

ARTICLE

De novo deletions and duplications detected by array CGH: a study of parental origin in relation to mechanisms of formation and size of imbalance

Charlene Sibbons^{1,2}, Joan K Morris³, John A Crolla^{1,2}, Patricia A Jacobs^{1,2} and N Simon Thomas^{*,1,2}

We report a large series of 173 patients with physical and/or neurological abnormalities and a *de novo* imbalance identified by array CGH. Breakpoint intervals were screened for the presence of low copy repeats (LCRs) to distinguish between rearrangements formed by non-allelic homologous recombination (NAHR) and rearrangements formed by other mechanisms. We identified significant differences in size and parental origin between the LCR-mediated and non-LCR groups. Non-LCR imbalances were evenly distributed among the four size intervals we defined, whereas LCR-mediated rearrangements had a narrow size distribution, predominantly between 1 and 5 Mb ($P=0.001$). Among the LCR-mediated rearrangements there were equal numbers of maternally and paternally derived cases. In contrast, for the non-LCR rearrangements there was a significant excess of paternal cases ($P=0.024$) over a wide size range including below 1 Mb. Our results provide novel evidence that unbalanced chromosome rearrangements are not only more frequent in males, but may also arise through different mechanisms than those seen in females. Although the paternal imbalances identified in our study are evenly distributed throughout the four size groups, there are very few maternal imbalances either <1 Mb or >10 Mb. Furthermore, a lower proportion of paternal imbalances are LCR mediated (13/71) compared with the maternal imbalances (12/30). We hypothesise that imbalances of maternal origin arise predominantly through NAHR during meiosis, while the majority of imbalances of paternal origin arise through male-specific mechanisms other than NAHR. Our data suggest that mitotic mechanisms could be important for the formation of chromosome imbalances; however, we found no association with increased paternal age.

European Journal of Human Genetics (2012) 20, 155–160; doi:10.1038/ejhg.2011.182; published online 28 September 2011

Keywords: array CGH; *de novo* chromosome abnormality; deletions and duplications; non-allelic homologous recombination; low copy repeats; parental and chromosomal origin

INTRODUCTION

Unbalanced *de novo* chromosome abnormalities are relatively common in humans and are associated with a wide variety of congenital malformations. Before the introduction of array CGH, two groups of structural chromosome imbalances were amenable to study: (1) recurrent imbalances identified by fluorescent *in-situ* hybridisation (FISH) which share a common size with clustered breakpoints; (2) non-recurrent imbalances visible by light microscopy which are distributed throughout the genome, and have different sizes and distinct breakpoint combinations. Although there is some overlap, the two groups of structural abnormality display clear differences in how and when they arise.

With a few exceptions, such as duplications causing type 1A Charcot-Marie-Tooth disease,¹ the common microdeletion and microduplication syndromes (eg, DiGeorge syndrome, DGS and Williams–Beuren syndrome, WBS) have approximately equal numbers of maternal and paternal cases.^{2,3} In contrast, among cytogenetically visible non-recurrent *de novo* imbalances, there is a clear paternal bias and in a study of 115 such abnormalities we showed that the overall proportion of paternally derived cases was 72%.⁴

For recurrent rearrangements, the predominant mechanism is non-allelic homologous recombination (NAHR) mediated by low copy repeats

(LCRs, also called segmental duplications or SDs).⁵ NAHR is a reciprocal process resulting in the gain or loss of the genomic region flanked by the LCRs and is responsible for a large number of genomic disorders including DGS and WBS. NAHR may involve both chromosome homologues (interchromosomal) or separate chromatids of only a single chromosome (intrachromosomal). An interchromosomal origin is likely to indicate a meiotic event while an intrachromosomal origin may be either meiotic or mitotic.^{2,3,5} Recurrent microdeletions and microduplications are mainly interchromosomal and are assumed to arise at meiosis.^{2,3}

Although more frequent, the formation of non-recurrent rearrangements is less well understood. Among cytogenetically visible deletions and duplications, there appear to be approximately equal numbers of interchromosomal and intrachromosomal abnormalities,⁴ although only relatively small numbers have been investigated. Until recently, it was assumed that most non-recurrent imbalances arise through non-homologous end joining, a process that joins double-stranded breaks in the absence of extensive sequence homology.⁵ However, alternative mechanisms have now been proposed such as fork stalling and template switching (FoSTes)⁶ and microhomology-mediated breakpoint-induced replication (MMBIR);⁷ these are mitotic mechanisms based on stalling of the replication fork during DNA replication.

