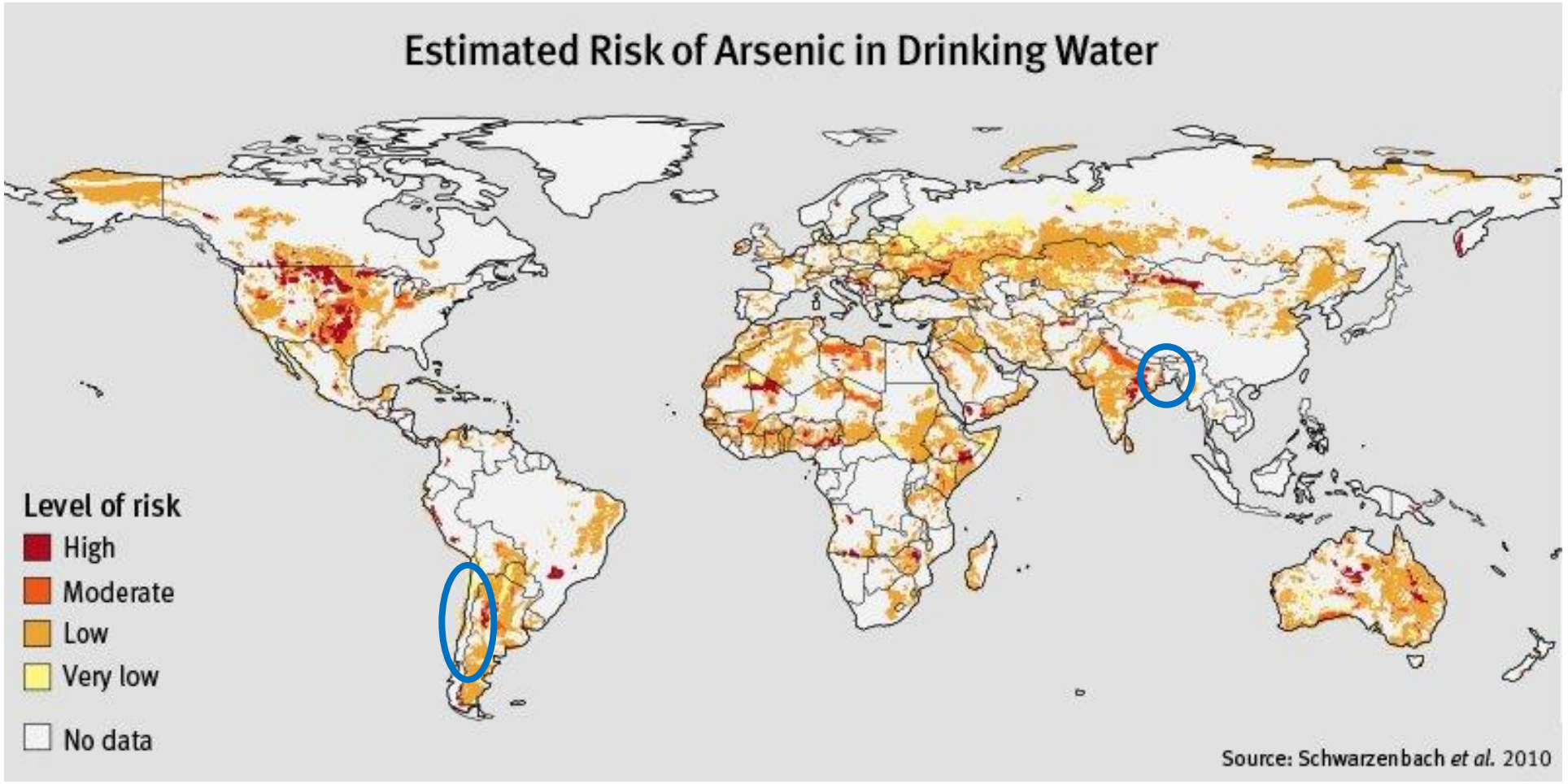
Arsenic Epigenetics **META:**
Meta-analysis of **E**pigenome Da**t**a on **A**rsenic

Risk e-Learning Webinar Series, August 3, 2021

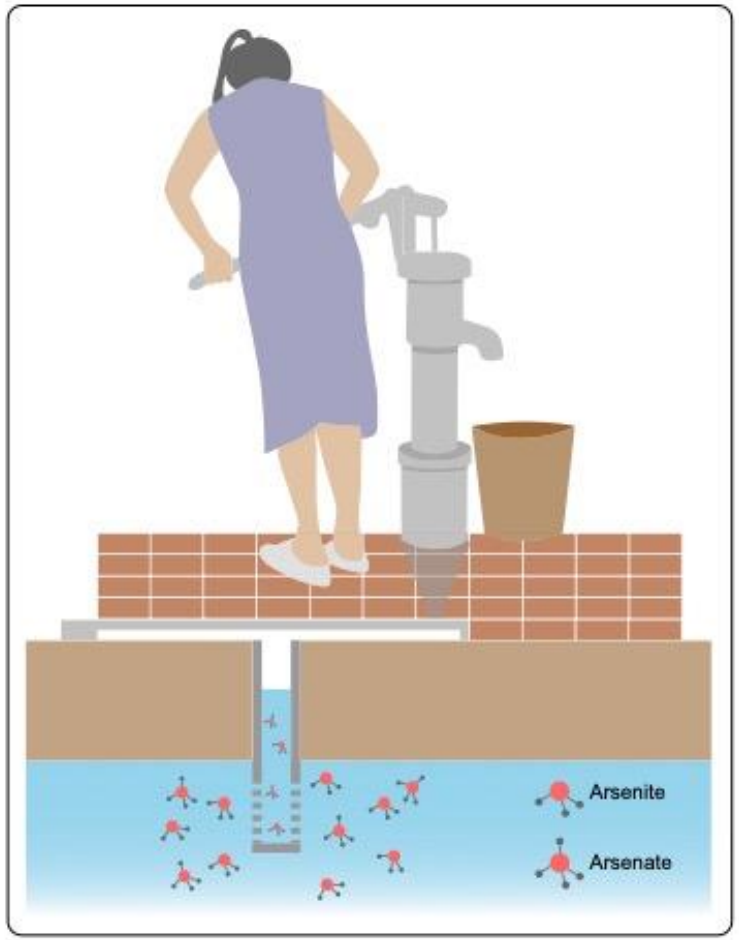
Anne Bozack, MPH, PhD
Andres Cardenas, MPH, PhD

