**SUPPLEMENTARY METHODS**

**High-intensity sequencing reveals the sources of plasma circulating cell-free DNA variants**

Razavi et al.

***Library preparation, target enrichment, and sequencing***

Buffy coat gDNA (50ng) and plasma cfDNA (≤75ng) were used for NGS library construction with a modified Illumina TruSeq DNA Nano protocol. The adapter includes 96 (cancer samples) or 218 (healthy controls) unique molecular identifier (UMI) sequences, of 6-mer in length, used to suppress technical assay errors. Amplified libraries were cleaned up using magnetic beads and quantified using the Fragment Analyzer Standard Sensitivity NGS kit.

Quantified libraries underwent hybridization-based capture with a research cancer panel targeting 508 cancer-related genes (2.13 Mb; GRAIL, Inc.; Menlo Park, CA). The panel included full exons except for the telomerase reverse transcriptase (*TERT*) gene, which only included promoter regions. Additional intronic regions were included for rearrangement detection of 28 genes and copy number aberration detection of 42 genes. Up to 2ug of gDNA libraries were used for target enrichment with a modified Illumina Nextera Rapid Capture protocol. For cfDNA samples, up to 4ug of libraries (two parallel enrichment reactions) were used to maximize capture efficiencies. The two enriched libraries from the same cfDNA library were pooled and quantified using a Fragment Analyzer Standard Sensitivity NGS kit.

Three libraries per flowcell or six libraries across two flowcells were clustered (pooled and loaded across all eight lanes of each flowcell) and sequenced on a HiSeq X (Illumina; San Diego, CA) at a nominal raw target coverage of 60,000X (median collapsed target coverage [unique molecule counts] was ~4,400X). Read lengths were set to 150, 150, 8, and 8, respectively for read 1, read 2, index read 1, and index read 2.

**cfDNA analysis pipeline**

### **Preprocessing and alignment**

Raw bcl files from the sequencer were demultiplexed and converted to fastq format using bcl2fastq version 2.18.0.6 (Illumina). Settings in the sample sheets were configured so that bcl2fastq trimmed the first 7 bp at the 5’ end of the reads containing the 6 bp UMI along with the ‘T’ base, following which, the UMI sequences were added to the read name. The fastqs were further processed to trim 3’ overhangs based on pairwise alignment of the paired-end reads. The reads were then aligned using BWA MEM (*56*) to a version of the human reference genome created by combining the UCSC hg19 assembly with decoy sequences.

### **Error correction and read pair stitching**

Error correction was performed by first identifying reads originating from the same source DNA fragment and building a consensus sequence from those reads. Aligned reads were bagged using mapping coordinates (5’ end of the paired-end reads) and UMIs added during library construction step. Each read pair within a bag represented a copy of either the reference or non-reference strand of the initial DNA fragment. A read-pair was derived from the reference (or the non-reference) strand of the initial fragment if the read pair was mapped in F1R2 (or F2R1) orientations. In bagging the reads, a small amount of *slack* (1 bp) was allowed in the position of the 5’ end of the reads, suggesting that read pairs in which the 5’ end of one of the two reads is off by 1 bp could still go into the same bag, provided that they shared the same UMI pair with the other read pairs. Additionally, up to two mismatches were allowed within the combined UMI length (12 bp) to account for sequencing errors. Bagging was performed in a greedy fashion by picking a founder pair and adding all read pairs that satisfied the 1 bp slack and UMI error requirements with the founder pair.

Reads within each bag were error-corrected by building a reference-free consensus of each set of reads in the bag. A proto-consensus was initialized by copying one of the reads in the set. This proto-consensus was adjusted by adding the rest of the reads in the bag one at a time and updating the proto-consensus to the pairwise alignment between the proto-consensus and the newly added read. At the end of this process, the proto-consensus pair was converted to a collapsed error-corrected pair by sliding through the proto-consensus one position at a time and selecting a consensus base and nominal quality score for that position based on the sum of quality scores for each base at that position. A collapsed read was called duplex-collapsed if both strands of the initial molecule were represented in the bag. The read was called non-duplex collapsed if only one of the two strands was represented in the bag. Duplex-collapsed reads had lower error rates because of a better chance of correcting PCR and sequencing errors when both strands were represented in the bag. Bags with a single read pair were omitted from the output of the error correction step, as error correction was not possible on those reads.

