Applicant: Lucas De Araujo Nogueira Melo (52242971)

Program: Ph.D. in Genome Science and Technology (VGDPHD-LE)

Entry period: September 2024

Application comments:

No comments available

Order of content:

Application form
Resume
Statement of Interest/Intent
Transcripts & Diplomas – Unofficial
eReference (eRef) Responses
Reference Letter

MELO, LUCAS DE ARAUJO NOGUEIRA ()
52242971

Degree Selection

Submission Date: 19/Nov/2023

Campus	Program (VGDPHD-LE)	Academic Year	Term	Term Start
Vancouver	Ph.D. in Genome Science and Technology	2024-2025	W1	Sep 2024

Source of Interest

How did you find out about UBC?
Referral from Professor

Personal and Contact Details

Student Number	Family Name (Surname)	Preferred Name
52242971	MELO	

Title	Given Name	Middle Name	Former Family Name (Surname)
MR	LUCAS	DE ARAUJO NOGUEIRA	

Date of Birth	Gender	Country of Birth	Country of Current Citizenship
01/Dec/1999	Male	Brazil	United States of America
Dual Citizenship	**Primary Spoken Language**	**Other Spoken Language**	**Visa Type**
Brazil	English		International Student

Address Line (1 & 2)
301 W 110TH ST 4V

City	Province, State or Region	Postal or Zip Code	Country
NEW YORK	NY	10026	United States of America

Day Telephone Number	Evening Telephone Number	Email Address
5103711053		lucas.melo@columbia.edu

Do you identify yourself as an Aboriginal person of Canada?
Do you identify yourself as a Racialized person?
No

Academic History

- Applicant indicates that they have only attended post-secondary institution(s) other than UBC.

Columbia University

Institution Country:	**United States of America**
Start Date:	01/Sep/2018
End Date (or Expected End):	01/May/2022
Program of Study:	Computer Science
Credential Status	Conferred / Complete
Date Conferred:	01/May/2022
Credential Received:	Bachelor's
Awards & Honours received with this degree:	» Bachelor of Science, summa cum laude » Computer Science Scholarship Award (2022) "for outstanding academic achievement in computer science" » Behring Foundation Award (2022) "for students that best exhibit academic excellence, have entrepreneurial interests, and possess leadership skills" » C. Prescott Davis Scholar (2018) "for entering students of extraordinary abilities to enjoy the advantage of co-curriculum enhancement programs to prepare them for excellence in their chosen professions"
Required to withdraw:	No
Self Reported GPA:	
Used for Basis of Admission to UBC:	Yes

GPA Calculations Summary

Calculation Name	Purpose	Date of Calculation	Minimum GPA Req'd	GPA Calculation	GPA Rank	Meets Progrm Requirements	Meets UBC Requirements	First Class Standing?
Computer Science	Admissions	29/11/2023		4.13		Yes	Yes	Yes

- No **UBC** academic history found for this student number (52242971)

Funding

Standard Questions

Primary Funding

SOURCE of the support	
DOLLAR amount	
Includes TUITION fees?	
WHEN the support will commence	
WHEN the support will end	

SOURCE of the support	As an international student in a PhD program, I will apply for fellowships once I join a lab.

Experience & Interests

Standard Questions

Areas of Interest

Faculty Members

De Boer, Carl
Wasserman, Wyeth

Please provide a brief statement of your academic and/or professional goals and how these align with this graduate program.
Please describe any research and/or work experience (including publications, etc.) you've undertaken that is relevant to your proposed field of study.
My research experiences are elaborated and contextualized into my academic goals in my Statement of Intent. Briefly, my main research experience is that I have been working in the lab of Prof. Bussemaker at Columbia University for the past several years developing computational methods for modeling high-throughput protein-ligand binding affinity data. Initial results from my work on in-vivo occupancy data were published previously (Rube et al. Nature Biotechnol 2022), with a follow-up publication on a new package that I created, PyProBound, soon to be submitted. My current work focuses on modeling protein-protein interactions using bacterial and phage display.

Program-Specific Questions

Briefly discuss your background in life sciences, including academic, work or other experiences that may assist the admissions committee. Please limit your response to one page.	During my undergraduate degree I also completed all of the requirements of the Biology major, including coursework in organic chemistry and biochemistry, as well as electives focusing on molecular biology and biophysics. In addition to the experiences listed in my Statement of Intent, I also participated in the Undergraduate Research Program at Cold Spring Harbor Laboratory, where I worked on a medical informatics project in the lab of Dr. Krasnitz. I also worked as a teaching assistant in the Biological Sciences department, where I taught a recitation for the Intro sequence and helped create a new course targeting first-years to support students in transitioning to an undergraduate-level life sciences education.
Briefly discuss your background in quantitative sciences (math, statistics, computer science, engineering, physics) including academic, work or other experiences that may assist the admissions committee. Please limit your response to one page.	As a Computer Science major with a focus on machine learning, I took extensive coursework in math, statistics, computer science, and physics, including discrete math, linear algebra, multivariate calculus, machine learning, and more. In addition to the experiences listed in my Statement of Intent, I worked at the Sean N. Parker Center at Stanford where I created a cloud-based computational pipeline to facilitate Big Data approaches in the center.

Referee 1

Name	Harmen J. Bussemaker
Job Title / Occupation	Professor of Biological Sciences
Institution / Company / Organization	COLUMBIA UNIVERSITY
Type of Reference	**Academic**
Address	1212 AMSTERDAM AVE MC2441NEW YORK NEW YORK United States of America 10027
Referee Email / Website	hjb2004@columbia.edu
Telephone #	
Notes to Referees	

Referee 2

Name	David A. Knowles
Job Title / Occupation	Assistant Professor of Computer Science
Institution / Company / Organization	COLUMBIA UNIVERSITY & NEW YORK GENOME CENTER
Type of Reference	**Academic**
Address	500 W 120 ST ROOM 450 MC0401NEW YORK NEW YORK United States of America 10027
Referee Email / Website	scurtiss@nygenome.org
Telephone #	
Notes to Referees	

Referee 3

Name	H. Tomas Rube
Job Title / Occupation	Assistant Professor in Applied Mathematics
Institution / Company / Organization	UNIVERSITY OF CALIFORNIA, MERCED
Type of Reference	**Academic**
Address	5200 NORTH LAKE RD ACS 371MERCED CALIFORNIA United States of America 95343
Referee Email / Website	trube@ucmerced.edu
Telephone #	
Notes to Referees	

Lucas Melo

301 W 110th St 4V, New York, NY 10026
+1 (510) 371-1053 • lucas.melo@columbia.edu

Education

Columbia University, *School of Engineering and Applied Science* New York, NY
Bachelor of Science in Computer Science (Intelligent Systems track), *summa cum laude* May 2022
Relevant Courses: Machine Learning, Computational Genomics, Macromolecular Structure

Research Experience

Columbia University, *Department of Biological Sciences* New York, NY
Research Staff Assistant – Bussemaker Lab June 2022 – Present

- Leading the analysis on a series of collaborations on receptor-ligand interactions
- Creating PyProBound, an easily extensible and efficient machine learning framework for quantitative methods on high-throughput sequencing data
- Developing approaches for isolating protein binding epitopes and predicting functional readouts
- Working closely with collaborators to optimize assays and reduce biological biases

Cold Spring Harbor Laboratory Cold Spring Harbor, NY
Undergraduate Research Program – Krasnitz Lab June – August 2021

- Identified clinically relevant copy number variation patterns in acute myeloid leukemia through unsupervised clustering

Columbia University, *Department of Biological Sciences* New York, NY
Summer Undergraduate Research Fellow – Bussemaker Lab June – August 2020

- Developed a fragment-level ChIP-seq motif discovery and cooperativity approach that does not require peak calling or read mapping, retaining biologically-relevant low-affinity binding sites

Publications & Posters

- **Melo LAN**, Li X, Bussemaker HJ (2023) "Assessing allele-specific binding variant effects with TF-DNA binding models" (*in preparation*)
- **Melo LAN**, Knowles DA (2022) "Predicting longitudinal BCR repertoires" Poster presented at: Machine Learning in Computational Biology, 21-22 November 2022
- Rube HT, Rastogi C, Feng S, Kribelbauer JF, Li A, Becerra B, **Melo LAN**, Do BV, Li X, Adam HH, Shah NH, Mann RS, Bussemaker HJ (2022) "Prediction of protein–ligand binding affinity from sequencing data with interpretable machine learning" *Nature Biotechnology*

