Introduction to the Third Epigenome Informatics Workshop

NIH Roadmap
Epigenomics Data Analysis and Coordination Center

March 5-6 2012
Houston, Texas

Workshop Objective:

Catalyze Conversion

of Epigenomic Profiling Data

into Biological Insights

through Integrative Analysis

- Methods (Assays, Data Processing)
- Standards (Metadata, Interoperability)
- Data Resources (Human Epigenome Atlas)
- Tools (Epigenomic Toolset, Genboree Workbench)
- Use Cases / Case Studies
- Collaborative Opportunities / Networking / Exchange of Experience

Workshop Participants

Monday, March 5th 2012

Session 1

9:00 – 9:45 am Introduction to the Workshop – Aleks Milosavljevic

9:45 – 10:30 am *Quantitative profiling of histone modifications, peak calling and segmentation of epigenomic signals, and small RNA analysis* – Cristi Coarfa (EDACC)

10:30 – 11:15 am *Methylome mapping using MeDIP-seq, WGBS and RRBS and analysis of allelic imbalances* - Alan Harris (EDACC)

11:15 – 11:30 am *Break*

11:30 – 12:15 pm *Whole-genome bisulfite sequencing: comparative analysis of programs for mapping bisulfite reads* – Govind K Ramamoorthy (EDACC, Rob Waterland Laboratory)

12:15 – 1:00 pm *Methylome mapping using Illumina Human Methylome 450K arrays*- Bekim Sadikovic (Art Beaudet Laboratory)

1:00 – 2:00 pm Boxed lunch (outside auditorium)

Monday, March 5th 2012

Session 1

9:00 – 9:45 am Introduction to the Workshop – Aleks Milosavljevic

9:45 – 10:30 am *Quantitative profiling of histone modifications, peak calling and segmentation of epigenomic signals, and small RNA analysis* – Cristi Coarfa (EDACC)

10:30 – 11:15 am *Methylome mapping using MeDIP-seq, WGBS and RRBS and analysis of allelic imbalances* - Alan Harris (EDACC)

11:15 – 11:30 am *Break*

11:30 – 12:15 pm *Whole-genome bisulfite sequencing: comparative analysis of programs for mapping bisulfite reads* – Govind K Ramamoorthy (EDACC, Rob Waterland Laboratory)

12:15 – 1:00 pm *Methylome mapping using Illumina Human Methylome 450K arrays*- Bekim Sadikovic (Art Beaudet Laboratory)

1:00 – 2:00 pm Boxed lunch (outside auditorium)

Monday, March 5th 2012

Session 1

9:00 – 9:45 am Introduction to the Workshop – Aleks Milosavljevic

9:45 – 10:30 am *Quantitative profiling of histone modifications, peak calling and segmentation of epigenomic signals, and small RNA analysis* – Cristi Coarfa (EDACC)

10:30 – 11:15 am *Methylome mapping using MeDIP-seq, WGBS and RRBS and analysis of allelic imbalances* - Alan Harris (EDACC)

11:15 – 11:30 am *Break*

11:30 – 12:15 pm *Whole-genome bisulfite sequencing: comparative analysis of programs for mapping bisulfite reads* – Govind K Ramamoorthy (EDACC, Rob Waterland Laboratory)

12:15 – 1:00 pm *Methylome mapping using Illumina Human Methylome 450K arrays*- Bekim Sadikovic (Art Beaudet Laboratory)

1:00 – 2:00 pm Boxed lunch (outside auditorium)

Monday, March 5th 2012

Session 1

9:00 – 9:45 am Introduction to the Workshop – Aleks Milosavljevic

9:45 – 10:30 am *Quantitative profiling of histone modifications, peak calling and segmentation of epigenomic signals, and small RNA analysis* – Cristi Coarfa (EDACC)

10:30 – 11:15 am *Methylome mapping using MeDIP-seq, WGBS and RRBS and analysis of allelic imbalances* - Alan Harris (EDACC)

11:15 – 11:30 am *Break*

11:30 – 12:15 pm *Whole-genome bisulfite sequencing: comparative analysis of programs for mapping bisulfite reads* – Govind K Ramamoorthy (EDACC, Rob Waterland Laboratory)

12:15 – 1:00 pm *Methylome mapping using Illumina Human Methylome 450K arrays*- Bekim Sadikovic (Art Beaudet Laboratory)

1:00 – 2:00 pm Boxed lunch (outside auditorium)

Monday, March 5th 2012

Session 2

2:00 – 2:45 pm *NIH Epigenomics Roadmap: reference epigenomes, metadata standards, interoperability, and integrative data analysis* – Aleks Milosavljevic (EDACC)

2:45 – 3:15 pm *Introduction to Genboree Workbench features that will be used in Case Studies* - Aleks Milosavljevic (EDACC)

3:15 – 3:30 pm *Break*

3:30 – 5:00 pm *Preparation for Case Studies on Day 2 – Setting up projects, databases, accessing files, navigating Genboree* - Matt Roth, Kevin Riehle, Chia-Chin Wu, Yuan Yuan, Cristi Coarfa, Alan Harris, Aleks Milosavljevic

6:00 pm Depart for Houston Livestock Show & Rodeo *or* dinner on your own

Monday, March 5th 2012

Session 2

2:00 – 2:45 pm *NIH Epigenomics Roadmap: reference epigenomes, metadata standards, interoperability, and integrative data analysis* – Aleks Milosavljevic (EDACC)

2:45 – 3:15 pm *Introduction to Genboree Workbench features that will be used in Case Studies* - Aleks Milosavljevic (EDACC)

3:15 – 3:30 pm *Break*

3:30 – 5:00 pm *Preparation for Case Studies on Day 2 – Setting up projects, databases, accessing files, navigating Genboree* - Matt Roth, Kevin Riehle, Chia-Chin Wu, Yuan Yuan, Cristi Coarfa, Alan Harris, Aleks Milosavljevic

6:00 pm Depart for Houston Livestock Show & Rodeo *or* dinner on your own

Tuesday, March 6th 2012

Session 3

9:00 – 9:30 am *Review of Day 1 and of Case Studies* – Matt Roth (EDACC)

9:30 – 11:30 pm ***Case Studies 1, 2, 5:*** *Epigenomic Variation between Tissues (Part 1)* – Matt Roth, Kevin Riehle, Chia-Chin Wu, Yuan Yuan, Cristi Coarfa, Alan Harris, Aleks Milosavljevic

11:30 - 11:45 pm Break

11:45 – 1:00 pm ***Case Studies 10, 9:*** *Epigenomic Variation Between Tissues (Part 2)* – Matt Roth, Kevin Riehle, Chia-Chin Wu, Yuan Yuan, Cristi Coarfa, Alan Harris, Aleks Milosavljevic

1:00 – 2:00 pm Boxed lunch (outside auditorium)

The data for several of the workshop Use Cases was kindly provided by Dr. Jonathan Mill (King's College London, UK), and is under review for this publication:

"Tissue-specific epigenetic variation across brain and blood: functional annotation of the human brain methylome". Matthew Davies[1], Manuela Volta[1], Abhishek Dixit[1], Simon Lovestone[1], Cristian Coarfa[2], R. Alan Harris[2], Aleksandar Milosavljevic[2], Claire Troakes[1], Safa Al-Sarraj[1], Richard Dobson[1], Leonard C. Schalkwyk[1], Jonathan Mill[1]*

[1]Institute of Psychiatry, King's College London. UK. [2]Baylor College of Medicine, Houston, Texas. USA. *Corresponding Author: Dr. Jonathan Mill, Address: Institute of Psychiatry, SGDP Centre, De Crespigny Park, Denmark Hill, London.

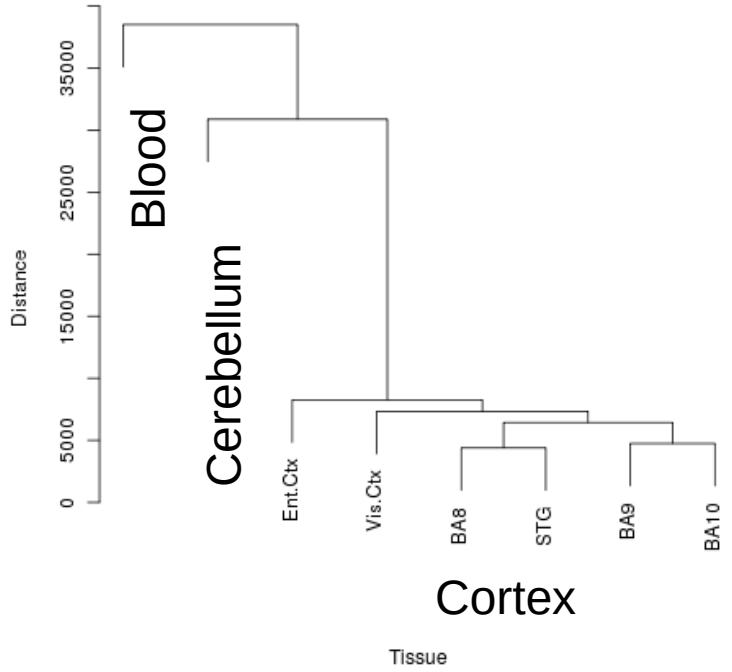
Since the paper is under review (Davies et al), it can not be shared with anyone outside of the workshop at this time, and we ask that you consider the data confidential. We will notify you when the data can be shared. Thank you for your understanding.

MEDIPS-processed signal averaged over 500bp windows genome-wide

Pairwise correlation between signals

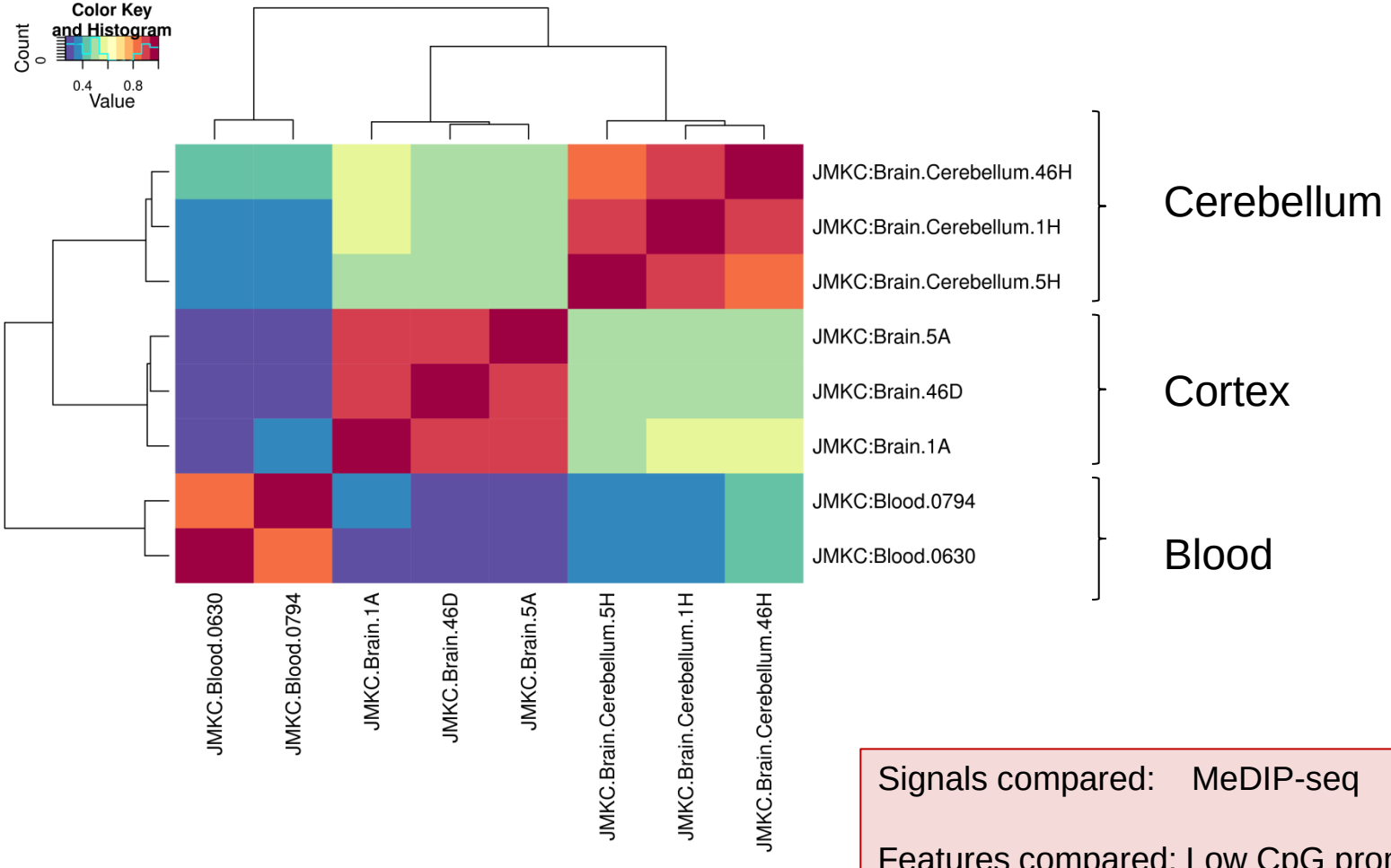
	BA9	BA10	BA8	STG	Ent Ctx	Vis Ctx	Cerebellum	Blood
BA9		0.97	0.95	0.94	0.92	0.92	0.57	0.64
BA10	0.97		0.96	0.95	0.94	0.93	0.59	0.66
BA8	0.95	0.96		0.96	0.92	0.92	0.57	0.58
STG	0.94	0.95	0.96		0.93	0.92	0.58	0.59
Ent Ctx	0.92	0.94	0.92	0.93		0.90	0.63	0.75
Vis Ctx	0.92	0.93	0.92	0.92	0.90		0.57	0.60
Cerebellum	0.57	0.59	0.57	0.58	0.63	0.57		0.49
Blood	0.64	0.66	0.58	0.59	0.75	0.60	0.49	

