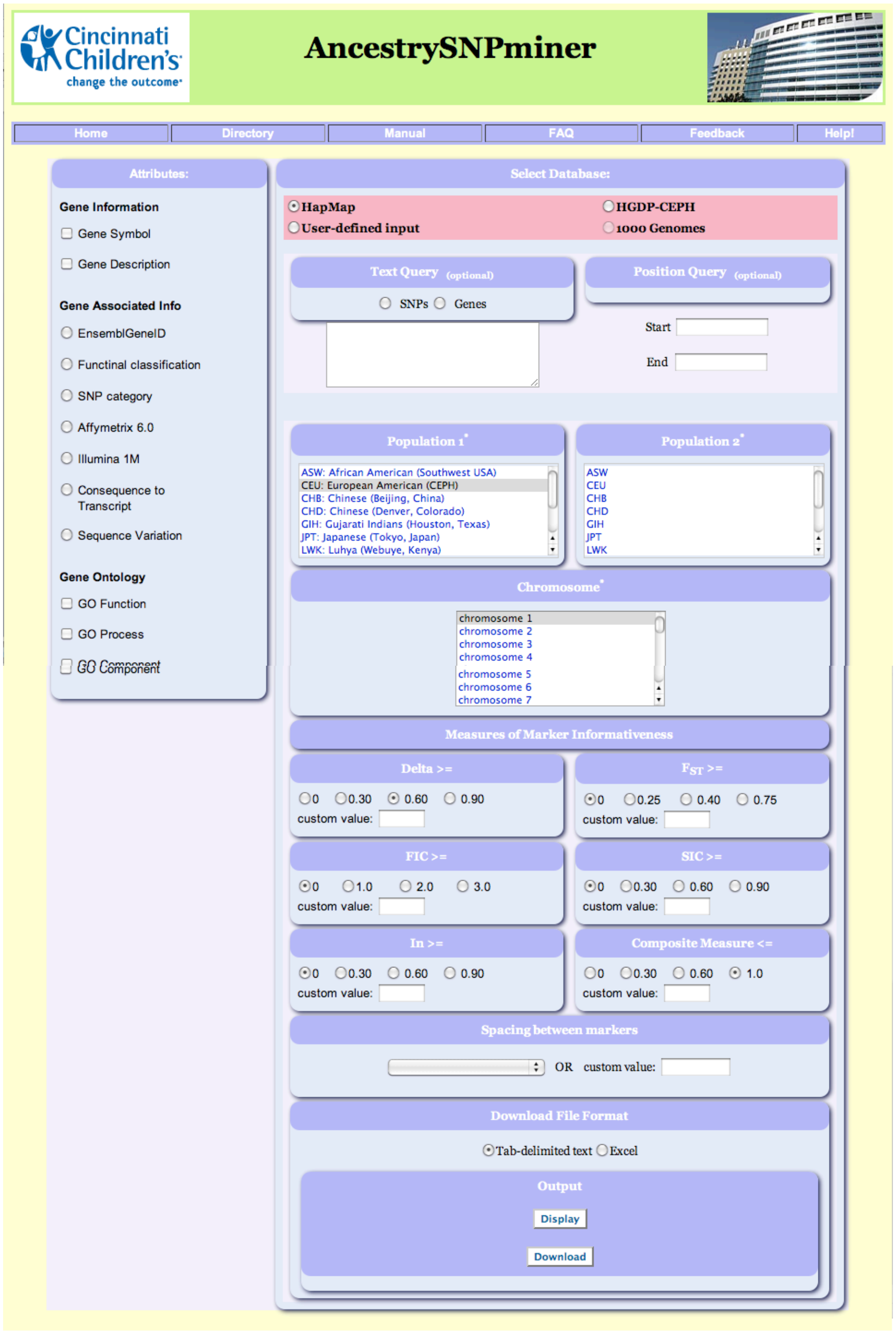
ANCESTRYSNPMINER MANUAL

AncestrySNPminer is a simple yet powerful preprocessing web based application, which is capable of mining millions of SNPs useful in ancestry and admixture mapping using multiple filters at a time. The power of this technique relies on mining ancestry informative markers (AIMs) between 66 different combinations of populations in HapMap database and 2809 different combinations of populations in Human Genome Diversity Project data to infer ancestry along the chromosomes of admixed individuals.



Select Database: ● HapMap

Delta >= ○0 ○0.30 ● 0.60 ○ 0.90 custom value:

F_{ST} >= ●0 ○0.25 ○ 0.40 ○ 0.75 custom value:

FIC >= ●0 ○1.0 ○ 2.0 ○ 3.0 custom value:

SIC >= ●0 ○0.30 ○ 0.60 ○ 0.90 custom value:

In >= ●0 ○0.30 ○ 0.60 ○ 0.90 custom value:

