

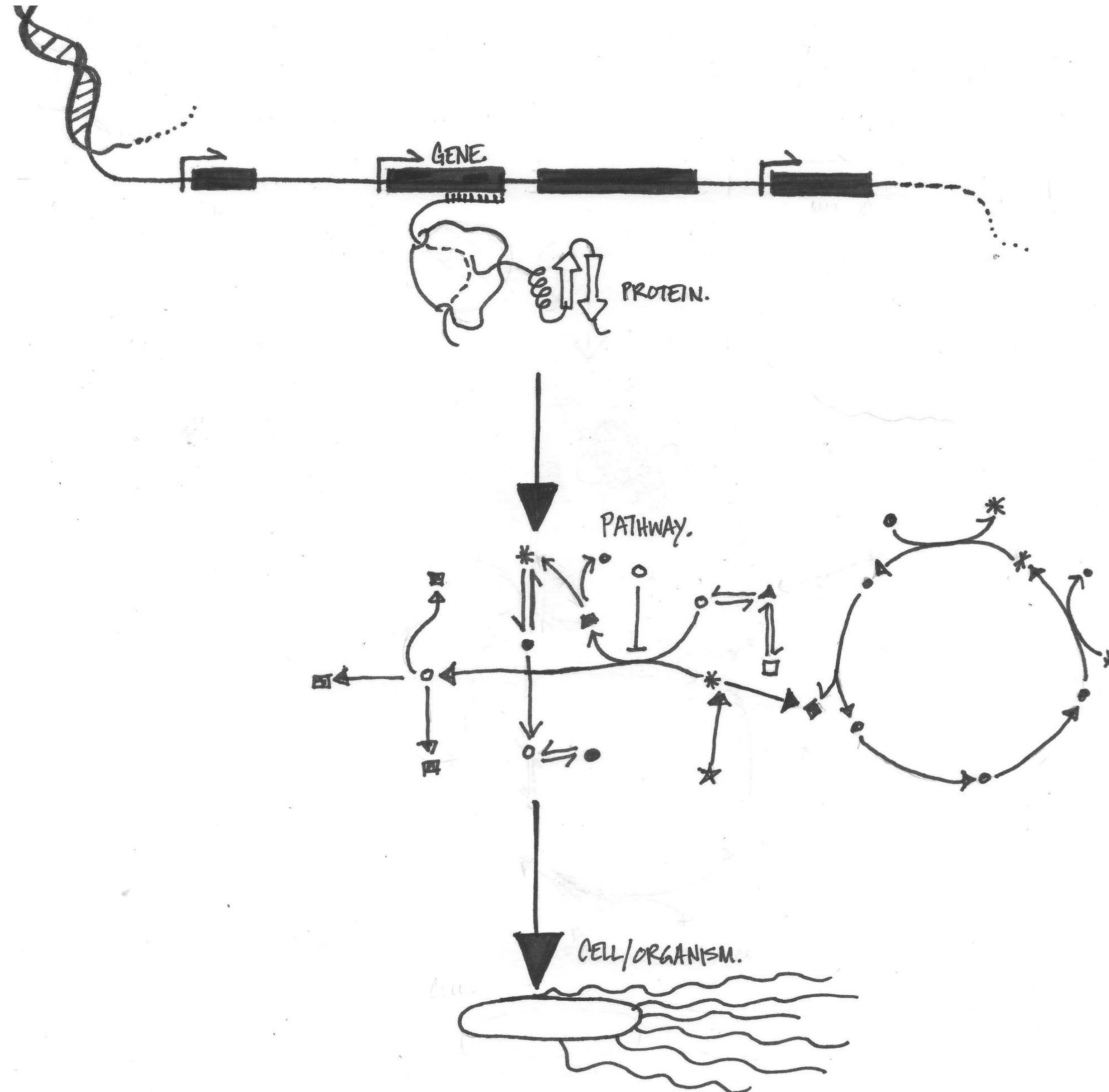
Welcome!

Python Programming for Enabling Your Science



STARS 2024
UT Southwestern Medical Center

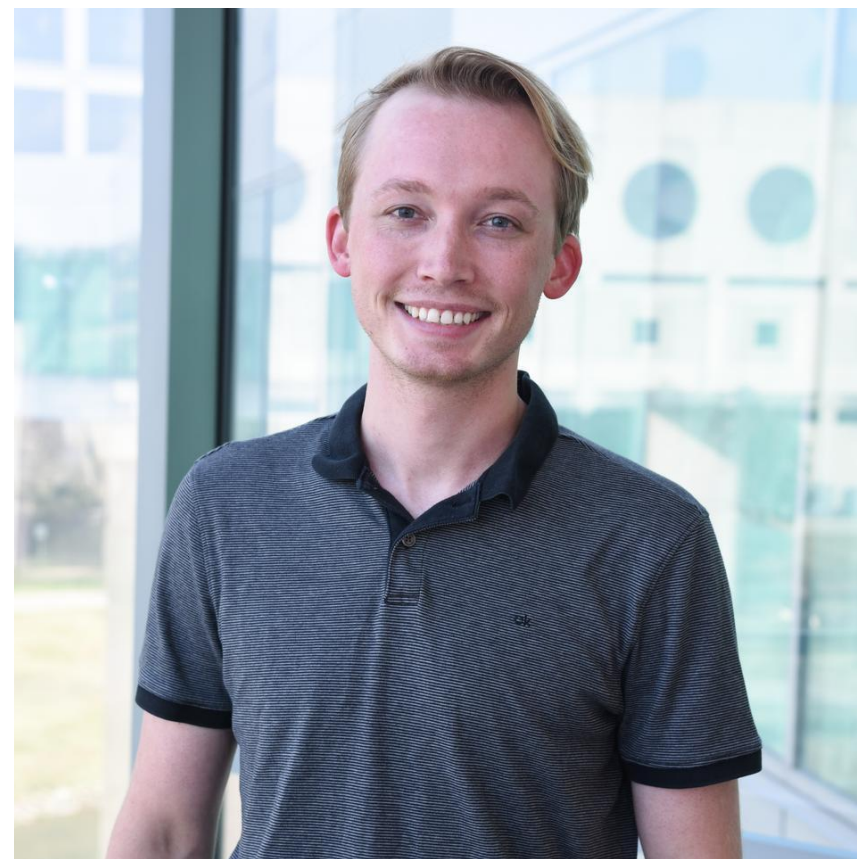
My own lab uses a combination of computation and experiment in our research.
We want to understand:



How individual mutations (variation) at the genetic level...

ultimately shape appearance, behavior, and growth
at the organismal level.

These are the folks that have made this class possible:



Ryan Otto



Dominique Lagasca



Suzette Palmer



Ashley Vu



Tasia Bos



Phil Brown



Jerry Dinan



Eryn Sale



Let's get started with a download: <https://www.anaconda.com/>



If you are using a Mac for this course you will want the 64 bit graphical installer. You will need to pay attention to whether or not you have an Apple silicon or intel processor.

If you wish to double check, you can go to the apple menu, then select:
“About This Mac”
Under **Processor**, it will specify the type

If you are using a PC, and not sure if you need 64 or 32 bit... you probably need 64.

If you wish to double check, you can go to the START button, then select:
Settings -> System -> About
Under **Device specifications**, see **System type**

Let's get started with a download: <https://www.anaconda.com/>



If you are using a Mac for this course you will want the 64 bit graphical installer.

If you are using a PC, and not sure if you need 64 or 32 bit... you probably need 64.

If you wish to double check, you can go to the START button, then select:

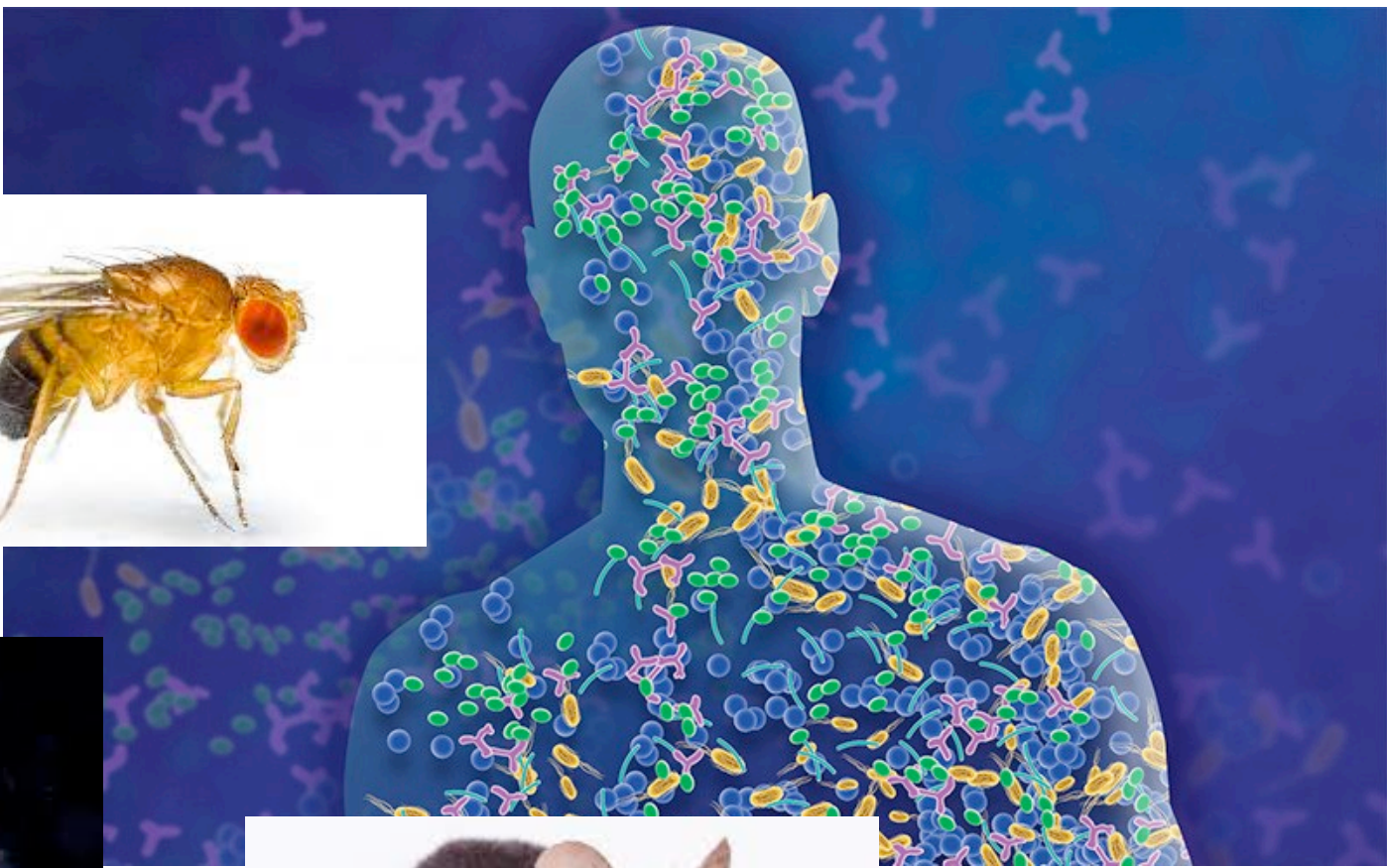
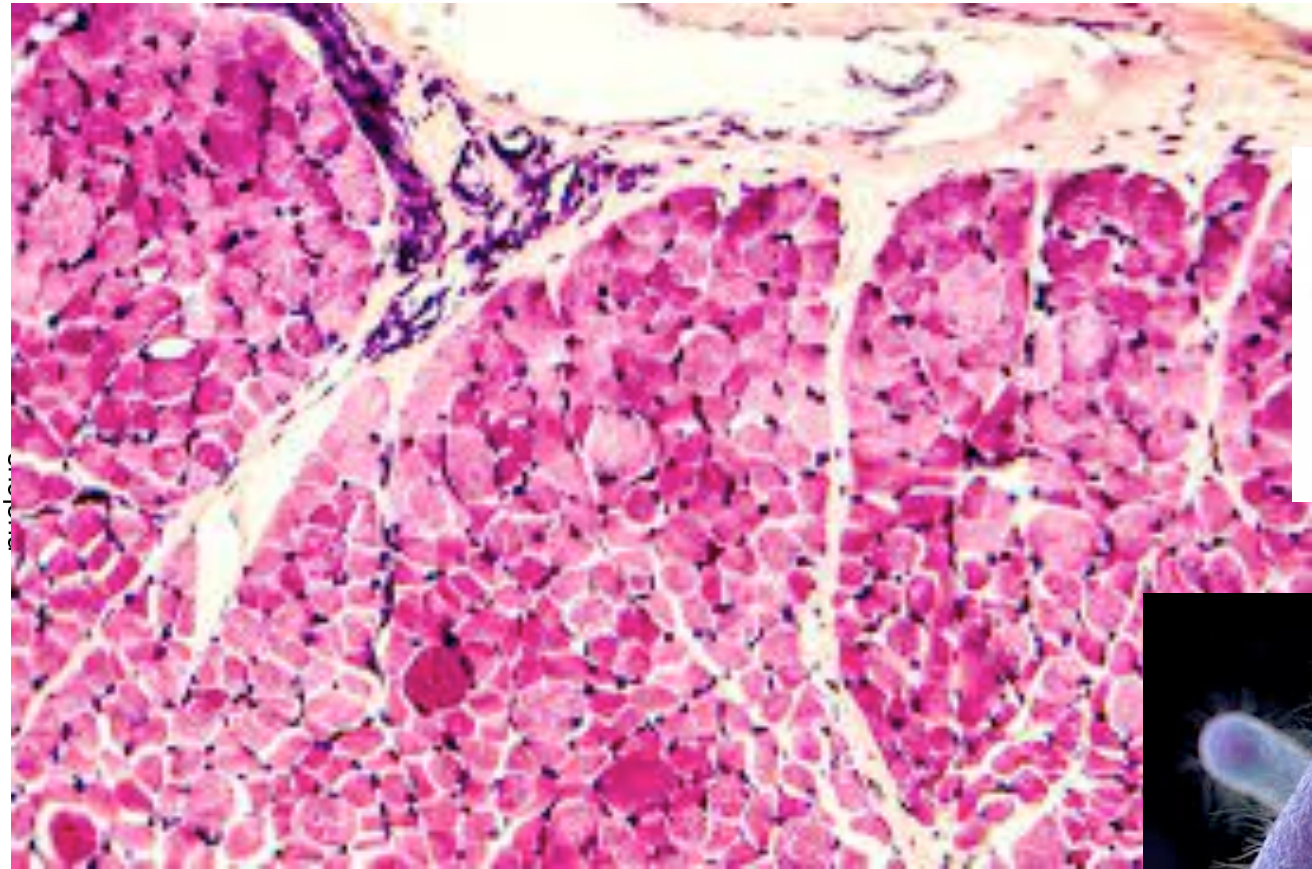
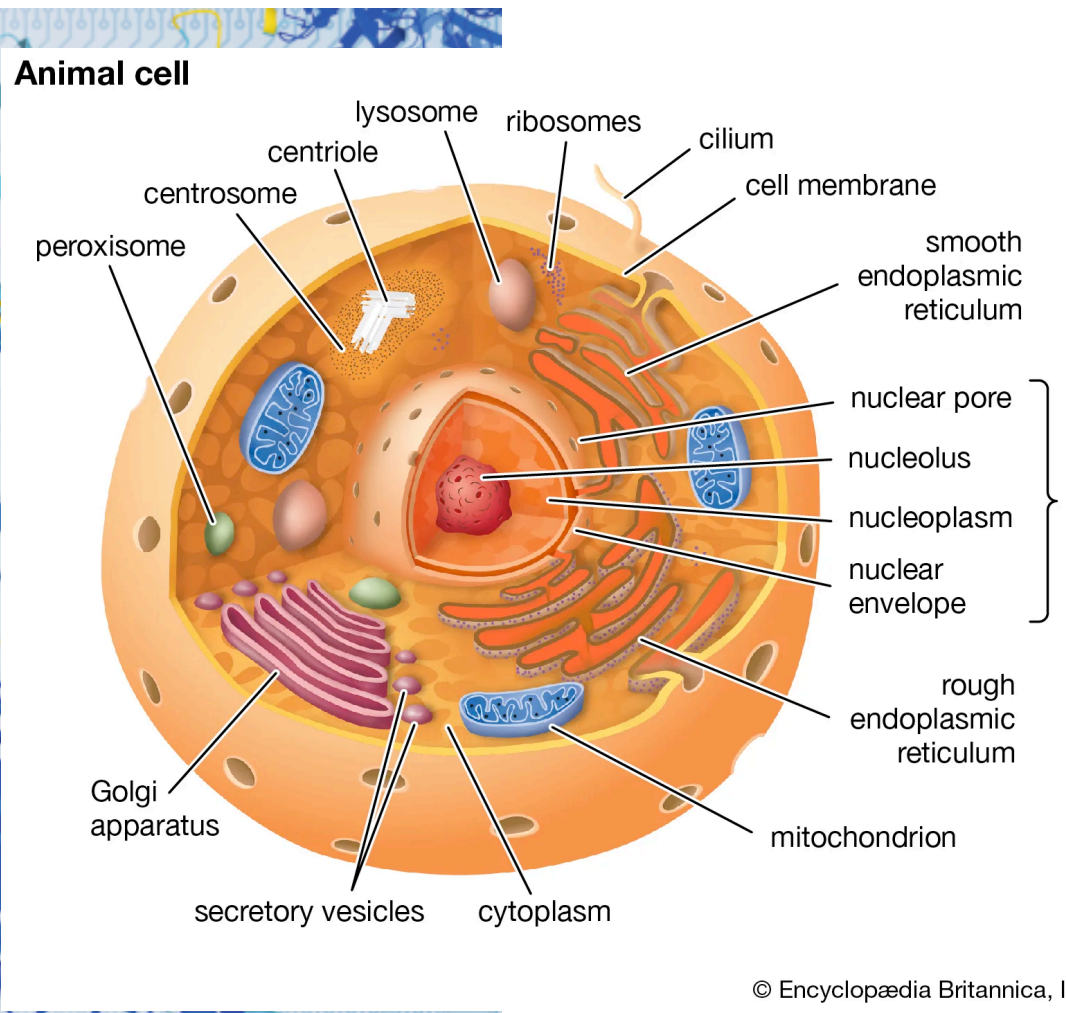
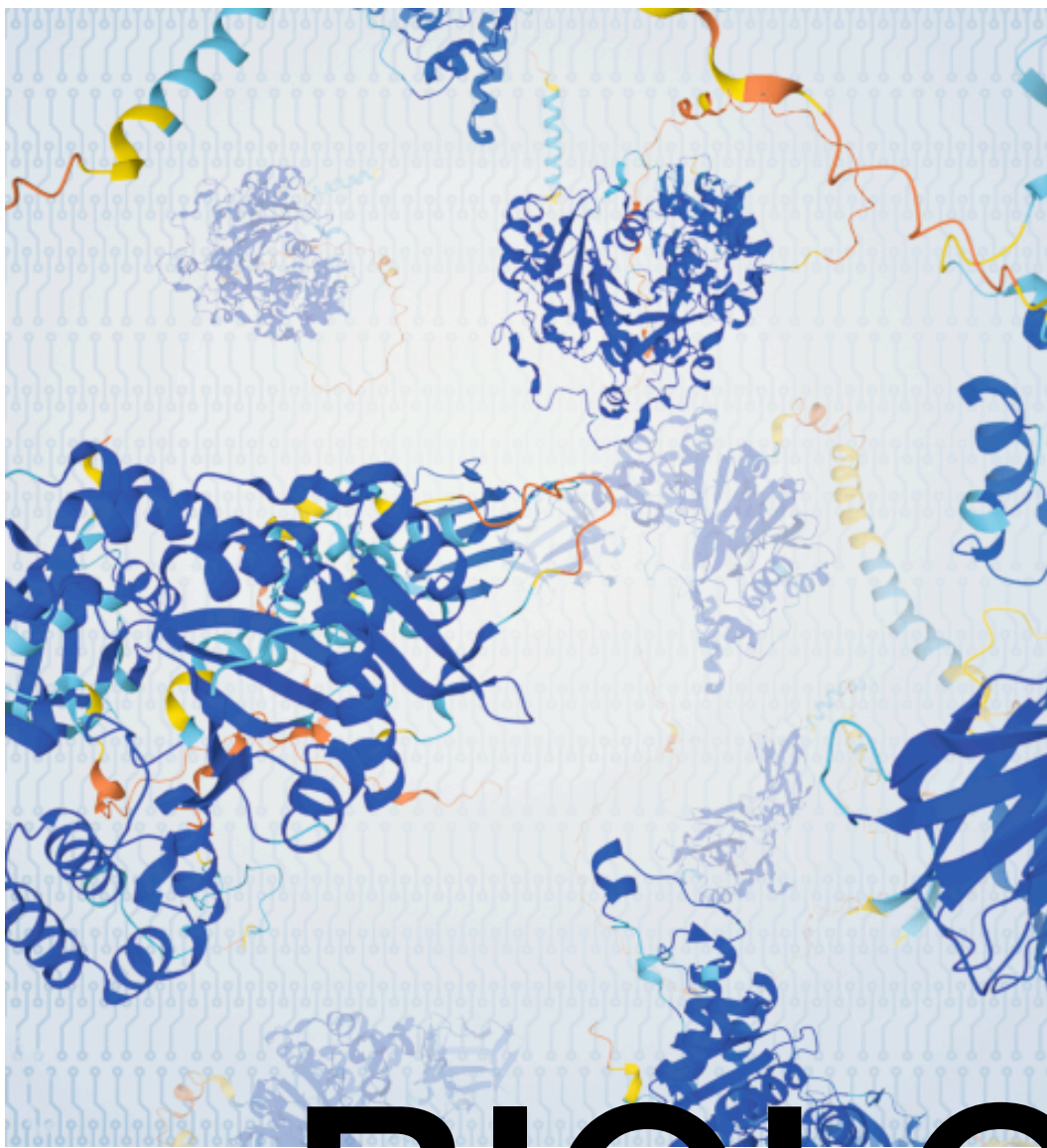
Settings -> System -> About