Hierarchical clustering discriminates tissues



Davies, Volta et al. In Review

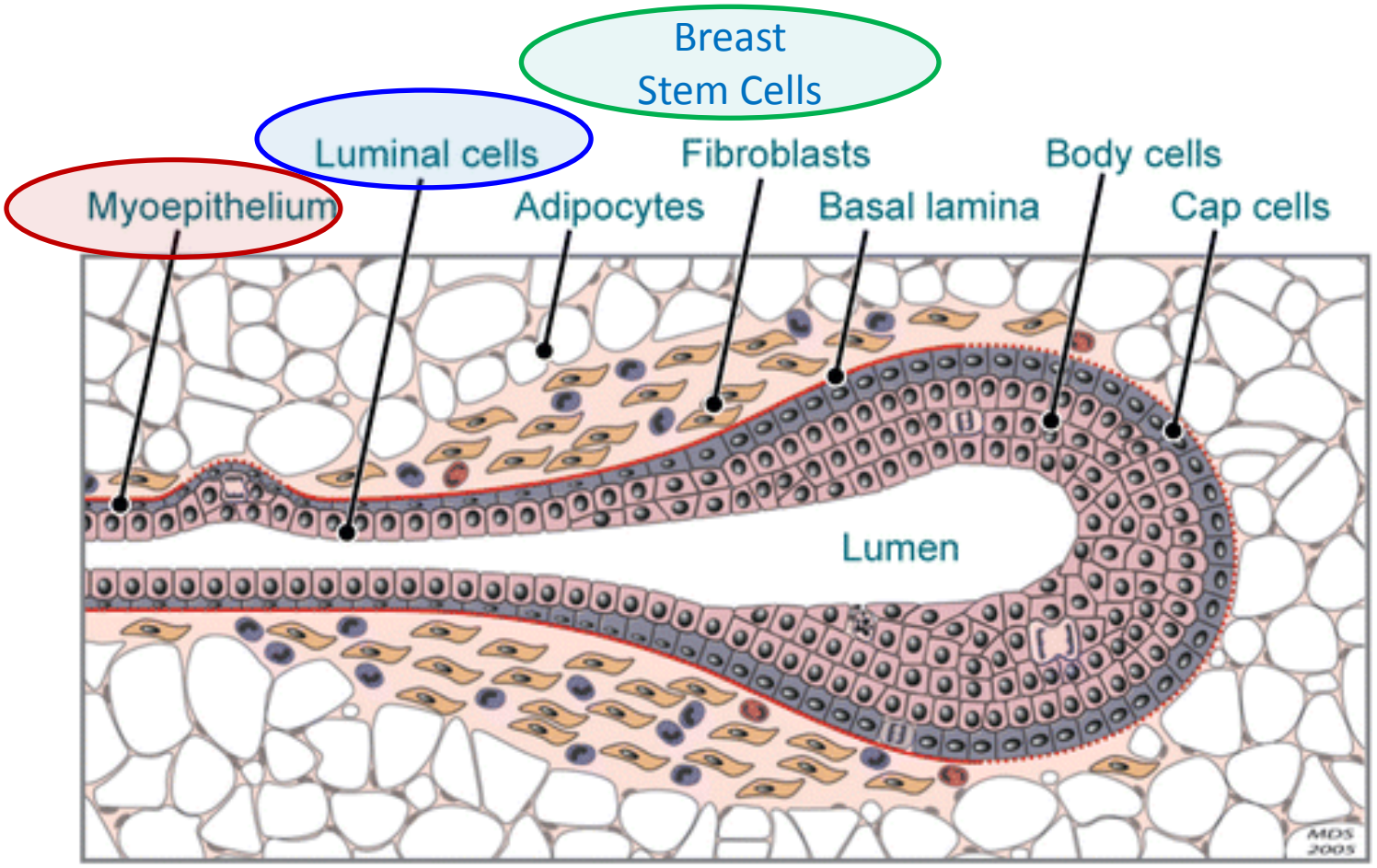
Use Case 1: Genomewide Patterns of Methylation can Distinguish Between Blood, Cerebellum, and Cortex



Signals compared: MeDIP-seq

Features compared: Low CpG promoters

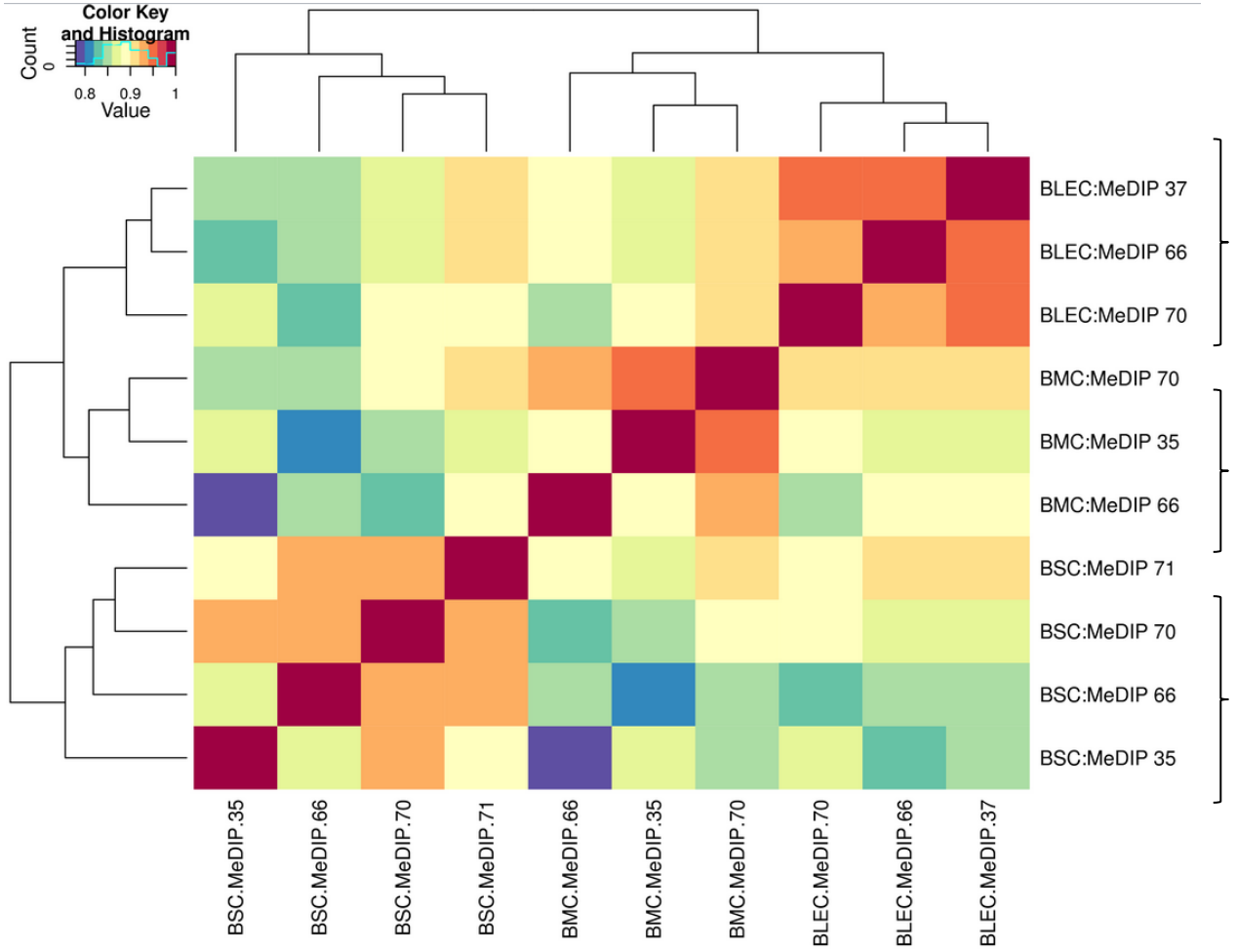
Use Case 2: Genomewide Patterns of Methylation can Distinguish Between Breast Cell Types



Hebner C, et al. 2008.
Annu. Rev. Pathol. Mech. Dis. 3:313–39

Use Case 2: Breast Cell Types Cluster Correctly Based on Their MeDIP-seq Profiles

(Epigenome Atlas and UCSF REMC data)



Breast luminal cells

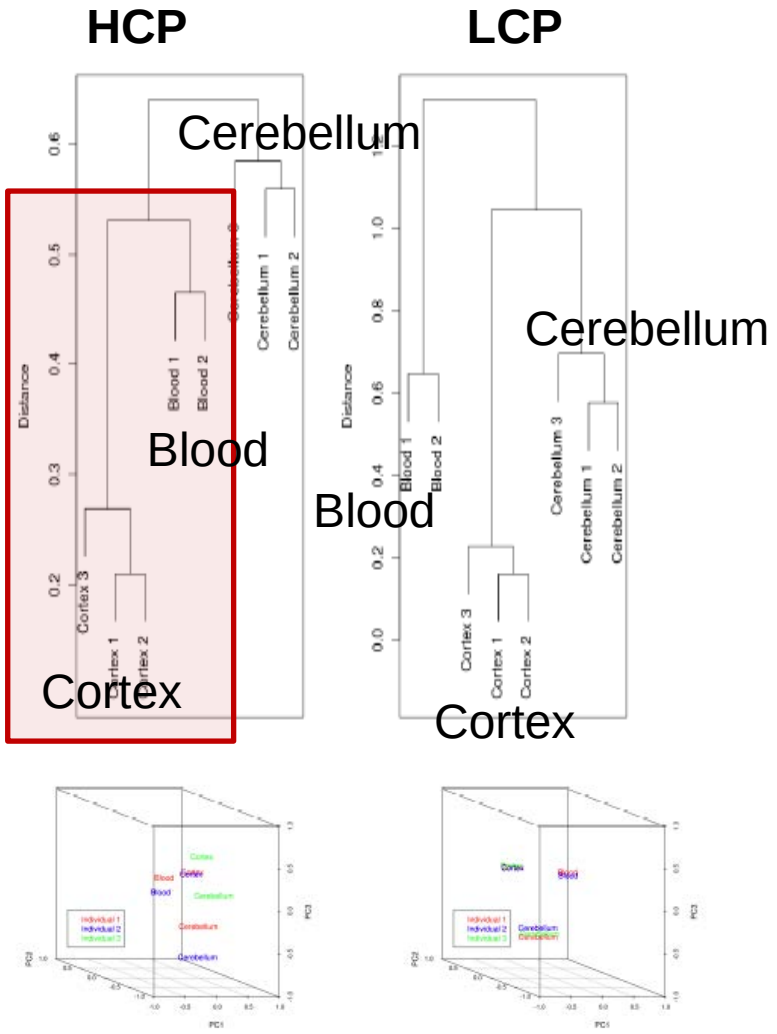
Breast myoepithelial cells

Breast stem cells

Data from: Epigenome Atlas, Release 5

Use Case 5: Methylation Profiles of Some Features More Informative About Cell Type Than Profiles of Others

Methylation of LCPs conveys more information about tissue type than methylation of HCPs.

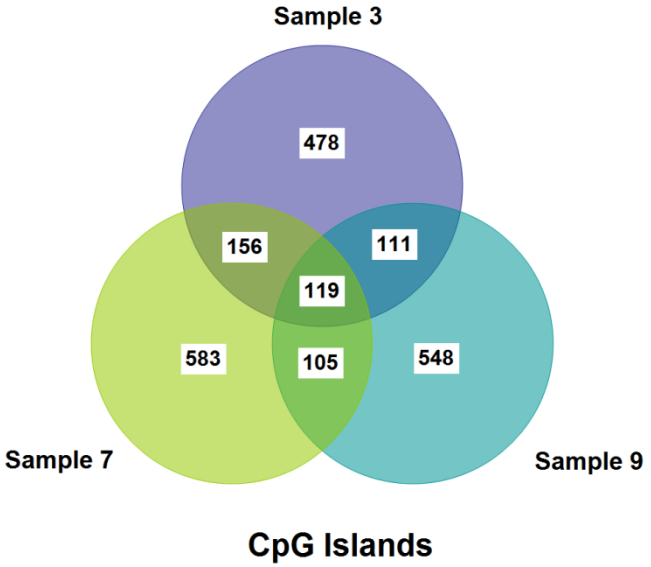


Use Case 10: Methylation changes during CD4+ T-cell maturation (Epigenome Atlas and UCSF REMC data)

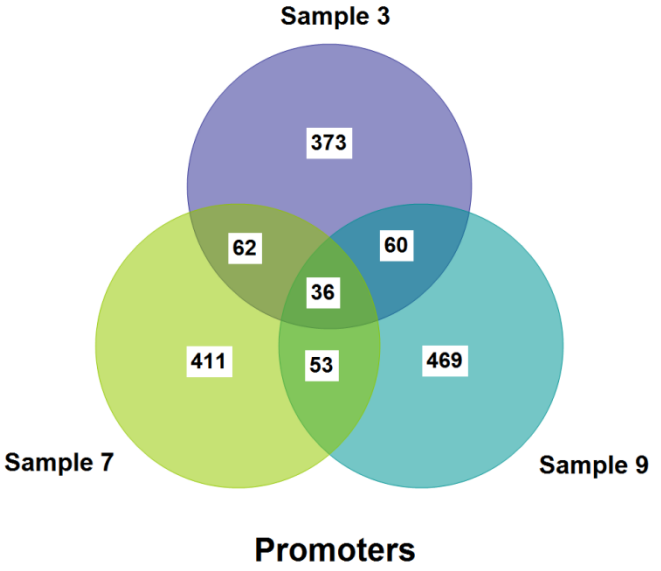
CD4+ Memory Primary T cells / CD4+ Naïve Primary T cells

methylation differences across 3 individuals

Genes with changes in associated CpG islands



Genes with changes in promoters



Use Case 10: Methylation changes during CD4+ T-cell maturation (Epigenome Atlas and UCSF REMC data)

