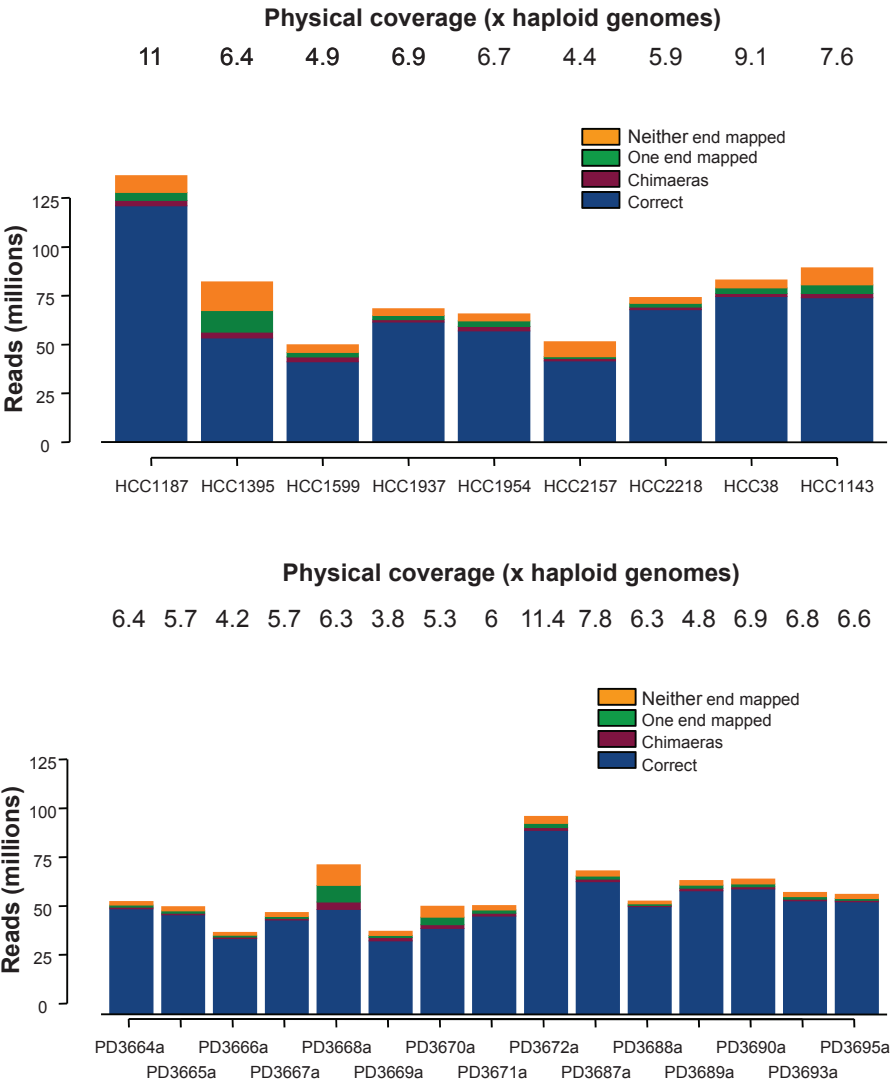
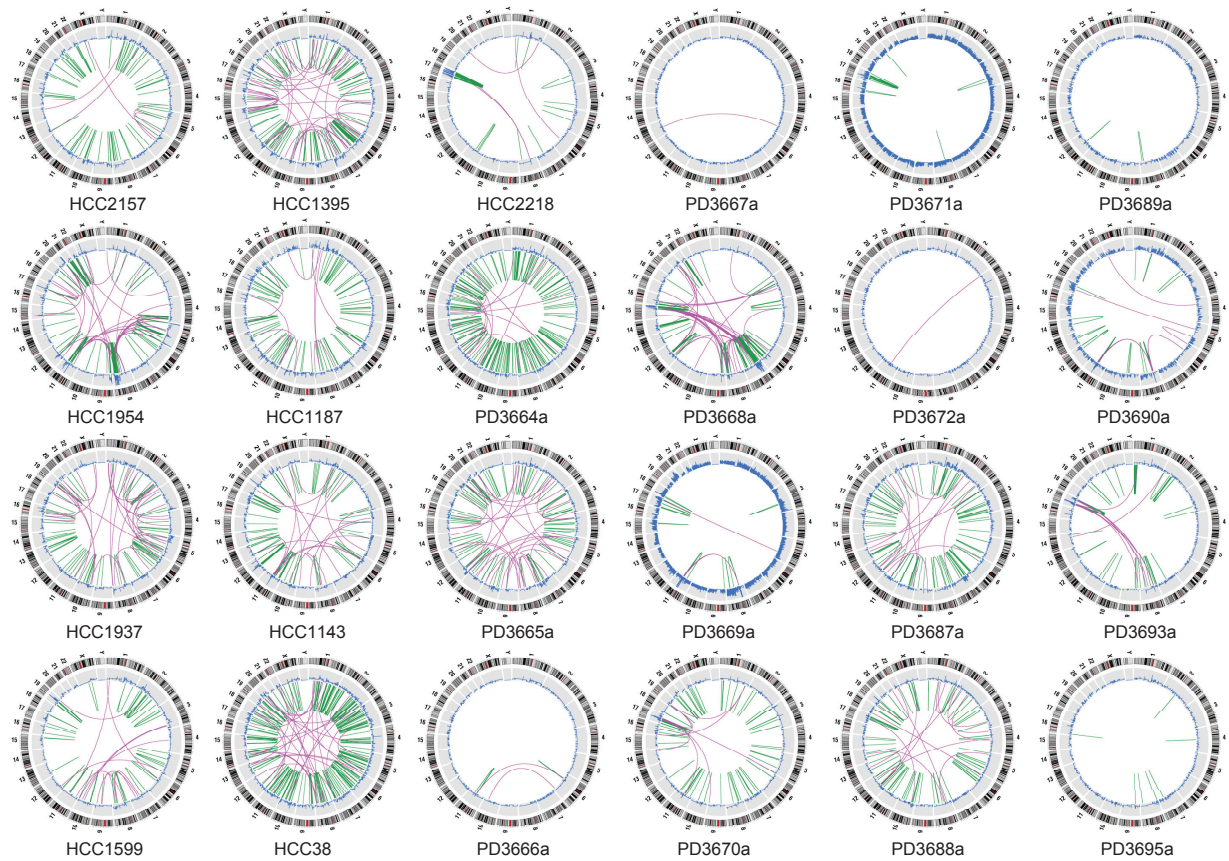


