

AutoSNPa walk-through

Data files

A SNP data file produced by the Affymetrix system is in the form of tab-delimited text. The file's **.xls** file extension will cause it to be opened by Excel by default. However, AutoSNPa can only use the flat text file, and cannot use data saved as an Excel spreadsheet. If you have saved your data as an Excel file, you should convert it back to tab-delimited text, via the **File...Save As** menu option in Excel.

When opened as a spreadsheet, the data are arranged in a number of columns, containing the allele frequencies, SNP position data, names and genotypes. The AutoSNPa programme does not use most of this data, much of which is therefore superfluous. The minimum dataset required by the programme is shown in **Table 1**. Each of these columns must contain the information indicated, while other columns can be deleted from the data file. [Exception: if you wish to make frequency files, used by Merlin, then the columns titled **Freq A Cauc**, **Freq A Asian** or **Freq A AfAm** should be left in the file. Although the position of these columns is not critical, the column title must be included.]

Column	Data
1st	SNP index
2nd	SNP ID
3rd	SNP database RS ID
4th	Chromosome
5th	Physical distance (bp)
7th	Genotype data

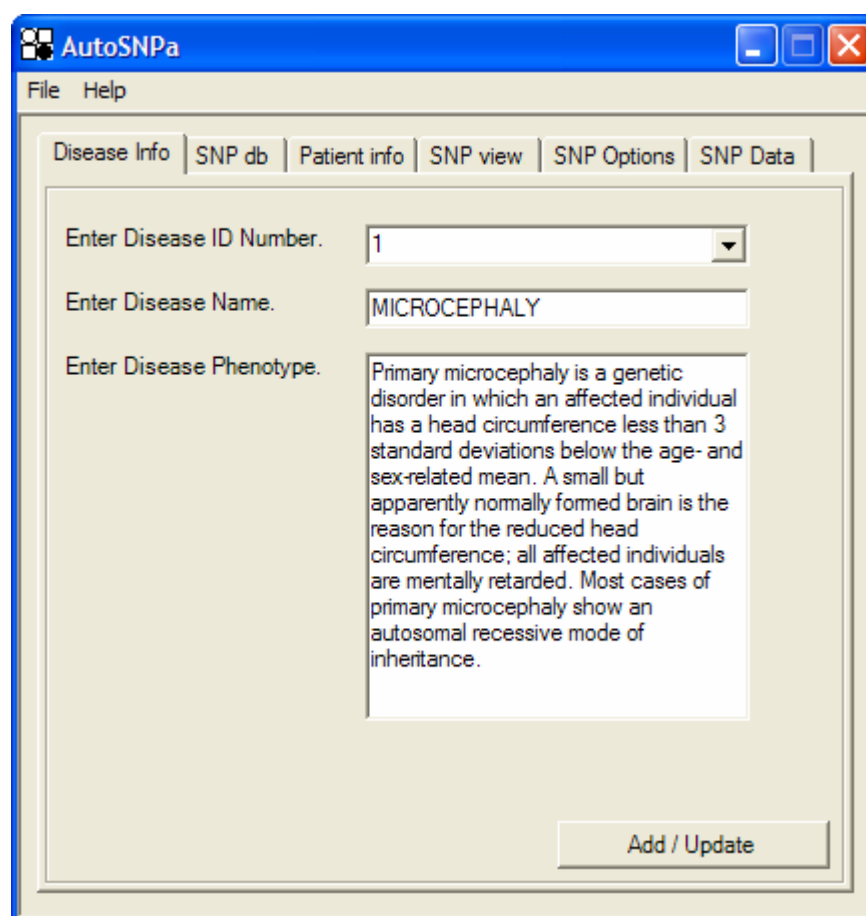
Table 1: Minimum file structure needed by AutoSNPa.

By default, SNP data are displayed with respect to physical position (5th column). If you wish to display markers relative to their Marshfield genetic positions, then the column entitled **marshfield avg** must also be included within the file, with its correct heading.

Creating the SNP database

Entering disease information

Before creating the AutoSNPa database, you must enter a disease ID number; additional disease information is optional. This information is accepted using the [Add/Update](#) button (**Figure 1**), and is used solely to verify that the correct file has been opened during subsequent data analysis.



The screenshot shows the AutoSNPa application window. The title bar reads "AutoSNPa" and includes standard window control buttons. The menu bar contains "File" and "Help". Below the menu bar is a tabbed interface with six tabs: "Disease Info", "SNP db", "Patient info", "SNP view", "SNP Options", and "SNP Data". The "Disease Info" tab is selected. It contains three input fields: "Enter Disease ID Number." with a dropdown menu showing "1", "Enter Disease Name." with a text box containing "MICROCEPHALY", and "Enter Disease Phenotype." with a text area containing a detailed description of primary microcephaly. At the bottom right is an "Add / Update" button.

Figure 1: The AutoSNPa opening screen.

Choosing SNP database production options

Underlying AutoSNPa is a relational database, consisting of a number of data tables. These tables are automatically configured, with the exception of the SNP information table. Every 3 months, Affymetrix updates its SNP information dataset, and removes SNPs that have failed quality control procedures. This may result in your Affymetrix

files containing different SNP datasets, even if your samples were sent for analysis together. However, there is considerable overlap between datasets. **Figure 2** shows the [SNP db](#) tab; this offers you three database configuration options:

1. **Data from single file:** if this option is chosen, the SNP information table is created from a single Affymetrix data file (selected by clicking the [Select SNP data file...](#) button). Genotyping data for a particular SNP in patient data files are only used if this file contains naming and positional data for that SNP.
2. **Data pooled from all files:** If this option is chosen, SNP information is pooled from all the data files entered into the database. Therefore, if SNP information is absent from one data file but present in a second, the SNP information table will be updated using information from the second file. This option results in use of the largest possible number of SNPs. However, including SNPs that have subsequently failed quality control may result in SNPs being placed in the wrong region, or erroneous genotypes being included in the database.

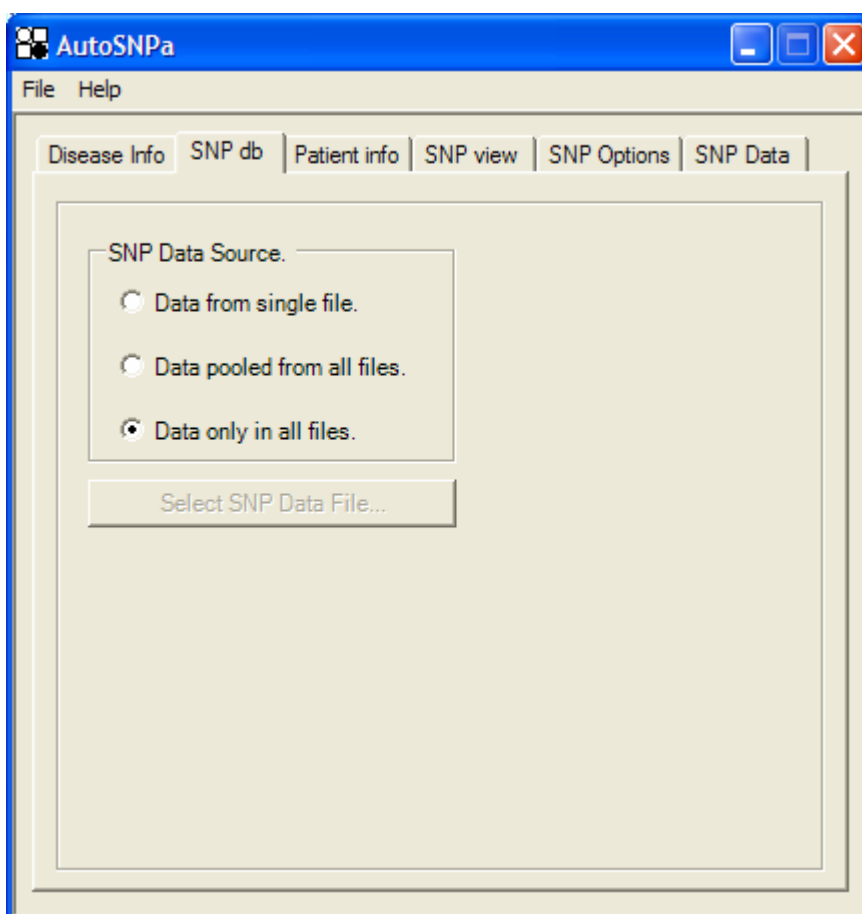


Figure 2: SNP data table configuration page.

3. **Data only in all files** (default): This option constructs the SNP data table from information common to all the data files. If one file doesn't contain information about a specific SNP, then genotypes for that SNP are not used in further analysis, even if all other files do contain information about that SNP. This option reduces the chance that the programme uses incorrect genotyping data, and while it also yields the smallest SNP dataset, in practice this doesn't create any problems.

Addition of pedigree and patient information

Pedigree and patient information can be entered either graphically or via the keyboard; both of these methods are accessed through the [Patient info](#) tab (**Figure 3**).

The screenshot shows the 'AutoSNPa' application window with the 'Patient info' tab active. The form contains the following data:

Disease ID Number.	1
Patient ID.	24
First Name.	Ian
Last Name.	Carr
Father ID.	15
Mother ID.	16

Gender: Male Female Affected?

Get Patient Data File... S0309_xba142_150305.xls

To add data graphical click Pedigree.

Buttons: Pedigree, Info, Add, Find, Delete

Figure 3: [Patient info](#) tab. This page is used to enter a person into the database; in this instance the individuals with IDs 15 (father) and 16 (mother) have an affected son (ID 24) named Ian Carr, whose genotype data are in file S0309_xba142_150305.xls.

Pedigree construction with the keyboard

To construct the pedigree database using the keyboard, first enter the individual's unique ID (this must be a number), first name and last name, and then select their parents from the [Father ID](#) and [Mother ID](#) drop-down lists. (If one of the two parents has not been added to the database, enter "0" for that parent.) If you wish to create a Merlin pedigree file, then all individuals must have either both parents in the pedigree or have neither (*i.e.* mother is "0" and father is "0"). The [Father ID](#) and [Mother ID](#) lists only contain the ID numbers of individuals previously entered into the pedigree; this requires an individual's parents to be entered, then that individual, and then their children. Although this constraint helps to reduce errors in pedigree construction, it can become difficult to construct complex consanguineous pedigrees in this manner. After entering the individual's sex and disease status, you can link him or her to an Affymetrix data file by clicking the [Get Patient Data File...](#) button. This file is only processed after the [Add](#) button is clicked.

Once entered, an individual's information can be modified by selecting his or her ID number from the [Patient ID](#) drop-down list and clicking [Find](#). If edited, the information is updated after clicking the [Update](#) button (which was previously labelled [Add](#)). If you change the Affymetrix data file linked to this person, the original genotyping data are discarded and the new data added. If you select a patient ID from the list and click the [Info](#) button, a dialog box appears giving a brief summary of the individual's nuclear family and SNP genotype data. Individuals can be removed from the database using the [Delete](#) button. However, this action does not remove the individual's descendants, and so may result in incorrectly formatted Merlin pedigree files and a pedigree that will not be drawn correctly.

Graphical construction of a pedigree

Clicking on the [Pedigree](#) button of the [Patient info](#) tab opens a second form (**Figure 4**), which allows you to create a pedigree graphically. If the database already contains patient information, this form will show the pedigree.

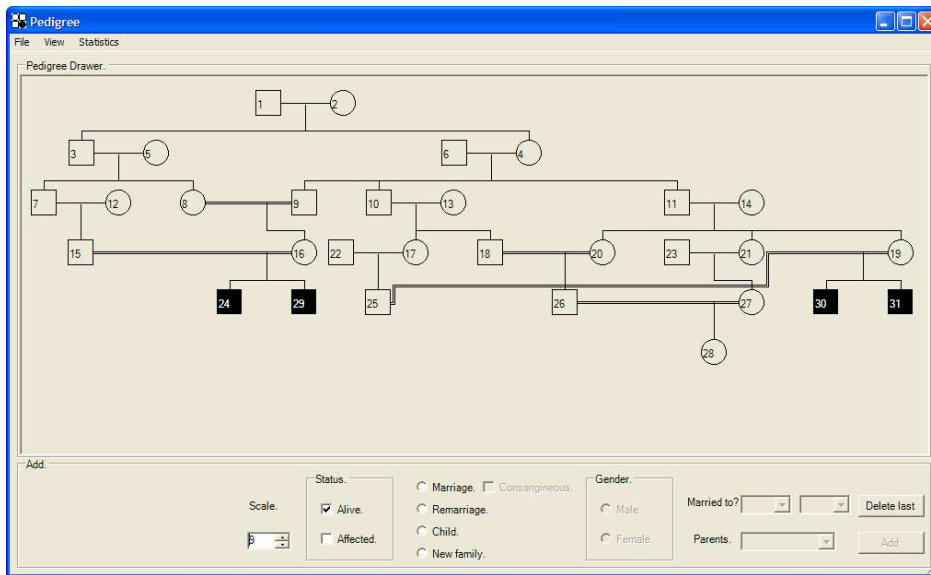


Figure 4: The [Pedigree](#) form, displaying a consanguineous pedigree.

Adding a new family

If no individuals are present in the database, the form is blank, and a founding couple must be added by clicking the [New family](#) radio button (**Figure 5**). It is possible to add more than one new family in this way, and the descendants of each family can marry.

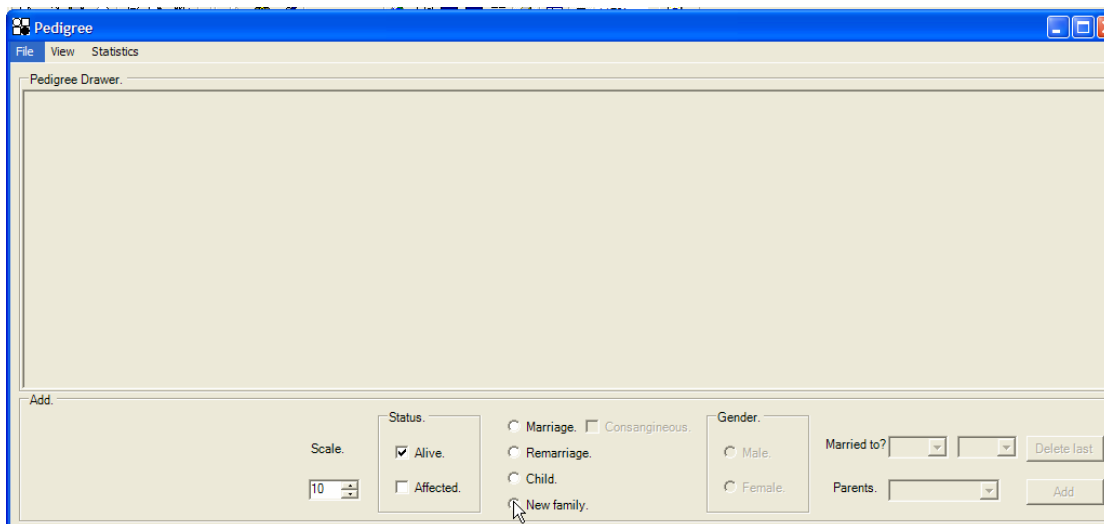


Figure 5: Addition of a new family.

Adding children

Children can be added once a founding couple has been added. First, click on the [Child](#) radio button above the [New family](#) radio button. This enables the [Gender](#) radio buttons; after selecting the child's sex, choose its parents from the [Parents](#) drop-down list. This list contains all marriages previously entered into the pedigree, each as two numbers linked by a hyphen. The first number is the husband's ID and the second is the mother's ID. Clicking the [Add](#) button (**Figure 6**) draws the child. When adding a new individual, you can specify whether that individual is affected and/or alive. This can also be done after the individual has been added, but it may be a useful option when drawing the early part of a pedigree in which the majority of the individuals have died, or when adding an affected child to a large family.

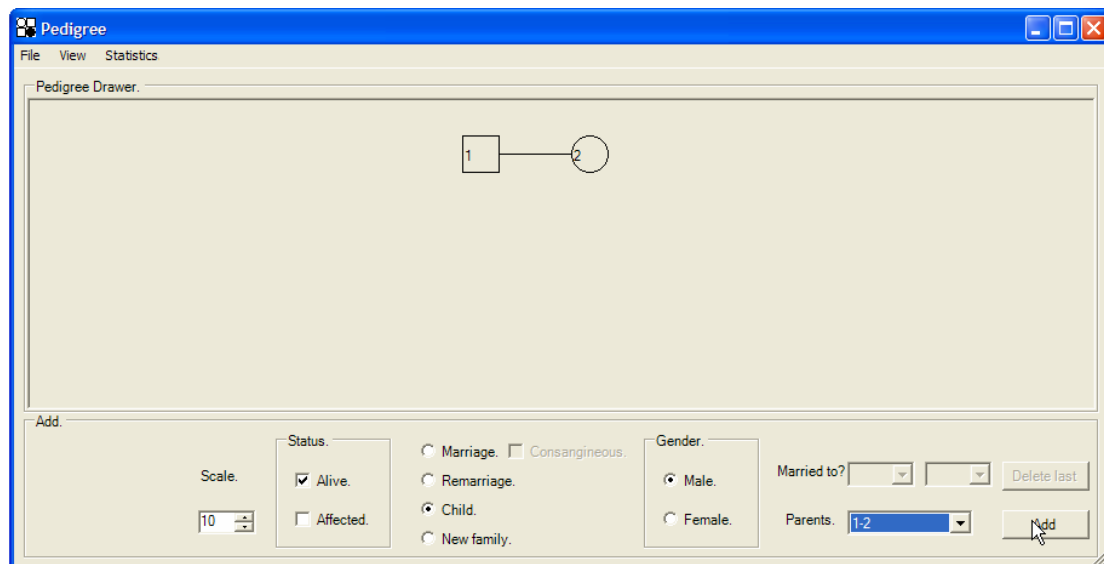


Figure 6: Adding a child. In this instance, a male unaffected child who is still alive is added to the family of individuals 1 (father) and 2 (mother).