After the initial error correction, collapsed reads were re-aligned to the reference genome. Collapsed read pairs with sufficient implied overlap were stitched together. This step builds a single contiguous sequence for the entire fragment, which provides better context for downstream assembly. In addition, this allows for further error correction when there is a discrepancy between the base calls from the two paired end reads within the overlap between the two reads. The paired-end collapsed reads that did not have sufficient overlap were unaffected by stitching.

### **Candidate variant generation by de novo assembly**

A *de novo* assembly was constructed for each contiguous target region based on the error corrected read pairs. The assembly provided a compact, reference-free representation of all the haplotypes observed in the reads. For each target region, an optimal *k*-mer size *k* ∈ (48, 64, 96) was dynamically selected based on the largest perfect repeat 𝜆 within that region. The selected *k* was the smallest *k* such that *k* ≥ 𝜆. Once *k* was selected, a De Bruijn graph was assembled based on *k*-mers from all reads overlapping the target region. The error-corrected reads were first localized to target regions by alignment to the reference genome sequence. The reads in each target region were assembled into a de Bruijn graph, the reads aligned to the graph, and a read support vector was populated for each *k*-mer in the graph based on the lowest quality base in each occurrence of that *k*-mer in an error-corrected read. The graph was iteratively pruned by applying heuristics to remove edges with weak support. Variant calling was performed by aligning the edges of the pruned graph to a reference and listing genomic coordinates where an edge was different from the reference. By extending the edge on either side with the edges with the highest support upstream and downstream, it was possible to provide a longer context for the alignment than what was possible based on a single read, which helped avoid alignment-related artifacts in variant calling.

***Reproducibility of the targeted DNA assay***

Two similar targeted DNA assay protocols (V1 and V2) for plasma cfDNA and matching WBC gDNA samples were employed in this study. The main differences were the UMI sequences in the library adapters and the reaction volumes in hybridization enrichment process, neither of which would be expected to influence results. To ensure assay performance equivalence, six patient samples with a large cfDNA yield were selected (**Supplementary Table 6**), allowing for reprocessing with both assay protocols, as well as droplet digital PCR (ddPCR). Bio-Rad ddPCR was used to measure canonical hotspot variants in five of the six patient samples, revealing good agreement with measurements using the targeted DNA assays (**Fig. 1b**). **Fig. 1c** and **Supplementary Fig. 3** show the measured allele fraction of variants called either using the V1 or V2 targeted DNA assay protocol. Measured VAFs between the two technical replicates for samples from five patients (**Fig. 1c**) showed a strong agreement (*R2* = 0.9997) as well as for one sample from a hypermutated case (**Supplementary Fig. 3,** *R2* = 0.9972).

### **Machine learning error model**

Quality scores were assigned to all called variants using a pair of probabilistic error models. For SNVs and MNVs, a hierarchical Bayesian model with one set of noise parameters per genomic position in the panel and alternative allele was trained using baseline control samples from the technical validation study. Candidate variants generated from the De Bruijn graph were scored relative to the collapsed pileup information without base alignment quality. At each genomic coordinate, , the expected alternate allele depth has a baseline noise rate where is the total depth at and is the allele frequency. The model assumes that is drawn from a Poisson distribution such that:

Each was assumed to be drawn from a mixture of Gamma distributions indexed by , the parameters of which varied based on discrete covariates such that:

The covariates included were: (1) whether was uniquely alignable, (2) whether was located within a known segmental duplication and (3) the trinucleotide context. These covariates were found to be indispensable given that the error rate can vary as a function of the coordinate , the type of mutation and its mutational context. **Supplementary Figs. RR2a** and **RR2b** below show the distribution of the mean posterior estimates of () across a representative subset of sites by type of substitution and trinucleotide context.

