The Fannie and John Hertz Foundation
Application for Graduate Fellowship

To apply for a Hertz Foundation Graduate Fellowship, you must provide us with all of the information requested. Doing so will likely require at least 15 but no more than 30 minutes. Please read the instructions carefully before filling in each box. It is important that you format your answers according to the instructions.

Please fill in your complete name:

Mr./Ms./Mrs.	First name	Middle name	Last name	Name adjunct
Ms.	Olga	Borisovna	Botvinnik	

Enter your date of birth using the format **mm/dd/yyyy,** e.g. **June 10, 1983** would be **06/10/1983:**

09 /12 /1988
Date of birth

Enter the name of the country where you were born:

Country of birth Russia

If you are you a US citizen, print US in the Citizenship box.
If you are a Permanent Resident, print PR.

US
Citizenship?

If you are neither a US citizen nor a Permanent Resident of the US, you are not eligible to apply for a Hertz Fellowship. (If you are not a US citizen, we will ask to see proof of Application for Citizenship or your Permanent Resident Card when you are interviewed).

Current School/Work Information

If you are currently attending school, enter that school's name in the "Present school" box. If you are currently out of school, put your current occupation in the "Currently out of school" box and leave the "Present school" box blank. In the City and State boxes, enter the city and state of your current location.

Univ Calif San Diego
Present school

Currently out of school

Please enter the location, (city and state only), of your present school in the next box.

La Jolla CA
Present school location (city, state)

What is your current status? Graduate
Current Status

Undergraduate Institution

If you are currently a Graduate student, name the school that awarded your undergraduate degree. If you are currently an Undergraduate, enter your current school.

Mass Institute Technology

UG School

UG School (if not listed)

Fall 2014 - Spring 2015 will be your first year as a graduate student.

What is the Academic Major (field of study) in which your undergraduate degree was awarded, or, if you are an undergraduate, the field in which you are currently enrolled? Academic Major means: physics, chemistry, mathematics, mechanical engineering, etc.

Mathematics and Biological Engineerin

Undergraduate Academic Major

Overall Undergraduate GPA (4.0 = A) 3.20

UG GPA

What is your undergraduate GPA in your major subject? 3.00

UG GPA Major

If your undergraduate university does not assign grades, please enter "not available." Your scores in the GRE Aptitude and Advanced Subject Tests will be especially important to the foundation in evaluating your application.

If you are currently a graduate student, what is your overall graduate GPA?

Grad GPA

Graduate Institution

If you know the educational institution at which you currently propose to hold your Fellowship, please name it. Otherwise select, “Not yet known.” If the school you select is not currently included in our list of Currently Participating Institutions (page 6 of the Instructions), please understand the process by which that school will be approved.

It is not necessary for you to have been accepted at this Institution at the time you submit your application. While a definite school choice will be required at the time of Fellowship acceptance, this choice may subsequently be changed by submitting a written request to the Foundation.

University of California, San Diego, La Jolla, California

School Choice

School Choice (if not tenable)

In the next box, please indicate what your proposed General Field of Study will be at the school selected above. Please refer to the instructions for a discussion of Hertz-tenable fields of study.

Quantitative Biology

General Field

Please note that the Foundation does not support students seeking professional degrees other than the PhD, (e.g., MD, LLB, MBA, etc. are not supported).

In the Specialty Area box, please indicate what specialty area you intend to study, (e.g. condensed matter physics, molecular biology-DNA analysis, electrical engineering). If you do not know what Specialty Area you'll be studying, enter "Unknown."

Bioinformatics

Specialty

Graduate Institution (continued)

Quantitative Biology
General field

Bioinformatics
Specialty area

Please list other graduate schools that you are seriously considering attending (in addition to your choice listed above):

Other Potential School Choices

Toward what **ultimate** degree will you be working while in Graduate School? PhD
Degree sought

During what calendar year do you expect to complete your Ph.D. studies? 2017
Completion date

Please list other multi-year Fellowships you have applied for, currently hold, or are planning to apply for:

National Defense Science and Engineering Graduate Fellowship

Other Multi-Year Fellowships

Preparation for Graduate Work in the Applied Physical Sciences

You should have had university-level courses in physics, mathematics and chemistry beyond the introductory level. Please indicate below the number of upper division courses in each category that you have taken already or plan to take as an undergraduate. Enter the numbers as decimals, with a year-long course counting as 1.0, a semester course as 0.5, and a quarter course as 0.33. Please count research courses as upper division courses.

Upper-Division Course Years

Physics	
Chemistry	1.5
Engineering (ME, ChE, EE, CE)	4.0
Mathematics	3.0
Computer Science	1.0
Earth Science	
Modern Biology	4.0

Your Location

During the application process, it is important for us to be able to reach you on short notice. We may need to arrange a technical interview, or we may have a question about your application. Please give careful thought to this information. Further along on this form you will have the opportunity to give several alternate telephone numbers and addresses where you might be reached during the next 6 months. If any of this information changes between the time you submit this application and **March 15, 2014**, please inform us promptly of the changes, by e-mail to askhertz@hertzfoundation.org.

What is your present **mailing** address?

4067 A Miramar Street
Present street address

La Jolla | CA | 92037
City | **State (abbr only)** | **Zip (first five digits only)**

In which state do you currently hold residency? CA United States
Residency

What are your present home and work phone numbers? (Be sure to include area codes.)

Cell phones	Home phones	Work phones
(650)2040912	(541)9532482	( )
(541)9532482	( )	( )

obot@alum.mit.edu
e-mail address

Do you have a second or alternate e-mail address? If so, enter it here:

olga.botvinnik@gmail.com
Alternate e-mail address

We have found e-mail to be an especially good way to reach Fellowship applicants. We will be using e-mail to correspond with you regarding the status of your application, and to let you know when your application arrives at our office. We may also be using e-mail to alert you to the possibility of a technical interview.

Until what date do you expect this contact information to be valid? Use the same format, **mm/dd/yyyy,** as was used to enter your day of birth (e.g. **June 10, 2007** would be **06/10/2007**).

Contact information valid until: 09 / 05 / 2013

In order to plan the technical interviewing process, we need to know what airport is nearest to your current location. Find and enter the most appropriate airport abbreviation (from the list included in the Instruction section) into the "closest airport" field.

SAN
Closest airport

Please enter the address where we should mail the final selection announcement letter on/before April 1, 2014.

4067 Miramar Street, Apt A
April 2007 street address

La Jolla | CA | 92037
City | **State (abbr only)** | **Zip (first five digits only)**

The Interview

You will be contacted in mid-November to schedule a technical interview if you are one of the approximately 25% percent of our applicants who are selected for continued consideration. This interview generally lasts 45 to 60 minutes. It is patterned after the PhD oral exam and you may be asked to perform calculations, discuss your previous research work, and to demonstrate the breadth and depth of your technical knowledge.

The interview may be held in one of a variety of locations, including hotel conference rooms, restaurants, airport conference rooms, or on a university campus. You may not be given much advance notice and you will have to arrange transportation to your interview. Your reasonable-and-ordinary travel expenses will be fully reimbursed by the Foundation if, at our request, you must travel over 5 miles. If you are in any of the following cities or broad surrounding regions, interviews will be held on the following dates in the cities indicated. You will be notified of the actual location when you are contacted for an interview.

Saturday, November 23, 2013, Boston
Saturday, December 7, 2013, Chicago
Saturday, January 18, 2014, New York Area
Saturday, January 18, 2014, San Francisco Bay Area
Saturday, January 25, 2014, Los Angeles

We expect to interview most applicants at one of these five venues, but interviews at other locations may sometimes be scheduled by the Foundation. Reasonable expenses will be paid for your travel to the nearest interview site. While you may choose to travel to a more distant site at your own expense, you may wish to keep these dates in mind when making your Fall/Winter travel plans.*

Approximately 50 Finalists will be selected in early February and these Finalists will be interviewed for a second time between early February and mid-March. The Finalist interviews will take place in a variety of metropolitan areas and as before, you may not be given much advance notice and you will need to arrange transportation to your interview. Your reasonable-and-ordinary travel expenses will be fully reimbursed by the Foundation if, at our request, you must travel over 5 miles.*

* We ask that you consult with us first if your travel expenses are likely to exceed $100.

Four very brief essays

In the spaces below, (i.e., in approximately 300 words), please provide concise responses to each of the following questions.

Question 1: Choice of Field and Future Expectations

How did you choose your field and what are your primary expectations of your future career? If you are currently in your second or later year of graduate school, you should make your case here for why receiving a Fellowship would result in exceptional leverage in the kind and quality of your graduate work, including your ability to pursue promising new ideas. Please understand before continuing this application, that such a case will have to be very strong to be considered further, and that new Hertz Fellowships are very rarely granted to students currently in their second year of graduate study or beyond.

