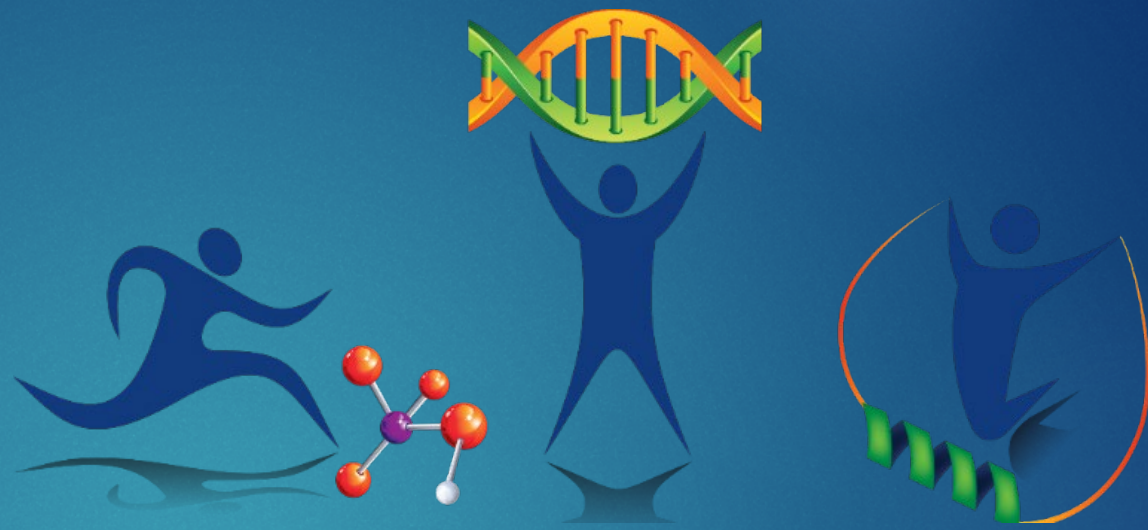
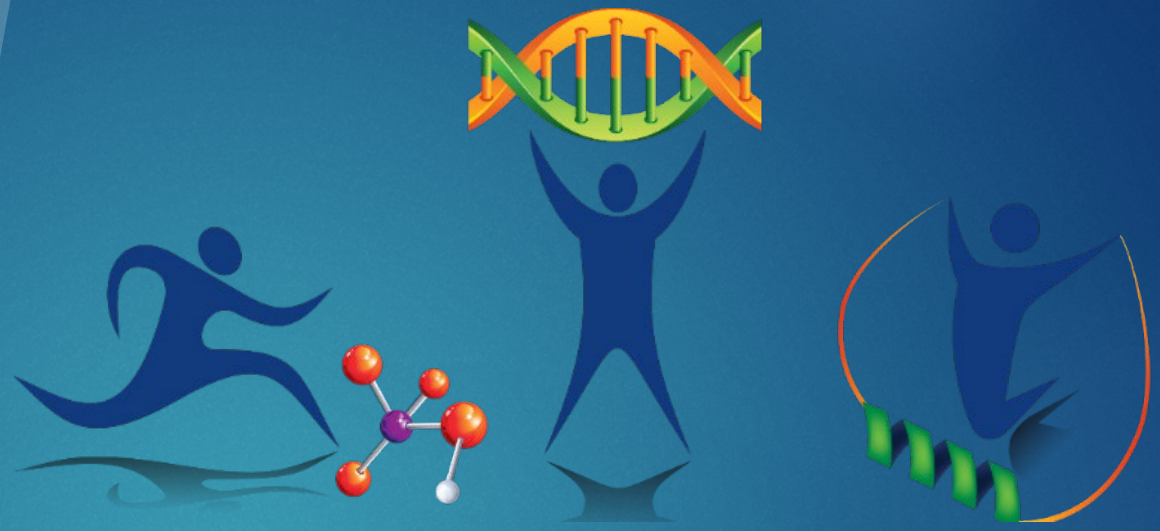


Early Insights From The MoTrPAC Consortium

Stephanie George, PhD, MPA, MA, NIAMS, NIH
Michael E. Miller, PhD, Wake Forest University School of Medicine
Scott Rushing, BS, Wake Forest University School of Medicine
Cindy Stowe, MPM, Wake Forest University School of Medicine
Matthew Wheeler, MD, PhD, Stanford University



Unique Design Aspects of the MoTrPAC Clinical Studies

Stephanie M. George, PhD, MPH, MA

Epidemiologist and Program Director

National Institute of Arthritis and Musculoskeletal and Skin Diseases

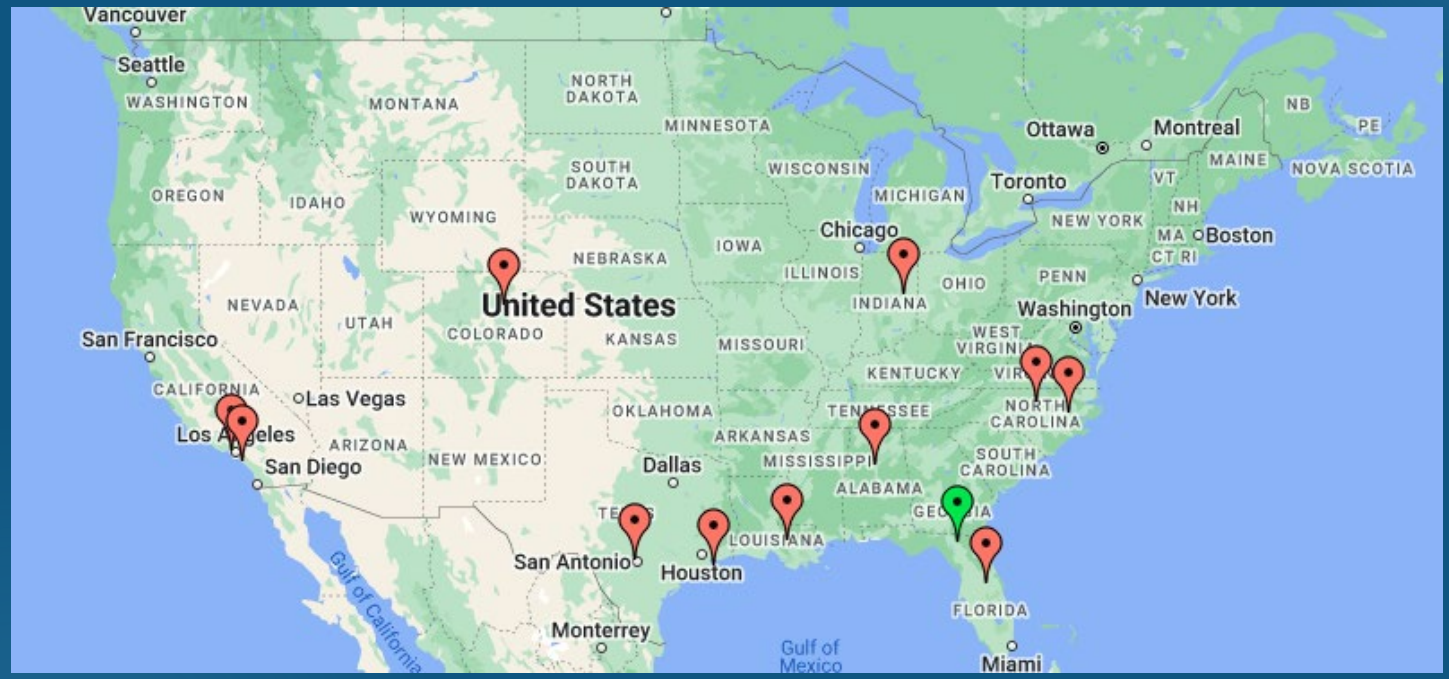
Disclosures

- No Relevant Disclosures

Why MoTrPAC was built

- We know physical activity is good for us, but how does it improve and preserve health in real-time?
- MoTrPAC is the largest NIH-funded study of physical activity to-date.
  - Designed to understand exactly what physical activity does in our bodies.
  - Assess molecular changes that occur in response to physical activity (PA).
  - The rich data will help us know how to optimize physical activity across the lifespan, from middle school children to older adults

MoTrPAC clinical sites



- AdventHealth Translational Research Institute for Metabolism and Diabetes, Orlando FL
- Ball State University, Muncie, IN
- Cedars-Sinai Medical Center, Los Angeles, CA
- Duke University, Durham, NC
- East Carolina University, Greenville, NC
- Pennington Biomedical Research Center, Baton Rouge, LA
- The University of Alabama at Birmingham, Birmingham, AL
- University of California, Irvine, CA
- University of Colorado, Denver, CO
- University of Texas Health Science Center, San Antonio, TX
- University of Texas Medical Branch, Galveston, TX

Executing multiple designs

SEPARATE BUT HARMONIZED ADULT AND PEDIATRIC STUDIES

Participants: Adult Study

~1,980 sedentary adults randomized to one of three intervention arms in which molecular transducers are examined in muscle, fat, and blood: endurance exercise (EE), resistance exercise (RE), or Control for 12 weeks

~300 highly active EE and RE adults in which similar molecular transducers are studied in response to either an acute bout of EE or RE

Participants: Pediatric Study

Cross-sectional phase in ~270 low active children ages 10-18 in which molecular transducers (obtained from blood sampling) are measured in response to an acute exercise challenge

An intervention phase that parallels the adults and involves ~170 low active children randomized to either EE or control

Response of a cohort of 50 highly active EE trained athletes to an acute bout of EE

Ensuring standardized measurements and operating procedures across studies and sites

UPHOLDING RIGOR AND REPRODUCIBILITY: A TOUR OF SOME KEY MEASURES

Visits

- Screening/Phenotyping
- Familiarizations
- Baseline Acute Test/Rest
- Adherence/Monitoring
- Follow-up Phenotyping (Post Intervention)
- Post-Intervention Acute Test/Rest

Characterizing exercise, movement, and strength

Cardiopulmonary Exercise Test (CPET)

- Measure cardiorespiratory fitness on an electronically braked cycle ergometer per the American College of Sports Medicine (ACSM) Testing Guidelines.
  - Oxygen consumption and carbon dioxide production are measured by indirect calorimetry.

Accelerometry

Adult study

- Triaxial wrist-worn ActiGraph
  - 28 days of 24-hour data capture among 1,980 sedentary participants.
  - 7 days of 24-hour data capture among 300 highly active participants
- ActivPAL
  - 28 days of 24-hour ActivPAL data capture among a subsample of sedentary participants

Pediatric Study

- Triaxial wrist-worn ActiGraph
  - 7 days of 24-hour data capture among 320 participants
  - 28 days of 24-hour data capture among 170 participants

Rationale for actigraphy in MoTrPAC

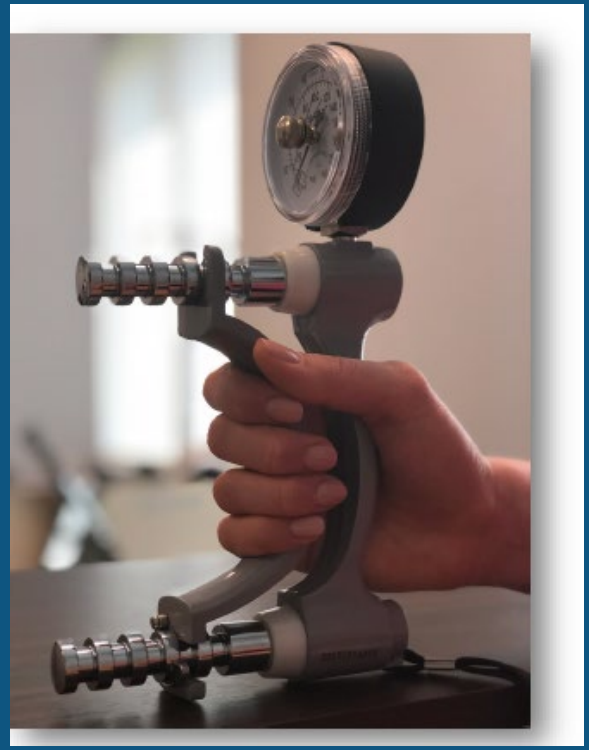
- To determine if free-living physical activity was impacted by supervised exercise training (particularly important in sedentary intervention study).
- To determine if free-living physical activity, beyond the supervised training, might account for some of the molecular response
- Because it is wrist-worn, it will give us some indication of sleep.

Self-reported physical activity

- Training and Activity History (PED)
  - Sport and training history by recreation or competition (current and past activity by age)
- Paffenbarger (SED, HA)
  - Assess activity across the past year
- International Physical Activity Questionnaire (SED, HA)
  - Captures job-related, transportation-related, housework, house maintenance, and caring for family and recreation, sport and leisure time and time spent sitting over the past week.

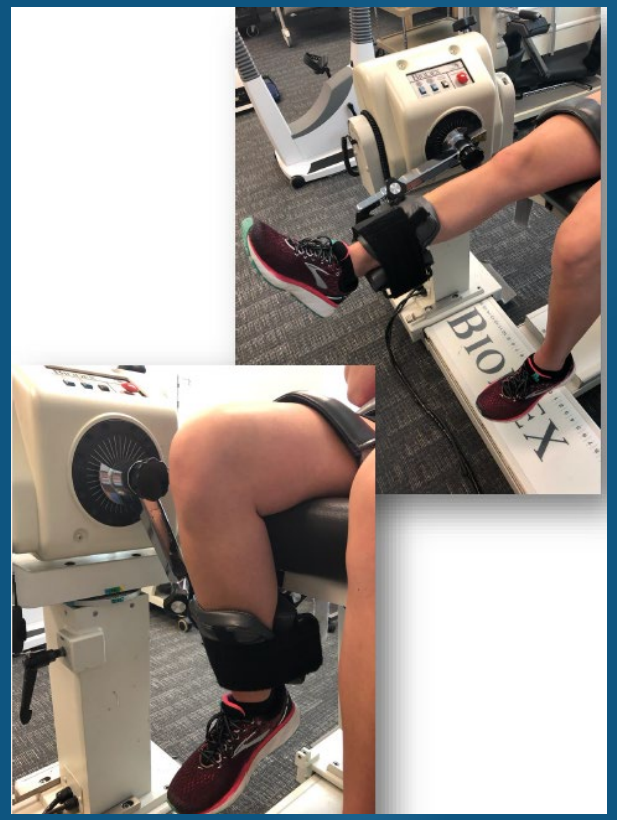
Strength - Grip Strength

- Measure isometric hand grip strength of the dominant hand using a handgrip dynamometer



Strength - Isometric Knee Extension

- Measure knee extensor strength via isometric maximum voluntary contraction at approximately 60° knee flexion



Strength - 1RM (Repetition Max)

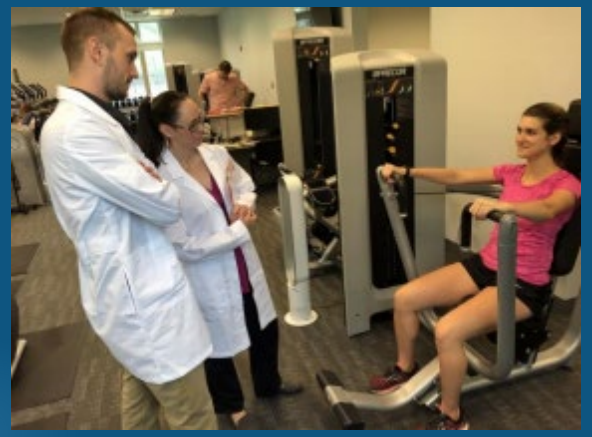
- Measure dynamic voluntary strength on one upper body and two lower body exercises

Acute Endurance Exercise Test

- After a 5 min warm-up, participants cycle for ~40 minutes at a pre-determined workload that corresponds to ~65% VO2peak
- Respiratory variables (by indirect calorimetry), RPM, Heart Rate, RPE, and Blood Pressure are collected periodically
- Specimens are collected before, during, and after the test (per randomization)

Acute Resistance Exercise Test

- After a 5 min warm-up, participants perform three sets of specified resistance exercises at target loads for each exercise and in a specified order.
  - Chest Press, Overhead Press, Seated Row, Triceps Extension, Biceps Curl, Leg Press, Leg Curl, Leg Extension
- Heart Rate data is monitored continuously
- Specimens are collected before and after the test (per randomization)



Control Rest

- Prescribed rest periods to coincide with the timing of the acute tests/biospecimen collection
  - Rest 1 = ~30 mins
  - Rest 2 = ~40 mins
  - Rest 3 = ~4 hours
  - Rest 4 = ~30 mins
- Specimens are collected before, during, and after the rest (per randomization)

