**Supplementary Methods**

**1. DNA extraction**

For tissue DNA extraction, 10 slides, 5 μm, were required for resected specimens whereas 20 slides, 5 μm, were needed for small biopsy samples. Tissue genomic DNA (gDNA) was extracted from formalin-fixed, paraffin-embedded (FFPE) tissues with the QIAamp sDNA FFPE Tissue kit (Qiagen, Hilden, Germany) according to the manufacturer's instructions and eluted in a 50 μL volume. DNA yield was evaluated using a DropSense16 Micro-Volume spectrometer (Trinean, Kingston, Canada) and Qubit 4.0 Fluorometer (Thermo Fisher Scientific, Waltham, MA). DNA size was examined using a 4200 TapeStation Instrument (Agilent Technologies, Santa Clara, CA). Specimens with a DNA yield over 200 ng and a median DNA fragment size of at least 350 bp were selected for targeted sequencing. If tumor tissue was not available, cell-free DNA (cfDNA) was extracted from plasma using the QIAamp circulating nucleic acid kit (Qiagen) with the QIAvac 24 Plus vacuum manifold, following the manufacturer's instructions. cfDNA was quantified using the Qubit 4.0 (Life Technologies, Burlington, Canada). cfDNA purity was checked using an Agilent High Sensitivity DNA Kit and the 2100 Bioanalyzer Instrument (Agilent Technologies). When required, additional purification was performed using Agencourt AMPure XP (Beckman Coulter, Brea, CA) to remove larger contaminating nucleic acid. cfDNA concentration was quantified with a Qubit 4.0 Fluorometer Fluorometer (Thermo Fisher Scientific) using the Agilent High Sensitivity DNA Kit (Agilent Technologies).

**2. Targeted sequencing and bioinformatics in each platform**

**1) SNUH FIRST Cancer Panel v3.01 [1]**

Targeted sequencing was performed using the SNUH FIRST Cancer Panel v3.01 which includes the whole exomes of 183 cancer-related genes and the intronic regions of 23 genes. Genomic DNA was sheared using a Covaris S220 (Covaris, Woburn, MA). The libraries were prepared with Agilent SureSelectXT target enrichment system for Illumina paired-end sequencing library protocol and sequenced in Illumina Hiseq 2500 platform.

We used FastQC and Trimmomatic (ver. 0.33) software for quality control of the resulting reads. Sequence reads were aligned against the human reference genome version 19 (GRCh37) using Burrows-Wheeler Aligner (ver. 0.7.12). We then marked PCR duplicates using Picard and recalibrated local realignment and base quality score with Genome Analysis Toolkit.

**(1) Mutation calling and filtering**

Single nucleotide variants (SNVs) were detected using MuTect (ver. 1.1.7) [2]. To exclude SNVs due to 8-oxoG artifact, OxoG filter [3] was implemented in house. Briefly, we calculated 'FoxoG' (the alternate reads fraction in OxoG configuration) using mutation data, which were detected independently in read 1 and read 2 by SAMtools mpileup. 'Tumor_lod' of OxoG filter were obtained from 'tumor_rod' score of MuTect, log odds that mutation arises from reference alleles. We selected only SNVs satisfying the following criteria. (1) total depth ≥ 50 (if hotspot, 10), allele depth ≥ 3, allele frequency ≥ 5% (if hotspot, ≥ 1%) (2) OxoG filter: Tumor_lod > -10+(100/3)*FoxoG (3) strand bias test: Fisher test p-value > $10^{-6}$ or allele depth ≥ 10 observed in both strands (4) if variants were in non-hotspot regions, the minimum allele depth > 1 was observed in both strands.

Small insertions and deletions (Indels) were called by IndelGenotyper v36.3336.

INDELs were selected with the criteria as described above for SNV except OxoG filter, Fisher test p-value ≥ 1.0E-20, and allele frequency ≥ 10% (if hotspot, ≥ 5%). The hotspot used for selection of SNVs and INDELs above includes (1) variants observed more than 100 times in COSMIC database [4]; (2) variants observed frequency > 20% in mutation data of The Cancer Genome Atlas (TCGA) cohort including breast, colon, gastric, liver and lung cancer; and (3) selected regions of *EGFR* (exon 19, exon 20) and *MET* (exon 13, intron 13, exon 14, exon 15, splicing sites).

The remaining variants were annotated by ANNOVAR [5] and further filtered out if: (1) not located within exonic or splicing regions; (2) annotated as synonymous SNV; (3) present with a minor allele frequency > 1% in 1000 Genomes Project Database[6], Exome Aggregation Consortium Database [7], NHLBI ESP6500 [8], and Korea Exome Information Database [9]; (4) located within segmental duplication regions and not observed in TCGA mutation data and COSMIC database [4]; and (5) frequently detected unknown variants with low allele frequencies.