I chose the field of bioinformatics, the study of biology using techniques of mathematics and computer science, because it ties together my desire to understand life's molecular processes with my love of quantification. Learning about the translation of RNA to protein in high school, I thought incredulously to myself, "This process is happening inside all of us, right now." Later in college, as a math major working in a biology lab, I was most excited by running the code which converted the data of a randomly polka-dotted microarray slide to information: the DNA letters bound by a protein. And I realized I could connect my passions for biology and quantification through bioinformatics.

My sub-field within bioinformatics is single-cell genomics, the study of individual cells within a population. This is a radical shift from previous biological research, which drew conclusions from average measurements over many cells at a time, distorting the true biological interactions.

I intend to become a professor at a top institution, where I will be surrounded by leaders in their field who will question, support and challenge my research. I'm also excited about the impact of genomics outside of the academic world and hope to create spin-off companies from my research and consult for biotech, government, and other industries.

As a second-year graduate student, the Hertz Fellowship will grant me the leverage to create collaborations, and the freedom to fully immerse myself in developing assembly algorithms and data analysis pipelines for single-cell genomics. Specifically, I would extend the single-cell bacterial genome assemblers developed by UCSD's Prof. Pavel Pevzner's second lab at St. Petersburg Academic University. I would work with Pevzner's programmers to develop my idea of simultaneous single-cell RNA and DNA sequence assembly of human genomes, which would help us at Prof. Gene Yeo's lab at UCSD by applying to the question of RNA heterogeneity as a function of DNA heterogeneity in neural tissue, and its effect on developmental diseases.

Question 2: Proposed Field of Study

How do your proposed field of study and career constitute an application of the **physical** sciences or engineering?

Bioinformatics is the application of mathematics and computer science to discover new biology and advance medicine. Without the aid of bioinformatics, the work of multiscale modeling to predict protein structure from X-ray crystallography and NMR spectroscopy, which was awarded this year's Nobel Prize in chemistry, could not have been possible.

Using mathematical principles, I study alternative mRNA splicing of single cells during motor neuron differentiation. Alternative mRNA splicing allows for creation of different versions of genes from the same original DNA sequence. It is estimated that over 90% of human genes undergo alternative splicing, and splicing is especially prevalent in human brain. Previous experiments measured splicing over an average of all cells, but it is unknown how the different versions are shared (or not shared) amongst cells of a single population. If the bulk measurement shows 60% of version A and 40% of version B, does that mean within each cell, there is a 60/40 version breakdown? Or do 60% of cells have version A, and 40% version B? I developed a method of categorizing gene versions using linear algebra. By this method, I found that regardless of developmental stage, most cells have either one version or another; and if there is a mixture of molecules A and B, it is usually not consistently shared amongst all cells.

The principles we are developing in the motor neuron differentiation system are a template for the many exciting future applications of single-cell genomics: early cancer detection, kitchen-top DNA sequencers to test for food safety, and discovering new bacterial species from dirt. Single-cell is the future of genomics, and I will be leading the charge.

Question 3: Choice of Graduate School

What are the considerations involved in your choice of graduate school?

I came to University of California, San Diego (UCSD) to work on single-cell genomics. UCSD has a unique mix of people working on the technology for extracting molecules from single cells (Roger Lasken), algorithms to assemble the sequences (Pavel Pevzner), and analysis pipelines to deal with the noisy and heterogeneous data (Gene Yeo).

In particular, my advisor Prof. Gene Yeo's lab was the first to prototype the Fluidigm C1, a microfluidic machine which efficiently and reproducibly separates single cells. The Yeo lab also has the experience in machine learning algorithms to tease apart the subtle differences between individual cells, and the biological ability to draw biological conclusions from the data. In addition, I hope to work with Dr. Roger Lasken from the J. Craig Venter Institute here in San Diego, who developed a method for simultaneous extraction of both RNA and DNA, allowing for simultaneous assembly of single-cell genomes and transcriptomes, a completely novel application as almost all current genomics data assumes all cells have exactly the same genome, even though we know this is not true. I hope to extend the single-cell bacterial genome assemblers developed by Prof. Pavel Pevzner's group, to single-cell human genome and transcriptome assembly. Joining forces within the collaborative environment of UCSD is key because ultimately, my goal is discovery of new, disease-relevant, biological knowledge, which the Yeo lab excels in. But the integration of the technology developed by Lasken, and algorithms by Pevzner, catalyzes our ability to uncover biological breakthroughs.

Besides my single-cell specific interests, I chose UCSD because I wanted a university that not only had fantastic biology, math, and computer science departments, but also already had a fabulous bioinformatics program which seamlessly integrates these three fields.

Question 4: Chronological Resume

Provide a concise resume, in chronological order, with dates, recapitulating significant periods of technical and other creative activity since high school graduation. Omit activities only distantly related to your professional development. Include workshops, summer schools, a general description of all courses of study pursued (e.g. "3 quarters of Differential Equations") and degrees expected or awarded (dates, institutions, fields). Separate your undergraduate activities from your graduate activities (if/as applicable) with a single dashed line.

------ Undergraduate ------
--- 2006 ---
- September -
Began study at the Massachusetts Institute of Technology (MIT), Cambridge, MA in Mathematics and Biological Engineering.

--- 2007 ---
- June to September -
Worked in Prof. Martha Bulyk's laboratory on transcription factor binding sites via protein binding microarrays, contributing to a publication in Cell journal.

--- 2008 ---
- January -
Participated in Undergraduate Professional Opportunities Program to hone networking and professional skills.

- June -
Work from Prof. Bulyk's laboratory was published in Cell journal: "Variation in homeodomain DNA binding revealed by high-resolution analysis of sequence preferences." Berger MF, Badis G, Gehrke AR, Talukder S, Philippakis AA, Peña-Castillo L, Alleyne TM, Mnaimneh S, Botvinnik OB, Chan ET, Khalid F, Zhang W, Newburger D, Jaeger SA, Morris QD, Bulyk ML, Hughes TR. Cell (2008) PMID: 18585359

- June to August -
Won a summer scholar position at Howard Hughes Medical Institute Janelia Farm Research Campus to work with Sean Eddy on using Hidden Markov Models to create a better null model for protein similarity search.

--- 2009 ---
- January to December -
Worked in Prof. David Gifford's laboratory: compared synthetic lethality networks of yeast strains, and modeled T-cell receptor enrichment, in silico.

--- 2010 ---
- January to June -
Worked in Prof. Sebastian Seung's laboratory on neuron orientation in rabbit retina.

- June -
Earned two S.B. degrees in Mathematics and Biological Engineering at MIT

MIT Course highlights, lower level:
1 semester each calculus, differential equations, linear algebra, computational science and engineering
1 semester each chemistry, organic chemistry, genetics, biochemistry
2 semesters physics with laboratory
1 semester computer programming

MIT Course highlights, upper level:
1 semester each probability, discrete math, number theory, computational molecular biology, physical mathematics, theoretical computer science
1 semester each thermodynamics; cell biology; biological engineering lab; biological instrumentation and measurement; biomechanics; fields, forces, and flows; biological engineering design; computational systems biology

Academic Honors

In the space provided below, list, in chronological order, academic honors and distinctions you have received and the time or time-interval of receipt. Separate your undergraduate from your graduate awards (if/as applicable) with a single dashed line (Include title, reason for award, and where/when received). Especially significant awards received in high school can also be included. Use no more than one line per award whenever possible (what, where/when received).

------ Undergraduate ------
Gordon-MIT Engineering Leadership Scholar (2008-2009)
One of two (out of a thousand) undergraduates to double major in Mathematics and Biological Engineering
------ Graduate ------
First person to finish 2-year M.S. in 9 months at UCSC
First 1st-year graduate student to TA a graduate class in UCSC BME program

Fellowships, Scholarships, etc.

List here, in chronological order, any fellowships, scholarships, teaching or other appointments held since entering college or university. Separate your undergraduate and graduate education intervals (if/as applicable) with a single dashed line. Put an asterisk (*) at the beginning of all lines indicating a national-level award (e.g., award of a National Merit Scholarship; NSF, NASA, or NDSEG Fellowship; election to Phi Beta Kappa, Tau Beta Pi, or Sigma Xi; etc.). Use one line per item whenever feasible (provide basic what, where, when data).


```
------ Undergraduate ------
*Howard Hughes Medical Institute Janelia Farm Research Campus Summer Scholar (2008)
*Cold Spring Harbor Undergraduate Research Program (2008 and 2009, declined for other
opportunities)
------ Graduate ------
UCSC Regent’s Fellowship (2011-2012)
*NSF Graduate Research Fellowship Honorable Mention (top 20% of applicants,
2011-2012)
*Hertz Fellowship Finalist (2012-2013)
*National Defense Science and Engineering Graduate (NDSEG) Fellow (2012-2013)
```

Previous Research

Please list the most significant research projects that you have pursued, in chronological order. (Include reference information for those that have been formally documented, presented at a conference, or submitted for publication.)

