THANK YOU TO OUR SPONSORS

CEGx

diagenode

illumina

KAPA BIOSYSTEMS
Arranged By

John Greally
Albert Einstein College of Medicine

Will Liao
New York Genome Center

Nathaniel Pearson
New York Genome Center

Zac Chatterton
Mount Sinai School of Medicine

Aslihan Dincer
Mount Sinai School of Medicine

Masako Suzuki
Albert Einstein College of Medicine

Ari Melnick
Weill-Cornell Medical College

Doron Betel
Weill-Cornell Medical College

Benjamin Tycko
Columbia University

Wireless Information

Name: NYGC-Guest
Password: welcome101

Social Media

Twitter/Facebook: #NYEpi15
8:00AM–9:00AM  
**REGISTRATION/BREAKFAST**

9:00AM–9:15AM  
**OPENING REMARKS**  
Robert Darnell, MD, PhD  
New York Genome Center/The Rockefeller University

9:15AM–11:00AM  
**EPGENOMICS IN CANCERS**  
Session Chair  
Nathan Pearson, PhD  
New York Genome Center  
Lessons Learned from the AML Epigenome  
Ari M. Melnick, MD  
Weill Cornell Medical College  
Targeting Epigenetic Mechanisms for Leukemia Therapy  
Scott A. Armstrong, MD, PhD  
Memorial Sloan Kettering Cancer Center  
Histone Variation in Cancer  
Emily Bernstein, PhD  
Icahn School of Medicine at Mount Sinai

11:15AM–1:00PM  
**EPGENOMICS IN CHILDHOOD DISEASES**  
Session Chair  
John Greally, MB, PhD  
Albert Einstein School of Medicine  
Sex Differences in Cytosine Methylation Associated with Abnormal Fetal Growth  
Francine H. Einstein, MD  
Albert Einstein College of Medicine  
Neuronal Epigenome Mapping in Developing and Diseased Cerebral Cortex  
Schahram Akbarian, MD, PhD  
Icahn School of Medicine at Mount Sinai  
Developmental Origins of Postnatal Lung Disease  
Wellington V. Cardoso, MD, PhD  
Columbia University Medical Center

11:00AM–11:15AM  
**COFFEE & NETWORKING**

1:00PM–2:00PM  
**LUNCHEON PRESENTATION**  
Off the Beaten Path: Newer Applications of NGS Technology  
Michael Smith, PhD  
Illumina

2:00PM–3:00PM  
**POSTER SESSION**

3:00PM–3:45PM  
**KEYNOTE**  
Epigenetic Drivers and Dependencies of Heterogeneous Tumor Cell States  
Bradley E. Bernstein, MD, PhD  
Massachusetts General Hospital

3:45PM–5:15PM  
**EMERGING EPGENOMIC TECHNOLOGIES**  
Session Chair  
Rahul Satija, PhD  
New York Genome Center/NYU  
Statistical Challenges in Epigenomics: Detecting Differentially Methylated Regions in the Presence of Unwanted Variability  
Rafael Irizarry, PhD  
Harvard University  
How Genomes Fold: Now Inside the Loop  
Erez S. Lieberman-Aiden, PhD  
Baylor College of Medicine  
ATACing chromatin structure: From Ensembles to Single-Cells, a window into cis- and trans-variation  
Jason Buenrostro  
Stanford University
Neuronal Epigenome Mapping in Developing and Diseased Cerebral Cortex

Schahram Akbarian
Icahn School of Medicine at Mount Sinai

The overwhelming majority of the neuronal constituents in the human cerebral cortex permanently exit from the cell cycle in early midgestation, but little is known about epigenetic regulation of neuronal and glial epigenomes during pre- and postnatal development and maturation. Here, I will discuss innovative neuroepigenetic approaches, including cell-type specific histone methylation profiling and chromosome conformation capture technology applicable to postmortem brain tissue. I will discuss initial insights on developmental regulation and disease-associated alterations of neuronal epigenomes from human and mouse prefrontal cortex. Cell-type specific epigenome, transcriptome and chromosomal conformation mappings in human brain, in conjunction with the emerging genetic risk architectures of psychiatric disease, are likely to illuminate the role of regulatory non-coding sequences in the neurobiology of autism and psychosis spectrum disorders.

Evaluation and Comparison of Three Base Pair Resolution CpG Methylation Enrichment Techniques

Alicia Alonso1, Thadeous Kackmarkzyk1, Yuan Xin1, Xihui Zhang1, Yushan Li1, Francine Garrett-Bakelman1, Doron Betel1,2

1 Epigenomics Core Facility, Department of Medicine, Division of Hematology/Oncology, Weill Cornell Medical College
2 Institute for Computational Biomedicine, Weill Cornell Medical College

The mammalian methylome, roughly 28M CpGs, provides important epigenetic marks required for normal development and involved in disease progression. Due to the prohibitive cost of whole genome bisulfite sequencing, we compared three current methylome enrichment techniques that sample approximately 10–20% of CpG sites at functionally annotated regions: a) Digestion-based enhanced reduced representation (ERRBS) covering 6.6M CpGs, b) Capture hybridization using stranded RNA baits (Agilent Sure Select) covering 3.7M CpGs and, c) Capture hybridization using strand-specific, methylation-specific probes (Roche CpGiant) covering 5.6M CpGs. We evaluated a) overlap of genomic regions b) concordance of CpG sites and coverage, and c) genomic annotation. Libraries, from human DNA representing different epigenetic backgrounds, were prepared using each of the protocols and sequenced to >10x depth. Additionally, we tested libraries derived from Formalin Fixed Paraffin Embedded DNAs, the prevalent material in clinical studies. Overall, SureSelect and CpGiant recovered >95% of their designed capture regions whereas ERRBS covered 70% of its expected fragments.

