Using the Sensitive Aneuploidy Detection (SAD™) Tool

John McGuigan, Ning Wan

Sample	13	18	21
Batch02_Sample01	(Normal, 0.993)	(Normal, 1.356)	(Normal, -0.656)
Batch02_Sample02	(Undetermined, 2.758)	(Normal, 0.254)	(Normal, 0.875)
Batch02_Sample03	(Normal, 0.088)	(Normal, 1.927)	(Trisomy, 8.470)
Batch02_Sample04	(Normal, 1.158)	(Normal, 0.483)	(Trisomy, 11.443)
Batch02_Sample05	(Normal, 1.566)	(Normal, 1.122)	(Normal, -0.392)
Batch02_Sample06	(Normal, 1.261)	(Normal, 0.680)	(Normal, -0.370)
Batch02_Sample07	(Normal, 0.307)	(Normal, 1.688)	(Trisomy, 7.565)
Batch02_Sample08	(Normal, 1.485)	(Normal, 0.392)	(Trisomy, 9.436)

Example analysis results for three chromosomes (13, 18, 21) in eight samples

Introduction

The SAD tool (Figure 1) is designed for analysis of low-coverage whole-genome-sequencing samples in order to detect whole-chromosome aneuploidies. The tool is streamlined for very simple operation and the normalization process allows for detection of aneuploidies even with a high background of non-aneuploid data. A set of at least 30 normal samples ("modeling controls") is used to set up an analysis model which automatically accounts for biases in coverage across the chromosomes. The model is based on comparing "ratio" values for each tested chromosome to a normal distribution of these "ratio" values in the controls, resulting in normalized chromosome values (NCVs) similar to z-scores. Cutoffs are used to make calls based on the NCVs. The model can be tested for sensitivity and specificity using samples with known genotypes.

One possible application of this tool is for non-invasive prenatal testing (NIPT). This kind of testing has been used in several studies with data from Illumina® (1,2), Ion Torrent™ (5), and Pac Bio (6) sequencers and is based on sequencing of maternal plasma samples to detect fetal aneuploidies.
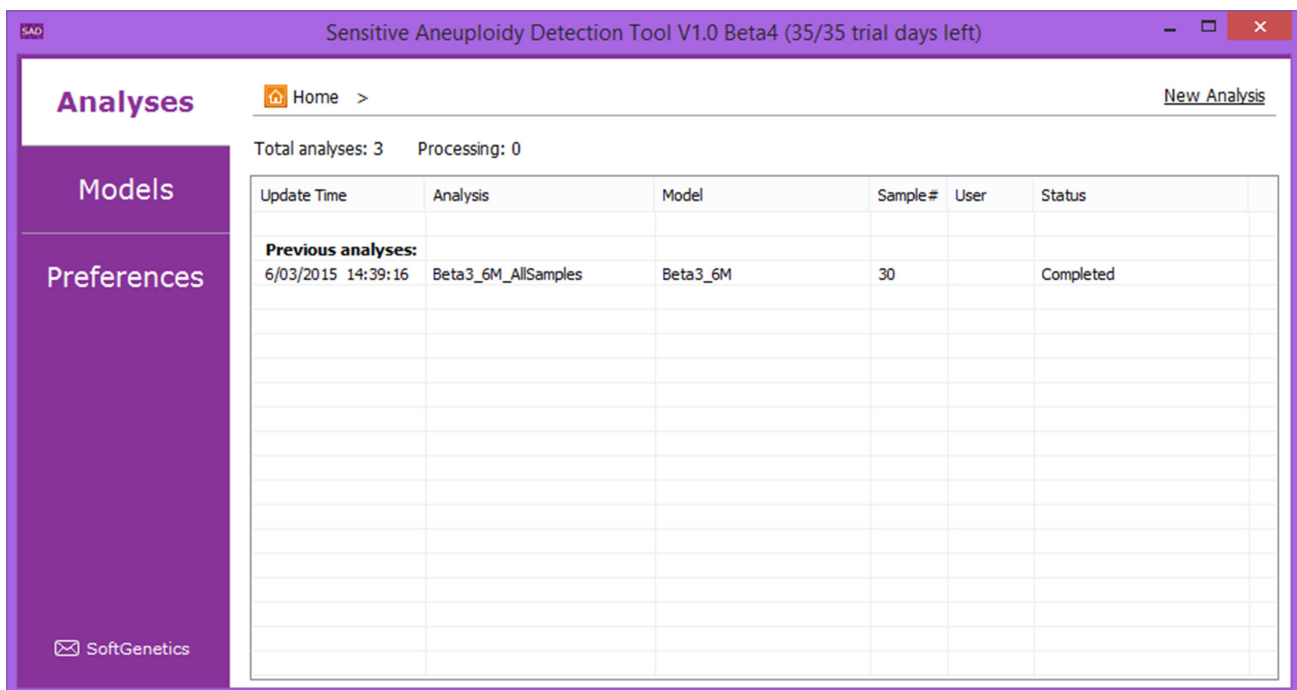

