Microarray analysis
on spontaneous abortions

Helga Hauksdóttir

Thesis for the degree of Master of Science
University of Iceland
Faculty of Medicine
School of Health Science

HÁSKÓLI ÍSLANDS

Örflögugreiningar á fósturlátum

Helga Hauksdóttir

Ritgerð til meistaragráðu

Umsjónarkennari: Jón Jóhannes Jónsson

Meistaranefnd: Magnús Karl Magnússon

Sigríður Valgeirsdóttir

Læknadeild

Heilbrigðisvísindasvið Háskóla Íslands

Júní 2014

Microarray analysis on spontaneous abortions

Helga Hauksdóttir

Thesis for the degree of Master of Science

Supervisor: Jón Jóhannes Jónsson

Masters committee: Magnús Karl Magnússon

Sigríður Valgeirsdóttir

Faculty of Medicine

School of Health Sciences

June 2014

Printing office: Háskólaprent

Reykjavík, Íceland 2014

Ágrip

Allt að 20% þekktra þungana enda í fósturláti og litningagallar eru greinanlegir með litningarannsókn í 40-50% tilfella. Með örflögugreiningu (array CGH) hafa greinst smærri eða sjaldgæfir eintakafjöldabreytileikar (undir greiningarhæfni smásjár) í um 20% fósturláta með eðlilega litningagerð. Klínísk þýðing þessara breytileika hefur í flestum tilfellum verið óljós.

Markmið þessarar rannsóknar var að kanna hvort við gætum með háskerpu heildarerfðamengis táknraða-miðaðri örflögu greint orsakir endurtekinna fósturláta í mönnum þar sem litningagerð var eðlileg. Tilgáta okkar var að með háskerpu örflögu mætti greina smærri eintakafjöldabreytileika en áður hafa verið rannsakaðir í fósturlátum. Þannig gætum við greint þekktar eða líklegar orsakir fósturláta og hugsanlega uppgötvað ný álitsgen fyrir fósturlátum.

Fóstursýni voru fengin frá konum sem komu á Landspítala-háskólasjúkrahús vegna fósturláts. Allir þátttakendur voru pör með endurtekin fósturlát (þrjú eða fleiri) og meðgöngulengd ≤20 vikur. Þau voru að gangast undir litningarannsókn og erfðaráðgjöf vegna endurtekinna fósturláta.

Fjörutíu og þrjú fósturlátasýni frá 34 pörum voru tekin með í rannsóknina. Litningagreining var afbrigðileg í 23 (53.5%) tilfella, eðlileg í 14 (32.6%) tilfellum og tókst ekki í sex (14.0% tilfellum). Við fundum markvert fleiri og smærri eintakafjöldabreytileika með táknraða-miðuðu örflögunni en í fyrri rannsóknum. Alls greindust 1723 eintakafjöldabreytileikar í 13 sýnum með eðlilega litningagerð, að meðaltali 133 eintakafjöldabreytileikar á sýni. Flestir þessara eintakafjöldabreytileika voru smáir með 60.2% <10 kb (bil 27 bp-1.36 Mb). Stór hluti eintakafjöldabreytileikanna náðu yfir gen (92.7%), þ.á.m. OMIM gen (66.6%) og OMIM morbid gen (16.0%). Hlutfall sjaldgæfra breytileika með <50% skörun við Database of Genomic Variants var 26.9%.

Við fundum engar þekktar skýringar fyrir fósturlátum með þessari örflögu. Hins vegar fundum við tvö álitsgen sem hugsanlegar nýjar skýringar á fósturlátum. Breytingarnar virtust vera 34 bp arfhreint tap á útröð í *TAF4* geninu og 4.6 kb arfhreint tap í *GDF6* geninu. Átta tilfelli voru arfblendin fyrir tap á genum sem hugsanlega eru banvæn á arfhreinu formi skv. leit í músagagnabanka. Kerfisgreining leiddi í ljós hugsanlegar vísbendingar um samverkandi arfblendni í sumum tilfellum, sem fól í sér samband á milli gena sem höfðu hlutverki að gegna í fósturþroska.

Þetta er fyrsta rannsóknin á táknraða-miðaðri örflögu sem notuð er fyrir rannsóknir á fósturlátum. Á heildina litið voru ekki nógu miklar sannanir fyrir meinvaldandi virkni eintakafjöldabreytileikanna til að nota gögnin í erfðaráðgjöf fyrir parið sem um ræðir. Með þessari flögu greindum við breytingar sem náðu yfir einstakar táknraðir og hugsanleg álitsgen fyrir fósturlátum. Samt sem áður eru miklar hindranir fyrir því að nota táknraða-miðaða örflögu fyrir rannsóknir á fósturlátum. Helstu ástæðurnar eru erfiðleikar við bæði líffræðilega og tæknilega túlkun á þeim mikla fjölda eintakafjöldabreytileika sem greinast, sérstaklega smáum sjaldgæfum eintakafjöldabreytileikum. Einnig eru erfiðleikar við að meta hvort um falsk jákvæðar breytingar sé að ræða, þar á meðal tæknilegir erfiðleikar og aukinn kostnaður við að framkvæma staðfestingarpróf á hugsanlegum álitsbreytingum.

Abstract

Spontaneous abortion (SA) occurs in up to 20% of recognized pregnancies and chromosomal abnormalities are dectable with conventional karyotypic analysis in 40-50% of cases. Array CGH analysis has revealed submicroscopic or rare CNVs in approximately 20% of SAs with normal karyotype, although clinical significance of CNVs in most cases has been unclear.

The aims of this study was to test if with a high resolution Exon-Focused CGH array we could detect causes of recurrent, spontaneous abortions in humans, where karyotype was normal. Our hypothesis was that this type of array would detect smaller copy number variants then have been previously studied in spontaneous abortions and with this approach we would be able to detect known or likely causes for SAs and possibly discover new candidate genes for SAs.

Fetal tissue samples were obtained from women with a SA at Landspitali-National University Hospital. All subjects were couples with recurrent abortions (three or more) and gestational age ≤20 weeks. They were already undergoing karyotyping and receiving genetic counseling due to recurrent abortions.

Forty three fetal tissue samples, from 34 couples, were included in the study. Conventional karyotyping was abnormal in 23 (53.5%) cases, normal in 14 (32.6%) cases, and failed in 6 (14.0%) cases. We found a significantly larger amount and smaller CNVs with the Exon-Focused CGH array than in previous studies. A total of 1723 copy number variants (CNVs) were identified in the 13 samples with normal karyotype with an average of 133 CNVs per sample. Majority of the CNVs were small, with 60.2% <10 kb (range 27 bp-1.36 Mb). A large proportion of the CNVs overlapped genes (92.7%), including OMIM genes (66.6%) and OMIM morbid genes (16.0%). The proportion of rare variants with <50% overlap with Database of Genomic Variants was 26.9%.

We did not find any known causes for SAs with this array. There were nevertheless two possible candidate genes as new causes for SAs. The variants were an apparent 34 bp exonic homozygous loss in the *TAF4* gene and an apparent 4.6 kb homozygous loss in the *GDF6* gene. Eight cases were heterozygous carriers for losses of genes, which are possibly lethal in a homozygous form based on a mouse database search. Network analysis revealed possible indications of synergistic heterozygosity in some cases, involving interaction of genes with a role in embryonic development.

This is the first study using an exon-focused array for analysis of SAs. Overall, there was not enough evidence of pathogenicity of the CNVs identified to use the data in genetic counseling for the couple involved. We detected exonic variants with this array and possible candidate genes for SAs. However, there are some major drawbacks to using this type of array for analysis of SAs. The main drawback are difficulties with both biological and technical interpretation of the high number of CNVs detected, especially small rare exonic CNVs. There are also difficulties with determining if the calls are false positive, including technical challenges and additional cost with performing verification tests of possible candidates.

Acknowledgements

First, I would like to thank my instructor, Jón Jóhannes Jónsson for his support and patience and for giving me the opportunity to perform the analysis at the Department of Genetics and Molecular Medicine (GMM) at Landspitali-University Hospital (LUH).

I would also like to thank the Prenatal Diagnostic Unit at LUH, especially Hildur Harðardóttir, for recruiting participants to the study. I would also like to thank Vigdís Stefánsdóttir at GMM for introducing the study to participants, obtaining informed consents, and completing questionnaires and pedigrees with participants.

I am also thankful to the Cytogenetics laboratory at LUH for preparing samples for the array CGH analysis and collecting demographic data and results from karyotype analysis, especially Ástrós Arnardóttir, Margrét Steinarsdóttir, and Jóhann Heiðar Jóhannsson. I am particularly grateful to Ástrós Arnardóttir for cooperation in the array CGH analysis as well.

I would also like to thank Roche NimbleGen´s service department for scanning the arrays and initial data analysis. I would also like to thank Sigríður Valgeirsdóttir and Magnús Karl Magnússon for being members of my MS committee. I am particularly grateful to Sigríður Valgeirsdóttir for reading this thesis and giving good advice.

I would also like to thank all staff-members and students at GMM LUH and the Department of Biochemistry and Molecular Biology, University of Iceland for helpful advice and discussion, especially Halldóra Sunna Sigurðardóttir, Jónína Jóhannsdóttir, Sif Jónsdóttir, Bjarki Guðmundsson and Hans Guttormur Þormar. I am particularly grateful to Martin Ingi Sigurðsson for making a javascript program to search for losses of imprinted genes.

I am also grateful to the LUH Scientific Fund for partial support for this project with grants.

Finally, I want to thank my family for their support during the difficult times of the work, my mother, Ingibjörg B. Sveinsdóttir, and sisters, Steinunn Hauksdóttir and Hulda Hauksdóttir.

Table of contents

Table of Figures

Table of Tables

Abbreviations

Array CGH = Comparative genomic hybridization with microarray analysis

ASD = Autism spectrum disorder

BAC array = Bacterial artificial chromosome array

BMD = Becker muscular dystropy

Bp = Base pair

CGH = Comparative genomic hybridization

CNV = Copy number variation

CRL= Crown rump length

Ct = Cycle threshold

CVS = Chorionic villus sampling

DD = Devolopmental delay

ddCt = delta - delta Ct

DECIPHER = Database of Chromosomal Imbalance and Phenotype in Humans using Ensembl Resources

DGV = Database of Genomic Variants

DMD = Duchenne muscular dystropy

DNA = Deoxyribonucleic acid

dNTP = Deoxyribonucleotide triphosphate

EDTA = Ethylenediaminetetraacetic acid

FASST = Fast Adaptive States Segmentation Tecnique

FCM = Flow cytometry

FISH = Fluorescence in situ hybridization

GO = Gene Ontology

HLA = Human leukocyte antigen

ID = Intellectual disability

ICCG = The International Collaboration for Clinical Genomics

ISCA = International Standards for Cytogenomic Arrays

Kb = Kilobase

LINE = Long interspersed element

MA = Maternal age

Mad.1dr = Median absolute deviation, 1st derivative

Mb = Megabase

MCA = Multiple congenital abnormalities

MCC = Maternal cell contamination

MEI = Mobile element insertions (MEI)

Mer = Chemical suffix attached to a prefix such as mono-, di-, poly-, tri-, etc., to indicate the smallest unit of a repeating structure, polymer

MGB **=** Minor groove binding

$MgCl_2$ = Magnesium chloride

MGI = Mouse Genome Informatics

MHC = Major histocompatabililty complex

MLPA = Multiplex ligation-dependent probe amplification

MP p-value = Markov Process (MP) p-value

NA= Not available

NAHR = Non-allelic homologous recombination

NEJM = New England Journal of Medicine

NGF = Nerve growth factor

NHEJ = Nonhomologous end-joining

OMIM = Online Mendelian Inheritance in Man

p75NTR = p75 neurotrophin receptor

PCR = Polymerase chain reaction

PMD = Pelizaeus-Merzbacher disease

P-value = Probability value

Q-bound p-value = False discovery rate (FDR)-corrected (for multiple testing) p-value

QC = Quality control

qPCR = Quantitative polymerase chain reaction

RA = Recurrent abortion

RNA = Ribonucleic acid

RQ = Relative quantification

SA = Spontaneous abortion

SE = Standard error

SNP = Single nucleotide polymorphism

String **=** Search Tool for the Retrieval of Interacting Genes/Proteins

TE = Tris-EDTA

UCSC = University of California, Santa Cruz

WG = Whole genome

1 Introduction

Spontaneous abortion (SA) is often defined as pregnancy loss before 20 weeks and it occurs in up to 20% of recognized pregnancies (1). Chromosomal abnormalities, detectable with conventional karyotypic analysis, are found in approximately 40-50% of sporadic SAs and approximately 40% of abortions from women with prior history of abortion, according to a recent review by van den Berg *et al.* (2012) (2). The majority of chromosomal abnormalities in sporadic cases were numerical abnormalities, including trisomies, monosomy X, and polyploidies (~90%) followed by structural abnormalities (~5%) and other abnormalities (~5%). The last group included mosaicism, double, triple and quadruple trisomies, autosomal monosomy and one trisomy plus a balanced translocation. The spectrum of chromosomal abnormalities in recurrent cases was similar (2).

Other causes for SAs are presumbly lethal *de novo* autosomal dominant single gene mutations, inherited autosomal recessive disorders, or result from poorly characterized maternal or environmental factors, including genetic factors, uterine abnormalities, antiphospholipid syndrome, hereditary thrombophilias, alloimmune factors, and endocrine abnormalities (3, 4).

Several factors influence the frequency of chromosomal abnormalities detected in SAs, such as gestational age, presence of malformations, and maternal age. Overall, the earlier in gestation the pregnancy loss, the greater the likelihood of an abnormal karyotype. Chromosomal abnormalities have been found in up to 90% of anembryonic samples decreasing to about 50% at 8-11 weeks gestation, and around 30% at 16-19 weeks gestation (5). The composition of detected abnormalities is also variable, depending on the gestational age as mortality rates of fetuses differ depending on the chromosomal abnormality involved, with the most harmful abnormalities causing abortion earlier in gestation. Abnormalities detected in the second trimester are similar to those detected in liveborns, with trisomies 21, 18, and 13 making up a majority. Later in gestation the proportion of losses caused by chromosomal abnormalities decreases, but other genetic causes, for example single gene mutations do not necessarily decrease in frequency. The frequency of chromosomal abnormalities detected also increase with presence of malformations and with older maternal age, primarily due to increased risk of trisomies.

1.1 Genetic testing in SAs

Conventional karyotypic analysis with G-banding of metaphase chromosomes, which reveals the light microscopic presentation of the chromosome set of a given cell, is the current standard method used to identify chromosomal abnormalities in SAs and in couples with history of recurrent abortions. Cytogenetic analysis of SAs may be invaluable as it may eliminate the need for further investigation and provide a better recurrence risk estimate for the couple. This technique is based on conventional tissue culture and has some limitations such as relatively low resolution (about 10 Mb), poor chromosome morphology, high rate of tissue-culture failure, maternal cell contamination (MCC). Karyotyping is also time intensive and operator-dependent.

Other techniques such as fluorescence in situ hybridization (FISH), multiplex ligation-dependent probe amplification (MLPA) and quantitative polymerase chain reaction (qPCR) can also be

used to detect chromosomal abnormalities, but they are site-specific, and only used if suggested by clinical information such as fetal morphology, family history or positive antenatal trisomy screening.

Comparative genomic hybridization (CGH) is a method that allows detection of gains and losses of DNA copy number accross the whole genome, without the need for culturing cells. It is based on the comparative hybridization of a fluorescently labelled test sample and differently labelled reference sample to metaphase chromosomes. Disadvantages are relatively low resolution (about 10 Mb) and inability to detect balanced rearrangements and polyploidy (2).

Comparative genomic hybridization with microarray analysis (array CGH) is based on the same principle as CGH, but the metaphase chromosomes are replaced by DNA clones or oligonucleotides. This technique is used to detect chromosomal abnormalities (copy number variations, CNVs), including those that are not visible by conventional karyotyping. It has overcome some of the limitations of G-banding analysis and CGH such as low resolution and, it has shown success in cases of tissue culture failure, and maternal cell contamination. It can also have quicker turn around time than G-banding analysis. Array CGH has some limitations, however, such as inability to detect truly balanced translocations and inversions, polyploidy (except in cases with male karyotype), and low-level mosaicism. It is, however, more costly than conventional karyotyping (2, 6).

Maternal cell contamination (MCC) is one limitation of cytogenetic analysis of prenatal samples. According to one study using qPCR genotyping, MCC was identified in 9.1% of direct and/or cultured amniotic fluid samples, 17.8% of which were not bloodstained (7). This may be less of an issue when performing array CGH analysis as maternal cell overgrowth or selection of chromosomal normal cells during cell culture may happen in conventional karyotypic analysis (8).

Some studies suggest that array CGH is more sensitive in detecting low-level mosaicism than conventional karyotypic analysis, primarily because of the need to analyse a large number of cells with karyotypic analysis. The lowest grade of mosaicism described is 8% monosomy 7 (9), detected with a 1 Mb BAC array. In general, SNP arrays are also claimed to be more sensitive than CGH arrays in detecting mosaicism, as they give information on genotype signal strength.

Multiple types of arrays are used for CNV analysis. The first arrays used large clones, typically bacterial artificial chomosome (BAC) clones, 80-200 kb in size, spaced 1 Mb apart resulting in low resolution (2). Today, the most widely used arrays are oligonucleotide arrays, with 25 to 85 mer synthesized probes, and single nucleotide polymorphism (SNP) arrays, which can detect CNVs as well as SNPs (2). The microarrays vary significantly in probe types, resolution, analytical tools used, and quality of signal. This variation complicates interpretation and generalization of results based on specifc platforms.

For simplicity the term CNVs will be used for all copy number variations (losses and gains) in this thesis, irrespective of clinical significance, i.e. whether they are benign or pathogenic. Array CGH has proved to be a powerful technique in detecting genomic disorders associated with CNVs (microdeletions and microduplications) in the human genome, and it is now rapidly replacing conventional karyotyping as the first-tier test in postnatal cytogenetic diagnostics (10). This technique has been documented to detect around 15% extra causally related abnormalities over conventional karyotyping in individuals with unexplained devolopmental delay (DD)/intellectual disability (ID), autism

spectrum disorder (ASD), or multiple congenital abnormalities (MCA). Array CGH is increasingly being applied in prenatal settings (see below).

1.1.1 Array CGH analysis of CNVs in SAs

So far limited research has been done on SAs using array CGH, with studies identified in PubMed limited to about 510 SAs worldwide in 12 papers (8, 11-21) (further summarized and reviewed in Results section 4.3.7 and Table 14). One paper was excluded from our review because of internal inconsistency in reported karyotypes (22). These studies have detected submicroscopic CNVs in SAs but the proportion of CNVs reported has been highly variable. It is quite difficult to compare these studies as they differ in many respects. The variability in diagnostic yield could be explained by many variables, such as differences in array types used with different genomic coverage and resolution, sample size, inclusion criteria (e.g. history of SAs, gestational age, presence of malformations), reference samples, and algorithms used. Lack of standardized reporting criteria also further complicates comparison.

In a recent review on the genetics of early SAs by Van de Berg *et al.* (2011) (2), results from seven studies on SAs were combined (8, 11-14, 16, 19). Submicroscopic CNVs were detected in 5% of 362 sporadic SAs. The study material in those studies was quite variable. The array CGH analysis was in some studies performed only on material from failed karyotypic analysis (12), in other studies only samples with normal karyotype were included (13, 18, 21), and some were blinded studies on unselected material (Table 14).

Initial array CGH studies on SAs used low resolution arrays, such as 1 Mb BAC arrays, and reported submicroscopic CNVs in 2-23% of SAs with a normal karyotype (8, 11-15) (Table 14). In most of these cases the exact size, and gene content of the CNVs were not reported, and location was usually only specified to a chromosome band corresponding to a single or few BAC clones.

More recent papers have detected a higher number of submicroscopic CNVs using higher resolution arrays, i.e. oligonucleotide or SNP arrays. This includes unique (non-polymorphic) CNVs reported in 10-100% of cases (Table 14) with normal or euploid karyotype (15-21). Warren *et al.* (2009) found a number of additional unique CNVs in each of the six samples (100%) tested with a higher resolution 244K Agilent array to verify CNVs detected with a Spectral 2600 BAC array (15). Most large CNVs have been shown to be rare, so these results might actually not be far from what would be expected in a normal population using 60-244K arrays. For example, Itsara *et al.* (2009) observed that 71% of individual CNVs (94% of CNV loci) larger than 100 kb are rare (<1% population frequency), and events >500 kb are heavily enriched for events seen in only one individual (23). It is noteworthy that two recent studies did not report any CNVs at all (17, 20), but the aim of these studies was to compare the array technique to cytogenetic analysis with respect to diagnosis of large chromosomal abnormalities.

Overall size range of reported CNVs in the reviewed studies was 12.9 kb – 4.3 Mb, but the majority of the CNVs reported were small. For example Rajcan-Separovic *et al.* (2010) (19) found that all of the unique CNVs they detected in SAs with structural abnormalities were <250 kb in size. Still, all but one CNV in the studies were larger than 20 kb in size (66/67 or 98.5%). The smaller CNV was originally estimated as 12.9 kb and was later reestimated to be 1.6 kb on a higer-resolution array (19).

This is a different size distribution from what is seen in cases of ID/DD, where majority of known pathogenic CNVs are >400 kb in size (24).

A limitation of the studies is that data on the origin of the CNVs by parental analysis, obstetric history, and embryonic morphology was usually missing. Determination of *de novo* origin of CNVs is often used as an indicator of pathogenicity, although this is not always reliable. Information on the *de novo* or inherited origin of CNVs was only reported in three studies (15, 18, 19). Warren *et al.* (2009) found *de novo* CNVs in 4 of 27 (14.8%) euploid cases. Pathogenicity of those CNVs is not clear, as they either didn´t overlap any genes or overlapped genes with unknown function (15). Rajcan-Separovic *et al.* (2010) (18) found 11 unique CNVs, that were all inherited, in 13 SAs from 8 couples with recurrent pregnancy loss. In a separate study they found the unique CNVs to be predominantly inherited, although *de novo* changes were found in 2 of 14 (14.3%) euploid cases with abnormal morphology (19).

Only one study specified that the study group was women with recurrent abortion (18), although in two studies some of the women have a history of pregnancy loss (14, 19). The study by Rajcan-Separovic *et al.* (2010) (19) is also the only one that provides data on embryonic morphology. They used material from euploid SAs with a range of structural abnormalities, as their previous embryoscopic evaluations had shown that structural abnormalities were detected in majority of both euploid (74%) and aneuploid or polyploid (90%) SAs, and therefore they suggested that lethal submicroscopic CNVs might be responsible for SAs of euploid embryos. They identified six unique CNVs in 5 of the 14 (35.7%) embryos studied with the 1 Mb BAC array, both inherited and *de novo*, although it is not evident if the CNVs are causative for the SAs.

Overall, the clinical relevance of the CNVs reported in those studies remains unclear. The number of cases and samples tested in this area is still relatively small so, further research is needed to determine the frequency and size distribution of CNVs in SAs, and also the proportion of *de novo* CNVs, and an estimate of whether these CNVs contribute to the cause of SAs.

1.1.2 Array CGH versus karyotypic analysis for chromosomal abnormalities in SAs

A few studies have compared array CGH and karyotypic analysis in SAs, some of which suggest that array CGH analysis gives results more often than karyotypic analysis (8, 11, 14, 17, 20) mainly due to its success with nonviable tissue, and also in cases of MCC (8, 20) or mosaicism (8, 11, 17).

Van den Berg *et al.* (2012) reviewed four studies with 264 SAs samples to determine the accuracy of array CGH compared to conventional karyotyping (8, 11, 14, 17). The studies differed in resolution of the platform used, and sample sizes and results differed with regard to proportion of chromosome abnormalities detected by array CGH (range of chromosome abnormalities 27–41%). These studies combined showed that less abnormalities remained undetected with array CGH compared to conventional karyotyping out of all abnormalities detected with either method (array-CGH missed 2% (95% CI: 0–5) compared to 10% (95% CI: 6–14) by karyotyping. According to the review, this could be explained by the fact that karyotyping has a higher failure rate (18% (95% CI: 7–30)) compared to array CGH (5% (95% CI: 0–10)) and array CGH detected some additional

submicroscopic abnormalities. The overall failure rate of conventional karyotyping was 21% (95% CI:13–30). Van den Berg *et al.* (2012) also found the proportion of chromosomal abnormalities to be equal with array CGH (31% (95% CI: 14–38) and conventional karyotyping (30% (95% CI: 23–37) out of successfully analysed samples with each method in sporadic SA samples.

Benkhalifa *et al.* (2005) and Zhang *et al.* (2009) also demonstrated efficacy of array CGH in cases of tissue culture failure in small studies on SAs. They were able to obtain results in all cases (26 and 58 respectively) and detected chromosomal abnormalities in 23-58% of cases with array CGH, some of which were submicroscopic. In addition, larger chromosomal abnormalities that were originally missed by karyotyping due to mosaicism or MCC were detected in a few studies (8, 11, 13, 17).