Concordance of annotations of CpG sites for genomic features (e.g. CpG islands and shores) and annotation of gene regions (e.g. promoter, exon, and intron) displayed roughly the same proportions of genomic features, where SureSelect and CpGiant are most similar, and cover marginally more annotated regions than ERRBS. Although the number of CpG sites shared by all methods was low (10–20%), they do provide independent validation across the different platforms. We conclude that captured based methods are largely consistent in terms of covered CpG loci, although ERRBS provides comparable data at a significantly reduced price. Furthermore, library preparation for ERRBS can be performed with 75ngs of starting material, whereas micrograms are needed for the capture hybridization techniques.

Epigenetic Drivers and Dependencies of Heterogeneous Tumor Cell States

Bradley E. Bernstein
Department of Pathology, Massachusetts General Hospital and Harvard Medical School, Broad Institute of Harvard and MIT, Howard Hughes Medical Institute

I will present an overview of our epigenomics research and ongoing technology development. I will then focus in on our recent studies of the malignant brain tumor
glioblastoma, which exemplify our efforts to characterize functional and epigenetic heterogeneity in cancer. Glioblastomas contain a primitive population of stem-like cells with high tumor-propagating potential. These stem-like cells are thought to underlie aggressive tumor growth and therapeutic resistance. We have extensively characterized patient-derived stem-like gliomaspheres that potently initiate tumors upon xenotransplantation. We identified a set of neurodevelopmental transcription factors and associated regulatory programs that drive tumorigenicity in this model. We also surveyed intratumoral heterogeneity in primary glioblastomas using single-cell RNA-seq. Cells within a single tumor vary extensively in their expression of regulatory programs related to cell cycle, hypoxia, stemness and established glioblastoma subtype classifiers. The single-cell RNA-seq data for glioblastoma tumors also reveal an in vivo population of stem-like cells that largely, but not perfectly, emulate the in vitro model. I will discuss these studies as well as ongoing efforts to establish more faithful glioblastoma models and to use these in vitro systems to better understand glioblastoma biology and advance therapeutic strategies.

Histone Variation in Cancer

Emily Bernstein
Icahn School of Medicine at Mount Sinai

Cutaneous melanoma is a highly aggressive skin cancer and one of the most challenging cancers in its therapeutic management. Emerging studies clearly demonstrate that cancer is a result of the concerted action of genetic and epigenetic alterations. Surprisingly, our understanding of the ‘epigenetic landscape’ of melanoma remains poorly understood. Our recent work has shed light on the underlying mechanisms of chromatin dynamics in development and cancer. We have identified a critical role for histone variants of the H2A family in melanoma progression. We identified that the histone variant macroH2A acts as a barrier to melanoma growth and metastasis by suppressing key oncogenic factors. Here I will discuss our recent efforts to further understand the role of histone variants in melanoma pathogenesis and provide insight into the epigenetic landscape of malignant melanoma.

ATACing Chromatin Structure: From Ensembles to Single-Cells, a Window into Cis- and Trans- Variation

Jason Buenrostro, Beijing Wu, Ulrike Litzenburger, Mike Snyder, Howard Chang, William Greenleaf
Departments of Genetics, Stanford University

Genome-wide maps of chromatin accessibility represents a catalogue of regulatory elements defined by transcription factor binding and associated transcriptional machinery. Such comprehensive and unbiased profiling of the “regulome” holds tremendous power for understanding transcriptional regulation. While single-cell gene expression methods have become a major workhorse for understanding phenotypic variability in complex cellular populations, the underlying regulatory variation that causes this phenotypic heterogeneity remains unexplored due to an acute lack of methods that probe the regulome of single cells. We have developed a robust method for probing the accessible genome of rare cellular populations using assay for transposase accessible chromatin (ATAC-seq) and have now extended the method to profile epigenomes of individual cells using single-cell ATAC-seq (scATAC-seq). We find that aggregate scATAC-seq maps of ~200 single-cells closely recapitulates accessibility profiles measured via DNase-seq from tens of millions of cells. Furthermore, by quantifying the accessibility variance across single-cells, we find cell-type specific high-variance trans-factors. These high-variance factors are associated with different cellular programs, including cell-fate decisions and stress response. Indeed these methods for understanding single cell regulatory variation promise to provide the “why” behind the “what” of single-cell gene expression variance, adding an orthogonal and synergistic genome-wide single-cell information sub-type. Going forward, we believe scATAC-seq will enable the interrogation of the epigenomic landscape of small or rare biological samples allowing for detailed, and potentially de novo, reconstruction of cellular differentiation or disease at the fundamental unit of investigation—the single cell.
Developmental Origins of Postnatal Lung Disease
Wellington V. Cardoso
Columbia Center for Human Development, Department of Medicine, Columbia University Medical Center

There is accumulated evidence of an association between fetal exposure to an adverse intrauterine environment (xenobiotics, physical stress, nutritional imbalances) and increased susceptibility to postnatal diseases. Abnormal developmental programming resulting from these exposures has been proposed to contribute to the pathogenesis of several postnatal conditions. Epidemiological studies implicate maternal Vitamin-A (VA) deficiency in decreased lung function during postnatal life; VA deficiency has been also linked with increased prevalence and severity of asthma. The mechanistic basis for this association has been elusive. Here we present evidence from mouse genetic, in vitro and nutrition models that disruption of VA-retinoic acid (RA) signaling for a short developmental window when airways are forming leads to aberrant differentiation of smooth muscle (SM) with increased ectopic expression of SM markers. Notably, this phenotype persists in the adult lung when mice are similarly VA-deprived but subsequently are left to develop under normal diet until adulthood. Although apparently normal, these mice show asthma-like abnormal pulmonary function with increase in airway resistance and exaggerated bronchoconstriction in response to cholinergic stimulation, without evidence of inflammation. Thus, during development RA restricts the SM gene expression program to prevent excessive ectopic SM differentiation in forming airways. Disruption of this mechanism leads to a structural defect carried throughout life regardless the postnatal vitamin A status. Preliminary studies suggest an involvement of RA-depended epigenetic mechanisms in place during development influencing the airway SM program.

