Targeted Chromatin Profiling Reveals Novel Enhancers in Ig H and Ig L Chain Loci

Alexander V. Predeus, Suhasni Gopalakrishnan, Yue Huang, Jun Tang, Ann J. Feeney, Eugene M. Oltz and Maxim N. Artyomov

*J Immunol* 2014; 192:1064-1070; Prepublished online 18 December 2013; doi: 10.4049/jimmunol.1302800

http://www.jimmunol.org/content/192/3/1064

---

**Supplementary Material**

http://www.jimmunol.org/content/suppl/2013/12/18/jimmunol.1302800.DCSupplemental

**References**

This article cites 30 articles, 11 of which you can access for free at:

http://www.jimmunol.org/content/192/3/1064.full#ref-list-1

**Subscription**

Information about subscribing to *The Journal of Immunology* is online at:

http://jimmunol.org/subscription

**Permissions**

Submit copyright permission requests at:

http://www.aai.org/About/Publications/JI/copyright.html

**Email Alerts**

Receive free email-alerts when new articles cite this article. Sign up at:

http://jimmunol.org/alerts
Targeted Chromatin Profiling Reveals Novel Enhancers in Ig H and Ig L Chain Loci

Alexander V. Predeus,*§,† Subhasni Gopalakrishnan,*§,† Yue Huang,*§,† Jun Tang,*§,† Ann J. Feeney,*§,† Eugene M. Oltz,*§,† and Maxim N. Artyomov,*§,†

The assembly and expression of mouse Ag receptor genes are controlled by a collection of cis-acting regulatory elements, including transcriptional promoters and enhancers. Although many powerful enhancers have been identified for Ig (Ig) and TCR (Tcr) loci, it remained unclear whether additional regulatory elements remain undiscovered. In this study, we use chromatin profiling of pro-B cells to define 38 epigenetic states in mouse Ag receptor loci, each of which reflects a distinct regulatory potential. One of these chromatin states corresponds to known transcriptional enhancers and identifies a new set of candidate elements in all three Ig loci. Four of the candidates were subjected to functional assays, and all four exhibit enhancer activity in B but not in T lineage cells. The new regulatory elements identified by focused chromatin profiling most likely have important functions in the creation, refinement, and expression of Ig repertoires. The Journal of Immunology, 2014, 192: 1064–1070.

Materials and Methods

Data collection and processing

We considered 16 different epigenetic modifications that can be classified into four groups, as follows: 1) histone modifications (H3K4me1, H3K4me2, H3K4me3, H3K27ac, H3K27me3, H3K36me3, and H3K9ac/K14ac), 2) key transcription factors (p300, Pu.1, Med1, c-Myc, Rad21, CTCF, EBF, E2A, and Pax5), 3) nucleosome-poor, transcribed regions (DNase I hypersensitivity [DHS] and RNA Pol II occupancy), and 4) mature transcriptional signal from RNA-Seq experiments. Fourteen genome-wide experiments were available in public databases. For RNA Pol II and H3K27ac, new chromatin immunoprecipitation (ChIP) analyses were performed on custom-made microarray covering all AgR loci (ChIP-Chip; see below). Supplemental Table I summarizes the sources of all experimental data.

All ChIP-Seq and DHS experiments were processed starting from SRA files. The binary SRA archives were converted into FASTQ files using the SRA toolkit, and then aligned with Bowtie (version 0.12.7) using “-m 1–v 3 –best –strata” options. The resulting alignment SAM file was converted into read BED files using BEDTools. RNA-Seq data were aligned with TopHat (version 1.4.1.1) using “--pre-filter-multithis–max-multithis 15 –segment-length 20” options, and GenBank annotated mRNAs as an alignment reference (–GTF option).

Peak calling

We applied the SICER (7) (v1.1) algorithm to read BED files and call peaks for all ChIP-Seq and DHS experiments. We used settings for narrow peaks (200-bp window size, 200-bp gap size, and false discovery rate of 0.01) in all cases except for H3K27me3 and H3K36me3, which have broad signal distributions (200-bp window size, 600-bp gap size, and 0.01 false discovery rate). Peak identification for RNA-Seq was performed by transcriptome assembly with Cufflinks (8) using no reference transcriptome, and exons of assembled transcripts with fragments per kilobase of transcript per million mapped reads >2 were considered as peaks. For ChIP-Chip experiments, peak calling was performed with MA2C (9) using a p value 0.01.