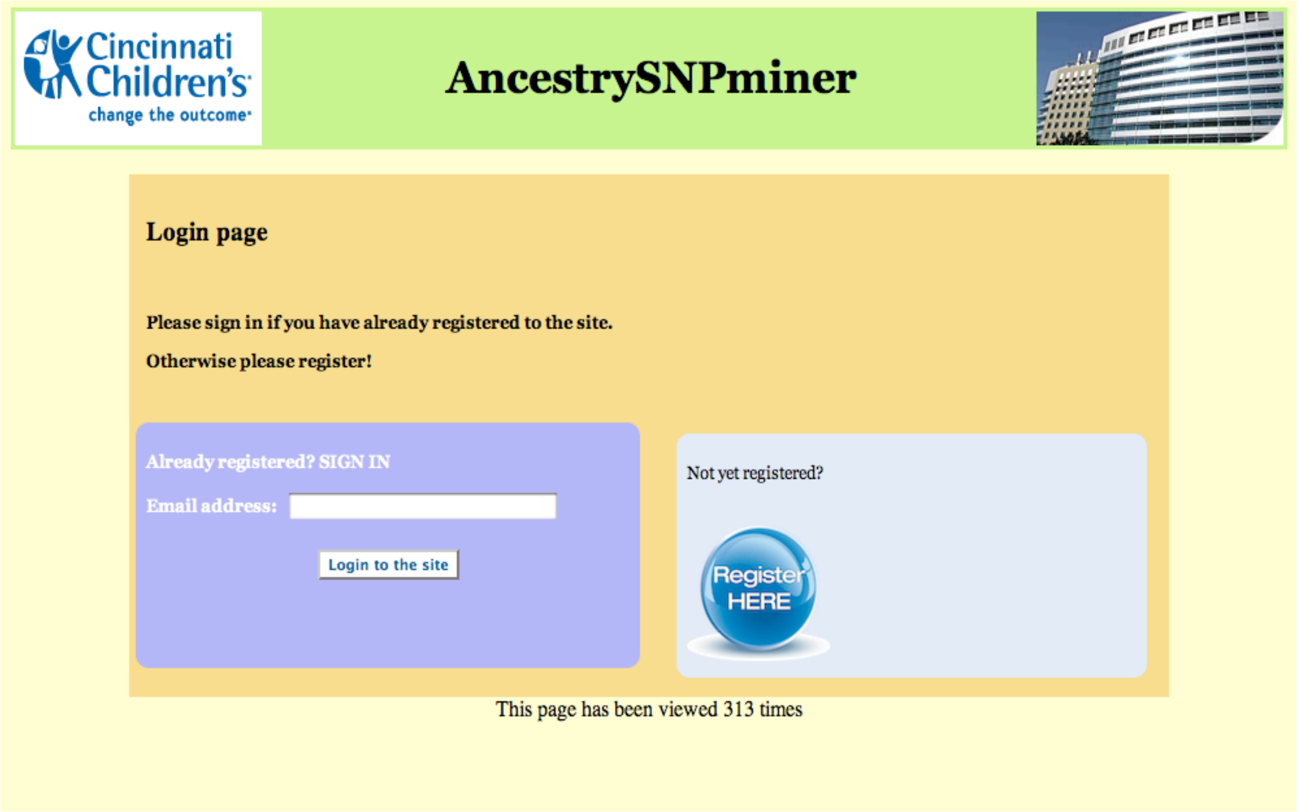
Composite Measure <= ○0 ○0.30 ○ 0.60 ● 1.0 custom value:

Download File Format: ● Tab-delimited text ○ Excel

Steps to use AncestrySNPminer:

Login page:

User has to sign-in with their email-id in order to access AncestrySNPminer application.



AncestrySNPminer

Login page

Please sign in if you have already registered to the site.

Otherwise please register!

Already registered? SIGN IN

Email address:

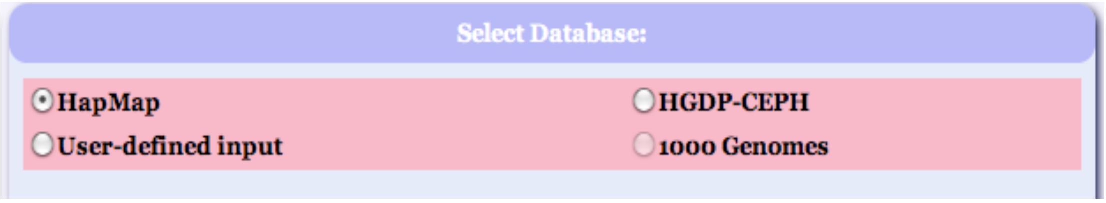
Not yet registered?

This page has been viewed 313 times

First time users need to register by filling out a quick registration form.

1. Select Database:

HapMap and HGDP-CEPH Human Genome Diversity Project are the databases available with AncestrySNPminer application. User-defined input allows user to upload the two ancestry population files of choice. We have future plans to include 1000 Genomes Project database.

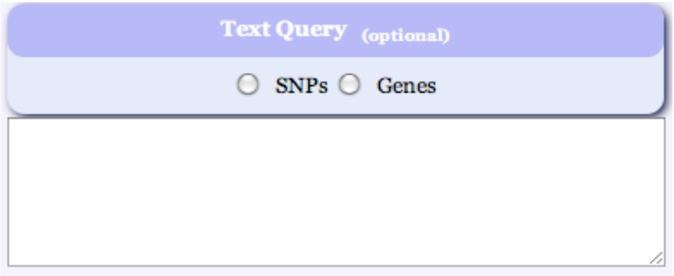


Select Database:

● HapMap ○ HGDP-CEPH
○ User-defined input ○ 1000 Genomes

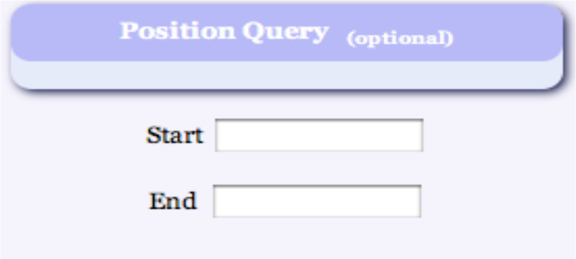
2. Text Query *(Optional)*:

User can type in or copy-paste the SNP ID's or Gene Symbols separated by a new line.
Note: For Query through Genes select Genes option. For query through SNPs select SNPs.



