Final Progress Report for Research Projects Funded by Health Research Grants

Instructions: Please complete all of the items as instructed. Do not delete instructions. Do not leave any items blank; responses must be provided for all items. If your response to an item is "None", please specify "None" as your response. "Not applicable" is not an acceptable response for any of the items. There is no limit to the length of your response to any question. Responses should be single-spaced, no smaller than 12-point type. The report **must be completed using MS Word**. Submitted reports must be Word documents; they should not be converted to pdf format.

1. **Grantee Institution:** The Pennsylvania State University

2. **Reporting Period (start and end date of grant award period):** 1/1/2010 - 12/31/2013

3. **Grant Contact Person (First Name, M.I., Last Name, Degrees):** John Anthony, MPA

4. **Grant Contact Person's Telephone Number:** 814 935 1081

5. **Grant SAP Number:** 4100050904

6. **Project Number and Title of Research Project**: 5. Epigenetic Changes in Myeloid Differentiation and Acute Myeloid Leukemia

7. **Start and End Date of Research Project:** 1/1/2010 - 6/30/2013

8. **Name of Principal Investigator for the Research Project:** Sergei Grigoryev, PhD

9. **Research Project Expenses**.

   9(A) Please provide the total amount of health research grant funds spent on this project for the entire duration of the grant, including indirect costs and any interest earned that was spent:

   $ 61,164

   9(B) Provide the last names (include first initial if multiple individuals with the same last name are listed) of **all** persons who worked on this research project and were supported with health research funds. Include position titles (Principal Investigator, Graduate Assistant, Post-doctoral Fellow, etc.), percent of effort on project and total health research funds expended for the position. For multiple year projects, if percent of effort varied from year to year, report in the % of Effort column the effort by year 1, 2, 3, etc. of the project (x% Yr 1; z% Yr 2-3).

Last Name, First Name	Position Title	% of Effort on Project	Cost
Correll, Sarah	Graduate assistant	41.7	11,180

9(C) Provide the names of **all** persons who worked on this research project, but who *were not* supported with health research funds. Include position titles (Research Assistant, Administrative Assistant, etc.) and percent of effort on project. For multiple year projects, if percent of effort varied from year to year, report in the % of Effort column the effort by year 1, 2, 3, etc. of the project (x% Yr 1; z% Yr 2-3).

Last Name, First Name	Position Title	% of Effort on Project
Grigoryev, Sergei	Principal Investigator	2
Loughran, Thomas	Principal Investigator	1
Claxton, David	Co-Investigator	2

9(D) Provide a list of **all** scientific equipment purchased as part of this research grant, a short description of the value (benefit) derived by the institution from this equipment, and the cost of the equipment.

Type of Scientific Equipment	Value Derived	Cost
None		

10. Co-funding of Research Project during Health Research Grant Award Period. Did this research project receive funding from any other source during the project period when it was supported by the health research grant?

Yes_____ No_____X_____

If yes, please indicate the source and amount of other funds:

11. Leveraging of Additional Funds

11(A) As a result of the health research funds provided for this research project, were you able to apply for and/or obtain funding from other sources to continue or expand the research?

Yes_____ No____X_____

If yes, please list the applications submitted (column A), the funding agency (National Institutes of Health—NIH, or other source in column B), the month and year when the application was submitted (column C), and the amount of funds requested (column D). If you have received a notice that the grant will be funded, please indicate the amount of funds to be awarded (column E). If the grant was not funded, insert “not funded” in column E.

Do not include funding from your own institution or from CURE (tobacco settlement funds). Do not include grants submitted prior to the start date of the grant as shown in Question 2. If you list grants submitted within 1-6 months of the start date of this grant, add a statement below the table indicating how the data/results from this project were used to secure that grant.

A. Title of research project on grant application	B. Funding agency (check those that apply)	C. Month and Year Submitted	D. Amount of funds requested:	E. Amount of funds to be awarded:
None	☐NIH ☐ Other federal (specify:_____) ☐ Nonfederal source (specify:_)		$	$

11(B) Are you planning to apply for additional funding in the future to continue or expand the research?

Yes____X_____ No_____

If yes, please describe your plans:

A collaborative RO1 grant application by S. Grigoryev (PI) and D. Claxton (Co-PI) tentatively titled "Epigenetic background of in myeloid disorders" is being prepared and planned to be submitted to the NIH in year 2014. Alternatively, a smaller application will be submitted to the Leukemia and Lymphoma society.

12. Future of Research Project. What are the future plans for this research project?

Our work has identified new groups of genes that are transcriptionally deregulated in AML and belong to chromosomal domains with long-range epigenetic changes. In the future, we plan to proceed to studying the role of particular epigenetic factors (HMTases) that affect gene expression at specific chromosomal locations in AML and use the whole-genome epigenetic mapping as a versatile tool to monitor the efficiency of the epigenetic therapies of various forms of leukemia including differences between individual patients. In this institution, our work coincided with creation of the Institute of Personalized Medicine that provides an excellent infrastructure for future genomic and epigenomic characterization of leukemia and other blood disorders for better diagnostics and more efficient personalized therapy. The future of this project is seen in epigenetic analysis of multiple patients' samples and better characterization of AML and other blood disorders amenable to efficient epigenomic screening.

13. New Investigator Training and Development. Did students participate in project supported internships or graduate or post-graduate training for at least one semester or one summer?

Yes____X_____ No_____

If yes, how many students? Please specify in the tables below:

	Undergraduate	Masters	Pre-doc	Post-doc
Male				
Female			1	
Unknown				
Total			**1**	

	Undergraduate	Masters	Pre-doc	Post-doc
Hispanic				
Non-Hispanic			1	
Unknown				
Total			**1**	

	Undergraduate	Masters	Pre-doc	Post-doc
White			1	
Black				
Asian				
Other				
Unknown				
Total			**1**	

14. **Recruitment of Out-of–State Researchers**. Did you bring researchers into Pennsylvania to carry out this research project?

    Yes_____ No____X _____

    If yes, please list the name and degree of each researcher and his/her previous affiliation:

15. **Impact on Research Capacity and Quality**. Did the health research project enhance the quality and/or capacity of research at your institution?

    Yes___X_____ No_____

    If yes, describe how improvements in infrastructure, the addition of new investigators, and other resources have led to more and better research.

    The development of ChIP-sequencing technique for analysis of individual AML samples helped to improve the infrastructure at the Genomic core facility at Penn State Hershey Institute for Personalized Medicine. This project allowed us to successfully introduce ChIP-seq method in analysis of patient's samples and to develop ChIP-seq standard operation protocols for use at the IPM. In addition, these resources contributed to acquiring and testing the automated library preparation and next-generation sequencing. For Sarah Correll, who successfully defended a PhD dissertation thesis in June 2011, the project helped to get research experience in

epigenetics and find a postdoctoral position at the Children's Hospital of Philadelphia. The acquisition of NextGen software allowed us to gain experience in the conduct of whole-genomic ChIP mapping and it strengthened a network of collaborations with our Bioinformatic core (Drs. Berg and Salzberg) and other research laboratories at Penn State Hershey (Drs. Barnstable, Spratt, Broach). This project also helped to recruit new faculty in genomics and bioinformatics (Drs. Kawasawa and Yue) by opening possibilities for collaboration on epigenetic analysis of a human disorder.

16. Collaboration, business and community involvement.

16(A) Did the health research funds lead to collaboration with research partners outside of your institution (e.g., entire university, entire hospital system)?

Yes___X_____ No_____

If yes, please describe the collaborations:

The results were reported at the annual Medicine Research Day in May 2013 that includes the physicians from the entire hospital system. This project enabled us to recruit a new collaborator, Jihua Cheng MD, PhD, a resident fellow of the Hematology/oncology division of the M.S. Hershey Medical Center who volunteered to participate in this project in addition to his clinical duties. Dr. Thomas Loughran MD, who has most recently moved to a new position as Director of the Cancer Institute at the University of Virginia, Charlottesville, continues to collaborate as a co-author on preparing a new manuscript based on this work.

16(B) Did the research project result in commercial development of any research products?

Yes_____ No____X_____

If yes, please describe commercial development activities that resulted from the research project:

16(C) Did the research lead to new involvement with the community?

Yes_____ No_____X_____

If yes, please describe involvement with community groups that resulted from the research project:

17. Progress in Achieving Research Goals, Objectives and Aims.

List the project goals, objectives and specific aims (as contained in the grant agreement). Summarize the progress made in achieving these goals, objectives and aims for the period that the project was funded (i.e., from project start date through end date). Indicate whether or not each goal/objective/aim was achieved; if something was not achieved, note the reasons why. Describe the methods used. If changes were made to the research goals/objectives/aims, methods, design or timeline since the original grant application was submitted, please describe the changes. Provide detailed results of the project. Include evidence of the data that was generated and analyzed, and provide tables, graphs, and figures of the data. List published abstracts, poster presentations and scientific meeting presentations at the end of the summary of progress; peer-reviewed publications should be listed under item 20.

This response should be a DETAILED report of the methods and findings. It is not sufficient to state that the work was completed. Insufficient information may result in an unfavorable performance review, which may jeopardize future funding. If research findings are pending publication you must still include enough detail for the expert peer reviewers to evaluate the progress during the course of the project.

Health research grants funded under the Tobacco Settlement Act will be evaluated via a performance review by an expert panel of researchers and clinicians who will assess project work using this Final Progress Report, all project Annual Reports and the project's strategic plan. After the final performance review of each project is complete, approximately 12-16 months after the end of the grant, this Final Progress Report, as well as the Final Performance Review Report containing the comments of the expert review panel, and the grantee's written response to the Final Performance Review Report, will be posted on the CURE Web site.

There is no limit to the length of your response. Responses must be single-spaced below, no smaller than 12-point type. If you cut and paste text from a publication, be sure symbols print properly, e.g., the Greek symbol for alpha (α) and beta (ß) should not print as boxes (□) and include the appropriate citation(s). DO NOT DELETE THESE INSTRUCTIONS.

Acute myeloid leukemia (AML) is a highly malignant blood cancer with an estimated 12,810 new cases in 2009 and the lowest survival rate (23.4%) among the four major types of leukemia. This disease is caused by a combination of genetic rearrangements and epigenetic changes. Based on our previous studies of developmentally-regulated epigenetic and transcriptional changes associated with myeloid differentiation and leukemia as well as other data known from the literature, we hypothesize that epigenetic changes manifested in mis-localization of a repressive histone modification, histone H3 dimethylation at lysine 9 (H3K9me2) may interfere with transcription during myeloid differentiation and promote AML.

In this project we will conduct a detailed whole-genomic characterization of the localization and expression of H3K9me2 and identify clusters of genes involved in epigenetic changes marked by this histone modification. This project has two specific aims.

Specific aim 1 is to map genome-wide topography of histone H3(K9) demethylation and its transitions in the process of normal myeloid differentiation. We will use ChIP-sequencing to map this histone modification over the genomes of normal myeloid precursors and differentiated granulocytes. Our analysis should reveal sites of H3K9 dimethylation involved in normal myeloid differentiation.

Specific aim 2 is to determine spatial and functional relationship between histone H3(K9) methylation and transcriptional changes in AML. The proposed work includes an innovative application of the cutting-edge genomic sequencing and bioinformatics analysis to study association of chromatin modification topographies with transcription in leukemia and is expected to reveal new chromosomal loci altered in AML that can be served for the disease prognosis as well as targets for its epigenetic therapy.

Techniques used for whole-genome ChIP-sequencing and bioinformatics analyses of histone modifications in AML and their optimization.

Cells and blood samples:
Normal polymorphonuclear granulocytes (predominantly neutrophils) were isolated from discarded fresh white blood cells (buffy coats) from unidentified healthy donors collected at Hershey Medical Center Blood Bank (IRB protocol # HY03-136EP-A) using standard OptiPrep density centrifugation [1] and resuspended in PBS buffer. Cryopreserved AML cells isolated from bone-marrow samples by Ficoll-Pacque density gradient centrifugation collected at Claxton laboratory (IRB protocol 2000-186) were thawed and resuspended in PBS for fixation immediately after thawing. Cultured cells (e.g. K562) in suspension were washed 2 x PBS with centrifugation 5 min at 200 g and then resuspended in PBS for fixation.

Formaldehyde fixation:
Living cell cultures or freshly thawed cells were suspended in PBS at ~ $5x10^7$ cell/ml at room temp., mixed with 1/10 of the final volume of 10% formaldehyde in PBS freshly prepared from 37% stock (Fisher, ACS reagent F79-500). Cells were incubated the 1% formaldehyde for 10-15 min at room temp on a wheel. Fixation was stopped by adding 1M glycine to final concentration 125 mM. Cells were immediately spun down in pre-chilled centrifuge at +4°C and 4000 g for 5 min. and washed 3 times with 20 ml of cold PBS (spinning each time as above). The fixed cells were counted to estimate the yield and stored overnight at +4°C.

Nuclear Isolation:
Cell nuclei were isolated using a protocol that employed physiological salt during the nuclear lysis and purification through sucrose layer as described [2]. The purified nuclei were resuspended in micrococcal nuclease digestion buffer (50 mM Tris-HCl, pH=7.6; 3 mM CaCl2) at ~1 mg/ml DNA.

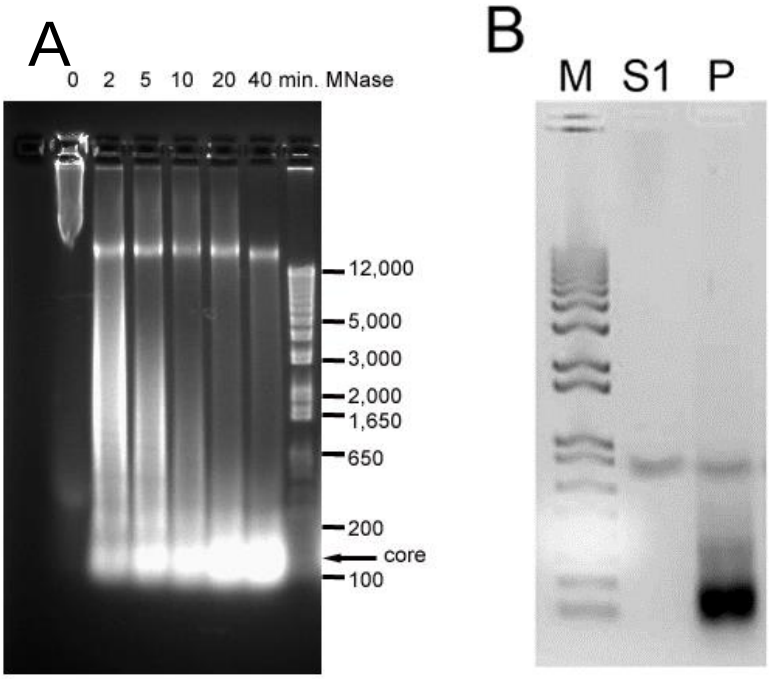

Figure 1. A: Agarose gel electrophoresis shows DNA from granulocyte nuclei in the process of Micrococcal nuclease (MNase) digestion for 0 – 40 min. as indicated. MNase converts granulocyte chromatin into nucleosome cores containing ~150 bp DNA fragments (indicated by the arrow) suitable for ChIP and genomic sequencing. B: DNA from solubilized fractions S1 and pellet (P) resuspended in L-ChIP buffer.

Micrococcal nuclease digestion and chromatin solubilization:
The nuclear preparation (1 ml) was digested with 60 units of Micrococcal nuclease (Roche) for 40 min. at + 37°C to ~150 bp nucleosome sizes (Fig. 1A) and digestion was stopped by adding 3 mM EDTA. After digestion, the supernatant (S1) was collected and discarded and the nuclear pellet (P, Fig. 1B) resuspended in SDS-containing lysis buffer with light sonication. A typical yield for the nuclear lysate was ~ 0.5 mg DNA for 10^8 cells.

Chromatin Immunoprecipitation (ChIP):
The ChIP protocol was modified from that previously used in our lab [3] to use with cryopreserved hematological tissues and MNase fragmentation. In this project, we have successfully used anti-H3K9me2 (ab1220), anti-H3K9me3 (ab8898), anti-H3 C-tail (ab8898) from Abcam; and Anti-H3K4me2 (07–030) from Upstate. For quality assessment of those antibodies see reference [4].