Typical workflows for detecting variants from single tumor samples and/or matched tumor normal pairs seldom include site-specific priors and if they do, it is usually through ad hoc filters for putative mutations detected in panels of normal samples (PMID: 23396013). Here instead, the posterior distribution of obtained through Markov chain Monte-Carlo resampling provides site-specific error rates with non-zero estimates at sites where no actual alternate allele counts were observed across samples in the training set. **Supplementary Fig. RR2c** of this response shows the scatter plot of site-specific mean posterior estimates against the observed for samples in the training set.

The posterior distribution of each was approximated to a Gamma distribution and translated to a negative Binomial predictive posterior for the observed number of alternate allele counts at each and each alternate allele such that:

where and are the rate and dispersion parameters. This provides the basis for recalibrated quality scores for each observed alternate allele at each such that:

where is the Phred-scaled quality score and is the alternate allele depth. **Supplementary Fig. RR2d** below compares the estimated to empirical probability of observing an event. The recalibrated quality scores provide an objective measure to define thresholds for calling variants. Using the mean and variance estimates of obtained by training the model on 43 healthy control individuals, we compute the site- and allele-specific quality scores comparing the variants thus detected against the matched tumor biopsy. **Supplementary Fig. RR2e** shows the recall rate per cancer type against the mean number of variants detected in healthy control individuals at different thresholds. By definition, scores are expected to allow one false positive per million bases and retained almost all biopsy-matched variants.

The hierarchical Bayesian model described above was applied to the raw variants scored both in cfDNA and WBC where the observed alternate allele depths, were assumed to correspond to the sum of negative Binomial noise and Poisson distributed signal. Formally:

with an equivalent model for WBC. The likelihood of the observed pair of and was computed conditional on the cfDNA and WBC total depths, noise parameters and alternate allele fractions. Combining the joint likelihood with a uniform prior on the alternate allele fractions allows one to assign the source of origin of the variant such that:

where describes the posterior probability of assigning a detected variant to cfDNA. The procedure used for choosing is provided in the supplementary methods under section “Joint variant analysis using the machine learning error model”. The posterior probability allows the model to be further optimized. **Supplementary Fig. RR3** illustrates the cancer specific trade-offs between the recall rate of biopsy-matched variants and allowing additional SNVs to be assigned to cfDNA in healthy control individuals at fixed . The particular thresholds were obtained through cross-validation by holding out each cancer type and selecting the threshold retaining most biopsy-matched variants whilst filtering out most variants of potential hematopoietic origin.

In summary, candidate variants in cfDNA and WBC generated through *de novo* assembly using the de Bruijn graph are assigned a quality score based on site-specific error rates and a posterior probability of originating from cfDNA based on the joint likelihood of the alternate allele depth in WBC. Using cross-validated cancer specific thresholds on the quality score and the posterior probability of assignment to cfDNA, the number of putative somatic variants is reduced to <2 non-synonymous variants per healthy control individual. **Supplementary Tables RR3-6** summarize the final number of variants whereby driver variants are defined as cancer-specific hotspot mutations and mutations considered to be pathogenic or likely pathogenic (PMID: 28890946).

Several groups have explored the space of high accuracy error-corrected sequencing. Integrated Digital Error Suppression (iDES) (PMID: 27018799) computes a per-base error rate which optimally balances error suppression with molecular depth as 2x10-5. The Safe-Sequencing System (Safe-SeqS) (PMID: 21586637) reported 0.9x10-5 supermutants (likely errors) per base pair. Both reports are similar to the per base error rate of 1x10-5 to 3x10-5 in our study. Additionally, Lanman *et al.* (PMID: 26474073) and Phallen *et al.* (PMID: 28814544) both point out that beyond per base error rates, effective filtering for false positives is necessary; both report stringent filtering on small, highly curated panels to produce no false positive mutation calls in 1.56x106 bases (20 samples with a panel size of ~78Kb) and fewer than 1 false positive mutation call per 3x106 bases (38 samples with a panel size of ~80kb) attempted, respectively, in a population of relatively young individuals. Our study included a panel that was more than ten-fold larger than the reports noted above, with 1.5 variant calls per age-matched control sample in approximately 1x106 bases per experiment.

