Supporting Information

SI Materials and Methods

Sample Information

Plasma was purified from 1348 healthy individuals and 883 patients with cancer using Qiagen kit catalog #937556 (QIAsymphony DSP Circulating DNA Kit) or Biochain kit catalog #K5011625MA. All individuals participating in the study provided written informed consent after approval by the institutional review board at the patients' participating institutions. Their demographic information is included in Dataset S6.

Primer Development

We first calculated the frequency all possible 6-mers (4^6 = 4096) in the genomic regions within the hg19 RepeatMasker track (Fig. S1). Next, we calculated the frequency of all possible 4-mers (4^4 = 256) within 75 bp upstream or downstream from the 6-mers. Joining the 6-mers with the 4-mers generated 2,097,152 candidate pairs. We narrowed these pairs based on the number of unique genomic loci expected from their PCR-mediated amplification, the average size between the 6-mer and its corresponding 4-mers, and the distribution of these sizes, aiming for a unimodal distribution. This filtering criteria generated 7 potential k-mer pairs that were each used to design primer pairs for PCR. (Supplementary Table 1). Each primer pair had with a 6-mer at the 3'-end of one primer and a 4-mer at the 3'-end of the other primer (Fig. S1). Nucleotides upstream of the 6-mers and 4-mers within the primers were then added, informed by the most common nucleotides at the predicted amplicons and their predicted melting temperatures; optimally, all amplicons generated with a specific primer pair should have identical melting temperatures. The seven primer pairs (Supplementary Table 1) that were generated in this way were then tested on genomic DNA in pilot tests. Two of the 7 primer pairs (REAL1 and REAL2) outperformed the remaining 5 primers in tests, as assessed by the number of unique loci that could be amplified and the size distribution of the amplicons. After further experimental testing on 100 peripheral blood samples, REAL-1 was chosen for all of the experiments described in the main text.

Bowtie2 was used to align reads of the amplicons to the human reference genome assembly GRC37 (1). Because we expected many amplicons to have non-perfect primer matching, we only searched for a section of the reverse primer (AGTC, CCCA, TACT, or ACTT) during primer stripping. With REAL-1 primers, an average of 51.1% of the total reads could be uniquely aligned in this fashion. In Fig. S2 and in Dataset S1, we report the size of the unique genomic regions that were identified by this alignment. The average sizes of these genomic regions (bases between primers) was 43 bp. This is an estimate that does

www.pnas.org/cgi/doi/10.1073/pnas.1910041117

not fully capture any amplicons that are very long or those in which there was more than one sequence that closely matched the reverse primer within a few hundred bases of the forward primer. A more precise way to calculate the size of each amplicon would be to perform paired end sequencing. From paired end sequencing of 40 different plasma and 40 different peripheral white blood cell samples, we found that 19% of the amplicons could be more accurately mapped if paired end sequencing were used. While paired end sequencing can thereby provide more accurate information of amplicon sizes, it is significantly more costly and time consuming that using single end sequencing. Moreover, the more accurate sizing of amplicons did not alter their mapping, which was the key to aneuploidy analysis. REAL1 was theoretically able to amplify up to 745,184 repetitive elements (Supplementary Data 1). However, an average of only 350,000 repetitive elements were observed among the 2231 plasma samples evaluated in this study. There are several potential reasons for the discrepancy between the potential number and the actual observed number of amplicons: 1) Polymorphisms within the amplified sequences may cause misalignment and result in "missing amplicons;" 2) Polymorphisms within the primers may result in amplification failures. 3) Each amplicon has a different PCR efficiency based on GC content, size and other variables. Low efficiency amplicons may be outcompeted during PCR and not be present. 4) Long amplicons may be absent in cell free DNA due to the small sizes of the DNA fragments in cell free DNA. 5) The amount of sequencing performed may not be high enough to observe every amplicon, especially those with low PCR efficiency. 6) Some repetitive elements may not be present in every individual.

Within REAL-1 primer-generated amplicons, we identified 52,762 polymorphisms. The average number of heterozygous sites per patient was 2,200 and these sites could be used to measure allelic imbalance, identify samples, and determine whether samples had been accidentally mixed together. Scripts for allelic imbalance are in a github repository and available upon request. Using the same SNPs, synthetic experiments were used to estimate that sample mixing could be detected when the amount of one DNA sample was >4% of the amount of the second DNA sample in a given mixture.

Experimental Protocol

PCR was performed in 25 uL reactions containing 7.25 uL of water, 0.125 uL of each primer, 12.5 uL of NEBNext Ultra II Q5 Master Mix (New England Biolabs cat # M0544S), and 5 uL of DNA. The cycling conditions were: one cycle of 98°C for 120 s, then 15 cycles of 98°C for 10 s, 57°C for 120 s, and 72°C for 120 s. Each plasma DNA sample was assessed in eight independent reactions, and the amount of DNA per reaction varied from ~0.1 ng to 0.25 ng. A second round of PCR was then performed to add dual indexes (barcodes) to each PCR product prior to sequencing. The forward and reverse primers used for the second round of PCR are listed in Supplementary Table 1. The second round of PCR was performed in 25 uL reactions containing 7.25 uL of water, 0.125 uL of each primer, 12.5 uL of NEBNext Ultra II Q5 Master Mix (New England Biolabs cat # M0544S), and 5 uL of DNA containing 5% of the PCR product from the first round. The cycling conditions were: one cycle of 98°C for 120 s, then 15 cycles of 98°C for 10 s, 65°C for 15 s, and 72°C for 120 s. Amplification products from the second round were purified

with AMPure XP beads (Beckman cat # a63880), as per the manufacturer's instructions, prior to sequencing. As noted above, each sample was amplified in eight independent PCRs in the first round. Each of the eight independent PCRs was then re-amplified using index primers in the second PCR round. The sequencing reads from the 8 replicates were summed for the bioinformatic analysis but could also be assessed individually for quality control purposes.

All oligonucleotides were purchased from IDT (Coralville, Iowa). Massively parallel sequencing was performed on an Illumina HiSeq 4000. During the first round of PCR, degenerate bases at the 5' end of one of the primers were used as molecular barcodes (unique identifiers, UIDs) to uniquely label each DNA template molecule (2). This ensured that each DNA template molecule was counted only once, as described in (2). In all instances in this paper, the term "reads" refers to uniquely identified reads (UIDs). Depending on the experiment, each read was sequenced on average 1.17 times. An average of 13.2 million reads per sample (IQR 7.9M to 15.2M) was assessed. If multiple reads had the same UID, we required at least 50% of the reads to map to the same genomic location. Reads with the same UID, but with discordant genomic locations were discarded from analysis.

ReqlSeqS Bioinformatic Analysis

The Within-Sample Aneuploidy DetectiOn (WALDO) approach was developed to detect the presence of aneuploidy in amplicon sequencing reads (3). Unlike conventional WGS approaches that assess aneuploidy, WALDO does not compare normalized read counts in a test sample to the normalized read counts in a panel of normal samples. Direct comparisons are subject to batch effects and can introduce new artifacts (4). WALDO attempts to mitigate problematic batch effects using a within sample comparison. We tailored this approach for our RealSeqS assay and made several analytical improvements (Fig. S2). The major modifications included a new normalization step, a new way to evaluate copy number changes involving small regions of chromosome arms, and an improved way to detect genome-wide aneuploidy, as described below. These analytical improvements coupled with the increased genomic density of amplicons achieved with REAL1 primers enabled greater sensitivity as well as the detection of focal amplifications and deletions less than 1 Mb in size, which was not possible with FAST-SeqS.

Normalization

We employed a new method of normalization which mitigated the impact of highly correlated chromosome region amplifications among samples and ultimately reduced the variability among samples. To perform this normalization, we employed the following steps:

Normalization Training: For all controls (n= C)

1) Bin read counts for each control sample into 5,344 autosomal intervals of 500 kb.
2) Divide the read counts by the total coverage to control for coverage differences.
3) Using the entire set of controls, project the 5,344 500kb intervals into PCA space using the built-in R function "prcomp". Principal component analysis is a statistical procedure that uses an orthogonal transformation to convert a set of observations of possibly correlated variables into a set of values of linearly uncorrelated variables called principal components. Each principal component is a linear combination of the 5,344 500kb intervals.
4) Store the linear combinations of the 5,344 500kb intervals so that test samples can be projected into the same PCA space.
5) Store the first 5 PCA dimensions as a 5xC matrix.
6) Define a new variable termed "correction factor".

   Unseen hidden variables can impact interval read counts. Because all control samples are euploid, we assume that read count variability for a particular interval will be caused by random noise and hidden variables.

   In the prior steps, we calculated the principal components which serve as proxies for unseen hidden variables.

   Ultimately, we would like to learn the relationship of hidden variables to read counts for a particular interval. By determining the relationship of principal components to the expected read counts for a particular interval, the read counts for a future unlabeled test sample can be appropriately adjusted given a its principal components.

   Note: you can only directly calculate the correlation factor term using the equation below for the euploid control samples. A future unlabeled test sample may or may not euploid. If the sample is not euploid, read count variability within the interval maybe caused by amplifications/deletions not just random noise or hidden variables.

$$Correction\ Factor\ for\ 500kb\ Interval_i = \frac{\mu_i}{Observed_i}$$

$$where:$$

$$\mu_i\ is\ the\ numerical\ mean\ of\ read\ counts\ (after\ controlling\ for\ coverage\ differences)$$

$$calculated\ for\ all\ controls\ in\ Interval_i$$

$$Observed_i\ is\ the\ observed\ read\ depth\ in\ Interval_i\ (after\ controlling\ for\ coverage\ differences)$$

7) For each control sample, calculate the "correction factor."
8) Store every correction factor as a 1xC vector.
9) Define a regression model using the following equation (i=1).
   a. Correction factors are calculated in Step 6.

b. PCA coordinates are defined in Steps 2-4.
c. β parameters are undefined.

$$Correction\ Factor\ for\ 500kb\ Interval_i = \beta_{oi} + \beta_{1i} * PCA_1 + \beta_{2i} * PCA_2 + \beta_{3i} * PCA_3 + \beta_{4i} * PCA_4 + \beta_{5i} * PCA_5$$

10) Estimate the β parameters using the built-in linear regression function "lm" in R. The correction factor is a 1xC vector from Step 8 and the PCA coordinates are a 5xC matrix from Step 5 (i=1).
11) Store the β parameters for Interval i (when i=1).
12) Repeat Steps 6-11 for the remaining 5,343 intervals (i=2…5344). At the conclusion of this process, we have a matrix of 6 columns (β parameters) by 5,344 (One for each 500kb interval).

Normalization of test samples:

1) Bin read counts for a new test sample into 5,344 500kb intervals.
2) Divide the observed read counts by the total coverage to control for coverage differences.
3) Using the predefined linear combinations defined above during PCA analysis (Step 4), calculate the test sample's principal components.
4) Using the regression β parameters from above, estimate the correction factor for the new test sample on Interval i=1 (ie how much should the sample differ from the average normal sample given its principal components).

$$Correction\ Factor\ for\ 500kb\ Interval_i = \beta_{oi} + \beta_{1i} * PCA_1 + \beta_{2i} * PCA_2 + \beta_{3i} * PCA_3 + \beta_{4i} * PCA_4 + \beta_{5i} * PCA_5$$

5) Multiply the test sample's observed read count in Interval 1 by the estimated correction factor.
6) Repeat Steps 4 and 5 for the remaining 5343 intervals.