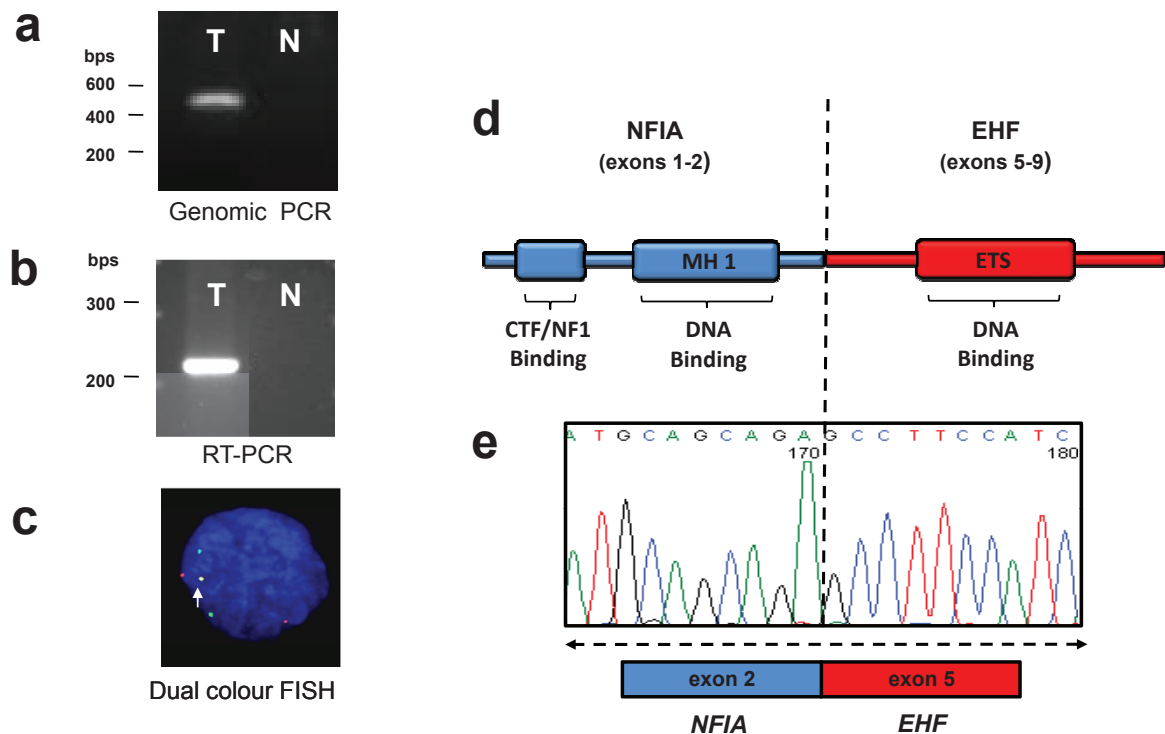
SUPPLEMENTARY INFORMATION



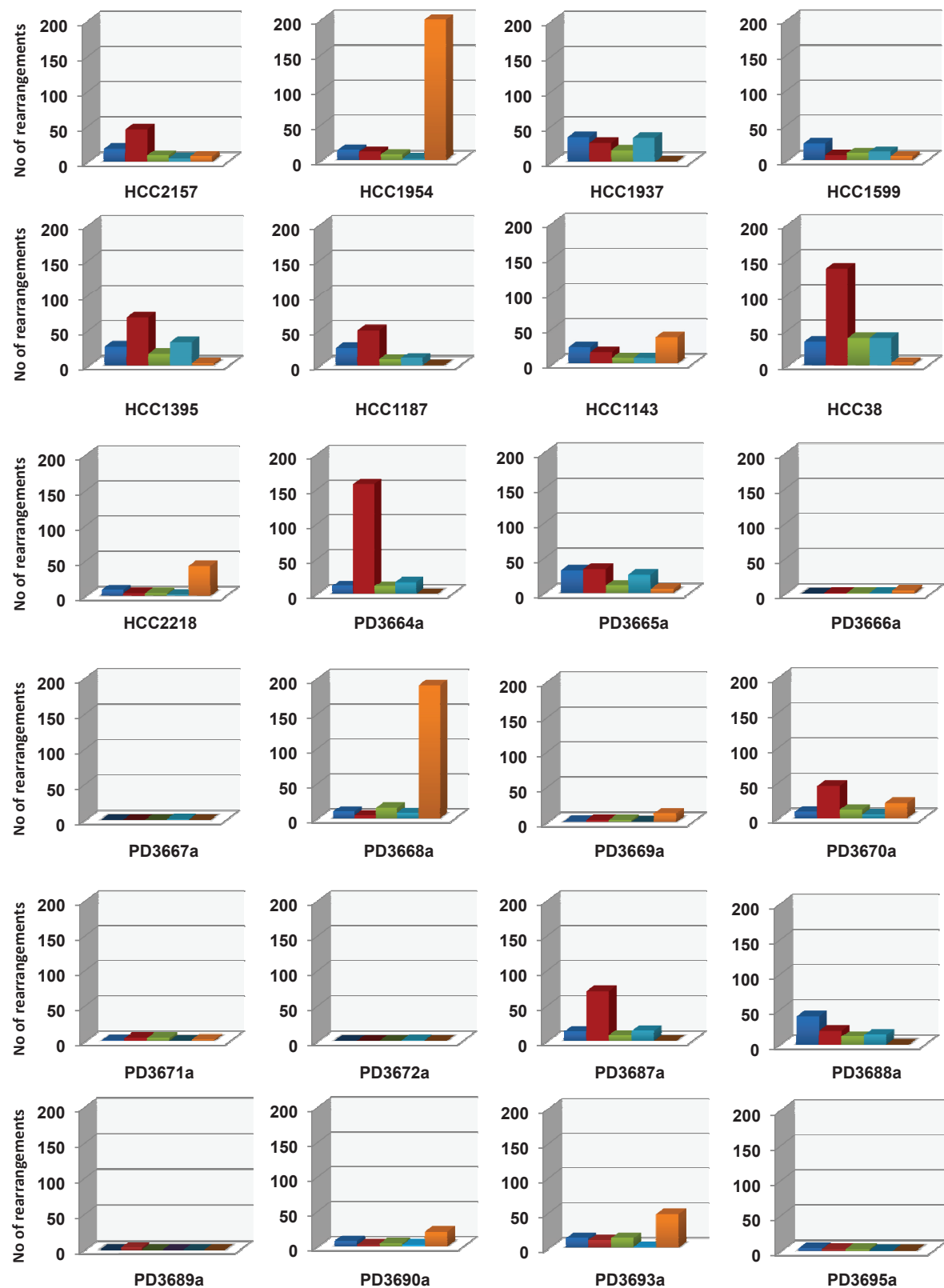
Supplementary Figure 1 Haploid physical coverage of breast cancer samples. Physical coverage indicates the number of DNA fragments of which both ends have been sequenced that on average overlie any position in the genome.



