

Transcription and methylation analyses of preleukemic promyelocytes indicate a dual role for PML/RARA in leukemia initiation

Coline Gaillard,^{1,5} Taku A. Tokuyasu,² Galit Rosen,³ Jason Sotzen,⁴ Adeline Vitaliano-Prunier,⁵ Ritu Roy,² Emmanuelle Passegué,⁶ Hugues de Thé,⁵ Maria E. Figueroa,⁴ and Scott C. Kogan¹

¹Department of Laboratory Medicine, University of California, San Francisco, CA, USA; ²Computational Biology Core, Helen Diller Family Comprehensive Cancer Center, University of California, San Francisco, CA, USA; ³Center for Cancer and Blood Disorders, Phoenix Children's Hospital, AZ, USA; ⁴Department of Pathology, University of Michigan, Ann Arbor, MI, USA; ⁵Institut Universitaire d'Hématologie, Université Paris-Diderot UMR 944/7212, France; and ⁶Department of Medicine, Eli and Edythe Broad Center of Regeneration Medicine and Stem Cell Research, University of California, San Francisco, CA, USA

©2015 Ferrata Storti Foundation. This is an open-access paper. doi:10.3324/haematol.2014.123018

The online version of this article has a Supplementary Appendix.

Manuscript received on December 23, 2014. Manuscript accepted on May 6, 2015.

Correspondence: scott.kogan@ucsf.edu

Supplemental Methods:

BrdU incorporation assays

For *in vitro* incorporation, promyelocytes were sorted into Myelocult M5300 (Stem Cell technologies#07904) containing 100ng/ml of recombinant IL3, IL6, GM-CSF and SCF (Peprotech#213-13, 216-16, 315-03 and 250-03). BrdU (60 μ M) or PBS were added for 1h at 37°C. For *in vivo* incorporation, mice were injected with 2mg of BrdU or PBS and animals euthanized 1h later, BM harvested and promyelocytes sorted. For detection, cells were fixed in 4%PFA, washed in PBS 50mM NH₄Cl and permeabilized with 0.2%Tritonx100. Cells were DNaseI-treated for 30min, stained for 30min with anti BrdU-PE (eBioscience#12-5071-41) at 1/40 (*in vivo*) or 1/400 (*in vitro*) and percentage of PE⁺ cells assessed by flow cytometry.

qRT-PCR

Online supplementary Figures 1B and 5B: Assays were run using the TaqMan Universal PCR Master Mix as per manufacturer's instructions (Life Technologies#4324018). Triplicate wells were run, plus one well

without template for each probe as control. To generate cDNA, RNA samples were DNaseI-treated (Life Technologies#AM2222), then incubated with random hexamers (Roche#58002113-01) and dNTPs (Bioline#39044). Reverse transcriptase reaction was carried out with Superscript III (Life Technologies#18080-044), RNase out (Life Technologies#10777-019) and DTT (Life Technologies#15508-013). Following the RT reaction, RNase H (Life Technologies #18021-071) was added to complete the process.

Online supplementary Figure 5A: Assays were run using the One-Step RT-PCR Master Mix as per manufacturer's instructions (Life Technologies#4309169). Quadruplicate wells were run, including one well without reverse transcriptase as control.

In vitro differentiation assay

Cells (10,000) were directly sorted onto a 96 well-plate containing Myelocult M5300 (Stem Cell technologies#07904) and 100ng/ml of recombinant murine SCF and G-CSF (Peprotech#250-03 and 250-05). Wells were harvested every 24h for cytopspins and stained with Wright-giemsa.

Gene set enrichment analysis (GSEA)

GSEA was performed using the command line tool v2-2.0.10, with the normalized array data used as input. Gene sets from the C2 and C5 collections of MSigDB version 3.1 were assessed, using 5,000 gene set permutations to compute the null distribution, otherwise using default settings. Three replicate runs were performed for each parameter setting. Gene sets that passed thresholds (false discovery rate $q\text{-value} < 0.05$ for C2; nominal $p\text{-value} < 0.05$ for C5) in all three runs were retained.

Heatmap generation**1- Transcriptome heatmap, Wt vs PML/RARA preleukemic, Figure 3A**

The normalized intensity values on the top ten percent of probes (4,119 total) by variance were selected. The data were median-centered per probe, and values beyond ± 3 were brought in to their respective cutoffs. Complete linkage hierarchical clustering of samples and probes with Euclidean distance measure was performed in R using the heatmap.2 function in the gplots package.

2- DNA methylation heatmap, Wt vs PML/RARA preleukemic, Figure 4A

From the raw ERRBS data files, chromosomal region and freqC (*i.e.* percentage methylated, range 0 to 100) columns were extracted. Over one million regions were interrogated per sample. Data from regions common to all samples were formed into a matrix and logit transformed. Changes in methylation between early and late promyelocytes were assessed using the limma R/Bioconductor package (*i.e.* disregarding the genotype factor). Applying a raw p-value threshold of 0.001 resulted in 1,615 differentially methylated regions. Hierarchical clustering of the logit-transformed data results in the cluster heatmap presented in Figure 4A where the colors represent the original freqC values. The heatmap was constructed using the heatmap.2, colorRampPalette, and reorder.dendrogram functions in R.

3- Lian granule genes heatmap (Lian *et al*, Blood 2001;98(3):513-524), Figure 6C

Gene symbols in the LIAN_NEUTROPHIL_GRANULE_CONSTITUENTS gene set were taken from the MSigDB v3.1 symbols file (msigdb.v3.1.symbols.gmt) and mapped to Entrez Gene identifiers

(egIDs) using the Bioconductor package `org.Mm.eg.db`. Agilent accessions were mapped to egIDs with the `org.Mm.eg.db` package and both sets were matched. For replicate instances of a given gene, the median of the data per sample was taken. The data per gene was median centered and the heatmap generated using the `heatmap.2` function of the `gplots` R package.

