**Step1: Design SNP probe**

To capture alterations of B allele frequency, SNPs with sufficiently high MAF values must be covered in your bait. Step 1 is about how to select candidate SNPs. If you already have properly designed baits, you can skip this step.

To design a probe set covering targeted regions and genome-wide SNPs, execute following command.

bash design\_probe.sh [SNPs] [Target] [TAG]

[SNPs] : Please select one of the following two sets of genome-wide SNPs.

“cnacs/probe/ver1.bed” (include 1250 genome-wide SNPs)

“cnacs/probe/ver1s.bed” (include 1000 genome-wide SNPs)

[SNPs] : Indicate the path to your bed file of targeted regions.

[TAG] : Indicate the name for output.

You will find the following output files under “cnacs/probe/[TAG]”.

[TAG].all.bed : all candidate SNPs

[TAG].1000.bed : candidate SNPs selected within each 1000bp bin in targeted regions. The same for bin width of 3000, 5000, 10000, 30000, 50000, and 100000.

[TAG].summary.txt : Summary of the categories (Tier1~8) of selected SNPs for each bin width. Smaller tier numbers mean better SNPs.

**Step2: Establish a control pool**

**2-1 : Preparation**

Before analyzing tumor sequencing data, you have to make control pool. To obtain higher S/N ratio in tumor analysis, appropriate controls should be included in the pool. Controls must be captured by the same bait as tumor samples. Controls sequenced together with tumor samples are better. To establish a control pool, make a directory under “cnacs/control/”, and put selected control bam files there as indicated below. Just putting symbolic links also works. We usually put 5~10 controls. Never “remove duplicates” in these bam files.

(e.g.)

cnacs/control/XXX/Controlsample1/Controlsample1.bam

cnacs/control/XXX/Controlsample2/Controlsample2.bam

…..

Prapare sex\_info.txt and put in the same directory.

(e.g.)

cnacs/control/XXX/sex\_info.txt

The format of “sex\_info.txt” should be as follows.

Controlsample1 M

Controlsample2 F

Controlsample3 F (tab-delimited)

**2-2: prepare bait information**

Under “cnacs/probe”, make directory for probe files. The name of this directory must be the same as the control directory you make in Step 1.

(e.g.) cnacs/probe/XXX

Prepare a bed file of the regions covered by your bait, and put it the directory you just made. It is convenient to put gene names or other information in the 4th column.

(e.g.) chr1 1718702 1718942 GNB1

**2-3 : Establish a control pool**

To establish control pool, execute following commands.

cd cnacs/script

bash ref\_count.sh [CONTROL\_TAG] [Path\_To\_BaitFile]

(e.g.) cd bash ref\_count.sh XXX /home/Path/To/Bait.bed

**2-4 : Filtering probes**

When the job finish, You find four pdf files under “cnacs/control/[TAG]/stats” directory.

“depth\_mean.pdf” : histogram of mean depths on all probes

“depth\_coefvar.pdf” : histogram of coefficient of variance of depths on all probes

“baf\_mean.pdf” : histogram of mean BAFs on all hetero-SNPs

”baf\_coefvar.pdf” : histogram of coefficient of variance of BAFs on all hetero-SNPs

Based on these histograms, please set appropriate threshold to filter out low quality probes (outliers or high variance), by editing “cnacs/control/[TAG]/stats/threshold.txt” file. This step is critical to make the result less noisy.

After editing the “threshold.txt” file, please execute following commands.

cd cnacs/script

bash ref\_install.sh [CONTROL\_TAG] [Path\_To\_BaitFile]

(e.g.) bash ref\_count.sh XXX /home/Path/To/Bait.bed

**In WES**, if you don’t want to discard some genes or regions in this step, make a gene file and indicate its path as below. Otherwise, just indicate path to a void file. Its format is the same as a bed file mentioned above (e.g. chr1 1718702 1718942 GNB1).

cd cnacs/script

bash ref\_install.sh [CONTROL\_TAG] [Path\_To\_BaitFile] [Path\_To\_GeneFile\_or\_VoidFile]

**Step3 : Analysis of tumor samples**

**3-1: Preparation**

Make a tumor directory under “cnacs/data/input/”.

(e.g., cnacs/data/input/YYY)

Put tumor bam files under this directory. Never “remove duplicates” in these bam files.

(e.g.)

cnacs/data/input/YYY/case1/case1.bam

cnacs/data/input/YYY/case2/case2.bam

cnacs/data/input/YYY/case2/case2.bam

**3-2: Run main program**

To run the main program, execute following commands.

cd cnacs/script

bash cnacs.sh [TUMOR\_TAG] [CONTROL\_TAG] [Path\_To\_BaitFile]

(e.g.) bash cnacs.sh YYY XXX /Path/To/Bait.bed

You will find the result in cnacs/data/output/[Test\_directory\_name] directory.

**Step4 : Adjustments of results**

**4-1 : Adjustment of ploidy**

Especially in samples with many CNAs (like whole genome doublings), CNACS sometimes make mistakes in ploidy calibration. When you find such miscalibrations, please adjust by executing following commands.

cd cnacs/script

bash set\_plpoidy.sh [TUMOR\_TAG] [sample\_ID] [region] [ploidy]

(e.g.) bash set\_plpoidy.sh XXX tumor\_1 chr1 2

[sample\_ID] : ID for the sample which you want to adjust. Just use the name of directory for that sample.

[region] : Indicate a region whose ploidy can be confidently determined as 1, 2, or 3. Use one of the three formats: chromosome names (chr1, chr2, …), arm names (1p, 1q, …), or genomic coordinates (chr1:10000-20000).

[ploidy] : Indicate the ploidy you selected above. Select from 1, 2, or 3.

**4-2 : Curation of CNA segments**

Automated copy-number call sometimes makes error calls or miss true CNAs. Through manual curation, these mistakes have to be corrected. To do that, please modify the information of CNA segments in “cnacs/data/output/[TUMOR\_TAG]/[TUMOR\_TAG]\_result.org.txt”. After this procedure, please execute the following command.

cd cnacs/script

bash finalize.sh ../data/output/[TUMOR\_TAG]/[TUMOR\_TAG] \_result.org.txt

You will find adjusted informations of CNA segments as “[TUMOR\_TAG]\_result.org.processed.txt", and adjusted plots under “[TUMOR\_TAG]/plot\_final/” directory.

**Step5 : Summarization of results**

To summarize the results of tumor samples, execute following commands.

cd cnacs/script

bash surmmarize.sh [ [TUMOR\_TAG]\_result.org.processed.txt ] [Path\_To\_BaitFile] [n\_samples]

[n\_samples] : The number of tumor samples you analyzed

After that, you find “[TUMOR\_TAG]\_summary” directory, including summarized information and figures of the tumor samples.

**Analysis of WGA samples :**

When you analyze samples after whole-genome amplification (WGA), please note these two points.

1. Please establish control pool consisting only control samples which were also treated with WGA.

2. CNACS has a mode for samples after WGA. To activate that mode, please use following command when you establish control pools or analyze tumor samples.

For establishing control pools:

bash ref\_count.sh -r [CONTROL\_TAG] [Path\_To\_BaitFile]

(e.g.) bash ref\_count.sh -r XXX /home/Path/To/Bait.bed

For analyzing tumor samples:

bash cnacs.sh -r [TUMOR\_TAG] [CONTROL\_TAG] [Path\_To\_BaitFile]

(e.g.) bash cnacs.sh -r YYY XXX /Path/To/Bait.bed

These points are very important to reduce noise due to WGA procedures.