For sensitive detection of CNVs, we used two read-depth based algorithms. One method takes an approach of comparison within a sample, applying Z-transformation to RPKM (read per kilobase per million mapped reads) values calculated by Conifer [10] for overall targeted regions. Change of copy number at each gene was estimated by averaging Z-scores of targeted regions to cover each gene. Amplifications were called at genes with Z-score ≥ 30 and gains at genes with range of 20 ≤ Z-score < 30 (or 10 ≤ Z-score < 30 for hotspot genes: *ERBB2*, *MET*). To reduce false-positives, we identified deep deletions at only hotspot genes (*CDKN2A/B*, *PTEN*, *RB1*, *TP53*) with Z-score < –5.5. The other method, CNVkit software identifies segments of copy number ratios from on- and off-target reads of tumor versus pooled normal (comprising

several normal samples for each version of panel as described in the following section) [11]. Amplifications were called at segments with ≥ 6 copies and homozygous deletions at 0 copies.

SVs were called using DELLY [12], and subsequently filtered if they did not have three or more paired-end reads with average mapping quality over 45. To discover confident SVs, all candidates were not only annotated and filtered using the in-house method but also manually reviewed using the Integrative Genomic Viewe r [13,14].

**(2) Estimation of tumor mutation burden**

A modified method of assessing tumor mutation burden (TMB) was implemented based on an earlier described method [15]. It was defined as the number of somatic, coding, and base substitution mutations per Mb of genome examined. Without matched normal samples, we used predicted somatic mutations for calculating TMB. Specifically, the somatic status of SNVs were predicted by using the program PureCN v.3.1 [16], which estimates tumor purity, copy number, and loss of heterozygosity (LOH), and classifies SNVs by somatic status and clonality. PureCN takes GC bias and assay-specific bias into account, which requires process-matched normal samples for bias correction. By using tumor and pool of normal samples, PureCN makes coverage profiles and estimates purity and ploidy of a tumor sample. At the same time, quality control for SNVs is performed, which includes removal of variants with low allele frequency and variants which occur recurrently in normal samples. Then PureCN estimates somatic status of SNVs based on SNV likelihood model, which uses prior knowledge (SNVs in dbSNP database [17] are likely germline and SNVs in the COSMIC database [4] are likely somatic), previously estimated purity and ploidy, and allele frequency. Then these predicted somatic mutations were additionally filtered by

using SNP databases as follows: NCBI dbSNP Database [17] (build150), 1000 Genomes Project Database [6] (1000G_201508), NHLBI ESP6500 [8], the Exome Aggregation Consortium Database (ExAC_0.3) [7], Kaviar Genomic Variant Database [18] (Kaviar_20150923), Korea Exome Information Database [9] and our internal database (PDX). We excluded germline-tagged variants in dbSNP and variants with a minor allele frequency of 1% or greater in other databases.

**2) K MASTER cancer panel v1.1 [19]**

Genomic DNA was sheared using a Covaris S220 (Covaris). Target capture was performed using the Sure-Select XT Reagent Kit, HSQ (Agilent Technologies) and a paired-end sequencing library was constructed with a barcode. Sequencing was performed on a HiSeq 2500 with 100-bp reads (Illumina, San Diego, CA). The paired-end reads were aligned to the human reference genome (hg19) using BWA-MEM v0.7.5. Samtools v0.1.18, GATK v3.1-1, and Picard v1.93 were used for bam file handling, local realignment, and removal of duplicate reads, respectively. Samples with a mean target coverage of less than 200× were excluded from further analysis.

**(1) Mutation calling and filtering**

To increase SNV detection sensitivity, we used two published methods, MuTect v1.1.4 [2], and Lowfreq v0.6.1 [20] with default parameters. The union of the variants identified by the two callers (with the high confidence [HC] set for MuTect) was used as the candidate set of variants. Small insertions and deletions (indels) that were less than 30 bp in size were detected using Pindel v0.2.5a4 [21]. We applied several filtering steps to filter these putative germline variants: (1) variants with very

high VAF (≥ 97%), except for the hotspot mutations; (2) variants with population allele frequency >3% in the > 400 normal samples in our database (this is important for removing ethnic-specific variants); and (3) other frequently detected variants that are likely to be alignment artifacts or are in hard-to-sequence regions, as curated by manual review and compiled in our database. The variants were annotated by ANNOVAR [5].