Awards & Honors

- Computer Science Scholarship Award – *Columbia Department of Computer Science* May 2022
- Behring Foundation Award – *Columbia Department of Computer Science* May 2022
- Genentech Fellow – *Columbia Department of Biological Sciences* June 2020
- C. Prescott Davis Scholar – *Columbia School of Engineering and Applied Sciences* April 2018

Teaching & Activities

Columbia University, *Department of Biological Sciences* New York, NY
Teaching Assistant III September 2020 – May 2022

- Foundations of Biology: Helped create a new course targeting first-year students who need additional support and preparation to achieve their goals in the life sciences
- Introductory Biology: Organized collaborative problem solving recitations of fifteen students to develop a molecular understanding of biology

Columbia Science Review New York, NY
Editorial Board Member January 2019 – May 2022

The wealth of high-throughput sequencing (HTS) data available for computational biology has led to an explosion of model interpretability approaches, in part due to the breadth of the field, which spans everything from quantitative *in vitro* experiments to assays of higher-order *in vivo* systems. Yet across these disparate applications, domain knowledge can inform crucial modeling insights and creative wet-lab experiments to decompose the signal from experimental artifacts and disentangle intricate relationships between biophysical parameters.

During my undergraduate studies at Columbia University in machine learning and molecular biology, I worked in the lab of Professor Harmen Bussemaker. Despite being in the Biology department, most came from a physics background, making it the ideal interdisciplinary environment for exploring HTS data with biophysically-informed sequence-to-function models.

With funding from the Genentech Fellowship in the summer 2020, I applied ProBound, a general ML method for *in vitro* assays, towards *in vivo* assays. Traditional ChIP-seq methods involve peak calling to identify only the most enriched sites. I instead quantitatively modeled the HTS reads directly, retaining all fragments so the entire affinity range can be interpreted. This approach created glucocorticoid receptor (GR) binding models of comparable quality to *in vitro* models while simultaneously discovering its cofactors, and was published as a section of the ProBound publication (Rube et al. *Nature Biotechnol* 2022). In the process, I learned how to creatively integrate a molecular understanding of wet-lab experiments into an ML framework.

I was soon coming up with more modeling ideas than I could implement with ProBound, so I later created the PyProBound package, a new easily extensible and efficient ML library for HTS of protein-ligand interactions. It is already being widely used in our lab and the labs of our collaborators and we hope it will advance quantitative methods in the broader field. I designed it so new layers and loss functions can be easily implemented to suit experimental designs. For example, a demonstration of the flexibility of PyProBound in the manuscript being prepared infers a biophysical binding model through a beta-binomial likelihood of allele-specific binding.

In my last year as an undergraduate, I wanted to push beyond protein-DNA interactions. As part of a course taught by Professor David Knowles, I created a new research project in immunology. Over the course of an infection, the population of antibodies evolves to counteract the replicating virus. I wanted to leverage this natural evolution to train a deep neural network for antibody-antigen binding by combining it with a dynamical system that models a time series of samples from COVID-19 patients. However, I discovered that limitations in sequencing technology meant I could not capture the vital relationship between heavy and light chains.

While I was initially frustrated to learn my approach would not work as envisioned, I learned how to creatively adapt a research project to new challenges. I instead used the model to explain the trajectory of antibody replication, validating against individual clonotypes, and identifying the sequence determinants of clonal expansion. What started as a class project became a poster at the 2022 Machine Learning in Computational Biology conference.

After graduating at the top of my CS cohort, I continued at the Bussemaker lab to advance a series of projects with several collaborators studying receptor-ligand interactions such as PD-1:PD-L2, GPCRs, and SH2 and PDZ domains using HTS of randomized peptide libraries

expressed and selected using bacterial and phage display. I am working closely with experimental scientists, using my models to help identify and address bottlenecks and sources of bias in their protocols, while in turn learning about experimental subtleties I can use to improve my modeling. This two-way feedback creates strong synergies, resulting in incredibly precise bacterial display screens with over 1M unique peptides even after several rounds of affinity selection, as well as highly accurate protein-protein binding models spanning over two orders of magnitude. Over the process I have delved deep into the intricacies of protein affinity screening, and I am currently applying this understanding to develop a new computational approach that leverages counterselection experiments to isolate the affinity contribution of different protein epitopes. These advances bring us closer towards single-shot protein engineering, and the sum of these projects has impressed upon me the importance of meaningful wet lab-dry lab collaboration in creating breakthroughs in computational biology.

As I continue my academic career, I am especially interested in advancing the quantitative skills I have developed in protein-DNA and protein-protein recognition modeling towards investigating the *cis*-regulatory code, which is often misregulated in heritable diseases and cancer. Experimental and computational innovations created at the University of British Columbia are providing new quantitative insights into complex gene regulatory networks. In particular, Professor Carl de Boer is pioneering the use of functional assays on synthetic DNA libraries with much greater diversity and lower bias than the human genome. Additionally, Professor Wyeth W. Wasserman is leveraging years of experience in gene regulation to develop interpretable deep neural networks such as ExplaiNN, which charts a way forward in bridging the gap between “black-box” and “white-box” models. My in-depth experience in biophysically-interpretable ML on HTS of randomized libraries has prepared me to hit the ground running and creatively integrate computational modeling with a molecular understanding of experiments and regulatory networks to contribute to and push forward on any of these endeavors.

The Genome Science and Technology graduate program at UBC is ideal for me because of its focus on cross-disciplinary technological innovation. With each new research project, from *in-vivo* assays to protein-ligand experiments, I quickly built up the domain knowledge necessary to create new analytical approaches and collaborate closely with experimental scientists to push the boundaries of what is possible with mixed computational/experimental approaches. Through initiatives such as the research rotation award and joint seminars, GSAT introduces students to faculty from a wide variety of disciplines, fostering the type of interdisciplinary collaborations that have been so valuable to my research over the years. Within the cohort I will provide a strongly quantitative background that melds together an in-depth understanding of both experimentational and computational techniques to catalyze these collaborations.

My research goal is to develop frameworks for integrating domain knowledge, experimental techniques, and modeling strategies to disentangle complex biological systems. The GSAT graduate program is the best place for me to develop these frameworks, and will prepare me to move on to postdoctoral positions where I can continue developing these interdisciplinary collaborations as I progress in my academic career.

The wealth of high-throughput sequencing (HTS) data available for computational biology has led to an explosion of model interpretability approaches, in part due to the breadth of the field, which spans everything from quantitative *in vitro* experiments to assays of higher-order *in vivo* systems. Yet across these disparate applications, domain knowledge can inform crucial modeling insights and creative wet-lab experiments to decompose the signal from experimental artifacts and disentangle intricate relationships within biological systems.

During my undergraduate studies at Columbia University in machine learning and molecular biology, I worked in the lab of Professor Harmen Bussemaker. Despite being in the Biology department, most came from a physics background, making it the ideal interdisciplinary environment for exploring HTS data with biophysically-informed sequence-to-function models.

In my first project, I applied ProBound, an ML framework designed for *in vitro* SELEX experiments, towards *in vivo* methods. ProBound models assays by combining a convolutional partition function of sequence recognition, biophysically-derived formulas of enrichment, and a principled loss function. I used this approach to directly model ChIP-seq reads without peak calling, producing glucocorticoid receptor (GR) binding models of similar quality to *in vitro*-derived models while simultaneously discovering its cofactors. My work became a section of the ProBound publication (Rube et al. *Nature Biotechnol* 2022), and in the process I learned how to integrate a molecular understanding of wet-lab experiments into an ML framework.

I was soon coming up with more modeling ideas than I could implement with ProBound, so I created PyProBound, an easily extensible and efficient ML library for HTS of protein-ligand interactions. I designed it so new model architectures can be easily implemented to suit experimental designs. For example, the manuscript in preparation shows how to improve models by adjusting for the fragmentation bias at the ends of reads, and how to infer binding models through a beta-binomial likelihood of allele-specific binding. It is already widely used in our lab and the labs of our collaborators, and we hope it will advance quantitative methods in the field.

In my senior year, I wanted to push beyond protein-DNA interactions. As part of a course taught by Professor David Knowles, I began a new research project in immunology. During an infection, the population of antibodies evolves to counteract the replicating virus. I wanted to leverage this natural evolution to train a deep neural network for antibody-antigen binding by combining it with a dynamical system that models a time series of samples from COVID-19 patients. However, I discovered that sequencing technology cannot capture both heavy and light chains at the throughput necessary to produce a generalizable model of antibody binding.

