We will begin shortly

BDC Community Hours

Art of the Start

January 17 at 2 PM ET

While you wait, register for February:

bit.ly/BDC-Feb

NIH National Heart, Lung, and Blood Institute | BioData CATALYST®

The Art of the Start

Tips from experienced researchers

Zelia Worman, Velsera, Director of Researcher Engagement/Education

"Try things outside of your comfort zone.

“There are lot of interesting things out there—inside and outside the academia —that are interesting, engaging, and need your expertise.

“Start by looking at detail opportunities, ways to volunteer, and explore how your interests could become your career. Above all, have fun!"

We will begin shortly

BDC Community Hours

Art of the Start

January 17 at 2 PM ET

While you wait, register for February:

bit.ly/BDC-Feb

NIH National Heart, Lung, and Blood Institute | BioData CATALYST®

The Art of the Start

Tips from experienced researchers

Ingrid Borecki, BDC Steering Committee Chair

“Mentors are extraordinarily valuable assets in the development of an investigator's career.

“Find people who embody who you want to be, where you want to go and engage them - they can come from various areas of your life or workplace, you can have several, and they may change over time.

“Their wisdom will help you envision and navigate the road ahead.”

We will begin shortly

BDC Community Hours

Art of the Start

January 17 at 2 PM ET

While you wait, register for February:

bit.ly/BDC-Feb

NIH National Heart, Lung, and Blood Institute | BioData CATALYST®

The Art of the Start

Tips from experienced researchers

Cera Fisher, Velsera/Seven Bridges User Engagement Specialist

"Even before you have your data access in place, or before you have fully identified what cohort you're going to work with, you can do test runs of your analysis with data like the data you're going to use.

"Use the public datasets, like 1000 Genomes, and make *tiny* test datasets that you can run all the way through your expected analysis pipeline, that will run fast (and fail fast if they're going to fail).

"The goal is to not do development/troubleshooting on your whole analysis dataset."

We will begin shortly

BDC Community Hours

Art of the Start

January 17 at 2 PM ET

While you wait, register for February:

bit.ly/BDC-Feb

NIH National Heart, Lung, and Blood Institute | BioData CATALYST®

The Art of the Start

Tips from experienced researchers

Dave Roberson, Velsera/Seven Bridges Community Engagement Manager

"After onboarding to the BDC Platform, the tendency is to wait to your data access permissions for your specific phenotypes and genotypes are approved AND that data is situated/ingested in BDC.

"To make the best of your start up time is to work on refining your analysis objectives and testing your analysis steps on the platform. Testing is best done with a subset of data anyway.

"I recommend either synthetic data on an open dataset. This way, when your data is fully available you can hit the ground running."

We will begin shortly

BDC Community Hours

Art of the Start

January 17 at 2 PM ET

While you wait, register for February:

bit.ly/BDC-Feb

NIH National Heart, Lung, and Blood Institute | BioData CATALYST®

The Art of the Start

Tips from experienced researchers

Tony Patelunas, Velsera, Director of Health Initiatives Partnerships

"Cloud computing is faster and more scalable than the high performance cluster. It's faster because there are more resources available and you can scale to what you need.

"With the cloud you are also not stuck in a queue. There was a job I was running one time...I had to wait two days because I needed a whole node and there were a lot of people in front of me using the node. On the cloud it is all simultaneous and you can get your analysis done fast."

We will begin shortly

BDC Community Hours

Art of the Start

January 17 at 2 PM ET

While you wait, register for February:

bit.ly/BDC-Feb

NIH National Heart, Lung, and Blood Institute | BioData CATALYST®

The Art of the Start

Tips from experienced researchers

Paul Kerr, RENCI, Project Manager BDC Coordinating Center

"New BDC users can apply for 500 cloud credits, known as pilot credits, and many analyses can be completed for that amount or less.

"Cloud credits can be used to determine how much a full analysis will cost, and can be applied to different workspaces...you can try different workspaces to make an informed decision about which is the best workspace for your research.

"And, if you like, you can transfer unused credit from any one workspace to another workspace of your choice."

