Step by Step Tutorial of CRCView Web Server

CRCView is a web based tool designed for clustering microarray gene expression data. Clustering is performed on genes based on their expression profiles across multiple experiments. A detailed step-by-step walk through is provided as follows.

1. Get started.

Point your browser to <u>http://crcview.hegroup.org/</u>. To start, click on "CRC analysis" in the yellow bar at the top as highlighted in Figure 1. At the next page, click "here" at the second line as highlighted in Figure 2. This will take you to the login page.

CRCView - Mozilla Firefox							
ile <u>E</u> dit <u>V</u> iew Hi <u>s</u> tory <u>B</u> ookmarks			4				
🛛 🕶 🔶 😴 🙆 🏠 🔪 http://helab	🧼 👻 📀 🏠 🔪 http://helab.bioinformatics.med.umich.edu/crcview/						
	CR	CView					
Home	RC Analysis Tutorial	Documentatio	on FAQs	Reference			
Welcome to CRCView CRCView is a web-based microarray cluster, a Dirichlet process model-ba expression analysis programs from g analysis, rich graphical illustration, a interpretation of clustering results. Recent updates: 2007-03-12: CRC engine upd removed nonsignific ant cluste 2007-03-09: Added probe to 1 2007-03-09: Added probe to 1 2007-03-02-27: Minimum express 2007-02-27: Minimum express 2007-01-16: Coefficient of var Click here to check all the change to	v data analysis and visualiz used clustering algorithm re <u>sloconductor</u> , CRCView allo nd integrated Gene Ontolo ated. Fixed a bug (occurs to rs from output. VCBI ENTREZ ID and gene , Bloconductor to 2.0, Gos hange in result page. ison level and fold change t iation filtering added. gs.	ation system. CR cently developed ws flexible input gy (GO)-based g when inversion fla symbol convertion ats to 2.1.12, an iltering added. A	CView is pow d by <u>Dr. Stew</u> data format, jene enrichm ag is turned d on d added chic step by step	ered by <u>CRC</u> , or Chinese <u>e Oin</u> . It also incoporates s automated model-based C ent for efficient annotation off but max shift is positive) ken annotation. tutorial on how to use CR	Restaurant everal gene CRC clustering and), and CView added.		
Login	Register Acknowledgen	ients Links	Contact Us	Disclaimer			
	UNIVERS	ITY OF MICH	HIGAN				
					393 since 8/2/20		
one				🥹 🌢 🚄			

Figure 1.



Figure 2.

2. Set up an account.

Follow the instruction to set up an account at the CRCView website and login as shown in Figure 3. This will take you to you own private account in CRCView system, where all your uploaded data and analysis results will be securely stored. As shown in Figure 4, all historical records are automatically saved in CRCView web server such that you can easily find out previous analyses you have conducted. The account is secure and the content is encrypted, no one other than the owner of the account is able to see the contents.

File Edit Yew Higtory Bookmarks Tools Help	Screw - Mozilla Firefox										
Image: Image	<u>Eile E</u> dit <u>V</u> iew Hi <u>s</u> tory <u>B</u> ookmarks <u>Tools H</u> elp					$\langle \rangle$					
CRC Analysis Tutorial Documentation FADs Reference Login Please login if you wish to submit data or to perform Bayesian analysis. If you have not registered for an account, then please on the registration page and sign up for an account. Upue have aiready registered for an account, then please enter your registered email address and password to login to the system. If you have an account but have forgotten your pagesword to login to the system. If you have an account but have forgotten your pagesword, then you may reset your password from here. Intermediate a system of the register and password demo to login. Intermediate account registered for an account implease enter your registered mail address and password from here. Intermediate account registered for an account implease of the register an account implease use email democe of up and password demo to login. Intermediate account registered for an account registered for an account registered for an account registered for an account register anaccount re	 	.med.umich.edu/	/crcview/sandbox/d	data/index.ph;	🔹 🔹 💽 🖌 Google	Q					
Home CRC Analysis Tutorial Documentation FAQs Reference Login Please login if you wish to submit data or to perform Bayesian analysis. If you have not registered for an account, then please go to the registration page and sign up for an account. If you have already registered for an account, then please enter your registered email address and password to login to the system. If you have an account but have forgotten your password to not new your any case your password from here. For demontration, please use email demone du and password demo to login. Image: Image: Image: Image: Contact Us Distainer Login Register Actnowledgements Links Contact Us Distainer	CRCView										
Login Pease login if you wish to submit data or to perform Bayesian analysis. If you have not clustered for an account, then please enter your registered email address and you may registered for an account, then please enter your registered email address and you may registered your password from here. To the work of the system. If you have an account but have forgotten your sussword to login to you may registered to una count. To demontration, please use email demonse du and password demo to login. Main account Image account Register an account. Image account Register an account. Register an account. Register and account. Register an account. Register an account. Register an account. Register and account. Register an account. Register an account. Register an account. Register an account. Register an account. Register an account. Register an account. Register analyses. Register analyses. Register analyses. Register analyses. Register anaccount. Reg	Home CRC Analysis	Tutorial	Documentatio	n FAQs	Reference						
W INIVERSITY OF MICHIGAN	Control Register Cont										
UNIVERSITY OF MICHIGAN		LINUTEDO	TY OF MICH	ICAN							
		UNIVERSI	TY OF MICH	IIGAN							
Done 🥘 🕘 🖒 📩 ,	Done				۵ 🔮						

Figure 3.

