

O'REILLY®

# BioCoder

BIO NEWSLETTER

SEPTEMBER 2016



---

## A Biotech Wave at SXSW

Karen Ingram

---

## miniPCR: Enabling the Era of Personal DNA

Zeke Alvarez Saavedra & Sebastian Kraves

---

## Are Your Ribosomes in a Twist?

Danfox Davies

---

## Design.bio

Orkan Telhan

---

# Did you enjoy this issue of BioCoder?

Sign up and we'll deliver future issues  
and news about the community for FREE.

<http://oreilly.com/go/biocoder-news>



# BioCoder #11

SEPTEMBER 2016

Beijing • Boston • Farnham • Sebastopol • Tokyo

**O'REILLY®**

## BioCoder #11

Copyright © 2016 O'Reilly Media, Inc. All rights reserved.

Printed in the United States of America.

Published by O'Reilly Media, Inc., 1005 Gravenstein Highway North, Sebastopol, CA 95472.

O'Reilly books may be purchased for educational, business, or sales promotional use. Online editions are also available for most titles (<http://safaribooksonline.com>). For more information, contact our corporate/institutional sales department: 800-998-9938 or [corporate@oreilly.com](mailto:corporate@oreilly.com).

**Editors:** Mike Loukides and Nina DiPrimio

**Interior Designer:** David Futato

**Production Editor:** Colleen Lobner

**Cover Designer:** Randy Comer

**Copyeditor:** Amanda Kersey

**Illustrator:** Rebecca Demarest

September 2016: First Edition

### Revision History for the First Edition

2016-09-01: First Release

The O'Reilly logo is a registered trademark of O'Reilly Media, Inc. *BioCoder #11*, the cover image, and related trade dress are trademarks of O'Reilly Media, Inc.

While the publisher and the authors have used good faith efforts to ensure that the information and instructions contained in this work are accurate, the publisher and the authors disclaim all responsibility for errors or omissions, including without limitation responsibility for damages resulting from the use of or reliance on this work. Use of the information and instructions contained in this work is at your own risk. If any code samples or other technology this work contains or describes is subject to open source licenses or the intellectual property rights of others, it is your responsibility to ensure that your use thereof complies with such licenses and/or rights.

978-1-491-97133-8

[LSI]

# Contents

Foreword | v

miniPCR: Enabling the Era of Personal DNA	i
Are Your Ribosomes in a Twist?	13
Building Artificial Connectomes	21
BioBright	29
Design.bio	33
Where Science-as-a-Service and Supercomputing Meet	45
Blockchain-Enabled Open Science Framework	51
Predicting Cures	57
A Biotech Wave at SXSW	61
HiveBio: Access Granted	69



# Foreword

*Nina DiPrimio*

Our mission is to publish content on what you are doing to make biology research more accessible, low-cost, efficient, and reproducible. The stories that we chronicle come from innovators in academic labs, startups, and at home, and this issue is no exception. Highlighted in this issue are multiple software tools for optimizing a variety of research processes. Atomwise developed a novel AI tool for drug discovery, BioBright is developing software tools for optimizing a researcher's productivity, Tim Busbice explores the idea of building artificial connectomes for multiple functionalities, and Danfox Davies introduces his software tool for the prediction of alternative start sites in protein coding and how that will help predict pathway functions.

We are excited to include an update by MiniPCR and how their apparatus is being used to detect Ebola, genotype truffles, conduct experiments on the ISS, and much more. Additionally, Orkan Telhan discusses Microbial Design Studio, built to facilitate the use of biology in design, and also announces Design.bio, a portal for sharing workflows.

Sindhu Ravuri interviewed Russell Poldrack of the Stanford Center for Reproducible Neuroscience and examines how they are revolutionizing data sharing and analysis in neuroscience, HiveBio updates us on their community lab, Karen Ingram reviews the evolution of biology at SXSW, and Vikram Dhillon researches the potential use of blockchain for open science.

To better align with our goal of sharing these stories with you, we transitioned from not only publishing in this quarterly magazine, but also publishing content on [our new O'Reilly Bio Ideas page](#) for easier communication of your projects and ideas. Additionally, we decided to publish content on a more frequent basis. Please keep the articles coming, we love sharing your stories! Contact us at [biocoder@oreilly.com](mailto:biocoder@oreilly.com) if you are interested in contributing.



# miniPCR: Enabling the Era of Personal DNA

*Zeke Alvarez Saavedra, PhD, and Sebastian Kraves, PhD*

Three years ago, we set out to make personal DNA tools that everyone could use. miniPCR started as a maker project, and after a design and prototyping phase, we launched the DNA Discovery System on Kickstarter in 2014. We exceeded our funding goal threefold, which enabled us to scale up production and mount distribution logistics. We have since built a team of biologists, designers, and engineers that operates out of the Harvard Innovation Launch Lab, a community of entrepreneurs and innovators on Harvard University's dynamic Allston Campus. Manufacturing operations are located just outside of Boston, from where we've now shipped DNA Discovery Systems to hundreds of researchers, educators, and DNA curious individuals in every continent.

The miniPCR DNA Discovery System is a portable, personal DNA lab that anyone can use. It contains the fundamental tools of DNA analysis: a PCR thermal cycler, a gel electrophoresis apparatus with a built-in blue-light transilluminator, and a variable volume micropipette. It is portable, user friendly, operable through smartphones or computers, and is priced at \$990. Its components are also available individually so users can build-to-fit their DNA labs. For example, lab-based researchers may need greater PCR capacity and educators may chose to start with the simpler techniques of DNA gel electrophoresis and visualization. Field-based users may opt to power miniPCR with rechargeable batteries, and couple it to downstream sequencers foregoing gel electrophoresis components. The miniPCR DNA Discovery System (pictured in [Figure 1-1](#)) is truly a modular ecosystem that can be configured to the user's needs.



Figure 1-1. The miniPCR DNA Discovery System.

When we first shared our story with the Biocoder community (fall 2014 issue), we didn't know the speed at which miniPCR and the field in general would move. We started miniPCR driven less by a specific goal than by sheer curiosity: what would the world look like if everyone had access to DNA science? What would people use miniPCR for? We were excited about enabling new possibilities, about what might happen if we removed barriers to experimentation, expanding channels for scientific innovation. Over the last three years, miniPCR has been in the hands of hundreds of new users. In this article, we share the stories of some of our inspiring users to illustrate the explosion of creativity and innovation that they are leading. There are many more stories that we'd like to share, but just won't fit the page.

## Reimagining Animal Health

Imagine you're a pig farmer (Figure 1-2). You live on a small farm, in the Philippines. Your animals are your family's sole source of income, as long as they're healthy.





*Figure 1-2. Reimagining animal health.*

You know that any day one of your pigs can catch the flu—the swine flu. Living in tight quarters, one pig coughing and sneezing can quickly lead to the next pig coughing and sneezing, until an outbreak of swine flu has taken over your farm. If it's a bad enough virus, the health of your herd can be gone in the blink of an eye. If you called in a veterinarian, she would visit your farm, take samples from your pigs' noses and mouths, and drive back to the city to analyze those samples in the lab. Two weeks later, you'd learn the results. Two weeks might be just the time it takes for infection to spread and take away your way of life.

But it doesn't have to be that way. Today farmers can collect those samples themselves. They can jump right into the pen and swab their pigs' mouths with a little filter paper. Then, they can take those paper swabs, place them in a test tube, and mix them with chemicals that help isolate genetic material from their pigs' mouths. Without leaving their farms, they can place a drop of genetic material into a little analyzer smaller than a shoebox, program it to detect DNA or RNA from the flu virus, and within one hour farmers can visualize the results.

## A Personal DNA Revolution

This reality is possible because we're living in an era of personal DNA technology ([Figure 1-3](#)). Several groups around the world are working to put DNA technology

in the hands of everyday people. This movement has been defined under different labels: biohackers, do-it-yourself (DIY) biologists, citizen scientists, etc. Behind these labels lies a common direction of change: through accessible DNA analysis tools, every one of us can test and tinker with DNA ourselves, in our basements and kitchens.



*Figure 1-3. A personal DNA revolution is underway.*

DNA is strung into a long and twisted double helix that in humans has a total of three billion base pairs beginning to end. The lines that carry genetic information, also known as genes, can be as short as a few dozen to several thousand base pairs long. So to find useful genetic information, such as DNA that can answer, “Is my pig sick?” we typically don’t need to read all three billion bases. That would be like craving pizza at night and having to read the phone book from cover to cover, pausing at every line, in search for the nearest pizza joint.

Three decades ago, humans started to invent tools that can very quickly find a specific chunk of genetic information, no matter how small. In a test tube, they can find any given line of DNA. But that DNA is still very small and surrounded by a lot of other DNA. In order to make that specific gene stand out above the rest, we can make copies of the gene, in a process known as the polymerase chain reaction (PCR). PCR machines (also known as thermal cyclers or thermocyclers) can very quickly make billions of copies of a target DNA sequence, and the copies accumulate until we can finally see that target piece of DNA, detect it, read it, and understand it (Figure 1-4). The process is a way to answer questions not only

about swine flu in pigs but also those encoded in our own DNA: Am I at risk for ovarian cancer? Am I of Irish descent? Is this child my son?

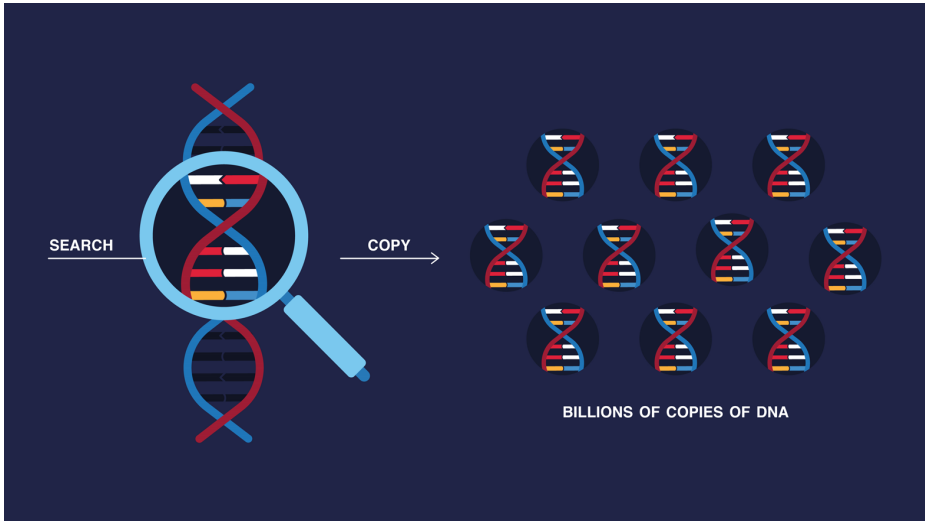


Figure 1-4. PCR is a method for finding and copying specific segments of DNA.

This ability to quickly make millions of copies of DNA by PCR, as simple as it sounds, has transformed our world. PCR machines help scientists detect disease, modify crops, decide when food is safe to eat or when it's contaminated with deadly bacteria. They help judges decide who goes to jail and who is released based on DNA evidence. For this invention, Kary Mullis was awarded the Nobel Prize in Chemistry in 1993.

Today, this power is accessible to every one of us, as PCR and other DNA technologies become increasingly available outside of professional labs.

## The Truffle Farmer That Needed DNA Analysis

People often ask us where this revolution in accessibility might take us, or how the world will change. However, it is notoriously hard to predict how new technology will transform society. For example, when we first developed miniPCR, we could not have predicted that a truffle farmer would have a use for DNA analysis. Dr. Paul Thomas grows truffles for a living, and we see him pictured in [Figure 1-5](#) holding the first UK-cultivated truffle.



*Figure 1-5. Dr. Paul Thomas holding the first UK-cultivated truffle. (Photo courtesy of Dr. Paul Thomas.)*

Truffles are a delicacy that stems from a fungus growing on the roots of living trees, and it's a really rare fungus: some species may fetch between \$3,000 to \$7,000 per kilogram, or even more. We've learned from Thomas that the stakes for a truffle farmer can be really high: When he sources new truffles to grow on his plantations, Thomas is exposed to knock-offs that even to an expert's eye can look and feel like the real thing. Even under a microscope, these lower quality truffle spores can look authentic. So in order to select the right truffles to grow on his farm, the ones that chefs all around the world will fight over, Thomas has to use DNA analysis. This is mind-blowing, and it's hard now for us to look at a black truffle risotto without thinking of its genes.

## Conservation Biology: Taking the DNA Lab into the Field

Personal DNA technology also means that professional scientists can now take their labs into the field. Elaine Guevara is pursuing her PhD in anthropology from Yale. She's interested in primate evolution and spends significant time in South-western Madagascar, studying a colony of sifaka lemurs on the Beza Mahafaly Reservation ([Figure 1-6](#)).



Figure 1-6. Mobile DNA lab at the Beza Mahafaly Reservation in Madagascar. (Photo courtesy of Elaine Guevara.)

This colony has only 700 lemurs left. So few that these lemurs now share a lot of genes in common. And if they keep inbreeding, they may face health risks or even extinction. To help prevent that, Guevara is monitoring their genetic diversity; but getting a bunch of endangered lemurs on a plane so you can study their genes back at Yale is not really a viable option—imagine the hassle at customs. With miniPCR, Guevara can now make billions of copies of these lemurs' genomes right in the Madagascar forest, powered by a simple solar battery, and bring back to her lab in the US just test tubes that contain the lemurs' complete genetic information via whole-genome amplification (WGA).

## Saving Lives: Ebola Detection in Makeni, Sierra Leone

We can also use personal DNA technology to save human lives. In 2015, Cambridge University professor Ian Goodfellow traveled to Sierra Leone to try and diagnose patients with Ebola (Figure 1-7). Doctors there were limited in what they could do; test results could take more than a week to come back, by which time it was too late for those suffering from the disease—and their relatives.





*Figure 1-7. Setting up an Ebola diagnostics and sequencing lab in Makeni, Sierra Leone. (Photo courtesy of Ian Goodfellow.)*

So Goodfellow decided to move his lab to Makeni—that’s him with 10 tons of equipment in the picture—and set it up in a pop-up tent which when fully functional, would be able to turn samples around in 24 hours or less. Only one problem: the high-end DNA analysis equipment didn’t work like it would under controlled lab conditions in the UK. We’re talking 35°C heat and over 90% humidity in Makeni. So instead, he used miniPCR machines, small enough to be placed in front of the air conditioning unit, to keep the virus sequencing workflow going, and keep saving lives.

## Genes in Space: DNA Analysis Beyond Earth

Goodfellow’s pop-up lab in Makeni may seem like an extreme place for DNA testing, but let’s travel to an even more extreme place, beyond Earth. Let’s talk about DNA analysis in outer space. Astronauts live aboard the International Space Station, 250 miles above the Earth, orbiting the planet at 17,000 miles per hour, so they often see 15 sunsets and sunrises every day. They live in microgravity, where human bodies can start to do weird things. For example, when astronauts are in space for an extended period of time, their immune systems get weakened, making them more prone to infection.

Why do astronauts experience immune system suppression? Could epigenetic changes be the cause? These were the questions that a high-school student from New York, Anna-Sophia Boguraev, wanted to answer. Anna-Sophia is one of hundreds of students who have participated in Genes in Space, a science competition organized by miniPCR and partners (Boeing, the Center for the Advancement of Science in Space, New England Biolabs, and Math for America).

Anna-Sophia designed a DNA experiment that uses miniPCR to test her hypotheses about immune suppression aboard the International Space Station. On April 8, 2016, miniPCR and Anna-Sophia's experiment were launched to space from NASA's Kennedy Space Center in Cape Canaveral. [Figure 1-8](#) shows Anna Sophia watching her experiment launch, and the stream of smoke is indeed the rocket that brought miniPCR to the International Space Station.



