

# SCIENTIFIC REPORTS



OPEN

## A clear bias in parental origin of *de novo* pathogenic CNVs related to intellectual disability, developmental delay and multiple congenital anomalies

Received: 25 October 2016  
Accepted: 08 February 2017  
Published: 21 March 2017

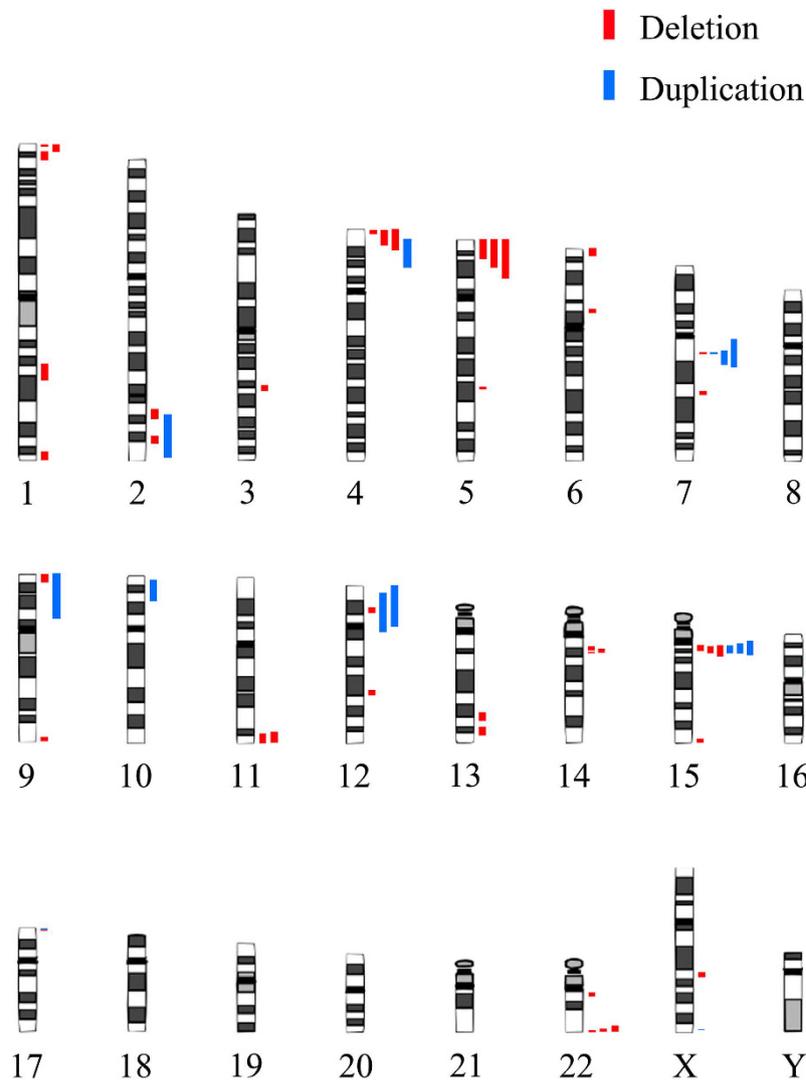
Ruiyu Ma, Linbei Deng, Yan Xia, Xianda Wei, Yingxi Cao, Ruolan Guo, Rui Zhang, Jing Guo, Desheng Liang & Lingqian Wu

Copy number variation (CNV) is of great significance in human evolution and disorders. Through tracing the parent-of-origin of *de novo* pathogenic CNVs, we are expected to investigate the relative contributions of germline genomic stability on reproductive health. In our study, short tandem repeat (STR) and single nucleotide polymorphism (SNP) were used to determine the parent-of-origin of 87 *de novo* pathogenic CNVs found in unrelated patients with intellectual disability (ID), developmental delay (DD) and multiple congenital anomalies (MCA). The results shown that there was a significant difference on the distribution of the parent-of-origin for different CNVs types (Chi-square test,  $p = 4.914 \times 10^{-3}$ ). An apparently paternal bias existed in deletion CNVs and a maternal bias in duplication CNVs, indicating that the relative contribution of paternal germline variations is greater than that of maternal to the origin of deletions, and vice versa to the origin of duplications. By analyzing the sequences flanking the breakpoints, we also confirmed that non-allelic homologous recombination (NAHR) served as the major mechanism for the formation of recurrent CNVs whereas non-SDs-based mechanisms played a part in generating rare non-recurrent CNVs and might relate to the paternal germline bias in deletion CNVs.

Germ cell, as the carrier of genetic information, is essential for human reproductive fitness and the development of offsprings. The health of germ cells is reflected not only in the morphology, structure and physiological function, but also in the integrity and stability of the genome they carry, which is indispensable for the constant transfer of genetic information. As a fact, during the whole life of germ cells, a wide variety of genomic variations will emerge, possibly affecting the stability of the genome. In terms of hominid evolution, these variations, including base mutations, variations of microsatellites and structural variations (SVs), play an important role in the biological diversity and the genome variability among individuals<sup>1-3</sup>. On the other hand, they are also recognized as significant contributors to the occurrence of human diseases.

