Manual CSA

Comparative Sequence Analysis (CSA) is a simple alternative method for directly comparing base-lined raw data from equivalent sequencing reactions on different samples without the need for base calling. The control sequence electropherogram is split into its four component traces (one for each base), which are overlaid with equivalent traces from a second reaction. The resulting format allows simple and rapid visualisation of any differences (mutations) between the two sets of traces. Because the assay is based on sequencing, mutations can be characterised as well as identified. Moreover, because there has been minimal software manipulation of the data, heterozygous mutations that are difficult to pick up using standard sequencing analysis techniques have been shown to be easily detectable in a single direction using CSA. Finally, and perhaps most importantly from a diagnostic point of view, CSA is a simple and rapid technique for verifying normal sequence data with a very high degree of confidence.

Platforms

Manual CSA can be performed on data generated by either ABI 377 or ABI 3100 platforms. The analysis is carried out using the Genescan software; on a MAC for the 377 and on a PC for the 3100. Because there are differences in the way the two versions of the software work, slightly different approaches are needed relating to the use of size standards.

In order to horizontally align sequences so that the individual base specific traces can be overlaid a size standard is defined for each lane analysed. The size standard can be in one of two forms:

- Internal size standard (MAC analysis only)
- External size standard (MAC or PC analysis)

Internal size standards

Internal size standards can only be used with the MAC version of GeneScan. In general this means analysis of runs performed on ABI 377 machines. It is however possible to analyse 3100 data using an internal size standard by converting the .ab1 files to MAC format using an applet available from Applied Biosystems. Sequencing reactions are carried out run as normal.

The size standard is created artificially during analysis by assigning scan numbers to each of the peaks one of the coloured traces (e.g. Red representing the C trace) in the control sample. Since this type of size standard is fragment specific, a new size standard will need to be made for each different fragment being analysed

The colour that represents the size standard is indicated by a diamond (eg red).

1. To change the size standard colour for a single sample hold down apple(MAC)/Ctrl(PC) and click the box representing the desired colour.

2. To change the size standard colour for all the samples hold down apple(MAC)/Ctrl(PC) and click the coloured box at the top of the column representing the desired colour.

3. Select ONE of the boxes representing a size standard (this will become coloured), and click the 'Analyse' button.

4. Open the analysed file by double clicking the appropriate file name listed in the 'Sample File ' column. Ignore the warning box: Click 'OK'.
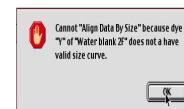
5. Click the 'Data Point' column header – this will highlight the whole column. Now copy the whole column by holding down apple(MAC)/Ctrl(PC) and pressing 'C' on the keyboard. Close the analysed file.

6. Open the drop menu in the 'Size Standard' column that relates to the sample you have analysed – select 'Define New'.

	B	G	Y	R	Sample File	Define New	ndard	▶	Parameters	▶
1			◈		Water blank 2F	● <None>		▶	<Analysis Parameters>	▶
2			◈		Water blank 2R	ais		▶	<Analysis Parameters>	▶
3			◈		Normal female 2F	vhl		▶	<Analysis Parameters>	▶
4			◈		Normal female 2R	<None>		▶	<Analysis Parameters>	▶

7. In the table select the top cell in the 'Size' column (a). Scroll to the bottom of the table and select the bottom cell whilst holding down the 'Shift' key (b) this will select the whole of the 'Size' column. Paste the size data into the column by pressing 'V' whilst holding down the apple(MAC)/Ctrl(PC)key. Check that the numbers in the 'Data points column match their partners in the 'Size' column (c). Name and save the size standard.

(a)

Peak	Data Points	Size
1	1225	O
2	1272	O
3	1359	O
4	1455	O
5	1545	O
6	1598	O
7	1800	O

(b)

Peak	Data Points	Size
2	1272	O
3	1359	O
4	1455	O
5	1545	O
6	1598	O
7	1800	O
8	1862	O

(c)

Peak	Data Points	Size
1	1225	1225
2	1272	1272
3	13	1359
4	1	55
5	5	5
6	1598	1598
7	1800	1800

External size standards

These comprise a set of fluorescently labelled fragments that range in size from about 50 bp to 500 bp (or greater if required) in 50 bp increments. This type of standard is run in every lane of sequencing (by addition to the loading buffer) and is entirely distinct from the sequencing reaction. These can be used in conjunction with analysis on either MAC or PC and generally give better alignment than internal size standards.

