SUPPLEMENTAL TABLES AND FIGURES

Supplemental Table 1. Comparison of pFosill vectors. Cloning vectors described in this work are available to academic researchers upon request.

vector	**pFosill-1**	**pFosill-2**	**pFosill-3**	**pFosill-4**
ILMN-F sequencing primer	SBS-3	SBS-3	SBS-3	SBS-3
ILMN-R sequencing primer	SBS-8	SBS-12	SBS-8	SBS-12
Multiplex Fosmid cloning[a]	No	No	Yes	Yes
Multiplex sequencing[b]	No	Yes	No	Yes
Useful cloning sites	*Eco*72I	*Eco*72I	*Sap*I	*Sap*I
Vector-derived bases at the beginning of sequencing reads	Yes[c]	Yes[c]	No[d]	No[d]
Optional cloning sites	*Bam*HI *Xcm*I *Sph*I	*Bam*HI *Xcm*I *Sph*I	*Bam*HI	*Bam*HI
Useful nicking sites	Nb.*Bbv*CI	Nb.*Bbv*CI	Nb.*Bbv*CI Nt.*Bsp*QI	Nb.*Bbv*CI Nt.*Bsp*QI
GenBank accession number	JX069761	JX069762	JX069763	JX069764

[a]Inline barcodes introduced by ligating *Sap*I adapters to genomic DNA fragments and read at the beginning of each paired-end sequencing read

[b]Barcodes introduced by tailed PCR primers at the end of the Fosill conversion process and read by a dedicated sequencing read away from the insert into the Illumina adapter

[c]Sequencing reads primed with standard Illumina primers start with four constant vector-derived bases. It is advisable to sequence Illumina libraries with such "mono-templates" at reduced cluster densities. Otherwise, the density of fluorescent clusters at the beginning of the read may exceed the limits of the imaging system for distinguishing and locating clusters during the first few sequencing cycles.

[d]Sequencing reads begin with the barcodes. To avoid "mono-template" issues during sequencing it is advisable to use more than one barcode even when multiplexing of the library is not required.

