CNVRuler

User Manual

V1.2

CNVRuler software is freely available with associated files and user manual in our website: http:// www.ircgp.com/CNVRuler/index.html

Contact to developer: Yeun-Jun Chung (<u>yejun@catholic.ac.kr</u>) and Ji-Hong Kim (<u>lomolith@gmail.com</u>) CNV-Ruler is designed for CNVR based association analysis with user-friendly graphic interface. All forms of major CNV call outputs from different segmentation tools such as Genotyping Console, Genome Studio, Genomic Workbench, BirdSuite, PennCNV and Nexus can be processed without additional converting steps. CNV-Ruler supports defining three different types of CNV regions (CNVRs) and four statistical methods for CNVR based association analysis. Users can analyze CNVR-phenotype associations with their preferable segmentation tools and can test various CNVR definitions and statistical methods suitable for their own study design.

1. Prerequisites

CNV-Ruler needs Java Run-time Environment of SUN Microsystems or equivalent (JRE 1.6.0 or higher). For all statistical analyses, R is used as a calculation core.

• JRE

If your system does not have Java Virtual Machine (JVM), you can download it from Oracle's Java home page (<u>http://www.oracle.com/technetwork/java/javase/downloads/index.html</u>).

For checking whether JVM is properly installed, type *java* –*version* on a terminal prompt. The version of your JVM will appear if it is correctly installed. Windows users can type the command on the command line window from Start button.

• R

CNV-Ruler needs R for its calculation process. You can download it from its project home page. <u>http://www.r-project.org/</u>

After selecting download mirror site and OS platform, you can download the distribution binaries. If your system doesn't have the R package on it, CNV-Ruler will prompt it and try to open the R download site.

NOTICE: If CNV-Ruler keeps warning that there is no R package after installation, you may add *PATH* variable manually. Test by typing *R* –-*version* on your terminal. Usually, Linux users do not need to change it.

2. Installation

The CNV-Ruler package consists of two executable files – CNVRuler.bin and CNVRuler.exe – and one text file – readme.txt – which is the change log of version history. Simply uncompress it and select executable by type of your OS – CNVRuler.bin for Linux, CNVRuler.exe for Windows.

Step 1) Download compressed zip file. Right click and select Extract All.



Step 2) Extract anywhere you want to put them by following instruction

Extraction Wizard	
Select a Destination Files inside the ZIP a choose.	archive will be extracted to the location you
	Select a folder to extract files to. Files will be extracted to this <u>directory:</u> <u>C:\download\CNVFuler</u> <u>Browse</u> <u>Password</u> Extracting
	< <u>₿</u> ack <u>N</u> ext > Cancel

Step3) Double click to execute CNVRuler.exe



NOTICE: Do not put programs nor data on the folder which has a name with 2-byte character (ex. Asian characters). It makes R occurs inside error (cannot make temporary directory)

After finishing installation, user interface will be appeared.

Clinical Information	Clinical infomation file					
	Sample ID 🗨 Gender 🗸					
	Phenotype Age 🗸					
	Covariate					
CNV	CNV call file					
	Remove smaller than 0 LOD (BirdSuite) 2					
	Segment mean cut-off (TCGA, NimbleScan) 0.3					
CNV Region	Method CNVR Recurrence 0.1					
	☐ Gain/Loss separated regions ✔ Include CNVs across low frequency areas					
Association Test	Method Logistic Regression					
	Use PCA as covariates for Population Stratification No 💌					
	Minor Allele Threshold 0.05					
	Separated p-values for Gain/Loss					

3. Data analysis

A) Data uploading

Clinical Information	Clinical infomation file]
	Sample ID Gender 📃 🗸]
	Phenotype 📃 🗸 Age]
	Covariate]
CNV	CNV call file	
	Remove smaller than 0 LOD (BirdSuite) Segment mean cut-off (TCGA, NimbleScan) 0.3	
	Segment mean curon (rCOA, NimpleScan) 0.3	

For CNV-Ruler analysis, two types of information (Clinical and CNV data) must be prepared.

• Step 1. Uploading clinical data

In the clinical information (CI) file, 4 items (sample ID, age, sex and phenotype) are to be included as separate columns in the CI txt file (see the example below). If age or sex information is not available, the users can do the association analysis with just sample ID and phenotype data. Phenotype means the dependent variable for regression analysis. After selecting the CI file for the analysis, you must choose the sample ID and phenotype columns in the user interface with other covariates. The sample ID should be matched to the name of the samples in the CNV data file. Phenotype status must have binary values 0 and 1 for logistic regression. For sex column, users can input values as a string 'male' or 'female'; 'm' or 'f'; 'man' or 'woman'; '1' or '0'; '1' or '2' and it is not case sensitive. In addition to the four basic CIs, other variables for logistic regression analysis can be added in your CI (see an example below). If you have more CIs than the 4 CI columns, click the *Covariates* button, then 'Covariates' pop up window will appear. You can select the extra variables as many as you want. Only the selected extra covariates will be included for the association analysis.

Example of clinical information file and data loaded screen:

Sample ID	Phenotype	Age	Gender	Smoking	Atopy
A016	3.06	53	0	1	0
A028	1.862	31	0	0	1
A042	2.009	45	1	0	0
A044	3.313	54	1	0	1
A061	1.681	37	0	1	1
A063	2.435	16	1	0	1
A065	3.035	54	1	1	1
A077	2.121	44	1	1	0
A084	1.072	65	1	0	0
A085	3.301	26	0	1	1
A121	3.313	60	0	0	0
A130	2.382	69	0	0	1
A142	0.412	49	0	0	1
A150	2.74	47	0	0	1
A161	1.993	75	0	0	0
A162	2.072	51	1	0	0

1) Analysis with sample ID and phenotype

CNV-Ruler ver 1.2.1beta					
Clinical Information	/workspace/pap	er/ACAT/120108/example_ci_manual.txt			
	Sample ID	Sample ID 🔻 Gender -	•		
	Phenotype	Phenotype 🔻 Age -	-		
	Covariate				

2) Analysis with four main CIs

CNV-Ruler ver 1.2.1beta					
Clinical Information	/workspace/pa	per/ACAT/120108/example_ci_manual.txt			
	Sample ID	Sample ID 🔻 Gender Gender	-		
	Phenotype	Phenotype 🔻 Age Age			
	Covariate				

