Published OnlineFirst August 17, 2016; DOI: 10.1158/0008-5472.CAN-15-3457

had only blood sample as control, 32 (3%) had both blood and tissue as normal controls, and 82 (9%) had only tumor-adjacent tissue as normal control (Supplementary Table S1). To understand the level of contamination of ctDNA in whole-exome sequencing of cancer patients' whole blood DNA and the extent to which it affects somatic mutation calls, we developed a method for the systematic identification of ctDNA mutations from a set of tumor/blood samples. The method we propose should be executed as an additional step to classic somatic mutation analysis for recovering bona fide somatic mutations for which the allele read count or frequency in the blood is higher than expected.

Materials and Methods

Simulated data and performance assessment

We first retrieved the whole exome sequencing data of 99 individuals of CEU population [Utah Residents (CEPH) with Northern and Western European Ancestry] in 1000 Genomes project phase III dataset (Supplementary Table S2). The bam files were then remapped to the reference hg19 (using bwa). To create a set of samples with comparable coverage, we included 66 samples with a mean coverage between 80 and 200.

Simulated tumor samples. We randomly assigned a number of variants ranging from 50 to 500 (SNPs and indels) to each sample with a mutant allele frequency following a Gaussian distribution (10). Because of the consideration of tumor heterogeneity and tumor sample purity, the mean of the Gaussian distribution was set between 0.2 and 0.45. We then chose all the variants from the COSMIC database with CNT ≥ 2 and added the variants to the bam files using Bamsurgeon (11).

Simulated contaminated whole blood samples. We first estimated ctDNA concentration in whole blood DNA from a set of 117 pan-cancer patients, where the cfDNA concentration in the plasma ($C_{\text{cfDNA,plasma}}$) and the allele frequency of somatic mutations identified in the cfDNA ($AF_{\text{somaticM,plasma}}$) were available (Supplementary Table S14; ref. 9). The ctDNA concentration in whole blood was calculated as: $C_{\text{ctDNA, whole blood}} = \frac{C_{\text{ctDNA, plasma}} \times 1\,\text{mL}}{Q_{\text{DNA, whoole blood}}}$, where $C_{\text{ctDNA,plasma}} = \max(AF_{\text{somaticM,plasma}}) \times C_{\text{cfDNA,plasma}}$ and $C_{\text{DNA,whole blood}} \sim 15 - 60\ \mu\text{g/mL}$ (DNA/RNA Mini Kit protocol, Qiagen), $C_{\text{plasma,whoole blood}} = 55\%$ and $Q_{\text{DNA, whoole blood}} = C_{\text{DNA,whole blood}} \times (1\ \text{mL}/C_{\text{plasma, whole blood}})$. On the basis of this estimation, we found that the maximum concentration of ctDNA in whole blood was above 2% (Supplementary Table S14). We then set the percentage of contamination to a Gaussian distribution of mean equals to 1.38% (for final values ranging from 0.08% to 2.99%). To simulate the reality where the normal samples are usually sequenced at a lower coverage compared with the tumor samples, and to avoid having the tumor sample greatly identical to the simulated whole blood sample, we mixed the tumor and the normal samples at half of the corresponding proportion to create a mixed sample of lower mean coverage (0.5 $\times$ mean coverage of tumor sample).

Performance. Somatic variants were identified with Mutect and Varscan2 with different sets of filters (maximum variant allele frequency ranging from 0 to 0.15 and maximum variant supporting reads ranging from 0 to 10; Supplementary Table S5). We then compared the performance of traditional somatic variant callers alone versus traditional somatic variant callers plus cmDetect by evaluating the Recall, Precision, and *F*-score (11). False negatives were defined as true somatic mutations incorrectly filtered out by mutations callers because of high detection level in the blood sample. The confidence interval (CI) of each call with different filters was calculated on the basis of a binomial distribution using Clopper–Pearson method (12).

Metastatic samples

We used 86 tumor–normal sample pairs from patients included in the SAFIR01 trial (NCT01414933; ref. 13), and 67 tumor–normal pairs from the MOSCATO trial (NCT01566019). Of these, 93 samples were from metastatic breast cancer and were sequenced on a Hiseq sequencer, whereas the remaining 60 were from a panel of different cancers and were sequenced on a NextSeq sequencer. Tumor DNA was extracted from frozen tissue from a biopsy sample taken in the context of the corresponding trial. A surgical pathologist reviewed the samples for diagnosis purpose and assessed tumor cell content (Supplementary Table S6) before whole-exome sequencing. The average tumor cell content was 61% for the 93 metastatic breast cancer samples and 52% for the 60 pan-cancer samples. Normal DNA was extracted from whole blood, taken at the same time as the biopsy. All patients gave their informed consent for translational research and genetic analyses of their germline DNA.

Whole exome sequencing. Genomic DNA was captured using Agilent in-solution enrichment methodology with their biotinylated oligonucleotides probes library (SureSelect Human All Exon v5–50 Mb; Agilent), followed by paired-end 75 bases massively parallel sequencing on Illumina HiSeq 2500 or NextSeq 500 sequencer.

HiSeq 2500. Sequence capture, enrichment, and elution are performed according to manufacturer's instruction and protocols (SureSelect; Agilent) without modification. Briefly, 600 ng of each genomic DNA are fragmented by sonication and purified to yield fragments of 150 to 200 bp. Paired-end adaptor oligonucleotides from Illumina are ligated on repaired, A-tailed fragments then purified and enriched by four to six PCR cycles. Five hundred nanograms of these purified Libraries are then hybridized to the SureSelect oligo probe capture library for 24 hours. After hybridization, washing, and elution, the eluted fraction is PCR amplified with 10 to 12 cycles, purified and quantified by qPCR to obtain sufficient DNA template for downstream applications. Each eluted-enriched DNA sample is then sequenced on an Illumina HiSeq 2500 as paired-end 75b reads. Image analysis and base calling is performed using Illumina Real Time Analysis Pipeline version 1.12.4.2 with default parameters.