¹Wessex Regional Genetics Laboratory, Salisbury District Hospital, Salisbury, UK; ²Division of Human Genetics, University of Southampton, Southampton, UK; ³Centre for Environmental and Preventive Medicine, Wolfson Institute of Preventive Medicine, London, UK

*Correspondence: Dr NS Thomas, Wessex Regional Genetics Laboratory, Salisbury District Hospital, Salisbury, SP2 8BJ, UK. Tel: +44 1722 429080; Fax: +44 1722 338095; E-mail: simon.thomas@salisbury.nhs.uk

Received 17 May 2011; revised 18 August 2011; accepted 23 August 2011; published online 28 September 2011

The introduction of array CGH has enabled very small, previously unrecognised imbalances to be investigated. Array CGH can define precisely the size of an imbalance allowing the presence of LCRs at the breakpoints to be investigated. This makes it possible to distinguish rearrangements formed by NAHR from rearrangements formed by the other mechanisms. We report a large series of patients referred because of physical and/or neurological abnormalities, who were found to have a *de novo* imbalance by array CGH. We compare those abnormalities attributable to NAHR with those that arose by other mechanisms in terms of size and parental origin, and looked for differences in the origin of maternally and paternally derived imbalances.

METHODS

Study population

The study population comprised a total of 173 patients referred to our diagnostic laboratory between 2008 and 2010 for array CGH testing who carried a *de novo* unbalanced structural chromosome abnormality. Just over half of the study population was originally ascertained through conventional cytogenetic analysis; array CGH was then carried out to confirm the cytogenetic abnormality and/or to define its extent and gene content. For the remaining cases, the structural abnormality was identified directly by array CGH: they either had an apparently normal karyotype and the array identified a cryptic imbalance or they were analysed by array CGH without prior cytogenetic testing.

All patients for whom array CGH was requested were eligible for the study, except for rearrangements involving the X and Y chromosomes and well-established neutral CNVs which were excluded. Patients referred exclusively for a specific microdeletion/duplication syndrome (eg, DGS, WBS, Prader–Willi syndrome and Angelman syndrome) were also excluded, as they were not tested by array CGH.

The 173 patients were classed into three groups according to the number and nature of their structural abnormalities: (single) 150 patients with a single *de novo* imbalance, consisting of 121 deletions (105 interstitial and 16 terminal) and 29 duplications (28 interstitial and 1 terminal); (multiple) six patients with two or more imbalances that were assumed to be independent; (complex) 17 patients with multiple imbalances that were assumed to have been formed from the same rearrangement event. Multiple imbalances were assumed to be independent if at least one was interstitial and they occurred on different chromosomes, or on the same chromosome but were not contiguous. Complex imbalances were assumed to be formed by a single event if they were contiguous or were shown by FISH to be two terminal imbalances likely to represent an unbalanced translocation.

Laboratory methods

All probands were tested by array CGH using either the Agilent/NGRL Wessex Array or 8X60K Oxford Genetics Technology ISCA array ([http://www.ngrl.org.uk/Wessex/downloads/pdf/nGRLW_aCGH_1\[1\].0.pdf](http://www.ngrl.org.uk/Wessex/downloads/pdf/nGRLW_aCGH_1[1].0.pdf)).

Parental origin was determined by amplifying polymorphic microsatellite repeats from each unbalanced chromosome segment selected from the UCSC genome database. Of the 173 patients, we were able to determine the parental origin of the *de novo* imbalances for 124. Results could not be obtained for 49 patients. For 27 patients insufficient DNA was available, for example, where follow-up parental testing was carried out only by FISH or cytogenetics. For the remaining 22 patients we were unable to identify informative microsatellite markers within the deleted interval.

For duplications the microsatellite results were also used to determine whether the rearrangement had arisen through an interchromosomal or intrachromosomal event.

Identification of LCRs

For every imbalance, the UCSC database was also used to look for the presence of LCRs in the breakpoint intervals mapped by array CGH. The breakpoint interval was defined by the maximum and minimum size of the imbalance, and varied in size from 5 to 250 kb according to the array platform used and the density of probes in that genomic region. The origin of an imbalance was

assumed to have been mediated by NAHR if paralogous LCRs spanned all or a large proportion of both the breakpoint intervals, while an imbalance was assumed to have arisen by a mechanism other than the NAHR if LCRs were absent from the breakpoint intervals. The orientation of the LCR was not taken into consideration. For three simple imbalances and two multiple imbalances, an LCR was present at only one breakpoint interval and these were also assumed to have arisen by a mechanism other than the NAHR.

Statistical methods

For clarity in the tables, imbalances are grouped by size into four classes: <1 Mb, between 1 and 5 Mb, between 5 and 10 Mb and greater than 10 Mb. The χ^2 -test and Fisher's exact test were used to examine the association of parental origin and the presence or absence of LCRs at both the breakpoints. The Kolmogorov–Smirnov test for equality of distributions was used to examine the size distribution of cases according to parental origin and the presence or absence of LCRs at both the breakpoints.