Supplemental Table 2. Detailed breakdown of sequencing reads

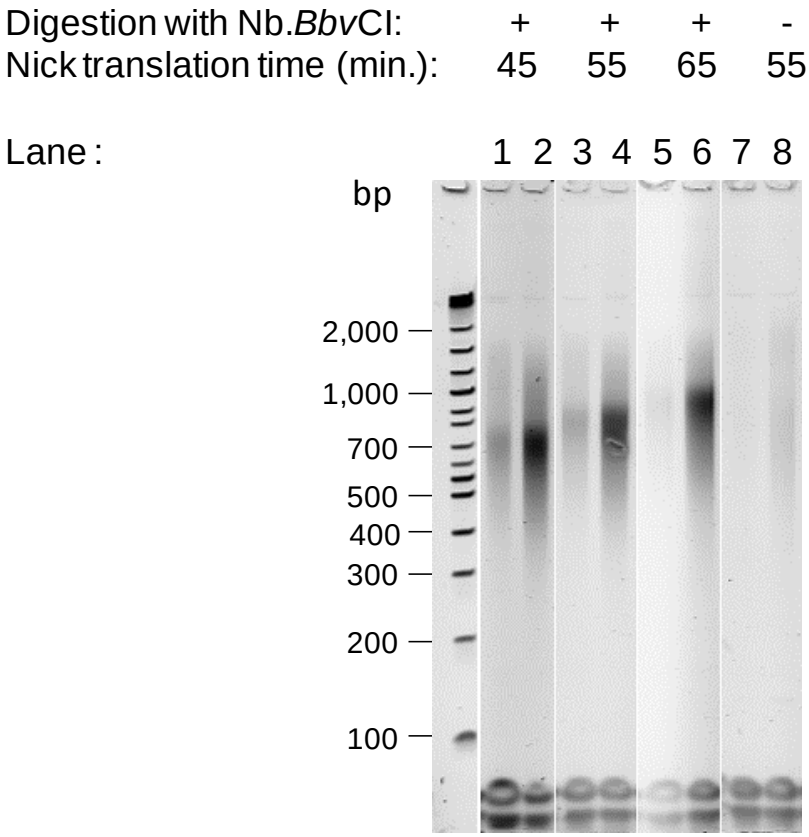
Organism:	**S. pombe**		**Human (K-562)**		**Mouse (C57B6J)**	
Estimated size of Fosmid library (million cfu)	1.4		6.6		7.5	
Fosill library	S		H1	H2	M	
Size selection	1x prep gel		low range	high range	2x prep gel	
Illumina GAII lanes	2		2	2	2	
Read length (bases)	101		101	101	101	
Total read pairs	**63,679,610**	***100%***	**46,349,056**	***100%***	**13,583,577**	***100%***	**23,599,670**	***100%***
Unaligned pairs	42,633,191	*67%*	3,979,087	*9%*	2,038,668	*15%*	2,970,052	*13%*
Only one read aligned	2,246,406	*4%*	7,871,684	*17%*	1,416,895	*10%*	1,298,900	*6%*
Multiple placements	694,036	*1%*	621,482	*1%*	396,288	*3%*	581,726	*2%*
Unambiguously placed pairs	18,105,977	*28%*	33,876,803	*73%*	9,731,726	*72%*	18,748,992	*79%*
Unambiguously placed pairs	**18,105,977**	***100%***	**33,876,803**	***100%***	**9,731,726**	***100%***	**18,748,992**	***100%***
Chimeric pairs	156,793	*0.9%*	706,750	*2.1%*	300,878	*3.1%*	180,757	*1.0%*
inverted orientation	2,320		39,193		39,927		3,493	
tandem orientation	3,487		13,420		7,058		4,857	
linking 2 reference contigs	99,084		596,381		229,919		155,534	
spacing >100 kb	51,902		57,756		23,974		16,873	
Spaced <30 kb or 50-100 kb	865,834	*4.8%*	2,836,671	*8.4%*	391,759	*4.0%*	158,105	*0.8%*
Pairs with correct spacing and orientation	17,083,349	*94.4%*	30,333,382	*89.5%*	9,039,089	*92.9%*	18,410,129	*98.2%*
Total unique pairs	**1,708,254**	***100%***	**6,956,308**	***100%***	**4,247,777**	***100%***	**5,866,705**	***100%***
Unique chimeric pairs	79,674	*4.7%*	385,992	*5.5%*	198,928	*4.7%*	140,335	*2.4%*
inverted orientation	736		15,085		18,883		1,705	
tandem orientation	1,201		3,840		3,292		1,996	
linking 2 reference contigs	51,685		340,657		161,982		125,674	
spacing >100 kb	26,052		26,410		14,771		10,960	
Unique pairs spaced 1-30 kb or 50-100 kb	52,727	*3.1%*	126,249	*1.8%*	80,397	*1.9%*	39,015	*0.7%*
Unique pairs spaced <1 kb	107,479	*6.3%*	933,693	*13.4%*	182,237	*4.3%*	55,645	*0.9%*
Unique pairs with correct spacing and orientation	1,468,374	*86.0%*	5,510,374	*79.2%*	3,786,215	*89.1%*	5,631,710	*96.0%*
Unique pairs with correct spacing and orientation	**1,468,374**		**6,933,573**		**5,631,710**	

Supplemental Table 3. Pair-wise matrix of inline barcodes at the beginning of read pairs derived from a Fosmid library prepared by pooling aliquots of differently tagged (*Sap*I-adapter-ligated), non-size selected sheared mouse DNA. Concordant read pairs are in bold type face.