3) Analysis with additional covariates

2	CNV-Ruler ver 1.2.1beta					٠	-	×
							_	٦
	Clinical Information	/workspace/pape	r/ACAT/120108/ex	ample_ci	_manual.txt			
		Sample ID	Sample ID 🔻	Gender	Gender	•		
		Phenotype	Phenotype 💌	Age	Age	•		
		Covariate						



₫	CNV-Ruler ver 1.2.1beta				
	Clinical Information	/workspace/pap	er/ACAT/120108/example_ci_manual.txt		
		Sample ID	Sample ID 🔻 Gender Gender	-	
		Phenotype	Phenotype 💌 Age Age	-	
		Covariate	Smoking,Atopy		

• Step 2. Uploading CNV data

Choose your CNV call output file for analyzing the CNVR based association. CNV-Ruler can read 10 types of CNV call outputs (see Table below) and a custom CNV call.

Format	Version Tested	Ref.
PennCNV	2011Jun16	Wang <i>et al.</i> , 2007
Nexus	5.1	www.biodiscovery.com
Genomic Workbench	6.5	www.agilent.com
CGHscape	1.5	Jeong <i>et al.</i> , 2008
TCGA files	Jun.2011	cancergenome.nih.gov
NimbleScan	2.6	www.nimblegen.com
Genome Studio	2011.1	www.illumina.com
QuantiSNP	2.0	Colella <i>et al.</i> , 2007
BirdSuite	1.5.5	Korn <i>et al.,</i> 2008
Genotying Console	4.1	www.affymetrix.com

If you want to use your own CNV list file, you must prepare a simple tabdelimited text file containing 5 columns as the example below. The names and order of column headings should be Chr, Start, End, Event and Sample_ID respectively.

Example of user own CNV data file:

Chr	Start	End	Event	Sample_ID
1	10430	10592	Loss	Syndrome_TypeA_01
1	12410	12900	Loss	Syndrome_TypeA_01
2	400	8210	Gain	Syndrome_TypeA_01
1	2430	2592	Loss	Syndrome_TypeA_02

• Filtering options

There are two filtering options in the CNV data uploading section.

CNV	CNV call file	
	Remove CNVs smaller size than	0
	Segment mean cut-off (TCGA, NimbleScan)	0.3

1) CNV size filter

Users can set their own threshold for minimum size to define the CNVs (unit: bp). A CNV which is smaller than the threshold will be excluded.

2) Mean signal intensity of the segment filter

This option will be only used when the input file is TCGA or NimbleScan data. TCGA and NimbleScan data do not have 'gain or loss' information but have mean value of segmentation. Therefore, a cut-off criterion is required to define the copy number 'gain' or 'loss' status. Default value is ± 0.3 , which means that

a CNV segment with mean value < -0.3 will be assigned as 'loss' and > 0.3 as 'gain CNV'. Users can set their own cut-off filter.

B) Defining CNVR

CNV Region	Method	CNVR -	Recurrence	0.1
		eparated regions s across low freque	ncy are	

CNV-Ruler supports three different definitions of CNV Regions (CNVRs): CNVR, RO, and Fragment. They produce similar but slightly different boundaries and each of them has its own advantages and limitations as described in the main text.

Method

Select one of the following 3 definitions of CNVRs.

- 1) CNVR (CNV region)
- 2) RO (Reciprocal Overlap)
- 3) Fragment

1) CNVR (CNV region)

CNVR is defined by merging of overlapping CNVs.

i. CNVR trimming threshold

Definition of CNVR is simple and straightforward, but this definition can overestimate the size and frequency of CNVR due to the potential false calls, which are usually rare and long-sized. CNV-Ruler can trim these extreme ones during merging process by CNV frequencies. In case of the CNVR method, users can trim the sparse area by using the regional density (recurrence) threshold. This option checks the regional density of participating CNVs base-wise and trimming the sparse area not satisfying the given density threshold (default: 0.1). This option does not affect RO nor Fragment method.

ii. Additional options for building CNVRs: Gain/Loss separated region

Using this option, the CNVR can be created with same types of CNVs, gain or loss type, within the considering area. If you select this option, CNVR outputs will be copy number gain CNVR or loss CNVR. If you don't select this option, all overlapped CNVs will be used for building CNVR regardless of their type.









2) Reciprocal overlap (RO)

CNV regions are determined by reciprocal overlap (RO) measure. First, CNVs which overlap at least one-base are grouped as initial CNV clusters. Within each cluster, RO is calculated for each CNV to the others. The pair of which RO is highest (default minimum threshold is >50%) will be merged and formed a CNV element (in orange). This process is repeated until every pair has RO of 50% or lower. The detailed process is illustrated in the following figure.



In st ep 1, since the pair consisting of CNV2 and CNV3 has the highest RO, these two CNVs are merged into a CNV element called CNV2-3. Similarly, in step 2, CNV2-3 and CNV4 are merged into CNV2-3-4. As the RO values of all the remaining pairs do not pass the RO threshold, three CNV regions are defined (in black).

The RO method can reduce the extent of size overestimation of CNV regions caused by CNVR method. However, compared with the other two methods, the RO method may increase the possibility of false negative results. For example, suppose that one locus embedded within CNVR3 (in the final step, figure above) is truly associated with the trait of interest. In both CNVR and fragment methods, the frequency of this locus is 2, but in the RO method, the frequency is 1, which may cause the true association to be statistically missed.

3) Fragment

The fragment method dissects overlapping regions which have different frequencies of CNVs from the neighboring regions into smaller, separate fragments.



Therefore, this method has the least probability of size overestimation compared with the other two methods. Also, the potential of false negative results may be lower than the RO method. By dissecting overlapping CNVs into smaller fragments, this method could generate a large number of smaller CNV elements which may increase the possibility of false positive associations as well as the calculation burden.

NOTICE: CNV region information is stored in *tmp* directory and can be used later. You can remove it safely by deleting *tmp* directory if error is occurred.