Scripts for normalization are in a github repository and available upon request.

Copy number variants affecting relatively small chromosome regions

The original WALDO method required the specification of a particular genomic region of interest (usually an entire chromosome arm) and then calculate the statistical significance of the desired region. For REALSeqS, we incorporated the ability to detect copy number variants affecting relatively small regions of a chromosome arm. To do so, we calculated the log ratio of the observed test sample and WALDO predicted values from every 500 kb interval across each chromosomal arm. Using the log ratio, we applied a circular binary segmentation algorithm (5) to find copy number variants throughout each chromosome arm. Any copy number variant ≤ 5Mb in size was flagged. Before calculating the statistical significance across each chromosome arm, these flagged CNVs were removed. Because we were

interested in chromosomal abnormalities affecting a large part of a chromosome arm, these small CNVs were not critical for achieving the main goals of the current study. However, the small CNVs could be used to assess microdeletions or microamplifications, such as those occurring in DiGeorge Syndrome (chromosome 22q11.2 or in breast cancers (chromosome 17q12) (Fig. 3 in main text), and for other applications of RealSeqS in the future.

Genome Wide Aneuploidy

A two-class support vector machine (SVM) (6) was trained to discriminate between euploid samples and aneuploid samples. The training set contained a negative class of 1348 presumably euploid plasma samples from normal individuals containing at least 2.5 M reads and 2651 aneuploid samples. The aneuploid class contained a mixture of 2016 *in silico* simulated samples and 635 actual aneuploid samples. SVM training was done with the e1071 package in R, using radial basis kernel and default parameters (7). Each sample had 39 Z-score features, representing chromosome arm gains and losses. During training, the positive class was randomly sampled so that the positive class was 10% the size of the negative class. The positive class was randomly sampled at a ratio of two real samples to one *in silico* simulated sample. Ten iterations of this procedure were performed. The final genome wide aneuploidy score was the average of the raw svm score across the 10 iterations.

Sample Exclusion Criteria

To ensure that all samples included in the results section of paper were of high quality, we developed several exclusion criteria. 1) Samples with less than 2.5M reads were excluded. 2) Samples with sufficient evidence of contamination from at least two genetically unrelated individuals were excluded. To be labeled as contaminated, the sample had to have at least 10 significant allelic imbalanced chromosome arms (z score >= 2.5) and fewer than ten significant chromosome arms gains or losses (z >= 2.5 or z<= -2.5). Allelic imbalance is determined from SNPs, while gains or losses were assessed through WALDO (3). A large number of chromosome arms with allelic imbalance in the absence of a large number of gains or losses indicated that the tested sample contained DNA from at least two genetically unrelated individuals. 3) We excluded samples in which more than 8.15% of the amplicons were larger than 50 base pairs between the forward and reverse primers. Such samples were likely to be contaminated with leukocyte DNA, as inferred from the comparison to RealSeqS data in leukocyte DNA and the distribution of this metric in plasma DNA (Fig. S3). This criterion excluded 16 samples from cancer patients and 27 samples from normal individuals (Fig. S3). 4) An additional QC metric was designed because reads that map to chromosome arms 2q, 3q, 4q, 5q, 6q, 8q, and 13q were highly correlated, as indicated by the matrix in SI Appendix Table S7. The QC metric was defined in the following way:

$$QC\ Dynamic\ Range\ Metric = \sum_{i}^{2q,3q,4q,5q,6q,8q,13q} \frac{Reads\ on\ chr_i}{\sum_{j=1}^{39} Reads\ on\ chr_j}$$

The distribution of this metric had long tails (Fig. S4). Samples which were outliers as defined by this metric (<0.2320 or > 0.2450). This QC metric removed 30 cancer samples and 24 normal samples. Nine samples failed both DNA size and QC.

Generation of *in silico* simulated samples harboring aneuploidy.

We first selected data from 84 presumably euploid plasma samples, each containing at least 10 million reads. *In Silico* simulated aneuploid samples were created by adding (or subtracting) reads from several chromosome arms to the reads from these normal DNA samples. We added or subtracted the reads from 1, 10, 15, or 20 chromosome arms to each sample. The additions and subtractions were designed to represent neoplastic cell fractions ranging from 0.5% to 1.5% and resulted in *in silico* simulated samples containing exactly ten million reads. The reads from each chromosome arm were added or subtracted according to the pseudocode in Fig S6 and S7. For example, when we modeled five chromosome arms that were lost, each was lost to the identical degree and we did not incorporate tumor heterogeneity into the model. Furthermore, we did not create *in silico* simulated samples containing more than three of any chromosome arm; e.g. 4 copies of chromosome 3p. This simplified approach did not comprehensively cover all biologically plausible aneuploidy events. However, limiting the possible combinations of altered arms made sample generation computationally tractable, and the resulting support vector machine appeared to work well in practice. The synthetically generated samples in which reads from only a single chromosome arm were added or subtracted enabled us to estimate the performance of WALDO when only a single chromosome arm of interest was gained or lost.

Comparison to Various Massively Parallel Sequencing Technologies

We selected 10 publicly available plasma samples from normal individuals on which whole genome sequencing had been performed (8). Each of the ten samples had been sequenced at a depth of ~144 M reads. The authors performed bioinformatic filters to remove highly polymorphic locations (9) and common insertions/deletions (10). Reads on the sex chromosomes, contigs, and acrocentric chromosome arms were dropped. The fraction of reads for each of the 39 non-acrocentric chromosome arms were calculated from the remaining reads. The fraction of reads for each of the 39 non-acrocentric autosomal chromosome arms is reproduced in Supplementary Table 2.

We then selected ten samples of normal individuals studied by FAST-SeqS with an average of 10 M reads and ten samples from normal individuals studied by ReqlSeqS with an average of 28 M reads. The fraction of reads on each chromosome arm is recorded in Supplementary Tables 3 and 4, respectively.

In silico simulated aneuploid samples were generated at 5% cell fraction to represent the amount of fetal DNA typically present in noninvasive prenatal testing (NIPT) to compare whole genome sequencing, FAST-SeqS, and ReqlSeqS. We evaluated trisomies and monosomies for each of the 39 chromosome arms as well as the 1.5 MB DiGeorge deletion (22:19009792-20509792). Next, we generated focal amplifications of ERBB2 (17:37819167-37911679 20 copies) at 1% cell fraction to represent the amount of tumor DNA typically present in liquid biopsies from late stage patients. FAST-SeqS does not have the spatial genomic density to cover the region of interest of ERBB2 and could not be evaluated for this purpose.

Aneuploidies were generated at various sequencing depths of 2M, 10M, 40M reads. The statistical significance was calculated using a simple z score defined below. Sensitivities were calculated based on Z > 2.575 and Z < -2.575 (alpha=0.01).

$$z_{chr@Depth} = \frac{Observed_{@Depth} - \mu_{normal\ panel}}{\sigma_{normal\ panel\ @Depth}}$$

Reduced Requirements for DNA Input

A major advantage of amplicon sequencing is the reduced requirement for input DNA. To test whether ReqlSeqS could reliably detect aneuploidy with less than one genome equivalent, we evaluated trisomy 21 samples and normal euploid DNA at various amounts of input DNA (3 to 225pg). The relationship of reads to DNA was based on negative controls (water wells with no DNA) and the known concentration of the euploid control (Sample DNA concentration in picograms=7e-5*Read Depth-0.5196). Trisomy 21 was scored as detected in samples with z>5. No other arms were aneuploid in the trisomy 21 samples at this z-score and no arms were aneuploid in the euploid controls at this z-score (Supplementary Data 2).

Detection of leukocyte DNA in cell free DNA

Leukocyte DNA (gDNA) has an average size of >1000 bp while cell-free plasma DNA has an average size of < 200 bp. DNA size impacts PCR efficiency and long amplicons may not be present in cfDNA. ReqlSeqS enables the detection of leukocyte DNA contamination by virtue of the amplicons generated with REAL-1 primers.

We selected 50 plasma samples aand 50 gDNA samples. We found 1241 amplicons (Supplementary Data 3) that were not present in the plasma samples (< 5 reads across all samples) but were present in the gDNA samples (>1000 reads across all samples).

We mixed DNA from leuokocytes into cell free DNA from a euploid plasma sample at various dilutions ranging from 4% to 54%. The fraction of reads mapping to the 1241 amplicons in the contaminated samples were more than 3 times higher than the euploid sample (Supplementary Table 5).

Comparison of Aneuploidy Detection and Somatic Mutation Detection

For this study, we evaluated aneuploidy in plasma from 1348 healthy individuals and 883 cancer patients. We selected a cutoff (0.441) that produced 99% specificity in 1348 healthy individuals and calculated sensitivity on the 883 cancer patients. A detailed table of the aneuploidy results is included in Supplementary Data 4.

Cohen et al evaluated 812 plasma samples from healthy patients for somatic mutation detection. We selected an omega cutoff (1.77) that produced 99% specificity and calculated sensitivity on the same 883 cancer patients that were used in our study.

Detection of cancer using a multi-analyte test

Cohen *et al*. (11) demonstrated that combining somatic mutations and protein markers can better predict cancer status of plasma samples than either type of marker alone. We determined whether aneuploidy could be integrated as an additional biomarker into the published logistic regression framework.

In Cohen *et al*., 812 plasma samples were from healthy individuals and 1005 were from cancer patients. Our study analyzed 1348 plasma samples from healthy individuals and 883 from cancer patients. Of the 1348 healthy samples, only 248 overlapped with the original study. All 883 cancer samples were included in the original study. The sample demographic information is provided in Supplementary Data 5.

Using the original 812 healthy samples and the 883 cancer samples, we trained a logistic regression model and assessed performance using ten rounds of tenfold cross validation. A full list of samples and their biomarker values was provided in Supplementary Data 6. Because 564 of the original healthy samples were not analyzed for aneuploidy, we randomly sampled the list of scores from the 1348 normal samples and assigned each missing sample an aneuploidy value. Ten rounds of analysis were performed and each new round, we randomly sampled the collection of 1348 normal scores again to assign the 564 samples a new score.

To account for variations in the lower limits of detection across different experiments, we found the 90th percentile feature value in the healthy training samples. We then found any feature value below that

threshold and set all values to the 90th percentile threshold. This transformation was done for all training and testing samples. This procedure was done for aneuploidy scores, somatic mutation scores, and protein concentrations. The 90th percentile thresholds were listed in Supplementary Table 6. The results from each round and each cross validation fold are included in Supplementary Data 7. The final feature coefficients from the logistic regression model are listed in Supplementary Table 6.