NextSeq 500. Sequence capture, enrichment, and elution are performed according to manufacturer's instruction and protocols (SureSelect; Agilent) without modification except for library preparation performed with NEBNext Ultra Kit (New England Biolabs). For library preparation, 600 ng of each genomic DNA are fragmented by sonication and purified to yield fragments of 150 to 200 bp. Paired-end adaptor oligonucleotides from the NEB Kit are ligated on repaired, A-tailed fragments then purified and enriched by eight PCR cycles. A total of 1,200 ng of these purified libraries are then hybridized to the SureSelect oligo probe capture library for 72 hours. After hybridization, washing, and elution, the

Downloaded from cancerres.aacrjournals.org on October 15, 2021. © 2016 American Association for Cancer Research.

Published OnlineFirst August 17, 2016; DOI: 10.1158/0008-5472.CAN-15-3457

eluted fraction is PCR amplified with nine cycles, purified, and quantified by qPCR to obtain sufficient DNA template for downstream applications. Each eluted-enriched DNA sample is then sequenced on an Illumina NextSeq 500 as paired-end 75b reads. Image analysis and base calling was performed using Illumina Real Time Analysis (RTA 2.1.3) with default parameters.

Somatic mutations calling. Fastq files were aligned to the reference genome hg19 with the Burrows-Wheeler Alignment tool (BWA) 0.7.5a mem algorithm (14). After alignment, the BAM files were treated for PCR duplicate removal then sorted and indexed with samtools (15) version 0.1.19 (options rmdup, sort and index) for further analyses. Base recalibration and local realignment around indels was done with GATK. For defining somatic mutations, we used the Mutect version 1.1.4 algorithm for identifying substitutions and the IndelGenotyper (IndelGenotyper.36.3336-GenomeAnalysisTK.jar) algorithm for identifying small insertions and deletions (indels). We defined the final list of somatic mutations with the following filters: frequency of the reads with the altered base in the tumor $\geq$0.1; number of reads with the altered base in the tumor $\geq$5; frequency of the reads with the altered base in the normal <0.03; number of reads with the altered base in the normal <2; not in dbSNP database except for variants that are also in COSMIC with a variant allele frequency in 1000G < 0.001 or not reported. The resulting somatic mutations were annotated with the snpEff 4.1c algorithm (16).

TCGA cases

Whole exome sequencing data of 60 primary solid tumors of breast invasive carcinoma and corresponding blood derived normal were randomly selected from TCGA. To avoid the bias of breast cancer subtypes, we selected 20 samples from each subtype (Her2 amplified, triple negative, hormone receptor positive with Her2 not amplified). A list of 4,286 curated somatic mutations was also retrieved from TCGA (genome.wustl.edu_BRCA.IlluminaGA_DNASeq.Level_2.1.1.0.curated.somatic.maf).

ctDNA mutation identification workflow

The method first retrieves heterozygous variants using GATK HaplotypeCaller in each tumor sample using hard filters (Quality By Depth QD > 2.0, strand bias FS < 60.0, mapping quality MQ > 40.0, MappingQualityRankSum > −12.5, ReadPosRankSum > −8.0, GenotypeQuality > 30 for SNPs and QD > 2.0, FS < 30, ReadPosRankSum > −20.0 for INDELs). The variants in coding regions and satisfying the following filters are then selected: $\geq$5 reads supporting the variant; total base depth $\geq$ 10 and variant allele frequency $\geq$ 0.1. For each position with a variant detected with these filters, the number of reads supporting the reference and the variant allele in the BAM files of the tumor and corresponding blood sample are retrieved using samtools mpileup (15) with a minimum score of 20 for both base and mapping qualities. Only the variants with at least one supporting read for the variant in the BAM file of the blood sample are kept. Patient polymorphisms are then filtered out by three strategies: (i) the difference of allele frequency in the tumor and corresponding normal samples is tested with a one-tailed Fisher exact test (FDR > 0.01); (ii) the probability of each variant to be germline is computed by comparing the variant read frequency in the normal sample to the read frequency distribution of germline heterozygous SNPs per patient (empirical P > 0.01), and (iii) the minor allele frequency (MAF) of each variant in the population under study is calculated as the number of patients with an allele frequency >0.1 in both the tumor and the blood divided by the total number of patients in the population (observed MAF > 0.01). Finally, we filtered out known polymorphisms as defined in dbSNP database (version 138) after excluding positions with variants present in COSMIC (version 67) unless the observed 1,000 Genomes MAF was higher than 0.001.

Sequencing bias. To distinguish a ctDNA mutation from a sequencing bias, we retrieved the variant-supporting reads in each blood sample for all patients for each of the positions detected in the previous step. We hypothesized that, in the absence of a true variant, the variant-supporting read counts N_v should follow a binomial distribution $B(N_t, p_{error})$, where N_t is the total number of reads at the position of interest and p_{error} is the sequencing error rate, estimated by the variant read frequency in the mixture of the blood samples of all the patients. A variant with n_v supporting reads for a total depth of n_t was

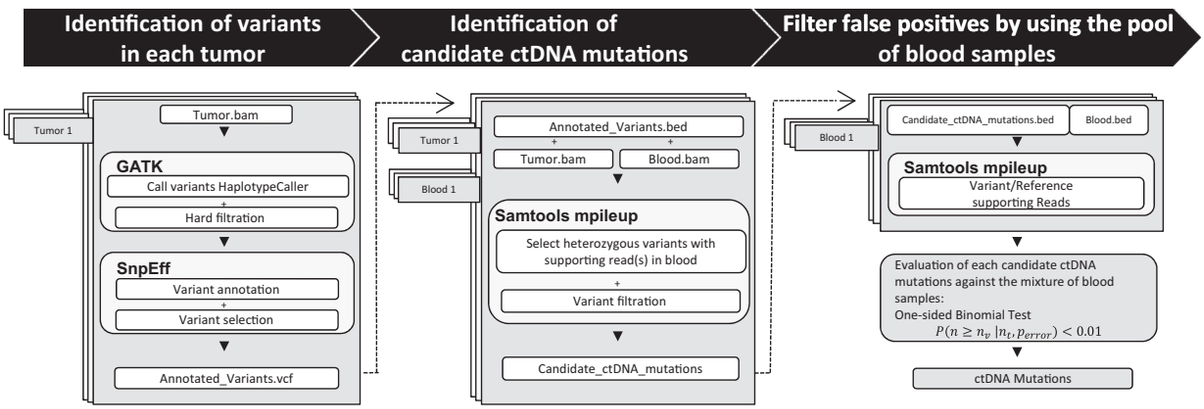

Figure 1.