• Examples of CNVR outputs from the same CNV data

The list below contains the CNVs in chromosome 1 identified from the 7 samples from Affymetrix Genotype Console. You can download the sample CNV file from our web site (www.ircgp.com/CNVRuler/index.html)

Chr	Start	End	Туре		
1	61723	228694	Gain		
1	61723	229063	Loss		
1	61723	229607	Gain		
1	61723	356530	Gain		
1	85924	229607	Gain		
1	235658	564621	Loss		
1	740857	1030307	Gain		
1	16830808	16935995	Gain		
1	16968362	17298496	Gain		
1	17029580	17245518	Gain		
1	17035208	17177033	Gain		
1	17036531	17182425	Gain		
1	17037085	17182425	Gain		
1	17045446	17190850	Gain		
1	63704937	63810371	Gain		
1	82461630	82644795	Gain		
1	104130168	104307231	Gain		
1	121343784	121482967	Gain		
1	121343784	121482967	Gain		
1	121343784	121482967	Gain		
1	144036737	144849544	Gain		
1	145206610	145398179	Gain		
1	148530424	148662751	Gain		
1	148530424	148953984	Gain		
1	148947698	149051903	Loss		
1	149086173	149202866	Loss		
1	149086551	149190306	Loss		
1	166574788	166966828	Gain		
1	182454823	182611606	Loss		
1	196706260	196812518	Gain		
1	196706260	196812518	Gain		
1	196711067	196812518	Gain		
1	243163830	243274530	Gain		

Total CNV List

1) CNVR (Recurrence Threshold is 0.1)

Chr	Start	End	Туре
1	61,723	564,621	Mixed
1	740,857	1,030,307	Gain
1	16,830,808	16,935,995	Gain
1	16,968,362	17,298,496	Gain
1	63,704,937	63,810,371	Gain
1	82,461,630	82,644,795	Gain
1	104,130,168	104,307,231	Gain
1	121,343,784	121,482,967	Gain
1	144,036,737	144,849,544	Gain
1	145,206,610	145,398,179	Gain
1	148,530,424	149,051,903	Mixed
1	149,086,173	149,202,866	Loss
1	166,574,788	166,966,828	Gain
1	182,454,823	182,611,606	Loss
1	196,706,260	196,812,518	Gain
1	243,163,830	243,274,530	Gain

Chr	Start	End	Туре
1	61,723	356,530	Mixed
1	235,658	564,621	Loss
1	740,857	1,030,307	Gain
1	16,830,808	16,935,995	Gain
1	16,968,362	17,298,496	Gain
1	63,704,937	63,810,371	Gain
1	82,461,630	82,644,795	Gain
1	104,130,168	104,307,231	Gain
1	121,343,784	121,482,967	Gain
1	144,036,737	144,849,544	Gain
1	145,206,610	145,398,179	Gain
1	148,530,424	148,662,751	Gain
1	148,530,424	148,953,984	Gain
1	148,947,698	149,051,903	Loss
1	149,086,173	149,202,866	Loss
1	166,574,788	166,966,828	Gain
1	182,454,823	182,611,606	Loss
1	196,706,260	196,812,518	Gain
1	243,163,830	243,274,530	Gain

3) Fragment

Chr	Start	End	Туре
1	61,723	85,923	Mixed
1	85,924	228,694	Mixed
1	228,695	229,063	Mixed
1	229,064	229,607	Gain
1	229,608	235,657	Gain
1	235,658	356,530	Mixed
1	356,531	564,621	Loss
1	740,857	1,030,307	Gain
1	16,830,808	16,935,995	Gain
1	16,968,362	17,029,579	Gain
1	17,029,580	17,035,207	Gain
1	17,035,208	17,036,530	Gain
1	17,036,531	17,037,084	Gain
1	17,037,085	17,045,445	Gain
1	17,045,446	17,177,033	Gain
1	17,177,034	17,182,425	Gain
1	17,182,426	17,190,850	Gain
1	17,190,851	17,245,518	Gain
1	17,245,519	17,298,496	Gain
1	63,704,937	63,810,371	Gain
1	82,461,630	82,644,795	Gain
1	104,130,168	104,307,231	Gain
1	121,343,784	121,482,967	Gain
1	144,036,737	144,849,544	Gain
1	145,206,610	145,398,179	Gain
1	148,530,424	148,662,751	Gain
1	148,662,752	148,947,697	Gain
1	148,947,698	148,953,984	Mixed
1	148,953,985	149,051,903	Loss
1	149,086,173	149,086,550	Loss
1	149,086,551	149,190,306	Loss
1	149,190,307	149,202,866	Loss
1	166,574,788	166,966,828	Gain
1	182,454,823	182,611,606	Loss
1	196,706,260	196,711,066	Gain
1	196,711,067	196,812,518	Gain
1	243,163,830	243,274,530	Gain

C) CNVR-Phenotype association analysis

Association Test	Method	Logistic Regression		-
	Use PCA a	s covariates for Population	Stratification	No 🔻
	Minor Allele	e Threshold		0.05
	🗌 Separate	ed p-values for Gain/Loss		

Methods

- A. Logistic regression
- B. Linear regression
- C. Chi-Squared
- D. Fisher's Exact Test

Users select one of the methods above.

Regarding the Chi-Squared test, users can select between 'Chi-Squared test' or 'Chi-squared test with Yates' continuity correction' based on the characteristics of their data.

Logistic Regression	•
Logistic Regression	
Linear Regression	
Chi-Square Test	
Chi-Square Test (Yate's continuity correction)	
Fisher's Exact Test	

Additional options for the association analysis 1) LRT

CNV Ruler supports -2 Log Likelihood Ratio Test (LRT) and calculates p value of chi-squared distribution of LRT. With this value, user can figure out the regression model used for association analysis is significantly better than null model or not. Currently, this option could be applied to logistic regression only.

2) Population Stratification by PCA

Since the association found could be due to the underlying structure of the population and not a disease associated locus, CNV Ruler can use Principal Component Analysis (PCA) to adjust population stratification. CNV Ruler calculates eigen vectors and uses up to 3 principal components as covariates for regression. Currently, this option could be applied to logistic regression only.

3) Separated p-values for Gain/Loss

If a region contains both type of CNV – Gain and Loss - CNV Ruler will calculate p-values for statistical test using only gain-type CNVRs or only loss-type ones with this option.