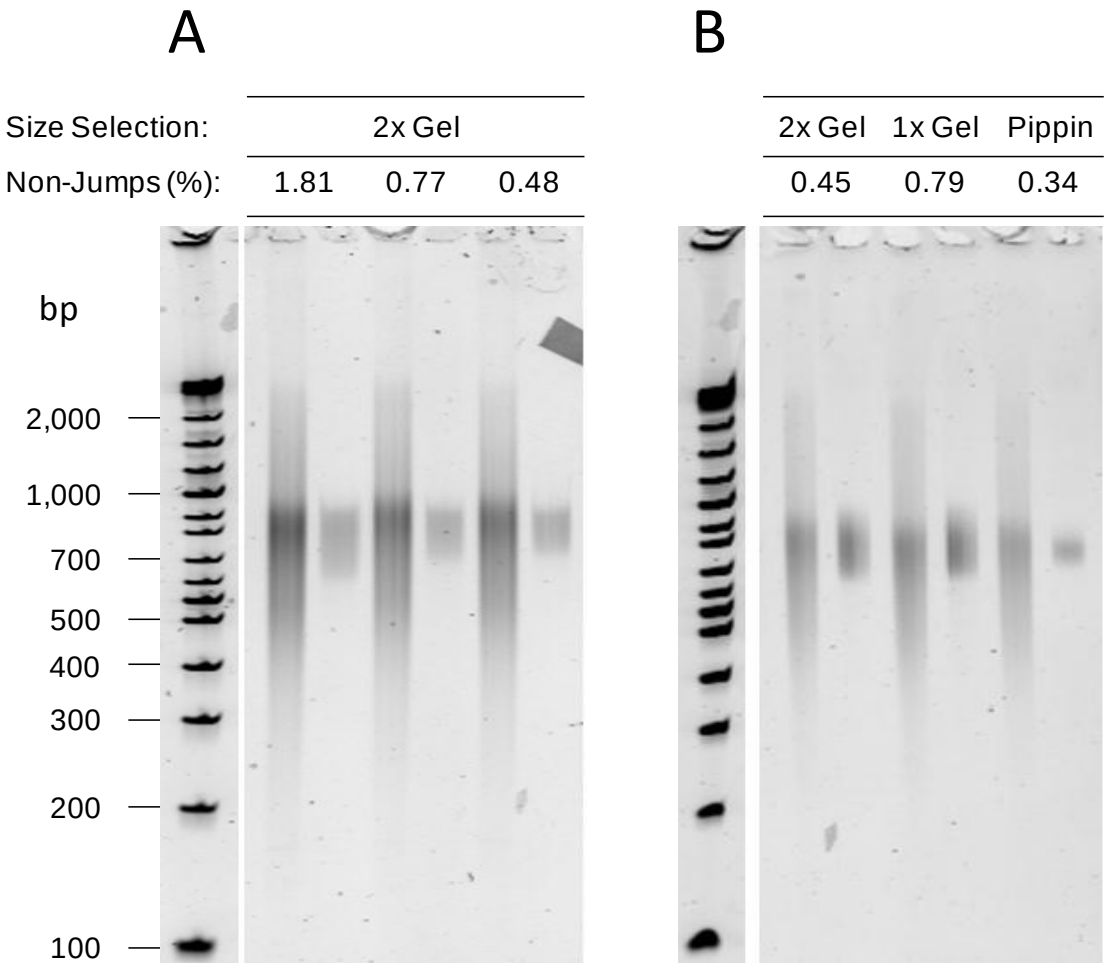
			READ 1			
	Barcode		A	B	C	D
		Sequence	AGTTGCTT	CCAGTTAG	CTACCAGG	TTGAGCCT
READ 2	A	AGTTGCTT	**9,004,871**	13,256	20,704	19,976
	B	CCAGTTAG	12,346	**6,442,973**	15,813	17,305
	C	CTACCAGG	21,245	15,880	**9,877,836**	24,329
	D	TTGAGCCT	19,469	15,580	23,796	**8,294,202**

Supplemental Table 4: Structural rearrangements detected in the K-562 cell line by *Fosill* sequencing. Rearrangements were ranked by the number of supporting unique read pairs. Events supported by less than 10 unique *Fosill* jumps were not included. Seven rearrangements were translocations (chromosome pairs involved denoted under headings chr1 and chr2). Intrachromosomal events were classified as tandem duplications (n=6), deletions (n=3), inversions (n=1), or simply as "long range" (n=4) when read pairs spanned >1 Mb in the reference genome. Forward and reverse reads aligning to the same strand of the reference genome sequence (columns str1 and str2) indicates an inversion. Notable gene features on either side of the breakpoints are listed under site1 and site2.

