génomique touchée afin d'en établir

la signification clinique chez le patient investigué. Tandis que le mécanisme pathogénique de certains CNVs est relativement bien connu et décrit¹⁶⁻¹⁹, d'autres ont un mécanisme toujours inconnu, mais n'en sont pas moins clairement associés à un phénotype anormal²⁰⁻²³. C'est pourquoi en plus d'examiner le contexte génomique d'un CNV pour prédire son effet sur l'expression génique, il est également essentiel de consulter la littérature scientifique et les bases de données pour établir les corrélations génotype-phénotype chez les patients testés. Le processus d'interprétation est détaillé dans les sections 3.3.2 et 4.3.5, et certaines précisions sont apportées dans la section 1.2.4.1. Les outils informatiques nécessaires à l'interprétation sont détaillés dans la section 1.2.4.2.

En plus de l'approche utilisée en génétique clinique, consistant à effectuer l'analyse chez un patient présentant un phénotype anormal afin de poser un diagnostic, des équipes de recherche adoptent une approche par association statistique pour identifier des CNVs présents dans de grandes cohortes de patients atteints d'une même pathologie à une plus grande fréquence que chez des individus sains. Alors que l'approche clinique s'est avérée particulièrement efficace pour les affections de type mendélien, ces études pangénomiques d'association (GWAS : *genome-wide association studies*) ont, en plus d'avoir identifié plusieurs CNVs pathogéniques, permis de mettre en lumière le rôle des CNVs dans les maladies complexes touchant les systèmes immunitaire et cardiovasculaire²⁴, et les désordres neuropsychiatriques^{25,26}. Ces résultats ont permis de renforcer l'idée que le modèle mendélien n'explique pas tout, et de commencer à mieux saisir l'origine de la variabilité d'expression et de la pénétrance incomplète observées dans certaines pathologies génétiques.

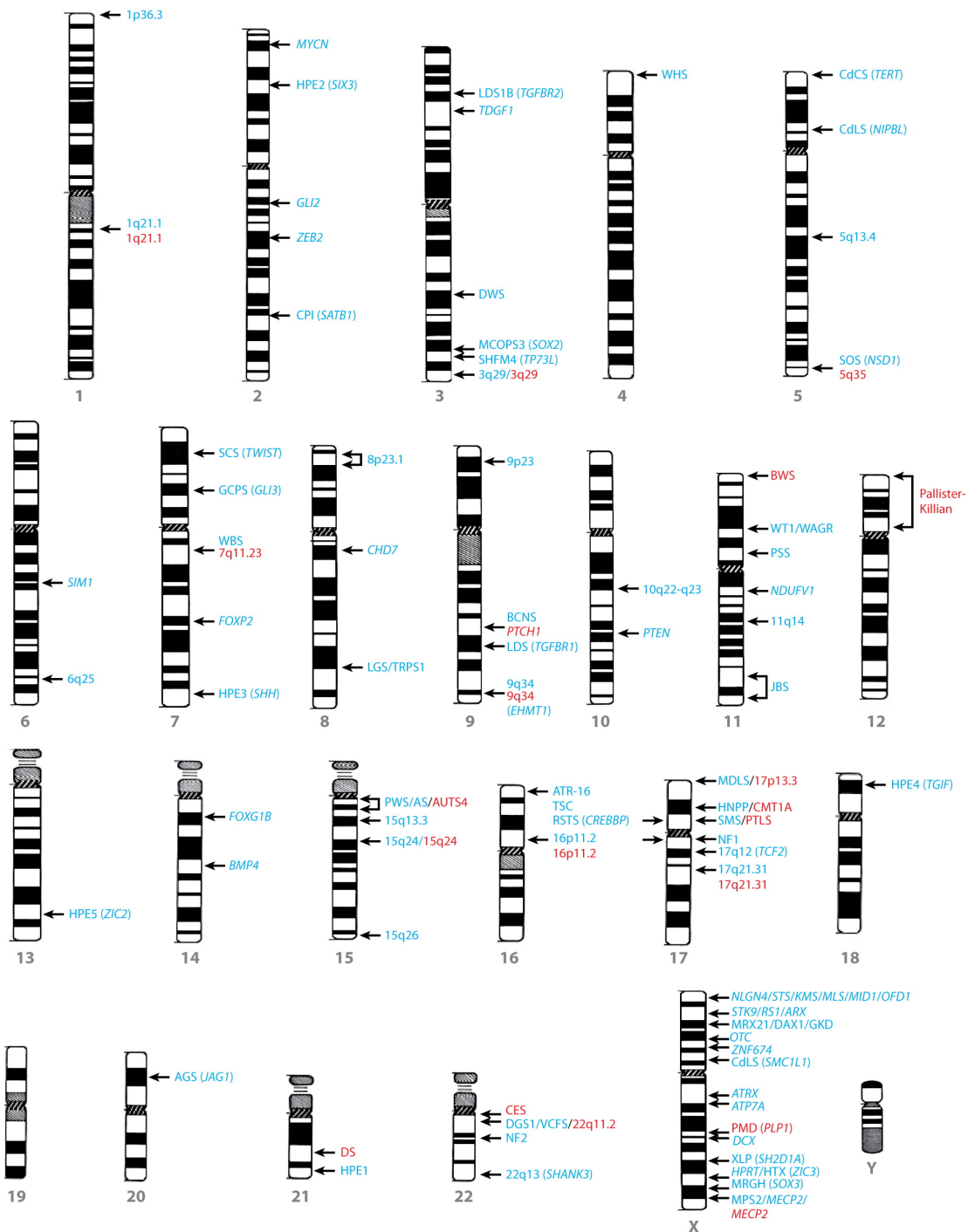


Figure 1-2 : Idéogramme des chromosomes humains avec des loci génomiques responsables de déficience intellectuelle syndromique et non syndromique.

Les duplications sont indiquées en rouge, les délétions en bleu. Tirée de Stankiewicz 2010²⁷.

1.2.4.1 Éléments compliquant l'interprétation des résultats

Les difficultés d'interprétation ne sont pas apparues avec le caryotype moléculaire, mais sont bien présentes depuis les débuts de la cytogénétique clinique. Les critères généralement associés à une anomalie chromosomique ayant un impact phénotypique incluent : anomalie touchant l'euchromatine, anomalie *de novo* ou hérité d'un parent atteint, anomalie de grande taille, et enfin — probablement le plus important — phénotype concordant observé chez d'autres individus porteurs de l'anomalie en question (dans la famille élargie ou dans la littérature scientifique). Si ces critères facilitent grandement le travail d'interprétation, ils ne sont toutefois pas absolus et plusieurs exceptions où le phénotype observé (normal) ne correspond pas à celui attendu (anormal ou atteint) ont été rapportées²⁸. Bien que certains de ces variants euchromatiques soient récurrents et aient été connus depuis longtemps comme n'ayant aucun impact phénotypique, il fut surprenant de constater qu'environ 40% des anomalies chromosomiques déséquilibrées non-récurrentes héritées n'avaient également aucun impact phénotypique. Il est vraisemblable que ces « exceptions » se multiplieront — et deviendront plutôt la norme — à mesure que la résolution du caryotype augmente et que la taille des anomalies détectées diminue.

Des directives cliniques ont été publiées par divers groupes²⁹⁻³¹, mais celles-ci restent imparfaites, et établir la signification clinique d'un CNV détecté chez un patient, même en suivant un algorithme précis, demeure un défi dans certains cas. En effet, certaines règles, qui étaient pratiquement absolues aux débuts du caryotype moléculaire, ont été délaissées — parce que démenties par les connaissances actuelles — au profit de nouvelles balises servant à faciliter le jugement du cytogénéticien.

Ainsi, contrairement aux anomalies chromosomiques visibles au microscope, le statut *de novo* d'un CNV n'est plus suffisant pour considérer avec une presque certitude que celui-ci est pathogénique, de même qu'un CNV hérité d'un parent sain n'est pas automatiquement éliminé comme étant la cause du phénotype anormal chez l'enfant³². Également, on sait depuis près de dix ans que tous les CNVs ne sont pas pathogéniques, mais que la plupart sont plutôt des polymorphismes³³⁻³⁵. Alors que certains d'entre eux ont été identifiés par plusieurs études différentes et sont relativement communs, avec une fréquence de l'allèle mineur $\geq 5\%$ (MAF : *minor allele frequency*), d'autres sont retrouvés à une fréquence plus basse (MAF

0.05-5%), et plusieurs — principalement les CNVs rares ($MAF \leq 0.05\%$) et les CNVs < 1 kb — n'auraient toujours pas été identifiés³⁶. Même s'il est vrai que les mutations ponctuelles (incluant les SNPs) ont longtemps été considérés comme étant le type de variation génétique le plus répandu du génome humain, il s'avère que les variations de structures (≥ 1 kb) affectent une plus grande portion du génome humain (jusqu'à 12%^{2,37}, vs 0,1% pour les SNPs³⁶). Leur fréquence moyenne dans la population est difficile à établir avec précision pour le moment, puisqu'elle varie en fonction de la taille, la technique d'analyse, la résolution et l'algorithme de détection. Mais à titre d'exemple, on sait que 5-10% de la population aurait au moins un CNV > 500 kb, et 65 à 80% en aurait au moins un > 100 kb³⁸. Également, on s'est récemment aperçu que les CNVs bénins n'étaient pas tous en-deçà de 1 Mb³⁸, et que la taille d'un CNV ne constituait donc pas un critère de pathogénicité absolu comme on le croyait pour les grands CNVs. Même les régions génomiques connues comme ayant un impact clinique représentent un défi du fait que certaines des pathologies y étant associées ne s'expriment pas de la même façon chez tous les patients, ou épargnent certains individus porteurs du CNV soi-disant pathogénique. On connaît bien entendu plusieurs désordres dont l'expressivité est variable d'un patient à l'autre, ou qui ont une pénétrance incomplète. Il est assez aisé dans ces cas d'en tenir compte dans l'interprétation des résultats. Le défi réside dans l'hypothèse, plus que probable, que plusieurs de ces syndromes à expressivité variable et pénétrance réduite restent à être identifiés parmi les CNVs rapportés plus récemment.

En effet, les cohortes de patients rapportées jusqu'à présent dans la littérature sont pour la plupart affectées d'un biais important : les patients sont généralement recrutés parce qu'ils ont un phénotype anormal. De plus, les CNVs détectés de façon fortuite chez des individus sains sont rarement publiés. L'information phénotypique disponible pour une anomalie donnée reflète donc souvent ce qui est observé chez les patients atteints plus sévèrement, excluant par le fait même — faussement — toute association avec un phénotype normal ou plus léger. Ce n'est pas sans rappeler ce qui s'est produit lorsque les premiers patients atteints d'anomalies des chromosomes sexuels ont été identifiés³⁹. Il importe de tenir compte de ce biais au moment d'interpréter les résultats du caryotype moléculaire et d'effectuer les corrélations phénotype-CNV avec précaution. Malgré tout, des progrès importants sont constamment faits, grâce entre autres au développement des bases de données, qui permettent l'accès aux résultats

obtenus dans diverses études, tant chez des patients que chez des individus sains, et facilitent le partage d'information entre les laboratoires.

1.2.4.2 Outils informatiques nécessaires

Actuellement, l'interprétation des résultats consiste à visualiser le CNV détecté, en inspecter le contenu génique, et vérifier s'il est situé dans une région génomique ayant un impact clinique connu. Étant donné le nombre de CNVs détectés chez chaque patient, cette étape pourrait difficilement se faire sur papier, et l'utilisation d'un logiciel de visualisation est nécessaire. Également, le procédé d'analyse du caryotype moléculaire s'effectuant sur support informatique, les données sont déjà sous cette forme et tous les manufacturiers fournissent un logiciel de visualisation des résultats. Ces logiciels ont différentes caractéristiques permettant de voir le signal pour chaque patient, et il n'est pas nécessaire de s'en tenir au logiciel fourni, même s'il est plus simple de le faire, puisque quelques compagnies produisent des logiciels permettant la visualisation des résultats de plusieurs plateformes.

S'il était auparavant relativement aisé de comparer la taille et la région chromosomique des anomalies détectées au microscope avec les anomalies rapportées antérieurement (assez bien documentées), la hausse de résolution des techniques cytogénétiques actuelles a complexifié le processus, puisqu'elle multiplie à l'infini le nombre de microremaniements pouvant être détectés et les régions génomiques impliquées. Se référer à des listes statiques de remaniements connus, tels que le *Catalogue of unbalanced chromosome aberrations in man*⁴⁰, n'est donc plus suffisant ni efficace, puisque les données changent à un rythme beaucoup trop rapide. C'est pourquoi il est essentiel d'avoir accès à différentes bases de données répertoriant, en temps réel, les CNVs détectés chez des patients et des individus sains.

Les bases de données les plus utilisées pour rechercher les altérations identifiées antérieurement chez des patients sont DECIPHER (*Database of Chromosomal Imbalance and Phenotype in Humans Using Ensembl Resources* : decipher.sanger.ac.uk), ISCA (*International Standards for Cytogenomic Arrays Consortium* : www.iscaconsortium.org), OMIM (*Online Mendelian Inheritance in Man* : omim.org) et PubMed (www.ncbi.nlm.nih.gov/pubmed). Les deux premières répertorient les données phénotypiques individuelles et la liste des CNVs

détectés chez des patients testés par divers laboratoires, qui y déposent leurs résultats après approbation des patients. Une recherche par phénotype ou par région génomique peut être effectuée, et les résultats peuvent être visualisés dans un navigateur génomique (*genome browser*). OMIM répertorie plutôt les gènes associés à un syndrome ou un désordre génétique dans une « *Morbid Map* » et fournit les détails cliniques pour chacun, en plus de liens vers les mutations connues et la littérature scientifique pertinente. Une troisième base de données peut également être utile, particulièrement quand un CNV est détecté dans une région pour la première fois, et que peu de données cliniques sont disponibles. PubMed recense les citations de la littérature biomédicale provenant de MEDLINE et permet de rechercher des articles scientifiques en fonction des gènes, régions chromosomiques, ou phénotypes pertinents. RefSeqGene (www.ncbi.nlm.nih.gov/refseq/rsg) répertorie tous les gènes séquencés, peu importe les connaissances scientifiques accumulées. Cette base de données devient particulièrement utile lorsqu'il n'y a que la recherche dans les autres bases n'a donné aucun résultat, et peut donc servir de point de départ pour une recherche de la littérature scientifique sur les gènes inclus dans l'altération détectée.

L'autre type d'information qui est essentiel à l'interprétation des résultats est un répertoire de CNVs bénins. La plus populaire est DGV (*Database of Genomic Variants*, projects.tcag.ca/variation), qui inventorie les CNVs identifiés dans la plupart des études effectuées dans des cohortes d'individus sains. Certaines limitations de DGV sont discutées dans la section 4.5.1. Également, des logiciels de visualisation commerciaux, comme Genoglyphix® (Signature Genomics), peuvent inclure un accès à leurs bases de données privées (CNVs bénins et pathogéniques), ce qui peut être une autre source importante d'information facilitant l'interprétation.

Une dernière base de données très importante est celle mise en place dans chaque laboratoire pour répertorier les anomalies identifiées fréquemment dans la population testée par le laboratoire, que ce soit les patients ou les individus sains. En effet, les différences génétiques entre les populations sont importantes³⁶, et un CNV fréquemment détecté chez des individus sains, originaires du Saguenay par exemple, peut ne jamais avoir été détecté ailleurs et donc ne pas être répertorié dans les bases de données externes. Le même principe s'applique aux CNVs pathogéniques hérités qui montrent une pénétrance réduite ou dont l'expressivité

est variable. De plus, chaque plateforme utilisant un ADN de référence qui lui est propre, certains CNVs peuvent être le résultat non pas d'une altération du nombre de copies chez les patients testés, mais plutôt de l'ADN de référence utilisé, ce qui sera rapidement détecté si une base de données répertoriant tous les résultats obtenus par le laboratoire est gardée à jour.

Les bases de données mentionnées précédemment sont accessibles en ligne, et peuvent être consultées à partir d'un lien dans le logiciel de visualisation, ou mieux encore, y être directement visualisées. La plupart permettent également le téléchargement des données qu'elles contiennent afin de pouvoir les importer dans le logiciel voulu. Également, un agrégateur de données, disponible gratuitement en ligne, est grandement utilisé : *UCSC Genome Browser* (genome.ucsc.edu/cgi-bin/hgGateway). Ce site web permet de consulter, de différentes façons (navigateur web, tableaux, listes), les données de la plupart des bases de données importantes, incluant OMIM, DGV, DECIPHER et RefSeq. L'outil le plus utile représente sans doute le navigateur web, qui permet de choisir les données à être visualisées sous forme de « pistes » (*tracks*) superposées avec la version de la séquence de référence humaine de son choix, les coordonnées génomiques et les bandes chromosomiques. Il est même possible de charger ses propres données pour les superposer aux pistes qui sont offertes. Ce site ne peut toutefois pas remplacer un bon logiciel de visualisation associé à une base de données interne, mais représente plutôt un outil supplémentaire pouvant être utilisé dans les cas où l'interprétation est problématique.

1.3 Pertes d'hétérozygotie

1.3.1 Définition

Au sens strict du terme, une perte d'hétérozygotie (LOH : *loss of heterozygosity*) réfère à l'absence de deux allèles différents à une position génomique donnée. Puisque les cellules humaines sont diploïdes, la majorité des séquences génomiques (sauf celles des chromosomes sexuels) est présente en deux copies. Chacun des parents transmettant une copie à leur enfant, ce dernier sera, pour chaque locus, soit hétérozygote (si les allèles reçus des parents sont différents) ou homozygote (si les allèles sont identiques). Le génome de chacun d'entre nous est donc un mélange d'hétérozygotie et d'homozygotie, qui peut être analysé par le génotypage de marqueurs tels que les microsatellites ou les SNPs, ou même par séquençage. De façon plus pratique, on parlera donc d'un segment de LOH si plusieurs marqueurs successifs montrent une absence d'hétérozygotie.

L'absence d'hétérozygotie n'est pas synonyme d'homozygotie, puisqu'elle peut aussi résulter d'une délétion. On parle alors d'hémizygotie. On utilise les termes « perte d'hétérozygotie à copie neutre » (CN-LOH : *copy neutral loss of heterozygosity*) ou « longues extensions contigües d'homozygotie » (LCSH : *long contiguous stretches of homozygosity*) pour faire référence à un LOH sur un segment génomique dont le nombre de copies est normal⁴¹. Les études effectuées chez des sujets sains ont plutôt tendance à parler de « suites », ou de « régions d'homozygotie » (ROH : *runs ou regions of homozygosity*)⁴²⁻⁴⁶. Toutes ces expressions signifient la même chose, et l'acronyme LCSH sera utilisé ici.

1.3.2 Ce que révèle la présence de pertes d'hétérozygotie

Le génome de pratiquement tous les individus contient, à divers degrés, des LCSHs⁴⁵. Lorsque l'on réfère à ces segments, le terme « autozygotie » serait plus juste puisque les allèles homozygotes constituant chacun d'eux ont une origine ancestrale commune⁴⁷. Ces LCSHs reflètent principalement le degré d'endogamie de la population dont un individu est issu, de même que le degré de consanguinité de ses parents, s'il y a lieu. Moins fréquemment, mais représentant un intérêt certain d'un point de vue clinique, les LCSHs peuvent également

signaler la présence de disomie uniparentale, à condition qu'au moins une portion du segment disomique soit isodisomique^{48,49}.

1.3.2.1 Endogamie & consanguinité

Plusieurs études ont identifié des LCSHs chez des individus sains dans diverses populations^{42-44,50,51}. Leur présence est le résultat de l'évolution des populations à partir d'ancêtres communs, au cours de laquelle le génome a été transmis de génération en génération via des blocs d'haplotypes générés par la recombinaison homologe durant la méiose. C'est pourquoi dans une population exogame — où les individus se reproduisent en choisissant un partenaire avec qui ils n'ont pas de lien de parenté ou qui est à l'extérieur de leur groupe social — le nombre et la taille des LCSHs sont variables, et ils sont répartis dans tout le génome⁴⁵. Lorsque le degré d'autozygotie retrouvé chez un individu est plus élevé, cela signale un certain degré de consanguinité parentale, qui peut être évaluée selon le pourcentage d'autozygotie présent^{41,47,50} (Tableau 1-1). Un pourcentage d'autozygotie intermédiaire peut également résulter de l'effet fondateur d'une sous-population et par conséquent être retrouvé chez des individus dont les parents n'ont aucun lien de parenté récent^{47,50,52}.

Tableau 1-1 : Corrélation entre le pourcentage de LCSH et le degré de parenté des parents.

Adapté de Kearney 2011 ⁴¹.

Lien de parenté	Degré	Coefficient d'endogamie	LCSH anticipé chez l'enfant (~%)^a
Parent/Enfant	Premier	0.25	25
Frères & Sœurs	Premier	0.25	25
Demi-frères & Demi-sœurs	Deuxième	0.125	12.5
Oncle/Nièce ou Tante/Neveu	Deuxième	0.125	12.5
Double cousin(e)s germain(e)s	Deuxième	0.125	12.5
Grand-parent/Petit-enfant	Deuxième	0.125	12.5
Cousin(e)s germain(e)s	Troisième	0.0625	6
Petit(e)s-cousin(e)s	Quatrième	0.03125	3
Arrière petit(e)s-cousin(e)s	Cinquième	0.015625	1.5
Arrière arrière petit(e)s-cousin(e)s	Septième	0.0039062	< 0.5

^a En supposant une population exogame

1.3.2.2 Disomie uniparentale

Normalement, chacun des parents transmet à son enfant un homologue de chaque paire de chromosomes. Lorsque les deux chromosomes homologues de l'enfant proviennent du même parent, on parle de disomie uniparentale (DUP; en anglais UPD : *uniparental disomy*). La DUP peut être soit iso (deux copies du même homologue parental) ou hétéro (une copie de chaque homologue parental). Également, la DUP peut affecter un chromosome en entier (disomie complète), ou seulement une portion du chromosome, auquel cas on parle de DUP segmentaire. La prévalence de la DUP est estimée à 1 naissance vivante sur 3500⁵³, et ne touche habituellement qu'un seul chromosome. En effet, bien qu'il soit théoriquement possible que plus d'un chromosome soit affecté par la DUP, cela est extrêmement rare et a été rapporté chez des individus chimériques ayant une lignée cellulaire normale, et une autre dont tout le complément chromosomique provient d'un seul parent (DUP pangénomique)⁴¹. Également, quelques cas de DUP pangénomiques en mosaïque ont récemment été rapportés chez des patients présentant un phénotype suggestif du syndrome de Beckwith-Wiedemann ou de Silver-Russell⁵⁴.

Les mécanismes menant à la DUP déterminent le type de disomie — complète vs segmentaire, hétérodisomie vs isodisomie — mais aussi le patron de LCSHs qui pourra être observé par l'analyse avec les micropuces SNP (seule l'isodisomie est directement détectable, voir section 1.4.2.1). Les mécanismes de formation de la DUP impliquent une erreur de division cellulaire ou de recombinaison pouvant se produire soit durant la méiose (Figure 1-3), ou durant la mitose (Figure 1-4). Dans tous les cas de non-disjonction, l'erreur résulte soit en une monosomie ou une trisomie, et c'est le sauvetage de cette erreur qui mène à la DUP.

La DUP résultant d'une non-disjonction méiotique touche le chromosome en entier et est soit isodisomique, hétérodisomique, ou une combinaison des deux. Si le complément chromosomique de départ du zygote comprend une monosomie suite à cette non-disjonction, peu importe que celle-ci ait eu lieu en méiose I ou II, le sauvetage s'effectue en dupliquant le chromosome présent, ce qui résulte en une isodisomie complète (Figure 1-3 E). Si le complément de départ comprend une trisomie, le sauvetage s'effectue via la perte d'un des trois chromosomes, et le type de DUP en résultant dépend de l'étape de la division à laquelle s'est produit la non-disjonction (méiose I ou II), et de s'il y a eu recombinaison homologue

durant la méiose. Une non-disjonction en méiose I résulte en une hétérodisomie complète s'il n'y a pas eu de recombinaison homologue (Figure 1-3 A), et en une hétérodisomie (incluant généralement la portion péricentromérique) combinée à un ou plusieurs segment(s) isodisomique(s) dépendamment du nombre de recombinaisons ayant eu lieu (Figure 1-3 B). Si la non-disjonction se produit en méiose II, la DUP est complètement isodisomique s'il n'y a pas eu de recombinaison homologue (Figure 1-3 C) ou isodisomique dans la région péricentromérique avec un ou plusieurs segment(s) hétérodisomique(s) s'il y a eu recombinaison (Figure 1-3 D).

La DUP résultant d'une erreur mitotique résulte toujours en une isodisomie, étant donné que le zygote avait au départ un complément chromosomique normal provenant des deux parents, et que le segment disomique provient soit d'une duplication ou de la réplication d'un homologue. Dans le cas d'une non-disjonction, la DUP résulte soit de la correction d'une monosomie par duplication du chromosome homologue présent (non illustré), ou de la correction d'une trisomie par perte de l'homologue unique (et donc conservation des deux copies de l'autre homologue) (Figure 1-4 B). La DUP segmentaire, quant à elle, peut provenir d'un événement de recombinaison mitotique ⁵⁵ (échange entre les chromatides des chromosomes homologues), et les deux lignées cellulaires résultantes présentent alors une isodisomie segmentaire qui est complémentaire (Figure 1-4 A). Même s'il s'agit d'isodisomie, ce type de DUP est indétectable par les méthodes moléculaires, à moins que l'échange ait eu lieu très tôt et que les deux lignées soient ensuite séparées (tissus embryonnaires vs extra-embryonnaires) ou qu'une des lignées ne soit pas viable et disparaisse ⁵⁶. Une disomie segmentaire peut également être le résultat d'une réparation de cassure effectuée via l'utilisation d'un brin sur le chromosome homologue comme matrice. Toutes les cellules-filles sont alors identiques et la disomie peut être détectée par les méthodes moléculaires.

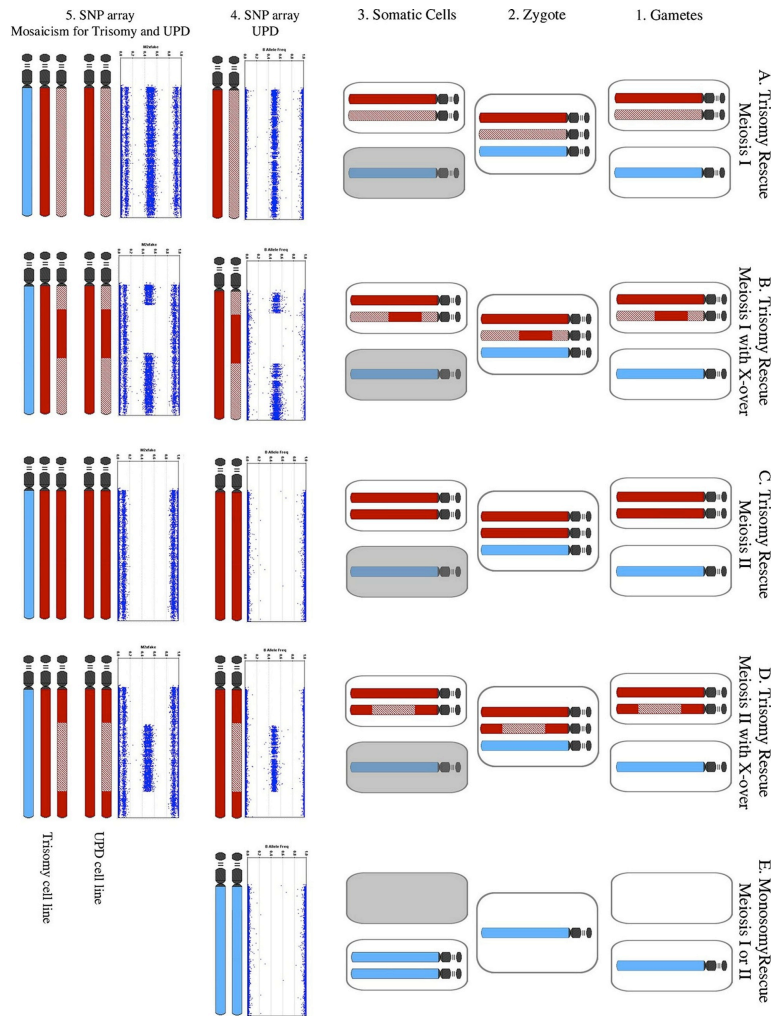


Figure 1-3 : Disomie uniparentale résultant d'erreurs méiotiques.

La couleur de chaque chromosome représente son origine. Rouge: homologue maternel provenant de la grand-mère; rose: homologue maternel provenant du grand-père; bleu pâle: chromosomes paternels (ici sans distinction d'homologues pour simplification). Les points bleus foncé correspondent aux patrons de fréquence de l'allèle B (BAF : *B Allele Frequency*) représentant le génotype SNP tel que détecté par une micropuce SNP. Les rangées A-E représentent chacune un type d'erreur méiotique: erreur en méiose I sans recombinaison homologue (A), ou avec recombinaison (B), erreur en méiose II sans (C), ou avec recombinaison (D), et sauvetage d'une monosomie résultant d'une erreur en méiose I ou II (E). Les colonnes 1-3 représentent la paire de chromosomes impliquée dans la disomie uniparentale à chaque stade de développement : des gamètes (1), au zygote (2), et à l'enfant (3). Les cellules ombragées en gris indiquent que les lignées cellulaires ne survivent pas et sont perdues. Les colonnes 4-5 représentent le patron de BAFs du génotypage SNP (à droite) résultant soit d'une disomie uniparentale complète (4), ou en mosaïque avec une lignée trisomique (5). Le complément chromosomique des homologues impliqués est modélisée à gauche du patron de BAFs pour chaque lignée cellulaire présente. *Tirée de Kearney 2011*⁴¹.

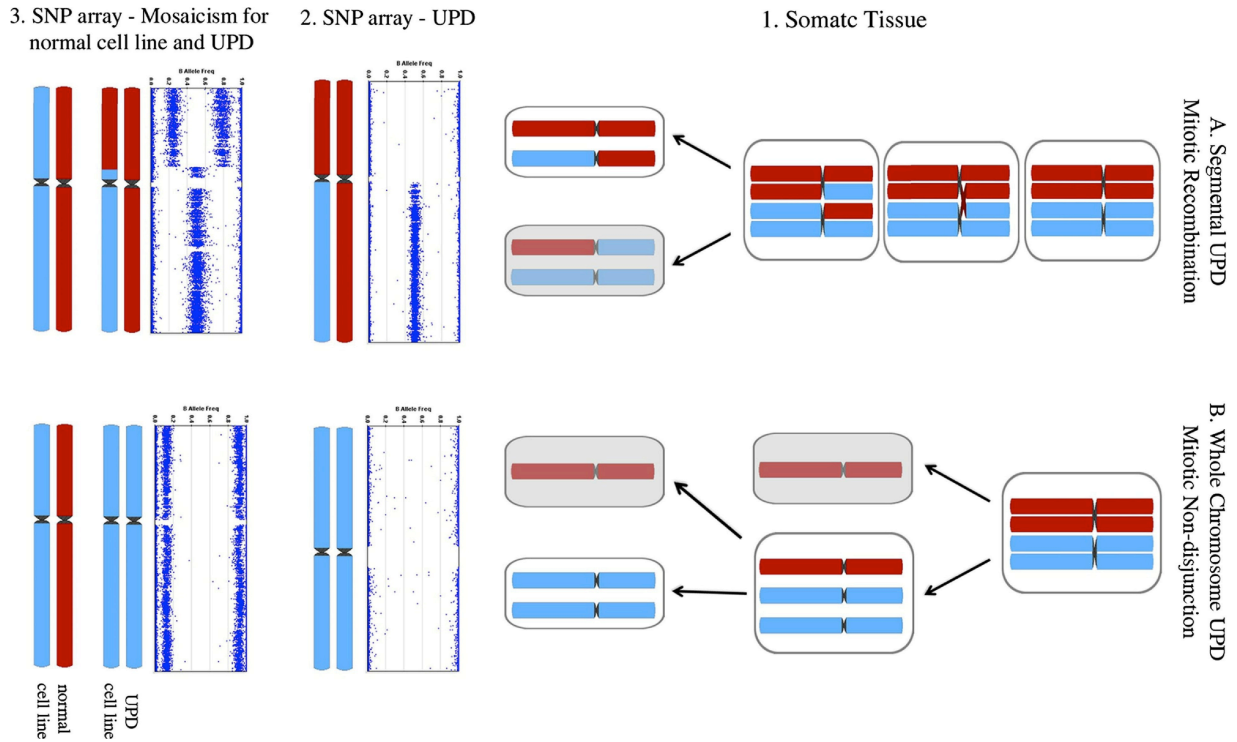


Figure 1-4 : Disomie uniparentale résultant d'erreurs mitotiques.

La couleur de chaque chromosome représente son origine. Rouge: homologue maternel; bleu pâle: homologue paternel. Les points bleus foncé correspondent aux patrons de fréquence de l'allèle B (BAF : *B Allele Frequency*) représentant le génotype SNP tel que détecté par une micropuce SNP. Les rangées A-B représentent chacune un type d'erreur mitotique: recombinaison mitotique (A), et non-disjonction mitotique (B). La colonne 1 représente la paire de chromosomes impliquée dans la disomie uniparentale au cours de la mitose, de la cellule-mère aux cellules-filles résultantes. Les cellules ombragées en gris indiquent que les lignées cellulaires ne survivent pas et sont perdues. Les colonnes 2-3 représentent le patron de BAFs du génotypage SNP (à droite) résultant d'une disomie uniparentale complète (2) ou en mosaïque avec une lignée normale (3). Le complément chromosomique des homologues impliqués est modélisée à gauche du patron de BAFs pour chaque lignée cellulaire présente.

*Tirée de Kearney 2011*⁴¹.

1.3.3 Mécanismes pathogéniques associés aux pertes d'hétérozygotie

L'intérêt de détecter les LCSHs en génétique clinique repose sur deux mécanismes pathogéniques : la transmission de mutations homozygotes, et le dérèglement de l'expression d'un ou plusieurs gène(s) soumis à l'empreinte parentale. Les pertes d'hétérozygotie résultant d'une consanguinité parentale ou d'une isodisomie parentale permettent d'identifier des régions génomiques pouvant potentiellement contenir des mutations homozygotes. De plus, dans le cas où la ou les perte(s) d'hétérozygotie détectée(s) correspond(ent) à une DUP (iso ou hétéro), les régions touchées pourraient contenir un ou des gène(s) soumis à l'empreinte parentale.

1.3.3.1 Transmission de mutations homozygotes

Les généticiens utilisent l'histoire familiale et le degré de consanguinité parentale depuis longtemps pour orienter leur diagnostic vers un désordre à transmission récessive lorsque c'est pertinent de le faire. Mais les moyens pour poser un diagnostic ou cibler les gènes à étudier chez un patient ne permettent de tester qu'un seul gène à la fois, ou nécessitent des analyses qui ne sont pas disponibles de routine dans un laboratoire clinique (cartographie par homozygotie). C'est pourquoi détecter ou confirmer le degré de consanguinité parentale est d'un intérêt clinique limité en soi, sauf s'il est également possible d'étudier directement le génome des segments autozygotes. En effet, il est connu depuis longtemps que la prévalence de certains désordres héréditaires, entre autres la déficience intellectuelle et les malformations congénitales, est plus élevée chez les enfants issus de parents ayant un certain degré de consanguinité ^{57,58}. Ceci est dû au fait que mathématiquement, un grand pourcentage d'autozygotie augmente le risque de transmission d'une mutation homozygote touchant un gène récessif ⁵⁹. Pouvoir à la fois détecter ces segments autozygotes et en étudier le contenu en gènes représente un avantage certain, surtout lorsque le tableau clinique d'un patient est non-spécifique et que le diagnostic s'avère difficile, puisque cela permet de réduire le nombre de gènes à étudier à ceux contenus à l'intérieur des segments autozygotes. Évidemment, l'identification d'un gène suspect dans un segment autozygote chez un enfant issu d'un couple

consanguin n'est pas suffisant pour établir un diagnostic, et les analyses moléculaires habituelles demeurent nécessaires pour confirmer le diagnostic suspecté.

Un autre mode de transmission d'une mutation homozygote s'effectue via une DUP, si celle-ci est iso. En effet, si la mutation hétérozygote d'un parent porteur est présente sur un chromosome ou un segment de chromosome isodisomique chez l'enfant, celui-ci aura alors deux copies de la même mutation, devenue par le fait même homozygote, et l'enfant sera affecté du désordre récessif malgré le fait qu'un seul de ses parents soit porteur⁵⁴. Encore une fois, une confirmation par les méthodes moléculaires habituelles est requise pour poser le diagnostic, mais l'identification d'un segment isodisomique peut grandement faciliter la recherche du gène en cause en réduisant le nombre de gènes à étudier. L'hétérodisomie parentale, quant à elle, ne peut résulter en la transmission d'une mutation homozygote (ou hétérozygote composée), à moins que le parent ne soit également atteint (parfois moins sévèrement que l'enfant), ou qu'il s'agisse d'un désordre à pénétrance réduite.

1.3.3.2 Gènes soumis à l'empreinte parentale

L'empreinte génomique — aussi dite parentale — est une marque épigénétique qui entraîne l'expression différentielle des deux allèles d'un gène selon qu'ils aient été hérités de la mère ou du père⁶⁰. Cette marque, réversible dans les cellules germinales, est établie durant la gamétogenèse, sur des régions génomiques différentes dans l'ovule et le spermatozoïde, et elle demeure intacte dans les cellules somatiques durant toute la vie embryonnaire et adulte⁶¹. Les résidus cytosine des régions génomiques soumises à l'empreinte parentale sont méthylés^{62,63}, réduisant au silence les gènes qui y sont localisés, ce qui fait qu'un seul allèle — maternel ou paternel — est exprimé⁶⁴. Cette empreinte parentale n'est pas le fruit du hasard et a une fonction importante, comme le démontrent les différentes pathologies causées par les disomies uniparentales de certains chromosomes⁶⁵, de même que les tératomes ovariens et les môles hydatiformes complètes résultant d'un complément chromosomique entièrement maternel⁶⁶ ou paternel^{67,68}, respectivement.

Plusieurs syndromes associés à l'empreinte parentale ont un phénotype bien défini. Dans la plupart des cas, on connaît le(s) gène(s), ou à tout le moins la région génomique en

cause⁶⁹. Par exemple, le premier locus identifié comme étant soumis à l’empreinte parentale, la région chromosomique 15q11.2-13, est associé aux syndromes de Prader-Willi (PWS : *Prader-Willi syndrome*) et d’Angelman (AS : *Angelman syndrome*) (Figure 1-5)⁷⁰. La sous-région PWS contient cinq gènes dont la copie maternelle, soumise à l’empreinte parentale, est normalement silencieuse, et la région AS contient deux gènes dont c’est au contraire la copie paternelle qui est normalement silencieuse. Une disomie maternelle de cette région ou du chromosome 15 en entier conduit donc au PWS, alors qu’une disomie paternelle mène au AS. Une délétion peut également mener à ces deux syndromes, et le phénotype résultant dépend alors de la copie manquante, une délétion maternelle résultant en un AS alors qu’une délétion paternelle cause un PWS.

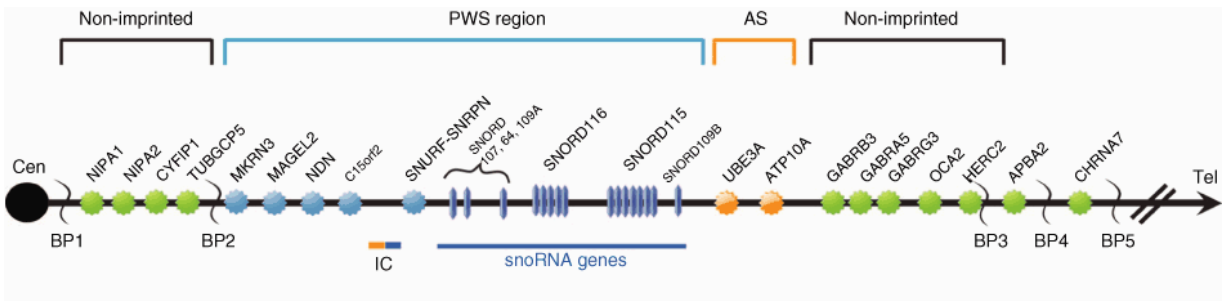


Figure 1-5 : Résumé de la carte génomique de la région chromosomique 15q11.2-q13.

La région PWS (en bleu) contient cinq gènes codant des protéines et six gènes codant des snoRNAs (*small nucleolar RNAs*) dont seule la copie paternelle est normalement exprimée. La région AS (en orange) contient deux gènes dont la copie paternelle est silencieuse, et ce, uniquement dans certains tissus. Les autres gènes (en vert) ne sont pas soumis à l’empreinte parentale et leur expression est donc biallélique. Tirée de Cassidy 2012⁷⁰.

1.4 Micropuces en cytogénétique clinique

Plusieurs techniques, allant du caryotype conventionnel au caryotype moléculaire, en passant par l'hybridation *in situ* en fluorescence, sont utilisées couramment dans les laboratoires de cytogénétique clinique (voir section 3.2). Si l'utilisation du marquage en bandes G (caryotype conventionnel) remonte aux années 1970, la technique de CGH est plus récente. D'abord effectuée sur des chromosomes en métaphase étalés sur une lame, puis sur des micropuces de BACs (*bacterial artificial chromosomes*), et enfin d'oligonucléotides, elle est maintenant également disponible sur des micropuces détectant à la fois le nombre de copies des séquences génomiques et le génotype de centaines de milliers de SNPs. Seront abordés dans cette section l'évolution de la CGH et des plateformes de génotypage de SNP, de même que les avantages et les limites associés à l'utilisation de chacune de ces techniques en cytogénétique clinique.

1.4.1 Hybridation génomique comparative (CGH)

1.4.1.1 Principe et évolution de la technique de CGH

L'hybridation génomique comparative a été mise au point dans les années 1990 ⁷¹. D'abord utilisée dans un cadre de recherche en oncologie, elle fut par la suite adoptée en cytogénétique clinique. Si les plateformes ont évolué avec le temps, le principe de base est demeuré le même. Il s'agit de mélanger en proportions égales l'ADN du patient à investiguer (marqué d'un fluorochrome vert par exemple) avec un ADN de référence (marqué en rouge par exemple), d'hybrider le mélange à des séquences génomiques « normales » et de comparer le signal émis par chacun afin de déterminer le nombre de copies des séquences génomiques chez le patient (Figure 1-6).

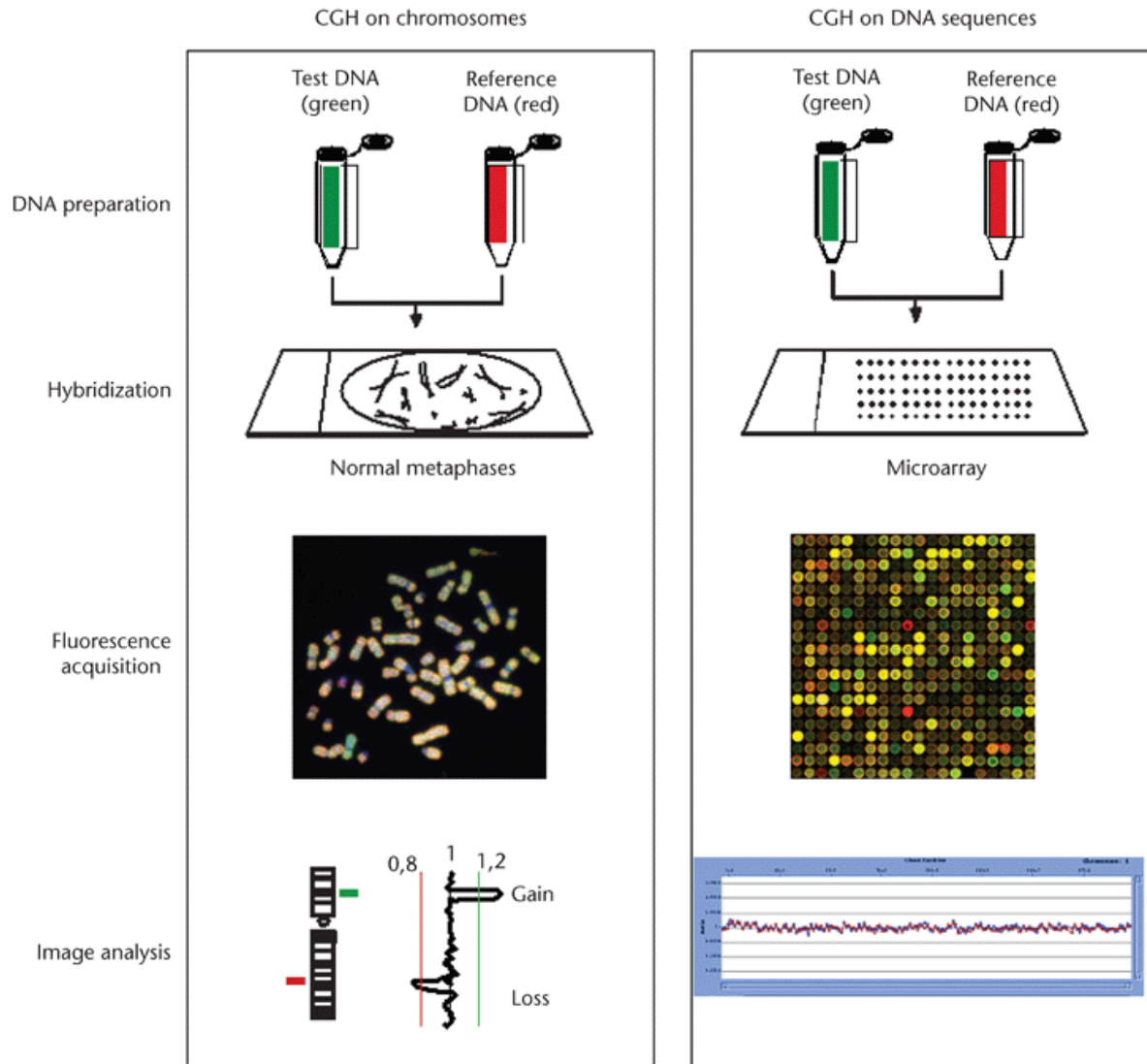


Figure 1-6 : Principe de la technique de CGH.

La technique de CGH, sur métaphase (à gauche) et sur micropuce (à droite), comprend quatre étapes: 1) la préparation de l'ADN à hybrider (mélange en proportions égales d'ADN du patient marqué en vert et d'ADN contrôle marqué en rouge), 2) hybridation du mélange d'ADN, soit sur une préparation de chromosomes normaux en métaphase ou sur une micropuce constituée de séquences d'ADN (BACs ou oligonucléotides), 3) acquisition de l'image à l'aide d'un scanner, 4) analyse de l'image afin de détecter les gains et les pertes de matériel génomique. *Tirée de Tachdjian 2009* ⁷².

La technique initiale consistait à hybrider le mélange d'ADNs marqués sur une préparation de chromosomes en métaphase étalés sur une lame ⁷³. On capturait ensuite une image digitale de la lame et, après normalisation des signaux, on mesurait l'intensité de chaque fluorochrome. Le ratio vert:rouge, quantifié sur toute la longueur de chaque chromosome, permettait de détecter une délétion (ratio vert:rouge 0.5:1) ou une duplication (ratio vert:rouge 1.5:1). Un nombre de copies égal entre l'ADN du patient et l'ADN de référence résultait en un ratio vert:rouge 1:1 et donc un signal de couleur jaune. Théoriquement, la résolution maximale de cette technique était d'environ 2 à 3 Mb ⁷⁴, mais dans les faits, il était difficile d'atteindre une résolution supérieure à 10 Mb en raison de la limite inhérente à la préparation chromosomique sur laquelle était faite l'hybridation ⁷⁵.

Vers la fin des années 1990, l'utilisation de micropuces constituées de milliers de PACs (*PI-derived artificial chromosomes*) ou de BACs a permis d'augmenter considérablement la résolution de la technique ^{76,77}. Au lieu d'hybrider le mélange d'ADN sur des chromosomes, et d'ensuite obtenir une image de ceux-ci, l'ADN est hybridé avec les sondes contenues dans des micropuits, le signal émis par chacun d'entre eux est ensuite mesuré, et un logiciel place chaque signal à la position chromosomique correspondant à la sonde. C'est ainsi que le caryotype moléculaire est reconstitué. Quelques années plus tard, la cartographie des BACs utilisés pour le séquençage du génome humain ^{78,79} a été mise à profit et a facilité la confection de micropuces couvrant le génome en entier, permettant d'obtenir un caryotype moléculaire dont la résolution réelle moyenne pouvait désormais atteindre 1 Mb ⁸⁰. La résolution des micropuces étant limitée par la taille des BACs, et par l'espacement (chromosomique) entre ceux-ci, il a été possible, en augmentant la densité de la couverture de certaines régions, d'en augmenter d'autant la résolution pour y détecter des anomalies aussi petites que 100 kb. Les micropuces sont dites « ciblées » lorsqu'elles ne couvrent que les régions chromosomiques associées à des syndromes connus, alors que celles couvrant tout le génome sont « pangénomiques ». Les BACs ont éventuellement été délaissés pour faire place aux oligonucléotides ^{33,81}, qui permettent d'avoir une couverture pangénomique dense et uniforme, tout en ciblant les régions souhaitées avec une couverture encore plus dense, à une résolution n'ayant en théorie pour seule limite que la longueur des oligonucléotides utilisés.

Des CNVs aussi petits que 0,5 kb ont en effet été identifiés grâce à des plateformes ciblant les exons⁸²⁻⁸⁴.

1.4.1.2 Avantages et limites de la CGH

La CGH comporte plusieurs avantages par rapport au caryotype conventionnel, le principal étant évidemment sa haute résolution. Le caryotype conventionnel a en effet une résolution limitée, allant d'environ 5 Mb (550 bandes) à 3 Mb pour un caryotype à haute résolution (850 bandes). Or, l'analyse d'un caryotype conventionnel à haute résolution demande du temps et un niveau élevé d'expertise, tant de la part du technicien que du cytogénéticien. Puisqu'il est beaucoup plus aisé de distinguer les anomalies chromosomiques au caryotype moléculaire, le travail d'analyse s'en trouve grandement réduit, de même que les risques de « manquer » une anomalie. Finalement, ce qui rend l'utilisation de la CGH réellement avantageuse, c'est sa capacité à détecter les anomalies chromosomiques cryptiques. Bien que d'autres méthodes de cytogénétique moléculaire, telles que la FISH, soient utilisées de routine pour la détection d'anomalies cryptiques, celles-ci nécessitent de cibler à l'avance les régions génomiques à tester, ou encore sont restreintes aux régions subtélomériques, alors que la CGH interroge le génome en entier en une seule expérience. Ceci présente un avantage certain, particulièrement pour investiguer les patients qui présentent un tableau clinique non spécifique ou compatible avec plus d'un syndrome génétique.

La nature de la technique amène d'autres avantages par rapport au caryotype conventionnel, mais aussi quelques inconvénients. En effet, les micropuces sont constituées de sondes (soit des BACs ou des oligonucléotides), et certaines séquences répétées (telles que celles retrouvées dans les centromères, les télomères et les satellites) en sont exclues. Les problèmes d'interprétation parfois associés aux chromosomes marqueurs et aux anneaux sont donc éliminés puisque seules les anomalies comprenant des séquences euchromatiques sont détectées. De plus, puisque l'analyse est généralement effectuée à partir d'ADN génomique provenant de cellules non cultivées, le résultat peut être disponible plus rapidement, et le risque de détecter des artéfacts de culture est éliminé. Pour la même raison, l'évaluation du pourcentage de mosaïcisme est plus représentative de la réalité. Également, comme le

caryotype moléculaire est reconstitué artificiellement à partir du signal émis par chaque sonde de la puce, les remaniements équilibrés ne sont pas détectés si tout le matériel génétique est présent. De ce fait, la CGH n'est pas appropriée dans les cas où l'on recherche ce type de remaniements (par exemple pour identifier les parents porteurs de translocation équilibrée), et c'est l'une des raisons pour lesquelles cette technique ne peut complètement remplacer le caryotype conventionnel. Il est toutefois à noter que plusieurs études ont montré que certains remaniements apparemment équilibrés au caryotype conventionnel ne l'étaient en fait pas⁸⁵⁻⁸⁷. Être capable de détecter ces anomalies cryptiques représente donc un atout pour l'investigation de patients présentant un phénotype anormal. Par ailleurs, la polyploïdie ne peut être détectée étant donné l'étape de normalisation du signal effectuée sur toute la lame. En effet, c'est la comparaison du signal de chaque puits avec la moyenne du signal mesuré sur toute la lame qui permet de déterminer qu'une région génomique est délétée ou dupliquée. Il est donc impossible de détecter un déséquilibre si c'est l'ensemble du génome qui est touché. Enfin, la détection de mosaïques, bien que possible^{88,89}, pourrait être problématique dans certains cas. En effet, la différence de signal pourrait être atténuée par la présence d'une lignée cellulaire normale, ou même complètement masquée dans les cas où deux lignées sont présentes dans des proportions similaires et ont des anomalies complémentaires (par exemple, une lignée 45,X et une lignée 47,XXX, présentes à 50% chacune). En utilisant des paramètres standard de détection, l'aCGH peut détecter une aneuploïdie en mosaïque présente dans plus de 20 à 30% des cellules, et une aneuploïdie segmentaire en mosaïque à 40 à 50%⁸⁹, bien que selon notre expérience, ces dernières peuvent être détectées à un niveau plus bas (25 à 30%). La sensibilité de l'aCGH est donc similaire à celle du caryotype conventionnel effectué selon les recommandations standard, qui détecte une mosaïque présente dans 26% des cellules avec un degré de confiance à 95%⁹⁰. Il est toutefois possible, en modifiant certains paramètres de l'aCGH, de détecter un niveau de mosaïcisme aussi bas que 10%, de la même façon qu'il est possible de détecter un niveau de mosaïcisme plus bas au caryotype conventionnel en analysant un plus grand nombre de cellules.

1.4.2 Génotypage SNP sur micropuce

1.4.2.1 Principe du génotypage SNP sur micropuce

La cartographie par homozygotie est une stratégie utilisée en recherche depuis de nombreuses années qui a permis l'identification de plusieurs gènes impliqués dans des désordres à transmission récessive⁹¹⁻⁹⁴. D'abord effectuée avec des méthodes analysant les microsatellites, les micropuces SNP ont maintenant pris le relais dans plusieurs centres à causes des nombreux avantages qu'elles présentent. Elles sont couramment utilisées dans le cadre d'études d'association pangénomiques et d'analyse de liaison, ou de cartographie par homozygotie, de même qu'en oncologie pour mieux caractériser les tumeurs et prédire la réponse aux traitements^{54,95}. Plus récemment, plusieurs laboratoires de cytogénétique se sont intéressés à ces micropuces étant donné leur haute résolution et l'information supplémentaire fournie par la détection des pertes d'hétérozygotie (voir section 1.3.2).