Adding marriages to the pedigree

Once children have been added, it is possible to add marriages to the pedigree. Click the [Marriage](#) radio button and then select the sex of the individual to be added; this enables a list box opposite the [Married to?](#) label. If you choose to add a female to the pedigree, the left hand box is enabled, allowing you to pick any unmarried males in

the pedigree (**Figure 7**). The marriage is added by clicking the [Add](#) button. Children can be added to these marriages in the same way as to the original founder family.

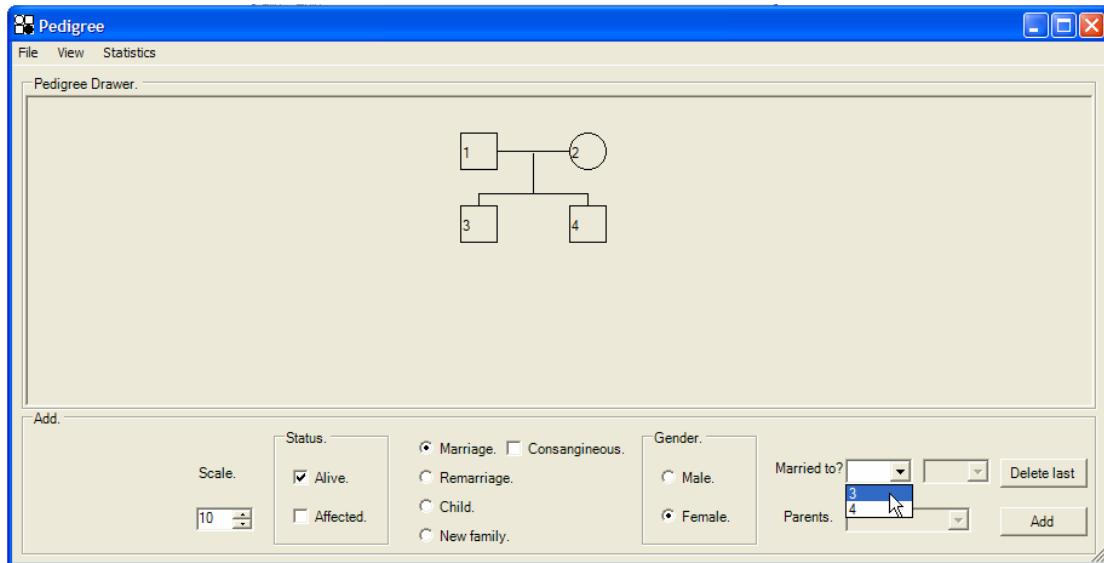


Figure 7: Adding a marriage to the pedigree. The figure shows the creation of a marriage between a new female and individual 3.

Adding a consanguineous marriage

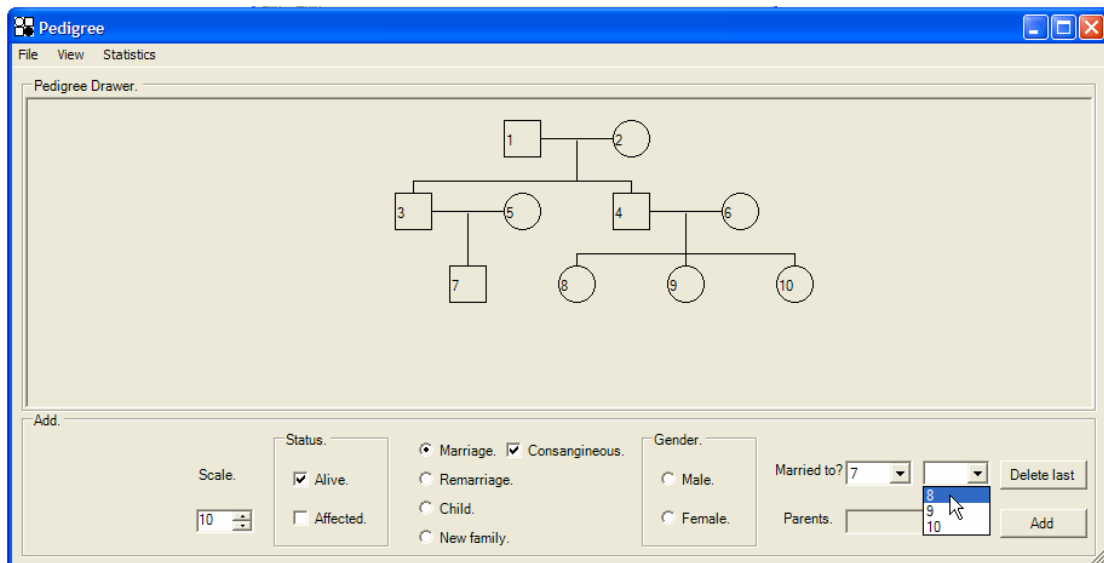


Figure 8: Adding a consanguineous marriage. The figure shows the addition of a consanguineous marriage between husband 7 and wife 8.

As more individuals are added to the pedigree, it becomes possible to create consanguineous marriages. To do this, select the [Marriage](#) radio button and then check the [Consanguineous](#) check box next to it. This enables both list boxes next to the [Married to?](#) label. Select the husband and wife from the list boxes and click [Add](#) to create the marriage (**Figure 8**).

If children are added to a consanguineous marriage, they always appear to the left of the right-most parent (**Figure 9**).

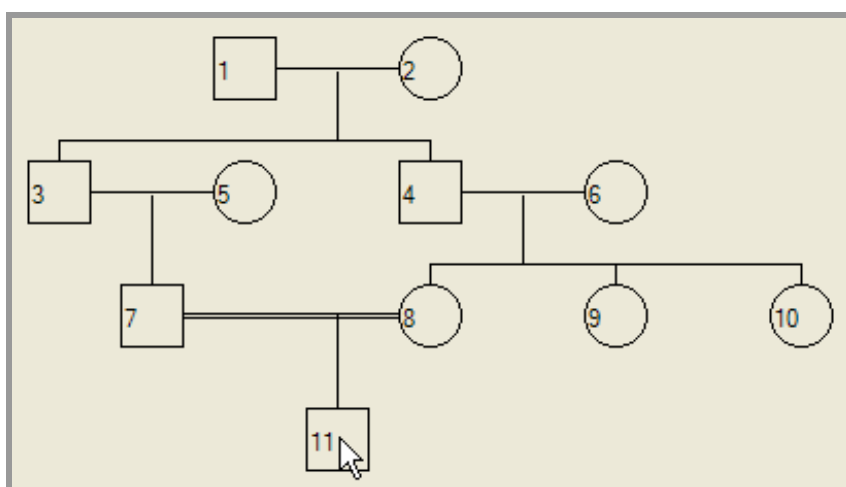


Figure 9: Children of consanguineous marriages are drawn to the left of the right-hand parent.

Adding a second marriage

To add a second marriage to the pedigree, click the [Remarriage](#) radio button, to display the form shown in **Figure 10**. Select the person you wish to remarry and then click the [New Person](#) radio button, followed by [Accept](#). This will add a new individual to the pedigree, married to an individual who is already married. (If you select an individual who is not already married, then that person will become married as if you selected the [Marriage](#) radio button described earlier.)

It is also possible to create a second marriage between individuals already present in the pedigree. This is done as described above, but instead of clicking [New Person](#),

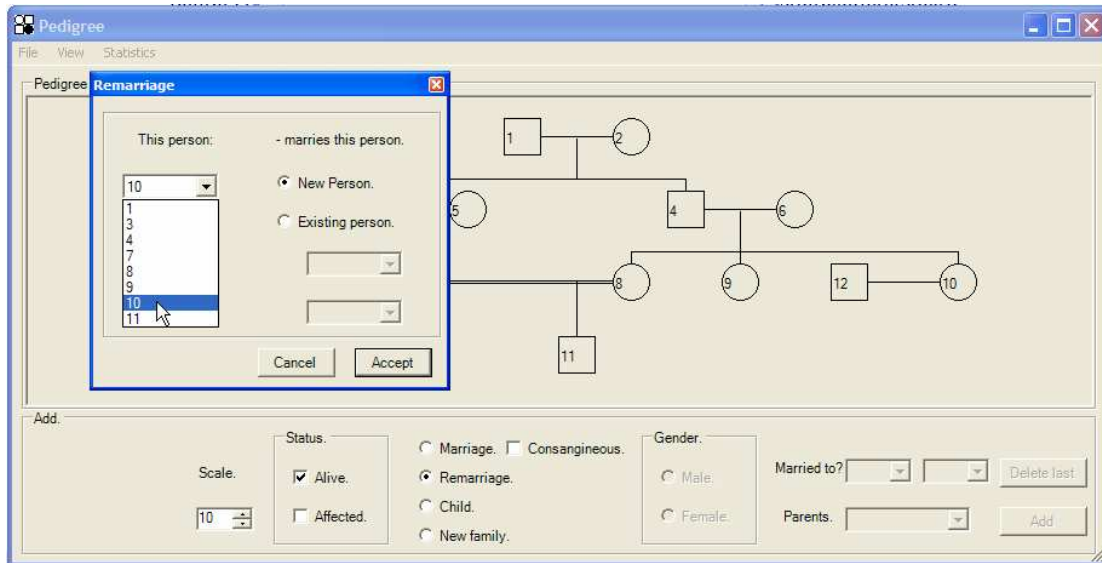


Figure 10: Adding a second marriage to the pedigree. The figure shows the addition of a second marriage between individual 10 and a new individual.

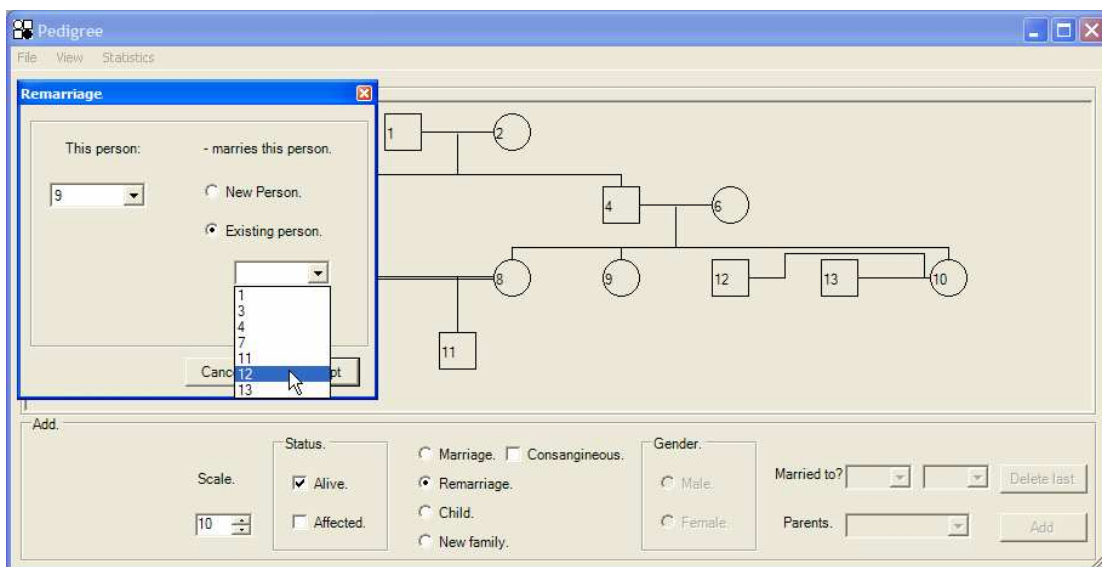


Figure 11: Addition of a second marriage between two pre-existing individuals. In this example, individual 12 is marrying his first wife's sister, individual 9.

select the [Existing person](#) radio button and then choose a partner from the enabled list box (**Figure 11**). The only limitation to forming remarriages is that at least one of the

individuals must be genetically related to other members of the pedigree (*i.e.* in **Figure 12**, individual 6 could not remarry individual 13).

Figure 12 shows the pedigree created above, with the addition of children to each marriage, to illustrate where each child is added. Child 14 is the son of 12 (father) and 9 (mother), child 15 is the daughter of 12 (father) and 10 (mother) and 13 (husband) and 10 (wife) have a daughter 16.

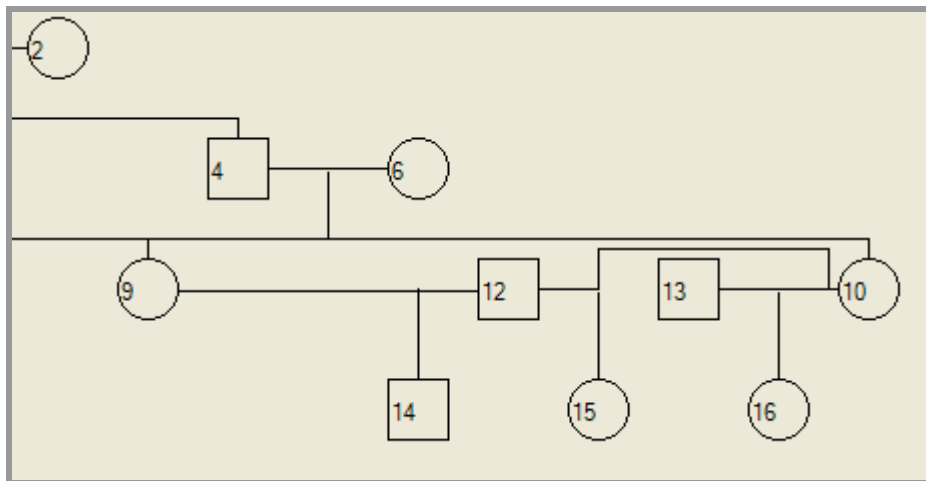


Figure 12: Position of children arising from remarriages within a pedigree.

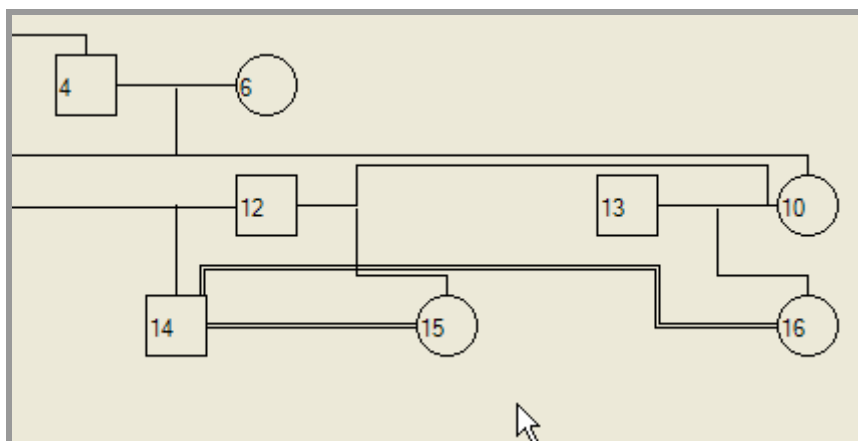


Figure 13: Remarriage between consanguineous individuals.

It is also possible to form second marriages that are consanguineous; the programme automatically detects when two individuals would produce a consanguineous union and adds double lines to signify it as a consanguineous marriage (**Figure 13**).

If multiple consanguineous marriages exist within the pedigree, it is possible that the lines showing the marriages will lie on top of each other, as shown in **Figure 14**.

