**De Novo** CNV Formation in Mouse Embryonic Stem Cells Occurs in the Absence of Xrcc4-Dependent Nonhomologous End Joining

Martin F. Arlt¹, Sountharia Rajendran¹, Shanda R. Birkeland², Thomas E. Wilson¹,²*, Thomas W. Glover¹*

¹ Department of Human Genetics, University of Michigan, Ann Arbor, Michigan, United States of America, ² Department of Pathology, University of Michigan, Ann Arbor, Michigan, United States of America

**Abstract**

Spontaneous copy number variant (CNV) mutations are an important factor in genomic structural variation, genomic disorders, and cancer. A major class of CNVs, termed nonrecurrent CNVs, is thought to arise by nonhomologous DNA repair mechanisms due to the presence of short microhomologies, blunt ends, or short insertions at junctions of normal and de novo pathogenic CNVs, features recapitulated in experimental systems in which CNVs are induced by exogenous replication stress. To test whether the canonical nonhomologous end joining (NHEJ) pathway of double-strand break (DSB) repair is involved in the formation of this class of CNVs, chromosome integrity was monitored in NHEJ-deficient Xrcc4⁻/⁻ mouse embryonic stem (ES) cells following treatment with low doses of aphidicolin, a DNA replicative polymerase inhibitor. Mouse ES cells exhibited replication stress-induced CNV formation in the same manner as human fibroblasts, including the existence of syntenic hotspot regions, such as in the Autos 2 and Wwox loci. The frequency and location of spontaneous and aphidicolin-induced CNV formation were not altered by loss of Xrcc4, as would be expected if canonical NHEJ were the predominant pathway of CNV formation. Moreover, de novo CNV junctions displayed a typical pattern of microhomology and blunt end use that did not change in the absence of Xrcc4. A number of complex CNVs were detected in both wild-type and Xrcc4⁻/⁻ cells, including an example of a catastrophic, chromothripsis event. These results establish that nonrecurrent CNVs can be, and frequently are, formed by mechanisms other than Xrcc4-dependent NHEJ.

**De Novo** CNV Formation in Mouse Embryonic Stem Cells Occurs in the Absence of Xrcc4-Dependent Nonhomologous End Joining.

PLOS Genet 8(9): e1002981. doi:10.1371/journal.pgen.1002981

Citation: Arlt MF, Rajendran S, Birkeland SR, Wilson TE, Glover TW (2012) De Novo CNV Formation in Mouse Embryonic Stem Cells Occurs in the Absence of Xrcc4-Dependent Nonhomologous End Joining. PLoS Genet 8(9): e1002981. doi:10.1371/journal.pgen.1002981

Editor: John H. J. Petriti, Memorial Sloan-Kettering Cancer Center, United States of America

Received May 31, 2012; Accepted August 1, 2012; Published September 20, 2012

Copyright: © 2012 Arlt et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Funding: This work was supported by a March of Dimes Foundation research grant to TWG (http://www.marchofdimes.com/research/researchgrants.html) and NIH/NCI grant R01-CA102563 to TEW (http://www3.cancer.gov/admin/gab/). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing Interests: The authors have declared that no competing interests exist.

* E-mail: wilsonte@umich.edu (TEW); glover@umich.edu (TWG)

**Introduction**

The importance of genomic copy number variant (CNVs), defined as submicroscopic deletions or duplications ranging in size from 50 bp to over a megabase [1], has become better understood in recent years. Normal polymorphic CNVs are a major contributor to human genomic variation and phenotypic diversity [2,3,4,5,6], while spontaneous CNVs are a very important and frequent cause of genetic and developmental disorders, including intellectual disability, neuropsychiatric disorders, and structural birth defects [7,8,9,10,11,12,13,14]. Their frequency further suggests a high involvement of a nonhomologous repair mechanism in their formation [22,23,24,25]. A number of different DNA repair mechanisms have been suggested to account for nonhomologous junctions, principally nonhomologous end-joining (NHEJ), alternative end-joining (alt-EJ), and forms of replication template switching [26].

Canonical NHEJ, along with homologous recombination (HR), is one of the two major mechanisms used to repair DNA double-strand breaks (DSBs) in eukaryotic cells. NHEJ directly joins two DSB ends without using extensive sequence homology to guide repair through the action of a well-defined set of proteins, including the Xrcc4-ligase IV complex, which is dedicated to and essential for this pathway [27]. The junctions formed are typically characterized by blunt ends, microhomologies, and small insertions, suggesting the involvement of a nonhomologous repair mechanism in their formation [22,23,24,25]. A number of different DNA repair mechanisms have been suggested to account for nonhomologous junctions, principally nonhomologous end-joining (NHEJ), alternative end-joining (alt-EJ), and forms of replication template switching [26].

CNVs have breakpoint junctions that are characterized by blunt ends, microhomologies, and small insertions, suggesting the involvement of a nonhomologous repair mechanism in their formation [22,23,24,25]. A number of different DNA repair mechanisms have been suggested to account for nonhomologous junctions, principally nonhomologous end-joining (NHEJ), alternative end-joining (alt-EJ), and forms of replication template switching [26].

...
Copy number variants (CNVs) are a major factor in genetic variation and are a common and important class of mutation in genomic disorders, yet there is limited understanding of how many CNVs arise and the risk factors involved. One DNA damage response pathway implicated in CNV formation is nonhomologous end joining (NHEJ), which repairs broken DNA ends by Xrcc4-dependent direct ligation. We examined the effects of loss of Xrcc4 and NHEJ on CNV formation following replication stress in mouse cells. Cells lacking NHEJ displayed unaltered CNV frequencies, locations, and breakpoint structures compared to normal cells. These results establish that CNV mutations in a cell model system, and likely in vivo, arise by a mutagenic mechanism other than canonical NHEJ, a pattern similar to that reported for model translocation events. Potential roles of alternative end joining and template switching are discussed.

