FERMI: A novel method for sensitive detection of rare mutations in somatic tissue

Authors: L. Alexander Liggett[1,2], Anchal Sharma[3], Subhajyoti De[3], and James DeGregori[1,2,4,5,6*]

***Corresponding Author. Email:** james.degregori@ucdenver.edu

Affiliations:

[1]Department of Biochemistry and Molecular Genetics, University of Colorado School of Medicine, Aurora, CO 80045, USA.

[2]Linda Crnic Institute for Down Syndrome, University of Colorado School of Medicine, Aurora, CO 80045, USA.

[3]Rutgers Cancer Institute, New Brunswick, NJ 08901, USA.

[4]Integrated Department of Immunology, University of Colorado School of Medicine, Aurora, CO 80045, USA.

[5]Department of Pediatrics, University of Colorado School of Medicine, Aurora, CO 80045, USA.

[6]Department of Medicine, Section of Hematology, University of Colorado School of Medicine, Aurora, CO 80045, USA.

Abstract

With growing interest in monitoring mutational processes in normal tissues, tumor heterogeneity, and cancer evolution under therapy, the ability to accurately and economically detect ultra-rare mutations is becoming increasingly important. However, this capability has often been compromised by significant sequencing, PCR and DNA preparation error rates. Here, we describe FERMI (Fast Extremely Rare Mutation Identification) - a novel method designed to eliminate majority of these sequencing and library preparation errors in order to significantly improve rare somatic mutation detection. This method leverages barcoded targeting probes to capture and sequence DNA of interest with single copy resolution. The variant calls from the barcoded sequencing data then further filtered in a position-dependent fashion against an adaptive, context-aware null model in order to distinguish true variants. As a proof of principle, we employ FERMI to probe bone marrow biopsies from leukemia patients, and show that rare mutations and clonal evolution can be tracked throughout cancer treatment, including during historically intractable periods like minimum residual disease. Importantly, FERMI is able to accurately detect nascent clonal expansions within leukemias in a manner that may facilitate the early detection and characterization of cancer relapse.

Background

Sequencing Challenges

The simultaneous growth in accuracy and reduction in cost of DNA sequencing has encouraged its use throughout many diverse areas of biology. Accompanying this explosion of applications for sequencing has been a natural demand for increasingly sensitive sequencing methods. While the detection of high frequency variants like germline SNPs is not particularly challenging by most sequencing technologies, sequencer and library preparation error rates are typically high enough to mask most rare or somatic variants. What is perhaps most challenging about library preparation is that the very isolation of DNA exposes it to oxidation that can change base identities (Shibutani, Takeshita, and Grollman 1991; Cheng et al. 1992), and DNA probing and amplification can also thermally alter nucleotide identities (Lindahl and Karlstrom 1973; Lindahl and Nyberg 1974).

Because of these sequencing and library preparation limitations, quantitative PCR (qPCR) and multiparameter flow cytometry (MFC) have still remained common methods of rare variant detection (Terwijn et al. 2013). More recent technologies such as high-throughput digital droplet PCR (Hindson et al. 2011; Sykes et al. 1992; Vogelstein and Kinzler 1999), COLD-PCR (Milbury et al. 2012; J. Li et al. 2008), and BEAMing (Dressman et al. 2003) have shown promise for rare mutation detection, but

are often limited to variant allele frequencies (VAFs) greater than 1 percent or are restricted to assaying only a few chosen mutations at a time.

A number of studies have sought improvements in sequencing technology accuracies by targeting and labeling small regions of genomic DNA such as sMIPs (Hiatt et al. 2013; Young et al. 2016), paired strand collapsing (Kennedy et al. 2014) and other targeting methods (Thol et al. 2018; Mansukhani et al. 2018; Onecha et al. 2018). Some groups have also incorporated error correction methods to eliminate sequencing and PCR errors, like PELE-Seq (Preston et al. 2016), and error correcting enrichment processes (Schmitt et al. 2015). While these targeting, and enrichment methods have certainly improved rare variant detection, they are still often limited to detecting variants that exist in at least 1% of a sample, or are limited to simultaneous detection of only a handful of variants.

Here, we describe a novel integrated genomic method that utilizes single molecule tagging and position specific background correction to push the limit of detection to variants existing in as little as 0.01% of a sample. Initial detection improvements come from the quantitative tracking ability of molecular barcodes that significantly reduce the effect of sequencer and PCR amplification errors. Combined with paired-end sequence collapsing, consensus reads are produced that contain reduced numbers of false variants.

In a similar manner to Capp-Seq (Chaudhuri et al. 2017), we then experimentally derive a background of expected errors for each position within the consensus reads. As we know that sequence context impacts nucleotide stability (Benzer 1961; Gaffney and Keightley 2005; Lercher, Williams, and Hurst 2001; Nachman and Crowell 2000; Hwang and Green 2004), we use this background to correct our consensus reads based on probability density functions created for each assayed nucleotide position. We build on previous methods by then extensively characterizing the background error probabilities that generally occur in our sequencing library preparations. These characterizations were sufficiently comprehensive that in some cases we were able to eliminate all background variants, and only detect known mutations within a sample i.e. reaching 100% specificity.

One recent application of especially sensitive sequencing technologies is assaying and understanding clonal evolution within cancerous tissues (Greaves and Maley 2012). The rarity of somatic mutations, even within the clonally expanding pool of cells that exists within a tumor, has limited the observation of changes that can occur. Such an understanding would be valuable, as cancer therapies often leave behind a small number of cells that can frequently lead to relapse. In leukemias, the state during which these small numbers of cells remain after initial treatment is referred to as minimal residual disease (MRD). During this MRD stage, residual leukemia cells continue to evolve, and successful detection of relapsing leukemia at early stages may

facilitate improved prognosis and treatment strategies (Ivey et al. 2016; Krönke et al. 2011; Chaudhuri et al. 2017).

As a proof of principle, we directly sample leukocyte genomic DNA, and demonstrate the ability of FERMI to detect oncogenic changes during the MRD state, and monitor clonal changes with time. We also show that by concurrently sampling a diverse panel of oncogenic regions, we can detect the expansion of new oncogenic variants during MRD. Such observations could be critically important in predicting relapse in patients, and may significantly impact survival rates.

Results

Method Overview

We devised FERMI (Fast Extremely Rare Mutation Identification) as a method to overcome current sequencing challenges facing rare mutation detection. FERMI is based on Illumina's TrueSeq Custom Amplicon and AmpliSeq Myeloid protocols, which are designed for rare mutation detection. In FERMI, sequences found within human genomic DNA (gDNA) are captured by targeted oligomer probes, which are then sequenced and analyzed for the presence of any existing mutations. We adapted the AmpliSeq process to target a much smaller number of regions of the genome (32 vs 1500 regions) in order to achieve a greater sequencing depth per location with a reduced sequencing cost. The gDNA used for capture and sequencing was purified

either from blood, cancer cell line, or sperm cells, though most of our work focused on peripheral blood cells. The method should be adaptable to any species.

Barcode-guided single molecule sequencing

Capture of gDNA, including any existing variants, begins by incubating double stranded gDNA together with oligomer probes designed to bind specified regions of the genome (Figure 1A). These probes span regions of approximately 150bp in length and contain two identifying indexes. The first index is a 16 bp sequence specific to each sample being processed, and the second is a 12 bp unique molecular identifier (UMI) of randomized DNA that should be unique to each captured strand of gDNA. Double stranded gDNA is melted apart to allow these targeting probes to bind the resulting single stranded DNA. Probe annealing is then achieved by slowly cooling the samples to allow for efficient targeting.

Following hybridization of the probes and gDNA, DNA polymerase is used to copy the template, and DNA ligase joins the strands together into a single contiguous amplicon. Using the sample indexes and the capture-specific UMIs to ensure each capture is tracked, amplicons are amplified by polymerase chain reaction (PCR) and pooled together for sequencing. Samples were sequenced using paired-end 150 bp sequencing, allocating approximately 30 million Illumina HiSeq or NovoSeq reads per sample. This coverage allocates on average about 1,000,000 capture reactions per sample, resulting in about 30X sequencing coverage for each capture (given an

average of 30,000 captures per probed region). Though capture efficiency was not uniform for the different probes, which show a 5-fold range in the numbers of successful unique captures, we show sufficient coverage at each probed location to capture mutations at least as rare as 0.01% (Figure 1B).

Assessment of background error profile

Following sequencing, reads are distributed into sample-specific bins by their sample index. Within these sample-specific bins, paired-end reads are combined into single consensus reads by eliminating all base calls that are mismatched. This approach yielded better results than elimination of pairs with some threshold of mismatches, as it retained substantially more sequencing information. These paired-end consensus reads are then sorted into capture-specific bins by their UMI sequences. These capture-specific bins are then collapsed into final consensus reads. In order to qualify for this final UMI-based consensus derivation, a UMI-specified capture is required to have at least 5 supporting sequencing reads, and the base at each position is only called if 75% of supporting reads agree with its identity (Supplemental Figure 1). The final consensus reads are then compared against an experimentally determined background to distinguish true-positive variants from false positive signal.

Though UMI barcode collapsing of sequencing probes is an important technique by which sequencing sensitivity and accuracy can be increased (Hiatt et al. 2013), we

find that UMI-collapsed data still retain a significant amount of false-positive variant signal. Using leukocyte gDNA purified from putatively healthy blood donors, we find approximately 5000 unique variants within our UMI collapsed consensus sequences in each individual. To estimate how much of this signal might be false-positive background, consensus sequences were computationally binned by the presence or absence of heterozygous SNPs found within our probed individuals. This sorting created bins of sequencing that should have originated from only a single allele. Theoretically, if rare variants were indicative of mutations that existed in-vivo, by their very nature of being rare, the mutations should exhibit an associative bias with only one of the two alleles. When we call variants within these two allele-specific bins however, we find that the variants associate quite uniformly across both alleles, suggesting that much of the variant signal found within our final consensus reads is erroneous (Figure 2A).

Further suggestive of a significant false-positive presence within consensus reads, we show that when the rare variants found within the blood of any two individuals are compared, the same variants are found in each sample at nearly the same allele frequencies (Figure 2B). This similarity is not limited to inter-blood sample comparisons, as sperm and lung gDNA show similar patterns (Supplemental Figure 2). Finally, being similar to that in blood, the somatic mutation load we observe in sperm cells is well above previous estimates of less than 100 mutation per genome (Lynch

2016). Furthermore, when blood from healthy individuals was compared, mutations were no more similar in repeats from the same individual than between individuals (Supplemental Figure 3). Combined, these observations suggest that a false-positive background exists relatively uniformly across samples and sample types, and invites the possibility for a correction algorithm to distinguish real from false signal in order to significantly improve sequencing detection limits.

Within our final consensus reads, single-nucleotide substitutions account for the majority of falsely identified variants, and within this group of variants, there is significant identity bias. We find that C>T substitutions account for nearly 50% of the variants present in our final consensus reads, while other changes like C>G and T>G are far more rare Figure 2C. Breaking down these substitutions into their trinucleotide contexts by including the bases located 5' and 3' of each change, we find that sequence context significantly impacts the probability of a false variant being identified Figure 2D. Among the trinucleotide contexts, false variants within CpG sites are overrepresented within our final consensus reads.