3. Add a new dataset.

To analyze a new dataset, click on "add a new dataset" as high lighted in Figure 4. The linked page is illustrated in Figure 5.

- 3A. First, enter a name for the new dataset, required as indicated by "*".
- 3B. A description of the dataset can be entered here. Useful information such as source of the data, experimental conditions are examples of entrees here.
- 3C. A sample dataset is provided next. Columns in pink (gene IDs (for example, Affy IDs) and expression profiles) are required, rows and columns in blue (column names in the first row and gene names (or secondary identifier) in second column) are optional. There are two ways of entering data, either by direct copy and paste or uploading a plain text file.

Important notes:

Each column has to be separated by a tab or comma (one or more spaces will not be enough). Missing data is allowed since CRC is able to handle missing data on the fly, no imputation step is needed. The default symbol for missing data is "NA", other symbols are allowed as long as one specifies at the bottom of this page. No space is allowed in the required gene ID and optional gene name (or secondary identifier). In another word, each gene ID or gene name has to be one word.

- 3D. CRCView allows flexible input formats which eliminates annoying reformatting when performing CRC clustering analysis. Here the user is able to specify the key characteristics of her data, and CRCView will automatically adjust the input data format internally. The first two questions relate to whether the optional row and/or column are present. The third question regards whether the data have been log transformed. CRC is performed on log transformed expression levels. An extra log transformation step will be performed if the data entered is not yet log transformed. The last question is the symbol for missing data. The default is "NA", but other symbols entered here can be used instead, for example, common symbols are "-999", "?", "-", etc. User can specify a corresponding microarray annotation data which will be used in the later analysis (probe to gene mapping and GOStats analysis).
- 3E. After data is entered, and format options are specified, click submit button, the data will be uploaded to the CRCView database, and the user will be taken to the CRC analysis setting page.

🖻 CRCView - Mozilla Firefox 📃 🗖 🔀											
<u>File E</u> dit <u>V</u> iew	Hi <u>s</u> tory <u>B</u> ookmarks <u>T</u> ools <u>H</u> elp						$\langle \rangle$				
🗣 • 🔶 • 🕑 🔅	Attp://helab.bioinformatics.r	ned.umich.edu/crcview/	sandbox/data/index.p	hp	• 🕨 🖸	Google	a)				
CRCView											
Home CRC Analysis Tutorial Documentation FAQs Reference											
1				System Admin	<u>My CRC Analysis Cha</u>	nge Password Up	odate Profile Logout				
My Data Sets	My Data Sets										
	Add a new data set				Refresh data set li	<u>st</u>					
Data Set Name	Description	Last Upate Time			Operations						
test1	This is a testing project	2007-03-07 17:44:23	Run CRC analysis	Show CRC results	Show data set detail	Update data set	Delete data set				
test2	test2	2007-02-26 11:38:24	Run CRC analysis	Show CRC results	Show data set detail	Update data set	Delete data set				
Oliver data	Oliver data	2007-03-07 17:43:24	Run CRC analysis	Show CRC results	Show data set detail	Update data set	Delete data set				
test4		2007-03-07 17:43:45	Run CRC analysis	Show CRC results	Show data set detail	Update data set	Delete data set				
test5		2007-03-07 17:44:02	Run CRC analysis	Show CRC results	Show data set detail	Update data set	Delete data set				
CRCView Demo	This a CRCView demonstration project.	2007-03-08 17:59:38	Run CRC analysis	Show CRC results	Show data set detail	Update data set	Delete data set				
Test 8	test	2007-03-07 17:44:12	Run CRC analysis	Show CRC results	Show data set detail	Update data set	Delete data set				
		Acknowledgements	Links Contact	Us Disclaimer							
	W UNIVERSITY OF MICHIGAN										
http://helab.bioinfo	rmatics.med.umich.edu/crcview/sandbox/	data/index.php				🧶 🕚 🖉	> 🙆 🙆 //				

Figure 4.