**Author Summary**

For both experimentally-induced CNVs and those occurring in humans in vivo, the proposed mechanisms have only been inferred from junction sequences; direct experimental tests have been lacking. Because of the strong functional implications of the potential alternative mechanisms of nonrecurrent CNV formation, we sought to definitively explore the role of the well-defined, canonical NHEJ pathway. We report studies of CNV formation using Xrc4^-/- mouse embryonic stem (ES) cells and the DNA polymerase inhibitor aphidicolin (APH). Xrc4 is an essential component of DNA ligase IV that is absolutely required for NHEJ [27]. We demonstrate that APH induces de novo CNVs in mouse ES cells as it does in human fibroblasts, but that there is no difference in the frequency or structure of spontaneous or induced CNVs between wild-type and Xrc4^-/- cells. CNV breakpoint junctions were characterized by blunt ends and microhomologies regardless of genotype, with no observed shift in microhomology lengths. We conclude that replication-associated CNVs in mouse ES cells are created through mechanisms other than canonical NHEJ and discuss the potential roles of alt-EJ and template switching in the context of both simple and complex CNVs observed in the presence and absence of Xrc4.

**Results**

Xrcc4 deficiency does not reduce the frequency of APH-induced CNVs

To demonstrate the presence of a homozygous inactivating Xrcc4 deletion mutation in the Xrc4^-/- mouse ES cells used in these studies. Supporting this, these cells also demonstrated a large decrease in survival after exposure to ionizing radiation (Figure S1), consistent with NHEJ deficiency. Prior to the experiments, parental cells of each genotype were expanded from a single clone to minimize the number of potentially mosaic CNVs in the starting ES cell population. To induce replication stress, wild-type and Xrcc4^-/- mouse ES cells were cultured in the presence of 0–0.6 μM APH for 72 hours prior to plating for isolation of clonal cell populations. This mild dose of APH does not block the cell cycle, but instead allows replication to proceed at a reduced rate. Individual clones were expanded and subjected to CNV analysis using Nimblegen 3x720K aCGH arrays (Figure S2). De novo CNVs were defined as a segmental gain or loss detected in a clone when using the parental cell population as a reference.

A total of 85 independent clones from untreated or APH-treated wild-type and Xrc4^-/- cells were analyzed in three independent experiments. In wild-type cells, de novo CNVs were found in untreated and APH-treated clones at a frequency of 0.43 and 5.19 CNVs per clone, respectively (p<10^-14) (Figure 1A), demonstrating that, just as in previous studies with human fibroblasts, de novo CNVs in mouse ES cells can arise spontaneously during culture but that their frequency is significantly increased following replication stress.

In Xrc4^-/- cells, de novo CNVs were identified in untreated and APH-treated clones at a frequency of 1.31 and 7.34 de novo CNVs per clone, respectively (p<10^-14) (Figure 1A). When all experiments are considered together as in Figure 1A, there appears to be a slight increase in CNV induction in Xrc4^-/- cells compared to wild-type cells. However, this effect was only seen in one experiment in which APH-treated, wild-type cells had an unusually low CNV frequency. It was not recapitulated in the two subsequent experiments (Figure S3). Thus, there was no consistent effect of Xrcc4 deficiency on the frequency of spontaneous or APH-induced de novo CNV formation in ES cells. Most importantly, Xrc4 deficiency did not reduce the frequency of
of deletions compared to duplications is consistent with results seen in normal human fibroblasts after replication stress, in which 63–82% of CNVs were deletions [42,43], and in humans in vivo [6].

There was no difference in overall de novo CNV size between wild-type and Xrcc4+/− cells (Figure 1B). De novo CNVs in wild-type cells were generally large, with a median size of 59 kb (11.6 kb to 1.4 Mb). These sizes are similar to de novo CNVs seen in Xrcc4−/− cells, which had a median size of 63 kb (7.7 kb to 26.2 Mb). We did note that CNVs arising in mouse ES cells (median = 62 kb) were 2.2-fold smaller than de novo CNVs seen in similar experiments with human fibroblasts (median = 138 kb) [42] (Figure 1C).

Locations of de novo CNVs in wild-type and Xrcc4−/− cells

Consistent with previous observations in human fibroblasts, spontaneous and APH-induced CNVs in both wild-type and Xrcc4−/− mouse ES cells were distributed throughout the genome, with most arising in distinct, nonoverlapping regions (Figure 2). Superimposed on this distribution pattern were hotspots containing five or more different, overlapping CNVs, a number identified as unexpected by simulation modeling (see Materials and Methods). Each CNV within these hotspots had unique boundaries, indicating that each one arose independently, supporting the hypothesis that these regions are especially sensitive to replication stress. A difference in the size distribution of CNVs at hotspots and non-hotspots was observed, with hotspot CNVs being on average 1.9-fold larger than non-hotspot CNVs (median sizes of 89.8 kb and 46.3 kb, respectively). In addition, the abundance of deletions over duplications was more pronounced at hotspots than at non-hotspots. At non-hotspots, 79.5% (194/244) of de novo CNVs were of the deletion type, whereas at hotspots, almost all de novo CNVs (98.5%; 131/133) were deletions (p<0.0001). Most importantly, there was no apparent difference in the spatial distribution of CNVs between wild-type and Xrcc4−/− cells, including that hotspots accounted for 35.0% (50/143) and 35.5% (83/234) of all de novo CNVs, respectively.