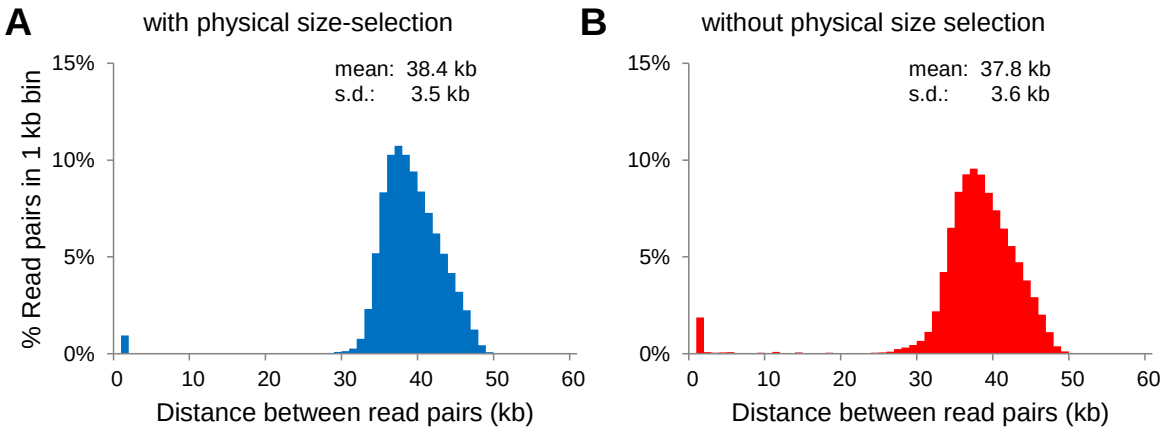
Rank	Read pairs	Rearrangement	chr1	str1	pos1	chr2	str2	pos2	site1	span (bp)	site2	Potential aberrant protein
1	887	translocation	9	1	133,607,091	22	0	23,632,602	Intron of *ABL1*	N/A	Intron of *BCR*	Fusion: in frame (BCR-ABL1)
2	131	inversion	9	0	131,109,854	9	0	131,464,347	Intron of *SLC27A4*	354,493	Promoter of *PKN3*	Antisense fusion
3	130	tandem duplication	6	1	31,619,152	6	0	31,833,732	Intron of *BAT3*	214,580	Exon 14 of *SLC44A4*	Fusion: mid-exon (BAT3-SLC44A4)
4	110	tandem duplication	16	1	88,592,357	16	0	88,860,888	Intron of *ZFPM1*	268,531	Intergenic	-
5	75	long range	9	1	123,552,783	9	0	131,103,311	5'-UTR of *FBXW2*	7,550,528	5'-UTR of *SLC27A4*	Antisense fusion
6	61	translocation	9	0	134,155,051	13	0	108,661,074	Intergenic	N/A	Intergenic	-
7	55	deletion	10	0	103,603,092	10	1	103,799,734	Intron of *KCNIP2*	196,642	Intron of *C10orf76*	Fusion: out of frame (C10orf76-KCNIP2)
8	51	translocation	3	0	48,228,049	10	0	87,847,570	Intron of *CDC25A*	N/A	Intron of *GRID1*	-
9	36	tandem duplication	17	1	26,900,245	17	0	27,272,846	3'-UTR of *ALDOC*	372,601	Intron of *PHF12*	-
10	29	deletion	10	0	126,639,489	10	1	126,731,062	Intron of *ZRANB1*	91,573	5'-UTR of *CTBP2*	Antisense fusion
11	25	tandem duplication	1	1	6,452,708	1	0	6,617,700	Intron of *ACOT7*	164,992	Intron of *TAS1R1*	Antisense fusion
12	23	long range	13	0	94,022,128	13	1	108,503,621	Intron of *GPC6*	14,481,493	Intron of *FAM155A*	Antisense fusion
13	22	translocation	6	1	38,108,755	16	1	78,260,859	Intron of *ZFAND3*	N/A	Intron of *WWOX*	-
14	20	tandem duplication	4	1	10,106,099	4	0	10,329,568	Intron of *WDR1*	223,469	Intergenic	-
15	18	translocation	9	0	134,074,293	22	0	17,300,213	Exon 29 of *NUP214*	N/A	5'-UTR of *XKR3*	Fusion: mid-exon (NUP214-XKR3)
16	15	deletion	17	0	64,347,281	17	1	64,506,647	Intron of *PRKCA*	159,366	Intron of *PRKCA*	Deletion of 1 exon: out of frame
17	11	translocation	6	1	16,753,402	16	0	85,643,398	5'-UTR of *ATXN1*	N/A	Promoter of *KIAA0182*	Antisense fusion
18	11	long range	9	1	123,565,059	9	0	132,658,152	Intergenic	9,093,093	Exon 16 of *FNBP1*	-
19	11	translocation	9	0	139,634,410	23	1	63,755,308	Exon 5 of *LCN10*	N/A	Intergenic	-
20	11	tandem duplication	14	1	23,217,247	14	0	23,806,209	Intergenic	588,962	Intergenic	-
21	11	long range	18	1	480,258	18	1	3,474,872	Intron of *COLEC12*	2,994,614	Intergenic	-



