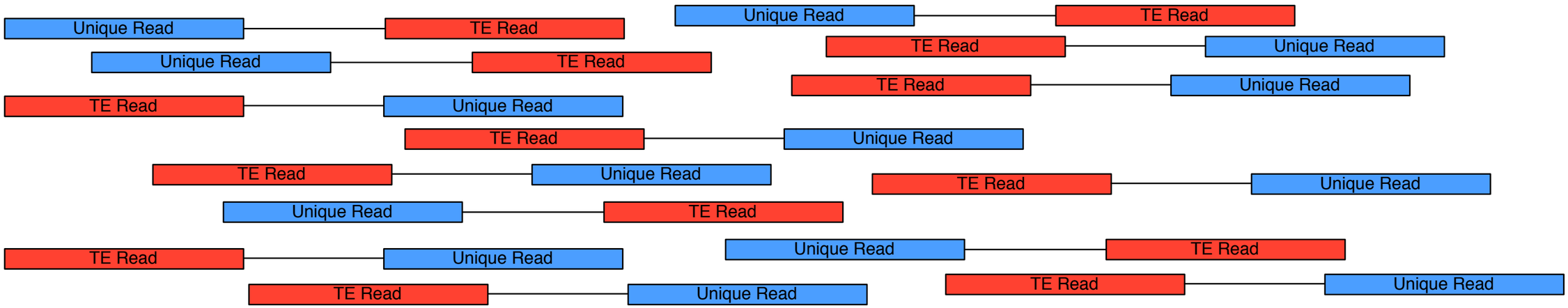
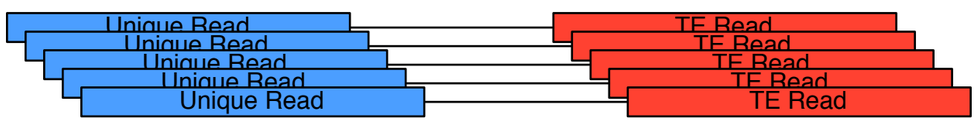
Supplementary Figure 1

1) Identify read-pairs where one read aligns uniquely to the genome and the other read aligns to a TE.

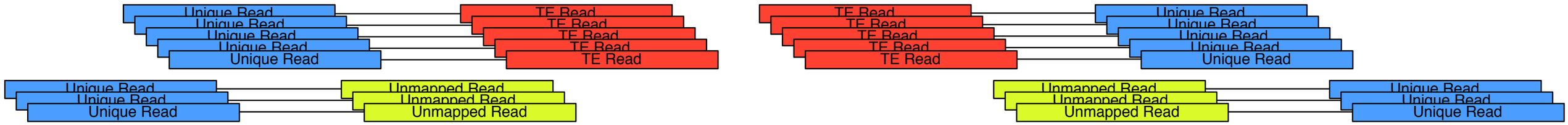


2) Cluster read pairs based on the alignment of the unique read.



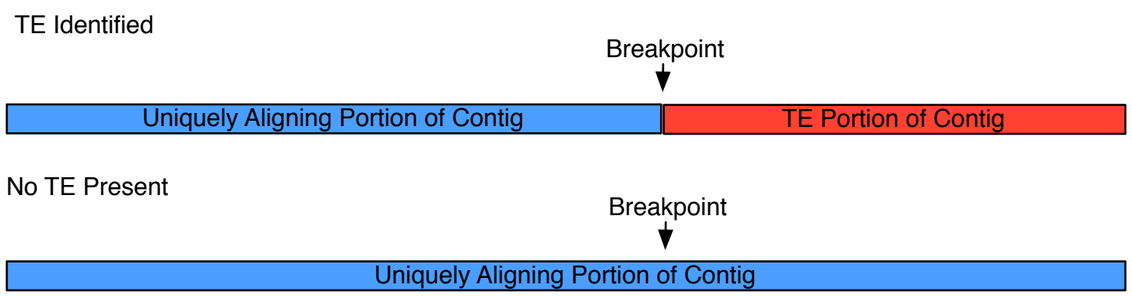
3) Join clusters where the unique reads are on the plus strand to clusters where the unique reads are on the minus strand to produce events.