3. Position Query *(Optional)*:

Position Query allows the user to mine SNPs near a region of interest by specifying the start position and end position of the chromosome. For example, below is the position query that returns all the records of SNPs present between start position 10905506 and end position 238767839 in the selected populations and chromosomes.

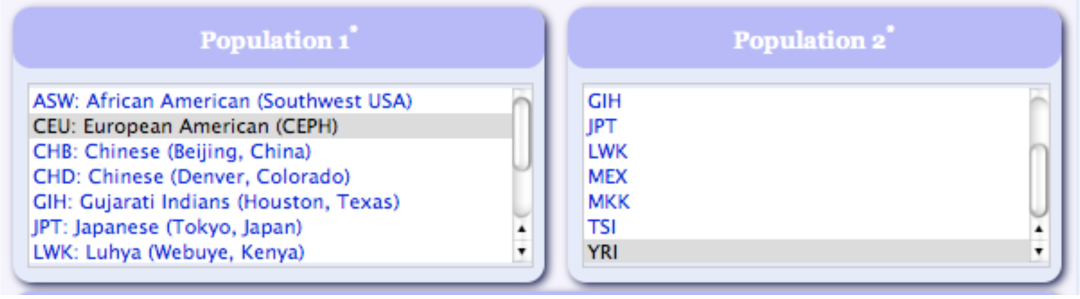


4. Choose Populations:

User has to select any two populations to retrieve Ancestry Informative Markers.

HapMap:

The web interface allows the user to select 66 different combinations of populations. In the below snapshot populations CEU and YRI are selected.



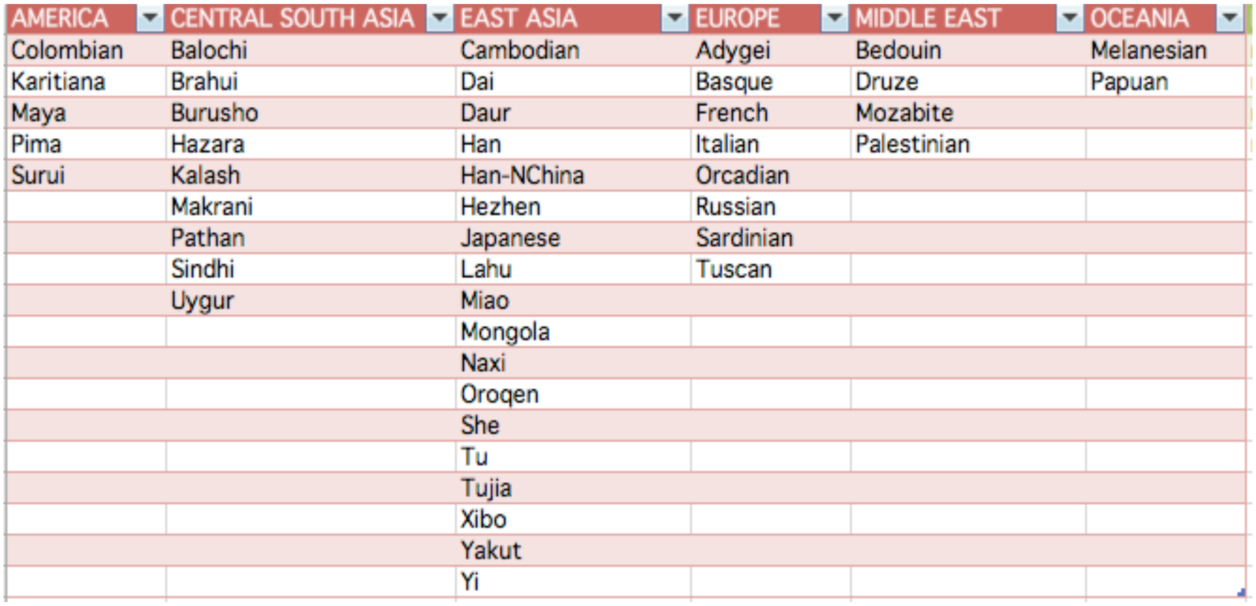
HapMap Phase III release includes 11 different populations:

Code	Populations
ASW	African ancestry in Southwest USA
CEU	Utah residents with Northern and Western European ancestry from the CEPH collection
CHB	Han Chinese in Beijing, China
CHD	Chinese in Metropolitan Denver, Colorado
GIH	Gujarati Indians in Houston, Texas
JPT	Japanese in Tokyo, Japan
LWK	Luhya in Webuye, Kenya
MXL	Mexican ancestry in Los Angeles, California
MKK	Maasai in Kinyawa, Kenya
TSI	Toscani in Italia
YRI	Yoruba in Ibadan, Nigeria

HGDP:

The web interface allows the user to select 2809 different combinations of population. In the below snapshot populations Colombian of America and Yoruba of Africa are selected.