Supplemental Figure 1. Time course of nick translation. Fosmid DNA (mouse library M) was nicked by incubation with Nb.*Bbv*CI. A mock-nick control reaction contained Fosmid DNA but no Nb.*Bbv*CI (lanes 7 and 8). The reaction products were subjected to a nick-translation reaction at 0˚C for 45, 55 or 65 min. followed by S1 digestion, end-repair and dilute ligation to circularize DNA fragments. Small-scale test inverse PCR reactions were run for 15 (lanes 1, 3, 5, 7) or 18 cycles (lanes 2, 4, 6, 8). Amplification products were analyzed on a SYBR-green-stained 5% non-denaturing polyacrylamide gel. The longer the time of nick translation, the longer the PCR products which form a long smear.



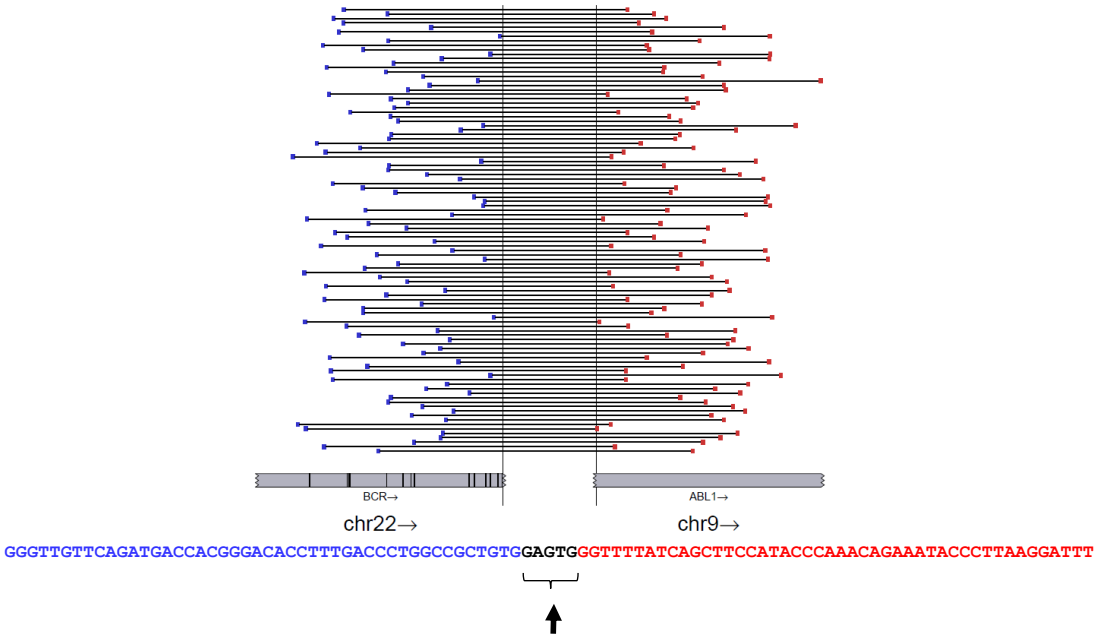
Supplemental Figure 2: Percent "non-jumps" in six differently size-selected *Fosill* libraries. The two SYBR-green-stained analytical 5% polyacrylamide gels show the size distribution of *Fosill* libraries before (left lane of each pair) and after electrophoretic size selection. 1x and 2x Gel indicates one and two consecutive preparative agarose gels, respectively. "Pippin" is an automated preparative gel-electrophoresis system. (*A*) Narrowing the size distribution by raising the lower cut-off size from about 450 to 550 bp decreased the percentage of unique "non-jumps", *i.e.*, read-pairs spanning less than 1 kb in the human reference genome, from 1.81% to 0.48%. (*B*) A single narrow size selection on the Pippen instrument (right-most lane) is at least as effective as two consecutive preparative agarose gels and currently our preferred protocol.



