Solanaceae Coordinated Agricultural Project

Working with Infinium Genotyping Data

Allen Van Deynze

UC Davis

November 16th, 2010

USDA United States Department of Agriculture National Institute of Food and Agriculture

Sol CAP

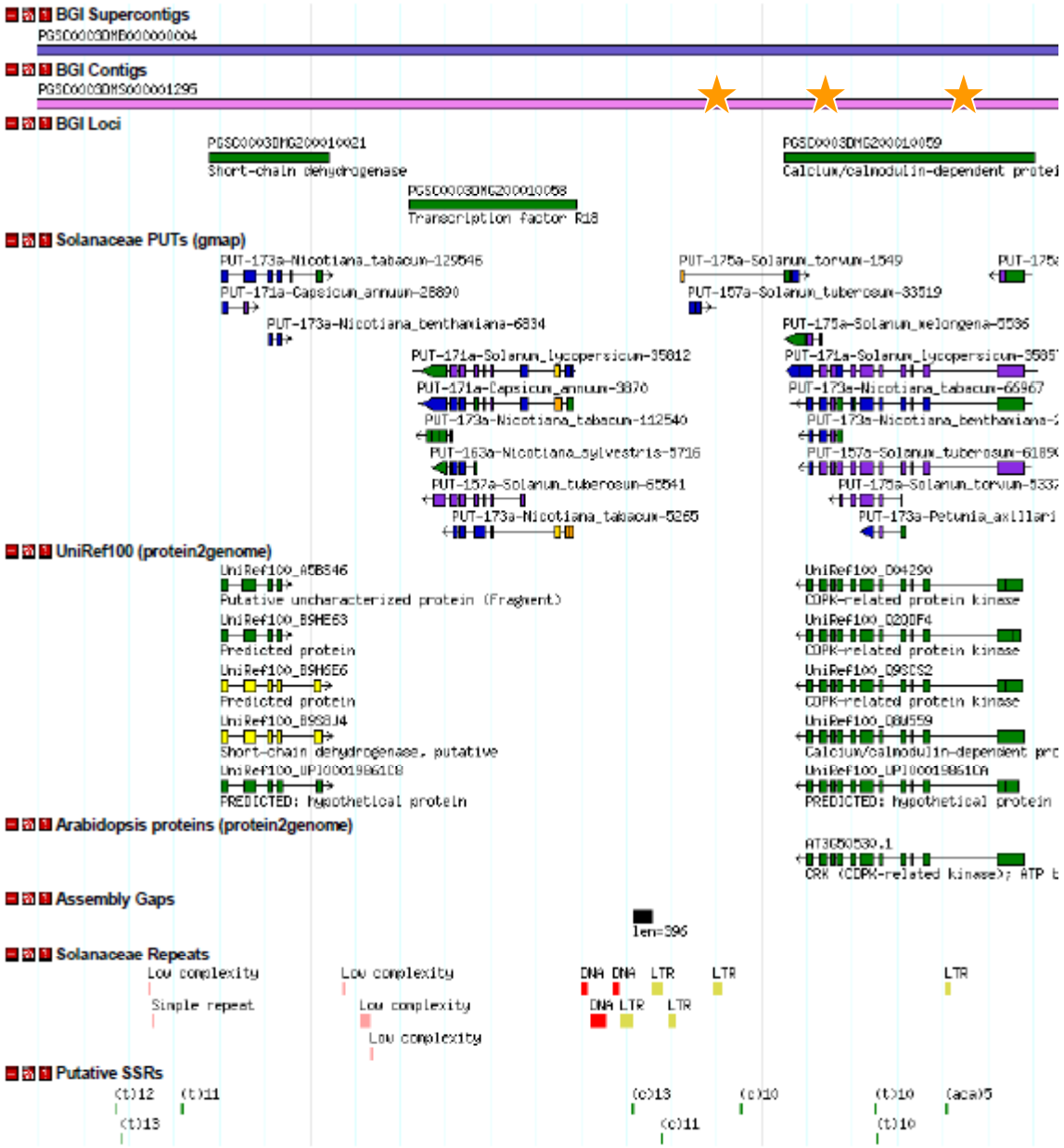
Marker development considerations

- ✓How to sequence?
- ✓What part of the DNA to sequence?

SNPs

- ➢What lines to sequence?
- ➢How many lines to sequence?
- ➢Qualifying genotypic data

Genome Browser

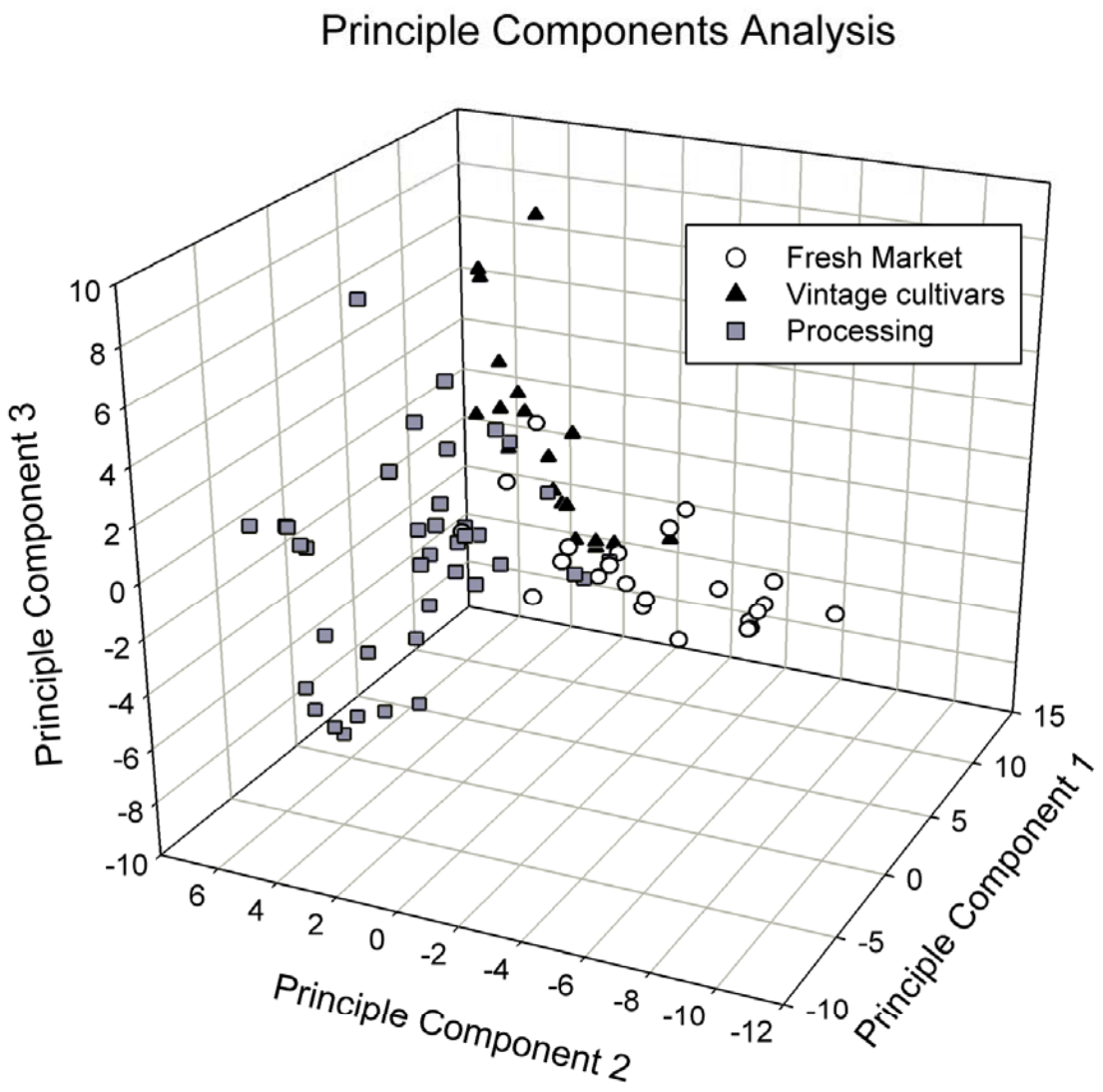


Converting Sequence to SNPs

Consider:

