Detection of Clonal Hematopoiesis of Indeterminate Potential in Clinical Sequencing of Solid Tumor Specimens

Eric A. Severson [1,9], Gregory M. Riedlinger [2,3,4,9], Caitlin F. Connelly [5,9], Jo-Anne Vergilio [5], Mendel Goldfinger [2,6], Shakti Ramkissoon [1,7], Garrett M. Frampton [5], Jeffrey S. Ross [5], Anthony Fratella-Calabrese [3], Laurie Gay [5], Siraj Ali [5], Vincent Miller [5], Julia Elvin [5], Mohammad Hadigol [2,3], Kim M. Hirshfield [2,6], Lorna Rodriguez-Rodriguez [2,8], Shridar Ganesan [2,3,6], Hossein Khiabanian [2,3,4]

[1] Foundation Medicine, Inc., Morrisville, NC, USA, [2] Rutgers Cancer Institute of New Jersey, Rutgers University, New Brunswick, NJ, USA, [3] Center for Systems and Computational Biology, Rutgers Cancer Institute of New Jersey, Rutgers University, New Brunswick, NJ, USA, [4] Department of Pathology and Laboratory Medicine, Rutgers Robert Wood Johnson Medical School, Rutgers University, New Brunswick, NJ, USA, [5] Foundation Medicine, Inc., Cambridge, MA, USA, [6] Department of Medicine, Rutgers Robert Wood Johnson Medical School, New Brunswick, NJ, USA, [7] Department of Pathology, Wake Forest Baptist Health and Comprehensive Cancer Center of Wake Forest University, Winston-Salem, NC, USA, [8] Department of Obstetrics and Gynecology, Rutgers Robert Wood Johnson Medical School, Rutgers University, New Brunswick, NJ, USA

[9] These authors contributed equally to this work

Supplemental Methods and Figures

Methods

Patient cohorts. At Foundation Medicine, Inc. solid tumor cases were evaluated for all classes of genomic alterations using comprehensive genomic profiling (CGP) by the FoundationOne® and FoundationOne®Heme platforms; 257 shared genes were interrogated via DNA sequencing in both assays as previously described.[1]

At Rutgers, under the IRB protocol 20170000252, anonymized patient clinical data was obtained from the Cancer Institute's Bioinfomatics Data Warehouse.[2] This data included CGP sequencing data, age at diagnosis, cancer type, treatment record, subsequent clinical states, and history of carcinogenic exposure, including smoking and prior therapy. Anonymized peripheral blood samples as well as Formalin-Fixed Paraffin-Embedded (FFPE) slides were obtained from the Cancer Institute's Biospecimen Repository Service. Among the patients with CHIP-associated alterations at VAFs lower than expected from tumor purity, 53% (19 out of 36) had a history of smoking and 58% (21 out of 36) had gone under leukemogenic chemotherapy or radiation prior to obtaining the sample that was sequenced; in total, 86% (31 out of 36) of the patients were exposed to known carcinogens before specimen collection for genomic profiling.

Disease group assignment. Diagnoses were assigned at time of sequencing based on the clinical information provided. Diagnoses were placed into groups dividing sarcomas, melanomas, mesotheliomas, glial neoplasms, and carcinomas. Carcinomas were further subdivided based on anatomic location. Diagnoses not fitting in one of these categories were placed in an 'other' category. This resulted in 21 disease groups.

Identifying patients with CHIP mutations. The FoundationOne® and FoundationOne®Heme assays were used for CGP of 132,872 samples of which 113,079 were solid tumors from patients older than 20 years. VAF was calculated by the number of sequencing reads divided by the total number of reads at that location. Computational tumor purity was calculated based on the sequencing data.[1] Specifically, each tumor sample was analyzed alongside a process-matched normal control (an internally-validated mixture of 10 heterozygous diploid HAPMAP samples), which allowed normalizing sequence coverage distribution across baited targets.[1] Purity and base ploidy were inferred simultaneously using empirical Bayesian sampling methodologies. Credibility and confidence intervals varied with sample data, but this method has shown high performance (sensitivity of 99%, with positive predictive value >99%) when tumor content is within 20-75%.[1] Moreover, since *TP53*-mutated tumors often lose wild-type allele while remaining copy-number neutral,[3] *TP53* mutations were used to provide additional evidence for purity estimates, when possible. Hybrid-capture, high-depth sequencing permits high-confidence identification of genomic alterations at allele frequencies. For substitutions, the VAF cutoff is 1% for known somatic variants (based on the COSMIC database)[4] and 5% for novel somatic variants. For indels, the VAF cutoff is 3% for known somatic variants and 10% for novel somatic variants.[1,5] Mutations were classified as associated with CHIP when they were verified by the NCCN Clinical Practice Guidelines in Oncology on Myelodysplastic Syndrome,[6] or were previously assessed as CHIP in healthy individuals.[7,8] All polymorphisms in SNP database were removed in all analyses.

Calculating variant clonality. The power to detect mutated, small clones depends on the sequencing depth, the relative abundance of the cell population that harbors them, and the tumor purity of the sample. Computational tumor purity assessments were used to calculate the probability of each mutation to be heterozygous, somatic, and at significantly smaller VAF relative to the tumor purity [9]. To this end, cumulative binomial probability for each mutation was calculated following $\mathrm{Bino}(f \times D | D, p/2)$, where *f* is VAF, *D* is the total number of sequencing reads covering the mutation, and *p* is the tumor purity. After false discovery rate (FDR) analysis based on the Benjamini-Hochberg method [10], mutations with FDR <0.05 were then classified as associated with CHIP when they were verified by the NCCN Clinical Practice Guidelines in Oncology on Myelodysplastic Syndrome [6], or were previously assessed as CHIP in healthy individuals [7,8].

