

Protocol

OMiCC: An expanded and enhanced platform for meta-analysis of public gene expression data



OMiCC (OMics Compendia Commons) is a biologist-friendly web platform that facilitates data reuse and integration. Users can search over 40,000 publicly available gene expression studies, annotate and curate samples, and perform meta-analysis. Since the initial publication, we have incorporated RNA-seq datasets, compendia sharing, RESTful API support, and an additional meta-analysis method based on random effects. Here, we provide a step-by-step guide for using OMiCC.

Publisher's note: Undertaking any experimental protocol requires adherence to local institutional guidelines for laboratory safety and ethics.

Candace C. Liu, Yongjian Guo, Kiera L. Vrindten, William W. Lau, Rachel Sparks, John S. Tsang

cliu72@stanford.edu (C.C.L.) john.tsang@nih.gov (J.S.T.)

Highlights

OMiCC (OMics Compendia Commons) is a free web-based tool for gene expression data reuse

Search publicly available studies to perform sample group comparisons to explore a disease

In meta-analysis, multiple studies are combined to identify coherent signals

OMiCC supports crowd-sharing and users can share their own analyses with the community

Liu et al., STAR Protocols 3, 101474 September 16, 2022 © 2022 https://doi.org/10.1016/ j.xpro.2022.101474

Protocol



OMiCC: An expanded and enhanced platform for meta-analysis of public gene expression data

Candace C. Liu,^{1,3,4,*} Yongjian Guo,^{1,4} Kiera L. Vrindten,¹ William W. Lau,^{1,6} Rachel Sparks,^{1,5} and John S. Tsang^{1,2,5,6,7,*}

¹Multiscale Systems Biology Section, Laboratory of Immune System Biology, National Institute of Allergy and Infectious Diseases, National Institutes of Health, Bethesda, MD 20892, USA

²NIH Center for Human Immunology, National Institute of Allergy and Infectious Diseases, National Institutes of Health, Bethesda, MD 20892, USA

³Present address: Immunology Graduate Program, School of Medicine, Stanford University, Stanford, CA 94305, USA

⁴These authors contributed equally

 $^5\mathrm{These}$ authors contributed equally

⁶Technical contact: lauwill@mail.nih.gov

⁷Lead contact

*Correspondence: cliu72@stanford.edu (C.C.L.), john.tsang@nih.gov (J.S.T.) https://doi.org/10.1016/j.xpro.2022.101474

SUMMARY

OMiCC (OMics Compendia Commons) is a biologist-friendly web platform that facilitates data reuse and integration. Users can search over 40,000 publicly available gene expression studies, annotate and curate samples, and perform meta-analysis. Since the initial publication, we have incorporated RNA-seq datasets, compendia sharing, RESTful API support, and an additional metaanalysis method based on random effects. Here, we provide a step-by-step guide for using OMiCC.

For complete details on the use and execution of this protocol, please refer to Shah et al. (2016).

BEFORE YOU BEGIN

Overview

Advances in microarray and RNA sequencing technologies have led to a rapid increase in the amount of gene expression data deposited in public data repositories such as the Gene Expression Omnibus (GEO) (Barrett et al., 2013) and Array Express (Rustici et al., 2013). Integration and meta-analysis of data across studies is a powerful tool to incorporate heterogeneity into analysis and increase the statistical power for generating and testing hypotheses (Andres-Terre et al., 2015; Chaussabel and Baldwin, 2014; Chen et al., 2014; Dudley et al., 2011; Granlund et al., 2013; Khatri et al., 2013; Segal et al., 2005; Sirota et al., 2011; Sweeney et al., 2015; Teslovich et al., 2010). The ever-growing volume of publicly available data enables such meta-analysis, but much of the data remains under-utilized (Haynes et al., 2017; Rung and Brazma, 2013), partly because integrating data across multiple studies is not trivial and requires the user to search databases for relevant studies, manually annotate samples, transform data into different formats, normalize the data, and finally perform meta-analysis (Ramasamy et al., 2008; Rung and Brazma, 2013). Significant statistical and computational experience is required for many of these steps, which may dissuade biologists with less computational training to perform such analyses.

To bridge the gap between the volumes of publicly available data and the integration of this data to generate new biological insights, we developed a biologist-friendly web platform for data reuse and





meta-analysis. OMiCC (OMics Compendia Commons) is a freely available tool that enables biologists with little computational training to search existing data sets easily, annotate and curate data, and perform differential expression analysis and meta-analysis (Shah et al., 2016). Users can add multiple publicly available datasets to a compendium, a collection of studies focused on the biological question of interest, and on which differential analysis and meta-analysis can be performed. To date, OMiCC contains pre-processed datasets from more than 40,000 studies from GEO and the Sequencing Read Archive (SRA) (Leinonen et al., 2011). OMiCC is continually supported and updated with new studies deposited in GEO and SRA. Since the original publication (Shah et al., 2016), we have added multiple new features, including incorporating RNA-seq data, sharing of compendia, RESTful API support, and an additional meta-analysis method. In this paper, we provide users with a clear, step-by-step workflow for performing a complete analysis in OMiCC, from study selection to meta-analysis, as well as walk users through the new features of OMiCC.

The metadata available in large public repositories is often not standardized, and annotating samples (i.e., assigning labels such as disease state, gender, age) and collating datasets (i.e., creating control and treatment groups) is a time-consuming step (Ramasamy et al., 2008; Rung and Brazma, 2013). To streamline this process and enable the generation of gene expression signatures based on two-group comparisons, OMiCC supports creating sample groups and comparison group pairs (CGPs) using a simple point-and-click interface. A CGP contains two sample groups, for example, a control group and a treatment group from the same experiment/study, on which differential gene expression analysis can be performed. CGPs can be added to a compendium, which represents a group of CGPs relevant to a biological question. Users can add annotations to sample groups specifying perturbation, disease, sample, or source type. Users can then perform two types of analysis, differential expression analysis within a study and meta-analysis across studies, using an easy-to-use web interface. Users are able to retrieve the analyzed data directly from OMiCC.

Moreover, an important aspect of OMiCC is its crowdsourcing feature. In OMiCC, annotations and curated data sets created by users are stored and can be made available to other OMiCC users, who can integrate and use these data sets for their own analyses. One of the new features allows sharing an entire compendium containing data, structured and reusable annotations, and analyses integrated across multiple datasets. Users are also able to provide professional information on their profile, such as a link to a professional or LinkedIn page; this allows others to determine if a user shares biological interests or level of expertise. Thus, OMiCC provides the broad biomedical research community with the capacity to participate in community-wide collaborations. In 2016, we conducted a crowdsourcing "jamboree" exercise within the National Institutes of Health, where groups were tasked with using OMiCC to assess transcriptomic signatures of several autoimmune diseases (Lau et al., 2016; Sparks et al., 2016). We reported encouraging findings, providing evidence that OMiCC can facilitate and accelerate the pace by which publicly available data can be used to generate new biological insights.

While we provide evidence that biologists with little computational experience can use OMiCC to perform meta-analysis, wider adoption requires biologists to invest the time to familiarize themselves with the platform and its features. As the OMiCC user community grows, more users will create and share their datasets. This protocol provides an important resource towards achieving wider adoption by providing a clear guide and walking users through a complete workflow.

Comparison with other methods

While there are other useful resources that can be applied for data reanalysis and meta-analysis, it takes considerable programming to connect these tools in a complete workflow (Ramasamy et al., 2008). Microarray Retriever searches and retrieves data from GEO and ArrayExpress (Ivliev et al., 2008). ProfileChaser (Engreitz et al., 2011) and ExpressionBlast (Zinman et al., 2013) are web tools that take a gene expression profile as input and query GEO studies by similarity to the provided data. NetworkAnalyst (Xia et al., 2015) takes a list of genes or proteins or gene expression data as



input and performs meta-analysis and network visualization. GenePattern (Kuehn et al., 2008) provides a web interface for a broad array of computational tools for analyzing gene expression data. These tools focus on one or a subset of the steps in the complete workflow, while OMiCC provides a simple user-interface for the entire workflow from searching for samples to performing metaanalysis.

A similar tool, Crowd Extracted Expression of Differential Signatures (CREEDS), is a web resource that contains disease signatures that were annotated and analyzed through a crowdsourcing exercise (Wang et al., 2016). Through an online course on Coursera, participants developed over 2,000 single-gene perturbation signatures, over 800 disease signatures, and over 900 drug perturbation signatures. This further demonstrates the value of crowdsourcing in meta-analyzing existing data sets. While CREEDS contains thousands of previously curated signatures, OMiCC allows users to develop their own signatures by creating their own CGPs and compendia. Therefore, OMiCC facilitates the creation of CREEDS-like signatures without any programming. Furthermore, communities can be built using the crowdsourcing features in OMiCC, as evidenced by our jamboree event (Lau et al., 2016).

Terminology used in OMiCC

- Sample group: A collection of samples from a single study. Two sample groups are required to create a comparison group pair. For example, one sample group can be made up of samples before an influenza challenge, while the second sample group can be made up of samples collected 12 h after an influenza challenge. Sample groups must be annotated using at least two MeSH (Medical Subject Headings) terms. MeSH is a controlled vocabulary used to index and search biomedical databases such as MEDLINE/PubMed (https://www.nlm.nih.gov/mesh/ meshhome.html).
- 2. Comparison group pair (CGP): Composed of two sample groups from the same study that a user wants to compare. To compare the changes in gene expression after an influenza challenge, a user can add the sample groups described above to a CGP. Multiple CGPs can be made from a single study. For example, if one wants to compare gene expression between males and females, a user can create a CGP from the study mentioned above with a healthy male sample group and a healthy female sample group.
- 3. Compendium: A collection of CGPs. Users can add CGPs from multiple studies to a compendium.
- 4. Differential expression profile (DEPs): Gene expression differences of all genes between the two sample groups in a CGP.
- Meta-analysis: Method for extracting statistically coherent signals from multiple CGPs, even when CGPs are from different studies or generated using different technology platforms. OMiCC provides two methods for meta-analysis, RankProd (Hong et al., 2006) and MetaIntegrator (Haynes et al., 2017).

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Software and algorithms		
OMiCC (OMics Compendia Commons)	(Shah et al., 2016)	OMiCC URL: https://omicc.niaid.nih.gov/
recount2	(Collado-Torres et al., 2017)	https://jhubiostatistics.shinyapps.io/recount/
RankProd	(Hong et al., 2006)	https://bioconductor.org/packages/RankProd/
MetaIntegrator	(Haynes et al., 2017)	https://CRAN.R-project.org/package= MetaIntegrator
Deposited data		
Gene Expression Omnibus (GEO)	(Barrett et al., 2013)	http://www.ncbi.nlm.nih.gov/geo/
Sequencing Read Archive (SRA)	(Leinonen et al., 2011)	https://www.ncbi.nlm.nih.gov/sra





MATERIALS AND EQUIPMENT

To use OMiCC, all required is a computer with an internet connection. We recommend Chrome or Safari internet browsers (tested on Chrome version 98.0.4758.80 and Safari version 14.1.2).

STEP-BY-STEP METHOD DETAILS

Herein we provide users with a clear, step-by-step workflow for performing a complete analysis in OMiCC, from study selection to meta-analysis, as well as walk users through the new features of OMiCC.

Register

© Timing: 5 min

This step describes how to register for an account on OMiCC.

- To utilize the full functionality of the website, register for a free account with OMiCC.
 a. Go to https://omicc.niaid.nih.gov/ and click the "Register" button.
- 2. Enter an account name, password, name, and email.
- a. Users enter information about their area of expertise and organization, and can link to a professional home page, PubMed search term, LinkedIn ID, or ResearchGate ID.
- 3. An email is sent to the address provided to complete the registration process.

Search for studies and existing CGPs or compendia using the "search" tab

© Timing: 15 min

This step describes how to search for gene expression studies on OMiCC, and how to search for existing CGPs and compendia created by other users.

Under the "Search" tab/menu at the top:

- 4. Select "On Study" to search for GEO or SRA studies using keywords extracted from the study and filtering by subject or technology platform (Figure 1).
 - a. The search can be restricted to human or mouse only, or to specific platforms.
 - b. Select "Studies with public Comparison Group Pairs" to search studies with at least one publicly available CGP.
 - c. After applying a filter, click "Search".
 - d. To get a list of all studies in the OMiCC database, click "Search" without any keywords or click "Browse All".
 - e. Expand the study by clicking on the "+" icon to view study details.
 - f. Click the icon under the "Study" column to be linked to the study on the GEO website to view experimental details, study design and contributors, and any available citations.
 - g. Save studies by clicking on "Save to my study list" to return to these studies later.
- 5. Select "On Sample Groups" to search for sample groups by using keywords extracted from sample group annotations or metadata associated with the sample group, such as the user ID of the owner.
 - a. Select the "Public Groups" option to show only sample groups that an owner makes public.
 - b. Click "Search" after applying a filter.
- Select "On Comparison Group Pairs" to search for comparison group pairs (CGPs) by using keywords extracted from the CGP title, the user ID of the owner, or the MeSH terms that are used to annotate the sample groups.
 - a. Select "Public CGPs" to show only CGPs that the owner makes public.
 - b. Click "Search" after applying a filter.

