

Computational Biology (BIOSC 1540)

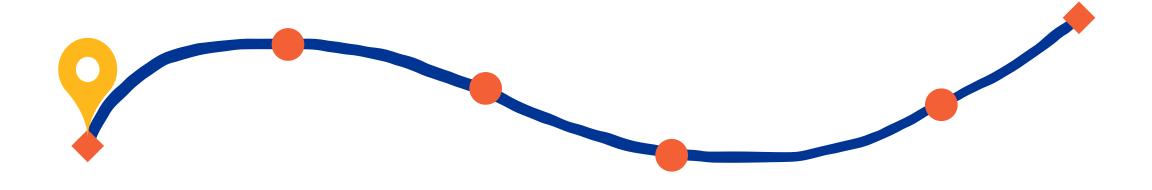
Lecture 01:

Introduction to computational biology

Aug 27, 2024



After today, you should be able to



1. Describe the course structure, expectations, and available resources for success.

- 2. Define computational biology and explain its interdisciplinary nature.
- 3. Identify key applications and recent advancements.
- 4. Understand the balance between applications and development.
- 5. Identify potential career paths and educational opportunities.

Meet your teaching team

Alex Maldonado, PhD he/him/his





B.S.E in Chemical Engineering, 2018 Western Michigan University



Ph.D. in Chemical Engineering, 2023 University of Pittsburgh

Office hours: Tuesday & Thursday from 11:30 am - 12:30 pm in Clapp Hall

Email: alex.maldonado@pitt.edu

Acceptable ways to address me: **Alex** (preferred) Dr. Maldonado Dr. Alex Dr. M

Reya Kundu (she/her/hers)

Major: Computational biology

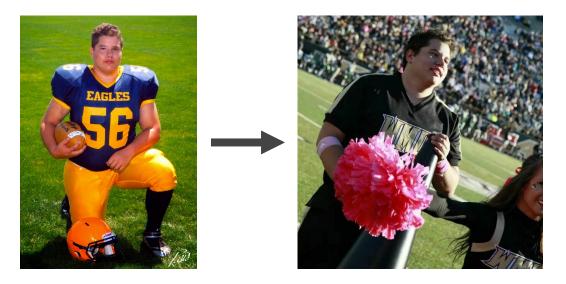


Justine Denby (she/her/hers)

Major: Computational biology

Alex's fun facts

Every male in my (maternal) family played football —I rebelled





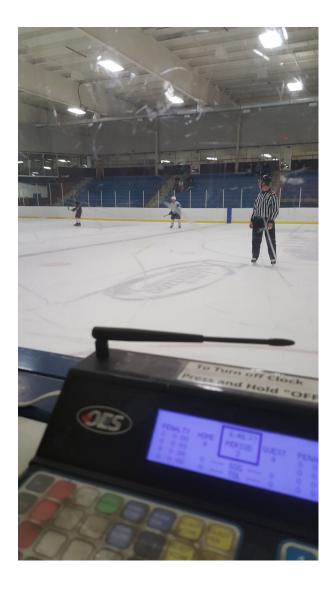
Alex's fun facts

Part-time jobs

- Construction
- UPS package handler
- Kent County Traffic safety
- Jimmy John's delivery driver
- Wings West ice events

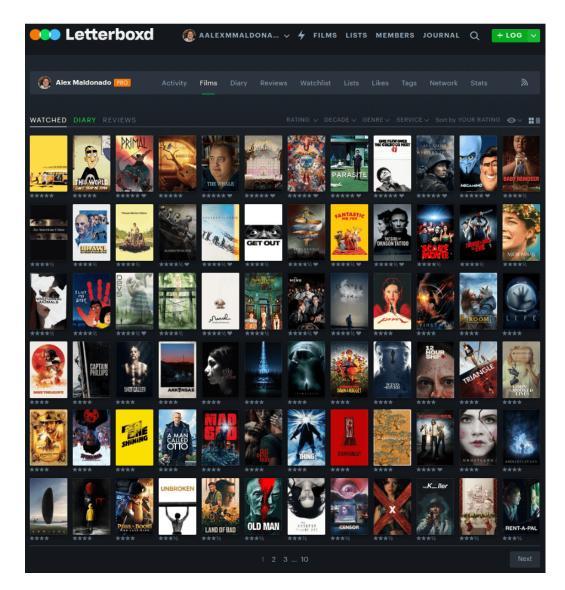




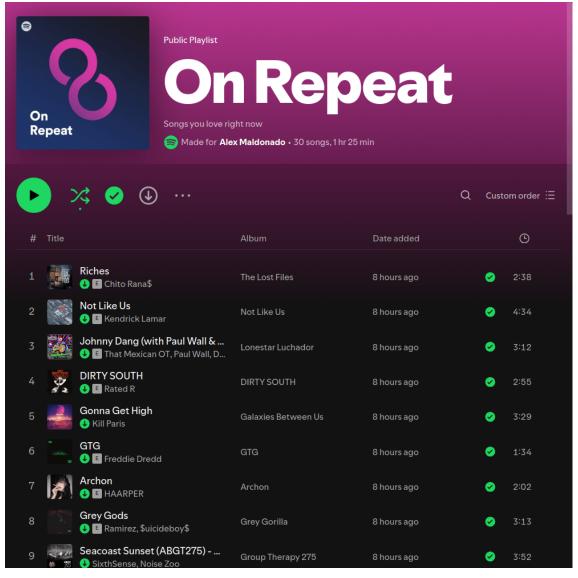


Get to know my ...

Movie taste



Music taste



Single source of truth

All course materials will be posted on this website: pitt-biosc1540-2024f.oasci.org/

Why? I am extra

There are few comprehensive resources for this rapidly changing field

Things that contain student information will be only on Canvas to be FERPA compliant

Assignments will be submitted on Gradescope

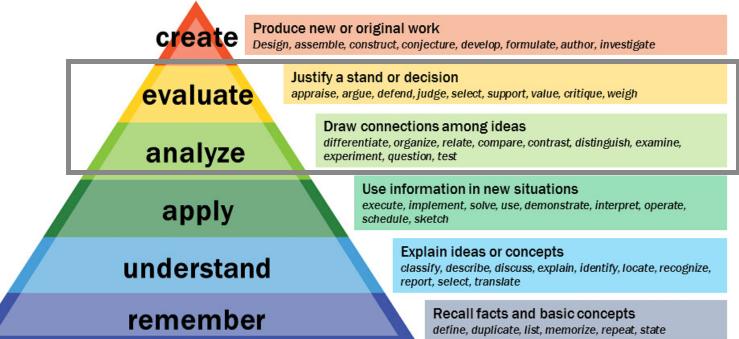
https://pitt-biosc1540-2024f.oasci.org/

My course philosophy

Critical thinking is paramount and happens outside your comfort zone

How does this influence my teaching?

