

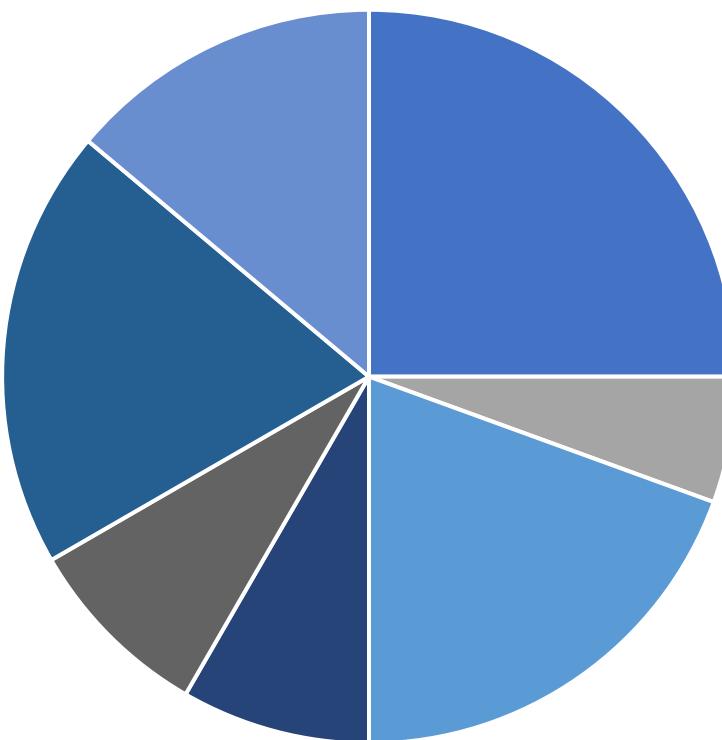
# Making HTE data useful

Jessica Sampson





# Estimated project breakdown over 2021-2023



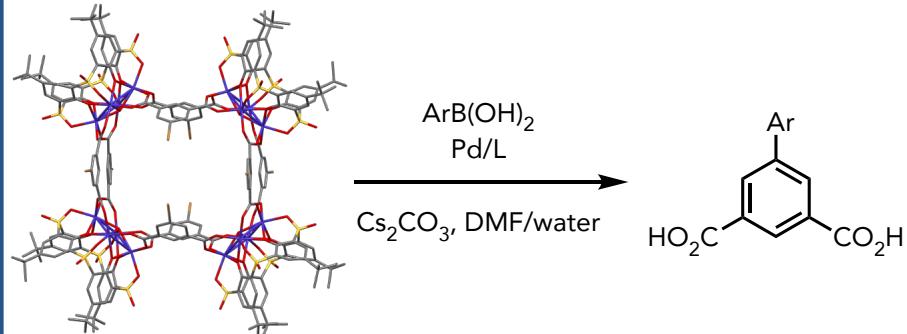
- Asymmetric catalysis development
- Methods for medicinal chemistry
- ML efforts
- Process chemistry and adjacent
- Analytical methods development
- Total synthesis
- Miscellaneous

Includes planned and/or proposed work; numbers are estimated number of projects



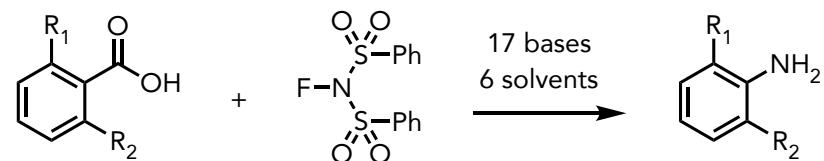
# Select projects from 2021-2023

## Post-Synthetic Modification of a Porous Cage



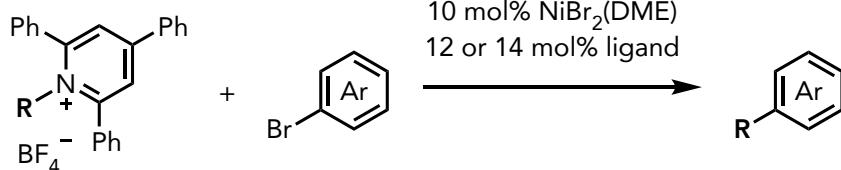
With Mike Dworzak and Prof. Eric Bloch (UD/IUB)

## Solvent/Base Effects in a Oxidative Amination



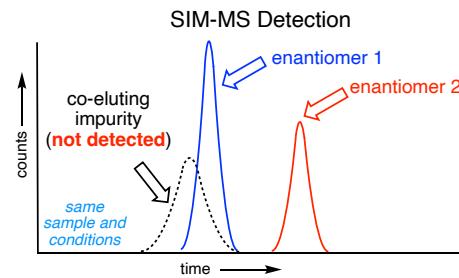
With Rabina Basnet and Prof. Jessica Hoover (WVU)

## Scope of Deaminative XEC



With Bria Garcia, Dr. Jiantao Fu, Prof. Mary Watson (UD), and Dr. Dipannita Kalyani (Merck)

## Quantitative MS for ee Determination in HTE



With Dr. Katerina Korch, Jacob Hayes, Raphael Kim, Dr. Austin Kelly, and Prof. Donald A. Watson (UD)

Storing data generated from HTE campaigns in an understandable and reusable format is the first step towards being able to understand broader trends