Multiple Mechanisms that Regulate MacroH2A1 Alternative Splicing are Perturbed in Cancer
Alyssa D. Casill and Matthew J. Gamble
Department of Molecular Pharmacology, Albert Einstein College of Medicine

Changes in alternative splicing patterns and in chromatin structure/function are hallmarks of cancer. MacroH2A1 (H2AFY) is a histone variant that undergoes significant changes in its alternative splicing patterns in tumors. Its two splice variants, macroH2A1.1 and macroH2A1.2, are products of a mutually exclusive alternative splicing event that leaves each isoform with a unique exon. MacroH2A1.1 can bind NAD+–derived ligands and therefore regulate processes such as transcription and cell proliferation while macroH2A1.2 cannot. It has been shown that macroH2A1.1 functions to suppress tumor cell proliferation and anchorage-independent growth and that macroH2A1.1 splicing is significantly reduced in many cancers. Computational analysis of data from over 7,000 samples across 25 types of cancer in The Cancer Genome Atlas (TCGA) shows that macroH2A1.1 splicing is reduced in uterine, lung, colon, rectal, breast, bladder, and prostate cancer. While macroH2A1.1 has been shown to function as a tumor suppressor in several contexts, the mechanisms regulating the macroH2A1 alternative splicing decision are still unknown. The overall objective of this study is to understand the regulation of macroH2A1 alternative splicing and its dysregulation in cancer.

To identify macroH2A1 splicing factors, we correlated the expression of putative splicing factors to macroH2A1.1 levels across samples in TCGA and identified several factors potentially involved in the alternative splicing decision. Conservation analysis identified intronic regions in H2AFY with binding sites for several of these factors. These factors will be further investigated for their role in macroH2A1 splicing using lentiviral-mediated knockdown as well as a splicing assay that utilizes an intronic-sequence-containing ectopic macroH2A1 expression vector to probe the cooperativity of cis- and trans-acting factors. Unexpectedly, TCGA analysis showed that while...
macroH2A1.1 splicing is reduced, the total level of macroH2A1 expression in several cancers is significantly increased. We have subsequently shown that changes in the rate of RNA Polymerase II (Pol II) elongation modulate the production of macroH2A1.1; decreasing the rate of Pol II elongation increases the splicing of macroH2A1.1. We hypothesized that the rate of Pol II elongation modulates the ability of splicing factors to bind the nascent RNA and therefore regulates macroH2A1 splicing. To evaluate the function of Pol II elongation rate in splicing factor binding, RNA-immunoprecipitation will be done in cells with different Pol II elongation rates (modulated by drugs such as DRB). We have also found that the expression of several transcription factors is correlated to both macroH2A1.1 splicing and macroH2A1 total expression, indicating that these factors are potentially involved in macroH2A1 transcription and macroH2A1.1 splicing (potentially by modulating Pol II elongation rates) in normal and cancer cells.

Investigating Non-Normality in the Cancer Transcriptome and Methylome
Laurence deTorrente

It is often assumed that gene expression follows a Normal distribution. The most common statistical methods that we employ for the analysis of gene expression data, such as the t test, linear regression and ANOVA models, are all predicated on this assumption. For methylation, data naturally follows a Bimodal distribution but due to the sampling of multiple cells, we often see various different distributions represented. In this project, we focused on identifying the prevalence of non-Normal distributions in the gene expression and methylation profiles of different cancer datasets from The Cancer Genome Atlas. We evaluated the number of genes/loci whose density could be explained by one of six statistical distributions: Normal, Lognormal, Bimodal, Pareto, Gamma and Cauchy. In order to assess the utility of acknowledging non-Normal distributions for modeling methylome data, we examined whether certain genomic features were associated with a particular density distribution. We also investigated the correlation between methylation and gene expression with respect to these genomic features, and density distributions. In brief, our findings suggest that:

- AML and Glioblastoma have loci with methylation profiles that follow a Normal and Bimodal distribution.
- There is not much difference between the distribution category when looking at methylated, intermediate or unmethylated loci.
- Non-Normal loci in the gene body are mostly methylated for AML and unmethylated for Glioblastoma.
- When looking at the genomic features, we have a greater number of positively-correlated (with expression) loci in the gene body and negatively-correlated (with expression) loci in the promoter region.
- Loci with methylation profiles that are Normally-distributed generally fall into distinct states of either low or high methylation as opposed to the non-Normally distributed ones that tend to have profiles that are more spread amongst the two states. Overall this work suggests that there is value, both in understanding the biology but also for ensuring the appropriateness of our models, in accounting for non-Normal distributions in gene expression and methylation profiling data.

Sex Differences in Cytosine Methylation Associated with Abnormal Fetal Growth
Francine Einstein
Albert Einstein School of Medicine

Extreme fetal growth is associated with increased susceptibility to a range of adult diseases. We tested whether heritable epigenetic processes in long-lived CD34+ hematopoietic stem/progenitor cells (HSPCs) showed evidence of a cellular memory associated with the extremes of fetal growth. We have found that both fetal growth restriction and overgrowth are associated with global shifts towards DNA hypermethylation, targeting cis-regulatory elements in proximity to genes involved in glucose homeostasis and stem cell function. A dimorphic response was also found based on infant sex, with growth restriction associated with substantially
greater epigenetic dysregulation in males but large for gestational age or overgrowth affecting females predominantly. The findings are consistent with extreme fetal growth interacting with variable fetal susceptibility to influence cellular aging and metabolic characteristics through epigenetic mechanisms, potentially generating biomarkers that could identify infants at higher risk for chronic disease later in life.