Nuclear lysate P (0.350 ml) was spun in Eppendorf centrifuge at maximal speed for 5 min. and the supernatant was diluted 10 fold with D-CHIP buffer (0.01% SDS; 1.1% Triton X-100; 1.2 mM EDTA; 16.7 mM Tris-HCl pH 8.0; 167 mM NaCl). 3 mkl of Protease inhibitor cocktail (general use, Sigma. Cat# P 2714), 3 mkl 100 mM PMSF, and ~5 mkg of antibody or preimmune serum were added to the solution and rotated at +4^0 C overnight in 15 ml Falcon tubes. 90 mkl of protein A Sepharose (Sigma cat.# P9424) beads slurry was washed 2 times with 0.4 ml of washing buffer (9:1 = D-CHIP: L-CHIP), resuspended in 0.4 ml washing buffer, mixed with 25 mkl 10mg/ml

salmon sperm DNA (0.5 mg/ml final) and 5 mkl 10mg/ml BSA (100 mkg/ml final) and rotated at $+4^0$ C overnight. L-CHIP (lysis) buffer: 1% SDS; 10mM EDTA; 50 mM Tris HCl pH 8.0.

On the next day, the beads were washed 2 times with D-CHIP buffer and resuspended in 300 mkl washing buffer (same as above = 9 D-CHIP: 1 L-CHIP). ~100 mkl beads slurry was added to each Falcon tube with IP and rotated at 4^oC for 2 hours. The Falcon tubes were gently centrifuged with IP reactions (2000 rpm, 5min), the unbound fraction discarded, and the beads resuspended in 1 ml LS-CHIP (low salt) buffer (0.1% SDS; 1% Triton X-100; 2 mM EDTA; 20 mM Tris-HCl pH 8.0; 150 mM NaCl). The beads were transferred to 1.5 ml Eppendorf tubes and washed 4 times with 1 ml of LS-CHIP buffer and 1 time with HS-CHIP (high salt) buffer (0.1% SDS; 1% Triton X-100; 2 mM EDTA; 20 mM Tris-HCl pH 8.0; 500 mM NaCl). The beads were spun at 2000 rpm in Eppendorf centrifuge after each wash. The washed material was eluted with 350 mkl of E-CHIP buffer and incubated by rotating for 10min at room temp., then spinning 2 min. at 12.000 rpm and collecting the liquid supernatant designated as IP CHIP. 50 mkl of starting lysate was mixed with 300 mkl E-ChIP buffer and designated as IP Input. 17.5 mkl RNase A (10 mg/ml) and 17.5 mkl Proteinase K (10mg/ml; final concentration for both 0.5 mg/ml) was added to each tube with IP CHIP or IP Input and incubated for 30 min at 37^0C. Finally, the cross-links were reversed by incubating the tubes at 65^oC overnight.

On the third day, the IP materials were extracted 2 times with phenol/chlorophorm/isoamyl alcohol and 1 times with chloroform by using Phase Lock tubes (5 PRIME, Hamburg, ref# 2302820). Each 350 mkl DNA solution was precipitated with 1.6 ml 100% ethanol, plus 4 mkl glycogen Sigma G1767 (20 mg/ml), plus 40 mkl Na acetate (3M, pH=5.5). The final DNA was dissolved in 50-100 mkl of fresh, DNase-free water. DNA concentration was measured using Qubit fluorometer (HS mode).

<u>Controls and yields:</u>
For typical controls, we conducted two parallel immunoprecipitation reactions with a) no antibody added; b) similarly prepared IgG from a none-immunized animal (e.g. rabbit IgG #2729, Cell Signaling). A typical yield from immunoprecipitation with antibodies against H3K9me2 (Abcam, ChIP Grade, ab1220) was more than 400 ng for 40 mkg of the input (ratio >1%). Nonspecific binding in the control samples was < 0.02%. The yields varied between different cells depending on the histone modification level. For example, for a number of primary AML cells, yields of ChIP-recovered DNA were in the range of 420 - 4200 ng and ratio of ChIP to input was between 0.45 and 3.4% (Table 1). These variations are consistent with natural variations of H3K9me2 levels in leukemia blood cells described by us before [1].

AML	4	5	6	7	8	9	10
Cell x10^8	1.2	1.2	1	1.2	1	1	0.9
DNA ng nuclei	1405000	186000	1205000	275000	760000	1137000	817000
DNA ng input	403200	92167	336000	86333	136267	125300	115500
DNA ng ChIP	4280	418.8	1840	924	2184	1824	1452
ChIP/input	0.0106	0.0045	0.0055	0.0107	0.0160	0.0146	0.0126

AML	11	12	13	14	15	16
Cell x10^8	1.6	1.2	1.1	1	0.96	0.78
DNA ng nuclei	711000	171000	630000	1022000	402500	456000
DNA ng input	70467	29400	62300	90067	73266	55533
DNA ng ChIP	1416	993.6	721.2	1344	1548	472.8
ChIP/input	0.0201	0.0338	0.0116	0.0149	0.0211	0.0085

Table 1. Yield of chromatin immunoprecipitation from AML-derived myeloid progenitors with H3K9me2 antibodies.

ChIP optimization:

We conducted several experiments to test antibody efficiency for immunoprecipitating chromatin labeled by H3K9me2. First, we conducted ChIP using two different commercially available types of antibodies at the same concentration: Abcam, ChIP Grade, ab1220 and Millipore 05-1249. We observed that for the same concentration of antibody used, the yield for Abcam antibodies was, on average, 1.84 fold higher so that in subsequent experiments we used only Abcam ab 1220. As Abcam antibodies are IgG2a type that could potentially react with protein A and protein G, we set experiments to test which type of immobilized protein is best suited for immunoprecipitation. For this analysis, we tested similar amount of sonicated material (36 mkg) to react with the same amounts of Abcam antibodies (1.8 mkg) and immunoprecipitated with a) Protein A beads b) protein G beads c) Protein A-G combination. We observed that protein A was the most efficient of immunoprecipitating H3K9me2 chromatin but in combination with protein G increased the efficiency by 20% (Fig. 2A). In all subsequent experiments we employed a combination of protein A-G beads. Next we tested the optimal concentration of antibodies for immunoprecipitating of fixed amounts of material (each sample contained 36 mkg of DNA) with variable concentration of antibodies (0 - 3.6 μg). We found that the maximal amount of antibody was most efficient to pull down material for deep sequencing (Fig. 2B). In all subsequent experiments we used 3.6 mkg of antibodies that appeared to be a cost-efficient way of conducting ChIP for H3K9me2. Finally, we estimated the number of cells providing necessary amount for ChIP by keeping the amount of antibody constant (2.7 mkg/ml) and changing the amount of material from 0 to 36 mkg per sample. We estimated that with 3.6 mkg of input DNA (corresponding to a yield from ~$2x10^6$ cells) we could generate 100 ng ChIP DNA sufficient for robust sequencing (Fig. 2C).

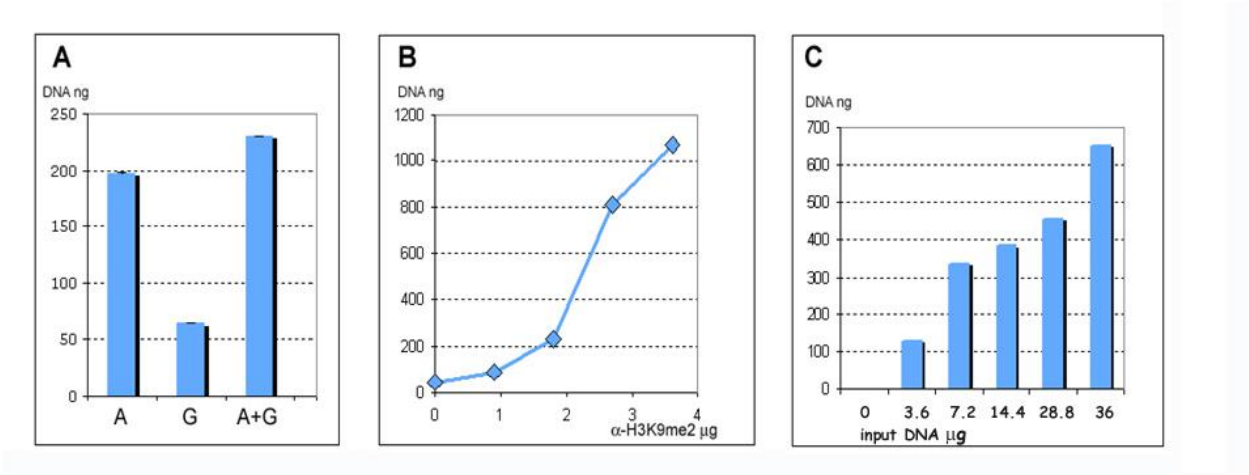

Figure 2. Testing and optimizing conditions for chromatin immunoprecipitation with H3K9me2 antibodies. A: similar amount of chromatin (36 mkg by DNA) was reacted with 1.8 mkg of Abcam ab 1220 antibodies and immunoprecipitated with Protein A beads, protein G bead, and Protein A+G combination. Control immunoprecipitation contained no antibodies and A+G beads. B: chromatin samples each containing 36 mkg of DNA were reacted with variable concentration of α-H3K9me2 antibodies as indicated. C: Chromatin samples containing different amounts of input DNA (indicated at the X-axis) were reacted each with 2.7 mkg α-H3K9me2 antibodies. Y-axis shows the amount of DNA recovered by ChIP.

Library preparation and Next-Generation sequencing:

For Illumina sequencing we used NEBNext ChIP-Seq library preparation reagent set for Illumina (cat #E6200S/L for 12/60 reactions) and NEBNext Multiplex Oligos for Illumina Index Primers Set 1 (cat #E7335S/L for 24/96 reactions). For one library, we took 10-30 ng of chromatin-immunoprecipitated or control DNA in ≤ 40 mkl of water or elution buffer. End repair of ChIP DNA, dA-Tailing, Adaptor ligation, and cleaning of DNA by AMPure XP magnetic beads (Agencourt part # A63880), were conducted according to NEBNext ChIP-Seq library preparation manual with the exception that after adaptor ligation and AMPure cleaning (step 4), the material was cleaned and redissolved in 20 mkl of 0.1 x TE buffer for agarose gel size selection. To run an agarose gel, we added 4 µl of 6X Gel Loading Dye to 20 µl of sample DNA and prepared a 50 ml, 2% agarose with SyBr Gold gel using 1 X TAE buffer. We then added 32 mkl TAE Buffer with 1 x SyBr Gold and 7 mkl of 6X Gel Loading Dye to 1 mkl of DNA ladder (100 bp DNA ladder with 6X gel loading dye – NEB N3231S) and loaded 20 mkl of the ladder solution onto one lane of the gel and the 20 mkl samples onto the other lanes of the gel, leaving a gap of at least one empty lane between samples and ladders.

We ran the gel at 90 V constant voltage for 90 minutes, viewed the gel on a Dark Reader transilluminator, excised a band from the gel spanning the width of the lane and ranging in size from 175-225 bp (150 bp nucleosome plus 50 bp adaptors) with a clean scalpel using the DNA ladder as a guide. We took a snapshot of the gel after band excision for a record (Figure 3A), purified the DNA on one Qiagen QIAquick (cat# 28704) gel extraction mini column and eluted in 50 mkl of sterile water. We took 23 mkl for one PCR enrichment of adaptor-ligated PCR conducted as described in NEB Next ChIP-Seq library preparation manual. We used one universal primer and one out of 12 index primers (NEB cat #E7335S) for multiplexing. We used 15-17 cycles of PCR

depending on the amount of starting material. After PCR, we cleaned the PCR-amplified library DNA by AMPure XP magnetic beads, redissolved in 20 mkl of 0.1 x TE buffer and measured DNA concentration using Qubit fluorometer (HS mode). Typical yield was 40 – 100 ng/mkl. For quality controls, we took x 1 mkl from each PCR reaction, added 9 µl TAE Buffer with 1 x SyBr Gold and 2 mkl of 6X Gel Loading Dye and ran DNA electrophoresis in TAE buffer/2% agarose (80 V, 40 min) with SyBr Gold (Figure 3B). For precise analysis of DNA size distribution, we diluted the library 50 fold with nuclease free water, and assessed the library quality on a Bioanalyzer® (Agilent high sensitivity chip). Note that the electrophoregram shows a narrow distribution with a peak size around 275 bp as expected (Figure 4)

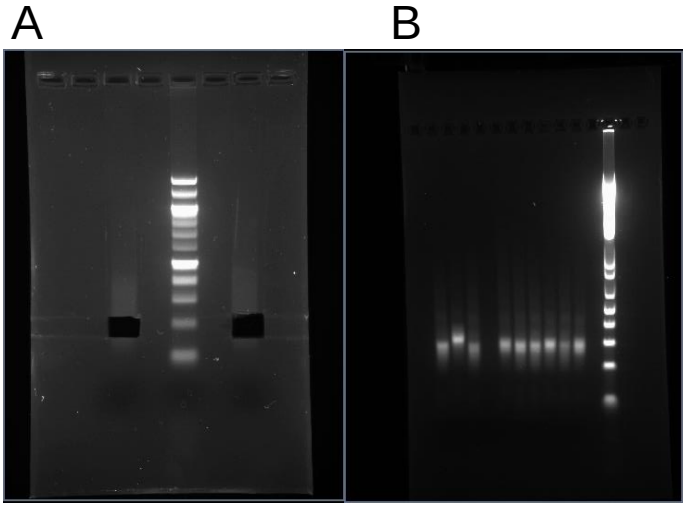

Figure 3. SyBr Gold stained 2% agarose gel after excision of two 175-225 bands corresponding to adaptor-conjugated library DNA (A) and after PCR amplifications (15-17 cycles) of adaptor-conjugated DNA libraries (B).

Next-gen genomic sequencing:

We submitted the DNA libraries for deep sequencing using Illumina Hi-Seq 2500 sequencer at Penn State Genomic core facility (University Park). We used rapid-run mode with 2 lanes, 4 samples per lane and 100 nucleotides per read. For each sample, Illumina sequencing provides a fastq file and fastqc report. For example, each of 8 fastq files obtained from 8 libraries sequenced in one run on two lanes contained 32-40 mln. of 100 bp - long reads. About 83-93% of all reads were successfully and uniquely aligned to the human genome using Bowtie software.

Human genome mapping and bioinformatic analysis:

For initial genome mapping of primary ChIP-seq reads generated by genome SOLiD, we used NextGENe software (Softgenetics, University Park, PA) specifically adapted for ChIP-sequencing analysis. We used *Alignment* tool to map the ChIP-recovered DNA sequences to the annotated human genome v.37. We used Expression report tool (in most cases after fragmenting the genome map into 50 Kb regions) to identify extended blocks of H3K9 methylation on human chromosomes. We also used the *Search* tool to locate genes of interest and *Expression* tool to compare peak intensity between the proliferating and differentiated cells. Cluster analysis was done with open source Gene Cluster 3.0 software[5]with hierarchical clustering by complete linkage and visualized with TreeView version 1.6 (2002).

