Supplementary Methods

Next generation sequencing and copy number variation analyses

Sequenced reads were mapped to the human reference genome (NCBI build 37) using the Burrows-Wheeler Aligner (0.5.9) [1], and demultiplexing was conducted using MarkDuplicates of the Picard package to remove polymerase chain reaction duplicates from the aligned read (http://broadinstitute.github.io/picard). Deduplicated reads were realigned at known indel positions using GATK IndelRealigner [2], and base qualities were then recalibrated using GATK TableRecalibration. Somatic variant calling for single nucleotide variants and short indels was conducted using VarDict [3]. Germline variants from candidates of somatic variants were filtered out with common dbSNP (build 141; found in ≥1% of samples) [4], a panel of normal samples, ExAC, gnomAD, and common variants of 1,100 healthy Korean population. somatic variants were annotated with Variant Effect Predictor (ver. 79) [5] and converted to MAF files using vcf2maf (https://github.com/mskcc/vcf2maf). The candidates of somatic variants were manually reviewed using Integrative Genomics Viewer (IGV) [6]. Copy number analyses were performed using CNVkit [7]. Copy numbers of tumors were adjusted using the tumor cellularity.

References

1. Li H, Durbin R. Fast and accurate short read alignment with Burrows–Wheeler transform. Bioinformatics. 2009;25:1754-60.
2. McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernytsky A, et al. The genome analysis toolkit: A MapReduce framework for analyzing next-generation DNA sequencing data. Genome Res. 2010;20:1297-303.

3. Lai Z, Markovets A, Ahdesmaki M, Chapman B, Hofmann O, McEwen R, et al. VarDict: a novel and versatile variant caller for next-generation sequencing in cancer research. Nucleic Acids Res. 2016;44:e108.
4. Sherry ST, Ward MH, Kholodov M, Baker J, Phan L, Smigielski EM, et al. dbSNP: the NCBI database of genetic variation. Nucleic Acids Res. 2001;29:308-11.
5. McLaren W, Pritchard B, Rios D, Chen Y, Flicek P, Cunningham F. Deriving the consequences of genomic variants with the Ensembl API and SNP Effect Predictor. Bioinformatics. 2010;26:2069-70.
6. Robinson JT, Thorvaldsdottir H, Winckler W, Guttman M, Lander ES, Getz G, et al. Integrative genomics viewer. Nat Biotechnol. 2011;29:24-6.
7. Talevich E, Shain AH, Botton T, Bastian BC. CNVkit: genome-wide copy number detection and visualization from targeted DNA sequencing. PLoS Comput Biol. 2016;12:e1004873.