# Unique data sets are generated in academic HTE

Reaction type	Average Pharma	Company/Institution							
		Amgen	AZ	BMS	GSK	Merck	Pfizer	Lilly	UPenn
Biocatalysis	19	19	14	20	24	22	14	19	0
Suzuki-Miyaura	11	23	9	9	11	3	7	16	27
Buchwald-Hartwig	8	13	6	7	6	5	3	19	16
Chiral Salt Resolution	8	19	4	7	0	3	6	13	0
Heterogeneous Catalysis	6	0	0	8	12	2	21	0	0
Solvent/Base	5	0	5	6	4	9	7	6	1
Scavenger Screening	4	8	8	4	1	10	1	0	0
Other Reaction Types	4	2	3	7	1	12	1	5	7
Asymmetric Hydrogenation	4	5	9	4	2	6	3	0	1
Pd-Borylation	4	0	2	4	11	1	3	3	4
Other Cross-Coupling	3	1	3	4	0	5	2	6	1
Ullmann Coupling	3	0	12	0	1	2	1	2	1
Lewis/Protic Acid	2	0	1	5	3	4	2	0	1
Pd-Carbonylation	2	1	3	1	6	1	2	0	1
Stoichiometric Amide Formation	2	0	0	4	2	1	4	2	0
Heck Coupling	2	0	4	0	3	2	3	1	1
Photo-Mediated Catalysis	2	4	1	1	1	6	0	0	27
Negishi Coupling	1	2	0	0	0	1	6	1	0
Pd C-H Activation	1	0	1	3	0	1	2	1	2
Organocatalysis	1	0	0	1	2	4	1	2	1
Sonagashira Coupling	1	0	1	1	2	1	1	3	1
Catalytic Amide Formation	1	0	3	0	2	0	0	0	0
Ir-Borylation	1	0	4	0	1	0	1	0	1

Figure from Mennen and coworkers, Org. Proc. Res. Dev. 2019, 23, 1213-1242.

X  
X

X  
X  
X  
X

X  
X

X



# Unique data sets are generated in academic HTE

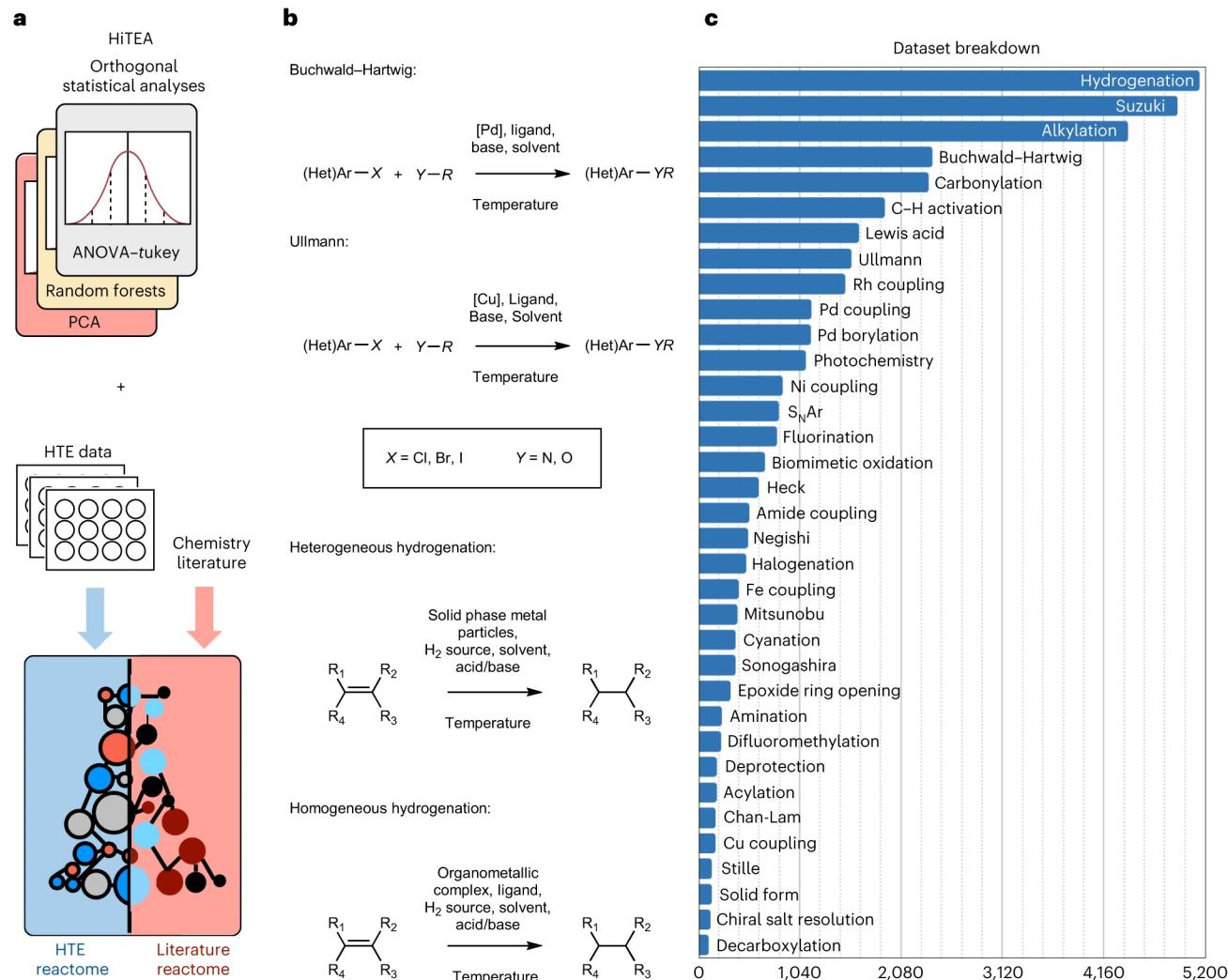
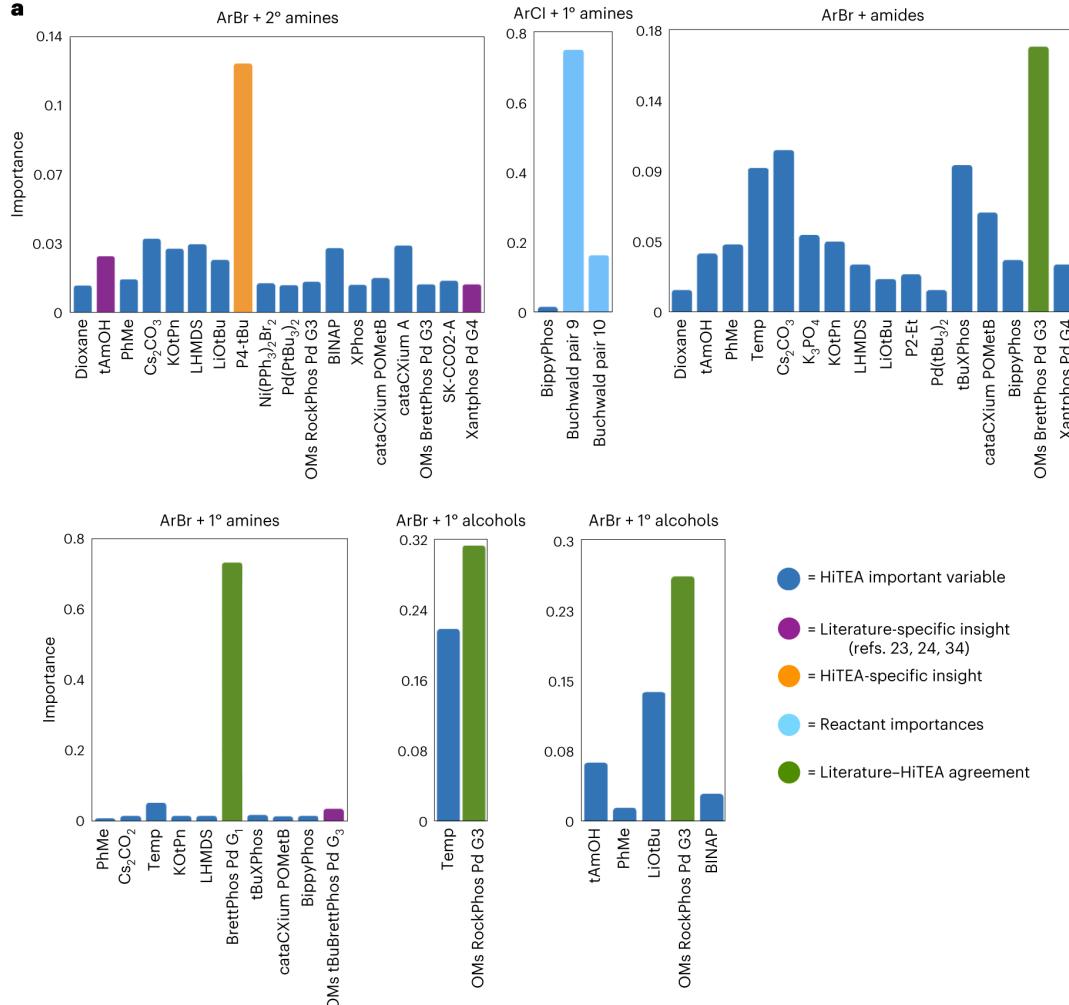