Logistic regression modeling. A generalized bimodal linear model was used to perform logistic regressions to evaluate the relationship between genomic alterations and patient age, allele frequency and sample tumor purity. The calculations were performed using the R glm function from the 'stats' package v 3.1.1.

Targeted deep sequencing of peripheral blood and macrodissected samples. Deidentified peripheral blood samples as well as macrodissected cells from FFPE slides were sequenced using the ThunderBolts™ Myeloid Panel (RainDance Technologies, Billerica, MA). This panel covers 548 amplicons to interrogate mutational hotspots in 49 genes commonly altered in hematological malignancies (**Supplemental Table 4**), using the Illumina MiSeq sequencer at >1,000x (2x150 bp). The primers for the amplicons in this panel were designed so that the paired-end reads substantially overlap with each other. Each read pair was merged to correct sequencing errors that occur in the instrument. The merged reads were mapped to the reference genome (hg19) using the BWA aligner,[11] and all sites different from the reference sequence were identified using an inclusive variant caller as part of the MERIT pipeline.[12] A Bayesian approach was used to detect true mutations against background sequencing errors, where each mutation was tested in each sample against all sequenced samples.[13] After correcting for multiple hypotheses and FDR analysis based on the Benjamini-Hochberg method, a list of variants that pass FDR <0·001 was generated. For the alterations tested in this study, the mutation-specific sensitivity varied from 0·13% to 1·0%. Unlike previous studies,[14] which limited CHIP mutations in peripheral blood to those with VAFs at least twice the VAF detected in the solid tumor, we did not add other filtering criteria.

Histologic analysis of tumor infiltrating lymphocytes. Available H&E stained FFPE sections of sequenced specimens were evaluated and scored for tumor-infiltrating lymphocytes by a board certified pathologist blinded to genomic data. To ensure unbiased analysis, 19 slides from patients harboring small *TET2* and/or *DNMT3A* mutations were combined with 9 slides from patients harboring clonal *TET2* and/or *DNMT3A* mutations, in addition to 19 slides from a randomly selected group of patients without CHIP associated mutations, which included a variety of solid tumors from specimens located at an assortment of sites. A total of 42 H&E slides were evaluated for the presence of infiltrating lymphocytes on a scale of 1 to 3 (1:low, <10% lymphocytic infiltration; 2:intermediate, 10-20% lymphocytic infiltration; 3:high, >20% lymphocytic infiltration) across the entire slide. The entire slide was evaluated as the whole slide section was used for sequencing.

Tumor and lymphocyte macrodissection. Manual macrodissection was performed to enrich for solid tumor or lymphocyte components for cases in which CHIP mutations were not detected in blood or we did not have blood available. Areas containing a high percentage of tumors or lymphocytes were selected by microscopic examination and circled with a slide-marking pen. Manual macrodissection was then carried out on unstained slides with a scalpel blade using the marked H&E slide as a guide and tissue was transferred to a microcentrifuge tube for downstream DNA extraction and deep sequencing.

Data availability. Sequence data from Foundation Medicine Inc. are not publicly available due to them containing information that could compromise research participant privacy.

References

1. Frampton GM, Fichtenholtz A, Otto GA, et al. Development and validation of a clinical cancer genomic profiling test based on massively parallel DNA sequencing. *Nat Biotechnol*. 2013;31(11):1023-1031.

2. Foran DJ, Chen W, Chu H, et al. Roadmap to a Comprehensive Clinical Data Warehouse for Precision Medicine Applications in Oncology. *Cancer Inform*. 2017;16:1176935117694349.

3. Alexandrova EM, Mirza SA, Xu S, Schulz-Heddergott R, Marchenko ND, Moll UM. p53 loss-of-heterozygosity is a necessary prerequisite for mutant p53 stabilization and gain-of-function in vivo. *Cell Death Dis*. 2017;8(3):e2661.

4. Forbes SA, Beare D, Boutselakis H, et al. COSMIC: somatic cancer genetics at high-resolution. *Nucleic Acids Res*. 2017;45(D1):D777-D783.

5. He J, Abdel-Wahab O, Nahas MK, et al. Integrated genomic DNA/RNA profiling of hematologic malignancies in the clinical setting. *Blood*. 2016;127(24):3004-3014.

6. Greenberg PL, Stone RM, Al-Kali A, et al. Myelodysplastic Syndromes, Version 2.2017, NCCN Clinical Practice Guidelines in Oncology. *J Natl Compr Canc Netw*. 2017;15(1):60-87.

7. Jaiswal S, Fontanillas P, Flannick J, et al. Age-related clonal hematopoiesis associated with adverse outcomes. *N Engl J Med*. 2014;371(26):2488-2498.

8. Jaiswal S, Natarajan P, Silver AJ, et al. Clonal Hematopoiesis and Risk of Atherosclerotic Cardiovascular Disease. *N Engl J Med*. 2017;377(2):111-121.

9. Khiabanian H, Hirshfield KM, Goldfinger M, et al. Inference of germline mutational status and evaluation of loss of heterozygosity in high-depth tumor-only sequencing data. *JCO Precision Oncology*. 2018(2):1-15.

