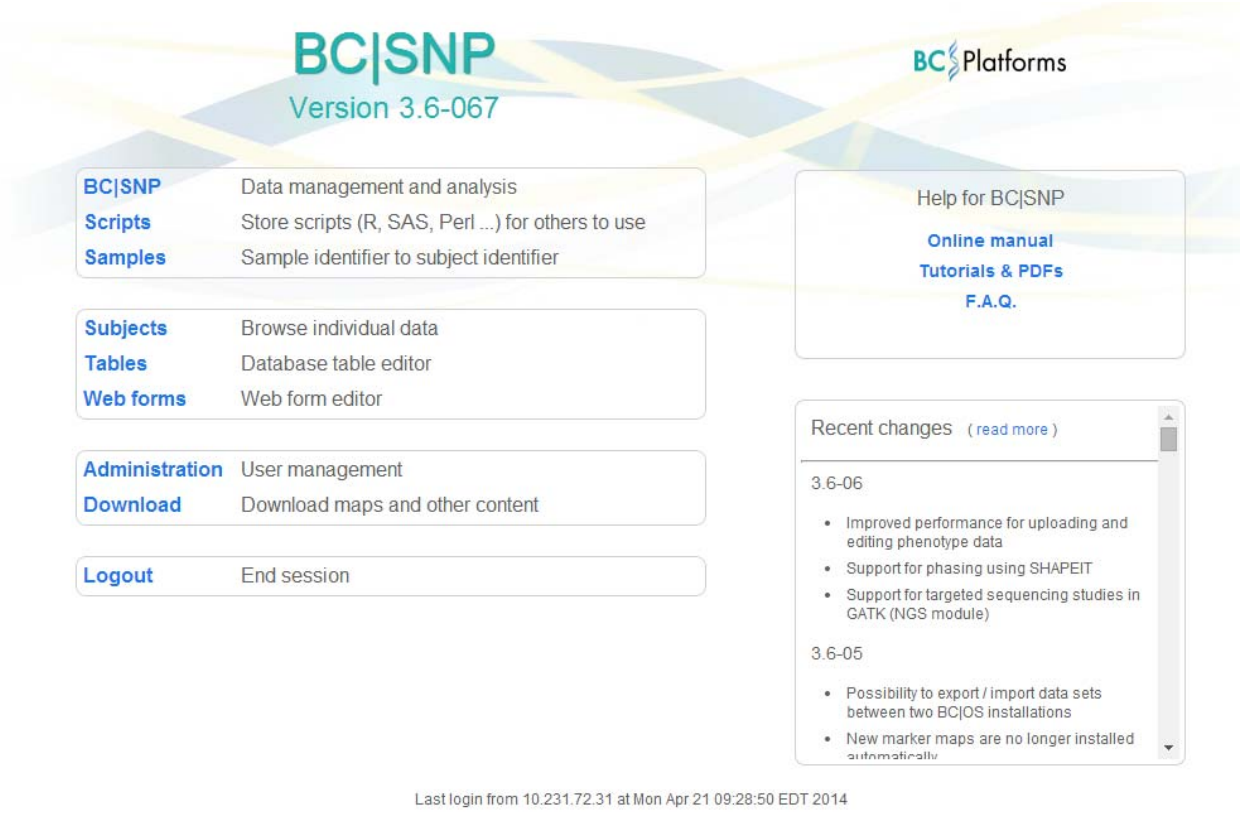
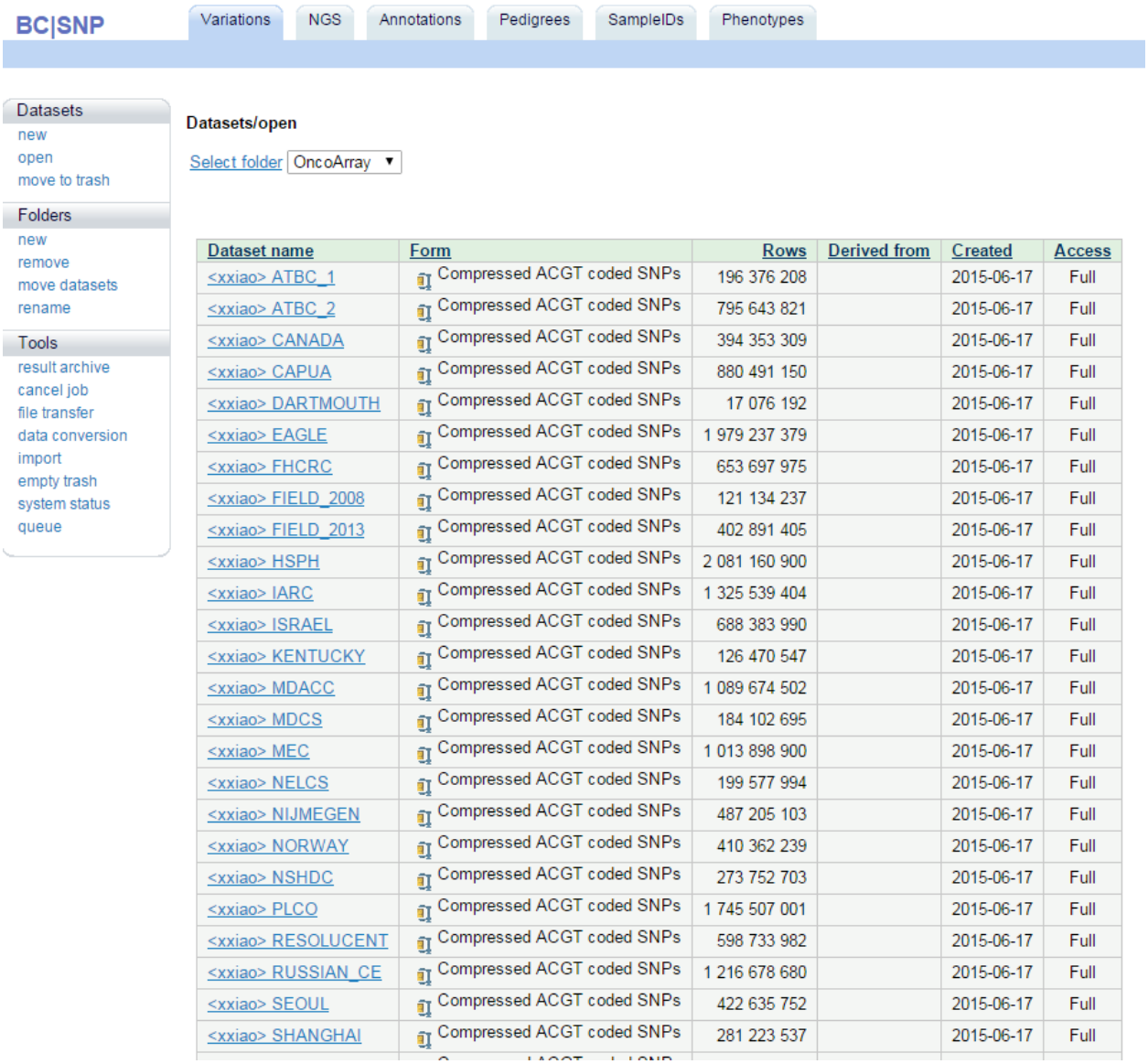
Descriptions of retrieving your requesting markers in OncoArray data