For indels, a similar Bayesian hierarchical model was employed (**Supplementary Fig. 10**). Each anchor position p and indel length l (positive for insertions, negative for deletions) in the panel was assumed to have a baseline noise rate dp 𝜆pl. Additionally, 𝜆pl = 𝜆pωlp. 𝜆p was assumed to be drawn from a Gamma distribution with parameters unique to the combination of proximity to segmental duplications, lack of unique alignability, and repeats as identified by RepeatMasker. 𝛚p (a probability vector) was assumed to be drawn from a Dirichlet distribution with parameters varied based on the same factors. Training was done via a similar Metropolis-Hastings within Gibbs algorithm, with analogous numbers of iterations and convergence diagnostics. The same dispersion approximation and re-estimation was applied to each 𝜆pl, using the marginal posterior obtained from the samples. Quality scores for observed indels were computed as described above using these approximate posterior predictive distributions. The estimation processes described above were carried out separately for cfDNA and WBC samples, resulting in four models in total (SNV and indel models for each of cfDNA and WBC).

### **Edge artifact scoring**

DNA damage processes similar to those observed in ancient DNA samples generated artifacts in samples used in this study. These “edge artifacts” were frequently near the ends of fragments. This error source was controlled by using a discrete mixture model. Specifically, for each putative artifact, the following metrics were considered: (1) the median distance of the putative artifact from the fragment end, (2) the Phred-scaled p-value (from the noise model) of the putative artifact, (3) the concentration towards fragment ends based on a Wilcoxon rank-sum test, and (4) the observed putative artifact allele fraction. Each metric was separately discretized to a resolution of 5 for the median distance and Phred-scaled p-value, and to a resolution of 0.01 for allele fractions. The model assumed that the observed distribution of these features for “at-risk” G>A and C>T variants consisted of a mixture of edge artifact and null (non-artifact) distributions. The proportion of SNVs originating from each mixture component was assumed to vary across samples. For training, SNVs, except for G>A and C>T, were considered as negative controls, assuming they come from the null distribution. Similarly, for training the model, variants considered at risk for artifactual origin were those with a median edge distance below 25 bp and a median absolute deviation in edge distance below 25 bp. Additionally, it was assumed that the discrete mixture distribution for the artifact component factorized across features. The null mixture component was estimated directly from the observed joint distribution of the given features in the negative control SNVs. The alternative mixture component was estimated via maximum likelihood, holding the null distribution estimate fixed. The model was trained using SNVs from held-out healthy controls and samples from the MSK site not used in this study.

The resulting mixture model was applied to each sample. On a per-sample basis, the proportion of at-risk candidate variants (G>A and C>T SNPs passing error model) was estimated via maximum likelihood, fixing the estimates of mixture components. These estimated proportions varied in training samples, from less than 1% to over 98%. This estimate was then applied to compute, for each such variant, the posterior probability that it originated from the edge artifact mixture component. SNVs with posterior probability greater than 1% were flagged as potential edge artifacts and removed from further analysis.

### **Joint variant analysis using the machine learning error model**

The noise models described above were used to assess the support for non-WBC variant origins. Starting from the approximate posterior predictive distributions estimated above, the observed alternative depth for each putative variant was assumed to correspond to the sum of negative Binomial noise, with parameters based on the noise model, and Poisson-distributed signal. Formally, assume ADcfDNA,p~ Pois(AFcfDNA,pdp) + NB(mpdp, rp) for cfDNA, with an analogous model for the WBC observation. From this assumption, the likelihood of the observed pair of cfDNA and WBC alternative depths was computed, conditional on the cfDNA and WBC total depths, noise parameters, and hypothesized alternative allele fractions. This provided a joint likelihood function parameterized by the pair of cfDNA and WBC alternative allele fractions (parameters). Combining this likelihood with a hypothesized joint uniform prior on the alternative allele fractions, the posterior probability P(AFcfDNA,p>= k \* AFWBC,p) was computed. Joint calling amounted to setting a threshold on this probability, and selecting a functional form for k.

