•Epigenomic analysis using massively-parallel sequencing

•Masako Suzuki DVM PhD

Center for Epigenomics

Department of Genetics

Albert Einstein College of Medicine, Bronx, NY

center for epigenomics | Einstein

EINSTEIN

Albert Einstein College of Medicine

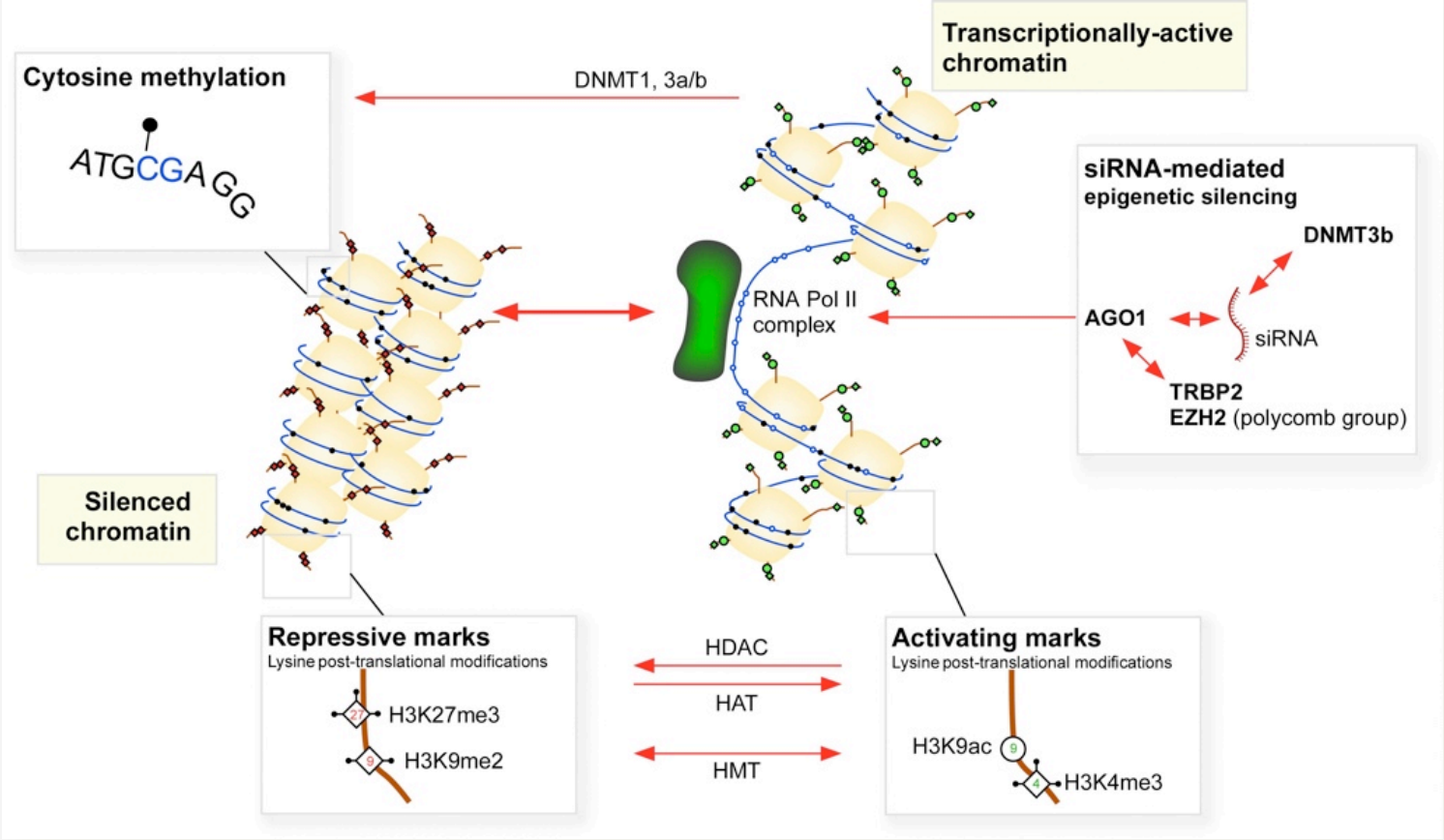
OF YESHIVA UNIVERSITY

Overview of presentation

- Introduction of Epigenomics
- Center for epigenomics Einstein
  - ✓ New DNA methylation assay development
- Data management using WikiLIMs
  - ✓ Sample submission and retrieving results
  - ✓ Electronic lab notebook

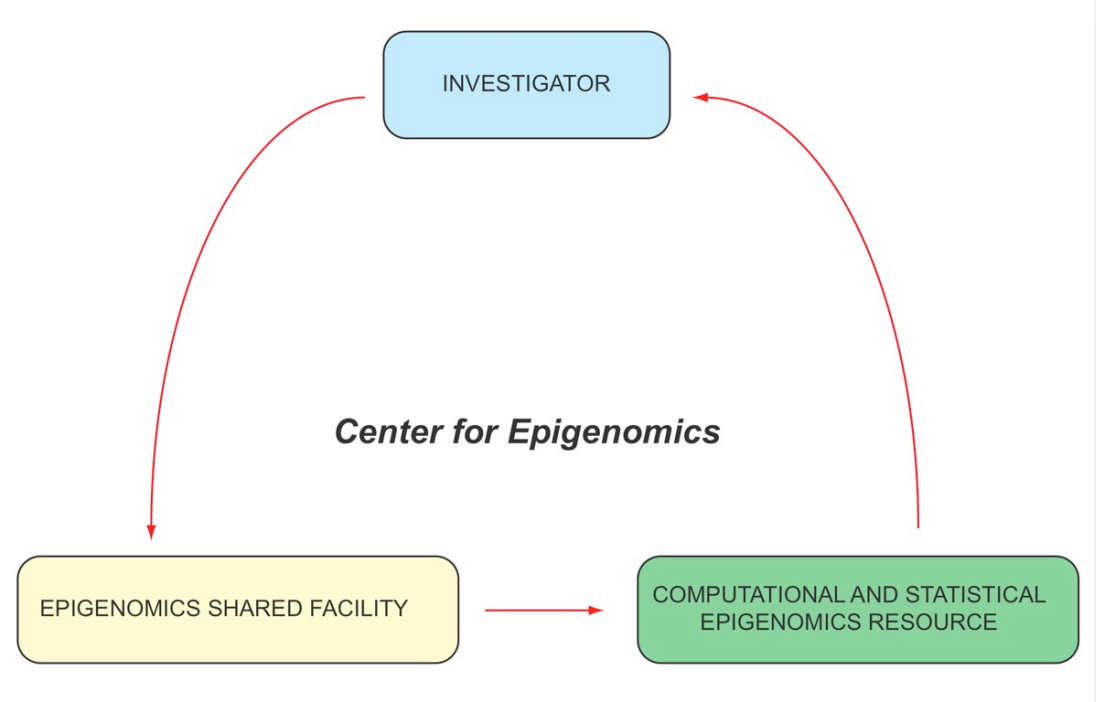
The epigenome, a complex system

- Epigenetic regulators



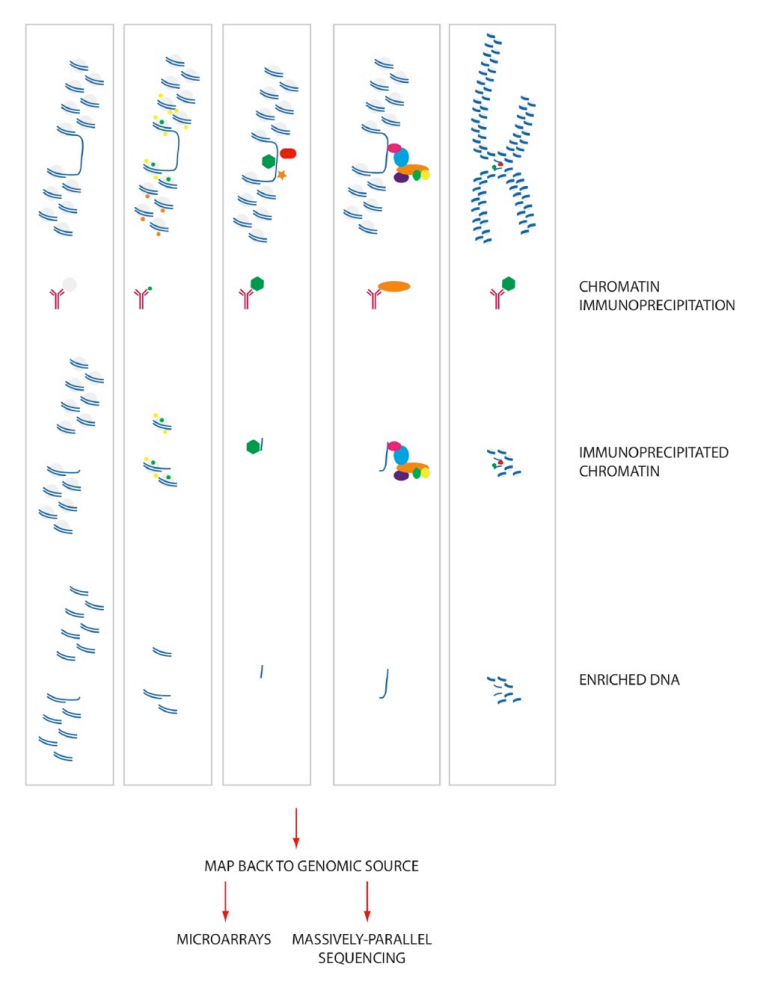
The Einstein Center for Epigenomics

- Epigenomic dysregulation in human disease
- Novel technologies
- Data management and analysis