Le génotypage de SNP sur micropuce peut se faire de plusieurs façons, et chaque fabricant a mis au point sa propre méthode. Seules les méthodes utilisées par Affymetrix et Illumina seront mentionnées ici, puisque les micropuces SNP testées dans le cadre des travaux rapportés dans ce mémoire provenaient de ces deux fabricants. D'abord, contrairement à la CGH, le génotypage de SNP sur micropuce ne nécessite pas l'hybridation d'un ADN de référence, mais seulement celui du patient. Pour les puces Illumina, les sondes présentes dans chacun des puits correspondent à la séquence génomique précédant le SNP à génotyper, auxquelles l'ADN du patient s'hybride. Une étape d'élongation est ensuite effectuée avec des nucléotides marqués de façon différentielle, et c'est le signal émis à ce moment qui permet de déterminer le génotype du patient pour ce SNP (Figure 1-7 A). Dans le cas d'Affymetrix, chaque SNP est représenté par une paire de sondes dans deux puits différents, correspondant aux deux génotypes possibles. C'est la présence ou l'absence d'un signal dans chaque puits qui permet de déterminer le génotype pour ce SNP (Figure 1-7 B). Cette information est détectée pour tous les SNPs testés, puis rapportée sous forme de fréquence de l'allèle B (BAF : *B Allele Frequency*) à la position génomique correspondante sur toute la longueur des chromosomes. De plus, l'intensité du signal dans chaque puits est mesurée, comparée *in silico* à un ADN de référence, et rapportée sous forme de log ratio à la

position génomique correspondante. La visualisation des BAFs permet ensuite de voir directement les pertes d'hétérozygotie, en plus de confirmer le nombre de copies déterminé grâce à l'intensité du signal (Figure 1-7 C). Les LCSHs, quant à eux, correspondent aux segments présentant simultanément une perte d'hétérozygotie et un nombre de copies égal à 2 (non-illustré).

1.4.2.2 Avantages et limites des micropuces SNP

Le caryotype moléculaire effectué à l'aide des micropuces SNP présente les mêmes avantages que lorsqu'il est effectué avec la technique CGH (détection d'anomalies cryptiques, non-détection des anomalies impliquant les régions hétérochromatiques), en plus d'éliminer certains inconvénients. En effet, la résolution des micropuces SNP atteint ou excède celle de la plupart des micropuces CGH, et elles aussi ne couvrent que les régions euchromatiques des chromosomes. De plus, la visualisation du patron de BAFs fournis par le génotypage de SNPs permet la détection de polyploïdie (Figure 1-8 A), contrairement à la CGH, en plus d'apporter une meilleure sensibilité que cette dernière pour la détection de mosaïcisme. En effet, il est possible de détecter le mosaïcisme à un niveau allant de 5 à 20%, dépendamment des paramètres et de l'anomalie chromosomique ⁹⁶ (Figure 1-8 B). Par contre, tout comme pour la CGH, la détection des remaniements équilibrés n'est pas possible.

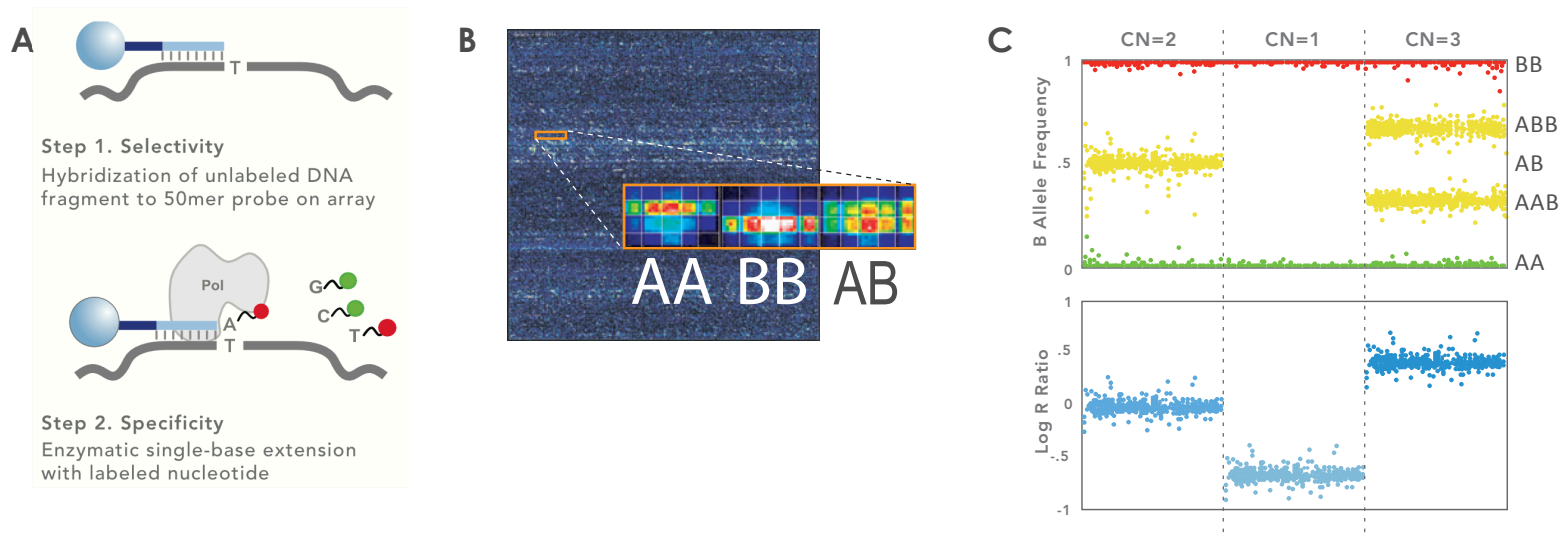


Figure 1-7 : Principe du génotypage SNP sur micropuce.

A | Illumina. L'ADN du patient est fragmenté puis hybridé sur la micropuce. Chaque fragment (segment gris foncé) s'hybride à la sonde complémentaire (bleu pâle) maintenue à la micropuce par une bille (cercle bleu pâle). Une polymérase (*Pol* en gris pâle) vient ensuite allonger la sonde avec le nucléotide (marqué en rouge ou vert) complémentaire au SNP présent chez le patient, ce qui produira le signal fluorescent correspondant au génotype (C). **B** | Affymetrix. La micropuce contient une paire de sondes pour chaque SNP (une sonde pour chaque génotype). L'ADN marqué du patient est hybridé à la micropuce, et le signal émis dans chaque puits permet de détecter la présence (signal rouge, orange, jaune ou vert) ou l'absence (bleu foncé) du génotype correspondant. **C** | Patterns des signaux mesurés (*B Allele Frequency* et *Log R Ratio*) en fonction du génotype SNP (BB, ABB, AB, AAB ou AA) et du nombre de copies (CN) présents. En haut : chaque point représente un allèle et son génotype exprimé en fréquence de l'allèle B (BAF) : rouge = homozygote BB (BAF=1); jaune = hétérozygote AAB, AB ou ABB (BAF= 0,33 à 0,66); ratio vert = homozygote AA (BAF=0). En bas : le nombre de copies est exprimé en \log_2 ratio, chaque point représentant une sonde à une position génomique spécifique : 1 copie = -1; 2 copies = 0; 3 copies = 0,5. *Adaptée de Illumina, Inc. 2012*⁹⁷ et *Affymetrix, Inc. 2009*⁹⁸.

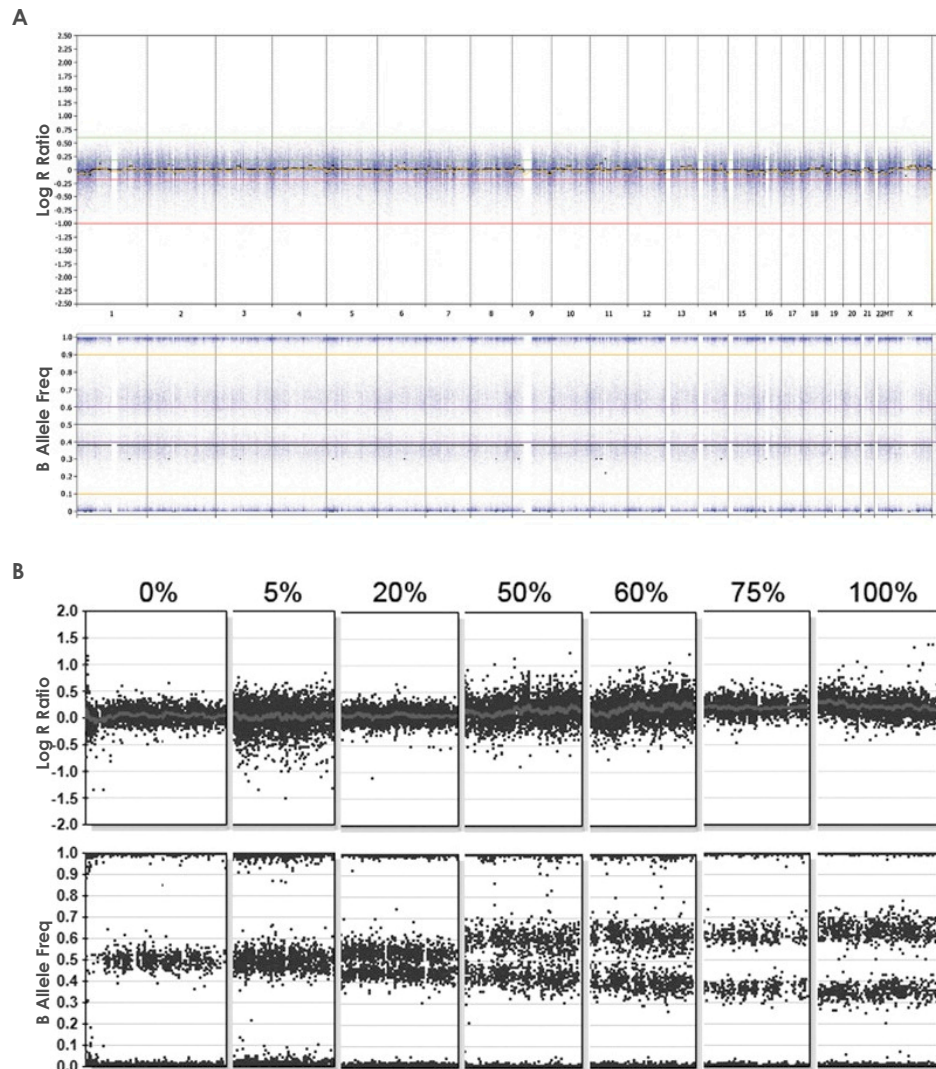


Figure 1-8 : Patterns de fréquence de l'allèle B (BAF : *B Allele Frequency*) et log ratios détectés par les micropuces SNP en cas de polyplôidie et de mosaïcisme.

Patterns des signaux mesurés (*B Allele Frequency* et *Log R Ratio*) en fonction du génotype SNP (BB, ABB, AB, AAB ou AA) et du nombre de copies (CN) présents. Chaque point violet (A) et noir (B) correspond soit au nombre de copies (exprimé en \log_2 ratio) à une position génomique spécifique (partie du haut), ou au génotype SNP (exprimé en BAF) pour un allèle donné (en bas). **A** | Triploïdie : 69,XXX. Le génome entier est représenté, chaque chromosome étant séparé par une ligne verticale. La triploïdie résulte en un \log_2 ratio normal à 0 (en haut) et un patron de BAFs dont les points s'alignent à environ 0,4 et 0,6 au lieu de 0,5, ce qui indique la présence de 3 allèles. **B** | Trisomie en mosaïque. Chaque paire d'encadrés représente une partie du génome chez 7 patients ayant un pourcentage différent de cellules avec trisomie (de 0 à 100%). Le \log_2 ratio augmente proportionnellement au pourcentage de trisomie présent, et le patron de BAFs passe progressivement de 0,5 à 0,33 et 0,66, ce qui indique la présence des 3 allèles (trisomie). *Adaptée de Srebniak 2011⁹⁹ et Conlin 2010⁹⁶.*

Un des principaux atouts des micropuces SNP est leur capacité à détecter les pertes d'hétérozygotie. Tel que discuté dans la section 1.3, leur identification peut révéler ou suggérer la présence d'une anomalie comme une DUP ou une mutation homozygote, et ainsi aider à établir le diagnostic. Bien que d'autres techniques permettent de détecter les DUPs, le caryotype moléculaire effectué sur une micropuce SNP permet d'effectuer cette analyse simultanément, sans analyse supplémentaire ou suspicion préalable. Il faut toutefois être prudent dans l'interprétation d'un résultat négatif, puisque seules les isodisomies sont détectables (voir section 1.3.2.2, Figure 1-3 et Figure 1-4). En effet, puisque ce sont les BAFs qui sont visualisés, il est impossible d'exclure une hétérodisomie (Figure 1-3 A) sans également effectuer l'analyse des parents et comparer les génotypes directement. En ce qui concerne les mutations homozygotes, il importe de rappeler que dans un contexte constitutionnel ou prénatal, les LCSHs détectés n'ont aucune valeur diagnostique en soit, mais permettent plutôt d'orienter les analyses moléculaires afin d'en arriver à un diagnostic. Toutefois, la possibilité d'appliquer la méthode de cartographie par autozygotie, de routine, à des patients pour qui le diagnostic s'avère difficile, représente — à tout le moins théoriquement — un avantage non négligeable. En plus d'être utiles pour la détection des pertes d'hétérozygotie, les patrons de BAFs peuvent également être utilisés pour caractériser certaines anomalies complexes et déduire le modèle de transmission, à condition de procéder à l'analyse de l'ADN parental ^{48,96}.

Malgré tous ces avantages, les micropuces SNP ont également d'importantes limites. Initialement, elles furent mises au point non pas pour détecter les CNVs, mais bien pour génotyper les SNPs dans le cadre d'études pangénomiques. De ce fait, la couverture génomique est biaisée vers les régions riches en SNPs et n'est pas uniforme. Par conséquent, les CNVs présents dans les régions moins bien couvertes risquent de ne pas être détectés ¹⁰⁰, même s'ils ont une signification clinique connue. De plus, puisque la sélection des oligonucléotides fut effectuée en fonction de leur performance à génotyper les SNPs, et qu'aucun ADN de référence n'est hybridé avec celui du patient, ces micropuces tendent à produire un important bruit de fond n'affectant pas le génotypage, mais réduisant la fiabilité du signal utilisé pour calculer le log ratio servant à la détection des CNVs ¹⁰¹.

1.4.3 Le caryotype moléculaire aujourd'hui

1.4.3.1 Combinaison CGH / génotypage SNP

Les plateformes de caryotypage moléculaire évoluent rapidement. D'une part, la qualité des micropuces basées sur la CGH est éprouvée et la résolution des plateformes les plus récentes atteint un niveau permettant la détection de CNVs aussi petits que 10 kb dans certaines régions chromosomiques. Par ailleurs, les micropuces SNP — plus denses — ont traditionnellement une résolution beaucoup plus élevée (jusqu'à 0,5 kb) et permettent en plus la détection des pertes d'hétérozygotie, ce qui génère un intérêt légitime pour leur utilisation en cytogénétique. Leur performance pour la détection des CNVs était toutefois bien inférieure aux plateformes CGH^{100,101}, si bien que les manufacturiers ont mis au point des micropuces « hybrides » élaborées spécifiquement pour la cytogénétique, combinant leur technologie de génotypage de SNP à des oligonucléotides non-polymorphiques sélectionnés strictement pour la détection de CNVs, et pour lesquels seule l'intensité du signal est mesurée.

1.4.3.2 Utilisation actuelle des micropuces en clinique

La majorité des laboratoires de cytogénétique clinique ont d'abord implanté les micropuces CGH, et leur objectif était en continuité avec celui du caryotype conventionnel, soit de détecter des anomalies chromosomiques avec la meilleure résolution possible. Le bénéfice apporté par cette hausse de résolution ne fait plus de doute chez les patients présentant une déficience intellectuelle (syndromique ou non), et les lignes directrices actuelles le reflètent puisque le caryotype moléculaire est l'analyse génétique de première intention recommandée chez ces patients (voir sections 3.2 et 4.2).

Par contre, le rendement du caryotype moléculaire dans un contexte prénatal n'est pas aussi bien défini et le rapport risques/bénéfices est controversé. En effet, les études investiguant des fœtus ont été effectuées soit de façon rétrospective, avec des micropuces ciblées, ou encore dans une population hétérogène, et rapportent un rendement variable, ce qui ne permet pas de connaître le rendement réel d'une micropuce CGH pangénomique chez des fœtus à haut risque d'anomalies chromosomiques. De plus, les problèmes d'interprétation

associés à la détection de CNVs dont la signification clinique est incertaine représentent un défi particulier dans un contexte de grossesse (voir section 3.2).

Jusqu'à tout récemment, l'utilisation des micropuces SNP était plus répandue en recherche que dans un contexte clinique. Plusieurs études ont investigué des cohortes de patients atteints de déficience intellectuelle ¹⁰²⁻¹⁰⁶, mais la plupart de ces études ne se sont attardées qu'à la détection de CNVs, et une seule d'entre elles a rapporté les résultats fournis par la détection des pertes d'hétérozygotie ¹⁰². De plus, la résolution des micropuces utilisée était inférieure à ce qui est actuellement disponible, et le seuil de détection des CNVs utilisé était généralement de 100 à 200 kb, ce qui est comparable ou même plus élevé que le seuil pouvant être utilisé avec certaines micropuces CGH. On dispose donc de peu d'information quant aux avantages réels que peuvent procurer la très haute résolution et la détection des LCSHs par les micropuces SNP dans un contexte clinique.

2 HYPOTHÈSES ET OBJECTIFS

2.1 Hypothèses

La technique de CGH sur micropuce permet de détecter une anomalie chromosomique cryptique chez 10% à 14% des patients atteints de déficience intellectuelle et/ou de malformations congénitales^{29,107,108}. En comparaison, le caryotype conventionnel détecte une anomalie chez 3,7% dans cette même population de patients¹⁰⁹. Depuis août 2008, le laboratoire de cytogénétique du CHU Sainte-Justine utilise une plateforme CGH pangénomique pour effectuer le caryotype moléculaire. Nous avons voulu étudier l'utilité clinique des micropuces en cytogénétique clinique dans les contextes prénatal et postnatal, et les travaux rapportés dans ce mémoire comportent donc deux volets.

2.1.1 CGH pangénomique en prénatal

Le caryotype conventionnel identifie une anomalie chromosomique dans 2% à 7% de toutes les grossesses pour lesquelles une amniocentèse est effectuée^{110,111}, mais cette proportion peut aller jusqu'à 35% chez les fœtus qui présentent des malformations multiples¹¹². Nous suggérons que l'utilisation d'une micropuce CGH pangénomique dans un contexte prénatal pourrait permettre d'identifier des anomalies chromosomiques pathogéniques qui ne peuvent être détectées par le caryotype conventionnel, dans une proportion similaire à ce qui est obtenu dans un contexte postnatal.

Nous avons choisi de restreindre la population prénatale à étudier pour plusieurs raisons. Premièrement, le caryotype conventionnel obtient un rendement diagnostique plus élevé chez les fœtus présentant des malformations détectées à l'échographie, et il semble que ce soit également le cas pour le caryotype moléculaire¹¹³. Ensuite, la proportion de résultats ayant une signification clinique incertaine n'étant pas encore établie pour les micropuces pangénomiques, nous croyons qu'il est éthiquement plus prudent de ne pas offrir cette analyse aux patientes portant un fœtus apparemment normal pour l'instant. Finalement, l'investigation

d'une population plus homogène permettra d'obtenir un rendement diagnostique plus précis et valable.

2.1.2 Combinaison CGH / géotypage SNP en postnatal

Le rendement diagnostique du caryotype moléculaire est maintenant bien établi chez les patients atteints de déficience intellectuelle et/ou de malformations. Malgré tout, près de la moitié de ces patients restent sans diagnostic ¹¹⁴. Les causes génétiques de la déficience intellectuelle sont variées et incluent les anomalies chromosomiques (visibles au microscope, cryptiques, en mosaïque), les mutations ponctuelles et dynamiques (répétitions de trinuécléotides), et la disomie uniparentale. Nous suggérons que pour ces patients, effectuer le caryotype moléculaire avec une micropuce à très haute résolution et offrant le géotypage de SNP pourrait améliorer le rendement diagnostique de deux façons : 1) en permettant la détection d'anomalies de très petite taille; 2) en utilisant l'information fournie par la détection des pertes d'hétérozygotie pour cibler les analyses moléculaires subséquentes, en fonction du phénotype et des gènes présents. Nous anticipons également que la hausse de résolution amènera la détection d'un plus grand nombre de CNVs non pathogéniques (de signification clinique incertaine et bénins) et souhaitons établir l'ampleur de cette hausse.

2.2 Objectifs

2.2.1 CGH pangénomique en prénatal

Ce volet consistait à évaluer l'utilité clinique d'une micropuce CGH pangénomique dans un contexte prénatal. Pour ce faire, nous souhaitions : 1) déterminer le rendement diagnostique d'une telle plateforme dans un contexte prénatal, en étudiant les résultats obtenus dans une cohorte de fœtus avec malformations majeures détectées à l'échographie, et pour qui le caryotype conventionnel est normal; 2) établir la proportion de fœtus chez qui un CNV de signification clinique incertaine (VOUS : *variant of uncertain clinical significance*) est détecté, puisque leur identification complique le conseil génétique.

2.2.2 Combinaison CGH / géotypage SNP en postnatal

Le second volet visait principalement deux objectifs, soit vérifier l'utilité clinique d'une micropuce à très haute résolution chez des patients atteints de déficience intellectuelle, et évaluer la pertinence de détecter les pertes d'hétérozygotie chez ces mêmes patients. Plus spécifiquement, nous voulions : 1) déterminer si la hausse de résolution offerte par les micropuces SNP permet de détecter un plus grand nombre de CNVs pathogéniques; 2) établir la fréquence à laquelle ces micropuces détectent des CNVs bénins et des VOUS; 3) définir la fréquence à laquelle la détection de pertes d'hétérozygotie est informative et permet d'orienter le diagnostic; 4) vérifier si les logiciels de visualisation fournis avec les différentes plateformes permettent une interprétation adéquate des résultats dans un contexte clinique; 5) évaluer la charge de travail et les défis associés à l'utilisation de ces micropuces.

3 WHOLE-GENOME ARRAY CGH IDENTIFIES PATHOGENIC COPY NUMBER VARIATIONS IN FETUSES WITH MAJOR MALFORMATIONS AND A NORMAL KARYOTYPE

Clinical Genetics 81:128–141 (2012) doi: 10.1111/j.1399-0004.2011.01687.x

Guylaine D'Amours^{a,b,c}, Zoha Kibar^b, Géraldine Mathonnet^a, Raouf Fetni^{b,c,d,e}, Frédérique Tihy^{a,b,c,e}, Valérie Désilets^{a,c,f}, Sonia Nizard^{a,c,f}, Jacques L. Michaud^{a,b,c,f}, Emmanuelle Lemyre^{a,b,c,f}

^a Service de génétique médicale, CHU Sainte-Justine, Montréal, QC, Canada, ^b Centre de recherche, CHU Sainte-Justine, Montréal, QC, Canada, ^c Faculté de médecine, Université de Montréal, Montréal, QC, Canada, ^d Département de pathologie, CHU Sainte-Justine, Montréal, QC, Canada, ^e Pathologie et biologie cellulaire, Université de Montréal, Montréal, QC, Canada, ^f Pédiatrie, Université de Montréal, Montréal, QC, Canada

Corresponding author

Emmanuelle Lemyre, MD, FRCPC, FCCMG

Service de génétique médicale

CHU Sainte-Justine

3175 Chemin de la Côte-Sainte-Catherine

Montréal, QC, Canada H3T 1C5

Tel: 1 (514) 345-4931 ext. 6342

Fax: 1 (514) 345-4766

3.1 Abstract

Despite a wide range of clinical tools, the etiology of mental retardation and multiple congenital malformations remains unknown for many patients. Array-based comparative genomic hybridization (aCGH) has proven to be a valuable tool in these cases, as its pangenomic coverage allows the identification of chromosomal aberrations that are undetectable by other genetic methods targeting specific genomic regions. Therefore, aCGH is increasingly used in clinical genetics, both in the postnatal and the prenatal settings. While the diagnostic yield in the postnatal population has been established at 10-12%, studies investigating fetuses have reported variable results. We used whole-genome aCGH to investigate fetuses presenting at least one major malformation detected on ultrasound, but for whom standard genetic analyses (including karyotype) failed to provide a diagnosis. We identified a clinically significant chromosomal aberration in 8.2% of tested fetuses (4/49), and a result of unclear clinical significance in 12.2% of tested fetuses (6/49). Our results document the value of whole-genome aCGH as a prenatal diagnostic tool and highlight the interpretation difficulties associated with CNVs of unclear significance.

Key Words

Array comparative genomic hybridization, chromosome aberrations, congenital abnormalities, copy number variations, fetus, genotype-phenotype correlations, prenatal diagnosis, prenatal ultrasonography.

3.2 Introduction

Congenital malformations are found in approximately 3% of liveborn infants, and GTG-banding karyotype identifies a chromosomal abnormality in 6% of them (1). While karyotyping identifies a chromosomal imbalance in 2% to 7% of all pregnancies undergoing amniocentesis (2, 3), fetuses with abnormal ultrasound findings have an abnormal karyotype in 9% to 35% of cases, depending on the presence of a single anomaly or multiple malformations (4-6). In addition to karyotyping, fluorescence *in situ* hybridization (FISH) with subtelomeric probes also allows the detection of chromosomal imbalances in individuals with unexplained multiple congenital anomalies and/or mental retardation (7) and fetuses with major malformations (8). Locus-specific probes also allow the diagnosis of interstitial cryptic imbalances if the observed features or ultrasound findings are suggestive of a particular microdeletion or microduplication syndrome. In addition to FISH, quantitative PCR based methods, such as multiplex ligation probe amplification (MLPA), and quantitative multiplex PCR of short fragments (QMPSF) can also be used with specific probes to detect or confirm cryptic chromosomal aberrations (9-12). But despite the wide use of these methods and other clinical investigative tools, the etiology of congenital malformations and mental retardation remains unknown in a large proportion of cases (13, 14). This makes genetic counseling difficult, particularly in the context of a pregnancy.

During the last decade, various studies employing array-based comparative genomic hybridization (aCGH) to identify chromosomal aberrations have reported promising results. As many as 10-12% of individuals with developmental delay (DD) or multiple congenital anomalies (MCA) have a submicroscopic chromosomal imbalance that can be identified by aCGH (15, 16). These results have recently led to the replacement of karyotyping by aCGH as the first-line genetic test for patients with unexplained DD or MCA (16). The detection of submicroscopic aberrations with aCGH has also been studied in the prenatal setting (17-23). The proportion of clinically significant results reported in those studies is highly variable, ranging from 1.3% to 16%; thus, the diagnostic yield to expect from aCGH in the prenatal setting is still unclear. Karyotyping finds the highest proportion of chromosomal aberrations in fetuses with multiple malformations, and it appears that aCGH also detects the highest proportion of pathogenic copy number variations (CNVs) in this subgroup of fetuses (24).

Furthermore, the detection of CNVs of unclear significance by whole genome aCGH represents an important counseling challenge that might increase the risk of terminating an otherwise normal pregnancy. Few studies have focused on fetuses with malformations (25-27), and only two have used a high-density array (28, 29). The aim of the present study was to assess the use of whole-genome aCGH as a diagnostic tool in the investigation of fetuses with unexplained major malformations detected by ultrasound, and to help establish the diagnostic yield, as well as the proportion of results of unclear clinical significance to expect. We found that whole-genome aCGH allowed the identification of new pathogenic chromosomal abnormalities that would not have been found otherwise in 12.2% of testes fetuses, while keeping the number of unclear results to a manageable level. We also review previous prenatal aCGH studies, focusing on the results obtained for karyotypically normal fetuses presenting major malformations, and compare them with our results.

3.3 Materials and Methods

3.3.1 Patients and samples

Patients referred for prenatal diagnosis at Centre Hospitalier Universitaire Sainte-Justine were offered aCGH analysis if they met the following criteria: 1) ultrasound showed at least one major malformation; 2) GTG-banded karyotype was normal at the resolution obtained (n=47) or could not be performed (n=2); and 3) the malformation(s) observed remained unexplained after standard investigations (Supporting information : Table 3-S1). Targeted FISH analyses, which were performed in a subset of patients, did not show any abnormality (details in Supporting information, section 3.8.1.2). Samples used for aCGH were obtained from amniocentesis (n=30), chorionic villus sampling (CVS) (n=4), cordocentesis (n=2), placenta (n=2), or tissue biopsy performed during fetal autopsy (n=11), as part of standard clinical procedures. Array CGH was performed either for prenatal diagnosis during an ongoing pregnancy (n=21), or for diagnosis and genetic counseling after interrupted pregnancy (spontaneous or terminated) (n=28).

3.3.2 Array CGH analysis and interpretation

All samples were analyzed with one of four custom-designed microarrays (Signature Genomics, Spokane, WA; Table 3-1), according to previously described methods (30), with minor modifications detailed in Supporting information (section 3.8.1.1). Each patient's data were accessed and interpreted using a custom aCGH analysis software (Genoglyphix[®]; Signature Genomics). All nucleotide positions mentioned refer to the March 2006 human reference sequence (NCBI Build 36.1).

Table 3-1 : Summary of the aCGH results obtained with each type of array.

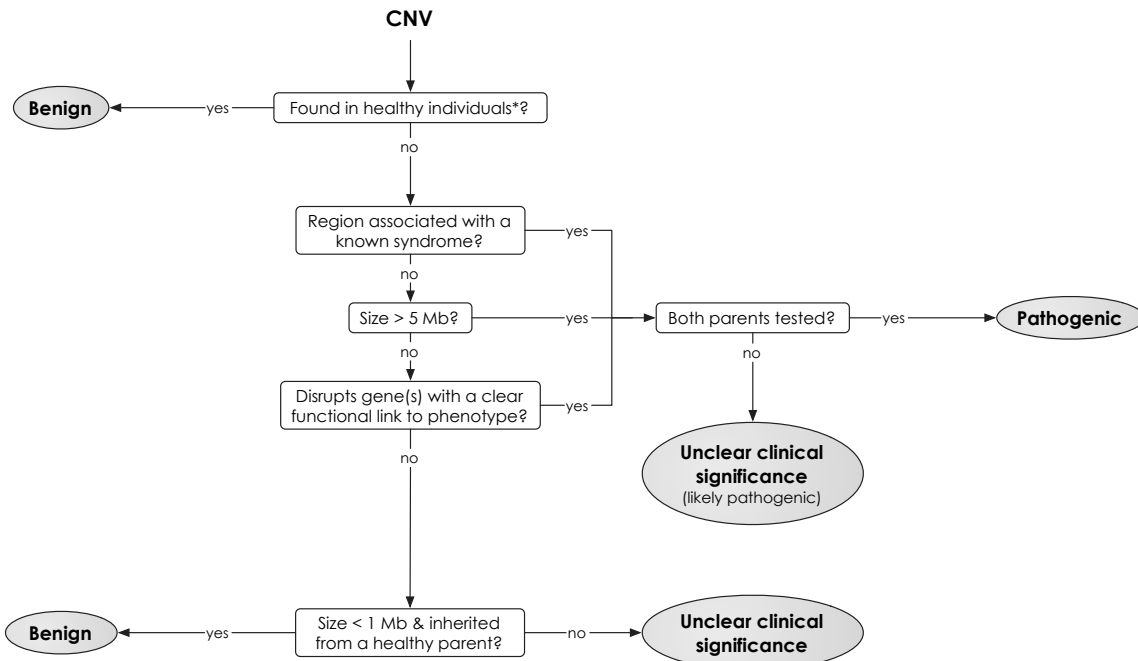
Array	Density	Result			Total fetuses
		Pathogenic CNV	Unclear CNV	Normal	
WG	4685 BACs	0	0	7	7
OS 1.1	105 000 oligos	1	2	13	16
OS 2.0 / CGX-12	135 000 oligos	3	4	19	26
Total		4	6	39	49

CGX-12, NimbleGen CGX-12; OS 1.1, SignatureChip Oligo Solution v1.1 Rev B; OS 2.0, SignatureChip Oligo Solution v2.0 12-plex; WG, SignatureChip WG v1.0.1.

Copy number gains and losses were interpreted individually by comparing each genomic region to information available in public and private databases (links available in Supporting Information, section 3.8.1.4), and classified into one of three categories: benign, pathogenic, or CNV of unclear clinical significance (Figure 3-1). First, a CNV was classified as benign and was not reported if it was previously reported in healthy subjects (DGV, Database of Genomic Variants; GCAD, Genoglyphix Chromosome Aberration Database (31); or our internal laboratory database). Secondly, we reported a CNV as pathogenic if: 1) it corresponded to a region of known clinical relevance (DECIPHER; Online Mendelian Inheritance in Man, OMIM; or scientific literature retrieved through MEDLINE) and the fetus clearly presented the expected phenotype; 2) it disrupted gene(s) (RefSeq) clearly linked to the patient’s phenotype; or 3) its size was larger than 5 Mb. Only CNVs for which we could establish inheritance status were classified as pathogenic. If a CNV was considered pathogenic but one or both parents were unavailable for testing, we deemed it likely pathogenic, but reported it as being of unclear clinical significance. Thirdly, a small CNV (< 1 Mb) that did not fit any of the previous criteria was reported as benign if it was inherited from a healthy parent. Finally, the remaining CNVs were reported as being of unclear clinical significance. Although a de novo CNV has more probability of being pathogenic than an inherited CNV, inheritance status of a CNV was not a criterion in itself for interpretation, but was evaluated on a case-by-case basis, depending on the genomic region involved.

3.3.3 Validation of CNVs

The inherited or *de novo* status of the CNVs of interest was established by performing aCGH or FISH studies in the parents. When possible, *de novo* CNVs found in amniotic fluid (AF) and CVS were confirmed by FISH analysis with labeled bacterial artificial chromosomes (BAC) clones (Centre for Applied Genomics, Hospital for Sick Children, Toronto, ON, Canada) selected from the RP11 library with the University of California Santa Cruz Genome Browser, based on the Human March 2006 Assembly (hg18).



*Database of Genomic Variants, Genoglyphix Chromosome Aberration Database or internal laboratory database

Figure 3-1 : Interpretation algorithm used to establish clinical significance of the CNVs detected by aCGH in our cohort.

Each CNV was visualized in Genoglyphix®. If the CNV was previously reported in healthy individuals (in at least two different studies in DGV, or at least two different individuals in GCAD or our internal database), the CNV was reported as benign. If the detected CNV was larger than 5 Mb in size, corresponded to a genomic region previously associated with a known syndrome, or if the genes within the genomic region had a clear functional link with the fetus’s phenotype, we reported it as pathogenic, unless one or both parents could not be tested, in which case it was deemed likely pathogenic but reported as a CNV of unclear clinical significance. If the CNV was smaller than 1 Mb, and inherited from a healthy parent, we reported it as benign. Otherwise, it was reported as a CNV of unclear clinical significance.

3.4 Results

Between July 2008 and November 2009, a total of 49 samples from fetuses with single (n=25) or multiple (n=24) malformation(s) were submitted for aCGH. Most cases (35/49; 71.4%) show normal results, either with a benign CNV previously found in healthy individuals (30/49; 61.2%), or without any copy number changes (5/49; 10.2%). Fifteen fetuses had copy number changes of defined (4/49; 8.2%), or unclear clinical significance (11/49; 22.4%). Three of the fifteen fetuses (17, 22 & 44) had more than one copy number change, for a total of 19 alterations with potential clinical significance. Following parental studies, nine CNVs in six fetuses were reclassified as benign, either because they were inherited from a healthy parent (fetuses 17, 22, 25, 40 & 44), or because the only involved gene is in a known deletion hotspot in the normal French-Canadian population (fetus 34) (32). The ten remaining CNVs were either *de novo* (fetuses 18, 20, 47 & 53), inherited (fetuses 36 & 48) or of undetermined inheritance because one or both parents were unavailable for testing (fetuses 12, 23 & 51). Once all analyses were completed, four fetuses (4/49; 8.2%) had a pathogenic CNV, and six (6/49; 12.2%) had a CNV that remained of unclear clinical significance, although two of them were considered likely pathogenic (fetuses 12 and 51). Phenotype information and aCGH results for these fetuses are summarized in the next sections and detailed in Table 3-2. Supplementary clinical information and discussion are also available as Supporting Information (section 0).

Table 3-2 : Details for fetuses with an aCGH result of defined or unclear clinical significance.

Fetus	Ultrasound	Autopsy / postnatal examination	Specimen	Karyotype	Array	aCGH result	Inheritance	Size & type	Genes	Interpretation
12	Diaphragmatic hernia (right), anomalous pulmonary venous return (right)	Lung hypoplasia, metatarsus eversus, multiple contractures, joints anomalies, bilateral camptodactyly, dysmorphic facies (coarse face, short neck, epicanthus inversus, deep philtrum, depressed nasal bridge, anteverted nares, downturned corners of mouth, and micrognathia)	FFT	46,XY	OS 2.0	arr 3p21.31 (48,325,068-48,983,044)x3	Unknown (parents not tested)	658 kb gain	SPINK8, FBXW12, PLXNB1, CCDC51, CCDC72, ATRIP, TREX1, SHISA5, PFKFB4, UCN2, COL7A1, UQCRC1, TMEM89, SLC26A6, CELSR3, NCKIPSD, IP6K2, PRKAR2A, SLC25A20, C3ORF71, ARIH2	Unclear clinical significance ^a
18	Cystic hygroma, increased nuchal and cervical soft tissues measurements, possible omphalocele, thin thorax, fetal death	Cleft palate, gastroschisis, aortic isthmus hypoplasia, arthrogryposis (inferior limbs)	CVS	46,XY	OS 1.1	arr 15q14 (33,340,370-32,628,739)x1	De novo	288 kb loss	SLC12A6, NOP10, LPCAT4, C15orf55, GOLGA8B	Unclear clinical significance
20	Alobar holoprosencephaly, facial abnormalities (cyclopia, proboscis, arhinia)	Single cerebral ventricular cavity, complete corpus callosum agenesis	FFT	46,XY	OS 1.1	arr 2p21 (44,949,357-45,184,052)x1	De novo	235 kb loss	SIX3, SIX2	Pathogenic
22	Cerebral ventriculomegaly, microcephaly, severe neuronal migration abnormalities, possible lissencephaly, SUA	Visceromegaly (liver, lungs, kidneys), thin corpus callosum	CB	46,XY	OS 2.0	arr 6q23.3q24.1 (138,848,815-139,632,692)x3	De novo	784 kb gain	NHSL1, CCDC28A, ECT2L, REPS1, C6orf115, HECA, TXLN8	Unclear clinical significance
23	Hydrocephalus with stenosis of the aqueduct of Sylvius	Clubfeet, persistent left superior vena cava, mullerian duct abnormalities, dysmorphic facies, SUA	FFT	46,XX	OS 1.1	arr 5p15.33 (132,145-480,969)x3	Unknown (parents not tested)	349 kb gain	Cri-du-Chat region, PLEKHG4B, LRRC14B, CCDC127, SDHA, PDCD6, AHRR	Unclear clinical significance
36	Increased nuchal fold measurement, short nasal bone, persistent superior vena cava draining to the coronary sinus, SUA	Birth at 33 weeks; healthy	CVS	46,XX	OS 2.0	arr 7p14.3p14.2 (32,874,023-36,540,594)x1	Maternal	3.67 Mb loss	KBTBD2, RP9P, FKBP9, NT5C3, RP9, BBS9, BMPER, NPSR1, AAA1, DPY19L1, DPY19L2P1, TBX20, HERPUD2, SEPT7, EEPD1, KIAA0895, ANLN, AOA1	Unclear clinical significance

Fetus	Ultrasound	Autopsy / postnatal examination	Specimen	Karyotype	Array	aCGH result	Inheritance	Size & type	Genes	Interpretation
47	Bilateral ventriculomegaly (mild), mega cisterna magna, cerebellum and vermis atrophy, upward rotation of the vermis, aortic stenosis, bicuspid aortic valve	Hepatomegaly	FFT	46,XY	CGX-12	arr 3q23q25.32 (143,856,784-158,118,852)x1	De novo	14.26 Mb loss	69 genes, including the Dandy-Walker malformation region	Pathogenic
48	Multicystic dysplastic kidney (right), median ureterocele, clubfeet	Birth at 28 weeks; patent ductus arteriosus, mild atrial septal defect, megaureter (right), hydroureter, hydronephrosis	AF	46,XX	CGX-12	arr 22q11.21 (19,064,244-19,790,658)x1	Maternal	726 kb loss	DG/VCF region, ZNF74, SCARF2, KLHL22, MED15, POM121L4P, TMEM191A, PI4KA, SERPIND1, SNAP29, CRKL, AIFM3, LZTR1, THAP7, FLJ39582, MGC16703, P2RX6, SLC7A4, P2RX6P, LOC400891	Pathogenic
51	Ventricular septal defect, two superior venae cava, pelvic kidney (right), clubfoot (right), SUA	Birth at 40 weeks	AF	46,XX	CGX-12	arr 10q24.32 (103,201,726-103,433,500)x3	Unknown (father not tested)	232 kb gain	SHFM3 region, BTRC, POLL, DPCD, FBXW4	Unclear clinical significance ^a
53	Hyperechogenic kidneys, bilateral pyelocaliectasis, vesical distension, urethral cyst, oligohydramnios	Macrosomia, Potter's facies, bilateral hydroureter, obstructive multicystic dysplastic kidneys, dilated sigmoid, mega rectum, imperforated anus	AF	46,XY	CGX-12	arr 10q26.13q26.3 (124,590,071-132,729,781)x1	De novo	8.14 Mb loss	42 genes	Pathogenic

AF, amniotic fluid; CB, cord blood; CGX-12, NimbleGen CGX-12; CVS, chorionic villus sampling; DG/VCF, DiGeorge/Velocardiofacial syndrome; FFT, frozen fetal tissue; IUGR, intrauterine growth retardation; OS 1.1, SignatureChip Oligo Solution v1.1 Rev B; OS 2.0, SignatureChip Oligo Solution v2.0 12-plex; SHFM3, Split-hand/foot malformation 3; SUA, single umbilical artery.

^aLikely pathogenic

3.4.1 Clinically significant results

Fetus 20 showed alobar holoprosencephaly and facial abnormalities, which were initially documented by prenatal ultrasound. Array CGH revealed a *de novo* 235 kb deletion of 2p21, encompassing SIX2 (MIM 604994) and SIX3 (MIM 603714) (Figure 3-2). This CNV was clearly responsible for the fetus's phenotype, as microdeletions encompassing *SIX3* are known to cause holoprosencephaly-2 (33-36).

Fetus 47 showed cerebral abnormalities, and aortic stenosis with bicuspid aortic valve at 18 weeks of gestation. Array CGH showed a *de novo* 14.26 Mb deletion ranging from 3q23 to 3q25.32. Careful review of the initial karyotype revealed that the imbalance was visible, but went unnoticed because of low resolution. The deletion found in this fetus was classified as pathogenic because of its size.

Fetus 48 showed severe clubfeet, multicystic dysplastic kidney, and ureterocele. A baby girl was prematurely born at 28 weeks. She showed respiratory distress, patent ductus arteriosus, mild atrial septal defect and megaureter, and later developed hydroureter and hydronephrosis. The only dysmorphic feature noted was microstomia. Array CGH showed a 726 kb deletion of 22q11.21, within the distal part of the classical DiGeorge/Velocardiofacial (DG/VCF; MIM 188400 & 192430) syndrome region. This alteration was inherited from the healthy mother. It was deemed pathogenic because the patient's phenotype is consistent with published cases, which are known to exhibit variable expressivity and reduced penetrance (37, 38).

Fetus 53 showed severe urological abnormalities and oligohydramnios. Autopsy revealed macrosomia, dysmorphic facies associated with Potter's sequence, and anal abnormalities. A *de novo* 8.14 Mb deletion of 10q26.13q26.3 was found. This CNV was classified as pathogenic because of its size. In addition, the fetus's phenotype is consistent with the features associated with 10q monosomy (39-41).

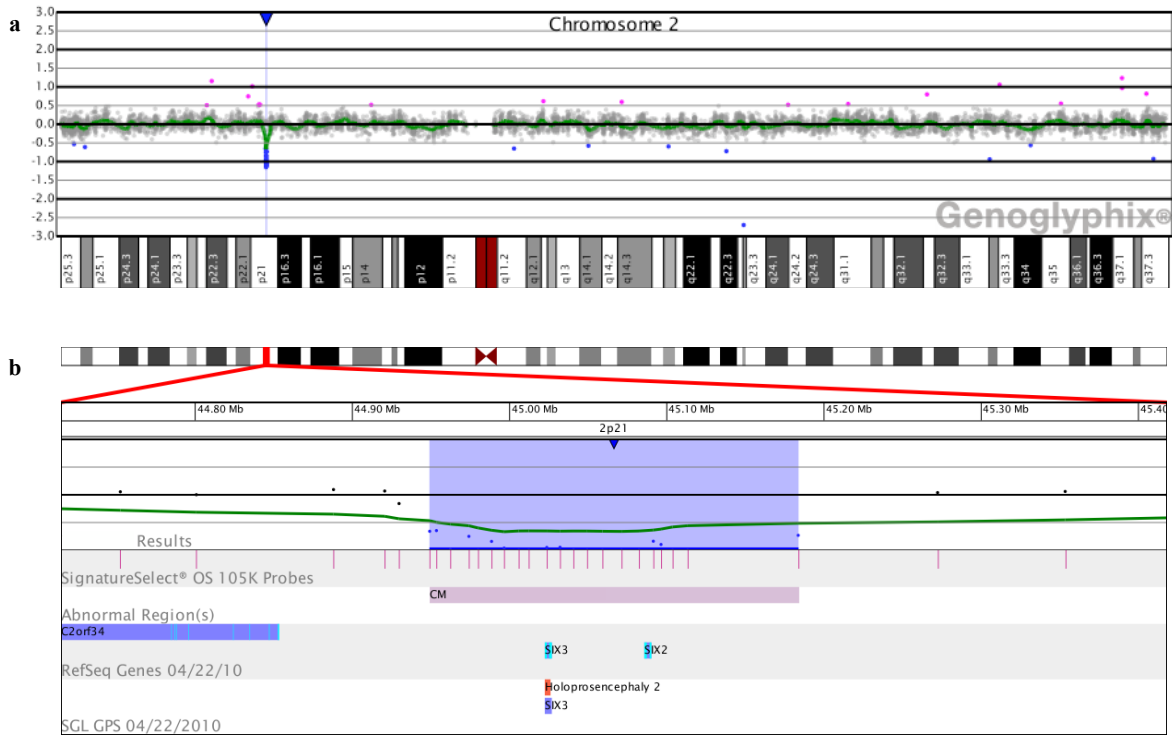


Figure 3-2 : Array CGH result for fetus 20.

Each dot represents one oligonucleotide and its position on the chromosome. (a) Log_2 ratio of means for oligonucleotides covering chromosome 2. A copy loss is clearly seen in band p21. (b) Close-up of the genomic region with the deleted segment. The 235 kb deletion is covered by 21 oligonucleotides (blue dots), and contains 2 genes: *SIX3* and *SIX2*. Point mutations and haploinsufficiency of *SIX3* have been associated with Holoprosencephaly 2 (MIM #157170).

3.4.2 Results of unclear clinical significance

Fetus 12 showed diaphragmatic hernia and anomalous pulmonary venous return at 23 weeks of gestation. Autopsy further revealed multiple contractures, joints anomalies, bilateral camptodactyly, metatarsus eversus, and dysmorphic facies. FISH excluded Pallister-Killian syndrome and array CGH detected a 658 kb duplication of 3p21.31. Although parents could not be tested and no clear link could be established with the diaphragmatic hernia, this CNV was deemed likely pathogenic because the dysmorphic features present in the fetus are associated with trisomy 3p syndrome (42-46).

Fetus 18 was tested because of cystic hygroma, possible omphalocele, thin thorax, and fetal death. Autopsy later revealed a cleft palate, gastroschisis, aortic isthmus hypoplasia, and arthrogryposis. Array CGH performed on CVS showed a *de novo* 288 kb deletion of 15q14, spanning 4 genes: *NOP10* (MIM 606471), *C15orf55* (MIM 608963), *LPCAT4* (MIM 612039), and *GOLGA8A*. In addition, both breakpoints probably interrupt a gene: *SLC12A6* (MIM 604878) proximally, and *GOLGA8B* (MIM 609619) distally. However, the closest probe with a normal signal is outside the gene boundaries, raising the possibility that either gene is completely deleted. FISH analysis with the RP11-1141P10 BAC clone on cultured CVS and amniocytes did not confirm the deletion. Although this BAC covers a 156 kb sequence that is rich in segmental duplications, and is known to harbor benign CNVs, none of them has the same boundaries as the one found in our patient. Therefore, either of the two segments adjacent to the one covered by the BAC could be deleted and cause a clinically significant genomic imbalance. The proximal segment contains part of *GOLGA8B*. The distal segment contains part of *SLC12A6*, *NOP10*, *C15orf55*, and *LPCAT4*. It was not possible to establish if this CNV is responsible for the malformations observed, thus it remained of unclear clinical significance.

Fetus 22 was characterized by severe brain malformations (abnormal gyration and microcephaly) discovered on magnetic resonance imaging at 34 weeks of gestation. Autopsy further suggested histological changes consistent with lissencephaly. A *de novo* 784 kb duplication of 6q23.3q24.1, spanning 7 genes, was detected. One breakpoint interrupts *NHSLI*, which encodes a member of the NHS proteins family that is expressed in the brain and in other organs during fetal development (47, 48). Members of this family of proteins

contain a domain homologous to the actin assembly regulation domain of the WAVE proteins, and the ectopic expression of one of them (NHS) inhibits lamellipod formation, which is essential for neuronal migration. Considering the *de novo* nature of this duplication and the functional link between NHSL1 and the fetus's phenotype, this CNV was deemed potentially pathogenic. However, it was reported as a CNV of unclear clinical significance, due to the lack of evidence supporting its actual effect on the patient's phenotype.

Fetus 23 initially showed massive hydrocephalus with stenosis of the aqueduct of Sylvius, on an ultrasound done at 20 weeks of gestation. Autopsy further revealed persistent left superior vena cava, mullerian duct abnormalities, clubfeet, and dysmorphic facies. Array CGH revealed a 349 kb duplication of 5p15.33, within the distal part of the Cri-du-Chat syndrome region (MIM 123450). The duplicated segment encompasses 5 genes: *PLEKHG4B*, *LRR14B*, *CCDC127*, *SDHA* (MIM 600857), *PDCD6* (MIM 601057); and interrupts one gene: *AHRR* (MIM 606517). Parents could not be tested. Although this region is rich in segmental duplications and harbors many CNVs found in healthy individuals, these vary in size and none corresponds to the CNV found in our patient. In addition, our fetus's phenotype is strikingly similar to the phenotype associated with trisomy 5p (49-54), and we cannot exclude that this duplication is responsible for the observed malformations. Thus, until further evidence is available, this CNV remains of uncertain clinical significance.

Fetus 36 was referred for increased nuchal fold measurement (7.9 mm) at 12 weeks. Ultrasound later showed a short nasal bone, microcephaly, persistent superior vena cava, and a single umbilical artery. A baby girl was prematurely delivered at 33 weeks, and was otherwise healthy. Array CGH showed a maternally inherited 3.67 Mb deletion of 7p14.3p14.2, which was confirmed by FISH with the RP11-23B6 BAC clone on both the mother's and her daughter's blood samples after birth. The deleted region encompasses 17 genes, including *BBS9* (MIM 607968), *RP9* (MIM 607331), *BMPER* (MIM 608699), *TBX20* (MIM 606061), and *SEPT7* (MIM 603151). A one year-old female with a 5.36 Mb deletion completely overlapping the deletion in our patient was reported in DECIPHER (Patient 248400). She had microcephaly, hypotonia, low birthweight, atrial septal defect, cleft palate & uvula, and clinodactyly. As in our case, the deletion reported in this patient was inherited from a healthy parent. Therefore, this CNV, if pathogenic, may exhibit variable expressivity or reduced

penetrance. This alteration was reported as being of unclear clinical significance, because of its size, its gene content, and the phenotype observed in individuals with a deletion overlapping the one found in our patient.

Fetus 51 was tested after prenatal ultrasound showed clubfoot, two superior venae cava, a pelvic kidney, and a single umbilical artery. Ventricular septal defect was also suspected and confirmed at birth. Array CGH showed a 232 kb duplication of 10q24.32, partially overlapping the Split-hand/foot malformation 3 (SHFM3, MIM 600095) region, which is duplicated in patients with SHFM3. The mother did not carry the duplication, and the father was unavailable for testing. Although our patient does not present the distinctive features of SHFM3, SHFM3 has been known to exhibit variable expressivity and reduced penetrance (55, 56), and cardiac and renal abnormalities have been reported in the syndromic form of SHFM3 (57, 58). Therefore, this CNV was deemed likely pathogenic.

3.4.3 Novel benign CNVs

Six patients had at least one alteration that was smaller than 1 Mb and could not be found in public variants databases. They were initially classified as of unclear clinical significance, but were reassigned to the benign category following further evaluation of the genomic region involved, or after parental testing showed they were inherited from a healthy parent (Supporting Information, section 3.8.2.3).

3.5 Discussion

We report our experience with the use of whole-genome aCGH for the analysis of 49 fetuses with unexplained major abnormalities detected on ultrasound. Array CGH successfully identified a pathogenic chromosomal alteration in four fetuses (4/49; 8.2%). Six additional fetuses (6/49; 12.2%) showed CNVs of unclear clinical significance. In comparable studies (21-23, 25-29), pathogenic CNVs were found in 0% to 12.5% of cases, while unclear results were reported in 0% to 12.3% of cases (Table 3-3).

Table 3-3 : Summary of previous and present prenatal studies investigating fetuses with an abnormal ultrasound.

Study	Retrospective or prospective cases	Karyotype	Inclusion criteria	Array(s)	Type of CNV reported	
					Pathogenic	Unclear significance
Le Caignec 2005	Retrospective: after TOP (n=41) or IFD (n=8)	Normal	≥ 3 structural malformations (n=49)	Targeted: GenoSensor Array 300: 287 clones (PI, BAC, PAC)	8.2% (4/49)	2% (1/49)
Shaffer 2008	Both: cases referred from external hospital centers & clinics	Normal	Abnormal US (n=110)	Targeted: Signature Prenatal chip (BAC array, ø details provided)	1.8% (2/110)	0-0.6% (1/151) ^a
Van der Veyver 2008	Both: cases referred from external hospital centers & clinics	Ongoing	Abnormal US (n=84)	Targeted: BAC array BCM v5 (853 clones) or v6 (1476 clones), BCM Oligo Chromosomal Microarray v6	0% (0/81) ^b	1.2% (1/81) ^b
Kleeman 2009	Prospective: ongoing pregnancies	Normal	Significant structural malformation and/or IUGR (excl. only soft markers) (n=50)	Targeted: SignatureChip 4.0 (n=26) 1887 clones; Whole genome: SignatureWG chip (n=24) 4685 clones (average gap 1.6 Mb) Resolution in targeted region (both chips): 80-100 kb	2% (1/50)	2% (1/50)
Coppinger 2009	Both: cases referred from external hospital centers & clinics	Normal or ongoing	Abnormal US ø details (Whole Genome: n=155, Targeted: n=33)	Targeted: SignatureChip v4.0 or Prenatal chip v1.0/v2.0 (n=33) 2100 BACs; Whole genome: SignatureChip WG v1.0 (n=137) 4670 BACs, SignatureChipOS v1.0 (n=18) 105K oligos	2.2% (4/186) ^b	0.5% (1/186) ^b
Tyreman 2009	Retrospective: randomly selected anonymised samples	Normal	Major US abnormality or multiple soft markers by US (n=106)	Whole genome: Affymetrix GeneChip 6.0 SNP array	10.4% (11/106) ^c	12.3% (13/106) ^c
Vialard 2009	Retrospective: after TOP	Normal (n=37)	≥ 2 malformations (n=39)	Targeted: GenoSensor Array 300: 287 clones (PI, BAC, PAC)	8.1% (3/37)	2.7% (1/37)
Valduga 2010	Retrospective: after TOP	Normal	≥ 3 significant malformations (n=42) or severe brain abnormality (n=8)	Whole genome: Agilent 44K oligo array (n=49): 43 kb median spacing, 24 kb in RefSeq genes; Agilent 244K (fetus 31)	10% (5/50)	0% (0/50)
Faas 2010	Retrospective: after TOP (n=20), IFD (n=4) or live birth (n=8)	Normal (n=32)	Significant structural malformation	Whole genome: Affymetrix GeneChip 250K Nspl SNP array	12.5% (4/32)	3.1% (1/32)
Our study	Both: after TOP/ IFD (n=28) & ongoing pregnancies (n=21)	Normal	≥ 1 major malformation diagnosed by US	Whole Genome: SignatureChip WG v1.0.1 (n=7) 4685 BACs; SignatureChipOS v1.1 rev B (n=16) 105K oligos; SignatureChipOS v2.0 12-plex (n=8) or NimbleGen CGX (n=18): 135K oligos	8.2% (4/49)	12.2% (6/49)

TOP, termination of pregnancy; IFD, intrauterine fetal death; IUGR, intrauterine growth retardation; US, ultrasound.

