Supporting Information

Arlt et al. 10.1073/pnas.1109272108

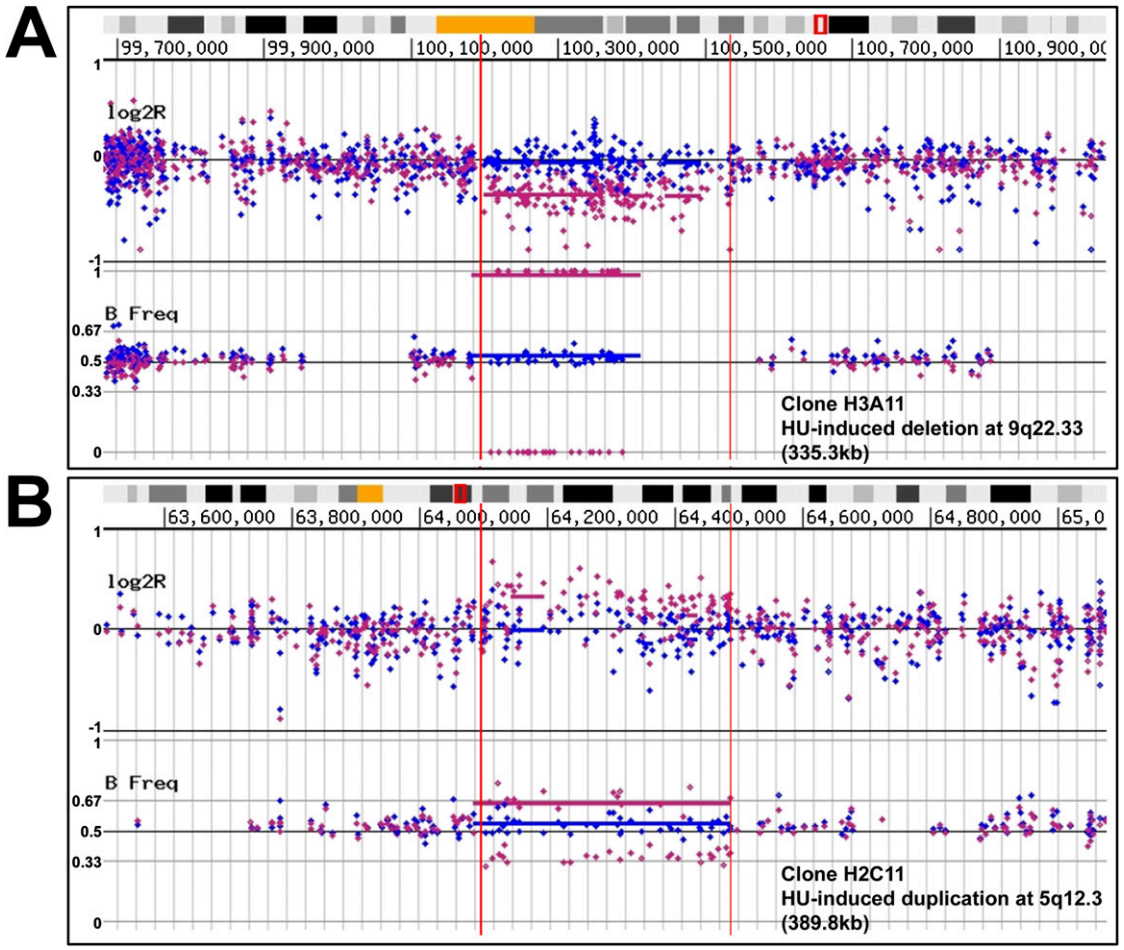

Fig. S1. Examples of hydroxyurea (HU)-induced copy number variants. Illumina SNP array intensity data (log2R) and B allele frequency (B Freq) are at the top and bottom of each panel, respectively. Each dot represents a single probe on the array. Blue dots represent data generated from the untreated control population. Red dots represent data from an HU-treated clone. (*A*) A 335.3-kb deletion at 9q22.33 in clone H3A11 is easily detected by a reduction in the log2R intensity and loss of heterozygous B allele frequencies, shifting from 0.5 to 0 or 1. (*B*) A 389.8-kb duplication at 5q12.3 in clone H2C11 can be identified by an increase in the log2R values and a shift in the B allele frequency from 0.5 to 0.33 or 0.67.

