

# Research Objects for Everyday Use

### Carl Kesselman

## Dean's Professor, Industrial and Systems Engineering University of Southern California



## What does it mean to have a scientific "result"



- Others have to "know" about it
- Others have to be able to validate it
  - Reproduce the method and achieve the same result
  - Achieve the same result via a different method
    Reuse the result in a new method
- "Non-reproducible single occurrences are of no significance to science."
  - Karl Popper, 1959. The logic of scientific discovery. Hutchinson, London, United Kingdom.



## Science must be reproducible..

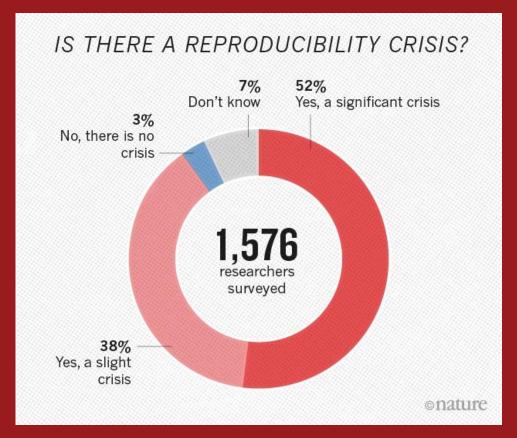
Nature May 25, 2016

 "More than 70% of researchers have tried and failed to reproduce another scientist's experiments, and more than half have failed to reproduce their own experiments. Those are some of the telling figures that emerged from *Nature*'s survey of 1,576 researchers who took a brief online questionnaire on reproducibility in research."



#### Where are we now?

#### Only 10% of published results are reproducible

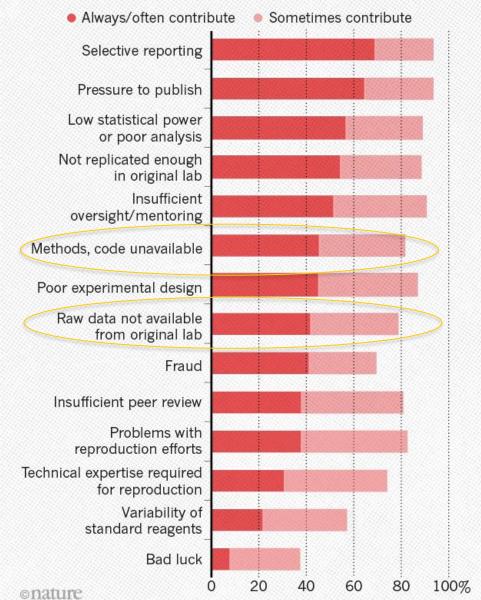






#### WHAT FACTORS CONTRIBUTE TO IRREPRODUCIBLE RESEARCH?

Many top-rated factors relate to intense competition and time pressure.





#### Information Sciences Institute

USC Viterbi School of Engineering

## Lets focus on the data...

**Reproducibility of data** 

- Data is only useful if we can figure out:
  - Where it is
  - What it is
  - How it was created
- Process and product: the data lifecycle
  - Initial data acquisition (perhaps from experiment)
  - Computing, and analysis
  - Publication



# **Reproducible Data is FAIR**



- Findable

   identified by a unique identifier, characterized by rich metadata
- Accessible
  - standard protocol with access control, metadata accessible even when the data is not,
- Interoperable
  - by standardized terms to describe it
- Reusable
  - Accurate and relevant attributes.



# **GOAL: FAIR Collaboration**



- Lets make all data produced in an investigation FAIR
  - Not just final "published" result, but all results
  - Need to scale application of FAIRness principals down to level of daily practice
- Requires accurate descriptions of data
  - -Characteristics of data element
  - Relationships between data elements
- Requires robust naming of the data products



# **Knowledge Turns and Publication**

## Publication

- Slow turn around, polished content and presentation
- -Human processes and audience
- µPublication
  - -Rapid turn around, incremental raw data exchange
  - -Human and machine processes and audience





# What does it mean to have research objects in daily use



Get users thinking about lots of publication events

- Prepare information for sharing
- Package it up so that its accessible
- Name the package so that it can be identified.

## What can we do to make these things easier?



# Identifiers are important, but too hard!!



Lets consider a range of identifier user cases

- We need to separate out issues of naming from persistence
- Digital Object Identifiers (DOIs) are designed for archival objects
  - You don't mind a DOI unless you really "mean it"
- What about identifying intermediate and temporary data
  - Can also benefit from unambiguous naming
- Two options:
  - Local identifiers, e.g. assession numbers (often not actionable)

Alternative identifier systems such as ARK



# Minimal Identifiers (Minids)

Lightweight identifiers that support simple creation and use

- Unique identifier (ARK)
  - E.g., /ark:/57799/b9040f
  - Or compact identifier (minid:b9040f)
- Minimal metadata (creator, date, name)
- Checksum ensures data is verifiable
- Service to provide the landing page
- Easy to use: CLI, Python SDK, R SDK, JSON-based REST API





# Why and When use Minids?

## Naming intermediate data

- Quickly associate a lightweight identifier
- Validate data integrity
- Lookup identifiers based on checksum
- CLI or Python/R client

😕 🗇 💿 🛛 kyle@ubuntu: ~/lymphoblast kyle@ubuntu:~/lymphoblast\$ ll total 47880 drwxrwxr-x 2 kyle kyle 4096 Oc drwxr-xr-x 44 kyle kyle 4096 Oc -rw-r--r-- 1 kyle kyle 49020928 Oc kyle@ubuntu:~/lymphoblast\$ kyle@ubuntu:~/lymphoblast\$ minid bd s "globus://galaxy#bdds/scratch/mad 2016-10-02 11:45:19,025 - INFO - Co @ 0x7f4f8bee5940> 2016-10-02 11:45:19,229 - INFO - Ch 54d33980b2aa already exists on the 2016-10-02 11:45:19,243 - INFO - St 2016-10-02 11:45:19,426 - INFO - Cr 2016-10-02 11:45:19,434 - INFO - Starting new HTTP connection (1): minid.bd2k.org 2016-10-02 11:45:21,662 - INFO - Created/updated minid: ark:/99999/fk45m6qp8t