Protocol



Mics Comp	endia Cor		ment About -	v	Velcome JDoe [My Profile]
Search Stu Keywords And	All Fields	¢ lupus	- +	Search [Simple Search]	Browse All
Studies with public Comparison Group Pairs	?		Add or remove search fields	Click here to switc to Simple Search	h back mode.
Filter on Platform	All Human Platf	orms (103) + Click "+ orms (47) + → list of p	" to show platforms.		
Show 25 Showing 1 to 25 of 149 e	Note: Only the	e counts on the popular platform. es	s are displayed - please see the Tu	torial for details.	my study list Take a Tour
Show 25 Showing 1 to 25 of 149 e	Note: Only the entries	e counts on the popular platform es Study 1	s are displayed - please see the Tu it le	torial for details. Save to	my study list Take a Tour
Show 25 Showing 1 to 25 of 149 e 2 2 Study 2 - GSE301	Note: Only the only th	e counts on the popular platform es Study 1	s are displayed - please see the Tu ittle	torial for details. Save to Summary Systemic lupus erythema	my study list Take a Tour

Figure 1. Search for studies

Users can search for studies using keywords on all fields or restricted data fields. The search results are presented in a list format, while detailed information can be displayed without leaving the search result list.

- c. Select CGPs and click the "Add to Compendium" button to add them to an existing compendium.
- 7. Select "On Compendia" to search for compendia using keywords extracted from the compendia name, description, or user ID of the owner.
 - a. Select "Public Compendia Only" to search on compendia that the owner makes public.
 - b. Click "Search" after applying a filter.
 - c. To add CGPs to your own compendium, click the "+" button in front of the compendium name, then check the specific CGPs to add.
 - d. To add all the CGPs in a compendium, select the associated checkbox to the left of the compendium name and click the green "Add to the Compendium" button.

Create sample groups

© Timing: 30 min

This step describes how to create sample groups from publicly available gene expression data.



OMic	s Cor	npendia Com	nmons (OM	licc)	
Home	Search	My Study Lists	My Compendia	Document - About -	Welcome JDoe [My Profile]
		Make Sample 0	Groups Make	Comparison Group Pairs 🔪 Add	Comparison Group Pairs to Compendium
Stu	dy: G	SE50772 🏧	o Source		
Availa Note	ble data set About using	s): RMA Normalized Data , the Normalized GEO Data	GEO Data (i.e., 'Series +	s Matrix File' from GEO), Normalized (GEO Data (quantile normalized version of the GEO Data)
	Title: Expres	ssion data in PBMCs from S	SI E patients and contro	als	
Sumn	hary: Periph	eral blood mononuclear ce	lls were collected from	SLE patients in an observational study	r performed at the University of Michigan Blood microar [View Detail]
Platf	orm: GPL57	70: [HG-U133_Plus_2] Affyn	netrix Human Genome	U133 Plus 2.0 Array	Evicting comple groups
Pro	bes: 41025 nes: 20011				
Comp				Create a Sam	ple Group Next > Sample Croups
Samp	ies t.comply	as to be grouped	Filter for a	amples .	Sample Groups
Minim	um 3 re	equired.	of inte	rest.	Group Name (Sample Count) 2
1 2	?	▲ Sample ID ?	Title	Source	GSE50772-control-PBMC (20) Detail
	+	GSM1228860 C	SAM607438 Control	PBMCs	GSE50772-SLE (59) Detail
0	+	GSM1228861 C	SAM607439 Control	PBMCs	GSE50772-SLE-PBMC (59) Detail
	+	GSM1228862 C	SAM607440 Control	PBMCs	PBMC IFN SLE (61) Detail
	+	GSM1228863 🖻	SAM607441 SLE	PBMCs	PBMC SLE IFN control (20) Detail
	+	GSM1228864 C	SAM607442 SLE	PBMCs	
	+	GSM1228865 🖻	SAM608074 SLE	PBMCs	
	ч				
See ad	ditional	sample			
in	formatio	on.			
New Sa	ample (Group			
					Conv Tage Paste Tag
G	roup Name	: GSE50772-SLE1	Update 🔿 Chai	nge group name.	
	Is Public	Yes Change			↓ Copy/paste annotat
A	sage Stats	: + : Perturbation: None	×		from one group to
		Time with perturbati	on:		another.
		Sample type(eg: Mor	nocytes): PBMC 🗶	Enter an annotation	and click "Add
		Source(cell or tissue	type; eg: PBMC): P	BMC X Annotations." Add a	at least 2 tags.
Add New A	nnotations	:			
		Perturbation:	Time	ur	
		Disease:	Samp	le type:	
		Source:	Other	:	
			Add Annotations		
Membe	er Samples	Sample ID		Title	Source
		GSM1228863 ¥		SAM607441 SLE	PBMCs
		GSM1228864 ¥		SAM607442 SLE	PBMCs
		GSM1228865 ¥		SAM608074 SLE	PBMCs

Figure 2. Create sample groups

(A) Operations on a specific study can be followed with the green and blue arrows at the top of the page. To view sample metadata, users can click the "+" and select samples to add to a Sample Group. Filter for samples of interest by inputting key words. Select the samples to include (a minimum of



Figure 2. Continued

3 samples) in a sample group by checking the boxes. Once created, the sample group name appears in the sample group list alongside previously created sample groups created by other OMiCC users. (B) Upon clicking the "Create a Sample Group" button, a pop-up panel appears where the user can add annotations (a minimum of 2 annotation tags) to the selected samples. To copy the annotation tags to another group, click the "Copy Tags" button, then the "Paste Tags" button in the other group. You can change the group name by clicking the "Update" button. After saving the Sample Group, the user can proceed to the "Make Comparison Group Pairs" step by clicking the "Next" button.

- 8. From the study search results, click on the Study ID to access the study page.
- 9. To create your own sample group, select samples to form a group (i.e., cases or controls).
 - a. Use the filter to identify samples of interest.
 - b. For example, if you are interested in samples from systemic lupus erythematosus (SLE) patients, enter "SLE" into the filter to display only those samples.
 - c. At least 3 samples are required for each sample group (Figure 2A).
- 10. After selecting the samples, click the "Create a Sample Group" button. A pop-up window appears for the new sample group (Figure 2B).
- 11. Enter at least two annotations for the sample group.
 - a. We provide a controlled vocabulary (MeSH) to annotate sample groups.
 - b. MeSH annotation suggestions appear as you type into the annotation box.
 - c. While any text is allowed, we encourage users to use the MeSH terms.
 - d. Click the "Add Annotations" button after entering in the annotations.
- 12. The sample group name generates automatically based on the annotations.
 - a. To change the sample group name, click the "Update" button, type in the desired group name, then click "Save".
- 13. To remove unwanted annotations, click the "X" next to the annotation term.
- 14. To copy the annotation tags to another group, click the "Copy Tags" button, then the "Paste Tags" button in the other group.
- 15. To make the sample group public, click the "Change" button next to the "Is Public" field.
 - a. Once a sample group is saved and public, the sample group cannot be edited.
 - b. If the sample group is not being used by any other OMiCC user, the owner of the group can change the status back to private and edit it.
- 16. Click "Save" to create the sample group.
 - a. If at least 3 samples are not added to the sample group, the "Save" button is opaque and cannot be clicked.
- 17. If other users create sample groups using the study, the sample groups show up in bold in a pane on the right side of the browser.
 - a. Click the purple "Detail" button to view annotations and sample details.
- 18. At least two sample groups are needed to create a CGP. Once at least two sample groups are listed under the "Sample Group" panel on the right side, click the "Next" button.

Create comparison group pair (CGP)

© Timing: 15 min

This step describes how to create comparison group pairs (CGPs) using the previously created sample groups.

- 19. Choose two sample groups to add to the CGP.
 - a. Select a sample group from the list of existing sample groups, then click on the "Add to Comparison Group Pair" button.
 - b. Assign the selected sample group as "Condition 1" or "Condition 2 (reference)" (Figure 3A).
- 20. After both conditions are assigned, a pop-up window appears (Figure 3B).
 - a. By convention, OMiCC treats the "Condition 2" sample group as the reference group for downstream differential expression analysis. A positive change in gene expression means



	Home Sea	rch 🛪 🛛 My Stu	dy Lists	My Compendia	Document -	About -		Welcome JD	oe [My Profile]
		Mak	e Sample Gr	pups Mak	e Comparison Gro	oup Pairs	Add Comparison Gro	oup Pairs to Compendium	
	Study:	GSE507	72 Link to	o Source					
	Available data	set(s): RMA Norn Ising the Normalize	nalized Data ed GEO Data	a, GEO Data (i.e., ' ╋	Series Matrix File'	from GEO), No	rmalized GEO Data (qı	uantile normalized version of t	he GEO Data)
	Title: E Summary: P D PubMed: 2 Platform: G Probes: 4 Genes: 20	pression data in P rripheral blood mor ttail] 861459 2L570: [HG-U133_ 025 0011	BMCs from \$ nonuclear ce Plus_2] Affyr	SLE patients and c Ils were collected netrix Human Gen	ontrols from SLE patients ome U133 Plus 2.	in an observati 0 Array	onal study performed a	t the University of Michigan Bl	ood microar [V
	Sample Grou	ips							
		Select samp	le group	to be add	ed as		Previous <	Add to Comparison Group	o Pair 🔹 🛛 Ne
	2 Group Na		or Con	altion 2 in	a CGP.	Tags 2		As Condition 1	:e)
	GSE5077	2-control-PBMC	Detail		20	control PE	BMC		
	GSE50772	2-Healthy1 Detail			20	None Hea	Ithy PBMC PBMC		
	GSE5077	2-SLE Detail			59	none SLE	PBMC PBMC		
	GSE5077	-SLE-PBMC Deta	il		59	SLE PBMC	3		
	GSE5077	2-SLE1 Detail			59	None SLE	PBMC PBMC		
		SLE Detail			61	SLE PBMC			
	Tag Categories Perturbatio	: n	Time with perturbation		Disease	Sam	ple type	Source	Other
	mparison	Group Pair	r (CGP)						
	CGP Nam Descriptio Is Publi Owne	e: GSE50772-S n: c: Yes cr	BLE1::GSE50)772-Healthy1				Remove Pair Inc	dex Delete Gro
	Usage Stat	s: +							
d	ition 1 Group: G	SE50772-SLE1				Condition	2 Group: GSE50772-F	lealthy1	
	▲ Sample ID	Title		¢	Pair Index	Pair Inde	ex ▲ Sample ID	▲ Title	
	GSM1228863	SAM6074	41 SLE		0	٢	GSM1228860	SAM607438 Control	
	GSM1228864	SAM6074	42 SLE			٢	GSM1228861	SAM607439 Control	

to indicate paired samples.

Figure 3. Create a comparison group pair (CGP)

(A) Users can select a sample group and add it either "As Condition 1" or "As Condition 2 (Reference)".

(B) Upon adding sample groups to both Condition 1 and Condition 2, a pop-up panel appears where the user can modify the CGP. If samples are paired, the user can add indices to indicate which samples are paired, which is used later in the analysis.



that the transcript level is higher in Condition 1 than Condition 2 and vice versa for down-regulated transcripts.

- 21. Optionally, samples in the two groups of a CGP can be paired (Figure 3B).
 - a. For example, samples obtained before and after a perturbation from the same subject should be paired.
 - b. If the samples are paired, paired analyses are performed downstream, for example, a paired t-test.
 - c. If the samples to be paired are in the same order, click on "Set Default Sample Pair Index" to automatically create the pair index.
 - d. The pair indices can also be manually changed.
 - e. Pairing can only be done when the CGP is private. If the CGP is public, it must be made private for pair indices to be changed.
 - f. Click "Save Sample Pair Index" to save the pair index for the samples.
- 22. To make the CGP public, click the "Change" button next to the "Is Public" field.
- 23. Click "Close" to save the CGP.
- 24. If other users create CGPs using the study, they appear in bold under "Comparison Group Pairs" at the bottom of the page.
 - a. Click the green "CGP Detail" button to view sample details (Figure 4A).
- 25. Click "Next" on the study page to add CGP(s) to a compendium.

Create compendium

© Timing: 5 min

This step describes how to create a compendium using the previously created CGPs.

- 26. Select the CGP(s) to add to a compendium (Figure 4A).
- 27. Choose an existing compendium or click the "Create New Compendium" button.
- 28. Click the green "Add to Compendium" button.
- 29. After adding CGPs to a compendium, you can go back to the search study page, create more sample groups in the current study, or go to the compendium to perform analyses. The toolbar on the top of the page can also be used to navigate the website.

Compute differential expression profiles (DEPs)

© Timing: 15 min

This step describes how to compute differential expression profiles (DEPs) between the sample groups in a CGP.

