Supplementary Information for:

A streamlined and high-throughput targeting approach for human germline and cancer genomes using Oligonucleotide-Selective Sequencing

Samuel Myllykangas[1†], Jason D. Buenrostro[2†], Georges Natsoulis[1], John M. Bell[2], Hanlee P. Ji[1,2]

† These authors contributed equally to this work.

[1]Division of Oncology, Department of Medicine, Stanford University School of Medicine, Stanford, CA, United States, 94305

[2]Stanford Genome Technology Center, Stanford University, Palo Alto, CA, United States, 94304

Supplementary Methods. Related programs are available in Supplementary Data Files.

Supplementary Tables 1 – 4. Supplementary Tables 5 and 6 are available as separate files.

Supplementary Figures 1 – 8.

Supplementary Methods

Overview of the molecular biology of OS-Seq. We have provided a detailed description of the underlying molecular biology of OS-Seq including hybridization, ligation and extension among other molecular parameters. The sequences of the specific oligonucleotides are also provided in the following pages. The basic reagents include target-specific oligonucleotides, a single adapter for creating sequencing libraries and standard cluster amplification reagents. There are three general steps for OS-Seq. For **Step 1**, target-specific oligonucleotides are used to modify flow cell primers to primer-probes. In the Illumina flow cell two types of primers (named C and D) are immobilized on a paired-end flow cell. In OS-Seq a subset of D primers are modified to primer-probes using complex library of oligonucleotides. Oligonucleotides have sequences that hybridize to type D flow cell primers. Hybridized oligonucleotides are then used as a template for DNA polymerase and D primers are extended. After denaturation, target-specific primer-probes are randomly immobilized on the flow cell.

For **Step 2**, genomic targets in a single-adaptor library are captured using primer-probes. These adapters incorporate sites for sequencing primers and immobilized flow cell primers. In OS-Seq, we use a modified adapter to prepare single-adapter libraries from genomic DNA. Targets in single-adaptor library are captured during high heat hybridization to their complementary primer-probes. Captured single-adapter library fragments are used as a template for DNA polymerase and primer-probes are extended. Denaturation releases template DNA from immobilized targets.

For **Step 3**, immobilized targets are rendered to be compatible with Illumina sequencing. In Illumina sequencing, solid-phase amplification of the immobilized sequencing library fragments using C and D primers is required. In OS-Seq, during low heat hybridization the single-adapter tails of the immobilized targets hybridize to type C primers on the flow cell surface, which stabilizes a bridge structure. The 3' ends of immobilized targets and C primers are extended using DNA polymerase. After denaturation, two complementary, immobilized sequencing library fragments are formed that contain complete C and D priming sites and are compatible with solid-phase amplification. After all of the steps of OS-Seq, immobilized targets are structurally identical to a standard paired-end Illumina library and are amplified and processed using Illumina's standard kits and protocols.

.

Primer and adapter sequences. Listed are the primers used in the OS-Seq approach. Modified nucleotides are noted.

X = phosphorothioate bond

P = 5'-phosphate

Oligonucleotide Name	Sequence	Note
Ad_top_FC_capture_A_tail	CGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGATCxT	This oligo is used for PCR amplification of the library.
Ad_bot_FC_capture_A_tail	p-GATCGGAAGAGCGTCGTGTAGGGAAAGAGTGTAGATCTCG	
Ad_top_FC_capture_TGCTAA_1	CGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGATCTTGCTAAxT	
Ad_top_FC_capture_AGGTCA_2	CGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGATCTAGGTCAxT	
Ad_top_FC_capture_AACCTG_3	CGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGATCTAACCTGxT	
Ad_bot_FC_capture_TGCTAA_1	p-TTAGCAAGATCGGAAGAGCGTCGTGTAGGGAAAGAGTGTAGATCTCG	
Ad_bot_FC_capture_AGGTCA_2	p-TGACCTAGATCGGAAGAGCGTCGTGTAGGGAAAGAGTGTAGATCTCG	
Ad_bot_FC_capture_AACCTG_3	p-CAGGTTAGATCGGAAGAGCGTCGTGTAGGGAAAGAGTGTAGATCTCG	
Microarray_oligo_amp_primer_1_U	GCTGACCTTAAACCTAACGCGAGGGCGGCAGTTGGGATTTCGTGACCTATGCACCAGACGU	
Microarray_oligo_amp_primer_2	CAAGCAGAAGACGGCATACGAGATCGGTCTCGGCATTCCTGCTGAACCGCTCTTCCGATCT	

Overview of the molecular biology and related DNA sequence events of OS-Seq. We list the oligonucleotide sequences and underlying molecular events that are part of OS-Seq and subsequent sequencing of targets.

0) Oligonucleotides

```
OS-Seq oligonucleotide:
5' - NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNAGATCGGAAGAGCGGTTCAGCAGGAATGCCGAGACCGATCTCGTATGCCGTCTTCTGCTTG - 3' (Generic capture oligonucleotide, N = unique 40-mer sequence)

Ad_top_FC_capture_A_tail:
5' - CGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGATCT - 3'

Ad_bot_FC_capture_A_tail:
5' - GATCGGAAGAGCGTCGTGTAGGGAAAGAGTGTAGATCTCG - 3'

Flow cell primer ‘C’:
5’ - PS-TTTTTTTTTTAATGATACGGCGACCACCGAGAUCTACAC - 3’ (U = 2-deoxyuridine)

Flow cell primer ‘D’:
5’ - PS-TTTTTTTTTTCAAGCAGAAGACGGCATACGAGoxoAT - 3’, (Goxo = 8-oxoguanine)

Sequencing primer 1:
5' - ACACTCTTTCCCTACACGACGCTCTTCCGATCT - 3'

Sequencing primer 2:
5’ - CGGTCTCGGCATTCCTGCTGAACCGCTCTTCCGATCT - 3’
```

---

1) Flow cell modification

Anneal

```
3' - GTTCGTCTTCTGCCGTATGCTCTAGCCAGAGCCGTAAGGACGACTTGGCGAGAAGGCTAGANNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN - 5' (OS-Seq oligonucleotide)
FC - CAAGCAGAAGACGGCATACGAGAT - 3’                                                                                  (Flow cell primer ‘D’)
```

Extension

```
3' - GTTCGTCTTCTGCCGTATGCTCTAGCCAGAGCCGTAAGGACGACTTGGCGAGAAGGCTAGANNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN - 5' (OS-Seq oligonucleotide)
FC - CAAGCAGAAGACGGCATACGAGATCGGTCTCGGCATTCCTGCTGAACCGCTCTTCCGATCTNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN - 3' (primer-probe)
```

Denature

```
FC - CAAGCAGAAGACGGCATACGAGATCGGTCTCGGCATTCCTGCTGAACCGCTCTTCCGATCTNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN - 3' (primer-probe)
```

---

2) Library prep

Fragmentation, end repair

```
5' - NNNNNNNNNNNNNNNNNNNN - 3' (genomic DNA)
3' - NNNNNNNNNNNNNNNNNNNN - 5' (genomic DNA)
```