Aneuploidy Tumor Concordance

As noted in the main test, we compared chromosomal gains or losses in the plasma to those observed in primary tumors from the same patients. If RealSeqS data indicating aneuploidy were "real", one would expect that those chromosome arms exhibiting gains in plasma would also exhibit gains in the corresponding primary tumors. Similarly, one would expect that those chromosome arms exhibiting losses in the plasma would also exhibit losses in the corresponding primary tumors. We were able to perform this analysis in 243 instances (214 patients) in which chromosome arm losses or gains were significant (z-scores >4 or <-4) in plasma DNA. Of these 243, 188 (77%) were found to be concordant in their respective tumors at a z-score >2 or <-2 when assessed by FAST-SeqS and previously reported (Supplementary Appendix and Dataset S9). Note that concordance was directional; if a gain of a chromosome arm was found in the plasma, a gain (rather than a loss) had to be identified in the primary tumor and vice versa. Concordance was calculated as the number of concordant arms in the plasma divided by the total number of significant gains and losses in plasma. Chromosome arms 19p and 19q were RealSeqS due to the high false positive rate on these arms.

We did not expect perfect concordance between the plasma and tumors for the following reasons. Plasma samples with focal amplifications or deletions can produce highly significant arm level scores. These focal changes may have been captured with REALSeqS performed on the plasma DNA as a result of the increased spatial genomic coverage but missed in FAST-SeqS performed on the primary tumor DNA. Additionally, using a within sample normalization has been shown to improve performance at low amounts of tumor DNA but the statistical assumptions underlying the normalization principle breakdown with increased amounts of tumor DNA. Evaluating widely aneuploid tumor samples using a within sample normalization approach can produce false negatives, particularly on the smaller chromosome arms.

1. Langmead B & Salzberg SL (2012) Fast gapped-read alignment with Bowtie 2. *Nature methods* 9(4):357.
2. Kinde I, Wu J, Papadopoulos N, Kinzler KW, & Vogelstein B (2011) Detection and quantification of rare mutations with massively parallel sequencing. *Proceedings of the National Academy of Sciences* 108(23):9530-9535.

3. Douville C*, et al.* (Detection of aneuploidy in patients with cancer through amplification of long interspersed nucleotide elements (LINEs). *Proceedings of the National Academy of Sciences* 115(8):1871-1876.
4. Straver R*, et al.* (WISECONDOR: detection of fetal aberrations from shallow sequencing maternal plasma based on a within-sample comparison scheme. *Nucleic acids research* 42(5):e31-e31.
5. Olshen AB, Venkatraman E, Lucito R, & Wigler M (2004) Circular binary segmentation for the analysis of array-based DNA copy number data. *Biostatistics* 5(4):557-572.
6. Cortes C & Vapnik V (1995) Support-vector networks. *Machine learning* 20(3):273-297.
7. Dimitriadou E*, et al.* (2006) The e1071 package. *Misc Functions of Department of Statistics (e1071), TU Wien*.
8. Leary RJ*, et al.* (2012) Detection of chromosomal alterations in the circulation of cancer patients with whole-genome sequencing. *Science translational medicine* 4(162):162ra154-162ra154.
9. Consortium GP (2010) A map of human genome variation from population-scale sequencing. *Nature* 467(7319):1061.
10. Mills RE*, et al.* (2011) Mapping copy number variation by population-scale genome sequencing. *Nature* 470(7332):59.
11. Cohen JD*, et al.* (Detection and localization of surgically resectable cancers with a multi-analyte blood test. *Science* 359(6378):926-930.

Table S1: The 7 primer pairs chosen for experimental analysis. The 6mer and 4mer used during design are in red font

cgacgtaaaacgacggccagtNNNNNNNNNNNNNNNNGGTGAAACCCCGTCTCTACA	REAL1a_z4	cacacaggaaacagctatgaccatgCCTCCTAAGTAGCTGGGACTACAG
cgacgtaaaacgacggccagtNNNNNNNNNNNNNNNNGGTGAAACCCCGTCTCTAC	REAL1b_z4	cacacaggaaacagctatgaccatgCCTCCTAAGTAGCTGGGACTACAG
cgacgtaaaacgacggccagtNNNNNNNNNNNNNNNNGGTGAAACCCCGTCTCTACT	REAL1c_z4	cacacaggaaacagctatgaccatgCCTCCTAAGTAGCTGGGACTACAG
cgacgtaaaacgacggccagtNNNNNNNNNNNNNNNNCATGCCTGTAGTCCCAGCTACT	REAL2_z4	cacacaggaaacagctatgaccatgTGCAGTGGCACGATCATAGCTCACTGCAGCCTTGA
cgacgtaaaacgacggccagtNNNNNNNNNNNNNNNNATAGTGAAACCCCATCTCTACAAAA	REAL3_z4	cacacaggaaacagctatgaccatgCTCCCGAGTAGCTGGGACT
cgacgtaaaacgacggccagtNNNNNNNNNNNNNNNNGGTGAAACCCCATCTCTACAA	REAL4_z4	cacacaggaaacagctatgaccatgCTCCCGAGTAGCTGGGACTAC
cgacgtaaaacgacggccagtNNNNNNNNNNNNNNNNATAGTGAAACCCCATCTCTACAAA	REAL5_z4	cacacaggaaacagctatgaccatgCCCGAGTAGCTGGGACTACA
cgacgtaaaacgacggccagtNNNNNNNNNNNNNNNNGAGGTGGGAGGATTGCTT	REAL6_z4	cacacaggaaacagctatgaccatgAGGCTGGAGTGCAGTGG
cgacgtaaaacgacggccagtNNNNNNNNNNNNNNNNACCAGCCTGGGCAACATA	REAL7_z4	cacacaggaaacagctatgaccatgCCACCATGCCTGGCTAA

Table S2: Fractional Representation of the the REAL1 prim Normal Samples with Whole Genome Sequencing that were used for the NGS technology comparison. Reproduced from Leary et al.

	N1_WGS	N2_WGS	N3_WGS	N4_WGS	N5_WGS	N6_WGS	N7_WGS	N8_WGS	N9_WGS	N10_WGS
1p	0.04817	0.04831	0.04804	0.04841	0.04790	0.04815	0.04817	0.04815	0.04798	0.04812
1q	0.03932	0.03926	0.03927	0.03948	0.03915	0.03929	0.03928	0.03928	0.03916	0.03928
2p	0.03003	0.03003	0.03008	0.03019	0.03001	0.03009	0.03007	0.03000	0.02999	0.03005
2q	0.06194	0.06182	0.06191	0.06190	0.06183	0.06192	0.06190	0.06176	0.06187	0.06175
3p	0.03740	0.03740	0.03730	0.03746	0.03741	0.03744	0.03745	0.03734	0.03745	0.03744
3q	0.04250	0.04237	0.04244	0.04248	0.04234	0.04247	0.04247	0.04239	0.04234	0.04235
4p	0.01790	0.01799	0.01793	0.01775	0.01806	0.01792	0.01792	0.01789	0.01806	0.01801
4q	0.05568	0.05542	0.05550	0.05533	0.05544	0.05541	0.05561	0.05553	0.05548	0.05534
5p	0.01765	0.01772	0.01774	0.01753	0.01785	0.01766	0.01764	0.01774	0.01789	0.01779
5q	0.05516	0.05509	0.05511	0.05510	0.05510	0.05514	0.05520	0.05495	0.05495	0.05508
6p	0.02457	0.02460	0.02458	0.02487	0.02444	0.02466	0.02465	0.02461	0.02445	0.02455
6q	0.04271	0.04254	0.04271	0.04254	0.04264	0.04269	0.04268	0.04267	0.04265	0.04261
7p	0.02024	0.02029	0.02039	0.02037	0.02034	0.02036	0.02032	0.02023	0.02039	0.02033
7q	0.03332	0.03323	0.03327	0.03303	0.03339	0.03322	0.03330	0.03342	0.03334	0.03319
8p	0.01207	0.01203	0.01207	0.01194	0.01212	0.01206	0.01208	0.01212	0.01210	0.01206
8q	0.03978	0.03972	0.03980	0.03964	0.03982	0.03978	0.03976	0.03971	0.03981	0.03974
9p	0.01041	0.01040	0.01034	0.01029	0.01044	0.01036	0.01040	0.01037	0.01041	0.01039
9q	0.02792	0.02798	0.02799	0.02816	0.02793	0.02801	0.02798	0.02790	0.02796	0.02799
10p	0.01516	0.01509	0.01520	0.01524	0.01510	0.01515	0.01518	0.01523	0.01510	0.01510
10q	0.03301	0.03296	0.03307	0.03311	0.03292	0.03302	0.03301	0.03285	0.03290	0.03297
11p	0.01817	0.01815	0.01810	0.01821	0.01814	0.01816	0.01816	0.01804	0.01813	0.01817
11q	0.03020	0.03032	0.03030	0.03026	0.03027	0.03033	0.03028	0.03018	0.03026	0.03032
12p	0.01181	0.01183	0.01180	0.01193	0.01174	0.01186	0.01185	0.01181	0.01179	0.01181
12q	0.04007	0.04001	0.03997	0.03990	0.04005	0.03988	0.04001	0.04008	0.03993	0.04003
13q	0.03760	0.03752	0.03757	0.03760	0.03764	0.03762	0.03748	0.03763	0.03774	0.03764
14q	0.03368	0.03365	0.03363	0.03364	0.03363	0.03360	0.03361	0.03362	0.03358	0.03357
15q	0.02791	0.02797	0.02793	0.02817	0.02785	0.02797	0.02794	0.02790	0.02790	0.02804

16p	0.00800	0.00803	0.00802	0.00787	0.00809	0.00801	0.00797	0.00808	0.00808	0.00808
16q	0.01682	0.01688	0.01689	0.01679	0.01692	0.01680	0.01680	0.01686	0.01687	0.01689
17p	0.00565	0.00567	0.00565	0.00559	0.00568	0.00562	0.00566	0.00571	0.00566	0.00567
17q	0.01748	0.01752	0.01737	0.01754	0.01735	0.01743	0.01748	0.01760	0.01734	0.01744
18p	0.00616	0.00617	0.00620	0.00618	0.00616	0.00613	0.00612	0.00616	0.00619	0.00614
18q	0.02520	0.02522	0.02529	0.02520	0.02530	0.02528	0.02522	0.02516	0.02531	0.02523
19p	0.00460	0.00464	0.00456	0.00449	0.00470	0.00458	0.00461	0.00485	0.00468	0.00467
19q	0.00835	0.00845	0.00839	0.00817	0.00853	0.00828	0.00827	0.00859	0.00854	0.00848
20p	0.00979	0.00991	0.00985	0.00983	0.00991	0.00988	0.00981	0.00982	0.00990	0.00990
20q	0.01279	0.01287	0.01282	0.01278	0.01284	0.01280	0.01280	0.01279	0.01281	0.01284
21q	0.01205	0.01212	0.01212	0.01218	0.01218	0.01216	0.01208	0.01214	0.01218	0.01215
22q	0.00877	0.00884	0.00881	0.00884	0.00878	0.00880	0.00876	0.00885	0.00886	0.00880
#Reads after filtering on 39 chr arms	83410100	89110671	88895663	88022967	119670123	126567284	116737767	116802698	126028642	119613813
#Reads after alignment	231,520,314	224,708,017	161,874,934	223,342,309	288,318,721	285,785,107	226,653,076	279,224,469	257,431,312	287,180,675
%Percent Usage after filtering	0.3603	0.3966	0.5492	0.3941	0.4151	0.4429	0.5151	0.4183	0.4896	0.4165
Assumed Alignment Rate	90.00%	90.00%	90.00%	90.00%	90.00%	90.00%	90.00%	90.00%	90.00%	90.00%