To identify somatic CNVs, we calculate the mean read depth at each exon, normalized by the coverage of the target regions in that sample using an in-house SV caller [22]. This normalized read depth is further standardized by dividing by the expected coverage for a normal individual (The expected coverage at each exon was taken to be the median of the read depth at that exon across a set of normal individuals). These steps account for the variability in capture efficiency and GC content at different exons. To infer the correct copy number, the amplitude of the copy numbers was then adjusted based on the estimated purity. If the adjusted amplitude of the copy change is greater than 1 or less than 1 (in log scale), the region is called as amplification or deletion, respectively.

Most fusions involve intronic breakpoints. To identify fusion using a gene panel, we tiled across the "hotspot" introns that are known to contain most breakpoints for a set of clinically relevant fusions. Introns of 23 genes were covered densely with capture probes. Since the average DNA fragment size was ~180 bp in our libraries (thus, with 100 bp reads, most fragments are fully sequenced), we expect each fusion to be reflected in multiple split reads. We require four split reads to make a fusion call, with at least two reads mapping to each side of the breakpoint. We also consider both primary and secondary alignments to increase sensitivity. Once the candidate fusions are identified, further filtering is performed using various features including mapping

quality, insert size, CIGAR string, strand direction, alignment information, local cluster coverage, and concordance of the read alignment direction. The split reads allow mapping of the breakpoints with base pair resolution.

**(2) Estimation of TMB**

TMB is the total number of non-synonymous mutations in a DNA coding region. We used the somatic nucleotide variant results of the K MASTER cancer panel sequencing platform to calculate TMB [19]. To obtain only non-synonymous mutations, we performed filtering processes. First, non-coding alterations were excluded. Then germline variants were removed using public data such as ExAC and KRGDB. Lastly, truncation mutations were excluded since K-MASTER Cancer Panel includes genes that are already known to function in cancer. The number of filtered mutations was divided by the length of the target coding region to produce the TMB.

**3) Axen cancer panel**

Targeted sequencing was performed using the Axen Cancer Panel 1 which includes the exomes of 88 cancer-related genes and the intronic regions of 3 genes. *Hybridization-based* enrichment Libraries were prepared manually following the manufacturer's protocol for Agilent *(SureSelectXT HS and XT Low input Target Enrichment System for Illumina Paired-End Sequencing Library).* In summary, 20ng cfDNA was amplified with individual index and molecular- barcode and hybridized with capture oligos during library preparation for Illumina sequencing. The captured sequences are then enriched with streptavidin-conjugated paramagnetic beads and further amplified before being subjected to Illumina sequencing. Fragment sizes for all libraries were measured using the 2100 Bioanalyzer (Agilent Technologies, Palo Alto,

CA), and qPCR was performed on the LightCycler 480 System (Roche, CA) with the Kapa library quantification kit (KK4854, KAPA Biosystems). Sequencing was performed using the Illumina NextSeq500 platform with an average read length of 2 × 150 bp, total depth 5,000×, as per the manufacturer's instructions for Illumina.

Adaptor sequences and low-quality bases in the raw sequencing reads were removed using Agilent Genomics Toolkit (AGeNT, Agilent Technologies). After trimming step, sequence reads were aligned to the human reference genome hg19 using Burrows-Wheeler Aligner-MEM (BWA-MEM) [23]. Using the molecular barcode information, the duplicate reads were marked by LocatIt (Agilent software). Poorly mapped reads were removed using sambamba [24] and base quality score recalibration (BQSR) was performed by Genome Analysis Tool Kit (GATK) [25]. Somatic mutations were detected by the MuTect2 algorithm [2] and variants were annotated using SnpEff & SnpSift [26]. To reduce the effect of false-positive variants, we applied additional filtration criteria: (1) Variants with a minor allele frequency (MAF) more than 5% in the genome aggregation database (gnomAD) [27] and Exome Aggregation Consortium (ExAC) database were excluded [7], (2) Variants with mutated read counts less than 5 were excluded, and (3) Variants with total read depth less than 200 were excluded.

## References

1. Park C, Kim M, Kim MJ, Kim H, Ock CY, Keam B, et al. Clinical application of next-generation sequencing-based panel to BRAF wild-type advanced melanoma identifies key oncogenic alterations and therapeutic strategies. Mol Cancer Ther. 2020;19:937-44.

2. Cibulskis K, Lawrence MS, Carter SL, Sivachenko A, Jaffe D, Sougnez C, et al. Sensitive detection of somatic point mutations in impure and heterogeneous cancer samples. Nat Biotechnol. 2013;31:213-9.

