# Bioinformatics pipelines, workflows, and resources

Marylyn D. Ritchie, PhD Professor, Biochemistry & Molecular Biology Center for Systems Genomics The Pennsylvania State University







# **GWAS: Genome-Wide Association Study**

Genome-wide association analysis of QRS duration using sex-adjusted linear regression





# PheWAS: Phenome-Wide Association Study

Genome-wide association analysis of QRS duration using sex-adjusted linear regression

Correso

2



Phenome-wide association analysis of



# PheWAS: Phenome-Wide Association Study

Genome-wide association analysis of QRS duration using sex-adjusted linear regression

Phenome-wide association analysis of SCN5A/SCN10A using logistic regression



# PheWAS

- Comprehensively calculating the association between:
  - Small to large number of SNPs
  - Diverse range of phenotypes from phenotypically rich datasets
- PheWAS has the potential to:
  - Provide novel mechanistic insights
  - Foster hypothesis generation
  - Identify pleiotropy
- PheWAS can be performed as part of:
  - Epidemiology cohort studies
  - Animal breeding studies
  - Prospective clinical trials
  - DNA Biobanks linked to electronic health records

### The Use of Phenome-Wide Association Studies (PheWAS) for Exploration of Novel Genotype-Phenotype Relationships and Pleiotropy Discovery

S.A. Pendergrass,<sup>1</sup> K. Brown-Gentry,<sup>1</sup> S.M. Dudek,<sup>1</sup> E.S. Torstenson,<sup>1</sup> J.L. Ambite,<sup>2</sup> C.L. Avery,<sup>3</sup>
S. Buyske,<sup>4,5</sup> C. Cai,<sup>2</sup> M.D. Fesinmeyer,<sup>6</sup> C. Haiman,<sup>7</sup> G. Heiss,<sup>3</sup> L.A. Hindorff,<sup>8</sup> C.-N. Hsu,<sup>2</sup> R.D. Jackson,<sup>9</sup>
C. Kooperberg,<sup>6</sup> L. Le Marchand,<sup>10</sup> Y. Lin,<sup>6</sup> T.C. Matise,<sup>5</sup> L. Moreland,<sup>11</sup> K. Monroe,<sup>7</sup> A.P. Reiner,<sup>6,12</sup>
R. Wallace,<sup>13</sup> L.R. Wilkens,<sup>10</sup> D.C. Crawford,<sup>1,14</sup> and M.D. Ritchie<sup>1,14\*</sup>

- Analysis pipeline for PheWAS
- Establish phenotype classes for analysis
- Data visualization for synthesizing results
- Challenges in interpreting results
  - Multiple testing
  - Correlations between phenotypes (as well as LD)
  - Replication suggested to avoid false positive results



# PheWAS in Epidemiology Cohorts

OPEN O ACCESS Freely available online

PLOS GENETICS

#### Phenome-Wide Association Study (PheWAS) for Detection of Pleiotropy within the Population Architecture using Genomics and Epidemiology (PAGE) Network

Sarah A. Pendergrass<sup>1</sup>, Kristin Brown-Gentry<sup>2</sup>, Scott Dudek<sup>2</sup>, Alex Frase<sup>1</sup>, Eric S. Torstenson<sup>2</sup>, Robert Goodloe<sup>2</sup>, Jose Luis Ambite<sup>3</sup>, Christy L. Avery<sup>4</sup>, Steve Buyske<sup>5,6</sup>, Petra Bůžková<sup>7</sup>, Ewa Deelman<sup>3</sup>, Megan D. Fesinmeyer<sup>8</sup>, Christopher A. Haiman<sup>9</sup>, Gerardo Heiss<sup>4</sup>, Lucia A. Hindorff<sup>10</sup>, Chu-Nan Hsu<sup>3</sup>, Rebecca D. Jackson<sup>11</sup>, Charles Kooperberg<sup>8</sup>, Loic Le Marchand<sup>12</sup>, Yi Lin<sup>8</sup>, Tara C. Matise<sup>5</sup>, Kristine R. Monroe<sup>9</sup>, Larry Moreland<sup>13</sup>, Sungshim L. Park<sup>12</sup>, Alex Reiner<sup>8,14</sup>, Robert Wallace<sup>15</sup>, Lynn R. Wilkens<sup>12</sup>, Dana C. Crawford<sup>2,16</sup>, Marylyn D. Ritchie<sup>1</sup>\*