Supplemental Figure 3: Comparison of length-distributions of Fosill jumps derived from Fosmid libraries constructed with or without physical size selection of the sheared genomic DNA on a pulsed-field gel. Shown is the distribution of unique read pairs in mouse Fosill library M (A) and in the mouse Fosill sub-library tagged with barcode AGTTGCTT (B). The mean and standard deviation of *bona fide* Fosmid-sized jumps (30-50 kb) are indicated.

A

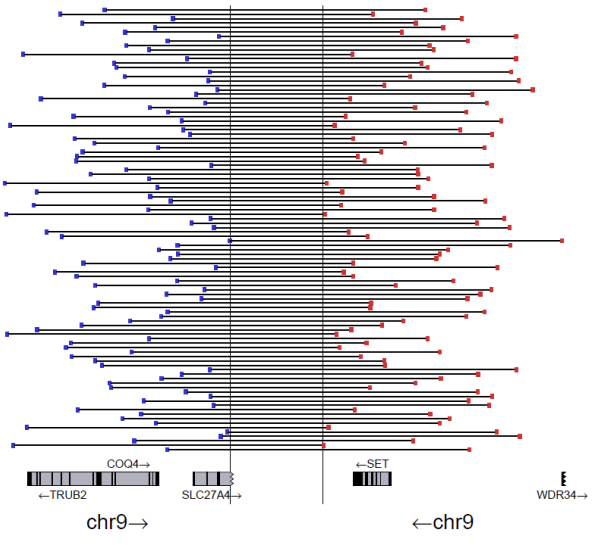
Translocation (9;22): BCR-ABL1



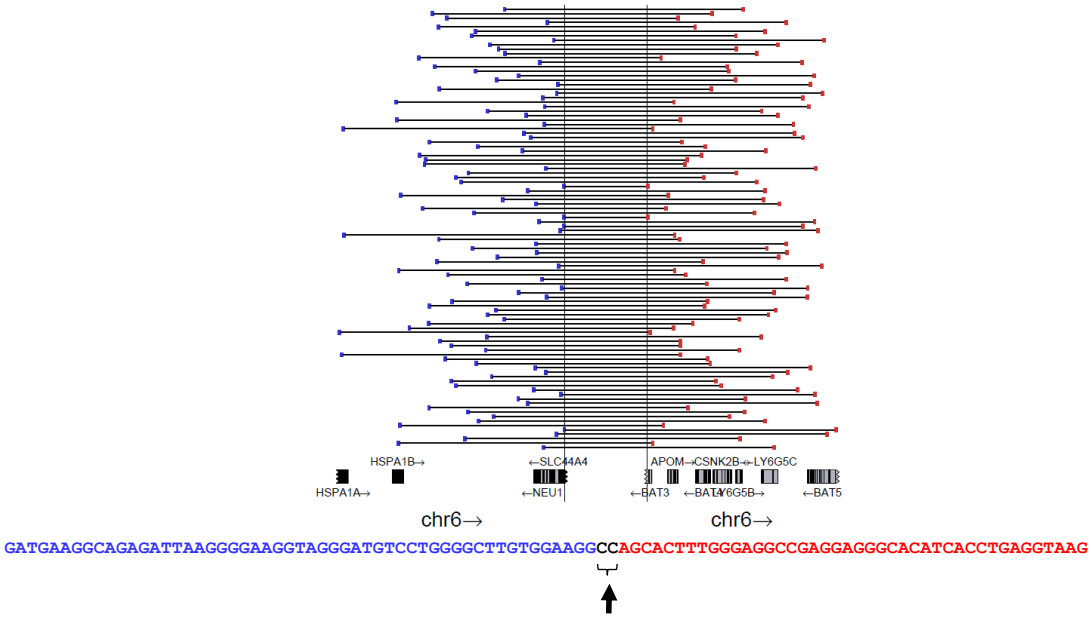
chr22:23632738-23632742/ /chr9:133607147-133607151