#### minid

identifi a-625 e847f: checksums [{'func eb6b6 identifier ark:/9 landing_page https:/ location [https 2-e08 metadata {'Title' visible_to ['publi 2 11:44 bdds_tr s_trena_lymphobla uri/bdds_trena_ly puting checksum f	globus:auth:identity:3b843349-4d4d-4ef3-916d-2a465f9740a9', 'urn:globus:auth: ty:94f0c387-9528-4bed-b373-4ad840f32661', 'urn:globus:auth:identity:4846ded ee-4456-9c84-1647e53d71e1', 'urn:globus:groups:id:23acce4c-733f-11e8-a40d-0 194132', 'urn:globus:auth:identity:b5614711-228d-414f-8092-b518a25b072f'] ction': 'sha256', 'value': 'fc56ef4c8d1af6acf24e958e0e2f2796bc8f40184c6e42afc5 57a73d6c90']] 99999/fk4U4TyRAKafWMB //identifiers.globus.org/ark:/99999/fk4U4TyRAKafWMB/landingpage s://s3.amazonaws.com/portal-sc17-nick-globuscs-info/b03f15ca-ffcd-4638-903 817e8bcfa2.zip ]
eb6b6 identifier ark:/9 landing_page https:/ location [ https 2-e08 metadata ('Title' visible_to ['publi 2 11:44 bdds_tr i_trena_lymphobla puting checksum f	57a73d6c90'}] 9999/fk4U4TyRAKafWMB //identifiers.globus.org/ark:/99999/fk4U4TyRAKafWMB/landingpage s://s3.amazonaws.com/portal-sc17-nick-globuscs-info/b03f15ca-ffcd-4638-903
landing_page https:/ location [ https 2-e08 metadata {'Title' visible_to ['publi 2 11:44 bdds_tr s_trena_lymphobla uri/bdds_trena_ly puting checksum f	//identifiers.globus.org/ark:/99999/fk4U4TyRAKafWMB/landingpage s://s3.amazonaws.com/portal-sc17-nick-globuscs-info/b03f15ca-ffcd-4638-903
location [https 2-e08 metadata ('Title' visible_to ['publi 2 11:44 bdds_tr 5_trena_lymphobla uri/bdds_trena_ly puting checksum f	s://s3.amazonaws.com/portal-sc17-nick-globuscs-info/b03f15ca-ffcd-4638-903
2-e08 metadata ('Title' visible_to ['publi 2 11:44 bdds_tr 5_trena_lymphobla uri/bdds_trena_ly puting checksum f	
visible_to ['publi 2 11:44 bdds_tr 5_trena_lymphobla pri/bdds_trena_ly puting checksum f	
2 11:44 bdds_tr 2 11:44 bdds_tr _trena_lymphobla uri/bdds_trena_ly outing checksum f	: 'Downsample CRAM/CRAI ID Number: NWD176325, NA19238'}
2 11:44 bdds_tr _trena_lymphobla ri/bdds_trena_ly uting checksum f	[c']
ri/bdds_trena_ly puting checksum f	ena_lymphoblast_bag.zlp
king if the TEST	st_bag.ziptestregistertitle "Lympho Bag"loca mphoblast_bag.zip" for bdds_trena_lymphoblast_bag.zip using <sha256 hash="" obj<="" td=""></sha256>
erver: http://min	entity 3dea066ca2c1d38a57ec85d149b5d844128787b69ea873e10 hid.bd2k.org/minid hnnection (1): minid.bd2k.org
ating new identif	

kyle@ubuntu:~/lymphoblast\$



## Large Multi-File Datasets: Big Data Bags

- Profile on the BagIt specification
  - Payload: arbitrary content
  - Tags: metadata describing payload
  - Checksums: to verify content
- Content may be "missing"
  - Missing content must be listed in "fetch.txt"
  - Fetch entries list local name in data directory, and URL of where to fetch data
- Enhancements to support big data
  - Different data access protocols (Globus, HTTP, iRODS, S3)
  - Research Object metadata



tion Sciences Institute



```
Bio_data_bag/
|-- data
| \-- genomic
| \-- 2a673.fastq
| \-- 2a673.fastq
| -- manifest-md5.txt
| afbfa231324812378123bfa data/genomic/2a673.fasta
| -- bagit.txt
Contact-Name: John Smith
```



### **Research Objects: rich metadata for bags**

```
"@context": {
   "@vocab": "http://purl.org/dc/terms/",
   "dcmi": "http://purl.org/dc/dcmitype/Dataset"
 },
 "@id": "../../data/numbers.csv",
 "@type": "dcmi:Dataset",
 "title": "CSV files of beverage consumption",
"description": "A CSV file listing the number of cups consumed per person."
}
```



## Why and When use Minids? Large, multi-file datasets

- Integration with BDBag allows for content independent verification (i.e., holey bags)
- Streamline download and references to complex multi-file datasets
- Example: Encode2Bag