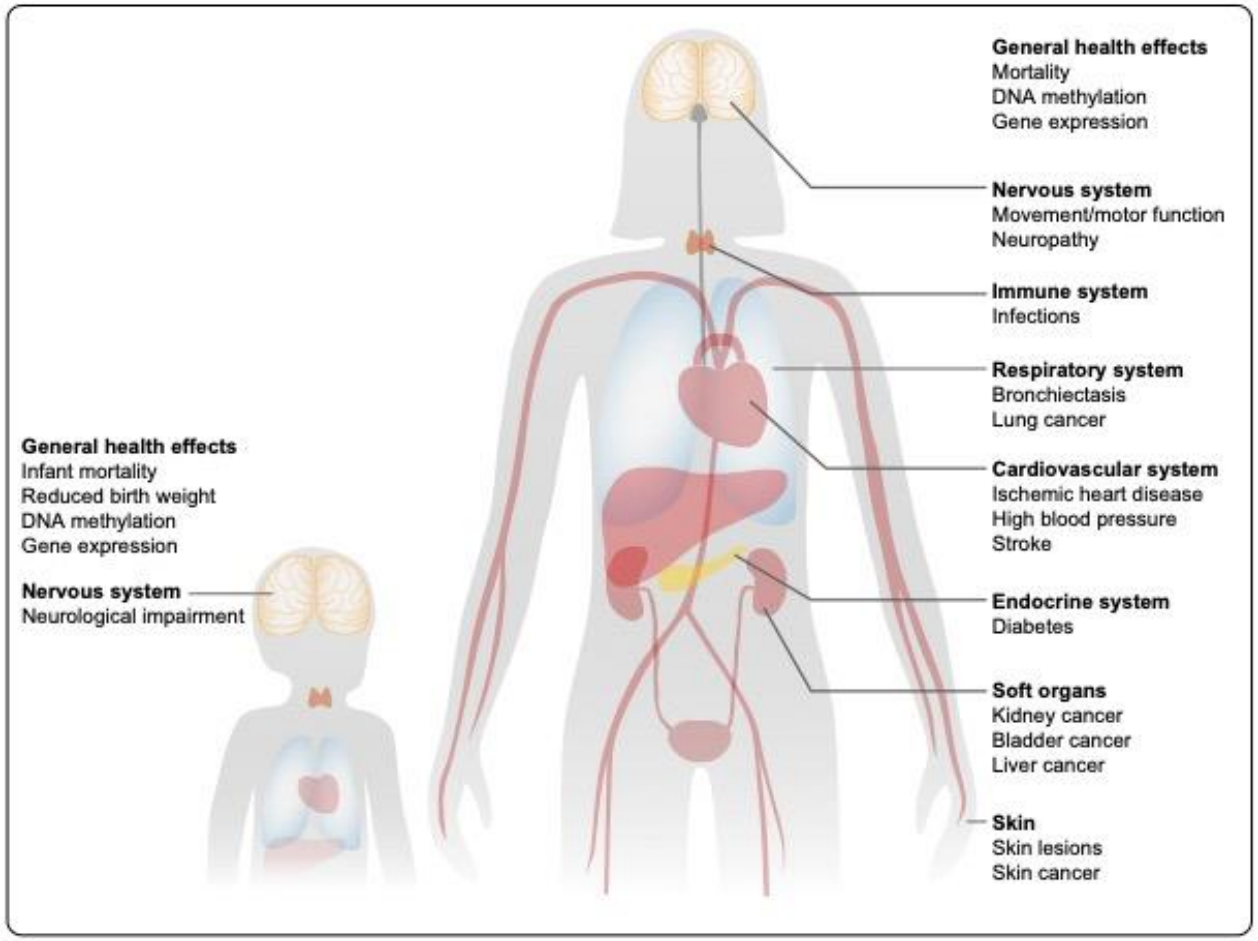
Berkeley
UNIVERSITY OF CALIFORNIA

Arsenic exposure and health effects



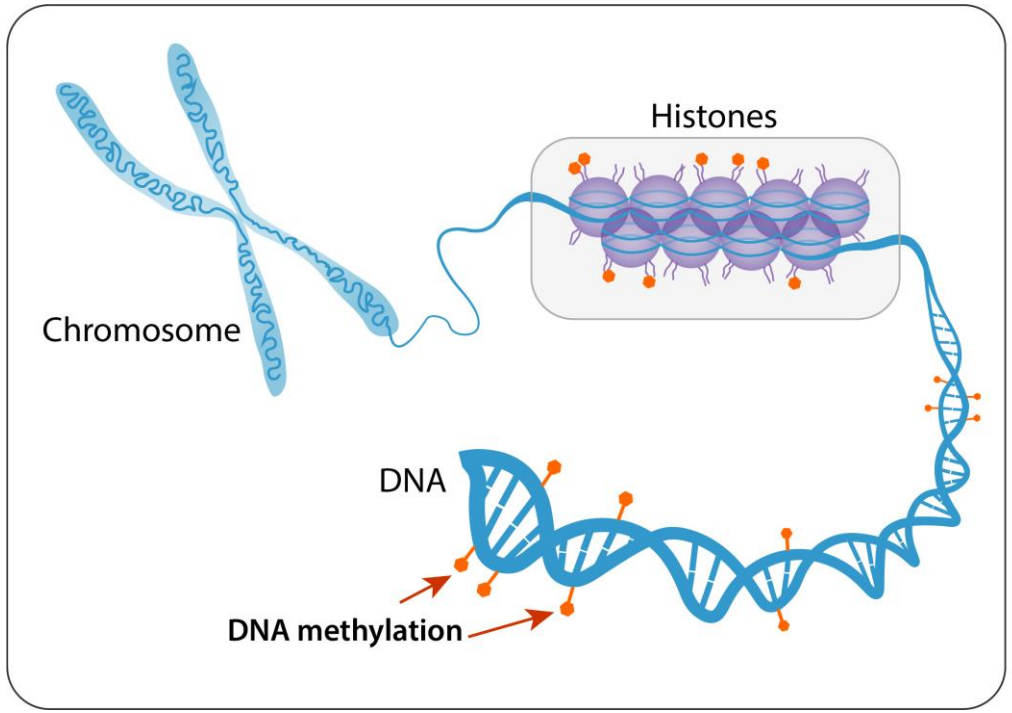
Arsenic exposure and health effects

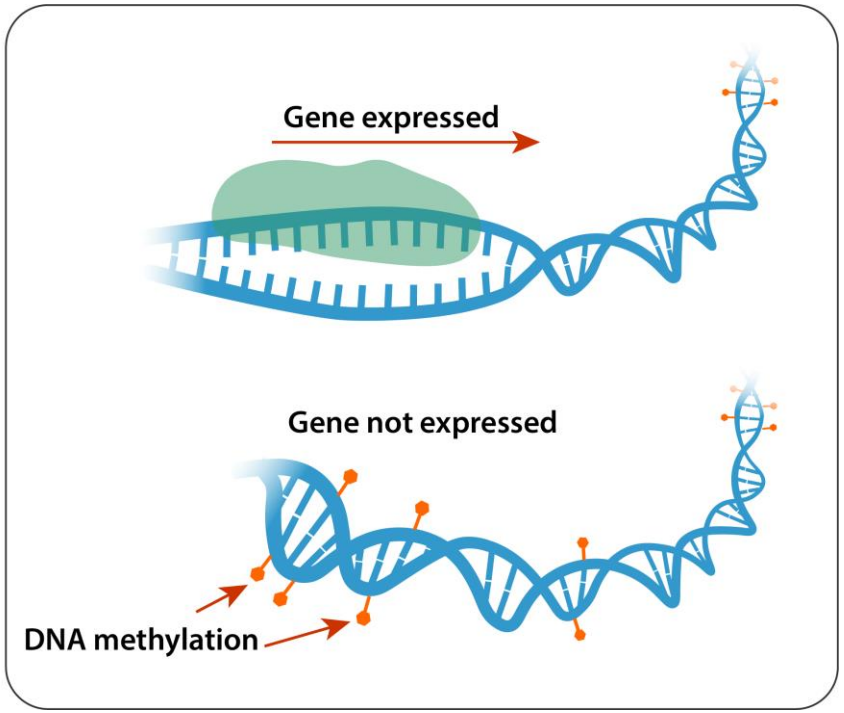




Arsenic exposure and health effects

- Arsenic-related health risks persist after exposure has ended.
  - Epigenetic dysregulation may be a mechanistic link between As and health outcomes.





General overview

- **Summary**
  - Leverage previously measured Epigenome-Wide DNA methylation data across SRP centers for a meta-analysis of arsenic exposure on the epigenome of human cohorts
    - Addresses the question as to whether epigenetic biomarkers of As exposure are generalizable
  - Goal is to develop a framework, protocols, open-source code, and associated workflow that can be utilized to meta-analyze multiple EWAS related to environmental exposures (Epigenetics Consortium of Environmental Exposures)

	University of California, Berkeley	Columbia University
Point of contact	Andres Cardenas	Mary Gamble
Lead project title	Exposomics and Arsenic Epidemiology	Impact of Nutrition on Arsenic-Induced Epigenetic Dysregulation
Other partners	Craig Steinmaus; Martyn Smith; Waverly Wei, Philippe Boileau	Ana Navas-Acien; **Anne Bozack**

Inputs and actions

Inputs