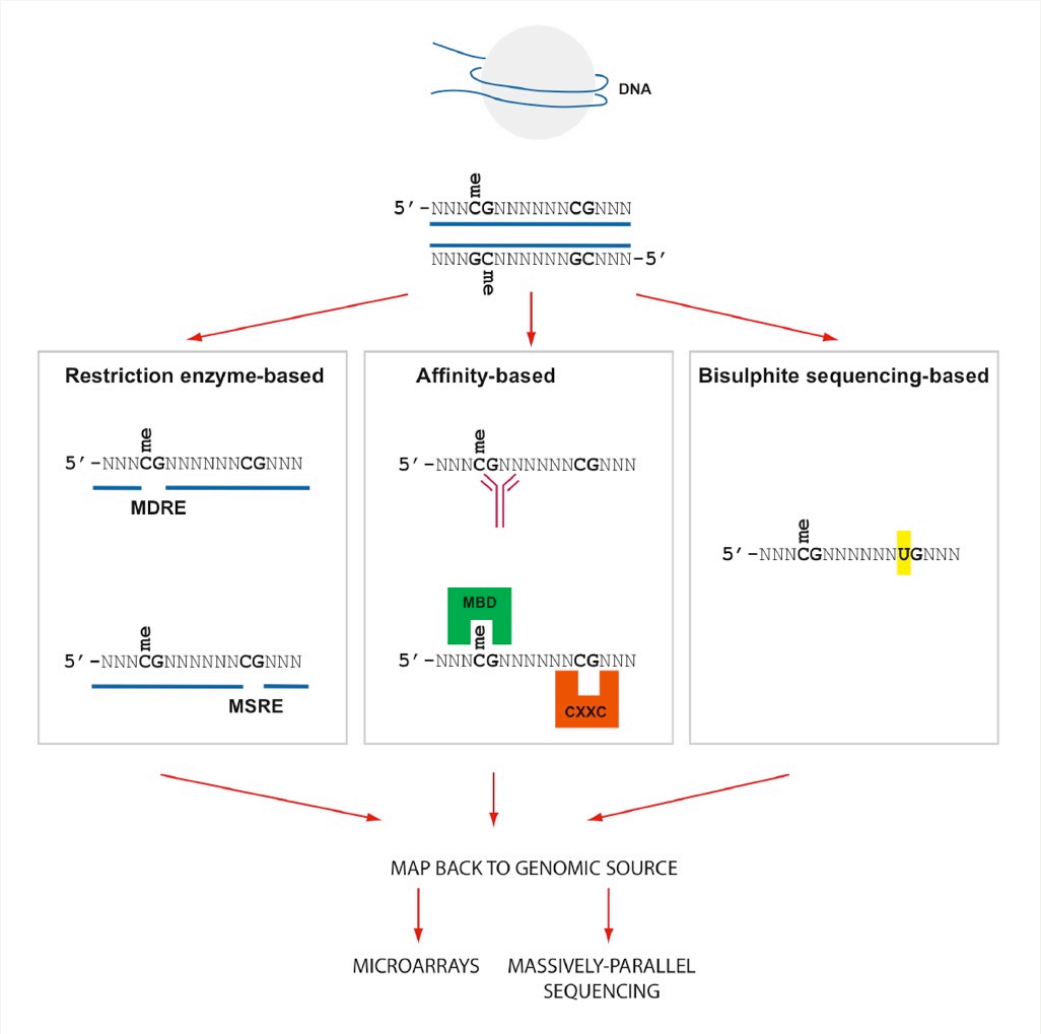
Epigenomic assays

- **Chromatin immunoprecipitation (ChIP) assays**



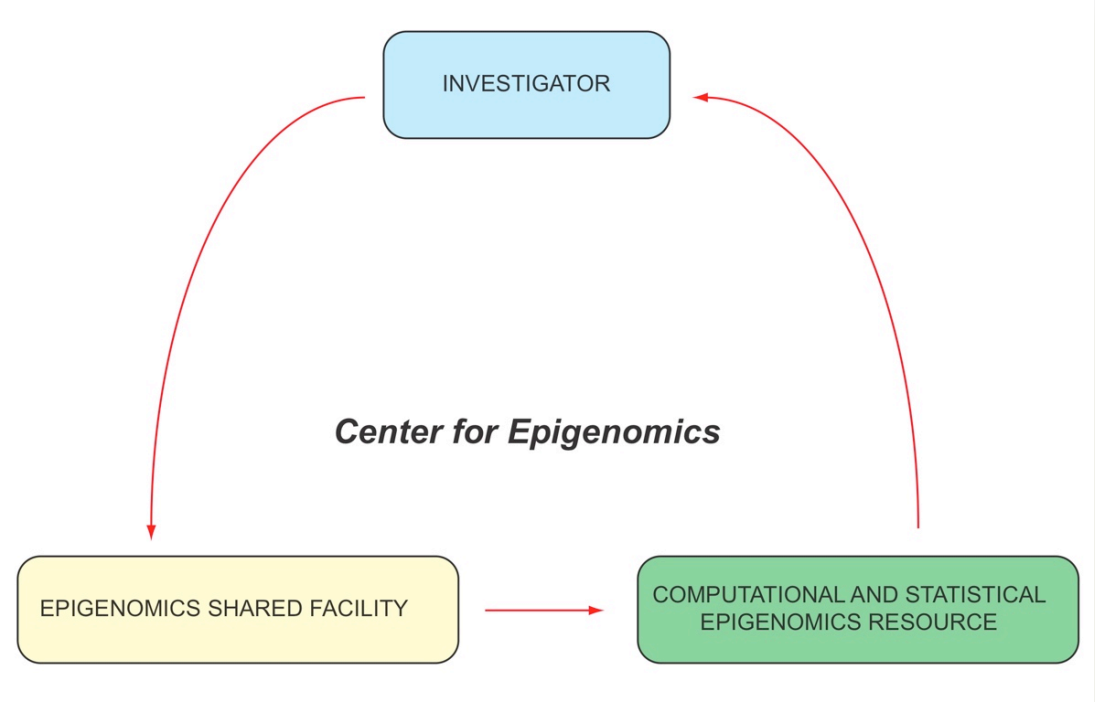
Epigenomic assays

- Cytosine methylation assays

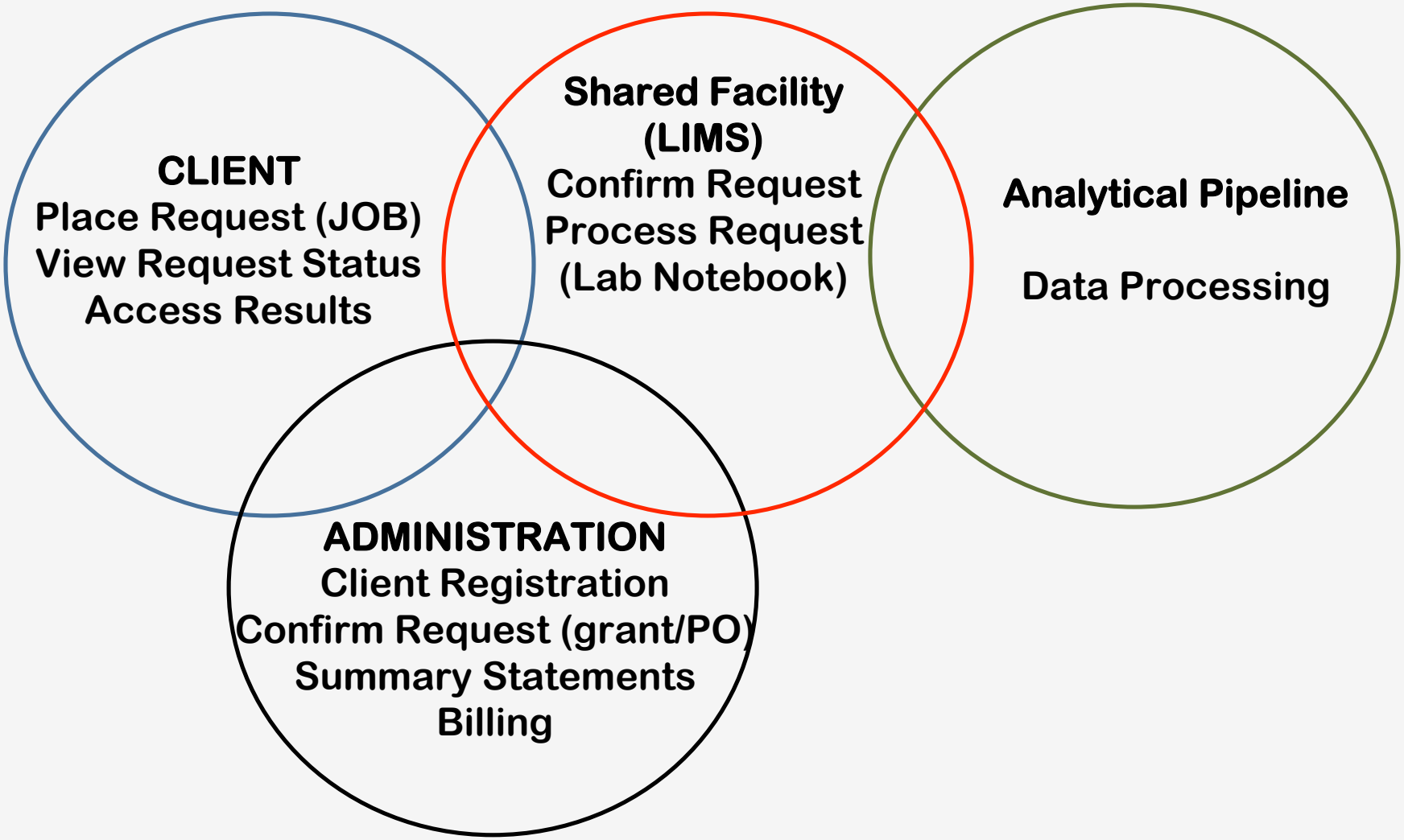


The Einstein Center for Epigenomics

- Sample submission & analysis need to be integrated
- Development of such a system at Einstein