*Figure 1-8. Anna-Sophia Boguraev watches her DNA experiment launch to the International Space Station from Cape Canaveral on April 8, 2016. (Photo credit: miniPCR.)*

A few days later, astronauts carried out the first DNA amplification experiment in space, and British astronaut Tim Peake conducted Anna-Sophia's experiment aboard the Space Station ([Figure 1-9](#)). We are now in the process of analyzing the results (we know that PCR in space was successful!). miniPCR is now aboard the International Space Station, where it can help monitor living conditions and protect the lives of astronauts. Its efficient design allowed miniPCR to open up a new era of genetics in space.



*Figure 1-9. Astronaut Tim Peake conducted Anna-Sophia's experiment aboard the ISS using miniPCR. (Photo by NASA.)*

## DNA Knocking on Everyone's Doors

A 17-year-old student designing a DNA analysis experiment to protect the lives of astronauts may sound like a rarity, or the mark of a child genius. But to us, this signals something bigger, that DNA science is finally becoming within the reach of everybody. Just like a few years ago, a college student armed with a personal computer could code an app in his dorm, an app that is now a social network with more than one billion users.

Could we be moving into a world of one personal DNA machine in every home? We already work with families that are living in such a reality. The Daniels family, for example, set up a DNA lab in the basement of their suburban Chicago home last summer (Figure 1-10). This family isn't made up of PhD scientists. Instead, they are like many other families who are looking for fun, creative ways to spend time together.





Figure 1-10. The Daniels Family set up a DNA lab in their suburban Chicago home. (Photo courtesy of C. Bryan Daniels.)

In fact, the dad of this family, Bryan, said all he wanted to do was to “keep his kids off the Xbox this summer,” so he replaced the Xbox for the personal DNA box. By day, Bryan is an executive at a private equity firm. On nights and weekends, he plays with miniPCR alongside his kids aged 7 and 9, as a way to explore the living world. Last time I spoke with them, they were testing home-grown produce from their backyard garden. They had picked fresh tomatoes, mixed them with a simple DNA extraction solution, and with their home DNA copier, they were ready to find out whether these tomatoes had genetically engineered traits. For them, the DNA Discovery System is the equivalent of the chemistry set for the 21st century.

## Are You DNA Curious?

In 2016, most of us may not yet be diagnosing genetic conditions in our kitchen sinks, or doing at-home paternity testing, but we’ve reached a point in history where you, working in your own kitchen, can take DNA into your own hands and copy it, paste it, analyze it, and extract meaningful information from it.

Explosive innovation is bound to happen. A powerful, transformative technology that was only available to a select few in the ivory tower can now be in your

hands, and in the hands of everyone, from farmers to school children. Think about when telephones stopped being attached to the wall via cords, or when computers left the mainframe and entered everyone's offices and homes. The ripples of this personal DNA revolution may be unexpected. But here's the thing about revolutions: they don't go backward. Access to DNA technology is spreading faster than our imagination. So if you're curious, get up close and personal with DNA today. It's in our DNA to be curious.

---

*Ezequiel (Zeke) Alvarez Saavedra, PhD, is a cofounder at miniPCR and at Genes in Space. Zeke is a geneticist trained at MIT and Stanford. He has conducted biomedical research alongside two Nobel laureates, and his work has been cited thousands of times and profiled in national and international media such as the New York Times, National Public Radio, and the BBC. He is also an inventor of gene-detection technologies. In his spare time, Zeke digs soil in search of new species (one so far!) He can be reached via email at [zeke@minipcr.com](mailto:zeke@minipcr.com).*

*Sebastian Kraves, PhD, is a cofounder at miniPCR and at Genes in Space. Sebastian is a molecular neurobiologist trained at Harvard. Previously he was a principal with The Boston Consulting Group (BCG), where he helped make biomedical technology accessible to low-resource settings for global health. He has published widely cited work on neural circuits and the genetic regulation of behavior. Sebastian first dreamed of becoming a biologist at age 12 after reading Darwin's *The Voyage of the Beagle*. You can email Sebastian at [seb@minipcr.com](mailto:seb@minipcr.com).*

# Are Your Ribosomes in a Twist?

**MAKING PROTEINS THE RIGHT WAY CAN BE A CHALLENGE**

*Danfox Davies, Vulpine Designs*

Tracing the intracellular contraptions of life is essential to understanding how to debug our life code. Alas, without a software development kit, API, or even a manual--let alone easy-to-understand source code--we have to reverse-engineer the machine code of living things. That's the only way to find the causes of genetic diseases and to not just fix them, but to do so without breaking other parts in the process. Current methods tend to involve testing on genetically altered animals and cultured cells in long and expensive studies to find out if a certain applied change has the expected effects on protein composition, function, and organization, and to infer from this the way the system works. This can seem to even the bravest of minds like trying to fathom a recursive Rube Goldberg or Heath Robinson contraption. In this article we look at the cellular equivalent of compiler problems and the ways my company is trying to help.

## In Tents

All biological cells have proteins to function. Proteins are strings of amino acids, folded up a particular way that serves a purpose, a bit like a pop-up tent. You can fit a pop-up tent in your car boot when it's folded away; but when you reach the campsite, you need to have it quickly pop up in the rain so you can peg it down and get camping. When you get home, to allow it to dry properly, you might pop it up again but not peg it down. But if you store it without its proper bag (perhaps the bag was not included with this tent because of a mistake at the factory) and other objects get crammed in there with it, you don't want it to get stuck in a twisted shape, out of alignment and tangled up. This could cause it not only to be damaged, but to damage and get tangled with other things around it, misaligned poles and rods finding their ways to their lowest energy configurations.

With proteins, getting stuck in the useless lowest energy configuration can be a major problem. Amyloids and prions (which are among the causes of neurodegenerative diseases) are deposits of misfolded proteins that destroy any other proteins that come into contact with their ever-growing tangle. Heat, pH, and other molecules can cause some proteins to get deformed (denatured). Indeed, if something vital to keeping the protein well shaped for when it might be needed is missing or the conditions don't allow the right kind of fold, its deformation is all the more likely, and the cell might not be able to do anything more with it (Chiti and Dobson 2006).

## I Fold

Calculating protein folding and errors thereof has been the preserve of Stanford University's [Folding@Home](#) (Larson et al. 2002), [FoldIt](#) (Cooper et al. 2010), and Samsung's [PowerSleep](#) (Anonymous 2014) for many years, using crowd computing to make an impossibly huge task possible. (More recently incentivized by CurieCoin [Cygnus-Xi and Vorksholk 2014]. The recent advancements in blockchain usage for financial applications [the most famous of these being Bitcoin] have led to a variety of 'proofs of work' being selected for transaction verification. A great mutual benefit was seen in making the calculation of protein folding a proof of work in itself, such that cryptocurrency could be mined in the process of distributed computational scientific endeavor.)

Every protein is specialized to certain tasks, conditions, and cell types. A Mongolian ger and a mountaineering tent are very different in their construction but cope with similar weather. One is meant to be semi-permanent; the other is meant for one or two nights at a given spot. A big top is different again, but they can all be described as tents. Cells have to provide equal versatility, both within the same cell during different conditions and in hundreds of different cell types. Protein folding has to be determined not only by environmental factors or utility, but by genetic and epigenetic instructions. Otherwise there would be too few ways for cells to fold their own proteins, and amyloids would run amok.

## Lost in Translation

Cells often need not just any of the several possible proteins and variants of them that a particular gene theoretically encodes, but the exact protein needed for a particular cell at that time of day, with that amount of glucose, that level of stress and whilst it is in that exact position among other cells, which are at various stages of their lives. That gene must be transcribed and translated according to the correct

setting of a collection of variables in both the DNA and the transcribed mRNA, as well as in the proteins handling, storing, and maintaining them.

Proteins can be coded by the same gene in several different ways at different stages of the protein's production. This helps explain the surprisingly small number of genes humans have, compared to other, less complex species (Lander et al. 2001, Venter 2001, Ezkurdia et al. 2013).

Transcription is the process in which the relevant DNA, which was wrapped around histone proteins that form the chromatin structure, is unspooled to reveal a gene. This is copied to shorter, exportable mRNA strands by the transcription complex. There are many ways in which transcription can selectively be varied within the same gene, which each have effects on the mRNA available for translation. This is an area of much research focus already, and tools are available to handle bioinformatic big data about most aspects of transcription.

Translation is the process by which the Messenger RNA (mRNA) copy is read by the ribosomes, proteins whose job is to assemble proteins based on the instructions in the mRNA. Ribosomes are like factories, which are assembled on demand for protein production. Their assembly and that of the proteins they make, like flat pack furniture, relies on all the lugs and holes aligning, all the folds and magnets and coded labels matching where they are meant to be. mRNA is often derived from a much reduced portion of the original DNA sequence. However, mRNA itself contains sequences that can affect how it folds in the cytoplasm of the cell, how quickly it degrades over time before it has to be remade, and where the ribosome is able to start reading from to make a protein. If part of a notepad is crumpled up, it becomes more physically difficult to straighten and read without mistakes. The ribosome will skip such crumpled parts of the mRNA and start reading at the first sentence that makes sense. That sentence might not normally make the most sense to start from, but the sentence that normally starts things off is unavailable. This is seen in mRNAs as alternative initiation codons (AICs). Codons are sets of three nucleotides (nucleotides being A, G, C, T in DNA and A, G, C, U in RNA) that correspond to amino acids via the transfer RNAs (tRNAs) that ribosomes pick up to build their proteins with. These can be arranged such that the ribosome has multiple choices of starting places on the mRNA and will clamp on (or bind) wherever the most obvious AIC is. If the usually most obvious AIC with the best sequences around it to clamp onto is not available, the ribosome will go for the next best and so forth, with reduced success rates compared to the optimum, until all reasonable possibilities are exhausted. An mRNA made exclusively of 'stop' codons, whose job it is to identify the end of a new amino acid chain and halt the ribosome, would be illogical. Most mRNAs will be used by cells to make something at least on occasion, unless cellular condi-

tions render their shape unreadable. An mRNA made exclusively of 'stop' codons, whose job it is to identify the end of a new amino acid chain and halt the ribosome, would be illogical. Most mRNAs will be used to make something at least on occasion, unless cellular conditions change to the point where some mRNAs are so folded up they can't be read. Therefore, contrary to the assertions of Kozak (1989), on whose 'Consensus' sequences many modern translatomic studies rest, it has been deduced that any codon which is not a stop can be a start, and it's not solely the Kozak-identified contextual codons which define this (Cowan et al. 2014, Ingolia et al. 2011, Lee et al. 2012). This is backed by the experimental data from those last three papers, the scientific details of which are too lengthy for this article.

## Codon, Code-off

If a mutation is found in a genome and seems to be related to cancer, scientists need to know which and how many of the proteins are affected in the cell and why. They will need to know which of those effects are good or bad and what is necessary in terms of treatments, to stop the bad effects without stopping the good ones. If a genetic edit is to be made, will it be possible for it to stop the negative effects of a mutation without causing negative effects of its own? Can the editing method be adjusted? And where on a gene do the AICs reside? Which ones are used by the proteins? We need a way to simulate this accurately if we are to ever cut the time it takes to solve genetic diseases (particularly those caused or affected by multiple mutations), edit genes or trace the effects of mutations. Given the dawn of CRISPR/Cas9 (Qi et al. 2013, Ran et al. 2013), TALENS (Boch 2011) and related methods of genetic editing, addition, and removal, a means to simulate the *effects* of the edits they make will be especially valuable.

The effects of AICs can be phenotype defining, as seen in orchids' *matK* genes, which have been previously called "pseudogenes" due to their different alternative initiation codon usages (Barthet MM et al. 2015a, 2015b). The Wilms' Tumour Suppressor Gene, *WT1*, where novel versions of the protein are produced depending on AICs, affecting where the protein goes and what it ends up doing (Bruening & Pelletier, 1996, Wegrzyn et al, 2008). Some of these can come from a wide variety of AICs affecting the same gene, such as dihydrofolate reductase (Peabody 1989, Wegrzyn et al 2008). In each case, understanding bioinformatically en masse how the AICs change the proteins and why the ribosomes respond differently in different cellular conditions to the same genes' mRNA transcripts will be invaluable (and a lot less time-consuming than present methods) to researchers wishing to avoid the cure being worse than the disease.

This is where we come in. Vulpine Designs is programming a software module called INITIATOR SET, which includes Initmine, based on the Intermine project from the University of Cambridge (Smith et al. 2012). This system maps AIC information to gene sequences, testing each AIC for viability of ribosome assembly, likelihood of mRNA folded structures, protein targeting to organelles etcetera. The more information is characterized, the more the system can be refined. As time goes on, we will add extra featuresx to it. The module will draw its data from existing databases and user input, and will make a collection of all the data on that gene and the effects of it and its transcription and translation.

It is an exciting time for scientists handling all sorts of genetic diseases as we become able to grasp the many variables between genes and resultant proteins and phenotypes. It's a privilege to be part of it, as accessibility of tools and information continues to increase through all our shared efforts. There is a possibility that one day we will be able to handle and design all aspects of the body at any age, and know the effects of our edits to the life code before we apply them.

## References

- Anonymous, 2014. Power Sleep. *Samsung Electronics Austria*, Universität Wien.
- Barthet, M.M., Moukarzel, K., Smith, K.N., Patel, J., Hilu, K.W., 2015a. Alternative translation initiation codons for the plastid maturase MatK: unraveling the pseudogene misconception in the Orchidaceae. *BMC Evol. Biol.* 15. doi:http://dx.doi.org/10.1186/s12862-015-0491-1
- Barthet, M.M., Moukarzel, K., Smith, K.N., Patel, J., Hilu, K.W., 2015b. Data from: Alternative translation initiation codons for the plastid maturase MatK: unraveling the pseudogene misconception in the Orchidaceae. *Dryad Digit. Repos.* doi:http://dx.doi.org/10.5061/dryad.j4489
- Boch, J., 2011. TALEs of genome targeting. *Nat. Biotechnol.* 29, 135–136. doi: 10.1038/nbt.1767
- Bruening W., Pelletier J., 1996. A non-AUG translational initiation event generates novel WT1 isoforms. *J. Biol. Chem.* 15.

