



Programming for Biomedical Informatics

Lecture 1 - Welcome & Getting Started

<https://github.com/tisimpson/pbi>

Ian Simpson
ian.simpson@ed.ac.uk



THE UNIVERSITY of EDINBURGH
informatics

Course Organisation

- **Lectures**

- Weeks 1-4, 6-11
- Tuesdays 11:10 - 12:00
- Thursdays 13:10 - 14:00
- Recorded and released same day
- All slides and videos will be put on GitHub

- **References & Reading**

- Relevant books, websites, papers, & any other sources will be in lecture slides and coding notebooks.
- These will be collated on dedicated pages on GitHub for easier access

- **Materials**

- Everything will be available via the course GitHub
- <https://github.com/tisimpson/pbi>

- **Assessment**

- 10x weekly assignments on GitHub Classroom (20% of overall course mark, 2% each - pass/fail - used to gauge understanding & engagement with the course.
- Intended to practice something small from the previous week and/or introduce a resource for you to look at
- As the course progresses these will be predominantly coding problems/examples - supposed to be fun (!)
- End of semester exam (80% of overall course mark)
- New course - will be introducing example exam questions and discussing model solutions through the course
- Exam prep session & Q&A in week 11

- **Communication**

- EdStem - all course discussion and questions (including private questions) please use it!

Course Topics

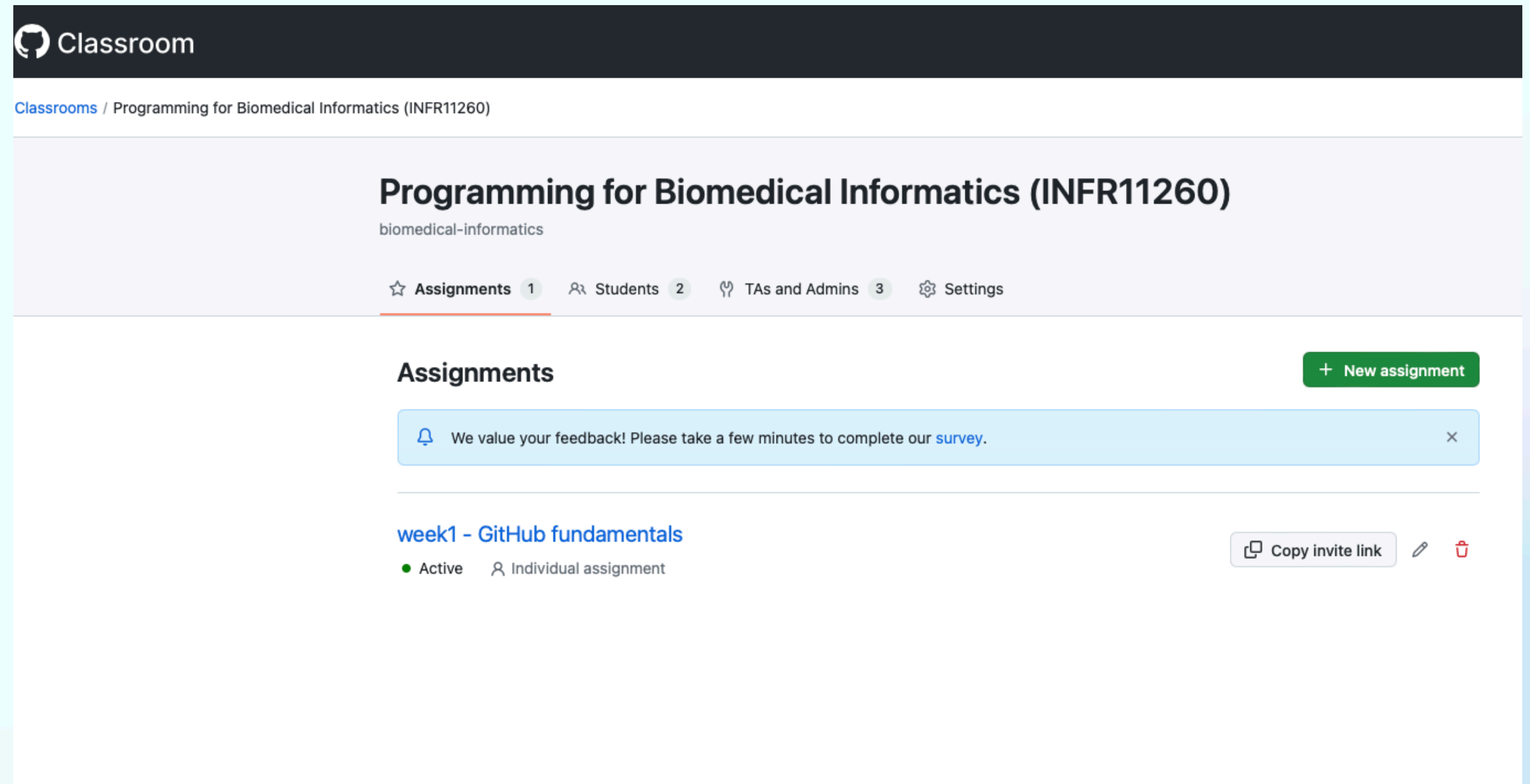
Week	Week Commencing	Weekly Topics (background, application)
1	16th September	(L1) Course Introduction & Setup, (L2) Working with Notebooks & Git
2	23rd September	(L3) Introduction to the Biomedical Dataverse, (L4) Finding & Fetching Data
3	30th September	(L5) Mapping and Harmonisation, (L6) Data Integration & Summary Analysis
4	7th October	(L7) Biomedical Evidence, (L8) Mining & Analysing Biomedical Literature
5	BREAK	
6	21st October	(L9) Measuring Gene Expression, (L10) Differential Gene Expression (GXD)
7	28th October	(L11) Biological Networks, (L12) Network Construction Techniques
8	4th November	(L13) Essential Network Methods, (L14) Network Analysis in Practice
9	11th November	(L15) Structuring Biomedical Data with Ontologies (L16) Functional Analysis
10	18th November	(L17) Working with Multiple Data Modalities, (L18) Modelling at the Patient Level
11	25th November	Course Review, Exam Prep, and Q&A session