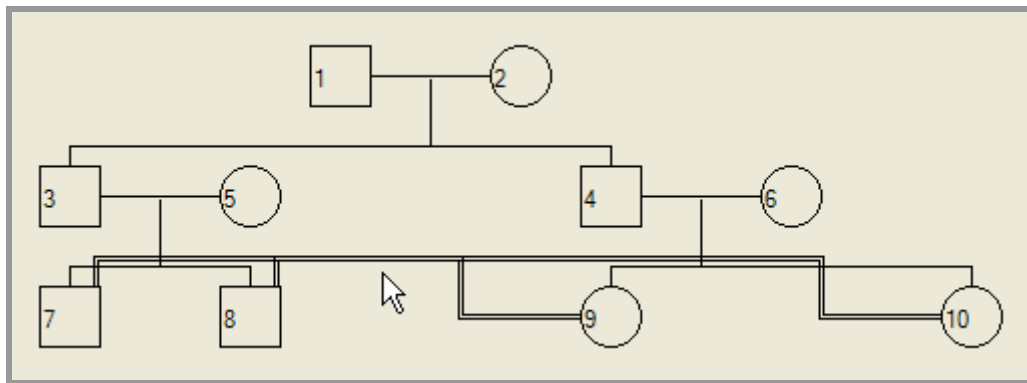


Figure 14: Example of the overlapping of two consanguineous unions, making it impossible to tell whether individual 9 is married to individual 7 or 8.

To overcome this problem, left-click on the symbol of an individual involved in the confusion; this displays the [Edit details](#) dialogue box (**Figure 15**). Using this dialogue,

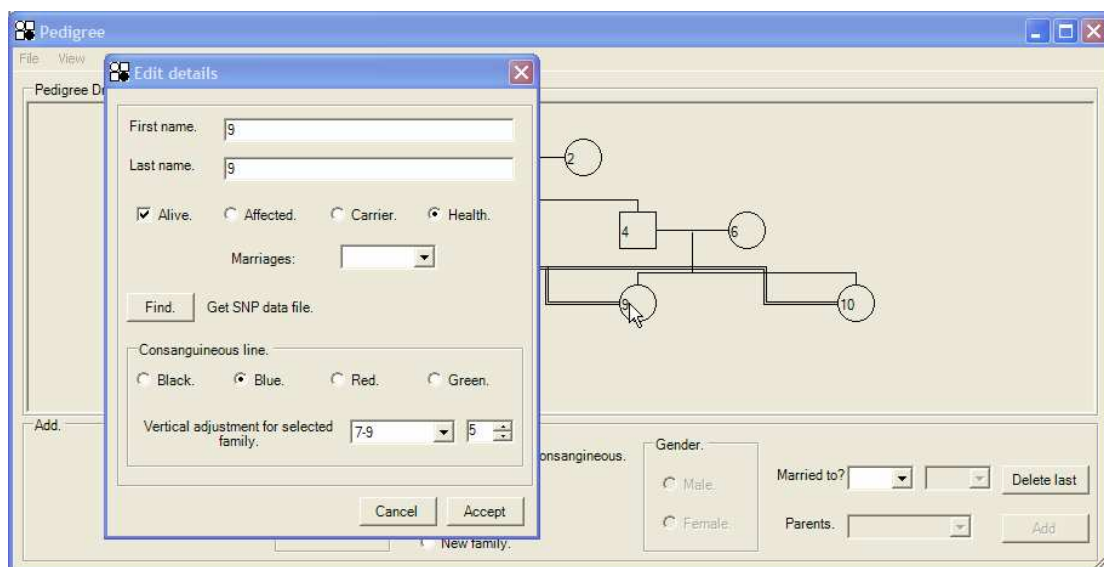


Figure 15: [Edit details](#) dialog box.

you can select one consanguineous union from the drop-down list next to the [Vertical adjustment for selected family](#) label. You can then either change the colour of the consanguineous union lines (radio buttons) or the vertical displacement of the line (using the box to the right of the marriage selection box). **Figure 16** shows the result of performing the adjustments selected in Figure 15.

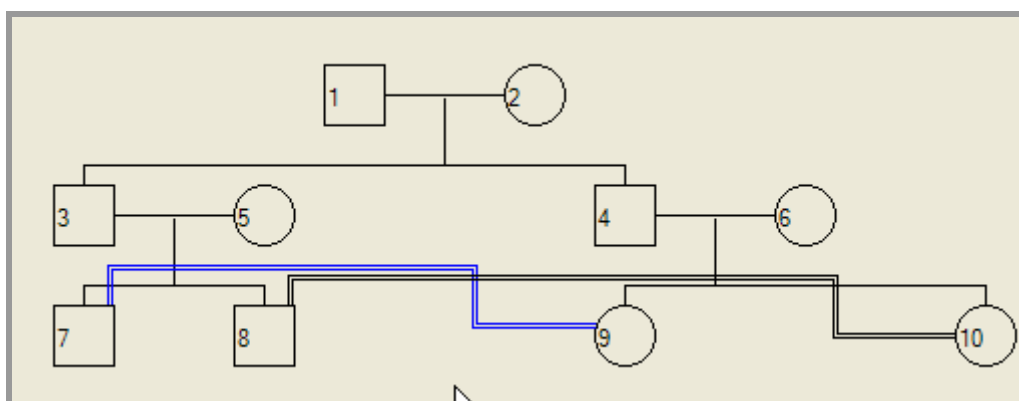


Figure 16: The display after the changes shown in Figure 15 were accepted.

Editing the details of individuals in a pedigree

The [Edit details](#) dialogue box can also be used to edit an individual's details. As above, left-clicking on an individual opens this dialogue box, as in Figure 15. It is now possible to change the disease status of the individual to [Healthy](#), [Carrier](#) or [Affected](#), as well as assigning him or her as [Alive](#) or not. It is also possible to name the individual. (By default, the first and last names are set to the individual's unique ID number.) The [View](#) menu on the [Pedigree](#) form can be used to select display of ID numbers or names of family members (**Figure 17**).

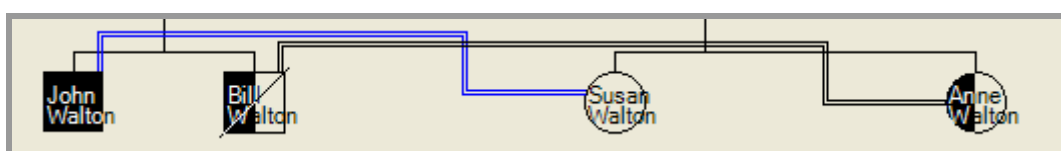


Figure 17: Affected, carrier and healthy individuals identified by their full

Disease status probabilities within the pedigree

If an individual has his or her phenotype set (using the [Edit details](#) dialogue), the probability can be calculated of his or her descendants being affected or a carrier.

Figure 18a shows a pedigree in which founder 1 has been assigned as a carrier, while **Figure 18b** shows the same pedigree with two carriers assigned.

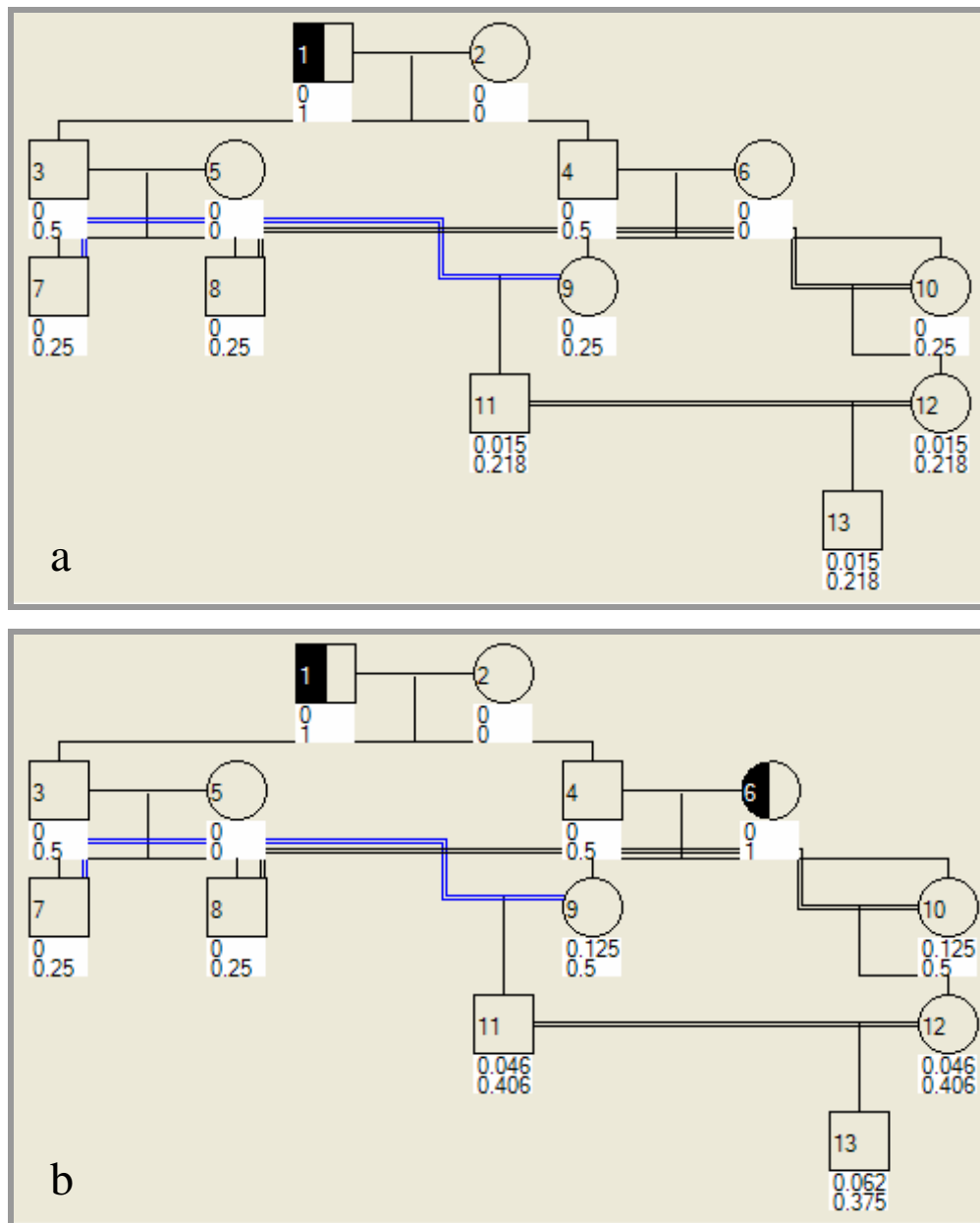


Figure 18: Display of probabilities that individuals in the pedigree are carriers or affected, by virtue of either individual 1 (a) or individuals 1 and 6 (b) being assigned as carriers.

The upper number is the probability that the individual is affected and the lower number that he or she is a carrier. For example, in Figure 18a, individual 11 has probabilities of 0.015 of being affected and 0.218 of being a carrier, while in Figure 18b, the same individual has probabilities of 0.046 of being affected and 0.406 of being a carrier. (These figures are truncated, not rounded, so that the exact values for individual 11 in Figure 18b are in the ranges 0.04600 to 0.04699 for being affected and 0.40600 to 0.40699 for being a carrier.)

Using the [Statistics](#) menu of the [Pedigree](#) form, the view can be changed to display any of the following: no probability data, the chance that an individual is a carrier, the chance that an individual is affected, the chance that an individual is either affected or a carrier, or (as in Figures 18a and 18) both the affected and the carrier probabilities.

Linking Affymetrix data files to individuals

In the [Edit details](#) form, click the button labelled [Find](#). This will allow you to browse to the location of the data file and link it to the individual. If you make a mistake, you can unlink the file by clicking the [Reset](#) button, which appears to the right of the file's name. No data are added to the database until the [Pedigree](#) form is closed. To close the form, use the [File](#) menu to either discard all the pedigree information or add it to the database. Only if the [Save and close...](#) menu option is used are the data saved to the database. If you edit a pedigree you made earlier, only the data files that you have changed will be read. Creating the database may take a couple of minutes, depending on the number of files to be read and the speed of the computer. During this time, the programme will stop responding to the mouse and keyboard, but its activity can be seen as a busy hard disk.

Pedigrees with more than one founding family

If a pedigree contains more than one founding family, members of different families can marry by selecting them as described above. If two families intermarry a number of times, AutoSNPa can detect which marriages are consanguineous and then draw them appropriately (**Figure 19**). If individuals in each family are designated as affected or carriers then AutoSNPa can still calculate the probabilities as described above (**Figure 20**).

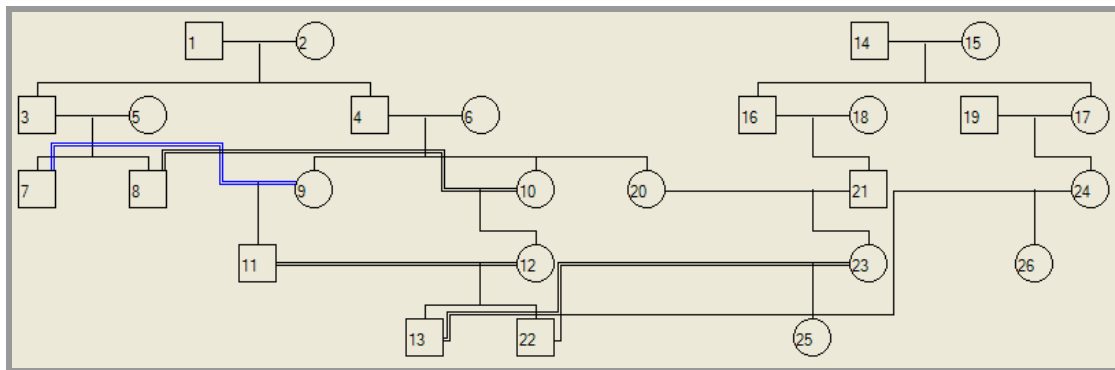


Figure 19: Two founding families (based on marriages 1-2 and 14-15) intermarry multiple times. Marriages 21-20 and 13-24 are non-consanguineous, while the union between individuals 13 and 23 is consanguineous.

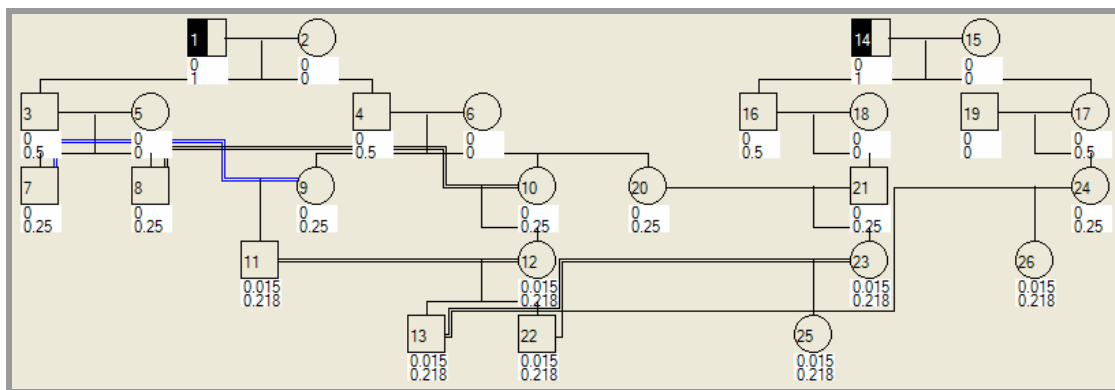


Figure 20: The chances of individuals in the two families being affected or carriers, as described in Figure 18.

Moving and changing the scale of the pedigree display

Families can be moved around the [Pedigree](#) window by holding down the left mouse button and dragging. The size of the pedigree can also be changed using the value in the [Scale](#) box.

Notes on drawing pedigrees

- This programme is primarily intended for SNP analysis, and so only marriages that contain children may be redrawn when a pedigree image is produced from a previously saved file.
- While the programme is able to draw complex pedigrees, it is possible to produce a pedigree that it is unable to draw. However, these pedigrees tend to include marriages between the children of multiple incestuous unions. In our experience, these kinds of pedigree rarely occur.
- If the window is resized, the image may appear cropped. To see all the individuals; move the pedigree in the window as described above.

Analysis of SNP data

Once the SNP database has been constructed, it is possible to view the SNP data. However, the speed at which the database can be accessed is vastly increased if it is first saved to disk and then reopened. You really should do this! Use the [File](#) menu on the AutoSNPa form (**Figure 21**).

