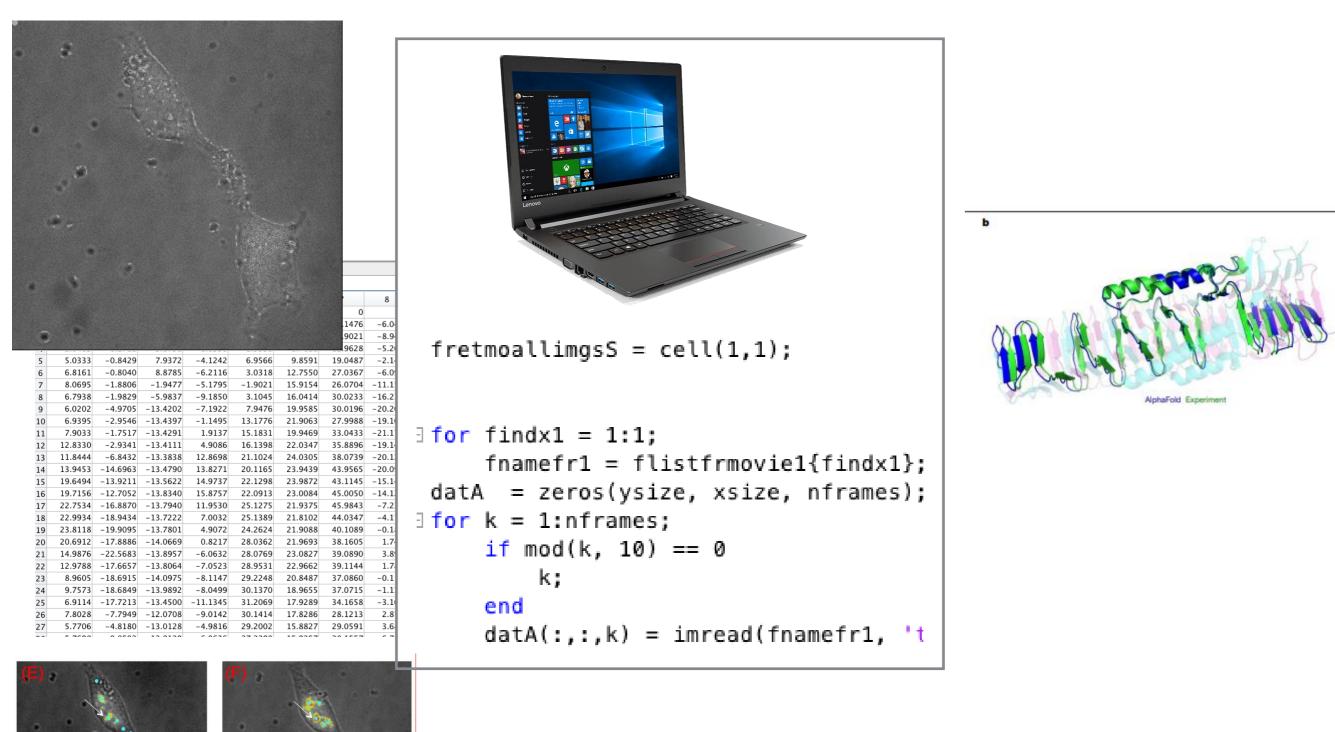
Introduction to Scientific Computation



Halil Bayraktar Lecture 1 - Introduction

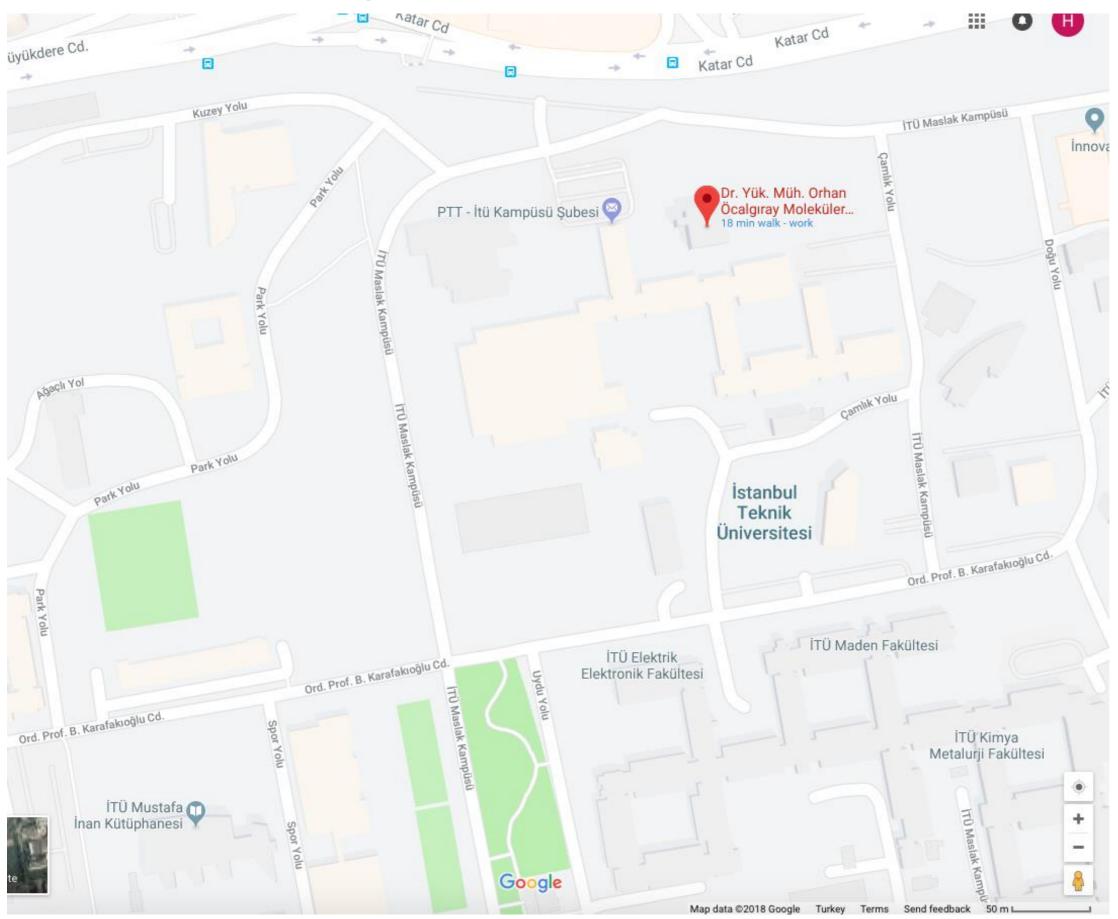
Course Instructor Halil Bayraktar



Associate Professor Molecular Biology and Genetic Department Office : MOBGAM 308 Course Teaching Assistant

Alican Topaloglu Umut Cagan Ucar

Molecular Biology and Genetics Department Building



Office hours : Monday14:30-15:00 pm and by appointment hbayraktar@itu.edu.tr

Course website

https://b2lab.wordpress.com/scientific-computing/

🕅 My Sites 🛛 🖬 Reader				Ct Write) 🌪
B2LAB	Home About	Research Pu	ublications Courses	 Contact Blog 	
new tools to explore cells			Scientific Computation 113	<	
			Biostatistics 271	<	
В	Biostatistics		Optical Imaging for Biology	<	
			Functional Imaging	I	
Edit	t		Introduction to Chemistry	<	
			Science	<	
	Syllabus Lecture Slides		Energy & Environment		
B2LAB / 🕅	Popular Readings Homework Assignm Other Materials	nents			

Grading

Percentage	Percentage	Date
Midterm 1	30	to be announced
Midterm 2	30	to be announced
Final	40	to be announced

A minimum average of 20 % from semester activities is required to attend final exam.

Books and other resources

No official textbook for this course



Please check other reserved books at ITU library and other online resources.