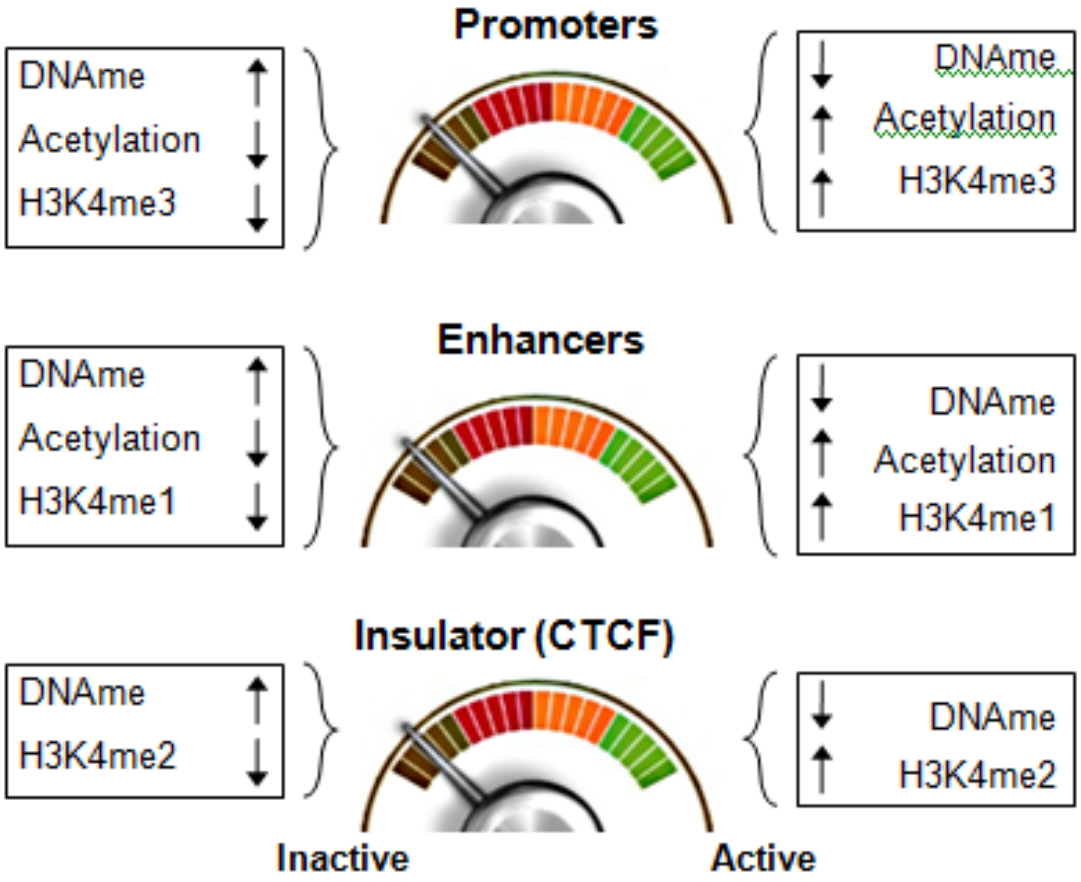
Methylation changes in **CpG islands**

KEGG pathways (via DAVID web site):

Sublist	Category	Term	RT	Genes	Count	%	P-Value	Benjamini
☐	KEGG_PATHWAY	T cell receptor signaling pathway	RT		5	5.9	2.3E-3	8.9E-2
☐	KEGG_PATHWAY	Natural killer cell mediated cytotoxicity	RT		5	5.9	4.9E-3	9.4E-2
☐	KEGG_PATHWAY	Wnt signaling pathway	RT		4	4.7	4.5E-2	4.6E-1
☐	KEGG_PATHWAY	Pathways in cancer	RT		5	5.9	9.2E-2	6.2E-1

Use Case 9: Coordinated Changes of Epigenomic Marks Across Tissue Types

"Active" or "inactive" states of TF binding elements (promoters, enhancers) may be inferred with certain probability based on the state of a handful of correlated epigenomic marks.



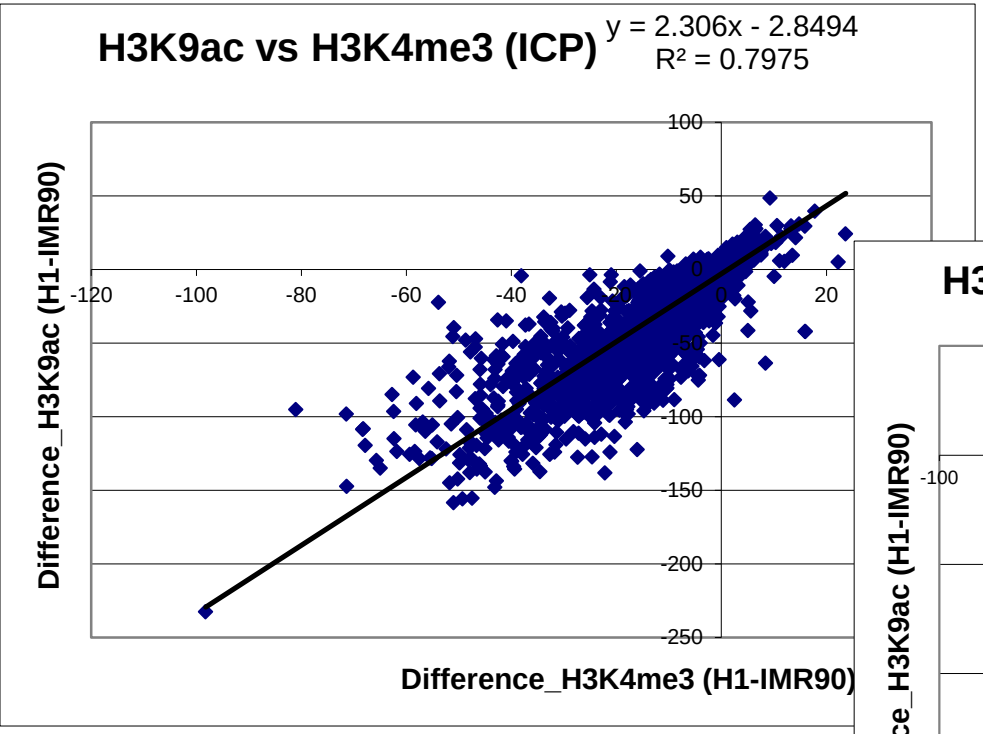
Use Case 9: Coordinated Changes of Epigenomic Marks Across Tissue Types

- High- and Intermediate CpG (HCP and ICP) promoters are regulated by different classes of TFs (including Polycomb complex) than Low-CpG (LCP) promoters.

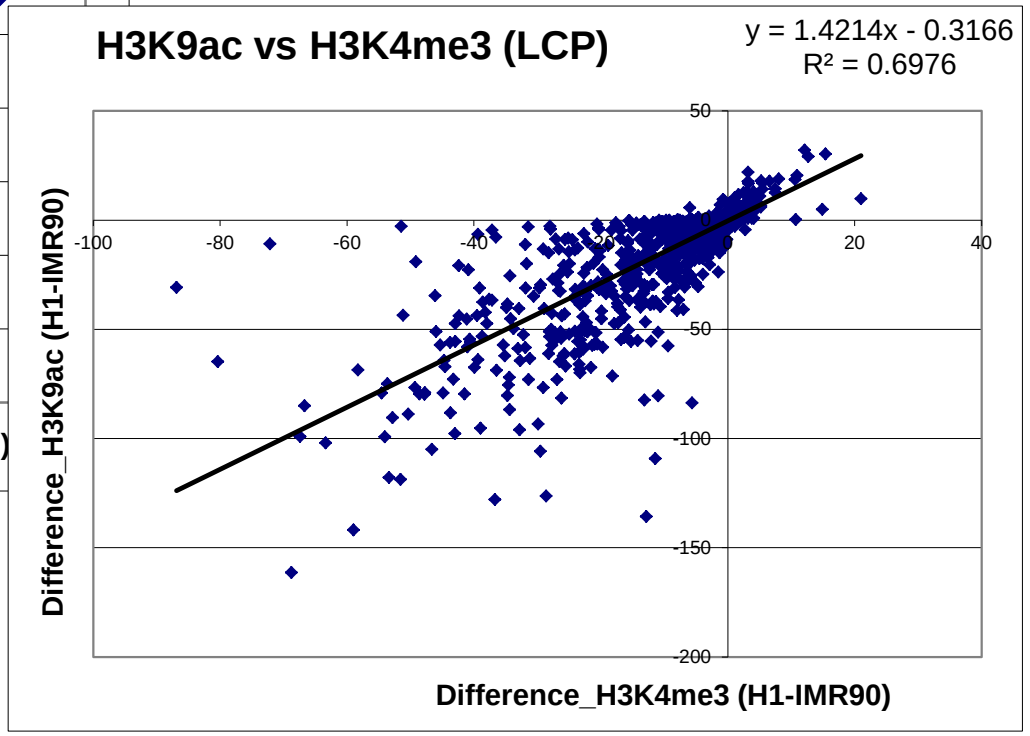
- It is therefore possible that HCP, ICP, and LCP promoters experience different patterns of epigenomic changes during cellular differentiation.

- Approach: **Examine changes in H3K9ac and H3K4me3 over HCP, ICP, and LCP promoters between an ES cell line (H1) and a fibroblast cell line (IMR90).**

Use Case 9: Coordinated Changes of Epigenomic Marks Across Tissue Types



Scatter Plots of H3K9ac and H3K4me3 Differences Between H1 and IMR90 Over ICP and LCP Promoters



Tuesday, March 6th 2012

Session 4

2:00 – 3:30 pm ***Case Study 7:*** *Epigenomic Variation Between Individuals* – Matt Roth, Kevin Riehle, Chia-Chin Wu, Yuan Yuan, Cristi Coarfa, Alan Harris, Aleks Milosavljevic

3:30 – 4:30 pm *Case Studies: Epigenomic Variation in Cancer* – Matt Roth, Kevin Riehle, Chia-Chin Wu, Yuan Yuan, Cristi Coarfa, Alan Harris, Aleks Milosavljevic

4:30 – 5:00 pm Open discussion and wrap-up

5:00 pm Adjourn

Use Case 7: Detecting Coordinated Changes in Cortex, Cerebellum and Blood Between Individuals

Data Source: Matthew Davies, M. et al. (submitted) *"Tissue-specific epigenetic variation across brain and blood: functional annotation of the human brain methylome".*

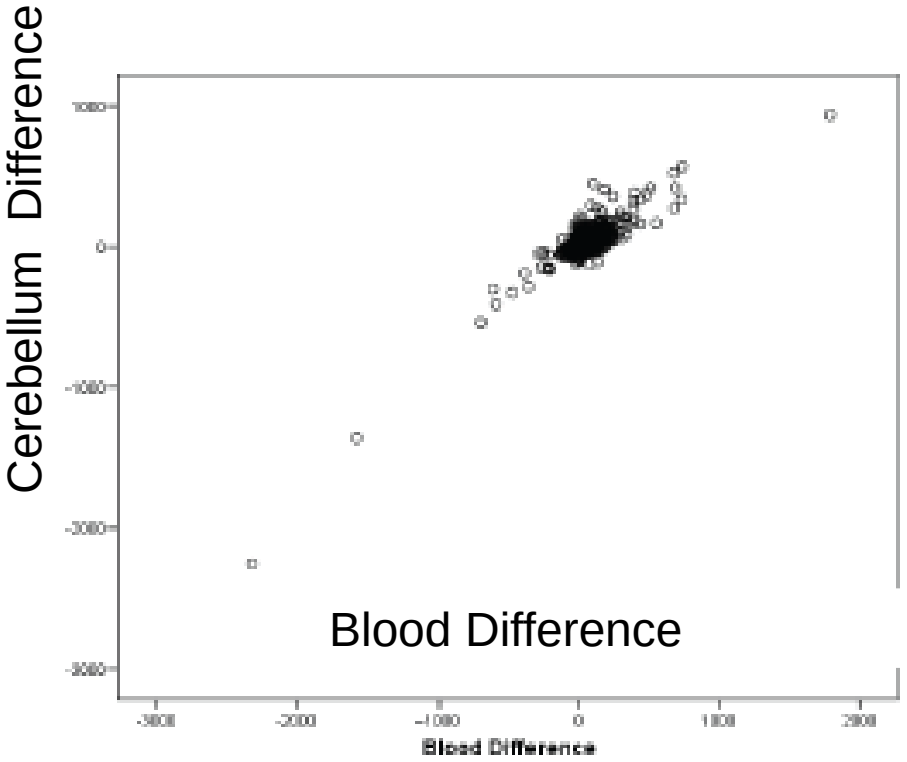
Use Case 7: Detecting Coordinated Changes in Cortex, Cerebellum and Blood Between Individuals

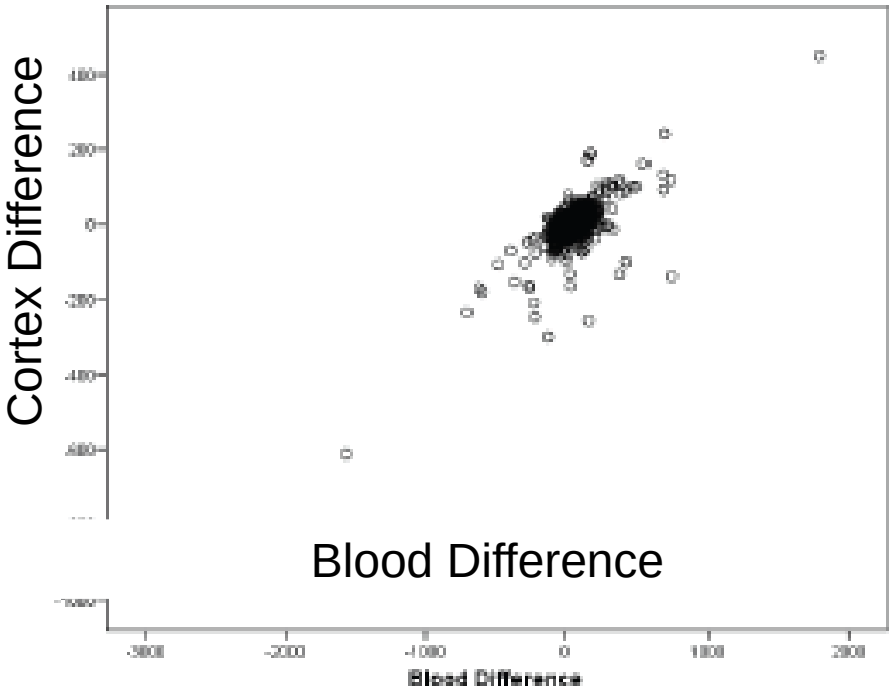
- To which extent the easily accessible peripheral tissue (e.g. whole blood) can be used to ask questions about **inter-individual phenotypic variation manifest in a phenotypically relevant but inaccessible tissue such as brain**.

- It is also of interest to delineate inter-individual differences that are genetic (say, due to copy-number polymorphisms) from those that may arise due to environmental, developmental, and physiological differences and disease processes.