CRCView

	ne CRC Analysis	s Tutorial	l Doc	umentati	on FA	Qs R	eference	
			Syste	m Admin <u>N</u>	ly CRC An	alysis <u>Ch</u>	ange Passw	ord Update Profile
d a New Data Set								
\longrightarrow	2 ۸							
me of Your Data Set *	JA							
escription:	2 P							
<u> </u>	JD							
iter Data (A. manual er	ntry or B. file uplo	ad):						
Instruction								
Input file cont probeset IDs.	ains gene expression Each subsequent of	on profiles. Ea	ach line sents da	represent ta out of a	s a gene. a hybridize	The first ed micora	column co arrav chip, v	ntains probe or which stores dene
expression lev	vels of individual ge	nes observe	d from th	nat chip (o	r an expe	riment re	plicate). En	tries are separate
or gene name	nma. Missing data i e/symbol. A sample	s represente input file is sl	a by NA	. Please R low:	EMOVE	any space	e character	Inside probeset IL
Probeset_ID	Gene_name/sym	bol Chip 1	Chip 2	Chip 3	Chip 4	Chip 5	Chip 6	Header line, option
YBL021C	HAP3	0.0125	0.0000	-0.0220	0.1572	0.1653	-0.0086	
YBL025W	RRN10	0.0229	-0.3708	-0.1411	0.1329	0.0536	-0.2374	
YBL072C	RPS8A	-0.0406	0.1674	0.1858	-0.4608	-0.4397	-0.0967	
YBL087C	RPL23A	-0.1879	0.2026	0.1636	-0.4881	-0.5087	-0.3331	
YBR055C	PRP6	0.0302	-0.2260	0.0182	0.0985	0.0590	-0.0715	
YBR123C	TFC1	0.0000	-0.1827	-0.0302	0.1222	0.0785	-0.0246	
YBR181C	RPS6B	0.1537	0.1016	0.1014	-0.3700	-0.3972	-0.1922	
YBR188C	NTC20	0.0932	NA	0.0541	0.1394	0.1054	0.0435	
YBR189W	RPS9B	-0.0952	0.1964	0.2025	-0.4473	-0.4843	-0.1850	
A. Enter your data	here: <u>Click here</u> to	get a sampl	e data s	et.				
A. Enter your data	nere: <u>Click here</u> to	get a sampl	e data s	et.				
A. Enter your data	here: <u>Click here</u> to	get a sample	e data se	et.	Brox	vse	>	
A. Enter your data B. Upload a data fi ea & tell us more abo	here: <u>Click here</u> to	get a sample	e data so	et.	Brox	vse	>	
A. Enter your data B. Upload a data fi safe tell us more abo Annotation Data: N	here: <u>Click here</u> to	get a sampli	e data si	et.	Brox	vse	>	
A. Enter your data B. Upload a data fi ea e tell us more abo Annotation bata. N The first line is a hea	here: <u>Click here</u> to	9 get a sample	e data si	et.		vse	>	
A. Enter your data B. Upload a data fi ea e tell us more abo Annotation bata. N The first line is a he The second row has	here: <u>Click here</u> to	9 get a sample 3 9 No 0 No 0 Yes	e data s BC	et.	Brox	vse	>	
A. Enter your data B. Upload a data fi B. Upload a data fi a tell us more abo Annotation Data: N The first line is a he The second row has Data are log transfo	here: <u>Click here</u> to	9 get a sample	e data s BC	et.	Brox	vse	>	
A. Enter your data A. Enter your data B. Upload a data fi eate tell us more abo Annotation Data: N The first line is a he. The second row has Data are log transfo Missing data point d	Ile:	9 get a sample 9 No 10 No 10 Yes 10	e data so BC	et.		vse	>	
A. Enter your data B. Upload a data fi B. Upload a data fi Annotation Data. N The first line is a he The second row has Data are log transfo Missing data point d	here: <u>Click here</u> to	 get a sample No Ols O Yes Io 	e data s BC	et.	Brox	vse	>	
A. Enter your data A. Enter your data B. Upload a data fi ease tell us more abo Annotation Data: N The first line is a he The second row has Data are log transfo Missing data point d	here: <u>Click here</u> to	9 get a sample 9 No 9 No 10 10 10 10 10 10 10 10 10 10	e data s BC	contact Us	Brox 3E	wse	>	
A. Enter your data B. Upload a data fi eate tell us more abo Annotation bata: N The first line is a he The second row has Data are log transfo Missing data point d	here: <u>Click here</u> to	9 NO OIS O YES IO UNIVER	e data si BC	Contact Us	Brow Brow Brow Brow Brow Brow Brow Brow	wee	>	

4. Filtering.

After a new dataset is uploaded, the users can initiate CRC analysis by clicking "Run CRC analysis". A filtering page will show up as illustrated in Figure 6. A user can run filtering based on minimum Coefficient of Variation (CV), or run filtering based on minimum CV, minimum value and minimum fold change.

- 4A. Click "Run filtering" if you wish to run filtering.
- 4B. Click "Go to next step" if you are satisfied with the filtering results.
- 4C. Click "Skip filtering and go to next step" if you wish to don't want to run filtering. All the filtering results will be ignored.