Assay         ChiP-seq of K562       Experiment status         Project       ChiP-seq of K562       Experiment status         Project indexado       Staturid       Project indexado         Corganism       Frage of X562       Experiment status         Project indexado       ChiP-seq of K562       Experiment status         Project indexado       Project index	Assay category DNA binding 561	Wew All Download Filter to 100 to resultare -	
ChiP-seq of K562         Exp           eCUP         178           genetic modification Disase-seq         40           -Sam mode         12           Project         ENCODE           ENCODE         561           Project Status         Encode           ENCODE         561           Project Status         Encode           ENCODE         200           Experiment status         Target Status           Inferseq of K562         Experiment status           Inferseq of K562         Experimat status           Inferse	in the second		
MRNA RIVA.seq         272 eCLIP         ChIP-seq of K562         Expense           Work Spins K54         Target 50:AF1         Bit Spins K542         Bit Spins K542           Interview Control         - See more         ChIP-seq of K562         Expense           Project         ENCODE         Spins K542         Bit Spins K542         Bit Spins K542           RFA         ENCODE         Spins K542         Bit Spins K542         Bit Spins K542           ENCODE         201         ChIP-seq of K562         Expense         Bit Spins K542         Bit Spins K542           ENCODE         202         ChIP-seq of K562         Expense         Expense         Bit Spins K542         Bit Spins K542           Interview Control         1         Bit Spins K542         Expense         Bit Spins K542         Bit Sp			
eCUP     178       sRNA FNA-seq     50       genetic modification DNase-seq     40       - Stemmer.     - Stemmer.       Project     FA       ENCODE     561       FFA     ENCODE       ENCODE     561       FFA     ENCODE       ENCODE     341       ENCODE     200       Experiment status     Fredesic Stanford       Index Status     Fredesic Stanford       Index Status     Fredesic Status       Index Status     ChIP-seq of KS62       Index Status     Experiment status       Index Status     Fredesic Status       Index St			
siRRA PNA-seq     50       peretic modification DRass-seq     50       Project     510       Project     510       Project     501       Project     502       Project     500       Project     500    <		ChIP-seq of K562	Expe
peretic modification DNase-seq         40	eCLIP 178	Homo sapiens K562	ENCSR
* Star more.     Project: ENCODE       Project     ChIP-seq of K562       Project: PSCODE     Experiment Status       Provide: Status     Project: Status       Project: Status     Project: Status	siRNA RNA-seq 50	Target: BCLAF1	n
• See more.         Project: ENCODE           Project:         Project: ENCODE         Experiment Status         Encode	genetic modification DNase-seg 40	Lab: Michael Snyder, Stanford	
ENCODE     961     Home sagiens XS82: Taget NOND     ENCODE       EXCODE2     240       EXCODE2     240       Experiment status     Project: ENCODE       Experiment status     ChIP-seq of K562       Experiment status     Experiment status       revolad     144       archived     24       Cenome assembly (visualization)     Taget SNR8       bq15     322       GRCh38     312       Target GrK52     Experiment SK52       Target GrK54     Experiment SK52       Target GrK562     Experiment SK52       GRCh38     312       Target GrK562     Experiment SK52       Target GrK562     Expe		Project: ENCODE	
ENCODE     961     Home sagiens XS82: Taget NOND     ENCODE       EXCODE2     240       EXCODE2     240       Experiment status     Project: ENCODE       Experiment status     ChIP-seq of K562       Experiment status     Experiment status       revolad     144       archived     24       Cenome assembly (visualization)     Taget SNR8       bq15     322       GRCh38     312       Target GrK52     Experiment SK52       Target GrK54     Experiment SK52       Target GrK562     Experiment SK52       GRCh38     312       Target GrK562     Experiment SK52       Target GrK562     Expe	Project	ChiP-sec of K552	Exce
FFA     Target NONO     Life Status       ENCODE2     220       Experiment status     Target NONO       Infeased     545       revolved     14       archived     14       bits     Status       Genome assembly (visualization)     ChIP-seq of K562       Corganism     Formo sapiens K562       Target of assay     Excore       transortefin fador     120       rescombinant protein     72       PNA binding protein     570       rescombinant protein     72       Biosample type     Immotalized cell line       Immotalized cell line     560       Biosample type     17       Immotalized cell line     560       Biosample type     17       Immotalized cell line     560       Biosample type     17       Immotalized cell and     562       Statig     Experimentalintemodeler       tin			
RFA     Lab: Mchaal Styder, Stanford       Project: ENCODE     240       Experiment status     Project: ENCODE       Infeased     545       Invelad     14       archived     14       archived     14       archived     14       archived     14       archived     14       archived     2       Genome assembly (visualization)     Project: ENCODE       Page: of assay     Experiment Status       Target Control     120       Homo sapiens K562     Experiment Status       Target Control     120       Interscription factor     411       control     120       rarget Control     121       rarget Control     120       rarget Control </td <td>501</td> <td></td> <td></td>	501		
BXCODE3     341     Project: BXCODE       Experiment status     Project: BXCODE       Interaction     545	DEA		1
ENCODE2     220       Experiment status     Fine-seq of K662       Interaction     545       Interaction     546       Interaction     2       Genome assembly (visualization) hp19     222       GRCh38     3122       Organism     Froject: ENCODE       Homo sapiens     560       Organism     Froject: ENCODE       Homo sapiens     560       Target Control     120       Interaction     120			
Experiment status     ChIP-acq of K662     Experiment status       Mono sapiens K662     Target SN38       revolvid     14       archived     14       archived     14       archived     14       htmos sapiens K662     Experiment status       Genome assembly (visualization)     Project BNCODE       Organism     ChIP-acq of K662       Homo sapiens K662     Experiment status       Organism     ChIP-acq of K662       Target of assay     Experiment status       transpione     140       Interview     140       advit     120       reserview     120   <		Project CNUVUE	
Experiment status     Percent     Per	220		
Experiment status     Provestagions K562     Experiment Status       Interset     Status     Project: ENCODE       Genome assembly (visualization)     Project: ENCODE       hybrid     322       Genome assembly (visualization)     ChIP-seq of K562       Organism     Project: ENCODE       Momo sagions K562     ENCER       Organism     Froject: ENCODE       Momo sagions K562     ENCER       Target Control     Lab: Xiang-Dong Fu, UCSD       Project: ENCODE     Encer       Target Control     Target Control       Target of assay     ChIP-seq of K562       Tranget Control     Target Control       Target Control     Target Control       Lab: Xiang-Dong Fu, UCSD     Exper       Momo sapiens K562     Exper       Homos aspiens K562     Exper	Even advected at a tank of	ChIP-seq of K562	Expe
Invested     395       Invested     14       archived     14       archived     2       Genome assembly (visualization)     Project ENCODE       bi19     322       GRCh38     312       Organism     Project ENCODE       Homo sapiens     560       Project ENCODE     Expe       Vision Conspilence     Momo sapiens       Manacoption factor     411       ChIP-seq of K562     Expe       Momo sapiens     560       Project ENCODE     Project ENCODE       Target Control     120       Target Control     120       Instructure     120       Project ENCODE     Expe       Nomo sapiens K642     ENCORE       Target Control     120       Life stage     140 <td></td> <td></td> <td>ENCSR</td>			ENCSR
Introduct     14       archived     2       Genome assembly (visualization) by 19     322       GROB     312       Organism     Frome sagients       Forme assembly (visualization) by 19     560       ChIP-seq of K552     Excel Biosample treatment interferon alpha       Biosample treatment interferon alpha     560       Eliesample treatment interferon alpha     560       ChIP-seq of K552     Excel Biosample treatment interferon alpha       Biosample treatment interferon alpha     560       ChIP-seq of K552     Excel Biosample treatment interferon alpha       fatel fatel fatel fatel fatel fatel fatel fatel fatel bam     560       ChIP-seq of K552     Excel Biosample treatment interferon alpha       fatel fatel fatel fatel bam     560       ChIP-seq of K552     Excel Biosample treatment interferon alpha       fatel fatel fatel bam     561       Formo sagiens K552     Excel Biosample treatment interferon alpha       fatel fatel bam     565       bay     577       big/big bad annon/Peak     577       big/big bad annon/Peak     577       big/big annon/Peak     577       big/big bay     577       big/big bay     577       big/big bay     577       big/big bay     577       big/big			