^a Including cases without US abnormalities (presence or absence of malformations for the fetus with an unclear result was not specified).

^b Excluding cases with an abnormal karyotype.

^c No parental analysis was performed.

3.5.1 Diagnostic yield

The differences in the proportion of significant and unclear results reported in the various cohorts are very likely due to the design of the arrays used, the unavailability of parental samples, and the varying severity of the malformations observed in the fetuses studied. In our cohort, the low-density array (SignatureChipWG) gave a normal result for all fetuses analyzed with this array ($n=7$), while the denser oligo-based array (SignatureChipOS v2.0™ / NimbleGen CGX) gave a clinically significant result in 11.5% of the cases (3/26). Although not statistically significant due to the small sample size (Exact $p = 0.583$), these numbers are consistent with the expectation that denser arrays will detect a greater proportion of pathogenic CNVs (15). However, increasing array resolution may also increase the number of unclear results (20), as observed in a recent study performed with a high-density array, which reported a CNV of unclear clinical significance in 12.3% of cases (28). Another factor probably contributing to the higher proportion of unclear results in the study by Tyreman *et al.* is the fact that parental DNA was not available. Obtaining a parental sample is not always possible in the real world, as our experience demonstrates. In half the cases with a result of unclear clinical significance, we were unable to analyse both parents. This may also have contributed to increase our rate of unclear results. Finally, the proportion of pathogenic CNVs reported in previous studies is much higher in retrospective cohorts investigating fetuses following termination of pregnancy (Le Caignec *et al.* 2005 (25), Vialard *et al.* 2009 (21), Valduga *et al.* 2010 (27); 8.2%, 8.1% and 10% respectively) than in prospective cohorts (Shaffer *et al.* 2008 (20), Van den Veyver *et al.* 2009 (22), Coppinger *et al.* 2009 (23); 1.8%, 0% and 2.2% respectively). The prospective cohorts might have included fetuses with milder abnormalities, which might in turn have a lower risk for chromosomal abnormalities. Although three other studies focusing on fetuses with major malformations reported variable proportions of pathogenic CNVs (Kleeman *et al.* 2009, Tyreman *et al.* 2009, Faas *et al.* 2010; 2%, 10.4% and 12.5% respectively), these differences can be attributed to the density of the arrays used, rather than the characteristics of the fetuses studied. In the present study, our cohort only included fetuses with major malformations diagnosed on ultrasound, of which a large proportion was tested retrospectively. Indeed, there was no difference in the diagnostic yield between fetuses with isolated or multiple malformations (8% vs 8.3%, Exact $p = 1.000$;

Table 3-S1). Also, the proportion of pathogenic CNVs was slightly higher (but not statistically significant) in the retrospective group compared to the group tested prospectively (10.7% vs 4.8%, Exact $p = 0.625$; Table 3-S1). In addition, most fetuses ($n=42$) were tested with a dense whole-genome oligo-based array. These factors probably explain the proportion of pathogenic CNVs reported in the present study, which is comparable to the rates found in previous retrospective studies, and confirm the value of whole-genome aCGH in the investigation of unexplained fetal malformations.

3.5.2 Size of detected chromosomal aberrations

In contrast to conventional karyotyping in blood samples, resolution is limited for prenatal samples, and some significantly large aberrations can be overlooked (fetuses 47 and 53). In these situations, aCGH can detect the aberration, unless it is balanced (translocation or inversion), allowing a reduction of false negative results. However, the main benefit of aCGH remains the detection of submicroscopic chromosomal aberrations with a clinical impact. Fetus 36 had a large aberration, although not detectable by conventional cytogenetics. Also, the pathogenic 235 kb deletion spanning the *SIX3* gene in fetus 20 would not have been identified by any other method, despite the specific phenotype that could have oriented the investigation. This is a clear illustration of the potential benefits of aCGH in the prenatal setting.

3.5.3 Interpretation challenges

Genotype-phenotype correlation is not always straightforward, as illustrated by the six CNVs of unclear clinical significance. In these cases, testing parents can be helpful, as *de novo* alterations have a higher probability of being pathogenic than inherited CNVs. In addition, establishing the inheritance status of chromosomal abnormalities allows more specific genetic counseling, especially if an abnormality is known to display reduced penetrance or variable expressivity (fetus 48), or is subject to parental imprinting. However, the deletion found in fetus 18 highlights how the *de novo* nature of an alteration, associated with a possible link

between the patient's phenotype and the function of the involved genes, is not sufficient to establish clinical significance. In this case, the presence of polymorphic segmental duplications in the CNV weakened the probability of its potential pathogenicity, but because causality could not be excluded, this CNV remains of uncertain significance. Another example is the 3.67 Mb deletion found in fetus 36. Despite its size and its gene content, this alteration remains of unclear significance, as it was inherited from a healthy parent. These specific cases clearly demonstrate that currently, some CNVs found by aCGH remain impossible to interpret, even when parental status is known, and highlight the importance of expanding and sharing genotype-phenotype CNV databases.

3.5.4 Implementation in the prenatal setting

It has been proposed that aCGH should be offered universally in the prenatal setting, instead of restricting its use to testing fetuses with ultrasound anomalies (20, 22, 23). While this may be an acceptable approach with targeted arrays, our results speak against such a testing paradigm. We believe that whole-genome aCGH should be reserved for fetuses presenting major malformations and a concomitant higher probability of detectable chromosomal imbalance, because of the uncertainties associated with CNVs of unclear clinical significance. Previous studies failed to demonstrate the benefit of aCGH, as far as the capacity to detect additional chromosomal aberrations in the general pregnant women population, since they did not detect any chromosomal abnormalities that would not have been detected by other methods (karyotype or ultrasound) (20, 22). Genetic counseling challenges associated with unclear aCGH results require that decisions regarding ongoing pregnancies be based on ultrasound findings, which have proven to better predict neonatal outcome (59, 60). As our knowledge of benign and pathogenic CNVs increases, and as databases (both private and public) become more complete, we might be able to expand whole-genome aCGH to fetuses without major abnormalities.

In conclusion, whole-genome aCGH is useful for the investigation of fetal malformations, as it identifies pathogenic chromosomal abnormalities that would otherwise go undetected, and thus improves diagnostic yield allowing better genetic counseling for families.

However, the clinical interpretation difficulties associated with unclear results demand that we proceed with caution during its implementation as a prenatal diagnostic tool.

3.6 Acknowledgements

The authors would like to thank parents who participated in this study. They also would like to thank the clinical geneticists and the cytogeneticists from CHU Sainte-Justine for their collaboration, as well as Joanie Gingras for her help with statistical analysis, and Mark E. Samuels for his review of the manuscript. G.D. is supported by a NSERC Alexander Graham Bell Canada Graduate Scholarship, and a FRSQ Master's training award.

3.7 References

1. Kalter H, Warkany J. Medical progress. Congenital malformations: etiologic factors and their role in prevention (first of two parts). *N Engl J Med* 1983; 308: 424-431.
2. Dallaire L, Michaud J, Melancon SB et al. Prenatal diagnosis of fetal anomalies during the second trimester of pregnancy: their characterization and delineation of defects in pregnancies at risk. *Prenat Diagn* 1991; 11: 629-635.
3. Benn PA, Egan JFX, Fang M et al. Changes in the utilization of prenatal diagnosis. *Obstet Gynecol* 2004; 103: 1255-1260.
4. Wilson RD, Chitayat D, McGillivray BC. Fetal ultrasound abnormalities: correlation with fetal karyotype, autopsy findings, and postnatal outcome--five-year prospective study. *Am J Med Genet* 1992; 44: 586-590.
5. Rizzo N, Pittalis MC, Pilu G et al. Distribution of abnormal karyotypes among malformed fetuses detected by ultrasound throughout gestation. *Prenat Diagn* 1996; 16: 159-163.
6. Tseng J-J, Chou M-M, Lo F-C et al. Detection of chromosome aberrations in the second trimester using genetic amniocentesis: experience during 1995-2004. *Taiwanese journal of obstetrics & gynecology* 2006; 45: 39-41.
7. Ledbetter DH, Martin CL. Cryptic telomere imbalance: a 15-year update. *Am J Med Genet Part C Semin Med Gen* 2007; 145C: 327-334.
8. Gignac J, Danis K, Tihy F et al. Prenatal detection of subtelomeric rearrangements by multi-subtelomere FISH in a cohort of fetuses with major malformations. *Am J Med Genet* 2006; 140: 2768-2775.
9. Rooms L, Reyniers E, Wuyts W et al. Multiplex ligation-dependent probe amplification to detect subtelomeric rearrangements in routine diagnostics. *Clin Genet* 2006; 69: 58-64.
10. Saugier-veber P, Goldenberg A, Drouin-Garraud V et al. Simple detection of genomic microdeletions and microduplications using QMPSF in patients with idiopathic mental retardation. *Eur J Hum Genet* 2006; 14: 1009-1017.
11. Gouas L, Goumy C, Véronèse L et al. Gene dosage methods as diagnostic tools for the identification of chromosome abnormalities. *Pathol Biol* 2008; 56: 345-353.

12. Goumy C, Gouas L, Pebrel-Richard C et al. Prenatal detection of cryptic rearrangements by multiplex ligation probe amplification in fetuses with ultrasound abnormalities. *Genet Med* 2010; 12: 376.
13. Knight SJ, Regan R, Nicod A et al. Subtle chromosomal rearrangements in children with unexplained mental retardation. *Lancet* 1999; 354: 1676-1681.
14. Moore KL, Persaud TVN, Torchia MG. *The Developing Human : Clinically Oriented Embryology*. Philadelphia, PA: Saunders/Elsevier, 2008.
15. Sagoo GS, Butterworth AS, Sanderson S et al. Array CGH in patients with learning disability (mental retardation) and congenital anomalies: updated systematic review and meta-analysis of 19 studies and 13,926 subjects. *Genet Med* 2009; 11: 139-146.
16. Miller DT, Adam MP, Aradhya S et al. Consensus statement: chromosomal microarray is a first-tier clinical diagnostic test for individuals with developmental disabilities or congenital anomalies. *Am J Hum Genet* 2010; 86: 749-764.
17. Schaeffer AJ, Chung J, Heretis K et al. Comparative genomic hybridization-array analysis enhances the detection of aneuploidies and submicroscopic imbalances in spontaneous miscarriages. *Am J Hum Genet* 2004; 74: 1168-1174.
18. Rickman L, Fiegler H, Shaw-Smith C et al. Prenatal detection of unbalanced chromosomal rearrangements by array CGH. *J Med Genet* 2006; 43: 353-361.
19. Sahoo T, Cheung SW, Ward P et al. Prenatal diagnosis of chromosomal abnormalities using array-based comparative genomic hybridization. *Genet Med* 2006; 8: 719-727.
20. Shaffer LG, Coppinger J, Alliman S et al. Comparison of microarray-based detection rates for cytogenetic abnormalities in prenatal and neonatal specimens. *Prenat Diagn* 2008; 28: 789-795.
21. Vialard F, Molina Gomes D, Leroy B et al. Array comparative genomic hybridization in prenatal diagnosis: another experience. *Fetal Diagn Ther* 2009; 25: 277-284.
22. Van den Veyver IB, Patel A, Shaw CA et al. Clinical use of array comparative genomic hybridization (aCGH) for prenatal diagnosis in 300 cases. *Prenat Diagn* 2009; 29: 29-39.
23. Coppinger J, Alliman S, Lamb AN et al. Whole-genome microarray analysis in prenatal specimens identifies clinically significant chromosome alterations without increase in

- results of unclear significance compared to targeted microarray. *Prenat Diagn* 2009; 29: 1156-1166.
24. Hillman SC, Pretlove S, Coomarasamy A et al. Additional information from array comparative genomic hybridization technology over conventional karyotyping in prenatal diagnosis: a systematic review and meta-analysis. *Ultrasound Obstet Gynecol* 2010; 37: 6-14.
 25. Le Caignec C, Boceno M, Saugier-veber P et al. Detection of genomic imbalances by array based comparative genomic hybridisation in fetuses with multiple malformations. *J Med Genet* 2005; 42: 121-128.
 26. Kleman L, Bianchi DW, Shaffer LG et al. Use of array comparative genomic hybridization for prenatal diagnosis of fetuses with sonographic anomalies and normal metaphase karyotype. *Prenat Diagn* 2009; 29: 1213-1217.
 27. Valduga M, Philippe C, Bach Segura P et al. A retrospective study by oligonucleotide array-CGH analysis in 50 fetuses with multiple malformations. *Prenat Diagn* 2010; 30: 333-341.
 28. Tyreman M, Abbott KM, Willatt LR et al. High resolution array analysis: diagnosing pregnancies with abnormal ultrasound findings. *J Med Genet* 2009; 46: 531-541.
 29. Faas BHW, van der Burgt I, Kooper AJA et al. Identification of clinically significant, submicroscopic chromosome alterations and UPD in fetuses with ultrasound anomalies using genome-wide 250k SNP array analysis. *J Med Genet* 2010; 47: 586-594.
 30. Duker AL, Ballif BC, Bawle EV et al. Paternally inherited microdeletion at 15q11.2 confirms a significant role for the SNORD116 C/D box snoRNA cluster in Prader-Willi syndrome. *Eur J Hum Genet* 2010; 18: 1196-1201.
 31. Neill NJ, Torchia BS, Bejjani BA et al. Comparative analysis of copy number detection by whole-genome BAC and oligonucleotide array CGH. *Mol Cytogenet* 2010; 3: 11.
 32. Bradley WEC, Raelson JV, Dubois DY et al. Hotspots of large rare deletions in the human genome. *PLoS ONE* 2010; 5: e9401.
 33. Bendavid C, Dubourg C, Gicquel I et al. Molecular evaluation of foetuses with holoprosencephaly shows high incidence of microdeletions in the HPE genes. *Hum Genet* 2006; 119: 1-8.

34. Geng X, Speirs C, Lagutin O et al. Haploinsufficiency of Six3 fails to activate Sonic hedgehog expression in the ventral forebrain and causes holoprosencephaly. *Dev Cell* 2008; 15: 236-247.
35. Bendavid C, Rochard L, Dubourg C et al. Array-CGH analysis indicates a high prevalence of genomic rearrangements in holoprosencephaly: an updated map of candidate loci. *Hum Mutat* 2009; 30: 1175-1182.
36. Paulussen AD, Schrandt-Stumpel CT, Tserpelis DCJ et al. The unfolding clinical spectrum of holoprosencephaly due to mutations in SHH, ZIC2, SIX3 and TGIF genes. *Eur J Hum Genet* 2010; 18: 999-1005.
37. Garcia-Miñaur S, Fantes J, Murray RS et al. A novel atypical 22q11.2 distal deletion in father and son. *J Med Genet* 2002; 39: e62.
38. Rauch A, Zink S, Zweier C et al. Systematic assessment of atypical deletions reveals genotype-phenotype correlation in 22q11.2. *J Med Genet* 2005; 42: 871-876.
39. Gorinati M, Zamboni G, Padoin N et al. Terminal deletion of the long arm of chromosome 10: case report and review of the literature. *Am J Med Genet* 1989; 33: 502-504.
40. Leonard NJ, Harley FL, Lin CC. Terminal deletion of chromosome 10q at band 26.1: follow-up in an adolescent male with high-output renal failure from congenital obstructive uropathy. *Am J Med Genet* 1999; 86: 115-117.
41. Motoyama O, Tokuyama M, Uga N et al. Urinary retention in a boy with terminal deletion of chromosome 10q at band 26.1. *Pediatr Int* 2008; 50: 603-604.
42. Rethoré MO, Lejeune J, Carpentier S et al. [Trisomy for the distal part of the short arm of the number 3 chromosome in 3 siblings. First example of chromosomal insertion: INS(7;3)(q 31;p 21 p 26)]. *Ann Genet* 1972; 15: 159-165.
43. Centerwall W, Mankinen C, Marsh W et al. The 3p trisomy syndrome: de novo case identified via banded prometaphase chromosomes. In: Bloom AD, James LS, Paul NW et al., eds. *The Fetus and the Newborn, Birth Defects original article series, Vol. 17*. New York: A.R. Liss, 1981: 77-78.
44. Kleczkowska A, Fryns JP, van den Berghe H. Partial trisomy of chromosome 3(p14----p22) due to maternal insertional translocation. *Ann Genet* 1984; 27: 180-183.
45. Orye E, Laureys G. Trisomy 3p syndrome. Report of a new case, due to a chromosomal insertion. *Ann Genet* 1984; 27: 245-247.

46. Zhang SZ, Wang Q. Partial serial duplication of the short arm of chromosome 3. *Chin Med J* 1984; 97: 425-428.
47. Brooks SP, Ebenezer ND, Poopalasundaram S et al. Identification of the gene for Nance-Horan syndrome (NHS). *J Med Genet* 2004; 41: 768-771.
48. Brooks SP, Coccia M, Tang HR et al. The Nance-Horan Syndrome protein encodes a functional WAVE Homology Domain (WHD) and is important for co-ordinating actin remodelling and maintaining cell morphology. *Hum Mol Genet* 2010; 19: 2421-2432.
49. Kleczkowska A, Fryns JP, Moerman P et al. Trisomy of the short arm of chromosome 5: autopsy data in a malformed newborn with inv dup (5) (p13.1---p15.3). *Clin Genet* 1987; 32: 49-56.
50. Gustavson KH, Lundberg PO, Nicol P. Familial partial trisomy 5p resulting from segregation of an insertional translocation. *Clin Genet* 1988; 33: 404-409.
51. Reichenbach H, Holland H, Dalitz E et al. De novo complete trisomy 5p: clinical report and FISH studies. *Am J Med Genet* 1999; 85: 447-451.
52. Velagaleti GV, Morgan DL, Tonk VS. Trisomy 5p. A case report and review. *Ann Genet* 2000; 43: 143-145.
53. Grosso S, Cioni M, Garibaldi G et al. De novo complete trisomy 5p: clinical and neuroradiological findings. *Am J Med Genet* 2002; 112: 56-60.
54. Sheen VL, Wheless JW, Bodell A et al. Periventricular heterotopia associated with chromosome 5p anomalies. *Neurology* 2003; 60: 1033-1036.
55. Basel D, DePaepe A, Kilpatrick MW et al. Split hand foot malformation is associated with a reduced level of Dactylin gene expression. *Clin Genet* 2003; 64: 350-354.
56. Roscioli T, Taylor PJ, Bohlken A et al. The 10q24-linked split hand/split foot syndrome (SHFM3): narrowing of the critical region and confirmation of the clinical phenotype. *Am J Med Genet* 2004; 124A: 136-141.
57. Elliott AM, Evans JA. The association of split hand foot malformation (SHFM) and congenital heart defects. *Birth Defects Res Part A Clin Mol Teratol* 2008; 82: 425-434.
58. Dimitrov BI, de Ravel T, Van Driessche J et al. Distal limb deficiencies, micrognathia syndrome, and syndromic forms of split hand foot malformation (SHFM) are caused by chromosome 10q genomic rearrangements. *J Med Genet* 2010; 47: 103-111.

59. Chitty LS, Hunt GH, Moore J et al. Effectiveness of routine ultrasonography in detecting fetal structural abnormalities in a low risk population. *BMJ* 1991; 303: 1165-1169.
60. Luck CA. Value of routine ultrasound scanning at 19 weeks: a four year study of 8849 deliveries. *BMJ* 1992; 304: 1474-1478.

3.8 Supporting information

3.8.1 Methods

Table 3-S1 : Cohort description.

Fetus	Prospective / Retrospective	Array	Tissue	Karyotype	aCGH result	Malformation	Ultrasound
1	Retrospective	SignatureChipWG v1.0.1	Fetal blood	46,XX	♀ Normal	multiple	Ascites, imperforated anus, sexual ambiguity, dilated bowel loops, severe oligohydramnios, single umbilical artery, facial abnormalities
2	Retrospective	SignatureChipWG v1.0.1	Frozen fetal tissue	not performed	♂ Normal	isolated	Cerebral ventriculomegaly
3	Retrospective	SignatureChipWG v1.0.1	Amniotic fluid	46,XX	♀ Normal	isolated	Semilobar holoprosencephaly, microcephaly
5	Retrospective	SignatureChipOS v1.1 Rev B	Amniotic fluid	46,XY	♂ Normal	isolated	Fetal hydrops, cystic hygroma, hyperechogenic bowel loops, intracardiac hyperechogenic foci (left & right ventricles), IFD
8	Retrospective	SignatureChipOS v1.1 Rev B	Amniotic fluid	46,XY	♂ Normal	multiple	Cystic hygroma, retrognathia, closed hands, rocker bottom feet, pterygia?, arthrogryposis, fetal immobility
9	Retrospective	SignatureChipOS v1.1 Rev B	Amniotic fluid	46,XY	♂ Normal	multiple	Multiple malformations syndrome suggesting heterotaxia: complex cardiopathy, cardiomegaly, mesocardia, biventricular hypertrophy, left atrial isomerism, interrupted inferior vena cava, persistent left superior vena cava, pericardial effusion, atrioventricular block, atrioventricular septal defect, transposition of great vessels. Short humerus.
10	Prospective	SignatureChipWG v1.0.1	Amniotic fluid	46,XY	♂ Normal	multiple	Twin pregnancy: 1) right kidney agenesis, left kidney hypoplasia, IFD; 2) clubfeet, cardiopathy (arterial canal, pulmonary valve stenosis)
11	Prospective	SignatureChipWG v1.0.1	Amniotic fluid	46,XX	♀ Normal	isolated	Left hemimegalencephaly, cerebral ventriculomegaly (unilateral), gyration anomalies, macrocrania, dysmorphic facies
12	Retrospective	SignatureChipOS v2.0 12-plex	Frozen fetal tissue	46,XY	♂ arr cgh 3p21.31(48,325,068-48,983,044)x3	isolated	Diaphragmatic hernia, right anomalous pulmonary venous return
13	Retrospective	SignatureChipWG v1.0.1	Amniotic fluid	46,XX	♀ Normal	multiple	Dandy-Walker malformation, increased nuchal fold, short femur and humerus length, rocker bottom feet
14	Retrospective	SignatureChipWG v1.0.1	Amniotic fluid	46,XY	♂ Normal	isolated	Recurrent fetal immobility & cystic hygroma
16	Prospective	SignatureChipOS v1.1 Rev B	Amniotic fluid	46,XY	♂ Normal	multiple	Multiple malformations syndrome: corpus callosum agenesis, abnormal posterior fossa, abnormal mega cisterna magna, mild ventriculomegaly, vermis agenesis, left cerebellar hypoplasia/atrophy, irregular gyration (internal hemispheric surface), closed fists, small thorax & stomach, penis deformation, dysmorphic facies, intrauterine growth retardation, single umbilical artery
17	Prospective	SignatureChipOS v1.1 Rev B	Amniotic fluid	46,XX	♀ arr cgh Xp22.33(1381-114,549)x1, Xp22.33(259,532-309,095)x1	multiple	Dandy-Walker malformation, cerebral ventriculomegaly, cerebellar hypoplasia, mega cisterna magna, abnormal biparietal diameter/femur length ratio
18	Prospective	SignatureChipOS v1.1 Rev B	Chorionic villus sampling	46,XY	♂ arr cgh 15q14(32,340,370-32,628,739)x1	isolated	Cystic hygroma (recurrent), thin thorax, possible omphalocele, IFD, previous holoprosencephaly
20	Retrospective	SignatureChipOS v1.1 Rev B	Frozen fetal tissue	46,XY	♀ arr cgh 2p21(44,949,357-45,184,052)x1	isolated	Alobar holoprosencephaly, facial abnormalities (cyclopia, proboscis, arhinia)
21	Retrospective	SignatureChipOS v1.1 Rev B	Amniotic fluid	46,XY	♂ Normal	multiple	Fetal akinesia (arthrogryposis), clubhands & feet, flexed limbs, accentuated cervico-dorsal angulation, inferior vertebrae anomalies, single umbilical artery
22	Retrospective	SignatureChipOS v2.0 12-plex	Cord blood	46,XY	♂ arr cgh 3q22.3(139,749,445-140,055,929)x3; 6q23.3q24.1(138,848,815-139,632,692)x3; 7q11.21(64,317,115-64,579,853)x1	isolated	Cerebral ventriculomegaly, severe neuronal migration abnormalities with lissencephaly, microcephaly, single umbilical artery
23	Retrospective	SignatureChipOS v1.1 Rev B	Frozen fetal tissue	46,XX	♀ arr cgh 5p15.33(132,145-480,969)x3	isolated	Hydrocephalus by stenosis of the cerebral aqueduct
24	Prospective	SignatureChipOS v2.0 12-plex	Placental tissue	46,XY	♂ Normal	isolated	Intrauterine growth retardation, oligohydramnios, severe uteroplacental insufficiency, hyperechogenic bowel loops
25	Retrospective	SignatureChipOS v2.0 12-plex	Placental tissue	46,XX	♀ arr cgh 11q25(132,338,283-132,769,193)x3	isolated	Cerebral ventriculomegaly, hyperechogenic ventricle wall, periventricular leucomalacias, possible gyration retardation
26	Retrospective	SignatureChipOS v1.1 Rev B	Amniotic fluid	46,XY	♂ Normal	isolated	Diaphragmatic hernia, cardiac compression & associated pulmonary sequestrum
27	Prospective	SignatureChipOS v1.1 Rev B	Amniotic fluid	46,XY	♂ Normal	isolated	Cardiac malformation (dysplastic tricuspid valve with insufficiency, aortic valve anomaly), pulmonary artery dilatation, abnormal ductus venosus insertion
28	Prospective	SignatureChipOS v1.1 Rev B	Amniotic fluid	46,XY	♂ Normal	multiple	Situs inversus, diaphragmatic hernia, increased nuchal fold, closed & contracted hands, overlapping fifth finger, contracted toes, dysmorphic facies

Table 3-S1 : Cohort description (continued).

Fetus	Prospective / Retrospective	Array	Tissue	Karyotype	aCGH result	Malformation	Ultrasound
29	Prospective	SignatureChipOS v1.1 Rev B	Amniotic fluid	46,XY	♂Normal	multiple	Tetralogy of Fallot, vertebral column segmentation abnormalities & abnormal rachis, bilateral cleft lip & palate, abnormal hands position, short & arched femurs, syndactyly, dysmorphic facies
30	Prospective	SignatureChipOS v1.1 Rev B	Amniotic fluid	46,XY	♂Normal	isolated	Intrauterine growth retardation (severe & symmetrical), hypoplastic kidneys, oligohydramnios
31	Prospective	SignatureChipOS v1.1 Rev B	Amniotic fluid	46,XY	♂Normal	isolated	Cystic hygroma, hydronephrosis, caliectasis, right pleural effusion, bilateral pyelectasis
32	Retrospective	SignatureChipOS v1.1 Rev B	Amniotic fluid	46,XX	♀Normal	isolated	Cerebral ventriculomegaly, triventricular hydrocephaly, cystic posterior fossa with cerebellar hypoplasia, mild spina bifida
33	Prospective	SignatureChipOS v2.0 12-plex	Amniotic fluid	46,XX	♀Normal	multiple	Left diaphragmatic hernia, cardiomegaly, atrioventricular septal defect
34	Prospective	SignatureChipOS v2.0 12-plex	Amniotic fluid	46,XY	♂ arr cgh 10q21.3(67,688,386-67,840,524)x1	isolated	Bilateral schizencephaly, septal agenesis
35	Prospective	SignatureChipOS v2.0 12-plex	Chorionic villus sampling	46,XX	♀Normal	multiple	Cystic hygroma, choroid plexus cysts, situs inversus, left atrial isomerism, inferior vena cava agenesis
36	Prospective	SignatureChipOS v2.0 12-plex	Chorionic villus sampling	46,XX	♀ arr cgh 7p14.3p14.2(32,874,023-36,540,594)x1 mat	multiple	Increased nuchal fold, short nasal bone, persistent superior vena cava draining to the coronary sinus, single umbilical artery
37	Retrospective	NimbleGen CGX-12	Frozen fetal tissue	46,XX	♀Normal	isolated	Non-immune hydrops fetalis
38	Prospective	NimbleGen CGX-12	Amniotic fluid	46,XX	♀Normal	isolated	Left diaphragmatic hernia
39	Retrospective	NimbleGen CGX-12	Amniotic fluid	46,XX	♀Normal	multiple	Clubfeet, cardiac malformation (ventricular septal defect, hypoplastic right heart, tricuspid stenosis), small & poorly differentiated kidneys, closed right hand, mild oligohydramnios, possible corpus callosum agenesis
40	Retrospective	NimbleGen CGX-12	Amniotic fluid	46,XY	♂ arr cgh 3p26.2(4,434,665-4,745,478)x3	multiple	Large hyperechogenic kidneys, postaxial polydactyly (feet & left hand), clinodactyly (fifth finger left hand)
41	Prospective	NimbleGen CGX-12	Amniotic fluid	46,XY	♂Normal	multiple	Polycystic dysplastic kidneys, polydactyly, intrauterine growth retardation, thorax/abdomen discordance (short ribs)
42	Retrospective	NimbleGen CGX-12	Frozen fetal tissue	46,XY	♂Normal	multiple	Multiple malformations syndrome: cardiac malformation (abnormal cavities, single great vessel, possible left heart hypoplasia), micrognathia, corpus callosum agenesis, increased nuchal fold
43	Prospective	NimbleGen CGX-12	Chorionic villus sampling	46,XX	♀Normal	isolated	Increased nuchal fold, tricuspid insufficiency, pulmonary valve stenosis & dysplasia, nasal bone not visualized
44	Retrospective	NimbleGen CGX-12	Frozen fetal tissue	46,XX	♀ arr cgh 9q22.31q22.32(95,439,577-95,975,663)x3, 20p11.21(25,336,781-25,576,750)x1	isolated	Triventricular hydrocephaly, posterior parietal encephalocele
45	Retrospective	NimbleGen CGX-12	Amniotic fluid	46,XY	♂Normal	multiple	Atypical omphalocele, hyperechogenic bowel loops, hydrops, hepatomegaly, IFD
46	Retrospective	NimbleGen CGX-12	Frozen fetal tissue	46,XY	♂Normal	multiple	Dandy-Walker malformation, nephromegaly, polycystic kidneys, increased nuchal fold
47	Retrospective	NimbleGen CGX-12	Frozen fetal tissue	46,XY	♂ arr cgh 3q23q25.32(143,856,784-158,118,852)x1	multiple	Cerebral bilateral ventriculomegaly, mega cisterna magna, cerebellum & vermis atrophy, upward rotation of the vermis, aortic stenosis, bicuspid aortic valve
48	Prospective	NimbleGen CGX-12	Amniotic fluid	46,XX	♀ arr cgh 22q11.21(19,064,244-19,790,658)x1	multiple	Severe clubfeet, right polycystic dysplastic kidney, median ureterocele
49	Retrospective	NimbleGen CGX-3	Frozen fetal tissue	not performed	♂Normal	isolated	Bilateral kidney agenesis, four previous pregnancies with Potter's syndrome
50	Retrospective	NimbleGen CGX-12	Amniotic fluid	46,XX	♀Normal	multiple	Frontal encephalomenocele, corpus callosum agenesis, holoprosencephaly, atrioventricular septal defect
51	Prospective	NimbleGen CGX-12	Amniotic fluid	46,XX	♀ arr cgh 10q24.32(103,201,726-103,433,500)x3	multiple	Right clubfoot, pelvic kidney, ventricular septal defect, two superior venae cava, single umbilical artery
52	Prospective	NimbleGen CGX-12	Amniotic fluid	46,XX	♀Normal	multiple	Vertebral segmentation abnormality, ventricular septal defect, aortic coarctation, single umbilical artery
53	Retrospective	NimbleGen CGX-12	Frozen fetal tissue	46,XY	♂ arr cgh 10q26.13q26.3(124,590,071-132,729,781)x1	isolated	Hyperechogenic kidneys, bilateral pyelocaliectasis, vesical distension, urethral cyst
54	Retrospective	NimbleGen CGX-12	Amniotic fluid	46,XX	♀Normal	isolated	Neuronal migration abnormalities, polymicrogyria, cerebral ventriculomegaly with possible septal rupture

3.8.1.1 Samples preparation

Amniotic fluid, chorionic villus sampling (CVS), and blood samples were cultured using standard procedures, and other tissues were frozen at -80°C. GTG-banded karyotype and FISH analysis were performed on *in situ* cultured cells using routine cytogenetic protocols. Purified DNA was obtained from cultured amniocytes, cultured CVS and frozen fetal tissues using the ArchivePure DNA Cell/Tissue Kit (5 Prime Inc., Gaithersburg, MD), and from blood using the ArchivePure DNA Blood Kit (5 Prime Inc., Gaithersburg, MD), according to manufacturer's instructions.

3.8.1.2 Targeted FISH

Targeted FISH analyses were performed for the following fetuses, when clinically indicated.

Table 3-S2 : List of fetuses tested with FISH and probes used.

FISH probe(s)	Manufacturer	Fetuses tested
TUPLE1	Vysis	10, 27, 28, 29, 39, 42, 47, 50, 51
Multiprobe-T subtelomere	Cytocell	12, 38, 39, 42, 43, 46
D12Z3	Cytocell	12, 26, 38
WHS	Vysis	29
ELN	Vysis	47

3.8.1.3 Arrays

The SignatureChip Whole Genome v1.0.1 contains 4685 BACs distributed on 1543 loci, and targets over 150 genetic syndromes and 500 functionally important genes. The average gap between BACs is 1.6 Mb throughout the genome, and 80-100 Kb in targeted regions. The SignatureChip Oligo Solution v1.1 rev B (Agilent Technologies, Santa Clara, CA) targets the same loci as the SignatureChipWG. It consists of 105 000 oligonucleotides, with an average spacing of 10 Kb between probes in targeted regions, and 65-75 Kb in the rest of the genome. The SignatureChipOS v2.0 12-plex and NimbleGen CGX arrays (Roche

NimbleGen, Madison, WI) are the same array with a different name, and contain 135 000 oligonucleotides. They target 675 genes and more than 200 important cytogenetic regions including subtelomeres, pericentromeric regions and 170 genetic syndromes with the same average spacing as the 105 K array.

All hybridizations were performed with a sex-matched commercial pooled reference DNA (Promega, Madison, WI). Array data were analyzed with DNA Analytics v4.0 (Agilent Technologies), and abnormal regions were called using the Aberration Detection Method 1 (ADM1) algorithm. Imbalances spanning a minimum of one BAC, or five oligonucleotides, with a mean \log_2 ratio above 0.4 (gains) or below -0.7 (losses) were considered and confirmed visually.

3.8.1.4 Web database links

DGV: <http://projects.tcag.ca/variation>

DECIPHER: <https://decipher.sanger.ac.uk/application>

OMIM: <http://www.ncbi.nlm.nih.gov/omim>

MEDLINE: <http://www.ncbi.nlm.nih.gov/pubmed>

RefSeq: <http://www.ncbi.nlm.nih.gov/RefSeq>

UCSC Genome Browser: <http://genome.ucsc.edu/cgi-bin/hgGateway>

3.8.1.5 Statistical analysis

Statistical significance was tested with a Likelihood Ratio Chi-Square Exact Test and was performed with SPSS v16.

3.8.2 Results & discussion

3.8.2.1 Clinically significant results: additional clinical & interpretation details

Fetus 20

SIX2 is involved in early nephrogenesis (16), and missense mutations in this gene have been identified in patients with renal hypodysplasia (17). To our knowledge, haploinsufficiency of *SIX2* has not been associated with renal abnormalities. This is consistent with the fact that our patient presented normal kidney development both on ultrasound examination, and at autopsy.

Fetus 47

Due to advanced autolysis, the neuropathological examination failed to confirm the cerebral abnormalities. The unbalanced segment includes the Dandy-Walker syndrome (DWS, MIM 220200) critical region, which is 7 Mb in size and encompasses 25 genes. The remaining segment distal to the DWS region is 6.86 Mb in size and encompasses 39 genes. The remaining proximal segment is 400 kb in size and encompasses 5 genes.

The 14.26 Mb 3q23q25.32 deletion found in fetus 47 includes the DWS critical region (3q24q25.1), consistent with the fetus's phenotype. Few patients with a deletion in this region have been reported. A number of patients with DWS have been reported with various chromosomal abnormalities (18), including 3q deletions (19-21). Fine-mapping of breakpoints in seven patients with DWS and carrying interstitial 3q deletions (22) has allowed the localization of a 7 Mb minimally deleted region at band 3q24, and pointed to two candidate genes for DWS: *ZIC1* (MIM 600470) and *ZIC4* (MIM 608948).

Fetus 48

FISH confirmed the CNV in the patient and her mother: .ish del(22)(q11.21q11.21)(RP11-801O20-).

Although the patient and her mother had no cardiac abnormalities, the urogenital abnormalities and clubfeet present in the patient are among the abnormalities commonly found in individuals with 22q11.2 deletion (23-26). Consistent with what is observed in DG

syndrome, this atypical deletion seems to exhibit variable expressivity and reduced penetrance. The deleted segment spans at least 19 genes, from *ZNF74* (MIM 194548) proximally, to LOC400891 distally. Because the adjacent regions are rich in low-copy repeats (LCR) and segmental duplications, they are not covered by the array used, and the deleted segment could be as long as 1.5 Mb, from *RTN4R* (MIM 605566) to *HIC2* (MIM 607712). The breakpoints in our patient fall within known LCRs: LCR-B proximally, and LCR-D distally (27). These breakpoints match an atypical distal deletion that has been reported in two individuals with DG/VCF syndrome (28, 29). One of them had tetralogy of Fallot and mild facial dysmorphism, while the other had no cardiac anomaly and presented significant developmental delay, hypotonia and facial features mildly suggestive of DG/VCF syndrome. A third individual was reported in DECIPHER with a slightly smaller deletion (Patient 251336). The observed phenotype includes mental retardation / developmental delay and seizures. No urological abnormalities were reported in any of those individuals.

Fetus 53

Urological abnormalities: hyperechogenic kidneys, bilateral pyelocaliectasis, vesical distension, and urethral cyst. Autopsy: imperforated anus, mega rectum, dilated sigmoid, bilateral hydroureter, obstructive polycystic dysplastic kidneys. Dysmorphic facies associated to Potter's sequence: anteverted nares and retrognathia.

FISH with the RP11-16P8 BAC clone showed that parents did not carry the CNV.

Individuals with 10q monosomy present a recognizable, but clinically heterogeneous phenotype: facial dysmorphism; cardiac, renal, urinary tract, and anogenital abnormalities; and diverse developmental deficits (growth failure, psychomotor and developmental delay, mental retardation) (41-57). Most reported cases are terminal deletions. A recent study reported five patients and reviewed previously published patients with 10q deletion involving the same region as the one deleted in our patient (57). The authors suggested three critical regions responsible for the 10q- phenotype. The first, ranging from q25.3 to q26.12, is associated with sex reversal; the second, in band 10q26.2, is involved in craniofacial, cardiac and neuropsychiatric anomalies; and a third, in 10q26.2q26.3, is involved in urogenital defects. The deleted segment found in our patient spans over 40 genes, including *DOCK1* (MIM

601403), which has been suggested as a candidate gene for the urinary, craniofacial and developmental anomalies associated with 10q deletion (57). *DOCK1* (Dedicator of cytokinesis-1) plays a role in cytoskeletal reorganization during apoptosis and cell migration (58-60), and is highly expressed in placenta, lung, kidney, pancreas and ovary (61). Our case further refines the critical region involved in urogenital abnormalities associated with 10q deletion, which will help identify the causal gene.

3.8.2.2 Results of unclear clinical significance: additional clinical & interpretation details

Fetus 12

FISH (Pallister-Killian syndrome) was performed on 200 interphase amniocytes using a chromosome 12 centromeric probe (D12Z3, Cytocell, Cambridge, UK).

The 658 kb 3p21.31 duplication found in fetus 12 is most likely responsible for the fetus's dysmorphic features, which are found in trisomy 3p syndrome, although the link with the diaphragmatic hernia and joints abnormalities remains unclear. Partial trisomy 3p has been associated with a variable phenotype, including psychomotor and mental retardation, dysmorphism (hypertelorism/telecanthus, frontal bossing, square face, short neck), congenital heart defect; and, less frequently, holoprosencephaly, orofacial clefts, brachycephaly, microcephaly, urogenital abnormalities, and other mild dysmorphic features (1, 2). Few cases of pure partial trisomy 3p have been reported (3-12), and none of these patients presented holoprosencephaly/cyclopia, although one family with a 3p21p26 trisomy presented cleft lip and palate (1/3), and two had congenital heart defects (3, 4). Six of the reported individuals with pure partial trisomy 3p share a common region with fetus 12, and most of his dysmorphic features are found in at least one of them. None of the patients with partial trisomy 3p (either pure or not) have been reported with diaphragmatic hernia or camptodactyly. These features can suggest Fryns syndrome (MIM 229850), however, fetus 12 lacks many typical features of this syndrome. Goumy *et al.* reported one fetus with a partial trisomy 3p21pter resulting from an unbalanced t(3;5)(p21;p13), who presented Larsen-like phenotype (13). He shares many characteristics with our fetus: multiple dislocations, down-turned mouth, short neck,

micrognathia, and weak pulmonary hypoplasia. Multiple congenital dislocations and clubfeet are characteristic features of Larsen syndrome (MIM 150250), which is an autosomal dominant disorder caused by missense mutations or in-frame deletions in *FLNB* (MIM 603381) located at 3p14.3 (14). Diaphragmatic hernia is not part of Larsen syndrome's clinical features, but it has been reported in one of two siblings with Larsen-like phenotype, both of whom also had joint dislocations (15).

Fetus 18

This 33 year-old French-Canadian mother previously had one healthy girl, one spontaneous abortion at 8 weeks, and one interruption of pregnancy for multiple fetal malformations at 20 weeks. The malformations in the previous pregnancy included cystic hygroma; intrauterine growth retardation; lobar holoprosencephaly with multiple nervous system malformations (microcephaly, corpus callosum and commissural agenesis, migration abnormalities including polymicrogyria, cerebellar hypogenesis, and possible syringomyelia); ventricular septal defect; inferior limbs hypoplasia; arthrogryposis; megacystis with bilateral hydronephrosis; and dysmorphic facies.

Both parents had a normal aCGH result. Allele-specific PCR determined that parents were not carriers of the 2436delG mutation in the *SLC12A6* gene causing Andermann syndrome in French-Canadians.

Homozygous mutations in the *SLC12A6* gene are a known cause of agenesis of the corpus callosum with peripheral neuropathy (Andermann syndrome, MIM 218000) (62), but parents did not carry the mutation usually found in French-Canadians with this syndrome. However, a different mutation could be present in one of the parents. Clefting anomalies and cryptorchidism are among the recurrent abnormalities observed in individuals with 15q14 deletions (63-67). To our knowledge, the smallest 15q14 deletion found in a patient presenting a cleft anomaly is reported in DECIPHER (Patient 1770). It is 920 kb in size, and completely overlaps with the deletion found in fetus 18. Despite the possible link between the phenotype observed in this fetus, the mother's history and the genes located in the deleted region, we could not confirm that this CNV is responsible for the phenotype, and it remains of uncertain clinical significance until further investigation is completed.

Fetus 22

Brain malformations: severe diffuse abnormal bilateral supratentorial gyration and microcephaly. Brainstem, vermis and cerebellum were normal.

Trisomy 6q syndrome is usually found in patients with terminal 6q duplication, and is well described: microcephaly, acrocephaly, psychomotor and mental retardation, characteristic facies, cardiac anomalies, joint contractures, clinodactyly, syndactyly, and clubfoot (68-71). The smallest region associated with trisomy 6q syndrome is a tandem duplication of 6q23.2q23.3 reported in a 12 year-old boy displaying most of the characteristic features, except microcephaly (72). The duplicated segment in our patient is just distal to this region, with no overlap. Brain malformations are not typically found in this syndrome. A stillborn boy with multiple congenital anomalies was reported with a terminal duplication of 6q21 (73). He had microcephaly, occipital encephalocele, and focal neuronal disorganization. Another patient with a terminal duplication of 6q23 had multiple malformations, including hydrocephalus, partial aplasia of the vermis of the cerebellum, and cerebral cortex atrophy (74). Our patient does not have any of the reported trisomy 6q features, except microcephaly, which has been linked to duplication of 6q23q24 (75). The *NHSL1* (Nance-Horan syndrome like-1) gene encodes a member of a new family of proteins that include NHS, NHSL1 and NHSL2. These proteins have a conserved domain homologous to the WAVE proteins, which regulate actin assembly (76). Mutations in *NHS* result in Nance-Horan Syndrome (MIM 302350), which is characterized by cataracts, dental anomalies, dysmorphism, and mental retardation. *NHS* expression pattern during fetal development is consistent with its probable involvement in the regulation of cell morphology and motility in the structures affected in Nance-Horan Syndrome (76-78). Ectopic overexpression of *NHS* inhibited lamellipod formation essential for neuronal migration. *NHSL1* is more broadly expressed during development, in multiple tissues including brain, liver, kidney, heart, and lung (78). A CNV duplicating the entire gene has been reported in a healthy control (79). However, the proximal breakpoint in our fetus interrupts *NHSL1*, which is thus only partially duplicated. *TXLNB* (taxilin beta) encodes a taxilin isoform involved in intracellular trafficking, which interacts with syntaxins (preferentially syntaxin 1A, STX1A) (80). *STX1A* expression is restricted to the brain, is required for synaptic vesicle docking, and may also be involved in synaptic plasticity

(81). The duplication found in fetus 22 interrupts *NHSL1*, and probably *TXLNB*. Thus, it could result in a fusion of *NHSL1* (exons 1, 2 and 3) and *TXLNB* (exons 5-10), which would encode a chimeric protein that might interfere with normal brain function or plasticity. In *NHSL1*, the conserved functional domain required for interaction with the Abl interactor family of proteins during actin remodeling is encoded by exon 1 (76), while the syntaxin-interacting domain in *TXLNB* is the long coiled-coil region in its C-terminal half (80). Alternatively, if the distal breakpoint is outside *TXLNB*, this gene could be completely duplicated and its expression increased. Mutated genes that are known to play a role in neuronal migration disorders include *LISI*, *DCX*, *TUBA1A*, *RELN*, and *FLNA*. These genes encode proteins that are involved in cytoskeleton stabilization and cell migration through their interaction with microtubule, actin or extracellular matrix (82, 83). Considering the possible roles of *NHSL1* and *TXLNB*, and their expression pattern in brain, liver, kidney, and lung, which were all affected in fetus 22, we deemed this CNV potentially pathogenic, but reported it as a CNV of unclear clinical significance.

Fetus 23

This 32 year-old mother previously had three healthy children, and a premature baby girl born at 23 weeks gestation who did not survive.

Mullerian ducts abnormalities: agenesis of the left kidney, ureter, and uterine tube. Dysmorphic facies: large anterior fontanel, hypertelorism, microretrognathia, anteverted nares, and low-set ears.

Trisomy 5p features include mental retardation, facial dysmorphism, clubfeet, and cardiac, renal, and intestinal malformations (84). Hydrocephalus and ventriculomegaly have also been reported in patients with complete (85, 86) or partial trisomy 5p (87, 88). It has been suggested that duplications involving 5p11p13 are associated with a more severe phenotype than 5p14p15 duplications (89, 90). A mother and her two children, carrying a distal 5p15.3 duplication, have been reported (91). The features observed in this family include developmental delay, learning and behavior problems, attention deficit disorder, clinodactyly and mild dysmorphism. Sheen *et al.* reported a patient with a 3 Mb 5p15.33 terminal duplication that presented atrial septal defect, minor craniofacial abnormalities, bilateral

clubfoot, clinodactyly, and periventricular heterotopias (92). Our fetus's phenotype is strikingly similar to the phenotype associated with trisomy 5p, and if this duplication is responsible for the observed malformations, it would be the smallest duplication causing a trisomy 5p phenotype reported so far. However, until further evidence is available, this CNV remains of uncertain clinical significance.

Fetus 36

At 12 months, this girl shows a normal weight, a slightly short height (10th percentile) and reduced head size (3rd percentile). She and her mother present mild dysmorphic features (high anterior forehead and anteverted nares).

Fetus 51

In addition to duplicating part of *FBXW4* (MIM 608071), the unbalanced segment also encompasses *DPCD*, *POLL* (MIM 606343) and part of *BTRC* (MIM 603482). The next distal probe with a normal signal is at 12 kb, farther than the 5' end of the *FBXW4* gene, thus, this gene could be completely duplicated. The closest normal proximal probe is at 32 kb, within the *BTRC* gene.

Split-hand/foot malformation, or ectrodactyly, is characterized by syndactyly and a deep median cleft in the hand and/or foot, with or without aplasia/hypoplasia of the phalanges, metacarpal and metatarsal bones. Although our patient does not present the distinctive features of SHFM3, SHFM3 has been known to exhibit variable expressivity and reduced penetrance (30, 31). Also, cardiac and renal abnormalities have been reported in the syndromic form of SHFM3 (32, 33). Microduplications found in patients with syndromic SHFM3 are at least 500 kb in size, and encompass the following genes: *LBX1*, *BTRC*, *POLL*, *DPCD* and *FBXW4* (33-36). A similar duplication has been found in two families previously reported with Distal limb deficiencies with micrognathia (MIM 246560) (33), suggesting these individuals might in fact express the most severe form of syndromic SHFM3. Although *FBXW4* has been the candidate gene of choice, because of its possible role in the maintenance of the apical ectodermal ridge and the phenotype observed in *Dactylaplasia* (*FBXW4* mutant) mice (30, 34, 37, 38), recent studies suggest that SHFM3 might result from a disruption in the interactions between many genes (including genes outside the minimally duplicated region, such as *TLX1* or *FGF8*) and

their long-range regulatory elements (33, 39, 40). This is the first report of a patient with clubfoot in association with 10q24.32 duplication. As aCGH analysis is performed on more patients, we can expect an expansion of the phenotypic features associated with this microduplication.

3.8.2.3 Benign CNVs: clinical & interpretation details

Fetus 17

This 32 year-old mother was referred upon finding Dandy-Walker malformation at 20 weeks on ultrasound examination in her female fetus, which also showed discordance between humeral length and biparietal diameter. She previously had two spontaneous abortions at 7 and 8 weeks, and one healthy boy. Autopsy confirmed the cerebral malformation, and revealed a bilobed right lung, bilateral clinodactyly of the fifth toe, a large anterior fontanel, and dysmorphic facies (hypertelorism, retrognathia, and low-set ears). Array CGH was performed on cultured amniocytes, and results showed two deletions in the pseudoautosomal region (Xp22.33) on the X chromosome. Neither alterations were found in the normal variants databases, but each one was inherited from a different phenotypically normal parent. The smallest alteration is 50 kb in size, interrupts *PPP2R3B* (MIM 300339), and was inherited from the healthy mother. The second alteration is 113 kb in size, contains no gene, and was inherited from the father. Thus, both were thought to be benign. The mother subsequently had a third spontaneous abortion at 8 weeks of pregnancy.

Fetus 22

In addition to the potentially pathogenic CNV reported above, this fetus also had a 306 kb duplication of 3q22.3, and a 263 kb deletion of 7q11.21, which were both inherited from the healthy mother, and were classified as benign.

Fetus 25

This 26 year-old mother was referred after an ultrasound performed at 19 weeks showed cerebral ventriculomegaly. Further imaging studies (ultrasound at 21.5 weeks, and

magnetic resonance imaging at 24 weeks) revealed multiple cerebral lesions suggestive of periventricular leukomalacia. Autopsy also showed microcephaly, abnormalities in neuronal migration (focal polymicrogyria in the right frontal lobe; bilateral heterotopias of medullary olive; and caudate dysplasia or fasciculation of the dentate nucleus), high-arched palate, and dysmorphic facies (hypertelorism and pointed chin). Array CGH performed on placental tissues showed a 431 kb paternally inherited duplication of 11q25. The unbalanced segment partially overlaps with the Jacobsen syndrome region, and is located within an intron of the *OPCML* gene (MIM 600632). A slightly larger CNV (481.5 kb) has been reported in two healthy controls (79). We also found this CNV in three other healthy individuals at our center. Therefore, it is thought to be benign.

Fetus 34

The mother was referred at 24 weeks gestation for bilateral schizencephaly and septal agenesis in her fetus. Autopsy also showed overlapping skull bones, a bulging membranous covered cyst in the region adjacent to the sylvian fissures, and dysmorphic facies (hypertelorism, receding forehead, retrognathia, short nose, depressed nasal ridge, smooth philtrum, and thin vermilion of the upper lip). Array CGH on cultured amniocytes showed a 152 kb deletion of 10q21.3, for which parents could not be tested. FISH on uncultured amniocytes using BAC RP11-791D13 was inconclusive (signal too weak). The deleted segment spans exons 12 and 13 of the *CTNNA3* gene (MIM 607667). This gene is subject to paternal imprinting in villus cytotrophoblast during the first trimester (93). Its product, α -T-catenin is expressed in adult neurons, particularly those in the fourth and fifth cortical layers (94). The α -catenins are necessary for the proper function of the cell-adhesion complexes by which neuronal positioning occurs during brain development (95, 96). It has also been suggested that one of them, α -N-catenin, might be necessary for the longitudinal dispersion of the Purkinje cells in the cerebellum (97).

Despite a possible link between *CTNNA3* function and the cerebral abnormalities observed in this fetus, it is likely that a deletion in one allele has no phenotypic effect. In fact, many benign CNVs are found in this region, and a deletion hotspot affecting coding exons in the 3' half of *CTNNA3* was identified in various cohorts, including normal individuals (98).

Furthermore, we later detected four more deletions in this gene in patients without cerebral abnormalities. Thus, after careful evaluation, we considered this CNV benign.

Fetus 40

This 36 year-old mother had two previous spontaneous abortions, and one healthy child. Ultrasound examination at 21 weeks of pregnancy showed post-axial polydactyly in both feet and in the left hand, clinodactyly of the fifth left finger, and bilateral hyperechogenic nephromegaly. Autopsy also showed dysmorphic facies (micrognathia and broad nose), polysplenia, and calcifications in the pancreatic head. Array CGH on cultured amniocytes showed an inherited 311 kb duplication of 3p26.2. Both breakpoints are located within genes: *SUMF1* (MIM 607939) and *ITPRI* (MIM 147265). This CNV was inherited from the healthy father. Thus, it was thought to be benign.