			Maximum Sample Size <sup>3</sup>				Minimum Sample Size <sup>3</sup>					
PMID <sup>2</sup>	Age Range	Sex	EA	AA	н	ΑΡΙ	EA	AA	н	ΑΡΙ	# SNPs⁴	# Phenotypes⁵
[36]	45–64	M/F	11,068	4,007	NA	NA	17	7	NA	NA	69	612
[37]	65–100 at baseline	M/F	4,487	820	NA	NA	151	116	NA	NA	46	341
[38]	12–95	M/F	2,628	2,107	2,071	NA	7	16	15	NA	236	327
[39]	45–75	M/F	3,893	4,749	6,863	6,810	33	27	40	13	74	63
[40]	50–79 at baseline	F	13,334	4,274	2,023	927	14	5	7	5	94	3,363
	PMID <sup>2</sup> [36] [37] [38] [39] [40]	Age Range           [36]         45-64           [37]         65-100 at baseline           [38]         12-95           [39]         45-75           [40]         50-79 at baseline	Age Range         Sex           [36]         45–64         M/F           [37]         65–100 at baseline         M/F           [38]         12–95         M/F           [39]         45–75         M/F           [40]         50–79 at baseline         F	Age Range         Sex         Maximum           [36]         45-64         M/F         11,068           [37]         65-100 at baseline         M/F         4,487           [38]         12-95         M/F         2,628           [39]         45-75         M/F         3,893           [40]         50-79 at baseline         F         13,334	Age RangeSexMaximum Sample[36]45-64M/F11,0684,007[37]65-100 at baselineM/F4,487820[38]12-95M/F2,6282,107[39]45-75M/F3,8934,749[40]50-79 at baselineF13,3344,274	Age RangeSexMaximum Sample Size³[36]45-64M/F11,0684,007NA[37]65-100 at baselineM/F4,487820NA[38]12-95M/F2,6282,1072,071[39]45-75M/F3,8934,7496,863[40]50-79 at baselineF13,3344,2742,023	Age RangeSexEAAAHAPI[36]45-64M/F11,0684,007NANA[37]65-100 at baselineM/F4,487820NANA[38]12-95M/F2,6282,1072,071NA[39]45-75M/F3,8934,7496,8636,810[40]50-79 at baselineF13,3344,2742,023927	PMID2Age RangeSexEAAAHAPIEA[36]45-64M/F11,0684,007NANA17[37]65-100 at baselineM/F4,487820NANA151[38]12-95M/F2,6282,1072,071NA7[39]45-75M/F3,8934,7496,8636,81033[40]50-79 at baselineF13,3344,2742,02392714	PMID²Age RangeSexEAAAHAPIEAAA[36]45-64M/F11,0684,007NANA177[37]65-100 at baselineM/F4,487820NANA151116[38]12-95M/F2,6282,1072,071NA716[39]45-75M/F3,8934,7496,8636,8103327[40]50-79 at baselineF13,3344,2742,023927145	Maximum Sample Size3Minimum Sample Size3Minimum Sample Size3PMID2Age RangeSexEAAAHAPIEAAAH[36] $45-64$ M/F $11,068$ $4,007$ NANA177NA[37] $65-100$ at baselineM/F $4,487$ $820$ NANA151116NA[38] $12-95$ M/F $2,628$ $2,107$ $2,071$ NA71615[39] $45-75$ M/F $3,893$ $4,749$ $6,863$ $6,810$ $33$ 2740[40] $50-79$ at baselineF $13,334$ $4,274$ $2,023$ $927$ 14 $5$ $7$	MaximuSampleSize <sup>3</sup> MiniSumpleSize <sup>3</sup> PMID <sup>2</sup> Age RangeSexEAAAHAPIEAAAHAPI[36]45-64M/F11,0684,007NANA177NANA[37]65-100 at baselineM/F4,487820NANA151116NANA[38]12-95M/F2,6282,1072,071NA71615NA[39]45-75M/F3,8934,7496,8636,81033274013[40]50-79 at baselineF13,3344,2742,02392714575	PMID2Age RangeSexEAAAHAPIEAAAHAPI# \$NP\$4[36]45-64M/F11,0684,007NANA177NANA69[37]65-100 at baselineM/F4,487820NANA151116NA40[38]12-95M/F2,6282,1072,071NA71615NA236[39]45-75M/F3,8934,7496,8636,8103327401374[40]50-79 at baselineF13,3344,2742,0239271457594