Notably, several ES cell hotspots were found in the syntenic regions corresponding to hotspots seen in human fibroblasts (Table 1, Figure 3, Figure S4). The mouse hotspot with the most frequent occurrence of CNVs was in the Auts2 gene at chromosome 5G2. One or more CNVs in Auts2 were seen in 28/55 APH-treated clones, and accounted for 8.5% of all de novo CNVs. In addition, a hotspot was seen in the Wwox gene at 8E1, with CNVs found in 12/55 APH-treated clones. Both of these hotspots corresponded to hotspots seen in human fibroblasts (Figure 3, Table 1). However, the most frequently observed hotspot in human fibroblasts, at 3q13.31, was not a hotspot in mouse ES cells. In fact, most (11/13) of the hotspots in mouse ES cells were not observed in human fibroblasts (Table 1). In addition, two hotspots corresponded to the common fragile sites Fra8E1 (FRA16D) and Fra14A2 (FRA3B), while others mapped to regions syntenic to human fragile sites. These results suggest that while there is some conservation in replication stress-induced CNV hotspots, differences are also seen due to cell type or species variation.

CNV breakpoints in Xrcc4−/− cells show blunt ends, short microhomologies, and small insertions

In the absence of canonical NHEJ, the pattern of breakpoint junction sequences provides the most precise structural signature for revealing altered utilization of different end joining repair mechanisms [28,36,37]. To examine this, we sequenced 24 CNV breakpoint junctions from Xrcc4−/− cells and 17 from wild-type cells (Table 2, Figure 4). All of the junctions from both wild-type and Xrcc4−/− cells were characterized by 0–5 bp of homology, while two junctions in each cell type also had small insertions of 1–
3 bp. The mean length of microhomology in CNVs from wild-type and \(Xrcc4^{+/+}\) cells was 2.0 bp and 2.1 bp, respectively, and the median length for both was 2.0 bp. The lack of a shift toward longer microhomologies in the absence of \(Xrcc4\) strongly argues against a shift from utilization of canonical NHEJ toward alt-EJ in \(Xrcc4\)-deficient cells, and therefore that these junctions were not formed by \(Xrcc4\)-dependent DSB repair, even in wild-type cells.

Similarly, none of the sequenced junctions had long stretches of homology that would suggest a shift toward HR in NHEJ-deficient cells. To explore this further, we examined the breakpoint regions of unsequenced deletions \textit{in silico} to determine if \(Xrcc4^{+/+}\) cells had an increased breakpoint frequency near segmental duplications that might suggest formation by HR. For each CNV, 10 kb windows of sequence from the left and right breakpoint regions were compared to each other, scoring instances of sequence identity \(>90\%\) along a stretch of sequence at least 1000 bp. Such large sequence homologies were associated with only 3.5\% and 4.0\% of CNVs in wild-type and \(Xrcc4^{+/+}\) cells, respectively \((p = 1.0)\), reinforcing that there is no apparent increase in sequence homology at breakpoint regions in \(Xrcc4\)-deficient cells.

**Similar complex CNVs occur in wild-type and \(Xrcc4^{+/+}\) cells**

Thirteen of the sequenced breakpoint junctions were from five complex CNVs that contained two to four breakpoint junctions.
each. These CNVs recapitulate the type of complex events seen in human fibroblasts [42,43] and in vivo [19,38,39,40]. Two of these complex CNVs were found in wild-type cells and three were from Xrcc4+/− clones, again suggesting no Xrcc4-dependent structural difference. These complex CNVs were initially scored as simple deletions based on aCGH data, but sequencing revealed the presence of small retained sequences, as well as duplications and inversions that were below the resolution limit of the array (Figure 5A, Figure S5). In addition, Xrcc4+/− clone X6-40 contained a 2.5 Mb region of chromosome XE3 containing at least 10 discrete deletions (Figure 5B). This CNV is similar to the recently-described chromothripsis class of structural alterations [45]. Finally, we note that we successfully sequenced CNV breakpoint junctions in only 41 out of 60 attempts (68%). The CNVs for which breakpoint cloning failed likely include some junctions with complex structures that are difficult to amplify. Accordingly, we expect that our six complex CNVs are an underrepresentation of the actual incidence of such events.

Discussion

The experiments reported here demonstrate that APH-induced replication stress creates de novo CNVs in mouse ES cells that mimic in vivo nonrecurrent CNVs in the same manner as in human fibroblasts, and that these and spontaneous CNVs arise independently of Xrcc4-dependent NHEJ. Neither the frequency nor any observable feature of location or structure of APH-induced CNVs was affected by Xrcc4 loss. Almost all de novo CNVs in both wild-type and Xrcc4−/− cells had breakpoint regions devoid of the extended sequence homology needed to drive HR. Detailed characterization of individual breakpoint junctions confirmed that the CNVs arose via a non-homologous mechanism characterized by blunt ends, short microhomologies or short insertions, regardless of Xrcc4 status. These results eliminate canonical NHEJ as a primary mechanism for de novo CNV formation in our cell system. Moreover, the identification of complex, chromothripsis-like events in Xrcc4−/− cells suggest this rearrangement can
**Table 1.** *De novo* CNV hotspots in mouse ES cells.