4- Transcriptome heatmap, PML/RARA preleukemic vs leukemic,

Online supplementary Figure 6B

The normalized intensity values on the top ten percent of probes by variance were selected. The data were median-centered per probe, and values beyond ± 4 were brought in to their respective cutoffs. Complete linkage hierarchical clustering of samples and probes with Euclidean distance measure was performed in R using the `heatmap.2` function in the `gplots` package.

5- Human normal promyelocytes vs M3 APL and humanized APL

samples heatmap (Payton *et al*, JCI 2009;119(6):1714-1726, Matsushita *et al*, PLoS One 2014; 9(11):e111082), Figure 6D

The Payton and Matsushita gene expression arrays were processed using ENSEMBL based CDF file (HG133Plus2_Hs_ENST_19.0.0)

from <http://brainarray.mbni.med.umich.edu/Brainarray>. The datasets were then quantile normalized (Bolstad *et al*, Bioinformatics 2003;19(2):185-193). The probesets were annotated with GENCODE v21 database by mapping ENSEMBL IDs. The human arrays were mapped to the mouse array with gene symbols. Differentially expressed (DE) genes were identified as those having unadjusted $p < 0.001$. The association of clinical variables with gene expression was made using moderated t-statistics using the limma package in R/Bioconductor (Gentleman *et al*, Genome Biology 2004;5(10):R80). Because of human/mouse differences, only 9 out of the 13 common genes from *Online supplementary Table 4-‘CommonEarly-Late’*, could be matched. For genes with more than one probe, the mean log₂ intensity was used. The heatmap was generated using the image function and custom code in graphics package in R. Genes were clustered using hierarchical clustering with Pearson correlation as the distance matrix and Ward as the agglomeration method.

Overlap analysis between differentially expressed (DE) genes, differentially methylated regions (DMRs) and PML/RARA binding sites

For each comparison, gene identifier lists were created using the most appropriate and reproducible mapping method, and compared for enrichment as detailed below:

1. DE genes versus DMRs, Figure 4D

Initial lists were compiled consisting of 1) unique DE mouse gene symbols and 2) unique mouse gene symbols associated with DMRs (see ERRBS methods for annotation details). Significance of the overlap was calculated using a one-sided Fisher's exact test assuming a gene universe size of 20,424. Specifically, the universe size corresponds to the number of Entrez Gene IDs that can be mapped from the Agilent mRNA accessions via `org.Mm.eg.db` `ACCNUM2EG`, plus one to account for the presence of a UCSF custom probe (murine *Cebpe*).

2. DE genes versus PML/RARA binding sites identified by Martens (Martens *et al*, *Cancer Cell* 2010;17(2):173-185), *Online supplementary Figure 7A*

Initial lists were compiled consisting of 1) unique DE mouse gene symbols and 2) unique human Ensembl IDs from 'Table S3 PMLRAR common' in Martens *et al*. The latter were mapped to gene symbols through the

following conversion steps: Ensembl IDs > Entrez Gene IDs > MGI IDs (mouse genome informatics IDs, <http://www.informatics.jax.org/>) > mouse Entrez IDs > mouse gene symbols. Specifically, mapping was performed using the org.Hs.eg.db R Bioconductor package (ENSEMBL2EG map), a custom annotations table downloaded from the HUGO gene nomenclature committee (HGNC, www.genenames.org), the org.Mm.eg.db package (MGI2PROBE map), and Agilent probe annotations for the mouse whole genome array. This resulted in a list of 432 unique non-blank mouse gene symbols representing the Martens binding sites.

Significance of the overlap was calculated using a one-sided Fisher's exact test assuming a gene universe size of 20,424.

3. DMRs versus PML/RARA binding sites identified by Martens (Martens *et al*, Cancer Cell 2010;17(2):173-185), *Online supplementary Figure 7B*

Initial lists were compiled consisting of 1) unique mouse gene symbols associated with DMRs (see ERRBS methods for annotation details) 2) human Ensembl IDs from 'Table S3 PMLRAR common' in Martens *et al*. The latter were mapped to mouse gene symbols through the following conversion steps: human Ensembl IDs > human Entrez Gene IDs >

MGI IDs (mouse genome informatics IDs, <http://www.informatics.jax.org/>) > mouse Entrez IDs > mouse gene symbols. Specifically, mapping was performed using the org.Hs.eg.db R Bioconductor package (ENSEMBL2EG map), and a custom annotations table downloaded from the HUGO gene nomenclature committee (HGNC, www.genenames.org), the org.Mm.eg.db package (MGI2PROBE map). This resulted in a list of 432 unique non-blank mouse gene symbols representing the Martens binding sites. Significance of the overlap was calculated using a one-sided Fisher's exact test assuming a gene universe size of 20,424.

Overlap analysis between genes identified as differentially expressed (DE) in this study vs genes identified as DE in a study of Cathepsin G PML/RARA mice (Wartman et al, PLoS One 2012;7(10):e46529), Online supplementary Table 5.

DE genes identified in this study (*Online supplementary Tables 1- 3*) were compared with DE genes identified in the study of CSG mice (ANOVA lists, Tables S3, S4, S6), with genes showing a similar direction of change selected. Significance of the overlap was calculated using a one-sided

Fisher's exact test assuming a gene universe size of 20,424, with similar results when smaller (12,000) or larger (30,000) gene universe sizes were considered.

Cytospins and images

Cytospins were prepared following a 4min spin at 400rpm and stained with Wright-giemsa staining. Photographs were taken on a Nikon Eclipse 80i microscope with a Nikon Digital Sight camera on a 100x lens using NIS-Elements F2.30 software at a resolution of 2560×920. Using Adobe Photoshop CS3, image quality was improved with unsharp mask, and images were re-sized and set at a resolution of 300 pixels/inch.

Chromatin Immunoprecipitation

Four independent experiments were performed on bone marrow cells harvested from primary PML/RARA leukemia. ChIP was performed using LowCell ChIP kit (Diagenode#C01010070) according to manufacturer's recommendations except that chromatin was first incubated with antibodies overnight and then for 2 hours with beads. Antibodies used were anti GFP-FL (Santa-Cruz#sc-8334), anti-RNA polymerase II CTD (Abcam#ab817) and anti-PML+RARA fusion (Abcam#ab43152).