While I was initially disappointed to learn my approach would not work as envisioned, I learned how to creatively adapt a research project to new challenges. I instead used the model to explain the trajectory of antibody replication, validating against individual clonotypes, and identifying the sequence determinants of clonal expansion. What started as a class project became a poster at the 2022 Machine Learning in Computational Biology conference.

After graduating at the top of my CS cohort, I continued at the Bussemaker lab to lead projects involving several collaborators studying receptors such as PD-1, GPCRs, and SH2 and PDZ domains using HTS of peptide libraries expressed with bacterial and phage display. I work closely with experimental scientists, using my models to identify and address bottlenecks and sources of bias in their protocols, while in turn learning about experimental subtleties to improve my models. This two-way feedback unlocks powerful synergies, resulting in bacterial display screens with over 1M unique peptides even after several rounds of selection, as well as models that accurately predict binding affinity over more than two orders of magnitude.

I am currently leveraging the knowledge I have developed in biopanning to develop a novel sequence-based computational approach combining multiple experiments to isolate the affinity contribution of distinct epitopes. These advances bring us closer towards single-shot protein engineering, and the sum of these projects has impressed upon me the importance of meaningful wet lab-dry lab collaboration in creating breakthroughs in computational biology.

With each research experience, I have found that a more quantitative molecular approach allows for greater creativity by providing a toolset for transformative thinking. A doctoral program will provide me with the training necessary to harness and grow that creativity.

The Genome Science and Technology program at UBC is ideal for me because of its focus on cross-disciplinary technological innovation. In each project, I quickly gained the domain knowledge needed to design and collaborate on new approaches. I value the opportunity to explore a variety of topics with a diverse group of students and faculty from different academic and personal backgrounds. Through initiatives such as the research rotation award and joint seminars, GSAT introduces students to faculty from a wide range of disciplines, fostering invaluable interdisciplinary collaborations. Within the cohort my interest in integrating experimental and computational techniques will catalyze these discussions and collaborations.

I am especially interested in applying the skills I have gained in protein recognition modeling towards investigating the *cis*-regulatory code, which is often misregulated in heritable diseases and cancer. Innovations from UBC are providing new insights into complex gene regulatory networks. Professor Carl de Boer is pioneering functional assays on synthetic DNA libraries with much greater diversity and lower bias than the genome, while Professor Wyeth W. Wasserman is leveraging years of experience in gene regulation to create interpretable neural networks such as ExplaiNN, bridging the gap between “black-box” and “white-box” models. Both thoughtfully combine ML with biological and experimental constraints to produce robust, generalizable models. My in-depth experience in biophysically-interpretable ML has prepared me to hit the ground running and push forward on these and future endeavors.

GSAT is the best place for me to develop frameworks for creatively integrating domain knowledge, experimental techniques, and modeling strategies to disentangle biological systems, and will prepare me for postdocs where I can nurture interdisciplinary collaborations.

COLUMBIA UNIVERSITY IN THE CITY OF NEW YORK

NAME: Lucas de Araujo Nogueira Melo
SSN#: XXX-XX-2962
SCHOOL: FU FOUNDATN SCHL OF ENGINEERING & APPLIED SCIENCE:UGRAD

DEGREE(S) AWARDED: Bachelor of Science
DATE AWARDED: May 18, 2022
MAJOR: COMPUTER SCIENCE
HONORS: SUMMA CUM LAUDE

PROGRAM TITLE: COMPUTER SCIENCE

SUBJECT	COURSE NUMBER	TITLE	POINTS	GRADE
Fall 2018				
CHEM	UN 1403	GENERAL CHEMISTRY I-LECTURES	4.00	A+
CHEM	UN 1407	GENERAL CHEMISTRY I - REC	0.00	
ECON	UN 1105	PRINCIPLES OF ECONOMICS	4.00	A
ECON	UN 1155	PRINCIPLES OF ECONOMICS-DISC	0.00	
ENGL	CC 1010	UW: READINGS IN HUMAN RIGHTS	3.00	A
MATH	UN 1102	CALCULUS II	3.00	A
PHYS	UN 1401	INTRO TO MECHANICS & THERMO	3.00	A
PHYS	UN 1404	INTRO TO MECH & THERMO - REC	0.00	
			GPA	4.077

HONORS: DEANíS LIST

Spring 2019				
APMA	E 2000	MULTIVARIABLE CALC ENG/APP SCI	4.00	A
APMA	E 2001	MV CALC RECITATION	0.00	
CHEM	UN 1404	GENERAL CHEMISTRY II-LECTURES	4.00	A+
CHEM	UN 1406	GENERAL CHEMISTRY II-REC	0.00	
CHEM	UN 1500	GENERAL CHEMISTRY LABORAT	3.00	A+
CHEM	UN 1501	GENERAL CHEMISTRY LAB-LECTURE	0.00	
COMS	W 1004	INTRO-COMPUT SCI/PROG IN JAVA	3.00	A+
ENGI	E 1102	THE ART OF ENGINEERING	4.00	A
PHYS	UN 1402	INTRO TO ELEC/MAGNETSM & OPTCS	3.00	A+
PHYS	UN 1405	INTRO TO ELEC/MAGNETSM/OPT-REC	0.00	
			GPA	4.204

HONORS: DEANíS LIST

Fall 2019				
BIOL	UN 2005	INTRO BIO I-BIOCHEM,GEN,MOLEC	4.00	A
BIOL	UN 2015	INTRO BIO I: BIOCHEM,GEN,PHYS	0.00	
COMS	W 3134	DATA STRUCTURES IN JAVA	3.00	A+
COMS	W 3203	DISCRETE MATHEMATICS	3.00	A
ENGI	E 1006	INTRO TO COMP FOR ENG/APP SCI	3.00	A+
PHED	UN 1001	PHYSICAL EDUCATION ACTIVITIES	1.00	P
PHYS	UN 1494	INTRO TO EXPERIMENTAL PHY	3.00	A+
			GPA	4.185

HONORS: DEANíS LIST

Spring 2020				

Due to the COVID-19 pandemic, Mandatory Pass/Fail grading was in effect for all regular, full-term courses for the spring 2020 semester.

BIOL	UN 2006	INTRO BIO II:CELL BIO,DEV	4.00	P
BIOL	UN 2016	INTRO BIO II:CELL BIO,DEV	0.00	
BIOL	UN 2501	CONTEMPORARY BIOLOGY LAB	3.00	P
COMS	W 3157	ADVANCED PROGRAMMING	4.00	P
COMS	W 3251	COMPUTATIONAL LINEAR ALGE	3.00	P
COMS	W 4111	INTRODUCTION TO DATABASES	3.00	P

Summer 2020				
COMS	W 4701	ARTIFICIAL INTELLIGENCE	3.00	A+
RSRH	C 0001	FULL-TIME SUMMER RESEARCH	0.00	
			GPA	4.330

Fall 2020				
BIOL	UN 3022	DEVELOPMENTAL BIOLOGY	3.00	A
CBMF	W 4761	COMPUTATIONAL GENOMICS	3.00	A
CHEM	UN 2443	ORGANIC CHEMISTRY I-LECTU	4.00	A
CHEM	UN 2445	ORGANIC CHEMISTRY - REC	0.00	
CHEM	UN 2493	ORGANIC CHEM. LAB I TECHN	0.00	
CHEM	UN 2495	ORGANIC CHEM. LABORATORY	1.50	A
CSEE	W 3827	FUNDAMENTALS OF COMPUTER	3.00	A
PHED	UN 1001	PHYSICAL EDUCATION ACTIVITIES	1.00	P
			GPA	4.000

This official transcript was produced on OCTOBER 16, 2023.