Fetus 44

This 27 year-old mother previously had one spontaneous abortion at 12 weeks, and one girl with lateral cleft lip. Ultrasound examination performed at 20 weeks showed triventricular hydrocephalus, and posterior parietal encephalocele. The pregnancy was terminated at 20 weeks. Autopsy revealed pulmonary hypoplasia, wide nasal base, anteverted nares, and placental hypoplasia. Neuropathological examination showed a 1 cm low occipital meningoencephalocele; partial hernia of the rhombencephalic structures, below the occipital hole; and complete corpus callosum agenesis. Dandy-Walker malformation was macroscopically excluded. Array CGH on cultured amniocytes showed two alterations: a 536 kb duplication of 9q22.31q22.32 and a 240 kb deletion of 20p11.21. The duplicated region on chromosome 9 encompasses two genes (*BARX1*, MIM 603260; *PTPDC1*) and interrupts *PHF2* (MIM 604351). The deleted region on chromosome 20 encompasses two genes (*NINL*, MIM 609580; *NANP*, MIM 610763), and interrupts *GINS1* (MIM 601608). Both CNVs were inherited from the healthy mother, and were thought to be benign. She subsequently became pregnant again, and ultrasound examinations (at 20, 24, 29 and 31 weeks) showed no abnormalities in her fetus.

3.8.3 References

1. Conte RA, Pitter JH, Verma RS. Molecular characterization of trisomic segment 3p24.1-->3pter: a case with review of the literature. *Clin Genet* 1995; 48: 49-53.
2. Chen C-P, Wang T-H, Lin C-C et al. Prenatal diagnosis of partial trisomy 3p (3p21-->pter) and partial monosomy 11q (11q23-->qter) associated with abnormal sonographic findings of holoprosencephaly, orofacial clefts, pyelectasis and a unilateral duplex renal system. *J Formos Med Assoc* 2008; 107: 822-826.
3. Rethoré MO, Lejeune J, Carpentier S et al. [Trisomy for the distal part of the short arm of the number 3 chromosome in 3 siblings. First example of chromosomal insertion: INS(7;3)(q 31;p 21 p 26)]. *Ann Genet* 1972; 15: 159-165.
4. Centerwall W, Mankinen C, Marsh W et al. The 3p trisomy syndrome: de novo case identified via banded prometaphase chromosomes. In: Bloom AD, James LS, Paul NW et al., eds. *The Fetus and the Newborn, Birth Defects original article series, Vol. 17*. New York: A.R. Liss, 1981: 77-78.
5. Kleczkowska A, Fryns JP, van den Berghe H. Partial trisomy of chromosome 3(p14----p22) due to maternal insertional translocation. *Ann Genet* 1984; 27: 180-183.
6. Orye E, Laureys G. Trisomy 3p syndrome. Report of a new case, due to a chromosomal insertion. *Ann Genet* 1984; 27: 245-247.
7. Zhang SZ, Wang Q. Partial serial duplication of the short arm of chromosome 3. *Chin Med J* 1984; 97: 425-428.
8. Watson MS, Dowton SB, Rohrbaugh J. Case of direct insertion within a chromosome 3 leading to a chromosome 3p duplication in an offspring. *Am J Med Genet* 1990; 36: 172-174.
9. Kotzot D, Krüger C, Braun-Quentin C. De novo direct duplication 3 (p25-->pter): a previously undescribed chromosomal aberration. *Clin Genet* 1996; 50: 96-98.
10. Smeets E, Vandenbossche L, Fryns JP. Partial distal trisomy 3p. A partial autosomal trisomy without major dysmorphic features. *Genet Couns* 2001; 12: 85-89.

11. Antonini S, Kim CA, Sugayama SM et al. Delimitation of duplicated segments and identification of their parental origin in two partial chromosome 3p duplications. *Am J Med Genet* 2002; 113: 144-150.
12. Stuart SW, King CH, Pai GS. Autism Spectrum Disorder, Klinefelter Syndrome, and Chromosome 3p21.31 Duplication: A Case Report. *Medscape General Medicine* 2007; 9: 60.
13. Goumy C, Beaufrère AM, Tchirkov A et al. Larsen-like phenotype associated with partial trisomy 3p and monosomy 5p. *Prenat Diagn* 2008; 28: 131-134.
14. Bicknell LS, Farrington-Rock C, Shafeghati Y et al. A molecular and clinical study of Larsen syndrome caused by mutations in *FLNB*. *J Med Genet* 2007; 44: 89-98.
15. Topley JM, Varady E, Lestringant GG. Larsen syndrome in siblings with consanguineous parents. *Clin Dysmorphol* 1994; 3: 263-265.
16. Self M, Lagutin OV, Bowling B et al. *Six2* is required for suppression of nephrogenesis and progenitor renewal in the developing kidney. *EMBO J* 2006; 25: 5214-5228.
17. Weber S, Taylor JC, Winyard P et al. *SIX2* and *BMP4* mutations associate with anomalous kidney development. *J Am Soc Nephrol* 2008; 19: 891-903.
18. Imataka G, Yamanouchi H, Arisaka O. Dandy-Walker syndrome and chromosomal abnormalities. *Congenital anomalies* 2007; 47: 113-118.
19. Willner JP, Vosatka RJ, Hirschhorn K. Interstitial deletion of chromosome 3q with Dandy-Walker Syndrome. *Am J Hum Genet* 1990; 47: A44.
20. Sudha T, Dawson AJ, Prasad AN et al. De novo interstitial long arm deletion of chromosome 3 with facial dysmorphism, Dandy-Walker variant malformation and hydrocephalus. *Clin Dysmorphol* 2001; 10: 193-196.
21. Zweier C, Guth S, Schulte-Mattler U et al. 9 Mb deletion including chromosome band 3q24 associated with unsuspecting facial gestalt, persistent ductus omphaloentericus, mild mental retardation and tic. *Eur J Med Genet* 2005; 48: 360-362.
22. Grinberg I, Northrup H, Ardinger H et al. Heterozygous deletion of the linked genes *ZIC1* and *ZIC4* is involved in Dandy-Walker malformation. *Nat Genet* 2004; 36: 1053-1055.
23. Devriendt K, Swillen A, Fryns JP et al. Renal and urological tract malformations caused by a 22q11 deletion. *J Med Genet* 1996; 33: 349.

24. Ming JE, McDonald-McGinn DM, Megerian TE et al. Skeletal anomalies and deformities in patients with deletions of 22q11. *Am J Med Genet* 1997; 72: 210-215.
25. Wu H-Y, Rusnack SL, Bellah RD et al. Genitourinary malformations in chromosome 22q11.2 deletion. *J Urol* 2002; 168: 2564-2565.
26. McDonald-McGinn DM, Emanuel BS, Zackai EH. 22q11.2 Deletion Syndrome. In: Pagon RA, Bird TC, Dolan CR et al., eds. *GeneReviews*, Accessed May 29, 2010. Seattle: University of Washington, 1997-2010.
27. Shaikh TH, Kurahashi H, Saitta SC et al. Chromosome 22-specific low copy repeats and the 22q11.2 deletion syndrome: genomic organization and deletion endpoint analysis. *Hum Mol Genet* 2000; 9: 489-501.
28. Garcia-Miñaur S, Fantes J, Murray RS et al. A novel atypical 22q11.2 distal deletion in father and son. *J Med Genet* 2002; 39: e62.
29. Rauch A, Zink S, Zweier C et al. Systematic assessment of atypical deletions reveals genotype-phenotype correlation in 22q11.2. *J Med Genet* 2005; 42: 871-876.
30. Basel D, DePaepe A, Kilpatrick MW et al. Split hand foot malformation is associated with a reduced level of Dactylin gene expression. *Clin Genet* 2003; 64: 350-354.
31. Roscioli T, Taylor PJ, Bohlken A et al. The 10q24-linked split hand/split foot syndrome (SHFM3): narrowing of the critical region and confirmation of the clinical phenotype. *Am J Med Genet* 2004; 124A: 136-141.
32. Elliott AM, Evans JA. The association of split hand foot malformation (SHFM) and congenital heart defects. *Birth Defects Res Part A Clin Mol Teratol* 2008; 82: 425-434.
33. Dimitrov BI, de Ravel T, Van Driessche J et al. Distal limb deficiencies, micrognathia syndrome, and syndromic forms of split hand foot malformation (SHFM) are caused by chromosome 10q genomic rearrangements. *J Med Genet* 2010; 47: 103-111.
34. de Mollerat XJ, Gurrieri F, Morgan CT et al. A genomic rearrangement resulting in a tandem duplication is associated with split hand-split foot malformation 3 (SHFM3) at 10q24. *Hum Mol Genet* 2003; 12: 1959-1971.
35. Kano H, Kurosawa K, Horii E et al. Genomic rearrangement at 10q24 in non-syndromic split-hand/split-foot malformation. *Hum Genet* 2005; 118: 477-483.

36. Everman DB, Morgan CT, Lyle R et al. Frequency of genomic rearrangements involving the SHFM3 locus at chromosome 10q24 in syndromic and non-syndromic split-hand/foot malformation. *Am J Med Genet* 2006; 140: 1375-1383.
37. Crackower MA, Motoyama J, Tsui LC. Defect in the maintenance of the apical ectodermal ridge in the Dactylaplasia mouse. *Dev Biol* 1998; 201: 78-89.
38. Sidow A, Bulotsky MS, Kerrebrock AW et al. A novel member of the F-box/WD40 gene family, encoding dactylin, is disrupted in the mouse dactylaplasia mutant. *Nat Genet* 1999; 23: 104-107.
39. Lyle R, Radhakrishna U, Blouin J-L et al. Split-hand/split-foot malformation 3 (SHFM3) at 10q24, development of rapid diagnostic methods and gene expression from the region. *Am J Med Genet* 2006; 140: 1384-1395.
40. Friedli M, Nikolaev S, Lyle R et al. Characterization of mouse Dactylaplasia mutations: a model for human ectrodactyly SHFM3. *Mamm Genome* 2008; 19: 272-278.
41. Shapiro SD, Hansen KL, Pasztor LM et al. Deletions of the long arm of chromosome 10. *Am J Med Genet* 1985; 20: 181-196.
42. Wulfsberg EA, Weaver RP, Cunniff CM et al. Chromosome 10qter deletion syndrome: a review and report of three new cases. *Am J Med Genet* 1989; 32: 364-367.
43. Rooney DE, Williams K, Coleman DV et al. A case of interstitial deletion of 10q25.2---q26.1. *J Med Genet* 1989; 26: 58-60.
44. Gorinati M, Zamboni G, Padoin N et al. Terminal deletion of the long arm of chromosome 10: case report and review of the literature. *Am J Med Genet* 1989; 33: 502-504.
45. Teyssier M, Charrin C, Dutruge J et al. Monosomy 10qter: a new case. *J Med Genet* 1992; 29: 342-343.
46. Turk J, Christie D, Sales J et al. The psychiatric, psychological and behavioural functioning of a boy with terminal deletion of the long arm of chromosome 10. *Dev Med Child Neurol* 1993; 35: 439-448.
47. Petersen B, Strassburg HM, Feichtinger W et al. Terminal deletion of the long arm of chromosome 10: a new case with breakpoint in q25.3. *Am J Med Genet* 1998; 77: 60-62.
48. Waggoner DJ, Chow CK, Dowton SB et al. Partial monosomy of distal 10q: three new cases and a review. *Am J Med Genet* 1999; 86: 1-5.

49. Leonard NJ, Harley FL, Lin CC. Terminal deletion of chromosome 10q at band 26.1: follow-up in an adolescent male with high-output renal failure from congenital obstructive uropathy. *Am J Med Genet* 1999; 86: 115-117.
50. Ogata T, Muroya K, Sasagawa I et al. Genetic evidence for a novel gene(s) involved in urogenital development on 10q26. *Kidney Int* 2000; 58: 2281-2290.
51. Irving M, Hanson H, Turnpenny P et al. Deletion of the distal long arm of chromosome 10; is there a characteristic phenotype? A report of 15 de novo and familial cases. *Am J Med Genet* 2003; 123A: 153-163.
52. Scigliano S, Grégoire MJ, Schmitt M et al. Terminal deletion of the long arm of chromosome 10. *Clin Genet* 2004; 65: 294-298.
53. Kehrer-Sawatzki H, Daumiller E, Müller-Navia J et al. Interstitial deletion del(10)(q25.2q25.3 approximately 26.11)--case report and review of the literature. *Prenat Diagn* 2005; 25: 954-959.
54. Courtens W, Wuyts W, Rooms L et al. A subterminal deletion of the long arm of chromosome 10: a clinical report and review. *Am J Med Genet* 2006; 140: 402-409.
55. Motoyama O, Tokuyama M, Uga N et al. Urinary retention in a boy with terminal deletion of chromosome 10q at band 26.1. *Pediatr Int* 2008; 50: 603-604.
56. Piccione M, Antona V, Piro E et al. 10qter deletion: a new case. *Am J Med Genet* 2008; 146A: 2435-2438.
57. Yatsenko SA, Kruer MC, Bader PI et al. Identification of critical regions for clinical features of distal 10q deletion syndrome. *Clin Genet* 2009; 76: 54-62.
58. Wu YC, Horvitz HR. *C. elegans* phagocytosis and cell-migration protein CED-5 is similar to human DOCK180. *Nature* 1998; 392: 501-504.
59. Park D, Tosello-Trampont A-C, Elliott MR et al. BAI1 is an engulfment receptor for apoptotic cells upstream of the ELMO/Dock180/Rac module. *Nature* 2007; 450: 430-434.
60. Tachibana M, Kiyokawa E, Hara S et al. Ankyrin repeat domain 28 (ANKRD28), a novel binding partner of DOCK180, promotes cell migration by regulating focal adhesion formation. *Exp Cell Res* 2009; 315: 863-876.

61. Hasegawa H, Kiyokawa E, Tanaka S et al. DOCK180, a major CRK-binding protein, alters cell morphology upon translocation to the cell membrane. *Mol Cell Biol* 1996; 16: 1770-1776.
62. Howard HC, Mount DB, Rochefort D et al. The K-Cl cotransporter KCC3 is mutant in a severe peripheral neuropathy associated with agenesis of the corpus callosum. *Nat Genet* 2002; 32: 384-392.
63. Autio S, Pihko H, Tengström C. Clinical features in a de novo interstitial deletion 15q13 to q15. *Clin Genet* 1988; 34: 293-298.
64. Galán F, Aguilar MS, González J et al. Interstitial 15q deletion without a classic Prader-Willi phenotype. *Am J Med Genet* 1991; 38: 532-534.
65. Erdogan F, Ullmann R, Chen W et al. Characterization of a 5.3 Mb deletion in 15q14 by comparative genomic hybridization using a whole genome "tiling path" BAC array in a girl with heart defect, cleft palate, and developmental delay. *Am J Med Genet* 2007; 143: 172-178.
66. Chen C-P, Lin S-P, Tsai F-J et al. A 5.6-Mb deletion in 15q14 in a boy with speech and language disorder, cleft palate, epilepsy, a ventricular septal defect, mental retardation and developmental delay. *Eur J Med Genet* 2008; 51: 368-372.
67. Brunetti-Pierri N, Sahoo T, Frioux S et al. 15q13q14 deletions: phenotypic characterization and molecular delineation by comparative genomic hybridization. *Am J Med Genet* 2008; 146A: 1933-1941.
68. Pivnick EK, Qumsiyeh MB, Tharapel AT et al. Partial duplication of the long arm of chromosome 6: a clinically recognisable syndrome. *J Med Genet* 1990; 27: 523-526.
69. Henegariu O, Heerema NA, Vance GH. Mild "duplication 6q syndrome": a case with partial trisomy (6)(q23.3q25.3). *Am J Med Genet* 1997; 68: 450-454.
70. Cappon SL, Duncan AM, Khalifa MM. Interstitial 6q duplication in an adult male without growth delay or severe mental retardation. *Med Sci Monit* 2000; 6: 581-585.
71. Goh DL, Tan AS, Chen JY et al. Dysmorphic sibs trisomic for the region 6q22.1-->6q23.3. *J Med Genet* 2000; 37: 889-892.
72. Causio F, Fischetto, Carnevale F et al. Trisomy 6q syndrome: a case with a << de novo >> 6q23 tandem duplication. *Genet Couns* 2001; 12: 145-150.

73. Stamberg J, Shapiro J, Valle D et al. Partial trisomy 6q, due to balanced maternal translocation (6;22) (q21; p13) or (q21; pter). *Clin Genet* 1981; 19: 122-125.
74. Erdel M, Duba HC, Verdorfer I et al. Comparative genomic hybridization reveals a partial de novo trisomy 6q23-qter in an infant with congenital malformations: delineation of the phenotype. *Hum Genet* 1997; 99: 596-601.
75. Brewer C, Holloway S, Zawalnyski P et al. A chromosomal duplication map of malformations: regions of suspected haplo- and triplolethality--and tolerance of segmental aneuploidy--in humans. *Am J Hum Genet* 1999; 64: 1702-1708.
76. Brooks SP, Coccia M, Tang HR et al. The Nance-Horan Syndrome protein encodes a functional WAVE Homology Domain (WHD) and is important for co-ordinating actin remodelling and maintaining cell morphology. *Hum Mol Genet* 2010; 19: 2421-2432.
77. Burdon KP, McKay JD, Sale MM et al. Mutations in a novel gene, NHS, cause the pleiotropic effects of Nance-Horan syndrome, including severe congenital cataract, dental anomalies, and mental retardation. *Am J Hum Genet* 2003; 73: 1120-1130.
78. Brooks SP, Ebenezer ND, Poopalasundaram S et al. Identification of the gene for Nance-Horan syndrome (NHS). *J Med Genet* 2004; 41: 768-771.
79. Pinto D, Marshall C, Feuk L et al. Copy-number variation in control population cohorts. *Hum Mol Genet* 2007; 16 Spec No. 2: R168-173.
80. Nogami S, Satoh S, Tanaka-Nakadate S et al. Identification and characterization of taxilin isoforms. *Biochem Biophys Res Commun* 2004; 319: 936-943.
81. Fujiwara T, Mishima T, Kofuji T et al. Analysis of knock-out mice to determine the role of HPC-1/syntaxin 1A in expressing synaptic plasticity. *J Neurosci* 2006; 26: 5767-5776.
82. Spalice A, Parisi P, Nicita F et al. Neuronal migration disorders: clinical, neuroradiologic and genetics aspects. *Acta Paediatr* 2009; 98: 421-433.
83. Verrotti A, Spalice A, Ursitti F et al. New trends in neuronal migration disorders. *Eur J Paediatr Neurol* 2010; 14: 1-12.
84. Velagaleti GV, Morgan DL, Tonk VS. Trisomy 5p. A case report and review. *Ann Genet* 2000; 43: 143-145.
85. Reichenbach H, Holland H, Dalitz E et al. De novo complete trisomy 5p: clinical report and FISH studies. *Am J Med Genet* 1999; 85: 447-451.

86. Grosso S, Cioni M, Garibaldi G et al. De novo complete trisomy 5p: clinical and neuroradiological findings. *Am J Med Genet* 2002; 112: 56-60.
87. Kleczkowska A, Fryns JP, Moerman P et al. Trisomy of the short arm of chromosome 5: autopsy data in a malformed newborn with inv dup (5) (p13.1----p15.3). *Clin Genet* 1987; 32: 49-56.
88. Gustavson KH, Lundberg PO, Nicol P. Familial partial trisomy 5p resulting from segregation of an insertional translocation. *Clin Genet* 1988; 33: 404-409.
89. Chia NL, Bousfield LR, Johnson BH. A case report of a de novo tandem duplication (5p) (p14----pter). *Clin Genet* 1987; 31: 65-69.
90. Baialardo EM, Torrado MdV, Barreiro CZ et al. Partial distal 5p trisomy resulting from paternal translocation (5;15)(p15.1;p13) in a boy with no mental retardation. *Clin Dysmorphol* 2003; 12: 257-259.
91. Riordan D, Vust A, Wickstrom DE et al. Identification of a dup(5)(p15.3) by multicolor banding. *Clin Genet* 2002; 61: 277-282.
92. Sheen VL, Wheless JW, Bodell A et al. Periventricular heterotopia associated with chromosome 5p anomalies. *Neurology* 2003; 60: 1033-1036.
93. van Dijk M, Mulders J, Könst A et al. Differential downregulation of alphaT-catenin expression in placenta: trophoblast cell type-dependent imprinting of the CTNNA3 gene. *Gene Expr Patterns* 2004; 5: 61-65.
94. Busby V, Goossens S, Nowotny P et al. Alpha-T-catenin is expressed in human brain and interacts with the Wnt signaling pathway but is not responsible for linkage to chromosome 10 in Alzheimer's disease. *Neuromolecular Med* 2004; 5: 133-146.
95. Lien W-H, Klezovitch O, Fernandez TE et al. alphaE-catenin controls cerebral cortical size by regulating the hedgehog signaling pathway. *Science* 2006; 311: 1609-1612.
96. Abe K, Chisaka O, Van Roy F et al. Stability of dendritic spines and synaptic contacts is controlled by alpha N-catenin. *Nat Neurosci* 2004; 7: 357-363.
97. Park C, Falls W, Finger JH et al. Deletion in *Catna2*, encoding alpha N-catenin, causes cerebellar and hippocampal lamination defects and impaired startle modulation. *Nat Genet* 2002; 31: 279-284.
98. Bradley WEC, Raelson JV, Dubois DY et al. Hotspots of large rare deletions in the human genome. *PLoS ONE* 2010; 5: e9401.

4 SNP ARRAYS: COMPARING DIAGNOSTIC YIELDS FOR FOUR PLATFORMS IN CHILDREN WITH DEVELOPMENTAL DELAY

Manuscrit prêt pour soumission à Human Genetics (Springer)

Guylaine D'Amours^{a,b,c}, Géraldine Mathonnet^a, Raouf Fetni^{b,c,d,e}, Sonia Nizard^{a,c,f}, Myriam Srour^b, Frédérique Tihy^{a,b,c,e}, Jacques L. Michaud^{a,b,c,f}, Emmanuelle Lemyre^{a,b,c,f}

^a Service de génétique médicale, CHU Sainte-Justine, Montréal, QC, Canada, ^b Centre de recherche, CHU Sainte-Justine, Montréal, QC, Canada, ^c Faculté de médecine, Université de Montréal, Montréal, QC, Canada, ^d Département de pathologie, CHU Sainte-Justine, Montréal, QC, Canada, ^e Pathologie et biologie cellulaire, Université de Montréal, Montréal, QC, Canada, ^f Pédiatrie, Université de Montréal, Montréal, QC, Canada

Corresponding author

Emmanuelle Lemyre, MD, FRCPC, FCCMG

Service de génétique médicale

CHU Sainte-Justine

3175 Chemin de la Côte-Sainte-Catherine

Montréal, QC, Canada H3T 1C5

Tel: 1 (514) 345-4931 ext. 6342

Fax: 1 (514) 345-4766

4.1 Abstract

Molecular karyotyping is now the first-tier genetic test for patients affected with unexplained intellectual disability and/or multiple congenital malformations, since it identifies a pathogenic CNV in 10-14% of them. High-resolution microarrays combining molecular karyotyping and SNP genotyping were recently introduced to the market. In addition to identifying CNVs, these platforms detect LOHs, which can indicate the presence of a homozygous mutation or of uniparental disomy. Since these abnormalities can be associated with intellectual disability or congenital abnormalities, their detection is of particular interest for patients whose phenotype remains unexplained. However, the diagnostic yield obtained with these platforms is not confirmed, and the real clinical value of LOH detection is not yet established. We tested 21 children affected with intellectual disability for whom standard genetic analyses failed to provide a diagnosis, and identified a pathogenic CNV in three patients, and an informative LOH in four (including two siblings). A homozygous mutation within the informative LOH was found in three of these patients. Therefore, we were able to increase the diagnostic yield from 14.3% to 28.6% as a result of the information provided by LOHs. This study shows the clinical usefulness of SNP arrays in children with intellectual disability.

Key Words

Array comparative genomic hybridization (aCGH), DNA copy number variation (CNV), Intellectual disability, Loss of heterozygosity (LOH), Microarray analysis, Single nucleotide polymorphism (SNP).

4.2 Introduction

Although intellectual disability (ID) and developmental delay (DD) affect 3% of children (Goldenberg and Saugier-Veber 2010), etiology remains elusive in close to 50% of cases (Curry et al. 1997; van Karnebeek et al. 2005), and a diagnosis cannot always be provided to patients and their families, even when a genetic cause is suspected. Many chromosomal aberrations have been associated with syndromic and non-syndromic ID; hence, G-banded karyotyping was, for over 30 years, the first-line genetic analysis for patients with unexplained ID and/or multiple congenital anomalies (MCA). Later, fluorescence *in situ* hybridization was added to the diagnostic arsenal, at first to test patients for whom the clinician suspected a specific microdeletion syndrome, and then to detect subtelomeric rearrangements, which can be found in 3-6% of patients with idiopathic ID (Ledbetter and Martin 2007). Technological advances in the last decade have changed this testing paradigm, and molecular karyotyping is now the recommended first-tier test for these patients (Manning et al. 2010; Miller et al. 2010; Stavropoulos et al. 2010), as it detects unbalanced chromosomal aberrations at a much higher resolution than conventional karyotype, thus allowing the detection of microdeletions and microduplications without the need to target a specific genomic region.

Initially, molecular karyotyping was performed using BACs in an array-based comparative genomic hybridization (aCGH). Then, oligonucleotides arrays allowed an increase in resolution and a better customization of the tested genomic regions. Array CGH led to the identification of new rare and recurrent submicroscopic pathogenic chromosomal aberrations (Koolen et al. 2008; Mefford et al. 2008; Sharp et al. 2006; Sharp et al. 2008), increasing the diagnostic yield in patients with unexplained ID and/or MCA from 3.7% (G-banded karyotype) (Shevell et al. 2003) to 10-14% (molecular karyotype) (Cooper et al. 2011; Miller et al. 2010; Sagoo et al. 2009). Conversely, an important number of chromosomal aberrations that are thought to be benign, or for which clinical significance currently cannot be determined, are being identified (Sagoo et al. 2009), requiring cytogenetics laboratories to store and share genotype-phenotype data in order to provide the most accurate clinical interpretation for each result.

Although high-resolution SNP arrays have been used extensively in genome-wide association studies (Heinzen et al. 2010; Weiss et al. 2008; Xu et al. 2008), cytogenetics laboratories have only recently begun using them in the clinical setting, and available reports are restricted to their use for CNV detection (Filges et al. 2012; Qiao et al. 2012; Srebniak et al. 2011). However, the SNP genotyping information provided by these arrays also allows loss of heterozygosity (LOH) detection, which can be the result of a deletion, autozygosity (identity-by-descent) or uniparental isodisomy (UPD). In a clinical setting, this is of interest because autozygous regions may either: 1) harbour disease-causing homozygous mutations, as illustrated by the higher rate of MCA, ID, and other serious medical conditions in consanguineous families (Madhavan and Narayan 1991; Sridhara Rama Rao and Narayanan 1976; Zlotogora and Shalev 2010); or 2) signal UPD, whether segmental or affecting a whole chromosome, which can cause syndromic conditions when it involves imprinted genes (Butler 2009). This extends the scope of molecular karyotyping beyond chromosomal aberrations detection, to include the detection of pathogenic genomic abnormalities that would usually require separate molecular analyses.

In this study, we test children affected with ID/DD on four high-resolution SNP arrays, and compare the results with the moderate-resolution oligo-array currently used in our laboratory. Our goal is to evaluate, in the clinical setting, the performance and workload of each platform, and to determine whether there is an additional value in using one of these arrays to test patients with ID and/or MCA. We compare the proportion of pathogenic and benign CNVs, as well as variants of uncertain clinical significance (VOUS); examine LOH larger than 5 Mb; and compare the characteristics of each vendor's provided visualization software.

4.3 Material & methods

4.3.1 Patients

Twenty-one patients, between the age of six months and 17 years old, were recruited consecutively upon evaluation by a clinical geneticist at the Centre Hospitalier Universitaire Sainte-Justine. Inclusion criteria were as follow: 1) the child presented DD (n=13), ID (n=5), or both (n=3); 2) the geneticist ordered aCGH as part of the clinical investigation; and 3) both biological parents were available for testing. A subset of children also had autism (n=3), or pervasive developmental delay (n=2), or presented dysmorphic features (n=3), malformation(s) (n=3), or both (n=7). Phenotypic details are available in Table 4-1. Patients were excluded from the study if a single-gene disorder seemed most probable, or if a chromosomal rearrangement (balanced or unbalanced) had previously been detected. Child's assent (if possible) and parents' informed consent were obtained, and phenotypic information was collected both for the children and their parents. This study was approved by the CHU Sainte-Justine's Research Ethics Board.

Table 4-1 : Phenotype information of patients included in the study.

Patient	Sex	Age^a	Diagnosis^b	Dysmorphism	Malformation(s)	Other
1.1	F	12 y	MR	Yes	Multiple: multicystic dysplastic kidney (left), ventriculomegaly, temporal white matter loss, pectus excavatum	Deafness, myopia, nystagmus, dysphagia, mild scoliosis
3.2	M	6 mo	Growth delay	Yes	Cerebral: microcephaly, gyral simplification, myelinization delay	Spastic quadriparesis
6.3	M	11 y	PDD, mild MR	Yes	Mild: pectus excavatum	Café au lait spots. Mother: school difficulties, epilepsy, murmur, arthrosis
10.4	M	7 y	Global DD, PDD	No	No	Hypotonia, café au lait spots, pes planus
11.4	M	13 y	Global DD	No	No	Simple febrile convulsions, hypotonia, diminished deep tendon reflexes
14.5	F	14 y	Mild MR, growth delay	Yes	Multiple: microcephaly, thick corpus callosum, malocclusion, filum terminale lipoma, sacral agenesis	Feeding difficulties
17.6	M	3 y	DD, autism	No	No	-
20.7	M	18 mo	Speech delay	Yes	Craniofacial: trigonocephaly, labiopalatine cleft	Heterochromia, facial asymmetry
23.8	M	13 y	DD, mild MR	No	Multiple: cerebellar atrophy, malocclusion, valgus feet	Ataxia
26.9	M	5 y	Global DD, mild MR	Yes	No	Obesity, ataxia, buccolingual dyspraxia, increased lactates (blood & LCR)
29.10	M	17 y	MR, autism	Yes	No	Obesity, retinitis pigmentosa
32.11	M	2 y	DD	Yes	Cerebral: microcephaly, pons size slightly reduced, 4 th ventricle size slightly increased	Short stature, epilepsy, gastroesophageal reflux

Patient	Sex	Age^a	Diagnosis^b	Dysmorphism	Malformation(s)	Other
35.12	M	17 mo	Global DD	Yes	Multiple: optic atrophy, thin corpus callosum, cerebral atrophy, myelination delay	Epilepsy, hypotonia, mild limb spasticity
38.13	F	6 y	MR	No	No	-
41.14	M	22 mo	Global DD	Yes	No	Ligament hypermobility. Father: learning disability
44.15	M	4 y	DD, autism	No	Craniofacial: submucosal palatine cleft	Velopharyngeal insufficiency
47.16	M	3 y	Global DD	No	Pulmonary valve stenosis	Gastroesophageal reflux
50.17	F	10 y	Global DD	No	No	Epilepsy
53.18	M	5 y	Global DD	No	No	Tall stature
56.19	F	11 y	Mild MR	No	No	Strabismus
59.20	F	34 mo	Atypical development	No	No	Mild facial asymmetry

^a y: years, mo: months

^b MR: mental retardation, PDD: pervasive developmental disorder, DD: developmental delay

4.3.2 Samples

Peripheral blood samples were drawn from each individual and kept at -20°C until DNA purification. Purified genomic DNA was obtained using the QIAamp DNA Blood Midi Kit (QIAGEN, Mississauga, ON, Canada) according to manufacturer's instructions and stored at 4°C. Absorbance measurements were performed with a NanoDrop 1000 spectrophotometer (Thermo Scientific, Wilmington, DE, USA) to verify DNA purity and adjust DNA concentration. Genomic DNA length was controlled for each DNA sample by running 50 ng of DNA diluted in 10 uL of TE buffer (QIAGEN) on a 1.2% agarose E-Gel (Invitrogen, Carlsbad, CA, USA).

4.3.3 Array genomic hybridization

DNA samples were sent to an external laboratory (Centre de Pharmacogénomique Beaulieu-Saucier, Montréal, Qc, Canada) for array genomic hybridization, which was performed according to manufacturer's instruction with the following four arrays: Affymetrix Genome-Wide Human SNP Array 6.0 (SNP 6.0) and Cytogenetics Whole-Genome 2.7M array (2.7M) (Affymetrix, Santa Clara, CA, USA), and Illumina HumanOmni1-Quad BeadChip (Omni1) and HumanCytoSNP-12 DNA Analysis BeadChip (CytoSNP) (Illumina, San Diego, CA, USA). Samples were also analysed with the NimbleGen CGX-12 Cytogenetics Array (CGX-12) (Roche NimbleGen, Madison, WI, USA) for all patients, as previously described (D'Amours et al. 2012). This CGH array contains 135 000 oligonucleotides targeting 675 genes and more than 200 important cytogenetic regions with an average spacing of 65-75 kb throughout the genome, and 10 kb in targeted regions. Parents were tested with this array only if required for CNV interpretation (n=2).

4.3.4 Visualization software settings

CNV and LOH were visualized in the software provided by each vendor: Affymetrix Chromosome Analysis Suite 1.1 (ChAS) for SNP 6.0 and 2.7M, Illumina KaryoStudio 1.3.1

for Omni1 and CytoSNP, and Signature Genomics Genoglyphix for CGX-12. For CGX-12, we used the thresholds currently in place in the clinical laboratory, which are as follows: imbalances spanning a minimum of five oligonucleotides, with a mean \log_2 ratio above 0.4 for gains or below -0.7 for losses. For Affymetrix and Illumina arrays, confidence was set at 80% for 2.7M (SNP 6.0 does not offer a confidence statistic) and 50% for the Illumina arrays. In ChAS, we used the following configuration: Smoothing enabled and Joining enabled (number of markers = 1). There is no such custom configuration offered in KaryoStudio. For CNV visualization, we used two different sets of thresholds, a standard set and an optimized set (Table 4-2). The standard set used parameters recommended by the manufacturer, and we anticipated that this approach would lead to a large number of CNVs. To reduce the number of CNVs to interpret, we explored how we could use the various filters available in both visualization softwares. In addition, we found that using a standard size threshold of 50 kb would miss smaller pathogenic CNVs as those found in our lab and in the literature (Boone et al. 2010; Bruno et al. 2011a). Our goals with the use of optimized thresholds were: 1) to reduce the total number of CNVs to interpret, while preserving potentially pathogenic alterations; 2) decrease the size threshold in order to detect small pathogenic CNVs; and 3) remove known benign CNVs. We accomplished this by: 1) increasing thresholds in genomic regions exempt of genes of known clinical relevance; 2) decreasing thresholds in regions with genes of known clinical relevance; and 3) filtering out known benign CNVs. We used the adjustable parameters in both softwares to perform steps 1 and 2, by applying different thresholds to regions with genes catalogued in OMIM Morbid Map (Online Mendelian Inheritance in Man Morbid Map, Feb 2009 hg19 assembly, accessed October 28, 2010) and regions without such genes. For the third step, we took advantage of a feature available in ChAS, which allows masking CNVs with a certain % of overlap with specific genomic regions, to filter out CNVs without genes of known clinical relevance that also had a 100% overlap with CNVs reported in DGV (Database of Genomic Variants, Build 36(hg18), released July 2009). In KaryoStudio, this filter was not used since it was not possible to apply it only in certain regions, excluding those overlapping OMIM genes. Both sets of thresholds are detailed in Table 4-2, as well as thresholds used for long contiguous stretches of homozygosity (LCSH) visualization.

Table 4-2 : ChAS and KaryoStudio thresholds used for CNV and LCSH visualization.

	Parameter	SNP 6.0	2.7M	Omni1	CytoSNP
CNV (standard)	Size	≥ 50 Kb	≥ 50 Kb	≥ 50 Kb	≥ 50 Kb
	Marker count ^a	≥ 25	≥ 25	≥ 7	≥ 7
CNV (optimized)	Size <i>Inside OMIM</i> ^b	≥ 10 Kb	≥ 10 Kb (loss) ≥ 50 Kb (gain)	≥ 10 Kb	≥ 10 Kb
		<i>Outside OMIM</i>	≥ 100 Kb	≥ 200 Kb	≥ 100 Kb
	Marker count <i>Inside OMIM</i>	≥ 10	≥ 10	≥ 5	≥ 5
		<i>Outside OMIM</i>	≥ 25	≥ 25	≥ 7
	DGV ^c <i>Outside OMIM</i>	Removed	Removed	N/A	N/A
LCSH	Size	≥ 5 Mb	≥ 5 Mb	≥ 5 Mb	≥ 5 Mb
	Marker count	≥ 100	≥ 100	≥ 20	≥ 20

N/A: non-applicable

^a Abnormal markers within the CNV or LCSH segment

^b *OMIM* : Regions containing genes of known clinical relevance from the NCBI *Online Mendelian Inheritance in Man* Morbid Map (February 2009 hg19 assembly, accessed October 28, 2010).

^c CNV completely overlapping with a CNV reported in *Database of Genomic Variants* (NetAffx Build 30.2)

4.3.5 CNV interpretation

Inheritance for detected CNVs was determined by visualizing a child and his parents' results side by side in the software, assigning an inherited status to a child's CNV if there was $\geq 80\%$ overlap with a parental CNV. Due to the high number of detected CNVs by some platforms using standard thresholds (see Table 4-2), we selected CNVs with a high probability of being pathogenic for the analysis with optimized thresholds: *de novo* CNVs, homozygous CNVs inherited from heterozygous parents, and maternally inherited CNVs on chromosome X found in boys. There were only 14 CNVs in the latter two categories, which are later referred to as *de novo* CNVs. Therefore, inherited CNVs do not include these. Each *de novo* CNV was then classified into one of three categories, based on its clinical significance: benign, pathogenic, or VOUS. Clinical significance was established by comparing each unbalanced genomic region to information from public & private databases (DGV, OMIM, and Genoglyphix Chromosome Aberration Database, GCAD (Neill et al. 2010)). CNVs reported in DGV or overlapping segmental duplications were considered benign. For CGX-12, CNVs previously reported in at least two healthy individuals from our internal database or GCAD were also considered benign. CNVs larger than 5 Mb or involving genomic regions of known clinical relevance (OMIM Morbid Map) compatible with the patient's phenotype were classified as pathogenic. Remaining CNVs were classified as VOUS.

4.3.6 LOH interpretation

For LOH interpretation, as homozygous regions can indicate hemizyosity, autozygosity (identity-by-descent), or UPD, which can suggest a recessive or imprinting disorder that would require confirmation by molecular analyses, we excluded LOHs resulting from a deletion and only interpreted LCSHs. For UPD detection, we looked for autosomal LCSHs spanning a whole chromosome, which would reveal uniparental isodisomy (iUPD), and autosomal LCSHs larger than 10 Mb, which could indicate a mix of uniparental isodisomy and heterodisomy, or segmental UPD (Papenhausen et al. 2011). For UPD detection purposes, we excluded large LCSHs present in patients who also had large LCSHs (> 10 Mb) in multiple chromosomes, as this is suggestive of parental consanguinity rather than UPD. For total

homozygosity calculations, we only considered autosomes (Kearney et al. 2011a). To identify LCSHs potentially indicative of a homozygous mutation, we studied LCSHs larger than 5 Mb, and compared gene content with information from the OMIM database. We considered a LCSH informative if it contained a gene associated with a recessive disorder consistent with the patient's phenotype. Other LCSHs were considered uninformative.

4.4 Results

The goal of this study was to evaluate, in a clinical context, the diagnostic performance of SNP arrays combining CNV detection and SNP genotyping information. Twenty-one patients were tested on four different platforms from two manufacturers (Affymetrix SNP 6.0 and 2.7M arrays; and Illumina Omni1 and CytoSNP arrays), and their results were compared with those obtained on the NimbleGen CGX-12 array, currently in use at our laboratory. In order to do this, we optimized the CNV detection thresholds for clinical application, and interpreted all *de novo* CNVs, as well as regions of homozygosity that could be of clinical interest. We also compared each visualization software's features, ease of use, and efficiency in a clinical setting.

4.4.1 Detected CNVs: standard thresholds

Using standard thresholds (Table 4-2), SNP 6.0 and 2.7M from Affymetrix respectively detected 503 and 368 CNVs for all 21 patients, while Omni1 and CytoSNP from Illumina detected 197 and 30 CNVs, respectively (Table 4-S1). The ratio of copy number gains versus copy number losses was close to 1:1 for Omni1 and CytoSNP (89:108, and 16:14, respectively), while it was roughly 6:1 for SNP 6.0 and 2.7M (430:73, and 314:54, respectively). The proportion of *de novo* CNVs detected varied from 17.3% (34/197) for Omni1 to 47.8% (176/368) for 2.7M. CytoSNP detected the lowest number of CNVs per patient, with an average of 1 (range 0-5), while SNP 6.0 detected the highest, with an average of 24 CNVs per patient (range 14-37). Overall, both Affymetrix arrays detected a very high number of CNVs, the majority (85%) being gains, while Illumina arrays detected less CNVs, and gains and losses were more balanced.

4.4.2 Detected CNVs: optimized thresholds

The approach with optimized thresholds resulted in an important reduction in the total number of detected CNVs by the Affymetrix arrays, while numbers remained relatively

similar for both Illumina arrays (Tables 4-S1 & 4-S2). The proportion of *de novo* CNVs detected also remained similar for the Illumina arrays, but increased from 21.7% (109/503) to 55.6% (139/250) for SNP 6.0, and from 47.8% (176/503) to 69.4% (150/216) for 2.7M. These differences in the overall number and the proportion of *de novo* CNVs were expected, since we removed benign CNVs reported in DGV (which would mostly be inherited) for the Affymetrix arrays, but we were technically unable to do so with KaryoStudio. The ratio of gains vs losses decreased from 6:1 to about 4:1 for Affymetrix arrays, while it slightly increased from 1:1 to 1.7:1 for CytoSNP, and remained unchanged at around 1:1 for Omni1. The average number of CNVs per patient was cut in half for Affymetrix arrays: from 24 to 12 for SNP 6.0 (range 4-30) and from 18 to 10 for 2.7M (range 1-32). There was a slight decrease in the average number of CNVs per patient for Omni1, which went from 10 to 9 (range 5-13) and there was no change for CytoSNP (1 CNV/patient, range 0-5). In comparison, CGX-12 detected 58 CNVs for all 21 patients, for an average of 3 CNVs per patient (range 0-7), and a ratio of gains vs losses of 0.9:1 (27:31). Inherited vs *de novo* proportions cannot be compared with the CGX-12 array because parents were only tested if required for CNV interpretation. Optimized thresholds resulted, as expected, in a reduction of the gains vs losses ratio and of the overall number of CNVs detected by the Affymetrix arrays, while it had almost no impact on the Illumina arrays.

4.4.3 CNV validation

In order to confirm detected CNVs, we compared results obtained with the optimized thresholds from each array, and considered a CNV validated if it was detected on at least two platforms with > 50% overlap. The list of confirmed CNVs is available in Table 4-S3. The confirmation rate was very low for most arrays except CytoSNP, for which 96.0% of the detected CNVs were also seen on at least another array (Table 4-3). In comparison, NimbleGen CGX-12 array had a confirmation rate of 46.6%. There was an important difference in the confirmation rate of gains vs losses on the SNP 6.0 and 2.7M arrays (13.9% vs 40.8%, and 7.6% vs 25.0%, respectively), while Omni1, CytoSNP and CGX-12 arrays showed a similar confirmation rate for gains and losses. Also, the confirmation rate was higher

for inherited CNVs than for *de novo* CNVs for SNP 6.0, 2.7M and CytoSNP (27.9% vs 12.2%, 19.7% vs 7.3%, and 100% vs 87.5%, respectively) while it was similar for Omni1 (30.0% of inherited vs 34.3% of *de novo*). In order to verify that the unconfirmed CNVs were not due to an inadequate detection algorithm or lack of coverage on the other arrays, we also examined, on all arrays, the signal of unconfirmed *de novo* CNVs that were later classified as pathogenic or VOUS. None of those CNVs had a detectable signal on another array, and the majority (80.9%) had adequate coverage on at least one additional array (data not shown). In summary, very few gains detected on the Affymetrix arrays were confirmed, while there was no such bias with the Illumina or the NimbleGen arrays. Also, the confirmation rate of all arrays, except Omni1, was higher for inherited CNVs than for *de novo* CNVs.

Table 4-3 : Number of confirmed CNVs and confirmation rate for each tested array.

Confirmed CNVs	SNP 6.0	2.7M	Omni1	CytoSNP	CGX-12	All arrays
Total number (%)^a	48 (19.2%)	24 (11.1%)	57 (30.8%)	24 (96.0%)	27 (46.6%)	180 (72.0%)
Gains (%)^b	28 (13.9%)	13 (7.6%)	30 (38.5%)	15 (93.8%)	14 (51.9%)	100 (20.2%)
Losses (%)^c	20 (40.8%)	11 (25.0%)	27 (25.2%)	9 (100.0%)	13 (41.9%)	80 (33.3%)
<i>De novo</i> (%)^d	17 (12.2%)	11 (7.3%)	12 (34.3%)	7 (87.5%)	25^e (44.6%)	72 (51.8%)
Inherited (%)^f	31 (27.9%)	13 (19.7%)	45 (30.0%)	17 (100.0%)	2 (100.0%)	108 (77.7%)

^a Confirmation rate = number of confirmed CNVs divided by number of detected CNVs by each array, expressed in %

^b Confirmation rate = number of confirmed gains divided by number of detected gains by each array, expressed in %

^c Confirmation rate = number of confirmed losses divided by number of detected losses by each array, expressed in %

^d Confirmation rate = number of confirmed *de novo* CNVs divided by number of detected *de novo* CNVs by each array, expressed in %

^e De novo or inheritance unknown (parents not systematically tested)

^f Confirmation rate = number of confirmed inherited CNVs divided by number of detected inherited CNVs by each array, expressed in %

4.4.4 Clinical significance of detected *de novo* CNVs

We interpreted all detected *de novo* CNVs (confirmed or not) on each platform individually, using the manufacturer's accompanying visualization software, and assigned each CNV one of three categories: pathogenic, benign, or VOUS. In order to have a true assessment of each array's performance in a clinical setting, we classified each CNV independently from the information gained on the other arrays. There was a total of 332 *de novo* CNVs detected on any of the four tested arrays for all 21 patients (Table 4-S2). CytoSNP detected 8 *de novo* CNVs, Omni1 detected 35, SNP 6.0 detected 139 and 2.7M detected 150 (Table 4-4). Seventeen of these were false positives that should not have been called or were inherited but the software did not call the parental CNV.

There were some concerns about the accuracy of the results obtained with the 2.7M platform version that we tested. Indeed, an extremely high proportion (139/150, 92.7%) of *de novo* CNVs were VOUS, which was unexpected even considering the high density of this array. In addition, only one of the 150 CNVs was benign, despite extensive coverage of polymorphic regions and segmental duplications, and only 24 of the 216 CNVs detected (inherited and *de novo*) were confirmed on another array. Therefore, we suspect a design or technical defect affecting CNV detection was responsible for these results. In order to have a meaningful comparison and for clarity purposes, we only report the results obtained with the other three arrays in the following section, Figure 4-1 and Figure 4-2. Information regarding *de novo* CNVs detected by 2.7M is included in all the other sections and in all tables.

Table 4-4 : Clinical significance of detected de novo CNVs.

De novo CNVs	SNP 6.0	2.7M	Omni	CytoSNP	CGX-12^a	All arrays
Total	139	150	35	8	56	388
Gains	108	121	7	4	25	265
Losses	31	29	28	4	31	123
Range per patient	0-17	0-20	0-4	0-2	0-7	10-30
Median per patient	7	6	2	0	3	16
Average per patient	6.6	7.1	1.7	0.4	2.7	18.5
Benign (% of total)	36 (25.9%)	1 (0.7%)	14 (40.0%)	1 (12.5%)	51 (91.1%)	103 (26.5%)
Gains	22	0	3	1	23	49
Losses	14	1	11	0	28	54
Range per patient	0-3	0-1	0-4	0-1	0-7	2-13
Median per patient	2	0	0	0	3	5
Average per patient	1.7	0.0	0.7	0.05	2.4	4.9
Nb of patients (%)	17 (81.0%)	1 (4.8%)	7 (33.3%)	1 (4.8%)	20 (95.2%)	21 (100.0%)
Unclear (% of total)	89 (64.0%)	139 (92.7%)	10 (28.6%)	3 (37.5%)	1 (1.8%)	242 (62.4%)
Gains	75	115	1	2	1	194
Losses	14	24	9	1	0	48
Range per patient	0-15	0-19	0-2	0-1	0-1	5-24
Median per patient	4	6	0	0	0	9
Average per patient	4.2	6.6	0.5	0.1	0.05	11.5
Nb of patients (%)	20 (95.2%)	19 (90.5%)	7 (33.3%)	3 (14.3%)	1 (4.8%)	21 (100.0%)
Pathogenic (% of total)	6 (4.3%)	7 (4.7%)	5 (14.3%)	4 (50.0%)	4 (7.1%)	26 (6.7%)
Gains	3	4	2	1	1	11
Losses	3	3	3	3	3	15
Range per patient	0-4	1-2	0-3	0-2	0-2	1-13
Median per patient	0.5	1	0.5	0.5	0.5	3
Average per patient	1.0	1.2	0.8	0.7	0.7	4.3
Nb of patients (%)	3 (14.3%)	6 (28.6%)	3 (14.3%)	3 (14.3%)	3 (14.3%)	6 (28.6%)
False positive (% of total)	8 (5.8%)	3 (2.0%)	6 (17.1%)	0 (0.0%)	0 (0.0%)	17 (4.4%)
Gains	8	2	1	0	0	11
Losses	0	1	5	0	0	6
Range per patient	0-2	0-1	0-2	0	0	1-4
Median per patient	1	0	0	0	0	1
Average per patient	0.8	0.3	0.6	0.0	0.0	1.7
Nb of patients (%)	6 (28.6%)	3 (14.3%)	4 (19.0%)	0 (0.0%)	0 (0.0%)	10 (47.6%)

Nb: number

^a De novo or inheritance unknown (parents not systematically tested)

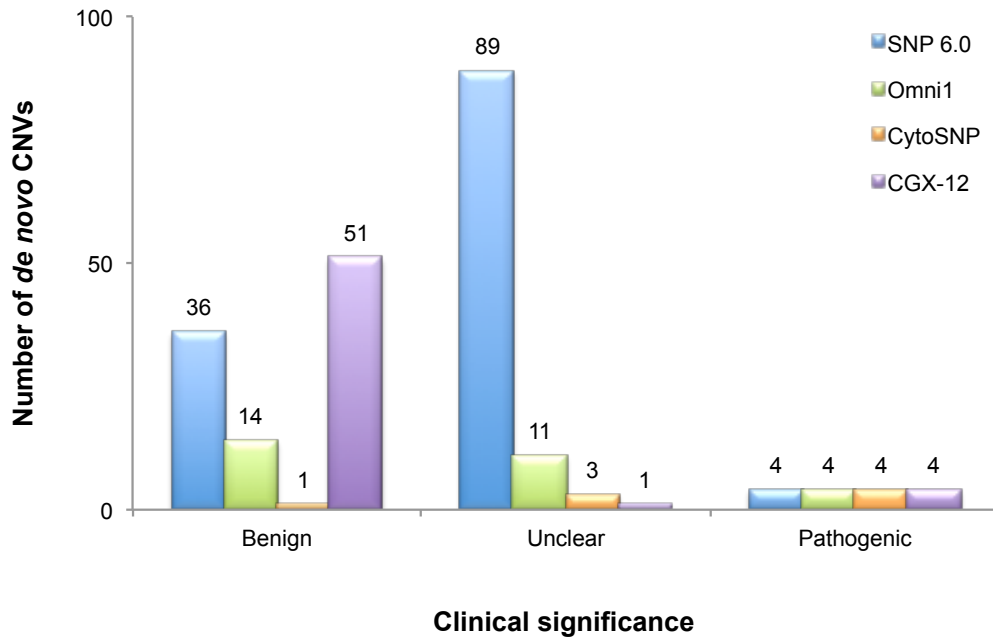


Figure 4-1 : Number of *de novo* CNVs detected by each array for all 21 patients, categorized according to clinical significance.

SNP 6.0 detected the highest number of benign CNV with 36 (Figure 4-1), for an average of 1.7 per patient, compared to 0.7 for Omni1 and 0.05 for CytoSNP (Table 4-4). If we consider the total number of *de novo* CNVs detected, Omni1 had the highest proportion of benign CNVs (40.0% (14/35) vs 25.9% (36/139) for SNP 6.0 and 12.5% (1/8) for CytoSNP). In comparison, 91.1% of CNVs detected on CGX-12 were benign (51/56), for an average of 2.4 per patient. But this was expected since parents are not systematically tested on this array, thus a high proportion of these CNVs were probably in fact inherited. SNP 6.0 also had the highest number of VOUS with 89 (vs 10 for Omni1 and 3 for CytoSNP), for an average of 4.2 per patient (vs 0.5 for Omni1 and 0.1 for CytoSNP). Correspondingly, the highest proportion of VOUS was found by SNP 6.0, with 64.0% (89/139), compared to 28.6% (10/35) for Omni1 and 37.5% (3/8) for CytoSNP. In comparison, CGX-12 detected only one VOUS out of 56 “*de novo*” CNVs (1.8%), for an average of 0.05 per patient. Most of the VOUS detected by SNP 6.0 were gains (75 gains vs 14 losses), while benign CNVs were more equally distributed (22 gains vs 14 losses). Omni1 showed the opposite trend, both for VOUS (1 gain vs 9 losses) and benign CNVs (3 gains vs 11 losses). Interestingly, only four of the 89 VOUS detected by SNP 6.0 were also detected on at least another array (Table 4-5), while a higher proportion of VOUS detected on CytoSNP and Omni1 were confirmed (2/3 and 4/10, respectively). In contrast, the confirmation rate of benign CNVs detected on SNP 6.0 was similar to the rate on Omni1 (7/36 confirmed for SNP 6.0, and 3/14 for Omni1), while it was 100% for CytoSNP (1/1). In addition, while size did not seem to impact the distribution of gains and losses for the benign CNVs (Figure 4-2 a), the bias towards gains for the VOUS detected on the SNP 6.0 array appeared to be inversely correlated with the size of the CNVs (Figure 4-2 b). In summary, the SNP 6.0 array detected a higher proportion of gains in the VOUS category, most of which were of small size, whereas there was no such bias for the benign CNVs. The CytoSNP array did not show a bias in either category, while the Omni1 array identified more losses than gains (both in the benign and VOUS categories). In addition, the confirmation rate of VOUS was lower for the SNP 6.0 array than for the Illumina arrays, while there was no difference in the confirmation rate of benign CNVs.

Table 4-5 : List of de novo CNVs confirmed on at least two arrays.