- **Existing data sets**: Columbia SRP DNA methylation data from Bangladeshi adults exposed to arsenic (**urinary and water**); UC Berkeley cohort from Northern Chile of adults exposed to arsenic early in life (**prenatal vs post**)
- **Variables**: High dimensional DNA methylation data (**450K** or **850K CpG** sites in the human genome); historical As exposure and biomarkers, and demographic characteristics
- **Repositories**: Data currently stored locally at each SRP center but not systematically preserved/annotated

Actions: how are we achieving F, A, I, and or R

- Analytical code is ***findable*** internally/**externally** by users at each center by navigating a well-annotated GitHub repository (https:///github.com/annebozack/SRP_arsenic_DNAm_metaanalysis)
- Summary results ***accessible*** by sharing our analytical protocol and code : https:///github.com/annebozack/SRP_arsenic_DNAm_metaanalysis
- We will increase ***interoperability*** as summary EWAS findings can be integrated with other omics results (OSF)
- By preserving our data and annotated code we will ensure data is ***reusable*** for trainees and investigators. (**Epigenetic Aging Biomarkers**)

Collaboration tools

GitHub

- Created a shared repository to collaborate on development of data processing and analysis pipeline
- Ensured that collaborators had access to the most recent code versions
- Repository made publicly available for other researchers to access data processing and analysis pipeline