Genome segmentations

BED files obtained after peak calling were binarized using BEDTools (10). Genome-wide files were prepared with 200-bp windows, and the overlaps...
of peak BED files and window files were calculated. If overlap constituted >50%, the bin was assigned 1. The exact regions of mouse genome (mm9 assembly) that were used for the analysis of AgR loci are as follows: chr6, 40838000–40845000; chr6, 40986000–41250000; chr6, 41476000–41555000 (Tcrb); chr13, 19245000–19490000 (Tcrγ); chr14, 52962000–54855000 (Tcrα/δ); chr12, 114435000–117280000 (Igh); chr6, 67490000–70715000 (Igk); chr16, 189710000–192850000 (Igl). Values for the chromosome outside of AgR boundaries were automatically set to 0, thus excluding all conventional genes from the segmentation. The resulting binarized input was then used in ChromHMM segmentation software (v1.10) (11) to generate hidden Markov models with the number of states ranging from 20 to 40, generating emission and transition probabilities, as well as segmentation BED files and HTML output. Corresponding BED files are available online at https://artemovolab.wustl.edu/publications/support_materials/AgR_2013/.

Conservation analysis

Individual states were overlapped with phyloP30WayPlacental track from UCSC table browser (downloadable as a WIG file; a complete description of how the conservation score was generated is provided at http://hgdownload-test.cse.ucsc.edu/goldenPath/mm9/multiz30way/multiz30way.html), and maximum conservation score for each interval was obtained using an in-house script by picking the highest value within each genomic interval. After this, the average maximum conservation score was calculated for each state by summing individual scores and dividing them by the number of intervals in the state.

ChIP-Chip experiments

Pro–B cells from RAG-deficient mice (C57BL/6, 4–6 wk) were purified using MACS in conjunction with CD19 microbeads (Miltenyi Biotec). ChIP experiments for H3K27ac and RNA Pol II were performed, as described (12), using the following Abs: H3K27ac (Abcam; ab47297) and Pol II (Abcam; ab5131). ChIP-DNA was purified using a QiaGen kit and subjected to whole genome amplification (Sigma-Aldrich), labeled, and hybridized to custom Nimblegen microarrays, according to the manufacturer’s protocol by Mogen. Total input DNA was used as the hybridization control.

Luciferase assays

The following cell lines were used: P5424 (RAG-1+/−, p53+/− pro–T cell line), 63-12 (RAG-2−/− A-MuLV–transformed pro–B cell line), and J558L (B myeloma cell line). All cell lines were cultured at 37°C with 5% CO2 in RPMI 1640 supplemented with 10% FCS, 2 mM L-glutamine, 1% penicillin/streptomycin, and 50 μM 2-ME. For transient transfection, cells were centrifuged at 100 × g for 5 min at room temperature, resuspended in serum-free RPMI 1640 at 10 6/ml. After this, 3 × 10 6 cells were mixed with 3 μg, respectively, Firefly plasmid and 30 ng Renilla control plasmid pRL-CMV (Promega), electroporated at 250 V/960 μF, transferred into 5 ml preincubated media, and cultured for 24 h. Then Firefly and Renilla activities were measured using a dual assay kit, and the fold changes were calculated following the technical manual (Promega E2920).

Candidate regulatory elements were amplified using PCR. A full list of cloning primers is provided in Supplemental Table II. The Igl enhancers EA24 and EA31 were amplified and cloned into the BamHI site of pGL3 (Promega). The regions of interest of upstream of conventional enhancers, denoted ARE1, ARE2, and ARE3, were cloned individually in the SalI site of pGL3. The VA1 promoter was cloned into the XhoI/HindIII sites of the Igk enhancer-containing luciferase construct. The bARE1, bARE2, and bARE1 regions were amplified and blunt-end cloned into the BamHI site of pGL3 promoter, which contains a SV40 promoter. Cells were cotransfected with a Renilla expression plasmid for normalization and analyzed, as described previously (12).