Figure from King-Smith and coworkers. *Nature Chemistry* 2024. DOI:

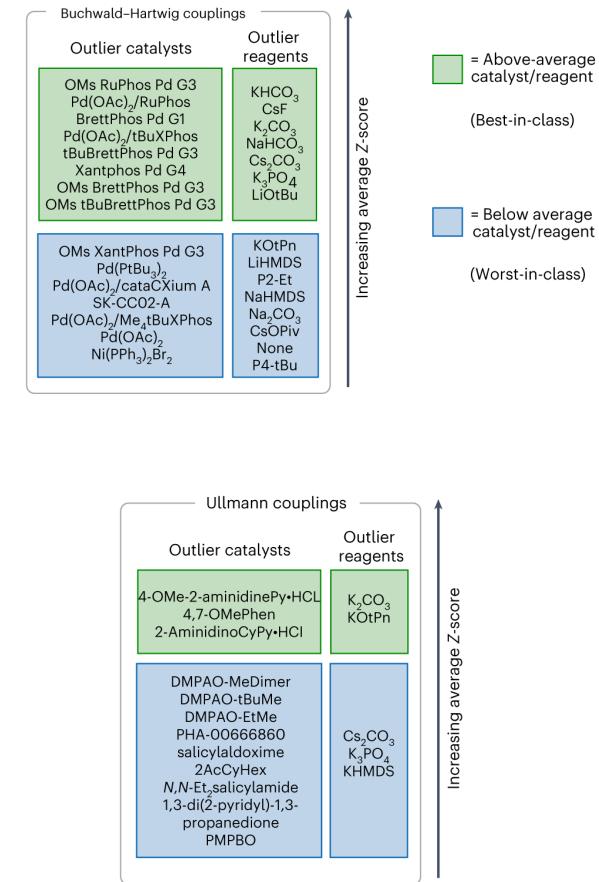


# Pooling all data allows us to ask big questions

**a**



**b**





# Standardized data practices are required

## Data Must Be:

- Centrally located and easily findable
- In an archival data format
- Interpretable by others
- Not require (substantial) conversion or additional manipulation



# FAIR data practices

## Findable:

- Meta data assigned a globally unique identifier
- Data described with rich metadata
- Data and metadata are registered and indexed in a searchable resource

## Accessible:

- Data can be retrieved by their identifier
- Retrieval protocol is accessible to users

## Interoperable:

- Data uses a formal, shared, and broadly applicable language for knowledge representation
- (Meta)data uses vocabularies that follow FAIR principles

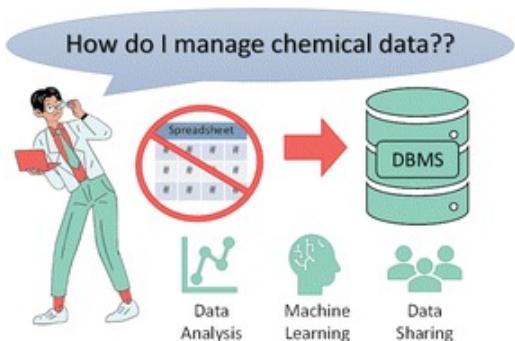
## Reusable:

- Data is richly described with many accurate and relevant attributes
- Data is associated with a detailed provenance
- Data meets domain relevant community standards



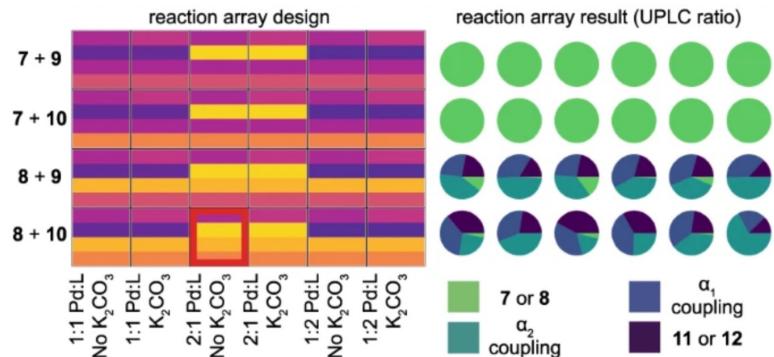
## What does this mean when applied to synthetic chemistry?

### Data Architectures to accelerate discovery



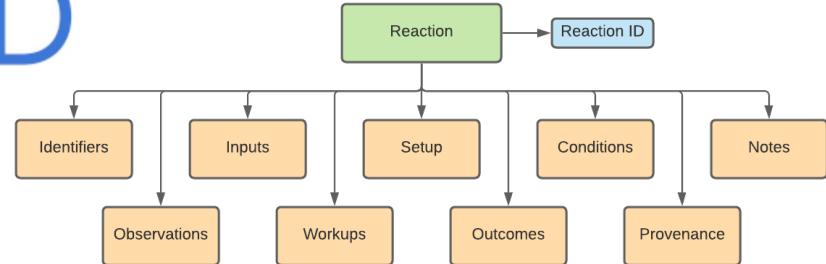
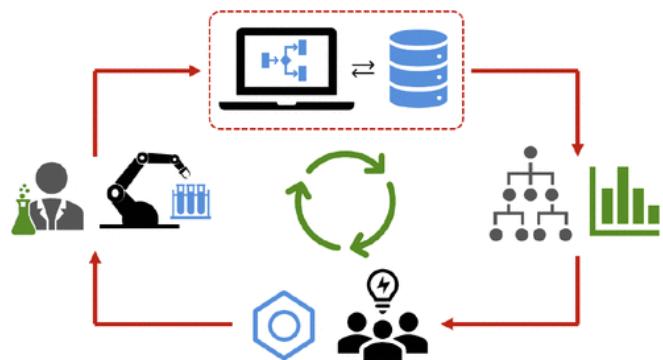
Chad Risko and coworkers. *Chem. Sci.* **2022**, 13, 13646-13656.

### Phactor: ELN for HTE data



Tim Cernak and coworkers. *Nature Commun.* **2023**, 14, 3924.

### Open Reaction Database



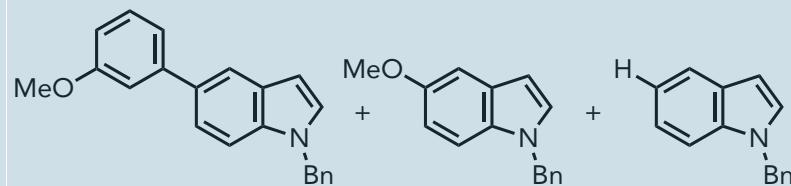
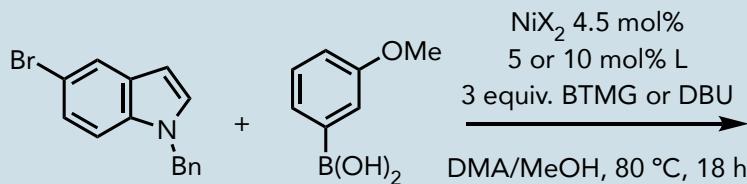
Steve Kearnes and Connor Coley and coworkers. *J. Am. Chem. Soc.* **2021**, 143, 18820-18826; *J. Chem. Inf. Model.* **2023**, 63, 4253-4265.



# Data generated in an HTE campaign

## Metadata:

- Experimenter (including group, notebook number, project, date and time)
- Experimental goals
- Analytical technique, how yield was determined, experimental procedure
- Vial size, whether it was run in the glovebox or not, light source, pressure, stir bar size, etc.



## Inputs:

- SMILES
- Reactant type (substrate, base, etc.)
- Loading ( $\mu\text{mol}$ ,  $\mu\text{L}$ , mg, etc.)
- Source
- Order of addition
- Electrochemical parameters

## Outputs:

- Original files
- SMILES
- Area count
- Analytical yield
- LC or GC area %
- Ratio of product peak to IS peak
- Selectivity measurement (ee, etc.)
- Other figures of merit (i.e.  $\lambda_{\max}$ )



## How do you assemble all that information for 24 or 96 reactions at a time?



## How do you enforce a schema using a spreadsheet?

- Extensive templating and streamlining using Excel's built in functions

**How do you ensure that the data can't be broken by Excel updates?**

  - Export and store a summary table as a CSV file

	11	12	13	14	15	16	17	18	19	20	21	22		
catalyst_2	henylphosp	DPPF	DCEPhos	DPPE	P(p-anis)3	Sphos	henylphosp	DPPF	DCEPhos	DPPE	P(p-anis)3	Sphos	base_1	row variable
catalyst_1														
Preplated ligand ID numbers below														
47	NiCl2(OH2)6												BTMG	55
48	NiBr2DME												BTMG	56
49	NiCl2(OH2)6												DBU	57
50	NiBr2DME												DBU	58
51	NiCl2(OH2)6												BTMG	59
52	NiBr2DME												BTMG	60
53	NiCl2(OH2)6												DBU	61
54	NiBr2DME												DBU	62
column variable														
	23	24	25	26	27	28	29	30	31	32	33	34		
column variable														
	35	36	37	38	39	40	41	42	43	44	45	46		