1) Suppose that you already have account in https://bc1.dartmouth.edu/bcos/

2) Go to BC|SNP section



3) Go to Variations section, and select folder “Oncoarray”

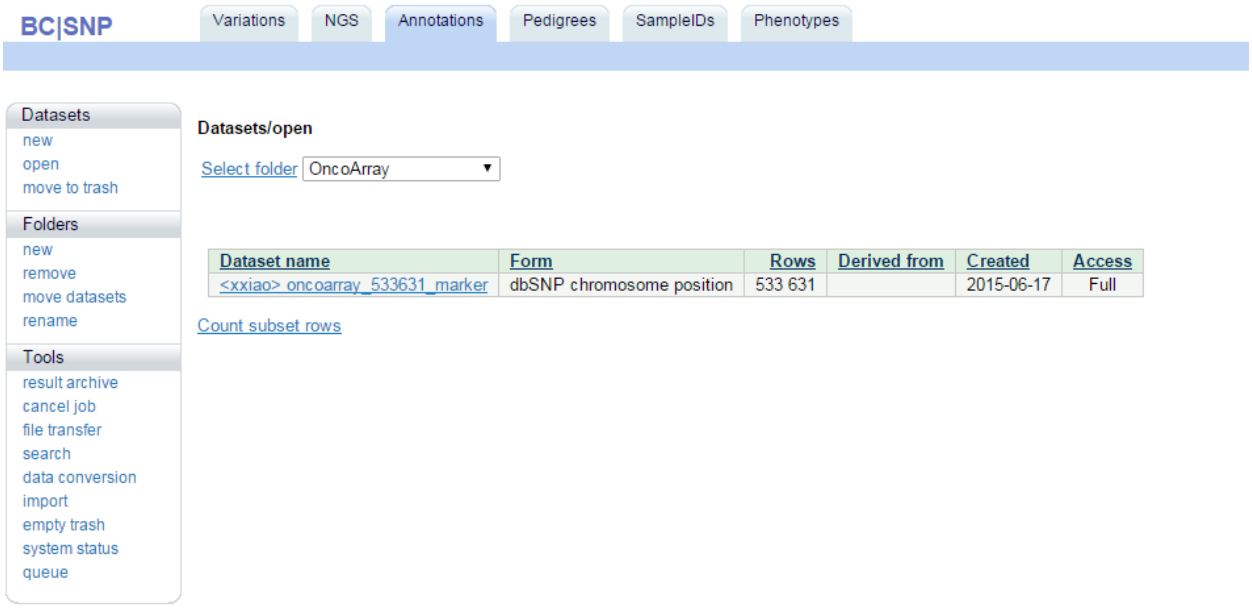


Dataset name	Form	Rows	Derived from	Created	Access
<xxiao> ATBC_1	Compressed ACGT coded SNPs	196 376 208		2015-06-17	Full
<xxiao> ATBC_2	Compressed ACGT coded SNPs	795 643 821		2015-06-17	Full
<xxiao> CANADA	Compressed ACGT coded SNPs	394 353 309		2015-06-17	Full
<xxiao> CAPUA	Compressed ACGT coded SNPs	880 491 150		2015-06-17	Full
<xxiao> DARTMOUTH	Compressed ACGT coded SNPs	17 076 192		2015-06-17	Full
<xxiao> EAGLE	Compressed ACGT coded SNPs	1 979 237 379		2015-06-17	Full
<xxiao> FHCRC	Compressed ACGT coded SNPs	653 697 975		2015-06-17	Full
<xxiao> FIELD_2008	Compressed ACGT coded SNPs	121 134 237		2015-06-17	Full
<xxiao> FIELD_2013	Compressed ACGT coded SNPs	402 891 405		2015-06-17	Full
<xxiao> HSPH	Compressed ACGT coded SNPs	2 081 160 900		2015-06-17	Full
<xxiao> IARC	Compressed ACGT coded SNPs	1 325 539 404		2015-06-17	Full
<xxiao> ISRAEL	Compressed ACGT coded SNPs	688 383 990		2015-06-17	Full
<xxiao> KENTUCKY	Compressed ACGT coded SNPs	126 470 547		2015-06-17	Full
<xxiao> MDACC	Compressed ACGT coded SNPs	1 089 674 502		2015-06-17	Full
<xxiao> MDCS	Compressed ACGT coded SNPs	184 102 695		2015-06-17	Full
<xxiao> MEC	Compressed ACGT coded SNPs	1 013 898 900		2015-06-17	Full
<xxiao> NELCS	Compressed ACGT coded SNPs	199 577 994		2015-06-17	Full
<xxiao> NIJMEGEN	Compressed ACGT coded SNPs	487 205 103		2015-06-17	Full
<xxiao> NORWAY	Compressed ACGT coded SNPs	410 362 239		2015-06-17	Full
<xxiao> NSHDC	Compressed ACGT coded SNPs	273 752 703		2015-06-17	Full
<xxiao> PLCO	Compressed ACGT coded SNPs	1 745 507 001		2015-06-17	Full
<xxiao> RESOLUCENT	Compressed ACGT coded SNPs	598 733 982		2015-06-17	Full
<xxiao> RUSSIAN_CE	Compressed ACGT coded SNPs	1 216 678 680		2015-06-17	Full
<xxiao> SEOUL	Compressed ACGT coded SNPs	422 635 752		2015-06-17	Full
<xxiao> SHANGHAI	Compressed ACGT coded SNPs	281 223 537		2015-06-17	Full