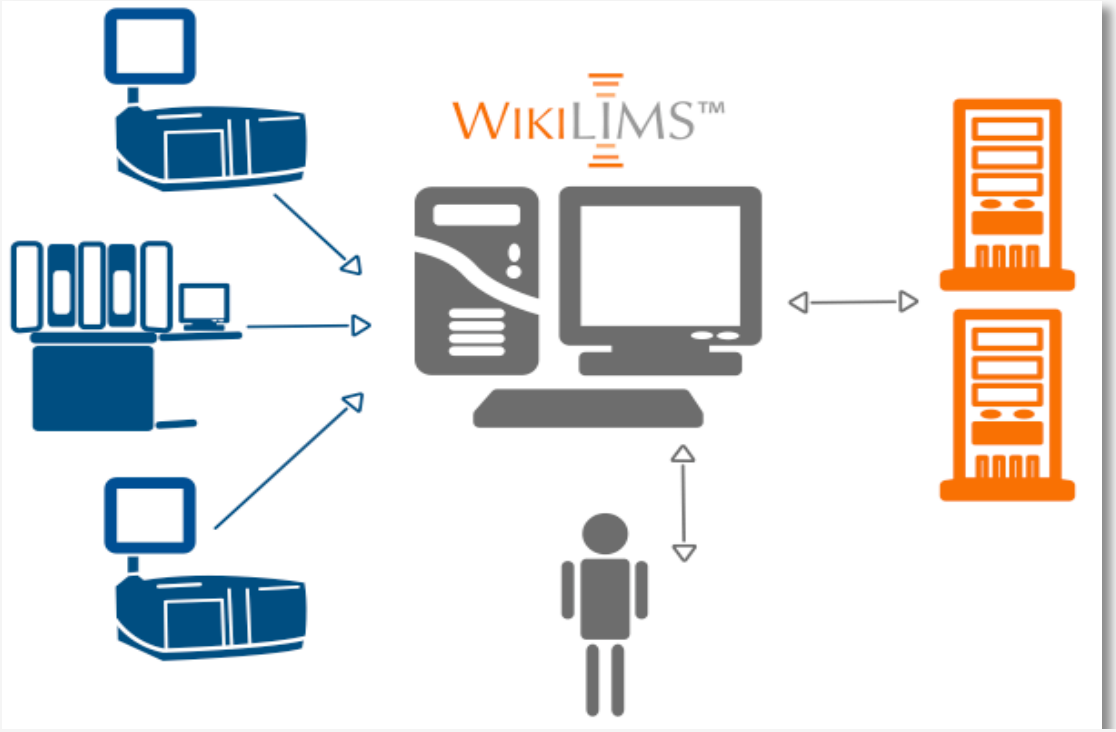


EPIGENOMICS SHARED FACILITY



WikiLIMs:

- A system that is quickly constructed and easily revised
- A system that can handle many data types
- A modular and transparent system
- A Web-based system