- Lines sequenced
- Duplications
- Heterozygosity

Basis of SNPs is important



Robbins et al. 2011

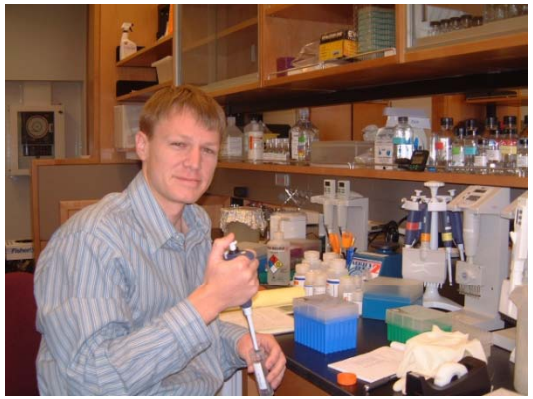
cDNA Libraries for Sequencing

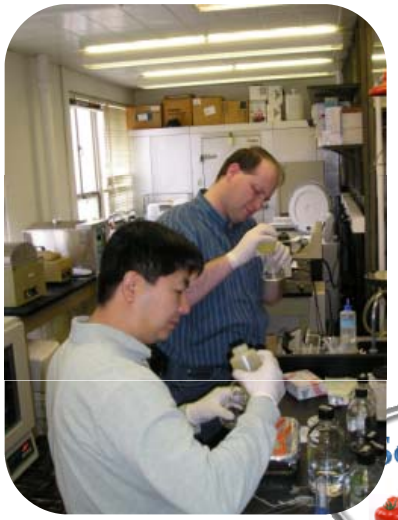
Using Illumina Genome Analyzer

- OH08-6405 (CA fresh-market)
- FL7600 (Florida fresh-market)
- NC84173 (NC fresh-market)
- OH9242 (Ohio processing)
- PI114490 (wild cherry tomato)
- PI128216 (wild current tomato)

Root
Leaf
Flower
Fruit
Callus

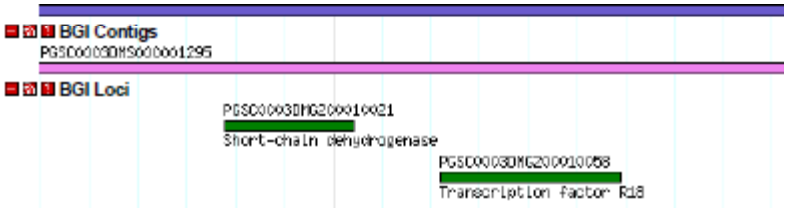
- H1706 (Midwest Processing; Genome sequence)
- TA496 (CA Processing; EST sequence)



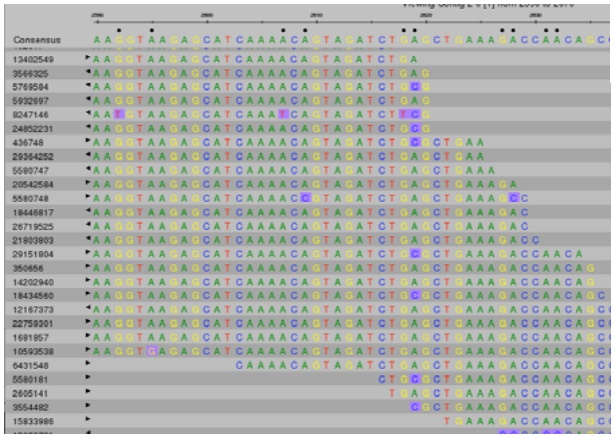


SNP Filters

Sequence Alignment



Allele depth



Allele Frequency

- OH08-6405 (CA fresh-market)
- FL7600 (Florida fresh-market)
- NC84173 (NC fresh-market)
- OH9242 (Ohio processing)
- PI114490 (wild cherry tomato)
- PI128216 (wild current tomato)

Candidate genes

Results of validation Test-96 Bead express

After filters: 28,380 putative SNPs
Cultivated: 9200

Category	% PM	Total Number
Total Number	97%	9,200
Cultivated	70%	6,500
Processing	60%	5,600
Fresh-Market	43%	4,000
Vintage	34%	3,500
S. pimpinellifolium	40%	3,750
M82 x LA716	30%	2,760

10,000 Bead-Infinium Tomato Chip Design

GAII SNPs cultivated	4,548
Unique Community SNPs	1,387
Unique Community Gene SNPs	552
Scaffold Coverage SNPs	2,297
Total SNPs	8,784
Transversion SNPs	1,216
Total Probes	10,000

SNPs from SolCAP, Trait Genetics and INRA

Design: John Hamilton, Robin Buell, David Francis, Allen Van Deynze

Solanaceae Coordinated Agricultural Project

Working with Infinium Genotyping Data

Allen Van Deynze

UC Davis

November 16th, 2010

USDA United States Department of Agriculture National Institute of Food and Agriculture

Sol CAP

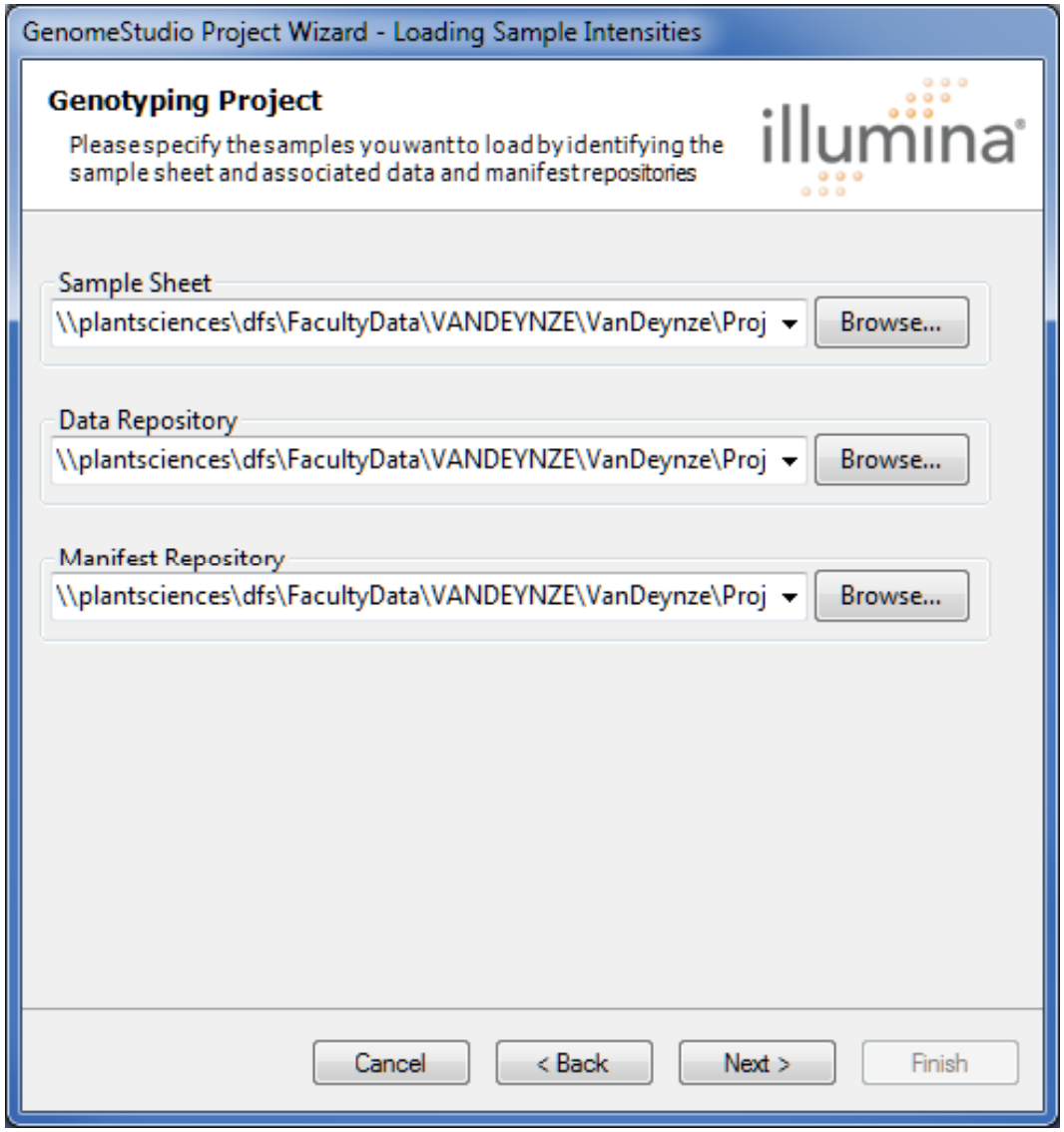
Marker development considerations

- ✓How to sequence?
- ✓What part of the DNA to sequence?