Course Objectives

Basic coding (MATLAB) knowledge to use in your research, other courses and workplace.

Learn coding/algorithm tools to evaluate scientific data

Enhance your data presentation skills for publications (graphs, pie charts, box plots vs.)

Improve your practical mathematics skills

Teaching skills of transition from raw data to presented data

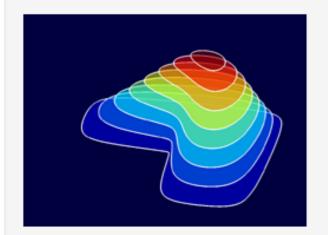
Why study scientific computation?

- Computer and numerical methods are essential to analyze complex data and provide analytical solutions.
- Presentation of your data with various plots such as line, histogram plot, polar plot, box plot etc.
- Analyze multidimensional data (matrix)
 Example, gene expression profile, microscope images, videos etc.
- Learn to simplify complex problems
- Guidelines to computation
- Learn how to write code to analyze your data

Textbooks by Cleve Moler

Cleve Moler is chairman and chief scientist at MathWorks. Moler was a professor of math and computer science for almost 20 years at the University of Michigan, Stanford University, and the University of New Mexico. In addition to being the author of the first version of MATLAB, Moler is one of the authors of the LINPACK and EISPACK scientific subroutine libraries. He is co-author of three textbooks on numerical methods.

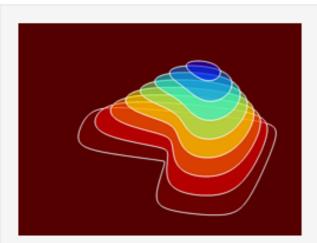




Numerical Computing with MATLAB

This lively textbook by Cleve Moler is designed for use in an introductory course in numerical methods, MATLAB, and technical computing.

- » View the textbook
- » Get curriculum tools and materials



Experiments with MATLAB

Under active development by Cleve Moler, this electronic book features chapters that supplement high school and early college courses in mathematics and technical computing, including calculus and matrix theory.

» View the e-book

Textbook and other resources

a) Numerical computing with Matlab by Cleve Moler, available for free download at

http://www.mathworks.com/moler/.

b) The Matlab demos are available at <u>https://www.mathworks.com/moler/exm/chapters.html</u>

c) Matlab help files

d) Matlab online resources (online videos)

https://www.mathworks.com/matlabcentral/answers/

What can you do with MatLab?

Advance computation/analysis

Data visualization/graphics

Code development

Modeling and simulation

Data acquisition

Graphical user interface

and more...

- Coding for data analysis
- Create your own functions
- Advance plotting
- Create user interfaces (not ideal)
- Interface with other programming languages (JAVA, Python, C/C++)