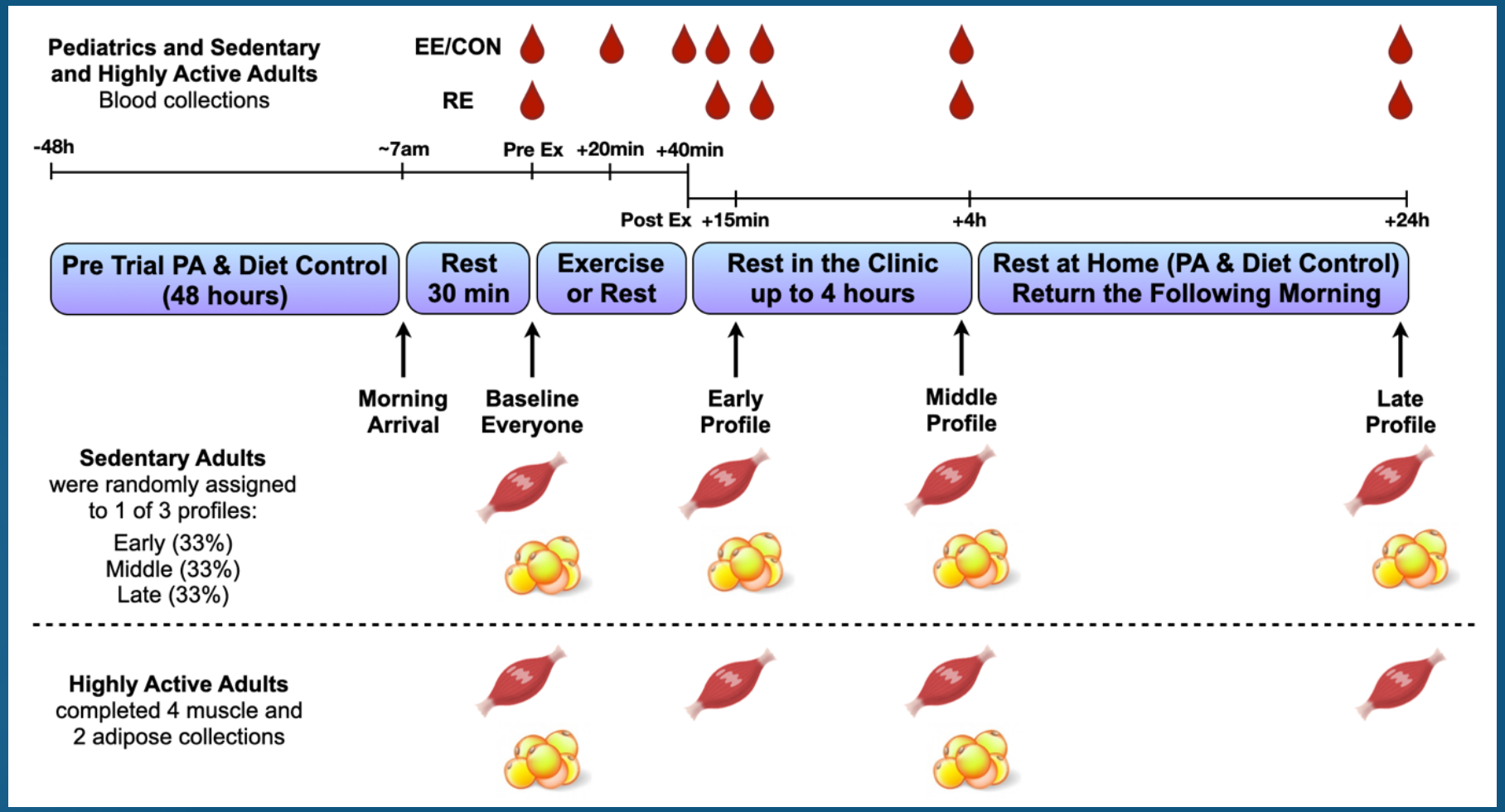
Characterizing the body's response to movement

Intervention

- Endurance and Resistance Exercise
  - All sessions supervised by an Exercise Professional
  - 3 sessions per week for 12 weeks
  - Endurance Exercise: Bicycle and Treadmill with HR monitoring; Progresses from 60-80% of Heart Rate Reserve
  - Resistance Exercise: 8 exercises – Chest press, overhead-press, seated row, triceps extension, bicep curl, leg press, leg curl, leg extension; 3 Sets to Fatigue – Target of 8-12 repetitions per set
- No Exercise Controls
  - Maintain current activity level and dietary habits
  - 3 check-in phone calls or in-person sessions over the course of 12 weeks

Biomarkers

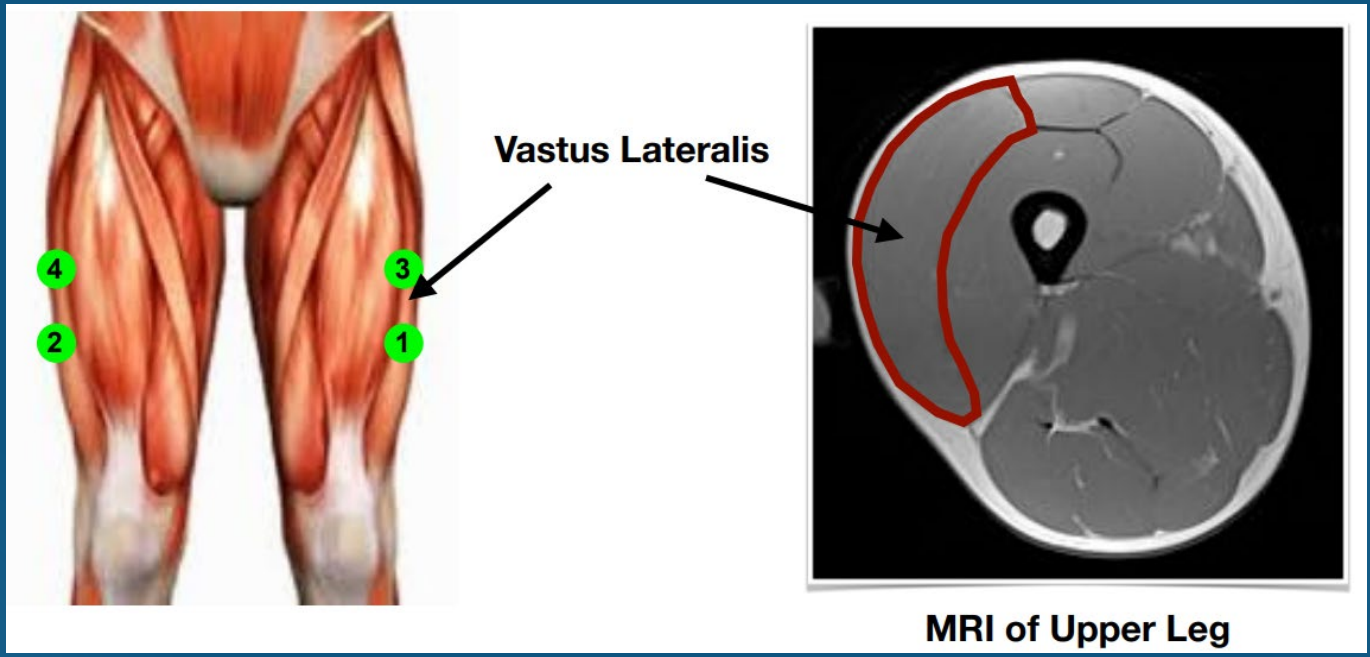
Blood, muscle, and adipose tissue collected prior to, during, and following a 40-minute acute bout of exercise



Blood Collection

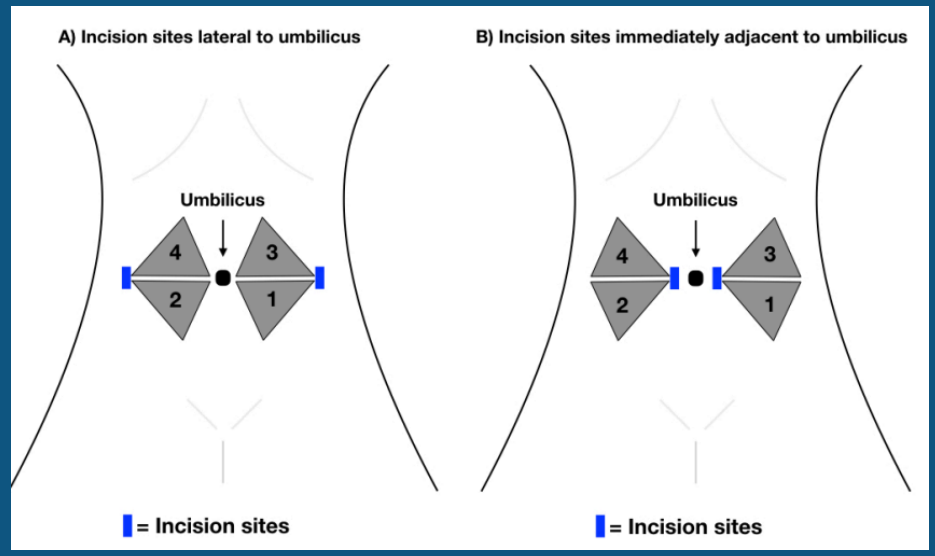
- Collected from via catheter or venipuncture (depending on participant and timepoint)
- Blood is collected via syringes and transferred to blood tubes

Muscle Collection



- Collected from both thighs (taken from the Vastus Lateralis) with Bergstrom style biopsy needle connected to syringe for suction
- Goal is ~150 mg per timepoint collection (2-4) for HA/SED

Adipose Collection



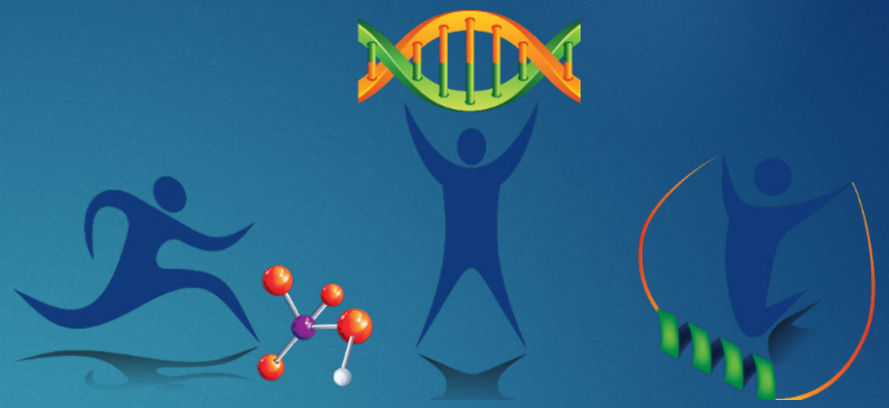
- Collected from peri-umbilical region) with Mercedes cannula connected to a collection syringe
- Goal is ~200 mg per timepoint collection (2) for HA
- Goal is ~600 mg per timepoint collection (1-2) for SED

Problem solving in a multidisciplinary large team

EXAMPLES OF KEY CHALLENGES

The value of working groups charged to tackle key issues from various angles

- Debates on needed design features given “hypothesis generating” purpose of overall consortium
  - Example 1: 1 per household/family
  - Example 2: Individual versus Group Based Intervention (IRT vs IRGT)
  - Example 3: Full time sequence of sampling in Controls at Follow-Up
  - Example 4: Use of Cycle vs Treadmill for testing
  - Example 5: Burden of number of biospecimen samples per individual
- Changes in inclusion/exclusion criteria or safety procedures



Managing the MoTrPAC Consortium Project: An Ever-Evolving Project Management Landscape

CINDY STOWE, MPM

MICHAEL E. MILLER, PhD

DATA MANAGEMENT, ANALYSIS AND QUALITY CONTROL CENTER, WAKE FOREST UNIVERSITY SCHOOL OF MEDICINE

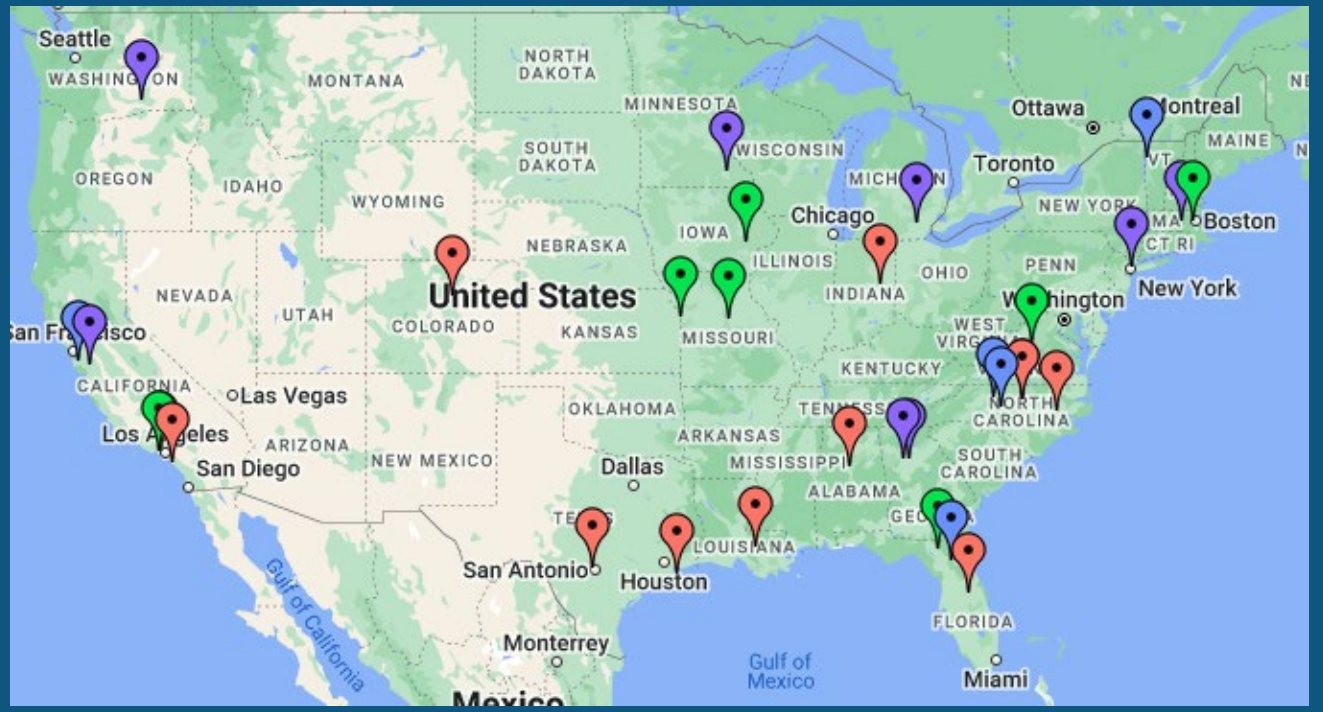
SCT
SOCIETY FOR CLINICAL TRIALS

Disclosures

- No Relevant Disclosures

MoTrPAC Project Management

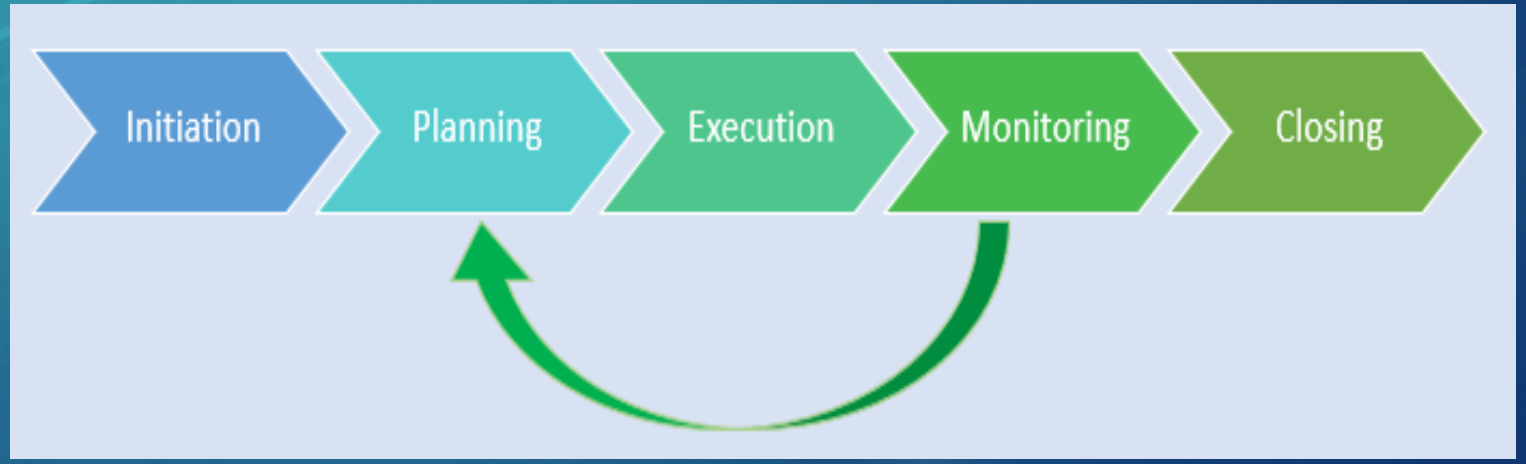
- Managing the timing/staging of multiple concurrent studies requires multiple levels of integration
  - Animal studies – two with one added after initial studies completed (overlap with human studies)
  - Adult studies – Sedentary and Highly Active
  - Pediatric studies – Similar to adult with nuanced changes



- Managing collaboration of ongoing activities at different entities with different responsibilities/scope of work
  - Varying goals/priorities of the CCC, Bioinformatics Center, Chemical Analysis Sites, Biorepository, Pre-Clinical Animal and Clinical Sites

Ever-Evolving Project Management

- Management of a project is composed of:
  - Multiple project phases
  - Many processes for organization and operations within a project
- Effective project management enables a project to:
  - Manage constraints of scope, time, cost/resources, and quality
  - Meet goals
  - Respond to changes
  - Deliver products