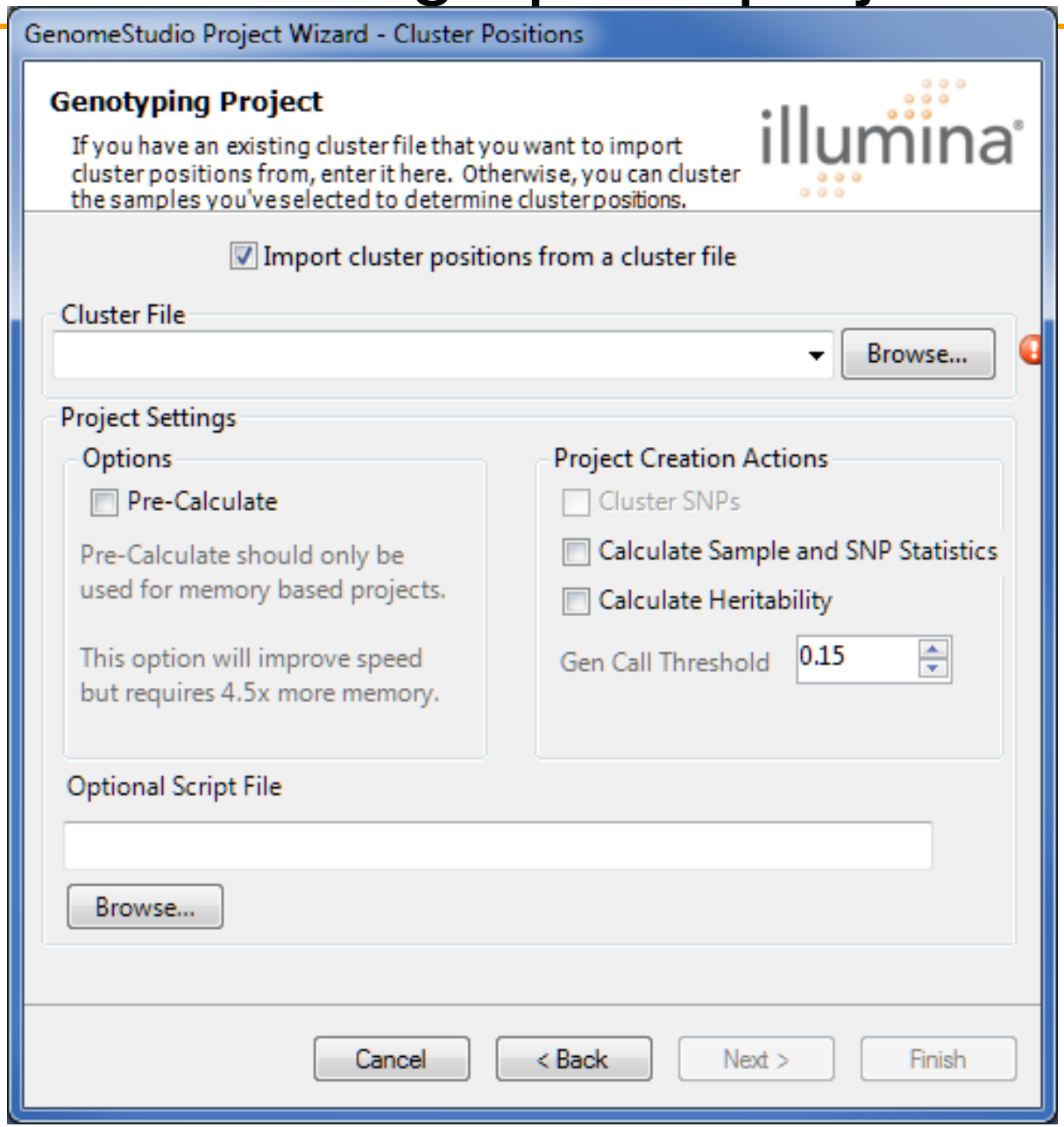
SNPs

- ➢What lines to sequence?
- ➢How many lines to sequence?
- ➢Qualifying genotypic data

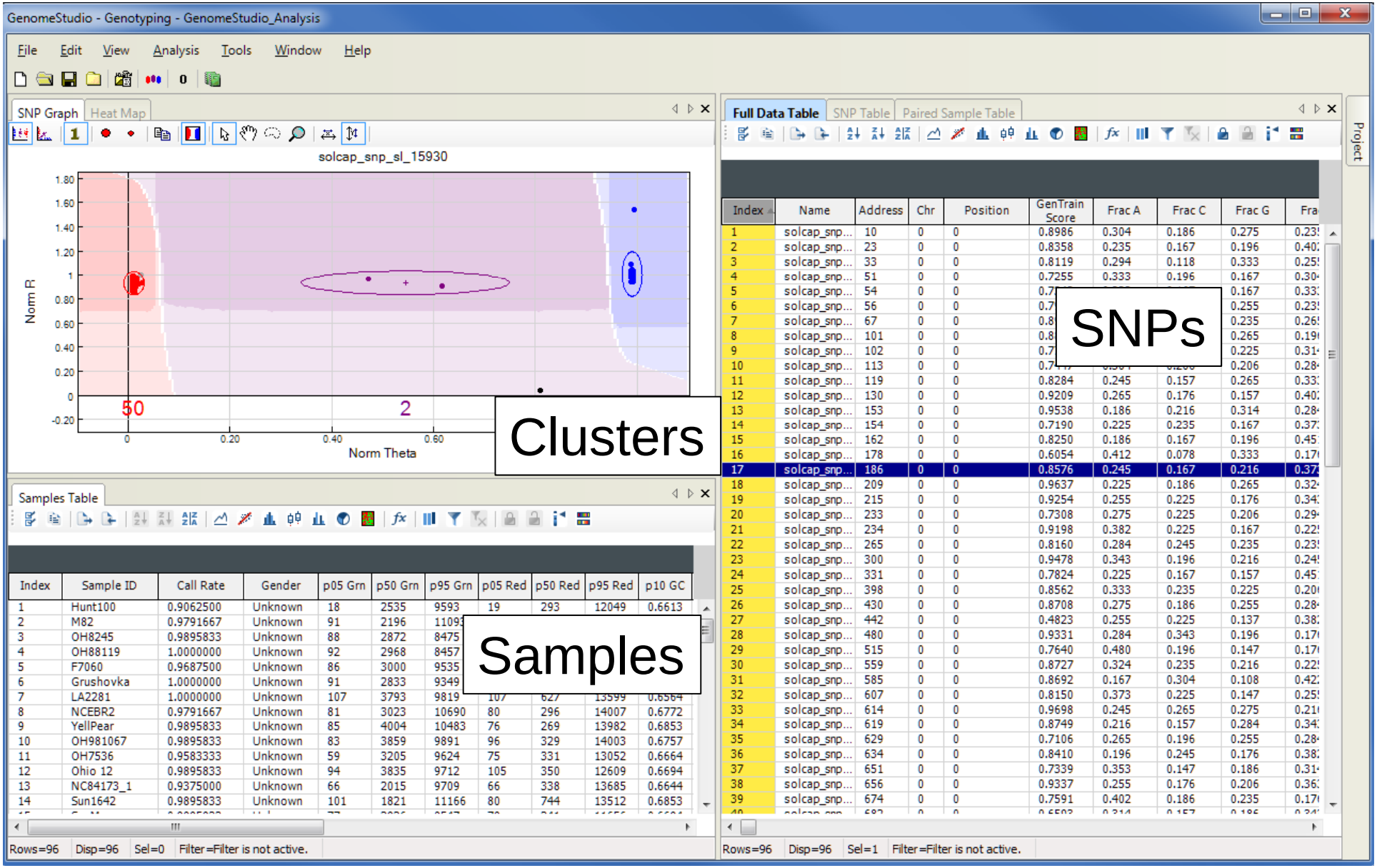
Qualifying genotyping data-not about bands