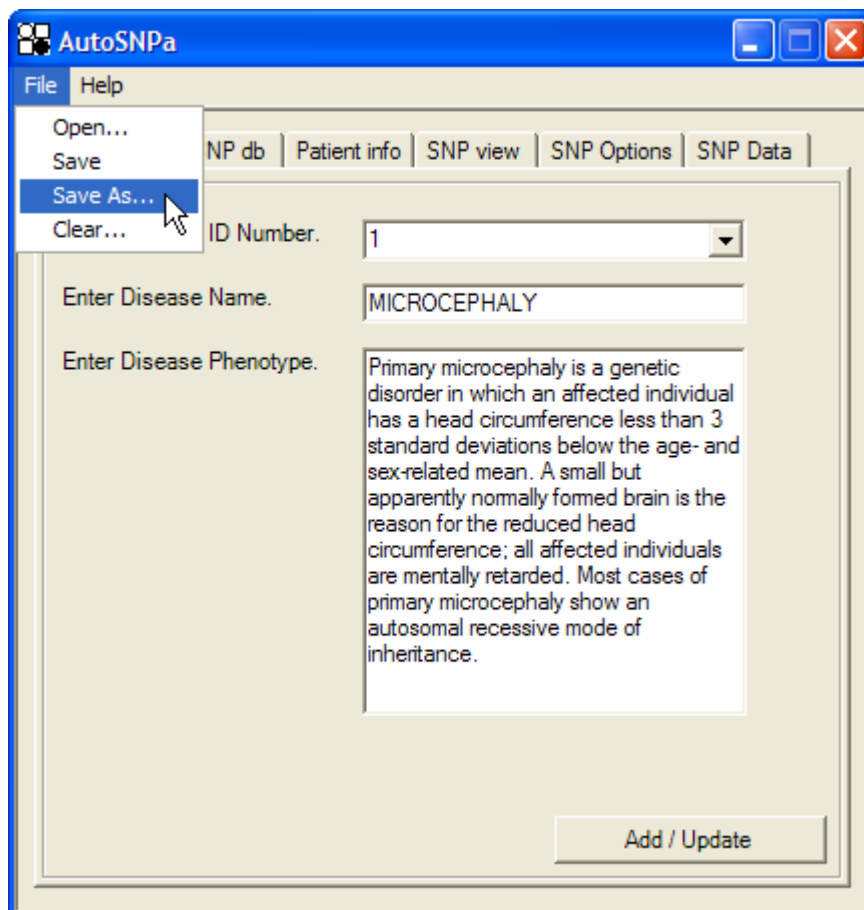


Figure 21: Options for opening and saving the SNP database.

Viewing the SNP data

The genotype data can be viewed by clicking the [View](#) button on the [SNP view](#) tab (**Figure 22**), which opens the [Chromosome View](#) window (**Figure 23**). (The [View large database](#) button also allows you open this window, but without drawing all the

SNPs; this is a useful option when viewing pedigrees containing large amounts of data on a slow computer.)

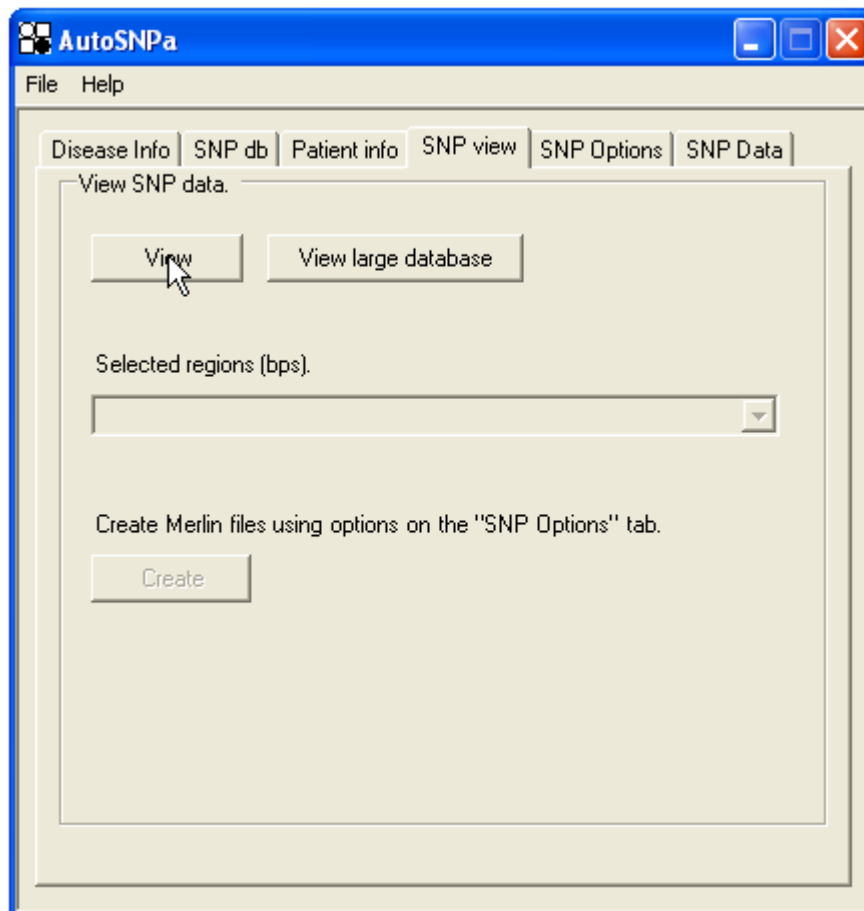


Figure 22: SNP data can be visualised by clicking the [View](#) button.

Each chromosome is displayed individually and accessed by clicking on the tab above the SNP data image (**Figure 23**). As the cursor is moved over the genotype data, the individual's name is displayed alongside the chromosome number (highlighted by the red ellipses in Figure 23). (If the individual's name has not been entered, then the person's unique ID number is shown instead.)

Initially, the genotyping data are shown as an array of colour-coded SNPs, placed along the chromosome ideogram according to physical map position (**Figure 25**). By selecting the [Centimorgan](#) option from the [Distance units](#) submenu under the [Options](#)

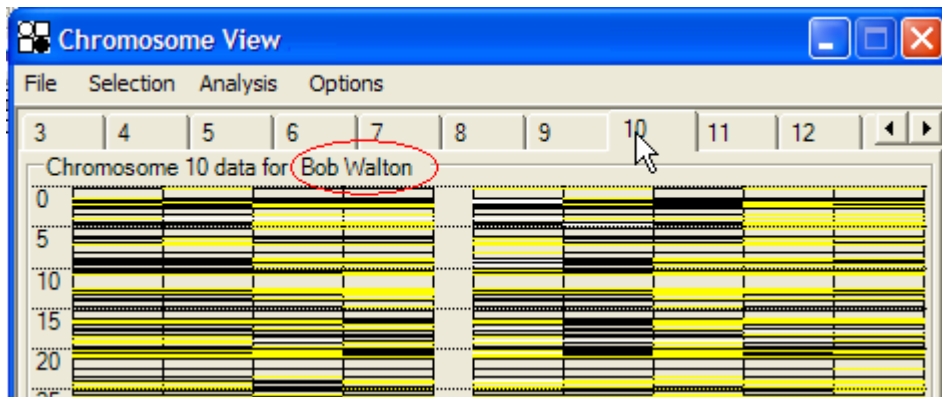


Figure 23: Each chromosome is selected by clicking the appropriate tab (In this figure the cursor has selected chromosome 10.)

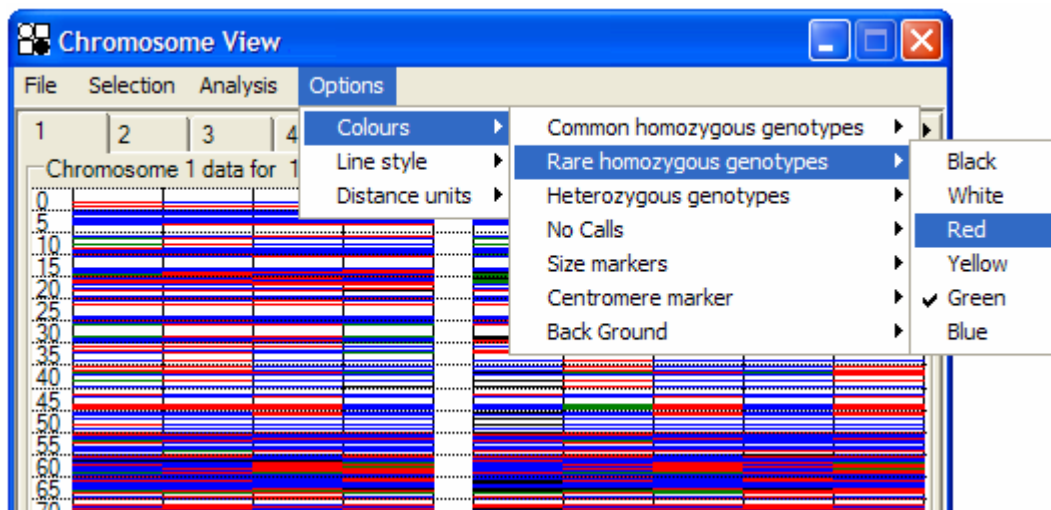


Figure 24: The genotype colour code can be changed via the [Options](#) menu.

menu, the genotyping data are instead displayed relative to genetic position (Marshfield averaged distance) (**Figure 26**). Figures 25 and 26 demonstrate the difference this can make, as the regions around the centromere and rRNA clusters are shortened and the telomeric regions expanded when viewed by genetic distance.

The data are divided into “affected” and “unaffected” groups, with affected individuals displayed to the left of unaffected. Each SNP genotype is colour-coded, depending on its categorization as a “no-call”, “heterozygous”, “common

homozygous” or “rare homozygous”. (If both AA and BB genotypes for a SNP are present in different affected individuals, the variant that is most common is designated

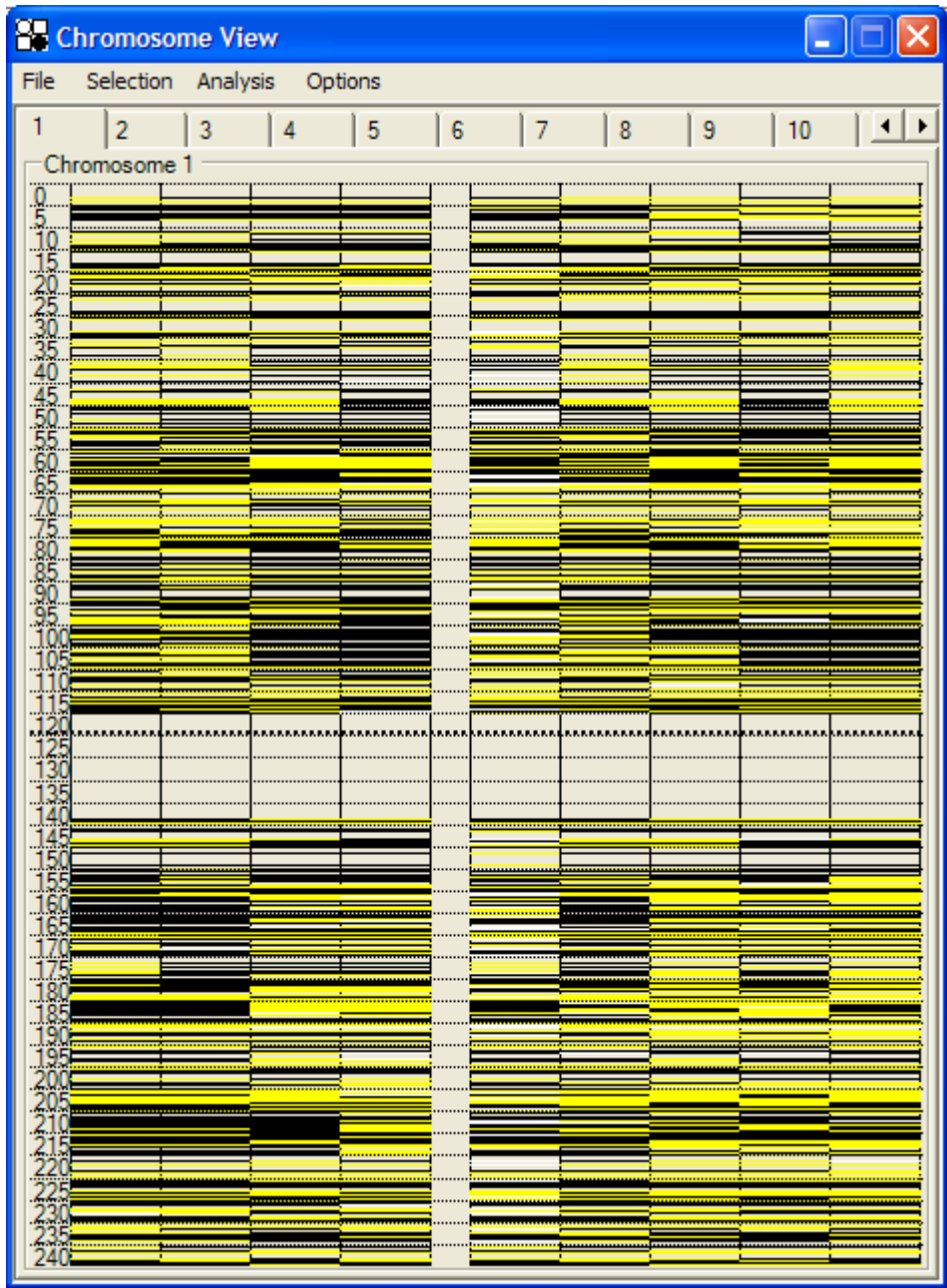


Figure 25: SNP data for chromosome 1, arranged along the chromosome by physical position; the numbers to the left show the distance from the **p** arm telomere in Mb.

the “common homozygous” genotype; genotypes of unaffected individuals are not used in arriving at this decision.) The colour assigned to each category can be changed by the user (**Figure 24**). However, by default, the “common homozygous” category is coloured black and the “rare homozygous” and “heterozygous” genotypes

are both yellow. This allows homozygous regions common to all patients to be identified by the absence of yellow markers.

Smaller regions of this display can be mouse-selected and then expanded for easier viewing (**Figure 27**), while clicking the right mouse button will identify those SNPs lying around the cursor position (**Figure 28**).

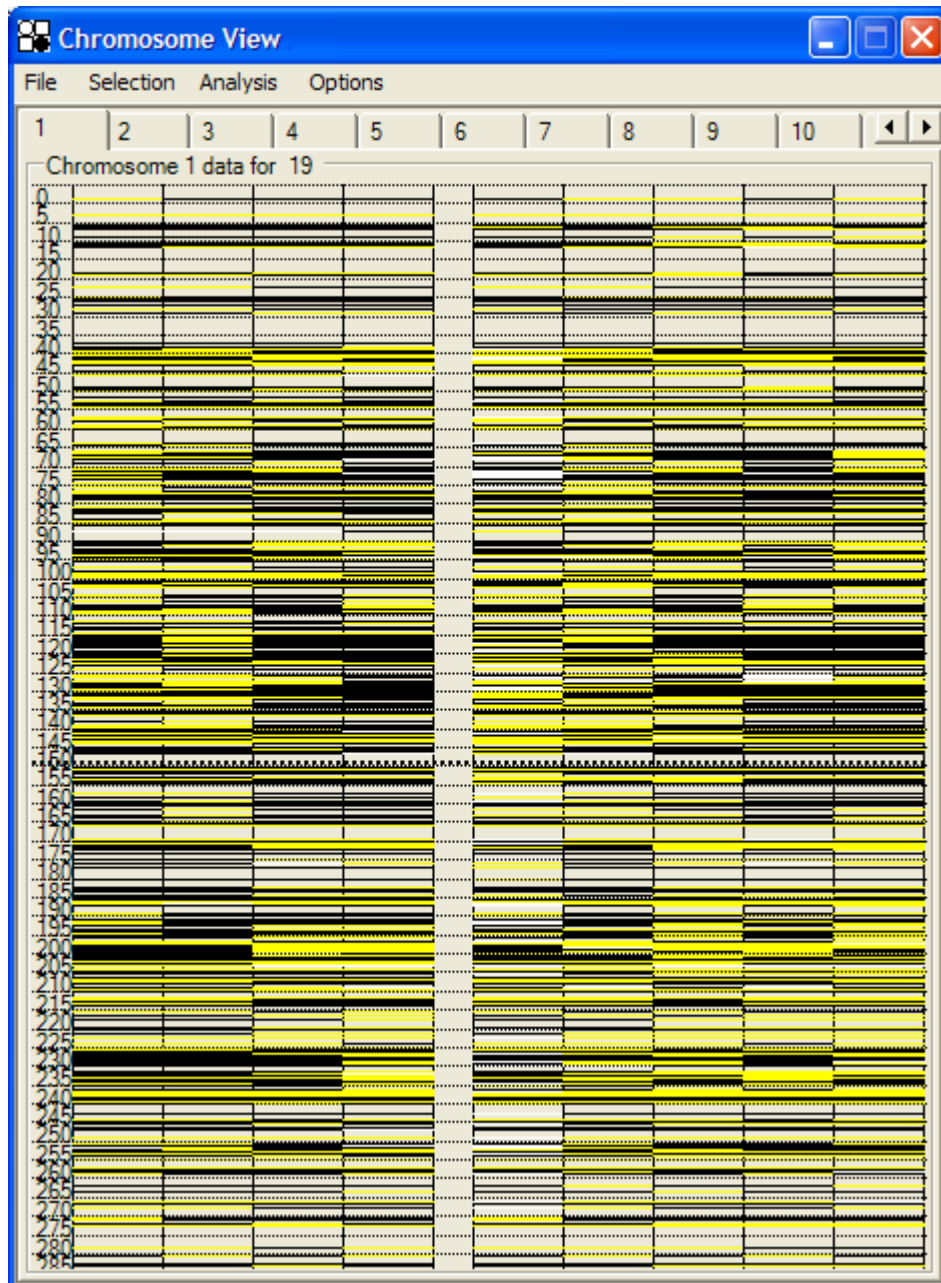


Figure 26: SNP data for chromosome 1 arranged along the chromosome by genetic position; the numbers to the left of the SNP data show the map distance in cM.