Results

Epigenetic landscapes of AgR loci

Genome-wide patterns of histone modifications have been characterized for numerous cell types using ChIP, followed by high-throughput sequencing (ChIP-Seq) (13). Bioinformatic integration of these data has emerged as a powerful method for the functional assignment of genomic regions, including the identification of promoters, enhancers, and microRNA sites (14–16). However, histone modifications can play additional, specialized roles at genetic loci. For example, the H3K4me3 modification is a hallmark of active promoters, but, at AgR loci, this epigenetic mark also enhances binding to RAG-2, an essential component of the V(D)J recombination machinery (17). The specialized roles of histone modifications at certain loci may produce unique epigenetic patterns that are impossible to unravel with supervised segmentation methods. To decipher such novel patterns, unsupervised algorithms have been used (11, 18). For example, the epigenome of CD4+ T lymphocytes was segmented into 53 functional states, including active and repressed promoters, enhancers, and gene bodies (19). These approaches rely on statistical enrichment of specific combinations of chromatin marks throughout the genome to identify reproducible patterns. However, because genes represent the major organizational unit of the genome, the most robust patterns identified with current approaches correspond to promoter/gene bodies.

The unique segmented organization of AgR loci, coupled with genome-scale statistical analyses, could potentially mask AgR-specific patterns that are rare or nonexistent in the remaining epigenome. To circumvent these potential complications in our search for new regulatory elements, we restricted combinatorial analysis of chromatin features to data covering only the seven AgR loci (Igh, Igk, Igl, Tcra/d, Tcrα/δ, Tcrβ, and Tcrγ). All data sets were obtained from purified pro–B cells harboring germline AgR loci (RAG deficient), with the exceptions of Med1 and PU.1 association, which correspond to ChIP-Seq data from a transformed pro–B cell line (3).

We first calculated the coverage of individual features at the seven AgR loci (histone modifications, factor binding, transcription) compared with the entire genome. As shown in Fig. 1, the epigenetic landscape of AgR loci is distinguished from the rest of the genome in several important respects. First, Ig and Tcr loci display a much lower density of the repressive H3K27me3 modification in pro–B cells relative to the entire genome. The dearth of this epigenetic mark suggests that Polycomb-mediated repression is less pronounced at AgR loci, even when a locus is silent for transcription/recombination. Second, the density of H3K36me3, a modification associated with transcriptional elongation, as well as transcripts themselves (RNA-Seq), is substantially decreased in AgR loci relative to the whole genome. This finding most likely reflects the predominance of gene segments, rather than conventionally expressed genes in AgR loci, as well as the limited amounts of transcription arising from the four Tcr loci in pro–B cells. Third, despite lower overall transcription, signals for the mediator component, Med1, and the transcription factors PU.1 and E2A are increased several-fold relative to the entire epigenome, suggesting a higher density of regulatory sites, potentially corresponding to enhancers. Finally, transcription factors c-Myc and EBF, which have important functions in pro–B cells, show substantially lower peak densities compared with the whole genome. This implies that the binding sites for these factors are mostly located outside of AgR loci. Overall, these initial analyses indicate that the distribution of important chromatin features in AgR loci differs substantially from the remainder of the genome. Therefore, identification of novel regulatory regions within AgR loci will benefit from a more focused computational analysis of chromatin states tailored to these regions.

Chromatin profiling of AgR loci in pro–B cells

The complex structure of Ig and Tcr loci requires advanced computational analysis to identify major AgR-specific chromatin patterns. For this purpose, we applied the ChromHMM algorithm (11), focusing on only the seven AgR loci. The resulting chromatin states may then be used to identify active and poised regulatory elements in an unbiased manner. Briefly, ChromHMM...
utilizes a hidden Markov model that captures two types of information—the co-occurrence frequency of individual features at either the same location (emission probabilities) or adjacent locations (transition probabilities)—yielding patterns of chromatin features defined as characteristic states.