- Chiti, F., Dobson, C.M., 2006. Protein Misfolding, Functional Amyloid, and Human Disease. *Annu. Rev. Biochem.* 75, 333–366. doi:10.1146/annurev.biochem.75.101304.123901
- Cooper, S., Treuille, A., Barbero, J., Leaver-Fay, A., Tuite, K., Khatib, F., Snyder, A.C., Beenen, M., Salesin, D., Baker, D., Popov(i)c, Z., 2010. The challenge of designing scientific discovery games.
- Cowan, J.L., Perry, L.S., Edwards, R.J., Damerall, D., Roworth, A.P., Johnston, H.E., Coldwell, M.J., 2014. Identification of non-AUG initiated, N-terminally extended open reading frames in human genes by an experimentally-informed bioinformatics workflow. *Nucleic Acids Res.* Unpublished Draft.
- Cygnus-Xi, J., Vorksholk, M.S., 2014. CureCoin. Curecoin LLC.
- Ezkurdia, I., Juan, D., Rodriguez, J.M., Frankish, A., Diekhans, M., Harrow, J., Vazquez, J., Valencia, A., Tress, M.L., 2013. The shrinking human protein coding complement: are there now fewer than 20,000 genes? ARXIV eprint arXiv:1312.7111.
- Ingolia, N.T., Lareau, L.F., Weissman, J.S., 2011. Ribosome Profiling of Mouse Embryonic Stem Cells Reveals the Complexity and Dynamics of Mammalian Proteomes. *Cell* 147, 789–802. doi:10.1016/j.cell.2011.10.002
- Kozak, M., 1989. Context effects and inefficient initiation at non-AUG codons in eucaryotic cell-free translation systems. *Mol. Cell. Biol.* 9, 5073–5080. doi: 10.1128/MCB.9.11.5073
- Lander, E.S., Linton, L.M., Birren, B., Nusbaum, C., Zody, M.C., Baldwin, J., Devon, K., Dewar, K., Doyle, M., FitzHugh, W., Funke, R., Gage, D., Harris, K., Heaford, A., Howland, J., Kann, L., Lehoczy, J., LeVine, R., McEwan, P., McKernan, K., Meldrim, J., Mesirov, J.P., Miranda, C., Morris, W., Naylor, J., Raymond, C., Rosetti, M., Santos, R., Sheridan, A., Sougnez, C., Stange-Thomann, N., Stojanovic, N., Subramanian, A., Wyman, D., Rogers, J., Sulston, J., Ainscough, R., Beck, S., Bentley, D., Burton, J., Clee, C., Carter, N., Coulson, A., Deadman, R., Deloukas, P., Dunham, A., Dunham, I., Durbin, R., French, L., Grafham, D., Gregory, S., Hubbard, T., Humphray, S., Hunt, A., Jones, M., Lloyd, C., McMurray, A., Matthews, L., Mercer, S., Milne, S., Mullikin, J.C., Mungall, A., Plumb, R., Ross, M., Shownkeen, R., Sims, S., Waterston, R.H., Wilson, R.K., Hillier, L.W., McPherson, J.D., Marra, M.A., Mardis, E.R., Fulton, L.A., Chinwalla, A.T., Pepin, K.H., Gish, W.R., Chissole, S.L., Wendl, M.C., Delehaunty, K.D., Miner, T.L., Delehaunty, A., Kramer, J.B., Cook, L.L., Fulton, R.S., Johnson, D.L., Minx, P.J., Clifton, S.W., Haw-



- kins, T., Branscomb, E., Predki, P., Richardson, P., Wenning, S., Slezak, T., Doggett, N., Cheng, J.-F., Olsen, A., Lucas, S., Elkin, C., Uberbacher, E., Frazier, M., Gibbs, R.A., Muzny, D.M., Scherer, S.E., Bouck, J.B., Sodergren, E.J., Worley, K.C., Rives, C.M., Gorrell, J.H., Metzker, M.L., Naylor, S.L., Kucherlapati, R.S., Nelson, D.L., Weinstock, G.M., Sakaki, Y., Fujiyama, A., Hattori, M., Yada, T., Toyoda, A., Itoh, T., Kawagoe, C., Watanabe, H., Totoki, Y., Taylor, T., Weissenbach, J., Heilig, R., Saurin, W., Artiguenave, F., Brottier, P., Bruls, T., Pelletier, E., Robert, C., Wincker, P., Rosenthal, A., Platzer, M., Nyakatura, G., Taudien, S., Rump, A., Smith, D.R., Doucette-Stamm, L., Rubenfield, M., Weinstock, K., Lee, H.M., Dubois, J., Yang, H., Yu, J., Wang, J., Huang, G., Gu, J., Hood, L., Rowen, L., Madan, A., Qin, S., Davis, R.W., Federspiel, N.A., Abola, A.P., Proctor, M.J., Roe, B.A., Chen, F., Pan, H., Ramser, J., Lehrach, H., Reinhardt, R., McCombie, W.R., de la Bastide, M., Dedhia, N., Blöcker, H., Hornischer, K., Nordsiek, G., Agarwala, R., Aravind, L., Bailey, J.A., Bateman, A., Batzoglu, S., Birney, E., Bork, P., Brown, D.G., Burge, C.B., Cerutti, L., Chen, H.-C., Church, D., Clamp, M., Copley, R.R., Doerks, T., Eddy, S.R., Eichler, E.E., Furey, T.S., Galagan, J., Gilbert, J.G.R., Harmon, C., Hayashizaki, Y., Haussler, D., Hermjakob, H., Hokamp, K., Jang, W., Johnson, L.S., Jones, T.A., Kasif, S., Kasprzyk, A., Kennedy, S., Kent, W.J., Kitts, P., Koonin, E.V., Korf, I., Kulp, D., Lancet, D., Lowe, T.M., McLysaght, A., Mikkelsen, T., Moran, J.V., Mulder, N., Pollara, V.J., Ponting, C.P., Schuler, G., Schultz, J., Slater, G., Smit, A.F.A., Stupka, E., Szustakowski, J., Thierry-Mieg, D., Thierry-Mieg, J., Wagner, L., Wallis, J., Wheeler, R., Williams, A., Wolf, Y.I., Wolfe, K.H., Yang, S.-P., Yeh, R.-F., Collins, F., Guyer, M.S., Peterson, J., Felsenfeld, A., Wetterstrand, K.A., Myers, R.M., Schmutz, J., Dickson, M., Grimwood, J., Cox, D.R., Olson, M.V., Kaul, R., Raymond, C., Shimizu, N., Kawasaki, K., Minoshima, S., Evans, G.A., Athanasiou, M., Schultz, R., Patrinos, A., Morgan, M.J., 2001. Initial sequencing and analysis of the human genome. *Nature* 409, 860–921. doi:10.1038/35057062
- Larson, S.M., Snow, Christopher D., Shirts, M.R., Pande, V.S., 2002. *Folding@home and Genome@Home: Using distributed computing to tackle previously intractable problems in computational biology*. Horiz. Press.
  - Lee, S., Liu, B., Lee, S., Huang, S.-X., Shen, B., Qian, S.-B., 2012. Global mapping of translation initiation sites in mammalian cells at single-nucleotide resolution. *Proc. Natl. Acad. Sci.* 109, E2424–E2432. doi:10.1073/pnas.1207846109
  - Peabody DS, 1989. Translation Initiation at non-AUG triplets in Mammalian Cells. *J. Biol. Chem.* 5031–5035.

- Qi, L.S., Larson, M.H., Gilbert, L.A., Doudna, J.A., Weissman, J.S., Arkin, A.P., Lim, W.A., 2013. Repurposing CRISPR as an RNA-Guided Platform for Sequence-Specific Control of Gene Expression. *Cell* 152, 1173–1183. doi: 10.1016/j.cell.2013.02.022
- Ran, F.A., Hsu, P.D., Lin, C.-Y., Gootenberg, J.S., Konermann, S., Trevino, A.E., Scott, D.A., Inoue, A., Matoba, S., Zhang, Y., Zhang, F., 2013. Double Nicking by RNA-Guided CRISPR Cas9 for Enhanced Genome Editing Specificity. *Cell* 154, 1380–1389. doi:10.1016/j.cell.2013.08.021
- Smith, R.N., Aleksic, J., Butano, D., Carr, A., Contrino, S., Hu, F., Lyne, M., Lyne, R., Kalderimis, A., Rutherford, K., Stepan, R., Sullivan, J., Wakeling, M., Watkins, X., Micklem, G., 2012. InterMine: a flexible data warehouse system for the integration and analysis of heterogeneous biological data. *Bioinformatics* 28, 3163–3165. doi:10.1093/bioinformatics/bts577
- Venter, J.C., 2001. The Sequence of the Human Genome. *Science* 291, 1304–1351. doi:10.1126/science.1058040
- Wegrzyn, J.L., Drudge, T.M., Valafar, F., Hook, V., 2008. Bioinformatic analyses of mammalian 5'-UTR sequence properties of mRNAs predicts alternative translation sites. *BMC Bioinformatics* 9.

---

**Danfox Davies** is a biodesign entrepreneur, completed a *BSc in biochemistry* at the *University of Southampton*, and is founder and CEO of *Vulpine Designs*. He is from West Yorkshire, frequents hackerspaces and DIYbio labs, such as *Forma Biolabs*, and wants to establish a lab near Cork, Ireland, with staff accommodation for research and development, in time for *Indie Bio EU* next year. His current work involves designing a bioprinter, putting ideas for software into UML for his colleagues, and investment pitches. His dream is a future in which everyone can edit their own genes in every cell of their body at home as easily as they can use a 3D printer today, if not better.

# Building Artificial Connectomes

*Timothy Busbice*

I have emulated the *C elegans* connectome (i.e., how a nervous system is wired) in robotics and shown that the connectome with a simple neuron model **displays behaviors similar to the animal itself**. With further emulations of other animal connectomes (e.g., Zebra fish), I have discovered certain rules and paradigms that govern connectomic engineering and that animal behavior is highly integrated into the connectomic structure. This means that least at this time, it's virtually impossible to remove unwanted behaviors.

Using animal connectome emulation, unwanted behaviors such as defensive mechanisms, will make it difficult to create robots and applications that are conducive to our desires and needs. For example, one would not want to create a robotic puppy for their child that bites. We would probably more want the robotic puppy to simply shut down and stop playing rather than chew on our children. This has led me to create artificial connectomes with the hope that we can teach behavior into the system and not have to try and remove behavior from existing and established animal connectomes. This product is what I label *connectomic artificial intelligence*.

## Connectomic Rules

There are certain rules to connectomic engineering that allow the emulation of sensory to cortical to motor output behaviors. These are the rules that I have observed, and many have been noted by other connectomic scientists. Some observations may be obvious and simple, but these rules set connectomic AI apart from other traditional artificial intelligence paradigms such as deep learning and Hierarchical Temporary Memory (HTM), and I believe is the best opportunity for humans to recreate our own intelligence in a computer system or network.

Here are connectomic observations and rules in no particular order. Please note that these are purely connectomic properties and do not place emphasis on the intracellular activity of individual neurons:

1. The connectome is always on. Our nervous systems never shut down; they are always active. Having emulated animal connectomes, I once ran a test of an emulated connectome with some initial sensory input that started the cortical processing; and after 24 hours, I turned it off since it was obvious that the neuron-to-neuron activity would never end. This creates computing issues as well as network issues.
2. The connectome is highly recursive, and in many cases, it is exponentially recursive. If we start with a cortical neuron and map out the connections to other neurons, we see that there are a few of those postsynaptic neurons that connect back to the originating neuron. This set of neurons connects to another set of neurons, and from that next set, many connect back to the originating neuron. This goes out to more layers and sets; and at each level, there are a number of connections back to the originating neuron. From my data analysis, in many cases, we see that this recursive connectivity grows exponentially back to the originating neuron. I believe this is why the connectome is always on. Further work with high-level organisms has shown that the recursive nature of the connectome is more so in local regions of nervous systems and present, but not as connected, between regions of cortex.
3. There are two basic temporal aspects in any nervous system:
  - a. The polarization timing where if a neuron, real or simulated, is left alone or not stimulated over a certain timespan, it polarizes or goes to a negative (or zero) value as an internal function of time; i.e., the all or none response.
  - b. The network topology of the connectome is a function of time as well. If certain neurons fire before others, a set of behaviors will occur; but if a set of those same neurons fire in a different sequence, we often see other behaviors as an external, network timing.
4. There is a dampening effect displayed on muscle output. When we apply connectomes, animal or artificial, to robotics, we observe a lot of neural muscular activity where values are being sent to the motors (muscles) continuously and

rapidly, but we do not see motors react in a jerky or erratic manner. We observe the motor activity as smooth and continuous. I speculate that our cerebellum plays a role in this smoothing of motor output, along with the coordination of cortex-to-motor response.

5. There is a left-to-right component to all (or most) nervous systems. This is obvious but rarely employed, if ever, when creating artificial intelligence, and there is an impact on how these Left-Right edges are configured. From data analysis and further emulation of artificial connectomes, we find that connectivity between Right-Right or Left-Left neurons is about two-thirds more than Left-Right or Right-Left connectivity. These hemispheric connections are very important in perceptual direction as well as in sequencing motor output.

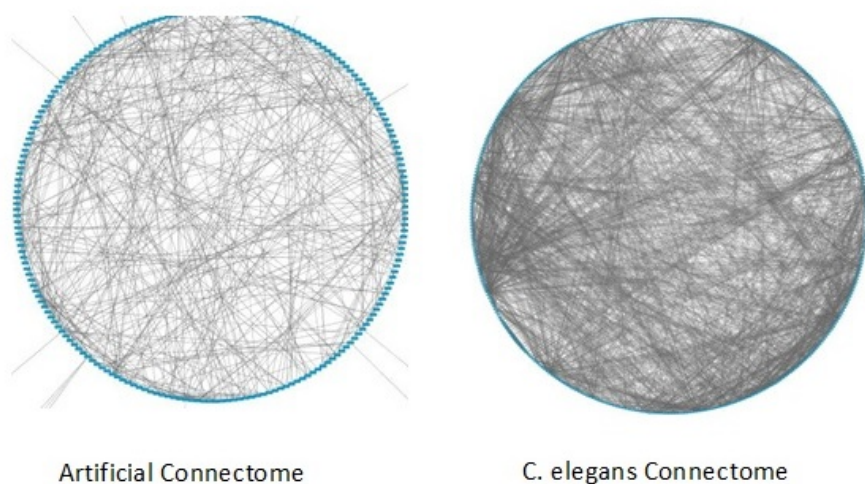
## Building Artificial Connectomes

Using these observations, I wanted to develop a connectome from scratch that could learn and adapt to whatever sensory input I wanted it to consume. I did some backward data analysis on motor-neuron-to-sensory-neuron pathways using animal connectomes and realized some patterns emerging. I could see the way in which neurons were connected was much like the concept purported by the Watts-Strogatz small world algorithm. I created a program that allowed me to construct a connectome using different parameters in the Watts-Strogatz algorithm and ran a simple classification test with each model created.

Running approximately 100 trials using different configurations, I found that there is a fine line between too much connectivity and too little connectivity. Too much would result in a model that would run continuously once stimulated but would output to all motor neurons/muscles simultaneously and constantly without any discrimination; for example, all muscles were active all the time. Too little connectivity would result in the connectome taking a very long time to activate any muscle activity, if any in some cases, and the system would cease to self-stimulate once any sensory stimulation was stopped. The mean degree matched against the rewiring probability was very narrow in nature, with a resulting connectome that would self-simulate once it had a short amount of sensory input, and the motor neuron to muscle output was discriminatory determined by the sensory input.

## Theoretical Artificial Nervous Systems

Each artificial connectome as shown in [Figure 3-1](#) can be connected to other connectomes; and I can create multiple regions, each highly recurrent, with a large degree of interconnectivity between the regions making recurrent interconnections. Therefore, in theory, we can expand our model of a simple artificial connectome to create an artificial brain as shown in [Figure 3-2](#).



*Figure 3-1. A simple artificial connectome compared against the C. elegans connectome. Squares around the circumference are nodes, and lines between are edges that connect the nodes.*

After experimenting with an artificial brain concept, I was reminded of efforts I developed about 10 years ago utilizing the six-layer concept of a cortical column of the human cortex. Using the concept of the interconnectivity of these cortical columns, I created three cortical regions, each containing several hundred cortical columns, and interconnected between the layers in a rudimentary fashion as defined by the functions of each layer. I then added sensory neurons that connected directly to the first region of cortical columns and left-right motor neurons that connected to sets of left-right muscles, as shown in [Figure 3-3](#). The six-layer approach worked very well with simple classification tests using a 10 × 10 grid of pixels representing numeric values.