Real-time quantitative PCR was performed using the Fast SYBR Green Master Mix (Life Technologies#4385610) on the 7500 Fast Real-Time PCR system (Life Technologies) and the subsequent oligonucleotides: *Rarb* as a positive control (forward: GTTGGGTCATTTGAAGGTTAGC; reverse: ATATATGCGAGTGAACCTTTCGG) and two sites on the *lactoferrin* locus, a DR19 repeat (forward: CCTGTTTCCCAAAGTGGCTACATC; reverse: TGGTCTTGTTTCAGCCACTGTACTC) and a DR1 repeat (forward: GTAACCTGAAGTGGACAGGCATACG; reverse: CATGGCATCAGCTCTGTTTGTCTG). The relative occupancy was calculated as fold enrichment over control antibody anti-GFP.

Retroviral transduction and transplantation

Procedures were carried out as described in Jones *et al* (Jones *et al*, JEM 2010;207(12):2581-2594).

Note regarding murine Rara

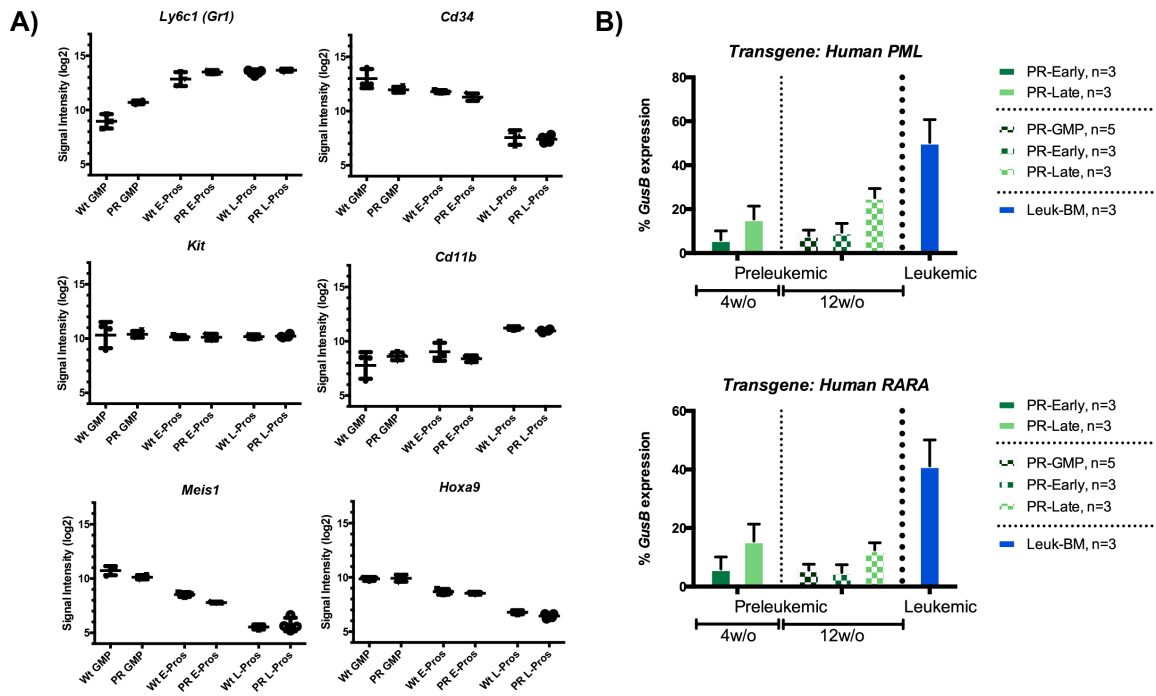
Rara appears as 6.5x up-regulated in the PML/RARA late promyelocyte samples. However, since the murine probe on the array presents 98% homology to the human *RARA* sequence, we believe that the differential signal is due to the probe detecting the human *PML/RARA*

transgene. For that reason, *Rara* was not considered a DE gene and was removed from Figures 3B-C, and from the overlap analysis.

Online supplementary Figures:

Online supplementary Figure 1

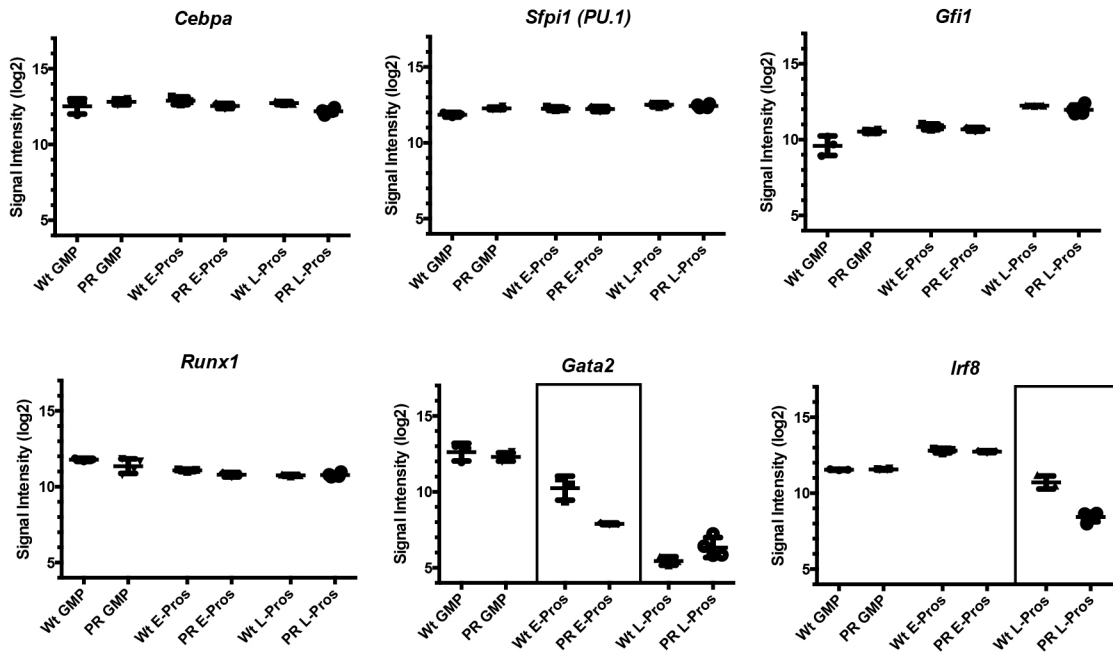
Validation of the sorting strategy: markers and transgene expression



