

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](http://www.ircgp.com/CNVRuler/index.html)

Contact to developer:
Yeun-Jun Chung (yejun@catholic.ac.kr) and Ji-Hong Kim (lomolith@gmail.com)

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 (<http://www.oracle.com/technetwork/java/javase/downloads/index.html>).

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. <http://www.r-project.org/>

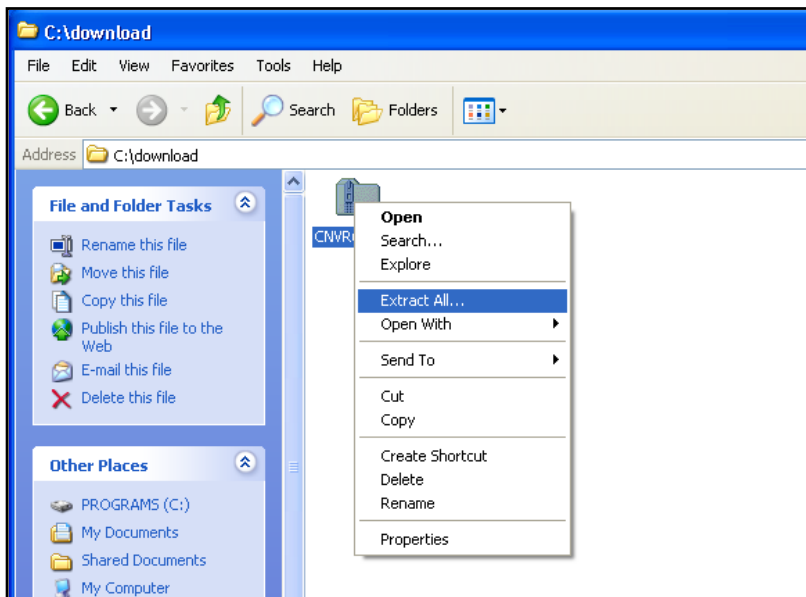
After selecting download mirror site and OS platform, you can download the distribution binaries. If your system doesn't have the R package on it, CNV-Ruler will prompt it and try to open the R download site.

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.

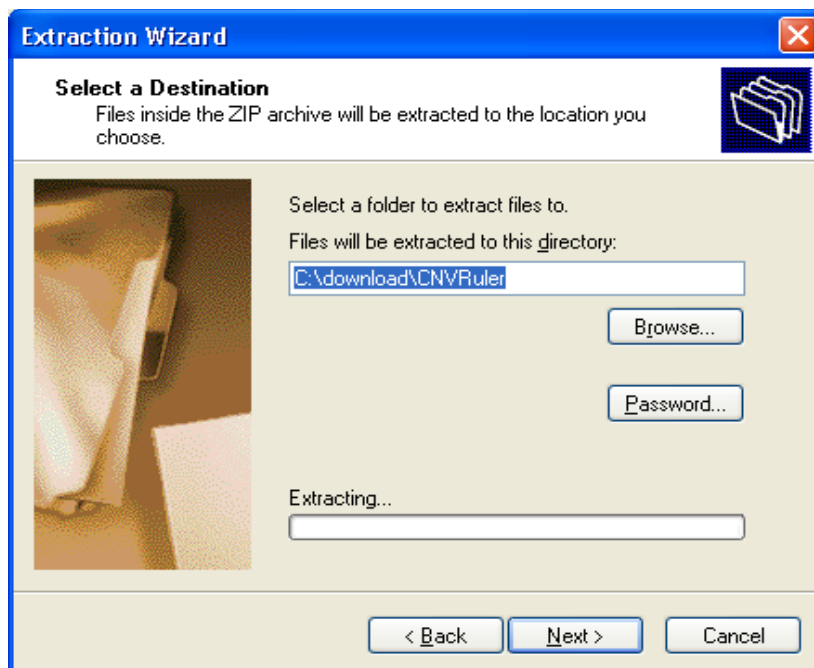
2. Installation

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.