Use Case 7: Detecting Coordinated Changes in Cortex, Cerebellum and Blood Between Individuals

Between-individual variation in DNA methylation is correlated between blood and brain





Davies, Volta et al. In Review

Tuesday, March 6th 2012

Session 4

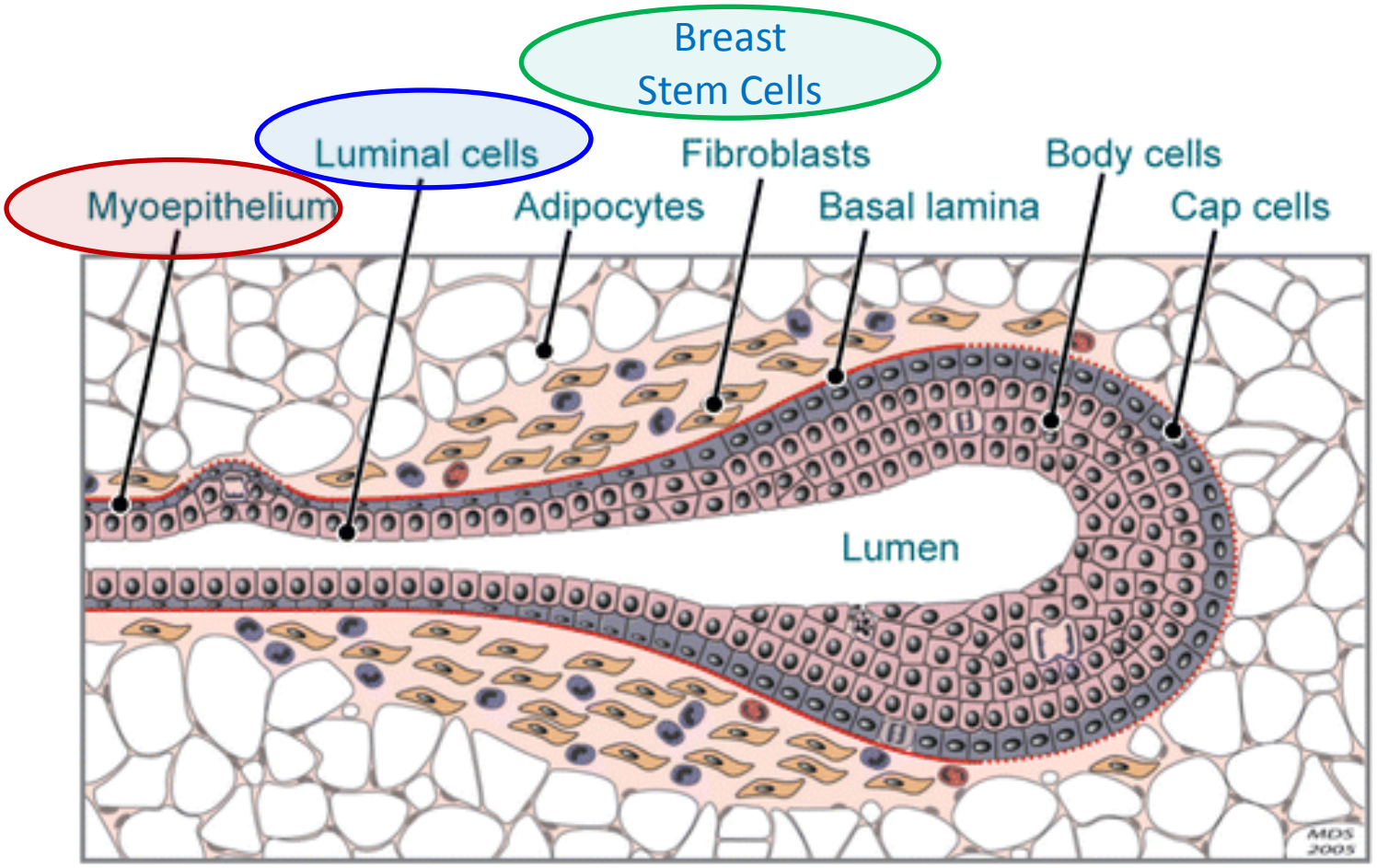
2:00 – 3:30 pm *Case Studies: Epigenomic Variation Between Individuals* – Matt Roth, Kevin Riehle, Chia-Chin Wu, Yuan Yuan, Cristi Coarfa, Alan Harris, Aleks Milosavljevic

3:30 – 4:30 pm ***Case Studies 12, 13:*** *Epigenomic Variation in Cancer* – Matt Roth, Kevin Riehle, Chia-Chin Wu, Yuan Yuan, Cristi Coarfa, Alan Harris, Aleks Milosavljevic

4:30 – 5:00 pm Open discussion and wrap-up

5:00 pm **Adjourn**

Use Case 12: Determining Cell Type of Origin of Breast Cancer



Hebner C, et al. 2008.
Annu. Rev. Pathol. Mech. Dis. 3:313–39

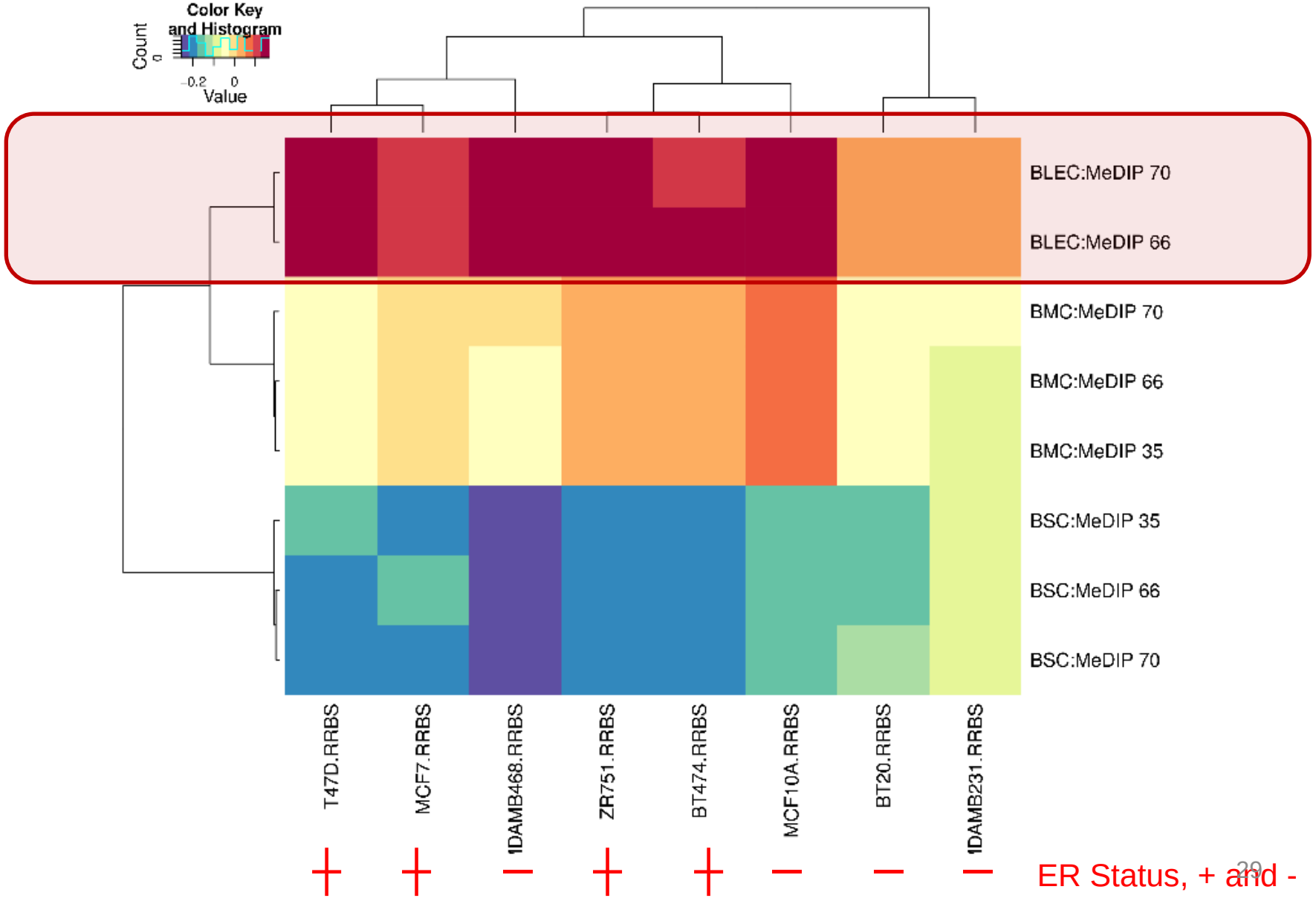
Use Case 12: Determining Cell Type of Origin of Breast Cancer

Compare profiles of 4 ER+ and 4 ER- breast cancer cell lines (RRBS profiles)[1] against the reference epigenomes of normal breast cell types (MeDIP-seq profiles) from the Human Epigenome Atlas.

Data Source:

1 Sun, Z., et al. (2011) Integrated Analysis of Gene Expression, CpG Island Methylation, and Gene Copy Number in Breast Cancer Cells by Deep Sequencing. PLoS ONE 6(2): e17490.

Use Case 12: Determining Cell Type of Origin of Breast Cancer

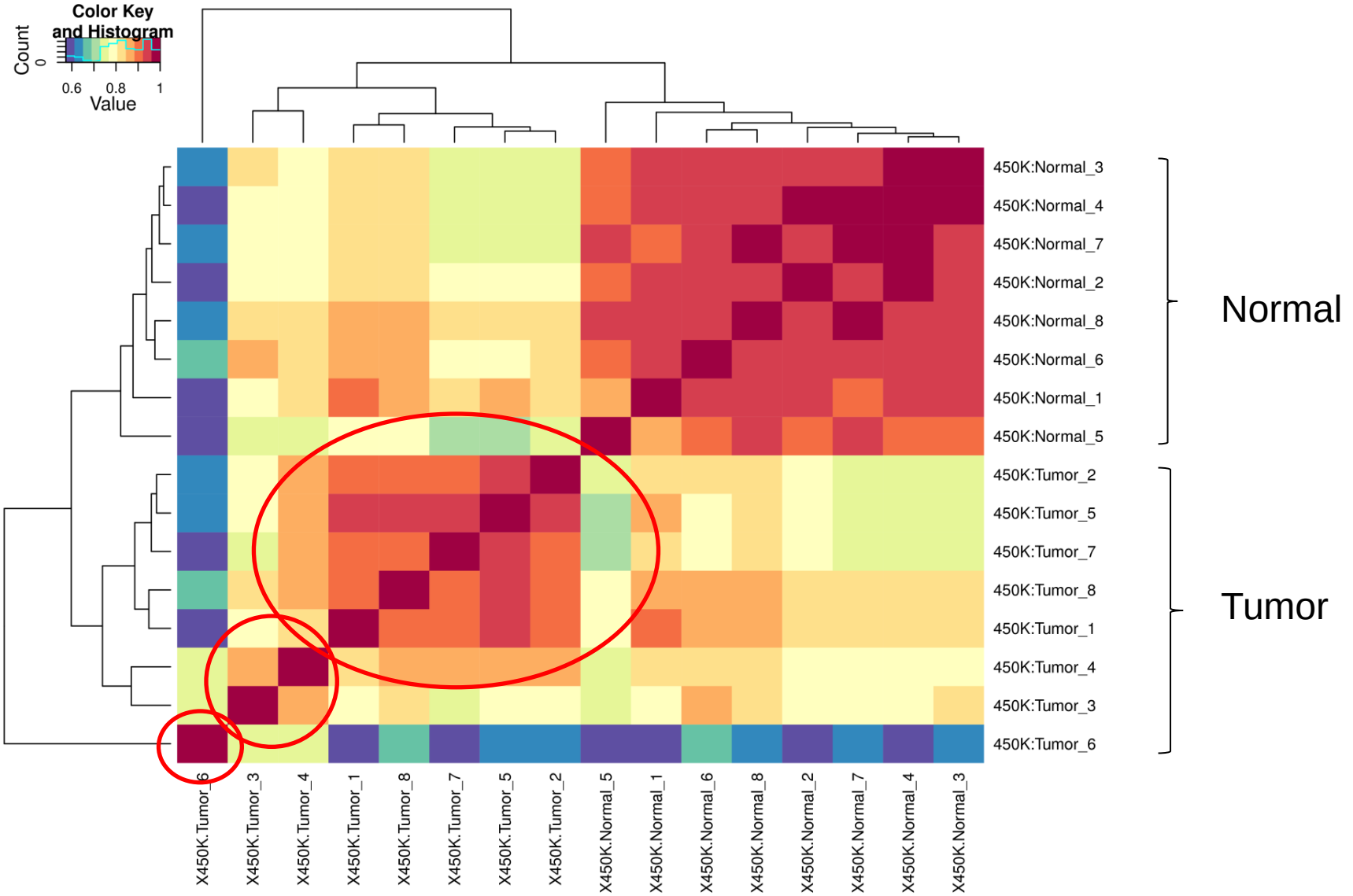


ER Status, + and -

Use Case 13: Analysis of Epigenomic Variation in Breast Tumors

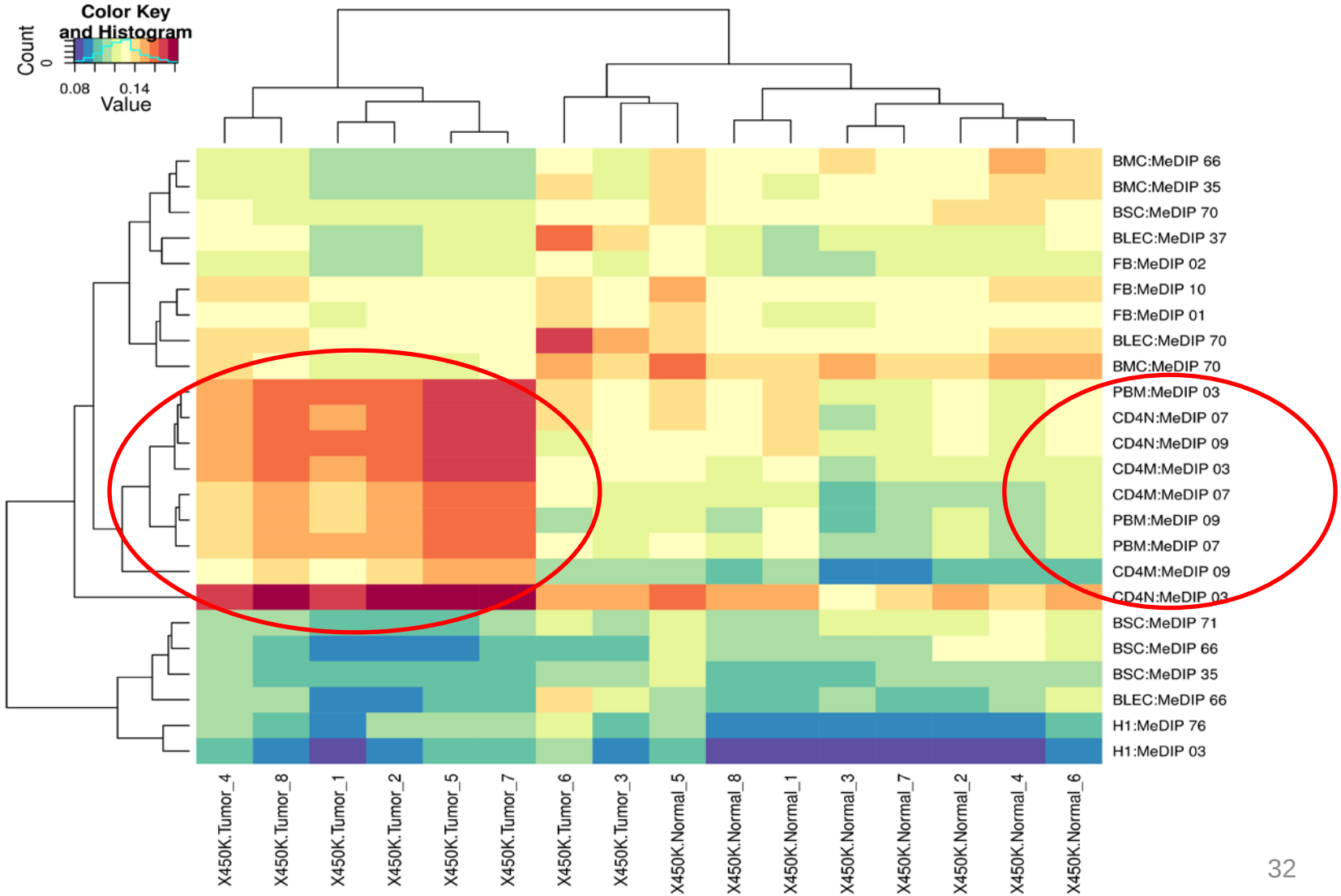
- Data Source: Dedeurwaerder, S.et al. (2011) "Evaluation of the Infinium Methylation 450K technology", Epigenomics 3(6):771-84.
- **16 breast tissue samples were profiled** (8 normal, 8 primary tumor samples)
- The paper evaluates 450K technology and does not report any analysis concerning cancer biology.
- We analyze the data to explore epigenomic states and cell type composition of normal and tumor samples.