Supplementary Figure 2 Genome wide circos plots of somatic rearrangements in all 24 breast cancers in the study.

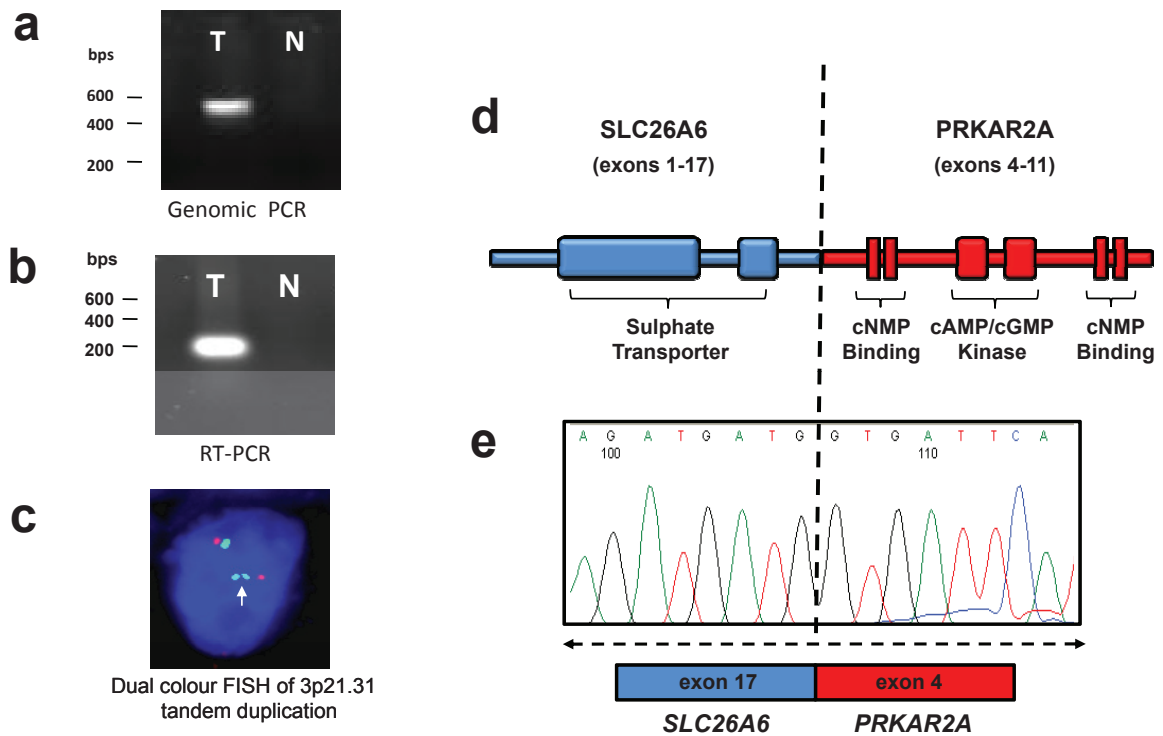


Supplementary Figure 3 *NFIA-EHF*, an expressed, in frame fusion gene caused by an interchromosomal rearrangement in breast cancer cell line HCC1937. (a) Across rearrangement PCR to confirm the presence of the somatic rearrangement in the cancer but not in normal DNA; (b) RT-PCR of RNA between *NFIA* exon 2 and *EHF* exon 5 to confirm the presence of a chimeric expressed transcript; (c) Representative picture of dual colour FISH confirming a translocation in HCC1937. Red probe corresponds to BAC RP11-364M11, chromosome 1: 61,064,196-61,228,554. Green probe corresponds to BAC RP11-277N08, chromosome 11: 34,772,104-34,965,946. (d) Schematic diagram of the protein domains fused in the predicted *NFIA/EHF* fusion protein. Domains from *NFIA* are blue, domains from *EHF* are red (e) Sequence from RT-PCR product shown in (b) confirming *NFIA* exon 2 fused to *EHF* exon 5.