A-tailing

```
5' -  NNNNNNNNNNNNNNNNNNNNA - 3' (genomic DNA after A-tailing)
3' - ANNNNNNNNNNNNNNNNNNNN  - 5' (genomic DNA after A-tailing)
```

Adaptor ligation

OS-Seq dsAdapter

```
5' -  GATCGGAAGAGCGTCGTGTAGGGAAAGAGTGTAGATCTCG - 3' (Ad_bot_FC_capture_A_tail)
3' - TCTAGCCTTCTCGCAGCACATCCCTTTCTCACATCTAGAGC - 5' (Ad_top_FC_capture_A_tail)
```

OS-Seq dsAd library (This is the structure of the OS-Seq-adaptor library, N = random genomic DNA sequence defined by fragmentation)

```
5' - CGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNNNNNNNNNNNNNNNNNNNNNAGATCGGAAGAGCGTCGTGTAGGGAAAGAGTGTAGATCTCG - 3'
3' - GCTCTAGATGTGAGAAAGGGATGTGCTGCGAGAAGGCTAGANNNNNNNNNNNNNNNNNNNNNNNTCTAGCCTTCTCGCAGCACATCCCTTTCTCACATCTAGAGC - 5'
```

Library PCR

OS-Seq adaptor library amplification (Ad_top_FC_capture_A_tail, single primer PCR is used to amplify the adaptor library)

```
5' - CGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGATCT - 3'                                                                  (Ad_top_FC_capture_A_tail)
3' - GCTCTAGATGTGAGAAAGGGATGTGCTGCGAGAAGGCTAGANNNNNNNNNNNNNNNNNNNNNNNTCTAGCCTTCTCGCAGCACATCCCTTTCTCACATCTAGAGC - 5' (OS-Seq library fragment)

5' - CGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNNNNNNNNNNNNNNNNNNNNNAGATCGGAAGAGCGTCGTGTAGGGAAAGAGTGTAGATCTCG - 3' (OS-Seq library fragment)
3' -                                                    TCTAGCCTTCTCGCAGCACATCCCTTTCTCACATCTAGAGC - 5' (Ad_top_FC_capture_A_tail)

5' - CGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNNNNNNNNNNNNNNNNNNNNNAGATCGGAAGAGCGTCGTGTAGGGAAAGAGTGTAGATCTCG - 3' (OS-Seq library fragment, amplified)
3' - GCTCTAGATGTGAGAAAGGGATGTGCTGCGAGAAGGCTAGANNNNNNNNNNNNNNNNNNNNNNNTCTAGCCTTCTCGCAGCACATCCCTTTCTCACATCTAGAGC - 5' (OS-Seq library fragment, amplified)
```

---

3) Capture

Anneal

OS-Seq adaptor library annealing (N = 40-mer specific capture site)

```
      3' -    GCTCTAGATGTGAGAAAGGGATGTGCTGCGAGAAGGCTAGAgenomicdna
                                                                NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNgenomicdnaTCTAGCCTTCTCGCAGCACATCCCTTTCTCACATCTAGAGC - 5' (OS-Seq library)
FC - CAAGCAGAAGACGGCATACGAGATCGGTCTCGGCATTCCTGCTGAACCGCTCTTCCGATCTNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN - 3'                                                                                  (primer-probe)
```

Extension

OS-Seq capture

```
3' -           GCTCTAGATGTGAGAAAGGGATGTGCTGCGAGAAGGCTAGAgenomicdna
                                                                NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNgenomicdnaTCTAGCCTTCTCGCAGCACATCCCTTTCTCACATCTAGAGC - 5' (OS-Seq library fragment, amplified)
FC - CAAGCAGAAGACGGCATACGAGATCGGTCTCGGCATTCCTGCTGAACCGCTCTTCCGATCTNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNgenomicdnaAGATCGGAAGAGCGTCGTGTAGGGAAAGAGTGTAGATCTCG - 3' (captured DNA)
```

Denature

OS-Seq library

```
FC - CAAGCAGAAGACGGCATACGAGATCGGTCTCGGCATTCCTGCTGAACCGCTCTTCCGATCTNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNgenomicdnaAGATCGGAAGAGCGTCGTGTAGGGAAAGAGTGTAGATCTCG - 3' (captured DNA)
```

4) Adapter finalizing

Hybridization in 40C

OS-Seq_Library (there is 12-mer homology between the OS-Seq adaptor and Oligo-C)

```
FC - CAAGCAGAAGACGGCATACGAGATCGGTCTCGGCATTCCTGCTGAACCGCTCTTCCGATCTNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNgenomicdnaAGATCGGAAGAGCGTCGTGTAGGGAAAGAGTGTAGATCTCG - 3'          (captured DNA)
                                                                                                                                                3' - CACATCTAGAGCCACCAGCGGCATAGTAA - FC (Oligo'C')
```

Extend

```
FC - CAAGCAGAAGACGGCATACGAGATCGGTCTCGGCATTCCTGCTGAACCGCTCTTCCGATCTNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNgenomicdnaAGATCGGAAGAGCGTCGTGTAGGGAAAGAGTGTAGATCTCGGTGGTCGCCGTATCATT - 3' (finalized DNA)
3' - GTTCGTCTTCTGCCGTATGCTCTAGCCAGAGCCGTAAGGACGACTTGGCGAGAAGGCTAGANNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNgenomicdnaTCTAGCCTTCTCGCAGCACATCCCTTTCTCACATCTAGAGCCACCAGCGGCATAGTAA - FC (finalized DNA)
```

Denature

```
FC - CAAGCAGAAGACGGCATACGAGATCGGTCTCGGCATTCCTGCTGAACCGCTCTTCCGATCTNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNgenomicdnaAGATCGGAAGAGCGTCGTGTAGGGAAAGAGTGTAGATCTCGGTGGTCGCCGTATCATT - 3' (finalized DNA)

3' - GTTCGTCTTCTGCCGTATGCTCTAGCCAGAGCCGTAAGGACGACTTGGCGAGAAGGCTAGANNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNgenomicdnaTCTAGCCTTCTCGCAGCACATCCCTTTCTCACATCTAGAGCCACCAGCGGCATAGTAA - FC (finalized DNA)
```

---

5) Cluster generation

Anneal

```
FC - CAAGCAGAAGACGGCATACGAGATCGGTCTCGGCATTCCTGCTGAACCGCTCTTCCGATCTNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNgenomicdnaAGATCGGAAGAGCGTCGTGTAGGGAAAGAGTGTAGATCTCGGTGGTCGCCGTATCATT - 3' (finalized DNA)
                                                                                                                                                3' - CACATCTAGAGCCACCAGCGGCATAGTAA - FC (Oligo'C')

FC - CAAGCAGAAGACGGCATACGAGAT - 3’                                                                                                                                                         (Oligo'D')
3' - GTTCGTCTTCTGCCGTATGCTCTAGCCAGAGCCGTAAGGACGACTTGGCGAGAAGGCTAGANNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNgenomicdnaTCTAGCCTTCTCGCAGCACATCCCTTTCTCACATCTAGAGCCACCAGCGGCATAGTAA - FC (finalized DNA)
```