| Chr | Hotspot Region Start | Hotspot Region End | Size | Associated genes | Number of CNVs at hotspot (n = 377) | Hotspot CNV frequency per APH-treated clone (n = 55) | Mouse fragile site | Human Syntenic Region | Human CNV hotspot? | Human fragile site |
|-----|----------------------|--------------------|------|------------------|-------------------------------------|-----------------------------------------------------|------------------|----------------------|-----------------|----------------------|
| 2C1.3 | 67,227,959 | 67,873,163 | 645,204 | Xsp2 | 15 (40%) | 27.3% | 2q24.3 | FRA2T |
| 2F3 | 140,550,331 | 141,557,233 | 1,006,902 | Macrod2 | 9 (2.4%) | 16.4% | 20p12.1 | FRA20B |
| 2H2 | 161,777,721 | 162,229,484 | 451,763 | Ptpn13 | 8 (2.1%) | 14.5% | 20q12 |
| 5A1 | 4,957,925 | 5,164,013 | 206,088 | Cdk14 | 5 (1.3%) | 9.1% | 7q21.13 |
| 5G2 | 132,196,231 | 132,935,046 | 738,815 | Auts2 | 32 (8.5%) | 58.2% | 7q11.22 | + |
| 6A2 | 21,333,255 | 21,447,908 | 11,653 | Kcnd2 | 6 (1.6%) | 10.9% | 7q31.31 |
| 8E1 | 117,160,242 | 117,608,147 | 447,905 | Wwox | 14 (3.7%) | 25.5% | Fra8E1 | 16p23.1 | + |
| 12A3 | 35,023,371 | 35,455,164 | 431,793 | Hdac9 | 10 (2.7%) | 18.2% | 7p21.1 |
| 12B1 | 38,642,601 | 39,183,139 | 540,538 | Dgkab | 5 (1.3%) | 9.1% | 7p21.2 |
| 14A1 | 10,761,778 | 10,964,749 | 202,971 | Fhit | 5 (1.3%) | 9.1% | Fra14A2 | 3p14.2 | FRA3B |
| 14E4 | 117,568,872 | 117,958,464 | 389,592 | Gpc6 | 10 (2.7%) | 18.2% | 13q31.3 | Unnamed |
| 17C | 50,995,219 | 51,146,542 | 151,323 | Tbc1d5 | 7 (1.9%) | 12.8% | 3p24.3 | FRA3A |
| 17E5 | 90,801,260 | 91,269,013 | 467,883 | Nhx1 | 7 (1.9%) | 12.8% | 2p16.3 | FRA2D |

(a) Arlt et al., 2009; Arlt et al., 2011.
(b) Mrasek et al., 2010.
(c) Fragile site not molecularly characterized, mapped to region cytogenetically.
(d) Human band corresponds to APH-induced fragile site in fibroblasts (Le Tallec et al, 2011).
doi:10.1371/journal.pgen.1002981.t001
occur in the absence of the NHEJ pathway. Instead, the findings together implicate alt-NHEJ and/or replication template switching as the principal mediator(s) of nonhomologous junction formation.

In many ways, the results of this CNV study are similar to observations made using a two-DSB translocation model system [35,46,47]. Jasin and colleagues have shown that alt-EJ rather than canonical NHEJ likely acts in the formation of translocations following DSB induction, even when a functional NHEJ pathway is present. Similar to results here, they found that loss of Xrcc4 does not change the nature of translocation breakpoint junctions, which, like those seen at APH-induced CNVs, are typically characterized by 0–4 bp of microhomology. In addition, translocation junctions were sometimes complex, containing multiple insertions that were duplicated from sequences that could be as much as 4 Mb away from the initiating DSB, suggesting that iterative DNA synthesis occurred prior to joint resolution. These similar results could indicate that alt-EJ is playing a role in the CNVs induced in our system. However, a key difference is that loss of Xrcc4 does not change the nature of translocation breakpoint junctions, which, like those seen at APH-induced CNVs, are typically characterized by 0–4 bp of microhomology.

Figure 4. Comparison of observed de novo CNV breakpoint junction sequence homology in wild-type and Xrcc4−/− cells. Histogram showing CNV breakpoint homology in wild-type (blue) and Xrcc4−/− cells (red), compared to the expected distribution if microhomology usage was random (gray).

doi:10.1371/journal.pgen.1002981.g004
A template switch upstream of the DSB end. In total, though, both alt-EJ and maturation of template copying events might be used to resolve one-ended replication DSBs in different CNV events.

The results from this study also relate to the nature of chromothripsis events and CNV hotspots. Chromothripsis is a recently-described, catastrophic chromosome rearrangement seen in 2–3% of cancers [45]. It is also seen as a constitutional event in humans, suggesting that it is not specific to aberrant DNA repair pathways seen in cancer [48,49]. These complex rearrangements are thought to occur as a single catastrophic event, rather than accumulating over time [50,51]. The detection of a chromothripsis-like event in a Xrcc4−/− clone suggests that these catastrophic rearrangements can occur via an NHEJ-independent pathway, in contrast to chromosome shattering followed by

Table 2. APH-induced de novo CNV breakpoint junctions.