The threshold on this posterior probability and functional form for k were jointly determined from control data and structured cross-validation across tumor types. The functional form of k() was restricted to a piecewise linear function of the observed WBC allele fraction. For allele fractions greater than 1/3, k was set to 3 to address the appearance of germline heterozygous variants within regions of altered copy number. A range of values for k(0) from 1 to 3 was evaluated, setting the slope as required to maintain continuity of the resulting function. Next, control of false positives was required, defined as minimizing deviation from target false positive range of 1 variant per sample in held-out healthy control samples with a binomial 95% upper confidence bound of at most 3 variants per sample. Within the set of k(0) and probability threshold combinations admitted by these criteria, cross-validation across tumor types was employed to select final parameters. For each tumor type, a training set of samples was created consisting of all other tumor types (e.g. the training set for breast cancer samples consisted of NSCLC and CRPC samples). The combination of k(0) and probability threshold that maximized sensitivity in the training set from the admissible set defined above was then selected. The selected parameters were then employed on the test set, consisting of the original samples (e.g. breast cancer samples in the preceding example).

***Estimation of VAF***

Modeled VAFs were obtained as the mean posterior estimate from the Bayesian hierarchical model after Markov chain Monte-Carlo resampling. VAFs were estimated from the raw pileup without BAQ filtering and with or without pseudocounts such that and where and are the alternate allele and total base counts, respectively (**Supplementary Figure RR5**). Whilst being more factual, raw VAFs without pseudocounts cannot be displayed accurately since variants detected in cfDNA with zero alternate allele count in WBC cannot be represented in a typical scatterplot with logarithmic axes. Similarly, whilst the raw VAFs with added pseudocounts is a possible alternative, the equivalent depiction using posterior estimates of VAF does not affect the interpretation of the underlying data and has the advantage of being part of the model rather than being transformed for display purposes.

## Reproducibility of the targeted DNA assay

Two similar targeted DNA assay protocols (V1 and V2) for plasma cfDNA and matching WBC gDNA samples were employed in this study. The main differences were the UMI sequences in the library adapters and the reaction volumes in hybridization enrichment process, neither of which would be expected to influence results. To ensure assay performance equivalence, six patient samples with a large cfDNA yield were selected (**Supplementary Table 6**), allowing for reprocessing with both assay protocols, as well as droplet digital PCR (ddPCR). Bio-Rad ddPCR was used to measure biopsy-matched canonical hotspot variants (*PIK3CA* E542K, *PIK3CA* H1047R, *KRAS* G12C, *KRAS* G12A, *EGFR* L861Q) in five of the six patient samples, revealing a 100% positive percent agreement (PPA) and 100% negative percent agreement (NPA) considering ddPCR as the benchmark whilst the VAF measured in cfDNA using either version of the assay protocol closely mirrors that obtained with ddPCR (**Fig. 1b**). In addition, we have performed ddPCR assays targeting VUSo (i.e. somatic cfDNA variants that were not matched to the paired tumors or WBCs) to evaluate the specificity of the cfDNA sequencing assay for low VAF non-tumor-matched somatic variants. For this analysis, we first identified the subset of VUSo for which a validated ddPCR assay was available. We then identified all the patients with residual extracted cfDNA and/or leftover pre-enrichment sequencing libraries harboring any of these mutations. Our final analysis included four ddPCR assays (*PIK3CA* E545K, *NRAS* Q61K, *CTNNB1* T41I, *CTNNB1* S33C) targeting VUSo in seven patients. For one patient, only cfDNA extracted from plasma was available. For two patients, both cfDNA and pre-enrichment sequencing libraries were available and tested, whereas for the remaining four patients, only pre-enrichment libraries were available. As a negative control, the pre-enrichment sequencing libraries from 12 patients where the target VUSo were not detected in cfDNA were used. All experiments were performed in triplicate. The results are shown in **Fig. 4x,** whereby the cfDNA sequencing assay had a 100% PPA and 100% NPA considering ddPCR as the benchmark. **Fig. 1c** and **Supplementary Fig. 3** show the measured allele fraction of variants called either using the V1 or V2 targeted DNA assay protocol. Measured VAFs between the two technical replicates for samples from six patients (**Fig. 1c**) showed a strong agreement (*R2* = 0.9997) as well as for one sample from a hypermutated case (**Supplementary Fig. 3,** *R2* = 0.9972). Additional validation was performed through a second replicate of three samples using V2. The pairwise comparison of VAFs between V1 vs V2 and V2 vs V2 for all the samples that have been retested are shown in **Supplementary Figure RR8**. Overall, across the five non-hypermutated patients, 170 of variants detected in V1 have been tested and 152 (89.4%) of these were also detected in V2. In the hypermutated MBC patient, 630 of 659 (95.6%) of variants detected in V1 were also detected in V2. Comparing version V1 and V2 for the three patients who have been retested, 45 of 51 (88.2%) variants were found present in two non-hypermutated patients whilst the corresponding number in the hypermutated MBC patient was 640 of 659 (97.1%).