Bowtie2 Supplementary Table 3 Reports Various Alignment Rates for 100bp unpaired reads ranging from 70% for Bowtie (original); 73% for SOAP2; 90-93% for BWA depending on quality score; and 73-97% for Bowtie2 depending on quality score

Table S3: Fractional Representation of the 10 Normal Samples with FAST-SeqS that were used for the NGS technology comparison

	FAST1_N	FAST2_N	FAST3_N	FAST4_N	FAST5_N	FAST6_N	FAST7_N	FAST8_N	FAST9_N	FAST10_N
1p	0.03534	0.03566	0.03553	0.03537	0.03559	0.03530	0.03545	0.03537	0.03567	0.03529
1q	0.03766	0.03745	0.03754	0.03747	0.03736	0.03729	0.03727	0.03731	0.03741	0.03754
2p	0.03148	0.03149	0.03151	0.03125	0.03142	0.03157	0.03161	0.03158	0.03170	0.03150
2q	0.05955	0.05968	0.05996	0.05961	0.05964	0.05956	0.05974	0.05976	0.05973	0.05966
3p	0.03579	0.03607	0.03602	0.03600	0.03618	0.03642	0.03633	0.03602	0.03617	0.03590
3q	0.04850	0.04801	0.04814	0.04826	0.04816	0.04829	0.04810	0.04853	0.04859	0.04842
4p	0.02095	0.02067	0.02099	0.02052	0.02062	0.02111	0.02065	0.02050	0.02088	0.02075
4q	0.06919	0.06893	0.06889	0.06902	0.06897	0.06904	0.06922	0.06870	0.06963	0.06937
5p	0.02200	0.02223	0.02237	0.02181	0.02197	0.02217	0.02217	0.02201	0.02218	0.02204
5q	0.05918	0.05882	0.05885	0.05928	0.05902	0.05924	0.05895	0.05912	0.05891	0.05917
6p	0.01689	0.01688	0.01685	0.01700	0.01691	0.01701	0.01678	0.01693	0.01670	0.01691
6q	0.05094	0.05112	0.05117	0.05121	0.05116	0.05119	0.05130	0.05089	0.05063	0.05127
7p	0.01884	0.01887	0.01872	0.01894	0.01890	0.01857	0.01873	0.01874	0.01873	0.01882
7q	0.03908	0.03906	0.03901	0.03900	0.03900	0.03910	0.03925	0.03917	0.03902	0.03889
8p	0.01534	0.01536	0.01540	0.01527	0.01527	0.01523	0.01504	0.01533	0.01535	0.01521
8q	0.04447	0.04482	0.04451	0.04458	0.04469	0.04445	0.04449	0.04448	0.04450	0.04444
9p	0.01942	0.01928	0.01938	0.01919	0.01948	0.01912	0.01933	0.01927	0.01926	0.01931
9q	0.02142	0.02148	0.02131	0.02151	0.02168	0.02160	0.02162	0.02152	0.02112	0.02144
10p	0.01455	0.01436	0.01465	0.01471	0.01439	0.01465	0.01435	0.01450	0.01434	0.01464
10q	0.03049	0.03042	0.03019	0.03040	0.03048	0.03018	0.03027	0.03059	0.03050	0.03024
11p	0.02194	0.02182	0.02188	0.02218	0.02206	0.02199	0.02189	0.02212	0.02195	0.02190
11q	0.03037	0.03042	0.03036	0.03077	0.03039	0.03029	0.03048	0.03044	0.03058	0.03046
12p	0.01367	0.01357	0.01359	0.01358	0.01351	0.01363	0.01352	0.01341	0.01341	0.01363
12q	0.03940	0.03909	0.03925	0.03894	0.03921	0.03917	0.03905	0.03899	0.03942	0.03946
13q	0.04004	0.04039	0.04024	0.04024	0.04009	0.04024	0.04011	0.04024	0.04030	0.04004
14q	0.03073	0.03094	0.03079	0.03077	0.03090	0.03056	0.03078	0.03037	0.03075	0.03089
15q	0.02354	0.02363	0.02371	0.02335	0.02331	0.02349	0.02353	0.02346	0.02347	0.02316
16p	0.00904	0.00902	0.00873	0.00906	0.00916	0.00914	0.00916	0.00922	0.00902	0.00908

16q	0.01157	0.01149	0.01163	0.01164	0.01145	0.01151	0.01155	0.01173	0.01155	0.01152
17p	0.00537	0.00541	0.00533	0.00543	0.00556	0.00535	0.00527	0.00549	0.00528	0.00537
17q	0.01220	0.01225	0.01229	0.01219	0.01213	0.01223	0.01219	0.01247	0.01212	0.01229
18p	0.00450	0.00462	0.00456	0.00444	0.00446	0.00449	0.00456	0.00450	0.00450	0.00458
18q	0.02279	0.02289	0.02298	0.02288	0.02289	0.02304	0.02329	0.02313	0.02293	0.02284
19p	0.00317	0.00314	0.00318	0.00333	0.00327	0.00330	0.00324	0.00323	0.00308	0.00319
19q	0.00687	0.00693	0.00681	0.00694	0.00689	0.00698	0.00693	0.00694	0.00683	0.00682
20p	0.00909	0.00903	0.00910	0.00918	0.00914	0.00901	0.00915	0.00914	0.00912	0.00920
20q	0.00602	0.00611	0.00605	0.00608	0.00607	0.00611	0.00612	0.00612	0.00605	0.00614
21q	0.01317	0.01324	0.01316	0.01320	0.01316	0.01309	0.01329	0.01324	0.01318	0.01323
22q	0.00543	0.00537	0.00537	0.00542	0.00545	0.00530	0.00525	0.00543	0.00542	0.00541
#Reads after filtering on 39 chr arms	10344939	9374057	8375864	9362636	11245352	8866124	9012516	9506659	8989128	7980360
#Reads after alignment	11352867	10281325	8832100	10267810	11855382	9341784	9885452	10021398	9475035	8751215
%Percent Usage after filtering	0.9112	0.9118	0.9483	0.9118	0.9485	0.9491	0.9117	0.9486	0.9487	0.9119
Assumed Alignment Rate	74.40%	74.40%	74.40%	74.40%	74.40%	74.40%	74.40%	74.40%	74.40%	74.40%

Table S4: Fractional Representation of the 10 Normal Samples with REAL-SeqS that were used for the NGS technology comparison.

	REAL1_N	REAL2_N	REAL3_N	REAL4_N	REAL5_N	REAL6_N	REAL7_N	REAL8_N	REAL9_N	REAL10_N
1p	0.05388	0.05380	0.05386	0.05369	0.05364	0.05385	0.05382	0.05387	0.05394	0.05397
1q	0.03919	0.03913	0.03911	0.03928	0.03915	0.03920	0.03923	0.03906	0.03918	0.03907
2p	0.03163	0.03148	0.03146	0.03163	0.03167	0.03143	0.03173	0.03155	0.03169	0.03174
2q	0.04511	0.04501	0.04516	0.04508	0.04520	0.04505	0.04517	0.04501	0.04512	0.04513
3p	0.02941	0.02951	0.02937	0.02940	0.02961	0.02955	0.02951	0.02948	0.02960	0.02943
3q	0.03298	0.03293	0.03284	0.03303	0.03299	0.03335	0.03326	0.03302	0.03318	0.03311
4p	0.01527	0.01533	0.01531	0.01523	0.01538	0.01539	0.01528	0.01519	0.01530	0.01527
4q	0.03455	0.03467	0.03468	0.03488	0.03486	0.03471	0.03475	0.03454	0.03481	0.03480
5p	0.01142	0.01140	0.01151	0.01155	0.01156	0.01144	0.01150	0.01140	0.01142	0.01148
5q	0.04058	0.04047	0.04042	0.04047	0.04072	0.04033	0.04047	0.04054	0.04049	0.04038
6p	0.02218	0.02238	0.02220	0.02232	0.02223	0.02233	0.02228	0.02221	0.02238	0.02244
6q	0.02991	0.02980	0.02973	0.02999	0.02985	0.02984	0.02995	0.02972	0.03001	0.02990
7p	0.01907	0.01891	0.01897	0.01903	0.01894	0.01896	0.01920	0.01903	0.01894	0.01885
7q	0.03823	0.03843	0.03825	0.03842	0.03841	0.03823	0.03818	0.03826	0.03828	0.03810
8p	0.01525	0.01523	0.01529	0.01524	0.01523	0.01498	0.01524	0.01517	0.01508	0.01507
8q	0.02802	0.02793	0.02791	0.02810	0.02797	0.02796	0.02798	0.02774	0.02804	0.02804
9p	0.01234	0.01236	0.01239	0.01223	0.01226	0.01233	0.01222	0.01229	0.01246	0.01241
9q	0.03104	0.03137	0.03114	0.03103	0.03110	0.03108	0.03088	0.03132	0.03107	0.03120
10p	0.01693	0.01679	0.01695	0.01689	0.01691	0.01680	0.01684	0.01684	0.01684	0.01686
10q	0.03234	0.03242	0.03236	0.03237	0.03223	0.03245	0.03246	0.03243	0.03226	0.03246
11p	0.01615	0.01603	0.01602	0.01610	0.01606	0.01611	0.01601	0.01605	0.01613	0.01611
11q	0.02836	0.02837	0.02845	0.02825	0.02841	0.02833	0.02824	0.02834	0.02831	0.02861
12p	0.01286	0.01279	0.01285	0.01277	0.01289	0.01285	0.01287	0.01284	0.01287	0.01287
12q	0.03992	0.03993	0.04006	0.03987	0.03990	0.03986	0.03979	0.03994	0.04003	0.03988
13q	0.02628	0.02613	0.02620	0.02629	0.02623	0.02636	0.02620	0.02629	0.02640	0.02624
14q	0.03410	0.03407	0.03410	0.03423	0.03419	0.03421	0.03405	0.03414	0.03419	0.03439
15q	0.03656	0.03623	0.03648	0.03601	0.03614	0.03639	0.03620	0.03633	0.03602	0.03600

16p	0.02080	0.02094	0.02057	0.02063	0.02070	0.02062	0.02056	0.02076	0.02071	0.02053
16q	0.02078	0.02085	0.02073	0.02072	0.02058	0.02050	0.02064	0.02064	0.02063	0.02060
17p	0.01685	0.01686	0.01699	0.01686	0.01676	0.01713	0.01687	0.01691	0.01685	0.01692
17q	0.03693	0.03699	0.03738	0.03691	0.03700	0.03709	0.03703	0.03722	0.03674	0.03696
18p	0.00576	0.00578	0.00572	0.00578	0.00577	0.00574	0.00579	0.00577	0.00574	0.00577
18q	0.01626	0.01625	0.01622	0.01644	0.01630	0.01619	0.01625	0.01612	0.01621	0.01630
19p	0.02525	0.02551	0.02559	0.02541	0.02536	0.02557	0.02545	0.02579	0.02554	0.02542
19q	0.02338	0.02339	0.02327	0.02350	0.02333	0.02347	0.02347	0.02354	0.02344	0.02345
20p	0.00812	0.00810	0.00816	0.00814	0.00815	0.00812	0.00811	0.00816	0.00801	0.00811
20q	0.01747	0.01745	0.01743	0.01741	0.01741	0.01731	0.01740	0.01753	0.01744	0.01734
21q	0.01142	0.01150	0.01141	0.01151	0.01153	0.01144	0.01156	0.01152	0.01144	0.01145
22q	0.02340	0.02349	0.02346	0.02330	0.02335	0.02346	0.02356	0.02345	0.02325	0.02335
#Reads after filtering on 39 chr arms	52359796	19912706	26513539	27738534	24302273	20829086	19693809	20953153	28359518	24923934
#Reads after alignment	55002204	20893575	27852787	29128768	25555370	21862716	20313496	21972946	29775871	26190523
%Percent Usage after filtering	0.9520	0.9531	0.9519	0.9523	0.9510	0.9527	0.9695	0.9536	0.9524	0.9516
Assumed Alignment Rate	51.10%	51.10%	51.10%	51.10%	51.10%	51.10%	51.10%	51.10%	51.10%	51.10%

Table S5: Prediction of gDNA contamination in plasma.