Examples:

1) "Controlled Fusion in the Shadow of the French Alps: My Summer Internship at Grenoble"; June-August 2002

2) "Peculiarities in Gene Transcription Regulation in wingless Drosophila Mutants", prepared/presented as part of the Student Research Seminar Series; U of Calif. Report, UCRL Report 1029-04, Sept 01-May 02.

3) "Deficiencies in Stereospecific Iodination of Thyroxine Precursors Isolated From Chernobyl-Region Voles", J. Exotic Endocrine Chem. 239, 3365 (June 2003), K. Early, M. Y. Name, L. Late, and M. Middle.

*****IMPORTANT*****

Finally, choose one or two projects that best exemplify your own creativity and discuss in more detail what you personally contributed to them.

Note that your answer to this question is a very important factor in our selection process. Please highlight what makes your personal contributions stand out, and use the first person wherever appropriate. Specific evidence of your personal creativity is more important to us than what your research group did as a whole, unless you contributed centrally to its leadership.

Please submit copies of your most significant scientific publications/reports using the upload capability on Page 13.

Number of:

B.S. Researches Pursued	5
B.S. Researches Documented	1
B.S. Researches Submitted to Refereed Publications	1
Grad. Researches Pursued	7
Grad. Researches Documented	2
Grad. Researches Submitted to Refereed Publications	1

Previous Research

1) "Variation in homeodomain DNA binding revealed by high-resolution analysis of sequence preferences." Berger MF, Badis G, Gehrke AR, Talukder S, Philippakis AA, Peña-Castillo L, Alleyne TM, Mnaimneh S, Botvinnik OB, Chan ET, Khalid F, Zhang W, Newburger D, Jaeger SA, Morris QD, Bulyk ML, Hughes TR. Cell (2008) PMID: 18585359

2) "Creating a better null model for HMMER protein domain search." PI: Dr. Sean Eddy. Location: Howard Hughes Medical Institute Janelia Farm Research Campus. June-August 2008.

3) "Differences in synthetic lethality networks in Saccharomyces cerevisiae strains S288C and Sigma1287b." PI: Prof. David Gifford. Location: MIT Computer Science and Artificial Intelligence Laboratory. January - June 2009.

4) "Detecting immune response by T-cell receptor sequence enrichment." PI: Prof. David Gifford. Location: MIT Computer Science and Artificial Intelligence Laboratory. June - December 2009

5) "Determination of neuron orientation via computational image analysis." PI: Prof. Sebastian Seung. Location: MIT Department of Brain and Cognitive Sciences. January - June 2010

6) "Uncovering of new cancer activators via integrated genomics." PI: Dr. Jill Mesirov. Location: Broad Institute of Harvard and MIT. November 2010 - September 2011

I worked at the Broad Institute of Harvard and MIT in Cambridge, MA, with Prof. Jill Mesirov to develop REVEALER, an algorithm that integrates genomic and functional data to infer new associations [A]. For example, a researcher may know that a certain oncogenic gene signature is overexpressed in many cancers, and that this overexpression can be caused by a mutation in the governing oncogene. However, there are many cases where there is no mutation in that oncogene and yet the pathway is highly expressed. REVEALER finds novel candidate activators of this signature by removing samples which already have a mutation in the original oncogene, and searching for the top genomic feature that explains the remaining samples using a mutual information (MI) metric to discern between top candidates. This flexible MI metric was also used to determine associations between sensitivity to the melanoma drug PLX-4720 and NFkB gene set expression [B], analysis of response to the myelodysplastic syndrome drug ezatiostat via microRNA and gene set expression [C], and in single-sample gene set enrichment analysis [D]. The REVEALER algorithm found that resistance to the oncogene KRAS RNAi knockdown is related to an amplicon in 8q23-24 in addition to KRAS mutation [E], a finding that had taken another researcher months of looking through Excel files. This ambitious project combines state-of-the-art computational methods with genomic analysis in the same way I hope to do in my future research.

Publications:

[A] Botvinnik OB, Tamayo P, Mesirov JP. Discovery of novel candidate oncogenic activators with REVEALER. Intelligent Systems for Molecular Biology Conference. Vienna, Austria (2011)

[B] Wood KC, Konieczkowski DJ, Johannessen CM, Boehm JS, Tamayo P, Botvinnik OB, Mesirov JP, Hahn WC, Root D.E, Garraway LA, Sabatini DM. Miniaturized functional screening reveals genetic modifiers of therapeutic response in melanoma. Sci Signaling (2012) PMID: 22589389

[C] Galili N, Tamayo P, Botvinnik OB, Mesirov JP, Brown G, Raza A. Prediction of response to therapy with ezatiostat in lower risk myodysplastic syndrome. J Clin Oncol (2012) PMID: 22559819

[D] Birger C, Botvinnik OB, Tamayo P, Mesirov JP. Single-sample GSEA compares gene set enrichment of multi-sample experiments. (in preparation)

[E] Botvinnik OB*, Kim JW*, Birger C, Rosenbluh J, Shrestha Y, Abazeed M, Hammerman PS, Abudayyeh O, DiCara D, Konieczkowski DJ, Johannassen C, Alizad-Ravhar AR, Alexe G, Aguirre A, Ghandi M, Greulich H, Vazquez F, Weir BA, Van Allen EM, Liberzon A, Tsherniak A, Shao DD, Zack TI, Noble M, Getz G, Beroukhim R, Garraway LA, Ardakani M, Romualdi C, Sales G, Barbie DA, Boehm J, Hahn WC, Mesirov JP, Tamayo P. Mapping Genomic Alterations to Functional Profiles of Pathway Activation, Gene Dependency and Drug Sensitivity. (submitted) *These authors contributed equally to this work

GRE Scores

Every Fellowship applicant is REQUIRED to take at least the aptitude portion of the Graduate Record Examination and to have the results (Verbal, Quantitative and Writing or Analytical scores) sent to the Foundation. This can be done by adding the Hertz Foundation ID Number, 4366, to those of the schools requiring the GRE for admission as a graduate student on your GRE form. If you have already taken the GRE, you can enter your scores in the boxes provided. We continue to accept GRE scores after the application deadline. Initial interview decisions will be made by mid-November. If you enter scores, you must also send documented verification of these GRE scores to the Foundation. MCAT scores may be accepted in lieu of GRE scores if you are already enrolled in medical school.

	Score	%ile	Date of Exam (mm/yyyy)	GRE Subject 1	Subject 1 Score	%ile	Date of Exam (mm/yyyy)
Verbal	570	80	12 / 2010				/
Quantitative	800	94	12 / 2010	GRE Subject 2	Subject 2 Score		
AWA/Writing	5.5	94	12 / 2010				/
Analytical			/				

Transcripts

You must submit 1 set of **official** transcripts of all your college academic records, undergraduate and (if applicable) graduate. Official transcripts are those submitted directly from the school to the Hertz Foundation. (You may collect these and send them, but be sure they are still in the sealed, signed envelopes, prepared by the school's Registrar.)

Please list here the schools you have attended, the dates you were enrolled, your GPA while attending each school, and any degrees earned. We will expect you to send an official transcript from each of these schools. (If you briefly attended an institution, e.g., for a summer course, or for a class that is unrelated to your major, e.g., a foreign language session, you do not need to furnish a transcript of this schooling.) Please indicate which of the schools in your list below will be providing transcripts by placing an "x" in the "Transcript ?" box on the right.

	Enrollment dates. mm/yyyy	GPA	Degree earned	Transcript ?
Transcripts 1 Massachusetts Institute of Technology	06 / 2010	4.20	SB	X
Transcripts 2 Massachusetts Institute of Technology	06 / 2010	4.20	SB	X
Transcripts 3 University of California, Santa Cruz	09 / 2011	3.91	MS	X
Transcripts 4 University of California, San Diego	09 / 2012	3.72		X
Transcripts 5	/			
Transcripts 6	/			
Transcripts 7	/			
Transcripts 8	/			

Other Opportunities

The Foundation is often approached by other institutions, including industrial organizations, universities and national laboratories, who are interested in considering our Applicants or Fellows for part-time or summer employment or in offering other types of opportunities. The Foundation also receives requests for information concerning Applicants from other funding organizations, both governmental and private charities, who are seeking well-qualified applicants to consider for fellowships. The Foundation will normally share information on a Fellowship application with potential employers and other funding organizations that the Foundation considers to be responsible, disinterested and acting in the public interest, provided that the Applicant has not asked that such information be withheld. If you would prefer that we NOT share the information in this application with these other organizations, please check the "Do NOT make other opportunities available" box below. (Your chances for selection as a Hertz Fellow are not influenced by this choice.)