| Clone  | Genotype | [APH] | CNV type | Chr | Left breakpoint (hg18) | Right breakpoint (hg18) | # bp homology at junction | Homologous bases | Inserted bases |
|--------|----------|-------|----------|-----|------------------------|------------------------|--------------------------|------------------|---------------|
| X6-5   | Xrcc4−/− | 0.6   | Deletion | 1   | 68,450,738             | 68,966,178             | 1                        | A                | -             |
| X6-10  | Xrcc4−/− | 0.6   | Deletion | 2   | 67,650,700             | 67,818,163             | 2                        | TT               | -             |
| X6-5   | Xrcc4−/− | 0.6   | Deletion | 2   | 140,718,860            | 141,009,164            | 3                        | CAA              | -             |
| WT6-31 | WT       | 0.6   | Deletion | 2   | 162,172,198            | 162,232,243            | 0                        | -                | -             |
| X6-11  | Xrcc4−/− | 0.6   | Deletion | 3   | 156,164,457            | 156,204,161            | 2                        | TT               | -             |
| WT6-33 | WT       | 0.6   | Complex  | 4   | 153,025,685            | 153,222,760            | 1                        | G                | -             |
| WT6-33 | WT       | 0.6   | Complex  | 4   | 153,025,694            | 153,222,677            | 0                        | -                | TG            |
| WT6-33 | WT       | 0.6   | Complex  | 4   | 153,222,732            | 153,025,763            | 2                        | TG               | -             |
| WT6-33 | WT       | 0.6   | Complex  | 4   | 153,222,767            | 153,222,638            | 4                        | AACA             | -             |
| WT6-3  | WT       | 0.6   | Complex  | 5   | 5,038,414              | 5,140,617              | 3                        | CAT              | -             |
| WT6-3  | WT       | 0.6   | Complex  | 5   | 5,141,891              | 5,167,320              | 4                        | GGTA             | -             |
| X6-10  | Xrcc4−/− | 0.6   | Deletion | 5   | 132,199,490            | 132,455,826            | 0                        | -                | -             |
| WT6-3  | WT       | 0.6   | Deletion | 5   | 132,226,382            | 132,474,192            | 0                        | A                | -             |
| WT6-1  | WT       | 0.6   | Deletion | 5   | 132,476,605            | 132,681,396            | 1                        | A                | -             |
| X6-37  | Xrcc4−/− | 0.6   | Deletion | 5   | 132,488,580            | 132,582,758            | 2                        | CA               | -             |
| WT6-33 | WT       | 0.6   | Deletion | 5   | 132,621,299            | 132,650,279            | 1                        | T                | -             |
| X6-11  | Xrcc4−/− | 0.6   | Complex  | 5   | 132,624,159            | 132,850,747            | 3                        | CAG              | -             |
| X6-11  | Xrcc4−/− | 0.6   | Complex  | 5   | 132,850,831            | 132,844,066            | 2                        | AC               | -             |
| X6-40  | Xrcc4−/− | 0.6   | Deletion | 6   | 77,628,529             | 77,694,182             | 0                        | -                | ATA           |
| X6-6   | Xrcc4−/− | 0.6   | Complex  | 8   | 30,060,662             | 30,079,853             | 3                        | TCA              | -             |
| X6-6   | Xrcc4−/− | 0.6   | Complex  | 8   | 30,103,083             | 30,107,692             | 4                        | TCTG             | -             |
| WT6-1  | WT       | 0.6   | Deletion | 8   | 28,994,436             | 29,024,833             | 2                        | GT               | -             |
| X6-7   | Xrcc4−/− | 0.6   | Duplication | 9 | 47,093,534             | 47,349,312             | 2                        | CAG              | -             |
| WT6-4  | WT       | 0.6   | Deletion | 11  | 25,829,101             | 25,853,015             | 5                        | AGGTC            | -             |
| X6-5   | Xrcc4−/− | 0.6   | Deletion | 11  | 49,786,986             | 49,809,524             | 2                        | AC               | -             |
| X6-20  | Xrcc4−/− | 0.6   | Deletion | 12  | 38,642,366             | 38,735,475             | 2                        | TT               | -             |
| WT6-32 | WT       | 0.6   | Deletion | 12  | 80,456,477             | 80,489,599             | 3                        | GAG              | -             |
| WT6-14 | WT       | 0.6   | Deletion | 14  | 23,303,767             | 23,354,878             | 0                        | -                | -             |
| X6-13  | Xrcc4−/− | 0.6   | Deletion | 14  | 30,009,560             | 30,139,884             | 4                        | ACTA             | -             |
| X6-35  | Xrcc4−/− | 0.6   | Complex  | 16  | 49,786,986             | 49,809,344             | 2                        | GG               | -             |
| X6-35  | Xrcc4−/− | 0.6   | Complex  | 16  | 49,809,524             | 49,805,150             | 3                        | GCA              | -             |
| X6-4   | Xrcc4−/− | 0.6   | Deletion | 16  | 66,978,951             | 67,305,221             | 1                        | C                | -             |
| WT6-14 | WT       | 0.6   | Deletion | 16  | 97,254,626             | 97,308,282             | 1                        | CA               | -             |
| X6-4   | Xrcc4−/− | 0.6   | Deletion | 18  | 72,162,955             | 72,296,910             | 0                        | -                | CA            |
| X6-9   | Xrcc4−/− | 0.6   | Deletion | X   | 40,273,857             | 40,335,846             | 1                        | C                | -             |
| WT6-4  | WT       | 0.6   | Deletion | X   | 58,904,820             | 59,057,091             | 3                        | AGG              | -             |
| WT6-43 | WT       | 0.6   | Deletion | X   | 80,613,171             | 80,738,690             | 0                        | -                | -             |
| X6-6   | Xrcc4−/− | 0.6   | Deletion | X   | 84,741,276             | 84,811,809             | 0                        | -                | -             |
| X6-19  | Xrcc4−/− | 0.6   | Deletion | X   | 110,651,646            | 110,711,246            | 2                        | TT               | -             |
| X6-4   | Xrcc4−/− | 0.6   | Deletion | X   | 126,700,368            | 126,786,727            | 0                        | -                | -             |