Sample Name	Total Reads	Reads that map to 1241 amplicons	Fraction of Reads that map to 1241 amplicons	Fold Change Compared to cell Free DNA	Fold Change compared to the average cell free DNA sample	p-value (Binomial Test for Proprotions)
gDNA control	9013937	37598	0.00417	62.284	40.89356	0.000000
euploid cell free DNA	9810368	657	0.00007	1.000	0.656568	1.000000
54% gDNA	16666542	31138	0.00187	27.897	18.31661	0.000000
37% gDNA	13990980	19264	0.00138	20.560	13.49889	0.000000
26% gDNA	10760112	10907	0.00101	15.136	9.937756	0.000000
19% gDNA	10976478	8769	0.00080	11.929	7.832256	0.000000
10% gDNA	9408703	3415	0.00036	5.420	3.558449	0.000000
5.5% gDNA	9904155	2058	0.00021	3.103	2.037172	0.000000
5.0% gDNA	9083013	1987	0.00022	3.267	2.144706	0.000000
4.5% gDNA	8470920	1790	0.00021	3.155	2.071678	0.000000
4% gDNA	8852813	2336	0.00026	3.940	2.58697	0.000000

Table S6: Logistic Regression Coefficients and Thresholds for the Aneu+Mutations+Proteins.

	90th Percentile Values	Coefficients
Intercept		-11.85518
Aneuploidy	0.12	8.01470
Omega	1.15	2.34313
AFP	2866.68	0.00001
CA.125	6.90	0.08521
CA19.9	22.67	0.01966
CEA	2063.14	0.00037
HGF	264.64	0.00453
OPN	53651.18	0.00002
Prolactin	21304.07	0.00005
TIMP.1	85363.92	0.00001

Table S7: Correlation Matrix for 39 autosomal chromosome arms used in the manuscript. We found several chromosome arms are highly correlated. We calculated the correlation (Pearson Correlation Coefficient) for the fraction of reads that map to a particular chromosome arm compared to another chromosome arm.

	1p	1q	2p	2q	3p	3q	4p	4q	5p	5q	6p	6q	7p	7q	8p	8q	9p	9q	10p	10q	11p	11q	12p	12q	13q	14q	15q	16p	16q	17p	17q	18p	18q	19p	19q	20p	20q	21q	22q
1p	1.0	0.1	-0.4	-0.8	-0.3	-0.7	-0.8	-0.9	-0.9	-0.8	0.4	-0.9	-0.4	-0.5	-0.5	-0.9	-0.7	0.7	-0.4	-0.3	-0.3	0.6	0.1	0.5	-0.9	0.1	0.3	0.7	0.6	0.8	0.9	-0.5	-0.8	0.7	0.6	-0.3	0.7	-0.6	0.9
1q	0.1	1.0	0.3	0.1	-0.1	0.3	-0.1	0.0	-0.1	0.0	0.4	0.1	-0.2	-0.3	-0.2	0.0	-0.2	0.1	0.1	0.1	-0.1	0.1	0.2	-0.2	0.1	0.0	0.0	-0.3	-0.3	-0.1	0.0	-0.1	0.0	-0.1	-0.2	-0.2	-0.1	0.1	0.0
2p	-0.4	0.3	1.0	0.5	0.0	0.6	0.3	0.4	0.3	0.4	0.3	0.5	0.0	-0.1	0.2	0.4	0.1	-0.2	0.5	0.4	0.0	-0.3	0.2	-0.5	0.5	-0.1	0.0	-0.6	-0.5	-0.4	-0.4	0.2	0.4	-0.5	-0.6	0.1	-0.3	0.4	-0.4
2q	-0.8	0.1	0.5	1.0	0.2	0.8	0.8	0.9	0.8	0.7	-0.2	0.9	0.2	0.3	0.6	0.9	0.6	-0.5	0.6	0.6	0.3	-0.5	0.0	-0.7	0.9	-0.2	-0.1	-0.8	-0.7	-0.8	-0.9	0.3	0.9	-0.9	-0.9	0.3	-0.7	0.5	-0.8
3p	-0.3	-0.1	0.0	0.2	1.0	0.2	0.3	0.4	0.4	0.5	-0.4	0.3	0.3	0.4	0.2	0.4	0.4	-0.4	0.0	0.0	0.2	-0.4	-0.2	0.1	0.2	0.1	0.3	-0.2	-0.2	-0.4	-0.4	0.3	0.3	-0.2	-0.1	0.1	-0.4	-0.1	-0.4
3q	-0.7	0.3	0.6	0.8	0.2	1.0	0.6	0.8	0.7	0.7	0.0	0.8	0.2	0.1	0.3	0.7	0.5	-0.5	0.5	0.3	0.2	-0.5	0.1	-0.6	0.8	-0.1	-0.2	-0.8	-0.8	-0.7	-0.7	0.4	0.7	-0.7	-0.7	0.1	-0.7	0.6	-0.7
4p	-0.8	-0.1	0.3	0.8	0.3	0.6	1.0	0.8	0.9	0.7	-0.4	0.8	0.3	0.4	0.5	0.8	0.7	-0.6	0.4	0.4	0.4	-0.6	-0.2	-0.5	0.8	-0.2	-0.2	-0.6	-0.6	-0.7	-0.8	0.4	0.8	-0.7	-0.6	0.2	-0.6	0.5	-0.8
4q	-0.9	0.0	0.4	0.9	0.4	0.8	0.8	1.0	0.9	0.8	-0.4	0.9	0.4	0.4	0.6	0.9	0.8	-0.7	0.6	0.4	0.4	-0.7	-0.1	-0.5	0.9	-0.1	-0.3	-0.8	-0.8	-0.9	-1.0	0.4	0.9	-0.8	-0.8	0.3	-0.8	0.5	-0.9
5p	-0.9	-0.1	0.3	0.8	0.4	0.7	0.9	0.9	1.0	0.8	-0.5	0.9	0.4	0.5	0.6	0.9	0.7	-0.7	0.5	0.4	0.4	-0.7	-0.2	-0.5	0.9	-0.1	-0.3	-0.7	-0.7	-0.8	-0.9	0.5	0.9	-0.8	-0.6	0.3	-0.8	0.5	-0.9
5q	-0.8	0.0	0.4	0.7	0.5	0.7	0.7	0.8	0.8	1.0	-0.4	0.8	0.4	0.4	0.5	0.8	0.7	-0.7	0.4	0.2	0.4	-0.6	0.0	-0.4	0.8	-0.1	-0.3	-0.6	-0.6	-0.8	-0.8	0.5	0.7	-0.7	-0.6	0.2	-0.7	0.4	-0.8
6p	0.4	0.4	0.3	-0.2	-0.4	0.0	-0.4	-0.4	-0.5	-0.4	1.0	-0.3	-0.4	-0.7	-0.3	-0.4	-0.5	0.5	0.1	0.1	-0.3	0.5	0.4	-0.2	-0.3	0.0	0.4	0.0	0.1	0.3	0.4	-0.2	-0.3	0.1	-0.1	-0.2	0.3	0.1	0.4
6q	-0.9	0.1	0.5	0.9	0.3	0.8	0.8	0.9	0.9	0.8	-0.3	1.0	0.3	0.3	0.6	0.9	0.7	-0.6	0.6	0.5	0.3	-0.6	0.0	-0.6	0.9	-0.1	-0.2	-0.8	-0.8	-0.9	-0.9	0.4	0.9	-0.9	-0.8	0.3	-0.8	0.6	-0.9
7p	-0.4	-0.2	0.0	0.2	0.3	0.2	0.3	0.4	0.4	0.4	-0.4	0.3	1.0	0.4	0.3	0.3	0.4	-0.5	0.0	0.1	0.2	-0.4	-0.1	0.0	0.3	-0.1	-0.3	-0.2	-0.2	-0.3	-0.4	0.3	0.3	-0.2	0.0	0.1	-0.4	0.1	-0.4
7q	-0.5	-0.3	-0.1	0.3	0.4	0.1	0.4	0.4	0.5	0.4	-0.7	0.3	0.4	1.0	0.4	0.5	0.5	-0.5	0.0	0.0	0.2	-0.5	-0.4	0.1	0.4	-0.1	-0.4	-0.1	-0.2	-0.4	-0.5	0.2	0.4	-0.2	-0.1	0.1	-0.4	0.0	-0.5
8p	-0.5	-0.2	0.2	0.6	0.2	0.3	0.5	0.6	0.6	0.5	-0.3	0.6	0.3	0.4	1.0	0.7	0.5	-0.3	0.5	0.5	0.3	-0.4	-0.2	-0.3	0.6	-0.2	0.0	-0.4	-0.3	-0.7	-0.7	0.3	0.7	-0.7	-0.6	0.5	-0.4	0.2	-0.6
8q	-0.9	0.0	0.4	0.9	0.4	0.7	0.8	0.9	0.9	0.8	-0.4	0.9	0.3	0.5	0.7	1.0	0.7	-0.6	0.6	0.5	0.4	-0.6	-0.2	-0.6	0.9	-0.2	-0.2	-0.7	-0.6	-0.9	-0.9	0.4	0.9	-0.9	-0.8	0.4	-0.7	0.5	-0.9
9p	-0.7	-0.2	0.1	0.6	0.4	0.5	0.7	0.8	0.7	0.7	-0.5	0.7	0.4	0.5	0.5	0.7	1.0	-0.6	0.3	0.2	0.4	-0.6	-0.2	-0.3	0.7	0.0	-0.3	-0.5	-0.5	-0.7	-0.8	0.4	0.7	-0.6	-0.5	0.2	-0.6	0.3	-0.8
9q	0.7	0.1	-0.2	-0.5	-0.4	-0.5	-0.6	-0.7	-0.7	-0.7	0.5	-0.6	-0.5	-0.5	-0.3	-0.6	-0.6	1.0	-0.1	0.0	-0.3	0.6	0.0	0.1	-0.6	0.0	0.4	0.4	0.5	0.6	0.6	-0.3	0.5	0.3	0.2	-0.1	0.6	-0.3	0.7
10p	-0.4	0.1	0.5	0.6	0.0	0.5	0.4	0.6	0.5	0.4	0.1	0.6	0.0	0.0	0.5	0.6	0.3	-0.1	1.0	0.7	0.1	-0.3	0.0	-0.6	0.6	-0.2	0.1	-0.6	-0.4	-0.6	-0.5	0.3	0.6	-0.8	-0.7	0.4	-0.4	0.4	-0.5
10q	-0.3	0.1	0.4	0.6	0.0	0.3	0.4	0.4	0.4	0.2	0.1	0.5	-0.1	0.0	0.5	0.5	0.2	0.0	0.7	1.0	0.1	-0.2	-0.1	-0.5	0.5	-0.3	0.2	-0.5	-0.3	-0.5	-0.4	0.1	0.5	-0.7	-0.7	0.3	-0.2	0.2	-0.3
11p	-0.3	-0.1	0.0	0.3	0.2	0.2	0.4	0.4	0.4	0.4	-0.3	0.3	0.2	0.2	0.3	0.4	0.4	-0.3	0.1	0.1	1.0	-0.2	0.0	-0.2	0.4	-0.1	0.0	-0.2	-0.2	-0.3	-0.4	0.1	0.4	-0.3	-0.3	0.2	-0.4	0.2	-0.4
11q	0.6	0.1	-0.3	-0.5	-0.4	-0.5	-0.6	-0.7	-0.7	-0.6	0.5	-0.6	-0.4	-0.5	-0.4	-0.6	-0.6	0.6	-0.3	-0.2	-0.2	1.0	0.1	0.2	-0.6	0.0	0.3	0.4	0.5	0.6	0.7	-0.4	-0.6	0.5	0.4	-0.2	0.5	-0.3	0.7
12p	0.1	0.2	0.2	0.0	-0.2	0.1	-0.2	-0.1	-0.2	0.0	0.4	0.0	-0.1	-0.4	-0.2	-0.2	-0.2	0.0	0.0	-0.1	0.0	0.1	1.0	0.0	-0.1	0.1	0.1	-0.1	-0.1	0.1	0.1	0.0	-0.1	0.0	-0.1	-0.2	0.0	0.1	0.1
12q	0.5	-0.2	-0.5	-0.7	0.1	-0.6	-0.5	-0.5	-0.5	-0.4	-0.2	-0.6	0.0	0.1	-0.3	-0.6	-0.3	0.1	-0.6	-0.5	-0.2	0.2	0.0	1.0	-0.6	0.2	-0.2	0.7	0.5	0.5	0.5	-0.3	-0.6	0.7	0.7	-0.2	0.4	-0.6	0.4
13q	-0.9	0.1	0.5	0.9	0.2	0.8	0.8	0.9	0.9	0.8	-0.3	0.9	0.3	0.4	0.6	0.9	0.7	-0.6	0.6	0.5	0.4	-0.6	-0.1	-0.6	1.0	-0.2	-0.2	-0.8	-0.7	-0.9	-0.9	0.4	0.9	-0.9	-0.8	0.3	-0.8	0.6	-0.9
14q	0.1	0.0	-0.1	-0.2	0.1	-0.1	-0.2	-0.1	-0.1	-0.1	0.0	-0.1	-0.1	-0.1	-0.2	-0.2	0.0	0.0	-0.2	-0.3	-0.1	0.0	0.1	0.2	-0.2	1.0	0.0	0.1	0.0	0.2	0.1	0.0	-0.2	0.2	0.2	-0.2	0.0	-0.1	0.1