Dilution Factor	
Reaction volume ( $\mu\text{L}$ ):	100
Final reaction volume after dilut:	150
First aliquot volume ( $\mu\text{L}$ ):	10
First total analytical volume ( $\mu\text{L}$ )	500
Second aliquot volume ( $\mu\text{L}$ ):	0
Second total analytical volume (	0
Injection volume for calibration (	2
Injection volume for data collect	2
Conversion factor (yield to final	0.133

General Notes	
Research group:	ORD
Notebook number:	JRS-2023-158
Analysis method:	LC-MS
Stir rate:	500
Time (h):	18
Temperature (°C):	80
Atmosphere:	Glovebox
Pressure (psi):	
Light source:	ambient
Light intensity:	
Solvent: (blank if variable)	DMA
Internal standard:	TRUE
Reaction scale ( $\mu\text{mol}$ ):	10
Prepared ligand loading:	10
Reaction vol ( $\mu\text{L}$ ):	100
Ligand plating ( $\mu\text{mol}$ ):	1

Tracked Compounds					Correction Factors		
Number	Name	SMILES	Yield ( $\mu\text{mol}$ )	Yield (mg)	[I] (mg/mL)	Slope	R-squared
standard	4,4'-di-tert-butylbenzene	CC(C)(C)C	2.5	0.66605	0.09		
1	1-benzyl-5-phenyl-1,3-dihydro-1,3-dioxole-2,4-dione	COC1=CC=CBrC1=CC=O	10	3.13	0.42	1.598	#DIV/0!
2	4-phenyl-1-(4-phenylbutyl)-1,3-dihydro-1,3-dioxole-2,4-dione	COC1=CC=CBrC1=CC=O	10	2.8617	0.38	0.71	#DIV/0!
3	methanol acetyl ester	COC1=CC=O	10		0.00	#####	#DIV/0!
4	proteodehyd	[H]C1=CC=O	10		0.00	#####	#DIV/0!
5					0.00	#####	#DIV/0!
6					0.00	#####	#DIV/0!
7					0.00	#####	#DIV/0!
8					0.00	#####	#DIV/0!
9					0.00	#####	#DIV/0!
10					0.00	#####	#DIV/0!
11					0.00	#####	#DIV/0!
12					0.00	#####	#DIV/0!
13					0.00	#####	#DIV/0!
14					0.00	#####	#DIV/0!
15					0.00	#####	#DIV/0!



# Online templates

The screenshot shows a web browser window with the URL [sites.udel.edu](https://sites.udel.edu/). The page title is "HTE Resources". The main content features a 96-well plate visualization with rows labeled A-H and columns labeled 1-12. Each well contains a circular icon with a blue outer ring and an orange inner circle, representing data points. To the right of this visualization is a section titled "Plate Analysis" with the following text:

**Plate Analysis**

Data analysis of a 96 well HTE plate can become highly complex very quickly. By developing a standardized, user-friendly spreadsheet, we can readily visualize multi-dimensional HTE data in R or Python and long-term ensure that data is machine-readable.

[DOWNLOAD](#)

Rather than use an archived copy of the spreadsheet, please download it from here every time you need a copy



# Online templates

Screenshot of a GitHub repository page for "HTE-Templates".

The repository was last updated 3 weeks ago by jtheligand, with 20 commits. It contains files for 24-well and 96-well plates, along with a README file.

**Excel templates to manage multidimensional HTE campaigns and order data in a machine-readable format**

- Readme
- Activity
- 0 stars
- 1 watching
- 0 forks

Report repository

**Releases**

No releases published

**Packages**

No packages published

## HTE-Templates

This repository contains a series of Excel notebooks that are designed to help users manage multidimensional HTE screens and campaigns. They were designed and implemented by Jessica Sampson at the University of Delaware and are under active development. Files should be re-downloaded for each screen as the ligand directory in these files will be updated as the ligand library is updated.

As of 10/21/23 the 24-well versions are up-to-date.

Notebooks can be uploaded to the UD HTE ELN by going to: <https://forms.gle/8m4Kee8XAc6q15dC6>

Rather than use an archived copy of the spreadsheet, please download it from here every time you need a copy



# Template selection

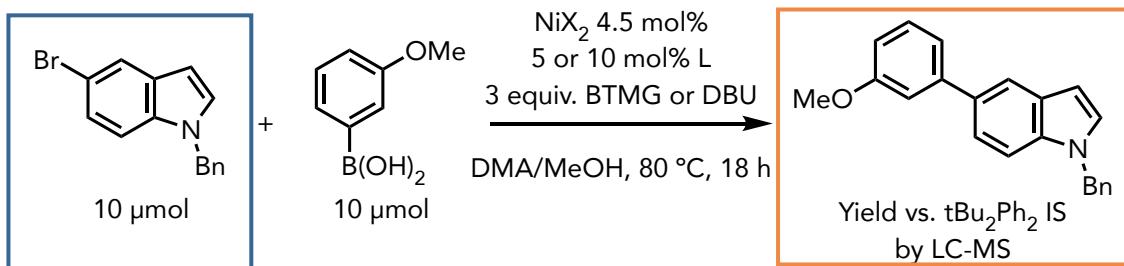
The screenshot shows a GitHub repository page for 'jtheligand/HTE-Templates'. The 'Code' tab is selected. On the left, there's a sidebar with file navigation. The main area shows a list of files in the '96-well-plates' folder. Several files are highlighted with different colors: 'ee\_noIS.xlsx' is blue, 'icap\_yield\_IS.xlsx' is orange, and 'icap\_yield\_IS\_notpreplated.xlsx' is red. Below the file list, there are three callout boxes with text:

- Quantifying ee and yield **OR** not using an internal standard when calculating yield
- Using an internal standard for yield and/or yield estimating by area %
- Using >12 of a non-preplated variable (additive, other ligand, substrate, etc.)