archived     2     Project: ENCODE       Genome assembly (visualization) hp19     ChIP-seq of K562     Expe Homo sapiens       GRCN38     312     Target: Control Lab: Xiang-Dong Fu, UCSD     Expe Homo sapiens       Target of assay Homo sapiens     ChIP-seq of K562     Expe Homo sapiens       Target of assay Homo sapiens     ChIP-seq of K562     Expe Homo sapiens       Target: Control econtributing protein     120 Target of assay     Expe Homo sapiens K562     Expe Homo sapiens       Biosample type Immodalized cell line     560     ChIP-seq of K562     Expe Homo sapiens K562       Biosample type Immodalized cell line     560     ChIP-seq of K562     Expe Homo sapiens K562       Biosample treatment Interfecon gamma     121 Lab: Xiang-Dong Fu, UCSD     Target: Control Lab: Xiang-Dong Fu, UCSD     Expe Homo sapiens K562       Biosample treatment Interfecon gamma     121 Lab: Mahale Snyber, Stanford Project: ENCODE     Expe Homo sapiens K562     Expe Homo sapiens K562       Available data fasiq     557 bam     ChIP-seq of K562     Expe Homo sapiens K562     Expe Homo sapiens K562       Biosample treatment Interfecon gamma     121 Lab: Mahale Snyber, Stanford Project: ENCODE     Expe Homo sapiens K562     Expe Homo sapiens K562       Biosample treatment Interfecon gamma     121 Lab: Mahale Snyber, Stanford Project: ENCODE     Target MiT	revoked 14		
hg19     322     ChIP-seq of K562     Expe       Organism     From sapiens K562     ENCSD       Organism     From sapiens K562     ENCSD       Target control     120     Expe       Target of assay     ChIP-seq of K562     Expe       Target control     120     Expe       Target of assay     ChIP-seq of K562     Expe       Target control     120     Expe       recombinant protein     72     Project ENCODE     Expe       Veronatin remodeller     45     Expe     Expe       *See mark     ChIP-seq of K562     Expe       *See mark     Folgect ENCODE     Interferon graphics       Biosample type     Target Control     Interferon graphics       Interferon graphics     14     Target Control       Interferon graphics     14     Expe       <	archived 2		
GRCh38     312     Homo sapiens K622     ENCRE       Organism     Momo sapiens     560     Project. ENCODE     Momo sapiens       Momo sapiens     560     ChIP-seq of K562     Expe       Momo sapiens K62     Encompositive     Encompositive     Encompositive       Target: Control     Target: Control     Encompositive     Encompositive       Momo sapiens     120     Target: Control     Encompositive       Momo sapiens K62     Encompositive     Encompositive     Encompositive       Momo sapiens     560     ChIP-seq of K562     Expe       Momo sapiens     562     Expe       Momo sapiens     562     Expe       Momo sapiens     562     Expe       Momo sapiens     562     Expe       Momo sapiens     562 <t< td=""><td>Genome assembly (visualization)</td><td>and the second se</td><td></td></t<>	Genome assembly (visualization)	and the second se	
Organism     Target Control     Lab: Xiany Dong Fu, UCSD       Project: ENCODE     Project: ENCODE       Target of assay     ChIP-seq of K562     Experimental difference       Target control     Target Control     Difference       Immotalized cell line     560     Experiment       Interfecton alpha     14     Target Control       Interfecton alpha     14     Difference       Interfecton alpha     14	hg19 322		
Organism     Lab: Xiang-Coop Fu, UCSD       /form sapiens     560       Target of assay     ChIP-seq of K562       Expendition factor     110       recombinant protein     72       PNAb Einding protein     72       PNAb Einding protein     72       PNAb Einding protein     72       Einsample type     Fraget Control       Immontalized onl line     560       Eliosample type     Fraget Control       Interfect alge     Encortel       adut     560       Eliosample traatment     560       Interfect alge     Encortel       adut     560       ChIP-seq of K562     Expe       Homo sapiens K562     Expe       Project: ENCODE     model       Kabiti     560       Eliosample traatment     Target Control       Interfect alge     ENCORE       Available data     Expe       fastig     545       bain     550       Endition sapiens K562     Expe       Lab: Mahael Snyder, Stanford     Expe       Project: ENCODE     Expe       Available data     Ethilds       fastig     545       bain     ChIP-seq of K562       Expe     Expe       Lab: Mahael	GRCh38 312		
Momo sapiens     560     Project: ENCODE       Target of assay     ChiP-seq of K562     Experimentation of the sapiens K562       recombinant protein     720     Target: Control     Experimentation       recombinant protein     720     Project: ENCODE     Experimentation       memortalized cell line     560     ChiP-seq of K562     Experimentation       Interference applies     141     ENCSR     Project: ENCODE       Available data     557     ChiP-seq of K562     Experimentation       faring     557     Lab: Mahael Snyder, Stanford     Project: ENCODE       by/Wig     429     Abot Mahael Snyder, Stanford     Project: ENCODE <td></td> <td></td> <td>re</td>			re
Target of assay     ChIP-seq of K562     Exception       Target of assay     ChIP-seq of K562     Exception       meconbinant protein     Target: Control     meconbinant       recombinant protein     57     Project: ENCODE       Biosample type     Immostalized cell line     560       Interferon applers     560     ChIP-seq of K562       Biosample traitment     560     Expetion       Interferon applers     560     ChIP-seq of K562       Available data     560     ChIP-seq of K562       Available data     560     ChIP-seq of K562       Available data     560     ChIP-seq of K562       Biosample treatment     Target: Control     Expetion       Interferon applers     562     Expetion       Available data     565     ChIP-seq of K562     Expetion       Istering     565     ChIP-seq of K562     Expetion       Available data     566     ChIP-seq of K562     Expetion       Istering     563     ChIP-seq of K562     Expetion       Available data     564     ChIP-seq of K562     Expetion       Istering     563     ChIP-seq of K562     Expetion       Istering     564     Expetion     Expetion       Istering     565     ChIP-seq of K562	Organism		
transcription factor     411     Home sapiens K562     ENCSPC       control     Target: Control     Target: Control     Transcription factor       recombinant protein     72     Lab: Xiang Doop Fu, UCSD     Transcription factor       recombinant protein     77     Project: ENCODE     Exception       e-See more     ChIP-seq of K562     Exception       Biosample type     Target: Control     Transcription factor       Immontalized cell line     560     ChIP-seq of K562     Exception       Biosample trage     Freight: Stating Doop Fu, UCSD     Traget: Control     Traget: Control       Immontalized cell line     560     ChIP-seq of K562     Exception       Biosample treatment     Target: Control     Traget: Control     Traget: Control       Interferon ajpha     14     Target: Control     Traget: Control       Interferon ajpha     14     Target: Control     Traget: Control       Interferon ajpha     14     Target: Control     Traget: Control       Interferon ajpha     14     Expetient Stote     Expetient Stote       Available data     ChIP-seq of K562     Expetient Stote       Biosample treatment     Target: Control     Traget: Control       Interferon ajpha     14     Lab: Motale: Stote     Expetient Stote       Avail		Project: ENCODE	
transcription factor     411     Home sapiens K592     ENCSP       recombinant protein     72     Lab: Xiang Dong Fu, UCSD     m       recombinant protein     57     Project: ENCODE     m       else more     ChIP-seq of K562     ENcsp       Biosample type     Target: Control     m       Immontalized cell line     560     ChIP-seq of K562     ENcsp       Biosample type     Target: Control     ENCSP       Biosample traatment     Froject: ENCODE     m       Interferen ajpha     14     ENCSP       Interferen ajpha     14     Encepter Norde       fastq     557     Home sapiens K552     Excpe       bay     257     Target: Control     Encepter Norde       fastq     557     Home sapiens K552     Excpe       bay     459     ChIP-seq of K562     Excpe       bay     577     Home sapiens K552     Encepter Norde       bay     577     Home sapiens K552     Encepter Norde       bay     459     ChIP-seq of K562     Excpe       bay     577     Home sapiens K552     Encepter Norde       bay     577     Home sapiens K552     Encepter Norde       bay     455     ChIP-seq of K562     Encepter Norde       ba	Target of assay	ChIP-seg of K562	Expe