Bloom's Taxonomy



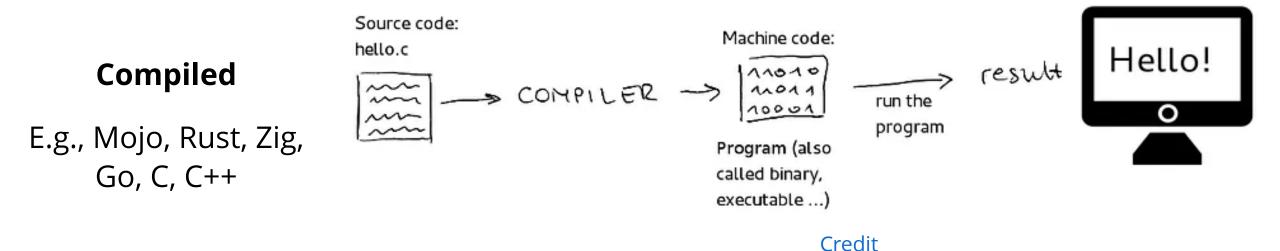
Few points

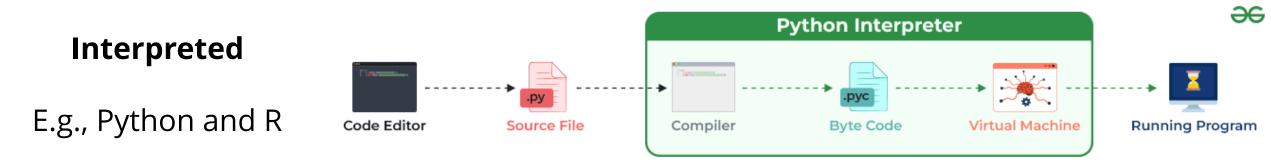
This is where my focus is (similar to computer science and engineering)

Many points



There are two types of programming languages*



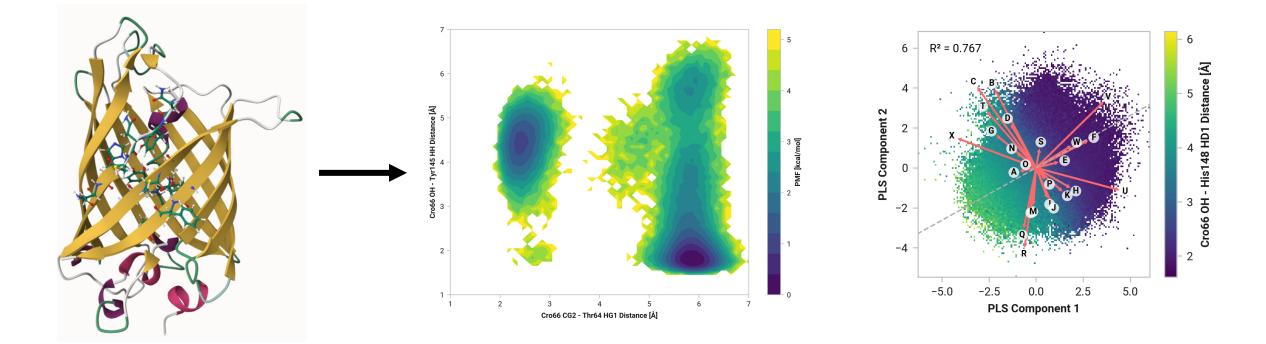


Credit

There are some exceptions: Java, Julia

Python is absolutely necessary for a career in computational biology

Programming is how you obtain, manage, and analyze data



Data

Results and insights

No coding will be necessary to successfully complete this course

Previous semesters used R or Python



We will emphasize learning the foundational principles instead of coding

There will be optional coding opportunities

Semester overview

Bioinformatics

Modules

Computational Structural Biology

Special interests/ Python?

Subfields we are not able to cover in detail

1. Genomics

3. Molecular simulations

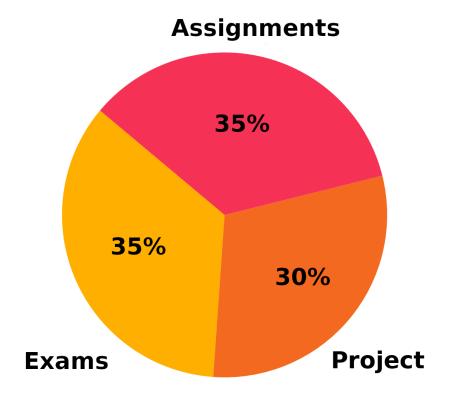
Where do we get our insight from?

Scientific python

2. Transcriptomics

4. Computer-aided drug design

Assessments



We will have ...

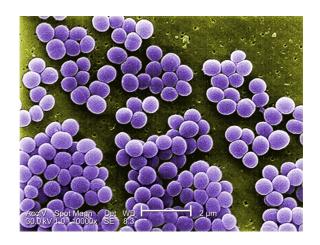
- Eight homework assignments
- One hands-on project
- Two exams

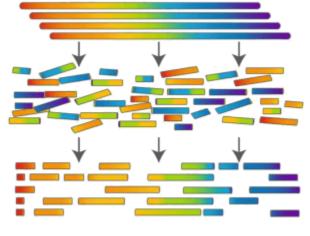
An optional cumulative final will be provided to replace lowest exam grade

Attendance is not mandatory, but encouraged

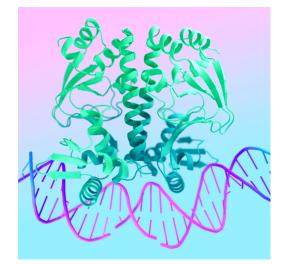
Project: Computer-Aided Drug Design for a Novel Pathogen

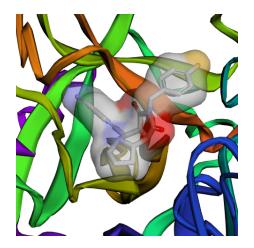
We will work through a complete, web-based workflow mirroring the steps researchers might take when confronted with a new pathogenic threat





GTTCCGATTAGGAAACCTATCTGTAACTGTTTCATTCAGTAAAAGGAGGAAATAT





DNA sequencing of *Staphylococcus aureus*

Assembled and annotated genome

Protein structure prediction

Protein-ligand docking

Other policies

Please read the rest of the syllabus on your own

I can also answer any questions now

Syllabus

Semester: Fall 2024 Meeting time: Tuesdays and Thursdays from 4:00 - 5:15 pm. Location: 1510 Posvar Instructor: Alex Maldonado, PhD (he/him/his) Email: alex.maldonado@pitt.edu Office hours:

| Day | Time | Location |
|----------|---------------------|----------------|
| Tuesday | 11:30 am - 12:30 pm | 102 Clapp Hall |
| Thursday | 11:30 am - 12:30 pm | 315 Clapp Hall |

Catalog description

This course gives students a broad understanding of how computational approaches can solve problems in biology. We will also explore the biological and computational underpinnings of the methods.

🕗 Note

The catalog course description mandates what material this course has to cover. How the material is covered is at the discretion of the instructor.

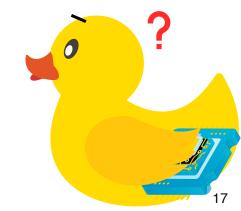


1. Describe the course structure, expectations, and available resources for success.

2. Define computational biology and explain its interdisciplinary nature.

- 3. Identify key applications and recent advancements.
- 4. Understand the balance between applications and development.
- 5. Identify potential career paths and educational opportunities.

What is computational biology?



What is computational biology?

My definition . . .

Any application of computational methods to obtain insight into biological phenomena.

My main categories . . .