The Role of H3K27 in IFNγ-mediated Gene Expression

Eugenia G. Giannopoulou, Yu Qiao, Celeste Fang, Lionel B Ivashkiv
Arthritis and Tissue Degeneration Program and David C. Rosensweig Center for Genomics Research, Hospital for Special Surgery

Interferon gamma (IFNγ) is a major cytokine responsible for lymphocyte activation in enhanced anti-microbial and anti-tumor responses. It is the major pro-inflammatory cytokine to drive monocyte and macrophage classical differentiation (pro-inflammatory phenotype), which is critical for intracellular pathogen clearance. How IFNγ induces gene expression in macrophages has been well studied: IFNγ activates JAKSTAT1 signaling pathway, and the transcriptional factor STAT1 binds and facilitates the transcription of genes with GAS (IFNγ activation sequence) motif. Recently, we found that IFNγ also remodels chromatin structure through the regulation of histone acetylation as another layer of gene transcriptional regulation. However, IFNγ not only induces hundreds of genes, but also represses hundreds of others that promote alternative macrophage differentiation and function. The mechanism of IFNγ-mediated gene repression remains elusive. The IFNγ repressive effect is mainly thought to be indirect, and the general understanding has been limited to the balance and competition between IFNγ-STAT1 and IL10-STAT3. In this study, we investigate epigenetic mechanisms in IFNγ-mediated gene repression and focus on the role of Polycomb Repressive Complex 2 (PRC2)-mediated H3K27me3 in comparison with H3K27ac. We first discovered that H3K27 marks (ac and me3) are good indicators of gene activity state in human primary monocytes. H3K27me3 occupies silenced genomic regions and H3K27ac active regions, which correlates with the expression levels of the associated genes. Genes in bivalent domains, with both H3K27ac and H3K27me3 marks, have intermediate expression levels. Upon IFNγ stimulation, the change of these two histone marks correlates well with gene expression: increase/decrease of the H3K27ac histone mark agrees with increased/decreased gene expression respectively, while an opposite pattern is observed for H3K27me3. We also found that unlike H3K27ac, which undergoes dramatic genome-wide shift during IFNγ response, H3K27me3 mark is much less subject to modulation. In fact, genome-wide analysis shows regulation of this mark at only a few specific loci, indicating high levels of specificity of the PRC2-H3K27me3 mechanism in IFNγ response. Further functional/pathway analysis revealed that the targets of H3K27me3 regulation are genes that play important roles in monocyte differentiation and’ function. These findings indicate that PRC2-mediated transcriptional regulation plays an important role in macrophage inflammatory response.

Reconstructing A/B Compartments Using Long-Range Correlations in Epigenetic Data

Kasper D. Hansen, Jean-Philippe Fortin

A Hi-C experiment produces a genome-wide contact matrix whose entries estimate how often two distinct loci interact with each other. Analysis of Hi-C contact matrices have shown that at a gross scale, the genome can be divided into two compartments—closed and open—and that this compartmentalization is cell-type specific. Recent work has shown that 36% of these compartments change during stem cell differentiation. Here we show that genome compartments can be reliably estimated using DNA methylation data from the Illumina 450k platform, an inexpensive and popular methylation microarray. To do so, we show that the long-range correlations of methylation levels are substantially higher for two loci that belong to the “closed” compartment (“closed-closed” interaction) than for the two other types of interactions (“open-open” and “open-closed”)
interactions). By applying principal component analysis to the methylation correlation matrix, we can estimate where the “closed-closed” interactions occur and thereby obtain the genome compartmentalization at a 100kb resolution. We show that we are able to recover differences between cell types. This work makes it possible to systematically examine genome compartments in primary samples.

A Hidden Markov Random Field Based Bayesian Method for the Detection of Long-Range Chromosomal Interactions in Hi-C Data

Ming Hu
PhD, Assistant Professor, Department of Population Health, Division of Biostatistics, New York University School of Medicine

Motivation: Advances in chromosome conformation capture and next-generation sequencing technologies are enabling genome-wide investigation of dynamic chromatin interactions. For example, Hi-C experiments generate genome-wide contact frequencies between pairs of loci by sequencing DNA segments ligated from loci in close spatial proximity. One essential task in such studies is peak calling, that is, detecting non-random interactions between loci from the two-dimensional contact frequency matrix. Successful fulfillment of this task has many important implications including identifying long-range interactions that assist interpreting a sizable fraction of the results from genome-wide association studies. The task—distinguishing biologically meaningful chromatin interactions from massive numbers of random interactions—poses great challenges both statistically and computationally. Model-based methods to address this challenge are still lacking. In particular, no statistical model exists that takes the underlying dependency structure into consideration.

Results: In this paper we propose a hidden Markov random field (HMRF) based Bayesian method to rigorously model interaction probabilities in the two-dimensional space based on the contact frequency matrix. By borrowing information from neighboring loci pairs, our method demonstrates superior reproducibility and statistical power in both simulation studies and real data analysis.

Statistical Challenges in Epigenomics: Detecting Differentially Methylated Regions in the Presence of Unwanted Variability

Rafael Irizarry
Harvard University

I will describe new data analysis challenges related to scan statistics. I will describe the important role modern statistical techniques play in finding regions of the genome that are consistently different between disease and normal groups and new challenges specifically related to next-generation sequencing data. I will also describe batch effects and cell heterogeneity and how we deal with it. Finally, I will present some interesting Biological results related to development and cancer.

hMe-Bead-Integrated Click-seq (hMe-BIC-seq): An Integrated NGS Approach to Genome Wide 5-Hydroxymethylcytosine Profiling for Limiting Levels of DNA

Jennifer Ishii1, Matt Teater2, Doron Betel1,3, Alicia Alonso1, Ari Melnick1,2
1 Department of Medicine, Division of Hematology-Oncology, Epigenomics Core Facility, Weill Cornell Medical College
2 Department of Medicine, Division of Hematology-Oncology, Weill Cornell Medical College
3 Institute for Computational Biomedicine, Weill Cornell Medical College

Genome-wide, NGS assays for 5-hydroxymethylcytosine require large quantities of input DNA, and in the case of base-resolution methods also carry a large sequencing burden. These features can be prohibitive to studies with samples that have limiting amounts of DNA, such as translational studies involving clinical samples and those with limited funds. To provide a genome-wide assay amenable to low input material and
multiplex sequencing, we present here hMe-BIC-seq. This method employs both selective labeling of 5hmC moieties by beta-glucosyltransferase and biotinylation by click chemistry and it improves and expands on similar affinity-based methods (Szulwach et al., 2011) by integrating the chemical affinity-capture with on-bead library preparation. We used hMe-BIC-seq on both human brain DNA and a leukemia cell line K562 to map 5hmC-enriched genomic regions and show increased CpG abundance relative to random similar regions. We titrated input DNA from 1µg to 25ng with hMe-BIC-seq and showed strong concordance of ~100,000 5hmC-enriched peaks throughout the range, indicating the robust nature of this approach.

