

Supporting Information

Arlt et al. 10.1073/pnas.1109272108

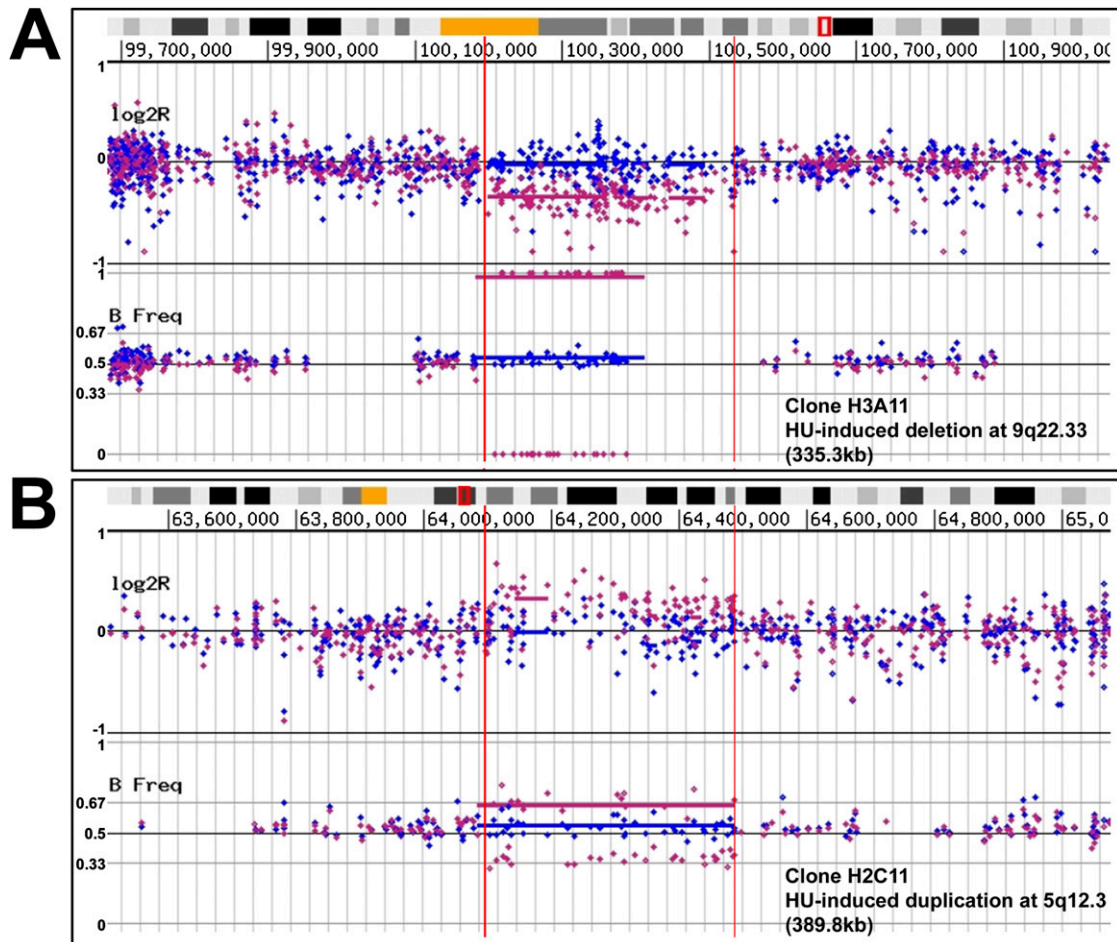


Fig. S1. Examples of hydroxyurea (HU)-induced copy number variants. Illumina SNP array intensity data (log₂R) and B allele frequency (B Freq) are at the top and bottom of each panel, respectively. Each dot represents a single probe on the array. Blue dots represent data generated from the untreated control 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.