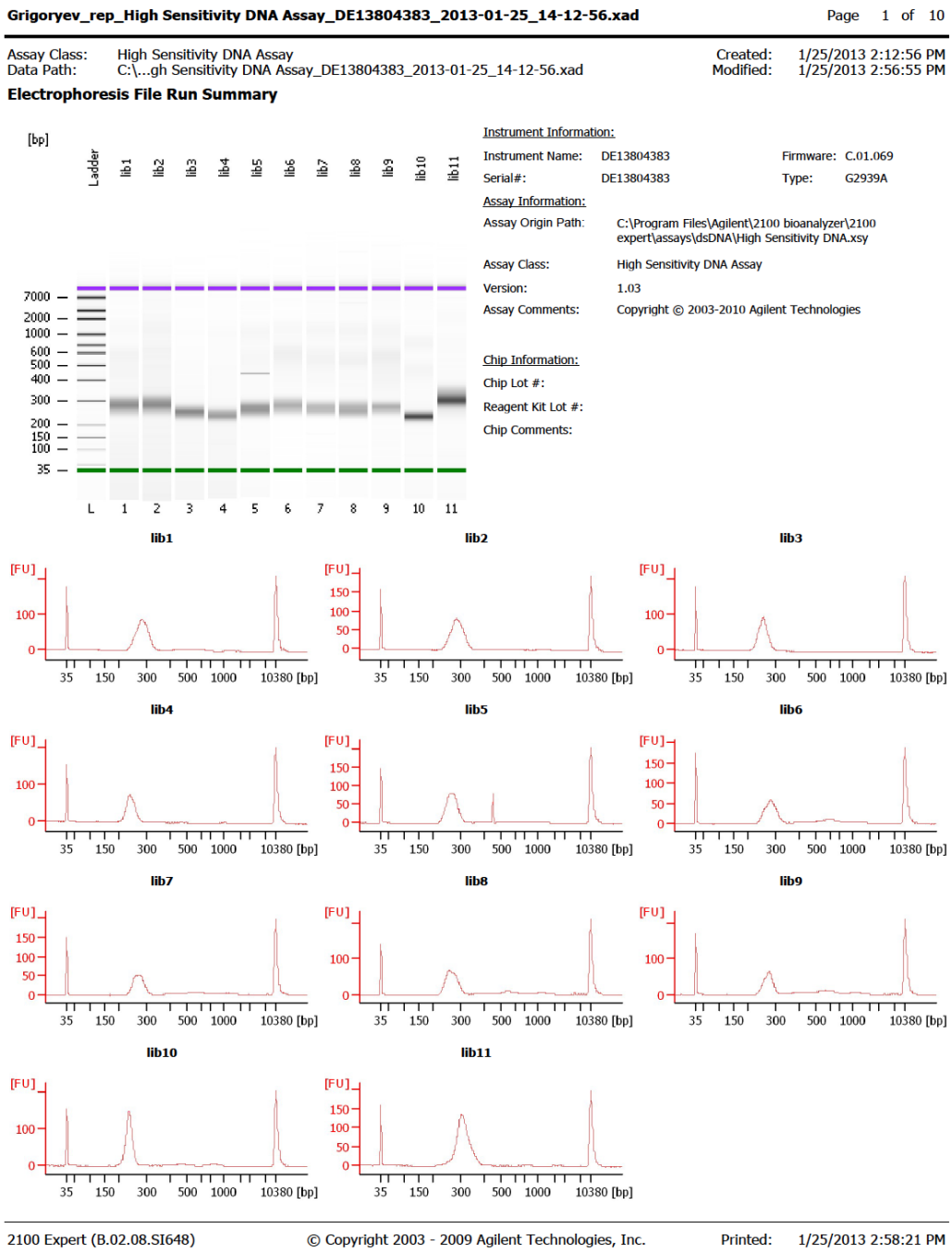

Figure 4. Bioanalyzer traces of PCR-amplified genomic DNA libraries. Note: DNA from samples 1, 2, 5-9 was extracted for 175-225 bp range resulting in library peaks at 275 bp. Samples 3, 4, 10 were extracted for 150-175 bp range resulting in library peaks at 250 bp. Sample 11 was extracted for 225-275 bp range resulting in library peaks at 320 bp.

Bioinformatic analysis was conducted in collaboration with Drs. Arthur Berg and Anna Salzberg of the Penn State Hershey Bioinformatics Core. Sequence reads (omitting repeated sequences) were

normalized to average values and plotted to the reference human genome hg19 using Bowtie program (version 0.12.9). Domains were called using RSEG software, version 0.4.8 RC [6] with provided deadzone file "deadzones-k41-hg19.bed" for the human samples. Other parameters used were: expected domain size of 10,000, keep duplicates. The inputs were the bowtie mappings except for the male samples whereby the X and Y mappings were duplicated. RSEG-derived enriched HMM domains were mapped into windows of size 10k, 50k, 100k, 500k and 1M. The score contribution of each domain for a window was set to the length of the domain falling within that window multiplied by the average read count for the domain as output by RSEG. Window scores were then normalized by the average score of all windows. Pearson correlations were calculated using R 2.15.1. For 10K windows every 100th window was extracted, for 50K windows every 20th window, for 100K every 10th, for 500k every 2nd and for 1M every window was used. In this way every window used in the correlation starts at multiples of 1M.

Differential Histone H3K9 methylation domains (DHKDs) were calculated by adding log ratios of each of the 10k window score for an AML sample to that of a control sample for every possible AML and control pair. These values were normalized by the number of pairwise terms in the sum (i.e. the number of control samples times the number of AML samples). The Y chromosome 10k window score only included pairwise terms corresponding to the male control and AML samples (and was normalized accordingly), and the X chromosome score was set to the sum of separately computed female and male such normalized scores. The threshold of significant positive DHKDs was set to the 99 percentile, and that of significant negative DHKDs to the 1 percentile. The latter calculation was performed using the 'quantile' function in R, with alpha = 0.99 and 0.01 respectively. Six different DHKD and DHKD sets were calculated: Granulocyte > AML (positive); AML > Granulocyte (negative); CD34+ > AML (positive); AML > CD34+ (negative); CD34+ > Granulocyte (positive), and Granulocyte > C34+ (negative).

Pathway, Network, and Gene Ontology Analyses:
These were conducted using Ingenuity Pathways, build version 212183 including the most enriched canonical pathways, networks, diseases and disorders, molecular and cellular functions, physiological system development and functions, and toxicity lists. It also returned the top upstream regulators. Ingenuity Pathway was used with the following parameters: reference set to Ingenuity Knowledge Base Genes Only, relationships to include set to direct and indirect, includes endogenous chemicals, and filter set to only consider molecules or relationships for the human species and where the confidence is set to experimentally observed. For each hypo and hyper DHKD set described above, the genes of the top 1%, 5% and 50% scoring DHKDs were extracted. An Ingenuity core analysis was run for each resulting gene list.

Work on Specific Aim 1: To map genome-wide topography of histone H3(K9) dimethylation and its transitions in the process of normal myeloid differentiation:

To study epigenetic transitions during myeloid differentiation, we isolated chromatin from normal granulocytes and normal CD34+ hematopoietic progenitors. These chromatin samples were immunoprecipitated with specific antibodies against H3K9me2 (Abcam, ChIP Grade, ab1220) and the immunoprecipitated and input DNA samples were isolated. We used the input DNA and H3K9me2-associated DNA from differentiated granulocytes and undifferentiated CD34+ cells to prepare DNA libraries for genomic sequencing using either Applied Biosystems SOLiD system or

New England Biolabs reagents for Illumina. 100 ng of DNA was used to prepare one library for SOLiD genomic sequencing using Applied Biosystems reagents. Each sample was individually barcoded; the barcoded libraries were combined and sequenced on one sequencing glass slide at a bead density allowing us to read $20x10^6$ 70 bp sequence reads per sample. Alternatively just 10-30 ng DNA were converted to a barcoded library and sequenced on Illumina 2500 Hi-seq genome sequencer at Penn State Genomics Core Facility allowing to read approximately $30x10^6$ 100 bp sequence reads per sample.

By mapping H3K9me2 over human chromosomes we were able to make several novel and specific observations for differentiating myeloid cells. We found that both the input DNA (Fig.5, panel 1) and H3K9me2 (panel 2) reads were distributed evenly among all chromosomes with the exception of the haploid X and Y chromosomes in males. Distribution of H3K9me2 was strikingly similar over the chromosomes of the primary granulocytes (Fig. 5, panels 3, 4) and CD34+ cells and (panel 5). This shows that the global, genome-wide distribution of H3K9me2 is conserved during myeloid differentiation including the low levels zones of H3K9me2 at the gene-dense regions and pronounced blocks of high H3K9me2 associated primarily with gene deserts depleted of protein-coding sequences (gene positions are shown below the panel 6). As an example, one extended region of increased H3K9me2 is prominent in the ChIP samples (panels 4, 5) but absent in the input DNA (panel 6). Similar large organized chromatin K9 dimethylation domains (LOCKs) were previously described for mouse ES and human placenta cells [7]. We also noted that the majority of the H3K9me2 domains that we mapped correspond to lamina-associated domains or LADs in non-myeloid human cells [8] and to regions depleted of protein-coding sequences (Annotated CDS) – see Fig. 6 consistent with a conserved H3K9me2 LOCK association with gene deserts and peripheral facultative heterochromatin.

As negative controls, we used antibodies against histone H3K4me2, an activatory epigenetic mark, and H3 C-tail, a non-modified histone epitope. Analysis of genome-wide correlation using windows from 10K to 1M confirmed a strong positive correlation between granulocytes, CD34+ cells and K562 H3K9me2 ChIP and clear discrimination from gene density, H3K4me2, and input DNA over a wide range of windows (Figure 7). ChIP against nonmodified histone H3 was practically identical to the input DNA (not shown).

To examine whether the H3K9me2 domains in myeloid cells show any biologically important correlations with known genomic data sets, we conducted correlation analysis with biological data available from UCSC and ENCODE databases by grouping the data in several increasing windows of 10K, 50K, 100K, 500K, and 1M. First, we observed a strong positive correlation between H3K9me2 distribution in myeloid cells and human T-cells (#1 on Fig.8; data from [9]) confirming that this histone modification is mostly conserved across different cell lines. We also observed a strong correlation between the myeloid H3K9me2 domains (LOCKs) and Lamina-associated domains LADs (#3 on Fig.8; data from [8]) as well as with regions of increased somatic cancer mutations (#2 and #4 on Fig. 8; data from [10]) and AML-specific SNV [11] contrasted with no correlation or negative correlations with most other biological variables (#5-37 on Fig.8). Thus we observed that the genome-wide distribution of H3K9me2 LOCKs is negatively correlated with gene density, is conserved within the human genome, and does not undergo a global remodeling during terminal differentiation.

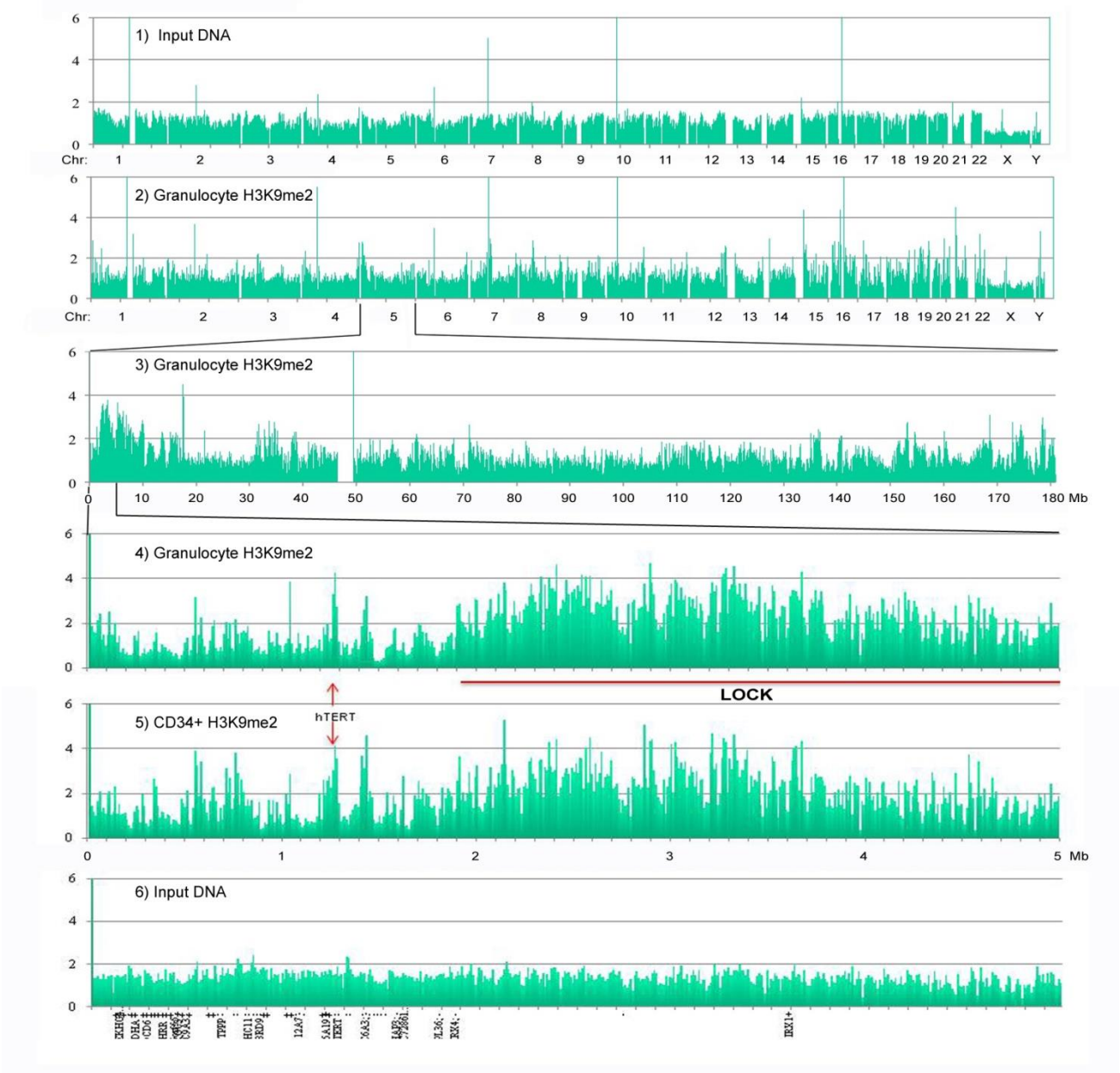

Figure 5. Mapping of input and ChIP-recovered sequence reads to the human reference genome hg19 at different resolution. 1, 2: sequence reads were derived from the input DNA (1) and H3K9me2 (2) ChIP-seq of male granulocytes and mapped to the whole genome at 1M windows. 3: sequence reads were derived from ChIP-seq of H3K9me2 in human granulocytes and mapped to chromosome 5 at 50K windows. 4-6: sequence reads were derived from ChIP-seq of H3K9me2 in human granulocytes (4), CD34+ progenitors (5) and input DNA of human male granulocytes (6) and mapped at 10K windows. Note a continuous enrichment of H3K9me2 over a large organized domain of H3K9 dimethylation (LOCK) [7] indicated by the red line. Red arrows indicate the position of hTERT gene associated with a peal of H3K9me2.