Use Case 13: Analysis of Epigenomic Variation in Breast Tumors



Use Case 13: Analysis of Epigenomic Variation in Breast Tumors

Most tumor samples appear to contain more blood and immune cells than normal tissue.



Use Case 13: Analysis of Epigenomic Variation in Breast Tumors

- **Most breast tumor samples appear to contain an excess of blood and immune cells.**
- Comparison of normal and tumor tissue should therefore reveal **differentially methylated genes that are involved in immunity-related pathways or biological processes.**

GOTERM_BP_FAT	neurological system process	RT	166	12.9	3.8E-18	1.2E-14
GOTERM_BP_FAT	sensory perception of smell	RT	76	5.9	1.1E-13	1.8E-10
GOTERM_BP_FAT	sensory perception	RT	115	9.0	2.1E-13	2.2E-10
GOTERM_BP_FAT	G-protein coupled receptor protein signaling pathway	RT	144	11.2	4.5E-13	3.5E-10
GOTERM_BP_FAT	cognition	RT	123	9.6	8.6E-13	5.4E-10
GOTERM_BP_FAT	sensory perception of chemical stimulus	RT	79	6.2	1.1E-12	6.0E-10
GOTERM_BP_FAT	cell surface receptor linked signal transduction	RT	207	16.1	1.6E-12	7.3E-10
GOTERM_BP_FAT	defense response	RT	92	7.2	4.4E-12	1.7E-9
GOTERM_BP_FAT	cell-cell signaling	RT	90	7.0	6.7E-12	2.3E-9
GOTERM_BP_FAT	cell adhesion	RT	98	7.6	3.9E-11	1.2E-8
GOTERM_BP_FAT	biological adhesion	RT	98	7.6	4.2E-11	1.2E-8
GOTERM_BP_FAT	immune response	RT	96	7.5	9.1E-11	2.4E-8
GOTERM_BP_FAT	homophilic cell adhesion	RT	30	2.3	2.1E-8	5.1E-6
GOTERM_BP_FAT	cell-cell adhesion	RT	46	3.6	8.0E-8	1.8E-5
GOTERM_BP_FAT	feeding behavior	RT	20	1.6	1.5E-7	3.2E-5
GOTERM_BP_FAT	behavior	RT	65	5.1	1.7E-7	3.4E-5
GOTERM_BP_FAT	synaptic transmission	RT	47	3.7	3.0E-7	5.5E-5
GOTERM_BP_FAT	transmission of nerve impulse	RT	52	4.1	4.3E-7	7.5E-5
GOTERM_BP_FAT	cell activation	RT	45	3.5	6.6E-7	1.1E-4
GOTERM_BP_FAT	positive regulation of immune system process	RT	38	3.0	3.6E-6	5.7E-4
GOTERM_BP_FAT	inflammatory response	RT	46	3.6	8.2E-6	1.2E-3

Tuesday, March 6th 2012

Session 4

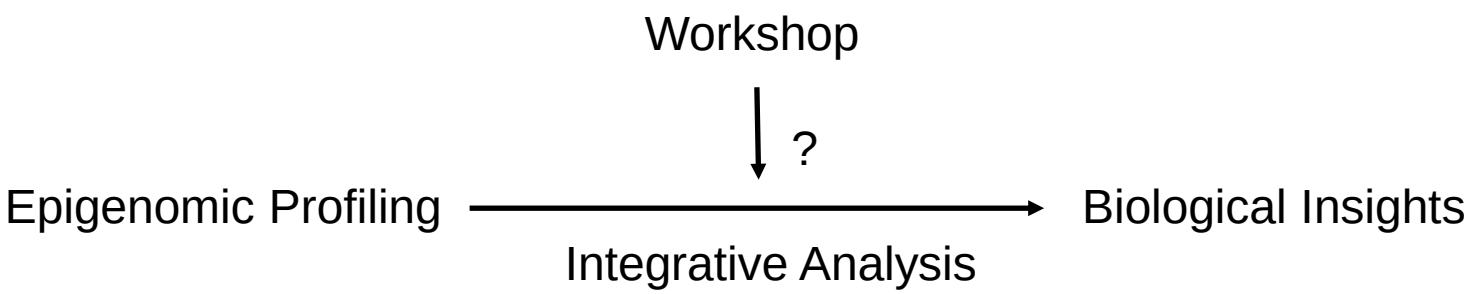
2:00 – 3:30 pm *Case Studies: Epigenomic Variation Between Individuals* – Matt Roth, Kevin Riehle, Chia-Chin Wu, Yuan Yuan, Cristi Coarfa, Alan Harris, Aleks Milosavljevic

3:30 – 4:30 pm *Case Studies: Epigenomic Variation in Cancer* – Matt Roth, Kevin Riehle, Chia-Chin Wu, Yuan Yuan, Cristi Coarfa, Alan Harris, Aleks Milosavljevic

4:30 – 5:00 pm Open discussion and wrap-up

5:00 pm **Adjourn**

Workshop Evaluation



- Methods (Assays, Data Processing)
- Standards (Metadata, Interoperability)
- Data Resources (Human Epigenome Atlas)
- Tools (Epigenomic Toolset, Genboree Workbench)
- Use Cases / Case Studies
- Collaborative Opportunities / Networking / Exchange of Experience

NIH Roadmap Epigenomics: Reference Epigenomes and the Human Epigenome Atlas

NIH Roadmap
Epigenomics Data Analysis and Coordination Center

March 5-6 2012
Houston, Texas

NIH Roadmap Epigenomics Project

Hypothesis:

Origins of health and susceptibility to disease are, in part, result of epigenetic regulation

Goal:

Transform biomedical research by

- **Developing comprehensive reference epigenome maps**
- **Developing new technologies for epigenomic analyses**

Including **cyberinfrastructure** for epigenomic research

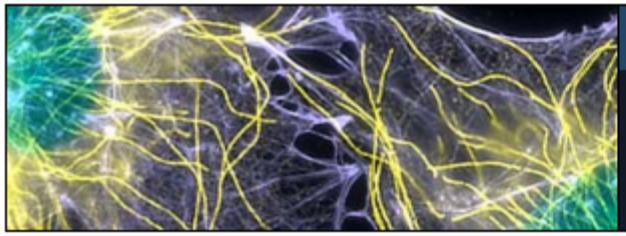
NIH Roadmap Epigenomics Project: Data Flow

Human Epigenome Atlas

Quarterly cumulative releases of the Human Epigenome Atlas

(Human Epigenome Atlas Release 6 being completed)

Epigenomics Portal at NCBI



Epigenomics

Genomics maps of stable, yet reprogrammable nuclear changes that control gene expression and influence our health.

How to...

- How to Use the Sample Browser
- How to Manage Collections of Samples
- How to View Genome Tracks
- How to Download Genome Tracks

Epigenomics Tools

- Browse Experiments
- Browse Samples
- Compare Samples Beta
- Advanced Search

Scientific Background

- About Epigenetics
- About DNA Methylation
- About Histone Modifications
- About Chromatin Structure

Latest Studies

CTCF Binding Sites by ChIP-seq from ENCODE/University of Washington [ESS000141]

Genome-wide remodeling of the epigenetic landscape during myogenic differentiation [ESS000133]

Histone Modifications by ChIP-seq from ENCODE/Broad Institute [ESS000130]

DNaseI hypersensitivity by digital DNaseI from ENCODE/University of Washington [ESS000129]

The Cohesin Complex Cooperates with Pluripotency Transcription Factors in the Maintenance of Embr [ESS000137]

See more...

NIH Roadmap Epigenomics

About the Project
Reference epigenome maps and their applications to human health.

- Roadmap Epigenomics Web Resources
- Roadmap Data in GEO
- Data Access Policies

Recent Review Articles

Epigenetics in disease and cancer. [Malays J Pathol. 2011]

Role of TGF-β and the tumor microenvironment during mammary tumorigenesis. [Gene Expr. 2011]

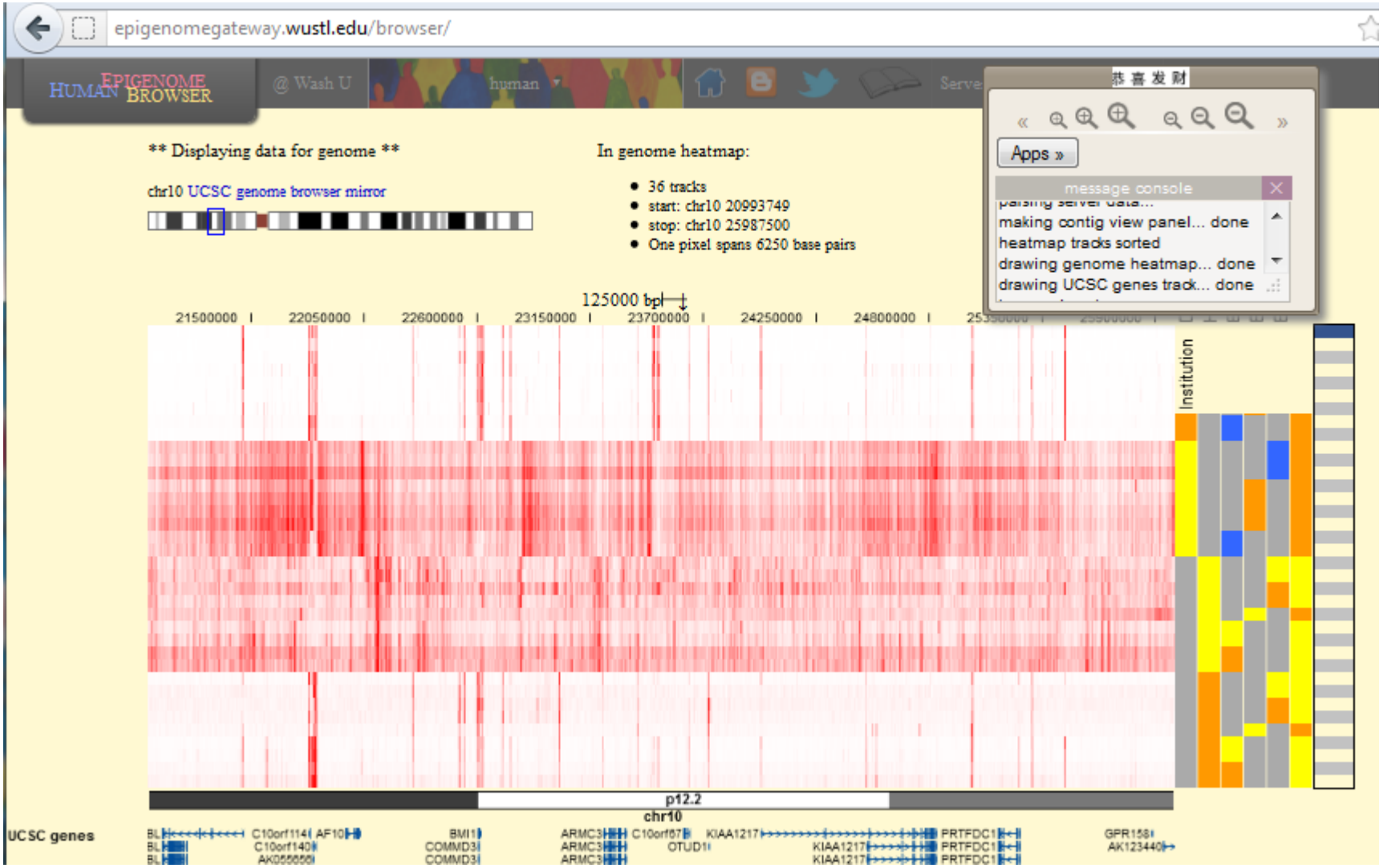
[Epigenetic regulation of transcription in heart development]. [Seikagaku. 2011]

[Neurobiology of suicidal behaviour]. [Psychiatr Pol. 2011]

Histone methylation makes its mark on longevity. [Trends Cell Biol. 2012]

See more...