A Case Study in Epigenomic Research Reproducibility: The Elusive Effect of KDM5b Depletion on Embryonic Stem Cell Transcription and Intragenic Promoter Induction

Andrew D. Johnston, Maria-Paz Ramos, John M. Greally

The difficulty with the reproducibility of genome-wide studies is becoming increasingly apparent. We took advantage of an unusual situation in which three separate groups performed highly-comparable experiments, depleting the lysine demethylase KDM5b in mouse embryonic stem cells and testing effects on transcription and histone H3 lysine 4 trimethylation (H3K4me3). We reproduced previously-used experimental approaches and then re-analyzed all four data sets using current best practices. We show that none of the prior results is reproducible either through re-analysis or through replication of the experiments themselves.

Specifically, we show that the depletion of KDM5b does not enrich H3K4me3 at sites of KDM5b incorporation into chromatin. Our studies also fail to show any evidence for transcriptional dysregulation and do not reveal any changes in DNA methylation patterns at KDM5b locations or gene promoters. We use this study to represent how multiple independent studies of the same experimental system can report findings that are inconsistent and misleading. Furthermore, we examined the sources of variability among the studies focusing on data analysis practices and offer a framework to improve the reproducibility of genome-wide studies.

We conclude that these KDM5b studies represent a broader problem with manuscripts reporting genome-wide studies, which are fundamentally unreviewable, causing a failure to recognize problems of over-interpretation of data by the authors, at the review stage and subsequent to publication. The implementation of solutions for this problem will represent major challenges for investigators, journals and funding agencies combined.

Multiplexing of ChIP-Seq Samples in an Optimized Experimental Condition Has Minimal Impact on Peak Detection

Thadeous J. Kacmarczyk¹, Caitlin Bourque², Xihui Zhang³, Yanwen Jiang¹, Yariv Houvras², Alicia Alonso¹, Doron Betel¹,³

¹ Department of Medicine, Division of Hematology/Oncology, Epigenomics Core Facility, Weill Cornell Medical College
² Departments of Surgery and Medicine, Weill Cornell Medical College
³ Institute for Computational Biomedicine, Weill Cornell Medical College

Multiplexing samples in sequencing experiments is a common approach to maximize information yield while minimizing cost. In most cases the number of samples that are multiplexed is determined by financial consideration or experimental convenience with limited understanding on the effects on the experimental results. Here we set to examine the impact of multiplexing ChIP-seq experiments on the ability to identify a specific epigenetic modification. We performed peak detection analyses to determine the effects of multiplexing. These include false discovery rates, size, position and statistical significance of peak detection, and changes in gene annotation. We found that, for histone marker H3K4me3, one can multiplex up to 8 samples (7 IP + 1 input) at ~21 million reads each and still detect over 90% of all
peaks found when using a full lane for sample (~181 million reads). Furthermore, there are no variations introduced by indexing or lane batch effects and importantly there is no significant reduction in the number of genes with neighboring H3K4me3 peaks. We conclude that, for a well characterized antibody and therefore, model IP condition, multiplexing 8 samples per lane is sufficient to capture most of the biological signal.

**Integration of DNA Methylation and Gene Expression Profiles Improves Prognostic Stratification in MCL**

Violetta V. Leshchenko¹, Jessica Overbey², Pei-Yu Kuo¹, Deepak Perumal¹, Francine E. Garrett-Bakelman¹, Kieron Dunleavy¹, Adrian Wiestner⁵, Samir Parekh¹

¹ Division of Hematology and Medical Oncology, Icahn School of Medicine at Mount Sinai
² Department of Population Health Science and Policy, Icahn School of Medicine at Mount Sinai
³ Department of Medicine/Hematology-Oncology, Weill Cornell Medical College
⁴ Lymphoid Malignancies Branch, Center for Cancer Research, National Cancer Institute, National Institutes of Health
⁵ Hematology Branch, National Heart, Lung, and Blood Institute, National Institutes of Health

Mantle Cell Lymphoma (MCL) is an aggressive and incurable B cell malignancy accounting for about 6% of non-Hodgkin’s lymphomas. Despite recent advances in the prognosis of Mantle Cell Lymphoma (MCL) using MCL International Prognostic Index (MIPI), more MCL-specific biological prognostic factors need to be identified using modern high-throughput sequencing techniques. More potentially important prognostic molecular risk factors have been suggested but their applications in MCL are still limited. Epigenetic changes such as methylation of gene promoters have been demonstrated to contribute to the pathogenesis of both solid and hematological malignancies including MCL. To understand the genomic and epigenomic basis of differences in patient outcomes, particularly progression free survival (PFS), we carried out high-resolution genome-wide methylation analysis using enhanced RRBS (ERRBS) and correlated cytosine methylation to gene expression (GE) and patient outcomes. For ERRBS, genomic DNA was extracted from CD19+ selected cells from lymph node biopsies or peripheral blood samples from a Phase II study of Bortezomib plus DA-EPOCH-Rituximab in newly diagnosed MCL patients led by National Cancer Institute, Bethesda, MD. Library fragment lengths of 150–250 bp and 250–400 bp were prepared and gel isolated per Akalin et al. (PLOS Genetics, 2012) and sequencing was performed on an Illumina Hi-seq 2000. 26 out of 26 patient samples passed quality control with methylation assayed at >3.3 million CpG dinucleotides per sample on average 80× coverage per cytosine.