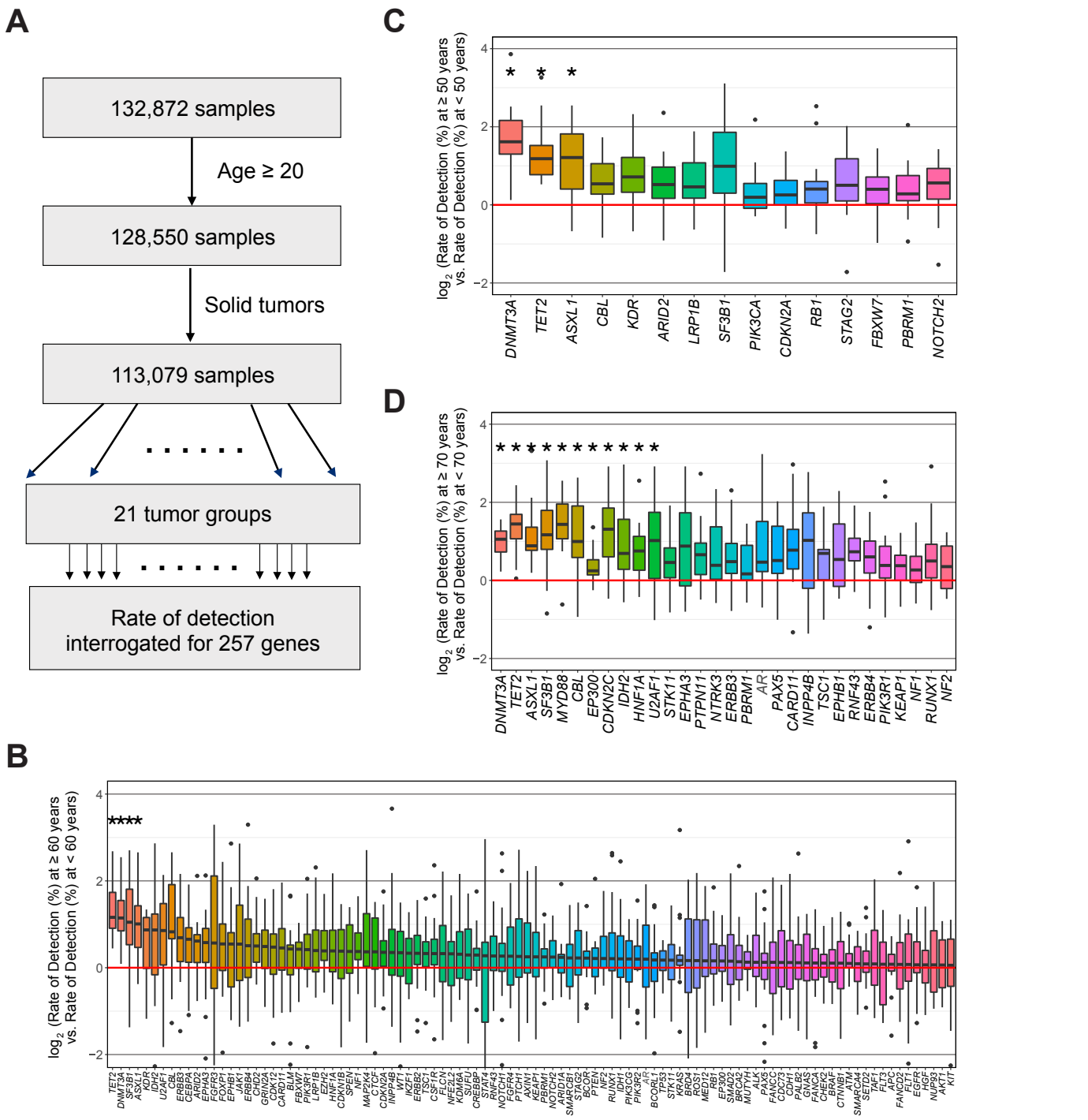
10. Benjamini Y, Hochberg Y. Controlling the False Discovery Rate - a Practical and Powerful Approach to Multiple Testing. *Journal of the Royal Statistical Society Series B-Methodological*. 1995;57(1):289-300.

11. Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics*. 2009;25(14):1754-1760.