B

Inversion on chr 9



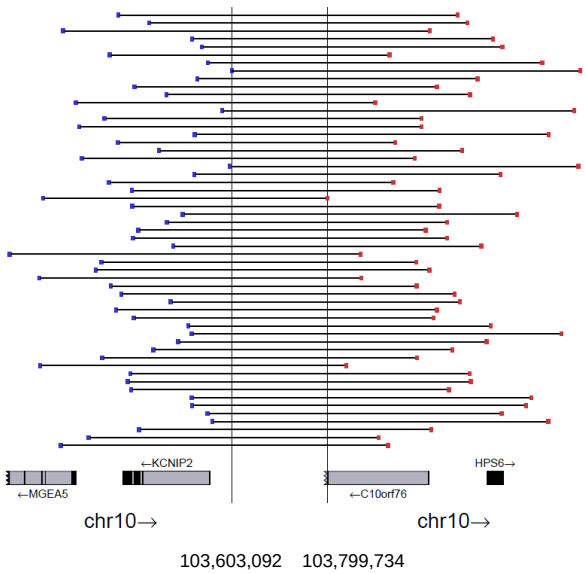
C Tandem duplication on chr 6: BAT3-SLC44A4

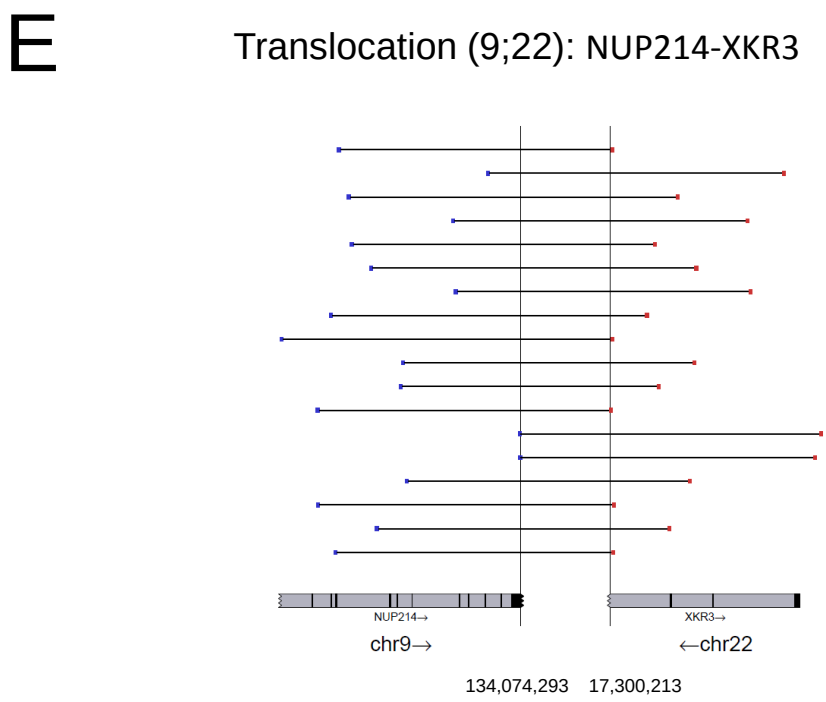


Breakpoint

chr6:31833714-31833715//chr6:31619064-31619065

D Deletion on chr 9





Supplemental Figure 4. dRanger maps showing the read pairs supporting the five structural rearrangements listed in Table 1 of the main text. (*A*) The t(9;22) translocation in the K-562 genome that gives rise to the *BCR-ABL* fusion gene was bracketed by a total of 887 unique *Fosill* read pairs of which only a random subset is shown. The break point itself was covered by 32 reads. One of them is shown at the bottom with sequence from chromosome 22 and chromosome 9 distinguished by blue and red type face, respectively. Due to micro homology of the two sequences involved in the translocation event, the five bases denoted in black cannot be assigned to one chromosome. (*B*) Shown is the map position of all 131 unique read pairs implicating an inversion on chromosome 9. No single read spanning the breakpoint could be identified in this case. The read starts of chimeric read pairs closest to the break point are indicated (chr9:131,109,854 and chr9:131,464,347). (*C*) A tandem duplication on chromosome 6 that gives rise to a *BAT3-SLC44A4* fusion gene (Berger et al. 2010) was framed by 130 unique read pairs. The break point was covered by 7 single reads. The micro homology between the fusion partners is 2 bp. (*D*) Fifty-five read pairs spanning more than 100 kb in the reference genome indicated a ~200-kb deletion on chromosome 10. (*E*) Twenty-two *Fosill* read pairs defined a 2nd t(9;22) translocation which gives rise to a fusion gene involving *NUP214* and *XKR3* (Levin et al. 2009; Berger et al. 2010). No single read crossing the translocation breakpoint was found.