Bioinformatics

Computational structural biology

<u>Bioinformatics</u> deals with untangling big data for biological insight

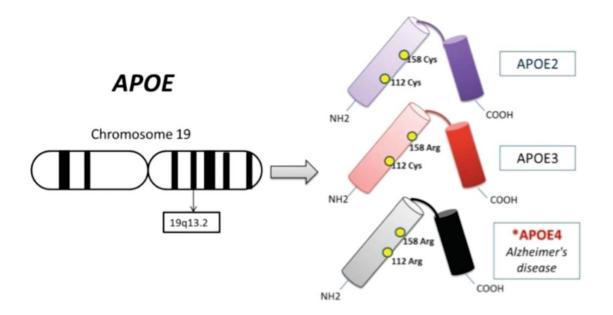
Data

Genetic sequences of healthy and Alzheimer patients

| | * | . : | | . * | : : : | |
|---|-----|-----|----|-----|-------|--|
| | | | | | | |
| | | | | | | |
| | | | | | | |
| | | | | | | |
| | | | | | | |
| | | | | | | |
| | | | | | | |
| MSGAG-SKRKKLFIEKATKLFTT YDKMIVAEADFVGSSOLOKIRKSIRGI-GAVLMGKKTMIRKVIRDLADSK-PELD MSGAG-SKRNVFIEKATKLFTT YDKMIVAEADFVGSSOLOKIRKSIRGI-GAVLMGKKTMIRKVIRDLADSK-PELD MSGAG-SKRNVFIEKATKLFTT YDKMIVAEADFVGSSOLOKIRKSIRGI-GAVLMGKKTMIRKVIRDLADSK-PELD | | | | | | |
| MS G AG - SKRKN VF IEKATKLFTT YDKMI VAE ADF VGS SOLOK IRKS IRG I - GAV LMGKKIMIRK VIRDLADSK PELD MKLSKQQK KQM YIEKLSSLIQQYSKILI VHVDNVGS NGMASVRKSLRGK - AT ILMGKNTRIRTALKKNLGAV - PQTE MI G LAVTTTKKIAKWK VDE VAE LE KLKTHKT III AN IEGFPADKLHE IRKKLRGK - AT IKVTKNTLFG IAAKNAG | | | | | | |
| MAKLSKQQKKQMYTEKLSSLIQQYSKILIVHVDNVGSNOMASVRKSLRGK-ATILMGKNTRIRTALKKNLQAV-PQIE MIGLAVTTTKKIAKWKVDEVALTTKLKTHKTIIANIEGFPADKLHEIRKKLRGK-ADIKVTKNNLFNIALKNAGYDUK MRIMAVITQERKIAKWKIEEVKELEQKLREYHTIIANIEGFPADKLHEIRKKLRGK-ADIKVTKNNLFNIALKNAGYDUK MKRLALALKQRKVASWKLEEVKELTELIKNSNTILIGNEEGFPADKLHEIRKKLRGK-ADIKVTKNTLFKIAKNAGIDJE MSVVSLVGQWYRKEKPIPEWKTLMRELEELFSKHRVULFADIRCTPFVVQRVRKKLWKK-YPMVAKKRIILRAMKAAGLELDDN MAERHHTEHIPQWKKDEINIKELIQSHKVFGWGIEGILATKNKKIRVCVQVAVLKVSKNTLFKIAAKAAGLELDDN | | | | | | |
| MICLAVTTKKIAKWK VDE VAELTEK LKTHKTIIIAN IEGFPADKLHE IEKK LEGK-AD IKVTKNNLFN IALKNAGYD TK MKIMAVITQE RKIAKWK VDE VAELE QK LRE YHTIIIAN IEGFPADKLHD IEKK MEGM-AEIK VTKNTLFG IAAKNAG | | | | | | |
| MRIMAVITQE RKIAKWKIE EVKELE OKLREYHTIIIANIEGFPADKLHDIRKKMRGM-AEIKVTKNTLFGIAAKNAGDUVS MKRLALLKQRKVASWKLEEVKELTEIKNSNTILIGNEEGFPADKLHEIRKKLRGK-ATIKVTKNTLFGIAAKNAGDUVS MSVVSIVGQMYKREKPIPEWKTLMLRELESLFSKHRVVLFADLTGTPTFVVRVRKKLWK-YPMVAKKRIIFGKAAKAAGDUTE MSVVSIVGQMYKREKPIPEWKTLMLRELESLFSKHRVVLFADLTGTPTFVVRVRVKKLWK-YPMVAKKRIIFGKAAKAAGDUTE MSVVSIVGQWYKREKPIFEWKTLMLRELESLFSKHRVVLFADLTGTPTFVVRVRVKKLWK-YPMVAKKRIIFGKAAKAGLDUT MASURGS | | | | | | |
| MKRLALALKQRKVASWKLEEVKELT BLIKNSNTILIGNLEGFPADKLHEIRKKLRGK-ATIKVTKNTLFKIAAKNAGIDIE MSVVSLVGQMYKREKPIDEWKTLMLRELEELFSKHRVULFADLTGTPFVVQRVRKLWKK-VPMVAKKRILFKIAMKAAGLELDDN -MMLAIGKRRYVTRQVPARKVKIVSATELLQKYPYVFLFDLHGLSSRILHEYRYKLWK-VPMVAKKRILFKIAMKAAGLELDDN -MAERNHTEHIPQWKKDEIENIKELIQSHKVFGMVGIEGILATKUKIRRDLKD-AVLKVSRNTLTERALNQLGETTP MAERNHTEHIPQWKKDEIENIKELIQSHKVFGMVGIEGILATKUKIRRDLKD-AVLKVSRNTLTERALNQLGETTP MAERNHTEHIPQWKKDEIENIKELIQSHKVFGMVGIEGILATKUKIRRDLKD-AVLKVSRNTLTERALNQLGETTP MAERNHTEHIPQWKKDEIENIKELIQSHKVFGMVGIEGILATKUKIRRDLKD-AVLKVSRNTLTERALNQLGETTP MAERNHTEHIPQWKKDEIENIKELIQSHKVFGMVGIEGILATKUKIRRDLKD-AVLKVSRNTLTERALNQLG | | | | | | |
| MSVVSLVGQMYKREKPIPEWKTLMLRELE SLFSKHRVVLFADLTGTPTVVGRVRKKLWKK-YPMMVAKKRIIL RAMKAAGLE LDDN -MMLAIGKRRVVTRQVPARKVKIVSEATELLQKYPYPLFDLHGLSRILHEYRYKLRKY-GVIKIIKPTLFKIAFTKVYGG IPAE MAEERHHTEH IPQWKKDE IEN IKELIQSHKVFGWGIEGILATKMK IRRDLKDV-AVLKVSRNTLTE RALNQLG ET IP MAEERHHTEH IPQWKKDE IEN IKELIQSHKVFGWGIEGILATKMK IRRDLKDV-AVLKVSRNTLTE RALNQLG ET IP MAEERHHTEH IPQWKKDE IEN IKELIQSHKVFGWGIEGILATKIK IRRDLKDV-AVLKVSRNTLTE RALNQLG ET IP MAEERHHTEH IPQWKKDE IEN IKELIQSHKVFGWGIEGILATKIKIKIRRDLKDV-AVLKVSRNTLTE RALNQLG ET IP | | | | | | |