foundation

application

Course Setup

GitHub Classroom






<https://github.com/biomedical-informatics>

Course Setup

Notable

<https://noteable.edina.ac.uk/>

[About Noteable](#)[Resources](#)[Pricing](#)[Status](#)[Contact Us ▼](#)


THE UNIVERSITY
of EDINBURGH

Create and share engaging coding lessons with Noteable, a cloud-based computational notebook service which works in your browser from any device.

Developed by [EDiNA](#) at the University of Edinburgh, Noteable hosts your computational notebooks in one simple online hub and can integrate with your [VLE](#).


Streamline your preparation, teaching and marking and build a better learning experience for your students with Noteable.

[Get a Free Trial today](#)




What our clients say


Noteable supports teaching in a variety of subject areas and institutions.



For larger courses, or those containing students with less computing experience, this is undoubtedly a huge benefit of using the Noteable service.



Interspersing live code blocks with narrative content makes for a wonderfully efficient and interactive classroom and online experience.



Very easy to use and allowed me to deliver working examples live in lectures and for students to work with at home. No complicated set-up, works straight from a browser.

Course Setup

VisualStudioCode

<https://code.visualstudio.com/>

bio1_coursework2.ipynb

Users > iansimpson > Library > CloudStorage > OneDrive-UniversityofEdinburgh > Teaching > Bio1 > 2023 > coursework > coursework2 > bio1_coursework2.ipynb > ...

+ Code

+ Markdown

Run All

Clear All Outputs

Outline

...