cmDetect workflow. The method consists of three main steps including: (i) tumor variant identification based on GATK and SnpEff; (ii) selection of non-germline variants with some evidence support in the corresponding blood sample; and (iii) filtration of candidate variants based on sequencing error from the pool of blood samples.

Downloaded from cancerres.aacrjournals.org on October 15, 2021. © 2016 American Association for Cancer Research.

Published OnlineFirst August 17, 2016; DOI: 10.1158/0008-5472.CAN-15-3457

considered as a true variant if $P(n \geq n_v \mid n_t, p_{error}) < 0.05$ using a one-sided binomial test (17).

Plasma mutations

Details of plasma DNA extraction and sequencing can be found in the publication by Jovelet and colleagues (9). Fastq files were treated with the Torrent Suite BaseCaller version 4.0 or 4.2. We retrieved hotspot variants using GATK HaplotypeCaller in each plasma sample and satisfying the following filters: strand bias FS < 30; variant supporting reads >4; total base depth >50; and variant allele frequency ≥0.1. We filtered out variants present in polymorphism databases (ESP, the Exome Sequencing Project or 1000G, from European samples in 1000 genome project; ref. 18) with a minor allele frequency >0.001.

Survival analyses

Overall survival was estimated by the Kaplan–Meier methods. Correlation between the number of ctDNA mutations and survival was assessed using Cox-proportional hazard models. Univariate analysis was performed using Log-rank test for categorized variables. Multivariate analysis was assessed using Cox proportional Hazard Modeling. All factors with $P < 0.10$ in univariate analysis were evaluated on multivariate analysis. All P values reported are two-sided. For all statistical tests, differences were considered significant at the 5% level. Stata 13.0 was used for all statistical analyses.

Results

ctDNA mutation detection workflow

We introduce a method, cmDetect (ctDNA mutation detection), for the systematic identification of ctDNA mutations by leveraging information from the tumor and blood samples (Fig. 1). cmDetect consists of three steps and is described in details in the method section. Briefly, the proposed workflow first retrieves heterozygous variants in gene coding regions in each tumor independently using the Genome Analysis ToolKit (GATK; ref. 19) and selects those variants with supporting read(s) in the corresponding blood sample. The patient's germline variants and common polymorphisms are then filtered out to obtain a set of mutations that are identified with high confidence in the tumor sample and lower incidence in the blood sample. At this step, as the read frequency of the selected variants in the blood may be very small (<0.02), it is important to be able to distinguish ctDNA mutations from sequencing biases. Therefore, the frequency of each selected variant is tested against the observed frequencies of

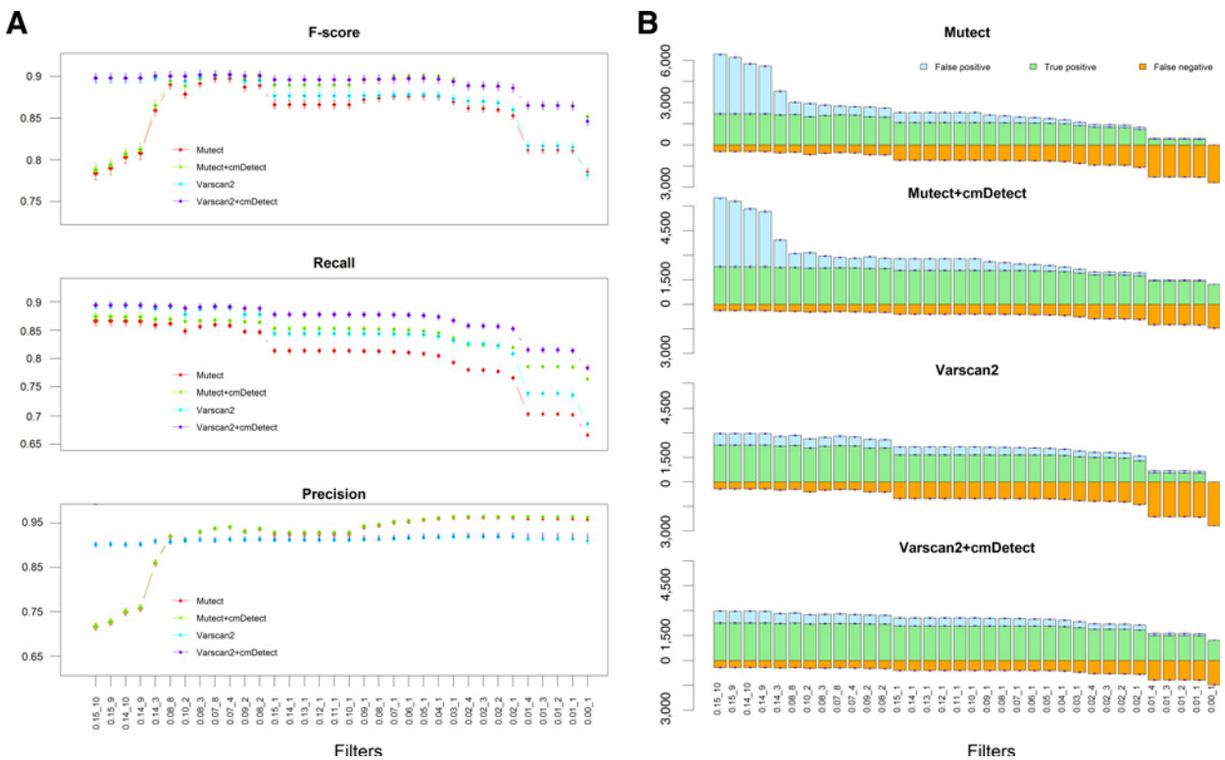

Figure 2.