Online supplementary Figure 1: A) Expression data from the microarray validates the sorting strategy utilized. The Y axis displays the raw signal intensity with standard deviation off the array for each biological replicate (log₂ scale). The X axis displays the 6 groups interrogated, from the more immature to the more mature (from left to right: GMP, E-Pros, L-Pros) stage, in the FVB/n (Wt) or PML/RARA (PR) samples. *Ly6c1 (Gr1)*, *Cd34*, and *Kit* values match expectations based on sorting strategy. *Cd11b* increases concomitant with decreasing *Cd34*, whereas *Meis1* and *Hoxa9* decrease as expected. **B)** qPCR data showing expression levels of the transgene at the GMP (PR-GMP; 12 week old only), early and late promyelocyte stages (PR-Early and PR-Late; 4 and 12 weeks old), as well as in leukemic promyelocytes (Leuk-BM). Human probes did not show signal when run on Wt samples (*data not shown*). Expression levels are displayed as a percentage of *Gusb* expression with standard deviation.

Online supplementary Figure 2

Expression of transcription factors in PML/RARA myeloid cells



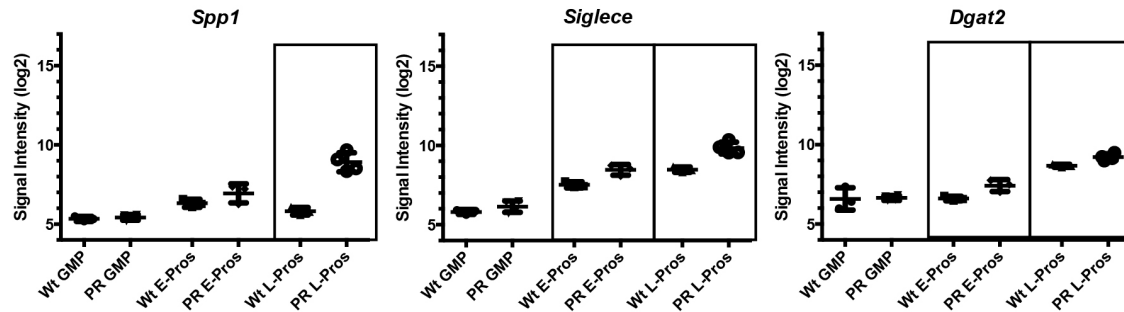
Online supplementary Figure 2: As observed in the microarray, PML/RARA does not alter expression of key myeloid development transcription factors (*Cebpa*, *PU.1*, *Gfi1* and *Runx1*) but differential expression was observed for *Gata2* in the E-Pro compartment and *Irf8* in the L-Pro compartment. The Y axis displays the raw signal intensity with standard deviation off the array for each biological replicate (log₂ scale). The X axis displays the 6 groups interrogated, from the more immature to the more mature (from left to right: GMP, E-Pros, L-Pros) stage, in the FVB/n (Wt) or PML/RARA (PR) samples.

Subsequent qPCR study (*Online supplementary Figure 5B*) demonstrates that the slight difference in *Cebpa* transcripts seen in the microarrays (not reaching $p < 0.001$) likely does reflect a biological difference in expression, consistent with prior reports implicating this gene in APL pathogenesis. The changes observed in *Gata2* and *Irf8* could conceivably contribute to APL pathogenesis, given that both of these genes have been identified as myeloid tumor suppressors (IRF8 reviewed in Abrams. *Immunol Res* 2010;46(1-3):59-71, GATA2 identified in Pasquet *et al*, *Blood* 2013;121(5):822-829). Whereas we subsequently observed that *Gata2* levels are increased in our

leukemic cells, *Irf8* levels are further decreased in leukemia (*Online supplementary Table 10*). Therefore, investigation of a possible role of *Irf8* in APL appears particularly warranted.

Online supplementary Figure 3

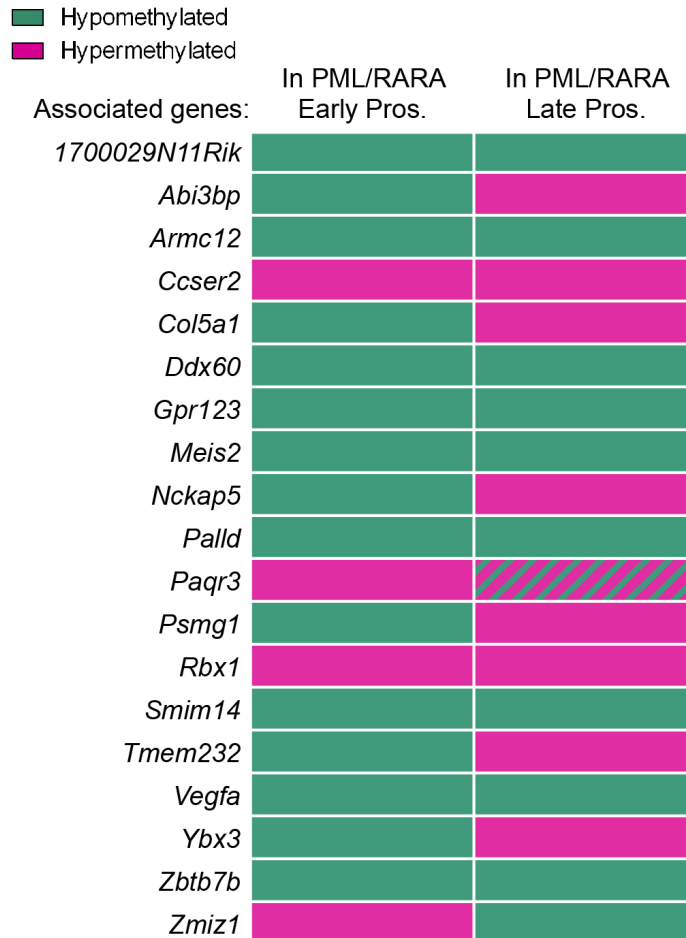
Gene up-regulation in PML/RARA promyelocytes



Online supplementary Figure 3: Expression data from the microarray showing transcript levels of the most up-regulated gene with PML/RARA (*Spp1*), along with two genes that were up-regulated both at the early and late promyelocyte stages (*Online supplementary Table 4*).

