

GenoKey combinatorial data mining platform



Case-control studies investigate existing sets of data, comparing cases with controls, and can look at longitudinal data over long periods. Case/control data can include genetic information (SNPs, DNA and sequence data, genetic tests), clinical data, information from electronic medical records, or blood test results, or even data from plant or animal studies.

An example of a small data set for a proof-of-concept study could be:

- 1000 patients
- 1000 controls
- 1000 biomarkers (e.g. SNPs).

However, data mining can be used for very large sets of data, such as tens of thousands of patients and controls and millions of markers. The data can be anonymised if required, without SNP, gene or patient IDs.

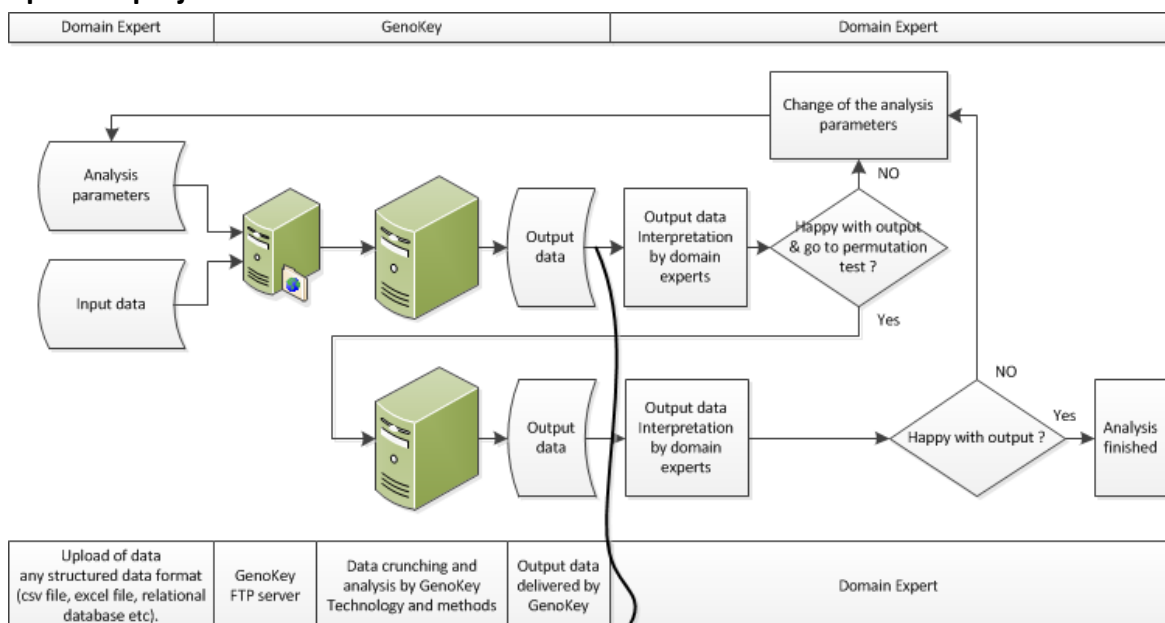
Before analysis, the data should be structured as tables of case and the control results. GenoKey's technology can handle any structured data format, including CSV files, Excel files, relational databases and others.

Projects can use a variety of analysis parameters, for example:

- N: Maximum number of biomarkers to be analyzed at a time (N-combinations)
- Cluster size: Minimum number of cases in each cluster defined by N-combinations

After analysis, GenoKey returns the combinatorial data mining output to you for further investigation by your in-house domain experts (subject matter experts).

An example of a project workflow



File Edit View Windows Session Log Action Options Tools Threads Help

display_cluster 0

Critical SNP: AVPR18_rs33976516

Critical genotype: 1

Number of patients in clusters: 41

Total number of patients: 41

SNP0	SNP1	SNP2	Genotypes	Patients
CNTNAP2_rs1024676	AVPR18_rs33976516	ANK3_rs1010556	0 1 2	6 57 94 132 304 409 5
AVPR18_rs33976516	ANK3_rs10994195	ANK3_rs1010556	1 1 0	42 51 88 114 149 153 166 :
CNTN2_rs16855045	P2RX7_rs1718161	AVPR18_rs33976516	1 1 1	51 149 166 210 231 248 330