We will begin shortly

BDC Community Hours

Art of the Start

January 17 at 2 PM ET

While you wait, register for February:

bit.ly/BDC-Feb

NIH National Heart, Lung, and Blood Institute | BioData CATALYST®

Interact with today's slides:

BDC Community Hours

Art of the Start

January 17 at 2 PM ET

NIH National Heart, Lung, and Blood Institute

BioData CATALYST®

Statement of Conduct

The BioData Catalyst Consortium is dedicated to **providing a harassment-free experience for everyone**, regardless of gender, gender identity and expression, age, sexual orientation, disability, physical appearance, body size, race, or religion (or lack thereof). We do not tolerate harassment of community members in any form. Sexual language and imagery is generally not appropriate for any venue, including meetings, presentations, or discussions.

Resource: Statement of Conduct

Hosts and Support

Jasmine Olvany, PhD
BDC Fellow and Alumna, currently Research Project Manager

Emily Hughes
Bioinformatics Systems Analyst, Harvard Medical School

Nathalie Volkheimer
User Engagement Specialist, BDC Coordinating Center

Amber Voght
Instructional Design Specialist, BDC Coordinating Center

Session Materials and Housekeeping

- We encourage you to submit unanswered questions, no matter how big or small, to our help desk
- Join the ecosystem: https://biodatacatalyst.nhlbi.nih.gov/contact/ecosystem
- Check your email inbox and the community forum materials, links and videos. Please pass them on to your colleagues and networks.

Serendipity, rabbits, and the challenge of big data

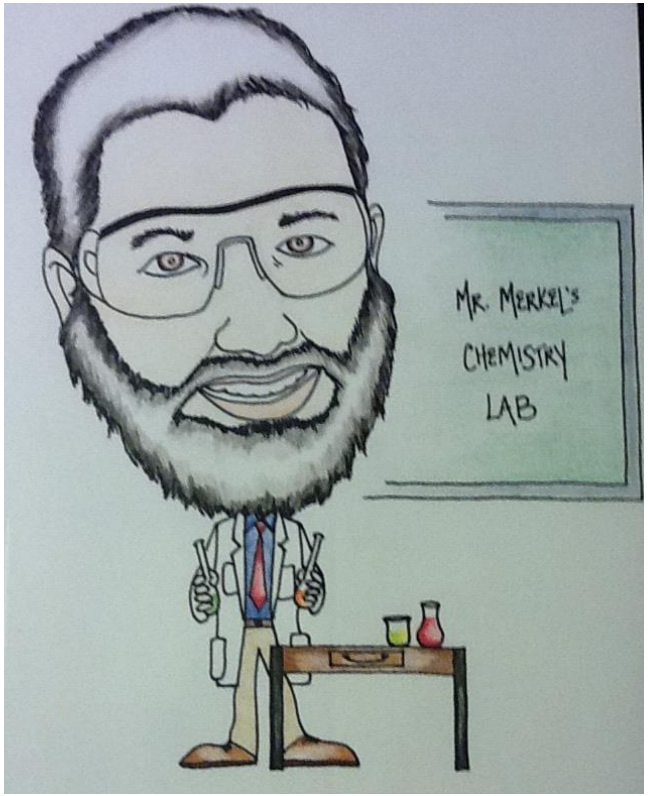
Jasmine Olvany, PhD
BDC Fellow and Alumna, currently
Research Project Manager

Meet Jasmine Olvany, Ph.D., NHLBI BioData Catalyst fellow alumna.