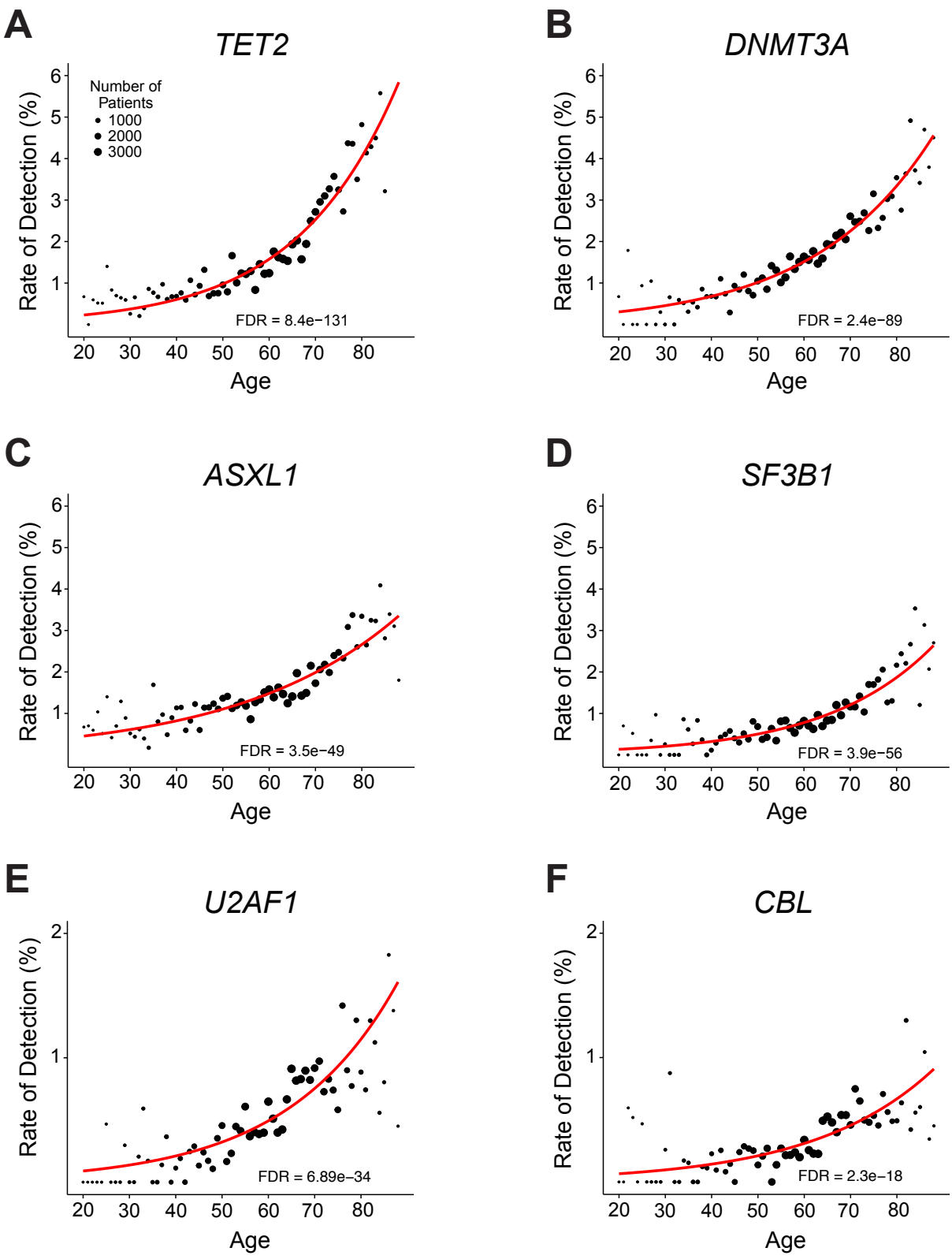
12. Hadigol M, Khiabanian H. MERIT: a Mutation Error Rate Identification Toolkit for Ultra-deep Sequencing Applications. *Preprint at bioRxiv*. 2017.

13. Rabadan R, Bhanot G, Marsilio S, Chiorazzi N, Pasqualucci L, Khiabanian H. On Statistical Modeling of Sequencing Noise in High Depth Data to Assess Tumor Evolution. *Journal of Statistical Physics*. 2017.

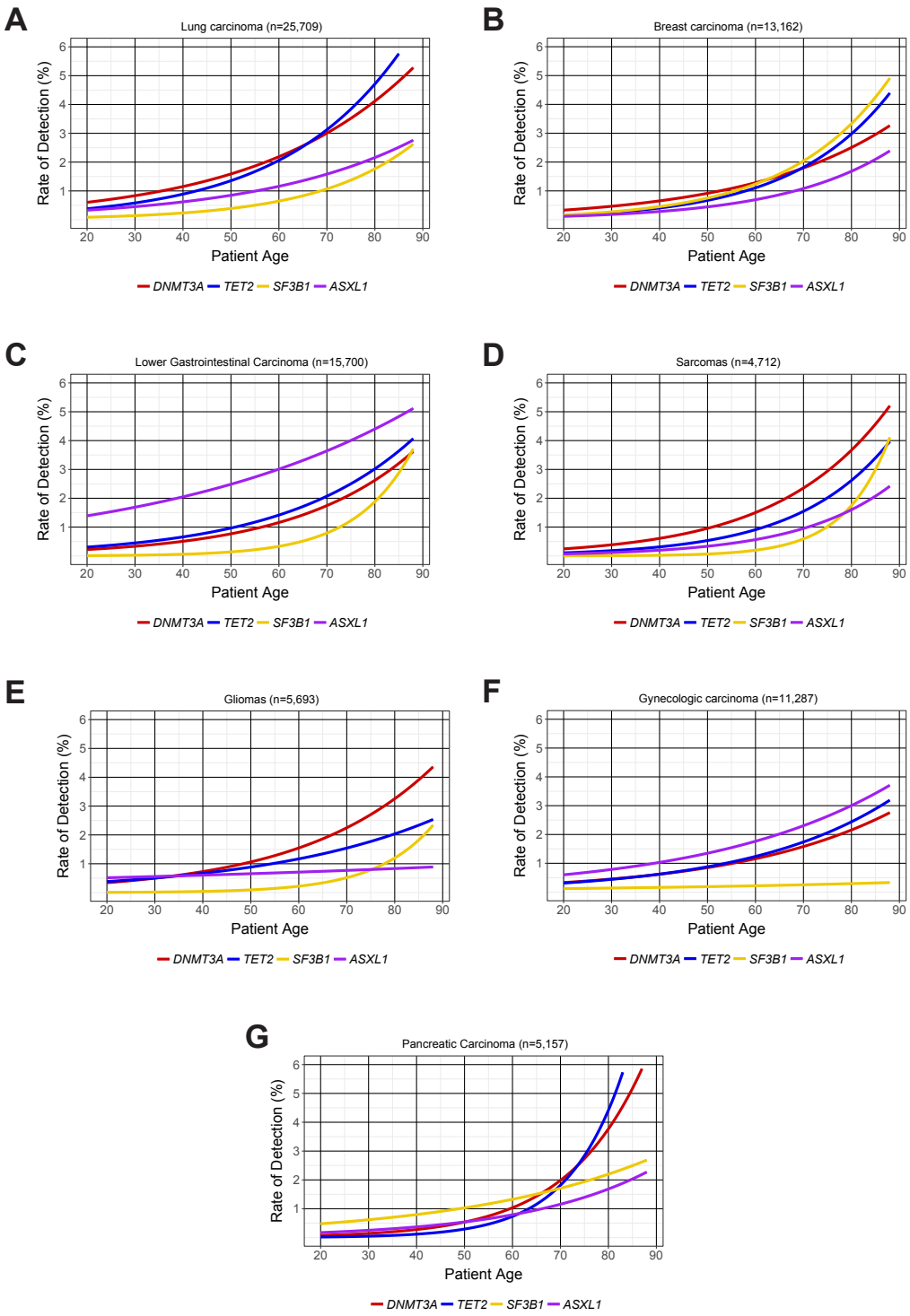
14. Coombs CC, Zehir A, Devlin SM, et al. Therapy-Related Clonal Hematopoiesis in Patients with Non-hematologic Cancers Is Common and Associated with Adverse Clinical Outcomes. *Cell Stem Cell*. 2017.



