

Science AMA Series: I'm Denis Bauer, a Team leader at Australia's government research organization, CSIRO. We develop BigData and cloud-based software to give researchers a 'CRISPR' look at genome engineering applications. AMA!

Denis_{Bauer}¹*andr/ScienceAMAs*¹

¹Affiliation not available

April 17, 2023

[REDDIT](#)

Science AMA Series: I'm Denis Bauer, a Team leader at Australia's government research organization, CSIRO. We develop BigData and cloud-based software to give researchers a 'CRISPR' look at genome engineering applications. AMA!

DENIS_BAUER [R/SCIENCE](#)

[removed]

[READ REVIEWS](#)

[WRITE A REVIEW](#)

CORRESPONDENCE:

DATE RECEIVED:

November 30, 2016

DOI:

10.15200/winn.148042.27438

ARCHIVED:

November 29, 2016

CITATION:

Denis_Bauer , r/Science ,
Science AMA Series: I'm Denis
Bauer, a Team leader at
Australia's government
research organization, CSIRO.
We develop BigData and cloud-
based software to give
researchers a 'CRISPR' look at
genome engineering
applications. AMA!, *The
Winnower* 3:e148042.27438 ,
2016 , DOI:
[10.15200/winn.148042.27438](https://doi.org/10.15200/winn.148042.27438)

© et al. This article is
distributed under the terms of
the [Creative Commons
Attribution 4.0 International
License](#), which permits
unrestricted use, distribution,
and redistribution in any
medium, provided that the
original author and source are
credited.

When do you envision human trials occurring/the tech being commercially viable, etc? basically whats the time frame for joe average being able to avail of the benefits of genetic engineering via CRISPR?

[randathrowaway1211](#)

University of Pennsylvania trial beginning next year was the first announcement for an application in [human health](#), however Sichuan University in Chengdu surprised the scientific community when they announced that they [already applied](#) the technology for trials in lung cancer. Both applications take cells from the patient, edit them in culture and re-inject them so they are proof of principle studies and mainly aimed at demonstrating that the CRISPR edited cells are safe for medical treatments.

Perhaps you know the youtube channel kurzgesagt. The channel is very good and sticks to the facts (in my experience and opinion). The channel has, some time ago, made a video about genetic engineering and CRISPR. Do you, with your personal knowledge and experience, think this video is accurate? Do you agree with the possible effects/outcomes the video describes? If not, what is your opinion of the possible effects CRISPR could have? Link: <https://youtu.be/jAhjPd4uNFY>

[Captionater](#)

Like [u/Lumene](#) has already explained (rather well!), the video gets the theory of CRISPR correct but oversells the ability of the CRISPR-Cas9 system in its current state. CRISPR has certainly revolutionized genetic research, making genome editing cheaper, simpler and more accessible for all scientists. While CRISPR is certainly simpler than previous methods of genome editing, it still requires significant optimization in terms of target site selection, maximizing on-target effects and minimizing off-target effects. Especially when used in systems more complex than cell lines. We are also limited by our current understanding of genetics. Regulation of the genome is incredibly complex, and changes at one site could have consequences on the regulation of multiple, distal sites. While changing a single nucleotide might fix a genetic disorder, before we make that change we have to be



certain that there will be no unintended side-effects of said mutation. The biggest benefit of CRISPR currently is to basic research. The simplicity of this system makes reverse genetic screens, where a mutation is introduced and the effect on the cell/organisms observed, much easier to perform. This will benefit research immensely, allowing researchers to better understand the regulation of the genome.

I have two very different questions:

How do you think the much-lamented "death of Moore's law" will affect bioinformatics, and particularly genetic research? At the moment it seems that for most research groups, private or department-level servers are sufficient for their current work, but clearly things like GT-Scan2 require much bigger resources. As we see data continue to grow and computing power stagnates, how will this affect the field?

Working in a public research organization and especially on a technology that aims to democratize identification of CRISPR targets, can you talk a little about what you think about the value of open source tooling and how you envision the funding environment for development of the next generation of bioinformatics software?