Importantly, the patterns we identify in the trinucleotide context-independent and context-dependent substitutions mirror those identified in other studies of both normal tissues and cancers (Blokzijl et al. 2016; Martincorena et al. 2015; Alexandrov et al. 2013). The similarity of these patterns provides a cautionary note for mutation detection, as obedience to known patterns does not necessarily provide confidence in

the accuracy of calls. It is possible that in-vivo mechanisms of mutation generation are similar to those experienced by template DNA ex-vivo, and therefore results in similar patterns within the background.

Nucleotide context insufficiently explains background signal

In search of universal patterns within our false positive background, we looked for surrounding sequence contexts that play a role in the prevalence of a false variant. While trinucleotide context does impact the probability that a substitution is found within our final consensus read pool, it often incompletely predicts the resulting variant allele frequency (VAF). We observe that many of the background substitutions found within our final consensus reads such as C>A within the contexts of CCA and ACA, exist within two relatively distinct VAF groups (Figure 3A). This indicates that within a given trinucleotide context, a substitution such as C>A will occur with either a high or a low frequency. Alternatively, some substitutions such as C>A within the CCG context largely occur with a low frequency, while other changes such as T>G in the context of CTA never exist at a high frequency. Both results suggest that trinucleotide context is not sufficient to predict background substitution rates at a given locus (Supplemental Table 1), consistent with recent reports that broader (epi)genomic contexts play key roles in replication errors, DNA damage, and repair (Coleman and De 2018). We do

find, that regardless of the substitution identity or the trinucleotide context, a substitution never exclusively occurs at high frequency.

If the background variants are separated by their presence in either the upper or the lower VAF population, we find that for some changes such as C>A, both the 5' and the 3' nucleotides of the trinucleotide context significantly impact the VAF of the change Figure 3B. This impact of the trinucleotide context is however not present for all changes, such as C>T substitutions, which show minimal bias of any of the possible trinucleotide contexts. We further searched for deterministic patterns within the 10bp upstream and downstream of a given change and find that only the triplet context showed any meaningful impact on mutation rate (Supplemental Figure 5).

Algorithmic background subtraction eliminates most false positive signal

Although the trinucleotide context alone does not provide a sufficient amount of contextual information to determine the frequency with which a background variant is observed, nucleotide position strongly impacts the VAF of a substitution within the final consensus sequences. Throughout the probed regions, each nucleotide locus shows a unique background signal pattern that is relatively conserved across individuals (Figure 4A). Some of the mutational patterns we observe within our background signal are similar to those found in The Cancer Genome Atlas (TCGA) (Supplemental Figure 6).

Notably, enrichment for previously defined signatures are evident in background variants, representing artifacts of damage to isolated gDNA.

Within our observed backgrounds, some nucleotide loci exhibit a strong bias towards a particular base change, showing only one type of substitution across all tested individuals. This effect is most commonly observed at nucleotides that exhibit a C>T substitution, where it is often the only observed change at that locus. Other nucleotide positions exhibit multiple different background substitutions, some changing to all three possible other bases. We noticed that while the background signal was surprisingly conserved across samples, that variability did exist (Figure 4B). It is often the case that a particular nucleotide locus will exhibit the same types of variants across different samples, but the allele frequencies vary.

Because each nucleotide locus tends to show a similar background across all tested individuals (Supplemental Figure 4), it was possible to derive a governing probability distribution for each observed substitution at every probed position. To create this distribution, a probability density function was created using a student's T continuous random variable function. This probability density function was then used to calculate the high and low VAF endpoints of a confidence range by using a specified alpha fraction of the distribution.

We compared a number of different types of backgrounds to understand which best allowed us to eliminate false variants (Figure 4C). Initially a generic background

was created by deriving a probability density function for each type of nucleotide change based on its neighbors (the three nucleotide patterns shown in Figure 2D), independent of genomic position. This was somewhat effective at eliminating background variants from samples, as it reduced the total variant calls by about 80 percent. By incorporating positional information, and deriving a density function for each observed substitution at all genomic loci, about 99.9 percent of variant calls were eliminated, providing very clean sequencing data. Importantly, experimental variability seems to play a significant role in the accuracy of the background. While a background derived from samples taken from a different experiment will result in about 10 variants being called as real in a given sample, a background created from samples in the same experiment will result in about 1-3 variants being called as real (Figure 4C). As expected variants are always retained, but total number of significant variants is minimized when using an internal background created from samples of the same experiment, internal backgrounds are used for all analysis.

To understand what alpha fraction of the probability density functions should be called as bonafide mutations, we used 10 healthy blood samples to derive a confidence interval range for each observed substitution across all probed nucleotide positions. These confidence intervals were then compiled into a comprehensive false-positive background against which experimental samples were then compared.

Using 5 healthy blood samples, variants were called from their derived consensus sequences, and then compared against the comprehensive background. Within these 5 samples, variants were called as real if their variant allele frequencies were high enough to fall within the specific alpha of their governing probability density function. As expected, as the confidence interval was increased from 0.9 (One 9) to 0.999999999999999 (Fifteen 9's) the number of variants called as real, exponentially decreases (Figure 4D). This method eliminates nearly all background signal by confidence intervals in the range of ten 9's. Furthermore, at higher confidence intervals even germline variants are often eliminated for being too close to background, indicating excessive stringency. Peripheral blood taken from leukemic patients at different points during therapy were also tested, and show similar exponential decreases in confident variant calls, though the overall numbers of variants are higher than in healthy blood (Figure 4E).

Assessing FERMI sensitivity and specificity

To help understand the specificity of FERMI in detecting only true mutations, Molm13 cells were expanded in-vitro after passing them through bottlenecks of 1, 100, or 1 million cells. We show that many mutations can be observed within the cell cultures started from 100 cells, given that the 100 cell bottleneck should create clones at approximately 1% frequency each with occasional variants in our probed region (Figure 5A). Heterozygous mutations are expected at allele frequencies of 0.005 at 2N

loci, and at lower VAFs if a variant falls in a region with greater ploidy. Indeed, most variants fall within this range. As expected, very few mutations are detected in gDNA isolated from the cell cultures started from 1 million cells, as most mutations will exist at rare allele frequencies (below our limit of detection). Similarly, we observed no mutations within the cultures initiated with single cells, consistent with the low odds that a mutation would occur in our probed region during the ~14 cell divisions required to generate the 10,000 cell limit of detection. The absence of mutations detected in the cultures initiated with 1 cell each, but the presence of mutations within the cultures initiated with 100 cells, indicates that we have sufficiently limited false positive variants, but retained true mutations.

To assess the sensitivity and limit of detection of FERMI, gDNA from human blood containing known heterozygous SNPs was serially diluted into blood lacking these SNPs. We find that the detection limits of tracking single dilutions to be somewhat variable as the level of background noise is position specific, but diluted germline variants were detected at frequencies at least as rare as 1:10,000 at the expected VAFs Figure 5B. In some cases, where the background is exceptionally high, a dilution series would not be detected as low as 1:10,000.

One of the samples used in the dilution series testing contained three heterozygous SNPs within the same probe region, on the same allele, allowing for an extra level of error correction. Within this sample, it was assumed that only those

consensus reads with all three SNPs or those without any of the SNPs were correct, and all other reads were eliminated. This analysis significantly reduced the background error rates at the SNP positions, and allowed detection of diluted mutations at least as low as 1:10,000 Figure 5C.

Oncogenic driver detection in leukemias

To test the ability of FERMI to detect and follow mutations throughout leukemia treatments, patient biopsies were collected at disease inception, throughout treatment, and during relapse when possible. Using FERMI, we are able to detect these mutations when they are present and observe clonal evolution as it occurs. In some cases we detect the principle oncogenic driver and watch it fluctuate in frequency in response to treatment without ever disappearing below background (Figure 6A). In one case, a JAK2 mutation is initially observed at high frequency, but treatment eliminates the clone. As relapse occurs, blast counts increase, but the initial JAK2 clone does not increase in frequency, as the genetics of the leukemia has clearly changed with treatment (Figure 6B). In a third sample, we observe a patient relapse with a previously undetected JAK2 mutation (Figure 6C). While this time point was taken at relapse, we detect it at a frequency significantly below that of most sequencing method sensitivities, and requiring only 5 ml of peripheral blood. The early detection of such a clone could allow treatment with a kinase inhibitor before overt disease relapse.

Discussion ~750 words

In this study, we designed a sensitive sequencing method that enables the accurate detection of rare variants and clonal evolution within cancer samples. In leveraging the quantitative power of capture-unique UMI barcodes, we achieve single-allele sequencing resolution from gDNA, and by then combining these sequencing results with a comprehensive analysis of expected background signal, we achieve an exceptional sequencing fidelity.

While capture-specific barcoding has been effectively used in the past, an inability to achieve sufficient capture numbers and high background have often held the theoretical limit of detection to variants existing at a frequency of >1/1,000. By probing about 50,000 different alleles for each region of interest per sample, we pushed our theoretical limit of detection to at least 1/10,000 (at least for positions where background signal is sufficiently low).

Paired-end collapsing has been successfully used to reduce the number of sequencing errors within sequencing data. Unfortunately, errors also occur during library preparation, and while the molecular barcodes assist with the elimination of these library preparation errors, mistakes made before or during the first round of PCR will typically appear indistinguishable from a heterozygous variant, such that neither paired-end collapsing or molecular barcode collapsing will be capable of eliminating them. This understanding prompted the development of an expected false positive

background that could be used to filter out common mistakes that occur during library preparation.

Our experimentally derived backgrounds proved vitally important in determining whether or not a mutation found in a sample was sufficiently elevated in allele frequency that it could be classified as truly existing in the in-vivo gDNA from which it originated. The idea of creating a background to correct for expected library preparation errors has been used in the past (Chaudhuri et al. 2017), but the correction algorithm designed for FERMI is particularly scalable, and maintains its accuracy regardless of the number of examined sites. This generalized background allowed us to not only detect expected mutations, but also discover new mutations within samples.

It is interesting to note that within our correction background, we observe similar mutation patterns to those observed for other studies, and even similar mutational signatures (Behjati et al. 2014; Alexandrov et al. 2013). This may indicate a surprising degree of similarity between the intrinsic mutagenic processes in-vivo, and error-causing processes involved in sample preparation. It is possible that this similarity is the result of the conserved behavior of error-inducing machinery like DNA polymerase and DNA ligase both in-vivo and in-vitro, or even similar mutagenic exposures such as oxidative damage. These observed similarities suggest caution

against mutation signature use as validation of the accuracy of sequencing data when attempting to identify rare variants.

The early detection of clonal evolution within cancer samples has held the promise of more comprehensive diagnoses and improved treatment strategies for patients. While deep sequencing has been applied to patient leukemias in the past, mutation discovery accuracies have typically limited these approaches to more of a validation role (Thol et al. 2018). We obtained a number of leukemic patient biopsies, taken at initial clinical presentation, and throughout treatment, and used FERMI to search for somatic mutations. We find that while the detection limit of FERMI is better than previous sequencing technologies, the greatest improvements are made through its mutation detection ability. Because we eliminate nearly all background variants, we can accurately detect unexpected relapse mutations and drivers of clonal expansions.

In requiring only around 5ml of blood, FERMI could be easily used in a clinical setting to quickly and easily identify important driver mutations and clonal evolution within patient's cancers. If relapse mutations were caught by FERMI when they are still rare, targeted therapies could be used to prevent them from clonally expanding to fixation and driving leukemic relapse.

Alexandrov, Ludmil B., Serena Nik-Zainal, David C. Wedge, Samuel A. J. R. Aparicio, Sam Behjati, Andrew V. Biankin, Graham R. Bignell, et al. 2013. “Signatures of Mutational Processes in Human Cancer.” *Nature* 500 (7463): 415–21.