CNVs, which are known as deletions or duplications of genomic materials that are larger than 1 kb in size in comparison with the reference genome<sup>4</sup>, have become a popular hotspot of the genomic SVs in the past decade. Many researches have reported about the close relationship between CNVs and complex genomic disorders such as ID, autism, schizophrenia and numerous developmental diseases<sup>5-8</sup>. And most of the known pathogenic and high-risk variants are *de novo*. SNP, with allele frequency at least 1%, is also one important type of variations throughout the genome. Comparatively, because of the larger size and more genetic materials involved, CNVs naturally will be more complicated. Nonetheless, whether *de novo* CNVs are more significant and abundant than SNPs in disease causation has already become a wide debate in recent years. Previous studies had emphasized that the proportion of nucleotide content in CNVs is far higher than that of SNPs per genome<sup>9-10</sup>. S.W. Scherer *et al.*<sup>11</sup> found that CNVs accounted for 1.28% of genetic variants between the first single individual genome (J Craig Venter's DNA named HuRef)<sup>12</sup> and the National Center for Biotechnology Information (NCBI) reference

State Key Laboratory of Medical Genetics, Central South University, Changsha, Hunan 410078, P.R. China. Correspondence and requests for materials should be addressed to D.L. (email: liangdesheng@sklmg.edu.cn) or L.W. (email: wulingqian@sklmg.edu.cn)



**Figure 1.** The distribution of 87 *de novo* CNVs on human chromosomes. Red bars represent deletions and blue bars represent duplications.

genome assembly, whereas SNPs encompassed only 0.1%. All these underscored one significant source of genetic variations in germ cell genome—CNVs.

Nowadays, it is widely shared that CNVs derived from a variety of ways during the formation of germ cells, such as DNA repair, replication errors, homologous recombination and chromosome separation errors. The generation mechanism underlying these pathogenic CNVs has been studied extensively. Yet, the relationship between the genesis of these CNVs and their parental origin is still rarely known. Here, we launched a research on the parent-of-origin of *de novo* pathogenic CNVs found in Chinese patients with ID, DD and MCA. Through tracing the origins and genesis mechanisms of these CNVs and exploring the relative contributions of the genome stability of sperms and eggs to reproductive health, we hope to know more about genome structure variations in germ cells, enhance the effective differential diagnosis and prenatal diagnosis of genomic disorders and to provide the potential for the prevention and treatment of related diseases in the future.

## Results

**CNVs analysis and selection.** Using SNP arrays and fluorescence *in situ* hybridization (FISH), 87 *de novo* suspicious pathogenic CNVs, including 69 deletions and 18 duplications, were identified in total in 87 unrelated patients with ID, DD and MCA. Ages of the probands span from 6 days to 34 years old. 65 patients carried CNVs overlapping with known chromosome syndromes like Wolf-Hirschhorn syndrome (WHS), Williams Beuren syndrome (WBS) and Angelman/Prader Willi syndrome (AS/PWS). The remaining 22 patients carried rare *de novo* CNVs, such as deletion area 12p12.2-p12.1, 13q31.3-q32.1 and duplication area 9p24.3-p13.1, 10p15.3-p12.33. These regions have not been defined as a syndrome or disease-related area presently. The comprehensive information of 87 families is summarized in Supplementary Table S1 and the distribution of the CNVs on chromosome is illustrated in Fig. 1. The genome sequence coordinates based on GRCh37/hg19 assembly.

Paternal hemizygous deletion				Maternal hemizygous deletion			
Patient ID	Patient	Father	Mother	Patient ID	Patient	Father	Mother
MD1459	<b>187</b>	189/191	185/ <b>187</b>	MD1908	<b>92</b>	89/ <b>92</b>	79/79
MD1506	<b>410</b>	393/430	<b>410</b> /418	MD2484	<b>100</b>	92/ <b>100</b>	92/92
MD1672	<b>199</b>	191/207	183/ <b>199</b>	MD2586	<b>136</b>	<b>136</b> /140	140/143
MD1975	<b>238</b>	221/226	226/ <b>238</b>	MD3164	<b>116</b>	105/ <b>116</b>	105/118
MD2389	<b>270</b>	268/272	<b>270</b> /272	MD3778	<b>C</b>	<b>C/T</b>	T/T
MD2750	<b>100</b>	117/119	<b>100</b> /117	MD4612	<b>144</b>	<b>144</b> /160	152/168
MD2994	<b>146</b>	150/150	<b>146</b> /150	MD5599	<b>283</b>	268/ <b>283</b>	259/268
MD4350	<b>120</b>	136/136	115/ <b>120</b>	MD6474	<b>170</b>	167/ <b>170</b>	167/167
MD4748	<b>205</b>	209/211	<b>205</b> /207	MD6894	<b>143</b>	<b>143</b> /147	150/159
MD5621	<b>151</b>	148/155	<b>151</b> /155	MD8177	<b>A</b>	<b>A/A</b>	C/C
MD6227	<b>204</b>	197/206	<b>204</b> /212	MD8557	<b>262</b>	<b>262</b> /300	254/303
MD6723	<b>222</b>	216/224	216/ <b>222</b>	MD8877	<b>115</b>	101/ <b>115</b>	101/101
MD7132	<b>124</b>	120/130	<b>124</b> /126	MD9792	<b>209</b>	195/ <b>209</b>	205/205
MD7454	<b>287</b>	283/295	275/ <b>287</b>	MD10290	<b>155</b>	155/159	143/151
MD8160	<b>A</b>	G/G	<b>A/A</b>	MD10369	<b>283</b>	<b>283</b> /287	280/287
MD8417	<b>254</b>	226/256	<b>254</b> /267	MD11049	<b>169</b>	<b>169</b> /175	164/167
MD9127	<b>A</b>	G/G	<b>A/G</b>	MD11905	<b>285</b>	262/ <b>285</b>	262/297
MD9478	<b>146</b>	132/144	132/ <b>146</b>	MD12173	<b>155</b>	151/ <b>155</b>	151/159
MD11042	<b>163</b>	161/161	161/ <b>163</b>	MD12563	<b>293</b>	262/ <b>293</b>	266/289
MD11193	<b>131</b>	129/137	<b>131</b> /131	MD12665	<b>95</b>	89/ <b>95</b>	91/93
MD11326	<b>140</b>	123/137	136/ <b>140</b>	MD13120	<b>212</b>	208/ <b>212</b>	208/210
MD11345	<b>G</b>	C/C	<b>G/C</b>	MD13380	<b>146</b>	<b>146</b> /150	142/150
MD11473	<b>G</b>	C/C	<b>G/G</b>	MD14271	<b>95</b>	<b>95</b> /97	92/92
MD11966	<b>A</b>	C/C	<b>A/C</b>	MD14442	<b>A</b>	<b>A/C</b>	C/C
MD11973	<b>C</b>	A/A	<b>A/C</b>	MD14816	<b>267</b>	263/ <b>267</b>	259/271
MD12285	<b>A</b>	G/G	<b>A/G</b>	MD15823	<b>262</b>	258/ <b>262</b>	258/266
MD12779	<b>C</b>	A	<b>A/C</b>	MD16014	<b>132</b>	129/ <b>132</b>	130/134
MD12807	<b>226</b>	248/261	<b>226</b> /230	MD16085	<b>G</b>	<b>G/G</b>	C/C
MD12812	<b>266</b>	242/262	<b>266</b> /266				
MD13097	<b>272</b>	266/274	262/ <b>272</b>				
MD13108	<b>150</b>	158/164	<b>150</b> /164				
MD13636	<b>166</b>	154/158	164/ <b>166</b>				
MD14047	<b>147</b>	133/139	133/ <b>147</b>				
MD14491	<b>139</b>	137/143	<b>139</b> /141				
MD14517	<b>A</b>	G/G	<b>A/G</b>				
MD14650	<b>121</b>	114/119	<b>121</b> /125				
MD15641	<b>144</b>	138/138	138/ <b>144</b>				
MD15761	<b>G</b>	A/A	<b>A/G</b>				
MD16107	<b>127</b>	122/122	113/ <b>127</b>				
MD16204	<b>A</b>	G/G	<b>A/G</b>				
MD16619	<b>C</b>	A/A	<b>A/C</b>				