doi:10.1371/journal.pgen.1002981.t002
Figure 5. Example of complex APH-induced de novo CNVs in mouse ES cells. (A) A complex CNV with two junctions at 5G2 in APH-treated \textit{Xrcc4}^{−/−} clone X6-11. Based on aCGH data, this CNV was called as a deletion, but sequencing of the breakpoint junctions revealed that this CNV was complex, containing a 219.9 kb deletion (red), as well as a duplication-insertion of 84 bp (blue) at the deletion boundary. (B) aCGH data demonstrating a region of complex CNV in APH-treated \textit{Xrcc4}^{−/−} clone X6-40 at XE3 containing 10 or more discrete deletions across a $\sim$2.5 Mb region. Data from the same genomic interval in a control clone (X6-38) is shown for comparison.
doi:10.1371/journal.pgen.1002981.g005
In summary, the experiments described here demonstrate that canonical, Xrcc4-dependent NHEJ is not involved in CNV formation in somatic cells cultured in vitro. Evidence from breakpoint junction structures further demonstrates that the CNVs did not form via HR. Instead, the data implicate a replication-dependent alt-EJ and/or template switching mechanism. Because of the strong similarity of the observed CNVs to the major classes of nonrecurrent normal and pathogenic CNVs seen in humans, we argue that these conclusions are generalizable to most de novo, nonrecurrent CNV formation in both germline and somatic human cell lineages, with the simple difference that event rates are higher in our model system because replication is exogenously stressed. Although not mutually exclusive, important features distinguish the remaining alt-EJ and template switching mechanisms, specifically the manner in which the strands are stably resolved. Also enigmatic is precisely which DNA intermediate is the substrate for template switching and which proteins are involved in executing the transfer when little or no microhomology is present. Major efforts moving forward should thus be to delineate the precise strand intermediates and protein mechanisms involved in mediating nonrecurrent CNV formation.

**Materials and Methods**

**Generation of cell clones containing replication stress-induced CNVs**

All experiments were performed with two isogenic male mouse ES cell lines. The first (TC1) was wild-type, while the second (Xrcc4<sup>−/−</sup>) was homozygous for a targeted inactivation of Xrcc4 [59]. Genomic DNA was prepared from cells using the Blood & Cell Culture DNA Mini Kit (Qiagen). ES Cells were grown irradiated fibroblast feeder cells in DMEM media supplemented with 15% FBS, 20 mM HEPES, and 1 mM sodium pyruvate. To create replication stress-induced CNVs, cells were treated with 0.6 μM APH. In three independent experiments, cells were treated for 72 hours followed by a 24 hour recovery period before plating at low density for single-cell clones. Cells were plated at a

---

**Figure 6. Models for replication-dependent, Xrcc4-independent CNV formation.** The induction of CNVs by replication stress strongly implicates stalled replication as a key intermediate (top). Template switching without fork collapse might directly create CNVs without DSB formation (left). Alternatively, fork collapse and end processing might lead to iterative template copying prior to final stable resolution of single-ended DSBs by either maturation of a one DSB end into a replication fork (MMBIR, middle) or joining of two distant DSBs by alt-EJ (right). In neither case are the single-ended DSBs good substrates for NHEJ. Results here establish that Xrcc4-dependent NHEJ is neither required for, nor suppresses, CNV formation via these inferred intermediates.

doi:10.1371/journal.pgen.1002981.g006
density of 100–500 cells per 100-mm culture dish and individual clones isolated with a pipette tip after 7–10 days.

**aCGH**

CNVs were detected using Nimblegen whole genome arrays containing 720,000 (720K) unique sequence oligonucleotides (Roche Applied Science). Arrays were prepared according to the manufacturer’s protocol. Arrays were scanned on an Axon 4000B scanner (Molecular Devices) with GenePix software at 532 and 655 wavelengths. Data extraction, normalization, and visualization were achieved by using manufacturer-provided software (NimbleScan and Signal-Map). Arrays were analyzed for copy number differences using SegMNT, part of Nimblegen’s Nim-Scan software package, as well as our software platform, VAMP, as previously described [19]. All clones were analyzed using the appropriate mixed parental cell population as the normalization reference. This approach routinely detects CNVs larger than 20 kb and can detect CNVs as small as a ~1 kb, depending on probe placement.

**CNV breakpoint junctions**

CNV breakpoint junctions were amplified using the Expand Long Template PCR System (Roche Applied Science). For deletions, PCR primer pairs were generated that flanked deletion breakpoints, whereas for duplications, primers were designed within the duplicated region, directed outward, as described previously [43]. PCR amplification generated a product that spanned the breakpoint junction. All products were then subjected to standard Sanger sequencing. The resulting sequence was compared to the reference genome (build mm9) to identify the breakpoint junctions.

**Statistical Methods**

CNVs in our model system are relatively rare events and therefore the numbers of CNVs per clone are expected to fit a Poisson distribution determined by the mean frequency of CNVs in all clones. Therefore, p values of treated vs. untreated samples were determined using the one-sided E-test of Krishnamoorthy and Thomson for comparing two Poisson mean rates [60].

To determine whether the observed clustering of CNVs within genome regions was non-random, we performed the Monte Carlo simulation summarized in Table S1, as previously described for human cells [42]. A simulation of 10,000 iterations was performed on the combined wild-type and Xrcc4−/− CNV sets. Regions with 5 or more overlapping CNVs were very rarely observed by random placement (p<0.01, Table S1) and were therefore scored as CNV hotspots in mouse ES cells. These hotspot regions are highlighted by shading in Table S2.

**Supporting Information**

**Figure S1** Confirmation of Xrcc4−/− mutant mouse ES cell line. (A) PCR confirmation of mutant Xrcc4 allele with deletion of exon 3 [52]. PCR primers: XFor1 GCTGAGTACTTAGATTGAGTAC; XRev1 ACCTGGGTGACCCCTACACG. (B) IR sensitivity of Xrcc4+/− ES cells. Wild-type and Xrcc4−/− cells were irradiated with indicated doses of X irradiation, cultured for 7 days, and surviving colonies were stained and counted. IR sensitivity is expressed as the percentages of surviving colonies over unirradiated controls.