Under **Device specifications**, see **System type**

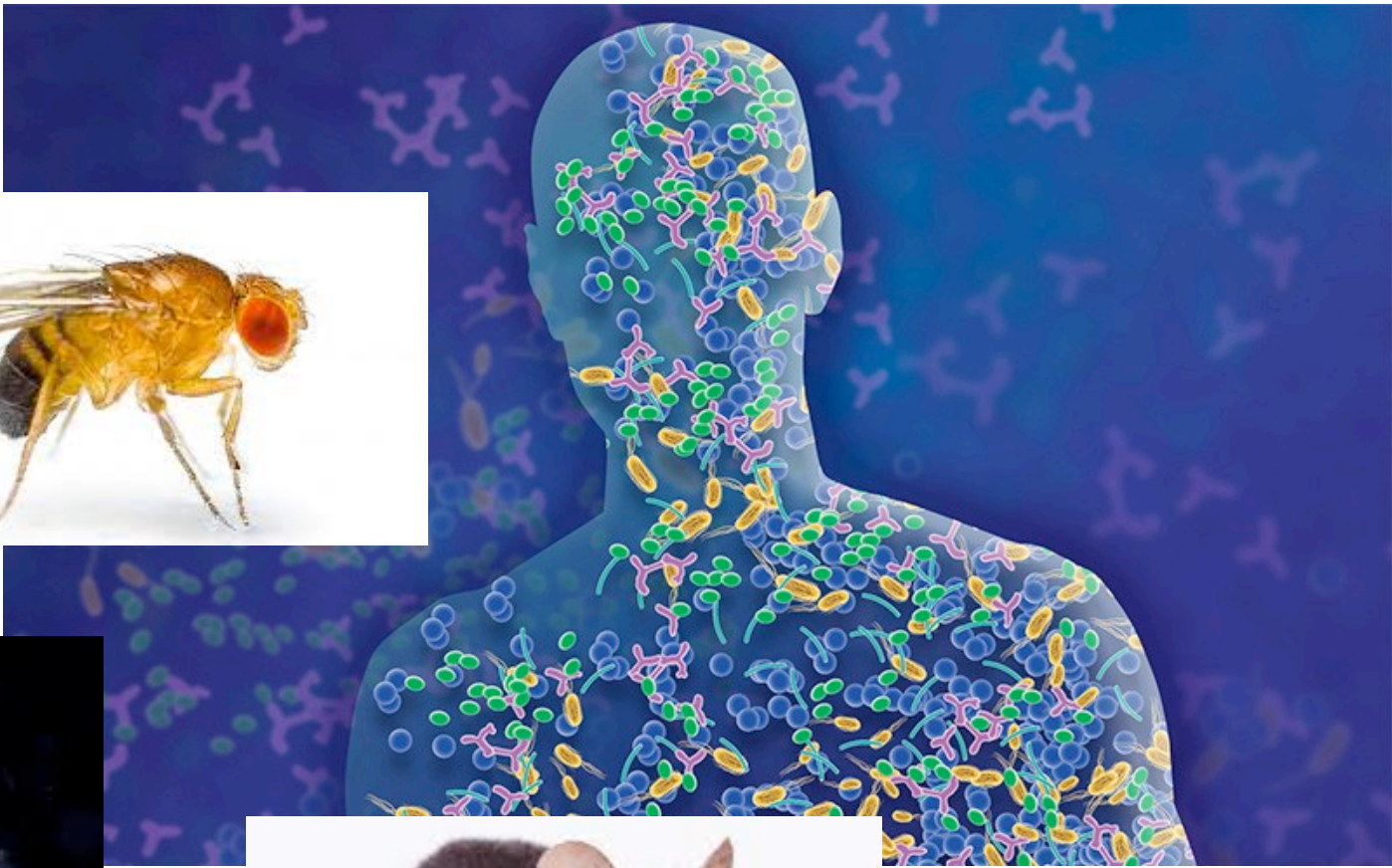
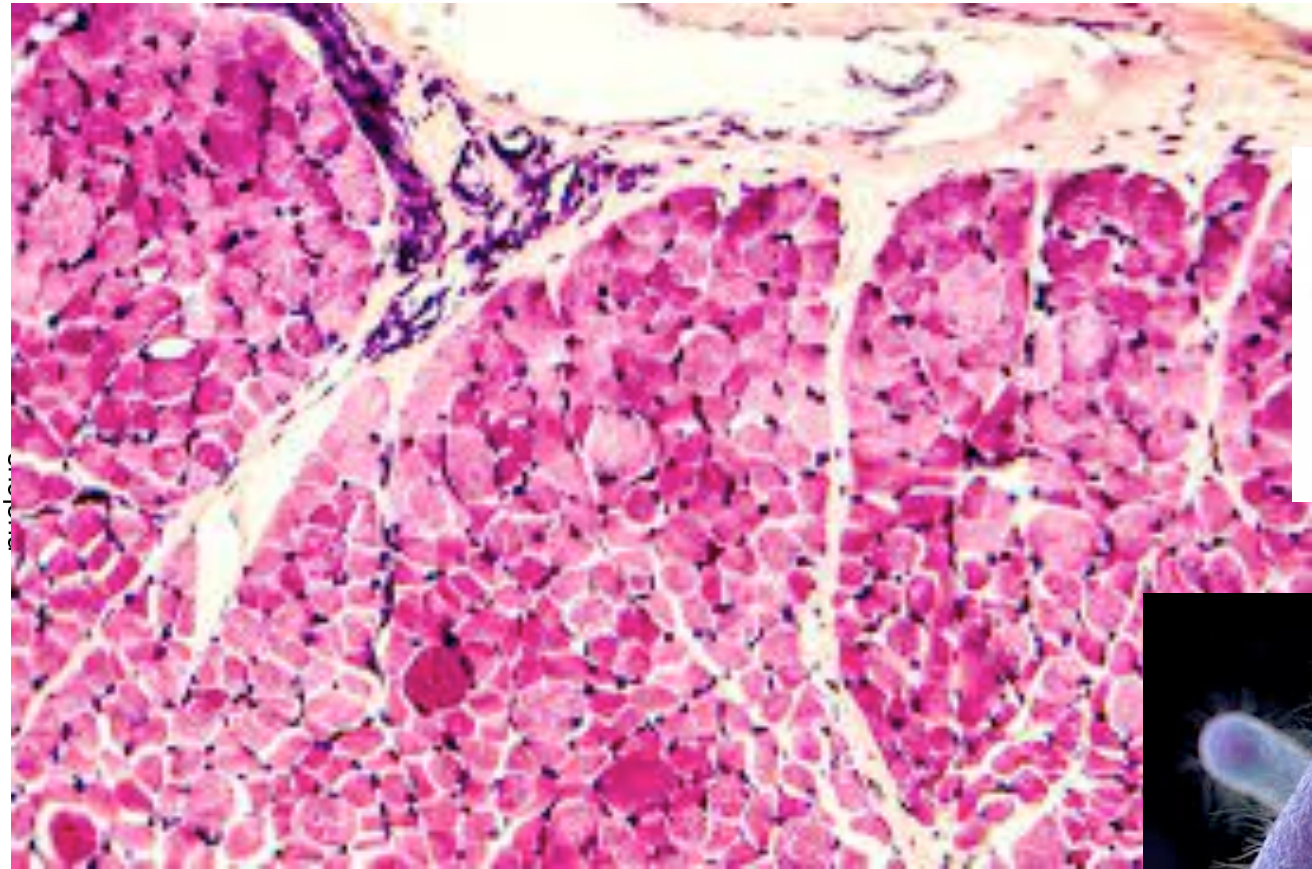
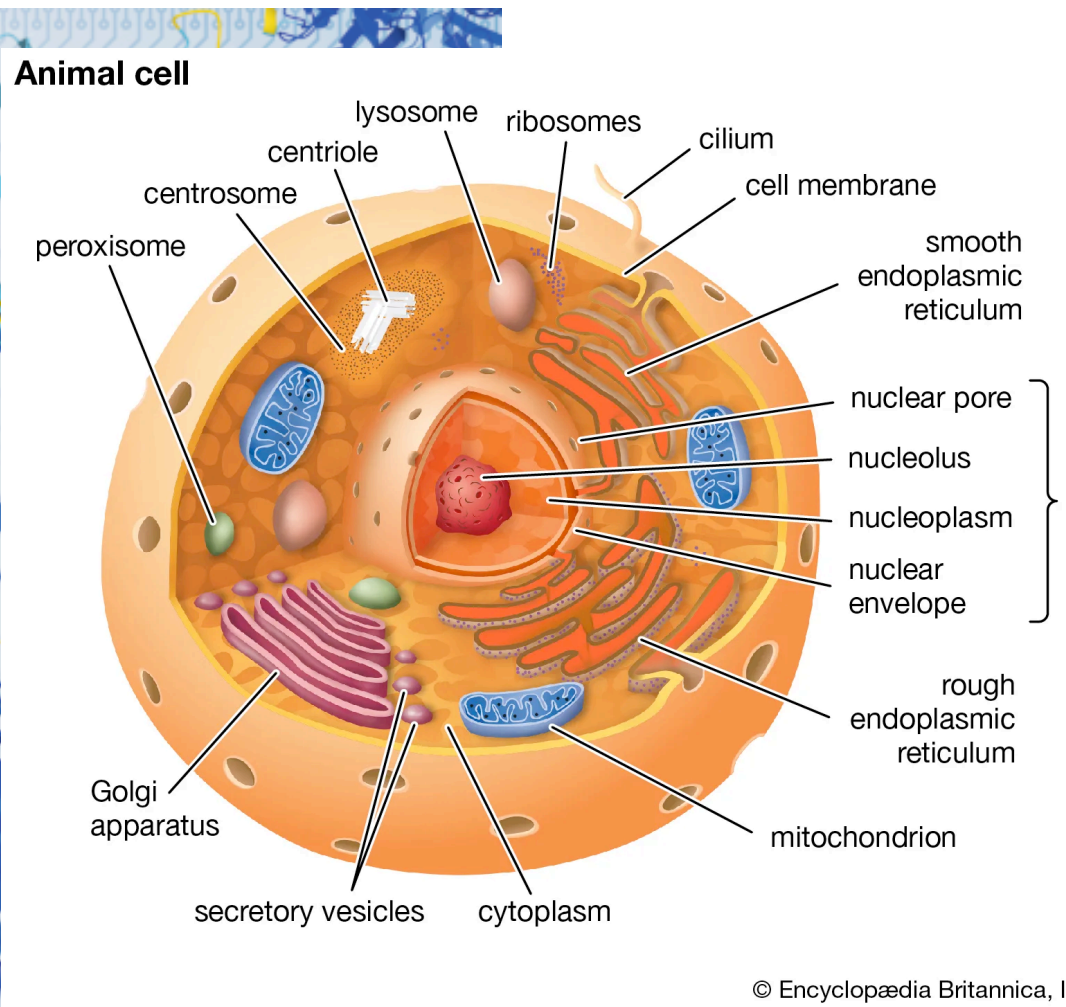
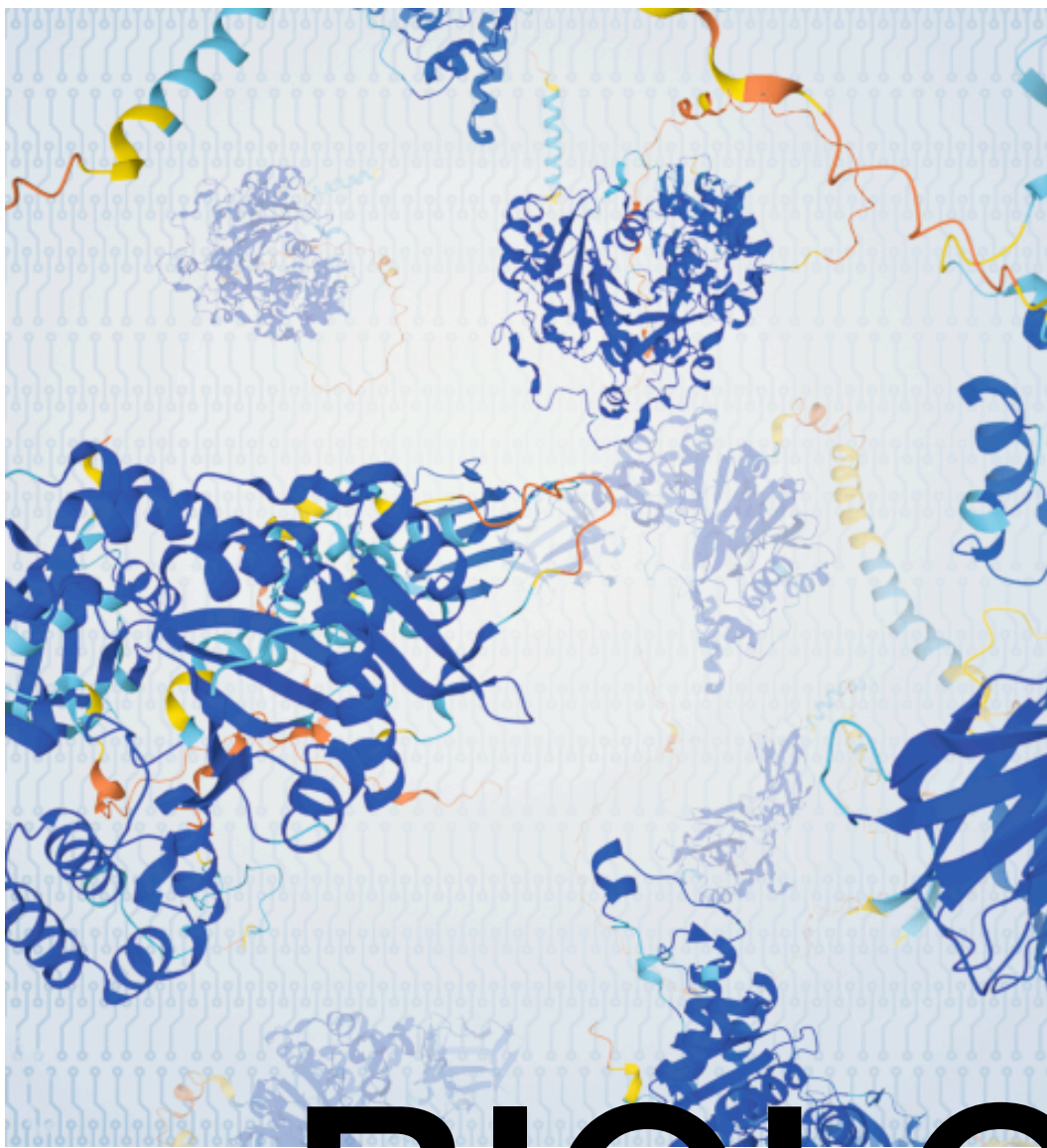
Anaconda is free.

