**Table S1. Human genome fosmid clone resource, related to Figure 2**

(A) Information from all 17 analyzed fosmid libraries is depicted. 'Best' placement refers to clones having both end-sequences mapping onto the reference genome with the proper orientation and separation. Variant sites were defined by multiple clones from the same individual that showed the same type of discordancy. Note that the later libraries have a smaller number of clones and have not been subjected to the same level of systematic validation as the initial libraries. (B) The number of structurally variant clones, as well as the nonredundant total of events, is given. A single sequence was chosen for further study when multiple clones captured the same variant. All sequence data have been deposited in GenBank and the NCBI trace archive under project ID 29893.

A

| Library | Population | Sample | Number of Reads | Number of Paired Clones | Average Q30 Bases | Total Q30 Bases | Estimated Sequence Coverage | Clones with 'Best' Placement | Mean *in silico* insert size (kb) | Estimated Physical Coverage (Best Clones) | StdDev (kb) | Size Threshold (kb) | Deletions | Insertions | Inversions |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| G248 | | NA15510 | 2,298,774 | 1,141,942 | 308 | 708,952,412 | 0.25 | 594,609 | 39.9 | 8.21 | 2.75 | 8.25 | 106 | 112 | 52 |
| ABC7 | Yoruba | NA18517 | 2,076,237 | 992,218 | 405 | 841,505,234 | 0.29 | 616,947 | 37.6 | 8.02 | 3.88 | 11.64 | 88 | 44 | 64 |
| ABC8 | Yoruba | NA18507 | 3,331,676 | 1,588,700 | 465 | 1,549,030,580 | 0.54 | 1,050,579 | 36.7 | 13.34 | 3.85 | 11.55 | 164 | 113 | 74 |
| ABC9 | Japan | NA18956 | 2,076,828 | 1,007,581 | 497 | 1,032,726,952 | 0.36 | 738,786 | 39.5 | 10.10 | 2.26 | 6.78 | 174 | 205 | 83 |
| ABC10 | Yoruba | NA19240 | 2,118,546 | 1,032,070 | 483 | 1,022,392,331 | 0.35 | 741,949 | 41.0 | 10.51 | 1.84 | 5.52 | 263 | 320 | 95 |
| ABC11 | China | NA18555 | 1,966,644 | 951,157 | 478 | 939,700,332 | 0.33 | 724,998 | 40.0 | 10.04 | 1.77 | 5.31 | 307 | 286 | 72 |
| ABC12 | CEPH | NA12878 | 2,168,656 | 1,039,478 | 464 | 1,005,800,422 | 0.35 | 755,087 | 39.8 | 10.39 | 1.4 | 4.2 | 290 | 299 | 83 |
| ABC13 | Yoruba | NA19129 | 2,053,392 | 1,009,530 | 528 | 1,083,273,138 | 0.37 | 757,837 | 39.3 | 10.30 | 1.77 | 5.31 | 268 | 291 | 91 |
| ABC14 | CEPH | NA12156 | 2,021,844 | 995,384 | 537 | 1,086,415,113 | 0.38 | 782,310 | 39.4 | 10.68 | 1.72 | 5.16 | 303 | 297 | 104 |
| JVI1 | China | NA18552 | 1,992,678 | 990,194 | 520 | 1,036,706,925 | 0.36 | 726,247 | 34.3 | 8.62 | 1.82 | 5.46 | 201 | 248 | 78 |
| ABC16 | Japan | NA18947 | 1,530,585 | 622,504 | 370 | 560,604,744 | 0.19 | 500,049 | 38.3 | 6.63 | 1.45 | 4.35 | 164 | 175 | 38 |
| ABC18 | CEPH | NA10847 | 1,203,756 | 444,663 | 369 | 444,126,739 | 0.15 | 361,412 | 39.2 | 4.90 | 1.81 | 5.43 | 21 | 20 | 6 |
| ABC21 | CEPH | NA11993 | 678,662 | 239,748 | 381 | 258,415,799 | 0.09 | 202,343 | 39.6 | 2.77 | 1.65 | 4.95 | 48 | 53 | 13 |
| ABC22 | CEPH | NA11840 | 777,200 | 256,967 | 343 | 266,283,260 | 0.09 | 212,681 | 38.7 | 2.85 | 1.91 | 5.73 | 41 | 24 | 11 |
| ABC23 | Yoruba | NA18523 | 1,504,416 | 566,334 | 379 | 569,946,866 | 0.20 | 467,955 | 38.7 | 6.26 | 1.72 | 5.16 | 107 | 113 | 32 |
| ABC24 | Yoruba | NA18502 | 1,380,651 | 544,470 | 368 | 507,621,602 | 0.18 | 453,587 | 38.9 | 6.11 | 1.52 | 4.56 | 167 | 165 | 39 |
| ABC27 | Japan | NA18942 | 1,227,159 | 458,599 | 350 | 429,089,775 | 0.15 | 374,887 | 38.9 | 5.05 | 1.43 | 4.29 | 126 | 131 | 25 |
| **Total** | | | **30,407,704** | **13,881,539** | **439** | **13,342,592,224** | **4.62** | **10,062,263** | | **134.79** | | | **2,838** | **2,896** | **960** |
| **Total (non-redundant)** | | | | | | | | | | | | | ***899*** | ***908*** | ***244*** |