[divvyflax](#)

Moore's law is the observation that the power of a computer core doubles every two years with recent improvements made by putting multiple core on a node. However, this "free" improvement may not be sufficient to deal with the flood of data and bioinformatics and computing application at large must start to improve the underlying algorithms and/or abandon the on-premise compute cluster as new technology likely require a complete replacement of infrastructure rather than incremental improvements (e.g. Hadoop, though there are efforts for making it compatible with high-performance-compute clusters). Hence breaking free from the underlying compute technology perhaps makes a "compute-power-per-dollar" metric more relevant. So overall compute power will continue to increase and the community will continue to learn new languages and paradigms to make the most of this.

Developing in open source is very important as technology moves too fast for individual (even large) organizations to keep up. Monetization hence needs to come from a different business model than selling the binaries e.g. from the service around it or continuous advancements of the technology.

For research software funding, I think the [model proposed](#) by Dr Vivien Bonazzi at NIH has great merit. NIH plans to put important datasets on a public cloud under a FAIR (Findable, Accessible, Interoperable, and Re-usable) agreement. This data common can be attached to compute with software access point via a marketplace. This enables a merit-based credit system where researchers write grant applications to obtain "NIH credits" that they then can spend on the data/software of their choice, which in turn funds the maintenance and software development of the popular resources with additional funding to new/strategic developments to maintain diversity.

Thanks for the ama, i think crispr is incredibly exciting! How cheap do you see this technology becoming in the future? Will it become cheap/mainstream enough that everyday people will be able to edit their genomes if their doctors tell them for example that they are predisposed to a certain disease ?

Also, do you see any potential use of this technology in anti aging research ?

[Hidden_Troll](#)

Thanks for the question! In answer to your first question: CRISPR has already become the cheapest option for genome editing. Zinc Fingers, a genome editing tool discovered ~10 years ago cost about

\$5000 or more to make. In contrast, CRISPR can be designed for as little as \$30. Similarly, the CRISPR has system has been rapidly adopted, with ~600 publications and >150 patents in 2014 alone. [This article](#) from 2015 sums it up nicely. The cheap price certainly makes it an attractive candidate for use in the clinic. Whether or not CRISPR becomes widely available will depend on the outcome of trials (both currently ongoing and future ones) as well as how it is regulated. However, "editing out" a predisposition is very difficult as the change will have to be introduced in every cell of your body or at least the relevant tissues. As for CRISPR's use in anti-aging research, people have already started looking at its potential use [see this interview with Harvard Professor George Church](#).

How do you get to work for such a company? What paths did you take in your career? Any tips you picked up on the road?

[jangeest](#)

I agree with you that CSIRO is a very exhilarating research organization to work for. As a background for people not familiar with CSIRO, we are Australia's government research agency. CSIRO is with more than 5,000 experts in the top one per cent in 15 of its 22 research fields ranging from Health to Information Technology to Mining and Manufacturing. Our [world-renowned successes](#) include WiFi, the Hendra vaccine and polymer banknotes.

CSIRO is a very inclusive organization and there are hence [multiple avenues in](#) from doing a student internships to postdoctoral positions (we have one [open now!](#)).

I did my undergraduate in Germany, then PhD and two PostDocs at research institutes in Australia before joining CSIRO and taking up the team leadership role two years ago. So as a general tip I'd say collecting experiences from different Institutes and even countries is important.

gt-scan2 looks nice and a great interface for CRISPR design. I have previously coded a basic dual cas9 target finder because it was impossible to find a tool that would work with non-model species.

How easy is it to incorporate another reference genome into your pipeline (locally), and do you intend to give options for non-human species through your webapp in the future?

[danielpass](#)

Thanks for your comment and question [/u/danielpass!](#) Technically, implementing another reference genome into our pipeline (locally or on GT-Scan2) is rather trivial—however as we use epigenetic data in our prediction model, we require this data to be available for any reference genome we incorporate. As this data becomes available, we will likely expand the list of available genomes as we've done with our specificity-predictor [GT-Scan](#), and which now has around 50 reference genomes available (for example different fungi, chordates, plants, fish).

Most discussions about CRISPR which I've seen focus on pros and cons of CRISPR application on a global scale (how can it affect medicine and industry, how can it change our species). There is also a lot of praise for CRISPR as a technique, but I haven't heard anyone criticising it. So my question is:

Could you tell us about some of the issues associated with using CRISPR?