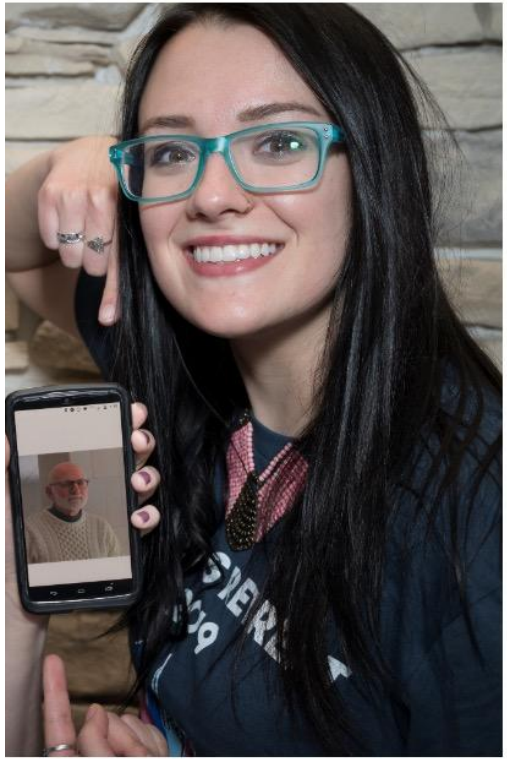
Dr. Jasmine Olvany studied the genomic sequencing of the *Plasmodium* parasite species that causes malaria, a devastating mosquito-borne disease that is a leading cause of death in developing countries.

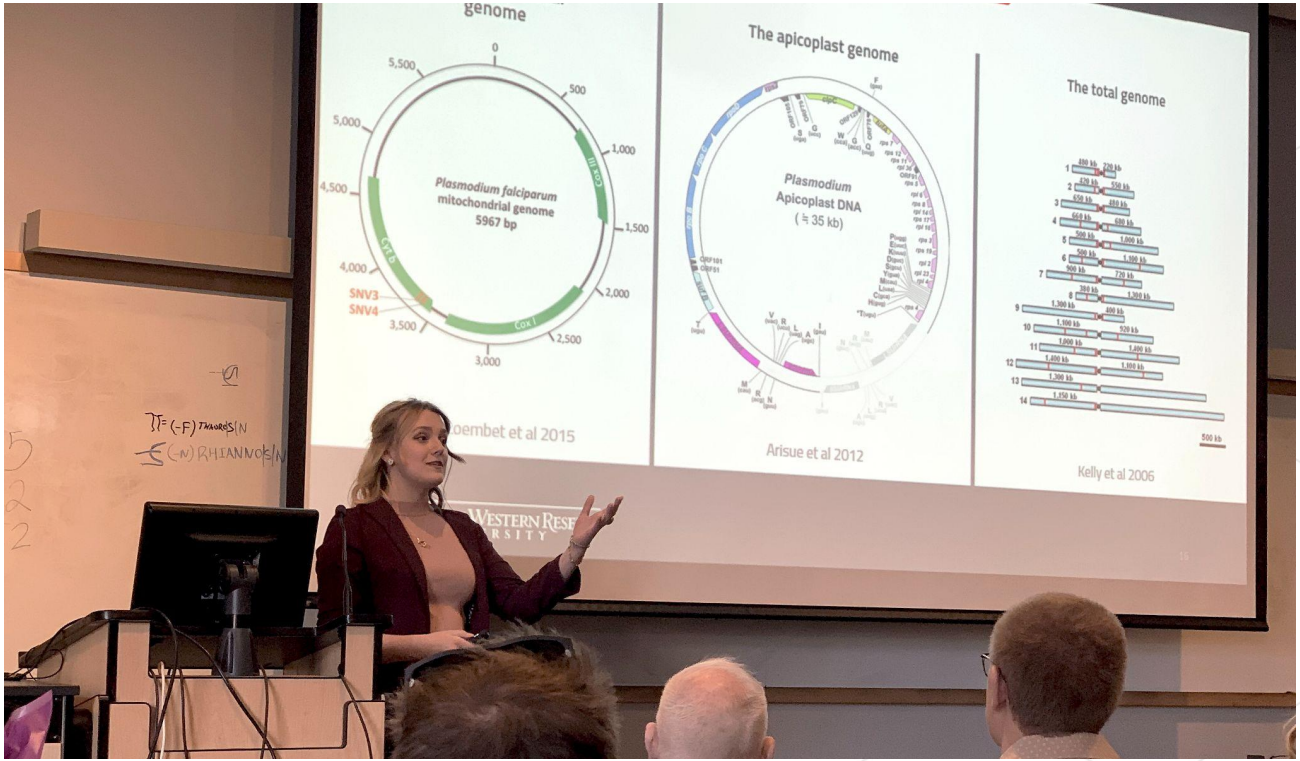
Dr. Olvany says that NHLBI BioData Catalyst® (BDC) has been a gamechanger for her research.