**Table 1. STR fragment size and SNP genotyping results of *de novo* deletion CNVs.** The genotyping data from one of informative markers was listed. The bold font indicates the same STR/SNP locus between patients and their parents.

**Parent-of-origin of *de novo* pathogenic CNVs.** Analysis of parental origin was performed on *de novo* pathogenic CNVs mentioned above. By means of STR profiling (69 cases) and SNP genotyping (18 cases) techniques, the parental origin of all the 87 cases were successfully determined. Among the 69 cases with deletions 41 had arisen paternally and 28 maternally (Table 1). Four types were observed in 18 cases with duplications, 10 cases were heterologous maternal allele duplication, 4 cases were homologous maternal allele duplication and 2 cases were heterologous or homologous paternal allele duplication respectively (Table 2). Two-tailed  $\chi^2$  tests on fourfold contingency table were carried out for statistical analyses.

**Analysis of genome sequence nearby breakpoints.** DNA sequences at all 87 CNVs breakpoints were preliminary analyzed in our research. According to the distribution of segmental duplications (SDs)/low-copy repeats (LCRs) at the fractured regions, CNVs were divided into the following three types: (1) SDs/LCRs existed

Homologous maternal allele duplication				Heterologous maternal allele duplication			
Patient ID	Patient	Father	Mother	Patient ID	Patient	Father	Mother
MD11928	A/C/C	A/A	A/C	MD1643	<b>268/275/301</b>	260/301	<b>268/275</b>
MD12320	270/ <b>289/289</b>	270/285	262/ <b>289</b>	MD4267	<b>185/187/195</b>	193/195	<b>185/187</b>
MD13935*	G/G	A	A/G	MD4278	<b>130/132/134</b>	122/134	<b>130/132</b>
MD15100	<b>181/181/190</b>	185/190	<b>181/181</b>	MD4477	<b>162/167/173</b>	167/179	<b>162/173</b>
				MD7381	<b>254/258/262</b>	262/266	<b>254/258</b>
				MD7484	<b>135/147/151</b>	151/155	<b>135/147</b>
				MD8152	<b>228/247/264</b>	256/264	<b>228/247</b>
				MD13015	<b>107/117/121</b>	107/111	<b>117/121</b>
				MD15558	<b>259/263/298</b>	259/282	<b>263/298</b>
				MD15714	<b>134/136/138</b>	136/148	<b>134/138</b>
Homologous paternal allele duplication				Heterologous paternal allele duplication			
Patient ID	Patient	Father	Mother	Patient ID	Patient	Father	Mother
MD10790	161/ <b>186/186</b>	174/ <b>186</b>	161/194	MD4571	<b>106/115/140</b>	<b>106/115</b>	125/140
MD15000	205/ <b>225/225</b>	220/225	205/234	MD8294	<b>165/182/194</b>	<b>165/182</b>	178/194

**Table 2. STR fragment size and SNP genotyping results of *de novo* duplication CNVs.** The genotyping data from one of informative markers was listed. The bold font indicates the same STR/SNP locus between patients and their parents. \*The patient is a male with a 0.5 Mb duplication on Xq28. The SNP loci within the region of Xq28 showed homozygous G from the mother which indicated a maternal allele duplication.

in both ends of the CNV; (2) SDs/LCRs existed in one end of the CNV; (3) None of SDs/LCRs existed in either end of the CNV. Results of classification are shown in Supplementary Table S1.