It is one (of several) possible Python distribution platforms.

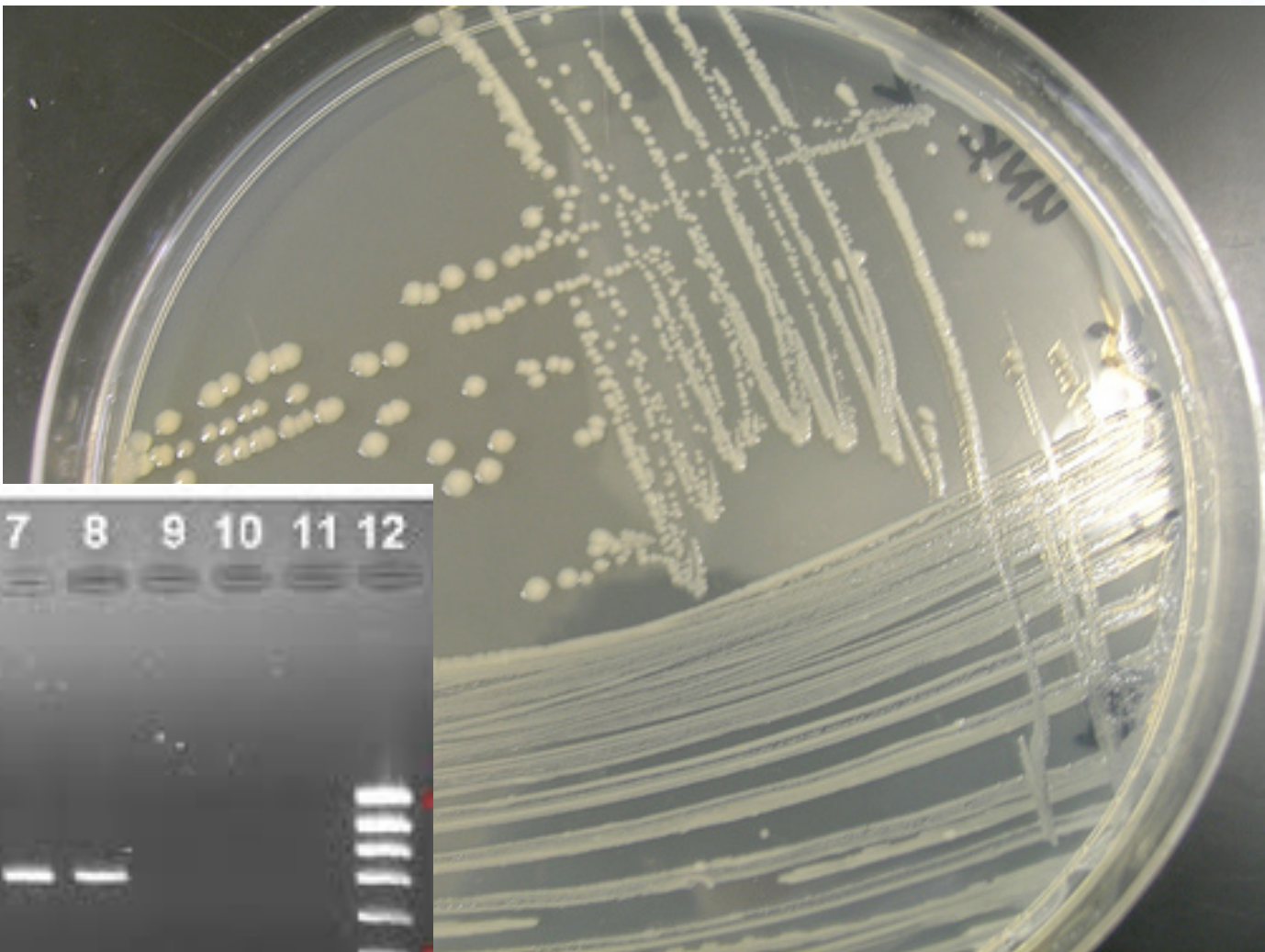
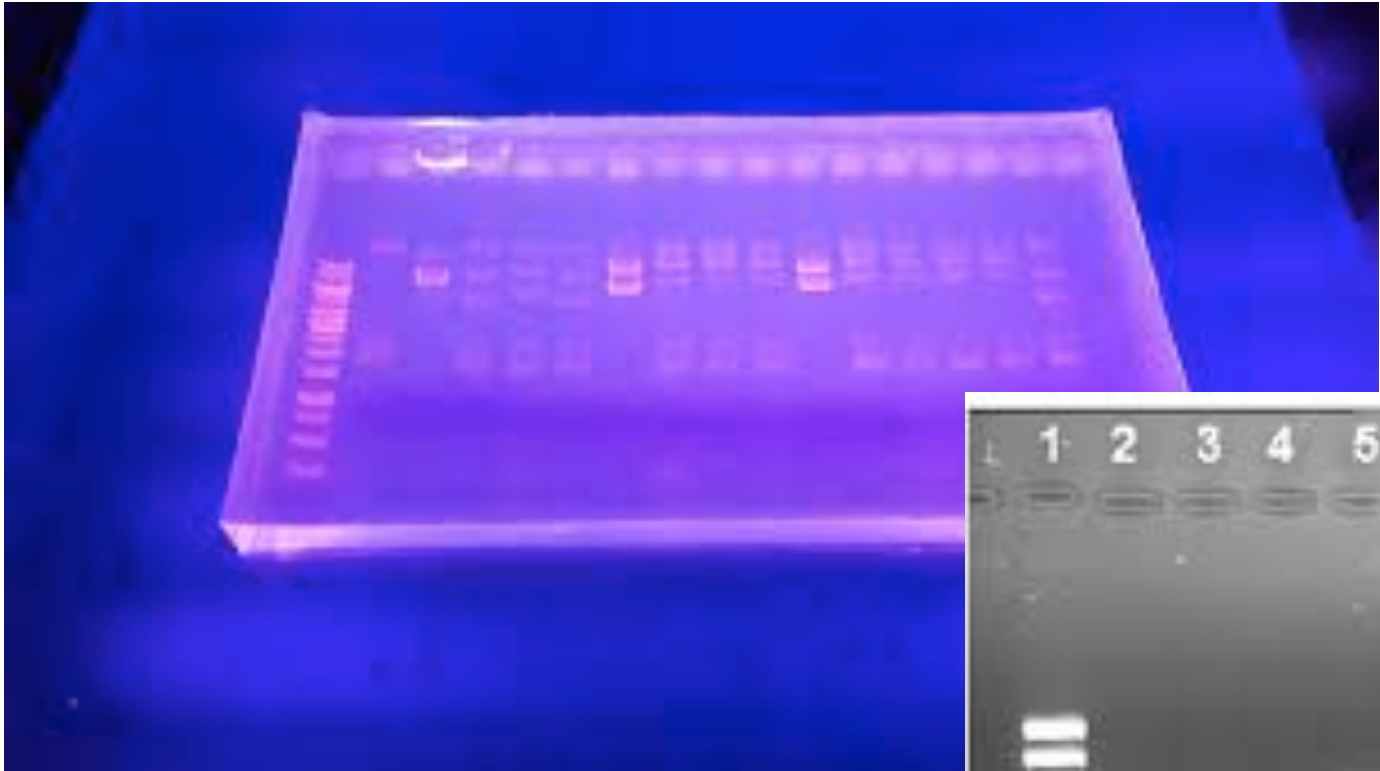
It includes a number of other tools that are useful to scientific computing and which we will make use of in this class.

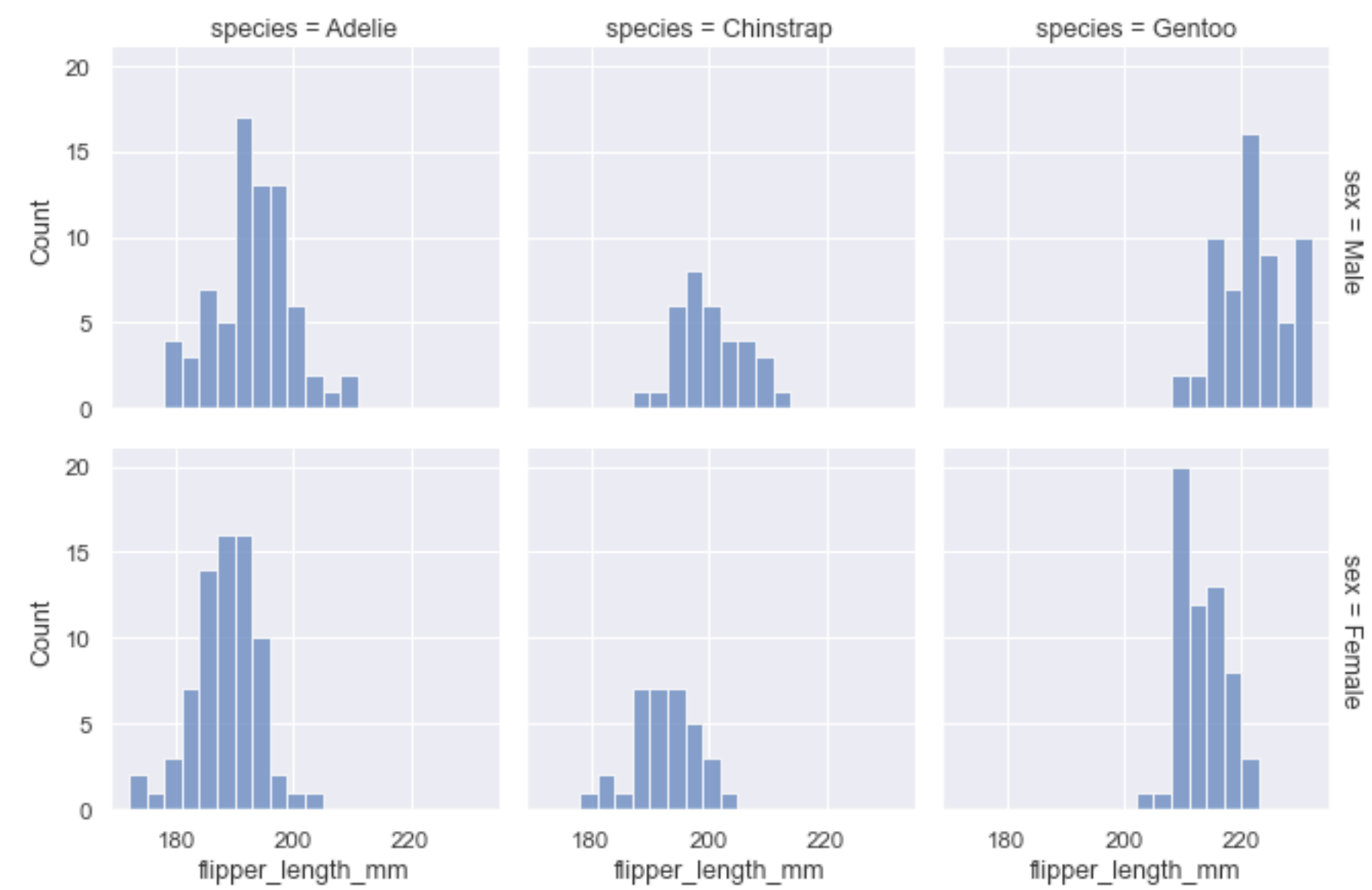


BIOLOGY...

