

# Gabriella Miller Kids First Pediatric Research Program

## *Public Webinar*

October 6, 2020

2:00 pm EDT



The Common  
Fund



# Webinar Agenda



**2:00pm** - Introduction; *NIH Kids First Staff*

**2:05pm** - Progress on the Kids First Study on Novel Cancer Susceptibility in Families; *Dr. Sharon Plon, Baylor College of Medicine*

- Questions for Dr. Plon

**2:35pm** - Kids First Data Resource Center;  
*Kids First DRC Staff*

- New Portal Features
- Kids First Variant Workbench Demo

**3:05pm** - NIH Common Fund Data Ecosystem (CFDE) presentation; *Dr. C. Titus Brown, UC Davis*

**3:25pm** - NIH Program Updates; *NIH Kids First Staff*

**3:40pm** - Questions & Answers



The Common  
Fund



## **Valerie Cotton**

Kids First Program Manager

*Eunice Kennedy Shriver* National Institute of  
Child Health and Human Development (NICHD)

# NIH Kids First Working Group

Kids First is an NIH Common Fund program coordinated by a trans-NIH Working Group, which is chaired by four institutes:

*Eunice Kennedy Shriver* National Institute of Child Health and Human Development (**NICHD**)

National Human Genome Research Institute (**NHGRI**)

National Heart, Lung, and Blood Institute (**NHLBI**)

National Cancer Institute (**NCI**)



The Common Fund



Other Working Group Representation:

**NIDCR**

**NIAAA**

**NIDDK**

**NEI**

**NIAID**

**ORIP**

**NIDA**

**NINDS**

**NIHHS**

**NIAMS**

**NCATS**

**CDC**

# How did Kids First get started?

- Initiated in response to the [2014 Gabriella Miller Kids First Research Act](#):
  - Signed into law on April 3, 2014
  - Ended taxpayer contribution to presidential nominating conventions
  - Transferred \$126 million into the Pediatric Research Initiative Fund
  - Authorized appropriation of \$12.6 million per year for 10 years to the NIH Common Fund for pediatric research; first appropriation was for FY2015

Form **1040** Department of the Treasury—Internal Revenue Service (99) **2014** U.S. Individual Income Tax Return OMB No. 1545-0074 IRS Use Only—Do not write or staple in this space.

For the year Jan. 1–Dec. 31, 2014, or other tax year beginning \_\_\_\_\_, 2014, ending \_\_\_\_\_, 2014. See separate instructions.

Your first name and initial \_\_\_\_\_ Last name \_\_\_\_\_ Your social security number \_\_\_\_\_

If a joint return, spouse's first name and initial \_\_\_\_\_ Last name \_\_\_\_\_ Spouse's social security number \_\_\_\_\_

Home address (number and street). If you have a P.O. box, see instructions. \_\_\_\_\_ Apt. no. \_\_\_\_\_

City, town or post office, state, and ZIP code. If you have a foreign address, also complete spaces below (see instructions).

Foreign country name \_\_\_\_\_ Foreign province/state/county \_\_\_\_\_ Foreign postal code \_\_\_\_\_

**Presidential Election Campaign**  
Check here if you, or your spouse if filing jointly, want \$3 to go to this fund. Checking a box below will not change your tax or refund.  
 You  Spouse

1  Single 4  Head of household (with qualifying dependent child who is your spouse's son, daughter, or stepchild)



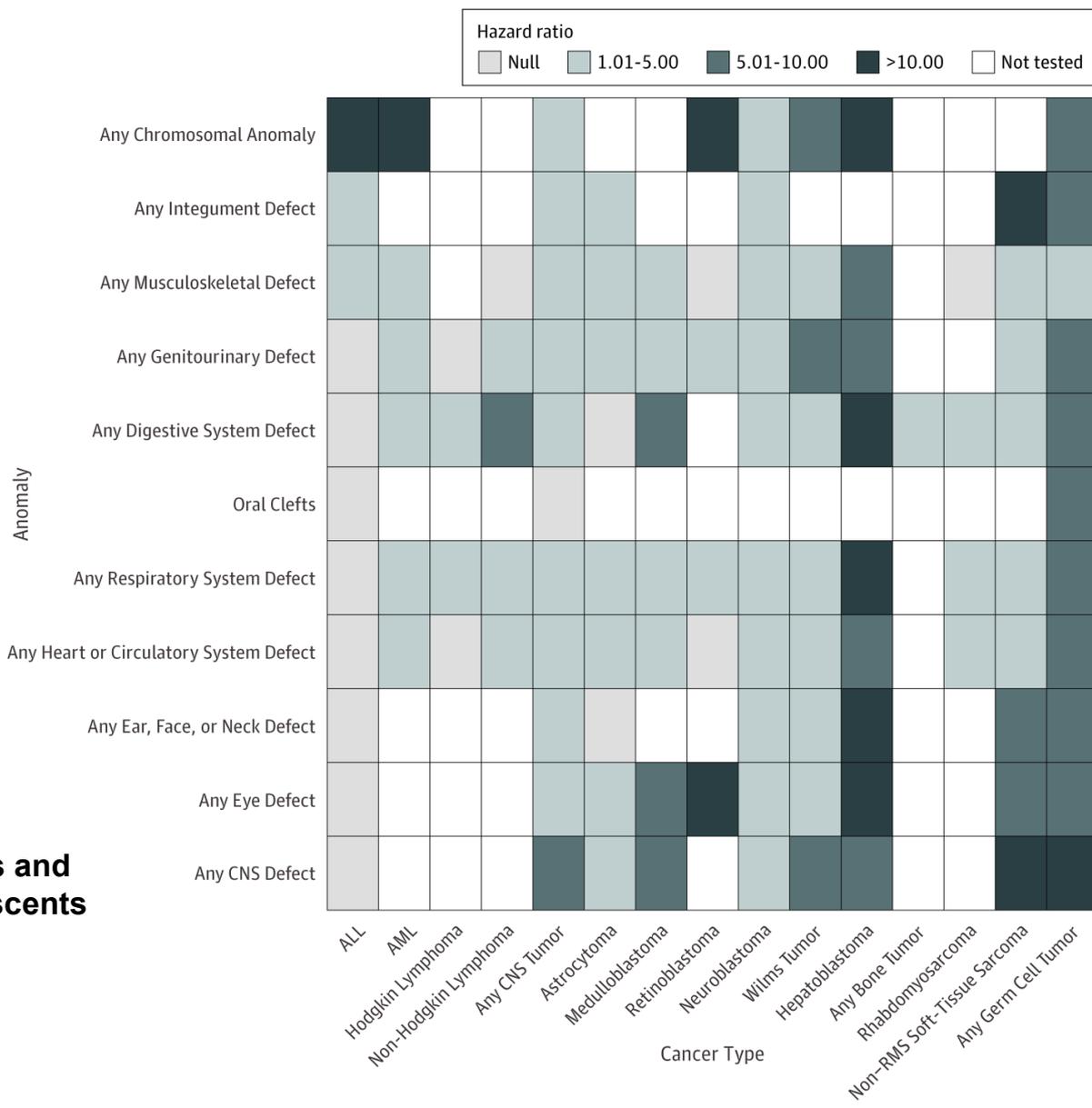
# Vision



Alleviate suffering from childhood cancer and structural birth defects by fostering **collaborative research** to uncover the etiology of these diseases and supporting **data sharing** within the pediatric research community.

# Why study childhood cancer & structural birth defects together?

- **Both are leading causes of childhood mortality**
- **Birth defects are associated with increased risk of cancer among children... suggesting shared genetic pathways**



From: **Association Between Birth Defects and Cancer Risk Among Children and Adolescents in a Population-Based Assessment of 10 Million Live Births**

Lupo et al, JAMA Oncol. 2019;5(8):1150-1158.  
doi:10.1001/jamaoncol.2019.1215

# Kids First Major Initiatives

Through 2021:

1. Identify & sequence cohorts of children with **childhood cancer and/or structural birth defects**.
2. Build the **Gabriella Miller Kids First Data Resource** to empower discovery

Year | 15 ✓ | 16 ✓ | 17 ✓ | 18 ✓ | 19 ✓ | 20 ✓ | 21 | 22 | 23 | 24



**Phase 2**

# The Kids First Dataset is Growing!

40 projects | 40,000 genomes | 16,000 cases | 14 released datasets



- Disorders of Sex Development
- Congenital Diaphragmatic Hernia
- Ewing Sarcoma
- Structural Heart & Other Defects
- Syndromic Cranial Dysinnervation Disorders
- Cancer Susceptibility
- Adolescent Idiopathic Scoliosis
- Neuroblastomas
- Enchondromatoses
- Orofacial Clefts in Caucasian, Latin American, Asian & African, Filipino populations
- Osteosarcoma
- Familial Leukemia
- Hemangiomas, Vascular Anomalies & Overgrowth
- Craniofacial Microsomia
- Intersection of childhood cancer & birth defects
- Microtia
- Esophageal Atresia and Tracheoesophageal Fistulas
- Kidney and Urinary Tract Defects
- Nonsyndromic Craniosynostosis
- Bladder Exstrophy
- Hearing Loss
- Cornelia de Lange Syndrome
- Intracranial & Extracranial Germ Cell Tumors
- Fetal Alcohol Spectrum Disorders
- Myeloid Malignancies + overlap with Down syndrome
- Congenital Heart Defects & Acute Lymphoblastic Leukemia in Children with Down Syndrome
- Structural Brain Defects
- Structural Defects of the Neural Tube (Spina Bifida: Myelomeningocele)
- CHARGE Syndrome
- Laterality Birth Defects
- T-cell Acute Lymphoblastic Leukemia
- Pediatric Rhabdomyosarcoma
- Valvar Pulmonary Stenosis

# The Kids First Data Resource for Collaborative Discovery

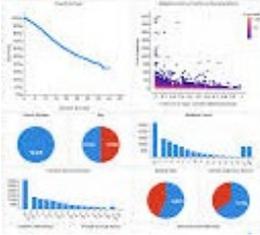
## Data Resource Portal

Entry point. Query, search, discover, build & visualize synthetic cohorts



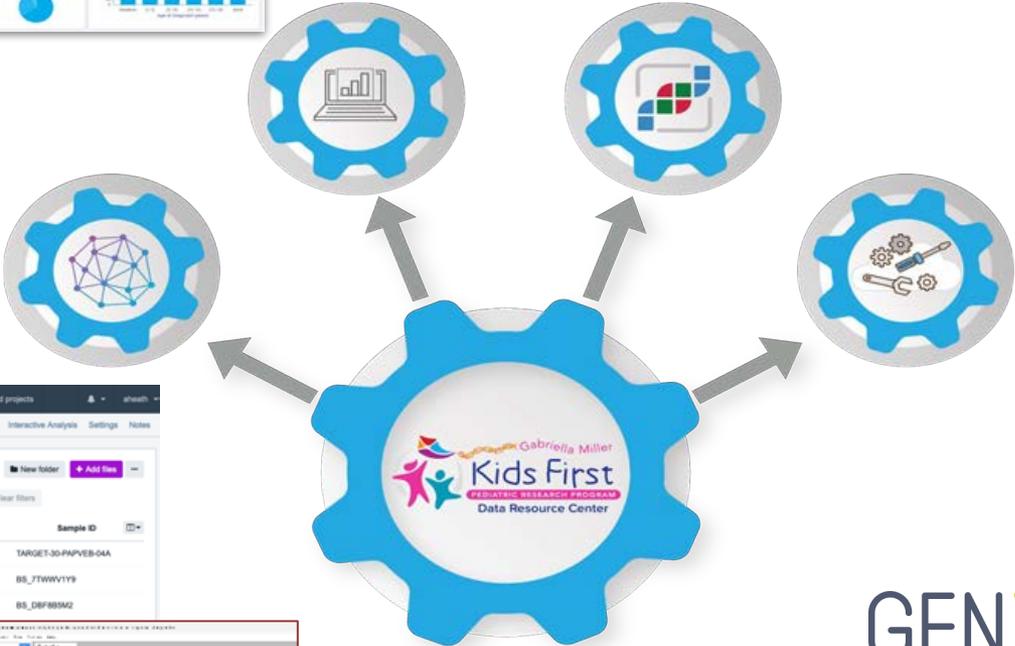
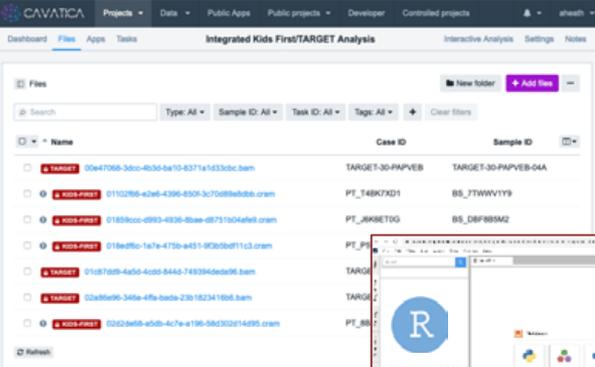
## Knowledge Base Integrations (PedcBioPortal)

Integrations with existing curated/published data visualizations



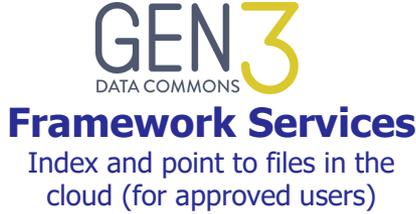
## Cavatica

Pull data from multiple sources into one workspace. Use notebooks, bring-your-own or use available workflows.



## Data Services

Model clinical data in FHIR-based data services for semantic interoperability and coordination



# More researchers are accessing Kids First data!

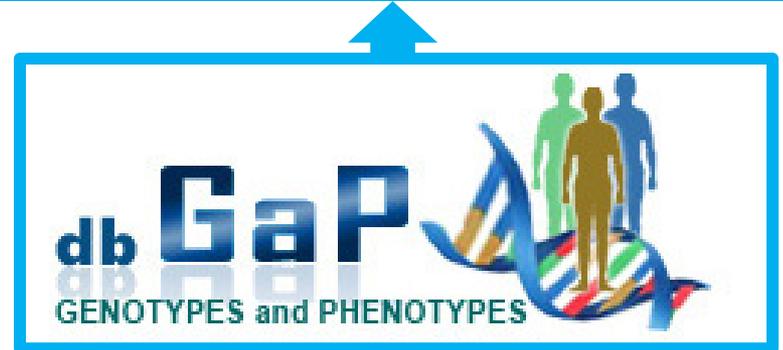
The screenshot displays the Kids First Data Repository interface. On the left, there are filter sections for 'Study Name' (listing categories like Pediatric Brain Tumors, Orofacial Cleft, etc.) and 'Diagnosis Category' (listing Cancer, Other, etc.). The main area shows a table with columns: File ID, Participants ID, Study Name, Proband, Family ID, Data Type, File Format, and File Size. The table lists 20 files, all related to 'Congenital Diaphragmatic Hernia'. Summary statistics at the top indicate 28,810 files, 5,621 participants, 1,625 families, and 750.47 TB size. On the right, there are 'Actions' and 'Data Analysis' sections.

>1700  
registered  
users since  
2018 launch



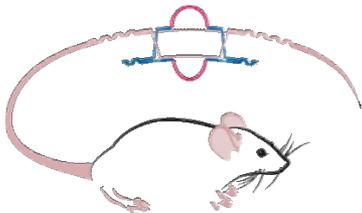
Individual-level sequence data  
>150 Data Access Requests  
approved by the Kids First Data  
Access Committee across 14 Kids  
First genomic datasets available

NIH Kids First  
Data Access Committee



# *Researchers are using Kids First data to answer new scientific questions*

- **14 awards for R03 for analyses of Kids First data** (PAR-16-348 ; PAR-18-733; PAR-19-069, [PAR-19-375](#))
- **2 awards for NIDCR R03** (PAR-16-070)
- **2 awards for INCLUDE R03** (RFA-OD-20-006)
  
- **3 awards for R01s** (PA-13-302, PAR-17-236)
  
- **Spurred new collaborations with KOMP2 & INCLUDE**



*Knockout Mouse Phenotyping Project (KOMP2)*



*INvestigation of Co-occurring conditions across the Lifespan to Understand Down syndrome (INCLUDE)*

# Researchers Are Making Discoveries

19 Journal Publications To-Date  
Average RCR: 1.36

<https://commonfund.nih.gov/publications?pid=40>



## Genomic analyses implicate noncoding de novo variants in congenital heart disease

Felix Richter<sup>1,31</sup>, Sarah U. Morton<sup>2,3,31</sup>, Seong Won Kim<sup>4,31</sup>, Alexander Kitaygorodsky<sup>5,31</sup>, Lauren K. Wasson<sup>4,31</sup>, Kathleen M. Chen<sup>6,31</sup>, Jian Zhou<sup>6,7,8</sup>, Hongjian Qi<sup>9</sup>, Nihir Patel<sup>9</sup>, Steven R. DePalma<sup>4</sup>, Michael Parfenov<sup>4</sup>, Jason Homsy<sup>4,10</sup>, Joshua M. Gorham<sup>4</sup>, Kathryn B. Manheimer<sup>11</sup>, Matthew Velinder<sup>12</sup>, Andrew Farrell<sup>12</sup>, Gabor Marth<sup>12</sup>, Eric E. Schadt<sup>9,11,13</sup>, Jonathan R. Kaltman<sup>14</sup>, Jane W. Newburger<sup>15</sup>, Alessandro Giardini<sup>16</sup>, Elizabeth Goldmuntz<sup>17,18</sup>, Martina Brueckner<sup>19</sup>, Richard Kim<sup>20</sup>, George A. Porter Jr.<sup>21</sup>, Daniel Bernstein<sup>22</sup>, Wendy K. Chung<sup>23</sup>, Deepak Srivastava<sup>24,32</sup>, Martin Tristani-Firouzi<sup>25,32</sup>, Olga G. Troyanskaya<sup>6,7,26,32</sup>, Diane E. Dickel<sup>27,32</sup>, Yufeng Shen<sup>5,32</sup>, Jonathan G. Seidman<sup>4,32</sup>, Christine E. Seidman<sup>4,28,32</sup> and Bruce D. Gelb<sup>9,29,30,32</sup> ✉



## Deleterious de novo variants of X-linked ZC4H2 in females cause a variable phenotype with neurogenic arthrogyposis multiplex congenita

Suzanna G.M. Frints ✉, Friederike Hennig, Roberto Colombo, Sebastien Jacquemont, Paulien Terhal, Holly H. Zimmerman, David Hunt, Bryce A. Mendelsohn, Ulrike Kordaß ... See all authors

First published: 17 June 2019 | <https://doi.org/10.1002/humu.23841>



Original Investigation | Open Access | Published: 17 December 2019

## Whole genome sequencing of orofacial cleft trios from the Gabriella Miller Kids First Pediatric Research Consortium identifies a new locus on chromosome 21

Nandita Mukhopadhyay, Madison Bishop, Michael Mortillo, Pankaj Chopra, Jacqueline B. Hetmanski, Margaret A. Taub, Lina M. Moreno, Luz Consuelo Valencia-Ramirez, Claudia Restrepo, George L. Wehby, Jacqueline T. Hecht, Frederic Deleyiannis, Azeez Butali, Seth M. Weinberg, Terri H. Beaty, Jeffrey C. Murray, Elizabeth J. Leslie, Eleanor Feingold & Mary L. Marazita ✉

*Human Genetics* 139, 215–226 (2020) | [Cite this article](#)

1039 Accesses | 11 Altmetric | [Metrics](#)

## PLOS GENETICS

OPEN ACCESS PEER-REVIEWED

RESEARCH ARTICLE

## De novo variants in congenital diaphragmatic hernia identify MYRF as a new syndrome and reveal genetic overlaps with other developmental disorders

Hongjian Qi ✉, Lan Yu ✉, Xueya Zhou ✉, Julia Wynn, Haoquan Zhao, Yicheng Guo, Na Zhu, Alexander Kitaygorodsky, Rebecca Herman, Gudrun Aspelund, Foong-Yen Lim, Timothy Crombleholme, Robert Cusick, [ ... ], Yufeng Shen ✉ [view all]

Version 2 | Published: December 10, 2018 • <https://doi.org/10.1371/journal.pgen.1007822>

Received: 18 April 2019 | Revised: 30 May 2019 | Accepted: 9 July 2019  
DOI: 10.1371/journal.pgen.1007822

ORIGINAL ARTICLE



## Phenotype delineation of ZNF462 related syndrome

Paul Kruszka<sup>1</sup> ✉, Tommy Hu<sup>1</sup> | Sungkook Hong<sup>1</sup> | Rebecca Signer<sup>2</sup> | Benjamin Cogné<sup>3</sup> | Bertrand Isidor<sup>3</sup> | Sarah E. Mazzola<sup>4</sup> | Jacques C. Giltay<sup>5</sup> | Koen L. I. van Gassen<sup>5</sup> | Eleina M. England<sup>6</sup> | Lynn Pais<sup>6</sup> | Charlotte W. Ockeloen<sup>7</sup> | Pedro A. Sanchez-Lara<sup>8,9</sup> | Esther Kinning<sup>10</sup> | Darius J. Adams<sup>11</sup> | Kayla Treat<sup>12</sup> | Wilfredo Torres-Martinez<sup>12</sup> | Maria F. Bedeschi<sup>13,19</sup> | Maria lascone<sup>14</sup> | Stephanie Blaney<sup>15</sup> | Oliver Bell<sup>8</sup> | Tiong Y. Tan<sup>16,17,18</sup> | Marie-Ange Delrue<sup>19</sup> | Julie Jurgens<sup>20</sup> | Brenda J. Barry<sup>6,21</sup> | Elizabeth C. Engle<sup>6,21,22</sup> | Sarah K. Savage<sup>23</sup> | Nicole Fleischer<sup>23</sup> | Julian A. Martinez-Agosto<sup>2</sup> ✉ | Kym Boycott<sup>24</sup> ✉ | Elaine H. Zackai<sup>4</sup> | Maximilian Muenke<sup>1</sup> ✉

# **Kids First X01:** Identifying novel cancer susceptibility mutations from unselected childhood cancer patient and parent trios

---



**Sharon E. Plon, MD, PhD, FACMG**  
Baylor College of Medicine



**Owen Hirschi**  
Baylor College of Medicine



# 1 X01 HL136994-01 - Identifying novel cancer susceptibility mutations from unselected childhood cancer patient and parent trios (BASIC3)

Sharon E. Plon, MD, PhD, FACMG

Departments of Pediatrics/Hematology-Oncology  
and Molecular and Human Genetics

Human Genome Sequencing Center

Baylor College of Medicine

## **Disclosures – Sharon E. Plon, MD, PhD**

- I have the following financial relationships to disclose:
  - I am a member of the Baylor Genetics Laboratory Scientific Advisory Board

# Three different approaches to understanding genetic basis of childhood cancer

- Several studies of clinical genome-scale testing of diverse cohorts of childhood cancer patients (BASIC3 and Texas KidsCanSeq study)
- Sequencing of large cohorts of specific pediatric cancer patients
- Precision oncology treatment trials of matched tumor/normal sequencing for patients with relapsed or recurrent tumors (NCI/COG Pediatric MATCH Trial)



# Clinical Sequencing Exploratory Research

*Moving the Genome Into the Clinic*

377 Researchers  
21 Institutions  
1 Consortium



# Baylor College of Medicine BASIC3 Key Team Members



Will Parsons  
Pediatric Oncology



Murali  
Chintagumpala



Stacey  
Berg



Richard  
Gibbs



Christine  
Eng



David  
Wheeler



Amy  
McGuire



Laurence  
McCullough



Richard  
Street



Sue  
Hilsenbeck



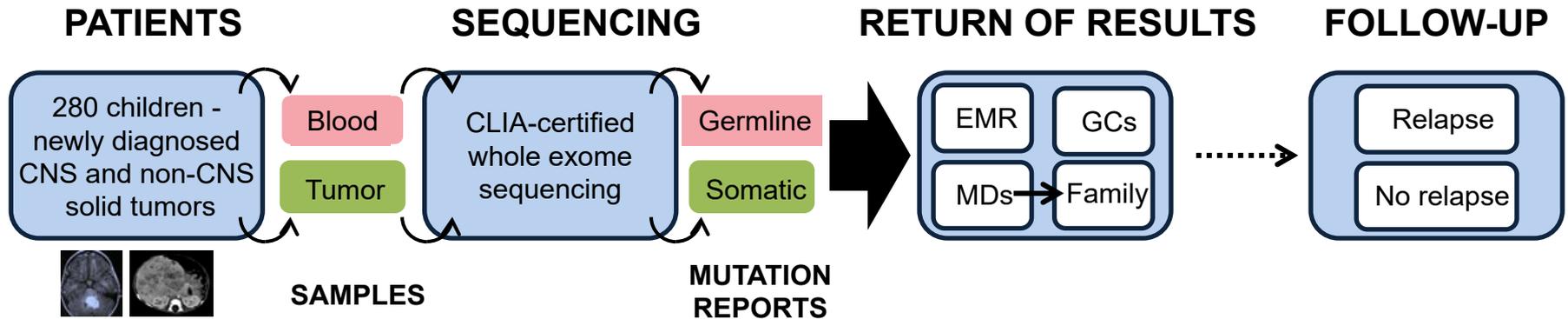
Angshumoy  
Roy



Dolores  
Lopez-Terrada

# BASIC<sup>3</sup>

Baylor College of Medicine **A**dvancing **S**equencing **I**nto **C**hildhood **C**ancer **C**are



## Study objectives:

- To integrate information from CLIA-certified germline and tumor exome sequencing into the care of newly diagnosed solid and brain tumor patients at Texas Children's Cancer Center
- To perform parallel evaluation of the impact of tumor and germline exomes on families and physicians



Will Parsons  
Pediatric Oncology

Research

Original Investigation

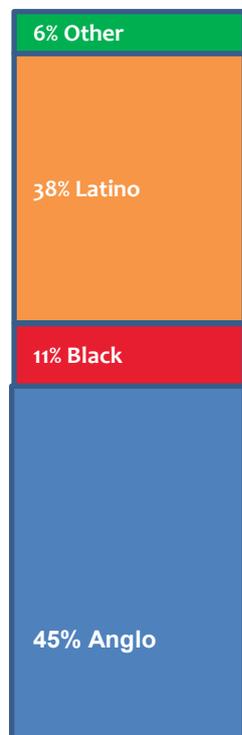
# Diagnostic Yield of Clinical Tumor and Germline Whole-Exome Sequencing for Children With Solid Tumors

D. Williams Parsons, MD, PhD; Angshumoy Roy, MD, PhD; Yaping Yang, PhD; Tao Wang, PhD; Sarah Scollon, MS, CGC; Katie Bergstrom, MS, CGC; Robin A. Kerstein, BS, MT; Stephanie Gutierrez, BS; Andrea K. Petersen, MD; Abhishek Bavle, MD; Frank Y. Lin, MD; Dolores H. López-Terrada, MD, PhD; Federico A. Monzon, MD; M. John Hicks, MD, PhD, DDS; Karen W. Eldin, MD; Norma M. Quintanilla, MD; Adekunle M. Adesina, MD, PhD; Carrie A. Mohila, MD, PhD; William Whitehead, MD; Andrew Jea, MD; Sanjeev A. Vasudevan, MD; Jed G. Nuchtern, MD; Uma Ramamurthy, PhD; Amy L. McGuire, JD, PhD; Susan G. Hilsenbeck, PhD; Jeffrey G. Reid, PhD; Donna M. Muzny, MSc; David A. Wheeler, PhD; Stacey L. Berg, MD; Murali M. Chintagumpala, MD; Christine M. Eng, MD; Richard A. Gibbs, PhD; Sharon E. Plon, MD, PhD

*JAMA Oncol.* doi:10.1001/jamaoncol.2015.5699

Published online January 28, 2016.