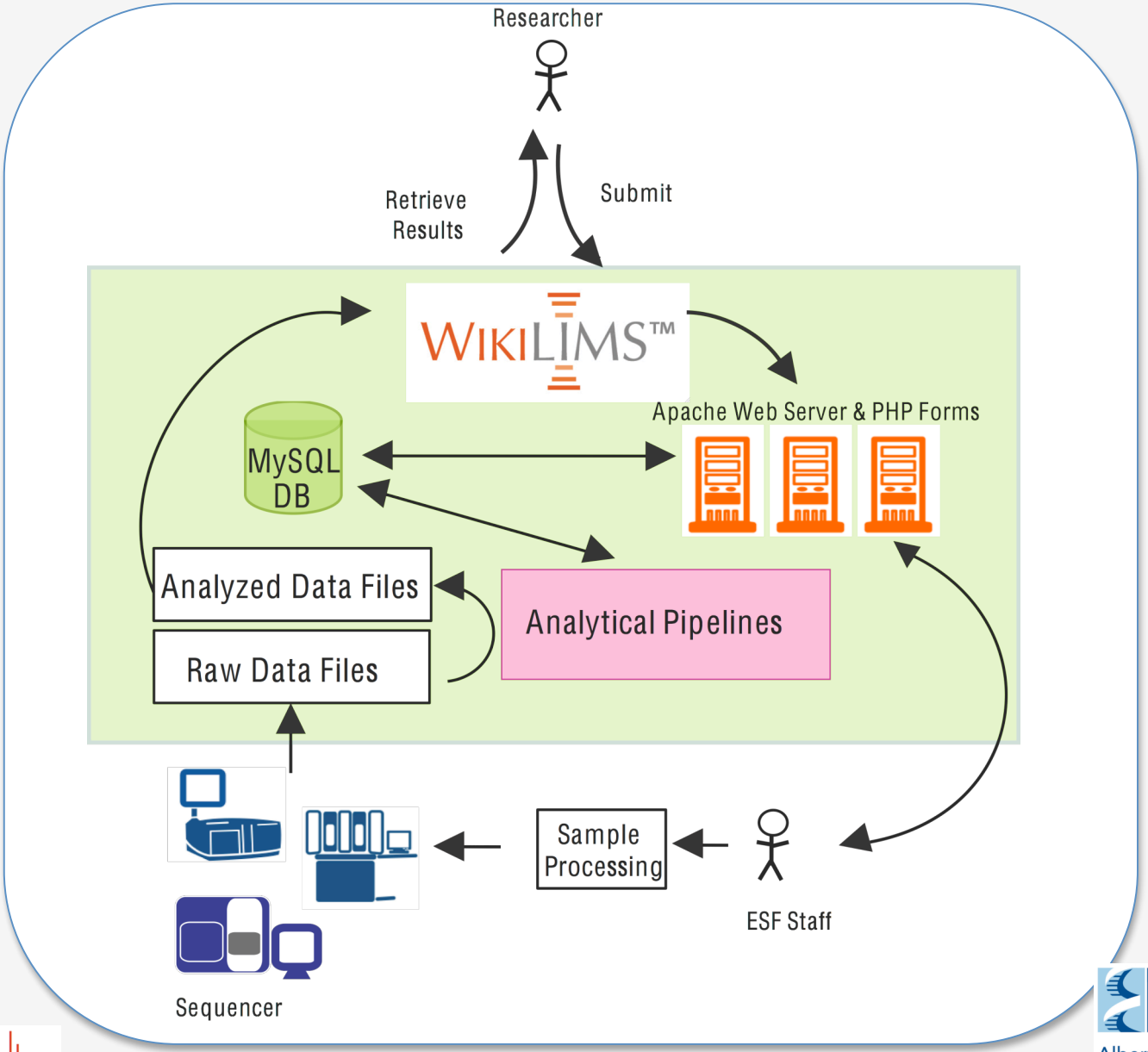
Powered By MediaWiki

BIOTEAM
Enabling Science

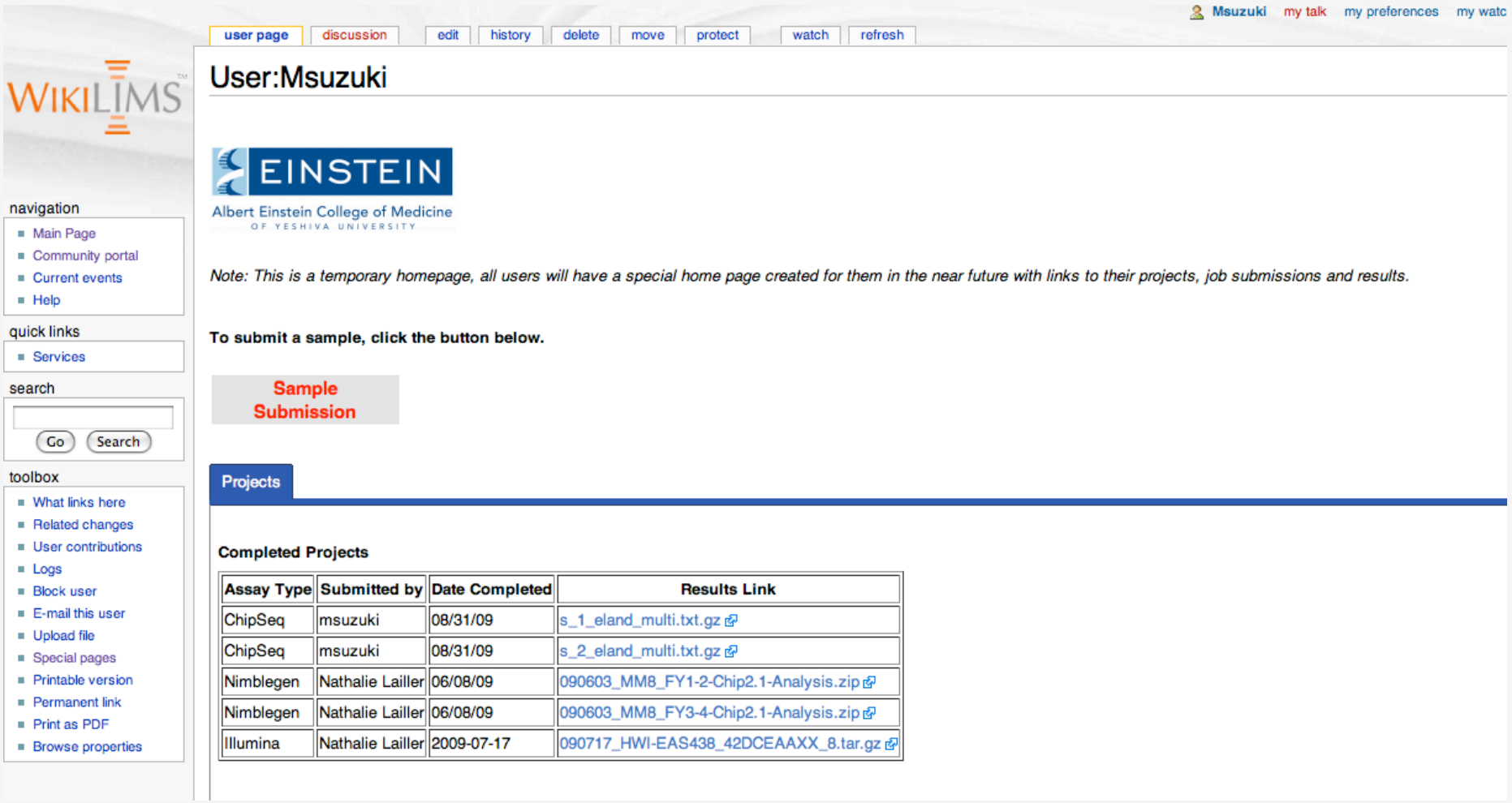
center for epigenomics | Einstein

EINSTEIN
Albert Einstein College of Medicine
OF YESHIVA UNIVERSITY

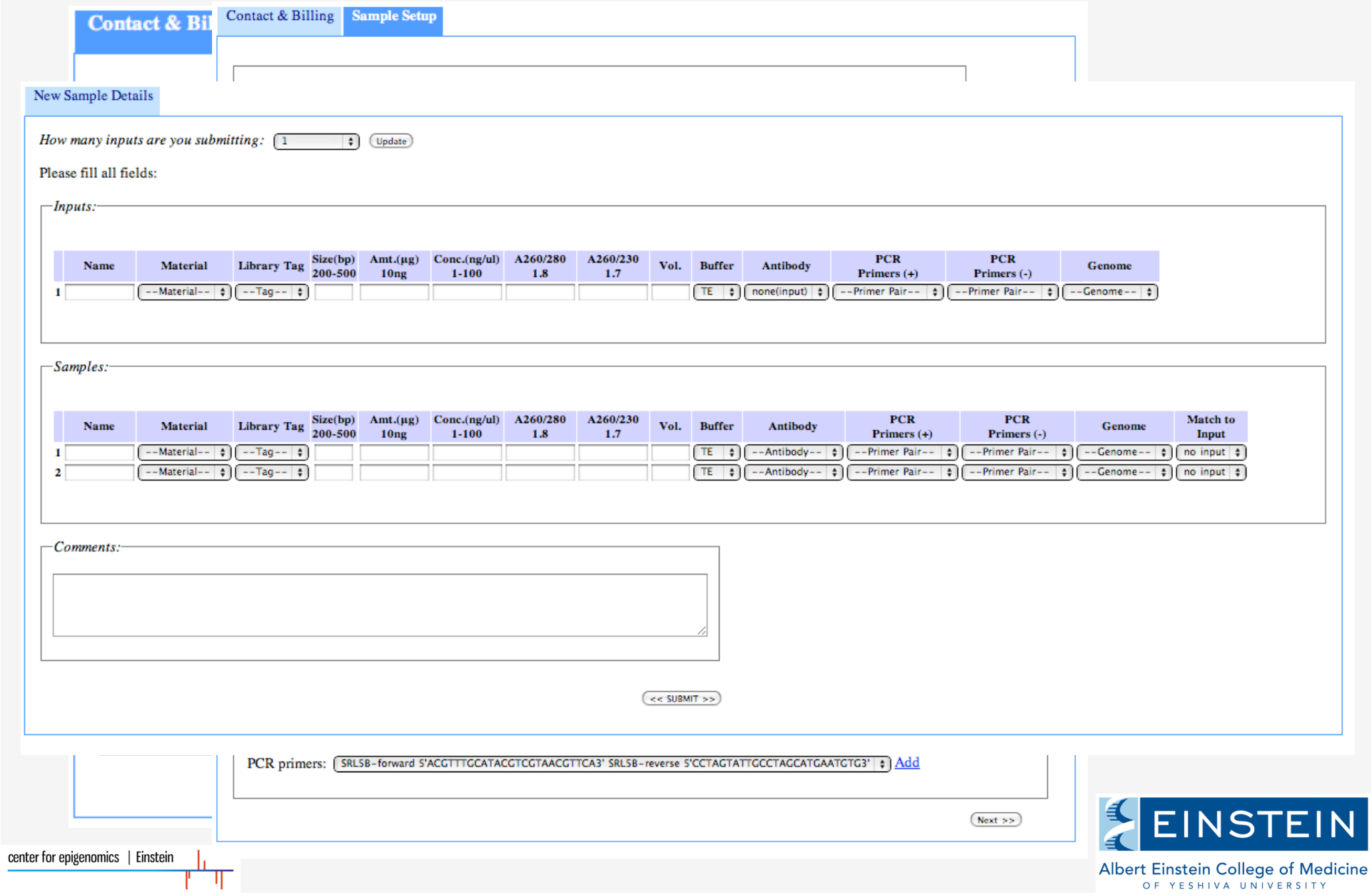
WikiLIMs:


Researcher
Retrieve Results
Submit
WIKILIMS™
Apache Web Server & PHP Forms
MySQL DB
Analyzed Data Files
Raw Data Files
Analytical Pipelines
Sample Processing
ESF Staff
Sequencer