# APOE/APOC1/C1P1/C2/C4, rs4420638, Coded Allele A



SNP rs4420638 has previously been associated with LDL cholesterol levels, triglycerides, Alzheimer's disease, coronary artery disease

Pendergrass et al. PLoS Genet. 2013;9(1):e1003087

Plot generated by PheWAS-View Diabetes. 2014 Sep;63(9):3149-58. doi: 10.2337/db13-1800. Epub 2014 Apr 10.

#### Pleiotropic Effects of Lipid Genes on Plasma Glucose, HbA1c, and HOMA-IR Levels.

Li N<sup>1</sup>, van der Sijde MR<sup>2</sup>; LifeLines Cohort Study Group, Bakker SJ<sup>3</sup>, Dullaart RP<sup>4</sup>, van der Harst P<sup>5</sup>, Gansevoort RT<sup>3</sup>, Elbers CC<sup>6</sup>, Wijmenga C<sup>2</sup>, Snieder H<sup>7</sup>, Hofker MH<sup>8</sup>, Fu J<sup>9</sup>.

#### Author information

#### Abstract

Dyslipidemia is strongly associated with raised plasma glucose levels and insulin resistance (IR), and genome-wide association studies have identified 95 loci that explain a substantial proportion of the variance in blood lipids. However, the loci's effects on glucose-related traits are largely unknown. We have studied these lipid loci and tested their association collectively and individually with fasting plasma glucose (FPG), glycated hemoglobin (HbA1c), and IR in two independent cohorts: 10,995 subjects from LifeLines Cohort Study and 2,438 subjects from Prevention of Renal and Vascular Endstage Disease (PREVEND) study. In contrast to the positive relationship between dyslipidemia and glucose traits, the genetic predisposition to dyslipidemia showed a pleiotropic lowering effect on glucose traits. Specifically, the genetic risk score related to higher triglyceride level was correlated with lower levels of FPG (P =  $9.6 \times 10(-10)$  and P = 0.03 in LifeLines and PREVEND, respectively), HbA1c (P =  $4.2 \times 10(-7)$  in LifeLines), and HOMA of estimated IR (P =  $6.2 \times 10(-4)$  in PREVEND), after adjusting for blood lipid levels. At the single nucleotide polymorphism level, 15 lipid loci showed a pleiotropic association with glucose traits (P < 0.01), of which eight (CETP, MLXIPL, PLTP, GCKR, APOB, APOE-C1-C2, CYP7A1, and TIMD4) had opposite allelic directions of effect on dyslipidemia and glucose levels. Our findings suggest a complex genetic regulation and metabolic interplay between lipids and glucose.

© 2014 by the American Diabetes Association. Readers may use this article as long as the work is properly cited, the use is educational and not for profit, and the work is not altered.

PMID: 24722249 [PubMed - in process]

### GALNT2, rs2144300, Coded Allele C



SNP published for HDL cholesterol and triglycerides in 2008 SNP published for circulating myeloperoxidase levels (serum) in 2013

Pendergrass et al. PLoS Genet. 2013;9(1):e1003087

Plot generated by PheWAS-View

# PheWAS

- Comprehensively calculating the association between:
  - Small to large number of SNPs
  - Diverse range of phenotypes from phenotypically rich datasets
- PheWAS has the potential to:
  - Provide novel mechanistic insights
  - Foster hypothesis generation
  - Identify pleiotropy
- PheWAS can be performed as part of:
  - Epidemiology cohort studies
  - Animal breeding studies
  - Prospective clinical trials
  - DNA Biobanks linked to electronic health records

# Electronic Health Record Based PheWAS

- Electronic Health Record Data (EHR)
  - Billing codes
    - International Statistical Classification of Diseases and Related Health Problems (ICD-9 codes)
    - Can create "case/control" status based on the presence/absence of diagnosis ICD-9 codes
      - Two instances of an ICD-9 code = case, otherwise control
  - Clinically measured lab variables