Behjati, Sam, Meritxell Huch, Ruben van Boxtel, Wouter Karthaus, David C. Wedge, Asif U. Tamuri, Iñigo Martincorena, et al. 2014. “Genome Sequencing of Normal Cells Reveals Developmental Lineages and Mutational Processes.” *Nature* 513 (7518): 422–25.

Benzer, S. 1961. “ON THE TOPOGRAPHY OF THE GENETIC FINE STRUCTURE.” *Proceedings of the National Academy of Sciences of the United States of America* 47 (3): 403–15.

Blokzijl, Francis, Joep de Ligt, Myrthe Jager, Valentina Sasselli, Sophie Roerink, Nobuo Sasaki, Meritxell Huch, et al. 2016. “Tissue-Specific Mutation Accumulation in Human Adult Stem Cells during Life.” *Nature* 538 (7624): 260–64.

Chaudhuri, Aadel A., Jacob J. Chabon, Alexander F. Lovejoy, Aaron M. Newman, Henning Stehr, Tej D. Azad, Michael S. Khodadoust, et al. 2017. “Early Detection of Molecular Residual Disease in Localized Lung Cancer by Circulating Tumor DNA Profiling.” *Cancer Discovery* 7 (12): 1394–1403.

Cheng, Keith C., David S. Cahill, Hiroshi Kasai, Susumu Nishimura, and Lawrence A. Loeb. 1992. “8-Hydroxyguanine, an Abundant Form of Oxidative DNA Damage, Causes G----T and A----C Substitutions.” *The Journal of Biological Chemistry* 267 (1): 166–72.

Coleman, Neil, and Subhajyoti De. 2018. “Mutation Signatures Depend on Epigenomic Contexts.” *Trends in Cancer Research* 4 (10): 659–61.

Dressman, Devin, Hai Yan, Giovanni Traverso, Kenneth W. Kinzler, and Bert Vogelstein. 2003. “Transforming Single DNA Molecules into Fluorescent Magnetic Particles for Detection and Enumeration of Genetic Variations.” *Proceedings of the National Academy of Sciences of the United States of America* 100 (15): 8817–22.

Fujita, Pauline A., Brooke Rhead, Ann S. Zweig, Angie S. Hinrichs, Donna Karolchik, Melissa S. Cline, Mary Goldman, et al. 2010. “The UCSC Genome Browser Database: Update 2011.” *Nucleic Acids Research* 39 (suppl_1): D876–82.

Gaffney, Daniel J., and Peter D. Keightley. 2005. “The Scale of Mutational Variation in the Murid Genome.” *Genome Research* 15 (8): 1086–94.

Garrison, Erik, and Gabor Marth. 2012. “Haplotype-Based Variant Detection from Short-Read Sequencing.” *arXiv [q-bio.GN]*. arXiv. http://arxiv.org/abs/1207.3907.

Greaves, Mel, and Carlo C. Maley. 2012. “Clonal Evolution in Cancer.” *Nature* 481 (7381): 306–13.

Hiatt, Joseph B., Colin C. Pritchard, Stephen J. Salipante, Brian J. O’Roak, and Jay Shendure. 2013. “Single Molecule Molecular Inversion Probes for Targeted, High-Accuracy Detection of Low-Frequency Variation.” *Genome Research* 23 (5): 843–54.

Hindson, Benjamin J., Kevin D. Ness, Donald A. Masquelier, Phillip Belgrader, Nicholas

J. Heredia, Anthony J. Makarewicz, Isaac J. Bright, et al. 2011. “High-Throughput Droplet Digital PCR System for Absolute Quantitation of DNA Copy Number.” *Analytical Chemistry* 83 (22): 8604–10.

Hwang, Dick G., and Phil Green. 2004. “Bayesian Markov Chain Monte Carlo Sequence Analysis Reveals Varying Neutral Substitution Patterns in Mammalian Evolution.” *Proceedings of the National Academy of Sciences of the United States of America* 101 (39): 13994–1.

Ivey, Adam, Robert K. Hills, Michael A. Simpson, Jelena V. Jovanovic, Amanda Gilkes, Angela Grech, Yashma Patel, et al. 2016. “Assessment of Minimal Residual Disease in Standard-Risk AML.” *The New England Journal of Medicine* 374 (5): 422–33.

Kennedy, Scott R., Michael W. Schmitt, Edward J. Fox, Brendan F. Kohrn, Jesse J. Salk, Eun Hyun Ahn, Marc J. Prindle, et al. 2014. “Detecting Ultralow-Frequency Mutations by Duplex Sequencing.” *Nature Protocols* 9 (11): 2586–2606.

Krönke, Jan, Richard F. Schlenk, Kai-Ole Jensen, Florian Tschürtz, Andrea Corbacioglu, Verena I. Gaidzik, Peter Paschka, et al. 2011. “Monitoring of Minimal Residual Disease in NPM1-Mutated Acute Myeloid Leukemia: A Study from the German-Austrian Acute Myeloid Leukemia Study Group.” *Journal of Clinical Oncology: Official Journal of the American Society of Clinical Oncology* 29 (19): 2709–16.

Lander, E. S., L. M. Linton, B. Birren, C. Nusbaum, M. C. Zody, J. Baldwin, K. Devon, et al. 2001. “Initial Sequencing and Analysis of the Human Genome.” *Nature* 409 (6822): 860–921.

Lercher, Martin J., Elizabeth J. B. Williams, and Laurence D. Hurst. 2001. “Local Similarity in Evolutionary Rates Extends over Whole Chromosomes in Human-Rodent and Mouse-Rat Comparisons: Implications for Understanding the Mechanistic Basis of the Male Mutation Bias.” *Molecular Biology and Evolution* 18 (11): 2032–39.

Li, Heng, and Richard Durbin. 2009. “Fast and Accurate Short Read Alignment with Burrows–Wheeler Transform.” *Bioinformatics* 25 (14): 1754–60.

Li, Heng, Bob Handsaker, Alec Wysoker, Tim Fennell, Jue Ruan, Nils Homer, Gabor Marth, Goncalo Abecasis, Richard Durbin, and 1000 Genome Project Data Processing Subgroup. 2009. “The Sequence Alignment/Map Format and SAMtools.” *Bioinformatics* 25 (16): 2078–79.

Li, Jin, Lilin Wang, Harvey Mamon, Matthew H. Kulke, Ross Berbeco, and G. Mike Makrigiorgos. 2008. “Replacing PCR with COLD-PCR Enriches Variant DNA Sequences and Redefines the Sensitivity of Genetic Testing.” *Nature Medicine* 14 (5): 579–84.

Lindahl, T., and B. Karlstrom. 1973. “Heat Induced Depyrimidation of DNA.” *Biochemistry* 25: 5151–54.

Lindahl, T., and B. Nyberg. 1974. “Heat-Induced Deamination of Cytosine Residues in

Deoxyribonucleic Acid." *Biochemistry* 13 (16): 3405–10.
Lynch, Michael. 2016. "Mutation and Human Exceptionalism: Our Future Genetic Load." *Genetics* 202 (3): 869–75.
Mansukhani, Sonia, Louise J. Barber, Dimitrios Kleftogiannis, Sing Yu Moorcraft, Michael Davidson, Andrew Woolston, Paula Zuzanna Proszek, et al. 2018. "Ultra-Sensitive Mutation Detection and Genome-Wide DNA Copy Number Reconstruction by Error-Corrected Circulating Tumor DNA Sequencing." *Clinical Chemistry*, August. https://doi.org/10.1373/clinchem.2018.289629.
Martincorena, Iñigo, Amit Roshan, Moritz Gerstung, Peter Ellis, Peter Van Loo, Stuart McLaren, David C. Wedge, et al. 2015. "Tumor Evolution. High Burden and Pervasive Positive Selection of Somatic Mutations in Normal Human Skin." *Science* 348 (6237): 880–86.
Milbury, Coren A., Mick Correll, John Quackenbush, Renee Rubio, and G. Mike Makrigiorgos. 2012. "COLD-PCR Enrichment of Rare Cancer Mutations prior to Targeted Amplicon Resequencing." *Clinical Chemistry* 58 (3): 580–89.
Nachman, M. W., and S. L. Crowell. 2000. "Estimate of the Mutation Rate per Nucleotide in Humans." *Genetics* 156 (1): 297–304.
Oliphant, T. E. 2007. "Python for Scientific Computing." *Computing in Science Engineering* 9 (3): 10–20.
Onecha, Esther, Maria Linares, Inmaculada Rapado, Yanira Ruiz-Heredia, Pilar Martinez-Sanchez, Teresa Cedena, Marta Pratcorona, et al. 2018. "Novel Deep Targeted Sequencing Method for Minimal Residual Disease Monitoring in Acute Myeloid Leukemia." *Haematologica*, August. https://doi.org/10.3324/haematol.2018.194712.
Preston, Jessica L., Ariel E. Royall, Melissa A. Randel, Kristin L. Sikkink, Patrick C. Phillips, and Eric A. Johnson. 2016. "High-Specificity Detection of Rare Alleles with Paired-End Low Error Sequencing (PELE-Seq)." *BMC Genomics* 17 (June): 464.
Schmitt, Michael W., Edward J. Fox, Marc J. Prindle, Kate S. Reid-Bayliss, Lawrence D. True, Jerald P. Radich, and Lawrence a. Loeb. 2015. "Sequencing Small Genomic Targets with High Efficiency and Extreme Accuracy." *Nature Methods*, no. November 2014: 1–4.
Shibutani, S., M. Takeshita, and A. P. Grollman. 1991. "Insertion of Specific Bases during DNA Synthesis Past the Oxidation-Damaged Base 8-oxodG." *Nature* 349 (6308): 431–34.
Sykes, P. J., S. H. Neoh, M. J. Brisco, E. Hughes, J. Condon, and A. A. Morley. 1992. "Quantitation of Targets for PCR by Use of Limiting Dilution." *BioTechniques* 13 (3): 444–49.
Tan, Adrian, Gonçalo R. Abecasis, and Hyun Min Kang. 2015. "Unified Representation of Genetic Variants." *Bioinformatics* 31 (13): 2202–4.
Terwijn, Monique, Wim L. J. van Putten, Angèle Kelder, Vincent H. J. van der Velden, Rik A. Brooimans, Thomas Pabst, Johan Maertens, et al. 2013. "High Prognostic

Impact of Flow Cytometric Minimal Residual Disease Detection in Acute Myeloid Leukemia: Data from the HOVON/SAKK AML 42A Study." *Journal of Clinical Oncology: Official Journal of the American Society of Clinical Oncology* 31 (31): 3889–97.

Thol, Felicitas, Razif Gabdoulline, Alessandro Liebich, Piroska Klement, Johannes Schiller, Christian Kandziora, Lothar Hambach, et al. 2018. "Measurable Residual Disease (MRD) Monitoring by NGS before Allogeneic Hematopoietic Cell Transplantation in AML." *Blood*, September. https://doi.org/10.1182/blood-2018-02-829911.

Vogelstein, B., and K. W. Kinzler. 1999. "Digital PCR." *Proceedings of the National Academy of Sciences of the United States of America* 96 (16): 9236–41.

Young, Andrew L., Grant A. Challen, Brenda M. Birmann, and Todd E. Druley. 2016. "Clonal Haematopoiesis Harbouring AML-Associated Mutations Is Ubiquitous in Healthy Adults." *Nature Communications* 7 (August): 12484.