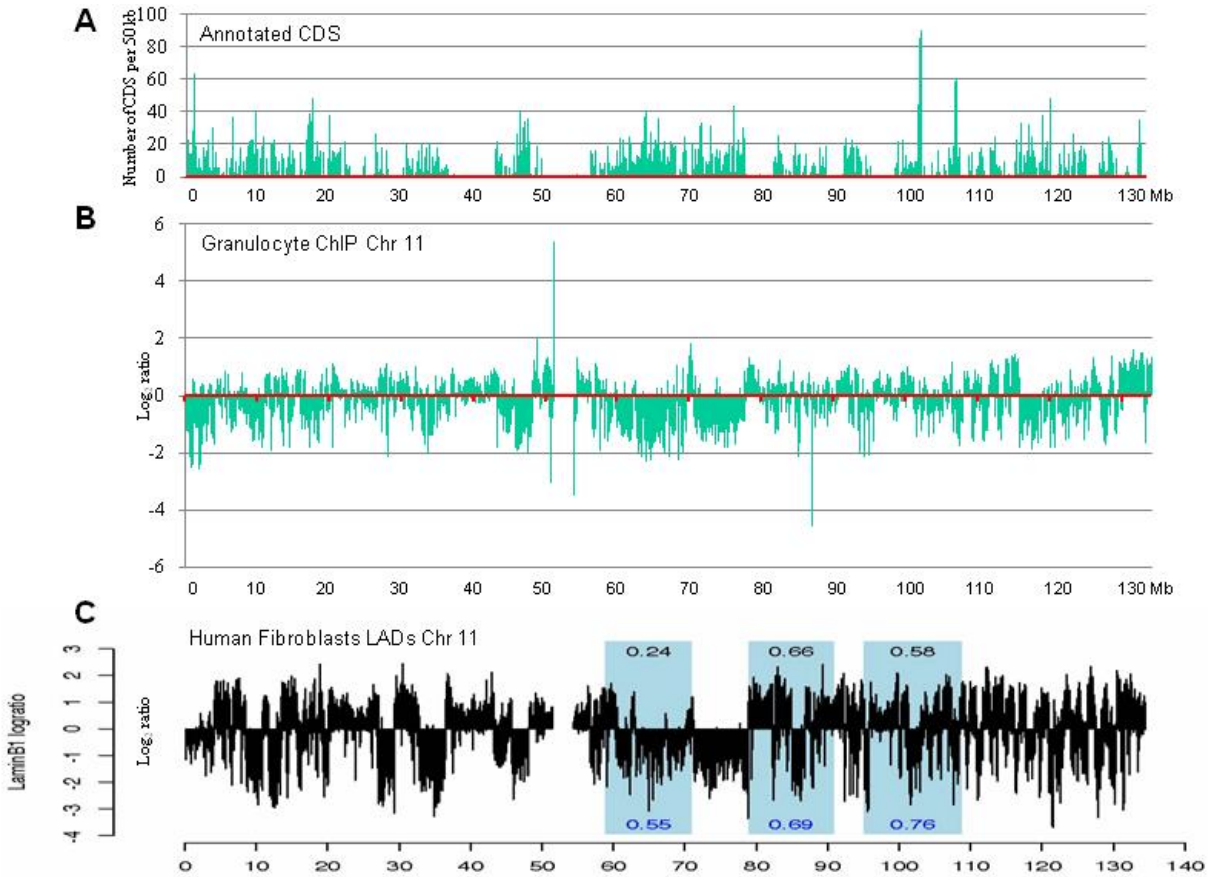

Figure 6. A: map of annotated coding sequences (CDS) over human chromosome 11 derived from NextGene human genome map v. 37_1. Y-axis shows the number of CDS per 50 Kb segments (green bars). B: map of human granulocyte sequence reads derived from H3K9me2 ChIP-seq and grouped by 50 Kb segments (green bars) over human chromosome 11. Sequence reads were normalized to average and plotted in Log_2 scale. C: Sequence reads were derived from mapping of nuclear lamin-associated domains (LADs) over human chromosome 11 taken from [8].

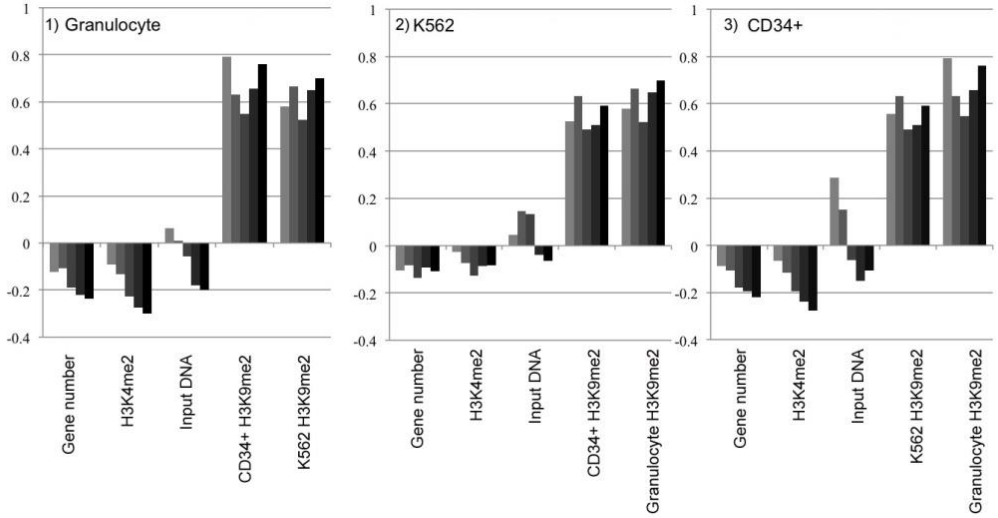

Figure 7. Graphs showing correlations between female granulocyte (1), K562 (2) and CD34+ (3) H3K9me2 ChIP vs. number of genes (gene density), H3K4me2 (active chromatin), input DNA, and H3K9me2 from the two other samples as indicated. The five columns in each sample show correlation coefficients calculated

with 10K, 50K, 100K, 500K, and 1M windows (left to right).

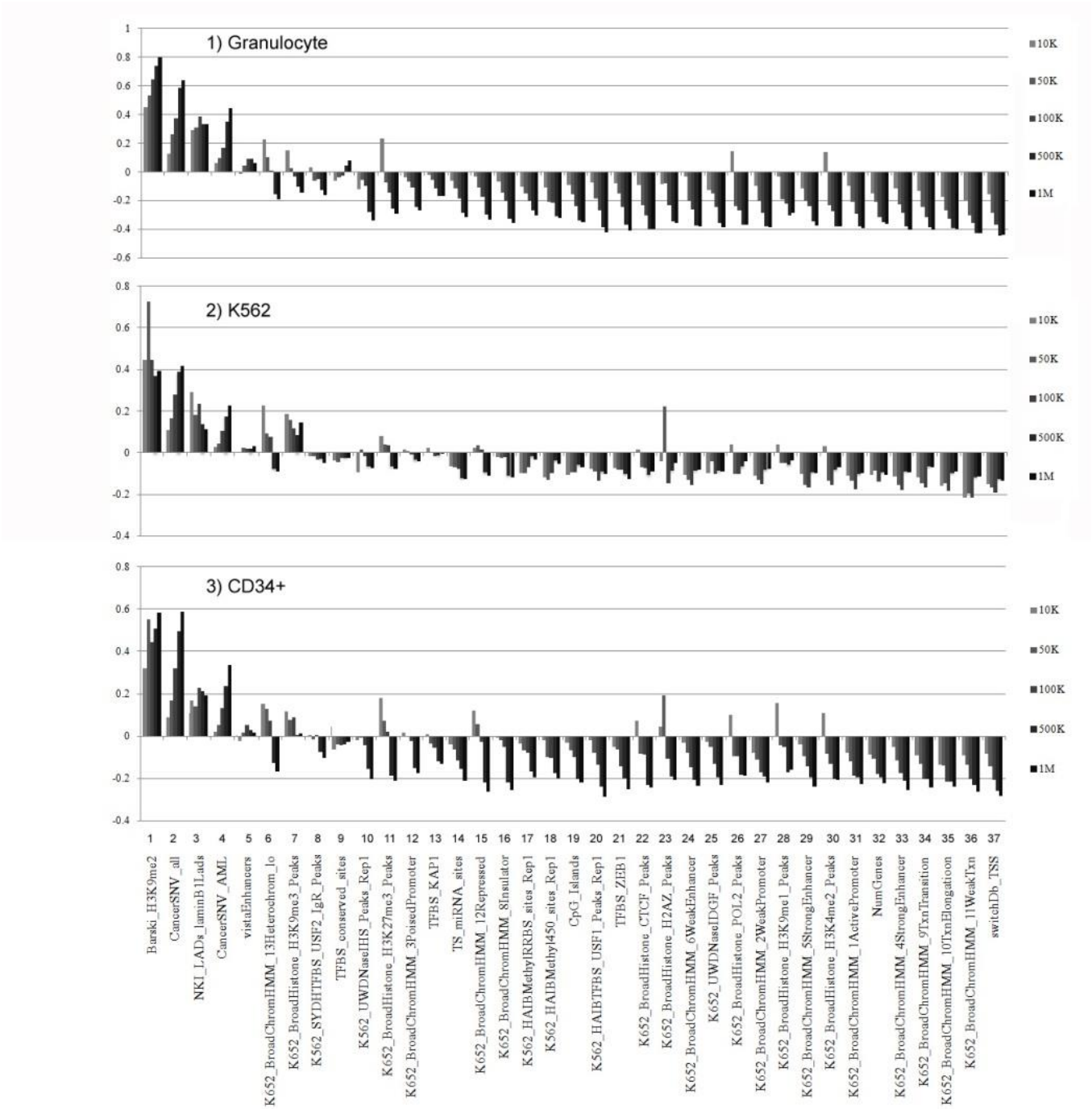

Figure 8. Graphs showing correlations between granulocyte (1), K562 (2), and CD34+ (3) H3K9me2 ChIP-seq reads vs. biological data entries (listed at the bottom) from UCSC database at 10K, 50K, 100K, 500K and 1M resolution.

Work on Specific Aim 2: Spatial and functional relationship between histone H3(K9) methylation and transcriptional changes in AML.

AML-derived myeloid progenitors: We have conducted ChIP-sequencing with H3K9me2 antibodies on 10 samples of AML-derived myeloid progenitors (predominantly myeloblasts) that have been previously isolated from AML human bone-marrow samples and cryopreserved at Claxton laboratory. We mapped the resulting sequence reads against the reference human genome Hg19. Reads mapping showed that the AML samples display the same pattern of conserved LOCKs of H3K9me2 distribution. In addition to the conserved LOCKs, we observed cell-specific domains of H3K9me2 that showed a dramatic difference between differentiated granulocytes and undifferentiated AML myeloblasts. One striking example is the changes in the H3K9me2 profile associated with the three chromosomes 19 loci: 19q13.12, 19q13.41, and 19q13.42 (shown by red arrows on Fig. 9, panels A,B). Each of these loci contains dozens of ZNF genes. ZNF is the largest family of transcription regulation genes that had been rapidly changed in human evolution. A sharp increase of heterochromatin levels suggests dramatic and coordinated change in expression of multiple ZNF genes during myeloid differentiation.

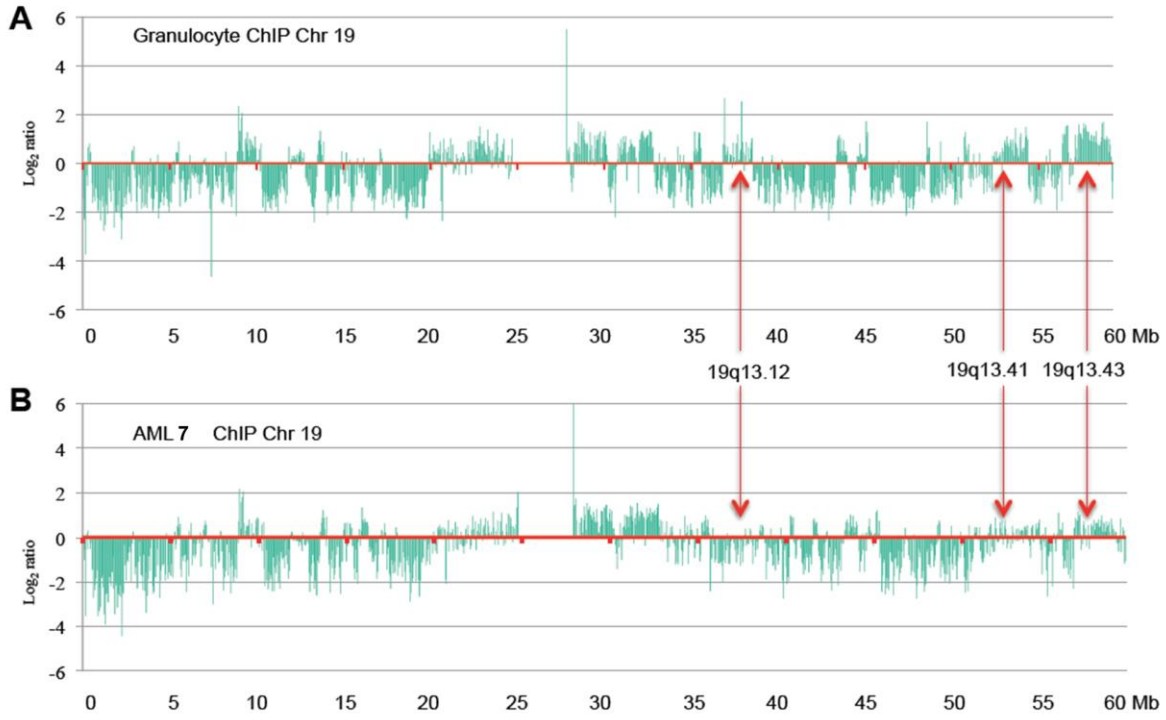

Figure 9. A,B: maps of sequence reads derived from H3K9me2 ChIP-seq of human granulocytes (A) and AML150 myeloblasts (B) grouped by 50 Kb segments (green bars) over human chromosome 19. Sequence reads were normalized to average and plotted in Log2 scale. Red arrows show difference in H3K9me2 levels at the chromosome 19 loci 19q13.12, 19q13.41, and 19q13.42.

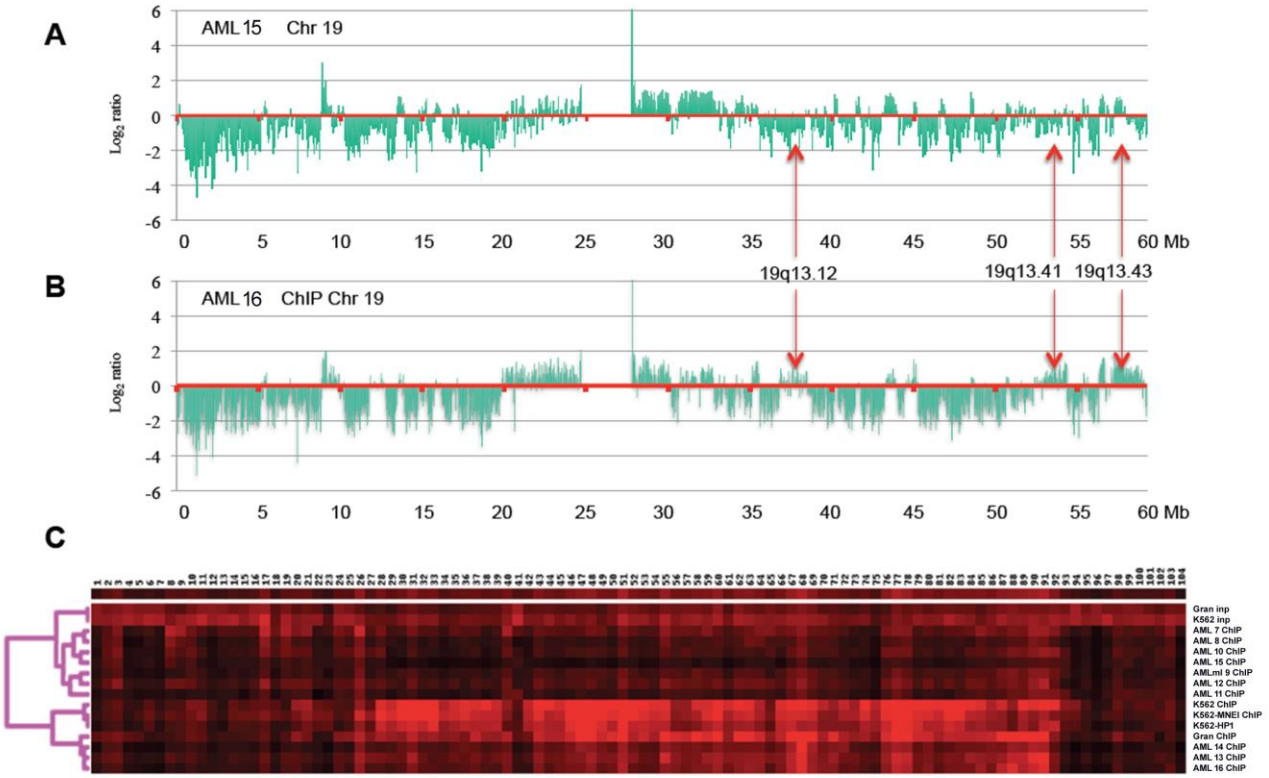

Figure 10: maps of sequence reads derived from H3K9me2 ChIP-seq of AML 472 (A) and AML 482 (B) myeloblasts grouped by 50 Kb segments (green bars) over human chromosome 19. C: H3K9me2 peak distribution over the 19.q13.41 locus from 16 ChIP-seq samples was subjected to cluster analysis using Gene Cluster 3.0 software with hierarchical clustering by complete linkage and visualized with TreeView version 1.6 (2002)

Also notable was a strong variability at the ZNF cluster between individual AML cases such as those between AML 472 and AML 482 (Fig 10, cf. panels A and B). A hierarchical cluster analysis of H3K9me2 of one such locus, 19.q13.41, across 10 AML samples as well as granulocytes and K562 cells allowed us to distinguish two groups of AML: group I including 7 AMLs with maximal changes from the most differentiated cells and group II including 3 AML with H3K9me2 distribution similar to that in most differentiated cells.