Patient	Alteration	Gain/ Loss	Array	CN	Chr	Chr band(s)	Coordinates		Size (kb)	Inheritance	Interpretation ^a
							Start	End			
3.2	1	Gain	SNP 6.0	3	2	p21	43848625	44069477	221	D	VOUS
			2.7M	3	2	p21	44001534	44065431	64	D	VOUS
			Omni1	3	2	p21	43844007	44248753	405	D	VOUS
			CytoSNP	3	2	p21	43884153	44225225	341	D	VOUS
10.4	10	Gain	SNP 6.0	3	14	q32.33	105082340	105867730	785	D	VOUS
			Omni1	3	14	q32.33	105446201	105855351	409	H	N/A
11.4	13	Gain	SNP 6.0	3	8	q11.1	47043274	47381308	338	H	N/A
			2.7M	3	8	q11.1	47042670	47597918	555	D	VOUS
14.5	14	Loss	SNP 6.0	1	1	p36.11	25477632	25535885	58	D	Benign
			Omni1	1	1	p36.11	25469887	25528125	58	H	N/A
	16	Gain	SNP 6.0	3	6	p25.3 p25.2	129591	2752759	2623	D	Pathogenic
			SNP 6.0	3	6	p25.2 p25.1	2779263	4643666	1864	D	Pathogenic
			SNP 6.0	3	6	p25.1	4646487	6416684	1770	D	Pathogenic
			2.7M	3	6	p25.3	109490	401749	292	D	Pathogenic
			2.7M	3	6	p25.3 p25.1	561662	6414958	5853	D	Pathogenic
			Omni1	3	6	p25.3	100815	610448	510	D	Pathogenic
			Omni1	3	6	p25.3 p25.2 p25.1	647434	6415891	5768	D	Pathogenic
	CytoSNP	3	6	p25.3 p25.2 p25.1	94609	6412865	6318	D	Pathogenic		
	CGX-12		6	p25.3 p25.1	128203	6399128	6270	U	Pathogenic		
	18	Loss	SNP 6.0	1	7	q36.3	154929838	158812469	3883	D	Pathogenic
			2.7M	1	7	q36.3	154962181	158811204	3849	D	Pathogenic
			Omni1	1	7	q36.3	154946906	158815293	3868	D	Pathogenic
CytoSNP			1	7	q36.3	154952401	158812247	3860	D	Pathogenic	
CGX-12				7	q36.3	154947683	158816094	3870	U	Pathogenic	
17.6	21	Gain	SNP 6.0	3	14	q32.33	105082340	105998615	916	D	VOUS
			Omni1	3	14	q32.33	105446084	105927451	481	H	N/A
20.7	26	Loss	SNP 6.0	1	19	q13.42	59994794	60055324	61	D	Benign
			Omni1	1	19	q13.42	59949314	60068550	119	D	Benign
23.8	27	Loss	SNP 6.0	1	15	q15.3	41680363	41726086	46	H	N/A
			2.7M	1	15	q15.3	41636500	41801729	165	D	Benign
			Omni1	1	15	q15.3	41670172	41735843	66	H	N/A
	28	Gain	SNP 6.0	3	22	q11.23	24035301	24246386	211	H	N/A
			Omni1	3	22	q11.23	23998784	24249492	251	H	N/A
CytoSNP	3	22	q11.23	24041667	24240879	199	D	Benign			
26.9	29	Gain	Omni1	3	17	q21.31	41519743	41713128	193	D	Benign
			CGX-12		17	q21.31	41541432	41631306	90	U	Benign

Patient	Alteration	Gain/ Loss	Array	CN	Chr	Chr band(s)	Coordinates		Size (kb)	Inheritance	Interpretation ^a
							Start	End			
32.11	35	Loss	Omni1	0	19	p12	20404485	20509513	105	H0	VOUS
			CytoSNP	0	19	p12	20398497	20514068	116	D	VOUS
			CGX-12		19	p12	20408868	20506175	97	U	Benign
38.13	39	Loss	SNP 6.0	1	6	p21.32	32589952	32626735	37	D	Benign
			Omni1	1	6	p21.32	32585397	32609122	24	H	N/A
41.14	42	Loss	SNP 6.0	1	1	p36.11	25455203	25519561	64	D	VOUS
			Omni1	1	1	p36.11	25469887	25537253	67	H	N/A
	45	Loss	SNP 6.0	1	12	q13.12 q13.13	48272109	51079253	2807	D	Pathogenic
			2.7M	1	12	q13.13	49027905	49487526	460	D	Pathogenic
			2.7M	1	12	q13.13	49504234	49868403	364	D	Pathogenic
			2.7M	1	12	q13.13	49936316	50469339	533	D	Pathogenic
			2.7M	1	12	q13.13	50559302	50665339	106	D	Pathogenic
			Omni1	1	12	q13.12 q13.13	48271477	51081854	2810	D	Pathogenic
			CytoSNP	1	12	q13.12 q13.13	48324559	51093547	2769	D	Pathogenic
CGX-12		12	q13.12 q13.13	48284132	51067934	2780	U	Pathogenic			
44.15	47	Loss	SNP 6.0	1	6	p21.32	32556106	32626735	71	D	Benign
			Omni1	0	6	p21.32	32563460	32613823	50	H	N/A
	49	Loss	SNP 6.0	1	22	q11.21	17256415	19795835	2539	D	Pathogenic
			2.7M	1	22	q11.21	17418054	18295326	877	D	Pathogenic
			Omni1	1	22	q11.21	17255869	18692668	1437	D	Pathogenic
			Omni1	1	22	q11.21	19038685	19795050	756	D	VOUS
			CytoSNP	1	22	q11.21	17269490	19792353	2523	D	Pathogenic
			CGX-12		22	q11.21	17299469	19790658	2490	U	Pathogenic
50.17	53	Loss	SNP 6.0	1	16	p12.2 p12.1	21441804	21746841	305	D	Benign
			2.7M	1	16	p12.2 p12.1	21306155	21741570	435	H	N/A
			Omni1	1	16	p12.2	21381301	21688143	307	D	Benign
			CytoSNP	1	16	p12.2	21208498	21647732	439	H	N/A
			CGX-12		16	p12.2	21479157	21647056	168	U	Benign
53.18	54	Loss	SNP 6.0	1	1	q23.3	159780382	159904813	124	D	Benign
			Omni1	0	1	q23.3	159768149	159910608	142	D	VOUS
	58	Gain	SNP 6.0	4	17	q12	31464078	31653797	190	H4	Benign
		Omni1	3	17	q12	31467924	31675965	208	H	N/A	

Chr: chromosome; CN: Copy Number state; D: de novo; H: inherited; H0: inherited from heterozygous parents & CN=0; H4: inherited from heterozygous parents & CN=4; U: unknown, VOUS: variant of uncertain clinical significance.

^a Inherited CNVs were not interpreted, except on the CGX-12 array or if they were inherited from heterozygous parents and resulted in a homozygous alteration.

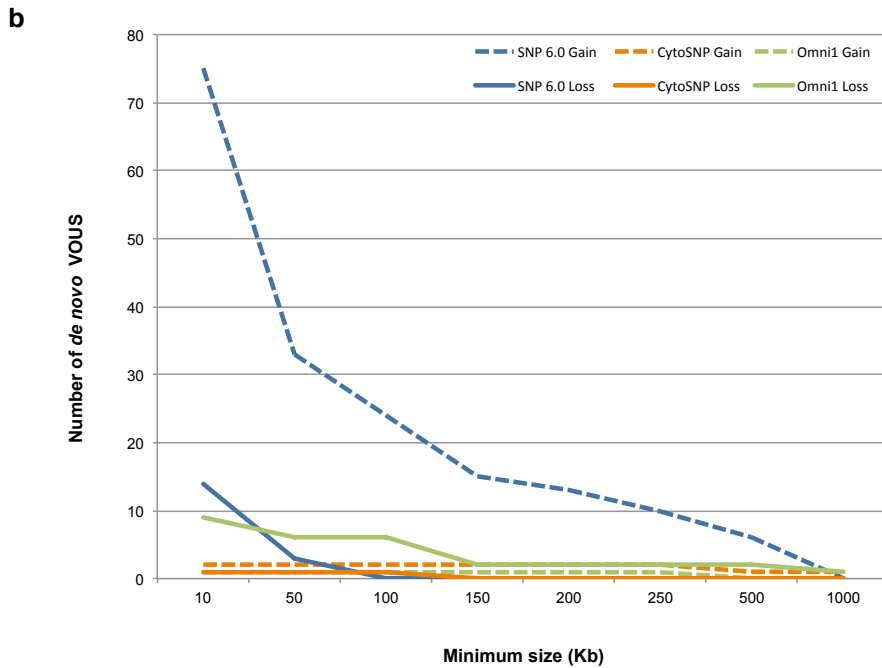
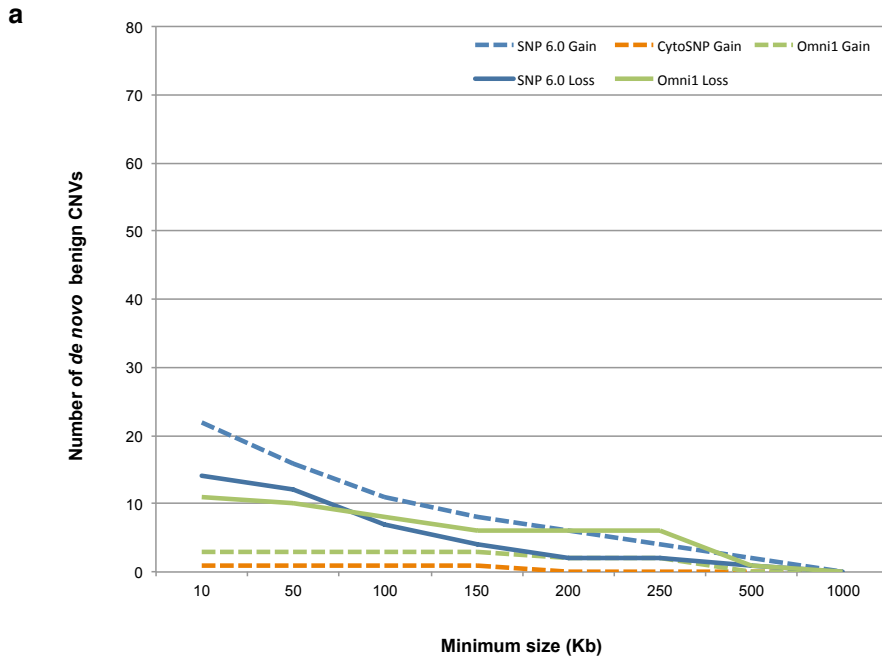


Figure 4-2 : Number of *de novo* CNVs detected by each array, as a function of minimum size.

(a) Benign CNVs. (b) VOUS.

Finally, SNP 6.0 found six pathogenic CNVs (6/139, 4.3%) while Omni1 found five (5/35, 4.7%) and CytoSNP found four (4/8, 50%), but these all correspond to the same four pathogenic alterations that were fragmented into smaller segments by SNP 6.0 and Omni1, thus none of the tested arrays identified a pathogenic alteration that was undetected by the other arrays. Therefore, these three arrays found a pathogenic CNV in 14.3% of patients (3/21, one patient had two CNVs), which corresponds to the results obtained on the CGX-12 array (details in the next section). The main differences resides in the total number of CNVs that had to be interpreted for each patient (Table 4-4), which ranged from an average of 1.7 per patient (range 0-4) for Omni1 and 0.4 (range 0-2) for CytoSNP, to 6.6 (range 0-17) for SNP 6.0, and the proportion of patients for whom we identified at least one VOUS, which was 14.3% for CytoSNP (3/21), 33.3% for Omni1 (7/21) and 95.2% for SNP 6.0 (20/21). In comparison, for CGX-12, the average number of CNVs to interpret for each patient was 2.7 (range 0-7), and the proportion of patients with a VOUS was 4.8% (1/21). In summary, all arrays detected the pathogenic CNVs, but the overall number of CNVs to interpret for each patient was higher for SNP 6.0.

4.4.5 Clinical significance of confirmed *de novo* CNVs

Of the 332 *de novo* CNVs detected on the four tested arrays, 47 were confirmed (detected on at least two arrays), and they represented 20 unique alterations (Table 4-5). Many of these showed variations in the breakpoints delineated by each array. Four of these alterations, in three patients, were classified as pathogenic, ten were classified as benign, and six were classified as VOUS. All pathogenic CNVs and two of the benign CNVs (patients 26.9 and 50.17) were also detected on CGX-12. One of the VOUS was deemed benign on CGX-12 (patient 32.11), because it was called as a heterozygous loss in a polymorphic region, but Omni1 and CytoSNP called a homozygous deletion (CN=0), hence the uncertain clinical significance interpretation. Only five CNVs were called on all four tested arrays: the four pathogenic CNVs, and one VOUS (patient 3.2, Figure 4-S1). This copy number gain could be seen retrospectively on CGX-12, but was not called by the algorithm.

In patient 14.5, CGX-12 found a 6.27 Mb gain of 6p25.3p25.1 and a 3.9 Mb loss of 7q36.3, resulting from an unbalanced 6;7 translocation (Figure 4-S2). All four tested arrays successfully detected these alterations, although 2.7M and Omni1 fragmented the chromosome 6 gain into two smaller segments, and SNP 6.0 fragmented it into three segments.

In patient 41.14, CGX-12 identified a 2.8 Mb deletion of 12q13.12q13.13 that was also detected by all tested arrays as a single segment, except for 2.7M, which fragmented this loss into 4 segments (Figure 4-S3).

In patient 44.15, CGX-12 detected a 2.5 Mb deletion in the DiGeorge/Velocardiofacial syndrome region (OMIM 188400/192430) on chromosome 22 (Figure 4-S4). SNP 6.0 and Cyto SNP both detected the deletion, while Omni1 fragmented this loss into two segments (1.43 Mb and 756 kb). The gap between these two segments is rich in segmental duplications and is not well covered by the other arrays. 2.7M only detected a 877 kb loss.

4.4.6 UPD

Three patients (1.1, 10.4 and 11.4), in two families, had LCSHs larger than 10 Mb. There were differences between arrays concerning the detection and the size of these LCSHs (Table 4-S4). All large LSCHs were successfully detected by SNP 6.0 (except one that was just below threshold at 9999.286 kb) and CytoSNP, with a few differences in size. However, Omni1 detected only 9 of the 21 large LSCHs, and 2.7M detected 18 of them. The remaining LCSHs were fragmented and thus below the 10 Mb threshold, and even below the 5 Mb threshold in some cases on the 2.7M array. All three patients had multiple large LCSHs on more than one chromosome, suggesting parental consanguinity. Since CytoSNP detected all LCSHs larger than 10 Mb, we report these results in the rest of this section. Patient 1.1 had nine LCSHs larger than 10 Mb, for a total of 150.239 Mb, and an estimated percentage of homozygosity of 5.2% (Table 4-S5). This was consistent with third degree consanguinity (coefficient of inbreeding (F): 0.0625). Patient 10.4 and 11.4, who are siblings, had six LCSHs each, for a total of 159.919 Mb and 168.894 Mb, respectively. The percentage of homozygosity was respectively 5.6% and 5.8%, which was also consistent with third degree

consanguinity. Parental consanguinity was confirmed with family history obtained from the parents.

4.4.7 Informative LCSHs

The four tested arrays detected a total of 112 LCSHs larger than 5 Mb, representing 28 unique homozygous regions in five patients (Table 4-S4). The 2.7M array detected the highest number with 31, of which 29 were confirmed on other arrays, but with important differences in size. The two remaining LCSHs were in fact detected on other arrays, but with a size below the 5 Mb threshold. The SNP 6.0 array detected 26 LCSHs larger than 5 Mb in five patients, all of which were detected by CytoSNP and 2.7M, except two: a 5.2 Mb LCSH on chromosome X (female patient 59.20) in a region sparsely covered by CytoSNP, and a 15.8 Mb LCSH on chromosome 16 (patient 1.1), which was fragmented into four segments smaller than the 5 Mb threshold by 2.7M. CytoSNP detected 25 LCSHs larger than 5 Mb, all of which were confirmed at least on the SNP 6.0 array. Finally, Omni1 detected 30 LCSHs in three patients (1.1, 10.4 and 11.4), many of which were fragments of the larger LCSHs found on the other arrays. Omni1 did not detect the LCSH in patient 59.20, despite adequate coverage, and detected the LCSH in patient 26.9, but fragmented it in two smaller segments that were below the 5 Mb threshold. Of the 25 confirmed LCSHs, 24 were found in patients born from consanguineous parents (patients 1.1, 10.4 and 11.4) and one was found in patient 26.9 (no consanguinity). Four of these LCSHs (in three families) were considered informative (Table 4-6), because they overlap a gene associated with a recessive condition compatible with the patient's phenotype. Three of them were confirmed to harbour a pathogenic homozygous mutation.

Patient 1.1 had a 32.2 Mb LCSH in 8q21.11q23.1 (Figure 4-S5), which contains *VPS13B* (OMIM 616817). Patients with Cohen syndrome (OMIM 216550) are homozygous or compound heterozygous for mutations in this gene. Cohen syndrome was amongst the suspected diagnoses for this patient, but exon sequencing only identified a synonymous mutation in exon 54, which is a known polymorphism found in healthy individuals.

Patient 11.4 had a 42.5 Mb LCSH in 11q14.1q24.2, which completely overlaps with a 32.0 Mb LCSH in 11q14.2q23.3 found in his brother, patient 10.4 (Figure 4-S6). The overlapping region contains *DPAGT1* (OMIM 191350) and *ALG9* (OMIM 606941). Homozygous or compound heterozygous mutations in either of these genes cause Congenital disorder of glycosylation, type 1j (OMIM 608093) and type11 (OMIM 608776), respectively. Both patients have DD, hypotonia, and abnormal transferrin glycosylation. Exome sequencing confirmed a homozygous A322G mutation in exon 3 of *DPAGT1* in both patients. This mutation has not been reported before.

Patient 26.9 had a 5 Mb LCSH in 2p21 (Figure 4-S7) that overlaps with *LRPPRC* (OMIM 607544). Homozygous and compound heterozygous mutations in this gene have been identified in patients affected with Leigh syndrome, French-Canadian type (OMIM 220111). This patient has DD, ataxia, lactic acidosis, ptosis, and obesity. He was confirmed to be homozygous for the A354V mutation in exon 9 of *LRPPRC*, which is found in most patients with French-Canadian type Leigh syndrome.

Table 4-6 : Informative LCSHs larger than 5 Mb.

Patient	LCSH	Array	Chr	Chr band(s)	Coordinates		Size (kb)	Gene(s) of interest & disorder	Confirmed?	
					Start	End				
1.1	7	SNP 6.0	8	q21.11	74259208	106471063	32 212	<i>VPS13B</i> , Cohen syndrome	No. Known polymorphism identified by exon sequencing (c.T9492C, Homozygous C).	
		2.7M		q21.11	74144714	103958979				
		Omni1		q22.3						29 814
				q21.11	74275315	106474543				
CytoSNP	q23.1			32 199						
	q21.11	74256250	106477669	32 221						
10.4	16	SNP 6.0	11	q14.2	86714942	118761985	32 047	<i>DPAGT1</i> , Congenital disorder of glycosylation, type 1j (CDG1J)	Yes. Homozygous c.A322G mutation in exon 3 of <i>DPAGT</i> identified by exome sequencing.	
		2.7M		q23.1	86640025	112683203				
		CytoSNP		q14.2	86714943	118727916				
				q23.3						26 043
11.4	24	SNP 6.0	11	q14.1	81709865	124226213	42 516	<i>DPAGT1</i> , Congenital disorder of glycosylation, type 1j (CDG1J)	Yes. Homozygous c.A322G mutation in exon 3 of <i>DPAGT</i> identified by exome sequencing.	
		2.7M		q23.3						
		Omni1		q23.3	118899401	124319838				35 050
				q24.2						5 420
CytoSNP	q21	93339399	103743297	10 404						
	q22.3			17 468						
	q22.3	106756061	124224315							
	q24.2			42 492						
26.9	27	SNP 6.0	2	p21	41858417	46908485	5 050	<i>LRPPRC</i> , Leigh syndrome, French-Canadian type (LSFC)	Yes. Homozygous c.C1119T mutation in exon 9, corresponding to the known major p.A354V present in 97.6% of affected LSFC.	
		2.7M		p21	41794186	46981017				5 187
		CytoSNP		p21	41850116	46912030				5 062

Chr: chromosome.

4.4.8 Interpretation software

Since establishing clinical significance of detected CNVs can be challenging, and the software used to visualize results influences interpretation efficiency, we also compared the cytogenetics software provided by the vendors in order to have a complete assessment of each platform's performance. ChAS (Affymetrix) and KaryoStudio (Illumina) allow array results analysis, interpretation and reporting by visualizing data in a chromosome and genome view. Both softwares offer various features that are detailed in Table 4-S6. In ChAS, every item visualized in "Detailed view" is clickable and links to either the detailed information in tables ("Called segments") or the external database entry (RefSeq gene, OMIM, DGV), which made it easier to follow the same interpretation workflow for each CNV and avoid missing important information. In KaryoStudio, the lack of links between the graphical view, the external databases (except RefSeq) and the information in tables made interpretation approximate, time-consuming and more prone to errors. In addition, although KaryoStudio offers the possibility of simultaneously visualizing results for a trio (patient and parents), called segments appear only for the patient, and have to be compared with the LogR or BAF to establish inheritance. For some CNVs, the parent's signal was clear, but for others it was not, and in these cases, there was no way to visually confirm if the same CNV was found in the parent. The only option was to search the "Detected Regions" table and compare genomic coordinates. One useful feature available in KaryoStudio is the ability to save interpretation information in a "Comments" column. This was not possible in ChAS, where there is an "Interpretation" column that can be exported with a table report, but the information entered cannot be saved within ChAS. Finally, both softwares offer a report feature that exports a graphical view and the details of the abnormal region. KaryoStudio exports results for all patients in the same batch with a simple click, while every abnormal region has to be exported individually in ChAS. However, with ChAS, the user can choose exactly which tracks to export, and how it is presented (individual or trio, zoomed or not). In summary, although KaryoStudio's interface is simpler, CNV interpretation was easier in ChAS.

4.5 Discussion

In the present study, we evaluated the diagnostic performance, in a clinical context, of four high-resolution SNP arrays by testing 21 patients affected with ID/DD and comparing the results obtained on each one. A diagnosis was successfully made in six patients (28.6%, 6/21) from 5 families (25.0%, 5/20): four pathogenic CNVs (including one unbalanced translocation) and three LCSHs harbouring a pathogenic homozygous mutation were detected. However, our results also highlight the challenges associated with the use of high-resolution arrays at low detection thresholds, and identify some of the issues specific to each array. Specifically, all arrays successfully picked up the pathogenic abnormalities, but with enough variation in size and breakpoints to have an impact on the quality and accuracy of the final reported results. Another major difference between arrays was observed in the overall number of CNVs detected by each platform, which has a significant impact on workload. Nonetheless, our results show that the tested SNP arrays are effective at identifying large CNVs, and that there is an added value in detecting and analysing LCSHs. Indeed, as a result of the homozygous mutations uncovered, the overall diagnostic yield was higher (28.6%, 6/21) than the yield we would have obtained with the platform we currently use (14.3%, 3/21), as well as the yield reported in previous studies using high resolution SNP arrays but restricted to CNV detection, which ranges from 4.4% to 16% (Bruno et al. 2009; Friedman et al. 2009; Gijbbers et al. 2009; Männik et al. 2011; McMullan et al. 2009).

4.5.1 CNVs

Recent studies performed with high-resolution arrays (Bruno et al. 2011a; Filges et al. 2012; Qiao et al. 2012) show the usefulness of these arrays by focusing on the pathogenic CNVs identified without taking into account the overall number of CNVs detected for each patient, therefore ignoring a critical aspect in a clinical setting: the time allocated to CNV interpretation and the challenges associated with result reporting. With standard thresholds, it was clear that both Affymetrix arrays detected too many CNVs for each patient (average of 24.0 and 17.5 per patient, for SNP 6.0 and 2.7M respectively, see Table 4-S1) to be used in a clinical laboratory. Indeed, we want to continue detecting small (<200 kb) pathogenic

alterations — the whole-genome array we currently use can detect pathogenic CNVs as small as 7 kb — while keeping the total number of CNVs to interpret for each patient to a manageable level. Filtering out benign variants and using different thresholds according to gene content proved to efficiently reduce the overall number of CNVs detected per patient (see Table 4-S2) to an average of 11.9 (SNP 6.0) and 10.3 (2.7M), which was comparable to what the Omni1 array detected (average 8.8), but still much higher than CytoSNP (average 1.2) and CGX-12 (average 2.8). Considering that the first three arrays have a much higher density of markers than CytoSNP and CGX-12, this higher number of detected CNVs was expected, and could nevertheless represent an advantage if it results in the detection of additional small pathogenic alterations.

It is noteworthy that although 80% of the CNVs detected on the Affymetrix arrays were gains (see Table 4-S2), a much lower proportion of gains (compared to losses) were confirmed (see Table 4-3). In contrast, there was no difference between gains and losses confirmation rates for the Illumina and the NimbleGen arrays. For the SNP 6.0 array, this bias towards gains was more important as the size of the alterations decreased, but only for VOUS (see Figure 4-2). In addition, no gains smaller than 150 kb on the Affymetrix arrays were confirmed, but larger gains and small losses were (see Table 4-5). This bias was seen in all tested patients, which suggests that there was a technical or design flaw affecting small gains (either false positives on the Affymetrix arrays, or false negatives on the Illumina arrays) that would limit the validity of the results obtained with the problematic array(s).

In addition to the overall number of CNVs detected by the Affymetrix arrays being an issue, the proportion of *de novo* VOUS was also a concern, especially considering the fact that most of them were not confirmed on another array. Although the problem seemed less important on SNP 6.0 than on 2.7M (92.7% of the *de novo* CNVs detected by 2.7M were VOUS, of which only two were confirmed), the SNP 6.0 array still detected a total of 89 VOUS (64.0%; 89/139), in 20 of the 21 patients, and only four of them were confirmed. This extremely low confirmation rate was not seen with the benign CNVs, for which Affymetrix had rates comparable to Illumina. Since the majority of these VOUS were gains, it supports a technical issue with the detection of copy number gains on the Affymetrix arrays. In contrast, CytoSNP identified a VOUS in only three patients and Omni1 found ten VOUS in seven

patients. Moreover, a higher proportion of them were confirmed on another array (2/3 and 4/10). Our results are consistent with the expectation that increased array resolution would lead to an increase in the overall number of CNVs detected, as well as VOUS. These proportions are more manageable on the Illumina arrays even if higher than CGX-12, which identified only one VOUS. This 45 kb gain was also seen on the SNP 6.0 and Omni1 arrays, but not called because it was below the size threshold (100 kb in regions without genes of known clinical significance). This last example confirms that using differential thresholds according to gene content is an effective approach to reduce the number of VOUS detected.

Even if SNP 6.0 and Omni1 detected a high proportion of benign CNVs, this was still lower than the results obtained on CGX-12. In addition, the proportion of confirmed benign CNVs was also low (see Table 4-5), but contrary to the confirmation rate of VOUS, the rates were similar for Affymetrix and Illumina arrays. Although detecting a high number of benign CNVs added to the interpretation workload, this was less of an issue since the mere fact of being able to classify them as benign means their interpretation was fairly straightforward. The fact that parents were not systematically tested, and the use of a pool of six individuals as reference DNA probably explain why most CNVs identified on CGX-12 were classified as benign. The reference DNA used was also likely responsible for the low confirmation rate and the discrepancies observed in the copy number states called by different arrays for a same segment in some patients, particularly in regions prone to variations in normal individuals. For example, CGX-12 detected a 103 kb gain at 8p11.23 that was deemed benign in patient 38.13, while Omni1 identified an overlapping 154 kb inherited loss in the same patient (see Table 4-S3). There were numerous examples of this, and it affected all arrays. This illustrates the importance of referring to normal variants databases that are complete and contain information from various studies, as well as in-house created databases.

We identified a total of four pathogenic CNVs in three patients, and all tested arrays detected them. However, in some instances, there were important differences in the breakpoints called by each array (data not shown). The 2.7M array was particularly problematic: it fragmented two of the four pathogenic CNVs, and called a third CNV less than half the size reported by all the other arrays. Omni1 also fragmented the pathogenic deletion identified in patient 44.15 into two smaller segments, separated by a 350 kb gap with a normal

copy number state. However, Omni1 is the only one of the five arrays that has a dense coverage of the genomic region in that gap, which is rich in segmental duplications and normal variants. This type of fragmentation in highly polymorphic regions is likely more representative of the true copy number state, and although its interpretation should be fairly easy, it could complicate result reporting. For this patient, the larger CNV would still have been considered pathogenic, but the smaller deleted segment would have been assigned an uncertain clinical significance. The other variations in breakpoints and fragmentation were less important, and would not have had any significant impact in the clinical interpretation of the results.

The high number of CNVs detected for each patient on some arrays may have been acceptable if it had also resulted in an increase in diagnostic yield, by allowing the detection of small pathogenic alterations that could not be seen on lower density arrays. However, that was not the case in our study, nor was it in another study applying uniform thresholds for all patients (Tucker et al. 2011). This may be due to a small sample size, since two other studies found pathogenic CNVs smaller than 100 kb in a cohort of 6500 and 82 patients, respectively (Bruno et al. 2011a; Qiao et al. 2012). However, these reports also suggest that decreasing detection thresholds is not sufficient to improve diagnostic yield. Indeed, the thresholds used in the first study (Bruno et al. 2011a) were not specifically defined, and they were possibly not the same for all tested patients since they were influenced by patients' phenotype information. In the second study (Qiao et al. 2012), a separate confirmation technique — in addition to various quality control parameters — was needed in order to reduce the false positive detection rate from 60% to 25%. This may be justifiable in a research setting, but we suspect the workload and the proportion of VOUS in these studies were tremendously important and would be difficult to sustain in a clinical setting. In addition, there should not be a need to confirm detected CNVs once an array has been properly validated (Kearney et al. 2011b; South and Brothman 2011), although doing so may be indicated to rule out mosaicism, or to visualize structural CNV location. These two studies and our results highlight the need to improve how we target potentially relevant genomic regions in order to increase the efficiency of high-resolution arrays, and use them to their full potential. Automating CNV interpretation with computational methods such as GECCO (Hehir-Kwa et al. 2010), and decreasing the

proportion of VOUS with more complete variants databases could also eventually allow the detection and interpretation of a higher number of CNVs for each patient and the identification of more pathogenic CNVs.

4.5.2 LCSHs & UPD

None of the patients tested in our study had UPD, therefore it was not possible to directly assess the arrays capacity at detecting it. Nevertheless, with the method we applied, results from SNP 6.0 and CytoSNP were highly concordant and lead to no false positives. Therefore, it is reasonable to believe that their performance would have been adequate. In fact, two prospective studies successfully detected UPD using these arrays with thresholds similar to ours (Bruno et al. 2011b; Papenhausen et al. 2011). The performance of 2.7M and Omni1 is more difficult to predict, since 2.7M missed three large LCSHs and Omni1 detected less than half of LCSHs larger than > 10 Mb. Because it tended to fragment LCSHs, using a method that takes into account smaller LCSHs (Tucker et al. 2012) could be more appropriate for Omni1, although doing so might increase the number of false positives, especially when patients come from a population with some degree of inbreeding. It is more plausible that higher LCSHs size thresholds, combined with individual chromosome size and LCSHs size average will be the best approach for UPD detection by SNP arrays. However, more prospective studies are needed to determine the optimal strategy for each array.

We were able to directly assess the capacity of each array to detect consanguinity, since parents were first cousins in two of the tested families. Three arrays successfully detected third degree consanguinity (SNP 6.0, 2.7M and CytoSNP). The percentages of homozygosity obtained with Omni1 were suggestive of consanguinity in two patients, but too ambiguous to establish at which degree (2.5% and 2.3%, see Table 4-S5), and the percentage was too low in the third patient (0.5%). However, decreasing the minimum size of LCSH included in the calculation to 3 Mb would have improved the performance of all arrays, including Omni1. It has been suggested that a threshold of 10 Mb is sufficient to identify consanguinity and that lowering the minimum size of LCSHs included in the calculation has a negligible impact on the determination of percentage of homozygosity (Kearney 2011). While

a 10 Mb threshold is probably adequate for first-degree consanguinity, perhaps it is not for more distant consanguinity, and a lower threshold would be more appropriate in these cases.

Lastly, gene content of LCSHs larger than 5 Mb was also studied to uncover potential homozygous mutations causing autosomal recessive disorders. Once again, SNP 6.0 and CytoSNP showed the best concordance and detected all LCSHs — except one on chromosome X — while Omni1 and 2.7M fragmented larger LCSHs into smaller segments, sometimes below the 5 Mb threshold. We followed up on four informative LCSHs — two in siblings, thus in the same genomic region — and successfully identified a pathogenic homozygous mutation in three of them. This resulted in a diagnosis for 14.3% of tested patients (3/21), and 66.6% of follow-ups (2/3). Previous studies reported a positive diagnosis in 0.08% of tested patients (Bruno et al. 2011b) — although in that report, the size threshold seems to have been variable — and in 10% of followed up LCSHs (Kearney et al. 2011a) (unknown thresholds). Our results clearly show the potential of SNP arrays in improving the diagnostic rate in patients with ID/DD, without unnecessarily burdening the workload, and compare advantageously to these reports. However, the expected yields might be overestimated because of our small sample size. The mutation suspected in the fourth LCSH, which was found in a patient with parental consanguinity, could not be confirmed. This LCSH was large and consequently, isolating a causative gene can be challenging.

Homozygosity mapping has been used for years in research, and more recently in clinical genetics. Combining it with molecular karyotype on SNP arrays undoubtedly facilitates its implementation and reduces overall costs (Alkuraya 2010). The challenge resides, as with CNVs, in the workload associated with phenotype-genotype correlations, as illustrated by our failure to identify a causative mutation in one family. Because of the high amount of homozygosity present in all individuals, and even higher in consanguineous families, directing the analysis with a precisely defined phenotype seems unavoidable (Bruno et al. 2011a), and might even allow the use of very low threshold if the list of genes to verify for a specific phenotype is short. In addition, using an approach with fixed thresholds may not be optimal for LCSHs, since their number varies greatly from patient to patient, depending on their genetic background, and smaller LCSHs can also harbour homozygous mutations. It may indeed be more productive to adjust size thresholds for each patient in order to get a

manageable number of LCSHs to further investigate. Automation can also play an important role in reducing the interpretation workload. As an example, Wierenga *et al.* recently developed an online tool that searches OMIM Clinical synopsis data to allow quick retrieval of the relevant genes within a genomic region, according to the patient's phenotype (Wierenga *et al.* 2012). More prospective clinical studies will be needed to establish a detailed algorithm and efficiently implement homozygosity mapping with SNP arrays in a clinical setting.

4.5.3 Interpretation software

Interpretation and reporting of array results is the last but most important step in a clinical cytogenetics laboratory. It is a challenge in some cases, and requires easy and quick access to various databases containing information that is accurate and up-to-date. ChAS and KaryoStudio each have their pros and cons. In summary, we found that ChAS was much more complete and customizable, which allowed easy and thorough CNV interpretation, while saving and reporting results was much simpler in KaryoStudio. However, both softwares lacked features that are needed to make them functional as a stand-alone software in a clinical laboratory. The increase in resolution of the current arrays has resulted in a rise in the number of CNVs that are detected, and the interpretation process needs to be that much more efficient. Neither of the provided softwares would be adequate as they are, especially because they cannot be linked to an in-house laboratory database to keep track of past results. This is essential for many reasons, such as future reinterpretation of results as knowledge expands, but especially in the maintenance of each laboratory's own database of normal variants, which vary depending on the population investigated or the platform used. There are a few softwares currently available that allow CNV (and/or LOH) visualization, interpretation, and database capabilities, such as Nexus Copy Number (BioDiscovery), CGHFusion (InfoQuant), and BENCH Lab CNV (Cartagenia) (Coe *et al.* 2009; de Leeuw *et al.* 2012), and that are successfully used by various clinical laboratories (Srebniak *et al.* 2012; Taapken *et al.* 2010; Wohlleber *et al.* 2010). While using them increases costs in the short term, arrays are constantly updated and improved, and laboratories needs may evolve. Therefore, this might

represent the best approach in the long term, since it facilitates switching from one vendor to another, and allows using a different array depending on the clinical indication.

4.6 Conclusion

High-resolution SNP arrays increase the diagnostic yield in patients with unexplained DD/MCA, compared to CGH microarrays, because of the information provided by the detection of LCSHs. However, the higher resolution does not increase the number of pathogenic CNVs detected, and there are differences in breakpoint accuracy, both for CNVs and LCSHs. In addition, the workload associated with the high number CNVs detected, the high proportion of VOUS identified (except for CytoSNP), and the inefficient visualization softwares provided by the vendors could make clinical implantation challenging. Therefore, *in silico* targeting of potentially pathogenic regions, interpretation automation, and the use of a commercial interpretation software and database may represent the best approaches that will allow the use of these arrays to their full potential in a clinical setting.

4.7 Acknowledgements

The authors would like to thank the children, and their parents, who participated in this study. They also would like to thank Michael S. Phillips and Mathieu Langlois from the Centre de Pharmacogénomique Beaulieu-Saucier for performing the SNP array genomic hybridization, Affymetrix and Illumina for graciously providing the arrays and reagents, and Mark E. Samuels for helpful discussions. G.D. is supported by a FRSQ Master's training award.

4.8 References

- Alkuraya FS (2010) Homozygosity mapping: one more tool in the clinical geneticist's toolbox. *Genet Med* 12:236–239. doi: 10.1097/GIM.0b013e3181ceb95d
- Boone PM, Bacino CA, Shaw CA, et al. (2010) Detection of clinically relevant exonic copy-number changes by array CGH. *Hum Mutat* 31:1326–1342. doi: 10.1002/humu.21360
- Bruno DL, Ganesamoorthy D, Schoumans J, et al. (2009) Detection of cryptic pathogenic copy number variations and constitutional loss of heterozygosity using high resolution SNP microarray analysis in 117 patients referred for cytogenetic analysis and impact on clinical practice. *J Med Genet* 46:123–131. doi: 10.1136/jmg.2008.062604
- Bruno DL, Stark Z, Amor DJ, et al. (2011a) Extending the scope of diagnostic chromosome analysis: detection of single gene defects using high-resolution SNP microarrays. *Hum Mutat* 32:1500–1506. doi: 10.1002/humu.21581
- Bruno DL, White SM, Ganesamoorthy D, et al. (2011b) Pathogenic aberrations revealed exclusively by single nucleotide polymorphism (SNP) genotyping data in 5000 samples tested by molecular karyotyping. *J Med Genet* 48:831–839. doi: 10.1136/jmedgenet-2011-100372
- Butler MG (2009) Genomic imprinting disorders in humans: a mini-review. *J Assist Reprod Genet* 26:477–486. doi: 10.1007/s10815-009-9353-3
- Coe BP, Lockwood WW, Chari R, Lam WL (2009) Comparative genomic hybridization on BAC arrays. In: Pollack JR (ed) *Microarray Analysis of the Physical Genome: Methods and Protocols*. Methods in molecular biology (Clifton, N.J.), pp 7–19
- Cooper GM, Coe BP, Girirajan S, et al. (2011) A copy number variation morbidity map of developmental delay. *Nat Genet* 43:838–846. doi: 10.1038/ng.909
- Curry CJ, Stevenson RE, Aughton D, et al. (1997) Evaluation of mental retardation: recommendations of a Consensus Conference: American College of Medical Genetics. In: *Am. J. Med. Genet.* pp 468–477
- D'Amours G, Kibar Z, Mathonnet G, et al. (2012) Whole-genome array CGH identifies pathogenic copy number variations in fetuses with major malformations and a normal karyotype. *Clin Genet* 81:128–141. doi: 10.1111/j.1399-0004.2011.01687.x

- de Leeuw N, Dijkhuizen T, Hehir-Kwa JY, et al. (2012) Diagnostic interpretation of array data using public databases and internet sources. *Hum Mutat*. doi: 10.1002/humu.22049
- Filges I, Suda L, Weber P, et al. (2012) High resolution array in the clinical approach to chromosomal phenotypes. *Gene* 495:163–169. doi: 10.1016/j.gene.2011.12.042
- Friedman J, Adam S, Arbour L, et al. (2009) Detection of pathogenic copy number variants in children with idiopathic intellectual disability using 500 K SNP array genomic hybridization. *BMC Genomics* 10:526. doi: 10.1186/1471-2164-10-526
- Gijbbers ACJ, Lew JYK, Bosch CAJ, et al. (2009) A new diagnostic workflow for patients with mental retardation and/or multiple congenital abnormalities: test arrays first. *Eur J Hum Genet* 17:1394–1402. doi: 10.1038/ejhg.2009.74
- Goldenberg A, Saugier-Verber P (2010) [Genetics of mental retardation]. *Pathol Biol* 58:331–342. doi: 10.1016/j.patbio.2009.09.013
- Hehir-Kwa JY, Wieskamp N, Webber C, et al. (2010) Accurate distinction of pathogenic from benign CNVs in mental retardation. *PLoS Comput Biol* 6:e1000752. doi: 10.1371/journal.pcbi.1000752
- Heinzen EL, Radtke RA, Urban TJ, et al. (2010) Rare deletions at 16p13.11 predispose to a diverse spectrum of sporadic epilepsy syndromes. *Am J Hum Genet* 86:707–718. doi: 10.1016/j.ajhg.2010.03.018
- Kearney HM, Kearney JB, Conlin LK (2011a) Diagnostic implications of excessive homozygosity detected by SNP-based microarrays: consanguinity, uniparental disomy, and recessive single-gene mutations. *Clin Lab Med* 31:595–613. doi: 10.1016/j.cll.2011.08.003
- Kearney HM, South ST, Wolff DJ, et al. (2011b) American College of Medical Genetics recommendations for the design and performance expectations for clinical genomic copy number microarrays intended for use in the postnatal setting for detection of constitutional abnormalities. *Genet Med* 13:676–679. doi: 10.1097/GIM.0b013e31822272ac
- Koolen DA, Sharp AJ, Hurst JA, et al. (2008) Clinical and molecular delineation of the 17q21.31 microdeletion syndrome. *J Med Genet* 45:710–720. doi: 10.1136/jmg.2008.058701

- Ledbetter DH, Martin CL (2007) Cryptic telomere imbalance: a 15-year update. *Am J Med Genet Part C Semin Med Gen* 145C:327–334. doi: 10.1002/ajmg.c.30149
- Madhavan T, Narayan J (1991) Consanguinity and mental retardation. *J Ment Defic Res* 35:133–139.
- Manning M, Hudgins L, Professional Practice and Guidelines Committee (2010) Array-based technology and recommendations for utilization in medical genetics practice for detection of chromosomal abnormalities. *Genet Med* 12:742–745. doi: 10.1097/GIM.0b013e3181f8baad
- Männik K, Parkel S, Palta P, et al. (2011) A parallel SNP array study of genomic aberrations associated with mental retardation in patients and general population in Estonia. *Eur J Med Genet* 54:136–143. doi: 10.1016/j.ejmg.2010.11.005
- McMullan DJ, Bonin M, Hehir-Kwa JY, et al. (2009) Molecular karyotyping of patients with unexplained mental retardation by SNP arrays: a multicenter study. *Hum Mutat* 30:1082–1092. doi: 10.1002/humu.21015
- Mefford HC, Sharp AJ, Baker C, et al. (2008) Recurrent rearrangements of chromosome 1q21.1 and variable pediatric phenotypes. *N Engl J Med* 359:1685–1699. doi: 10.1056/NEJMoa0805384
- Miller DT, Adam MP, Aradhya S, et al. (2010) Consensus statement: chromosomal microarray is a first-tier clinical diagnostic test for individuals with developmental disabilities or congenital anomalies. *Am J Hum Genet* 86:749–764. doi: 10.1016/j.ajhg.2010.04.006
- Neill NJ, Torchia BS, Bejjani BA, et al. (2010) Comparative analysis of copy number detection by whole-genome BAC and oligonucleotide array CGH. *Mol Cytogenet* 3:11. doi: 10.1186/1755-8166-3-11
- Papenhausen P, Schwartz S, Risheg H, et al. (2011) UPD detection using homozygosity profiling with a SNP genotyping microarray. *Am J Med Genet* 155:757–768. doi: 10.1002/ajmg.a.33939
- Qiao Y, Tyson C, Hrynychak M, et al. (2013) Clinical application of 2.7M Cytogenetics array for CNV detection in subjects with idiopathic autism and/or intellectual disability. *Clin Genet* 83:145–154. doi: 10.1111/j.1399-0004.2012.01860.x

- Sagoo GS, Butterworth AS, Sanderson S, et al. (2009) Array CGH in patients with learning disability (mental retardation) and congenital anomalies: updated systematic review and meta-analysis of 19 studies and 13,926 subjects. *Genet Med* 11:139–146. doi: 10.1097/GIM.0b013e318194ee8f
- Sharp AJ, Hansen S, Selzer RR, et al. (2006) Discovery of previously unidentified genomic disorders from the duplication architecture of the human genome. *Nat Genet* 38:1038–1042. doi: 10.1038/ng1862
- Sharp AJ, Mefford HC, Li K, et al. (2008) A recurrent 15q13.3 microdeletion syndrome associated with mental retardation and seizures. *Nat Genet* 40:322–328. doi: 10.1038/ng.93
- Shevell M, Ashwal S, Donley D, et al. (2003) Practice parameter: evaluation of the child 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* 60:367–380.
- South ST, Brothman AR (2011) Clinical laboratory implementation of cytogenomic microarrays. *Cytogenet Genome Res* 135:203–211. doi: 10.1159/000331425
- Srebniak M, Boter M, Oudesluijs G, et al. (2011) Application of SNP array for rapid prenatal diagnosis: implementation, genetic counselling and diagnostic flow. *Eur J Hum Genet* 19:1230–1237. doi: 10.1038/ejhg.2011.119
- Srebniak MI, Boter M, Oudesluijs GO, et al. (2012) Genomic SNP array as a gold standard for prenatal diagnosis of foetal ultrasound abnormalities. *Mol Cytogenet* 5:14. doi: 10.1186/1755-8166-5-14
- Sridhara Rama Rao BS, Narayanan HS (1976) Consanguinity and familial mental retardation. *J Med Genet* 13:27–29.
- Stavropoulos J, Shago M, Bruyere H, et al. (2010) CCMG Guidelines for Genomic Microarray Testing. Canadian College of Medical Geneticists
- Taapken SM, Nisler BS, Sampsell-Barron TL, Montgomery KD (2010) Automation in an array comparative genomic hybridization laboratory improves throughput and data quality. *Journal of the Association for Laboratory Automation* 15:390–395.

- Tucker T, Montpetit A, Chai D, et al. (2011) Comparison of genome-wide array genomic hybridization platforms for the detection of copy number variants in idiopathic mental retardation. *BMC Med Genomics* 4:25. doi: 10.1186/1755-8794-4-25
- Tucker T, Schlade-Bartusiak K, Eydoux P, et al. (2012) Uniparental disomy: can SNP array data be used for diagnosis? *Genet Med*. doi: 10.1038/gim.2012.35
- van Karnebeek CDM, Scheper FY, Abeling NG, et al. (2005) Etiology of mental retardation in children referred to a tertiary care center: a prospective study. *Am J Ment Retard* 110:253–267. doi: 10.1352/0895-8017(2005)110[253:EOMRIC]2.0.CO;2
- Weiss LA, Shen Y, Korn JM, et al. (2008) Association between microdeletion and microduplication at 16p11.2 and autism. *N Engl J Med* 358:667–675. doi: 10.1056/NEJMoa075974
- Wierenga KJ, Jiang Z, Yang AC, et al. (2012) A clinical evaluation tool for SNP arrays, especially for autosomal recessive conditions in offspring of consanguineous parents. *Genet Med*. doi: 10.1038/gim.2012.136
- Wohlleber E, Kirchhoff M, Zink AM, et al. (2010) Clinical and molecular characterization of two patients with overlapping de novo microdeletions in 2p14-p15 and mild mental retardation. *Eur J Med Genet* 54:67–72. doi: 10.1016/j.ejmg.2010.09.012
- Xu B, Roos JL, Levy S, et al. (2008) Strong association of de novo copy number mutations with sporadic schizophrenia. *Nat Genet* 40:880–885. doi: 10.1038/ng.162
- Zlotogora J, Shalev SA (2010) The consequences of consanguinity on the rates of malformations and major medical conditions at birth and in early childhood in inbred populations. *Am J Med Genet* 152A:2023–2028. doi: 10.1002/ajmg.a.33537

4.9 Supporting information

Table 4-S1 : Number of CNVs detected by each tested array using standard thresholds.

Number of CNVs	SNP 6.0	2.7M	Omni1	CytoSNP	All arrays
Total	503	368	197	30	1098
Gains (% ^a)	430 (85.5%)	314 (85.3%)	89 (45.2%)	16 (53.3%)	849 (77.3%)
Losses (% ^a)	73 (14.5%)	54 (14.7%)	108 (54.8%)	14 (46.7%)	249 (22.7%)
<i>De novo</i> (% ^a)	109 (21.7%)	176 (47.8%)	34 (17.3%)	10 (33.3%)	329 (30.0%)
Inherited (% ^a)	394 (78.3%)	192 (52.2%)	163 (82.7%)	20 (66.7%)	769 (70.0%)
All					
Range per patient	14-37	3-66	6-15	0-5	29-98
Median per patient	22	12	9	1	50
Average per patient	24.0	17.5	9.4	1.4	52.3
<i>De novo</i>					
Range per patient	1-12	0-29	0-4	0-2	3-35
Median per patient	5	6	2	0	14
Average per patient	5.2	8.4	1.6	0.5	15.7
Inherited					
Range per patient	12-29	0-39	4-11	0-5	22-63
Median per patient	19	6	8	1	32
Average per patient	18.8	9.1	7.8	1.0	36.6

^a % of total

Table 4-S2 : Number of CNVs detected by each tested array using optimized thresholds.

Number of CNVs	SNP 6.0	2.7M	Omni1	CytoSNP	CGX-12	All arrays
Total	250	216	185	25	58	734
Gains (% ^a)	201 (80.4%)	172 (79.6%)	78 (42.2%)	16 (64.0%)	27 (46.6%)	494 (67.3%)
Losses (% ^a)	49 (19.6%)	44 (20.4%)	107 (57.8%)	9 (36.0%)	31 (53.4%)	240 (32.7%)
<i>De novo</i> (% ^a)	139 (55.6%)	150 (69.4%)	35 (18.9%)	8 (32.0%)	N/A ^b	332 (45.2%) ^c
Inherited (% ^a)	111 (44.4%)	66 (30.6%)	150 (81.1%)	17 (68.0%)	2 (3.4%)	344 (46.9%) ^c
All						
Range per patient	4-30	1-32	5-13	0-5	0-7	22-51
Median per patient	11	9	9	1	3	32
Average per patient	11.9	10.3	8.8	1.2	2.8	35.0
<i>De novo</i>						
Range per patient	0-17	0-20	0-4	0-2	N/A ^b	10-23 ^c
Median per patient	7	6	2	0	N/A ^b	13 ^c
Average per patient	6.6	7.1	1.7	0.4	N/A ^b	15.8 ^c
Inherited						
Range per patient	1-13	0-12	4-11	0-5	0-1	9-25 ^c
Median per patient	5	2	7	1	0	16 ^c
Average per patient	5.3	3.1	7.1	0.8	0.1	16.4 ^c

^a % of total

^b Parents not systematically tested

^c Excluding CNVs detected on CGX-12

Table 4-S3 : List of CNVs confirmed on at least two arrays.