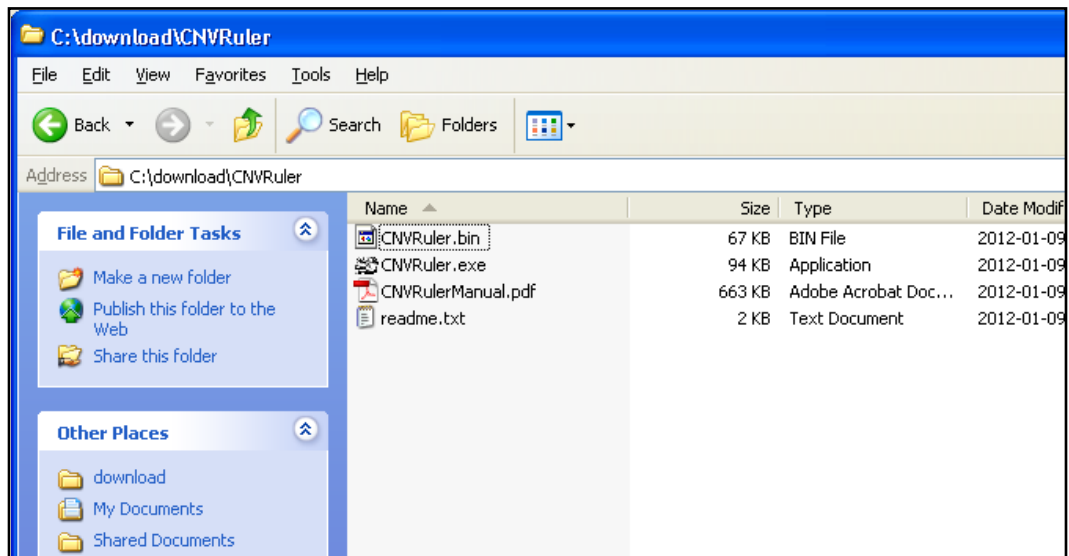
Step 1) Download compressed zip file. Right click and select *Extract All*.