To determine precise size distribution and boundaries of large H3K9me2 domains and reveal their quantitative change associated with AML, we used RSEG, an unbiased procedure for selecting domain boundaries based on revealing hidden Markov model (HMM) in genome-wide ChIP-seq read distribution [6].We took ChIP-seq data and calculated HMM with subtracted deadzones for 10 human AML samples and examined genome-wide correlations for all samples in a range of windows between 10K and 1M. With all windows, we observed strong genome-wide correlation between AML and the “normal” human samples – granulocytes and CD34+ progenitors as well as cultured K562 cells (Figure 11).

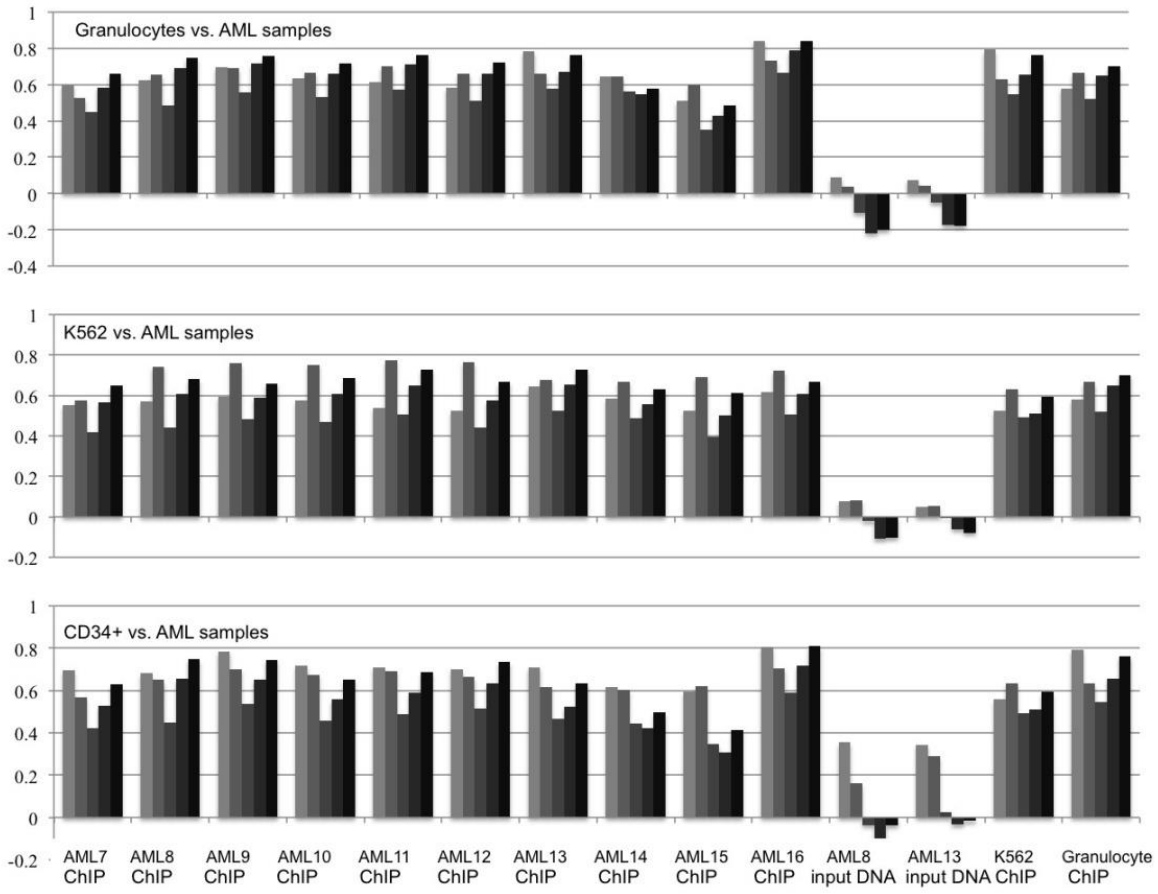

Figure 11. Graphs showing correlations between H3K9me2 in granulocytes (top) K562 (center), and CD34+ (bottom) H3K9me2 and 10 AML samples each at 10K, 50K, 100K, 500K, and 1M resolution (left to right).

Despite the overall conservation, high-resolution mapping of HMM domains of H3K9me2 HMM domains between the normal and cancer samples confirmed that some chromosomal domains such as seen on chromosome 19 (see Fig. 9) had strong variations between the normal and AML samples (Figure 12, cf. chromosome 19 from granulocytes (panel 1) and CD34+ progenitors (panel 2) vs. AML#8 (panel 3)). To range all H3K9me2 domains by the "weight" of their developmentally-regulated and AML-related changes, we grouped all AML domains (samples 7-16) and calculated the Differential H3K9 dimethylation domains (DHKDs) between 6 comparison categories: Granulocyte > AML; AML > Granulocyte; CD34+ > AML; AML > CD34+; Granulocyte > C34+ and CD34+ > Granulocyte. Significance of the observed change in each given domain was calculated as "quantile" function in R, with alpha representing the probability of a given change to be non-random. Within each of the six categories, each DHKD is given a number depending on the Sum Log Ratios. The DHKDs with highest Sum Log Ratio changes ($\alpha < 0.0001$ and $\alpha > 0.9999$) are shown in Table 2.

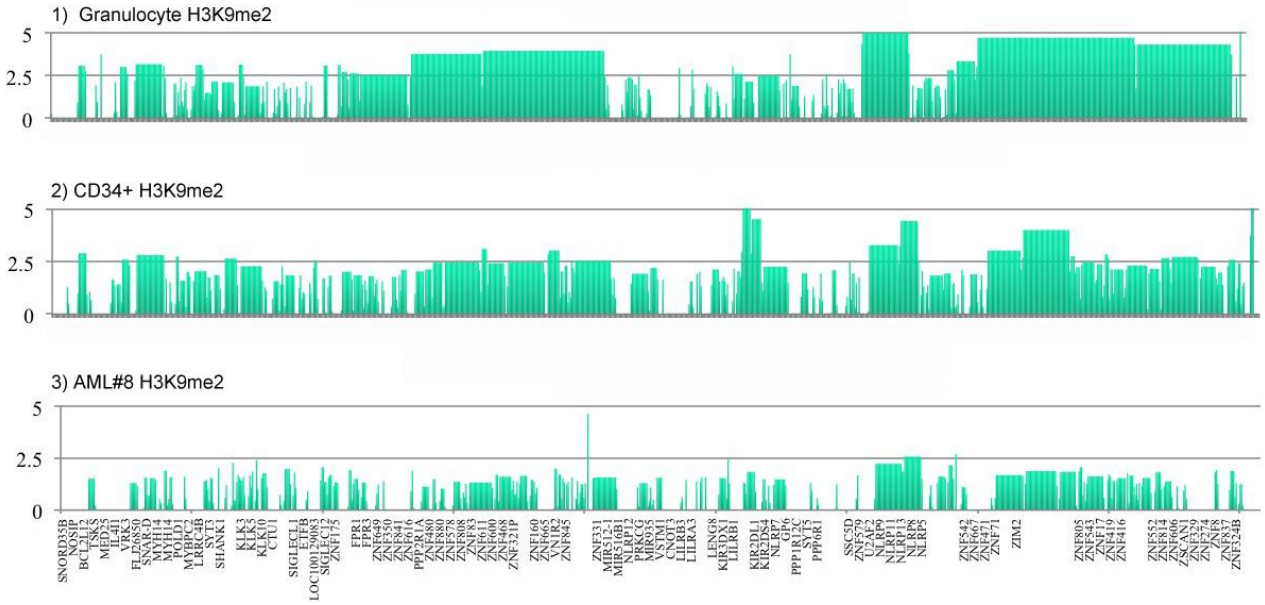

Figure 12: High resolution maps (10 K windows) of normalized HMM domains on human chromosome 19 locus between 50 and 59 Mb containing two AML-hypomethylated H3K9me2 DHKDs at 19q13.42 and 19q.13.43. HMM domains were derived from H3K9me2 ChIP-Seq of granulocytes (1), CD34+ cells (2) and AML8 (3).

Remarkably, the DHKDs most strongly hypomethylated in AML (Granulocyte > AML and CD34+ > AML) contained multiple genes (Table 2) in contrast to the gene-poor conserved H3K9me2 LOCKs. The most prominent among those genes are clusters containing repeats of ZNF (KRAB-ZFP) genes on chromosome 19 already noted above (Fig. 10). Also of interest, the AML - hypomethylated gene clusters also included repeats of small nucleolar RNA SNORD116 and SNORD 115 on chromosome 15q11-q13 known as an imprinted chromatin locus whose deletion is associated with Prader-Willy syndrome [12] and a number of cadherin and procadherin genes previously shown to be coordinately regulated by H3K9me2-specific histone methyltransferase G9a in epithelial to mesenchymal transition [13] (not shown).

Category	#	locus	Start	End	SumMinusLog	alpha	Gene number	1st Gene
Gran > AML	1	19q13.43	58010001	59050000	20.61507737	> 0.99999	47	ZNF211
	2	19q13.41	52740001	53770000	17.64917002	> 0.9999	29	ZNF321P
	3	19q13.12	37040001	37760000	11.99553525	> 0.9999	15	ZNF420
	4	19p13.2	11870001	12770000	11.0460077	> 0.9999	25	MAN2B1
	5	16p13.3	840001	1330000	9.147934813	>0.999	13	CACNA1H
AML > Gran	1	10q22.2	77170001	77810000	-7.534063671	< 0.00001	1	C10orf11
	2	5q35.1	169130001	169640000	-6.271680829	< 0.0001	3	DOCK2
	3	7p14.2	36730001	37210000	-5.233786436	< 0.0001	3	AOAH
	4	8p21.3	19270001	19570000	-5.113040625	< 0.0001	1	CSGALNACT1
	5	15q21.1	48690001	48930000	-4.504764962	< 0.0001	1	FBN1
CD34+ > AML	1	16p13.3	840001	1330000	16.19076483	> 0.99999	13	CACNA1H
	2	21q22.3	45770001	46190000	9.77820114	> 0.9999	9	C21orf90
	3	8q24.3	143130001	143540000	9.507677545	> 0.9999	3	MIR4472-1
	4	Xq28	152630001	153020000	9.083124606	> 0.9999	11	BCAP31
	5	22q13.32	48790001	49120000	7.661235208	> 0.999	2	LOC284933
AML > CD34+	1	1q43	239440001	240060000	-5.007083988	< 0.00001	1	CHRM3
	2	7q11.22	69080001	69810000	-4.541770226	< 0.0001	1	AUTS2
	3	2q14.3	124340001	124800000	-4.183061584	< 0.0001	1	CNTNAP5
	4	19q13.31	43170001	43580000	-4.075585917	< 0.0001	9	PSG11
	5	4q13.2	69370001	69490000	-3.783760139	< 0.0001	1	UGT2B17
CD34+ > Gran	1	5q13.2 duplicated	68830001	70390000	11.10618592	> 0.99999	17	SERF1A
	2	16p13.3	840001	1330000	7.042830024	> 0.9999	13	CACNA1H
	3	10q22.2	77170001	77810000	6.171933861	> 0.99999	1	C10orf11
	4	2q37.3	240260001	240550000	6.022509492	> 0.99999	2	HDAC4
	5	5q35.1	169130001	169640000	5.932457822	> 0.999	3	DOCK2
Gran > CD34+	1	19q13.43	58010001	59050000	-15.62495492	< 0.00001	47	ZNF211
	2	19q13.41	52740001	53770000	-11.95355841	< 0.0001	29	ZNF321P
	3	19q13.12	37040001	37760000	-10.8912562	< 0.0001	15	ZNF420
	4	19p13.2	11870001	12770000	-8.745980609	< 0.0001	25	MAN2B1
	5	19q13.31	43170001	43580000	-7.011448316	< 0.0001	9	PSG11

Table 2: Top-scoring DHKD domain variations (SumMinusLog difference) calculated for 6 comparison categories: Granulocyte > AML; AML > Granulocyte; CD34+ > AML; AML > CD34+; Granulocyte > C34+ and CD34+ > Granulocyte. Only DHKDs with alpha values representing the probability of a given change to be non-random ($\alpha < 0.0001$ and $\alpha > 0.9999$) are shown.

We used Interactive pathway analysis (Ingenuity systems) to identify the annotated top scoring functions of the genes in DHKDs and their upstream regulators. Functional annotations of the genes derived from the AML-hypermethylated DHKDs, AML > Granulocyte and AML >CD34+ (Table 3) revealed a number of top networks including various non-blood cancers, and myeloproliferative disorder that may provide new clues for understanding AML progression.

	Functions Annotation	p-Value	# Molecules
	assembly of synapse	2.04E-12	11
	insulin-dependent diabetes mellitus	5.20E-10	38
Gran > AML	development of gap junctions	1.86E-08	12
	formation of plasma membrane	1.83E-07	14
	assembly of intercellular junctions	5.98E-07	13
	lung adenocarcinoma	1.94E-07	156
	lung cancer	2.08E-07	167
AML > Gran	adenocarcinoma	2.31E-07	184
	carcinoma in lung	3.39E-07	163
	myeloproliferative disorder	7.00E-06	21
	chronic myeloproliferative disorder	1.64E-05	13
	insulin-dependent diabetes mellitus	1.03E-17	52
	assembly of synapse	4.76E-12	11
CD34+ >AML	synaptogenesis	1.22E-09	12
	diabetes mellitus	2.95E-09	68
	development of gap junctions	3.72E-09	13
	lung adenocarcinoma	2.14E-11	107
	adenocarcinoma	2.74E-11	123
	lung cancer	3.22E-11	113
AML > CD34+	carcinoma in lung	3.89E-11	111
	Cancer	3.25E-09	162
	carcinoma	7.69E-09	145
	epithelial neoplasia	9.59E-09	146
	insulin-dependent diabetes mellitus	4.61E-17	42
	diabetes mellitus	1.50E-15	65
CD34+ > Gran	glucose metabolism disorder	1.39E-14	67
	systemic autoimmune syndrome	1.03E-08	77
	arthritis	7.59E-08	73
	rheumatoid arthritis	1.03E-06	59
	adenocarcinoma	4.74E-17	171
	lung adenocarcinoma	6.32E-15	144
	lung cancer	2.90E-14	151
Gran > CD34+	carcinoma in lung	1.20E-13	147
	solid tumor	3.23E-12	198
	carcinoma	3.67E-12	197
	epithelial neoplasia	6.60E-12	198
	Cancer	1.37E-10	214

Table 3: Annotated top scoring biological functions of genes calculated using Interactive Pathway analysis (Ingenuity systems) for 6 comparison DHKD categories: Granulocyte > AML; AML > Granulocyte; CD34+ > AML; AML > CD34+; Granulocyte > C34+; and CD34+ > Granulocyte.

A search for the top upstream regulators revealed a number of regulatory pathways that may be involved in AML progression (Figure 13). The top-scoring regulatory molecules, TRIM28, ERG,

Mdm2 and Twist1 have oncogenic properties and are likely candidates to be directly involved in epigenetic regulation by recruiting histone methyltransferases specific for histone H3K9. TRIM28 (KAP-1), a common co-repressor of KRAB-Zfp genes, is consistent with their striking enrichment in the granulocyte/AML hypermethylated domains, acts by recruiting HMTase SETDB1 [14]. ERG is a transcriptional regulator that may act by recruiting histone methyltransferase SETDB1and subsequent heterochromatization of local chromatin structure [15]. MDM2, another strongly associated regulator is involved in recruitment of an H3K9me2-specific histone methyltransferase EHMT1 [16]. Thus, our analysis of variable H3K9me2 domains reveals gene clusters and regulatory pathways undergoing coordinated epigenetic and gene regulatory changes in AML.