Acknowledgments: We would like to thank Ruth Hershberg of Technion University and Jay Hesselberth and Robert Sclafani of the University of Colorado School of Medicine for useful suggestions and for review of the manuscript. We would like to thank Craig Jordan and Amanda Winters for the leukemia treatment samples used in the oncogenic mutation detection experiments. These studies were supported by grants from the National Cancer Institute (R01CA180175 to J.D.), NIH/NCATS Colorado CTSI Grant Number UL1TR001082CU (seed grant to J.D.), a SCOR grant from the Leukemia and Lymphoma Society (to Craig Jordan), F31CA196231 (to L.A.L.), the Linda Crnic Institute for Down Syndrome (to J.D. and L.A.L.), and P30-CA072720 (to A.S. and S.D.). The research utilized services of the Cancer Center Genomics Shared Resource, which is supported in part by NIH grant P30-CA46934. L.A.L. and J.D. developed the concept of this project, planned the experiments, analyzed results, and wrote the manuscript. L.A.L. processed and prepared samples from blood biopsy to sequencing, and wrote the software used for analysis. A.S. and S.D. analyzed results, and contributed to writing of the manuscript.

Figures

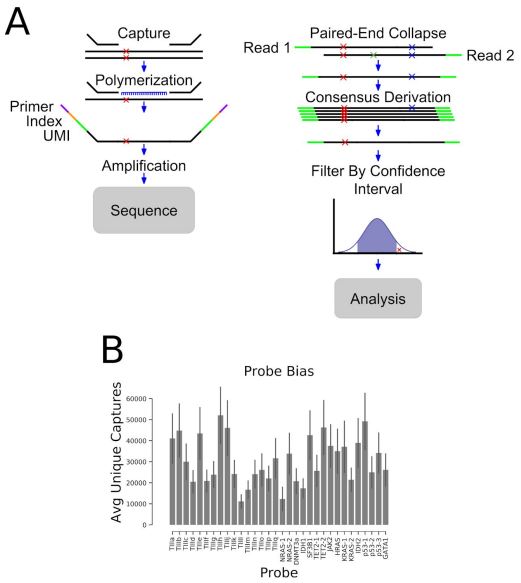


Figure 1

Method overview. A) Schematic representing the steps involved in identifying mutations with FERMI. B) The average number of unique captures varies by probe location. Error is standard deviation across 20 samples.

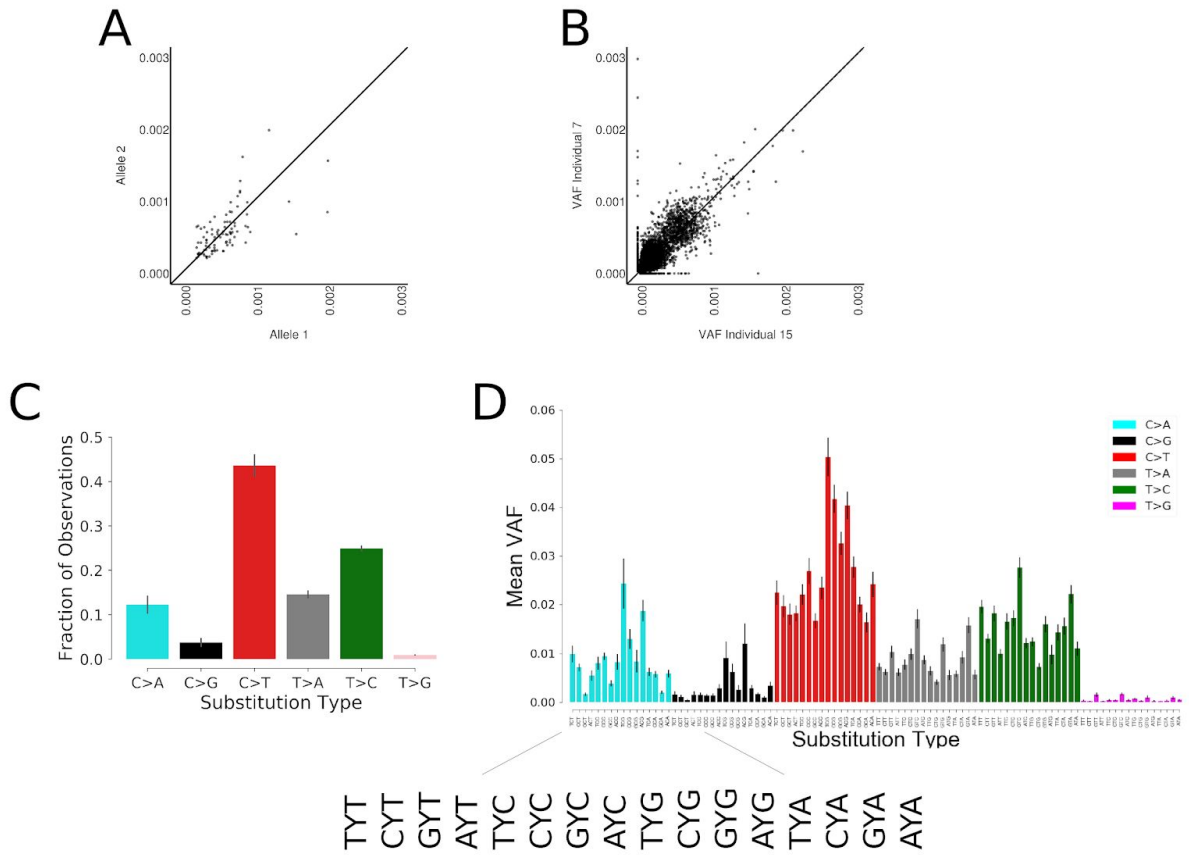


Figure 2

The background of false positive variants is similar across individuals. A) Using heterozygous SNPs to identify different alleles in human leukocyte gDNA, rare variants within consensus reads equally associate each of the alleles suggesting they are occurring randomly ex-vivo. B) Allele frequencies of background signal is similar within leukocyte gDNA of two individuals. C) Relative prevalence of observed substitutions within background signal found in leukocyte gDNA consensus reads. Complementary changes such as C>T and G>A are combined. Error is standard deviation across 20

individuals. D) Relative prevalence of observed substitutions classified by the neighboring upstream and downstream nucleotides. Error is standard deviation across 20 individuals.

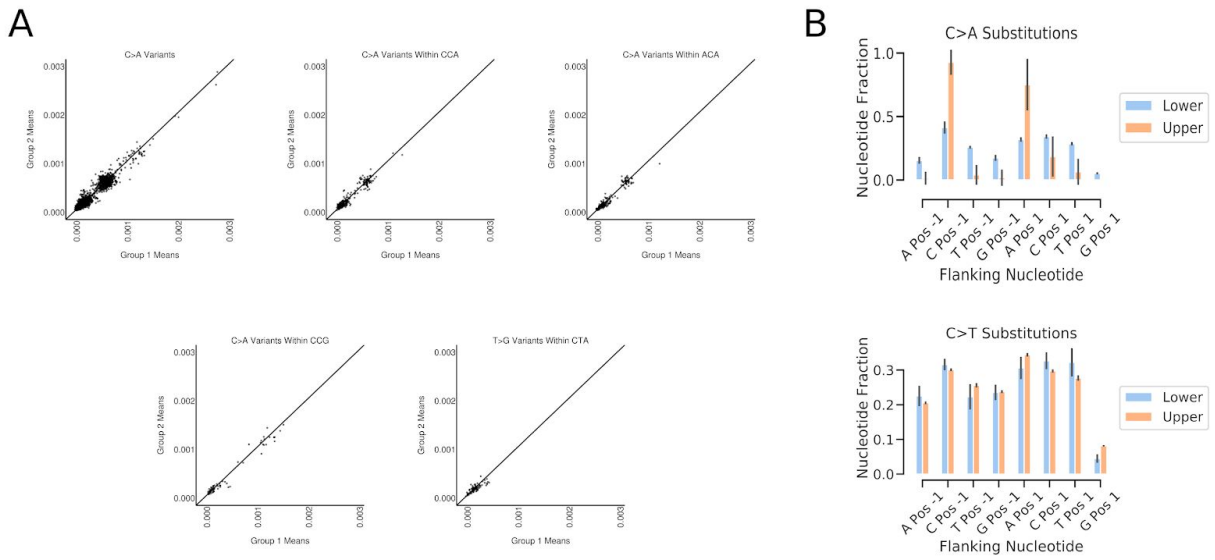


Figure 3

Trinucleotide context is insufficient to predict VAF. A) Mean VAF of background variants calculated from two different groups of 10 individuals. Variants are either classified by substitution identity alone or within a particular trinucleotide context. B) Relative substitution rates for different substitutions classified by triplet context. Error is standard deviation across 20 individuals.

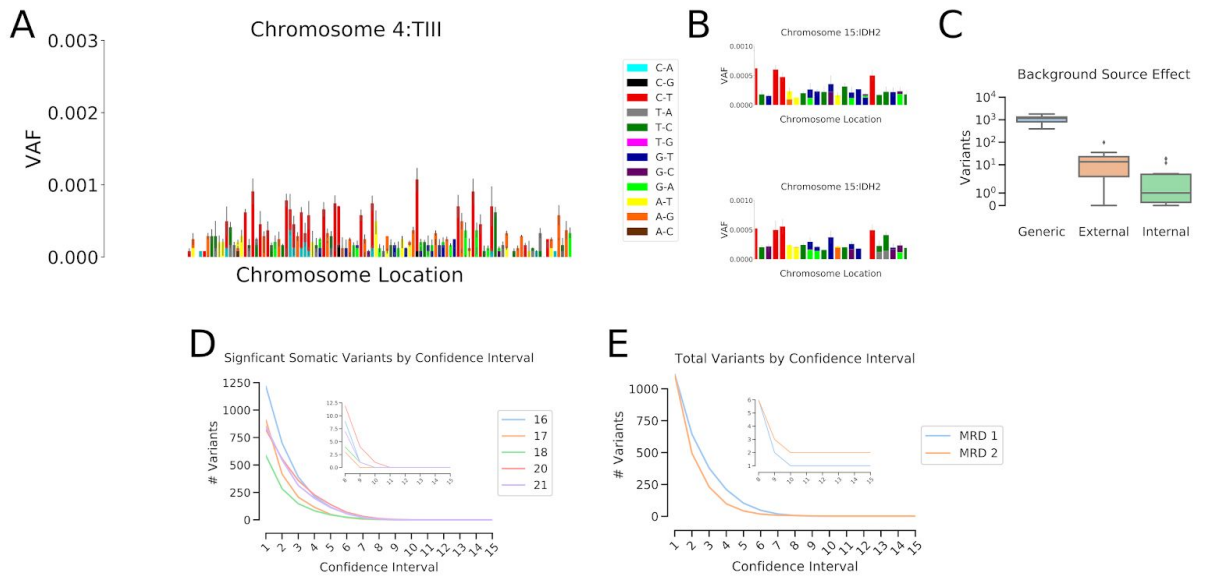


Figure 4

Background confidence intervals eliminate most variants. A) Observed substitution VAFs for a TIII probe, illustrating the varying mutation presence at each nucleotide locus. Error is standard deviation across 20 individuals. B) Subsets of the IDH2 probe region from two different groups of 10 individuals illustrates the degree of similarity and differences that are commonly observed between samples. Error is standard deviation across the individuals. C) Total number of variants deemed significantly above background when only triplet context was used to generate expected background substitution rates (Generic), or when position-specific substitution rates generated from a different experiment (External) or the same experiment (Internal) are used. D) The total numbers of variants called as significantly above background for 5 individuals at confidence intervals from 0.9 - 0.999999999999999. E) The total numbers of variants

called as significantly above background for two leukemic patients using sample 4 from MRD1 and sample 3 from MRD2 (See Figure 6), which are both points that had followed treatment, and at which leukemic burden was low.