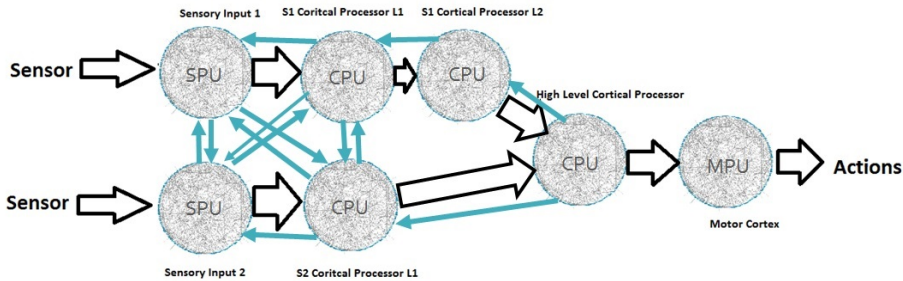


Figure 3-2. A theoretical artificial brain. Sensor data is encoded and sent to sensory processing units (SPU) which connect to cortical processing units (CPU) which connect to motor processing units (MPU) that stimulate muscles or appropriate actions. Each region of SPUs and CPUs are interconnected and send pathways forward and back. Each processing unit region is an artificial connectome in and of itself.

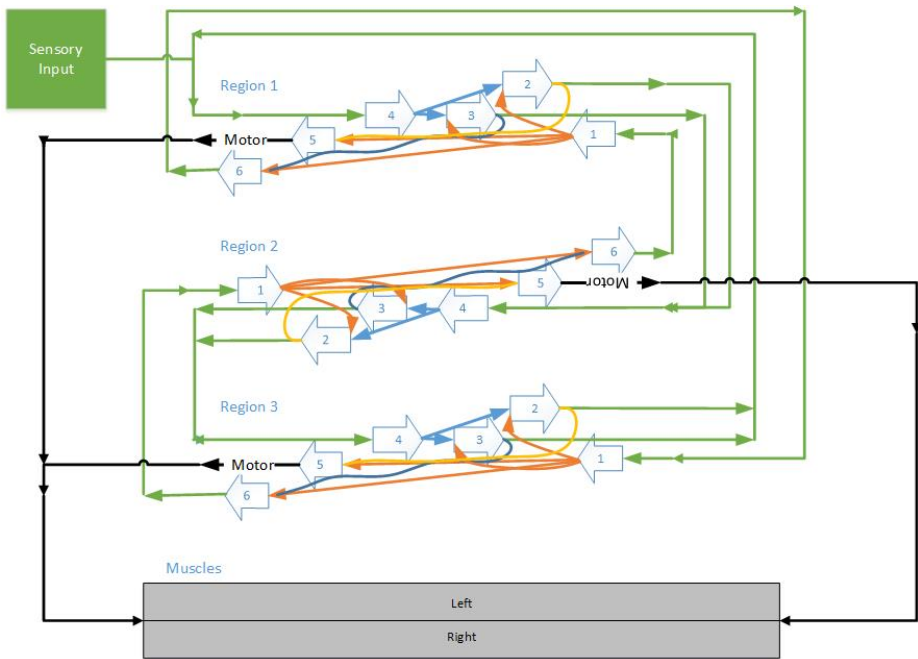


Figure 3-3. A simple schematic of an artificial connectome using a six-layer cortical column approach where each neuron or node in a particular layer interconnects to other layers; and layers 2 and 3 connect to layer 4 of other regions, layer 6 connects back to previous regions and layer 5 connects to motor output.

## Computational Methods

In each phase of the research, I use *adjacency matrices*, where each row represents a connectivity (edge) vector that can be used to increase weighted values as individual neurons are fired. In addition to an adjacency matrix to contain the connections and weights between those connections, there are five other arrays where each cell of the array represents a specific neuron. Those five arrays are:

- An array to contain the accumulation of weights
- An array to contain a timestamp of the last time a neuron was stimulated
- An array containing the threshold of a particular neuron
- An array containing a “learning” accumulator
- An array containing a “growth” accumulator (note that the learning and growth accumulators trigger plasticity in the system)

As with my early experiments in emulating animal connectomes, I use an internal temporal aspect of each neuron so that if a neuron has not been stimulated in a few microseconds, all accumulators (i.e., weight, learning, and growth) are set to zero and the threshold is set to a default value. For a neuron to fire (to reach its action potential), the weight accumulation must exceed the threshold. Any time a neuron fires, its weight accumulation is set to zero (I have also experimented with negative numbers, but the outcome is pretty much the same). The weight vector of the adjacency matrix is added to the accumulation array each time a neuron fires, and the weights are accumulated. A timer function constantly runs to check when accumulation is greater than the threshold and fires each of the neurons that meets this criteria.

To simulate synaptic fatigue, each time a neuron fires, the threshold for that neuron is incremented by a small amount so that a constant activation of a neuron gets a little harder to exceed the threshold. As mentioned, the threshold is reset to a default value if a neuron does not fire within a set time frame.

Plasticity is formed in two ways: learning and growth. A “learning” array is incremented when a neuron fires; and if that learning accumulation exceeds a learning threshold, weight is increased on the adjacency matrix so where a connection between neuron A to B may have a weight of one (1), it could be increased by one so now it would be two (2). This simulates connectivity strength. A “growth” array is incremented when the learning accumulation exceeds the learning threshold. When the growth accumulation exceeds a growth threshold, I add a



synaptic connection to a nearby neuron, which simulates synaptic growth. As mentioned, if a specified time has passed since a neuron was stimulated, both the learning and growth accumulators are set to zero; that is to say, all accumulation is a function of time.

## Conclusion

This technology can be used for unsupervised learning in domain-agnostic environments. The real goal is real-time, streaming information that allows the AI to organize, optimize, and make predictions, then test outcomes against these hypotheses for its own reinforcement. There is still much research to do; but having a framework, and having successfully completed simple tests, my goal is to apply this technology to develop reaction to more complex input both from an application perspective and a robotic perspective. On the application side, there are many mundane tasks that can be automated with intelligence and with a system that can learn while unsupervised. Good examples are functions that invoke business continuity and active auditing. On the robot side, I am working to create an open system with many sensors to be able to navigate through hostile environments without human intervention where the robot can operate on its own to carry out primary objectives while avoiding harm. This technology can be used for many mundane, robotic tasks, including weed removal, picking fruit and vegetables to maintaining your lawn. Imagine a robot lawn mower or trimmer that senses a sprinkler head and knows not to run over it, not because you have programmed it to not run over sprinkler heads, but because the robot can sense there is an unidentified object that it has no reason or purpose to run over.

For us to advance AI beyond the current status quo, we need to explore new ways to give machines intelligence beyond programming rules and beyond current artificial neural networks. I believe connectomic AI holds a key to making this happen.

---

*Timothy Busbice is an author, connectomic engineer, and entrepreneur who applies biological nervous systems into complex networks through simulation and real-world adaptation. He has combined his expertise in neurobiology and computer science to create full sensory to motor nervous system simulations and continues to grow those results into more complex, inorganic organisms. Busbice is the cofounder and acting CTO of Connectomix, a startup dedicated to the creation of artificial nervous systems for the use in nonspecific domain learning and execution.*



# BioBright

*Glen Martin*

Of all the challenges implicit in biotechnology, one in particular galls Charles Fracchia: the laboratory. It is so basic to the field that it seems immutable. And that's the problem, says Fracchia, a fabricator of molecular machines, a former MIT Media Lab associate, and the founder of the biomedical research firm, BioBright. To a large degree, the lab is immutable. For no apparent reason other than inertia, much of its equipment and many of its techniques have remained unchanged for decades.

"I mean, come on," Fracchia says. "People are still using paper notebooks. *Notebooks?* In 2016? Really?"

It's not just that outmoded tools and antiquated, entrenched protocols waste time and energy, observes Fracchia. They also yield irreproducible results due to human error and improperly accrued and analyzed data. At the least, this means a great deal of wasted effort and capital. And at worst? It's potentially catastrophic. In the ever-accelerating pressure to publish, flawed studies are promoted as established fact. Respected journals are forced to retract articles. The entire peer review process is questioned. Not only are reputations ruined: human health and public safety can be threatened.

An obvious response to the lab conundrum, of course, is automation. And despite such homey and persistent anomalies like notebooks and timers, labs indeed are becoming more automated. But that's not enough, Fracchia maintains.

"The deficiencies of the laboratory are well-known, even if people tend to cling to certain relics," says Fracchia. "So the thinking is that we can solve a lot of the problems by taking the human out of the loop. But that's not the right approach either. Biology is not computer science. It *benefits* from computer science, but it's a different field. You can't just turn a handle and crank out numbers. If that were the case, Big Pharma would have solved all our problems decades ago. My argument is that we don't need an automated lab. We need a *smart* lab, one in which the human being is optimized, not replaced."

Such a lab, says Fracchia, divides tasks among humans and machines depending on relative strengths and weaknesses.

“In most labs, we’re asking human brains to do things they’re really, really bad at,” he says. “We’re asking people to do repetitive, deterministic things, things that require impeccable memories. But our memories are inherently limited, and we’re lousy at repetition. Machines, on the other hand, are great at that kind of stuff, of creating and verifying longitudinal data. What we—human beings—are good at is context and pattern recognition. We’re good at taking verified data and figuring out what it means in regard to a given project or issue, and how to apply it to a specific goal. So what we’re talking about is not a robot lab, but a kind of Iron Man lab—a suit, if you will, that the researcher puts on that greatly enhances outcomes.”

So Fracchia and his partners—Ryan Harrison, Joel Dapello, Hao Zhang, Nate Johnson, Adam Marblestone, and Jen Weber—founded BioBright to design such a lab. Or rather, labs; because a truly smart lab, Fracchia avers, must be customized for specific goals. A lab that’s positively brilliant for one mission may be relatively stupid for another.

“The idea isn’t to create a cookie-cutter lab, a one-size-fits-all lab,” he says. “Biologists not only work at different disciplines—they work in different ways; each has a unique approach. So we’re not asking biologists to accommodate themselves to the lab. We’re focused on creating bespoke applications that accommodate and augment the biologist.”

That involves proprietary devices that essentially assume the scut work of the lab.

“By way of example, we have a little, completely encrypted wireless sensor that you can plug into any platform,” says Fracchia. “It can be mounted to any experiment. All the data flows to it automatically, and it can sample as often as you want. One month, two months or more later, you can go back and review the historical values.”

Getting back to notebooks: as Fracchia observes, they’re ubiquitous.

“Despite their profound shortcomings, everybody still uses them,” he says, “and that can lead to serious errors, even for the most meticulous researcher.”

There are electronic notebooks, Fracchia acknowledges, but they require human inputs; people still have to “write” in them.

“So we developed a platform that functions as a self-writing notebook,” Fracchia says. “It eliminates the very real drudgery of notebook updating, it eliminates human error, and it optimizes the one thing about notebooks that is absolutely essential: the creation of an established and reliable record. Our platform allows

researchers to maintain a precise and verifiable log of the procedures and values of an experiment.”

Perhaps the most pressing need for the BioBright approach is in health care, where vast quantities of data on millions of individuals must be collected, analyzed, and correlated, a situation where errors are inevitable, given the clunky and ill-matched interfaces between machines and people that still characterize most medical labs.

“One of our clients is the University of Texas at Austin, which is the first tier-one university to build a new full-stack medical school in 50 years,” says Fracchia. “They contacted us, and we’re working with them to create infrastructure specifically designed to help doctors with analytical work. We’re helping build the smart health care facilities of the future. It’s incredibly exciting because we’re starting from scratch, starting clean, not trying to remedy existing and deficient systems.”

Along with medicine, Fracchia also is deeply interested in the “food space.” He recently participated in a conference on the future of food production sponsored by the Culinary Institute of America and MIT Media Lab.

“Medicine and food production both have deep connections to biotechnology,” says Fracchia, “and when it comes to lab work, both fields employ similar or identical processes. So our systems apply wherever biology applies.”

If Fracchia is highly skeptical of established lab protocols, he’s equally dissatisfied with the way most tech startups are funded these days. While venture capital usually is seen as the only practical means for getting new ideas and products to the marketplace, Fracchia maintains VC is more often than not an anchor, an inhibition to technological progress. Accordingly, BioBright is supported by revenues from clients and a few angel investors who, in Fracchia’s words, “are involved because they believe we need to change the way biology is done.”

VC’s demands for relatively quick returns, says Fracchia, “can absolutely kill companies before they have time to deploy their products or services. That’s especially true for fields other than hardware, like biology. I don’t want to go to a venture capitalist and say, ‘Give me \$20 million to build this,’ knowing it will take me eight years to complete the project, and then the response is ‘What is my return in 4.5 years?’ If I say, ‘I need eight years,’ I won’t get funded. Ultimately, these rigid, profit-oriented timelines destroy innovation, and we can’t afford that. Biology touches on real life and death issues: cures for Zika and Ebola, cures for cancer, food production. We need faster rates of innovation, and we won’t get that by building silos.”

And speaking of silos: Fracchia is concerned that overspecialization and walled-off data will make it difficult to staff smart labs with accomplished researchers.

“We’re not going to save the world by producing ever larger hordes of PhDs with narrow disciplines,” he says. “Look at the current numbers: only 15 percent of biology post docs are able to establish an academic career. A lot of them end up as technicians at pharma companies for \$45,000 a year. They have fantastic brains, but because they haven’t been trained properly, and because of insufficient open-source technology, they hit dead ends.”

Biologists, Fracchia continues, understand the problems implicit in biotechnology to a far greater degree than computer scientists, “but they don’t really have the tools to solve them. Computer scientists, on the other hand, have mastered a lot of the tools biologists need, but they don’t understand the essential problems. Somehow, we have to meld computer science and electronics and biology into a single disciple, and we have to make technology open and accessible. Cross-trained researchers with full access to the resources they need will be able to take full advantage of the smart labs that are coming. They’re the ones who will make the dramatic breakthroughs.”

---

***Glen Martin** covered science and the environment for the San Francisco Chronicle for 17 years. His work has appeared in more than 50 magazines and online outlets, including Discover, Science Digest, Wired, The Utne Reader, The Huffington Post, National Wildlife, Audubon, Outside, Men’s Journal, and Reader’s Digest. His latest book, *Game Changer: Animal Rights and the Fate of Africa’s Wildlife*, was published in 2012 by the University of California Press.*

# Design.bio

## BIOLOGY AT THE DESIGN STUDIO

### *Orkan Telhan*

As scientists' understanding and ability to manipulate biological systems have improved, there is growing interest in the design community to work with synthetic biology and biological design. In the hands of speculative designers, bacteria is turning into gold, furniture is grown out of mushrooms, specs of meat become leather jackets or discussed in cooking shows. Many of these bold ideas are realized as creative design statements or symbolic gestures that make us think about the cultural and environmental implications of working with biology. Sometimes due to the limits in knowledge or resources—or by simply by intention—these designs do not become part of everyday life and remain as intellectual endeavors featured in books, exhibitions, or competitions. This disconnect gives such designs the ability to be radical, unconstrained from regulations, feasibility, cost, or market demands. In return, however, they gain the power to shape public opinion, create discussion and debate, and inspire many more designers to work with biology.

On the other hand, there is a growing community of designers who would like to see biology as part of their practices in architecture, fashion design, food design, and so on (Figure 5-1). Biology for them is not a new niche or a realm for speculation, but rather a radical rethinking of aesthetics and manufacturing to eliminating pollution, addressing labor abuse, and income equality. The products that emerge from this mindset are on their way to becoming probiotic cosmetics, self-healing concrete, cow free-milk, or spider silk that can be woven into clothing. Unlike speculative design, these ideas are measured against pragmatics, safety concerns, and regulations. They are vetted by business models, market trends, or investor preferences like every other product that matters in our lives. They must consider safety and they are subject to regulations.



Figure 5-1. *Design.bio: new ways to bring design and biology together*

Today, there is a disparity between the growing demand to use biology and access to the resources to practice it safely outside academic institutions. Biological materials and knowledge are becoming increasingly available at informal learning centers—makerspaces, science museums, or community labs—but even at those places designers have to invest significant time and energy to learn biology and gain the skills to be able to do the simplest experiments.