Robust replication of genotype-phenotype associations across multiple diseases in an electronic medical record. Ritchie et al. *AJHG* 2010; *PMID:* 20362271

PheWAS: demonstrating the feasibility of a phenome-wide scan to discover gene-disease associations Denny, Ritchie et al. *Bioinformatics* 2011; *PMID:* 20335276



Part of the eMERGE network (Electronic Medical Records in Genomics)

- Vanderbilt University *BioVU*
- Geisinger Health System *MyCode*

# Electronic Health Record Based PheWAS

IMPUTE2 Im	nputed Data
Dataset	# Genotyped Samples
Merged eMERGE-I 1M	2,634
Merged eMERGE-I 660	16,029
Geisinger	3,111
Group Health/U Wash	731
Marshfield/Essentia/PSU	616
Mayo Clinic	3,121
Mount Sinai	6,290
Northwestern	2,951
Vanderbilt	7,616
CCHMC/BCH	5,346
СНОР	6,850
Total – All IMPUTE2 Imputed Samples	55,292





### Electronic Health Record Based PheWAS - ImmunoPheWAS

	Genotype Da	ita		Imputed Data			Post-Imputation Quality Check				Filter	ed Genotype	e Data	
	MyCode	BioVU			MyCode BioVU			- Sex Check					MyCode	BioVU
Samples	3,111	3,375		Samples	3,111	- Marker Call Rate (99%) - Sample Call Rate (99%) Samples			Samples	3,029	2,900			
SNPs	729,078	558,980		SNPs	38,054,243	38,041,351		- MAF Filte - IBD estim	r (>0.01) ation			SNPs	4,636,178	4,163,988
													1	
	Phenotype D	ata		Quality Check			Filtere	d Phenotyp	e Data		J	Final Datase	t	
	MyCode	BioVU		- Only Diag	nosis Code				MyCode	BioVU			MyCode	BioVU
Samples	4,065	10,819		- Assigning	Case-Contr	ol Status		Samples	3,024	2,899		Samples	3,024	2,899
ICD9 Code	es 6,525	1,206		- Intersectio Phenotype s	on of Genoty samples	/pe and		ICD9 Codes	477	381		SNPs	95,448	87,690
L		<u> </u>	1											
		PheWAS	Result	ts					Re	esult Analys	sis and	Visualizatio	'n	
Replication Criteria :				- 222 association results replicating for same ICD9 code										

Synthesis-View

Phenogram

- 3,317 association results replicating from the same ICD9 code group.

Replication with GWAS Catalog

Pleiotropy

Verma A, Okula A, Pendergrass SA, in preparation

P-value < 0.01

Same coded allele

Same direction of effect

Same ICD9 Code or group



rs8111071 Paroxysmal tachycardia (Sudden cardiac arrest)

rs2277862 Heart failure (Total cholesterol)

rs1008953 Chronic bronchitis (Psoriasis)

46307406

34152782

43980726

Plot generated by Synthesis-View



rs2277862 Heart failure (Total cholesterol)

rs1008953 Chronic bronchitis (Psoriasis)

34152782 24 43980726 Plot generated by Synthesis-View

### Immune/Autoimmune Diagnosis Code Replications



# Evidence of pleiotropy - ImmunoPheWAS



Verma A, Okula A, Pendergrass SA, in preparation

Plot generated by Phenogram

# EHR-only PheWAS

- PheWAS does not REQUIRE genotype data
- EHR phenotypes/variables can be the independent variables → associate with one or more clinical outcomes of interest





# EHR-only PheWAS

- PheWAS does not REQUIRE genotype data
- EHR phenotypes/variables can be the independent variables → associate with one or more clinical outcomes of interest
- Use statistical or machine learning methods to look for patterns of clinical variables that are predictive of adverse clinical outcomes
- Use one EHR as a training set and others as validation

# Building pipelines and workflows

- Commercial and free software exists for setting up pipelines and workflows
- Important considerations:
  - Will I be running the same thing every time?
    - You do not want to do the same thing repeatedly without some automation in the pipeline.
  - How much human interaction is needed during the pipeline?
    - Will you need to implement stopping points to look at the data?
  - Who all will be using this pipeline?
    - Is it for one person, a lab, a unit, an institution?