# Race/Ethnicity of BASIC3 Subjects are Representative of Houston Population



Texas

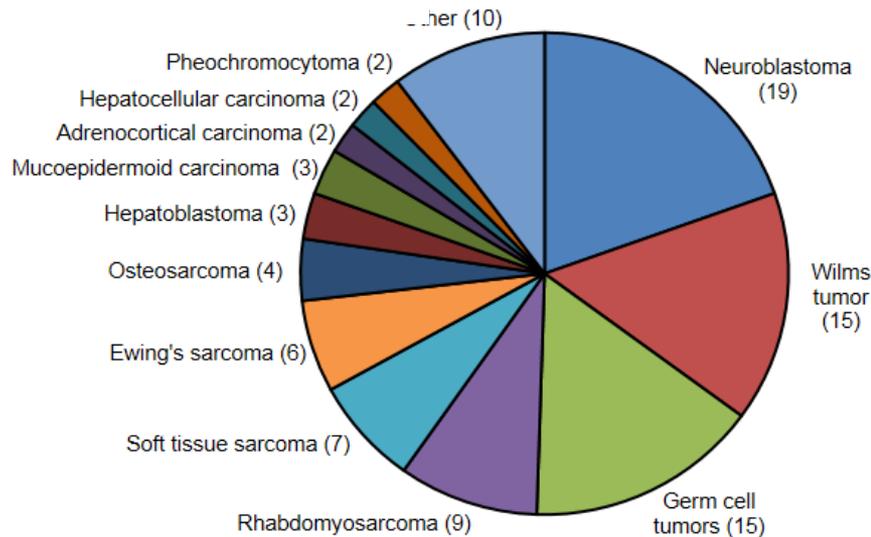
## Characteristics of patients enrolled and not enrolled on study - updated

Characteristic - no. (%)	Enrolled (n=239)	Declined (n=103)	P Value
Ethnicity			0.54
Hispanic	111 (46%)	41 (40%)	
Non-Hispanic	119 (50%)	52 (50%)	
Not reported	10 (4%)	10 (10%)	
Race			0.11
White	141 (59%)	74 (72%)	
Black or African American	25 (10%)	12 (12%)	
Asian	7 (3%)	4 (4%)	
American Indian or Alaska Native	10 (4%)	2 (2%)	
Multiple	14 (6%)	---	
Not reported	42 (18%)		

Updated from Scollon et al., *Genome Medicine* 2014

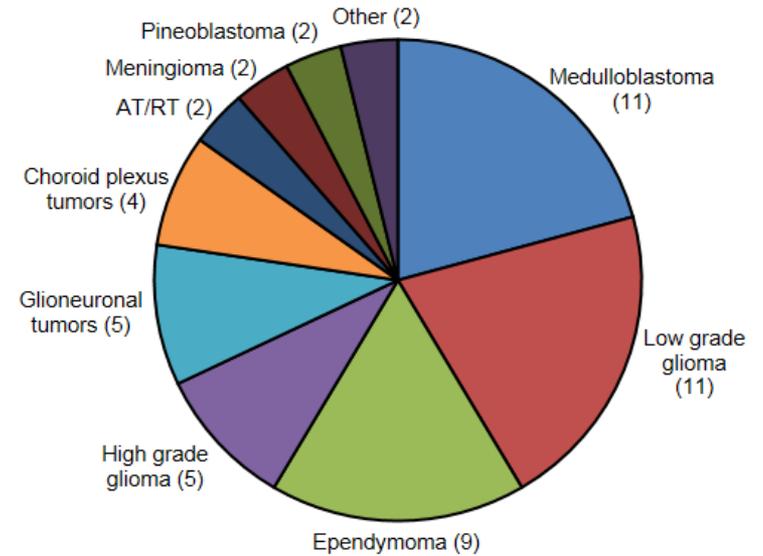
# BASIC<sub>3</sub> DIVERSE PEDIATRIC TUMOR DIAGNOSES

## NON-CNS



**81/94 (86%)**

## CNS

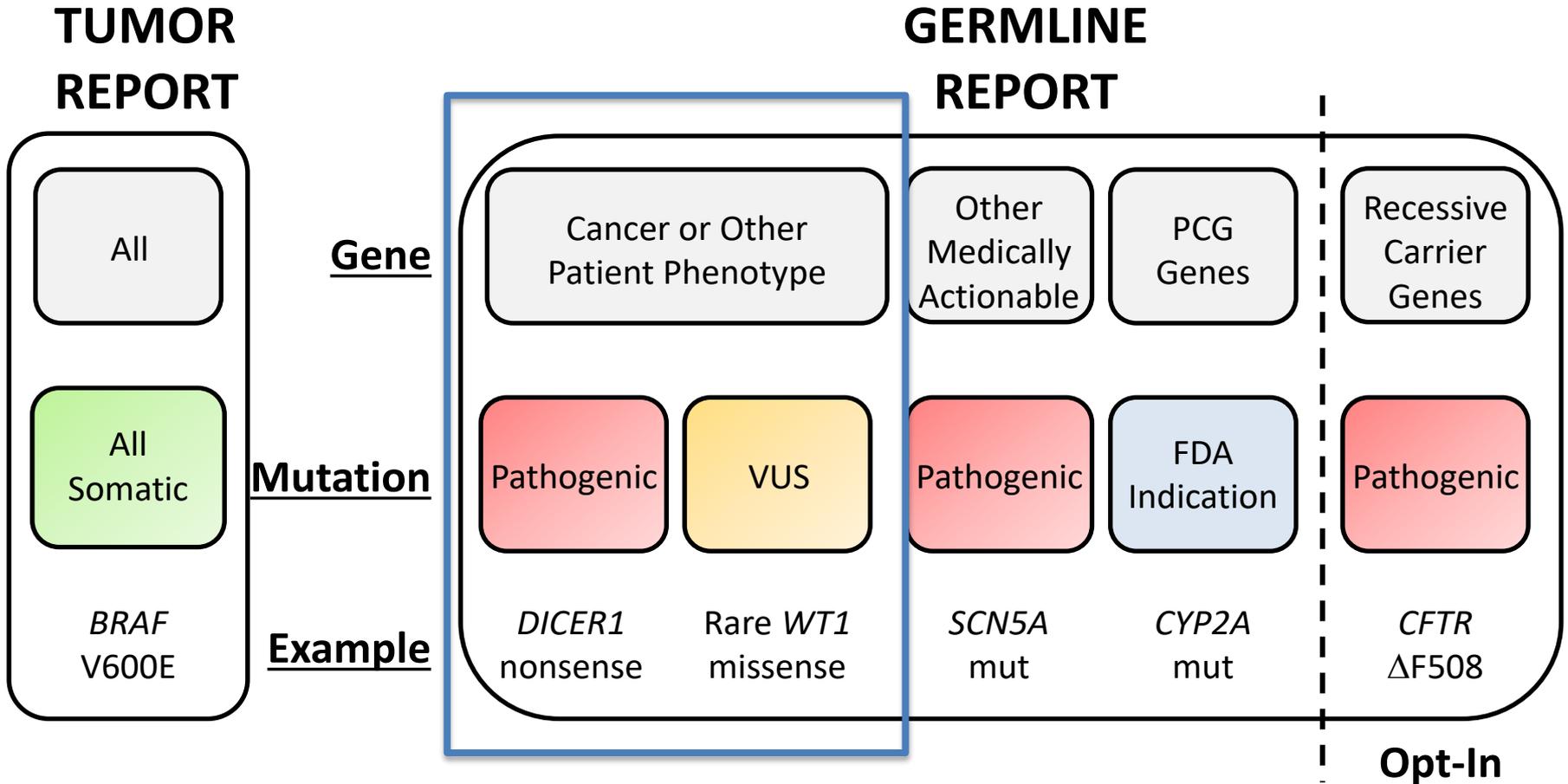


**40/56 (71%)**

---

**Tumor available for WES**

# Diversity of clinical exome results returned



# Cancer susceptibility molecular diagnosis in 9.8% (27/278) pediatric cancer patients

<b>Autosomal dominant (P/LP)</b>	<b>26</b>	<b>19 different genes</b>
Genes associated w/ specific childhood cancer	15	<i>Examples include DICER1, VHLx3, MSH2, WT1x2, TP53x3</i>
Genes not previously associated w/ specific childhood cancer	11	<i>Examples include BRCA1x2, BRCA2, PALB2, CHEK2x2, FLCN, SMARCA4</i>
<b>Autosomal recessive (biallelic)</b>	<b>1</b>	<b>TJP2</b>
<b>No one gene was reported in more than 3 BASIC3 patients: 3 each for VHL and TP53.</b>		

# Rationale for WGS from Heterogeneous Dataset

- Several examples from BASIC3 exome data of identifying a rare variant in a single individual that was then enlarged to identify other rare cases outside the study resulting in new germline or somatic drivers.
  - *TJP2* deficiency and risk of hepatocellular carcinoma
  - Internal tandem duplication of *BCOR* as major somatic driver of clear cell sarcoma of the kidney

HEPATOLOGY

Official Journal of the American Association for the Study of Liver Diseases

CLINICAL OBSERVATIONS IN HEPATOLOGY

## Hepatocellular Carcinoma Associated With Tight-Junction Protein 2 Deficiency

Shengmei Zhou, M.D.,<sup>1,2</sup> Paula M. Hertel,<sup>3</sup> Milton J. Finegold,<sup>4</sup> Larry Wang,<sup>1,2</sup> Nanda Kerkar,<sup>2,5</sup> Jing Wang,<sup>6</sup> Lee-Jun C. Wong,<sup>6</sup> Sharon E. Plon,<sup>7</sup> Melissa Sambrotta,<sup>8</sup> Pierre Foskett,<sup>9</sup> Zhiyv Niu,<sup>6</sup> Richard J. Thompson,<sup>8</sup> and A.S. Knisely<sup>9</sup>



ARTICLE

Received 15 Jun 2015 | Accepted 14 Oct 2015 | Published 17 Nov 2015

DOI: 10.1038/ncomms9891

OPEN

## Recurrent internal tandem duplications of *BCOR* in clear cell sarcoma of the kidney

Angshumoy Roy<sup>1,2,3,4,5</sup>, Vijetha Kumar<sup>1,2</sup>, Barry Zorman<sup>4</sup>, Erica Fang<sup>1,2</sup>, Katherine M. Haines<sup>6</sup>, HarshaVardhan Doddapaneni<sup>6,7</sup>, Oliver A. Hampton<sup>6,7</sup>, Simon White<sup>7</sup>, Abhishek A. Bavle<sup>4</sup>, Nimesh R. Patel<sup>1,2</sup>, Karen W. Eldin<sup>1,2</sup>, M. John Hicks<sup>1,2,3,4,5</sup>, Dinesh Rakheja<sup>8,9,10</sup>, Patrick J. Leavey<sup>10</sup>, Stephen X. Skapek<sup>10</sup>, James F. Amatruda<sup>10</sup>, Jed G. Nuchtern<sup>3,5,11,12</sup>, Murali M. Chintagumpala<sup>3,4,5</sup>, David A. Wheeler<sup>6,7</sup>, Sharon E. Plon<sup>3,4,5,6,7</sup>, Pavel Sumazin<sup>3,4</sup> & D. Williams Parsons<sup>3,4,5,6,7</sup>

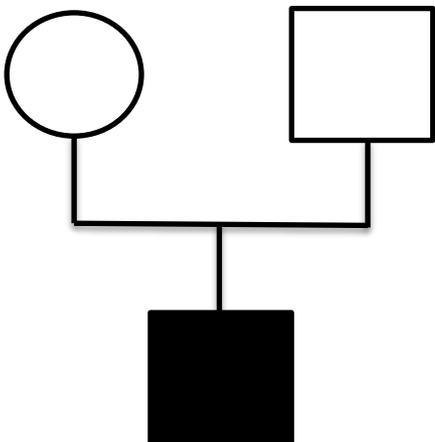
# Goal of KidsFirst Project

- Utilize the rich existing dataset from BASIC3 to make further discoveries using whole genome sequencing of proband and parent trios.
- Existing data:
  - Germline exomes of probands only
  - Tumor exomes from probands
  - RNAseq for a subset of tumors
  - DNA from patient and parents isolated from blood samples
  - Lymphoblastoid cell lines available from almost all patients and parents for subsequent functional or splicing studies as needed.

# Approach to KidsFirst Analysis

- Perform sequencing of patient/parent BASIC3 trios with following priority:
  - Complete trios with adequate DNA
  - Patients with unusual histories (multiple malignancies, birth defects)
  - Some solved cases to look for evidence of genomic instability
    - For example, number of de novo variants in child of *TP53* carrier
- Analysis has focused on both single nucleotide variants (SNV) and structural variants (SV)
  - Prioritized *de novo* variants in both situations
  - Also analyzed known cancer susceptibility genes for “missed” variants from exome analysis
  - Performed spliceAI analysis of rare variants for cryptic splicing variants

# De novo SNV calling in BASIC3 WGS trios (n=54 complete trios)



## Platypus

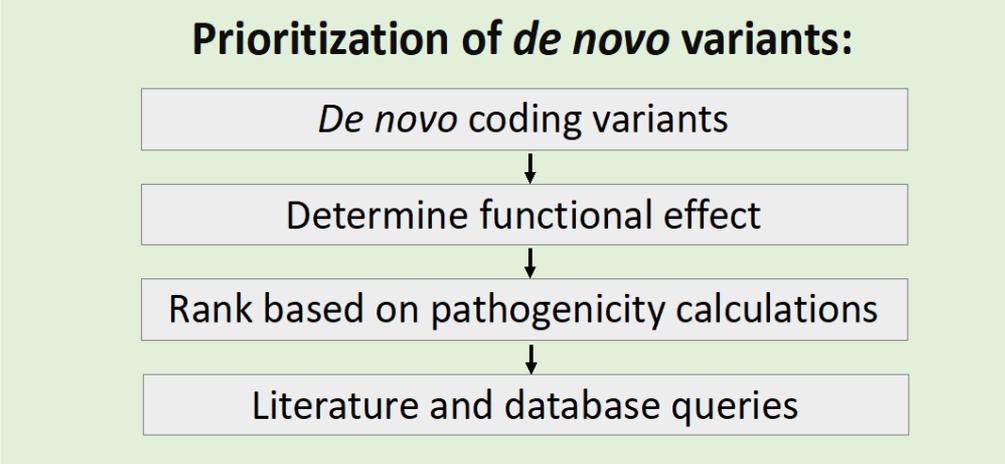
Published: 13 July 2014

**Integrating mapping-, assembly- and haplotype-based approaches for calling variants in clinical sequencing applications**

Andy Rimmer<sup>1 na1</sup>, Hang Phan<sup>1 na1</sup>, Iain Mathieson<sup>1</sup>, Zamin Iqbal<sup>1</sup>, Stephen R F Twigg<sup>2</sup>, WGS500 Consortium, Andrew O M Wilkie<sup>2</sup>, Gil McVean<sup>1,3 na1</sup> & Gerton Lunter <sup>1</sup>

Nature Genetics 46, 912-918(2014) | Cite this article

## Prioritization:

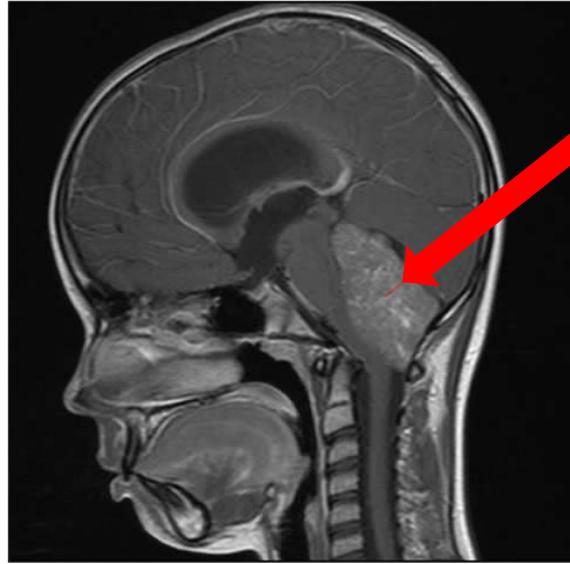


## Outcome:

- SNV analysis completed on 54 proband-parent trios
- The pipeline resulted in an expected number of variants per trio

Variant Type	Frequency per trio
Genome-wide <i>de novo</i>	60 to 190
Coding <i>de novo</i>	0 to 4

# De novo heterozygous missense variant in *KMT5C* in pediatric ependymoma



- Mutation is rare, occurs in highly conserved functional domain, and has a high CADD (28.3) and REVEL (.662) score

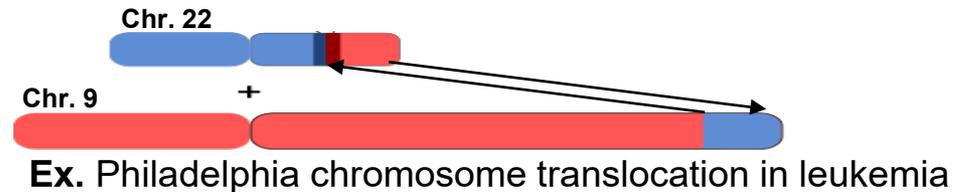


# **STRUCTURAL VARIANT CALLING IN BASIC<sub>3</sub>/KIDSFIRST WGS DATA**

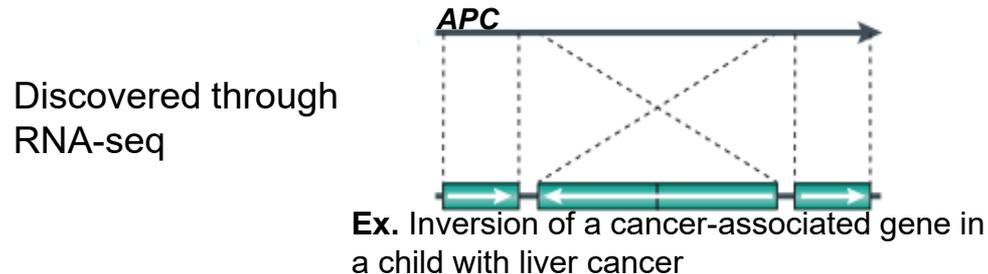
Owen Hirschi  
Graduate Student  
Genetics and Genomics Program  
Baylor College of Medicine

# Identification of germline structural variation in childhood cancer is limited

**Somatic structural variants** play a critical role in cancers



**Germline structural variants** are implicated in pediatric cancers



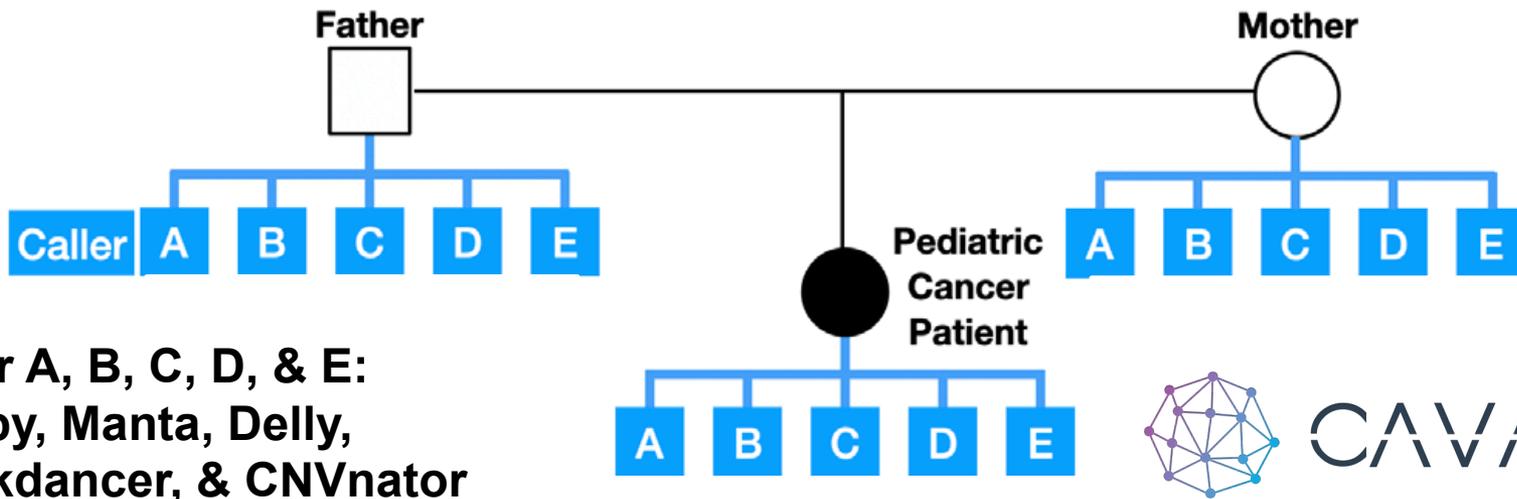
Haines K, et al. 2019

Since 2010 **over 40** programs have been created to identify SV from short-read WGS - primarily built for identification in tumors

- Sensitivity ranges from **10-70%** and false positive rate **up to 89%**

Alkan C, et al. 2011  
Mahmoud M, et al. 2019

# Developed new pipeline for more effective germline structural variant calling

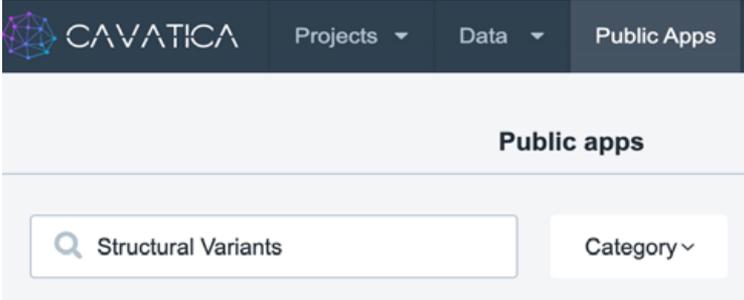


Caller A, B, C, D, & E:  
Lumpy, Manta, Delly,  
Breakdancer, & CNVnator

Using multiple structural variant callers allowed for more specificity and sensitivity in variant identification

The pipeline is built for identification of *de novo* structural variants and is being tuned for inherited variants

# Analysis of structural variant on Cavatica requires multiple features of the platform



## Explore genomics data

Understand complex genomics data with interactive analysis tools.