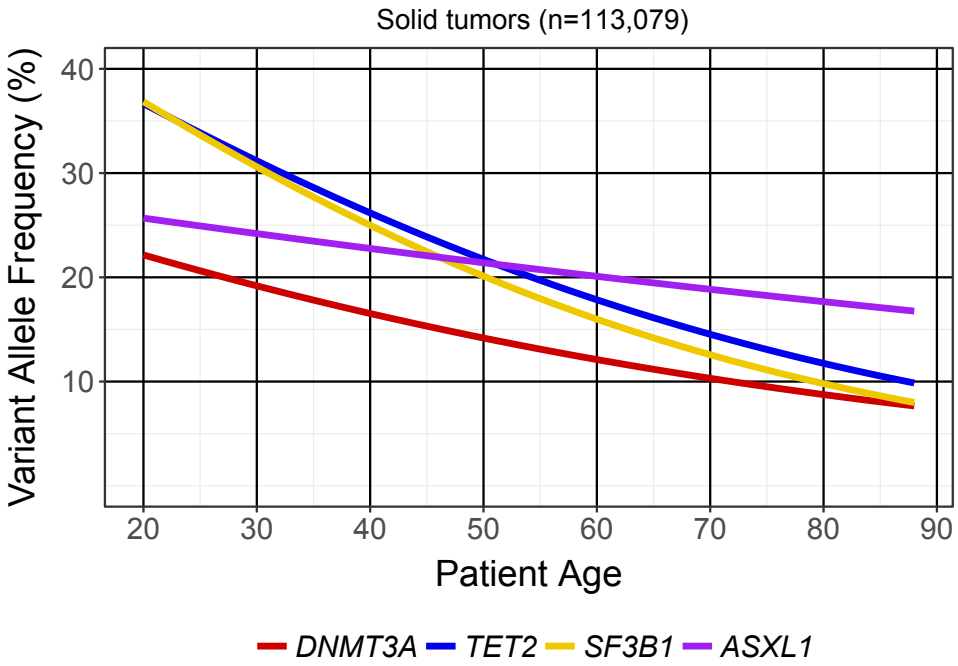
Supplemental Figure 1. Analysis of genomic alterations by age across tumor type reveals CHIP-associated genes. A) Genomic profiles were filtered for patient age ≥ 20 years and solid tumor types, yielding a cohort of 113,079 samples. These samples were divided into 21 tumor groups for further analysis. B) The log odds ratio for rate of detecting genomic alterations in 257 genes in each tumor group, in patients older than 60 years relative to those younger than 60 years. Each data point on the boxplot represents the log odds for a single disease group in a single gene. The top 100 genes, as ordered by the median odds ratio, are shown (* indicates p <0.05). C) The log odds ratio calculated using an age cutoff of 50 year (* indicates p <0.05). D) The log odds ratio calculated using an age cutoff of 70 years (* indicates p <0.05).



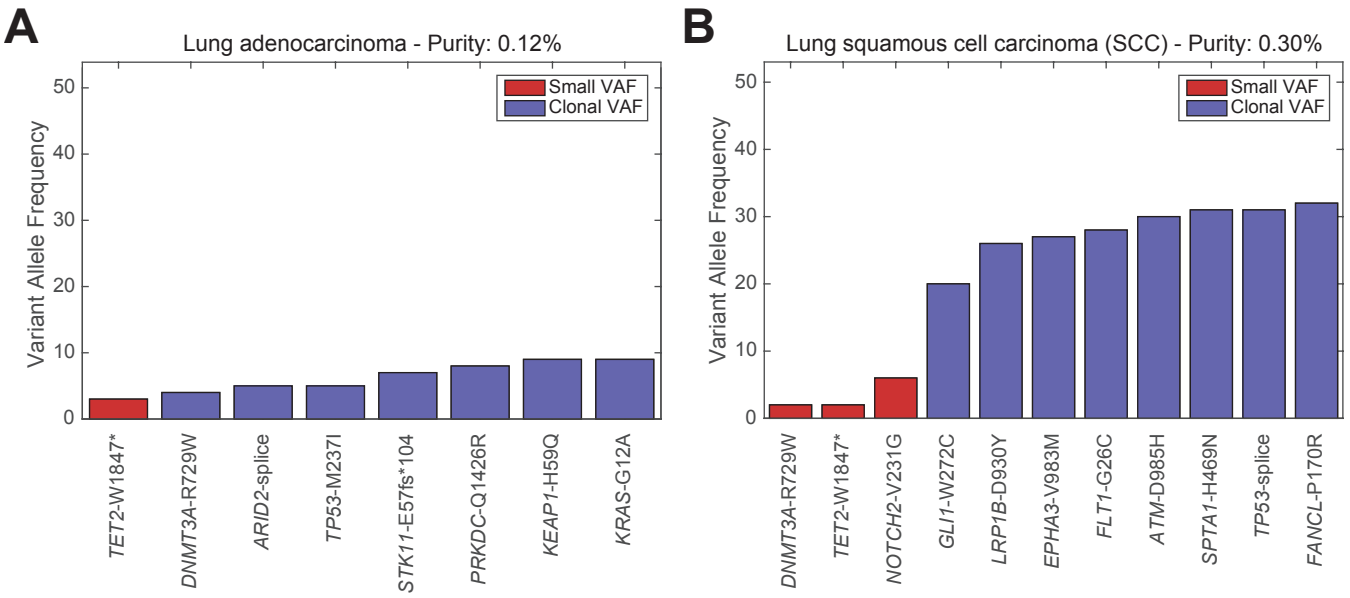
Supplemental Figure 2. Detection of genomic alterations is significantly correlated with patient age in CHIP-associated mutations. A-D) Logistic regression across all samples for *DNMT3A*, *TET2*, *SF3B1*, and *ASXL1* identified using an age cutoff of 60 years. E-F) Logistic regression across all samples for *CBL* and *U2AF1*, identified using an age cutoff of 70 years. Circle size indicates the number of patients in the data set at that age.