# Galaxy

http://main.g2.bx.psu.edu/

#### 💳 Galaxy

### Data intensive biology for everyone.

<u>Galaxy</u> is an open, web-based platform for data intensive biomedical research. Whether on the <u>free public server</u> or <u>your own instance</u>, you can perform, reproduce, and share complete analyses.



Use the free public server



Install <u>locally</u> or in the cloud Learn Galaxy



Screencasts, Galaxy 101, ....



And the second second second second	-				
- Ball-Forums & Taylor (1997)		-		Last Past	-
Managers and the second state of the second	0		1	Lines to New Call	
- All the second	0		4	Labor In Convertient	1.1
A Disettist and addition from report of the	0		4	Author to long not matter	
A state of sector of state Text	-		4	Line to Derror Salar	
A Draw Come Conference by Contract Autom			1	3.30et to Ontrop Address	
A Name Agent Agent and in Preserv, Mar	0		4	Rep. 22 To Common Station	1.1
And the log to the Color	0		4	Regular to Chernese States?	
<ul> <li>Bustless and solving with ref. current solutions, and solving instanting for busing later.</li> </ul>			4	MAX In Concessions	
A family only only only on the	0		Å	Tel. III in Paul Contin-	
<ul> <li>ante lastes la regiliadas la</li> <li>Andres Sinte</li> </ul>	0		٨	BALK IN TANK, MAR	
<ul> <li>Rest Galaxy have a lost in cards would be failed.</li> </ul>	0		٨	No.10 19 19 19	
Martin And In No. 211			1	Autor Contraction	

Mailing lists, Tool Shed, wiki

The <u>Galaxy Team</u> is a part of <u>BX</u> at <u>Penn State</u>, and the <u>Biology</u> and <u>Mathematics and Computer</u> <u>Science</u> departments at <u>Emory University</u>. The Galaxy Project is supported in part by <u>NSF</u>, NHGRI, The Huck Institutes of the Life Sciences, The Institute for CyberScience at Penn State,



**Taverna** is an open source and domain-independent Workflow Management System – a suite of tools used to design and execute scientific workflows and aid *in silico* experimentation.

Taverna has been created by the myGrid team and is currently funded though FP7 projects BioVeL, SCAPE and Wf4Ever.



#### http://www.taverna.org.uk/



### Taverna

Introduction Doc

Documentation

Download

Developers Cite

Google<sup>™</sup> Custom Search

News

Collaborations

About

#### Introduction

- Taverna features
- myExperiment integration
- BioCatalogue integration
- Why use workflows?
- What is 'in silico' experimentation?
- What is a Workflow Management System?
- How can Taverna help me?
- Services in Taverna
- Taverna in use
- By domain
- As a server
- On a grid
- On a cloud
- Behind a portal
- Bundled with products
- Taverna in use chronology
- Related projects
- Works with...
  - Bioconductor
  - SoapLab
  - R
  - SADI
  - Biomart

### Genome and gene expression

Home » Introduction » Taverna in use » Genome and gene expression

Taverna is being used by various projects and researchers for the analysis of genomes and gene expression. These include:

- Next Generation Sequencing using Taverna 2 Server on Amazon cloud
- TavernaPBS next generation sequencing analysis using a computational cluster that uses a PBS queuing system and Taverna 2 Workbench
- Coordination and Sustainability of International Mouse Informatics Resources (CASIMIR) workflows to associate mouse genome and phenome
- Developmental Gene Expression Map (DGEMap)- analysis of human gene expression during development
- · Graves disease identification of genes responsible
- MicroArray analysis using R statistical analysis of gene expression
- · SIGENAE development of workflows to analyse breeding animal data
- Trypanosomiasis identification of genes responsible for sleeping sickness
- Williams-Beuren syndrome automation and confirmation of gene characterization
- Integration of plant genome resources (PLANET)
- Annotation of genomes
- Shared Genomics

TAVA	XY				contact us at support@tavaxy.org
About	Use Tavaxy	Help and Resources	Software Packages	Changes and updates	

Tavaxy is a pattern based workflow system for the bioinformatics domain, focusing on genome comparison and sequence analysis. The basic motivation behind developing Tavaxy is to enable the design of workflows composed of Taverna and Galaxy sub-workflows in addition to other workflow nodes within a single environment. This provides an easy solution to run parts of the workflow on local infrastructure and other parts remotely through web-services. Tavaxy uses the concept of workflow patterns, which are workflow language constructs, in order to facilitate the design and execution of workflows and to enable the integration of Taverna and Galaxy workflows.