Figure 1: The main window of the SAD Tool

SOFTGENETICS®
Software PowerTools for Genetic Analysis

NextGENe®
Next Generation Sequencing Software

Procedure

Configuration via the 'Preferences' Tab

The preferences section (figure 2) contains four settings:
- Workspace Directory – the location of analysis results
- Optional Custom Model directory – the location of saved model information.
  - By default this is the same as the workspace directory.
- Genome Reference
  - A reference has been created for use with the SAD tool and can be downloaded from:
    ftp://SAD:sadtool@ftp.softgenetics.com/SAD_Build_37.zip
  - The reference should be extracted, resulting in a folder named "SAD_Build_37".
  - This setting should point to the folder containing the "SAD_Build_37" folder.
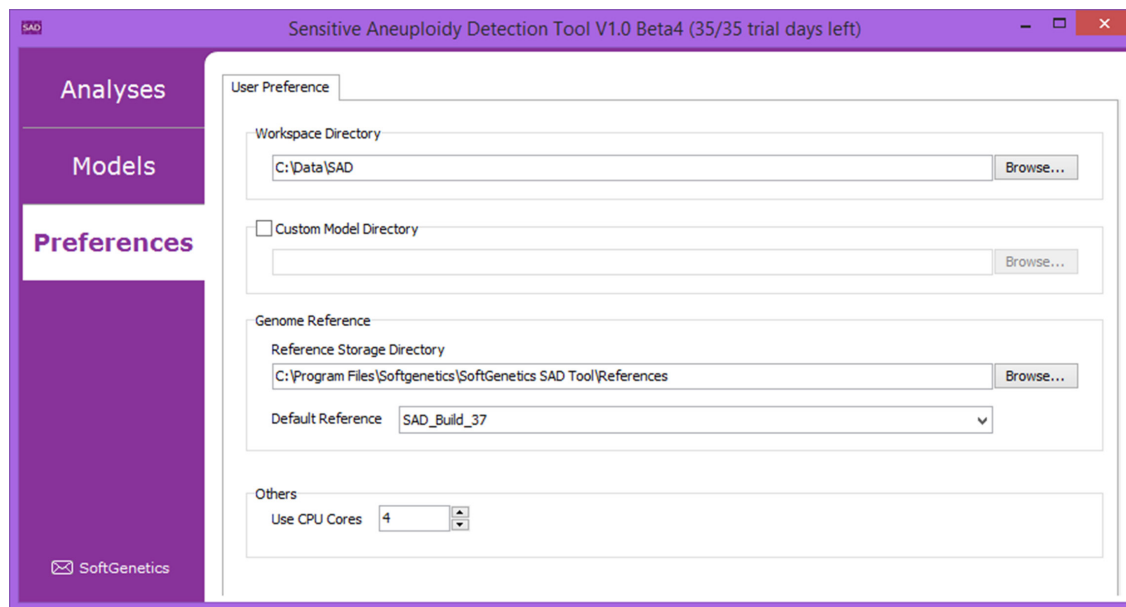- Number of CPU cores to use during processing

Figure 02: SAD Preferences Section

Setting up a Model

The "Models" section allows users to create new control models or examine existing ones. Figure 3A shows the interface for creating a new project before running, and 3B shows the same view after the model has been run. Creating a new model has the following steps:

- Naming the model (or using the default name)
- Adding modeling (negative) controls
  - If analyzing chromosomes X and/or Y the gender of modeling controls must be specified.
- Optionally adding test (positive or negative) controls
- Selecting which chromosomes will be used for aneuploidy analysis (or using the default 13, 18, and 21).
- Adjusting NCV cutoff values for making calls. These can also be adjusted after running.