## Discussion

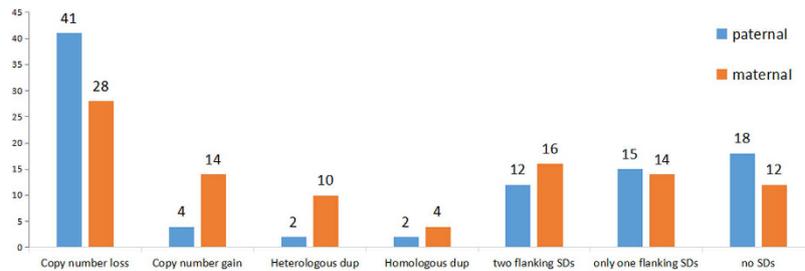
As mentioned above, variations of germ line cells may arise in many stages during embryonic development like premeiotic, meiotic, or postzygotic cells. Whenever occurred, these variations would only involve germ line cells, without affecting somatic cells in the parents. Therefore, when analyzing blood samples from parents, we would not find them carrying the same pathogenic CNV as the patient. Such CNVs are classified as *de novo* in our research. Until now, systematic researches on parent-of-origin of *de novo* pathogenic CNVs were relatively few. In 2006, Thomas NS published their results on 115 *de novo* unbalanced abnormalities and indicated a pronounced paternal bias. The author thought the bias varied by chromosome and recombination patterns<sup>13</sup>. Later, Hehir-Kwa JY and colleagues<sup>14</sup> also reported a paternal bias in 118 autosomal *de novo* CNVs discovered in 108 ID patients. The difference between them is that the latter declared the bias was independent of CNV type and length. Unlike these findings, the research from Itsara A, *et al.*<sup>15</sup> showed no evidence for a parent-of-origin preference in a relatively small sample (47 *de novo* CNVs). For some familiar recurrent genomic disorders, Thomas NS, *et al.*<sup>16</sup> determined the parental and chromosomal origins of several groups of patients involving 7q11.23, 15q11-q13 and 22q11, and no significant parental bias was observed except for duplications of 15q11-q13, which represented a maternal bias. And they also found a slight overall excess (just reach statistical significance,  $p = 0.05$ ) of maternal bias for 22q11 deletions when combining their data with those from previous studies. This bias was then replicated and confirmed by Delio *et al.*<sup>17</sup> which a total of 810 combined results revealed a highly significant maternal bias occurring in the origin of the *de novo* 22q11.2 deletion. Besides, some other specific micro-deletions/duplications have also been reported with a discrepant proportion of maternal or paternal origins, such as *NF1* region with an excess of maternal deletions<sup>18</sup>, the ~550 kbp autism-associated region on 16p11.2 with a maternal bias for both *de novo* deletions and duplications<sup>19</sup>, Sotos syndrome and 22q13 deletion syndrome with an excess of paternal deletions<sup>20–23</sup>.

In our study, the parent-of-origin of all the 87 *de novo* pathogenic CNVs were analyzed. Among them 45 (51.7%) derived from the paternal germ line and 42 (48.3%) from the maternal germ line, with no obvious bias (two-tailed binomial test,  $p = 0.83$ ). However, we found a clear bias when analyzing the parental origin in different CNV types. The cases of hemizygous deletions were found to be paternal in 59.4% (41/69) and maternal in 40.6% (28/69). Meanwhile, the duplications were found to be paternal in 22.2% (4/18) and maternal up to 77.8% (14/18) (Table 3, Fig. 2). The value of chi-square test ( $\chi^2 = 7.911$ ,  $p = 4.914 \times 10^{-3}$ ) suggested that the distribution of the parent-of-origin varied significantly in different CNVs types. Combining with these data, we inferred that a paternal germline bias existed in deletions, and an evident excess of maternal origin existed in the type of duplications. For the reasons that deletions are much more likely than duplications to result in moderate-to-severe ID and major congenital anomalies, and duplications are more frequently associated with incomplete penetrance and highly variable phenotypes, it is difficult to obtain an adequate sample for duplications comparing with deletions (only 18 in our study) in this regard. For the sample size of 87, we had 95% power to detect a difference of 0.20 from the null hypothesis of 0.5. We need to point out that our observations are not a “general trend” in all *de novo* pathogenic CNVs on account of sampling bias. Our data and discussion are restricted to patients with ID, DD and MCA mentioned in Table S1.

From the perspective of the entire development process of human germline, this sex differences in parent-of-origin can be largely explained by the relative greater numbers of DNA replication and cell divisions

	Parent-of-origin		Total number	Two-Tailed p value*
	Paternal	Maternal		
Deletions	41	28	69	$4.914 \times 10^{-3}$
Duplications	4	14	18	
Heterologous	2	10	12	
Homologous	2	4	6	
Total number of CNVs	45	42	87	

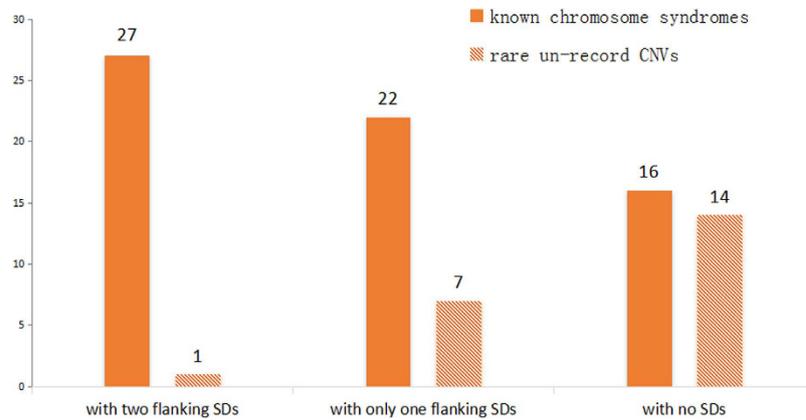
**Table 3.** The distribution of the parent-of-origin in different CNVs types. \*Chi-square test using IBM SPSS Statistics 20.0.