WikiLIMs: User main page



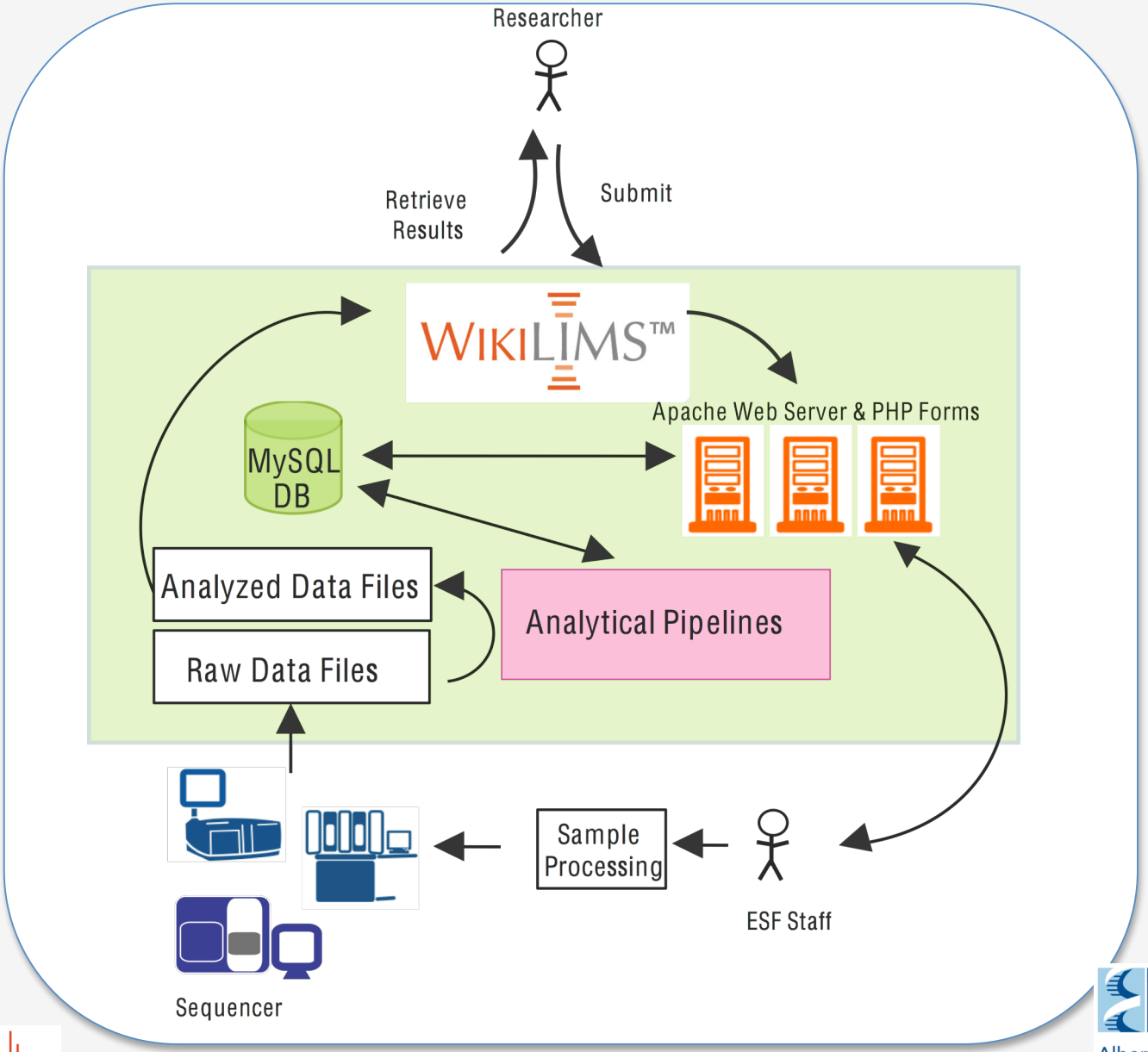
WikiLIMs:Sample submission management



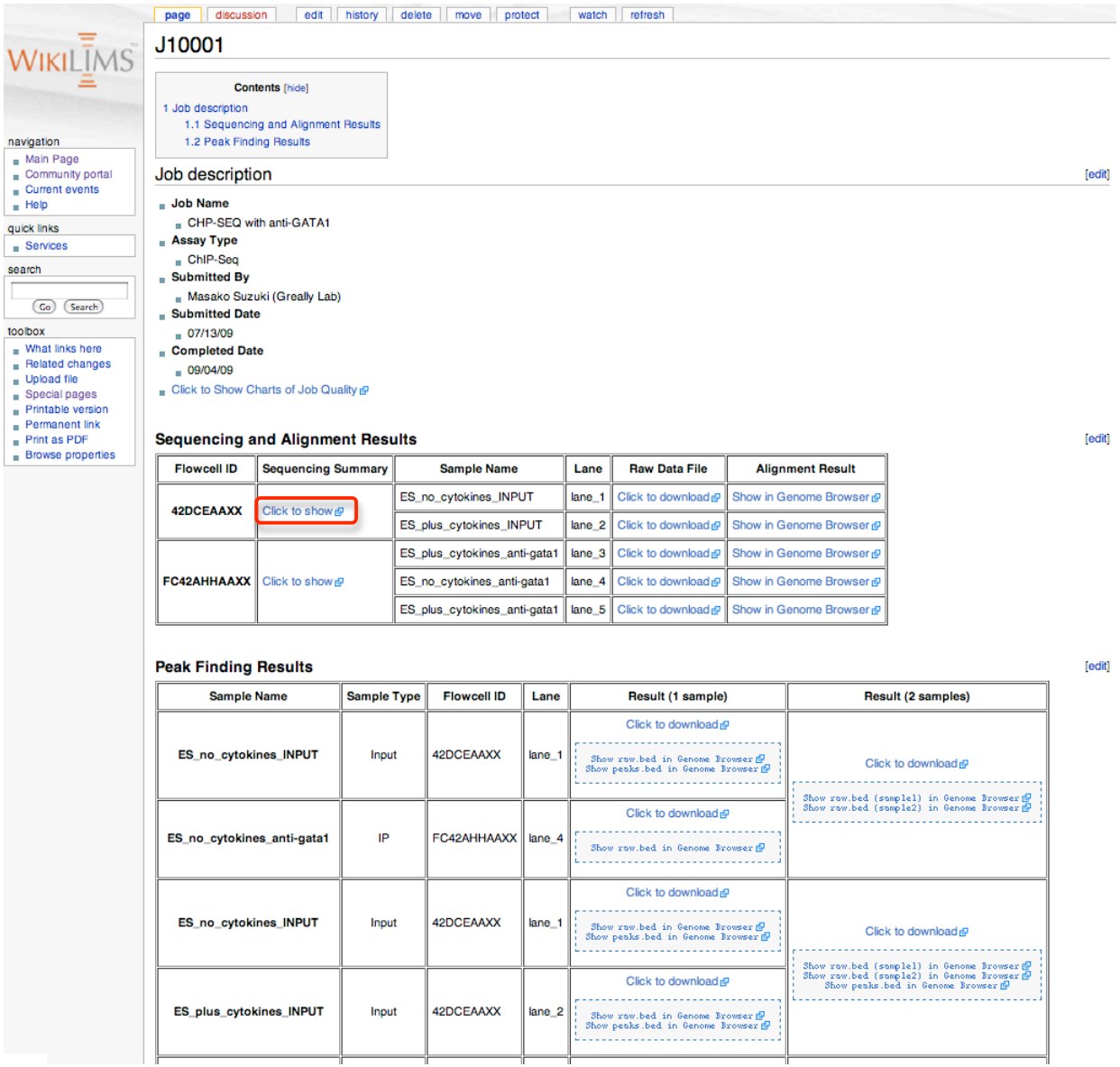
center for epigenomics | Einstein

EINSTEIN
Albert Einstein College of Medicine
OF YESHIVA UNIVERSITY

WikiLIMs:


Researcher
Retrieve Results
Submit
WIKILIMS™
Apache Web Server & PHP Forms
MySQL DB
Analyzed Data Files
Raw Data Files
Analytical Pipelines
Sample Processing
ESF Staff
Sequencer

WikiLIMs: Sequence data management



page | discussion | edit | history | delete | move | protect | watch | refresh

WIKILIMS

navigation
- Main Page
- Community portal
- Current events
- Help

quick links
- Services

search
Go Search

toolbox
- What links here
- Related changes
- Upload file
- Special pages
- Printable version
- Permanent link
- Print as PDF
- Browse properties

J10001

Contents [hide]
1 Job description
1.1 Sequencing and Alignment Results
1.2 Peak Finding Results

Job description [edit]

- **Job Name**
  - CHP-SEQ with anti-GATA1
- **Assay Type**
  - ChIP-Seq
- **Submitted By**
  - Masako Suzuki (Greally Lab)
- **Submitted Date**
  - 07/13/09
- **Completed Date**
  - 09/04/09
- Click to Show Charts of Job Quality

Sequencing and Alignment Results [edit]

Flowcell ID	Sequencing Summary	Sample Name	Lane	Raw Data File	Alignment Result
42DCEAAXX	Click to show	ES_no_cytokines_INPUT	lane_1	Click to download	Show in Genome Browser
		ES_plus_cytokines_INPUT	lane_2	Click to download	Show in Genome Browser
FC42AHHAAXX	Click to show	ES_plus_cytokines_anti-gata1	lane_3	Click to download	Show in Genome Browser
		ES_no_cytokines_anti-gata1	lane_4	Click to download	Show in Genome Browser
		ES_plus_cytokines_anti-gata1	lane_5	Click to download	Show in Genome Browser

Peak Finding Results [edit]

Sample Name	Sample Type	Flowcell ID	Lane	Result (1 sample)	Result (2 samples)
ES_no_cytokines_INPUT	Input	42DCEAAXX	lane_1	Click to download Show raw.bed in Genome Browser Show peaks.bed in Genome Browser	Click to download Show raw.bed (sample1) in Genome Browser Show raw.bed (sample2) in Genome Browser
ES_no_cytokines_anti-gata1	IP	FC42AHHAAXX	lane_4	Click to download Show raw.bed in Genome Browser	
ES_no_cytokines_INPUT	Input	42DCEAAXX	lane_1	Click to download Show raw.bed in Genome Browser Show peaks.bed in Genome Browser	Click to download Show raw.bed (sample1) in Genome Browser Show raw.bed (sample2) in Genome Browser Show peaks.bed in Genome Browser
ES_plus_cytokines_INPUT	Input	42DCEAAXX	lane_2	Click to download Show raw.bed in Genome Browser Show peaks.bed in Genome Browser	