Alteration	Array	Patient	CN	Gain/ Loss	Chr	Chr band(s)	Coordinates start	Coordinates end	Size (kb)	Inheritance	Interpretation^a
1	SNP 6.0	3.2	3	Gain	2	p21	43848625	44069477	221	D	VOUS
	2.7M	3.2	3	Gain	2	p21	44001534	44065431	64	D	VOUS
	Omni1	3.2	3	Gain	2	p21	43844007	44248753	405	D	VOUS
	CytoSNP	3.2	3	Gain	2	p21	43884153	44225225	341	D	VOUS
2	SNP 6.0	3.2	1	Loss	21	q21.1	22094265	22199637	105	H	N/A
	Omni1	3.2	1	Loss	21	q21.1	22094903	22195946	101	H	N/A
3	Omni1	3.2	3	Gain	22	q11.1	15453066	15801989	349	H	N/A
	CytoSNP	3.2	3	Gain	22	q11.1	15461707	15791899	330	H	N/A
4	Omni1	6.3	1	Loss	2	q37.3	242512401	242750507	238	H	N/A
	CytoSNP	6.3	1	Loss	2	q37.3	242566407	242678246	112	H	N/A
	CGX-12	6.3		Loss	2	q37.3	242513576	242654701	141	U	Benign
5	Omni1	6.3	3	Gain	7	q11.21	64365027	64748676	384	H	N/A
	CytoSNP	6.3	3	Gain	7	q11.21	64316996	64824912	508	H	N/A
	CGX-12	6.3		Gain	7	q11.21	64317115	64511622	195	H	VOUS
6	Omni1	6.3	3	Gain	10	q11.22	47004551	47139283	135	H	N/A
	CytoSNP	6.3	3	Gain	10	q11.22	47066810	47221596	155	H	N/A
7	Omni1	6.3	1	Loss	10	q21.3	67938287	68044090	106	H	N/A
	CytoSNP	6.3	1	Loss	10	q21.3	67929325	68040186	111	H	N/A
8	Omni1	6.3	1	Loss	22	q11.23	23994408	24249492	255	H	N/A
	CytoSNP	6.3	1	Loss	22	q11.23	24056992	24240879	184	H	N/A
9	SNP 6.0	10.4	3	Gain	1	q24.2	167557554	167702604	145	H	N/A
	Omni1	10.4	3	Gain	1	q24.2	167555394	167700744	145	H	N/A
10	SNP 6.0	10.4	1	Loss	2	q23.3	153829472	153962506	133	H	N/A
	Omni1	10.4	1	Loss	2	q23.3	153823451	153962506	139	H	N/A
	CGX-12	10.4		Loss	2	q23.3	153830942	153959766	129	U	Benign
11	SNP 6.0	10.4	3	Gain	14	q32.33	105082340	105867730	785	D	VOUS
	Omni1	10.4	3	Gain	14	q32.33	105446201	105855351	409	H	N/A
12	SNP 6.0	11.4	1	Loss	2	q23.3	153829472	153962506	133	H	N/A
	Omni1	11.4	1	Loss	2	q23.3	153823451	153962506	139	H	N/A
13	SNP 6.0	11.4	3	Gain	3	p11.1	89771865	90008100	236	H	N/A
	SNP 6.0	11.4	3	Gain	3	p11.1	90034987	90370639	336	H	N/A
	2.7M	11.4	3	Gain	3	p11.1	89772434	90060996	289	H	N/A
	2.7M	11.4	3	Gain	3	p11.1 q11.2	90123470	95002793	4879	H	N/A
	Omni1	11.4	3	Gain	3	p11.1	89759430	90401980	643	H	N/A

<i>Alteration</i>	<i>Array</i>	<i>Patient</i>	<i>CN</i>	<i>Gain/ Loss</i>	<i>Chr</i>	<i>Chr band(s)</i>	<i>Coordinates start</i>	<i>Coordinates end</i>	<i>Size (kb)</i>	<i>Inheritance</i>	<i>Interpretation^d</i>
	CytoSNP	11.4	3	Gain	3	p11.1	89780708	90525615	745	H	N/A
	CGX-12	11.4		Gain	3	p11.1	89772351	90275193	503	U	Benign
14	SNP 6.0	11.4	3	Gain	8	q11.1	47043274	47381308	338	H	N/A
	2.7M	11.4	3	Gain	8	q11.1	47042670	47597918	555	D	VOUS
15	SNP 6.0	14.5	1	Loss	1	p36.11	25477632	25535885	58	D	Benign
	Omni1	14.5	1	Loss	1	p36.11	25469887	25528125	58	H	N/A
16	SNP 6.0	14.5	1	Loss	2	p12	82866297	83152322	286	H	N/A
	2.7M	14.5	1	Loss	2	p12	82865835	83147096	281	H	N/A
	Omni1	14.5	1	Loss	2	p12	82867397	83146076	279	H	N/A
	CGX-12	14.5		Loss	2	p12	82866933	83141928	275	U	Benign
17	SNP 6.0	14.5	3	Gain	6	p25.3 p25.2	129591	2752759	2623	D	Pathogenic
	SNP 6.0	14.5	3	Gain	6	p25.2 p25.1	2779263	4643666	1864	D	Pathogenic
	SNP 6.0	14.5	3	Gain	6	p25.1	4646487	6416684	1770	D	Pathogenic
	2.7M	14.5	3	Gain	6	p25.3	109490	401749	292	D	Pathogenic
	2.7M	14.5	3	Gain	6	p25.3 p25.1	561662	6414958	5853	D	Pathogenic
	Omni1	14.5	3	Gain	6	p25.3	100815	610448	510	D	Pathogenic
	Omni1	14.5	3	Gain	6	p25.3 p25.2 p25.1	647434	6415891	5768	D	Pathogenic
	CytoSNP	14.5	3	Gain	6	p25.3 p25.2 p25.1	94609	6412865	6318	D	Pathogenic
	CGX-12	14.5		Gain	6	p25.3 p25.1	128203	6399128	6270	U	Pathogenic
18	Omni1	14.5	4	Gain	7	q11.21	61639074	61755397	116	H	N/A
	CytoSNP	14.5	4	Gain	7	q11.21	61631605	61776647	145	H	N/A
19	SNP 6.0	14.5	1	Loss	7	q36.3	154929838	158812469	3883	D	Pathogenic
	2.7M	14.5	1	Loss	7	q36.3	154962181	158811204	3849	D	Pathogenic
	Omni1	14.5	1	Loss	7	q36.3	154946906	158815293	3868	D	Pathogenic
	CytoSNP	14.5	1	Loss	7	q36.3	154952401	158812247	3860	D	Pathogenic
	CGX-12	14.5		Loss	7	q36.3	154947683	158816094	3870	U	Pathogenic
20	Omni1	17.6	0	Loss	8	p11.23 p11.22	39352501	39506336	154	H	N/A
	CGX-12	17.6		Loss	8	p11.23	39365946	39490567	125	U	Benign
21	SNP 6.0	17.6	3	Gain	8	q11.1	47046239	47473783	428	H	N/A
	2.7M	17.6	3	Gain	8	q11.1	47042670	47589707	547	H	N/A
22	SNP 6.0	17.6	3	Gain	14	q32.33	105082340	105998615	916	D	VOUS
	Omni1	17.6	3	Gain	14	q32.33	105446084	105927451	481	H	N/A
23	SNP 6.0	17.6	2	Gain	Y	q11.223	23765338	24269772	504	H	N/A
	SNP 6.0	17.6	2	Gain	Y	q11.223 q11.23	24313856	26866713	2553	H	N/A
	2.7M	17.6	2	Gain	Y	q11.223	23664966	23856606	192	H	N/A

<i>Alteration</i>	<i>Array</i>	<i>Patient</i>	<i>CN</i>	<i>Gain/ Loss</i>	<i>Chr</i>	<i>Chr band(s)</i>	<i>Coordinates start</i>	<i>Coordinates end</i>	<i>Size (kb)</i>	<i>Inheritance</i>	<i>Interpretation^d</i>
	2.7M	17.6	2	Gain	Y	q11.223 q11.23	25301370	25489188	188	H	N/A
24	CytoSNP	20.7	3	Gain	8	p23.2	2258259	2408353	150	H	N/A
	CGX-12	20.7		Gain	8	p23.2	2257338	2413416	156	U	Benign
25	Omni1	20.7	3	Gain	10	q11.21	44536118	44679489	143	H	N/A
	CGX-12	20.7		Gain	10	q11.21	44532360	44677919	146	U	Benign
26	Omni1	20.7	3	Gain	17	q21.31	41519627	41713128	194	H	N/A
	CGX-12	20.7		Gain	17	q21.31	41541432	41631306	90	U	Benign
27	SNP 6.0	20.7	1	Loss	19	q13.42	59994794	60055324	61	D	Benign
	Omni1	20.7	1	Loss	19	q13.42	59949314	60068550	119	D	Benign
28	SNP 6.0	23.8	1	Loss	15	q15.3	41680363	41726086	46	H	N/A
	2.7M	23.8	1	Loss	15	q15.3	41636500	41801729	165	D	Benign
	Omni1	23.8	1	Loss	15	q15.3	41670172	41735843	66	H	N/A
29	SNP 6.0	23.8	3	Gain	22	q11.23	24035301	24246386	211	H	N/A
	Omni1	23.8	3	Gain	22	q11.23	23998784	24249492	251	H	N/A
	CytoSNP	23.8	3	Gain	22	q11.23	24041667	24240879	199	D	Benign
30	Omni1	26.9	3	Gain	17	q21.31	41519743	41713128	193	D	Benign
	CGX-12	26.9		Gain	17	q21.31	41541432	41631306	90	U	Benign
31	SNP 6.0	29.1	3	Gain	1	p22.1	92059761	92385499	326	H	N/A
	Omni1	29.1	3	Gain	1	p22.1	92070475	92381440	311	H	N/A
	CGX-12	29.1		Gain	1	p22.1	92061886	92371388	310	H	Benign
32	SNP 6.0	29.1	3	Gain	2	q13	110061320	110340958	280	H	N/A
	Omni1	29.1	3	Gain	2	q13	110201336	110339819	138	H	N/A
	CytoSNP	29.1	3	Gain	2	q13	110201336	110328316	127	H	N/A
	CGX-12	29.1		Gain	2	q13	110183551	110335513	152	U	Benign
33	SNP 6.0	29.1	3	Gain	8	q11.1	47046239	47381308	335	H	N/A
	2.7M	29.1	3	Gain	8	q11.1	47042670	47594607	552	H	N/A
34	SNP 6.0	32.11	1	Loss	5	p13.3	30333033	30470952	138	H	N/A
	Omni1	32.11	1	Loss	5	p13.3	30351424	30471111	120	H	N/A
	CGX-12	32.11		Loss	5	p13.3	30347121	30469474	122	U	Benign
35	SNP 6.0	32.11	4	Gain	6	p21.32	32564648	32635250	71	H	N/A
	Omni1	32.11	3	Gain	6	p21.32	32577151	32613823	37	H	N/A
36	Omni1	32.11	0	Loss	19	p12	20404485	20509513	105	H0	VOUS
	CytoSNP	32.11	0	Loss	19	p12	20398497	20514068	116	D	VOUS
	CGX-12	32.11		Loss	19	p12	20408868	20506175	97	U	Benign
37	SNP 6.0	35.12	1	Loss	14	q21.3	44201738	44403334	202	H	N/A
	2.7M	35.12	1	Loss	14	q21.3	44199396	44404021	205	H	N/A

<i>Alteration</i>	<i>Array</i>	<i>Patient</i>	<i>CN</i>	<i>Gain/ Loss</i>	<i>Chr</i>	<i>Chr band(s)</i>	<i>Coordinates start</i>	<i>Coordinates end</i>	<i>Size (kb)</i>	<i>Inheritance</i>	<i>Interpretation^d</i>
	Omni1	35.12	1	Loss	14	q21.3	44199614	44404737	205	H	N/A
	CGX-12	35.12		Loss	14	q21.3	44227138	44387982	161	U	Benign
38	SNP 6.0	35.12	3	Gain	14	q32.33	105100669	106005581	905	H	N/A
	Omni1	35.12	3	Gain	14	q32.33	105451690	105978228	527	H	N/A
39	SNP 6.0	35.12	1	Loss	16	p12.1	27245100	27258188	13	H	N/A
	2.7M	35.12	1	Loss	16	p12.1	27247601	27258480	11	H	N/A
40	SNP 6.0	38.13	1	Loss	6	p21.32	32589952	32626735	37	D	Benign
	Omni1	38.13	1	Loss	6	p21.32	32585397	32609122	24	H	N/A
41	Omni1	38.13	1	Loss	8	p11.23 p11.22	39352501	39506336	154	H	N/A
	CGX-12	38.13		Gain	8	p11.23	39387824	39490567	103	U	Benign
42	Omni1	38.13	3	Gain	17	q21.31	41521621	41723989	202	H	N/A
	CytoSNP	38.13	3	Gain	17	q21.31 q21.32	41519743	41922779	403	H	N/A
	CGX-12	38.13		Gain	17	q21.31	41541432	41631306	90	U	Benign
43	SNP 6.0	41.14	1	Loss	1	p36.11	25455203	25519561	64	D	VOUS
	Omni1	41.14	1	Loss	1	p36.11	25469887	25537253	67	H	N/A
44	SNP 6.0	41.14	1	Loss	6	q12	66456937	66599725	143	H	N/A
	Omni1	41.14	1	Loss	6	q12	66459576	66600314	141	H	N/A
	CGX-12	41.14		Loss	6	q12	66470628	66593618	123	U	Benign
45	SNP 6.0	41.14	3	Gain	8	p11.1	43856033	47381308	3525	H	N/A
	2.7M	41.14	3	Gain	8	p11.1	43535936	47458730	3923	H	N/A
46	SNP 6.0	41.14	1	Loss	12	q13.12 q13.13	48272109	51079253	2807	D	Pathogenic
	2.7M	41.14	1	Loss	12	q13.13	49027905	49487526	460	D	Pathogenic
	2.7M	41.14	1	Loss	12	q13.13	49504234	49868403	364	D	Pathogenic
	2.7M	41.14	1	Loss	12	q13.13	49936316	50469339	533	D	Pathogenic
	2.7M	41.14	1	Loss	12	q13.13	50559302	50665339	106	D	Pathogenic
	Omni1	41.14	1	Loss	12	q13.12 q13.13	48271477	51081854	2810	D	Pathogenic
	CytoSNP	41.14	1	Loss	12	q13.12 q13.13	48324559	51093547	2769	D	Pathogenic
	CGX-12	41.14		Loss	12	q13.12 q13.13	48284132	51067934	2780	U	Pathogenic
47	Omni1	41.14	3	Gain	15	q11.2	19798009	20091543	294	H	N/A
	CytoSNP	41.14	3	Gain	15	q11.2	18818086	20317047	1499	H	N/A
48	SNP 6.0	44.15	1	Loss	6	p21.32	32556106	32626735	71	D	Benign
	Omni1	44.15	0	Loss	6	p21.32	32563460	32613823	50	H	N/A
49	Omni1	44.15	0	Loss	8	p11.23 p11.22	39370187	39506336	136	H	N/A
	CGX-12	44.15		Loss	8	p11.23	39365946	39490567	125	U	Benign
50	SNP 6.0	44.15	1	Loss	22	q11.21	17256415	19795835	2539	D	Pathogenic
	2.7M	44.15	1	Loss	22	q11.21	17418054	18295326	877	D	Pathogenic

<i>Alteration</i>	<i>Array</i>	<i>Patient</i>	<i>CN</i>	<i>Gain/ Loss</i>	<i>Chr</i>	<i>Chr band(s)</i>	<i>Coordinates start</i>	<i>Coordinates end</i>	<i>Size (kb)</i>	<i>Inheritance</i>	<i>Interpretation^d</i>
	Omni1	44.15	1	Loss	22	q11.21	17255869	18692668	1437	D	Pathogenic
	Omni1	44.15	1	Loss	22	q11.21	19038685	19795050	756	D	VOUS
	CytoSNP	44.15	1	Loss	22	q11.21	17269490	19792353	2523	D	Pathogenic
	CGX-12	44.15		Loss	22	q11.21	17299469	19790658	2490	U	Pathogenic
51	SNP 6.0	44.15	3	Gain	22	q11.23	23991375	24247730	256	H	N/A
	Omni1	44.15	3	Gain	22	q11.23	23991390	24244593	253	H	N/A
	CytoSNP	44.15	3	Gain	22	q11.23	24041667	24254444	213	H	N/A
52	SNP 6.0	47.16	4	Gain	2	p11.2	88899204	89242149	343	H	N/A
	Omni1	47.16	3	Gain	2	p11.2	88922426	89101325	179	H	N/A
53	SNP 6.0	47.16	4	Gain	3	p12.2	82916842	83087402	171	H	N/A
	Omni1	47.16	3	Gain	3	p12.2	82914933	83092216	177	H	N/A
	CGX-12	47.16		Gain	3	p12.2	82915480	83084183	169	U	Benign
54	SNP 6.0	50.17	1	Loss	16	p12.2 p12.1	21441804	21746841	305	D	Benign
	2.7M	50.17	1	Loss	16	p12.2 p12.1	21306155	21741570	435	H	N/A
	Omni1	50.17	1	Loss	16	p12.2	21381301	21688143	307	D	Benign
	CytoSNP	50.17	1	Loss	16	p12.2	21208498	21647732	439	H	N/A
	CGX-12	50.17		Loss	16	p12.2	21479157	21647056	168	U	Benign
55	SNP 6.0	53.18	1	Loss	1	q23.3	159780382	159904813	124	D	Benign
	Omni1	53.18	0	Loss	1	q23.3	159768149	159910608	142	D	VOUS
56	SNP 6.0	53.18	3	Gain	8	q11.1	46966686	47386013	419	H	N/A
	2.7M	53.18	3	Gain	8	q11.1	47042670	47474410	432	H	N/A
57	SNP 6.0	53.18	1	Loss	8	q13.3	73256174	73377311	121	H	N/A
	Omni1	53.18	1	Loss	8	q13.3	73266402	73377033	111	H	N/A
	CytoSNP	53.18	1	Loss	8	q13.3	73258535	73377033	118	H	N/A
58	Omni1	53.18	3	Gain	12	p11.21	31187486	31300846	113	H	N/A
	CytoSNP	53.18	3	Gain	12	p11.21	31086539	31292645	206	H	N/A
59	SNP 6.0	53.18	4	Gain	17	q12	31464078	31653797	190	H4	Benign
	Omni1	53.18	3	Gain	17	q12	31467924	31675965	208	H	N/A
60	SNP 6.0	56.19	4	Gain	2	p11.2	88899204	89281576	382	H	N/A
	Omni1	56.19	3	Gain	2	p11.2	88941152	89101325	160	H	N/A
61	SNP 6.0	56.19	3	Gain	8	q11.1	47049925	47294570	245	H	N/A
	2.7M	56.19	3	Gain	8	p11.1	43517316	47597918	4081	H	N/A
62	SNP 6.0	56.19	3	Gain	14	q32.33	105100669	105930091	829	H	N/A
	Omni1	56.19	3	Gain	14	q32.33	105400732	105590401	190	H	N/A
63	Omni1	56.19	3	Gain	X	p22.33	1130687	1464278	334	H	N/A
	CGX-12	56.19		Gain	X	p22.33	1033739	1431315	398	U	Benign

<i>Alteration</i>	<i>Array</i>	<i>Patient</i>	<i>CN</i>	<i>Gain/ Loss</i>	<i>Chr</i>	<i>Chr band(s)</i>	<i>Coordinates start</i>	<i>Coordinates end</i>	<i>Size (kb)</i>	<i>Inheritance</i>	<i>Interpretation^a</i>
64	Omni1	59.2	3	Gain	10	q11.21	44523664	44679489	156	H	N/A
	CytoSNP	59.2	3	Gain	10	q11.21	44557214	44679489	122	H	N/A
	CGX-12	59.2		Gain	10	q11.21	44532360	44677919	146	U	Benign

Chr: chromosome, CN: Copy Number state, D: de novo, H: inherited, H0: inherited from heterozygous parent & CN = 0, H4: inherited from heterozygous parent & CN = 4, VOUS: variant of uncertain clinical significance.

^a Inherited CNVs were not interpreted, except on the CGX-12 array or if they resulted in a homozygous alteration.

Table 4-S4 : List of LCSHs larger than 5 Mb.

<i>Patient</i>	<i>LCSH</i>	<i>Array</i>	<i>Chr</i>	<i>Chr band(s)</i>	<i>Coordinates start</i>	<i>Coordinates end</i>	<i>Size (kb)</i>	<i>Interpretation</i>
1.1	1	SNP 6.0	1	q31.1 q32.1	185704276	202879427	17 175	NI
		2.7M		q31.1 q32.1	185298759	199220118	13 921	NI
		Omni1		q31.3 q32.1	192631284	200867175	8 236	NI
		CytoSNP		q31.1 q32.1	185710374	202875817	17 165	NI
	2	SNP 6.0	2	q14.2 q21.2	121442514	132997313	11 555	NI
		2.7M		q14.2 q21.2	121778868	133035506	11 257	NI
		Omni1		q14.3 q21.2	127405306	132997684	5 592	NI
		CytoSNP		q14.2 q21.2	121449946	132994153	11 544	NI
	3	SNP 6.0	3	q21.2 q22.3	126317406	139299507	12 982	NI
		2.7M		q21.2 q22.3	126228738	139771911	13 543	NI
		Omni1		q21.2 q22.3	127158201	137503665	10 345	NI
		CytoSNP		q21.2 q22.3	126326179	139310057	12 984	NI
	4	2.7M	3	q25.33 q26.1	160147220	167468263	7 321	NI
	5	SNP 6.0	4	q21.23 q22.3	85153753	95901447	10 748	NI
		2.7M		q21.23 q22.3	84771766	95997323	11 226	NI
		Omni1		q22.1 q22.3	89270941	95889528	6 619	NI
		CytoSNP		q21.23 q22.3	85149254	95868284	10 719	NI
	6	SNP 6.0	5	p15.31	8126377	16464810	8 338	NI
		2.7M		p15.31 p15.1	8020227	16544052	8 524	NI
		CytoSNP		p15.31 p15.1	8130509	16471829	8 341	NI
	7	SNP 6.0	8	q21.11 q23.1	74259208	106471063	32 212	Informative
		2.7M		q21.11 q22.3	74144714	103958979	29 814	Informative
		Omni1		q21.11 q23.1	74275315	106474543	32 199	Informative
		CytoSNP		q21.11 q23.1	74256250	106477669	32 221	Informative
	8	SNP 6.0	10	q23.31 q24.2	91941308	101144430	9 203	NI
		2.7M		q23.31 q24.2	91859746	101210063	9 350	NI
		CytoSNP		q23.31 q24.2	91929328	101128516	9 199	NI
	9	SNP 6.0	16	p13.3 p13.11	26670	15797518	15 771	NI
		Omni1		p13.3 p13.11	1971	15225201	15 223	NI
		CytoSNP		p13.3 p13.11	50925	15795766	15 745	NI
	10	SNP 6.0	18	q12.1 q21.31	25404113	52540576	27 136	NI
		2.7M		q12.1 q21.31	25293658	52708225	27 415	NI
		Omni1		q12.1 q12.3	28453349	36511380	8 058	NI
		Omni1		q12.3 q21.31	37132113	52394552	15 262	NI
		CytoSNP		q12.1 q21.31	25410998	52455677	27 045	NI
	11	SNP 6.0	22	q11.1 q12.1	16236959	26236244	9 999	NI
		CytoSNP		q11.1 q12.1	16235554	26244054	10 009	NI

<i>Patient</i>	<i>LCSH</i>	<i>Array</i>	<i>Chr</i>	<i>Chr band(s)</i>	<i>Coordinates start</i>	<i>Coordinates end</i>	<i>Size (kb)</i>	<i>Interpretation</i>
	12	SNP 6.0	22	q12.2 q13.2	29283520	42092301	12 809	NI
		CytoSNP		q12.2 q13.2	29290412	42097023	12 807	NI
10.4	13	2.7M	X	q13.1 q21.1	71084969	77716485	6 632	NI
	14	SNP 6.0	3	p14.2 p12.3	61597189	74629918	13 033	NI
		2.7M		p14.2 p12.3	61364515	74850408	13 486	NI
		Omni1		p14.1 p12.3	67649944	74608906	6 959	NI
		CytoSNP		p14.2 p12.3	61610634	74608906	12 998	NI
	15	SNP 6.0	8	q11.23 q22.3	53273604	102232276	48 959	NI
		2.7M		q11.23 q22.3	53251463	102358627	49 107	NI
		Omni1		q12.1 q12.3	57814044	64131874	6 318	NI
		Omni1		q12.3 q13.3	64142279	71879417	7 737	NI
		Omni1		q21.11 q21.3	75529385	90955185	15 426	NI
		Omni1		q21.3 q22.1	90961719	96940947	5 979	NI
		Omni1		q22.1 q22.3	97183808	102224589	5 041	NI
		CytoSNP		q11.23 q22.3	53274939	102224589	48 950	NI
		16	SNP 6.0	11	q14.2 q23.1	86714942	118761985	32 047
		2.7M		q14.2 q23.1	86640025	112683203	26 043	Informative
		CytoSNP		q14.2 q23.3	86714943	118727916	32 013	Informative
17	SNP 6.0	12	q13.13 q21.1	48680684	70029255	21 349	NI	
	2.7M		q13.13 q21.1	48667426	70248329	21 581	NI	
	Omni1		q13.13 q13.2	48736243	53975084	5 239	NI	
	Omni1		q14.1 q15	61225824	67535354	6 310	NI	
	CytoSNP		q13.13 q21.1	48696800	70027331	21 331	NI	
18	SNP 6.0	13	q13.2 q14.11	32987752	67439378	34 452	NI	
	2.7M		q13.2 q14.11	32950098	44197080	11 247	NI	
	2.7M		q14.12 q21.33	44316584	67603604	23 287	NI	
	Omni1		q13.2 q13.3	32989610	38247459	5 258	NI	
	Omni1		q14.12 q14.3	44327434	49492391	5 165	NI	
	Omni1		q14.3 q21.1	49972694	55222980	5 250	NI	
	Omni1		q21.2 q21.33	59833428	67403871	7 570	NI	
	CytoSNP		q13.2 q21.33	33002405	67717268	34 715	NI	
19	SNP 6.0	13	q21.33 q22.2	67890615	74624577	6 734	NI	
	2.7M		q21.33 q22.2	67781276	74746569	6 965	NI	
	Omni1		q21.33 q22.1	68174740	73786929	5 612	NI	
	CytoSNP		q21.33 q22.2	67872384	74712633	6 840	NI	
20	SNP 6.0	22	q12.3 q13.31	33353411	43434248	10 081	NI	
	Omni1		q12.3 q13.2	35830589	41924990	6 094	NI	
	CytoSNP		q12.3 q13.31	33153828	43386132	10 232	NI	
11.4	21	SNP 6.0	5	q23.1 q31.3	115840660	141547982	25 707	NI

<i>Patient</i>	<i>LCSH</i>	<i>Array</i>	<i>Chr</i>	<i>Chr band(s)</i>	<i>Coordinates start</i>	<i>Coordinates end</i>	<i>Size (kb)</i>	<i>Interpretation</i>
		2.7M		q23.1 q31.1	115497671	134570772	19 073	NI
		2.7M		q31.1 q31.3	134861767	141498288	6 637	NI
		Omni1		q23.1 q23.3	120684093	127363932	6 680	NI
		Omni1		q23.3 q31.2	127465300	136877472	9 412	NI
		CytoSNP		q23.1 q31.3	115826400	141544829	25 718	NI
	22	SNP 6.0	11	p14.2 p11.12	26293367	51084159	24 791	NI
		2.7M		p11.2 p11.12	44610852	51383794	6 773	NI
		2.7M		p13 p11.2	34214491	44519520	10 305	NI
		2.7M		p14.2 p13	26276823	34098104	7 821	NI
		Omni1		p14.1 p11.12	29951041	51199751	21 249	NI
		CytoSNP		p14.2 p11.12	26244233	51398585	25 154	NI
	23	SNP 6.0	11	q11 q14.1	54584637	77408373	22 824	NI
		2.7M		q11 q12.2	54809921	60952865	6 143	NI
		2.7M		q13.4 q14.1	70896911	77506826	6 610	NI
		Omni1		q12.1 q13.1	57624716	65397482	7 773	NI
		CytoSNP		q11 q14.1	54847844	77408373	22 561	NI
	24	SNP 6.0	11	q14.1 q24.2	81709865	124226213	42 516	Informative
		2.7M		q14.1 q23.3	81631409	116681560	35 050	Informative
		2.7M		q23.3 q24.2	118899401	124319838	5 420	NI
		Omni1		q21 q22.3	93339399	103743297	10 404	NI
		Omni1		q22.3 q24.2	106756061	124224315	17 468	Informative
		CytoSNP		q14.1 q24.2	81730470	124222740	42 492	Informative
	25	SNP 6.0	12	p13.31 p11.1	6789829	34237698	27 448	NI
		2.7M		p13.31 p11.1	6624778	34290302	27 666	NI
		Omni1		p12.1 p11.21	22154486	31792017	9 638	NI
		CytoSNP		p13.31 p11.21	6811282	31789630	24 978	NI
	26	SNP 6.0	12	q11 q14.2	36144017	61751530	25 608	NI
		2.7M		q11 q13.12	36184809	47635644	11 451	NI
		2.7M		q13.13 q14.2	48621894	61768808	13 147	NI
		Omni1		q11 q13.2	36162667	54318130	18 155	NI
		CytoSNP		q12 q14.2	36702406	61746824	25 044	NI
26.9	27	SNP 6.0	2	p21	41858417	46908485	5 050	Informative
		2.7M		p21	41794186	46981017	5 187	Informative
		CytoSNP		p21	41850116	46912030	5 062	Informative
59.20	28	SNP 6.0	X	q11.1 q12	61865213	67040497	5 175	NI
		2.7M		q11.1 q12	61934834	67143616	5 209	NI

Chr: chromosome, NI: non informative.

Table 4-S5 : Percentage of homozygosity^a calculated from copy-neutral LOHs.

Patient	> 10 Mb				> 5 Mb				> 3 Mb			
	SNP 6.0	2.7M	CytoSNP	Omni1	SNP 6.0	2.7M	CytoSNP	Omni1	SNP 6.0	2.7M	CytoSNP	Omni1
1.1	4,9	3,7	5,2	2,5	5,9	4,6	5,9	3,5	6,0	5,0	6,0	4,6
3.2	-	-	-	-	-	-	-	-	0,1	0,8	-	-
6.3	-	-	-	-	-	-	-	-	-	-	-	-
10.4	5,6	5,0	5,6	0,5	5,8	5,3	5,8	3,3	6,4	5,7	6,3	4,8
11.4	5,9	4,1	5,8	2,3	5,9	5,4	5,8	3,5	6,0	5,6	5,9	4,7
14.5	-	-	-	-	-	-	-	-	-	0,1	-	-
17.6	-	-	-	-	-	-	-	-	0,4	-	-	-
20.7	-	-	-	-	-	-	-	-	-	0,1	-	-
23.8	-	-	-	-	-	-	-	-	0,1	0,2	-	-
26.9	-	-	-	-	0,2	0,2	0,2	-	0,4	0,6	0,3	0,1
29.10	-	-	-	-	-	-	-	-	-	-	-	-
32.11	-	-	-	-	-	-	-	-	-	-	-	-
35.12	-	-	-	-	-	-	-	-	0,1	0,2	0,1	-
38.13	-	-	-	-	-	-	-	-	-	0,2	-	-
41.14	-	-	-	-	-	-	-	-	0,1	0,3	0,1	0,1
44.15	-	-	-	-	-	-	-	-	0,1	-	-	-
47.16	-	-	-	-	-	-	-	-	-	-	-	-
50.17	-	-	-	-	-	-	-	-	0,2	0,1	0,1	-
53.18	-	-	-	-	-	-	-	-	-	-	-	-
56.19	-	-	-	-	-	-	-	-	0,1	0,2	0,1	-
59.20	-	-	-	-	-	-	-	-	0,1	0,3	-	-

^a Excluding chromosome X

Table 4-S6 : Software features and their impact on result interpretation.

Feature	ChAS	KaryoStudio
Karyotype view	Yes	No
Chromosome view	Yes	Yes
Detailed view	Yes, configurable	Yes, not configurable
Results tables	Segments, CytoRegions, Overlap Map & Graphs	Detected Regions, Known Regions & Database of Genomic Variants
Type of data available for visualization	Copy Number State, Log2 Ratio, Allele Difference, LOH, Smooth Signal, Weighted Log2 Ratio, Mosaicism, LCSH, Allele Peaks	LogR, Smoothed LogR, B Allele Freq
Data style	Bar, points, big dots, line, min/max/avg, stairstep, heatmap; Adjustable range & height	Line (LogR), points (B Allele Freq); Optional AutoScale Log axis
Result files	One file per individual	One file for many individuals (batch)
Trio visualization	Automatic, requires loading 3 files	Possible but laborious, called segments cannot be directly compared
Segment manipulation	Segment Smoothing and Joining	None
Available tracks		
OMIM genes	Yes	Some via Known regions
RefSeq genes	Yes, cannot see orientation	Yes, not visible in trio view
Database of Genomic Variants	Collapsed or extended view; Gains in blue, losses in red	Condensed view only; Gains and losses in same color
Array markers	Yes	Yes
Segmental duplications	Yes	No
Sno/miRNA	Yes	No
FISH clones, BACs	Yes	No
Filters	2 sets: CytoRegions and Genome; Segment Filters: size, marker count, % confidence and % overlap (Overlap Map)	2 sets: Known Regions and Genome; Found Regions Filter Form: size, marker count and % confidence
Regions of interest	CytoRegions, any track or custom file can be used; Possible to link to external database (ex: OMIM); Restricted Mode available	Known Regions, provided file (245 Cytogenetic regions) or a custom file can be used; No link to external database
Masked regions	Overlap Map, any track or custom file can be used to mask CNVs with an adjustable % overlap	Ignore Regions, custom file can be used to mask CNVs with complete overlap; DGV track cannot be used
Links to external databases	RefSeq genes, OMIM, DGV, FISH clones; All accessible from detailed view	RefSeq genes accessible from detailed view; Known Regions & DGV accessible from table view only

Feature	ChAS	KaryoStudio
Inheritance	Simple to visualize & establish	Visualization not straightforward, difficult to compare patient's results with parents
Interpretation	Can be written and exported, not saved	Can be saved in Comments column
Configuration	Last configuration and filters used automatically reloaded	Known Regions file and Filters need to be reloaded every time
Search function	Can search any item in any track or custom file to visualize region in loaded patient(s)	Can search gene or cytoband to visualize region in selected patient(s)
Report	Completely customizable; Exports visualized region with chosen tracks, one report per abnormality	Not customizable; Exports one report per patient in the batch, which includes checked abnormalities

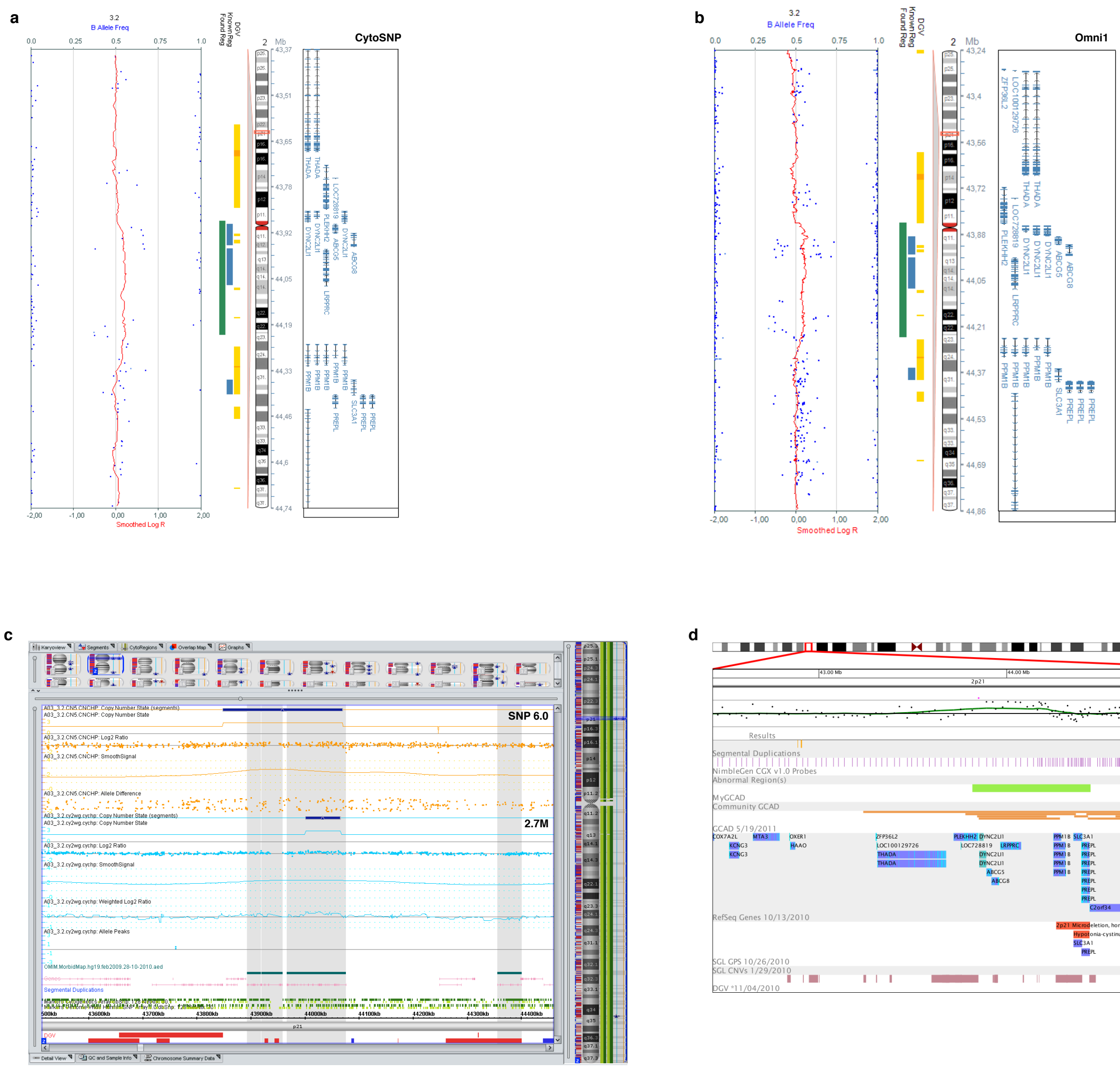


Figure 4-S1 : VOUS detected in patient 3.2, as visualized in each software.

2p21 gain. (a) CytoSNP. (b) Omni1. (c) SNP 6.0 and 2.7M. (d) CGX-12.

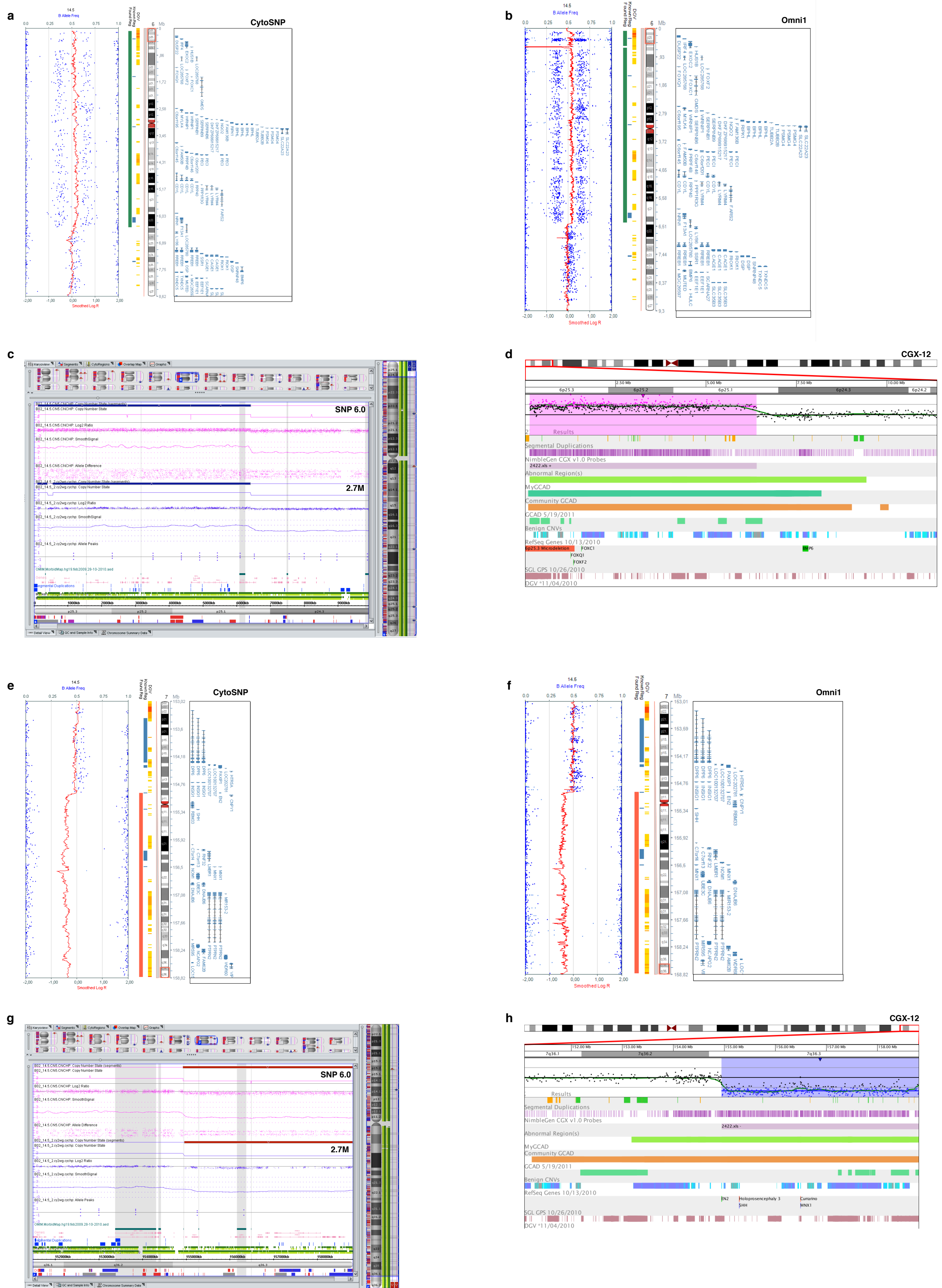


Figure 4-S2 : Pathogenic CNVs detected in patient 14.5, as visualized in each software.

Der(7)t(6;7)(p25.1;q36.3). (a-d) 6p25.1pter gain. (a) CytoSNP. (b) Omni1. (c) SNP 6.0 and 2.7M. (d) CGX-12. (e-h) 7q36.3qter loss. (e) CytoSNP. (f) Omni1. (g) SNP 6.0 and 2.7M. (h) CGX-12.

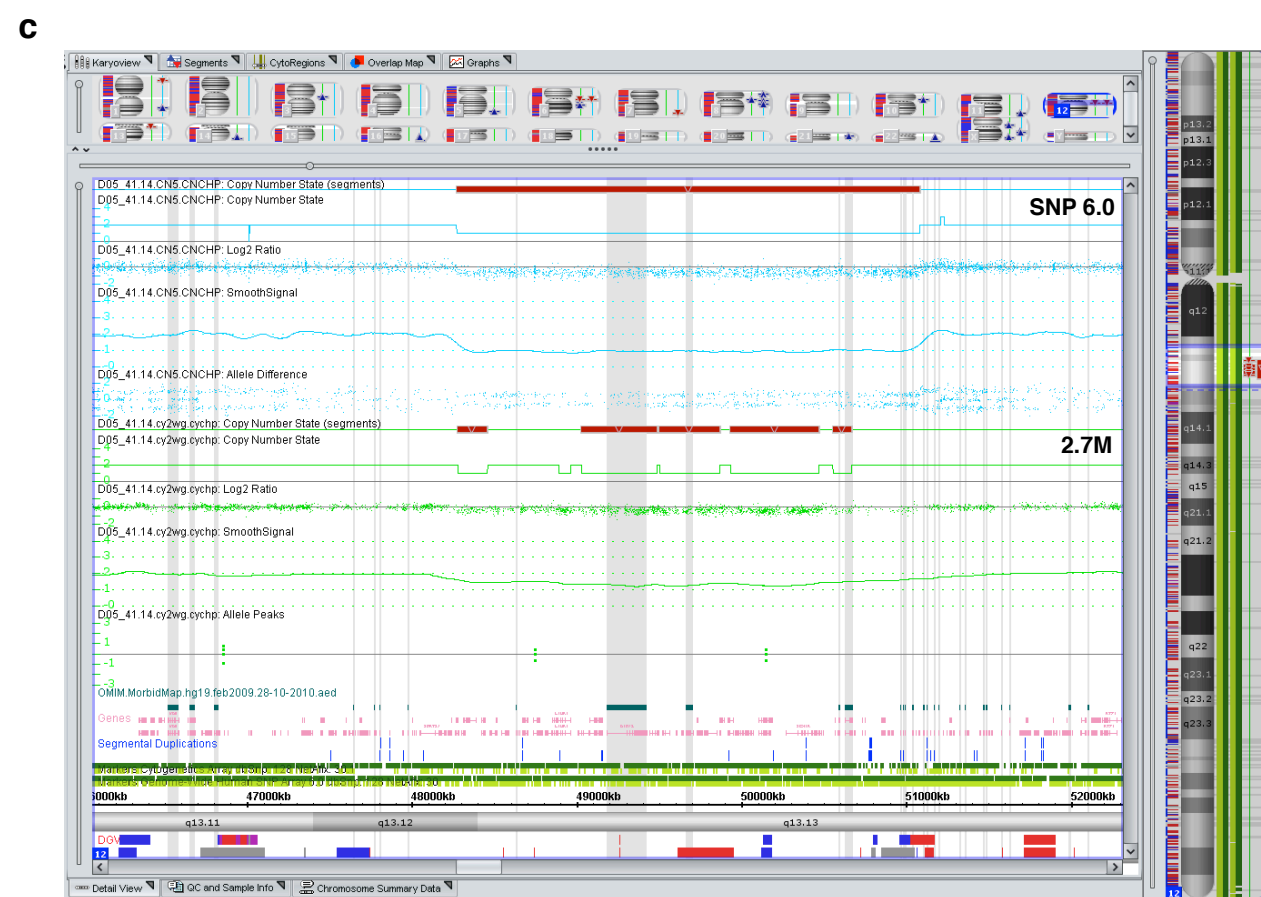
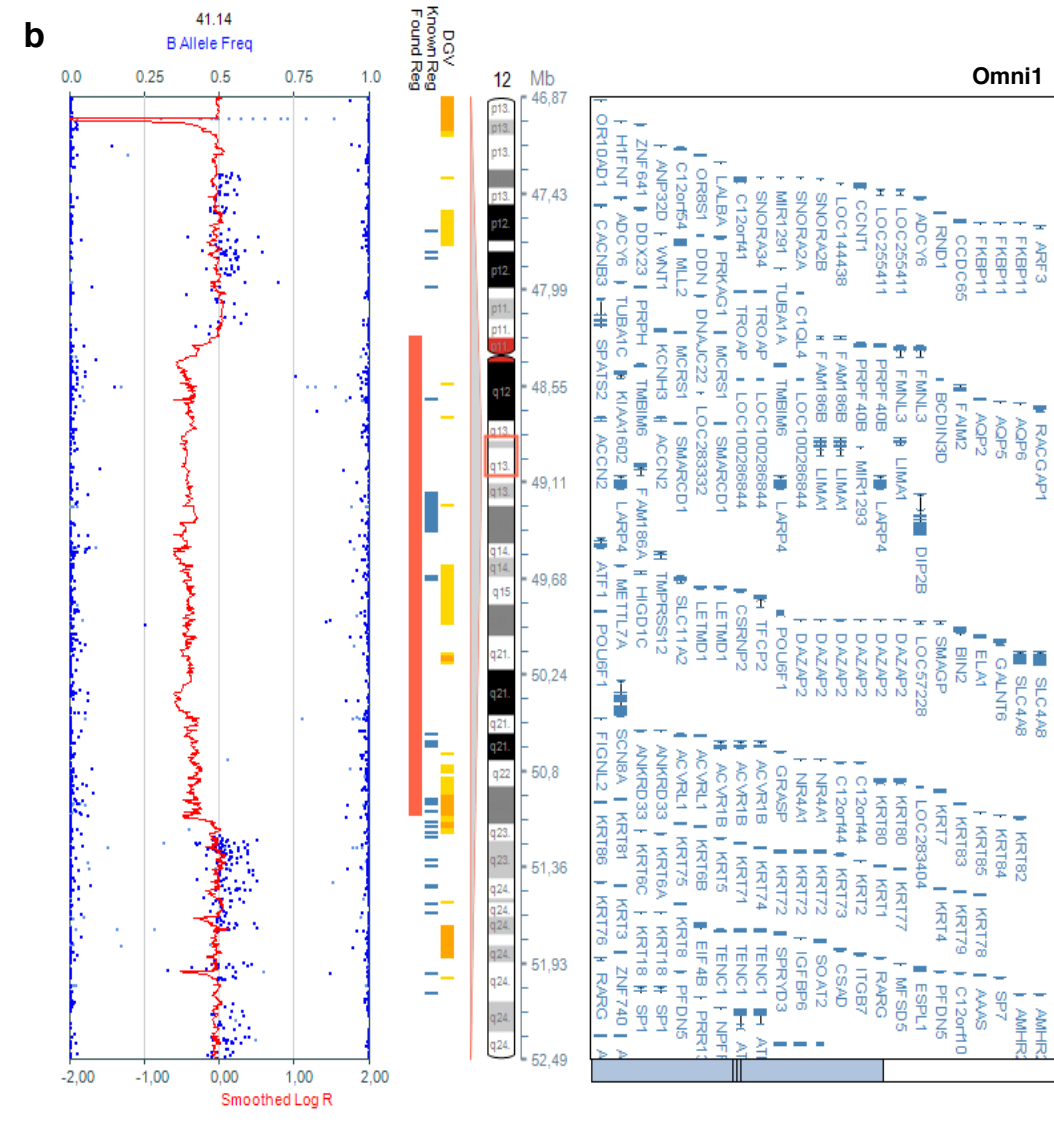
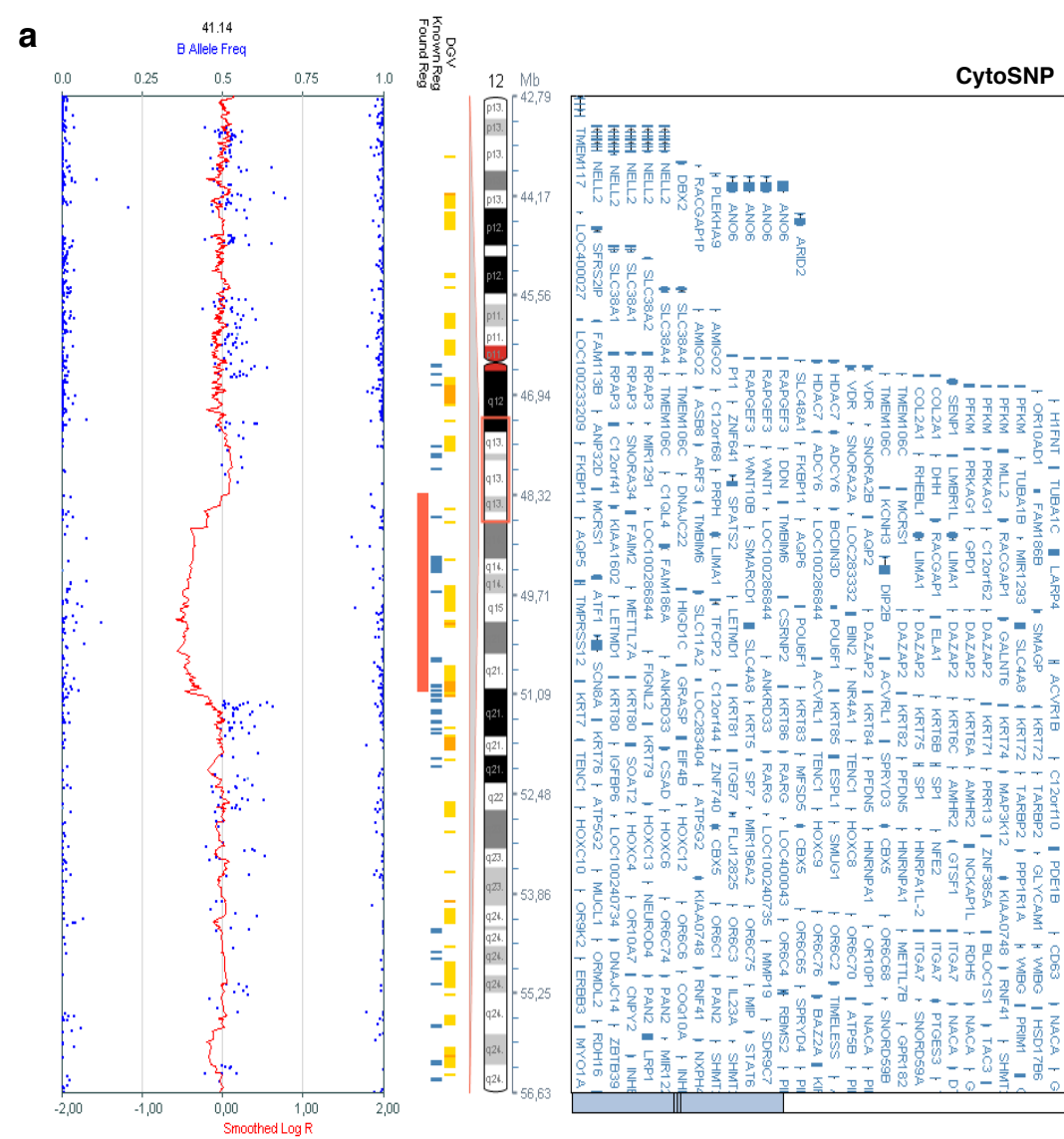


Figure 4-S3 : Pathogenic CNV detected in patient 14.5, as visualized in each software.

12q13.12q13.13 loss. (a) CytoSNP. (b) Omni1. (c) SNP 6.0 and 2.7M. (d) CGX-12

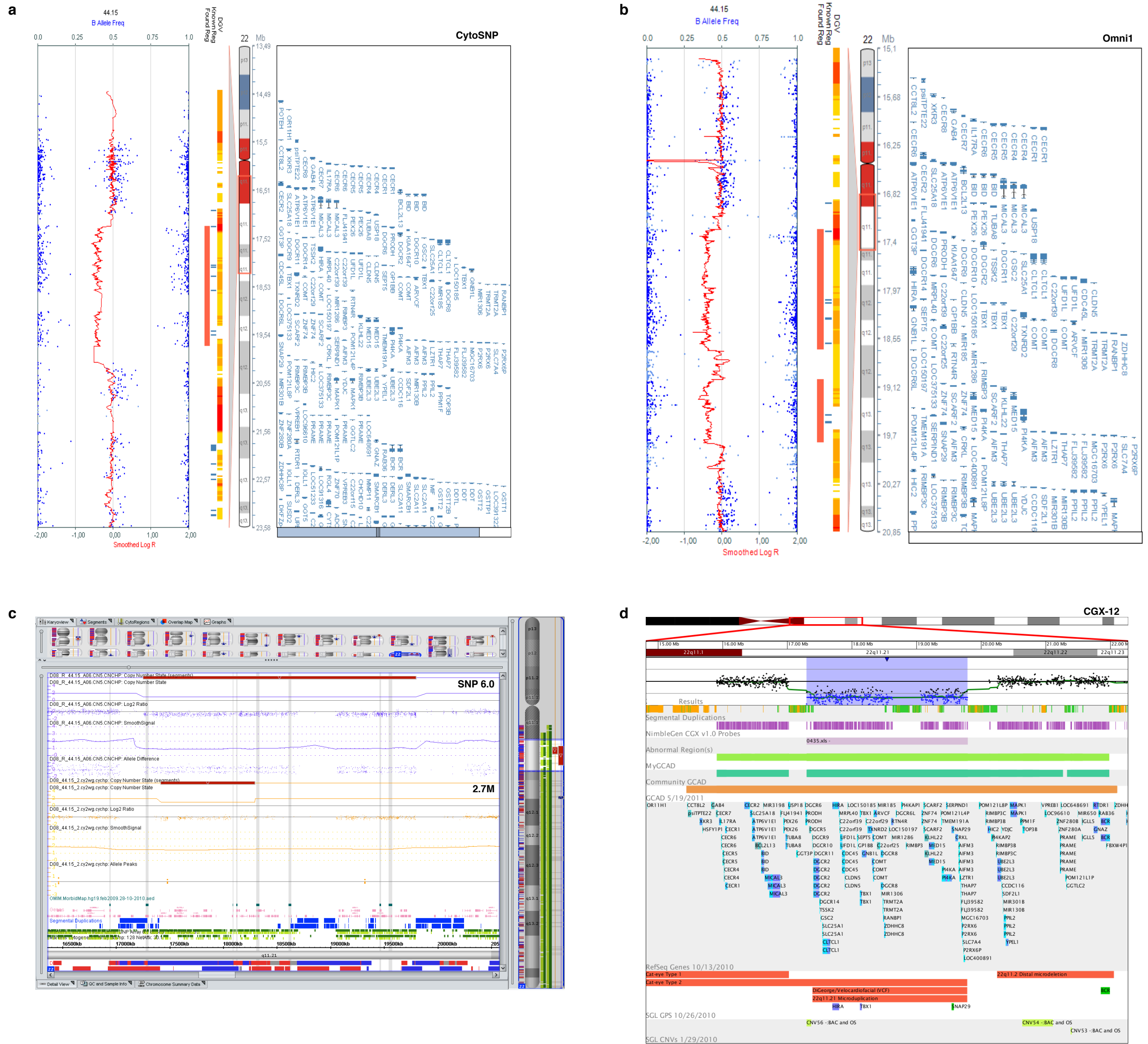


Figure 4-S4 : Pathogenic CNV detected in patient 44.15, as visualized in each software.

22q11.21 loss. (a) CytoSNP. (b) Omni1. (c) SNP 6.0 and 2.7M. (d) CGX-12.