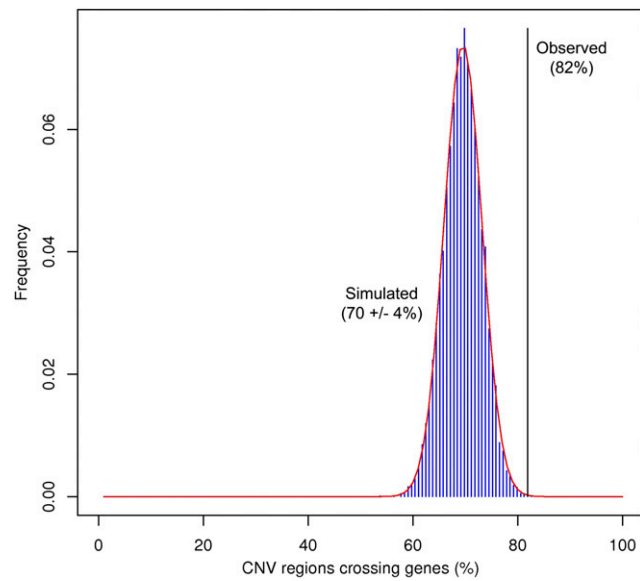


Fig. S2. Overlap of copy number variant (CNV) regions with genes. Histogram (blue bars) shows the frequency of iterations in a 10,000-iteration simulation that had different percentages of simulated regions crossing RefSeq genes, fitted to a Gaussian distribution (red line). Black line shows the percentage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