Do NOT make other opportunities available to me: ☐

Optional Survey

How did you hear about the Hertz Foundation Fellowship? Referred by current Fellow

If you heard about the Hertz Foundation through a press release, news story, or other source, please use the box below to provide a brief description:

Were you a Goldwater Scholar or Nominee?
Choose one of the following: None of the above

Reference Reports

Please list here the names, addresses, telephone numbers and email addresses of the four persons who best know your academic and professional ability and whom you are requesting to provide Reference Reports on your behalf. Note that these Reports must be received by the Foundation by Monday, November 4, 2013 or your application is not assured of full consideration by the Foundation. IT IS YOUR RESPONSIBILITY TO ASSURE TIMELY RECEIPT OF ALL FOUR OF THESE REFERENCE REPORTS BY THE FOUNDATION. Once your referees are registered, they will receive an email from Embark explaining how to submit the Reference Report. This email will include the necessary access information and password.

Reference Report 1

	title	first name	last name
	Dr.	Jill	Mesirov
title/position	Director of Computational Biology and Bioinformatics		
school or business	Broad Institute		
address	7 Cambridge Center Cambridge MA 02142 USA		
phone	(617) 7147070		
email	mesirov@broadinstitute.org		

Reference Report 2

	title	first name	last name
	Prof.	Nader	Pourmand
title/position	Associate Professor		
school or business	University of California, Santa Cruz		
address	1156 High St Santa Cruz CA 95060 USA		
phone	(831) 5027315		
email	pourmand@soe.ucsc.edu		

Reference Report 3

	title	first name	last name
	Dr.	Sean	Eddy
title/position	Group Leader		
school or business	Howard Hughes Medical Institute Janelia Farm Research Camp		
address	19700 Helix Dr Ashburn VA 20147 USA		
phone	(571) 2094163		
email	eddys@janelia.hhmi.org		

Reference Report 4

	title	first name	last name
	Prof.	Gene	Yeo
title/position	Assistant Professor		
school or business	University of California, San Diego		
address	2880 Torrey Pines Scenic Dr. La Jolla CA 92037		
phone	(858) 5349322		
email	geneyeo@ucsd.edu		

Contact Information

After examination of all applicants for the Fannie and John Hertz Foundation Graduate Fellowships, approximately 25% of the applicants will be offered interviews. These interviews are technical in nature, covering topics in physics, chemistry, mathematics, engineering, quantitative biology, etc.

Most interviewers are volunteers, fitting the meetings into their work or travel schedules. Interviews may take place at hotels, schools, airports, restaurants, or other convenient locations. If we request that you travel more than five miles to meet with a Foundation interviewer, we will fully reimburse your reasonable-and-ordinary travel expenses to attend the interview. **However, please note the Hertz Foundation will only reimburse up to $400 for airfare roundtrip from outside of the United States.**

Because the volunteers sandwich interviews into their other activities, there may be only two or three days notice of your interview. We must know how to let you know of a possible interview from **November 15, 2013 to March 15, 2014**.

In the boxes below, please enter telephone numbers where we can reach you, or an answering machine, or a friend, or a relative, during this time, so that you will be able to contact the Hertz Foundation within 24 hours of our attempt to reach you. Please identify each number, e.g., "my work #," "my parent's #," "my dad's # at work," "my advisor's #." We will try the numbers in the order you give them, but the more options you give us, the more likely it is that we will be able to reach you when we need to. Please also include email addresses, as we will be using email as our first line of communication in most cases.

My Cell #	self	(541) 9532482	olga.botvinnik@gmail.com
My Dad's work #	father	(541) 3465636	botvinn@math.uoregon.edu
My close friend's #	close friend	(814) 8836535	kwasi.nti@gmail.com
		( )	
		( )	
		( )	
		( )	
		( )	
		( )	
		( )	

ADDITIONAL INFORMATION

Olga	Botvinnik	34454788

I used to think that data was the same as knowledge. That knowing the fact that in DNA, A binds T and G binds C meant you were smart. But there is deeper information behind this pattern, the how of AT and GC binding, such as that A and T form two hydrogen bonds, but G and C form three. So A cannot bind with either G or C, and vice versa. But even deeper, what does the implication of the three bonds in GC versus two bonds in AT have for biological function, or evolutionary conservation? Knowledge is not what, the exact answer, or even how it was obtained, but the "why?" Integrating with other data, adding context, plus a lot of deep thinking, is what produces knowledge.

In high school, we visited a cadaver lab and the technician described a cadaver which was completely missing part of a leg muscle. Immediately, I thought to myself that this should be easily detectable through a genetic test. If only we had the data of the person's genome sequence, we would have the knowledge of the presence or absence of this person's muscle. Later, I learned that the genome sequence alone doesn't tell you whether a gene is on or off, but it is only data. Valuable data, yes, but only one step. If we have that person's genome sequence plus a "reference" human genome sequence, maybe then we could pinpoint what places are different in this person's genome. But we would be a far cry from knowledge. We would still need to know where the muscle genes are, how leg-specific genes are turned on during embryogenesis, and how the muscle development process can be disrupted.

In the quest to obtain knowledge from data, one common step is visualization, but most scientific visualization software leaves much to be desired, and may even distort interpretation of the data. I truly believe that scientific progress is impeded when improper data visualizations are used. Improving data visualizations has two main advantages. First, the researcher can understand their work better and faster, enabling rapid prototyping of hypotheses. Second, the researcher can spend more time discussing the science instead of losing the audience with a confusing figure. There is established research in data visualizations through color perception studies by Cynthia Brewer, and Edward Tufte's work in simplifying graphs to their essence: data. However, the standards for data visualization in the scientific world are lacking. Data visualization is an afterthought for many researchers, but I'm working on this forgotten part and attempting to automate reasonable plotting defaults for myself and others.

To help establish reasonable standards for scientific visualization, I developed a Python package, "prettyplotlib," which builds on Python's existing plotting library, "matplotlib," to painlessly create clean default figures using principles by Brewer and Tufte. Importantly, most users are not interested in creating the most beautiful plot of their data -- they are focused on getting the data visualized, and do care about design. This package keeps the focus on the data, while making it easy to create clear figures. Users find prettyplotlib easy to display their results, without the painstaking process of retouching every tiny aspect of the plot. Since prettyplotlib has been released, it has had thousands of views from around the world, and has been starred by more than 300 people on GitHub, a code-sharing website. Users have commented on its ease and practicality of use: "finally a wrapper for matplotlib with sensible defaults -- great plots without the pain," and "FINALLY. prettyplotlib might just make matplotlib usable (because by default, it's not even close)" Throughout graduate school, I will continue advocating for sound data visualizations and developing prettyplotlib.

While visualization helps with converting data to information, it is easy to get stuck in this information-generating phase and be fooled that information is knowledge. It may be easy to generate many different figures of your data versus other data, but showing correlation may not lead to any new knowledge. It is important to not just plot data for the sake of plotting, but to filter the data into potentially interesting subsets, and then visualize. But even those figures are not yet knowledge - we must think deeply about what these plotted relationships may mean.

Research ties data to knowledge through information. But there is no end to knowledge. Eventually, knowledge may become as simple as data, that A binds T and G binds C. But when Watson and Crick first noticed the pattern of the proportions of A exactly matching the

Table of Contents

Personal Website
olgabotvinnik.com

Blog
blog.olgabotvinnik.com

On my blog, I discuss productivity in the academic world and write Python tutorials.

