KJIM

SUPPLEMENTARY MATERIALS

Sample preparation and next-generation sequencing

One hundred and twenty-three samples taken at initial diagnosis and follow-up samples from bone marrow after treatment (n = 49) were sequenced. Deep-sequencing was performed by targeting the coding regions of 51 genes known to have recurrent driver mutations, as observed in large cohort studies of acute myeloid leukemia (AML) and other myeloid malignancies [1,2]. Genomic DNA was extracted using QIAamp DNA blood mini-kits (Qiagen, Valencia, CA, USA) according to the manufacturer's protocol. After library preparation, all samples were subjected to targeted sequencing per the manufacturer's instructions. Using the Agilent (SureSelect Target Enrichment, CITY, STATE, COUNTRY) custom probe set targeting the coding region of 51 genes (COMPANY, Santa Clara, CA, USA), all samples were sequenced using an Illumina Hiseq 2500 (COMPANY, San Diego, CA, USA) with a 101-bp pair-end mode.

Next-generation sequencing read processing and procedure for variant calling

The read processing and variant calling procedure were followed as in our previous studies [3,4]. For samples without corresponding T-cell samples, the same read processing was used. For variant calling of those samples, we also searched for variants from samples with corresponding T-cell samples in addition to variants in the list compiled from two other AML and MDS studies [1,2], and ran the DeepSNV R package, which implements the Shearwater algorithm using its default parameters [5]. Once lists of candidate somatic mutations were compiled, we filtered out germline mutations and rare SNPs as in our previous studies [3,4] and based on population frequency (over 0.1% according to NHLBI GO Exome Sequencing Project [ESP]) [6,7].

All sequencing data have been deposited at the European Nucleotide Archive (Accession number: PRJEB48636).

SUPPLEMENTARY REFERENCES

1. Papaemmanuil E, Gerstung M, Malcovati L, et al. Clinical and biological implications of driver mutations in myelodysplastic syndromes. Blood 2013;122:3616-3627; quiz 3699.
2. Cancer Genome Atlas Research Network; Ley TJ, Miller C, et al. Genomic and epigenomic landscapes of adult de novo acute myeloid leukemia. N Engl J Med 2013;368:2059-2074.
3. Kim T, Tyndel MS, Kim HJ, et al. The clonal origins of leukemic progression of myelodysplasia. Leukemia 2017;31:1928-1935.
4. Kim T, Yoshida K, Kim YK, et al. Clonal dynamics in a single AML case tracked for 9 years reveals the complexity of leukemia progression. Leukemia 2016;30:295-302.
5. Gerstung M, Papaemmanuil E, Campbell PJ. Subclonal variant calling with multiple samples and prior knowledge. Bioinformatics 2014;30:1198-1204.
6. Fu W, O'Connor TD, Jun G, et al. Analysis of 6,515 exomes reveals the recent origin of most human protein-coding variants. Nature 2013;493:216-220.
7. Wang K, Li M, Hakonarson H. ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. Nucleic Acids Res 2010;38:e164.