Parental age analysis was carried out using the method described in Thomas *et al.*⁸ Logistic regression was used to compare parental ages from this study with data on paternal and maternal ages obtained from the UK Office for National Statistics for all jointly registered births. Cases of paternal origin were adjusted for both date of birth and maternal age, by selecting the national distribution of paternal ages for each patient's year of birth and their mother's age.

RESULTS

Single imbalances

The 150 single imbalances (121 deletions and 29 duplications) were divided into 36 cases with LCRs at both breakpoints and 114 non-LCR cases. Figure 1 shows the sizes of each imbalance according to the presence or absence of LCRs. Deletions and duplications displayed a very similar size distribution pattern and, as there are many more deletions than duplications, they have been grouped together for subsequent analysis.

LCR-mediated single imbalances

Analysis of the 36 LCR-mediated cases showed that the vast majority either corresponded to, or overlapped with, known microdeletion/duplication syndromes (Supplementary Table 1). The only exceptions are two deletions of chromosome 7p22.1, which share a common distal breakpoint but have a different proximal breakpoint; we could find no microdeletion syndrome reported for this cytogenetic band. Many of the microdeletion/duplications identified in this study have only recently been reported in the literature. However, there were also five patients with a classic 22q11.21 deletion (DGV) and one patient with a WS deletion on chromosome 7.

The mean size of the LCR-mediated imbalances was 2.15 Mb: 7 (19%) were smaller than 1 Mb; 26 (72%) were between 1 and 5 Mb in size; three (8%) were between 5 and 10 Mb and none were larger than 10 Mb in size (Figure 1 and Table 1). Parental origin was determined for 25 imbalances, of which 13 were paternal and 12 maternal in origin. The size distribution was similar for imbalances of both maternal and paternal origin (Figure 2 and Table 1).

Non-LCR-mediated single imbalances

The 114 non-LCR single imbalances were, on average, larger than rearrangements mediated by LCRs, with a mean of 5.61 Mb, and were more evenly distributed among the four size groups (Figure 1 and Supplementary Table 2). In all, 30 (27%) were less than 1 Mb, 38 (33%) were between 1 and 5 Mb, 25 (22%) were between 5 and 10 Mb and 21 (18%) were above 10 Mb in size. (Figure 1 and Table 1). There was evidence that the size distribution of the LCR and non-LCR-mediated rearrangements differed ($P=0.001$ Kolmogorov–Smirnov test for equality of distribution functions).

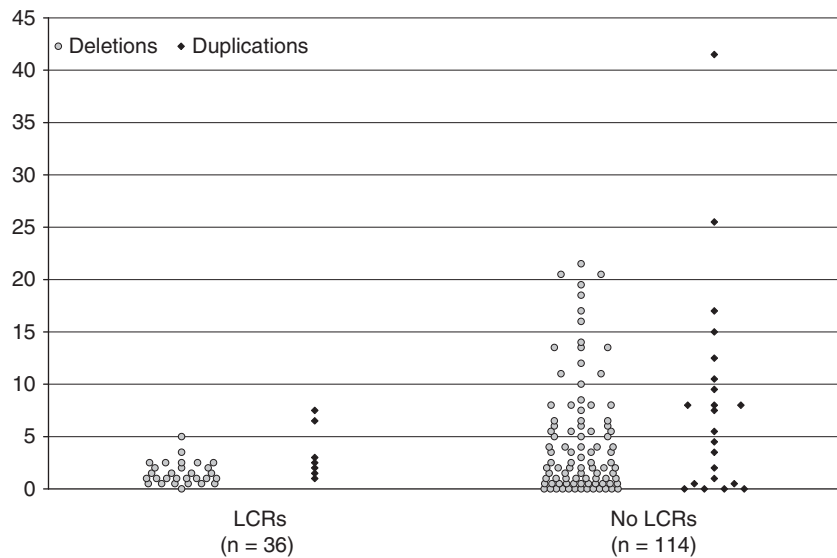


Figure 1 Size distribution of LCR and non-LCR imbalances. Plot of imbalance sizes in Mb for LCR-mediated (LCRs) and non-LCR-mediated (no LCRs) deletions and duplications. Light grey circles represent deletions and black diamonds represent duplications.