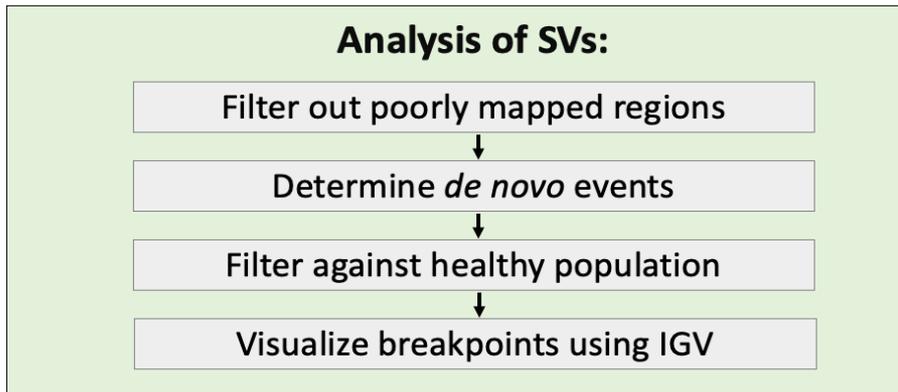
The 'Data Cruncher' interface displays a JupyterLab or RStudio environment. It features a code editor at the top and a plot area below showing a distribution of data points. A blue 'Open' button is located at the bottom right of the interface.

The screenshot shows a job completion notification. At the top, a black box with white text says 'Completed'. Below it, the job details are: 'BATCH 165 Delly - Call run - 01-25-20 21:17:01'. Further down, it states 'Executed on Jan. 25, 2020 15:22 by owenhirschi | Batch by: File'. Resource usage is listed as 'Spot Instances: On', 'Memoization: Off', and 'Price: \$68.37'. At the bottom, a dropdown menu shows 'App: Delly - Call - Revision: 0'.

```
1 #!/bin/bash
2
3
4 #DEL
5 n=1
6 while [ $n -le 165 ]
7 do
8     # Stack all the filtered SV calls per sample,
9     # and perform some additional filtering
10
11     # update: small filter size change from 1kb to 100bp
12     # update: use new exclude region
13     # update: specify caller in script
14     # update: directly decide reciprocal overlap using new script
15     # update: remove vqoutta
16     # update: remove sample name and role in output
17     # update: intersect with healthy control stack data to identify denovo variant
18     # update: need output that stack all healthy calls
19
20     event=DEL
21     callers=(cnvnator delly lumpy manta breakdancer)
22     sizeMax=1000000
23     sizeMin=100
24     rootDir=/sbgenomics/workspace
25     listFiles=$rootDir/FullList.list
26     infoFiles=$rootDir/BASIC3.info
27     sample=$(sed -n "$n"p" $listFile)
28     healthyList=$rootDir/Parent.list
29     outputDir=/sbgenomics/workspace/CNV_v02/Sevent/stack_by_sample/$sample
30     output=$outputDir/stack.bed
31     filterOutput=$outputDir/stack.filter.bed
32     recipOut=$outputDir/stack.filter.recip
33     excludeRegion=/sbgenomics/workspace/RLCRs_no_Repeat_Masker_3B.txt
34     healthyStack=$rootDir/CNV_v02/Sevent/stack_by_sample/stack.allHealthy.bed
35
36     ##### make healthyStack if not exist #####
```

# De novo structural variant calling in BASIC3 WGS trios (n=54 trios)

## Analysis:



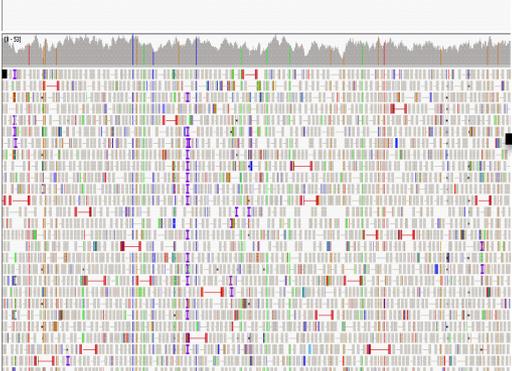
## Outcome:

- SV analysis completed on 54 proband-parent trios
- The pipeline identified between 0 to 1 SVs per trio

Structural Variant Type	# of SVs identified
Deletions	10
Duplications	7
Inversions	0

# Germline *de novo* heterozygous 20kbp deletion of *NF2* in a meningioma patient

Mother



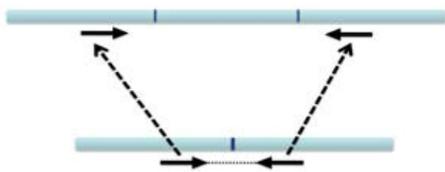
Father



Proband



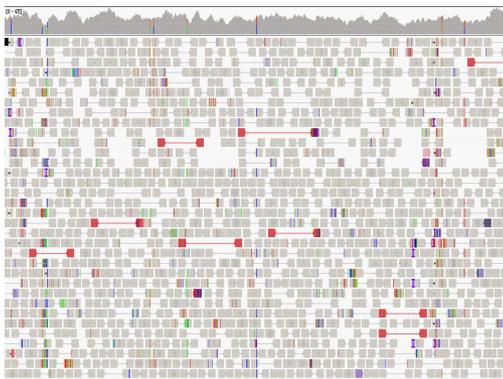
*NF2* exon:5-9



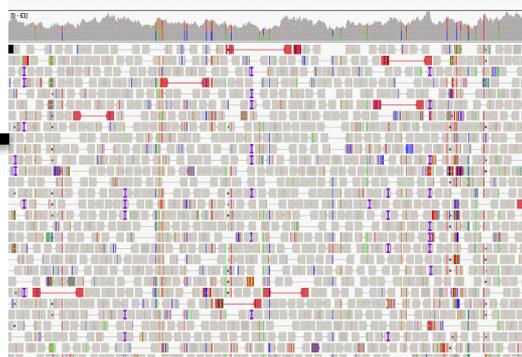
Potentially Somatic  
Mosaic  
6/40 reads

# Germline *de novo* heterozygous 8kbp duplication of *PTGR2* in an ependymoma patient

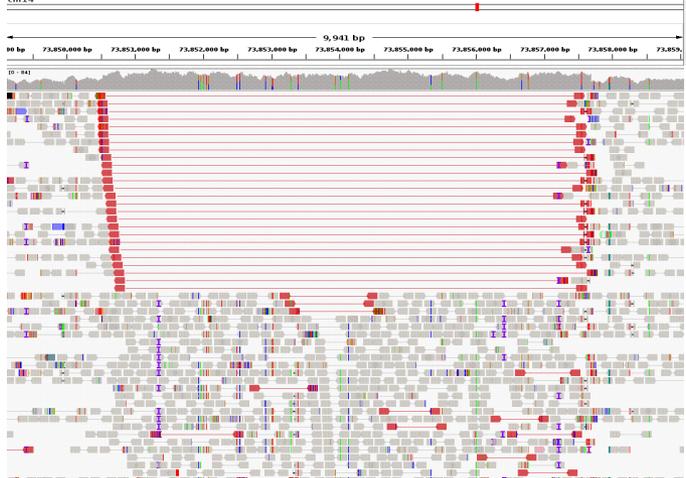
Mother



Father

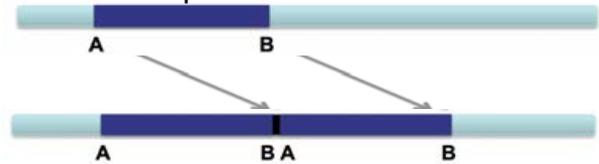


Proband



Head-to-tail insertion

*PTGR2* promoter and exon 1



# PTGR2 duplication includes 5 predicted promoter-like regions

	DNase max-Z	H3K4me3 max-Z	H3K27ac max-Z	CTCF max-Z	chr	start	length	nearest genes: protein-coding / all	predicted function
P 	5.50	4.63	4.48	3.11	chr14	73,851,666	347	<b><i>PTGR2</i></b>	<b>Promoter</b>
P 	5.15	4.75	4.64	2.76	chr14	73,852,048	168	<b><i>PTGR2</i></b>	<b>Promoter</b>
P 	4.37	4.38	4.35	1.67	chr14	73,852,538	348	<b><i>PTGR2</i></b>	<b>Promoter</b>
P 	4.31	3.86	4.25	1.81	chr14	73,850,993	238	<b><i>PTGR2</i></b>	<b>Promoter</b>
P 	3.18	1.96	2.63	1.29	chr14	73,850,000	218	<b><i>PTGR2</i></b>	<b>Promoter</b>

-  High H3K4me3
-  High H3K27ac
-  High CTCF
-  High DNase
-  Z-score < 1.64

P/D Proximal/Distal to a Transcription Start Site

# Whole genome sequencing revealed novel germline structural variants

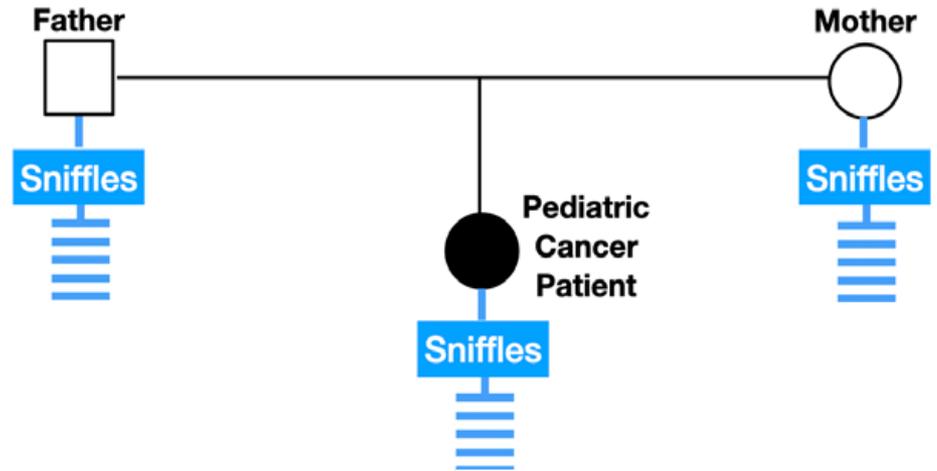
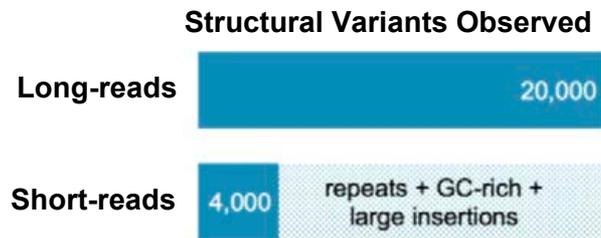
We have developed a pipeline with the aim to better identify germline *de novo* structural variation in pediatric cancer

We have identified a germline *de novo* heterozygous duplication in a pediatric ependymoma that we are beginning to functionally assess

We have benefited from additional research studies involving both tumor and germline sequencing of *BASIC3*

# Long-read sequencing of BASIC3 through Kids First

Long-read sequencing allows for greater detection of SV



## Goals of this sequencing

study will compare the SV results of the long-read sequencing to those of the short-read sequencing using our pipeline

Enabling us to identify:

- Limitations and accuracy of short-read structural variant in trio studies
- Causative SVs only found via long-read sequencing



**QUESTIONS?**



# Gabriella Miller Kids First Data Resource Center





**Adam Resnick, PhD**  
Children's Hospital of Philadelphia  
Principal Investigator,  
Gabriella Miller Kids First  
Data Resource Center



**Vincent Ferretti, PhD**  
Sainte-Justine University Hospital  
Principal Investigator,  
Gabriella Miller Kids First  
Data Resource Portal

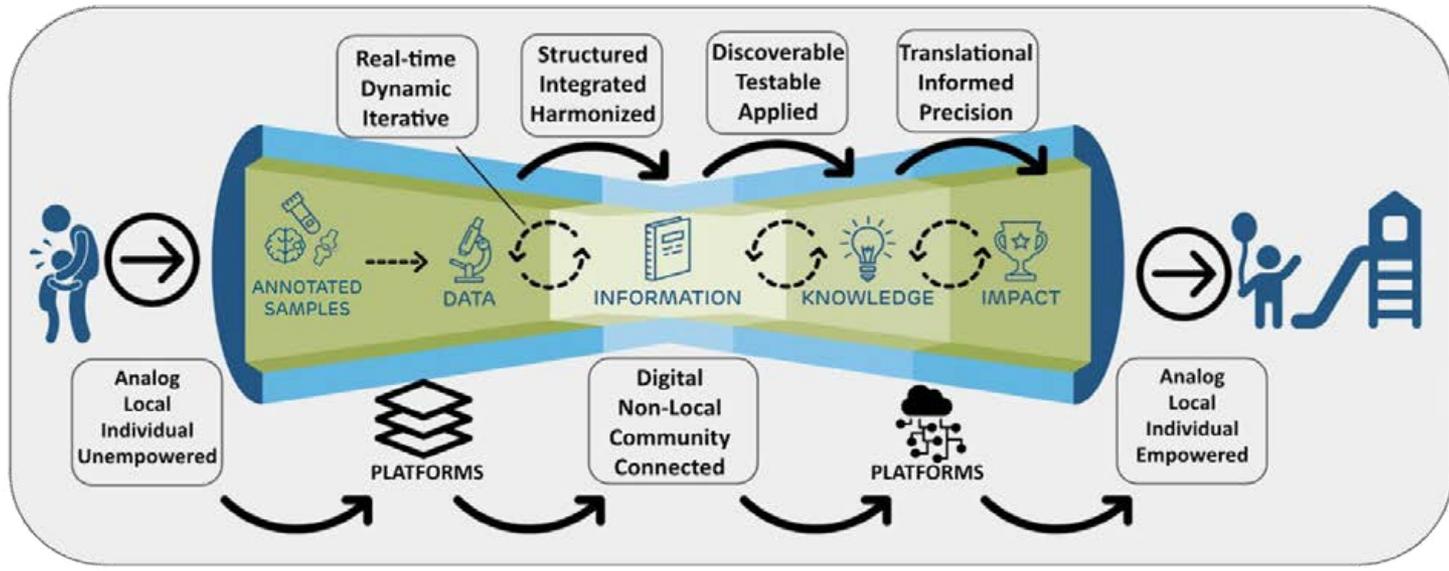


# KFDRC Portal Update



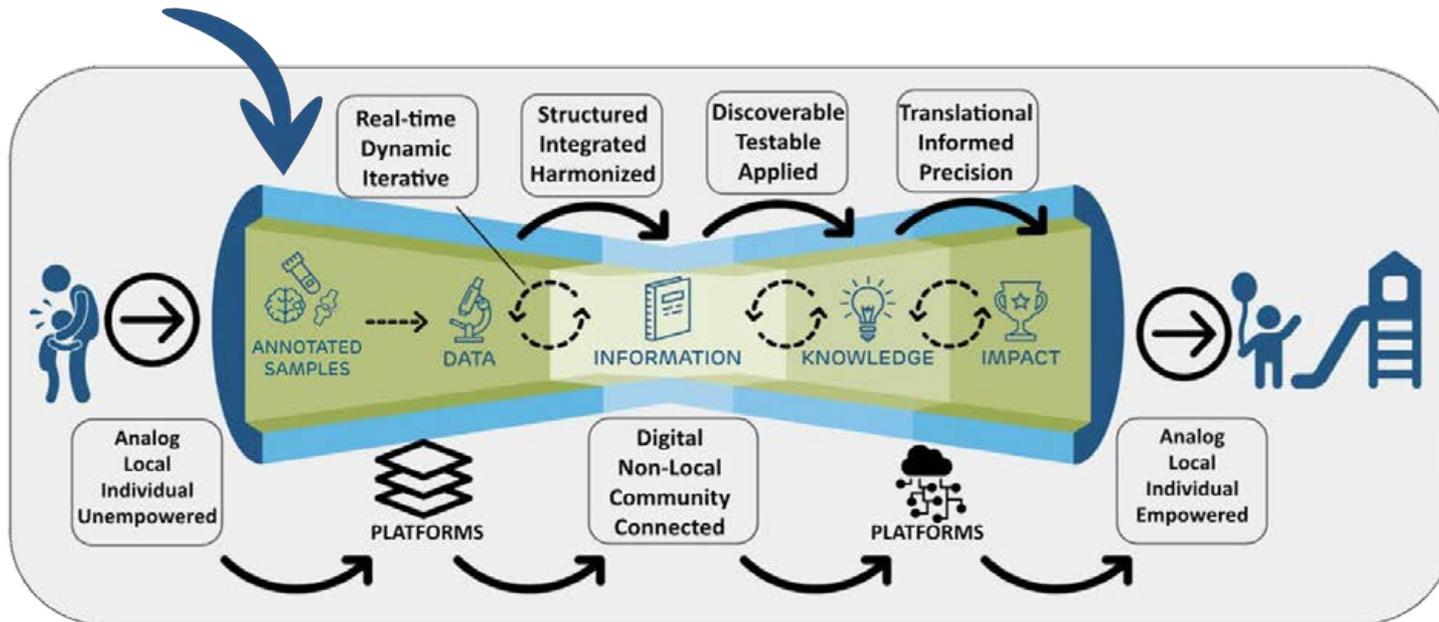


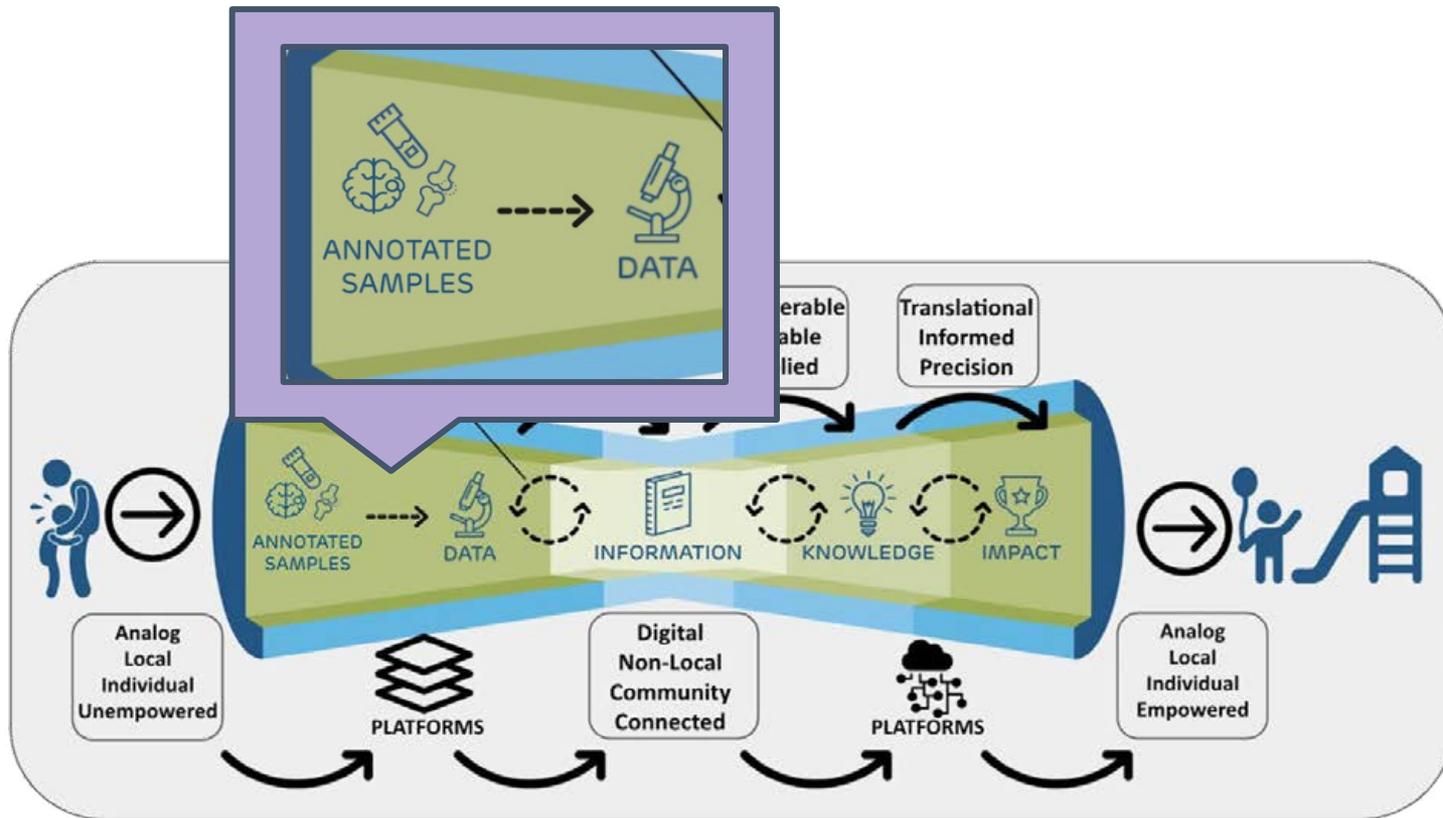
# Methodology





## Double "funnel" model







## Released Studies

Study Name	Funding Year	Category	Participants	Trios	Other Families
Neuroblastoma	FY17	Cancer	1612	463	121
Congenital Diaphragmatic Hernia	FY15 & FY16 & FY17	Structural Birth Defects	2245	714	60
Congenital Heart Defects (PCGC)	FY15 & FY16	Structural Birth Defects	2133	711	n/a
Orofacial Clefts - European Ancestry	FY15	Structural Birth Defects	1295	380	90
Orofacial Clefts - Asian & African Ancestry	FY17	Structural Birth Defects	725	238	6
Novel Cancer Susceptibility (from BASIC3)	FY16	Cancer	291	66	48