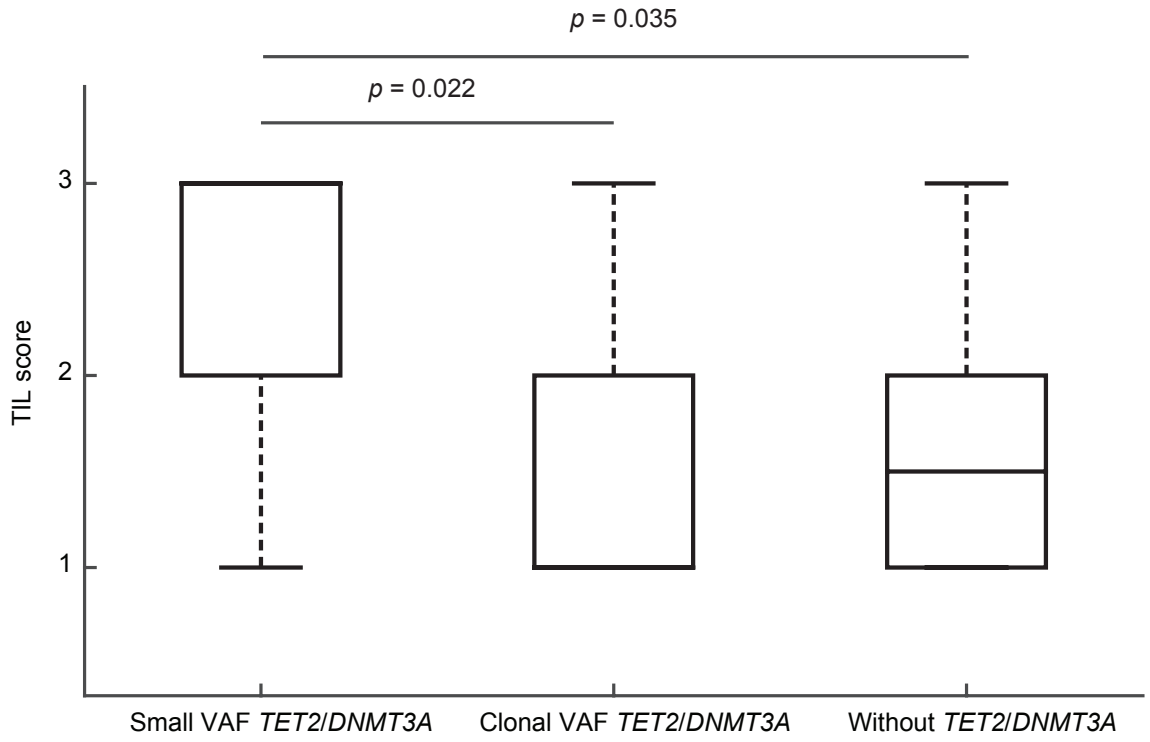
Supplemental Figure 3. Detection of CHIP-associated alterations is significantly correlated with patient age across tumor types. A-G) Logistic regression of genomic alteration rate of detection, by percent of cases, for DNMT3A, TET2, SF3B1, and ASXL1 by patient age for all disease groups with greater than 4,000 samples. Disease groups and the number of cases are listed in Supplemental Table 1. Rates of detection of all four genes were higher in lung carcinoma samples compared to glioma samples, possibly due to differences in vascularity and degree of immune infiltrates. Of note, ASXL1 mutations had a baseline mutation rate of approximately 1·5% in lower gastrointestinal carcinomas, with an age-associated increase occurring in addition to the background alteration rate. The single exception to these trends was that SF3B1 alterations were rarely detected in gynecologic carcinomas.