Edit: Formatting

[DarthSmart](#)

The biggest issue with CRISPR is balancing on and off-target effects. While the CRISPR-Cas9 system can be targeted to almost any region of the genome, not every target site is as effective. This can be due to the sequence of the target site itself or the environment the target site exists in. It is therefore important to be able to identify what target sites are the most amenable to editing with the CRISPR-Cas9 system. It's also important to minimize the potential for off-target sites. If the sequence of a target site is similar to other regions in the genome there is the potential for other regions of the genome to be unintentionally edited. Balancing these two requirements can be difficult (and is what we aim to address with [GT-scan2](#)).

My question:

What advice would you provide to a computer science graduate interested in entering the field of Bioinformatics?

My opinion about the benefits:

I personally think that this will completely revolutionize the field of medicine and make traditional drugs obsolete.

[zencodr](#)

Bioinformatics is a very diverse field covering a wide spectrum from biological research using computational tools all the way to designing theoretical algorithms. As a computer scientist you probably want to stay up to date with the latest compute advancement but also develop a solid understanding of the contemporary biology. It is also important to partner with the right people that can complement your area of expertise.

I'm an aspiring physician, and also interested in data science.

Is it possible to combine the 2 into one career?

That is, use data science while still being a physician.

[Malknar](#)

Learning from previous and related cases underpins evidence based medicine. In my opinion, data science provides the technical tools to scale this up to larger cohorts. As more medical records become electronic and get shared within the health care system, mining this data will likely become second nature for physicians. CSIRO's e-health research scientists work with physicians to develop tools from data in electronic medical records (EMR) for identifying risk of readmission of patients through to understanding the efficiency of hospital services. EMR data will also accelerate clinical research - with cloud based tools for genomics and imaging available through new initiatives such as SMART on FHIR apps for EMRs. I therefore think your goal of joining the two disciplines is very promising. This [JAMA](#) paper may be a good starting point.

Will it ever be possible for geneticists to truly 'create', instead of just 'cut-copy-paste'ing genetic information from organism to organism or modifying existing genes? With even CRISPR-Cas9, for all the glowing accolades it draws, being lifted from the natural world, I've noticed a certain reluctance amongst geneticists to ever truly push the bar and engineer capabilities which are new to life. Is this a lack of ability or simply a lack of imagination?

[417kh](#)

The complexity of genes and proteins means it's often easier and more efficient to adopt processes

already present in nature. For example, Cas9 was originally adopted from bacteria, [new versions of the enzyme have been developed for different applications](#).

However, there is research into developing completely novel biological pathways. Synthetic biology combines biology and engineering in order to design and develop artificial biological pathways/components. It's possible to combine distinct components (genes, regulatory elements, functional protein domains etc) in novel ways to solve problems. For example, [a synthetic pathway for the fermentation of artemisininic acid \(a precursor for the potent anti-malarial artemisinin\) was introduced to budding yeast](#), allowing the drug to be mass produced cheaply and quickly. On a larger scale, [The Yeast 2.0 project](#) aims to synthesize an entirely synthetic eukaryotic genome by redesigning the yeast genome.

Just graduated engineering in Australia (Newcastle) and I would love a job at CSIRO but there never seems to be a graduate position. Do you see this changing in the near future? The organisation was one I admired throughout my undergraduate.

[Shazbah91](#)

There are multiple options to [join CSIRO](#). It is always a good idea to get in touch with the research group beforehand and potentially do a student internship before starting a graduate position.

Having two degrees from a good computer engineering school and 10 years working in industry on various technologies, what (if one exists) do you think is a viable path to retrain some of my skills for this new and growing field? Go get another masters degree? Try to find something entry level in the field?

[curiousgeorge84](#)

Solid skills in computing and biology are both essential in modern bioinformatics. However, gaining practical experience from working in a research group might be more beneficial at this stage.

I have a question that is more of a CS question. What type of database back-end are you using? Is it NoSQL or SQL style Database? I know you said you are using AWS for your computing needs. Is this because crowd sourced computing, example being the CERN project reference:

<https://home.cern/about/updates/2013/09/crowd-sourced-computing-platform-reaches-one-trillion-events>, was hard to implement or was/is there a worry about privacy/property information being gleaned?

[caperneignis](#)

For our database backend, we are using a NoSQL solution (AWS DynamoDB) with its ability to easily fit in with our "serverless" solution.

The CERN project you are referring to is a distributed computing solution; we chose to go with microfunctions instead for the parallelisation so it would not have been applicable.

Do you happen to have a RESTful interface to your tool? That would be wonderful for the work I'm doing, an automated CRISPR prediction tool would be wonderful! I'm looking at many hundreds of thousands of sites at once.