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

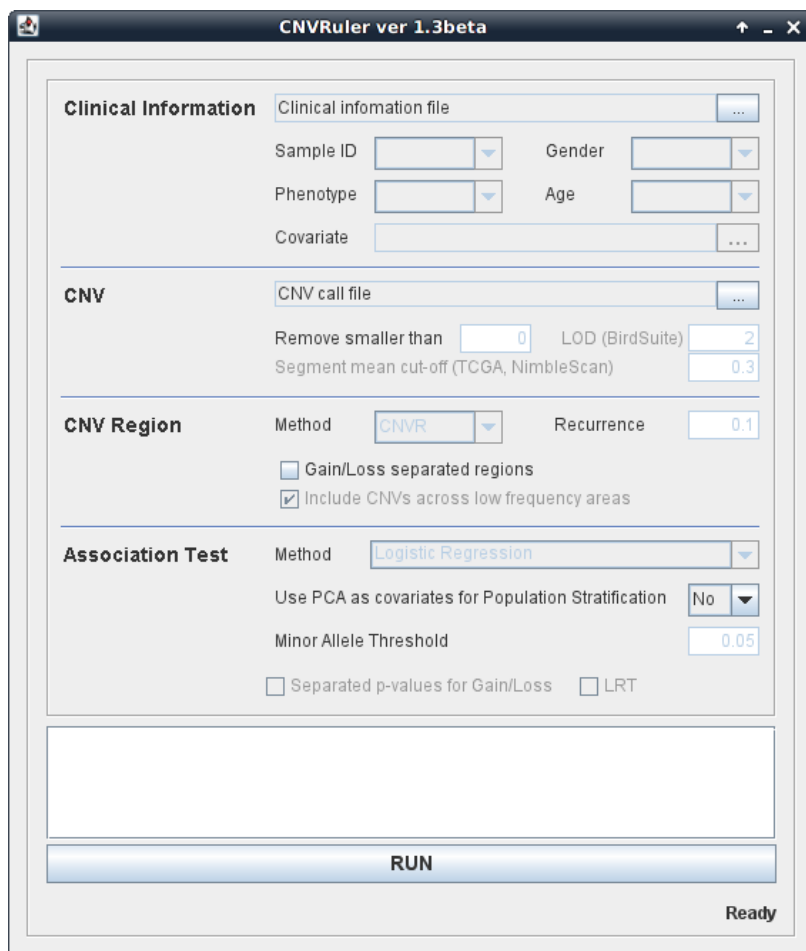


Step3) Double click to execute CNVRuler.exe



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.



3. Data analysis

A) Data uploading

The screenshot shows a web-based interface for data uploading. It is divided into two main sections: 'Clinical Information' and 'CNV'.
Under 'Clinical Information', there is a text input field for 'Clinical information file' with a browse button (...). Below this are four dropdown menus: 'Sample ID', 'Gender', 'Phenotype', and 'Age'. At the bottom of this section is a text input field for 'Covariate' with a browse button (...).
Under 'CNV', there is a text input field for 'CNV call file' with a browse button (...). Below this are three input fields: 'Remove smaller than' with a value of '0', 'LOD (BirdSuite)' with a value of '2', and 'Segment mean cut-off (TCGA, NimbleScan)' with a value of '0.3'.