#### The basic features of Tavaxy include

#### • Interface

- · A simple and intuitive user interface. The workflow components are similar to that of flowchart.
- · User can choose to run their tools either on local infrastructure or remotely using web-services or cloud computing
- · User can easily set the parameters of the tools and even change them in the running time

#### Set of advance workflow patterns

- Tavaxy supports control and data patterns (constructs) that facilitate the design and execution of the workflow.
- The pattern set includes, among others, conditionals, iterations, sub-workflows, lists, data selects, data merges.

#### Integration of Taverna and Galaxy workflows

- · Workflows written in Taverna or in Galaxy can be imported and rendered in Tavaxy
- The user can compose and execute a hybrid workflow composed of Taverna and Galaxy subwrkflows in addition to Tavaxy workflows.
- Remote calls in Taverna workflow can be automatically replaced with calls using locally installed tools.
- A repository of pre-imported Taverna and Galaxy workflows, where the user can use these workflows or integrate them in a new workflow.

#### Use of cloud computing

- Tavaxy provides three modes of using cloud computing
  - Whole system instantiation, where Galaxy as a whole is hosted on cloud computing infrastructure (currently Amazon Web Service). The system can
    automatically configure a computer cluster for the user.
  - Sub-workflow instantiation, where the user has a Tavaxy version on his machine and likes to delegate the execution of sub-workflow in the run time to the cloud

#### http://www.tavaxy.org/

# **Getting Genetics Done**

Getting Things Done in Genetics & Bioinformatics Research

#### About GGD

Many resources (e.g. Nature Reviews Genetics) offer a 10,000-foot view of the current trends in the field, reviews of various technologies, and guidelines on how to effectively design, analyze, and interpret experiments in human genetics and bioinformatics research. By comparison very few resources focus on the mundane, yet critical know-how for those on the ground actually *doing* the science (i.e. grad students, postdocs, analysts, and junior faculty). Getting Genetics Done aims to fill that gap by featuring software, code snippets, literature of interest, workflow philosophy, and anything else that can boost productivity and simplify getting things done in human genetics research.

I highly encourage you to leave comments. You can comment on any post without having to register, so please share your thoughts with us and other readers. By its very nature, there may be much easier or better ways to do things than what is covered here, so feel free to share your own ideas.

All the written content is licensed under cc-by-sa and all the code here is open source (GPLv2). So feel free to wholesale copy, modify, share, or redistribute anything you find here, just include a link back to the site.

GettingGeneticsDone.blogspot.com can also be reached by going to GettingGeneticsDone.com. You can keep up with GGD by signing up for an e-mail whenever new content is posted, subscribing to the RSS feed, or following @genetics\_blog on Twitter. If you would like to contribute, or if you have any topics, suggestions, praise, complaints, or anything else you'd like to share, please email me.

Thank you, and welcome!

#### Stephen Turner

#### SEARCH GGD

Search

#### THE GGD TEAM



Stephen Turner is an assistant professor of public health sciences and director of the Bioinformatics Core at the

University of Virginia.

Will Bush is an assistant professor in the Vanderbilt University Center for Human Genetics Research

About GGD @genetics\_blog on Twitter Subscribe via e-mail Subscribe via RSS

#### ARCHIVE

2014 (20)

■ Contombor (2)

http://gettinggeneticsdone.blogspot.com



#### https://www.biostars.org/



BIOO SCIENTIFIC'	Fighting to Reduce Small RNA-Seq Bias	
LEARN MORE RAN	DOMIZED ADAPTERS REDUCE LIGATION BIAS	

•

👕 SEQanswers Home				User Name User Name	Jser Name Remember Me
Register	FAQ	Community 🔻	Calendar	Today's Posts	Search

You are currently viewing the SEQanswers forums as a guest, which limits your access. <u>Click here to register now</u>, and join the discussion