Human Epigenome Browser at Wash U



Epigenomic Data

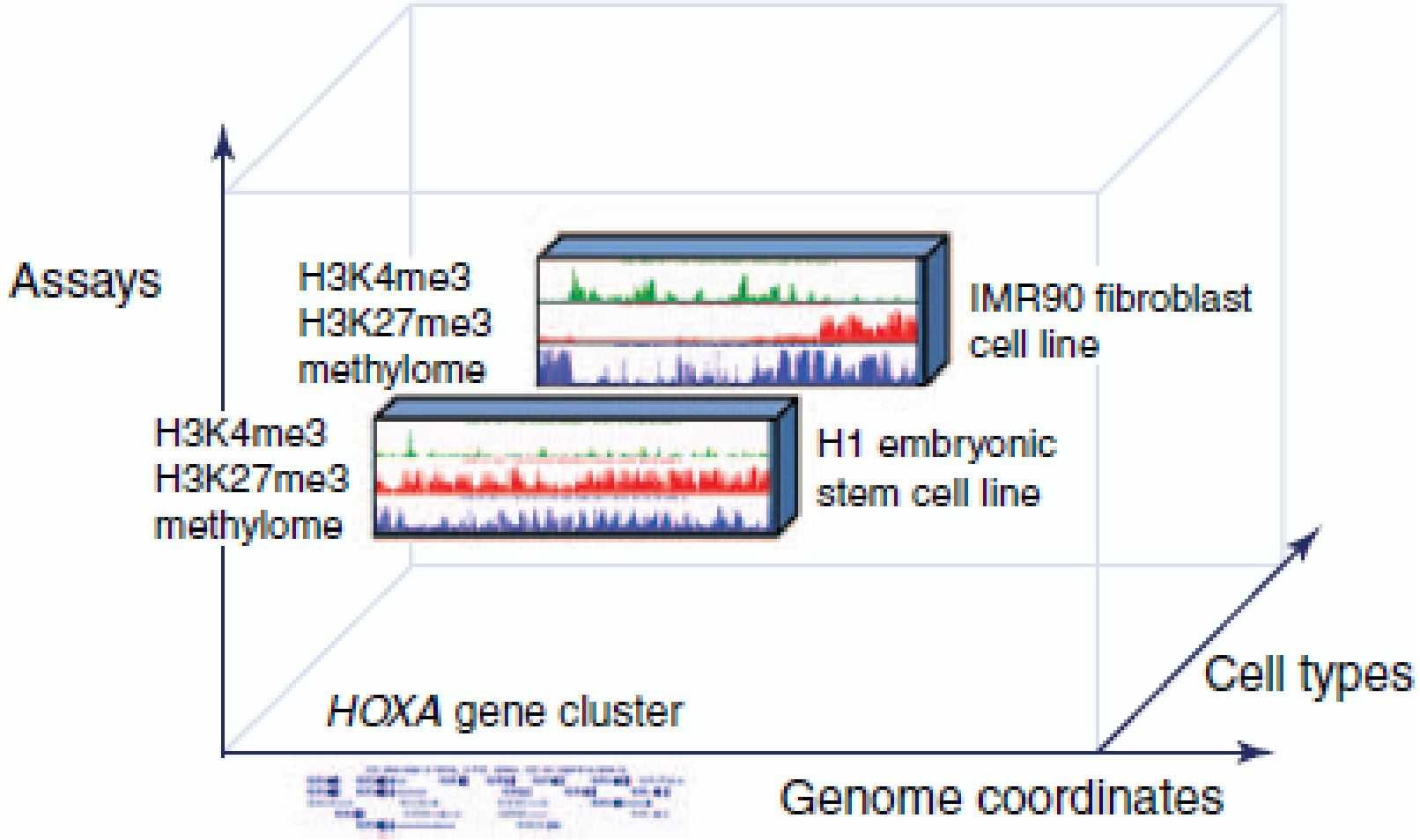
	“chip” data	**“seq” data**
Level 0	image	reads
Level 1	extracted features	mapped reads
Level 2	normalized intensities	read density maps
Level 3	epigenomic state (per feature such as enhancer or genomic segments outside of features)	
Level 4	comparative analysis results (e.g., cell type – specific state)	

Epigenomic Data Standards
www.ihec-epigenomes.org

Human Epigenome Atlas

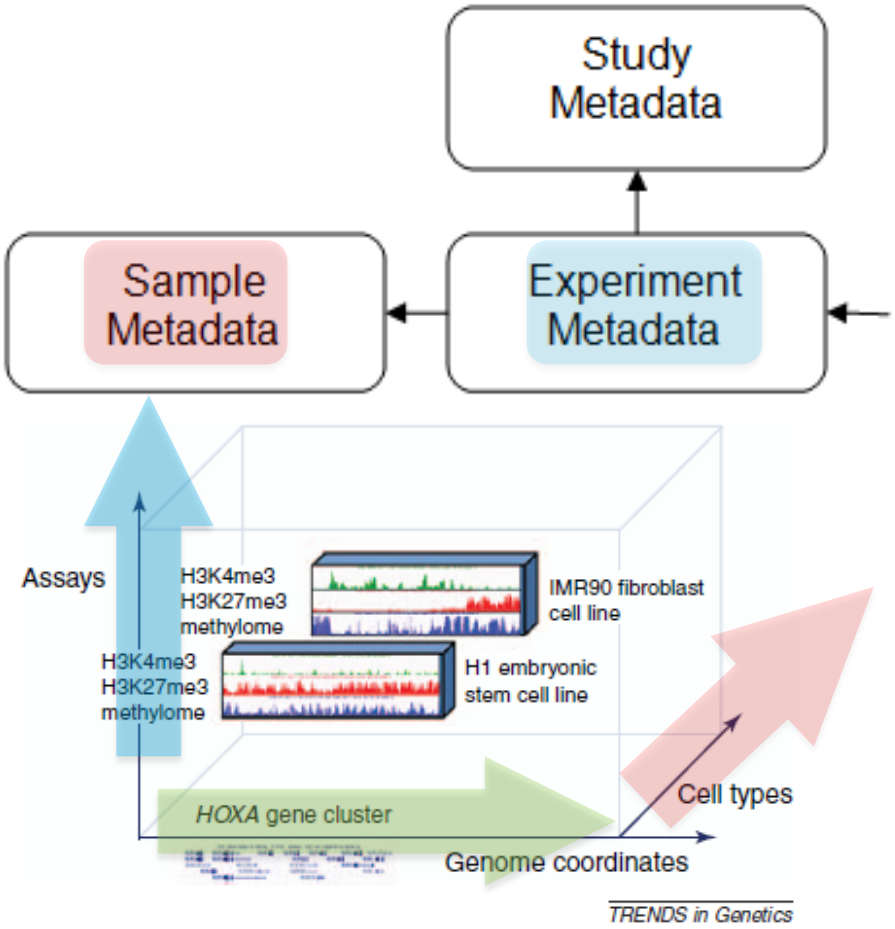
ROADMAP epigenomics PROJECT

www.epigenomeatlas.org



TRENDS in Genetics

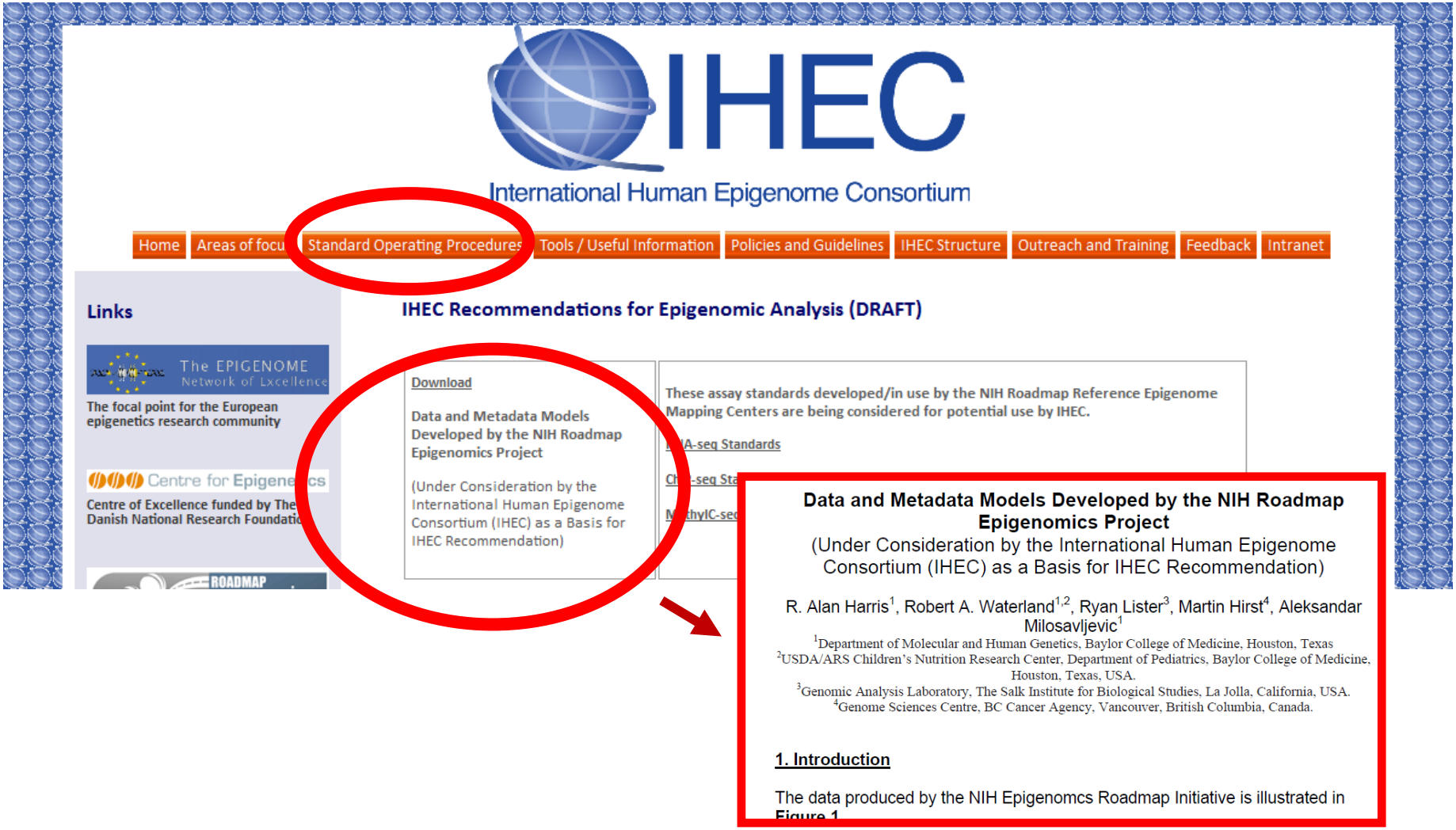
Epigenomic Metadata



Level	Data	Metadata
0	Fastq / SRF	Run

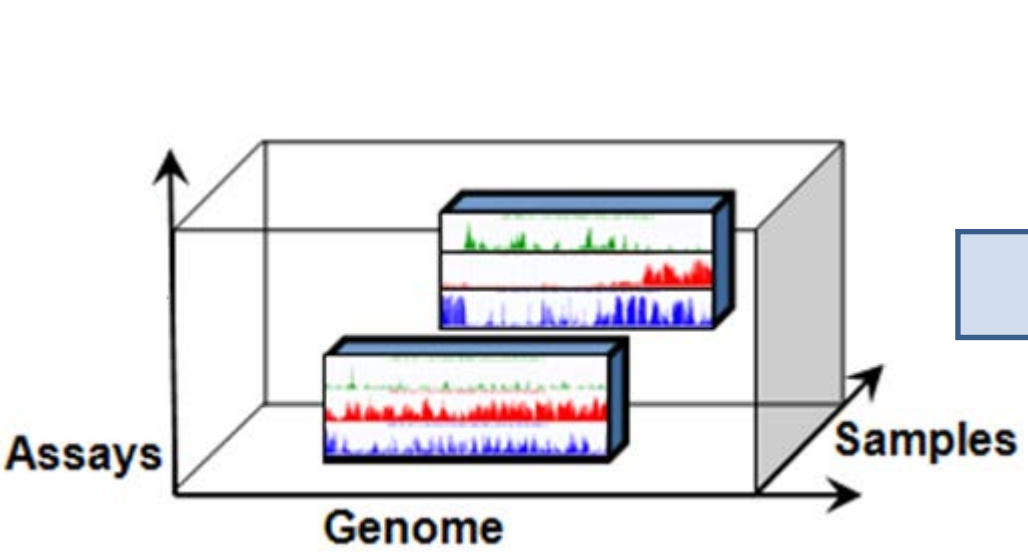
Level	Analysis	Metadata
1	BED / BAM	Software Parameters QC
2	Wig / Counts	Software Parameters QC
3	Peaks / GFF3 / LFF	Software Parameters QC
4	Comparative Analysis / GFF3 / LFF	Software Parameters QC