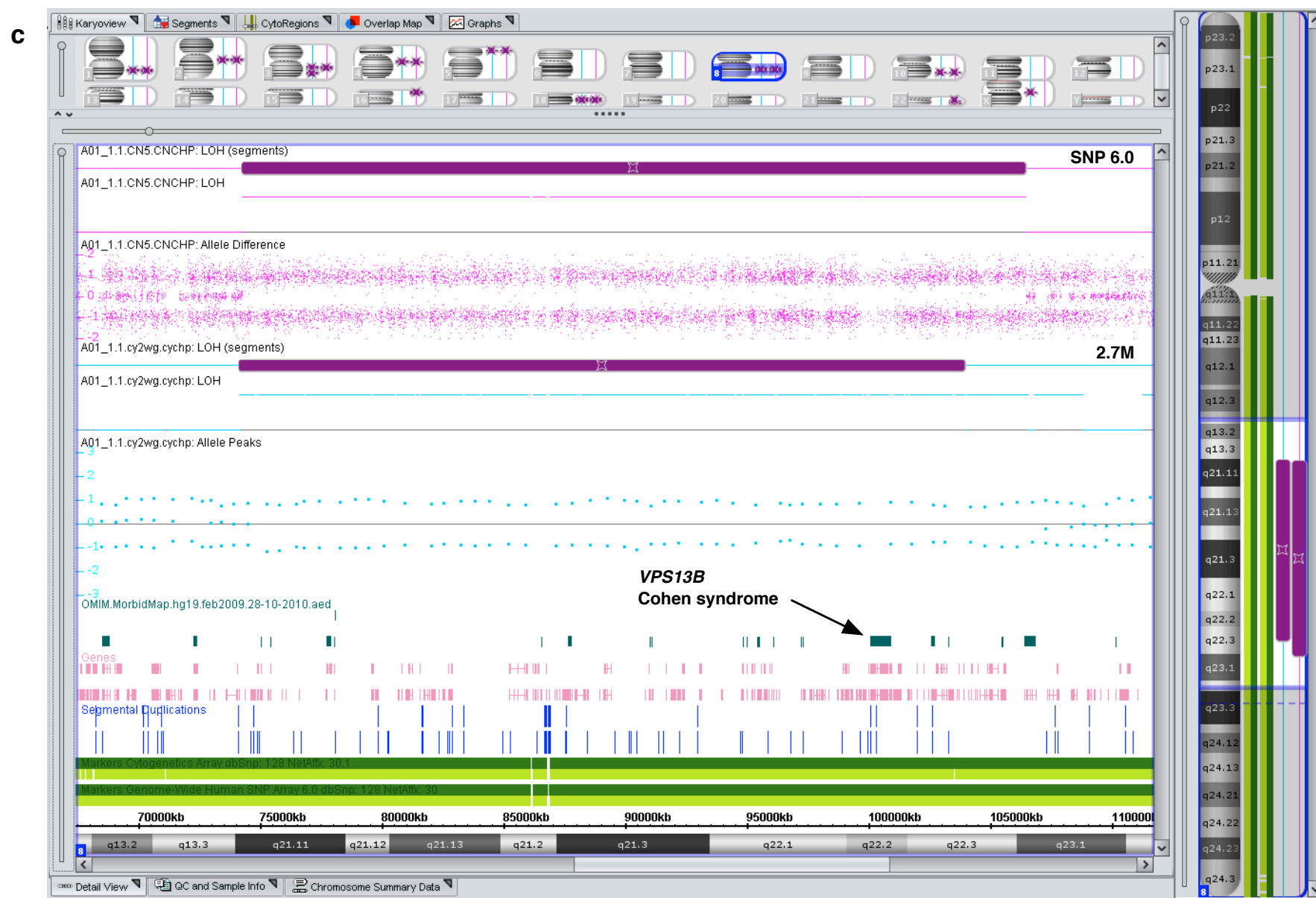
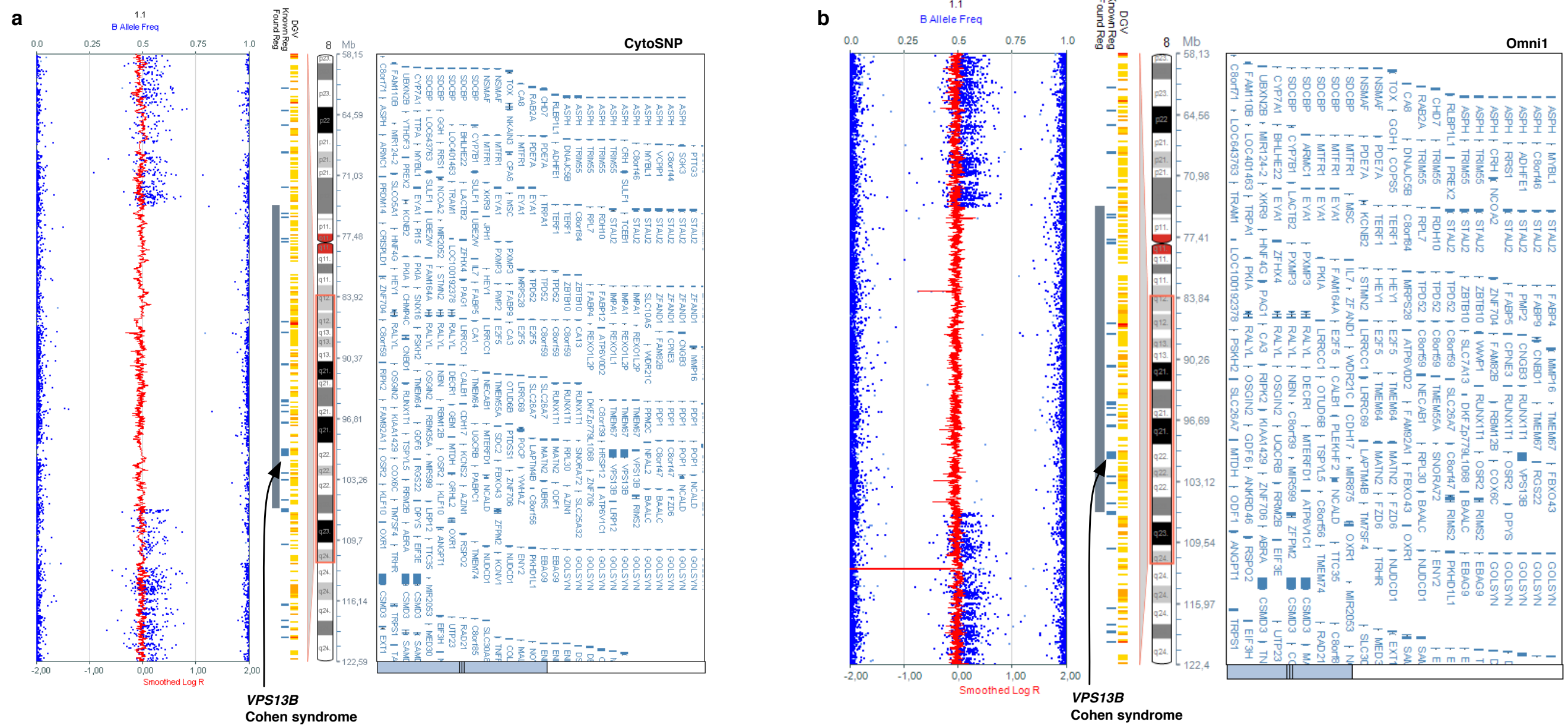


Figure 4-S5 : Informative LCSH detected in patient 1.1, as visualized in each software.

32.2 Mb LCSH in 8q21.1q23.1, encompassing *VPS13B*. (a) CytoSNP. (b) Omni1. (c) SNP 6.0 and 2.7M.

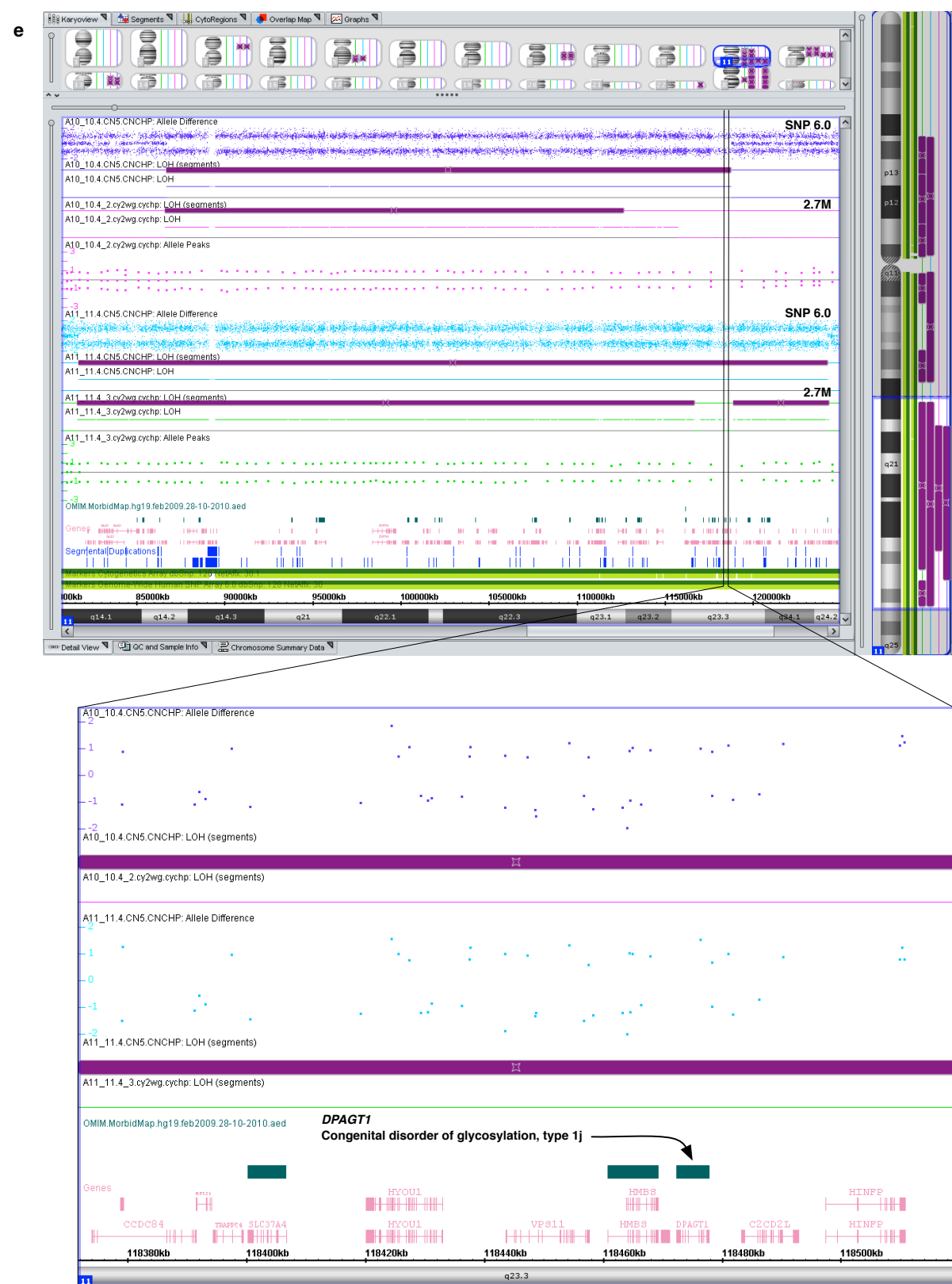
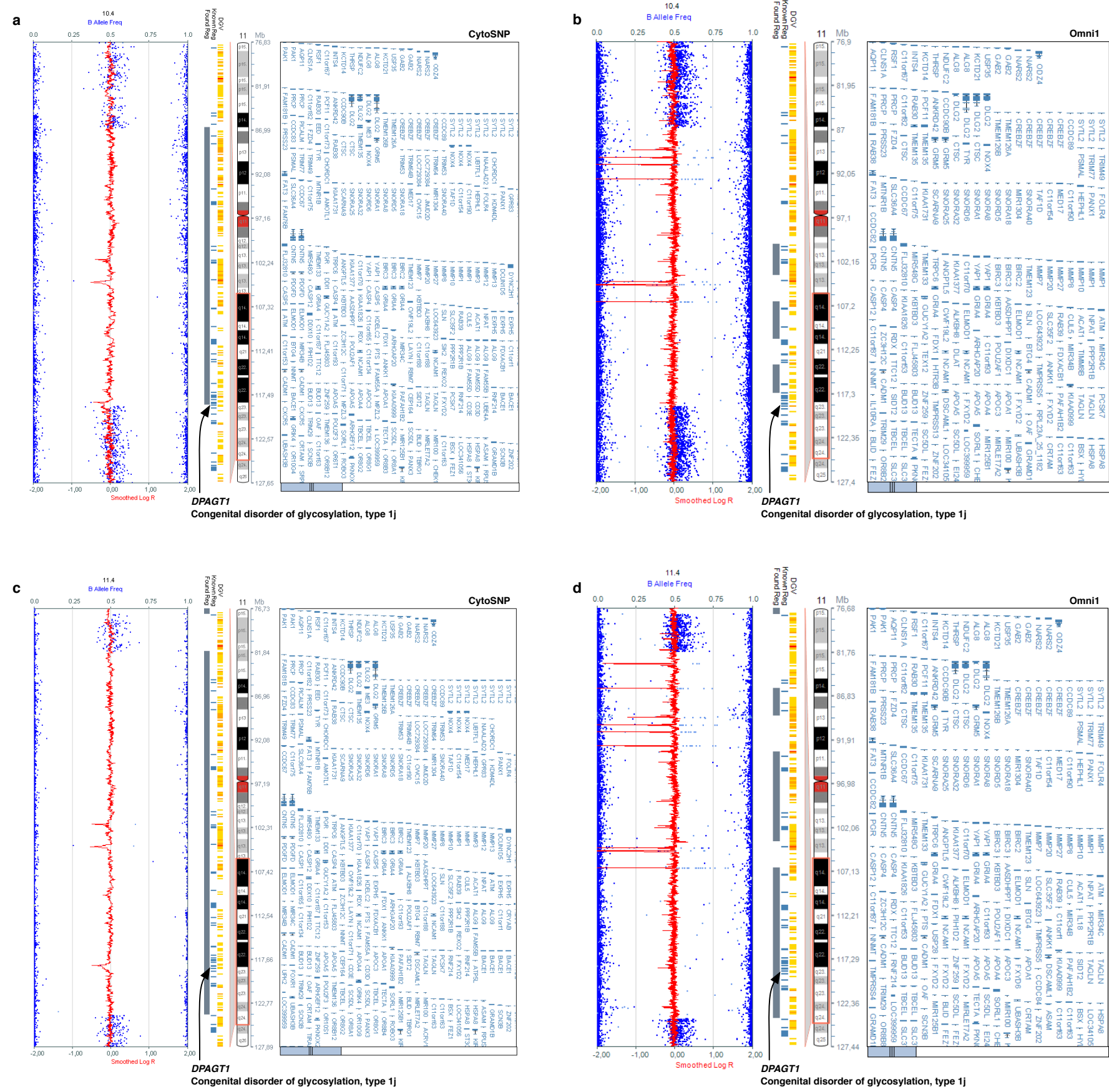


Figure 4-S6 : Informative LCSHs detected in patients 10.4 and 11.4, as visualized in each software.

32 Mb and 42 Mb LCSHs in 11q14.1q24.2, encompassing *DPAGT1*. (a-b) Patient 10.4. (a) CytoSNP. (b) Omni1. (c-d) Patient 11.4. (c) CytoSNP. (d) Omni1. (e) Patient 10.4 (upper part) and 11.4 (lower part), SNP 6.0 and 2.7M.

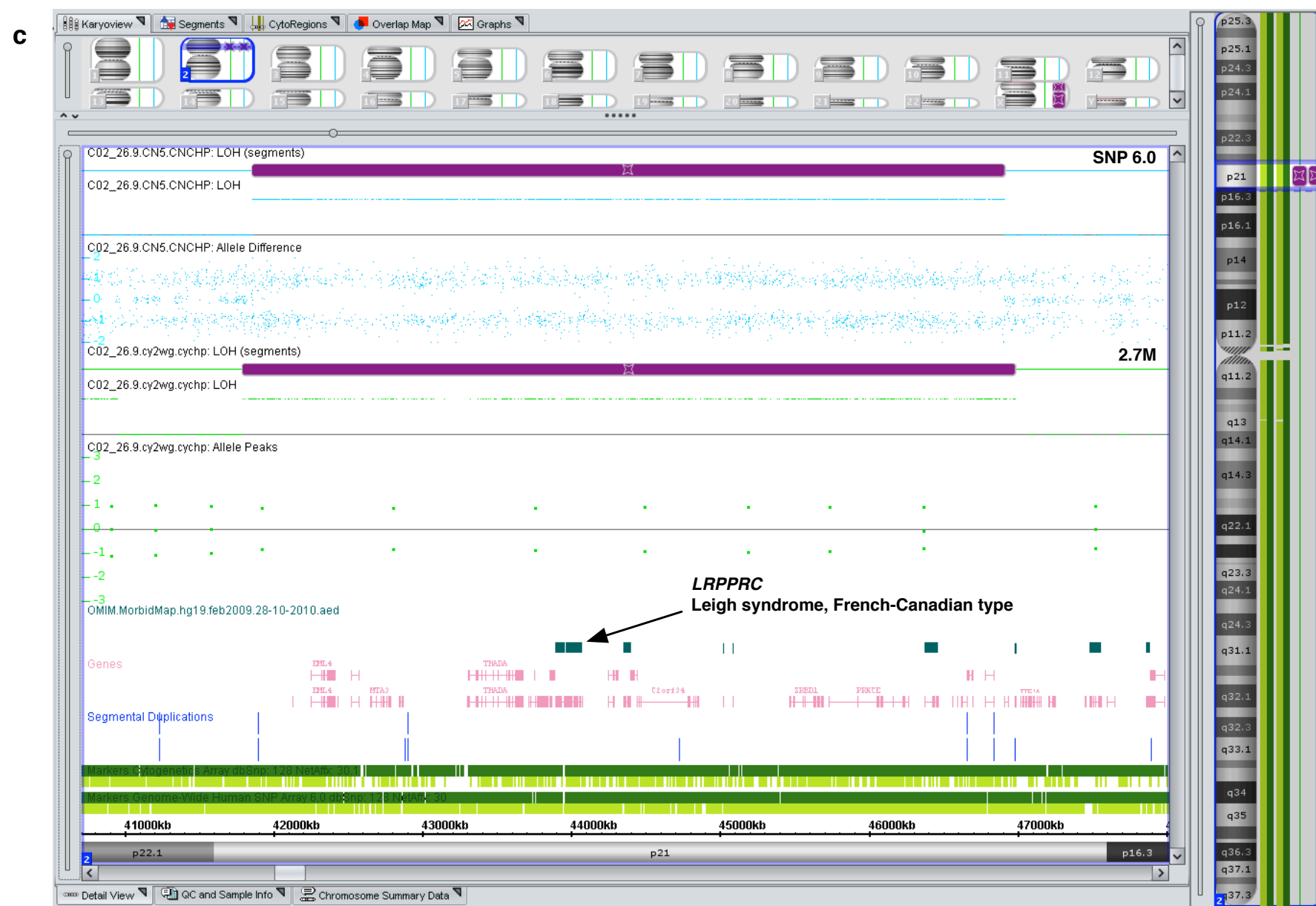
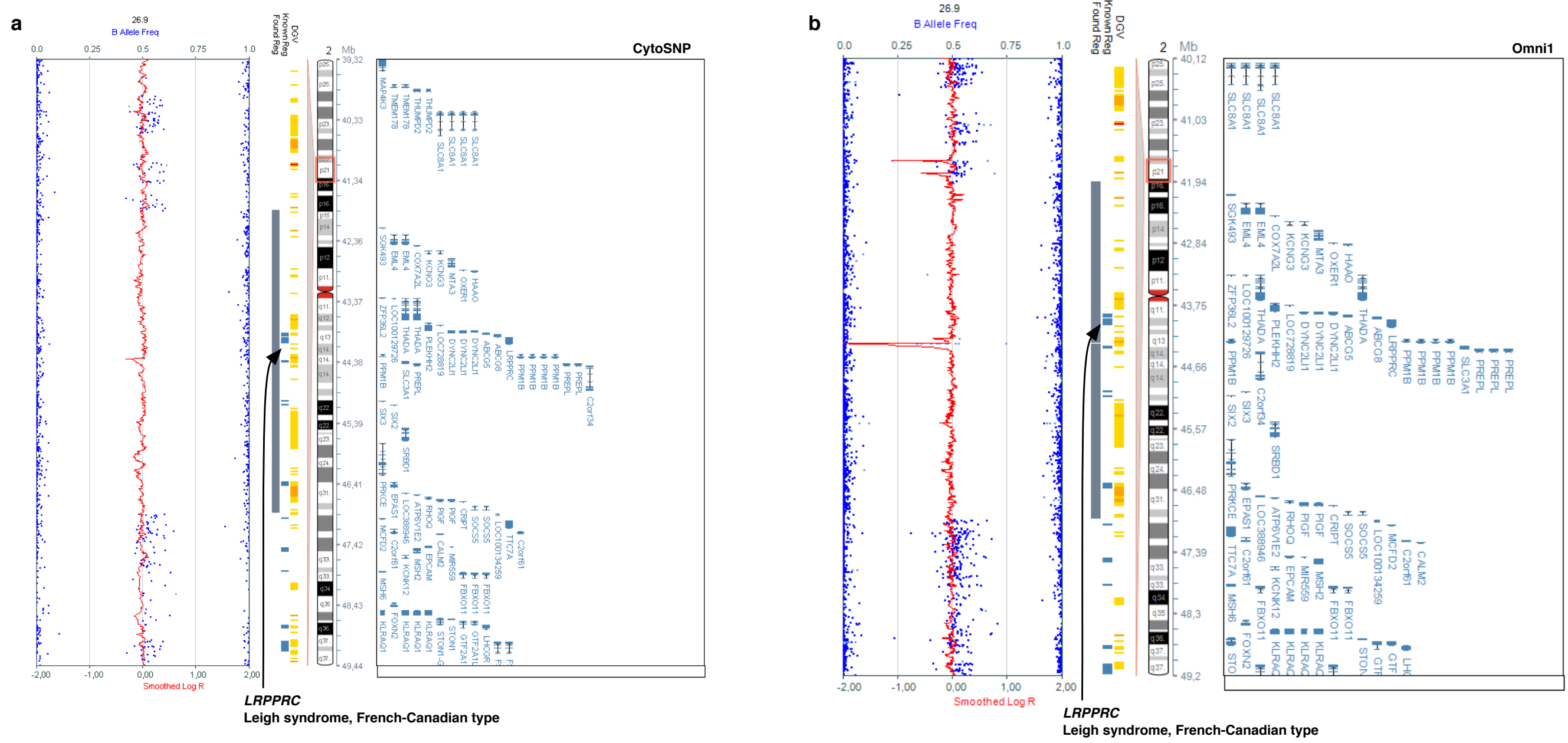


Figure 4-S7 : Informative LCSH detected in patient 26.9, as visualized in each software.

5 Mb LCSH in 2p21, encompassing *LRRPCC*. (a) CytoSNP. (b) Omni1. (c) SNP 6.0 and 2.7M

5 SYNTHÈSE ET CONCLUSION

L'objectif principal de ce projet était d'évaluer l'utilité clinique de diverses plateformes de caryotype moléculaire dans les contextes prénatal et postnatal. Le premier volet a évalué le rendement diagnostique d'une micropuce CGH pangénomique chez des fœtus avec malformations majeures détectées à l'échographie, en plus de déterminer la proportion de fœtus pour lesquels on détecte un CNV de signification clinique incertaine. Le second volet a évalué le rendement diagnostique (CNVs et pertes d'hétérozygotie) de quatre micropuces SNP à très haute résolution chez des enfants atteints de déficience intellectuelle, défini la fréquence de détection de CNVs de signification clinique incertaine chez ces patients, évalué les logiciels fournis par les manufacturiers, et mis en lumière les défis associés à l'utilisation de ces micropuces en contexte clinique.

5.1 CGH pangénomique en prénatal

5.1.1 Rendement diagnostique

Tel que rapporté dans la section 3.4.1, nous avons pu établir le rendement diagnostique d'une micropuce CGH pangénomique à 8,2% dans la population de fœtus avec malformations détectées à l'échographie que nous avons étudiée. La comparaison de nos résultats avec ceux des études prénatales antérieures était difficile puisque certaines ont utilisé des critères d'inclusion variés (âge maternel avancé, anxiété parentale, malformations sévères ou non); des micropuces pangénomiques, ciblées, ou les deux; et n'avaient pas nécessairement effectué un caryotype conventionnel au préalable (section 3.5.1). Nous avons toutefois pu effectuer une comparaison valable en extrayant de ces études les résultats concernant les fœtus avec malformations détectées à l'échographie ayant un caryotype conventionnel normal (Table 3-3). Le rendement diagnostique que nous avons obtenu est comparable à celui retrouvé dans les études rétrospectives (8,1% à 10%), qui pour la plupart étudiaient les fœtus suite à une interruption de grossesse ou une perte fœtale, donc possiblement plus sévèrement atteints. Ceci confirme que, tout comme le caryotype conventionnel, le caryotype moléculaire identifie la plus grande proportion de CNVs pathogéniques chez les fœtus présentant des malformations. Récemment, Wapner *et coll.* ont également obtenu des résultats allant en ce sens, puisqu'ils ont identifié un CNV pathogénique chez 2,8% des fœtus ayant une anomalie à l'échographie, contre seulement 0,5% des fœtus ayant été testés pour une autre raison (âge maternel avancé, histoire familiale, dépistage de trisomie positif) ¹¹⁵. Également, notre rendement diagnostique est plus élevé que celui obtenu avec la majorité des micropuces ciblées (0% à 2,2%) — excepté la plateforme GenoSensor Array 300 qui a donné un rendement autour de 8% — ce qui va dans le sens d'un meilleur rendement avec les micropuces pangénomiques. Enfin, la détection d'anomalies de plus de 5 Mb chez deux fœtus de notre cohorte illustre la capacité du caryotype moléculaire à détecter des anomalies non détectées au caryotype conventionnel à cause d'une mauvaise résolution, ce qui est particulièrement pertinent en contexte prénatal (section 3.5.2). Dans le premier cas (fœtus 47), l'examen a posteriori du caryotype initial a révélé que l'altération était visible à la résolution obtenue (375 bandes), mais qu'elle n'avait pas été repérée. Dans le second cas (fœtus 53),

l'anomalie n'était pas visible au caryotype standard à la résolution obtenue (400 bandes), même après le réexamen minutieux des métaphases initialement analysées.

5.1.2 CNVs de signification clinique incertaine

Nous avons trouvé un résultat de signification clinique incertaine chez 12,2% des fœtus investigués (section 3.4.2), ce qui se situe dans la portion supérieure de l'intervalle (0% à 12,3%) pour les études effectuées chez des fœtus avec malformations, et semble confirmer qu'une augmentation de la densité des micropuces mène aussi à la détection d'un plus grand nombre de CNVs dont la signification clinique est incertaine (section 3.5.1). Ce résultat illustre l'importance du conseil génétique (pré-test et post-test), de même que celle de l'analyse parentale, qui ne peut malheureusement pas toujours être effectuée (section 3.5.3). Il confirme également que l'approche consistant à restreindre la CGH pangénomique aux fœtus présentant des malformations semble appropriée, du moins pour l'instant (section 3.5.4), ce qui est en accord avec les plus récentes recommandations du Collège Canadien de Généticiens Médicaux et de la Société des obstétriciens et gynécologues du Canada ¹¹⁶. En effet, nous avons pu augmenter le rendement diagnostique tout en conservant une proportion de résultats incertains qui est acceptable pour ce groupe de fœtus, mais qui serait problématique dans le cas de grossesses normales. Bien que certains groupes de recherche aux États-Unis supportent l'usage des micropuces en remplacement du caryotype conventionnel pour toutes les grossesses, il est probable que le contexte de pratique différent (couverture privée des soins et des analyses, risque de poursuites en cas de non-diagnostic) influence leur opinion, et nous pensons qu'il n'est éthiquement pas acceptable d'offrir la CGH pangénomique à toutes les patientes enceintes au Canada. L'utilisation d'une micropuce ciblée serait plus appropriée pour les fœtus à plus bas risque, afin de minimiser l'obtention de résultats dont la signification clinique est incertaine.

5.2 Combinaison CGH / géotypage SNP en postnatal

5.2.1 Rendement diagnostique

Premièrement, les résultats obtenus semblent indiquer que l'utilisation d'une micropuce à très haute résolution pour détecter les CNVs ne permet pas d'augmenter le rendement diagnostique (section 4.5.1). En effet, aucune des micropuces n'a détecté un CNV pathogénique qui n'aurait pas été vu sur la micropuce CGH utilisée comme standard de comparaison, qui a permis d'obtenir un diagnostic chez 14,3% des patients, ce qui correspond au rendement auquel on doit s'attendre dans cette population^{29,107,108}. Par ailleurs, les micropuces testées ont toutes détecté les CNVs pathogéniques, mais avec des points de cassure variés, et certaines plateformes ont fragmenté les altérations en segments plus petits (Figures 4-S2, 4-S3 & 4-S4). Or, il est possible que cette fragmentation soit plus représentative de la réalité, puisque dans certains cas la plateforme ayant fragmenté l'altération avait la meilleure couverture de la région génomique en cause. Les anomalies chromosomiques pathogéniques identifiées avaient toutes une taille supérieure à 2 Mb, ce qui permet de confirmer que ces plateformes détectent les grandes anomalies adéquatement, mais ne permet pas de se prononcer quant à leur capacité à détecter des anomalies plus petites. Toutefois, on peut penser qu'elles seraient en mesure de le faire, étant donné qu'elles ont pu détecter les CNVs bénins de petite taille.

Le second aspect du rendement diagnostique concernait les pertes d'hétérozygotie, et dans quelle mesure leur détection pouvait mener à un diagnostic (section 4.5.2). Nous avons identifié un LCSH informatif chez quatre patients, incluant deux frères (Table 4-6 & Figures 4-S5, 4-S6 & 4-S7), et avons pu grâce à cette information identifier la mutation homozygote responsable du phénotype anormal chez trois de ces enfants (patients 10.4, 11.4 et 26.9). L'utilisation d'une micropuce SNP a donc permis d'augmenter le rendement diagnostique de 14,3% à 28,6% par rapport à une micropuce CGH. Ce rendement de près de 15% des patients, et 66% des LCSHs considérés comme étant informatifs est beaucoup plus élevé que ce qui a été obtenu par d'autres groupes de recherche, qui ont identifié une mutation homozygote pathogénique chez 0,08% (5/6500) des patients testés¹¹⁷ et 6,8% à 10% des LCSHs informatifs investigués^{41,118}. Une autre étude effectuée chez 35 patients ayant un résultat

normal au caryotype moléculaire a identifié quatre LCSH informatifs, sans toutefois pouvoir confirmer de mutation homozygote pathogénique ¹¹⁹. Également, même s'il n'a pas été possible d'évaluer directement la capacité des micropuces SNP à détecter la DUP — aucun de nos patients ne présentait cette anomalie — nous avons quand même pu vérifier que l'approche utilisée évitait les faux positifs. Par ailleurs, nous avons pu démontrer que toutes les plateformes, sauf Omni1, pouvaient identifier la consanguinité parentale, et en déterminer le degré de façon assez précise, sans faux positifs encore une fois (Table 4-S5). La pertinence d'obtenir cette information via l'analyse moléculaire est toutefois incertaine, et c'est un aspect important dont le conseil génétique devra tenir compte.

5.2.2 CNVs bénins & de signification clinique incertaine

Même si le rendement diagnostique d'une plateforme constitue l'élément principal permettant de décider de son utilité clinique, il est important de s'assurer que ce rendement n'est pas accompagné d'une spécificité démesurément réduite, c'est-à-dire d'un nombre trop élevé de résultats dont la signification clinique est incertaine. Les études récentes effectuées avec les micropuces à très haute résolution ne rapportent pas cette information ^{117,120,121}, qui est pourtant cruciale en contexte clinique. Une hausse importante du nombre total de CNVs détectés était anticipée dans notre étude, puisque nous voulions exploiter la très haute résolution offerte par trois des quatre micropuces testées pour détecter des CNVs pathogéniques aussi petits que 10 kb. C'est effectivement ce qui s'est produit, particulièrement pour les puces Affymetrix, et ce malgré nos efforts pour réduire cette hausse (section 4.5.1). Nous avons en effet utilisé les différentes options de chacun des logiciels afin de filtrer automatiquement les CNVs en fonction de divers facteurs tels la taille, le contenu en gènes, et la présence de variants bénins connus (section 4.4.2 et Table 4-2). Cela nous a permis d'obtenir, pour toutes les micropuces, un nombre de CNVs bénins inférieur à ce qui a été trouvé par la micropuce CGH de comparaison (Figure 4-1 et Table 4-4). Mais cette approche, bien que partiellement efficace, n'a malheureusement pas suffi à réduire le nombre de CNVs dont la signification clinique est incertaine à un niveau acceptable sur toutes les plateformes (Figure 4-1 et Table 4-4). En effet, la micropuce Omni1 d'Illumina a détecté un nombre de

VOUS plus élevé — mais potentiellement tolérable — que la micropuce CGH, alors que CytoSNP en a détecté un nombre comparable. Mais les puces Affymetrix en ont détecté un nombre qui ne pourrait être acceptable dans un contexte clinique. Selon l'information obtenue du fabricant, un problème de design serait à l'origine de ce résultat pour la micropuce 2.7M, mais ce n'est vraisemblablement pas l'unique raison puisque la plateforme SNP 6.0 a également détecté un nombre inacceptable de VOUS. De plus, la majorité de ces CNVs étaient des gains (Figure 4-2b) et n'ont pas été confirmés sur une deuxième plateforme (Table 4-3 & Table 4-5), alors que les CNVs bénins ne présentaient pas de biais (Figure 4-2a) et ont été confirmés à un taux similaire aux micropuces Illumina et NimbleGen (Table 4-3 & Table 4-5). Il est donc probable qu'il y ait un problème affectant spécifiquement la détection des gains sur cette plateforme, mais nous n'avons pas été en mesure d'en identifier l'origine.

5.2.3 Logiciels de visualisation

Un autre aspect ayant un impact important sur l'utilité clinique des différentes plateformes est l'efficacité avec laquelle l'interprétation des résultats peut être effectuée. Comme cette étape requiert l'utilisation de logiciels pour visualiser les résultats, nous avons évalué les options disponibles dans les deux logiciels de visualisation accompagnant les micropuces, de même que leur facilité d'utilisation, et nous avons vérifié s'ils permettaient d'effectuer l'interprétation des anomalies détectées de façon adéquate. Bien que ChAS et KaryoStudio offrent plusieurs options intéressantes et offrent un certain degré de configuration permettant de l'adapter aux besoins du laboratoire (Table 4-S6), aucun n'arrive au niveau du logiciel actuellement utilisé (section 4.5.3). ChAS est hautement configurable, et permet tout de même d'effectuer une interprétation des anomalies détectées de manière très efficace, alors que KaryoStudio n'offre que peu d'items pouvant être configurés selon les besoins, et est plus efficace pour produire des rapports de résultats que pour effectuer une interprétation adéquate. La lacune la plus importante — qui touche les deux logiciels — demeure l'absence d'une base de données pour enregistrer les résultats. Une solution à ce problème serait d'établir un lien avec une base de données externe, ce qui demanderait toutefois d'avoir recours à des spécialistes en informatique pour mettre le tout en place. Une autre solution pourrait être de

remplacer le logiciel de visualisation par un logiciel commercial prenant en charge diverses micropuces, et offrant une base de données. Ces deux options entraîneraient toutefois un coût supplémentaire.

5.2.4 Défis & charge de travail associés aux micropuces SNP

Nous avons évalué la charge de travail et les défis spécifiques aux micropuces SNP que nous avons testées, puisque ces éléments doivent être pris en considération avant de décider si une plateforme peut être utilisée efficacement ou non dans un contexte clinique. D'abord, il est clair que pour trois des quatre micropuces (SNP 6.0, 2.7M et Omni1), la charge de travail associée à la détection des CNVs serait insoutenable dans un contexte clinique (section 4.5.1). La raison principale, dans le cas des puces Affymetrix, est le nombre total de CNVs détectés — qui est beaucoup trop élevé — alors que pour Omni1, c'est le nombre total de CNVs — élevé mais potentiellement tolérable advenant une amélioration du rendement diagnostique — combiné à un logiciel inefficace qui rendrait son utilisation trop lourde. De plus, les micropuces Affymetrix ont détecté quelques CNVs faux positifs — facilement identifiables — et les plateformes Illumina quelques faux négatifs — plus difficile à identifier — qui ont nécessité un temps d'interprétation plus important, ce qui augmenterait significativement la charge de travail.

Par contre, même si elle s'est avérée imparfaite (voir section 5.3), l'approche de filtrage et de seuils différentiels que nous avons utilisée (4.4.2 et Table 4-2) a tout de même permis de réduire considérablement le nombre total de CNVs à interpréter (Tables 4-S1 et 4-S2), tout en ciblant les régions potentiellement pathogéniques avec des seuils comparables à ceux actuellement utilisés sur la micropuce CGH. Toutefois, l'utilisation des micropuces SNP à leur plein potentiel en cytogénétique clinique demandera une optimisation des paramètres d'utilisation. Pour l'instant, combiner une couverture ciblée à une couverture pangénomique semble être l'approche la plus avantageuse, puisque cette méthode permettrait de détecter les CNVs de très petite taille dans les régions connues comme étant cliniquement significatives, tout en réduisant le nombre de VOUS détectés. Certaines plateformes offrent un tel design, mais il est aussi possible de cibler *in silico* les régions génomiques voulues sur une micropuce

pangénomique, ce qui nous apparaît comme plus fructueux à long terme étant donné la rapidité avec laquelle les connaissances évoluent. Compléter les bases de données de variants bénins contribuera également à faciliter l'interprétation des résultats pour permettre de réduire les seuils de détection, tout comme l'utilisation de méthodes d'interprétation automatisées, que certains groupes commencent à développer ¹²², et qui s'inscrit dans le même esprit que ce que nous avons tenté de faire.

Finalement, la hausse du rendement diagnostique que nous avons pu obtenir grâce à la détection des pertes d'hétérozygotie semble appuyer l'implantation de cette analyse dans les laboratoires cliniques. Toutefois, il faudra préalablement déterminer les paramètres d'utilisation optimaux, de même que cibler les patients pour qui cette analyse est effectuée, ou pour qui les résultats sont analysés, puisque cela représente un travail imposant. À un seuil de 5 Mb, le plus petit des LCSHs peut en effet contenir plus de 30 gènes. Certains LCSHs seront plus grands, et parfois plus nombreux, particulièrement dans les cas de consanguinité parentale. Cibler le gène à investiguer risque d'être particulièrement difficile dans ces cas, tel qu'illustré par notre propre échec à obtenir un diagnostic chez une famille dont les parents étaient cousins (sections 4.4.7 et 4.5.2).

5.3 Limites et améliorations

Certaines contraintes ont affecté les conditions dans lesquelles nous avons effectué cette étude, et ont pu avoir un impact sur les résultats obtenus.

Premièrement, l'étude prénatale ayant été effectuée dans un contexte clinique, il n'a pas été possible d'obtenir l'ADN parental dans tous les cas où cela était requis pour l'interprétation des résultats. Dans deux de ces cas (fœtus 12 et 51), il était donc difficile de considérer avec certitude le CNV identifié comme étant pathogénique, puisque nous n'avons pas pu confirmer si un parent était porteur ou non (Table 3-2). C'est pourquoi chacun de ces CNVs a été considéré pathogénique probable, et classé parmi les VOUS, ce qui a probablement contribué à augmenter le taux que nous rapportons. La comparaison avec les résultats obtenus dans les autres études mentionnées (Table 3-3) demeure valide, puisque plusieurs de celles-ci n'ont pas pu obtenir l'ADN parental pour tous les fœtus (Van den Veyver *et coll.*, Kleeman *et coll.*, Coppinger *et coll.*, Tyreman *et coll.*, et Vialard *et coll.*). Également, bien que le fait de savoir si un CNV est hérité ou non puisse aider à interpréter les résultats, ceci ne constitue ni un critère absolu, ni une nécessité dans tous les cas (section 3.5.3). Le taux de VOUS que nous rapportons demeure donc valable, puisqu'il représente le taux réel auquel on peut s'attendre dans un contexte clinique — imparfait — où il n'est pas toujours possible d'obtenir l'ADN des deux parents.

Plusieurs raisons expliquent le fait qu'un échantillon des parents n'ait pas été prélevé d'emblée pour tous les fœtus testés. D'abord, il faut rappeler que cette étude a été faite dans le contexte de la réalité clinique. Ainsi, la mise en place d'une procédure de prélèvement automatique des parents pour tous les CGH sur micropuce effectués entraînerait un nombre très élevé d'échantillons à traiter et à stocker au laboratoire. L'analyse de l'ADN parental n'est pas nécessaire si le résultat du fœtus est normal ou si le CNV détecté est clairement bénin, ce qui était le cas pour la majorité des fœtus que nous avons testés (71%, 35/49). Le laboratoire a donc choisi de ne pas prélever d'emblée les parents afin de faciliter la gestion des échantillons et du travail technique. Par contre, dans certains cas, les parents n'ont pas souhaité être prélevés par la suite pour clarifier la signification clinique du CNV trouvé. Aussi, certaines patientes ayant bénéficié de l'analyse n'étaient plus en contact avec le père de l'enfant, ce qui

rendait impossible l'obtention d'un prélèvement. Finalement, l'analyse par CGH sur micropuce n'est pas toujours la méthode de choix pour évaluer si un parent est porteur de l'anomalie retrouvée chez le fœtus. En effet, il est souvent plus avantageux d'effectuer les analyses parentales par FISH, lorsque possible, afin de permettre d'éliminer la présence d'un remaniement équilibré chez un parent et ainsi mieux préciser le risque de récurrence. Or, ces analyses s'effectuent sur des lymphocytes en culture, requérant l'obtention d'un échantillon frais. Prélever d'emblée tout parent d'un fœtus pour lequel une analyse de CGH sur micropuce est faite nécessiterait donc qu'un prélèvement pour mise en culture soit aussi fait, ce qui serait très lourd pour le laboratoire. Toutefois, l'analyse des parents constitue un aspect très important pour réduire le taux de résultats incertains obtenu (section 3.5.1), et il est donc essentiel de tenter d'optimiser le pourcentage d'analyses parentales obtenues.

Par contraste, la disponibilité de l'ADN des deux parents était un critère d'inclusion dans l'étude CGH / SNP, et c'est la nécessité d'avoir l'ADN parental qui pourrait constituer une contrainte en clinique. Le taux de VOUS que nous rapportons pourrait par conséquent sous-estimer le taux réel obtenu en contexte clinique. Nous avons exclu les CNVs hérités de notre analyse (section 4.3.5) à cause du nombre élevé d'altérations détectées, ce qui ne serait pas acceptable dans un contexte clinique (section 4.5.1). Or, les résultats que nous avons obtenus montrent qu'il serait probablement difficile d'effectuer une interprétation adéquate des résultats sans les comparer avec les résultats obtenus chez les parents. Contrairement à l'analyse CGH, pour laquelle on peut procéder à l'analyse parentale seulement en cas de besoin, les micropuces Affymetrix — même en excluant les CNVs hérités — ont identifié au moins un VOUS chez pratiquement tous les patients (Table 4-4). Les puces Illumina ayant identifié un VOUS chez moins d'un tiers des patients, on pourrait penser que ce problème serait moins important pour ces plateformes. Or, 81% des CNVs détectés sur la micropuce Omni1 étaient hérités (Table 4-S2), et cette plateforme a détecté un nombre total de CNV s'approchant de celui détecté sur les micropuces Affymetrix. Il est donc prévisible que l'ADN parental soit nécessaire pour tous les patients, ce qui pourrait s'avérer problématique dans certains cas, et triplerait les coûts d'analyse.

Ensuite, l'approche de filtrage des CNVs bénins que nous avons utilisée comporte certaines limites. Premièrement, il n'a pas été possible de l'appliquer pour les micropuces

Illumina puisque KaryoStudio ne permet pas de restreindre le filtre aux régions sans gènes OMIM (section 4.3.4). Il est probable que l'ajout d'une telle option dans KaryoStudio modifierait nos conclusions concernant la performance des micropuces de ce manufacturier. Ensuite, même si ChAS permet d'examiner les détails de chacun des CNVs rapportés dans DGV (gain/perte, nom et lien vers les études rapportant le variant), le logiciel ne prend pas ces informations en considération lors du masquage des CNVs détectés chez les patients. Il est peu probable que ceci ait mené au masquage d'altérations potentiellement pathogéniques, puisque le filtre était appliqué seulement dans les régions exemptes de gènes OMIM. On ne peut toutefois exclure que cela se soit produit et il faudra résoudre cette limitation du logiciel advenant une utilisation en clinique.

De plus, il est possible que l'utilisation des données fournies par la compagnie Affymetrix comme ADN de référence ait eu un impact sur la qualité de la détection des CNVs par les micropuces SNP 6.0 et 2.7M. Comme nous l'avons vu à la section 5.2.2, la proportion de gains et d'altérations non-confirmées était très élevée pour ces deux micropuces, et il est possible que ce problème ait été causé ou amplifié par les différences entre la population de qui provenait l'ADN de référence et la population que nous avons testée, de même que par des variations techniques. Une solution à explorer pour remédier à ce problème serait l'utilisation, comme ADN de référence, de données obtenues avec des individus provenant de la population étudiée, et testés dans les mêmes conditions : même laboratoire, mêmes appareils, mêmes lots de réactifs et de micropuces. Toutefois, comme cet ADN de référence doit provenir d'au moins 44 patients, c'est une approche coûteuse qu'il n'a pas été possible d'utiliser dans notre étude, mais qui pourrait se justifier une fois la décision prise d'implanter la micropuce SNP.

Enfin, la taille de notre échantillon limite potentiellement l'exactitude du rendement diagnostique que nous avons obtenu. En effet, il est possible que celui-ci ait été surestimé, particulièrement dans le cas des LCSHs. Nous avons obtenu un diagnostic grâce à la détection des pertes d'hétérozygotie chez trois patients, résultant en un rendement diagnostique de 14,3%, ce qui est beaucoup plus élevé que le taux rapporté jusqu'à présent dans la littérature^{117,119}. Cependant, il est aussi possible que notre résultat soit plus représentatif du rendement que l'on pourrait obtenir au Québec — étant donné l'importance de l'effet fondateur chez une

grande partie de la population — et dans les populations où la consanguinité parentale est fréquente.

5.4 Perspectives

5.4.1 Implantation en clinique

L'implantation du caryotype moléculaire dans les laboratoires de cytogénétique clinique nord-américains n'est pas complète — bien qu'avancée dans la majorité des grands centres urbains — et cette analyse est loin d'avoir totalement remplacé le caryotype conventionnel. D'abord, si les lignes directrices actuelles concernant son utilisation sont relativement claires, particulièrement pour la population pédiatrique, il y a toujours de vifs débats quant aux femmes à qui offrir l'analyse dans un contexte prénatal. La majorité des experts s'entendent pour dire qu'il est actuellement préférable de restreindre l'analyse aux fœtus à haut risque d'anomalies chromosomiques, mais il est concevable qu'elle soit éventuellement offerte à toutes les femmes enceintes pour qui le caryotype et/ou la détection d'aneuploïdie par FISH sont indiqués. Il faudra toutefois que l'avancement des connaissances sur le génome humain et ses variations ait permis la réduction du taux de VOUS en deçà du taux obtenu au caryotype conventionnel et à la FISH interphasique avant que le caryotype moléculaire devienne l'analyse de choix en contexte prénatal. Une réinterprétation des résultats obtenus dans notre cohorte de fœtus, avec la littérature la plus récente, indique qu'il y a déjà eu une amélioration considérable à cet égard. En effet, cette réanalyse nous a permis de réduire de moitié le nombre de VOUS (6/49 à 3/49), ce qui résulte en un taux de résultats incertains de 6,1%. Wapner *et coll.* ont également pu réduire de moitié la proportion de résultats incertains en réinterprétant les VOUS détectés jusqu'à cinq ans plus tôt ¹¹⁵. L'impossibilité à détecter la triploïdie sur les plateformes CGH demeure toutefois un inconvénient non négligeable et représente un argument important en faveur des micropuces SNP pour effectuer le caryotype moléculaire en contexte prénatal.

Ensuite, il n'existe pas de consensus quant à la plateforme idéale, et chaque laboratoire doit donc faire un choix en fonction de ses besoins, des ressources disponibles, des populations testées et de son confort avec les résultats incertains. Des recommandations quant à la performance attendue des plateformes de caryotype moléculaire ont été publiées et peuvent en faciliter l'évaluation et le choix ¹²³. Peu importe la plateforme choisie, celle-ci doit être dûment validée par le laboratoire avant d'être utilisée ³⁰.

Enfin, malgré les avantages certains offerts par les micropuces SNP, il n'existe actuellement pas de recommandations quant à la détection des pertes d'hétérozygotie dans un contexte clinique. Il est envisageable d'établir des paramètres précis et efficaces pour la détection de la disomie uniparentale assez rapidement. Par contre, l'approche utilisée pour les CNVs — avec des paramètres de détection uniformes pour tous les patients — pourrait ne pas fonctionner pour les LCSHs, et il sera probablement plus productif de les ajuster à chaque situation, par exemple en ciblant certaines régions spécifiques en fonction du phénotype du patient. D'autres études seront nécessaires pour établir des lignes directrices claires.

5.4.2 Interprétation des résultats : les défis

Le défi le plus important de l'implantation clinique des micropuces est associé à l'interprétation des résultats. Malgré les efforts pour utiliser divers critères d'interprétation des CNVs tels que la taille, l'hérédité, le contenu en gènes, la présence chez des individus sains ou affectés (sections 3.3.2, 3.5.3 et 4.3.5), la mise au point d'un algorithme pour établir la signification clinique des CNVs s'avère difficile. L'expérience clinique montre que les trois catégories « classiques » de CNVs — pathogénique, bénin et variant de signification clinique incertaine (VOUS) — ne permettent pas de rapporter les résultats avec précision dans tous les cas, la catégorie VOUS servant un peu de catégorie fourre-tout pour rapporter tout CNV qui n'entre pas dans l'une des deux autres catégories au moment du rapport. Or, il existe de grandes variations quant à la probabilité qu'un VOUS soit pathogénique ou non, et une subdivision de cette catégorie s'impose afin de mieux représenter le spectre des VOUS détectés et de rapporter la meilleure information disponible au médecin requérant. L'American College of Medical Genetics a récemment publié des recommandations concernant les catégories de CNVs, incluant pathogénique et bénin, mais aussi trois sous-catégories de VOUS : pathogénique probable, bénin probable, et signification clinique incertaine ³¹. L'utilisation de ces catégories permet de mieux refléter la signification clinique réelle des CNVs rapportés, même si elle demeure imparfaite.

Bien qu'il s'agisse d'un idéal souhaitable, il est peu probable que l'on puisse établir un algorithme d'interprétation capturant toutes les possibilités et permettant d'éliminer les VOUS

totalemment. En effet, la complexité du génome et la multiplication des bases de données contribuent à rendre l'atteinte de cet objectif difficile (section 1.2.4.1). Le modèle du « double-hit »^{124,125} n'est qu'un des exemples qui illustrent bien la raison pour laquelle l'interprétation des résultats sera difficilement automatisable et continuera de requérir un bon niveau d'expertise et de jugement de la part du cytogénéticien. En effet, comment systématiser l'interprétation de « paires de CNVs »? Le recensement de ces CNVs à lui seul représente un travail colossal. Il est donc probable que des CNVs que l'on classe aujourd'hui comme étant bénins puissent en fait être pathogéniques chez les individus portant une seconde altération ou même une mutation ponctuelle. Également, le processus d'interprétation des CNVs s'attarde principalement au contenu en gènes. Pourtant, on sait déjà que la présence de remaniements chromosomiques dans certaines régions intergéniques contenant des éléments régulateurs peut altérer l'architecture régulateur-gène(s) (section 1.2.3) et résulter en un phénotype anormal. Or, il est beaucoup plus complexe d'effectuer des corrélations phénotype-CNV dans ces régions puisque leur structure et leur fonction sont moins bien décrites. Toutefois, un consortium s'y est récemment attaqué et a publié un catalogue des éléments d'ADN dans le génome humain (ENCODE : *Encyclopedia of DNA Elements*)¹²⁶, et Savarese *et coll.* ont mis au point une micropuce aCGH ciblant 1250 *enhancers*¹²⁷, ce qui devrait contribuer à accumuler des données et faciliter l'interprétation des CNVs identifiés dans ces régions. Il faudra donc éventuellement tenir compte de ces informations et cesser de se limiter aux régions géniques dans la recherche de nouveaux CNVs pathogéniques.

De plus, même si l'accessibilité aux bases de données est de plus en plus facilitée, l'information disponible n'en est pas moins éparpillée, plusieurs groupes ayant créé leur propre base de données afin de pallier aux défauts de celles déjà existantes. On se retrouve donc aujourd'hui avec une quinzaine de bases de données publiques¹²⁸, en plus des bases commerciales, répertoriant des CNVs identifiés dans diverses populations — « normales » ou affectées de différents désordres — et fournissant une information variable dont la validité est parfois incertaine: phénotype, hérédité, points de cassure, nombre d'individus porteurs... Ces informations ont une valeur inestimable, mais il est devenu difficile d'en faire le tour d'une manière systématique lorsque vient le temps d'interpréter un CNV chez un patient. Chaque laboratoire doit donc établir son propre processus d'interprétation qui tienne compte de la

micropuce et du logiciel de visualisation utilisés, et des bases de données disponibles dans ce dernier. Tel que mentionné dans la section 1.2.4.2, l'agrégateur *UCSC Genome Browser* est également très utile pour réduire le nombre de sites à consulter, puisque les plus importantes bases de données peuvent y être visualisées.

Malgré tout, d'autres outils pourront faciliter le travail d'interprétation en améliorant les corrélations génotype-phénotype et en rendant le processus plus uniforme pour tous les patients. Les logiciels permettant d'automatiser la recherche dans les bases de données phénotypiques telles que OMIM Clinical synopsis, par exemple, présentent des avantages certains, particulièrement pour l'interprétation des VOUS et des LCSHs. L'outil web développé par Wierenga *et coll.* pour la priorisation des gènes retrouvés dans les régions d'homozygotie identifiées chez des familles avec consanguinité en est un bon exemple ¹²⁹. Celui-ci peut être utilisé tant pour cibler les gènes à investiguer chez un patient avec un LCSH, que pour préciser la signification clinique probable d'un CNV rare ou d'un VOUS. Par ailleurs, un consortium s'est récemment affairé à élaborer un processus basé sur des preuves pour rendre l'interprétation des CNVs plus systématique, peu importe le laboratoire ou la plateforme utilisée ¹³⁰. Bien que cette méthode ne soit pas un algorithme d'interprétation en tant que tel, mais plutôt une évaluation de la pathogénicité des CNVs basée sur les différents niveaux de preuve disponible dans la littérature, il offre des balises pouvant guider les cytogénéticiens afin que la signification clinique de chaque altération soit évaluée et rapportée de façon adéquate, et qu'elle soit bien comprise par les médecins requérants. Le but ultime de ce consortium est d'établir une liste de CNVs pathogéniques et bénins qui couvre le génome en entier et à laquelle on peut référer pour l'interprétation. Cette liste mise à jour est disponible en ligne: www.ncbi.nlm.nih.gov/projects/dbvar/ISCA.

Les connaissances actuelles sur le génome humain laissent entrevoir que les efforts de systématisation dans l'interprétation des CNVs parviendront difficilement à capturer toutes les altérations pathogéniques. Il deviendra de plus en plus difficile de tenir compte de toutes les possibilités dans le choix des critères à utiliser pour départager les altérations pathogéniques de celles qui sont bénignes ou qui sont des facteurs de risque pour le développement de certaines pathologies. Même en se basant sur des données probantes pour le choix de ces

critères, une part de subjectivité demeure et il sera important de bien comprendre les limites associées à chacun des algorithmes et des outils d'interprétation qui seront utilisés.