The impact of a rockstar chemistry teacher



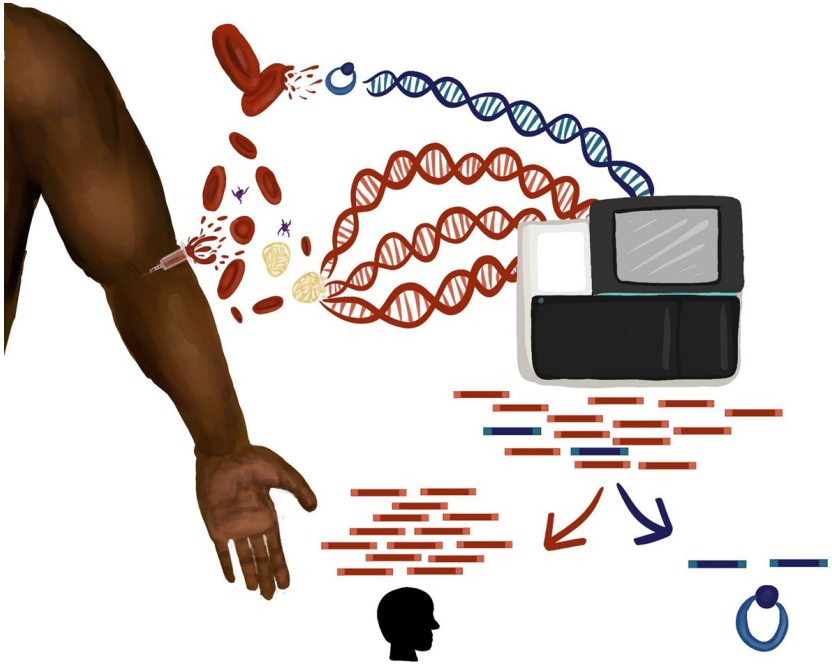
Enter BioData Catalyst



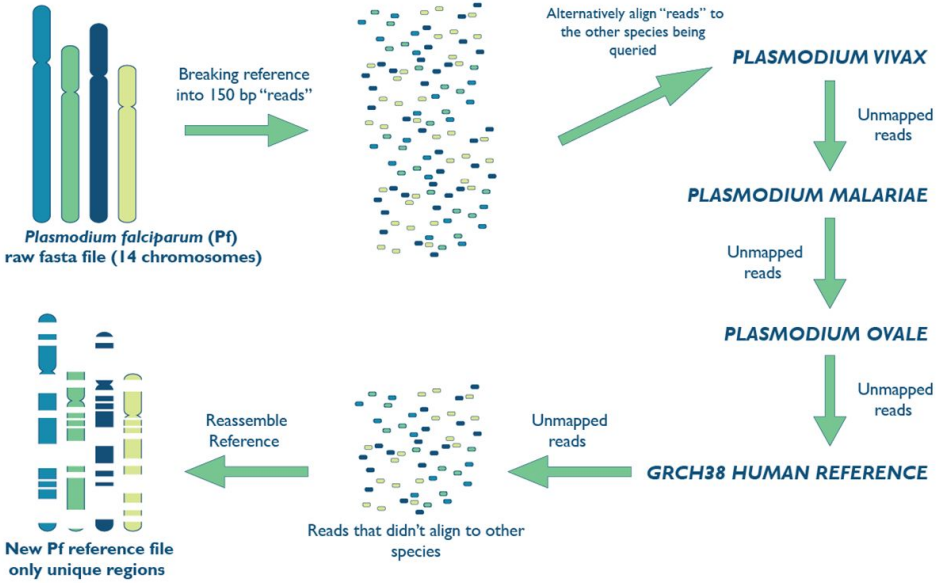


BDC Research- The Theory

Sourcing the data:



Finding the target:

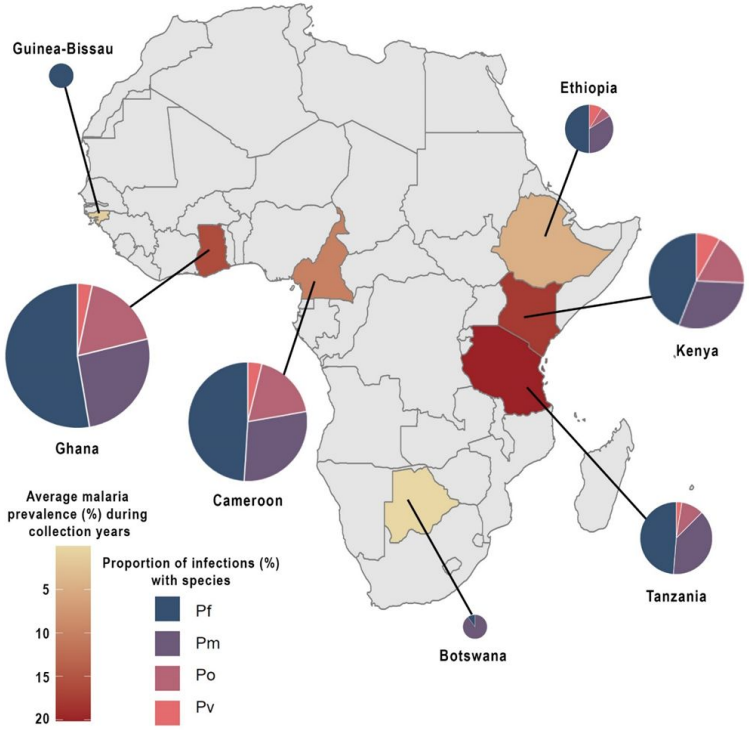


BDC Research- Cloud Platform driving research

In comparison to the gold standard using unaligned reads from WGS performed reasonably well:

86.2% agreement for binary calls (infected/uninfected)

BDC Research- Big Picture Results



PF: 11.1%-100%

PM: 0%-100%

PO: 0%-31.3%

PV: 0%-14.3%

BDC Research- Pushing it a little farther

Not only can we pick up the presence of the parasite from the human WGS sequence, but we can even detect variation in the parasitic genome with global health importance

Gene	Codon	Wild type amino acid	Mutant amino acid	Resistant allele frequency (%) by country					
				Cameroon	*Ethiopia*	*Ghana*	*Guinea-Bissau*	*Kenya*	*Tanzania*
sulfadoxine-pyrimethamine									
pf-dhps	436	S	A	53.8	0	80.4	0	9.1	0
	437	A	G	100	50	82.7	0	90.9	57.1
	540	K	E	0	0	0	0	100	100
	581	A	G	18.2	0	0	0	0	25
	613	A	S	15.4	0	17.1	0	0	0