Epigenomic MetaData Standards
www.ihec-epigenomes.org



Epigenome Atlas Grid

www.epigenomeatlas.org



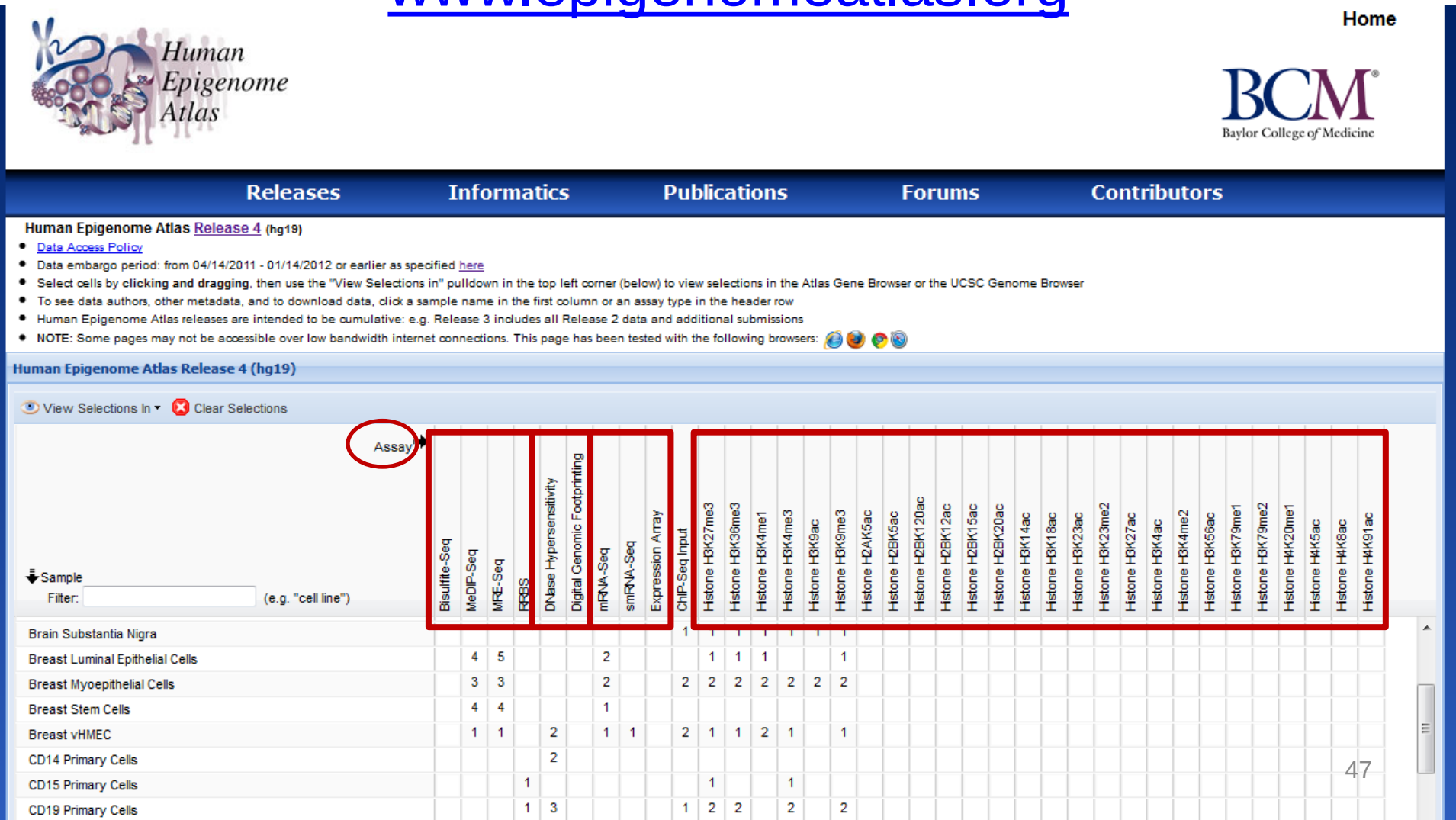
Select tracks in the Assay/Sample grid

Select genes or features

Pathway browsing
Genome browsing

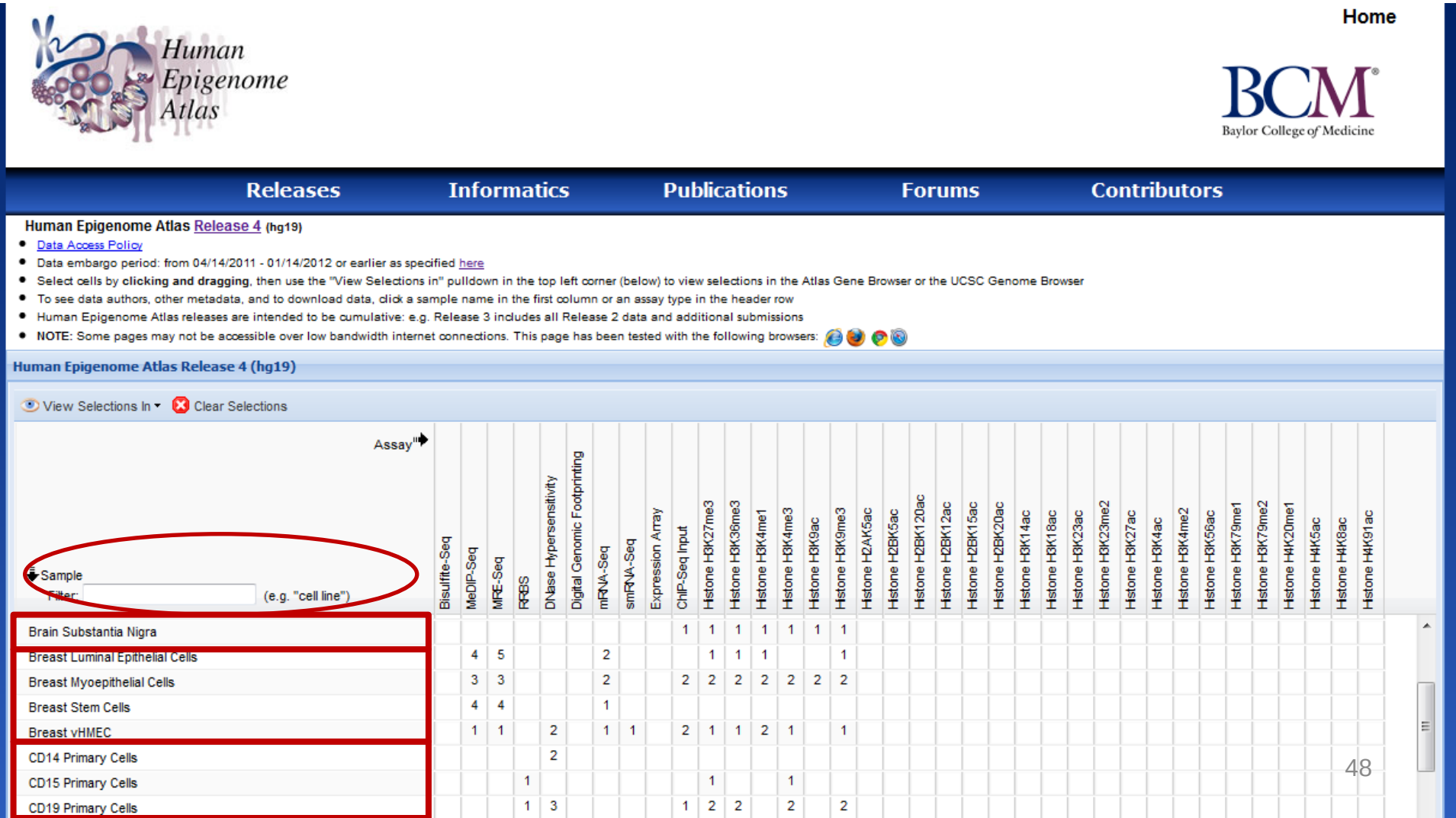
Epigenome Atlas Release 5 over 1500 experiments