control     120     Traget: Control     Traget: Control       modifying troth     72     Lab: Xiang-Dong Fu, UCSD     Traget: Control       PNA binding protein     77     Lab: Xiang-Dong Fu, UCSD     Expe       Biosample type     Traget: Control     Traget: Control     Traget: Control       Immontatized cell line     560     ChIP-seq of K562     Expe       Addit     560     ChIP-seq of K562     Expe       Available data     ChIP-seq of K562     Expe       Interferon agma     12     Traget: Control     Traget: Control       Available data     ChIP-seq of K562     Expe       Israg     545     ChIP-seq of K562     Expe       Available data     ChIP-seq of K562     Expe       Israget     ChIP-seq of K562     Expe       Abardit     545     ChIP-seq of K562     Expe       Available data     ChIP-seq of K562     Expe       Israget: Advall     Traget: Control     Traget: Control       Israget     245     ChIP-seq of K562			
recombiant protein 72 RNA binding protein 57 RNA binding protein 57 Project ENCODE Biosample type Biosample type adult 560 Biosample treatment 560 Biosample treatment 560 Biosample treatment 7652 Available data fariq 565 bam 557 Homo sapiens K562 ChIP-seq of K562 Target: Control Target: Control ChIP-seq of K562 Target: Control Target: Control ChIP-seq of K562 ChIP-seq of K56			ENCORD
PNA binding protein devonatin remodeller     57 45emmet.     Project: ENCOGE       issample type immortalized cell line     560 560 560 560 560 560 560 560 560 560			
dromatin remodeller     45       - Seemare     ChIP-seq of K562       Biosample type     ENCSR       Immotalized cell line     560       Life stage     Excell       advit     560       Biosample treatment     Interferon alpha       interferon alpha     14       interferon alpha     14       interferon alpha     14       fastq     545       bam     537       by/ing     429       bigdot narouPeek     397			
* See more.     ChIP-seq of K562     Expe       Biosample type     Target: Control     ENCSRI       Immotalized cell line     560     Project: ENCODE       Life stage     Encode     Encode       adut     560     ChIP-seq of K562     Expe       Biosample treatment     Interferon apha     14     Encode       interferon apha     12     Project: ENCODE     Encode       Available data     ChIP-seq of K562     Expe       big/ing     429     Target of K562     Expe       big/ing     429     Target, Stanford     Expe       big/ong     429     Target, Stanford     Project: ENCODE       big/ong     429     Target, Stanford     Project: ENCODE			0
Biosample type Immotalized cell line     Home sagierns K562     ENCSR4       Life stage aduit     560     ChIP-seq of K562     Expendence       Biosample treatment     Target: Control     Target: Control     Target: Control       Interferon gamma     12     Project: ENCODE     Expendence       Available data     ChIP-seq of K562     Expendence       Interferon gamma     545     Home sagierns K562     Expendence       Bolosample treatment     ChIP-seq of K562     Expendence       Available data     ChIP-seq of K562     Expendence       Istigan     545     ChIP-seq of K562     Expendence       Bolosample treatment     Lab: Michael Snyder, Stanford     Project: ENCODE       Available data     Expendence     Encode       Istigan moviPeak     397     Target: Control     Target: Control		A Second s	11241/074
Biosample type     TargeE control     TargeE control       immodalized cell line     560     Expe       adult     560     ChIP-seq of K562     Expe       Biosample treatment     TargeE control     ENCSR       interfecon alpha     14     TargeE control     ENCSR       interfecon alpha     12     Project: ENCODE     Expe       Available data     ChIP-seq of K562     Expe       fatag     545     ChIP-seq of K562     Expe       barn     537     Homo saplers K562     ENCSR       bijfden anrowPeak     397     Lab: Michael Sinyder, Stanford     ENCSR	+ See more		
Immontalized cell line         560         Lab: Xiang-Dong Fu, UCSD Project: ENCODE           Life stage adult         560         Expe More stapients K562         Expe More stapients K562 stably expressing fusion protein         ENCSP Biosample treatment           Biosample treatment interferon gamma         14         Target: Control Lab: Michael Snyder: Stanford         Expe Project: ENCODE           Available data         ChilP-seq of K562         Expe More stapients K562         Expe ENCSPE           barn         537         Homo stapients K562         Expe Project: ENCODE           bighd narrowPeak         397         Project: ENCODE		Homo sapiens K562	ENCSR4
Life stage         Project: ENCODE           aduit         569           Biosample treatment interferon aipha interferon agrims         12           Available data faste barn         537           big/Mig         439           big/Mig         397           big/big narrowPeak         397			14
Life stage     660     ChIP-seq of K562     Expe       Biosample treatment     Target: Centrol     ENCSR       Interferon gamma     12     Project: ENCODE     Target: Centrol       Available data     ChIP-seq of K562     Expe       Bam     537     Homo sapiens K562     Expe       Biogen and S37     Homo sapiens K562     Expe       bigMing     239     Target: MNT     m       bid narros/Peak     397     Project: ENCODE     m	immortalized cell line 560		
adult         560         ChIP-seq of K562         Expe           Biosample treatment Interferon alpha         14         Target: Control         Biosample treatment           Interferon alpha         14         Target: Control         Biosample treatment         Biosample treatment           Interferon alpha         14         Target: Control         Biosample treatment         Biosample treatment           Available data         F         ChIP-seq of K562         Exception           barm         537         Homo sapiens K562         ENCSR           big/Mig         439         Target: MNT         Biosample treatment           big/Bot anrowPeak         397         Project: ENCODE         Biosample treatment		Project ENCODE	
Biosample treatment         ChilP-seq of K562         Expe           Interfeton ajcha         14         Lab: Mchail Snyder, Stanford         ENCSR           Available data         ChilP-seq of K562         Expe           Barn         537         Homo sapiens K562         ENCSR           big/Mg         429         Target: ENCODE         ENCSR           big/Mg         439         Target: MNT         Enclose           big/do narrow/Peak         397         Project: ENCODE         Interfetor			
Interferon alpha     14     Target Control     Interferon       Interferon gamma     12     Lab: Michail Snyder, Stanford     Interferon       Available data     ENCRSC     Expect       fastq     545     ChIP-seq of K562     Expect       barm     537     Homo sapiens K562     ENCRSC       big/Mg     429     Target: NNT     Interferon       big/denarion/Peak     397     Lab: Michail Snyder, Stanford	aduat 560		Expe
Interferon alpha     14     Target: Control     Interferon       Interferon gamma     12     Lab: Michael Snyder: Stanford     Interferon       Available data     Project: ENCODE     Expe       fastq     545     ChIP-seq of KS62     Expe       barn     537     Homo saplens KS62     ENCSR3       big/fig     429     Target: MNT     Interferon       big/bed narrox/Peak     397     Project: ENCODE	Biosample treatment	Homo sapiens K562 stably expressing fusion protein	ENCSR
Interferon gamma     12     Lab: Mohael Snyder. Stanford       Available data     Fariq     545     Expe       fariq     545     ChIP-seq of K562     Expe       bam     637     Homo saplens K562     ENCSR:       big/Mg     429     Target. MNT     m       big/Bot narowPeak     397     Lab: Mohael Snyder. Stanford     m			n
Available data     Project: ENCODE       fastq     545       bam     537       Homo saplens K562     ENCSR3       bed narrowPeak     397       bigBed narrowPeak     397			
fastq     545     ChIP-seq of K562     Expe       bam     537     Homo saplens K562     ENCSR:       big/Mg     429     Target: NNT     m       big/Mg     429     Target: NNT     m       big/Mg     439     Lab: Michael Snyder, Stanford     m       bigBed narrowPeak     397     Project: ENCODE     m	TZ	Project: ENCODE	
barn         537         Homo saplens K562         ENCSR0           big/Mg         429         Target: MNT         rr           bed narrowPeak         397         Lab: Mchaal Snyder, Stanford         rr           bigBod narrowPeak         397         Project: ENCODE         rr			
big/Mg         429         Target MNT         r           bed narrow/Paak         397         Labs Michael Snyder, Stanford         regited narrow/Paak         397           bigBed narrow/Paak         397         Project: ENCODE         regited narrow/Paak         107	fastq 545	ChIP-seq of K562	Expe
bigWig         429         Tanget: MNT         r           bed namov/Peak         397         Lab: Michael Snyder, Stanford         r           bigBed namov/Peak         397         Project: ENCCODE         r			
bed narrowPeak 397 Lab: Michael Snyder, Stanford bigBed narrowPeak 397 Project: ENCODE			EPROPO N
bigBed narrowPeak 397 Project ENCODE			
crypers narrow was			