SEAL OF COLUMBIA UNIVERSITY IN THE CITY OF NEW YORK

Barry S. Kane
Associate Vice President and University Registrar

TO VERIFY AUTHENTICITY OF DOCUMENT, THE BLUE STRIP BELOW CONTAINS HEAT SENSITIVE INK WHICH DISAPPEARS UPON TOUCH

SCHOOL: FU FOUNDATN SCHL OF ENGINEERING & APPLIED SCIENCE:UGRAD

NAME: Lucas de Araujo Nogueira Melo
SSN#: XXX-XX-2962

SUBJECT	COURSE NUMBER	TITLE	POINTS	GRADE
		Spring 2021		
BIOL	UN 3031	GENETICS	3.00	A
CHEM	UN 2444	ORGANIC CHEMSTRY II-LECTU	4.00	A-
CHEM	UN 2446	ORGANIC CHEMISTRY - REC	0.00	
CHEM	UN 2494	ORGANIC CHEM. LAB II SYNT	0.00	
CHEM	UN 2496	ORGANIC CHEM. LABORATORY	1.50	A
COMS	W 4705	NATURAL LANGUAGE PROCESSI	3.00	P
COMS	W 4771	MACHINE LEARNING	3.00	A
				GPA 3.885
		Summer 2021		
BCHM	UN 3511	BIOCHEM I-STRUCT/METABOLS	0.00	
BCHM	GU 4501	BIOCHEM I-STRUCTURE/METAB	3.00	A
HUMA	S 1123	MASTERPIECES OF WESTERN M	3.00	A+
				GPA 4.165
		Fall 2021		
BMEN	E 4110	BIOSTATISTICS FOR ENGINEE	4.00	A
COCI	CC 1101	CONTEMP WESTERN CIVILIZAT	4.00	A+
COMS	W 3261	COMPUTER SCIENCE THEORY	3.00	A+
COMS	W 4775	CAUSAL INFERENCE	3.00	A
PSYC	UN 1001	THE SCIENCE OF PSYCHOLOGY	3.00	A
				GPA 4.135

HONORS: DEANíS LIST

		Spring 2022		
BIOC	GU 4512	MOLECULAR BIOLOGY	3.00	A+
BIOL	GU 4002	MACROMOL STRUCTURE & INTE	4.00	A+
COCI	CC 1102	CONTEMP WESTRN CIVILIZATI	4.00	A+
COMS	E 4762	ML FOR FUNCTIONAL GENOMIC	3.00	A+
				GPA 4.330

REMARKS

Cumulative GPA: 4.134
15.00 Credits Transferred from College Bd: Advanced Placement

FINAL

SEAL OF COLUMBIA UNIVERSITY
IN THE CITY OF NEW YORK

Barry S. Kane
Associate Vice President and University Registrar

TO VERIFY AUTHENTICITY OF DOCUMENT, THE BLUE STRIP BELOW CONTAINS HEAT SENSITIVE INK WHICH DISAPPEARS UPON TOUCH

OFFICE OF THE UNIVERSITY REGISTRAR
STUDENT SERVICE CENTER
1140 AMSTERDAM AVENUE
205 KENT HALL, MAIL CODE 9202
NEW YORK, NEW YORK 10027
(212) 854-4400

SEAL OF COLUMBIA UNIVERSITY
IN THE CITY OF NEW YORK

Columbia College, Engineering and Applied Science, General Studies, Graduate School of Arts and Sciences, International and Public Affairs, Library Service, Human Nutrition, Nursing, Occupational Therapy, Physical Therapy, Professional Studies, Special Studies Program, Summer Session
A, **B**, **C**, **D**, **F** (excellent, good, fair, poor, failing). NOTE: Plus and minus signs and the grades of **P** (pass) and **HP** (high pass) are used in some schools. The grade of **D** is not used in Graduate Nursing, Occupational Therapy, and Physical Therapy.

American Language Program, Center for Psychoanalytic Training and Research, Journalism
P (pass), **F** (failing). Grades of **A**, **B**, **C**, **D**, **P** (pass), **F** (failing) — used for some offerings from the American Language Program Spring 2009 and thereafter.

Architecture
HP (high pass), **P** (pass), **LP** (low pass), **F** (failing), and **A**, **B**, **C**, **D**, **F** — used June 1991 and thereafter **P** (pass), **F** (failing) — used prior to June 1991.

Arts
P (pass), **LP** (low pass), **F** (fail). **H** (honors) used prior to June 2015.

Business
H (honors), **HP** (high pass), **P1** (pass), **LP** (low pass), **P** (unweighted pass), **F** (failing); plus (+) and minus (-) used for H, HP and P1 grades Summer 2010 and thereafter.

College of Physicians and Surgeons
H (honors), **HP** (high pass), **P** (pass), **F** (failing).

College of Dental Medicine
H (honors), **P** (pass), **F** (failing).

Law
A through **C** [plus (+) and minus (-) with **A** and **B** only], **CR** (credit - equivalent to passing). **F** (failing) is used beginning with the class which entered Fall 1994. Some offerings are graded by **HP** (high pass), **P** (pass), **LP** (low pass), **F** (failing). **W** (withdrawn) signifies that the student was permitted to drop a course, for which he or she had been officially registered, after the close of the Law School's official Change of Program (add/drop) period. It carries no connotation of quality of student performance, nor is it considered in the calculation of academic honors.
E (excellent), **VG** (very good), **G** (good), **P** (pass), **U** (unsatisfactory), **CR** (credit) used from 1970 through the class which entered in Fall 1993.

Any student in the Law School's Juris Doctor program may, at any time, request that he or she be graded on the basis of Credit-Fail. In such event, the student's performance in every offering is graded in accordance with the standards outlined in the school's bulletin, but recorded on the transcript as Credit-Fail. A student electing the Credit-Fail option may revoke it at any time prior to graduation and receive or request a copy of his or her transcript with grades recorded in accordance with the policy outlined in the school bulletin. In all cases, the transcript received or requested by the student shall show, on a cumulative basis, all of the grades of the student presented in single format – i.e., all grades shall be in accordance with those set forth in the school bulletin, or all grades shall be stated as Credit or Fail.

Public Health
A, **B**, **C**, **D**, **F** - used Summer 1985 and thereafter. **H** (honors), **P** (pass), **F** (failing) — used prior to Summer 1985.

Social Work
E (excellent), **VG** (very good), **G** (good), **MP** (minimum pass), **F** (failing).
A though **C** is used beginning with the class which entered Fall 1997. Plus signs used with **B** and **C** only, while minus signs are used with all letter grades. The grade of **P** (pass) is given only for select classes.

OTHER GRADES USED IN THE UNIVERSITY

AB = Excused absence from final examination.

AR = Administrative Referral awarded temporarily if a final grade cannot be determined without additional information.

AU = Audit (auditing division only).

CP = Credit Pending. Assigned in graduate courses which regularly involve research projects extending beyond the end of the term. Until such time as a passing or failing grade is assigned, satisfactory progress is implied.

F* = Course dropped unofficially.

IN = Work Incomplete.

MU = Make-Up. Student has the privilege of taking a second final examination.

R = For the Business School: Indicates satisfactory completion of courses taken as part of an exchange program and earns academic credit.

R = For Columbia College: The grade given for course taken for no academic credit, or notation given for internship.

R = For the Graduate School of Arts and Sciences: By prior agreement, only a portion of total course work completed. Program determines academic credit.

R = For the School of International and Public Affairs: The grade given for a course taken for no academic credit.

UW = Unofficial Withdrawal.

UW = For the College of Physicians and Surgeons: Indicates significant attempted coursework which the student does not have the opportunity to complete as listed due to required repetition or withdrawal.

W = Withdrew from course.

YC = Year Course. Assigned at the end of the first term of a year course. A single grade for the entire course is given upon completion of the second term. Until such time as a passing or failing grade is assigned, satisfactory progress is implied.

OTHER INFORMATION

NOTE: All students who cross-register into other schools of the University are graded in the **A**, **B**, **C**, **D**, **F** grading system regardless of the grading system of their own school, except in the schools of Arts (prior to Spring 1993) and in Journalism (prior to Autumn 1992), in which the grades of P (pass) and **F** (failing) were assigned. Notations at the end of a term provide documentation of the type of separation from the University.

% of A Effective fall 1996: Transcripts of Columbia College students show the percentage of grades in the **A** (**A+**, **A**, **A-**) range in all classes with at least 12 grades, the mark of **R** excluded. Calculations are taken at two points in time, three weeks after the last final examination of the term and three weeks after the last final of the next term. Once taken, the percentage is final even if grades change or if grades are submitted after the calculation. For additional information about the grading policy of the Faculty of Columbia College, consult the College Bulletin.

KEY TO COURSE LISTINGS

A course listing consists of an area, a capital letter(s) (denotes school bulletin) and the four digit course number (see below).