**How Genomes Fold: Now Inside the Loop**

Erez S Lieberman-Aiden
Baylor College of Medicine

We use in situ Hi-C to probe the 3D architecture of genomes, constructing haploid and diploid maps of nine cell types. The densest, in human lymphoblastoid cells, contains 4.9 billion contacts, achieving 1 kb resolution. We find that genomes are partitioned into contact domains (median length, 185 kb), which are associated with distinct patterns of histone marks and segregate into six subcompartments. We identify ~10,000 loops. These loops frequently link promoters and enhancers, correlate with gene activation, and show conservation across cell types and species. Loop anchors typically occur at domain boundaries and bind CTCF. CTCF sites at loop anchors occur predominantly (>90%) in a convergent orientation, with the asymmetric motifs “facing” one another. The inactive X chromosome splits into two massive domains and contains large loops anchored at CTCF-binding repeats.
A Survey of DNA Methylation Polymorphism in the Human Genome Identifies Environmentally Responsive Co-Regulated Networks of Epigenetic Variation

Andrew J. Sharp, Paras Garg, Ricky Joshi, Corey Watson
Department of Genetics and Genomic Sciences, Mount Sinai School of Medicine

Understanding the causes and consequences of genomic variation in populations is one of the major goals in the field of genetics. Over the past decade, studies such as the Hapmap and 1000 Genomes Projects have resulted in detailed maps of patterns of genetic variation in human populations that have facilitated studies of human disease and provided deep insights into diverse genomic processes. However, alterations of the primary DNA sequence are not the only type of genomic variations that occur among humans. In particular there are now well documented examples of epigenetic marks such as DNA methylation and histone modifications that show significant inter-individual variation. In contrast to sequence polymorphism, to date there are no robust maps of common epigenetic variation in humans, and, as a result, few insights have been made into their significance or underlying biology.

Here we set out to define sites of common epigenetic variation in humans and to characterize these loci, that we term Variable Methylation Regions (VMRs). In order to avoid the potential confounder of cellular heterogeneity, we focused our analyses on DNA methylation data derived from two populations of individuals assayed with the Illumina 450k array (n=58 to 111 individuals), representing five purified cell types: T-cells, B-cells, fibroblasts, neurons and glia. Using a robust approach designed to detect highly polymorphic loci, we identify hundreds of VMRs in each cell type that show common variability in DNA methylation levels in the normal population.

We find that VMRs occur preferentially at enhancers and in 3’ UTR regions, consistent with a functional role in regulating gene expression levels. We observed that at the majority of VMRs methylation levels show high heritability, indicating that many are associated with underlying genetic variation, and VMR loci are significantly enriched for mQTLs. However, we also observed a subset of VMRs distributed across the genome that show highly correlated variation in trans. In contrast to most other epigenetically variable regions, VMRs that form these co-regulated networks tend to have low heritability, differ between cell types and are enriched for specific biological pathways of direct functional relevance to each tissue. For example, in T-cells we defined a network of 61 co-regulated VMRs enriched for genes that form the T-cell receptor complex and play roles in T-cell activation; in fibroblasts a network of 21 co-regulated VMRs comprising all four HOX gene clusters that is highly enriched for control of tissue growth; and in glia a network of 66 VMRs enriched for roles in postsynaptic membrane organization. These VMR networks share common transcription factor binding sites that are significantly enriched within each network, indicating that the epigenetic state of these VMR networks is likely responsive to molecular signaling cascades induced by environmental cues. By culturing fibroblasts under varying conditions of nutrient deprivation and cell density, we experimentally demonstrate that methylation of the HOX gene cluster VMR network is responsive to environmental conditions, with methylation levels at these loci changing in a coordinated fashion in trans dependent on cellular growth.

Our study provides the first detailed map of common epigenetic variation in the human genome, showing that both genetic and environmental causes underlie this variation.

O-GlcNAc: A Regulator of RNA Polymerase III Transcription in Mammals and Protozoa?

Natalie C. Silmon de Monerri, Ian Willis, Kami Kim
Albert Einstein College of Medicine

O-GlcNAc is a posttranslational modification that is added to Serine and Threonine residues of target proteins, of which there are many, including chromatin-
associated proteins. By modifying transcription factors and histones, O-GlcNAc acts as both an activator and a repressor of transcription. *Toxoplasma gondii* is an obligate intracellular protozoan parasite, responsible for significant morbidity and mortality in immunocompromised individuals, and can cause birth defects if an immunologically naïve woman is infected during pregnancy. It shares many features of eukaryotic chromatin biology and has a homologue of the O-GlcNAc transferase enzyme responsible for the addition of O-GlcNAc to target proteins.

We confirmed the presence of O-GlcNAc on *T. gondii* histones by mass spectrometry and by western blot using O-GlcNAc specific antibodies. To study the function of O-GlcNAc on chromatin, we performed genome-wide native ChIP-seq, and were surprised to find an enrichment of O-GlcNAc at tRNA loci. We performed micro-RNAseq to determine whether the tRNA genes that are associated with the O-GlcNAc modification are expressed. We confirmed that the majority of O-GlcNAc peaks are found at active tRNA genes. To determine whether this phenomenon might occur in higher eukaryotes, in which the O-GlcNAc modification is better understood, we performed native ChIP-seq on human foreskin fibroblasts and mouse embryonic fibroblasts. Indeed, O-GlcNAc was enriched at tRNA loci in all cell types tested. Transcription of tRNA genes is mediated by RNA polymerase III, which is recruited to chromatin by polymerase-specific transcription factors. We compared the locations of the O-GlcNAc peaks to published ChIP-seq data for RNA polymerase III and its transcription factors and found that O-GlcNAc is present at the same loci. Furthermore, the shape of the O-GlcNAc ChIP-seq read pile-ups around active tRNA genes suggest that O-GlcNAc occurs on a RNA polymerase III transcription factor, rather than on the polymerase. Current studies are investigating the function of this modification and its potential role in the regulation of RNA polymerase III function.

**Spectacle: Fast Chromatin State Annotation Using Spectral Learning**

Jimin Song, Kevin C. Chen

Epigenomic data from ENCODE can be used to associate specific combinations of chromatin marks with regulatory elements in the human genome. Hidden Markov models and the expectation-maximization (EM) algorithm are often used to analyze epigenomic data. However, the EM algorithm can have overfitting problems in data sets where the chromatin states show high class-imbalance and it is often slow to converge. Here we use spectral learning instead of EM and find that our software Spectacle overcame these problems. Furthermore, Spectacle is able to find enhancer subtypes not found by ChromHMM but strongly enriched in GWAS SNPs.