Since the sequencing reaction itself requires four colour collection (one for each base) an external size standard has to be labelled in a fifth colour. With ABI 377 platforms (MAC), Collection software version 2.6 or above and GeneScan version 3.0 or above are required in order to collect and analyse data comprising five colours. All software associated with the ABI 3100 is compatible with 5 colour collection.

ABI GeneScan LIZ500 (PN 4322682) is a suitable size standard for sequences up to 500bp (0.2ul per reaction added to the loading buffer). Sequences of greater length can be analysed using LIZ500 although the alignment becomes less accurate in relation to the length beyond 500bp. Alternatively if greater size range is required an 'in house' size standard can be made by fluorescently labelling a range of suitably sized fragments. Bodipy 630/650 (Molecular Probes; Leiden, Netherlands) or other dye with equivalent absorption/emission spectra is suitable for this purpose.

To set the machine up for using a LIZ size standard (5 colour collection) run matrix standard DS-02 following the instructions in the kit to define dye set E5. The version of DS-02 will depend on the platform being used. If a 'home made' size standard is to be used run matrix standard DS-01, with the addition of the fifth colour size standard, following the instructions in the kit to define dye set E5. Again the version of DS-01 will depend on the platform being used. Dye set E5 should be used for all subsequent CSA runs involving a LIZ size standard.

Values for the size standard peaks are assigned according to the standard procedure for GeneScan (peaks at 250 and 340 should not be used). Because this size standard is generic it can be used for all future CSA runs.

Handling Files.

ABI 377

The 377 outputs gel file incorporating data for all fragments run. This can be opened from within Genescan by using 'File…Open…Gel'. Once the gel file has been opened in GeneScan it is irreverably altered so that it can no longer be opened in Sequencing analysis. For this reason it is desirable to make a copy of the Gel file before opening; prefixing one with 'csa' and one with 'seq'.

The format of the sample sheet is different for GeneScan and Sequencing Collection. Genescan uses four lines per sample rather than one for the Sequencing sample sheet. This means that the extraction process will incorrectly assign sample names from the sample sheet to the data (lane 2 will be called sample 5 on the Sequencing sample sheet. To prevent this a new GeneScan sample sheet will need to be made using 'New…Sample Sheet'. This should then be applied to the gel using 'Gel...Install New Sample Sheet'.

ABI 3100

The 3100 outputs individual files for each analysis. In order to open them in GeneScan the extension .ab1 needs to be changed to .fsa.   Once this is done a usable .ab1 file can not be recreated. For this reason it is desirable to make a copy of the .ab1 files into a separate folder – for convenience call this folder fsa.   He file extensions can be rapidly changes using the DOS command prompt.

1. Make a note of the path (eg C:\Documents and Settings\My Documents\csa\run070703\fsa) – if there is an address bar in your browser this can be copied this by selecting it and 'Ctrl+C'.

2. Open the DOS command line by selecting 'Run' on the start menu. Type 'cmd' in the dialogue box and click ok. There should now be a black window with the command prompt (C:\ > or something similar). Navigate to the fsa folder containing the copied .abi files by using the change directory command – C:\ > cd [path name] then press return. The path name can either be typed long hand or pasted from the clipboard by right clicking the mouse and selecting paste (you can not use 'Ctrl V'). There should now be a command prompt including the full address of the fsa folder containing the copied .abi files (e.g. C:\Documents and Settings\My Documents\csa\run070703\fsa >). All the .ab1 files can be renamed using a single rename command – > ren *.ab1 *.fsa

3. Close the command window.

4. The .fsa files can be opened in GeneScan as normal.

Analysis

If samples do not appear in the Analysis Control in the same order as listed in the sample sheet, the order can be changed using 'Settings…Project Options…Sample File Sorting' (this is not essential). Before analysis perform the following:

1. ***Install a matrix:*** *(this is only applicable to 377 data and may not be necessary. Try without first; the GeneScan will give a warning if this step is required).* Select all the files by clicking the 'Sample file' column header. Install the appropriate sequencing matrix by selecting 'Gel…Install New Gel Matrix'.