- 30. Click on "My Compendia" on the toolbar to see a list of all compendia.
 - a. Click on the desired compendium.
- 31. On the "Compendium" page, the "CGPs" tab lists the CGPs that are in the compendium.
 - a. This tab includes the number of samples in each condition of the CGP, the number of features, the number of genes, the study of origin, platform, and whether it is public.
 - b. Click the "Make Public" button to make CGPs public.
- 32. To export the raw expression data, select the desired CGPs and click the "Export Raw Data On" button and choose whether to export the data in probe or gene space.
 - a. When the selected CGPs originate from different platforms or from RNA-seq studies, users can only export data in gene space.
- 33. To compute DEPs with default settings, click the "Compute DEPs with Default Settings" button.
 - a. The default settings use limma with BH multiple-testing correction, using normalized GEO data.



	Search 👻	My Study Lists	s My Compendia Doo	cument Ab	out 🛪	Welcome JDoe [My Profile]
		Make Samp	le Groups Make Com	parison Group Pa	irs Add Comparison Gro	up Pairs to Compendium
Stu	dy: GSE	50772	Link to Source			
Availa Note	ble data set(s): RI About using the N	MA Normalized Normalized GEO	Data, GEO Data (i.e., 'Series Data ∔	Matrix File' from (GEO), Normalized GEO Data (qu	antile normalized version of the GEO Data)
T Summ PubM Platfo Prot Ger	itte: Expression ary: Peripheral b Detail] led: 25861459 rm: GPL570: [H pes: 41025 nes: 20011 arison Group F	data in PBMCs f olood mononucle G-U133_Plus_2] Pairs	rom SLE patients and controls ar cells were collected from SI Affymetrix Human Genome U	LE patients in an o	observational study performed at	the University of Michigan Blood microar [Vie
Select co	CGPs to ad mpendium.	d to	Previous <	dd to Compendiu	Case Study-SLE	Or Create New Compendium Nex
- 2	Condition 1 Gro	oup ?	Condition 2 Group ?	Paired ?	In Compendia ?	
	GSE50772-SLE	-PBMC	GSE50772-control-PBMC	No		
	GSE50772-SLE	CGP Detail	GSE50772-Healthy	No		
	GSE50772-SLE	1 CGP Detail	GSE50772-Healthy1	No	Case Study-SLE X	
	PBMC IFN SLE	CGP Detail	PBMC SLE IFN control	No		
ompendi Name: Description:	PBMC IFN SLE um Case Study-SLE Upstate	CGP Detail	PBMC SLE IFN control	No C • Compendium	Compendium Name: Case Study-SLE Used Description: Turcore	to Take a Taur Oversi
ompendi Description Is Public Compendium	PBMC IFN SLE	CGP Detail	PBMC SLE IFN control	No C	Compendium Name: Case Study-SLE use Description: Used Is Public No Conver Compendium:	as Take Take
ompendi Name: Description: Is Public Compendium: Number of CGPs:	PBMC IFN SLE	CGP Detail	PBMC SLE IFN control	No C	Compendium Name: Case Study-SLE Used Description: Used Is Public: No Cargo Compendit: 3	te Tour Overs
Compendia Compendia Compendia Compendia Number of CQPs: Compute Di Th doubt about c	PBMC IFN SLE	CGP Detail	PBMC SLE IFN control	No C sCompandium	Compendium Name: Case Study-SLE use Description: used Compendix No Compendix Study-SLE Number of COPs: 3 COP Compute Differential Expression Profiles (C	as Take a Taxe Deviced EP(a) Meta-analysis Analysis Results tation phases see are "bioled and/or proped with your hipidetermative evolution
Compendia Description: Numer of CGPs: Compendia Number of CGPs: Compute Dia Market of Landau Company of Landau Compendia Company of Landau Company of Landa	PBMC IFN SLE	(DEP) Mete-analysis retation, pieses see our Turk DEPs under most of sconsars a details. The outputs	PBMC SLE IFN control	No C (Congentium 2045 tris. Please torial.	Compendium Name: Case Study-SLE use Description: Used Secondaria: No care Compendium: 3 CGPs Compute Differential Expression Profiles (D CGP I doubt about data analysis and result Interpre For more information about Meta-analysis, planase op	The analysis Analysis Results EPa) Meta-analysis Analysis Results tation, please see our Tutorial and/or consult with your bioinformatics colleage: through the Pubmed reference - Please consult with your local bioinformatics from at this section of the Tutorial Analysis Results
Compendia Name Description Namber GROPA Namber GROPA Compandum Namber GROPA Compandum Namber GROPA	PBMC IFN SLE	(CGP Detail) data (DEPa) Meta-analysis Meta-analysis DEPa under most of scorearia so details. The outputs of DID EFE under most of scorearia so details. The outputs of DID EFE under most of scorearia	PBMC SLE IFN control	No C In Comparadium gues rs. Please torial.	Compendium Name: Case Study-SLE use Description: Use Compendium: 3 Number of COPe: COPP Compute Differential Expression Profiles (C COP Compute Differential Expression Profiles (C COP Compute Differential Expression Profiles (C COP Compute Differential Expression Profiles (C	Base Store Owner EPa) Meta-analysis Analysis Results EPa) Meta-analysis Analysis Results tation, please see our Tutorial and/or consult with your bioinformatics colleage through the Pubmed reference . Please consult with your local bioinformatics found at this section of the Tutorial. Warning: MetaIntegrator has not been
Compension Compension Particle Pa	PBMC IFN SLE	CGP Detail (CEPa) Meta-analysis retation, please see our Tuth CEPa under most of sconaria a details. The outputs of DI CEPa under most of sconaria a details. The outputs of DI	PBMC SLE IFN control	No C C C C C C C C C C C C C C C C C C C	Compendium Mare: Case Study-SLE use Description: Compendium Description: Compendium Description: Compendium Compendium: Compendium C	Bee stor Over Cover Cover Cover Cover Cover Cover
Compension Particle Part	PBMC IFN SLE	CGP Detail CEPa Meta-analysis CEPa Meta-analysis retation, please see our Tuth EEPa under most of scenaria a details. The outputs of DI EEPa under most of scenaria a	PBMC SLE IFN control	No C C C C C C C C C C C C C C C C C C C	Compendium Mare: Case Study-SLE use Description: Surger Description: Surger Compendium: Surger Reme I doubt about data analysis and result interpre- tions of the doubt about data analysis and result interpre- formore information about Meta-snalysis can be attinicade details. The outputs of Meta-snalysis can be attinicade details.	
Compension Market Ma	PBMC IFN SLE	(CGP Detail) CEPa) Meta-analysis (CEPa) Meta-analysis retation, please see our Tuth EPPs under most of sconaria a details. The outputs of DI EPPs under most of scona	PBMC SLE IFN control	No C C C C C C C C C C C C C C C C C C C	Compendium Mare: Case Study-SLE use Description: <u>Usern</u> Description: <u>Usern</u> Description: <u>Usern</u> Compendium: <u>3</u> Rumber of CGPH: <u>3</u> CGP Compute Differential Expression Profiles (D Compute D Compute D Comp	
Compension Compension Company Compan	PBMC IFN SLE	(CCP Detail) CCP Detail (CPPa) Mets-analysis (CPPa) Mets-analysis CCP and the outputs of DI CCP output most of scoraria contails. The outputs of DI CCP output most of scoraria contails. The outputs of DI CCP output the output to DI CCP output to DI CCP output to DI CCP outpu	PBMC SLE IFN control	No C C C C C C C C C C C C C C C C C C C	Compendium Mare: Case Study-SLE use Description: Scare Compendium:	

Figure 4. Create a compendium and run analysis

(A) Users can add CGPs to an existing compendium or create a new compendium.

(B) Users can compute Differential Expression Profiles (DEPs). Users can choose which studies or platforms to include, as well as modify the statistical significance testing method, multiple testing correction, statistic to use for DEP matrix generation, and differential gene thresholds. (C) Users can perform meta-analysis. Users can choose the meta-analysis method to use, as well as the differential gene thresholds.

- b. See "quantification and statistical analysis" section below for more information on normalization methods) and using t-statistics to generate a gene-by-CGP matrix (i.e., the DEP matrix).
- c. CGPs with studies using platforms with missing probe-to-gene mapping information or where normalized GEO data is unavailable cannot be analyzed using this "one-click" approach.



- Protocol
- 34. To change the default settings used to compute DEPs, click on the "Compute Differential Expression Profiles (DEPs)" tab (Figure 4B).
- 35. Filter by organism, studies, or platform using the drop-down menus.
- 36. Choose the statistical parameters for performing differential analysis. Select from the following (Table 1):
- 37. Click on the check box next to the CGP to include that CGP in the DEP analysis.
- 38. If the CGP is generated from a study using a microarray platform, choose the data source (GEO, RMA normalized, or normalized GEO) and whether to perform analysis in gene or probe space.
- 39. If the CGP is generated from RNA-seq data, the data is normalized internally, and analyses are done in gene space.
- 40. Click the "Compute DEP(s)" button.

Perform meta-analysis

© Timing: 15 min

This step describes how to perform meta-analysis of the studies in a compendium. Users can choose between two methods - RankProd and MetaIntegrator.

- 41. On the "Compendium" page, click on the "Meta-analysis" tab (Figure 4C).
- 42. Select which meta-analysis method to use, RankProd (rank product method) or MetaIntegrator (random effect model).
- 43. If the CGPs in the compendium are generated from studies using a microarray platform, select to perform the analysis in gene or probe space.
 - a. If working with RNA-seq data, analysis is performed in gene space.
- 44. If using RankProd, select p-value or adjusted p-value and a threshold for differential expression analysis. Default is an adjusted p-value threshold of 0.05.
 - a. Selecting the adjusted p-value uses the percentage of false positive predictions (pfp) value. RankProd calculates the pfp value as the estimated percentage of false predictions.
- 45. If using MetaIntegrator, select FDR or effect size and a threshold for differential expression analysis. Default is an FDR threshold of 0.05.
- 46. Click on the check box next to the CGP to select CGPs to include in the analysis.
- 47. If the CGP is generated using a study using a microarray platform, select the data source (GEO, RMA normalized, or normalized GEO).
- 48. Select the reference condition for each CGP so the comparison is biologically consistent across CGPs.
 - a. For example, one CGP has patients with a disease status labeled as "Condition 1" and healthy subjects as "Condition 2". In a second CGP, healthy subjects are "Condition 1" and disease subjects are "Condition 2".
 - b. The reference conditions need to be made uniform to either the healthy or disease groups.

△ CRITICAL: The reference conditions need to be standardized correctly for the results to have biological significance.

49. Click the blue "Run Meta-analysis" button.

Table 1. Statistical parameters for performing differential analysis					
Parameter	Options				
Statistical Significance Testing Method	limma, Mann-Whitney test, student's t-test				
Multiple Testing Correction	Benjamini & Hochberg, Benjamini & Yekutieli, Holm, Hochberg, Hommel, Bonferroni				
Statistic to Use for DEP Matrix Generation	log fold-change, average expression, t-statistic, b-statistic, p-value, adjusted p-value, -log(p-value), -log(adjusted p-value)				
Differential Gene Threshold	Adjusted p-value, p-value				