4) Identify additional reads that align near the clustered read pairs.

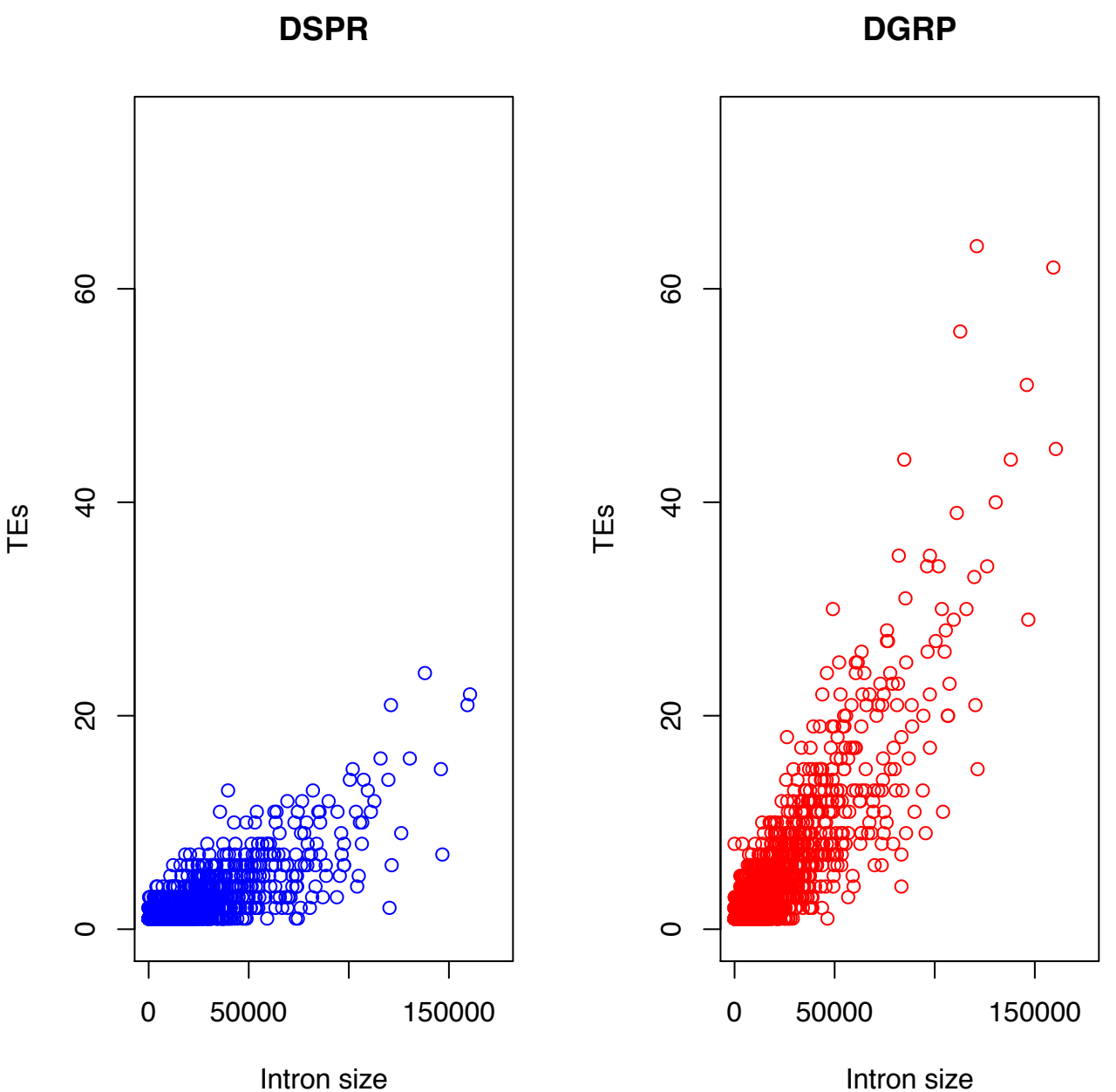


5) Run Phrap to generate contigs from reads.