A, combining cmDetect with somatic mutation caller(s) improves the performance for somatic mutation identification. *F*-score, recall, and precision (*y*-axis) between Mutect (red), Mutect+cmDetect (green), Varscan2 (blue), and Varscan2+cmDetect (violet) at different filters (*x*-axis) are shown. Error bars correspond to 95% CIs. **B,** sensitivity of somatic mutation calling by adding cmDetect to Mutect(Varscan2) for different filters. Bar plots showing the numbers of false positives (blue), true positives (green), and false negatives (yellow) called by Mutect only, Mutect+cmDetect, Varscan2 only, and Varscan2+cmDetect with different filters for somatic mutations with at least one supporting read in the blood. The vertical axis shows the number of mutations, true positives are shown above 0, and false negatives are shown below 0. The horizontal axis corresponds to the different filters applied for somatic mutation calling, from the most stringent filter (left) to the least stringent filter (right).

Downloaded from cancerres.aacrjournals.org on October 15, 2021. © 2016 American Association for Cancer Research.

Published OnlineFirst August 17, 2016; DOI: 10.1158/0008-5472.CAN-15-3457

the same variant in the mixture of blood samples to estimate its probability of being a false positive (due to sequencing bias or bad alignment).

Benchmarking

To estimate the statistical power of the cmDetect workflow, a set of 66 tumor/blood pairs of samples was simulated on the basis of real whole-exome sequencing data from the 1000 Genomes project (Supplementary Table S2; ref. 19). Briefly, for each sample, we derived a tumor sample by introducing cancer-causing mutations (COSMIC; ref. 20) at different read frequencies and a normal blood sample contaminated with tumor DNA at various levels (see Materials and Methods; Supplementary Table S2). This way, we introduced a total of 12,469 somatic mutations including 2,676 ctDNA mutations (Supplementary Table S3). We identified ctDNA mutations with cmDetect and retrieved somatic mutations with Mutect (21) and Varscan2 (22) for a range of filters on maximum detection level of the mutation in the normal sample. Application of cmDetect identified 1,219 ctDNA mutations, all true positives (Supplementary Table S4). In addition, 1,458 (54%) ctDNA mutations were not identified by cmDetect for the following reasons: (i) the mutations did not have enough support evidence in the normal sample to be called a ctDNA mutation; (ii) the read frequency in the tumor and corresponding normal was comparable; and (iii) the read frequency in the normal could not be distinguished from a polymorphism (see the polymorphism filtering section in Methods and Supplementary Fig. S1). However, a large proportion (>50%) of the ctDNA mutations missed due to a very low coverage in the normal were correctly identified by the somatic mutation callers [749 (51%) by Mutect and 830 (57%) by Varscan2]. By evaluating the performance of the somatic mutation callers with and without cmDetect at the various filters, we show that without decreasing the sensitivity, adding ctDNA mutations to the results of traditional somatic variant callers decreased the number of false negatives for all the different filters tested (Supplementary Table S5; Fig. 2A). Importantly, the observed recall was not dependent on the initial filters applied in Mutect or Varscan2 (Fig. 2B), indicating that cmDetect can be efficiently combined with somatic mutation callers. This also demonstrated that the gain in sensitivity cannot be achieved by relaxing the initial mutation caller filters, for example by increasing the maximum allowed allele frequency of the variant in the blood, which will have dramatic effects on specificity.

ctDNA is detectable in whole-exome sequencing of metastatic breast cancer patients' blood

We applied our method to a panel of 93 whole-exome sequenced pairs of breast cancer metastases/blood samples from the SAFIR01 (NCT01414933; ref. 13) and MOSCATO (NCT01566019) clinical trials (see Supplementary Tables S6 and S7, for clinical information and sequence quality metrics). We first applied Mutect and indelGenotyper and identified