The main limitations of array CGH compared to conventional karyotyping is difficulty in detecting balanced chromosomal abnormalites and polyploidy. Zhang *et al.* (2009) and Menten *et al.* (2009) were partly able to overcome this as they used a combination of array CGH and microsatellite genotyping (16) or flow cytometry (FCM) (14) for detection of ploidy status, and thus significantly increased detection rate of chromosomal abnormalities in SAs compared to conventional karyotyping by up to 53% (14).

1.1.3 Array CGH versus conventional karyotypic analysis for chromosomal abnormalities in stillbirths and prenatal diagnosis

Conventional karyotyping is currently the standard method for identifying chromosomal abnormalities in prenatal diagnosis and stillbirths, although the array CGH technology is increasingly being applied in prenatal settings. Two papers were recently published in the New England Journal of Medicine (NEJM) on the comparison of array CGH and conventional karyotyping for chromosomal abnormalities in a large multi-center clinical trial on 4406 prenatal diagnosis cases (25) and in the second paper on 532 stillbirths (26).

The study on prenatal diagnosis cases (25) showed that array CGH analysis detected additional, clinically significant cytogenetic information compared to karyotyping, and it was suggested that array CGH would become a standard part of prenatal testing. The investigators used two array platforms, i.e. a custom 44K Agilent array covering targeted regions of known disease association and Affymetrix Genome-Wide Human SNP Array 6.0, containing 1.8 million oligonucleotide probes. Array CGH was especially valuable in cases with a structural anomaly and a normal karyotype based on cytogenetic analysis, where it detected clinically relevant losses or gains in 6.0% of cases. It also detected clinically relevant losses or gains in 1.7% of cases missed by karyotyping where indications were advanced maternal age or positive Down´s syndrome screening result. It was equally efficacious as karyotyping in identifying aneuploidies and unbalanced rearrangements, but did not identify balanced translocations and triploidies.

In the study on stillbirths (26), defined as pregnancy loss after 20 weeks, array CGH analysis was more likely than conventional karyotyping to provide a genetic diagnosis, primarily because of its success with nonviable tissue. These investigators used the Affymetrix Genome-Wide Human SNP Array 6.0. Array containing 1.8 million probes and they included all CNVs ≥500 kb in the analysis. A result

was obtained from array CGH analysis in 24.0% more cases than with karyotyping (87.4% vs. 70.5%, P<0.001), and it provided a better detection of chromosomal abnormalities (aneuploidy or pathogenic copy number variants, 8.3% vs. 5.8%, P=0.007). Array CGH also identified more abnormalities among antepartum stillbirths (8.8% vs. 6.5%, P=0.02). Array CGH was especially valuable in analysis on stillbirths with congenital anomalies where it also detected more chromosomal abnormalities (29.9% vs. 19.4%, P=0.008). Karyotype analysis failed in 157 of 532 stillbirth cases (29.5%) which is comparable to data on karyotype failure in SAs (see above).

1.2 Recurrent abortion

Recurrent abortion (RA) is typically defined as the loss of at least two or three pregnancies, and the sequence of the abortions does not have to be consecutive (27). Up to 5% of all couples will experience RAs (28), which is greater than would be expected by chance alone as the risk of three consecutive SAs is about $(0.20)^3$ or 0.80% if the risk of pregnancy loss is about 20% in any given pregnancy.

Maternal age and previous number of abortions are two major independent risk factors for RAs. The risk for abortion increases with advancing maternal age as the frequency of chromosomal abnormalities (i.e. trisomies) increases. A history of pregnancy loss has also shown to increase risk of abortion at subsequent pregnancies, with the empiric risk estimates increasing from 12% without any prior history, to 24% after one loss, 32% after three losses, and 53% after six or more losses (29). This suggests some underlying operative mechanisms as causes of SA.

Positive family history of SAs can also be a risk factor for RAs (30, 31). The risk of abortion has been reported to be enhanced for siblings of RA patients (30, 31), and a few factors associated with abortion have been identified. An example is the discovery that sisters that are HLA-identical-by-descent with probands with RAs exhibit a higher risk of abortion than those who are not HLA identical, which suggests that genes in the HLA region are involved in recurrent and probably also sporadic abortions (32).

1.2.1 HLA antigens and RAs

The human leukocyte antigen (HLA) gene family is the human version of the major histocompatabililty complex (MHC). It plays a central role in the immune system. This locus is located on chromosome 6, and is divided into three main classes, i.e. class I, II and III. The main class I genes in humans are *HLA-A*, *HLA-B* and *HLA-C* and they play a major role by presenting peptides derived from the endoplasmic reticulum lumen to the immune system. The main class II genes in humans are *HLA-DPA1*, *HLA-DPB1*, *HLA-DQA1*, *HLA-DQB1*, *HLA-DRA*, and *HLA-DRB1*. They present peptides derived from extracellular proteins to the immune system.

A number of studies have revealed an association of HLA antigens with RAs, e.g. polymorphisms in the *HLA-G* locus have been linked to increased risk of RAs (33). Decades ago, Komlos *et al.* (1977) hypothesized that HLA sharing among couples was associated with RAs (34). Evidence has, however, remained divided on the role HLA sharing among couples. Several studies have shown that there is increased sharing of HLA antigens among couples with RAs, including the *HLA-A* and *HLA-DR* antigens (35-45). Other studies have, however, not found an association with

HLA sharing and RAs (46, 47). A limitation of the studies is that they were small and used couple-sharing as a proxy measure of maternal-foetal sharing.

Furthermore, some studies have shown that there is a lack of antibodies against partner´s HLA antigens (maternal blocking antibodies) among women with RAs (40, 41, 44, 48, 49). Unander *et al.* (1983) hypothesized that increased HLA compatability between the mother and the fetus was linked to a deficiency in development of antifetal antibody during development (40) As a consequence, the fetus might be deprived of the protection by maternal blocking antibody, which might allow maternal cytotoxic reactions to cause abortion (40). It has also been suggested that sharing of HLA antigens could be in linkage disequilibrium with other genes of the same region, which are lethal for the embryo in the homozygous state (43, 50).

Despite careful evaluation of possible causes for pregnancy loss, including known genetic, anatomic, endocrinologic, immunologic, and infectious factors, as many as 50% of cases are classified as idiopathic, i.e. having unknown etiology (31). The fact that there is a positive family history in many of the idiopathic cases, suggests that there is some genetic factors in RAs that remain to be identified.

1.2.2 Genetic Causes of RAs

Genetic factors that have been associated directly with RAs include both single gene mutations, and parental chromosome abnormalities. A relatively small proportion of chromosomal abnormalities detected in the first trimester have been reported as inherited, in which cases the recurrance rate estimate for chromosomal abnormalities depend on the abnormality involved.

Typically, chromosome abnormalities are sporadic events, although recurrence rates are higher than expected by chance (51). They are remarkably frequent in RAs or around 40% (2). According to some studies, the prognosis of subsequent pregnancy outcomes is better after an aneuploid abortion than an euploid abortion (52). However, one study shows that fetal aneuploidy contributes to RAs in a small proportion of patients, and that 15% of patients will have repeat aneuploidy (3).

Structural rearrangements often present a greater risk for recurrence. The most common known example is a balanced rearrangement in parents, identified in 3-5% of couples with RAs (4). The rearrangements include reciprocal and Robertsonian translocations, which predispose to a genetically unbalanced chromosome constitution, because of unequal segregation of chromosome material during meiosis, resulting in either pregnancy loss or a live birth of a child with serious birth defects. Another example of parental abnormalities that have been identified as causes of pregnancy loss are inversions, detected in about 0.3% of RAs (29).

For assessment of pathogenicity of copy number variants in patients with ID/DD or MCA, one of the primary criteria is that the CNV is *de novo*. However, as *de novo* cases are sporadic, they are not likely to explain RAs, except in very rare cases with high-level of gonadal mosaicism. It is now becoming increasingly recognized that *de novo* origin of CNVs is not always necessary for pathogenicity. There are a number of situations in which a CNV could potentially lead to RAs, which are discussed in section 1.5.

1.3 Structural variation

Structural variation is generally defined as genomic alteration (e.g. inversion, balanced translocation, or copy number variation) of a DNA segment of approximately 1 kb or larger. Copy number variation involves a DNA segment that is either deleted or duplicated compared to a reference genome. Recent high resolution genome maps have, however, revealed CNVs smaller than 1 kb. They are common among healthy individuals and commonly referred to as indels (insertions or deletions). The Database of Genomic Variants (DGV) website at (http://dgvbeta.tcag.ca/dgv/app/) now defines the size limit of structural variation as 50 bp or larger.

In the past few years advances in technology such as array CGH and high-throughput sequencing have lead to the discovery of widespread copy number variation in the human genome, both novel pathogenic copy number variants (24, 53) and an extensive amount of polymorphic variation in healthy individuals (54-60). Many of the CNVs identified in healthy individuals contain genes that are involved in environmental responses, for example sensory perception and immune system response (61). There is currently great interest in studying possible associations between CNVs and disease risk.

Estimates of the extent of CNVs in the phenotypically normal human genome differ. In Database of Genomic Variants (http://dgvbeta.tcag.ca/dgv/app/), which is a CNV database for healthy individuals, there are now 109.863 merged CNVs in population control samples that collectively cover around 2235 Mb of nucleotides or 72.2% of the human genome (22/4/2014). This accounts for more genomic variability than single nucleotide polymorphisms (SNPs). However, this may be an overestimate as many of the studies are based on BAC arrays that tend to significantly overestimate the size of variants. At the same time there are presumably still many smaller CNVs (<30 kb) that remain to be identified.

Although multiple studies based on CGH/SNP arrays and sequencing have reported a significant contribution of CNVs to human variation, there has been limited overlap between independent studies, even based on the same DNA source (62, 63). Therefore it is still unclear to which extent two genomes differ with respect to CNVs. Conrad *et al.* (2010) identified an average of 1098 validated CNVs (>500 bp) and a cumulative locus length of 24 Mb (0.78% of the genome) when comparing two genomes by using 10x4.2M NimbleGen arrays (64). Pang *et al.* (2010) re-analysed existing whole-genome sequencing data and combined with new microarray data, using different array types (Agilent 10x2.4M, NimbleGen 10x4.2M, Affymetrix 6.0 with 1.8M probes, Illumina 1M and Custom Agilent 244K). The K and M units refer to the number of features or probes per array. Using this approach they found a total non-SNP variation content of 48.8 Mb in a single genome, and genomic differences from the consensus reference sequence by approximately 1.2% when considering indels/CNVs, 0.1% by SNPs and approximately 0.3% by inversions (62). The total amount of CNVs identified with each array type was, however, highly variable (range ~0.1-15 Mb), reflecting for example different probe coverage of the arrays.

1.3.1 Interpretation of clinical significance of CNVs

The discovery of this widespread copy number variation in healthy individuals presents a significant challenge to clinical laboratory geneticists in distinguishing pathogenic CNVs from those that are considered to be benign and less likely to contribute to an affected individual´s clinical phenotype. The phenotypic effects of many rare CNVs are still largely unknown as well as the overall proportion of dosage-sensitive genes in the genome. As many of the CNVs are rare and have not yet been reported in any CNV databases for affected individuals, interpretation can often be difficult and leads to a group of variants being classified with unknown clinical significance.

A number of databases have been formed to assist clinical laboratory geneticists in interpretation of CNV data and to make genotype-phenotype correlations. DECIPHER and ICCG are prominent CNV databases for affected individuals and Database of Genomic Variants (DGV) the major CNV database for healthy individuals. Several papers have also published guidelines for interpretation of CNVs. A leading example is the American College of Medical Genetics standards and guidelines for interpretation and reporting of postnatal constitutional copy number variants (65). These guidelines can be used as a reference for interpretation of CNVs in SAs, but with some modifications (see below).

Large size of CNVs is often used as an indicator of pathogenicity, although rare and large CNVs also occur in healthy individuals at a relatively high frequency. For example Itsara *et al.* 2009 detected a CNV > 500 kb in 5-10% of individuals and a CNV > 1 Mb in 1-2% (23). It can also be problematic to use the *de novo* origin of CNVs as a strict criteria of pathogenicity as discussed before. Commonly used criteria to indicate pathogenicity of CNVs (66) are that the CNV:

i) overlaps a syndromic region in a CNV database for affected individuals (for example DECIPHER or ICCG)
ii) is not present in a CNV database for healthy individuals (for example Database of Genomic Variants)
iii) is a loss rather than gain
iv) is a homozygous loss (or high copy number gain)
v) is gene rich /large rather than gene poor or small
vi) contains a morbid OMIM gene
vii) contains an imprinted gene
viii) is *de novo*, i.e. not inherited from a parent

Basically the same criteria can be used to identify candidate genes and CNVs for SAs, although that strategy might be problematic. For example it is not certain that CNVs listed as pathogenic in databases such as DECIPHER or OMIM would be likely to cause SAs as they are already known in live births. The gestational age also need to be taken into consideration when interpreting the clinical significance of CNVs. The fact that most studies on SAs are using material where gestational age ranges from 5-20 weeks may possibly restrict findings as the composition and frequency of chromosomal abnormalities varies with gestational age. For example, it has been shown that chromosomal abnormalities are most common early in gestation (5), which suggests that lethal

CNVs might also be more common before 5 weeks of gestation. It has also already been shown that array CGH is valuable in detecting pathogenic CNVs in stillbirths (gestational age >20 weeks) where the abnormalities detected are similar to those detected in liveborns (26). This leaves a time gap where the composition and frequency of pathogenic CNVs may be different.

A criterium for identifying candidate genes for SAs could also be that a mouse model shows lethality if homozygous/hemizygous for a loss/mutation. Other criteria that apply to SAs are that the genes in the CNV area play a role in embryogenesis or placental development, and are dosage sensitive and expressed in vital organs. An other possibility is that genes overlapping CNVs in the same material of SA share a biological pathway, which could cause SA by synergistic effects.

Synthetic lethality is defined as synergistic or combined effects of two or more genetic events that lead to cellular or organismal death. It has been described in several organisms ranging from yeast to human cells (67-70). Synthetic lethality applies more commonly to losses or where there is loss of function of a gene, although it may also apply to gains. Synthetic lethality can also explain the sensitivity of cancer cells to certain drugs (67). Similarly, a phenomenon called synergistic heterzygosity or the combined effects of multiple partial defects in one or more pathways can lead to disease in humans e.g.metabolic disorders (71). Recently, Girirajan *et al.* (2012) also proposed a two-hit model wherein the severity and variability of genomic disorders are due to alteration (CNV or point mutation) of multiple functionally relevant genes (72). Further insight on the biological function of the CNVs in SAs can be gained by using bioinformatic approaches such as gene ontology and network analysis to analyse the functions of genes that are deleted or duplicated.

1.3.2 Mechanism and frequency of *de novo* copy number variation

One of the key questions in current human genetics is finding the mechanism and frequency of *de novo* CNVs in human constitutional genomic DNA. Several major mechanism for human genomic rearrangements have been proposed, including non-allelic homologous recombination (NAHR), nonhomologous end-joining (NHEJ), replicative mechanisms, long interspersed element (LINE)-mediated retrotransposition or mobile element insertions (MEI) (73). The relative contribution of the different mechanisms to (both benign and pathogenic) CNVs in humans has not been well characterized, partly due to difficulties in sequencing the breakpoints, although some progess has been made (74).

Despite its importance, the *de novo* CNV mutation rate has remained elusive due to limited sample size and source material (75). Itsara *et al.* (2010) estimated the genome wide CNV mutation rate to be $\mu = 1.2 \times 10^{-2}$ per genome per transmission (or $\mu = 6.5 \times 10^{-3}$ for CNVs > 500 kb) at a resolution of ~30 kb through the direct identification of *de novo* events in a large number of trios of asthmatic individuals and their parents (75). Conrad *et al.* (2010) estimated the mutation rate to be significantly higher or $\mu = 3 \times 10^{-2}$ (for CNVs >500 bp) per haploid genome, per generation, using 10x4.2M NimbleGen arrays and population genetic approaches. In both studies the CNV mutation rate may be an underestimate for two reasons. First, small CNVs are missed. Second, this is an estimate based on live births in individuals without major developmental abnormalities so it does not (or only partially) account for purifying selection as most large CNVs are estimated to be deleterious (76). Purifying selection in this scenario could both mean that some CNVs are lost due to fetal lethality or

because individuals with large inherited CNVs causing neurodevelopmental disorders are less likely to reproduce. Supporting this, there has also been found an increased amount of *de novo* CNVs in individuals with some neurodevelopmental disorders. For example a fourfold enrichment of *de novo* CNVs was identified in cases of multiplex autism versus unaffected siblings, suggesting that the many CNVs contribute a risk for autism (75).

When comparing the contribution of CNVs and SNPs to genomic variation in humans, CNVs account for a vast majority of mutated base pairs; 8-25 kbp per gamete (16-50 kbp per birth) are affected by large CNVs (>100 kb), vs. an average of 30.5 bp per gamete (61 bp per birth) observed for SNPs (76). By contrast, new large CNVs (>100 kb) are relatively rare compared with SNPs. There is one new large CNV per 42 births (95% Poisson CI: 23–97) compared with an average 61 new SNPs per birth (95% CI of the mean: 58–64) (76). The SNP data in this review was, however, based on mutation frequency in males, where mutation rate is increased.

Using a different, interesting approach Wang *et al.* (2012) used single cell high throughput sequencing to analyse recombination activity and *de novo* mutation rates in human sperms (77). They found a recombination of 22.8±0.4 SE (±3.7 SD) events per cell (which agrees well with the average male results implied from other methods) and a *de novo* point mutation rate of 2-4x10^{-8}, which is higher than data from recent estimates of the human germline sequence mutation rates of 1.2-2.3x10^{-8} (76, 78, 79).

1.4 Genomic imprinting and SAs

Genomic imprinting is an epigenetic process by which certain genes are expressed in a parent-of-origin dependent manner. The major epigenetic processes (i.e. modifications of DNA or chromatin that affect genetic function without altering DNA sequence) involved in imprinting are base methylation and histone modification. Methylation of CpG sites of imprinted genes causes expression to be turned off in either the developing egg or the sperm and thus making an epigenetic mark or “imprint” on the next generation. Imprinted expression can also differ between tissues and developmental stages (80).

There are over 200 human genes that have been predicted to be imprinted according to one of the most well known imprinted database (www.geneimprint.com), and many of these genes have been shown to play a role in embryonic or placental development (81). Imprinted genes are susceptibility targets for a number of human pathologies because only a single genetic or epigenetic change is needed to disrupt their function and potentially cause disastrous health effects. Errors in genomic imprinting can cause a variety of developmental disorders, (e.g. Prader Willi and Angelman syndromes), and have also been linked to some complex diseases and cancers (82, 83). In some cases imprinting errors can be lethal (84).

Although it is well recognized that imprinted genes play an important role in both embryonic and placental development, relatively few studies have provided a comprehensive analysis of imprinting disorders in SAs (for example (85-94)). The majority of these studies were performed in mice and only focused on a couple of genes. Most of them showed deviation at the expression or methylation level, and did not specify copy number status. Very limited research has been done using high throughput technique to screen for losses (or mutations) in imprinted genes in RAs in humans. Rajcan-Separovic *et al.* (2010) (18) identified CNVs in two imprinted genes (*TIMP2* and *CTNNA3*) in a

small array CGH study on SAs, and they suggested that investigations of CNVs involving genes that are imprinted in placenta in women with RAs could be worthwhile (18). This could be an important approach to identifying new genetic causes of RAs as “genomic variation” in imprinted genes in either of the parents could significantly increase the risk for SAs,see below.

1.5 Recurrence risk of SAs associated with CNVs

As discussed before there are a number of situations in which a CNV could potentially lead to RAs with up to 50% recurrence risk for SAs. A few examples are discussed below.

i) Incomplete penetrance or variable expressivity of the gene covered by the CNV

These phenomena can result in a large difference in the severity of a phenotype caused by the same variation, even within the same family (95), so that a CNV inherited from an apparantly healthy parent can cause an abnormal phenotype or even be lethal in the offspring.

ii) Germ line mosaicism in parents for the CNV

Mosaicism in parents for a mutation of a certain gene has been observed or suggested in the germ line for a variety of diseases, but the recurrence risk is only known for a few of these. This would manifest as a presumably *de novo* CNV in the fetus or child as the CNV is not detected in parental blood. The best known recurrence risk estimates due to germ line mosaicism are for Duchenne and Becker muscular dystropy (DMD/BMD). In a comprehensive study by Helderman-van den Enden *et al.* (2009) the recurrence risk for DMB/BMD was estimated to be 15.6% for proximal losses and 6.4% for distal losses if the risk haplotype was transmitted (96). The recurrence risk for unhaplotyped *de novo* DMD/BMD was estimated to be 4.3%.

However, the recurrence risk for most *de novo* copy number variations appears to be much lower, and limited data on recurrence risk is available. Rötthlisberger *et al.* (2007) collected data from the literature on the recurrence risk for *de novo* structural or combined structural and numeric chromosomal rearrangements, and calculated the recurrence risk to be less than 1% for a *de novo* i(21q) and even less than 0.3 % for all other rearrangements (97). This suggests that germ line mosaicism for copy number variations is at least not a common cause for RAs.

iii) The CNV (loss) unmasks a recessive allele

A loss could unmask a recessive allele on the other chromosome, resulting in loss of function of a gene in the fetus. The loss could be either *de novo* or inherited, causing an increased recurrence risk for abortion. If inherited the recurrence risk would be 25% or the same as for other recessive genomic disorders. The risk of this kind of variation causing up to three consecutive abortions is, however, low or equal to: $(¼)^3 = 1/64$.

iv) Parents share the same CNVs (losses) leading to homozygous loss in the fetus

Two parents could share the same hemizgyous losses leading to a homozygous loss in the fetus possibly resulting in SA. According to a study by McCarroll *et al.* (2008) the majority of CNVs between two individuals (~80%) arises from a limited set of common copy number polymorphisms (CNPs with

allele frequency >5%) (98). This suggests that it might be common for couples to share the same losses. Also, many common losses are listed in DGV in a hemizygous state, but it is not known what the effects are if the losses are homozygous. The recurrence risk is the same as in iii), i.e. this mechanism would be unlikely to cause up to three consecutive abortions.

v) The CNV contains an imprinted gene

When either of the parents (e.g. the mother) carries a loss of a gene that is imprinted in the other parent (e.g. the father), where the expression is turned off, this could lead to a SA with quite high recurrence risk or 50%. This would also give a relatively high risk for a loss causing up to three consecutive abortions or: $(1/2)^3 = 1/8$ making it a real possibility that losses of genes imprinted in the other parent could explain some RAs. This type of inheritance would also give a specific pattern of inheritance as the loss in the parent (for example the mother) with RAs would need to be inherited from a parent of the opposite gender or otherwise the result would be lethality. This could lead to a family pattern where members of the same gender (e.g. sisters) experience RAs, and they would be related through members of the opposite gender who are unaffacted carriers. This inheritance pattern could be exhibited in a larger family tree, depending on the origin of the loss. For example if a father, who has daughters with RAs, inherited the loss from his father, the grandfather could have more than one son who have daughters with RAs (Fig. 1).

vi) Synergistic effects of a number of CNVs or CNV(s) and non-allelic variation

The combined or synergistic effects of two (or more) CNVs, in particular losses, in a fetus could possibly lead to SAs. A similar two-hit model has been proposed for several neurodevelopmental disorders by Girirajan *et al*. (2012) as discussed before where the overall burden of CNVs creates differing sensitized backgrounds during development leading to different thresholds and disease outcomes (72).

vii) The CNV is not identical in size to that in the parent

In rare cases CNVs have been found to undergo further modification (e.g. expansion of a loss) when transmitted from carrier parent to an affected child (99). When traditional parental follow-up studies are performed with an alternative method such as FISH or qPCR this rare possibility can not be excluded. This would however not be likely to be a frequent cause of RAs.

viii) The CNV in a male fetus is X-linked and the mother is a non-manifesting carrier

This could cause an abnormal or lethal phenotype in a male fetus, for example in cases where the mother has skewed X-inactivation. The recurrence risk for this would be 25% if the transmission of the affected chromosome is not skewed.