Table 1 All simple LCR-mediated and non-LCR imbalances divided into four size categories and showing the numbers for which parental origin results were obtained

| Size (Mb) | LCR mediated | | | | Non-LCR mediated | | | |
|-----------|--------------|-----|-----|----|------------------|-----|-----|----|
| | Total | Mat | Pat | ND | Total | Mat | Pat | ND |
| <1 | 7 | 1 | 0 | 6 | 30 | 3 | 14 | 13 |
| 1–5 | 26 | 9 | 13 | 4 | 38 | 9 | 16 | 13 |
| 5–10 | 3 | 2 | 0 | 1 | 25 | 5 | 14 | 6 |
| >10 | 0 | 0 | 0 | 0 | 21 | 1 | 14 | 6 |
| Total | 36 | 12 | 13 | 11 | 114 | 18 | 58 | 38 |

Abbreviations: Mat, maternal origin; Pat, paternal origin; ND, result could not be determined.

Parental origin was determined for 76 of the non-LCR imbalances (Figure 2). There was a significant excess of paternally derived cases, with 58 being paternal and only 18 maternal in origin ($P=0.024$ Fisher's exact test), and this was evident in all the four size categories (Figure 2 and Table 1). There was evidence that the size distribution differed between maternally and paternally derived cases ($P=0.03$ Kolmogorov–Smirnov test for equality of distribution functions).

Multiple imbalances

Six patients had multiple imbalances that were assumed to have arisen from independent events (Table 2). In three patients all imbalances were derived from the same parent, maternal in two cases and paternal in one, while one imbalance was derived from each parent in the remaining three patients. In total, there were eight maternal imbalances and five paternal imbalances. A slightly higher proportion of the multiple imbalances was mediated by LCRs (31%) compared with the simple imbalances (23%).

Complex imbalances

The imbalances identified in the 17 cases were believed to have been formed from the same rearrangement and consistent with this assumption, the parental origin was the same for both the imbalances in all patients. The majority of patients (71%) had imbalances of

paternal origin, a similar proportion to the single non-LCR-mediated rearrangements (Table 3).

Chromosomal origin

Chromosomal origin was determined for 6 single LCR-mediated duplications, 11 single non-LCR duplications and 6 multiple duplications (Table 4). Eight of the duplications had an interchromosomal origin and 15 had an intrachromosomal origin. The 12 cases of maternal origin were equally split between interchromosomal and intrachromosomal, while 9 of the 11 paternal cases were intrachromosomal. The majority of the LCR-mediated duplications were interchromosomal (6/8) while the majority of the non-LCR cases were intrachromosomal (13/15).

Parental age

We looked for an association between parental age and the formation of imbalances categorised by class of rearrangement, presence of LCRs, size and parental origin. Because of their small numbers, the multiple and complex imbalance groups were combined. None of the analyses of paternal imbalances reached statistical significance. Three analyses of maternal imbalances were significant for increased maternal age: simple deletions and duplication including both the LCR mediated and non-LCR ($n=30$, $P=0.026$, OR 2.20); simple deletions including LCR mediated and non-LCR ($n=23$, $P=0.037$, OR 2.36); and simple deletions and duplications mediated by LCRs ($n=11$, $P=0.010$, OR 4.00). When simple duplications were excluded, the analysis was no longer significant ($n=7$, $P=0.111$, OR 2.65).

Using the same statistical approach, we have also re-analysed the parental age data of all categories of structural abnormalities from our two papers published in 2006: 122 patients with a *de novo* microdeletion³ and 115 patients with a non-recurrent cytogenetic rearrangement.⁴ This analysis showed no significant effect of increased parental age in any of the classes of abnormality (Morris, Thomas and Jacobs, unpublished data).

DISCUSSION

Many factors contribute to the formation of *de novo* structural rearrangements and the importance of some of these factors has

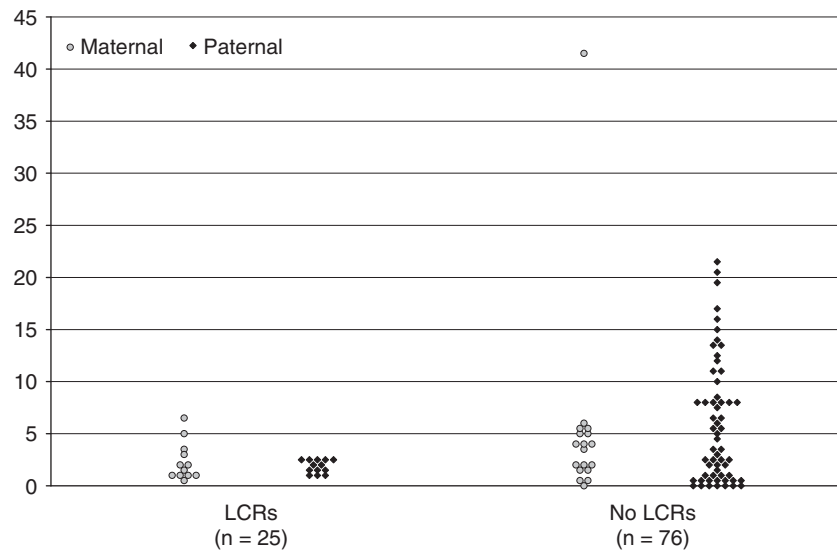


Figure 2 Size distribution of imbalances by parental origin. Plot of imbalance sizes in Mb for LCR mediated (LCRs) and non-LCR mediated (no LCRs) according to parental origin. Light grey circles represent imbalances of maternal origin and black diamonds represent imbalances of paternal origin.