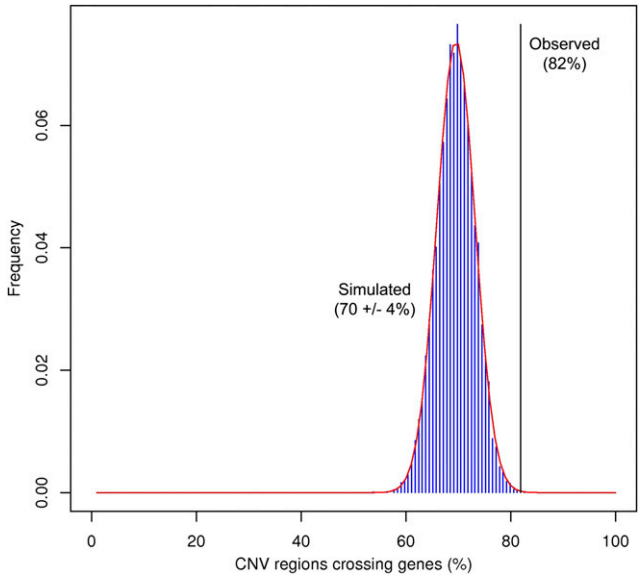

Fig. S2. Overlap of copy number variant (CNV) regions with genes. Histogram (blue bars) shows the frequency of iterations in a 10,000-iteration simulation that had different percentages of simulated regions crossing RefSeq genes, fitted to a Gaussian distribution (red line). Black line shows the percentage of observed copy number variant regions that overlap genes. Details and discussion in main text (*Materials and Methods*).

Fig. S3. Copy number variant breakpoint junctions from hydroxyurea (HU)-treated and untreated clones. Breakpoints from 12 deletions and 2 duplications from hydroxyurea-treated cells and 4 deletions and 1 duplication from untreated cells were sequenced, revealing microhomologies and blunt ends. Sequences from the left and right breakpoint regions are red and blue, respectively. Regions of perfect homology at the junction are underlined and highlighted in yellow.

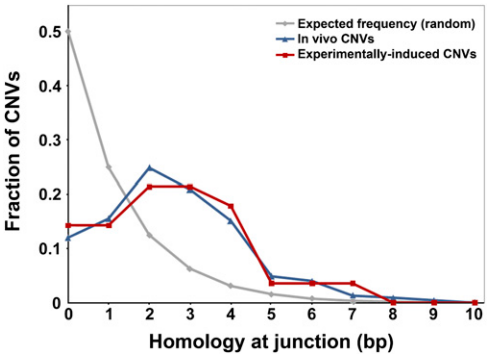

Fig. S4. Comparison of breakpoint sequences in hydroxyurea- and aphidicolin-induced copy number variants (CNVs) in vitro and germline CNVs in vivo. The frequencies of CNV breakpoints with 0–10 bp of microhomology are compared between all replication stress-induced CNVs from this and previous studies (red line) (1, 2) and germline CNVs seen in vivo (blue line) (3, 4). The expected frequency if microhomologous junctions occurred randomly is also shown (gray line).

1. Arlt MF, et al. (2009) Replication stress induces genome-wide copy number changes in human cells that resemble polymorphic and pathogenic variants. *Am J Hum Genet* 84:339–350.
2. Arlt MF, et al. (2011) Comparison of constitutional and replication stress-induced genome structural variation by SNP array and mate-pair sequencing. *Genetics* 187:675–683.
3. Vissers LE, et al. (2009) Rare pathogenic microdeletions and tandem duplications are microhomology-mediated and stimulated by local genomic architecture. *Hum Mol Genet* 18: 3579–3593.
4. Conrad DF, et al. (2010) Mutation spectrum revealed by breakpoint sequencing of human germline CNVs. *Nat Genet* 42:385–391.

Table S1. Monte Carlo simulation to identify copy number variant (CNV) hotspots

	Observed CNVs		Simulation mean			
CNVs in region (N)	Regions (R_{obs})	CNVs	Regions (λ_N)	CNVs	$P(>0)$	$P(R_{obs})$
1	103	103	193	193	1.0	1.0
2	11	22	6.7	13.4	1.0	0.08
3	5	15	0.28	0.84	0.2	0.00001
4	3	12	0.011	0.044	0.01	0.0000002
6	1	6	0.0001	0.0006	0.0001	0.0001
7	1	7	0	0	<0.0001	<0.0001
11	1	11	0	0	<0.0001	<0.0001
31	1	31	0	0	<0.0001	<0.0001

Details in main text (*Materials and Methods*).

Other Supporting Information Files

Dataset S1 (XLSX)