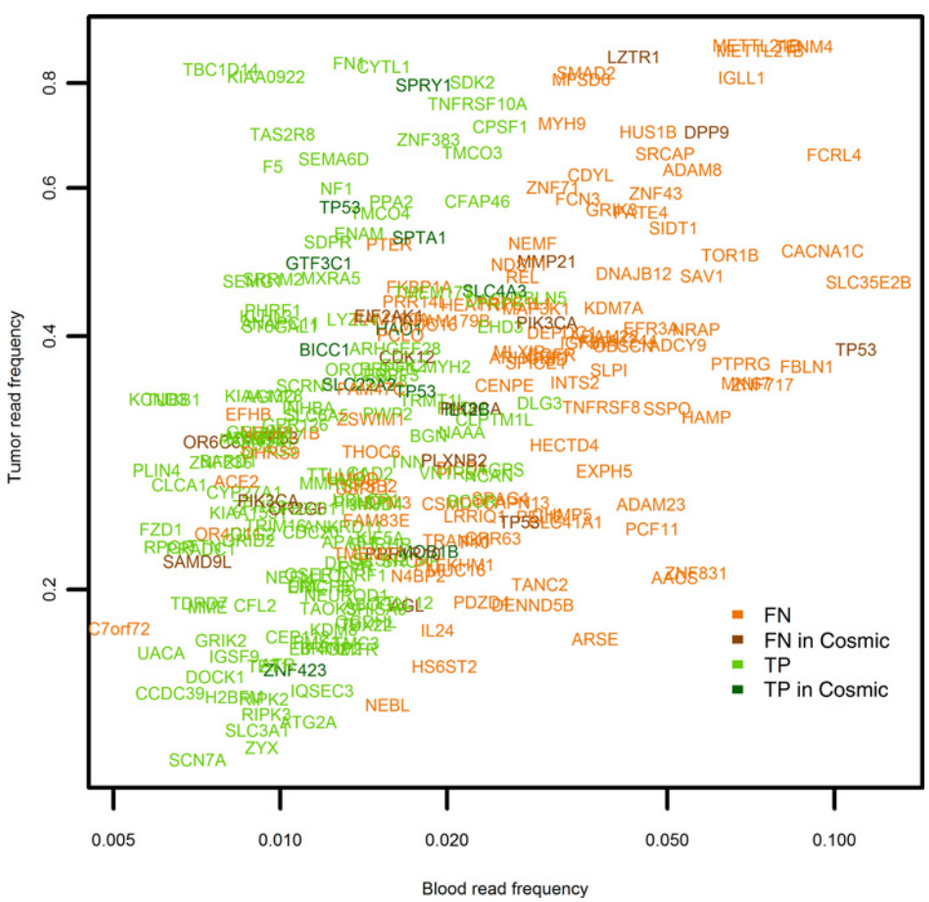

Figure 3.
Observed read frequencies of ctDNA mutations identified by cmDetect in breast cancer metastatic samples. The color of the gene corresponds to the status of the call of the mutation according to Mutect applied with default filters as described in Materials and Methods. Shown is the allele frequency in the blood (*x*-axis) and in the corresponding tumor (*y*-axis). False negative (FN; orange) corresponds to the ctDNA mutations identified by cmDetect but not by Mutect. True positive (TP; green) corresponds to the ctDNA mutations identified by cmDetect and Mutect. The mutations that are documented in COSMIC are shown in darker color.

Downloaded from cancerres.aacrjournals.org on October 15, 2021. © 2016 American Association for Cancer Research.

Published OnlineFirst August 17, 2016; DOI: 10.1158/0008-5472.CAN-15-3457

7,334 somatic mutations (Supplementary Table S8) for the 93 pairs of metastasis/blood samples (see Materials and Methods). The application of cmDetect identified 263 ctDNA mutations (Supplementary Table S9) distributed in 50 patients (54%). Among these 263 mutations, 141 were correctly identified by the somatic mutation analysis whereas 122 were false negatives, defined as true somatic mutations incorrectly filtered out by the mutation callers due to the high level of detection in the control blood sample, representing 1.67% of the total number of somatic mutations. Importantly, among the false negatives we found 12 mutations reported in clinical databases (COSMIC) including two *PIK3CA* missense mutations (V344M, H1047R) and two stop-gain *TP53* mutations (R306*, E349*), with a read frequency as high as 0.11 with seven supporting reads in the corresponding blood sample (E349*; Fig. 3). Other mutations of interest among the false negatives included one *EGFR* missense mutation (G322S) and one small insertion in *TP53* (N288fs). In the following, we consider a total of 7,457 somatic mutations among the 93 pairs of metastases/blood samples.

The 50 patients had in average 5.26 and a maximum of 38 ctDNA mutations, most of them (82%) having less than 10 ctDNA mutations identified. Patients with detectable ctDNA in the blood, identified as patients with at least one ctDNA mutation, had more somatic mutations in their tumor than patients with no detectable ctDNA (*t*-test $P = 0.0003$; Supplementary Fig. S2). However, there was not a direct correlation between number of ctDNA mutations and total number of mutations (Pearson $\rho = 0.12$; $P = 0.24$), indicating that the mutational load of the tumor does not reflect the amount of ctDNA detectable in the whole blood sample. Importantly, ctDNA mutations represented up to 53% of the total number of somatic mutations whereas false negative rates ranged from 0% to 35% of the total number of mutations per patient (Fig. 4). We also confirmed in the 86 SAFIR01 cases that patients with detectable ctDNA mutations were associated to poor outcome, marginally when only considering the presence/absence of ctDNA mutations ($P = 0.068$; Fig. 5), but very significantly when considering ctDNA mutations quantitatively ($P < 0.001$). Indeed, the number of mutations were highly significant in both univariate analysis (HR = 1.14; $P < 0.001$) and in multivariate analysis (HR = 1.15; 95% CI, 1.07–1.23; $P < 0.001$). Finally, we found that patients who received a prior chemotherapy (Supplementary Table S6) presented more ctDNA mutations (mean = 3.07) than the patients who did not (mean = 0.56; *t* test $P = 0.002$; Supplementary Fig. S3).

To further evaluate the extent of ctDNA contamination in early-stage disease, we applied cmDetect to a set of 60 primary tumor–blood pairs of whole-exome sequencing data from the TCGA breast cancer collection (see Materials and Methods; ref. 23).