Table 2 Characteristics of the six patients with multiple imbalances showing type of imbalance and chromosome, results for parental and chromosomal origin, and size and presence of LCRs for each breakpoint interval

| Case | Imbalance | Chromosome | Origin | Size (Mb) | Presence of LCRs | Chromosomal origin |
|------|-------------|------------|----------|-----------|------------------|--------------------|
| 94 | Deletion | 16 | Maternal | 0.27 | None | n/a |
| 94 | Duplication | 16 | Paternal | 2.65 | Both breakpoints | Interchromosomal |
| 127 | Deletion | 5 | Paternal | 1.72 | None | n/a |
| 127 | Deletion | 5 | Paternal | 6.94 | None | n/a |
| 141 | Deletion | 11 | Paternal | 2.63 | None | n/a |
| 141 | Deletion | 19 | Maternal | 0.75 | None | n/a |
| 83 | Deletion | 7 | Maternal | 1.44 | Both breakpoints | n/a |
| 83 | Duplication | 15 | Maternal | 0.49 | Both breakpoints | Interchromosomal |
| 83 | Duplication | 17 | Maternal | 0.15 | None | Interchromosomal |
| 102 | Duplication | 9 | Maternal | 5.79 | One breakpoint | Intrachromosomal |
| 102 | Duplication | 16 | Maternal | 0.79 | One breakpoint | Intrachromosomal |
| 93 | Deletion | 4 | Paternal | 3.45 | Both breakpoints | n/a |
| 93 | Duplication | 13 | Maternal | 8.27 | None | Intrachromosomal |

Table 3 Details of the 17 patients with complex imbalances showing a breakdown by class of rearrangement and parental origin results

| | Maternal | Paternal |
|--------------------------|----------|----------|
| Complex inversion | 1 | 4 |
| Complex insertion | 0 | 1 |
| Unbalanced translocation | 3 | 7 |
| Other complex | 1 | 0 |
| Total | 5 | 12 |

been determined experimentally. The formation of recurrent microdeletion/duplication syndromes is mediated by NAHR between LCRs, predominantly during meiosis.⁵ At most loci studied in detail, the number of maternally and paternally derived cases is approximately equal.^{2,3} Array CGH has defined many new microdeletion syndromes, and although these generally involve relatively small numbers of cases with little or no information on parental origin, the data presented in this paper suggest that these are also equally likely to be paternal or maternal in origin.

In contrast, the formation of non-recurrent chromosome imbalances appears to be much more heterogeneous. Our study of 115 cytogenetically visible cases identified a clear excess of paternal cases and, although the effect was genome wide, the proportion of paternal cases varied between genomic regions and among different classes of structural rearrangement.⁴ In all, 5 of 10 deletions and 6/11 duplications had an interchromosomal origin. Therefore at least a proportion arose during meiosis, although this study could not differentiate between LCR and non-LCR-mediated imbalances. Array CGH has greatly increased the number of non-recurrent rearrangements identified, and in particular small imbalances; however, very few studies have investigated parental origin.

We present a large study of *de novo* chromosome imbalances characterised by array CGH. Approximately 50% of the rearrangements were visible cytogenetically. Therefore, our study included imbalances with a very wide size range and our population overlaps with studies of both visible rearrangements, (eg, Thomas *et al*⁴), and more recent studies carried out using array CGH (eg, Itsara *et al*⁹). However, our study differs from nearly all those carried out using array CGH because our patient cohort comprised diagnostic referrals,

Table 4 Summary of results for all *de novo* duplications including parental and chromosomal origin