CRCView - Mozilla Firefo	x											
<u>Eile E</u> dit <u>V</u> iew Hi <u>s</u> tory	<u>B</u> ookmarks <u>T</u> ools <u>H</u> elp											1
🄄 🕫 - 🔶 - 🕼 🔪	http://helab.bioinformatics.	med.umich.edu/crcview	/sandbox/	data/filte	r.php?c_	data_fileI	C 🔹 🕨	G-	Google			0
CBCView												
UKUVIEW												
Home CRC Analysis Tutorial Documentation FAQs Reference												
System Admin My CRC Analysis Change Password Update Profile Logo												
Coefficient of Variation Filtering												
If you want to run coeffici	ent of variation filtering, p	lease specify the min	imum coe	efficient	of varia	tion, and	d click "F	Run filte	ring". Yo	ou can tr	y differe	nt
minimum coefficient of va	ariation and run filtering m	ultiple times. Once yo	ou are sat	ticified v	vith the	iltering.	please of	lick "G	o to nex	t step" to	continu	le.
If you don't want to run a	ny filtering, please click "S	kip filtering and go to	o next ste	p".								
Check here if you als	o wish to run filtering base	ed on minimum expre	ssion lev	el and f	old char	ige						
Remove the lower 10	% of the probesets at	d for each probeset	the mini	mum fo	ld chanc	e shoul	d be 15					
Remove the lower 10	% of the probesets at	to for each probeset		muni io	iu chang	e snoui	u be [1.5					
Minimum coefficient of va	ariation: 10 Runt	filtering	A									
Number of probes in vo	ur dataset. 7076.											^
Probe	Gene_name	0	2	4	6	8	10	12	14	16	18	1
A28102_at	A28102_at	0.959	<mark>1.10</mark> 9	-0. <mark>4</mark> 815	-0.3695	-0.855	0.585	-0.575	0.1545	0.267	-0.3315	
AB000114_at	AB000114_at	-0.895	0.356	-0.1075	-1.1725	-0.107	-1.0335	-0.617	-0.2	0.9115	2.07	
AB000115_at	AB000115_at	0.082	0.5545	-0.5795	-0.438	-1.17	-0.0835	2.16	-0.485	0.0815	1.452	
AB000220_at	AB000220_at	1.334	-0.423	-0.158	1.6825	1.1665	0.6365	-0.172	-0.005	0.33	-1.622	
AB000381_at	AB000381_at	-1.202	0.9575	0.6405	-0.7995	0.0355	0.209	1.677	0.8135	-0.5975	-0.223	
AB000409_at	AB000409_at	0.987	-1.437	0.3905	-0. <mark>14</mark> 75	0.0255	-0.629	-1.091	0.083	-0.8985	-1.052	
AB000410_at	AB000410_at	1.037	-0.7205	0.533	1.066	-0.086	0.1585	1.815	-0.144	-0.3025	-2.276	
AB000449_at	AB000449_at	-0.971	-0.877	-0.2395	-0.596	-0.821	-0.5955	-1.459	-0.3895	-0.483	0.155	
AB000450_at	AB000450_at	1.531	-0.2665	-0.3745	-0.123	-0.842	-0.177	-1.13	1.1175	-0.824	-1.921	
AB000460_at	AB000460_at	1.139	-0.6865	-0.5125	1.286	-0.4755	0.9485	0.094	0.8905	-0.2805	0.769	
AB000462_at	AB000462_at	-1.062	-1.119	0.477	-0.207	0.762	-1.4325	0.306	-0.4065	0.876	-0.663	
AB000464_at	AB000464_at	1.412	0.4395	-1.1965	0.3175	-0.08	0.2295	0.44	0.075	-0.2235	-0.202	
<	1 D										>	Ĩ
Go to next step	tD											
OUT OT TO TRACT STORE												
Skip filtering and go to nex	t step D(Angering result	s will be ignored)										
	Ackn	owledgements Link	s Con	tact Lis	Discla	mer						
	Actin	UNIVERSIT	TY OF N	Лісн	IGAN	inci						
		CIVIVERSI		men	IGIAN				15 C		-	
ne									99	0	0 2	2

Figure 6.

5. CRCView Analysis settings.

After a new dataset is uploaded, the users will be taken to the CRC analysis setting page as illustrated in Figure 7. A few parameters need to be specified here.