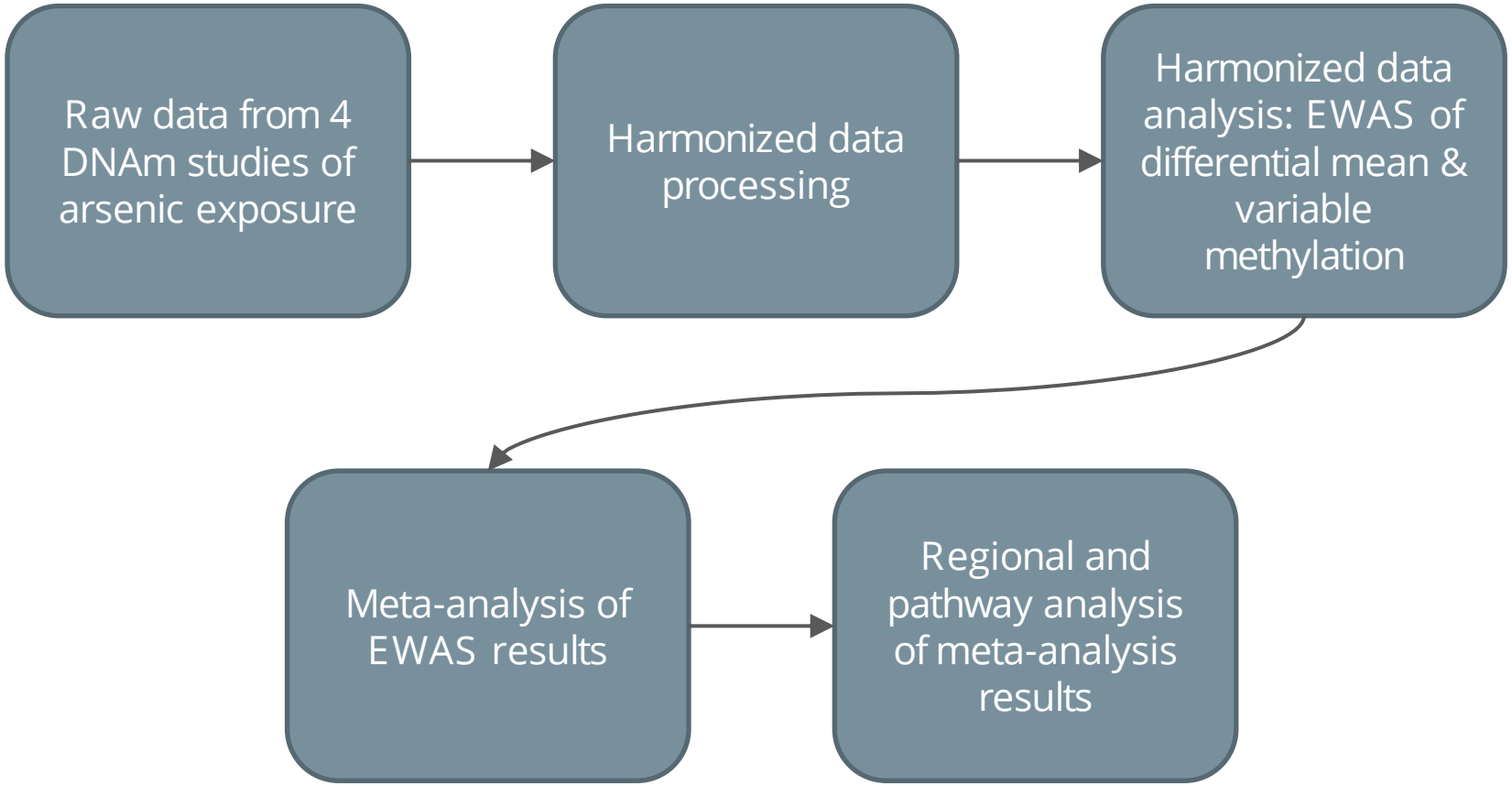
Box

- Used to securely store/transfer EWAS results between centers
- Convenient upload/download of large datasets (e.g., output from ~450,000 and 850,00 analyses)

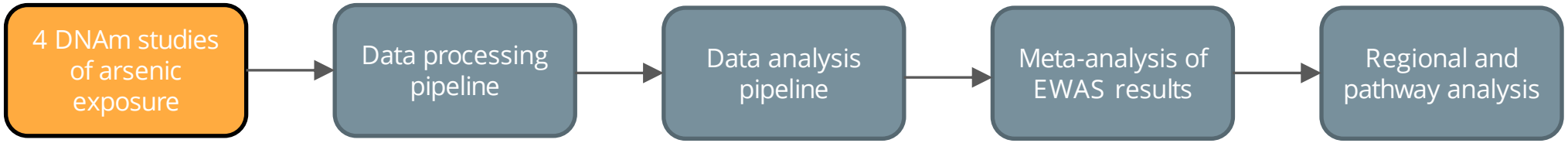
Google docs

- Allowed for collaboration and version control during manuscript preparation

Actions: Data processing and analysis pipeline



Actions: Data processing and analysis pipeline

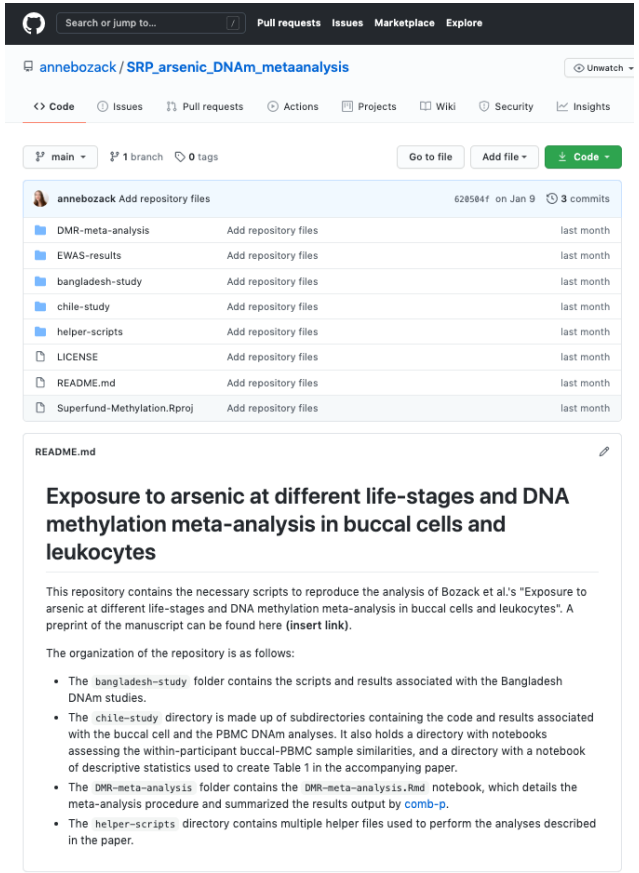


- Two study locations: Bangladesh and Chile
- Chile: two different tissues (buccal and blood cells)
- Bangladesh: two epigenomics platforms: 450K and EPIC (850K)
- Comparison of adult chronic exposure (Bangladesh) vs high fetal exposure (Chile)

