Supplementary Methods

1. DNA extraction

Circulating DNA was extracted from plasma using a QIAamp Circulating Nucleic Acid Kit (Qiagen, Santa Clarita, CA). An AllPrep DNA/RNA Mini Kit (Qiagen) was used to purify genomic DNA (gDNA) from formalin-fixed paraffin-embedded tissue. DNA concentrations and purity were quantified using a NanoDrop 8000 UV-Vis spectrometer (Thermo Fisher Scientific, Waltham, MA) and a PicoGreen fluorescence assay on a Qubit 2.0 fluorometer (Thermo Fisher Scientific). The fragment size distribution was measured using a 2200 TapeStation Instrument (Agilent Technologies, Santa Clara, CA).

2. Library preparation

Purified gDNA was sonicated into 150-200 bp fragments using a Covaris S2 (Covaris Inc., Woburn, MA). Tissue samples previously acquired for diagnosis were used to create reference libraries and were subjected to targeted sequencing. The tumor biopsy sample libraries were constructed using a SureSelect XT reagent kit (HSQ, Agilent Technologies) according to the manufacturer's instructions. Then, gDNA and cell-free DNA libraries were created using a KAPA Hyper Prep Kit (Kapa Biosystems, Woburn, MA).

3. Sequencing data processing

The base recalibration process was performed using the BaseRecalibrator and ApplyBQSR functions in the GATK software. Picard was used to identify the unique identifier (UID) family in each group of polymerase chain reaction duplicates. After identifying the UID family, we applied in-house Python (v. 2.7.9, Python Software Foundation, Wilmington, DE) scripts to process the duplicate reads. We modified the integrated digital error suppression methods described by Newman et al. [1] and created scripts to perform the method, which combined *in silico* elimination of stereotypical background errors with a molecular barcoding strategy for sensitive detection of circulating tumor DNA. Parallel sequencing of matched white blood cells was performed to exclude mutations associated with clonal hematopoiesis and enable appropriate variant calling. During processing, discordant pairs and off-target reads were filtered out.

4. Detection of somatic mutations

Somatic mutations were identified by a targeted deep sequencing method. Before

analysis, sequences with low quality scores were removed, and only positions with sequencing depth greater than 500× were used for variant detection. Somatic mutations were identified using the digital error suppression method to minimize sequencing background errors with minor modifications [1]. The detailed process for identifying somatic mutations was described in our previous study [2]. Briefly, peripheral blood leukocyte gDNAs from all patients were used as matched normals to filter out patient-specific germline mutations in the matched samples. To distinguish low–variant allele frequency (VAF) mutations from background noise and to remove false positives, an error distribution was generated from matched normal samples. Variants were annotated for their effects, and synonymous mutations were excluded from the analysis. Mutations with VAF ≥ 0.15% were selected and used for analysis. We performed a Z-test to identify variants that were present at a significantly higher frequency than the corresponding background errors in the normal samples (Bonferroni adjusted p-value < 0.05). We further applied the following thresholds: allele frequency ≥ 0.5% and alternative allele count ≥ 20. For biopsy specimens, single-nucleotide variant profiling was performed with different thresholds: total depth ≥ 100×, allele frequency ≥ 2%, and alternative allele count ≥ 10. Indels were called in all samples using Somatic Strelka2 and MANTA with default parameters [3,4]. Variants that passed the filter in Strelka2 were considered further.

References

1. Newman AM, Lovejoy AF, Klass DM, Kurtz DM, Chabon JJ, Scherer F, et al. Integrated digital error suppression for improved detection of circulating tumor DNA. Nat Biotechnol. 2016;34:547-55.
2. Shin SH, Kim YJ, Lee D, Cho D, Ko YH, Cho J, et al. Analysis of circulating tumor DNA by targeted ultra-deep sequencing across various non-Hodgkin lymphoma subtypes. Leuk Lymphoma. 2019;60:2237-46.
3. Chen X, Schulz-Trieglaff O, Shaw R, Barnes B, Schlesinger F, Kallberg M, et al. Manta: rapid detection of structural variants and indels for germline and cancer sequencing applications. Bioinformatics. 2016;32:1220-2.
4. Kim S, Scheffler K, Halpern AL, Bekritsky MA, Noh E, Kallberg M, et al. Strelka2: fast and accurate calling of germline and somatic variants. Nat Methods. 2018;15:591-4.