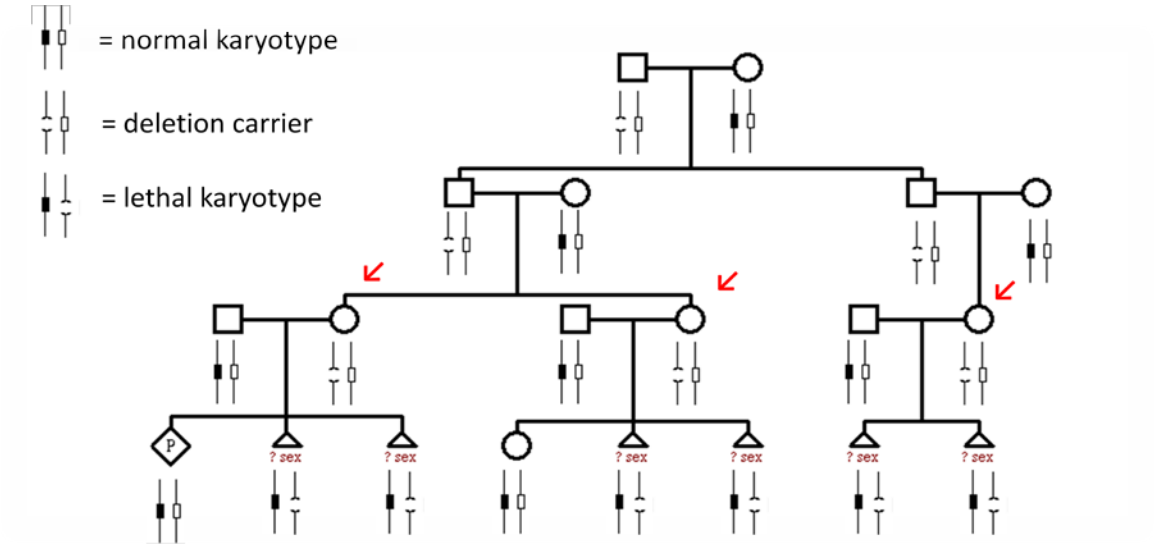


Figure 1. Pedigree showing a model of inheritance pattern of a loss of an imprinted gene resulting in a lethal karyotype and RAs. A grandfather transmits a loss of an imprinted (maternally expressed) gene to his sons, who further transmit it to their daughters, who have RAs (red arrows). In this case male family members are unaffected carriers and female family members experience RAs. An analogous pedigree with opposite genders would be seen if the mother passed on the imprinted gene. Black square = expression is turned off. White square = Gene is expressed. Gap = loss.

1.6 Exon-focused array CGH analysis of SAs

The main advantage of using high-resolution whole genome exon-array is that it allows detection of CNVs down to the exon-level, and allows clinicians to search for „submicroscopic miscarriage genes“. The likelihood of a CNV being pathogenic increases if it is of *de novo* origin, larger in size and includes genes that have been associated with disease (18). Also, this technique provides additional information on the frequency, size distribution, and gene content of CNVs in SAs.

In this study we used a 720K whole genome exon-focused CGH array from NimbleGen to screen for copy number variations in SA material from couples who have experienced recurrent abortions. This array had 720 thousand oligonucleotide probes covering the whole genome (probe length 50-75 mer), and increased density of probes in exon regions, with 65 bp median probe spacing in exons, and 7291 bp median probe spacing in the genomic backbone. This gives an extremely high resolution coverage, and the ability to detect small copy number changes down to the exon-level. We are not aware of an other study using such high resolution exon-focused array for analysis of SAs. Our hypothesis was that our array would detect smaller copy number variants than have been previously studied, and possibly discover new causes of SAs.

At the Department of Genetics and Molecular Medicine, Landspitali-University Hospital at Læknagarður we have set up a microarray facility for research and diagnostics. In order to find a suitable array CGH platform that could possibly be used for both research and diagnostic purpoes we tried a few platforms. This study on SAs was done as part of this effort. We chose this array as we wanted to test a high resolution array with the ability to detect CNVs in single genes and this array was just released when we started this project. Such a high resolution exon-focused array had never been used for analysis on SAs and it was produced by a major company in the field.

2 Aims

The aims of this study was to test if with a high resolution exon-focused CGH array we could detect causes of recurrent SAs in humans, where karyotype was normal.

Our hypothesis was that this array would detect smaller copy number variants then have been previously studied in SAs and with this approach we would be able to detect known or likely causes for SAs and possibly discover new candidate genes for SAs.

3 Materials and Methods

3.1 Subjects

Fetal tissue samples (chorionic villus samples (CVS), fibroblasts or umbilical cord) were obtained from women with a SA at Landspitali-National University Hospital during the investigation period (February 2009-July 2012). Blood samples were also collected from both parents.

All subjects were couples with RAs (three or more) and gestational age ≤20 weeks. They were already undergoing karyotyping, and receiving genetic counseling due to RAs. A questionnaire was completed with the parents (Appendix) about the current and previous abortions, as well as family history of SAs, and genetic disorders. Medical reports were examined as necessary. All personal identifiers were removed and samples plus information were coded.

This study was approved by the National Bioethics Committee of Iceland (no. 07-032) and the Data Protection Authority (no. S3307). Informed consent from both parents was obtained.

3.2 Cytogenetic Analysis

All fetal tissue samples were grossly examined, and separated from maternal deciduae using a dissecting microscope. Analysis of G-banded metaphase chromosomes was performed on all cultured fetal tissue samples and lithium-heparin blood samples from parents using standard procedures of the cytogenetics laboratory at Landspitali.

3.3 DNA Isolation

DNA from fetal samples was isolated directly from either chorionic villus sample or fibroblasts, and from EDTA blood samples from both parents, with a Gentra PureGene Blood kit from Qiagen according to the manufacturer´s instructions. In exceptional cases DNA was isolated from cultured fetal tissue cells instead of direct isolation. Samples were dissolved in 1 X TE buffer, pH 8.0 and, concentration measured on a NanoDrop ND-1000 Spectrophotometer (NanoDrop Technologies, Inc.).

3.4 Array CGH Analysis

Samples, with a normal karyotype, were analysed on a Human CGH 3x720K Whole-Genome Exon-Focused Array (Roche NimbleGen, Inc.). Each of three subarrays contained 720 K features, providing high-resolution analysis of the entire genome (7291 bp median probe spacing), and increased density of probes in exon regions (65 bp median probe spacing). Probe length was 50-75 mer.

Genomic DNA (1 µg) from fetus and reference sample (pools of six individuals from Promega Biotech AB, male cat no. G1471 and female cat no. G1521) was labeled with Cy3 (test) and Cy5 (reference) random nonamers (NimbleGen Dual-Color DNA Labeling Kit, product no. 0637025001), according to online NimbleGen Array User´s Guide:CGH Arrays & CGH/LOH Arrays. The labelled test and reference samples were combined, denatured, and hybridized to the array at 42°C for 40 +/-2 hours in the Maui hybridization system. After hybridization the arrays were washed according to online NimbleGen Array User´s Guide and scanned at 2 µm with the MS 200 Microarray scanner and the MS

200 Data Collection Software at the Roche NimbleGen Service Laboratory. DNA sequence coordinates were mapped to the National Center for Biotechnology Information Build 36 (UCSC hg18).

A custom site-specific 4x44K CGH array (Agilent Technologies) was designed for the *GDF6* gene. Probes were selected with SureDesign from the Agilent High Definition Database. The array had increased probe density (78.7% coverage) in a 100 kb area covering the *GDF6* gene. Median probe spacing in this area was 86 bp and each probe was replicated twice. Labeling, hybridization, and washing was performed according to Perkin Elmer´s protocol. Promega Biotech AB male sample (cat no. G1471) was used as reference. The array was scanned at 2 µm with the MS 200 NimbleGen Microarray scanner and analysed with the Agilent Cytogenomics software using a default analysis method.

3.4.1 Data Analysis

Grid alignment, feature extraction, and primary data analysis was performed with Nimblescan software by Roche NimbleGen´s Service Laboratory. Normalized intensity ratios for the 720K exon-focused array were imported into Nexus Copy Number v6.1 software for CNV calling, and statistical analysis. The FASST2 algorithm was used for analysis, with an average stringency. Systematic correction was applied to remove wave artifacts. Default threshold settings were used, and were as follows: significance threshold $1.0x10^{-5}$, $\log_2$ threshold were 0.23 for gains, -0.5 for losses, 1.0 for high gains, -1.1 for big/homozygous losses, 1.2 for 3:1 sex chromosome gain, and 1.7 for 4:1 sex chromosome gain. Three percent of outliers were removed. CNVs with <3 probes were filtered out, and also CNVs with over 50% overlap with DGV. Probes were recentered with median settings. Candidate rare (<50% overlap with DGV), large (≥50 kb) CNVs were further filtered based on visual inspection, due to waviness. Samples were considered to have an acceptable quality if they had a QC (quality control) score ≤0.15-0.20 calculated in the Nexus Copy Number software and/or a mad.1dr (median absolute deviation, 1st derivative) ≤0.23.

3.4.2 Database Search

The CNVs were inspected with regard to size, gene content, variation type, imprinting status, and presence in healthy controls in Database of Genomic Variants (http://www.projects.tcag.ca/variation/). Gene content of filtered CNVs were analysed in the UCSC Genome Browser (http://genome.ucsc.edu/), and/or OMIM (http://www.omim.org/). Search for losses (and gains) of imprinted and imprinting candidate genes was performed with a Javascript program (courtesy of Dr. Martin I. Sigurðsson) as well as with the Nexus Copy Number software. Large (>50 kb) and rare (or unique) CNV events (<50% overlap with DGV) were compared to known pathogenic CNVs in the International Standards for Cytogenomic Arrays (ISCA) database (http://www.iscaconsortium.org/) (now at http://www.iscaconsortium.org/) and DECIPHER (http://www.decipher.sanger.ac.uk/). Small CNVs (<50 kb) were considered as candidate variations for causing SAs if the CNV produces the same type of mutations that are known to be lethal or cause disease relevant to embryonic development in OMIM or in the Mouse Genome Informatics (MGI) database (http://www.informatics.jax.org/). The gene content of all unfiltered losses was furthermore inspected in the (MGI) database.

3.4.3 Gene Ontology (GO) and Network Analysis

Enrichment analysis on selected regions was performed in the Nexus Copy Number software to identify GO terms that are significantly overrepresented and genes annotated to these terms within the aberrant regions. The analysis was first performed separately on all CNVs, and all losses and GO terms within the classification group biological process were identified. The dataset was filtered to include only rare CNV events (<50% overlap with DGV) and the enrichments analysis was repeated as before. Three p-value measures were provided, i.e. a standard p-value, Markov Process (MP) p-value, and a False discovery rate (FDR)-corrected (for multiple testing) p-value (Q-bound). Only Q-bound values <0.05 were considered as significant.

Network analysis was performed on all losses in all fetuses collectively as well as on all CNVs, and all losses in each fetus separately on a list of genes overlapping CNVs (all and losses) using the String v9.1 database (http://www.string-db.org/), which mines protein-protein interaction data for direct and indirect interactions. Functional enrichment analysis for GO terms associated with biological process was also performed for each sample separately.

3.5 Quantitative-PCR (qPCR)

3.5.1 TaqMan Copy Number Assay

CNVs identified with Array CGH were confirmed with quantitative PCR using predesigned TaqMan Copy Number Assays from Applied Biosystems, and according to the manufacturer´s protocol or with SYBR GreenI-based methods designed in house (Supplementary Table 1, Appendix).

The TaqMan Copy Number Assay, which consists of a TaqMan minor groove binding (MGB) probe labeled with a FAM dye and unlabeled PCR primers, was run simultaneously with a TaqMan Copy Number Reference Assay (VIC dye-labeled TAMRA probe for the RNaseP H1RNA gene) in a duplex real time polymerase chain reaction. The copy number assay detects a target gene or genomic segment of interest, and the reference assay detects a sequence known to exist in two copies in a diploid genome, i.e. the RNaseP H1RNA gene. A calibrator sample, consisting of pools of six individuals (Promega Biotech AB, male cat no. G1471 and female cat no. G1521), was used as an external standard for comparison.

Duplex reactions were performed in four replicates in a 20 µl reaction volume with 20 ng of genomic DNA, 1X TaqMan Genotyping Master Mix, and 1X TaqMan Copy Number/Reference Assay. The Assay was performed in a 96 well MicroAmp Optical Reaction Plate (Applied Biosystems) on an ABI 7500 Real Time PCR System using the Relative Quantitation – Comparative CT (ddCt) setup. The thermal cycling conditions were: 95°C for 10 min, followed by 40 cycles of denaturation at 95°C for 15 sec, and amplification at 60°C for 60 sec. Analysis of results was performed using the 7500 Real Time PCR System (version 2.0) software, with the settings: Manual CT threshold – 0.2 and autobaseline, according to manufacturer´s protocol, using the comparative CT (ddCt) method.

3.5.2 SYBR Green I Copy Number Analysis

New quantification assays based on SYBR Green I were established for candidate CNV genes on a LightCycler 2.0 Carousel-based instrument from Roche. SYBR GreenI Assay was generated for loss

of the *KIAA1217* gene. The *LRRC8D* gene was used as a reference (as it was known to be present in two copies in a diploid genome and devoid of CNVs in healthy individuals). A calibrator sample, consisting of a pool of 10 random individuals (of the same gender) was used as an external standard for comparison.

Primers for candidate CNVs were designed by with the Primer3 software (http://www.frodo.wi.mit.edu/) (Supplementary Table 2, Appendix). Reactions were performed in three replicates in a 10 µl reaction volume in 20 µl capillaries with 20 ng of genomic DNA, 0.1 µM of each primer (Eurofins MWG GmbH), 1X Idaho buffer with 4 mM $MgCl_2$ (Idaho Technology), 0.2 mM of each dNTP (Fermentas), 0.05 U DreamTaq polymerase (Fermentas), and 1X SYBR Green I dye (BMA. BioWhittaker Molecular Applications). The thermal cycling conditions were: denaturation at 95°C for 3 min, 45 cycles of 5 sec at 95°C, 5 sec at 62°C and 7 sec at 72°C. This quantification reaction was followed by melting curve analysis (to verify product specificity) at the following conditions: starting temperature at 95°C with temperature gradient - 20°C/sec, followed by 67°C for 2 min, and temperature increase with 0.1°C/sec ending at 95°C. The analysis was performed with the Relative Quantification (Comparative CT (ddCt) method) in the LightCycler software (version 4.1).

3.6 Meta-analysis on previous studies on SAs

A PubMed search was performed looking for studies using array CGH analysis on SAs. The proportion of SAs reported with either a CNV and/or a unique (rare) CNV with <50% overlap with DGV was calculated in SAs that had a normal karyotype based on karyotypic analysis. Less that 50% overlap with DGV is a commonly used indicator of clinically significant CNVs (18, 100).

4 Results

4.1 Demographic information

Forty three SA samples (from thirty four couples) were included in the study (Table 1). One couple had five samples in the study and five couples had two samples. Mean maternal age at the time of pregnancy loss was 35 years (range 24-43) and range of gestational age was 5-20 weeks. None of the couples who had a fetus with normal karyotype based on conventional karyotypic analysis had a family history (in other members of family) of three or more SAs (two couples had family history of two or more SAs). Out of the couples who had a fetus with normal karyotype, there were two women with bicornuate uteruses, one of which also had deep vein thrombosis and factor XIII deficiency. Information including answers on questionnaire was missing from one couple with a normal karyotype.

4.2 Conventional karyotypic analysis

Karyotyping was successful on 37 samples, but analysis failed in six samples (14%) due to e.g. tissue culture failure, MCC or suboptimal quality of tissue (Tables 1 and 2). Three samples were suspected to have MCC and confirmed by microsatellite genotyping (data not shown). Two of those samples consisted entirely of maternal cells and one had low-level MCC (Table 1). Overall, 23 of 37 (62.2%) successfully karyotyped samples (or 53.5% of all samples) had an abnormal karyotype based on conventional karyotypic analysis. Out of all samples 21 (56.8%) karyotypes were assumed to be causative for the pregnancy loss (Tables 1 and 2). The two abnormal karyotypes, not assumed to be lethal, were mos 47,XYY[28]/46,XY[27] (case 39) and 46,XX,inv(12)(q15q24) (case 42). The majority of chromosomal abnormalities were aneuploidies (12 of 23 or 52.5%). There was one polyploidy, one inherited inversion on chromosome 12, and nine other abnormalities (39.1%), including e.g. mosaicism and double aneuploides. A normal karyotype was found in 14 of 37 samples (37.8%), and the male to female ratio of normal karyotypes was 0.75 (or 6 to 8). Conventional karyotypic analysis was performed for 32 out of 34 couples, and was normal in all cases, except case 42, where the mother had the same chromosome 12 inversion as was detected in the fetus. This case was included in the study as the causes for the two SAs from this couple where not explained by the karyotype (one sample had a 46,XX,inv(12)(q15q24) karyotype (case 42) and the other one 46,XX (case 43)).

4.3 Array CGH analysis

4.3.1 CNV findings and genomic distribution

Exon-focused array CGH analysis was performed on 14 SA samples (from 10 couples) with an apparently normal karyotype based on cytogenetic analysis. This included one couple which had four SAs with normal karyotype and one couple with two SAs and eight couples with one SA each. The Quality control parameter calculated in Nexus Copy Number was within an acceptable range (<0.15-0.20) in all samples, except for one (case no. 13 with a quality score ~0.26) (Table 3).

Table 1. Results from karyotypic analysis.

Case no.	Sample type	Karyotype	MA (years)	Lethal karyotype	Presumably detectable by array CGH	Comments
1	CVS	46,XX	35	No	Yes	
2	CVS	45,X	43	Yes	Yes	
3	CVS	mos92,XXXX[7]/46,XX[5]	32	Yes	No	
4	CVS	47,XX,+20 plenty of heterochromatin in one chr20	38	Yes	Yes	
5	CVS	47,XX,+22 (one cell 48,XX,+9,+22)	30	Yes	Yes, at least trisomy 22	
6	CVS	mos46,XX[45]/45,X[4] /47,XXX[2][g]	36	Probably	?	Low level MCC
7	Unknown	46,XX[h]	24	NA	No	MCC
8	Unknown	46,XX[h]	30	NA	No	MCC
9	CVS	47,XY,+18	37	Yes	Yes	
10	CVS	47,XX+4[10]/48,XX, +4,22[2]	37	Yes	Yes, at least trisomy 4	
11	CVS	47,XX,+add(2)(p22)	39	Yes	Yes	Cells cultured for analysis
12	CVS	47,XY,+13	39	Yes	Yes	
13	CVS	46,XX	33	No	Yes	Cells cultured for analysis
14	NA	NA	26	NA	No	Fetal tissue not detectable
15	Skin fibroblasts	NA	37	NA	Yes	Cells did not grow in vitro
16	CVS	47,XX,+16	36	Yes	Yes	
17	CVS	68,XXX,-22	33	Yes	Partly, at least chr22 monosomy	
18	CVS	46,XX	32	No	Yes	
19	CVS	47,XY,+4	34	Yes	Yes	
20	CVS	47,XX,+1[10]/47XX, t(2;20)+16[2]	35	Yes	Yes, at least trisomy 1	
21	CVS	mos47,XX,+11[10] /46,XX[2]	28	Yes	Yes	
22	CVS	46,XX	30	No	Yes	
23	CVS	46,XY	36	No	Yes	
24	CVS	70,XXY,+16[7]/69,XXY[7] /46,XY[2]	38	Yes	Yes, at least suspect triploidy	
25	CVS	46,XX	31	No	Yes	
26	NA	NA			No	No fetal tissue sent for analysis
27	CVS	47,XY,+15	38	Yes	Yes	
28	NA	NA		NA	No	Fetal tissue not fit for analysis
29[a]	CVS	mos47,XX,+10[13] /46,XX[2]	35	Yes	Yes	
30[a]	CVS	47,XY,+21	35	Yes	Yes	
31[b]	CVS	46,XY	34	No	Yes	Bicornuate uterus
32[b]	CVS	46,XY	35	No	Yes	Bicornuate uterus
33[c]	CVS	47,XY,+22	29	Yes	Yes	

Table 1. Continued.

34[c]	CVS	46,XX	40	No	Yes	
35[c]	Skin fibroblasts	46,XX	41	No	Yes	
36[c]	CVS	46,XY	41	No	Yes	
37[c]	Skin fibroblasts	46,XY	42	No	Yes	
38[d]	CVS	45,X	27	Yes	Yes	
39[d]	CVS	mos 47,XYY[28]/46,XY[27]	29	No	Probably	
40[e]	CVS	69,XXY	31	Yes	Yes/suspect	
41[e]	CVS	46,XY	33	No	Yes	
42[f]	CVS	46,XX,inv(12)(q15q24)	34	No	?	Inversion inherited from mother. Bicornuate uterus, deep vein thrombosis, factor 13 deficiency
43[f]	CVS	46,XX	35	No	Yes	

[a-f] Each superscript letter (a-f) represents SAs from the same couple. [g]Low-level MCC suspected and confirmed with genotyping. [h]Cases of maternal cell contamination (MCC). Unknown tissue consists of maternal cells only according to genotyping. Abbreviations: MA= maternal age, NA= not available.

Table 2. Summary of results from karyotypic analysis

Karyotypic analysis (Number)	Percentage (%)
Normal (14)	32.6
Abnormal (23)	53.5
- Aneuploidy (12)	27.9
- Polyploidy (1)	2.3
- Inherited inversion (1)	2.3
- Others (9)*	20.9
Failure of karyotypic analysis (6)	14.0
- Tissue culture failure (1)	2.3
- MCC (2)	4.7
- Fetal tissue not detectable/ fit for analysis (2)	4.7
- Fetal tissue not sent (1)	2.3
Total # (43)	100

*Others include mosaic aneuploidies (3), double aneuploidiy mosaicism (2), double aneuploidy mosaicism plus a balanced translocation (1), aneuploidy mosaicism plus polyploidy (1), polyploidy mosaicism (1), and polyploidy plus monosomy (1).

Case 13 that failed the quality parameter was excluded from statistical analysis on all SAs (Tables 3-5, Fig. 4 and 8A). However, we analysed this sample individually and decided to analyse some CNVs further (Fig. 2, 6A and 7B, Tables 6, 7, and 9-11). A total of 112 CNVs were identified in case 13.

A total of 1723 CNVs were identified in 13 samples with an average of 133 CNVs per sample (range 1-218). CNVs covered a total of 5.91 Mb of nucleotides on average per sample (5.18 Mb of gains and 0.74 Mb of losses) and range of nucleotide coverage per sample was 537 bp-12.3 Mb. The majority of CNVs identified were gains and the ratio of gain (one or more copies) to loss (one or two copy) was 6.4 (1490 to 233). Tables 4A, B and C summarize the overlap of CNVs with various relevant genomic features and the proportion that disrupted genes. A large proportion of the CNVs overlapped genes (1598 of 1723 or 92.7%), including OMIM genes (1148 of 1723 or 66.6%) of which 298 (16.0% of all CNVs) were OMIM morbid genes (Table 4A). The average number of genes (including microRNAs) that were overlapped by CNVs per sample was 299 genes (range 1-458) and the average number of genes (including microRNAs) per CNV was 2.26. Majority of the CNVs disrupted genes i.e. had at least one breakpoint inside a gene (Table 4C). Most of the CNVs were common (1259 of 1723 or 73.1%), although 464 of 1723 (or 26.9%) had <50% overlap with DGV. Out of these rare CNVs there were 392 (or 84.5%) that overlapped with OMIM genes and 105 (or 22.6%) that overlapped with OMIM morbid genes (Table 4B). There was a slight overlap of CNVs with DECIPHER syndromes (117 of 1723 or 6.8%) and majority of them were gains (7.4% for all gains vs. 3.0% for all losses), but the CNVs only overlapped a small portion of the syndrome areas. Out of the rare CNVs (<50% overlap with DGV) there were 15 of 464 (or 3.2%) that overlapped with DECIPHER syndromes. The average size and median size of those CNVs was smaller than for CNVs overall (22.8 and 5.6 kb average and median size for rare CNVs vs. 44.6 and 6.5 kb for CNVs overall respectively). There was no overlap of rare losses with OMIM morbid genes and DECIPHER syndromes.

A larger proportion of gains overlapped genes compared to losses (1415 of 1490 or 95.0% of all gains vs. 183 of 233 or 78.5% of all losses), although the difference was less significant when calculating the relative nucleotide coverage of gains (88.1% of all gains) vs. losses (79.2% of all losses) that overlapped genes. However, when only homozygous (two copy) losses are compared to gains there is much less overlap with genes and syndrome areas. Only 50.0% of homozygous losses overlapped genes and they were all common CNVs with >50% overlap with DGV. An exception of a rare CNV (with <50% overlap with DGV) was a loss of 4.6 kb overlapping part of the *GDF6* gene in case 13, which was excluded from the summary of CNVs overall (see Discussion in section 4.3.4). None of the homozygous losses overlapped DECIPHER syndromes. It is also noteworthy that the average size of two copy gains was significantly smaller (~9 kb) than the other CNV types.

None of the losses overlapped imprinted genes based on the Nexus Copy Number software, but some gains did (39 of 1723 or 2.3%). However, results from questionnaires and genetic counseling revealed one pedigree where members of the same gender (two sisters) had experienced RAs (at least two) (Fig. 2). This family could possibly fit our model for inheritance of losses (or mutations) of imprinted genes.

Table 3. Summary of CNV analysis statistics in SAs with normal karyotype.

Case no.	Quality*	One copy gain	One copy loss	Two or more copy gain	Two copy loss	Total CNVs
1	3.31E-02	1	0	0	0	1
32	3.63E-02	54	4	1	0	59
18	2.21E-02	79	17	6	5	107
35	1.54E-02	30	3	1	0	34
31	1.47E-02	188	15	11	4	218
36	1.54E-02	173	19	9	3	204
23	1.17E-02	142	27	16	9	194
34	1.97E-02	146	17	4	2	169
22	1.43E-02	96	20	5	9	130
37	1.51E-02	137	9	3	5	154
25	1.16E-02	100	21	8	6	135
43	1.33E-02	89	11	3	6	109
41	1.40E-02	183	14	5	7	209
SUM		**1418**	**177**	**72**	**56**	**1723**
MEAN		109	13.6	5.5	4.3	133
MEDIAN		100	15	5	5	135
MIN		1	0	0	0	1
MAX		188	27	16	9	218
AVERAGE SIZE		47003	40819	9005	42243	44625
MEDIAN SIZE		6143	20796	3018	33101	6544
MIN SIZE		61	27	335	34	27
MAX SIZE		1357210	399300	43961	161559	1357210

*This quality score is calculated in the Nexus Copy Number software and represents the probe to probe variance in a sample, see Materials and Methods.
Size of CNVs is given in bp.