| -MILA IGKRYVRT RQYPARK VKI VSEAT BLLQKYPYVFLFDLHGLSSRILHEYRYRLRRY-GVIKIIKPTLFKIAFTKVYGGIPAB MAEERHHTEHIPQWKKDEIENIKELIQSHKVFGWGGIEGILATKMQKIRRDLKDV-AVLKVSRNTLTERALNQLGETIP MAERHHTEHIPQWKKDEIENIKELIQSHKVFGWGIEGILATKMQKIRRDLKDV-AVLKVSRNTLTERALNQLGESIP MAERHHTEHIPQWKKDEIENIKELIQSHKVFGWGIEGILATKMQKIRRDLKDV-AVLKVSRNTLTERALNQLGESIP MAERCHHTEHIPQWKKDEIENIKELIQSHKVFGWGIEGILATKMQKIRRDLKDV-AVLKVSRNTLTERALNQLGESIP MAERCHHTEHIPQWKKDEIENIKELIQSHKVFGWGIEGILATKMQKIRRDLKDV-AVLKVSRNTLIERALNQLGESIP MAERCHHTEHIPQWKKDEIENIKELKMISSKPVAIVSFRNVPAGOMOKIRRDLRGLAVKSKRTLIERALNQLGGDKL MAVKAKGOPPSGYPKVARWERVELKELKMDYVALVSFRNVPAGOMOKIRREFRGK-AEIKVVKNTLLERALDALG | | | | | | |
| MAEERHHTEHIPQWKKDEIENIKELIQSHKVFGWVCIEGILATKMOKIRRDLKDV-AVLKVSRNTLTERALNQLGETIP MAEERHHTEHIPQWKKDEIENIKELIQSHKVFGWVRIEGILATKIOKIRRDLKDV-AVLKVSRNTLTERALNQLGETIP MAEKRGSDPEYKVRAVEELKRMISSKPVVAIVSFRNVPAGOMKIRREFRGK-AEIKVVKNTLTERALNQLGESIP MAVKAKGQPESGYEPKVAEWKREVKELKEIMDEYENVGLVDLEGIPAPDIOEIRAKLREDDIIRMSKNTLTERALNQLGCOVL MAVKAKGQPESGYEPKVAEWKREVKELKEIMDEYENVGLVDLEGIPAPDIOEIRAKLREDDIIRMSKNTLMRIALEKLDERPELE MAHVAEWKKEVQELHDLIKGYEVVGIANLADIPAROLOKIRDKIRGILROSALIKMSKNTLIERAIKEVAEETGNPEFA MITAESEHKIAPWKIEEVNLKELLKNQUVALVDMMEVPAROLOEIRDKIR-OMUKKMSNTLIERAIKEVAEETGNPEFA MITAESEHKIAPWKIEEVNLKELLKNQUVALVDMMEVPAROLOEIRDKIR-OMUKKMSNTLIIRAALEVAEETGNPEFA MITAESEHKIAPWKIEEVKILKSINVIALIDMEVPAROLOEIRDKIR-OKUKKMSNTLIIRAKKAGELAEKAGELENVD MITAESEHKIAPWKIEEVKILKSINVIALUDMEVPAROLOEIRDKIR-OKUKKMSNTLIIRAALEVAEETGNPEFA | | | | | | |
| MAEERHHTEH LPQWKKDE LEN IKELIQSHKYFGMYRIEGILATKICKIRRD LKDY-AVLKYSRNTLTERALNQLGESIP MAEYRGSPPEYKYRRYEEIKRMISSKPVAIVSFRNYPAGOMCKIRREFRGK-AEIKYVKNTLLERALDALGGDVL MAYKAKGQPPSGYEPKVAEWKREFVKELKELMDEYENVGLVDLEGIPAPOLGEIRAKLREFDGK-AEIKYVKNTLLERALDALGGDVL MAHYAEWKKEVQELHDLIKGYEVVGLVDLEGIPAPOLGEIRAKLREFDGK-AEIKYVKNTLLERALDALGGDVL MITAESEHKIAPWKIEEVNKLKELLKSANVIALIDMEVPAPOLGEIRDKIR-GTMILKMSRNTLIERATKEVAEETGNPEFA MITAESEHKIAPWKIEEVNKLKELLKSANVIALIDMEVPAPOLGEIRDKIR-GTMILKMSRNTLIERATKEVAEETGNPEFA MITAKSEHKIAPWKIEEVNKLKELLKSANVIALIDMEVPAPOLGEIRDKIR-GUMTKMSRNTLIERATKEVAEETGNPEFA | | | | | | |
| MAAYRGSPPE TKYRAYEE IKRMISSKP YVAIVSFRNVPAGOMOKIRREFRGK-AEIKYVKNTLLE RALDALGGDYL MAYKAKGOPPSGYEPKVAEWKRREVKELKELMDE YENVGLVDLEGIPAPOLOEIRAKLRENDTIIRMSKNTLMRIALEEKLDER-PELE MAHVAEWKKKEVQELHDLIKGYEVVGIANLADIPAROLOKMRQTIRDS-ALIRMSKKTLISLALEKAGREL-ENVD MITAESEHKIAPWKIEEVNKLKELLKNGQIVALVDMMEVPAROLOEIROKIR-DYNLIKMSKNTLIRAIALEEKLDER-PELE MITAESEHKIAPWKIEEVNKLKELLKNGQIVALVDMMEVPAROLOEIROKIR-DYNLIKMSKNTLIRAIKAAVEEVAEETGNPEFA MITAESEHKIAPWKIEEVNKLKELLKSANVIALIDMMEVPAROLOEIROKIR-DYNLIKMSKNTLIRAAKEAGREL-ENVD MITAESEHKIAPWKIEEVNLKELLKSANVIALIDMMEVPAROLOEIROKIR-DYNLIKMSKNTLIRAAKEAGELCNPEFA | | | | | | |
| MAVKAKGQPPSGYEPKVAEWKRREVKELKELMDEYENVGLVDLEGIPAPOLOEIRAKLRERDIIIRMSRMTLMRIALEEKLDERPELE MAHVAEWKKKEVQELHDIIKGYEVVGIANLADPARDLQKMRQTLRDS-ALIRMSKKILISLALEKAGREL-ENVD MITAESEHKIAPWKIEEVNKLKELLKNGQIVALVDMEVPARQLQEIRDKIR-GTMTLKMSRMTLIERAIKEVAEETGNPEFA MITAESEHKIAPWKIEEVNALKELLKNGVIVALVDMEVPARQLQEIRDKIR-GTMTLKMSRMTLIERAIKEVAEETGNPEFA MITAESEHKIAPWKIEEVNALKELLKNGVIVALVDMMEVPARQLQEIRDKIR-GTMTLKMSRMTLIERAIKEVAEETGNPEFA MITAESEHKIAPWKIEEVNALKELLKNGVIVALVDMMOVPAPQLQEIRDKIR-DVWLLKMSRMTLIERAIKEVAEETGNPEFA | | | | | | |
| MAHVAEWKKKEVQELHDLIKGYEVVGIANLADIPARQLOKMRQTIRDS-ALIRMSKKILISLALEKAGRELENVD MITAESEHKIAPWKIEEVNKLKELLKNGQIVALVDMMEVPARQLOEIRDKIR-GTMTLKMSRNTLIERATKEVAEETGNPEFA MITAESEHKIAPWKIEEVNKLKELLKNGQIVALVDMMEVPARQLOEIRDKIR-DQMTLKMSRNTLIERATKEVAEETGNPEFA MITAKSEHKIAPWKIEEVNLKELLKSAVIALIDMMEVPARQLOEIRDKIR-DKWTLKMSRNTLIIRATKEVAEETGNPEFA METKVKAHVAPWKIEEVKTLKGIIKSKPVALVDMMDVPAPQLOEIRDKIR-DKVKIRMSRNTLIIRATKEVAEETGNPEFA | | | | | | |