My most popular productivity post on Eric Lander: http://bit.ly/PMMkWz

AP Biology curriculum materials
http://compbio.soe.ucsc.edu/binf-in-AP/

RNA-Sequencing Differential Expression package
https://github.com/olgabot/rna-seq-diff-exprn

blog post on usage: bit.ly/PMMiOc

Prettyplotlib

My most popular blog post ever, a tutorial on prettyplotlib, the Python plotting library I wrote: bit.ly/17675RG

And the source code: https://github.com/olgabot/prettyplotlib

Cell

Resource

Variation in Homeodomain DNA Binding Revealed by High-Resolution Analysis of Sequence Preferences

Michael F. Berger,[1,3,8] Gwenael Badis,[5,8] Andrew R. Gehrke,[1,8] Shaheynoor Talukder,[5,8] Anthony A. Philippakis,[1,3,6] Lourdes Peña-Castillo,[4] Trevis M. Alleyne,[5] Sanie Mnaimneh,[4] Olga B. Botvinnik,[1,7] Esther T. Chan,[5] Faiqua Khalid,[4] Wen Zhang,[5] Daniel Newburger,[1] Savina A. Jaeger,[1] Quaid D. Morris,[4,5] Martha L. Bulyk,[1,2,3,6,*] and Timothy R. Hughes[4,5,*]
[1]Division of Genetics, Department of Medicine
[2]Department of Pathology
Brigham and Women's Hospital and Harvard Medical School, Boston, MA 02115, USA
[3]Committee on Higher Degrees in Biophysics, Harvard University, Cambridge, MA 02138, USA
[4]Banting and Best Department of Medical Research
[5]Department of Molecular Genetics
University of Toronto, Toronto, ON M5S 3E1, Canada
[6]Harvard/MIT Division of Health Sciences and Technology (HST), Harvard Medical School, Boston, MA 02115, USA
[7]Department of Mathematics, Massachusetts Institute of Technology, Cambridge, MA 02139, USA
[8]These authors contributed equally to this work
*Correspondence: mlbulyk@receptor.med.harvard.edu (M.L.B.), t.hughes@utoronto.ca (T.R.H.)
DOI 10.1016/j.cell.2008.05.024

SUMMARY

Most homeodomains are unique within a genome, yet many are highly conserved across vast evolutionary distances, implying strong selection on their precise DNA-binding specificities. We determined the binding preferences of the majority (168) of mouse homeodomains to all possible 8-base sequences, revealing rich and complex patterns of sequence specificity and showing that there are at least 65 distinct homeodomain DNA-binding activities. We developed a computational system that successfully predicts binding sites for homeodomain proteins as distant from mouse as *Drosophila* and *C. elegans*, and we infer full 8-mer binding profiles for the majority of known animal homeodomains. Our results provide an unprecedented level of resolution in the analysis of this simple domain structure and suggest that variation in sequence recognition may be a factor in its functional diversity and evolutionary success.

INTRODUCTION

The approximately 60 amino acid homeobox domain or "homeodomain" is a conserved DNA-binding protein domain best known for its role in transcription regulation during vertebrate development. The homeodomain can both bind DNA and mediate protein-protein interactions (Wolberger, 1996); however, the precise mechanisms that dictate the physiological function and target range of individual homeodomain proteins are in general either unknown or incompletely delineated (Banerjee-Basu et al., 2003; Svingen and Tonissen, 2006). In several cases, functional specificity can be traced to the homeodomain itself (Chan and Mann, 1993; Furukubo-Tokunaga et al., 1993; Lin and McGinnis, 1992), indicating that individual homeodomains have distinct protein- and/or DNA-binding activities. Since many homeodomains have similar DNA sequence preferences, much attention has been paid to the role of protein-protein interactions in target definition (Svingen and Tonissen, 2006), despite evidence that the sequence specificity of monomers contributes to targeting specificity (Ekker et al., 1992) and that binding sequences do vary, particularly among different subtypes (Banerjee-Basu et al., 2003; Ekker et al., 1994; Sandelin et al., 2004). Indeed, it has been proposed that the DNA-binding specificity of homeodomains is determined by a combinatorial molecular code among the DNA-contacting residues (Damante et al., 1996).

Efforts to understand the physiological and biochemical functions of homeodomains have been hindered by the fact that most have only a few known binding sequences, if any. Position weight matrices (PWMs) have been compiled for 63 distinct homeodomain-containing proteins from human, mouse, *D. melanogaster*, and *S. cerevisiae* in the JASPAR (Bryne et al., 2008) and TRANSFAC (Matys et al., 2003) databases. These matrices are based on 5 to 138 individual sequences (median 18), presumably capturing only a subset of the permissible range of binding sites for these factors. Further, the accuracy of PWM models has been questioned (Benos et al., 2002), and there are many examples in which transcription factors bind sets of sequences that cannot be described in a conventional PWM representation (Blackwell et al., 1993; Chen and Schwartz, 1995; Overdier et al., 1994).

Moreover, the sequence preferences of the individual proteins can, in some cases, be altered by the binding context: For instance, the binding specificity of the complex of *Drosophila* Hox-Exd homeodomain proteins is remarkably different from

CANCER

MicroSCALE Screening Reveals Genetic Modifiers of Therapeutic Response in Melanoma

Kris C. Wood,[1,2]* David J. Konieczkowski,[2,3] Cory M. Johannessen,[2,3] Jesse S. Boehm,[2] Pablo Tamayo,[2] Olga B. Botvinnik,[2] Jill P. Mesirov,[2] William C. Hahn,[2,3] David E. Root,[2] Levi A. Garraway,[2,3] David M. Sabatini[1,2,4]*

Cell microarrays are a promising tool for performing large-scale functional genomic screening in mammalian cells at reasonable cost, but owing to technical limitations they have been restricted for use with a narrow range of cell lines and short-term assays. Here, we describe MicroSCALE (Microarrays of Spatially Confined Adhesive Lentiviral Features), a cell microarray–based platform that enables application of this technology to a wide range of cell types and longer-term assays. We used MicroSCALE to uncover kinases that when overexpressed partially desensitized *B-RAF*^{V600E}–mutant melanoma cells to inhibitors of the mitogen-activated protein kinase kinase kinase (MAPKKK) RAF, the MAPKKs MEK1 and 2 (MEK1/2, mitogen-activated protein kinase kinase 1 and 2), mTOR (mammalian target of rapamycin), or PI3K (phosphatidylinositol 3-kinase). These screens indicated that cells treated with inhibitors acting through common mechanisms were affected by a similar profile of overexpressed proteins. In contrast, screens involving inhibitors acting through distinct mechanisms yielded unique profiles, a finding that has potential relevance for small-molecule target identification and combination drugging studies. Further, by integrating large-scale functional screening results with cancer cell line gene expression and pharmacological sensitivity data, we validated the nuclear factor κB pathway as a potential mediator of resistance to MAPK pathway inhibitors. The MicroSCALE platform described here may enable new classes of large-scale, resource-efficient screens that were not previously feasible, including those involving combinations of cell lines, perturbations, and assay outputs or those involving limited numbers of cells and limited or expensive reagents.

INTRODUCTION

Gain- and loss-of-function screens are powerful experimental approaches that can be leveraged to reveal the mechanistic underpinnings of a wide range of mammalian cellular biological processes (*1*–*5*). However, the costs and logistical challenges of current screening techniques restrict the range and extent of their application. These restrictions are particularly evident in large-scale screens involving multiple cell lines, genetic and pharmacological perturbations, or assay outputs, or alternatively in cases where cells, detection reagents, or other material resources are limited. Arrayed screens in multiwell plates can provide discrete, multiplexed measurements in proliferation- and image-based assays, but cost, labor requirements, and the need for specialized screening facilities equipped with appropriate biosafety and fluid-handling equipment limit their deployment (*3*, *5*, *6*). Pooled screening approaches address a number of these limitations and enable the screening of larger numbers of genetic constructs at lower cost (*5*, *7*–*10*). However, pooled screening is incompatible with image-based assays and typically requires large quantities of cells and reagents and libraries of genetic overexpression or knockdown reagents that are barcoded. Further, the scale of tissue culture required for pooled screens may present logistical challenges when large numbers of screens, for example, of many different cell lines or environmental perturbations, are required.

Cell microarrays have the potential to combine the best features of existing screening technologies, such as the assay versatility of multiwell plate–based screening and the high efficiency of pooled screening (*11*–*14*). Cell microarrays consist of hundreds to thousands of distinct genetic reagents printed as individual, microscopic "features" on glass slides, which are then seeded with adherent cells that attach and become treated with the reagent present on each feature (*13*, *15*). In principle, these systems make it possible to perform and analyze many parallel genetic perturbations on a single slide with high screening throughput and low cell and reagent consumption. In practice, although they have been featured in a number of interesting proof-of-concept studies describing variations on the basic cell microarray design with potential applications in several areas of cell biology (*13*, *14*, *16*–*21*), these systems have only rarely been adopted in large-scale screening applications because of several technical limitations (*22*, *23*). Namely, microarrays based on chemically delivered DNA expression plasmids or small interfering RNAs can only be used with cell types that are easily transfectable (for example, human embryonic kidney 293T cells) and with screening assays that are short in duration (typically 1 to 3 days) because of the inefficient and transient nature of nonviral transgene expression (*13*–*15*). Conversely, microarrays based on virally delivered open reading frames (ORFs) or short hairpin RNAs (shRNAs) can stably and efficiently infect a wide range of cell types but require large libraries of concentrated, purified viruses that cannot be prepared in a labor- or time-efficient way with existing methods (*24*). Both formats are further limited because cells tend to migrate away from the features on which they originally land and intermix with cells on neighboring features, an effect that considerably decreases the achievable spatial density of printed features

[1]Whitehead Institute for Biomedical Research, 9 Cambridge Center, Cambridge, MA 02142, USA. [2]Broad Institute of Harvard and Massachusetts Institute of Technology, 7 Cambridge Center, Cambridge, MA 02142, USA. [3]Department of Medical Oncology, Dana-Farber Cancer Institute, Harvard Medical School, 450 Brookline Avenue, Boston, MA 02215, USA. [4]Howard Hughes Medical Institute, Department of Biology, Massachusetts Institute of Technology, 77 Massachusetts Avenue, Cambridge, MA 02139, USA.
*To whom correspondence should be addressed. E-mail: kcwood@alum.mit.edu (K.C.W.); sabatini@wi.mit.edu (D.M.S.)