Extend

```
FC - CAAGCAGAAGACGGCATACGAGATCGGTCTCGGCATTCCTGCTGAACCGCTCTTCCGATCTNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNgenomicdnaAGATCGGAAGAGCGTCGTGTAGGGAAAGAGTGTAGATCTCGGTGGTCGCCGTATCATT - 3' (finalized DNA)
3' - GTTCGTCTTCTGCCGTATGCTCTAGCCAGAGCCGTAAGGACGACTTGGCGAGAAGGCTAGANNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNgenomicdnaTCTAGCCTTCTCGCAGCACATCCCTTTCTCACATCTAGAGCCACCAGCGGCATAGTAA - FC (finalized DNA)
```

Denature

```
FC - CAAGCAGAAGACGGCATACGAGATCGGTCTCGGCATTCCTGCTGAACCGCTCTTCCGATCTNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNgenomicdnaAGATCGGAAGAGCGTCGTGTAGGGAAAGAGTGTAGATCTCGGTGGTCGCCGTATCATT - 3' (Clustered DNA)

3' - GTTCGTCTTCTGCCGTATGCTCTAGCCAGAGCCGTAAGGACGACTTGGCGAGAAGGCTAGANNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNgenomicdnaTCTAGCCTTCTCGCAGCACATCCCTTTCTCACATCTAGAGCCACCAGCGGCATAGTAA - FC (Clustered DNA)
```

---

6) Sequencing

```
FC - CAAGCAGAAGACGGCATACGAGATCGGTCTCGGCATTCCTGCTGAACCGCTCTTCCGATCTNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNgenomicdnaAGATCGGAAGAGCGTCGTGTAGGGAAAGAGTGTAGATCTCGGTGGTCGCCGTATCATT - 3' (Clustered DNA)
3' -                                                                                                         < - - - -TCTAGCCTTCTCGCAGCACATCCCTTTCTCACA                          - 5' (Sequencing Primer
1)

5' -                         CGGTCTCGGCATTCCTGCTGAACCGCTCTTCCGATCT - - - - >                                                                                                    (Sequencing Primer
2)
3' - GTTCGTCTTCTGCCGTATGCTCTAGCCAGAGCCGTAAGGACGACTTGGCGAGAAGGCTAGANNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNgenomicdnaTCTAGCCTTCTCGCAGCACATCCCTTTCTCACATCTAGAGCCACCAGCGGCATAGTAA - FC (Clustered DNA)
```

Cluster station programs for OS-Seq. OS-Seq is run in two parts with two scripts: OSSEQ_program_p1.xml and OSSEQ_program_p2.xml. The programs are available in **Supplementary Data Files** and at our website (http://dna-discovery.stanford.edu).

The prompts in these scripts inform the user which reagents to load in each step. These prompts were written for use of PhiX in lane 4, if not using PhiX then we recommend one disregards the prompts specific to lane 4. To begin the OS-Seq process, we load the program ending with “p1” using “Illumina ClusterStation Software”. To proceed, we follow the prompts as indicated in the scripts. The second hybridization step is for 20 hours. We generally that the “p1” script be run between 12-5 pm, otherwise the hybridization will be completed early morning the next day. If using PhiX, we denature the template with the standard Illumina method and concentrations during the 2nd “Extension-Amplification” of “p1”. Once “p1” is complete, we load script beginning in “p2”, run, and follow the prompts. Once “p2” is complete, we proceed with the standard Illumina protocol for sequencing.

Indexing assignment. The following perl scripts were used for indexing assignment of sequence data produced from OS-Seq: 6b_tag_dist.pl and 6b_binner.pl. The programs are available in **Supplementary Data Files** and at the following website (http://dna-discovery.stanford.edu). The qseq.txt scripts produced by the Illumina off-line base caller (OLB 1.9) are concatenated into one file for each lane. The script 6b_tag_dist.pl is run on this concatenated file. Its two outputs contain the names "6b_index" and "6b_tag_dist". The output containing "6b_tag_dist" shows the tag counts, which serves as a useful quality control step. The file "6b_index.txt" is one of two input files for the script "6b_binner.pl". The script "6b_binner.pl" is run with the "6b_index" file and the Illumina export file as inputs. This produces a separate export file for each of the 16 tags.

An example is listed below.

Assume the qseq files have been created and GERALD.pl run to create the aligned files, including export files:
cat > 20120101_l1.all_seq.txt s_1_1_????_qseq.txt
6b_tag_dist.pl 201201_l1.all_seq.txt

The above produces: 20120101_l1.6b_tag_dist.txt and 20120101_l1.6b_index.txt.

6b_binner.pl 20120101_l1.6b_index.txt s_1_1_export.txt

This produces seventeen export files (including a garbage file with _OTH in the name) with identifiers such as s_1_1_export_AGGTCA.txt.

Estimating the primer-probe to target ratio. We used information about hybridization efficiency of PhiX, a standard sequencing control for Illumina systems, as a basis for our estimation. In Illumina GAIIx sequencing, PhiX was used in 10 pM concentration. Volume of the Illumina GAIIx flow cell is 10 µl. Thus, there were 60,221,400 Phix sequencing library fragments in the hybridization mix inside a single flow cell lane. Immobilization of the PhiX library resulted in 7,124,567 reads (clusters derived from single library molecules) that were unique matches. Therefore, the efficiency of immobilization was 11.8%.

We then used PhiX library immobilization efficiency to estimate the number of immobilized primer-probes. We used 100 nM targeting oligonucleotides to prepare OS-Seq-366. Overall, there were $6x10^{11}$ oligonucleotide molecules in 10 µl flow cell lane during hybridization. We used similar conditions as in PhiX lane to prepare the primer-probes. If similar efficiency is achieved using synthetic oligonucletides as in PhiX experiments, we estimated that $7.1x10^{10}$ primer-probes were immobilized on the flow cell.

We prepared a single-adapter sequencing library (30 ng/µl) using human genomic DNA. 10 µl of that library was used in hybridization to flow cell. In all, there were DNA equivalent of 90,000 human diploid genomes in the flow cell lane during capture. For OS-Seq-366 there are 366 targets in a human diploid genome, suggesting that there were 32,940,000 potential target molecules in the hybridization mix. For every target molecule in the hybridization there were 2,163 primer-probe molecules. We captured 1,629,751 molecules that resulted in uniquely matched reads (**Table 1**). Therefore, the efficiency of capture using OS-Seq-366 and GAIIx was 4.9%.

Supplementary Table 1. OS-Seq coverage metrics. We report the coverage classes for OS-Seq-366 which involves ten genes and OS-Seq-11K which covers 344 genes. Two samples were sequenced including a normal germline genome (NA18507) and a matched normal-tumor colorectal cancer sample.