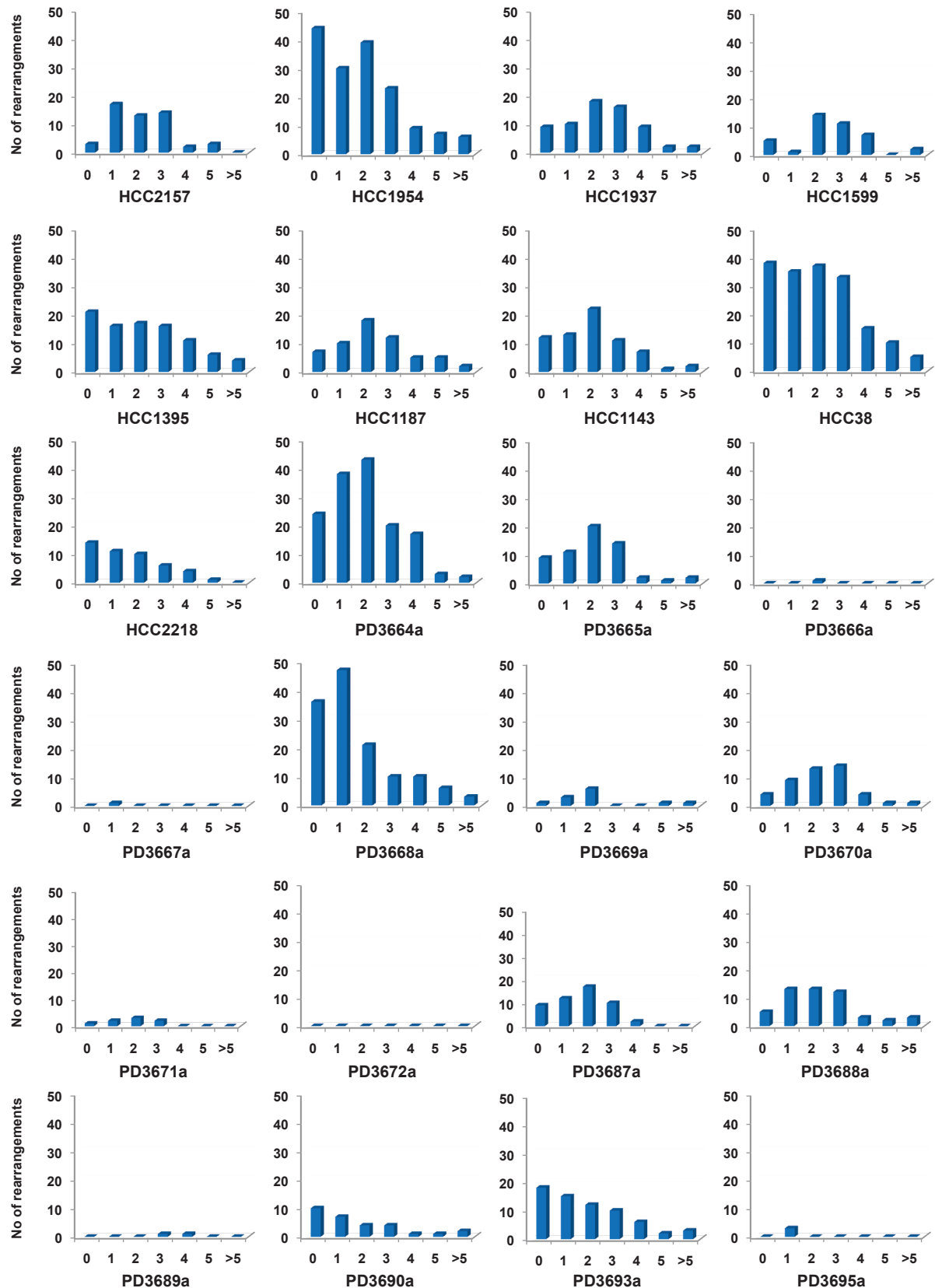


Supplementary Figure 4 *SLC26A6-PRKAR2A*, an expressed, in-frame fusion gene generated by a tandem duplication in the breast cancer cell line HCC38. (a) Across

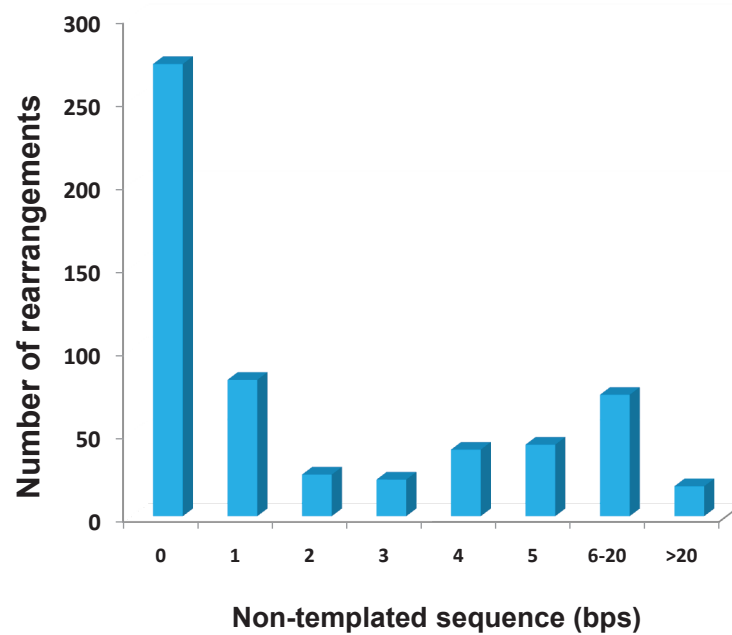
rearrangement PCR to confirm the presence of the somatic rearrangement in the cancer but not in normal DNA; (b) RT-PCR of RNA between *SLC26A6* exon 17 and *PRKAR2A* exon 4 to confirm the presence of a chimeric expressed transcript; (c) Dual colour FISH confirming the 3p21.31 tandem duplication in HCC38. Green-labelled BAC RP11-148G20 is within the tandem duplication. Red-labelled BAC RP11-527M10 is located ~3 Mb telomeric of the tandem duplication. (d) Schematic diagram of the protein domains in the predicted *SLC26A6-PRKAR2A* fusion protein. Domains from *SLC26A6* are blue, domains from *PRKAR2A* are red. (e) Sequence from RT-PCR product shown in (b) confirming *SLC26A6* exon 17 fused to *PRKAR2A* exon 4.



Supplementary Figure 5 Prevalence of architectures of rearrangements in all 24 breast cancers in the study: Deletion (dark blue), tandem duplication (red), inverted orientation (green), interchromosomal (light blue), breakpoint(s) within an amplicon (orange).



Supplementary Figure 6 Number of nucleotides of overlapping microhomology at rearrangement junctions in the 24 breast cancers.



Supplementary Figure 7 Nucleotides of non-templated DNA sequence at rearrangement breakpoints in the 24 breast cancers.