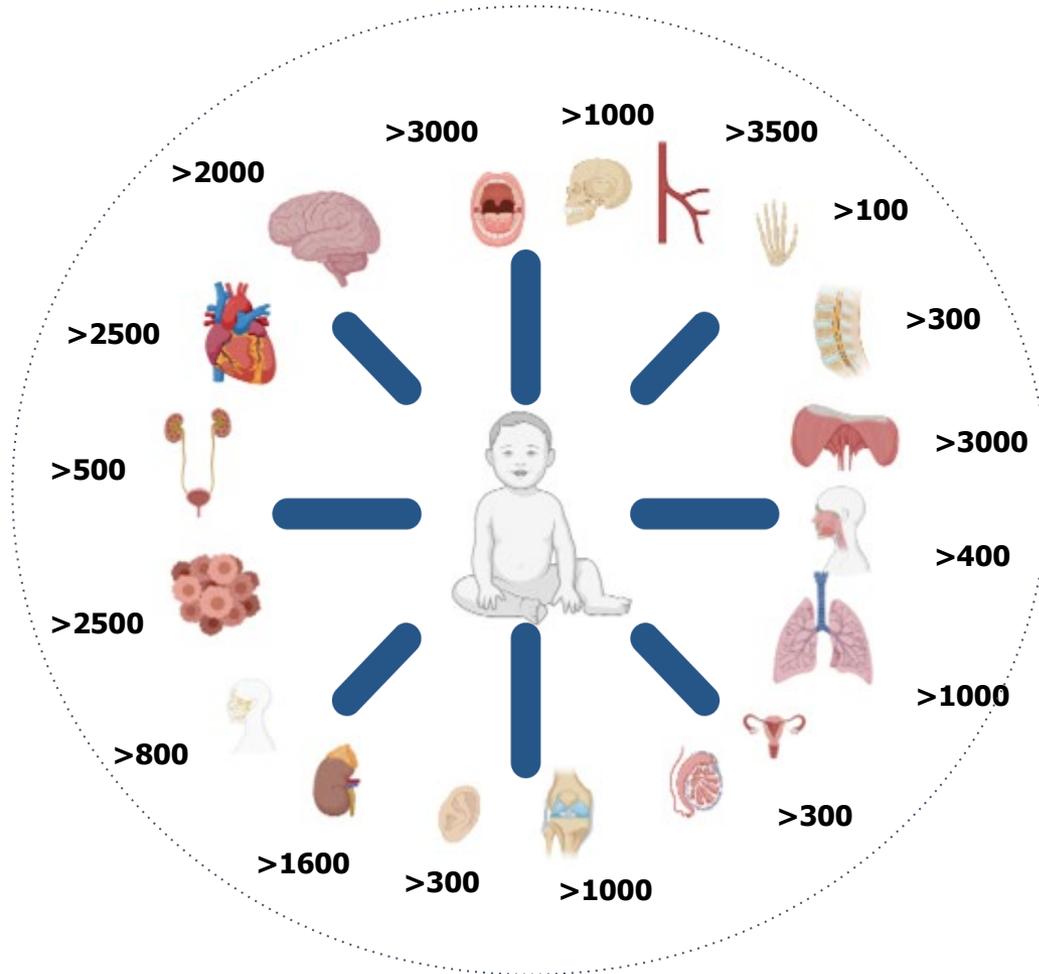


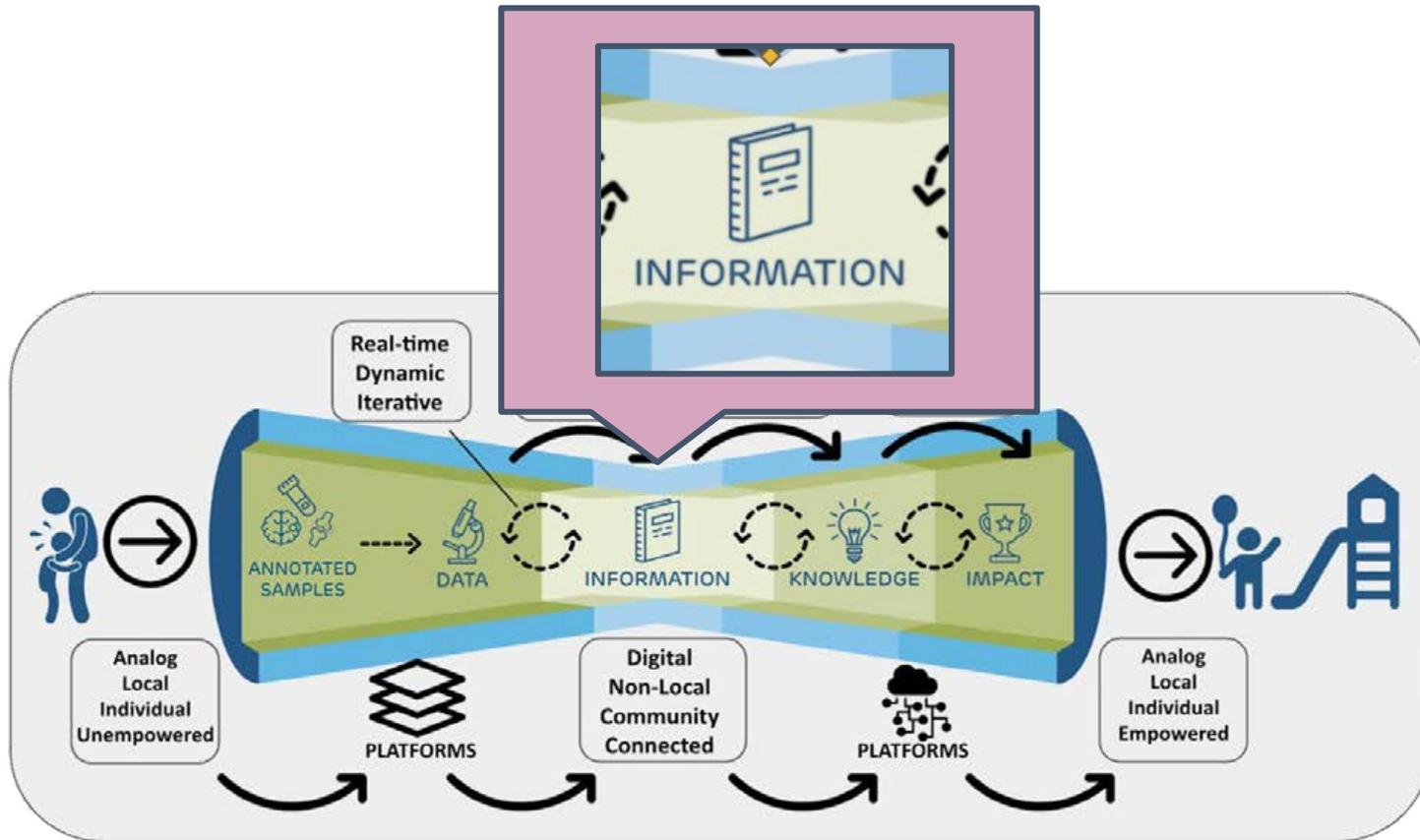
## Released Studies (Cont'd)

Study Name	Funding Year	Category	Participants	Trios	Other Families
Ewing Sarcoma	FY15	Cancer	1047	290	153
Syndromic Cranial Dysinnervation	FY15	Structural Birth Defects	801	248	n/a
Orofacial Cleft - Latin American Ancestry	FY16	Structural Birth Defects	804	262	9
Adolescent Idiopathic Scoliosis	FY16	Structural Birth Defects	262	65	7
Familial Leukemia	FY16	Cancer	365	53	3
Enchondromatoses	FY17	Structural Birth Defects	79	25	2
Disorders of Sex Development	FY15	Structural Birth Defects	300	91	3
Osteosarcoma	FY15	Cancer	84	proband-only study	n/a



# Approaching 30,000 samples released and in the que





# The Kids First Data Resource for Collaborative Discovery

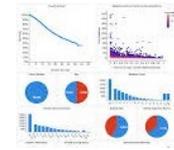
## Data Resource Portal

Entry point. Query, search, discover, build & visualize synthetic cohorts



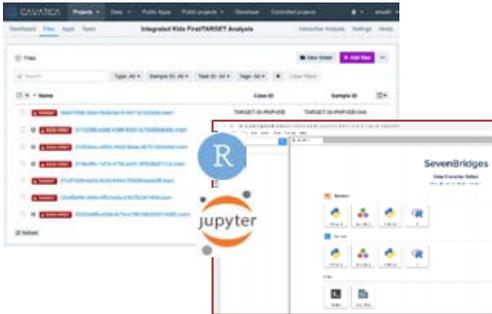
## Knowledge Base Integrations (PedcBioPortal)

Integrations with existing curated/published data visualizations



## Cavatica

Pull data from multiple sources into one workspace. Use notebooks, bring-your-own or use available workflows.



## Data Services

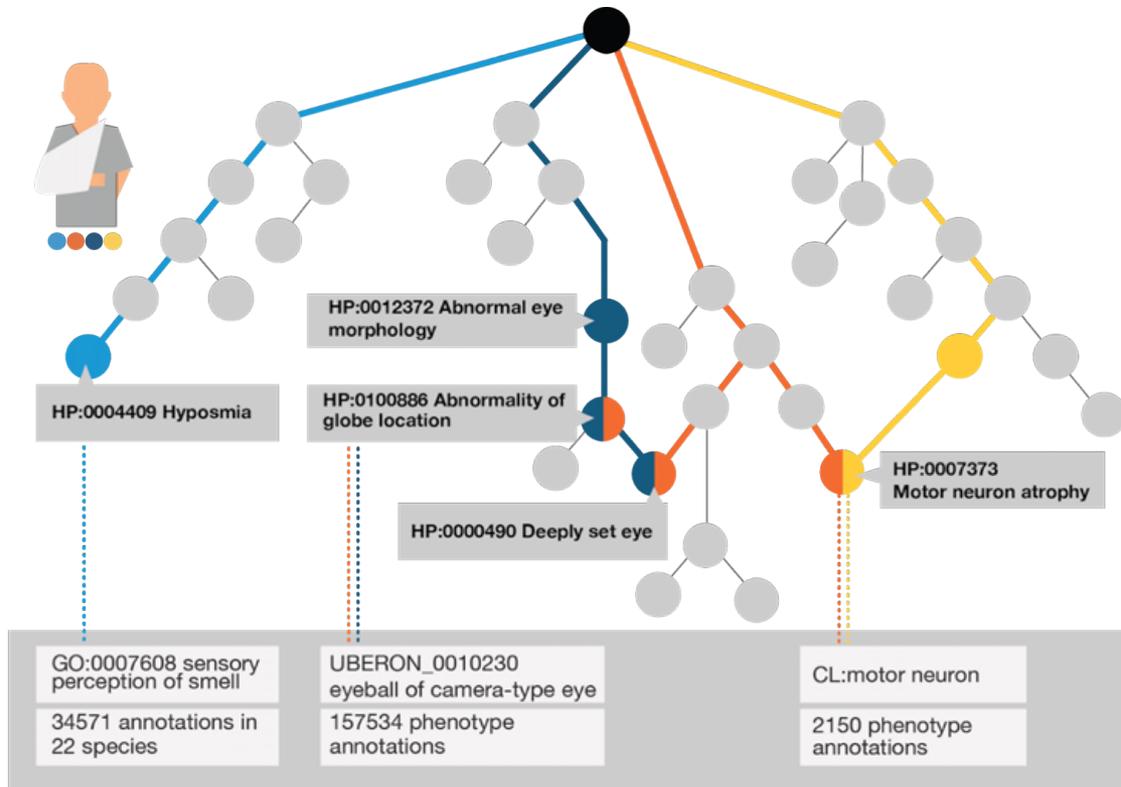
Model clinical data in FHIR-based data services for semantic interoperability and coordination



aws  
STRIDE  
S

GEN3  
DATA COMMONS  
Framework Services  
Index and point to files in the cloud (for approved users)

# Ontologies Enable Computable Models

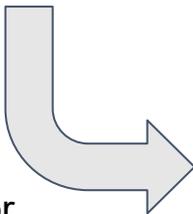


[hpo.jax.org](http://hpo.jax.org)



# Curation of Kids First data sets

2	source_text	hpo_id	hpo_official_text	HPO_ID	HPO_Label	HPO_Modifier_ID	HPO_Modifier_Label
50	Anteriorly Placed Anus	HP:0001545	Anteriorly placed anus				
51	Asplenia	HP:0001746	Asplenia				
52	Assymetric crying facies	HP:0011333	Asymmetric crying face				
53	Asthma	HP:0002099	Asthma				
54	Asymmetric mouth movement; hypoplastic left jaw	HP:0009118	Aplasia/Hypoplasia of the mandible	HP_0000347	Micrognathia	HP_0012835	Left
55	Atresia of left external ear	HP:0000413	Atresia of the external auditory canal			HP_0012835	Left
56	Atresia of the large intestine	HP:0010448	Colonic atresia				
57	Atrial Septal Defect	HP:0001631	Atrial septal defect				
58	Autistic spectrum disorder	HP:0000729	Autistic behavior				
59	Behavior/Mood disorder	None	None		Behavioral		
60	Belpharoptosis	HP:0000508	Ptosis				
61	Bicornuate Uterus	HP:0000813	Bicornuate uterus				

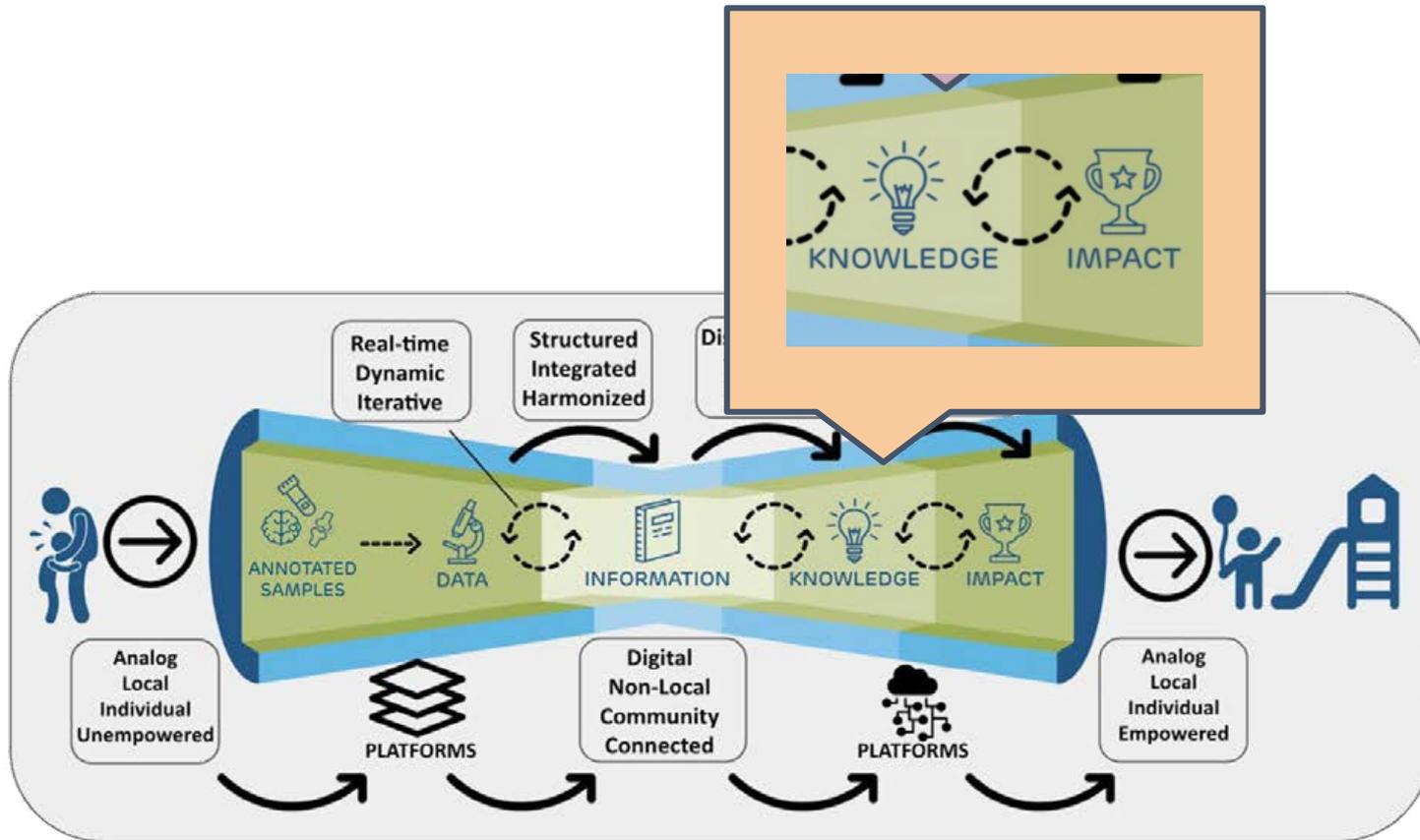


Engage with community for additions and/or recommendations

5 Open ✓ 0 Closed Author ▾ Labels ▾ Projects ▾ Milestones ▾ Assignee ▾ Sort ▾

- ! **NTR: meatus stenosis** Kid's First new term request   
#4180 opened an hour ago by nicolevasilevsky
- ! **NTR: hepatopulmonary fusion** Kid's First new term request   
#4179 opened 2 hours ago by nicolevasilevsky
- ! **NTR: asymmetrical gluteal crease** Kid's First new term request   
#4178 opened 2 hours ago by nicolevasilevsky
- ! **NTR: absent fallopian tube?** Kid's First new term request   
#4176 opened a day ago by nicolevasilevsky
- ! **NTR? wide space btw 4th and 5th toe** Kid's First new term request   
#4175 opened a day ago by nicolevasilevsky





# Variant Workbench

The screenshot shows the Kids First Variant Workbench interface. At the top, there is a navigation bar with the Kids First logo (Pediatric Research Program Data Resource Center) on the left, and navigation links for Dashboard, Explore Data, Variant Workbench (highlighted with a 'New' badge), File Repository, and Members. On the right of the navigation bar, there are links for Resources and a user profile for Allison.

## Kids First Variant Workbench

 [Build a SPARK cluster](#)

**Apache Zeppelin**

The Variant Workbench is a cloud-based analysis platform for querying and analyzing Kids First data.

The Variant Workbench is powered by web-based Zeppelin Notebooks. Using Zeppelin, bioinformaticians can create interactive data analytics and collaborative documents with SQL, Scala, Python, and more.

The Workbench contains germline variant calls and clinical data for probands and families registered in Kids First studies. The same variants found in the harmonized gVCF files provided by the Kids First DRC have been loaded into tables that can be explored using several programming languages.

Additionally, the Workbench is loaded with predicted consequences for these variants on genes. Clinical information and family relationships have also been uploaded, with all of these fields paired by Kids First Participant ID.

By combining clinical and genomic data together in one tool, the Variant Workbench allows users to query across these data types in one cloud-based environment, accelerating research in the field of pediatric cancer and structural birth defects.

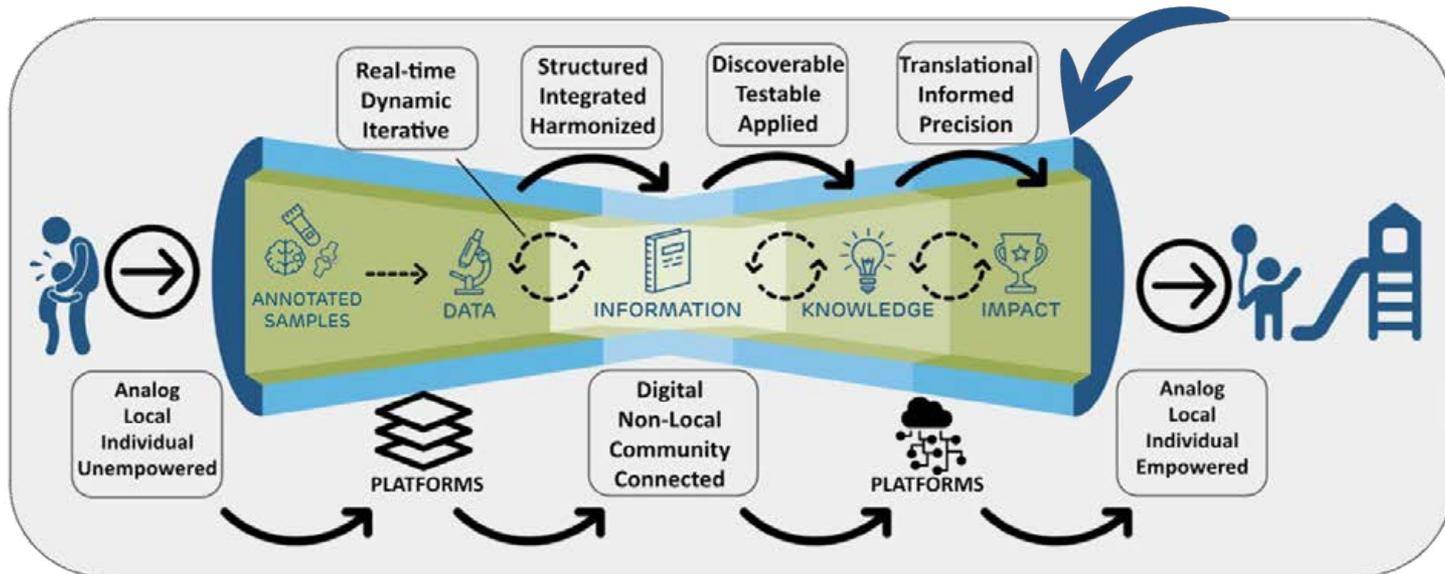
<i>Data Release 1</i>	May 13, 2020
Studies	8
Participants	8,100
Distinct Variants	251,801,242
Occurrences	42,513,213,093

[Feedback](#)



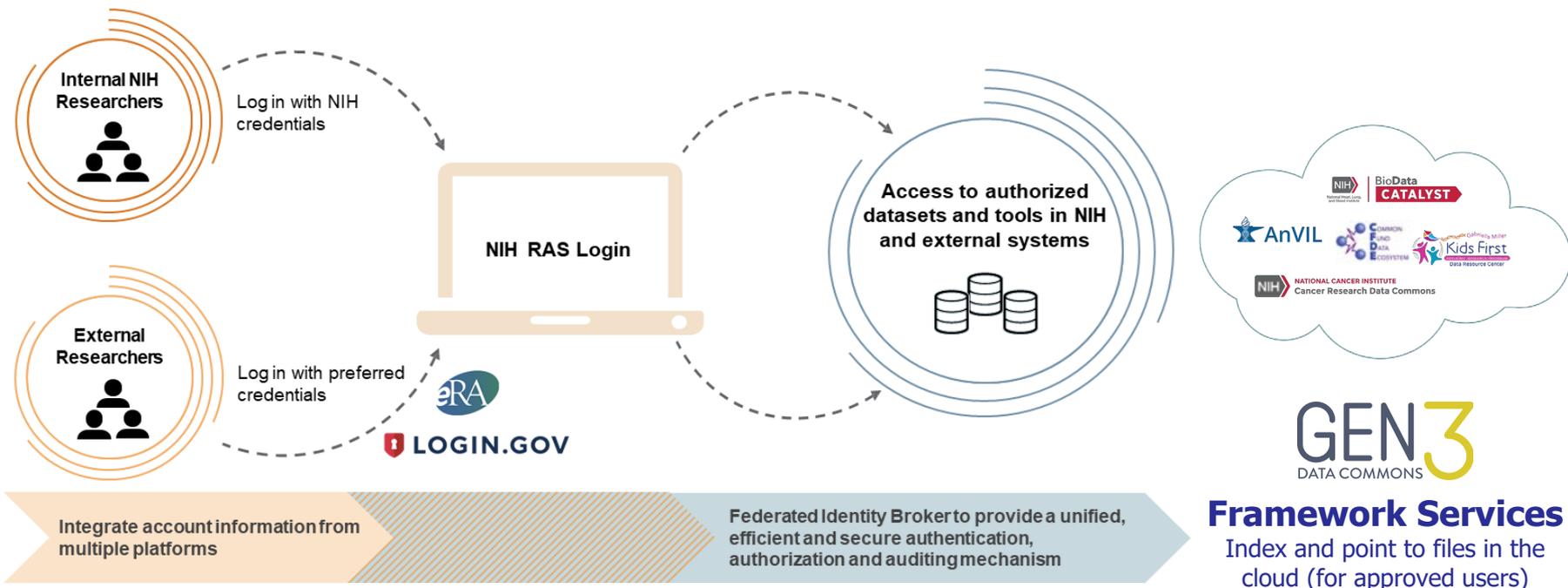
Gabriella Miller  
**Kids First**  
PEDIATRIC RESEARCH PROGRAM  
Data Resource Center

Double "funnel" model



# Interoperability Across NIH Resources

NIH RAS is a unified, efficient, and secure authentication and authorization service deploying in late 2020 provided by NIH CIT that enables streamlined researcher access to NIH-funded data assets and data repositories across multiple systems and provides standardized methods of logging and auditing such access.