Assay	**OS-Seq-366**	**OS-Seq-11K**	**OS-Seq-11K**	**OS-Seq-11K**
Sample	NA18507	NA18507	Normal (matched)	Tumor (colorectal)
Number of primer-probes	366	11,742	11,742	11,742
Exon bases	30,246	947,689	947,689	947,689
Exon bases covered at ≥ 1	26,300	917,360	900,574	908,609
(Proportion of exon bases covered at ≥ 1)	(87.0%)	(96.8%)	(95.0%)	(95.9%)
Exon bases covered at ≥ 10	25,574	641,866	625,815	619,159
(Proportion of exon bases covered at ≥ 10)	(84.6%)	(67.7%)	(66.0%)	(65.3%)
Exon bases covered at ≥ 20	25,349	433,297	436,686	413,500
(Proportion of exon bases covered at ≥ 20)	(83.8%)	(45.7%)	(46.1%)	(43.6%)

Supplementary Table 2. SNV concordance from OS-Seq and NA18507. To assess the variant calling performance of OS-Seq-366 and OS-Seq-11k assays, we conducted a targeted sequencing analysis on NA18507, a Yoruban individual who has undergone whole genome sequencing analysis. For SNV calling with either OS-Seq assay, we analyzed on-target positions with genotype quality scores greater than 50 and a minimum of 10X coverage. We extracted the published NA18507 SNVs and other reported SNPs that occurred in these same high quality regions. In comparison, 97% of the OS-Seq-366 and 95.7% of the OS-Seq-11k had previously been reported (**Table 1**). For OS-Seq-366 and OS-Seq-11k the sensitivity of variant detection was 0.97 and 0.95 respectively based on the reported SNPs.

Sample	**18507**	**18507**
Reported SNP data	Bentley et al. (2008) and dnSNP131	Bentley et al. (2008) and dnSNP131
OS-Seq assay	OS-Seq-366	OS-Seq-11K
Total OS-Seq SNVs	105	985
OS-Seq SNVs concordant with reported SNP position	105	943
OS-Seq SNVs not reported elsewhere	-	42
Reported NA18507 SNPs not called by OS-Seq	3	54
OS-Seq SNV Sensitivity	0.968	0.947

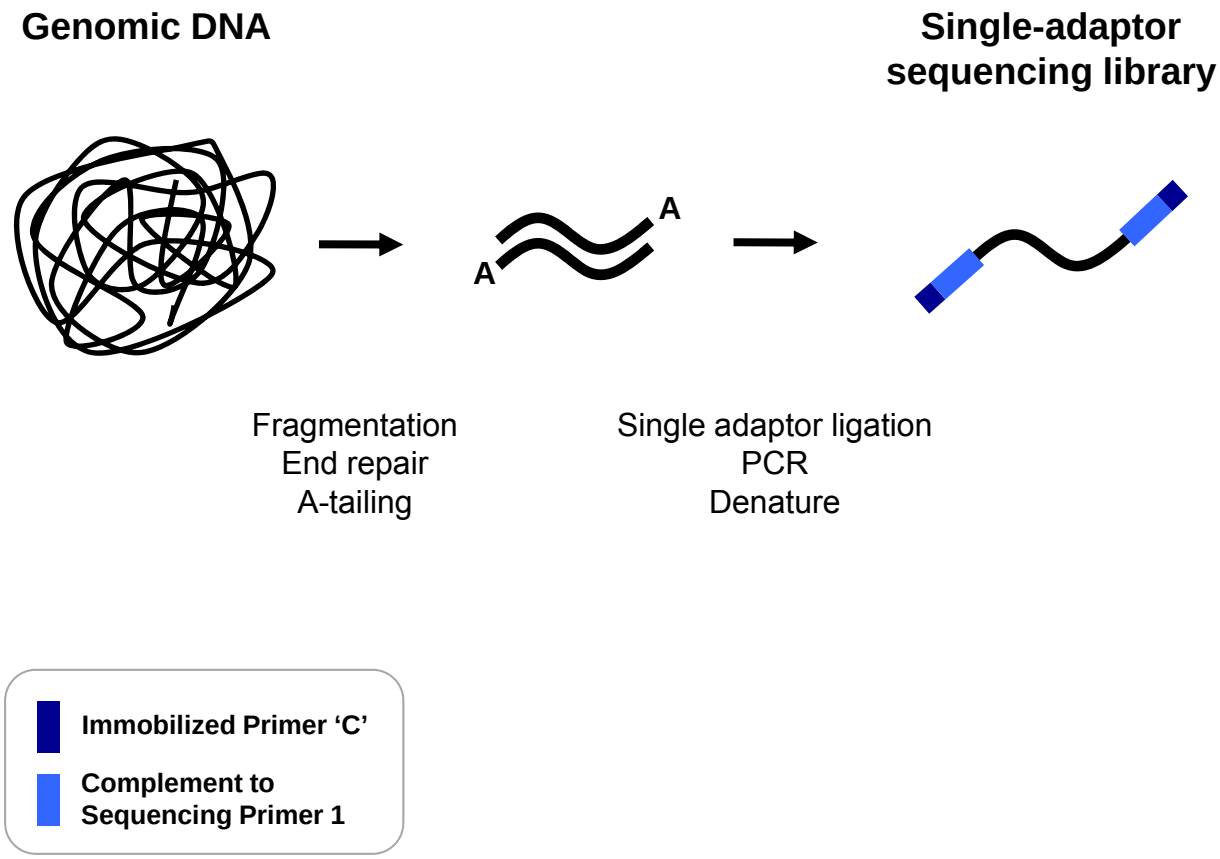
Supplementary Table 3. SNV concordance from OS-Seq and SNP genotyping arrays on a normal tumor pair. We applied OS-Seq-11k analysis to genomic DNA derived from a matched normal – colorectal carcinoma tumor pair. For comparison, we genotyped the two samples with the Affymetrix SNP 6.0 array. In comparing the OS-Seq SNVs to Affymetrix SNPs, we observed a high concordance of 99.8% for the normal and 99.5% for the tumor.

Sample	2722A	2736A
Source of array SNP data	Affymetrix SNP 6.0	Affymetrix SNP 6.0