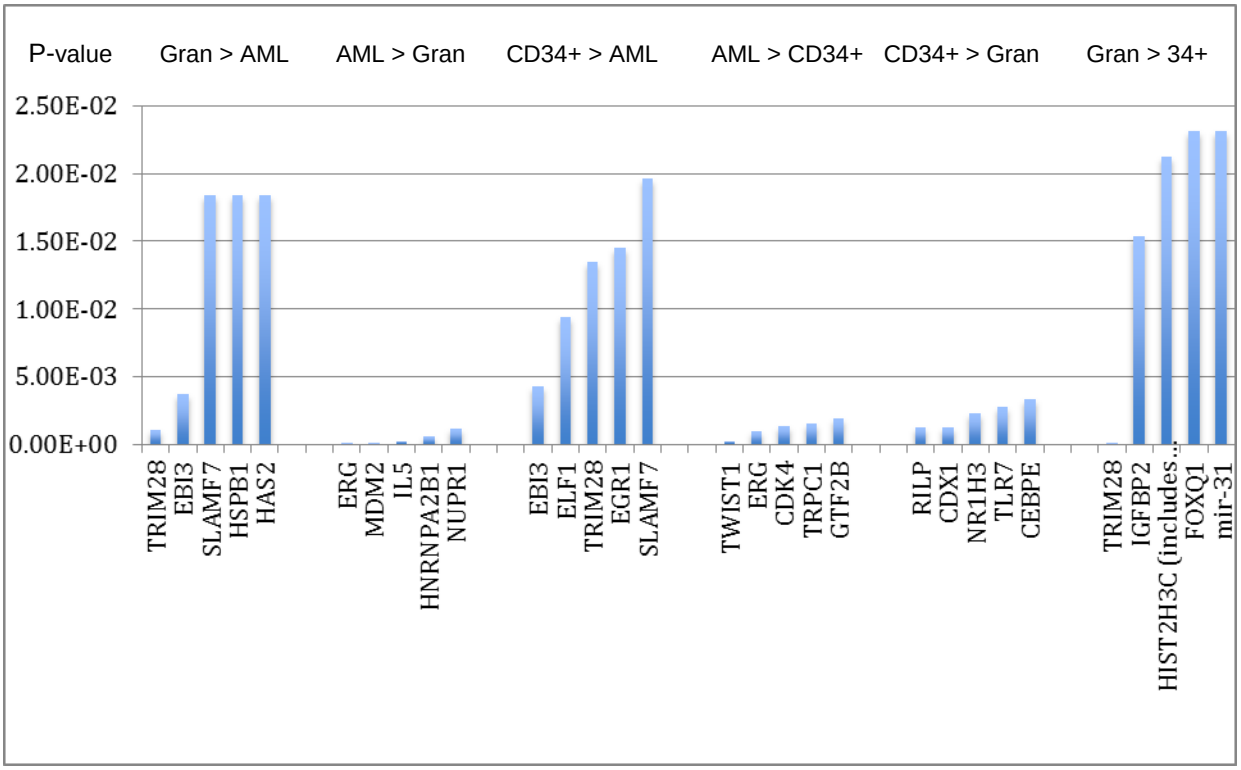

Figure 13. Bar histograms showing top upstream regulators for 6 comparison DHKD categories (Granulocyte > AML; AML > Granulocyte; CD34+ > AML; AML > CD34+; Granulocyte > C34+ and CD34+ > Granulocyte) as determined by Interactive pathway analysis (Ingenuity systems).

From the gene expression array analysis of multiple AML cases including 6 samples taken in our ChIP study and 12 normal CD34+ myeloid progenitors previously obtained by Drs. T. Loughran and D. Claxton, using Illumina HT12 arrays (37084 entries per one sample) we now identified 75 genes associated with AML-specific DHKDs and repressed or activated more than 1.5 fold in AML vs. CD34+ progenitors (Table 4). The majority of these genes (52 out of 75) showed a substantially increased H3K9me2 level associated with gene repression as expected. However, notable were groups of genes, in particular the cluster of ZNF genes associated with chromosome 19 and the cluster of Hist1 genes at chromosome 6 that were downregulated within DHKDs with increased H3K9me2 or upregulated within a DHKD with decreased H3K9me2 (Table 4). These surprising findings require further investigation.

Table 4.

Genes in AML-hypomethylated DHKDs (CD34+ > AML)

chromosome	position	Gene	SumLogDifference	AML/CD34 expression rates
2	71805000	DYSF	3.266189681	1.835618352
2	241455000	ANKMY1	6.133663824	1.786280915
4	875000	GAK	2.478169463	1.626287465
5	140105000	VTRNA1-3	2.418546156	1.545896149
5	140625000	PCDHB19P	4.742640694	1.675358651
5	176315000	HK3	2.504052895	5.777576769
19	12005000	ZNF69	2.300027083	2.298952326
19	12075000	ZNF763	2.300027083	1.674356879
19	12665000	ZNF564	2.300027083	1.575998171
19	53455000	ZNF816	5.695611609	1.573378504
19	55385000	FCAR	1.97999094	3.176177577
19	56475000	NLRP8	1.917174437	1.770077291
19	59035000	ZBTB45	4.990122453	1.610872034
X	48455000	WDR13	3.218943575	1.760683495
X	48545000	WAS	3.218943575	2.07675891
X	48775000	PIM2	6.223786135	1.900222499
X	48845000	GRIPAP1	6.223786135	2.10603869
X	70905000	BCYRN1	1.864161849	1.924835702
X	128925000	SASH3	2.017764943	1.573647715
X	153655000	ATP6AP1	5.895029854	2.706394075
X	153665000	GDI1	5.895029854	1.593067587
X	153675000	FAM50A	5.895029854	2.593944984
X	153695000	PLXNA3	5.895029854	1.570722774
X	153715000	UBL4A	5.895029854	1.904889265
X	153715000	SLC10A3	5.895029854	1.564576192
X	153735000	FAM3A	5.895029854	1.604088045
X	153755000	G6PD	5.895029854	2.570187981
X	153785000	IKBKG	2.902091753	1.564780011
Total upregulated genes:		**28**		

3	134795000	EPHB1	1.872491135	0.648235478
4	6285000	WFS1	1.848799328	0.415098098
9	137555000	COL5A1	3.232343929	0.273390799
19	11975000	ZNF439	2.300027083	0.621223675
19	53135000	ZNF83	5.695611609	0.469025344
19	58105000	ZIK1	4.990122453	0.551863275
19	58215000	ZNF154	4.990122453	0.608927527
19	58395000	ZNF814	4.990122453	0.539373985
19	58435000	ZNF418	4.990122453	0.613336974
19	58455000	ZNF256	4.990122453	0.373663846
19	58585000	ZNF135	4.990122453	0.542639517
19	58745000	ZNF544	4.990122453	0.490755838
22	50015000	C22orf34	2.623198864	0.592884804
X	48645000	GATA1	6.223786135	0.597438753
Total repressed genes:		**13**		

Genes in AML-hypermethylated DHKDs (AML > CD34+)

chromosome	position	Gene	SumLogDifference	AML/CD34 expression rates
3	169035000	MECOM	-1.841388833	1.758342937
6	26195000	HIST1H3D	-1.142499484	1.551816191
6	26195000	HIST1H2BF	-1.142499484	2.314648532
6	26215000	HIST1H2BG	-1.142499484	3.230242296
6	26215000	HIST1H2AE	-1.142499484	1.865670515
6	26255000	HIST1H3F	-1.142499484	1.560598541
7	36745000	AOAH	-1.53249867	2.358272161
7	102565000	LRRC17	-1.381348066	1.585665285
18	22805000	ZNF521	-2.338757035	2.384641796
18	25645000	CDH2	-1.355088106	1.755400083
Total upregulated genes:		**10**		

1	61755000	NFIA	-1.456658642	0.456783868
1	184465000	C1orf21	-1.245269798	0.601930774
2	73955000	TPRKB	-1.070741864	0.575612568
2	192705000	SDPR	-1.014878433	0.504098851
3	21605000	ZNF385D	-1.54852208	0.646999048
3	66455000	LRIG1	-1.162432387	0.272588242
3	168065000	EGFEM1P	-1.836154241	0.613939161
4	5485000	STK32B	-1.02792126	0.525222121
4	82045000	PRKG2	-1.309667528	0.649280121
4	146745000	ZNF827	-0.93334986	0.528446358
6	90915000	BACH2	-1.230115154	0.626312745
6	138195000	TNFAIP3	-1.118625305	0.437656642
7	110415000	IMMP2L	-1.689414537	0.62370952
8	131295000	ASAP1	-1.015049847	0.659774789
9	25675000	TUSC1	-1.339576901	0.565471346
9	109635000	ZNF462	-1.070555699	0.639947745
12	94675000	PLXNC1	-1.203391268	0.644680419
12	94715000	CCDC41	-1.203391268	0.62892353
13	41195000	FOXO1	-0.934972525	0.515958315
14	59755000	DAAM1	-0.996627001	0.604371503
15	95005000	MCTP2	-1.328703592	0.553725942
18	5435000	EPB41L3	-1.133282725	0.557972627
21	27875000	CYYR1	-2.176357587	0.390900387
X	114845000	PLS3	-1.103971491	0.648752249
Total repressed genes:		**24**		

Table 4. Ratios of mRNA expressed in AML (average of 6 AML samples) vs. normal bone marrow myeloid progenitors (CD34+) for 75 most up-regulated and repressed genes (more than 1.5 times) and Sum of Log difference between H3K9me2 levels for the same genes in CD34+>AML DHKDs (top 41 gene) and AML>CD34+ DHKDs (bottom 34 genes) also showing their chromosomal positions in the reference human genome hg19.

Discussion of the work completed.

We have completed the planned biochemical experiments and ChIP-sequencing experiments outlined in the Strategic Plan and, using extensive bioinformatics analysis, were able to identify significant gene-specific changes associated with AML. Specifically, we mapped extended epigenomic domains of histone H3K9me2 dimethylation similar to the large organized chromatin K9 modifications (LOCKs) previously described in non-myeloid cells [7], and uncovered myeloid-specific variable H3K9me2 domains (DHKDs) altered in AML.

Histone H3 lysine 9 methylation is involved in repression of multiple genes associated with acute myeloid leukemia[18-20]. During normal cell differentiation, H3K9me2 is significantly increased in its levels and spreads over repressed chromosomal domains as shown for differentiating mouse lymphocytes [21] and stem cells [7]. In cancer, disregulated H3K9me2 domains are associated with growth advantage for the affected cells [22]. Previous studies have shown functionally important genome-wide association between H3K9me2 domains in non-myeloid human and mouse cells [7] and Lamin-Associated domains or LADs [8, 23]. When we mapped and compared the myeloid H3K9me2 domains with human LADs, we observed a strong positive correlation (Fig. 8). Also striking was the correlation of the myeloid H3K9me2 domains with somatic cancer mutations [10] and AML-specific SNV [11]. Remarkably, H3K9me2 was significantly stronger correlated with LADs than the two other repressive heterochromatin marks, H3K9me3 and H3K27me2 in the human genome database. Correlation of H3K9me2 within myeloid and non-myeloid tissues was also dramatically stronger than that between H3K9me2 vs. H3K9me3 and H3K27me2. The H3K9me2 domains in myeloid cells are thus distinct from other repressive heterochromatin markers and may play a unique role in developmental regulation of the underlying transcription units and the DNA repair process. We concluded that in myeloid cells the H3K9me2 domains form a conserved set of large organized chromosomal domains (LOCKs) whose association with gene deserts, the nuclear periphery, and regions of increased occurrence of SNV reflecting their condensed state relatively inaccessible to transcription and DNA repair, is consistent with a heterochromatin state.

We observed that between the normal and differentiated myeloid cells, certain H3K9me2 domains (DHKDs) may expand, retract, or occupy new chromosomal positions. Spreading of H3(K9) methylation and heterochromatin is controlled by an epigenetic mechanism known as the histone code [24] that involves a positive feedback loop formed by an HMTase (histone code writer), HP1 protein (histone code reader), and histone H3 methylation (the code). Actively transcribed genes are protected from H3(K9) methylation by special boundary elements [25]. Our experiments had resulted in identifying sets of constitutive and variable LOCKs and the boundary regions separating them from the active chromosomal domains that contain no H3K9me2 domains. This opens a way for detailed analysis of not only the regulatory mechanism that affect the expression of DHKDs but also the boundary DNA regions that limit the domain spreading across the chromosomes including the presence of the insulator elements and insulating protein factors such as CTCF and USF1/2 [26]. In collaboration with the Penn State Hershey Bioinformatics Core, we are developing more combinatorial computational methods for precise mapping of the extended epigenomic domain boundaries revealed by our ChIP-seq data and comparing the boundaries of these domains with whole genome and leukemia-associated changes in transcription.

In the course of this work, we also mapped the chromosomal positions of H3K9me2 domains in K562 cells by overexpressing two nuclear proteins, ubiquitous heterochromatin factors, HP1 and granulocyte-specific MNEI that we previously suggested to replace HP1 during terminal myeloid differentiation [1]. We observed multiple H3K9me2 domains affected by overexpressing either HP1 or MNEI or both (data not shown). These experiments are consistent with the proposed role of histone code reader proteins, the general heterochromatin factor HP1, and myeloid-specific nuclear protein MNEI in regulating heterochromatin spreading in myeloid cells as we suggested previously [1]. Future work may show if the altered levels of HP1 and MNEI could affect H3K9me2 and gene expression in primary AML cells.

In our previous works, we found that H3K9me2 modification marks the condensed facultative heterochromatin that forms during normal cell differentiation and is relaxed in proliferating and undifferentiated cells [1, 28, 29]. Here we used ChIP-sequencing to conduct a detailed whole-genomic characterization of the localization and expression of a H3K9me2 and compare its genome-wide organization in normal and AML-derived myeloid cells. Our work showed that on the global scale, H3K9me2 LOCKs are conserved so that the changes in the structural organization in chromatin do not involve a massive H3K9me2 relocation. However, we were able to reveal multiple local changes in H3K9me2 and their association of with transcriptional changes in leukemia. This includes 75 genes associated with DHKDs whose transcription is either activated or repressed in AML (Table 4).

Our identification of the transcriptional changes associated with altered H3K9me2 domains in AML leads to a question if there is a common regulatory mechanism responsible for those changes. Most of H3K9me2 modification in mammalian cells is mediated by Histone Methyltransferase G9a/GLP that is essential for mouse embryo development [30] and involved in human hematopoiesis [17] and gene deregulation in AML [20]. Through interactions of G9a with DNA methyltransferases [31], H3K9me2 is linked to DNA methylation, another very important epigenetic marker [32] deregulated in AML [33]. It is most likely that the majority of the genes downregulated in DHKDs are repressed by G9a/GLP, the principal HMTase regulating LOCK and LAD stability in the nucleus [34] while the up-regulated genes may be activated by some demethylases removing the H3K9 dimethylation such as KDM3B that has been shown to mediate transcriptional activation in myeloid leukemia [19]. Our new data implies that genome-wide positioning of G9a and KDM3B should be mapped in the normal and AML cells to detect any changes in the distribution of these epigenetic factors on chromosomes of leukemia cells. Interestingly, we observed that a number of ZNF genes on the chromosome 19q13.43 associated with one of the strongest DHKDs undergoes the most significant drop in H3K9me2 (fig. 9) and, at the same time, a coordinated downregulation in AML despite their strongly decreased H3K9me2 level. This is consistent with a recent finding that the ZNF genes at this chromosomal locus are indeed upregulated rather downregulated by G9a and H3K9 dimethylation in normal hematopoietic progenitors[17]. The surprising mechanism of G9a-dependent activation of genes on a selected chromosome locus by otherwise repressive epigenetic factors in AML remains to be investigated.