2. ***Check Analysis Parameters:*** Double click one of the <Analysis Parameters> in the right hand column.
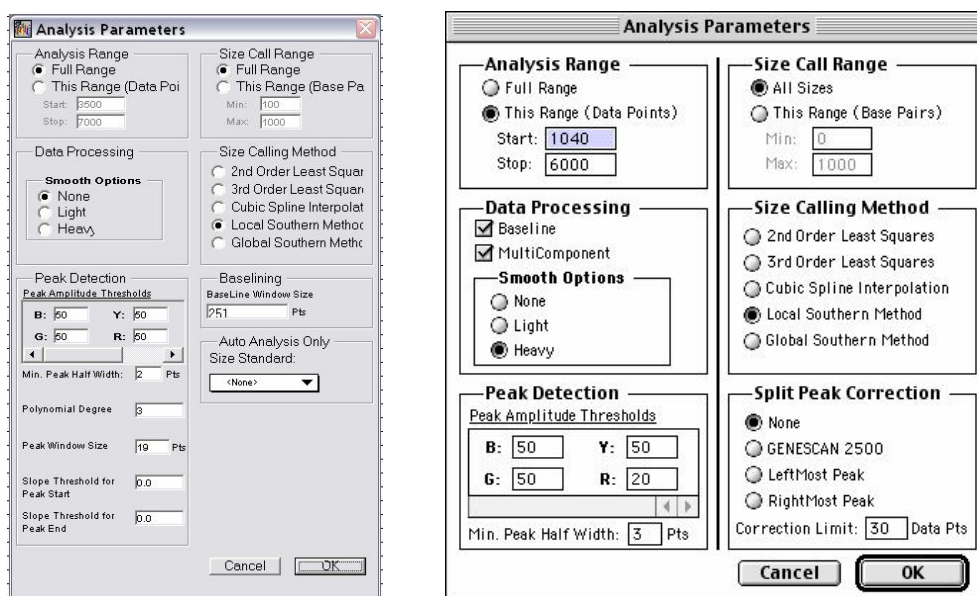
 Analysis Range – ensure this covers all data points on your gel.

 Data Processing – make sure both the 'Baseline' and 'Multicomponent' options are ticked. It is up to you what 'Smooth Option' you use (I use 'Heavy').

 Peak Analysis Thresholds – Should be set for at least 50 for each colour. Peaks below this threshold are liable to be unreliable.

 Size Call Range – select the 'All Sizes' option.

 Size Calling Method – This does not make much difference, in general use 'Local' or 'Global' Southern.

 Split Peak Correction – select 'None'.

3. ***Apply an appropriate size standard to each sample***: If using internal size standards a different one will be required for each fragment (refer to internal size standards for details). Only one size standard is required for all samples if an external size standard is being used.

Viewing CSA traces.

1. **Open results control:** Select 'Results Control' from the window menu

2. **Check setup:**

   ➢ Check there are 4 panels

   ➢ Ensure 'Quick Tile' is on.

   ➢ Select 'Custom Plot Colors' under 'project Options' in the 'Settings' menu. Check that colour 1. Is 'Black' and colour 2. is 'Red'.

   ➢ Select 'Results Display' under 'Preferences' in the 'Settings' menu. In the 'Default Display Settings' select 'Align By Size', 'Show Legends' and "Custom Plot Colours'.

   ➢

3. **View CSA:** Select two samples to compare by clicking on the numbered boxes on the left hand side of the results control. With the 'Quick Tile' on this will automatically place the four coloured traces in separate panels. Click 'Display'. If the first sample selected is always the control, this will always appear black in the results display. The second sample selected should appear red.