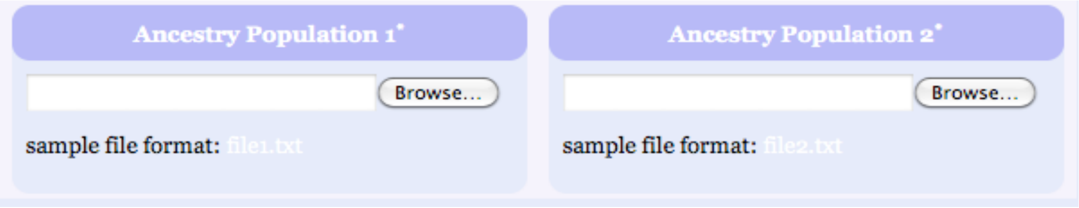
HGDP-CEPH Human Genome Diversity Project consists of 53 populations from 7 different regions:



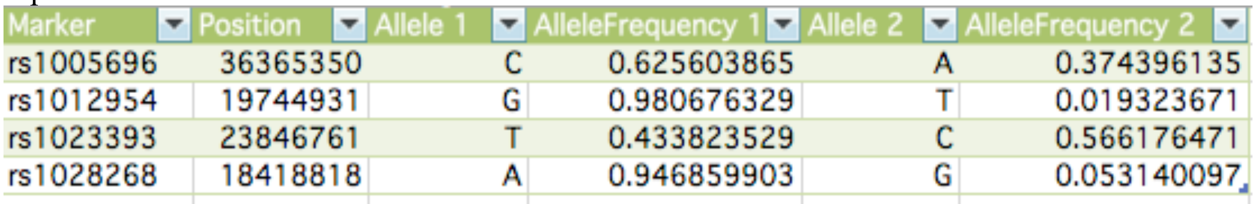
AMERICA	CENTRAL SOUTH ASIA	EAST ASIA	EUROPE	MIDDLE EAST	OCEANIA
Colombian	Balochi	Cambodian	Adygei	Bedouin	Melanesian
Karitiana	Brahui	Dai	Basque	Druze	Papuan
Maya	Burusho	Daur	French	Mozabite	
Pima	Hazara	Han	Italian	Palestinian	
Surui	Kalash	Han-NChina	Orcadian		
	Makrani	Hezhen	Russian		
	Pathan	Japanese	Sardinian		
	Sindhi	Lahu	Tuscan		
	Uygur	Miao			
		Mongola			
		Naxi			
		Oroqen			
		She			
		Tu			
		Tujia			
		Xibo			
		Yakut			
		Yi			

User-Defined Input:

User can upload population 1 and population 2 files in a specific format as given below.



Input File Format:

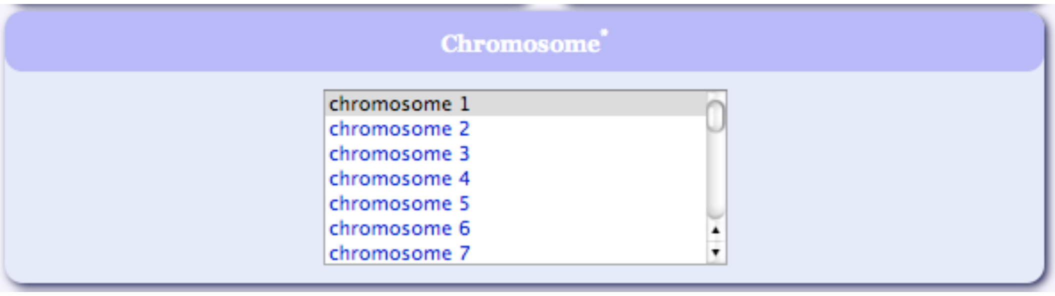


Marker	Position	Allele 1	AlleleFrequency 1	Allele 2	AlleleFrequency 2
rs1005696	36365350	C	0.625603865	A	0.374396135
rs1012954	19744931	G	0.980676329	T	0.019323671
rs1023393	23846761	T	0.433823529	C	0.566176471
rs1028268	18418818	A	0.946859903	G	0.053140097

5. Select Chromosomes:

The user has options to select either any one chromosome or multiple chromosomes or all the chromosomes.

Note: Hold Control key in case of windows OS and command key in the case of Mac OS to select multiple chromosomes. Use shift key to select all the chromosomes.

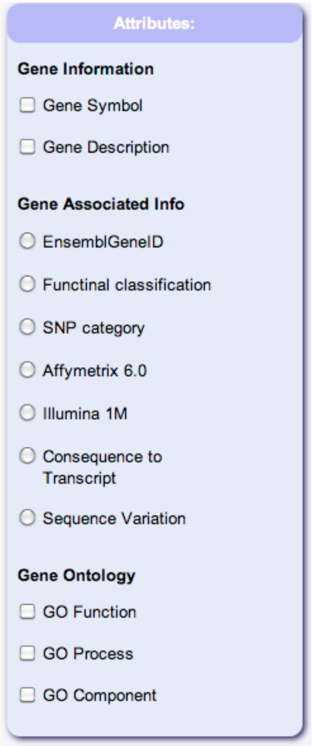


Chromosome*

chromosome 1
chromosome 2
chromosome 3
chromosome 4
chromosome 5
chromosome 6
chromosome 7

6. Attributes *(Optional)*:

User is provided with an option to select different attributes related to the SNPs.
Gene Information provides Gene Symbol and description linked with the SNP.