15q	0.3	0.0	0.0	-0.1	-0.3	-0.2	-0.2	-0.3	-0.3	-0.3	0.4	-0.2	-0.3	-0.4	0.0	-0.2	-0.3	0.4	0.1	0.2	0.0	0.3	0.1	-0.2	-0.2	0.0	1.0	0.0	0.2	0.2	0.2	-0.1	-0.1	-0.1	-0.2	0.1	0.3	-0.1	0.3
16p	0.7	-0.3	-0.6	-0.8	-0.2	-0.8	-0.6	-0.8	-0.7	-0.6	0.0	-0.8	-0.2	-0.1	-0.4	-0.7	-0.5	0.4	-0.6	-0.5	-0.2	0.4	-0.1	0.7	-0.8	0.1	0.0	1.0	0.7	0.7	0.7	-0.4	-0.7	0.8	0.8	-0.2	0.6	-0.6	0.7
16q	0.6	-0.3	-0.5	-0.7	-0.2	-0.8	-0.6	-0.8	-0.7	-0.6	0.1	-0.8	-0.2	-0.2	-0.3	-0.6	-0.5	0.5	-0.4	-0.3	-0.2	0.5	-0.1	0.5	-0.7	0.0	0.2	0.7	1.0	0.6	0.7	-0.3	-0.6	0.6	0.6	-0.1	0.6	-0.5	0.7
17p	0.8	-0.1	-0.4	-0.8	-0.4	-0.7	-0.7	-0.9	-0.8	-0.8	0.3	-0.9	-0.3	-0.4	-0.7	-0.9	-0.7	0.6	-0.6	-0.5	-0.3	0.6	0.1	0.5	-0.9	0.2	0.2	0.7	0.6	1.0	0.9	-0.4	-0.9	0.8	0.8	-0.4	0.7	-0.4	0.9
17q	0.9	0.0	-0.4	-0.9	-0.4	-0.7	-0.8	-1.0	-0.9	-0.8	0.4	-0.9	-0.4	-0.5	-0.7	-0.9	-0.8	0.6	-0.5	-0.4	-0.4	0.7	0.1	0.5	-0.9	0.1	0.2	0.7	0.7	0.9	1.0	-0.5	-0.9	0.8	0.7	-0.3	0.8	-0.5	0.9
18p	-0.5	-0.1	0.2	0.3	0.3	0.4	0.4	0.4	0.5	0.5	-0.2	0.4	0.3	0.2	0.3	0.4	0.4	-0.3	0.3	0.1	0.1	-0.4	0.0	-0.3	0.4	0.0	-0.1	-0.4	-0.3	-0.4	-0.5	1.0	0.4	-0.4	-0.3	0.2	-0.4	0.4	-0.5
18q	-0.8	0.0	0.4	0.9	0.3	0.7	0.8	0.9	0.9	0.7	-0.3	0.9	0.3	0.4	0.7	0.9	0.7	-0.5	0.6	0.5	0.4	-0.6	-0.1	-0.6	0.9	-0.2	-0.1	-0.7	-0.6	-0.9	-0.9	0.4	1.0	-0.9	-0.8	0.4	-0.7	0.5	-0.8
19p	0.7	-0.1	-0.5	-0.9	-0.2	-0.7	-0.7	-0.8	-0.8	-0.7	0.1	-0.9	-0.2	-0.2	-0.7	-0.9	-0.6	0.3	-0.8	-0.7	-0.3	0.5	0.0	0.7	-0.9	0.2	-0.1	0.8	0.6	0.8	0.8	-0.4	-0.9	1.0	1.0	-0.4	0.6	-0.5	0.7
19q	0.6	-0.2	-0.6	-0.9	-0.1	-0.7	-0.6	-0.8	-0.6	-0.6	-0.1	-0.8	0.0	-0.1	-0.6	-0.8	-0.5	0.2	-0.7	-0.7	-0.3	0.4	-0.1	0.7	-0.8	0.2	-0.2	0.8	0.6	0.8	0.7	-0.3	-0.8	1.0	1.0	-0.3	0.5	-0.5	0.6
20p	-0.3	-0.2	0.1	0.3	0.1	0.1	0.2	0.3	0.3	0.2	-0.2	0.3	0.1	0.1	0.5	0.4	0.2	-0.1	0.4	0.3	0.2	-0.2	-0.2	-0.2	0.3	-0.2	0.1	-0.2	-0.1	-0.4	-0.3	0.2	0.4	-0.4	-0.3	1.0	-0.2	0.1	-0.3
20q	0.7	-0.1	-0.3	-0.7	-0.4	-0.7	-0.6	-0.8	-0.8	-0.7	0.3	-0.8	-0.4	-0.4	-0.4	-0.7	-0.6	0.6	-0.4	-0.2	-0.4	0.5	0.0	0.4	-0.8	0.0	0.3	0.6	0.6	0.7	0.8	-0.4	-0.7	0.6	0.5	-0.2	1.0	-0.4	0.8
21q	-0.6	0.1	0.4	0.5	-0.1	0.6	0.5	0.5	0.5	0.4	0.1	0.6	0.1	0.0	0.2	0.5	0.3	-0.3	0.4	0.2	0.2	-0.3	0.1	-0.6	0.6	-0.1	-0.1	-0.6	-0.5	-0.4	-0.5	0.4	0.5	-0.5	-0.5	0.1	-0.4	1.0	-0.5
22q	0.9	0.0	-0.4	-0.8	-0.4	-0.7	-0.8	-0.9	-0.9	-0.8	0.4	-0.9	-0.4	-0.5	-0.6	-0.9	-0.8	0.7	-0.5	-0.3	-0.4	0.7	0.1	0.4	-0.9	0.1	0.3	0.7	0.7	0.9	0.9	-0.5	-0.8	0.7	0.6	-0.3	0.8	-0.5	1.0

Table S8: List of the Well Barcode Indices Used in the second round of PCR for RealSeqS.