   Note: The PC version of Genescan has a bug – when the results dis[ply is opened samples are not displayed with the 'custom colours' even if this has been set as default.

4. **Multiple alignments:** More than two samples can be selected for multiple alignments. Colours displayed can be controlled using 'Settings… project Options…Custom Plot Colors'. Colours must be adjusted manually each time a display is opened by selecting 'Veiw…Plot Colors'.

   Note: The PC version of Genescan has a bug sometimes buttons within the 'Analysis Control' or the 'Results Control' do not do what they are supposed to do (e.g. 'clear all' opens the print menu. This can be rectified by ensuring both the programme window and the control window are the default size or are maximised.

5. *What* things *do:*
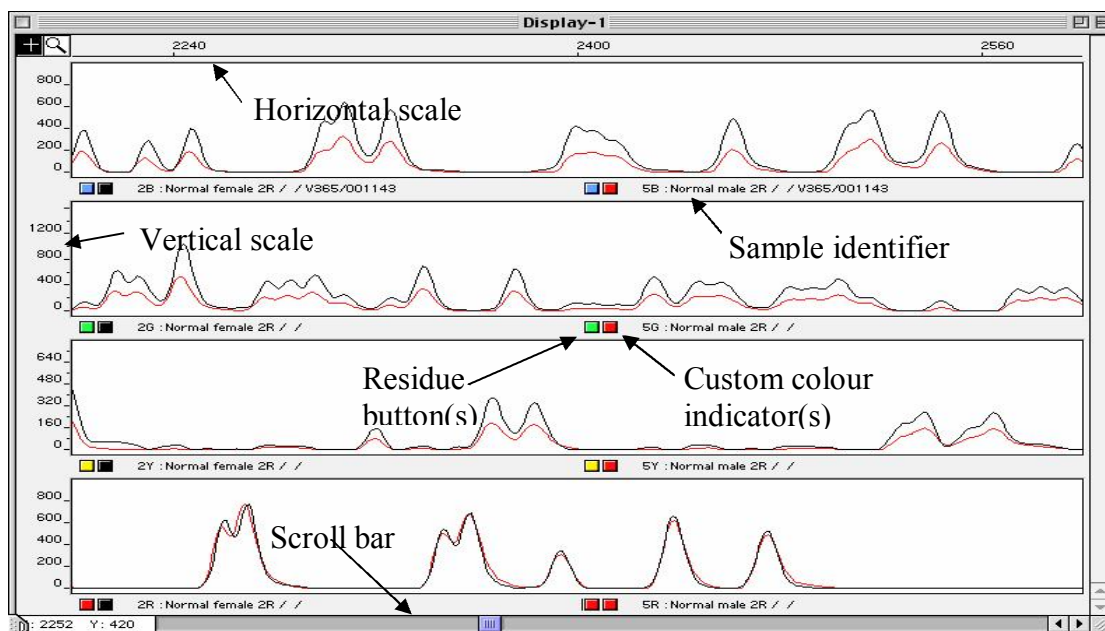
➢ *Horizontal scale:* defines what region of the traces are visible – Double clicking the mouse over this area ⌗ opens a dialogue box in which the horizontal scale can be changed. A simpler way to zoom in and out is to hold down apple(MAC)/Ctrl(PC) and pressing + (zoom in) or – (zoom out).

➢ *Vertical scale:* (one for each panel) all four of these work the same as the horizontal scale (double click when the cursor changes to a ⌗)

➢ *Residue button:* shows the true colour of the traces represented by the peaks in the panel above. NB: Colours used to represent bases in CSA are not the same as the standard sequencing colours (View colour table)

   Note: that the two residue buttons in each panel are the same colour. Therefore the top panel represents the BLUE traces (or G residues), the next the GREEN traces (A residues), the next the YELLOW traces (T residues) and the bottom represents the RED traces (C residues).

➢ *Custom colour indicator:* this in conjunction with the *sample identifier* shows which sample is represented by which coloured trace in the panel above. In the example above the black traces represent a 'Normal female' and the red traces represent a 'Normal male'.