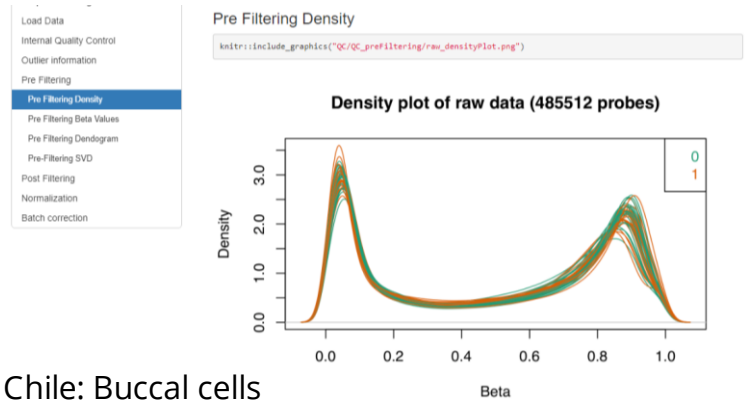
	Chile, PBMCs (N = 40) [a]		**Chile, buccal cells (N = 39) [a]**		**Bangladesh, 450K (N = 48)**		**Bangladesh, 850K (N = 32)**	
	n	**%**	**n**	**%**	**n**	**%**	**n**	**%**
Age, years, mean (SD)	48.7	(4.7)	48.7	(4.7)	39.7	(8.1)	41.7	(6.3)
Male	21	52.5	20	51.3	48	100.0%	32	100.0%
Ever smoker	16	40.0	16	41.0	21	43.8%	20	62.5%
Prenatal/early life arsenic exposure	20	50.0	19	48.7	-	-	-	-
High arsenic exposure [b]	-	-	-	-	23	47.9%	11	34.4%

a. 850K; PBMC and buccal cell samples from the same study participants. b. ≥ 100 µg/L water arsenic for 450K analyses and 104 µg/L water arsenic for 850K analyses.

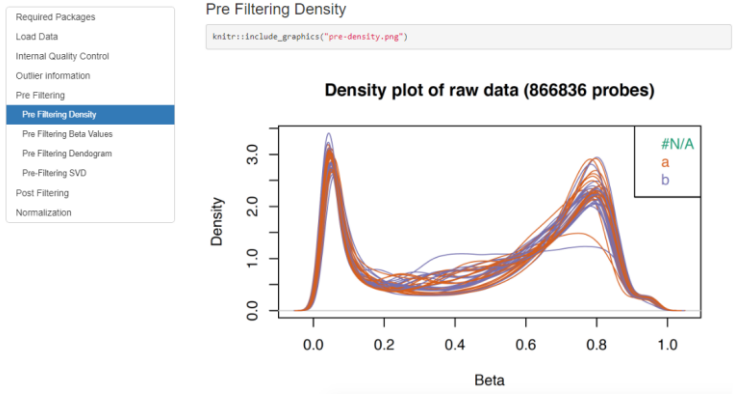
Actions: Data processing and analysis pipeline