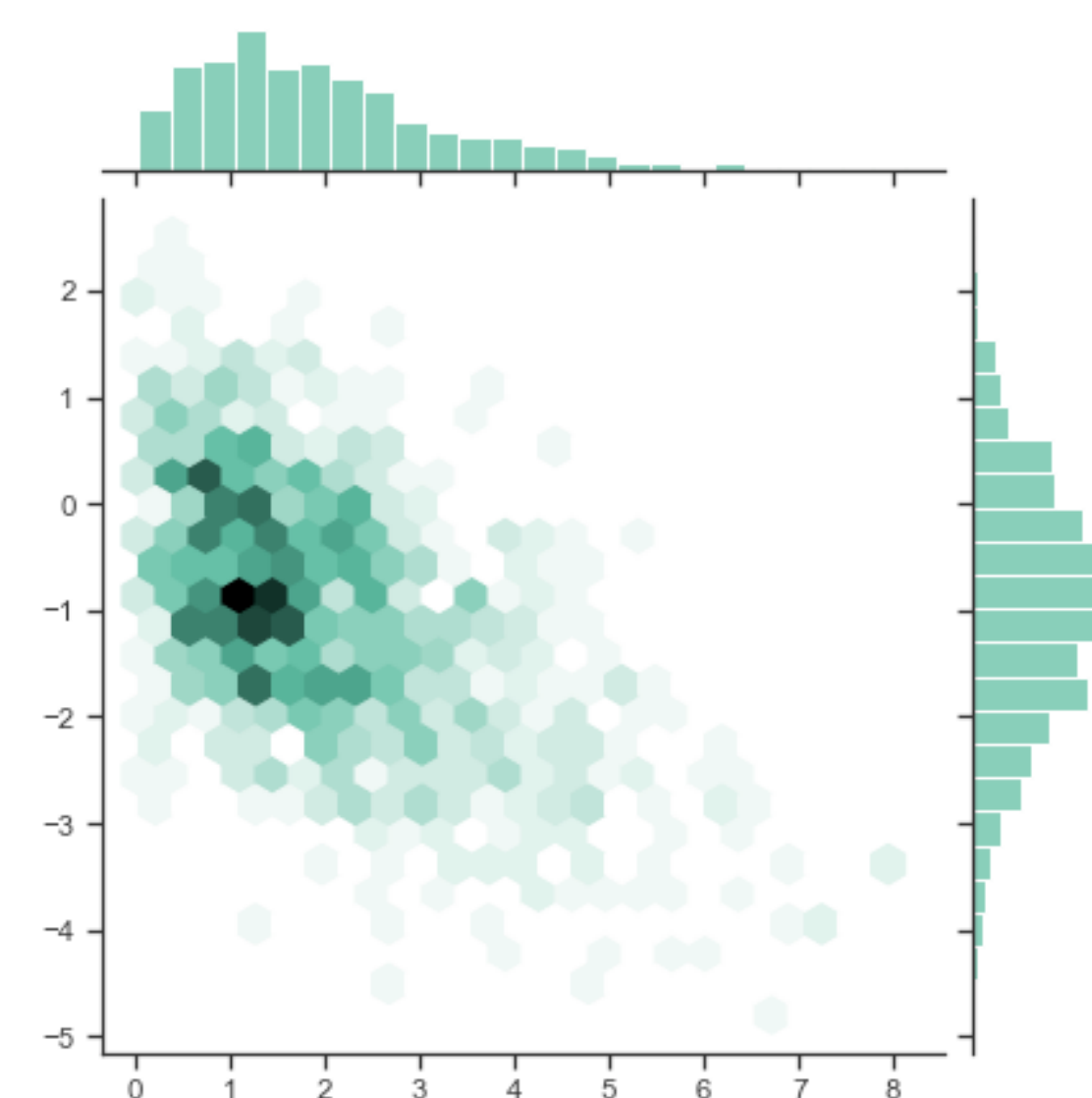
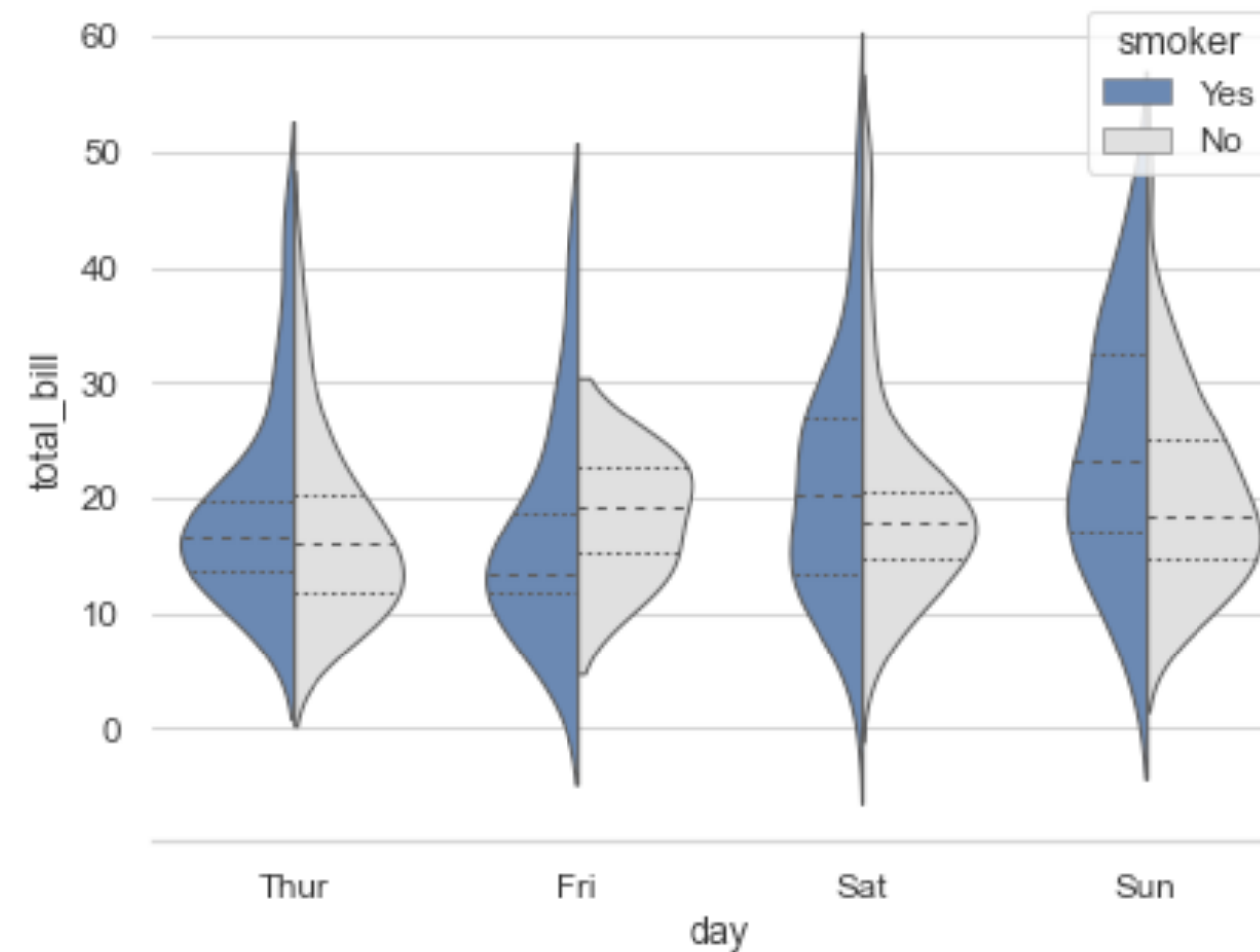
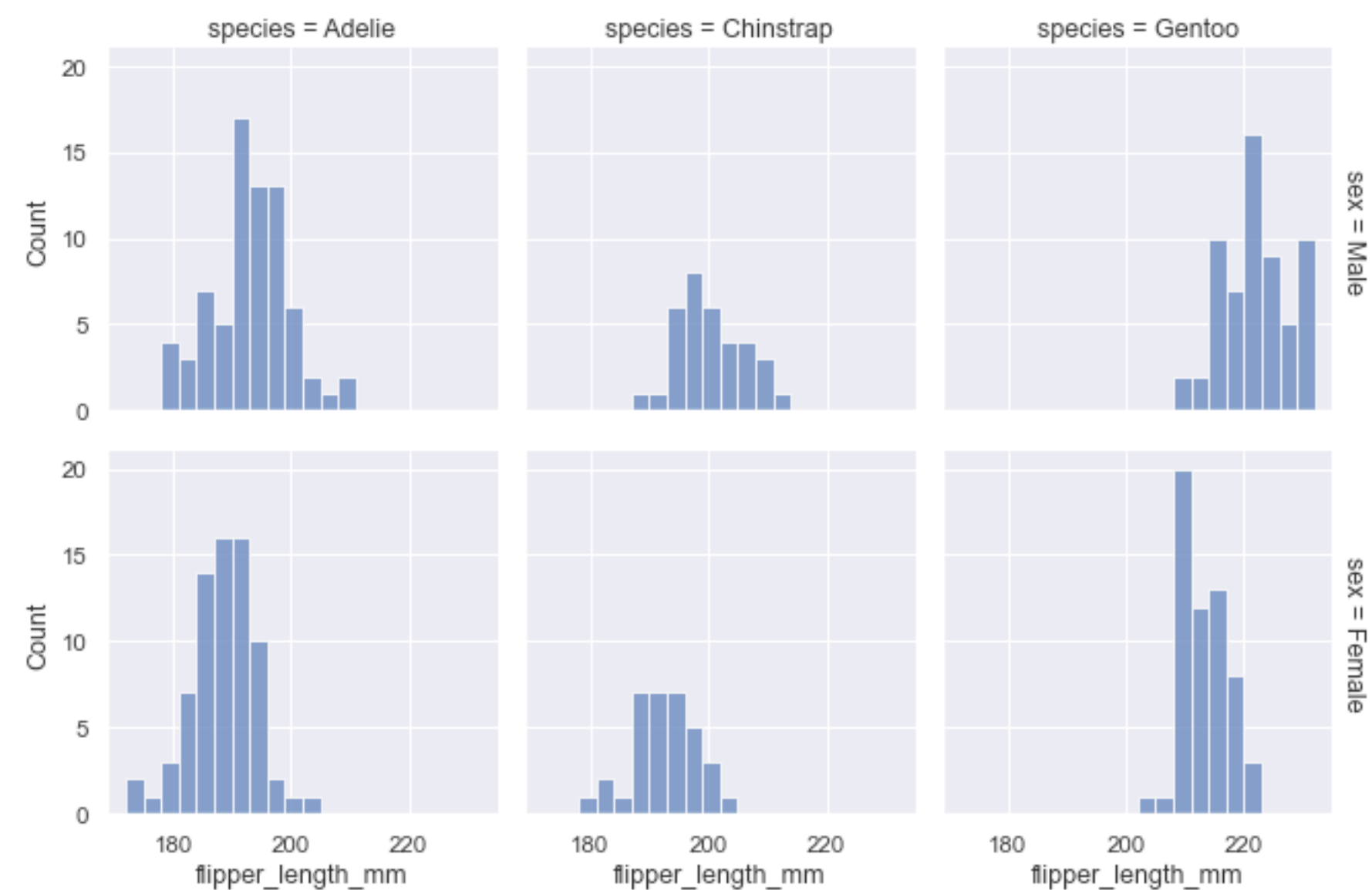


BIOLOGY... WHY CODE?

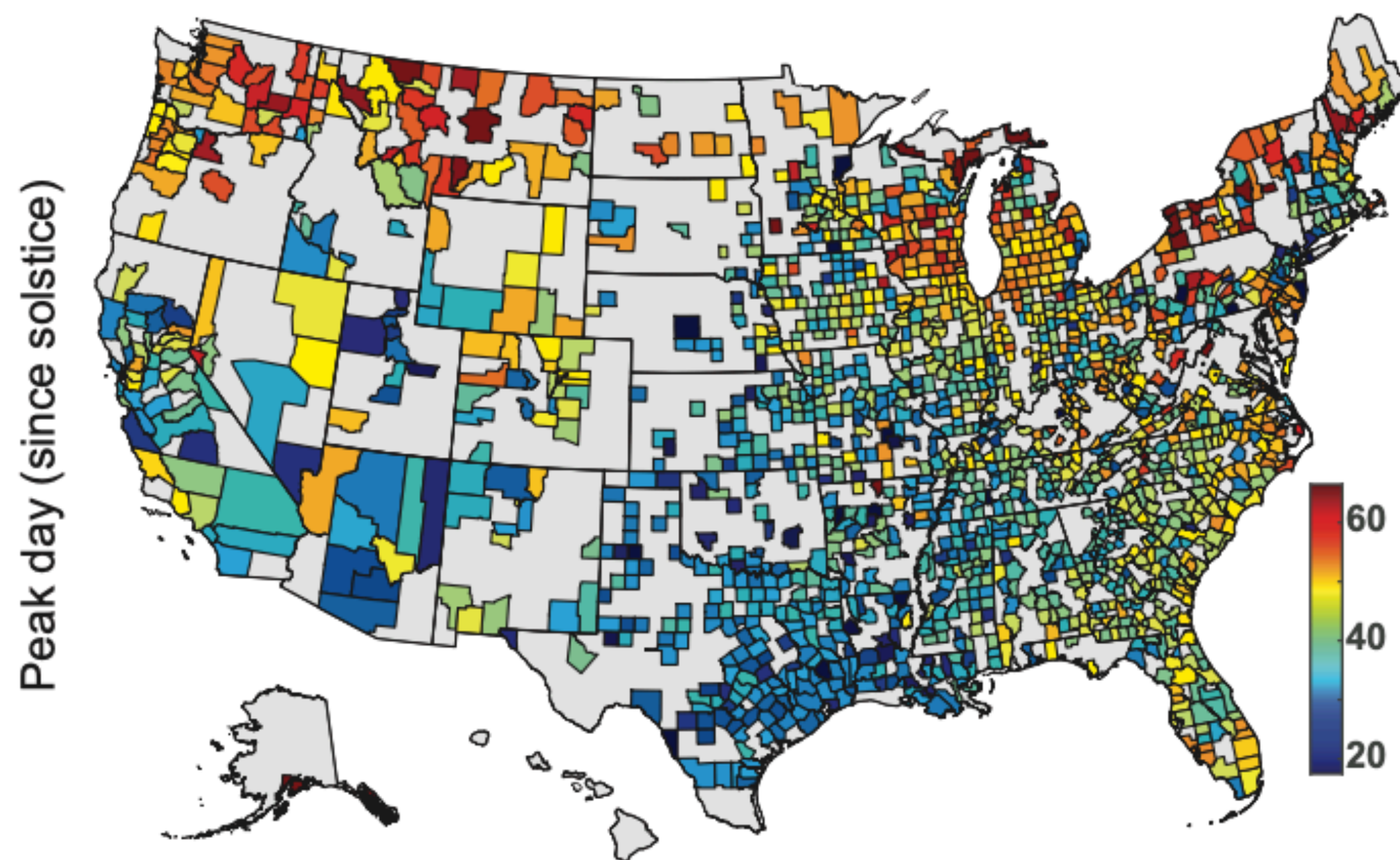


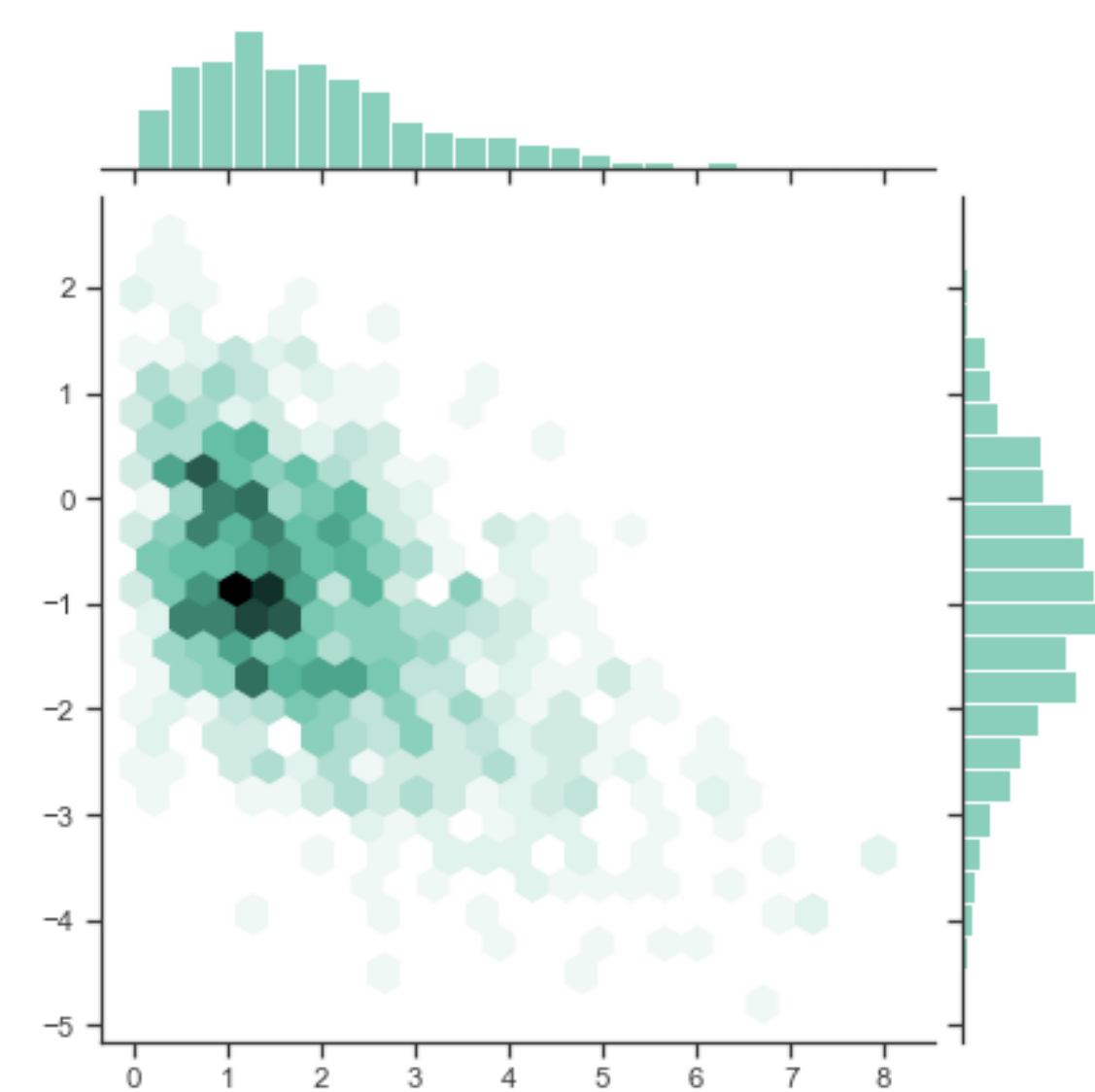
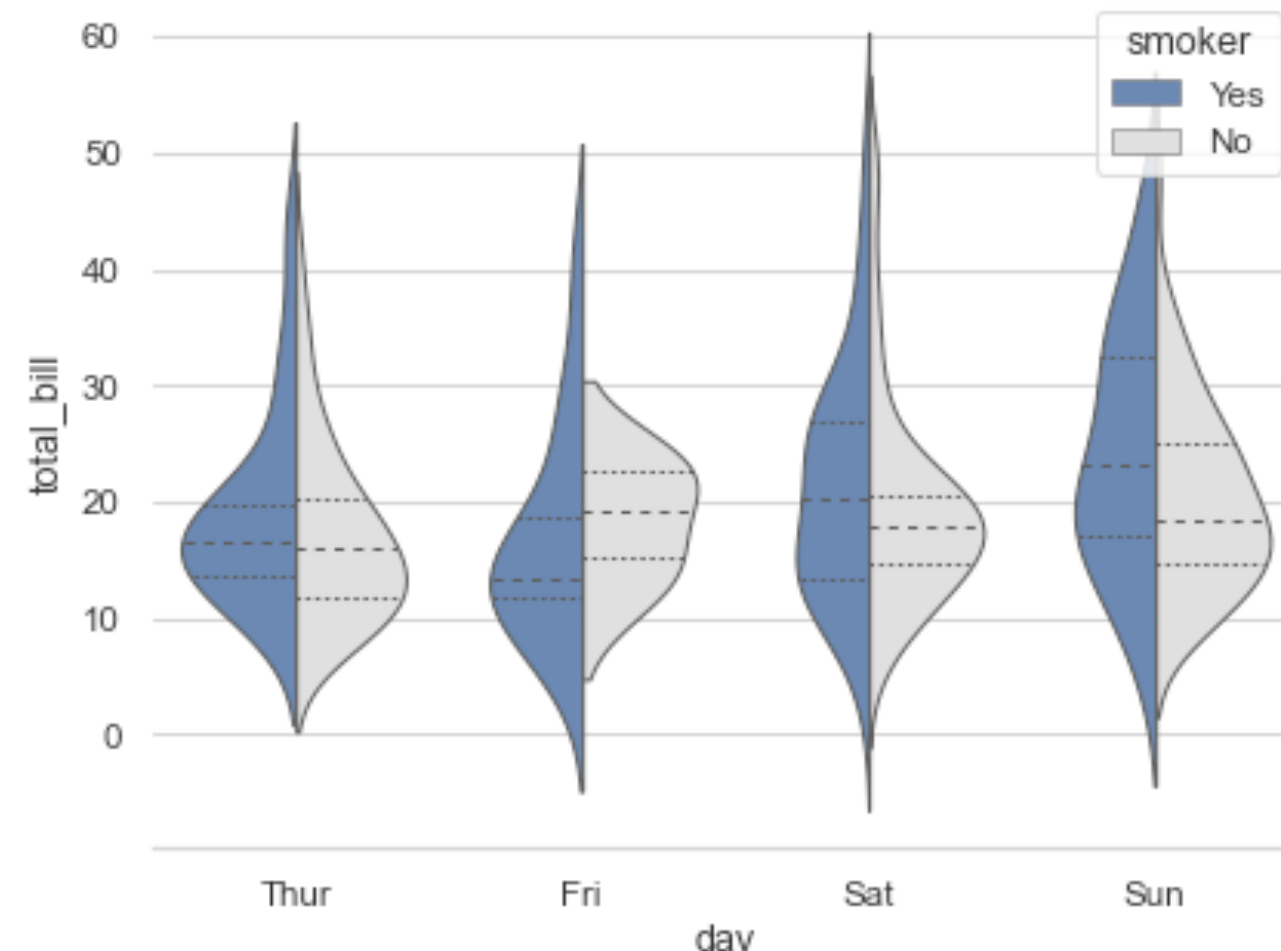
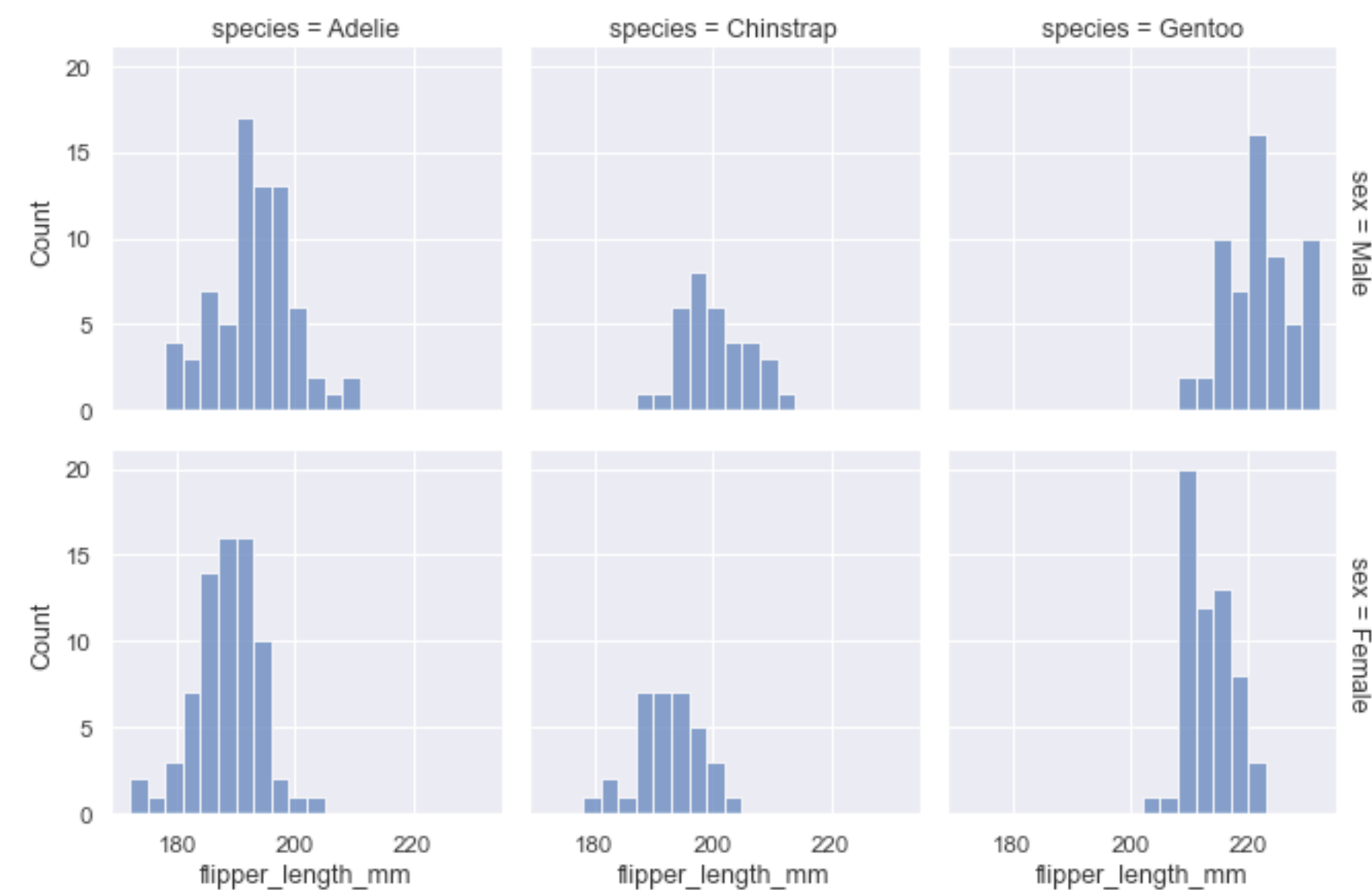