   Note: if GeneScan has been set up as described above, the first sample selected in the results control will be shown in BLACK and the second sample will be shown in RED. It is simplest to always use one colour (i.e. BLACK) to represent the 'normal' leaving RED for the 'test' sample.

➢ *Scroll bar:* enables scrolling along the traces in zoomed in view.

6. ***Scaling traces (vertical alignment):*** To alter the height of peaks in a particular trace double click the appropriate residue button. This opens a dialogue box. Entering 2 in the box doubles the height of the peaks in that trace whereas entering 0.5 will halve the height of the peaks – any number can be entered in this box. The peaks of the RED trace in the bottom panel have been scaled by a factor of 2 by double clicking on the right hand residue button below the panel. This has aligned the RED and BLACK traces on top of one another.

7. ***Mutations:*** Heterozygous mutations will always be characterised by two indicators:

   (a) A 'half height' peak indicating the 'lost' base.▼

   (b) A 'lone' peak representing the mutant base.▼

C>A point mutation   2 Bp deletion (frame shift)
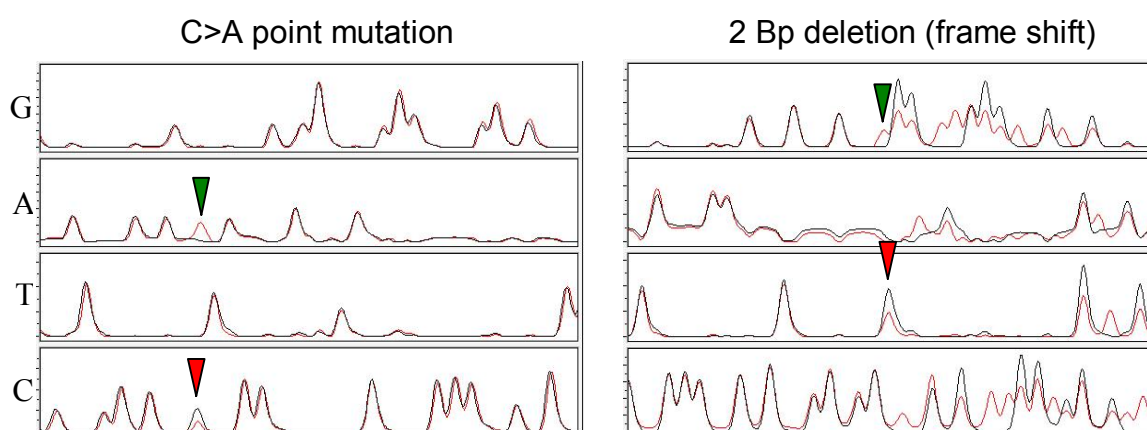
8. ***To view another sample:*** Close the current 'Display window' – this will return you to the 'Results Control'. Click 'Clear All' to remove your previous selection and select a new pair of samples as previously described.

9. ***Printing:*** You can only print what is shown on the screen. To do this select 'Page Setup' from the 'File' menu. Make sure 'Paper' is set to 'A4' and 'Orientation' is 'Landscape'. Now select 'Print One" from the 'File' menu.

REMEMBER THE DATA IS JUST SEQUENCE DATA. It is therefore not possible to have more than one peak at any one position (except in the case of a heterozygous mutation). If traces *do* have more than one peak at various positions it is likely that the data is 'noisy' and therefore unreliable. To confirm this double click on the relevant sample name in the 'Analysis Control', this opens up all four traces in a single panel and should look like normal sequence data. It will be clear from this view if the data is 'noisy'.

Appendix A: CSA Base Colour Table

Base	Sequencing	CSA
A	Green	Green
C	Blue	Red
G	Yellow (Black)	Blue
T	Red	Yellow (Black)

Appendix B: Web links

NGRL (Wessex): http://www.ngrl/wessex/workprogrammes.html

Molecular probes: http://www.probes.com/servlets/product?item=10004

Applied Biosystems: http://www.appliedbiosystems.com/products/