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

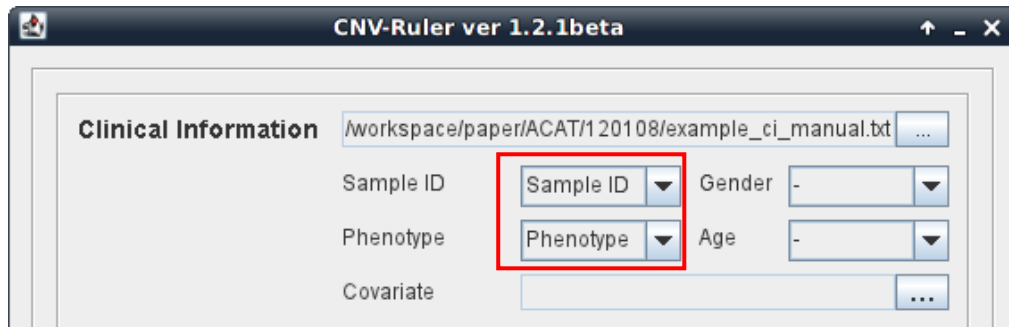
- **Step 1. Uploading clinical data**

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

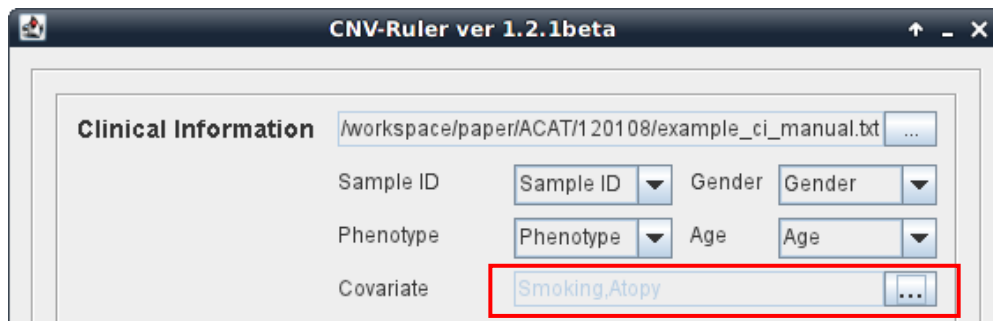
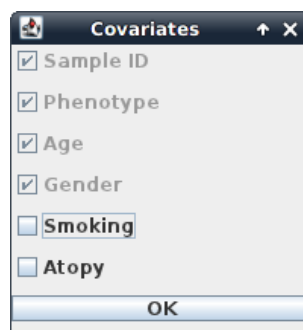
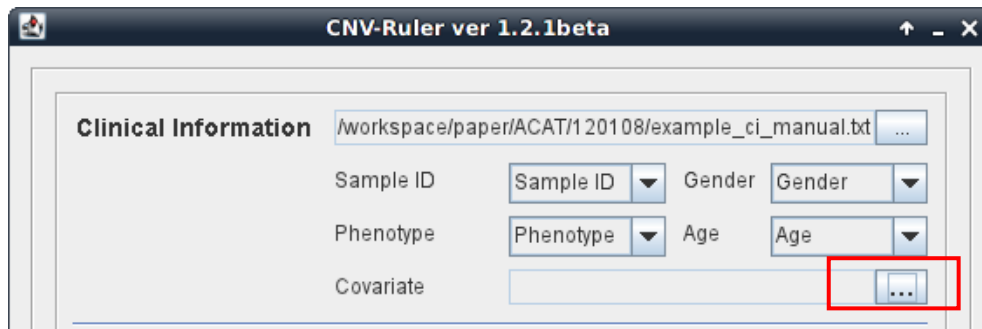
1) Analysis with sample ID and phenotype



2) Analysis with four main CIs



3) Analysis with additional covariates



- **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
Genotyping Console	4.1	www.affymetrix.com

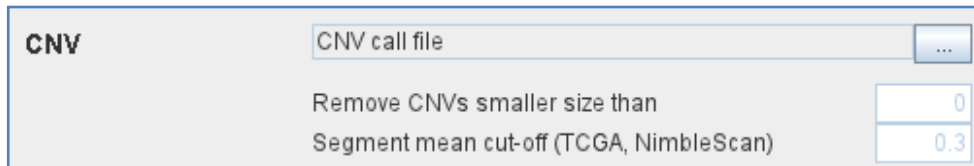
If you want to use your own CNV list file, you must prepare a simple tab-delimited text file containing 5 columns as the example below. The names and order of column headings should be Chr, Start, End, Event and Sample_ID respectively.

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.



The screenshot shows a user interface for CNV data uploading. It features a header 'CNV' on the left. To the right, there is a text input field labeled 'CNV call file' with a browse button ('...'). Below this, there are two rows of filtering options. The first row is 'Remove CNVs smaller size than' with a text input field containing the value '0'. The second row is 'Segment mean cut-off (TCGA, NimbleScan)' with a text input field containing the value '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	<input type="text" value="CNVR"/>	Recurrence	<input type="text" value="0.1"/>
	<input type="checkbox"/>	Gain/Loss separated regions		
	<input checked="" type="checkbox"/>	Include CNVs across low frequency 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.

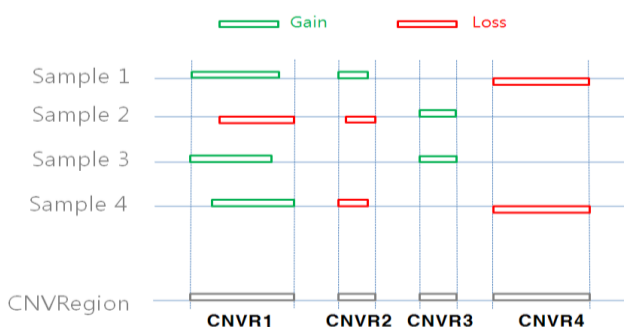
i. CNVR trimming threshold

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.