4) Minor allele frequency

Default value is 0.05 (5%). This means that CNVRs with less than 5% allele frequency will be excluded from the downstream association analysis. Alternatively, users can set their own threshold. For example, by setting the minor allele threshold to '0', users can observe the association result of all CNVRs regardless of the allele frequency.

D) Running

After selecting statistical methods and setting allele frequency threshold level, click 'Run' key. Then the output of statistical calculation will be displayed in the report screen.

Report Screen

A. CNVR Report

The CNV region determined by user's preference will be displayed in this window. The same list is also stored as a tab-delimited text file on the same directory. The file name consists of original name and region type.

These are example tables for three different type of region

CNVR

RO

Fragment

NMR_1 Y 2,985,939 3,068,364 Gain P NMR_2 Y 3,352,707 3,558,704 Gain P NMR_3 Y 3,613,770 3,578,4134 Gain P NMR_4 Y 3,613,770 3,578,4134 Gain P NMR_5 Y 4,610,751 4,550,549 Gain P NMR_5 Y 4,610,751 4,550,549 Gain P NMR_5 Y 4,610,751 4,550,549 Gain P NMR_5 Y 4,600,883 4,700,229 Gain P PAG Y 4,400,721 4,528,282 Gain NMR_9 Y 5,563,469 5,650,482 Gain P PAG Y 4,410,751 4,550,549 Gain NMR_11 16,633,082 1,639,082 1,729,266 Gain P 7,712,702 7,522,678 Loss NMR_12 10,723 564,621 meed N 7,712,702 7,522,678 Loss NMR_11 16,633,0262,1729,6466 Gain P		example_ci_	GC.cmt0x1s	0c0x3.table	e	+ _ X	2	example_ci_GC.rmt0x1	0c0x3.table		+ _ ×	2	example_ci_	GC.fmt0x1s	0c0x3.table	
CNMP, 2 Y 3,352,707 3,358,704 Gain CNMP, 3 Y 3,613,770 3,352,707 3,552,707 6,350 M2 M2 <th></th> <th></th> <th></th> <th></th> <th></th> <th></th> <th>CNVR ID</th> <th></th> <th></th> <th>Туре</th> <th></th> <th>CNVR ID</th> <th>Chromoso</th> <th>Start</th> <th>End</th> <th>TVD</th>							CNVR ID			Туре		CNVR ID	Chromoso	Start	End	TVD
NNR, 3 Y 3.613.770 3.784.346 Gain NNR, 4 4.181.592 4.224.000 2.33.70 3.784.348 Gain NNR, 4 4.410.751 4.550.549 Gain FRAG.3 Y 3.423.828 J.471.481 Gain NNR, 5 Y 4.410.751 4.550.549 Gain FRAG.4 Y 3.423.828 J.471.481 Gain NNR, 5 Y 4.500.549 Gain FRAG.6 Y 3.423.828 J.471.481 Gain R0, 4 Y 4.410.751 4.550.540 Gain FRAG.6 Y 3.423.828 J.471.482 3.587.704 Gain R0, 7 Y 5.363.615 S.563.461 S.560.821 Gain FRAG.7 Y 3.783.387 J.784.346 Gain R0, 7 Y S.353.461 S.560.821 Gain FRAG.6 Y 4.190.524 J.44.808 Gain R0, 1 Y Y.524.092 Y.525.461 Loss FRAG.10 Y 4.10.7		Y				^					A	FRAG 1	γ	2,985,939	3,086,364	Gain
NNR.4 Y 4.181.582 4.324.606 Gain NNR.5 Y 4.410.751 4.550.549 Gain NNR.6 Y 4.670.883 4.780.229 Gain NNR.7 Y 4.500.883 4.780.229 Gain NNR.7 Y 4.670.883 4.780.229 Gain NNR.9 Y 5.633.346 5.471.500 Gain NNR.9 Y 5.633.461 5.850.821 Gain NNR.1 Y 7.522.678 Loss RO.8 Y 7.522.678 Loss NNR.11 Y 7.522.678 Loss RO.9 Y 7.127.027 7.522.678 Loss NNR.11 1.6.6370.4927 Gain RO.9 Y 7.127.027 7.522.678 Loss NNR.11 1.6.6370.4927 Gain.0 NO.11 1.6.1723 S64.621 m.808 1.6.935.995 Gain NNR.12 1.6.6.637.64.291 M.975 Gain RO.12 1.2.256.58 S64.621 Loss NNR.12 1.6.6.830 RO.15 1.6.936.9217.286.486 <td< td=""><td></td><td>Y</td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td>Gain</td><td></td><td>FRAG 2</td><td>Y</td><td>3.352.707</td><td>3,423,827</td><td>Gain</td></td<>		Y								Gain		FRAG 2	Y	3.352.707	3,423,827	Gain
RNR_1 R181.082 V 4.130.78 Cain FRAG V 3.613.770 3.665.082 Gain FRAG V 3.613.770 3.665.082 3.778.386 Gain FRAG V 3.613.770 3.665.082 3.778.386 Gain FRAG V 4.140.752 V 3.63.869 S.473.1500 Gain FRAG V 4.140.752 V 3.63.869 S.473.1500		1								Gain		FRAG 3	Y	3,423,828	3,471,481	Gain
RNR 7 (4 670 883 (780 22) Cannot RNR 7 (4 670 883 (780 22) Cannot NNR 8 (4 70 883 (470 25) (580 44) (580 64) (780 64) NNR 9 (553 461 (580 64) (580 64) (780 64) (780 64) (780 66) (780 83) (780 85) (=		Y			Gain
PixeR Y 4.000.027 5.013.153 Gain FRAG K 4.670.883 4.780.229 Gain FRAG FRAG K 5.653.365 3.778.385 Gain NMR Y 5.653.461 5.850.821 Gain Ro Y 4.670.883 4.780.229 Gain FRAG Y 3.778.386 Gain NMR Y 7.524.092 7.697.801 Loss FRAG Y 3.778.386 Gain NMR Y 7.524.092 7.697.801 Loss FRAG Y 4.410.751 4.550.549 Gain NMR, 11 Y 7.524.092 7.697.801 Loss FRAG Y 4.410.751 4.450.886 4.780.229 Gain NMR, 12 1.61.023 S66.421 NmR						_							v			