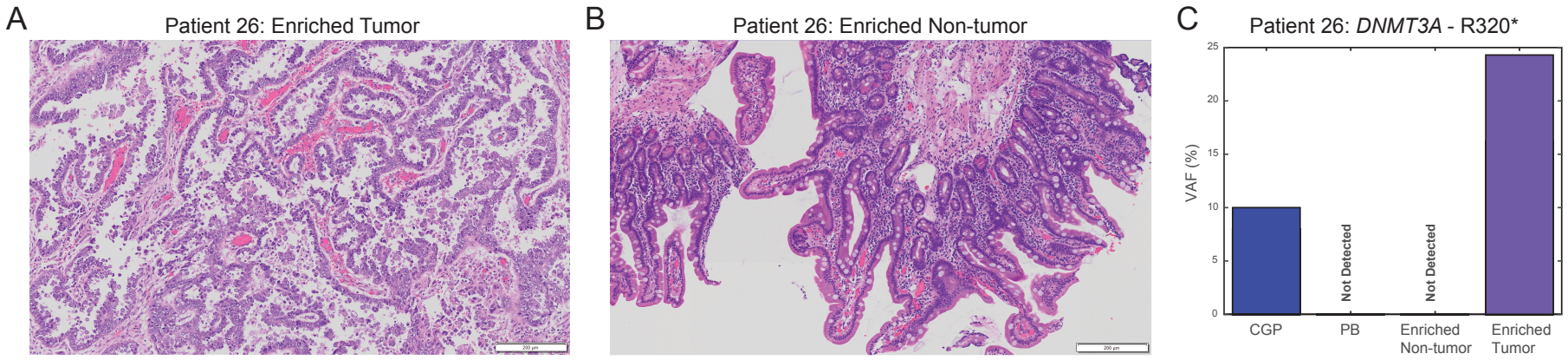
Supplemental Figure 4. The VAFs of CHIP-associated alterations are inversely correlated with age. Logistic regression of variant allele frequency for *DNMT3A, TET2, SF3B1,* and *ASXL1* by patient age (*DNMT3A p* <1.3e-13; *TET2 p* <2.1e-30; *ASXL1 p* <1.6e-7; *SF3B1 p* <4.1e-26), which is consistent with the findings in Figure 1D.



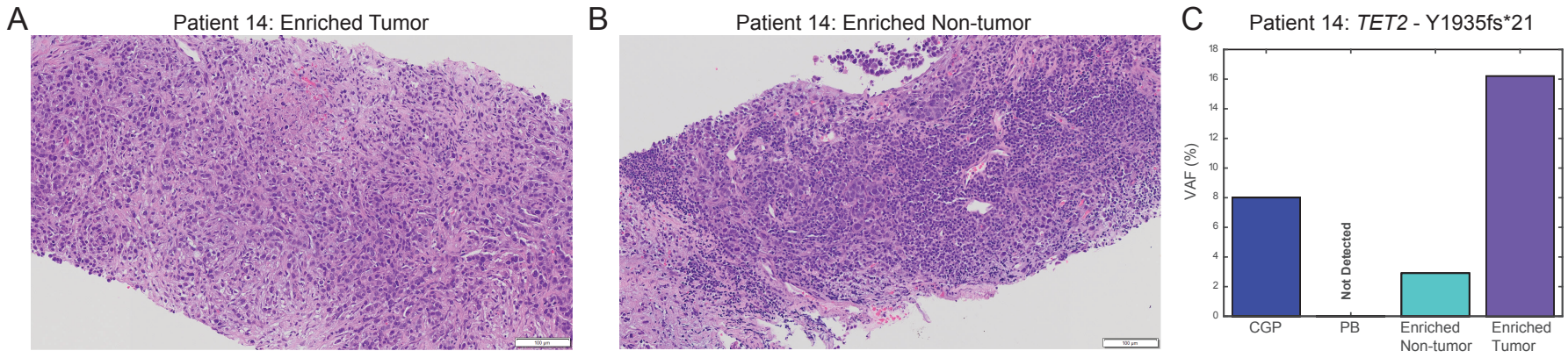
Supplemental Figure 5. Index patient: distinct mutational patter in synchronously diagnosed primary lung cancers. A) CGP results for lung adenocarcinoma. B) CGP results for lung squamous cell carcinoma. The VAF of a *NOTCH2* genomic alteration was also significantly lower, but this mutation was not detected in hematopoietic elements. VAFs for the *TET2* genomic alteration were significantly below the expected from tumor purity in both specimens (adenocarcinoma: p = 8·5e-04, SCC: p = 6·6e-40). For the *DNMT3A* mutation, VAF was significantly less than expected in the SCC (p = 1·8e-26), but due to very low tumor purity, its VAF was not significantly less than expected in the adenocarcinoma (p = 9·6e-02).



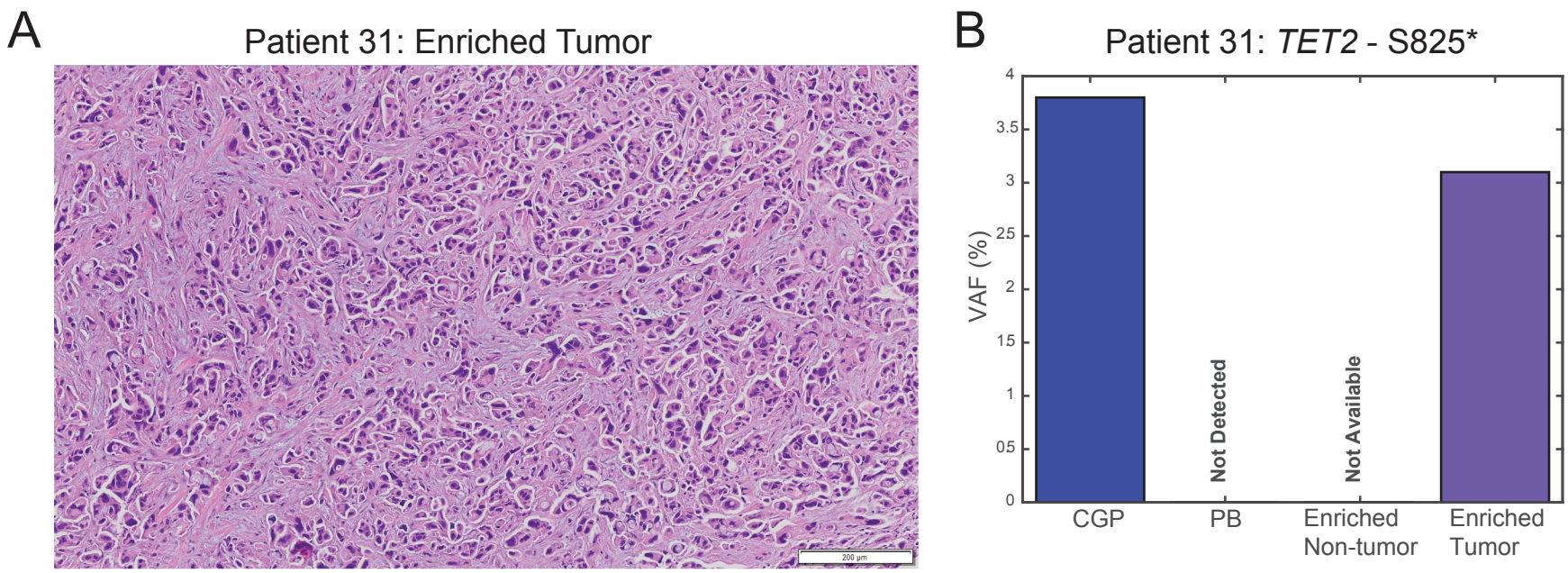
Supplemental Figure 6. Significant tumor-infiltrating lymphocytes (TIL) in samples with small CHIP-associated mutations. H&E slides from 19 patients with small VAF *TET2* and/or *DNMT3A* mutations were scored for TIL and compared with the scores for 9 patients with clonal *TET2* and/or *DNMT3A* genomic alterations and 19 patients with no CHIP-associated genomic alterations. Low score of 1 corresponds to <10%, intermediate score of 2 corresponds to 10-20%, and high score of 3 corresponds to >20% lymphocytic infiltration.