| Class | Mechanism | | Origin | Chr | Cytogenetic band(s) | Size (Mb) | Family |
|----------|--------------|-----|------------------|-----|---------------------|-----------|--------|
| Multiple | LCR-mediated | Mat | Interchromosomal | 15 | q13.31 | 0.49 | 083 |
| Multiple | LCR-mediated | Pat | Interchromosomal | 16 | p13.11–p12.3 | 2.65 | 094 |
| Multiple | Non-LCR | Mat | Interchromosomal | 17 | p13.3 | 0.15 | 083 |
| Multiple | Non-LCR | Mat | Intrachromosomal | 9 | p13.3–p13.1 | 5.79 | 102 |
| Multiple | Non-LCR | Mat | Intrachromosomal | 13 | q12.11–q12.2 | 8.27 | 093 |
| Multiple | Non-LCR | Mat | Intrachromosomal | 16 | p11.2–p11.1 | 0.79 | 102 |
| Simple | LCR-mediated | Mat | Interchromosomal | 15 | q11.2–q13.1 | 6.54 | 166 |
| Simple | LCR-mediated | Mat | Interchromosomal | 17 | q11.2 | 1.31 | 103 |
| Simple | LCR-mediated | Mat | Interchromosomal | 22 | q11.21 | 2.47 | 165 |
| Simple | LCR-mediated | Mat | Intrachromosomal | 17 | p11.2 | 3.37 | 092 |
| Simple | LCR-mediated | Pat | Interchromosomal | 3 | q29 | 1.83 | 031 |
| Simple | LCR-mediated | Pat | Intrachromosomal | 22 | q11.21 | 2.85 | 090 |
| Simple | Non-LCR | Mat | Interchromosomal | 20 | p13 | 0.33 | 053 |
| Simple | Non-LCR | Mat | Intrachromosomal | 1 | q32.1–q44 | 41.59 | 029 |
| Simple | Non-LCR | Mat | Intrachromosomal | 5 | p13.11 | 2.37 | 047 |
| Simple | Non-LCR | Pat | Intrachromosomal | 2 | q24.1–q24.3 | 12.68 | 030 |
| Simple | Non-LCR | Pat | Intrachromosomal | 10 | q23.33–q25.1 | 17.19 | 032 |
| Simple | Non-LCR | Pat | Intrachromosomal | 02 | q31.1–q24.3 | 8.39 | 046 |
| Simple | Non-LCR | Pat | Intrachromosomal | 12 | p11.23 | 1.47 | 051 |
| Simple | Non-LCR | Pat | Intrachromosomal | 21 | q22.11–q22.31 | 8.38 | 056 |
| Simple | Non-LCR | Pat | Intrachromosomal | 1 | q41–q42.2 | 7.99 | 099 |
| Simple | Non-LCR | Pat | Intrachromosomal | 13 | q32.34 | 15.15 | 132 |
| Simple | Non-LCR | Pat | Intrachromosomal | 3 | p25.3–p24.3 | 4.58 | 139 |

Full details of these duplications are given in Table 1 and Supplementary Tables 1 and 2.

and in the great majority of cases the *de novo* imbalance identified was considered to be the cause of their abnormal phenotype.

The array coordinates defined breakpoint intervals of between 5 and 250 kb for each imbalance, and we investigated these intervals for the presence of LCRs. In all, 36 of the single imbalances had LCRs at both the breakpoints, of which 34 involved regions with known microdeletion syndromes, while 114 were not mediated by LCRs. We restricted our analysis to LCRs because, although many other repetitive sequences such as LINEs and *Alus* can mediate NAHR events,^{5,10} these sequences are much smaller and more widely distributed than LCRs and would be present by chance in most regions of the genome. As the minimum length of sequence homology required for NAHR in humans is thought to be only 200 bp,¹¹ a proportion of the non-LCR rearrangements could also have been formed through NAHR. However, it should be noted that some mutation events mediated by short repetitive sequences, such as *Alus*, could also be considered as microhomology-mediated replication errors rather than as NAHR.^{12,13}

The LCR-mediated rearrangements had a narrow size distribution with 72% between 1 and 5 Mb and none larger than 10 Mb in size. In a study of *de novo* CNV rates, Itsara *et al*,⁹ also observed that LCR-mediated imbalances were rare among small imbalances but constituted the majority of imbalances above 1 Mb. For efficient NAHR to occur there may be a correlation between the LCR size and the distance separating them,^{2,14} and this may explain the size constraint on LCR-mediated imbalances observed in this study. *In silico* analysis has defined potential recombination hotspots throughout the genome with LCRs between 50 and 10 Mb apart.¹⁵ Our data suggest that *in vivo* the majority of LCR-mediated rearrangements will be between 1 and 5 Mb, although our study would not have picked up very small imbalances below the level of resolution of the array platforms used.

There were equal numbers of maternally and paternally derived LCR-mediated imbalances and their size appeared to be independent of parental origin. In contrast, there was a strong paternal bias (76%)

for non-LCR-mediated imbalances. This was most pronounced below 1 Mb and above 10 Mb, because while paternal imbalances were evenly distributed across the four size categories, maternal imbalances were mainly confined to the 1–5 Mb and 5–10 Mb classes. Therefore, the LCR-mediated and non-LCR groups have very different characteristics in terms of both size and parental origin. Interestingly, the size distribution of the LCR-mediated imbalances was very similar to that of the maternal non-LCR-mediated imbalances.