Home 3	earcn - My	Study	Lists I	My Co	mpendia	Document -	About -			Welcome JDoe [My	y Profile]
My Compend	dia	Coi	mper	ndiu	Jm						
Name	Count			Name:	Case Stu	udy-SLE	Jpdate			Take a Tour De	elete Compendi
Case Study.	3		Descri	ption:	Up	odate					
Create New C	ompendium		ls l	Public	No	Change					
		Ν	Compen	dium:	3						
				our s.							
		CGPs	s Comp	ute Dif	ferential Ex	pression Profiles	s (DEPs) Meta-analy	sis Analysis R	esults		
	1	Note: //	f in doubt a	bout d	ata analysis	s and result inter	pretation, please see ou	r Tutorial and/or c	consult with your	bioinformatics coll	leagues
		An	alysis Run	Туре:	All		\$				Delete Jo
		Anal	ysis Run S	tatus:	All		\$				
			Run ID	R	un Type 🍦	Run Status	Scheduled Time	Launched	Time 🔶 F	inished Time 🕴	
		0 41	13	Met ana	a- Ivsis	Completed	04/28/22 17:36	04/28/22 17:3	36 04/2	8/22 17:36	×
		0 41	14	DEF	,	Completed	04/28/22 17:37	04/28/22 17:3	37 04/2	8/22 17:38	×
		U		DE.		Completed	o WEOVEE THIOT	o WEOREE THR	0.12		
							Name: Status: Public Access	Case Study-SLE Completed Generate			
Input Parameters:	T Differential Corr	Testing Testing Co	Method: mrection: 0.05	a Benjamini i sted p-valu	& Hochberg (19 ie	195)]	Public Access Input Parameters:	Case Study-SLE Completed Generate	Gene or Probe?:	Gene fdr 0.05	
Input Parameters:	T Differential Gen Differential Gen Statistical value(i	Testing festing Co the Thresho e Thresho in the DEP	Method: BH [E adjus prrection: 0.05 old Field: t-stat P matrix):	a Benjamini i sted p-valu sistic	& Hochberg (19 e	95)]	Comparison	Case Study-SLE Completed Generate Differential Ge	Gene or Probe?: ne Threshold Field: ne Threshold Value:	Gene Idr 0.05	
Differential Expression Profile (DEP) Analysis	T Differential Gen Differential Gen Statistical value(i CGP Name	Testing Testing Co the Thresho Thresho in the DEP Paired	Method: BH [E adjus rrection: 0.05 old Field: Matrix): P matrix):	a Benjamini i Isted p-valu Iistic PataOutput	& Hochberg (19 e t Result	95))	Compendant Name: Public Access URL: Input Parameters: Comparison Group Pairs:	Case Study-SLE Completed Generate Differential Ge Differential Ge CGP Name GSE3447-Healthy-	Gene or Probe?: ne Threshold Field: ne Threshold Value: Normalized Data Normalized GEO	Gene Idr 0.05 Non-reference Cond C2: GSE3447-	itiBeference co C1: GSE3447-
Input Parameters: Differential Expression Profile (DEP) Analysis Results:	T Differential Gen Differential Gen Statistical value(f CGP Name GSE3447-Healthy- PBMC::GSE3447- Lupus_Erythematosus, Systemic-PBMC	Testing Testing Co the Thresho Thresho in the DEP Paired No	limm: BH [E] prrection: adjus Jold Field: 0.05 Jold Field: 1.5tat P matrix): P Normalized GEO Data	a Benjamini i Isted p-valu listic JataOutpul Gene	& Hochberg (19) te te DEP statistics genes/probes determining D	95)] s for all s (statistics for 2E gene list below):	Comparison Status: Public Access URL: Input Parameters: Comparison Group Pairs:	Case Study-SLE Completed Generate Differential Generate CGP Name GSE3447-Healthy- PBMC::GSE3447- Lupus_Erythematosu s_Systemic-PBMC Ends In Strive	Gene or Probe?: ne Threshold Field: ne Threshold Value: Normalized Data Normalized GEO Data	Gene fdr 0.05 Non-reference Cond C2: GSE3447- Lupus_Erythematosu s_Systemic-PBMC	iti Bef erence co C1: GSE3447 Healthy-PBM(
Differential Expression Profile (DEP) Analysis Results:	Differential Gen Differential Gen Statistical value(i CGP Name GSE3447-Healthy- PBMC:GSE3447-Healthy- PBMC:GSE3447-Healthy- PBMC:	Testing Testing Co te Thresho e Thresho e Thresho n the DEP Paired'	Imma BH [E adjus old Field:	a Benjamini i tistic PataOutput Gene	& Hochberg (19) te Result DEP statistics genes/probes determining D J Differentially E Probe List: Up Down-regulat	s for all (statistics for E gene list below): Expressed Gene or p-regulated: ▲ ed: ▲	Comparison Status: Public Access URL: Input Parameters: Comparison Group Pairs:	Case Study-SLE Completed Generate Differential Gen Differential Gen GSE3447-Healthy- PBMC::GSE3447- Lupus_Erythematosu s_Systemic-PBMC GSE20864- Lupus Erythematosu Lupus Erythematosu	Gene or Probe?: ne Threshold Field: ne Threshold Value: Normalized GEO Data	Gane fdr 0.05 Non-reference Cond C2: GSE3447- Lupus, Erythematosu s_Systemic-PBM C1: GSE20864- Lupus Erythematosu	itiBeference co C1: GSE3447 Healthy-PBMC C2: GSE2086 Healthy-
Differential Expression Profile (DEP) Analysis Results: ts for GP.	T Differentil 6en Differentil 6en Statistical value(i CGP Name GSE3447- tupus, Eyritemiotosys Systemic-PBMC Cir GSE3447- Lupus, Eyritemiotosys Systemic-PBMC Cir GSE3447- thoitity- PBMC Cir GSE3447-thoitity- PBMC Cir GSE3447-thoitity- PBMC Cir GSE3447-thoitity- PBMC	Testing Cesting Co the Thresho e Thresho in the DEP Paired No	Imma BH [6] BH [6] Ind Field 2.05 Ind Field 2.05 Ind Field 2.05 Normalized D GEO Data	a Benjamini i Ited p-valu iistic VataOutpul Gene	Hochberg (19) ie Besult DEP statistics genes/probes determining D Differentially E Drobe List: Up Down-regulat Heatmap (100 genes/probes GenePattern I	s for all (statistics for begins in the state of the state Expressed Gene or pregulated; ▲ dot ▲ tot ▲ To most DE); ■ ▲ ACT▲	Comparison Status: Public Access URL: Input Parameters: Comparison Group Pairs:	Case Study-SLE Completed Generate Differential Ge Differential Ge CGP Name GSE3447-Healthy- PBMC::GSE3447- Lupus_Erythematosu s_Systemic-PBMC GSE20844- Lupus_Erythematosu s_Systemic- PBMC::GSE20864- Healthy-	Gane or Proba?: nen Threshold Field: ne Threshold Value: Normalized GEO Data Normalized GEO Data	Sene Idr 0.05 Non-reference Cond C2: GSE3447- Lupus, Erythematosu sSystemic-PBMC C1: GSE20864- Lupus, Erythematosu sSystemic-PBMCs	C1: GSE3447 Healthy-PBMC C2: GSE2086- Healthy- Lupus_Eryther s_Systemic
Differential Expression Profile (DEP) Anilysis (DEP) Anilysis (DEP	T Differential Gen Differential Gen Statistical value(CGP Name GSE3447-tealthy- PBMC::GSE3447- Lupus, Erythematosus, Systemic-PBMC C2: GSE3447- Lupus, Erythematosus, .Systemic-PBMC GSE20844-	Testing Co lesting Co e Thresho in the DEP Paired ⁷ No	Imm Bri [E Bri [Z Bri [Z Bri] Bri] Bri [Z Bri] Bri [Z Bri] Bri] Bri [Z Bri] Bri] Bri [Z Bri] Bri	a Benjamini i tited p-valu listic Gene Gene	8. Hochberg (19) e Result DEP statistics genes/probes determining D Differentially E Probe List: Up Down-regulat Beans/probes Gene9/attern I DEP statistics DEP statistics	a for all (statistics for 2° gene list below): ∑regulated: ▲ ort. ▲ most DE): ■ ▲ Input Files: GCT▲ a for all	Comparison Name: Status: Public Access URL: Input Parameters: Comparison Group Pairs:	Case Study-SLE Completed Generate Differential Ge Differential Ge GE2447-Healthy- PPMC-GSE3447- Lupus_Erythematosu s_Systemic-PBMC GSE20864- Lupus_Erythematosu s_Systemic-PBMC GSE20864- Lupus_Erythematosu s_Systemic PBMC to Study	Gene or Proba?: ne Threshold Field: ne Threshold Field: Normalized Data Normalized GEO Data	Gene fdr 0.05 C2: GSE3447- Lupus Erythematogu sSystemic-PBMC C1: GSE20864- Lupus Erythematogu sSystemic-PBMCs	ItiBeference co C1: GSE3447 Healthy-PBMC C2: GSE20866 Healthy- Lupus Erythe s_Systemic
nput Parameters: Differential Expression Profile (DEP) Analysis Results: ES fOr GP.	T Differential Gen Differential Gen Statistical value(CGP Name CGP Name CGP Name CGS2447-Haalthy- PBMC-2GS2447-Haalthy- PBMC-2GS2447- Lupus, Erythematosus, Systemic-PBMC CSS20844- Lupus, Erythematosus, Systemic- PBMC- CSS20844- Lupus, Erythematosus, Systemic- PBMC- Systemic- PBMC- Systemic- Sys	Testing Co tee Thresho e Thresho e Thresho No No	IIImm Method: adjuster adjuster Method: adjuster I object Method Method GEO Data	a Benjamini i ted p-valu iistic ataOutput Gene Gene	Result Result DEP statistics genes/probes determining D Image: the statistics determining D mining D mining D mining D Down-regulat mining D DEP statistics genes/probes determining D DEP statistics determining D Defendation	s for all is datatics for be gene list below): Expressed Gene or pr-egulated: ▲ most DE is in the files: GCT▲ s for all is to below): E gene list below):	Comparison Status: Public Access URL: Input Parameters: Comparison Group Pairs:	Case Study-SLE Completed Generate Differential Generate CGP Name GSE3447-Healthy- PBMC::GSE3447- Lupus_Erythematosu S_Systemic-PBMC GSE20844 Lupus_Erythematosu S_Systemic PBMCS::GSE20844 Healthy- Lupus_Erythematosu S_Systemic Deck To Study GSE50874- BES (SSE2084- Lupus_Erythematosu S_Systemic Deck To Study GSE50772- SLE1::GSE50772-	Gene or Probe?: men Threshold Field: ne Threshold Value: Normalized Data Normalized GEO Data Normalized GEO Data	Gene Idd 0.05 Non-reference Cond C2: GSE3447- Lupus_Erythematosu s_Systemic-PBMCs C1: GSE20864- Lupus_Systemic-PBMCs Systemic-PBMCs C1: GSE50772-SLE1	ItitBeference co C1: GSE3447. Healthy-PBMC C2: GSE2086. Healthy- Lupus_Eryther s_Systemic C2: GSE50777. Healthy1
Differential xpression Profile (DEP) Analysis Results: ts for GP.	T Differential Gen Differential Gen Statistical value(CGP Name GSE3477-Ikathy- BMC::GSE3477- Lupus, Erythematosus, Systemio-PBMC C2: GSE347- Lupus, Erythematosus, Systemio-PBMC GSE20864- Lupus, Erythematosus, Systemio- Lupus, Erythematosus, Systemio- Lupus, Erythematosus, Systemio- Lupus, Erythematosus, Systemio- Beack to Study	Testing Co te Thresho e Thresho e Thresho No No	limm Method: BH [E vrrection: adjus for adjus for adjust for adjust rectance adjust rectanco rectance adjust r	a Benjamini i ited p-valu iistic BataOutput Gene Gene	Hochberg (19/e e E Result DEP statistics genes/probes determining D Øriferentially E Probe List: U Down-regulat Heatmap (100 genes/probes GenePattern I CLSA DEP statistics genes/probes determining D Øriferentially E Probe List: U Down-regulat	s for all (statistics for big Begine list below): Expressed Gate or pregulated ± dot ± most DE (statistics for DE (statistics for All (statistics for All (statistics for All (statistics for (statistics for All (statistics for All (statistics for (statistics for All (statistics for All (statistics for (statistics for All (statistic))))))))))))))))))))))))))))))))))	Comparison Status: Public Access URL: Input Parameters: Comparison Group Pairs:	Case Study-SLE Completed Generate Differential Generate CGP Name GSE3447- Lupus, Erythematosu s, Systemic-PEMC (GSE20864- Lupus, Erythematosu s, Systemic-PEMC (GSE20864- Lupus, Erythematosu s, Systemic PEMCS::GSE20864- Lupus, Erythematosu s, Systemic Exot to Study (GSE20772- SLE1::GSE50772- SLE1::GSE50772- SLE1::GSE50772-	Gene or Proba?: ne Threshold Field: ne Threshold Field: Normalized GEO Data Normalized GEO Data	Gene (dr 20.05) Non-reference Cond C2: GSE3447. Lupus Eythematosu s.Systemic-PBMC C1: GSE20864- Lupus Eythematosu s.Systemic-PBMCs C1: GSE20864- Lupus Eythematosu s.Systemic-PBMCs C1: GSE20864- Lupus Eythematosu s.Systemic-PBMCs	ISBN ference co C1: GSE3447. Healthy-PBMC C2: GSE2086- Healthy- Lupus_Eythenic S_Systemic C2: GSE50777 Healthy1