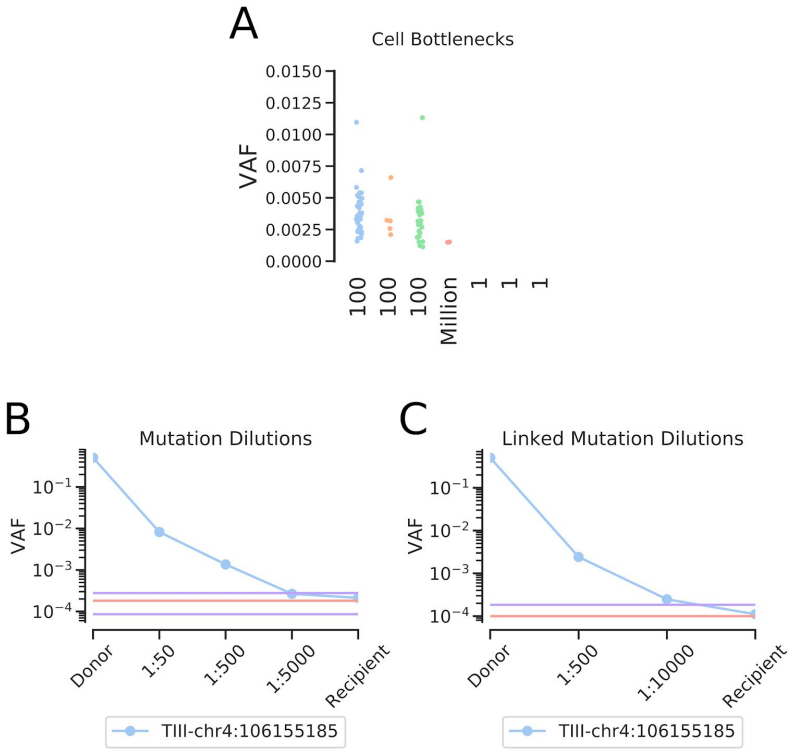


Figure 5

Assessing FERMI specificity and accuracy. A) MOLM-13 cells grown from 1, 100, or 1 million initial cells were expanded to a pool of 1 million cells, and then probed with FERMI. Samples illustrate how clonality impacts mutation detection by FERMI by altering the VAFs of somatic variants. B) Observed frequencies of serially diluted mutation show successful detection at allele frequencies from 1/2 to 1/10,000. Background signal mean and standard deviation shown in red and purple respectively. C) Limit of detection improvements observed when multiple mutations in linkage disequilibrium are leveraged to eliminate erroneous reads.

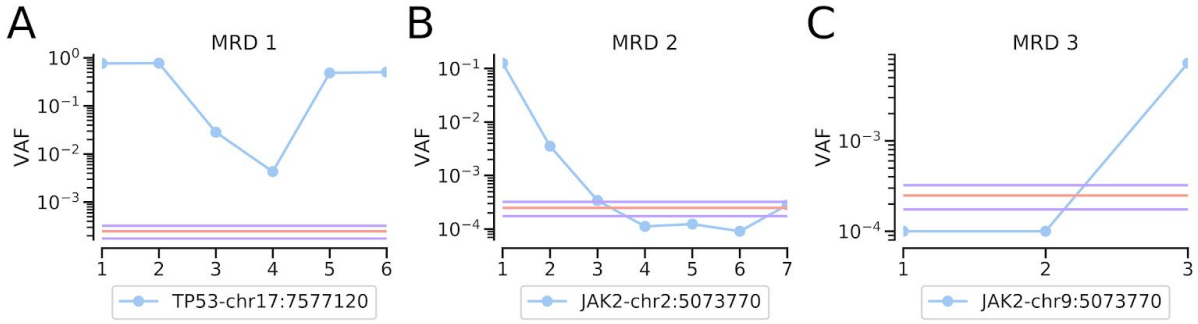


Figure 6

Oncogenic mutation detection during leukemia treatments A) Oncogenic driver detection using FERMI on leukemic liquid biopsies taken at 6 different timepoints starting with clinical presentation and ending with relapse (x-axis is in chronological order of leukemia samplings). B) Oncogenic driver detection throughout leukemia treatment in a case where relapse was not driven by a mutation within our panel. C) Example of a leukemic relapse in which the detected driver was not unknown prior to blood sampling.

Supplementary Materials:

Materials and Methods

Supplementary Text

Supplementary Figures 1-10

Table S1

References (36-44)

Materials and Methods:

Amplicon Design

Amplicon probes for targeted annealing regions were created using the Illumina Custom Amplicon DesignStudio (https://designstudio.illumina.com/). UMIs were then added to the designed probe regions and generated by IDT using machine mixing for the randomized DNA. Probes were PAGE purified by IDT. All probes are listed below along with binding locations and expected lengths of captured sequence.

Gene	Probe Up	Probe Down	Probe Start	Probe End	Length
JAK2	AGTTTACACTGACAC CTAGCTGTGATC	CCATAATTTAAAACC AAATGCTTGTGAGAA	chr9:5073733	chr9:5073887	155
TP53-1	TCATCTTGGGCCTGT GTTATCTCCTA	ATCCTCACCATCATC ACACTGGAAGAC	chr17:7577504	chr17:7577635	132
TP53-2	CCCTCAACAAGATG TTTTGCCAACTG	ATGAGCGCTGCTCA GATAGCGATGGT	chr17:7578369	chr17:7578544	176
TP53-3	GGACAGGTAGGACC TGATTTCCTTACT	TGTCCTGGGAGAGA CCGGCGCACAGA	chr17:7577084	chr17:7577214	131
NRAS-1	CAATAGCATTGCATT CCCTGTGGTTTT	GTACAGTGCCATGA GAGACCAATACAT	chr1:115256496	chr1:115256680	185
NRAS-2	GAAGGTCACACTAG GGTTTTCATTTCC	AAAAGCGCACTGAC AATCCAGCTA	chr1:115258713	chr1:115258897	185
HRAS	TCCTTGGCAGGTGG GGCAGGAGACCC	GCAAGAGTGCGCTG ACCATCCA	chr11:534258	chr1:534385	128
KRAS-1	AGGTACTGGTGGAG TATTTGATAGTGT	CAAGAGTGCCTTGA CGATACAGCTAATT	chr12:25398247	chr12:25398415	169
KRAS-2	GACTGTGTTTCTCCC TTCTCAGGATTC	TACAGTGCAATGAG GGACCAGTACATG	chr12:25380242	chr12:25380368	127
TET2-1	CCATGTTTTGGCTCA TTCATGCTCTTA	ACGGCCACTCCCCC AATGTCAG	chr4:106197237	chr4:106197405	169
TET2-2	CTTTTGAAAGAGTGC CACTTGGTGTCT	GGTGATGGTATCAG GAATGGACTTAGTC	chr4:106155137	chr4:106155275	139
DNMT3A	TGTGTGGTTAGACG GCTTCCGGGCA	AGGCAGAGACTGCT GGGCCGGTCA	chr2:25457211	chr2:25457364	154
IDH1	CAAATGTGGAAATCA CCAAATGGCACC	TGGGGATCAAGTAA GTCATGTTGGCA	chr2:209113077	chr2:209113239	163
IDH2	GAAGAAGATGTGGA AAAGTCCCAATGG	CATGGCGACCAGGT AGGCCAGG	chr15:90631809	chr15:90631969	161

GATA1	CTTCCAGCCATTTCT GAGATATCCTCA	CAGCTGCAGCGGTG GCTGTGCT	chrX:48649667	chrX:48649849	183
SF3B1	GTGAACATATTCTGC AGTTTGGCTGAA	ACCATCAGTGCTTTG GCCATTGC	chr2:198266803	chr2:198266967	165
TIIIA	CATCTATTCTGTGCT AGGCATTGTGTG	CAGACCTAGCATCT GTGCCAGAC	chr1:115227814	chr1:115227978	165
TIIIB	CAGTCTGGGTTTTG GAGCAATGATATC	GCAGTGAGCTCAGC CTTGATTTT	chr2:223190674	chr2:223190820	147
TIIIC	CCTGGTGCTTAGTC CTGTTCTGAAATT	AGTCTTCTATAATGC CACAACCTGTAT	chr2:229041101	chr2:229041289	189
TIIID	GAACAGAACACTTG GTAGTTGACCATG	AGACAGGGAACTGG CATGAAGAGTTT	chr4:110541172	chr4:110541302	131
TIIIE	GCCTAGAACAGGCA CCATACATTCAAT	AGATGGTGTTGCTGT GCCGGATAGGAG	chr4:112997214	chr4:112997386	173
TIIIF	TGGCACTATGTGGA GATGTTAGTACAG	GGATGTTGGTGCTAT CAGTAGCCATA	chr4:121167756	chr4:121167884	129
TIIIG	CTCTAGGCTTAGTG GTCAAGGAATGAA	AGAAGCAGGACTGT GCTTCCAAACAA	chr4:123547743	chr4:123547901	159
TIIIH	CTTGGTGGTAGCCT AGGCAGTAATTAA	CACGTGGTTGGGAA GAGAAAGTG	chr4:124428637	chr4:124428767	131
TIIIJ	TTCTATAGCACTGGT GACCAGGACACT	CTGGCCACAGTGCC TGGTTTCC	chr11:2126256	chr11:2126420	165
TIIIK	AGACAGGAGGAAGG AGCAATTCAGAAG	CATGGAGATCTCGT CCCCTCAGA	chr11:2389983	chr11:2390171	189
TIIIL	TAGGCCAGAAAACA CACAGTGTCGGG	AACTCCGGTAAGTG GCGGGTGGGGGT	chr11:2593889	chr11:2594074	186
TIIIM	ATCTGGGAACAGAC CTTCCTCAGGCAT	GTTCTAAGTTACTCT GTGTACTTGACT	chr11:11486596	chr11:11486728	133
TIIIN	AGCCTAGTTACCATA GACGGATTCAAC	GAATATCTTCTAACT GGACTTAGAAAACC	chr15:92527052	chr15:92527176	125
TIIIO	CCAACATGTTCTAAA	TGGGTCTCAGCCAT	chr16:73379656	chr16:73379832	177

	TTCTGGCCACAG	CCCATTACTG			
TIIIP	CTAACATCTCACTTC TACCCTACGCTA	TAAGTGCCCACTAC CCCATCCTTAAT	chr16:82455026	chr16:82455164	139
TIIIQ	TCATGACCCAGGCC TCCCAGAACTGAG	ATCTGTGAAGCCGG AGTGAAAACAAC	chr16:85949137	chr16:85949299	163

Genomic DNA Isolation

Human blood samples were purchased from the Bonfils Blood Center Headquarters of Denver Colorado. Our use of these deidentified samples was determined to be “Not Human Subjects” by our Institutional Review Board. Biopsies were collected as unfractionated whole blood from apparently healthy donors, though samples were not tested for infection. Samples were approximately 10 mL in volume, and collected in BD Vacutainer spray-coated EDTA tubes. Following collection, samples were stored at 4°C until processing, which occurred within 5 hours of donation. To remove plasma from the blood, samples were put in 50 mL conical tubes (Corning #430828) and centrifuged for 10 minutes at 515 rcf. Following centrifugation, plasma was aspirated and 200 mL of 4°C hemolytic buffer (8.3g NH_4Cl, 1.0g $NaHCO_3$, 0.04 Na_2 in 1L ddH_2O) was added to the samples and incubated at 4°C for 10 minutes. Hemolyzed cells were centrifuged at 515 rcf for 10 minutes, supernatant was aspirated, and pellet was washed with 200 mL of 4°C PBS. Washed cells were centrifuged for at 515rcf for 10 minutes, from which gDNA was extracted using a DNeasy Blood & Tissue Kit (Qiagen REF 69504).