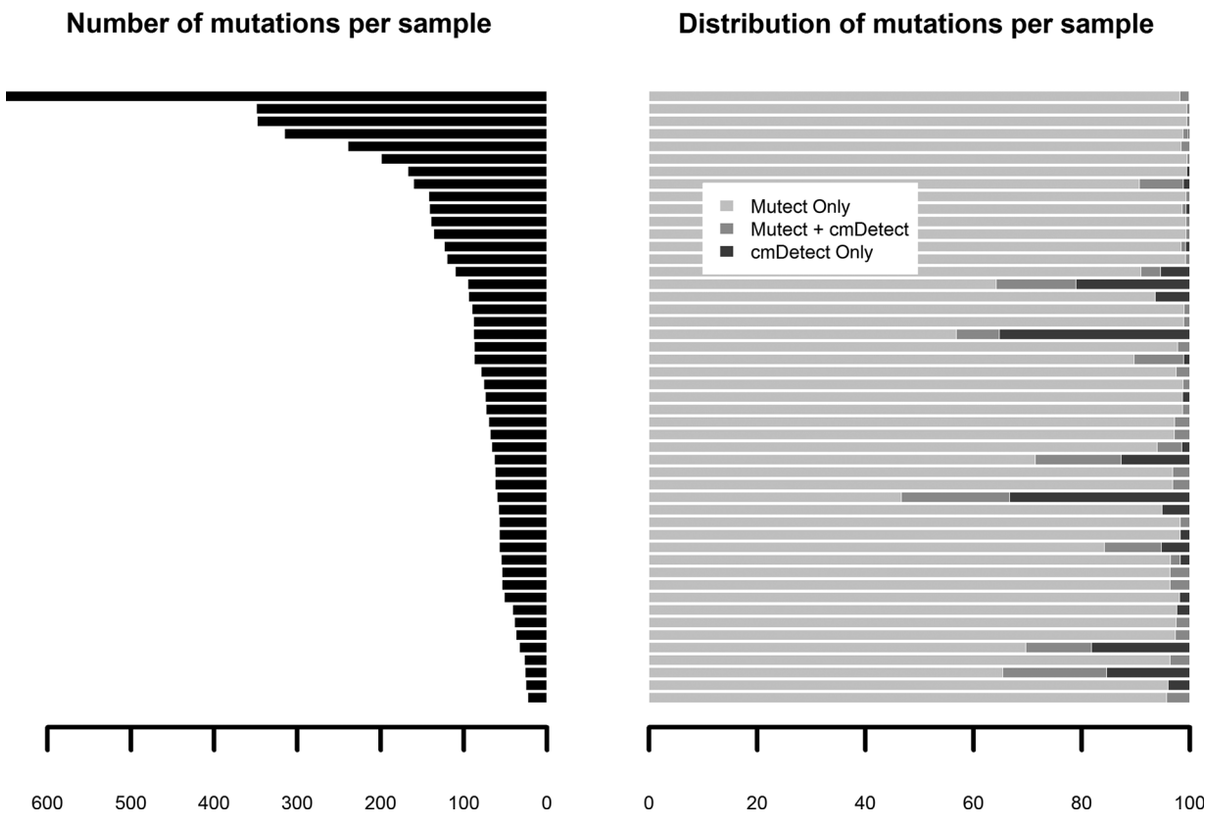

Figure 4.

Somatic mutation profile of breast cancer metastatic patients with detectable ctDNA in whole blood. Left, total number of somatic mutations per sample; right, percentage of somatic mutation types. Mutect only, mutations identified by Mutect but not cmDetect; Mutect+cmDetect, ctDNA mutations identified by Mutect and cmDetect; cmDetect only, ctDNA mutations identified by cmDetect but not Mutect.

Downloaded from cancerres.aacrjournals.org on October 15, 2021. © 2016 American Association for Cancer Research.

Published OnlineFirst August 17, 2016; DOI: 10.1158/0008-5472.CAN-15-3457

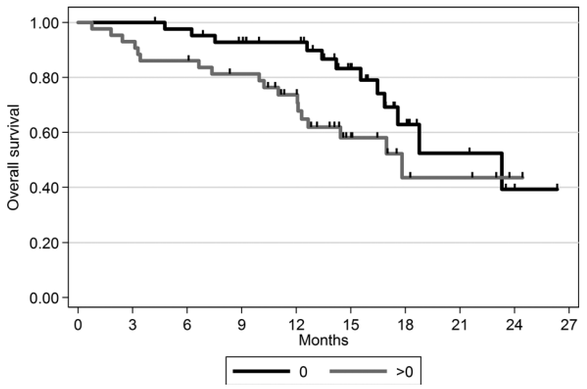

Figure 5.

Patients survival in the SAFIR01 trial according to detectable ctDNA status. The black line (0) represent patients with no detectable ctDNA, whereas the gray line (>0) contains the patients with at least one ctDNA mutation identified.

We identified 41 ctDNA mutations in 18 patients (30%) harboring from 1 to 14 ctDNA mutations with an average of two mutations per patient (Supplementary Table S10). Among these 41 mutations, 10 were present in the somatic mutation results file from TCGA whereas 31 (0.7% of the total number of somatic mutations) were false negatives including nine that were reported as clinical variants (COSMIC; Supplementary Table S10).

Validation with plasma samples

We applied cmDetect to a set of 60 pairs of metastasis/blood samples from the MOSCATO clinical trial sequenced on a NextSeq500, including cancers of different tissues of origin (Supplementary Table S7). For 46 MOSCATO patients, 43 in this cohort and three in the breast cancer metastasis cohort, targeted sequencing data of the plasma was also available (Supplementary Table S11) and reported 12 COSMIC mutations (Supplementary Table S12), validating the presence of ctDNA for a total of nine patients. In parallel, cmDetect identified ctDNA mutations from the whole-exome sequencing data for eight of these nine (89%) patients, confirming the sensitivity of our approach (Supplementary Table S13). We found that two of the 12 COSMIC mutations identified in the plasma were also identified by cmDetect from the whole-blood sample: one *TP53* stop-gained mutation (E349*) in patient BC93 with an allele frequency of 0.76 in the plasma, 0.39 in the tumor, and 0.12 in total blood, and one *TP53* stop-gained mutation (R213*) in patient PCAN39 with an allele frequency of 0.2 in the plasma, 0.35 in the tumor, and 0.013 in total blood. It is interesting to note that BC93 presented with the highest mutated DNA fraction (0.76) in the plasma and also had the highest number of ctDNA mutations (38) detected by cmDetect.