Create a BDBag from an ENCODE search.	
For example enter the following search:	
https://www.encodeproject.org/search/?type=Experiment&assay_title=RNA-seq&replicates.library.biosample.biosample_type=stem+cell	
Or paste in an Encode metadata file.	
Encode Search Query Encode Metadata File	
Enter your search for encode data	
Create BDBag	

305



### **Scientific Digital Asset Management**



Discovery Environment for Relational Information and Versioned Assets (DERIVA)

- Discovery as process of creating and updating contextualized digital assets.
- Adaptive and extensible



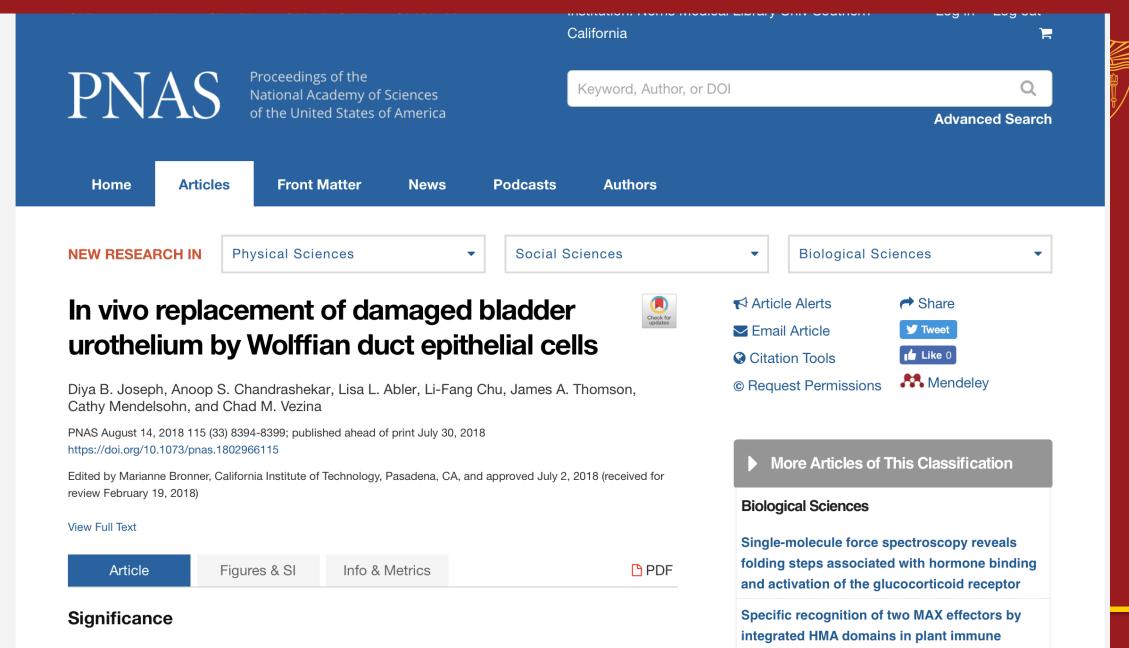
### Scientific Asset Management System



## Discovery Environment for Relational Information and Versioned Assets (DERIVA)

- DERIVA promotes FAIR data production by:
  - F: providing rich metadata using an Entity-Relationship model to express relationships between diverse data elements;
  - A: offering rich access control and access to metadata via standard HTTP web service interfaces;
  - I: integrating with standardized terms defined by collaborators, consortium or communities; and
  - R: supporting dynamic model evolution so that the data presented accurately represents the current structure and state of knowledge within an investigation.