Initiation and Planning Phases

Initiation Phase Definition

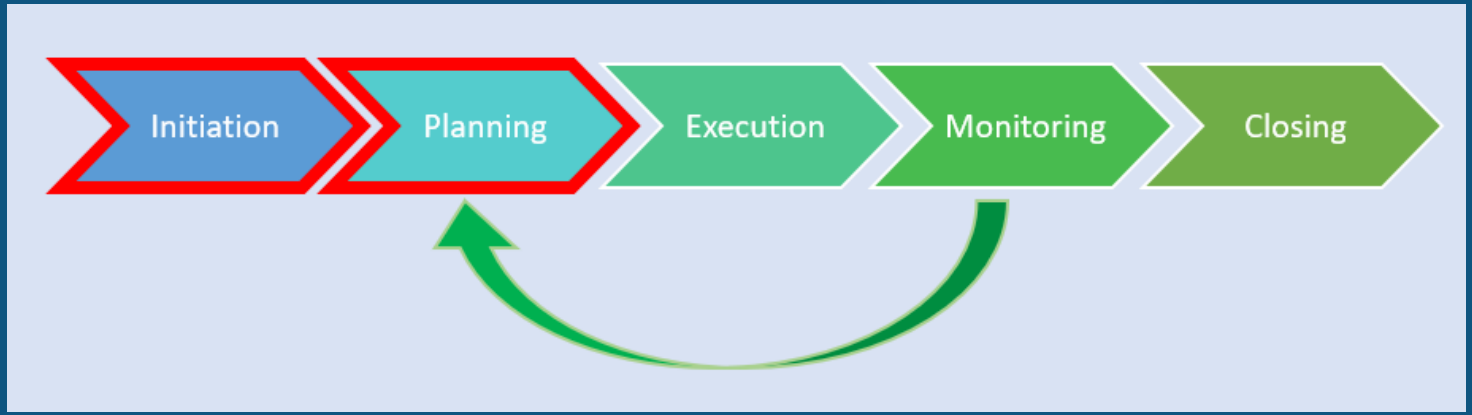
- Starting the Project
- Defining the Project
- Establishing the Scope
- Obtaining Authorization to Start

Planning Phase Definition

- Organizing and preparing
- Refining Objectives
- Operationalizing the plan

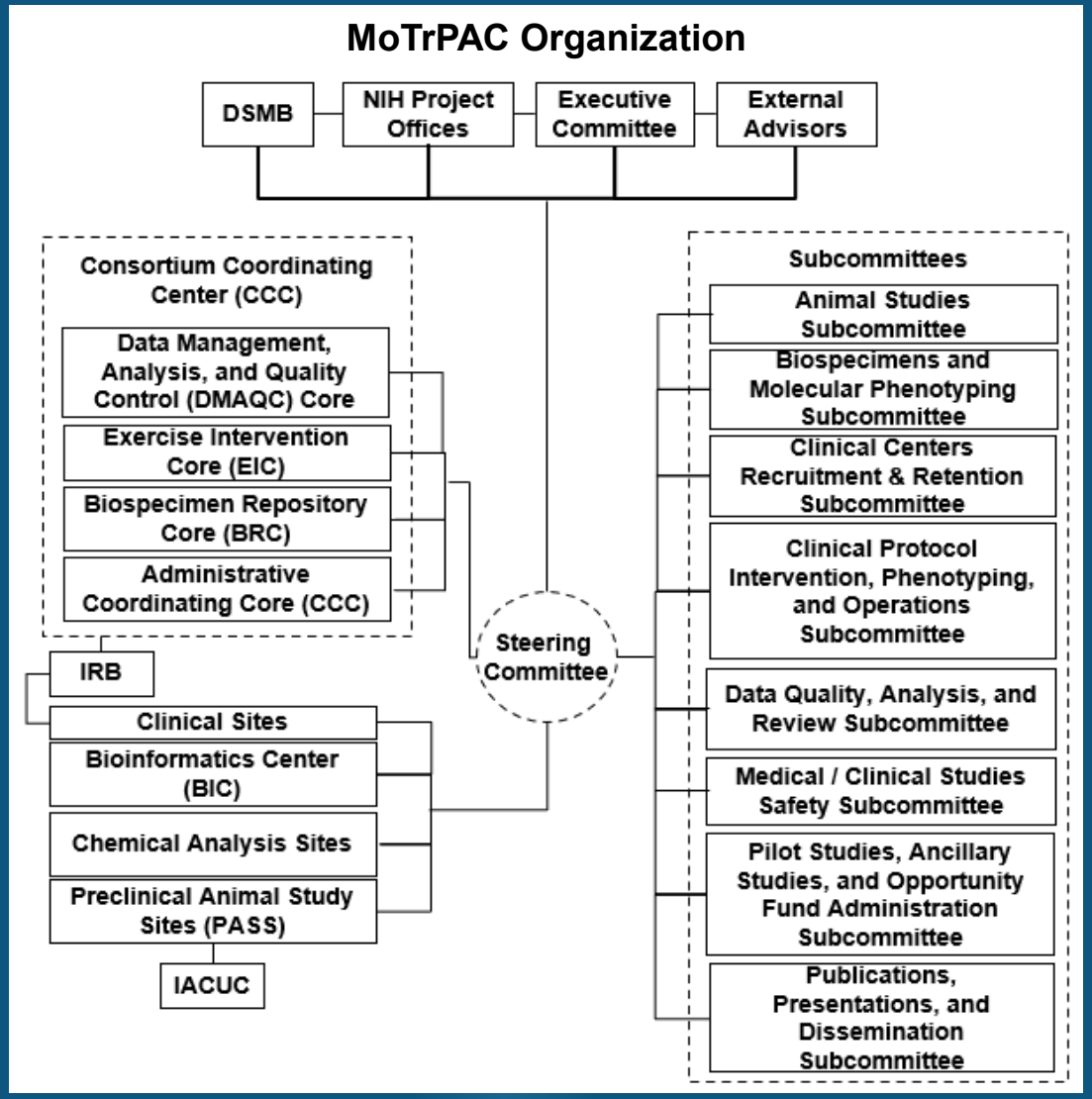
Included in these Phases

- Protocol development
- Single IRB considerations
- Communication plan
- MOP development
- Data collection needs (CRFs)
- External Data requirements



Initiation and Planning Phase Challenges

- Many, many stakeholders with many ideas about protocol content
- Terminology – everyone has their own "language"
- Broad scope that had to be narrowed
- Balancing Scope with the constraints of time and resources/cost

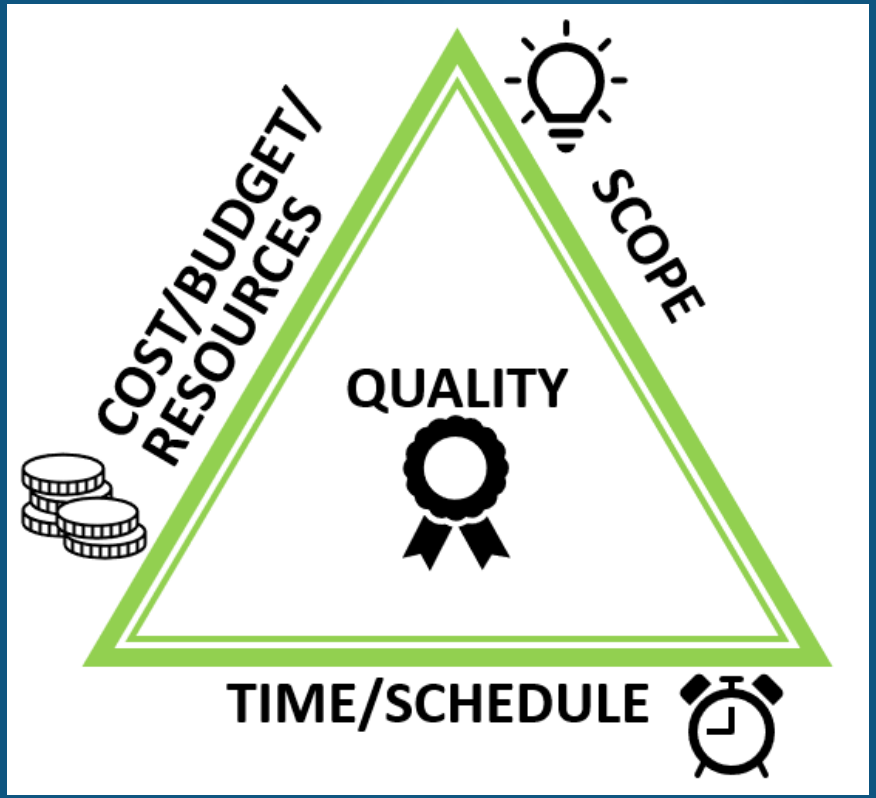


Initiation Phase Lessons Learned

- Requires coordination and effective communication with other stakeholders (NIH, CAS, PASS, CCs, BIC)
- Tools to use
  - Consortium Policies and Procedures
  - Communication Plan
  - Central Website
  - Basecamp
  - Slack
  - Google Drive, docs, sheets

Planning Phase Lessons Learned

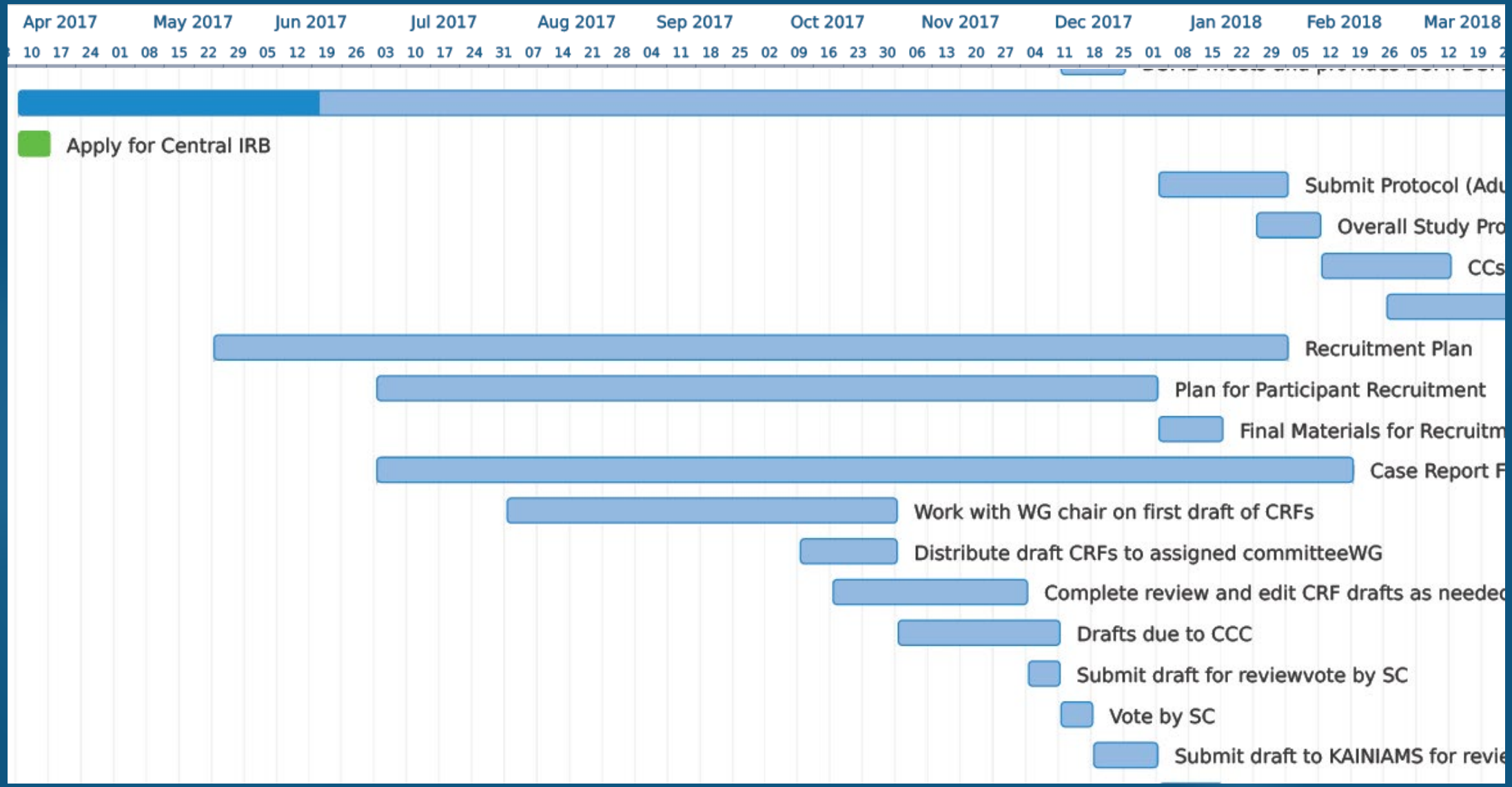
- Requires organization/time
- Change happens
- Must involve balancing constraints
- Tools to Use
  - Scope Statement
  - Timeline
  - Gantt
  - RACI
  - Project Management Software
  - Change Management Plan



Timeline

Category	Phase needed	Activity	Begin	End
Data	Phase 1	*Develop PID and Sample ID structure • What will be the components of the ID structure? Is site imbedded in some manner? • How many do we need? • How many digits? Check digits? • Same for Sample Tracking?	2/1/2017	4/15/2017
Protocol	Phase 1	Develop Protocol	2/1/2017	5/1/2017
Sample Tracking	Phase 1	*Program Repository Tracking System (sample collection, shipping, and storage) • Details on integration of software with the Biorepository software? >>Assuming Phase 1 starts in July	3/1/2017	6/1/2017
CRFs	Phase 1	*Develop and finalize Case Report Forms (by committee) for Phase 2 use • How many forms will there be? • Will there be translations to other languages? • In what format will we get each type of data? >>Assuming that they start Phase 2 in October	4/1/2017	6/1/2017
MOP	Phase 1	MOP Development	5/1/2017	6/1/2017
Admin	Phase 2?	Get started with Single IRB agreements	2/17/2017	7/1/2017
Sample Tracking	Phase 1	*Testing of Repository Tracking System	6/1/2017	7/1/2017

Gantt Chart


Apr 2017
May 2017
Jun 2017
Jul 2017
Aug 2017
Sep 2017
Oct 2017
Nov 2017
Dec 2017
Jan 2018
Feb 2018
Mar 2018
10 17 24 01 08 15 22 29 05 12 19 26 03 10 17 24 31 07 14 21 28 04 11 18 25 02 09 16 23 30 06 13 20 27 04 11 18 25 01 08 15 22 29 05 12 19 26 05 12 19
Apply for Central IRB
Submit Protocol (Adu
Overall Study Pro
CCs
Recruitment Plan
Plan for Participant Recruitment
Final Materials for Recruitm
Case Report F
Work with WG chair on first draft of CRFs
Distribute draft CRFs to assigned committeeWG
Complete review and edit CRF drafts as needed
Drafts due to CCC
Submit draft for reviewvote by SC
Vote by SC
Submit draft to KAINIAMS for revie

RACI – Responsible, Accountable, Consulted, Informed

Tasks	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W
Accelerometry API/Download (webhook)				R		A												R			I		
Metadata API	C					A									I	I		R		C	I		
Purge of test data	I					I									R	RA				C	I		
CRF SAS queries							R			R	R			R		R				C	C		
Transfer of data to BIC	C					RA	R	R		I	I			I		C				C	C		
PROMIS scoring data storage															R	A				C			
DXA setup of backend (upload functionality)						C									R						I		
DXA storage						C									R						I		
Accelerometry QC			A	R																	C		
Accelerometry Summary files (create)			A	R																			
Randomization Tables - Pediatric			A																			I	
Reports - DSMB (overarching reports)	RA		R				R			R	R	R		R						I	I	I	I
Sample Selection for LCBR (to CAS)	A		R				R													C			
Explore PASS ad hoc QC issues		A								R	R												
Reports - CRF QC (Clinical QC)		A										R								C	C		
BIC data hub review		R				A																	
Committee - Ancillary Studies		RA																		C			
Abnormal ECG data dictionary (classification)							R	R												A			
Abnormal ECG QC data transfer						RA																	
Accelerometry data transfer (1 min)						RA															I		
Accelerometry data transfer (summary)						RA															I		
Accelerometry Site Setup (acccounts, MOP, etc.)						A														C	R		
ASA24 download (cumulative)												RA											
ASA24 QC							C					RA										C	
ASA24 storage (files)												RA											
ASA24 storage space acquisition						RA																	
ASA24 transfer						RA																	
Data Documentation - change log							RA																
Data Documentation - data dictionaries	C							R															C
Data Documentation - file naming						A		R															