Differential xpresion Profile (DEP) Analysis Results: LS FOT GP.	T Differential Gen Differential Gen Statistical value(CGP Name GSE3447-Healthy- PBMC::GSE347-Healthy- PBMC: C2: GSE347-Healthy- PBMC C2: GSE347-Healthy- PBMC C2: GSE347-Healthy- PBMC::GSE32084- Lupus, Erythematosus, Systemic-PBMCs: GSE2084- Lupus, Erythematosus, Systemic-PBMCs: C1: GSE2084- Lupus, Erythematosus, Systemic-PBMCs: C1: GSE2084- C1: GSE20	Testing Co testing Co te Thresho In the DEP Paired1 No	Imm Method: BH [E adjustmettion: adjustmethod: adjustmethod: diffield: 1-stat matrix: Normalized GEO Data	a Benjamini i Listic detaOutpul Gene Gene	Hochberg (19) in	s for all (callises for (callises for Eigene list below): Expressed Gane or pregulated: \bot or most DE (callises for DE gene list below): Expressed Gane or DE gene list below): Expressed Gane or regulated: \bot or all \bot or all \bot or all \bot \bot or all \bot \bot or below): Expressed Gane or \bot \bot \bot \bot \bot \bot \bot \bot	Comparison Status: Public Access URL: Input Parameters: Comparison Group Pairs:	Case Study-SLE Completed Generate Differential Generate GSE3447-Healthy- PBMC::GSE3447- Lupus_Erythematosu s_Systemic-PBMC GSE20844 Lupus_Erythematosu s_Systemic- PBMCS::GSE2084- Healthy- Lupus_Erythematosu s_Systemic Back to Study GSE50772- Back to Study GSE50772- Healthy- Eack to Study	Gene or Probe?: ne Threshold Field: normalized Data Normalized GEO Data Normalized GEO Data Normalized GEO Data Normalized GEO Data	Gane fdr C2: GSE3447- Lupus, Erythematosu s_Systemic-PBMC C1: GSE20864- Lupus, Erythematosu s_Systemic-PBMCs C1: GSE20864- Lupus, Erythematosu s_Systemic-PBMCs C1: GSE30772-SLE1	ItiBeference cc C1: GSE3447 Healthy-PBM C2: GSE2088 Healthy- Lupus, Erythe sSystemic C2: GSE5077; Healthy1 Healthy1 Healthy1
Input Parameters:	To Differential Gen Statistical value() CGP Name CGP	Testing Geting Co. te Thresho Thresho In the DEP Paired: No	limm Method: BH [E vrrection: adjus of Field: 0-5 Normalized D Normalized GEO Data	a adenjamini i denjamini i denjamini i denjamini i denjamini i denjamini i denjamini denjamini denjamini denjami d Denjami denjami d	Hochberg (19) e	a for all is fatalities for Expressed Gane or progulate: ▲ Droot DE Stransfer Corta or call or call below: Expressed Gane or progulates for Stransfer Delow: Expressed Gane or progulates for Stransfer Delow: Stransfer Delow: Strans	Comparison Status: Public Access URL: Comparison Group Pairs:	Case Study-SLE Completed Generate Differential Generate CGP Name GSE3447-Healthy- PBMC::GSE3447- Lupus_Erythematosu s_Systemic-PBMC GSE20844 Lupus_Erythematosu s_Systemic PBMCS::GSE20844 Healthy- DBMCS::GSE20844 Healthy- BMCS::GSE20844 Lupus_Erythematosu s_Systemic Back to Study GSE50772- SLE1::GSE50772- Healthy1 Back to Study CSE50772- SLE1::GSE50772- Healthy1 Back to Study CSE50772- SLE1::GSE50772- Healthy1 Back to Study	Gene or Probe?: Iner Threshold Field: Iner Threshold Value Normalized Data Normalized GEO Data Normalized GEO Data Normalized GEO Data Anormalized GEO Data	Gene Idd 0.05 Non-reference Cond C2: GSE3447- Lupus_Erythematosu s_Systemic-PEMCs C1: GSE20864- C1: GSE2086- C1: GSE2086- C1: GSE2086- C1: GSE208- C1: GSE208-	ItilBaference co C1: GSE3447 Heatthy-PBMC C2: GSE2086 Heatthy- Lupus_Eryther s_Systemic C2: GSE50777 Heatthy- C2: GSE50777 Heatthy- theatthy- theatthy- theatthy- theatthy- theatthy- theatthy- model CGEI Intor
Differential Expression Profile (DEP) Analysis Results: ts for GP.	T Differential Gen Differential Gen Statistical value(CGP Name GSE3447-Healthy- PBMC: GSE3447- Lupus, Erythematosus, Systemic-PBMC C2: GSE3447- Lupus, Erythematosus, Systemic-PBMCs: GSE2084- Healthy- Lupus, Erythematosus, Systemic-PBMCs: GSE2084- Lupus, Erythematosus, Systemic-PBMCs: GSE2084- Lupus, Erythematosus, Systemic-PBMCs: GSE2084- Lupus, Erythematosus, Systemic-BMCs C2: GSE2084- Lupus, Erythematosus, Systemic-BMCs GSE3072- Systemic-Statistical Systemics Systemic-Statistical Systemics Systemic	Testing Contracting Contractin	IImm Method: BH [E vrrection: adjus vrrection: adjus for adjus for adjust vestice of the second district of the second geo Data	a Benjamini jamini ja Listic LetaOutputa Gene Gene	Hochberg (19) ie	a for all (statistics for by Egene list below): Expressed Gate or progulated ▲ for all (statistics for reliance or progulated = for all (statistics for reliance or progulated ± Expressed Gate or progulated ± (statistics for Direct Disc (statistics for): ■ ▲ A (statistics for	Neme: Status: Public Access URL: Input Parameters: Comparison Group Pairs: Meta-analysis results:	Case Study-SLE Completed Generate Differential Ge Differential Ge GE3447-Healthy- PPMC::GSE3447- Lupus_Erythematosu s_systemic-PBMC GSE20864- Lupus_Erythematosu s_systemic-PBMC GSE20864- Lupus_Erythematosu s_systemic Back to Study GSE20864- Lupus_Erythematosu s_systemic Back to Study GSE50772- SLE1::GSE0772- Healt to Study GSE50772- SLE1::GSE0772- Healt to Study Complementally Expre	Gene or Probe?: me Threshold Field: hormalized Data Normalized GEO Data Normalized GEO Data Normalized GEO Data Normalized GEO Data ed* genes correspond tive to the reference of seed Gene List: [Ana negative	Gene (dr 0.05 Non-reference Cond C2: GSE3447- Lupus Erythematosu sSystemic-PBMCs C1: GSE20864- Lupus Erythematosu sSystemic-PBMCs C1: GSE50772-SLE1 It of those with higher expro- ordition and vice versa for	ItiBeference co C1: GSE3447 Healthy-PBMC C2: GSE2086 Healthy-PBMC C2: GSE2086 Healthy-C2: GSE5077 Healthy1 C2: GSE5077 Healthy1 Healthy1 Healthy1 ToppGene Suite
Differential Expression Profile (DEP) Analysis Results: ts for GCP.	To Differential Gen Differential Gen Statistical value(CGP Name SS2447-14althy- PENC-3653447-16althy- PENC-3653447-16althy- PENC-3653447-16althy- PENC-3653447-16althy- PENC-3653447-16althy- Lupus, Erythematosus, Systemic-PENC GSE20844- Lupus, Erythematosus, Systemic- PENC-36536947- Lupus, Erythematosus, Systemic- Cal GSE2084- Lupus, Erythematosus, Systemic- Cal GSE2084- Lupus, Erythematosus, Systemic- Cal GSE2084- Lupus, Erythematosus, Systemic- Cal GSE2084- Healthy- Lupus, Erythematosus, Systemic- Cal GSE2084- Healthy- Lupus, Erythematosus, Systemic- Cal GSE2084- Healthy- Lupus, Erythematosus, Systemic- Cal GSE2077-SLE1	Testing feeting Co. In the DEP Paired? No No	IIIrmm Method: BH [E irrrection: adjus of Felde: 0-54 (Second Sec	a Baenjamini i Sunda S	Hochberg (19) Be Hochberg (19) Be Hochberg (19) Be Hochberg (19) DEP statistics genes/probes denemining D Derestatistics genes/probes GenePattern CLS4 DEP statistics genes/probes GenePattern CLS4 DEP statistics genes/probes GenePattern DEP statistics genes/probes determining D DEP	PS5) of or all is datalises for Egene list below): Expressed Gene or progulated: ▲ act. ▲ most DE ▲ is datalises for E gene list below): E gene list below): E gene list below): is a table of the list below): E gene list below): E gene list below): E gene list below): E gene list below):	Mene: Status: Public Access URL: Input Parameters: Comparison Group Pairs: Meta-analysis results:	Case Study-SLE Completed Generate Differential Generate GSE3447-Healthy- PBMC::GSE3447- Lupus_Erythematosu s_systemic-PBMC GSE20844 Lupus_Erythematosu s_systemic- PBMCS::GSE20844 Healthy- Lupus_Erythematosu s_systemic PBMCS::GSE20844 Healthy- Lupus_Erythematosu s_systemic Benet to Study GSE5072- SLE1::GSE50772- Healthy1 Benet to Study (SGE50772- Healthy1 Benet to Study) (SGE50772- Healthy1 Benet to Study)	Gene or Probe?: me Threshold Field: ne Threshold Value: Normalized Data Normalized GEO Data Normalized GEO Data Normalized GEO Data ed" genes correspond tive to the reference of ssed Gene List [Ana neså Geneså	Gana Idd 0.05 Non-reference Cond C2: GSE3447- Lupus_Erythematosu s_Systemic-PBMC C1: GSE20864- Lupus_Erythematosu s_Systemic-PBMCs C1: GSE20864- Lupus_Erythematosu s_Systemic-PBMCs C1: GSE20864- Lupus_Erythematosu s_Systemic-PBMCs C1: GSE20772-SLE1 C1: GSE50772-SLE1 Lot those with higher expr condition and vice versa for Down	ItilBaference co C1: GSE3447- Healthy-PBMC C2: GSE2086- Healthy-Lupus Eryther sSystemic C2: GSE50772 Healthy1 ression in the non-reg mited CCP Infor
Input Parameters:	CGP Name CGP	Testing feeting Co threshold In Threshold No No	IIImm Method: BH [E irrection: adjus irrection: adjus reading adjust irrection: adju	a anapannini inter prvalu istic prvalu istic Gene Gene	Hochberg (19) Hochber	a for all (statistics for Egene list below): Expressed Gene or progulated: ▲ tot ↓ The most DE (statistics for Egene list below): Expressed Gene or progulated: ↓ of crall (statistics for Egene list below): Expressed Gene or progulated: ↓ of crall (statistics for Egene list below): Expressed Gene or progulated: ↓ Statistics for Egene list below): Expressed Gene or progulated: ↓ Statistics for Expressed Gene or Progulated: ↓ Expressed Gene or Expressed Gene or Expr	Meta-analysis results:	Case Study-SLE Completed Generate Differential Ge Differential Ge GE3447-Healthy- PBMC::GSE3447- Lupus_Erythematosu s_Systemic-PBMC GSE20844 Lupus_Erythematosu s_Systemic PBMCS::GSE20844 Heaty- PBMCS::GSE20844 Lupus_Erythematosu s_Systemic Back To Study GSE20844 Lupus_Erythematosu s_Systemic Back To Study GSE50772- SLE1::GSE50772- Heaty1 GSE50772- SLE1::GSE50772- Heaty1 GSE50772- SLE1::GSE50772- Heaty1 GSE50772- SLE1::GSE50772- Heaty1 GSE50772-	Gene or Probe?: The Threshold Field: Normalized GEO Data Normalized GEO Data Normalized GEO Data Normalized GEO Data Normalized GEO Data Normalized GEO Data Mormalized GEO Data Mormalized GEO Data Mormalized GEO Data Mormalized GEO Data Mormalized GEO Data Mormalized GEO Data	Gene Idr 0.05 Non-reference Cond C2: GSE3447- Lupus_Erythematosu s_Systemic-PEMC C1: GSE2086-4 C1: GSE2086-4 C1: GSE20872-SLE1 I to those with higher expr ondition and vice versa for Upper list using "	Iti Baference co C1: GSE3447- Heatthy-PBMC C2: GSE2086/ Heatthy- Lupus_Eyther s_Systemic C2: GSE50777 Heatthy- C2: GSE50777 Heatthy- C2: GSE50777 Heatthy- C2: GSE50777 Heatthy- norther down-reg more CCPLinton