B

| Sample | Population | Deletions | Insertions | Inversions |
|---|---|---|---|---|
| NA15510 | -- | 62 | 27 | 29 |
| NA18517 | Yoruba | 13 | 2 | 11 |
| NA18507 | Yoruba | 22 | 4 | 6 |
| NA18956 | Japan | 182 | 174 | 7 |
| NA19240 | Yoruba | 89 | 89 | 14 |
| NA18555 | China | 80 | 40 | 12 |
| NA12878 | CEPH | 59 | 31 | 10 |
| NA19129 | Yoruba | 69 | 41 | 12 |
| NA12156 | CEP | 55 | 19 | 8 |
| Total (nonredundant events) | | 589 | 384 | 81 |

**Table S4. Repeat enrichment at homologous breakpoints, related to Figure 4**

Enrichment for variants having the indicated elements at both breakpoints was calculated relative to breakpoint pairs randomly sampled from the genome. Analysis was limited to the 590 class I insertion-deletion variants (A) and 74 class I inversions (B). Alternatively, an event was counted if at least one of the breakpoints mapped within the indicated repeat class (C and D). Enrichment was calculated relative to random genome sampling.

**A Insertions and Deletions**

| Elements at Breakpoints | Number of Events | Percent of Events | Enrichment |
|---|---|---|---|
| Alu - Alu | 114 | 19.3% | 5.2 |
| SegDup - SegDup | 84 | 14.2% | 3.2 |
| L1 - L1 | 74 | 12.5% | 2.0 |
| LTR - LTR | 36 | 6.1% | 2.9 |

**B Inversions**

| Elements at Breakpoints | Number of Events | Percent of Events | Enrichment |
|---|---|---|---|
| Alu - Alu | 16 | 21.6% | 1.3 |
| SegDup - SegDup | 39 | 52.7% | 14.3 |
| L1 - L1 | 29 | 39.2% | 2.6 |
| LTR - LTR | 10 | 13.5% | 1.6 |

**C Insertions and Deletions**

| Repeat | Number of Events | Percent of Events | Enrichment |
|---|---|---|---|
| Alu | 192 | 32.5% | 1.4 |
| AluY | 59 | 10.0% | 2.6 |
| AluSg | 25 | 4.2% | 1.8 |
| AluSp | 15 | 2.5% | 1.8 |
| AluSq | 24 | 4.1% | 1.6 |
| AluSx | 70 | 11.9% | 1.4 |
| SegDup | 92 | 15.6% | 2.5 |
| LTR-ERVK | 13 | 2.2% | 3.4 |

**D Inversions**

| Repeat | Number of Events | Percent of Events | Enrichment |
|---|---|---|---|
| Alu | 20 | 27.0% | 0.6 |
| AluY | 6 | 8.1% | 0.7 |
| AluSg | 0 | 0.0% | 0.0 |
| AluSp | 2 | 2.7% | 0.5 |
| AluSq | 5 | 6.8% | 0.8 |
| AluSx | 9 | 12.2% | 0.6 |
| SegDup | 41 | 55.4% | 7.8 |
| LTR-ERVK | 1 | 1.4% | 1.0 |

**Table S5. Mapped location of additional junction sequences, related to Figure 4**

The matching genomic location for junction sequences from class II deletions is given.