Designers are often expected to become capable amateur biologists before applying biology. Learning biology in this way is a daunting task, particularly when design possibilities are unavoidably vague at the outset; it requires some knowledge of biology to know what is possible, and some knowledge of what is possible to begin the process of designing. For science and engineering students, learning the basic principles of working with living organisms may suffice. But it is harder for design students to engage with the abstract world of organisms, which all look indifferent to the eye in small vials and test tubes, without having the necessary means to culture them. Learning how to make bacteria smell like bananas or emit fluorescent colors are attractive first steps, but maintaining the interest often proves to be difficult if designers cannot directly apply their knowledge to a domain they already know. Architects, for example, usually do not learn programming by studying computer science. They learn it as part of their architecture education. Coding is incorporated into parametric design tools where



modeling and fabrication processes are scripted. Designers see the immediate application of their knowledge as part of their design, whether it is modeling a car or generating patterns for fabrics. Designers also need to see biology in form and action—as materials or applications—that can be sketched, prototyped, tested in iterations.

It is important for designers to engage with biology in a lab and understand the sceptic and labor-intensive culture of working with life. Yet design almost always happens in studios where computers, software, and various prototyping tools—laser cutters, plotters, 3D printers—work next to one another. These tools help us quickly test ideas in various forms and get feedback. More importantly they connect us to the large world of resources—images, fonts, code, 3D models, circuit designs—that are created by others can quickly be incorporated into our designs. The diversity of these resources is the very DNA of design studios in which ways to work with biology are currently absent, but increasingly on demand.

Microbial Design Studio brings biology to the design studio. It is a design tool that minimizes the learning curve to work with biology to create sufficiently complex applications in food, material, fashion, graphic design, or architecture. The studio functions as a portable wetlab that automates the design and prototyping process of working with microorganisms such as yeast and E.coli. Designers can grow these organisms to synthesize new chemicals and use the chemicals or the organisms directly as part of sensing and remediation applications ([Figure 5-2](#)).

Designers load bacteria and DNA from one end and receive their incubated products—such as food ingredients, dyes, limestone, or biopolymers, or biosensors—on the other end. They can immediately combine the bioproducts or the organisms with their designs at the studio. MDS lets designers work with CRISPR and other advanced gene editing technologies to directly manipulate genomes. However, in addition to genetically transforming organisms, the tool can be used to combine different organisms to design microbial communities, which can then be tested in the development of foods and flavors or in microbiome work. Designers can use the platform to analyze the microbial signatures of built environments and explore ways to design biofilms that can be incorporated to architecture, furniture, and objects that will yield beneficiary results to the users.

MDS is a combinatorial and parametric design tool. It allows designers to make experiments in which they can insert different combinations of gene sequences to different microorganisms to try multiple designs at the same time. After the DNA is inserted to the bacteria, the designs can be manipulated further through adding nutrients, reagents, and media, or by controlling incubation conditions by varying temperature or oxygen intake. The automation process elimi-

nates the need for manual labor in moving liquid across different devices to transform, incubate, purify, and measure outcomes at different stages. Designs are monitored in a single workflow through sensors that measure growth and check whether the intended features are realized or not. The outcomes of every experiment are recorded and stored on a website where they can be visualized, historicized, and shared with others.



Figure 5-2. *Microbial Design Studio*

The platform also connects to other machines and an online repository in which different protocols—or design recipes—can be downloaded to the machine. This feature allows designers to replicate an existing design that is already verified by other designers or experts. A network of design studios can also be programmed to work together to produce more of the same bioproduct or its variations, such that machines can work on distributed experiments.

MDS is distributed by Design.bio, a web portal that focuses on simplifying workflows for designing with biology by providing the tutorials, consumables, and control protocols for Microbial Design Studio. The site works as a content hub to demonstrate what designers can achieve with MDS.

One common problem biological designers experience is a lack of design examples that are simple enough to try on their own or in community labs. There are now kits, curricula (e.g., BioBuilder), and incubation platforms (e.g., Amino) on the market that can help new audiences learn about the basic principles of genetics or synthetic biology. We want to complement these efforts by making biology accessible to the design market. Our goal is to simplify the processes of

working with biology such that product and fashion designers and architects can immediately start working from examples and quickly develop their own ideas. Instead of spending years learning advanced levels of biology and complex, manual techniques such as pipetting, designers can focus on applications by building on existing designs in much the same way that electronics designers accelerate their work by using circuitry examples or programmers quickly develop code by building on source code samples. Manipulating existing examples is a great way of customizing designs for new applications as well as expanding knowledge in creative fields.

Today, when new applications are introduced to the community—through videos, research papers, or online write-ups—it is often very difficult to get the information and know-how. The limits of knowledge and the lack of familiarity with the jargon make scientific papers difficult to parse for non-biologists. Even if the protocols are made accessible through video demonstrations and visualizations, it is often difficult to achieve the same outcome, as results obtained through one set of equipment many not necessarily yield the same results through another. Also when manual labor is involved, the process is often prone to mistakes due to lack of skills, handling measurements, liquid, or tasks.

Design.bio streamlines the workflow so that designers have access to a standardized set of consumables, media, organisms, and a hardware platform so they can just focus on realizing their designs. It works like sites that distribute cooking recipes (e.g., [molecularrecipes.com](http://molecularrecipes.com)) or electronic kits (e.g., [Makershed](http://Makershed)) by becoming a single source for ingredients, equipment, and information necessary to work with biology. Here, designers can order applications created by other designers, who provide example use cases and systematic ways to vary outcomes such that they can be customized for different uses.

While currently the activities are centered on the MDS, we expect that the site will grow beyond a single platform to feature tools and technologies geared toward other domains of biological design initiated by other vendors. We see MDS and Design.bio as first steps to facilitate designer-to-designer interaction and allow the fields of biological design and synthetic biology to gain more traction in the mainstream fields. That's because product design, fashion, gastronomy, or architecture can harbor many creative and critical applications of working with living organisms.

As an example, I would like to share with you how MDS can be used to design probiotic donuts. This application was designed by Taylor Caputo, a product designer who explores the role of augmented yeast products in food and flavor design.



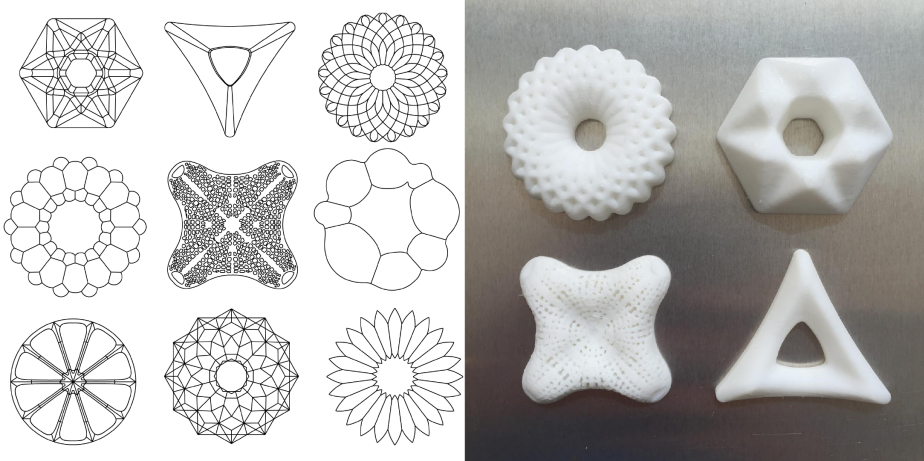


Figure 5-4. Selected shapes are 3D printed to cast silicon molds

The yeast inside the dough can make a big difference in taste. Commercial yeast cultures use standardized strains, which are generally selected for fast leavening and long shelf life. In the United States and Canada, genetically modified yeast is commonly used to improve performance in industrial production as well as in home bread-making machines. In addition to being optimized for speed, the cultures are manipulated to synthesize ascorbic acid (vitamin C), which can become a preserving agent for the dough when it is produced by the yeast.

In her donuts, Caputo works with golden bread yeast obtained from Bio-Builder.<sup>1</sup> The yeast (see Figures 5-5 and 5-6) carries a special genome that allows it to produce beta-carotene. This strain is specially designed to create golden bread that addresses vitamin A deficiency in countries where there are high infant mortality rates due to malnutrition. Vitamin A, contained within baked products, is a potentially a reliable source. It does not pose risks to the environment in the way a genetically modified plant could, since the organisms do not interact directly with nature. Instead they are grown in an isolated environment, such as an industrial fermenter similar to those used in beer and cheese production. Once beta-carotene is synthesized, it can be extracted from the organisms through purification methods, after which the organism is safely terminated.

1 More information about the science behind golden yeast, can be found at <http://biobuilder.org/golden-bread-students>.



*Figure 5-5. Golden yeast cultured on plate showing the variable outcomes of the unstable strain*



*Figure 5-6. Golden yeast cultured with microbial design studio*

While the golden bread yeast is engineered to produce the color orange, it is genetically unstable and can produce white or red colors as well. An unstable strain or a growth process interrupted before the completion of the metabolic pathway may not yield beta-carotene (orange) but an intermediary molecule such



as lycopene, which makes the culture appear red. When grown with MDS, the yeast can be incubated at different temperatures and with different reagents that can speed up or slow down the growth. This process can be driven by algorithms that periodically measure the amount of yeast in the system and terminate the growth based on the desired output.

Golden yeast, on the other hand, is not as efficient as a commercial yeast culture. For a more airy taste, it is better to mix it with more standardized, off-the-shelf baker's yeast. However, the mixing process can be systematized using MDS by combinatorially mixing different strains at different ratios under different growth conditions. Multiple strains can be simultaneously grown together to do a comparative study to find out which donut will taste better. As the leavened dough is ultimately baked (Figures 5-7 and 5-8), the final donuts are free of any live culture. Unless one saves a batch, there is often no way to save the know-how. Instead of saving a starter culture for the next experiment, MDS can keep track of the analytics. The amount of cultures that are mixed, the type of strains, temperature, and reagents can all be recorded automatically as variables of a recipe that can be used later. These recipes then can be shared with other designers, who can replicate the same taste without shipping any materials.



Figure 5-7. Donuts are baked in custom shapes using silicon molds and dough augmented with golden yeast



Figure 5-8. A series of parametrically designed donuts

The icing of the donut (Figure 5-9) is the last step of the design process. In addition to choosing from many different kinds of toppings, one could use flavors grown out of various microorganisms. In her application, Caputo demonstrates the use of Farnesol, a sweet smelling compound that is often present in sandalwood oil. She uses a special strain of odor-free *E. coli* K12 (strain YYC912) sourced from [Yale E. coli Genetic Stock Center](#). The odor-free *E. coli* is transformed with pMBIS plasmid obtained from Addgene plasmid repository ([Plasmid #17817](#)). Like yeast, *E. coli* K12 is also a non-pathogenic strain, which can be safely handled by the user. The Farnesol producing bacteria can be grown in the MDS and mixed with a sweetened icing to top the golden donut with a sandalwood-like smell.

Caputo's microbial donuts at this stage are proof of concepts. They have not been evaluated for their taste. Due to regulatory requirements, the donuts cannot be consumed within the University of Pennsylvania, where the experiments are designed. However, the design application will be available in [Design.bio](#) with a variety of strains of food-grade yeast and bacteria and combinatorial and parametric design options. In the meantime, the full tutorial of microbial donuts example, ingredients, and design instructions can be found at [design.bio/microbialdonuts](#).





Figure 5-9. Donuts are decorated with icing mixed with *E. coli*

Designers who are interested in participating in Microbial Design Studio's alpha testing program can also reach us at [info@biorealize.com](mailto:info@biorealize.com).

---

**Orkan Telhan, PhD**, is an interdisciplinary designer whose investigations focus on the design of interrogative objects, interfaces, and media engaging with critical issues in social, cultural, and environmental responsibility. Telhan is Assistant Professor of Fine Arts - Emerging Design Practices at University of Pennsylvania, School of Design. He holds a PhD in Design and Computation from Massachusetts Institute of Technology with a focus on Synthetic Biology and Biological Design. He was part of the Sociable Media Group at the MIT Media Laboratory and the Mobile Experience Lab at the MIT Design Laboratory.



# Where Science-as-a-Service and Supercomputing Meet

**A DEEPER LOOK: RUSSELL POLDRACK AND THE CENTER FOR REPRODUCIBLE NEUROSCIENCE**

*Sindhu Ravuri*

With the meteoric advancement of technology, there is increasing scrutiny about how science is getting done. The Internet has enabled scientific results to be publicized, disseminated, modified, and expanded within minutes. Social media can subsequently propagate that information universally, allowing virtually anyone with access to WiFi to influence—and, therefore, potentially skew—data collection procedures and results. Ultimately, this means that reproducibility of modern science is under attack.

Russell Poldrack and his colleagues at the Stanford Center for Reproducible Neuroscience (for brevity's sake, we will refer to this as SCRn) are tackling that problem head on.

“We think that the kind of platform that we’re building has the chance to actually improve the quality of the conclusions that we can draw from science overall,” said Poldrack, codirector of SCRn and professor of psychology at Stanford.

SCRn, founded last year, empowers researchers to easily access tools to analyze or share their data and analysis workflows, while paying specific attention to the reproducibility of research. In particular, SCRn focuses on neuroimaging research which, due to its complexity, offers individuals flexibility when analyzing it.

“One of the big issues centers around harnessing that flexibility so that people don’t just punt for one analysis out of 200 and then publish it,” said Poldrack. “We want to provide researchers the ability to see how robust their results are across lots of different analysis streams, and also record the provenance of their

analysis. That way, when they find something, they know the  $n$  different analyses they did in order to get their results. You have some way to quantify the exploration that went into data analyses.”

But we already have platforms like USC’s LONI Pipeline workflow application. So what makes SCRn unique from other data sharing projects? The answer lies in its universal availability.

“[SCRn] is open such that anybody can come on [to] analyze data and share it,” said Poldrack. “That’s the big difference—ours is going to offer curated workflows. If somebody just shows up and says, ‘I have a data set; I want to run it through the best of breeds standard workflow,’ we’ll have that ready for them.”

Similarly, if a researcher wants to run an analysis through 100 different workflows and look across those for a consensus, Poldrack said therein lies the goal of the SCRn.

“So, rather than just saying, ‘Here’s the platform; come do what you will,’ we want also offer people the ability to say basically, ‘I have a data set, and I want to do the gold standard analysis on it,’” said Poldrack.

Users with absolutely no programming ability are one hundred percent welcome to upload data values to SCRn. The researcher would upload raw data, which is comprised of both images (from MRIs, for example) and other measurements (such as those of the participant’s behavior during MRI scanning). She would also include the necessary metadata to describe their image data, such as the details of MRI acquisition.

“Our beta site already has the ability to upload a whole data set, validate the data set, and then perform some processing on it,” said Poldrack. “Right now, the main processing...is quality assurance on the data, but ultimately the idea is that the researcher should upload the data set, have it run through the entire standard processing pipeline, be able to download those results, and ultimately share those results. Therefore, the goal is to make it as easy as possible for people without a lot of technical experience to be able to use state of the art analyses.”

However, the main target audience of SCRn includes researchers in neuroscience, neurology, radiology, and medical-related fields that are working with neuroimaging data.

“What we’d like to ultimately do is generate some competitions for development of machine learning tools, for decoding particular features from images,” Poldrack said.

SCRn, though currently in beta mode, extends two other data sharing platforms, OpenfMRI and NeuroVault, conceived seven years ago by Poldrack.

Since its inception, OpenfMRI, primarily catering to audiences in academia rather than industry, has targeted the secure sharing of complete raw data sets,

providing opportunity for more accurate medical diagnosis and other types of analysis. In fact, if you compute the amount of money on data collection that's been saved up by reuses of OpenfMRI, it is on the order of \$800,000 worth of savings.

