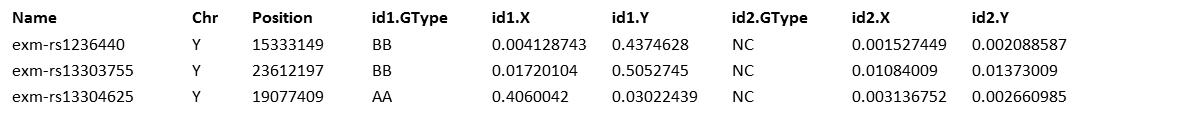
**SOP: Using Z-call on raw Genotypes data (idat files)**

This SOP specifies the use of zCall to post-process Gencall results. This SOP uses Version 3.4 “GenomeStudio - Thresholds derived from GenomeStudio report”. The zCall distribution is available for download at: (<https://github.com/jigold/zCall>)

**zCall Requires a tab-delimited GenomeStudio** **file generated using Illumina’s GenomeStudio software as follows:**

1. **SOP: Running GenomeStudio to generate SNP-wise report file idat files**
2. Capture IDATs from iScan scanner, along with sample sheet.
3. GenomeStudio is available at Broad via Terminal Services (virtual desktop), currently accessed on "rolas" (<https://it.broadinstitute.org/wiki/Terminal_services>)
4. Call using Illumina Genome Studio (Gencall module in the BeadStudio/GenomeStudio software).
   1. Manifest file (HumanCoreExome-12v1-0\_B.bpm).
   2. Import cluster positions from the standard Illumina Cluster file (HumanCoreExome-12v1-0\_B.egt).
   3. Use the default Gen Call Threshold of 0.15.
   4. Output data to a SNP-wise zCall matrix as required by zCall. (See zCall file InstructionsForGeneratingGSreport.txt) The main input format for zCall is a tab-delimited file generated using Illumina’s GenomeStudio software as follows:
   5. In GenomeStudio select ‘Full Data Table’ tab.
   6. Click on ‘Column Chooser’ icon.
   7. In Displayed Columns select ‘Name’, ‘Chr’,'Position’ & all samples. In Displayed Subcolumns select ‘GType’, ‘X’ and ‘Y’.
   8. Hit OK then click on ‘Export displayed data to a file’ icon
   9. Multi-Scanner protocol. Genome Studio
5. Output GenoeStudio SNP details to determine and set orientation:
   1. Select “SNP Table” and use column chooser to select SNP details: Name, Chr, Position, SNP, Plus/Minus Strand, ILMN Strand
   2. Output file as reference
   3. Illumina SNP A/B alleles can be updated using “SNP” variable, where A=first allele and B=second allele (e.g. [T/G] would update A to T and B to G).
   4. To set all SNPs as TOP, select SNPs that have “ILMN Strand” =”BOT” for flipping
   5. Step performed after z-call update complete
6. **SOP: Running zCall**
7. Run Basic QC Filters on Gencall Data, filtering must be done in order to obtain accurate results (quality in, quality out)
   1. Basic Filter Based on Subject Call rate (zCall package script),

**python qcReport.py -R my.input -C 0.99 > my.qc.input**

* 1. The GenomeStudio report can be converted/extracted into PED for more detailed Sample level QC using plink via the zCall python script:

**convertReportToTPED.py**

Samples identified by detailed QC can be removed via the zCall python script:

**dropSamplesFromReport.py**

***Note:*** *Per zcall recommendations, there is no need to run SNP exclusions at this stage – zCall internally performs a strict SNP QC as part of the script findMeanSD.py (incl. call rate > 99%, MAF > 5%, HWE > 0.00001).*

1. Find the mean and standard deviations of each homozygote cluster for common SNPs (zCall python script)

**python findMeanSD.py -R my.qc.input > my.mean.sdD**

***Note:*** *Calculation performed only SNPs that pass a strict SNP QC as part of the script findMeanSD.py (incl. call rate > 99%, MAF > 5%, HWE > 0.00001).*

1. Find the relationship between the X and Y intensities for both the means and standard deviations (zCall r script)

***Rscript findBetas.r my.mean.sd my.betas***

1. Find the thresholds (Z value) used for calling No Calls:

**python findThresholds.py -R my.qc.input -B my.betas -Z 7 -I 0.2 > my.thresholds.7**

1. Recall all No Calls (zCall python script)

**python zCall.py -R my.input -T my.thresholds.7 -O my.output**

1. Calibrating Z: zCall authors have found that using the default Z value of 7 works well for most datasets. The steps below, can be used to optimize the value of Z:
   1. Find thresholds for various values of Z (zCall bash code/shell script)

**for z in $(seq 3 15); do**

**python findThresholds.py -R my.qc.input -Z $z -B my.betas > my.thresholds.$z**

**done**

* 1. Calculate the accuracy of a given value of Z

**for z in $(seq 3 15); do**

**python calibrateZ.py -R my.qc.input -T my.tresholds.$z > calibrate.$z**

**done**

* 1. Choose value of Z: Once all of the statistics for each value of z have been calculated, choose the one that maximizes the global concordance while still have acceptable sensitivity for calling heterozygote calls and false positive rates.

1. Post zCall
   1. Convert to Plink ped format using

**convertReportToTPED.py**

* 1. Update chip TOP allele annotation to forward strand of build 37/hg19
  2. Perform post call QC

**References**:

<http://wgsaboston.wordpress.com/2014/01/14/genotype-calling-zcall-with-external-data/>

<http://wgsaboston.wordpress.com/2014/01/14/genotype-calling-zcall-with-broad-intensity-files/>

<https://github.com/jigold/zCall>

**EXOME-CHIP QUALITY CONTROL SOP Version 5, 2012-11-20 (Exome-chip QC SOP v5.pdf)**