Table 4A. Overlaps of CNVs with relevant genomic features.

	One copy gain Number (%)	One copy loss Number (%)	Two or more copy gain Number (%)	Two copy loss Number (%)	Total CNVs Number (%)
Overlapping genes (including microRNAs)	1355 (95.6)	155 (87.6)	60 (83.3)	28 (50.0)	1598 (92.7)
Overlapping imprinted genes	39 (2.8)	0 (0.0)	0 (0.0)	0 (0.0)	39 (2.3)
Overlapping OMIM genes	991 (69.9)	95 (53.7)	44 (61.1)	18 (32.1)	1148 (66.6)
Overlapping OMIM morbid genes	239 (16.9)	24 (13.6)	10 (13.9)	3 (5.4)	276 (16.0)
Overlapping DECIPHER syndromes	108 (7.6)	7 (4.0)	2 (2.8)	0 (0.0)	117 (6.8)
<50% overlap with DGV	422 (29.8)	22 (12.4)	20 (27.8)	0 (0.0)	464 (26.9)
Total number of CNVs	1418	177	72	56	1723 (100)

Table 4B. Overlaps of rare CNVs (<50% overlap with DGV) with relevant genomic features.

	One copy gain Number (%)	One copy loss Number (%)	Two or more copy gain Number (%)	Two copy loss Number (%)	Total CNVs Number (%)
Overlapping genes (including microRNAs)	415 (98.6)	21 (95.5)	20 (100)	0 (0.0)	456 (98.3)
Overlapping imprinted genes	2 (0.48)	0 (0.0)	0 (0.0)	0 (0.0)	2 (0.43)
Overlapping OMIM genes	364 (86.5)	10 (45.5)	18 (90.0)	0 (0.0)	392 (84.5)
Overlapping OMIM morbid genes	102 (24.2)	0 (0.0)	3 (15.0)	0 (0.0)	105 (22.6)
Overlapping DECIPHER syndromes	15 (3.6)	0 (0.0)	0 (0.0)	0 (0.0)	15 (3.2)
Total number of CNVs	421	22	20	0	464 (100)

Table 4C. Overlapping CNVs that disrupt genes with one or two breakpoints.

Disrupting genes	One copy gain Number (%)	One copy loss Number (%)	Two or more copy gain Number (%)	Two copy loss Number (%)	Total CNVs Number (%)
Overlapping genes (including microRNAs)	1210 (89.3)	132 (85.2)	60 (100)	26 (92.9)	1428 (89.4)
Overlapping imprinted genes	30 (76.9)	0 (0.0)	0 (0.0)	0 (0.0)	30 (76.9)
Overlapping OMIM genes	907 (91.5)	78 (82.1)	44 (100)	16 (88.9)	1045 (91.0)
Overlapping OMIM morbid genes	226 (94.6)	15 (62.5)	10 (100)	3 (100)	254 (92.0)
Overlapping DECIPHER syndromes	89 (82.4)	4 (57.1)	2 (100)	0 (0.0)	95 (81.2)
<50% overlap with DGV	394 (93.3)	18 (81.8)	20 (100)	0 (0.0)	432 (93.1)

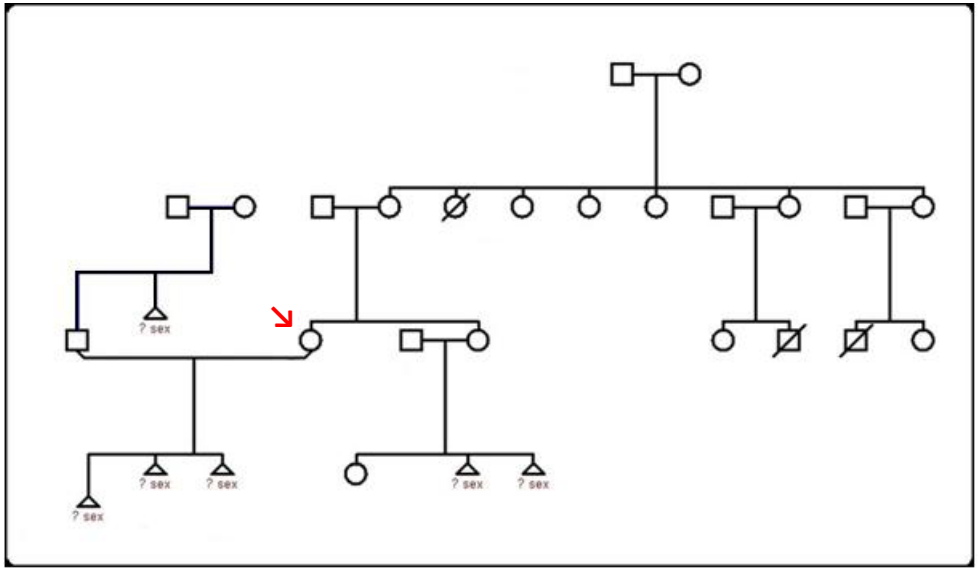


Figure 2. Pedigree chart of case 13 (red arrow) showing two sisters with RAs.

In cases where the same couple had more than one SA in the study (one couple had four samples and one had two), we checked if there were common CNVs among all samples. This could have been indicative of a common cause for the abortions. We did not find CNVs unique to the four SA cases from the same couple. Out of rare CNVs, there was one unique CNVs to the couple from the couple with two cases, i.e. a gain overlapping the *SP9* gene, although it was thought to be probably benign.

Overall the CNVs identified in this study mapped to all chromosomes (Fig. 3 and Table 5) although the distribution was somewhat uneven. The highest proportion of CNVs mapped to chromosome 1 (206 or 12.0%) and the lowest to the Y chromosome (7 or 0.41%), but there was a relatively high proportion of CNVs that mapped to chromosome 19 (149 or 8.65%). The proportion of nucleotides covered with redundant CNVs was highest for chromosome 15 (8.13%) and lowest for chromosome 13 (0.30%). There were only three losses that mapped to the X chromosome, all of which were in female fetuses.

4.3.2 Size distribution of CNVs

The average size of CNVs was 44.625 bp and median size 6544 bp (range 27-1.357.210 bp) (Table 3). Size distribution of CNVs is shown in Fig. 4. The majority of the CNVs identified were small with 80.4% smaller than 50 kb and 60.2% smaller than 10 kb. Also, there was a considerable amount of very small CNVs < 1 kb or 62 of 1723 (3.6%). There was a slight difference in the size distribution between gains and losses as the highest relative number of gains were in the 1-10 kb size range (61.7%) whereas most losses were in the 10-50 kb size range (45.9%) (Fig. 4B,C). However, there was also an excess of large CNVs compared to other studies on populations of living subjects (23) as 8 samples of 13 (61.5%) had a CNV > 500 kb and 2 of 13 samples (15.4%) had a CNV > 1 Mb (see Discussion in section 5.2.2).

4.3.3 Rare, large CNVs in SAs

Table 6 shows eight non-redundant large (>50 kb) rare (<50% overlap with DGV) CNVs found in ten SA cases, including case 13. These CNVs ranged in size from 76-399 kb and mapped to six autosomal chromosomes except two that mapped to chromosome X (Table 6). The CNVs overlapped 24 genes (including microRNAs) in total, and three on average (range 1-6). Only ten genes were OMIM genes and none of them were OMIM morbid genes. Two possible candidate CNVs for causing SA were selected based on size and gene content (i.e. 289 kb gain in chrXq22.1-q22.2 and 242 kb loss in chr10p12.1) (Fig. 5A, B) and verified with qPCR and tested in parents to determine the origin of the imbalance (Tables 7 and 8). The 242 kb loss in chr10p12.1 found in case 13 was found to be paternally inherited (Fig. 5A, Table 7). It overlapped the *KIAA1217* gene, which codes for the sickle tail isoform 1 protein. This gene is required for normal development of intervertebral disks.

The 289 kb gain in chrXq22.1-q22.2 was found in three (two females and one male) out of four karyotypic normal SAs from the same couple (cases 34-36) and it was found to be maternally inherited (Fig. 5B, Table 8). It overlapped the Pelizaeus-Merzbacher disease duplication/deletion area, and contained six genes, i.e. *NGFRAP1*, *RAB40A*, *TCEAL3*, *TCEAL4*, *TCEAL7* and *WBP5*.

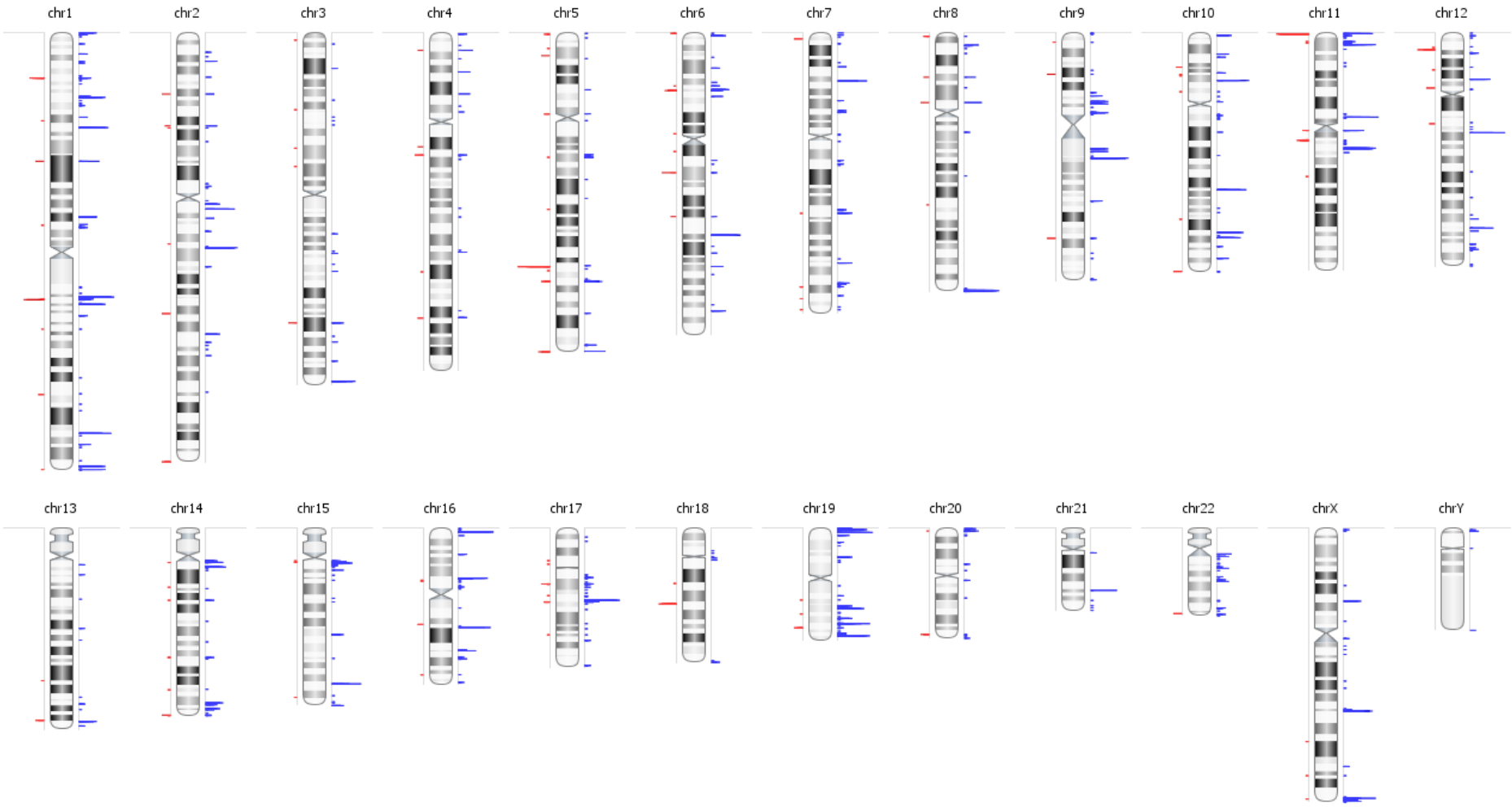


Figure 3. Whole genome view showing distribution of all gains (blue) and losses (red) on chromosomes.

Table 5. Distribution of CNVs in the genome.

Chr.	# Gains	# Losses	# CNVs	% of all CNVs	Nucleotides covered*	Chr. length	% Covered
1	182	24	206	12.0	5616159	247249719	2.27
2	80	11	91	5.28	2006846	242951149	0.83
3	31	7	38	2.21	2140758	199501827	1.07
4	34	18	52	3.02	2449331	191273063	1.28
5	56	31	87	5.05	2706705	180857866	1.50
6	66	19	85	4.93	2224734	170899992	1.30
7	70	7	77	4.47	5894919	158821424	3.71
8	68	7	75	4.35	3295973	146274826	2.25
9	70	7	77	4.47	10178797	140273252	7.26
10	76	8	84	4.88	2027951	135374737	1.50
11	105	24	129	7.49	1491398	134452384	1.11
12	61	13	74	4.29	2040291	132349534	1.54
13	22	4	26	1.51	338303	114142980	0.30
14	50	12	62	3.60	6416083	106368585	6.03
15	60	4	64	3.71	8159727	100338915	8.13
16	82	5	87	5.05	3084126	88827254	3.47
17	68	10	78	4.53	4775125	78774742	6.06
18	10	7	17	0.99	1243528	76117153	1.63
19	144	5	149	8.65	1622508	63811651	2.54
20	23	4	27	1.57	584594	62435964	0.94
21	14	0	14	0.81	165041	46944323	0.35
22	40	3	43	2.50	2592303	49691432	5.22
X	71	3	74	4.29	3541310	154913754	2.29
Y	7	0	7	0.41	2292221	57772954	3.97
Total	1490	233	1723	100	76888731	3080419480	2.50

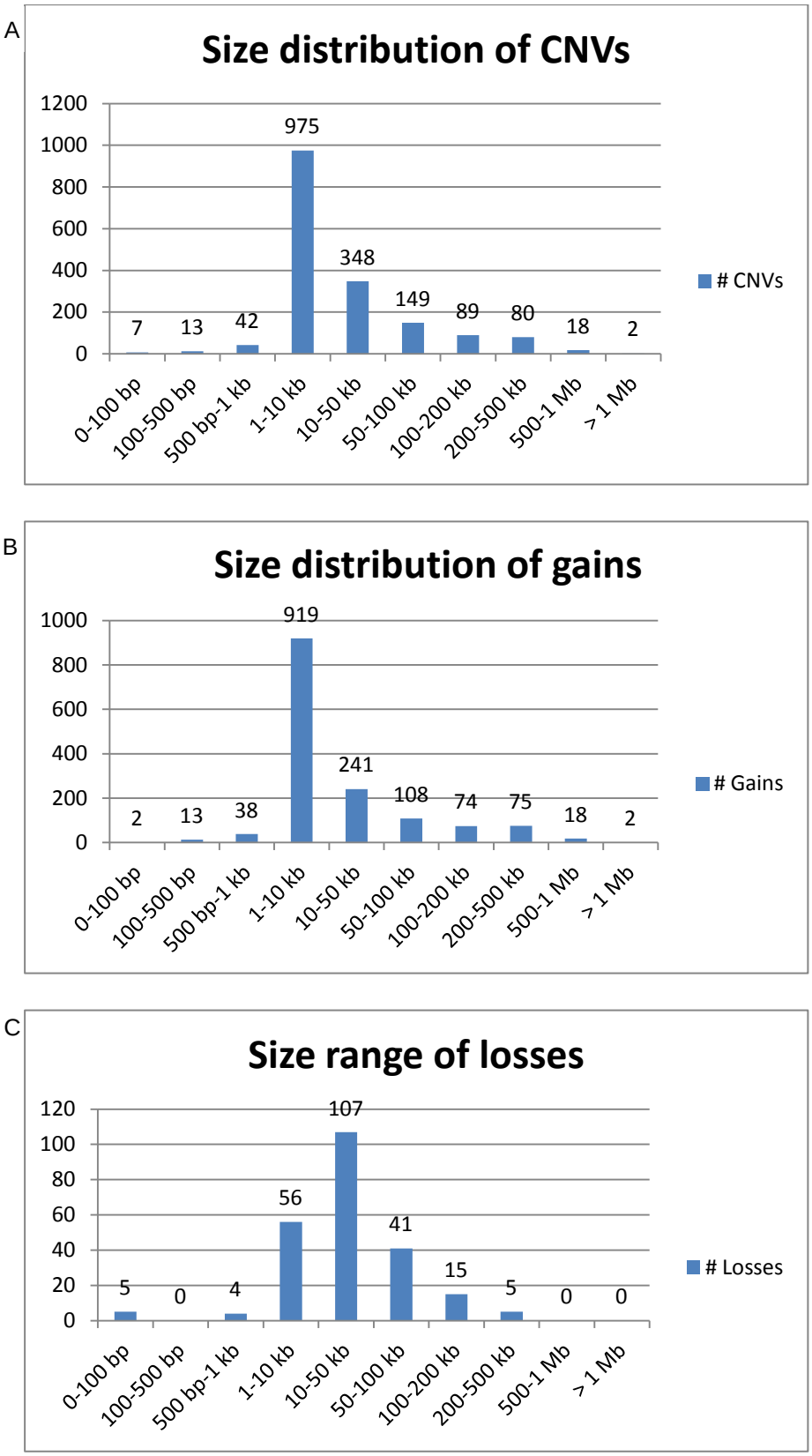


Figure 4. Size distribution of CNVs. A. All CNVs. B. All gains. C. All losses. X-axis shows size ranges. Y-axis shows number of CNVs.

Table 6. Rare, large copy number variants (>50 kb) with <50% overlap with Database of Genomic Variants.

Case no.	Region (Mb)	Cytoband Location	Region Length bp	Event	% of CNV Overlap	Log2 ratio	# Genes	Gene Symbols
13	chr10:24,248,484-24,490,604	p12.1	242.120	CN Loss[c]		-0.76	1	*KIAA1217*
18	chrX:86,404,371-86,803,183	q21.31	398.812	CN Gain	12.7	0.40	1	*KLHL4*
22	chr18:12,699,535-12,784,271	p11.21	84.736	CN Gain	4.77	0.37	2	*PSMG2, PTPN2*
23	chr1:41,245,289-41,400,335	p34.2	155.046	CN Gain	3.95	0.54	6	*CTPS, CTPS1, MIR5095, SCMH1, SLFNL1, SLFNL1-AS1*
32	chr16:82,885,845-82,961,963	q24.1	76.118	CN Loss	0.0	-0.60	2	*ATP2C2, WFDC1*
34, 35, 36[b]	chrX:102,462,042-102,750,723	q22.1 - q22.2	288.681	CN Gain[c]	22.4	0.45, 0.31, 0.81	6	*NGFRAP1, RAB40A, TCEAL3, TCEAL4, TCEAL7, WBP5*
35,37[b]	chr3:113,633,988-113,999,321	q13.2	365.333	CN Gain	2.76	0.34, 0.52	4	*ATG3, BTLA, CCDC80, SLC35A5*
41[a]	chr1:49,682,644-49,768,447	p33	85.803	CN Loss	23.9	-0.87	1	*AGBL4*

[a-b] Each superscript letter (a-b) represents SAs from the same couple. [c] These CNVs were verified with qPCR.

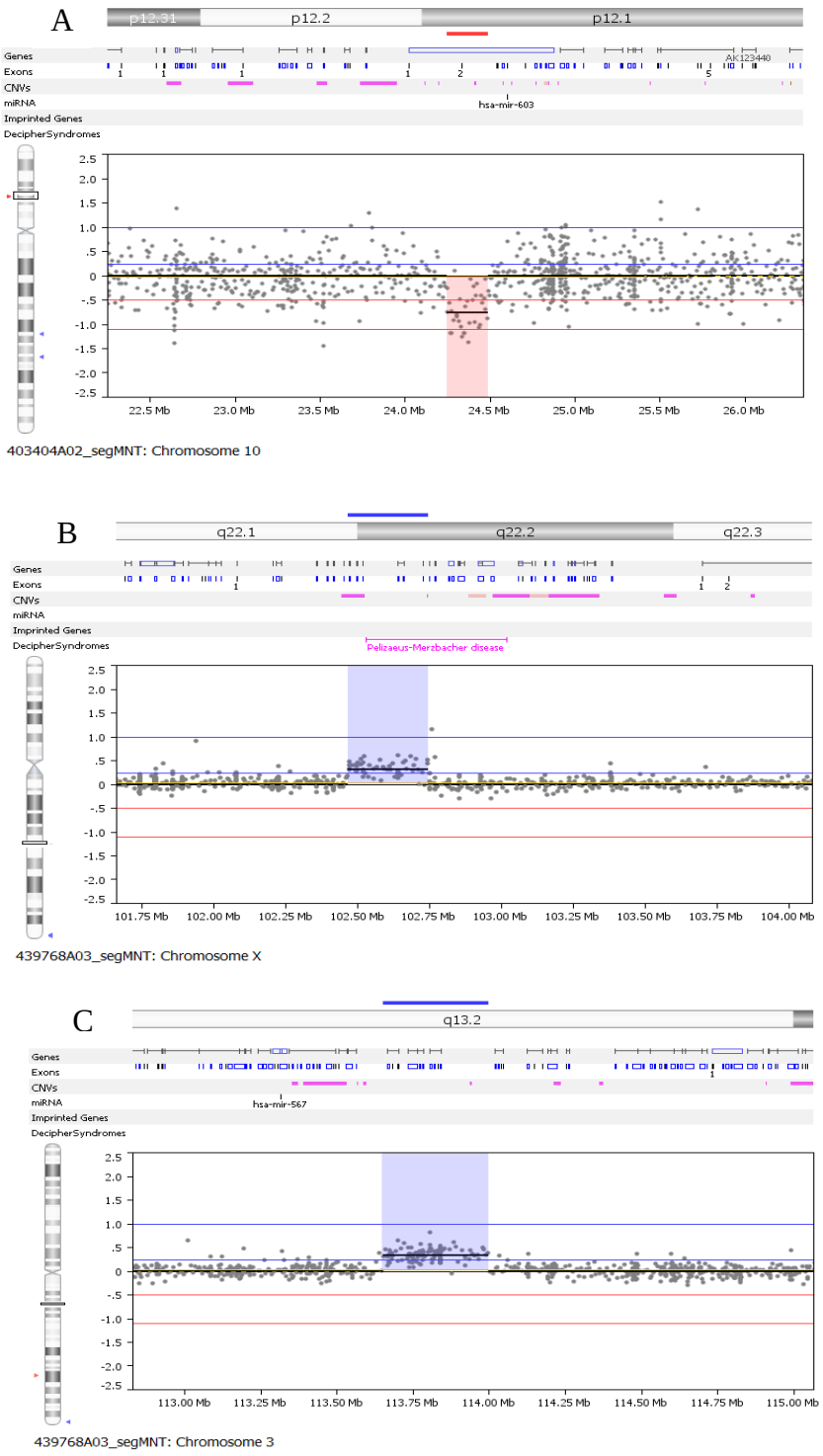


Figure 5. Graphical view of three rare, large CNVs. A. 242 kb loss of *KIAA1217* gene in chr10p12.1 in case 13. B-C. Graphical view of gains in case 35, detected in three (B) or two (C) samples (out of four with normal karyotype) from the same couple. B. 289 kb gain of *NGFRAP1, RAB40A, TCEAL3, TCEAL4, TCEAL7* and *WBP5* genes in chrXq22.1-22.2. This gain partly overlaps with Pelizaeus-Merzbacher disease. C. 365 kb gain of *ATG3, BTLA, CCDC80* and *SLC35A5* in chr3q13.2. Red highlighted areas indicate loss and blue highlighted areas indicate gain. Bars above show chromosomal position, and overlap of genes, exons, CNVs in DGV, miRNA, imprinted genes, and DECIPHER syndromes.

Table 7. Verification of a 242 kb loss in chr10p12.1 overlapping the KIAA1217 gene with SYBR GreenI analysis.

Case no.	Relative quanitification (RQ)	ΔCт SE
13	0.48	0.081
Mother	1.31	0.033
Father	0.61	0.103
Calibrator	1	0.171

Abbreviations: RQ = Relative quantification, Cт = Cycle threshold, SE = Standard error.

Table 8. Verification of a 289 kb gain in chrXq22.1-22.2overlapping the *NGFRAP1* gene with TaqMan Copy Number Assay.

Case no.	Gender	Relative quanitification (RQ)	ΔCт SE
34	Female	1.49	0.024
35	Female	1.34	0.034
36	Male	1.07*	0.060
Mother	Female	1.30	0.092
Father	Male	0.49	0.050
Calibrator	Female	1.00	0.031
Reference	Male	0.44	0.023

*The RQ value in this male is consistent with two copies (or a gain) when compared to a female calibrator. Abbreviations: RQ = Relative quantification, Cт = Cycle threshold, SE = Standard error.