JOURNAL OF HEMATOLOGY & ONCOLOGY

RESEARCH Open Access

Prediction of response to therapy with ezatiostat in lower risk myelodysplastic syndrome

Naomi Galili[1*], Pablo Tamayo[2], Olga B Botvinnik[2], Jill P Mesirov[2], Margarita R Brooks[1], Gail Brown[3] and Azra Raza[1]

Abstract

Background: Approximately 70% of all patients with myelodysplastic syndrome (MDS) present with lower-risk disease. Some of these patients will initially respond to treatment with growth factors to improve anemia but will eventually cease to respond, while others will be resistant to growth factor therapy. Eventually, all lower-risk MDS patients require multiple transfusions and long-term therapy. While some patients may respond briefly to hypomethylating agents or lenalidomide, the majority will not, and new therapeutic options are needed for these lower-risk patients. Our previous clinical trials with ezatiostat (ezatiostat hydrochloride, Telentra®, TLK199), a glutathione S-transferase P1-1 inhibitor in clinical development for the treatment of low- to intermediate-risk MDS, have shown significant clinical activity, including multilineage responses as well as durable red-blood-cell transfusion independence. It would be of significant clinical benefit to be able to identify patients most likely to respond to ezatiostat before therapy is initiated. We have previously shown that by using gene expression profiling and grouping by response, it is possible to construct a predictive score that indicates the likelihood that patients without deletion 5q will respond to lenalidomide. The success of that study was based in part on the fact that the profile for response was linked to the biology of the disease.

Methods: RNA was available on 30 patients enrolled in the trial and analyzed for gene expression on the Illumina HT12v4 whole genome array according to the manufacturer's protocol. Gene marker analysis was performed. The selection of genes associated with the responders (R) vs. non-responders (NR) phenotype was obtained using a normalized and rescaled mutual information score (NMI).

Conclusions: We have shown that an ezatiostat response profile contains two miRNAs that regulate expression of genes known to be implicated in MDS disease pathology. Remarkably, pathway analysis of the response profile revealed that the genes comprising the jun-N-terminal kinase/c-Jun molecular pathway, which is known to be activated by ezatiostat, are under-expressed in patients who respond and over-expressed in patients who were non-responders to the drug, suggesting that both the biology of the disease and the molecular mechanism of action of the drug are positively correlated.

Background

Myelodysplastic syndrome (MDS) is a clonal stem cell disorder resulting in bone marrow failure and variable cytopenias. Development of new treatment strategies has greatly improved the outlook for patients with MDS. There are three FDA-approved drugs for therapy of patients who have become transfusion-dependent, including two hypomethylating drugs (HMAs), azacitidine and decitabine, and the thalidomide derivative lenalidomide. Patients with higher-risk disease have been shown to benefit from HMA therapy [1,2], while patients with lower-risk disease with a karyotype of clonally restricted deletion of the long arm of chromosome 5 (deletion 5q or del[5q]) are highly responsive to lenalidomide [3,4]. Only 26% of transfusion-dependent lower-risk patients without del(5q) will also become transfusion-independent while on treatment [5], but the FDA has not approved lenalidomide for these patients. There are few treatment options for the majority of transfusion-dependent MDS patients with lower-risk disease. This situation represents a significant unmet medical need. Once disease-modifying therapy is

* Correspondence: ng2368@columbia.edu
[1]Department of Medicine, Division of Hematology and Oncology, Columbia University Medical Center and New York Presbyterian Hospital, 177 Fort Washington Ave., New York, NY 10032, USA
Full list of author information is available at the end of the article

Goncearenco et al. BMC Bioinformatics 2012, 13(Suppl 18):A1
http://www.biomedcentral.com/1471-2105/13/S18/A1

BMC Bioinformatics

INTRODUCTION Open Access

Highlights from the Eighth International Society for Computational Biology (ISCB) Student Council Symposium 2012

Alexander Goncearenco[1*], Priscila Grynberg[2], Olga B Botvinnik[3], Geoff Macintyre[4], Thomas Abeel[5,6]

From Eighth International Society for Computational Biology (ISCB) Student Council Symposium 2012
Long Beach, CA, USA. 13-14 July 2012

About the Student Council and the symposium

The Student Council (SC) is part of ISCB and is led by students and postdocs. Our mission is to nurture and assist the next generation of computational biologists. We offer networking opportunities and soft skill training to scientists in bioinformatics who are in the early stages of their career.

The SC Symposium series has been running for eight consecutive years: Vienna 2011 [1], Boston 2010 [2], Stockholm 2009 [3], Toronto 2008 [4], Vienna 2007 [5], Fortaleza 2006 and Madrid 2005. Every second year when ISMB and ECCB are not jointly organized, an additional symposium is held called the European Student Council Symposium (ESCS). This has been running since 2010 and this year ESCS was held in conjunction with ECCB in Basel.

Scope and format of the meeting

The Student Council Symposium is a day-long meeting held in conjunction with the ISMB/ECCB conferences every year. The goal of our activities at ISMB is to help fellow students build their career in computational biology. We do this by creating opportunities to meet peers from all over the world, promote the exchange of ideas and provide networking opportunities. The 8th ISCB Student Council Symposium (SCS8) started with a scientific speed dating session in which participants introduced themselves and their science to a new person every two minutes. The traditional scientific component of the meeting consisted of three sessions, each with a keynote talk and several student presentations. In the evening, everybody got the opportunity to present their work during the poster session.

This year's keynote lectures were kindly delivered by Dr. Robin Dowell (University of Colorado at Boulder), Dr. Matthew Hibbs (University of Maine at The Jackson Laboratory), and Dr. Jonathan Eisen (University of California, Davis). Furthermore, three institutional partners gave short presentations about career opportunities at their respective institute: NICTA (http://www.nicta.com.au), EBI (http://www.ebi.ac.uk) and EMBO (http://www.embo.org).

This year, the symposium received 103 submissions from students. These submissions were peer-reviewed by 56 independent reviewers. Ten abstracts were selected for oral presentation and approximately 50 additional abstracts were accepted for poster presentations. Abstracts of the oral presentations are included in this meeting report. Abstracts of the poster presentations are available online in the symposium booklet (http://symposium.iscbsc.org/content/booklet).

Keynotes

The day began with a keynote by Dr. Matthew Hibbs, who linked discrepancies between RNA and DNA, known as RNA-editing, to micro-RNA (miRNA) via stem cell development. According to Dr. Hibbs, RNA-editing events lead to novel miRNAs and functional relationship networks in the developing mouse embryo are quite different from cell to cell. He also introduced the StemSight website as a means of visualizing these differences.

After the lunch break, Dr. Jonathan Eisen regaled us with stories of his graduate career, encouraged us to keep our science open-access, take advantage of random opportunities, "what you don't know can hurt you," and how phylogeny is a driving force behind bioinformatics.

* Correspondence: alexander.goncearenco@uni.no
[1]Computational Biology Unit, Department of Informatics, University of Bergen, Bergen, Norway
Full list of author information is available at the end of the article

Mapping Genomic Alterations to Functional Profiles of Pathway Activation, Gene Dependency and Drug Sensitivity

O. B. Botvinnik[1,2,3,4,*], J. W. Kim[1,5,*], C. Birger[1], J. Rosenbluh[1,5], Y. Shrestha[1,5], M. Abazeed[1,6], P. S. Hammerman[1,7,8], O. Abudayyeh[1,5,8], D. DiCara[1], D. J. Konieczkowski[1,5], C. Johannessen[1,5], A. R. Alizad-Rahvar[10], G. Alexe[1,11,12,13], A. Aguirre[1,5], M. Ghandi[1], H. Greulich[1,5,14], F. Vazquez[1,5], B. A. Weir[1], E. M. Van Allen[1,5], A. Liberzon[1], A. Tsherniak[1], D. D. Shao[1,5], T. I. Zack[1,17,18], M. Noble[1], G. Getz[1], R. Beroukhim[1,5,14,17], L. A. Garraway[1,5,14], M. Ardakani[10], C. Romualdi[9], G. Sales[9], D. A. Barbie[1,5], J. Boehm[1], W. C. Hahn[1,5,14,15], J. P. Mesirov[1] and P. Tamayo[1].