"There's already been a number of papers where people took data from OpenfMRI and did novel analyses on it to come up with new answers to a particular question," said Poldrack.

Nevertheless, accompanying each MRI scan is both data off the scanner as well as metadata. This relatively large data set (hundreds of megabytes per participant in a study), requires a substantial amount of organization.

"The amount of work that goes into curating a data set like that and preparing it has kind of limited the degree to which people are willing to go to that effort," said Poldrack.

Enter NeuroVault, a project that aims at a different level of data sharing by providing researchers a method to spread more derivative data. Consequently, NeuroVault has seen a substantially larger audience than OpenfMRI.

In other words, when analysis is done on neuroimaging data, what generally results is a statistical map across the whole brain. In essence, this illustrates every little spot on the brain, as well as the evidence for a response to whatever the study is studying. For example, if you are studying how people make decisions, and you want to compare easy versus hard decisions. The statistical map would tell you the statistical evidence for each particular response in the brain for both decisions.

"Those are much easier to share because instead of a whole huge data set, it's a single image. The amount of metadata you would need to interpret that image is much less," said Poldrack.

Additionally, NeuroVault very efficiently shares a statistical map, in addition to rendering a viewable browser and a permalink for that image. That way, researchers can actually put it into their publication for readers to see the full data themselves.

SCRN's research interface is based on a data format called the Brain Imaging Data Structure (BIDS). If a researcher has a data set formatted according to that standard, the SCRN website can automatically upload it. From there, they can choose from the particular workflows that they want to run the data through. Each time they do that, they select a specific snapshot of the data, so we know exactly which data went into exactly which analysis.

As with any new project, one of the most difficult parts is actually getting it out there and garnering users. In addition to spreading awareness about SCRN to friends and testing it out with them, Poldrack and his team want to make sure

that groups who may not have access to supercomputing take full advantage of the SCRN.

“We want to show people that the tools they can get by using this platform and the answers it will allow them to obtain about their data are so much better than what they could do on their own,” said Poldrack. “Especially to groups that don’t have access to supercomputing, we are basically going to be providing access to such computing. Instead of a group having to develop its whole infrastructure, it can basically just take advantage of our platform and say, ‘We’re going to put our data on it, analyze it, and we’ll be ready to share it once we publish the data.’”

Planning how to publicize and configure SCRN into the current scientific landscape is not the only hurdle the team faced, though. A lot of the struggle emerged in deciding what technology platforms would best work with the scheme of reproducibility.

“In general, we throw around lots of ideas and see what sticks,” said Poldrack. “Energy can be pretty intense, because we all come from scientific backgrounds where our standard mode is to question everything; everyone is willing to challenge any idea, regardless of who it comes from.”

One of the ideas largely debated and discussed was which platform was optimal for the computing and data storage. The platform had to “scale across heterogeneous computing environments (including both academic computing and commercial cloud systems,” according to Poldrack. The team chose the Agave ‘science-as-a-service’ API, which has become the framework to the project’s infrastructure.

But that still begs the question: how does the SCRN team plan to ensure that data sharing via their project will become an automated, innate part of the research industry? Will the project actually compel scientists to upload and share their data as frequently as possible? Well, first things first—you need incentives. You need people to think that doing so is worth their time and effort.

“In part, that’s a structural issue about whether people will give out money for research or to people who publish their research, and whether those people want to incentivize researchers to share data,” said Poldrack.

The secondary facet of tackling this issue is the need to actually make it easy for people. It is certainly still relatively challenging.

“I think that our new BIDS will actually help with that,” Poldrack said. “I’m actually working on tools to be able to take in any data set—even if it’s not structured in BIDS format—and sort of reformat it into BIDS with a little bit of hand-holding.”

Basically, it comes down to incentive and tools. People need to be incentivized to share—improve their career, and it’s worth their while—and they need tools to make it as easy as possible.

“The tools aspect is obviously much easier than the incentives, which requires much bigger changes in the scientific landscape,” said Poldrack.

Although SCRN hinges on neuroimaging data, there is possibility for it to expand into other areas of cutting-edge scientific research, as long as it is dealing with large and complex data where there are relatively standard workflows and ways of describing the data.

“We don’t have any plans to expand it into other areas yet,” said Poldrack. “We first want to show that it works well in the neuroimaging domain. Ultimately, I think we’ll hope to expand it at least to other types of neuroimaging, for example, electroencephalogram imaging; but right now the goal is to make it work really well in our domain.”

In order to expand the tools that are available within the project, there will be a four-day coding sprint at Stanford August 1–4, 2016, where programmers from NeuroScout, Agave API, SPM, and FSL, to name a few, will come and work with SCRN to implement their tools within their workflows and the particular infrastructure that they use.

“One of the most exciting things is to see how quickly the technology has been able to come together to get data in and process them using this science as a service interface to the supercomputing systems,” Poldrack said. “We’ve actually been much more successful than I would have thought in pushing that out.”

SCRN is very close to its finishing line.

“One thing to say is that we all realize that there is increasing scrutiny about how science is getting done,” Poldrack said. “We realize there are problems and we want to fix them.”

We will see it all unfold in approximately five months time, when the full pipeline will be released. For more information, check out the [SCRN website](#).

---

*Sindhu Ravuri is a rising undergraduate sophomore majoring in molecular and cellular biology at the University of California, Berkeley. She covered science in society for The Triple Helix international organization, and won first place at the annual conference at Cambridge University for her research on schizophrenia. Her articles have earned her a national YoungArts Award, a literary scholarship at UC Berkeley, and internships at leading newspapers like the San Jose Mercury News.*





# Blockchain-Enabled Open Science Framework

*Vikram Dhillon*

Open dissemination of scientific research and data is a prerequisite for solving the reproducibility crisis. The problem of low reproducibility is plaguing all disciplines, but its impact is much worse in preclinical research and development. Further clinical R&D into drug candidates and targets is turning up wasteful because published results can't be replicated (Freedman 2015, Begley 2012). There are similar problems with clinical research: Drug trials are published with mostly positive data supporting the claims, and negligible negative data (Ioannidis 2015). In this article, the discussion will mostly remain limited to preclinical research outcomes.

The lack of replicability can be attributed to two broad reasons: a lack of consensus on protocols being used in research labs and the level of access to tools and equipment required for performing experiments. There are numerous methods that can be used to solve a problem, and domain-specific researchers often have their own versions of protocols which work optimally in their lab. This creates obvious problems for independent verification of results and claims. Moreover, not every lab has access to high-end equipment. Many researchers have to substitute lower-end techniques which may affect the overall accuracy. To solve these problems and standardize research equipment, we need the scientific community to reach a consensus on specific protocols and also provide generalized methods with sufficient flexibility for minor substitutions.

To get higher returns, investors specializing in the life-science research domain are interested in investing at the preclinical stages. But they often require some level of screening before making an investment. Traditionally, startups looking for investment have to partner with established institutions and go through thorough due diligence. Once those partnerships are established, the investors are more willing to consider making an investment. Currently, there is no independ-

ent vetting mechanism that would let biopharma companies demonstrate their ability to translate early research into promising drug candidates.

In this article, I propose making the process of commercializing preclinical research more reproducible and transparent by basing it on a blockchain. This effort will rely on the blockchain for communication to carry out peer reviews and publicly report the results. The program will be discussed thoroughly in a later section. Let us begin by reviewing three major initiatives currently in place to enhance reproducibility.

## Reproducibility Project: Cancer Biology (RP:CB)

**RP:CB** is a collaboration between the Center for Open Science and Science Exchange (under the CEO Elizabeth Iorns) which reexamines the most impactful publications in cancer biology from 2010-2012 and replicate them (Errington 2014). The replication is being conducted by expert members from the Science Exchange network, in accordance with standards from the Center of Open Science. The results of these experiments and RP:CB will be made publicly available through the open-access publisher eLife. The high impact studies under consideration will go through two stages of review, and the results from each stage will be published in a new, innovative format. The first phase will culminate in the creation of a replication plan for a study, in the form of a registered report. The protocols proposed in this report will have been reviewed by the scientific community prior to starting the experiments. The second phase is the replication study itself which will be conducted using the protocols specified in the Registered Report and completed by the Science Exchange network of scientific labs. In the end, both reports will be peer-reviewed by the eLife reviewers and made available online, along with all the methods and data (Mourouzis, 2015).

## Minimum Publishing Standards

Print-based publications used to limit the amount of space allocated to each section of a paper. The investigators would prioritize reporting results and making their claims, instead of detailing the protocols, especially in the methods section, where attention to detail is essential. This section generally provides the **instructions** that other researchers need in order to replicate an experiment. Most journals have now moved online, and space is not an issue. However, even with supplemental materials, the quality of methods and procedures reported has not increased significantly.

In some cases, the investigators choose to not disclose the details of their protocols and only vaguely cover the techniques used in their experiments, keeping

the exact procedure a trade secret. While there might be some value in doing so, it certainly doesn't help increase reproducibility. [BioMed Central](#) has sought to standardize the publication of procedures by providing a checklist of minimum standards that must be met before a paper can be published. The checklist outlines specific criteria that can be used as a guideline to write the entire methods section. Therefore, if all the requirements are followed, it would embed a high degree of replication in the methods.

The idea of publish or perish is still largely a reality of academia, and recognizing this, BioMed Central has launched new types of articles that will enhance reproducibility. This is the idea of launching support articles that are full publications to ensure authors receive credit for all their work—not just the publication but also protocols as registered reports, case studies, and data notes. The online platform for the journal also links related articles and resources such as the data notes or protocols which have been published in addition to the main research paper.

## Data Discovery Index

The National Institutes of Health (NIH) recognizes the crucial importance of reproducibility in life sciences, and the realistic complexities that arise during experiments which contribute to the lack of reproducibility. A complex array of other factors are involved, but we will focus on two elements that the NIH identified: lack of incentives for publishing negative data or other relevant experimental data and poor training in experimental design that would lend itself to high replicability. Let's discuss potential solutions to both issues.

Researchers only report a portion of data generated from their experiments, because their goal is to support the claims made in a publication. This often leads to scientists not publishing any negative data or other relevant data points that support the experiments. Recently, cloud-based data sharing services like figshare have become more popular for researchers to upload additional data corresponding to their publications, but the adoption rate remains low. A larger data set would allow for more consistent conclusions and statistical analysis. In addition, negative data can be very useful in determining potential avenues or pitfalls to avoid.

Data often goes missing in action because once a publication is live, the investigators have no motivation to dedicate additional time and resources to uploading supplementary data. NIH created a solution to this problem by creating the [Data Discovery Index](#), which allows researchers to upload unpublished databases (Collins 2014). If other researchers use this data in their own work, the original data

set can be cited. This becomes a new measure of scientific contribution that will allow researchers to gain additional citations for making experimental data available.

The second problem pointed out by the NIH is insufficient training in experimental design that translates easily to replicable studies. To remedy this, the NIH is releasing training modules to help researchers understand common pitfalls to avoid while designing experiments. In this paper, I propose a new vetting and disclosure mechanism for research labs interested in commercializing their research and attracting private investment by demonstrating rigorous preclinical reproducibility. My hope is that such an effort will promote reasonable disclosure of published research by private entities (such as biotech startups) and prevent colossal failures like Theranos from happening. The program will be based on an open and publicly available ledger called the blockchain, and will enable blockchain-based research tracking.

## **My Proposition: Use the Blockchain**

How can the blockchain enable research data tracking and communication? The answer is to use the blockchain as a backend “file-drawer” to store links. These links are simply references that in turn point to real data sources, and there are several advantages to storing only links on the blockchain. The first advantage is that the data uploaded to the blockchain is tamper-evident, meaning that any attempts made to modify it will be recorded. The blockchain relies on cryptographic signatures to maintain the integrity of the network, and any information pegged to it can be verified using the signatures. The second advantage is that any data stored in this manner becomes very easy to share publicly. To understand why, let’s follow an analogy. Imagine that we have a room full of packages, where each package has a sticker-label. Then, we can create a catalog of all the stickers. Once we have that, it is very easy to show someone the contents of a particular package. In the same sense, once we have links stored in the blockchain, we can easily and publicly share the metadata on the information stored by that link. Now if that catalog was electronic, we could search through it digitally. Similarly, we would have to develop frontend applications to search through the blockchain for links.

So what’s the practical advantage of a blockchain? The blockchain is a secure, tamper-evident, and verifiable mechanism for metadata storage. Public or private labs, early-stage startups, and private entities looking to commercialize their research can use this mechanism to report deep due diligence around the basic sciences. We can rely on the blockchain as a data structure to maintain the integ-

rity of the reported information. Novel modalities or new scaffold identification and optimization often take longer than planned. Biotech startups interested in next steps on the road to commercialization can use the blockchain to provide reliable and consistent history of updates. These updates and the platform can become a sophisticated system for providing public disclosures to potential investors or funding bodies. Moreover, startups are often required to already be vetted through partnerships prior to raising venture funding, and the blockchain can be used as a mechanism to partially automate the due diligence and review process. The reported methods, supplementary data, figures and experimental design can be uploaded to the blockchain and shared with scientists from the partnering firm to be analyzed for evidence of reproducibility and being on par with their standards of research. Blockchain-based histories and disclosure reviews can become the gold standard of vetting research labs or private entities before raising venture funding.

A comprehensive, blockchain-based solution is beyond the scope of this paper. However, a major source of innovation would be the applications built on top of the blockchain that can rely on a provable mechanism of data storage. One such example is a decentralized reputation system that can be built on top of the blockchain for rewarding groups with high standards of publications. The published documents can also become a primary source for journalists, who otherwise rely on press releases. I speculate that once more labs and biotech startups join such an effort, different tiers of review and reporting will emerge that can be advantageous to journalists and the general public. This mechanism can assist well-funded startups or companies to publicly disclose their progress, and stealth companies can appoint a reviewer to supplement an executive summary of the review process. These guidelines based on the blockchain can reduce the likelihood of another disaster like Theranos from occurring by requiring reasonable peer review and scrutiny.

## Closing Remarks

The validation of basic sciences in an independent, public, and steadily improving manner demonstrates a strong commitment to improving the quality of preclinical research being carried out by a research lab. It may become necessary for labs to obtain private funding to accelerate the development of drug candidates after obtaining promising animal-model data. Releasing updates publicly not only improves the research being done, but also boosts the confidence of private investors toward funding independently verified research. This commitment to repro-

ducibility can become a cornerstone for public-private partnerships between pharmaceutical companies and research labs in academia.

Carlota Perez describes the nature of technological revolutions (in her wonderful [book](#)) as being characterized by two distinct phases: the installation phase, when a technology is introduced to the market and the required infrastructure is built; and the deployment phase, when the technology has been widely adopted and next-generation applications are built on top. An entire file-system built on top of the blockchain containing metadata descriptors that are easy to share is an example of a scientific disclosure application deployed using the blockchain as a file-linking service. The blockchain is going through the installation phase presently, and this article presents a use case for the future deployment era.

## References

- Begley, C.G. and Ellis, L.M., 2012. Drug development: Raise standards for pre-clinical cancer research. *Nature*, 483(7391), pp.531-533.
- Collins, F.S. and Tabak, L.A., 2014. NIH plans to enhance reproducibility. *Nature*, 505(7485), p.612.
- Errington, T.M., Iorns, E., Gunn, W., Tan, F.E., Lomax, J. and Nosek, B.A., 2014. An open investigation of the reproducibility of cancer biology research. *Elife*, 3, p.e04333.
- Freedman, L.P., Cockburn, I.M. and Simcoe, T.S., 2015. The economics of reproducibility in preclinical research. *PLoS Biol*, 13(6), p.e1002165.
- Ioannidis, J.P., 2005. Why most published research findings are false. *PLoS Med*, 2(8), p.e124.
- Mourouzis, I. and Pantos, C., 2015. The Biomedical Data Journal in the New Era of Open Access to Scientific Information. *Biomed Data J*, 1(1).