We next assessed the associations between sequencing depth, target coverage and mutation VAFs. In brief, 215 somatic mutations labelled as either one of `Biopsy-matched`, `Biopsy-subthreshold`, `WBC-matched` or `VUSo` at collapsed depth >10,000X. Of which, 121 (56.3%) and 20 (9.3%) were detected in samples from patients MSK-VB-0023 and MSK-VB-0050, respectively, both of which were samples from hypermutated cancers (**Supplementary Fig. RR14**). Only 74 mutations occurred with depth >10,000X, of which 57 (77%) were also detected in the matched tumor biopsy or WBC. **Supplementary Figs. RR14b-RR14c** demonstrate that the total depth of somatic mutations detected in cfDNA is a function of the mean collapsed target coverage in cfDNA, which is itself a function of the amount of input DNA used for library preparation. No association between the VAF and the sequencing depth of variants was observed, irrespective of source of origin. As the majority of variants detected at >10,000X occurred in patient MSK-VB-0023, we investigated whether these co-occurred with copy number gains or amplifications **(Supplementary Figs. RR15).** A gain of chromosome arm 1q and a hemizygous loss of chromosome arm 16q were detected in cfDNA and the matched tumor biopsy MSK-VB-0023; no other copy number alterations were identified. These findings are consistent with the notion that the mutations detected at >10,000X depth cannot be attributed to gains or amplifications of the corresponding loci **(Supplementary Figs. RR15)**.

***Microsatellite instability detection in high depth-of-read cfDNA assays***

The standard program MSIsensor 39 detects microsatellite instability by the following steps: (1) using a catalogue of sites built from the human reference genome, all read pairs with at least one read mapping within 2 kb of the sites are retrieved from the tumor and normal samples, (2) at candidate sites, a histogram of k-mer alleles with different repeat lengths is constructed by enumerating observed instances of the k-mer in the tumor and normal samples separately, and (3) finally, the difference in the distribution of counts is used to define a candidate site with ≥20 reads in the tumor and normal as being unstable using a standard 𝝌2 goodness-of-fit test. MSIsensor reports the percentage of sites classified unstable as the MSI score.

However, the default settings of MSIsensor when applied to cfDNA data generated in this study led to suboptimal performance, producing inappropriately elevated MSI scores across samples (**Supplementary Fig. 11a**). This required reexamining the assumptions of the MSIsensor algorithm to optimize MSI detection in high depth-of-read, error-corrected, cfDNA data.