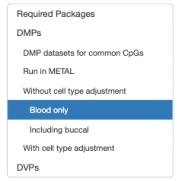


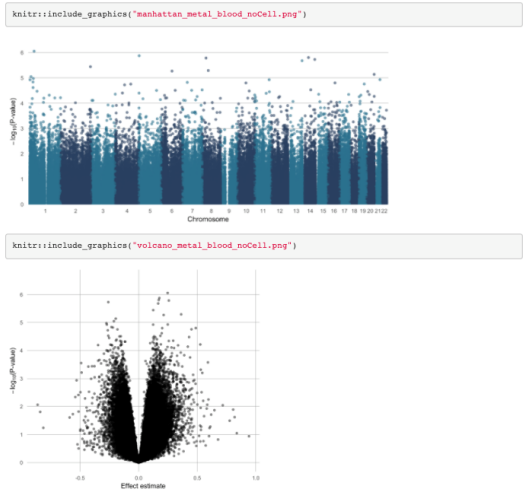
Bangladesh: Leukocytes



Chile: Buccal cells





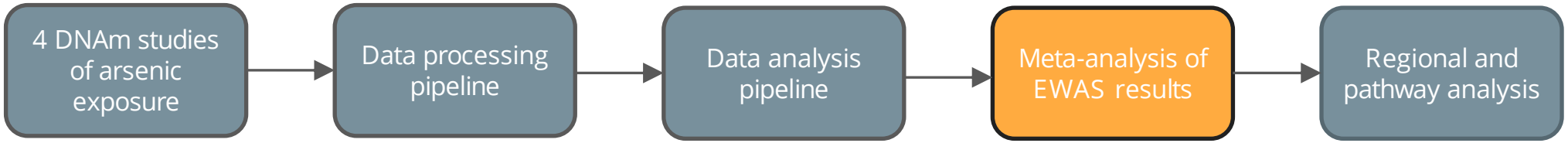


Top 100 probes

	Name	Effect	StdErr	P.value	Direction	P.FDR	P.Bonf	chr	pos	UCSC_RefGene_N
93	cg00004257	0.2473	0.0503	0.0000009	+++	0.1315697	0.3325217	chr1	34629175	CSMD2
316516	cg22938488	0.1765	0.0365	0.0000013	-++	0.1315697	0.5045183	chr5	1501670	LPCAT1
114376	cg07490485	0.1711	0.0356	0.0000016	+++	0.1315697	0.5924411	chr14	50432614	
194923	cg13490635	0.2572	0.0537	0.0000016	+++	0.1315697	0.6207424	chr8	30242021	RBPMS;RBPMS;RE
17421	cg01050273	-0.2598	0.0545	0.0000019	--+	0.1315697	0.7064011	chr14	101120705	
129233	cg08528486	0.1663	0.0350	0.0000021	-+-	0.1315697	0.7894183	chr13	113648767	MCF2L;MCF2L
248565	cg17342864	0.1239	0.0268	0.0000036	++-	0.1950905	1.0000000	chr2	239973967	HDAC4
39707	cg02483029	0.3626	0.0795	0.0000051	++-	0.2263687	1.0000000	chr8	48297271	KIAA0146
359036	cg26299756	0.2130	0.0468	0.0000054	+++	0.2263687	1.0000000	chr6	85478807	
73319	cg04685632	-0.1960	0.0437	0.0000073	--	0.2749002	1.0000000	chr20	62082611	KCNQ2;KCNQ2;KC
182880	cg12547807	-0.2121	0.0478	0.0000090	--	0.2996167	1.0000000	chr1	9473751	
46132	cg02892755	-0.2760	0.0627	0.0000106	--	0.2996167	1.0000000	chr1	28103358	STX12
233149	cg16190478	0.2482	0.0566	0.0000118	+++	0.2996167	1.0000000	chr21	45789122	TRPM2

https:///github.com/annebozack/SRP_arsenic_DNAm_metaanalysis

Actions: Data processing and analysis pipeline



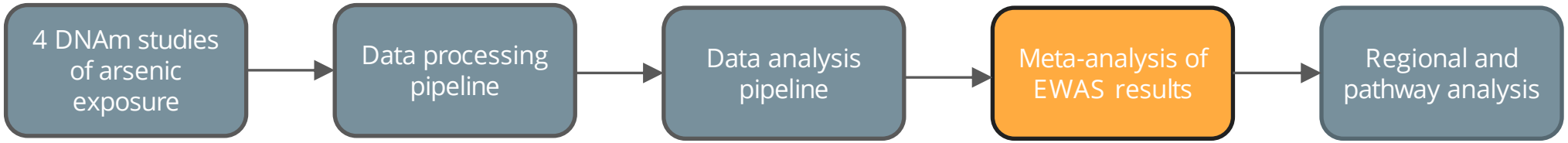
Summary of results of individual EWAS

	Common probes [a]	
DMPs	**p < 0.05**	
Chile, PBMCs	23,116	
Chile, buccal cells	21,336	
Bangladesh, 450K	18,301	
Bangladesh, 850K	7,954	
DVPs	**p < 0.05**	**FDR < 0.05**
Chile, PBMCs	23,487	3
Chile buccal, cells	20,735	4
Bangladesh, 450K	16,904	2
Bangladesh, 850K	26,155	24

DMP: differentially methylated position; DVP: differentially variable position. Adjusted for cell type proportions, age, and smoking status. a. 377,351 included in all four EWAS.

No DMPs at FDR < 0.05 identified in individual EWAS.

Actions: Data processing and analysis pipeline



Summary of results of individual EWAS

	Common probes [a]	
DMPs	**p < 0.05**	
Chile, PBMCs	23,116	
Chile, buccal cells	21,336	
Bangladesh, 450K	18,301	
Bangladesh, 850K	7,954	
DVPs	**p < 0.05**	**FDR < 0.05**
Chile, PBMCs	23,487	3
Chile buccal, cells	20,735	4
Bangladesh, 450K	16,904	2
Bangladesh, 850K	26,155	24

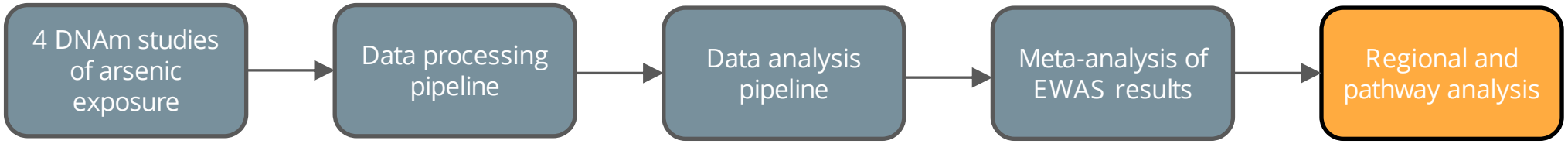
DMP: differentially methylated position; DVP: differentially variable position. Adjusted for cell type proportions, age, and smoking status. a. 377,351 included in all four EWAS.

Summary of results of meta-analyses

	p < 0.05	FDR < 0.05	λ
DMPs			
PBMCs	23,361	1	1.07
PBMCs + buccal cells	22,612	3	1.06
DVPs			
PBMCs	28,578	23	1.17
PBMCs + buccal cells	28,399	19	1.18

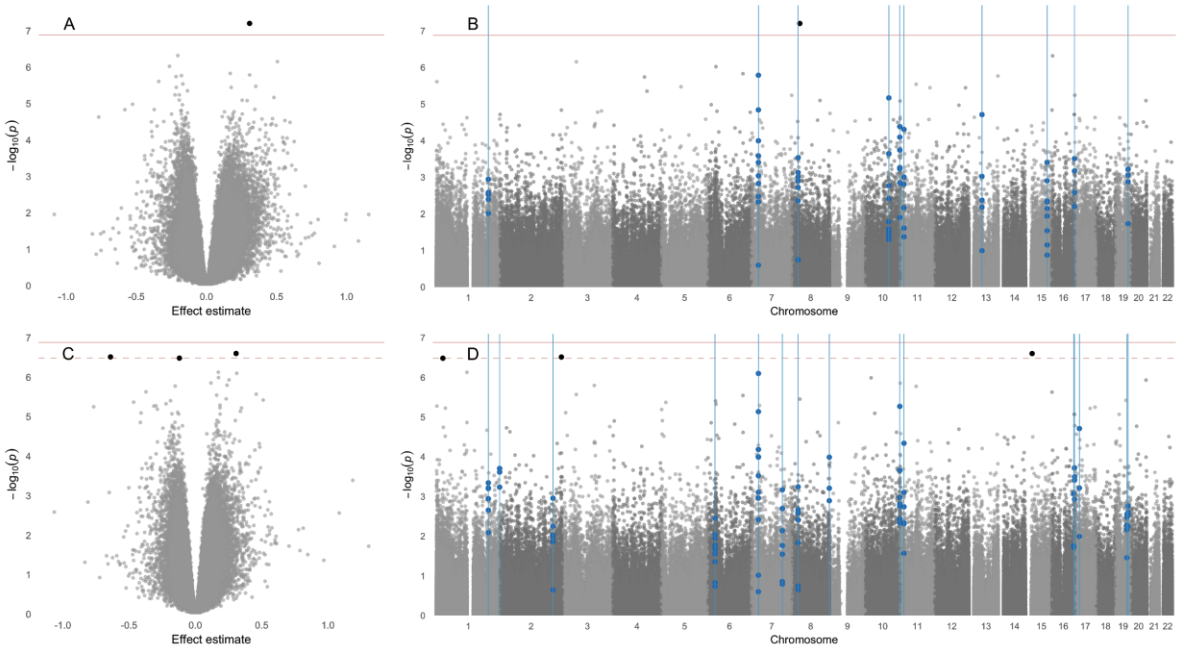
Adjusted for cell type proportions, age, and smoking status. DMP: differentially methylated position; DVP: differentially variable position.