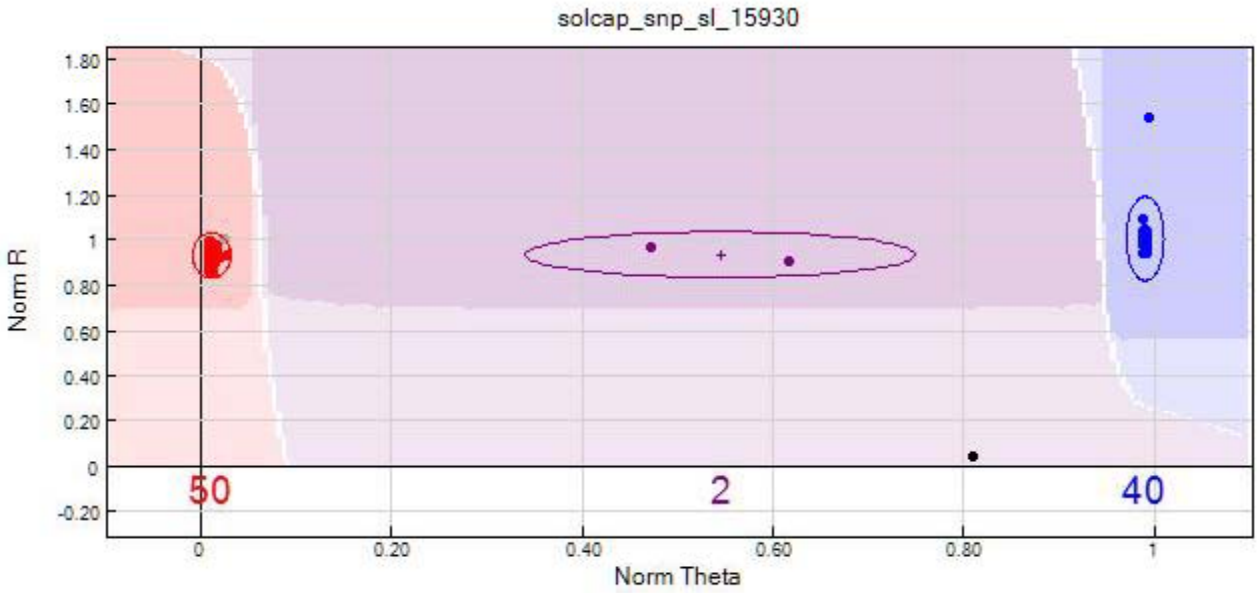
Setting up the project



The workbench

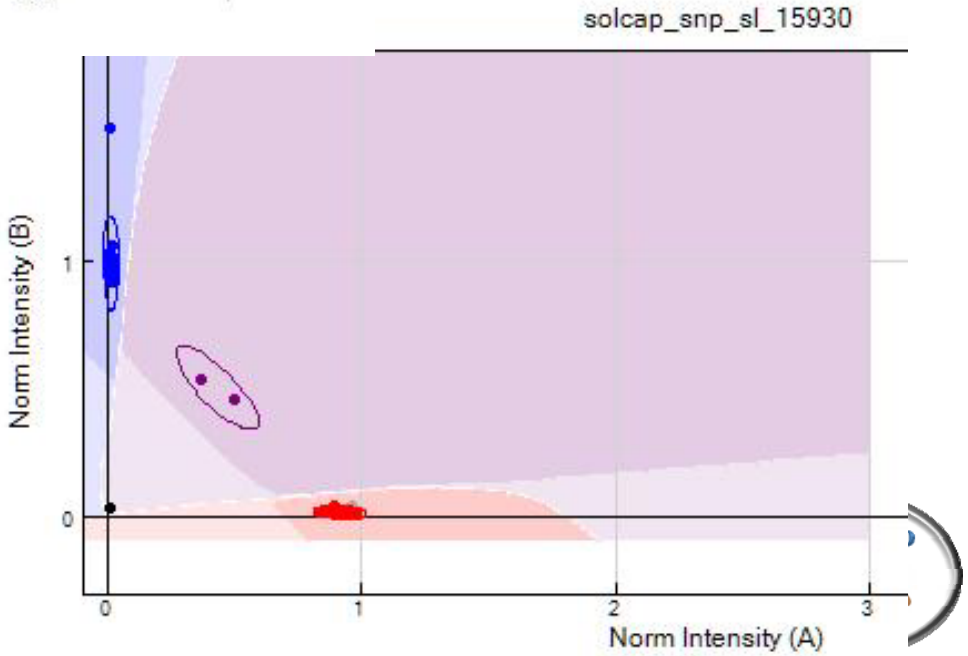


Views

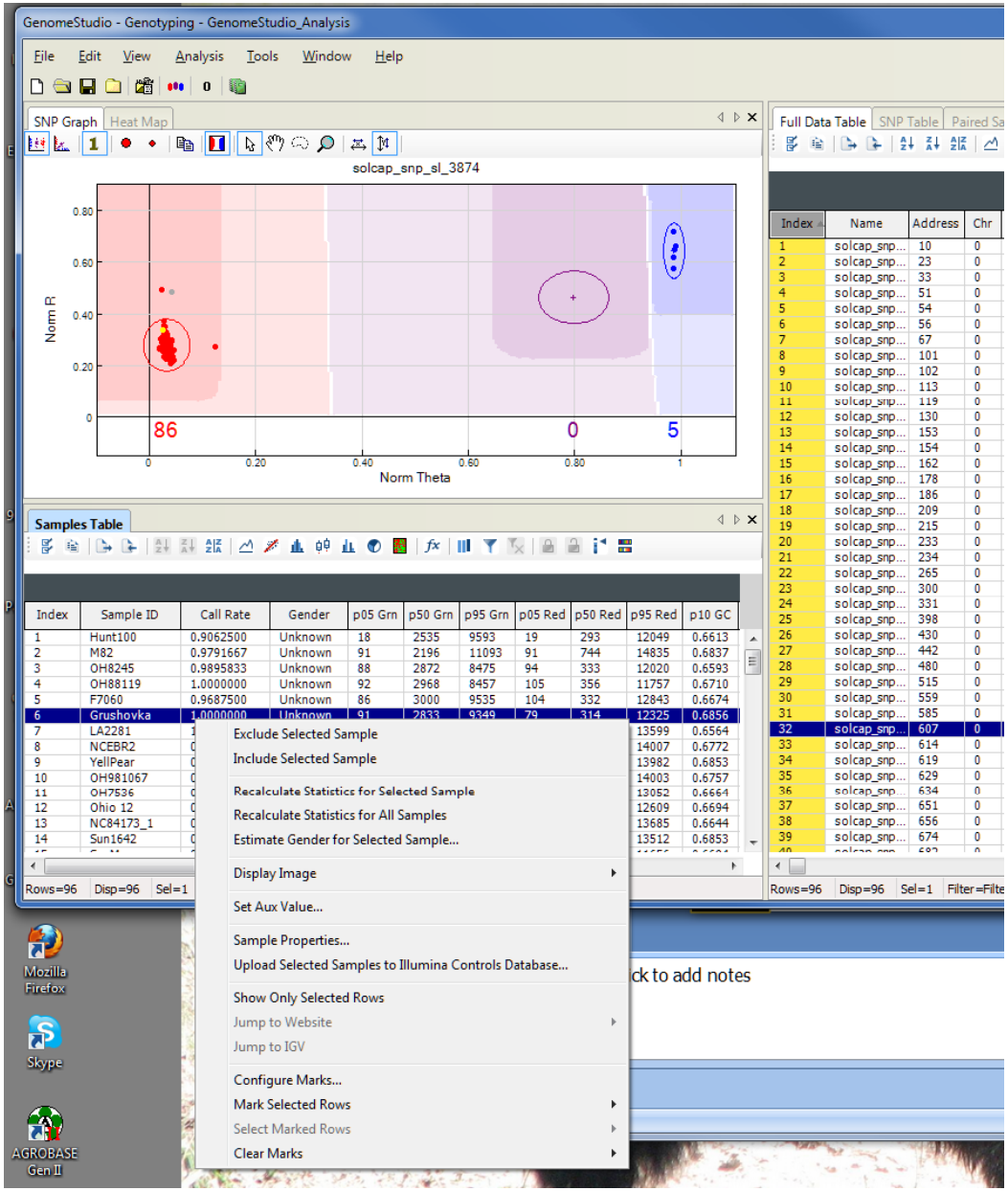


Polar

Cartesian

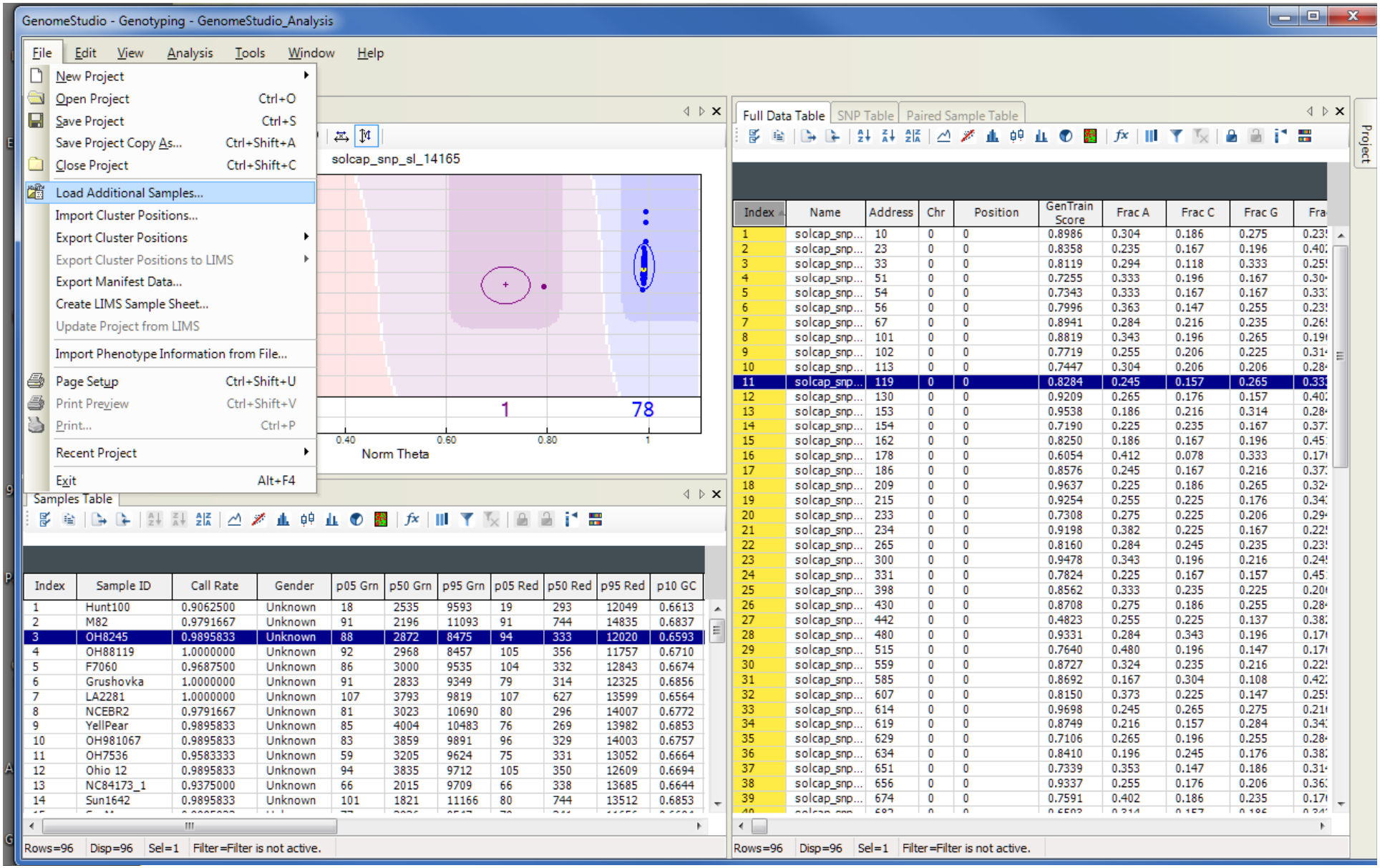


Managing samples

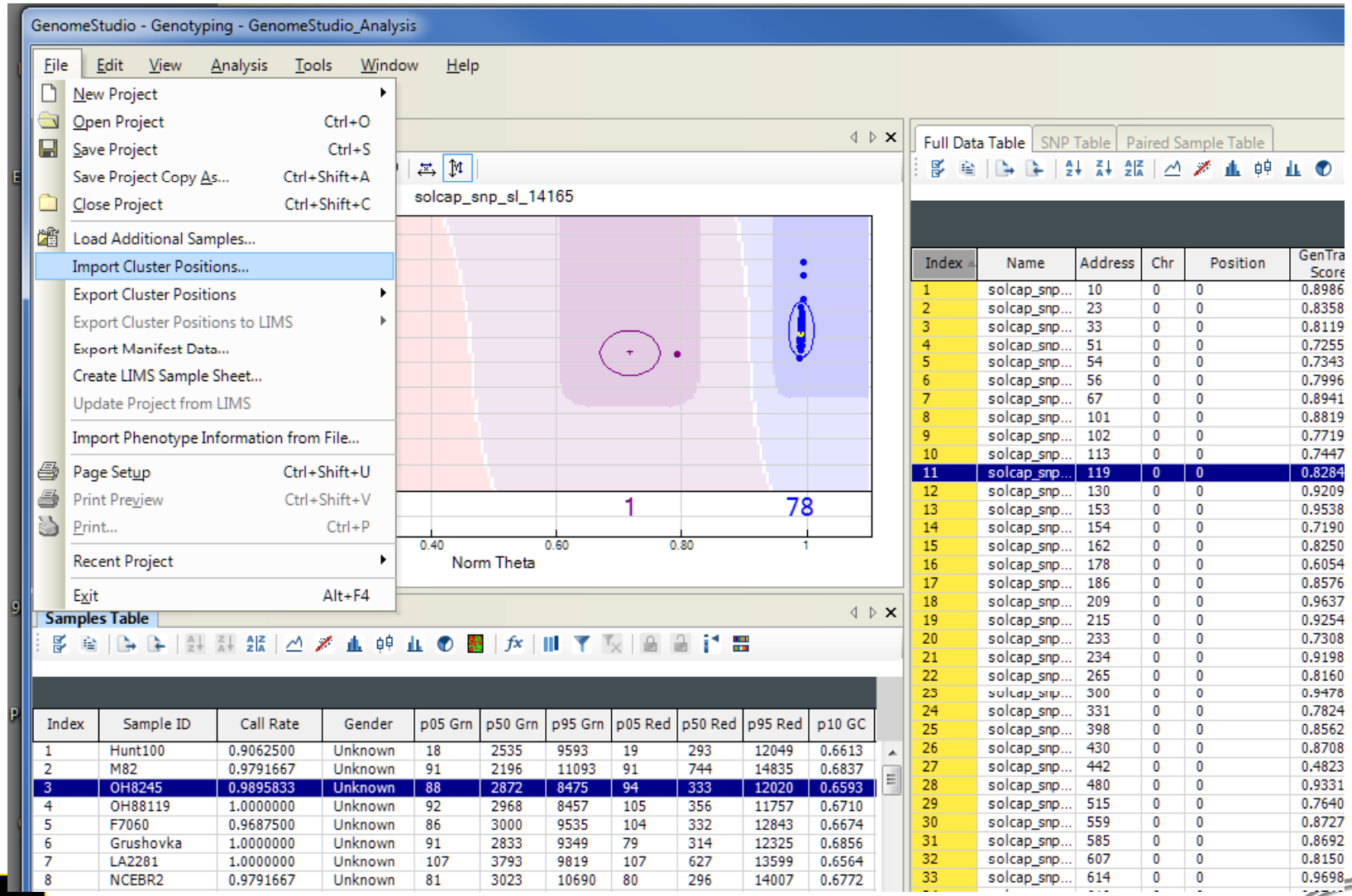


- Display
- Call rate-graph
- Sort
- Exclude
- Configure

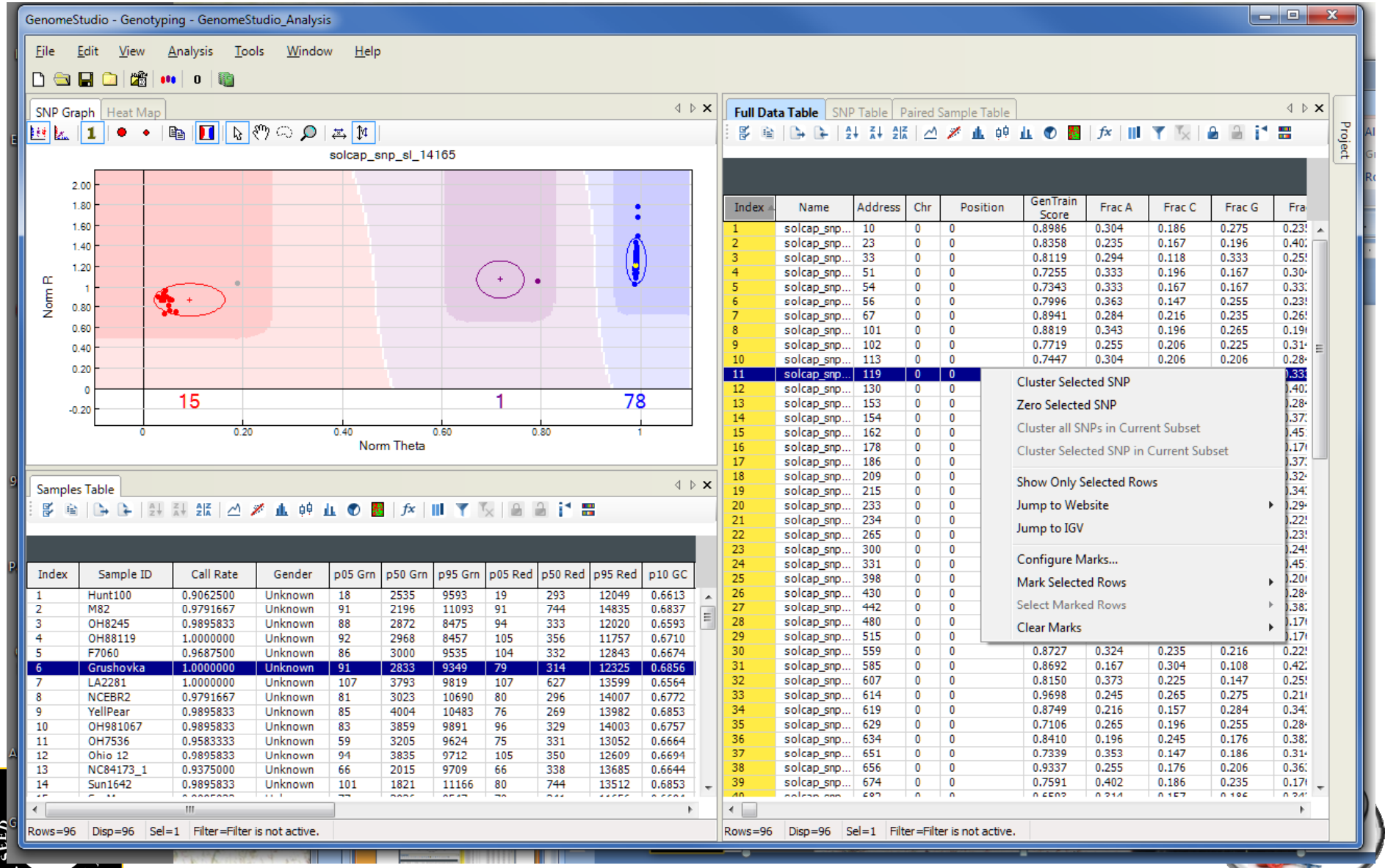
Load additional samples



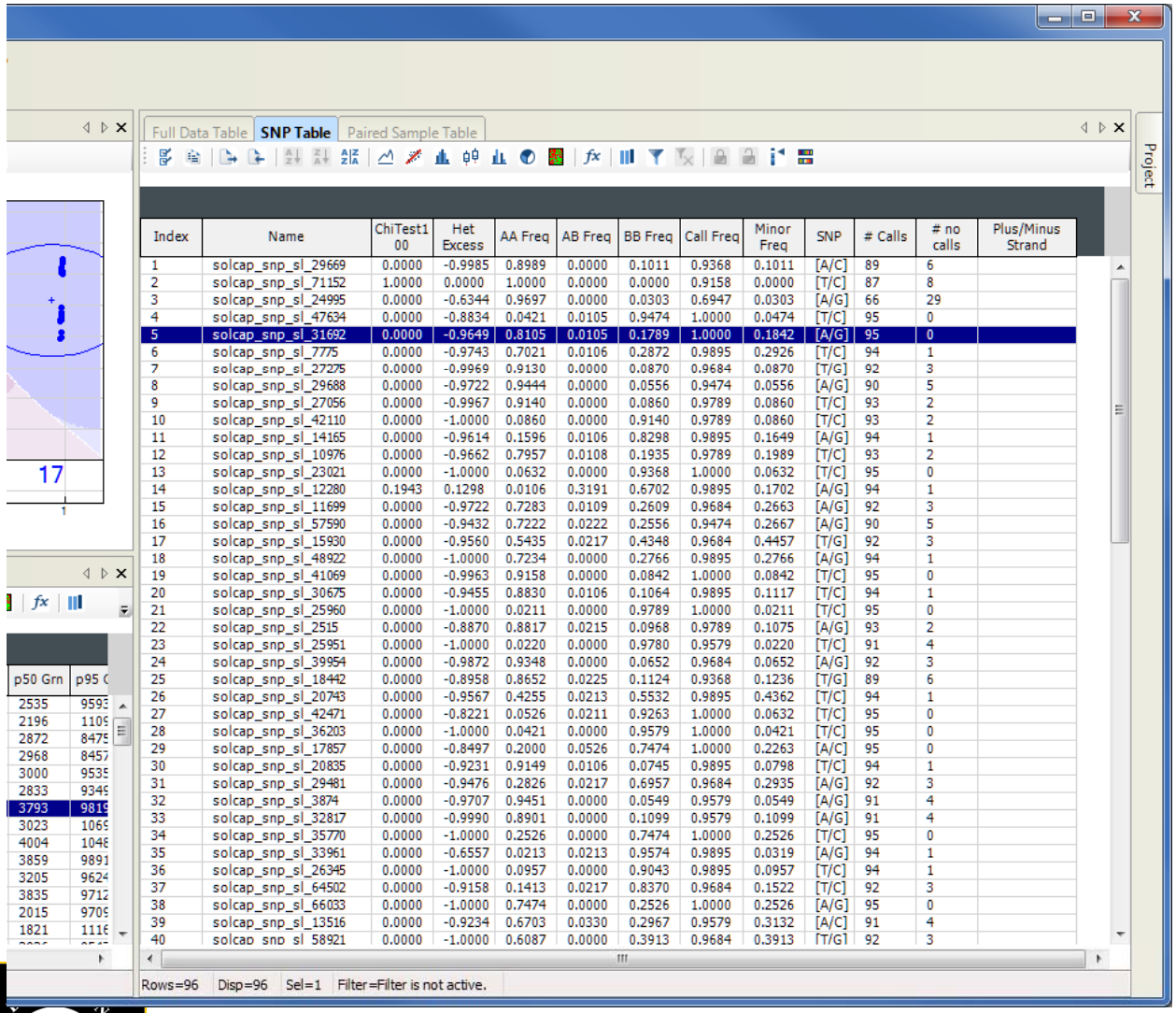
Import clusters



Managing SNPs



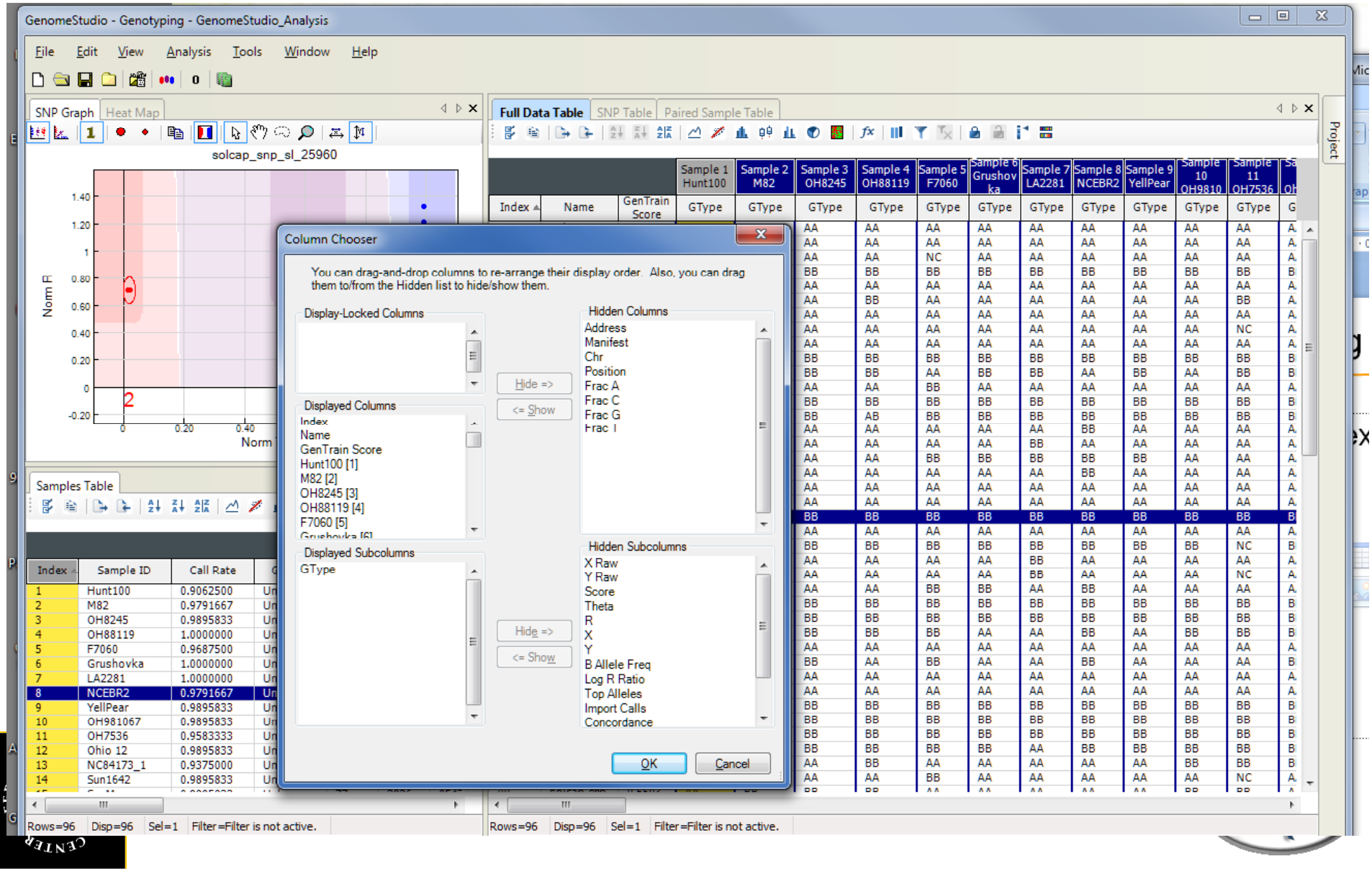
SNP Table



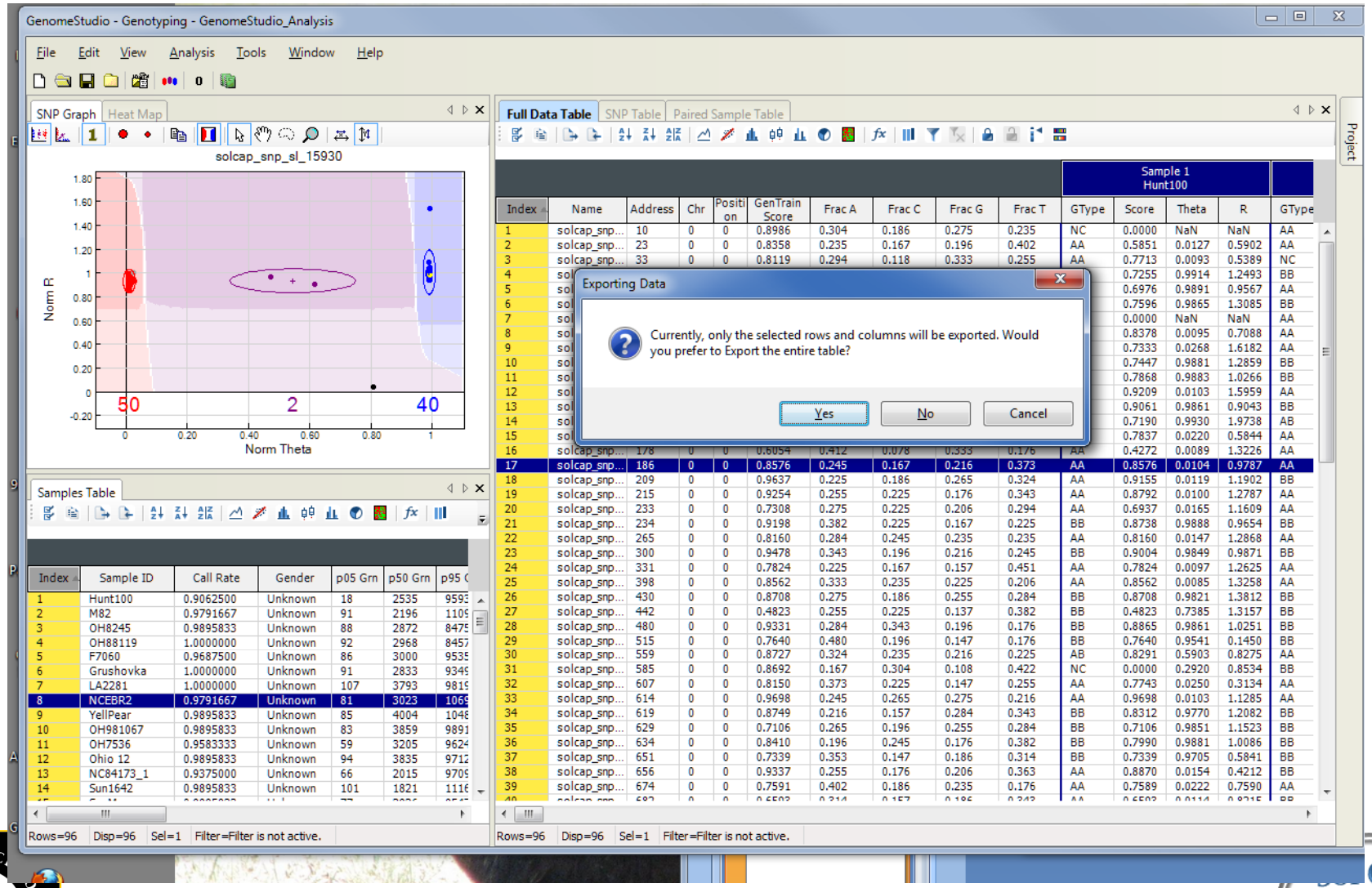
Index	Name	ChiTest100	Het Excess	AA Freq	AB Freq	BB Freq	Call Freq	Minor Freq	SNP	# Calls	# no calls	Plus/Minus Strand
1	solcap_snp_sl_29669	0.0000	-0.9985	0.8989	0.0000	0.1011	0.9368	0.1011	[A/C]	89	6	
2	solcap_snp_sl_71152	1.0000	0.0000	1.0000	0.0000	0.0000	0.9158	0.0000	[T/C]	87	8	
3	solcap_snp_sl_24995	0.0000	-0.6344	0.9697	0.0000	0.0303	0.6947	0.0303	[A/G]	66	29	
4	solcap_snp_sl_47634	0.0000	-0.8834	0.0421	0.0105	0.9474	1.0000	0.0474	[T/C]	95	0	