**Figure 2.** A multiple comparison between paternal CNVs and maternal CNVs. A clear difference was found when comparing the number of *de novo* CNVs according to CNV types. Paternal CNVs were noticeably more in deletions, and maternal were dominant in duplications especially in heterologous group.

in male than female<sup>24</sup>. During a male's life, self-renewing spermatogonial stem cells continuously produce new mature sperms, while female genetic material in oogonia is already fixed before birth and will stay in early days of meiosis I for a long time until adolescence. Considering that all of 87 patients in our study were sporadic with the absence of relevant family history, and that mosaicism is rare, it can be speculated that the CNVs should mostly occur in the period of meiosis. For 87 *de novo* pathogenic CNVs with clearly parental origin, we attempted to infer their formation times. In meiosis I, primary spermatocytes and oocytes finished homologous recombination<sup>25</sup>. The unequal exchange between non-sister chromatids could produce deletion and heterologous allele duplication. The interchromatid breakage and exchange during the second meiosis would also produce deletion but with homologous allele duplication. So, it can be concluded that the CNV type of allele hemizygous deletion can form at any time in meiosis, heterologous allele duplications should form before the separation of homologous chromosomes in meiosis I, while homologous allele duplications might form in the interval between the separation of homologous chromosomes and the separation of sister chromatids in meiosis II. Our previous study on a patient with 2q partial trisomy presenting an inverted duplicated neocentric marker chromosome also reached consistent conclusion<sup>26</sup>.

Therefore, based on the replication-dependent variation in human, it is not difficult to understand the obvious paternal bias in deletions. However, if replication-based mechanisms are responsible for most *de novo* CNV formations, a paternal bias might also be observed for duplications. The maternal bias we found in duplications did not seem in line with this theory. For further details, we consulted previous publications and found there were still similar findings with a maternal bias for duplications in some specific micro-duplication syndromes like 15q11-q13 duplication syndrome<sup>16</sup> and duplications in 16p11.2<sup>19</sup>. Furthermore, in our study, we also found difference in the degree of maternal bias between heterologous and homologous duplications, with more maternal CNVs in the heterologous duplications (Table 2). The reason would be that, as mentioned, heterologous allele duplications happened mainly in early period of meiosis I yet that is the longest residence time in the life of oocytes. One more thing should be cautioned is that our observation stands in conflict with some previous publications mentioned above which have noted maternal biases in specific micro-deletion syndromes. In two of these studies, the skewing of recombination rates between female and male was suggested to be the most likely explanation for the maternal biases in 22q11.2 and 16p11.2 regions<sup>17,19</sup>. In order to explore whether the maternal bias we found in duplications was affected by a higher female recombination rate, we examined the data from published recombination maps<sup>27-29</sup> and calculated mean recombination rates (cM/Mb) of females and males from three studies (Marshfield, deCODE, and Genethon) in each duplication region. Higher mean female recombination rates were found in 10 (55.5%) duplications. Compared with the increase of 27.8% in detected maternal duplications, the increase of 5.5% in female-to-male recombination rates was only slightly higher for these regions. Therefore, the recombination rate could not entirely explain the observed bias in our study. Hypothesis of male-driven molecular evolution proposed by Miyata *et al.*<sup>30-32</sup> claimed that the male germline mutations served as the major factor motivating the human molecular evolution. Nevertheless, different mechanisms participate disproportionately to various types of structural variations. Many other mechanisms may also have an influence on chromosome unbalanced recombination and the parental origin of *de novo* CNVs. The distinction at the type



**Figure 3.** The existence of flanking SDs nearby the CNV breakpoints.

of recombination, the specific loci on chromosomes involved, synapsis and errors repair processes during meiotic, the severity of the phenotype, genomic imprinting, recombination rates and gender-specific recombination hotspots between males and females are all likely to produce bias on this kind of research. It is noteworthy that our study also shed light on the probable bias of paternal origin in WHS and maternal origin in 1p36 deletion syndrome and 15q11-q13 duplication syndrome, but each is small in size. Overall, current researches on the parental origin of CNVs are rare and based on different race and diseases, and even the results from various studies are inconsistent. Considering many other inevitable factors, an exact consensus remains further exploration.

To clarify the possible formation mechanisms of the CNVs detected, we intercepted and checked the sequence at  $\pm 10$  kb of both proximal and distal breakpoints, and found 57 of 87 CNVs having at least one breakpoint directly flanked by SDs. Combining the data from our study, the size of *de novo* suspicious pathogenic CNVs ranged from 0.5 Mb to 42.2 Mb, most of which were chromosome syndromes that had been well-known. Half of these syndromes often share with the similar breakpoints and rearranged fragments in unrelated individuals, called recurrent genomic disorders, such as 22q11.2 deletion syndrome, WBS and AS/PWS. As there are the same contiguous genes existed in deletion or duplication intervals, it is easier to analyze the genotype-phenotype correlations<sup>33</sup>. As is well known, recurrent rearrangements are mainly mediated by non-allelic homologous recombination (NAHR), the first mechanism recognized in the human genome abnormality<sup>34</sup>. Yet, as a common genomic architecture for these rearrangements, SDs/LCRs usually enriched near the breakpoints. In our study, the breakpoints of 49 (86%) CNVs causing well-known chromosome syndromes were clustered inside SDs, while only 8 (14%) of rare non-record *de novo* CNVs. And it can be seen from Fig. 3 that for those non-record suspicious pathogenic CNVs, nearly 2 out of 3 were not flanked by SDs in both breakpoints, and the remaining showed characteristic SDs only at one breakpoint except for one case. Besides, SDs were not attendant upon a small amount of known syndromes. Interestingly, we noticed that the breakpoints of these syndromes are scattered in the genome, some even with no common region of overlap. It just coincided well with the concept of the non-recurrent rearrangements. The absence of SDs suggested that the generation mechanism of these rearrangements was probably not NAHR but some of others like non-homologous endjoining (NHEJ), fork stalling and template switching (FoSTeS) or microhomology-mediated break-induced replication (MMBIR), in which SDs were not an indispensable requirement<sup>35</sup>. In portion of non-recurrent rearrangements, some shorter and highly homologous repetitive sequences such as long interspersed nuclear elements (LINEs), long terminal repeats (LTRs) and *Alus* are also functioning as substrates for NAHR as well<sup>35,36</sup>, which could explain why we found quite a number of CNVs in our study only be flanked by one SD rather than two. To our knowledge, several disorders have been proved to be non-recurrent events occurred via NHEJ, FoSTeS or MMBIR, including Miller-Dieker lissencephaly syndrome<sup>37</sup>, 17p13.3 duplication syndrome<sup>38</sup>, Pelizaeus-Merzbacher disease (PMD)<sup>39</sup> and MECP2 duplication syndrome<sup>40</sup>. Our analysis of the sequence surrounding the breakpoints regarding these disorders also reached the same conclusion. In addition, our results of SDs/LCRs analysis nearby CNVs in different parental origins (Fig. 2) revealed that the paternal origin bias we found in deletion CNVs might have relations with mechanisms of NHEJ, FoSTeS or MMBIR, because of the higher proportion of paternal *de novo* CNVs with no SDs around. And this is also corresponding with our earlier discussion on paternal bias as these mechanisms are all replication-dependent. Of course, our analysis on sequence homology near the CNVs breakpoints was a rough estimate. The specific relationship between pathogenesis and formation mechanism of CNVs remains to be explored with precise breakpoint position mapping and related functional studies.