The tool will automatically align, count reads, and perform GC correction (LOESS-based correction using 50kb bins). Analysis results are shown for the model in a pane at the bottom of the window. Results for each tested chromosome are displayed on a separate tab. Clicking on a sample in the list will highlight the results for that sample to in order to allow for fast visual comparison of the NCV values. Each chromosome results tab includes (from left to right):

- Summary information on the calculated ratios, which chromosomes were used as part of the normalization procedure, and the p-value from the Shapiro-Wilk normality test.
- A boxplot of the ratios in the control samples
- A quantile-quantile plot comparing the NCV distribution to a normal distribution
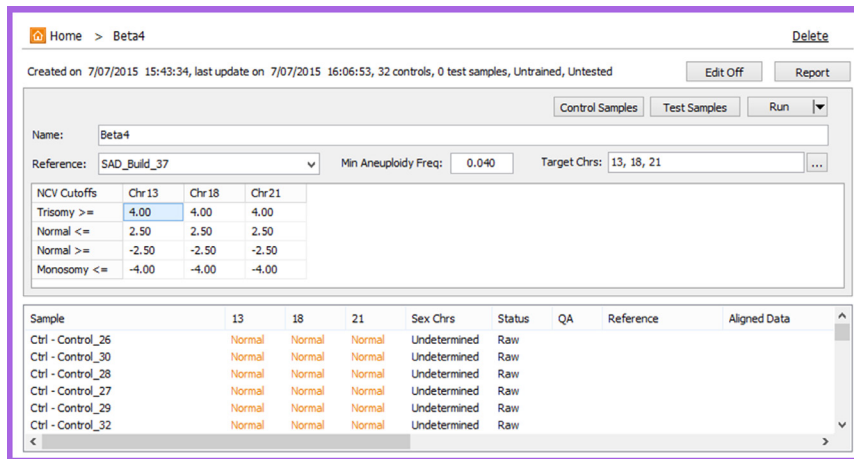- A plot of the Normalized Chromosome Values (NCVs) for each control sample

SoftGenetics LLC 100 Oakwood Ave. Suite 350 State College, PA 16803 USA
Phone: 814/237/9340 Fax 814/237/9343
www.softgenetics.com email: info@softgenetics.com

SOFTGENETICS®
Software PowerTools for Genetic Analysis

NextGENe®
Next Generation Sequencing Software

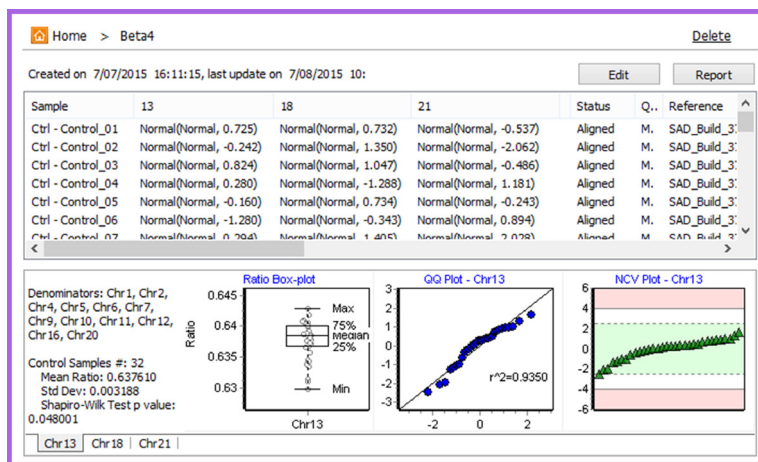**Figure 3A: New Model Dialog before Running**