*The **capital letter** indicates the University school, division, or affiliate offering the course:*

A	Graduate School of Architecture, Planning, and Preservation
B	School of Business
BC	Barnard College
C	Columbia College
D	College of Dental Medicine
E	School of Engineering and Applied Science
F	School of General Studies
G	Graduate School of Arts and Sciences
H	Reid Hall (Paris)
J	Graduate School of Journalism
K	School of Library Services/Continuing Education (effective Fall 2002)
L	School of Law
M	College of Physicians and Surgeons, Institute of Human Nutrition, Program in Occupational Therapy, Program in Physical Therapy, Psychoanalytical Training and Research
N	School of Nursing
O	Other Universities or Affiliates/Auditing
P	School of Public Health
Q	Computer Technology/Applications
R	School of the Arts
S	Summer Session
T	School of Social Work
TA-TZ	Teachers College
U	School of International and Public Affairs
V	Interschool Course
W	Interfaculty Course
Y	Teachers College
Z	American Language Program

UNDER THE PROVISION OF THE FAMILY EDUCATION RIGHTS AND PRIVACY ACT OF 1974, THIS TRANSCRIPT MAY NOT BE RELEASED OR REVEALED TO A THIRD PARTY WITHOUT THE WRITTEN CONSENT OF THE STUDENT.

*The **first digit** of the course number indicates the level of the course, as follows:*

0	Course that cannot be credited toward any degree
1	Undergraduate course
3	Undergraduate course, advanced
4	Graduate course open to qualified undergraduates
5	Graduate course open to qualified undergraduates
6	Graduate course
7	Graduate course
8	Graduate course, advanced
9	Graduate research course or seminar

Note: Level Designations Prior to 1961:
1-99 Undergraduate courses
100-299 Lower division graduate courses
300-999 Upper division graduate courses

The term designations are as follows:
X=Autumn Term, Y=Spring Term, S=Summer Term
Notations at the end of a term provide documentation of the type of separation from the University.

THE ABOVE INFORMATION REFLECTS GRADING SYSTEMS IN USE SINCE SPRING 1982. THE CUMULATIVE INDEX, IF SHOWN, DOES NOT REFLECT COURSES TAKEN BEFORE SPRING
ALL TRANSCRIPTS ISSUED FROM THIS OFFICE ARE OFFICIAL DOCUMENTS. TRANSCRIPTS ARE PRINTED ON TAMPER-PROOF PAPER, ELIMINATING THE NEED FOR SIGNATURES AND STAMPS ON THE BACK OF ENVE
CERTIFICATION PURPOSES, A REPRODUCED COPY OF THIS RECORD SHALL NOT BE VALID. THE HEAT-SENSITIVE STRIP, LOCATED ON THE BOTTOM EDGE OF THE FACE OF THE TRANSCRIPT, WILL CHANGE FRO
CLEAR WHEN HEAT OR PRESSURE IS APPLIED. A BLUE SIGNATURE ALSO ACCOMPANIES THE UNIVERSITY SEAL ON THE FACE OF THE TRANSCRIPT.

COLUMBIA UNIVERSITY IN THE CITY OF NEW YORK

NAME: Lucas de Araujo Nogueira Melo
SSN#: XXX-XX-2962
SCHOOL: FU FOUNDATN SCHL OF ENGINEERING & APPLIED SCIENCE:UGRAD

DEGREE(S) AWARDED: Bachelor of Science
DATE AWARDED: May 18, 2022
MAJOR: COMPUTER SCIENCE
HONORS: SUMMA CUM LAUDE

PROGRAM TITLE: COMPUTER SCIENCE

SUBJECT	COURSE NUMBER	TITLE	POINTS	GRADE
Fall 2018				
CHEM	UN 1403	GENERAL CHEMISTRY I-LECTURES	4.00	A+
CHEM	UN 1407	GENERAL CHEMISTRY I - REC	0.00	
ECON	UN 1105	PRINCIPLES OF ECONOMICS	4.00	A
ECON	UN 1155	PRINCIPLES OF ECONOMICS-DISC	0.00	
ENGL	CC 1010	UW: READINGS IN HUMAN RIGHTS	3.00	A
MATH	UN 1102	CALCULUS II	3.00	A
PHYS	UN 1401	INTRO TO MECHANICS & THERMO	3.00	A
PHYS	UN 1404	INTRO TO MECH & THERMO - REC	0.00	
			GPA	4.077

HONORS: DEANíS LIST

Spring 2019				
APMA	E 2000	MULTIVARIABLE CALC ENG/APP SCI	4.00	A
APMA	E 2001	MV CALC RECITATION	0.00	
CHEM	UN 1404	GENERAL CHEMISTRY II-LECTURES	4.00	A+
CHEM	UN 1406	GENERAL CHEMISTRY II-REC	0.00	
CHEM	UN 1500	GENERAL CHEMISTRY LABORAT	3.00	A+
CHEM	UN 1501	GENERAL CHEMISTRY LAB-LECTURE	0.00	
COMS	W 1004	INTRO-COMPUT SCI/PROG IN JAVA	3.00	A+
ENGI	E 1102	THE ART OF ENGINEERING	4.00	A
PHYS	UN 1402	INTRO TO ELEC/MAGNETSM & OPTCS	3.00	A+
PHYS	UN 1405	INTRO TO ELEC/MAGNETSM/OPT-REC	0.00	
			GPA	4.204

HONORS: DEANíS LIST

Fall 2019				
BIOL	UN 2005	INTRO BIO I-BIOCHEM,GEN,MOLEC	4.00	A
BIOL	UN 2015	INTRO BIO I: BIOCHEM,GEN,PHYS	0.00	
COMS	W 3134	DATA STRUCTURES IN JAVA	3.00	A+
COMS	W 3203	DISCRETE MATHEMATICS	3.00	A
ENGI	E 1006	INTRO TO COMP FOR ENG/APP SCI	3.00	A+
PHED	UN 1001	PHYSICAL EDUCATION ACTIVITIES	1.00	P
PHYS	UN 1494	INTRO TO EXPERIMENTAL PHY	3.00	A+
			GPA	4.185

HONORS: DEANíS LIST

Spring 2020				

Due to the COVID-19 pandemic, Mandatory Pass/Fail grading was in effect for all regular, full-term courses for the spring 2020 semester.

BIOL	UN 2006	INTRO BIO II:CELL BIO,DEV	4.00	P
BIOL	UN 2016	INTRO BIO II:CELL BIO,DEV	0.00	
BIOL	UN 2501	CONTEMPORARY BIOLOGY LAB	3.00	P
COMS	W 3157	ADVANCED PROGRAMMING	4.00	P
COMS	W 3251	COMPUTATIONAL LINEAR ALGE	3.00	P
COMS	W 4111	INTRODUCTION TO DATABASES	3.00	P

Summer 2020				
COMS	W 4701	ARTIFICIAL INTELLIGENCE	3.00	A+
RSRH	C 0001	FULL-TIME SUMMER RESEARCH	0.00	
			GPA	4.330

Fall 2020				
BIOL	UN 3022	DEVELOPMENTAL BIOLOGY	3.00	A
CBMF	W 4761	COMPUTATIONAL GENOMICS	3.00	A
CHEM	UN 2443	ORGANIC CHEMISTRY I-LECTU	4.00	A
CHEM	UN 2445	ORGANIC CHEMISTRY - REC	0.00	
CHEM	UN 2493	ORGANIC CHEM. LAB I TECHN	0.00	
CHEM	UN 2495	ORGANIC CHEM. LABORATORY	1.50	A
CSEE	W 3827	FUNDAMENTALS OF COMPUTER	3.00	A
PHED	UN 1001	PHYSICAL EDUCATION ACTIVITIES	1.00	P
			GPA	4.000

This official transcript was produced on OCTOBER 16, 2023.