When the bladder's specialized epithelial lining is damaged by infection or injury, its own basal and intermediate cell progenitors are called upon to restore a functional barrier. Here we show that when these progenitor cells are depleted in conditional *Dnmt1* mutant mice,

#### tute

receptors involves distinct binding surfaces

to flee engulfment by macrophages

Phospholipid flippases enable precursor B cells

#### Methods

#### Data Dissemination.

To increase rigor, reproducibility, and transparency, raw image files and other data generated as part of this study were deposited into the GUDMAP consortium database and are fully accessible at: https://doi.org/10.25548/W-QXXC (**25**).

#### Conditional Dnmt1 Mutants.

Mice were housed as previously described (26). All procedures performed on mice were approved by the University of Wisconsin–Madison Animal Care and Use Committee and were carried out in accordance with the Guide for the Care and Use of Laboratory Animals. *Shh*<sup>cre</sup> alleles (B6.Cg*Shh*<sup>tm1(EGFP/cre)Cjt/J</sup>) (11) were used to conditionally inactivate *Dnmt1* using *Dnmt1flox* alleles (B6.129S4-*Dnmt1*<sup>tm2Jae/Mmucd</sup>) in *Shh* lineage cells marked



#### Information Sciences Institute

**O** Referen

Figures & SI

PDF

Info & Metrics

### In vivo replacement of damaged bladder urothelium by Wolffian duct epithelial cells COLLECTION

🖌 Show All Related Records 🛛 💆 Export 🗸 🕑 Share

				_		
RID	W-QXXC					Contents
Title	In vivo replac	ement of damage	ged bladder urothelium by Wo	olffiar	n duct epithelial cells	n duct epithelial cells Main
Description	Figures and c Joseph et al.		e PNAS 2018 paper titled "I	n viva	o replacement of damaged bladder urothelium by Wolffian duct epithelial cells" by	o replacement of damaged bladder urothelium by Wolffian duct epithelial cells" by <b>He Slide Co</b> Specimen
Details	The following	table shows the	mapping of figures and ima	ge r	record IDs (RID) presented in the paper.	
	Figure	Reference	Additional Images			
	1A	W-QXXW	W-QXZ4, W-QY2C			
	1B	W-QY34	W-QY38, W-QY3C			
	1C	W-QY66	W-QY6T, W-QY86			
	1D	W-QY8Y	W-QY9A, W-QYA6			$\leftrightarrow$
	1E	W-QYB6	W-QYBP, W-QZ6T			
	1F	W-QYC2	W-QYCE, W-QYCT			
	1G	W-QYDP	W-QYDY, W-QYEA			
	1H	W-QYEP	W-QYF2, W-QYFE			
	11	W-QYGP	W-QYH2, W-QYHE			
	1J	W-QYHT	W-QYJ6, W-QYJJ			
	1K	W-QYKP	W-QYM2, W-QYME			

## In vivo replacement of damaged bladder urothelium by Wolffian duct epithelial cells COLLECTION

🖌 Show All Related Records 🛛 💆 Export 👻 🖒 Share

										-	Contents
Require	DOI?	Yes	Yes								
Persiste	ent ID	https:	//doi.org/10	0.25548/W-QXXC							Main
Principa	al Investiga	ator Chad	Vezina				He Slide Collection (13)				
Data Pro	ovider	Unive	rsity of Wisc	onsin							Specimen Collection (25)
Consort	tium	GUDN	<i>I</i> AP								ł
Creation	n Time	2018-	-05-17 21:19	9:14						_	
Last Mo	odified Tim	<b>1e</b> 2018-	-07-09 22:13	3:01						_	
	Slide Colley	ction (showing	all 13 result	te)					\//	-	
		, J							View More		
View	RID 🛓	Thumbnail I1	Name 1		Species ‡1	Tissue 11	Age Stage ↓↑	Gender 11	Image File ↓↑		
	W-R01Y		•	,	Mus musculus	urogenital sinus	18.5 embryonic days	Male	20160826ShhcreDnmtiLOFME18.5U	•	→
	W-R02A		-	,	Mus musculus	urogenital sinus	18.5 embryonic days	Male	20160826ShhcreDnmtiLOFME18.5U		
	W-R02P			,	Mus musculus	urogenital sinus	18.5 embryonic days	Male	20160826ShhcreDnmtiLOFME18.5U		
	W-R02Y		•		Mus musculus	urogenital sinus	18.5 embryonic days	Male	20160826ShhcreDnmtiLOFME18.5U		
	W-R036	_	0	al sinus from Conditional Dnmt1 Shhcre/+; Dnmt1flox/flox) (2 of 3)	Mus musculus	urogenital sinus	18.5 embryonic	Male	20160826ShhcreDnmtiLOFM		

In vivo replacement of damaged bladder urothelium by Wolffian duct epithelial cells

#### Log In

🖌 Show All Related Records 🛛 💆 Export 🗸 🕑 Share

#### COLLECTION Contents W-R056 Bladder from Dnmt1 conditional knockout Mus bladder 18.5 Male 20160928H&EShhcreDnmt1LOFBIM embryo (Shhcre/+; Dnmt1flox/flox) (4 of 4) musculus embryonic Main days He Slide Collection (13) **Specimen Collection (25)** ➤ Specimen Collection (showing first 25 results) View More RID IÎ Species 11 Stage 1 Images 11 Genes 11 Assay Type 11 **Anatomical Sources** View Image 1 of 1 $\odot$ W-Mus **TS20** • urogenital sinus IHC R6QP musculus $\bigcirc$ **TS20** Image 1 of 1 Mus urogenital sinus IHC W-R6R2 musculus Image 1 of 1 **TS21** • urogenital sinus IHC W-Mus $\odot$ QYDP musculus INTO mation Sciences Institute