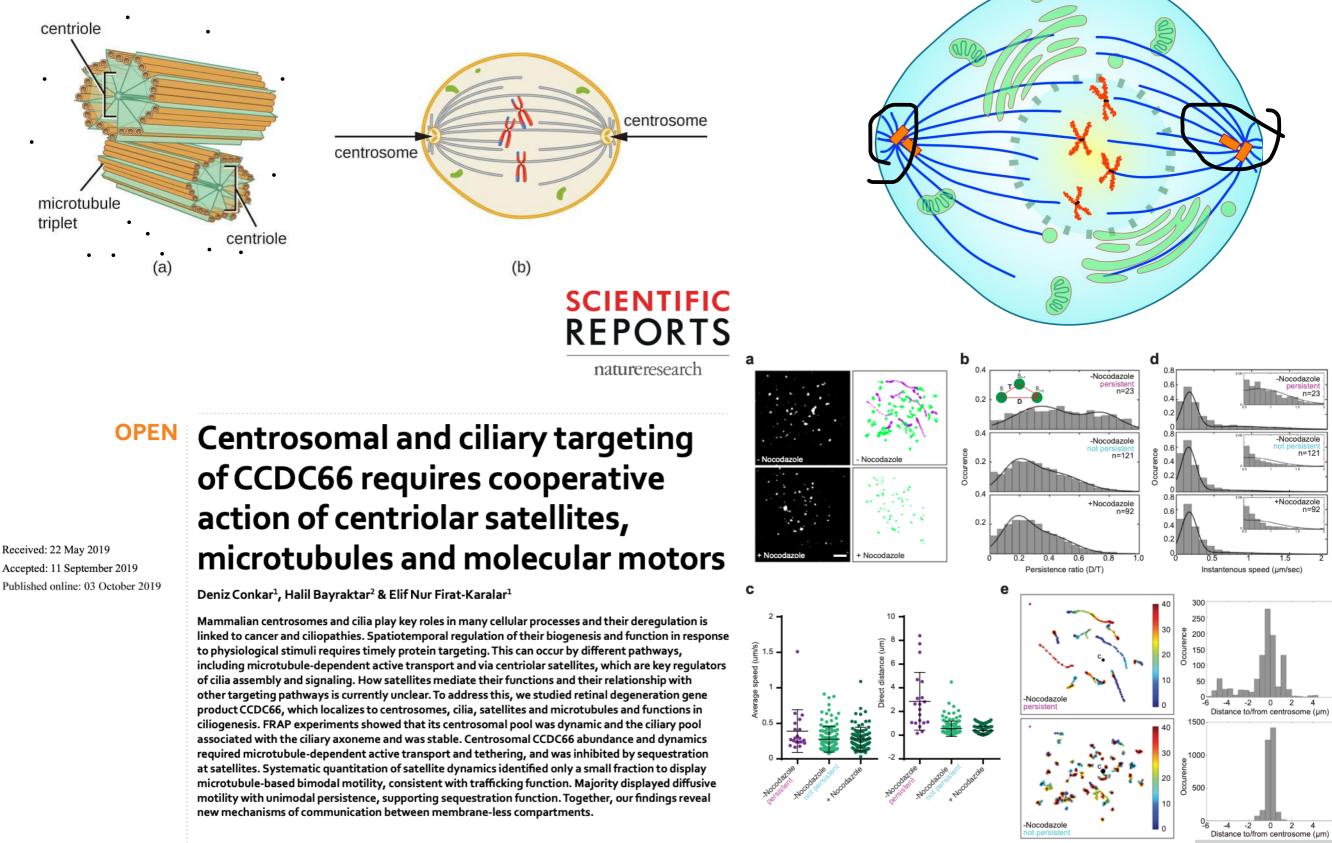
Raw data

```
Presented data
```

∫Afr ≍									
45x45 double	2							-	
1	2	3	4	5	6	7	Server Arren Barren Barren	1.04	
L 0	0	0	0	0	0	0		1.04	
1.8719	2.1662	1.9119	-6.1040	0.9919	0.7463			e	
3 2.8526	0.2256	3.8202	-4.0920	0.1165	5.7883		94	1.02 Inoresence I	
2.0403	0.2782	5.8758	-6.0550	1.9590	8.7028		26	S	ale at
5 5.0333	-0.8429	7.9372	-4.1242	6.9566	9.8591				let, e. a. Ale de la Ale de la Carde
6.8161	-0.8040	8.8785	-6.2116	3.0318	12.7550				ALL THE CONTRACTOR OF A DATA
7 8.0695	-1.8806	-1.9477	-5.1795	-1.9021	15.9154	26.0704 -1			talka an
6.7938	-1.9829	-5.9837	-9.1850	3.1045	16.0414	30.0233 -1		E I I I I I I I I I I I I I I I I I I I	
6.0202			-7.1922	7.9476	19.9585	30.0196 -2		E 0.98	
0 6.9395		-13.4397	-1.1495	13.1776	21.9063	27.9988 -1 33.0433 -2			
1 7.9033 2 12.8330		-13.4291 -13.4111	1.9137 4.9086	15.1831 16.1398	19.9469 22.0347	33.0433 -2 35.8896 -1	all a start a		
3 11.8444		-13.3838	12.8698	21.1024	22.0347	38.0739 -2		0.96	
4 13.9453			13.8271	20.1165	23.9439	43.9565 -2	<pre>fretmoallimgsS = cell(1,1);</pre>	- I shiked a	
5 19.6494	-13.9211		14.9737	22.1298	23.9455	43.1145 -1	· · · · · · · · · · · · · · · · · · ·		· · · · · · · · · · · · · · · · · · ·
6 19.7156	-12.7052		15.8757	22.0913	23.0084	45.0050 -1		0 400 800 120	00 1600
7 22.7534	-16.8870		11.9530	25.1275	21.9375	45.9843 -		Time(ms)	
8 22.9934			7.0032	25.1389	21.8102	44.0347 -		11110(1110)	
9 23.8118			4.9072	24.2624	21.9088	40.1089 -	<pre>for findx1 = 1:1;</pre>		
0 20.6912			0.8217	28.0362	21.9693	38.1605	-		
1 14.9876	-22.5683	-13.8957	-6.0632	28.0769	23.0827	39.0890	fnamefr1 = flistfrmovie1{findx1};		
2 12.9788	-17.6657	-13.8064	-7.0523	28.9531	22.9662	39.1144	<pre>datA = zeros(ysize, xsize, nframes);</pre>		
3 8.9605	-18.6915	-14.0975	-8.1147	29.2248	20.8487	37.0860 -			
4 9.7573	-18.6849		-8.0499	30.1370	18.9655	37.0715 -	<pre>for k = 1:nframes;</pre>		
5 6.9114		-13.4500	-11.1345	31.2069	17.9289	34.1658 -	-		
6 7.8028	-7.7949	-12.0708	-9.0142	30.1414	17.8286	28.1213	if mod(k, 10) == 0	()	
7 5.7706	-4.8180	-13.0128	-4.9816	29.2002	15.8827	29.0591	k;	(Tell)	
5 7 C 7 C 7 C 7 C 7 C 7 C 7 C 7 C 7 C 7	0.0503	12 0120	C 0000	27 2200	15 0057	201557	N,		
							end	Congratulations, A.C.	
								(it only took you 2 1 200	
							datA(:,:,k) = imread(fnamefr1, 't	65298 seconds	0.0
								20101 - 1	
								A Barrow March	
								See all the second	

merer jolpatica uk

Scientific computation is a key tool to decipher the mechanism of biology systems.



negative: to the centrosome (C) cositive: away from the centrosome (C)



Adaptive tracking algorithm for trajectory analysis of cells and layer-by-layer assessment of motility dynamics

Mohammad Haroon Qureshi^{a,b}, Nurhan Ozlu^a, Halil Bayraktar^{c,*}

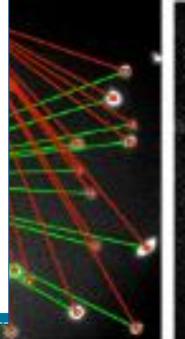
^a Department of Molecular Biology and Genetics, Koç University, Rumelifeneri Yolu, Sariyer, 34450, Istanbul, Turkey
 ^b Center for Translational Research, Koç University, Rumelifeneri Yolu, Sariyer, 34450, Istanbul, Turkey
 ^c Department of Molecular Biology and Genetics, Istanbul Technical University, Maslak, Sariyer, 34467, Istanbul, Turkey

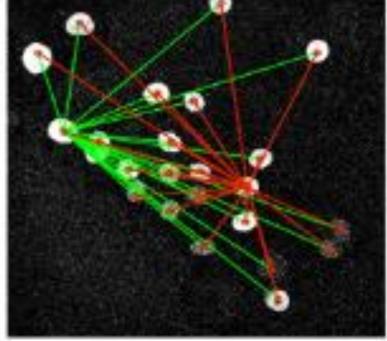
A R T I C L E I N F O

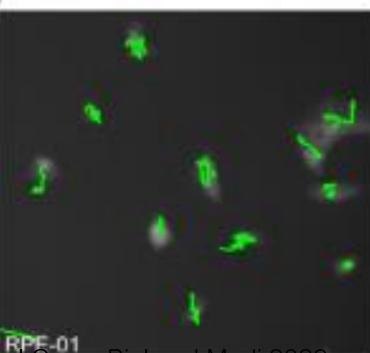
ABSTRACT

Keywords: Tracking Segmentation Algorithms Image processing Cell trajectories Linkage analysis Networks Open source software Motility dynamics Single cell

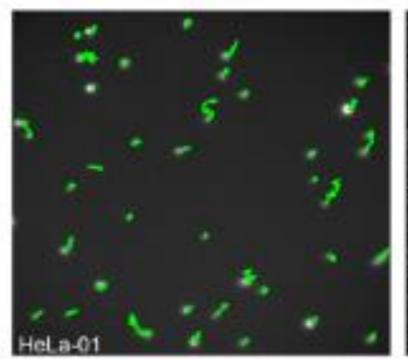
Tracking biological objects such as cells or subcellular components imaged with time-lapse microscopy enables us to understand the molecular principles about the dynamics of cell behaviors. However, automatic object detection, segmentation and extracting trajectories remain as a rate-limiting step due to intrinsic challenges of video processing. This paper presents an adaptive tracking algorithm (Adtari) that automatically finds the optimum search radius and cell linkages to determine trajectories in consecutive frames. A critical assumption in most tracking studies is that displacement remains unchanged throughout the movie and cells in a few frames are usually analyzed to determine its magnitude. Tracking errors and inaccurate association of cells may occur if the user does not correctly evaluate the value or prior knowledge is not present on cell movement. The key novelty of our method is that minimum intercellular distance and maximum displacement of cells between frames are dynamically computed and used to determine the threshold distance. Since the space between cells is highly variable in a given frame, our software recursively alters the magnitude to determine all plausible matches in the trajectory analysis. Our method therefore eliminates a major preprocessing step where a constant distance was used to determine the neighbor cells in tracking methods. Cells having multiple overlaps and splitting events were further evaluated by using the shape attributes including perimeter, area, ellipticity and distance. The features were applied to determine the closest matches by minimizing the difference in their magnitudes. Finally, reporting section of our software were used to generate instant maps by overlaying cell features and trajectories. Adtari was validated by using videos with variable signal-to-noise, contrast ratio and cell density. We compared the adaptive tracking with constant distance and other methods to evaluate performance and its efficiency. Our algorithm yields reduced mismatch ratio, increased ratio of whole cell track, higher frame tracking efficiency and nt of motility to characterize single

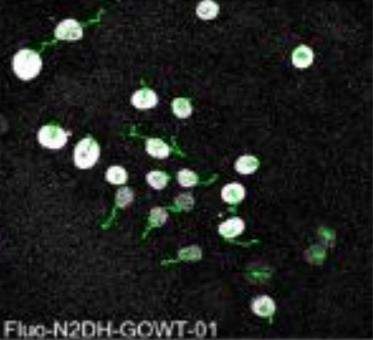






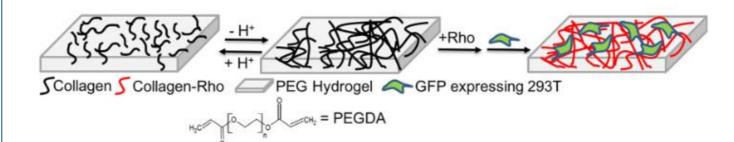
Qureshi et al Comp Biol and Medi 2022

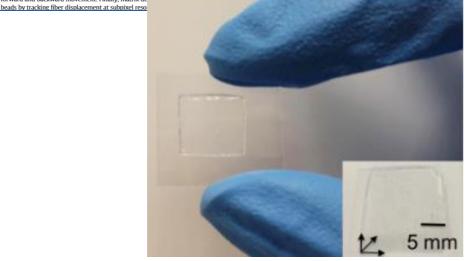


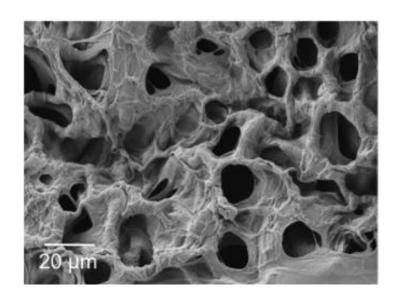


Material science and computation: Adaptive tracking to study cell motility at PEG-collagen hydrogels