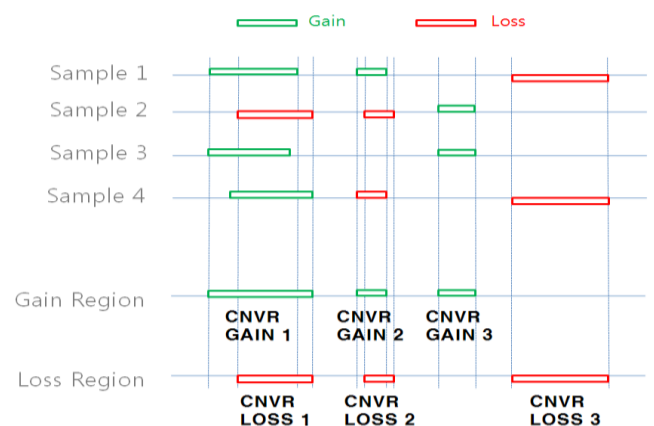
ii. Additional options for building CNVRs: Gain/Loss separated region

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.

Ex) Option is off:

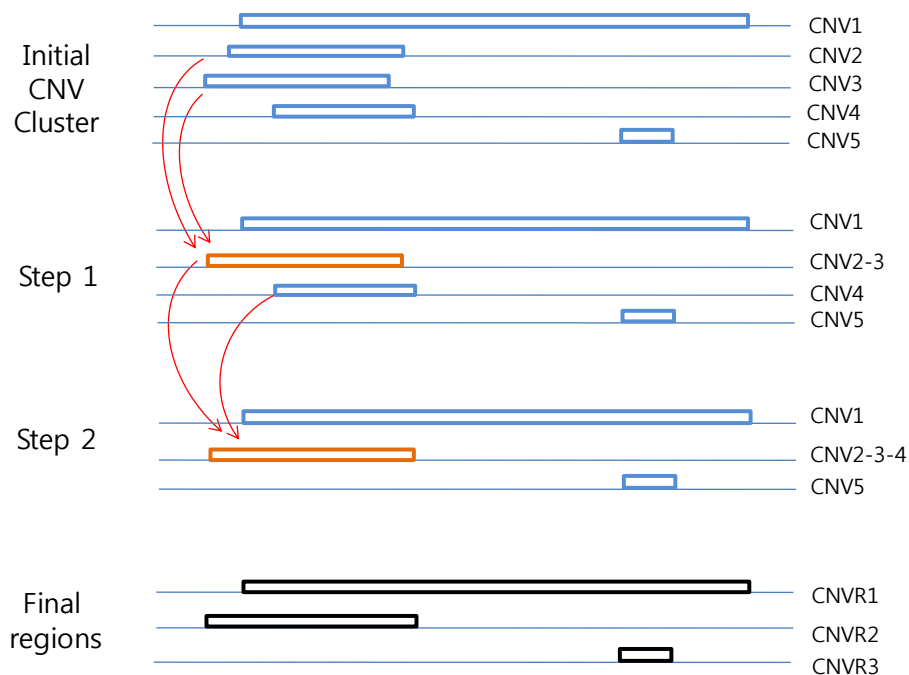


Gain/Loss option is on:



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.

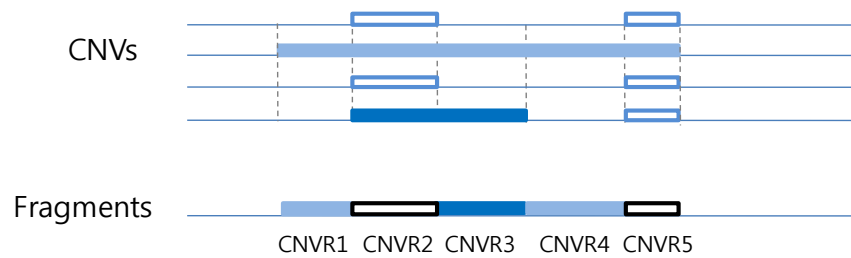


In step 1, since the pair consisting of CNV2 and CNV3 has the highest RO, these two CNVs are merged into a CNV element called CNV2-3. Similarly, in step 2, CNV2-3 and CNV4 are merged into CNV2-3-4. As the RO values of all the remaining pairs do not pass the RO threshold, three CNV regions are defined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