- 5A. Number of chains. CRC uses MCMC to infer unknown cluster memberships for all genes. Typically, multiple independent Markov chains with different initial states are used in this MCMC schemes. The main benefit of using multiple independent Markov chains is to ensure more thorough exploration of the entire sample space to avoid trapping at local mode. More chains potentially give better result but will take more computing time (linear increase). A good strategy is to start with low number to get quick preliminary results for evaluation, and then use a relatively large number to get the best result possible. The decision should also be made according to computing time. Recommended value is 10, the upper limit allowed is 100.
- 5B. Number of cycles or number of iterations. Within each Markov chain, CRC repeatedly goes through every gene and reassigns each gene's membership. The total number of assignments = number of cycles × total number of genes. More cycles means the Markov chain is more likely to be converged but will be more time consuming. The Bottom line is to ensure the markov chain has converged long before the chain ends. It is recommended that one starts with low number of cycles to get preliminary results, and then increase it later, say, 100, for final result. If one believes the MCMC procedure converges fast (can be seen from the log likelihood trace plot), lower number of cycles is fine. Selection of this number depends on computation time. The upper limit is 1000.
- 5C. Inversion flag. This parameter tells CRC whether to look for inverted correlation pattern when clustering. This maybe of interest when dealing with time course experiments or experiments where genes maybe antagonistically regulated. Default value for this parameter is 0 which indicates

no inversion pattern, alternatively, user can change this to 1 if she believes some inversion correlation patterns are expected.

- 5D. Maximum Shift. This parameter tells CRC whether to look for time-delayed correlation pattern when clustering. This maybe of interest when dealing with time course experiments where genes may display correlation with a time-shift. Default value for this parameter is 0 which indicates no time-shift pattern is allowed, alternatively, user can change this to a positive integer if she believes time-shifted correlation patterns are expected. Only low number such as 1 or 2 is recommended unless time between neighboring time-course experiments are very short. The Upper limit is 5.
- 5E. Probability cut-off. This parameter tells CRC to filter genes shown in a result cluster. Only genes with posterior probability of belonging to this cluster greater than this cut-off value will be retained in the result cluster report. The reason of this filtering step is that high posterior assignment probability indicates better fit of the expression profile with its cluster and is more likely to be biological meaningful. Ability to calculate this probability is an important advantage of model-based clustering methods. Users can use this information to prioritize their clustering results. Default value for this parameter is 0 which indicates no filtering (all genes in a cluster are reported). Users can increase this value such that only genes shown tight correlation pattern within a cluster will be reported.
- 5F. Note. Users can write down comments on their analysis here such as "preliminary". "final", etc. Information in this box is optional.
- 5G. After entering all information in 5A-F, click here to initiate the CRC run.

😻 CRCView - Mozilla Firefox									
Ele Edit View Higtory Bookmarks Iools Help									
🔹 🔶 👻 💿 🏠 🔪 http://helab.bio	informatics.med.umich.edu/crcvi	w/sandbox/data/crcSett	tings.php?c	data 🔹 🕨 🔽 Google	Q				
CRCView									
Home	CRC Analysis Tutorial	Documentation	FAQs	Reference					
CRC Analysis Settings		<u>System Ac</u>	<u>Imin</u> <u>My CF</u>	RC Analysis Change Password Update Pro	<u>ofile Logout</u>				
Your Settings for Current Data Set: R	lefresh Settings								
Sumber of Chains 0 Anteger, nu	umber of parallel chains to ru	n. Recommended val	ue is 10	2 More details					
Number of Cycles 20 5 Beger, n	umber of cycles to run in eac	h Markov chain. Reco	mmended	value is 20. <a>Image: More detail					
Evertion Flag* 1 Cheans use full model complex	e no invert relations, and ign relationships are considere	ore nonsynexpression	relationsh	ips such as inverted or time-shifted. 1	means				
for time course data only)	2,5, 0 means no time-shift	pattern allowed, n>0 i	ndicates t	me-shift of n units. <a>(2) More details					
erobability Cutoff* result. 2000 result	ails	or probability threade	na ior a ge	The to be included in a cluster in the in	itai				
Mote 5F			Your note	about current analysis.					
* Required fileds.									
Tip: For experiments conducted over time inverted and/or time-delayed, may be exp invertion flag and max shift to 1. Otherwis	e, such as cell cycle experim bected, it is recommended th se, fix invertion flag and max	ents, where complex c at the pattern selction shift to 0.	orrelation option of	relationships other than synexpression the program to be turned on. That is, s	1, e.g., set				
	Save Setting	5G Run CRC	Analysis	>					
	Acknowledgements	inks Contact Us	Disclaimer						
	Univer	SITY OF MICHIC	GAN						
Done				19 🔿 🖄	a 👛 ";				
		_							

Figure 7.

6. Monitor CRC run progress.

After CRC run was initiated, a new page will appear as shown in Figure 8 acknowledging that the CRC job has been submitted. During CRC run, one may click the place highlighted to monitor the progress of the job. A sample progress report page is illustrated in Figure 9. It shows how many chains have been completed. This page can be refreshed to get the updated progress by click the highlighted place. When

the job is finished, an automatic email will be sent to you to inform you the analysis is completed, you may click the link provided in the email to directly access the result page.