# Kids First DRC Team





# Kids First DRC Team





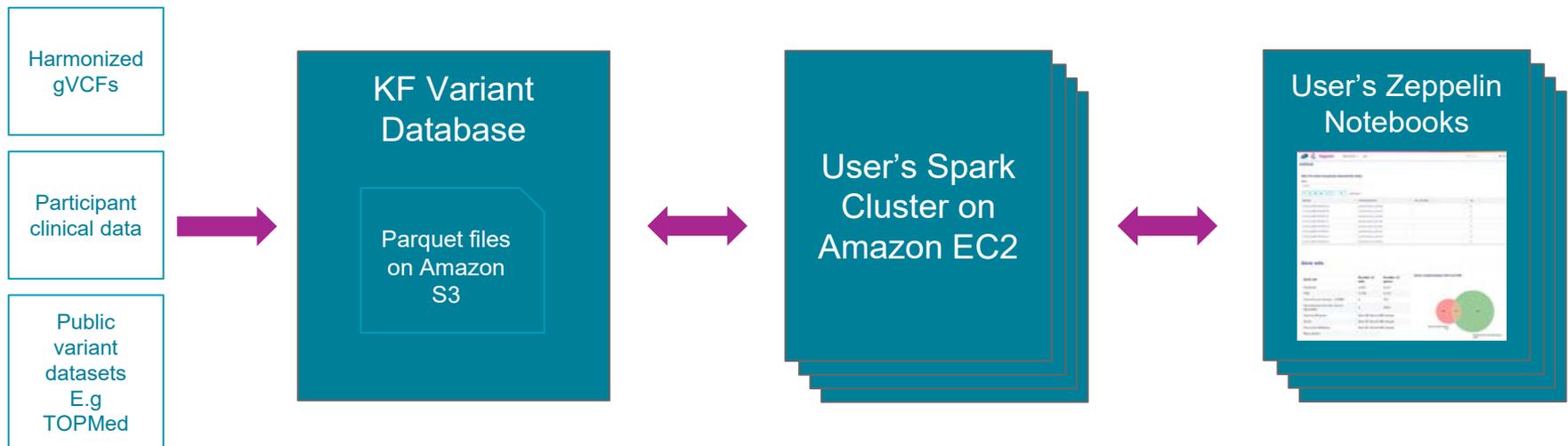
# The KF Germline Variant Workbench Project

- Prototype presented at the last webinar in May
- Productionizing software and infrastructure
  - Security, data quality, robustness, bioinformatics, cluster management and software deployment tools
- Objective
  - Extract, annotate and import variant data from harmonized gVCF files to a *queryable* database
- Challenges
  - Complex data, big data, confidentiality and security
- Production version to be released
  - Phase I: X01 investigator's teams, Oct-Nov 2020
  - Phase II: All Kids First Portal members, Jan 2021



## Two main components

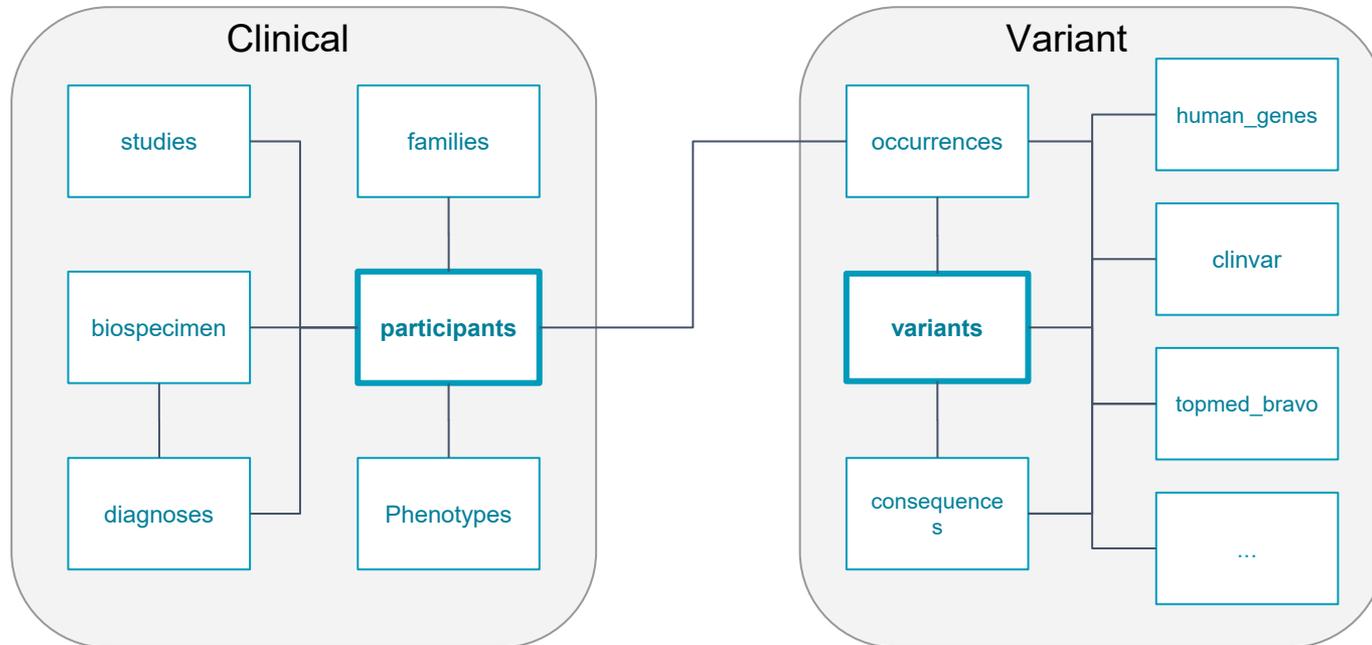
- An annotated variant database
- User private compute environment with Spark and Zeppelin





# Variant Database

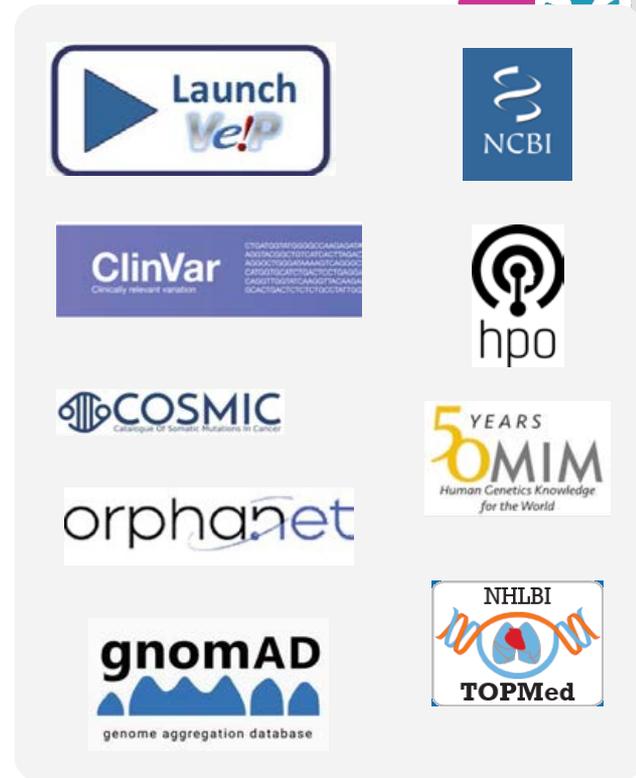
- Data is organized into tables like in a traditional relational database
- 31 tables in current version





# Current Annotations

- From VEP and dbsnfp
  - AA change, genes, transcripts,...
  - > 360 different annotations (e.g. predictive functional scores, gtex)
- Gene ids and alias from the NCBI
- Clinvar, dbSnp
- Gene panels
  - Cosmic, HPO, Omim, Orphanet
- Populational databases
  - TOPMed, gnomAD 2.1 and 3.0, 1000 genomes
- KF annotations
  - Allele frequencies, parent's genotypes, zygosity, QC metrics



Example: Find all **de-novo rare missense** variants with **high functional impact** in **low grade glioma** patients affected by any **cardiovascular abnormalities**



# Data Currently Loaded for Initial Beta Phase

-  **11** studies
-  **11 thousand** participants
-  **297 million** unique variants
-  **53 billion** occurrences

- Each variant occurs on average in 178 participants
- Each participant has on average 4.8 million variants
- 52% (153M) of the KF variants are in TOPMed and/or gnomAD

# The Zeppelin Data Analytic Environment

**Provides programmatic access to the variant database from web browsers**

- Accessible from the Portal
- Powered by individual Spark clusters on AWS
- User notebook workspace for creating and managing notebooks
- Users can only access variant datasets they have been authorized to

The image shows two overlapping screenshots. The top screenshot is the 'Kids First Germline Variant Database' portal. It features a navigation bar with 'Dashboard', 'Explore Data', 'Variant DB', 'File repository', 'Members', 'Resources', and a user profile 'Lucas'. The main content area includes the Apache Zeppelin logo and a description: 'Kids First is providing members with their own SPARK cluster running a web-based Zeppelin notebooks dashboard to explore, query and visualize its germline variant datasets. Using Zeppelin, bioinformaticians can create interactive data analytics and collaborative documents with SQL, Scala, Python, and more.' A 'Data Release 1' summary table is also present:

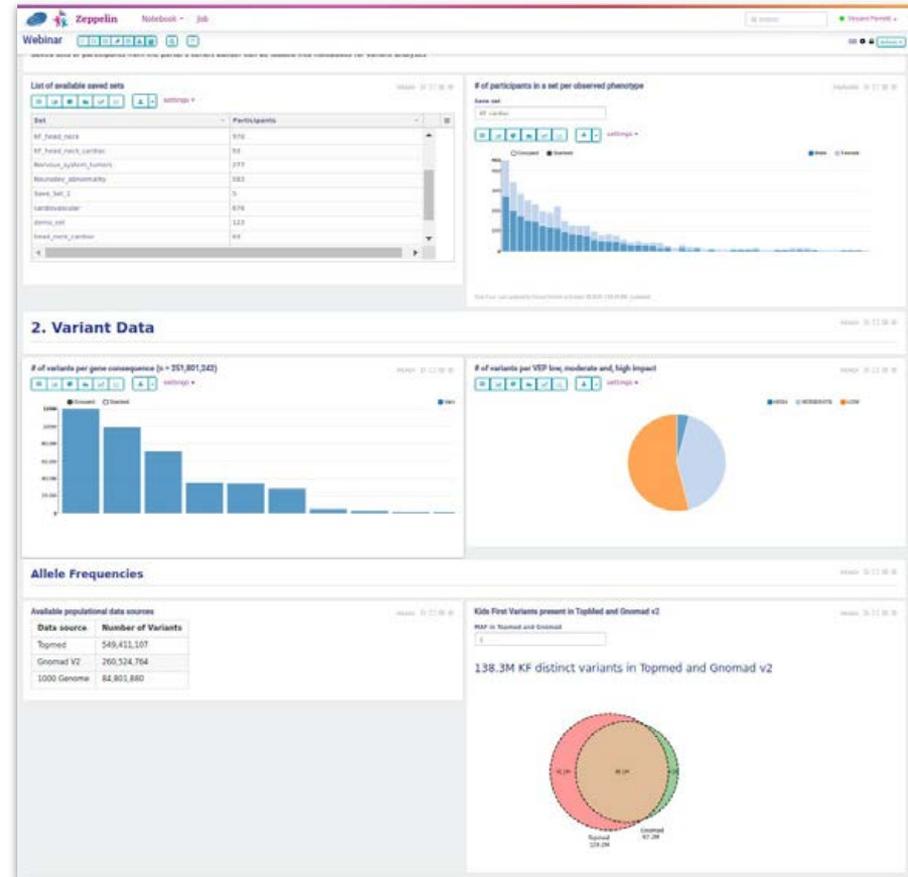
Category	Value
Studies	11
Participants	9,518
Unique Variants	300,976,211
Occurrences	70,864,456,268

The bottom screenshot shows the Zeppelin notebook interface. It displays a 'Welcome to the Kids First Germline Variant Database!' message and a sidebar with a 'Notebook' list containing items like 'jeremy', 'vincent', 'DemoBook', 'emma\_test', 'example\_variant\_by\_gene\_participant', 'frequencies', 'heatmap', 'patate', 'Untitled Note 1', 'Untitled Note 2', 'webinar', and 'webinar\_emma'. A 'Help' section and a 'Community' section are also visible.



# Zeppelin Notebooks

- Support for various programmatic languages (SQL, Python, R, Scala, Markdown)
- Can use different languages in the same Notebook
- Built-in chart plotting capability
- Interactive forms
- Multi-user capability
- Pre-installed libraries
  - **Hail**, Pandas, Numpy, Matplotlib, Plotly, Seaborn





# Cohort builder & variant database integration

Virtual cohorts built using the Portal cohort builder can be saved as *Participant sets*

Study Demographic Clinical Biospecimens Available Data Upload IDs

CLEAR ALL

Observed Phenotype (HPO) is any of Abnormality of he... 71

AND Diagnosis (Mondo) is any of tetralogy of fallot... 64

Query #1 71 Participants 71 Families 264 Files

Save participants set Download

Observed Phenotypes

8448 Phenotypic abnormality (HP:0000118)

Studies

Probands Other Participants

Kids First: Orofacial Cle... Kids First: Orofacial Cle... Kids First: Orofacial Cle... Kids First: Syndromic Cra... Kids First: Congenital He... Pediatric Brain Tumor AEL...

Name	Count
Cancer	2995
Cleft lip - Trio	856
Ewing sarcoma - Male	207
Lymphatic System	27
Neuroblastoma	1681
WGS - Tumor Nervous System	9

User dashboard

Save participants set

- + Save as new set
- ⌘ Add to existing set
- ⌘ Remove from existing set



# User's saved sets as part of the variant database

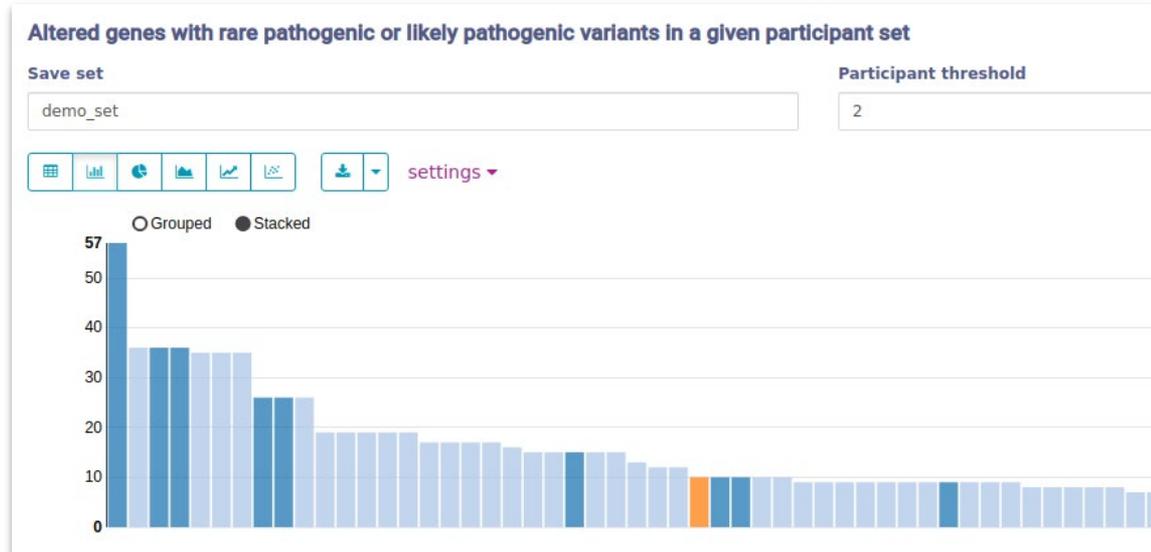
**List of available saved sets**

```
%sql
select tag as Set, count(*) as Participants from saved_sets group by tag order by tag
```

Set	Participants
Cranial_Dysinnervation_Musculature	103
Cranial_Dysinnervation_Musculature_Cardiac	123
Head_neck_Abnormalities	643
HeadandNeckandCardiac	93
KF_cardiac	643
KF_head_neck	970
KF_head_neck_cardiac	93
Nervous_system_tumors	277
Neurodev_abnormality	583

Variant workbench

Saved sets can be joined with other variant tables for genomic analyses of virtual cohorts





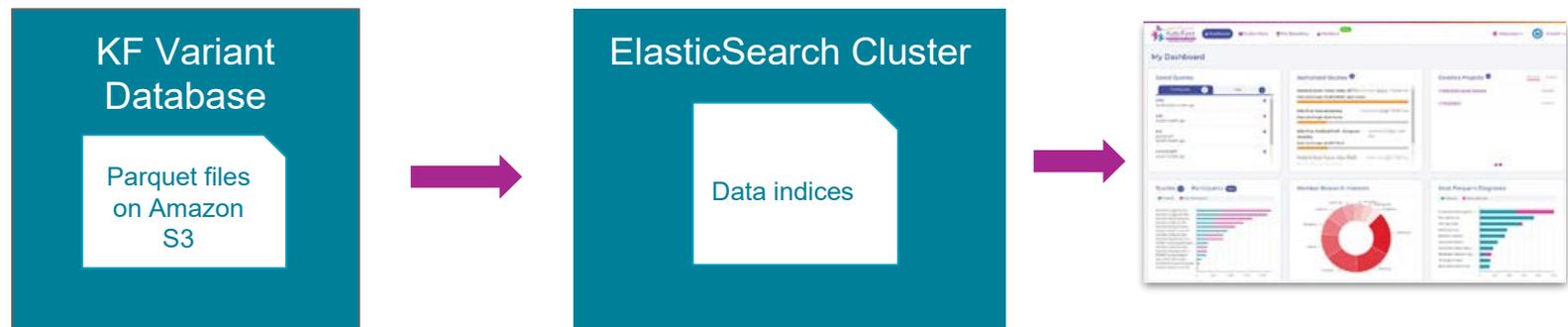
# Who is the variant workbench for?

- Initially:
  - KF DRC for data quality control
  - Bioinformaticians who know programming in SQL, Python, R, etc.
- **Next Aim:** Increase the user base by offering a series of interactive and comprehensive notebooks directly accessible from the portal
  - e.g. a gene-centric analysis notebook
- Users will be able to
  - input analysis parameters (Zeppelin forms)
  - import their own data to their notebooks (e.g. gene panels, variant datasets)
  - modify notebooks as they need and save them in their workspace
- A good tradeoff between standard data portal static analyses and the complexity of programming notebooks from scratch



## Further work

- Indexing the variant database with Elasticsearch
- Building data querying interfaces within the portal (integrating notebook use cases)
  - GA4GH-like Beacon service that returns yes/no answers on variant occurrences
  - GnomAD-like interface that returns aggregations e.g. allele frequencies
  - Integration with the Cohort Builder enabling complex queries with *both* clinical and genomic criteria
- Adding more annotations supporting variant prioritization





# Special Thanks To

## CHU Ste-Justine

- **Jeremy Costanza**, Lead software architect and developer
- Developers
  - Adrian Paul, Evans Girard, Francis Lavoie
- UX
  - Lucas Lemonnier

## CHOP

- DevOps
  - **Alex Lubneuski**
- Bioinformatics
  - **Yuankun Zhu**, Yiran Guo, Miguel Brown, David Higgins



# COMMON FUND DATA ECOSYSTEM

AN OVERVIEW

OCTOBER 6<sup>TH</sup>, 2020

C. TITUS BROWN, UC DAVIS

[nih-cfde.org](https://nih-cfde.org)



CFDE - CC  
Owen White, PI

UMB  
Owen White, PI



USC/ISI  
Carl Kesselman, PI



UCHI  
Ian Foster, PI



RTI  
Becky Boyles, PI



UMB  
Bob Carter, PM



Oxford  
Susanna Sansone, PI



ICAHN/Mt. Sinai  
Avi Ma'ayan, PI



UCD  
Titus Brown, PI



UCD  
Amanda Charbonneau, PM





## MISSION: INCREASING DATA REUSE

- Enhance
  - Reuse, and comparison to NIH biomedical resources
  - Access to current and sunsetted programs
  - Application of tools / systems across (CF) programs
  - Ability to ask scientific questions across (CF) data sets
- Increase the on-ramp for more researchers with training

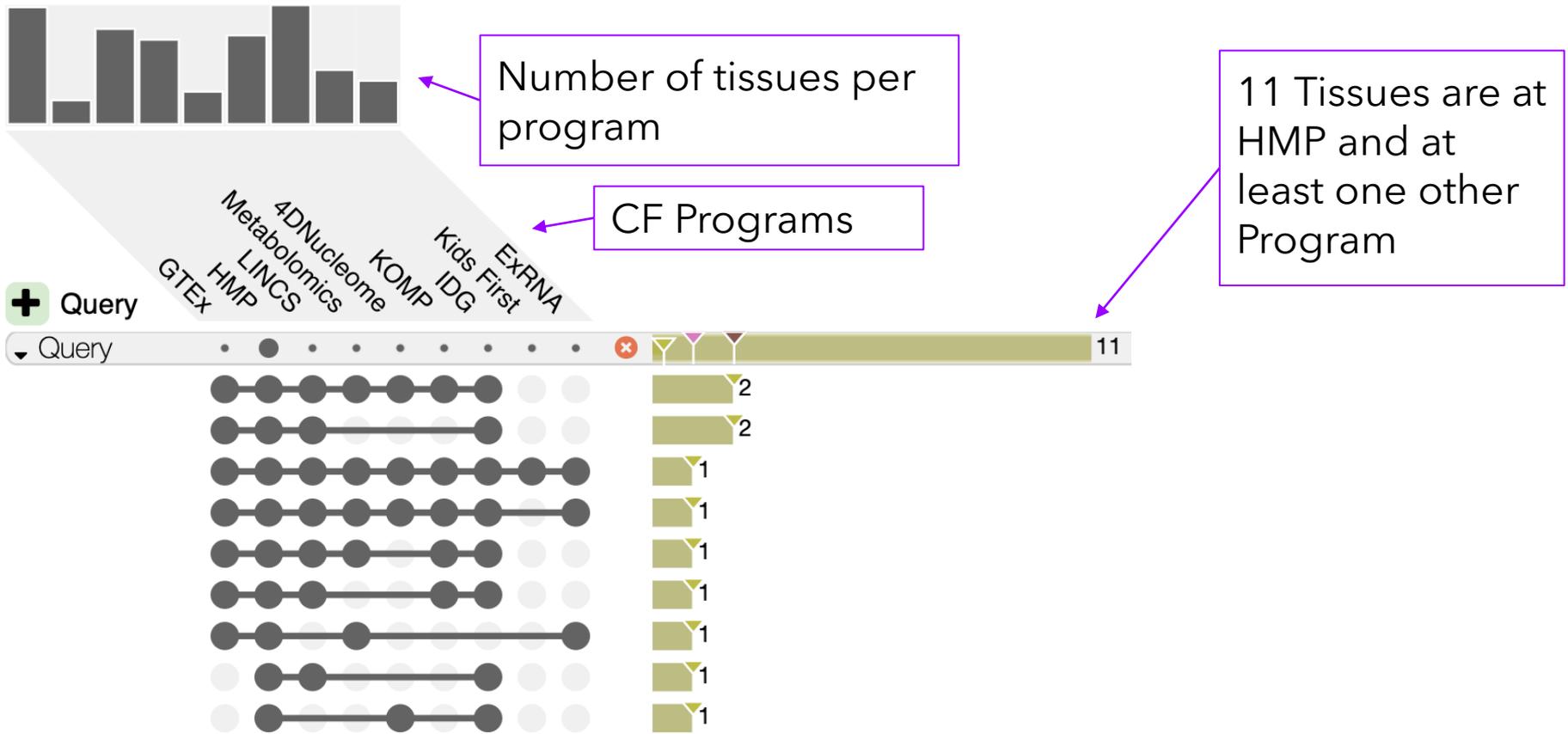
# WHY: COMMON FUND DATA ECOSYSTEM DIVERSE PORTFOLIO

A.		4D Nucleome GTEx HMP / iHMP HuBMAP Kids First LINCS Metabolomics MoTrPAC SPARC								
		4D Nucleome	GTEx	HMP / iHMP	HuBMAP	Kids First	LINCS	Metabolomics	MoTrPAC	SPARC
Clinical Data		X	X		X	X				
Whole Genome/Exome Sequence		X	X		X			P		
Transcriptomics	X	X	X	P	X	X		P	P	
Histology Images					X					
Radiology Images					X					
Metatranscriptomics			X					P		
Metaproteomics			X							
Marker Sequence Metagenomics			X					P		
Microbial Reference Genomes			X					P		
ChIPseq	X					X				
FISH	X			P						
ATACseq	X			P		X				
Hi-C	X									
ChIA-PET	X									
Proteomics			X	P		X		P	P	
KINOMEScan						X				
Metabolomics			X	P			X	P		
Lipidomics				P						
scDNAseq				P						
Epigenomics			X	P		X		P		