30 genotype files were generated in this folder, each file is for one SITE.

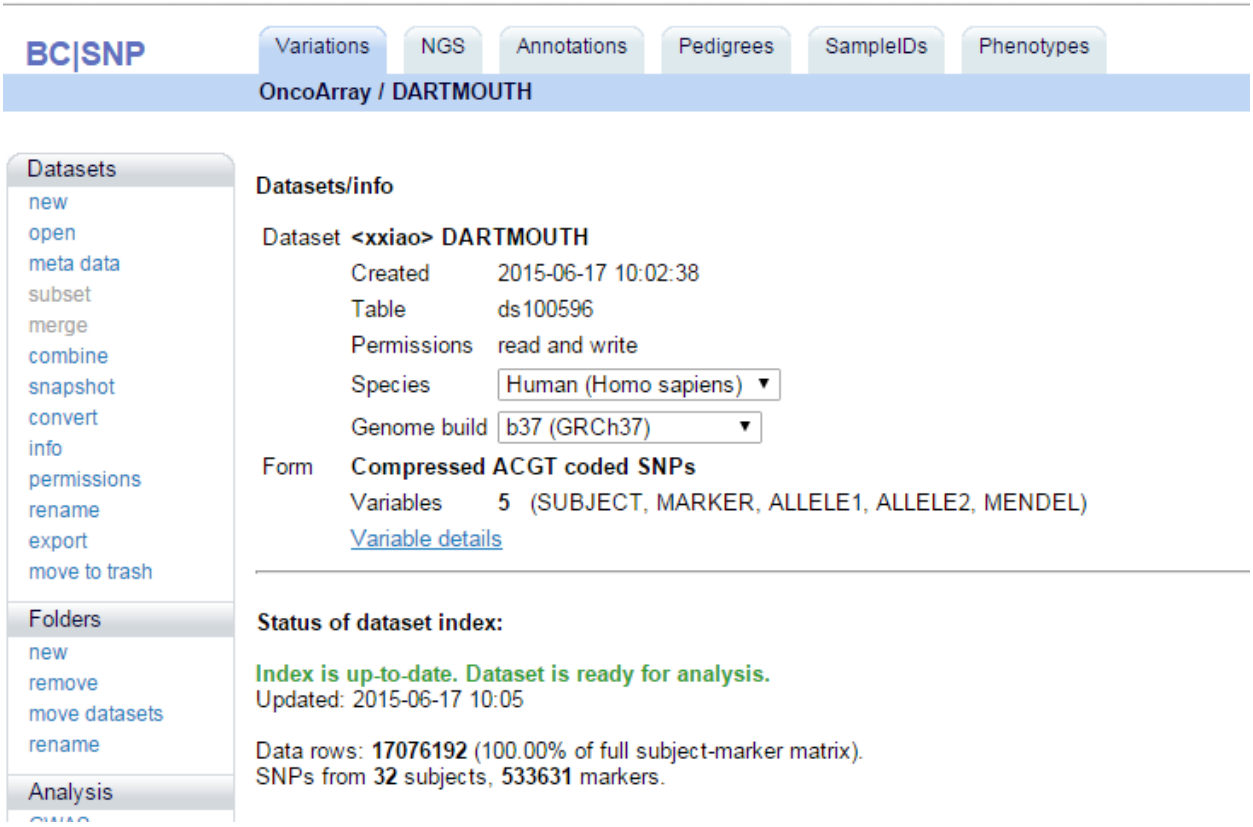
For security reason, I will set permission for each file based on request, therefore, users of every PI/ study SITE can only see one file existing in this folder.