Lessons Learned on BDC

1. The computational power behind cloud platforms like BDC is invaluable to big data research

2. Presenting to people who are excited about your research helps keep you passionate about the work

3. Community support is sometimes the only way to solve difficult problems, so put your pride aside and ask

The rabbit closes the deal

From big data to clinical trials

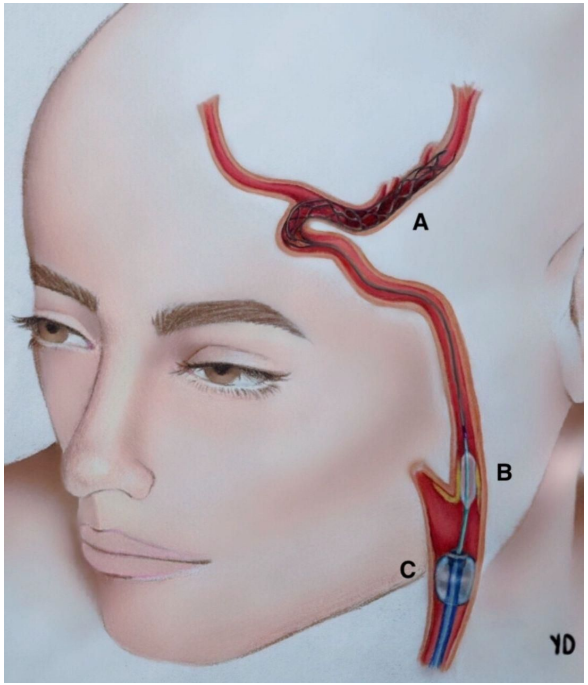
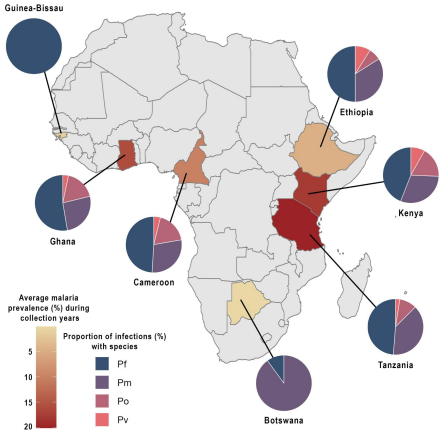
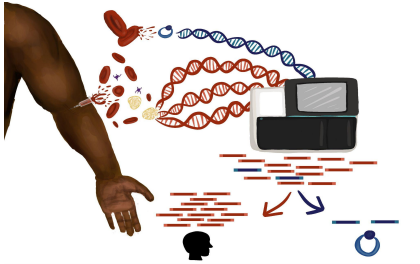


Illustration taken from Sultan-Qurraie et. al, *SEIMLESS,* Stroke, 2019

ICASSO
The Tandem Trial

Tip 1: Chase the people, not the science

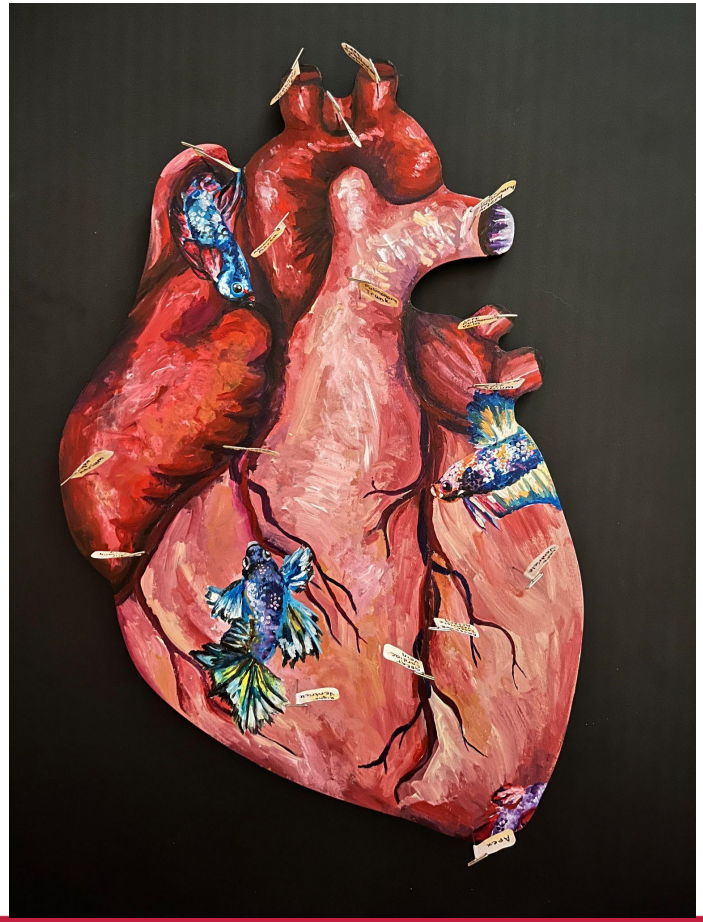




Tip 2: Build your community - and resilience

NIH National Heart, Lung, and Blood Institute | BioData CATALYST

Tip 3: Life happens - be creative and connect



Thank you, Jasmine!

New feature live demo: BDC *Powered By* Open PIC-SURE

Emily Hughes
Bioinformatics Systems Analyst,
Harvard Medical School

Meet Emily Hughes, one of the platform representatives on the BDC team.

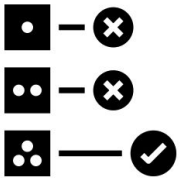
Emily studied bioinformatics and has experience optimizing pipelines for genetic colocalization and statistical analyses.