In summary, this is the first research on the parent-of-origin of *de novo* pathogenic CNVs of the Chinese population with ID, DD and MCA. We herein provide a clue on the significant role of human reproductive cell in the generation of *de novo* genome SVs, prompting that deletion CNVs are ordinarily paternal in origin and the relative contribution of paternal germline variations is greater in the origin of deletion CNVs, while maternal germline variations seem to be more important to duplication CNVs, especially the heterologous duplications. The analyses of the sequences flanking the breakpoints confirm that majority of recurrent CNVs arise by NAHR which relies on SDs. And non-SDs-based replication-dependent mechanisms serve as major characters in rare non-recurrent CNVs and may relate to the paternal germline bias in deletion CNVs. However, it is worth noting that our findings are not entirely consistent with what have been observed in previous literature. These

discrepancies may be due to the differences in race, the composition of cohort, the design of experiment and other inevitable factors. With the rapid development of deep sequencing and bioinformatic analyses on a global genome scale, there is no doubt that we will learn more about human genomic disorders in the future and do better in the prediction and prevention of birth defects in newborns.

## Materials and Methods

**Patient samples.** Clinical genetic testing for more than 1000 probands with ID, DD and MCA were completed from 2009 to 2015 at the State Key Laboratory of Medical Genetics, Central South University (Hunan, China) and Hunan Jiahui Genetics Hospital (Hunan, China). Using SNP arrays and FISH, 240 patients were found to have suspicious pathogenic CNVs. Excluding families that the parental samples were not available, a total of 87 *de novo* suspicious pathogenic CNVs identified in 87 unrelated patients in the cohort above were included in our study. DNA samples from both parents were available in all the 87 families. For each family, an informed consent was signed before testing, and then 5 ml\*2 peripheral blood (EDTA anticoagulant and heparin anticoagulant respectively) were collected from the patient and the parents. The study was approved by the Ethics Committee of the state key laboratory of medical genetics of China.

**SNP array analysis.** Genomic DNA of the family members was prepared from fresh peripheral blood lymphocytes using the standard phenol-chloroform method. SNP array analysis of the extracted DNA was performed using the Human OmniZhongHua-8 Beadchip or HumanCytoSNP-12 BeadChip array (Illumina Inc., San Diego, CA) following the manufacturer's recommended protocols ([www.illumina.com](http://www.illumina.com)). The two kinds of microarrays contain about 900,000 and 300,000 probes respectively. CNVs were detected with the *cnvPartition* algorithm and the data including B allele frequency and log R ratio were calculated using *GenomeStudio* software (*cnvPartition* Plugin v3.1.6). We picked out CNVs with a call rate >0.99 and a confidence >35 for further analyses.

**CNVs identification and analysis.** FISH analysis was performed on metaphase or interphase chromosomes from the patients and their parents to determine the authenticity and origin of CNVs. For each CNV, two BAC clone probes with different fluorescence were used. One located within the deletion or duplication region as a test probe, and another located within normal region in the same chromosome as a control probe. Alternatively, for small CNVs which were less than 1 Mb and hardly identified by regular FISH analysis, SNP Array analysis of the parents' DNA was utilized. Judgment of the pathogenicity of *de novo* CNVs has always been an intractable thing for clinical staffs and geneticist. Following the recommendations for systematic evaluation and clinical interpretation of CNVs from Miller *et al.* and American College of Medical Genetics (ACMG)<sup>41,42</sup>, we considered about various characteristics including CNV size, origin, inheritance mode and genomic content (including genes, exons and regulatory elements) in the CNV interval, and referenced latest published literature and publicly external databases such as OMIM (<http://omim.org/>), DECIPHER (<https://decipher.sanger.ac.uk/>), ISCA (<http://dbsearch.clinicalgenome.org/search/>) and DGV (<http://dgv.tcag.ca/dgv/app/home>). Ultimately, only *de novo* suspicious pathogenic CNVs were picked out according to the criteria or features below: (1) The CNV overlaps a critical interval with well-characterized genomic disorders (even if inherited from normal parents); (2) The CNV region contains haploinsufficient or triple sensitive diseases-associated genes; (3) Size of the CNV is large (generally >1 Mb, but also consider gene content) and it is not been observed in general population, especially *de novo*; (4) A similar CNV has been studied in large cohort showing significant association with a defined phenotype; (5) The CNV is recurrent in patients but lack of statistical power temporarily due to its rarity; (6) Deletion in a gene may cause disease by loss of heterozygosity (LOH).