» Site Navigation	» New Posts				8	» Our Sponsors	۲
» About SEQanswers	Title, Username, & Date	Last Post	Replies	Views	Forum	100 m	
» <u>Next Gen Summaries</u> » <mark>Forums</mark>	Problems with Bio-linux/Ubuntu 14.04 () Kimist99	Today 01:14 AM by <u>dpryan</u> 🔊	4	139	Bioinformatics	NEBNext®	
» <u>Wiki</u>	heatmap.2 for two samples without replication @ dwb0211	Today 12:10 AM by paolo.kunder 🔊	1	58	Bioinformatics	for Small RNA	1
» Instrument Map	<u>subread</u> hanshart	Yesterday 09:54 PM by <u>ndaniel</u> 🔊	31	3,645	General		
» Log in 🛞	iprlookup parameter in interproscan5 ?! bongbimit	Yesterday 09:54 PM by <u>bongbimit</u> 🔊	0	63	Bioinformatics	Low input high yield	
User Name:	CLC Genomics Workbench slow in de novo assembly jyuems	Yesterday 09:40 PM by <u>luc</u> 🔊	5	2,492	Complete Genomics		ΞN
Password:	<u>simulate illumina WGS run</u> nheb	Yesterday 09:24 PM by Luc 🔊	6	210	Bioinformatics	Your	-

### http://seqanswers.com/



### PENN STATE | ONLINE

Search

GO

ABOUT US HOME

DEGREES AND CERTIFICATES

HOW ONLINE LEARNING WORKS | ADMISSIONS | TUITION AND FINANCIAL AID | INFO FOR

### Graduate Certificate in Applied Bioinformatics

**REQUEST INFO** 

APPLY NOW

Home » Degrees and Certificates » Graduate Certificate in Applied Bioinformatics » Overview

#### **OVERVIEW**

~	0			~	-	~
C	( )		ы	~	⊢	~
~	$\sim$	v	1.	-	-	9

COSTS

HOW TO APPLY

#### CONTACT THE PROGRAM

World Campus was very inclusive *—there were people constantly* checking in with me. It was a very pleasant surprise!

Angelik Leslie Bachelor of Arts in Law and Conintu



#### Lead the Application of Genomic Information

With ongoing public attention on the Human Genome Project, understanding, analyzing, and interpreting genomic data continue to be a focus across many disciplines. By merging biology, computer science, and information technology, bioinformatics allows us to combine mathematics and computers to gain a better understanding of biological processes.

Program Summar	y 🗢
Credits Required	11
Tuition per Credit	\$930



KRIEGER SCHOOL of ARTS & SCIENCES ADVANCED ACADEMIC PROGRAMS

**Bioinformatics** 

Home Tools Contact Us

L.

+ Request Info

+ Share This Page

SEARCH

ABOUT US

ACADEMICS PROSPECTIVE STUDENTS

CURRENT STUDENTS

ATS ALUMNI

FACULTY RESOURCES

Giving

#### < Back to Graduate Degree Programs

Admissions Requirements Degree Requirements Course Descriptions Tuition and Fees Career Opportunities FAQ

Course Schedule Our Community The Experience

> Request Information

- > Contact Us
- > Apply Now



Click here to RSVP for the Nov. 11 Online Open House

Johns Hopkins University offers an innovative graduate degree program that prepares students for success in the field of bioinformatics. Drawing upon the strengths of the Johns Hopkins' Biotechnology and Engineering, the program creates a rigorous bioinformatics curriculum that brings together the computer science, biosciences and bioinformatics disciplines.

#### Quick Stats

Course Locations	Baltimore, MD; Rockville, MD; Online				
Available 100% Online	Yes.				
Entry Terms	Fall, Spring or Summer semester				
Degree Requirements	11 courses				



### Northeastern University

College of Science

### **Bioinformatics – Master of Science**



2

Search





# UC San Diego

# Bioinformatics Algorithms (Part 1)

This course is the first in a two-part series covering some of the algorithms underlying bioinformatics. It will cover some of the algorithms underlying the following fundamental topics in bioinformatics: assembling genomes, comparing DNA and protein sequences, finding regulatory motifs, analyzing genome rearrangements, identifying proteins, and many other topics.