Discussion

We introduced the first method for identifying somatic mutations from whole-exome sequencing data that takes into account the possible presence of ctDNA. We propose to use cmDetect as an additional step to traditional somatic mutations analysis pipelines when analyzing whole-exome sequencing of large sets of pairs of tumor/whole blood samples but it can also be used as a standalone workflow for the identification of ctDNA mutations. We showed that cmDetect is very sensitive but presented some limitations. First, cmDetect will not identify ctDNA mutations with blood read frequencies comparable to polymorphisms' observed read frequencies. Indeed, the maximum blood read frequency for a ctDNA mutation detected by cmDetect in the breast cancer metastasis cohort was 0.11 (Fig. 3). Second, cmDetect will also miss ctDNA mutations having comparable tumor and blood read frequencies as it uses a Fisher exact test to differentiate frequencies in tumor and blood samples, a filtering step that is also applied in traditional somatic mutation callers. Finally, ctDNA mutations with very low coverage in the blood sample will usually be missed by cmDetect, as they will not be distinguishable from sequencing errors. However, we demonstrated that most of these mutations will be identified by somatic mutation callers with default filters for the variant detection in the blood. We demonstrated that metastatic breast cancer patients' whole blood may contain ctDNA mutations detectable at low frequencies by whole-exome sequencing, even at relatively low coverage (mean coverage was 70× in the blood samples). Indeed, some of the variants identified at low frequency in the blood samples were known cancer mutations and were validated in the plasma of two patients, demonstrating that our method was sensitive enough to retrieve somatic mutations in the blood at a frequency as low as 0.013. Importantly, we demonstrated that, in combination with cmDetect, traditional mutation callers can be used with stringent filters on blood read count and frequency supporting the alternate allele, therefore reducing the number of false positives while increasing the number of true positive. Although it is always possible to recover known cancer-related mutations by using databases such as COSMIC, other not-documented mutations may be completely missed if they present read counts or frequencies in the whole blood higher than expected. An alternative approach to limit the contamination of normal germline DNA by ctDNA consist in using either blood cells pellet after discarding plasma, or purified peripheral blood mononuclear cell (PBMC). The use of normal tissues such as skin biopsies or fibroblast culture is also possible but appears to be a heavy procedure in standard biological samples' collection, whereas the bioinformatics method proposed may overcome these problems with an analytical solution and provided additional prognostic information associated to presence of ctDNA. Indeed, the number of ctDNA mutations per patient was strongly associated to survival in metastatic breast cancer patients and may reflect the amount of mutated DNA in circulation. This is consistent with previous analyses that have shown that increasing levels of ctDNA and CTC counts were associated with inferior survival in metastatic breast cancer (24, 25). Although we found a significant number of ctDNA mutations in metastatic cases, we also showed that ctDNA mutations may be identified in early-stage disease, but to a lower extent. The proposed approach was developed for the purpose of identifying tumor DNA contamination in germline DNA from whole blood, but it might also be useful for detecting contamination in blood samples from Heme-malignancies with minimum residual disease (MRD) or in DNA extracted from tumor-adjacent normal tissue.

Disclosure of Potential Conflicts of Interest

M. Campone has received speakers bureau honoraria from Novartis, AstraZeneca, Pfizer, and is a consultant/advisory board member for Pfizer and AstraZeneca. T. Bachelot has received speakers bureau honoraria from

Downloaded from cancerres.aacrjournals.org on October 15, 2021. © 2016 American Association for Cancer Research.

Published OnlineFirst August 17, 2016; DOI: 10.1158/0008-5472.CAN-15-3457

Roche and Novartis and is a consultant/advisory board member for Roche, Novartis, and AstraZeneca. J.-C. Soria is a consultant/advisory board member for AstraZeneca. No potential conflicts of interest were disclosed by the other authors.

Authors' Contributions

Conception and design: Y. Fu, J. Garrabey, J.-C. Soria, L. Lacroix, F. André, C. Lefebvre

Development of methodology: Y. Fu, Y. Luo, L. Lacroix, C. Lefebvre

Acquisition of data (provided animals, acquired and managed patients, provided facilities, etc.): C. Massard, M. Campone, C. Levy, V. Diéras, T. Bachelot, J. Garrabey, J.-C. Soria, L. Lacroix

Analysis and interpretation of data (e.g., statistical analysis, biostatistics, computational analysis): Y. Fu, T. Filleron, M. Pedrero, N. Motté, Y. Boursin, Y. Luo, V. Diéras, L. Lacroix, F. André, C. Lefebvre

Writing, review, and/or revision of the manuscript: Y. Fu, T. Filleron, N. Motté, Y. Boursin, C. Massard, M. Campone, V. Diéras, T. Bachelot, J. Garrabey, J.-C. Soria, F. André, C. Lefebvre

Administrative, technical, or material support (i.e., reporting or organizing data, constructing databases): Y. Fu, C. Jovelet, N. Motté, Y. Boursin

Study supervision: L. Lacroix, C. Lefebvre

Acknowledgments

We thank Marta Jimenez from UNICANCER; Marc Deloger and Guillaume Meurice of the bioinformatics core facility of Gustave Roussy and Mélanie Letexier and Emmanuel Martin from Integragen for their assistance. We thank the NIH TCGA project for granting us access to the sequencing data (under project #9082).

Grant Support

This work was supported by Breast Cancer Research Foundation (F. André), Fondation Lombard-Odier "Philanthropia" (F. André), Odyssea (F. André), Operation Parrains Chercheurs (F. André), Dassault Foundation (F. André), French NCI: INCa-DGOS-INSERM 6043 (F. André), SIRIC Socrate (J.C. Soria), and Fondation ARC (to UNICANCER).

The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked *advertisement* in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

Received December 24, 2015; revised June 30, 2016; accepted July 28, 2016; published OnlineFirst August 17, 2016.