Execution and Monitoring Phases

Execution Phase Definition

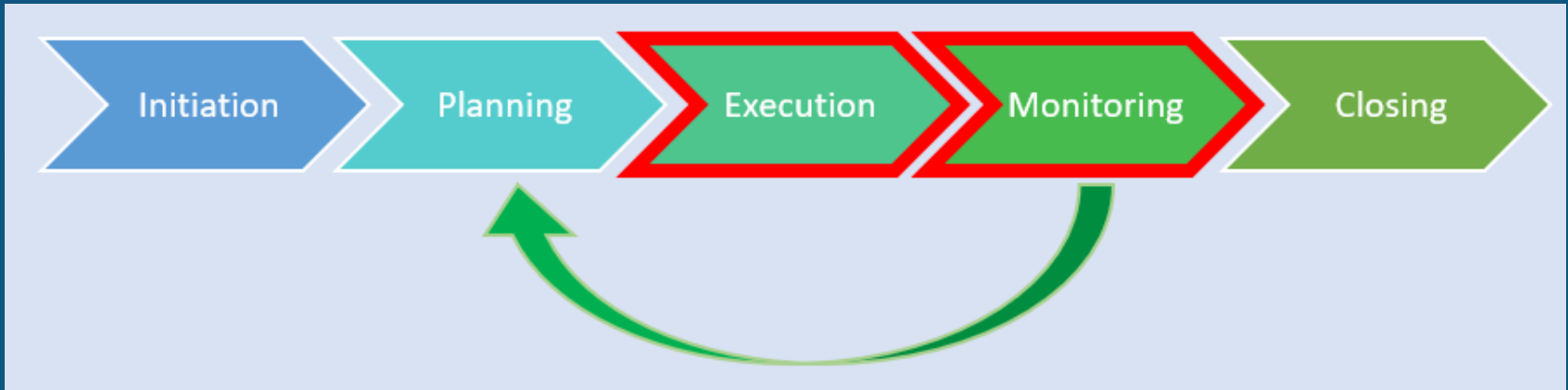
- Carrying out the work

Monitoring Phase Definition

- Track, review and regulate the project

Included in these Phases

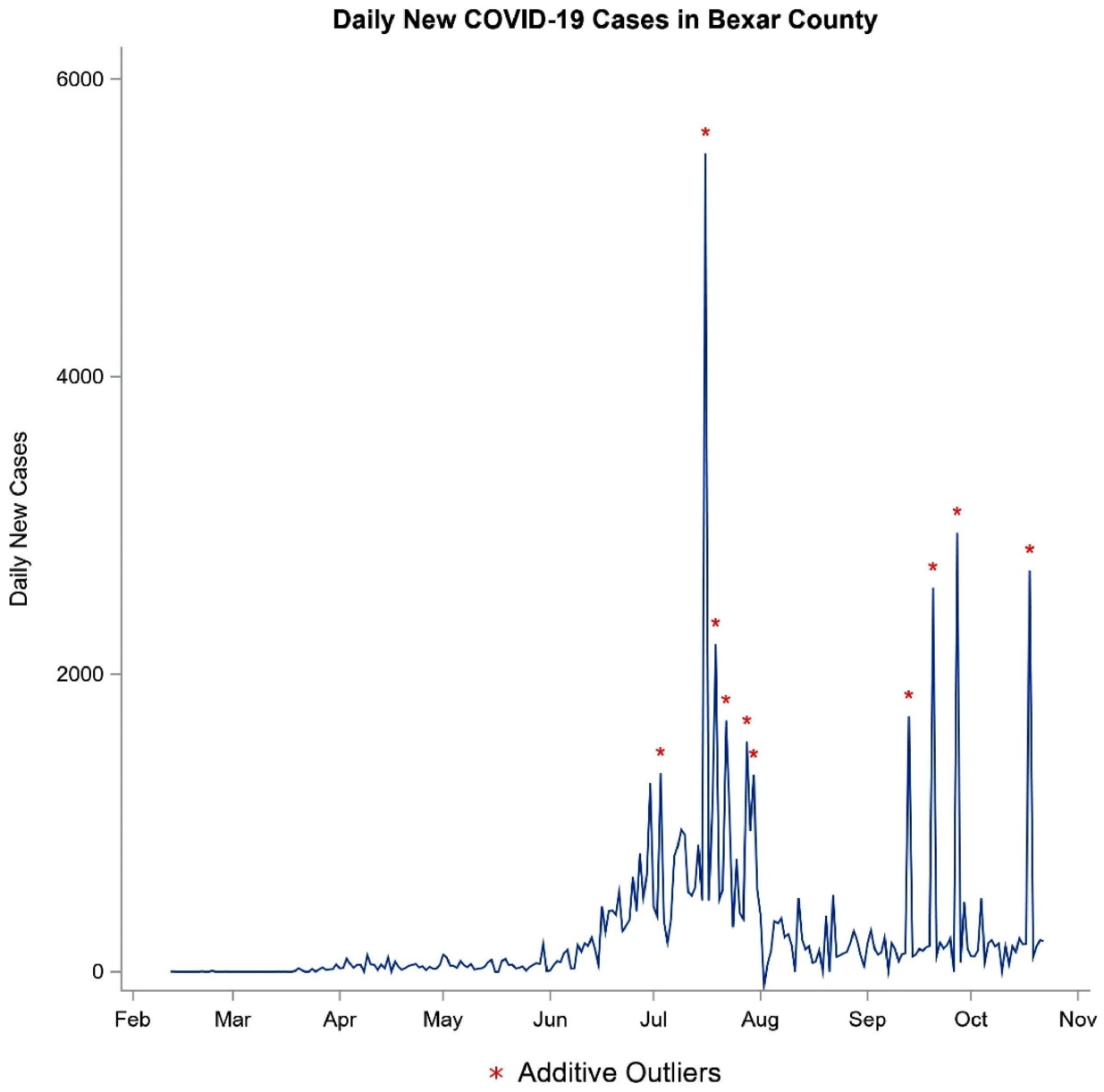
- Data Collection
- Study monitoring and reporting
- Data QC
- Sample Tracking



Execution and Monitoring Phase Challenges

- COVID Management mitigation
- Collection, tracking and analysis of hundreds of thousands of biological samples between multiple sites (batching metadata, selection of tranches)
- Complex integration of ancillary study sample collection processes
- Staffing changes and the virtual work environment

Tracking COVID rates



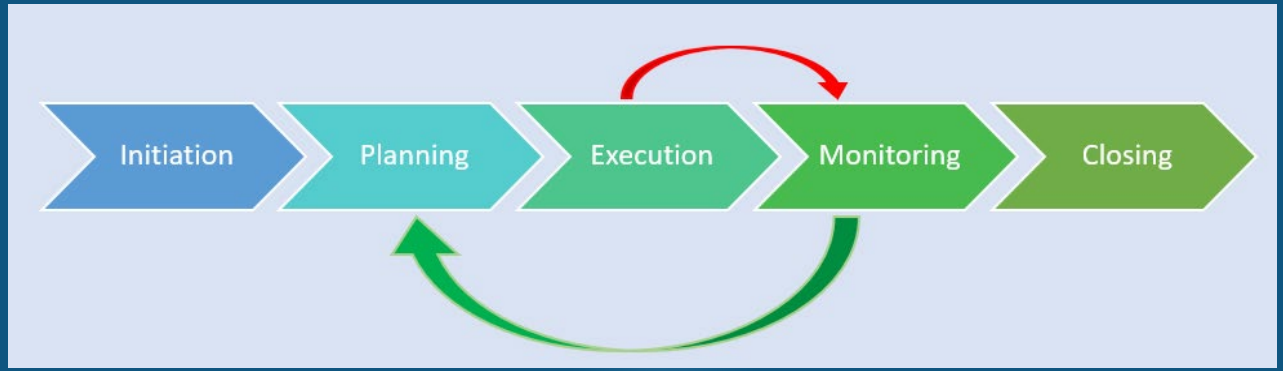
George SM, Chen H, Miller ME, et al. Rapid report on using data to make standardized decisions about enrollment during the COVID-19 pandemic: perspectives from the MoTrPAC study. *Ann Epidemiol*. 2021;62:19-21. doi:10.1016/j.annepidem.2021.04.016

MoTrHisto ancillary

- Built the system to be flexible enough to easily integrate ancillary studies into the sample tracking system already in use

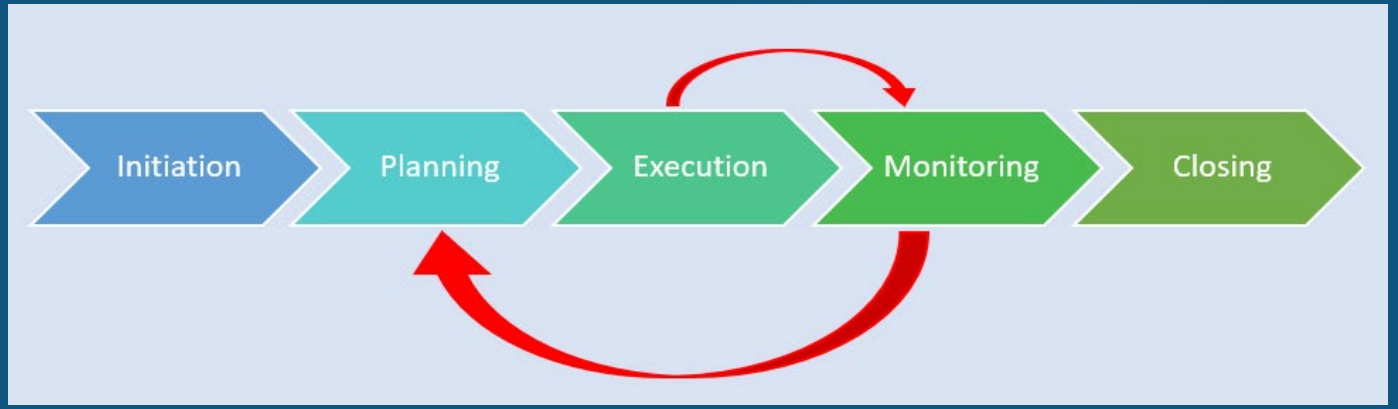
Execution Phase Lessons Learned

- Standardization is key to success
- Ability to integrate and being fluid is necessary
- Change is inevitable
- Tools to Use
  - Change Control process
  - Documentation of change approvals
  - Communication plan
  - Committees/WGs



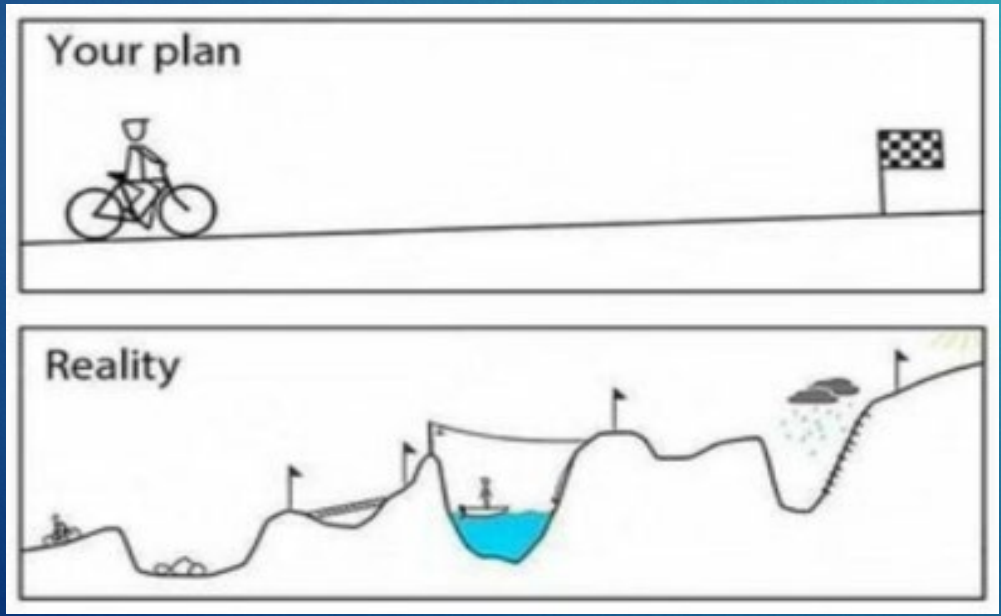
Monitoring Phase Lessons Learned

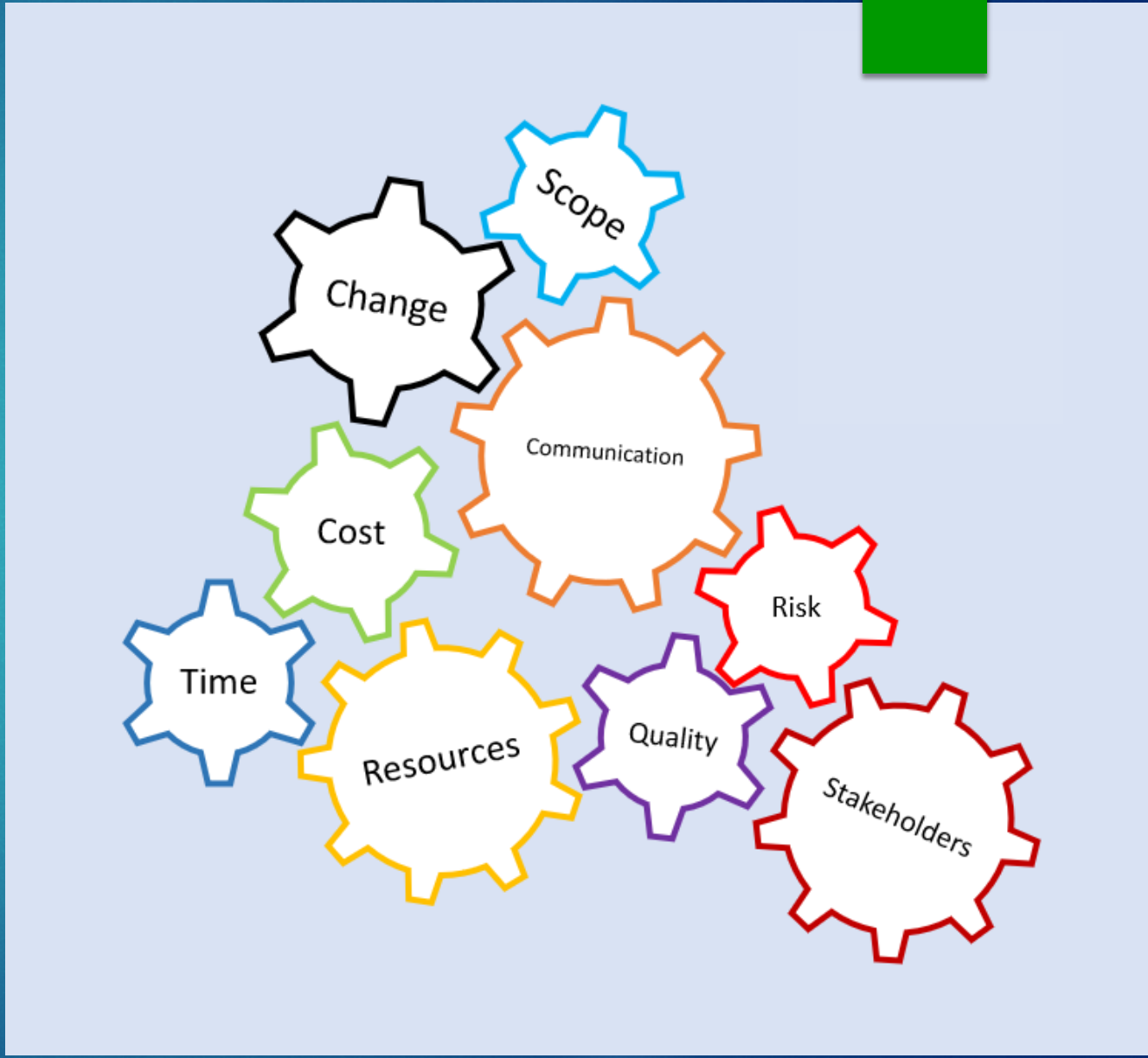
- Have a plan
- Tools to Use
  - Data Safety Monitoring Plan
  - Data Sharing Plan
  - Site Visits
  - Data Queries/Bulk QC
  - QC WGs Monitoring
  - Dynamic Reporting
  - Report A Problem/Issue Logs/Exception Logs
  - Online Tools and Tracking Systems (PPD, Ancillary, Overrides, safety)



Overarching Lesson Learned

- Clear communication is critical
- Integration is important
- Flexibility is fundamental





MoTrPAC CCC Team


Jerry Barnes


Kate Boyer


Karyn Esser


Mike Miller


Russ Tracy


Jack Rejeski


Pete Brubaker


Brandon Bukas


Haiying Chen


Shyh-Huei Chen


Cindy Stowe

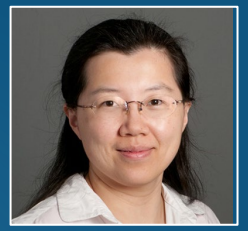

Scott Rushing


Jess Rooney


Jane Lu


Peter Durda


Shannon Emilson


Fang Chi Hsu


Byron Jaeger


Sandra May


John Nichols

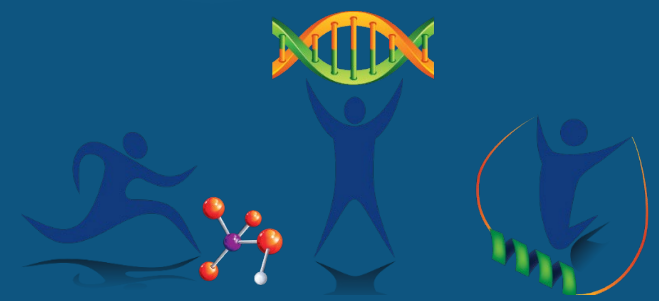

Barb Nicklas


June Pierce


Eric Reynolds


Joe Rigdon


Jeremy Rogers




Santiago Saldana


Courtney Simmons

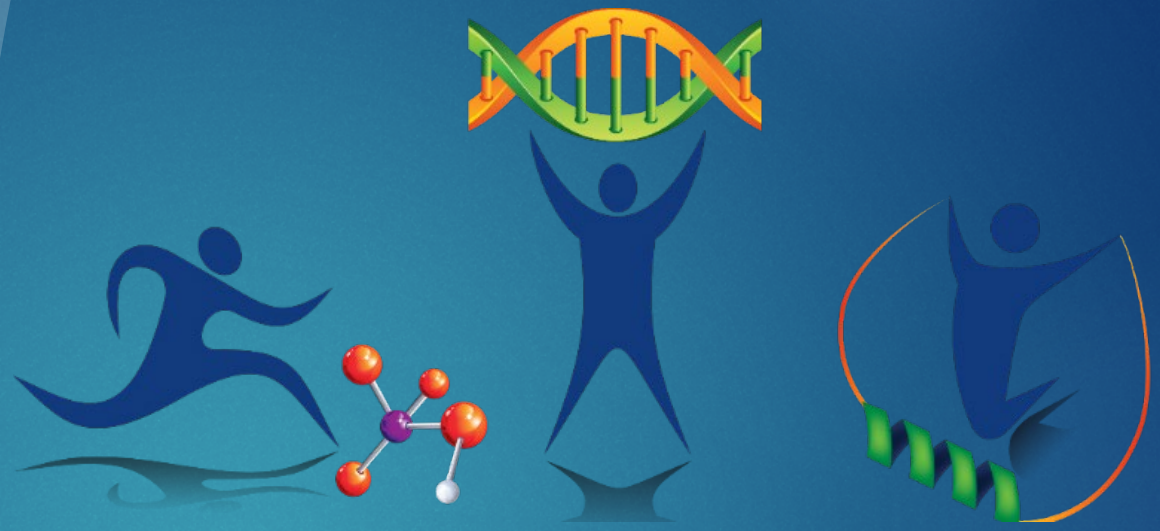

Jenn Talton


Mike Walkup


Christopher Webb


Sawyer Welden


Marilyn Williams



Clinical Information Technology Challenges and Solutions

Scott Rushing, Data Management, Analysis and Quality Control Center,
Wake Forest University School of Medicine