Attributes:

Gene Information

☐ Gene Symbol
☐ Gene Description

Gene Associated Info

○ EnsemblGeneID
○ Functinal classification
○ SNP category
○ Affymetrix 6.0
○ Illumina 1M
○ Consequence to Transcript
○ Sequence Variation

Gene Ontology

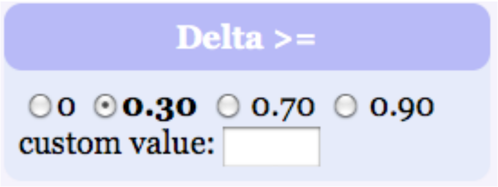
☐ GO Function
☐ GO Process
☐ GO Component

Gene associated information provides Ensembl GeneID, Biotype, dbSNP Synonym, Affymetrix chip ID, Illumina chip ID, Consequence to Transcript or Consequence Allele information.
Gene Ontology information gives the function, biological process and cellular component for the SNPs.

FILTERS:

7. Delta:

User can use either the provided filter value by selecting the radio button or may input a custom value. A marker with delta = 1 provides perfect information regarding ancestry whereas a marker with delta = 0 carries no information. Significant cutoff value for delta is 0.3.
Note: Make sure to unselect the radio button if you are using custom value.

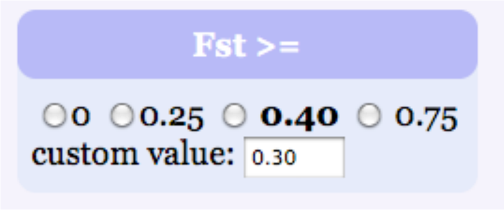


Delta >=

○ 0 ● **0.30** ○ 0.70 ○ 0.90
custom value:

8. F Statistics (F_{ST}):

User has an option to select either radio button or provide custom value. Values of FST can range between 0 and 1. A high FST value implies a considerable degree of differentiation between populations. Significant cutoff value for FST is 0.4.
Note: Make sure to unselect the radio button if you are using custom value.

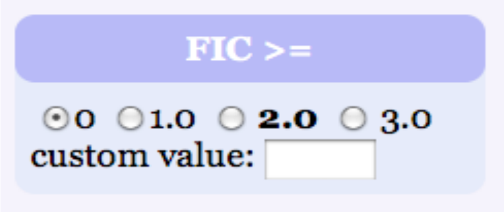


Fst >=

○ 0 ○ 0.25 ○ **0.40** ○ 0.75
custom value: 0.30

9. Fisher Information Content (FIC):

User has an option to select either radio button or provide custom value. Significant cutoff value for FIC is 2.0.
Note: Make sure to unselect the radio button if you are using custom value.

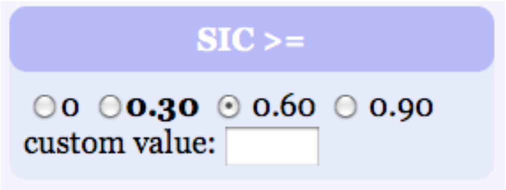


FIC >=

● 0 ○ 1.0 ○ **2.0** ○ 3.0
custom value:

10. Shannon Information Content (SIC):

User has an option to select either radio button or provide custom value. Significant cutoff value for SIC is 0.3.
Note: Make sure to unselect the radio button if you are using custom value.



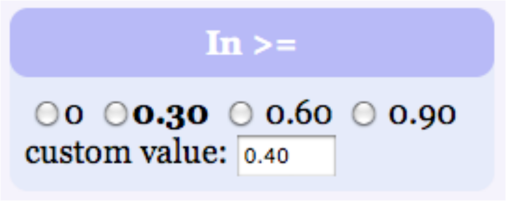
SIC >=

○ 0 ○ 0.30 ● 0.60 ○ 0.90

custom value:

11. Informativeness for assignment measure (In):

User has an option to select either radio button or provide custom value. Significant cutoff value for In is 0.3.
Note: Make sure to unselect the radio button if you are using custom value.



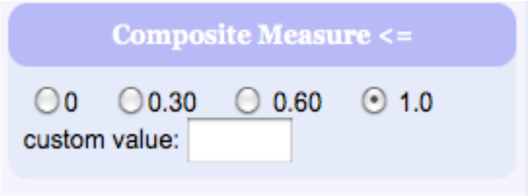
In >=

○ 0 ○ 0.30 ○ 0.60 ○ 0.90

custom value: 0.40

12. Composite Measure:

Composite measure is calculated by ranking each measures individually, dividing average of ranks for the five measures by highest average rank.
Composite measure of score 0.0 means most informative marker while 1.0 means the least.