Total CNV List

Chr	Start	End	Type
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

1) CNVR (Recurrence Threshold is 0.1)

Chr	Start	End	Type
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

2) RO

Chr	Start	End	Type
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	Type
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 as covariates for Population Stratification	No
	Minor Allele Threshold	0.05
	<input type="checkbox"/> Separated p-values for Gain/Loss <input type="checkbox"/> LRT	

- **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

CNVR ID	Chromoso...	Start	End	Type
CNVR_1	Y	2,985,939	3,086,364	Gain
CNVR_2	Y	3,352,707	3,538,704	Gain
CNVR_3	Y	3,613,770	3,784,346	Gain
CNVR_4	Y	4,181,582	4,324,808	Gain
CNVR_5	Y	4,410,751	4,550,549	Gain
CNVR_6	Y	4,670,883	4,780,229	Gain
CNVR_7	Y	4,909,027	5,013,153	Gain
CNVR_8	Y	5,363,369	5,471,500	Gain
CNVR_9	Y	5,653,461	5,850,821	Gain
CNVR_10	Y	7,127,027	7,522,678	Loss
CNVR_11	Y	7,524,092	7,697,801	Loss
CNVR_12	1	61,723	564,621	mixed
CNVR_13	1	740,857	1,030,307	Gain
CNVR_14	1	16,830,808	16,935,995	Gain
CNVR_15	1	16,968,362	17,298,496	Gain
CNVR_16	1	63,704,937	63,810,371	Gain
CNVR_17	1	82,461,630	82,644,795	Gain
CNVR_18	1	104,130,1...104,307,2...		Gain
CNVR_19	1	121,343,7...121,482,9...		Gain
CNVR_20	1	144,036,7...144,849,5...		Gain
CNVR_21	1	145,206,6...145,398,1...		Gain
CNVR_22	1	148,530,4...149,051,9...		mixed
CNVR_23	1	149,086,1...149,202,8...		Loss
CNVR_24	1	166,574,7...166,966,8...		Gain
CNVR_25	1	182,454,8...182,611,6...		Loss
CNVR_26	1	196,706,2...196,812,5...		Gain
CNVR_27	1	243,163,8...243,274,5...		Gain
CNVR_28	2	29,443	145,899	Gain
CNVR_29_1	2	89,133,112	89,500,461	Gain
CNVR_29_2	2	89,900,061	91,790,736	Gain
CNVR_30	2	112,464,0...112,649,8...		Gain
CNVR_31	2	114,130,3...114,236,6...		Gain
CNVR_32	2	131,932,4...132,170,8...		Gain

RO

CNVR ID	Chromoso...	Start	End	Type
RO_0	Y	2,985,939	3,086,364	Gain
RO_1	Y	3,352,707	3,538,704	Gain
RO_2	Y	3,613,770	3,784,346	Gain
RO_3	Y	4,181,582	4,324,808	Gain
RO_4	Y	4,410,751	4,550,549	Gain
RO_5	Y	4,670,883	4,780,229	Gain
RO_6	Y	4,909,027	5,013,153	Gain
RO_7	Y	5,363,369	5,471,500	Gain
RO_8	Y	5,653,461	5,850,821	Gain
RO_9	Y	7,127,027	7,522,678	Loss
RO_10	Y	7,524,092	7,697,801	Loss
RO_11	1	61,723	356,530	mixed
RO_12	1	235,658	564,621	Loss
RO_13	1	740,857	1,030,307	Gain
RO_14	1	16,830,808	16,935,995	Gain
RO_15	1	16,968,362	17,298,496	Gain
RO_16	1	63,704,937	63,810,371	Gain
RO_17	1	82,461,630	82,644,795	Gain
RO_18	1	104,130,1...104,307,2...		Gain
RO_19	1	121,343,7...121,482,9...		Gain
RO_20	1	144,036,7...144,849,5...		Gain
RO_21	1	145,206,6...145,398,1...		Gain
RO_22	1	148,530,4...148,662,7...		Gain
RO_23	1	148,530,4...148,953,9...		Gain
RO_24	1	148,947,6...149,051,9...		Loss
RO_25	1	149,086,1...149,202,8...		Loss
RO_26	1	166,574,7...166,966,8...		Gain
RO_27	1	182,454,8...182,611,6...		Loss
RO_28	1	196,706,2...196,812,5...		Gain
RO_29	1	243,163,8...243,274,5...		Gain
RO_30	2	29,443	145,899	Gain
RO_31	2	89,133,112	89,500,461	Gain
RO_32	2	89,428,082	90,023,495	Gain