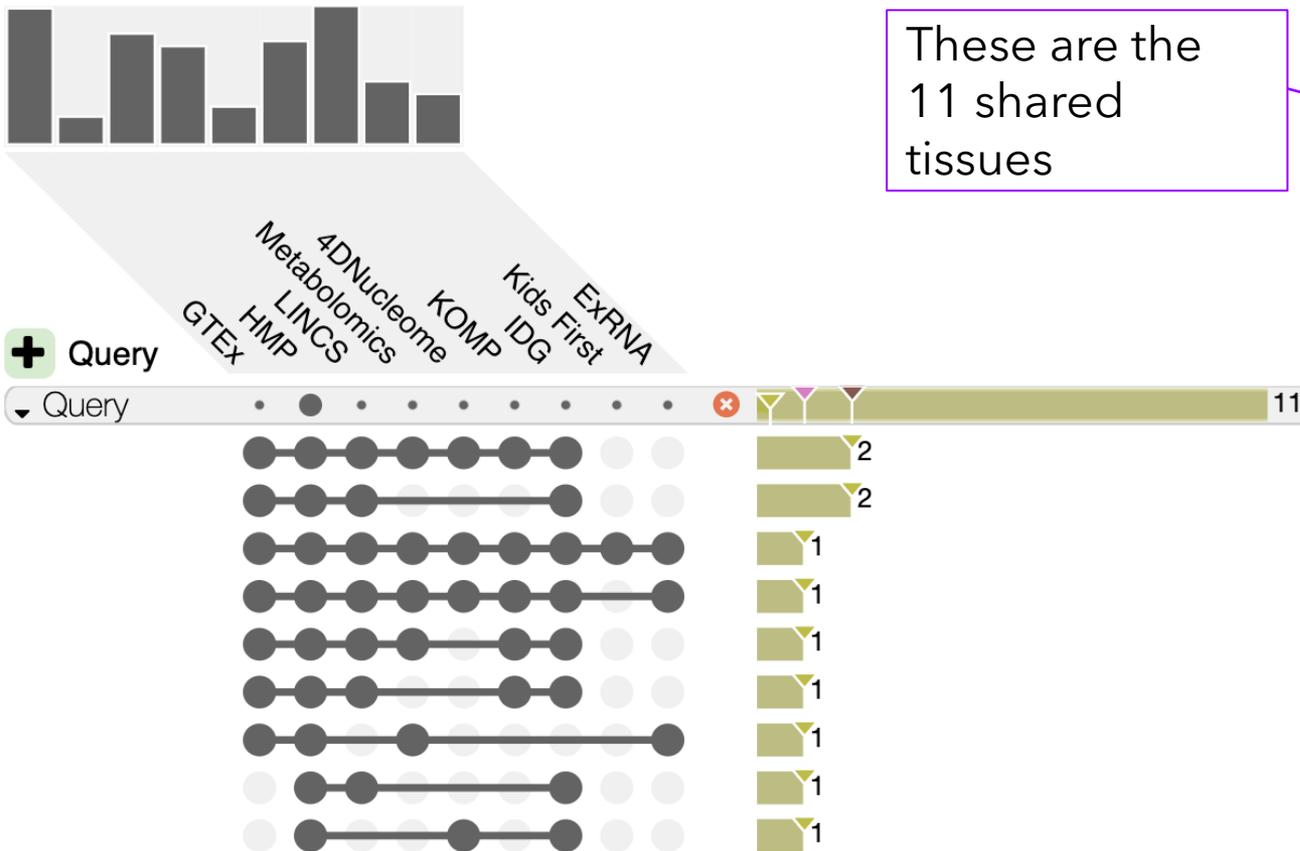
  

B.		Systems Organs Cells Molecules			
		Systems	Organs	Cells	Molecules
MoTrPAC	X	X			
SPARC	X	X			
HubMap		X	X		
LINCS			X	X	
4D Nucleome			X	X	
GTEx				X	
KidsFirst				X	
HMP/iHMP				X	
Metabolomics				X	

# TISSUE TYPES SHARED BY HMP AND OTHER CF PROGRAMS



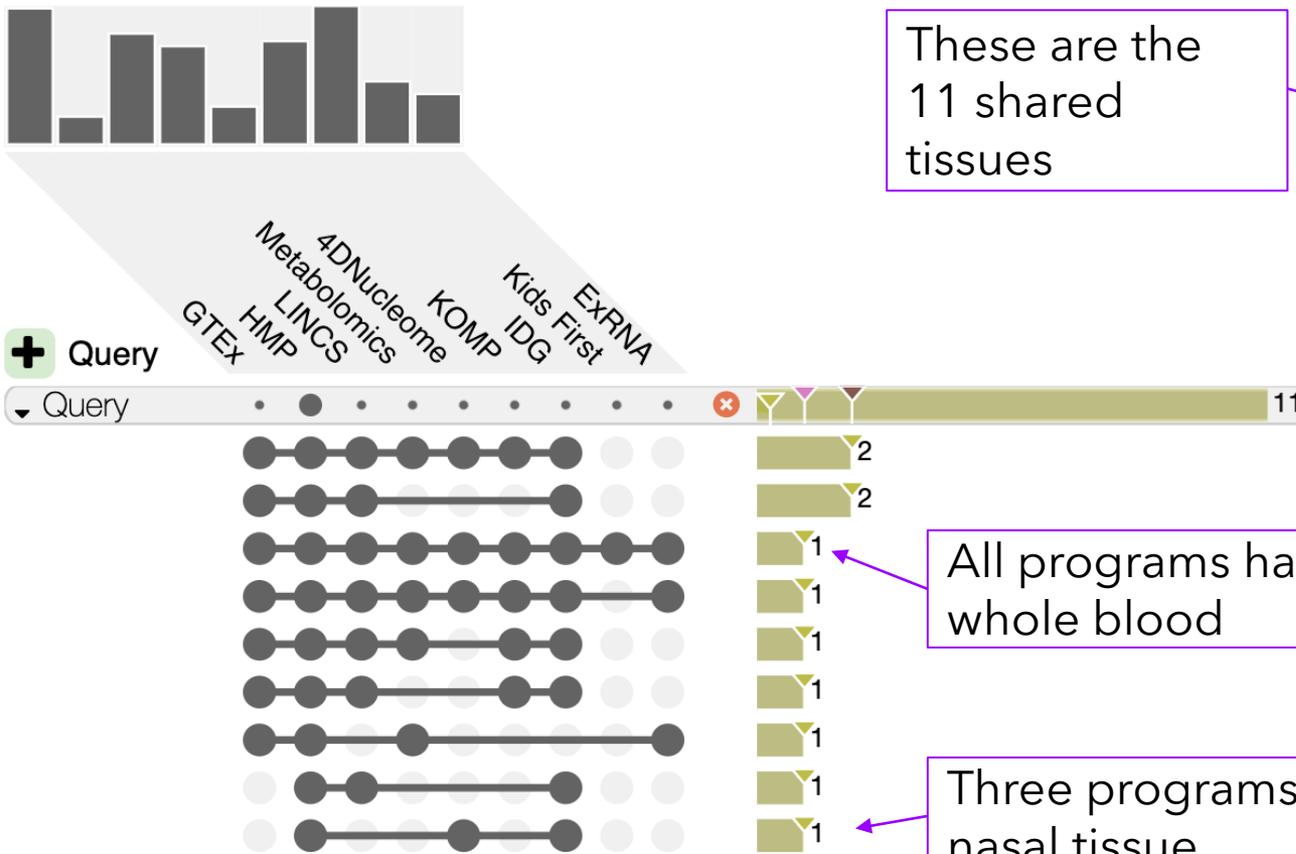
# TISSUE TYPES SHARED BY HMP AND OTHER CF PROGRAMS



Query Results

Tissue	Set Count
Cervix - Ectocervix	4
Colon - Sigmoid	9
Colon - Transverse	10
Esophagus - Mucosa	5
Minor Salivary Gland	5
Nasal	3
Rectum	3
Skin - Not Sun Exposed (Suprapubic)	7
Skin - Sun Exposed (Lower leg)	6
Vagina	4
Whole Blood	11

# TISSUE TYPES SHARED BY HMP AND OTHER CF PROGRAMS



These are the 11 shared tissues

Query Results	
Tissue	Set Count
Cervix - Ectocervix	4
Colon - Sigmoid	9
Colon - Transverse	10
Esophagus - Mucosa	5
Minor Salivary Gland	5
Nasal	3
Rectum	3
Skin - Not Sun Exposed (Suprapubic)	7
Skin - Sun Exposed (Lower leg)	6
Vagina	4
Whole Blood	11

All programs have whole blood

Three programs have nasal tissue

## CFDE INTEGRATION: HOW?

### THE CROSSCUT METADATA MODEL (C2M2)

**Goal:** DCCs to share structured, detailed **metadata** about their **experimental resources** across the ecosystem.

**Not** a warehouse

**No** data replication

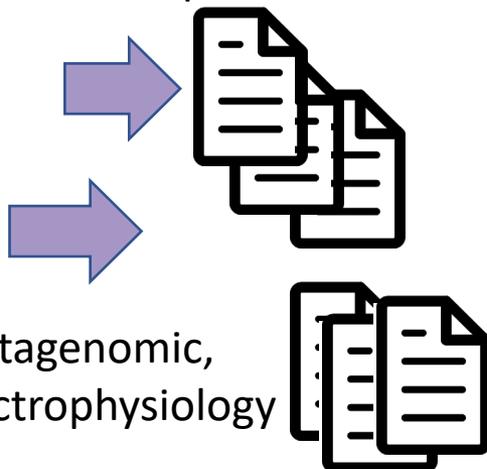
Users directed to **DCCs as primary resource**



NIH Human Microbiome Project



RNAseq, Variant files



Metagenomic,  
Electrophysiology

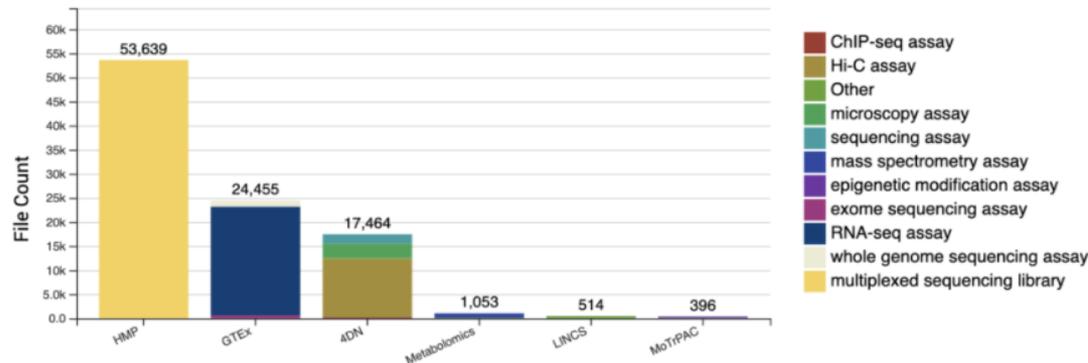
Metadata ONLY

- File type
- Organism
- Assay
- Patient information

CFDE Catalog



# Common Fund Data Ecosystem Web Portal



## Publications

[FAIRshake: Toolkit to Evaluate the FAIRness of Research Digital Resources](#)

[An Open Ecosystem for Pervasive Use of Persistent Identifiers](#)

## News & Events

[Fall 2020 Cross Pollination Series starts 9/29/2020. Click for more info..](#)

Data portal website coming in November 2020.

The NIH Common Fund has switched to rolling submission of Engagement Plans, effective May 2020. More information available on the [Engagement page](#).

- [Technical Documentation](#)
- [Use Case Library](#)
- [Training](#)
- [Tools](#)

## Tweets by @CFdeNIH

NIH\_CFDE Retweeted

**Lappalainen Lab**  
@tuuliel\_lab

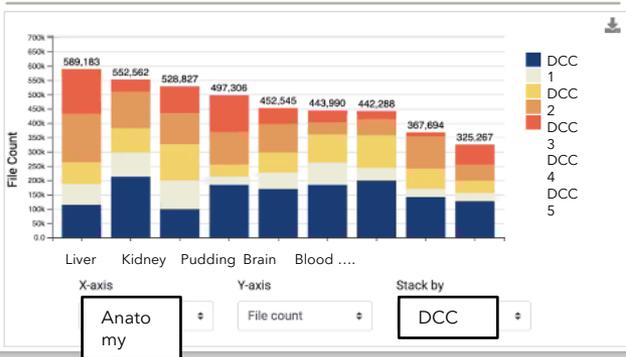
Very happy and proud to present: The GTEx papers. This set of 15 papers published today describes the final phase of this 10-year effort, providing the genomics community an atlas of genetic regulatory variants and a deep dive into the biology behind it. [sciencemag.org/collections/ge...](https://sciencemag.org/collections/ge...)



[Embed](#) [View on Twitter](#)

<http://nih-cfde.org>

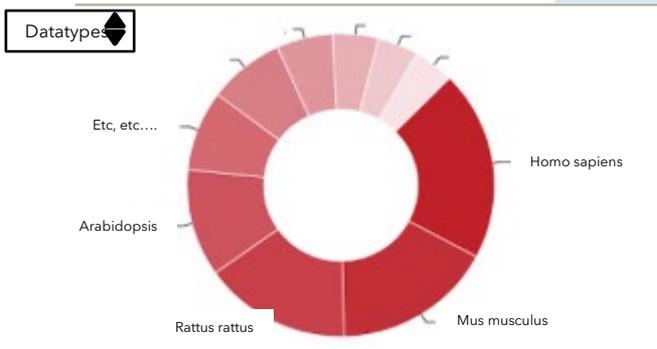
## Anatomy by CF Program



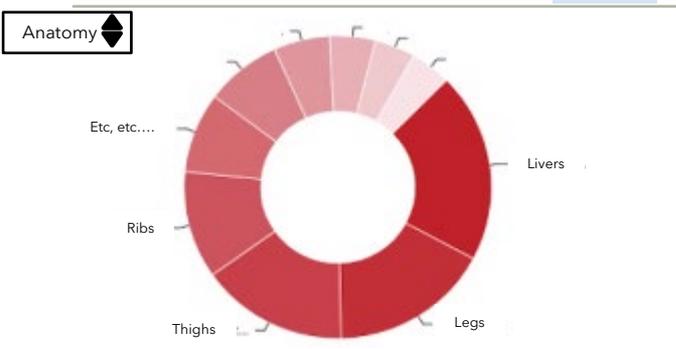
## Assays at Metabolomics



## RNAseq by Species across Common Fund



## Anatomy by CF Program



## CFDE Portal for cross data set search - public beta at end of October!!

Collection<sup>®</sup> Export Permalink

Search all columns

Refine search Hide panel Displaying first 25 of 2,990 records

Collection Creation Time <sup>Ⓞ</sup>

All records with value <sup>Ⓞ</sup>

**From:**

Date: 2011-07-14

Time: 20:00:00

**To:**

Date: 2020-06-11

Time: 20:00:00

**> File Creation Time <sup>Ⓞ</sup>**

**> Biosample Creation Time <sup>Ⓞ</sup>**

**> Defined By Project <sup>Ⓞ</sup>**

**> Subject Granularity <sup>Ⓞ</sup>**

**> Subject Role <sup>Ⓞ</sup>**

**> Subject Taxonomy <sup>Ⓞ</sup>**

**> Anatomy <sup>Ⓞ</sup>**

**> Assay Type <sup>Ⓞ</sup>**

**> Part of Collection <sup>Ⓞ</sup>**

View <sup>Ⓞ</sup>	ID Namespace <sup>Ⓞ</sup>	ID <sup>Ⓞ</sup>	Name <sup>Ⓞ</sup>	Description <sup>Ⓞ</sup>	Creation Time <sup>Ⓞ</sup>
	4D Nucleome	/experiment-set-replicates/4DNESENTEIC/	pADamID on RPE - 4DNESENTEIC	pADamID for LaminB1 on RPE wildtype cells	2020-05-28 11:36:17
	4D Nucleome	/experiment-set-replicates/4DNESG2OY6X6/	pADamID on RPE - 4DNESG2OY6X6	Dam control for accessibility and amplification biases on RPE wildtype cells	2020-05-28 11:36:16
	4D Nucleome	/experiment-set-replicates/4DNESXGXZEZ6/	Non-enriched ChIA-Drop on GM19239 - 4DNESXGXZEZ6	Replicates of non-enriched ChIA-Drop on GM19239 cells.	2020-05-22 01:48:00
	4D Nucleome	/experiment-set-replicates/4DNESF829JOW/	Single nucleus Hi-C on WTC-11 with GFP tagged AAVS1 - 4DNESF829JOW	Single nucleus Hi-C on Modified WTC-11 (GM25236) with GFP tagged AAVS1 locus - clone 28, select nucleus from G1 phase only	2020-04-24 03:40:39
	4D Nucleome	/experiment-set-replicates/4DNESJQ4RXY5/	Single nucleus Hi-C on WTC-11 with GFP tagged AAVS1 - 4DNESJQ4RXY5	Single nucleus Hi-C on Modified WTC-11 (GM25236) with GFP tagged AAVS1 locus - clone 6, select nucleus from G1 phase only	2020-04-24 03:40:38
	4D Nucleome	/experiment-set-replicates/4DNESW1G42GW/	ChIP-seq on G1E-ER4 - 4DNESW1G42GW	Biological replicates of PoII ChIP-seq on G1E-ER4 cells in late G1 phase	2020-03-03 15:08:16
	4D Nucleome	/experiment-set-replicates/4DNESY2UQH5B/	ChIP-seq on G1E-ER4 - 4DNESY2UQH5B	Biological replicates of PoII ChIP-seq on G1E-ER4 cells in mid-G1 phase	2020-03-03 15:08:15
	4D Nucleome	/experiment-set-replicates/4DNESHA64KOK/	ChIP-seq on G1E-ER4 - 4DNESHA64KOK	Biological replicates of PoII ChIP-seq on G1E-ER4 cells in early G1 phase	2020-03-03 15:08:14
	4D Nucleome	/experiment-set-replicates/4DNESWCWOS1Y/	ChIP-seq on G1E-ER4 - 4DNESWCWOS1Y	Biological replicates of PoII ChIP-seq on G1E-ER4 cells in anaphase/teIophase	2020-03-03 15:08:13
	4D Nucleome	/experiment-set-replicates/4DNESWQLLRLT/	ChIP-seq on G1E-ER4 - 4DNESWQLLRLT	Biological replicates of PoII ChIP-seq on G1E-ER4 cells in prometaphase	2020-03-03 15:08:12
	4D Nucleome	/experiment-set-replicates/4DNESRT3AFT8/	ChIP-seq on G1E-ER4 - 4DNESRT3AFT8	ChIP-seq input only control from G1E-ER4 cells in late G1 phase	2020-03-03 15:08:11
	4D Nucleome	/experiment-set-replicates/4DNESER4SPJ6/	ChIP-seq on G1E-ER4 - 4DNESER4SPJ6	ChIP-seq input only control from G1E-ER4 cells in mid-G1 phase	2020-03-03 15:08:10
	4D Nucleome	/experiment-set-	ChIP-seq on G1E-ER4 -	ChIP-seq input only control on G1E-	2020-03-03 15:08:09

## MISSION: INCREASING DATA REUSE

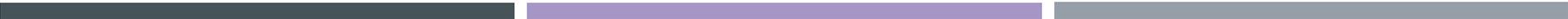
- Enhance
  - Reuse, and comparison to NIH biomedical resources
  - Access current and sunsetted programs
  - Application of your tools / systems
  - Ability to ask scientific questions across data sets

■ Increase the on-ramp for more researchers with training



## A CRITICAL TOOL: USER TRAINING

- Using DCC-specific resources
- Finding, accessing and combining resources between DCCs
- Analyzing data sets in the cloud
- Data-type specific training, e.g. WGS for clinicians
- Modalities include Web sites, videos, teleconference training, and (eventually) in-person.



## A CRITICAL TOOL: USER TRAINING **FOR KIDS FIRST:**

We are planning tutorials and training on:

- Logging into and using the KF DRC portal to find data sets
- Running data analyses in the Cavatica cloud platform
  - WGS?
  - RNAseq?
- Data-type specific training, e.g. WGS for clinicians
- Modalities include Web sites, videos, teleconference training, and (eventually...) in-person.

## A CRITICAL TOOL: USER TRAINING **FOR KIDS FIRST:**

We are planning tutorials and training on:

- Logging into and using the KF DRC portal to find data sets
- Running data analyses in the Cavatica cloud platform
  - WGS?
  - RNAseq?
- **Data-type specific training, e.g. WGS for clinicians**
- Modalities include Web sites, videos, teleconference training, and (eventually...) in-person.
  
- Also happy to talk with people who are developing tools about testing, documenting, training, etc! :)



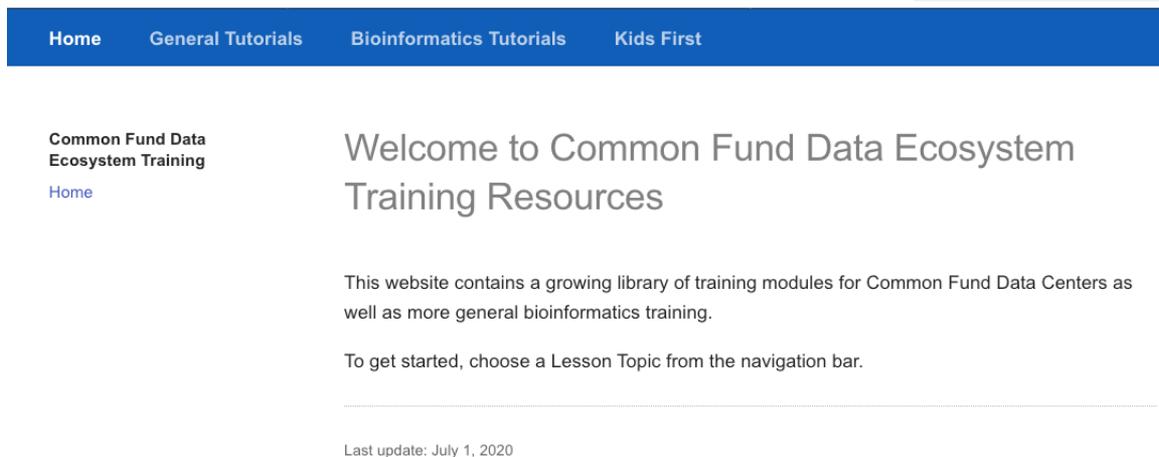
## SUPPORT **FOR KIDS FIRST:**

We will also support Kids First DRC in joining forces with other NIH programs:

- Cross-program analysis (WGS, RNAseq) - e.g. GTEx
- scRNA (HuBMAP) and/or drug effects (LINCS) and/or ...

## PUBLIC CFDE RESOURCES: TRAINING WEB SITE

- × Training Website <https://cfde-training-and-engagement.readthedocs-hosted.com/>
  - + Modules for learning how to use CFDE resources, tutorials on general bioinformatics workflows, and DCC specific tutorials



The screenshot shows the homepage of the Common Fund Data Ecosystem Training website. At the top is a blue navigation bar with the following links: Home, General Tutorials, Bioinformatics Tutorials, and Kids First. Below the navigation bar, on the left, is the site's logo: "Common Fund Data Ecosystem Training" with a "Home" link underneath. The main heading is "Welcome to Common Fund Data Ecosystem Training Resources". Below this, a paragraph states: "This website contains a growing library of training modules for Common Fund Data Centers as well as more general bioinformatics training." A second paragraph says: "To get started, choose a Lesson Topic from the navigation bar." At the bottom, a horizontal line is followed by the text "Last update: July 1, 2020".

## PUBLIC CFDE RESOURCES: TRAINING WEB SITE

- × Training Website <https://cfde-training-and-engagement.readthedocs-hosted.com/>
  - + Modules for learning how to use CFDE resources, tutorials on general bioinformatics workflows, and DCC specific tutorials

### Bioinformatics tutorials

All currently available bioinformatics tutorials can be found here.

- [GWAS in the Cloud](#)
- [Snakemake](#)

### Kids First Lessons

All the currently available lessons for the Kids First Data Resource Center

Setting up your KF Portal Permissions:

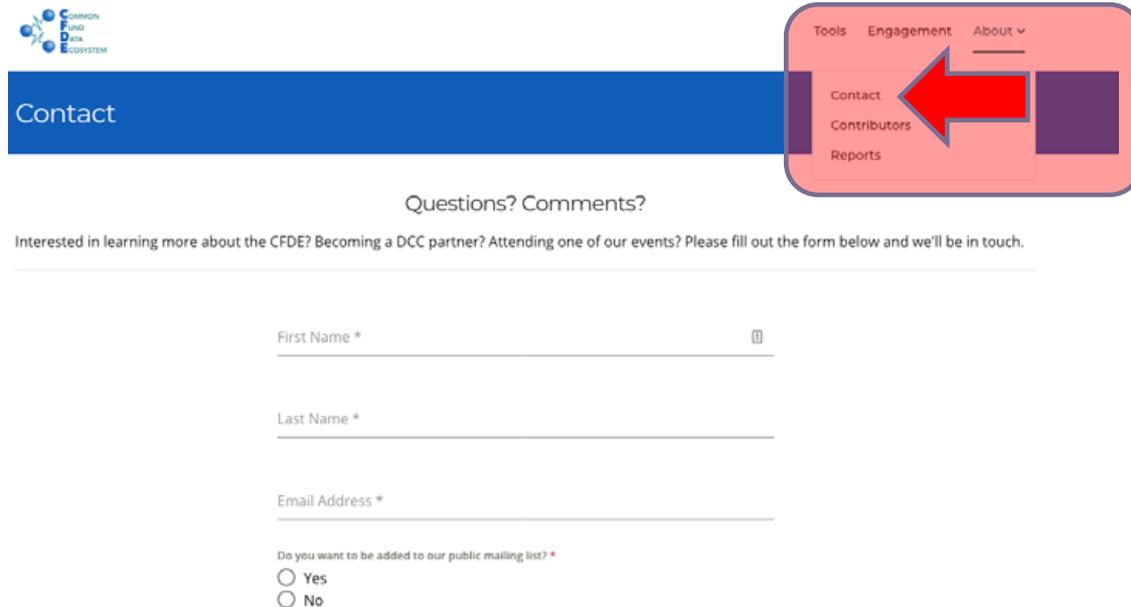
- [Portal Overview](#)
- [Registration](#)
- [Connecting Accounts](#)
- [Accessing Data](#)

# WE WELCOME FEEDBACK!

**ctbrown@ucdavis.edu**

Please let us know:

- What kind of training you are interested in!
- What kind of data set combinations you are interested in!



