

1 **Additional File 2**

2 **Supplementary material**

3 **Running CNVKit, CoNIFER, ExomeDepth and CopywriteR**

4 *Alignment and post-processing*

5 Post sequencing, Illumina basecalls were converted to FASTQs and aligned to the human genome
6 (HS37d5) using BWA mem 0.7.10. Picard MarkDuplicates (v. 1.47), was used to remove PCR duplicates
7 with validation stringency set to lenient. Read groups were added to the BAM file using Picard
8 AddOrReplaceReadGroups. The Genome Analysis Toolkit (GATK) version 1.6 was further applied to
9 refine the alignments using RealignerTargetCreator and IndelRealigner. The resulting BAM files are used
10 as inputs to CNVkit, CoNIFER, ExomeDepth, and CopywriteR.

11
12 *CNVKit*

13 CNVKit v8.1 was first executed in access mode (cnvkit.py access) to create an access bed file (-s10000).
14 Next, CNVKit was run in batch mode (cnvkit.py batch) using all normal, sorted bam files. We specified an exome
15 specific bed file, the access bed file, and used the --split option. After the normal reference file
16 was created, CNVKit was run in batch mode for all samples, using the appropriate reference file to create the cns
17 and cnr files, which were used for all further analysis.

18
19 *CoNIFER*

20 To run CoNIFER (version 0.2.2), we first calculated the RPKM values for the normal samples. Once these
21 files were available, RPKM values for each tumor sample were calculated, and analyzed alongside the
22 normal RPKM files. CoNIFER allows the user to remove singular value decomposition values. We chose
23 to remove 0, 1, 2, and 3 SVD components respectively and used the Z-RPKM values to assess which
24 regions have CNV event.

25
26
27
28
29
30
31
32
33
34
35
36
37
38

ExomeDepth

ExomeDepth (version 1.1.8) was implemented on R (version 3.1.3). It was run by providing a complete list of indexed BAM files, associated with both the tumor and normal samples, for each separate study. Following creation of the counts matrix, CNV calls were made using default parameters and a developer suggested transition probability of 10^{-4} and expected CNV length of 50,000.

CopywriteR

CopywriteR (version 2.12.0) was implemented on R (version 3.5.1). We ran the CopywriteR tool on theTCGA Prostate and AML datasets using the matched tumor/normal samples. MMRF samples, which did not have matched normal tissues, were processed as the tumor only. Segments identified by CopywriteR were then processed by DNACopy to run circular binary segmentation and determine CNV events using log2 thresholds of - 0.25/0.2 for TCGA AML and MMRF, and -0.3/0.3 for TCGA Prostate.