| Variant | Deletion Coordinates | | | Match Position for Junction Sequence | | | Seperation |
|---|---|---|---|---|---|---|---|
| AC212490 | chr3 | 53,002,188 | 53,013,939 | chr3 | 53,002,158 | 53,002,186 | 2 |
| AC207966 | chr1 | 185,731,451 | 185,733,353 | chr1 | 185,733,350 | 185,733,102 | 3 |
| AC215277 | chr8 | 75,525,426 | 75,529,555 | chr8 | 75,529,518 | 75,529,568 | 13 |
| AC210963 | chr7 | 156,079,934 | 156,087,081 | chr7 | 156,079,888 | 156,079,845 | 46 |
| AC217324 | chr3 | 133,190,943 | 133,196,075 | chr3 | 133,191,100 | 133,191,051 | 108 |
| AC214988 | chrX | 146,651,366 | 146,657,504 | chrX | 146,656,838 | 146,657,187 | 317 |
| AC226181 | chr2 | 216,798,188 | 216,801,357 | chr2 | 216,800,757 | 216,801,035 | 322 |
| AC210709 | chr2 | 88,941,533 | 89,223,181 | chr2 | 88,942,038 | 88,942,188 | 505 |
| AC203599 | chr12 | 45,314,843 | 45,319,576 | chr12 | 45,316,890 | 45,316,737 | 1,894 |
| AC216238 | chr4 | 115,727,270 | 115,733,314 | chr4 | 115,730,492 | 115,730,552 | 2,762 |
| AC226060 | chr5 | 114,282,831 | 114,289,330 | chr5 | 114,286,135 | 114,285,984 | 3,153 |
| AC226700 | chr3 | 147,867,880 | 147,873,095 | chr3 | 147,877,554 | 147,877,855 | 4,459 |
| AC215328 | chr3 | 165,784,887 | 165,794,633 | chr3 | 165,789,417 | 165,789,506 | 4,530 |
| AC225387 | chr6 | 69,744,415 | 69,748,554 | chr6 | 69,739,284 | 69,739,813 | 4,602 |
| AC207578 | chr4 | 116,386,356 | 116,396,630 | chr4 | 116,391,481 | 116,391,406 | 5,050 |
| AC207431 | chr13 | 29,113,837 | 29,126,124 | chr13 | 29,119,859 | 29,120,118 | 6,006 |
| AC206483 | chr9 | 112,064,397 | 112,069,787 | chr9 | 112,076,915 | 112,077,102 | 7,128 |
| AC211773 | chr2 | 4,190,551 | 4,201,354 | chr2 | 4,183,309 | 4,183,144 | 7,242 |
| AC193142 | chr11 | 55,121,095 | 55,214,327 | chr11 | 55,191,072 | 55,188,127 | 23,255 |
| AC193145 | chr14 | 105,311,004 | 105,451,267 | chr14 | 105,398,170 | 105,401,512 | 49,755 |
| AC231961 | chr1 | 191,237,059 | 191,260,247 | chr1 | 191,408,901 | 191,408,515 | 148,268 |
| AC212839 | chr19 | 9,135,510 | 9,145,364 | chr19 | 8,903,157 | 8,903,200 | 232,310 |
| AC226629 | chr1 | 246,740,696 | 246,749,190 | chr10 | 116,925,675 | 116,925,960 | NA |
| AC217141 | chr12 | 9,524,170 | 9,623,304 | chr5 | 64,495,645 | 64,498,607 | NA |
| AC215990 | chr7 | 93,165,078 | 93,170,218 | chrX | 120,157,245 | 120,156,932 | NA |
| AC213239 | chr2 | 79,185,033 | 79,193,091 | chrY | 57,445,023 | 57,444,967 | NA |
| AC209239 | chr19 | 20,387,650 | 20,509,814 | chr6 | 53,277,247 | 53,277,296 | NA |
| AC209204 | chr2 | 208,059,397 | 208,067,617 | chrY | 57,431,660 | 57,431,686 | NA |
| AC208508 | chr13 | 22,530,435 | 22,535,303 | chrY | 26,900,639 | 26,900,712 | NA |
| AC196529 | chr1 | 143,607,657 | 143,618,167 | chr1_random | 580,876 | 581,009 | NA |