NNR 9 Y 5,363.369 5,471.500 Gain NNR 9 Y 5,653.461 5,861.00 Gain NNR 10 Y 7,127.027 5,522.678 Loss NNR 11 Y 7,524.092 7,697.801 Loss NNR 11 Y 7,524.092 7,697.801 Loss NNR 11 Y 7,524.092 7,697.801 Loss NNR 11 1 6,808.061 6,357.955.66 Gain NNR 11 1 6,808.061 6,357.955 S64.621 Loss RO 11 1 16,17.23 S66,50 Hived NR<0.11		V						4,070,000					v			
NNR 30 Y 5.653.461 5.850.821 Gain NNR 10 Y 7.127.027 7.522.678 Loss NNR 11 Y 7.524.092 7.697.701 Loss N.9 N.9 Y 5.653.461 5.850.821 Gain NNR 12 10.732 564.621 mixed N.9 Y 7.522.678 Loss NNR 13 1 16.890.808 16.935.995 Gain N.0 Y 7.524.678 Loss NNR 14 1 16.890.8084 16.935.995 Gain N.0 Y 7.524.092 7.697.801 Loss NNR 15 1.6.698.3624 17.298.496 Gain N.0 Y 5.653.00 mixed NNR 14 1 16.137.04.937 63.810.371 Gain N.0 1.569.844 5.683.81 Gain NNR 14 1 12.14.898.63 1.644.795 Gain N.1 1.683.080.81 6.931.971 Gain NNR 12 1.244.0957 1.444.955 1.663.704.937 G3.810.371 Gain N.1 1.683.080.821.149.20.02 N.92.20 N		Y						1 10001027					V			
NNR 10 Y 7,127,027 7,522,678 Loss R0.8 Y 5,653,461 5,850,621 Gain NNR 11 1 7,522,067 Loss R0.9 Y 7,127,027 7,522,678 Loss NNR 12 1 61,723 564,621 mixed R0.9 Y 7,220 7,522,678 Loss NNR 13 1 740,857 1,030,307 Gain R0.1 1 61,723 566,530 mixed NNR 15 1 1.6,968,362 17,298,496 Gain R0.14 1 6,800,808 16,935,995 Gain RVR 16 1 0.4,709,873 6.810,737 Gain R0.12 1.235,658 564,621 Loss RVR 11 1 1.632,0808 16,935,995 Gain RAG 13 Y 5,653,461 5,683,415 5,683,451 5,683,451 5,683,451 5,683,451 5,683,451 5,683,416 5,683,416 5,683,416 5,683,416 5,683,416 5,683,416 5,683,416 5,683,416 5,683,416 5,683,416 5,683,416 5,683,416 5,683,416 5,683								Y 5,363,369					T V			
NNR.11 Y 7.524,092 7.697,801 Loss FRAG I Y 7.524,092 7.697,801 Loss FRAG I Y										Gain			ř V			
NNR 12 1 61/23 564,021 NNR 24 1 101/23 1													Y			
NNR_13 1 1,030,307 Gain NNR_14 1 16,683,088 16,935,995 Gain NNR_15 1 16,698,362 17,299,496 Gain NNR_16 1 16,300,808 16,935,995 Gain NNR_11 16,498,362 17,299,496 Gain NNR_11 104,130 104,130,1104,307,2 Gain NNR_18 104,130,1 104,307,2 Gain NNR_18 104,130,1 104,307,2 Gain NNR_10 121,2434,7 121,482,9 Gain NNR_12 144,0367, 144,943, Gain NNR_20 144,0367, 144,842, Gain NNR_22 149,0861 149,228 Gain NNR_22 1243,658 143,259 Gain NNR_22 1244,648 186,574,74 Gain NNR_22 1244,648 186,614 Gain NNR_22 1243,1638 144,945,90 Gain NNR_22 1244,548 186,574,74 Gain		1											Y			
NNR_15 1 1000000000000000000000000000000000000	NVR_13	1					R0_11			mixed			Y			
NNR 16 1 63704937 63810.371 Gain NNR 16 1 64704937 63810.371 Gain NNR 17 1 82.461.308 82.644.795 Gain NNR 19 1 121.343.7121.482.9Gain R0 15 1 6.983.902 17.298.496 Gain NNR 20 1 144.305144.945Gain R0 16 16.370.4937 Gain RAG 15 Y 5.838.453 Gain NNR 20 1 144.307144.849.5Gain R0 18 1.04.307164.849.5Gain RAG 17 Y 7.524.092 7.697.801 Loss NNR 21 1 165.704149.495Gain R0 19 1 121.343.7121.482.9Gain RAG 17 Y 7.524.092 7.697.801 Loss NNR 22 1 145.504149.051.9mixed R0 21 1.45.504149.853.9Gain RAG 18 1 1229.064 229.076 Mixee NR 25 1 196.745196.81.126loss R0 22 1 145.530.4149.853.9Gain RAG 221 229.064 229.063 Mixee RVR 25 1 196.970.2		1					R0_12	1 235,658	564,621	Loss			Y			
NNR 17 1 82.461.630 82.644.795 Gain NNR 18 1 104.130104.307.2Gain FRAG 15 Y \$5.838.454 \$5.90.821 Gain NNR 19 1 121.3437121.4823Gain FRAG 16 16 6.63.704.937 65.81.0.371 Gain NNR 20 1 144.5367144.894.56Gain FRAG 16 18.24.44.795 Gain NNR 21 1.45.206.6145.398.1Gain RO 18 1 104.307.2Gain FRAG 18 1 61.723 85.9.23 Mkee NNR 22 1 146.508							R0_13	1 740,857	1,030,307	Gain			Y	5,653,461	5,683,811	Gain
NNR 18 1 104,130 104,307.2 Gain NNR 19 1 121,343.7 121,482.9 Gain NNR 20 1 144,036.7 144,849.5 Gain NNR 22 1 145,206.6 145,398.1 Gain NNR 22 1 145,206.6 145,398.1 Gain NNR 22 1 146,066.8 Gain R0 18 1 104,130 NNR 22 1 146,066.8 Gain R0 20 1 144,036.7 144,849.5 Gain NNR 23 1 104,056 146,057 144,849.5 Gain R0 20 1 248,0652.0 Gain RV 22 1 144,056.7 144,849.5 Gain R0 20 1 228,695 229.064 Mkee NNR 25 1 126,544.8 1229,608 235,657 Gain R0 22 1 144,953.0 148,953.9 Gain RA_22 1 229,608 229,008 235,657 Gain NR 24 2 243,163.8							R0_14	1 16,830,808	16,935,995	Gain		FRAG_14	Y	5,683,812	5,838,453	Gain