	HP_0000006...										
2	GENCC_000101-HGNC_16636-OMIM_118210-HP_0000006...	HGNC:16636	KIF1B	MONDO:0007308	Charcot-Marie-Tooth disease type 2A1	OMIM:118210	Charcot-Marie-Tooth disease, type 2A1	GENCC:100004	Limited	HP:0000006	...
3	GENCC_000101-HGNC_17939-OMIM_617532-HP_0000007...	HGNC:17939	SLC45A1	MONDO:0044322	intellectual developmental disorder with neuro...	OMIM:617532	Intellectual developmental disorder with neuro...	GENCC:100004	Limited	HP:0000007	...
4	GENCC_000101-HGNC_11071-OMIM_616291-HP_0000007...	HGNC:11071	SLC9A1	MONDO:0014572	Lichtenstein-Knorr syndrome	OMIM:616291	Lichtenstein-Knorr syndrome	GENCC:100004	Limited	HP:0000007	...

5 rows x 30 columns

#comapre the contents of the column 'classification_title' and 'submitted_as_classification_name'

gencc_data['classification_title'].equals(gencc_data['submitted_as_classification_name'])

print the proportion of rows where the two columns are not equal

print("Proportion of rows where the two columns are not equal: ",sum(gencc_data['classification_title'] != gencc_data['submitted_as_classification_name'])/len(gencc_data

print the rows where they are not equal

gencc_data[gencc_data['classification_title'] != gencc_data['submitted_as_classification_name']]

[2]

... Proportion of rows where the two columns are not equal: 0.18617596195417208

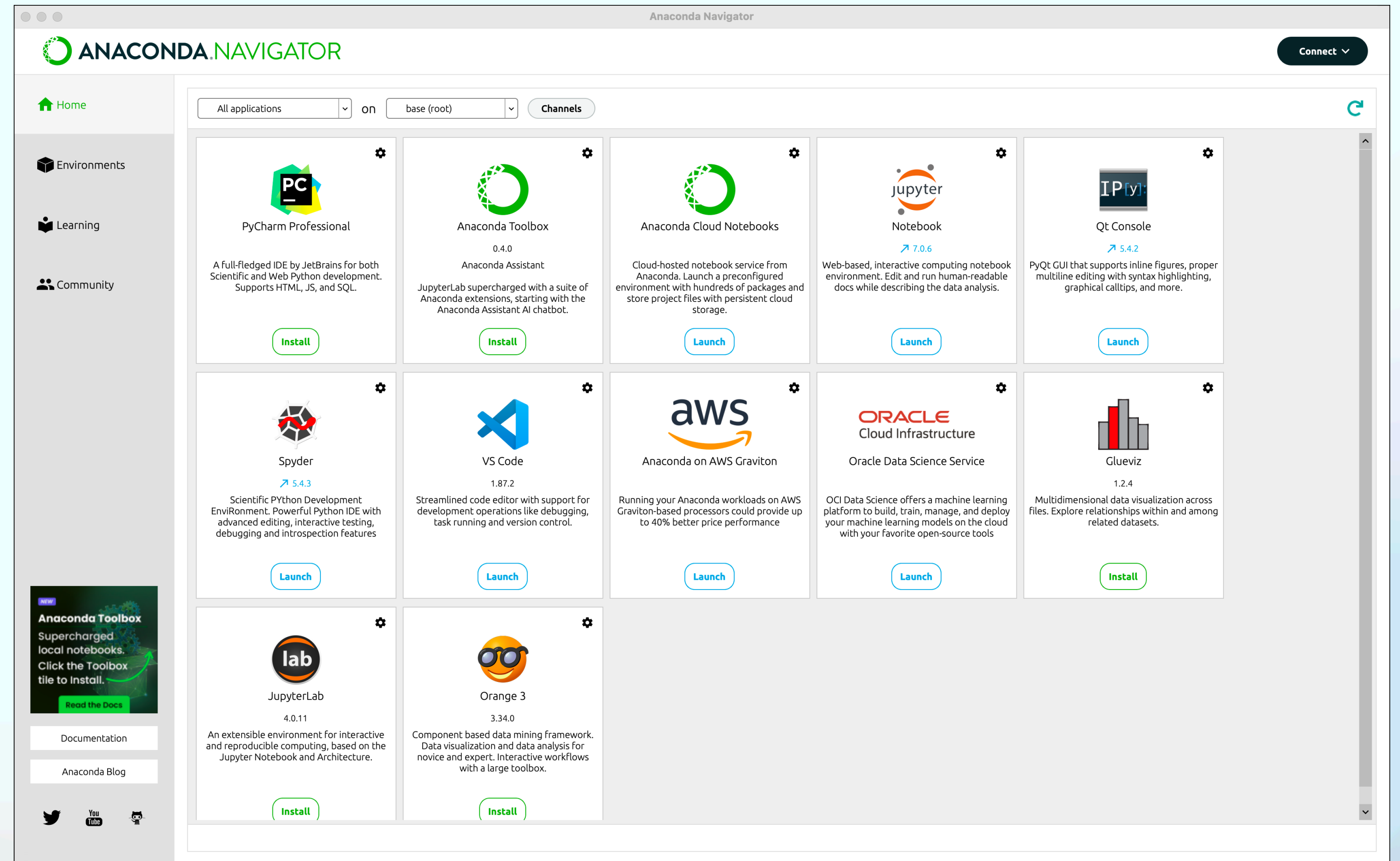
...