5	solcap_snp_sl_31692	0.0000	-0.9649	0.8105	0.0105	0.1789	1.0000	0.1842	[A/G]	95	0	
6	solcap_snp_sl_7775	0.0000	-0.9743	0.7021	0.0106	0.2872	0.9895	0.2926	[T/C]	94	1	
7	solcap_snp_sl_27275	0.0000	-0.9969	0.9130	0.0000	0.0870	0.9684	0.0870	[T/G]	92	3	
8	solcap_snp_sl_29688	0.0000	-0.9722	0.9444	0.0000	0.0556	0.9474	0.0556	[A/G]	90	5	
9	solcap_snp_sl_27056	0.0000	-0.9967	0.9140	0.0000	0.0860	0.9789	0.0860	[T/C]	93	2	
10	solcap_snp_sl_42110	0.0000	-1.0000	0.0860	0.0000	0.9140	0.9789	0.0860	[T/C]	93	2	
11	solcap_snp_sl_14165	0.0000	-0.9614	0.1596	0.0106	0.8298	0.9895	0.1649	[A/G]	94	1	
12	solcap_snp_sl_10976	0.0000	-0.9662	0.7957	0.0108	0.1935	0.9789	0.1989	[T/C]	93	2	
13	solcap_snp_sl_23021	0.0000	-1.0000	0.0632	0.0000	0.9368	1.0000	0.0632	[T/C]	95	0	
14	solcap_snp_sl_12280	0.1943	0.1298	0.0106	0.3191	0.6702	0.9895	0.1702	[A/G]	94	1	
15	solcap_snp_sl_11699	0.0000	-0.9722	0.7283	0.0109	0.2609	0.9684	0.2663	[A/G]	92	3	
16	solcap_snp_sl_57590	0.0000	-0.9432	0.7222	0.0222	0.2556	0.9474	0.2667	[A/G]	90	5	