🕲 CRCView - Mozill	a Firefox									
<u>File E</u> dit <u>V</u> iew H	Hi <u>s</u> tory <u>B</u> o	ookmarks <u>T</u> ools <u>H</u>	<u>t</u> elp				$\langle \rangle$			
🛊 🕶 👻 😨 🚳 🔪 http://helab.bioinformatics.med.umich.edu/crcview/sandbox/data 💌 🕨 💽 Google 🔍										
	CRCView									
	Home	CRC Analysis	Tutorial	Documentation	FAQs	Reference				
			Sy	stem Admin My CRC /	Analysis Ch	ange Password Up	idate Profile Logout			
Run CRC Analy	sis									
A new CRC Analy	sis ioh ha	s been submitted	We will send	d vou an email once	the job is fi	nished				
Click here to chec	k detailed	analysis status/re	sults	a you an onici onco	and job to th					
Go back to my dat	a set list	and you states to	iouno							
Go back to my da	ta sot list									
		Acknowledge	ments Li	nks Contact Us	Disclaimer					
		M	JNIVERS	ITY OF MICHIC	GAN					
Done						🥴 🌒 🚄) 🙆 🙆 "			

Figure 8.



Figure 9.

7. CRCView result.

In CRCView, rich information is provided on the clustering results. All results can be accessed from the result summary page shown in Figure 10.

7A. For each cluster generated by CRCView, basic statistics are provided here which include number of clusters and two types of cluster quality measurements: Log Bayes ratio and average co-

occurrence. Log Bayes ratio is defined as:

Log BR = logP(all genes belong to the same multivariate normal distribution)

-log(each gene belongs to its own multivariate normal distribution)

Higher value indicates that the genes in this cluster are more likely to share their expression profile hence are more likely to be functionally related. Co-occurrence for a pair of genes is defined as the proportion of assignments during the second half of the Markov chain (i.e., last N/2 cycles, N is the number of cycles) in which these two genes are assigned into the same cluster. Average co-occurrence is defined as the average co-occurrences for all pairs of gene in a cluster. This statistic measures the stability of the cluster. Values close to 1 means that genes within this cluster are always grouped together which indicates the cluster is very stable. Lower values indicate genes frequently move in and out of this cluster which indicates poor stability of the cluster.

- 7B. For each cluster generated by CRCView, a thumbnail trace plot of the expression profiles is shown here to allow users to quickly assess the basic characteristics of the cluster, such that more detailed display or re-analysis can be planned afterwards.
- 7C. CRCView also provides the traditional summary file that listed all member genes in a cluster along with the posterior probabilities of their assignment. This is the main result file generated by CRC, and can be obtained by clicking the highlighted area next to output file. A sample output file can be found in Figure 11. As one can see, for each member gene in a cluster identified by its gene ID, a number in parenthesis indicates the correlation pattern. "+0" means positive correlation with no time-shift, "-1" means inverted correlation with one time point shift when its expression profile is compared with that of the cluster. The value inside the square bracket indicates the posterior assignment probabilities. The higher the probability, it is believed that more likely this gene belong to this cluster.
- 7D. A cluster member file can be obtained by clicking the link here. A sample cluster member file is shown in Figure 12. This is a tab-delimitated plain text file. Each line represents a gene. The first number indicates the cluster ID, followed by the order of the member gene in the original dataset, gene ID and a number indicates its correlation relationships with the cluster expression pattern. This number ranges from 1 to 4, indicating positive correlation with no time-shift (1), negative correlation with no time-shift (2), positive correlation with time-shift (3) and negative correlation with time-shift (4). This file maybe useful for getting summary statistics or additional graphical outputs using software such as R.
- 7E. A trace plot of the log likelihoods during the CRC iterations can be obtained from the link shown here. This is to monitor the convergence of the Markov chains. A fast increasing, then stabilized pattern as shown in Figure 13 is what to be expected. Different Markov chain is illustrated with lines with different color. Multiple lines converge to approximately the same value is an indicator of likely convergence.



Figure 10.



Figure 11.

₩02	ina Firelox							JL
<u>F</u> ile <u>E</u>	dit <u>V</u> iew	History Bookmark	s <u>T</u> ools	<u>H</u> elp				
-	🔶 - 🕑 🗯	3 🚮 🔪 http://H	relab.bioin	formatics.med.um	nich.edu/crcv 🔻	G.	Google	
1	190	RPS12 1	>					
1	192	RFL21B 1						
1	193	RPS9A 1						
1	194	RPS6A 1						
1	195	RPL33A 1						
1	196	RPL7B 1						
1	198	RPL1A 1						
1	199	RPL43A 1						
1	202	RPS23B 1						
1	4	RPS8A 1						
1	5	RPL23A 1						
1	8	RPS6B 1						
1	10	RPS9B 1						
1	11	RPL21A 1						
1	20	RPS29B 1						
1	21	RPL31A 1						
1	22	RPL13A 1						
1	23	RPS16B 1						
1	24	RPL35B 1						
1	25	RPL35A 1						
1	27	RPS13 1						
1	33	YDR341C 1						
1	38	RPS18A 1						
1	39	RPL27B 1						
1	41	RPL37B 1						
1	47	RPS24A 1						
1	48	RPS8B 1						
								3

Figure 12.