**Ash2L and p300 Mediate Histone H3 Modifications at EGFR during its Developmental Silencing and Re-Expression in Gliomas**

Jessica Tome-Garcia

Gliomas are the most frequent and aggressive primary brain tumors, and are universally fatal once they progress to glioblastomas. Glioma cells commonly share abnormalities in pathways that normally control proliferation, migration and differentiation of glial progenitors; one such example is signaling through the epidermal growth factor receptor (*EGFR*). We show that *EGFR* is highly expressed in neural progenitors during fetal brain development, and that it becomes silenced during adulthood in most glia, except for those in the subventricular zone (SVZ) where adult neural progenitors are thought to reside. Interestingly, most low- and high-grade diffuse gliomas show strong re-expression of *EGFR*. The mechanism of this pathological overexpression is not clearly understood and cannot be accounted for by genomic abnormalities alone, since less than half of primary glioblastomas and almost no low-grade gliomas contain *EGFR*-activating gene amplification or mutations. We hypothesized that *EGFR* silencing...
during normal glial differentiation and its aberrant re-expression in gliomas is at least partly mediated via local epigenetic mechanisms at its promoter. To test this hypothesis, we have performed targeted epigenetic analyses at the **EGFR** promoter, measuring both DNA methylation levels via bisulfite sequencing and histone H3 methylation (me) and acetylation (ac) levels via chromatin immunoprecipitation (ChIP) using primary human samples. Surprisingly, we find that DNA methylation patterns at **EGFR** are conserved regardless of the **EGFR** expression status in non-neoplastic glia and in gliomas. In contrast, ChIP analysis reveals enrichment of the activating modifications H3K27ac and H3K4me3 during fetal brain development when **EGFR** is highly expressed, and their loss in adult white matter where **EGFR** is silenced. In the SVZ, where some glial cells continue to express **EGFR** at lower levels, these modifications are moderately enriched. The repressing modification H3K27me3, in contrast, is low during fetal development but becomes highly enriched in adult white matter. Interestingly, we also observe robust enrichment of H3K27ac and H3K4me3 in adult glioma specimens. Furthermore, we find significant binding of the histone methyltransferase complex Ash2L at **EGFR** only in fetal and glioma samples that show high levels of H3K4me3, and of the histone acetyltransferase p300 only in samples showing enrichment of H3K27ac, implicating these two histone modifying enzymes in the dynamic epigenetic regulation of **EGFR** expression during glial differentiation and glioma formation.

Our studies use endogenous human material and point to an important, region-specific role for chromatin remodeling in **EGFR** expression during gliomagenesis, which will undoubtedly open a broad range of new potential therapeutic tumor targets.

---

**Host Epigenetic Effects as a Transcriptional Regulatory Mechanism during Toxoplasma Gondii Infection**

Netha Ulahannan, Inessa Gendina, Masako Suzuki, Jessica Mar, Amit Verma, N. Ari Wijetunga, Fabien Delahaye, Matthew Croken, John Greally, Kami Kim

Department of Microbiology and Immunology, Department of Genetics, Albert Einstein College of Medicine

**T. gondii** alters host transcription during infection, presumably to facilitate its survival. However, the mechanisms by which it causes these alterations remain poorly understood. Epigenetic processes are involved in transcriptional regulatory changes that occur without changing the underlying DNA sequence. Any alterations to the host epigenome could mediate **T. gondii**’s transcriptional regulatory effects. We performed genome-wide studies of the host cell epigenome to determine if **T. gondii** infection affected DNA methylation or DNA hydroxymethylation of infected host cells.

We detected a global loss of 5-hydroxymethylcytosine (5-hmC) within the host during infection with **T. gondii**. Furthermore, we find an enrichment of dysregulated loci within CpG islands, CpG shores, and the promoter and enhancer regions of genes, as well as within several specific transcription factor binding sites, thus demonstrating that the epigenetic modifications occur within cis-regulatory regions. 5-methylcytosine is converted into 5-hmC through the action of the TET family of enzymes. We found that TET enzymatic activity was significantly decreased following the infection, with metabolomic studies showing a shift in host metabolism towards an increased rate of glycolysis. These observations have caused us to hypothesize further that the detected epigenetic modifications occur as a result of induction of a “Warburg-like” state within the infected host cell, resulting in a depletion of alpha-ketoglutarate, normally produced by the tricarboxylic acid (TCA) cycle and a co-factor for TET enzymatic activity. Consistent with a primary role of the Warburg Effect, we find dysregulation of expression of genes for the key.
A Trimming-and-Retrieving Alignment Scheme for Reduced Representation Bisulfite Sequencing

Xuefeng Wang1,2,3, Xiaoqing Yu4, Wei Zhu3, W. Richard McCombie5, Eric Antoniou5, R. Scott Powers6,7, Nicholas O. Davidson1, Ellen Li8, Jennie Williams1

1 Department of Preventive Medicine, Stony Brook University
2 Department of Biomedical Informatics, Stony Brook University
3 Department of Applied Mathematics and Statistics, Stony Brook University
4 Department of Biostatistics, Yale University
5 Cold Spring Harbor Laboratory
6 Department of Pathology, Stony Brook University
7 Department of Medicine, Washington University
8 Department of Medicine, Stony Brook University

Currently available bisulfite sequencing tools frequently suffer from low mapping rates and low methylation calls, especially for data generated from the Illumina sequencer, NextSeq. Here, we introduce a sequential trimming-and-retrieving alignment approach for investigating DNA methylation patterns, which significantly improves the number of mapped reads and covered CpG sites. The method is implemented in an automated analysis toolkit for processing bisulfite sequencing reads. The introduced package (BStools) also provides a one-stop integrated analysis workflow for Bisulfite sequencing data generated from other platforms: from alignment to DMR detection and annotation, and which is complemented by many useful graphic tools.