17	solcap_snp_sl_15930	0.0000	-0.9560	0.5435	0.0217	0.4348	0.9684	0.4457	[T/G]	92	3	
18	solcap_snp_sl_48922	0.0000	-1.0000	0.7234	0.0000	0.2766	0.9895	0.2766	[A/G]	94	1	
19	solcap_snp_sl_41069	0.0000	-0.9963	0.9158	0.0000	0.0842	1.0000	0.0842	[T/C]	95	0	
20	solcap_snp_sl_30675	0.0000	-0.9455	0.8830	0.0106	0.1064	0.9895	0.1117	[T/C]	94	1	
21	solcap_snp_sl_25960	0.0000	-1.0000	0.0211	0.0000	0.9789	1.0000	0.0211	[T/C]	95	0	
22	solcap_snp_sl_2515	0.0000	-0.8870	0.8817	0.0215	0.0968	0.9789	0.1075	[A/G]	93	2	
23	solcap_snp_sl_25951	0.0000	-1.0000	0.0220	0.0000	0.9780	0.9579	0.0220	[T/C]	91	4	
24	solcap_snp_sl_39954	0.0000	-0.9872	0.9348	0.0000	0.0652	0.9684	0.0652	[A/G]	92	3	
25	solcap_snp_sl_18442	0.0000	-0.8958	0.8652	0.0225	0.1124	0.9368	0.1236	[T/G]	89	6	
26	solcap_snp_sl_20743	0.0000	-0.9567	0.4255	0.0213	0.5532	0.9895	0.4362	[T/C]	94	1	
27	solcap_snp_sl_42471	0.0000	-0.8221	0.0526	0.0211	0.9263	1.0000	0.0632	[T/C]	95	0	
28	solcap_snp_sl_36203	0.0000	-1.0000	0.0421	0.0000	0.9579	1.0000	0.0421	[T/C]	95	0	