SEAL OF COLUMBIA UNIVERSITY IN THE CITY OF NEW YORK

Barry S. Kane
Associate Vice President and University Registrar

TO VERIFY AUTHENTICITY OF DOCUMENT, THE BLUE STRIP BELOW CONTAINS HEAT SENSITIVE INK WHICH DISAPPEARS UPON TOUCH

SCHOOL: FU FOUNDATN SCHL OF ENGINEERING & APPLIED SCIENCE:UGRAD

NAME: Lucas de Araujo Nogueira Melo
SSN#: XXX-XX-2962

SUBJECT	COURSE NUMBER	TITLE	POINTS	GRADE
		Spring 2021		
BIOL	UN 3031	GENETICS	3.00	A
CHEM	UN 2444	ORGANIC CHEMSTRY II-LECTU	4.00	A-
CHEM	UN 2446	ORGANIC CHEMISTRY - REC	0.00	
CHEM	UN 2494	ORGANIC CHEM. LAB II SYNT	0.00	
CHEM	UN 2496	ORGANIC CHEM. LABORATORY	1.50	A
COMS	W 4705	NATURAL LANGUAGE PROCESSI	3.00	P
COMS	W 4771	MACHINE LEARNING	3.00	A
				GPA 3.885
		Summer 2021		
BCHM	UN 3511	BIOCHEM I-STRUCT/METABOLS	0.00	
BCHM	GU 4501	BIOCHEM I-STRUCTURE/METAB	3.00	A
HUMA	S 1123	MASTERPIECES OF WESTERN M	3.00	A+
				GPA 4.165
		Fall 2021		
BMEN	E 4110	BIOSTATISTICS FOR ENGINEE	4.00	A
COCI	CC 1101	CONTEMP WESTERN CIVILIZAT	4.00	A+
COMS	W 3261	COMPUTER SCIENCE THEORY	3.00	A+
COMS	W 4775	CAUSAL INFERENCE	3.00	A
PSYC	UN 1001	THE SCIENCE OF PSYCHOLOGY	3.00	A
				GPA 4.135

HONORS: DEANíS LIST

		Spring 2022		
BIOC	GU 4512	MOLECULAR BIOLOGY	3.00	A+
BIOL	GU 4002	MACROMOL STRUCTURE & INTE	4.00	A+
COCI	CC 1102	CONTEMP WESTRN CIVILIZATI	4.00	A+
COMS	E 4762	ML FOR FUNCTIONAL GENOMIC	3.00	A+
				GPA 4.330

REMARKS

Cumulative GPA: 4.134
15.00 Credits Transferred from College Bd: Advanced Placement

FINAL

SEAL OF COLUMBIA UNIVERSITY
IN THE CITY OF NEW YORK

Barry S. Kane
Associate Vice President and University Registrar

TO VERIFY AUTHENTICITY OF DOCUMENT, THE BLUE STRIP BELOW CONTAINS HEAT SENSITIVE INK WHICH DISAPPEARS UPON TOUCH

OFFICE OF THE UNIVERSITY REGISTRAR
STUDENT SERVICE CENTER
1140 AMSTERDAM AVENUE
205 KENT HALL, MAIL CODE 9202
NEW YORK, NEW YORK 10027
(212) 854-4400

SEAL OF COLUMBIA UNIVERSITY
IN THE CITY OF NEW YORK

Columbia College, Engineering and Applied Science, General Studies, Graduate School of Arts and Sciences, International and Public Affairs, Library Service, Human Nutrition, Nursing, Occupational Therapy, Physical Therapy, Professional Studies, Special Studies Program, Summer Session
A, **B**, **C**, **D**, **F** (excellent, good, fair, poor, failing). NOTE: Plus and minus signs and the grades of **P** (pass) and **HP** (high pass) are used in some schools. The grade of **D** is not used in Graduate Nursing, Occupational Therapy, and Physical Therapy.

American Language Program, Center for Psychoanalytic Training and Research, Journalism
P (pass), **F** (failing). Grades of **A**, **B**, **C**, **D**, **P** (pass), **F** (failing) — used for some offerings from the American Language Program Spring 2009 and thereafter.

Architecture
HP (high pass), **P** (pass), **LP** (low pass), **F** (failing), and **A**, **B**, **C**, **D**, **F** — used June 1991 and thereafter **P** (pass), **F** (failing) — used prior to June 1991.

Arts
P (pass), **LP** (low pass), **F** (fail). **H** (honors) used prior to June 2015.

Business
H (honors), **HP** (high pass), **P1** (pass), **LP** (low pass), **P** (unweighted pass), **F** (failing); plus (+) and minus (-) used for H, HP and P1 grades Summer 2010 and thereafter.

College of Physicians and Surgeons
H (honors), **HP** (high pass), **P** (pass), **F** (failing).

College of Dental Medicine
H (honors), **P** (pass), **F** (failing).

Law
A through **C** [plus (+) and minus (-) with **A** and **B** only], **CR** (credit - equivalent to passing). **F** (failing) is used beginning with the class which entered Fall 1994. Some offerings are graded by **HP** (high pass), **P** (pass), **LP** (low pass), **F** (failing). **W** (withdrawn) signifies that the student was permitted to drop a course, for which he or she had been officially registered, after the close of the Law School's official Change of Program (add/drop) period. It carries no connotation of quality of student performance, nor is it considered in the calculation of academic honors.
E (excellent), **VG** (very good), **G** (good), **P** (pass), **U** (unsatisfactory), **CR** (credit) used from 1970 through the class which entered in Fall 1993.

Any student in the Law School's Juris Doctor program may, at any time, request that he or she be graded on the basis of Credit-Fail. In such event, the student's performance in every offering is graded in accordance with the standards outlined in the school's bulletin, but recorded on the transcript as Credit-Fail. A student electing the Credit-Fail option may revoke it at any time prior to graduation and receive or request a copy of his or her transcript with grades recorded in accordance with the policy outlined in the school bulletin. In all cases, the transcript received or requested by the student shall show, on a cumulative basis, all of the grades of the student presented in single format – i.e., all grades shall be in accordance with those set forth in the school bulletin, or all grades shall be stated as Credit or Fail.

Public Health
A, **B**, **C**, **D**, **F** - used Summer 1985 and thereafter. **H** (honors), **P** (pass), **F** (failing) — used prior to Summer 1985.

Social Work
E (excellent), **VG** (very good), **G** (good), **MP** (minimum pass), **F** (failing).
A though **C** is used beginning with the class which entered Fall 1997. Plus signs used with **B** and **C** only, while minus signs are used with all letter grades. The grade of **P** (pass) is given only for select classes.

OTHER GRADES USED IN THE UNIVERSITY

AB = Excused absence from final examination.

AR = Administrative Referral awarded temporarily if a final grade cannot be determined without additional information.

AU = Audit (auditing division only).

CP = Credit Pending. Assigned in graduate courses which regularly involve research projects extending beyond the end of the term. Until such time as a passing or failing grade is assigned, satisfactory progress is implied.

F* = Course dropped unofficially.

IN = Work Incomplete.

MU = Make-Up. Student has the privilege of taking a second final examination.

R = For the Business School: Indicates satisfactory completion of courses taken as part of an exchange program and earns academic credit.

R = For Columbia College: The grade given for course taken for no academic credit, or notation given for internship.

R = For the Graduate School of Arts and Sciences: By prior agreement, only a portion of total course work completed. Program determines academic credit.

R = For the School of International and Public Affairs: The grade given for a course taken for no academic credit.

UW = Unofficial Withdrawal.

UW = For the College of Physicians and Surgeons: Indicates significant attempted coursework which the student does not have the opportunity to complete as listed due to required repetition or withdrawal.

W = Withdrew from course.

YC = Year Course. Assigned at the end of the first term of a year course. A single grade for the entire course is given upon completion of the second term. Until such time as a passing or failing grade is assigned, satisfactory progress is implied.

OTHER INFORMATION

NOTE: All students who cross-register into other schools of the University are graded in the **A**, **B**, **C**, **D**, **F** grading system regardless of the grading system of their own school, except in the schools of Arts (prior to Spring 1993) and in Journalism (prior to Autumn 1992), in which the grades of P (pass) and **F** (failing) were assigned. Notations at the end of a term provide documentation of the type of separation from the University.

% of A Effective fall 1996: Transcripts of Columbia College students show the percentage of grades in the **A** (**A+**, **A**, **A-**) range in all classes with at least 12 grades, the mark of **R** excluded. Calculations are taken at two points in time, three weeks after the last final examination of the term and three weeks after the last final of the next term. Once taken, the percentage is final even if grades change or if grades are submitted after the calculation. For additional information about the grading policy of the Faculty of Columbia College, consult the College Bulletin.

KEY TO COURSE LISTINGS

A course listing consists of an area, a capital letter(s) (denotes school bulletin) and the four digit course number (see below).

