

Predictive Analytics in Healthcare

HI 743

Department of Health Informatics and Administration
Zilber College of Public Health
University of Wisconsin - Milwaukee

January 23, 2025

Overview

1. HI 743

2. Introductions

- 2.1 About Me
- 2.2 About You

3. Syllabus

- 3.1 Syllabus Review
- 3.2 Weekly Itinerary

HI 743 - Predictive Analytics in Healthcare

- **Instructor:** Ryan Gallagher
- **Time & Place:** Thursdays 2:30PM - 5:00PM in EH 103
- **Textbooks:** *Fundamentals of Machine Learning for Predictive Data Analytics: Algorithms, Worked Examples, and Case Studies* by John D. Kelleher, Brian Mac Namee and Aoife D' Arey

Introduction to Statistical Learning with Applications in R (Second Edition) by Gareth James, Daniela Witten, Trevor Hastie, Robert Tibshirani
(Freely Distributed at statlearning.com)
- **Required Materials:** Textbook, A Computer (Laptop preferably), R (software), & RStudio (IDE for R)

About Me



Ryan Gallagher

Biostatistician

MCW/CW Advanced Genomics Lab

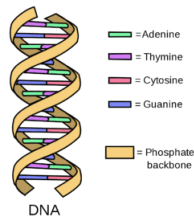
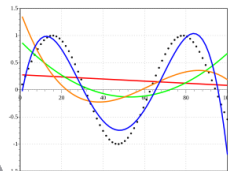
M.A. in Biostatistics & Data Science (MCW)

B.S. in Statistics & Applied Physics, double major (UWEC)



What I Do

- Biostatistics
 - Hypothesis Testing
 - Model Fitting
 - Differential Expression Analysis
- Bioinformatics
 - Pipeline Development
 - Genomic Data Analysis
- Programming
 - R for Statistics
 - Python for Data Processing Pipeline
 - Bash for Linux Cluster



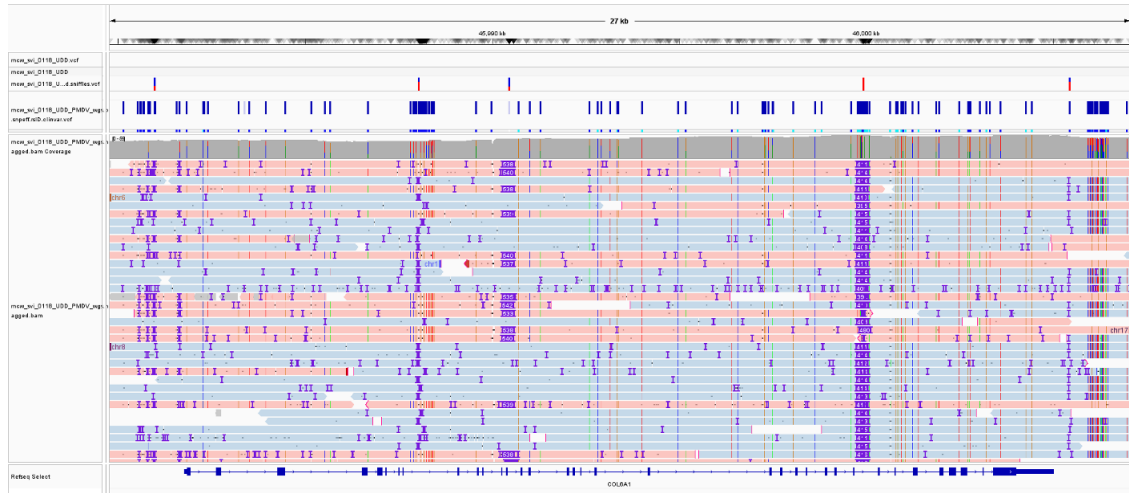
What I Do

I work on my lab's **Structural Variance Initiative**, which:

1. Recruits & Consents Pediatrics Patients
2. Sequences Entire Genome from Biological Sample
 - 2.1 Oxford Nanopore Long-Read Sequencing
 - 2.2 Bionano Optical Genome Mapping
3. Processes Raw Sequencing Reads (Python + Bash)
 - 3.1 Sequencing Data Management ($\sim 1.5\text{TB}$)
 - 3.2 Raw Sequence Alignment + QC
 - 3.3 Identify SNPs + Small INDELs, CNVs, SVs
4. Annotate Identified Variants
5. Identify Potential Disease Causing Variants



What I Do



Licensed

Ryan Gallagher ▾

Help ▾

🏠 > Subject: mcw_svi_0109 > Analysis: TCW241220125033 (Long Reads Singleton Analysis) > Analyze

Open

 Not Assigned

Info: [Subject](#) [Clinical](#) [VCF Samples](#)

Fast Track (Lines: 82 Locations: 79 Genes: 52)

Recessive HOM/HEMI

Recessive Compound HET *

Dominant HET

Recessive HET

X Linked

Mitochondria

Secondary Findings

CNV/SVs

Filters and Tools			GENOMIC AND GENETIC DATA										ACMG		VARIANT CALLING Q&R					CLINICAL EVIDENCE					IN HOUSE			MA...	
RELEVANCE	PATHOGENIC	NOTES	Tag	LOCATION	GENE	REF	ALT	AA	HGVSp	ZYG	REFSEQ	DOM	REC	Q&R	DP2	AL...	PHENO	M...	CLINVAR	OMIM	OMIM L...	LITV...	V	G	AF ...				
1	VUS			14:64795...	SPTB	C	T	E525K	c.1573G>A	HET	NM_001355436.2	LB	VUS	Med	17, 20	54.05	3.59	2/11	VUS	Anemia, ...	AD, AR								
1	VUS			17:40354...	RARA	G	T	A311S	c.931G>T / ...	HET	NM_000964.4	VUS	VUS	Med	18, 15	45.45	20.47	4/11		Leukemi...	UN								
1	VUS			6:306704...	DHX16	C	T	R211H	c.632G>A / ...	HET	NM_003587.5	VUS	VUS	Med	15, 14	48.28	3.30	1/11		Neurom...	AD								
2	VUS			6:1332417...	EYA4	G	A		c.-66G>A	HET	NM_004100.5	VUS	VUS	Med	22, ...	38.89	--			?Cardio...	AD								
3				16:89563...	RPL13	G	C	Q20...	c.615G>C / ...	HET	NM_000977.4	LB	VUS	Med	17, 10	37.04	--			Spondyl...	AD								
3				4:155996...	FBXL5	A	T		c.*6024T>A	HET	NM_012161.4	VUS	VUS	Med	19, ...	51.28	--		PAT, LP										
3	VUS			5:179634...	HNRNP1	G	C		c.-579C>G	HET	NM_001364225.2	VUS	VUS	Med	7, 20	74.07	10.34	2/11		Neurode...	AD				31.25				
1				1:2157282...	USH2A	C	T	W39...	c.11864G>...	HET	NM_206933.4	VUS	PAT	Med	21, 19	47.50	1.55	2/11	PAT	Retinitis ...	AR								
1				10:49500...	ERCC6	G	C	Q52...	c.1570C>G ...	HET	NM_000124.4	VUS	VUS	High	18, 9	33.33	16.01	4/11		?De San...	AR, AD, ...								
1	VUS			10:95693...	TCTN3	TCA	T	CD92	c.276_277d...	HET	NM_015631.6	PAT	PAT	High	23, 21	47.73	11.04	2/11	PAT, LP	Joubert ...	AR								
1				11:124639...	SIAE	T	G	L323F	c.969A>C / ...	HET	NM_170601.5	VUS	VUS	Med	17, 11	39.29	--		VUS	{Autoim...	UN								
1				12:14622...	GUCY2C	A	G	I859T	c.2576T>C ...	HET	NM_004963.4	VUS	VUS	Med	11, 18	62.07	7.24	1/11		Diarrhea...	AD, AR								
1				13:113118...	F7	C	T	R35...	c.1057C>T / ...	HET	NM_019616.4	BEN	VUS	Med	10, 10	50.00	--		LP	{Mycopar...	AR								
1				13:113118...	F10	C	T		c.-4126C>T	HET	NM_000504.4	BEN	VUS	Med	10, 10	50.00	--		LP	Factor X ...	AR								
1				16:702518...	AARS1	G	A		c.*839C>T	HET	NM_001605.3	VUS	VUS	Med	13, 11	45.83	7.50	2/11		?Leukoe...	AD, AR								
1				17:217029...	KCNJ18	G	A	R43H	c.128G>A / ...	HET	NM_001194958.2	VUS	VUS	Med	31, 14	31.11	1.99	1/11		{Thyroto...	AD					37.50			
1				17:217032...	KCNJ18	C	A	E320K	c.435C>A / ...	HET	NM_001194958.2	BEN	VUS	Med	37, 13	33.50	1.99	1/11		{Thyroto...	AD					34.35			