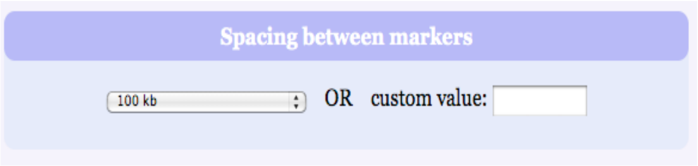
Composite Measure <=

○ 0 ○ 0.30 ○ 0.60 ● 1.0

custom value:

13. Spacing between markers *(Optional)*:

User can set a value for spacing between markers by selecting the values provided or by entering a custom value.

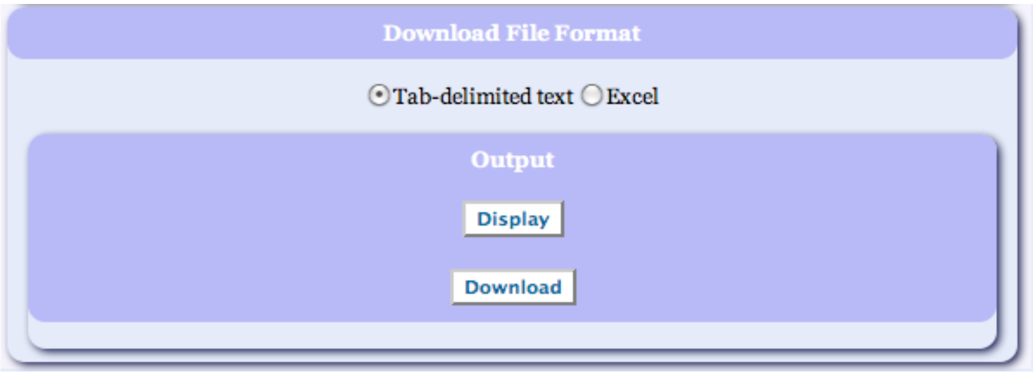


Spacing between markers

100 kb OR custom value:

14. Output:

User has an option to view the result on the web browser or to download. Select display to get the result on the browser or select download and choose the file format to save on local hard drive.

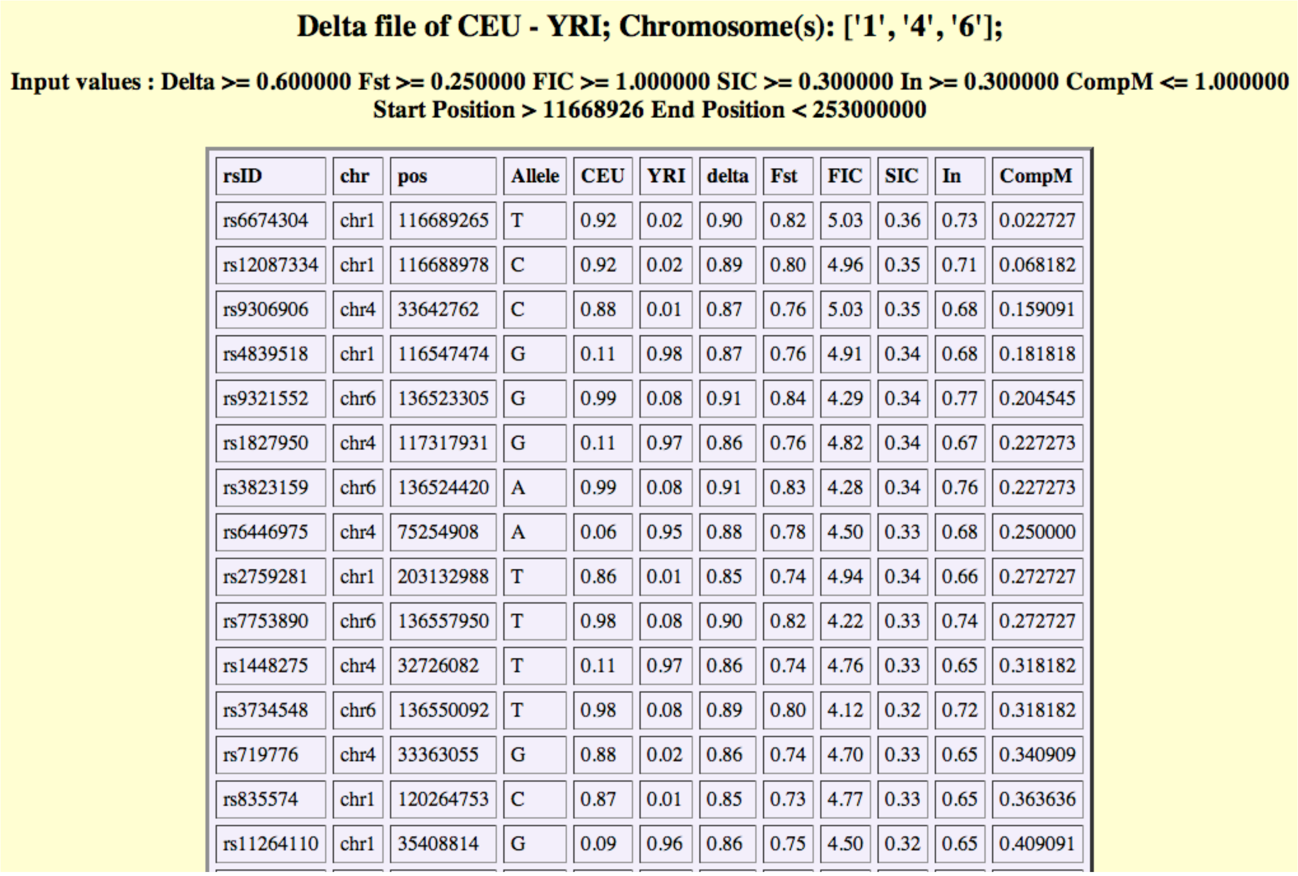


Download File Format

⦿ Tab-delimited text ○ Excel

15. Display Output:

Snapshot of the result displayed on the browser window when user clicks the display button using chromosome one between populations CEU and YRI. The input values of the filters are also displayed for verification. Output is sorted based on composite measure i.e., most informative to least informative markers.