This study has confirmed the paternal bias among large *de novo* imbalances previously identified for cytogenetically visible rearrangements.⁴ For the first time, we now show that this paternal bias operates across all sizes of imbalance including those below 1 Mb. An excess of paternal cases has been reported by most, but not all, studies of *de novo* non-recurrent rearrangements. The most pronounced paternal bias was reported for reciprocal translocations that arise almost exclusively in the male germ line.⁸ A large proportion of *de novo* reciprocal translocations associated with an abnormal phenotype also have a cryptic *de novo* deletion identifiable by array CGH,^{16–19} and to date all of the *de novo* deletions examined have been paternal in origin.^{17,18,20} In contrast, no significant parental origin bias was reported in a study of *de novo* copy number variants (CNVs).⁹ Of the 47 cases for which the parental origin was determined (median size 150 kb), there were 26 paternal cases and 21 maternal cases. However, looking at a breakdown of these data in Supplementary Table 16 in the Itsara paper,⁹ among the LCR-mediated imbalances (referred to as SDs by Itsara) there were seven maternal cases and six paternal cases while among those cases equivalent to our non-LCR category, there was a slight paternal excess with 20 of paternal origin and 14 of maternal origin. Thus, although the size of imbalance was generally smaller compared with our study, the results of the Itsara paper show the same trend as our data, but on a smaller number of cases.

The complex imbalance group also had an excess of paternal cases, suggesting that there is a paternal bias across different classes of

structural rearrangement as well as over a wide size range. In contrast, there was no paternal excess for the multiple imbalance group, although this comprised only six patients and 13 imbalances. The proportion of LCR-mediated cases (4/13) was similar to the single imbalances, although the proportion of duplications (6/13) was higher. Thus these small data give no suggestion of a 'predisposition' to the production of multiple independent imbalances, although this would have to be confirmed in a much larger series of patients.

Two lines of evidence can be used to identify rearrangements which are likely to have arisen by NAHR during meiosis: the presence of LCRs at both the breakpoints and for duplications an interchromosomal origin. Although they are not synonymous, they show good concordance in our data: 6/8 LCR-mediated rearrangements had an interchromosomal origin while only 2/15 rearrangements without LCRs were interchromosomal. The high proportion of intrachromosomal non-LCR duplications suggests that mitotic mechanisms, such as FoSTeS and MMBIR, could potentially be important in the formation of this class of structural rearrangement. Combining the 23 chromosomal origins from this study with the 21 from our 2006 study,⁴ there is a slightly higher rate of interchromosomal origin for maternal cases (10/17, 59%) compared with the paternal cases (9/27, 33%).

Male gametogenesis may be particularly susceptible to the formation of chromosome rearrangements, because there are far more divisions in spermatogenesis than oogenesis.²¹ Templado *et al*²² found a higher rate of structural abnormalities in the sperm of older males (6.6%) compared with younger males (4.9%). We have previously shown that non-recurrent reciprocal translocations have a strong paternal bias and are associated with a significant increase in paternal age.⁸ However, using the same statistical approach we found no association with increased paternal age in this study even allowing for size, class of imbalance and whether or not they were formed by the NAHR. Analysis of maternal cases was significant for the very small number of LCR-mediated duplications and we assume this result to be an artefact. Repeating the same analysis on our much larger microdeletion data set from 2006 was also not significant.

For the first time this study has identified differences in the characteristics of maternally and paternally derived imbalances. Paternal imbalances are evenly distributed throughout the four size groups; in contrast there are few very small or very large maternal imbalances. Among the maternal imbalances in this study, 40% are LCR-mediated (12/30) compared with only 18% (13/71) of the paternal imbalances. Although maternal imbalances are equally likely to have an interchromosomal or an intrachromosomal origin, paternal imbalances are predominantly intrachromosomal. Interestingly, maternal imbalances showed a size distribution pattern characteristic of imbalances formed by LCR-mediated NAHR. NAHR appears to be a common mechanism that works equally well in both male and female meiosis. We hypothesise that in females all or nearly all *de novo* imbalances are the result of NAHR, but that in males NAHR accounts for only a minority of paternal imbalances. Instead there are additional male-specific mechanisms, probably mitotic, which contribute to the formation of most structural chromosome imbalances in males. This hypothesis can be tested by greatly increasing the resolution of breakpoint interval mapping to look for the presence of smaller repetitive sequence elements, which could mediate NAHR.

In summary, we show that imbalances generated by NAHR have different characteristics compared with those generated by other

mechanisms, and this study has also identified preliminary evidence that there are differences in the mechanisms through which maternal and paternal imbalances arise. The results of this study represent a significant increase in our understanding of how chromosome imbalances arise, and now require replication in larger series of *de novo* imbalances.