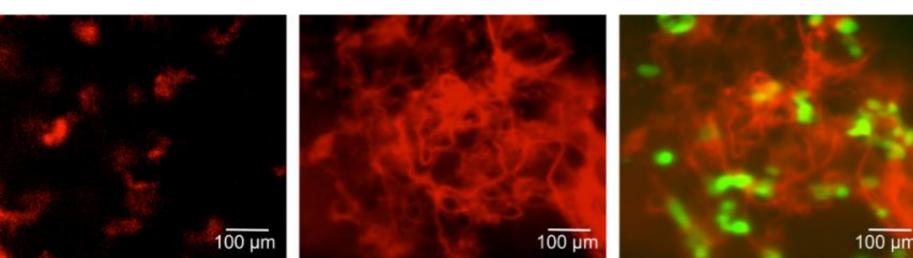
International Journal of Biological Macromolecules 157 (2020) 414-423 Contents lists available at ScienceDirect International Journal of Biological Macromolecules journal homepage: http://www.elsevier.com/locate/ijbiomac Alteration of cell motility dynamics through collagen fiber density in photopolymerized polyethylene glycol hydrogels Ozge Begum Akalin^a, Halil Bayraktar^{b,*} ^a Biomedical Science and Engineering, Koç University, Rumelifeneri Yolu, Sariyer, 34450 Istanbul, Turkey ^b Department of Molecular Biology and Genetics, Istanbul Technical University, Maslak, Sariyer, 34467 Istanbul, Turkey ARTICLE INFO ABSTRACT Article history: Received 16 December 2019 Received in revised form 3 April 2020 Accepted 18 April 2020 Available online 25 April 2020 Polyethylene glycol (PEG) hydrogels that have natural fibers mimicking extracellular matrix can be used as a model to understand the role of substrate properties on cell growth and migration. Due to the dependence of cell movement to adhesion, characterization of motility is needed to prepare biocompatible substrates. We dem-onstrated a method to encapsulate collagen into PEG hydrogel crosslinked via photopolymerization and studied the effect of fiber density on motility dynamics. Porous hydrogel immersed into collagen solution was coated with fibers after neutralizing solution. We provided a detailed study of cell instantaneous/average speed total displacement, persistence and angular displacement. We found that cells demonstrated a biphasic motility where a maximum speed of 17.4 μ m/h with a total distance of 215 μ m and persistence of 0.43 were obtained Keywords: Hydrogel PEG Collagen Single-cell tracking Motility dynamics Parrieteneo at 1.2 mg/ml collagen. High occurrence of low angular displacement observed at intermediate fiber density sug gests that cells tend to move forward along hydrogels_Increased anisotrony at low density was an indication of forward and backward movement. Finally, matrix de











Machine Learning and AI

- Supervised learning: Build a model based on input and output data
- Unsupervised learning: Build a model based on input data

Supervised learning uses classification and regression tech to develop a model Example,

Can a model predict the diving and nondiving cells in images? Can a model predict the negative or positive feedbacks from a text message? Can a model predict 3D structure of proteins? Can a model discover a new drug?

Process of learning

•Collection: Gathering data

- •Data Preprocessing: Cleaning and preparing the data for machine learning algorithms.
- •Model Training: Building a model by fitting it to the prepared data.
- •Model Evaluation: Assessing the performance of the trained model on unseen data.
- •Prediction: Using the trained model to make predictions on new data.

Protein Structure prediction with AI

AlphaFold developed by deepmind

It precits the protein's #D structure based on its amino acid sequence

AlphaFold Protein Structure Database (ebi.ac.uk)

AlphaFold - Google DeepMind

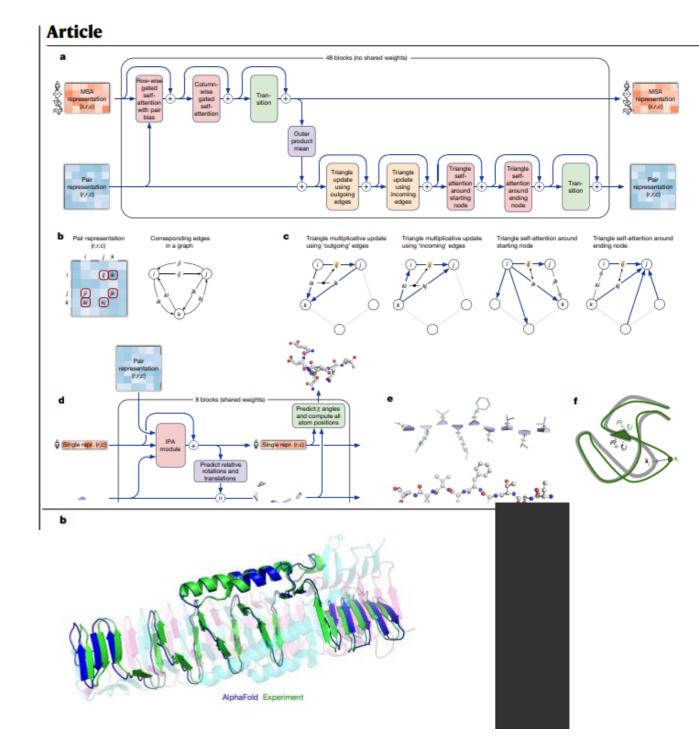
Article

Highly accurate protein structure prediction with AlphaFold

https://doi.org/10.1038/s41586-021-03819-2
Received: 11 May 2021
Accepted: 12 July 2021
Published online: 15 July 2021
Open access
Check for updates

John Jumper^{1,4}, Richard Evans¹⁴, Alexander Pritzel^{1,4}, Tim Green¹⁴, Michael Figurnov¹⁴, Olaf Ronneberger^{1,4}, Kathryn Tunyasuvunakool¹⁴, Russ Bates^{1,4}, Augustin Židek^{1,4}, Anna Potapenko¹⁴, Alex Bridgland¹⁴, Clemens Meyer¹⁴, Simon A. A. Kohl¹⁴, Andrew J. Ballard¹⁴, Andrew Cowie¹⁴, Bernardino Romera-Paredes^{1,4}, Stanislav Nikolov¹⁴, Rishub Jain^{1,4}, Jonas Adler¹, Trevor Back¹, Stig Petersen¹, David Reiman¹, Ellen Clancy¹, Michal Zielinski¹, Martin Steinegger^{2,3}, Michalina Pacholska¹, Tamas Berghammer¹, Sebastian Bodenstein¹, David Silver¹, Oriol Vinyals¹, Andrew W. Senior¹, Koray Kavukcuoglu¹, Pushmeet Kohli¹ & Demis Hassabis^{1,4}

Proteins are essential to life, and understanding their structure can facilitate a mechanistic understanding of their function. Through an enormous experimental effort1-4, the structures of around 100,000 unique proteins have been determined5, but this represents a small fraction of the billions of known protein sequences⁶⁷. Structural coverage is bottlenecked by the months to years of painstaking effort required to determine a single protein structure. Accurate computational approaches are needed to address this gap and to enable large-scale structural bioinformatics. Predicting the three-dimensional structure that a protein will adopt based solely on its amino acid sequence-the structure prediction component of the 'protein folding problem'8-has been an important open research problem for more than 50 years9. Despite recent progress¹⁰⁻¹⁴, existing methods fall far short of atomic accuracy, especially when no homologous structure is available. Here we provide the first computational method that can regularly predict protein structures with atomic accuracy even in cases in which no similar structure is known. We validated an entirely redesigned version of our neural network-based model, AlphaFold, in the challenging 14th Critical Assessment of protein Structure Prediction (CASP14)15, demonstrating accuracy competitive with experimental structures in a majority of cases and greatly outperforming other methods. Underpinning the latest version of AlphaFold is a novel machine learning approach that incorporates physical and biological knowledge about protein structure, leveraging multi-sequence alignments, into the design of the deep learning algorithm.