Disclosures

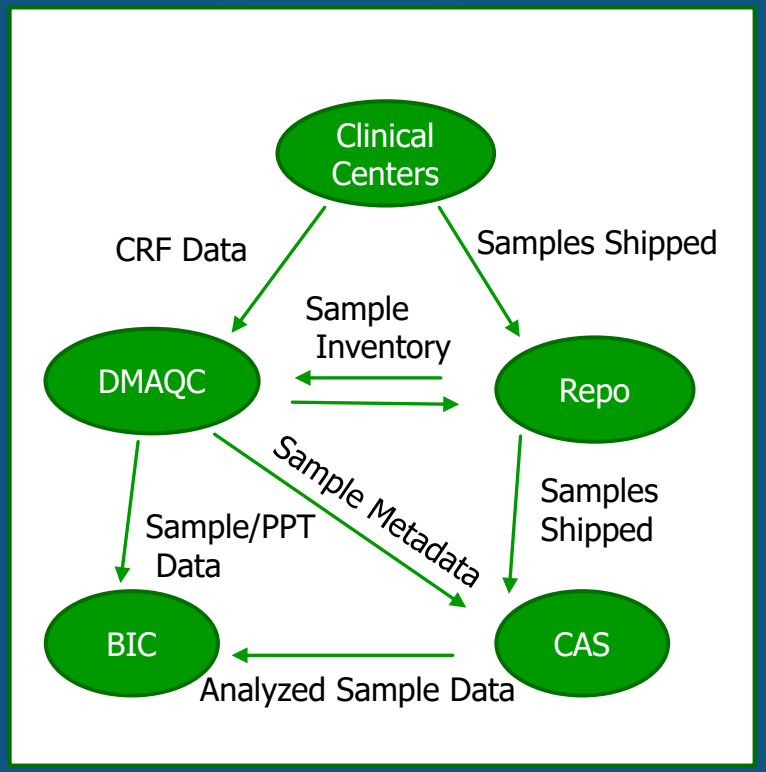
- No Relevant Disclosures

Acronyms

- DMAQC - Data Management and Quality Control
- BIC - BioInformatics Center
- BRC – Biorepository Center
- CC - Clinical Centers
- CAS - Chemical Analysis Sites
- PID – Participant Identifier

Responsibilities for DMAQC

- Responsible for
  - Developing the data management system that collects 100+ CRFs, ensuring the quality and integrity of the data
  - Developing a comprehensive biospecimen management system
  - Transferring data to the BIC to support main trial releases and QC activities
  - Providing a means for analytic groups to obtain sample metadata needed to stage and QC experiments
  - Integrating with the repositories for sample management and reconciliation
  - Integrating with the sites to gather local machine-driven data



IT Challenges for DMAQC

- Covid Suspension
- Eligibility criteria 1 person/household
- Accelerometry/DXA data
- Biospecimen Tracking
- Data Sharing

COVID Suspension from an IT Standpoint

- **Challenge:** Need to distinguish 'incomplete' data due to COVID versus other reasons and have the ability to prepare a separate datasets of pre/post COVID
- **IT Solution:** Decided to systematically close those existing logical sites and open new Post-Covid sites for each as a way to be able to more easily work with pre-suspension versus post-suspension data
  - In the data management system, participants are associated with a clinical site identifier. Those siteIDs align with the clinical locations.
  - Created new clinical SiteIDs that were linkable back to the Pre-Covid clinical SiteID --- Introduced the concept of 'hubs'
  - Participant IDs were random with no clinical site IDs embedded that we had to deal with

COVID Suspension from an IT Standpoint

This caused issues downstream:

- Creating a second clinical site identifier for each clinic location meant modification of numerous reports where combined totals were needed
- We had pre-assigned bulk PIDs to clinical sites (due to some external integrations) which meant we needed to reallocate unused PIDs across the new clinical siteID
- We also needed to map any external data that used clinical siteIDs back to the clinical siteIDs of the new sites
- All our data sharing processes had to take these changes into account

Eligibility: 1 per household/family

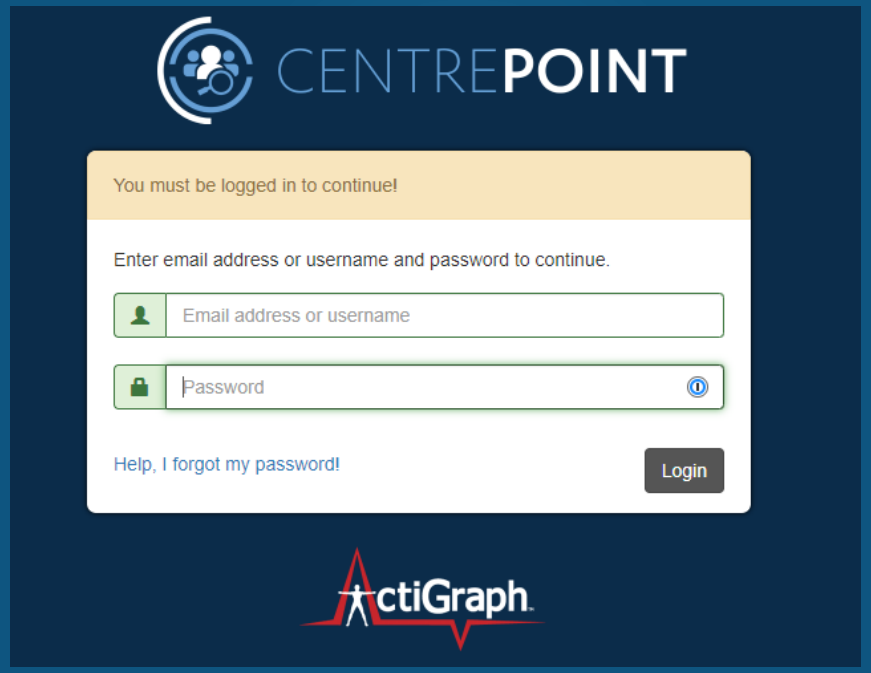
- **Challenge:** Needed to obtain address information from the participant that is easily comparable to other participants due to misspellings and abbreviations

- **IT Solution:**
  - Integrated with the USPS Address validator
  - Call an USPS API, passing in the address as we received it
  - That returns JSON object providing a standardized address object
  - We then do a comparison of that to other enrolled participants and alert the site if that household is already assigned to a participant
  - In some cases (apartments/dorms) we have the ability to override


```
{
    "addresses": [
        {
            "id": "us_ver_c7cb63d68f8d6",
            "recipient": "LOB.COM",
            "primary_line": "210 KING ST STE 6100",
            "secondary_line": "",
            "urbanization": "",
            "last_line": "SAN FRANCISCO CA 94107-1728",
            "deliverability": "deliverable",
            "valid_address": true,
            "components": { … },
            "deliverability_analysis": { … },
            "lob_confidence_score": { … },
            "object": "us_verification"
        }
    ],
    "errors": false
}
```

External Data Source: Accelerometry Data

- **Challenge:** Accelerometry data collection and management using local software at the sites is very problematic, need to streamline
- **IT Solution:** Enter the cloud, Actigraph CENTREPOINT Cloud solution
  - Ability to reduce the IT burden on the sites
  - Ability to standardize device configuration and data collection
  - Ability to easily access the data in the cloud



External Data Source: Accelerometry Data

Implementation Details:

- At DMAQC, we needed to account for each wearing of the device by participants, so we needed additional QC:
  - Established data collection procedures utilizing both a CRF as well as CP exports to monitor the process
    - Would download summary 1 minute epoch, milestones, and other subject data from CenterPoint

- Merge those with our CRF data that tracks participant visit and CRF data and run algorithms to determine where we had invalid or missing data uploads
- Produce reports for the PMs and the Sites on what inconsistencies existed

External Data Source: Accelerometry Data

- **Common Issues:**
  - Devices were initialized prior to date of visit
  - Sites failed to set the device to "stop collecting" once returned
  - If the dates on the device didn't match the data on the CRF, it creates challenges trying to align the data
  - Each of these cases also created gaps of recorded non-wear time in the data that had to be accounted for
- **Solutions:**
  - Enabled milestones within CenterPoint
  - Modified the CRFs to identify more clearly the date related data (date initialized, date assigned, date returned)
  - Continued to remind CCs about the MOP for device collection

External Data Source: DXA Data

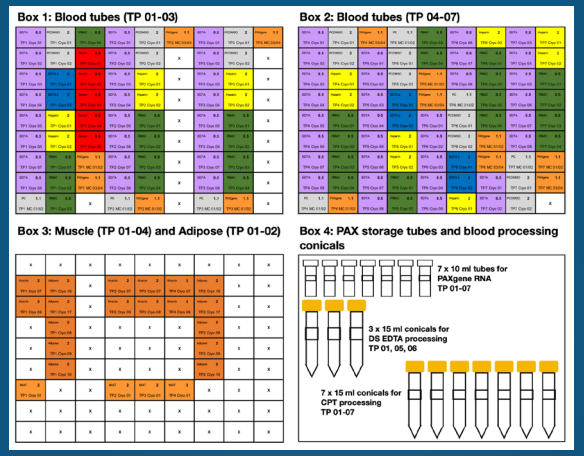
- **Implementation Strategy:**
  - Initial strategy was Checklist + Raw data upload from sites
  - We would process the uploads and compare RAW to what the Checklist said
  - Differences needed to be resolved by the site
- **Implementation Challenges:**
  - Vendor upgrades broke the custom upload procedures which lead to tedious work to identify the issues and fix
  - The complexity of each vendors data made it very challenging to align the data in a way to use it
  - Sites would sometimes fail to properly export data

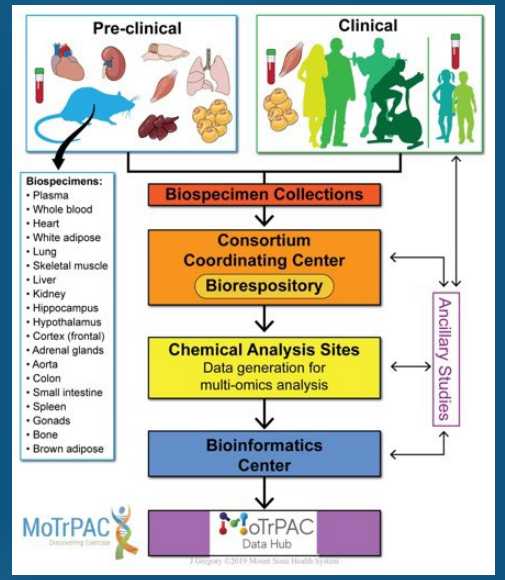
External Data Source: DXA Data

- **Resolution:**
  - After many discussions
    - Decision to stop relying on the raw data
    - Created stand-alone CRF which recorded key data
    - Each site continues to maintain local raw DXA data so that we can potentially pull those in the future

Biospecimens Management

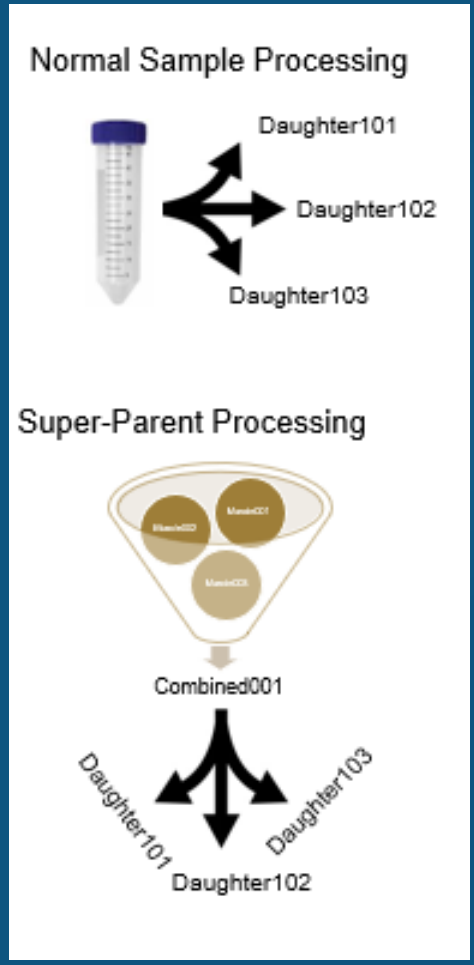
- Biospecimen management requires accounting for:
  - Over 170 vials per participant
  - over 70 sample types
  - Completeness of samples collection (parents)
  - Addition of sub-samples (daughters)
  - Collection of all required sample metadata
  - Facilitate shipping of samples between BRC/CC/CAS
  - Facilitate sharing of shipping and sample data to the BIC and CAs





Biospecimen Management

- Sample lineage
  - Many samples are processed into daughters
  - Some parents are combined (super-parents) and then split into daughters
  - Metadata from the parents are needed to be associated with the daughter
  - Need to generate and represent that lineage in the data

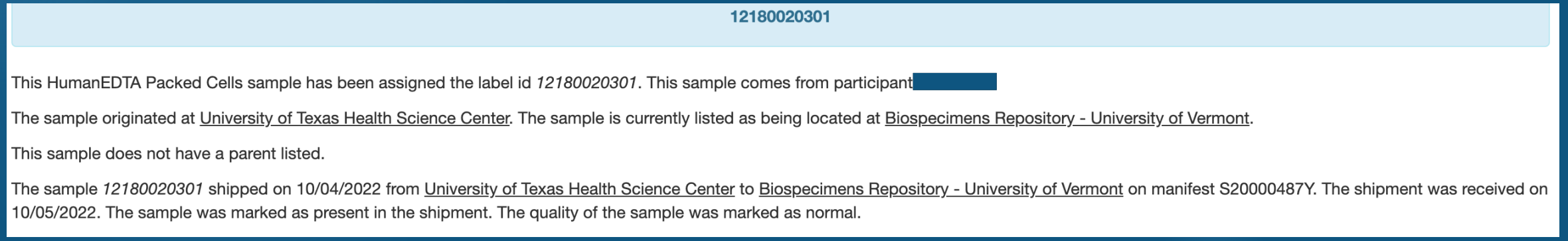


Biospecimen Management

- Common issues:
  - Shipments arrive at Repository with samples missing or with unexpected samples that were not record on the manifest
  - Because of the volume of work, CCs began to stage vials in advance which sometimes led to inaccurate reporting of what was or was not collected and when
  - The Initial shipping system assumed all samples for a collection would be shipped at the same time which was not the case

Biospecimen Management

- DMAQC was always monitoring and improving the system
  - Re-engineered the collection and reconciliation systems to shorten the amount of time it took to enter the data.
    - Was vial by vial entry; Now is exception driven
  - Developed integrated tools for DMAQC, Biorepository and Sites to use to correct shipping and reconciliation issues
    - Label History Tool



12180020301

This HumanEDTA Packed Cells sample has been assigned the label id *12180020301*. This sample comes from participant [redacted]

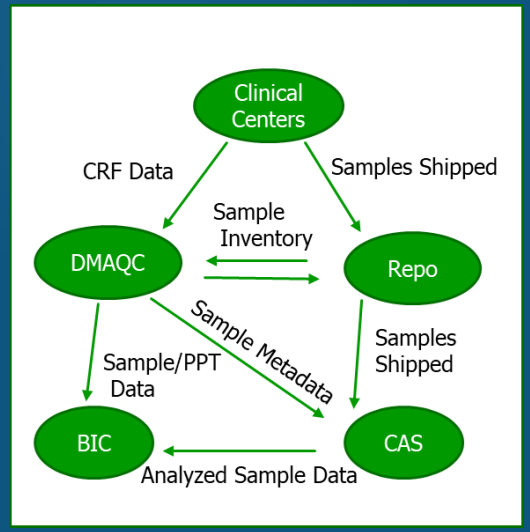
The sample originated at University of Texas Health Science Center. The sample is currently listed as being located at Biospecimens Repository - University of Vermont.

This sample does not have a parent listed.

The sample *12180020301* shipped on 10/04/2022 from University of Texas Health Science Center to Biospecimens Repository - University of Vermont on manifest S20000487Y. The shipment was received on 10/05/2022. The sample was marked as present in the shipment. The quality of the sample was marked as normal.