In total, we considered 19 distinct chromatin features in pro-B cells for this analysis (Fig. 2A), derived from published or new data sets, including histone modifications, key transcription factors, nucleosome density, and transcription (Supplemental Table I). Using ChromHMM, we compared individual emission probabilities for models in which the combinatorial number of states was varied from 20 to 40 and found that 38 states optimally described the epigenetic landscape of AgR loci. A higher model dimensionality produced redundancies, whereas distinct states, corresponding to active or poised regulatory elements, were merged when <38 were considered. Each state in the model corresponds to either a single feature or combinations of features, yielding an unbiased description of the AgR epigenetic landscape. A full list of states in the optimized model can be found in Supplemental Table III. The probabilistic relationship between chromatin features and an individual state is important to note. For example, state 13 is defined by simultaneous presence of H3K36me3, H3Ac, DNAse, E2A, Med1, and some other marks, all with probabilities of nearly 1 (dark blue, Fig. 2A), indicating that all state 13 regions have these chromatin features. However, the probability of observing p300 in state 13 is intermediate (0.4, light blue), reflecting the fact that only some state 13 regions associate with p300 (e.g., Eμ), whereas others do not.

For the 38-state model (Fig. 2A), AgR chromatin can be divided broadly into three nonredundant categories. The first category includes 12 chromatin states that are defined by the presence of a single feature, such as H3K4me1, H3K27me3, or Pax5, suggesting a limited regulatory potential for these regions in pro-B cells (states 15, 29, and 34, respectively). A second category...
corresponds to states associated with only two or three chromatin features, which may reflect a partially active, or poised, configuration (e.g., states 1, 2, 20, and 26). Finally, 6 of the 38 chromatin states show strong enrichment for multiple histone modifications or other features of active chromatin (states 3, 4, 5, 7, 8, and 13). Regions assigned to these chromatin states most likely harbor active regulatory elements because they are also nucleosome poor (DHS peaks) and have other modifications that characterize promoters or enhancers (H3K4me3, H3ac, etc.). Notably, state 4 is characterized by its robust enrichment for 10 chromatin features (hereafter, enrichment indicates that the probability for a feature is >0.5), including association with the transcription factors p300, PU.1, and Med1. Given these characteristics, regions within AgR loci categorized as state 4 most likely encompass cis elements with a high regulatory potential. Indeed, the chromatin states most highly enriched for activation features (states 3, 4, 5, and 13) are predominantly localized to Ig loci, particularly to Igk, which are more active (or poised) in pro–B cells compared with Tcr (Fig. 2B). Together, focused epigenetic analysis of AgR loci defines a unique set of chromatin states, some of which most likely reflect functionality in the context of gene regulation and recombination.

Chromatin state functions

To garner functional insights, we first assessed whether classes of known AgR elements segregate into different chromatin states. As shown in Fig. 2C, we parsed the AgR loci into seven functional categories corresponding to the following annotated regions (±500 bp): 1) Igk V segments (including upstream promoters); 2) Igk V segments plus promoters; 3) all D segments (Ig and Tcr); 4) all J segments; 5) all C regions; 6) all known enhancers; and 7) Pax5-activated intergenic repeat elements, a set of promoters that direct antisense transcription within specific regions of the Igk V cluster (20, 21). Strikingly, most of the annotated AgR regions segregate from one another into a small number of individual chromatin states, most likely reflecting the relationships between epigenetic features and their functionality. For example, Igk V regions, whose associated promoters exhibit varying degrees of activity in pro–B cells (22), belong to five different chromatin states. Conversely, Igk V regions belong to only two states (21, 22), distinct from those of Igk V, which display a highly restricted set of chromatin features, presumably reflecting their poised status in pro–B cells (H3K4me2 and PU.1). Moreover, most of the Pax5-activated intergenic repeat antisense promoters belong to chromatin state 3, which recapitulates known features of these regulatory elements, including their simultaneous association with Pax5, CTCF, and Rad21 (Fig. 2A, 2C). Most notably, 8 of the 12 known AgR enhancers belong to chromatin state 4, which is most enriched for activating features (Fig. 3 and see below). One exception, Eq, belongs to state 13 (Fig. 3A), most likely because of its dual function as a strong promoter. A second exception, Ey4, should be excluded from consideration because its epigenetic profile is masked by a proximal gene (Stard3nl) that is highly expressed in pro–B cells (Fig. 3F).