#### About the Course

The sequencing of the human genome fueled a computational revolution in biology. As a result, modern biology produces as many new algorithms as any other fundamental realm of science. Accordingly, the newly formed links between computer science and biology affect the way we teach applied algorithms to computer scientists.

Genome sequencing is just one of hundreds of biological problems that have become inextricable from the computational methods required to solve them. In this course, we





### Meetings & Courses Program

**Cold Spring Harbor Laboratory** 

Home

Meetings

Courses

#### PROGRAMMING FOR BIOLOGY October 13 - 28, 2014 Application Deadline: July 15, 2014

How To Apply

Selection Process and Stipends

Travel

General Information

Campus Information Instructor: Simon Prochnik, DOE Joint Genome Institute

#### See the <u>Roll of Honor</u> - who's taken the course in the past

Web-based tools are no longer enough for today's biologist who needs to access and analyze large datasets from myriad sources in desparate formats. The need to design and program custom analysis pipelines is becoming ever more important as new technologies increase the already exponential rate at which biological data is generated. Designed for lab biologists with little or no programming experience, students will leave the two-week **Programming for Biology** course with the bioinformatics and scripting skills necessary to exploit this abundance of biological data. The prerequisite for the course is basic knowledge of UNIX; some scripting experience is helpful. Lectures and problem sets from previous years that cover this background material are available online and students can study this material before starting the course.

The course teaches Perl, a scripting language that is easy to learn and efficient to use. Perl also has a vast array of ready-







🎹 Apps 🖪 🎔 뾪 🖻 🔞 🚫 🛡	al 😑 🐘 📐 🛐 🥩 🗱 Ritchie Lab Blogs 🕻 Ritchie Lab Wiki 🗱 Ritchie Lab Forum 🐺 PASS Explorer 🗀 PSU-Websites 🦳 Interesting-Websit
Stephen D. Turner, Ph.D.	Bioinformatics Workshops & Training
la CV	Jump to: In-person courses, online material, or other resources.
🛁 E-mail	
🛩 Twitter	In-person Courses
	Last updated: 2014-03-18
ັກ Blog	Note that many of the courses shown below also have course materials from past years publicly available online.
<b>O</b> Github	Cold Spring Harbor Courses: meetings.cshl.edu/courses.html
Slides	Cold Spring Harbor has been offering advanced workshops and short courses in the life sciences for years. Relevant workshops includ Genetics of Complex Human Diseases, Statistical Methods for Functional Genomics, Programming for Biology, Computational and Co Unlike most of the others below, you won't find material from past years' CSHL courses available online.
e Edu	Canadian Bioinformatics Workshops: bioinformatics.ca/workshops
	Bioinformatics.ca through its Canadian Bioinformatics Workshops (CBW) series began offering one and two week short courses in bioi proteomics in 1999. The more recent workshops focus on training researchers using advanced high-throughput technologies on the lat computational biology to deal with the new data. Course material from past workshops is freely available online, including both audio/vi include microarray analysis, RNA-seq analysis, genome rearrangements, copy number alteration, pathway analysis, genome visualizat functional annotation, data analysis using R, statistics for metabolomics, and much more.
	UC Davis Bioinformatics Training Program: training.bioinformatics.ucdavis.edu
	The UC Davis Bioinformatics Training program offers several intensive short bootcamp workshops on RNA-seq, data analysis and visu a focus on Amazon's computing resources. They also offer a week-long Bioinformatics Short Course, covering in-depth the practical th next-generation sequencing techniques. Every course's documentation is freely available online, even if you didn't take the course.
	MSU NGS Summer Course: http://bioinformatics.msu.edu/ngs-summer-course-2014

# Other web resources...

- <u>http://biocyc.org/</u>
- <u>http://metacyc.org/</u>

# Acknowledgements



Yuki Bradford, analyst Scott Dudek, software developer Alex Frase, software developer Molly Hall, PhD student Dokyoon Kim, Post-doc Ruowang Li, PhD student Anastasia Lucas, undergrad researcher Anna Okula, PhD student Donna Panasiti, administrative asst. Sarah Pendergrass, Research Scientist Suzy Unger, program coordinator Anurag Verma, bioinformatics programmer Shefali Verma, bioinformatics analyst John Wallace, software developer