School of Engineering

#### **GUDMAP:W-R6QP** 🖌 Show All Related Records 🛛 Export 👻 🖸 Share SPECIMEN Contents W-R6QP RID **Species** Mus musculus Main Images (1) **TS20** Stage **Specimen Collection (1) Chronological Age** E12.5 IHC **Assay Type** Preparation section **Anatomical Source** urogenital sinus Table Display | View More 4% Paraformaldehyde Fixation Embedding Paraffin Strain Mixed Genotype Shhcre/+; Dnmt1flox/+ (Control) **Principal Investigator** Chad Vezina Consortium **GUDMAP Creation Time** 2018-05-25 16:58:58 **Last Modified Time** 2018-10-11 00:57:59 ➤ Images (showing all 1 results) View More Thumbnail URL 1 Original File URL 1 Notes 11 View 0 Blue-DAPI, Green-Shh lineage label (EYFP), Red-PAX2 20171009aShhcreDnmt1LOFE12.5MalePAX2EYFPHET.tif

#### GUDMAP:W-R6QP

#### 🕂 Create 💉 Edit 🔏 Copy 💼 Delete 🥣 Hide Empty Related Records 🔮 Export 🗸 🕑 Share

RID	W-R6QP		Contents
Gene	None Table Displa	ay   View Mo	Main
Protein	None	dit   View More	Acknowledgement (0)
Species	Mus musculus		Images (1)
Stage	TS20		<b>Derived Specimens (0)</b>
Chronological Age	E12.5		Specimen Expression (0)
Assay Type	IHC		Specimen Result (0)
Preparation	section		Specimen Allele (0)
Anatomical Source	urogenital sinus Edit   Ac	Id   View More	Specimen Probe Association (0)
Specimen Cell Type	None Edit   Ad	Id   View More	Probes (0)
Fixation	4% Paraformaldehyde		<sup>↔</sup> Genes (0)
Embedding	Paraffin		Specimen Antibody (0)
Strain	Mixed		Anchor Gene Specimen (0)
Genotype	Shhcre/+; Dnmt1flox/+ (Control)		Marker Gene Specimen
Experiment Note	None Edit   Ad	Id   View More	
<b>Record Status Detail</b>	Complete		Mouse Strain
Curation Status	Release		Characterized by this Specimen (0)
Principal Investigator	Chad Vezina		Mouse Strains Contributing to this
Consortium	GUDMAP		Specimen (0)
School o	of Engineering		

GUDMAP/RBK Res	SOURCES Search - Create -	Dev-Onlv search 👻 Dashboards (requires login) 👻 Help 👻 Feedback	>		carlkesselman@globusid.org 👻
Whole-mount	<b>3D views of the hum</b>	Share Share Link	X	e 🦽 Hide Empty Related Rec	ords 🔄 Export - 🕑 Share
<b>IID</b> ïitle	Q-3K5A Whole-mount 3D views of the	https://dev.gudmap.org/chaise/record/#2/Common:Collection/RID=Q-3K5A <b>Citation</b>			Contents Main
Description	A collection of human embryo Related to JASN https://doi.c	Andrew McMahon GUDMAP Consortium https://doi.org/10.25548/BURB-6P44 ( Download Citation:	(2017).	e Kidney Organogenesis.	He Slide Collection (0) Specimen Collection (0)
lequire DOI? ersistent ID lecord Status Detail	Yes https://doi.org/10.25548/BUI	BibTex	_		IF Slide Collection (13) IF Video Collection (0)
uration Status	Complete Release				Sequencing Study Collection (0)
rincipal Investigator ata Provider	Andrew McMahon, USC University of Southern Californi	a			
Consortium Creation Time	GUDMAP 2017-09-23 02:18:02				+
ast Modified Time	2017-09-23 02:18:02				
<ul> <li>✓ He Slide Collection</li> <li>Actions RID ↓↑</li> </ul>	n (no results found) Thumbnail It Name It	Species ↓↑ Tissue ↓↑ Age Stage ↓↑ Gend	der 🕴 Ima	Add   View More age File J1 Record State	
			No Resi	ults Found	
✓ Specimen Collecti	ณ.(กก เคริมประโณเทศ)			Add LViow Moro	

## In vivo replacement of damaged bladder urothelium by Wolffian duct epithelial cells

#### + Create 🖍 Edit 🔏 Copy 🛍 Delete 🥣 Hide Empty Related Records 🔮 Export - 🕑 Share

J

RID	W-QXXC				CSV BAG
Title	In vivo replacem	Main			
Description	Figures and data Joseph et al.	He Slide Collection (13) Specimen Collection (25)			
Details	The following ta	ble shows the	mapping of figures and image re	ecord IDs (RID) presented in the paper.	IF Slide Collection (0)
	Figure	Reference	Additional Images		IF Video Collection (0)
	1A	W-QXXW	W-QXZ4, W-QY2C		Sequencing Study Collection (0)
	1B	W-QY34	W-QY38, W-QY3C		
	1C	W-QY66	W-QY6T, W-QY86		
	1D	W-QY8Y	W-QY9A, W-QYA6		↔
	1E	W-QYB6	W-QYBP, W-QZ6T		
	1F	W-QYC2	W-QYCE, W-QYCT		
	1G	W-QYDP	W-QYDY, W-QYEA		
	1H	W-QYEP	W-QYF2, W-QYFE		
	11	W-QYGP	W-QYH2, W-QYHE		
	1J	W-QYHT	W-QYJ6, W-QYJJ		



Name	Size	Kind	Date Added
Collection	<del></del>	Folder	Today at 12:35 AM
🕨 🚞 metadata		Folder	Today at 12:35 AM
🔻 💼 data		Folder	Today at 12:35 AM
🖬 SequenciIlection.csv	324 bytes	Commet (.csv)	Today at 12:35 AM
🖬 IF Video Cection.csv	291 bytes	Commet (.csv)	Today at 12:35 AM
IF Slide Collection.csv	706 bytes	Commet (.csv)	Today at 12:35 AM
🖈 Specimenection.csv	176 KB	Commet (.csv)	Today at 12:35 AM
🖬 He Slidellection.csv	9 KB	Commet (.csv)	Today at 12:35 AM
Collection.csv	9 KB	Commet (.csv)	Today at 12:35 AM
📄 fetch.txt	3 KB	Plain Text	Today at 12:35 AM
manifest-md5.txt	2 KB	Plain Text	Today at 12:35 AM
📄 tagmanifest-md5.txt	238 bytes	Plain Text	Today at 12:35 AM
bag-info.txt	299 bytes	Plain Text	Today at 12:35 AM
bagit.txt	55 bytes	Plain Text	Today at 12:35 AM







#### Summary



- derivacloud.org
- https://fair-research.org
- Carl@isi.edu