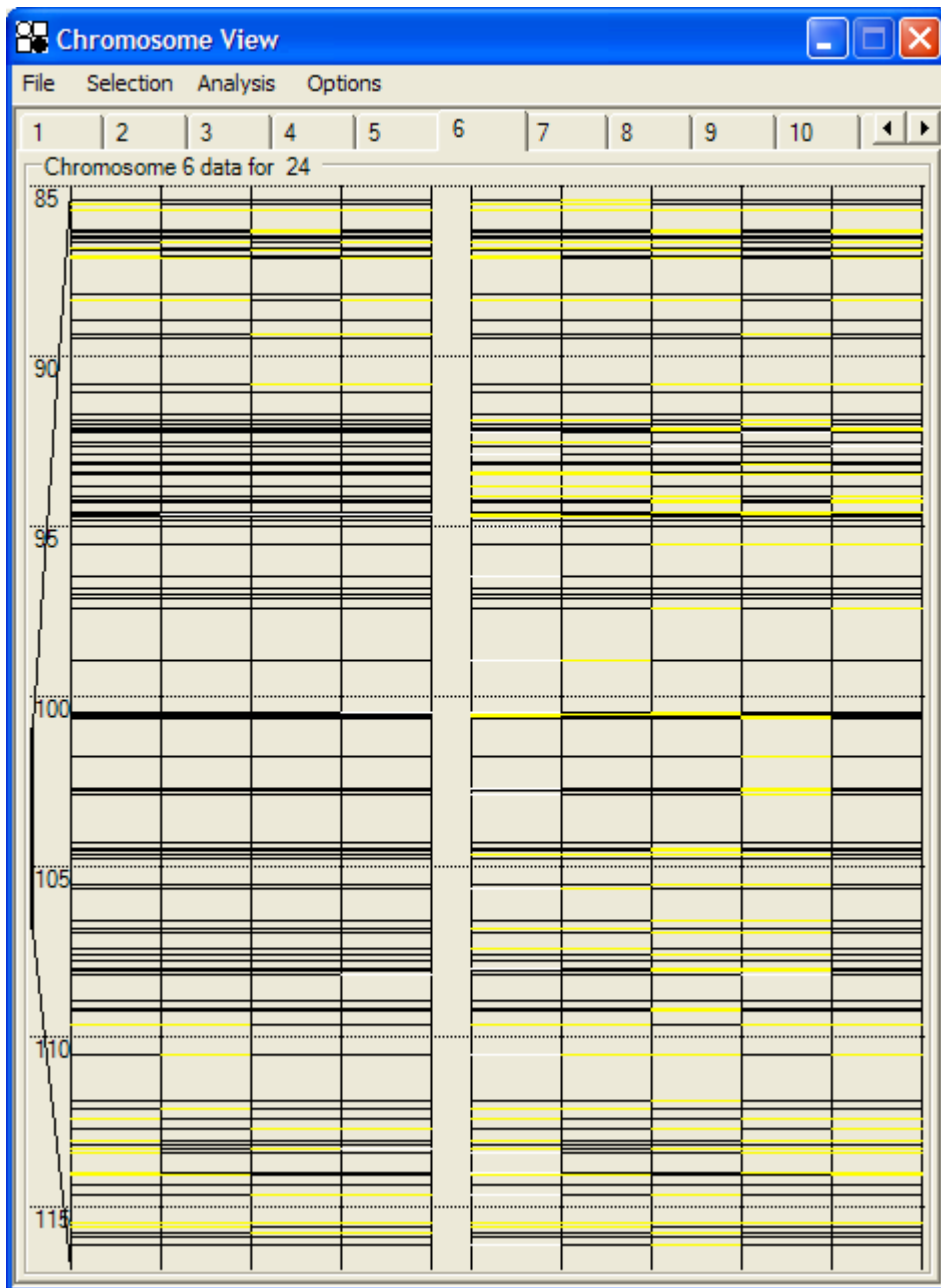


Figure 27: The chromosome view can be expanded by selecting a region

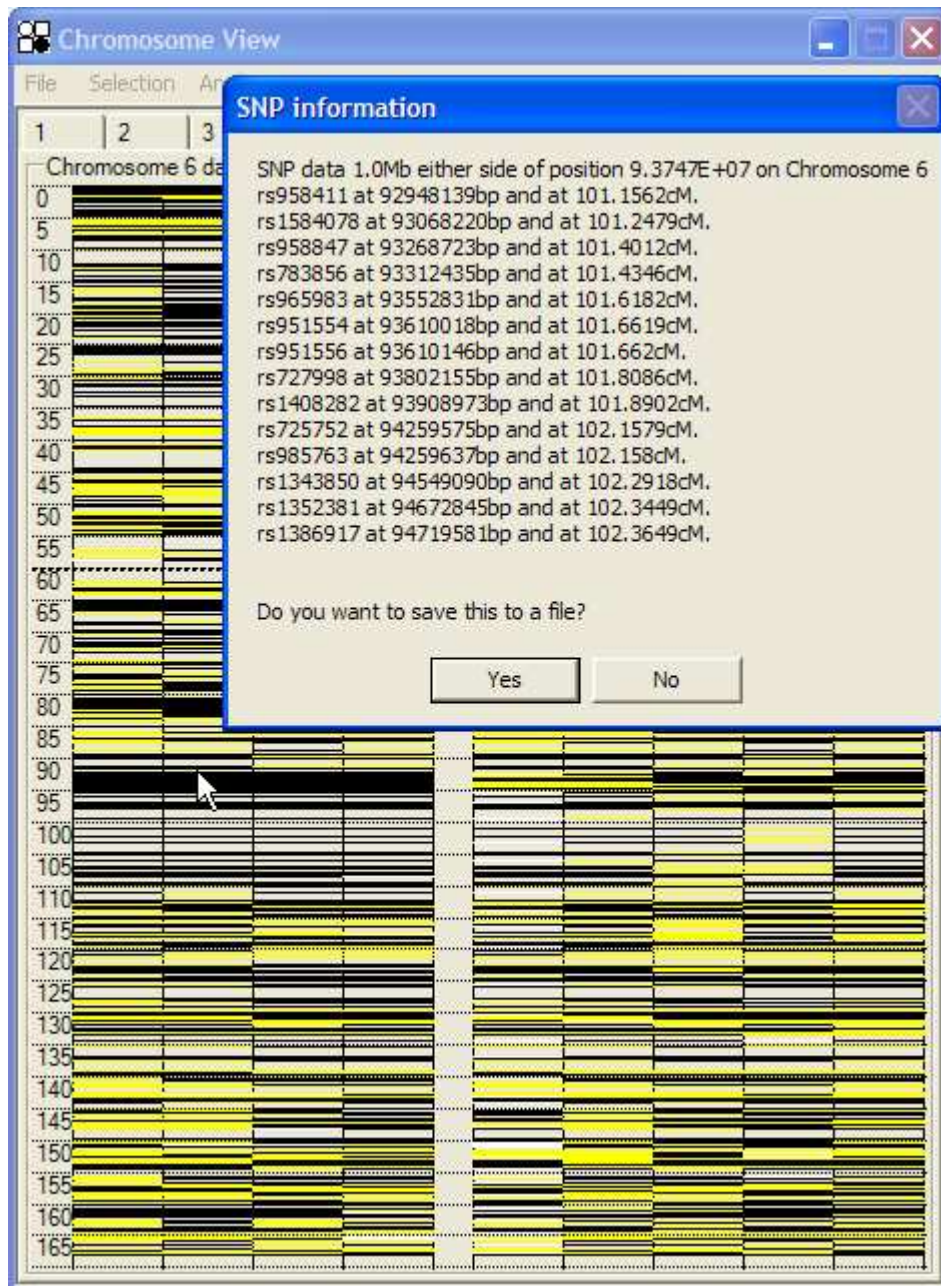


Figure 28: Right-clicking on the form generates a dialogue box showing details of the markers lying around the cursor.

Different methods for viewing SNP data

A consequence of drawing the data to scale with the physical or genetic map is that multiple SNP loci may occupy the same vertical position on the screen, with one result masking the others. To overcome this problem, it is possible to view three different graphical representations of the SNP data, or a text file detailing each region

of interest. Each of the three graphics represents the data in a different way from the initial chromosome display, but uses the same y axis scale and distance units (cM or Mb), so that both views can be directly compared. All of these functions are accessed via the [Analysis](#) menu (**Figure 29**). The [Analysis](#) menu also allows you to select a number of non-visual functions as well.

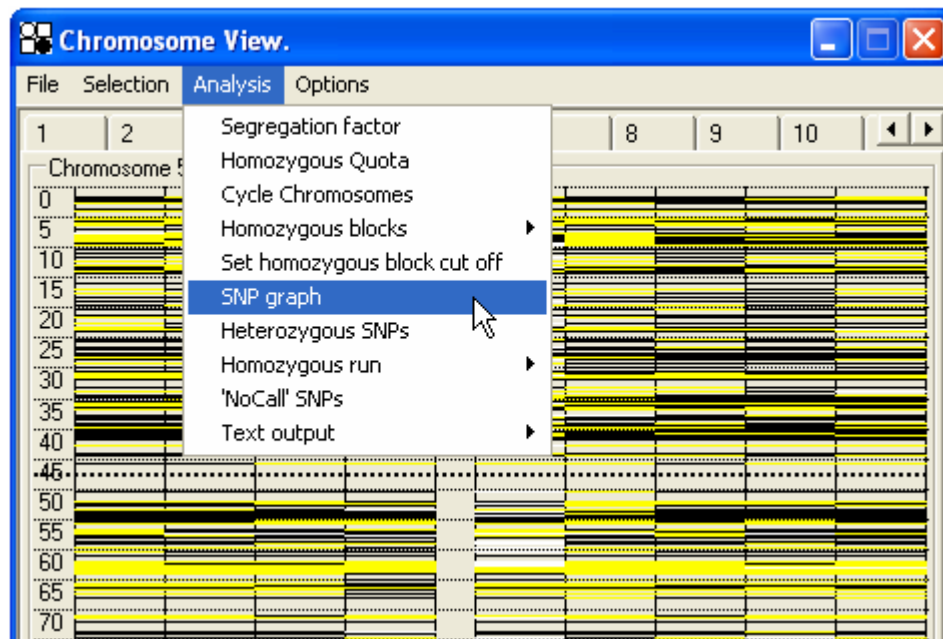


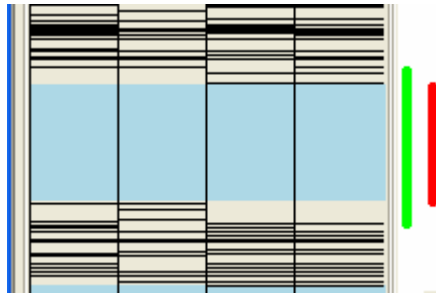
Figure 29: The [Analysis](#) menu.

Visual analysis options for SNP genotype data

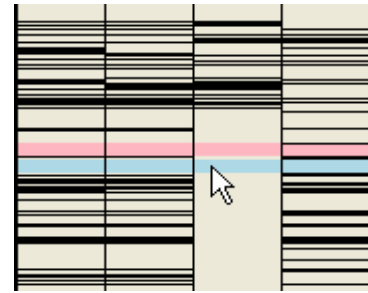
[Heterozygous SNPs:](#)

This view is very similar to the initial chromosome view, but only heterozygous SNPs are shown, enabling rapid identification of homozygous regions (**Figure 30c**). Extended regions that are homozygous for the same SNP genotype in all affected patients are highlighted by a pale blue box, whereas regions homozygous for different genotypes are shown in pink (Figure 30b). These features can be turned on/off via the [Homozygous blocks](#) menu. The preset default is to highlight runs longer than five SNPs, but this cut-off value can be changed to suit each pedigree via the [Set homozygous block cut off](#) option (**Figure 29**).

a.



b.



c.

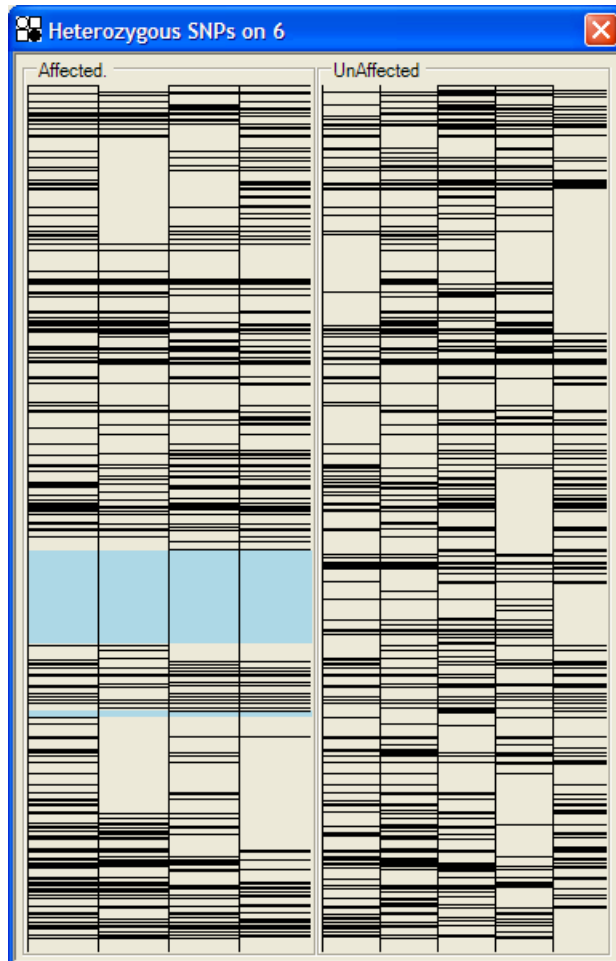


Figure 30: [Heterozygous SNPs](#) view. In Figure 30a, the green line illustrates the extent of the SNP data saved to file if the right mouse button is pressed when the cursor is over the blue box, and the red line shows the extent of the SNP data returned if the left mouse button is activated. Pale blue shading highlights regions where all affected patients share the same genotype; pink regions are where all patients are homozygous, but for any allele.

For a detailed view of the underlying genotype data, clicking the mouse buttons produces either a colour-coded Excel spreadsheet or a tab-delimited text file organised in columns, showing the SNPs' IDs, physical map positions and genotypes, for each individual. This output file also contains the URLs and search strings for the Ensembl and UCSC Golden Path genome browsers, allowing the rapid identification of genes within this chromosomal region. Clicking the left mouse button displays data for the minimum homozygous region (red bar in Figure 30a), whereas the right mouse button retrieves data for a region delimited by the longest runs of homozygous SNPs above and below the selected region (green bar in Figure 30a). This latter feature is designed to overcome the possible artefactual truncation of a homozygous region by a wrongly-called SNP. Examples of the files produced can be seen in the Appendix.

SNP graph:

In this view, all the SNPs that lie superimposed at each y axis pixel are shown as a colour-coded horizontal bar, using the same default colours as described above (**Figure 31**). Affected individuals are again shown in a panel to the left of the unaffected individuals. Genuinely homozygous regions then appear as black bars, whereas regions that harbour “rare homozygous” and heterozygous genotypes contain a black bar tipped by a yellow end. (“No-calls” are shown as white bars). As with the SNP graph, the underlying SNP data can be interrogated via the mouse buttons.

This view is useful for revealing heterozygous SNPs that lie very close to homozygous SNP runs, since some SNPs may be separated by less than 200bp. Because of the greater density of information, it is especially useful when viewing 50k SNP data.