MoTrPAC Data Sharing Description

- Multiple MoTrPAC groups rely on data from DMAQC
  - Quarterly transfers sent to the BIC:
    - QC data needed by CAS to validate their analysis activities
    - Sample shipping information needed by the BIC to ensure we receive all analysis results
  - Tranches of complete data for specific participant groups sent to the BIC to support broader data sharing internally and externally to MoTrPAC
  - Sample metadata needed by CAS to stage their experiments
    - Accessed via secure APIs, CAS received CSV/XML/JSON formatted files based on shipments

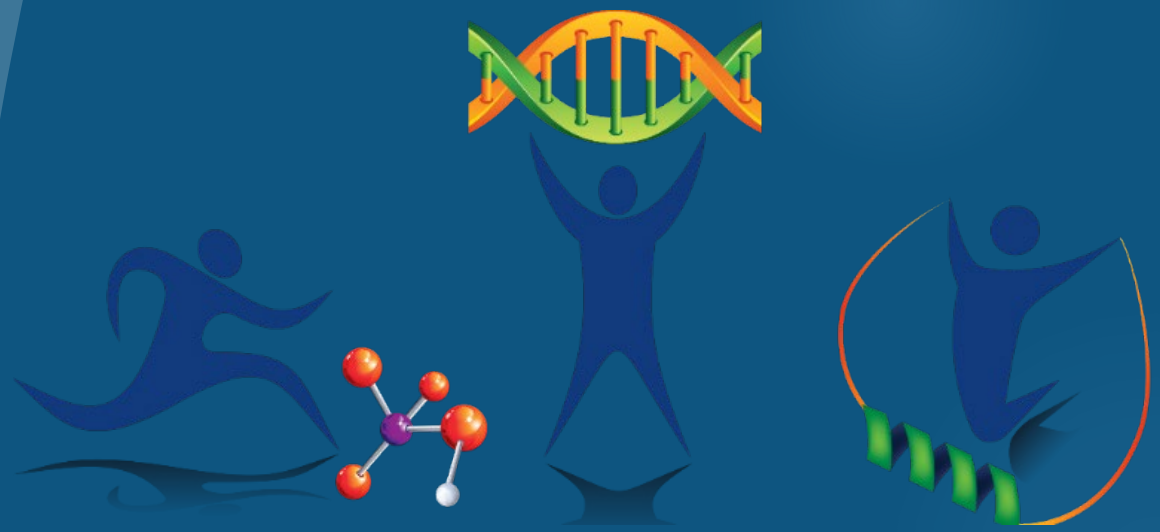


Challenging Issues With Data Sharing

- **Challenge:** MoTrPAC is sharing data in real-time, not at the end of the study when data cleaning is complete.
  - Decisions must be made for each type of transfer whether it will be run against a frozen copy of data or against live data?
- **IT Solution:** Sample shipping data and QC metadata files use live data, so we have the most up to date sample data included
  - Larger tranche transfers are complex and require additional time to complete all the final QC.
    - Implement a series of QC processes to identify if there are any outstanding issues
    - Request a freeze from the DBA; usually completed the next day.
    - But between the initial review and completion of data freezing, data can change causing additional QC steps to fail and requiring another QC review and freeze

Data Sharing Strategies in MoTrPAC

- **Further IT Solutions:**
  - A cross disciplined team developed an 'All Clear' query tool
  - We developed procedures for comparing previous output vs new to identify changes which were reviewed prior to transferring the data
  - Developed a data exception log to allow for documentation of known issues
- **Communication and Documentation**
  - ReadME files
  - Data dictionaries were developed to describe the raw output file contents
  - Established routine Zoom meetings with major groups we interact with to review recent transfers and any questions
  - Communicate regularly through email, Teams and Slack



Bioinformatics Challenges and Data Sharing Compromises

Matthew T. Wheeler, MD PhD

Assistant Professor of Medicine

Stanford University

SCT

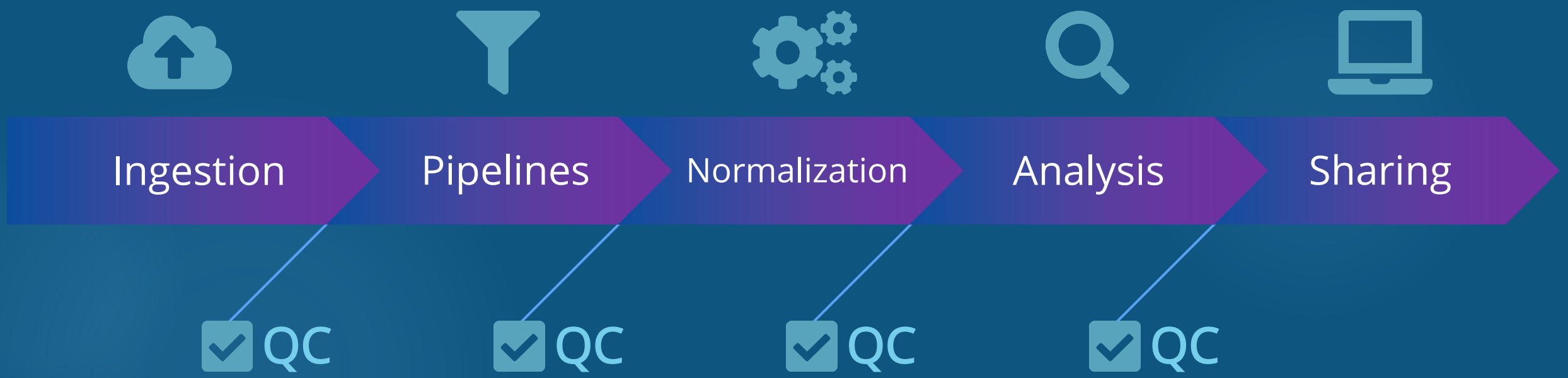
SOCIETY FOR CLINICAL TRIALS

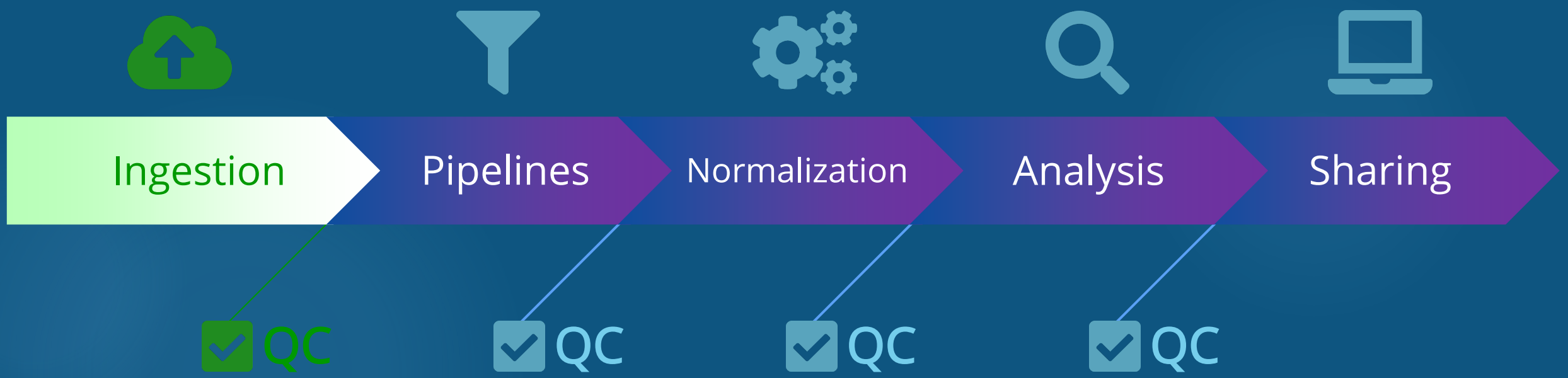
oTrPAC Bioinformatics Center

Disclosures

- No Relevant Disclosures

Data flow in MoTrPAC




Ingestion
Pipelines
Normalization
Analysis
Sharing
QC
QC
QC
QC

Ingestion Automation and QC

Transferred Data Audit System Application

Why do need automated tools?

identification

time

errors

When do we need it?

monitor data ingestion

Checking submitted data

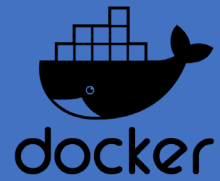
all omics

structured /unstructured

What do we audit?

python

serverless



docker

How do we automate?

Data Transfer Guidelines

- We enforce data transfer standards to ensure data quality
- Required folder structure and a required manifest file
- A new manifest must be uploaded with each data addition
- Data transfer is automatically QCd to assess adherence
- Iterative process with data submitters to develop and meet standards

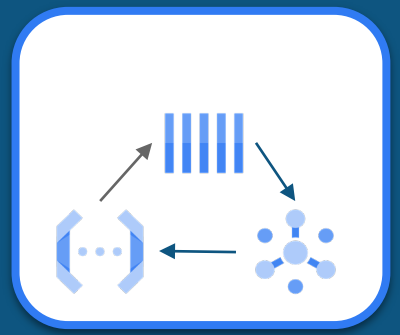
Cloud-native QC automation

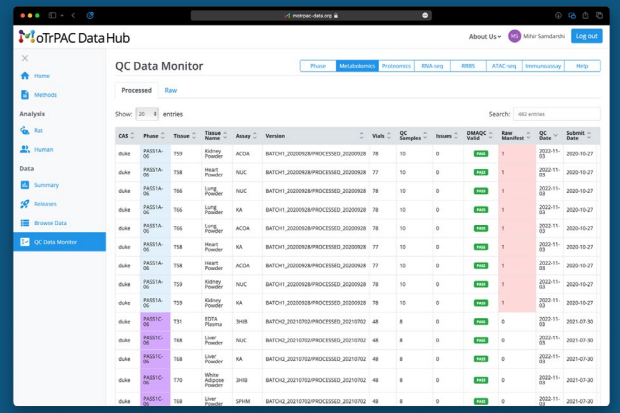
Automated QC pipelines to audit data ingestion and report QC statuses/results.

CAS uploads data and file manifest

Ingestion service gets automatic notifications

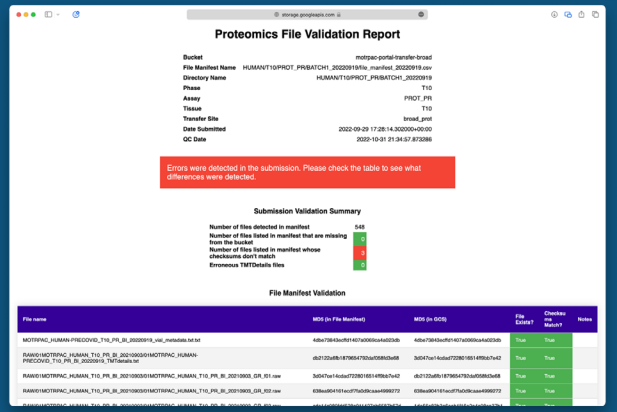
QC checks are queued to run after data submission



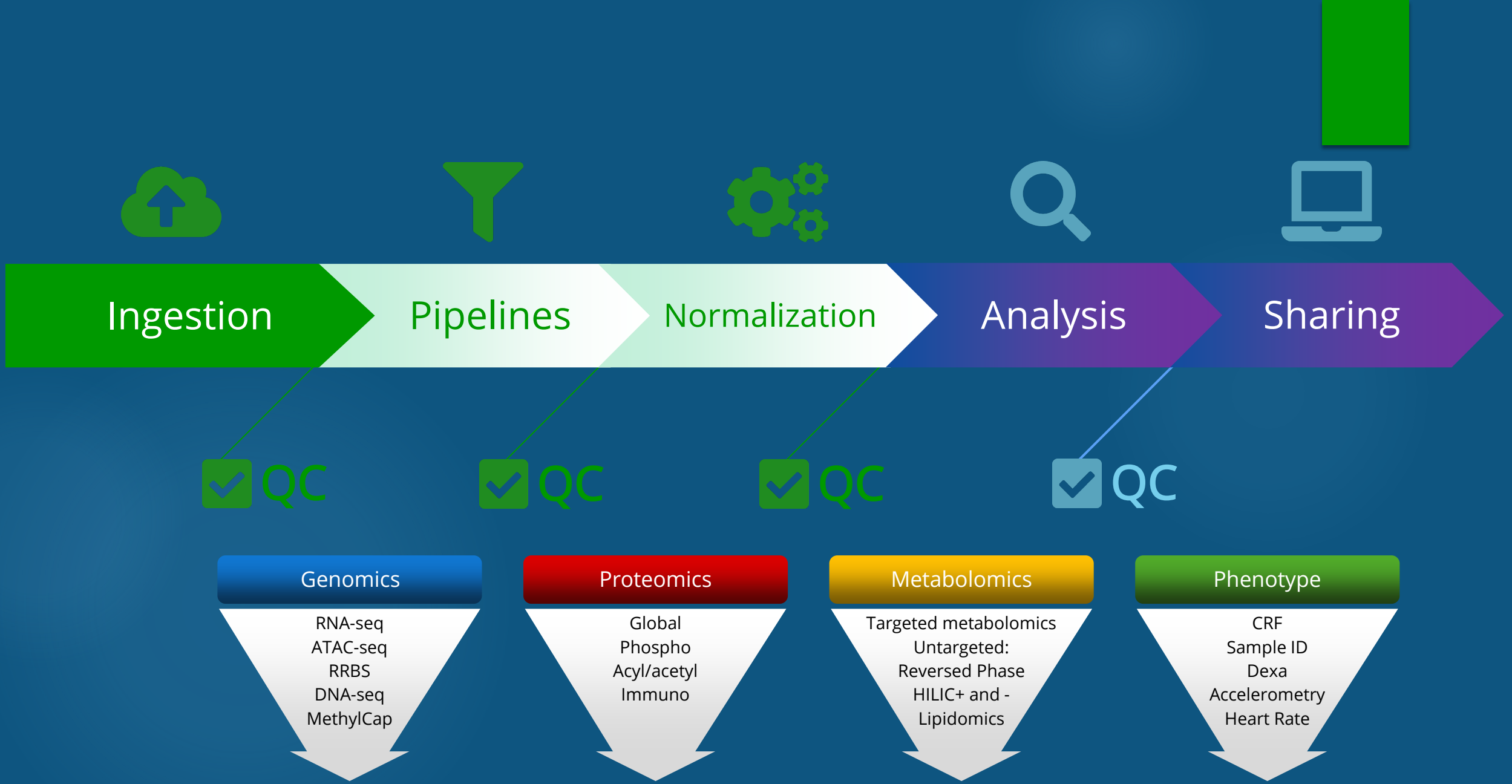


QC results automatically available on data portal

Notification of QC outcome via Slack

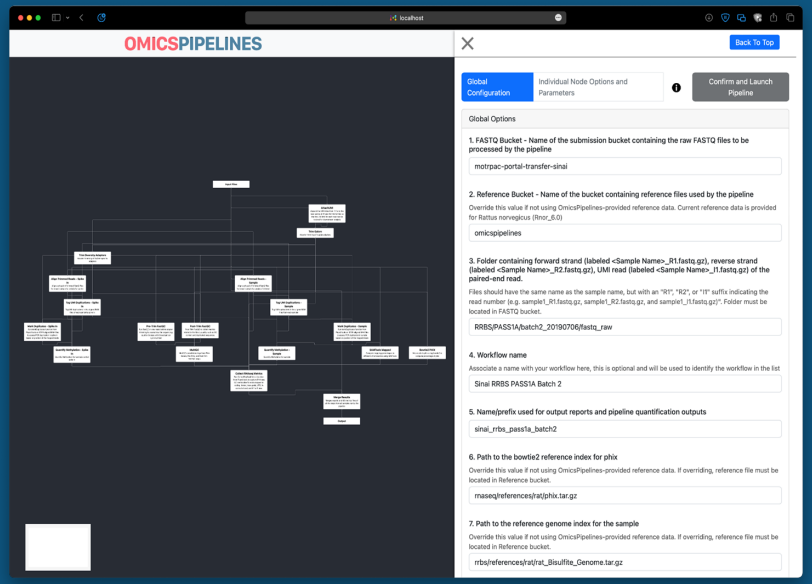


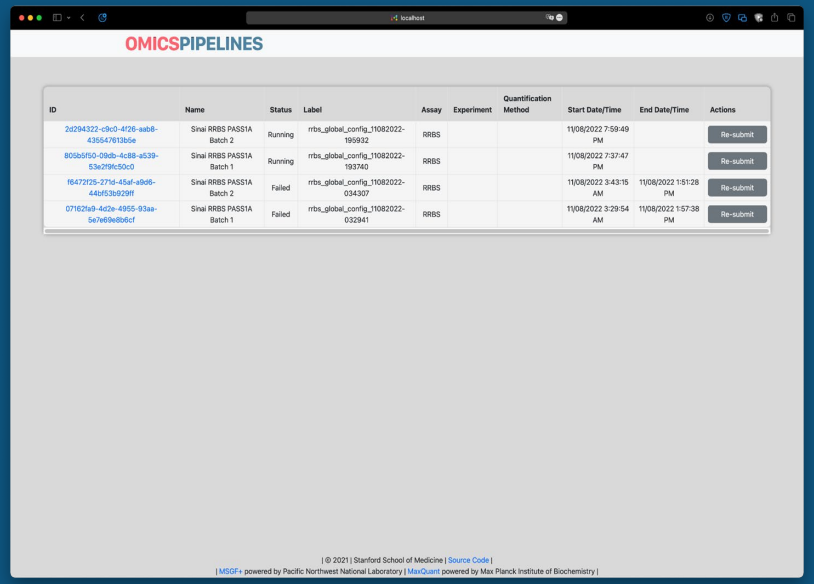
Rapid and detailed QC report for improved observability

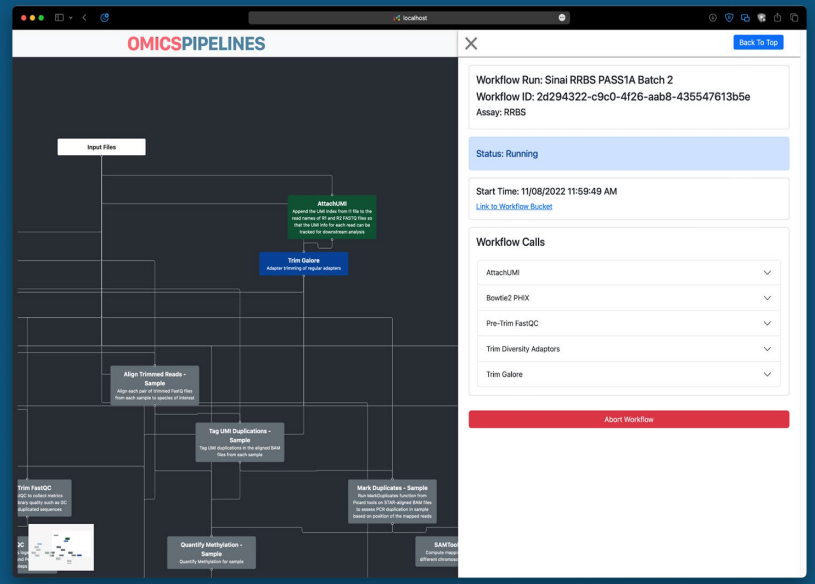

Ingestion
Pipelines
Normalization
Analysis
Sharing
QC
QC
QC
QC
Genomics
RNA-seq
ATAC-seq
RRBS
DNA-seq
MethylCap
Proteomics
Global
Phospho
Acyl/acetyl
Immuno
Metabolomics
Targeted metabolomics
Untargeted:
Reversed Phase
HILIC+ and -
Lipidomics
Phenotype
CRF
Sample ID
Dexa
Accelerometry
Heart Rate

