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

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

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

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

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.

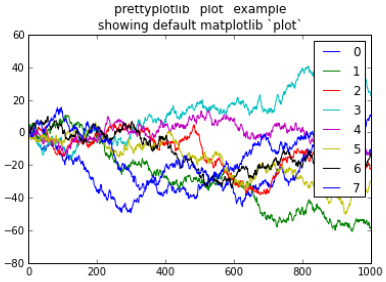
I truly believe that scientific progress is impeded when improper data visualizations are used. I spent a lot of time tweaking my figures to make them more understandable, and realized the scientific world could be a better place if the default parameters for plotting libraries followed recent advances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

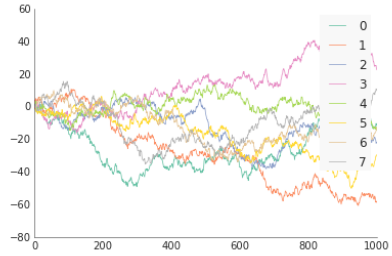
Requirements:

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

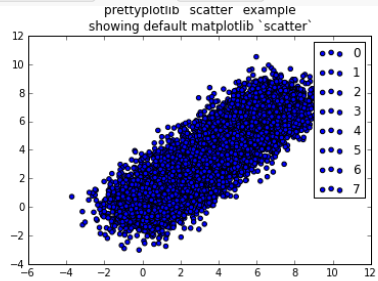
Comparison to `matplotlib`

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

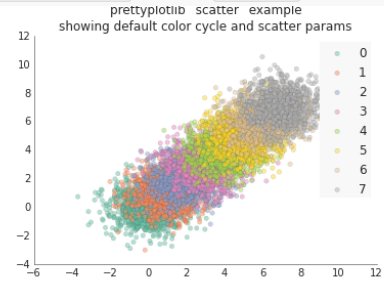




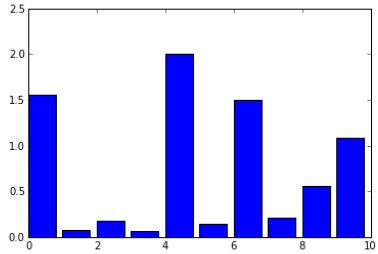
`matplotlib` default `scatter`



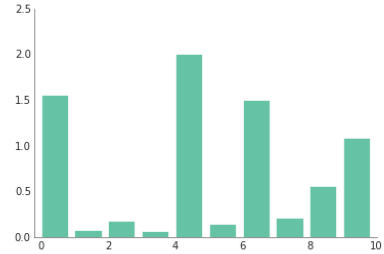
`prettyplotlib` default `scatter`



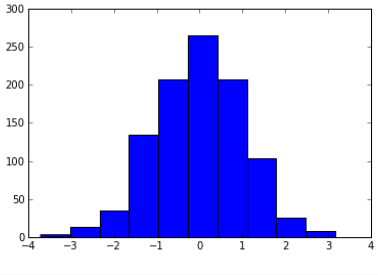
`matplotlib` default `bar`



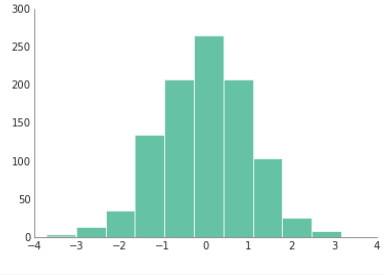
`prettyplotlib` default `bar`



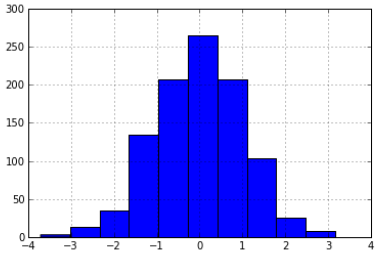
`matplotlib` default `hist`



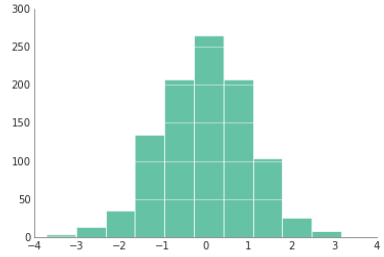
`prettyplotlib` default `hist`



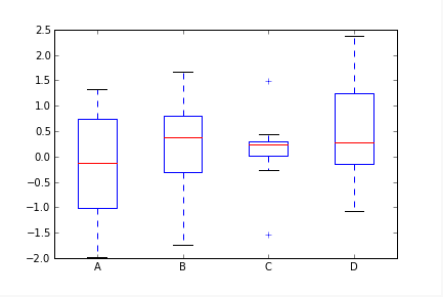
`matplotlib` default `hist`
with grid



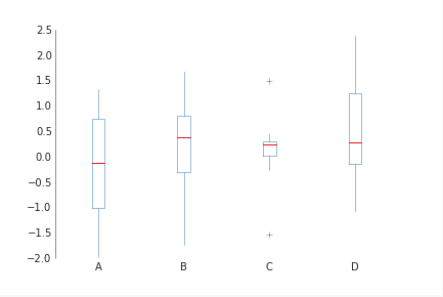
`prettyplotlib` default `hist`
with grid



`matplotlib` default `boxplot`

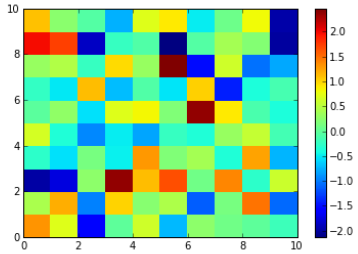


`prettyplotlib` default `boxplot`



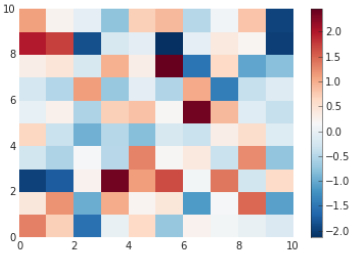
`matplotlib` default `pcolormesh`

positive and negative data



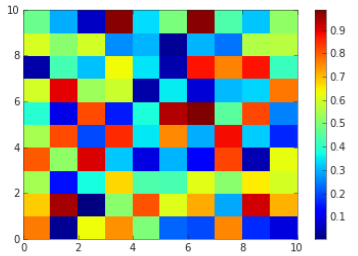
`prettyplotlib` default `pcolormesh`

positive and negative data



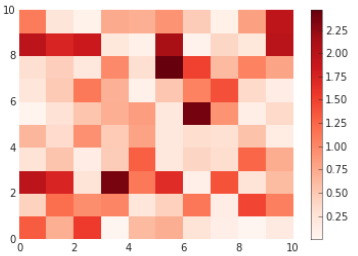
`matplotlib` default `pcolormesh`

positive data only



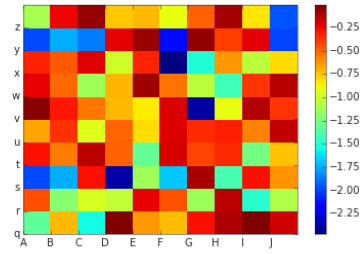
`prettyplotlib` default `pcolormesh`

positive data only



`matplotlib pcolormesh`

negative-valued data with labels



`prettyplotlib pcolormesh`

negative-valued data with labels