However, Pelizaeus-Merzbacher disease is in a majority of cases caused by a gain of the *PLP1* gene, which was not duplicated in our case (see Discussion in section 5.2.3).

The same couple also had two SAs with a 365 kb gain in chr3q13.2 (cases 35 and 37) which overlapped four genes, i.e. *ATG3*, *BTLA*, *CCDC80* and *SLC35A5* (Fig. 5C). The most likely candidate gene is *ATG3*, which encodes an ubiquitin-like-conjugating enzyme, which is known to play a role in autophagy during cell death (see Discussion in section 5.2.3).

4.3.4 Heterozygous losses in SAs

There were 177 heterozygous (or one copy) losses identified in 12 of the 13 SAs included in the analysis on all SAs together (Table 3). None of them overlapped imprinted genes, but 95 (53.7%) overlapped OMIM genes and 24 (13.6%) overlapped OMIM morbid genes (Table 4). However the phenotypic effects of none of losses were clearly related to SAs based on OMIM database search. The genes that were lost were also analysed in the MGI database to check if the phenotypic effects in the mice could reveal any candidate genes for a SA. Twelve heterozygous losses in eight cases included genes that were either homozygous embryonic or neonatal lethal in mice or were associated with severe structural abnormalities (Table 9), which suggests that they could be lethal in case there was a loss of function mutation (or an other loss) on the other allele. Case 13, which was analysed separately, was included in this table and the loss was further analysed, see section 4.3.3. Seven of those had 100% of overlap with DGV, one had 3.42% overlap, and two had no overlap. Most of these losses were small (eight out of ten were around 10 kb or smaller). These losses overlapped 1-3 genes (at least partly).

4.3.5 Homozygous losses in SAs

Overall, there were 56 homozygous losses identified in 11 of the 13 SAs analysed, included in the statistical analysis on all SAs together (Table 3). The homozygous losses had limited overlap with

genes (50.0%) as discussed before. Table 10 shows a list of homozygous losses that overlap with genes. Case 13 was included in this table and the loss was further analysed (see below). Nineteen homozygous losses overlapped a total of 12 OMIM genes in nine cases. Three of those genes were OMIM morbid genes, i.e. *HLA-DRB1, HLA-DRB5,* and *GDF6.* The homozygous losses were also all common CNVs, except for one loss overlapping the *GDF6* gene in case 13. A few of those CNVs known to be common (in a hemizygous state) according to DGV were nevertheless selected as possible candidates for causing SAs.

One small apparent homozygous loss of 34 bp in the *TAF4* gene (case 22), which codes for TAF4 RNA polymerase II, TATA box binding protein (TBP)-associated factor (Fig. 6A). TAF4 is one of the larger subunits of the transcription factor IID (TFIID) and it´s expression varies during development and cell differentiation. It has been suggested to have a connection to neurodegenerative disorders by aberrant binding to expanded polyglutamine stretches (101). Mice homozygous for a loss of this gene die at the embryo stage. The 34 bp loss was in exon 1, which is in the only known protein-coding transcript of the gene, making it a possible candidate as a cause for SA. There are, however, eight other shorter transcript of the gene, some of which are putative protein-coding transcripts, that don´t overlap with the lost sequence. This loss overlapped a similar variant in DGV (a heterozygous loss identified in 3 of 451 individuals (0.67%) (64)), suggesting that it is possible that the parents could be heterozygous carriers. Two other cases in this study (cases 25 and 36) also had an overlapping heterozygous loss in this area (Table 9), suggesting that this may be a more common variant (or that these variants contributed to the SAs). An effort was made trying to verify this loss with sequencing over the breakpoints. This area appeared to be quite repeat-rich making primer-design challenging and four primers that were designed for this area produced self-primed products. Two primer pairs with binding site overlapping the deleted area were also designed, but either produced unspecific or no products.

An other apparent homozygous loss of 4.6 kb overlapped exon 2 in the *GDF6* gene (case 13) which codes for growth differentiation factor 6, which is a member of the bone morphogenetic protein (BMP) family and the TGF-beta superfamily of secreted signaling molecules (Fig. 6B). It is required for normal formation of some bones and joints in the limbs, skull, and axial skeleton. Mutations in this gene are e.g. associated with microphthalmia and autosomal dominant Klippel-Feil syndrome, which is a congenital disorder of spinal segmentation. Homozygous null mice show multiple joint and skeletal patterning defects affecting the extremities, inner ear, and skull. This loss does not overlap CNVs in DGV. Two qPCR assays showed a relative quantification (RQ) value of 0.13 and 0.15 for this loss (Table 11). Repeated analysis showed comparable results. Furthermore, both parents seemed to be heterozygous or have a lower copy number compared to a calibrator for this loss as well (Table 11) (see Discussion in section 5.2.4 and 5.5). It is noteworthy that the RQ value in the fetus, was however, significantly lower than in the parents. This could indicate that the fetus had a homozygous loss that was inherited from heterozygous parents. One possible explanation for this could have been that there was an amplification of one of the *GDF6* paralogs in the qPCR reactions.

Table 9. List of heterozygous losses covering genes that are homozygous lethal or causative of severe structural abnormalities in mice.

Case no.	Chromosome Region	Cytoband	Length (bp)	% of CNV Overlap	Log2 ratio	Probes	# Genes	Gene Symbols	# OMIM Genes	# OMIM Morbid genes
13	chr7:150,283,459-150,287,092	q36.1	3634	100.0	-0.90	9	1	*KCNH2*	1	1
13	chr11:627,481-634,595	p15.5	7115	100.0	-0.88	14	2	*DEAF1, DRD4*	2	1
13	chr5:134,391,871-134,399,878	q31.1	8008	100.0	-0.84	11	2	*AK026965, PITX1*	2	1
13	chr10:24,248,484-24,490,604	p12.1	242121	3.42	-0.76	33	1	*KIAA1217*	0	0
18	chr2:119,318,782-119,322,154	q14.2	3373	0.0	-0.63	9	1	*EN1*	1	0
18	chr10:23,519,362-23,522,756	p12.2	3395	100.0	-0.63	9	1	*PTF1A*	1	1
34	chr14:33,339,116-33,343,670	q13.1	4555	100.0	-0.69	7	1	*NPAS3*	1	0
25	chr12:20,913,416-20,923,760	p12.2	10345	100.0	-0.59	7	3	*LST-3TM12, LST3, SLCO1B3*	1	0
25	chr20:60,074,157-60,074,190	q13.33	34	100.0	0.73	*5*	1	*TAF4*	1	0
36	chr20:60,074,164-60,074,190	q13.33	27	100.0	-0.68	*4*	1	*TAF4*	1	0
43	chr3:4,055,587-4,119,404	p26.2	63818	100.0	-0.82	9	1	*SUMF1*	0	0
41	chr10:33,238,302-33,240,452	p11.22	2151	0.0	-0.82	4	1	*ITGB1*	1	0

Table 10. List of homozygous losses covering genes based on array CGH criteria.

Case no.	Chromosome Region	Cytoband	Length (bp)	# Probes	# Genes	Gene Symbols
22, 34, 37, 43	chr1:150,821,455-150,840,050	q21.3	18596	5	1	*LCE3C*
23, 36	chr1:150,821,455-150,855,012	q21.3	33558	11	2	*LCE3B, LCE3C*
25	chr1:150,825,032-150,840,050	q21.3	15019	4	1	*LCE3C*
31	chr1:167,488,806-167,514,328	q24.2	25523	3	1	*NME7*
22, 25, 43	chr14:105,605,656-105,641,495	q32.33	35840	5	1	abParts
41	chr20:1,513,132-1,543,839	p13	30708	4	1	*SIRPB1*
22	chr20:60,074,157-60,074,190	q13.33	34	5	1	*TAF4*
37, 43	chr3:163,990,112-164,107,775	q26.1	117664	16	1	*BC073807*
23, 41	chr4:69,053,485-69,098,352	q13.2	44868	7	1	*UGT2B17*
23	chr4:69,114,915-69,116,487	q13.2	1573	5	1	*UGT2B17*
41	chr4:69,114,915-69,170,973	q13.2	56059	15	1	*UGT2B17*
34, 37	chr5:180,309,732-180,361,100	q35.3	51369	8	3	*BTNL3, BTNL8, LOC646227*
18	chr6:29,959,049-30,015,539	p21.33	56491	8	8	*AK097625, BC035647, HCG2P7, HCG4B, HCG4P6, HLA-A, HLA-A*0226, HLA-H*
22, 41	chr6:32,561,016-32,593,658	p21.32	32643	6	1	*HLA-DRB5*
23	chr6:32,597,755-32,603,926	p21.32	6172	3	2	*HLA-DRB1, HLA-DRB5*[a]
18, 23, 31	chr8:39,349,320-39,510,878	p11.23-p11.22	161559	22	3	*ADAM3A, ADAM5P, tMDC*
13	chr8:97,222,184-97,226,750	q22.1	4567	8	1	*GDF6*[b]

[a]qPCR analysis showed gain in reference samples or [b]amplification in test samples.

Another possible explanation could be that there was MCC in the sample. This was tested for with genotyping analysis, which showed no detectable MCC (data not shown), making this explanation more unlikely although still possible. A third possible explanation could be that this was a polymorphic area that has not been reported in DGV. It should, however, be noted that the qPCR SE value for the parents were unusually high (Table 11) so it is not yet proven that they are heterozygous carriers. Another array CGH analysis was also performed on case 13 and parents with a custom Agilent array with increased density of probes in this area. The losses were not verified with this array analysis, but there was high background and skewed distribution of signals with that array. Further studies are needed to verify pathogenicity of this loss in the fetus.

An other apparent homozygous loss of interest was a 25.5 kb loss that overlapped part of the *NME7* gene (case 31), which codes for nucleoside-diphosphate kinase 7 isoform a (Fig. 6C). Mice homozygous for a mutation exhibit hydrocephaly, domed skulls and 50% exhibit situs inversus. However, this loss only covers an intron, and seems to be a common variant in DGV, which makes it a less likely candidate.

There were a few apparent homozygous losses that overlapped some HLA genes. A 6.1 kb homozygous loss in chr6p21.32 in case 23 overlapped part of both the *HLA-DRB1* and *HLA-DRB5* genes (Fig. 7A, Table 12A). The *HLA-DRB1* and *HLA-DRB5* genes code for major histocompatibility complex (MHC), class II, DR beta 1 and beta 5. MHC class I and II molecules play a central role in the immune system by presenting peptides derived from the endoplasmic reticulum lumen (Class I) and from extracellular proteins (Class II). However, qPCR analysis showed that this was an apparantly paternally inherited heterozygous loss when compared with a Promega female reference sample. The analysis also showed that the Promega male sample used in the array CGH analysis had a gain compared to the female reference sample, explaining the apparent homozygous loss in case 23.

A complex 32.6 kb loss in cases 18, 22 and 41, overlapping part of the *HLA-DRB5* gene (Fig. 7B, Table 12B) was also of interest. This CNV was considered to be part of a complex CNV, i.e. consisting of more than one overlapping CNVs. There was a homozygous loss adjacent to the *HLA-DRB5* gene overlapping it only with one probe and a heterozygous loss next to it overlapping the *HLA-DRB1* and *HLA-DRB5* genes. This homozygous loss was *de novo* in case 18, and either maternally or paternally inherited in cases 22 and 41, as both parents carried a homozygous loss (Table 12B). There was also a 56.5 kb homozygous loss in chr6p21.33 in case 18 overlapping e.g. the MHC class I *HLA-A* pseudogene, which seemed to be inherited from both parents, who where heterozygous carriers according to qPCR analysis (Fig. 7C, Table 12C). All of these variants overlap common variants in DGV.

4.3.6 Gene Ontology and network analysis

Enrichment analysis was performed in the Nexus Copy Number software to identify GO terms that are significantly overrepresented in all SAs collectively to see if there could be any common pathways affected among the SAs. Analysis on all unfiltered CNV (see Materials and Methods) and all gains showed significant enrichment when corrected for multiple testing of six GO terms associated with biological process, including homophilic cell adhesion, brain development and immune response (Table 13).

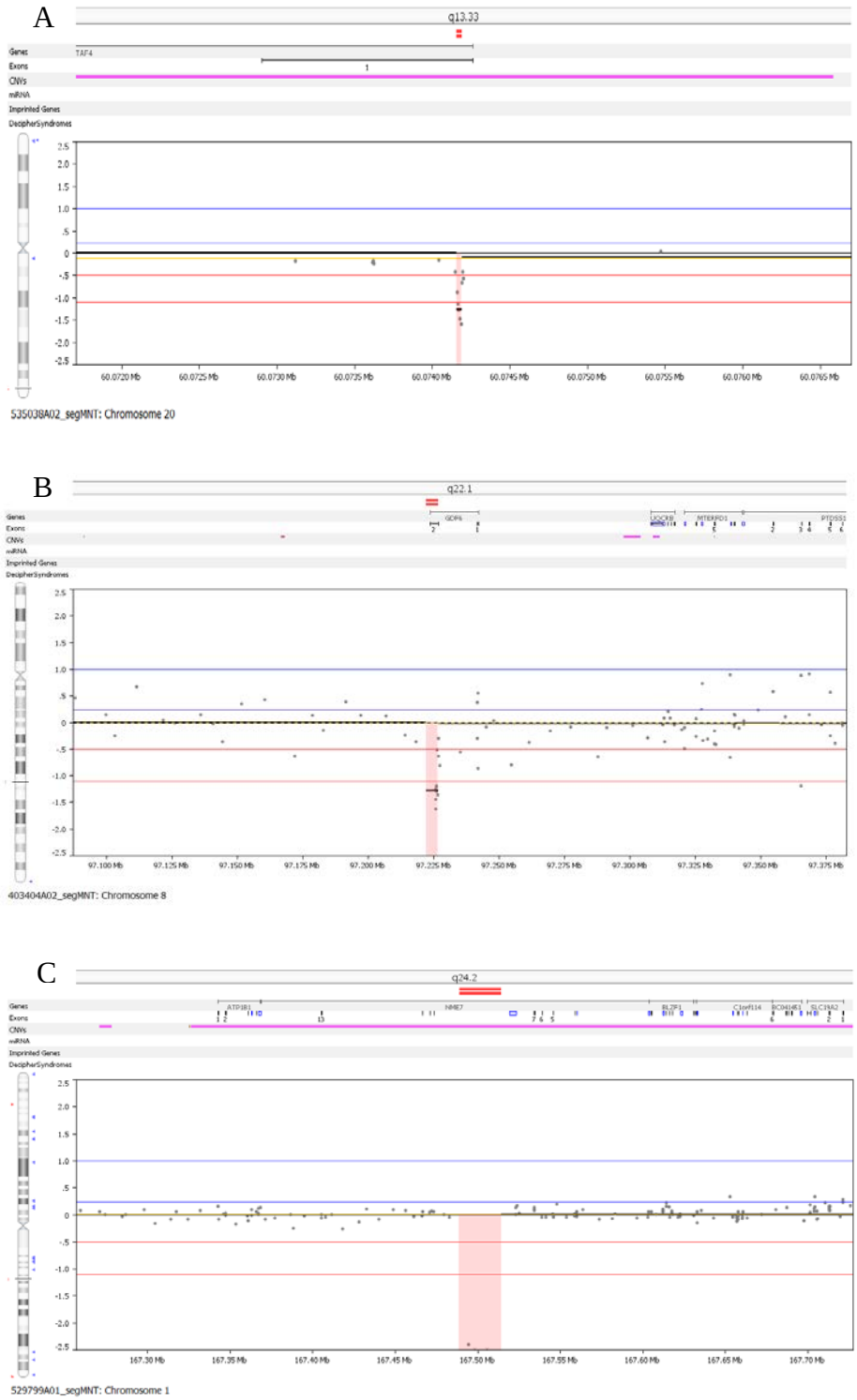


Figure 6. Graphical view of three homozygous losses. A. 34 bp homozygous loss overlapping one exon in the *TAF4* gene (case 22). B. 4.6 kb homozygous loss overlapping one exon in the *GDF6* gene (case 13). C. 25.5 kb loss that overlaps part of the *NME7* gene (case 31). Red highlighted areas indicate loss. Bars above show chromosomal position, and overlap of genes, exons, CNVs in DGV, miRNA, imprinted genes, and DECIPHER syndromes.

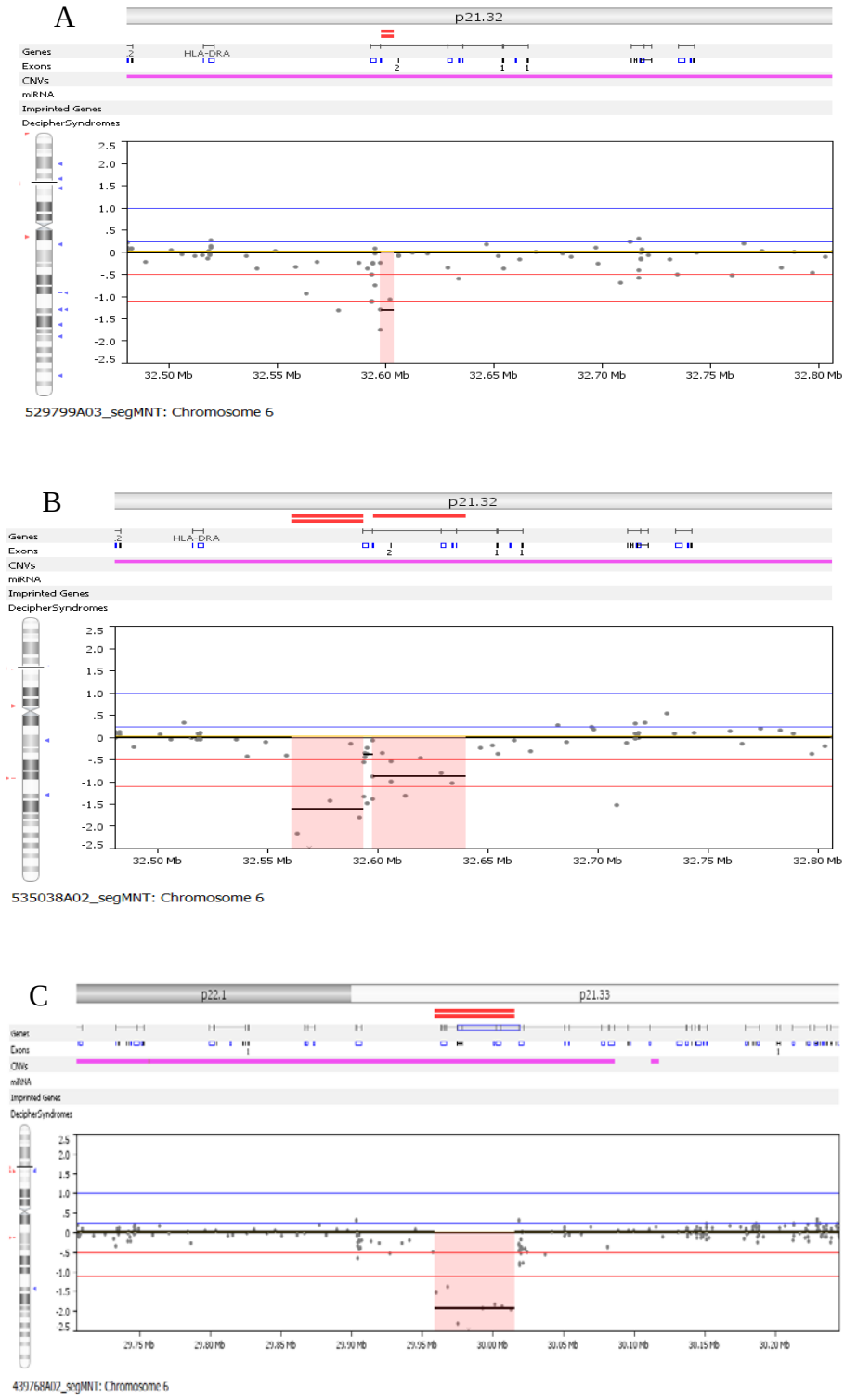


Figure 7. Graphical view of three homozygous losses overlapping HLA-genes. A. Apparent 6.1 kb homozygous loss overlapping the *HLA-DRB1* and *HLA-DRB5* genes in chr6p21.32 (case 23). B. 32.6 kb homozygous loss overlapping part of the *HLA-DRB5* gene in chr6p21.32 (case 22). C. 56.5 kb homozygous loss overlapping the *HLA-A* pseudogene in chr6p21.33 (case 18). Red highlighted areas indicate loss. Bars above show chromosomal position, and overlap of genes, exons, CNVs in DGV, miRNA, imprinted genes, and DECIPHER syndromes.

Table 11. Analysis of an apparent 4.6 kb homozygous loss overlapping the *GDF6* gene in chr8q22.1 with TaqMan Copy Number Assay.

Case no.	Gender	Relative quanitification (RQ)[a]	ΔCт SE[a]	Relative quanitification (RQ)[b]	ΔCт SE[b]
13	Female	0.15	0.164	0.13	0.058
Mother	Female	0.32	0.446	0.27	0.455
Father	Male	0.32	0.224	0.59	0.275
Calibrator	Male	1.00	0.097	1.00	0.241
Reference	Female	1.17	0.152	0.58	0.030

[a]First assay (Hs01657484_cn). [b]Second assay (Hs02148305_cn).
Abbreviations: RQ = Relative quantification, Cт = Cycle threshold, SE = Standard error

Table 12A. Analysis a 6.1 kb homozygous loss overlapping the *HLA-DRB1* and *HLA-DRB5* genes in chr6p21.31 with TaqMan Copy Number Assay.

Case no.	Gender	Relative quantification (RQ)	ΔCт SE
23	Male	0.67	0.041
Mother	Female	1.37	0.067
Father	Male	0.37	0.034
Calibrator	Female	1	0.081
Reference	Male	1.54	0.076

Abbreviations: RQ = Relative quantification, Cт = Cycle threshold, SE = Standard error.

Table 12B. Verification of a 32.6 kb homozygous loss in an area adjacent to the *HLA-DRB5* gene in chr6p21.32 with TaqMan Copy Number Assay.

Case no.	Gender	Relative quanitification (RQ)	ΔCт SE
18	Female	No amplification	
Mother	Female	4.15	0.194
Father	Male	4.82	0.067
22	Female	No amplification	
Mother	Female	No amplification	
Father	Male	0.02	1.169
41	Male	No amplification	
Mother	Female	0.06	0.468
Father	Male	0.03	1.130
Reference	Female	1.95	0.129
Calibrator	Male	1	0.031

Abbreviations: RQ = Relative quantification, Cт = Cycle threshold, SE = Standard error.

Table 12C. Verification of a 56.5 kb homozygous loss overlapping the *HLA-A* pseudogene in chr6p21.33 with TaqMan Copy Number Assay.

Case no.	Gender	Relative quantification (RQ)	ΔCт SE
18	Female	0.001	0.215
Mother	Female	0.63	0.019
Father	Male	0.73	0.130
Reference	Female	0.92	0.079
Calibrator	Male	1	0.030

Abbreviations: RQ = Relative quantification, Cт = Cycle threshold, SE = Standard error.

For losses there was only one significant GO term associated with biological process, i.e. homophilic cell adhesion, although the GO term for nervous system development was borderline with a Q-bound value of 0.0517 (Table 13). When the data was filtered to include only rare CNVs with <50% overlap with DGV (including gains and losses) there was an enrichment for eight GO terms associated with biological process, including regulation of transcription, blood coagulation, brain development and regulation of cell cycle (Table 13). For rare losses there was only one significant GO term, i.e. oligopeptide transport (Table 13).

Network analysis was performed in the String database to see if there were indications of synergistic effects that could be causative of SAs. Analysis on all losses collectively in all fetuses further showed a protein-protein interaction between genes involved in sensory perception and immune response, with direct protein binding only between a number of olfactory receptors and chemokine ligands (*CCL3* and *CCL4*) (Fig. 8A). Network analysis on all CNVs separately in each case revealed some interesting interactions. For example there was evidence of protein binding between gains overlapping parts of seven genes (*GLI2*, *ZIC2*, *MECP2*, *YY1*, *BMI1* and *JUND*) more or less involved in embryonic development (Fig. 8B) in case 41. The genes include *GLI2*, which is thought to play a role during embryogenesis, and mutations in it are associated with holoprosencephaly-9. Mutations in *ZIC2* are associated with holoprosencephaly-5. Mutations in *SKI* are associated with Shprintzen-Goldberg syndrome. The *MECP2* gene is essential for embryonic development and mutations in it are associated with Rett syndrome. The *YY1* gene codes for a ubiquitously distributed transcription factor belonging to the GLI-Kruppel class of zinc finger proteins and it has fundamental roles in embryogenesis, differentiation, replication, and cellular proliferation. The *BMI1* codes for a BMI1 polycomb ring finger oncogene. It is a component of a Polycomb group (PcG) multiprotein PRC1- like complex, a complex class required to maintain the transcriptionally repressive state of many genes, including Hox genes, throughout development. The *JUND* gene codes for a jun D proto-oncogene and it has been proposed to protect cells from p53-dependent senescence and apoptosis. Out of heterozygous losses, it was of interest that two genes (*EN1* and *PTF1A*) in case 18 were associated with the same GO term, i.e. hindbrain development, and homozygous mutations in both genes were neonatal lethal in mice (Table 9). The *EN1* gene codes for engrailed homeobox 1 protein, and it has been implicated in control of pattern formation during development of the central nervous system. The *PTF1A* gene codes for pancreas specific transcription factor 1a, and it is part of a complex known to have a role in pancreatic development.