Figure 3B: New Model Dialog after Running

Running an Analysis

The "New Analysis" button in the "Analyses" section can be used to set up analysis of untested data. The samples to be tested (aligned data or raw data) are loaded and the previously created model is selected. Running several analyses at once will automatically create a processing queue, and progress can be observed in the run window. Detailed progress information is available by clicking on the text above the progress bar (figure 4). Final results for a run are shown in figure 5.
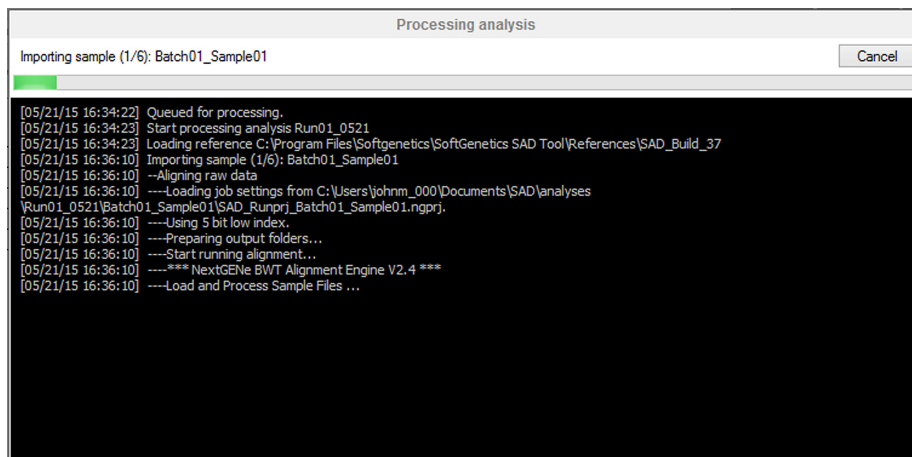

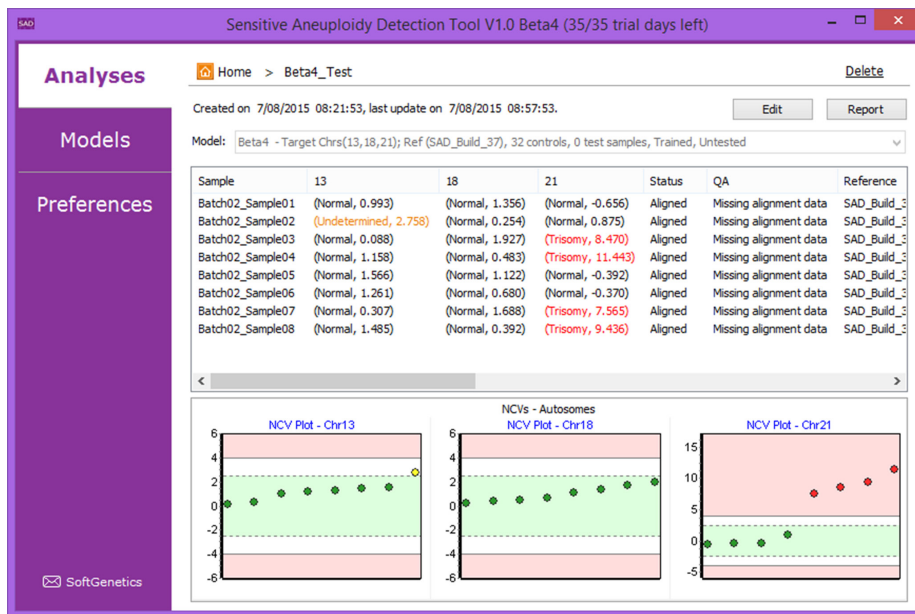**Figure 4: Detailed Processing Log**

Figure 5: Results after running an analysis.