**Figure S2** Examples of APH-induced CNVs showing Nimblegen aCGH intensity data (log2R). Each dot represents a single probe on the array. (A) A 107.0 kb deletion at 8E1 in clone X6-21 is easily detected by a reduction in the log2R intensity. (B) A 486.6 kb duplication at 9C–D in clone X6-7 can be identified by an increase in the log2R values.

**Figure S3** Box and whisker plot illustrating APH-induced CNV formation in wild-type (“WT”), blue and Xrcc4−/− (red) cells, in each of three experiments. It is evident that wild-type cells from Experiment 1 formed unusually low numbers of de novo CNVs compared to all other experimental groups. As a result, when data are combined, there is an apparent increase in CNV formation in Xrcc4−/− cells (Figure 1A).

**Figure S4** CNV coverage at all hotspots in mouse ES cells. The x-axis shows the position along the chromosome, while the y-axis indicates that fraction of hotspot CNVs that crossed a particular 10 kb genomic window.

**Figure S5** Demonstration of complex CNV rearrangements in wild-type and Xrcc4−/− cells. Each of these CNVs was called as a deletion based on aCGH data. Breakpoint junction sequencing revealed small duplications (blue), interrupted deletions (red), and inversions (gray).

**Table S1** Monte Carlo simulation to identify CNV hotspots.

**Table S2** List of de novo CNVs.

**Acknowledgments**

We thank JoAnn Sekiguchi for supplying us with the mouse ES cell lines used in this paper, as well as for insightful discussions and comments on the manuscript. We also thank Keisha McSweeney for assistance with PCR amplification of breakpoint junctions and Robert Lyons in the University of Michigan DNA Sequencing Core for providing Sanger sequencing.

**Author Contributions**

Conceived and designed the experiments: MFA TEW TWG. Performed the experiments: MFA SR SRB TEW. Analyzed the data: MFA SR SRB TEW TWG. Wrote the paper: MFA TEW TWG.

**References**

1. Mills RE, Walter K, Stewart C, Handshaker RE, Chen K, et al. (2011) Mapping copy number variation by population-scale genome sequencing. Nature 470: 56–66.

2. Iafrate AJ, Feuk L, Rivera MN, Listewnik ML, Donahoe PK, et al. (2004) Large-scale copy number variations detected by DNA array comparative genomic hybridization. Science 304: 1470–1472.

3. Redon R, Ishikawa S, Fitch KR, Feuk L, Perry GH, et al. (2006) Global variation in copy number in the human genome. Nature 444: 448–454.

4. Sebat J, Lakshmi B; Troge J, Alexander J, Young J, et al. (2004) Large-scale copy number polymorphism in the human genome. Science 305: 525–528.
18. Marshall CR, Noor A, Vincent JR, Lionel AC, Feuk L, et al. (2008) Structural variation of chromosomes in autism spectrum disorder. Am J Hum Genet 82: 477–488.

19. Christian SL, Bebb CW, Sudi J, Kumar RA, Liu S, et al. (2008) Novel submicroscopic chromosomal abnormalities detected in autism spectrum disorder. Biol Psychiatry 63: 1111–1117.

20. Cook EH, Jr., Scherer SW (2000) Copy-number variations associated with neuropsychiatric conditions. Nature 455: 919–923.

21. Kirov G, Grozeva D, Norton X, Kovanen E, Manniripragada KK, et al. (2009) Support for the involvement of large copy number variants in the pathogenesis of schizophrenia. Hum Mol Genet 18: 1497–1503.

22. Stankiewicz P, Beaudet AL (2007) Use of array CGH in the evaluation of dysmorphology, malformations, developmental delay, and idiopathic mental retardation. Curr Opin Genet Dev 17: 182–192.

23. Zhang F, Gu W, Hurles ME, Luski RJ (2009) Copy number variation in human health, disease, and evolution. Annu Rev Genomics Hum Genet 10: 451–481.

24. Campbell PJ, Stephens PJ, Pleasance ED, O’Meara S, Li H, et al. (2008) Identification of complete coding sequences in the human genome. Nature 451: 872–876.

25. Lee JA, Carvalho CM, Lupski JR (2007) A DNA replication mechanism for generating nonrecurrent rearrangements associated with genomic disorders. Cell 131: 1233–1247.

26. Campbell PJ, Stephens PJ, Pleasance ED, O’Meara S, Li H, et al. (2008) Identification of somatically acquired rearrangements in cancer using genome-wide massively parallel paired-end sequencing. Nat Genet 40: 722–729.

27. Mahaney BL, Meek K, Wala K, Park SS, et al. (2011) Diverse mutational mechanisms cause pathogenic subtelomeric rearrangements. Hum Mol Genet 20: 3769–3778.

28. Lee JA, Carvalho CM, Lupski RJ (2007) A DNA replication mechanism for generating nonrecurrent rearrangements associated with genomic disorders. Cell 131: 1233–1247.

29. Shaw CJ, Luski RJ (2005) Non-recurrent 1p11.2 deletions are generated by homologous and non-homologous mechanisms. Hum Genet 116: 1–7.

30. Hastings PJ, Luski RJ, Rosenblum SM, Ira G (2009) Mechanisms of change in human copy number variation. Curr Opin Genet Dev 19: 543–554.

31. Mahaney BL, Meek K, Lass-Miller SP (2009) Repair of ionizing radiation-induced DNA double-strand breaks by non-homologous end-joining. Biochem J 417: 639–650.

32. McVey M, Lee SE (2008) MMEJ repair of double-strand breaks (director’s cut): deleted sequences and alternative endings. Trends Genet 24: 529–530.

33. Boubakurov-Azzouz I, Rischetti M (2008) Low joining efficiency and non-conservative repair of two distant double-strand breaks in mouse embryonic stem cells. DNA Repair (Amst) 7: 149–161.