Early and Progressive DNA Methylation Differences at PRC2 Targets Suggest a Role for Polycomb-Group Proteins in Progression to and Detection of Cervical Cancer

N. Ari Wijetunga, Mark H. Einstein, Nicolas Schlecht, Miriam Ben-Dayan, Robert D. Burk, John M. Greally

Cervical cancer development involves a multistep pathogenesis with alterations in both genetic and epigenetic regulation of genes. There is limited knowledge of the extent to which widespread epigenetic disturbances occur and whether epigenome-wide effects are associated with disease progression. Examination of the relationship between epigenetic regulation and development of cervical cancer would improve our understanding of the disease, reveal potential early patterns of disease progression, and define new drug targets for novel therapies.

We obtained 128 patient biopsies representing normal cervix, Cervical Intraepithelial Neoplasia (CIN), and cancer from HIV positive and negative women. We quantified methylation at ~2 million CpGs using HELP-tagging, an unbiased restriction enzyme-based sequencing approach to assay genome-wide DNA methylation. We built multivariable polynomial regression models to obtain covariate-adjusted effects of the relationship between DNA methylation and lesion grade.

We observed substantial methylation perturbation genome-wide and identified approximately 4000 unique CpG’s with progressive disruption in DNA methylation. Increasing DNA methylation was associated with 457 gene promoters, and we observed a striking enrichment of 111 (24.3%) Polycomb Repressive Complex 2 (PRC2) target genes. This targeted promoter methylation occurs during early CIN, increases with lesion grade, and occurs in lesion-adjacent normal cervical tissue. Our findings indicate that an epigenetic field defect to PRC2 gene targets occurs reflecting a possible stem-cell like phenotype of epithelial cells during the development of cervical cancer. Additionally, early and progressive DNA methylation defects within PRC2 targets occurring
Integrative Analysis of DNA Methylation and Gene Expression Data Reveals Complex Regulation of Gastric Cancer

Seungyeul Yoo1,2, Suet-Yi Leung3, Jun Zhu1

1 Icahn Institute of Genomics and Multiscale Biology, Icahn School of Medicine at Mount Sinai School of Medicine, New York, USA
2 Department of Genetics and Genomic Sciences, Icahn School of Medicine at Mount Sinai School of Medicine, New York, USA
3 Department of Pathology, The University of Hong Kong, Pokfulam, Hong Kong

Gastric cancer is a heterogeneous disease where diverse genetic and epigenetic alternations can accumulate in different molecular and histological subtypes. We applied our recently developed causality test between genome-wide DNA methylation and gene expression profiles to two independent cohorts (97 tumor samples in Hong Kong University (HKU) and 365 samples in TCGA stomach adenocarcinoma (TCGA)). We focused on methylation variations within CpG islands, where global hypermethylation was observed, and identified 44 and 538 key regulators in HKU and TCGA dataset respectively. There were 28 common key regulators in both dataset (Fisher’s exact test p-value = 2.0×e-28) whose methylation variations had high impact on mRNA level changes of large number of downstream genes. Several key regulators were known for the association between their epigenetic disruption and the disease (for example, BNIP3, CDO1, TCF21, ZSCAN18, and so on) while other genes have not implicated in the gastric cancer previously. Further clustering key regulators based on their downstream genes overlaps revealed that there were two distinct groups of downstream genes commonly regulated by these key regulators and the expression of these two groups were anti-correlated. One group was enriched for cell cycle related genes and the other group was enriched for genes involved in immune responses. This result indicates that cell cycle and immune response functions were inversely regulated by methylation variations of the same set of genes. It is worth to note that methylation patterns of some key regulators were subtype dependent and the subtype specific methylation patterns were only observed in tumor samples, but not in adjacent normal tissues. Based on integrative analysis of genome-wide DNA methylation and gene expression profiles within two independent gastric cancer dataset, we identified a set of key regulators whose methylation changes might play a ‘causal’ role in the transcriptional regulation associated with the gastric cancer. Further experiments are needed to validate and dissect these putative candidate genes’ roles in tumorigenesis and progression of this complex and heterogeneous disease.

Off the Beaten Path: Newer Applications of NGS Technology

Michael Smith
PhD, Illumina

Over the past few years, Next Generation Sequencing has been used to determine the genomic sequence of thousands of organisms, identify disease-causing variants by focusing on coding content within exomes and monitor gene expression differences between cells and tissues by RNA-seq. While much effort has been devoted to these common applications, many researchers are interested in other aspects of the genome including epigenetics, origins of replication, cis -and trans- regulatory relationships and more. Here, we will provide a survey of some interesting, newer applications of this powerful technology.
About NYGC

The New York Genome Center (NYGC) is an independent, nonprofit at the forefront of transforming biomedical research and clinical care with the mission of saving lives. As a consortium of renowned academic, medical and industry leaders across the globe, NYGC focuses on translating genomic research into clinical solutions for serious disease. Our member organizations and partners are united in this unprecedented collaboration of technology, science, and medicine. We harness the power of innovation and discoveries to improve people’s lives - ethically, equitably, and urgently. Member institutions include: Albert Einstein College of Medicine, American Museum of Natural History, Cold Spring Harbor Laboratory, Columbia University College of Physicians and Surgeons, Cornell University/Weill Cornell Medical College, Hospital for Special Surgery, The Jackson Laboratory, Memorial Sloan Kettering Cancer Center, Icahn School of Medicine at Mount Sinai, NewYorkPresbyterian Hospital, The New York Stem Cell Foundation, New York University, North Shore-LIJ, The Rockefeller University, Roswell Park Cancer Institute, Stony Brook University and IBM. For more information, visit: www.nygenome.org.

Our Mission

The New York Genome Center (NYGC) is at the forefront of transforming biomedical research and clinical care with the mission of saving lives.

As a consortium of renowned academic, medical and industry leaders across the globe, NYGC focuses on translating genomic research into clinical solutions for serious disease.

Our member organizations are united in this unprecedented collaboration of technology, science, and medicine. We harness the power of innovation and discoveries to improve people’s lives –ethically, equitably, and urgently.