6) Generate a list of breakpoints identified in any line. Extract reads and construct contigs for each identified breakpoint in every line. Check contigs for evidence of a TE breakpoint or a contig that crosses the breakpoint indicating no TE insertion.



Supplementary Figure 2 : Number of TEs vs Intron size.



Supplementary Table 1: Included DGRP Lines

Line	Included	DGRP ≥ 25
RAL-21	Yes	No
RAL-26	Yes	No
RAL-28	No	No
RAL-38	No	No
RAL-40	Yes	Yes
RAL-41	No	No
RAL-42	Yes	No
RAL-45	Yes	No
RAL-49	Yes	No
RAL-57	Yes	Yes
RAL-59	Yes	No
RAL-69	Yes	No
RAL-73	Yes	No
RAL-75	Yes	No
RAL-83	Yes	No
RAL-85	Yes	No
RAL-88	Yes	No
RAL-91	Yes	No
RAL-93	Yes	No
RAL-101	Yes	No
RAL-105	Yes	No
RAL-109	Yes	No
RAL-129	Yes	No
RAL-136	Yes	No
RAL-138	No	No
RAL-142	Yes	No
RAL-149	Yes	No
RAL-153	No	No
RAL-158	Yes	No
RAL-161	Yes	No
RAL-176	Yes	Yes
RAL-177	Yes	Yes
RAL-181	Yes	Yes
RAL-195	Yes	No
RAL-208	Yes	Yes
RAL-217	Yes	No
RAL-227	Yes	No
RAL-228	Yes	No
RAL-229	Yes	No
RAL-233	Yes	No
RAL-235	Yes	No
RAL-237	Yes	No
RAL-239	Yes	No
RAL-256	Yes	No
RAL-272	No	No
RAL-280	Yes	No
RAL-287	Yes	No
RAL-309	Yes	No
RAL-310	Yes	No
RAL-313	No	No
RAL-317	Yes	No
RAL-318	No	No
RAL-320	Yes	Yes
RAL-321	Yes	Yes
RAL-325	No	No
RAL-332	Yes	Yes
RAL-338	Yes	No
RAL-350	Yes	No
RAL-352	No	No
RAL-356	Yes	No
RAL-357	Yes	No
RAL-358	No	No
RAL-359	Yes	No
RAL-362	No	No
RAL-365	Yes	No
RAL-367	Yes	No
RAL-370	Yes	No
RAL-371	Yes	No
RAL-373	Yes	No
RAL-374	Yes	No
RAL-375	Yes	Yes
RAL-377	Yes	No
RAL-378	Yes	Yes
RAL-379	No	No
RAL-380	Yes	Yes
RAL-381	Yes	Yes
RAL-383	Yes	No
RAL-386	No	No
RAL-391	Yes	Yes
RAL-392	Yes	No
RAL-398	No	No
RAL-399	Yes	No
RAL-405	Yes	No
RAL-406	Yes	Yes
RAL-409	Yes	No
RAL-426	Yes	No
RAL-427	Yes	No
RAL-437	Yes	No
RAL-439	No	No
RAL-440	Yes	No
RAL-441	Yes	No
RAL-443	Yes	Yes
RAL-461	No	No
RAL-491	Yes	No
RAL-492	Yes	Yes
RAL-502	Yes	Yes
RAL-508	Yes	No
RAL-509	Yes	No
RAL-513	Yes	No
RAL-517	No	No
RAL-531	Yes	No
RAL-535	Yes	No
RAL-554	No	No
RAL-555	Yes	No
RAL-563	Yes	No
RAL-589	No	No
RAL-591	No	No
RAL-595	Yes	No
RAL-639	Yes	No
RAL-642	Yes	No
RAL-646	Yes	No
RAL-703	Yes	No
RAL-705	Yes	No
RAL-707	Yes	No
RAL-712	No	No
RAL-714	Yes	No
RAL-716	Yes	No
RAL-721	Yes	No
RAL-727	Yes	Yes
RAL-730	No	No
RAL-732	Yes	No
RAL-737	Yes	Yes
RAL-738	Yes	Yes
RAL-757	Yes	Yes
RAL-761	Yes	No
RAL-765	Yes	No
RAL-774	Yes	No
RAL-776	Yes	No
RAL-783	No	No
RAL-786	Yes	No
RAL-787	Yes	No
RAL-790	Yes	No
RAL-796	Yes	No
RAL-799	No	No
RAL-801	No	No
RAL-802	No	No
RAL-804	Yes	No
RAL-805	Yes	No
RAL-808	Yes	No
RAL-810	No	No
RAL-812	Yes	No
RAL-818	Yes	No
RAL-820	Yes	No
RAL-822	Yes	No
RAL-832	No	No
RAL-837	Yes	No
RAL-852	Yes	No
RAL-855	Yes	No
RAL-857	Yes	No
RAL-859	Yes	No
RAL-861	No	No
RAL-879	Yes	No
RAL-882	No	No
RAL-884	Yes	No
RAL-887	Yes	No
RAL-890	Yes	No
RAL-892	Yes	No
RAL-894	Yes	No
RAL-897	Yes	Yes
RAL-907	Yes	No
RAL-908	Yes	No
RAL-911	Yes	No