*The **capital letter** indicates the University school, division, or affiliate offering the course:*

A	Graduate School of Architecture, Planning, and Preservation
B	School of Business
BC	Barnard College
C	Columbia College
D	College of Dental Medicine
E	School of Engineering and Applied Science
F	School of General Studies
G	Graduate School of Arts and Sciences
H	Reid Hall (Paris)
J	Graduate School of Journalism
K	School of Library Services/Continuing Education (effective Fall 2002)
L	School of Law
M	College of Physicians and Surgeons, Institute of Human Nutrition, Program in Occupational Therapy, Program in Physical Therapy, Psychoanalytical Training and Research
N	School of Nursing
O	Other Universities or Affiliates/Auditing
P	School of Public Health
Q	Computer Technology/Applications
R	School of the Arts
S	Summer Session
T	School of Social Work
TA-TZ	Teachers College
U	School of International and Public Affairs
V	Interschool Course
W	Interfaculty Course
Y	Teachers College
Z	American Language Program

UNDER THE PROVISION OF THE FAMILY EDUCATION RIGHTS AND PRIVACY ACT OF 1974, THIS TRANSCRIPT MAY NOT BE RELEASED OR REVEALED TO A THIRD PARTY WITHOUT THE WRITTEN CONSENT OF THE STUDENT.

*The **first digit** of the course number indicates the level of the course, as follows:*

0	Course that cannot be credited toward any degree
1	Undergraduate course
3	Undergraduate course, advanced
4	Graduate course open to qualified undergraduates
5	Graduate course open to qualified undergraduates
6	Graduate course
7	Graduate course
8	Graduate course, advanced
9	Graduate research course or seminar

Note: Level Designations Prior to 1961:
1-99 Undergraduate courses
100-299 Lower division graduate courses
300-999 Upper division graduate courses

The term designations are as follows:
X=Autumn Term, Y=Spring Term, S=Summer Term
Notations at the end of a term provide documentation of the type of separation from the University.

THE ABOVE INFORMATION REFLECTS GRADING SYSTEMS IN USE SINCE SPRING 1982. THE CUMULATIVE INDEX, IF SHOWN, DOES NOT REFLECT COURSES TAKEN BEFORE SPRING
ALL TRANSCRIPTS ISSUED FROM THIS OFFICE ARE OFFICIAL DOCUMENTS. TRANSCRIPTS ARE PRINTED ON TAMPER-PROOF PAPER, ELIMINATING THE NEED FOR SIGNATURES AND STAMPS ON THE BACK OF ENVE
CERTIFICATION PURPOSES, A REPRODUCED COPY OF THIS RECORD SHALL NOT BE VALID. THE HEAT-SENSITIVE STRIP, LOCATED ON THE BOTTOM EDGE OF THE FACE OF THE TRANSCRIPT, WILL CHANGE FRO
CLEAR WHEN HEAT OR PRESSURE IS APPLIED. A BLUE SIGNATURE ALSO ACCOMPANIES THE UNIVERSITY SEAL ON THE FACE OF THE TRANSCRIPT.

UNIVERSITY OF CALIFORNIA MERCED

BERKELEY • DAVIS • IRVINE • LOS ANGELES • MERCED • RIVERSIDE • SAN DIEGO • SAN FRANCISCO

SANTA BARBARA • SANTA CRUZ

UNIVERSITY OF CALIFORNIA MERCED
5200 N. LAKE ROAD | MERCED, CA 95343
TEL: 209.228.4400

Asst. Prof. H. Tomas Rube
Department of Applied Mathematics
trube@ucmerced.edu
650-521-4565 (cell phone)

November 30th, 2023

Recommendation Letter for Lucas Melo

It is my great pleasure to recommend Lucas Melo for the Genome Science and Technology PhD program at the University of British Columbia. I am an Assistant Professor in the Department of Applied Mathematics at University of California, Merced. I first came to know Lucas in 2019 when he joined Harmen Bussemaker's lab at Columbia University as an undergraduate researcher. At the time, I had just moved from Columbia (where I had been a postdoc in Harmen's lab) to University of California Merced, and I was in the progress of finishing up several projects. Lucas quickly became involved in these, and I have had the great pleasure to see his work and his growth over the past four years.

In the first project Lucas worked on, we developed an algorithm that quantifies how transcription factors (TFs) recognize DNA sequences in terms of biophysically rigorous quantitates, including the energetic impact of base-base interactions, TF-TF interactions, and chemical DNA modifications. Because Lucas was only a sophomore at the time, we decided that a good warm-up project would be for him to implement some code for plotting these energetic parameters in R. When assigning the problem, I did not expect him to succeed, both because the model (which use a combination of biophysical modeling and statistical learning) are challenging to grasp, and because other undergraduate students had previously failed. I was therefore pleasantly surprised when Lucas soon returned with fully a functional code. This code was integrated into the production pipeline and used to generate main figures for the final Nature Biotechnology paper. Lucas next helped me apply the algorithm (which had been developed to quantify *in vitro* TF binding) to learn well-calibrated TF binding models from ChIP-seq data, and the results were included in the abovementioned paper.

In parallel with our work on TF-DNA interactions, we realized that our algorithm also can be used to characterize protein-protein interactions. Chaitanya Rastogi (a postdoc in the Bussemaker lab) spearheaded an effort to build a startup company around the algorithm to accelerate the development of biologics. As a part of this effort, Lucas studied several high-throughput sequencing screens and developed computational strategies for estimating the energetics of the interactions. For example, he developed a method that integrates multi-time point T-cell receptor data to infer the sequence grammar of antigen recognition. Moreover, he re-implemented the main algorithm in using the PyTorch framework. In doing so, he introduced a number of generalizations that allows the algorithm to be applied to more sophisticated selection experiments. For example, in an ongoing project we are quantifying how peptide binding domains

UNIVERSITY OF CALIFORNIA MERCED

BERKELEY • DAVIS • IRVINE • LOS ANGELES • MERCED • RIVERSIDE • SAN DIEGO • SAN FRANCISCO

SANTA BARBARA • SANTA CRUZ

UNIVERSITY OF CALIFORNIA MERCED
5200 N. LAKE ROAD | MERCED, CA 95343
TEL: 209.228.4400

recognize phosphorylated sequences using a high-throughput library screen. Using his generalized algorithm, Lucas developed a highly non-trivial analysis that jointly analyzes the main sequencing experiment and control experiments, thus correcting for biases. Looking forward, I anticipate that Lucas will publish a first-author paper describing the generalized algorithm and his continued work on ChIP-seq analysis, and a near-first-author paper focused on the peptide-binding domains. Looking back, it has been incredibly rewarding to see Lucas' growth, first studying the underlying theory, then mastering a very complex algorithm, and finally developing important generalizations that I myself did not anticipate.

On a personal level, Lucas is a humble, no-drama person that is pleasant to work with. By now, Lucas has spent significant time conducting research and he is developing into a mature researcher. For example, when encountering problems interpreting the peptide-binding data, he often explores different solutions by his own initiative and returns with a not only a problem description but also a working solution. Overall, I believe this maturity will put him at a great advantage as he enters graduate school.

In closing, I am giving Lucas my strongest possible recommendation. I would accept him into my research group without a moment of hesitation, and I am most confident that he would excel in your PhD program. *Please do not hesitate* to contact me if any further information or assistance is needed.

Sincerely,

H. Tomas Rube,
Assistant Professor, Applied Mathematics,
School of Natural Sciences,
University of California, Merced

COLUMBIA UNIVERSITY
IN THE CITY OF NEW YORK

DEPARTMENT OF BIOLOGICAL SCIENCES

October 30, 2023

Dear Colleagues,

It is a tremendous pleasure to recommend that **Lucas Melo** be considered for admission to your PhD program, and I do so in the strongest possible terms. I can hardly imagine anybody who would be more deserving of admission to even the most selective PhD programs in the country.

Lucas started working for my lab part-time in the fall of 2019. His first assignment was to fix a tricky Java compatibility issue that caused our most widely used R package to be dropped from Bioconductor. He quickly got to work on this, and without requiring any detailed direction, found a solution to the problem. During the remainder of the academic year, alongside his heavy course load, he continued to spend regular time chatting with the people in my lab and becoming familiar with their projects.

For summer 2020, Lucas originally had an internship lined up in the SIBMI program at Harvard, but that got cancelled due to COVID, as did the Cold Spring Harbor Laboratory Undergraduate Research Program which he had also been admitted to. Luckily for me and my lab, Lucas asked whether he could do a (remote) summer project in my lab instead. He was officially admitted to our department's summer research (SURF) program, and suddenly had an entire summer ahead of him to work on projects with me.