5.4.3 Séquençage à très haut débit

Les CNVs ont d'abord été identifiés grâce aux techniques de cytogénétique moléculaire (sections 1.4 et 3.2). Toutefois, le séquençage à très haut débit (dit « de nouvelle génération »), d'abord mis au point et utilisé en biologie moléculaire pour détecter les SNPs et les mutations ponctuelles, peut également être utilisé pour détecter les variations de structure. S'il était auparavant impensable d'utiliser le séquençage de manière pangénomique en contexte clinique, la réduction du coût d'analyse permet maintenant d'envisager cette approche chez les patients pour qui un diagnostic n'a pu être obtenu avec les méthodes standards d'analyse génétique. Cette technologie pourrait donc éventuellement remplacer le caryotype moléculaire — tant l'aCGH que les micropuces SNP — et le caryotype conventionnel. Elle permettrait, en théorie, de combiner la détection des remaniements chromosomiques déséquilibrés et équilibrés, à celle des pertes d'hétérozygotie, de la disomie uniparentale, et des mutations ponctuelles, tout en fournissant la plus haute résolution possible, soit une paire de bases. Cependant, bien que différentes approches soit actuellement utilisées pour identifier les remaniements chromosomiques par séquençage, aucune d'entre elles ne permet de détecter tous les types de remaniements avec exactitude¹³¹.

La plupart des approches reposent sur un (ou une combinaison de plusieurs) des quatre principes suivants (Figure 5-1): la profondeur de lecture (de l'anglais *read depth*), les paires de lectures (de l'anglais *read pair*, *mate-paired*, ou *paired end*), les lectures fractionnées (de l'anglais *split reads*) et l'assemblage de séquences (de l'anglais *sequence assembly*). Suite au séquençage, deux étapes sont généralement nécessaires pour appliquer chacun de ces principes d'analyse : 1) l'alignement à un génome de référence; 2) l'identification de patrons/signatures discordants indiquant la présence d'un CNV.

Brièvement, la profondeur de lecture consiste à déterminer la couverture de séquençage obtenue pour chaque région génomique, pour ensuite identifier les régions ayant une couverture plus élevée (indiquant une duplication) ou plus basse (indiquant une délétion) que

le génome de référence (Figure 5-1 B). Cette méthode est toutefois sujette à deux importants biais : le pourcentage de GC, qui varie d'une région génomique à l'autre, et la présence de séquences répétées réparties à travers le génome¹³². Le nombre de lectures effectuées étant dépendant de la proportion de GC, il y aura des variations indépendantes du nombre de copies dans la couverture de séquençage, ce qui peut fausser les résultats. Également, l'alignement des lectures à de multiples positions sur le génome de référence peut rendre difficile l'analyse des régions couvrant des séquences répétées. Certains algorithmes excluent donc ces régions de l'analyse pour la restreindre aux séquences uniques, ce qui réduit l'application de cette technique.

Pour la méthode utilisant les paires de lectures, l'ADN est fragmenté au hasard et les fragments sont ensuite sélectionnés en fonction de la taille (Figure 5-1 A)¹³¹. Les deux bouts (paire de lectures) de chaque fragment — d'une taille connue — sont séquencés puis alignés sur un génome de référence. Une discordance avec la distance attendue entre les paires de lectures indique la présence d'un CNV à l'intérieur de la paire. Une distance plus courte qu'attendue indique une duplication ou une insertion, alors qu'une distance plus longue indique une délétion. Puisque le sens des lectures est également disponible, cette méthode permet l'identification de remaniements équilibrés, de même que la localisation et l'orientation des duplications, en plus de détecter les remaniements déséquilibrés. La précision du nombre de copies est excellente avec cette méthode. Par contre, la résolution obtenue n'est pas toujours très élevée, et la détection des remaniements impliquant les régions riches en séquences répétées est difficile¹³¹.

Le principe des lectures fractionnées est similaire à celui des paires de lectures, excepté que tout le génome est séquencé puis aligné sur un génome de référence^{131,132} (Figure 5-1 C). C'est l'alignement d'une lecture à plus d'un endroit dans le génome qui indique que cette séquence contient le point de cassure du remaniement, séparant la lecture en deux segments ou plus (d'où l'appellation *split read*). Cette méthode ne peut donc être utilisée que pour détecter les remaniements impliquant des séquences uniques. Une autre limitation associée à cette méthode est la capacité informatique nécessaire pour analyser l'énorme volume de données généré¹³². Une alternative consiste donc à combiner cette méthode avec celle des lectures pairées pour diminuer la quantité de données à traiter¹³³.

La dernière méthode repose sur l'assemblage *de novo* de séquences chevauchantes pour reconstituer le génome en entier, qui est ensuite comparé avec un génome de référence pour identifier les CNVs (Figure 5-1 D) ^{131,132}. Dans les faits, il n'est pas encore possible de réaliser l'assemblage strictement par chevauchement, à cause des nombreuses séquences répétées ¹³². L'augmentation de la longueur des lectures permettra éventuellement de pallier à ce problème, mais pour l'instant, l'assemblage requiert encore l'utilisation d'un génome de référence. C'est la méthode qui présente le plus de potentiel, puisqu'elle permettra de détecter tous les types de remaniements.

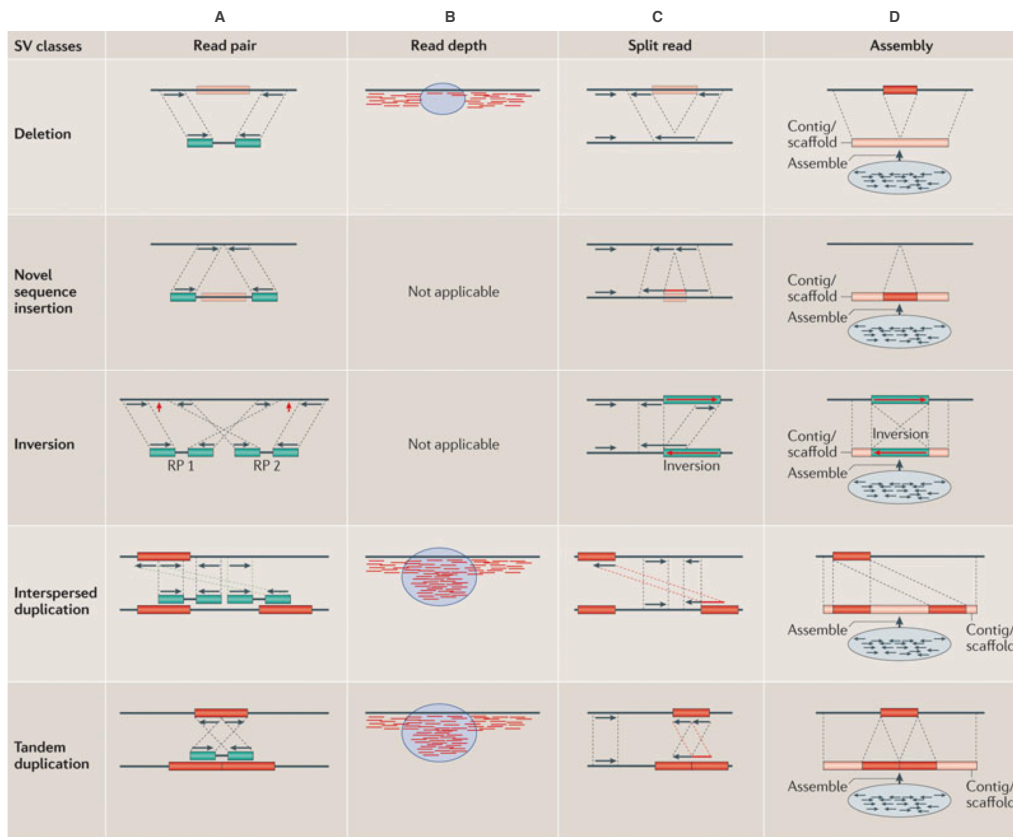


Figure 5-1 : Principes de détection de remaniements chromosomiques par séquençage à très haut débit.

Chaque rangée représente un type de remaniement chromosomique avec les approches analytiques utilisées (A-D) pour le détecter grâce aux lectures (flèches noires et lignes rouges) effectuées par séquençage à très haut débit. Dans chaque encadré, la ligne noire du haut correspond à la séquence du génome de référence et celle du bas à la séquence du génome du patient (un rectangle saumon indique que la séquence a été obtenue par assemblage des contigs); les rectangles rouges (opaques et translucides) représentent les segments délétés, dupliqués, ou insérés; les rectangles verts représentent les segments déplacés; et les rectangles verts contenant une flèche rouge représentent des segments inversés. Les pointillés alignent les extrémités des lectures à leur position sur la séquence du patient par rapport au génome de référence, et les flèches rouges indiquent les points de cassure. **A** | Paires de lectures. Les délétions, duplications et inversions causent une discordance entre la distance et l'orientation des paires de lectures obtenues chez le patient par rapport à celles du génome de référence. **B** | Profondeur de lecture. Le nombre de lectures effectuées pour chaque séquence génomique indique le nombre de copies présentes. **C** | Lectures fractionnées. Un rapprochement, un éloignement et/ou un changement d'orientation des lectures dont on connaît l'espace permet de détecter les délétions, les duplications et les inversions. **D** | Assemblage de séquences. Les différences entre le génome du patient et le génome de référence indiquent directement les délétions, duplications et inversions. *Adaptée de Alkan 2011* ¹³¹.

5.4.4 Stratégie d'analyse : ciblée ou non?

Que ce soit pour les techniques sur micropuces ou le séquençage, il est possible d'interroger tout le génome ou de cibler certaines régions seulement, telles que les gènes associés à un phénotype connu, les exons, ou les séquences régulatrices. Chacune de ces stratégies comporte des avantages et des inconvénients. En effet, si interroger le génome en entier augmente la probabilité d'identifier un CNV pathogénique, cela est également plus coûteux (particulièrement pour le séquençage) et génère un volume important de données que les outils bioinformatiques actuels ne peuvent pas nécessairement traiter ou dont les résultats ne peuvent être interprétés. À l'inverse, cibler seulement les régions déjà connues comme étant associées à un phénotype anormal exclut la possibilité d'identifier un CNV pathogénique n'ayant jamais été rapporté ou étant situé dans les régions intergéniques (voir section 1.2.3), mais garantit que l'interprétation des résultats sera sans ambiguïtés. Cela peut également permettre de réduire les seuils de détection et ainsi permettre l'identification de CNVs de très petite taille ou même intragéniques. Une approche intermédiaire semble de plus en plus populaire, tant avec les micropuces que le séquençage : cibler l'ensemble des séquences du génome codant des protéines (les exons, d'où le terme « exome »).

Alors que cibler l'exome par aCGH peut se faire simplement en utilisant une micropuce constituée de sondes ne couvrant que les exons, l'analyse de l'exome par séquençage nécessite une étape préalable appelée « capture des exons ». Celle-ci peut être effectuée de diverses façons : par PCR, par hybridation sur micropuces ou en solution, ou par « circularisation sélective » avec des sondes moléculaires inversées¹³⁴. L'approche par hybridation est intéressante puisqu'elle permet de gérer une grande quantité de séquences — l'exome humain a une taille d'environ 30 Mb¹³⁵ — et c'est l'hybridation en solution qui est actuellement la plus utilisée, principalement parce qu'elle nécessite moins d'ADN que l'hybridation sur micropuce¹³⁶. Des CNVs intragéniques pathogéniques ont déjà été identifiés en analysant l'exome par aCGH^{84,137}. La mise au point des outils bioinformatiques et des stratégies de détection de CNVs est actuellement en cours pour permettre de faire de même avec le séquençage de l'exome^{132,138,139} et l'identification de CNVs pathogéniques grâce à cette méthode dans un contexte clinique ne saurait tarder.

Si plusieurs des premières plateformes de CGH sur micropuce ont utilisé une approche ciblée, particulièrement pour l'utilisation dans un contexte prénatal, l'approche pangénomique est devenue de plus en plus répandue à mesure que la performance des plateformes s'est améliorée et que les données se sont accumulées. La difficulté de gérer une grande quantité de données n'est pas étrangère à la tendance à vouloir cibler les régions génomiques ayant une plus grande probabilité d'être reliées aux phénotypes retrouvés chez les patients (l'exome) ou celles étant déjà associées à des phénotypes spécifiques. Il est toutefois prévisible qu'à mesure que les coûts d'analyse diminueront et que le traitement des données deviendra plus efficace, la détection de CNVs par séquençage pourra être faite de façon pangénomique, tout comme cela s'est produit pour les micropuces. Il est également plausible que la diversification des méthodes d'identification de CNVs amènera certains laboratoires à personnaliser la technique utilisée en fonction de la population testée.

5.5 Conclusion

Nous avons démontré l'utilité clinique de la CGH pangénomique chez les fœtus avec malformations majeures détectées à l'échographie, en plus de préciser le rendement diagnostique attendu dans cette population, et de mettre en lumière les difficultés d'interprétation associées aux CNVs de signification clinique incertaine. C'est pourquoi nous pensons que cette analyse devrait être réservée aux fœtus avec malformations, à tout le moins pour le moment.

Nous avons également démontré que les plateformes SNP à très haute résolution permettent d'augmenter le rendement diagnostique obtenu chez les enfants atteints de déficience intellectuelle et/ou de malformations congénitales, grâce à la détection des pertes d'hétérozygotie (et non du nombre de CNVs pathogéniques, qui lui demeure inchangé). Nous croyons donc que les patients pour qui un diagnostic n'a pu être établi grâce aux analyses génétiques standards pourraient bénéficier dès maintenant de la détection des pertes d'hétérozygotie, puisque les résultats permettent d'identifier des régions d'intérêt et d'orienter les analyses moléculaires subséquentes.

Néanmoins, l'implantation clinique de micropuces SNP à très haute résolution représente un défi. En effet, leur utilisation résulte en une charge de travail trop grande à cause du nombre élevé de VOUS détectés et des logiciels d'interprétation qui sont inefficaces. Les micropuces SNP de résolution intermédiaire pourraient être plus appropriées pour l'instant, si leur rendement pour la détection des pertes d'hétérozygotie est satisfaisant. Une approche « hybride » (micropuce pangénomique à haute résolution combinée à une couverture ciblée *in silico*) pourrait également être une bonne façon de diminuer la proportion de VOUS tout en permettant de tirer profit de la haute résolution des plateformes à mesure que les connaissances avanceront. Il pourrait également être justifié d'investir dans un logiciel commercial compatible avec plusieurs plateformes — malgré les coûts — puisque cela augmenterait l'efficacité du processus d'interprétation, tout en réduisant les contraintes quant au choix de la ou des plateforme(s), particulièrement à long terme.

Au cours des prochaines années, il faudra travailler à optimiser et standardiser l'interprétation des résultats, tant pour les CNVs que pour les pertes d'hétérozygotie.

L'établissement de directives claires quant au niveau de preuve requis pour catégoriser les CNVs, l'utilisation de sous-catégories de VOUS et l'automatisation de certaines étapes d'interprétation constituent des stratégies intéressantes qui devront être explorées davantage. De plus, s'il est maintenant établi que le caryotype moléculaire est le premier test génétique indiqué chez les patients atteints de déficience intellectuelle et/ou de malformations congénitales, d'autres études devront déterminer quels patients pourraient tirer le plus grand bénéfice de la détection des pertes d'hétérozygotie, en plus d'en optimiser le processus afin de réduire la charge de travail tout en ayant un rendement diagnostique justifiant le coût de l'analyse.

La frontière traditionnelle entre cytogénétique et biologie moléculaire devient de plus en plus floue, les technologies développées dans chaque discipline étant désormais fréquemment utilisées pour détecter différents types de variations génétiques. Ainsi, le séquençage de nouvelle génération représente une technologie très prometteuse qui pourrait théoriquement combiner le caryotype moléculaire, la détection de disomie uniparentale et l'identification de mutations ponctuelles en une seule expérience.

Toutefois, la détection des CNVs par séquençage n'est pas encore au point et demeure très coûteuse, même en se limitant à l'exome. De plus, les équipements requis sont très dispendieux et leur implantation dans les centres hospitaliers n'en est qu'à ses débuts. De plus, si un certain niveau d'incertitude peut être acceptable dans une population pédiatrique, il en est tout autrement dans un contexte prénatal, et les résultats doivent avoir un degré de fiabilité au moins comparable à la CGH, ce qui n'est pas encore le cas. Pour toutes ces raisons, l'implantation de cette technique dans les laboratoires cliniques ne pourra se faire avant quelques années. Pour le moment, il est plus probable que le séquençage soit réservé pour la détection des mutations ponctuelles. Les micropuces demeurent donc la méthode de choix pour détecter les CNVs en cytogénétique clinique, tant en prénatal qu'en postnatal, et permettent de facilement compléter les analyses moléculaires grâce à la détection des pertes d'hétérozygotie.

BIBLIOGRAPHIE

1. Feuk, L., Carson, A.R., and Scherer, S.W. (2006). Structural variation in the human genome. *Nat Rev Genet* 7, 85–97.
2. Conrad, D.F., Pinto, D., Redon, R., Feuk, L., Gokcumen, O., Zhang, Y., Aerts, J., Andrews, T.D., Barnes, C., Campbell, P., et al. (2010). Origins and functional impact of copy number variation in the human genome. *Eur J Hum Genet* 18, 704–712.
3. Eichler, E.E. (2001). Recent duplication, domain accretion and the dynamic mutation of the human genome. *Trends Genet* 17, 661–669.
4. Stankiewicz, P., and Lupski, J.R. (2002). Genome architecture, rearrangements and genomic disorders. *Trends Genet* 18, 74–82.
5. Shaikh, T.H., Kurahashi, H., Saitta, S.C., O'Hare, A.M., Hu, P., Roe, B.A., Driscoll, D.A., McDonald-McGinn, D.M., Zackai, E.H., Budarf, M.L., et al. (2000). Chromosome 22-specific low copy repeats and the 22q11.2 deletion syndrome: genomic organization and deletion endpoint analysis. *Hum Mol Genet* 9, 489–501.
6. Lieber, M.R., Lu, H., Gu, J., and Schwarz, K. (2008). Flexibility in the order of action and in the enzymology of the nuclease, polymerases, and ligase of vertebrate non-homologous DNA end joining: relevance to cancer, aging, and the immune system. *Cell Res* 18, 125–133.
7. Gu, W., Zhang, F., and Lupski, J.R. (2008). Mechanisms for human genomic rearrangements. *Pathogenetics* 1, 4.
8. Zhang, F., Carvalho, C.M.B., and Lupski, J.R. (2009). Complex human chromosomal and genomic rearrangements. *Trends Genet* 25, 298–307.
9. Hastings, P.J., Ira, G., and Lupski, J.R. (2009). A microhomology-mediated break-induced replication model for the origin of human copy number variation. *Plos Gen* 5, e1000327.
10. Weischenfeldt, J., Symmons, O., Spitz, F., and Korbel, J.O. (2013). Phenotypic impact of genomic structural variation: insights from and for human disease. *Nat Rev Genet* 14, 125–138.

11. Lee, C., and Scherer, S.W. (2010). The clinical context of copy number variation in the human genome. *Expert Rev Mol Med* 12, e8.
12. Kim, V.N., Han, J., and Siomi, M.C. (2009). Biogenesis of small RNAs in animals. *Nat Rev Mol Cell Biol* 10, 126–139.
13. Pauler, F.M., Koerner, M.V., and Barlow, D.P. (2007). Silencing by imprinted noncoding RNAs: is transcription the answer? *Trends Genet* 23, 284–292.
14. Malcolm, S. (1996). Microdeletion and microduplication syndromes. *Prenat Diagn* 16, 1213–1219.
15. Shapira, S.K. (1998). An update on chromosome deletion and microdeletion syndromes. *Curr Opin Pediatr* 10, 622–627.
16. Yagi, H., Furutani, Y., Hamada, H., Sasaki, T., Asakawa, S., Minoshima, S., Ichida, F., Joo, K., Kimura, M., Imamura, S.-I., et al. (2003). Role of TBX1 in human del22q11.2 syndrome. *Lancet* 362, 1366–1373.
17. Geng, X., Speirs, C., Lagutin, O., Inbal, A., Liu, W., Solnica-Krezel, L., Jeong, Y., Epstein, D.J., and Oliver, G. (2008). Haploinsufficiency of Six3 fails to activate Sonic hedgehog expression in the ventral forebrain and causes holoprosencephaly. *Dev Cell* 15, 236–247.
18. Lettice, L.A., Horikoshi, T., Heaney, S.J.H., van Baren, M.J., van der Linde, H.C., Breedveld, G.J., Joosse, M., Akarsu, N., Oostra, B.A., Endo, N., et al. (2002). Disruption of a long-range cis-acting regulator for Shh causes preaxial polydactyly. *Proc Natl Acad Sci U S A* 99, 7548–7553.
19. Demars, J., Rossignol, S., Netchine, I., Lee, K.S., Shmela, M., Faivre, L., Weill, J., Odent, S., Azzi, S., Callier, P., et al. (2011). New insights into the pathogenesis of Beckwith-Wiedemann and Silver-Russell syndromes: contribution of small copy number variations to 11p15 imprinting defects. *Hum Mutat* 32, 1171–1182.
20. Dimitrov, B.I., de Ravel, T., Van Driessche, J., de Die-Smulders, C., Toutain, A., Vermeesch, J.R., Fryns, J.P., Devriendt, K., and Debeer, P. (2010). Distal limb deficiencies, micrognathia syndrome, and syndromic forms of split hand foot malformation (SHFM) are caused by chromosome 10q genomic rearrangements. *J Med Genet* 47, 103–111.

21. Nelson, M., Quinonez, S., Ackley, T., Iyer, R.K., and Innis, J.W. (2011). Multiple congenital anomalies and developmental delay in a boy associated with a de novo 16p13.3 deletion. *Am J Med Genet 155A*, 612–617.
22. Grasshoff, U., Bonin, M., Goehring, I., Ekici, A., Dufke, A., Cremer, K., Wagner, N., Rossier, E., Jauch, A., Walter, M., et al. (2011). De novo MECP2 duplication in two females with random X-inactivation and moderate mental retardation. *Eur J Hum Genet 19*, 507–512.
23. Klopocki, E., and Mundlos, S. (2011). Copy-number variations, noncoding sequences, and human phenotypes. *Ann Rev Genomics Hum Genet 12*, 53–72.
24. Almal, S.H., and Padh, H. (2012). Implications of gene copy-number variation in health and diseases. *J Hum Genet 57*, 6–13.
25. Zhang, F., Gu, W., Hurles, M.E., and Lupski, J.R. (2009). Copy number variation in human health, disease, and evolution. *Ann Rev Genomics Hum Genet 10*, 451–481.
26. Girirajan, S., Campbell, C.D., and Eichler, E.E. (2011). Human copy number variation and complex genetic disease. *Annu Rev Genet 45*, 203–226.
27. Stankiewicz, P., and Lupski, J.R. (2010). Structural variation in the human genome and its role in disease. *Annu Rev Med 61*, 437–455.
28. Barber, J.C.K. (2005). Directly transmitted unbalanced chromosome abnormalities and euchromatic variants. *J Med Genet 42*, 609–629.
29. Miller, D.T., Adam, M.P., Aradhya, S., Biesecker, L.G., Brothman, A.R., Carter, N.P., Church, D.M., Crolla, J.A., Eichler, E.E., Epstein, C.J., et al. (2010). Consensus statement: chromosomal microarray is a first-tier clinical diagnostic test for individuals with developmental disabilities or congenital anomalies. *Am J Hum Genet 86*, 749–764.
30. Stavropoulos, J., Shago, M., Bruyere, H., Chan, M., Chernos, J., Chun, K., Cote, G., Craddock, K., Dawson, A., and Duncan, A. (2010). CCMG Guidelines for Genomic Microarray Testing. Canadian College of Medical Geneticists.

31. Kearney, H.M., Thorland, E.C., Brown, K.K., Quintero-Rivera, F., South, S.T., Working Group of the American College of Medical Genetics Laboratory Quality Assurance Committee (2011). American College of Medical Genetics standards and guidelines for interpretation and reporting of postnatal constitutional copy number variants. *Genet Med* 13, 680–685.
32. Vermeesch, J.R., Balikova, I., Schrandt-Stumpel, C., Fryns, J.-P., and Devriendt, K. (2011). The causality of de novo copy number variants is overestimated. *Eur J Hum Genet* 19, 1112–1113.
33. Sebat, J., Lakshmi, B., Troge, J., Alexander, J., Young, J., Lundin, P., Månér, S., Massa, H., Walker, M., Chi, M., et al. (2004). Large-scale copy number polymorphism in the human genome. *Science* 305, 525–528.
34. Iafrate, A.J., Feuk, L., Rivera, M.N., Listewnik, M.L., Donahoe, P.K., Qi, Y., Scherer, S.W., and Lee, C. (2004). Detection of large-scale variation in the human genome. *Nat Genet* 36, 949–951.
35. Sharp, A.J., Locke, D.P., McGrath, S.D., Cheng, Z., Bailey, J.A., Vallente, R.U., Pertz, L.M., Clark, R.A., Schwartz, S., Seagraves, R., et al. (2005). Segmental duplications and copy-number variation in the human genome. *Am J Hum Genet* 77, 78–88.
36. International HapMap 3 Consortium, Peltonen, L., Dermitzakis, E., Schaffner, S.F., Yu, F., Dermitzakis, E., Bonnen, P.E., Altshuler, D.M., de Bakker, P.I.W., Deloukas, P., et al. (2010). Integrating common and rare genetic variation in diverse human populations. *Nature* 467, 52–58.
37. Redon, R., Ishikawa, S., Fitch, K.R., Feuk, L., Perry, G.H., Andrews, T.D., Fiegler, H., Shapero, M.H., Carson, A.R., Chen, W., et al. (2006). Global variation in copy number in the human genome. *Nature* 444, 444–454.
38. Itsara, A., Cooper, G.M., Baker, C., Girirajan, S., Li, J., Absher, D., Krauss, R.M., Myers, R.M., Ridker, P.M., Chasman, D.I., et al. (2009). Population analysis of large copy number variants and hotspots of human genetic disease. *Am J Hum Genet* 84, 148–161.
39. Abramsky, L., and Chapple, J. (1997). 47,XXY (Klinefelter syndrome) and 47,XYY: estimated rates of and indication for postnatal diagnosis with implications for prenatal counselling. *Prenat Diagn* 17, 363–368.

40. Schinzel, A. (2001). Catalogue of unbalanced chromosome aberrations in man (Berlin [u.a.]: Walter de Gruyter).
41. Kearney, H.M., Kearney, J.B., and Conlin, L.K. (2011). Diagnostic implications of excessive homozygosity detected by SNP-based microarrays: consanguinity, uniparental disomy, and recessive single-gene mutations. *Clin Lab Med* 31, 595–613.
42. McQuillan, R., Leutenegger, A.-L., Abdel-Rahman, R., Franklin, C.S., Pericic, M., Barac-Lauc, L., Smolej-Narancic, N., Janicijevic, B., Polasek, O., Tenesa, A., et al. (2008). Runs of homozygosity in European populations. *Am J Hum Genet* 83, 359–372.
43. Kirin, M., McQuillan, R., Franklin, C.S., Campbell, H., McKeigue, P.M., and Wilson, J.F. (2010). Genomic runs of homozygosity record population history and consanguinity. *PLoS ONE* 5, e13996.
44. Nothnagel, M., Lu, T.T., Kayser, M., and Krawczak, M. (2010). Genomic and geographic distribution of SNP-defined runs of homozygosity in Europeans. *Hum Mol Genet* 19, 2927–2935.
45. Ku, C.S., Naidoo, N., Teo, S.M., and Pawitan, Y. (2011). Regions of homozygosity and their impact on complex diseases and traits. *Hum Genet* 129, 1–15.
46. Teo, S.M., Ku, C.S., Salim, A., Naidoo, N., Chia, K.S., and Pawitan, Y. (2012). Regions of homozygosity in three Southeast Asian populations. *J Hum Genet* 57, 101–108.
47. Alkuraya, F.S. (2010). Autozygome decoded. *Genet Med* 12, 765–771.
48. Jinawath, N., Zambrano, R., Wohler, E., Palmquist, M.K., Hoover-Fong, J., Hamosh, A., and Batista, D.A.S. (2011). Mosaic trisomy 13: understanding origin using SNP array. *J Med Genet* 48, 323–326.
49. Papenhausen, P., Schwartz, S., Rishog, H., Keitges, E., Gadi, I., Burnside, R.D., Jaswaney, V., Pappas, J., Pasion, R., Friedman, K., et al. (2011). UPD detection using homozygosity profiling with a SNP genotyping microarray. *Am J Med Genet* 155, 757–768.
50. Li, L.-H., Ho, S.-F., Chen, C.-H., Wei, C.-Y., Wong, W.-C., Li, L.-Y., Hung, S.-I., Chung, W.-H., Pan, W.-H., Lee, M.-T.M., et al. (2006). Long contiguous stretches of homozygosity in the human genome. *Hum Mutat* 27, 1115–1121.

51. Pemberton, T.J., Absher, D., Feldman, M.W., Myers, R.M., Rosenberg, N.A., and Li, J.Z. (2012). Genomic patterns of homozygosity in worldwide human populations. *Am J Hum Genet* 91, 275–292.
52. Meerman, Te, G.J., Van der Meulen, M.A., and Sandkuijl, L.A. (1995). Perspectives of identity by descent (IBD) mapping in founder populations. *Clin Exp Allergy* 25 *Suppl* 2, 97–102.
53. Yamazawa, K., Ogata, T., and Ferguson-Smith, A.C. (2010). Uniparental disomy and human disease: An overview. *Am J Med Genet* 154C, 329–334.
54. Lapunzina, P., and Monk, D. (2011). The consequences of uniparental disomy and copy number neutral loss-of-heterozygosity during human development and cancer. *Biol Cell* 103, 303–317.
55. German, J. (1964). Cytological Evidence for Crossing-Over in vitro in Human Lymphoid Cells. *Science* 144, 298–301.
56. Kotzot, D. (2001). Complex and segmental uniparental disomy (UPD): review and lessons from rare chromosomal complements. *J Med Genet* 38, 497–507.
57. Sridhara Rama Rao, B.S., and Narayanan, H.S. (1976). Consanguinity and familial mental retardation. *J Med Genet* 13, 27–29.
58. Madhavan, T., and Narayan, J. (1991). Consanguinity and mental retardation. *J Ment Defic Res* 35, 133–139.
59. Tadmouri, G.O., Nair, P., Obeid, T., Ali, Al, M.T., Khaja, Al, N., and Hamamy, H.A. (2009). Consanguinity and reproductive health among Arabs. *Reprod Health* 6, 17.
60. Reik, W., and Walter, J. (2001). Genomic imprinting: parental influence on the genome. *Nat Rev Genet* 2, 21–32.
61. Sha, K. (2008). A mechanistic view of genomic imprinting. *Ann Rev Genomics Hum Genet* 9, 197–216.
62. Li, E., Beard, C., and Jaenisch, R. (1993). Role for DNA methylation in genomic imprinting. *Nature* 366, 362–365.
63. Edwards, C.A., and Ferguson-Smith, A.C. (2007). Mechanisms regulating imprinted genes in clusters. *Curr Opin Cell Biol* 19, 281–289.
64. Hall, J.G. (1990). Genomic imprinting: review and relevance to human diseases. *Am J Hum Genet* 46, 857–873.

65. Jiang, Y.-H., Bressler, J., and Beaudet, A.L. (2004). Epigenetics and human disease. *Ann Rev Genomics Hum Genet* 5, 479–510.
66. Linder, D., McCaw, B.K., and Hecht, F. (1975). Parthenogenic Origin of Benign Ovarian Teratomas. *N Engl J Med* 292, 63–66.
67. Kajii, T., and Ohama, K. (1977). Androgenetic origin of hydatidiform mole. *Nature* 268, 633–634.
68. Lawler, S.D., Fisher, R.A., Pickthall, V.J., Povey, S., and Evans, M.W. (1982). Genetic studies on hydatidiform moles. I. The origin of partial moles. *Cancer Genet Cytogenet* 5, 309–320.
69. Butler, M.G. (2009). Genomic imprinting disorders in humans: a mini-review. *J Assist Reprod Genet* 26, 477–486.
70. Cassidy, S.B., Schwartz, S., Miller, J.L., and Driscoll, D.J. (2012). Prader-Willi syndrome. *Genet Med* 14, 10–26.
71. Bryndorf, T., Kirchhoff, M., Rose, H., Maahr, J., Gerdes, T., Karhu, R., Kallioniemi, A., Christensen, B., Lundsteen, C., and Philip, J. (1995). Comparative genomic hybridization in clinical cytogenetics. *Am J Hum Genet* 57, 1211–1220.
72. Tachdjian, G. (2009). Comparative genomic hybridization. *eLS*.
73. Kallioniemi, A., Kallioniemi, O.P., Sudar, D., Rutovitz, D., Gray, J.W., Waldman, F., and Pinkel, D. (1992). Comparative genomic hybridization for molecular cytogenetic analysis of solid tumors. *Science* 258, 818–821.
74. Piper, J., Rutovitz, D., Sudar, D., Kallioniemi, A., Kallioniemi, O.P., Waldman, F.M., Gray, J.W., and Pinkel, D. (1995). Computer image analysis of comparative genomic hybridization. *Cytometry* 19, 10–26.
75. Bentz, M., Plesch, A., Stilgenbauer, S., Döhner, H., and Lichter, P. (1998). Minimal sizes of deletions detected by comparative genomic hybridization. *Genes Chromosomes Cancer* 21, 172–175.
76. Solinas-Toldo, S., Lampel, S., Stilgenbauer, S., Nickolenko, J., Benner, A., Döhner, H., Cremer, T., and Lichter, P. (1997). Matrix-based comparative genomic hybridization: biochips to screen for genomic imbalances. *Genes Chromosomes Cancer* 20, 399–407.

77. Pinkel, D., Se Graves, R., Sudar, D., Clark, S., Poole, I., Kowbel, D., Collins, C., Kuo, W.L., Chen, C., Zhai, Y., et al. (1998). High resolution analysis of DNA copy number variation using comparative genomic hybridization to microarrays. *Nat Genet* 20, 207–211.
78. Cheung, V.G., Nowak, N., Jang, W., Kirsch, I.R., Zhao, S., Chen, X.N., Furey, T.S., Kim, U.J., Kuo, W.L., Olivier, M., et al. (2001). Integration of cytogenetic landmarks into the draft sequence of the human genome. *Nature* 409, 953–958.
79. McPherson, J.D., Marra, M., Hillier, L., Waterston, R.H., Chinwalla, A., Wallis, J., Sekhon, M., Wylie, K., Mardis, E.R., Wilson, R.K., et al. (2001). A physical map of the human genome. *Nature* 409, 934–941.
80. Vissers, L.E.L.M., de Vries, B.B.A., Osoegawa, K., Janssen, I.M., Feuth, T., Choy, C.O., Straatman, H., van der Vliet, W., Huys, E.H.L.P.G., van Rijk, A., et al. (2003). Array-based comparative genomic hybridization for the genomewide detection of submicroscopic chromosomal abnormalities. *Am J Hum Genet* 73, 1261–1270.
81. Carvalho, B., Ouwerkerk, E., Meijer, G.A., and Ylstra, B. (2004). High resolution microarray comparative genomic hybridisation analysis using spotted oligonucleotides. *J Clin Pathol* 57, 644–646.
82. Saillour, Y., Cossée, M., Leturcq, F., Vasson, A., Beugnet, C., Poirier, K., Commere, V., Sublemontier, S., Viel, M., Letourneur, F., et al. (2008). Detection of exonic copy-number changes using a highly efficient oligonucleotide-based comparative genomic hybridization-array method. *Hum Mutat* 29, 1083–1090.
83. Bashiardes, S., Kousoulidou, L., van Bokhoven, H., Ropers, H.-H., Chelly, J., Moraine, C., de Brouwer, A.P.M., Van Esch, H., Froyen, G., and Patsalis, P.C. (2009). A New Chromosome X Exon-Specific Microarray Platform for Screening of Patients with X-Linked Disorders. *The Journal of Molecular Diagnostics* 11, 562–568.
84. Boone, P.M., Bacino, C.A., Shaw, C.A., Eng, P.A., Hixson, P.M., Pursley, A.N., Kang, S.H.L., Yang, Y., Wiszniewska, J., Nowakowska, B.A., et al. (2010). Detection of clinically relevant exonic copy-number changes by array CGH. *Hum Mutat* 31, 1326–1342.

85. Astbury, C., Christ, L.A., Aughton, D.J., Cassidy, S.B., Kumar, A., Eichler, E.E., and Schwartz, S. (2004). Detection of deletions in de novo “balanced” chromosome rearrangements: further evidence for their role in phenotypic abnormalities. *Genet Med* 6, 81–89.
86. Gribble, S.M., Prigmore, E., Burford, D.C., Porter, K.M., Ng, B.L., Douglas, E.J., Fiegler, H., Carr, P., Kalaitzopoulos, D., Clegg, S., et al. (2005). The complex nature of constitutional de novo apparently balanced translocations in patients presenting with abnormal phenotypes. *J Med Genet* 42, 8–16.
87. Higgins, A.W., Alkuraya, F.S., Bosco, A.F., Brown, K.K., Bruns, G.A.P., Donovan, D.J., Eisenman, R., Fan, Y., Farra, C.G., Ferguson, H.L., et al. (2008). Characterization of apparently balanced chromosomal rearrangements from the developmental genome anatomy project. *Am J Hum Genet* 82, 712–722.
88. Ballif, B.C., Rorem, E.A., Sundin, K., Lincicum, M., Gaskin, S., Coppinger, J., Kashork, C.D., Shaffer, L.G., and Bejjani, B.A. (2006). Detection of low-level mosaicism by array CGH in routine diagnostic specimens. *Am J Med Genet* 140, 2757–2767.
89. Scott, S.A., Cohen, N., Brandt, T., Toruner, G., Desnick, R.J., and Edelman, L. (2010). Detection of low-level mosaicism and placental mosaicism by oligonucleotide array comparative genomic hybridization. *Genet Med* 12, 85–92.
90. Hook, E.B. (1977). Exclusion of chromosomal mosaicism: tables of 90%, 95% and 99% confidence limits and comments on use. *Am J Hum Genet* 29, 94–97.
91. Omran, H., Häffner, K., Völkel, A., Kuehr, J., Ketelsen, U.P., Ross, U.H., Konietzko, N., Wienker, T., Brandis, M., and Hildebrandt, F. (2000). Homozygosity mapping of a gene locus for primary ciliary dyskinesia on chromosome 5p and identification of the heavy dynein chain DNAH5 as a candidate gene. *Am. J. Respir. Cell Mol. Biol.* 23, 696–702.
92. Ruf, R., Rensing, C., Topaloglu, R., Guay-Woodford, L., Klein, C., Vollmer, M., Otto, E., Beekmann, F., Haller, M., Wiedensohler, A., et al. (2003). Confirmation of the ATP6B1 gene as responsible for distal renal tubular acidosis. *Pediatr Nephrol* 18, 105–109.

93. Brooks, A.S., Bertoli-Avella, A.M., Burzynski, G.M., Breedveld, G.J., Osinga, J., Boven, L.G., Hurst, J.A., Mancini, G.M.S., Lequin, M.H., de Coo, R.F., et al. (2005). Homozygous nonsense mutations in KIAA1279 are associated with malformations of the central and enteric nervous systems. *Am J Hum Genet* 77, 120–126.
94. Sprecher, E., Ishida-Yamamoto, A., Mizrahi-Koren, M., Rapaport, D., Goldsher, D., Indelman, M., Topaz, O., Chefetz, I., Keren, H., O'brien, T.J., et al. (2005). A mutation in SNAP29, coding for a SNARE protein involved in intracellular trafficking, causes a novel neurocutaneous syndrome characterized by cerebral dysgenesis, neuropathy, ichthyosis, and palmoplantar keratoderma. *Am J Hum Genet* 77, 242–251.
95. Dutt, A., and Beroukhi, R. (2007). Single nucleotide polymorphism array analysis of cancer. *Curr Opin Oncol* 19, 43–49.
96. Conlin, L.K., Thiel, B.D., Bonnemann, C.G., Medne, L., Ernst, L.M., Zackai, E.H., Deardorff, M.A., Krantz, I.D., Hakonarson, H., and Spinner, N.B. (2010). Mechanisms of mosaicism, chimerism and uniparental disomy identified by single nucleotide polymorphism array analysis. *Hum Mol Genet* 19, 1263–1275.
97. Illumina, Inc. (2012). *Molecular Cytogenetics* (Illumina, Inc.).
98. Affymetrix, Inc. (2009). *Genome-Wide Human SNP Array 6.0 Data Sheet* (Affymetrix, Inc.).
99. Srebniak, M., Boter, M., Oudesluijs, G., Joosten, M., Govaerts, L., Van Opstal, D., and Galjaard, R.-J.H. (2011). Application of SNP array for rapid prenatal diagnosis: implementation, genetic counselling and diagnostic flow. *Eur J Hum Genet* 19, 1230–1237.
100. Cooper, G.M., Zerr, T., Kidd, J.M., Eichler, E.E., and Nickerson, D.A. (2008). Systematic assessment of copy number variant detection via genome-wide SNP genotyping. *Nat Genet* 40, 1199–1203.
101. Winchester, L., Yau, C., and Ragoussis, J. (2009). Comparing CNV detection methods for SNP arrays. *Brief Funct Genomic Proteomic* 8, 353–366.

102. Bruno, D.L., Ganesamoorthy, D., Schoumans, J., Bankier, A., Coman, D., Delatycki, M., Gardner, R.J.M., Hunter, M., James, P.A., Kannu, P., et al. (2009). Detection of cryptic pathogenic copy number variations and constitutional loss of heterozygosity using high resolution SNP microarray analysis in 117 patients referred for cytogenetic analysis and impact on clinical practice. *J Med Genet* 46, 123–131.
103. McMullan, D.J., Bonin, M., Hehir-Kwa, J.Y., de Vries, B.B.A., Dufke, A., Rattenberry, E., Steehouwer, M., Moruz, L., Pfundt, R., de Leeuw, N., et al. (2009). Molecular karyotyping of patients with unexplained mental retardation by SNP arrays: a multicenter study. *Hum Mutat* 30, 1082–1092.
104. Gijbbers, A.C.J., Lew, J.Y.K., Bosch, C.A.J., Schuurs-Hoeijmakers, J.H.M., van Haeringen, A., Hollander, den, N.S., Kant, S.G., Bijlsma, E.K., Breuning, M.H., Bakker, E., et al. (2009). A new diagnostic workflow for patients with mental retardation and/or multiple congenital abnormalities: test arrays first. *Eur J Hum Genet* 17, 1394–1402.
105. Friedman, J., Adam, S., Arbour, L., Armstrong, L., Baross, A., Birch, P., Boerkoel, C., Chan, S., Chai, D., Delaney, A.D., et al. (2009). Detection of pathogenic copy number variants in children with idiopathic intellectual disability using 500 K SNP array genomic hybridization. *BMC Genomics* 10, 526.
106. Bernardini, L., Alesi, V., Loddo, S., Novelli, A., Bottillo, I., Battaglia, A., Digilio, M.C., Zampino, G., Ertel, A., Fortina, P., et al. (2010). High-resolution SNP arrays in mental retardation diagnostics: how much do we gain? *Eur J Hum Genet* 18, 178–185.
107. Sagoo, G.S., Butterworth, A.S., Sanderson, S., Shaw-Smith, C., Higgins, J.P.T., and Burton, H. (2009). Array CGH in patients with learning disability (mental retardation) and congenital anomalies: updated systematic review and meta-analysis of 19 studies and 13,926 subjects. *Genet Med* 11, 139–146.
108. Cooper, G.M., Coe, B.P., Girirajan, S., Rosenfeld, J.A., Vu, T.H., Baker, C., Williams, C., Stalker, H., Hamid, R., Hannig, V., et al. (2011). A copy number variation morbidity map of developmental delay. *Nat Genet* 43, 838–846.

109. Shevell, M., Ashwal, S., Donley, D., Flint, J., Gingold, M., Hirtz, D., Majnemer, A., Noetzel, M., Sheth, R.D., Neurology, Q.S.S.O.T.A.A.O., et al. (2003). Practice parameter: evaluation of the child 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* 60, 367–380.
110. Dallaire, L., Michaud, J., Melancon, S.B., Potier, M., Lambert, M., Mitchell, G., and Boisvert, J. (1991). Prenatal diagnosis of fetal anomalies during the second trimester of pregnancy: their characterization and delineation of defects in pregnancies at risk. *Prenat Diagn* 11, 629–635.
111. Benn, P.A., Egan, J.F.X., Fang, M., and Smith-Bindman, R. (2004). Changes in the utilization of prenatal diagnosis. *Obstet Gynecol* 103, 1255–1260.
112. Rizzo, N., Pittalis, M.C., Pilu, G., Perolo, A., Banzi, C., Visentin, A., and Bovicelli, L. (1996). Distribution of abnormal karyotypes among malformed fetuses detected by ultrasound throughout gestation. *Prenat Diagn* 16, 159–163.
113. Hillman, S.C., Pretlove, S., Coomarasamy, A., McMullan, D.J., Davison, E.V., Maher, E.R., and Kilby, M.D. (2011). Additional information from array comparative genomic hybridization technology over conventional karyotyping in prenatal diagnosis: a systematic review and meta-analysis. *Ultrasound Obstet Gynecol* 37, 6–14.
114. van Karnebeek, C.D.M., Scheper, F.Y., Abeling, N.G., Alders, M., Barth, P.G., Hoovers, J.M.N., Koevoets, C., Wanders, R.J.A., and Hennekam, R.C.M. (2005). Etiology of mental retardation in children referred to a tertiary care center: a prospective study. *Am J Ment Retard* 110, 253–267.
115. Wapner, R.J., Martin, C.L., Levy, B., Ballif, B.C., Eng, C.M., Zachary, J.M., Savage, M., Platt, L.D., Saltzman, D., Grobman, W.A., et al. (2012). Chromosomal Microarray versus Karyotyping for Prenatal Diagnosis. *N Engl J Med* 367, 2175–2184.
116. Duncan, A., Langlois, S., SOGC Genetics Committee, CCMG Prenatal Diagnosis Committee (2011). Use of array genomic hybridization technology in prenatal diagnosis in Canada. *J Obstet Gynaecol Can* 33, 1256–1259.

117. Bruno, D.L., Stark, Z., Amor, D.J., Burgess, T., Butler, K., Corrie, S., Francis, D., Ganesamoorthy, D., Hills, L., James, P.A., et al. (2011). Extending the scope of diagnostic chromosome analysis: detection of single gene defects using high-resolution SNP microarrays. *Hum Mutat* 32, 1500–1506.
118. Sund, K.L., Zimmerman, S.L., Thomas, C., Mitchell, A.L., Prada, C.E., Grote, L., Bao, L., Martin, L.J., and Smolarek, T.A. (2012). Regions of homozygosity identified by SNP microarray analysis aid in the diagnosis of autosomal recessive disease and incidentally detect parental blood relationships. *Genet Med* 15, 70–78.
119. Siggberg, L., Sirpa, A.-M., Tarja, L., Kristiina, A., Ilari, S., Kati, K., Päivi, L., Marja, H., Liisa, M., Esa, K., et al. (2012). High-resolution SNP array analysis of patients with developmental disorder and normal array CGH results. *BMC Med Genet* 13, 84.
120. Filges, I., Suda, L., Weber, P., Datta, A.N., Fischer, D., Dill, P., Glanzmann, R., Benzing, J., Hegi, L., Wenzel, F., et al. (2012). High resolution array in the clinical approach to chromosomal phenotypes. *Gene* 495, 163–169.
121. Qiao, Y., Tyson, C., Hrynychak, M., Lopez-Rangel, E., Hildebrand, J., Martell, S., Fawcett, C., Kasmara, L., Calli, K., Harvard, C., et al. (2013). Clinical application of 2.7M Cytogenetics array for CNV detection in subjects with idiopathic autism and/or intellectual disability. *Clin Genet* 83, 145–154.
122. Hehir-Kwa, J.Y., Wieskamp, N., Webber, C., Pfundt, R., Brunner, H.G., Gilissen, C., de Vries, B.B.A., Ponting, C.P., and Veltman, J.A. (2010). Accurate distinction of pathogenic from benign CNVs in mental retardation. *PLoS Comput Biol* 6, e1000752.
123. Kearney, H.M., South, S.T., Wolff, D.J., Lamb, A., Hamosh, A., Rao, K.W., Working Group of the American College of Medical Genetics (2011). American College of Medical Genetics recommendations for the design and performance expectations for clinical genomic copy number microarrays intended for use in the postnatal setting for detection of constitutional abnormalities. *Genet Med* 13, 676–679.
124. Girirajan, S., Rosenfeld, J.A., Cooper, G.M., Antonacci, F., Siswara, P., Itsara, A., Vives, L., Walsh, T., McCarthy, S.E., Baker, C., et al. (2010). A recurrent 16p12.1 microdeletion supports a two-hit model for severe developmental delay. *Nat Genet* 42, 203–209.

125. Girirajan, S., Rosenfeld, J.A., Coe, B.P., Parikh, S., Friedman, N., Goldstein, A., Filipink, R.A., McConnell, J.S., Angle, B., Meschino, W.S., et al. (2012). Phenotypic Heterogeneity of Genomic Disorders and Rare Copy-Number Variants. *N Engl J Med* *367*, 1321–1331.
126. ENCODE Project Consortium, Dunham, I., Kundaje, A., Aldred, S.F., Collins, P.J., Davis, C.A., Doyle, F., Epstein, C.B., Frietze, S., Harrow, J., et al. (2012). An integrated encyclopedia of DNA elements in the human genome. *Nature* *489*, 57–74.
127. Savarese, M., Piluso, G., Orteschi, D., Di Fruscio, G., Dionisi, M., del Vecchio Blanco, F., Torella, A., Giugliano, T., Iacomino, M., Zollino, M., et al. (2012). Enhancer Chip: Detecting Human Copy Number Variations in Regulatory Elements. *PLoS ONE* *7*, e52264.
128. Sneddon, T.P., and Church, D.M. (2012). Online resources for genomic structural variation. *Methods Mol Biol* *838*, 273–289.
129. Wierenga, K.J., Jiang, Z., Yang, A.C., Mulvihill, J.J., and Tsinoremas, N.F. (2012). A clinical evaluation tool for SNP arrays, especially for autosomal recessive conditions in offspring of consanguineous parents. *Genet Med*.
130. Riggs, E.R., Church, D.M., Hanson, K., Horner, V.L., Kaminsky, E.B., Kuhn, R.M., Wain, K.E., Williams, E.S., Aradhya, S., Kearney, H.M., et al. (2012). Towards an evidence-based process for the clinical interpretation of copy number variation. *Clin Genet* *81*, 403–412.
131. Alkan, C., Coe, B.P., and Eichler, E.E. (2011). Genome structural variation discovery and genotyping. *Nat Rev Genet* *12*, 363–376.
132. Xi, R., Lee, S., and Park, P.J. (2012). A survey of copy-number variation detection tools based on high-throughput sequencing data. *Curr Protoc Hum Genet Chapter 7*, 7.19.1–7.19.15.
133. Ye, K., Schulz, M.H., Long, Q., Apweiler, R., and Ning, Z. (2009). Pindel: a pattern growth approach to detect break points of large deletions and medium sized insertions from paired-end short reads. *Bioinformatics (Oxford, England)* *25*, 2865–2871.

134. Mertes, F., ElSharawy, A., Sauer, S., van Helvoort, J.M.L.M., van der Zaag, P.J., Franke, A., Nilsson, M., Lehrach, H., and Brookes, A.J. (2011). Targeted enrichment of genomic DNA regions for next-generation sequencing. *Briefings in Functional Genomics* *10*, 374–386.
135. Ng, S.B., Turner, E.H., Robertson, P.D., Flygare, S.D., Bigham, A.W., Lee, C., Shaffer, T., Wong, M., Bhattacharjee, A., Eichler, E.E., et al. (2009). Targeted capture and massively parallel sequencing of 12 human exomes. *Nature* *461*, 272–276.
136. Asan, Xu, Y., Jiang, H., Tyler-Smith, C., Xue, Y., Jiang, T., Wang, J., Wu, M., Liu, X., Tian, G., et al. (2011). Comprehensive comparison of three commercial human whole-exome capture platforms. *Genome Biol* *12*, R95.
137. Celestino-Soper, P.B.S., Shaw, C.A., Sanders, S.J., Li, J., Murtha, M.T., Ercan-Sencicek, A.G., Davis, L., Thomson, S., Gambin, T., Chinault, A.C., et al. (2011). Use of array CGH to detect exonic copy number variants throughout the genome in autism families detects a novel deletion in TMLHE. *Hum Mol Genet* *20*, 4360–4370.
138. Karakoc, E., Alkan, C., O’Roak, B.J., Dennis, M.Y., Vives, L., Mark, K., Rieder, M.J., Nickerson, D.A., and Eichler, E.E. (2012). Detection of structural variants and indels within exome data. *Nat Methods* *9*, 176–178.
139. Shi, Y., and Majewski, J. (2013). FishingCNV: a graphical software package for detecting rare copy number variations in exome sequencing data. *Bioinformatics* (Oxford, England).

ANNEXE

Annexe 1 : Apport des auteurs aux articles

1^{er} article : Guylaine D'Amours, Zoha Kibar, Géraldine Mathonnet, Raouf Fetni, Frédérique Tihiy, Valérie Désilets, Sonia Nizard, Jacques L. Michaud, Emmanuelle Lemyre. **Whole-genome array CGH identifies pathogenic copy number variations in fetuses with major malformations and a normal karyotype.**

Les résultats rapportés dans cet article découlent d'analyses effectuées dans un cadre clinique par le personnel du laboratoire de cytogénétique prénatale du CHU Sainte-Justine. Z.K. a fourni les appareils utilisés pour effectuer l'analyse CGH sur micropuce. G.M. a effectué l'analyse CGH. V.D., S.N., J.L.M. et E.L. ont recueilli les informations cliniques. G.M., R.F., F.T., S.N. et E.L. ont effectué l'interprétation initiale des résultats individuels dans un cadre clinique. J.L.M. et E.L. ont révisé le manuscrit.

Ma contribution à cet article a consisté à colliger les données cliniques des patientes (et des enfants, le cas échéant), refaire l'interprétation des CNVs détectés, analyser les données, compiler les résultats et rédiger le manuscrit.

2^e article : Guylaine D'Amours, Géraldine Mathonnet, Raouf Fetni, Sonia Nizard, Myriam Srour, Frédérique Tihiy, Jacques L. Michaud, Emmanuelle Lemyre. **SNP arrays: comparing diagnostic yields for four platforms in children with developmental delay.**

L'analyse des échantillons avec la puce CGX-12 a été effectuée dans un cadre clinique par le personnel du laboratoire de cytogénétique prénatale du CHU Sainte-Justine. G.M. a effectué l'analyse CGH sur micropuce. S.N., M.S., J.L.M. et E.L. ont recruté les participants à l'étude et ont recueilli les informations cliniques. G.M., R.F., F.T., S.N. et E.L. ont effectué

l'interprétation des résultats individuels dans un cadre clinique. J.L.M. et E.L. ont révisé le manuscrit.

La technique d'hybridation génomique sur les micropuces SNP a été effectuée par un laboratoire externe, tel que mentionné dans les remerciements, qui a effectué l'hybridation, la lecture des micropuces et la création des fichiers de données brutes. Ces fichiers m'ont par la suite été transmis pour l'analyse des données.

Ma contribution à cet article a consisté à — pour les micropuces SNP — préparer les échantillons pour la technique d'hybridation (incluant l'extraction d'ADN), établir les paramètres de détection optimaux, configurer les logiciels de visualisation et interpréter les CNVs et LOHs sans connaître les résultats des analyses cliniques. Pour les micropuces SNP et CGX-12, j'ai compilé et comparé les résultats, en plus de rédiger le manuscrit.