NNR 19 1 121,437,121,482.9Gain NNR 20 121,437,121,482.9Gain NR 20 121,437,121,482.9Gain NNR 20 1 146,0367,144,895Gain RO 17 182,441,650 82,441,650 82,441,650 82,641,850 Bain NNR 22 1 145,026149,021 145,020 145,020 7,027,020 7,97,801 Loss NNR 22 1 146,057, 148,050,							RO 15	1 16,968,362	17,298,496	Gain		FRAG 15	Y	5,838,454	5,850,821	Gain
NNR 20 1 144,056,7 144,945,Gain 1,224,032 1,234,032 1,043,07,2Gain NNR 21 1 145,206,6 145,398,1Gain 1,043,07,2Gain 1,043,07,2Gain NNR 22 1 146,530,4 149,051,9mixed 1,043,07,2Gain 1,043,07,2Gain NNR 22 1 146,530,4 149,051,9mixed 1,043,07,2Gain 1,043,07,2Gain NNR 23 1 149,056,149,022,8Loss 1,043,07,2Gain 1,043,07,2Gain NNR 24 1.166,574,7166,066,8Gain 1,043,07,2Gain 1,043,07,2Gain NNR 25 1 122,454,8182,611,6Loss 1,043,07,2Gain NNR 26 1.062,153,0Gain 1,043,07,2Gain NNR 27 1 243,163,8243,274,5Gain NNR 28 2.0,441 148,930,4149,951,9Loss RO 24 1.149,947,6149,952,9Loss RO 25 1.190,961,1149,922,8Loss RO 28 1.96,706,2196,812,5Gain NNR 30 2.12,2464,0124,08Gain NNR 31 114,130,3		1					RO 16	1 63,704,937	63,810,371	Gain		FRAG 16	γ	7,127,027	7,522,678	Loss
NMR_20 1 144,036 144,949.5Gain 144,949.5Gain NMR_21 1 145,206 145,3981Gain 146,1723 85,923 Mixed NMR_22 1 146,504 149,051.9mixed R0 19 121,343.7 124,482.9Gain FRAG_19 1 85,924 228,694 Mixed NMR_22 1 166,574.7 166,966.8Gain R0 144,036.7 144,495.50Gain FRAG_20 1228,694 Mixed NMR_24 1 166,574.7 166,966.8Gain R0 21 144,5206.6 145,3981Gain NMR_25 1 196,706.2 196,706.2 196,706.2 196,510.5.0. Mixed NMR_26 1 166,574.7 166,966.8 Gain R0<23		1					R0 17	1 82,461,630	82,644,795	Gain		FRAG 17	Y	7.524.092	7.697.801	Loss
NVR.21 1 145,399,1163in 145,399,1163in 145,399,1163in NVR.22 1 146,500,149,020,80ss NVR.23 1 149,086,149,202,80ss NVR.24 122,843,7121,482,9Gain NVR.23 1 146,506,86ain NVR.23 1 145,200,6143,399,16ain RO 20 1 144,036,7143,399,16ain NVR.24 1 166,574,7166,966,86ain NVR.25 1 1245,206,6145,399,16ain RO 22 1 145,504,149,953,96ain NVR.25 1 126,5126ain RO 22 1 145,504,149,953,96ain RO 22 1 249,643 156,531 566,630 Mixed NVR.26 2 29,443 145,899 Gain RO 22 1 149,953,96ain RO 223 1 245,616,610 NKR RO 24 1 149,953,96ain RO 23 1 229,608 235,657 Gain NVR.29 2 29,443 156,893 1 166,974,7166,966,86ain RO 27 1 186,951,95 Gain RO 27 1 186,951,95 Gain RO 29 <td></td> <td>1</td> <td></td> <td></td> <td></td> <td></td> <td>R0 18</td> <td>1 104.130.1.</td> <td>. 104.307.2</td> <td>Gain</td> <td></td> <td></td> <td>1</td> <td></td> <td></td> <td>Mixed</td>		1					R0 18	1 104.130.1.	. 104.307.2	Gain			1			Mixed
NMR 22 1 149,530.4149,051.9Mixed NMR 23 1 149,051.9Mixed 1 144,036.7144,849.5Gain NMR 24 1 166,574166,966.8Gain 780,201 145,206.6143,939.1Gain NMR 25 1 122,745Gain 780,221 144,8350.4148,652.6Gain 780,221 124,8530.4148,953.98Gain NMR 26 1 126,574166,966.8Gain 780,221 144,8530.4148,953.98Gain 780,221 1229,064 229,067 Gain NMR 26 1 166,574164,92.5Gain 780,221 144,8530.4148,953.98Gain 780,221 229,043 249,042 229,063 356,550 Mixed NMR 27 2 29,443 145,899 Gain 780,221 124,966.1149,051.9Coss 780,221 235,658 356,530 Mixed NMR 29 2 29,443 145,899 Gain 780,221 146,966.816ain FRAG 24 128,651.6Loss R0 27 1 128,454.8		1					R0 19	1 121.343.7.	. 121.482.9	Gain			1			
NVR.23 1 149,086,1149,024,81085 NVR.23 1 149,086,1149,024,81085 NVR.23 1 149,026,1149,024,81085 NVR.23 1 149,026,1149,024,81085 NVR.25 1 1229,068,806ain FRAG 22 1 229,0607 Gain NVR.25 1 196,706,2196,812,506ain RO 22 1 148,530,4148,953,906ain FRAG 22 1 229,0607 Calin NVR.25 1 196,706,2196,812,506ain RO 23 1 148,530,4148,953,906ain FRAG 23 1 229,608 235,657 Gain NVR.25 2 243,163,8243,274,506ain RO 24 1 149,906,1149,202,806s5 FRAG 23 1 225,658 356,530 564,621 Loss NVR.29 2 89,90,061 91,790,736 Gain RO 26 1 166,574,7166,966,806ain FRAG 24 1 356,531 564,621 Loss NVR.30 2 114,130,3114,230,506ain RO 27 1 182,454,8182,611,60585 FRAG 27 1 16,968,362 17,025,50 Gain NVR.		1					R0 20						1			