The image shows a screenshot of the CFDE website's contact page. At the top left is the CFDE logo. Below it is a blue navigation bar with the word "Contact" in white. To the right of the navigation bar is a red rounded rectangle containing a menu with items: "Tools", "Engagement", "About", "Contact", "Contributors", and "Reports". A large red arrow points from the "Contact" menu item back to the blue navigation bar. Below the navigation bar is the text "Questions? Comments?". Underneath is a paragraph: "Interested in learning more about the CFDE? Becoming a DCC partner? Attending one of our events? Please fill out the form below and we'll be in touch." Below this is a contact form with three input fields: "First Name \*", "Last Name \*", and "Email Address \*". At the bottom of the form is a question: "Do you want to be added to our public mailing list? \*" with two radio button options: "Yes" and "No".

**<http://nih-cfde.org>**

# Gabriella Miller Kids First Pediatric Research Program

## *Program Updates*



The Common  
Fund



# **FY20 Long-Read Sequencing Pilot**

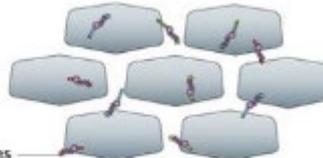
# 2020 Kids First X01 Long Read Sequencing Pilot



**SMRTbell template**  
Two hairpin adapters allow continuous circular sequencing

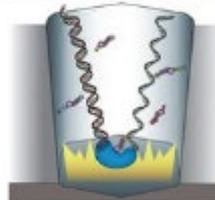


**ZMW wells**  
Sites where sequencing takes place



**Labelled nucleotides**  
All four dNTPs are labelled and available for incorporation

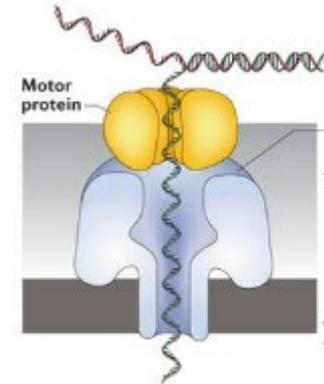
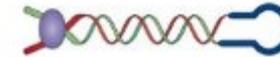
**Modified polymerase**  
As a nucleotide is incorporated by the polymerase, a camera records the emitted light



**PacBio output**  
A camera records the changing colours from all ZMWs; each colour change corresponds to one base

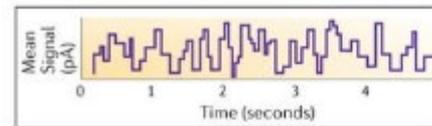


**Leader-Hairpin template**  
The leader sequence interacts with the pore and a motor protein to direct DNA, a hairpin allows for bidirectional sequencing



**Alpha-hemolysin**  
A large biological pore capable of sensing DNA

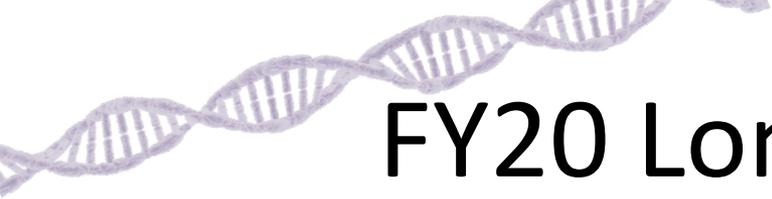
**Current**  
Passes through the pore and is modulated as DNA passes through



**ONT output (squiggles)**  
Each current shift as DNA translocates through the pore corresponds to a particular k-mer

**PACIFIC BIOSCIENCES**

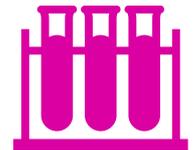
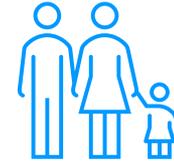
**OXFORD NANOPORE**



# FY20 Long-Read Projects

<https://commonfund.nih.gov/kidsfirst/longreadprojects>

- **Nonsyndromic Craniosynostosis**
  - PI: Simeon Boyd
- **Congenital Diaphragmatic Hernia**
  - PI: Wendy Chung
- **Structural Heart & Other Defects**
  - PI: Bruce Gelb
- **Bladder Exstrophy Epispadias Complex (BEEC)**
  - PI: Angie Jelin
- **Acute Myeloid Leukemia**
  - PI: Soheil Meshinchi
- **Novel Cancer Susceptibility**
  - PI: Sharon Plon
- **Ollier Disease & Maffucci Syndrome**
  - PI: Nara Sobreira



# **Strategic Planning & Phase 2 Initiatives**

# *What is next for Kids First?*

## 2018 Strategic (Re-)Planning Exercise:



- **2018 Program Survey** launched at **ASHG: Kids First** & external investigators
- **Kids First Steering Committee**
- **External Program Consultants**
- **DRC Admin & Outreach Core**  
(feedback from the public, patients, families)
- **NIH Kids First Working Group & IC Director Co-Chairs**



# 7 Consensus Recommendation Themes

## 1. Innovation: Resource, infrastructure, or tool development.

*Activities: Data Visualization tools; other tools for clinical/phenotypic data*



## 2. Clinical/phenotypic data extraction, harmonization, & curation.

*Activities: Collect, extract, organize, curate, harmonize, and submit deep clinical and phenotypic data; annotate variants with pathogenicity, ClinGen scores.*



## 3. Collaborative validation and discovery.

*Activities: Building synthetic cohorts; identify structural variants; test pipelines.*

*\*Engage trainees in data analysis projects\*\*Bring users to the platform\**



## 4. Integration and interoperability of external pediatric datasets.

*Activities: Using DRC workflow and best practices to harmonize external pediatric datasets; Building tools that can operate across multiple spaces*



## 5. Consent and data sharing.

*Activities: Re-consenting cohorts in line with our data sharing expectations*



## 6. Validation with model organisms.

*Activities: validating KF findings/variants, deep phenotyping of animal models*



## 7. Continue WGS & data generation, invest in long-read, consider other –

omics. Reissues of: <https://grants.nih.gov/grants/guide/pa-files/PAR-19-104.html>



# NIH Council of Councils

September 11, 2020

## **Common Fund Concept Clearance: Gabriella Miller Kids First Pediatric Research Program: Plans for FY22-24 (Phase 2)**

James Coulombe, Ph.D.

Chief, Developmental Biology and Structural Variation Branch

*Eunice Kennedy Shriver* National Institute of Child Health and Human Development

**Goal of Phase 2 Initiatives:** Enhance the value and impact of the Kids First Data Resource to accelerate pediatric research to improve preventative measures, diagnostics, and therapeutic interventions.

The archived videocast of the Council of Councils meeting is publicly available and can be viewed [here](#) (Kids First discussion begins at 4:48:00).

The presentation materials are available [here](#).



# ***Phase 2 Initiatives Approved***

## ***\$12.6M/year (FY22-24)***

### **1) Additional generation of childhood cancer and structural birth defects-related -omics data**

- Add epigenomic and proteomic assays



### **2) Continue development & improvement of the Data Resource**

- Plan for sustaining the Data Resource beyond FY24



### **3) Expert-driven activities to increase the value of Kids First data**

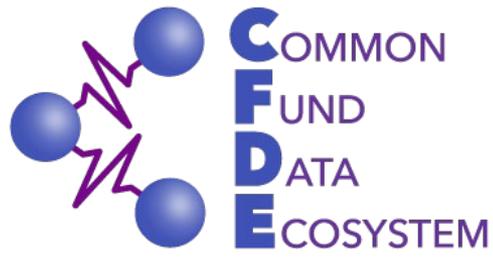
- Engage Kids First & community experts in activities such as integration, curation, and/or harmonization of rich clinical and phenotypic data



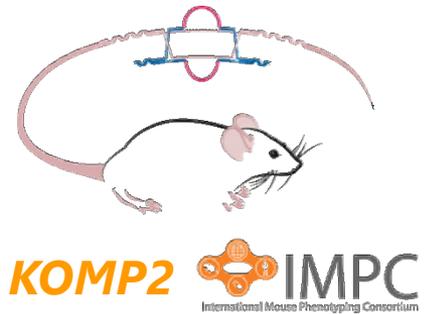
# Collaborations

*In parallel, continue trans-NIH  
collaborations to address  
common goals*

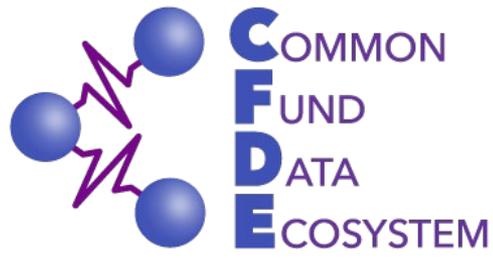
# Collaborations amplify the impact of Kids First



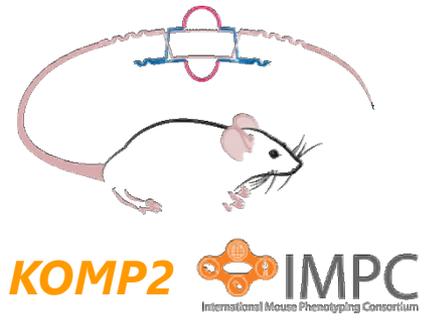
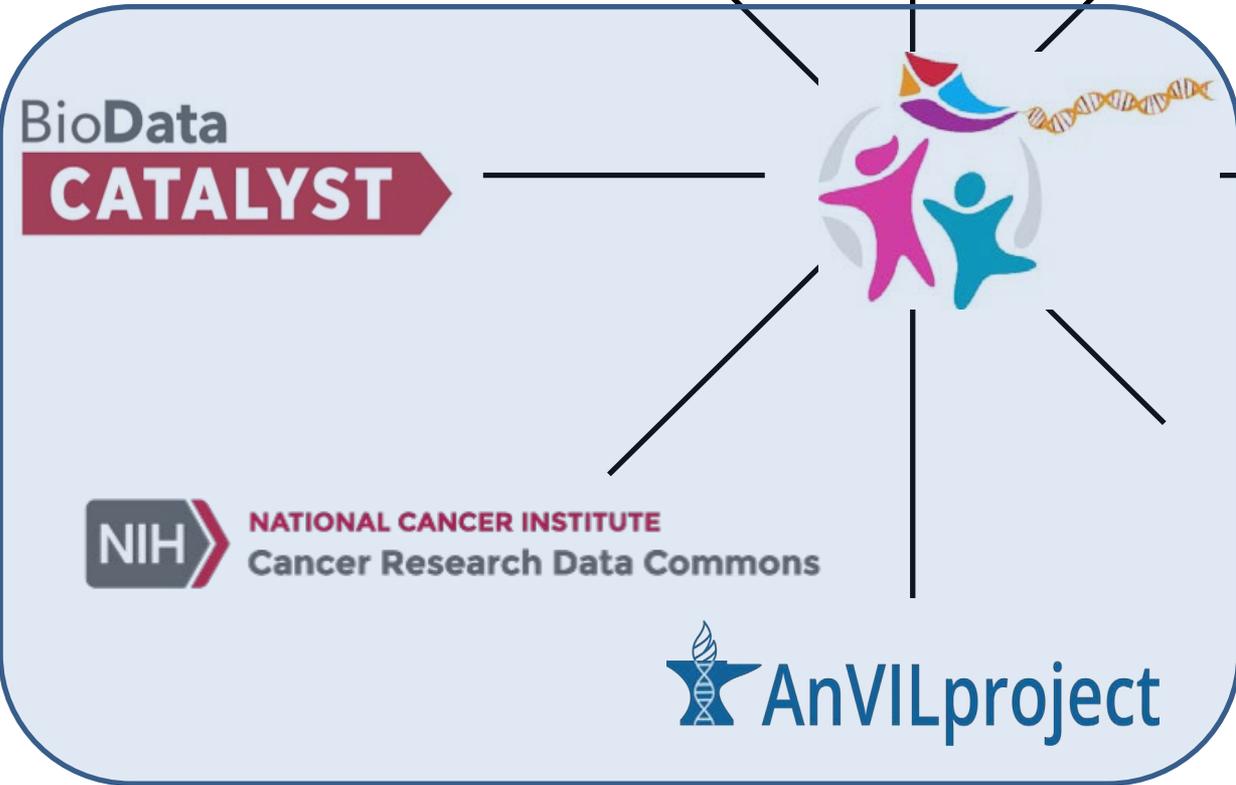
*INvestigation of Co-occurring conditions across the Lifespan to Understand Down syndrome (INCLUDE)*



# Collaborations amplify the impact of Kids First

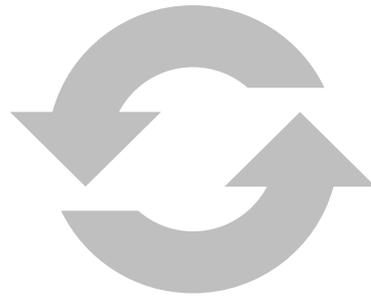


*INvestigation of Co-occurring conditions across the Lifespan to Understand Down syndrome (INCLUDE)*



# NIH Cloud Based Platforms Interoperability (NCPI)

*Goal: Empower end-user analyses across platforms through federation & interoperability*



# New NCPI Webpage!

<https://datascience.nih.gov/nih-cloud-platform-interoperability>



## NIH Cloud Platform Interoperability Effort

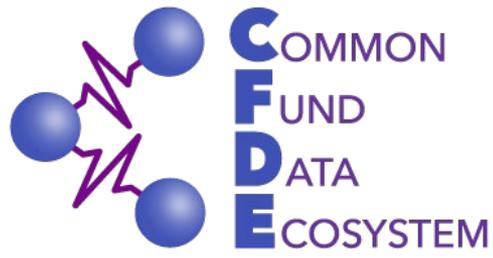
### About the NIH Cloud Platform Interoperability (NCPI) Effort

Connecting NIH's various data systems is a critical step toward improving researchers' access to all types of data. The [NIH Cloud Platform Interoperability \(NCPI\) effort](#) seeks to create a federated genomic data ecosystem and is a collaborative project between NIH and external partners comprising [five working groups](#).

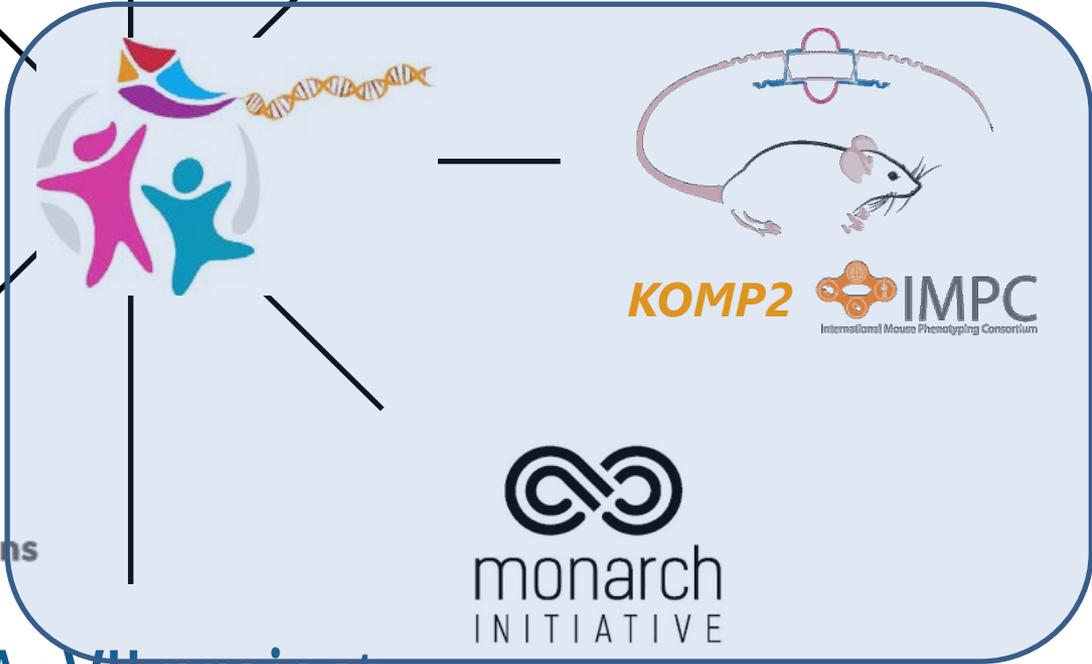
When researchers obtain data from a specific platform, there is no guarantee that the data will be readily usable alongside data from a different platform. By focusing on interoperability, the NCPI effort is ensuring that researchers can both find and integrate data more easily from the following four participating platforms:



# Collaborations amplify the impact of Kids First



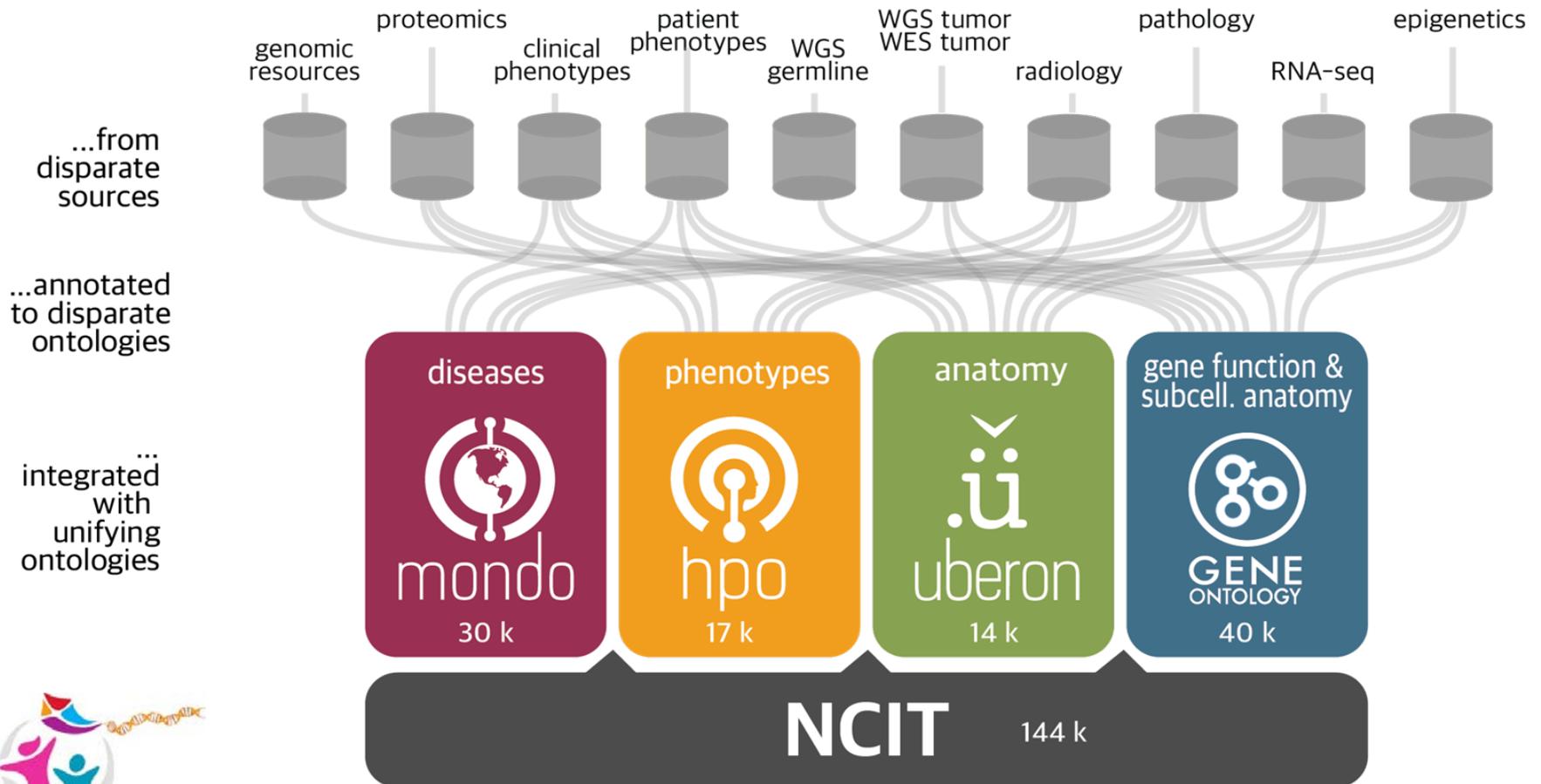
*INvestigation of Co-occurring conditions across the Lifespan to Understand Down syndrome (INCLUDE)*



# Innovation across the Phenotypic Translational Divide Webinar

Information: <https://monarch-initiative.github.io/phenomics/pages/clin-phen-webinar.html>

Curation with ontologies that support heterogenous data types in Kids First



# *Innovation across the Phenotypic Translational Divide Webinar Series*

## Part 1

<p>Pediatric Cardiac Genomics Consortium (PCGC) - Betsy Goldmuntz</p>	<p>Down Syndrome -Congenital Heart Disease - Joaquin Espinosa</p>	<p>Neuropsychological Data Harmonization - Stephanie Sherman</p>
<p>Down Syndrome-ALL &amp; Rhabdomyosarcoma - Phillip Lupo</p>		<p>Enchondromatosis and Related Malignant Tumors - Nara Sobreira</p>
<p>Orofacial Clefts - Mary Marazita</p>	<p>Adolescent Idiopathic Scoliosis - Carole Wise</p>	<p>Cornelia de Lange Syndrome - Sarah Raible</p>

## Part 2

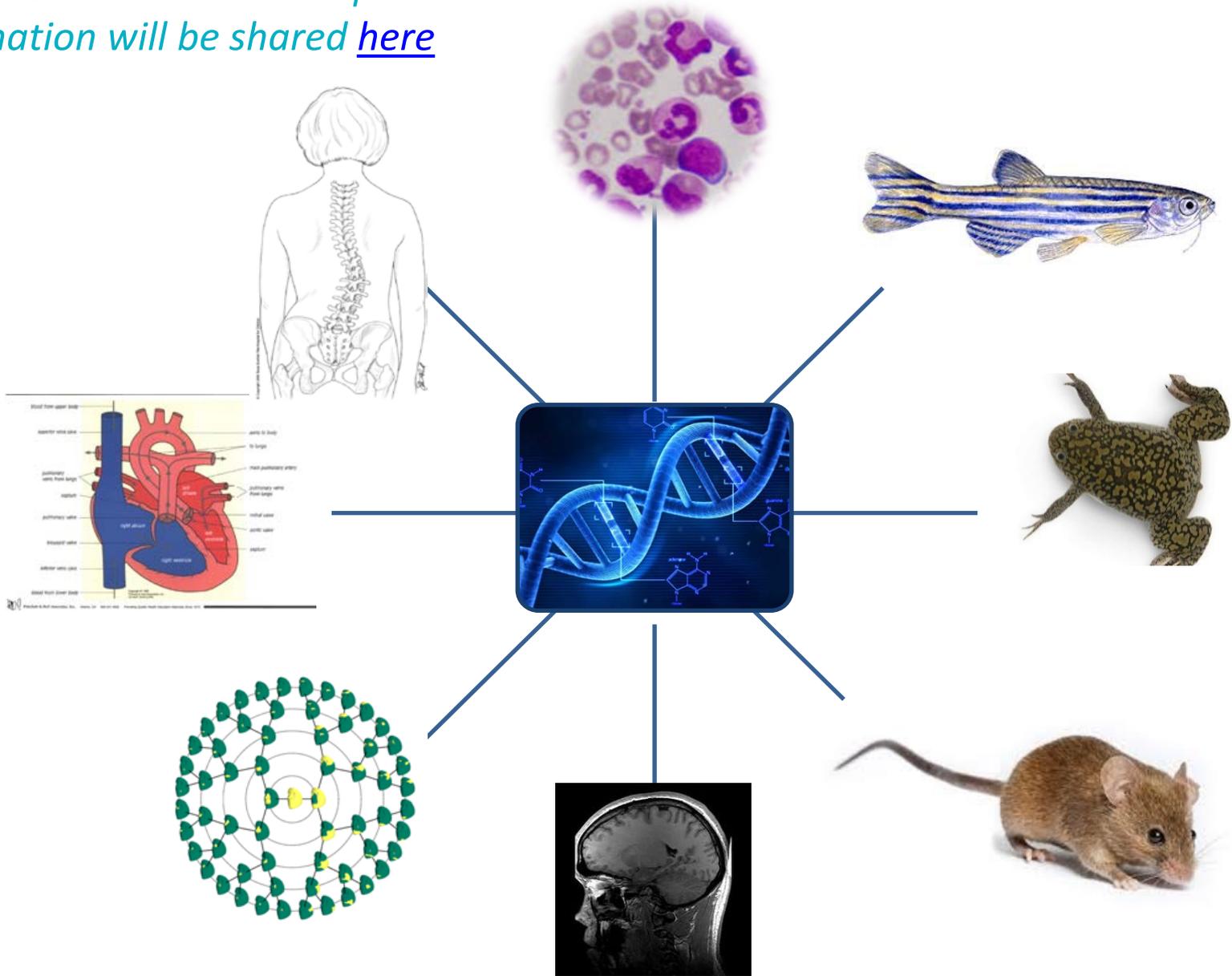
	<p><b>Kidney and Urinary Tract Defects</b> Ali Gharavi, Columbia University</p>
	<p><b>Structural birth defects and childhood cancer</b> Xiang Wang, Children's Hospital of Philadelphia</p>
	<p><b>Neuroblastoma</b> Sharon Diskin, Children's Hospital of Philadelphia</p>
	<p><b>CHARGE syndrome</b> Donna Martin, Michigan Medicine</p>
	<p><b>Laterality Birth Defects</b> Stephanie Ware, Indiana University</p>
	<p><b>Structural Brain Defects, Neural Tube Defects</b> Joe Gleeson, University of California, San Diego</p>

