Three multi-PIs: Gerstein, Kraus, Milosavljevic (contact)

Gerstein: Develop data processing pipelines. Define data Quality Control metrics. Perform knowledge integration. Participate in cloud computing implementation. Develop models for understanding the multidimensional, multimodality data, and developing high quality connectivity maps showing the interrelations between the various data types. Design and implement integrative analysis strategies.

Kraus:

1. Participate in metadata standards development (RFA: "coordinate implementation of data and ontology-based metadata standards"; "Selecting or developing common data elements to enable uniform aggregation of data") 0.2 FTE
2. Perform metadata "wrangling" and curation (RFA: "Developing, or using pre-existing and well-defined, data curation standards and methods"; "Accumulating, integrating, and storing as necessary physiological, metabolic, and metadata from the Clinical Centers and PASS, and metabolomic, proteomic, genomic, and transcriptomic data from the Chemical Analysis Sites") 1.5 FTE
3. Act as liaison with clinical studies – and co-develop plans for replication and validation studies (RFA: "Working with the Steering Committee to develop plans for replication/validation studies as needed.") 0.1FTE
4. Participate in the integration of biospecimen data 0.1 FTE
5. Co-develop data portal, 0.1 FTE
6. Co- develop models for understanding the multidimensional, multimodality data, and participate in developing connectivity maps showing the interrelations between the various data types 0.6 FTE
7. Participate in preliminary data analysis (RFA: "and conduct preliminary data analysis of the diverse datasets submitted by other MoTrPAC elements.") 0.4 FTE

Milosavljevic: Co-develop data and metadata standards. Develop a "virtual biorepository" to integrate biospecimen data across the consortium. Implement the infrastructure for metadata and data processing. Deploy data processing pipelines. Perform data processing. Enable data sharing. Implement provenance tracking. Perform data deposition into public archives. Develop on-line supporting material sand implement user training. Develop methods and tools to facilitate and participate in integrative analysis in the context of background knowledge of pathways and networks.

Confirmed co-Investigators:

Alex Pico (UCSF, networks, pathways, visualization)
Kei-Hoi Cheung (Yale, ontologies, Linked Data)
Hongyu Zhao (Yale, statistics)
Kim Huffman (Duke, metabolomics, exercise, physiology, metadata gene expression)
Svati Shah (Duke, metabolomics, exercise, QC, metadata)

Initial Thoughts on an Outline:

Grant Outline
12 pages overall (6 - 4 - 2 effort division)

* 1 pg intro

* 6 pg to DCC
(incl. 1/3 page to be slotted in on MG experience in privacy, cloud computing, integration w/ lit.)

* 4 pg to DAC

- Prelim. Results [2 pg]
- Upstream processing [1 pg]
+ Setting up standards & pipelines [.5 pg]
+ Integration w/ encode, exRNA [.5 pg]
(incl. sensitivity to extracellular & cellular information)
- developing integrative omics models & tools for these [1pg]
+ developing approaches for normalization, registration & analysis of temporal data
+ developing tools to analyze dynamic data & longitudinal variation
+ developing approaches for tissue de-convolution (⅙ page from Aleks)
- integration of different types of omics data with genetic variants ("star-QTL") & allelic analysis [.5 pg]

* 1 pg on consortium activities

Specific Aims

Overarching Goal: Implement public data resource that any researcher can access to develop hypotheses regarding the molecular mechanisms through which physical activity can improve or preserve health.

	Aim 1: develop informatic infrastructure and deploy it as a cloud-based service~~atabase design~~	**Aim 2: develop~~ment~~ ~~of~~ methods and tools for data processing and analysis**	**Aim 3: develop data analysis strategy and perform data processing and preliminary data analysis**	**Aim 4: support ~~other~~ consortium activities**
provide a	Aim 1a			

database for storage and integration of clinical physiological and multiple types of “omics” data				
develop a framework for integrating ~~integrate and deploy~~ pipelines and tools for analysis and visualization of data	Aim 1b			
provide rapid access to accumulated data and tools through the use of cloud-based computing	Aim 1c			
oversee standardization of data and metadata; develop and maintain SOPs		Aim 2a		
develop and deploy ~~deploy~~ data processing pipelines, including QC metrics and report generation		Aim 2b~~c~~		
develop and deploy data analysis tools		Aim 2c		

develop and deploy methods and tools for integrative analysis and visualization in the context of pathways and networks		Aim 2d		
process data and metadata, populate the database, and submit the data for permanent archiving		~~Aim 2d~~	Aim 3a	
Working with consortium members, develop analytical models and theoretical constructs to establish a 'molecular map' of physical activity in humans			Aim 3b~~a~~	
conduct preliminary data analysis of “acute response to exercise ” the diverse datasets submitted by other MoTrPAC elements			Aim 3c~~b~~	
conduct preliminary data analysis of the “durable response to			Aim 3d~~c~~	

exercise ” using diverse datasets submitted by other MoTrPAC elements				
~~Work with Consortium members and the SC to develop plans for replication studies.~~			~~Aim 3d~~	
Provide integration of biospecimen data and other resources across the consortium				Aim 4a
Provide expertise in data management and analytics				Aim 4b
Work with Consortium members and the SC to develop plans for replication studies.				Aim 4c