Amplicon Capture

For amplicon capture from gDNA, we modified the Illumina protocol called "Preparing Libraries for Sequencing on the MiSeq" (Illumina Part #15039740 Revision D). DNA was quantified with a NanoDrop 2000c (ThermoFisher Catalog #ND-2000C). 500ng of input DNA in 15µl was used for each reaction instead of the recommended quantities. In place of 5µl of Illumina 'CAT' amplicons, 5µl of 4500ng/µl of our amplicons were used. During the hybridization reaction, after gDNA and amplicon reaction mixture was prepared, sealed, and centrifuged as instructed, gDNA was melted for 10 minutes at 95°C in a heat block (SciGene Hybex Microsample Incubator Catalog #1057-30-O). Heat block temperature was then set to 60°C, allowed to passively cool from 95°C and incubated for 24hr. Following incubation, the heat block was set to 40°C and allowed to passively cool for 1hr. The extension-ligation reaction was prepared using 90 µl of ELM4 master mix per sample and incubated at 37°C for 24hr. PCR amplification was performed at recommended temperatures and times for 29 cycles. Successful amplification was confirmed immediately following PCR amplification using a Bioanalyzer (Agilent Genomics 2200 Tapestation Catalog #G2964-90002, High Sensitivity D1000 ScreenTape Catalog #5067-5584, High Sensitivity D1000 Reagents Catalog #5067-5585). PCR cleanup was then performed as described in Illumina's protocol using 45 µl of AMPure XP beads. Libraries were then

normalized for sequencing using the Illumina KapaBiosystems qPCR kit (KapaBiosystems Reference # 07960336001).

Sequencing

Prepared libraries were pooled at a concentration of 5 nM. Libraries were sequenced on the Illumina HiSeq 4000 at a density of 12 samples per lane with 5% PhiX DNA included, or on the Illumina NovaSeq 6000, allocating approximately 30 million reads per sample.

Bioinformatics

The analysis pipeline used to process sequencing results can be found under FERMI here: http://software.laliggett.com/ or here: https://github.com/liggettla/FERMI. For a detailed understanding of each function provided by the analysis pipeline, please refer directly to the software. The overall goal of the software built for this project is to analyze amplicon captured DNA that is tagged with equal length UMIs on the 5' and 3' ends of captures, and has been paired-end sequenced using dual indexes. Input fastq files are either automatically or manually combined with their paired-end sequencing partners into a single fastq file. Paired reads are combined by eliminating any base that does not match between Read1 and Read2, and concatenating this consensus read with the 5' and 3' UMIs. A barcode is then created for each consensus read from the 5'

and 3' UMIs and the first five bases at the 5' end of the consensus. All consensus sequences are then binned together by their unique barcodes. The threshold for barcode mismatch can be specified when running the software, and for all data shown in this manuscript one mismatched base was allowed for a sequence to still count as the same barcode. Bins are then collapsed into a single consensus read by first removing the 5' and 3' UMIs. Following UMI removal, consensus sequences are derived by incorporating the most commonly observed nucleotide at each position, so long as the same nucleotide is observed in at least a specified percent of supporting reads (75% of reads was used for results in this manuscript) and there are least some minimum number of reads supporting a capture (5 supporting reads was used for results in this manuscript). Any nucleotide that does not meet the minimum threshold for read support is not added to the consensus read, and alignment is attempted with an unknown base at that position. From this set of consensus reads, experimental quality measurements are made, such as total captures, total sequencing reads, average capture coverage, and estimated error rates. Typically we required 5 total captures of a variant to be observed for the variant to be counted as real.

Derived consensus reads are then aligned to the specified reference genome using Burrows-Wheeler (H. Li and Durbin 2009), and indexed using SAMtools (H. Li et al. 2009). For this manuscript consensus reads were aligned to the human reference genome hg19 (Lander et al. 2001; Fujita et al. 2010) (though the software should be

compatible with other reference genomes). Sequencing alignments are then used to call variants using the Bayesian haplotype-based variant detector, FreeBayes (Garrison and Marth 2012). Identified variants are then decomposed and block decomposed using the variant toolset vt (Tan, Abecasis, and Kang 2015). Variants are then filtered to eliminate any that have been identified outside of probed genomic regions. If necessary, variants can also be eliminated if below certain coverage or observation thresholds such that variants must be independently observed multiple times in different captures to be included.

The final variants called from the consensus sequences were then compared to experimentally derived confidence intervals for each probed position. These confidence intervals were created by using FERMI to sequence control peripheral blood samples from the same experiment as test samples. Following the logic described in Results, it was assumed that low frequency variants that are detected across multiple individuals (including in blood and sperm, where few variants are expected) were not real signal but rather false positive background. All of the variants from these control samples were thus used to construct a standard background. This background was calculated for each position at which a variant was observed within the standard control samples, and was uniquely calculated for each type of change. A student's T continuous random variable function was used to create a probability density function that describes the background distribution for each substitution type

at every probed locus (Oliphant 2007). By specifying a particular alpha fraction of the distribution, high and low VAF endpoints were derived that were then used to determine if an experimental signal was significantly above background.

Elimination of false positive sequencing and library creation artifacts

A number of steps have been included within sample preparation and bioinformatics analysis specifically to distinguish between true positive signal and false positive signal. Using the dilution series shown in Figures 1C-D, we can show sufficient sensitivity to identify signal diluted to levels as rare as 10^{-4}. While these dilutions show significantly improved sensitivity over many current sequencing methods, they do not address our background error rate. Unfortunately, because both endogenous and exogenous DNA synthesis is error prone, it is challenging to find negative controls that can be used to estimate background error rates with a method of mutation detection as sensitive as FERMI. Nevertheless, we have a number of steps that should eliminate most sources of false signal. The two largest sources of erroneous mutation when sequencing DNA will typically be from PCR amplification mutations (caused both by polymerase errors and exogenous insults like oxidative damage), and sequencing errors.

The steps are the following:

- *Elimination of first round PCR amplification errors*
- *Elimination of subsequent PCR amplification errors*
- *Elimination of sequencing errors*

Elimination of first round PCR amplification errors in consensus reads

The first round of PCR amplification performed during library preparation causes mutations that are challenging to distinguish from those that occurred endogenously. Since there is little difference between those mutations that occur during the first round of PCR amplification and those that occurred endogenously, we rely on probability to eliminate these errors. Since we are performing sequencing of individually captured alleles, we can ask whether requiring that a mutation be observed in multiple captured alleles before it is called as a true positive signal alters the frequency of variants identified. We expect about 400 first round PCR amplification errors, and the probability that the identical mutation will occur in multiple cells becomes exponentially unlikely (Supplementary Figure 1). By requiring a mutation be observed in just three captures before it is called as real signal, only about 1-2 first round PCR amplification errors should make it into the final consensus reads. In contrast, when we process our data requiring from 1 to 5 independent observations of a mutation, the overall mutation spectrum does not change, apart from a loss of the most rarely observed variants. This observation led us to include all variants that were observed even once.

Elimination of subsequent PCR amplification errors

Elimination of PCR amplification errors after the first round of PCR is done using UMI collapsing (Figure 1A). Each time a strand is amplified, the UMI will keep track of its identity. Any mutations that occur after the first round of PCR will be found on average in 25% of the reads (or fewer for subsequent rounds). This allows us to collapse each unique capture and eliminate any rarely observed variants (<75%) associated with a given UMI. Utilizing the UMI in this way allows us to essentially eliminate any PCR amplification errors that occurred after the first round of PCR. The method should also eliminate most errors resulting from DNA oxidation in vitro.

Elimination of sequencing errors

Sequencing errors are eliminated in two ways. This first method is by using paired-end sequencing to read each strand of a DNA fragment (Figure 1A). The sequence of these reads (Read1 and Read2) should match if no sequencing errors have been made. For an error to escape elimination it would need to occur at the same position (changing to the same new base) within both Read1 and Read2. Therefore, when the base call differs at a position on Reads 1 and 2, these changes are eliminated from the final sequence. This collapsing should eliminate most sequencing errors, although sequencing errors of the same identity occurring at the same position will

escape. These errors should be removed when collapsing into single capture bins (Figure 1A). As with the logic when eliminating subsequent PCR amplification errors, most sequences associated with each UMI pair should be identical. Therefore, sequencing errors passing through Read1 and Read2 will be very unlikely to match other sequenced strands from the same capture event, and are eliminated during consensus sequence derivation.

Derivation of Standard Background

For sequenced samples , confidence intervals above twelve 9's tend to be stringent enough that heterozygous and homozygous SNPs are not significantly elevated above background. As the confidence interval of ten 9's is not so strict as to eliminate germline SNPs, but most non-germline variants are eliminated (including for samples such as sperm DNA), this was used as the standard by which we called variants as real mutations. It is important to note however, that this confidence could be easily altered to target a useful confidence interval for any different uses of FERMI.

Supplemental Figures:

a

Supporting Captures	Duplex	Mock-Duplex	% Vars Eliminated
4	4240	4264	0.56285
3	4912	4928	0.32468
2	5704	5734	0.52319
1	6760	6794	0.50044

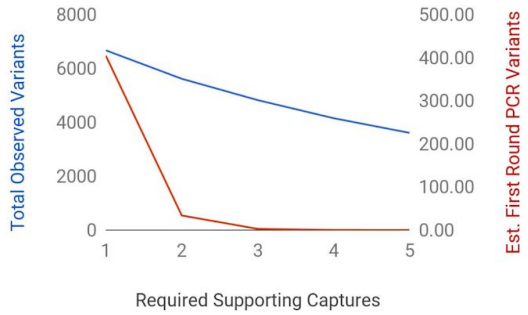
b

Enzyme	Error Rate (mut/base)	Unique UMIs	Captures per UMI	Total Amplicon Size	# Bases In First Amplification	Total Errors
Phusion HF Buffer	0.00000044	2818388	88075	4838	426105036	187
Phusion GC Buffer	0.00000095	2818388	88075	4838	426105036	405

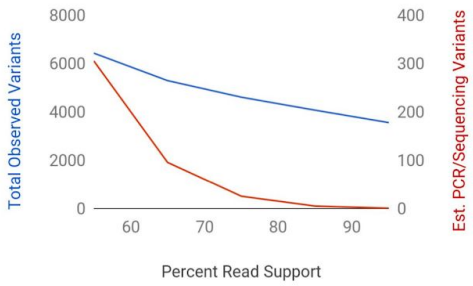
c

Supporting Captures	1	2	3	4	5
	187.49	7.27	0.28	0.01	0.00
	404.80	33.87	2.83	0.24	0.02

d



e



Supplemental Figure 1

Estimation of false-positive rates due to sequencing and PCR errors.