Input Parameters: Differential Expression Profile (DEP) Analysis Results: Its for CGP.	To Differential Gen Differential Gen Statistical value(CGP Name GSE3447-Healthy- PPMC: GSE347-Healthy- PPMC: GSE347-Healthy- PPMC: GSE3547-Healthy- PPMC C2: GSE347-Healthy- PPMC: GSE52084- Lupus, Erythematosus, Systemic-PBMCs: GSE2084- Healthy- PMCs: GSE2084- Lupus, Erythematosus, Systemic-PBMCs: GSE2084- Lupus, Erythematosus, Systemic-PBMCs: GSE2084- Lupus, Erythematosus, Systemic-PBMCs: GSE35084- Lupus, Erythematosus, Systemic-PBMCs: GSE35084- Lupus, Erythematosus, Systemic-PBMCs: GSE35084- Lupus, Erythematosus, Systemic-PBMCs: GSE35084- Lupus, Erythematosus, Systemic-PBMCs: GSE350772- Lic GSE350772- Lic GSE50772- Lic GSE50772- Lic GSE50772- Healthy1	No No	Imm Method: BH [E wrection adjus diversion adjus for adjus for adjustic state of the state of th	a sonjamini njegovje sonjamini njegovje sonjamini njegovje sonjamini njegovje sonjamini sonja so	Hochberg (19) ie Result DEP statistics genes/probes determining D DEP statistics genes/probes determining D DEP statistics genes/probes determining D Deven-regulat Heatmap (100 genes/probes determining D Deven-regulat Heatmap (100 genes/probes determining D Nown-regulat Heatmap (100 genes/probes GenePattern (2.54	a for all (statistics for Use gene list below): Expressed Gane or pregulated: ▲ tor most DE (statistics for rall (statistics for rall (statistic	Neme: Status: Public Access URL: Input Parameters: Comparison Group Pairs: Meta-analysis results:	Case Study-SLE Completed Generato Differential Generato GSE3447-Healthy- PPMC::GSE3447- Lupus_Erythematosu s_systemic-PBMC GSE20844- Lupus_Erythematosu s_systemic- PBMCS::GSE2084- Healthy- Lupus_Erythematosu s_systemic Back to Study GSE5072- Back to Study GSE5072- Healthy- Eack to Study GSE50772- Healthy- Eack to Study GSE50772- Healthy- Eack to Study Cottes The "up-regulated Down-regulated	Gene or Probe?: ne Threshold Field: hormalized Data Normalized GEO Data Normalized GEO Data Normalized GEO Data Normalized GEO Data seed Gene List: [Ana rest Genes 4 Meta-anal	Gene fdr C2: GSE3447- Lupus, Erythematosu sSystemic-PBMC C1: GSE20864- Lupus, Erythematosu sSystemic-PBMCs C1: GSE20864- Lupus, Erythematosu sSystemic-PBMCs C1: GSE50772-SLE1	ItiBeference coo
Input Parameters:	CGP Name SS2447-1-Authy- PBMC:SS2447-1-Authy- PBMC:SS2447-1-Authy- PBMC:SS2447-1-Authy- PBMC:SS2447-1-Authy- PBMC:SS2447-1-Authy- PBMC:SS2447-1-Authy- PBMC:SS2447-1-Authy- PBMC:SS2447-1-Authy- PBMC:SS2447-1-Authy- Lupus.Erythematosus, Systemic- PBMC:SS2447-1-Authy- Lupus.Erythematosus, Systemic- SS2084-1- Lupus.Erythematosus, SS2084-1- Lupus.Erythematosus, SS2084-1- Lupus.Erythematosus, SS2084-1- Lupus.Erythematosus, SS2084-1- Lupus.Erythematosus, SS2084-1- Lupus.Erythematosus, SS2084-1- Lupus.Erythematosus, SS2084-1- Lupus.Erythematosus, SS2084-1- Lupus.Erythematosus, SS2084-1- Lupus.Erythematosus, SS2084-1- Lupus.Erythematosus, SS2084-1- Lupus.Erythematosus, SS2084-1- Lupus.Erythematosus, SS2084-1- Lupus.Erythematosus, SS2084-1- Lupus.Erythematosus, SS2084	Testing Threshold	Imm Method: BH [E adjustication of the second of Field: 1-start matrix: Normalized GEO Data Normalized GEO Data	a anglannin in tea anglanni in tea anglanni tea anglan	Hochberg (19) ie Result DEP statistics genes/probes determining D Differentially E Probe List: Up Disetaming 100 genes/probes demers/probes GenePattern TCLS4 DEP statistics genes/probes GenePattern TCLS4	PS() For all (statistics for Degree list below): Expressed Gene or p-regulated: ↓ a for all (statistics for Degree list below): Expressed Gene or p-regulated: ↓ area in the below in the	Comparison Status: Public Access URL: Input Parameters: Comparison Group Pairs: Meta-analysis results:	Case Study-SLE Completed Generato Differential Ge Differential Ge GSE3447-Healthy- PBMC::GSE3447- Lupus_Erythematosu s_systemic-PBMC GSE20844 Lupus_Erythematosu s_systemic PBMCs::GSE20844 Healthy- Lupus_Erythematosu s_systemic Benet to Study? GSE20844 Lupus_Erythematosu s_systemic Benet to Study? GSE50772- Benet to Study? GSE50772- Healthy1 Benet to Study? INCES The 'up-regulated Download	Gene or Probe?: me Threshold Field: ne Threshold Field: Normalized Data Normalized GEO Data Normalized GEO Data Normalized GEO Data seed Gene List [Ana neg4 Genes4 meta-ana]	Gene Idd 0.05 Non-reference Cond C2: GSE3447- Lupus_Erythematosu s_Systemic-PBMC C1: GSE20864- Lupus_Erythematosu s_Systemic-PBMCs C1: GSE20864- Lupus_Erythematosu s_Systemic-PBMCs C1: GSE20864- Lupus_Erythematosu s_Systemic-PBMCs C1: GSE20864- Lupus_Erythematosu s_Systemic-PBMCs Sy	ItilBaference co C1: GSE3447- Healthy-PBMC C2: GSE20864 Healthy-Lupus_Eyther sSystemic C2: GSE50772 Healthy1 ression in the non-reg ression in the non-reg res
Input Parameters:	CGP Name CGP	Testing Testing Control Testing Threshold Testing No No No No No	IIImm Method: BH [E irrection: adjus irrection: adjus for the second of Field: 0-32 Record adjust Record adjust Re	a analization of the second se	Hochberg (19) He Result DEP statistics genes/probes determining D Probe List: Up Down-regulat Probe List: Up DeP statistics genes/probes determining D Differentially E Probe List: Up Differentially E Deen statistics genes/probes GenePattern I CLS-4 asamples in C1 Cl, vs. the refer relative to C2	a for all (statistics for Egrene list below): Expressed Gene or pregulated: ⊥ b and the below): Expressed Gene or pregulated: ⊥ and C2 of the CGP aread c(2). A gene and C2 of the CGP and vice versa.	Loon Name Status Public Access URL Comparison Group Pairs: Meta-analysis results:	Case Study-SLE Completed Generate Differential Ge Differential Ge GE3447-Healthy- PBMC::GSE3447- Lupus_Erythematosu s_Systemic-PBMC GSE2084- Lupus_Erythematosu s_Systemic PBMCS::GSE2084- Lupus_Erythematosu s_Systemic Back to Study GSE50772- SLE1::GSE50772- Head to Study GSE50772- Head to Study GSE50772- Head to Study CSE50772- Head to Study	Gene or Probe?: me Threshold Field: Normalized GEO Data Normalized GEO Data Normalized GEO Data Normalized GEO Data Mormalized GEO Data	Gene Idr 0.05 Non-reference Cond C2: GSE3447- Lupus_Erythematosu s_Systemic-PEMCs C1: GSE2086-4 C1: GSE2086-4 C1: GSE20872-SLE1 I to those with higher expr ondition and vice versa for I to those suith higher expr whyze the gene list using ' ysis results.	Iti Baference co C1: GSE3447- Heatthy-PBMC C2: GSE2086+ Heatthy- Lupus_Eyther s_Systemic C2: GSE50772 Heatthy- Lupus_Eyther s_Systemic C2: GSE50772 Heatthy- theatthy- theatthy- nor d CGPI inton ToppGene Suite
Input Parameters:	CGP Name Statistical value(CGP Name GSS.247.748179, FPMC GSS.247.748179, FPMC GSS.247.748179, FPMC GSS.247.748179, FPMC GSS.247.748179, FPMC GSS.247.748179, FPMC GSS.247.748179, FPMC GSS.247, Systemic GSS.20844 Lupus, Eyrthematosus, Systemic Ends fo Story C1: GSS.2084- Lupus, Eyrthematosus, Systemic C1: GSS.2084- Lupus, Eyrthematosus, Systemic C1: GSS.2084- Lupus, Eyrthematosus, Systemic C1: GSS.2087-24- Li: GSS.2077- Heathyl C1: GSS.	means whe	Imm Method: BH [E mrection adjuster digits of Field: 0-5 did Values - 5-aut matrix: Phormalized GEO Data Normalized GEO Data Normalized GEO Data Normalized GEO Data	a a conjamini ni sistic vistic	Hochberg (19) ie Result DEP statistics genes/probes dearmining D DEP statistics genes/probes dearmining D DEP statistics genes/probes dearmining D Dewn-regulat Heatmap (100 genes/probes deares/attern I CLS-4	afor all (statistics for r) Egene list below): Expressed Gare or progulated to most DE (statistics for rill (statistics for rill (statistics for rill (statistics for r) Expressed Gare or progulated: 1 (statistics for Expressed Gare or progulated: 1 (statistics for E gene list below): Expressed Gare or progulated: 1 (statistics for E gene list below): Expressed Gare or progulated: 1 most DE (statistics for E gene list below): E gene list belo	Name: Status: Public Access URL: Input Parameters: Comparison Group Pairs: Meta-analysis results:	Case Study-SLE Completed Generato Differential Ge Differential Ge (GSE3447-Healthy- PBMC::GSE3447- Lupus_Erythematosu s_systemic-PBMC (GSE20844- Lupus_Erythematosu s_systemic PBMCS::GSE20844- Healthy- Lupus_Erythematosu s_systemic Back to Study (GSE2084- Healthy- Lupus_Erythematosu s_systemic Back to Study (GSE50772- Healthy- Back to Study) (GSE50772- Healthy- Back to Study) (GSE5072- Healthy- Back to Study) (GSE50772- Healthy- Back to Study) (GSE50772- Healthy- Healthy- Healthy- (GSE50772- Healthy- Healthy- Healthy- (GSE50772- Healthy- (GSE50772- Healthy- (GSE50772- Healthy- (GSE50772- Healthy- (GSE50772- Healthy- (GSE50772- Healthy- (GSE50772- Healthy- (GSE50772- Healthy- (GSE50772- Healthy- (GSE50772- Healthy- (GSE50772- Healthy- (GSE50772- Healthy- (GSE50772- Healthy- (GSE50772- Healthy- (GSE50772- Healthy	Gene or Probe?: me Threshold Field: ne Threshold Value: Normalized Data Normalized GEO Data Normalized GEO Data Normalized GEO Data def genes correspond tive to the reference of seed Gene List: [Ana nes= Genes=	Gane tdr dr dr dr dr dr dr dr dr dr	ItilBeference co C1: GSE3447- Healthy-PBMC C2: GSE20864 Healthy-Lupus_Eryther s_Systemic C2: GSE50772 Healthy1
Input Parameters: Differential Expression Profile (DEP) Analysis Results: Lts for CGP.	CGP Name CGP	Testing Testing Comparison in the DEP Paired? No No No No No No	IIImm Method: BH [E Immediate adjustmediate adjustmediate adjustmediate of Field: 0-54 Record adjustmediate Record	a a conjanni i ni teted p-valu	Kecklerg (19) e	a for all is (statistics for p-regulated. ⊥ Expressed Gane or p-regulated. ⊥ D root DE is (statistics for input Files: GCT⊥ S for all (statistics for p-regulated. ⊥ and C2 of the CGP and Ucso verse. conditions C1 and cover-regulated.	Name: Status: Public Access URL: Input Parameters: Comparison Group Pairs: Meta-analysis results:	Case Study-SLE Completed Generate Differential Ge Differential Ge GSE3447-Healthy- PBMC::GSE3447- Lupus_Erythematosu s_Systemic-PBMC GSE20844 Lupus_Erythematosu s_Systemic-PBMC GSE20844 Lupus_Erythematosu s_Systemic PBMCS::GSE20844 Healthy- Lupus_Erythematosu s_Systemic SEC all Science Second Se	Gene or Probe?: me Threshold Field: ne Threshold Field: Normalized Data Normalized GEO Data Normalized GEO Data Normalized GEO Data ed' genes correspond tive to the reference of correct A meta-anal	Gene Idd 0.05 Non-reference Cond C2: GSE3447- Lupus_Enythematosu s_Systemic-PEMCG C1: GSE20864- Lupus_Enythematosu s_Systemic-PEMCS C1: GSE20872- C1: GSE50772-SLE1 It to those with higher exproondition and vice versa for condition and vice versa for system service versa for power the gene list using " ysis results.	ItilBaference co C1: GSE3447- Healthy-PBMC C2: GSE20864 Healthy- Lupus_Eryther sSystemic C2: GSE50772 Healthy1 C2: GSE50772 Healthy1 ression in the non-reg more CCP infort