The classification of a candidate site as unstable was designed for relatively shallow depth of sequencing and uses a standard 𝝌2 statistic to determine if the distribution of counts between the tumor and normal sample is different. At high read depths, even small differences in the distribution of k-mer counts between experiments will be detected as statistically significant in the absence of a biologically significant event. Additionally, if at a candidate site, the 𝝌2 test is positive, MSIsensor classifies the site as unstable without determination of the direction of the effect (i.e., normal tissue apparently unstable with respect to tumor).

Therefore, a set of filters was used to correct the MSIsensor analysis for the high depth-of-read cfDNA data as follows: (1) at candidate sites, the direction of instability as measured by the difference in entropy between the distributions of k-mers had to indicate tumor as the more unstable of the pair with respect to normal tissue, (2) the absolute magnitude of the instability as measured by the square root of the Jensen-Shannon divergence (information radius) had to exceed a cutoff chosen to remove spurious variation, (3) heterozygous variants in normal tissue were excluded, as copy number variations could render them apparently MSI unstable, and (4) the standard Benjamini-Hochberg multiple testing correction was applied where the documented version of MSIsensor uses a minor variation thereof.

Using the distributions obtained from MSIsensor and applying the updated filters, more robust results were obtained in both tumor-normal utilizing MSK-IMPACT and the higher depth-of-read cfDNA-WBC samples (**Supplementary Fig. 11b**). These results suggest that the high depth-of-read cfDNA data generated in this study are suitable for detecting MSI in cancer, and that MSI detection can be further improved in shallow sequencing biopsies.

**Gene copy number variation (CNV) detection**

For the CNV exploratory analysis, we first computed the Log2 Ratios, absolute copy numbers, purity and ploidy of the tumor biopsies using FACETS (PMID: 27270079). The raw read counts of the cfDNA and WBC samples were extracted from deduplicated uncollapsed BAM files and processed using CNVkit (PMID: 27100738) to obtain an estimate of Log2 Ratios for both on- and off-target regions. The raw read counts were corrected for library size, GC content and target length after which the WBC samples were averaged and used as baseline to normalize each cfDNA sample. The resulting Log2 Ratios were smoothed using the median absolute deviation and segmented by penalized least squares regression using the R/Bioconductor library copy number (PMID: 23442169) with default parameter values, except for the minimum number of probes in each segment and the penalty parameter for allowing discontinuities in Log2 Ratios. Absolute copy numbers were inferred from the segmented Log2 Ratios based on the ctDNA fraction estimates using the following equation:

where is the absolute copy number of a given segment, is the ctDNA fraction estimate of the sample, is the ploidy of the cfDNA sample, is the Log2 Ratio of the given segment and is a compression ratio. Since is unknown, we use an iterative approach to minimize the weighted sum of squared errors (SSE) over a range of ploidy values according to the following equation:

where is the length of segment and is the number of segments in a given sample. The optimal ploidy is given by:

Amplifications and homozygous deletions were inferred for the 410 genes included in the MSK-IMPACT assay based on the amplitude of and as follows:

where is the ordinal copy number call of a given gene, and are ploidy specific thresholds such that:

The Log2 Ratios of representative healthy male and female controls are shown in **Supplementary Figs. RR17a-17f**. The low variance of raw Log2 Ratios observed is expected due to the ultra-high depth sequence coverage achieved for the cfDNA and WBC samples. Given that no systematic technical artefact was observed in the healthy controls, these data could be used to derive parameter estimates for the analysis of cancer samples. There is, usually, a platform specific compression of Log2 Ratio (PMID: 20837533), which describes the ratio of observed to expected Log2 Ratio for a clonal loss of one copy in a diploid background in a sample with 100% purity. The compression ratio was estimated from the mean segmented Log2 Ratio of chromosome X across healthy male controls (**Supplementary Figs. RR17g-17h**).