Google deepmind

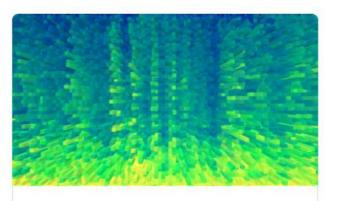
 \rightarrow

https://deepmind.google/technologies/

🗚 🏠 🧧 🤤 🗘 🛱 🖉



Imagen 2 Our most advanced text-to-image technology



SynthID

Robust and scalable tool for watermarking and identifying Al-generated images.

 \rightarrow



AlphaFold

Breakthrough AI system accurately predicts the 3D models of protein structures — and...

 \rightarrow



Phenaki Realistic video generation from open-domain textual descriptions



AlphaZero and MuZero

Powerful, general AI systems that mastered a range of board games and video games —...



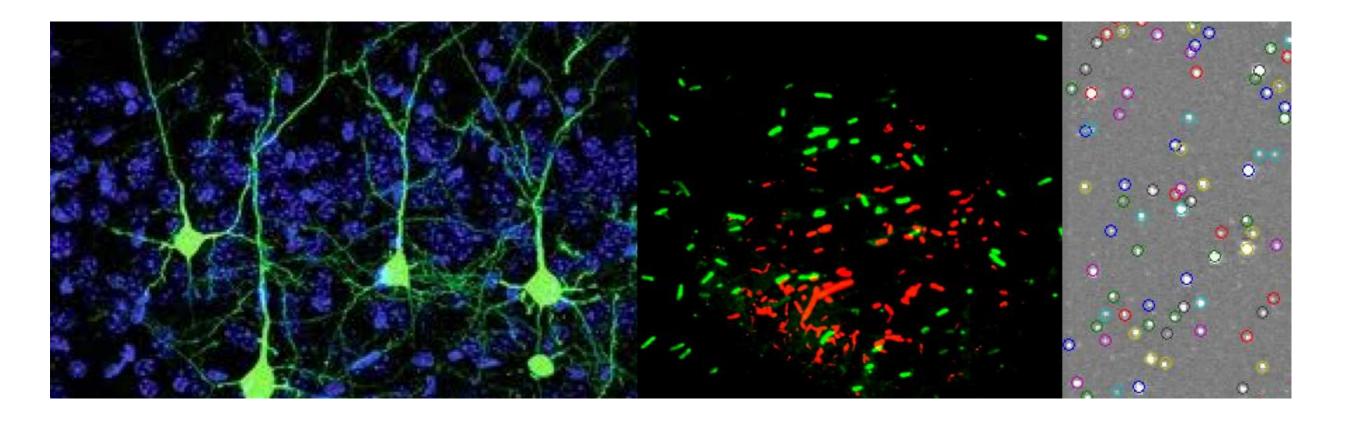
AlphaGo

Novel AI system mastered the ancient game of Go, defeated a Go world champion, and...

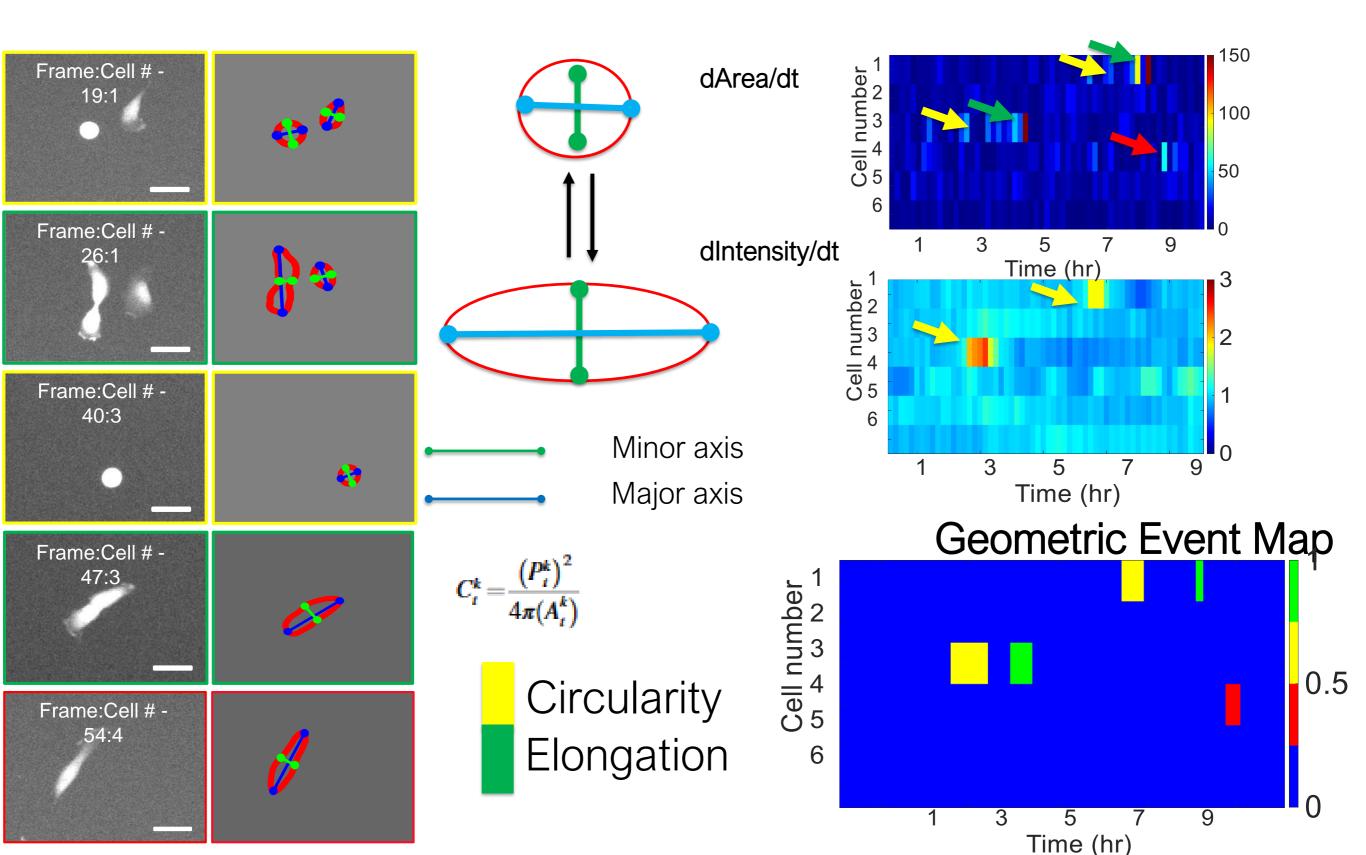
Application of ML

- Image recognition: Face recognition, medical image, self-driving cars.
- Natural language processing: Text analytics, sentiment analysis.
- Recommendation systems: Product recommendations, music recommendations, movie recommendations.
- Banking: Identifying fraudulent transactions, credit score.
- Protein folding
- Drug discovery

How many cell, bacteria or single molecule is present in each picture?