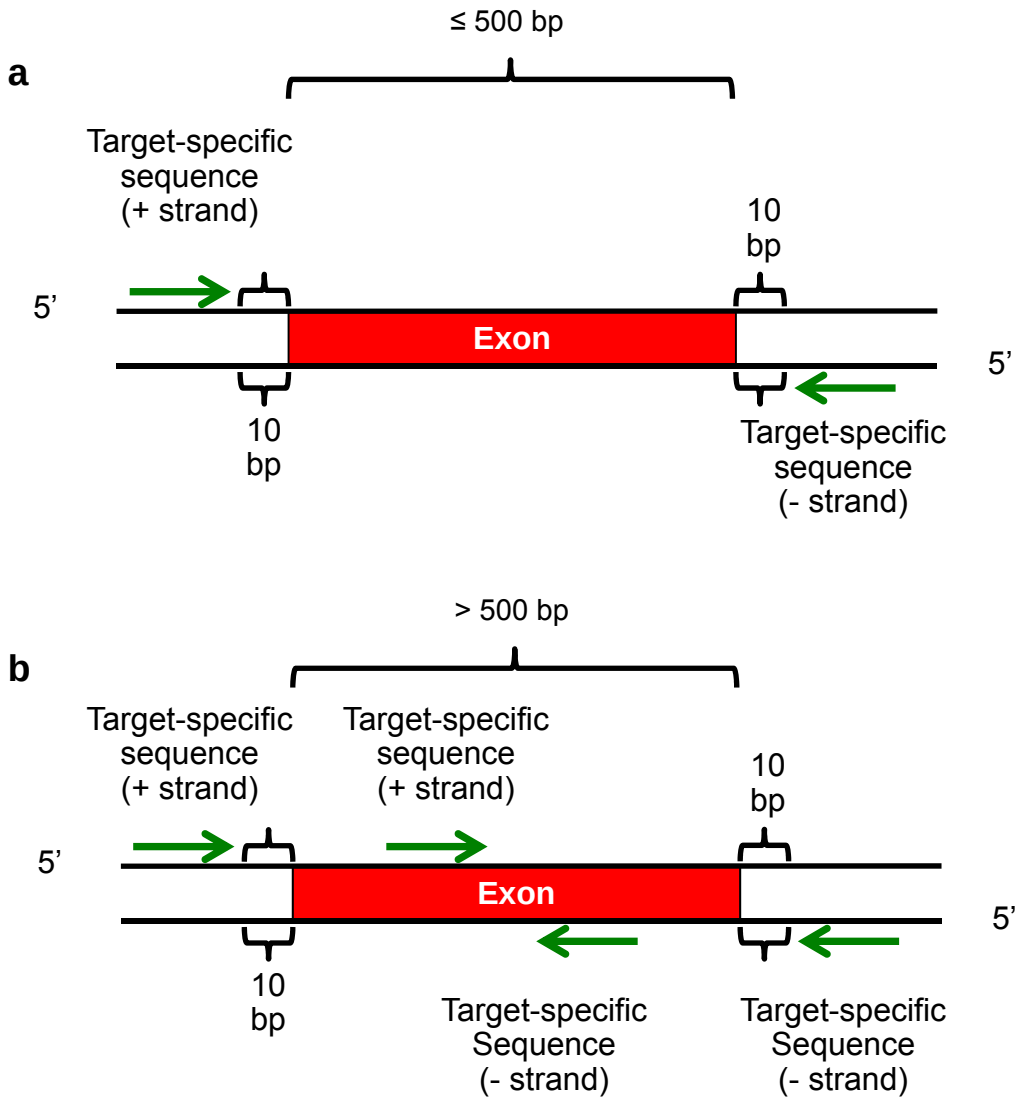
OS-Seq assay	OS-Seq-11K	OS-Seq-11K
Total OS-Seq SNVs	871	727
OS-Seq SNVs concordant with array SNPs	546	418
Array SNPs not called by OS-Seq	1	2

Supplementary Table 4. OS-Seq-366 primer probe sequences. This is provided as a separate table. This table lists the oligonucleotides used for OS-Seq-366.

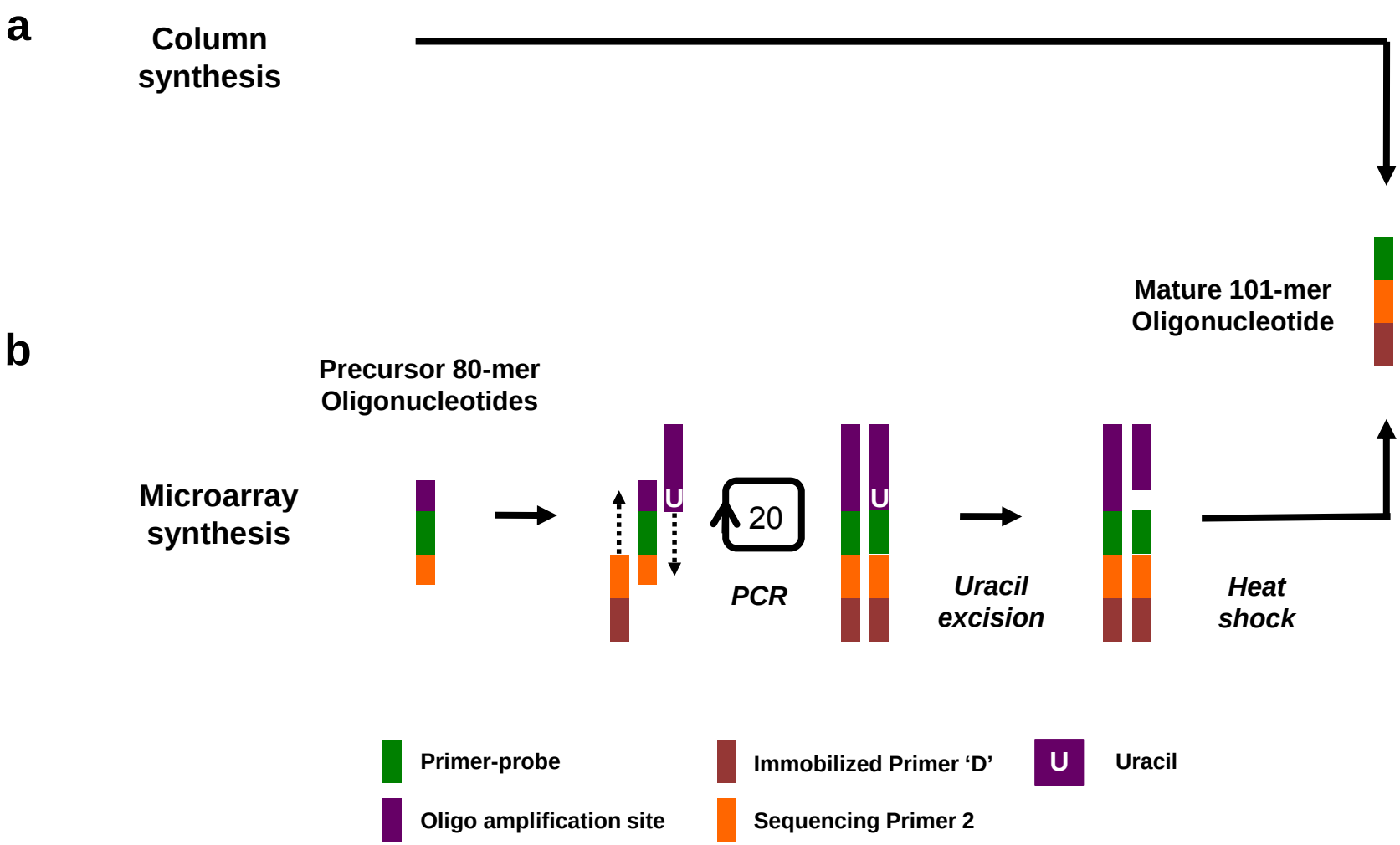
Supplementary Table 5. OS-Seq-11k primer probe sequences. This is provided as a separate table. This table lists the oligonucleotides used for OS-Seq-11k.