Actions: Data processing and analysis pipeline



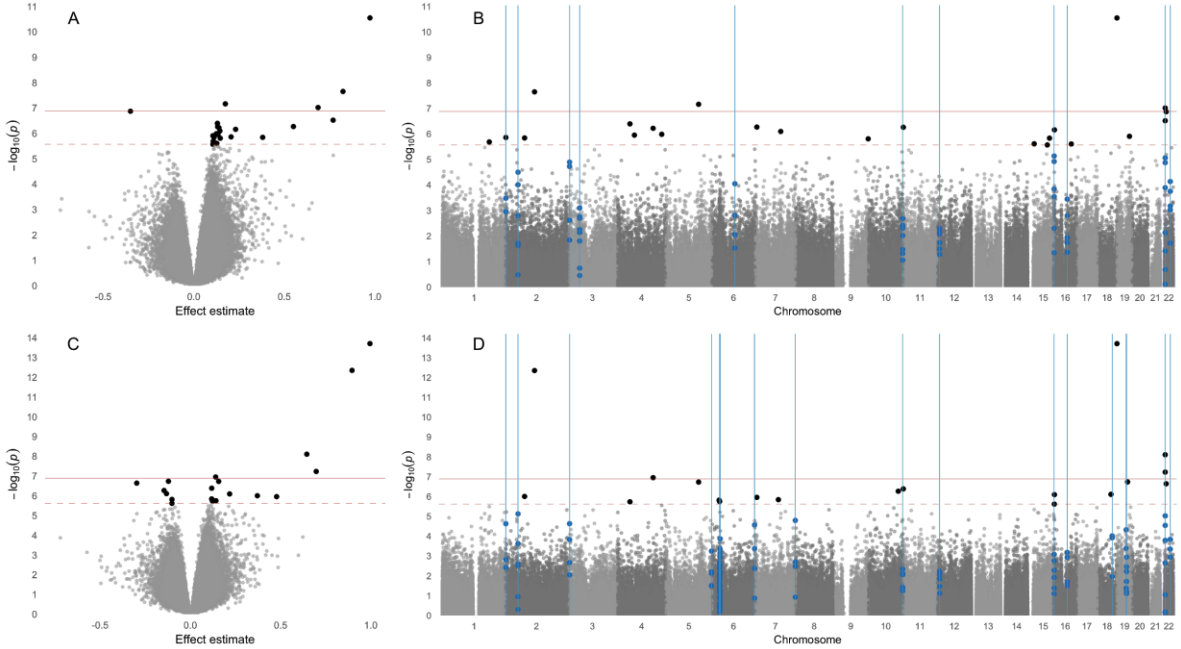
Differential mean methylation

Top row: PBMC EWAS; Bottom row: PBMC + buccal cell EWAS

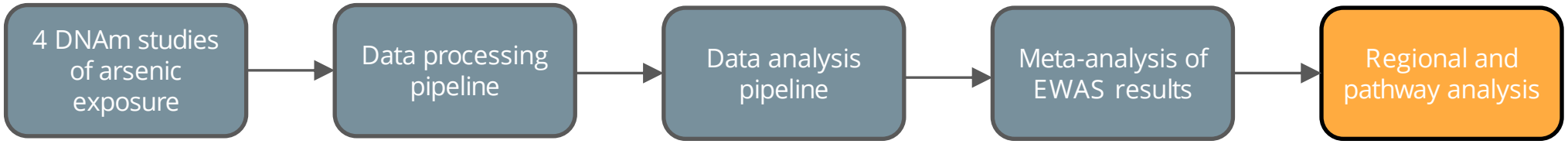


Differential variability in methylation

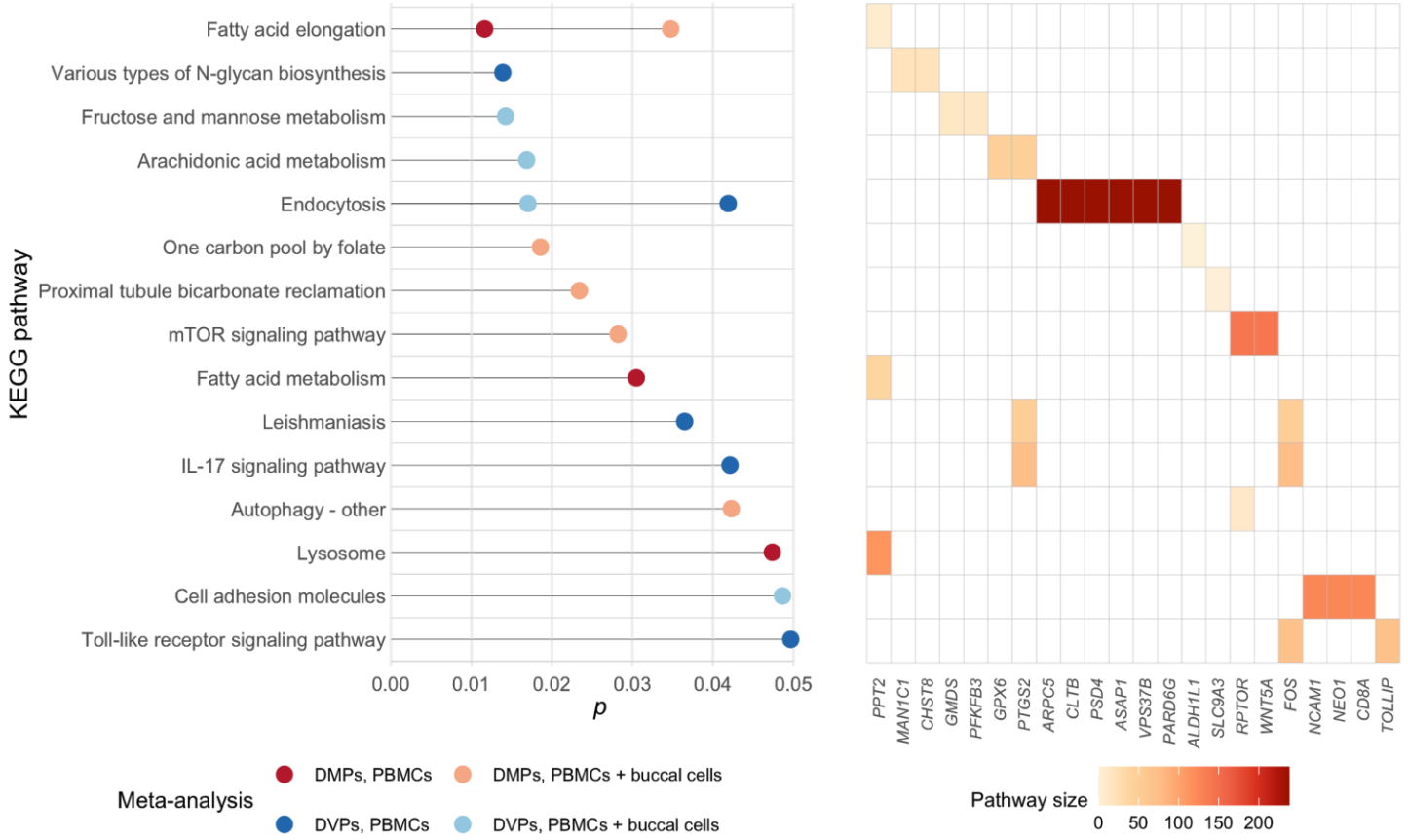
Top row: PBMC EWAS; Bottom row: PBMC + buccal cell EWAS



Actions: Data processing and analysis pipeline



KEGG pathway analyses



Actions

Platforms: Data processing/analysis pipelines and results available on GitHub; will transfer to Open Science Framework

- GitHub repository: https:///github.com/annebozack/SRP_arsenic_DNAm_metaanalysis

Integrating datasets: Established consistent classification of exposure across datasets; epigenetic measurements and QC

Communication: In-person project planning meeting; virtual symposium; weekly virtual meetings

- Virtual symposium: https://www.youtube.com/watch?v=J3-myoAVlU0
- GitHub repository to collaborate on developing code
- Google docs to work on manuscript

Collaborations: Established ongoing collaboration between UC Berkeley and Columbia SRPs around arsenic-induced epigenetic dysregulation

Outcomes and deliverables

Short-term

- Analytical approach for conducting meta-analyses of EWAS across different populations, platforms, and exposures

Intermediate

- Harmonized data processing and analysis pipeline
- Repository for code and results
- Virtual metal epigenetics symposium: https://www.youtube.com/watch?v=J3-myoAVlU0

Long-Term

- Manuscript describing EWAS meta-analysis approach and findings (*Environ Health*. 2021 Jul 9; 20(1): 79. doi: 10.1186/s12940-021-00754-7)
- Code and summary results publicly available
- Possible collaborations with other groups with arsenic and epigenomic data
- Creation of an Environmental Epigenetic Consortium (future)
- Collaboration between Biostatistics students and EHS scientist

Lessons learned

- Collaboration is key (multiple stakeholders), and reuse of data improves data FAIRness