| MITAESEHKIAPWKIEEVNKLKSLLKNGQIVALVDMMEVPAROLOEIRDKIR-GTMTLKMSRNTLIERAIKEVAEETGNPEFA MIDAKSEHKIAPWKIEEVNALKELLKSANVIALIDMMEVPAROLOEIRDKIR-DOMIKMSRNTLIERAIKEVAEETGNPEFA METKVKAHVAPWKIEEVNTLKGLIKSKPVAIVDMDVPAPOLOEIRDKIR-DKVKLRMSRNTLIKRAVEEVAEETGNPEFA | | | | | | |
| MIDAKSEHKIAPWKIEEVNALKELLKSANVIALIDMMEVPAVOLOEIRDKIR-DQMTLKMSRNTLIKRAVEEVAEETGNPEFA METKVKAHVAPWKIEEVKTLKGLIKSKPVAIVDMMDVPAPOLOEIRDKIR-DKVKLRMSRNTLIKRAVEEVAEETGNPEFA | | | | | | |
| METKVKAHVAPWKIE EVKTLKGLIKSKPUVAIVDMMDVPAPOLOEIRDKIR-DKVKLRMSRHTLIIRALKEAAEELNNPKLA | | | | | | |
| MAHVAEWKKKEVEELANLIKSYPVIALVDVSSMPAYPLSQMRRLIRENGGLLRVSRHTLIELAIKKAAQELGKPELE | | | | | | |
| MAHVAEWKKKEVEELAKLIKS YP VIALVDVSSMPAYPLSOMRRLIRENGGLLRVSRNTLIELAIKKAAKELGKPELE | | | | | | |
| MAHVAEWKKKEVEELANLIKSYP VALVDVSSMPAYPLSOMRRLIRENNGLLRVSRHTLIELAIKKVAOELGKPELE MAHVAEWKKKEVEELANIIKSYP VALVDVACVPAYPLSKMRDKLR-GKALLRVSRHTLIELAIKKAAOELGOPELE MSAESERKTETIPEWKQEEVDAIVEMIESYESVGVVNIAGIPSROLODMRRDIHGT-AELRVSRHTLIELAIKRAAOELGOPELE MSESEVROTEVIPOWKREEVDELVDFIESYESVGVVGVAGIPSROLODMRRDIHGS-AAVRMSRHTLVNRALDEVNDGFE MSAEQRTTEEVPEWKRQEVAELVDLIETYDSVGVVNUTGIPSKOLODMRRGIHGQ-AALRMSRHTLVNRALDEVNDGFE MSAEQRTTEEVPEWKRQEVAELVDLIETYDSVGVVNUTGIPSKOLODMRRGIHGQ-AALRMSRHTLVNRALEEAGDGLD MKINPKKEIVSELAQDITKSKAVAIVDIKGVRTRODIRGKNRGK-UNLKVIKKTLFKALENLGDEKLS MKINPKKEIVSELAQDITKSKAVAIVDIKGVRTROMODIRAKNRDK-VKIKVVKKTLLFKALDSINDEKLT | | | | | | |
| MAHVAEWKKKEVEELANIIKSYPVIALVDVAGVPAYPLSKMRDKLR-GKALLRVSRHTLIELAIKRAAQELGOPELE MSAESERKTETIPEWKQEEVDAIVEMIESYESVGVVNIAGIPSROLODMRRDLHGT-AELRVSRHTLIERALDDVDDGLE MSAESERQTEVIPOWKREEVDELVDFIESYESVGVVGVGVAGIPSROLOSMRRELHGS-AAVRMSRHTLVRALDEVNDGLE MSAEEORTTEEVEWKROEVAELVDLIETYDSVGVVNTGIPSKOLODMRRGLHGG-AALRMSRHTLIVRALDEVNDGLD MSEEORTTEEVEWKROEVAELVDLIETYDSVGVVNTGIPSKOLODMRRGLHGG-AALRMSRHTLIVRALDEVNDGLD MKEVSOOKKELVNEITORIKASRSVAIVDIRGIPSROLODMRRGLHGV-TILVVIKKTLIFKALENLGDEKLT | | | | | | |
| MSAESERKTET IPEWKQEEVDA IVEMIES YESVGVVNIAG IPERQLODMRRD LHGT-AELRVSRNTLLE RALDDVDDGLE MSESEVRQTEVIPQWKREEVDELVDFIESYESVGVGVAGIPSRQLOSMRRELHGS-AAVRMSRNTLVNRALDEVNDGFE MSAEGRTTEEVPEWKRQEVAELVDLIETYDSVGVNVYGIPSKQLODMRRGLHGQ-AALRMSNNTLVRALDEAGDGLO MKEVSQQKKELVNEITORIKASRSVAIVDTAGIPSKQLODMRRGLHGV-AALRMSNNTLVRALEEAGDGLO MKEVSQQKKELVNEITORIKASRSVAIVDTAGIPSKQLODMRRGLHGV-AALRMSNNTLVSKILLFKALENGDEKLS MKTINPKKEIVSELAODITKSKAVAIVDIKGVRTROMODIRAKNBCK-VKIKVVKKTLLFKALDSINDEKLT | | | | | | |
| MSESEVRQTEVIPQWKREEVDELVDFIESYESVGVVGVAGIPSRQLQSMRRELHCS-AAVRMSRNTLVNRALDEVNDGFE MSAEEQRTTEEVDEWKRQEVAELVDLLETYDSVGVVNYTCIPSKQLQDMRGLHGQ-AALRMSRNTLLVRALEEAGDGLD MKEVSQQKKELVNEITQRIKASRSVAIVDTRGITRROTQIRGKNRGK-INLKVIKKTLLFKALENLGDEKLS MKINPKKEIVSELAQDITKSKAVAIVDIKGVRTROMQDIRAKNRDK-VKIKVVKKTLLFKALDSINDEKLT MTEPAQWKIDFVKNLENEINSRVVAAIVSIKGLRNNFOKIRNSIRDK-ARIKVSRARLLRLATENTGK-NNIV | | | | | | |
| MSAEEQRTTEEVPEWKRQEVAELVDLLETYDSVGVVNVTGIPSKOLODMRRGIHGQ-AALRMSRNTLLVRALEEAGDGLD | | | | | | |
| MKEVSQQKKELVNE IT OR IKASRSVAIVO <mark>TAG IRTRO IODIRGKNRGK-INLKVIKKT</mark> LLFKALENLGDEKLS MRKIN <mark>P</mark> KKKE IVSELAODITKSKAVAIVDIKGVRTROMODIRAKNRDK-VKIKVVKKTLLFKALDSINDEKLT MTEPAQWKIDFVKNLENE INSRKVAAIVSIKGLRNNEFOKIRNSIRDK-ARIKVSRARLLRLAIENTGKNNIV | | | | | | |
| MRKIN <mark>P</mark> KKKE IVSELA <mark>O</mark> DITKSKAVAIVDIKGVR <mark>I</mark> ROMODIRAKNRDK-VKIKVVKKILLFKALDSINDEKLT MTEPAQWKIDFVKNLENE INSRKVAAIVSIKGLRNNEFOKIRNSIRDK-ARIKVSRARLLRLAIENIGKNNIV | | | | | | |
| | | | | | | |
| 110203040 | | | | | | |
| | 110 | | 40 | 50 | | |