You do not need to know how to do the calculations on area %, etc., the spreadsheet will do them for you  
You only need to supply it with your correction factor and the area counts



# What does mean for metadata collection?



## Substrate 1

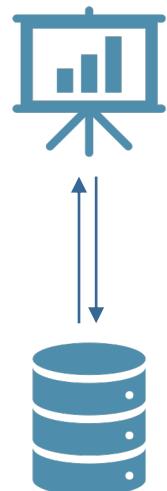
- 1-benzyl-5-bromo-1*H*-indole
- BrC1=CC2=C(N(CC3=CC=CC=C3)C=C2)C=C1
- Added in DMA
- Combi-Blocks
- Added third

## Product 1

- 1-benzyl-5-(3-methoxyphenyl)-1*H*-indole
- COC1=CC(C2=CC3=C(N(CC4=CC=CC=C4)C=C3)C=C2)=CC=C1
- Yield determined by LC-MS using a calibration curve against an IS



# Ok, but what about solving the original problem? Capturing all user data in a central archive



## HTViz: Initial build

- Continue to generate input/output summaries through Excel
  - Archival data written/read as CSV files
  - Metadata written to central index table
- Retrieve tags on *preplated* ligands from database
- Users can select from 5 different visualizations of their datasets



# HTViz: Uploading new data sets

Screenshot of the UD HTE Center - Data Visualization Archive interface on localhost.

The interface includes a navigation bar with icons for various applications and a main menu with options like Tabular data, Heatmap, Pie charts, Factor sorting, and Ligand filtering.

The main form for uploading a dataset is displayed:

- Visualize Dataset** (selected) and **Import New Dataset** buttons.
- Group**:
- Experiment number**:
- Analysis method**:
- Plate goals**:
- Reaction type**:
- Project**:
- Vessel size**:
- Limiting reagent**:
- Experimental procedure**: A large text area for notes.
- Filename**:
- Upload new data set** button (highlighted in blue).
- Status**: No data loaded!
- Clear form** button.



# HTViz: Uploading new data through Google Form

docs.google.com

HTE File Upload

Workbook will be checked and results uploaded within 1 week

jrsampso@udel.edu [Switch account](#)

The name, email, and photo associated with your Google account will be recorded when you upload files and submit this form

\* Indicates required question

Email \*

Record jrsampso@udel.edu as the email to be included with my response

Submitter UDID \*

Your answer

Group \*

Chain  
 DAW  
 Fox  
 Messina  
 MPW  
 Other:

!