8. More CRCView display options.

In CRCView, the user have many options to customize the display of clusters they are interested. These options are highlighted in Figure 14.

- 8A. Selecting clusters for more display or analysis options. One may want to concentrate on only a subset of clusters for further analysis. CRCView allows users to display or perform analysis on any subset of clusters. To select, one can either click "select all" to include all clusters, or click on the "select" box above each cluster's thumbnail trace plot.
- 8B. CRCView also allows users to rearrange the order of displayed clusters according to several criteria. This can be done by select criteria available at the order cluster by pull down menu at the top. The four available criteria to choose from are "Cluster ID", "Number of genes", "Log Bayes ratio" and "Average co-occurrence". The cluster ID is the original order. The users can choose to display them in ascending or descending orders according to these criteria.
- 8C. After clusters of interest have been chosen; the user can click the "download selected cluster images as one zip file" to export all graphical outputs for publication or further analysis.

- 8D. CRCView provides heatmap (aka Eisen plot) for selected clusters. This can be achieved simply by clicking the "show heatmap of selected cluster" button. A sample heatmap is shown in Figure 15.
- 8E. To get a list of genes in each cluster, click the "show gene list of selected cluster" button. A sample gene list is shown in Figure 16. It is a plain text file, each line contains a gene name, different clusters are separated by an empty line.
- 8F. CRCView allows user to generate trace plots of selected clusters. This can be achieved by click the "show selected cluster in one plot" button. The users can even arrange the format of the plot by specifying number of cluster per row in the nearby pull down menu. A sample plot is shown in Figure 17.



Figure 14.



Figure 15.



Figure 16.



Figure 17.

9. GOStats analysis.

An important step in clustering analysis is validating/interpreting the clustering results. CRCView provides automatic GO Term enrichment analysis for each cluster the user selected. Figures 17 and 18 illustrated how such type of analysis can be performed.

- 9A. The first step in GO term enrichment analysis is to select an appropriate annotation data file. This can be done from choosing among the available files in the "Annotation Data" pull down menu. These files are collected from Bioconductor package. Currently there are 75 available, including Human, mouse, rat.
- 9B. Cut-off value: this is the significance level for the hypergeometric test.
- 9C. Users can choose from the three GO categories: Molecular function, Biological process and Cellular Compartment. A sample GO term enrichment analysis result is shown in Figure 20.

🕲 CRCView - Mozilla Firefox									
<u>File Edit View History Bookmarks To</u>	ols <u>H</u> elp								
🖤 🔻 🐨 🥥 🖬 🔪 Trup://rieBo.bioinformatics.med.umicn.edu/crcview/sandbox/data/snowkesuts.pnp/C_D 💌 🌶 💽 Google									
CRCView									
Home	CRC Analysis	Tutorial D	ocumentation	FAQs Ref	erence				
			System Adm	in My CRC Analy	sis Change Password Update Profile Logour				
CRC Analysis Results									
Analysis ID: 124 number of clusters	returned: 4								
Select all Unselect all Order clusi	ters by Cluster ID	~	Ascending 💌	Update					
Cluster 1 Select Clu	uster 2 🗹 Select		Cluster 3 Sel	lect	Cluster 4 🗹 Select				
number of genes = 83 nul	mber of genes = 94	1	number of genes	= 14	number of genes = 14				
average co-occurrences = 1 ave	erage co-occurrent	ces = 0.99461	average co-occur	rrences = 1	average co-occurrences = 1				
Cuator 1	Date 1			atu 1	Date 4				
1	A. M	A			AAAA				
Arabidopsis ag: Affymetrix Arabidop	psis Genome Array Annot	ation Data (ag)	D	and a					
C. elegans	elegans Genome Array	Annotation Data (ce	elegans)	#					
Chicken chicken: Affymetrix ch	iken Annotation Data		oggene)		les .				
Operations: (Pleas Drosophila drosgenome 1: Affymet	trix Drosophila Genome A	may Annotation Dat	a (drosgenome 1)						
Show heatmap drosophila2: Affymetrix indac: INDAC FlyChip	[Drosophila Genome 2.0]ong_oligonucleotide_0	2 (FL002) Annotation D	ata (drosophila2) on Data (indac)	if sele	ected clusters				
Number of clusters h10kcod: CodeLink U h20kcod: CodeLink U	niSet Human I Bioarray (IniSet Human 20k I Bioar	~10 000 human ger ray Annotation Data	nes) Annotation Data						
hcg110: Affymetrix Hu hgfocus: Affymetrix Hu	man Cancer G110 Array Iman Genome Focus Arra	Annotation Data (ho ay Annotation Data	cg110) (hgfocus)	line li	testilesed				
hgu 133a: Affymetrix H hgu 133a: Affymetrix H	Human Genome U133 Set Human Genome U133A : Human Genome U133 Set	2.0 Array Annotation Apportation Data (n	igu 133a) 1 Data (hgu 133a2) 1 gu 133b)	login	Keiniood				
GOstats Analsyis: hgu95a: Affymetrix Hu	rix Human Genome U133 man Genome U95 Set A	Plus 2.0 Array Ann notation Data (hgu	otation Data (hgu133plus 195a)	s2)					
Annotation library I N/A	numan Genome USS Sei	Annotation Date (*	gu95av2)						
Cutoff: 0.05 GO Hierarch	Nigg: Moleculor Fund	tion 🔢	Run GC	stats of selected	clusters				
Go back to my data set list									
Go pacivito my data set list									
	Acknowledge	ments Links	Contact Us E	Disclaimer					
	N	JNIVERSIT	Y OF MICHIG	AN					
Done					: 么 🜭 🌒				