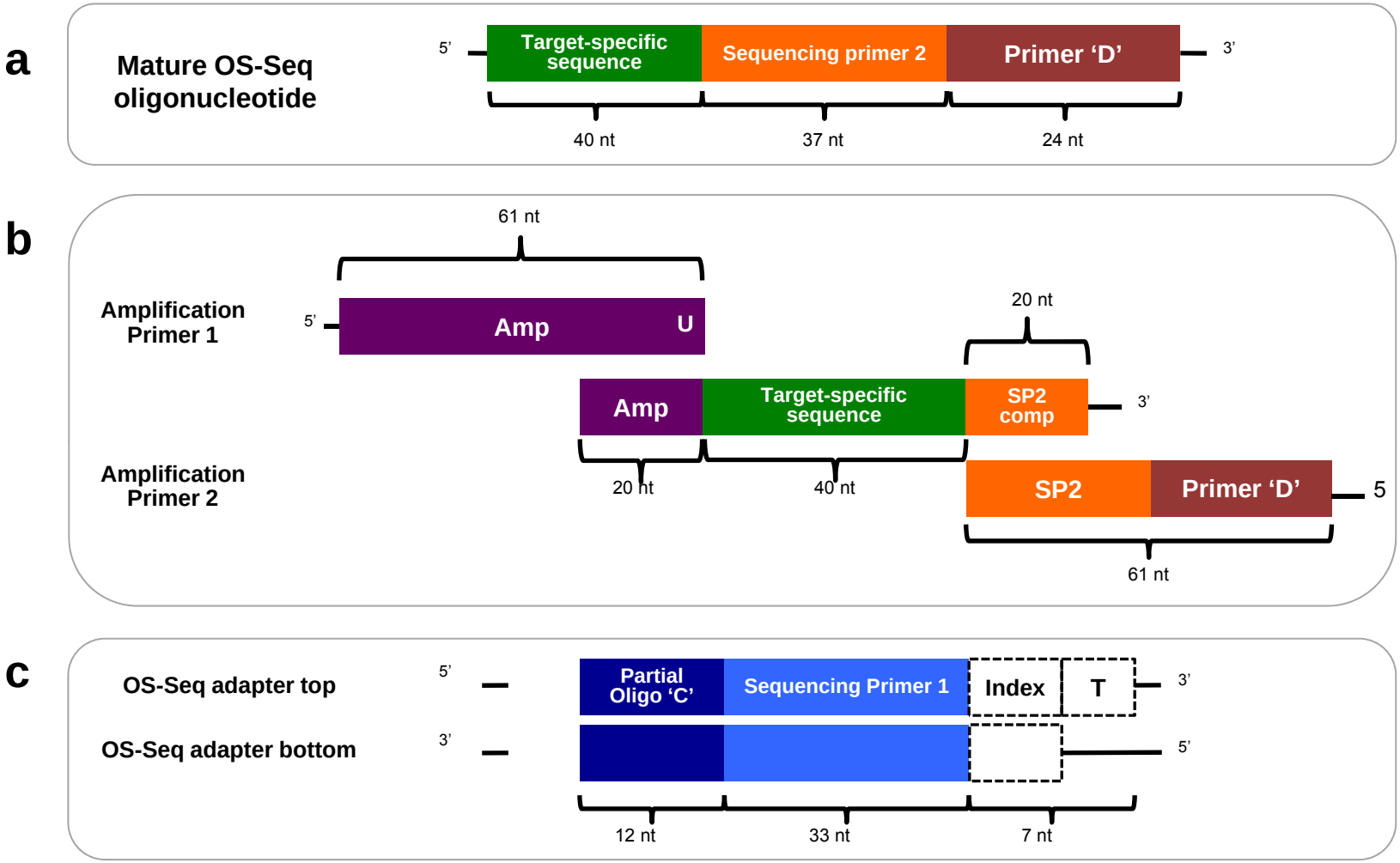
Supplementary Figure 1. Sequencing library preparation for OS-Seq. A general scheme of genomic DNA fragmentation, end repair, A-tailing, adaptor ligation and PCR used in the preparation of OS-Seq libraries.



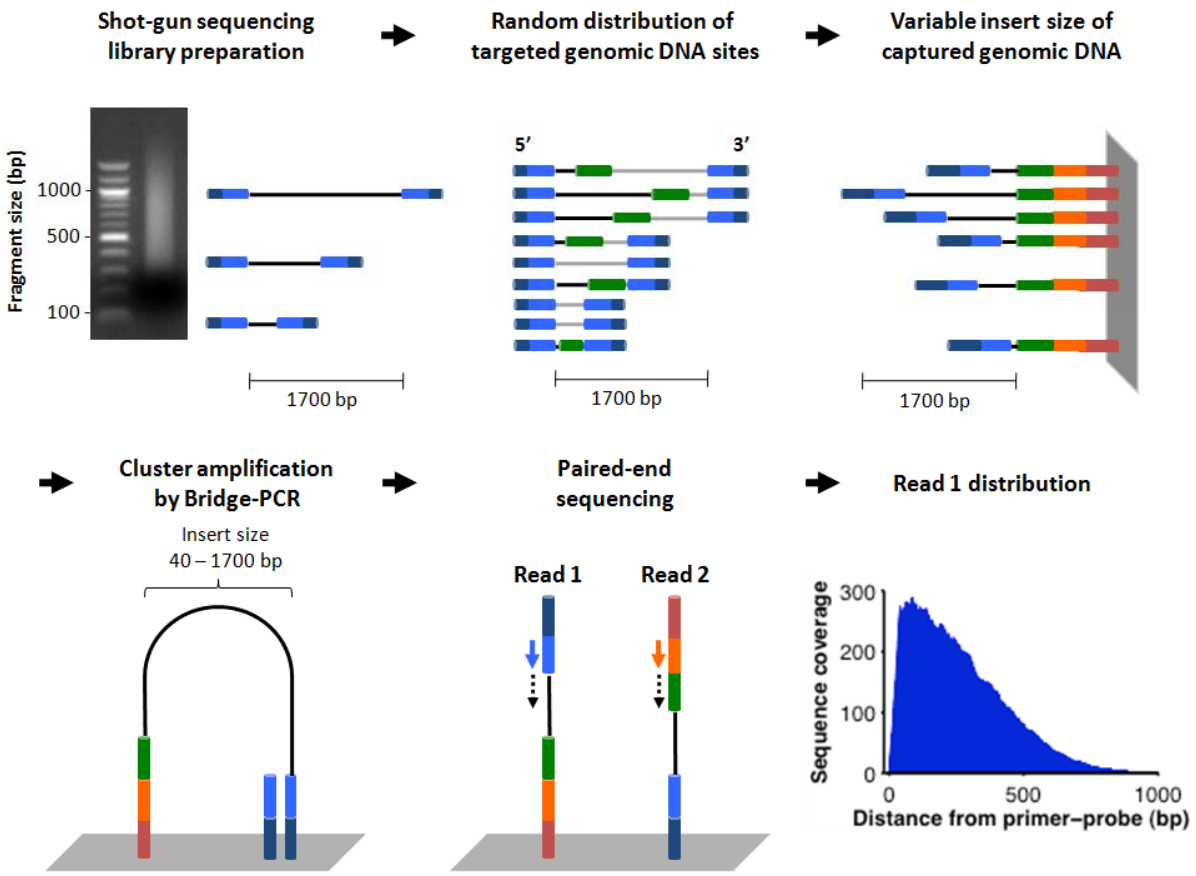
Supplementary Figure 2. Design strategies for OS-Seq. **(a)** Primer-probes were placed 10 bases from the exon or **(b)** tiled every 500 bases inside large exons.