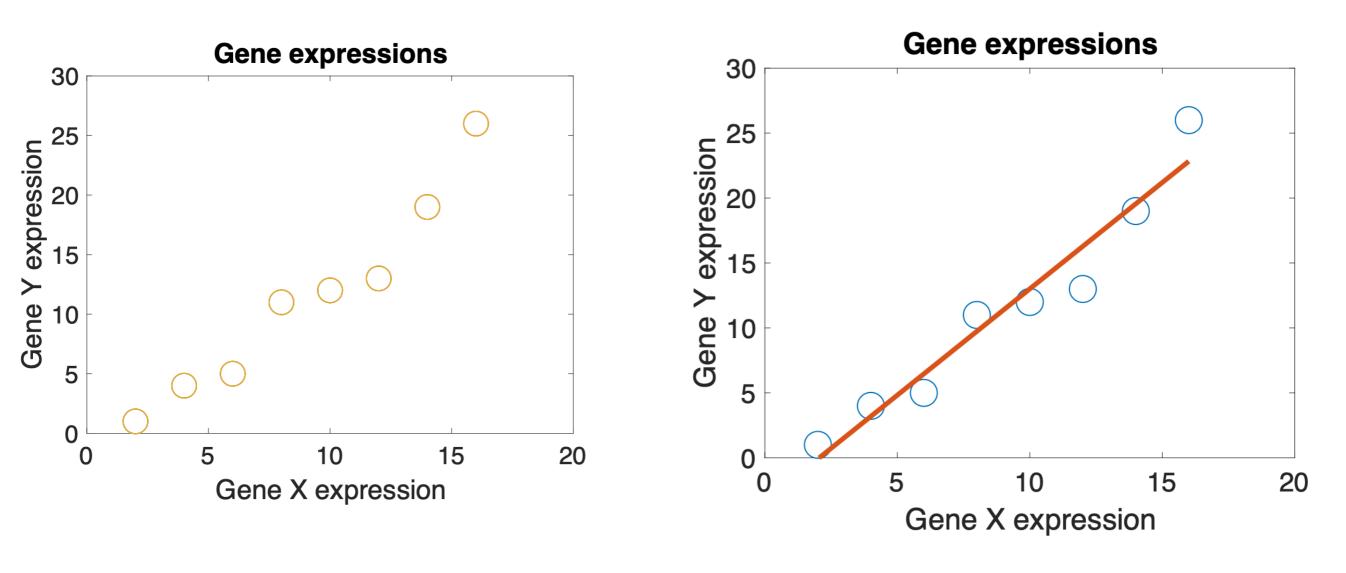


Time course of geometric events: event map generation for identification of cell splitting



Linear Regression: Least Squares Fit

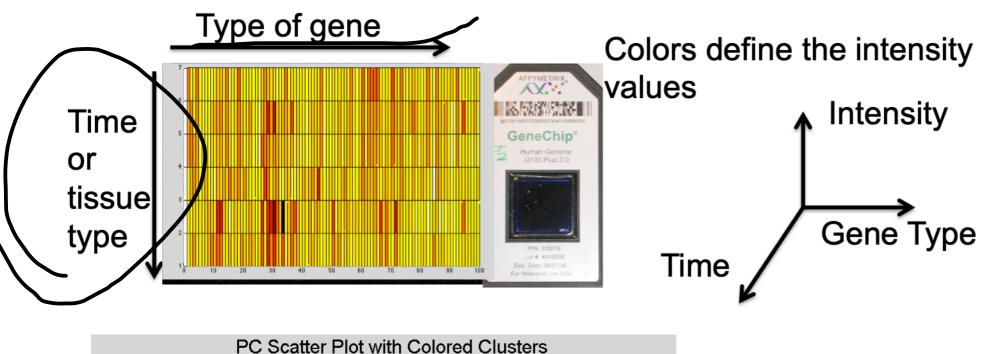
Fitting data y best in a least-square method.

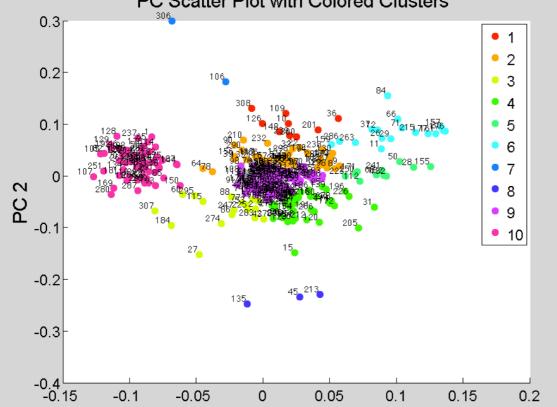


Which of the genes are expressed at the same time in cancer vs normal cells?

What are the genes that are expressed in cancer cells but not normal cells?

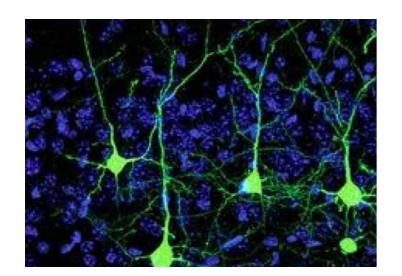
How the expression changes over time?