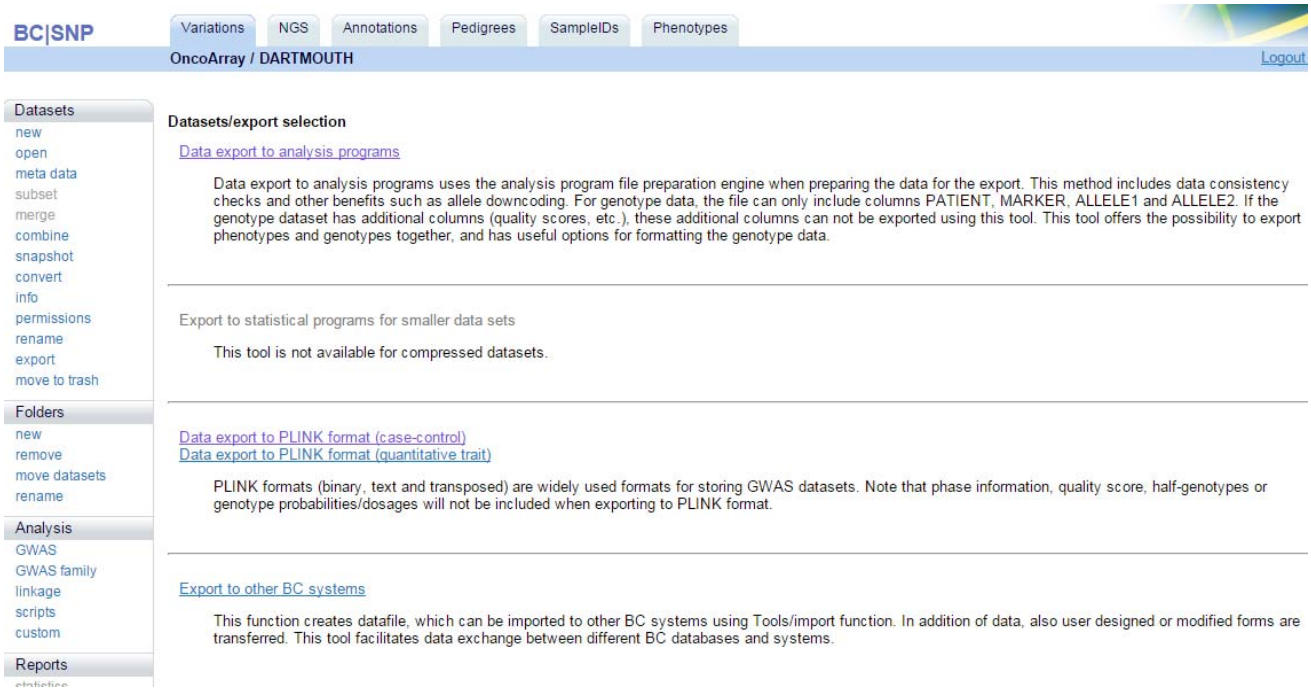
4) Go to Annotation section, and enter OncoArray folder, you will see a file:



This file: oncoarray_533631_marker has information for marker name, chromosome, and position. Currently, I do not have any more information, such as alleles and associated genes, I will update information whenever it is available.