Notwithstanding, the vast majority of known AgR enhancers, whether active (3′ RR; see below) or inactive in pro–B cells (Eβ, Eα, Ey2, 3′Ex, Eα31, Eα24; Figs. 3, 4), were assigned to state 4 using this unbiased analysis. Based on the aforementioned characteristics, chromatin state 4 provides a focused set of candidates

---

**FIGURE 3.** Chromatin states for selected regions of Ig and Tcr loci. Tracks for the indicated epigenetic features (ChIP-Seq) or transcription (RNA-Seq) as visualized in the IGV browser. Annotations for known elements and their corresponding chromatin states are shown in the bottom two tracks. Genomic coordinates are shown above the tracks (build mm9). State 13 is characteristic of actively transcribed elements (highlighted in light green) and harbors two enhancers, Eq (A) and Ey4 (F). State 4, which is characterized by a lack of transcription, the presence of activating chromatin marks, and binding by E2A, Pax5, PU.1, p300, and Med1, coincides with most known AgR enhancers, including 3′Ex (B), Eβ (C), Eα (D), and Ey2 (E). Regions identified as chromatin state 4 are highlighted in gray.
for novel regulatory elements in the AgR cistrome. Overall, state 4 is limited to 41 segments, spanning 42 kb of the 9.2 Mb that encompasses all AgR loci (~0.4%). The priority status of state 4 as an identifier of enhancers is further supported by its enrichment for the p300 histone acetyltransferase, which is considered to be a general feature of enhancers, as well as its enrichment in three key transcription factors for pro–B cell gene expression programs (PU.1, E2A, and Pax5). Accordingly, most of the putative cis elements belonging to state 4 are situated in active (46% in Igh) or poised loci (36% in Igk and Igl), whereas only 17% of such regions are located in the four Tcr loci (Fig. 2B). Additionally, most of the state 4 regions are well conserved at the level of DNA sequence (average conservation score is 0.682 for 4; see Materials and Methods and Supplemental Tables III, IV). We conclude that chromatin state 4, which encompasses most of the known AgR enhancers whether active, poised, or inactive in pro–B cells, will provide a rich source of candidate elements for functional analyses.

Characterization of novel enhancers in Ig L chain loci

Leveraging the predictive power of our AgR chromatin analyses, we selected three state 4 regions from Igk or Igl for functional assays. These L chain loci exhibit only modest transcriptional activity in primary pro–B cells and most likely reside in a poised chromatin configuration (23). We first tested a state 4 element, situated in the large cluster of Vk gene segments, for potential enhancer function (Fig. 4A; kRE1). Expression of luciferase reporters harboring the SV40 promoter was robustly augmented (7-fold) in a pro–B cell line (63-12) upon inclusion of kRE1 (Fig. 4C). In contrast, this region was devoid of enhancer activity when tested in pro–T or plasma cell lines (P5424 and J558L, respectively). Thus, the new Igk cis element is a stage- and cell type-specific enhancer, suggesting a role in controlling the recombination potential of some mouse Vk gene segments.

The mouse Igl locus has two highly conserved enhancers, termed E13 and E24, located distal to each of the VlJl cassettes (24). In addition, we identified two state 4 regions lying even more distal to the VAJl cassettes (24). In contrast, we identified two state 4 regions lying even more distal to the VAJl cassettes (24). Indeed, in conjunction with the VAJl promoter, each of these regions augments reporter gene expression in pro–B and plasma cells, but not in a mouse pro–T cell line (Fig. 4D). In the J558L plasmacytoma, EARE1 also boosts the function of its nearby enhancer, EARE3, in an additive manner. As a control, the IglRE2 region (Fig. 4B), which associates with PU.1 but identifies with chromatin states 20 and 21, fails to augment reporter gene expression in either pro–B or plasma cells (Fig. 4D). Taken together, assignment as chromatin state 4 accurately predicts the location of novel enhancers in Ig L chain loci.
Characterization of a superenhancer in Igh