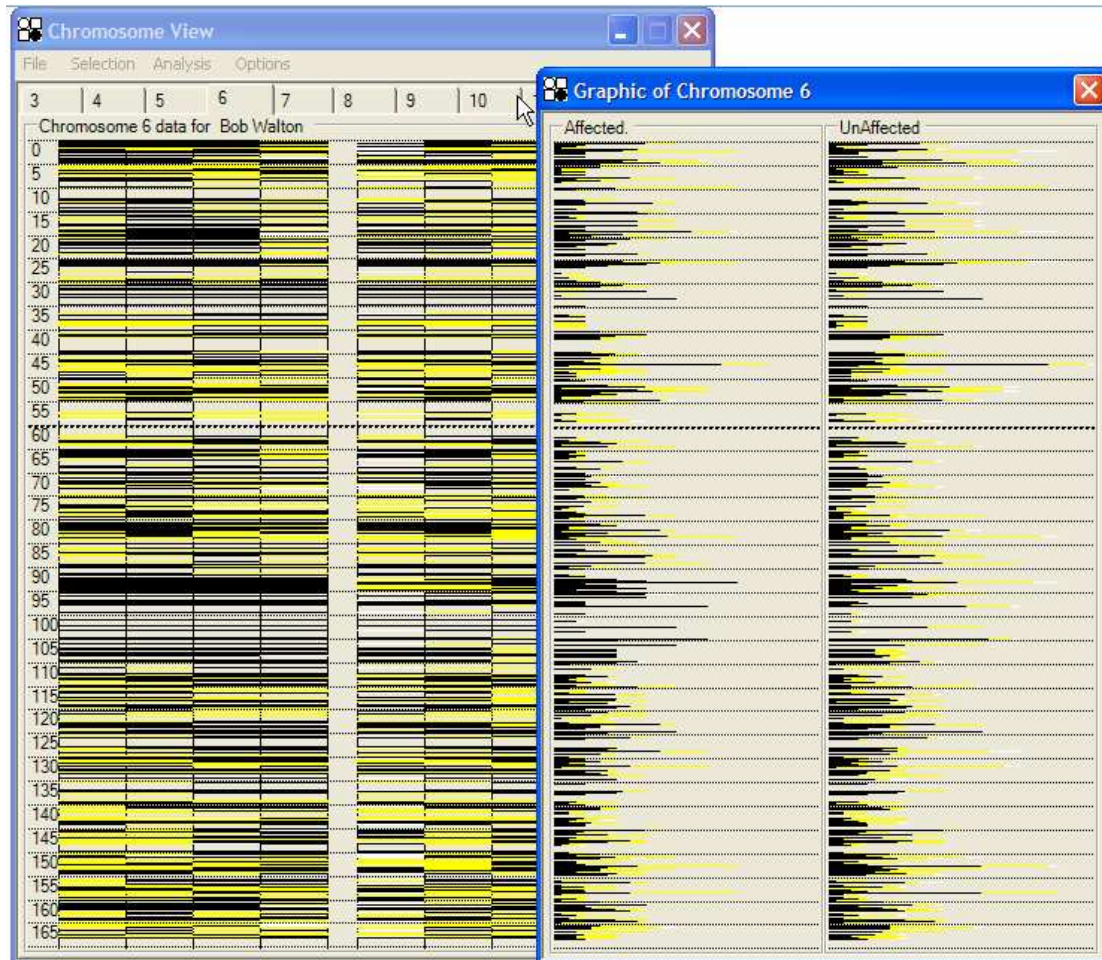


Figure 31: [SNP graph](#) view of the genotype data for chromosome 6.

Homozygous runs:

In this view, runs of homozygous SNPs are displayed as continuous lines (**Figure 32a, 32b**). The displacement of the line along the x axis depends on a simple estimate of the statistical significance of a run of this length. The lower the probability that such a run represents a random event, the farther to the right the line is drawn; a line to the extreme right has a nominal probability of $<10^{-4}$ of being a chance occurrence. This probability can be calculated either from the proportion of homozygous SNPs on that individual's chromosome (**Figure 32a**) or from the allele frequencies supplied by Affymetrix (**Figure 32b**). We emphasize that neither of these estimates is biologically robust, both because of the unknown true allele frequencies and because of unknown degrees of linkage disequilibrium within marker runs. Rather, the intention is to give a

qualitative indication of which homozygous runs may be chance occurrences, and which may be due to inheritance from a common ancestor. As described above (see [Heterozygous SNPs](#)), runs of homozygous SNPs common to all affected individuals are shown by blue or pink rectangles, and the underlying SNP data can be accessed via the mouse buttons. This allows the user to make rapid judgements about the interest of a region, taking into account factors such as the number and density of homozygous SNPs.



a.

b.

Figure 32: [Homozygous runs](#) views, in which each vertical line shows the length of a homozygous region, while its displacement to the right indicates the nominal probability that the region is the result of chance. This nominal probability can be calculated using either the proportion of homozygous SNPs that a patient has on the chromosome under analysis (**a**) or from the predetermined population allele frequencies (**b**).

NoCall SNPs:

This view is very similar to the heterozygous SNPs view, but instead of showing the heterozygous SNPs, only SNPs that produced a 'NoCall' result are shown. This allows the identification of datasets or regions that contain high levels of ambiguous SNP calls. This may be useful, for example, in the case of regions that are involved in duplications and deletions, where ambiguous fluorescence signals on the SNP chip result from aneuploid gene dosage (**Figure 33**).

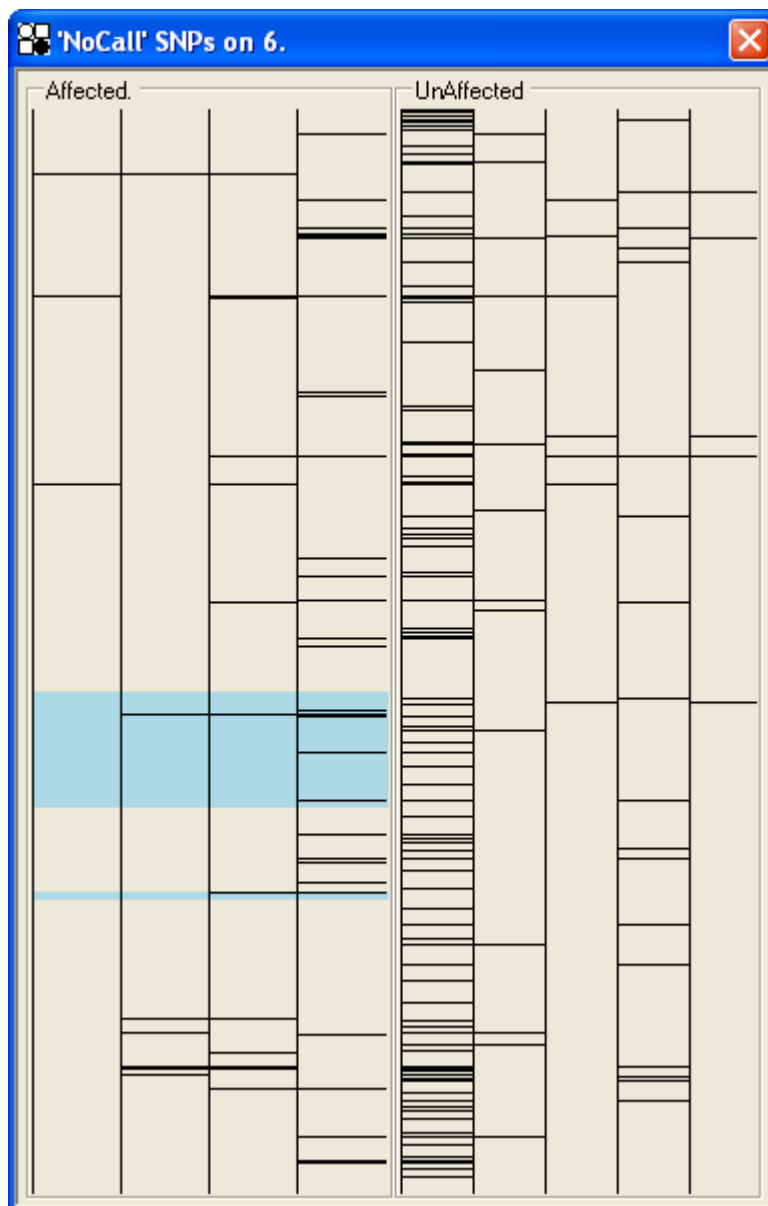


Figure 33: [NoCall SNPs](#) view; each horizontal line identifies the position a NoCall SNP result. The first unaffected individual shows a comparatively high NoCall rate.

Text-based options for analysis of SNP genotype data

[Homozygous blocks:](#)

These functions are accessed via the [Text output](#) menu (**Figure 29**) and generate text files that detail each homozygous region of interest (those shown as blue or pink rectangles in the graphical views described above). These can be either flat text files or colour-coded Excel spreadsheets. The former option makes it easier to include the data in a report, while the latter enables regions of SNP genotypes common to the affected individuals and an unaffected individual to be rapidly identified.

There are three different homozygous block file output options, [Identical genotypes](#), [Non-identical genotypes](#) and [Grouped genotypes](#). The [Identical](#) and [Non-identical](#) genotype options produce text files which contains the data equivalent to the blue and pink rectangles, respectively. The [Grouped genotypes](#) option allows you to place affected individuals into hereditary units which represent different pedigrees. Each unit is then analysed individually for homozygous runs of the same genotype across all individuals within that unit. The results for all of the units are then compared, and regions that have homozygous regions across all units are written to the file.

These groups or units are created using the window shown in Figure 34a. Each individual is selected from the top drop-down list, and then linked to a group, chosen from the lower list, by pressing the [Link](#) button. Patients can then be unlinked, either by placing them in a different group, or by selecting their ID/name and clicking the [Unlink](#) button. The current groupings can be seen via the [View](#) button, which creates a message box listing the patients and their grouping (Figure 34b). The output file is then produced by pressing the [Done](#) button. When using the [Grouped genotypes](#) option, any patients not explicitly linked into a group are disregarded when selecting regions for export. This allows individual patients with suspect SNP data to be arbitrarily excluded from the analysis.

a



b.

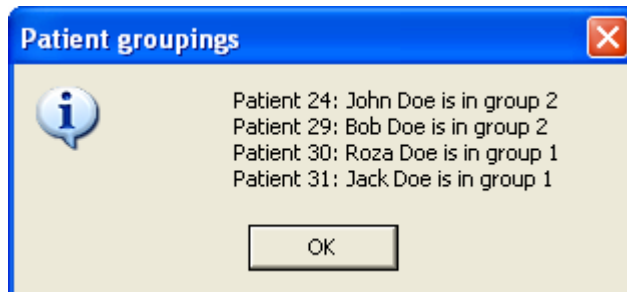


Figure 34: The [Family groups](#) window (a) allows you to place individuals into different hereditary units or groupings, while the [Patient groupings](#) window (b) shows information on the current patient groupings.

If the [Cycle chromosomes](#) menu option is checked, the programme cycles through each chromosome in turn, allowing them to be rapidly analysed. If this menu is checked and the [Homozygous blocks](#) analysis function is chosen, the selected data from all the chromosomes are placed in the resultant text file. The [Homozygous quota](#) analysis option calculates the number of homozygous SNPs on the current chromosome, for each individual. This gives an indication of the degree of inbreeding that each individual has for the current chromosome. This may be used to inform the relative importance assigned to homozygous runs on that chromosome.

The [Segregation factor](#) gives an indication of the likely power of a pedigree, based on simple inheritance patterns, and is calculated without reference to SNP data. The calculated value assumes that the given pedigree describes the total extent of

inbreeding within the family. This assumption is likely to be invalid, as consanguineous families tend to belong to more widely inbreeding populations, and identical-by-descent chromosomal regions may be inherited via individuals not seen in the pedigree (or via wrongly-specified inbreeding loops). Nonetheless, the underlying data that this value is based on can be saved to file and used to guide the selection for genotyping of those individuals that would be the most informative. (For example, two affected siblings are likely to exclude fewer regions than two affected cousins.) To calculate this value, click the [Segregation factor](#) menu (**Figure 29**) and enter the ID(s) of the individual(s) that you believe to be heterozygous for the disease allele. The resulting dialogue box shows the segregation value and offers to save this information to file. (Appendix 1 shows two examples).

Export of Merlin input files

While AutoSNPa is primary aimed at consanguineous pedigrees that are difficult to analyse by conventional statistical methods, it can also produce files for use by Merlin (<http://www.sph.umich.edu/csg/abecasis/Merlin/index.html>). These files can contain data from across the entire genome, whole chromosomes or a chromosomal region. The [SNP Options](#) tab allows you to choose whether to make a frequency file, and if so which allele frequency data are used (**Figure 35**). Merlin frequency files can only be created if the selected allele frequency data are present in the Affymetrix files.

Regions selected in the [Chromosome view](#) window and added through the [Selection](#) menu (**Figure 36a**) are saved to the [Selected regions](#) drop-down list in the [SNP view](#) tab (**Figure 36b**). If a region is then highlighted and the [Create](#) button clicked, Merlin files are created. Each file has the same name, but a different file extension (**Table 2**).

File name	File type	Merlin switches
<filename>.dat	SNP data files.	-d
<filename>.ped	Patient pedigree and genotype files.	-p
<filename>.map	SNP map files.	-m
<filename>.freq	SNP allele frequency files.	-f

Table 2 File extensions created by AutoSNPa for use with Merlin and the corresponding command line switches used by Merlin.

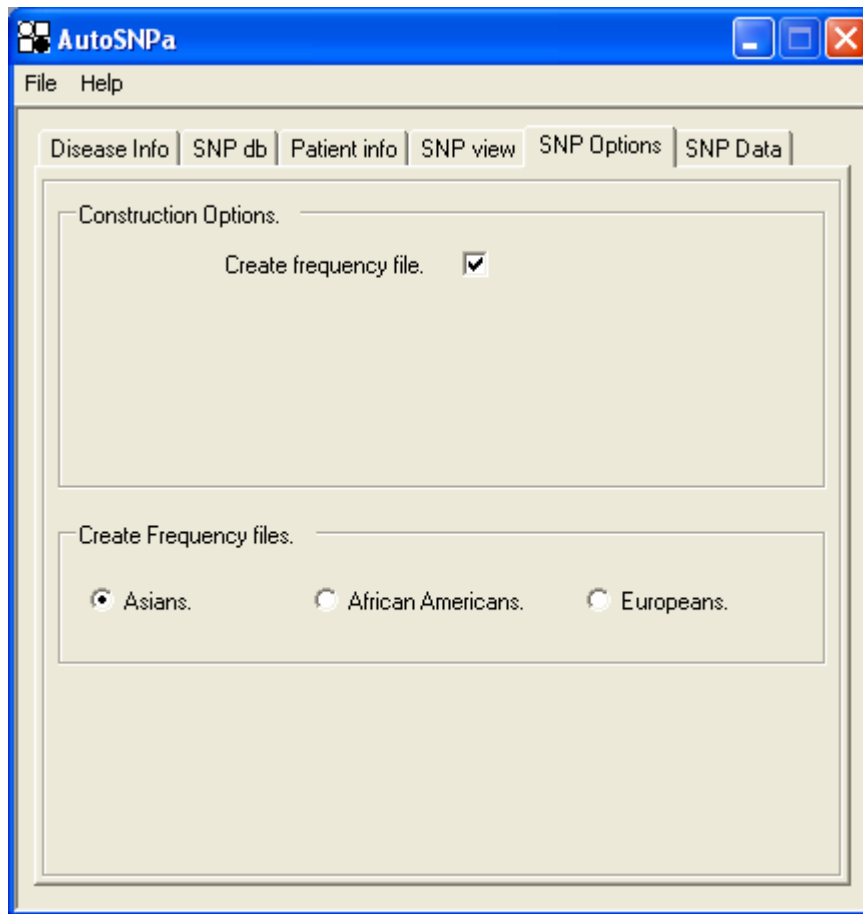
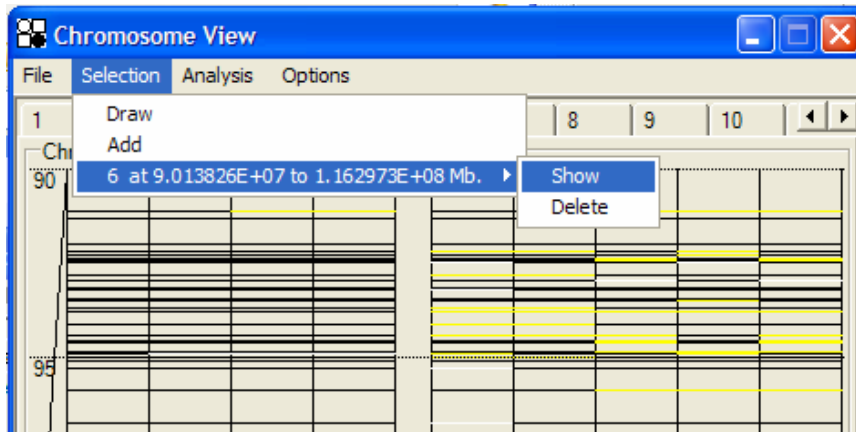
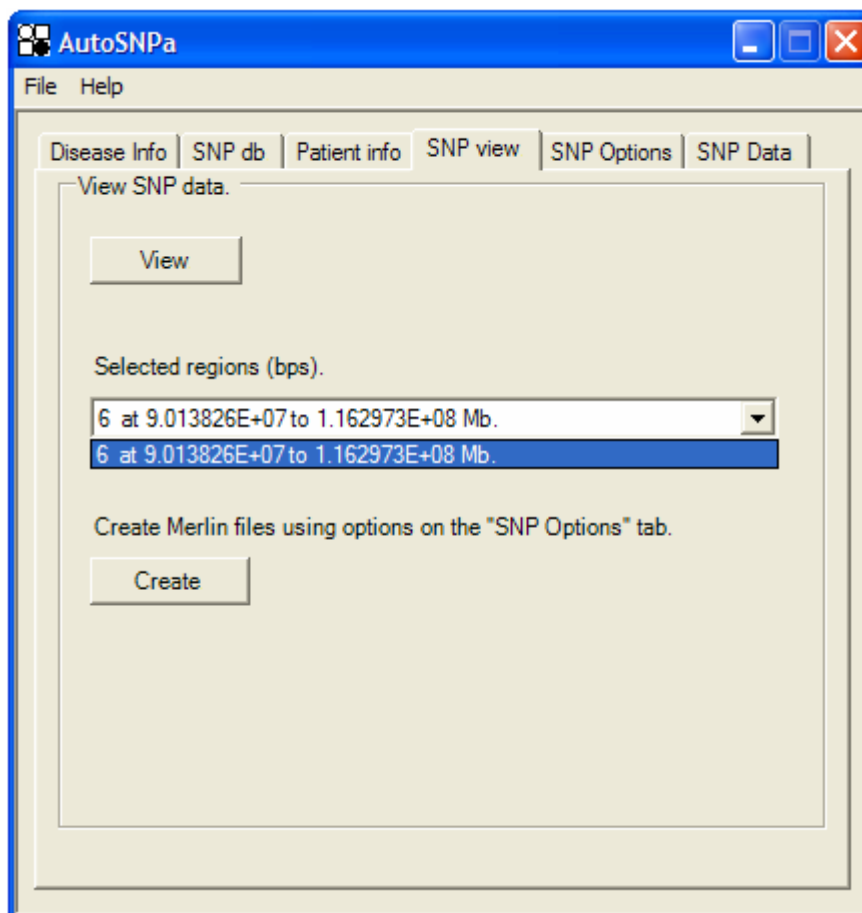


Figure 35: Merlin file SNP and frequency options.



a.



b.