[dashtonal](#)

Fantastic question, yes submitting multiple queries and automating how you get the results back is very important. We have an API which allows you to batch submit queries and fetch the result in an automated fashion. Get in touch with us directly to discuss how to use this feature as it is not annotated yet.

How does this tool compare to [chopchop](#)?

[errevs](#)

According to the CHOPCHOP [instructions](#), chopchop scores the efficiency of targets based on their sgRNA sequence. GT-Scan2, however, takes epigenetic information of specific cell lines and tissues into consideration as well as the sgRNA sequence.

How much shame do you feel over the pun in the title?

More seriously, what is/are the CRISPR-related advance(s) you're the most excited about in the next 5-10 years?

[oliwhail](#)

None ;) It got your attention.

The biggest advancement will come from a better understanding of how the cell responds to CRISPR-Cas9 editing. Once the double strand break is induced in the DNA, the cell can repair it via one of two pathways: the error prone Non-Homologous End-Joining pathway (NHEJ: which introduces point mutations and can disrupt genes and other elements) or through the more accurate Homology Directed Repair (HDR: which can be used to introduce brand new sequences to the genome) (see [Mei et al](#) for a good review). Most research so far has focused on how to optimize CRISPR-editing through the NHEJ pathway. While we are still able to use HDR to introduce new sequences, it's at a lower efficiency. While some research has attempted to address this imbalance (see [Chu et al](#), hopefully future research will further optimize how we exploit this pathway, allowing us more control over where and how well we can introduce sequences into the genome.

Will GT-Scan2 be free?

[Lode25](#)

Yes, GT-Scan2 is freely available to the research community. Hence feedback on how to improve it or integrate nicer in experimental workflows are very welcome.

What are your personal aspirations for a scalable cas9 system?

[super_symmetry](#)

I assume you are asking about how to make the computational target site search scalable for increasing number of CRISPR-Cas9 application cases: GT-Scan2 uses serverless AWS Lambda functions, which break down large task into smaller subtasks. For an average run, GT-Scan2 hence triggers 500-1000 individual Lambda functions, which simultaneously update the target site scores for the different putative targets. By massively parallelizing tasks and using Lambda, which instantaneously recruits the appropriate amount of functions, GT-Scan2 can scale to large numbers of users with large numbers of queries.

Thoughts on AWS+Lambda vs Google Cloud+Big Query/machine learning APIs.

[brian073](#)

The key advantage of [AWS Lambda](#) over the comparable services in Azure and Google is that Lambda can [execute Linux binaries](#). This was essential for being able to run some of our components. In contrast [Azure functions](#) run Windows hence executables must be compiled for Window, which is problematic for a large number of academic bioinformatics tools. And [Google functions](#) is currently just in alpha.

Hi Denis,

First of all thank you for doing this AMA!

What benefits does your software have in comparison to other CRISPR target finders (e.g. CHOPCHOP)? Also: how accurate do you think your "activity prediction" is?

More importantly: how big is the significance of better activity prediction? From my experience as a user most people will always test a couple of sgRNAs and for us at least one in four works like a charm ...

All the best

[iloveununoctium](#)

The benefit of GT-Scan2 is that it uses tissue and cell-line specific epigenetic information to inform its predictions (while other target-site finders only use information about the target site sequence). In terms of accuracy, we've tested our model thoroughly using independent datasets and it has a higher accuracy than other published methods we've tested against. As for the significance of better predictions, the real benefit will be for large-scale CRISPR-Cas9 applications. While a researcher looking to edit a single gene has the time and resources to test multiple CRISPR-Cas9 target sites, when you're looking at a larger scale (e.g. a GeCKO study, which targets every gene in the genome) it's not feasible to test multiple targets for each gene. Having the best predictions is therefore vital in order to ensure time and resources aren't wasted.

Do you ever worry that "democratizing" research may be a bad thing? GT-Scan2 seems like a tool that inexperienced researchers who may know a little but not enough could use to produce poor-quality research. Is there a threshold of knowledge that must be reached to use this tool?

[zombiegingerjesus](#)

The research community has always had the collective duty to call out poor research. Using a computational tool is no different than individuals publishing poorly researched outcomes in other disciplines. By providing GT-Scan2 to the research community as a whole we are helping to better identify and call-out the small cases of uneducated or fraudulent results.