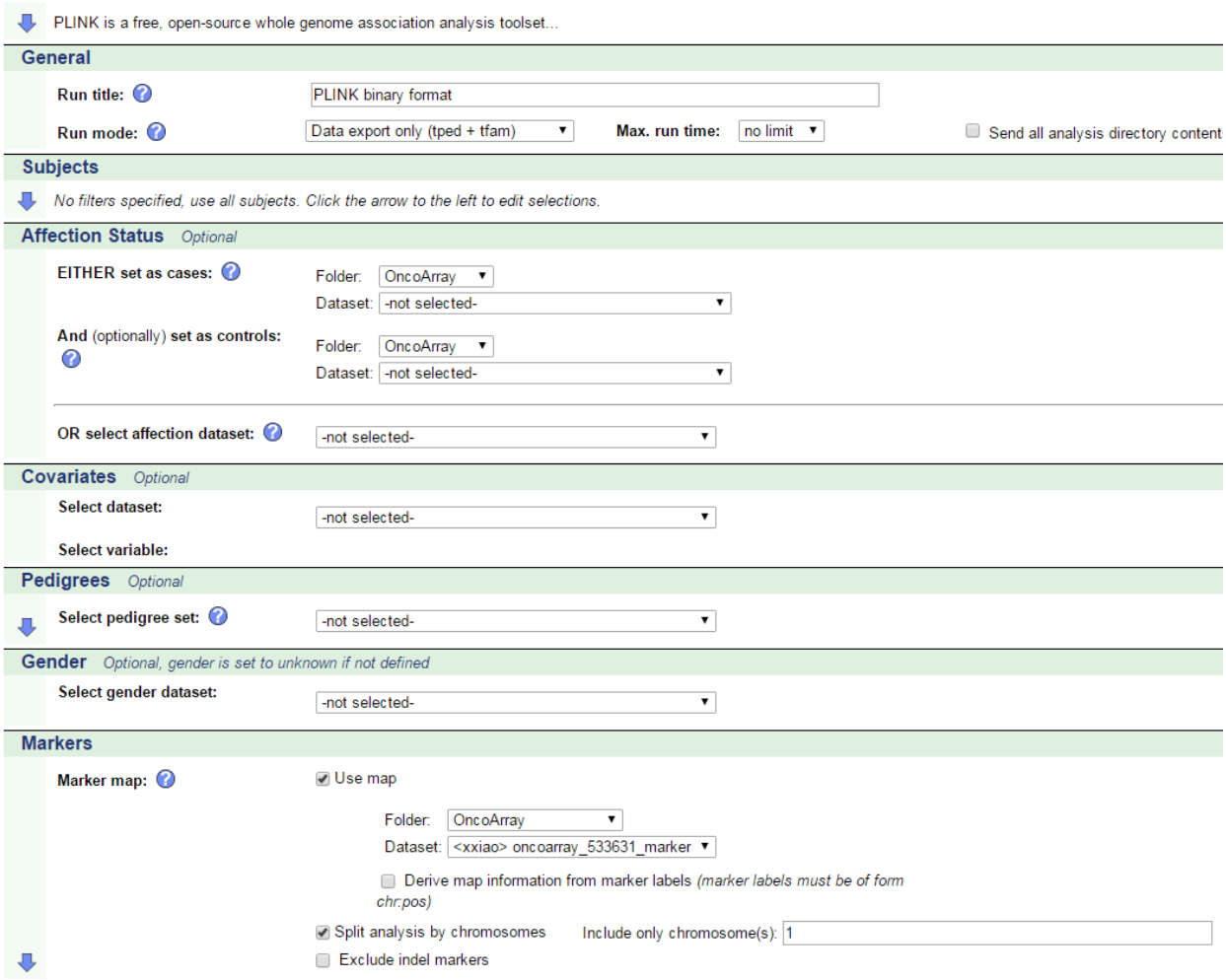
5) Now, I give an example to show how to export genotypes of plink format.
Go to variation section again, and select DARTMOUTH (an example, you will see genotype of your SITE name), select export