29	solcap_snp_sl_17857	0.0000	-0.8497	0.2000	0.0526	0.7474	1.0000	0.2263	[A/C]	95	0	
30	solcap_snp_sl_20835	0.0000	-0.9231	0.9149	0.0106	0.0745	0.9895	0.0798	[T/C]	94	1	
31	solcap_snp_sl_29481	0.0000	-0.9476	0.2826	0.0217	0.6957	0.9684	0.2935	[A/G]	92	3	
32	solcap_snp_sl_3874	0.0000	-0.9707	0.9451	0.0000	0.0549	0.9579	0.0549	[A/G]	91	4	
33	solcap_snp_sl_32817	0.0000	-0.9990	0.8901	0.0000	0.1099	0.9579	0.1099	[A/G]	91	4	
34	solcap_snp_sl_35770	0.0000	-1.0000	0.2526	0.0000	0.7474	1.0000	0.2526	[T/C]	95	0	
35	solcap_snp_sl_33961	0.0000	-0.6557	0.0213	0.0213	0.9574	0.9895	0.0319	[A/G]	94	1	
36	solcap_snp_sl_26345	0.0000	-1.0000	0.0957	0.0000	0.9043	0.9895	0.0957	[T/C]	94	1	
37	solcap_snp_sl_64502	0.0000	-0.9158	0.1413	0.0217	0.8370	0.9684	0.1522	[T/C]	92	3	
38	solcap_snp_sl_66033	0.0000	-1.0000	0.7474	0.0000	0.2526	1.0000	0.2526	[A/G]	95	0	
39	solcap_snp_sl_13516	0.0000	-0.9234	0.6703	0.0330	0.2967	0.9579	0.3132	[A/C]	91	4	
40	solcap_snp_sl_58921	0.0000	-1.0000	0.6087	0.0000	0.3913	0.9684	0.3913	[T/G]	92	3	