1. Eli and Edythe Broad Institute of Harvard and MIT, Cambridge MA.
2. Bioinformatics and Systems Biology Program, University of California at San Diego, La Jolla, CA
3. Department of Cellular and Molecular Medicine, University of California at San Diego, La Jolla, CA.
4. Stem Cell Program and Institute for Genomic Medicine, University of California at San Diego, La Jolla, CA.
5. Department of Medical Oncology, Dana-Farber Cancer Institute, Boston MA.
6. Translational Hematology and Oncology Research, Cleveland Clinic, Cleveland, OH.
7. Department of Medicine, Dana Farber Cancer Institute, Boston MA.
8. Harvard Medical School, Boston MA.
9. Department of Biology, University of Padova, Via Ugo Bassi 58/B, 35121 Padova, Italy.
10. Department of Electrical and Computer Engineering, University of Alberta, Edmonton, AB, Canada.
11. Department of Pediatric Oncology, Dana-Farber Cancer Institute, Boston MA.
12. Boston Children's Hospital, Boston MA.
13. Bioinformatics Graduate Program, Boston University, Boston MA.
14. Department of Medicine, Brigham and Women's Hospital, Boston MA.
15. Center for Cancer Genome Discovery, Dana-Farber Cancer Institute, Boston, MA.
16. Harvard-MIT Health Sciences and Technology, MIT, Cambridge, MA.
17. Department of Cancer Biology, Dana-Farber Cancer Institute, Boston MA.
18. Program in Biophysics, Harvard University, Boston, MA.

* These authors contributed equally to this work.

Abstract

Systematic efforts to sequence the cancer genome have identified many of the recurrent mutations and copy number alterations in tumors. However, the role played by these alterations is not obvious and necessitates an effective functional characterization of the pathways and networks that these genomic alterations regulate. Here we introduce REVEALER, an analysis method that enables the discovery of an ensemble of mutually exclusive or complementary genomic alterations correlated with "functional" phenotypes, e.g., the activation or dependency of oncogenic pathways. We use REVEALER to identify a small number of genomic alterations that account for a large fraction of the "activated" or "dependent" samples with respect to four targets: the transcriptional activation of β-catenin and NRF2, MEK-inhibitor sensitivity and KRAS (RNAi) dependency. REVEALER was able to "re-discover" several known features, as well as identify a number of novel findings, demonstrating the power of combining functional profiles and extensive characterizations of cancer genomes.

Science, meet productivity

How to be fabulously successful: Eric Lander

Eric Lander, the visionary behind the Human Genome Project, a Professor of Biology at MIT and Systems Biology at Harvard, recipient of a MacArthur "genius" grant, and director of a world-renowned genome research institute, was recently interviewed for the New York Times newspaper. As a computational biologist and an avid fan of his, I devoured the article and hungrily watched the video interview. His story is remarkable: genius mathematician, frustrated with the solitude of mathematics, stumbles upon biology and changes the field forever by applying math to understand genes and disease.

At first, it seems like bunch of chance circumstances that are completely unique to Prof. Lander, such as teaching economics at Harvard, exposing him to a brilliant scientific community that supported his endeavors in biology. Or happening to attend the one talk at the one conference where the idea of mapping all the genes in the human genome was first suggested. But if I aim to someday impact bioinformatics research at the same level as Lander, I want to know exactly what he did, day by day, to accomplish all that he has. So I delved deeper.

1. Be stubborn through struggle

Watch the below interview with Eric Lander from the New York Times (link).

The New York Times

More Video »

Eric Lander

Lander emphasizes the social aspect of science, which as a computational biologist, a field which is defined as an intersection of disciplines, I wholeheartedly support collaboration as an integral part of modern science. However, I want to highlight what he said at the end (emphasis mine):

> "A lot of times in my life, when I've been casting about, trying to figure out what's the next step, what's bothering me, where to go. In a way, it's not so different from working on a math problem, where if you try to take it head on, you usually can't make progress. But when you take on a challenge like 'what do I want to go do now?' or 'what's the field got to do right now?' or 'what's the big issue in cancer,' whatever. **You keep struggling with it, and eventually, the structure of the problem becomes clear**. And then **the path through it becomes clear**. But all those **moments of insight come from long periods of casting about and seeing all the pieces**. You just have to be **patient enough to wait until all the pieces really come together**. You gotta be stubborn."

This last note is exactly what I mean by "constancy and moderation."Here is a man who is, by all standards, is a genius. And yet he describes **struggling** with a problem. Really getting to know it, spending time with this problem and thinking about all its parts. This can be done alone or collaboratively, but the struggle must happen. In Outliers by Malcolm Gladwell, he discussed case studies of seemingly chance successes only to uncover the pattern of consistent opportunities to practice. For example, the Beatles performed *eight hours a night* at a club in Hamburg, Germany. Eight hours! Before, they had prepared an hour's worth of material for gigs and initially, the group struggled to find enough to play for such a long period of time. These huge stretches forced the group to constantly change their sound and have long, dynamic set lists to keep their audience engaged. They struggled through eight hour set lists and eventually became the legendary band they are today.

Cal Newport, author of Study Hacks, has written about the value of hard focus such as this,

which Lander exemplifies in his description of struggle. As a result of his struggles, Lander attained huge amounts of practice in mathematics as evidenced by his high school math team and leading up to his doctorate in mathematics, which qualified him to teach at an major research institution such as Harvard and be surrounded by experts in every possible discipline. **The "chance circumstances" Lander experienced were as a result of his grit and persistence through mathematics.**

And then, there are some **moments of insight**, which come from these struggles, come from putting in the time and energy onto a problem. The Beatles improvised, invented, and found their sound that led to them to become one of the most successful and prolific bands of all time. Lander worked with colleagues to invent new methods of analyzing biological information.

2. Have a reasonable workday

Let's dig deeper into how he actually accomplishes this. From the accompanying article, Lander's workday doesn't seem so remarkable:

> "After his morning workout, he sometimes goes to a local bakery where he can work quietly. He arrives at the Broad between 8 and 10 a.m. In the fall, he teaches introductory biology to a class of 700 M.I.T. students on Monday, Wednesday and Friday mornings. He often meets with graduate students and postdoctoral fellows in the afternoon to discuss their work.
>
> Then he has his administrative duties and his meetings with philanthropists, trying to raise more money. He also spends 20 percent of his time in yet another role, as co-chairman of President Obama's Council of Advisers on Science and Technology, which deals with topics like influenza vaccines, health information technology, science education and energy policy.
>
> In the evening, around 6:30 or 7, he has dinner with his family."

So how does he accomplish such remarkable things? Despite the apparent normality of this schedule, it has several gems.

3. Hard focus first

First, I'm guessing he uses the morning bakery time to "*work quietly*" on his hardest task. The one where he must make a creative leap from old paradigms to new ideas, which requires a great deal of focus and energy. This is in line with Leo Babauta, author of Zen Habits, advises to work on your most important thing first thing in the morning, before any interruptions might come up. Lander does exactly this. He could be writing a manuscript, thinking about a problem that a graduate student proposed, or preparing a brand-new lecture. He's refreshed after his morning workout and probably getting some calories and caffeine - a great recipe for focus.

In any case, Lander's definitely not checking his facebook or email first thing in the morning. And if you want to be successful like Eric Lander, you shouldn't, either.

Second, he then does less "hard focus" tasks. He then teaches, which is probably building from his many previous lessons, so he's well prepared. Then he meets with graduate students/post-

docs/collaborators, which is certainly intellectually stimulating but probably doesn't require quite as much energy as his morning task. If him and his colleagues hit a wall, he can always think about it in the next morning or meet up again later.

Finally, he gets some rest. He finishes his day at a reasonable time and has dinner with his family. Connecting with his family stimulates his social mind and rejuvenates his analytical brain. He doesn't expect to work late into the night, because he knows he already accomplished his most important thing in the morning. He touched base with his graduate students and connected with his colleagues.

Day by day, and little by little, Lander builds his success through daily focus, struggle, and he does it for those moments of insight that transform a field forever. And that's what I'll be doing.

- #advice
- #bioinformatics
- #case study
- #computational biology
- #eric lander
- #hard focus
- #most important thing
- #productivity

- 1 year ago
- 8
- 0 Comments
- Permalink
- Share

About


I'm Olga Botvinnik, a cello-playing, loose-leaf tea drinking, japanese pen-wielding matplotlib whisperer and bioinformatics and systems biology PhD student at UCSD.