NMR_22 100,2454100,2601.61003 NMR_22 102,2454100,2601.61003 NMR_22 102,2454100,2601.61003 NMR_22 106,2454100,2601.61003 NMR_22 106,2741661035 NMR_22 106,274162168148,953.0148,953.0168 NMR_22 29,443 148,899 Gain NMR_22 29,443 145,899 Gain RO_22 1148,953.0149,051.9Loss RO_22 196,774166,966.8Gain NMR_29 29,443 144,80.01,12.6440112,40.0.8Gain RO_22 1148,953.0149,0567I69.0658.8Gain RO_22 198,900,061 NRR_30 212,640124,90.8Gain RO_27 1182,454.8182,611.6Loss RO_29 1243,163.8243,274.5Gain RO_29 1243,163.8		1											1			
NNR.25 1 162,034,8 162,014,0 1033 NNR.25 1 1282,048 162,014,0 1033 NNR.25 1 1282,048 1282,048 1282,0108 233,037 Gain NNR.25 1 1243,165.8 243,274,5 Gain NNR.25 148,954,9 148,954,9 Ia82,014 148,954,9 Ia82,014 148,954,9 Ia82,014 148,954,9 Ia82,014 Ia82													1			
NMR_20 1 260,700,510,500,233,508,300,700,736 Mixed NMR_21 2 233,163,8,232,74,5Gain Nixed Nixed <td< td=""><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td>1</td><td></td><td></td><td></td></td<>													1			
RNR_28 2 20,43 145,899 Gain NNR_28 2 20,43 145,899 Gain NNR_29 2 89,133,112 89,500,461 Gain NNR_29 2 89,000,61 91,790,736 Gain NNR_30 2 112,464,0 112,649,8 Gain NNR_30 2 114,130,3 114,236,6 Gain NNR_30 2 113,932,4 132,170,8 Gain NNR_30 2 114,130,3 114,236,6 Gain R0_28 1 243,163,8 243,274,5 Gain R0_30 2 29,43 145,899 Gain R0_31 2 89,133,112 89,500,461 Gain R0_32 2 89,428,082 90,023,495 Gain FRAG_32 1 7,035,651 17,035,650 17,035,650 R0_32 2 89,428,082 90,023,495 Gain FRAG_32 1 17,037,084 Gain FRAG_32 1 17,037,084 Gain FRAG_32 1		1											1			
NVR 29 1 2 89.133.112 89.500.461 Gain NVR 29 2 2 89.900.661 91.790.736 Gain RO.27 1 166.574.7 166.966.8 Gain FRAG.25 1 740.857 1,030.70 Gain NVR 30 2 112.464.0112.c49.8 Gain RO.27 1 182.454.8 182.611.6 Loss NVR 30 2 114.130.3 114.236.8 Gain RO.29 1 243.163.8 1243.163.8 243.274.5 Gain FRAG.27 1 16.968.9080 16.935.995 Gain NVR 32 2 131.932.4 132.170.8 Gain RO.29 1 243.163.8 243.274.5 Gain FRAG.29 1 7.035.207 Gain RO.32 2.94.43 145.899 Gain FRAG.29 1 7.035.201 7.035.207 Gain RO.32 2.99.428.082 90.023.495 Gain FRAG.29 1 7.035.201 7.037.084 Gain RO.32 2 89.428.082 90.023.495 Gain FRAG.32 1		2											1			
NNR 29 2 2 89,900,061 91,790,736 Gain NNR 30 2 112,464,0 112,649,8 Gain RO 27 1 182,454,8 182,611,6 Loss FRAG 26 1 16,830,808 16,935,995 Gain NNR 30 2 112,464,0 112,649,8 Gain RO 28 1 196,706,2 Is812,55 Gain FRAG 28 1 17,029,580 17,035,207 Gain NNR 31 2 131,932,4 132,170,8 Gain RO 30 2 29,443 145,899 Gain FRAG 28 1 17,029,580 17,035,501 Gain RO 31 2 89,133,112 89,500,461 Gain FRAG 30 1 17,037,084 Gain FRAG 32 1 17,037,084 6 RO 32 2 89,428,082 90,023,495 Gain FRAG 31 1 17,037,084 Gain FRAG 32 1 17,045,446 17,177,033 Gain FRAG 32 1 17,045,446 17,177,033 Gain													1			
NNR 30 2 112,464.012,649.8Gain NNR 31 2 114,130.3114,236.6Gain NNR 32 2 131,932.4132,170.8Gain NNR 32 131,932.4132,170.8Gain NNR 32 2 131,932.4132,170.8Gain NR 32 2 131,932.4132,170.8Gain NR 32 2 131,932.4132,170.8Gain NR 32 2 89,133,112 89,500,461 Gain FRAG_29 1 17,035,208 17,037,084 Gain R0 32 2 89,428,082 90,023,495 Gain FRAG_30 1 17,036,530 I7,037,084 Gain R0 32 2 89,428,082 90,023,495 Gain FRAG_31 1 17,037,085 I7,045,445 Gain R0 32 2 89,428,082 90,023,495 Gain FRAG_32 1 17,045,445 Gain R0 32 2 89,428,082 90,023,495 Gain FRAG_32 1 17,045,445 Gain R0 32 2 89,428,082 90,023,495 Gain FRAG_32 1 <td< td=""><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td></td<>																
NNR.31 2 114,130.3114,236.6Gain FRAG_29 1 17,029,580 17,035,207 Gain NNR.32 2 131.932.4132.170.8Gain RO<30												FRAG_27	1	16,968,362	17,029,579	Gain
NVR_32 2 131,932,4132,170,8Gain Image: Constraint of the state of												FRAG_28	1	17,029,580	17,035,207	Gain
RO 31 2 89,133,112 89,500,461 Gain FRAG 30 1 17,036,531 17,037,084 Gain RO 32 2 89,428,082 90,023,495 Gain FRAG 1 17,037,085 17,045,445 Gain FRAG 32 2 89,428,082 90,023,495 Gain FRAG 1 17,037,085 17,045,445 Gain FRAG 32 1 17,045,446 17,177,033 Gain FRAG 32 1 17,045,446 17,177,033 Gain												FRAG 29	1	17,035,208	17,036,530	Gain
RO 2 89,428,082 90,023,495 Gain FRAG 31 17,037,085 17,045,445 Gain FRAG 32 10,428,082 90,023,495 Gain FRAG 1 17,045,446 17,177,033 Gain		12				- -						FRAG 30	1	17.036.531	17.037.084	Gain
FRAG_32 1 17,045,446 17,177,033 Gain																
							NU_32	2 03,428,082	30,023,495		-					
												FRAG 33				