4.3.7 Meta-analysis on previous studies on SAs

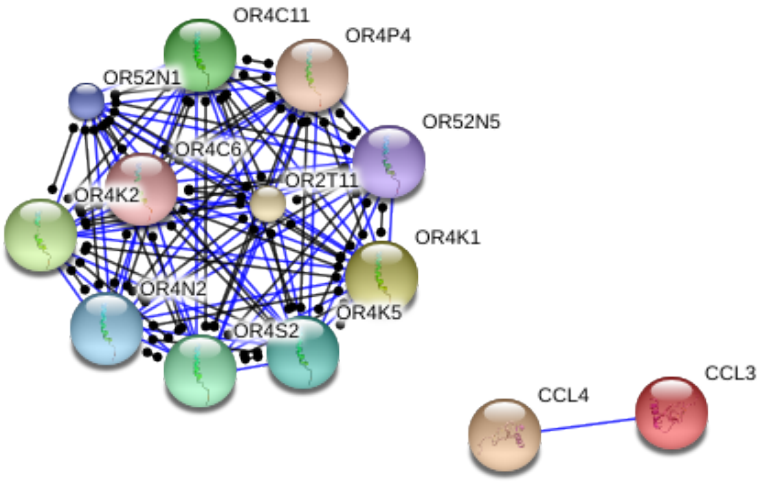
A meta-analysis was performed on 12 studies on SAs in order to get an overview of the frequency of CNVs identified in previous studies on SAs with normal karyotype. Results showed that the proportion of SAs reported with submicroscopic CNVs and normal karyotype was 18.8% (67 in 356) (Table 14). The proportion of SAs with CNVs is higher than was calculated in a recent review by Van de Berg *et al.* (2011) (2), where 5% of SAs where reported with submicroscopic CNVs. The study material in Van de Berg´s review was, however, quite variable and, included not only samples with normal karyotype. It is, however, difficult to compare those studies, as they differ in so many respects, e.g. platforms used, study material and indications for the study.

Table 13. Significant GO terms associated with biological function.

Dataset	Term	P-Value	MP P-Value	Q-Bound	Present	Total
All CNVs	Homophilic cell adhesion	2.77E 08	0.0038	2.307E 12	23	140
	Defense response to bacterium	9.19E 08	1.41E 10	0.0038	17	99
	Brain development	2.88E 09	3.55E 01	0.0073	21	155
	Regulation of immune response	3.51E 10	1.71E 09	0.0073	14	76
	Calcium-dependent cell-cell adhesion	1.14E 11	0.0069	0.0190	8	27
	Immune response	2.30E 10	2.39E 06	0.0320	31	322
All losses	Homophilic cell adhesion	8.72E 05	0.0035	7.26E 10	11	140
	Nervous system development*	1.24E 11	0.0010	0.0517	10	296
Rare CNVs	Negative regulation of transcription; DNA-dependent	9.90E 03	1.56E 02	8.25E 07	23	419
	Transcription; DNA-dependent	5.07E 07	1.52E 06	2.11E 12	41	1767
	Positive regulation of transcription; DNA-dependent	1.37E 08	2.68E 02	3.81E 11	19	483
	Negative regulation of transcription from RNA polymerase II promoter	3.39E 10	3.57E 05	0.0071	16	437
	Blood coagulation	6.65E 09	6.66E 08	0.0111	16	461
	Positive regulation of transcription from RNA polymerase II promoter	1.22E 11	1.21E 06	0.0169	18	597
	Brain development	1.42E 11	3.06E 07	0.0169	9	155
	Regulation of cell cycle	2.96E 11	1.52E 10	0.0308	6	65
Rare losses	Oligopeptide transport	3.13E 10	4.10E 12	0.0261	2	5

Three p-value measures were provided, i.e. a standard p-value, Markov Process (MP) p-value, and a false discovery rate (FDR)-corrected (for multiple testing) p-value (Q-bound). Only Q-bound values <0.05 were considered as significant. The numbers in the Present and Total Column represent the number of genes corresponding to each GO term present in this dataset (Present) and in the whole genome (Total). *Borderline significance.

A



B

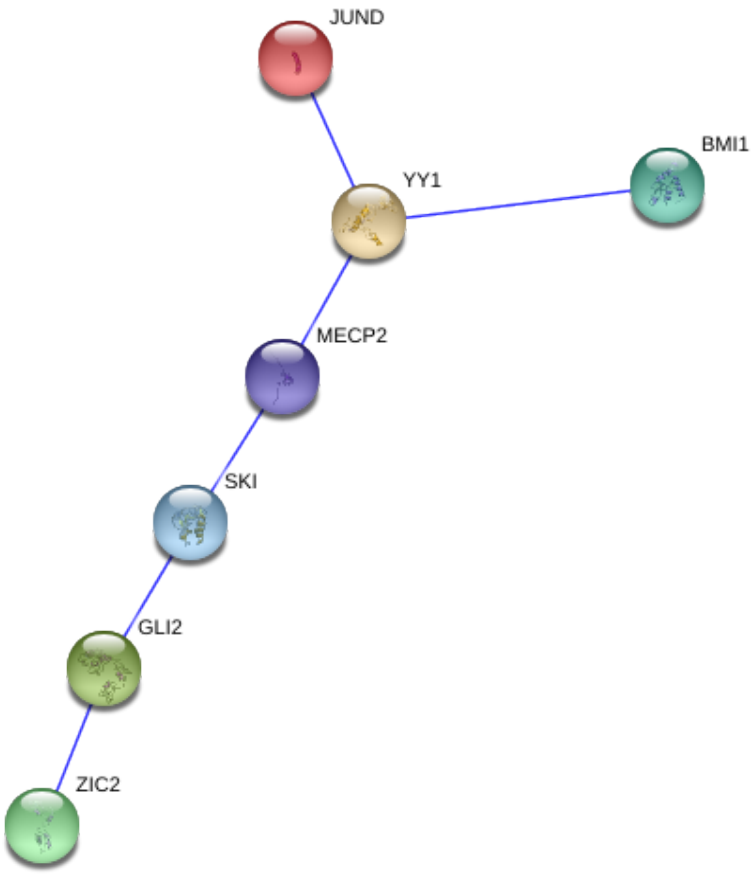


Figure 8. Predicted protein-protein interaction by String database. A. All losses in all SAs combined. B. Gains overlapping genes involved in embryonic development (case 41). Different line colors represent different types of protein associations. Blue lines indicate protein binding and black lines indicate reaction.

Table 14. Reported submicroscopic findings identified with array CGH in SAs with euploid (normal) karyotype based on conventional karyotyping.

Author	Material	Total # SAs n	Array platform	Submicroscopic CNVs with normal karyotype n (%)	CNV size
Schaeffer *et al.* (2004) (11)	Unselected, CVS, ≤ 20w	41	Genosensor Array 300 (287 clones) WG	1/25 (4.0)	2 clones
Benkhalifa *et al.*(2005) (12)	Culture failure, Fetal samples, 9-11w	26	1 Mb BAC WG Array (2600 BAC/PAC clones)	3/13 (23.1)[a]	1-2 clones
Shimokawa *et al.* (2006) (13)	Normal karyotype, CVS, 5-12w	20	2173 WG BAC clone array	1/19 (5)	1.4 Mb
Zhang *et al.*(2009) (16)	Normal karyotype or culture failure, Spontaneous/theraputic abortion, CVS, first-trimester	58	244K WG Agilent array	5/50 (10.0) Unique	108-1460 kb
Menten *et al.*(2009) (14)	Normal karyotype or culture failure, CVS, fetal skin, umbilical cord, first-og second trimester (mors in utero)	100	1 Mb WG BAC array	3/77 (3.9)	1-7 bands
Robberecht *et al.* (2009) (8)	Unselected, placental villi, membranes or fetal tissue	103	1 Mb WG BAC array (3534 clones)	1/64 (1.6)	787.5 kb
Warren *et al.* (2009) (15)	Normal karyotype (n=9) or no conventional cytogenetic testing (n=26), fetal tissue	35	1 Mb WG BAC array (2600 clones) 244K WG Agilent array	4/27 (15) BAC array 6/6 (100) unique - 244K array	93-289 kb -BAC array 20-1310 kb -Agilent
Rajcan-Separovic *et al.* (2010) (19)	Euploid karyotype and structural abnormalities, CVS, CRL 2-30 mm	14	1 Mb WG BAC array, 105K WG Agilent array, Custom Agilent array	5/14 (35.7)[b] Unique	12.9-214 kb 1.6 kb Custom array
Rajcan-Separovic *et al.* (2010) (18)	Normal karyotype, fetal tissue, <20w	23	105K WG Agilent array, Custom Agilent array – selected regions	13/27 (48.1)[c] Unique	27-1593 kb
Deshpande *et al.* (2010) (17)	Unselected, CVS	20	Focus WG Cytochip (BlueGnome) 1 Mb res in backbone, 100 kb res in selected regions	0	
Lathi *et al.* (2012) (20)	Unselected, CVS, 7-12w	30	Illumina CytoSNP-12 (1 SNP per 10kb)	0	
Viaggi *et al.* (2013) (21)	Normal karyotype, CVS, 7-11w	40	8x60K WG Agilent	31/40 (77.5)[d] 13/40 (32.5)[d] unique	120 kb-4.3 Mb
Total #		510		67	
Proportion				67/356 (18.8)	

[a] Only samples from culture failure. Number of SA with normal karyotype is estimated. [b]6 CNVs in 5 SAs with euploid karyotype. [c] 11 CNVs in 13 SAs from 8 couples. [d]45 CNV in 40 SAs. 45 CNVs in 31 SAs with euploid karyotype, including 14/45 unique CNVs in 13 SAs. Abbreviations: CVS=Chorionic villus sampling, CRL=Crown rump length. WG = Whole genome.

5 Discussion

Overall, we found a significantly larger amount of CNVs in SAs with normal karyotype than previous studies (8, 11-21). The possible causes for this difference are discussed in section 5.2.1. Majority of the CNVs identified were small and did not yet have a clear established genotype-phenotype correlation, even though there were some possible candidates. Many CNVs overlapped OMIM genes (66.6%), OMIM morbid genes (16.0%) and some partly overlapped DECIPHER syndromes (6.8%). Overall, however, there was not enough evidence of pathogenicity of the CNVs identified to use the data in genetic counseling for the couple involved. Difficulties with biological and technical interpretation of the high number of CNVs detected as well as technical challanges and additional cost with verification tests of possible candidates further made the analysis with this array difficult. Nevertheless, this array provides an opportunity to identify exonic variants and possible candidate genes for SAs. We did not find any known causes for SAs with this array. We found two possible candidate genes as new causes for SAs. The variants were an apparent 34 bp exonic homozygous loss overlapping the *TAF4* gene and an apparent 4.6 kb homozygous loss overlapping the *GDF6* gene.

5.1 Conventional karyotypic analysis

The proportion of abnormal karyotypes based on conventional karyotypic analysis was somewhat higher (over 60% of successfully karyotyped samples) than has been reported in other studies, where abnormal karyotypes are identified in approximately 40-50% of cases (2). The number of failed karyotypic analysis was 6 of 43 (or 14.0%), which is similar or lower to other studies (2). This, however, lead to fewer available samples for our array CGH study (i.e. with normal karyotype) or only 14 of 43 or 36.8% of all samples in the study. The composition of abnormalities consisted of many unusual karyotypes (9 of 23 or 39.1% of all abnormalities) such as mosaicism, double aneuplodies, etc. However, these results are probably coincidental since the study group is too small to draw conclusions from this or the proportion of SAs with abnormal karyotype. The male-to-female ratio of normal karyotypes was within the expected range, 0.75 (or 6 to 8) of successfully karyotyped samples, suggesting that MCC was not a major problem in our study, although it was suspected and confirmed in three cases (7.0%) in total.

5.2 Array CGH analysis

5.2.1 CNV findings and genomic distribution

The number of CNVs per individual (average 133) identified in this study was significantly higher than in other studies on SAs where submicroscopic CNVs were reported in 18.8% of cases overall (Table 14). There are three main possible causes for this difference. First, higher resolution of our array compared to arrays used in other studies on SAs and different analytical tools and reporting criteria used. Second, possibly higher frequency of false positives in our study, and third, our study population was somehow different from others.

First, our array had the highest resolution and was the only exon-focused array. As expected we identified a high number of small CNVs with the highest frequency of CNVs in the 1-10 kb size

range (55.1%), and the size distribution was similar to overall CNVs in DGV. For comparison, there was only one CNV < 20 kb reported among the other studies on the SAs reviewed (19). To try to see if our data would be more comparable to the other studies we put a size threshold of 10 and 20 kb to our data and found submicroscopic CNVs in 12 of 13 cases (92.3%) and an average of 57 and 46 CNVs per sample respectively. Most of the studies that used oligonucleotide arrays, however, reported rare or unique CNVs specifically in 10-100% of cases (15, 16, 18, 19, 21). When our data was analysed with a 10 or 20 kb size threshold we found rare CNVs (with <50% overlap with DGV) in 12 (92.3%) cases which is comparable to the other studies.

Second, it is possible that our data had a larger number of false positives then the other studies on SAs. We found the number (133) and average total size of CNVs (6.04 Mb) to be relatively high, although it is within the range that has been reported by other studies. For example, 12-724 CNVs covering ~0.1-15 Mb were identified in a single healthy individual in a recent paper using five different high-resolution microarrays (62). It is noteworthy that there was a higher ratio of gains vs. losses in our study than expected (or 6.4), which could indicate that there are some false positive gains in our data. The ratio of gains and losses identified in each study varies with the methodology used, but overall ratio seems to be equal or close to one according to Pang´s *et al.* study, where they combined whole genome sequencing data with results from five high resolution microarrays (62). The different ratio in our study can be explained by a higher stringency of the algorithm for calling of losses. The quality of our data and likelihood of false positives is further discussed in section 5.4.2.

Third, our study population could be different from others. As an example we used material from women with recurrent abortions instead of sporadic abortions whereas only one other study specified that the study group was women with recurrent abortions (18). This would be expected to diminish the likelihood of finding causative CNVs as most pathogenic CNVs are *de novo*. We wanted to test the applicability of array CGH analysis in RAs as genetic testing and counseling is only routinely performed in couples with RAs and not sporadic abortions.

The proportion of CNVs that overlapped genes was unusually high (1598 of 1723 or 92.7% of all CNVs overlapped genes) in our study, whereas only 52.3% of CNVs overlap genes in DGV (12/8/13). This could be explained by the design of our array, as there is a significantly increased coverage of probes in exons (median probe spacing 65 bp in exons. vs. 7291 in the genomic backbone).

There was was less overlap between homozygous losses with genes (28 of 56 or 50.0%) and DECIPHER syndromes (0.0%) than other CNV types. Such a bias has been assumed to reflect a greater purifying selection acting on complete losses of genes when seen in healthy adults (102). The fact that most of the homozygous losses were common variants with ≥50% overlap with DGV, may suggest that they are usually not harmful, although CNVs among healthy individuals in the hemizygous state are sometimes pathogenic in a homozygous state. Some homozygous losses might also be lethal in the embryonic state, and inherited from heterozygous carrier parents.

The distribution of CNVs throughout the genome was somewhat uneven, although they mapped to all of the chromosomes. It is noteworthy that a relatively high number of CNVs, in particular gains, mapped to chromosome 19, which is a very GC-rich chromosome. It is a known limitation of

array CGH data that it may sometimes show some “waviness” that shifts with GC content in the human genome and affecting clinical specificity and sensitivity of array platforms (103). This is often more apparent for GC-rich chromosome such as chromosome 19. Our data was corrected for this waviness, but the fact that there was a relatively high number of CNVs on this chromosome suggests that there may some false positives. There was, however, not a significantly higher relative number of CNVs on other GC-rich chromosomes, such as chromosomes 15-17, 20 and 22.

5.2.2 Size distribution of CNVs

Size distribution of CNVs identified in this study shows that the number and size of CNVs are negatively correlated (Fig. 4). However, although the majority of CNVs were small, we also found some excess of large CNVs. For example we found CNVs > 500 kb in 8 of 13 (61.5%) and a CNV > 1 Mb in 2 of 13 samples (15.4%). This is a considerably higher proportion than in reports on both healthy individuals and SAs, e.g. 5-10% of individuals had a CNV > 500 kb and 1-2% a CNV > 1 Mb in a study by Itsara *et al.* (2009) (23). However, in a study by Viaggi *et al.* (2013) on 40 SAs with normal karyotype, they found a CNV > 500 kb in 19 cases (47.5%) and a CNV > 1 Mb in 15 cases (37.5%) (21). The clinical significance of the CNVs in Viaggi´s study was, however, considered to be benign or unclear. The difference in proportion of individuals with large CNVs could be explained by the fact that our study group is very small. Furthermore, these large CNVs all overlapped variants in DGV and the gene content in those CNVs did not suggest that they were likely to cause SAs.

5.2.3 Rare, large CNVs in SAs

Ten samples (71.4%) carried large (>50 kb), rare (<50% overlap with DGV) CNVs (Table 6). They might be classified as possible candidate genes for causing SAs based on size (range ~76-399 kb) and absence in healthy individuals. There was, however, an absence of rare CNVs larger than 400 kb as is for example seen in the majority of microdeletion/microduplication syndromes (24) in our study as well as in majority of the studies reviewed on SAs. Also, the average number of genes (including microRNAs) per CNV was low or equal to three (range 1-6). Furthermore, the function of the genes within the CNVs was not known to be associated with disease or defective embryonic development, which makes them less likely to be causative/contributive for the abortions. A few of the CNVs identified in this category are further discussed below.

<u>242 kb loss in chr10p12.1</u>

This paternally inherited loss, found in one sample, overlapped the *KIAA1217* gene, which is required for normal development of intervertebral disks. Mice homozygous for a gene-trapped allele display malformations of the notochord and caudal vertebrae and may exhibit caudal tail kinks. However, as the loss is inherited from a healthy father it is not likely to be causative for the abortion, unless it unmasks a mutation on the other chromosome, is imprinted or has synergistic effects.

<u>289 kb gain in chrXq22.1-q22.2</u>

This maternally inherited gain found in three out of four karyotypic normal samples (two female and one male) from the same couple overlapped the Pelizaeus-Merzbacher (PMD) disease

duplication/deletion area. PMD is an X-linked recessive neurodevelopmental disorder, which has a range of phenotypes of different severity. This disease is in a majority of cases caused by a gain of the *PLP1* gene, which was not duplicated in our case, and this greatly diminishes the likelihood that the fetuses were affected by this disease. In exceptional cases, however, a PMD-like disorder has been linked to mutations in other genes or position effect rearrangements in *PLP1* neighboring genes (104). Most females are unaffected carriers, although there are some cases of affected females, that have shown lack of skewed X-chromosome inactivation (105). It might also be possible that a gain of PMD area would affect females in such a way that they are at increased risk of having RAs. Out of the six genes within this CNV (i.e. *NGFRAP1*, *RAB40A*, *TCEAL3*, *TCEAL4*, *TCEAL7* and *WBP5*), *NGFRAP1* (which codes for nerve growth factor receptor (TNFRSF16)) is of special interest based on gene function. According to the UCSC Genome Browser (http://www.genome.ucsc.edu/) it may be a signaling adapter molecule involved in p75NTR- mediated apoptosis induced by NGF. Also it plays a role in zinc-triggered neuronal death (by similarity) and may play an important role in the pathogenesis of neurogenetic diseases. Taking this information into account it is possible, although unlikely, that this gain contributed to the abortions.

The same couple also had two SAs with a 365 kb gain in chr3q13.2, which overlapped four genes, including the *ATG3* gene, which encodes an ubiquitin-like-conjugating enzyme, which is known to play a role in autophagy during cell death. However, there is not enough evidence to assume that it is likely to be causative for the abortions, although there are some reports that suggest that combined effect of two or more CNVs may lead to a more serious phenotype (72).

5.2.4 Heterozygous and homozygous losses in SAs and possible candidate genes

As discussed before our results indicate that some of our cases might be heterozygous carriers for a lethal condition according to mice phenotypes in MGI, but there is not enough evidence to assume that they would be a likely cause for the abortions.

Some homozygous losses were of interest as discussed before, in particular exonic losses of *TAF4* and *GDF6*, but further work would be needed both to verify if the loss in the *TAF4* gene is a truly homozygous loss and to prove the pathogenicity of both losses in SAs in larger cohorts (see further discussion in section 5.5). Results on the *GDF6* gene could indicate that the fetus (case 13) may have suffered from the autosomal dominant Klippel-Feil syndrome and microphthalmia or that the loss overlaps a polymorphic area that is not dosage sensitive.

The homozygous losses overlapping the MHC class I and II genes, *HLA-A*-pseudogene (class I), and *HLA-DRB1* and *HLA-DRB5* (class II) were also of special interest as these genes have been implemented with SAs before (35-45). However, the homozygous losses identified in our study were less likely to be pathogenic as they resulted in being either a heterozygous loss for the *HLA-DRB1* and *HLA-DRB5* genes (case 23), overlapping *HLA-DRB5* with only one probe (case 18, 22 and 41) and being identified as homozygous in parents, or overlapping the *HLA-A*-pseudogene in an area upstream of *HLA-A* (case 18). A null-haplotype for *HLA-DRB1* or *HLA-DRB5* has never been identified in screening of approximately 5000 individuals and in only one individual for the *HLA-A* gene (written

communication with Alfred Arnason and Kristjana Bjarnadottir, Blood Bank in Iceland). According to publications in DGV the frequency of loss covering the *HLA-A* gene is around 30% and *HLA-DRB5* around 16.7%. As the null-haplotype for both of these genes is much rarer than would have been expected by chance, this suggests that complete loss would be lethal. According to the Hardy-Weinberg principle (where p=normal allele, 2pq= heterozygous allele, q=homozygous allele) $p+q = (p+q)^2 = p^2 + 2pq + q^2 = (0.83)^2 + 0.3 + (0.17)^2 = 1$, so the allele frequency of the homozygous loss (q) for the *HLA-A* gene would be equal to $(0.17)^2 = 0.0289$ or 3%. The frequency of homozygous loss or null-allele is much rarer or 0.02% according to information from Alfred Arnason and Kristjana Bjarnadottir, which suggests that this genotype may often be lethal. Several studies have shown that HLA sharing (including *HLA-A* and *HLA-DR* antigens) among couples are associated with RAs (35-45). HLA typing of individuals does not distinguish between homozygous genotypes and heterozygous losses. Therefore we suggest that the couples that showed shared HLA antigens could in some instances both be heterozygous carriers of loss of the same gene. They would then have a 25% risk of producing offspring with a homozygous loss, which would be lethal. This would also explain the lack of antifetal antibody in the mother´s blood in previous studies as the paternal allele would be missing (40, 41, 44, 48, 49)

5.2.5 Gene Ontology and network analysis

The biological processes that were enriched overall in our samples were consistent with previous findings involving common CNVs (64, 106), but it is of interest that in the filtered data with rare CNV there was an enrichment of GO terms involving regulation of transcription and cell cycle, brain development and blood coagulation (gains), and oligonucleotide transport (losses). It is, however, not uncommon that GO terms involving brain development get a significant score when searching in the whole genome as genes that have a role in brain development tend to be larger on average (107). It can also be difficult to predict both the effects CNVs have on protein function (in particular for gains) as well as the phenotypic effects. Most gains affecting parts of genes result in loss of function and nonsense-mediated decay, although there are examples of gain of function of a protein. It is, however, not possible to predict the implications the combination of a number of gains in case 41 or losses in case 18.

5.3 Limitations of the study

5.3.1 Study material

There were some limitations of the study regarding the study material. First of all, a low number of samples with a normal karyotype according to conventional karyotypic analysis was a limitation to our study. The reason for this is that SAs with a normal karyotype are in many cases not caused by CNVs. Second, we are selecting against possible findings by choosing material from couples with RAs, as we would expect most pathogenic CNVs to be *de novo*. An exception would be e.g. if there was a loss of an imprinted gene. Also, by choosing material where conventional karyotypic analysis had already been performed, we would not be detecting the most common known genetic cause of RAs, i.e. unbalanced chromosomal rearrangements. Third, we were looking at a restricted time period of

gestation (~5-20 weeks), which may limit findings as discussed before. Fourth, our samples were derived from post-mortem tissue (except for case 13), although the cells were viable in most cases. As a consequence it is possible that some DNA damage may have occurred in the samples affecting the quality of the array CGH analysis. There is limited data available on this subject, but it may be interesting to analyse this in the future. Previous studies on SAs and stillbirths have, however, shown advantages of using array CGH over conventional karyotyping, primarily due to success with nonviable tissue (2, 26). Finally we did not have a suitable control group, which complicated interpretation. It would also have been of advantage if we would have had access to data from normal individuals analysed with the 720K exon-focused array. We were, however, only able to find two publications with this array, one on congenital glaucoma in Korean patients (108) and the other one on and epileptic encephalopathies (100). In Lee´s *et al.* (2011) (108) study the study group was Korean patients, which was not an optimal control group as it may differ from the Icelandic population. The investigators used the same software (Nexus Copy Number) and algorithm (FASST) for analysis, but with a slightly more stringent threshold for gains (log2 ratio 0.3) and they excluded CNVs with <5 probes or CNVs ≤500 bp in length. They found a lower amount of variants compared to our study or 156 CNVs in 149 samples, but interestingly they found the ratio of gains to losses to be 10.9, which is comparable to our results. In Mefford´s *et al.* (2009) (100) study on epileptic encephalopathies the investigators used a different software and algorithm for analysis, and they only reported possible candidate variants.