Pages

- olgabotvinnik.com
- R Bloggers

Me, Elsewhere

- @olgabot on Twitter

olgabot / prettyplotlib

Painlessly create beautiful matplotlib plots. http://olgabot.github.io/prettyplotlib — Edit

90 commits **3** branches **0** releases **4** contributors

branch: **master** prettyplotlib /

Merge pull request #12 from gotgenes/python3-compatibility ...

olgabot authored 15 days ago latest commit e33c1e1d78

examples	fixed links on front page, and all plotting functions now return the …	a month ago
prettyplotlib	Replaces .next() method calls with function calls.	15 days ago
tests	Python 3 compatibility for uppercase/lowercase characters.	15 days ago
.gitignore	removed egg info	a month ago
CHANGES.txt	started writing tests	4 months ago
LICENSE.txt	started writing tests	4 months ago
MANIFEST.in	reuploading for oolite	3 months ago
README	made soft links for README.md to README and README.txt	3 months ago
README.md	fixed links on front page, and all plotting functions now return the …	a month ago
README.txt	made soft links for README.md to README and README.txt	3 months ago
setup.py	Updates setup.py to indicate Python 3 compatibility.	15 days ago

README.md

prettyplotlib

Python matplotlib-enhancer library which painlessly creates beautiful default `matplotlib` plots. Inspired by Edward Tufte's work on information design and Cynthia Brewer's work on color perception.

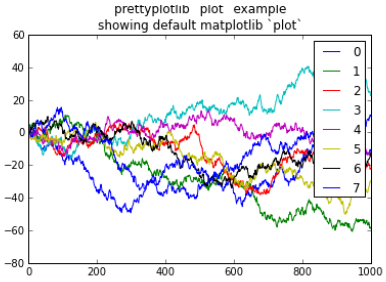
I truly believe that scientific progress is impeded when improper data visualizations are used. I spent a lot of time tweaking my figures to make them more understandable, and realized the scientific world could be a better place if the default parameters for plotting libraries followed recent advances in information design research. And thus `prettyplotlib` was born.

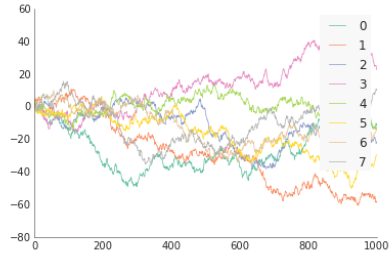
Requirements:

- `matplotlib`. Can be installed via `pip install matplotlib` or `easy_install matplotlib`
- `brewer2mpl`. Can be installed via `pip install brewer2mpl` or `easy_install brewer2mpl`

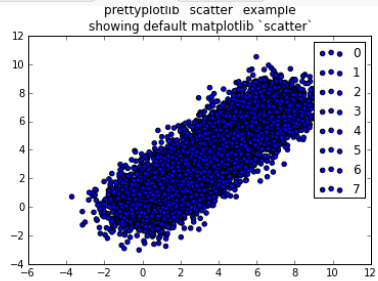
Comparison to `matplotlib`

`matplotlib` default `plot`	`prettyplotlib` default `plot`

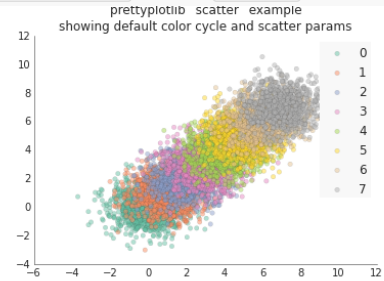




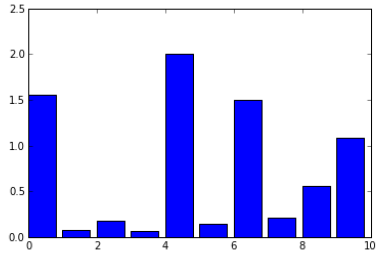
`matplotlib` default `scatter`



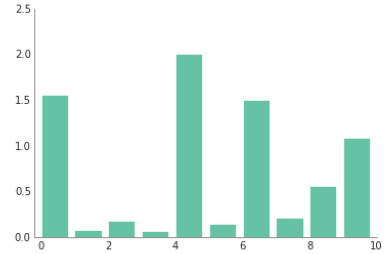
`prettyplotlib` default `scatter`



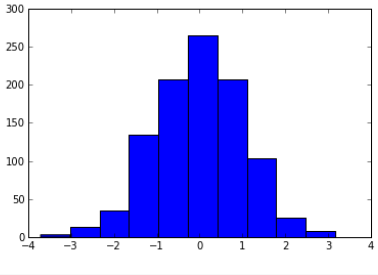
`matplotlib` default `bar`



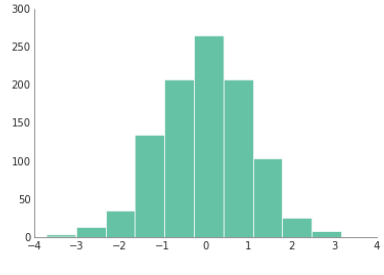
`prettyplotlib` default `bar`



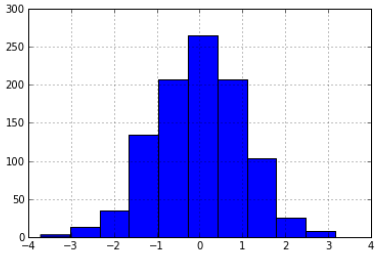
`matplotlib` default `hist`



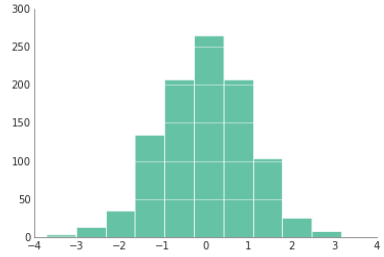
`prettyplotlib` default `hist`



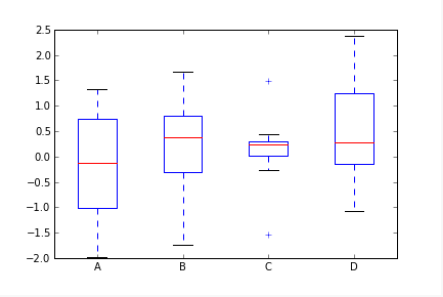
`matplotlib` default `hist`
with grid



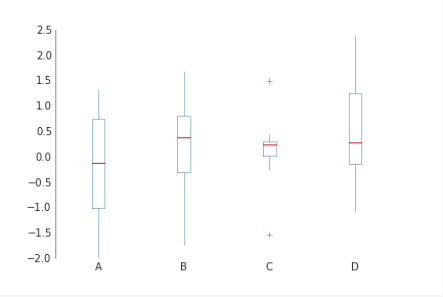
`prettyplotlib` default `hist`
with grid



`matplotlib` default `boxplot`

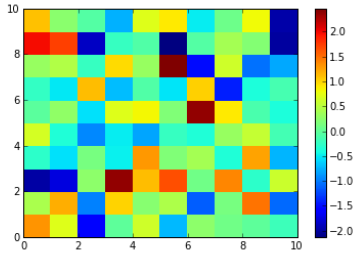


`prettyplotlib` default `boxplot`



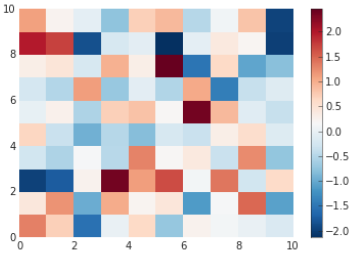
`matplotlib` default `pcolormesh`

positive and negative data



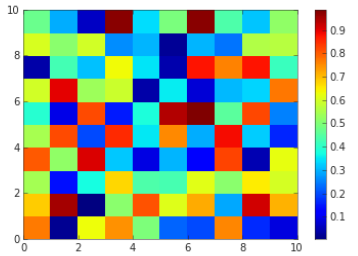
`prettyplotlib` default `pcolormesh`

positive and negative data



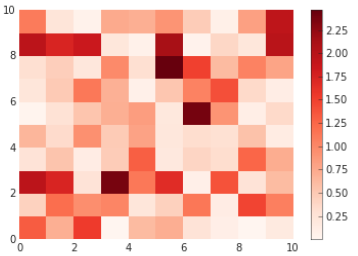
`matplotlib` default `pcolormesh`

positive data only



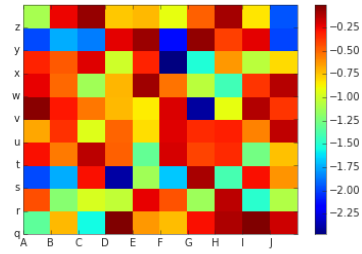
`prettyplotlib` default `pcolormesh`

positive data only



`matplotlib pcolormesh`

negative-valued data with labels



`prettyplotlib pcolormesh`

negative-valued data with labels