H1A11 2p22.2 deletion (100µM HU)	36,479,534	H2C41 12p13.33 deletion (200µM HU)	1,095,921
Bkpt#1 CATACTTAATAAAATAGTCACGTCGTTAA <u>ACTGATGATAAGTCTTACTAGTTTA</u>		Bkpt#1 TGGAATCTGTTAAAGTTCAGATAGCT <u>CTAATTTTAAACCACAGATTATTAG</u>	
Del CATACTTAATAAAATAGTCACGTCGTTAA <u>ACTGATGATAAGTCTTACTAGTTTA</u>		Del TGGAATCTGTTAAAGTTCAGATAGCT <u>TATAGGCGTGAGCCGCTGCGCCACG</u>	
Bkpt#2 <u>TTGTGGGGCCAGGACTTAGTGT</u> TTAA <u>ACTGTGATAAGTCTTACTAGTTTA</u>		Bkpt#2 <u>CCTCGGCCCTCCCTGAGTGTGGGAT</u> TT <u>TATAGGCGTGAGCCGCTGCGCCACG</u>	
	36,487,221		1,463,394
H2A31 3q13.31 deletion (200µM HU)	117,635,151	H3B11 3q13.31 deletion (300µM HU)	118,095,904
Bkpt#1 CAGTATGAATAACCCAGTTT <u>AGTCAAA</u> TAAATGTAAGTATATGCACATTG		Bkpt#1 TTAATAAATCAAGTTTTCAATCAA <u>AGACAAATAAAATCTGAAATACTGCAG</u>	
Del CAGTATGAATAACCCAGTTT <u>AGTCAAA</u> TAAATGTAAGTATATGCACATTG		Del TTAATAAATCAAGTTTTCAATCAA <u>AGAAATGTCACTTACAATGTCTTAGACC</u>	
Bkpt#2 <u>GTTTTACAAGTTTCTGTCTT</u> <u>TGAGAAA</u> <u>GAACCTCCACGCTGTGT</u> <u>TACCTCT</u>		Bkpt#2 <u>CAATACTAGATGAAAATTAGGCTC</u> <u>AGAAATGTCACTTACAATGTCTTAGACC</u>	
	117,889,426		118,200,900
H2C41 5p15.2 deletion (200µM HU)	9,128,809	NTB31 14q23.1 deletion (NT)	58,263,350
Bkpt#1 ACGGGTATCAGGGAGGTGTTGCAAA <u>AGG</u> AAATGAGGAAATGGGTTGATGGA		Bkpt#1 TGTTTCAGAGGCAGAGGCTCGGCT <u>GCAGTCACAAAACTAACAGAAAGAGA</u>	
Del ACGGGTATCAGGGAGGTGTTGCAAA <u>AGG</u> AAATGAGGAAATGGGTTGATGGA		Del TGTTTCAGAGGCAGAGGCTCGGCT <u>GCATCATAACAGCATGCAGCTCCTGAC</u>	
Bkpt#2 <u>AGGCAGTCTCTGTGGCTACTGGTCA</u> <u>AGT</u> <u>TAGCTGAACAGCAGCCATAGATG</u>		Bkpt#2 <u>TCTGTTTCCCACCTGTAACATGG</u> <u>GCATCATAACAGCATGCAGCTCCTGAC</u>	
	9,363,325		58,309,487
H2A31 13q21.1 deletion (200µM HU)	54,416,275	NTC21 3q13.31 deletion (NT)	117,153,488
Bkpt#1 CTGGAAGTCGCTCTGGGTAAGTCA <u>CTG</u> AGTGAAGTGGCGAGTCTTGTGAAG		Bkpt#1 CAAAAGAAGATATACAAATGGCCA <u>GCACAAATGAAAAATGCTCAACATCAC</u>	
Del CTGGAAGTCGCTCTGGGTAAGTCA <u>CTG</u> AGTGAAGTGGCGAGTCTTGTGAAG		Del CAAAAGAAGATATACAAATGGCCA <u>GCACATGCCTGTAATCCAGCTACTTG</u>	
Bkpt#2 <u>TTTTCTTTTTTGTATATGCCTTTGT</u> <u>CTG</u> <u>TTTTTGGTATCAGGGTAATACTGG</u>		Bkpt#2 <u>ATACAAAAATGGCTGGGCATGTG</u> <u>GCACATGCCTGTAATCCAGCTACTTG</u>	
	54,556,077		117,814,247
H2A61 2p22.2 deletion (200µM HU)	36,506,331	NTG51 7p15.1 deletion (NT)	31,756,287
Bkpt#1 TGGTGACAAATCCATTATCAGCAA <u>CATT</u> CTCATCAGAGGCTCAGATGAG		Bkpt#1 AGTGGAAATTTGGAAGAATAAAG <u>CAATT</u> AATGTGTGTGGAGAAGGAGATA	
Del TGGTGACAAATCCATTATCAGCAA <u>CATT</u> CTCATCAGAGGCTCAGATGAG		Del AGTGGAAATTTGGAAGAATAAAG <u>CAATT</u> CAAAAAGATCTCCACAGCACATT	
Bkpt#2 <u>CCAGTCCATGTAGAACAATATT</u> <u>CAAA</u> <u>CATT</u> <u>GCTGCTACTTTGGCCATCACCA</u>		Bkpt#2 <u>GCTCAAAGATTCACAAACAGATA</u> <u>CAATT</u> <u>CAAAAAGATCTCCACAGCACATT</u>	
	36,589,228		31,874,852
H2A61 18q22.1 deletion (200µM HU)	64,194,028	NTG51 2q32.3 deletion (NT)	196,823,287
Bkpt#1 GAAAAGTGTTTATCATTAAATTC <u>TGCA</u> TTTGAACAAGTAATTTGAGTCA		Bkpt#1 GACAGCTCTGCAAATCAGGGTTCG <u>TAT</u> TGTTCCAGTTGTTAATGGTTCAAAT	