a, The use of sequencing information found within Read 1 and Read 2 of paired-end sequencing is often used to correct sequencing errors. We performed paired-end collapsing prior to consensus read derivation (Figure 1A), though the effect was surprisingly mild. In this table, the number of identified variants are shown when duplex collapsing is used or not in consensus read derivation (mock duplexing processes the collapsing in the exact same way as duplex collapsing without eliminating variants for

not being in both reads). These variant counts are shown while also varying the number of required independent supporting captures for a variant to pass filtering. The logic behind this analysis is that the fewer captures in which a variant is found, the less confidence we have that it represents true biological signal. Lower confidence variants should be more likely to be eliminated by duplex collapsing reads, if other filters were otherwise insufficient. We show that whether reads are first duplex collapsed or not, there is little effect on the percent of variants that are eliminated, suggesting that our other filtering parameters appear to adequately eliminate sequencing errors. **b**, While the filters used for FERMI should eliminate the majority of errors introduced during PCR amplification and those errors arising from sequencing mistakes, errors made in the first round of PCR amplification could be identified as false positives. If there is a sufficient number of PCR errors made within the first round of amplification, these errors could create artificial patterns within the data. Using one supporting capture as the lower limit for variants to be identified as true signal, the expected number of errors were estimated from amplification using Phusion polymerase and are shown in the table (two estimations are included because Illumina's reaction mixtures are proprietary and we do not know the exact reaction conditions). **c**, When only requiring one supporting capture, 3-6% of variants should be derived from first round PCR errors, although more than half of these will be eliminated by the requirement that 55% of reads for a capture support the variant (errors from subsequent PCR rounds will be

even more efficiently eliminated by the 55% cutoff). If we require that the same variant be present at the same location across multiple captures before it is included in the final results, it becomes exponentially more unlikely that a first round PCR error would get included. In contrast, increased capture number requirements have a much more modest effect on variants called. **d**, While increasing the number of required supporting captures eliminates rare variants as well as first round PCR errors, the numbers of identified variants only decreases modestly for all individuals (blue line, left y-axis). In contrast, the number of variants expected to be identified as a result of first round PCR amplification errors exponentially decreases with each extra capture requirement (red line, right y-axis). When compared to the number of variants that pass all filters and processing, the first round PCR errors appear to have minimal effect even when only a single capture is required. Expectedly, as we increase the number of required captures supporting a variant, the total number of variants also decreases, and after two required captures should essentially not include mutations created by PCR amplification. Throughout most of this paper, a single capture is used, so as to not bias results by variant representation. Nonetheless, the patterns of mutations identified look very similar when greater numbers of supporting captures are required. **e**, As shown in Figure 1A, when deriving consensus reads, variants are eliminated for being rarely observed across reads supporting a given capture. The cutoff we use throughout most of this manuscript is 55%, such that a given variant must be present in at least 55

percent of sequencing reads supporting a capture or they are ignored. The logic behind this chosen cutoff is that more stringent cutoffs largely do not alter the observed mutation spectra, but result in a significant loss in putatively true positive signal. With this cutoff, the expected number of sequencing errors can be estimated. We observe that 9 percent of bases are mismatched within reads supporting a given capture. Each capture is approximately 150bp in length and is supported by an average 13.5 reads. This yields an average of 182.25 errors within each sequenced capture.

Applying the requirements that 55-95 percent of reads must support a given variant (shown as m), the number of false positive signals that pass filtering for each prepared blood sample can be computed. Within each capture there are approximately 450 total possible changes, and an average of 18 reads supporting each capture:

$$E_{seq} = m * 18\ reads/capture)^{\frac{182.5\ PCR\ err}{450\ bp}} * 1200000\ captures/sample$$

$$m = 0.55 : E_{seq} = 155.95\ errors/sample$$

$$m = 0.65 : E_{seq} = 31.48\ errors/sample$$

$$m = 0.75 : E_{seq} = 6.19\ errors/sample$$

$$m = 0.85 : E_{seq} = 1.22\ errors/sample$$

$$m = 0.95 : E_{seq} = 0.24\ errors/sample$$

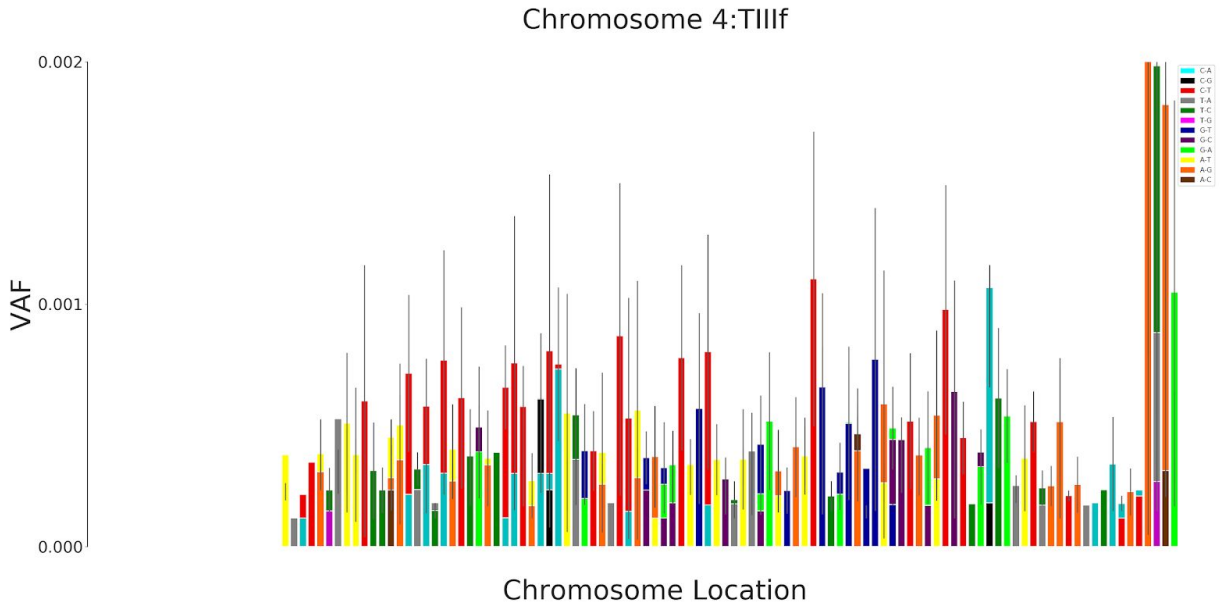
$$E_{tot} = 0.09 * 150\ bp * 13.5\ reads$$

$$E_{tot} = 182.25$$

The number of expected PCR amplification errors to pass all cutoffs is then estimated using a Gaussian distribution. The logic is that the first round of PCR amplification will create errors that will be at an allele frequency near 50 percent as an error will be created in one of two strands of a captured sequence. Using a Gaussian distribution with a mean at 50, the number of all PCR amplification errors expected to pass the 1 supporting capture and 55-95 percent of sequencing reads criteria can be calculated by integrating under the Gaussian distribution. Since we expected about 405 first round PCR amplification errors, and subsequent errors will exist at much smaller allele frequencies, the expected number of variants expected to pass criteria is calculated as follows:

$$E_{tot} = 405 * \int_{c}^{100} f(x) + m_c$$

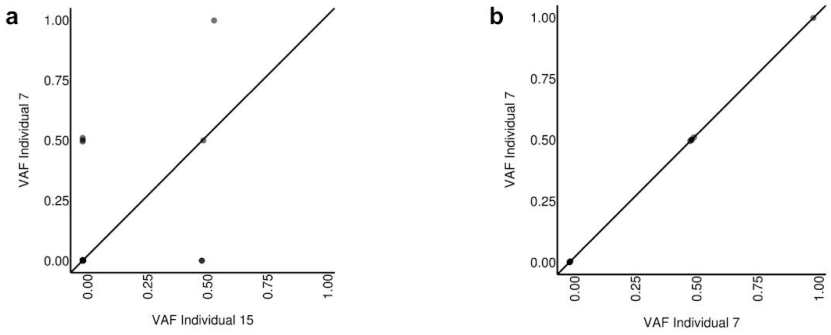
Above we integrate from the support allele frequency *c* to *100* under the Gaussian distribution *f(x)*, multiply this by the expected total number of first round PCR amplification errors, and add to this the number of expected sequencing errors *m* as a function of the support frequency *c*. As shown here, when variants must be supported by at least one unique capture and at least 55 percent of supporting reads, we anticipate only about 150 total variants false variants to make through all FERMI analysis. We believed this to be an acceptable amount of noise given that we see about 6000 total variants from each sample and generated most of the data in this manuscript with these criteria.



Supplemental Figure 2

Sperm gDNA shows similar background to blood samples.

Use of FERMI to sequence sperm gDNA shows a significant signal within uncorrected consensus reads. Work from other groups robustly shows that any given sperm cell should contain about 100 mutations, confirming the presence of a widespread background signal within our uncorrected consensus reads (Lynch 2016).



c

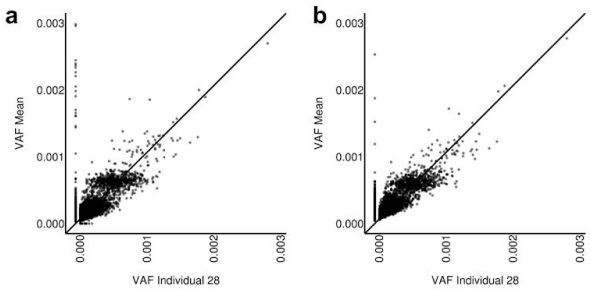
Individual	0mo vs 1mo
Individual A	0.460348
Individual B	0.538478
Individual C	0.436766
Individual D	0.522387
Individual E	0.519219
Individual F	0.482805

Supplemental Figure 3

Resequenced samples are not more similar to each other than to other individuals.