Information

Find genetic risk factors

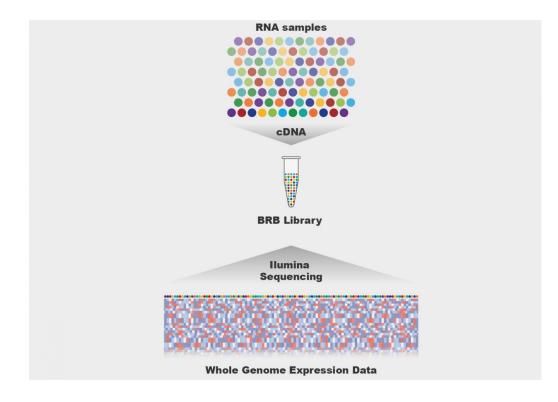


Credit

<u>Bioinformatics</u> deals with untangling big data for biological insight

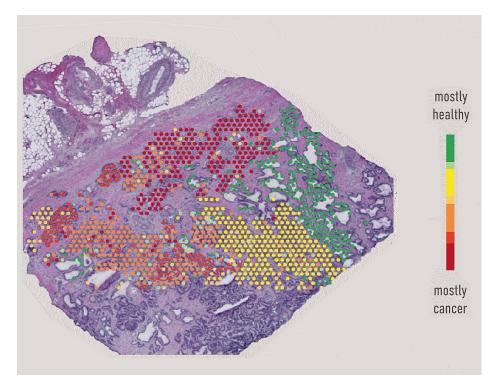
Data

mRNA of cancer cells in a tumor



Information

Early detection of benign to cancerous cell transition



Credit

Credit

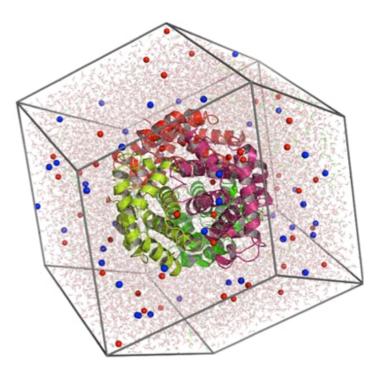
<u>Modeling</u> employs representations that mimic key biological phenomena

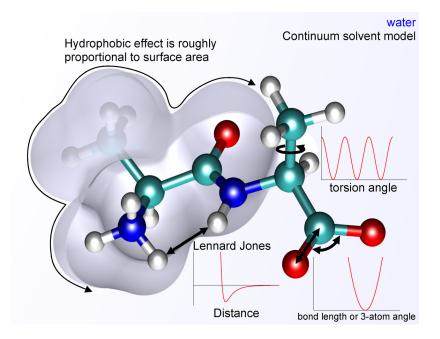
Phenomena

Representation

Protein-protein binding

Classical force fields

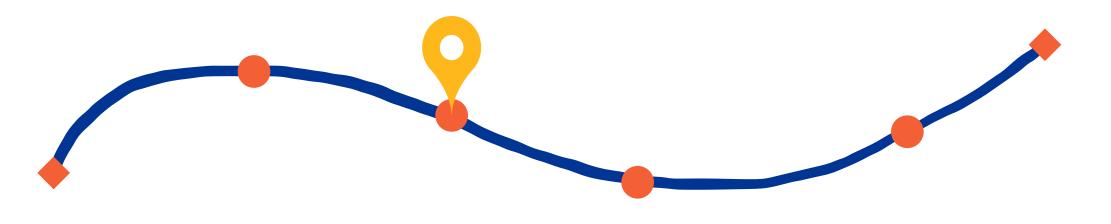




Credit

Credit

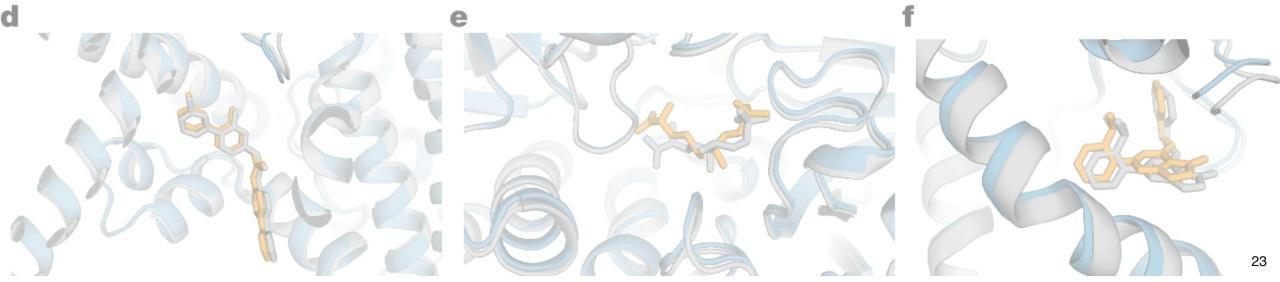
After today, you should be able to



- 1. Describe the course structure, expectations, and available resources for success.
- 2. Define computational biology and explain its interdisciplinary nature.
- 3. Identify key applications and recent advancements.
- 4. Understand the balance between applications and development.
- 5. Identify potential career paths and educational opportunities.

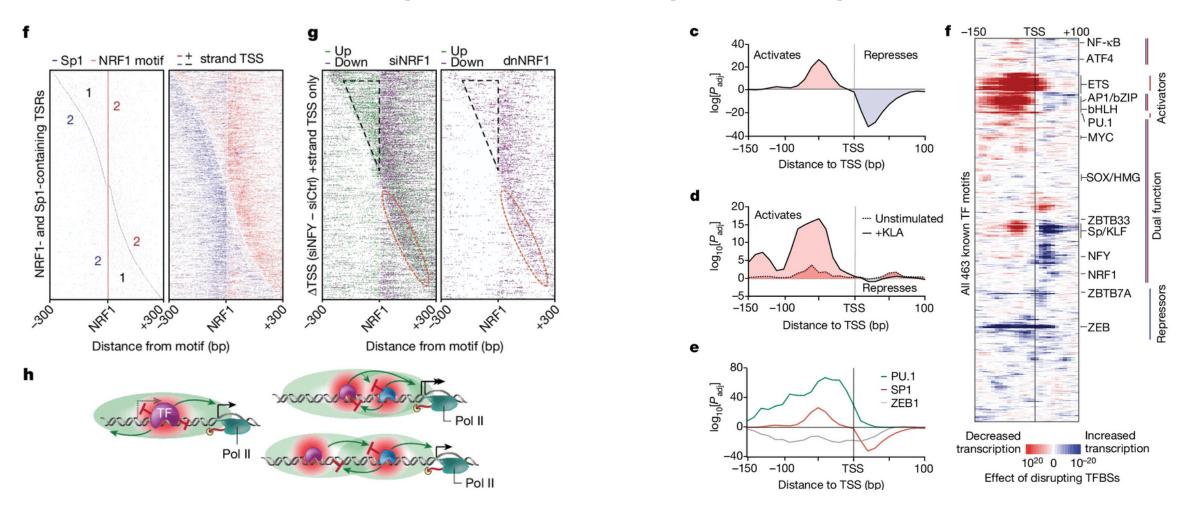
^bAlphaFold 3

"AlphaFold 3 can predict the joint structure of complexes including proteins, nucleic acids, small molecules, ions, and modified residues."