Corresponding Well	P5 index	P7 index
A1_1	CGTGCAGG	ACAAGTAT
A2_2	GATCGGCG	CGATGTAT
A3_3	ATCTCAGT	TTAGGCAT
A4_4	TTAGCAGT	TGACCAAT
A5_5	CGTATGAC	ACAGTGAT
A6_6	TGTATATT	GCCAATAT
A7_7	ATGTCAAC	CAGATCAT
A8_8	CCGAATAG	ACTTGAAT
A9_9	ATGCGGCA	GATCAGAT
A10_10	AGGTAGCC	TAGCTTAT
A11_11	ACGCTATT	GGCTACAT
A12_12	TTGAAGGC	CTTGTAAT
B1_13	TGGCGCGA	GAACCGAT
B2_14	TAGCATGG	CGACAAAT
B3_15	CGAATCGT	TACGAAAT
B4_16	CTAGCGGC	TCGGACAT
B5_17	GTTCTAAT	TTCCATAT
B6_18	AATTCATT	AGTGCAAT
B7_19	GAAGGTGA	CCGTCGAT
B8_20	ACATACTC	CAGTCCAT
B9_21	ACCGAACT	AGTTAGAT
B10_22	GACTAAGA	GACTTGAT
B11_23	TTACCTAC	GCCATAAT
B12_24	TAAGACTG	CGAAGCAT
C1_25	CGACCTCG	TATAAGAT
C2_26	CGATGGCA	ATCAGCAT
C3_27	AACACGTA	TTCTGTAT

C4_28	TGAGTAAG	CTAGTTAT
C5_29	AGCCACGC	ATACCTAT
C6_30	AGTCATGA	GACCATAT
C7_31	CTACAAGC	ATTCATAT
C8_32	TCAATAAC	TGATAAAT
C9_33	GGCAACAC	CTGTATAT
C10_34	GCGCCTAA	ATTGAAAT
C11_35	CAGGAATC	GTCTGCAT
C12_36	GTAGAATG	ACTAATAT
D1_37	AAGCCTGT	CTACCGAT
D2_38	CACATTAT	AGGTCAAT
D3_39	TTGTGGCT	ATGTGAAT
D4_40	AGGCGGTT	TGCGACAT
D5_41	TACGAGCG	AGAACGAT
D6_42	GTCATACA	TTCACAAT
D7_43	CCATTGAC	CGTACTAT
D8_44	GCAGTTAA	CCAGCGAT
D9_45	ATGGCGCT	TCTTCAAT
D10_46	CAGATAGT	CCGGAAAT
D11_47	TAGCCACC	GCTCAAAT
D12_48	ACGGACGG	GCGCAGAT
E1_49	AACGTCGA	CGCTTGAT
E2_50	GACAGTCG	GCCTGGAT
E3_51	GCCGGTTG	ATTGCTAT
E4_52	CTGTCATT	AATTGCAT
E5_53	CATTCCAG	ATTCGAAT
E6_54	GCTTCTCA	GATTCAAT
E7_55	AGTCTACC	TTCTACAT
E8_56	TGCTAACT	CTGAGGAT
E9_57	TCCGCGAT	TGTGATAT

E10_58	CTAGTCCG	TCATGTAT
E11_59	AATGCGAA	CCACTTAT
E12_60	GTGAGTCA	GTGGTGAT
F1_61	AGCGTGCG	CGTCTCAT
F2_62	CTCGAACA	AGAGGTAT
F3_63	GCAGTGGT	CTCCGCAT
F4_64	GTGTTAAG	TACTCGAT
F5_65	GCTGTAGC	AGGTGTAT
F6_66	ATACTTAG	TCTCGCAT
F7_67	CCGCGTAC	GGTAATAT
F8_68	GTAATTCC	GTCGAGAT
F9_69	TTATCGTC	AAGCGCAT
F10_70	GGTCGTAT	GGCCGTAT
F11_71	TCAAGGTT	TCCAGAAT
F12_72	ATTGCCAG	CTTAGCAT
G1_73	AACTTGAA	TTAGCGAT
G2_74	CACACATG	AGTTCCAT
G3_75	AGATGACT	CCTGACAT
G4_76	AATTATTG	GGCAGGAT
G5_77	ACGCATTA	TGACACAT
G6_78	CCGCTGTA	CTTCACAT
G7_79	CTCTCGAA	AACTCTAT
G8_80	CGGACATC	TACAGTAT
G9_81	TTGACCGA	CAAGGAAT
G10_82	TGCTTCTT	GTACTTAT
G11_83	TTGTGTAC	TAGAGCAT
G12_84	CAGCTAAC	GTCTCAAT
H1_85	TGGAAGAT	GACTAAAT
H2_86	TATAGGTG	CCGCTCAT
H3_87	TCCGTATG	TGCGGAAT

H4_88	CTTCTAGA	GCGTATAT
H5_89	TCGTGTCA	GCGGCTAT
H6_90	GTAGCACA	TCCTCCAT
H7_91	GGCGTGTC	ACCAGGAT
H8_92	ACCATAAG	GACACGAT
H9_93	ACTTGTCC	AACATAAT
H10_94	AGAACACA	GTTGGCAT
H11_95	GACGATGT	CATGAGAT
H12_96	AATATCAA	GCTGTCAT
A1_101	CGTGCAGG	CTATTAAT
A2_102	GATCGGCG	CGAGAGAT
A3_103	ATCTCAGT	GGATCGAT
A4_104	TTAGCAGT	GATCCTAT
A5_105	CGTATGAC	TATTGGAT
A6_106	TGTATATT	GAGCCAAT
A7_107	ATGTCAAC	CAGGACAT
A8_108	CCGAATAG	CGAGCCAT
A9_109	ATGCGGCA	ACACGAAT
A10_110	AGGTAGCC	TCCACGAT
A11_111	ACGCTATT	GAAGATAT
A12_112	TTGAAGGC	GGTACCAT
B1_113	TGGCGCGA	TGGCATAT
B2_114	TAGCATGG	CTATGCAT
B3_115	CGAATCGT	TAATGAAT
B4_116	CTAGCGGC	TAAGTCAT
B5_117	GTTCTAAT	AAGCATAT
B6_118	AATTCATT	TCAAGCAT
B7_119	GAAGGTGA	ATCCGGAT
B8_120	ACATACTC	AACCTTAT
B9_121	ACCGAACT	TGCATAAT

B10_122	GACTAAGA	ACGAACAT
B11_123	TTACCTAC	CATACGAT
B12_124	TAAGACTG	CTGTTCAT
C1_125	CGACCTCG	AATCGGAT
C2_126	CGATGGCA	ACCGCTAT
C3_127	AACACGTA	TATTCTAT
C4_128	TGAGTAAG	GGATTCAT
C5_129	AGCCACGC	AGTCAAAT
C6_130	AGTCATGA	ATCCACAT
C7_131	CTACAAGC	CGAATTAT
C8_132	TCAATAAC	GTAAGAAT
C9_133	GGCAACAC	CTGGAGAT
C10_134	GCGCCTAA	GAGATAAT
C11_135	CAGGAATC	CACATTAT
C12_136	GTAGAATG	GGACTAAT
D1_137	AAGCCTGT	TCTCTGAT
D2_138	CACATTAT	GGAGGCAT
D3_139	TTGTGGCT	CGTCAGAT
D4_140	AGGCGGTT	GTACGCAT
D5_141	TACGAGCG	ACGCTAAT
D6_142	GTCATACA	CTCGTCAT
D7_143	CCATTGAC	TTACGTAT
D8_144	GCAGTTAA	GAGAAGAT
D9_145	ATGGCGCT	TTGAATAT
D10_146	CAGATAGT	CAGCCTAT
D11_147	TAGCCACC	CGTGAAAT
D12_148	ACGGACGG	AAGGCGAT
E1_149	AACGTCGA	CATAACAT
E2_150	GACAGTCG	CAAGCTAT
E3_151	GCCGGTTG	ACCTGCAT

E4_152	CTGTCATT	CACCTAAT
E5_153	CATTCCAG	AGATGCAT
E6_154	GCTTCTCA	AGGCTTAT
E7_155	AGTCTACC	AAGAGGAT
E8_156	TGCTAACT	TTAGTAAT
E9_157	TCCGCGAT	GCACATAT
E10_158	CTAGTCCG	CCATGAAT
E11_159	AATGCGAA	TCGTGGAT
E12_160	GTGAGTCA	CCTGCTAT
F1_161	AGCGTGCG	GCATAAAT
F2_162	CTCGAACA	AGTAGGAT
F3_163	GCAGTGGT	TATGGTAT
F4_164	GTGTTAAG	GTTACAAT
F5_165	GCTGTAGC	GAGTGAAT
F6_166	ATACTTAG	ACGTAAAT
F7_167	CCGCGTAC	CAACGGAT
F8_168	GTAATTCC	GAGTCGAT
F9_169	TTATCGTC	AACTAGAT
F10_170	GGTCGTAT	CCTCGGAT
F11_171	TCAAGGTT	ACAGCCAT
F12_172	ATTGCCAG	TACGGCAT
G1_173	AACTTGAA	TGTGTGAT
G2_174	CACACATG	GACCGAAT
G3_175	AGATGACT	TTCGTTAT
G4_176	AATTATTG	CATTAAAT
G5_177	ACGCATTA	AAGTTGAT
G6_178	CCGCTGTA	AATGTAAT
G7_179	CTCTCGAA	TCAATGAT
G8_180	CGGACATC	AGGACTAT
G9_181	TTGACCGA	GCTTCTAT

G10_182	TGCTTCTT	CGTTGCAT
G11_183	TTGTGTAC	GAGCACAT
G12_184	CAGCTAAC	ACTAGCAT
H1_185	TGGAAGAT	AGACGGAT
H2_186	TATAGGTG	AACAATAT
H3_187	TCCGTATG	GGTGGTAT
H4_188	CTTCTAGA	AGCCTAAT
H5_189	TCGTGTCA	CCGACTAT
H6_190	GTAGCACA	GGCGTTAT
H7_191	GGCGTGTC	GAATACAT
H8_192	ACCATAAG	CAGTAGAT
H9_193	ACTTGTCC	ATGCCGAT
H10_194	AGAACACA	GGCCAGAT
H11_195	GACGATGT	CCTTATAT
H12_196	AATATCAA	ATGGCCAT

Schematic of final sequencing library (5’ to 3’)

Universal primer sequence:

Binding site for the second round PCR amplification and universal sequencing primers

Illumina Grafting Sequence:

Facilitates hybridization to the sequencing instrument**First Round of PCR:**

Forward Primer for amplifying ~350,000 regions; the 6-mers used for design are highlighted:

cgacgtaaaacgacggccagtNNNNNNNNNNNNNNNN GGTGAAACCCCGTC***TCTACA***

Reverse Primer for amplifying ~350,000 regions; the 4-mers used for design are highlighted:

cacacaggaaacagctatgaccatgCCTCCTAAGTAGCTGGGACT*ACAG*

Example amplicon containing a 37 bp sequence amplified by the forward and reverse primers (black font)

cgacgtaaaacgacggccagtNNNNNNNNNNNNNNNN GGTGAAACCCCGTC**TCTACA**
AAAAATACAAAAATTAGCTGGCCGTGGTGGCGCATGC CTGTAGTCCCAGCTACTTAGGAGG
catggtcatagctgtttcctgtgtg

Second Round of PCR:

Forward, with 8 bp sample bar code in red font:

AATGATACGGCGACCACCGAGATCTACACCGTGCAGGcgacgtaaaacgacggccagt

Reverse, with 8 bp sample bar code in red font:

CAAGCAGAAGACGGCATACGAGATACAAGTATcacacaggaaacagctatgaccatg

A full list of indices used in the manuscript is included in SI Appendix Table S8.