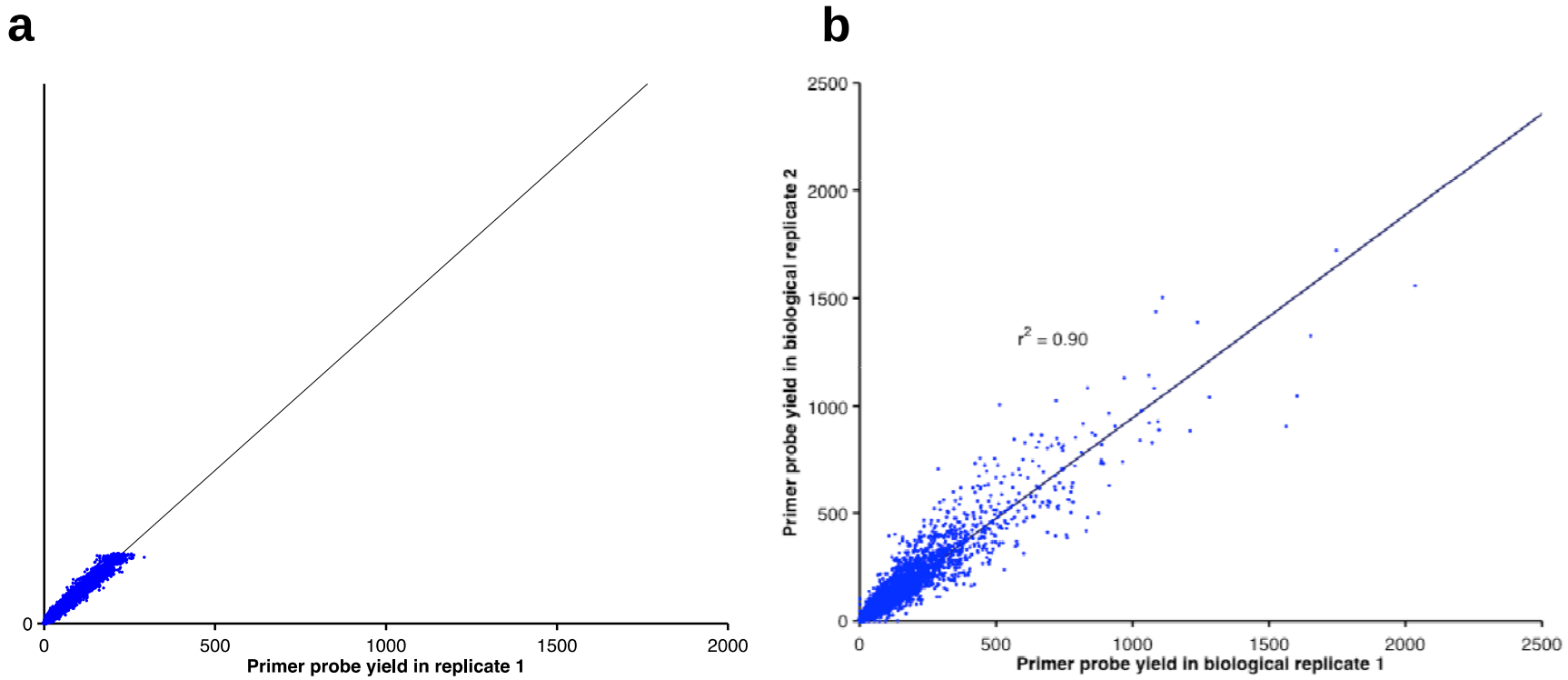
Supplementary Figure 3. Generation of OS-Seq oligonucleotides. **(a)** Column-synthesis yielded large amount of mature 101-mer OS-Seq oligonucleotides that were readily usable in the assay. **(b)** Microarray-synthesis was applied to generate high-content oligonucleotide pools. Precursor oligonucleotides were amplified using primers that incorporated additional sequences into oligonucelotides. Uracil excision was applied to cleave the amplification primer site from the coding strands of the OS-seq oligonucleotides.



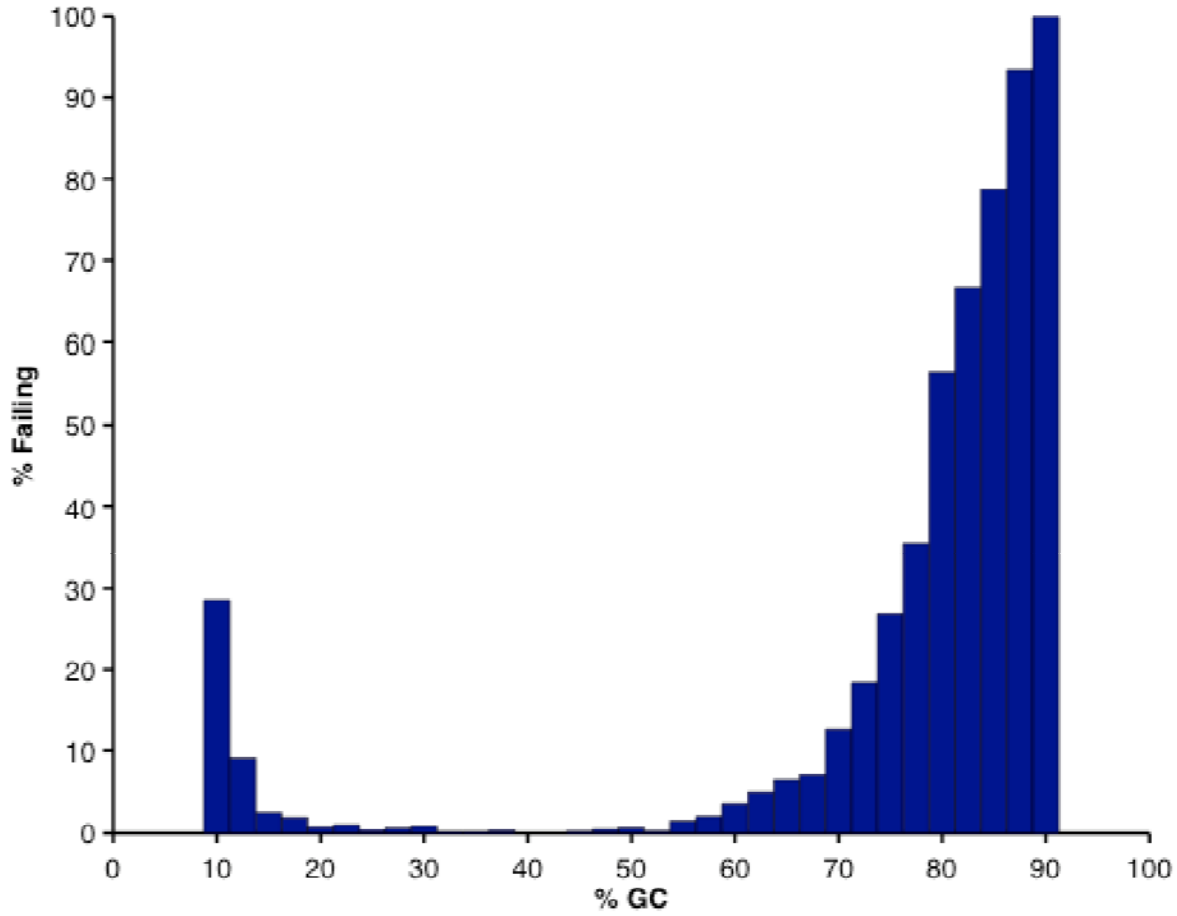
Supplementary Figure 4. Structures of oligonucleotide components in OS-Seq. **(a)** Mature 101-mer OS-Seq oligonucleotides contained target-specific site and sequences encoding for sequencing primer 2 and flow cell primer 'D'. **(b)** Microarray-synthesized oligonucleotides were amplified using primers that incorporated uracil to the 5' end of the OS-Seq oligonucleotide and additional active sites for sequencing. **(c)** Adapter for OS-Seq contained T-overhang for sticky-end ligation to the A-tailed genomic fragments. In addition, indexing sequences as well as flow cell primer 'C' site were present in the dsDNA adapter.