HOMER2

"We show that the effect of transcription factor binding on transcription initiation is position dependent."



Miniprot: protein-genome aligner

"Miniprot [...] is tens of times faster than existing tools while achieving comparable accuracy on real data."

Where?

Nucleotide-level annotation

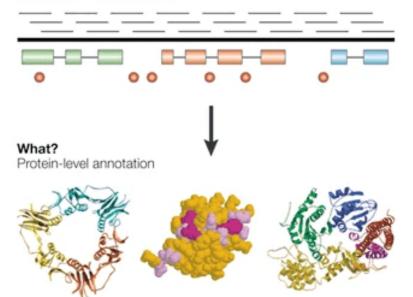
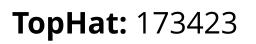


Table 1. Evaluating protein-to-genome alignment

| Genome species | Human | Human | Human | Human | Human |
|------------------------|-----------|-----------|-----------|-----------|-----------|
| Protein species | Zebrafish | Zebrafish | Zebrafish | Zebrafish | Zebrafish |
| Aligner | Miniprot | Miniprot | Spaln2 | Spaln2 | GeMoMa |
| Splice model | Human | General | Human | Default | N/A |
| Elapsed time (s) | 267 | 257 | 10 708 | 11 097 | 8718 |
| Peak RAM (GB) | 21.8 | 22.5 | 9.3 | 8.9 | 146.9 |
| No. of protein | 25 007 | 25 007 | 25 007 | 25 007 | 25 007 |
| No. of multi-exon | 16 866 | 17 104 | 13 643 | 13 854 | 23 109 |
| No. of predicted junc. | 157 918 | 161 295 | 151 388 | 209 312 | 204 764 |
| No. of non-ovlp. junc. | 482 | 802 | 1206 | 15 658 | 5712 |
| No. of confirmed junc. | 145 545 | 144 734 | 136 916 | 129 645 | 153 781 |
| % confirmed junc. | 92.16 | 89.73 | 90.44 | 61.94 | 75.10 |
| % base SN | 63.11 | 63.16 | 57.16 | 55.74 | 67.02 |
| % base SP | 95.43 | 94.91 | 95.11 | 86.75 | 88.70 |

Why would we use protein-genome instead of genome-genome mapping?

A. Protein-genome mapping is more sensitive for detecting distant homologs
B. Genome-genome mapping is too slow for large-scale comparisons
C. Protein-genome mapping allows for the detection of RNA editing events
D. Genome-genome mapping cannot handle intron-exon structures



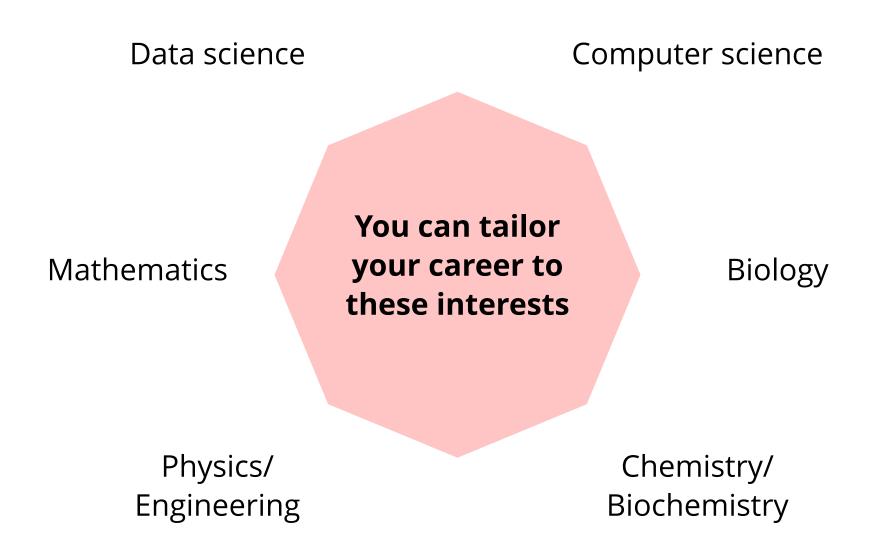
(Not for points)



After today, you should be able to

- 1. Describe the course structure, expectations, and available resources for success.
- 2. Define computational biology and explain its interdisciplinary nature.
- 3. Identify key applications and recent advancements.
- 4. Understand the balance between applications and development.
- 5. Identify potential career paths and educational opportunities.

Computational Biology is broad



We will touch on all of these topics in this course

Method development or applying tools?



Typically, it is harder to pick up after the fact (a different way of thinking)

Many, many, many specalities

Both separately are pretty saturated

After today, you should be able to

- 1. Describe the course structure, expectations, and available resources for success.
- 2. Define computational biology and explain its interdisciplinary nature.
- 3. Identify key applications and recent advancements.
- 4. Understand the balance between applications and development.
- 5. Identify potential career paths and educational opportunities.

Bioinformatics Scientist

Description: Develops software tools and approaches for analyzing biological data, particularly genomic and proteomic data.

Qualifications:

- PhD in Bioinformatics, Computational Biology, or related field
- Strong programming skills (Python, R, C++)

Example companies: UPMC, Illumina, 23andMe, Genentech, Regeneron Pharmaceuticals, Broad Institute

Expected Salary: \$80,000 - \$130,000

Computational Biologist

Description: Applies computational methods to study biological systems, often focusing on modeling complex biological processes.

Qualifications:

- PhD in Computational Biology, Systems Biology
- Expertise in mathematical modeling and simulation
- Strong programming and data analysis skills

Example companies: Moderna, Vertex Pharmaceuticals, Biogen, Allen Institute for Brain Science, Flatiron Health

Expected Salary: \$75,000 - \$135,000

Biostatistician

Description: Applies statistical methods to analyze biological and health-related data, often in clinical trials or epidemiological studies.

Qualifications:

- Master's or PhD in Biostatistics or related field
- Strong background in statistics and mathematical modeling
- Proficiency in statistical software (R, SAS, STATA)

Example companies: Pfizer, Merck, Johnson & Johnson, IQVIA, Fred Hutchinson Cancer Research Center

Expected Salary: \$72,000 - \$119,000

Molecular Modeler

Description: Uses computational methods to model and simulate molecular structures and interactions, often in drug discovery.

Qualifications:

- PhD in Computational Chemistry, Biophysics, or related field
- Experience with molecular dynamics simulations
- Knowledge of drug design principles

Example companies: Schrödinger, Novartis, GlaxoSmithKline (GSK), Atomwise, Dassault Systèmes BIOVIA

Expected Salary: \$85,000 - \$140,000

If these careers sound interesting, a PhD should be on your radar

Note: There tend to be more jobs in bioinformatics than simulation and modeling

Okay, but what about a Bachelor's degree?