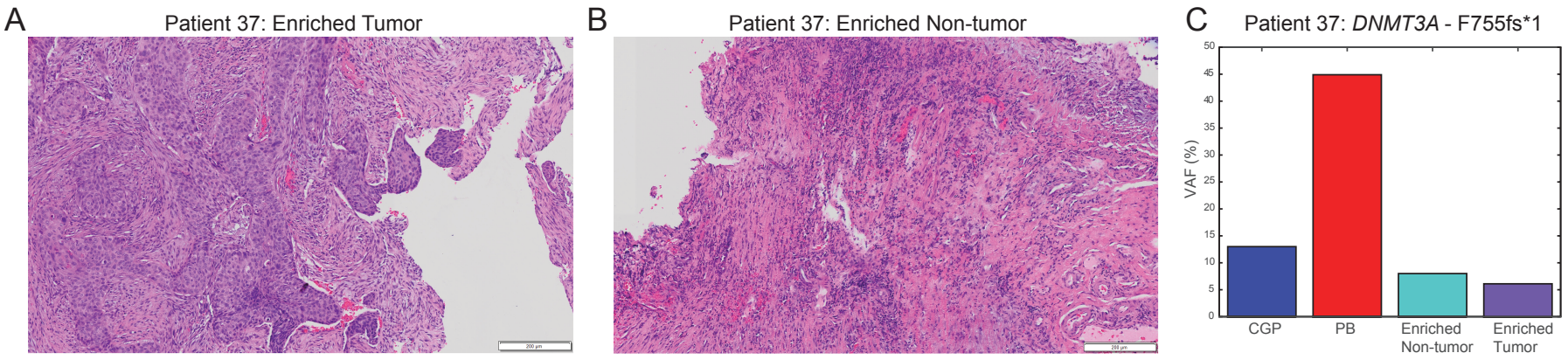
Supplemental Figure 7. Small CHIP-associated mutations were detected in tumor cells of a patient with uterine serous adenocarcinoma. A) Enriched tumor cell population. B) Non-tumor cell population from a duodenal biopsy showing chronic inflammation and no histologic evidence of tumor in patient 26. C) CGP detected a *DNMT3A*-R320* genomic alteration at VAF of 10%, but this mutation was not detected in peripheral blood. We detected this genomic alteration in enriched tumor cells at a VAF of 24·3% and it was not detected in the enriched lymphocyte sample without evidence of tumor.



Supplemental Figure 8. Small CHIP-associated mutations were detected in tumor cells of a patient with breast carcinoma (NOS). A) Enriched tumor cell population. B) Enriched lymphocyte population of from a needle core biopsy in patient 14. C) CGP detected a *TET2*-Y1935fs*21 genomic alteration at VAF of 8%. This genomic alteration was not detected in peripheral blood, but was detected in enriched tumor cells at a VAF of 16·2% and at significantly lower level in enriched macrodissected enriched lymphocytes at a VAF of 2·9%.



Supplemental Figure 9. Small CHIP-associated mutations were detected in tumor cells of a patient with bladder urothelial transitional cell carcinoma. A) Enriched tumor cell population in patient 31. An enriched lymphocyte population was not macrodissected for this patient as only a very limited number of lymphocytes were present in the sample. B) CGP detected a *TET2*-S825* genomic alteration at VAF of 3·8%. This genomic alteration was not detected in peripheral blood, but was detected in enriched tumor cells at similar frequency of 3·1%.



Supplemental Figure 10. Clonal CHIP-associated mutation was present in the hematopoietic elements of a patient with anus squamous cell carcinoma. A) Enriched tumor cell population. B) Enriched lymphocyte population in patient 37. C) CGP detected a *DNMT3A*-F755fs*1 genomic alteration at VAF of 13%, which not significantly different than expected from estimated tumor purity of 20% in the specimen. This genomic alteration was detected in peripheral blood at VAF of 44%, in enriched tumor cells at 6·1%, and in enriched macrodissected lymphocytes at 8%.