CONFLICT OF INTEREST

The authors declare no conflict of interest.

ACKNOWLEDGEMENTS

This study was partly funded by the Salisbury Hospital Foundation. This study was carried out with ethical approval from the Southampton and West Hampshire Research Ethics Committee.

- Lopes J, Vandenbergh A, Tardieu S *et al*: Sex-dependent rearrangements resulting in CMT1A and HNPP. *Nat Genet* 1997; **17**: 136–137.
- Baumer A, Dutly F, Balmer D *et al*: High level of unequal meiotic crossovers at the origin of the 22q11.2 and 7q11.23 deletions. *Hum Mol Genet* 1998; **7**: 887–894.
- Thomas NS, Durkie M, Potts G: Parental and chromosomal origins of microdeletion and duplication syndromes involving 7q11.23, 15q11-q13 and 22q11. *Eur J Hum Genet* 2006b; **14**: 831–837.
- Thomas NS, Durkie M, Van Zyl B *et al*: Parental and chromosomal origin of unbalanced *de novo* structural chromosome abnormalities in man. *Hum Genet* 2006a; **119**: 444–450.
- Gu W, Zhang F, Lupski JR: Review: mechanisms for human genomic rearrangements. *PathoGenetics* 2008; **1**: 4.
- Lee JA, Carvalho CM, Lupski JR: A DNA replication mechanism for generating nonrecurrent rearrangements associated with genomic disorders. *Cell* 2007; **131**: 1235–1247.
- Hastings PJ, Ira G, Lupski JR: A microhomology-mediated break-induced replication model for the origin of human copy number variation. *PLoS Genet* 2009a; **5**: e1000327.
- Thomas NS, Morris JK, Baptista J: *De novo* apparently balanced translocations in man are predominantly paternal in origin and associated with a significant increase in paternal age. *J Med Genet* 2010; **47**: 112–115.
- Itsara A, Wu H, Smith JD *et al*: *De novo* rates and selection of large copy number variation. *Genome Res* 2010; **11**: 1469–1481.
- Shaw CJ, Lupski JR: Non recurrent 17p11.2 deletions are generated by homologous and non-homologous mechanisms. *Hum Genet* 2005; **116**: 1–7.
- Inoue K, Lupski JR: Molecular mechanisms for genomic disorders. *Annu Rev Genomics Hum Genet* 2002; **3**: 199–242.
- Zhang F, Seeman P, Liu P *et al*: Mechanisms for nonrecurrent genomic rearrangements associated with CMT1A and HNPP: rare CNVs as a cause for missing heritability. *Am J Hum Genet* 2010; **86**: 892–903.
- Vissers LELM, Bhatt SS, Janssen IM *et al*: Rare pathogenic microdeletions and tandem duplications are microhomology-mediated and stimulated by local genome architecture. *Hum Mol Genet* 2009; **18**: 3579–3593.
- Lupski JR: Genomic disorders: structural features of the genome can lead to DNA rearrangements and human disease traits. *Trends Genet* 1998; **14**: 417–422.
- Sharp AJ, Hansen S, Selzer RR: Discovery of previously unidentified genomic disorders from the duplication architecture of the human genome. *Am J Hum Genet* 2006; **38**: 1038–1042.
- Ciccone R, Giorda R, Gregato G *et al*: Reciprocal translocations: a trap for cytogeneticists? *Hum Genet* 2005; **117**: 571–582.
- Baptista J, Mercer C, Prigmore E *et al*: Breakpoint mapping and array CGH in translocations: comparison of a phenotypically normal and an abnormal cohort. *Am J Hum Genet* 2008; **82**: 1–10.
- De Gregori M, Ciccone R, Magini P *et al*: Cryptic deletions are a common finding in "balanced" reciprocal and complex chromosome rearrangements: a study of 59 patients. *J Med Genet* 2007; **44**: 750–762.
- Schluth-Bolard C, Delobel B, Sanlaville D *et al*: Cryptic genomic imbalances in *de novo* and inherited apparently balanced chromosomal rearrangements: array CGH study of 47 unrelated cases. *Eur J Med Genet* 2009; **52**: 291–296.
- Gribble SM, Prigmore E, Burford DC *et al*: The complex nature of constitutional *de novo* apparently balanced translocations in patients presenting with abnormal phenotypes. *J Med Genet* 2005; **42**: 8–16.
- Crow JF: The origins, patterns and implications of human spontaneous mutation. *Nature* 2000; **1**: 40–47.
- Templado C, Donate A, Giraldo J: Advanced age increases chromosome structural abnormalities in human spermatozoa. *Eur J Hum Genet* 2011; **19**: 145–151.

Supplementary Information accompanies the paper on European Journal of Human Genetics website (<http://www.nature.com/ejhg>)