WikiLIMs: Sequence data management

Summary Information For Experiment 090717_HWI-EAS438_42DCEAAXX

Chip Summary

Machine	HWI-EAS438
Run Folder	090717_HWI-EAS438_42DCEAAXX
Chip ID	unknown

Chip Results Summary

Clusters	Clusters (PF)	Yield (kbases)
89175116	33045159	1189626

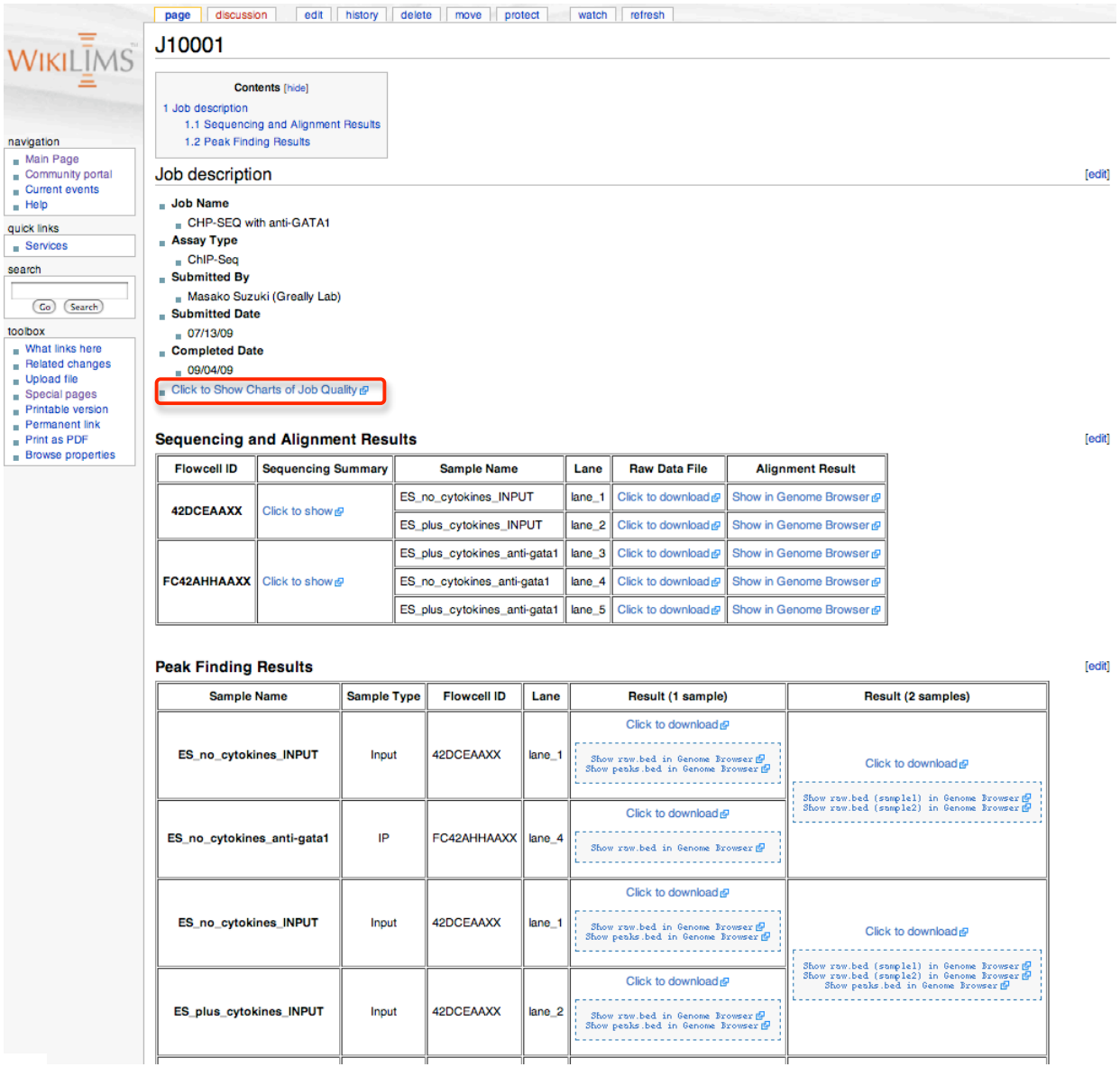
Lane Parameter Summary

Lane	Sample ID	Sample Target	Sample Type	Length	Filter	Chast. Thresh.	Num Tiles	Tiles
1	unknown	ucsd_mm_9	ELAND_EXTENDED	35	'((FAILED_CHASTITY<=1.000000))'	0.600000	100	Lane 1
2	unknown	ucsd_mm_9	ELAND_EXTENDED	35	'((FAILED_CHASTITY<=1.000000))'	0.600000	100	Lane 2
3	unknown	ucsd_mm_9	ELAND_EXTENDED	35	'((FAILED_CHASTITY<=1.000000))'	0.600000	100	Lane 3
4	unknown	ucsd_mm_9	ELAND_EXTENDED	35	'((FAILED_CHASTITY<=1.000000))'	0.600000	100	Lane 4
5	unknown	Hs19	ELAND_EXTENDED	35	'((FAILED_CHASTITY<=1.000000))'	0.600000	100	Lane 5
6	unknown	Hs19	ELAND_EXTENDED	35	'((FAILED_CHASTITY<=1.000000))'	0.600000	100	Lane 6
7	unknown	Scer288c	ELAND_EXTENDED	35	'((FAILED_CHASTITY<=1.000000))'	0.600000	100	Lane 7
8	unknown	ucsd_mm_9	ELAND_EXTENDED	35	'((FAILED_CHASTITY<=1.000000))'	0.600000	100	Lane 8

Lane Results Summary

Lane Info		Tile Mean +/- SD for Lane							
Lane	Lane Yield (kbases)	Clusters (raw)	Clusters (PF)	1st Cycle Int (PF)	% intensity after 20 cycles (PF)	% PF Clusters	% Align (PF)	Alignment Score (PF)	% Error Rate (PF)
1	85438	75939 +/- 9990	23972 +/- 6961	12 +/- 3	72.20 +/- 12.78	31.73 +/- 8.04	0.92 +/- 0.07	0.44 +/- 0.06	7.23 +/- 0.49
2	139730	106688 +/- 9424	38813 +/- 7938	15 +/- 4	72.12 +/- 14.20	36.46 +/- 6.93	1.38 +/- 0.12	0.74 +/- 0.08	5.68 +/- 0.51

WikiLIMs: Sequence data management



page | discussion | edit | history | delete | move | protect | watch | refresh

WIKILIMS

navigation
- Main Page
- Community portal
- Current events
- Help

quick links
- Services

search
Go Search

toolbox
- What links here
- Related changes
- Upload file
- Special pages
- Printable version
- Permanent link
- Print as PDF
- Browse properties

J10001

Contents [hide]
1 Job description
1.1 Sequencing and Alignment Results
1.2 Peak Finding Results

Job description [edit]

- **Job Name**
  - CHP-SEQ with anti-GATA1
- **Assay Type**
  - ChIP-Seq
- **Submitted By**
  - Masako Suzuki (Greally Lab)
- **Submitted Date**
  - 07/13/09
- **Completed Date**
  - 09/04/09
- Click to Show Charts of Job Quality

Sequencing and Alignment Results [edit]

Flowcell ID	Sequencing Summary	Sample Name	Lane	Raw Data File	Alignment Result
42DCEAAXX	Click to show	ES_no_cytokines_INPUT	lane_1	Click to download	Show in Genome Browser
		ES_plus_cytokines_INPUT	lane_2	Click to download	Show in Genome Browser
FC42AHHAAXX	Click to show	ES_plus_cytokines_anti-gata1	lane_3	Click to download	Show in Genome Browser
		ES_no_cytokines_anti-gata1	lane_4	Click to download	Show in Genome Browser
		ES_plus_cytokines_anti-gata1	lane_5	Click to download	Show in Genome Browser

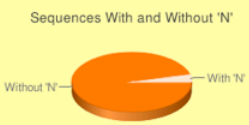
Peak Finding Results [edit]

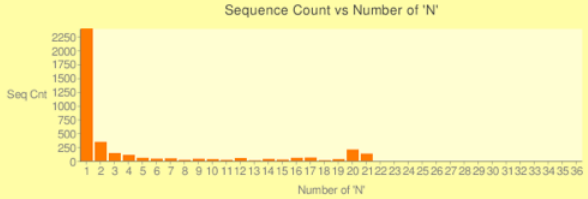
Sample Name	Sample Type	Flowcell ID	Lane	Result (1 sample)	Result (2 samples)
ES_no_cytokines_INPUT	Input	42DCEAAXX	lane_1	Click to download Show raw.bed in Genome Browser Show peaks.bed in Genome Browser	Click to download Show raw.bed (sample1) in Genome Browser Show raw.bed (sample2) in Genome Browser
ES_no_cytokines_anti-gata1	IP	FC42AHHAAXX	lane_4	Click to download Show raw.bed in Genome Browser	
ES_no_cytokines_INPUT	Input	42DCEAAXX	lane_1	Click to download Show raw.bed in Genome Browser Show peaks.bed in Genome Browser	Click to download Show raw.bed (sample1) in Genome Browser Show raw.bed (sample2) in Genome Browser Show peaks.bed in Genome Browser
ES_plus_cytokines_INPUT	Input	42DCEAAXX	lane_2	Click to download Show raw.bed in Genome Browser Show peaks.bed in Genome Browser	

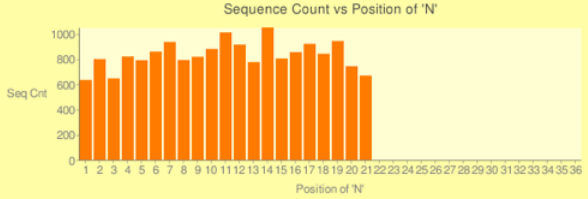
WikiLIMs: Sequence data management

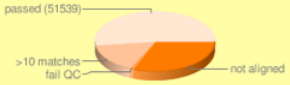
Sequence Quality Statistics

ES no cytokines INPUT (flowcell: 42DCEAAXX lane 1)