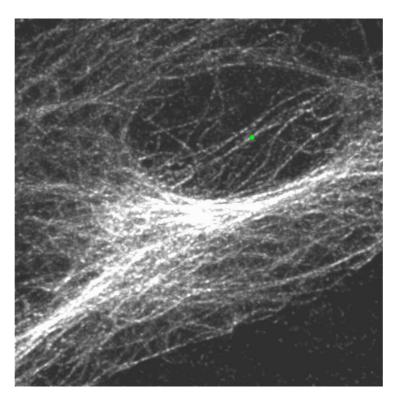


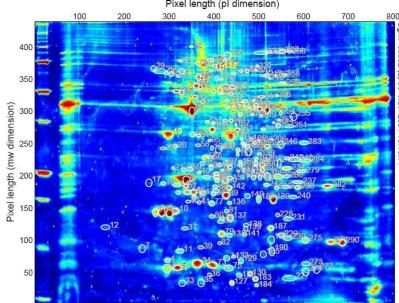


Genetics



Neuroscience





http://bioinformatics.musc.edu/Web2DE

Bioinformatics

Molecular Biology

Computing is essential to solve many of these research problems

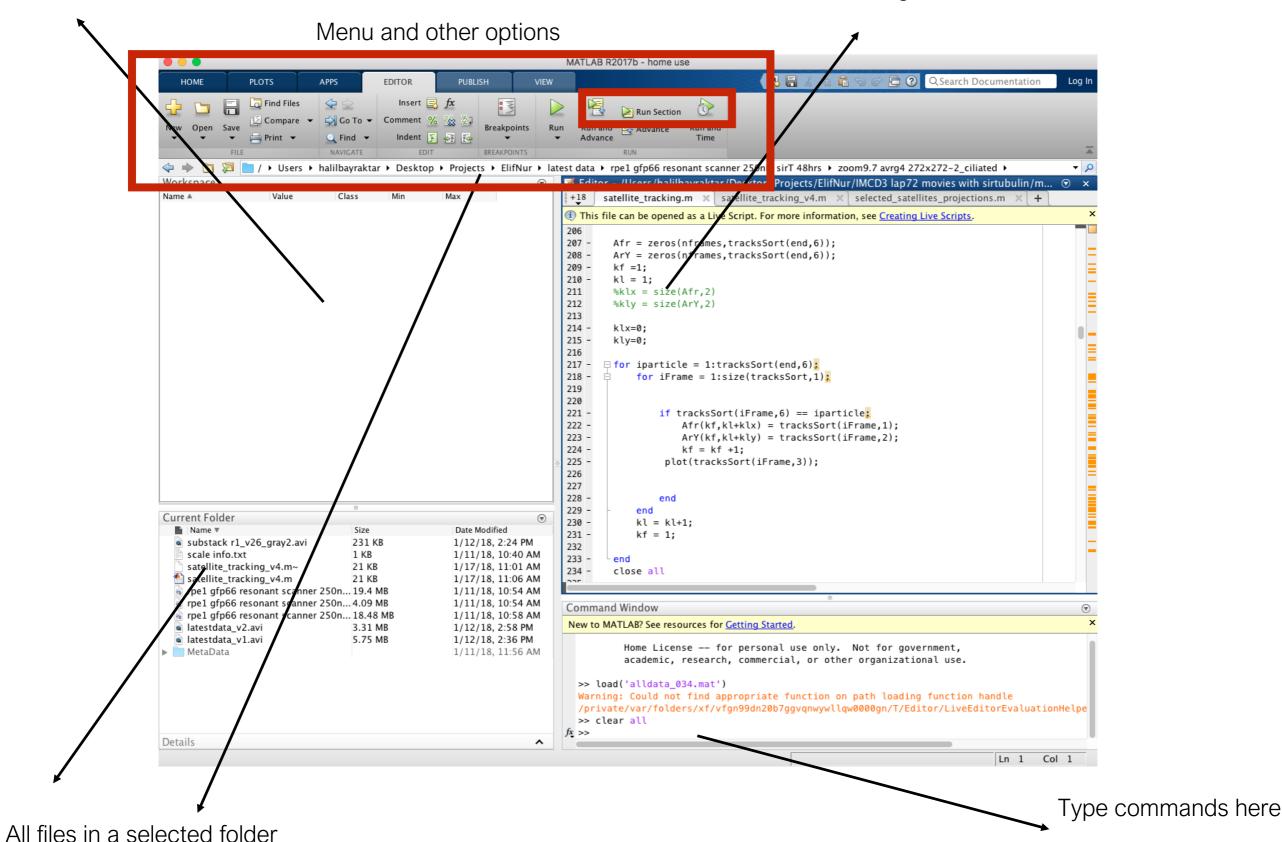


Artificial Intelligence

MATLAB USER INTERFACE

Workspace to show all variables

Coding



Python Layout (open source platform)

	0		-	yder (Python 2.7)									
		^r ≔ (≓)> ■ : ■	8 🗙 🍾 🤇	🕨 : 🗲 🄶 /Us	ers/halilbayra	ktar/Docume			projects and c	ode		<u> </u>	1
G File explorer	00		Editor					00	B	Variab	le explorer		
😋 Previous 💿 Next 🕜 Parent 📫	🗘 🕞 🐵 🛛 GUI.py	⊗ celltracking.p ⊗	game_v1.py 🛛 🖾	imageproces 🕲	loops.py	<mark>⊗ n < ></mark>	Q -	Ł 🖹 🛛	" /				Ф,
Name adder.py birdwatching.txt.sb-afb1d345-SIWVQA brownianmotion 2.py brownianmotion_@.py brownianmotion.py celltracking.py converter.py dataanalysis_v1.py database_v1.py face.png function1.py gui_2.pyx gui_2.pyx gui_3.py gui_3s.py gui_5.py GUI.py imagearray.py imageprocessing_load.py ins.py Introduction To Computing Using Pytho	46 47 os.chdir(' 48 49 retval = 0 50 files=os.l 51 52 53 54 #%% 55 print "Dir 56 57 e=56 58 59 60 files=os.l 61 62 #%% 63 img = Imag 64 imgf = img 65 figure(33) 66 subplot(1, 67 imshow(img 68 69 #%% 70 71 72 from PIL i 73 import num 74 75 def read_t 76 77 img = 78 images	<pre>/Users/halilbayraktar/ s.getcwd() istdir(retval) rectory changed success istdir(cwdb) e.open('B959K8NG_F0000 _as_float(img) 2,1) f) mport Image py as np iff(path, n_images): Image.open(path)</pre>	'Desktop/Nurhan	n data/nurhandata			-	am ▲ Ty	rpe Size		Value		
Outline	00		IPython cons	ole				00			Help		
Celltracking.py celltracking.py read_tiff read_tiffone read_tiff read_tiff	Type "copyright" IPython 5.3.0 ? -> In %quickref -> Qu help -> Py	Anaconda custom (x86_64 ', "credits" or "licens - An enhanced Interacts troduction and overview ick reference. thon's own help system tails about 'object', o	se" for more i ive Python. w of IPython's use 'object??'	nformation.			*	Source	Us Hei by eith Hei aut par car Pre	Object age re you can get hel pressing Cmd+I i her on the Editor of tomatically after w renthesis next to a n activate this beh eferences > Help. New to Spyder? Rea	n front of it, or the Console. wn riting a left an object. You avior in ad our <u>tutorial</u>		
-		IPython	console Pyth	on console						History log	Help		
				Permissions: R	W End-o	f-lines: LF	E	ncoding:	UTF-8	Line: 52	Column: 1	Memory: 72 9	6

Matlab help

•	Help
🜩 🍓 🛧 🛠	
Other uses of mean	×
Documentation	Search Help Q
CONTENTS	Close
Documentation Home	mean
MATLAB	Average or mean value of array collapse all in page
Coata Import and Analysis Coscriptive Statistics	Syntax
< MATLAB < Functions	<pre>M = mean(A) M = mean(A,dim) M = mean(,outtype)</pre>
mean ON THIS PAGE	<pre>M = mean(, nanflag)</pre>
Syntax Description	Description
Examples	M = mean(A) returns the mean of the elements of A along the first array dimension whose size does not equal 1.
Input Arguments More About	 If A is a vector, then mean(A) returns the mean of the elements.
Extended Capabilities	 If A is a matrix, then mean(A) returns a row vector containing the mean of each column.
See Also	 If A is a multidimensional array, then mean(A) operates along the first array dimension whose size does not equal 1, treating the elements as vectors. This dimension becomes 1 while the sizes of all other dimensions remain the same.
	M = mean(A,dim) returns the mean along dimension dim. For example, if A is a matrix, then mean(A,2) is a column vector containing the mean of each row.
	M = mean(,outtype) returns the mean with a specified data type, using any of the input arguments in the previous syntaxes. outtype can be 'default', 'double', or 'native'.

Why MATLAB?

A single platform to load, sort, analyze, plot and save your data.

Beginner level syntax structure

Easy coding

Highly preferred in academia

Many open resources is available for help

Many functions are included

.mat and m. files

.m : Files that contain the computer code. There are two kinds of .m files. Scripts and function. Script files do not need any input value to execute, however the function requires one or more arguments to return the output

