Supplementary Material

Supplementary Figure S1

Because acoustical shearing of DNA generates random ends that will have variable degrees of complementarity to the primers on the flow cell surface, a series of oligonucleotides was synthesized to test hybridization properties. A 29-mer was attached to an epoxide-containing flow cell surface via a 5' C12 amino link. The 3' end of this molecule can serve as a primer to read the sequence of any molecule that stably hybridizes to it. Thirty oligonucleotides were synthesized with unique 5' ends that could be distinguished by sequencing and connected to variable 3' ends that were complementary to 9-29nt of the surface-bound primer. Oligonucleotides were made complementary to 9, 14, and 19-29nt (corresponding to gaps of 0 to 20nt) of the primer. Additionally, oligonucleotides were made with perfect 29nt matches to the primer and 3'-overhangs of 1-10, 15 and 20nt. Duplicate oligonucleotides were made for -20, -15, -10, -5, 0, +5, +10, +15, and +20 relative to the primer (sequences provided in Table S2). All oligonucleotides except those containing only 9 and 14nt of complementarity provided significant amounts of sequence data with standard hybridization conditions and 1% formamide (dashed lines, open circles). At 25% formamide (solid lines, filled circles), the yield of oligos with only 19-22nt complementarity drops relative to lower formamide. In the presence of 1 μg of sheared human genomic DNA, this effect is magnified as the oligos show more specificity in the background of non-specific DNA (data not shown). Though there is an obvious length dependence of the hybridization, this effect is not monotonic as longer stretches of complementarity sometimes generate lower yield. In some cases, this appears to be due to internal secondary structure in the incoming

oligonucleotides as some have structures that might be stable in these conditions and inhibit binding to the surface. For example, the oligonucleotide with 23nt complementarity has a stable intramolecular structure that includes the 3' end which likely reduces its hybridization efficiency. Interestingly, these differences tend to disappear at 25% formamide, suggesting that the intramolecular structures melt and the entire sequence is available for intermolecular interactions. In addition, oligonucleotides with 3' overhangs of up to 20nt (the longest that was tested) that go beyond the 5' end of the primer hybridize as well or better than oligonucleotides that have no overhang. Interestingly, the overhang sequence affects the yield of reads as all the oligonucleotides whose overhang starts with the sequence AACCT yield more reads than the two that start with TTAGG. This may be due to interactions with the surface or the hybrid stability could be affected by the ease with which the single-stranded DNA can turn to avoid the surface or stack with the double-stranded DNA. This behavior is mirrored in the *BRCA1* sequences where it is found that the sequence just beyond the capture primer has an effect on hybridization efficiency. On average, the highest yielding capture primers have the sequence TT just beyond the 5' end or, AA on the incoming DNA, mirroring the effect of the synthetic oligonucleotides.

Supplementary Figure S2

When standard sequencing using tailed DNA on an oligo dT50 flow cell is carried out on the HeliScope Sequencing system, the DNA to be sequenced is blocked at the 3' end so that the only end available for sequence extension is the 3' end of surface-attached primer. Oligonucleotides were used to study this in more detail. One set of

oligonucleotides was blocked at the 3' end with a phosphate so they could not extend. A fraction of this mix was phosphatased to provide free 3' ends that could theoretically be extended during the sequencing reaction and yield sequence signal from both DNA strands, causing apparent insertions. Increased errors are observed with underhang oligonucleotides that are not fully complementary to the capture primer. The insertion (solid black squares) and substitution (gray triangles) rate is slightly higher for the molecules with a 3' OH relative to a 3' PO_4. The deletion rate (open diamonds) is unaffected. In contrast, oligonucleotides with a 3' overhang have exactly the same error rate whether they contain a 3' OH or PO_4. Thus, when no template is present, there is no need to block the 3' end. Since extended molecules hybridize efficiently and are likely to make up the majority of hybridizing molecules when using randomly sheared DNA, blocking the 3' end of samples is not necessary.

Supplementary Figure S3

Average coverage for 19 samples across the target region, demonstrating high reproducibility of probe capture efficiency among different samples. Coverage was calculated in bin sizes of 50nt. Exon numbering is shown. Intronic size is not to scale.

Supplementary Figure S4

The start site for observed sequence reads is shown for the standard direct capture (solid line) and with primer extension (JumpStart Sequencing, dashed line) prior to sequencing. Most reads start within one base of the expected start site for standard sequencing but

vary by more than 20nt when extended prior to sequencing. The length of this variable start can be changed by altering polymerization time or nucleotide concentration.

Supplementary Table 1

Sequences for the oligonucleotides used in the construction of the flow cell are listed as well as the exons for which they were designed.

Supplementary Table 2

Sequences for the oligonucleotides used for testing hybridization efficiency are shown with the length of under/overhang in each.