Online supplementary Figure 4

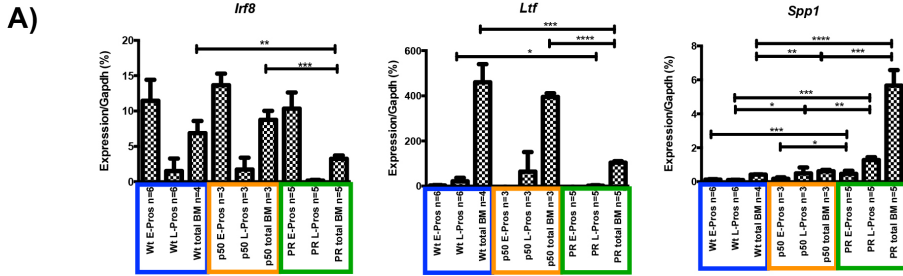
Common differentially methylated regions in both Early and Late PML/RARA promyelocytes (vs Wt)



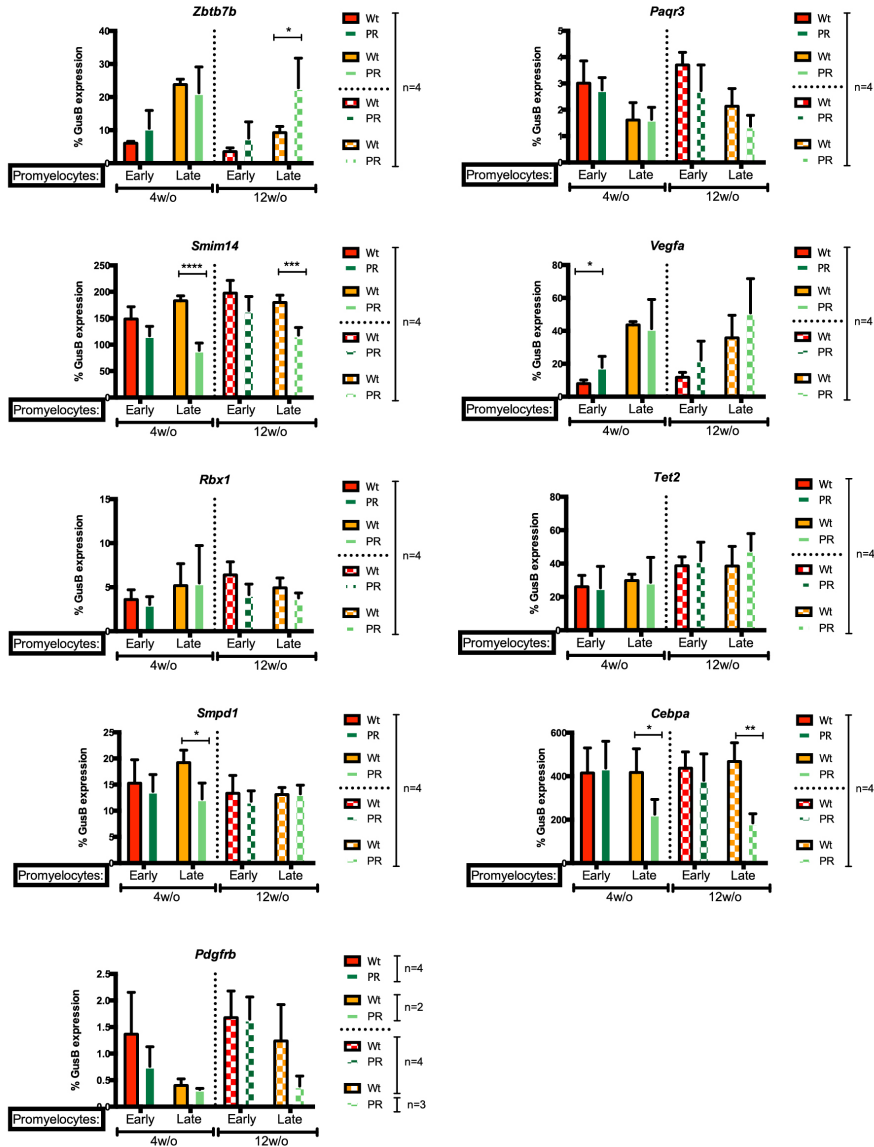
Online supplementary Figure 4: Methylation changes in presence of PML/RARA for the 19 DMRs identified both in the early and late promyelocyte compartments are not consistent at the two stages of maturation.

Online supplementary Figure 5

qPCR expression of selected DE genes in the Wt, p50/RARA and PML/RARA background



B) qPCR expression of selected DMR-associated genes at 4 vs 12 weeks old in Wt and PML/RARA promyelocytes