edit



# HTViz: Retrieving old data

localhost

UD HTE Center- Data Visualization Archive

Visualize Dataset Import New Dataset

Group: ORD

User: JRSAMPSON

Project: All

Experiment number: JRS-2023-165, JRS-2023-171, JRS-2023-181, JRS-2023-190

Visualize data set

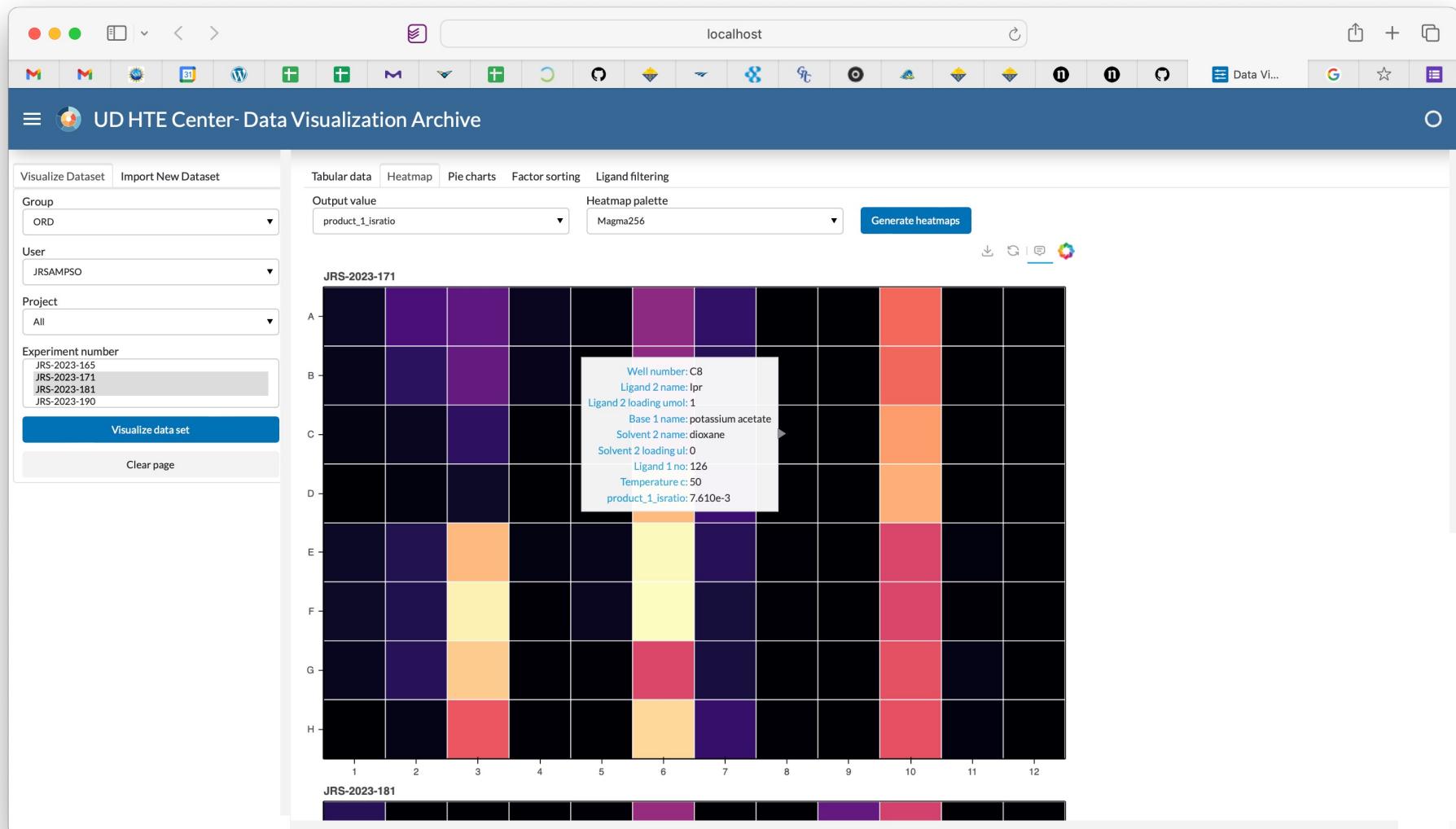
Clear page

Tabular data Heatmap Pie charts Factor sorting Ligand filtering

well_no	substrate_1_name	substrate_2_name	ligand_2_name	ligand_2_loading_umol	base_1_name	solvent_2_name	solvent_2_loading_uL	ligand_1_no	temperature_C	product_1_isratio	product_2
A1	Methyl acrylate	1-benzyl-5-bromindole	Triphenylphosphine	1.0	Potassium carbonate	dioxane	0	0	50	2.115144e-01	137
A2	Methyl acrylate	1-benzyl-5-bromindole	DPPF	0.5	Potassium carbonate	dioxane	0	0	50	7.274519e-01	132
A3	Methyl acrylate	1-benzyl-5-bromindole	DCEPhos	0.5	Potassium carbonate	dioxane	0	0	50	8.805675e-01	119
A4	Methyl acrylate	1-benzyl-5-bromindole	DPPE	0.5	Potassium carbonate	dioxane	0	0	50	1.803946e-01	132
A5	Methyl acrylate	1-benzyl-5-bromindole	P(p-anis)3	1.0	Potassium carbonate	dioxane	0	0	50	7.461613e-02	123
A6	Methyl acrylate	1-benzyl-5-bromindole	SPhos	1.0	Potassium carbonate	dioxane	0	0	50	1.234401e+00	91
A7	Methyl acrylate	1-benzyl-5-bromindole	RuPhos	1.0	Potassium carbonate	dioxane	0	24	50	5.590117e-01	111
A8	Methyl acrylate	1-benzyl-5-bromindole	Ipr	1.0	Potassium carbonate	dioxane	0	126	50	5.882005e-03	109
A9	Methyl acrylate	1-benzyl-5-bromindole	PCy3	1.0	Potassium carbonate	dioxane	0	4	50	1.050608e-07	119
A10	Methyl acrylate	1-benzyl-5-bromindole	DavePhos	1.0	Potassium carbonate	dioxane	0	28	50	2.106599e+00	79
A11	Methyl acrylate	1-benzyl-5-bromindole	tBuXPhos	1.0	Potassium carbonate	dioxane	0	13	50	7.603398e-03	123
A12	Methyl acrylate	1-benzyl-5-bromindole	DTBPF	1.0	Potassium carbonate	dioxane	0	461	50	4.296001e-03	120
B1	Methyl acrylate	1-benzyl-5-bromindole	Triphenylphosphine	1.0	Potassium phosphate	dioxane	0	0	50	1.529457e-01	132
B2	Methyl acrylate	1-benzyl-5-bromindole	DPPF	0.5	Potassium phosphate	dioxane	0	0	50	5.053776e-01	132
B3	Methyl acrylate	1-benzyl-5-bromindole	DCEPhos	0.5	Potassium phosphate	dioxane	0	0	50	9.329085e-01	118
B4	Methyl acrylate	1-benzyl-5-bromindole	DPPE	0.5	Potassium phosphate	dioxane	0	0	50	1.628024e-01	130
B5	Methyl acrylate	1-benzyl-5-bromindole	P(p-anis)3	1.0	Potassium phosphate	dioxane	0	0	50	4.326621e-02	127
R4	Methyl acrylate	1-benzyl-5-	SDPhos	1.0	Potassium	dioxane	0	0	50	1.230073e+00	04



# HTViz: Visualizing old data





# HTViz: Visualizing old data

localhost

UD HTE Center- Data Visualization Archive

Visualize Dataset Import New Dataset

Group: ORD

User: JRSAMPSO

Project: All

Experiment number: JRS-2023-165, JRS-2023-171, JRS-2023-181, JRS-2023-190

Tabular data Heatmap Pie charts Factor sorting Ligand filtering

Output notebook: JRS-2023-171

Output type: Yield

Chart type: Pie chart

Chart exterior values: product\_2\_yield | x

Palette: ['#3c0d03', '#8d1c06', '#e67424', '#ed9b49', '#f5c34d']

Chart interior values

Show interior:

Interior color:

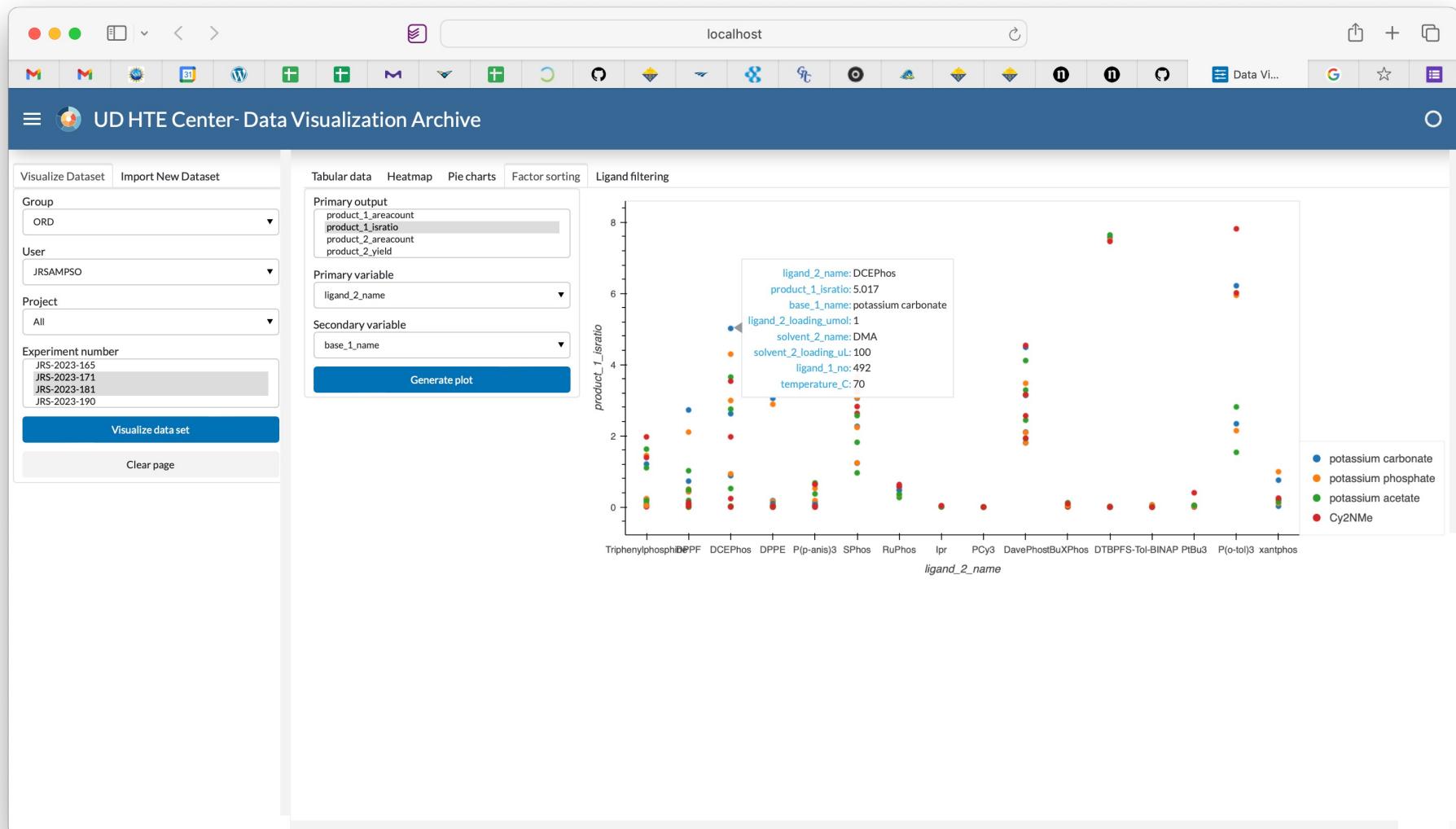
Generate chart

A B C D E F G H

1 2 3 4 5 6 7 8 9 10 11 12



# HTViz: Visualizing old data





# Path towards “born-digital” HTE core facility



## Immediate goals

- More expansive code for data validation and verification
- Some inventory management capabilities, particularly for ligand library
- Better tools for sorting and filtering preplated ligands

## Long-term goals

- Move away from Excel dependence for data organization
- Integration with robotics and analytical instrumentations



# Ok, but what about all that old data?



Please submit **all** of your old spreadsheets! (in whatever form or template!)

<https://forms.gle/8m4Kee8XAc6q15dC6>



# Ok, but what about all that old data?

A screenshot of a GitHub repository page for "HTE-Templates".

The repository was created by **jtheligand** and has 20 commits. The last commit was made 3 weeks ago.

The repository contains several files and folders:

- 24-well-plates**: added another tooltip 3 weeks ago
- 96-well-plates**: added another tooltip 3 weeks ago
- .DS\_Store**: added another tooltip 3 weeks ago
- .Rhistory**: added fields for source 5 months ago
- .gitattributes**: Initial commit 9 months ago
- 20230216\_HTE.pdf**: added pdf of sop 4 months ago
- README.md**: update readme 3 months ago

The **README** file is currently selected.

## HTE-Templates

This repository contains a series of Excel notebooks that are designed to help users manage multidimensional HTE screens and campaigns. They were designed and implemented by Jessica Sampson at the University of Delaware and are under active development. Files should be re-downloaded for each screen as the ligand directory in these files will be updated as the ligand library is updated.

As of 10/21/23 the 24-well versions are up-to-date.

Notebooks can be uploaded to the UD HTE ELN by going to: <https://forms.gle/8m4Kee8XAc6q15dC6>

Excel templates to manage multidimensional HTE campaigns and order data in a machine-readable format

- Readme
- Activity
- 0 stars
- 1 watching
- 0 forks

Report repository

### Releases

No releases published

### Packages

No packages published



# Conclusions





# Acknowledgements

Don Watson

Mary Watson

DAW Group

MPW Group

Jackson Burns

Jacob Letnauchyn

Pat McMahon

Justin Vitelli

Nessa Carson

Austin Kelly

Dr. Jiantao Fu

Dr. Dipannita Kalyani

## Funding:

UNIDEL: High Throughput Instrumentation

Core to Accelerate Discovery

NIGMS IDeA program

NIH COBRE P20GM104316



## Reproducibility Project:

Brandon Orzolek

Dr. Christopher Roberts

Jacob Letnauchyn

Guillermo Correa

Ben Deadman

Prof. Mary Watson

Prof. Don Watson

Prof. Marisa Koslowski

Prof. Mimi Hii

Prof. Abby Doyle

Prof. Connor Coley

Dr. Steve Kearnes

Dr. Spencer Dreher