OmicsPipelines

Simple, informative, and extendable web application for running cloud-based GET and Proteomics bioinformatics pipelines







Configure and run pipelines in under 5 minutes

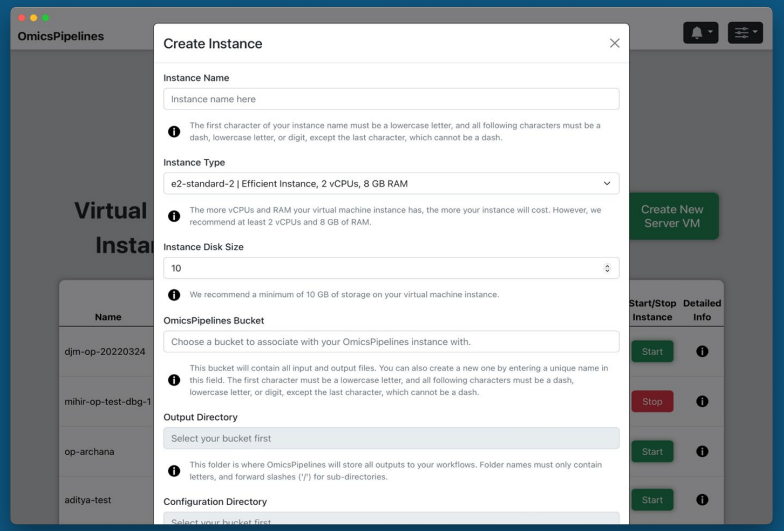
View and reuse previous pipelines, enabling reproducible data analysis

Gain real-time insights into running workflows, easier debugging of failed pipelines

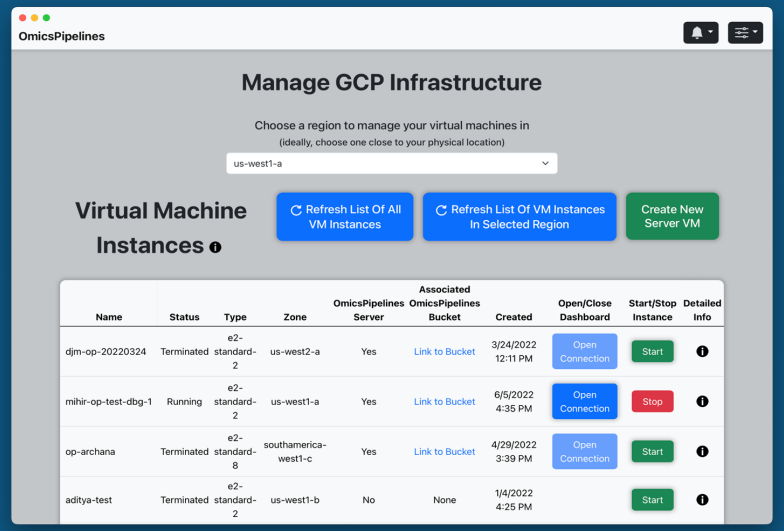
- Modern tech - written in Python and Typescript/React
- Easily extendable - adding a new workflow takes ~15 minutes as WDL files are automatically parsed

OmicsPipelines

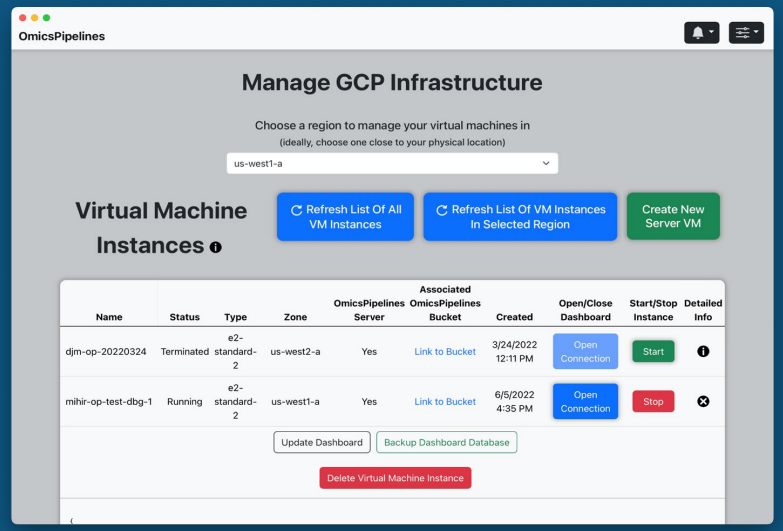
Lightweight desktop application designed to simplify creating, managing, and connecting to complicated cloud infrastructure needed to run cloud-based bioinformatics pipelines



Setup secure infrastructure in your own cloud account (Google or Amazon) with a single click



View existing infrastructure, and launch private, secure connections to virtual machines



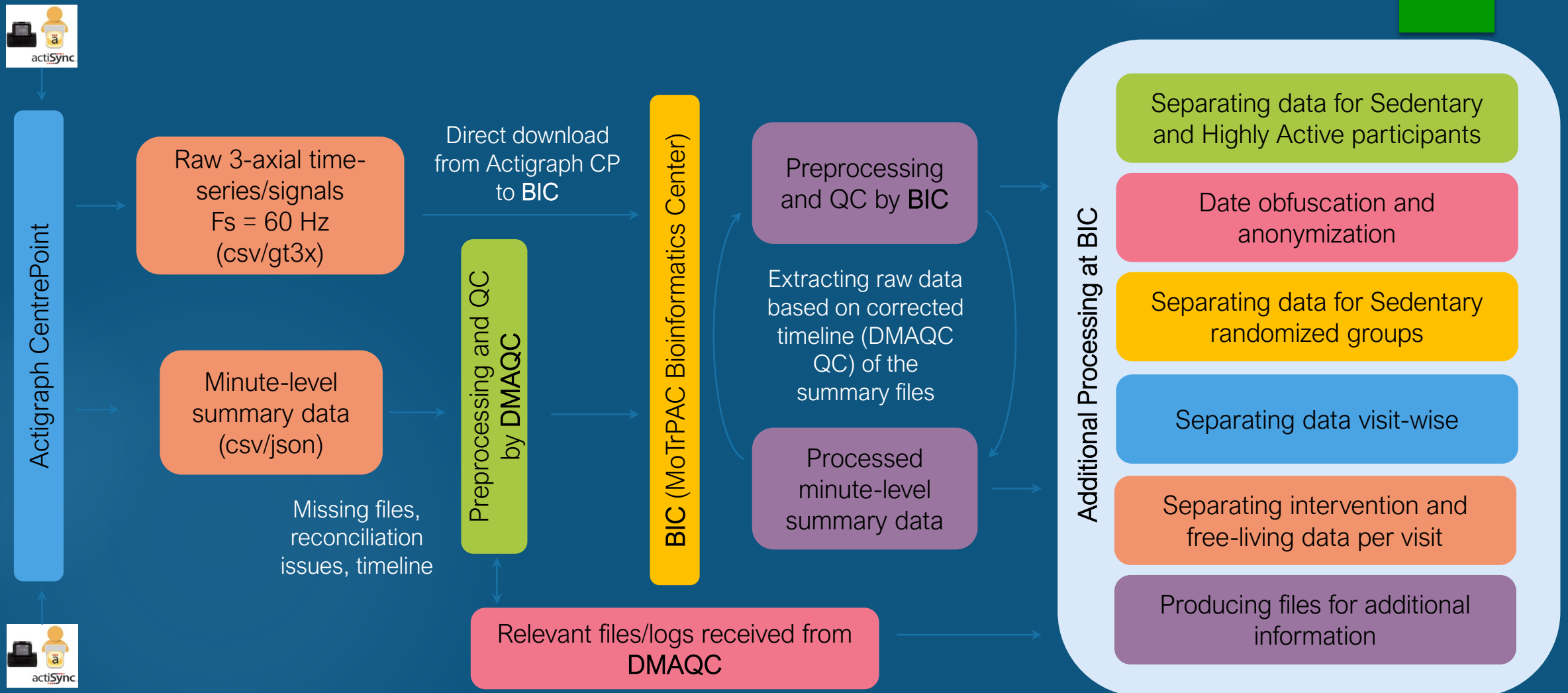
One-click actions for common tasks like starting, stopping, updating, or backing up virtual machines

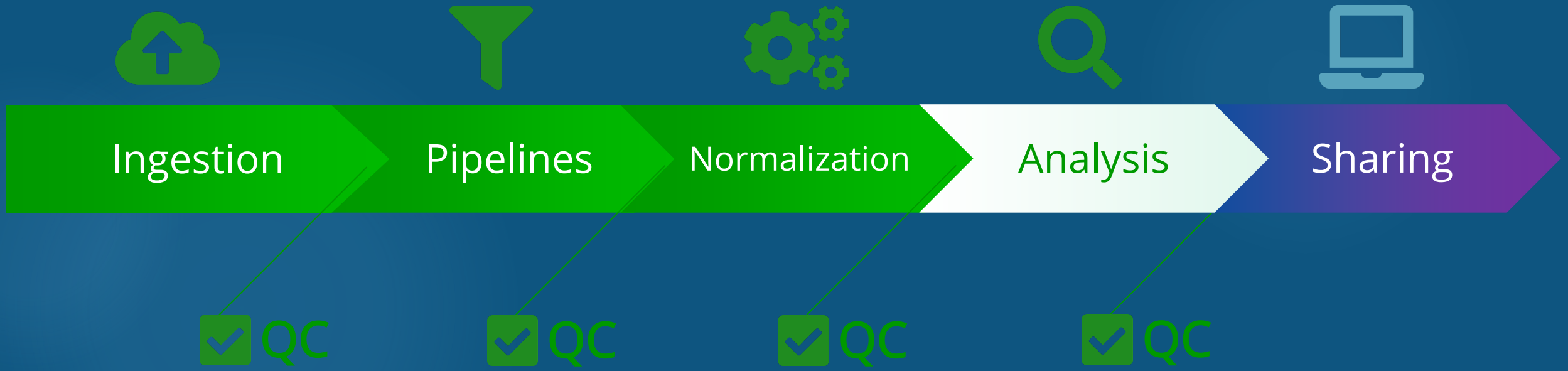
Heterogenous & High dimensional phenotype data

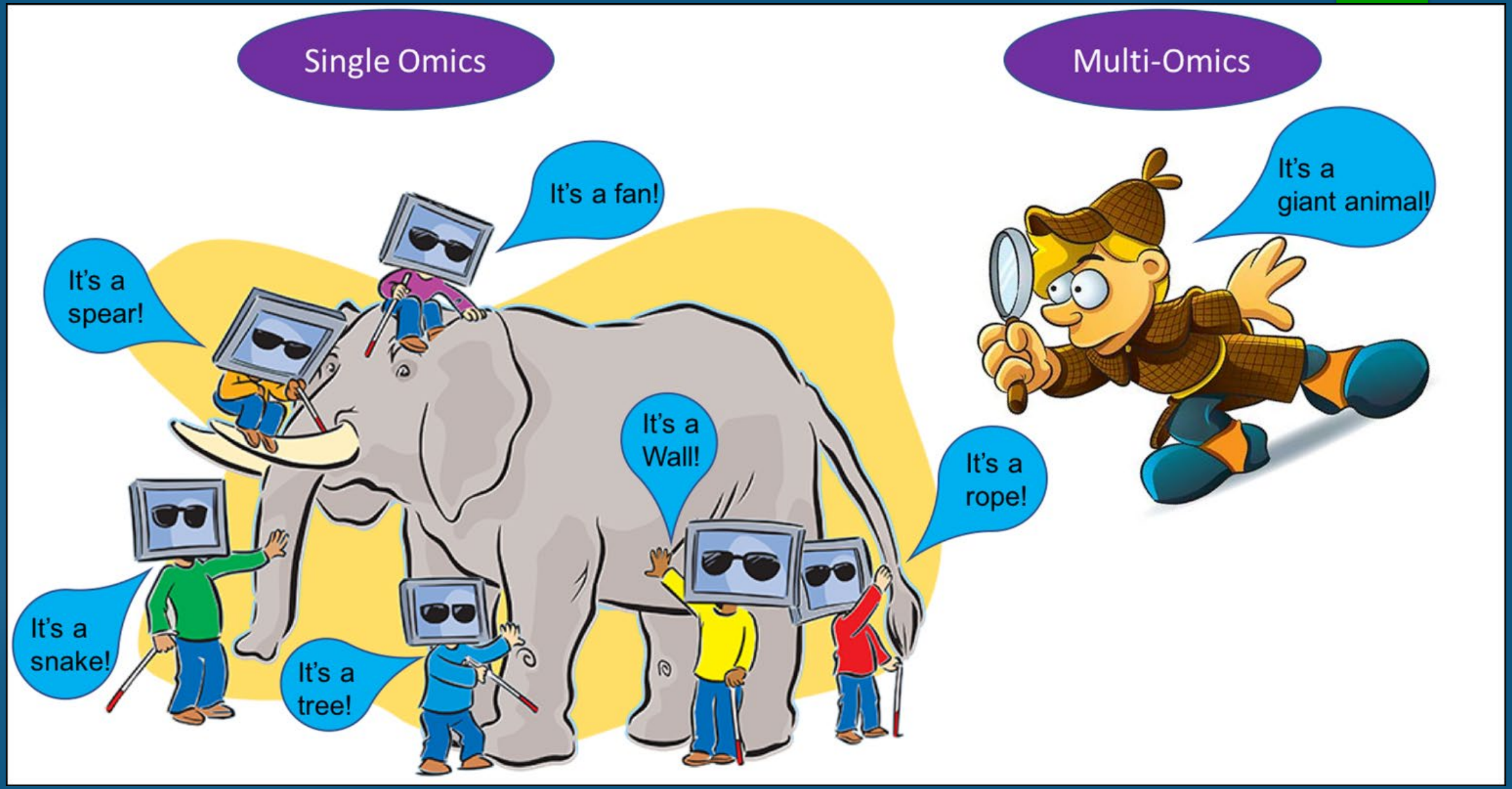
- Heterogeneous phenotype data
  - Dexa
  - Cardiopulmonary exercise testing

- High dimensional phenotype data
  - Accelerometry
  - Continuous heart rate data during training

Accelerometry Data Workflow




Ingestion
Pipelines
Normalization
Analysis
Sharing
QC
QC
QC
QC


Single Omics
Multi-Omics
It's a fan!
It's a spear!
It's a Wall!
It's a rope!
It's a snake!
It's a tree!
It's a giant animal!