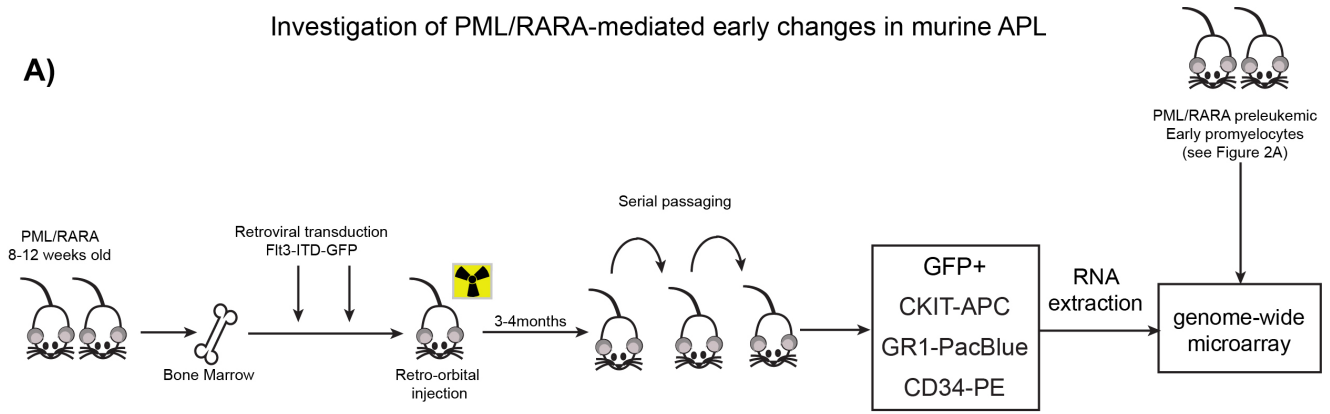
Online supplementary Figure 5: qPCR validation and investigation of expression levels of genes identified as differentially expressed or differentially methylated. **A)** Levels of *Irf8*, *Ltf* and *Spp1* in sorted promyelocytes or total BM from FVB/n (Wt, blue rectangles), p50/RARA (p50, orange rectangles) or PML/RARA (PR, green rectangles) samples. As seen, the expression profile in the p50/RARA background resembled that of Wt, validating that gene deregulation at those loci correlates with the ability to initiate leukemia. A two-tailed t-test was performed to assess statistical significance of differences between groups (* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$, **** $p < 0.0001$). **B)** qPCR quantification of selected genes associated with DMRs. We asked if differential methylation in our 4 weeks old preleukemic samples was an initial step to altered expression later during the preleukemic phase. To this aim, we harvested 12 weeks old early and late promyelocyte samples from Wt and PML/RARA mice, and performed qPCR analysis for a selected set of genes identified as differentially methylated in our ERRBS study. 4 weeks old samples were also studied to obtain baseline expression by qPCR to compare with results at 12 weeks. A two-tailed t-test was performed to assess statistical significance of differences between groups (* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$, **** $p < 0.0001$). Only 1 of 9 genes (*Zbtb7b*) showed a difference between PR and Wt levels of expression

at 12 weeks that was not apparent in the 4 weeks sample. Of interest in this supplemental study was the finding that qPCR analysis of L-Pros showed *Cebpa* to be down-regulated consistently at both time points, even though this transcription factor was not identified as differentially expressed in our microarray analyses. Comparison of the microarray (*Online supplementary Figure 2*) and qPCR (*Online supplementary Figure 5B*) data for *Cebpa* revealed that the less than 2-fold change observed by qPCR was present, but below our threshold for statistical significance, in the microarray study.

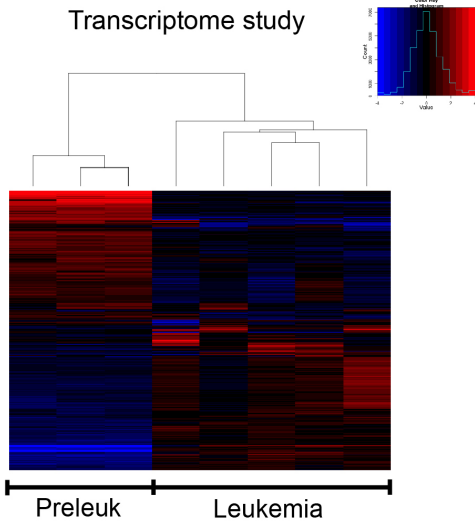
Online supplementary Figure 6

Investigation of PML/RARA-mediated early changes in murine APL

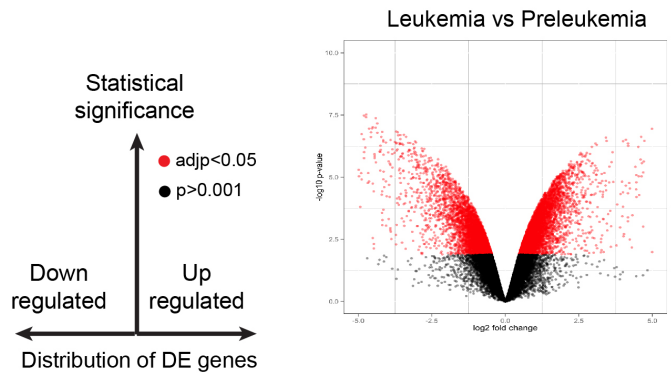
A)



B)



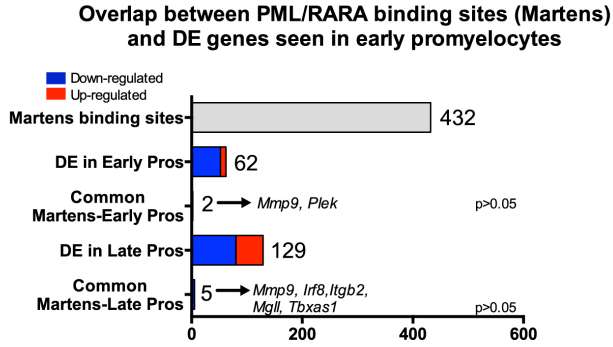
C)



Online supplementary Figure 6: Investigation of the dysregulome in fully-transformed murine leukemic cells. To this aim, we performed microarray analysis on *in vivo*-passaged murine APLs that arose in PML/RARA marrow transduced with activated Flt3 (designated as ‘Flt3-ITD’, carrying an internal tandem duplication as described in Sohal *et al*, Blood 2003;101(8):3188-3197). **A)** Experimental strategy to assess transcriptome deregulation of PML/RARA-Flt3-ITD leukemic samples compared to PML/RARA preleukemic early promyelocytes. **B)** Median-centered heatmap (displaying the top ten percent of probes by variance) of early PML/RARA promyelocytes and PML/RARA-Flt3-ITD leukemia samples. **C)** Volcano plot showing distribution of DE genes in PML/RARA-Flt3-ITD leukemia vs PML/RARA early promyelocytes, colored according to their statistical significance.