Del GAAAAGTGTTTATCATTAAATTC <u>TGCA</u> TTTGAACAAGTAATTTGAGTCA		Del GACAGCTCTGCAAATCAGGGTTCG <u>TAT</u> ATGCTGGGAGTCTCTGGATCTTAAC	
Bkpt#2 <u>TGCAATTACAATGAGCGTGAAGA</u> <u>TGCA</u> <u>GTGCTGGAATAAGTTGACTAAGT</u>		Bkpt#2 <u>AGGCTCTGAAAAGGAAAGGCTCA</u> <u>TAT</u> <u>ATGCTGGGAGTCTCTGGATCTTAAC</u>	
	64,220,275		197,031,352
H2B11 10q21.1 deletion (200µM HU)	53,638,925	H1A51 3q27.3 duplication (100µM HU)	189,205,577
Bkpt#1 TGGAGTAGGGGTCATAAACTTTT <u>CTG</u> CAGTGAAGCTTTGTAGTGAATAT		Bkpt#1 TGTCAGAATGCCTGGTGGATGCCT <u>GGTGGGACACCTGCCAAGGCATACAG</u>	
Del TGGAGTAGGGGTCATAAACTTTT <u>CTG</u> CAGTGAAGCTTTGTAGTGAATAT		Dup TGTCAGAATGCCTGGTGGATGCCT <u>GGTGGGACACCTGCCAAGGCATACAG</u>	
Bkpt#2 <u>TTTAGGACCTTAGGATTGGAGGACAGGAT</u> <u>TGG</u> <u>GAAGGGGACGATTGAGTAG</u>		Bkpt#2 <u>TTTTAAATAATGCCTTTTTGAGGAA</u> <u>CTAGCT</u> <u>TAGGCTCTCTATCCTGGG</u>	
	53,671,033		188,457,031
H2C21 3q13.31 deletion (200µM HU)	118,084,102	H2B41 17p12 duplication (200µM HU)	14,348,849
Bkpt#1 TCAGTGTAATAATGAGAACTATA <u>AACTCCTAGGAGAACTCCTAGGATTG</u>		Bkpt#1 CCAGCCCTGGCTTCAATGCTCTGG <u>TT</u> CACCTCCGTCAGACTGGTCTCTCTC	
Del TCAGTGTAATAATGAGAACTATA <u>AACTCCTAGGAGAACTCCTAGGATTG</u>		Dup CCAGCCCTGGCTTCAATGCTCTGG <u>TT</u> TTTTTTTTTTTTTTTTTGAATTTCTAT	
Bkpt#2 <u>CCTCAGTCTCCCAAAGTCTGGGAT</u> <u>TATAGACATGAGCCACCATGCCTGGC</u>		Bkpt#2 <u>AATTTTTAATGGACTTACCCTAC</u> <u>TT</u> <u>TTTTTTTTTTTTTTTTTGAATTTCTAT</u>	
	118,135,927		14,349,353
H2C21 7q11.22 deletion (200µM HU)	68,890,457	NTA11 10q24.13 duplication (NT)	102,832,429
Bkpt#1 GCATATTTTTGAATAACCTTCACTA <u>GC</u> ATTATTATAGTAGTTATTTTTATT		Bkpt#1 GTAAGGGATCTCTCCAGGTTGGA <u>TTTT</u> CCCAAATAGATGAGACCCTCAAT	
Del GCATATTTTTGAATAACCTTCACTA <u>GC</u> ATTATTATAGTAGTTATTTTTATT		Dup GTAAGGGATCTCTCCAGGTTGGA <u>TTTT</u> TATTATTGAGTTGTATGTTCTGG	
Bkpt#2 <u>ACTGTTAGAATCAATATTTTTGTG</u> <u>GAAGCCAAAGTTTATAGGTAAGGCA</u>		Bkpt#2 <u>TATTTTTGAATATAGTTATTTGTC</u> <u>TT</u> <u>TTTTATTATTGAGTTGTATGTTCTGG</u>	
	69,202,211		102,710,396
H2C31 7p21.3 deletion (200µM HU)	7,996,602		
Bkpt#1 TATTTTAGAAAATAAAGCATAATGC <u>TAT</u> CCTCCAATATGTTATGACACTTT			
Del TATTTTAGAAAATAAAGCATAATGC <u>TAT</u> CCTCCAATATGTTATGACACTTT			
Bkpt#2 <u>CTAGACAGTTAAGTCTGGAGTCTG</u> <u>T</u> <u>TTCTTTAATACTGTGGATAATTGA</u>			
	8,078,982		

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.

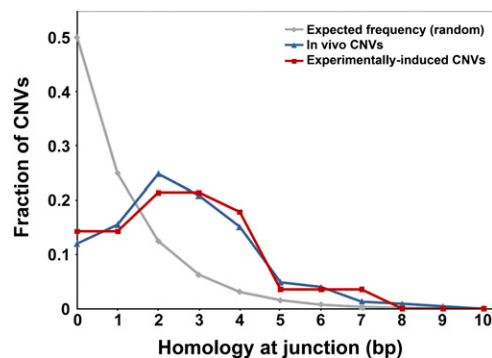


Fig. S4. Comparison of breakpoint sequences in hydroxyurea- and aphidicolin-induced copy number variants (CNVs) in vitro and germline CNVs in vivo. The frequencies of CNV breakpoints with 0–10 bp of microhomology are compared between all replication stress-induced CNVs from this and previous studies (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

CNVs in region (N)	Observed CNVs		Simulation mean		$P(>0)$	$P(R_{obs})$
	Regions (R_{obs})	CNVs	Regions (λ_N)	CNVs		
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 \(XLSX\)](#)