Data Visualization.

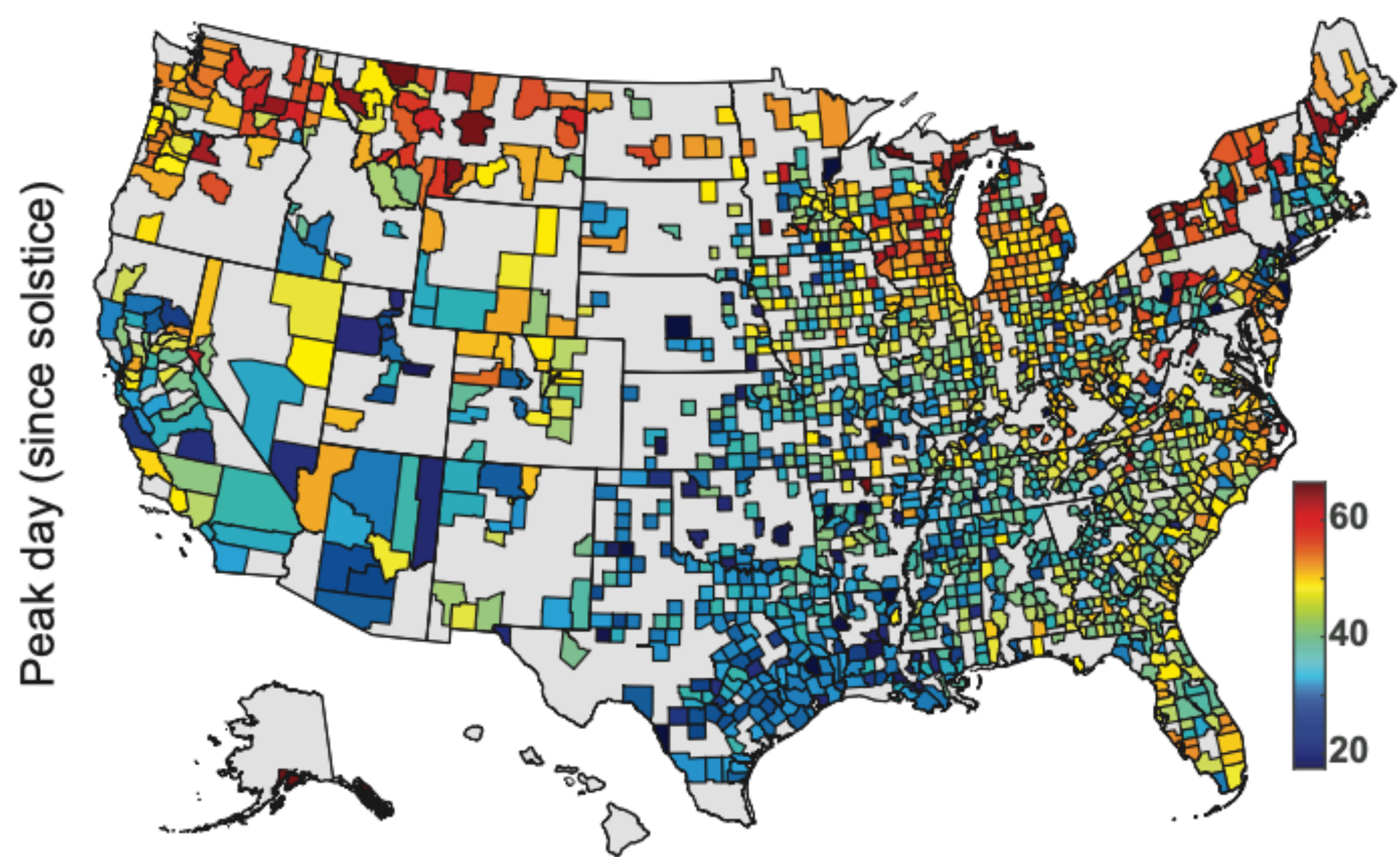


Data Visualization.





Data Visualization.

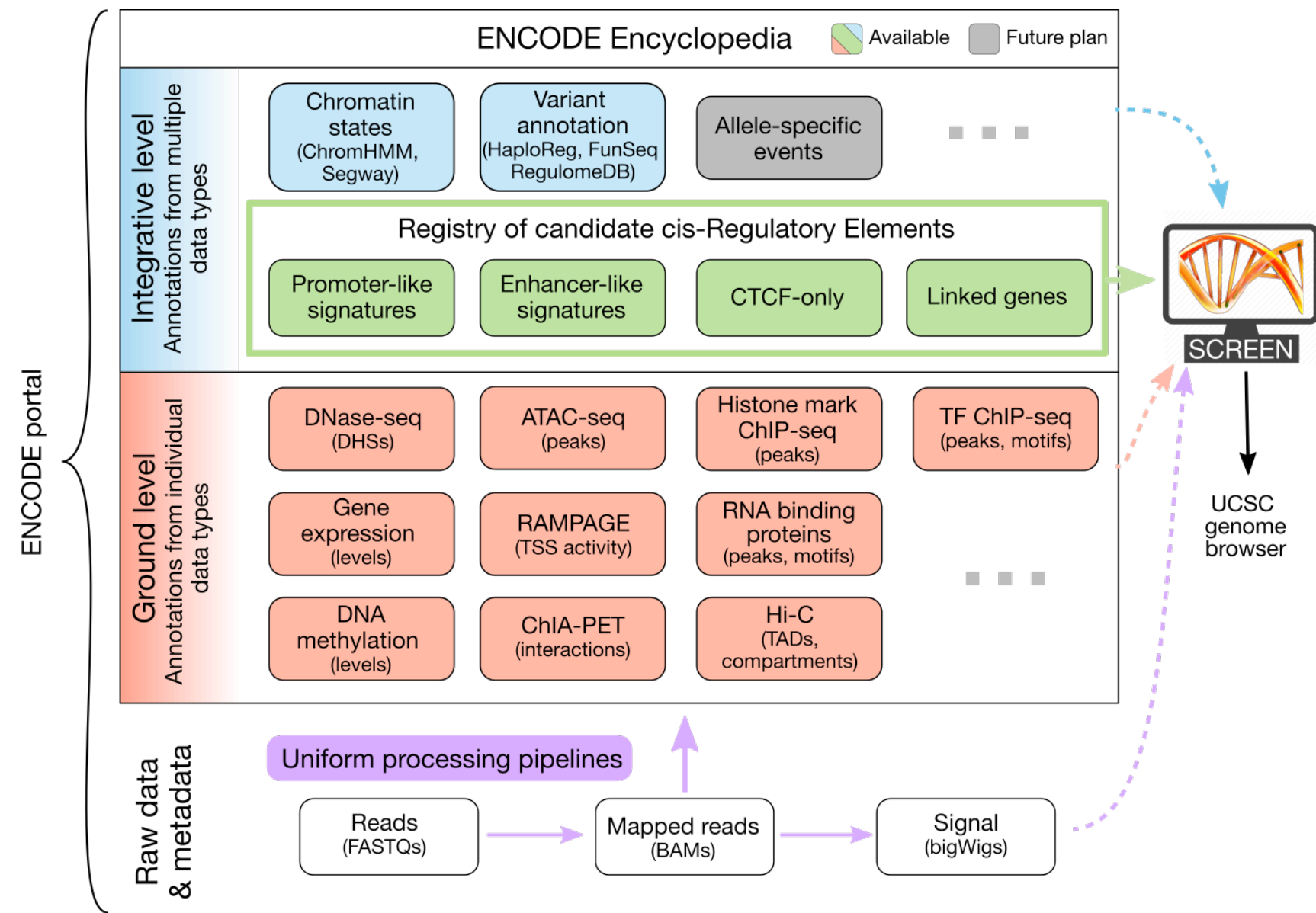


matplotlib



seaborn





Big Data!

NATIONAL CANCER INSTITUTE THE CANCER GENOME ATLAS

TCGA BY THE NUMBERS

TCGA produced over

2.5
PETABYTES
of data

To put this into perspective, **1 petabyte** of data is equal to

212,000
DVDs

TCGA data describes

33
DIFFERENT
TUMOR TYPES

...including

10
RARE
CANCERS

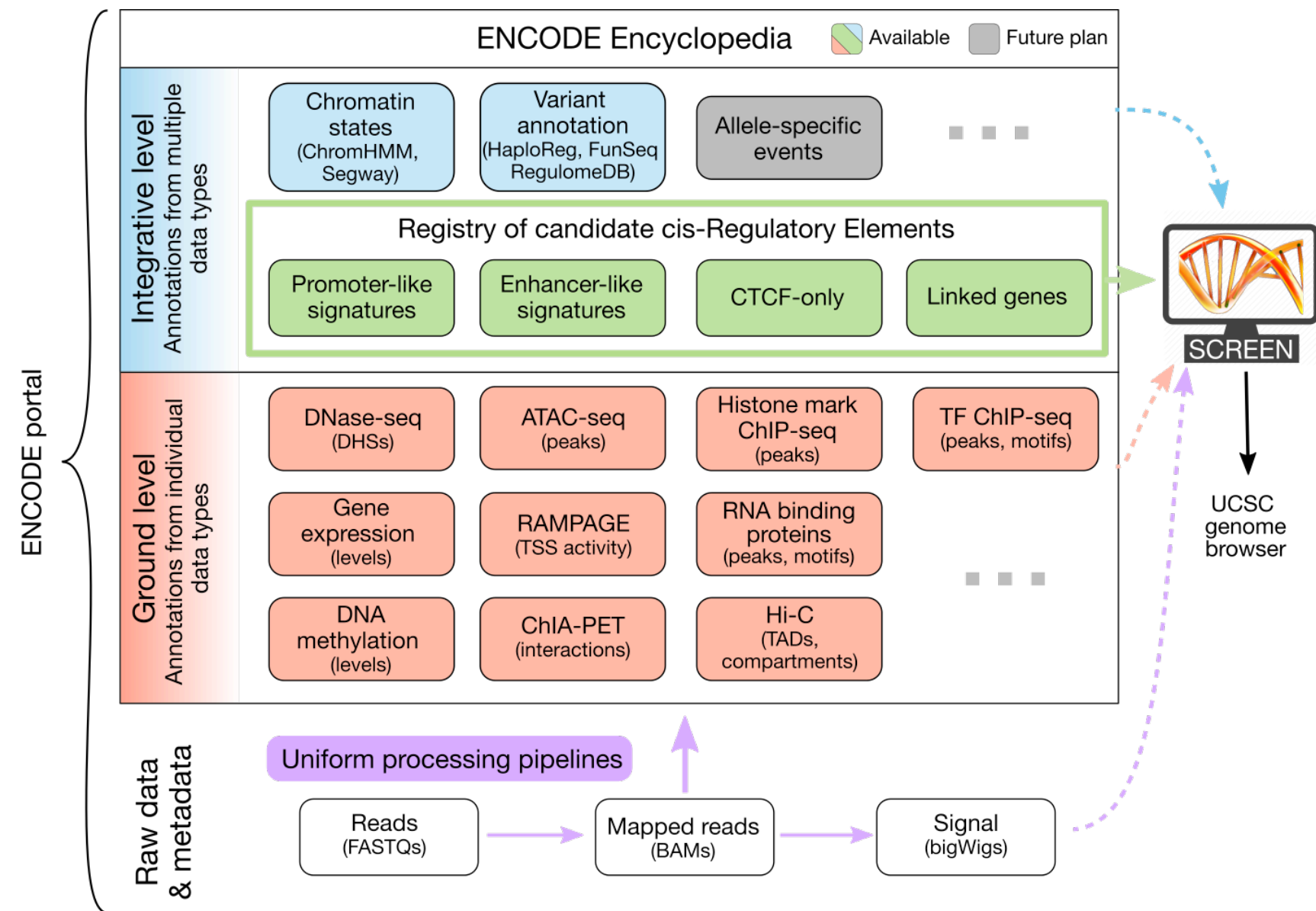
...based on paired tumor and normal tissue sets collected from