Results

100% sensitivity was seen for eight chr21 and two chr18 trisomies when testing non-invasive prenatal testing (NIPT) data randomly down sampled to 6 million reads per project. Thirty-two normal controls were used in the model. No other aneuploidy calls were made in the 30 tested samples, although two "Undetermined" calls were made. These calls could be successfully made "Normal" without a loss in sensitivity (of the known positive samples) by adjusting the default NCV cutoff values slightly.

Discussion

Quality Control

In the model view or test results view it is possible to hover the mouse on the "QA" column and see a pop-up that provides some basic quality information (Figure 6). This includes quality trimming and alignment performance, a plot of the raw basecall quality per base, and a small plot showing GC correction results.
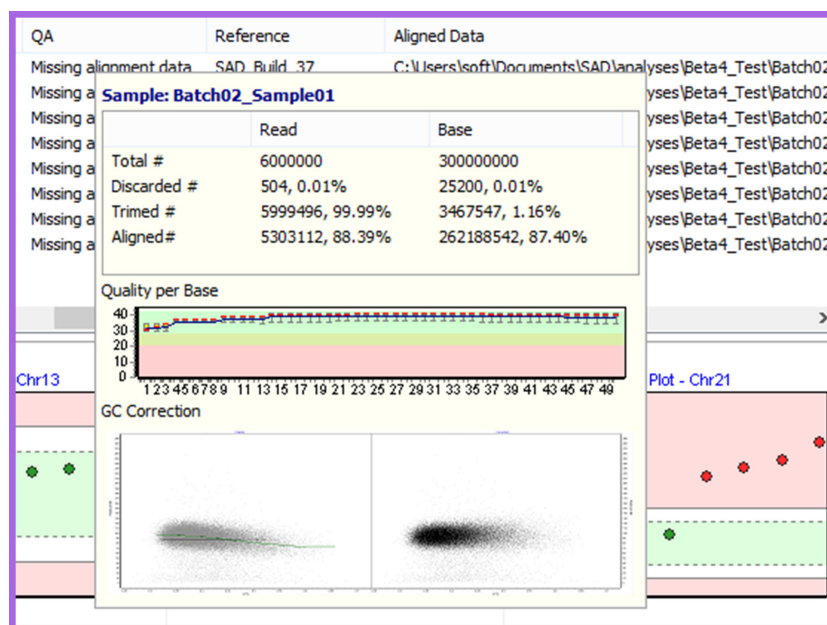

Figure 6: QA Summary Section

Normalization and Detection

The theoretical basis for the statistical test is that ratio values for a given chromosome are normally distributed amongst the different samples or controls and the presence of an aneuploidy will increase the ratio value relative to the value found in normal control samples. The extent of this increase is related to the mean ratio, the standard deviation, and the fraction of the sample derived from cells with aneuploidy. Models are selected based on these parameters. An expected anuploid fraction can be set (default = 4%) so that positive samples which increase coverage by this amount or more will (on average) have an NCV value of eight or more. Combined with the default cutoff of 4, this allows for positive samples to (on average) have ratio values 4 standard deviations from the cutoff. This would theoretically allow sensitivity and specificity greater than 99% if the distribution of ratios in tested samples is the same as that in the controls.

Cutoffs used for making a call can also be adjusted. Lowering the cutoff will increase sensitivity at the expense of specificity, while increasing it will do the opposite.

It is highly recommended to:
- Run multiple positive and negative controls in order to determine the actual sensitivity and specificity of any given model.
- Test samples to determine an expected aneuploidy fraction. For example, measuring the fetal fraction present in NIPT samples. Sensitivity in samples with lower (< 8%) aneuploidy fraction is generally lower than in samples with a higher fraction (4). For example, approximately 10 million reads per sample are needed for sensitive detection of trisomy in chromosome 21 when the aneuploid fraction is around 4% (3).