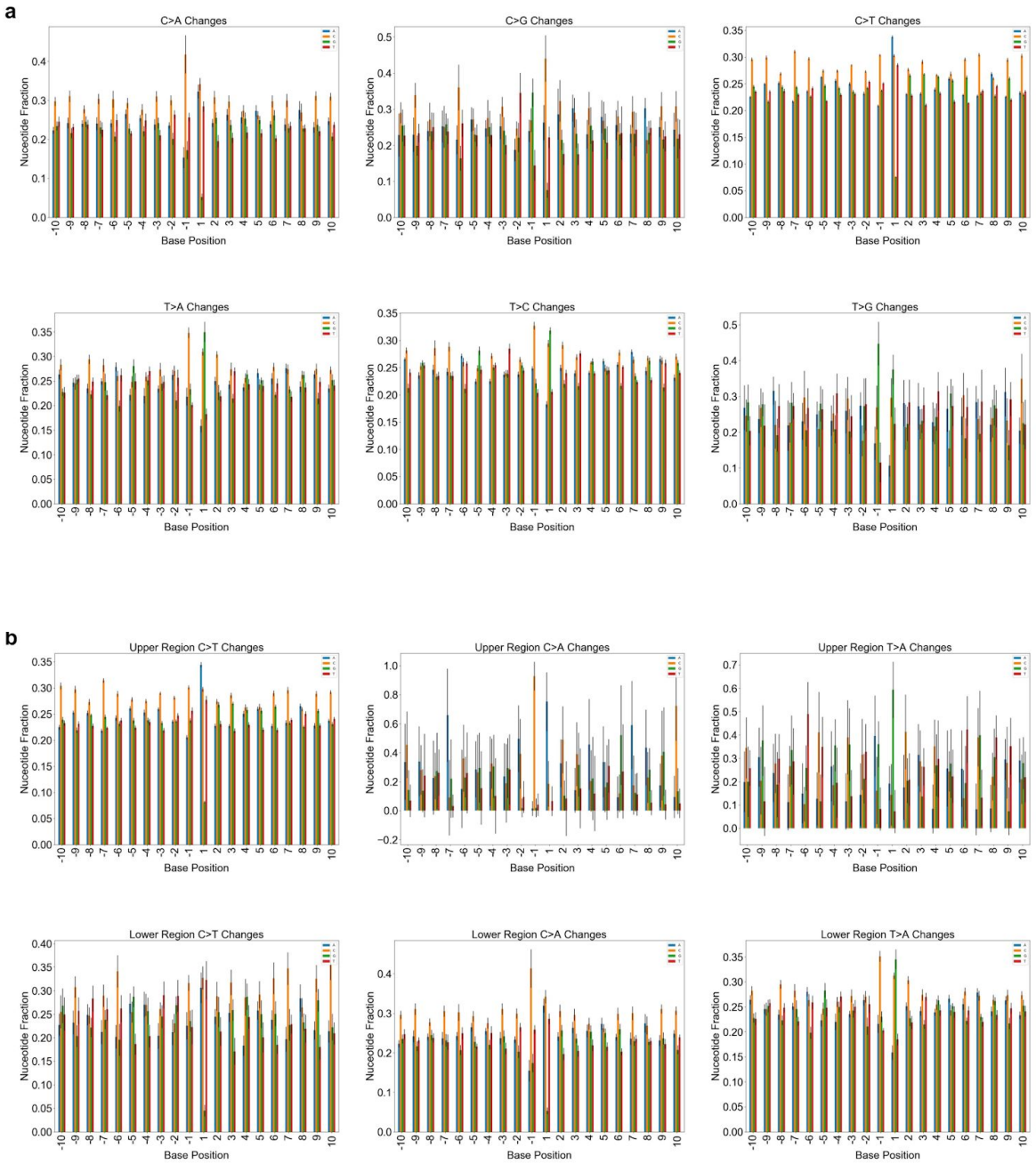
To understand how somatic mutation loads change with time, individuals were biopsied twice, 30 days apart. **a**, When comparing different individuals, low frequency variants tend to exist close to a y=x line, while high frequency SNPs differ. As expected, such SNPs cluster around frequencies of 0.5 and 1. **b**, When comparing the biopsies taken from the same individual variants show a high degree of similarity, both among SNPs and more rare variants. **c**, Though repeat sequencing of individuals typically results in close matches of rare variants with VAFs lower than 0.003, repeats

do not more closely each other than they match the VAF population mean or any other typical sample. This suggests that the differences observed between samples is likely due to sampling differences than to real differences in individual mutation patterns.



Supplemental Figure 4. Somatic mutation patterns are similar across multiple experiments.

For consistency, all samples used in the main analysis derived from a single bulk library preparation and sequencing run, forming Cohort 1. To ensure that the observed trends were not the result of some bias specific to this single preparation, the entire process was independently repeated, with eleven different blood biopsies to form Cohort 2. **a**, Individuals from Cohort 2 closely resembled mean VAFs from the Cohort 1, shown through one of the individuals from Cohort 2 (R-squared = 0.455316, p-value = 0.000000). **b**, Similarly, individuals from Cohort 2 closely resemble the overall VAF mean created from Cohort 2 samples as shown through one of the individuals from Cohort 2 (R-Squared = 0.615327, p-value = 0.000000).

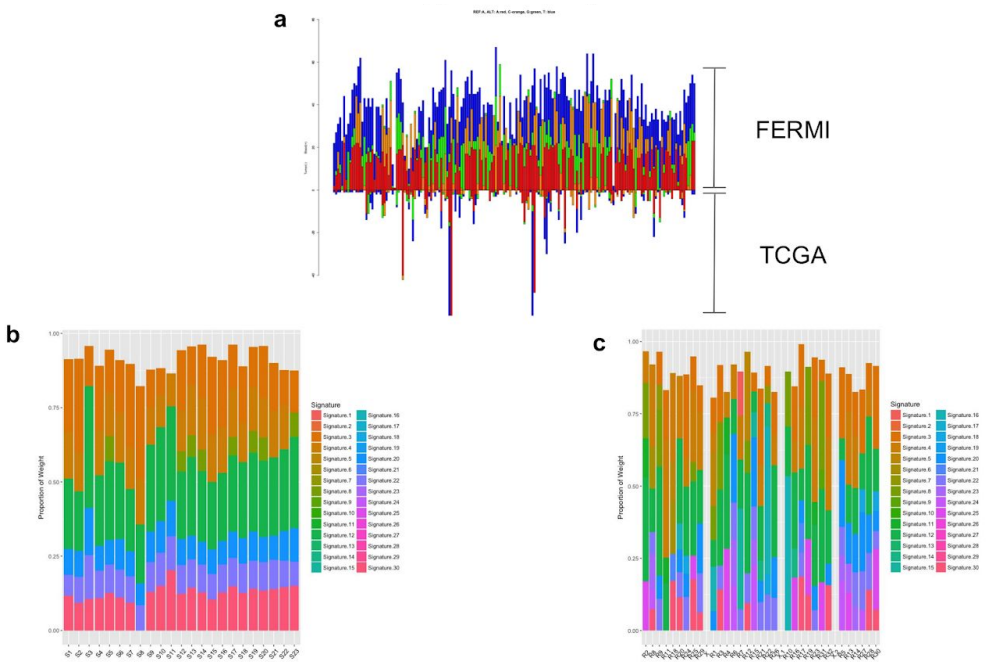


Supplemental Figure 5

Trinucleotide sequence is more impactful on substitution type than more distant surrounding bases.

a, Shown for each of the 6 possible substitution types are the relative frequencies at which each of the various nucleotides are found at a given position surrounding the mutated nucleotide (plotted as means with standard deviation between individuals as error; n = 22). Within each of the subplots, the mutation is not shown and exists at position 0. It appears that the greatest skewing of nucleotide representation occurs at positions -1 and 1, suggesting that they have the greatest impact on how a base will mutate when it suffers a substitution. Note that for C changes, underrepresentation of G at position 1 is expected based on low representation of CpGs in the captured regions. **b**, As seen in Fig 2a-b, substitutions tend to exist within an upper or lower region of allele frequencies. To understand if flanking nucleotide sequence plays a role in this, the populations were analyzed separately for each of 6 base changes at Cs and Ts. Suggesting the actual mutation plays a role in resulting VAF, most substitutions exist disproportionately in either the upper or lower population rather than being equally distributed between the two. For some comparisons, this resulted in larger error within one of the populations, rendering some comparisons not feasible. For C>T changes, the flanking base sequence was largely conserved between the two populations. Other substitutions show differences in flanking sequence when they exist

at higher or lower VAFs, as observed for T>A at positions -1 and 1 and C>A at position 1.



Supplemental Figure 6

Blood mutational patterns exhibit previously identified signatures distinct from those in cancers.

a, We focused on the amplicons in coding regions, and integrated Pan cancer somatic mutation data from exome sequencing in the TCGA to analyze patterns of base substitutions at genomic positions in the target regions which were mutated in both blood and tumor genomes. Substitution frequency and substitution patterns were both significantly different between blood and tumors, both at highly mutated sites (mutation count > 10; Chi square test; FDR adjusted p-value <0.05) and across all such sites (Mantel test; p-value < 1e-5), with substitution patterns in tumor genomes being more

skewed. It is possible that selection during cancer evolution (as opposed to nearly neutral evolution in terminally differentiated blood cells) contribute to the observed patterns. **b**, Integrating trinucleotide contexts of the substitutions, we determined the contributions of different mutation signatures previously identified. Out of 30 previously identified signatures, our data showed overrepresentation of only 7 of them (Signatures 3, 4, 8,12, 20, 22 and 30) across different samples. Out of seven signatures, Signature 12, 3 and 4 had maximum contributions. Signature 3 and 4 are known to be associated with failure of DNA double stranded break repair by homologous repair mechanism and tobacco mutagens respectively, whereas the aetiology of Signature 12 remains unknown. **c**, There was no systematic difference in mutation signatures between amplicons when grouped by their genomic context, and they also showed similar pattern of enrichment of few signatures as compared to others, with signature 12, 3 and 4 having maximum contributions. Signature 12 and 4 exhibits transcriptional strand bias for T>C and C>A substitutions respectively, whereas signature 3 is associated with increased numbers of large InDels.

Supplementary Table 2

	C>A		C>G		C>T		T>A		T>C		T>G	
	R-Squared	p-value	R-Squared	p-value	R-Squared	p-value	R-Squared	p-value	R-Squared	p-value	R-Squared	p-value
TNT	0.942188	0	0.283838	0.000008	0.942188	0	0.165416	0.001692	0.582888	0	0.582888	0
CNT	0.942336	0	0.072766	0.022908	0.942336	0	0.373684	0	0.373684	0	0.838034	0
GNT	0.914237	0	0.999999	0	0.914237	0	0.50136	0	0.50136	0	0.595753	0
ANT	0.952932	0	0.638111	0	0.952932	0	0.556291	0	0.556291	0	0.488577	0
TNC	0.958693	0	0.313046	0	0.958693	0	0.34038	0	0.34038	0	0.622346	0
CNC	0.884787	0	0.248655	0.00042	0.884787	0	0.503585	0	0.503585	0	0.690602	0
GNC	0.907913	0	0.641254	0	0.907913	0	0.46375	0	0.46375	0	0.999989	0
ANC	0.919144	0	0.404024	0	0.919144	0	0.82306	0	0.82306	0	0.871428	0
TNG	0.98003	0	0.98297	0.083316	0.98003	0	0.34038	0	0.560237	0	0.658504	0
CNG	0.999966	0	0.218539	0.04357	0.999966	0	0.491406	0	0.491406	0	0.77013	0
GNG	0.937916	0	0.589258	0.000012	0.937916	0	0.57064	0	0.57064	0	0.77013	0
ANG	0.935708	0	0.452428	0.006004	0.935708	0	0.928586	0	0.928586	0	0.627575	0
TNA	0.939775	0	0.470265	0	0.939775	0	0.960272	0	0.960272	0	0.350091	0
CNA	0.927451	0	0.459003	0	0.927451	0	0.933907	0	0.933907	0	0.999998	0
GNA	0.907147	0	0.229226	0.000016	0.907147	0	0.750916	0	0.750916	0	0.739989	0
ANA	0.929315	0	0.530768	0	0.929315	0	0.536645	0	0.536645	0	0.408415	0

Supplemental Table 1

Trinucleotide context is not sufficient to predict base mutability.

To understand how well the trinucleotide context of each unique nucleotide substitution predicts base mutability, all biopsied individuals were split into two groups (Group 1 and Group 2), which were similar in ages. Within these groups, each substitution was sorted by nucleotide and trinucleotide identity. Sorted substitutions were then plotted by their VAF and compared between Group 1 and Group 2. If trinucleotide context is sufficient to predict how often and to what a nucleotide mutates, it would be expected that the comparison between Groups 1 and 2 would result in a uniformly clustered set of variants. If this were the case, the R-squared value would be small as the variant population would not fit a line. Alternatively, if factors other than just trinucleotide context were important in determining the mutability of a

particular context, it would be expected that variant comparisons between Groups 1 and 2 would strongly adhere to a y=x line and therefore have a high R-squared value. For each context, the R-squared values are shown for the comparisons between Groups 1 and 2. With most comparisons showing a high R-squared value, it is clear that trinucleotide context is not sufficient to predict base mutability.