.mat files: files that contains the data and variable names

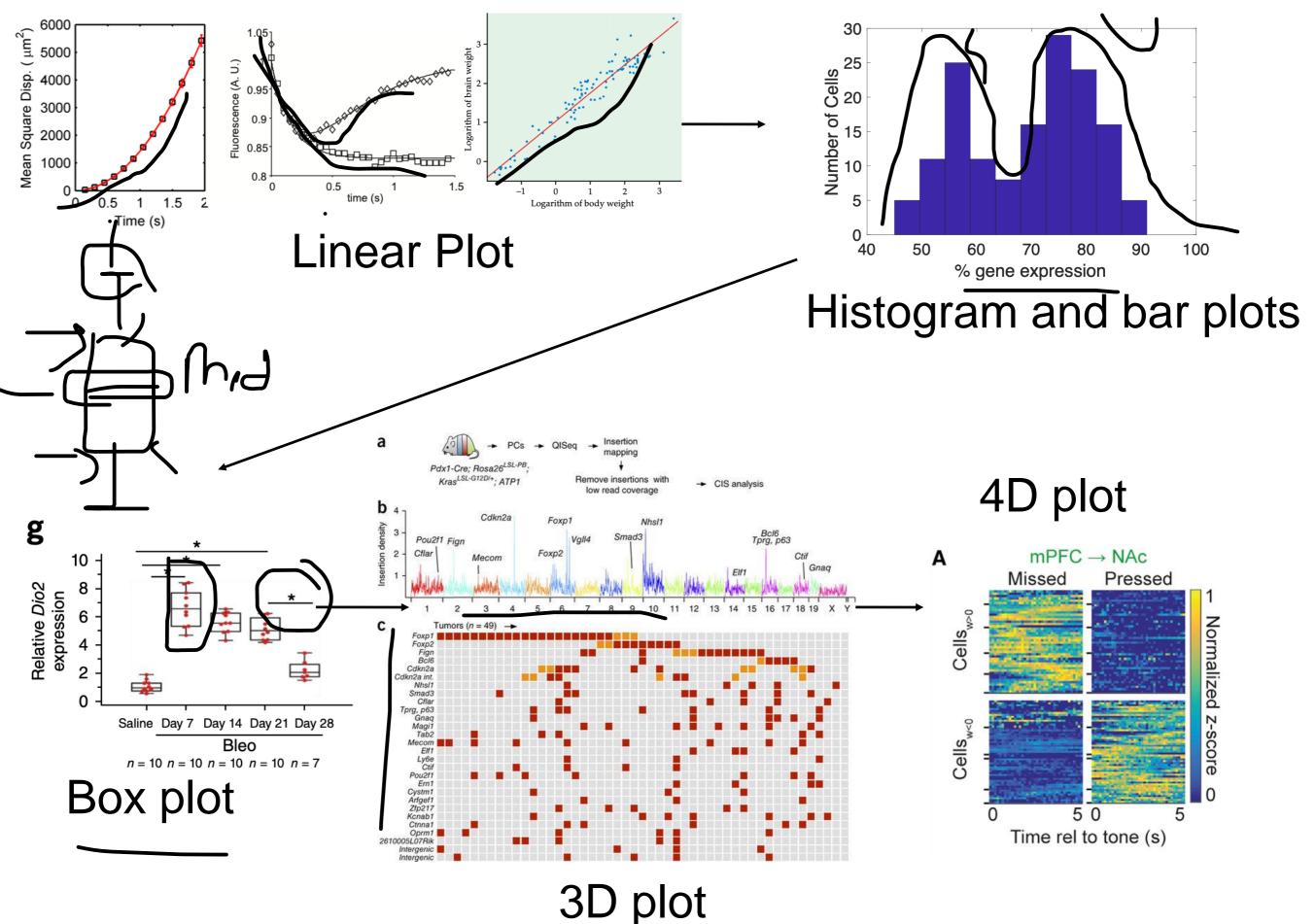
Writing code in computer vs playing notes in piano





You need practice, practice and practice to advance in writing codes.

Description and Visualization of Statistical Data



Introduction to Matlab

MATLAB=Matrix laboratory developed by MathWorks Inc

High level language for computing, data visualization, modeling and programming

Matrices (1D, 2D etc) are the main components

NAME OF PUNCTUATION MARKS USED in MATLAB

Square brackets []

hyphen -

curly brackets {}

Brackets, parentheses ()

Semi colon;

Colon :

Comma,

Dot.

Apostrophe ' or single quote

Double quote ""

Single quote "

Backslash $\$

Slash /

Create a variable

Create a number variable

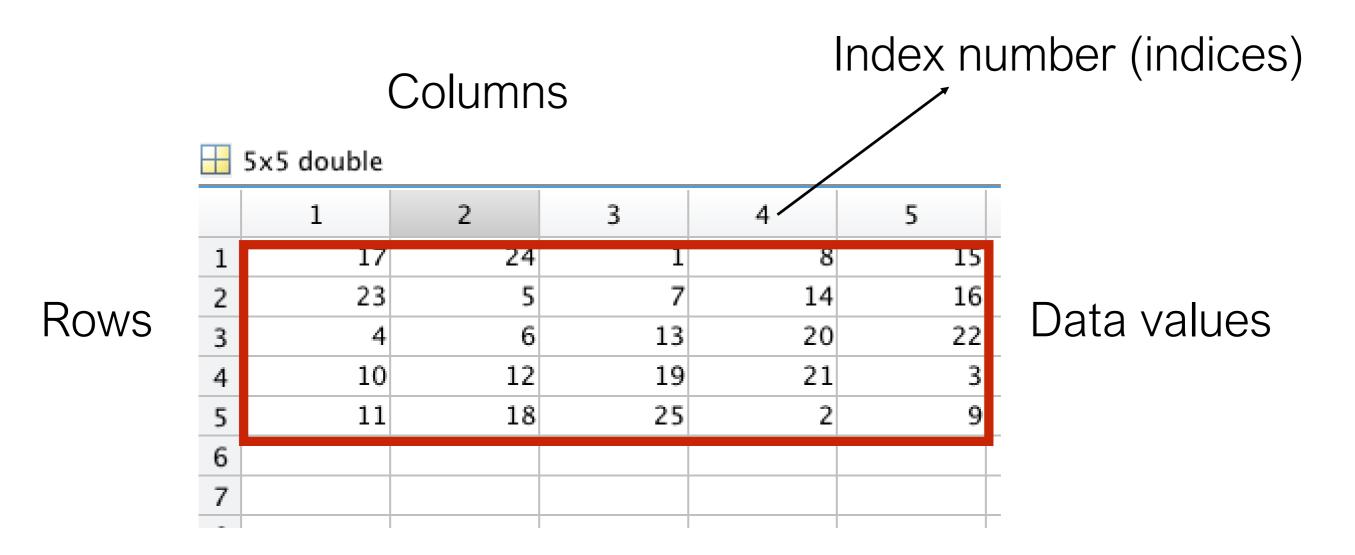
$$x=5;$$
 2
y=10;
z= x+y; 5
z = 2
y = 2
y = 2
z = 2

Create a char variable

7

x = 'Matlab'

You should use



5 by 5 matrix, 5x5

Create a vector

Generating a row and column vector

rowvec=[5,6,7,8,12] rowvec =

colvec=[3;4;7;9;14] 5 6 7 8 12

colvec =

vector slicing

a=0:2:20 b=a(1,1:5) c=a(1,5:end)

a = 0 2 4 6 8 10 12 14 16 18 20 b = 0 2 4 6 8 c = 8 10 12 14 16 18 20 Create a vector

x = 0,2,4,...,100

size(x)

Ans row and column 1 51 Indexing into a matrix is a means of selecting a subset of elements from

>> fcel

fcel =

Finding an element in row or column vector % First entry is row index number % Second entry is column index number

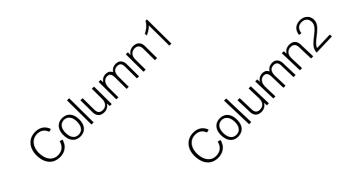
frel= rowvec(1,5) frel = fcel=colvec(3,1) 12

rowvec=[5,6,7,8,12]

colvec=[3;4;7;9;14]

Operators

Sum Division Multiplication Select a region from matrix



row 1 row 2

A		1	2	3	4	5
	1	17	24	1	8	15
2	2	23	5	7	14	16
	3	4	6	13	20	22
	4	10	12	19	21	3
	5	11	18	25	2	9
	6					
	7					

Row Column b=a([3,4],[1,2,3]) b =

> 4 6 13 10 12 19

Finding and selecting elements in a matrix z=magic(4) z([1,2],[3,4]) z(1,1) z(2)

Ζ =	ans =	ans =	ans =
$\begin{array}{rrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrr$	3 13 10 8	16	5

Some Common Scalar Mathematical Operations

	Operation	M-file	
Subtraction	Х-у	Х-у	
Addition	x + y xx+y+y	x + y	
Multiplication	ху	x*y	
Division	x/y	x/y	
Power	X ^y	x^Y	
Exp	e×	exp(x)	
Log	log10(x)	log10(x)	
Nat Log	ln(x)	log(x)	
	cos(X) 0	Cos(x)	
	Sin(x)	Sin(x)	
	X ^{1/2}	sqrt(x)	