B. Association analysis Report

This is the window for the output of the association test. You can sort it by any column by clicking its header. It is also stored as a tab-delimited text file. The detailed option information is written in the header of the file.

9			exa	ample_ci_G	C.20111217	112307				<u>+</u> -
CNVR ID	Chr	Start	End	Size	Freq. (Co	Freq. (Ca	Description	p value	-2 Log LRT	LRT.
CNVR_73	9	68,683,835	69,942,276	1,258,442	2	1	Gain	0.220824	7.023	0.26
CNVR_93	14	19,002,112	20,422,583	1,420,472	1	3	mixed	0.300951	7.863	0.3825
CNVR_139	22	18,626,234	18,887,369	261,136	1	3	Gain	0.300951	7.863	0.3825
CNVR_51	5	68,867,282	70,178,835	1,311,554	2	1	Loss	0.362467	8.179	0.4874
CNVR_40	3	129,715,	129,914,	199,134	1	3	Gain	0.362467	8.179	0.4874
CNVR_145	22	22,864,059	23,258,994	394,936	1	3	Gain	0.362467	8.179	0.4874
CNVR_110	16	32,113,670	32,573,464	459,795	2	1	mixed	0.376290	8.227	0.5142
CNVR_66	8	7,222,169	7,809,894	587,726	2	1	mixed	0.376290	8.227	0.5142
CNVR_137	22	16,055,171	16,386,602	331,432	1	2	Gain	0.398321	8.251	0.4745
CNVR_151	Х	88,861,135	89,182,355	321,221	2	3	Gain	0.442138	8.38	0.549
CNVR_70	9	39,313,808	41,480,601	2,166,794	1	1	Gain	0.442138	8.38	0.549
CNVR_60	7	143,917,	144,066,	149,347	1	2	Gain	0.545384	8.68	0.532
CNVR_118	17	44,394,400	44,794,572	400,173	1	2	mixed	0.585999	8.725	0.540
CNVR_56	7	61,063,962	61,839,758	775,797	1	1	Gain	0.628756	8.816	0.608
CNVR_68	8	47,012,218	47,262,143	249,926	1	1	Gain	0.628756	8.816	0.608
CNVR_123	19	90,898	258,072	167,175	1	1	Gain	0.628756	8.816	0.608
CNVR_108	16	21,412,391	21,620,547	208,157	1	1	Loss	0.687884	8.9	0.6236
CNVR_88	12	8,357,507	8,601,982	244,476	1	1	Gain	0.687884	8.9	0.6236
CNVR_83	11	3,426,602	3,624,237	197,636	1	1	Gain	0.687884	8.9	0.6236
CNVR_44	4	69,338,450	69,489,323	150,874	2	3	Loss	0.687884	8.9	0.623
CNVR_79	10	51,231,564	51,479,639	248,076	1	1	Gain	0.706006	8.918	0.6271
CNVR_45	4	132,545,	132,780,	234,793	1	2	Gain	0.714802	8.929	0.5836
CNVR_34	2	132,873,	133,136,	262,772	2	2	Gain	0.746615	8.96	0.5911
A SQ	7	1/3 211	1/13 576	364 411	1	1	mixed	0 842918	9.026	0 649

References

Bae, JS. *et al.* (2010) Genome-wide association analysis of copy number variations in subarachnoid aneurysmal hemorrhage. *J. Hum. Genet.*, **55(11)**, 726-30

Barnes, C. *et al.* (2008) A robust statistical method for case-control association testing with copy number variation. *Nat. Genet.*, **40**, 1245–1252.

Colella, S. *et al.* (2007) QuantiSNP: an Objective Bayes Hidden-Markov Model to detect and accurately map copy number variation using SNP genotyping data. Nucleic Acids Res., **35**, 2013–2025.

Forer,L. *et al.* (2010) CONAN: copy number variation analysis software for genome-wide association studies. *BMC Bioinformatic*, **11**, 318

Joeng, Y. *et al.* (2008) CGHscape: A Software Framework for the Detection and Visualization of Copy Number Alterations. Genome & Informatics, **6(3)**, 126-129

Korn, J.M. *et al.* (2008) Integrated genotype calling and association analysis of SNPs, common copy number polymorphisms and rare CNVs. *Nat. Genet.*, **40**, 1253–1260.

Pique-Regi, R. *et al.* (2010) R-Gada: a fast and flexible pipeline for copy number analysis in association studies. *BMC Bioinformatics*, **11**, 380

Purcell, S. *et al.* (2007) PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am. J. Hum. Genet.*, **81**, 559–575.

Subirana, I. *et al.* (2011) CNVassoc: Association analysis of CNV data using R. *BMC Medical Genomics*, **4**:47

The Cancer Genome Atlas (TGCA) research Network (2008) Comprehensive genomic characterization defines human glioblastoma genes and core pathways. Nature, **455**,1061–1068.

Wang,K. *et al.* (2007) PennCNV: An integrated hidden Markov model designed for high-resolution copy number variation detection in whole-genome SNP genotyping data. *Genome Res.*, **17**, 1665.

Wittig, M. et al. (2010) CNVineta: a data mining tool for large case–control copy number variation datasets. *Bioinformatics*, **26**, 2208-2209