Emily is part of Paul Avillach's lab at Harvard Medical School. Here, she works to improve search, exploration, and integration of many heterogeneous data sources to enable science.

From Emily's Journey: Lessons Learned

Teachers & Mentors

Are you the next rockstar teacher/mentor for a potential researcher in your life?



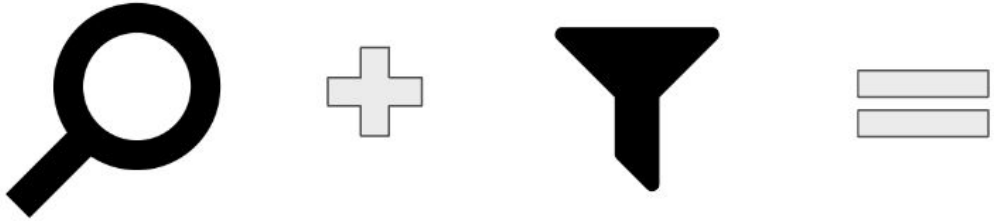
Trial & Error

What doesn't work is as important as what does.

Try everything and rule out what's not for you.

Empowering researchers to access data

BDC Powered by PIC-SURE facilitates approachable research for all skill levels.



Search at the variable value and genomic variant level

Apply filters to create a cohort

Dataframe ready for research without opening any files or mapping to data dictionaries

Open PIC-SURE (without login) in the BDC Ecosystem

Open PIC-SURE promotes **equity** to data exploration by **allowing all researchers** to discover and interact with data available on BDC. As a publicly available website hosted in the FISMA boundary, Open PIC-SURE **decreases the barrier** to data exploration.

Researchers can:

- Search the entire BDC data dictionary
- Apply variable value-level filters
- Retrieve obfuscated, aggregate counts of cohorts
- Conduct feasibility studies for exploration and GRANT writing at any time
- View variable distributions based on a selected cohort

Open vs. Authorized PIC-SURE

	Open PIC-SURE	Authorized PIC-SURE
Overview	Allows any user to search any clinical variable in PIC-SURE and return aggregate, obfuscated participant counts	Allows users with dbGaP authorization to access participant-level data
Access authorization	No approval or account required; publicly available	eRA Commons account required
Data types	Destigmatized clinical variables	All phenotypic and genomic data
Use case	Explore datasets to request access to based on query of interest	Filter datasets to create cohort of interest to run analyses

Last Tip: Ask Questions.

Often smart people feel silly asking questions, but research journeys are as much about questions as answers.

So, please ask!

Until Next Time

- This week's materials coming to you via email: pass them on!
- Subscribe to NHLBI on LinkedIn and Twitter/X
- Repost on your socials using **#BioDataCatalyst**
- NHLBI Social Resources for Heart Month (February) are online now

- February is also African American History Month, join us in celebrating inspiring scientists like Dr Charles R Drew: responsible for the blood bank.

Join us
Wednesday February 28th at 1pm

Register now: https://bit.ly/BDC-Feb

Community Hours with Clinician Scientist (Cardiologist) Jose Vargas

- Previous Postdoc Fellow at NIH Intramural Research Program
- Cardiology Fellow at Johns Hopkins
- Currently at US Department of Veterans' Affairs and working with Jorge Kizer's Lab at UCSF
- Combining heart imaging (for phenotype and covariates) and SNP data (TOPMed MESA and SOL).

NIH National Heart, Lung, and Blood Institute | BioData CATALYST

Join us
Wednesday February 28th at 1pm

Register now: https://bit.ly/BDC-Feb

Community Hours with Clinician Scientist (Cardiologist) Jose Vargas

Insights from Genome-Wide Association Studies: Linking Genetics and Imaging

This talk will go through an overview of basic GWAS and imaging studies and how they are being used in state of the art genetic analyses of cardiovascular disease. A special emphasis will be placed on the importance of diversity in genetic studies.

NIH National Heart, Lung, and Blood Institute | BioData CATALYST

Join the ecosystem

Join the BDC Community

https://biodatacatalyst.nhlbi.nih.gov/contact/ecosystem