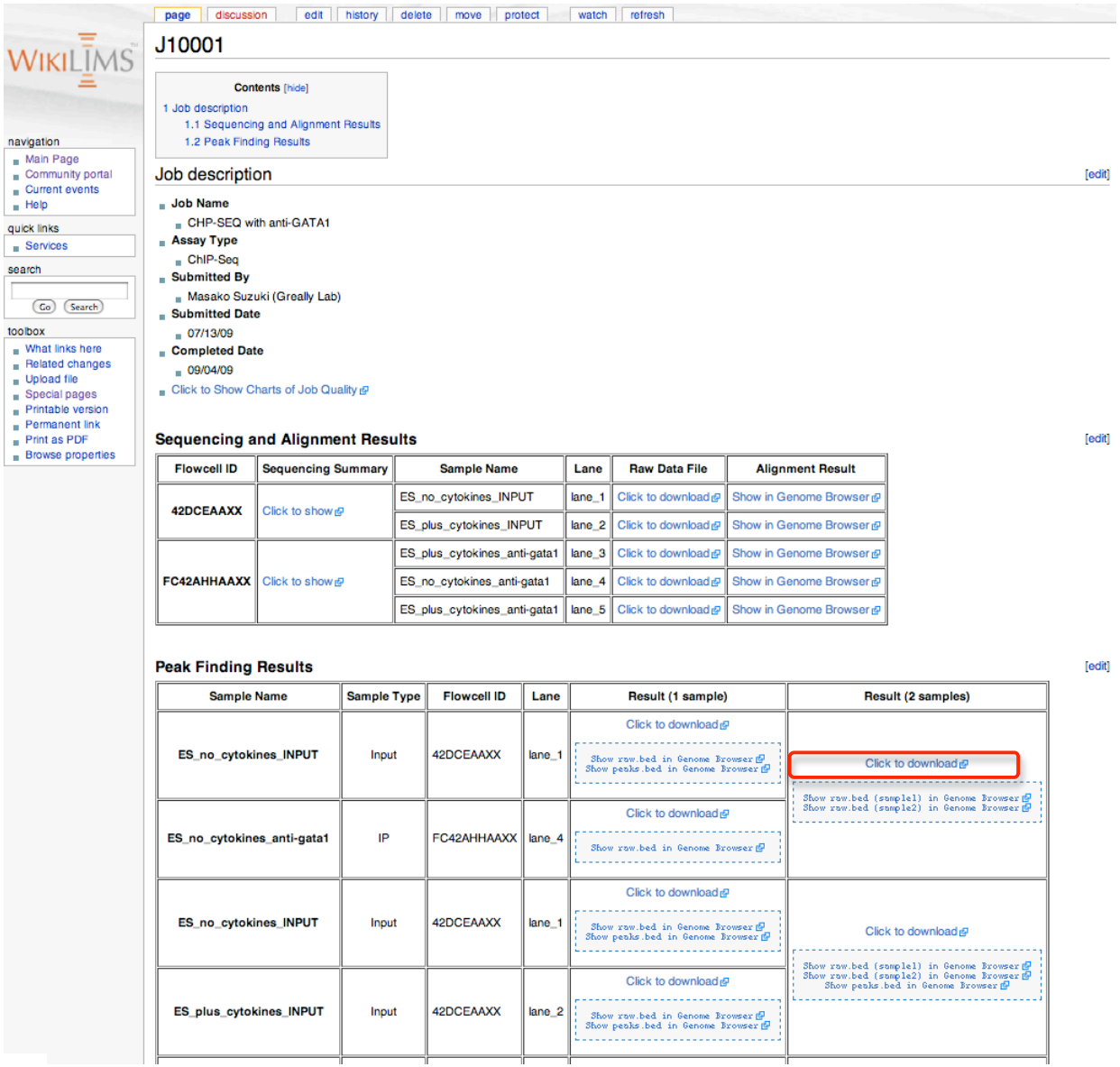


Run Quality Parameters

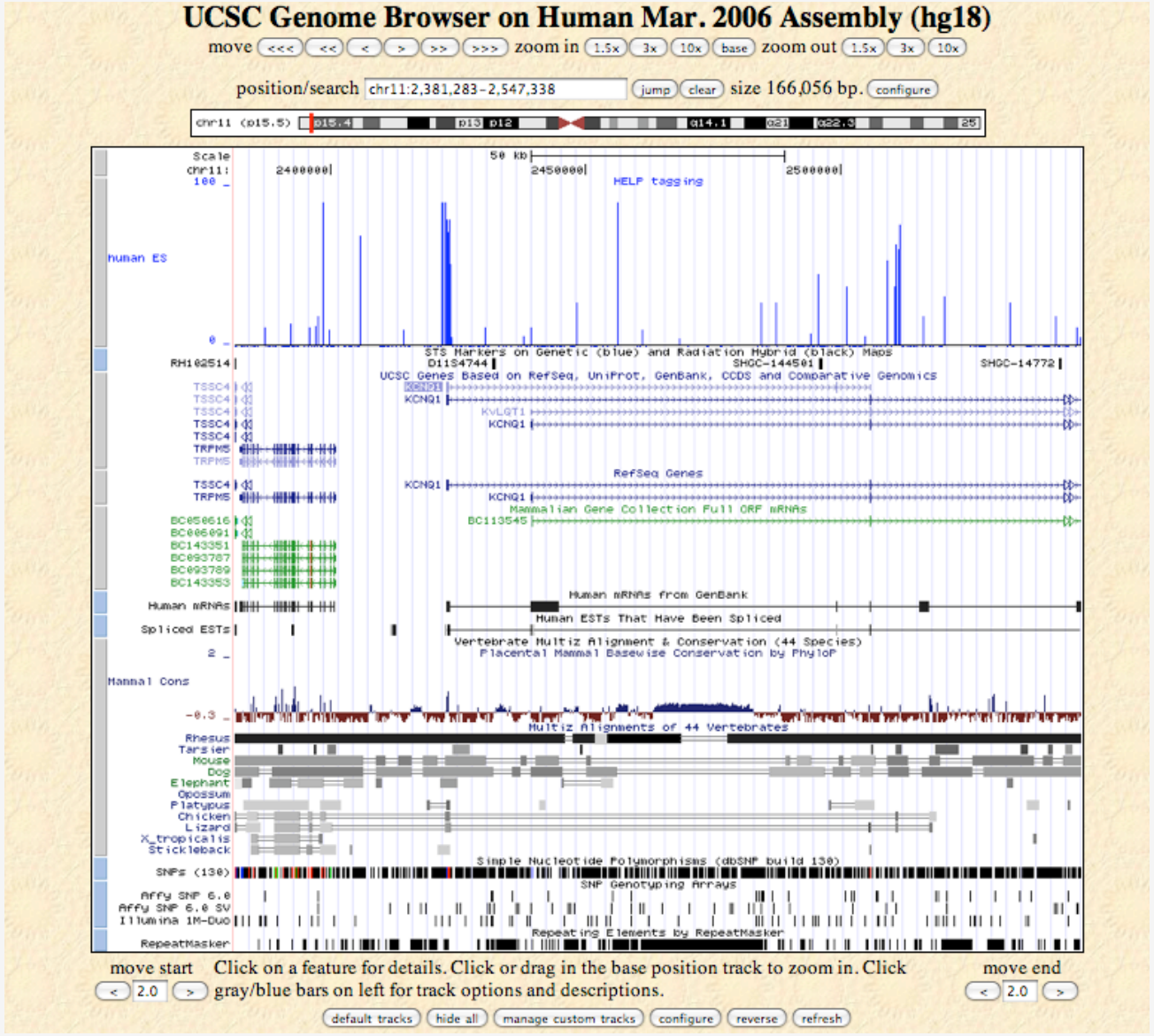
These quality metrics are based principally on single read 36bp sequencing of human DNA. The optimal results will vary according to experiment type.

Metric	Result	Uniformity (across tiles)	Notes
Total Yield			Result is 139730 (target is >1Gbases).
Raw Cluster Count			Result is 106688 +/- 9424 (target is >20,000).
% Clusters Passing Filter (PF)			Result is 36.46 +/- 6.93 (target is > 70%). Low % may be indicative of high cluster number (clusters too close together to obtain a clean signal in early cycles)
% Clusters PF that Align Uniquely to Reference			Result is 1.38 +/- 0.12. Optimal value dependent on read-length, genome sequenced and completeness of reference. For 30mers and the human genome, < 80% may be normal.
% Error Rate of Clusters PF			Result is 5.68 +/- 0.51. Should be ~1.5% but in any case, as low as possible.
% Phasing			Result is 0.6900. Should be ~0.5% to no more than 1% but in any case, as low as possible.
% Prephasing			Result is 0.5400. Should both be ~0.5% to no more than 1% but in any case, as low as possible.
First Cycle Intensity			Result is 15 +/- 4. Should be >1000.
20th Cycle Intensity as % of First			Result is 72.12 +/- 14.20. Should be >50%. If too high, suspect relatively low first cycle intensity