- Standard QC practices helped us compare data directly
- Improved data curation practices, annotation and storage
- Long-term storage of data with detail information will facilitate reuse
- Center specific analyses allows for equal partnership and shared governance

Advantages of collaboration and data sharing

- **Scientific question**: reproducibility of arsenic associated epigenetic dysregulation?
  - Pooling data enabled us to increase statistical power
  - Improved generalizability of findings
  - Meta-analyses can yield robust human epigenetic biomarkers

- Two cohorts and multiple tissues improved interpretability of epigenetic signature
- Results differed (*i.e.* cohort specific signals vs. common epigenetic signatures)
- Including more studies could address chronic vs acute exposure signatures

- Future questions that remain are *i*) chronic vs acute As epigenetic signature *ii*) reliability of arsenic exposure biomarker *iii*) expanding to other cohorts

Recommendations

- **Training**
  - Increasing data FAIRness for all research projects (PIs and trainees)
  - Application of data science methods to existing problems

- **What future activities** are needed to ensure success?
  - Provide incentives for collaborations (i.e., supplemental funds)
  - Increase participation of statisticians and bioinformaticians within and across centers
  - Increase activities/training among statisticians/data scientist and lab scientists

- **What future activities are** needed to foster and advance data sharing?
  - Provide incentives for collaborations (i.e., supplemental funds)
  - Increase participation of statisticians and bioinformaticians within and across centers

UC BERKELEY SUPERFUND RESEARCH PROGRAM
SCIENCE FOR A SAFER WORLD

Questions?

As SRP Globally Reducing Arsenic Exposure | Columbia Superfund Research Program

Contact: andres.cardenas@berkeley.edu; anne.bozack@berkeley.edu

Funding: NIEHS P42 ES004705, P42 ES010349

NIH NIEHSSRP

Berkeley UNIVERSITY OF CALIFORNIA