Figure 18.



Figure 19.

🕹 CRCView - I	Mozilla Firefo	x									
<u>File E</u> dit <u>V</u> ie	ew Hi <u>s</u> tory	<u>B</u> ookmarks <u>1</u>	ools <u>H</u> elp)				5° 5			
🤹 • 📦 • 🧿	F 🖂 🟠 🖪	http://helab	.bioinforma	tics.med.u	imich.edu/crcview/sa	ndbox 🔻 🕨 💽 🗸 Go	pogle	Q			
CRCView											
	Home CPC Analysis Tutorial Documentation EAOs Reference										
System Admin I My CRC Analysis I Change Password I Update Profile I Logout											
GOstats Ar	nalysis										
Cluster 7	• number of as	000 - 26 Ioa D	ovec rotio -	44 4006		000 - 0 E1E041					
gene list: M64 U20758_ma1 X56841_at, X6 HG4157-HT44 GO Hi	4930_at, M875 _at, U33017_at 0484_at, X809 27_at, J00209 erarchies: I	07_at, M93221 , U39573_at, U 23_at, X82068_ _f_at, L01664_ Biological Pr	_at, S69115 47686_at, I _at, D25539 at, L05515_ TOCESS	5_at, S769 J70321_a J_at, D283 _at, L0560	92_at, S78873_at, U t, U72515_at, X0519 83_at, D83597_at, H 6_at, L12468_at, L13	01337_at, U08316_at, U 6_at, X05246_at, X51698 G2260-HT2349_at, HG3: 258_at, L14430_at	18422_at, ⊾at, X51954_at, 31-HT331_at,				
GO T	erm	Description	6	n.values	Number significant	Gene Detail	1				
GO:00	06006 alucas	e metabolism		0.007	3	Show gene information					
GO:00	19320 hexose	catabolism		0.008	2	Show gene information					
GO:00	16052 carboh	vdrate catabolis	sm	0.008	2	Show gene information					
GO:00	06096 alvcolv	sis		0.008	2	Show gene information					
GO:00	06007 alucos	e catabolism		0.008	2	Show gene information					
GO:00	46365 monos	accharide cata	bolism	0.008	2	Show gene information					
GO:00	46164 alcoho	catabolism		0.008	2	Show gene information					
GO:00	44275 cellular	carbohydrate o	atabolism	0.008	2	Show gene information					
	GO:0006006						1				
	Gene Symbol	NCBI Gene ID		Probe S	iet						
	UGP2	<u>7360</u>	L14430_a	t, U00954_	_at, U27460_at						
	ALDOC	230	X05196_a	t							
	PGK2	<u>5232</u>	X05246_a	t							
	GO:0019320										
	Gene Symbol	NCBI Gene ID	Probe Set								
	ALDOC	<u>230</u>	X05196_a	t							
	PGK2	<u>5232</u>	X05246_a	t							
	GO:0016052		20								
	Gene Symbol	NCBI Gene ID	Probe Set								
	ALDOC	230	X05196_a	t							
	PGK2	<u>5232</u>	X05246_a	t							
	1			-				~			
Done						🥴 🅚		<u></u>			

Figure 20.

Reference

Qin ZS. (2006) Clustering microarray gene expression data using weighted Chinese restaurant process. *Bioinformatics* **22**(16):1988-1997.

Contact

The CRC system has been created and maintained at the University of Michigan. Please contact us if you have any questions or comments:

Zuoshuang Xiang, Unit for Laboratory Animal Medicine, Email: zxiang@umich.edu, Phone: (734) 615-2455.

Dr. Steve Qin, Department of Biostatistics, Email: qin@umich.edu, Phone: (734) 763-5965.

Dr. Yongqun He, Unit for Laboratory Animal Medicine, Department of Microbiology and Immunology, Bioinformatics Program, Email: yongqunh@med.umich.edu, Phone: (734) 615-8231.