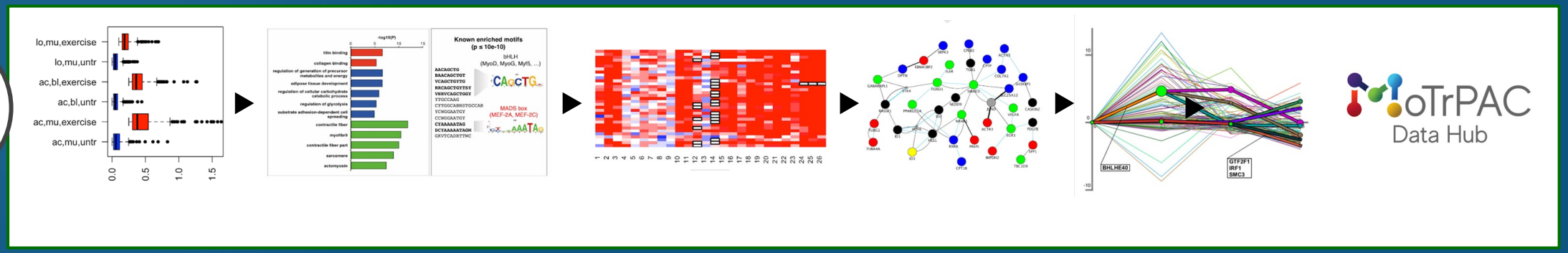
Integrative Data Analysis

CAS/ DMAQC

Genomics	Proteomics	Metabolomics	Phenotype
RNA-seq ATAC-seq RRBS DNA-seq MethylCap	Global Phospho Acyl/acetyl Immuno	Targeted metabolomics Untargeted: Reversed Phase HILIC+ and - Lipidomics	CRF Dictionary Rules Datasets

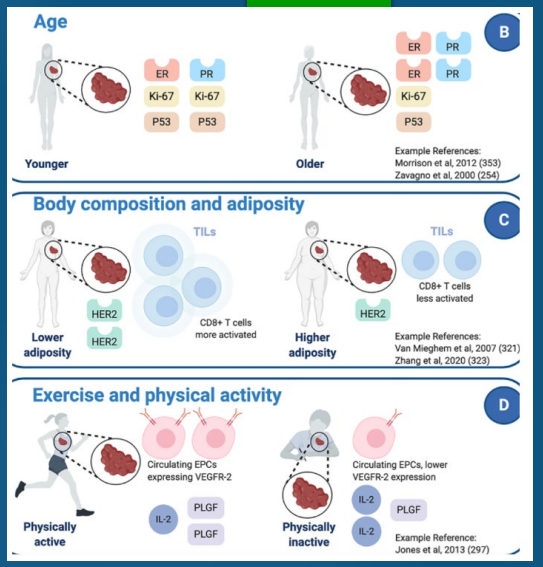
Ingestion → **Pipelines** → **Normalization** → **Analysis** → **Sharing**

AWGs

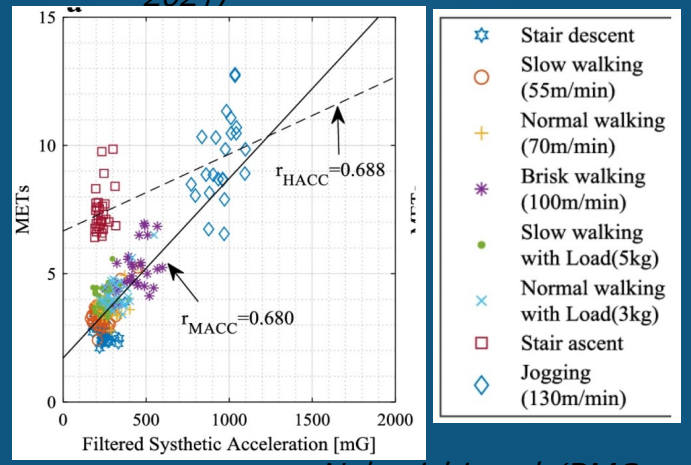


Accelerometry Analysis

- Study Compliance Analysis
  - Evaluating sedentary and intervention pattern
  - Comparing between visits/milestones activities
- Association Analysis
  - Association analysis between physical activity features and phenotypes and clinical biomarkers
- Predictive Analysis
  - Based on the association analysis, non-invasive prediction of physical (or molecular) changes with accelerometry data/features



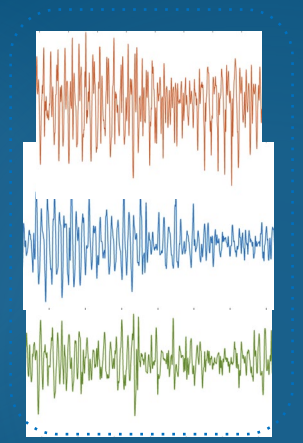
Echarri et al. (Frontiers Immunol., 2021)



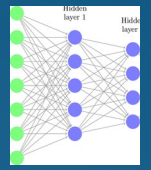
Nakanishi et al. (BMC, 2018)

Analysis exploration and integration

Measured parameters, e.g., steps, vector magnitude, axis counts, etc

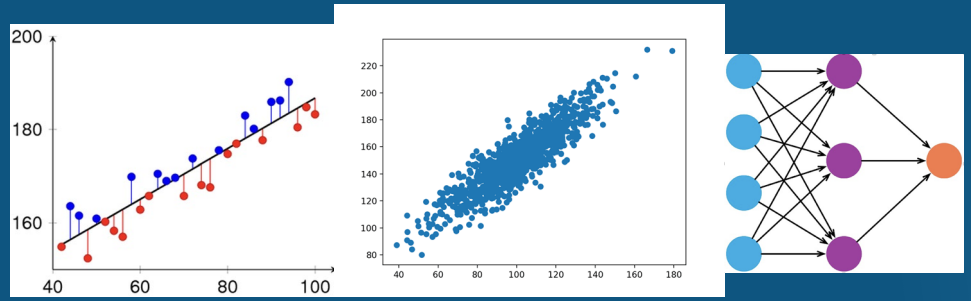


Raw Signal



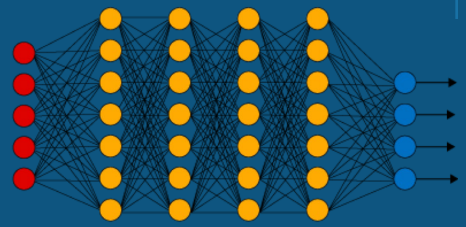
Extracted Features

Digital Biomarkers



Association/Correlation/Regression Analysis
Regressive machine learning/deep learning

Activity-wise association analysis



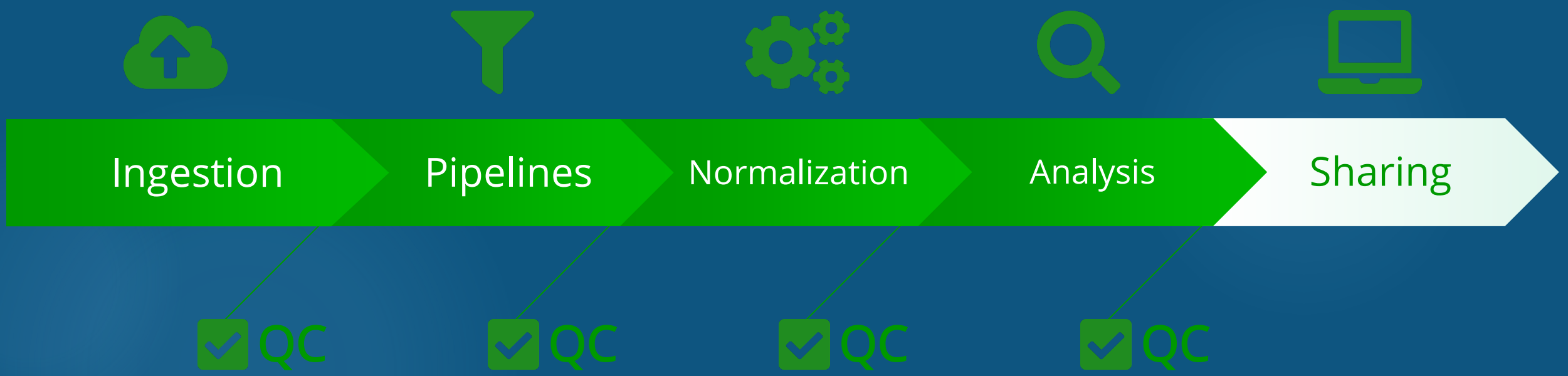
Physical Activity Recognition

Phenotypes/Clinical biomarkers

Molecular markers

Molecular networks

Molecular integration


Ingestion
Pipelines
Normalization
Analysis
Sharing
QC
QC
QC
QC

Data access and exploration

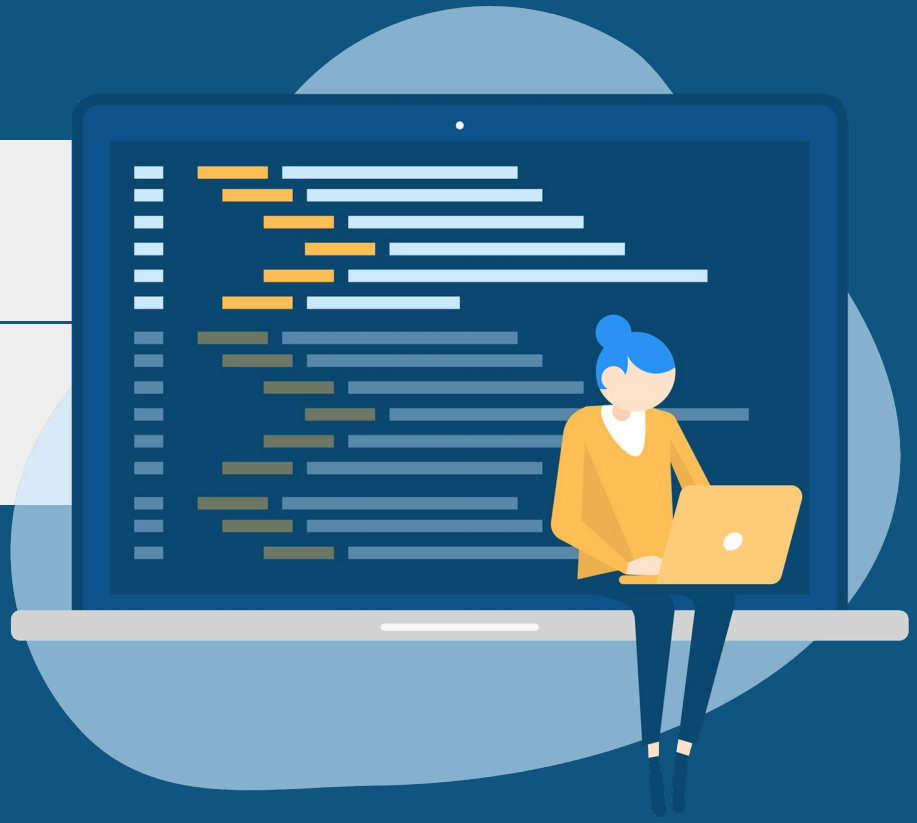
Search

Download

Visualization

Big Data

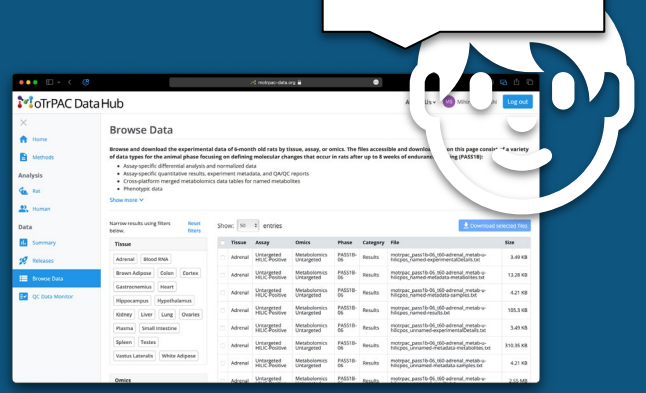
Result matrix size	117 M cells
Total data lake	1.5 Petabytes



Flexible data downloads

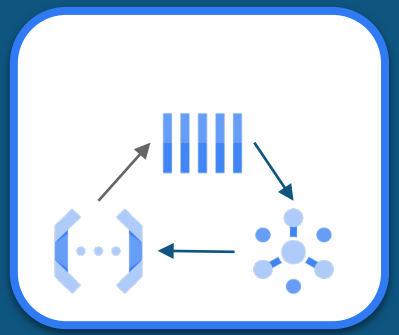
services to download any combination of data files on-demand

User requests to download files from data portal



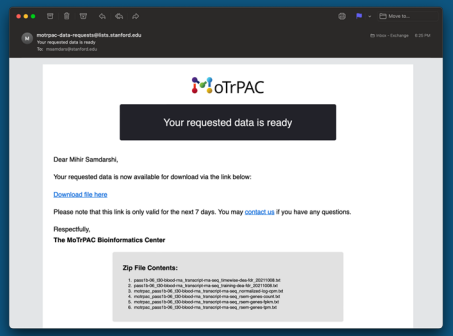
BIC get notified of request

Depending on total file size, microservices either generate a .zip file immediately or enqueue a larger compute resource to handle it



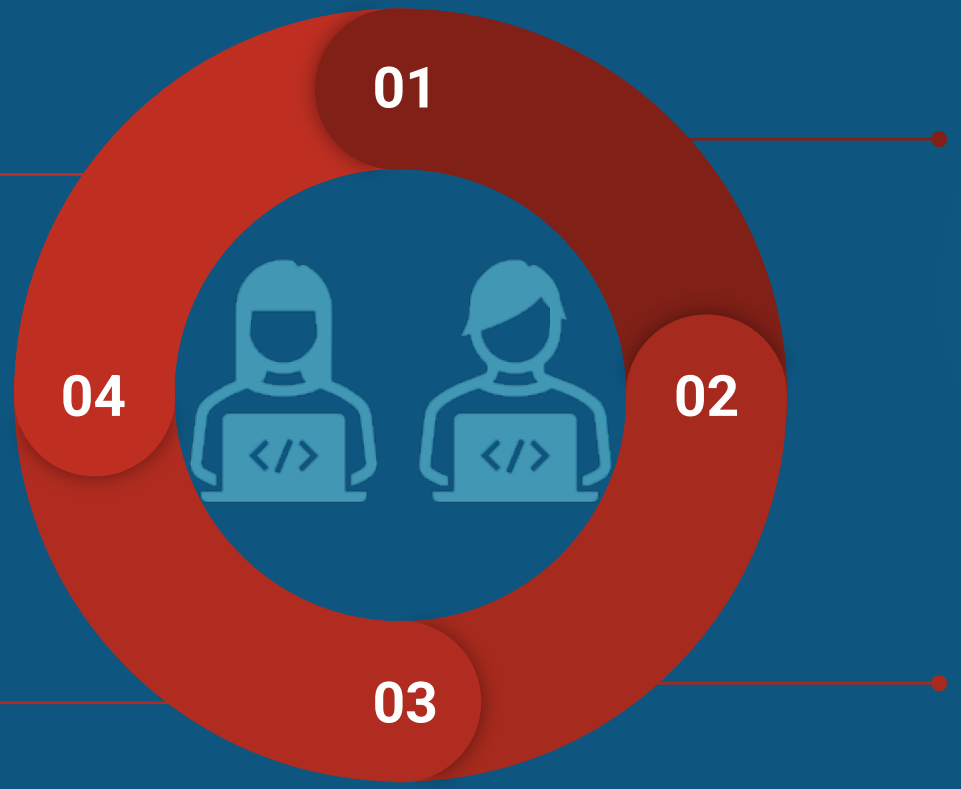
.zip file is uploaded to Google Cloud Storage, cached for future requests of the same set of files

Email with link to download file is generated and sent to the requester



Making data available to external users

Cloud Storage



motrpac-data.org

Externally accessible when Landscape paper published.

Packages

Publicly available on GitHub.

Public Repositories

Fostering Reproducibility

Packages
Publicly available on GitHub.

Consortium projects

- Results data and code shared through **R packages**

Companion projects

- **Github** repo for code
- **Zenodo** for data sharing
- Encourage early code sharing, documentation and collaboration

When does data sharing happen?

Clinician

When the study is done, data analysed and published by the investigators

Omicist

When the data is produced and data QCd and appropriate for analysis

Physiologist

When the data is produced and data analysed and published

When does data sharing happen?

Clinician

When the study is done and data analysed and published by the investigators

When trial is done, once

Omicist

When the data is produced and data QCd and appropriate for analysis

On a regular schedule, annually

Physiologist

When the data is produced and data analysed and published

With each manuscript

How is data sharing happening?

Clinician	Omicist	Physiologist
"published manuscript"	"preprint"	"preprint"
"Available on request"	"Public access at repository"	"Supplementary data"
"Clinical Trials. Gov"	"GitHub"	"protocols.io"
"BioLINCC"	"dbGAP"	

MoTrPAC BIC Team


Euan Ashley


Matthew Wheeler


Jonathan Myers


Rob Tibshirani


Trevor Hastie


Daniel Katz


Malene Lindholm


David Jimenez-Morales


Shruti Marwaha


David Amar


Archana Raja


Mihir Samdarshi


Laurens van de Wiel


Jeff Christle


Samiya Shimly


Samuel Montalvo


Terra Coakley

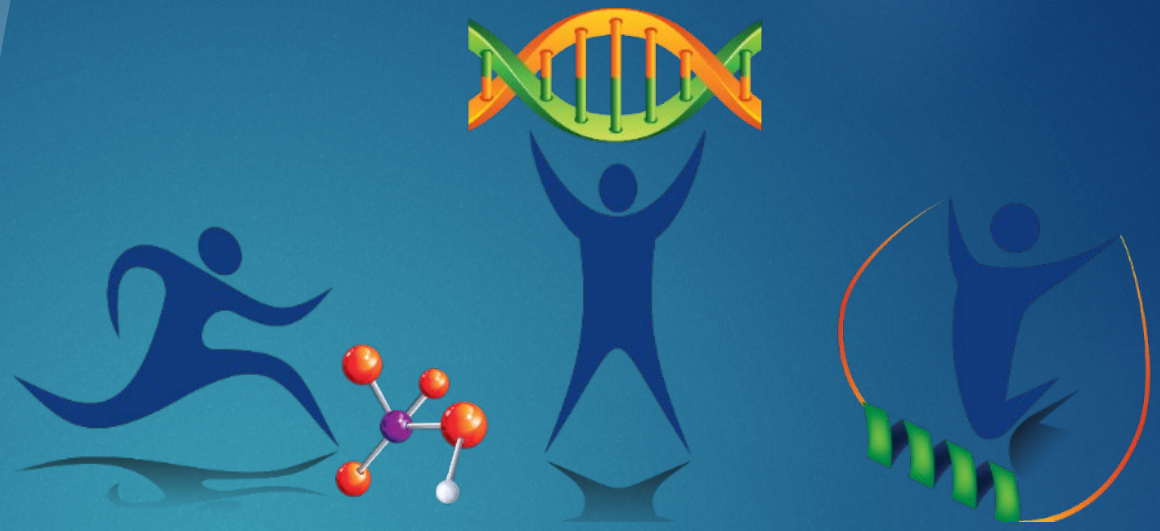

Brooke Zelnik


Jimmy Zhen


Chris Teng


Jay Yu

SCT
SOCIETY FOR CLINICAL TRIALS



Panel Discussion: Lessons Learned from The MoTrPAC Pilot Period

Stephanie George, PhD, MPA, MA NIH

Michael E. Miller, PhD, Wake Forest University School of Medicine

Scott Rushing, Wake Forest University School of Medicine

Matthew Wheeler, MD, Stanford University

What Adaptations Have Been Necessary

- **Data Sharing** – many studies are now focusing on both phenotypic and –omics data. What recommendations can be made based on the MoTrPAC experiences?

- **Outcomes –** MoTrPAC lacks a pre-specified outcome, but relies on best practices of randomized trials to standardize measurements, conceal randomization, and create a resource that can provide unbiased estimates of EE vs RE. Where has this created issues?

- **Central Repositories and Sample/Outcome Tracking** -- What are the key lessons learned relative to studies that use a central repository along with multiple chemical analysis sites?