Protocol



Figure 5. Access analysis results

(A) Users can view results of any analyses under the "Analysis Results" tab in a compendium.

(B) For DEP analysis, users can download results for each individual CGP by clicking on the black download icons corresponding to each type of result. If multiple CGPs were included, additional results are available.

(C) For meta-analysis, users can download the output file of the chosen meta-analysis method, as well as lists of differentially up-regulated or downregulated genes or probes.

Access and interpret analysis results

© Timing: 10 min

This step describes how to view and download the results from both differential expression analysis and meta-analysis. The outputs of both analyses are described here.

- 50. On the "Compendium" page, click on the "Analysis Results" tab to see all analyses performed for a compendium (Figure 5A).
- 51. Filter analyses by run type (Derive DEPs or Meta-analysis) or run status (queued, running, completed, failed).
- 52. Click on a run ID to see the detailed results for those analyses.
- 53. Click the green "Generate" button in the "Public Access URL" field to generate a URL to share the analysis results.
- 54. For DEP analysis, download results for each CGP or the DEP matrix (Figure 5B).
 - a. Click the black download icon next to each output type to download the results for each CGP.
 - b. Click the heatmap icon to view a heatmap of differentially expressed genes in a pop-up window.
 - c. These are the outputs (Table 2):
 - d. Click the green "Download CGP Information" button to download an Excel file with information about each CGP used in the analysis (e.g., the number of subjects, platform, samples included).
 - e. Click the green "Download Zip File For" button and select to download data for all differential expression profiles, all differentially expressed genes or probes, all heatmaps, or all GenePattern input files.
 - f. Additional outputs are available if two or more CGPs are included. Click the black download icon to download files. Click on the heatmap icons to view the graphs in a pop-up window.g. These are the outputs (Table 3):
- 55. For meta-analysis, click on the black download icons to download the output of the chosen method (Figure 5C).
 - a. The output file has these fields if the user selects the RankProd method (Table 4):

Table 2. Output of DEP analysis				
Output name	Description			
Differential Expression Profile (DEPs)	A file containing analysis result statistics of each gene or probe from condition 1 vs. condition 2 comparison. The statistics returned depend on the statistical method used. For example, if using the limma method, results include log(fold change), average expression, t-statistic, p-value, adjusted p-value, and b-statistic.			
Differentially Expressed Gene or Probe List	A list of differentially expressed genes or probes determined based on the defined statistical cut-off. This list can be used to perform gene-set enrichment analysis.			
Heatmap (100 most DE genes/probes)	A heatmap of the top 100 differentially expressed genes or probes showing all samples in the CGP. The values shown in the heatmap reflect the data source the user selected. If the "one-button" approach was used, by default the "Normalized GEO Data" is shown.			
GenePattern Input Files	Two files (GCT and CLS formatted) are provided containing gene expression values and sample grouping information for down-stream analysis outside of OMiCC. This format is supported by tools such as Gene Pattern, Gene Set Enrichment Analysis, and Integrative Genomics Viewer.			

CellPress OPEN ACCESS

STAR Protocols Protocol

Table 3. Additional outputs for two or more CGPs

Output name	Description
DEP × CGP matrix file	A merged matrix of DEPs (using the user-selected statistics) with genes or probes as rows and CGPs as columns.
Heatmap (500 most varying genes/probes)	A clustered heatmap of the 500 most varying genes or probes across CGPs. The value displayed is the user-selected statistic for the DEP \times CGP matrix.
Number of Significant DE Genes per CGP Bar Plot	A bar plot showing the number of statistically significant differentially expressed genes for each CGP.

- b. The output file has these fields if the user selects the MetaIntegrator method (please see MetaIntegrator paper (Haynes et al., 2017) for more details) (Table 5):
- c. Click on the black download icons to download lists of differentially up-regulated or downregulated genes or probes.
- d. Click the green "Download CGP Information" button to download an Excel file with information about each CGP in the analysis (e.g., the number of subjects, platform, samples included).
- 56. Open any of the downloaded tables in Excel or a similar program to inspect the results of your analysis further.

RESTful API

© Timing: 10 min

This step describes how to access OMiCC data using an API. A RESTful API (application programming interface) defines a set of functions through which users can interface directly with the OMiCC software from their computer to receive information from OMiCC.

- 57. In the toolbar, click "Document" then "RESTful APIs" for a list of API functions to access OMiCC's internal data, such as GEO studies, samples, platforms, public sample groups and CGPs.
- 58. Using the APIs requires users to acquire an access token from OMiCC. Use the following URL to get the token: https://omicc.niaid.nih.gov/api/login with your username and password in the request. For details, please refer to the RESTful API documents.

EXPECTED OUTCOMES

The results from any analysis run can be accessed through the "Analysis Results" tab of a compendium. The results of DEP analysis include the differential expression profiles of each CGP, a list of differentially expressed genes, and a heatmap of the 100 most differentially expressed genes or probes. GCT and CLS files are provided that can be used in downstream tools, such as GenePattern (Reich et al., 2006), Gene Set Enrichment Analysis (GSEA), and the Integrative Genomics Viewer (IGV). In addition to the CGP-specific results, DEP analysis returns a matrix of DEP \times CGP, a heatmap of the 500 most varying genes or probes across CGPs, and a bar plot showing the number of DE genes per CGP. The results of meta-analysis include the output file specific to the meta-analysis method chosen (described above in the step-by-step method details), as well as lists of differentially expressed genes.

Table 4. Outputs using the RankProd statistical method					
Parameter	Description				
pfp	Estimated percentage of false positive (pfp) up to the position of each gene per direction of change (i.e., up-regulated and down-regulated genes). Pfp is similar to the false discovery rate (FDR).				
Pval	Estimated p-value for being up-regulated or down-regulated for each gene				
Fc.avg	Log fold change of average expression in condition 1 over average expression in condition 2				
Fc (per CGP)	Log fold change for each CGP (labeled by the name of the CGP)				



D	цė.	<u> </u>	÷	<u> </u>	~	<u> </u>	I
		U		U	<u> </u>	U	

Table 5. Outputs using the MetaIntegrator statistical method						
Parameter	Description					
effectSize	Summary effect size computed using a random effect model					
effectSizeStandardError	Standard error for the summary effect size					
effectSizePval	Given summary effect size and standard error, p-value calculated based on a standard normal distribution					
effectSizeFDR	Benjamini-Hochberg FDR correction for multiple hypothesis testing					
tauSquared	Inter-dataset variation as estimated by the DerSimonian-Laird method					
numStudies	Number of studies in which the gene was present					
cochranesQ	Cochrane's Q value for evaluating heterogeneity of effect size estimates between studies					
heterogeneityPval	p-value of Cochrane's Q calculated against a chi-squared distribution					
fisherStatUp	Log sum of p-values that each up-regulated gene using Fisher's method					
fisherPvalUp	p-value under a chi-squared distribution					
fisherFDRUp	Benjamini-Hochberg FDR correction for multiple hypothesis testing					
fisherStatDown	Log sum of p-values that each gene is down-regulated using Fisher's method					
fisherPvalDown	p-value under a chi-squared distribution					
fisherFDRDown	Benjamini-Hochberg FDR correction for multiple hypothesis testing					

An example case study on performing a meta-analysis to investigate differential gene expression between patients with systemic lupus erythematosus, also called lupus, and healthy individuals without lupus is available to users (Methods videos S1, S2, S3, and S4).

This protocol walks users through data reuse and meta-analysis on publicly available gene expression data using the OMiCC web interface. We envision that as the user base of OMiCC grows, the number of annotated studies and publicly available CGPs and compendia will grow as well, facilitating more efficient data integration. OMiCC thus enables "virtual" collaborations in the broader biomedical research community and promotes the generation of new biological insights.

QUANTIFICATION AND STATISTICAL ANALYSIS

Both microarray and RNA-seq data are available in OMiCC. For microarray data, more than 40,000 human and mouse data sets and associated metadata were retrieved from GEO and incorporated into the OMiCC database. Three microarray data types are available in OMiCC depending on data availability in GEO: GEO series matrix data set; quantile normalized version of the series matrix data set; and for Affymetrix platforms, RMA normalized data derived from the raw CEL files. By default, OMiCC uses the quantile-normalized GEO series matrix file, but the user has the option of selecting other data types. Quality control is performed on all GEO studies and samples flagged as outliers by our pipeline are removed. A major update since our initial publication is the addition of RNA-seq data to the OMiCC database from the recount2 platform (Collado-Torres et al., 2017), a resource of RNA-seq data downloaded from SRA and uniformly processed using a Rail-RNA pipeline (Nellore et al., 2017) (Nellore et al., 2017). Gene count data from recount2 is transformed from coverage to read counts, then incorporated into the OMiCC database. If the user chooses RNA-seq data to use in their analysis, TMM (trimmed mean of M-values) normalization (Robinson and Oshlack, 2010), which adjusts for differences in total RNA production, and logCPM (counts per million) are performed on the data.

There are many tools in R, Python, and other environments that can perform statistical analysis, but they often require programming on the user's part (Tseng et al., 2012). Assembling the data into data structures can be a demanding task for biologists with little computational training. OMiCC assembles these data structures for the users through an easy-to-use interface and can also perform two types of analysis, differential expression analysis and meta-analysis. Differential expression analysis is performed on each individual CGP according to the user-specified parameters. Using drop-down menus, users can choose the statistical significance testing method, multiple testing correction method, and differential gene threshold. The three significance testing methods that can be





used in OMiCC are limma (Ritchie et al., 2015), an empirical Bayesian approach; the Mann-Whitney U test, also known as the Wilcoxon rank-sum test; and the Student's t-test.

There are many methods to perform meta-analysis on gene expression data, including combining ranks, combining effect sizes, combing p-values, and directly merging raw data using advanced normalization techniques (Tseng et al., 2012). While each has its own advantages, we use two methods to incorporate into OMiCC. RankProd is an R package that extends the rank product method and detects differential expression using a non-parametric test (Hong et al., 2006). An advantage of non-parametric methods is that they do not assume normality. While microarray data is continuous and can be assumed as normal distribution after normalization, RNA-seq data is a count distribution and thus cannot be assumed as normal distribution (Tseng et al., 2012). A disadvantage of RankProd is that it does not output a combined effect size across studies. Therefore, another new feature we incorporated is the meta-analysis package MetaIntegrator, which uses a random-effects model with the assumption that the results from each study in the analysis are drawn from a single distribution and that inter-study differences are a random effect (Haynes et al., 2017). It performs a DerSimonian and Laird random-effects meta-analysis and a Fisher's sum-of-logs test between cases and controls for each study and requires that a gene is significant using both methods. MetaIntegrator returns q-values, which evaluates whether each gene is differentially expressed between cases and controls in the included studies. An advantage of MetaIntegrator is that it returns a combined effect-size. Users should be aware that MetaIntegrator was developed for microarray studies and has not been extensively tested for RNA-seq data. While OMiCC does allow users to choose RNA-seq studies to use with MetaIntegrator and performs a TMM and CPM normalization on RNA-seq data, users should use cautious when interpreting these results. The results from either meta-analysis method can be downloaded directly from the OMiCC web interface.

LIMITATIONS

While OMiCC was developed to assist biologists without computational training to analyze publicly available data, it is not meant to replace collaborations with bioinformaticians or computational biologists. It can be dangerous to cherry-pick results and statistical expertise is often necessary to interpret meta-analysis results. If in doubt, we recommend users consult with a bioinformatics expert to determine which multiple-testing correction method and statistical test for differential expression analysis are appropriate for the studies of interest.

Selecting studies to include in the analysis is an important step. Certain microarray platforms are only used in a small number of studies or cover a small number of genes and may not be easily comparable with other platforms. To assist the user in choosing studies, OMiCC highlights the most popular microarray platforms. OMiCC also contains both microarray and RNA-seq data. While OMiCC allows the comparison of data across the two technologies, users should be aware that data generated from different technologies may not be directly comparable. The meta-analysis methods RankProd and MetaIntegrator were developed for microarray studies and have not been extensively tested for RNA-seq data. While OMiCC does allow users to choose RNA-seq studies to use with these methods and performs a TMM and CPM normalization on RNA-seq data, users should use caution when interpreting these results. Users can use both RankProd and MetaIntegrator methods and look for results consistent with both approaches. Users should also be aware that different library preparation protocols may heavily influence RNA-seq results (Alberti et al., 2014; Kumar et al., 2017; Sun et al., 2013).

Metadata is often not standardized and sometimes not reported in public databases, so it may be helpful to refer to the original publication for information on experimental design and batching. Meta-analysis methods are meant to mitigate the effect of biological and technical heterogeneity across studies, and therefore results that show a coherent signal across multiple studies and platforms are more robust.

Protocol



TROUBLESHOOTING

Problem 1 The sample group cannot be edited (step 15).

Potential solution

If a sample group is public and being used by another OMiCC user, the group cannot be edited. Unfortunately, you cannot edit a sample group if it is being used by another user.

Problem 2

Some CGPs cannot be included in generation of a DEP or meta-analysis (steps 37 and 46).

Potential solution

OMiCC uses the gene symbol as a common link across studies. If the probe-to-gene map is not available, there is no way to generate a cross-study analysis. Unfortunately, there is no way to use this CGP in a cross-study analysis.

Problem 3

Error when opening text file in Excel indicating text file is SYLK file (step 56).

Potential solution

This error occurs because the text file begins with the term "ID". You can ignore this error. Click OK and proceed to inspect the file.

Problem 4

When creating a sample group, the "Save" button is opaque and cannot be clicked (step 16).

Potential solution

This error occurs because there is an insufficient number of samples in the sample group. OMiCC requires at least three samples per sample group. The only solution is to add more samples to the sample group.

Problem 5

When searching for studies, the number of displayed studies does not change when trying to limit the results to human or mouse platforms (step 4).

Potential solution

Click "Search" again after checking or unchecking boxes to identify specific platforms of interest on which to search.

Problem 6

There is no button visible to add a compendia to your personal compendia collection after searching on compendia (step 7).

Potential solution

Ensure that you are logged into OMiCC. One can search studies, sample groups, CGPs and compendia when not logged in to OMiCC, but cannot create sample groups, CGPs, or copy compendia without being logged in.

Problem 7

In the meta-analysis results, the fold-change of a gene is very different across CGPs (step 55).

CellPress



Potential solution

In the CGPs, confirm that "Condition 1" and "Condition 2" are assigned consistently across studies (step 19). When performing meta-analysis, confirm that the correct group is assigned under "Reference Condition" (step 48). Users can also try to adjust the differential gene threshold or meta-analysis method. When in doubt, consult with your local bioinformatician.