Display
Stats
Select columns

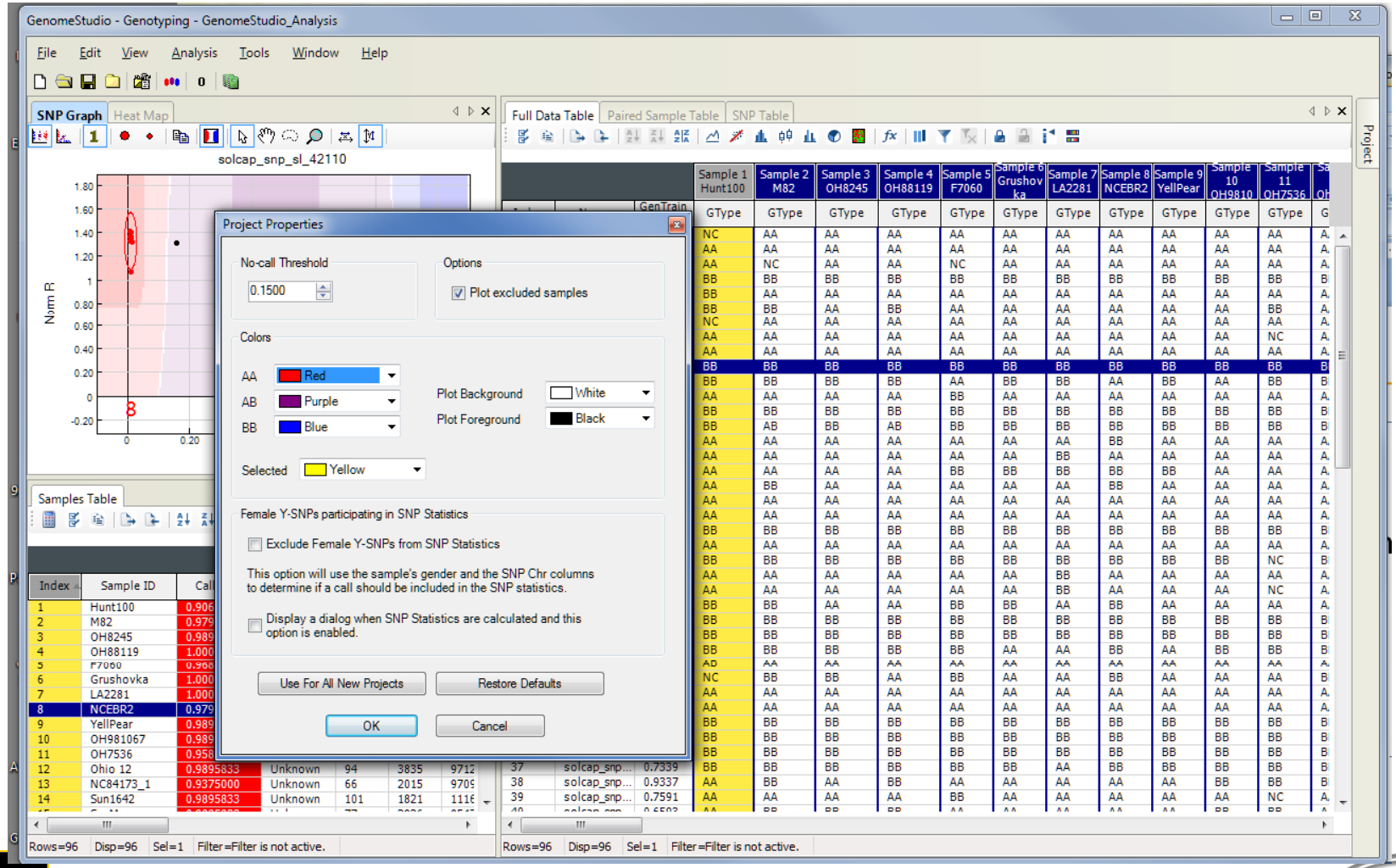
Selecting data to export



Exporting Genotypes



Options


GenomeStudio - Genotyping - GenomeStudio_Analysis
File Edit View Analysis Tools Window Help
SNP Graph Heat Map
solcap_snp_sl_42110
Norm R
Full Data Table Paired Sample Table SNP Table
Project
Project Properties
No-call Threshold
0.1500
Options
Plot excluded samples
Colors
AA Red
AB Purple
BB Blue
Plot Background White
Plot Foreground Black
Selected Yellow
Female Y-SNPs participating in SNP Statistics
Exclude Female Y-SNPs from SNP Statistics
This option will use the sample's gender and the SNP Chr columns to determine if a call should be included in the SNP statistics.
Display a dialog when SNP Statistics are calculated and this option is enabled.
Use For All New Projects
Restore Defaults
OK
Cancel
Samples Table
Index Sample ID
1 Hunt100
2 M82
3 OH8245
4 OH88119
6 Grushovka
7 LA2281
8 NCEBR2
9 YellPear
10 OH981067
11 OH7536
12 Ohio 12
13 NC84173_1
14 Sun1642
Sample 1 Hunt100
Sample 2 M82
Sample 3 OH8245
Sample 4 OH88119
Sample 5 F7060
Sample 6 Grushovka
Sample 7 LA2281
Sample 8 NCEBR2
Sample 9 YellPear
Sample 10 OH9810
Sample 11 OH7536
GType
Rows=96 Disp=96 Sel=1 Filter=Filter is not active.

QC

- Sample
- SNPs
- Allele Frequency