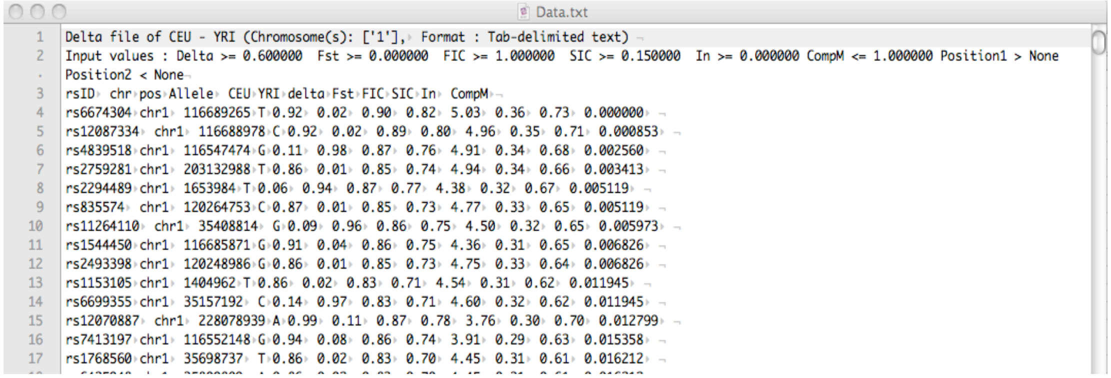
Delta file of CEU - YRI; Chromosome(s): ['1', '4', '6'];

Input values : Delta >= 0.600000 Fst >= 0.250000 FIC >= 1.000000 SIC >= 0.300000 In >= 0.300000 CompM <= 1.000000 Start Position > 11668926 End Position < 253000000

rsID	chr	pos	Allele	CEU	YRI	delta	Fst	FIC	SIC	In	CompM
rs6674304	chr1	116689265	T	0.92	0.02	0.90	0.82	5.03	0.36	0.73	0.022727
rs12087334	chr1	116688978	C	0.92	0.02	0.89	0.80	4.96	0.35	0.71	0.068182
rs9306906	chr4	33642762	C	0.88	0.01	0.87	0.76	5.03	0.35	0.68	0.159091
rs4839518	chr1	116547474	G	0.11	0.98	0.87	0.76	4.91	0.34	0.68	0.181818
rs9321552	chr6	136523305	G	0.99	0.08	0.91	0.84	4.29	0.34	0.77	0.204545
rs1827950	chr4	117317931	G	0.11	0.97	0.86	0.76	4.82	0.34	0.67	0.227273
rs3823159	chr6	136524420	A	0.99	0.08	0.91	0.83	4.28	0.34	0.76	0.227273
rs6446975	chr4	75254908	A	0.06	0.95	0.88	0.78	4.50	0.33	0.68	0.250000
rs2759281	chr1	203132988	T	0.86	0.01	0.85	0.74	4.94	0.34	0.66	0.272727
rs7753890	chr6	136557950	T	0.98	0.08	0.90	0.82	4.22	0.33	0.74	0.272727
rs1448275	chr4	32726082	T	0.11	0.97	0.86	0.74	4.76	0.33	0.65	0.318182
rs3734548	chr6	136550092	T	0.98	0.08	0.89	0.80	4.12	0.32	0.72	0.318182
rs719776	chr4	33363055	G	0.88	0.02	0.86	0.74	4.70	0.33	0.65	0.340909
rs835574	chr1	120264753	C	0.87	0.01	0.85	0.73	4.77	0.33	0.65	0.363636
rs11264110	chr1	35408814	G	0.09	0.96	0.86	0.75	4.50	0.32	0.65	0.409091