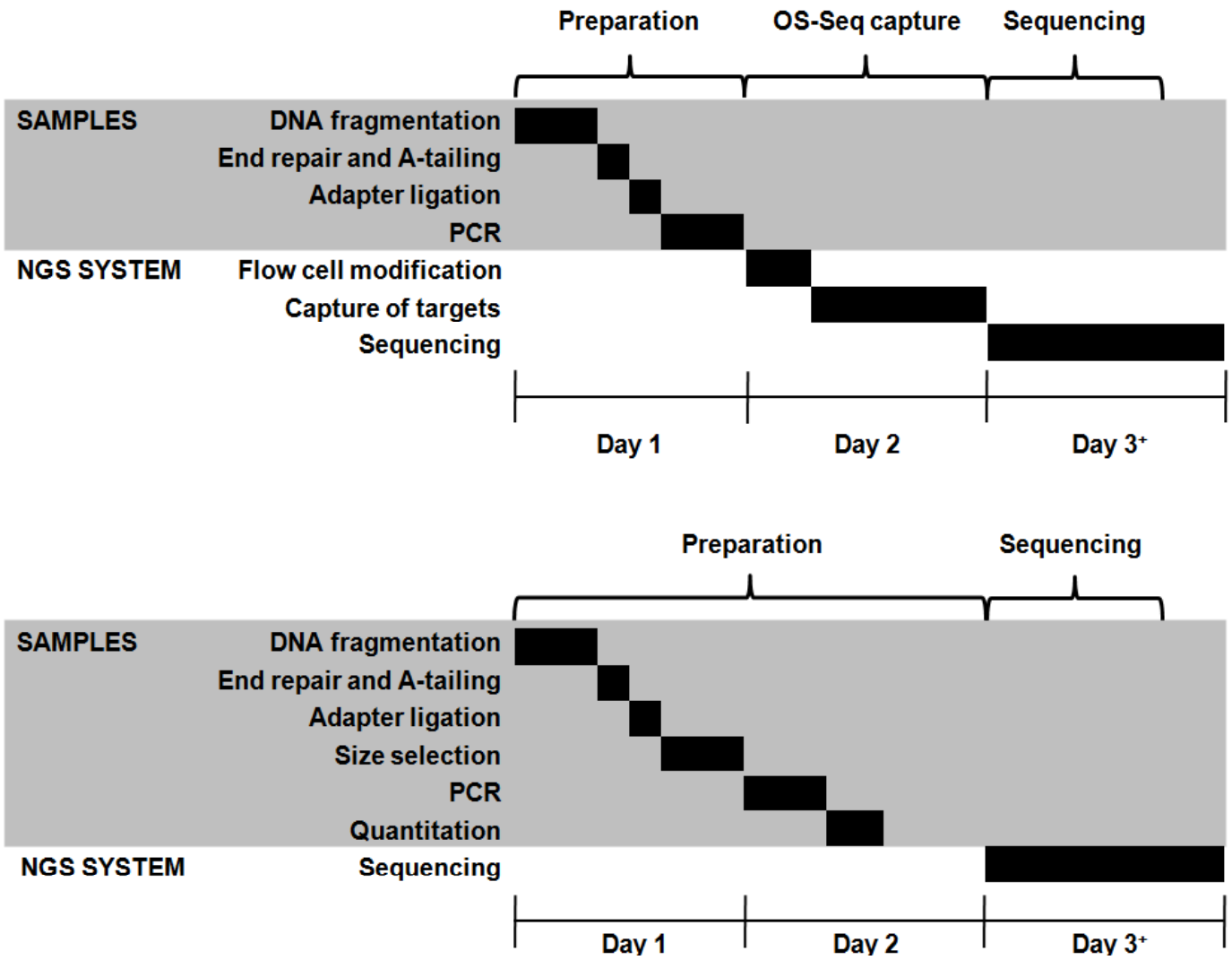
Supplementary Figure 5. Description of insert size distributions encountered in OS-Seq data. We produced genomic DNA fragments between 200 and 2kb. Sequencing library preparation adds common adapter to the ends of the fragments. PCR amplification distorts the fragment size distribution further. Target sites are randomly distributed within the single-adapter library fragments. Library fragments were immobilized on the flow cell and the distance between primer-probe and adapter defined the size of a genomic DNA insert. Bridge-PCR is applied to amplify immobilized target DNA (generally, solid-phase PCR preferentially amplifies shorter fragments). After cluster amplification and processing, immobilized fragments are sequenced using two sites. Read 1 originates from the genomic DNA and Read 2 is derived from the synthetic primer-probes. Read 1 is used for assessing the genomic DNA sequence from OS-Seq data.



Supplementary Figure 6. Reproducibility of OS-Seq. **(a)** Technical reproducibility of OS-Seq. Two identical libraries were analyzed using OS-Seq. Sequencing yields of individual primer-probes were compared between technical replicates. **(b)** Biological reproducibility of OS-Seq. Two different genomic DNA libraries were prepared using indexed adapters. Libraries were analyzed in the same OS-Seq experiment. In the figure, primer-probe specific capture yields are compared between two independent biological replicates.



Supplementary Figure 7. Effect of GC content on targeting yield. To analyze the effect of GC content in the efficiency of primer-probes, we determined the GC content of each target-specific primer-probe sequence. We classified primer-probes that were failing (captured 0 targets). Proportions of failing primer-probes were compared between different %CG content categories. The X-axis represents the percentages of the sorted CG categories and y-axis represents the proportion of failed primer-probes within each GC content category.



Supplementary Figure 8. Comparison of the processing workflow for OS-Seq and genome shotgun library creation methods. This workflow comparison shows that OS-Seq targeted sequencing is comparable to complete genome sequencing process.