A comparison of Log2 Ratios of three tumor biopsy and matched cfDNA sample pairs, one of each cancer types, where amplifications (*FGFR1* and *CCND1* in the breast cancer case, *EGFR* in the lung cancer case) and a homozygous deletion (*BRCA2* in the prostate cancer case) were detected in the tumor biopsies is shown in **Supplementary Figs. RR18a-18f**. Based on these initial observations, the concordance between the tumor biopsies and cfDNA samples was evaluated using two complementary approaches:

1. As the correlation of segmented Log2 Ratios for segments overlapping ≥75%. The distribution of Pearson’s *r* is shown in **Supplementary Figs. RR18g-18h**. There was a statistically significant association between the Pearson’s *r* and the ctDNA fraction (one-sided Jonckheere-Terpstra test p = 1e-4). A larger number of the breast and prostate cancer cases displayed high ctDNA fractions and Pearson’s *r* > 0.5. There was no association between the Pearson’s *r* and the purity of the matched tumor biopsy.
2. Based on the previous observation that the Log2 Ratios estimated from the tumor biopsy and cfDNA are comparable only for samples with high ctDNA fraction estimates, we compared amplifications and homozygous deletions in any of the 410 genes included in the MSK-IMPACT assay for only N = 49 tumor biopsy and cfDNA sample pairs with ctDNA fraction estimate ≥10%. The ROC curves for predicting amplifications or homozygous deletions from the absolute copy numbers in cfDNA are shown in **Supplementary Figs. RR19a-19b**. Pairwise comparisons between the cancer cohorts showed that none of the AUC for amplifications were significantly different (DeLong two-sided test for AUC; p>0.05) except comparing breast and prostate cancers (DeLong two-sided test for AUC; p = 0.042). This was expected given that breast cancers had the highest number of amplifications, whereas prostate cancers harbored the lowest. Similarly, comparing homozygous deletions across the different cancer cohorts showed that CRPCs had a statistically higher AUC than MBCs (p = 3.95e-3) and NSCLCs (p = 8.56e-2) possibly explained by the higher number alterations of this nature in CRPCs. We also estimated the probability of detecting an amplification in cfDNA as a function of the amplitude of the corresponding copy number in the matched tumor biopsy for sample pairs with concordant diploid or near diploid genome mass. The probit regressions are shown in **Supplementary Fig. RR19c**.

Overall, 34 of 49 (69%) patients had at least one amplification or homozygous deletion reported in their matched tumor biopsies. Among those, 26 (76.5%) patients displayed at least one concordant alteration. Importantly, 154 amplifications in 100 unique genes were detected in the 34 tumor biopsies and 106 (68.8%) of these were also detected in cfDNA. Only 23 of 51 (45.1%) homozygous deletions detected in the same tumor biopsies, mostly in prostate cancer patients, were also detected in cfDNA. **Supplementary Fig. RR20** illustrates all the amplifications and homozygous deletions detected in cfDNA and the paired tumor tissue. Whilst in many cases, concordant alterations were found, CNVs present in the cfDNA but not in tumor biopsies were detected, consistent with the tumor CNV heterogeneity extensively documented in cancers of advanced stage (e.g. PMID: 22397650, 20981102, 23539594).

Finally, we attempted to evaluate the concordance between the tumor biopsies and cfDNA samples of clinically actionable alterations (OncoKB Level 1; PMID: 28890946) for all patients irrespective of the ctDNA fraction. Four breast cancer patients harbored an amplification of *ERBB2* on 17q and one lung cancer patient had an amplification of *MET* on 7q. The Log2 Ratio of the tumor biopsies and matched cfDNA samples are shown in **Supplementary Fig. RR21**. Three of the five actionable CNVs could be detected. One of the four metastatic *HER2* amplified breast cancer cases did not have detectable *HER2* amplification in the cfDNA, and the *MET* amplified metastatic lung cancer did not have detectable *MET* amplification in the cfDNA. In these cases, the ctDNA fractions estimated were 1.3% and 1.9%, respectively, providing an explanation as to why these could not be detected in cfDNA using the high-intensity cfDNA assay.