Challenging for computational biology jobs, but other options are available

Focus on one half of your major

Biology

Computer Science

I'm unfamiliar with options here (your advisors are well-versed) Software engineer, data science, machine learning, web development

To be honest: Engineering degrees give the highest chance for a well-paying job after graduation

What will help you prepare for_

Everyone applying for the same positions has a college degree

Distinguish yourself with extracirriculars

Employers and graduate schools do not care about the classes you took, they care about **what you can do**

How to do this? Show and tell

Show what you can do

Contribute to open-source projects

| aalexmmaldonado commented on Dec 12, 2019 Contributor Reviewers The ORCA MP2 module uses a different flag other than "CARTESIAN GRADIENT" for gradients. Here, I just added another block to the ORCA parser that catches these MP2 gradients. Reviewers Image: Solution of the ORCA parser that catches these MP2 gradients. Image: Solution of the ORCA parser that catches these MP2 gradients. Reviewers Image: Solution of the ORCA parser that catches these MP2 gradients. Image: Solution of the ORCA parser that catches these MP2 gradients. Image: Solution of the ORCA parser that catches these MP2 gradients. Image: Solution of the ORCA 4.2 test files 186e157 Image: Solution of the ORCA parser the ORCA parser Image: Solution of the ORCA parser the ORCA parser Image: Solution of the ORCA MP2 gradients to ORCA parser 026b834 Projects None yet Image: Solution of the ORCA MP2 gradients 026b834 Projects None yet Image: Solution of the ORCA MP2 gradients 00 -403,6 f +403,25 00 def splitter(line): Milestone Image: Solution of the ORCA of the ORCA parser of the ORCA of | +25 <mark>-6</mark> 🔳 | | | F. Checks 0 主 Files changed 2 | -o- Commits 7 | 10 | vorsatio |
|--|-----------------------|---|----------------------------|---|-------------------------------|-----------------|-----------|
| The ORCA MP2 module uses a different flag other than "CARTESIAN GRADIENT" for gradients. Here, I just added another block to the ORCA parser that catches these MP2 gradients. Image: String of the ORCA parser that catches these MP2 gradients. Image: String of the ORCA parser that catches these MP2 gradients. Image: String of the ORCA parser that catches these MP2 gradients. Image: String of the ORCA parser that catches these MP2 gradients. Image: String of the ORCA parser that catches these MP2 gradients. Image: String of the ORCA and the ORCA parser the ORCA parser the ORCA parser the ORCA parser to ORCA parser the ORCA parser to ORC | | | | E Pries changed 2 | -> commits / E3 | 10 | iversatio |
| The ORCA MP2 module uses a different flag other than "CARTESIAN GRADIENT" for gradients. Here, I just added another block to the ORCA parser that catches these MP2 gradients. Image: block to the ORCA parser that catches these MP2 gradients. Image: block to the ORCA parser that catches these MP2 gradients. Image: block to the ORCA parser that catches these MP2 gradients. Image: block to the ORCA parser that catches these MP2 gradients. Image: block to the ORCA parser that catches these MP2 gradients. Image: block to the ORCA 4.2 test files Image: block to ORCA 4.2 test files Image: block to ORCA 4.2 test files Image: block to ORCA parser Image: block to ORCA MP2 gradients Image: block to ORCA MP2 gradients Image: block to the ORCA theorem that the original term of ter | | | Contributor | 12, 2019 | do commented on Dec 12, 2 | naldonad | aalexn |
| Image: Second particular and the | | | Here, I just added another | 5 | | | |
| Adding some ORCA 4.2 test files 186e157 Adding some ORCA 4.2 test files 186e157 Adding MP2 gradients to ORCA parser d9bd909 Adding ORCA MP2 gradients 026b834 Projects None yet Shivupa requested changes on Dec 22, 2019 View reviewed changes Cclib/parser/orcaparser.py Outdated 00 -403, 6 +403, 25 00 def splitter(line): 403 403 if not hasattr(self, 'grads'): | | 🏈 berquist | | hese MP2 gradients. | A parser that catches these l | the ORCA | block t |
| alexmmaldonado added 3 commits 5 years ago Adding some ORCA 4.2 test files Adding MP2 gradients to ORCA parser Adding ORCA MP2 gradients to ORCA parser Adding ORCA MP2 gradients Adding orca MP2 gradients | | Assignees | | | | | \odot |
| Adding some ORCA 4.2 test files Adding MP2 gradients to ORCA parser Adding ORCA MP2 gradients to ORCA parser Adding ORCA MP2 gradients Adding ORCA M | | No one assigned | | | | | |
| Adding MP2 gradients to ORCA parser Adding ORCA MP2 gradients Adding ORCA MP2 gradients Adding ORCA MP2 gradients Adding ORCA MP2 gradients Milestone Cclib/parser/orcaparser.py Outdated 00 -403,6 +403,25 00 def splitter(line): 403 403 if not hasattr(self, 'grads'): | | Labels | | | | | |
| Adding ORCA MP2 gradients Adding ORCA MP2 gradients shivupa requested changes on Dec 22, 2019 View reviewed changes Cclib/parser/orcaparser.py Outdated 00 -403, 6 +403, 25 00 def splitter(line): 403 403 if not hasattr(self, 'grads'): | | ORCA regression | 186e157 | es | ome ORCA 4.2 test files | Adding so | ÷ 🔞 |
| Image: Second | | | d9bd909 | parser | P2 gradients to ORCA parse | Adding MP | ÷ 🔕 |
| Image: Shivupa requested changes on Dec 22, 2019 View reviewed changes Milestone cclib/parser/orcaparser.py Outdated v1.6.3 403 403 if not hasattr(self, 'grads'): Development | | - | 026b834 | | RCA MP2 gradients | Adding OR | - 🔕 |
| cclib/parser/orcaparser.py Outdated v1.6.3 00 -403,6 +403,25 00 def splitter(line): 403 403 if not hasattr(self, 'grads'): | | None yet | | | | | |
| 00 -403,6 +403,25 00 def splitter(line): 403 403 if not hasattr(self, 'grads'): | | Milestone | View reviewed changes | 22, 2019 | iested changes on Dec 22, 20 | upa requ | 🗈 shi |
| 403 403 if not hasattr(self, 'grads'): Development | | v1.6.3 | | tdated | er/orcaparser.py Outdated | lib/pars | c |
| The hadded (set), grads). | | | | | | | |
| 404 404 self.grads = [] Successfully merging this pull reque | iest may clo | Development Successfully merging this pull request | | ot hasattr(self, 'grads'): self.grads = [] | | | |
| 405405self.grads.append(grads)baccessical participation | est may cro | | | 5 13 | | | |

Show what you can do

Hackathons and competitions





Join Competition

Leash Bio - Predict New Medicines with **BELKA**



Predict small molecule-protein interactions using the Big Encoded Library for Chemical Assessment (BELKA)

Data Models Discussion Leaderboard Rules Overview Code

Overview

Start

In this competition, you'll develop machine learning (ML) models to predict the binding affinity of small molecules to specific protein targets - a critical step in drug development for the pharmaceutical industry that would pave the way for more accurate drug discovery. You'll help predict which drug-like small molecules (chemicals) will bind to three possible protein targets.

14 days ago

Competition Host Leash Bio

Prizes & Awards

\$50,000 Awards Points & Medals

Participation

Close

3 months to go

2,571 Entrants 330 Participants 304 Teams 1,927 Submissions

Your marketable skills are learned outside the classroom

Classes give foundational knowledge to learn hands-on skills in research and internships

Computational biology: You will get a small taste of this in classes; you need some research or project experience

Computer science: Python, GitHub, machine learning

Graphic design: Illustrator/Inkscape, Photoshop/Gimp, Blender

Communication: Writing and presenting

Before the next class, you should



Lecture 01: Introduction to computational biology

Lecture 02: DNA sequencing

- Check that you are subscribed to Canvas notifications
- Make an account on the following sites (for the project):
 - usegalaxy.org
 - alphafoldserver.com