www.epigenomeatlas.org



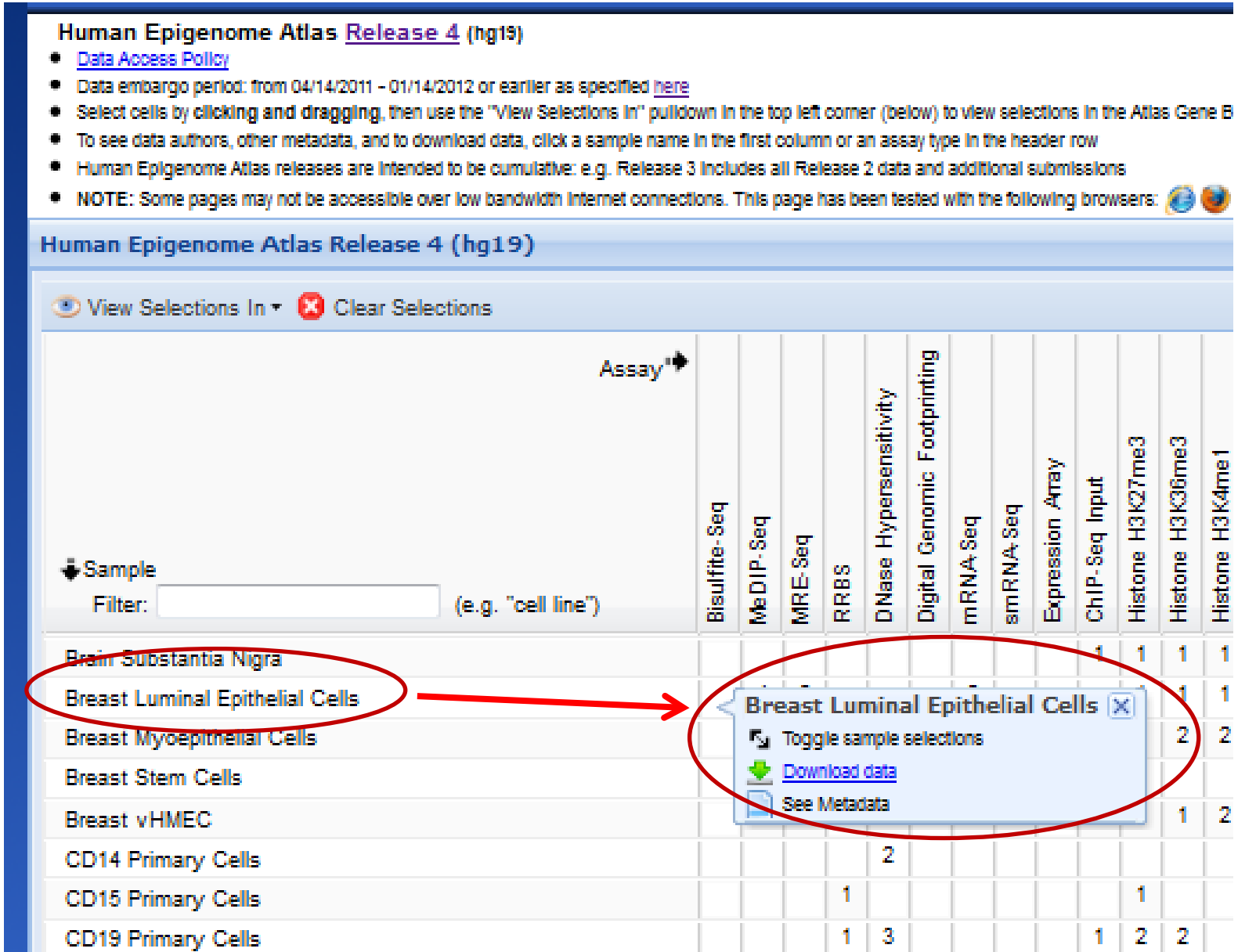
Epigenome Atlas Release 5

www.epigenomeatlas.org



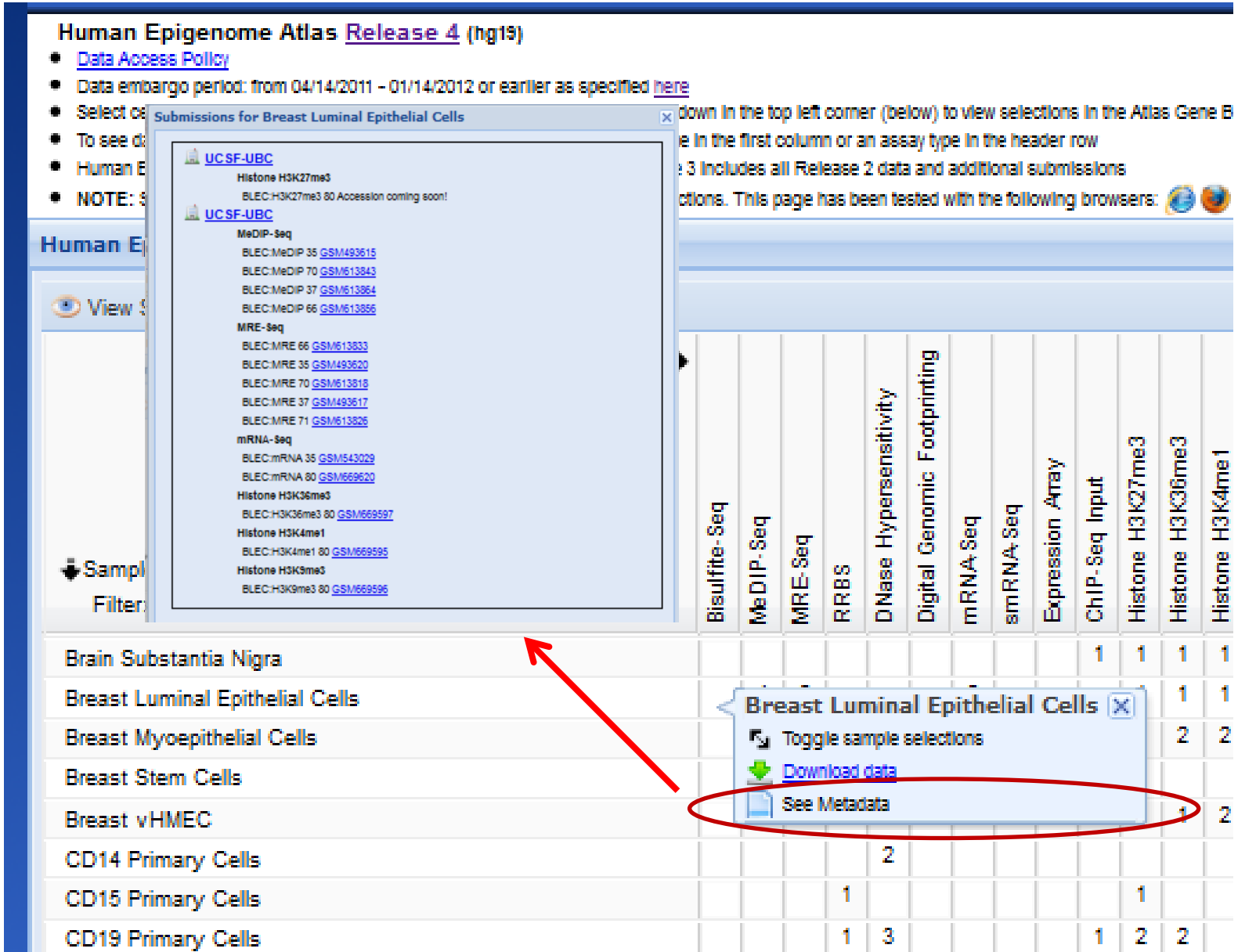
Epigenome Atlas Release 5

www.epigenomeatlas.org



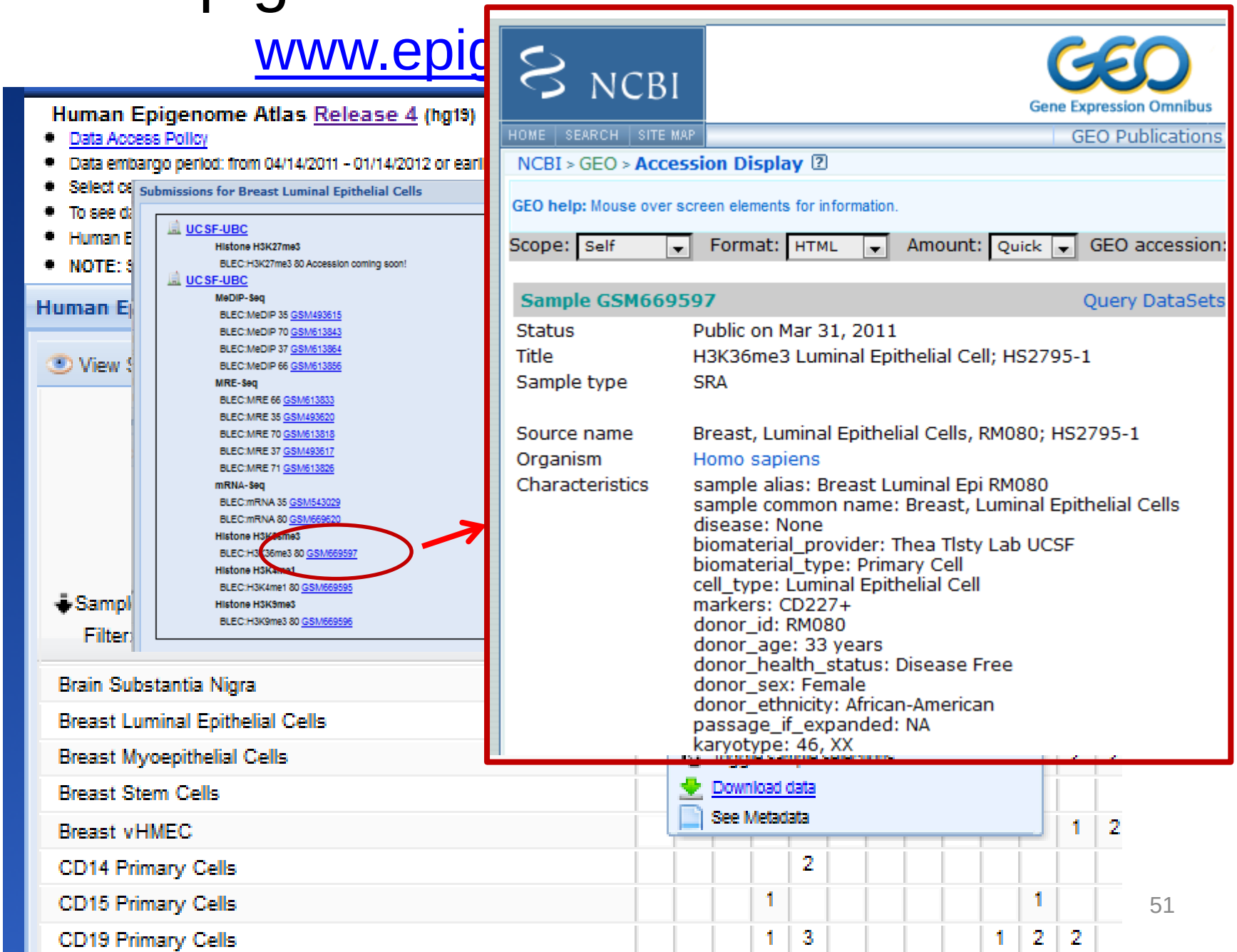
Epigenome Atlas Release 5

www.epigenomeatlas.org

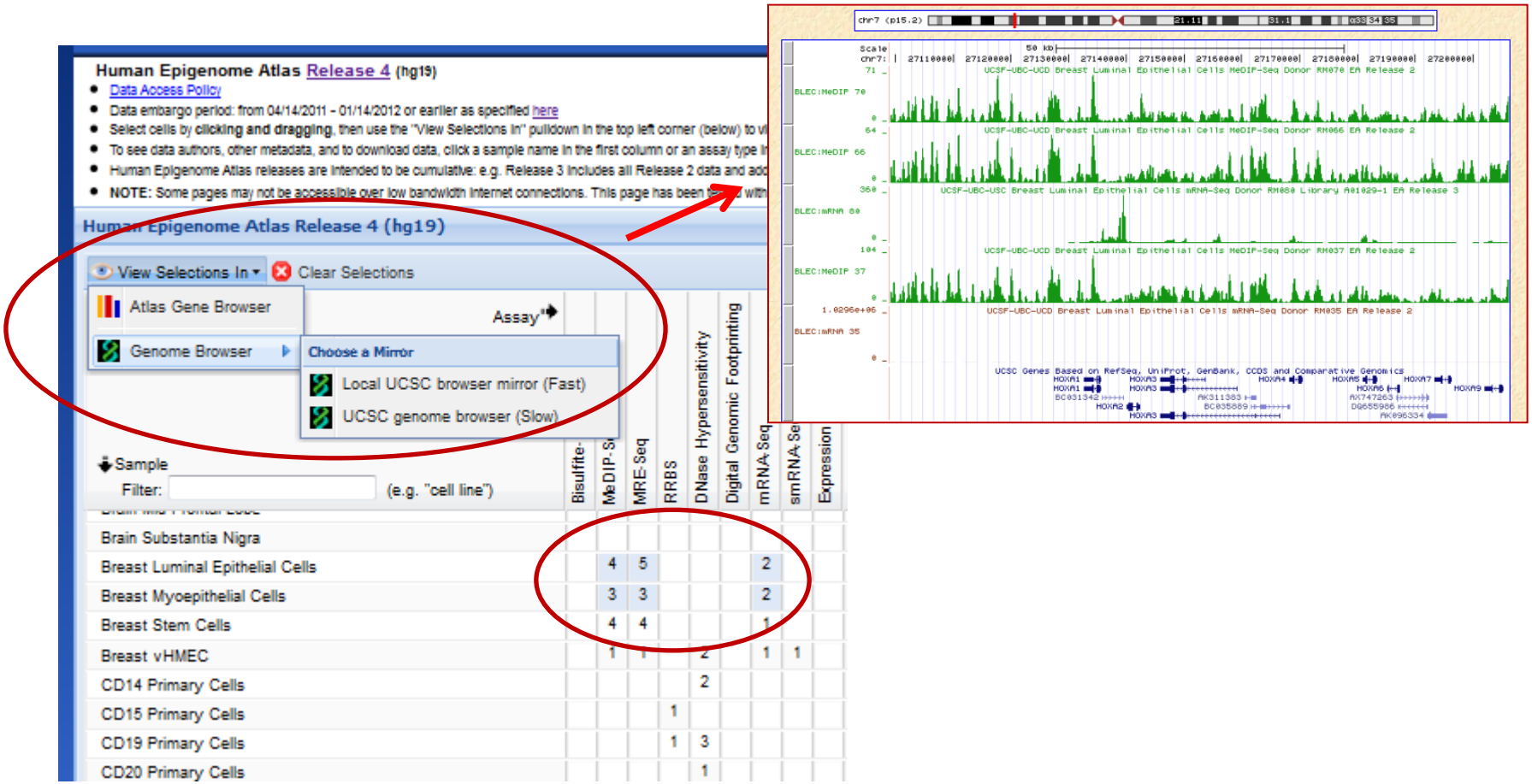


Epigenome Atlas Release 5

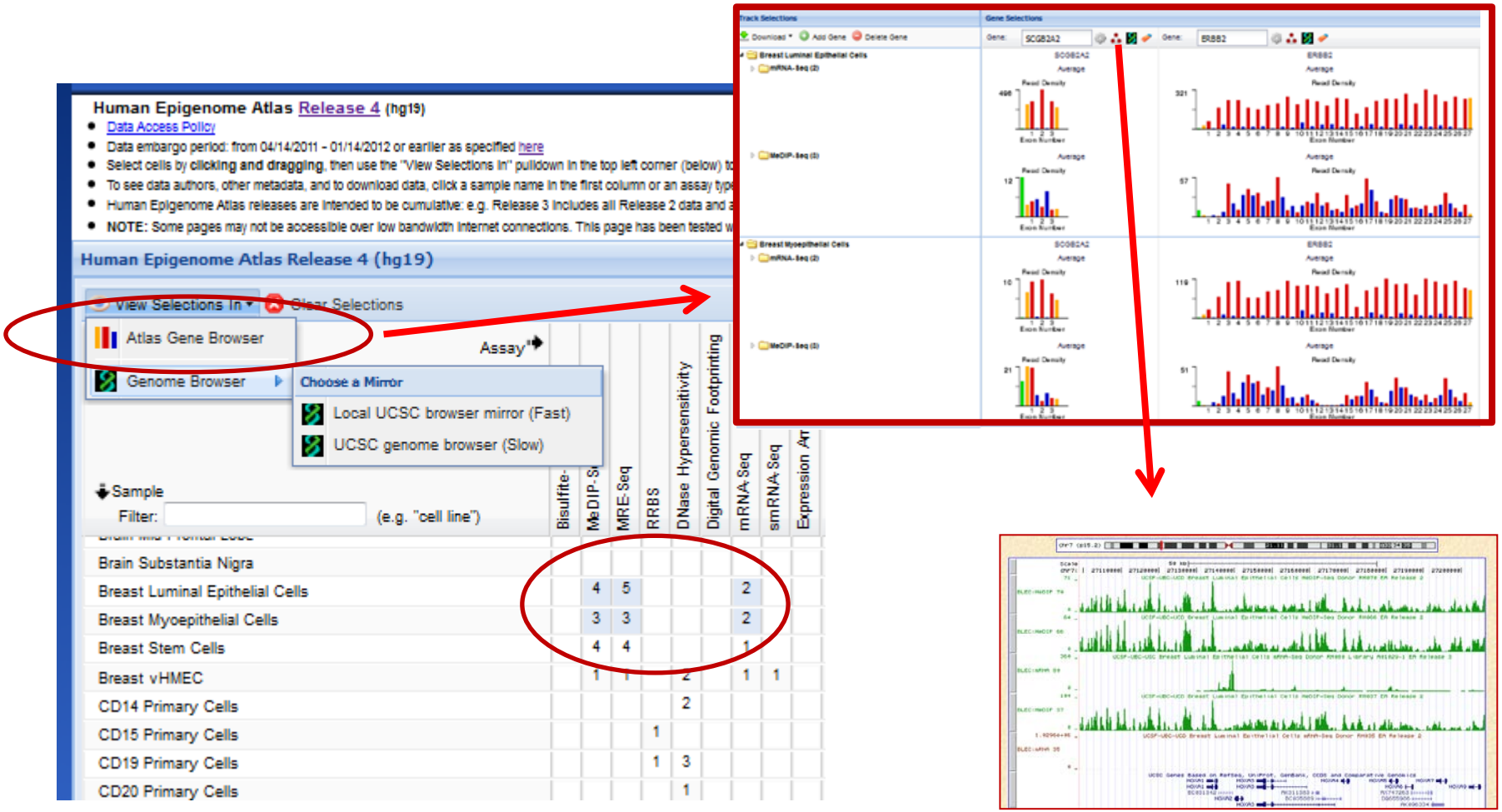
www.epig



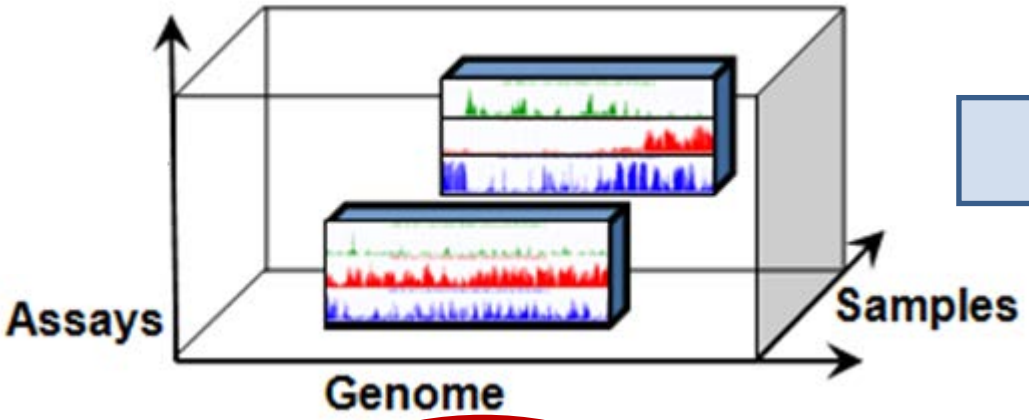
Viewing selections



Viewing selections



Epigenome Atlas Grid and the Atlas Gene Browser



Select tracks in the Assay/Sample grid

Pathway browsing

Genome browsing

Select genes or features

Atlas Gene Browser

Download scores into spreadsheet

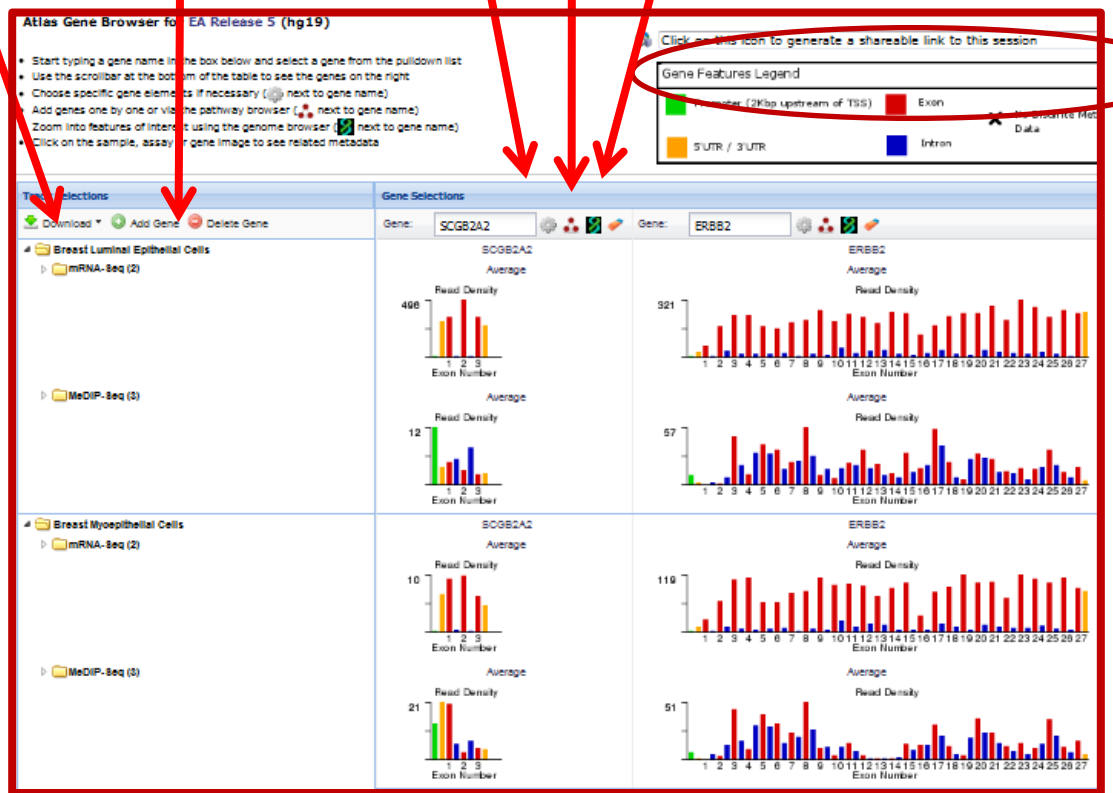
Configure gene elements

Add/delete genes

Select genes from shared KEGG pathways

Share view (URL)

View in UCSC genome browser



Including More Genes in the Same Pathway