**STR profiling/SNP genotyping.** Two approaches were used to estimate parental origin of *de novo* suspicious pathogenic CNVs detected in the patients. For STR profiling, proper microsatellites within the deleted or duplicated region were selected. Information about the specific location, heterozygosity in the crowd, primers and range of variation of microsatellites were taken from Marshfield Genetic Maps (<http://www.bli.uzh.ch/BLI/Projects/genetics/maps/marsh.html>) and UCSC Genome Browser (<http://genome.ucsc.edu/>). Analysis of markers was conducted on ABI 3100 Genetic Analyzer. By analyzing the different peaks of the same microsatellite loci among the proband and the parents in each pedigree, we obtained relevant information on the origin of CNVs. The second method based on SNPs was applied for CNVs without encompassing informative microsatellite loci. We defined the parent-of-origin by comparing the difference between the SNP loci in family members with two assisting methods: Sanger sequencing of the genes located in the interval of SNP-rich CNVs or analyzing the genotyping calls and B allele frequency in each CNV regions exported via *GenomeStudio* software. In general, we identified the parental origin based on at least two informative STR or SNP loci, whereas in a handful of cases only one was informative. All the methods were carried out in accordance with the approved guidelines.

**Structural analysis of genome sequence near the CNVs breakpoints.** To investigate the sequence homology near the CNVs breakpoints and reveal possible mechanisms of the disease-related CNVs formation, the genome sequence within a 20 kb window ( $\pm 10$  kb in either end) of all the 87 CNVs breakpoints were analyzed.

## References

1. Sudmant, P. H. *et al.* Evolution and diversity of copy number variation in the great ape lineage. *Genome Res.* **23**, 1373–1382 (2013).
2. Sebat, J. *et al.* Large-scale copy number polymorphism in the human genome. *Science* **305**, 525–528 (2004).
3. Zarrei, M., MacDonald, J. R., Merico, D. & Scherer, S. W. A copy number variation map of the human genome. *Nat. Rev. Genet.* **16**, 172–183 (2015).
4. Feuk, L., Carson, A. R. & Scherer, S. W. Structural variation in the human genome. *Nat. Rev. Genet.* **7**, 85–97 (2006).
5. Weiss, L. A. *et al.* Association between microdeletion and microduplication at 16p11.2 and autism. *N. Engl. J. Med.* **358**, 667–675 (2008).
6. Pinto, D. *et al.* Functional impact of global rare copy number variation in autism spectrum disorders. *Nature* **466**, 368–372 (2010).