Fig. S1: Schematic of final sequencing library (5’ to 3’)

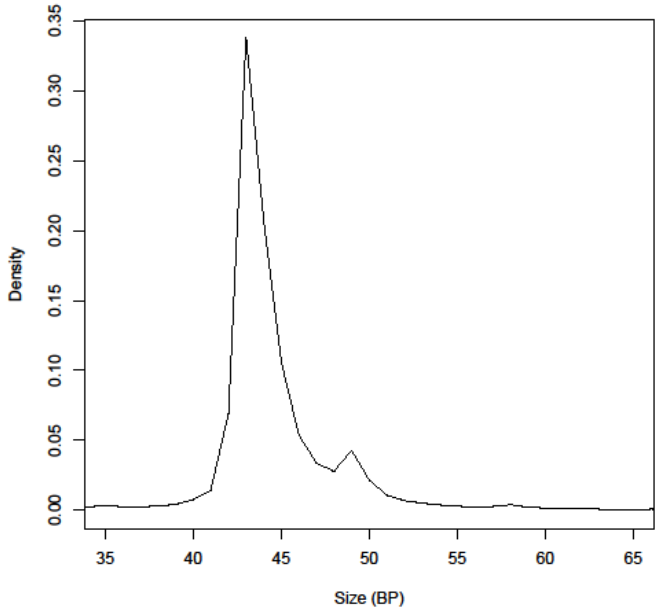

Fig. S2. Distribution of amplicon sizes obtained by ReqlSeqS

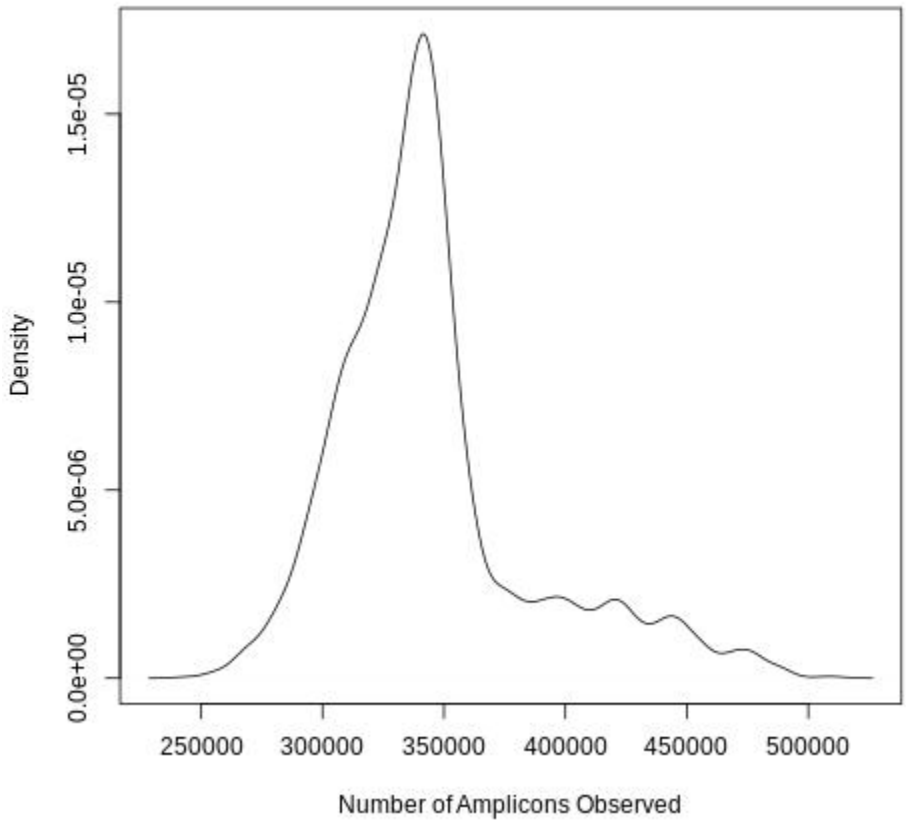

Fig. S3. Distribution of the number of amplicons observed in cell free DNA from 2231 plasma samples.

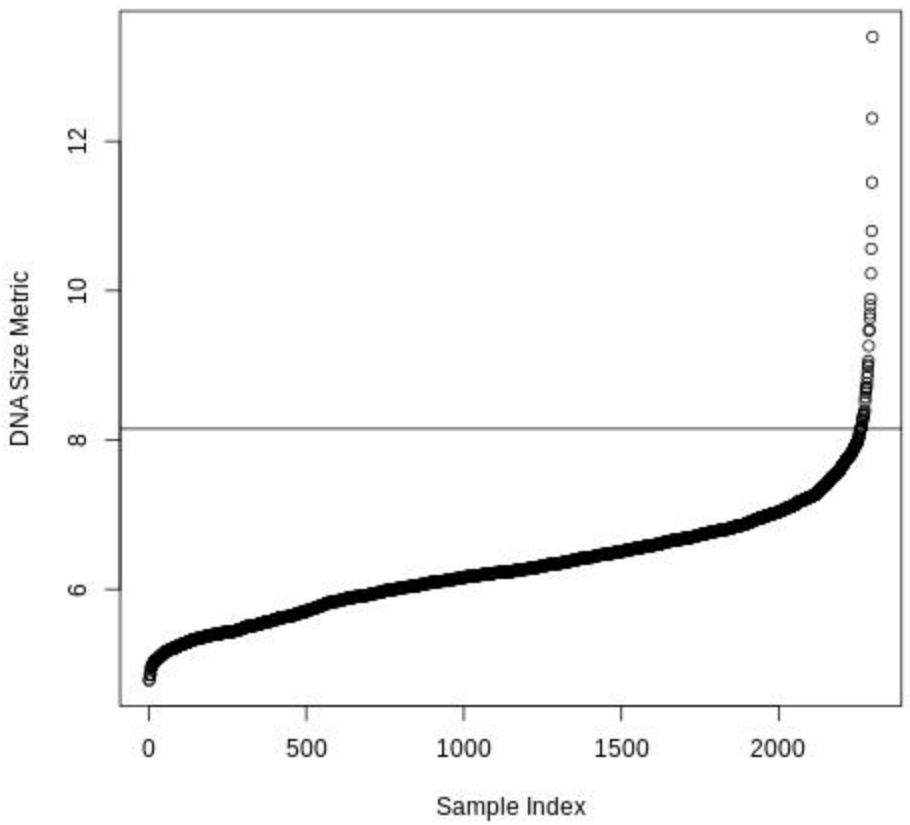

Fig. S4. Distribution of the DNA Size Metric used by RealSeqS to identify outlier samples. Samples with a metric greater than 8.15%, indicated by the horizontal line, were excluded.

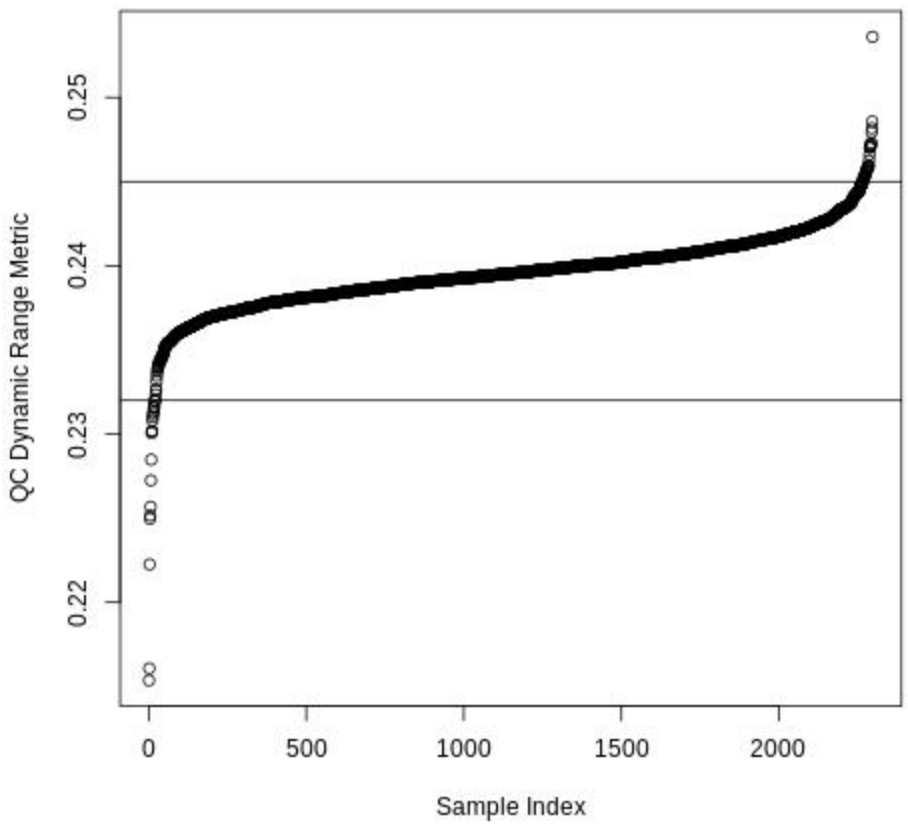

Fig. S5. The distribution of the QC Dynamic Range metric used by RealSeqS to identify outlier samples. Samples with values < 0.2320 or >0.2450, indicated by the horizontal black lines, were excluded.

***In Silico* simulated samples with monosomy or trisomy of one chromosomal arm**

```
N <- number of normal samples used
f <- desired neoplastic cell fraction
r <- chromosome arm (1p..22q)
r_type <- alteration type (gain or loss)
rho <- desired unique read depth of synthetic sample
j<-desired number of repeats
For j=1:5
        For i=1:N
                s <- sample[i]
                s_fraction<-vector that contains the fraction of reads that map to each of the 39 chr arm
                        for sample s
                For f in 1:F
                        For r in 1p:22q
                                For r_type in gain loss
                                        h <- new in silico sample
                                        t <- copy of s
                                        t_fraction<- vector that contains the fraction of reads that map
                                                to each of the 39 chr arm for sample t
                                        if r_type== gain
                                                t_fraction[r]<-t_fraction[r]*3/2 ###trisomy on arm r
                                        else (r_type == loss)
                                                t_fraction[r]<-t_fraction[r]*1/2###monosomy on arm r
                                        Endif
                                        h_normal<-weighted random select (1-f)*rho where the
                                                weights are r_fraction
                                        h_aneuploid<-weighted random select f*rho where the weights
```

```
                                        are  t_fraction
                                    h<-h_normal+h_aneuploid
                                End
                            End
                        End
                    End
End
```

Fig. S6. Pseudocode to generate the *in silico* trisomy and monosomy samples used for the comparison of whole genome sequencing, FAST-SeqS, and ReqlSeqS.

***In Silico* simulated samples with alterations of many chromosome arms**

N <- number of normal plasma samples used

d <- desired degree of aneuploidy (number arms altered)

f <- desired neoplastic cell fraction

r <- chromosome arm (1p..22q)

r_type<-arm alteration type (gain or loss)

p(r) <- probability that an arm is gained or lost in cancer (Estimated from (11))

rho <- desired unique read depth of *in silico* simulated sample (10M default)

j<- desired number of repeats

```
For j=1:5
    For i=1:N
        s <- sample[i]
        U <- get reads from s
        For f in 1:F
            h <- new in silico sample
            For d in (10,15,20) #desired numbers of altered arms
                t <- copy of s
                For a in 1:d #select alteration types and spike-in to t
                    r, r_type <- weighted random selection where the weights [p(r)]
                    are the likelihood that an arm is gained or loss in cancer
                    u_all <- get all Reads that that map to chr arm r
                    if r_type== gain
                        t <- add 50% of u_all reads
                    else (r_type == loss)
                        t <- subtract 50% of u_all reads
                    Endif
                    h_normal<-randomselect (1-f)*rho reads from s
                    h_aneuploid<-randomselect f*rho reads from t
                    h<-h_normal+h_aneuploid
```

End

End

End

End

End

Fig. S7. Pseudocode to Generate *in silico* simulated samples with multiple arm alterations that were used in the Genome Wide Aneuploidy SVM training set.