3. Costello M, Pugh TJ, Fennell TJ, Stewart C, Lichtenstein L, Meldrim JC, et al. Discovery and characterization of artifactual mutations in deep coverage targeted capture sequencing data due to oxidative DNA damage during sample preparation. Nucleic Acids Res. 2013;41:e67.

4. Forbes SA, Beare D, Boutselakis H, Bamford S, Bindal N, Tate J, et al. COSMIC: somatic cancer genetics at high-resolution. Nucleic Acids Res. 2017;45:D777-83.

5. Wang K, Li M, Hakonarson H. ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. Nucleic Acids Res. 2010;38:e164.

6. 1000 Genomes Project Consortium; Auton A, Brooks LD, Durbin RM, Garrison EP, Kang HM, et al. A global reference for human genetic variation. Nature. 2015;526:68-74.

7. Lek M, Karczewski KJ, Minikel EV, Samocha KE, Banks E, Fennell T, et al. Analysis of protein-coding genetic variation in 60,706 humans. Nature. 2016;536:285-91.

8. Fu W, O'Connor TD, Jun G, Kang HM, Abecasis G, Leal SM, et al. Analysis of 6,515 exomes reveals the recent origin of most human protein-coding variants. Nature. 2013;493:216-20.

9. Kwak SH, Chae J, Choi S, Kim MJ, Choi M, Chae JH, et al. Findings of a 1303 Korean whole-exome sequencing study. Exp Mol Med. 2017;49:e356.

10. Krumm N, Sudmant PH, Ko A, O'Roak BJ, Malig M, Coe BP, et al. Copy number variation detection and genotyping from exome sequence data. Genome Res. 2012;22:1525-32.

11. 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.

12. Rausch T, Zichner T, Schlattl A, Stutz AM, Benes V, Korbel JO. DELLY: structural variant discovery by integrated paired-end and split-read analysis. Bioinformatics. 2012;28:i333-9.

13. Robinson JT, Thorvaldsdottir H, Winckler W, Guttman M, Lander ES, Getz G, et al. Integrative genomics viewer. Nat Biotechnol. 2011;29:24-6.

14. Thorvaldsdottir H, Robinson JT, Mesirov JP. Integrative Genomics Viewer

(IGV): high-performance genomics data visualization and exploration. Brief Bioinform. 2013;14:178-92.

15. Chalmers ZR, Connelly CF, Fabrizio D, Gay L, Ali SM, Ennis R, et al. Analysis of 100,000 human cancer genomes reveals the landscape of tumor mutational burden. Genome Med. 2017;9:34.

16. Riester M, Singh AP, Brannon AR, Yu K, Campbell CD, Chiang DY, et al. PureCN: copy number calling and SNV classification using targeted short read sequencing. Source Code Biol Med. 2016;11:13.

17. 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.

18. Glusman G, Caballero J, Mauldin DE, Hood L, Roach JC. Kaviar: an accessible system for testing SNV novelty. Bioinformatics. 2011;27:3216-7.

19. Shin HT, Choi YL, Yun JW, Kim NKD, Kim SY, Jeon HJ, et al. Prevalence and detection of low-allele-fraction variants in clinical cancer samples. Nat Commun. 2017;8:1377.

20. Wilm A, Aw PP, Bertrand D, Yeo GH, Ong SH, Wong CH, et al. LoFreq: a sequence-quality aware, ultra-sensitive variant caller for uncovering cell-population heterogeneity from high-throughput sequencing datasets. Nucleic Acids Res. 2012;40:11189-201.

21. Ye K, Schulz MH, Long Q, Apweiler R, Ning Z. Pindel: a pattern growth approach to detect break points of large deletions and medium sized insertions from paired-end short reads. Bioinformatics. 2009;25:2865-71.

22. Shin HT, Kim NKD, Yun JW, Lee B, Kyung S, Lee KW, et al. Junction Location Identifier (JuLI): accurate detection of DNA fusions in clinical sequencing for precision oncology. J Mol Diagn. 2020;22:304-18.

23. Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics. 2009;25:1754-60.

24. Tarasov A, Vilella AJ, Cuppen E, Nijman IJ, Prins P. Sambamba: fast processing of NGS alignment formats. Bioinformatics. 2015;31:2032-4.

25. 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.

26. Cingolani P, Platts A, Wang le L, Coon M, Nguyen T, Wang L, et al. A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of Drosophila melanogaster strain w1118; iso-2; iso-3. Fly (Austin). 2012;6:80-92.
27. Karczewski KJ, Francioli LC, Tiao G, Cummings BB, Alfoldi J, Wang Q, et al. The mutational constraint spectrum quantified from variation in 141,456 humans. Nature. 2020;581:434-43.