---

**Vikram Dhillon** ([@opsbug](#)) is a research fellow at the University of Central Florida, at the Institute of Simulation and Training, where he is studying implementation of emerging technologies into digital health, with a focus on Ethereum and smart contracts.

# Predicting Cures

## ATOMWISE HARNESSES AI AND MACHINE LEARNING TO HASTEN DRUG DISCOVERIES

*Glen Martin*

It's hardly news that drug development is a horrendously expensive undertaking, one requiring painstaking physical synthesis and testing. That's because the process demands a massive commitment of equipment, time, and human labor; there's no other way, it seems, to assure newly identified molecules will properly address specific disorders without causing unacceptable side effects. Though it exists on the nano scale, the biochemical realm is vast in its granularity, and exploring and confirming its configurations is a Herculean task.

Correct that: there *was* no other way. In recent years, dramatic advancements in AI and machine learning have pointed to an alternative path, one that leads out of the wet lab to a virtual arena where the potential efficacy of candidate drugs can be predicted with a high degree of accuracy, greatly reducing the time necessary for the development, testing, and dissemination of new medicines.

"On average, it takes 14 years and about \$2 billion to discover and refine a new drug, take it through clinical trials, and get it into pharmacies," says Abraham Heifets, the CEO of San Francisco-based Atomwise. "And for every hundred projects that are initiated, only a few ultimately are approved, but the costs must be covered for every one of the failures. We have to do better. We have to reduce the number of flameouts."

Atomwise uses proprietary deep-learning algorithms and supercomputers to do just that: search for potentially beneficial compounds at the rate of millions of analyses per day.

"We aren't claiming that what we do is a substitute for clinical testing," says Heifets. "A good analogy is aircraft design. Computer simulations will tell you relatively quickly if your idea is practical in terms of basic benchmarks, but you're not going to just build a plane from the simulations and start flying people from

coast to coast. You're still going to test any wing you design in a wind tunnel. We see our job as winnowing out the likely failures and identifying the best candidates for testing."

Heifets notes that two main issues must be addressed in the development of any drug: efficacy and toxicity. But the relationship between the two isn't clear-cut. It's a palette of many shades of gray rather than a chiaroscuro of black and white. Many chemotherapy drugs, for example, are highly toxic, but they may be effective in treating specific cancers. The challenge lies in creating compounds that produce desired results with minimal side effects.

Heifets compares proteins in the human body to machines on an assembly line. Each such "machine" will transform a given input and pass it on to another proteinaceous machine. Diseases can occur, Heifets observes, when a protein fails to perform its assigned task, or messages are somehow garbled.

"Say cell division function is expressed but not turned off," Heifets says, "so you end up with a tumor. We see our job as identifying the machines that are running poorly, and monkey wrenching them—blocking their function. Ideally, you want a monkey wrench that fits the configuration of any given rogue protein really well so it chomps on it for a long time. But you want that wrench to work only for the target protein. You want all other proteins to remain unaffected. When they are affected, it's expressed as 'toxicity' or 'side effects.' So the goal is to produce molecules that stick really well to the bad protein, but leave all other proteins alone."

Coming up with candidate monkey wrenches doesn't necessarily mean starting from scratch. Atomwise researchers spend a lot of their time trying to repurpose or reposition existing drugs, evaluating them for applications other than their original purpose.

"All medicines have side effects, but not all side effects are necessarily bad," says Heifets. "Some are beneficial. Aspirin was originally marketed as a headache remedy, but later it was discovered it also had applications for heart disease. Viagra was first promoted as a treatment for hypertension; then it was determined it could counter sexual dysfunction. We've found that exploring alternative applications for existing medicines can be both productive and cost-effective."

Atomwise's most dramatic work to date involves Ebola.

"We got involved with an Ebola project in 2014, right when the pandemic was raging," says Heifets. "Everybody was in emergency mode, and we coordinated with a structural virologist to identify the mechanism that Ebola uses to invade healthy cells. Once we determined that, we started thinking of ways to block entry. There was no time to launch a full discovery process—we couldn't go through the usual 14-year cycle on this. So we started looking at promising molecules that



already had human safety data that had been through Phase II or later human trials.”

Atomwise’s researchers then ran analyses and came up with a short list of molecules that looked like good Ebola-inhibiting candidates.

“We tested just the top rankings of that list,” Heifets says, “and we found two molecules that blocked Ebola entry very effectively.”

The Atomwise team then conducted a pseudovirus-based inhibition assay to determine the molecules’ efficacy in stymieing Ebola.

“That turned out very promising,” Heifets says, “as did subsequent tests of cell metabolism to make sure the cells were staying healthy. So now we’re looking for a BSL-4 facility to run tests using the live virus and mice. Bottom line: We’ve cut that 14-year research cycle down massively. That’s incredibly heartening, and not just for Ebola. This project translates to all other infectious diseases. Just look at the list of pandemics since 2000: Ebola, SARS, Monkeypox, H1N1 influenza. These pathogens are constantly evolving, and pandemics will inevitably be part of our future. So we need better tools that will allow us to dramatically shorten the drug discovery and approval timeline.”

Heifets says algorithm-and-supercomputer-driven research will also play a crucial role in “addressing diseases that we thought we had cured but are coming back—like tuberculosis. WHO recently announced that TB has surpassed HIV as the world’s number one infectious disease. We’re in this Red Queen’s race with TB, and we need to do everything we can to improve our advantage.”

Orphan diseases will also benefit from AI and machine learning, Heifets maintains.

“There are many rare diseases that have never had effective medicines because it has been hard to justify big R&D budgets from a business standpoint,” Heifets says. “But our methods could provide medicines both quickly and relatively cheaply. And it’s here that DIY bio and advocacy groups play a particularly active role, both by bringing attention to the issues and helping with the research. We’ve partnered with some of these organizations, and they’re highly motivated. They’re doing good work, and we want to do everything we can to support them.”

Heifets takes pains to emphasize that Atomwise’s basic mission is not new. Rather, he says, the company’s successes are the fruition of decades of work by hundreds of researchers, most of whom remain unsung. The potential of computer technology in determining disease pathways and predicting effective remedies has been recognized for 40 years, says Heifets, but the machines and data that were available for most of that period weren’t up to the task.

“Now, finally, the technology is catching up with our ambitions,” he says. “But we have to remember we didn’t create this by ourselves. We’re standing on the shoulders of giants.”

---

**Glen Martin** covered science and the environment for the San Francisco Chronicle for 17 years. His work has appeared in more than 50 magazines and online outlets, including *Discover*, *Science Digest*, *Wired*, *The Utne Reader*, *The Huffington Post*, *National Wildlife*, *Audubon*, *Outside*, *Men’s Journal*, and *Reader’s Digest*. His latest book, *Game Changer: Animal Rights and the Fate of Africa’s Wildlife*, was published in 2012 by the University of California Press.

# A Biotech Wave at SXSW

*Karen Ingram*

South By Southwest (SXSW) is a music, film, and interactive festival that happens in Austin, Texas every March. Since its inception in 1987, it has evolved to become a launching pad for a wide range of innovative technologies (Figure 9-1).

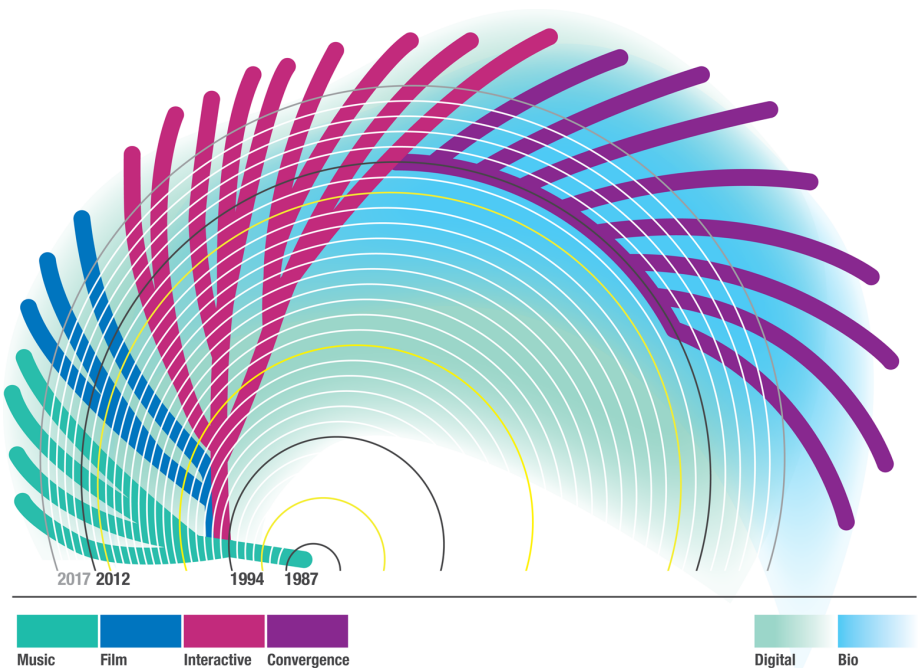


Figure 9-1. The SXSW festival, imagined in an illustration as a phylogenetic tree, reflects cultural attitudes toward tech adoption.

The rumblings of a biotech wave is indicated in a closer look at the evolving programming of the ecosystem. Biotech labs, startups, academic labs, investors,

and incubators like Craig Venter, 23andMe, UBiome, Genspace, La Pailasse, Ginkgo Bioworks, BioBuilder, Synbiota, Open BCI, Glowing Plant (TAXA), Backyard Brains, New Harvest, Perlstein Lab, JBEI, Bioeconomy Capitol, the crowd-funding platform Experiment, and more all indicate that a storm surge is rising.

Hugh Forrest, executive director of SXSW, forecasts:

*We see biotech as a strong area of growth for SXSW in 2017 and beyond. This belief stems from a lot of different factors. One is the explosion of wearable technology at SXSW, a lot of which has a strong connection with health and biotech. Also important is that so many of the startups we see at SXSW have a biotech focus—and as more of these startups find success in Austin in March, they should attract even more biotech related entrepreneurs to the event.*

## The Evolution of the Festival

### From Live Music to Social Impact

The SXSW “ecosystem” originated 30 years ago as a music conference. It reflected the spirit of Austin, whose official city slogan is “The Live Music Capitol of the World.” In 1994, two new pillars to the festival were introduced; Interactive (then called Multimedia) and Film. Over the past few years the Interactive festival has grown to be the largest of the three, with 37,660 festival attendees in 2016. Also in the festival family are SXSW Eco and SXSWedu.

Through my involvement with the festival since the early 2000s, I’ve begun to think of SXSW Interactive as a festival of convergence. Forrest, director of the SXSW Interactive since its inception, observes, “Today, these elements [music, film, and interactive] are very much intertwined and the lines that once separated them have become completely blurred.” In the past two years, Interactive has sprouted “convergence” topic areas such as government, style, health, food, social impact, design, journalism, sports (the list goes on) to allow for overlapping arenas focused around rapid emerging technologies (Figure 9-2). A cascade of exhibit venues such as the Gaming Expo, SXSW Create (a maker/tinkerer festival that’s hosted many a biohacker), and the MedTech Expo have also emerged in recent years.

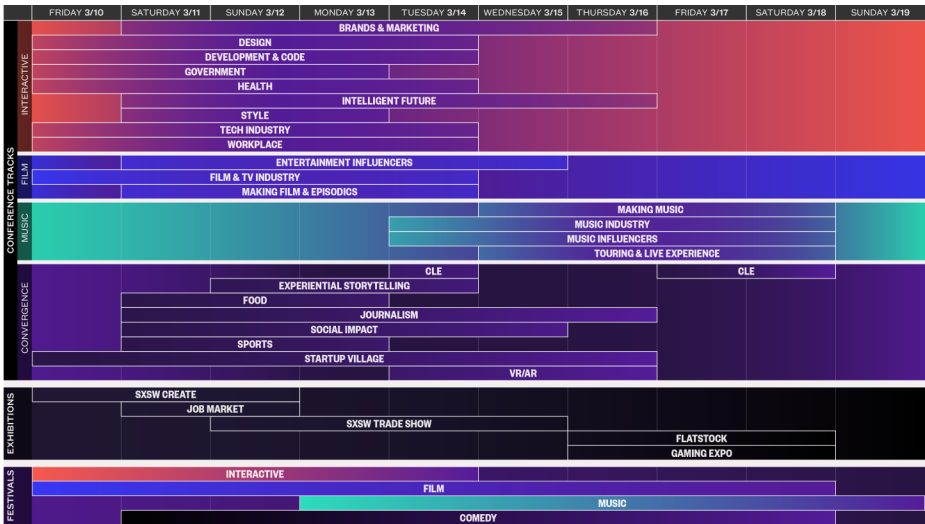


Figure 9-2. The plan for the 2017 conference series. (Source: SXSW.com.)

## The Early Adopters Convene

The first days of SXSW saw droves of visionaries, early adopters, software developers, coders, and designers. Forrest said, “SXSW Interactive launched as SXSW Multimedia in 1994. At that point, most of our programming was focused on how CD-ROMs would revolutionize the modern world.”

Sarah Campbell, a digital strategist who attended the very first SXSW Multimedia, said, “At the time, a lot of people thought the Internet was a fad. They didn’t understand the point of the event; there were very few people.” The early days of the festival were so intimate that the prolific Austin-based sci-fi writer Bruce Sterling would host a “come-to-my-house” party, open to all festival attendees. Now he holds the closing keynote to a live audience of thousands.

Those early days of the Internet were filled with skepticism, Campbell said, around claims of how important the Internet would become. “Nobody thinks it’s real; it’s all very sci-fi,” he said. “SXSW was a haven for techno-dreamers. But it’s built [the SXSW community] on the fundamental belief that we need to evolve and grow, to always think about the future.”

## A Launch Pad Launches

As the Internet matured and user-focused content platforms became more prevalent than handcoded sites, and bloggers came into the fray. Platforms expanded, creating greater access to technology and more diverse voices. Storytellers, content

producers, journalists, marketers, and brands all found a reason to come to Austin for SXSW. Social issues bubbled up, human rights, diversity, and the planet became only some of the topics that the technology festival addressed. One could say that SXSW grew not unlike a rapidly evolving phylogenic tree of life, powered by technological advancement.

The festival began to gain a reputation as a launching spot for hot applications. Twitter launched as a blog in 2007, and posed the question “What are you doing?” It’s now a major destination for what’s happening. In 2009 two geolocation apps launched: Austin bred Gowalla and Foursquare from NY, as did the ride-share app Uber. SXSW became a springboard for thought leadership, social good, and business practices, too. A fascinating evolution of SXSW is available on its website.

SXSW remains true to its tech-enthusiast DNA. “These days, we address everything from social media to sports to self-driving cars,” Forrest said. “But the bottom line of creative people discovering innovative new approaches really hasn’t changed all that much.” Paola Antonelli, Neil DeGrasse Tyson, Elon Musk, Craig Venter, Anne Wojcicki, Tim O’Reilly, NASA, Burning Man, DARPA, and countless designers, artists, entrepreneurs, filmmakers, marketers, thinkers, do-ers, makers, movers and shakers, have presented, attended, and delighted in the innovation hash that happens in early March in Austin each year.

## A Profile Of Attendees

The festival’s programming and audience growth clearly show the reach of emerging technologies to all areas of business. “In recent years, the audience for SXSW has moved from technology enthusiasts and first adopters to more attendees who reflect more mainstream business interests,” Forrest said (Figure 9-3). “This shift in audience reflects how new media technologies have become more and more integral to all business operations.”