Transcription, V(D)J recombination, and Igh class switching are controlled by a set of enhancers and promoters, most of which have presumably been uncovered (26). Our chromatin analysis assigned the two classical Igh enhancers as state 13 (E\textsubscript{m} and hs4) (Fig. 5A). Several independent enhancers and a CTCF-rich region, together termed as 3′RR, have distinct but important functions in the B cell lineage, including transcription and recombination of adjacent DHJH gene segments (E\textsubscript{m}) and control of class-switch recombination (E\textsubscript{m} and 3′RR) (27). Notably, 3′RR-proximal region identified as state 4 is also enriched in CTCF (Fig. 5A), consistent with its role as insulator, yet indicating possibly more complex role played by this region. A third stretch within Igh, embedded between C\textgamma{}1 and C\textgamma{}2b, was recently described as a superenhancer (3). Young and colleagues (3) define superenhancers using several parameters, including an exaggerated intensity of Med1 and PU.1 binding relative to other ChIP-Seq peaks, implicating these regions as key regulatory elements controlling cell identity genes. To identify superenhancers, the authors find regions of overlap for master transcription factors, such as PAX5 and PU.1 in pro–B cells, which also colocalize with the most intense and broadest peaks for the general transcription factor, Med1. Accordingly, we performed an unbiased analysis of Med1 distributions using SICER (7), which allows accurate peak calling for broad chromatin features (see Materials and Methods). We discovered that the 3′RR, E\textsubscript{m}, and the new superenhancer, heretofore called Igh-SE, all appear as outliers in both width and read density for Med1, when compared with the entire epigenome (Fig. 5B).

FIGURE 5. Functional definition of a novel Igh superenhancer. (A) Tracks for the indicated chromatin features as visualized in the IGV browser (see Fig. 2). Chromatin states 3 (black), 4 (blue), 5 (green), and 13 (lavender) are shown in the bottom four tracks. The location of candidate enhancer elements hRE1 and hRE2 is highlighted as red boxes. (B) The distribution of Med1 peaks (identified using SICER) in pro–B cells by width (left panel) and read count-to-width ratio (right panel). Arrows indicate positions of Med1 peaks overlapping the three superenhancer regions within Igh, highlighting their extreme breadth (left panel) and read densities (right panel). (C) Luciferase data for hRE1 and hRE2, as described in Fig. 3C.

Importantly, our focused AgR chromatin analysis splits the Igh-SE into two active regions that belong to states 4 and 5 (Fig. 5A; hRE1 and hRE2, respectively). As shown in Fig. 5C, only the hRE1 region functions as an enhancer in pro–B cells when monitored by luciferase reporters, but is devoid of enhancer activity in pro–T or plasma cell lines. The other region, hRE2, belongs to state 5 and most likely corresponds to the C\textgamma{}2b germline promoter, which is active in pro–B cells based on its enrichment for H3K4me3 and the presence of sterile I\textgamma{}2b transcripts (Fig. 5A). The new hRE1 enhancer region is highly conserved (0.723) and interacts physically with other Igh regulatory elements (26), strongly suggesting an important but unknown function during the early stages of B cell development.

Discussion

We have used tailored computational approaches to assign chromatin states throughout all seven AgR loci in pro–B cells. Although the functional significance of many chromatin states remains to be defined, state 4 was found to accurately predict sites corresponding to AgR regulatory regions, both known and novel. The set of potential regulatory regions identified by state 4 also includes AgR enhancers that are inactive or only poised in pro–B cells (e.g., E\textbeta{} and E\textalpha{}24, respectively), broadening the scope of this chromatin-guided approach for enhancer discovery. Indeed, all state 4 regions tested in this study (4 of 4), which were derived from each of the three Ig loci, have enhancer activity in pro–B cells.

Strikingly, the only tested region that was found to be inactive, hRE2, was assigned to a separate chromatin state that is enriched...
Cg2b (28). We suspect that enhancer region hRE1 plays a role in which is clearly active in pro–B cells judging from RNA-seq and the germline Cg2b promoter (located near the hRE1 enhancer), portion of the 3
recombination to all 9
is thought to be the 3
mary enhancer region directing CSR in activated, mature B cells regulation of these elements, will produce a more complete picture of AgR elements, including hRE1, which can now be studied in vivo for their roles in Ig gene assembly, expression, isotype switching, and somatic hypermutation.

In summary, we have developed an unbiased epigenome-based approach to define the regulomes of AgR and other complex loci, such as those encoding NK cell receptors or MHC molecules (31). Whereas our functional validation of new enhancers focused on regions belonging to chromatin state 4, other states may also harbor important regulatory elements. These include state 5, which was enriched for promoters, and state 13, which spans Eu and a flanking portion of the 3’RR. Future validations, including targeted disruption of these elements, will produce a more complete picture of AgR regulomes in the context of lymphocyte development and activation.

Acknowledgments
We are grateful to Dr. Stephanie Kolar for helpful discussions.

Disclosures
The authors have no financial conflicts of interest.

References