[Disclaimer](#) [Terms of Use](#) [Privacy Policy](#) [Cite Us](#)

© Genevix Genomex Ltd. 2025 (v6.1 Build 3952)

About You

[Name]

[Program / Area of Focus]

[Aspirations Post Education?]

[Familiarity with Statistics? (Scale 1-10)]

[Familiarity with Programming? (Scale 1-10) (Languages?)]

[Do you have a Laptop? (Which OS?)]

Syllabus

Lets look at our syllabus

Week 2 - Course Motivations

Lecture:

- What is Predictive Data Analytics?
- What is Machine Learning?
- How does Machine Learning work?
- How is Machine Learning relevant to my field of interest?

Week 3 - Intro to R Language

Lecture:

- What is R?
- Why use R?
- Brief History of R

R Lab:

- Hands-on Tutorial in R
- Data Manipulation with *Tidyverse* Package
- R Figures

Week 4 - Data Exploration

Lecture:

- Data Quality
- Data Preparation
- Missing Values / Outliers

R Lab:

- Hands-On Data Cleaning
- Data Visualization

Week 5 - Error Based Regression

Lecture:

- Simple Linear Regression
- Multiple Linear Regression & Interpretation
- Nonlinear Regression & Interpretation

R Lab:

- Multiple & Nonlinear Regression
- Interpreting R Output
- Introduction to RMarkdown for Reporting

Week 6 - Classification / Logistic Regression

Lecture:

- Logistic Regression
- Generalized Linear Models
- Multinomial Logistic Regression

R Lab:

- Logistic Regression in R
- `glm()` function

Week 7 - Case Study (or Classification Continued)

This week will either be a continuation of the Classification methods.

Or

We will look at a case study which applies Classification methods to real-world data.

Project Idea is Due Following Week.

Week 8 - Information Based Learning

Lecture:

- Basics of Decision Trees
- Random Forest, Bagging, Boosting, etc..
- Tree "Pruning"

R Lab:

- Fitting Classification Trees
- Fitting Regression Trees
- Bagging Algorithms

Week 9 - SPRING BREAK!

No Class!

Week 10 - Unsupervised / Similarity-based Learning

Lecture:

- Principal Component Analysis
- Clustering Methods
- Challenges of Unsupervised Learning

R Lab:

- K-Means Clustering
- PCA Example

Week 11 - Probability / Bayesian Methods in ML

Lecture:

- Bayes' Theorem
- Naive Bayes Model
- Probability Density Functions

R Lab:

- Naive Bayes Worked Examples

Week 12 - Neural Networks and Deep Learning

Lecture:

- What is Deep Learning?
- Single Layer Neural Networks
- Multilayer Neural Networks
- Convolutional Neural Networks

R Lab:

- Single, Multi, and Convolutional Neural Networks Applied

Week 13 - Model Evaluation

Lecture:

- Model Performance Measures
- Resampling Methods
- Cross Validation

R Lab:

- Validation Set Approach
- Cross-Validation
- Bootstrapping

Week 14 - Practical ML Application

Lecture:

- When to use Machine Learning vs. Typical Methods
- Choosing a Machine Learning Approach
- Matching Approaches to Data

Lab portion will be dedicated to working on Final Project.

Week 15 - Work on Final Project

**Classtime will be dedicated to working on Final Project.
Project Due May 9th.**