References

1. Wang Q, Jia P, Li F, Chen H, Ji H, Hucks D, et al. Detecting somatic point mutations in cancer genome sequencing data: a comparison of mutation callers. Genome Med 2013;5:91.
2. Ignatiadis M, Dawson SJ. Circulating tumor cells and circulating tumor DNA for precision medicine: dream or reality? Ann Oncol 2014; 25:2304–13.
3. Bettegowda C, Sausen M, Leary RJ, Kinde I, Wang Y, Agrawal N, et al. Detection of circulating tumor DNA in early- and late-stage human malignancies. Sci Transl Med 2014;6:224ra24.
4. Diehl F, Schmidt K, Choti MA, Romans K, Goodman S, Li M, et al. Circulating mutant DNA to assess tumor dynamics. Nat Med 2008;14: 985–90.
5. Spindler KL, Pallisgaard N, Andersen RF, Brandslund I, Jakobsen A. Circulating free DNA as biomarker and source for mutation detection in metastatic colorectal cancer. PLoS One 2015;10:e0108247.
6. Haber DA, Velculescu VE. Blood-based analyses of cancer: circulating tumor cells and circulating tumor DNA. Cancer Discov 2014;4:650–61.
7. Leary RJ, Kinde I, Diehl F, Schmidt K, Clouser C, Duncan C, et al. Development of personalized tumor biomarkers using massively parallel sequencing. Sci Transl Med 2010;2:20ra14.
8. McBride DJ, Orpana AK, Sotiriou C, Joensuu H, Stephens PJ, Mudie LJ, et al. Use of cancer-specific genomic rearrangements to quantify disease burden in plasma from patients with solid tumors. Genes Chromosomes Cancer 2010;49:1062–9.
9. Jovelet C, Ileana E, Le Deley MC, Motté N, Rosellini S, Romero A, et al. Circulating cell-free tumorDNA analysis of 50 genes by next-generation sequencing in the prospective MOSCATO trial. Clin Cancer Res 2016; 22:2960–8.
10. Alexandrov LB, Nik-Zainal S, Wedge DC, Aparicio SA, Behjati S, Biankin AV, et al. Signatures of mutational processes in human cancer. Nature 2013;500:415–21.
11. Ewing AD, Houlahan KE, Hu Y, Ellrott K, Caloian C, Yamaguchi TN, et al. Combining tumor genome simulation with crowdsourcing to benchmark somatic single-nucleotide-variant detection. Nat Methods 2015;12:623–30.
12. Clopper S, Pearson ES. The use of confidence or fiducial limits illustrated in the case of the binomial. Biometrika 1934;26:404–13.
13. Andre F, Bachelot T, Commo F, Campone M, Arnedos M, Dieras V, et al. Comparative genomic hybridisation array and DNA sequencing to direct treatment of metastatic breast cancer: a multicentre, prospective trial (SAFIR01/UNICANCER). Lancet Oncol 2014;15:267–74.
14. Li H, Durbin R. Fast and accurate long-read alignment with Burrows-Wheeler transform. Bioinformatics 2010;26:589–95.
15. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, et al. The Sequence Alignment/Map format and SAMtools. Bioinformatics 2009; 25:2078–9.
16. Cingolani P, Platts A, Wang le L, Coon M, Nguyen T, Wang L, et al. A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of Drosophila melanogaster strain w1118; iso-2; iso-3. Fly (Austin) 2012;6:80–92.
17. Bansal V. A statistical method for the detection of variants from next-generation resequencing of DNA pools. Bioinformatics 2010; 26:i318–24.
18. Abecasis GR, Auton A, Brooks LD, DePristo MA, Durbin RM, Handsaker RE, et al. An integrated map of genetic variation from 1,092 human genomes. Nature 2012;491:56–65.
19. McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernytsky A, et al. The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. Genome Res 2010; 20:1297–303.
20. Forbes SA, Beare D, Gunasekaran P, Leung K, Bindal N, Boutselakis H, et al. COSMIC: exploring the world's knowledge of somatic mutations in human cancer. Nucleic Acids Res 2015;43:D805–11.
21. Cibulskis K, Lawrence MS, Carter SL, Sivachenko A, Jaffe D, Sougnez C, et al. Sensitive detection of somatic point mutations in impure and heterogeneous cancer samples. Nat Biotechnol 2013;31:213–9.
22. Koboldt DC, Zhang Q, Larson DE, Shen D, McLellan MD, Lin L, et al. VarScan 2: somatic mutation and copy number alteration discovery in cancer by exome sequencing. Genome Res 2012;22:568–76.
23. Cancer Genome Atlas Network. Comprehensive molecular portraits of human breast tumours. Nature 2012;490:61–70.
24. Dawson SJ, Tsui DW, Murtaza M, Biggs H, Rueda OM, Chin SF, et al. Analysis of circulating tumor DNA to monitor metastatic breast cancer. N Engl J Med 2013;368:1199–209.
25. Bidard FC, Peeters DJ, Fehm T, Nolé F, Gisbert-Criado R, Mavroudis D, et al. Clinical validity of circulating tumour cells in patients with metastatic breast cancer: a pooled analysis of individual patient data. Lancet Oncol 2014;15:406–14.

Downloaded from cancerres.aacrjournals.org on October 15, 2021. © 2016 American Association for Cancer Research.

Published OnlineFirst August 17, 2016; DOI: 10.1158/0008-5472.CAN-15-3457

Cancer Research
The Journal of Cancer Research (1916–1930) | The American Journal of Cancer (1931–1940)

AACR American Association for Cancer Research

Improving the Performance of Somatic Mutation Identification by Recovering Circulating Tumor DNA Mutations

Yu Fu, Cécile Jovelet, Thomas Filleron, et al.

Cancer Res 2016;76:5954-5961. Published OnlineFirst August 17, 2016.

Updated version	Access the most recent version of this article at: doi:10.1158/0008-5472.CAN-15-3457
Supplementary Material	Access the most recent supplemental material at: http://cancerres.aacrjournals.org/content/suppl/2016/08/17/0008-5472.CAN-15-3457.DC1

Cited articles	This article cites 25 articles, 6 of which you can access for free at: http://cancerres.aacrjournals.org/content/76/20/5954.full#ref-list-1
Citing articles	This article has been cited by 4 HighWire-hosted articles. Access the articles at: http://cancerres.aacrjournals.org/content/76/20/5954.full#related-urls

E-mail alerts	Sign up to receive free email-alerts related to this article or journal.
Reprints and Subscriptions	To order reprints of this article or to subscribe to the journal, contact the AACR Publications Department at pubs@aacr.org.
Permissions	To request permission to re-use all or part of this article, use this link http://cancerres.aacrjournals.org/content/76/20/5954. Click on "Request Permissions" which will take you to the Copyright Clearance Center's (CCC) Rightslink site.

Downloaded from cancerres.aacrjournals.org on October 15, 2021. © 2016 American Association for Cancer Research.