Detailed Analysis Results

Finished Analysis projects have an "Edit" button that can be used to re-run the statistical analysis only, or both the alignment and statistical analysis. There is also a "Report" button which allows access to the raw calculation data (starting with read counts) for validation purposes. Larger plots of the GC correction results are also available in this section (Figure 7).
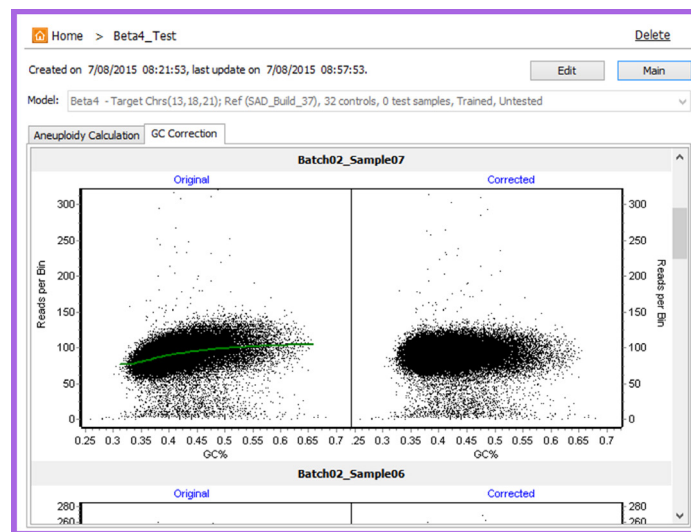

Figure 7: GC Correction Results

Testing the X and Y chromosomes

Testing of chromosome X and Y is also possible, but sensitivity and specificity may be lower than for the default chromosomes (21, 18, 13). It is suggested to have a large number of modeling controls of each gender in order to obtain accurate results. Testing is performed in the same manner as other chromosomes, but against male and female controls individually. Different calls are made against male and female controls for both ChrX and ChrY (table 1), and then these are combined into a single call.

Control Gender, Chromosome	Male, ChrX	Male, ChrY	Female, ChrX	Female, ChrY
Detected Increase	XX+	YY	XXX	Y+
Detected Normal	X	Y	XX	-
Detected Decrease	-	-	X	-
Undetermined	Undetermined	Undetermined	Undetermined	Undetermined

Table 1: Potential individual calls for ChrX and ChrY Analysis

SOFTGENETICS®
Software PowerTools for Genetic Analysis

NextGENe®
Next Generation Sequencing Software

Conflicting calls like an increase of ChrX relative to male controls ("XX+" for two or more copies of X) and a decrease of ChrX relative to female controls ("X" for a single copy of X) result in "Undetermined" results for that chromosome. Compatible calls are combined into a single call. For example, "Y" from male controls (normal = one copy) and "Y+" from female controls (increased = one or more copies) results in a call of a single copy of Y. This is because the male controls can distinguish an increase or decrease from a single copy, while the female controls can only detect the presence of one or more copies.

References

[1] Bianchi, Diana W., et al. "Genome-wide fetal aneuploidy detection by maternal plasma DNA sequencing." Obstetrics & Gynecology 119.5 (2012): 890-901.

[2] Bianchi, Diana W., et al. "DNA sequencing versus standard prenatal aneuploidy screening." New England Journal of Medicine 370.9 (2014): 799-808.

[3] Fan, H. Christina, and Stephen R. Quake. "Sensitivity of noninvasive prenatal detection of fetal aneuploidy from maternal plasma using shotgun sequencing is limited only by counting statistics." PLoS One 5.5 (2010): e10439.

[4] Palomaki, Glenn E., et al. "DNA sequencing of maternal plasma to detect Down syndrome: an international clinical validation study." Genetics in medicine 13.11 (2011): 913-920.

[5] Rava, Richard P., et al. "Circulating fetal cell-free DNA fractions differ in autosomal aneuploidies and monosomy X." Clinical chemistry 60.1 (2014): 243-250.

[6] van den Oever, Jessica ME, et al. "Single molecule sequencing of free DNA from maternal plasma for noninvasive trisomy 21 detection." Clinical chemistry58.4 (2012): 699-706

SoftGenetics LLC 100 Oakwood Ave. Suite 350 State College, PA 16803 USA
Phone: 814/237/9340 Fax 814/237/9343
www.softgenetics.com email: info@softgenetics.com

SOFTGENETICS®
Software PowerTools for Genetic Analysis

NextGENe®
Next Generation Sequencing Software