Online supplementary Figure 7

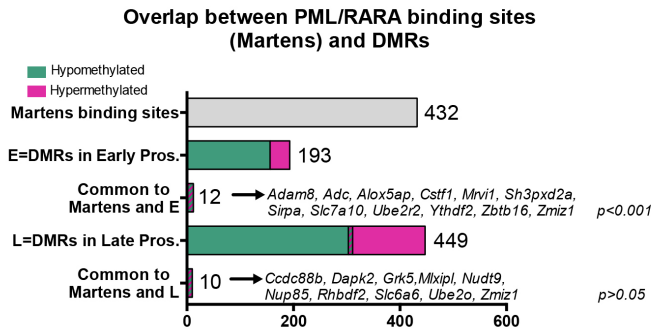
A)



Details of overlapping genes:

Gene	PML/RARA effect on gene expression	
	Down	Fold change
<i>Mmp9</i>	8.50, 5.51	
<i>Irf8</i>	4.82, 2.96	
<i>Itgb2</i>	2.38	
<i>Mgli</i>	1.80, 1.57	
<i>Tbxas1</i>	1.70	

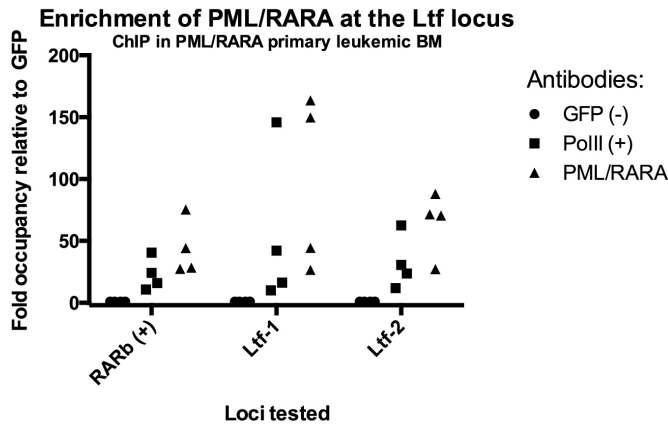
B)



Details of overlapping genes:

Gene	PML/RARA effect on DNA methylation: Early Promyelocytes			Gene	PML/RARA effect on DNA methylation: Late Promyelocytes		
	Hypo	Location/ gene	Distance to TSS		Hypo	Location/ gene	Distance to TSS
	Hyper				Hyper		
<i>Adam8</i>	Green	Gene body	7887	<i>Ccdc88b</i>	Pink	Gene body	1201
<i>Adc</i>	Green	Downstream	34161	<i>Dapk2</i>	Green	Gene body	61468
<i>Alox5ap</i>	Green	Upstream	-20134	<i>Grk5</i>	Green	Gene body	2349
<i>Cstf1</i>	Green	Downstream	10498	<i>Mixipl</i>	Pink	Downstream	41240
<i>Mrvi1</i>	Green	Gene body	65201; 101476	<i>Nudt9</i>	Green	Upstream	-22530
<i>Sh3pxd2a</i>	Green	Gene body	124401	<i>Nup85</i>	Green	Promoter	-758
<i>Sirpa</i>	Green	Upstream	-46431; -46076	<i>Rhbdf2</i>	Pink	Downstream	34068; 36833
<i>Slc7a10</i>	Pink	Upstream	-38404	<i>Slc6a6</i>	Pink	Gene body	29440
<i>Ube2r2</i>	Pink	Upstream	-48054	<i>Ube2o</i>	Pink	Promoter	-8240
<i>Ythdf2</i>	Green	Promoter	-330	<i>Zmiz1</i>	Green	Gene body	51830
<i>Zbtb16</i>	Green	Gene body	90050				
<i>Zmiz1</i>	Pink	Gene body	51830				

C)



Online supplementary Figure 7: Differentially expressed (DE) genes and differentially methylated regions (DMRs) correlate poorly with described PML/RARA binding sites. In figure panels, the included table presents details on the transcript fold change (multiple values if multiple probes) and methylation changes and location (TSS= transcriptional start site). Refer to supplemental methods for details on lists generation, overlaps and statistics.

A) Overlap analysis between published PML/RARA binding sites identified by Martens and DE genes in the early and late promyelocyte compartments (DE with $p < 0.001$). **B)** Overlap analysis between published PML/RARA binding sites identified by Martens and DMRs. Of note, the lack of overlap of binding sites described by Martens *et al.* and our DE genes and DMRs may reflect the marked differences in experimental design including species, cell types, and methods. **C)** CHIP experiment in primary murine PML/RARA leukemia shows that sequences pulled-down with PML/RARA are enriched for the lactoferrin sequence, the most strongly down-regulated gene identified in our microarray.

Online supplementary Tables legends:

Online supplementary Table 1-DE-GMP: List of the differentially expressed (DE) genes between Wt and PML/RARA GMPs with unadjusted $p < 0.001$. The list is sorted by down-regulated or up-regulated genes, and by fold change or p-value in different worksheets. For some genes, more than one probe on the arrays showed differential expression.

Online supplementary Table 2-DE-EarlyPros: List of the differentially expressed (DE) genes between Wt and PML/RARA early promyelocytes with unadjusted $p < 0.001$. The list is sorted by down-regulated or up-regulated genes, and by fold change or p-value in different worksheets. For some genes, more than one probe on the arrays showed differential expression.

Online supplementary Table 3-DE-LatePros: List of the differentially expressed (DE) genes between Wt and PML/RARA late promyelocytes with unadjusted $p < 0.001$. The list is sorted by down-regulated or up-regulated genes, and by fold change or p-value in different worksheets. For some

genes, more than one probe on the arrays showed differential expression.