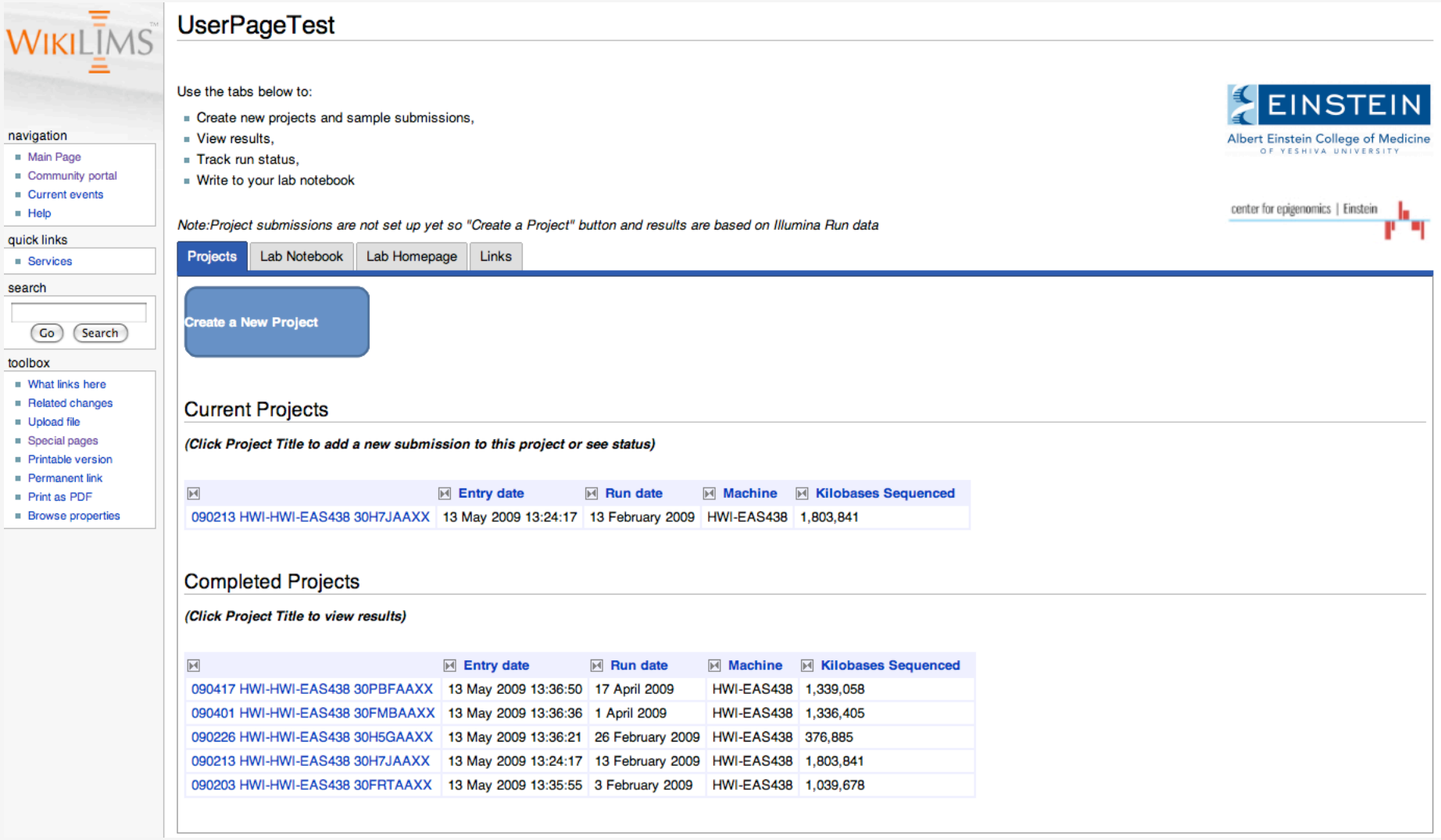
WikiLIMs: Sequence data management



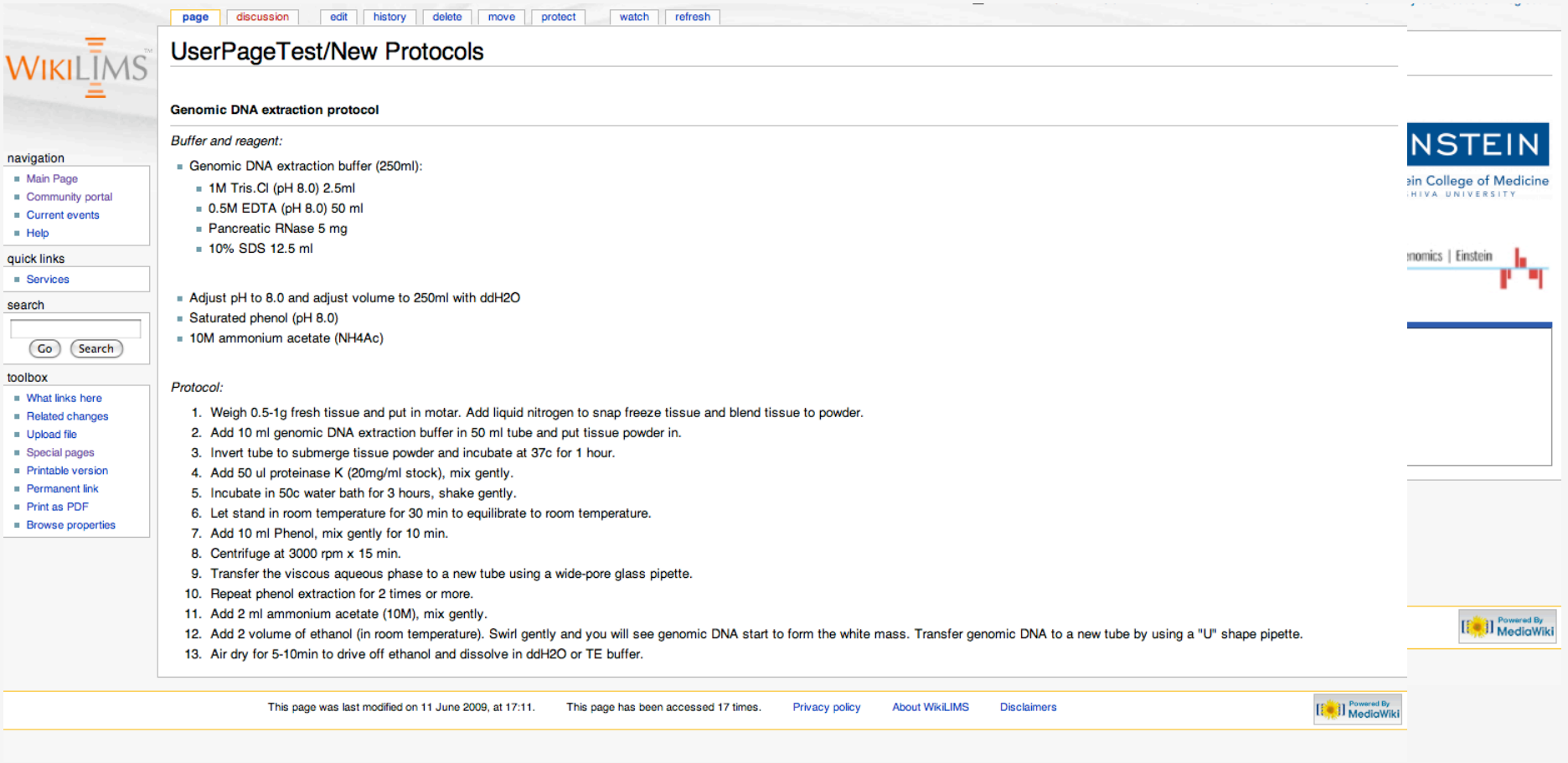
WikiLIMs: Sequence data management



WikiLIMs: electronic lab notebook



WikiLIMs: electronic lab notebook



center for epigenomics | Einstein

EINSTEIN
Albert Einstein College of Medicine
OF YESHIVA UNIVERSITY

WikiLIMs: electronic lab notebook

page | discussion | edit | history | delete | move | protect | watch | refresh

WIKILIMS

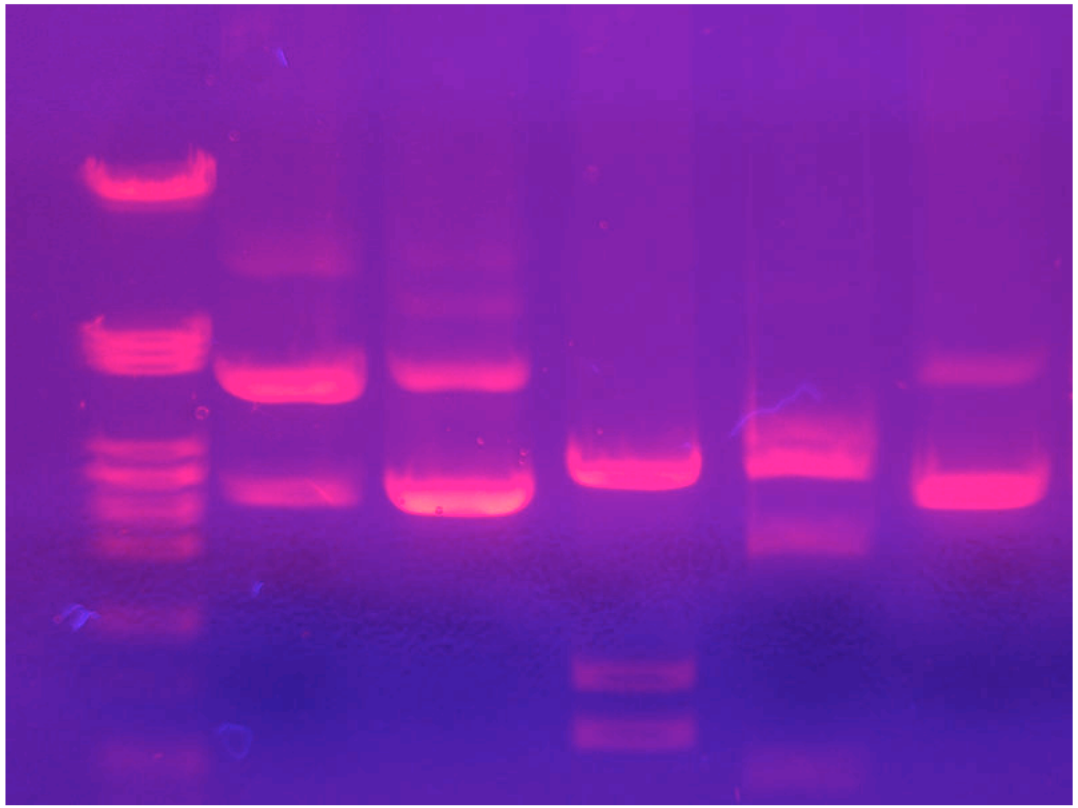
navigation
- Main Page
- Community portal
- Current events
- Help

quick links
- Services

search
Go Search

toolbox
- What links here
- Related changes
- Upload file
- Special pages
- Printable version
- Permanent link
- Print as PDF
- Browse properties

UserPageTest/Presentation Images



Description

Gel electrophoresis: 6 "DNA-tracks". In the first row (left), DNA with known fragment sizes was used as a reference. Different bands indicate different fragment sizes (the smaller, the faster it travels, the lower it is in the image); different intensities indicate different concentrations (the brighter, the more DNA). DNA was made visible using ethidium bromide and ultraviolet light.

Summary

Development of new epigenomic assays critically dependent on analytical capacity

Provision of these assays to researchers depends further on integration of data generation and data analysis services

WikiLIMS used to provide interface to analytical services

Automated analytical pipelines increase efficiency of services

These systems are inherently expandable and adaptable

Greally Lab

Dr. John Greally
Dr. Niki Athanasiadou
Dr. Niru Narayanan
Edyta Stasiek
Marién Pascual
Maria-Paz Ramos
Esther Berko
Kevin Lau

Former lab personnel

Dr. Mayumi Oda
Dr. Khulan Batbayar
Dr. Jacob Glass
Dr. Priti Tewari
Dr. Reid Thompson

Einstein

Dr. Melissa J. Fazzari

Computational and Statistical Epigenomic Resources

Dr. Andrew McLellan
Dr. Robert Dubin
Pilib Ó Broin
Qiang Jing (A.J.)

Bouhassira Lab
Dr. Eric Bouhassira
Dr. Emmanuel Olivier

Epigenomics Shared Facility
Dr. Shahina Maqbool
Raul Orea
Gael Westby

Bioteam

Dr. Stan Gloss
Dr. Brian Osborne
Dr. Adam Kraut

center for epigenomics | Einstein

EINSTEIN

Albert Einstein College of Medicine
OF YESHIVA UNIVERSITY