Problem 8

In the meta-analysis results, the downloaded files of up-regulated or down-regulated genes are empty (step 55c). For the differential expression analysis, the summary of the analysis run states that there are no genes or probes that passed the threshold for differential expression (step 52).

Potential solution

For the meta-analysis, these files will be empty when none of the tested genes pass the significance threshold. Download and open the meta-analysis results (steps 55 and 56) to evaluate the full results of the meta-analysis. For the differential expression analysis, download and open the "DEP statistics for all genes/probes" file (steps 54 and 56) to evaluate the full results.

RESOURCE AVAILABILITY

Lead contact

Further information and requests for resources should be directed to and will be fulfilled by the lead contact, John S. Tsang (john.tsang@nih.gov).

Materials availability

This study did not generate new unique materials or reagents.

Data and code availability

No new datasets were generated or analyzed during this study.

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at https://doi.org/10.1016/j.xpro.2022.101474.

ACKNOWLEDGMENTS

This work was supported by the Intramural Research Program of NIAID, NIH. We thank Laura Failla for her review of the paper. Graphical abstract created with BioRender.com.

AUTHOR CONTRIBUTIONS

Conceptualization, J.S.T.; Methodology, C.C.L, Y.G., W.W.L., and J.S.T.; Software, Y.G., W.W.L., and C.C.L.; Formal Analysis, Y.G., W.W.L., and C.C.L.; Investigation, C.C.L., W.W.L., K.L.V., and R.S.; Writing – Original Draft, C.C.L., W.W.L., K.L.V., and R.S.; Writing – Review & Editing, C.C.L., W.W.L., K.L.V., R.S., and J.S.T.; Visualization, C.C.L., K.L.V., and R.S.; Supervision, R.S. and J.S.T.; Funding Acquisition, J.S.T.

DECLARATION OF INTERESTS

The authors declare no competing interests.

REFERENCES

Alberti, A., Belser, C., Engelen, S., Bertrand, L., Orvain, C., Brinas, L., Cruaud, C., Giraut, L., Da Silva, C., Firmo, C., et al. (2014). Comparison of library preparation methods reveals their impact on interpretation of metatranscriptomic data. BMC Genom. *15*, 912. https://doi.org/10.1186/1471-2164-15-912. Andres-Terre, M., McGuire, H.M., Pouliot, Y., Bongen, E., Sweeney, T.E., Tato, C.M., and Khatri, P. (2015). Integrated, multi-cohort analysis identifies conserved transcriptional signatures across multiple respiratory viruses. Immunity 43, 1199–1211. https://doi.org/10.1016/j.immuni.2015. 11.003. Barrett, T., Wilhite, S.E., Ledoux, P., Evangelista, C., Kim, I.F., Tomashevsky, M., Marshall, K.A., Phillippy, K.H., Sherman, P.M., Holko, M., et al. (2013). NCBI GEO: archive for functional genomics data sets—update. Nucleic Acids Res. *41*, D991–D995. https://doi.org/10.1093/nar/ aks1193.

Chaussabel, D., and Baldwin, N. (2014). Democratizing systems immunology with modular transcriptional repertoire analyses. Nat. Rev. Immunol. 14, 271–280. https://doi.org/10.1038/ nri3642.

Chen, R., Khatri, P., Mazur, P.K., Polin, M., Zheng, Y., Vaka, D., Hoang, C.D., Shrager, J., Xu, Y., Vicent, S., et al. (2014). A meta-analysis of lung cancer gene expression identifies PTK7 as a survival gene in lung adenocarcinoma. Cancer Res. 74, 2892–2902. https://doi.org/10.1158/0008-5472.CAN-13-2775.

Collado-Torres, L., Nellore, A., Kammers, K., Ellis, S.E., Taub, M.A., Hansen, K.D., Jaffe, A.E., Langmead, B., and Leek, J.T. (2017). Reproducible RNA-seq analysis using recount2. Nat. Biotechnol. 35, 319–321. https://doi.org/10.1038/nbt.3838.

Dudley, J.T., Sirota, M., Shenoy, M., Pai, R.K., Roedder, S., Chiang, A.P., Morgan, A.A., Sarwal, M.M., Pasricha, P.J., and Butte, A.J. (2011). Computational repositioning of the anticonvulsant topiramate for inflammatory bowel disease. Sci. Transl. Med. 3, 96ra76. https://doi.org/10.1126/ scitranslmed.3002648.

Engreitz, J.M., Chen, R., Morgan, A.A., Dudley, J.T., Mallelwar, R., and Butte, A.J. (2011). ProfileChaser: searching microarray repositories based on genome-wide patterns of differential expression. Bioinformatics 27, 3317–3318. https://doi.org/10. 1093/bioinformatics/btr548.

Granlund, A.v.B., Flatberg, A., Østvik, A.E., Drozdov, I., Gustafsson, B.I., Kidd, M., Beisvag, V., Torp, S.H., Waldum, H.L., Martinsen, T.C., et al. (2013). Whole genome gene expression metaanalysis of inflammatory bowel disease colon mucosa demonstrates lack of major differences between Crohn's disease and ulcerative colitis. PLoS One 8, e56818. https://doi.org/10.1371/ journal.pone.0056818.

Haynes, W.A., Vallania, F., Liu, C., Bongen, E., Tomczak, A., Andres-Terrè, M., Lofgren, S., Tam, A., Deisseroth, C.A., Li, M.D., et al. (2017). Empowering multi-cohort gene expression analysis to increase reproducibility. Pac. Symp. Biocomput. Pac. Symp. Biocomput. 22, 144–153. https://doi. org/10.1142/9789813207813_0015.

Hong, F., Breitling, R., McEntee, C.W., Wittner, B.S., Nemhauser, J.L., and Chory, J. (2006). RankProd: a bioconductor package for detecting differentially expressed genes in meta-analysis. Bioinformatics 22, 2825–2827. https://doi.org/10. 1093/bioinformatics/btl476.

Wliev, A.E., Hoen, P.A.C.'t, Villerius, M.P., den Dunnen, J.T., and Brandt, B.W. (2008). Microarray retriever: a web-based tool for searching and large scale retrieval of public microarray data. Nucleic Acids Res. 36, W327–W331. https://doi.org/10. 1093/nar/gkn213.

Khatri, P., Roedder, S., Kimura, N., De Vusser, K., Morgan, A.A., Gong, Y., Fischbein, M.P., Robbins, R.C., Naesens, M., Butte, A.J., and Sarwal, M.M. (2013). A common rejection module (CRM) for acute rejection across multiple organs identifies novel therapeutics for organ transplantation. J. Exp. Med. 210, 2205–2221. https://doi.org/10. 1084/jem.20122709.

Kuehn, H., Liberzon, A., Reich, M., and Mesirov, J.P. (2008). Using GenePattern for gene expression analysis. Curr. Protoc. Bioinforma. *Chapter 7*, Unit 7.12. https://doi.org/10.1002/0471250953. bi0712s22.

Kumar, A., Kankainen, M., Parsons, A., Kallioniemi, O., Mattila, P., and Heckman, C.A. (2017). The impact of RNA sequence library construction protocols on transcriptomic profiling of leukemia. BMC Genom. 18, 629. https://doi.org/10.1186/ s12864-017-4039-1.

Lau, W.W., Sparks, R.; OMiCC Jamboree Working Group, and Tsang, J.S. (2016). Meta-analysis of crowdsourced data compendia suggests pandisease transcriptional signatures of autoimmunity. F1000Research 5, 2884. https://doi.org/10.12688/ f1000research.10465.1.

Leinonen, R., Sugawara, H., and Shumway, M. (2011). The sequence read archive. Nucleic Acids Res. 39, D19–D21. https://doi.org/10.1093/nar/ gkq1019.

Nellore, A., Collado-Torres, L., Jaffe, A.E., Alquicira-Hernández, J., Wilks, C., Pritt, J., Morton, J., Leek, J.T., and Langmead, B. (2017). Rail-RNA: scalable analysis of RNA-seq splicing and coverage. Bioinformatics 33, 4033–4040. https:// doi.org/10.1093/bioinformatics/btw575.

Ramasamy, A., Mondry, A., Holmes, C.C., and Altman, D.G. (2008). Key issues in conducting a meta-analysis of gene expression microarray datasets. PLoS Med. 5, e184. https://doi.org/10. 1371/journal.pmed.0050184.

Reich, M., Liefeld, T., Gould, J., Lerner, J., Tamayo, P., and Mesirov, J.P. (2006). GenePattern 2.0. Nat. Genet. *38*, 500–501. https://doi.org/10.1038/ ng0506-500.

Ritchie, M.E., Phipson, B., Wu, D., Hu, Y., Law, C.W., Shi, W., and Smyth, G.K. (2015). Limma powers differential expression analyses for RNAsequencing and microarray studies. Nucleic Acids Res. 43, e47. https://doi.org/10.1093/nar/gkv007.

Robinson, M.D., and Oshlack, A. (2010). A scaling normalization method for differential expression analysis of RNA-seq data. Genome Biol. 11, R25. https://doi.org/10.1186/gb-2010-11-3-r25.

Rung, J., and Brazma, A. (2013). Reuse of public genome-wide gene expression data. Nat. Rev. Genet. 14, 89–99. https://doi.org/10.1038/nrg3394.

Rustici, G., Kolesnikov, N., Brandizi, M., Burdett, T., Dylag, M., Emam, I., Farne, A., Hastings, E., Ison, J., Keays, M., et al. (2013). ArrayExpress update trends in database growth and links to data analysis tools. Nucleic Acids Res. 41, D987–D990. https:// doi.org/10.1093/nar/gks1174.

Segal, E., Friedman, N., Kaminski, N., Regev, A., and Koller, D. (2005). From signatures to models: understanding cancer using microarrays. Nat. Genet. 37, S38–S45. https://doi.org/10.1038/ ng1561.

Shah, N., Guo, Y., Wendelsdorf, K.V., Lu, Y., Sparks, R., and Tsang, J.S. (2016). A crowdsourcing approach for reusing and meta-analyzing gene expression data. Nat. Biotechnol. *34*, 803–806. https://doi.org/10.1038/nbt.3603.

Sirota, M., Dudley, J.T., Kim, J., Chiang, A.P., Morgan, A.A., Sweet-Cordero, A., Sage, J., and Butte, A.J. (2011). Discovery and preclinical validation of drug indications using compendia of public gene expression data. Sci. Transl. Med. 3, 96ra77. https://doi.org/10.1126/scitranslmed. 3001318.

Sparks, R., Lau, W.W., and Tsang, J.S. (2016). Expanding the immunology toolbox: embracing public-data reuse and crowdsourcing. Immunity 45, 1191–1204. https://doi.org/10.1016/j.immuni. 2016.12.008.

Sun, Z., Asmann, Y.W., Nair, A., Zhang, Y., Wang, L., Kalari, K.R., Bhagwate, A.V., Baker, T.R., Carr, J.M., Kocher, J.-P.A., et al. (2013). Impact of library preparation on downstream analysis and interpretation of RNA-seq data: comparison between illumina PolyA and NuGEN ovation protocol. PLoS One *8*, e71745. https://doi.org/10. 1371/journal.pone.0071745.

Sweeney, T.E., Shidham, A., Wong, H.R., and Khatri, P. (2015). A comprehensive time-coursebased multicohort analysis of sepsis and sterile inflammation reveals a robust diagnostic gene set. Sci. Transl. Med. 7, 287ra71. https://doi.org/10. 1126/scitranslmed.aaa5993.

Teslovich, T.M., Musunuru, K., Smith, A.V., Edmondson, A.C., Stylianou, I.M., Koseki, M., Pirruccello, J.P., Ripatti, S., Chasman, D.I., Willer, C.J., et al. (2010). Biological, clinical and population relevance of 95 loci for blood lipids. Nature 466, 707–713. https://doi.org/10.1038/nature09270.

Tseng, G.C., Ghosh, D., and Feingold, E. (2012). Comprehensive literature review and statistical considerations for microarray meta-analysis. Nucleic Acids Res. 40, 3785–3799. https://doi.org/ 10.1093/nar/gkr1265.

Wang, Z., Monteiro, C.D., Jagodnik, K.M., Fernandez, N.F., Gundersen, G.W., Rouillard, A.D., Jenkins, S.L., Feldmann, A.S., Hu, K.S., McDermott, M.G., et al. (2016). Extraction and analysis of signatures from the gene expression Omnibus by the crowd. Nat. Commun. 7, 12846. https://doi. org/10.1038/ncomms12846.

Xia, J., Gill, E.E., and Hancock, R.E.W. (2015). NetworkAnalyst for statistical, visual and networkbased meta-analysis of gene expression data. Nat. Protoc. 10, 823–844. https://doi.org/10.1038/nprot. 2015.052.

Zinman, G.E., Naiman, S., Kanfi, Y., Cohen, H., and Bar-Joseph, Z. (2013). ExpressionBlast: mining large, unstructured expression databases. Nat. Methods 10, 925–926. https://doi.org/10.1038/ nmeth.2630.