	uuid	gene_curie	gene_symbol	disease_curie	disease_title	disease_original_curie	disease_original_title	classification_curie	classification_title	moi_curie	
56	GENCC_000105-HGNC_9801-MONDO_0030913-HP_000000...	HGNC:9801	RAC1	MONDO:0030913	intellectual disability, autosomal dominant 48	MONDO:0030913	intellectual disability, autosomal dominant 48	GENCC:100002	Strong	HP:0000006	
57	GENCC_000105-HGNC_7882-MONDO_0007318-HP_0000000...	HGNC:7882	NOTCH2	MONDO:0007318	Alagille syndrome	MONDO:0007318	Alagille syndrome	GENCC:100003	Moderate	HP:0000006	
87	GENCC_000105-HGNC_15766-MONDO_0014379-HP_000000...	HGNC:15766	ADNP	MONDO:0014379	ADNP-related multiple congenital anomalies - i...	MONDO:0014379	ADNP-related multiple congenital anomalies - i...	GENCC:100001	Definitive	HP:0000006	
92	GENCC_000105-HGNC_1321-MONDO_0032634-HP_0000000...	HGNC:1321	TIMMDC1	MONDO:0032634	mitochondrial complex 1 deficiency, nuclear ty...	MONDO:0032634	mitochondrial complex 1 deficiency, nuclear ty...	GENCC:100003	Moderate	HP:0000007	
196	GENCC_000105-HGNC_2174-MONDO_0005258-HP_0000000...	HGNC:2174	CNTN4	MONDO:0005258	autism spectrum disorder	MONDO:0005258	autism spectrum disorder	GENCC:100005	Disputed Evidence	HP:0000006	
...
18484	GENCC_000112-HGNC_1750-OMIM_211380-HP_0000007-...	HGNC:1750	CDH11	MONDO:0008885	Elsahy-Waters syndrome	OMIM:211380	Elsahy-Waters syndrome	GENCC:100001	Definitive	HP:0000007	
18485	GENCC_000112-HGNC_9416-OMIM_619636-HP_0000007-...	HGNC:9416	PRKG2	MONDO:0030553	acromesomelic dysplasia 4	OMIM:619636	Acromesomelic dysplasia 4	GENCC:100002	Strong	HP:0000007	

Course Setup

Anaconda/miniconda

<https://www.anaconda.com/>





Programming for Biomedical Informatics

Next Lecture this Thursday - “Working with Notebooks & Git”

Please Bring your Laptop!

Ask Questions on the EdStem Discussion Board

<https://github.com/tisimpson/pbi>