Recent progress in understanding chromatin remodeling and gene function provided new tools for “epigenetic therapy” of acute myeloid leukemia such as inhibitors of DNA methylation 5-azacytidine and decitabine [35, 36], inhibitors of histone deacetylases (HDACs) [37, 38], and inhibitors of histone demethylase LSD1 [39]. These histone modifying agents do not selectively target specific genes but may affect all genes regulated by epigenetic mechanisms. Therefore the nonspecific epigenetic drugs may interfere with normal activity and impair differentiation in hematopoietic cells thus inflicting potential side effects. For example, in a previous work we showed that HDAC inhibitors negatively affect terminal differentiation of erythroid cells [40]. Among the 75 genes that we found to be associated with both epigenetic and transcriptional deregulation in AML (Table 4) there is a number of genes directly involved in myeloid proliferation and leukemia such as MECOM (Evi1), FOXO1, and GATA1 and a number of other genes (in particular clusters of ZNF transcriptional regulators) whose involvement in myeloid cell growth and differentiation should be further investigated. Importantly, we observed that changes in the

repressive heterochromatin mark might have caused both activation and repression of underlying genes. To make the "epigenetic therapy" of AML and other myeloid disorders most efficient in future, we should proceed to studying the role of particular epigenetic factors (HMTases) that affect genes expression at specific chromosomal locations in AML and use the whole-genome epigenetic mapping as a versatile tool to monitor the efficiency of the epigenetic therapies of various forms of leukemia.

Conclusions:

We used ChIP-sequencing to map this histone modification over the genomes of normal myeloid precursors and differentiated granulocytes and identified large chromosomal domains of increased H3(K9) dimethylation (LOCKs) that are conserved through normal myeloid differentiation and are correlated with conserved features of repressed chromatin in the human genome.

By differential mapping of whole-genomic H3K9me2 profiles between granulocytes and myeloid progenitors, we identified differential H3K9 domains (DHKDs) ranging from ~10,000 to more than 1,000,000 bp that contain pronounced developmentally-regulated epigenetic changes.

By using comparing the whole-genomic H3K9me2 profiles with differential expression gene array data, we determined spatial and functional relationship between histone H3(K9) methylation topographies and transcriptional changes in AML.

By using genomic and bioinformatic approaches, we integrated multiple data from whole-genome mapping of a histone modification, comprehensive gene expression array, and gene function and pathway annotation analysis to identify genes associated with DHKDs including those whose expression is associated with acute myeloid leukemia progression (MECOM, FOXO1, and GATA1) and predicted several upstream regulators that are the candidates for mediating the combined epigenetic and transcriptional changes (ERG, TRIM28, Twist1 and MGM2).

References:

1. Popova, E.Y., et al., *Epigenetic heterochromatin markers distinguish terminally differentiated leukocytes from incompletely differentiated leukemia cells in human blood.* Exp Hematol, 2006. **34**(4): p. 453-62.
2. Nekrasov, M., et al., *Histone H2A.Z inheritance during the cell cycle and its impact on promoter organization and dynamics.* Nature structural & molecular biology, 2012. **19**(11): p. 1076-83.
3. Istomina, N.E., et al., *Insulation of the Chicken β–globin Chromosomal Domain from a Chromatin-Condensing Protein, MENT.* Mol Cell Biol, 2003. **23**(18): p. 6455-6468.
4. Egelhofer, T.A., et al., *An assessment of histone-modification antibody quality.* Nat Struct Mol Biol, 2011. **18**(1): p. 91-3.
5. de Hoon, M.J., et al., *Open source clustering software.* Bioinformatics, 2004. **20**(9): p. 1453-4.
6. Song, Q. and A.D. Smith, *Identifying dispersed epigenomic domains from ChIP-Seq data.* Bioinformatics, 2011. **27**(6): p. 870-1.
7. Wen, B., et al., *Large histone H3 lysine 9 dimethylated chromatin blocks distinguish differentiated from embryonic stem cells.* Nat Genet, 2009. **41**(2): p. 246-50.

8. Guelen, L., et al., *Domain organization of human chromosomes revealed by mapping of nuclear lamina interactions.* Nature, 2008. **453**(7197): p. 948-51.
9. Barski, A., et al., *High-resolution profiling of histone methylations in the human genome.* Cell, 2007. **129**(4): p. 823-37.
10. Schuster-Bockler, B. and B. Lehner, *Chromatin organization is a major influence on regional mutation rates in human cancer cells.* Nature, 2012. **488**(7412): p. 504-7.
11. Welch, J.S., et al., *The origin and evolution of mutations in acute myeloid leukemia.* Cell, 2012. **150**(2): p. 264-78.
12. Bortolin-Cavaille, M.L. and J. Cavaille, *The SNORD115 (H/MBII-52) and SNORD116 (H/MBII-85) gene clusters at the imprinted Prader-Willi locus generate canonical box C/D snoRNAs.* Nucleic acids research, 2012. **40**(14): p. 6800-7.
13. McDonald, O.G., et al., *Genome-scale epigenetic reprogramming during epithelial-to-mesenchymal transition.* Nat Struct Mol Biol, 2011. **18**(8): p. 867-74.
14. Schultz, D.C., et al., *SETDB1: a novel KAP-1-associated histone H3, lysine 9-specific methyltransferase that contributes to HP1-mediated silencing of euchromatic genes by KRAB zinc-finger proteins.* Genes Dev, 2002. **16**(8): p. 919-32.
15. Blackburn, M.L., et al., *Genomic structure and expression of the mouse ESET gene encoding an ERG-associated histone methyltransferase with a SET domain.* Biochimica et biophysica acta, 2003. **1629**(1-3): p. 8-14.
16. Chen, L., et al., *MDM2 recruitment of lysine methyltransferases regulates p53 transcriptional output.* The EMBO journal, 2010. **29**(15): p. 2538-52.
17. Chen, X., et al., *G9a/GLP-dependent histone H3K9me2 patterning during human hematopoietic stem cell lineage commitment.* Genes & development, 2012. **26**(22): p. 2499-511.
18. Hoemme, C., et al., *Chromatin modifications induced by PML-RARalpha repress critical targets in leukemogenesis as analyzed by ChIP-Chip.* Blood, 2008. **111**(5): p. 2887-95.
19. Kim, J.Y., et al., *KDM3B is the H3K9 demethylase involved in transcriptional activation of lmo2 in leukemia.* Molecular and cellular biology, 2012. **32**(14): p. 2917-33.
20. Son, H.J., et al., *Negative regulation of JAK2 by H3K9 methyltransferase G9a in leukemia.* Molecular and cellular biology, 2012. **32**(18): p. 3681-94.
21. Su, R.C., et al., *Dynamic assembly of silent chromatin during thymocyte maturation.* Nat Genet, 2004. **36**(5): p. 502-6. Epub 2004 Apr 18.
22. Timp, W. and A.P. Feinberg, *Cancer as a dysregulated epigenome allowing cellular growth advantage at the expense of the host.* Nature reviews. Cancer, 2013. **13**(7): p. 497-510.
23. Peric-Hupkes, D., et al., *Molecular maps of the reorganization of genome-nuclear lamina interactions during differentiation.* Molecular Cell, 2010. **38**(4): p. 603-13.
24. Jenuwein, T. and C.D. Allis, *Translating the histone code.* Science, 2001. **293**(5532): p. 1074-80.
25. Felsenfeld, G., et al., *Chromatin boundaries and chromatin domains.* Cold Spring Harb Symp Quant Biol, 2004. **69**: p. 245-50.
26. Huang, S., et al., *USF1 recruits histone modification complexes and is critical for maintenance of a chromatin barrier.* Mol Cell Biol, 2007. **27**(22): p. 7991-8002.
27. *Genomic and epigenomic landscapes of adult de novo acute myeloid leukemia.* The New England journal of medicine, 2013. **368**(22): p. 2059-74.

28. Grigoryev, S.A., Y.A. Bulynko, and E.Y. Popova, *The end adjusts the means: heterochromatin remodelling during terminal cell differentiation.* Chromosome Res, 2006. **14**(1): p. 53-69.
29. Popova, E.Y., et al., *Developmentally Regulated Linker Histone H1c Promotes Heterochromatin Condensation and Mediates Structural Integrity of Rod Photoreceptors in Mouse Retina.* The Journal of biological chemistry, 2013.
30. Tachibana, M., et al., *G9a histone methyltransferase plays a dominant role in euchromatic histone H3 lysine 9 methylation and is essential for early embryogenesis.* Genes Dev, 2002. **16**(14): p. 1779-91.
31. Tachibana, M., et al., *G9a/GLP complexes independently mediate H3K9 and DNA methylation to silence transcription.* Embo J, 2008. **27**(20): p. 2681-90.
32. Stancheva, I., *Caught in conspiracy: cooperation between DNA methylation and histone H3K9 methylation in the establishment and maintenance of heterochromatin.* Biochem Cell Biol, 2005. **83**(3): p. 385-95.
33. Schoofs, T., W.E. Berdel, and C. Muller-Tidow, *Origins of aberrant DNA methylation in Acute Myeloid Leukemia.* Leukemia, 2013.
34. Kind, J., et al., *Single-cell dynamics of genome-nuclear lamina interactions.* Cell, 2013. **153**(1): p. 178-92.
35. Lubbert, M., et al., *Efficacy of a 3-day, low-dose treatment with 5-azacytidine followed by donor lymphocyte infusions in older patients with acute myeloid leukemia or chronic myelomonocytic leukemia relapsed after allografting.* Bone Marrow Transplant, 2009.
36. Chowdhury, S., S. Seropian, and P.W. Marks, *Decitabine combined with fractionated gemtuzumab ozogamicin therapy in patients with relapsed or refractory acute myeloid leukemia.* Am J Hematol, 2009. **84**(9): p. 599-600.
37. Recillas-Targa, F., *Epigenetic regulation and therapeutic approaches in cancer.* Curr Top Med Chem, 2008. **8**(15): p. 1367-78.
38. Jain, N., A. Rossi, and G. Garcia-Manero, *Epigenetic therapy of leukemia: An update.* Int J Biochem Cell Biol, 2009. **41**(1): p. 72-80.
39. Schenk, T., et al., *Inhibition of the LSD1 (KDM1A) demethylase reactivates the all-trans-retinoic acid differentiation pathway in acute myeloid leukemia.* Nature medicine, 2012. **18**(4): p. 605-11.
40. Popova, E.Y., et al., *Chromatin condensation in terminally differentiating mouse erythroblasts does not involve special architectural proteins but depends on histone deacetylation* Chromosome Res, 2009. **17**: p. 47-64.

18. Extent of Clinical Activities Initiated and Completed. Items 18(A) and 18(B) should be completed for all research projects. If the project was restricted to secondary analysis of clinical data or data analysis of clinical research, then responses to 18(A) and 18(B) should be "No."

18(A) Did you initiate a study that involved the testing of treatment, prevention or diagnostic procedures on human subjects?

_____Yes
___X__No

18(B) Did you complete a study that involved the testing of treatment, prevention or diagnostic procedures on human subjects?

_____Yes
___X__No

If "Yes" to either 18(A) or 18(B), items 18(C) – (F) must also be completed. (Do NOT complete 18(C-F) if 18(A) and 18(B) are both "No.")

18(C) How many hospital and health care professionals were involved in the research project?

_____Number of hospital and health care professionals involved in the research project

18(D) How many subjects were included in the study compared to targeted goals?

_____Number of subjects originally targeted to be included in the study
_____Number of subjects enrolled in the study

Note: Studies that fall dramatically short on recruitment are encouraged to provide the details of their recruitment efforts in Item 17, Progress in Achieving Research Goals, Objectives and Aims. For example, the number of eligible subjects approached, the number that refused to participate and the reasons for refusal. Without this information it is difficult to discern whether eligibility criteria were too restrictive or the study simply did not appeal to subjects.

18(E) How many subjects were enrolled in the study by gender, ethnicity and race?

Gender:
_____Males
_____Females
_____Unknown

Ethnicity:
_____Latinos or Hispanics
_____Not Latinos or Hispanics

_____Unknown

Race:
_____American Indian or Alaska Native
_____Asian
_____Blacks or African American
_____Native Hawaiian or Other Pacific Islander
_____White
_____Other, specify:_____
_____Unknown

18(F) Where was the research study conducted? (List the county where the research study was conducted. If the treatment, prevention and diagnostic tests were offered in more than one county, list all of the counties where the research study was conducted.)

19. Human Embryonic Stem Cell Research. Item 19(A) should be completed for all research projects. If the research project involved human embryonic stem cells, items 19(B) and 19(C) must also be completed.

19(A) Did this project involve, in any capacity, human embryonic stem cells?
_____Yes
___X_ No

19(B) Were these stem cell lines NIH-approved lines that were derived outside of Pennsylvania?
_____Yes
_____ No

19(C) Please describe how this project involved human embryonic stem cells:

20. Articles Submitted to Peer-Reviewed Publications.

20(A) Identify all publications that resulted from the research performed during the funding period and that have been submitted to peer-reviewed publications. Do not list journal abstracts or presentations at professional meetings; abstract and meeting presentations should be listed at the end of item 17. **Include only those publications that acknowledge the Pennsylvania Department of Health as a funding source** (as required in the grant agreement). List the title of the journal article, the authors, the name of the peer-reviewed publication, the month and year when it was submitted, and the status of publication (submitted for publication, accepted for publication or published.). Submit an electronic copy of each publication or paper submitted for publication, listed in the table, in a PDF version 5.0.5 (or greater) format, 1,200 dpi. Filenames for each publication should include the number of the research project, the last name of the PI, and an abbreviated title of the publication. For example, if you submit two

publications for Smith (PI for Project 01), one publication for Zhang (PI for Project 03), and one publication for Bates (PI for Project 04), the filenames would be:

Project 01 – Smith – Three cases of isolated
Project 01 – Smith – Investigation of NEB1 deletions
Project 03 – Zhang – Molecular profiling of aromatase
Project 04 – Bates – Neonatal intensive care

If the publication is not available electronically, provide 5 paper copies of the publication.

Note: The grant agreement requires that recipients acknowledge the Pennsylvania Department of Health funding in all publications. Please ensure that all publications listed acknowledge the Department of Health funding. If a publication does not acknowledge the funding from the Commonwealth, do not list the publication.

Title of Journal Article:	Authors:	Name of Peer-reviewed Publication:	Month and Year Submitted:	Publication Status (check appropriate box below):
1. None				☐Submitted ☐Accepted ☐Published

20(B) Based on this project, are you planning to submit articles to peer-reviewed publications in the future?

Yes____X_____ No_____

If yes, please describe your plans:

An abstract "Genome-wide mapping of large organized heterochromatin domains reveals hotspots of epigenetic and transcriptional changes associated with myeloid differentiation and acute myeloid leukemia" co-authored by Sergei A. Grigoryev, Anna C. Salzberg, Arthur Berg, Abigail Harris-Becker, Thomas P. Loughran, and David F. Claxton has been submitted for a presentations at the ASBMB meeting in San Diego in April 2014. A manuscript under the same title is being prepared and is planned to be submitted to a peer-reviewed journal (Blood) in spring 2014.

21. Changes in Outcome, Impact and Effectiveness Attributable to the Research Project.
Describe the outcome, impact, and effectiveness of the research project by summarizing its impact on the incidence of disease, death from disease, stage of disease at time of diagnosis, or other relevant measures of outcome, impact or effectiveness of the research project. If there were no changes, insert "None"; do not use "Not applicable." Responses must be single-spaced below, and no smaller than 12-point type. DO NOT DELETE THESE INSTRUCTIONS. There is no limit to the length of your response.

None

22. Major Discoveries, New Drugs, and New Approaches for Prevention Diagnosis and Treatment. Describe major discoveries, new drugs, and new approaches for prevention, diagnosis and treatment that are attributable to the completed research project. If there were no major discoveries, drugs or approaches, insert "None"; do not use "Not applicable." Responses must be single-spaced below, and no smaller than 12-point type. DO NOT DELETE THESE INSTRUCTIONS. There is no limit to the length of your response.

