



TIME WAITS FOR NO ONE

Enlist the experts at Bio X Cell for  
Antibody Production Services

EXPLORE

RECEIVE 10% OFF NOW with code: CONTRACT22JI



## 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

This information is current as of February 26, 2022.

*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**, 6 of which you can access for free at:  
<http://www.jimmunol.org/content/192/3/1064.full#ref-list-1>

**Why *The JI*? Submit online.**

- **Rapid Reviews! 30 days\*** from submission to initial decision
- **No Triage!** Every submission reviewed by practicing scientists
- **Fast Publication!** 4 weeks from acceptance to publication

*\*average*

**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,<sup>\*,†,1</sup> Suhasni Gopalakrishnan,<sup>\*,†,1</sup> Yue Huang,<sup>\*,†</sup> Jun Tang,<sup>\*,†</sup>  
Ann J. Feeney,<sup>‡,§</sup> Eugene M. Oltz,<sup>\*,†</sup> and Maxim N. Artyomov<sup>\*,†</sup>

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.

Many of the strategies employed by developing lymphocytes to regulate gene expression share features with mechanisms that control the stepwise assembly of Ag receptor (AgR) loci (1). Both processes require highly orchestrated interfacing between *cis*-regulatory elements, transcription factors, covalent modification of histones, changes in chromatin accessibility, and recruitment of machinery that drives transcription or recombination. In AgR loci, enhancer and promoter elements play crucial roles in modulating chromatin associated with V, D, and J segments to control their recombination potential at each stage of lymphocyte development (2). Accordingly, most of the *cis* elements associated with AgR loci are lineage and stage specific.

In addition to classical enhancers, recent studies identified a novel class of elements, termed superenhancers, which are thought to regulate the expression of genes that serve as primary determinants of cell identity (3, 4). Superenhancers are focal points for lineage-specifying transcription factors and for the ubiquitous mediator complex, which is required for activator-dependent gene expression. Moreover, superenhancers are centered within unusually large stretches of activating histone modifications, such as acetylation of histone <sup>3</sup>H at the lysine 27 position (H3K27Ac). Three regions harboring superenhancers have been identified within

the *Igh* locus (3). However, the collection of *cis* elements, known as the *cistrome*, which govern AgR gene assembly and expression during the early stages of lymphocyte development, remains incomplete. In this study, we identify novel enhancers within all three *Ig* loci, which exhibit activity specific for precursor B-lymphocytes, using focused computational analyses of publicly available and new chromatin data.

## 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 a custom-made microarray covering all AgR loci (ChIP-Chip; see below). Supplemental Table 1 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 (5) (version 0.12.7) using “-m 1 -v3 -best -strata” options. The resulting alignment SAM file was converted into read BED files using BEDTools. RNA-Seq data were aligned with TopHat (6) (version 1.4.1.1) using “-pre-filter-multihits-max-multihits 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

\*Department of Pathology, Washington University School of Medicine, St. Louis, MO 63110; <sup>†</sup>Department of Immunology, Washington University School of Medicine, St. Louis, MO 63110; <sup>‡</sup>Department of Immunology, The Scripps Research Institute, La Jolla, CA 92037; and <sup>§</sup>Department of Microbial Science, The Scripps Research Institute, La Jolla, CA 92037

<sup>1</sup>A.V.P. and S.G. contributed equally to this work.

Received for publication October 18, 2013. Accepted for publication November 20, 2013.

This work was supported by National Institutes of Health Grants AI 079732, AI 081224, CA 156690 (to E.M.O.), and AI 082918 (to A.J.F.).

Address correspondence and reprint requests to Dr. Maxim N. Artyomov and Dr. Eugene M. Oltz, Department of Pathology and Immunology, Washington University School of Medicine, 509 South Euclid Avenue, St. Louis, MO 63110. E-mail addresses: martyomov@pathology.wustl.edu (M.N.A.) and eoltz@pathology.wustl.edu (E.M.O.)

The online version of this article contains supplemental material.

Abbreviations used in this article: AgR, Ag receptor; ChIP, chromatin immunoprecipitation; CSR, class-switch recombination; DHS, DNase I hypersensitivity.

Copyright © 2014 by The American Association of Immunologists, Inc. 0022-1767/14/\$16.00

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, 19245000–19449000 (*Tcrp*); chr14, 52962000–54855000 (*Tcrad*); chr12, 114435000–117280000 (*Igh*); chr6, 67490000–70715000 (*Igk*); chr16, 18971000–19285000 (*Igl*). Values for the genome 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://artymovlab.wustl.edu/publications/supp\\_materials/AgR\\_2013/](https://artymovlab.wustl.edu/publications/supp_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; ab4729) 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<sup>-/-</sup>, p53<sup>-/-</sup> pro-T cell line), 63-12 (RAG-2<sup>-/-</sup> A-MuLV-transformed pro-B cell line), and J558L (B myeloma cell line). All cell lines were cultured at 37°C with 5% CO<sub>2</sub> 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<sup>7</sup>/ml. After this, 3 × 10<sup>6</sup> 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 Eλ24 and Eλ31 were amplified and cloned into the BamHI site of pGL3 (Promega). The regions of interest upstream of canonical enhancers, denoted λRE1, λRE2, and λRE3, were cloned individually in the SalI site of pGL3. The Vλ1 promoter was cloned into the XhoI/HindIII sites of the Igλ enhancer-containing luciferase constructs. The hRE1, hRE2, and κRE1 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<sup>+</sup> 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.