Fragment

CNVR ID	Chromoso...	Start	End	Type
FRAG_1	Y	2,985,939	3,086,364	Gain
FRAG_2	Y	3,352,707	3,423,827	Gain
FRAG_3	Y	3,423,828	3,471,481	Gain
FRAG_4	Y	3,471,482	3,538,704	Gain
FRAG_5	Y	3,613,770	3,665,082	Gain
FRAG_6	Y	3,665,083	3,778,385	Gain
FRAG_7	Y	3,778,386	3,784,346	Gain
FRAG_8	Y	4,181,582	4,324,808	Gain
FRAG_9	Y	4,410,751	4,550,549	Gain
FRAG_10	Y	4,670,883	4,780,229	Gain
FRAG_11	Y	4,909,027	5,013,153	Gain
FRAG_12	Y	5,363,369	5,471,500	Gain
FRAG_13	Y	5,653,461	5,683,811	Gain
FRAG_14	Y	5,683,812	5,838,453	Gain
FRAG_15	Y	5,838,454	5,850,821	Gain
FRAG_16	Y	7,127,027	7,522,678	Loss
FRAG_17	Y	7,524,092	7,697,801	Loss
FRAG_18	1	61,723	85,923	Mixed
FRAG_19	1	85,924	228,694	Mixed
FRAG_20	1	228,695	229,063	Mixed
FRAG_21	1	229,064	229,607	Gain
FRAG_22	1	229,608	235,657	Gain
FRAG_23	1	235,658	356,530	Mixed
FRAG_24	1	356,531	564,621	Loss
FRAG_25	1	740,857	1,030,307	Gain
FRAG_26	1	16,830,808	16,935,995	Gain
FRAG_27	1	16,968,362	17,029,579	Gain
FRAG_28	1	17,029,580	17,035,207	Gain
FRAG_29	1	17,035,208	17,036,530	Gain
FRAG_30	1	17,036,531	17,037,084	Gain
FRAG_31	1	17,037,085	17,045,445	Gain
FRAG_32	1	17,045,446	17,177,033	Gain
FRAG_33	1	17,177,034	17,182,425	Gain

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.

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	X	88,861,135	89,182,355	321,221	2	3	Gain	0.442138...	8.38	0.5495
CNVR_70	9	39,313,808	41,480,601	2,166,794	1	1	Gain	0.442138...	8.38	0.5495
CNVR_60	7	143,917,...	144,066,...	149,347	1	2	Gain	0.545384...	8.68	0.5325
CNVR_118	17	44,394,400	44,794,572	400,173	1	2	mixed	0.585999...	8.725	0.5408
CNVR_56	7	61,063,962	61,839,758	775,797	1	1	Gain	0.628756...	8.816	0.6085
CNVR_68	8	47,012,218	47,262,143	249,926	1	1	Gain	0.628756...	8.816	0.6085
CNVR_123	19	90,898	258,072	167,175	1	1	Gain	0.628756...	8.816	0.6085
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	0.6236
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
CNVR_58	7	143,711	143,576	364,411	1	1	mixed	0.812918	9.026	0.6467

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 (TCGA) 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) CNVneta: a data mining tool for large case–control copy number variation datasets. *Bioinformatics*, **26**, 2208-2209