Webinar Information:

<https://monarch-initiative.github.io/phenomics/pages/clin-phen-webinar.html>

# Part 3: Cross-Species Genotype-Phenotype Analysis

*Save-the-Date: Dec 9<sup>th</sup> 3-6pm ET*  
*Information will be shared [here](#)*



# Q & A

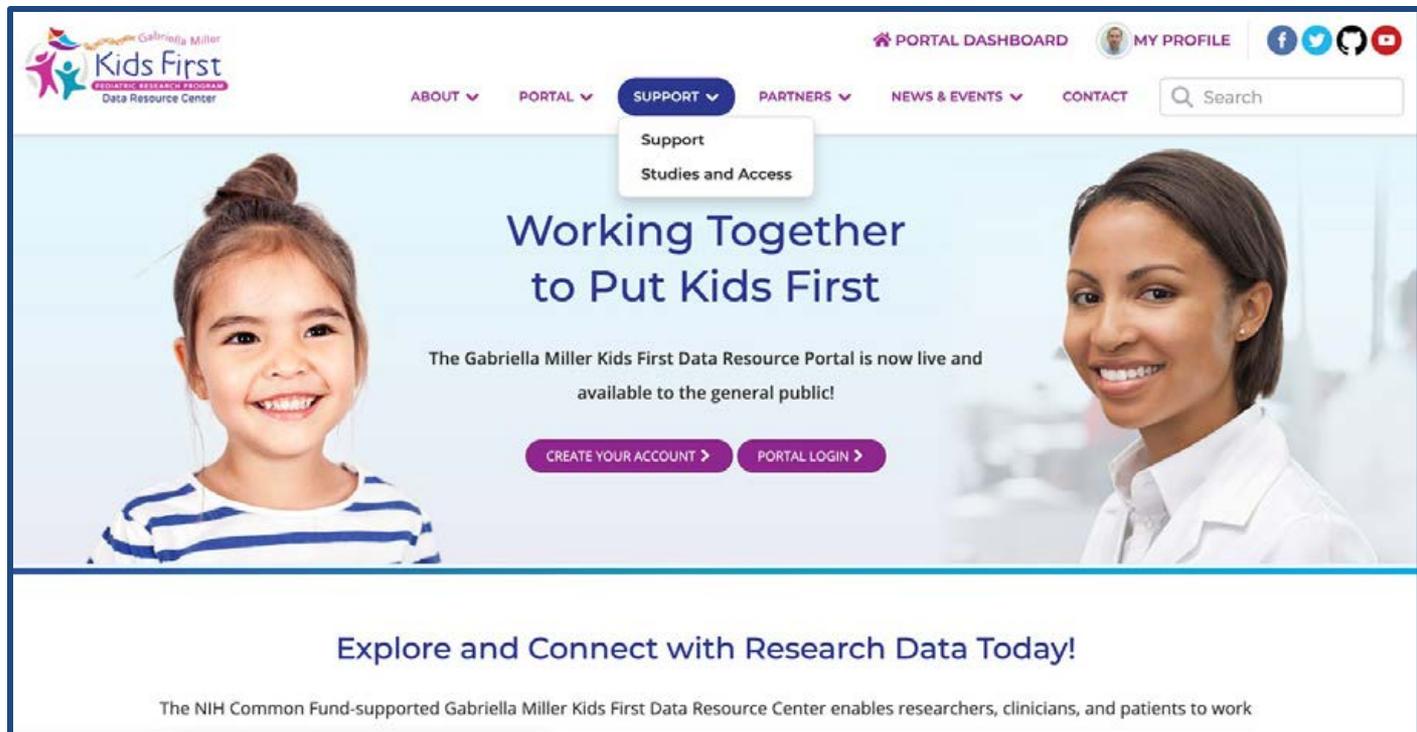
- Use the Q&A bar (lower right of your screen) to send your questions to “**All Panelists**”. We will read your questions out loud and answer them.
- You can also use the “chat”  service to send private messages to the host or presenters.

**What support is available for learning how to use the Kids First platforms?**



# What support is available for learning how to use the Kids First platforms?

## Support Pages on [kidsfirstdrc.org](https://kidsfirstdrc.org)



The screenshot displays the website for the Gabriella Miller Kids First Pediatric Research Program Data Resource Center. The top navigation bar includes links for PORTAL DASHBOARD, MY PROFILE, and social media icons. A dropdown menu for SUPPORT is open, showing options for Support, Studies, and Access. The main banner features a young girl and a woman in a lab coat, with the text "Working Together to Put Kids First" and "The Gabriella Miller Kids First Data Resource Portal is now live and available to the general public!". Below the banner are buttons for "CREATE YOUR ACCOUNT" and "PORTAL LOGIN". At the bottom, there is a call to action: "Explore and Connect with Research Data Today!" and a brief description of the center's mission.

**Support**  
Studies and Access

### Working Together to Put Kids First

The Gabriella Miller Kids First Data Resource Portal is now live and available to the general public!

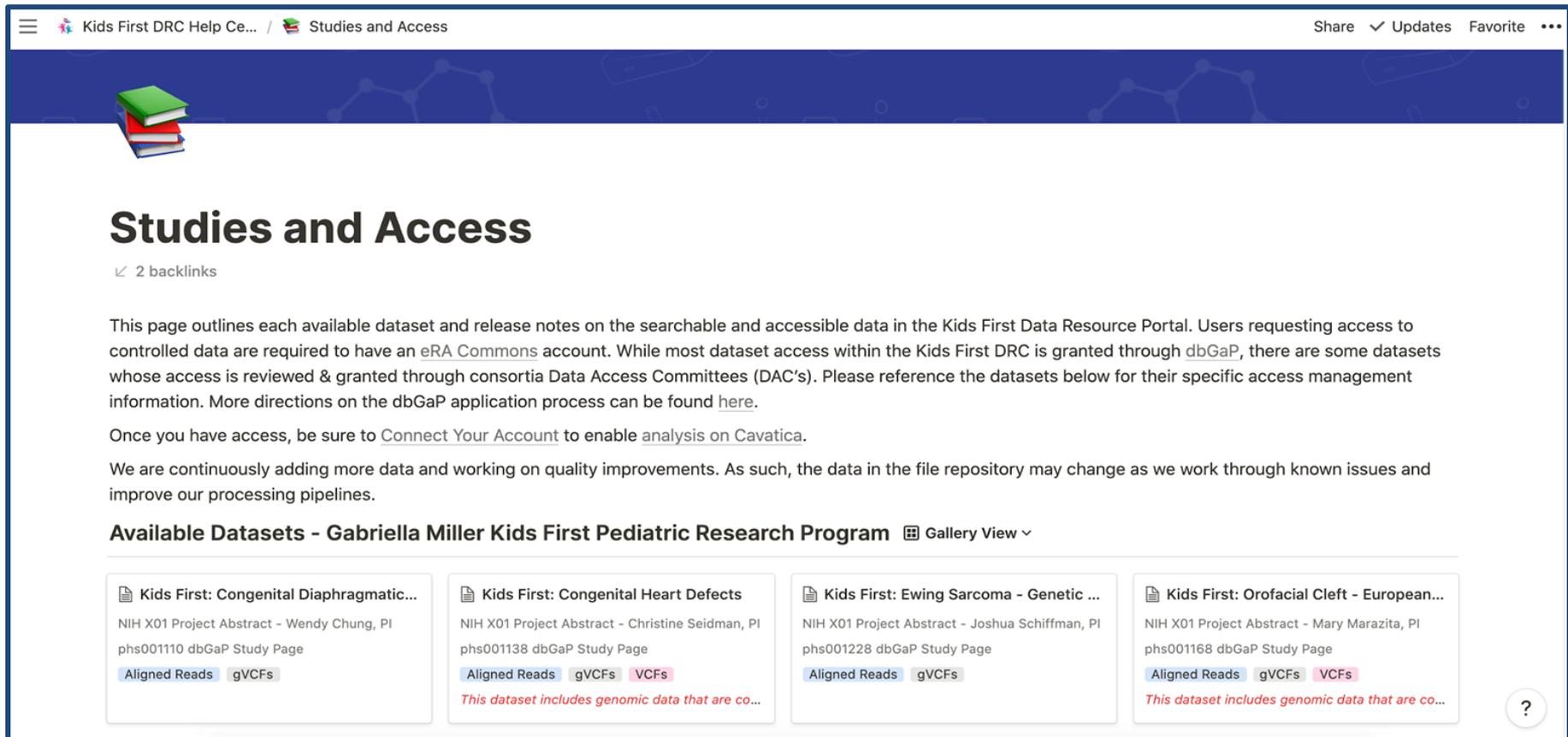
[CREATE YOUR ACCOUNT](#) [PORTAL LOGIN](#)

**Explore and Connect with Research Data Today!**

The NIH Common Fund-supported Gabriella Miller Kids First Data Resource Center enables researchers, clinicians, and patients to work



# What support is available for learning how to use the Kids First platforms?



The screenshot shows the 'Studies and Access' page in the Kids First DRC Help Center. The page has a blue header with a hamburger menu, the title 'Kids First DRC Help Ce... / Studies and Access', and navigation options 'Share', 'Updates', and 'Favorite'. Below the header is a blue banner with a book icon. The main content area has the title 'Studies and Access' and '2 backlinks'. The text explains that the page outlines available datasets and release notes, mentioning that access to controlled data requires an eRA Commons account and that some datasets are reviewed by Data Access Committees (DAC's). It also notes that data is continuously added and improved. Below the text is a section titled 'Available Datasets - Gabriella Miller Kids First Pediatric Research Program' with a 'Gallery View' dropdown. There are four dataset cards, each with a title, PI name, NIH X01 Project Abstract ID, dbGaP Study Page, and tags for 'Aligned Reads', 'gVCFs', and 'VCFs'. The last two cards include a red note: 'This dataset includes genomic data that are co...'. A question mark icon is in the bottom right corner.

Kids First DRC Help Ce... / Studies and Access Share ✓ Updates Favorite ...

## Studies and Access

2 backlinks

This page outlines each available dataset and release notes on the searchable and accessible data in the Kids First Data Resource Portal. Users requesting access to controlled data are required to have an [eRA Commons](#) account. While most dataset access within the Kids First DRC is granted through dbGaP, there are some datasets whose access is reviewed & granted through consortia Data Access Committees (DAC's). Please reference the datasets below for their specific access management information. More directions on the dbGaP application process can be found [here](#).

Once you have access, be sure to [Connect Your Account](#) to enable [analysis on Cavatica](#).

We are continuously adding more data and working on quality improvements. As such, the data in the file repository may change as we work through known issues and improve our processing pipelines.

### Available Datasets - Gabriella Miller Kids First Pediatric Research Program Gallery View

- Kids First: Congenital Diaphragmatic...**  
NIH X01 Project Abstract - Wendy Chung, PI  
phs001110 dbGaP Study Page  
Aligned Reads gVCFs
- Kids First: Congenital Heart Defects**  
NIH X01 Project Abstract - Christine Seidman, PI  
phs001138 dbGaP Study Page  
Aligned Reads gVCFs VCFs  
*This dataset includes genomic data that are co...*
- Kids First: Ewing Sarcoma - Genetic ...**  
NIH X01 Project Abstract - Joshua Schiffman, PI  
phs001228 dbGaP Study Page  
Aligned Reads gVCFs
- Kids First: Orofacial Cleft - European...**  
NIH X01 Project Abstract - Mary Marazita, PI  
phs001168 dbGaP Study Page  
Aligned Reads gVCFs VCFs  
*This dataset includes genomic data that are co...*



**Studies and Access page in the Kids First DRC Help Center**

What support is available for learning how to use the Kids First platforms?

## Kids First Support Inbox

*Send any questions to...*

**Kids First Support Inbox**  
**[Support@KidsFirstDRC.org](mailto:Support@KidsFirstDRC.org)**



# What support is available for learning how to use the Kids First platforms?

## Kids First Office Hours

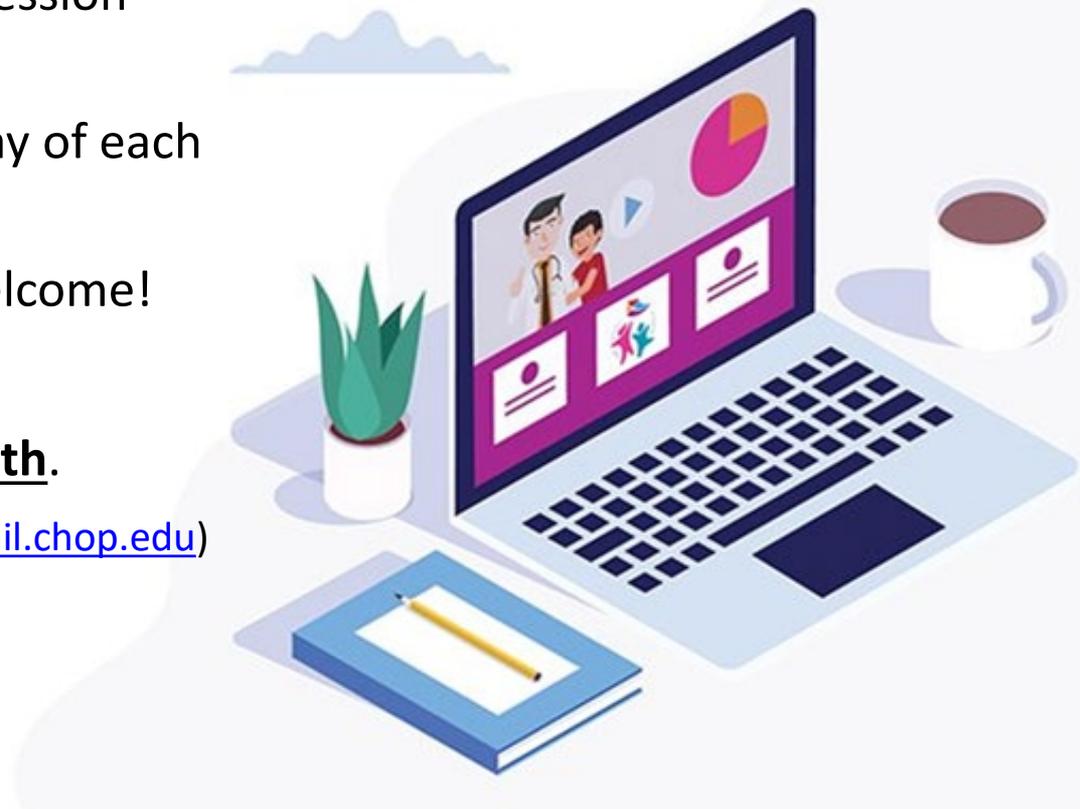
**What:** Monthly user support session hosted by the Kids First DRC

**When:** 3pm on the 2nd Tuesday of each month

**Who:** Users of all levels are welcome!

Our next session is **October 13th.**

Contact David Higgins ([HigginsD@email.chop.edu](mailto:HigginsD@email.chop.edu)) for login information.



# Q & A

- Use the Q&A bar (lower right of your screen) to send your questions to **“All Panelists”**. We will read your questions out loud and answer them.
- You can also use the “chat”  service to send private messages to the host or presenters.

What funding opportunities are available?

How do I access data?

How can I get involved?



# ***What funding opportunities are available?***

## Discovery of the Genetic Basis of Childhood Cancers and of Structural Birth Defects: Gabriella Miller Kids First Pediatric Research Program (X01 Clinical Trial Not Allowed)

### **Kids First cohort sequencing opportunity:**

- Propose samples from cohorts for whole genome sequencing at a Kids First-supported sequencing center
  - DNA from germline/normal from affected child and parents/family members (if available)
  - DNA/RNA from tumors and/or affected tissue
- 1 more reissue of [PAR-19-390](#) for 2021
- View FY20 materials [here](#)
- Considerations include broad data sharing & value of incorporating the data into the Data Resource



# ***What funding opportunities are available?***

See: FAQs for Funding Opportunities Announcements (FOAs) to Support Data Analyses of Kids First Datasets (<https://commonfund.nih.gov/kidsfirst/FAQ>)

- **Analyze Kids First data with support from:**

- **“Kids First R03 PAR”:** [PAR-19-375](https://commonfund.nih.gov/kidsfirst/FAQ)
- NIH “Parent” R01: [PA-19-056](https://commonfund.nih.gov/kidsfirst/FAQ)
- NIH Parent R03: [PA-19-052](https://commonfund.nih.gov/kidsfirst/FAQ)

- **Validate variants with support from:**

- ORIP’s Development of Animal Models and Related Biological Materials for Research (R21): <https://grants.nih.gov/grants/guide/pa-files/PA-16-141.html>
- Mechanistic Studies of Gene-Environment Interplay in Dental, Oral, Craniofacial, and Other Diseases and Conditions (R01) ([PAR-19-292](https://commonfund.nih.gov/kidsfirst/FAQ)).
- Development of Novel and Robust Systems for Mechanistic Studies of Gene-Environment Interplay in Dental, Oral, Craniofacial, and Other Diseases and Conditions (R21) ([PAR-19-293](https://commonfund.nih.gov/kidsfirst/FAQ)).
- To pursue collaborations with the [Knockout Mouse Phenotyping Program \(KOMP2\)](https://commonfund.nih.gov/kidsfirst/FAQ), contact: [KidsFirstKOMP@nih.gov](mailto:KidsFirstKOMP@nih.gov)

- **To receive updates about future Kids First opportunities, sign up for the listserv:**

- <https://commonfund.nih.gov/kidsfirst/register>



# How do I access data?

Anyone can register & login to the portal to filter, search, visualize datasets



Dashboard | File Repository

Kids First Data Resource Center

28,810 Files | 5,621 Participants | 1,625 Families | 750.47 TB Size

Showing 1 - 20 of 28,810 files

File ID	Participants ID	Study Name	Proband	Family ID	Data Type	File Format	File Size
GF_W0B3KSH	PT_J824PK7	Congenital Diaphra...	No	FM_Q85FMJ8	Aligned Reads	cram	15.53 GB
GF_BT35C7YV	PT_95T516RP	Congenital Diaphra...	No	FM_JADBN593	gVCF	gVCF	4.3 GB
GF_PTYBTP23	PT_2P1852YW	Congenital Diaphra...	No	FM_7CXDVHEP	gVCF	gVCF	5.94 GB
GF_RHQ4Q4C5	PT_5VXQJRA4	Congenital Diaphra...	No	FM_88T040VF	gVCF	gVCF	4.91 GB
GF_TDP3Q71	PT_Y2C44N7	Congenital Diaphra...	Yes	FM_33MY1VDM	Aligned Reads	bam	63.33 GB
GF_V031CSX	PT_RHW06ACA	Congenital Diaphra...	Yes	FM_FTQZWR1	gVCF	gVCF	5.37 GB
GF_BBEMPER	PT_5NV37967	Congenital Diaphra...	No	FM_5BGRV3	Aligned Reads	cram	16.87 GB
GF_GVB13YXN	PT_4ZBHQAM	Congenital Diaphra...	Yes	FM_HFSQFX6	Aligned Reads	bam	63.74 GB
GF_SAYKAVDW	PT_JPV99EDB	Congenital Diaphra...	No	FM_C2C8K05	Aligned Reads	cram	20.77 GB
GF_BY3WS2X	PT_QQ3MBPM	Congenital Diaphra...	Yes	FM_J0SDXHE	Aligned Reads	bam	62.31 GB
GF_00Q3K3SH	PT_2BHHBN57	Congenital Diaphra...	No	FM_7CXDVHEP	Aligned Reads	cram	20.62 GB
GF_FEB15QRD	PT_QQ31MEW3	Congenital Diaphra...	No	FM_PYH2RAJ2	Aligned Reads	bam	64.63 GB
GF_FNMQQ55G	PT_07B67CK2	Congenital Diaphra...	Yes	FM_4CSQ04FW	Aligned Reads	cram	20.26 GB
GF_5Y83Q23C	PT_ARGH0XBP	Congenital Diaphra...	Yes	FM_PHSTBST4	Aligned Reads	cram	20.95 GB

Actions

If you have not selected any files, all files in your query will be included in the actions.

Data Analysis

**COPY FILES TO CAVATICA**

Download

MANIFEST

BIOSPECIMEN

CLINICAL

NIH Kids First  
Data Access Committee

1. Find available data on the [portal](#) or [Kids First X01 page](#)
2. Submit [dbGaP Data Access Requests \(DARs\)](#) for individual-level sequence data
3. Push approved sequence data to Cavatica from the portal: <https://kidsfirstdrc.org/support/analyze-data/>



# Individual-level sequence data

- To learn more about submitting dbGaP Data Access Requests (DARs) watch:

[https://www.youtube.com/watch?v=39cba0gF2tw&index=3&t=503s&list=PLoXwgZfIAe4aMwWpVQU\\_WVeWHzyhI3BCu](https://www.youtube.com/watch?v=39cba0gF2tw&index=3&t=503s&list=PLoXwgZfIAe4aMwWpVQU_WVeWHzyhI3BCu)



**Also see:**

[https://dbgap.ncbi.nlm.nih.gov/aa/dbgap\\_request\\_process.pdf](https://dbgap.ncbi.nlm.nih.gov/aa/dbgap_request_process.pdf)



Submitting an Approvable  
dbGaP Data Access Request

Vivian Ota Wang, Ph.D  
Office of Data Sharing  
NCI

# *How can I get involved?*

- **Connect with and provide feedback to the DRC:**  
[support@kidsfirstdrc.org](mailto:support@kidsfirstdrc.org)
- **Contact the program for questions or feedback:**  
[kidsfirst@od.nih.gov](mailto:kidsfirst@od.nih.gov)
- **Learn more about the program & DRC:**  
<https://commonfund.nih.gov/kidsfirst> &  
<https://kidsfirstdrc.org/>
- **Search data available through the Kids First Data Resource Portal:** <https://portal.kidsfirstdrc.org/>



# Q & A

- Use the Q&A bar (lower right of your screen) to send your questions to “**All Panelists**”. We will read your questions out loud and answer them.
- You can also use the “chat”  service to send private messages to the host or presenters.



# Thank You!

Email Additional Questions and Comments to  
the Kids First Mailbox: [kidsfirst@od.nih.gov](mailto:kidsfirst@od.nih.gov)



# Kids First Investigators: Past Presentations

- **Congenital Diaphragmatic Hernia,**  
PI: Wendy Chung (April 2019):  
<https://www.youtube.com/watch?v=3CS6AphmCp0&t=978s>
- **Neuroblastoma,**  
PI: Sharon Diskin (September 2019):  
<https://www.youtube.com/watch?v=Gq8kK2UGI4s>
- **Orofacial Clefts**  
PI: Mary Marazita (May 2020)  
<https://www.youtube.com/watch?v=Tddklx3lZpl>