5.3.2 Quality of array CGH data

A limitation of our study is that we only performed verification tests on a few CNVs detected, so we don´t know the proportion of false positive calls. Although, the quality control parameters were met for most of the samples (all but one) (Table 3), we often found it difficult to determine if calls where true or false positive. It was also often difficult to map the breakpoints due to noise in this array. Poor characterization of the reference samples further complicated interpretation. An example was the apparent homozygous loss overlapping part of the *HLA-DRB1* and *HLA-DRB5* genes in case 23, which proved to be a heterozygous loss compared to a reference sample containing a gain.

The number and size of CNVs identified was within the range of what has been reported in other studies (62, 63), although it was difficult to find comparable studies using e.g. the same array. It is also been challenging to determine the expected amount of CNVs per genome, and also to determine the “gold standard” for analysis as there has been limited overlap between studies using the same DNA source. For example, Pinto *et al.* (2011) found that different analytical methods applied to the same raw data typically yielded CNV calls with <50% concordance when comparing CNV detection on 11 high resolution microarrays (63). They also found that reproducibility between replicate experiments was <70% for most platforms (63).

There was a higher ratio of gains vs. losses in our study than expected (or 6.4), where a ratio closer to one is more commonly found. This could partly be explained by the fact that we had a somewhat higher stringency for calling of losses (log2 ratio threshold of -0.5) than for gains (log2 ratio of 0.23), although we used default settings and an average stringency of the algorithm (instead of stringent). We did this as we wanted to get a good sensitivity in identifying true CNV calls while

compromising specificity as little as possible. The fact that there was a high number of gains on chromosome 19, which is a GC-rich chromosome could also suggest that there was some waviness in the data resulting in false positives. Another possible explanation for a bias in ratio of gains and losses identified could be due to dye-bias during the labeling reaction, which has by some laboratories been prevented by using dye-swap experiments.

In general there tends to be lower specificity in detection of small CNVs with array CGH, which might also suggest that considerable proportion of CNVs detected in our analysis could be false positive as majority of the variants in our study were small (79.7% of the CNVs detected were smaller than 50 kb and 58.5% smaller than 10 kb). Also, we did not use any size threshold for reporting of variants as some laboratories do. A size threshold would have been contrary to the idea of using an array designed to detect CNVs at the exon-level.

Another issue is that the quality of probes in exon-focused arrays might not be as good in as other CGH arrays as there is less flexibility for design of probes given their high density. In fact other studies using exon-focused arrays, have found a high proportion of false positive calls. For example one study failed to confirm 193 of 267 CNVs in autism trio cases (72%) using a custom 1M Agilent exon-focused CGH array (109).

5.3.3 Verification of array CGH results

One of the challenges in array CGH studies is the need to verify results with an alternative method and finding a suitable verification test. For large CNVs, FISH is the standard method to use for verification of results, and it has the advantage of identifying the genomic location of CNVs in cases of genomic rearrangements. For smaller CNVs (less than hundreds of kb in size), other methods such as qPCR or long range PCR are needed, which can often involve time consuming and difficult steps of probe selection and optimization. Also, it is more difficult to identify one copy gains than losses with these methods. Long range PCR would generally not work for gains, and the relative difference between the test and reference sample is smaller (3:2 for gains vs. 1:2 for losses). As the standard deviation of results with most qPCR methods is quite high it can be difficult and in some cases impossible to distinguish between two and three copies of DNA using that technique.

The SYBR GreenI qPCR method is an unspecific method of quantifying the difference in copy number as the SYBR GreenI dye binds to all double-stranded products in the sample. As a consequence, unspecific regions, such as primer-dimers are magnified as well as the fragments you are interested in.

The Applied Biosystem´s TaqMan qPCR Assay is a more specific method for copy number quantification as it consists of a TaqMan minor groove binding (MGB) probe labeled with a FAM dye and a quencher which only binds to the genomic sequence of interest. In our experience, the predesigned TaqMan qPCR assay produced more reproducible results and was also less time consuming than the SYBR GreenI analysis as no optimization steps were necessary. So, although the TaqMan assay is more expensive than the SYBR GreenI anlysis, it payed off due to better quality data and quicker turn around time. It is, however, still difficult and costly to confirm all CNVs using a TaqMan assay.

MLPA is an alternative method commonly used to verify array CGH results. A drawback of that method is that it is time consuming and difficult to develop or set up new assays. Now, there is also a new technique on the market, called Digital PCR, which offers a new and more accurate approach to nucleic acid detection and quantification (110).

5.3.4 Interpretation of array CGH data

A great challenge facing clinical laboratory geneticists is classification of CNVs into groups based on clinical significance as pathogenic, benign, or CNVs with unknown clinical significance. As the phenotypic effects of the majority of rare CNVs (and even some common CNVs) are still largely unknown many CNVs are classified into a group of CNVs with unknown clinical significance. This fact poses challenges for genetic counseling and has been one of the main limiting factors for implementation of array CGH in prenatal diagnosis, although much knowledge has been gathered in the last few years. A recent study by Wapner *et al.* (2012) on 4406 prenatal cases showed that CNVs of uncertain clinical significance are identified in 3.4% of cases (1.5% afer re-evaluation) using a 44K Agilent CGH array. This is a reasonably manageable number to deal with for genetic counseling. The fact that this was a relatively low resolution array, however, further decreased the number of CNVs with uncertain clinical significance, since only large CNVs were detected.

There are three main factors that determine the proportion of CNVs classified as of unknown clinical significance, i.e. type of array and algorithms used for analysis, skills of the clinical laboratory geneticists interpreting the data, and different policies and reporting criteria among laboratories.

We discovered a relatively large amount of rare CNVs (<50% overlap with DGV) in our study or 501 CNVs (26.9% of all CNVs identified in the study). Majority of them were small and did not yet have a clear genotype-phenotype correlation established, although the gene function was in many cases associated with embryonic or placental development. It can also often be difficult to predict the affect of maternally inherited CNVs on reproductive future (e.g. in PMD-like duplication carriers). It can be difficult to distinguish if the variant is affecting the fetuses or the mother herself, causing her to abort. Overall, we did not have enough evidence of pathogenicity for any of the CNVs identified to use the data in genetic counseling for the couple involved.

5.4 Array CGH versus conventional karyotyping for analysis of SAs

Conventional karyotypic analysis is the current standard method to identify chromosomal abnormalities in SAs. However, guidelines on the topic of recurrent abortion differ with regard to recommended evaluations. The European Society of Obstetrics and Gynaecology (ESHRE) advises genetic evaluation of SA samples only within the setting of scientific studies, but they advice karyotypic analysis of parents (111). In contrast the Royal College of Obstetricians and Gynaecologists (RCOG) recommends karyotypic analysis on fetal samples (112). The main reasons for applying cytogenetic analysis to fetal remains from SAs is to eliminate further investigation and to provide a better explanation and recurrence risk estimate for the couple.

The main advantages of using array CGH over conventional karyotyping is higher resolution for detection of CNVs, success in cases of tissue culture failure, and maternal cell contamination, and quicker turn around time. These advantages are practical in our material. It has some disadvantages

such as difficulty in detecting balanced translocations and inversions, polyploidy and low-level mosaicism, and higher cost.

One of our cases supports this as we detected a trisomy 21 in one sample (case 15), where karyotype was not possible due to tissue culture failure (1 of 43 or 2.3% of cases), using a NimbleGen 135K CGX array (data not shown). Furthermore, we were able to detect a mosaic trisomy 10 (case 29) on an other array (data not shown).

Also, by using array CGH only, we would expect to find most aneuploidies that were detected by conventional karyotyping in the study, although some would be missed (Table 1). Out of the 23 abnormal karyotypes we would expect to miss at least one (4.3%) abnormality, i.e. a female mosaic polyploidy (case 3). In two cases it was unclear if the abnormalities would be detectable with array CGH, i.e. an inversion on chromosome 12 (case 42) and a case of low-level mosaicism for double aneuploidy (karyotype mos46,XX[45]/45,X[4]/47,XXX[2]) (case 6). Recent data has shown that a large proportion of apparently balanced inversions and translocations according to conventional karyotypic analysis are not truly balanced, and that small losses are detectable at the breakpoints. As the inversion (case 42) was inherited from a healthy mother it was, however, not considered to be the cause of the abortion. The mosaicism (case 6) would probably have been missed by array CGH, although it is not certain as level of mosaicism might differ with the two methods, e.g. due to a selection of cells during culture. In two cases of male polyploidy (cases 24 and 40) (one of which (case 24) had a more complex karyotype, 70,XXY,+16[7]/69,XXY/46,XY[2]), a polyploidy and possibly mosaicism for trisomy 16 might have been suspected. Finally, in six cases of mosaicisms or combined numerical and structural abnormalities (cases 5, 10, 17, 20, 21, and 39), we would expect to detect at least a part of the abnormalities, i.e. aneuploidies. Polyploidies and translocations would be missed if they were truly balanced. Overall, we could expect to miss approximately three chromosomal abnormalities (13.0%) by using array CGH instead of conventional karyotyping, i.e. mos92,XXXX[7]/46,XX[5] in case 3, 46,XX,inv(12)(q15q24) in case 42 and mos46,XX[45]/45,X[4]/47,XXX[2] in case 6. Only one of those (1 of 24 or 4.2% of all abnormal karyotypes) was certain to be the cause of abortion. To summarize, an equal number of lethal karyotypes would be expected to have been missed by conventional karyotyping and our exon-focused array CGH, i.e. at least one abnormality (4.2%) with each method, giving a detection rate of 95.8%. Conventional karyotyping missed a trisomy 21 in case 15, where cells did not grow *in vitro* and array CGH would have missed mos92,XXXX[7]/46,XX[5] in case 3.

As discussed before, the main reasons for applying cytogenetic analysis to fetal remains from SAs is to eliminate further investigation and to provide a better explanation and recurrence risk estimate for the couple. The identification of a causative *de novo* CNV with array CGH in a SA could be of huge relief to couples as it would give them a low recurrence risk (~1%) for future pregnancies. Array CGH can also identify inherited imbalances, e.g. unbalanced chromosomal rearrangements or dup-del syndromes (below the resolution of G-banding), where one of the parents is a carrier for a balanced translocation and transmits one of the rearranged chromosomes to the fetus. This would give the couple a recurrence risk for either having an other SA due to the same condition or

alternative imbalanced karyotypes, which could result in a live birth of a child with serious birth defects. These conditions could be tested for prenatally in future pregnancies.

As discussed before, some of the main limitations of array CGH is inability to detect balanced translocations and inversions, polyploidy and low level mosaicism. Balanced translocations and inversions occur in approximately 0.08-0.09% of prenatal diagnostic samples (25). An inherited balanced rearrangement is not likely to be the cause of SA, but it is relevant to future reproductive counseling, as parents may be carriers and have increased risk of SA. Further investigation could be implemented to quantify the residual risk of a balanced rearrangements in cases of a normal array CGH analysis and to determine when and whether additional genomic analysis is necessary (25).

Triploidy is not detectable with array CGH except in cases of male karyotypes (69,XXY or 69,XYY) as the amount of DNA of test and reference sample is evened out before hybridization. This problem can be overcome by use of a combination of array CGH and microsatellite genotyping (16) or flow cytometry (FCM) (14) for detection of ploidy status. Also, SNP arrays or combined array CGH and SNP arrays can detect both polyploidy and absence of heterozygosity, a useful feature in the analysis of SAs, stillbirths and prenatal diagnosis cases.

Array CGH has been shown to miss some low-level mosaicism cases, although some studies suggest that the sensitivity might not be lower than with G-banding. This is mainly because there is a need to analyse a great number of cells to identify low-level mosaicism. Also, the clinical implications of very low-level mosaicisms are often unknown. The use of arrays with SNP probes may partly solve this problem as they are more sensitive to detection of low-level mosaicism. The cost of array CGH analysis is still higher than for cytogenetic analysis, although this is expected to decrease and it may be offset by the higher yield of chromosomal abnormalities detected.

Another limitation/challenge of array CGH analysis in SAs is the identification of many CNVs with unknown clinical significance as discussed before. However, the challenge of counseling couples about unpredictable outcomes is not new (26) and as more knowledge will be gained the number of variants with unknown function is expected to decrease. Array CGH analysis could also detect some unanticipated CNVs that are unrelated to the cause of abortion, but might be of concern to the couple. These CNVs might reveal a carrier recessive status for a recessive condition, be diagnostic or predictive of an adult onset/ presymptomatic or undiagnosed condition or be associated with the risk of neoplasia (65). Therefore it could be good practice to inform couples of all possible outcomes before analysis and ask if they are interested in receiving such information and at what time.

In summary, we believe that array CGH may replace conventional karyotyping for analysis of SAs in the future. Detection rates of chromosomal abnormalities has been shown to be equal or improved compared to conventional karyotyping, mainly because of success with nonviable tissue in cases of tissue culture failure. One possibility would be to implement array CGH as an additional test to conventional karyotyping, e.g. in cases of tissue culture failure. The addition of SNP probes would also be of advantage, for detection of polyploidy, low-level mosaicism, and absence of heterozygosity, which could indicate uniparental disomy, and an increased risk of imprinting disorders or lethal recessive disease. Alternatively array CGH and karyotyping could be considered complementary techniques, one possibly following the other if the first test does not reveal a cause for the SA.

5.5 Future considerations

In many cases the clinical significance of CNVs for SAs was unknown, although two candidates were found, i.e. the apparent 34 bp homozygous loss overlapping part of one exon in the *TAF4* gene in case 22 and the apparent homozygous loss overlapping part of an exon in the *GDF6* gene. Future studies in an effort to confirm the loss in the *TAF4* gene could involve qPCR analysis using a primer with a binding site within the deleted area. In order to confirm the pathogenicity of such a homozygous loss we would also need to analyse an unselected pool of living subjects to establish the frequency of heterozygous and homozygous losses. It may also be of interest to see the phenotypic effects of this exact homozygous loss in mice. It would also be possible to analyse loss with a custom high resolution microarray with increased probe density in the deleted area. Further studies are also needed to confirm the possible heterozygous or homozygous loss of the *GDF6* gene. These studies may include long-range PCR and sequencing over the breakpoints. It is possible that the parents both carry heterozygous losses of different size and that the loss in the fetus is partly heterozygous and partly homozygous. Therefore it could also be possible to use another TaqMan assay with a probe binding site at the point of lowest log2 ratio score on the array CGH analysis. It would also be possible to anlyse the loss further with another custom high resolution site-specific microarray.

The high number of CNV calls, especially small CNVs with unknown clinical significance poses a challenge to interpretation. However, it would be possible to apply more stringent criteria to the algorithm and use a size threshold or a specific filter for reviewing and reporting of CNVs. For example Reddy *et al.* (2012) only included variants of 500 kb or larger in their analysis on 532 stillbirths (26). Some laboratories also use specific arrays and/or add specific filters to the data before analysis of prenatal diagnosis cases. These filters are designed to mask a large proportion of the genome so that only regions that have been associated with clinically recognized genetic syndromes are reviewed. Another option would be to use an other array CGH platform with medium-high resolution and possibly including SNP probes. As more information on clinical significance of CNVs and gene function is acquired with high-throughput sequencing it may be more relevant to use high resolution arrays and/or high-throughput sequencing for clinical assessment of SAs.

For future research with high resolution exon-focused arrays on SAs it would be an advantage to analyse a control group as well, as it would help in classifying the variants as benign or clinically significant, and also be useful for the gene ontology analysis. We did, however, not have access to any suitable control group. One possible group could be the parents, which would also give information on the origin of the variants. However, as this was a small study we believed it would be more informative to analyse more samples from SAs rather than analysing trios from fewer samples. It may be a matter of consideration if parents need to be analysed as well if the intention is to use high resolution exon-focused arrays to identify exonic variants that could cause SAs. That could greatly minimize the number of variants with unknown clinical significance as *de novo* variants are more likely to be causative for the abortions, although there is still an issue with reproducibility. Another possible control group could be from CVS samples from healthy live births, which would resemble the genetic material in found CVS tissue of SAs more than blood.

As the phenotypic effect of many CNVs is still unknown it could also be interesting for future research on SAs to incorporate some gene functional studies into CNV analysis with exon-focused arrays. In these cases RNA from live cells could be kept for gene functional analysis to further assess a genotype-phenotype correlation of the CNVs identified.

It could also be of interest to analyse sporadic SAs instead of SAs from couples with RAs, as this may increase the likelihood of finding *de novo* CNVs. SAs from couples with several children and no history of pregnancy loss would be preferential. Testing sporadic SAs could possibly also give information about the frequency of *de novo* CNVs in humans and an idea of to what extent *de novo* CNVs are a cause of SAs.

6 Conclusions

The advantages of using high resolution exon-focused CGH arrays for analysis of SAs is the possibility of finding new and small CNVs (possibly exon-level changes) that could explain the cause of the SA and give relevant information for future risk assessment for the couple. While testing the high resolution exon-focused CGH array we found a significantly larger amount of CNVs in SAs with normal karyotype than previous studies. Overall, however, there was not enough evidence of pathogenicity of the CNVs identified to use the data in genetic counseling for the couple involved. Our conclusion is that we do not recommend use of this 720K exon-focused CGH array in research or for clinical assessment of SAs. The main reasons for this is the high number of CNVs detected, specially small rare exonic CNVs with unknown clinical significance. Also, because of technical challenges with determining if the calls are false positive. The technical challenges and additional cost with performing verification tests are also a further drawback.

We did not find any known causes of SAs with this type of array. Nevertheless, this is the first study to use exon-focused high resolution array for analysis on SAs and we have shown the possibility of finding possible candidate genes for new causes of SAs.

References

1. Griebel CP, Halvorsen J, Golemon TB, Day AA. Management of spontaneous abortion. American family physician. 2005;72(7):1243-50. Epub 2005/10/18.
2. van den Berg MM, van Maarle MC, van Wely M, Goddijn M. Genetics of early miscarriage. Biochimica et biophysica acta. 2012;1822(12):1951-9. Epub 2012/07/17.
3. Carp HJ. Recurrent miscarriage: genetic factors and assessment of the embryo. The Israel Medical Association journal : IMAJ. 2008;10(3):229-31. Epub 2008/05/23.
4. Horne AW, Alexander CI. Recurrent miscarriage. The journal of family planning and reproductive health care / Faculty of Family Planning & Reproductive Health Care, Royal College of Obstetricians & Gynaecologists. 2005;31(2):103-7. Epub 2005/06/01.
5. Warren JE, Silver RM. Genetics of pregnancy loss. Clinical obstetrics and gynecology. 2008;51(1):84-95. Epub 2008/02/28.
6. South ST, Lee C, Lamb AN, Higgins AW, Kearney HM, Working Group for the American College of Medical G, et al. ACMG Standards and Guidelines for constitutional cytogenomic microarray analysis, including postnatal and prenatal applications: revision 2013. Genetics in medicine : official journal of the American College of Medical Genetics. 2013;15(11):901-9. Epub 2013/09/28.
7. Stojilkovic-Mikic T, Mann K, Docherty Z, Mackie Ogilvie C. Maternal cell contamination of prenatal samples assessed by QF-PCR genotyping. Prenatal diagnosis. 2005;25(1):79-83. Epub 2005/01/22.
8. Robberecht C, Schuddinck V, Fryns JP, Vermeesch JR. Diagnosis of miscarriages by molecular karyotyping: benefits and pitfalls. Genetics in medicine : official journal of the American College of Medical Genetics. 2009;11(9):646-54. Epub 2009/07/21.
9. Menten B, Maas N, Thienpont B, Buysse K, Vandesompele J, Melotte C, et al. Emerging patterns of cryptic chromosomal imbalance in patients with idiopathic mental retardation and multiple congenital anomalies: a new series of 140 patients and review of published reports. Journal of medical genetics. 2006;43(8):625-33. Epub 2006/02/24.
10. Miller DT, Adam MP, Aradhya S, Biesecker LG, Brothman AR, Carter NP, et al. Consensus statement: chromosomal microarray is a first-tier clinical diagnostic test for individuals with developmental disabilities or congenital anomalies. American journal of human genetics. 2010;86(5):749-64. Epub 2010/05/15.
11. Schaeffer AJ, Chung J, Heretis K, Wong A, Ledbetter DH, Lese Martin C. Comparative genomic hybridization-array analysis enhances the detection of aneuploidies and submicroscopic imbalances in spontaneous miscarriages. American journal of human genetics. 2004;74(6):1168-74. Epub 2004/05/06.
12. Benkhalifa M, Kasakyan S, Clement P, Baldi M, Tachdjian G, Demirol A, et al. Array comparative genomic hybridization profiling of first-trimester spontaneous abortions that fail to grow in vitro. Prenatal diagnosis. 2005;25(10):894-900. Epub 2005/08/10.
13. Shimokawa O, Harada N, Miyake N, Satoh K, Mizuguchi T, Niikawa N, et al. Array comparative genomic hybridization analysis in first-trimester spontaneous abortions with 'normal' karyotypes. American journal of medical genetics Part A. 2006;140(18):1931-5. Epub 2006/08/15.
14. Menten B, Swerts K, Delle Chiaie B, Janssens S, Buysse K, Philippe J, et al. Array comparative genomic hybridization and flow cytometry analysis of spontaneous abortions and mors in utero samples. BMC medical genetics. 2009;10:89. Epub 2009/09/16.
15. Warren JE, Turok DK, Maxwell TM, Brothman AR, Silver RM. Array comparative genomic hybridization for genetic evaluation of fetal loss between 10 and 20 weeks of gestation. Obstetrics and gynecology. 2009;114(5):1093-102. Epub 2010/02/20.

16. Zhang YX, Zhang YP, Gu Y, Guan FJ, Li SL, Xie JS, et al. Genetic analysis of first-trimester miscarriages with a combination of cytogenetic karyotyping, microsatellite genotyping and arrayCGH. Clinical genetics. 2009;75(2):133-40. Epub 2009/02/14.
17. Deshpande M, Harper J, Holloway M, Palmer R, Wang R. Evaluation of array comparative genomic hybridization for genetic analysis of chorionic villus sampling from pregnancy loss in comparison to karyotyping and multiplex ligation-dependent probe amplification. Genetic testing and molecular biomarkers. 2010;14(3):421-4. Epub 2010/04/23.
18. Rajcan-Separovic E, Diego-Alvarez D, Robinson WP, Tyson C, Qiao Y, Harvard C, et al. Identification of copy number variants in miscarriages from couples with idiopathic recurrent pregnancy loss. Hum Reprod. 2010;25(11):2913-22. Epub 2010/09/18.
19. Rajcan-Separovic E, Qiao Y, Tyson C, Harvard C, Fawcett C, Kalousek D, et al. Genomic changes detected by array CGH in human embryos with developmental defects. Molecular human reproduction. 2010;16(2):125-34. Epub 2009/09/26.
20. Lathi RB, Massie JA, Loring M, Demko ZP, Johnson D, Sigurjonsson S, et al. Informatics enhanced SNP microarray analysis of 30 miscarriage samples compared to routine cytogenetics. PloS one. 2012;7(3):e31282. Epub 2012/03/10.
21. Viaggi CD, Cavani S, Malacarne M, Floriddia F, Zerega G, Baldo C, et al. First-trimester euploid miscarriages analysed by array-CGH. Journal of applied genetics. 2013;54(3):353-9. Epub 2013/06/20.
22. Gao J, Liu C, Yao F, Hao N, Zhou J, Zhou Q, et al. Array-based comparative genomic hybridization is more informative than conventional karyotyping and fluorescence in situ hybridization in the analysis of first-trimester spontaneous abortion. Molecular cytogenetics. 2012;5(1):33. Epub 2012/07/17.
23. Itsara A, Cooper GM, Baker C, Girirajan S, Li J, Absher D, et al. Population analysis of large copy number variants and hotspots of human genetic disease. American journal of human genetics. 2009;84(2):148-61. Epub 2009/01/27.
24. Cooper GM, Coe BP, Girirajan S, Rosenfeld JA, Vu TH, Baker C, et al. A copy number variation morbidity map of developmental delay. Nature genetics. 2011;43(9):838-46. Epub 2011/08/16.
25. Wapner RJ, Martin CL, Levy B, Ballif BC, Eng CM, Zachary JM, et al. Chromosomal microarray versus karyotyping for prenatal diagnosis. The New England journal of medicine. 2012;367(23):2175-84. Epub 2012/12/12.
26. Reddy UM, Page GP, Saade GR, Silver RM, Thorsten VR, Parker CB, et al. Karyotype versus microarray testing for genetic abnormalities after stillbirth. The New England journal of medicine. 2012;367(23):2185-93. Epub 2012/12/12.
27. van den Boogaard E, Kaandorp SP, Franssen MT, Mol BW, Leschot NJ, Wouters CH, et al. Consecutive or non-consecutive recurrent miscarriage: is there any difference in carrier status? Hum Reprod. 2010;25(6):1411-4. Epub 2010/04/13.
28. Stephenson M, Kutteh W. Evaluation and management of recurrent early pregnancy loss. Clinical obstetrics and gynecology. 2007;50(1):132-45. Epub 2007/02/17.
29. Schreck R, Silverman NS. Fetal Loss. In: Rimoin DL, Emery AEH, editors. Emery and Rimoin's principles and practice of medical genetics. 5th ed. Philadelphia, Pa.: Churchill Livingstone Elsevier; 2007. p. 875-86.
30. Christiansen OB, Mathiesen O, Lauritsen JG, Grunnet N. Idiopathic recurrent spontaneous abortion. Evidence of a familial predisposition. Acta obstetricia et gynecologica Scandinavica. 1990;69(7-8):597-601. Epub 1990/01/01.
31. Kolte AM, Nielsen HS, Moltke I, Degn B, Pedersen B, Sunde L, et al. A genome-wide scan in affected sibling pairs with idiopathic recurrent miscarriage suggests genetic linkage. Molecular human reproduction. 2011;17(6):379-85. Epub 2011/01/25.