Figure 36: Regions selected in the [Chromosome view](#) window (a) are saved to the [Selected regions](#) list (b) and can then be used to create Merlin files containing SNPs within this region.

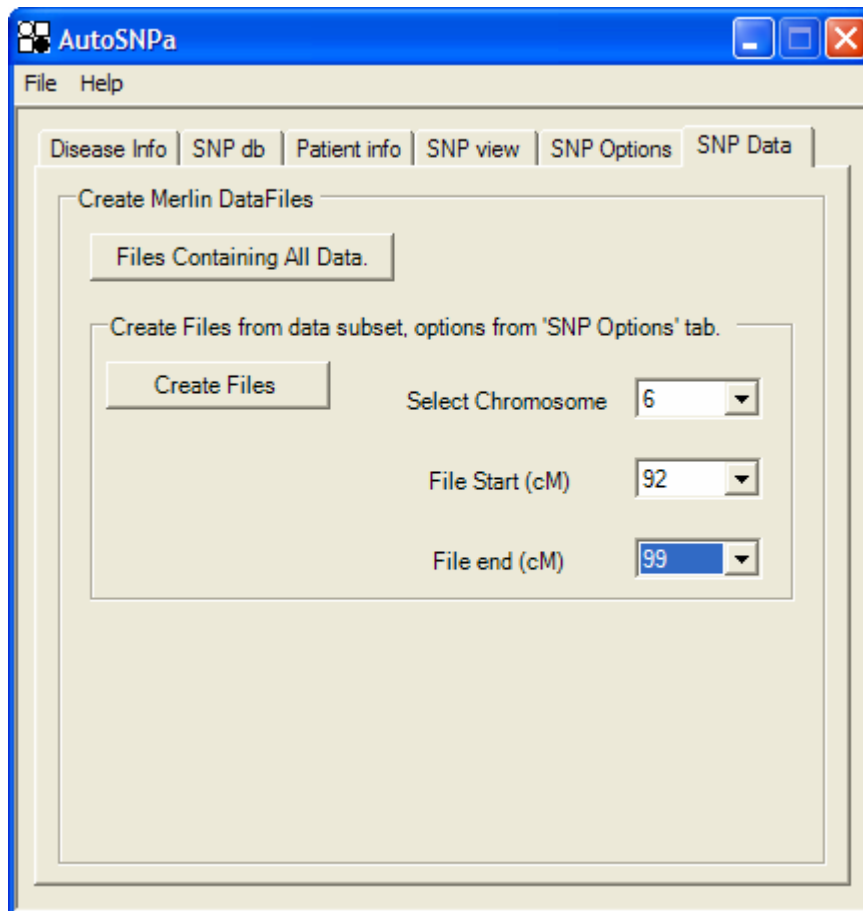


Figure 37: The [SNP Data](#) tab allows you either to create Merlin files containing all the SNPs in the genome or from a sub-region of a chromosome.

The [SNP Data](#) tab (**Figure 37**) allows you to create Merlin files corresponding to the entire genome (click the [Files containing all data](#) button) or a region of interest; the [Create files](#) button will produce files containing SNPs within the region defined by the [File start](#) and [File end](#) on the selected chromosome.

Computer hardware and software requirements

AutoSNPa should run on any computer that has the Microsoft .NET framework version 1.1; we have tested it on Windows 2000 and Windows XP platforms. The .NET framework is available free of charge from Microsoft (search www.microsoft.com for “Microsoft .NET Framework Version 1.1 Redistributable Package” or install via Windows Update). There is no minimum hardware specification for AutoSNPa, other than that needed to run the .NET framework, but obviously, the slower the CPU and smaller the RAM, the slower the programme will run. The forms produced by the programme cannot be resized, and so to view them the screen must have a vertical resolution greater than 600 pixels. In practice, we have found that with 10k SNP data, AutoSNPa will run satisfactorily even on a 266 MHz Pentium II with 128MB memory running Windows 2000.

AutoSNPa does not have intrinsic limitations on the size of data files, and can load and display data from 50k and 250k arrays effectively. However, with large datasets, hardware limitations need to be considered. If the overall size of the SNP database exceeds available RAM, the programme will (like any other) slow down as it swaps information to disk. This is true, irrespective of whether the database is composed of a few 250k files or a larger number of 10k files. If the [Chromosome View](#) window takes an excessive time to display, you should use the [View large database](#) button. This allows you to load this window and access its menu without having to draw the SNP data (which is very processor intensive). If loading the XML datafiles takes more than a few minutes, this may mean that you need more RAM, since the programme holds a large amount of data in memory while running.

To use the functions that create Excel files, Microsoft Excel must be installed on the computer and the file **Interop.Excel.dll** must be in the same folder as the **AutoSNPa.exe** file. Neither AutoSNPa.exe nor Interop.Excel.dll needs to be registered with the operating system, and AutoSNPa can consequently run on any computer with the .NET Framework, even without Administrator rights to install software (e.g. a computer in a departmental cluster).

Appendix

The following data include long lines, and are best viewed by copying and pasting into a simple text editor such as Notepad, with word-wrap disabled.

Segregation factor

Data for the pedigree in Figure 4, with individual 1 set as a carrier

There is a 1 in 8192 chance of an unrelated DNA region co-segregating with the disease locus, if 1 is the founder.

1 in 8192 is $\text{Log}_{10}(-)3.91338994363176$

Unaffected individual 1 has a chance of 0 of being heterozygous for the DNA region. Parental IDs are: mother ID: 20000 and father ID; 10000

Unaffected individual 2 has a chance of 0 of being heterozygous for the DNA region. Parental IDs are: mother ID: 0 and father ID; 0

Unaffected individual 3 has a chance of 0.5 of being either heterozygous or homozygous for the DNA region. Parental IDs are: mother ID: 2 and father ID; 1

Unaffected individual 4 has a chance of 0.5 of being either heterozygous or homozygous for the DNA region. Parental IDs are: mother ID: 2 and father ID; 1

Unaffected individual 5 has a chance of 0 of being heterozygous for the DNA region. Parental IDs are: mother ID: 0 and father ID; 0

Unaffected individual 6 has a chance of 0 of being heterozygous for the DNA region. Parental IDs are: mother ID: 0 and father ID; 0

Unaffected individual 7 has a chance of 0.25 of being either heterozygous or homozygous for the DNA region. Parental IDs are: mother ID: 5 and father ID; 3

Unaffected individual 8 has a chance of 0.25 of being either heterozygous or homozygous for the DNA region. Parental IDs are: mother ID: 5 and father ID; 3

Unaffected individual 9 has a chance of 0.25 of being heterozygous for the DNA region. Parental IDs are: mother ID: 4 and father ID; 6

Unaffected individual 10 has a chance of 0.25 of being heterozygous for the DNA region. Parental IDs are: mother ID: 4 and father ID; 6

Unaffected individual 11 has a chance of 0.25 of being heterozygous for the DNA region. Parental IDs are: mother ID: 4 and father ID; 6

Unaffected individual 12 has a chance of 0 of being heterozygous for the DNA region. Parental IDs are: mother ID: 0 and father ID; 0

Unaffected individual 13 has a chance of 0 of being heterozygous for the DNA region. Parental IDs are: mother ID: 0 and father ID; 0

Unaffected individual 14 has a chance of 0 of being heterozygous for the DNA region. Parental IDs are: mother ID: 0 and father ID; 0

Unaffected individual 15 has a chance of 0.125 of being either heterozygous or homozygous for the DNA region. Parental IDs are: mother ID: 12 and father ID; 7

Unaffected individual 16 has a chance of 0.234375 of being either heterozygous or homozygous for the DNA region. Parental IDs are: mother ID: 8 and father ID; 9

Unaffected individual 17 has a chance of 0.125 of being either heterozygous or homozygous for the DNA region. Parental IDs are: mother ID: 13 and father ID; 10

Unaffected individual 18 has a chance of 0.125 of being either heterozygous or homozygous for the DNA region. Parental IDs are: mother ID: 13 and father ID; 10

Unaffected individual 19 has a chance of 0.125 of being either heterozygous or homozygous for the DNA region. Parental IDs are: mother ID: 14 and father ID; 11

Unaffected individual 20 has a chance of 0.125 of being either heterozygous or homozygous for the DNA region. Parental IDs are: mother ID: 14 and father ID; 11

Unaffected individual 21 has a chance of 0.125 of being either heterozygous or homozygous for the DNA region. Parental IDs are: mother ID: 14 and father ID; 11

Unaffected individual 22 has a chance of 0 of being heterozygous for the DNA region. Parental IDs are: mother ID: 0 and father ID; 0

Unaffected individual 23 has a chance of 0 of being heterozygous for the DNA region. Parental IDs are: mother ID: 0 and father ID; 0

Affected patient 24 has a chance of 0.0078125 of being homozygous for the DNA region. Parental IDs are: mother ID: 16 and father ID; 15

Unaffected individual 25 has a chance of 0.0625 of being heterozygous for the DNA region. Parental IDs are: mother ID: 17 and father ID; 22

Unaffected individual 26 has a chance of 0.1210938 of being either heterozygous or homozygous for the DNA region. Parental IDs are: mother ID: 20 and father ID; 18

Unaffected individual 27 has a chance of 0.0625 of being heterozygous for the DNA region. Parental IDs are: mother ID: 21 and father ID; 23

Unaffected individual 28 has a chance of 0.09179688 of being either heterozygous or homozygous for the DNA region. Parental IDs are: mother ID: 27 and father ID; 26

Affected patient 29 has a chance of 0.0078125 of being homozygous for the DNA region. Parental IDs are: mother ID: 16 and father ID; 15

Affected patient 30 has a chance of 0.001953125 of being homozygous for the DNA region. Parental IDs are: mother ID: 19 and father ID; 25

Affected patient 31 has a chance of 0.001953125 of being homozygous for the DNA region. Parental IDs are: mother ID: 19 and father ID; 25

Since patient 29 has an affected sib their chance is not 0.125, but 1/4.

Since patient 31 has an affected sib their chance is not 0.015625, but 1/4.

Data for the pedigree in Figure 4, with individuals 1 and 6 set as carriers

There is a 1 in 260 chance of an unrelated DNA region co-segregating with the disease locus, if 1, 6 is the founder.

1 in 260 is $\text{Log}_{10}(-)2.41507938984216$

Unaffected individual 1 has a chance of 0 of being heterozygous for the DNA region. Parental IDs are: mother ID: 20000 and father ID; 10000

Unaffected individual 2 has a chance of 0 of being heterozygous for the DNA region. Parental IDs are: mother ID: 0 and father ID; 0

Unaffected individual 3 has a chance of 0.5 of being either heterozygous or homozygous for the DNA region. Parental IDs are: mother ID: 2 and father ID; 1

Unaffected individual 4 has a chance of 0.5 of being either heterozygous or homozygous for the DNA region. Parental IDs are: mother ID: 2 and father ID; 1

Unaffected individual 5 has a chance of 0 of being heterozygous for the DNA region. Parental IDs are: mother ID: 0 and father ID; 0

Unaffected individual 6 has a chance of 0 of being heterozygous for the DNA region. Parental IDs are: mother ID: 0 and father ID; 0

Unaffected individual 7 has a chance of 0.25 of being either heterozygous or homozygous for the DNA region. Parental IDs are: mother ID: 5 and father ID; 3

Unaffected individual 8 has a chance of 0.25 of being either heterozygous or homozygous for the DNA region. Parental IDs are: mother ID: 5 and father ID; 3

Unaffected individual 9 has a chance of 0.625 of being either heterozygous or homozygous for the DNA region. Parental IDs are: mother ID: 4 and father ID; 6

Unaffected individual 10 has a chance of 0.625 of being either heterozygous or homozygous for the DNA region. Parental IDs are: mother ID: 4 and father ID; 6

Unaffected individual 11 has a chance of 0.625 of being either heterozygous or homozygous for the DNA region. Parental IDs are: mother ID: 4 and father ID; 6

Unaffected individual 12 has a chance of 0 of being heterozygous for the DNA region. Parental IDs are: mother ID: 0 and father ID; 0

Unaffected individual 13 has a chance of 0 of being heterozygous for the DNA region. Parental IDs are: mother ID: 0 and father ID; 0

Unaffected individual 14 has a chance of 0 of being heterozygous for the DNA region. Parental IDs are: mother ID: 0 and father ID; 0

Unaffected individual 15 has a chance of 0.125 of being either heterozygous or homozygous for the DNA region. Parental IDs are: mother ID: 12 and father ID; 7

Unaffected individual 16 has a chance of 0.453125 of being either heterozygous or homozygous for the DNA region. Parental IDs are: mother ID: 8 and father ID; 9

Unaffected individual 17 has a chance of 0.375 of being either heterozygous or homozygous for the DNA region. Parental IDs are: mother ID: 13 and father ID; 10

Unaffected individual 18 has a chance of 0.375 of being either heterozygous or homozygous for the DNA region. Parental IDs are: mother ID: 13 and father ID; 10

Unaffected individual 19 has a chance of 0.375 of being either heterozygous or homozygous for the DNA region. Parental IDs are: mother ID: 14 and father ID; 11

Unaffected individual 20 has a chance of 0.375 of being either heterozygous or homozygous for the DNA region. Parental IDs are: mother ID: 14 and father ID; 11

Unaffected individual 21 has a chance of 0.375 of being either heterozygous or homozygous for the DNA region. Parental IDs are: mother ID: 14 and father ID; 11

Unaffected individual 22 has a chance of 0 of being heterozygous for the DNA region. Parental IDs are: mother ID: 0 and father ID; 0

Unaffected individual 23 has a chance of 0 of being heterozygous for the DNA region. Parental IDs are: mother ID: 0 and father ID; 0

Affected patient 24 has a chance of 0.015625 of being homozygous for the DNA region. Parental IDs are: mother ID: 16 and father ID; 15

Unaffected individual 25 has a chance of 0.1875 of being heterozygous for the DNA region. Parental IDs are: mother ID: 17 and father ID; 22

Unaffected individual 26 has a chance of 0.3398438 of being either heterozygous or homozygous for the DNA region. Parental IDs are: mother ID: 20 and father ID; 18

Unaffected individual 27 has a chance of 0.1875 of being heterozygous for the DNA region. Parental IDs are: mother ID: 21 and father ID; 23

Unaffected individual 28 has a chance of 0.2636719 of being either heterozygous or homozygous for the DNA region. Parental IDs are: mother ID: 27 and father ID; 26

Affected patient 29 has a chance of 0.015625 of being homozygous for the DNA region. Parental IDs are: mother ID: 16 and father ID; 15

Affected patient 30 has a chance of 0.01757813 of being homozygous for the DNA region. Parental IDs are: mother ID: 19 and father ID; 25

Affected patient 31 has a chance of 0.01757813 of being homozygous for the DNA region. Parental IDs are: mother ID: 19 and father ID; 25

Since patient 29 has an affected sib their chance is not 0.4375, but 1/4.

Since patient 31 has an affected sib their chance is not 0.140625, but 1/4.