**Table S6. Mechanism annotation for events sequenced in this study and targeted or captured in Conrad et al., 2010, related to Table 1**

| | | Annotation based on sequenced fosmids | | | |
|---|---|---|---|---|---|
| | **Total** | **VNTR** | **Class II** | **Microhomology or no homology (Class I, <=20 bp homology)** | **NAHR (Class I, >20 bp homology)** |
| Targeted | 237 | 14 | 56 | 98 | 69 |
| Has sequence from capture | 70 | 0 | 21 | 49 | 0 |

**Table S7. Genotyping structural variants using diagnostic k-mers, related to Figure 1**

A sequence tag corresponding to the breakpoint was tested for its ability to uniquely identify and, therefore, genotype each variant allowing either one mismatch (e1) or no mismatch (e0) when compared to the human reference sequence. The number of variants uniquely typable for various k-mer lengths is given. Analysis was limited to the 1,024 events not classified as VNTRs.

| Variant Type | Total Variants | Search Threshold | k=36 | k=50 | k=75 | k=100 |
|---|---|---|---|---|---|---|
| Deletion | 583 | e1 | 429 | 464 | 499 | 519 |
| | | e0 | 489 | 523 | 551 | 559 |
| Insertion | 360 | e1 | 265 | 295 | 328 | 338 |
| | | e0 | 309 | 331 | 347 | 351 |
| Inversion | 81 | e1 | 32 | 35 | 43 | 45 |
| | | e0 | 51 | 56 | 60 | 62 |

**Table S8. Related to supplementary experimental procedures**

(A) Unresolved variants that affect genes (B) Motif breakpoint discrimination. The area under the ROC curve (the ROC score) was used as a metric to determine whether classes of breakpoints could be readily distinguished based on frequency of k-mers (for k = 1 to 6) in a 200-bp window centered on one breakpoint from each event. The mean and standard deviation ROC score from 10-fold cross validation is shown. (C) K-mers enriched for events with homologous breakpoints (>10 bp) The top 12 k-mers that discriminate between microhomology (1-10 bp) and extended homology (>10 bp) containing insertion-deletion variants are shown. In each case, the k-mer overrepresented in breakpoints having >10 bp of homology is shown. The ROC score is given for each k-mer separately. Note that no single k-mer has as much discriminating ability as the combined set.

A

| Clone Accession | Classification | Genes |
|---|---|---|
| AC215346 | Duplicated Sequences | *AMY1A* |
| AC208386 | Tandem Insertion | *APBA1* |
| AC213216 | Duplicated Sequences | *CCL3L1* |
| AC153478 | Tandem Insertion | *CLK2,HCN3* |
| AC216062 | Tandem Insertion | *CLPS* |
| AC208162 | Duplicated Sequences | *CYP2D6* |
| AC193104 | Duplicated Sequences | *DEFA1* |
| AC193152 | Duplicated Sequences | *DHRS4,DHRS4L2* |
| AC193093 | Duplicated Sequences | *FCGBP* |
| AC193097 | VNTR | *HRNR* |
| AC196503 | Duplicated Sequences | *LPA* |
| AC193138 | Duplicated Sequences | *NBPF10* |
| AC196505 | Duplicated Sequences | *NBPF14* |
| AC195774 | Duplicated Sequences | *PGA4,PGA3* |
| AC212592 | Duplicated Sequences | *PSG1,PSG6,PSG7,PSG11* |
| AC153472 | Duplicated Sequences | *REXO1L1* |
| AC213264 | Duplicated Sequences | *TCEB3C,TCEB3CL* |
| AC203621 | Tandem Insertion | *UCHL3* |

B

| Comparison | Mean ROC | ROC StdDev |
|---|---|---|
| zero bp vs 1-10 bp microhomology | 0.487 | 0.121 |
| 1-10 bp microhomology vs >10 bp homology | 0.761 | 0.069 |
| microhomology vs 10X genome sampling | 0.521 | 0.053 |
| >10 bp homology vs 10X genome sampling | 0.786 | 0.049 |

C

| *k*-mer | ROC Score | Direction |
|---|---|---|
| C/G | 0.73 | + |
| CC/GG | 0.73 | + |
| CG | 0.71 | + |
| CGC/GCG | 0.71 | + |
| CCC | 0.69 | + |
| CCG/CGG | 0.69 | + |
| GCC/GGC | 0.68 | + |
| GC | 0.68 | + |
| CCCG/CGGG | 0.67 | + |
| CGCC/GGCG | 0.67 | + |
| CACC | 0.67 | + |
| CTCC | 0.67 | + |