11,000
PATIENTS

...using

7
DIFFERENT
DATA TYPES





NATIONAL CANCER INSTITUTE THE CANCER GENOME ATLAS

TCGA BY THE NUMBERS

TCGA produced over

2.5
PETABYTES
of data

To put this into perspective, **1 petabyte** of data
is equal to

212,000
DVDs

TCGA data describes

33
DIFFERENT
TUMOR TYPES

...including

10
RARE
CANCERS

...based on paired tumor and normal tissue sets
collected from

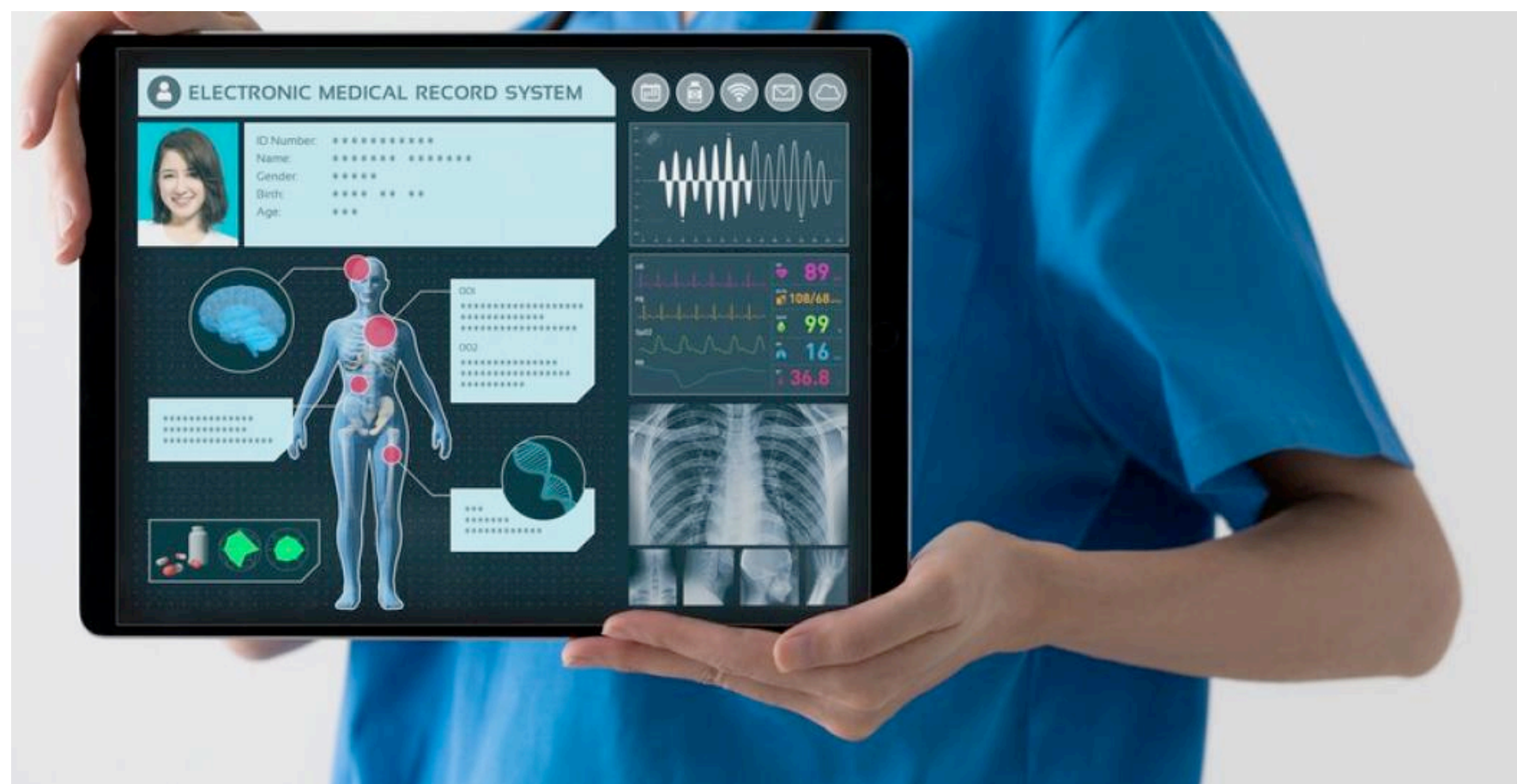
11,000
PATIENTS

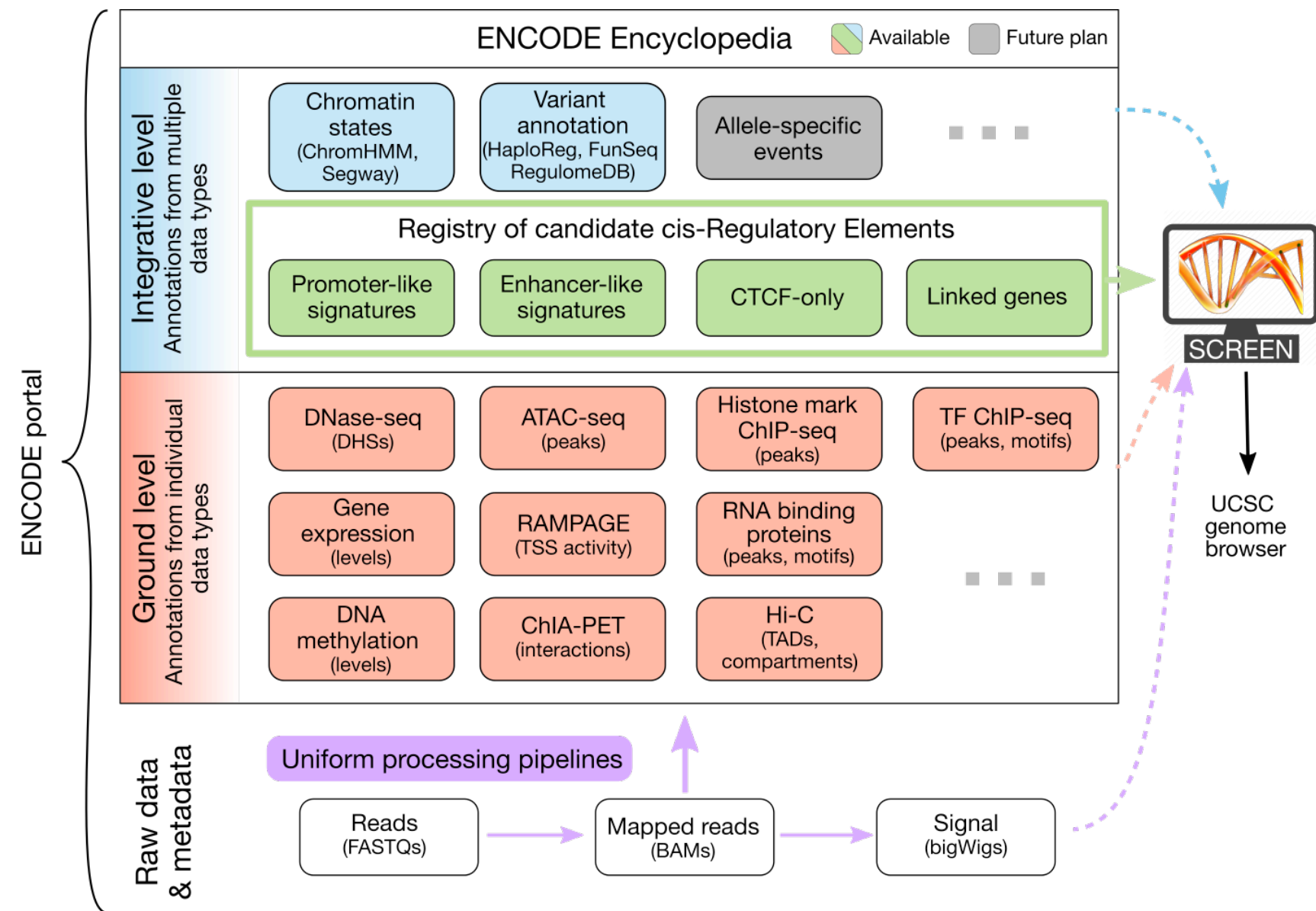
...using

7
DIFFERENT
DATA TYPES



Big Data!





NATIONAL CANCER INSTITUTE THE CANCER GENOME ATLAS

TCGA BY THE NUMBERS

TCGA produced over

2.5
PETABYTES
of data

To put this into perspective, **1 petabyte** of data is equal to

212,000
DVDs

TCGA data describes

33
DIFFERENT
TUMOR TYPES

...including

10
RARE
CANCERS

...based on paired tumor and normal tissue sets collected from

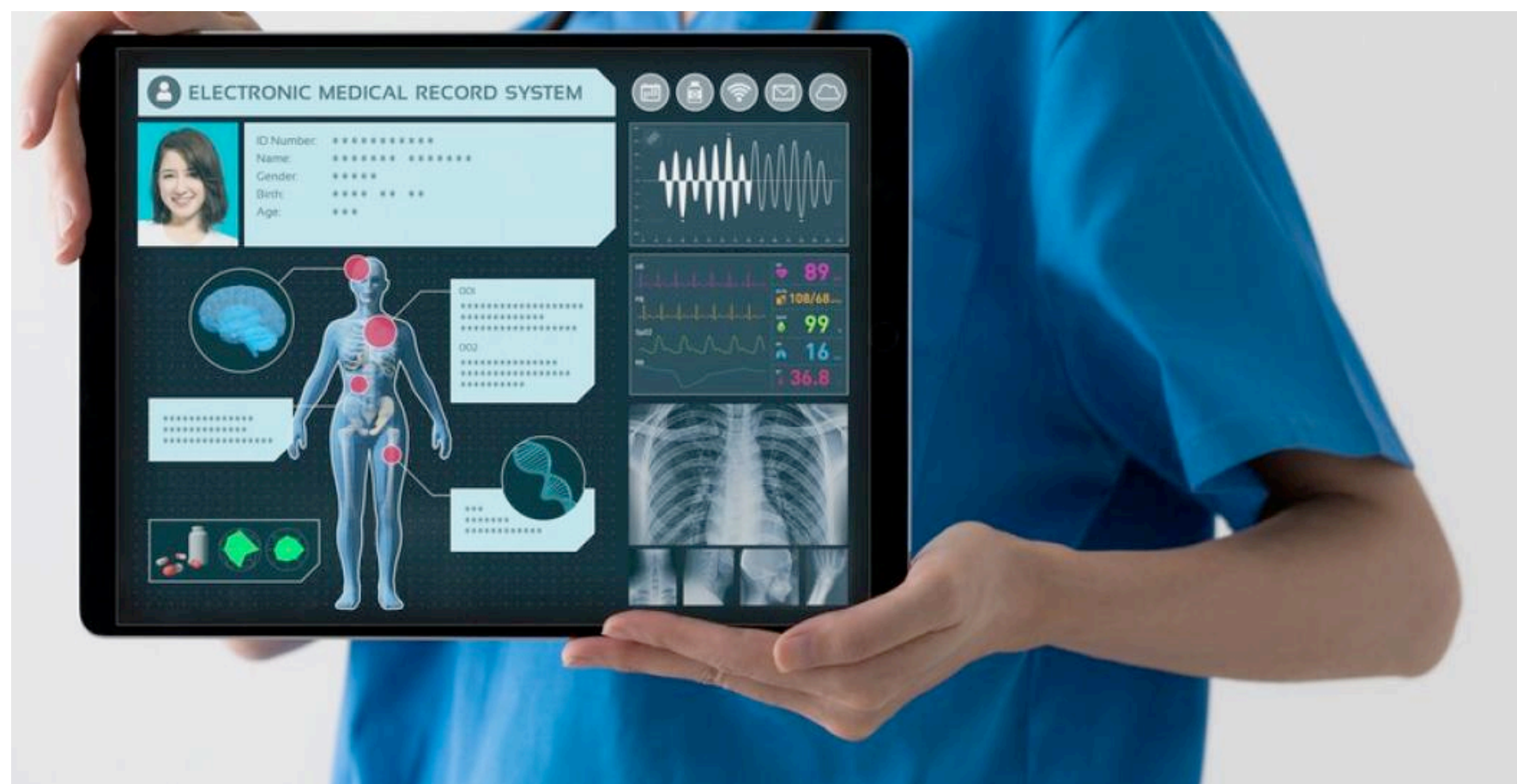
11,000
PATIENTS

...using

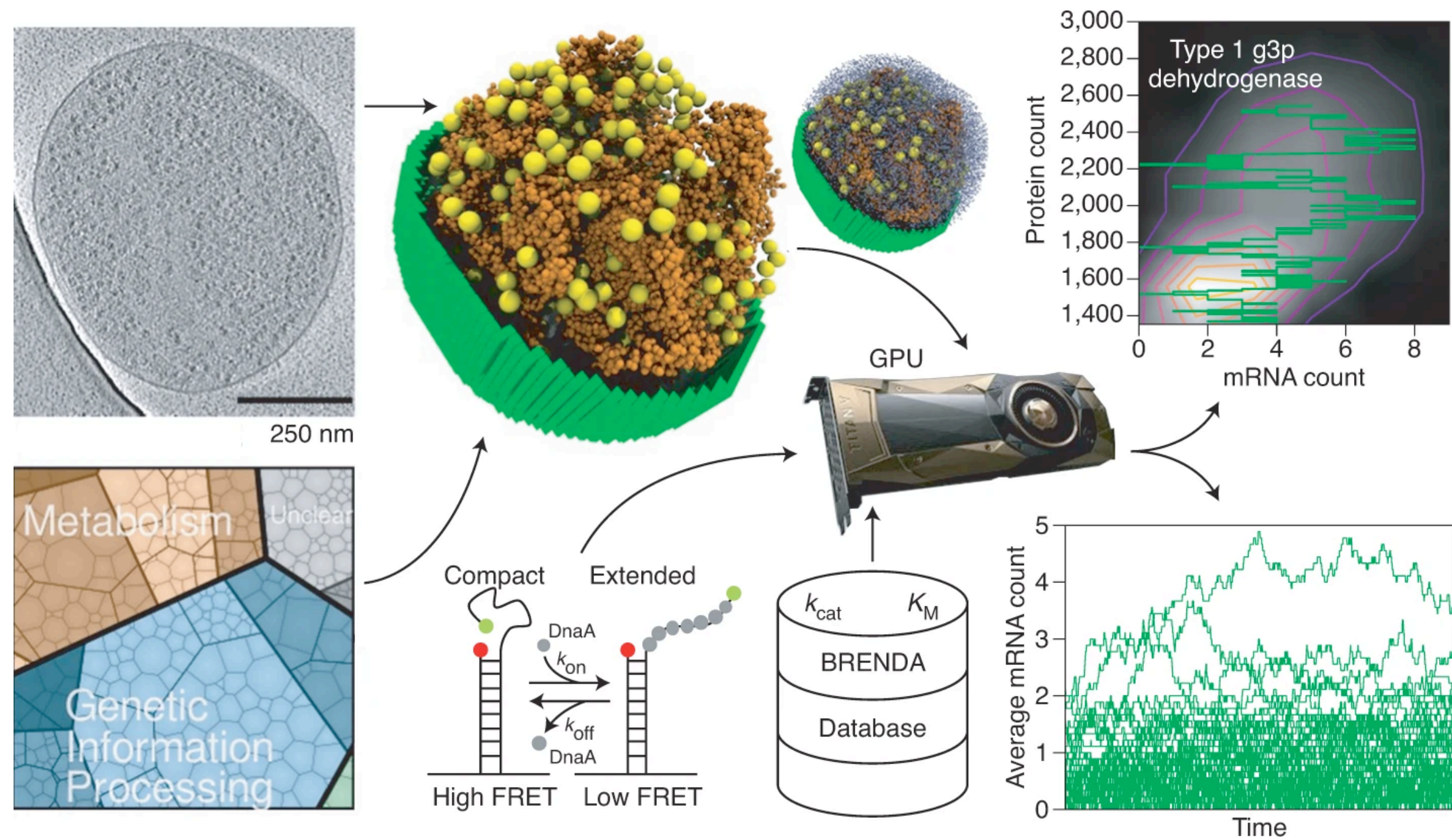
7
DIFFERENT
DATA TYPES



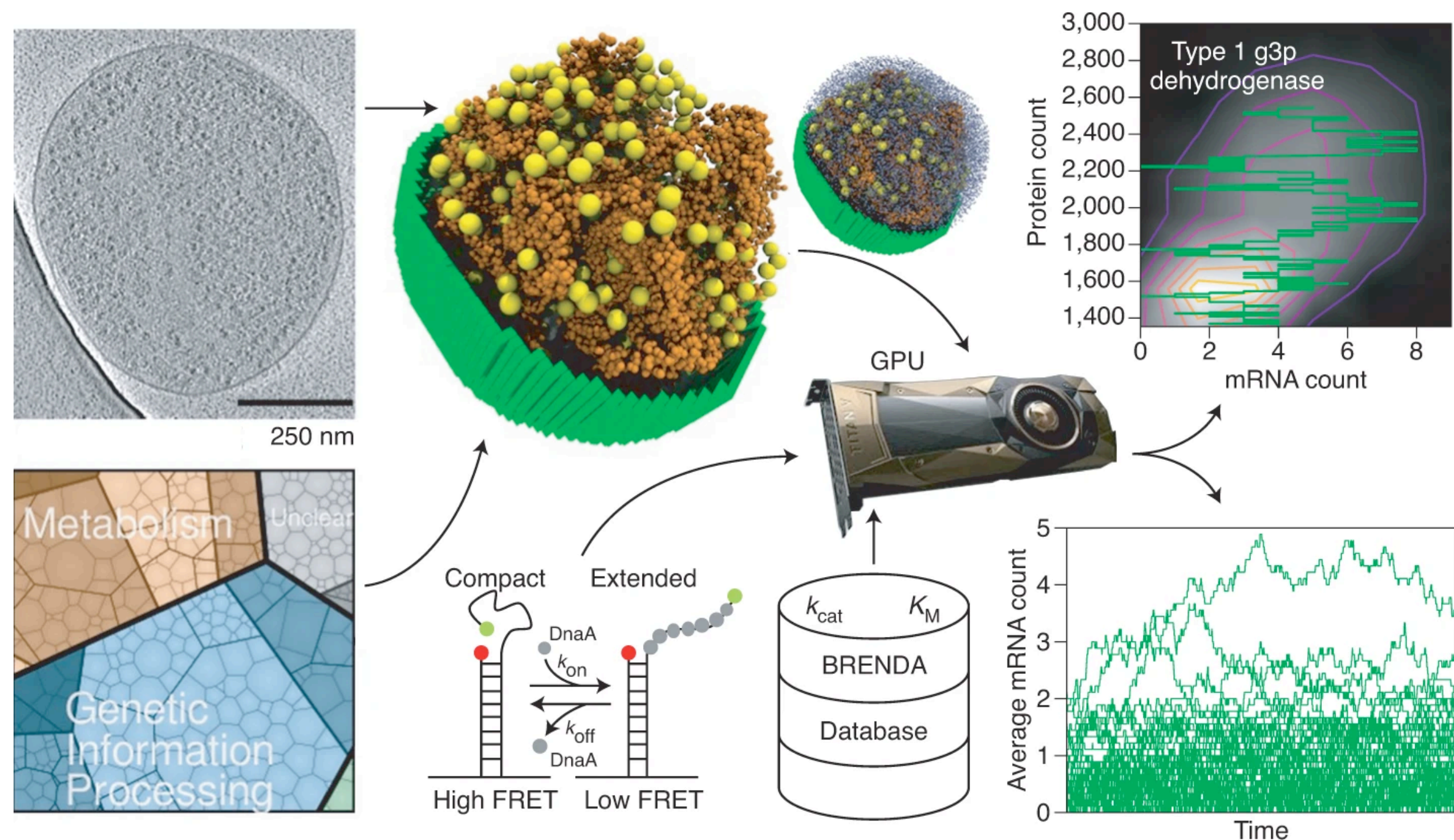
Big Data!