34. Inoue K, Osaka H, Thurston VC, Clarke JT, Yoneyama A, et al. (2002) Genomic rearrangements resulting in PLP1 deletion occur by nonhomologous end joining and cause different dysmyelinating phenotypes in males and females. Am J Hum Genet 71: 858–853.

35. Simsek D, Jasim M (2010) Alternative end-joining is suppressed by the canonical NHEJ component Xrc+ligase IV during chromosomal translocation formation. Nat Struct Mol Biol 17: 410–416.

36. Daley JM, Wilson TE (2005) Rejoining of DNA double-strand breaks as a function of overhang length. Mol Cell 23: 986–906.

37. Hastings PJ, Ira G, Luski RJ (2009) A microhomology-mediated break-induced replication model for the origin of human copy number variation. PLoS Genet 5: e1000327. doi:10.1371/journal.pgen.1000327.

38. Peric H, Ben-Dor A, Talucho A, Samgas N, Rodriguez-Revenga L, et al. (2008) The fine-scale and complex architecture of human copy-number variation. Am J Hum Genet 82: 685–695.

39. Yakin B, Wong K, Bhomra A, Gooden M, Kraane TM, et al. (2012) The fine-scale architecture of structural variants in 17 mouse genomes. Genome Biol 13: R18.

40. Liu P, Carvalho CM, Hastings P, Luski RJ (2012) Mechanisms for recurrent and complex human genomic rearrangements. Curr Opin Dev Biol. 24: 107–114.

41. Toffolatti L, Cardazzo B, Nobile C, Danieli GA, Gualandi F, et al. (2002) End joining and cause different dysmyelinating phenotypes in males and females. DNA Repair (Amst) 7: 149–161.

42. Conrad DF, Bird C, Blackburne B, Lindsay S, Mamanova L, et al. (2010) Genome-wide association study identifies rare copy number variation. Nature 463: 836–840.

43. Christian SL, Brune CW, Sudi J, Kumar RA, Liu S, et al. (2008) Novel submicroscopic chromosomal abnormalities detected in autism spectrum disorder. Biol Psychiatry 63: 1111–1117.

44. Egan CM, Sridhar S, Wigler M, Hall IM (2007) Recurrent DNA copy number variation in schizophrenia. J Autism Dev Disord. 37: 52–62.

45. Toffolatti L, Cardazzo B, Nobile C, Danieli GA, Gualandi F, et al. (2002) End joining and cause different dysmyelinating phenotypes in males and females. DNA Repair (Amst) 7: 149–161.

46. Zhang F, Gu W, Hurles ME, Luski RJ (2009) Copy number variation in human health, disease, and evolution. Annu Rev Genomics Hum Genet 10: 451–481.

47. Stankiewicz P, Beaudet AL (2007) Use of array CGH in the evaluation of dysmorphology, malformations, developmental delay, and idiopathic mental retardation. Curr Opin Genet Dev 17: 182–192.

48. Toffolatti L, Cardazzo B, Nobile C, Danieli GA, Gualandi F, et al. (2002) End joining and cause different dysmyelinating phenotypes in males and females. DNA Repair (Amst) 7: 149–161.

49. Christiansen SL, Brune CW, Sudi J, Kumar RA, Liu S, et al. (2008) Novel submicroscopic chromosomal abnormalities detected in autism spectrum disorder. Biol Psychiatry 63: 1111–1117.

50. Ozeri-Galai E, Lebofsky R, Rahat A, Bester AC, Bensimon A, et al. (2011) Chromothripsis as a mechanism driving complex de novo structural rearrangements in the germline. Hum Mol Genet 20: 1916–1924.

51. Christiansen SL, Brune CW, Sudi J, Kumar RA, Liu S, et al. (2008) Novel submicroscopic chromosomal abnormalities detected in autism spectrum disorder. Biol Psychiatry 63: 1111–1117.

52. Stankiewicz P, Inoue K, Bi W, Walz K, Park SS, et al. (2011) Diverse mutational mechanisms cause pathogenic subtelomeric rearrangements. Hum Mol Genet 20: 3769–3778.

53. Lee JA, Carvalho CM, Luski RJ (2007) A DNA replication mechanism for generating nonrecurrent rearrangements associated with genomic disorders. Cell 131: 1233–1247.

54. Campbell PJ, Stephens PJ, Pleasance ED, O’Meara S, Li H, et al. (2008) Identification of somatically acquired rearrangements in cancer using genome-wide massively parallel paired-end sequencing. Nat Genet 40: 722–729.

55. Shaw CJ, Luski RJ (2005) Non-recurrent 1p11.2 deletions are generated by homologous and non-homologous mechanisms. Hum Genet 116: 1–7.

56. Hastings PJ, Luski RJ, Rosenblum SM, Ira G (2009) Mechanisms of change in human copy number variation. Curr Opin Genet Dev 19: 543–554.

57. Mahaney BL, Meek K, Lass-Miller SP (2009) Repair of ionizing radiation-induced DNA double-strand breaks by non-homologous end-joining. Biochem J 417: 639–650.

58. McVey M, Lee SE (2008) MMEJ repair of double-strand breaks (director’s cut): deleted sequences and alternative endings. Trends Genet 24: 529–530.

59. Boubakurov-Azzouz I, Rischetti M (2008) Low joining efficiency and non-conservative repair of two distant double-strand breaks in mouse embryonic stem cells. DNA Repair (Amst) 7: 149–161.

60. Inoue K, Osaka H, Thurston VC, Clarke JT, Yoneyama A, et al. (2002) Genomic rearrangements resulting in PLP1 deletion occur by nonhomologous end joining and cause different dysmyelinating phenotypes in males and females. Am J Hum Genet 71: 858–853.