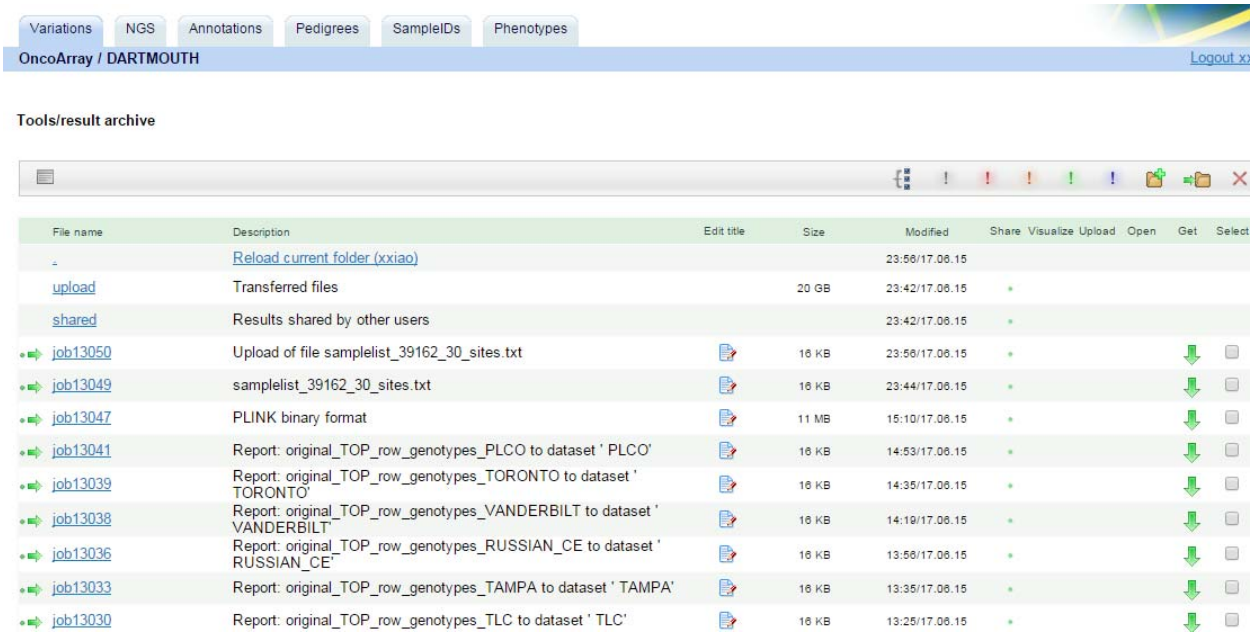


You will find two plink choices of exporting, now select "Data export to PLINK format (case-control)", note that there is no different between two PLINK choices, since there is no affection status in this example.

Now choose the format of PLINK format you want, for example, tped+tfam. Do not do anything on Affection, Covariates and Pedigrees parts. Then, click “Use map”, choose markers generated in folder OncoArray. You can split analysis by chromosomes or export all markers.



6) Run this process and you will find your results in “result archive”, for example: job13047, PLINK binary format, and then you can download datasets.



7) For sample information of each SITE, go to phenotypes section in folder OncoArray, you will find a file: sample-information-site-JUNE-18-2015**.** Open this file, you will find that there are four variables:

ORIGINAL_ID	CIDR	SITE	SAMPLEID

Original ID stands for in total 53600 samples (we received), the unique ID assigned for each sample, you need not to use it, database needs a primary key for this table.
CIDR stands for local ID
SITE stands for site name
SampleID stands for unique ID for each SITE, the order is exactly matched to genotype’s order. For example, if you generate tped (transposed plink format, row for maker, column for samples), the order of samples is matched to sampleid.

Note that we will frequently update this file with adding extra columns following four variables described above. In each file’s description box, there will be detailed information.