16. Tab delimited text file:

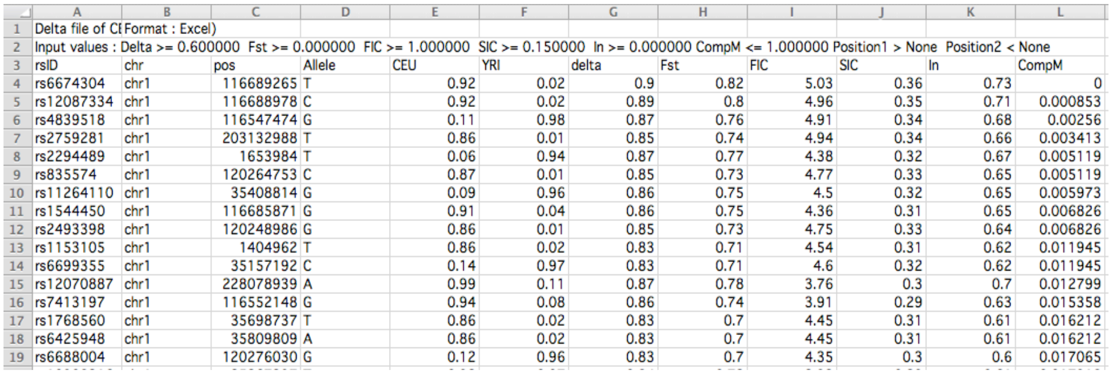
Snapshot of tab delimited text file when the user selects tab delimited text format and clicks download button.



```
Delta file of CEU - YRI (Chromosome(s): ['1'],	Format : Tab-delimited text)
Input values : Delta >= 0.600000  Fst >= 0.000000  FIC >= 1.000000  SIC >= 0.150000  In >= 0.000000 CompM <= 1.000000 Position1 > None
Position2 < None
rsID	chr	pos	Allele	CEU	YRI	delta	Fst	FIC	SIC	In	CompM
rs6674304	chr1	116689265	T	0.92	0.02	0.90	0.82	5.03	0.36	0.73	0.000000
rs12087334	chr1	116688978	C	0.92	0.02	0.89	0.80	4.96	0.35	0.71	0.000853
rs4839518	chr1	116547474	G	0.11	0.98	0.87	0.76	4.91	0.34	0.68	0.002560
rs2759281	chr1	203132988	T	0.86	0.01	0.85	0.74	4.94	0.34	0.66	0.003413
rs2294489	chr1	1653984	T	0.06	0.94	0.87	0.77	4.38	0.32	0.67	0.005119
rs835574	chr1	120264753	C	0.87	0.01	0.85	0.73	4.77	0.33	0.65	0.005119
rs11264110	chr1	35408814	G	0.09	0.96	0.86	0.75	4.50	0.32	0.65	0.005973
rs1544450	chr1	116685871	G	0.91	0.04	0.86	0.75	4.36	0.31	0.65	0.006826
rs2493398	chr1	120248986	G	0.86	0.01	0.85	0.73	4.75	0.33	0.64	0.006826
rs1153105	chr1	1404962	T	0.86	0.02	0.83	0.71	4.54	0.31	0.62	0.011945
rs6699355	chr1	35157192	C	0.14	0.97	0.83	0.71	4.60	0.32	0.62	0.011945
rs12070887	chr1	228078939	A	0.99	0.11	0.87	0.78	3.76	0.30	0.70	0.012799
rs7413197	chr1	116552148	G	0.94	0.08	0.86	0.74	3.91	0.29	0.63	0.015358
rs1768560	chr1	35698737	T	0.86	0.02	0.83	0.70	4.45	0.31	0.61	0.016212
```

17. Excel Format File:

Snapshot of Excel file when user selects Excel format and clicks download button.



	A	B	C	D	E	F	G	H	I	J	K	L
1	Delta file of CI	Format : Excel)										
2	Input values : Delta >= 0.600000 Fst >= 0.000000 FIC >= 1.000000 SIC >= 0.150000 In >= 0.000000 CompM <= 1.000000 Position1 > None Position2 < None											
3	rsID	chr	pos	Allele	CEU	YRI	delta	Fst	FIC	SIC	In	CompM
4	rs6674304	chr1	116689265	T	0.92	0.02	0.9	0.82	5.03	0.36	0.73	0
5	rs12087334	chr1	116688978	C	0.92	0.02	0.89	0.8	4.96	0.35	0.71	0.000853
6	rs4839518	chr1	116547474	G	0.11	0.98	0.87	0.76	4.91	0.34	0.68	0.00256
7	rs2759281	chr1	203132988	T	0.86	0.01	0.85	0.74	4.94	0.34	0.66	0.003413
8	rs2294489	chr1	1653984	T	0.06	0.94	0.87	0.77	4.38	0.32	0.67	0.005119
9	rs835574	chr1	120264753	C	0.87	0.01	0.85	0.73	4.77	0.33	0.65	0.005119
10	rs11264110	chr1	35408814	G	0.09	0.96	0.86	0.75	4.5	0.32	0.65	0.005973
11	rs1544450	chr1	116685871	G	0.91	0.04	0.86	0.75	4.36	0.31	0.65	0.006826
12	rs2493398	chr1	120248986	G	0.86	0.01	0.85	0.73	4.75	0.33	0.64	0.006826
13	rs1153105	chr1	1404962	T	0.86	0.02	0.83	0.71	4.54	0.31	0.62	0.011945
14	rs6699355	chr1	35157192	C	0.14	0.97	0.83	0.71	4.6	0.32	0.62	0.011945
15	rs12070887	chr1	228078939	A	0.99	0.11	0.87	0.78	3.76	0.3	0.7	0.012799
16	rs7413197	chr1	116552148	G	0.94	0.08	0.86	0.74	3.91	0.29	0.63	0.015358
17	rs1768560	chr1	35698737	T	0.86	0.02	0.83	0.7	4.45	0.31	0.61	0.016212
18	rs6425948	chr1	35809809	A	0.86	0.02	0.83	0.7	4.45	0.31	0.61	0.016212
19	rs6688004	chr1	120276030	G	0.12	0.96	0.83	0.7	4.35	0.3	0.6	0.017065