32. Christiansen OB, Riisom K, Lauritsen JG, Grunnet N, Jersild C. Association of maternal HLA haplotypes with recurrent spontaneous abortions. Tissue antigens. 1989;34(3):190-9. Epub 1989/09/01.
33. Wang X, Jiang W, Zhang D. Association of 14-bp insertion/deletion polymorphism of HLA-G gene with unexplained recurrent spontaneous abortion: a meta-analysis. Tissue antigens. 2013;81(2):108-15. Epub 2013/01/22.
34. Komlos L, Zamir R, Joshua H, Halbrecht I. Common HLA antigens in couples with repeated abortions. Clinical immunology and immunopathology. 1977;7(3):330-5. Epub 1977/05/01.
35. Gerencer M, Kastelan A, Drazancic A, Kerhin-Brkljacic V, Madjaric M. The HLA antigens in women with recurrent abnormal pregnancies of unknown etiology. Tissue antigens. 1978;12(3):223-7. Epub 1978/09/01.
36. Schacter B, Muir A, Gyves M, Tasin M. HLA-A,B compatibility in parents of offspring with neural-tube defects or couples experiencing involuntary fetal wastage. Lancet. 1979;1(8120):796-9. Epub 1979/04/14.
37. Aoki K. HLA-DR compatibility in couples with recurrent spontaneous abortions. Nihon Sanka Fujinka Gakkai zasshi. 1982;34(10):1773-80. Epub 1982/10/01.
38. Unander AM, Olding LB. Easily suppressed lymphocytes and absence of cytotoxic antibody in three women with habitual abortion. American journal of reproductive immunology : AJRI : official journal of the American Society for the Immunology of Reproduction and the International Coordination Committee for Immunology of Reproduction. 1982;2(5):254-9. Epub 1982/10/01.
39. McIntyre JA, Faulk WP. Recurrent spontaneous abortion in human pregnancy: results of immunogenetical, cellular, and humoral studies. American journal of reproductive immunology : AJRI : official journal of the American Society for the Immunology of Reproduction and the International Coordination Committee for Immunology of Reproduction. 1983;4(4):165-70. Epub 1983/12/01.
40. Unander AM, Olding LB. Habitual abortion: parental sharing of HLA antigens, absence of maternal blocking antibody, and suppression of maternal lymphocytes. American journal of reproductive immunology : AJRI : official journal of the American Society for the Immunology of Reproduction and the International Coordination Committee for Immunology of Reproduction. 1983;4(4):171-8. Epub 1983/12/01.
41. McIntyre JA, McConnachie PR, Taylor CG, Faulk WP. Clinical, immunologic, and genetic definitions of primary and secondary recurrent spontaneous abortions. Fertility and sterility. 1984;42(6):849-55. Epub 1984/12/01.
42. Beer AE, Semprini AE, Zhu XY, Quebbeman JF. Pregnancy outcome in human couples with recurrent spontaneous abortions: HLA antigen profiles; HLA antigen sharing; female serum MLR blocking factors; and paternal leukocyte immunization. Experimental and clinical immunogenetics. 1985;2(3):137-53. Epub 1985/01/01.
43. Bolis PF, Soro V, Martinetti Bianchi M, Belvedere M. HLA compatibility and human reproduction. Clinical and experimental obstetrics & gynecology. 1985;12(1-2):9-12. Epub 1985/01/01.
44. Casciani CU, Pasetto N, Forleo R, Adorno D, Valeri M, Piazza A. HLA sharing in couples with recurrent abortion. Experimental and clinical immunogenetics. 1985;2(2):65-9. Epub 1985/01/01.
45. Johnson PM, Barnes RM, Risk JM, Molloy CM, Woodrow JC. Immunogenetic studies of recurrent spontaneous abortions in humans. Experimental and clinical immunogenetics. 1985;2(2):77-83. Epub 1985/01/01.

46. Moghraby JS, Tamim H, Anacan V, Al Khalaf H, Moghraby SA. HLA sharing among couples appears unrelated to idiopathic recurrent fetal loss in Saudi Arabia. Hum Reprod. 2010;25(8):1900-5. Epub 2010/06/23.
47. Sagot P, Bignon J, Cesbron A, Cheneau ML, Boog G, Muller JY. Lack of evidence for a role of HLA-DP in unexplained recurrent spontaneous abortion. Transfusion clinique et biologique : journal de la Societe francaise de transfusion sanguine. 1995;2(3):145-50. Epub 1995/01/01.
48. Power DA, Catto GR, Mason RJ, MacLeod AM, Stewart GM, Stewart KN, et al. The fetus as an allograft: evidence for protective antibodies to HLA-linked paternal antigens. Lancet. 1983;2(8352):701-4. Epub 1983/09/24.
49. Kishore R, Agarwal S, Halder A, Das V, Shukla BR, Agarwal SS. HLA sharing, anti-paternal cytotoxic antibodies and MLR blocking factors in women with recurrent spontaneous abortion. The journal of obstetrics and gynaecology research. 1996;22(2):177-83. Epub 1996/04/01.
50. Hedrick PW. HLA-sharing, recurrent spontaneous abortion, and the genetic hypothesis. Genetics. 1988;119(1):199-204. Epub 1988/05/01.
51. Simpson JL. Causes of fetal wastage. Clinical obstetrics and gynecology. 2007;50(1):10-30. Epub 2007/02/17.
52. Li TC, Makris M, Tomsu M, Tuckerman E, Laird S. Recurrent miscarriage: aetiology, management and prognosis. Human reproduction update. 2002;8(5):463-81. Epub 2002/10/26.
53. Shaffer LG, Bejjani BA, Torchia B, Kirkpatrick S, Coppinger J, Ballif BC. The identification of microdeletion syndromes and other chromosome abnormalities: cytogenetic methods of the past, new technologies for the future. American journal of medical genetics Part C, Seminars in medical genetics. 2007;145C(4):335-45. Epub 2007/10/03.
54. Tuzun E, Sharp AJ, Bailey JA, Kaul R, Morrison VA, Pertz LM, et al. Fine-scale structural variation of the human genome. Nature genetics. 2005;37(7):727-32. Epub 2005/05/17.
55. Conrad DF, Andrews TD, Carter NP, Hurles ME, Pritchard JK. A high-resolution survey of deletion polymorphism in the human genome. Nature genetics. 2006;38(1):75-81. Epub 2005/12/06.
56. Hinds DA, Kloek AP, Jen M, Chen X, Frazer KA. Common deletions and SNPs are in linkage disequilibrium in the human genome. Nature genetics. 2006;38(1):82-5. Epub 2005/12/06.
57. McCarroll SA, Hadnott TN, Perry GH, Sabeti PC, Zody MC, Barrett JC, et al. Common deletion polymorphisms in the human genome. Nature genetics. 2006;38(1):86-92. Epub 2006/02/10.
58. Redon R, Ishikawa S, Fitch KR, Feuk L, Perry GH, Andrews TD, et al. Global variation in copy number in the human genome. Nature. 2006;444(7118):444-54. Epub 2006/11/24.
59. Korbel JO, Urban AE, Affourtit JP, Godwin B, Grubert F, Simons JF, et al. Paired-end mapping reveals extensive structural variation in the human genome. Science. 2007;318(5849):420-6. Epub 2007/09/29.
60. Genomes Project C, Abecasis GR, Auton A, Brooks LD, DePristo MA, Durbin RM, et al. An integrated map of genetic variation from 1,092 human genomes. Nature. 2012;491(7422):56-65. Epub 2012/11/07.
61. Girirajan S, Campbell CD, Eichler EE. Human copy number variation and complex genetic disease. Annual review of genetics. 2011;45:203-26. Epub 2011/08/23.

62. Pang AW, MacDonald JR, Pinto D, Wei J, Rafiq MA, Conrad DF, et al. Towards a comprehensive structural variation map of an individual human genome. Genome biology. 2010;11(5):R52. Epub 2010/05/21.
63. Pinto D, Darvishi K, Shi X, Rajan D, Rigler D, Fitzgerald T, et al. Comprehensive assessment of array-based platforms and calling algorithms for detection of copy number variants. Nature biotechnology. 2011;29(6):512-20. Epub 2011/05/10.
64. Conrad DF, Pinto D, Redon R, Feuk L, Gokcumen O, Zhang Y, et al. Origins and functional impact of copy number variation in the human genome. Nature. 2010;464(7289):704-12. Epub 2009/10/09.
65. Kearney HM, Thorland EC, Brown KK, Quintero-Rivera F, South ST. American College of Medical Genetics standards and guidelines for interpretation and reporting of postnatal constitutional copy number variants. Genetics in medicine : official journal of the American College of Medical Genetics. 2011;13(7):680-5. Epub 2011/06/18.
66. Lee C, Iafrate AJ, Brothman AR. Copy number variations and clinical cytogenetic diagnosis of constitutional disorders. Nature genetics. 2007;39(7 Suppl):S48-54. Epub 2007/09/05.
67. Nijman SM. Synthetic lethality: general principles, utility and detection using genetic screens in human cells. FEBS letters. 2011;585(1):1-6. Epub 2010/11/26.
68. Dedes KJ, Wilkerson PM, Wetterskog D, Weigelt B, Ashworth A, Reis-Filho JS. Synthetic lethality of PARP inhibition in cancers lacking BRCA1 and BRCA2 mutations. Cell cycle. 2011;10(8):1192-9. Epub 2011/04/14.
69. Le Meur N, Gentleman R. Modeling synthetic lethality. Genome biology. 2008;9(9):R135. Epub 2008/09/16.
70. Simons AH, Dafni N, Dotan I, Oron Y, Canaani D. Genetic synthetic lethality screen at the single gene level in cultured human cells. Nucleic acids research. 2001;29(20):E100. Epub 2001/10/16.
71. Vockley J, Rinaldo P, Bennett MJ, Matern D, Vladutiu GD. Synergistic heterozygosity: disease resulting from multiple partial defects in one or more metabolic pathways. Molecular genetics and metabolism. 2000;71(1-2):10-8. Epub 2000/09/26.
72. Girirajan S, Eichler EE. Phenotypic variability and genetic susceptibility to genomic disorders. Human molecular genetics. 2010;19(R2):R176-87. Epub 2010/09/03.
73. Liu P, Carvalho CM, Hastings PJ, Lupski JR. Mechanisms for recurrent and complex human genomic rearrangements. Current opinion in genetics & development. 2012;22(3):211-20. Epub 2012/03/24.
74. Conrad DF, Bird C, Blackburne B, Lindsay S, Mamanova L, Lee C, et al. Mutation spectrum revealed by breakpoint sequencing of human germline CNVs. Nature genetics. 2010;42(5):385-91. Epub 2010/04/07.
75. Itsara A, Wu H, Smith JD, Nickerson DA, Romieu I, London SJ, et al. De novo rates and selection of large copy number variation. Genome research. 2010;20(11):1469-81. Epub 2010/09/16.
76. Campbell CD, Eichler EE. Properties and rates of germline mutations in humans. Trends in genetics : TIG. 2013. Epub 2013/05/21.
77. Wang J, Fan HC, Behr B, Quake SR. Genome-wide single-cell analysis of recombination activity and de novo mutation rates in human sperm. Cell. 2012;150(2):402-12. Epub 2012/07/24.
78. Kong A, Frigge ML, Masson G, Besenbacher S, Sulem P, Magnusson G, et al. Rate of de novo mutations and the importance of father's age to disease risk. Nature. 2012;488(7412):471-5. Epub 2012/08/24.

79. Sun JX, Helgason A, Masson G, Ebenesersdottir SS, Li H, Mallick S, et al. A direct characterization of human mutation based on microsatellites. Nature genetics. 2012;44(10):1161-5. Epub 2012/08/28.
80. Reik W, Walter J. Genomic imprinting: parental influence on the genome. Nature reviews Genetics. 2001;2(1):21-32. Epub 2001/03/17.
81. Isles AR, Holland AJ. Imprinted genes and mother-offspring interactions. Early human development. 2005;81(1):73-7. Epub 2005/02/15.
82. Camprubi C, Monk D. Does genomic imprinting play a role in autoimmunity? Advances in experimental medicine and biology. 2011;711:103-16. Epub 2011/06/02.
83. Semi K, Matsuda Y, Ohnishi K, Yamada Y. Cellular reprogramming and cancer development. International journal of cancer Journal international du cancer. 2013;132(6):1240-8. Epub 2012/11/28.
84. Swales AK, Spears N. Genomic imprinting and reproduction. Reproduction. 2005;130(4):389-99. Epub 2005/09/27.
85. Biniszkiewicz D, Gribnau J, Ramsahoye B, Gaudet F, Eggan K, Humpherys D, et al. Dnmt1 overexpression causes genomic hypermethylation, loss of imprinting, and embryonic lethality. Molecular and cellular biology. 2002;22(7):2124-35. Epub 2002/03/09.
86. Bajaj V, Markandaya M, Krishna L, Kumar A. Paternal imprinting of the SLC22A1LS gene located in the human chromosome segment 11p15.5. BMC genetics. 2004;5:13. Epub 2004/06/04.
87. Kaneda M, Okano M, Hata K, Sado T, Tsujimoto N, Li E, et al. Essential role for de novo DNA methyltransferase Dnmt3a in paternal and maternal imprinting. Nature. 2004;429(6994):900-3. Epub 2004/06/25.
88. Ono R, Nakamura K, Inoue K, Naruse M, Usami T, Wakisaka-Saito N, et al. Deletion of Peg10, an imprinted gene acquired from a retrotransposon, causes early embryonic lethality. Nature genetics. 2006;38(1):101-6. Epub 2005/12/13.
89. Ostojic S, Pereza N, Volk M, Kapovic M, Peterlin B. Genetic predisposition to idiopathic recurrent spontaneous abortion: contribution of genetic variations in IGF-2 and H19 imprinted genes. Am J Reprod Immunol. 2008;60(2):111-7. Epub 2008/06/25.
90. Doria S, Sousa M, Fernandes S, Ramalho C, Brandao O, Matias A, et al. Gene expression pattern of IGF2, PHLDA2, PEG10 and CDKN1C imprinted genes in spontaneous miscarriages or fetal deaths. Epigenetics : official journal of the DNA Methylation Society. 2010;5(5):444-50. Epub 2010/05/21.
91. Pliushch G, Schneider E, Weise D, El Hajj N, Tresch A, Seidmann L, et al. Extreme methylation values of imprinted genes in human abortions and stillbirths. The American journal of pathology. 2010;176(3):1084-90. Epub 2010/01/23.
92. Ankolkar M, Patil A, Warke H, Salvi V, Kedia Mokashi N, Pathak S, et al. Methylation analysis of idiopathic recurrent spontaneous miscarriage cases reveals aberrant imprinting at H19 ICR in normozoospermic individuals. Fertility and sterility. 2012;98(5):1186-92. Epub 2012/09/11.
93. Carrell DT. Aberrant methylation of the H19 imprinting control region may increase the risk of spontaneous abortion. Epigenomics. 2013;5(1):23-4. Epub 2013/04/02.
94. Zheng HY, Tang Y, Niu J, Li P, Ye DS, Chen X, et al. Aberrant DNA methylation of imprinted loci in human spontaneous abortions after assisted reproduction techniques and natural conception. Hum Reprod. 2013;28(1):265-73. Epub 2012/10/09.
95. Rosenfeld JA, Coe BP, Eichler EE, Cuckle H, Shaffer LG. Estimates of penetrance for recurrent pathogenic copy-number variations. Genetics in medicine : official journal of the American College of Medical Genetics. 2012. Epub 2012/12/22.

96. Helderman-van den Enden AT, de Jong R, den Dunnen JT, Houwing-Duistermaat JJ, Kneppers AL, Ginjaar HB, et al. Recurrence risk due to germ line mosaicism: Duchenne and Becker muscular dystrophy. Clinical genetics. 2009;75(5):465-72. Epub 2009/05/29.
97. Rothlisberger B, Kotzot D. Recurrence risk in de novo structural chromosomal rearrangements. American journal of medical genetics Part A. 2007;143A(15):1708-14. Epub 2007/07/03.
98. McCarroll SA, Kuruvilla FG, Korn JM, Cawley S, Nemesh J, Wysoker A, et al. Integrated detection and population-genetic analysis of SNPs and copy number variation. Nature genetics. 2008;40(10):1166-74. Epub 2008/09/09.
99. South ST, Rope AF, Lamb AN, Aston E, Glaus N, Whitby H, et al. Expansion in size of a terminal deletion: a paradigm shift for parental follow-up studies. Journal of medical genetics. 2008;45(6):391-5. Epub 2008/04/17.
100. Mefford HC, Yendle SC, Hsu C, Cook J, Geraghty E, McMahon JM, et al. Rare copy number variants are an important cause of epileptic encephalopathies. Annals of neurology. 2011;70(6):974-85. Epub 2011/12/23.
101. Shimohata T, Nakajima T, Yamada M, Uchida C, Onodera O, Naruse S, et al. Expanded polyglutamine stretches interact with TAFII130, interfering with CREB-dependent transcription. Nature genetics. 2000;26(1):29-36. Epub 2000/09/06.
102. International HapMap C, Altshuler DM, Gibbs RA, Peltonen L, Altshuler DM, Gibbs RA, et al. Integrating common and rare genetic variation in diverse human populations. Nature. 2010;467(7311):52-8. Epub 2010/09/03.
103. Leo A, Walker AM, Lebo MS, Hendrickson B, Scholl T, Akmaev VR. A GC-wave correction algorithm that improves the analytical performance of aCGH. The Journal of molecular diagnostics : JMD. 2012;14(6):550-9. Epub 2012/08/28.
104. Muncke N, Wogatzky BS, Breuning M, Sistermans EA, Endris V, Ross M, et al. Position effect on PLP1 may cause a subset of Pelizaeus-Merzbacher disease symptoms. Journal of medical genetics. 2004;41(12):e121. Epub 2004/12/14.
105. Inoue K, Tanaka H, Scaglia F, Araki A, Shaffer LG, Lupski JR. Compensating for central nervous system dysmyelination: females with a proteolipid protein gene duplication and sustained clinical improvement. Annals of neurology. 2001;50(6):747-54. Epub 2002/01/05.
106. Park H, Kim JI, Ju YS, Gokcumen O, Mills RE, Kim S, et al. Discovery of common Asian copy number variants using integrated high-resolution array CGH and massively parallel DNA sequencing. Nature genetics. 2010;42(5):400-5. Epub 2010/04/07.
107. Webber C. Functional enrichment analysis with structural variants: pitfalls and strategies. Cytogenetic and genome research. 2011;135(3-4):277-85. Epub 2011/10/15.
108. Lee JH, Ki CS, Kim HJ, Suh W, Lee ST, Kim JW, et al. Analysis of copy number variation using whole genome exon-focused array CGH in Korean patients with primary congenital glaucoma. Molecular vision. 2011;17:3583-90. Epub 2012/01/06.
109. Celestino-Soper PB, Shaw CA, Sanders SJ, Li J, Murtha MT, Ercan-Sencicek AG, et al. Use of array CGH to detect exonic copy number variants throughout the genome in autism families detects a novel deletion in TMLHE. Human molecular genetics. 2011;20(22):4360-70. Epub 2011/08/26.
110. Pinheiro LB, Coleman VA, Hindson CM, Herrmann J, Hindson BJ, Bhat S, et al. Evaluation of a droplet digital polymerase chain reaction format for DNA copy number quantification. Analytical chemistry. 2012;84(2):1003-11. Epub 2011/11/30.
111. Jauniaux E, Farquharson RG, Christiansen OB, Exalto N. Evidence-based guidelines for the investigation and medical treatment of recurrent miscarriage. Hum Reprod. 2006;21(9):2216-22. Epub 2006/05/19.

112. Royal College of Obstetricians and Gynaecologists (RCOG), The investigation and treatment of couples with recurrent miscarriage, 2003. London, United Kingdom. Ref Type: Generic.

Appendix

Supplementary Table 1: TaqMan Copy Number Assays.

Gene name	Assay ID	Amplicon length
NGFRAP1	ID: Hs01597380_cn	102 bp
GDF6	ID: Hs01657484_cn	97 bp
GDF6	ID: Hs02148305_cn	110 bp
HLA-DRB1 and *HLA-DRB5*	ID: Hs04317542_cn	106 bp
Area adjacent to *HLA-DRB5*	ID: Hs03590782_cn	109 bp
HLA-A pseudogene	ID: Hs03587795_cn	76 bp

Supplementary Table 2: SYBR Green I primers.

Gene name	Primer	Sequence (5´ → 3´)	Amplicon size
KIAA1217 (target)	Forward	CAG AAG AGG TGC CTC TCA GC	113
	Reverse	AGC CTG GAA CAT TGG TGA AC	
LRRC8D (reference)	Forward	GAC CAA AGT TCC CTC CAA CA	129
	Reverse	AGT TCC AGC TCA GCG ACA TT	

VÍSINDARANNSÓKN Á ERFÐUM FÓSTURLÁTA.
Staðlað vinnublað til að skrá upplýsingar.

1. **Fjöldi fósturláta**_____
Hvenær á meðgöngu áttu fósturlátin sér stað?
Fyrsta fósturlát_____vikur/mán.
Annað fósturlát_____vikur/mán.
Þriðja fósturlát_____vikur/mán.
Fleiri?

Öll með sama maka?

Skýring fósturláta ef hún er þekkt:
Fyrsta fósturlát_____
Annað fósturlát_____
Þriðja fósturlát_____

2. **Hvenær á meðgöngu átti nýafstaðið fósturlát sérstað:_____vikur/mán**
Skýring fósturláts:_____
Var krufning gerð? _____
Var gerð litningarannsókn eða önnur rannsókn?

3. **Hafa orðið fósturlát hjá öðrum í fjölskyldunni?**
Ef já, þá hver mörg?

	Fjölskylda konunnar	**Fjölskylda mannsins**
a) Foreldrar konu	Fjöldi____	Fjöldi____
b) Systkini konu	Fjöldi____	Fjöldi____
c) Foreldrar manns	Fjöldi____	Fjöldi____
d) Systkini manns	Fjöldi____	Fjöldi____

4. **Hversu mörg börn áttu foreldrar konu:** _____

5. **Hversu mörg börn áttu foreldrar manns:** _____

6. **Hefur kona eftirfarandi sjúkdóma sem auka líkur á fósturláti:**

a. Sykursýki
Já ___Nei ___

b. Flogaveiki
Já ___Nei ___

7. **Er saga um einhverja erfðasjúkdóma í fjölskyldunni?**

Fjölskylda konu
Já ___Nei ___

Fjölskylda manns
Já ___Nei ___

Ef já, hvernig er sá einstaklingur skyldur?
Konu:_____
Manni:_____

Hvaða erfðasjúkdóm er um að ræða?
_____

8. Eru aðrir einstaklingar með sköpulagsgalla (fæðingargalla) í ættinni?

Fjölskylda konu	**Fjölskylda manns**
Já ___Nei ___	Já ___Nei ___

Ef já, hvernig er sá einstaklingur skyldur?
Konu:_____
Manni:_____

Hvers konar sköpulagsgalla er um að ræða?
_____
_____

9. Er kunnugt um einstaklinga með þroskaskerðingu í ætt?

Fjölskylda konu	**Fjölskylda manns**
Já ___Nei ___	Já ___Nei ___

Ef já, hvernig er sá einstaklingur skyldur?
Konu:_____
Manni:_____

Skýring ef hún er þekkt:
_____
_____

10. Er konan eða maðurinn með einhverja langvarandi sjúkdóma?
Nei___
Já___

Ef já, hvaða sjúkdóma?
Kona:_____
Maður:_____

11. Er þekktur skyldleiki milli pars ?

Veit ekki_____

Ef já, hvernig er honum háttað?
_____