The timing was perfect: Two brilliant postdocs in my lab – Dr. Tomas Rube, who is now an Assistant Professor at UC Merced, and Dr. Chaitanya Rastogi, a former PhD student in my lab with whom I recently co-founded a startup company – had just developed a new biophysically interpretable machine learning approach named *ProBound*. Its original purpose was to build accurate models of protein-DNA recognition from high-throughput *in vitro* transcription factor (TF) binding data (HT-SELEX), but Lucas started exploring whether it could also be used to analyze *in vivo* TF occupancy (ChIP-seq) data. Again with only high-level guidance and zero hand-holding, Lucas quickly made himself familiar with the complex java codebase of *ProBound*. Soon he had some intriguing results that showed that it was possible to build DNA binding specificity models ("motifs") from ChIP-seq data in a way that completely avoids the peak calling step used by all existing motif discovery approaches. Not only did this analysis become a key component of the paper about *ProBound* that we published in Nature Biotechnology last year (Rube et al., 2022), on which Lucas deservedly is a co-author; it is also the foundation of ongoing work in my lab, and the starting point for a new NIH grant proposal that I am currently developing.

While this achievement by itself would already have been impressive for an undergraduate summer project, Lucas also came up with an entirely new algorithm that same summer: I had become interested in understanding the evolution of immune repertoires in response to infection or immunization. After I had pointed Lucas to some data sets, he implemented an algorithm that could infer how clonal expansion of B or T cells depends on the sequence of the CDR sequence. At a technical level, this required the on-the-fly numerical solution of coupled partial differential equations in combination with maximum-likelihood fitting of biophysical binding energy parameters. Lucas figured out more or less completely by himself where to find the relevant background papers, which equations to solve, how to write the code to solve them, and how to interpret the results. I was blown away when he first presented this in our internal group meeting.

SHERMAN FAIRCHILD CENTER FOR LIFE SCIENCES
1212 Amsterdam Avenue New York, NY 10027

Another key example of Lucas' creativity and technical fearlessness is the following: When he joined my lab for two years as a Research Assistant after his graduation to further developed and apply *ProBound* to generate proof-of-concept for our startup project, he casually told Chaitanya and me that as a side-project, to make his life easier, he was reimplementing *ProBound* – which took multiple years to develop – from scratch (!) using the Pytorch framework, as he had found modifying the complex Java code cumbersome. Through a series of careful regression tests, Lucas showed that we were still getting exactly the same results as before. However, he suddenly could move much more quickly on expanding the functionality of ProBound. After little more than a year working as a full-time RA in my lab, and alongside countless other analyses he has performed for our startup project, Lucas has already developed news ways to understand allele-specific TF binding, including an entirely new algorithm that builds on the infrastructure of his Pytorch *ProBound* software to infer DNA binding specificity models directly from allele-aware ChIP-seq data. We are currently writing up a manuscript describing these results, on which Lucas will be shared first author, and which we are hoping to submit by early December.

I have had the privilege to mentor several exceptional undergraduate students during my two decades at Columbia, but Lucas has been the most impressive of them all. In addition to what I already outlined above, he is actively exploring complex and novel featurization methods very different from currently popular concepts such as transformers. He deeply understands the biological questions we are trying to solve, and has been exposed to our lab's way of thinking enough to come up with truly new contributions that make great sense to us once he tells us about it. Only Brian Trippe, who also worked in my lab as an undergraduate, and then went on to do a successful PhD at MIT and a short postdoc with David Baker, and who is now applying for faculty positions, is truly in the same rare category in my opinion.

Lucas is a gentle person, and was a bit reticent when he first joined my lab, but has grown tremendously. These days, when I walk into the lab in the morning, I often see Lucas having an impromptu meeting with another lab member (usually to help them rather than the other way around). Also, I routinely rely on him to lead the discussion of new results when we meet with our various external collaborators; there is no need to rehearse, or iterate on the slides beforehand. Lucas is also a confident scientific sparring partner for me and everybody else in the lab, with a great sense of scientific integrity.

In conclusion, I wholeheartedly recommend that Lucas Melo be admitted to your PhD program. In many regards, Lucas has been functioning in my lab as an advanced graduate student and in terms of technical independence even as a postdoc. I expect to him to do spectacular things during his PhD, and end up as a faculty member at a top-tier university.

With best wishes,

Harmen J. Bussemaker. Ph.D.
Professor of Biological Science and Systems Biology

COLUMBIA UNIVERSITY
IN THE CITY OF NEW YORK

NEW YORK GENOME CENTER®

Department of Computer Science
Columbia University, 500 W 120th St, New York, NY 10027
23 October 2023

To the PhD Admissions Committee,

It is my great pleasure to give Lucas Melo my extremely strong recommendation for admission to your PhD program. I would first like to introduce myself. I did my PhD studies developing Bayesian nonparametric models with Zoubin Ghahramani at the University of Cambridge and variational inference methods with Tom Minka and John Winn at Microsoft Research. During my postdoc at Stanford with Daphne Koller, Jonathan Pritchard and Sylvia Plevritis, I used machine learning methods to analyze large-scale genomics datasets to understand genetic effects on gene expression and alternative splicing, an important cellular process by which non-coding "junk" regions of genes are removed after transcription. I started my lab in January 2019, jointly between the Columbia University Department of Computer Science and the New York Genome Center (NYGC). We use computational and experimental approaches to investigate how genetic differences between individuals impact molecular phenotypes and how the resulting changes influence disease risk. I hold additional affiliations with the Data Science Institute and Systems Biology.

I came to know Lucas during his senior year through my course, COMS W4762 Machine Learning for Functional Genomics, which is offered for graduate and advanced undergraduate students. Among the nearly 100 students who take my course each year over the past several years, I would place Lucas among the brightest and most promising. His unique experiences in biophysically-informed machine learning address an important and growing question in the field, and he has the skills and potential to succeed in graduate school and become a leading researcher.

Lucas stood out early in the semester by reaching the top of the leaderboard in the course's Kaggle competition on predicting genomics data. I invited him to give an oral presentation during lecture to discuss his model, which he delivered with great clarity, conciseness, and poise. While most students who succeed at the competition use the latest techniques in deep learning to improve the performance of their models, Lucas' approach was unique. He carefully explained to a mixed audience of computational and biological backgrounds how he used principles from molecular biophysics to inform the model enhancements he designed. These insights evince the careful consideration with which he approaches every project, as well as the depth of his understanding in both biology and computational science.

In the project component of the course, Lucas was again unique among my students in the quality and originality of his proposal. In a short period of time, he developed an impressively thorough research project leveraging a dynamical model of clonal expansion in longitudinal immune sequencing to train a deep learning architecture of antibody binding. While he hit some initial stumbling blocks in dealing with a dataset which neither he nor I had previous experience in, Lucas demonstrated his resourcefulness by adapting the project to focus on the interpretability and validation of his model. The novelty of his approach was refreshing to see, going well beyond a typical project-based course submission. Seeing his enthusiasm for his project, I recommended that he publish it, and it was accepted as a poster to the 2022 Machine Learning for Computational

COLUMBIA UNIVERSITY
IN THE CITY OF NEW YORK

NEW YORK GENOME CENTER®

Biology conference, which I co-organize. Reviewers agreed on the potential of his methodology, writing that "*Integrating a biophysical model and a deep neural network is very interesting and a good example of physics-informed machine learning applications*."

Since completing his undergraduate degree Lucas has been working with my colleague Harmen Bussemaker who I believe is also providing a reference letter. Lucas contributed core computational components the Bussemaker lab's recent PROBOUND work, published in Nature Biotechnology. It is a testament to Lucas' knowledge, understanding and ability that he made such an important contribution even given his junior position in the group.

That Lucas has an excellent academic record goes without saying. Upon graduation, he was selected by the department for both the Behring Foundation Award and the Computer Science Scholarship Award in recognition of his academic achievements and leadership potential. These prestigious awards from a large and competitive department are a testament to how his capabilities rank well above that of his peers. Lucas is an exceptionally thoughtful and dedicated student, and I give him my highest endorsement.

Lucas is in the top 2% of undergraduate students I've interacted with at Cambridge, Stanford and Columbia. He'd rank among the best PhD students I've known, even given the early stage of his studies. I give him my strong recommendation.

Yours sincerely,

David A Knowles, PhD. {Computer Science, Systems Biology, Data Science Institute} @ Columbia University. Core Faculty Member @ New York Genome Center.