Supplementary Table 2: PCR validation of a subset of TEs present in the *D. melanogaster* reference in nine of the DGRP lines.

Line	Coverage	# of TE Present by PCR	Pipeline Agree	# of TE Absent by PCR	Pipeline Agree	Pipeline No Call	Total
375	38.1	47	0.91	136	0.96	0.02	187
555	21.6	46	1.00	129	0.98	0.06	186
639	19.4	51	0.94	124	1.00	0.07	188
705	18.9	44	0.95	134	0.98	0.06	189
707	20.5	46	1.00	125	0.99	0.08	185
714	20.0	52	0.94	126	0.99	0.06	189
732	18.9	45	1.00	134	0.99	0.04	187
765	19.7	46	1.00	130	0.99	0.07	189
820	17.0	46	0.98	126	0.99	0.08	187

Supplementary Table 3: TE presence and absence calls made by local reconstruction of TEs.

	Total	Presence	Absence	No Call	% Calls
DGRP	2485070	52805	2361918	70216	97.17%
DGRP25	143106	10753	132320	2724	99.98%
DSRP	111300	10630	99270	1385	98.74%

Supplementary Table 6: Genes with different TE frequencies.

FBgn	Gene	DSPR present	DSPR absent	DGRP present	DGRP absent	Bonferroni corrected p-value
FBgn0087007	*bbg*	9	6	13	118	2.24E-005
FBgn0041092	*tai*	9	6	14	117	3.52E-005
FBgn0024753	*Flo-2*	10	5	19	112	3.42E-005
FBgn0037363	*CG1347*	15	0	58	73	2.74E-005
FBgn0037295	*dpr16*	15	0	55	76	6.80E-006
FBgn0259242	*CG42340*	12	3	21	110	8.18E-007
FBgn0051536	*Cdep*	15	0	56	75	8.62E-006
FBgn0259152	*Clbn*	11	4	13	118	2.14E-007
FBgn0086689	*Hph*	15	0	54	77	5.34E-006
FBgn0000166	*bcd*	10	5	8	123	1.12E-007
FBgn0031116	*CG1695*	6	9	1	130	2.73E-006
FBgn0052817	*CG32817*	12	3	3	128	1.58E-012
FBgn0053202	*dpr11*	7	8	5	126	1.62E-005
FBgn0030600	*hiw*	6	9	1	130	2.73E-006
FBgn0003715	*Tkr*	15	0	38	93	5.89E-008
FBgn0004370	*Ptp10D*	12	3	20	111	5.34E-007
FBgn0083950	*CG34114*	9	6	7	124	6.65E-007
FBgn0011224	*heph*	15	0	39	92	8.16E-008
FBgn0053507	*dpr2*	10	5	9	122	2.29E-007
FBgn0037292	*CG2022*	8	7	6	125	3.46E-006
FBgn0083949	*CG34113*	10	5	14	117	4.05E-006