7. Malhotra, D. & Sebat, J. CNVs: harbingers of a rare variant revolution in psychiatric genetics. *Cell* **148**, 1223–1241 (2012).
8. Sharp, A. J. *et al.* A recurrent 15q13.3 microdeletion syndrome associated with mental retardation and seizures. *Nat. Genet.* **40**, 322–328 (2008).
9. Redon, R. *et al.* Global variation in copy number in the human genome. *Nature* **444**, 444–454 (2006).
10. Conrad, D. F. *et al.* Origins and functional impact of copy number variation in the human genome. *Nature* **464**, 704–712 (2010).
11. Pang, A. W. *et al.* Towards a comprehensive structural variation map of an individual human genome. *Genome Biol.* **11**, R52 (2010).
12. Levy, S. *et al.* The diploid genome sequence of an individual human. *PLoS Biol.* **5**, e254 (2007).
13. Thomas, N. S. *et al.* Parental and chromosomal origin of unbalanced de novo structural chromosome abnormalities in man. *Hum. Genet.* **119**, 444–450 (2006).
14. Hehir-Kwa, J. Y. *et al.* De novo copy number variants associated with intellectual disability have a paternal origin and age bias. *J. Med. Genet.* **48**, 776–778 (2011).
15. Itsara, A. *et al.* De novo rates and selection of large copy number variation. *Genome Res.* **20**, 1469–1481 (2010).
16. Thomas, N. S. *et al.* Parental and chromosomal origins of microdeletion and duplication syndromes involving 7q11.23, 15q11-q13 and 22q11. *Eur. J. Hum. Genet.* **14**, 831–837 (2006).
17. Delio, M. *et al.* Enhanced maternal origin of the 22q11.2 deletion in velocardiofacial and DiGeorge syndromes. *Am. J. Hum. Genet.* **92**, 439–447 (2013).
18. López Correa, C., Brems, H., Lázaro, C., Marynen, P. & Legius, E. Unequal meiotic crossover: a frequent cause of NF1 microdeletions. *Am. J. Hum. Genet.* **66**, 1969–1974 (2000).
19. Duyzend, M. H. *et al.* Maternal modifiers and parent-of-origin bias of the autism-associated 16p11.2 CNV. *Am. J. Hum. Genet.* **98**, 45–57 (2016).
20. Miyake, N. *et al.* Preferential paternal origin of microdeletions caused by prezygotic chromosome or chromatid rearrangements in Sotos syndrome. *Am. J. Hum. Genet.* **72**, 1331–1337 (2003).
21. Tatton-Brown, K. *et al.* Multiple mechanisms are implicated in the generation of 5q35 microdeletions in Sotos syndrome. *J. Med. Genet.* **42**, 307–313 (2005).
22. Wilson, H. L. *et al.* Molecular characterisation of the 22q13 deletion syndrome supports the role of haploinsufficiency of SHANK3/PROSAP2 in the major neurological symptoms. *J. Med. Genet.* **40**, 575–584 (2003).
23. Luciani, J. J. *et al.* Telomeric 22q13 deletions resulting from rings, simple deletions, and translocations: cytogenetic, molecular, and clinical analyses of 32 new observations. *J. Med. Genet.* **40**, 690–696 (2003).
24. Drost, J. B. & Lee, W. R. Biological basis of germline mutation: comparisons of spontaneous germline mutation rates among drosophila, mouse, and human. *Environ. Mol. Mutagen.* **25** (Suppl. 26), 48–64 (1995).
25. Hou, Y. *et al.* Genome analyses of single human oocytes. *Cell* **155**, 1492–1506 (2013).
26. Ma, R. *et al.* Partial trisomy 2q33.3-q37.3 in a patient with an inverted duplicated neocentric marker chromosome. *Mol. Cytogenet.* **8**, 10 (2015).
27. Kong, A. *et al.* A high-resolution recombination map of the human genome. *Nat. Genet.* **31**, 241–247 (2002).
28. Broman, K. W., Murray, J. C., Sheffield, V. C., White, R. L. & Weber, J. L. Comprehensive human genetic maps: individual and sex-specific variation in recombination. *Am. J. Hum. Genet.* **63**, 861–869 (1998).
29. Dib, C. *et al.* A comprehensive genetic map of the human genome based on 5,264 microsatellites. *Nature* **380**, 152–154 (1996).
30. Miyata, T., Hayashida, H., Kuma, K., Mitsuyasu, K. & Yasunaga, T. Male-driven molecular evolution: a model and nucleotide sequence analysis. *Cold Spring Harb. Symp. Quant. Biol.* **52**, 863–867 (1987).
31. Miyata, T., Hayashida, H., Kuma, K. & Yasunaga, T. Male-driven molecular evolution demonstrated by different rates of silent substitutions between autosome- and sex chromosome-linked genes. *Proc. Japan Acad., Ser. B* **63**, 327–331 (1987).
32. Miyata, T., Kuma, K., Iwabe, N., Hayashida, H. & Yasunaga, T. Different rates of evolution of autosome-, X chromosome-, and Y chromosome-linked genes: hypothesis of male-driven molecular evolution in *Population biology of genes and molecules* (eds. Takahata, N. & Crow, J.) 341–357 (Baifukan, 1990).
33. Watson, C. T., Marques-Bonet, T., Sharp, A. J. & Mefford, H. C. The genetics of microdeletion and microduplication syndromes: an update. *Annu. Rev. Genomics Hum. Genet.* **15**, 215–244 (2014).
34. Stankiewicz, P. & Lupski, J. R. Genome architecture, rearrangements and genomic disorders. *Trends Genet.* **18**, 74–82 (2002).
35. Weckselblatt, B. & Rudd, M. K. Human structural variation: mechanisms of chromosome rearrangements. *Trends Genet.* **31**, 587–599 (2015).
36. Gu, W., Zhang, F. & Lupski, J. R. Mechanisms for human genomic rearrangements. *Pathogenetics* **1**, 4 (2008).
37. Nagamani, S. C. *et al.* Microdeletions including YWHAE in the Miller-Dieker syndrome region on chromosome 17p13.3 result in facial dysmorphism, growth restriction, and cognitive impairment. *J. Med. Genet.* **46**, 825–833 (2009).
38. Bi, W. *et al.* Increased LIS1 expression affects human and mouse brain development. *Nat. Genet.* **41**, 168–177 (2009).
39. Lee, J. A., Carvalho, C. M. & Lupski, J. R. A. DNA replication mechanism for generating nonrecurrent rearrangements associated with genomic disorders. *Cell* **131**, 1235–1247 (2007).
40. Carvalho, C. M. *et al.* Complex rearrangements in patients with duplications of MECP2 can occur by fork stalling and template switching. *Hum. Mol. Genet.* **18**, 2188–2203 (2009).
41. Miller, D. T. *et al.* Consensus statement: chromosomal microarray is a first-tier clinical diagnostic test for 16 individuals with developmental disabilities or congenital anomalies. *Am. J. Hum. Genet.* **86**, 749–764 (2010).
42. Kearney, H. M. *et al.* American College of Medical Genetics standards and guidelines for interpretation and reporting of postnatal constitutional copy number variants. *Genet. Med.* **13**, 680–685 (2011).

## Acknowledgements

We are very grateful to all the families and clinicians that participated in our research. We thank Lei Dai for helpful support and discussion on the genome sequence analysis. This work was supported by grants from the National Key Basic Research Program of China (2012CB944600), the National Key Technology R&D Program of China (2014BAI06B03) and the Basic work of Science and Technology Program of China (2014FY110700).

## Author Contributions

R.M. drafted the manuscript. R.M., L.D. and Y.X. performed the experiments. X.W. and Y.C. revised the manuscript critically for important intellectual content. R.G., R.Z. and J.G. assisted in analyzing the data. D.L. and L.W. initiated the study, made substantial contribution to conception design and revised the manuscript critically. All authors reviewed the manuscript.

## Additional Information

**Supplementary information** accompanies this paper at <http://www.nature.com/srep>

**Competing Interests:** The authors declare no competing financial interests.

**How to cite this article:** Ma, R. *et al.* A clear bias in parental origin of *de novo* pathogenic CNVs related to intellectual disability, developmental delay and multiple congenital anomalies. *Sci. Rep.* 7, 44446; doi: 10.1038/srep44446 (2017).

**Publisher's note:** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



This work is licensed under a Creative Commons Attribution 4.0 International License. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in the credit line; if the material is not included under the Creative Commons license, users will need to obtain permission from the license holder to reproduce the material. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>

© The Author(s) 2017