SNP genotype data accessed via the mouse on forms [SNP Graph](#), [Homozygous runs](#) and [Heterozygous SNPs](#)

Left mouse button

Ensemble string: http://www.ensembl.org/Homo_sapiens/contigview?chr=6®ion=&start=91450040&end=110088856

GoldenPath string: chr6:91450040-110088856 at <http://genome.ucsc.edu/cgi-bin/hgTracks>

SNP	Mb distance	cM distance	24	29	30	31	15	16	19	25	28	
rs1923064	91450041	98.15902	AA	AA	AA	AA		AA	AA	ab	AA	ab
rs2093109	91450238	98.15928	AA	AA	AA	AA		AA	AA	AA	AA	AA
rs2144363	92098625	99.01501	BB	BB	BB	BB		BB	BB	BB	BB	BB
rs2180125	92098763	99.0155	BB	BB	BB	BB		BB	BB	BB	BB	BB
rs4129544	92305121	99.75694	AA	AA	AA	AA		ab	ab	AA	ab	AA
rs1201867	92397523	100.0889	BB	BB	BB	BB		BB	BB	BB	ab	BB
rs976763	92523890	100.543	AA	AA	AA	AA		NoCall	AA	ab	AA	ab
rs976762	92523979	100.5433	AA	AA	AA	AA		AA	AA	ab	AA	ab
rs2325395	92524832	100.5464	BB	BB	BB	BB		BB	BB	ab	BB	ab
rs1603727	92566077	100.6946	AA	AA	AA	AA		AA	AA	ab	AA	ab
rs1391496	92624521	100.9045	AA	AA	AA	AA		NoCall	AA	AA	NoCall	AA
rs958411	92948139	101.1562	AA	AA	AA	AA		ab	ab	AA	AA	AA
rs1584078	93068220	101.2479	AA	AA	AA	AA		AA	AA	NoCall	AA	NoCall
rs958847	93268723	101.4012	AA	AA	AA	AA		AA	AA	AA	AA	AA
rs783856	93312435	101.4346	BB	BB	BB	BB		NoCall	BB	BB	BB	BB
rs965983	93552831	101.6182	BB	BB	BB	BB		BB	BB	BB	BB	BB
rs951554	93610018	101.6619	AA	AA	AA	AA		ab	ab	ab	AA	ab
rs951556	93610146	101.662	BB	BB	BB	BB		BB	BB	BB	ab	BB
rs727998	93802155	101.8086	AA	AA	AA	AA		ab	ab	AA	AA	AA
rs1408282	93908973	101.8902	AA	AA	AA	AA		ab	ab	ab	ab	ab
rs725752	94259575	102.1579	AA	AA	AA	NoCall		ab	ab	AA	AA	AA
rs985763	94259637	102.158	BB	BB	BB	BB		ab	ab	BB	BB	BB
rs1343850	94549090	102.2918	AA	AA	AA	AA		ab	ab	ab	AA	ab
rs1352381	94672845	102.3449	AA	AA	AA	AA		AA	AA	ab	AA	ab
rs1386917	94719581	102.3649	BB	BB	BB	BB		BB	BB	ab	BB	ab
rs423246	94982708	102.4778	BB	BB	BB	BB		BB	BB	BB	BB	BB
rs2021314	94994454	102.4828	AA	AA	AA	AA		AA	AA	ab	ab	ab
rs952577	95044774	102.5044	AA	NoCall	NoCall	NoCall		ab	AA	AA	BB	AA

rs62827	95125887	102.5392	BB	BB	BB	BB	ab	ab	BB	BB	BB
rs2207430	95212976	102.5765	AA	AA	AA	NoCall	AA	AA	AA	AA	AA
rs2224003	95213139	102.5766	BB	BB	BB	BB	BB	BB	BB	BB	BB
rs2502004	95413900	102.6627	AA	AA	AA	AA	NoCall	AA	AA	AA	AA
rs2246786	95965958	102.8995	BB	BB	BB	BB	BB	BB	ab	ab	ab
rs2387156	96879221	103.2912	AA	AA	AA	AA	NoCall	AA	AA	AA	AA
rs1113233	97244293	103.4478	AA	AA	AA	AA	AA	AA	AA	AA	AA
rs953878	97434550	103.5294	BB	BB	BB	BB	ab	NoCall	BB	BB	BB
rs1359932	97436898	103.5304	AA	AA	AA	AA	NoCall	AA	AA	ab	AA
rs1928968	97437070	103.5304	AA	AA	AA	AA	AA	AA	AA	AA	AA
rs2143389	97529402	103.57	BB	BB	BB	BB	BB	BB	ab	ab	ab
rs2179537	97529651	103.5701	AA	AA	AA	AA	AA	AA	AA	AA	AA
rs2388039	97853965	103.7092	AA	AA	AA	AA	AA	AA	ab	AA	ab
rs4131464	99364448	105.1737	AA	AA	AA	AA	NoCall	ab	AA	AA	AA
rs1354685	100861479	106.5254	AA	AA	AA	NoCall	NoCall	AA	ab	AA	AA
rs718268	100965880	106.6197	BB	BB	BB	BB	ab	ab	ab	BB	BB
rs1857859	101001308	106.6517	BB	BB	BB	BB	ab	BB	BB	ab	BB
rs2001102	101085072	106.7273	BB	BB	BB	BB	BB	BB	BB	ab	BB
rs1856133	102207593	108.2534	AA	AA	AA	AA	AA	AA	AA	ab	AA
rs1934009	103122488	110.0321	AA	AA	AA	AA	NoCall	AA	AA	ab	AA
rs1933989	103171750	110.1037	AA	AA	AA	AA	AA	AA	AA	AA	AA
rs1934008	103171853	110.1038	BB	BB	BB	BB	BB	BB	BB	ab	BB
rs967405	103289331	110.2745	AA	AA	AA	AA	AA	AA	AA	ab	AA
rs1431213	103305549	110.2981	BB	BB	BB	BB	NoCall	BB	BB	ab	BB
rs1073233	104687791	112.009	AA	AA	AA	AA	AA	AA	AA	AA	AA
rs559295	104858181	112.1917	AA	AA	AA	AA	AA	AA	ab	AA	AA
rs1341112	104919391	112.2574	AA	AA	AA	AA	AA	AA	ab	AA	AA
rs271857	105058386	112.4065	AA	AA	AA	AA	ab	ab	ab	ab	AA
rs271858	105058568	112.4067	AA	AA	AA	AA	ab	ab	ab	ab	AA
rs1416042	105174572	112.5312	BB	BB	BB	BB	BB	BB	BB	BB	BB
rs720225	105949023	113.2462	BB	BB	BB	BB	BB	BB	ab	ab	BB
rs722366	106080157	113.3379	AA	AA	AA	AA	NoCall	ab	AA	AA	AA
rs1417352	107005919	114.3331	AA	AA	AA	AA	ab	ab	AA	AA	AA
rs1417353	107006104	114.3334	AA	AA	AA	AA	AA	AA	ab	ab	AA
rs2077953	107264726	114.6821	BB	BB	BB	BB	ab	ab	BB	BB	BB

rs723318	107387007	114.847	AA	AA	AA	AA	AA	AA	AA	ab	ab	AA
rs319088	107846733	115.4669	BB	BB	BB	BB	BB	ab	ab	BB	BB	BB
rs2243869	108027221	115.7102	BB	BB	BB	BB	BB	BB	BB	ab	ab	BB
rs2354095	108196930	115.9391	BB	BB	BB	BB	BB	BB	BB	BB	BB	BB
rs2354094	108197032	115.9392	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA
rs2355205	108446020	116.323	BB	BB	BB	BB	BB	NoCall	BB	ab	ab	BB
rs718174	108479948	116.5157	AA	AA	AA	AA	AA	AA	AA	ab	ab	AA
rs217538	108590163	117.1417	AA	AA	AA	NoCall	AA	AA	AA	AA	NoCall	AA
rs1040849	109349779	117.4965	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA
rs566164	109613154	117.5706	AA	AA	AA	AA	AA	AA	AA	ab	AA	AA
rs59673	109655406	117.5825	BB	BB	BB	BB	BB	BB	BB	ab	BB	BB
rs2016207	110088853	117.7045	ab	ab	BB	BB	AA	ab	BB	BB	ab	ab

Right mouse button

Ensemble string: http://www.ensembl.org/Homo_sapiens/contigview?chr=6®ion=&start=89751827&end=113107351

GoldenPath string: chr6: 89751827-113107351 at <http://genome.ucsc.edu/cgi-bin/hgTracks>

SNP	Mb distance	cM distance	24	29	30	31	15	16	19	25	28	
rs2207553	89751827	95.96333	BB	BB	ab	ab	BB	BB	BB	BB	AA	BB
rs1590957	89867130	96.15065	BB	BB	BB	BB	BB	BB	BB	BB	BB	BB
rs1353227	91233622	97.87446	BB	BB	ab	ab	BB	BB	ab	AA	ab	
rs1923064	91450041	98.15902	AA	AA	AA	AA	AA	AA	ab	AA	ab	
rs2093109	91450238	98.15928	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA
rs2144363	92098625	99.01501	BB	BB	BB	BB	BB	BB	BB	BB	BB	BB
rs2180125	92098763	99.0155	BB	BB	BB	BB	BB	BB	BB	BB	BB	BB
rs4129544	92305121	99.75694	AA	AA	AA	AA	ab	ab	AA	ab	AA	
rs1201867	92397523	100.0889	BB	BB	BB	BB	BB	BB	BB	BB	ab	BB
rs976763	92523890	100.543	AA	AA	AA	AA	NoCall	AA	ab	AA	ab	
rs976762	92523979	100.5433	AA	AA	AA	AA	AA	AA	ab	AA	ab	
rs2325395	92524832	100.5464	BB	BB	BB	BB	BB	BB	ab	BB	ab	
rs1603727	92566077	100.6946	AA	AA	AA	AA	AA	AA	ab	AA	ab	
rs1391496	92624521	100.9045	AA	AA	AA	AA	NoCall	AA	AA	NoCall	AA	
rs958411	92948139	101.1562	AA	AA	AA	AA	ab	ab	AA	AA	AA	

rs1584078	93068220	101.2479	AA	AA	AA	AA	AA	AA	AA	NoCall	AA	NoCall
rs958847	93268723	101.4012	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA
rs783856	93312435	101.4346	BB	BB	BB	BB	BB	NoCall	BB	BB	BB	BB
rs965983	93552831	101.6182	BB	BB	BB	BB	BB	BB	BB	BB	BB	BB
rs951554	93610018	101.6619	AA	AA	AA	AA	AA	ab	ab	ab	AA	ab
rs951556	93610146	101.662	BB	BB	BB	BB	BB	BB	BB	BB	ab	BB
rs727998	93802155	101.8086	AA	AA	AA	AA	AA	ab	ab	AA	AA	AA
rs1408282	93908973	101.8902	AA	AA	AA	AA	AA	ab	ab	ab	ab	ab
rs725752	94259575	102.1579	AA	AA	AA	NoCall	NoCall	ab	ab	AA	AA	AA
rs985763	94259637	102.158	BB	BB	BB	BB	BB	ab	ab	BB	BB	BB
rs1343850	94549090	102.2918	AA	AA	AA	AA	AA	ab	ab	ab	AA	ab
rs1352381	94672845	102.3449	AA	AA	AA	AA	AA	AA	AA	ab	AA	ab
rs1386917	94719581	102.3649	BB	BB	BB	BB	BB	BB	BB	ab	BB	ab
rs423246	94982708	102.4778	BB	BB	BB	BB	BB	BB	BB	BB	BB	BB
rs2021314	94994454	102.4828	AA	AA	AA	AA	AA	AA	AA	ab	ab	ab
rs952577	95044774	102.5044	AA	NoCall	NoCall	NoCall	NoCall	ab	AA	AA	BB	AA
rs62827	95125887	102.5392	BB	BB	BB	BB	BB	ab	ab	BB	BB	BB
rs2207430	95212976	102.5765	AA	AA	AA	NoCall	NoCall	AA	AA	AA	AA	AA
rs2224003	95213139	102.5766	BB	BB	BB	BB	BB	BB	BB	BB	BB	BB
rs2502004	95413900	102.6627	AA	AA	AA	AA	AA	NoCall	AA	AA	AA	AA
rs2246786	95965958	102.8995	BB	BB	BB	BB	BB	BB	BB	ab	ab	ab
rs2387156	96879221	103.2912	AA	AA	AA	AA	AA	NoCall	AA	AA	AA	AA
rs1113233	97244293	103.4478	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA
rs953878	97434550	103.5294	BB	BB	BB	BB	BB	ab	NoCall	BB	BB	BB
rs1359932	97436898	103.5304	AA	AA	AA	AA	AA	NoCall	AA	AA	ab	AA
rs1928968	97437070	103.5304	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA
rs2143389	97529402	103.57	BB	BB	BB	BB	BB	BB	BB	ab	ab	ab
rs2179537	97529651	103.5701	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA
rs2388039	97853965	103.7092	AA	AA	AA	AA	AA	AA	AA	ab	AA	ab
rs4131464	99364448	105.1737	AA	AA	AA	AA	AA	NoCall	ab	AA	AA	AA
rs1354685	100861479	106.5254	AA	AA	AA	NoCall	NoCall	NoCall	AA	ab	AA	AA
rs718268	100965880	106.6197	BB	BB	BB	BB	BB	ab	ab	ab	BB	BB
rs1857859	101001308	106.6517	BB	BB	BB	BB	BB	ab	BB	BB	ab	BB
rs2001102	101085072	106.7273	BB	BB	BB	BB	BB	BB	BB	BB	ab	BB
rs1856133	102207593	108.2534	AA	AA	AA	AA	AA	AA	AA	AA	ab	AA

rs1934009	103122488	110.0321	AA	AA	AA	AA	NoCall	AA	AA	ab	AA
rs1933989	103171750	110.1037	AA	AA	AA	AA	AA	AA	AA	AA	AA
rs1934008	103171853	110.1038	BB	BB	BB	BB	BB	BB	BB	ab	BB
rs967405	103289331	110.2745	AA	AA	AA	AA	AA	AA	AA	ab	AA
rs1431213	103305549	110.2981	BB	BB	BB	BB	NoCall	BB	BB	ab	BB
rs1073233	104687791	112.009	AA	AA	AA	AA	AA	AA	AA	AA	AA
rs559295	104858181	112.1917	AA	AA	AA	AA	AA	AA	ab	AA	AA
rs1341112	104919391	112.2574	AA	AA	AA	AA	AA	AA	ab	AA	AA
rs271857	105058386	112.4065	AA	AA	AA	AA	ab	ab	ab	ab	AA
rs271858	105058568	112.4067	AA	AA	AA	AA	ab	ab	ab	ab	AA
rs1416042	105174572	112.5312	BB	BB	BB	BB	BB	BB	BB	BB	BB
rs720225	105949023	113.2462	BB	BB	BB	BB	BB	BB	ab	ab	BB
rs722366	106080157	113.3379	AA	AA	AA	AA	NoCall	ab	AA	AA	AA
rs1417352	107005919	114.3331	AA	AA	AA	AA	ab	ab	AA	AA	AA
rs1417353	107006104	114.3334	AA	AA	AA	AA	AA	AA	ab	ab	AA
rs2077953	107264726	114.6821	BB	BB	BB	BB	ab	ab	BB	BB	BB
rs723318	107387007	114.847	AA	AA	AA	AA	AA	AA	ab	ab	AA
rs319088	107846733	115.4669	BB	BB	BB	BB	ab	ab	BB	BB	BB
rs2243869	108027221	115.7102	BB	BB	BB	BB	BB	BB	ab	ab	BB
rs2354095	108196930	115.9391	BB	BB	BB	BB	BB	BB	BB	BB	BB
rs2354094	108197032	115.9392	AA	AA	AA	AA	AA	AA	AA	AA	AA
rs2355205	108446020	116.323	BB	BB	BB	BB	NoCall	BB	ab	ab	BB
rs718174	108479948	116.5157	AA	AA	AA	AA	AA	AA	ab	ab	AA
rs217538	108590163	117.1417	AA	AA	AA	NoCall	AA	AA	AA	NoCall	AA
rs1040849	109349779	117.4965	AA	AA	AA	AA	AA	AA	AA	AA	AA
rs566164	109613154	117.5706	AA	AA	AA	AA	AA	AA	ab	AA	AA
rs59673	109655406	117.5825	BB	BB	BB	BB	BB	BB	ab	BB	BB
rs2016207	110088853	117.7045	ab	ab	BB	BB	AA	ab	BB	ab	ab
rs911022	110089298	117.7047	ab	ab	BB	BB	AA	ab	BB	ab	ab
rs1989634	110952941	117.9478	AA	ab	AA	AA	NoCall	ab	ab	AA	ab
rs1409836	112276043	118.3857	BB	BB	BB	BB	BB	BB	ab	BB	BB
rs910683	112289980	118.3907	AA	AA	AA	AA	AA	AA	ab	AA	AA
rs764196	112547827	118.4831	BB	ab	BB	BB	ab	ab	BB	BB	BB
rs949578	112812960	118.5781	ab	BB	BB	BB	ab	BB	BB	ab	AA
rs1378702	113107351	119.4909	BB	BB	ab	ab	BB	BB	ab	BB	ab