Online supplementary Table 4-DE-Common: Overlap between genes DE in the GMP, E-Pros and L-Pros compartment (unadjusted $p < 0.001$), highlighted according to their statistical significance.

Online supplementary Table 5-MRP8vsCSG-Overlap: Overlap of DE genes identified in our MRP8 study vs those identified in a previous study using the CSG-PML/RARA model (Wartman *et al*, PLoS One 2012;7(10):e46529). In this table, only DE genes with similar directionality are reported.

Online supplementary Table 6-DMRs-EarlyPros: List of the differentially methylated regions (DMRs) between Wt and PML/RARA early promyelocytes. The list is sorted by hypomethylation and hypermethylation, and by methylation difference, genomic location (with the promoter region being defined as spanning 10Kb upstream of the TSS) or q value in different worksheets. For some genes, more than one tile on the arrays showed

differential methylation and depending on the location, one tile can be annotated to several transcripts.

Online supplementary Table 7-DMRs-LatePros: List of the differentially methylated regions (DMRs) between Wt and PML/RARA late promyelocytes. The list is sorted by hypomethylation and hypermethylation, and by methylation difference, genomic location (with the promoter region being defined as spanning 10Kb upstream of the TSS) or q value in different worksheets. For some genes, more than one tile on the arrays showed differential methylation and depending on the location, one tile can be annotated to several transcripts.

Online supplementary Table 8-GSEA-C5: GSEA C5 (GO gene sets) analysis of DE genes at the GMP, early and late promyelocyte stages. Orange highlights indicate gene sets linked to granule protein function associated with the Wt phenotype. Dark orange highlight indicates the top hit enriched at both the early and late promyelocyte stages, from which enrichment plots and heatmaps of leading edge genes were extracted (see Figure 6B). Green highlights indicate gene sets linked to cell cycle and

mitosis associated with the PML/RARA phenotype. Dark green indicates the strongest hit from which the enrichment plot and heatmap of leading edge genes was extracted (see Figure 5A). NES= normalized enrichment score.

Online supplementary Table 9-GSEA-C2: GSEA C2 (curated gene sets) analysis of DE genes at the GMP, early and late promyelocyte stages. Blue highlight indicates the gene set most significantly associated with the Wt phenotype and depleted from the PML/RARA phenotype (see Figure 6C). NES= normalized enrichment score.

Online supplementary Table 10-DE-Leuk: List of the differentially expressed (DE) genes between PML/RARA early promyelocytes and PML/RARA-Flt3-ITD leukemia with unadjusted $p < 0.001$. The list is sorted by down-regulated or up-regulated genes, and by fold change or p-value in different worksheets. For some genes, more than one probe on the arrays showed differential expression. For this table, the mean over sample replicates was taken (both biological and technical replicates). Since PML/RARA early promyelocyte data was generated from equally mixed male and female samples, whereas the PML/RARA-Flt3-ITD

leukemias are all female samples, genes known to be differentially expressed due to gender differences (Jansen *et al*, BMC Genomics 2014;15:33) are highlighted.

Online supplementary Table 11-GSEA-C2-C5-Leuk: GSEA C2 (curated gene sets) and C5 (GO gene sets) analysis of DE genes between PML/RARA early promyelocytes and PML/RARA-Flt3-ITD leukemia. NES= normalized enrichment score. Note: genes previously described as sex-specific genes (Jansen *et al*, BMC Genomics 2014;15:33) were removed for the analysis.

Online supplementary Table 12-RARA-sites: Scanning of RARA half-sites along the gene bodies of our top 4 DE genes, and in a selected sets of DMRs which were also investigated for quantification by qPCR (*Online supplementary Figure 5B*). Of note regarding the DMRs scanning, paired sites were observed in only 3 of the 9 genes examined, presumably in part due to the limited genomic region we analyzed adjacent to the DMRs.

Online supplementary Table 13-Antibodies-Probes-Buffers: Methods:
antibodies, probes, and buffers.

Supplemental Discussion:

Curious about alternative murine APL models, we compared our lists of DE genes to those obtained in a prior study (Wartman *et al*, PLoS One 2012;7(10):e46529) using the cathepsin G (CSG) model (*Online supplementary Table 5*). A modest but consistent correspondence was found between the DE genes identified in our L-Pro population and the DE genes identified in the CSG CMP and GMP populations, notably the down-regulation of *Irf8*. Of interest, the overall small overlap observed between the two models highlights the important impact that different strains, sorting strategies, kinetics and level of transgene expression may have on the experimental outcome.

Although the goal of our study was to examine the early changes mediated by PML/RARA in myeloid cells, we also performed microarray analysis on *in vivo*-passaged murine APLs that arose in PML/RARA marrow transduced with activated Flt3 (*Online supplementary Figure 6A*). As expected, the samples segregated by disease stage (*Online supplementary Figure 6B*), with much broader changes in leukemic cells compared to that observed for the Wt vs preleukemic comparisons

(*Online supplementary Figure 6C* vs *Figure 3B*), and consisting of a mix of up-regulated and down-regulated genes (*Online supplementary Table 10*). Interestingly, only 2 out of the 13 genes deregulated by PML/RARA both at the early and late promyelocyte stages were among the genes differentially expressed between leukemic cells and their preleukemic counterpart, suggesting that leukemic transformation is not simply an amplification of the changes mediated by PML/RARA at the preleukemic stage, and further highlights the importance of secondary lesions in mediating the broad effects in fully transformed cells. Paralleling these findings, and although we have not examined genome-wide DNA methylation in our leukemic promyelocytes by ERRBS, in a limited study colleagues used our PML/RARA leukemic model to examine methylation in a region of mouse chromosome 12 and found a predominant pattern of hypermethylation (Manodoro *et al*, *Blood* 2014;123(13):2066-2074). Similarly, hypermethylation in human APL samples has been described by several groups (Figueroa *et al*, *Cancer Cell* 2010;17(1):13-27, Schoofs *et al*, *Blood* 2013;121(1):178-187). As we observed that PML/RARA is initially associated with hypomethylation, these data suggest that cells undergo an epigenetic switch during leukemia progression.