Figure 9-3. SXSW 2016 Top 10 Types of Business, Interactive Conference Registrants. (Source: SXSW Interactive Demographics, v 4.0.)

# The Current Biotech Landscape at SXSW: A Mainstream Audience of Tech Enthusiasts

## Storytelling

Director of MedTech, Dana Abramovitz, cited the panel topic “Hacking the Brain to Treat Paralysis” as an example of convergence in the medical realm. Patients share their stories and experiences alongside researchers, employing “storytelling” through personal experience as a means to explain cutting-edge technology. The panel included a young man who had experienced traumatic brain injury in high school and had undergone neural bypass surgery in 2014, as well as the fascinating team who had aided in his remarkable recovery: a neurosurgeon, a psychiatrist, and an electrical engineer. Abramovitz noted the result of attending programming such as this. “People come, get inspired, and are incentivized to put together a panel for the following year,” she said.

## Inspiration, Serendipity, and Collaboration

Inspiration is always a key theme at SXSW. Desktop Genetics, winner of the SXSW Interactive Innovation Award in the Health, Med, and Biotech category, concurred with that sentiment in a post-festival blog post. Planning committee members and staff strongly suggest that attendees deviate from their area of expertise to fuel the spark of innovation and creativity. Serendipity is core to the fabric of SXSW. As Abramovitz puts it, with the newly minted Health track, doctors who are involved in proposed programming take the perspective, “I don’t know why I am here.” However, once they become immersed in the experience, they get it and see the value. Speaking from the perspective of a physician, Abramovitz said, “Where else can I [a medical professional] be in a room with designers and developers? Doctors, providers, payers [health insurance], patients, have a voice... designers, entrepreneurs, regulators—everybody—should be at the table.”

## Public Engagement

In direct alignment with the spirit of serendipity, the hacker/maker arm of SXSW, SXSW Create, launched in 2012. Create offers hackers and makers a demonstration space in a grass-roots setting, apart from the conference (Figure 9-4). I hosted the first “Biohacker Meetup” at Create in 2014, which consisted of an impromptu panel composed of myself, BioHacker Josiah Zayner, DIYBio founder Jason Bobe, and UT Austin professor and iGEM team leader Jeffery Barrick. The upshot was that there is a great wealth of curiosity around biotechnology among the cadre of

tech enthusiasts that attend SXSW. The response was attentive and inquisitive, and not at all squeamish or fearful. It's apparent that the growing biohacking community provides unprecedented access to techno-enthusiasts of all shades, as well as education and awareness that address curiosity over fanning the flames of fear around emerging biotechnologies.

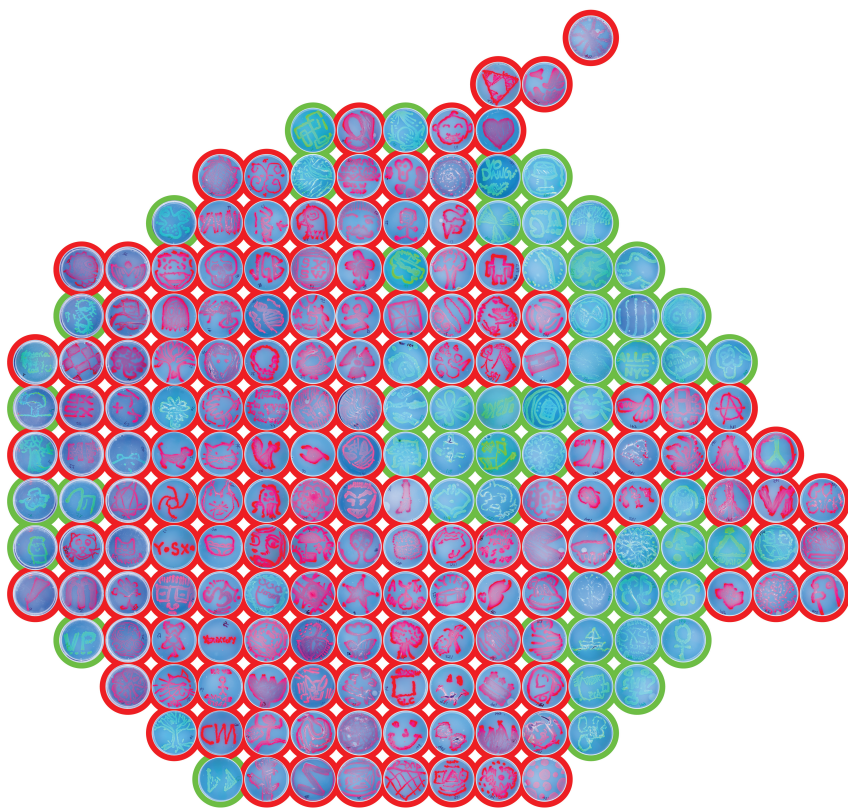


Figure 9-4. “Unicorn Mosaic” composed from bacterial paintings created by SXSW attendees at SXSW Create, 2013. Facilitated by the Brooklyn-based bioart collective, Cut/Paste/Grow.

Other arenas in which biotech has an emerging presence include SouthBites by SXSW, SXSWGGood, SXSW Eco, and SXSWedu.



## Untapped Opportunities

Based on the positive attendance from a fashion and technology 2012 Meetup, SXSW launched SXstyle in 2014. Kelly Krause, SXstyle director, breaks down the content as “...the future of retail, innovative design, fashion startups and apps, and sustainability.” SXSW 2017 includes a Style track that will run March 11–17.

“The Fashion and MedTech worlds are already colliding,” said Krause, “from L’Oreal creating a patch to help detect skin cancer to Under Armour working on connected apparel to track fitness; it’s really neat to see the innovation.” Many textile innovations come in the form of conductive fabrics, like Google’s “Made With Code” initiative, a collaboration between fashion engineer Madison Maxey and designer Zac Posen.

“Biotech is a burgeoning trend and was represented throughout SXSW Interactive, predominately in our Health and MedTech sessions,” Krause said. “We didn’t have any style-related biotech programming in 2016.” Perhaps 2017 will prove to be different, as biotech startups crop up, and applications for biomaterials find a home in fashion and textile design.

## Participating in the Festival

The bulk of SXSW programming is selected through PanelPicker. PanelPicker allows for the community as well as the staff and advisory board to vote on what types of programming should be accepted to the festival for the following year. For 2017 programming consideration, the panel picker was open from June 28–22.

Announcements for festival programming usually roll in in early fall, with the exception of some keynote speakers. Jennifer Doudna, co-inventor of CRISPR-Cas9, has been announced as a keynote speaker for 2017.

To whet your curiosity, examples of topic ideas that were accepted to previous SXSW conferences include the following:

- “Reprogramming the Genome with CRISPR”
- “Is Your Biological Data Safe?”
- “The Future of Citizen Science”
- “Collaborative Biohacking”
- “Surviving the Red Planet”
- “Synthetic Biology: Learn, Do, and Dream”

- “Decoding Our Bodies: A New Era of Citizen Health”
- “Inner Space: Bioelectronics and Medicine’s Future”
- “Why the World Needs Biological Design”
- “Teachers: The “Secret Sauce” for Innovation in the Classroom”

If you find yourself wanting to attend SXSW in 2017—and I hope you do—book accommodations early. With the massive influx of people in Austin for the festival, hotels fill up quickly. Badge sales for the 2017 festival began August 1, 2016. For more information, visit [sxsw.com](http://sxsw.com).

---

*Karen Ingram is a designer, artist, and creative director. She is a planning committee member of SXSW Interactive and a 2015 Synthetic Biology LEAP fellow. Ingram recently completed work on visual elements for Biobuilder (O’Reilly, 2015). Follow her [@krening](https://twitter.com/krening) on Twitter.*

# HiveBio: Access Granted

*Max Showalter*

Bergen McMurray has cherry-red hair and shoulders painted with black ink tattoos. You probably wouldn't be surprised to learn that she's worked as an artist, a photographer, and a graphic designer. But McMurray has a few more titles that set her apart. She's a loving mother, she's [published in \*Nature\*](#), and in her spare time, she happens to lead a buzzing colony of do-it-yourself biologists.

Do-it-yourself biology, or *DIYbio*, is part of a growing movement which seeks to give people the tools to learn without formal education. "Punk," "maker," or "hacker"—the DIY community can go by many names, all of which embody the same directive: democratize science. DIYbio empowers the individual to seek scientific knowledge through trial and error, getting hands dirty in books and bench work alike. In the same way computer tinkering and garage electronics powered the human element of the digital revolution, DIYbio provides the average individual a pathway to biological literacy.

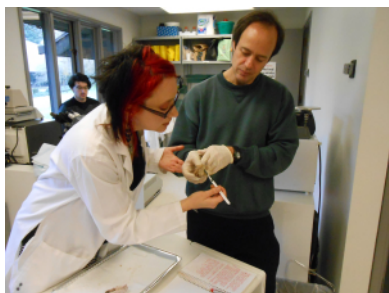
McMurray's philosophy aligns to the ethic of the DIY community. "I believe that everyone, regardless of education level or income or age, should [have] access to the resources needed to learn and build and create," she said.

But for biology, finding this access can be hard to come by. Expensive and potentially hazardous chemicals, specialized equipment, particular safety and environmental contamination protocols: all of these can present insurmountable hurdles. Even the most basic equipment—a simple pipette, for example—can cost hundreds of dollars, while autoclaves or necessary chemicals are locked behind laboratory doors accessible only to academics or industry types.

In Seattle, and other cities, the answer comes in the form of a DIYbio community lab.

## Community Labs: A Public Space for Science

Tucked away in a quiet corner of the city, HiveBio bustles with activity on any given Saturday as curious Seattleites of all ages and backgrounds gather for the chance to experience biology hands on. In the laboratory, two tables take up most of the room, while on the walls shelves hold donated equipment, latex gloves, and Rubbermaid containers filled with multicolor, plastic, microcentrifuge tubes. Around these tables, DIY enthusiasts and curious minds gather as McMurray guides a tour of the human brain (Figure 10-1).



*Figure 10-1. HiveBio CEO and Cofounder Bergen McMurray uses hands-on learning to teach about the brain.*

HiveBio is one of a growing rank of spaces known as community labs. These institutions are the forefront of the DIYbio movement, offering a place to learn fundamentals of biology, a network to meet fellow DIYbio enthusiasts, and an infrastructure to develop and execute your own biology project. You might think of them like community rec centers, but instead of swimming in a pool or taking your child to a day camp, you're exercising your brain with projects and educating your kids with classes. "Community labs are the hub where citizen scientists can find each other, learn new skills, and come together to create something greater than the sum of its parts," McMurray said.

In 2013, this was just the kind of place McMurray was looking for. When her son was diagnosed with bipolar II disorder, she sought to learn everything she could about the human brain. But without a formal degree in science, she hit a road block.

"[N]o lab would grant me access," she said. "It was possible to rent lab space, but the cost was extremely high."

What's a maker at heart to do but build her own solution to the problem?

McMurray set about gathering her dream team. Kathriona Guthrie-Honea, an impassioned and motivated high-school student seeking space for a class project,

became cofounder. Dr. Michal Galdziki, a veteran of the DIYbio scene, was to be chief science officer. And the role chief operations officer was championed by Elizabeth Scallon, founder of a green lab movement focused on recycling unused laboratory equipment. By the fall of 2013, HiveBio Community Laboratory was open for the public.

## A Model of Engagement

HiveBio acts as a DIYbio hub for the Pacific Northwest's maker community, providing a lab space, offering classes, and maintaining active projects. Like any community hub, HiveBio has grown to serve the community's needs. Whereas a [community lab in Brooklyn](#) might cater to local artists, or a [space in Silicon Valley](#) might offer advice for new start-up companies, HiveBio retains a uniquely Seattle feel.

Take Citizen Salmon, the current flagship project of HiveBio. In a region where salmon is king, tracking the origin of store-bought fish connects consumers to the local fishing economy.

While learning lab basics like PCR amplification, coupled with bioinformatic techniques, citizen scientists use genetic markers unique to regional spawns to approximate the geographic origin of their food. Similarly, in a tech-wise city, additional projects at HiveBio focus on open source instrument development on a DIY budget: QPCR and liquid handling instruments top the marquee.

The bread and butter of HiveBio, however, are the 18 unique classes regularly offered on weekends. On the list: squid dissections, building a smartphone microscope, modeling protein folding, the organic chemistry of smell, and, of course, a hands-on tour of the brain. The courses are taught by volunteers, mostly professors at nearby universities, local industry experts, or members of the DIYbio community, and they have proven hugely popular. In the 2013–2014 fiscal year, HiveBio saw over 100 people enroll in its classes.

Of the citizen scientists active in HiveBio, many have little to no formal training in biology, and many, like the ones pictured in [Figure 10-2](#), were young students.

“Educationally, my background is more in the business side,” said Zach Mueller, a local tech employee and longtime DIYbio enthusiast. “It’s definitely been a very diverse set of people. There’s a lot of people like me that don’t have the bio background and are interested in [DIYbio] out of curiosity.”



Figure 10-2. Young students participating in an eyeball discussion typify HiveBio's class format.

HiveBio is also making strides to engage the community beyond its doors. Monthly community discussion groups enfold new members of the DIYbio community, while dedicated volunteers liaise with local educators to bring hands-on STEM education to K12 students.

## The Future Is DIY

Community labs are gaining traction. According to [diybio.org](http://diybio.org), the online organization of global DIYbiologists, at least 34 bio-themed maker spaces exist in the United States, with dozens more across the globe. As an additional measure of making it in the DIYbio world, community labs even [have their own track](#) in the global iGEM competition.

The success of community labs is driven in part by the coordination and efforts of those who've found the secret to making a functioning organization. For its part, HiveBio has initiated an annual conference, SWARM, designed for maker space coordinators in the Pacific Northwest. In 2015, the first annual SWARM brought together 30 organizers from Portland to Vancouver to discuss ways to boost community participation and interest in DIY spaces.

What makes a community lab sustainable? McMurray said that, as with most things, finances are the limiting reagent for HiveBio.

"As with any grassroots organization, money is the largest factor holding us back."

HiveBio, which operates as a 501c3 nonprofit organization, receives most of its revenue from class fees, grant money, and membership dues, a standard trope

for community labs across the US. Equipment comes mostly from donations, and the organization is fueled entirely by volunteer dedication.

For a future when money is no object, HiveBio hopes to expand current operations to touch more lives, aiming to develop a scholarship department to take active learning of biosciences to more at risk communities.

As for Mueller—he envisions his own DIYbio future as an entrepreneur.

“I do have an interest, long, long term, to maybe start a company in the field of biotech. [HiveBio] is just helping me build up my skill set now to understand what the bounds of possibility would be.”

Maybe the future will see “business incubator” added to HiveBio’s resume, but for now, McMurray said HiveBio will keep toward its goal of bringing DIYbio access to all.

“[DIYbio’s] most important role is showing the world that anyone, regardless of age, socioeconomic status, or education level can perform scientific research and be contributing members of the greater scientific community,” she said. “Most of us grow up with idea that science is exclusive to genius. Community labs are breaking down these barriers and teaching people of all ages that science is for everyone.”

You can learn more about HiveBio by visiting the website [hivebio.org](https://hivebio.org) and following [@HiveBio](#) on Twitter and [HiveBio Community Labs](#) on Facebook.

---

**Max (Gordon) Showalter** is a graduate student studying bacterial adaptation to harsh polar conditions for a dual PhD in oceanography and astrobiology at the University of Washington. He is the current editor-in-chief for HiveBio Community Lab and a former Science Communication Fellow at the Pacific Science Center in Seattle, WA. Max has a passion for studying extreme bacterial life on Earth and in the Universe, and for communicating stories in science through written and spoken media. Follow him on Twitter at [@gshowalt](#) and through his website, [gmshowalter.com](https://gmshowalter.com).