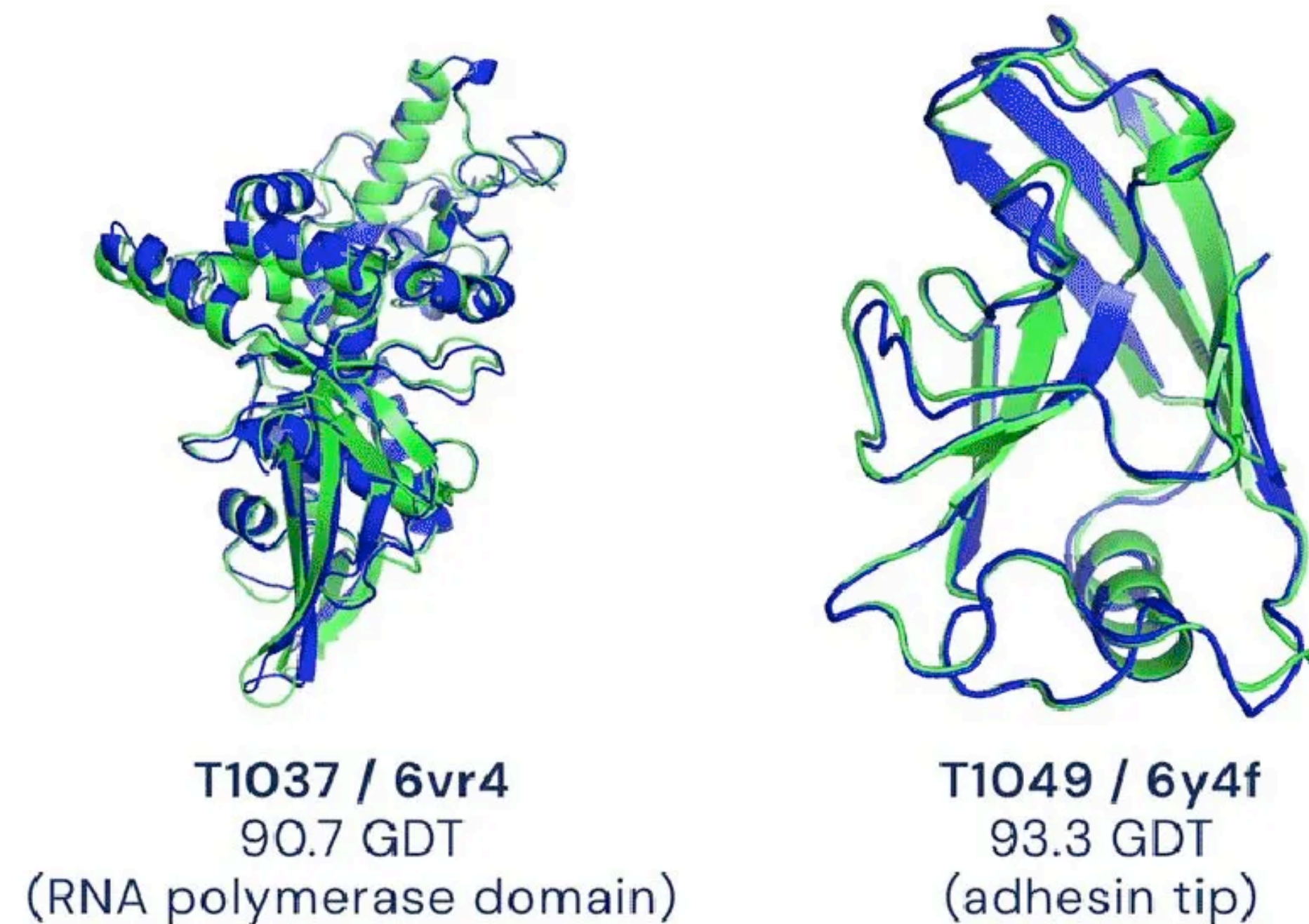
 **pandas**



Modeling, simulation, and AI



Modeling, simulation, and AI



- Experimental result
- Computational prediction



A widely used programming language that is valued for its read-ability and ease of use.

Is a common (if not the most common) language in data science, and interfaces well with a number of code libraries for machine learning and AI.

Our learning objectives for this summer:

- Build familiarity with Python data types (integers, float, boolean, list, dictionary) and flow control (if/then/else, while , for loops)
- Learn to read in and write out data
- Use Python to perform basic statistics (mean, variance)
- Use Python to create well-annotated plots

We will orient our lessons around a single (toy) data set.

Toy (or idealized, or fake) data play a useful role in research.
You can often use them to build your analyses or think about how to approach a problem before considering more challenging and less well-behaved real data.

But it is essential to be clear that they are not real, and are never presented as real

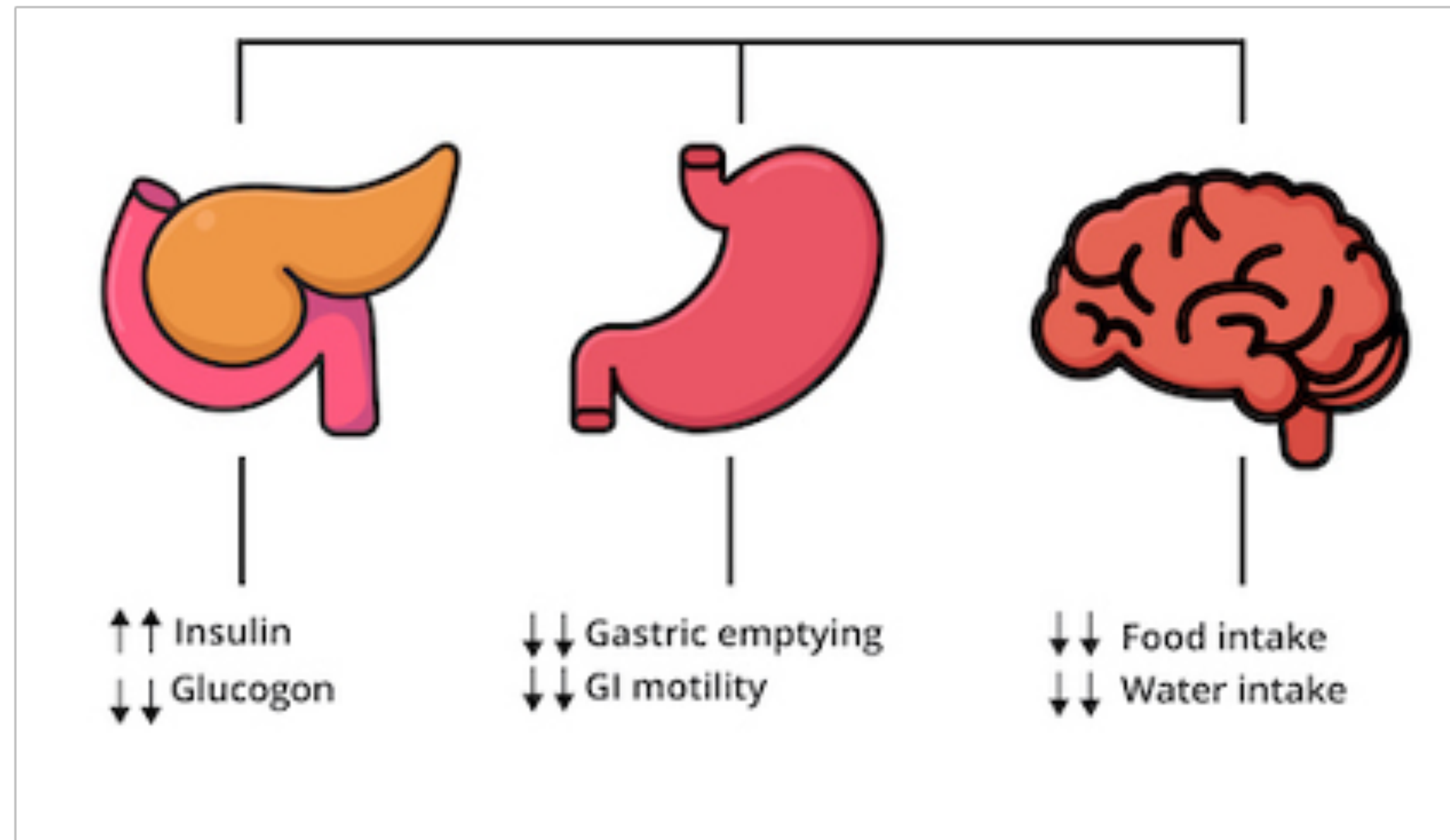
Mock Data Set: Mouse Obesity

Glucagon-Like Peptide -1

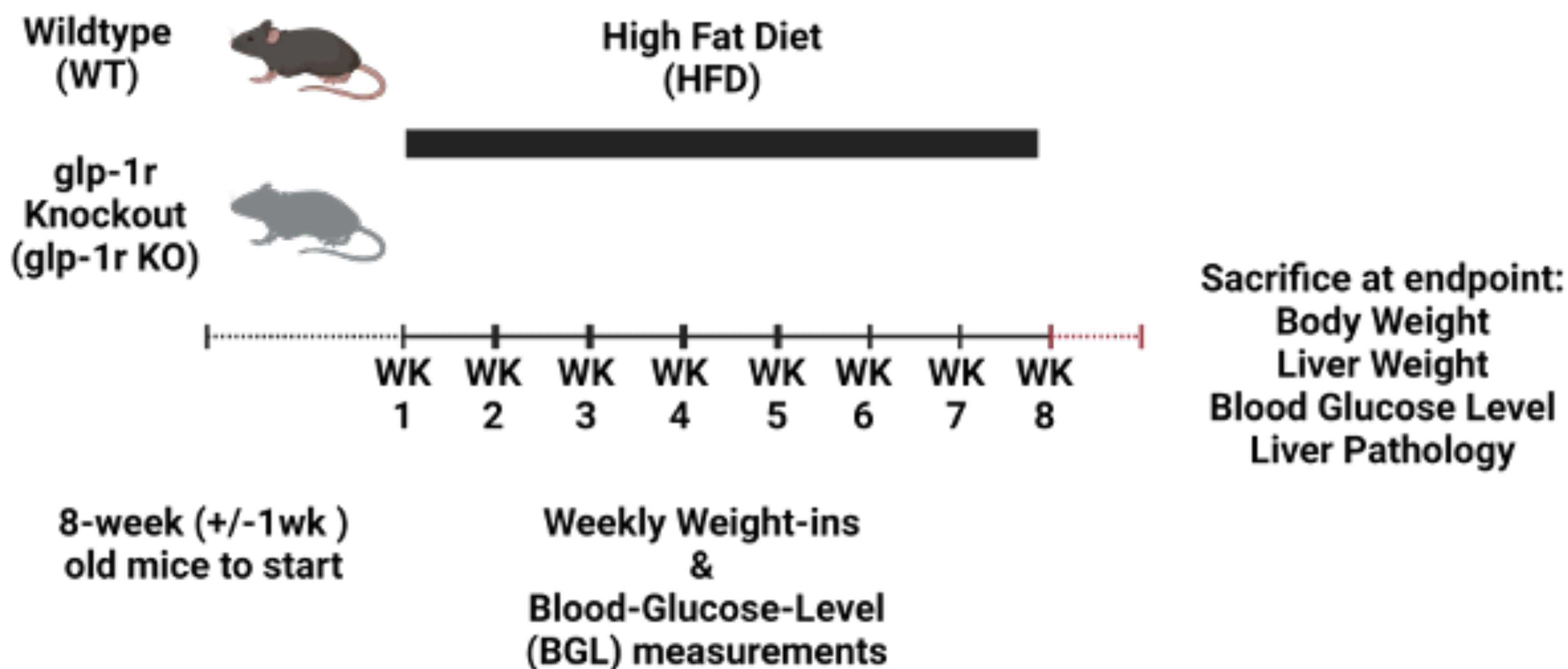
GLP-1

GLP-1 Agonist
(e.g. Liraglutide, Ozempic, etc.)

GLP-1 Receptor



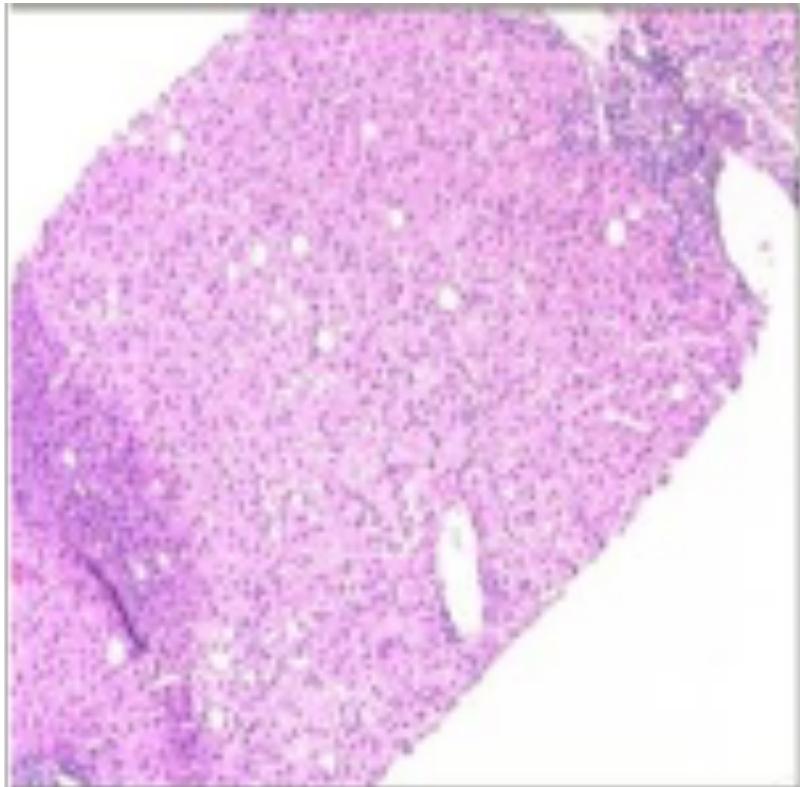
Mock Data Set: “Experiment” Layout



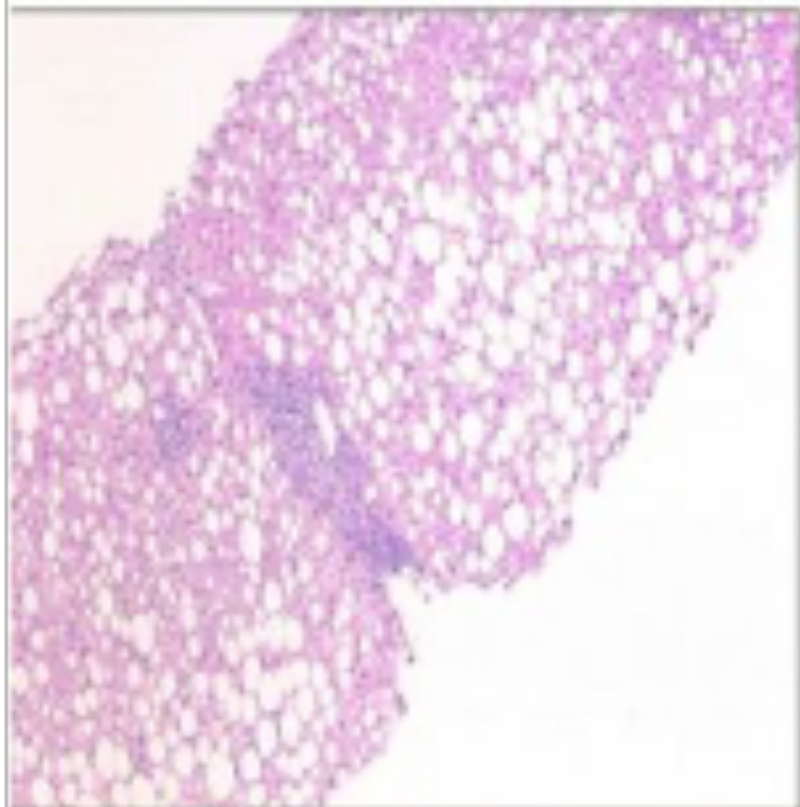
Mock Data Set:

Autotask CM Mouse Data - Saved to this PC Search																					
File Home Insert Page Layout Formulas Data Review View Autotask Help																					
Clipboard		Font				Alignment				Number				Styles				Cells		Editing	
Paste		Calibri 11				Merge & Center				General				Normal		Red		Insert		Autosum	
Format Painter		B I U								\$ %				Good		Neutral		Delete		Clear	
A1 ID																					
A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S			
ID	Sex	Genotype	Diet	Age (Weeks)	Final Weight (g)	Liver Weight (mg)	Final BGL (mg/dL)	Liver Pathology Note	BGL_wk1	BGL_wk2	BGL_wk3	BGL_wk4	BGL_wk5	BGL_wk6	BGL_wk7	BGL_wk8	Weight_wk1	Weight_wk2			
1	Male	gfp-tr_KO	HFD	15	40.51146472	1.55550808	134.678088	Steatotic	101.8058122	107.449056	120.260040	131.299495	140.5084264	150.8536249	161.4127812	168.926213	24.11470258	24.794953			
2	Female	gfp-tr_KO	HFD	17	34.6822048	1.29585803	177.8208432	Healthy	95.87959649	106.5099918	115.0126778	127.5696548	138.293380	150.3448071	157.4267265	169.4661799	21.88946166	22.5915426			
3	Male	WT	HFD	17	43.86713713	1.487167261	200.3633347	Steatotic	95.3856608	109.6880616	124.1798022	137.9606548	149.2034434	161.4983864	177.5085917	185.5657569	26.45253468	26.4464867			
4	Female	WT	HFD	16	36.86039164	1.190290341	196.8133095	Steatotic	98.4088899	109.4512382	120.4948729	133.6671329	147.4640012	160.138831	176.2666669	186.2765417	18.06383472	22.3952615			
5	Female	gfp-tr_KO	HFD	17	36.0861602	1.249046704	138.5615711	Fibrotic	95.48188683	108.9738037	115.5631237	127.6620669	136.0629312	146.0882614	164.212031	166.5080501	21.09376863	22.5275548			
6	Female	gfp-tr_KO	HFD	16	33.86949547	1.026632913	138.9138288	Healthy	98.34862718	107.7181863	117.0284033	126.363453	138.4540514	147.1401988	155.0377177	166.0770216	17.47931566	22.4064903			
7	Male	gfp-tr_KO	HFD	15	40.09168133	1.673888278	180.7101154	Steatotic	95.19188578	110.3032688	127.8171032	138.4985947	151.0188051	160.0718	169.4038333	180.8020154	24.54196471	26.264233			
8	Male	WT	HFD	16	40.53893679	1.453811417	202.1737805	Steatotic	100.5182489	111.57345	126.1134055	138.0544855	150.00047	160.8858117	177.3917379	188.3207019	24.58151177	26.1888267			
9	Female	gfp-tr_KO	HFD	16	39.50177634	1.207280718	174.2573917	Steatotic	94.65257158	108.3887434	117.8970788	129.582052	137.4935383	147.2657082	154.8242837	166.8847044	21.2397508	20.1389318			
10	Male	WT	HFD	16	41.23894134	1.388033881	201.4198013	Steatotic	101.8412482	117.7015288	129.8155812	135.4758117	149.1333851	161.7538773	176.1251376	186.828114	23.58144533	26.0888788			
11	Female	WT	HFD	15	33.259098	1.26803858	198.5014619	Steatotic	96.4993711	109.1096788	127.175777	138.40865	146.8858884	158.7488547	177.3978779	184.7788382	20.27882485	21.8886714			
12	Female	gfp-tr_KO	HFD	16	37.49880408	1.138846781	138.087775	Steatotic	95.5748788	109.571017	117.4288433	127.8435228	137.8914935	147.0135716	158.5258816	168.8885455	20.3532889	20.8813881			
13	Male	gfp-tr_KO	HFD	16	43.87191757	1.678614257	138.4075525	Steatotic	100.0789	118.7488835	117.3728779	137.4788882	148.5877794	157.8346548	168.8818779	187.9076681	24.60886734	27.7777784			
14	Female	gfp-tr_KO	HFD	16	33.43838798	1.174991385	138.8188889	Steatotic	98.67475843	110.4088517	118.8727529	130.584719	134.8188889	150.4582986	157.5058817	168.3887213	21.54813419	20.488495			
15	Female	WT	HFD	17	38.08877132	1.298877845	196.0897785	Steatotic	97.08481031	108.2947471	123.8881571	131.018853	149.3377763	158.8848147	171.8988748	181.588108	21.0789526	20.8158125			
16	Female	WT	HFD	16	38.23871443	1.272878711	171.078888	Steatotic	97.0817813	117.8182186	127.888877	136.4115917	147.388885	158.888888	176.071881	181.888882	23.8813881	21.8818881			
17	Male	gfp-tr_KO	HFD	16	41.71731844	1.498817716	182.8793816	Fibrotic	101.2718818	119.8888817	115.888888	137.088887	137.888887	157.888884	160.1718827	183.8718877	23.88812637	26.4918881			
18	Female	gfp-tr_KO	HFD	16	38.38888108	1.33888135	176.085881	Fibrotic	96.79887918	108.8788819	115.7888876	128.988882	134.888888	148.388884	157.8888879	168.288888	21.58888671	20.7188848			
19	Male	gfp-tr_KO	HFD	17	48.28938799	1.488883372	174.8793882	Steatotic	98.4718812	108.8278814	127.278126	137.8718812	151.0797383	160.8893831	180.0727886	188.788131	26.0788888	25.3888888			
20	Female	WT	HFD	16	35.28887998	1.301790571	177.4388791	Fibrotic	95.1011887	118.181887	118.878881	138.378881	148.3777882	159.4538855	171.8845882	185.8878887	18.8888788	19.1887888			
21	Male	gfp-tr_KO	HFD	16	42.25278148	1.37148288	188.1181388	Healthy	95.9788832	109.5827882	121.5971885	138.584888	148.111427	169.342188	180.881885	187.7788888	26.8887882	27.7477785			
22	Female	gfp-tr_KO	HFD	16	32.1178889	1.259186728	186.188887	Healthy	98.88818278	109.3188882	118.591371	138.8887885	137.3588213	144.188888	158.888888	168.877883	21.818888	21.7888882			
23	Male	WT	HFD	16	48.4788885	1.488887711	199.8177888	Steatotic	101.717784	111.717883	124.8888883	138.8178828	151.798888	167.288888	173.8888887	188.459783	23.888814	26.8813779			
24	Female	WT	HFD	16	37.88888882	1.148576599	200.7988879	Steatotic	95.8878872	107.1788879	121.888185	138.158887	149.8188148	158.8881329	171.1581276	185.8888889	19.88888873	19.7888888			
25	Female	WT	HFD	16	38.08417843	1.248858888	195.3188888	Steatotic	98.78852571	109.9858121	120.8888889	133.1288738	148.1288881	158.9788885	177.5788853	183.4278312	21.88885813	21.8888888			
26	Male	gfp-tr_KO	HFD	15	48.98818445	1.399126811	181.2188889	Fibrotic	102.1588888	109.5737885	120.781385	133.3478815	143.2917882	150.3888884	160.8817588	171.5888881	21.887888	25.5232882			
27	Female	WT	HFD	15	30.55381515	1.113487998	196.2078888	Fibrotic	97.8818888	108.8888813	127.8157888	137.8881382	147.5881889	160.8888843	173.8888872	185.8888883	17.88213885	21.0272789			
28	Male	WT	HFD	16	43.88888188	1.438888311	201.0018738	Steatotic	100.778848	112.8128814	127.8831224	141.3828817	147.2888889	162.9777878	178.988772	185.8888888	26.2888888	29.3382829			
29	Male	WT	HFD	15	40.88828129	1.528888281	201.8315717	Fibrotic	100.5977889	111.8288877	129.3188788	138.2282352	149.2387137	161.1782218	175.7888884	184.8817882	21.4991888	26.2758888			
30	Female	gfp-tr_KO	HFD	16	31.38888888	1.388881224	175.2887888	Healthy	100.5977889	107.1788812	126.1881353	137.8887884	148.2732214	157.888888	168.4281579	18.9777882	20.9487881				
31	Female	gfp-tr_KO	HFD	16	36.74215427	1.183047112	172.5388887	Fibrotic	95.8818889	107.8288884	117.3847718	129.8827887	138.5788817	148.821114	158.8979821	165.2197777	18.5387889	21.4993842			
32	Female	gfp-tr_KO	HFD	16	35.22488817	1.084138487	177.8228888	Healthy	98.87888882	104.7879434	116.3888887	122.7318887	139.7887774	148.4143386	157.8881814	167.5888888	19.88788126	21.0881283			
33	Female	gfp-tr_KO	HFD	16	32.84738137	1.243271882	175.8354371	Fibrotic	95.8887789	108.8442317	116.3188883	127.2528884	134.4257513	147.4788181	158.1888889	168.5888881	19.38888429	20.4732881			
34	Female	WT	HFD	16	35.72888888	1.147181583	196.2388888	Fibrotic	95.88812729	109.5488878	123.847189	133.8848213	148.822888	159.8925138	173.1888884	184.8484882	19.85218852	22.1588889			
35	Male	gfp-tr_KO	HFD	16	42.88888883	1.488881578	178.8884882	Steatotic	102.5847888	108.8188888	121.958888	138.829139	148.8184484	161.8429887	167.788888	17.81848131	25.8888884				
36	Male	gfp-tr_KO	HFD	17	43.0412987	1.318888919	180.8428888	Fibrotic	99.7881442	114.8888814	118.7723134	131.8228888	137.742583	148.8977889	161.5288721	171.8213845	27.51297884	27.9888841			

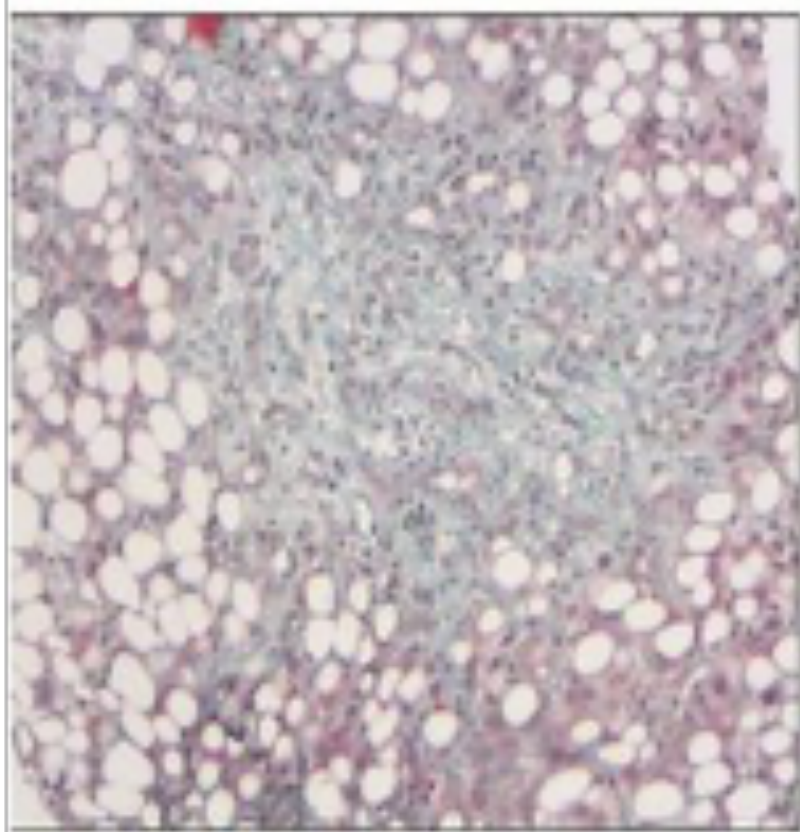
Normal Liver
(Healthy)



Fatty Liver
(Steatotic)



Liver Fibrosis
(Fibrotic)



Ok, so with this in mind, let's get started with installing anaconda.

Step-by-step:

- Navigate to the directory where you downloaded anaconda
- Double click to open the installer
- Click “next” to start the installation
- Agree to the license agreement
- Select to install the software for “just me” (not All Users/system wide)
- Choose Install Location. The default is usually a good choice (on Mac this will be your /Applications folder). Take note of where it is being installed.
- Do not add Anaconda to your PATH, but do register Anaconda as your default Python
- Wait for the install...
- You don't need to check the “learn more” buttons, you can click finish after the install is completed.

<https://www.geeksforgeeks.org/how-to-install-anaconda-on-windows/>

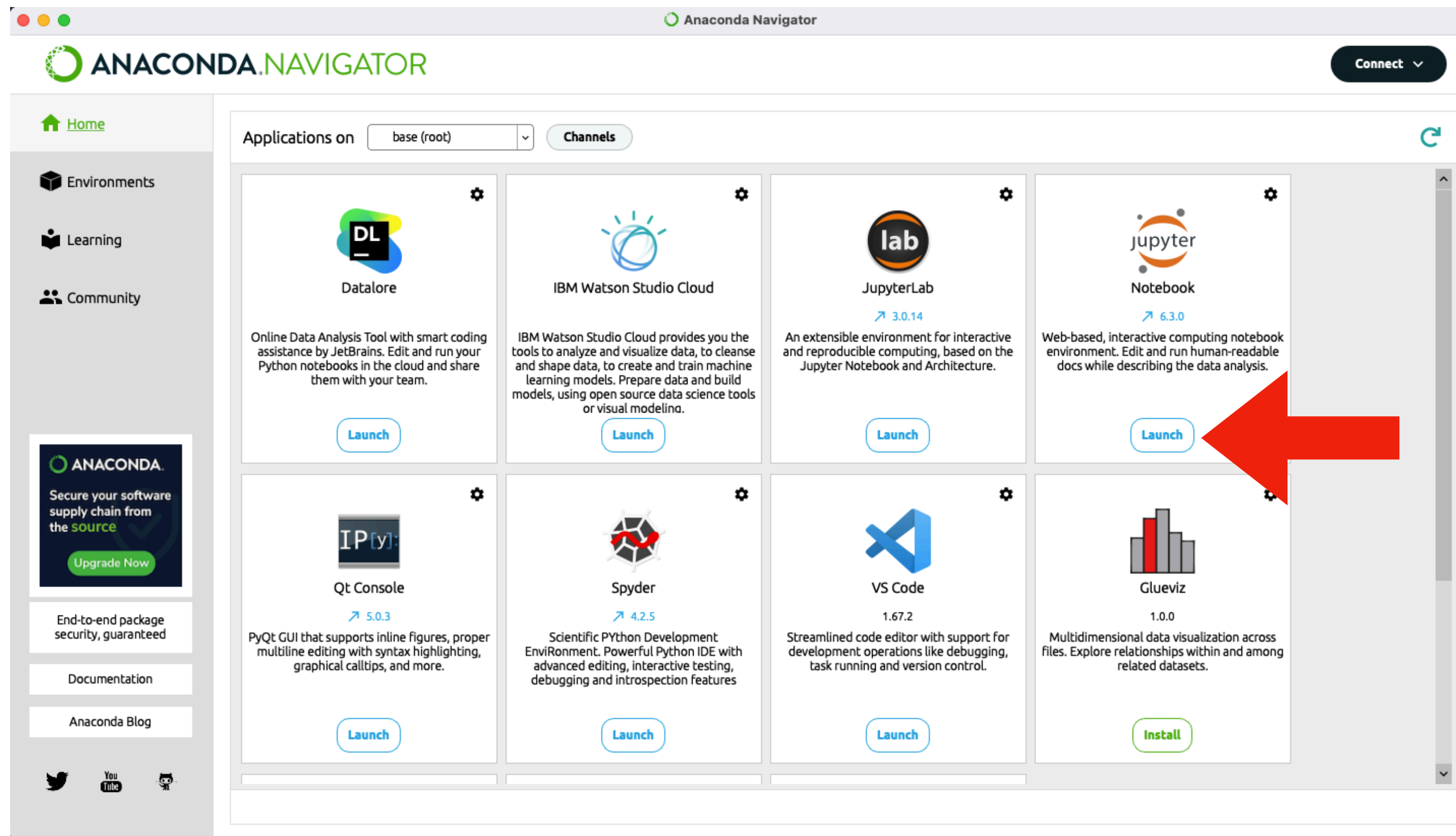
Now that you have anaconda installed, let's start by loading a Jupyter notebook!

On Mac:

Use the Finder to navigate to the applications folder
Double click to open the Anaconda-Navigator

On Windows:

Use the start menu to find the anaconda navigator
Double click to open the Anaconda-Navigator



So what is a Jupyter notebook?

It is a notebook for code, comments and plots that runs within a browser window.

Even though it is in a browser, it is not on the internet. It is running locally on your computer, and only you see your notebook

You can use the file navigator to find a spot where you want to create your notebook

Then click New -> Python 3

Now for our final exercise... let's get ready for the next class.

- * Close your Jupyter notebook and the Anaconda navigator
- * Download next week's notebook from here: <https://reynolds-lab.net/python-workshop-2024/>
- * Re-open the Anaconda navigator and open next week's notebook