None

23. Inventions, Patents and Commercial Development Opportunities.

23(A) Were any inventions, which may be patentable or otherwise protectable under Title 35 of the United States Code, conceived or first actually reduced to practice in the performance of work under this health research grant? Yes_____ No___X____

If "Yes" to 23(A), complete items a – g below for each invention. (Do NOT complete items a - g if 23(A) is "No.")

a. Title of Invention:

b. Name of Inventor(s):

c. Technical Description of Invention (describe nature, purpose, operation and physical, chemical, biological or electrical characteristics of the invention):

d. Was a patent filed for the invention conceived or first actually reduced to practice in the performance of work under this health research grant?
Yes_____ No____

If yes, indicate date patent was filed:

e. Was a patent issued for the invention conceived or first actually reduced to practice in the performance of work under this health research grant?
Yes_____ No____
If yes, indicate number of patent, title and date issued:
Patent number:
Title of patent:
Date issued:

f. Were any licenses granted for the patent obtained as a result of work performed under this health research grant? Yes_____ No____

If yes, how many licenses were granted?_____

g. Were any commercial development activities taken to develop the invention into a commercial product or service for manufacture or sale? Yes___ No____

If yes, describe the commercial development activities:

23(B) Based on the results of this project, are you planning to file for any licenses or patents, or undertake any commercial development opportunities in the future?

Yes_____ No____X_____

If yes, please describe your plans:

24. Key Investigator Qualifications. Briefly describe the education, research interests and experience and professional commitments of the Principal Investigator and all other key investigators. In place of narrative you may insert the NIH biosketch form here; however, please limit each biosketch to 1-2 pages.

BIOGRAPHICAL SKETCH

Provide the following information for the Senior/key personnel and other significant contributors in the order listed on Form Page 2.
Follow this format for each person. **DO NOT EXCEED FOUR PAGES.**

NAME	POSITION TITLE
Grigoryev, Sergei Alexandrovich	Professor
eRA COMMONS USER NAME (credential, e.g., agency login) SGRIGORYEV	

EDUCATION/TRAINING *(Begin with baccalaureate or other initial professional education, such as nursing, include postdoctoral training and residency training if applicable.)*

INSTITUTION AND LOCATION	DEGREE *(if applicable)*	MM/YY	FIELD OF STUDY
M.V.Lomonosov Moscow State University, Moscow, Russia	M.S.	1975-1980	Biochemistry
M.V.Lomonosov Moscow State University Moscow, Russia	Ph.D.	1981-1985	Molecular Biology

A. Positions and Honors

Positions and Employment

1985-1991 Research Scientist, Department of Molecular Biology, Moscow State University.
1991-1992 Research Associate, University of Massachusetts, Amherst. (Exchange scientist on leave from Moscow State University). With Dr. C.Woodcock.
1992-1995 Research Fellow in Biology, California Inst. of Technology, Pasadena. With Dr. A.Varshavsky.
1995-2000 Research Assistant Professor (non-tenure track), Biology Department, University of Massachusetts at Amherst.
2000-present Assistant, Associate, Full Professor; Department of Biochemistry and Molecular Biology, Penn State University College of Medicine.
2010 Visiting fellow, J.Curtin School of Medical Research, Australian Natl. University, Canberra.

Other Experience and Professional Memberships

Platform talks at major scientific meetings: ACSB meeting (2000), Gordon Conference on Chromatin (2004); International Symposium on SERPINs: (1999, 2002, 2005), BioMaPS Chromatin Structure and Gene Regulation (2008), SUNY-Albany, the XVIth conversation (2009), International Symposium on Physicochemical Field for Genetic Activities (2011), NCI Symposium on Chromosome Biology, NIH, Bethesda, November 2011, Workshop "Evolutionary Dynamics and Information Hierarchies in Biological System", Aspen, CO, August 2012 , Workshop: "Chromatin Structure and Dynamics ", Telluride, CO, August 2012

Journal reviewer: Science, Nature Struc. Mol. Biol., Nature Communications, PNAS, PLOS Biology, Molecular and Cellular Biology, Bioessays, Journal of Molecular Biology, Journal of Biological Chemistry, Biochemistry, Biophysical Journal, Experimental Cell Res., Molecular Genetics and Metabolism, Journal of Neurochemistry, Chromosome Res.

Grant/fellowship reviewer: National Science Foundation (panelist), The Wellcome Trust (UK), Medical Research Council (UK), Penn State COM Scientific Review Committee member.

Membership: The American Society for Cell Biology, American Society for Microbiology, American Association for the Advancement of Science

Honors

1994 James A. Shannon award from the National Institute of Health, USA.
2004 GSA Excellence in Teaching award, Penn State College of Medicine
2006 Award of tenure from Penn State University
2010 Graduate Student's Association award for the "Best overall lecturer "
2010 Dean's award for Excellence in Teaching.
2013 JBC Editorial board invitation (from July 2014)

D. Research Support

Ongoing Research Support (year 2010-2014)

MCB-1021681 (Grigoryev) 9/1/10 - 9/1/14 2.0
NSF $173,373
Internucleosomal interactions in chromatin and metaphase chromosomes
The goal of this proposal is to determine inter-nucleosome interactions that mediate chromatin higher order folding and chromatin condensation in interphase nuclei and metaphase chromosomes. *There is no overlap between this grant and the current proposal.*
Role: PI

Completed Research Support (year 2010-2013)

SAP # 4100050904 (Grigoryev) 1/1/10 - 6/30/13 0.24
PA Tobacco Settlement Fund $40,000
Epigenetic changes in myeloid differentiation and acute myeloid leukemia
The goal of this grant is to develop a hypothesis towards the role of heterochromatin changes in development of human Acute Myeloid Leukemia.
Role: PI

MCB-0615536 (Grigoryev) 9/1/06 - 9/1/10 2.4
NSF $92,964
Molecular and Structural Determinants of Heterochromatin
The goal of this proposal is to determine inter-nucleosome interactions and protein factors that mediate chromatin higher order folding and chromatin condensation in heterochromatin.
Role: PI

SAP#4100037822 (Grigoryev) 7/1/08 - 6/30/10 1.2
PA Tobacco Settlement Fund $30,000
Chromatin proteins involved in myogenic differentiation and myocyte hypertrophy
The goal of this grant is to develop a hypothesis towards the role of chromatin-associated epigenetic factors Suv39h1 and H2A.Z in cardiac muscle cell differentiation and primary cardiomyocyte hypertrophy.
Role: PI

BIOGRAPHICAL SKETCH

Provide the following information for the key personnel and other significant contributors in the order listed on Form Page 2.
Follow this format for each person. **DO NOT EXCEED FOUR PAGES.**

NAME	POSITION TITLE
Loughran, Jr., Thomas Patrick	Director, Penn State Hershey Cancer Institute
eRA COMMONS USER NAME (credential, e.g., agency login) tloughra	

EDUCATION/TRAINING *(Begin with baccalaureate or other initial professional education, such as nursing, and include postdoctoral training.)*

INSTITUTION AND LOCATION	DEGREE *(if applicable)*	YEAR(s)	FIELD OF STUDY
Ursinus College, Collegeville, PA	BS	1975	Biology
Hahnemann Medical School, Philadelphia, PA	MD	1979	Internal Medicine

Please refer to the application instructions in order to complete sections A, B, and C of the Biographical Sketch.

A. Positions and Honors.

Positions and Employment

1979-1982 Internal Medicine Internship and Residency; Thomas Jefferson University Hospital; Philadelphia, PA
1982-1985 Senior Oncology Fellow; Fred Hutchinson Cancer Research Center; University of Washington; Seattle, WA
1983-1985 Transplantation Biology Division; Fred Hutchinson Cancer Research Center; Laboratory of Dr. Rainer Storb
1985-1986 Associate in Clinical Research; Fred Hutchinson Cancer Research Center; Seattle, WA
10/86-1/87 Visiting Scientist, National Cancer Institute-Frederick Cancer Research Facility, Laboratory of Dr. Francis Ruscetti; Frederick, MD
1986-1990 Assistant Member, Clinical Research; Fred Hutchinson Cancer Research Center; Seattle, WA
1990-1992 Associate Member, Clinical Research; Fred Hutchinson Cancer Research Center, Seattle, WA
1992-1996 Professor of Medicine and Microbiology and Immunology
Associate Director, Bone Marrow Transplant Program SUNY Health Science Center and Chief of Hematology, Veterans Administration Medical Center; Syracuse, NY
1996-2003 Program Leader, Hematologic Malignancies, H. Lee Moffitt Cancer Center and Research Institute, and Professor of Medicine, University of South Florida; Tampa, FL
2003-present Director, Penn State Hershey Cancer Institute, and Professor of Medicine, Penn State Milton S. Hershey Medical Center, Penn State College of Medicine; Hershey, PA

Honors

1975 B.S. Cum Laude, Departmental Honors (Biology), Ursinus College Sigma Xi National Research Honor Society; Cub and Key Honor Society, Ursinus College
1976 Honors for Basic Sciences curriculum (top 10% of class), Hahnemann Medical School
1979-1980 Intern of the Year, Thomas Jefferson University Hospital
1983-1984 American Cancer Society Clinical Fellow
1985-1987 Leukemia Society of America Fellow
1987-1990 Special Fellow of the Leukemia Society of America

B. Selected peer-reviewed publications (in chronological order).
(Publications selected from 174 peer-reviewed publications)

C. Research Support.

Ongoing Research Support

U54 Application RR019397 Loughran (PI) 9/1/03-8/31/09
NIH
Bone Marrow Failure Clinical Research Center
The major goal of this project is to develop clinical research protocols for rare hematologic diseases, including aplastic anemia, MDS, and LGL leukemia.
Role: PI

R01 CA133525 Loughran (PI) 3/1/08-2/28/13
NIH
Targeted Therapeutics of LGL Leukemia Utilizing Ceramide Nanoliposomes
The major goal of this project is to design, develop and optimize the next-generation of pegylated C_6-ceramide cationic liposomal formulations for treatment of large granular lymphocyte (LGL) leukemia.
Role: PI

R01 CA94872 Loughran (PI) 7/1/03-6/30/11
NCI
Survival Mechanisms in Leukemic NK Cells
The major goal of this project is to understand the signaling mechanisms responsible for survival of leukemic NK cells.
Role: PI

Completed Research Support (during last 3 years)

R01 CA112112 Loughran (Co-PI) 8/15/05-8/14/09
NCI (8/1/08 completion at out site; ongoing at Moffitt)
Targeted Drug Therapy for LGL Leukemia
The major goal of this project is to assess therapeutic efficacy of targeted drug therapy with farnestyltransferase inhibitor for LGL leukemia.
Role: Co-PI

R01 CA90633 Loughran (PI) 4/1/01-3/31/06
NCI
Therapeutic Response in LGL Leukemia
The major goal of this project is to determine mechanism of treatment efficacy by monitoring samples obtained from patients on ECOG trial of methotrexate.
Role: PI

VA Merit Review Grant Loughran (PI) 4/1/00-9/30/06
VA
Pathogenesis of LGL Leukemia
The major goal of this project is a laboratory assessment of Fas resistance in leukemic LGL.
Role: PI

BIOGRAPHICAL SKETCH

Provide the following information for the key personnel and other significant contributors in the order listed on Form Page 2.
Follow this format for each person. **DO NOT EXCEED FOUR PAGES.**

NAME **Claxton, David F.**	POSITION TITLE Professor of Medicine
eRA COMMONS USER NAME DCLAXTON	

EDUCATION/TRAINING (*Begin with baccalaureate or other initial professional education, such as nursing, and include postdoctoral training.*)

INSTITUTION AND LOCATION	DEGREE *(if applicable)*	MM/YY	FIELD OF STUDY
McGill University, Montreal, Quebec	College Diploma	05/73	Medicine
McGill University, Montreal, Quebec	MD	05/78	Medicine

A. Personal Statement

I am Professor of Medicine, Director of the Hematological Malignancies Group, Penn State Hershey Cancer Institute, Hershey, PA. My research interests focus on biology and developmental therapeutics for leukemias and myelodysplastic syndrome. These activities include laboratory investigation in collaboration with basic scientists, and clinical investigation. I have authored over 100 articles, book chapters, abstracts and editorials. I am fully committed of this application, and will facilitate its goals in every way possible.

B. Positions and Honors

Positions and Employment

1978 – 1979	Rotating Internship, Royal Jubilee Hospital, Victoria, B.C.
1979 – 1980	General Practice – Emergency Room, Hotel Dieu St. Jerome, Rue Principale, St. Jerome, Quebec
1981 – 1984	Internal Medicine Residency, Royal Victoria Hospital, Montreal, Quebec
1984 – 1987	Hematology Fellowship, Royal Victoria Hospital, Montreal, Quebec
1987 – 1990	Fellow, Medical Oncology, The University of Texas M.D. Anderson Cancer Center, Houston, Texas
1990 – 1992	Faculty Associate, The University of Texas M.D. Anderson Cancer Center, Houston, Texas
1992 – 1998	Assistant Professor, Department of Hematology, The University of Texas M.D. Anderson Cancer Center, Houston, Texas
1998 – 2000	Associate Professor, Department of Blood and Marrow Transplantation, The University of Texas M.D. Anderson Cancer Center, Houston, Texas
2000 – 2005	Associate Professor with Tenure, Department of Medicine, Division of Hematology/ Oncology, Milton S. Hershey Medical Center, Hershey, Pennsylvania
2005 – Present	Professor of Medicine, Division of Hematology/Oncology, Milton S. Hershey Medical Center, Hershey, Pennsylvania

Other Experiences and Professional Memberships

6/08 – 9/09	Interim Chief, Division of Hematology/Oncology, Milton S. Hershey Medical Center, Hershey, Pennsylvania

Honors

1972	Academic Scholarship, McGill University
1987 – 1988	Clifton D. Howe Award for Clinical Excellence in Medical Oncology
1989 – 1990	Upjohn Research Award for Outstanding Contributions in Cancer Research

CYC682-06 Claxton (PI) 07/07/08 – 07/06/10
Cyclacel Pharmaceuticals, Inc.
A Randomized Phase II Study of Oral Sapacitabine in Elderly Patients with Acute Myeloid Leukemia Previously Untreated or in First Relapse. The primary objective of this study is to evaluate 1-year survival rate of 3 dosing schedules in elderly patients with previously untreated or first relapsed acute myeloid leukemia.
Role: PI

Investigator Initiated Trial iCLO125 Claxton (PI) 06/09/09 – 12/08/10
Genzyme Corporation.
Clofarabine, Etoposide, and Mitoxantrone for Relapsed and Refractory Acute Leukemias. The purpose of this study is to establish toxicity, DLT and an MTD/ recommended Phase 2 dose of Clofarabine in combination with Etoposide and Mitoxantrone for therapy of relapsed or refractory acute leukemias.
Role: PI

Investigator Initiated Trial iCLO085 Claxton (PI) 07/30/07 – 07/29/10
Genzyme Corporation.
The purpose of this study is to establish the safety of Clofarabine and cyclophosphamide preceding allogeneic hematopoietic engraftment and to assess the efficacy of Clofarabine and cyclophosphamide as conditioning for promoting allogeneic hematopoietic engraftment.
Role: PI

GEM016 Claxton (PI) 04/30/08 – 12/05/10
Gemin X.
A Multicenter, Open-Label, 2-Stage, Phase II Study of Single0Agent Obatoclax Mesylate Administered for 3 Consectutive Days Every 2 Weeks to Older Patients with Previously Untreated Acute Myeloid Leukemia (AML). The primary objective of this study is to determine the rate of morphologic CR of obatoclax single-agent therapy in older patients with previously untreated AML.
Role: PI

0001A3-300-GL Claxton (PI) 04/16/09 – 10/15/10
Antisoma Research Limited.
Phase 3 Open-Label Randomized Study of Amonafide L-Malate in Combination with Cytarabine compared to Daunorubicin in Combination with Cytarabine in Patients with Secondary Acute Myeloid Leukemia (AML). The purpose of this study is to establish the rate of confirmed complete remission with or without hematopoietic recovery (Cri); this includes complete cytogenetic remission (CRc) and complete remission with persistent morphologic dysplasia (CRd).
Role: PI

CL-024300606 Claxton (PI) 01/29/07 – 01/28/11
Genzyme Corporation.
A Phase II Study of Single Agent Clofarabine in Previously Untreated Older Adult Patients with Acute Myelogenous Leukemia (AML) for Whom Standard Induction Chemotherapy is Unlikely to be of Benefit. This study assessed the efficacy of clofarabine in previously untreated adult patients ≥ 60 years old with acute myelogenous leukemia (AML) for whom standard induction chemotherapy is unlikely to be of benefit.
Role: PI