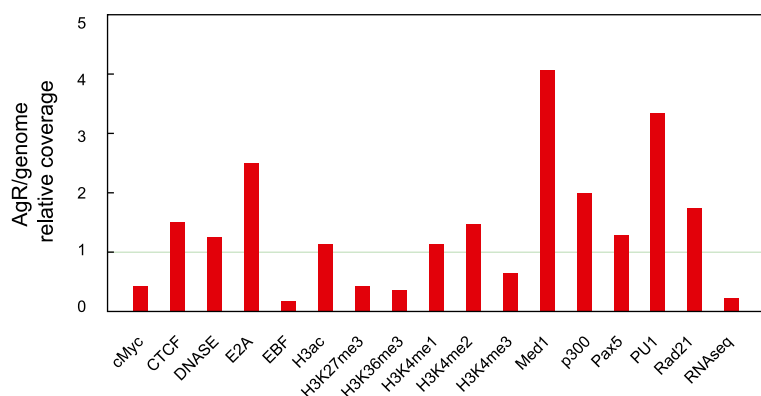
The unique segmented organization of AgR loci, coupled with genome-scale statistical analyses, could potentially mask AgR-specific patterns that are rare or nonexistent in the remaining epigenome. To circumvent these potential complications in our search for new regulatory elements, we restricted combinatorial analysis of chromatin features to data covering only the seven AgR loci (*Igh*, *Igk*, *Igl*, *Tcrad*, *Tcrb*, and *Tcrp*). 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

**FIGURE 1.** Unique epigenetic characteristics of mouse AgR loci. The y-axis represents the ratio of DNA space covered by a feature within AgR loci relative to the entire genome. A value of 1.0 corresponds to an equal distribution of that chromatin feature in AgR loci and the entire genome. Seventeen genome-wide ChIP-Seq, RNA-Seq, and DHS experiments were incorporated into the computational analyses.

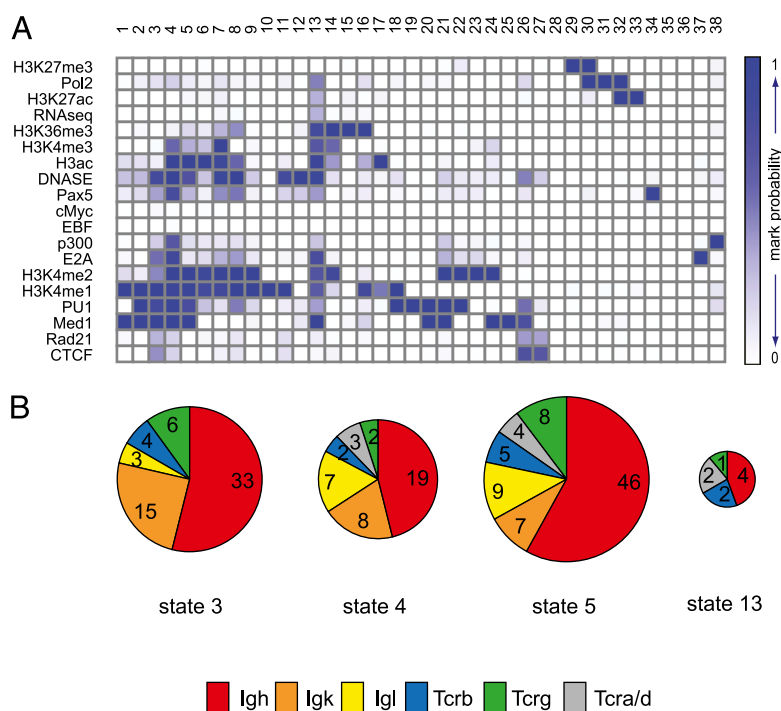


utilizes a hidden Markov model that captures two types of information—the co-occurrence frequency of individual features at either the same location (emission probabilities) or adjacent locations (transition probabilities)—yielding patterns of chromatin features defined as characteristic states.

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.,  $\mu$ ), 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



**FIGURE 2.** Unbiased characterization of the AgR epigenetic landscape. **(A)** The 38-state model of chromatin for AgR loci in pro-B cells. The hidden Markov model was based on the distribution of 19 chromatin features over all 7 AgR loci: 17 genome-wide features shown on Fig. 1 were narrowed to AgR and two features, Pol II and H3K27Ac, profiled by ChIP-Chip of AgR loci. The shades of blue represent numerically determined emission probabilities that range from 0.0 to 1.0 and describe the precise composition of each state, or the probability to find a certain chromatin mark or transcription factor in the region defined as a particular state. A mark is considered enriched in a particular state if its emission probability in the model is  $>0.50$ . **(B)** Distribution across AgR loci for states with the highest regulatory potentials. Pie charts are scaled to the total numbers of regions corresponding to each state. Significant enrichment of these chromatin states is observed for Ig loci, suggesting lineage-specific activities for these regulatory states. **(C)** Enrichment of individual states for specific AgR elements. Each chromatin state was evaluated for its composition with respect to the indicated elements ( $\pm 500$  bp from their annotated borders). Shades of blue correspond to hypergeometric probability of enrichment compared with random distribution across the entire collection of elements.



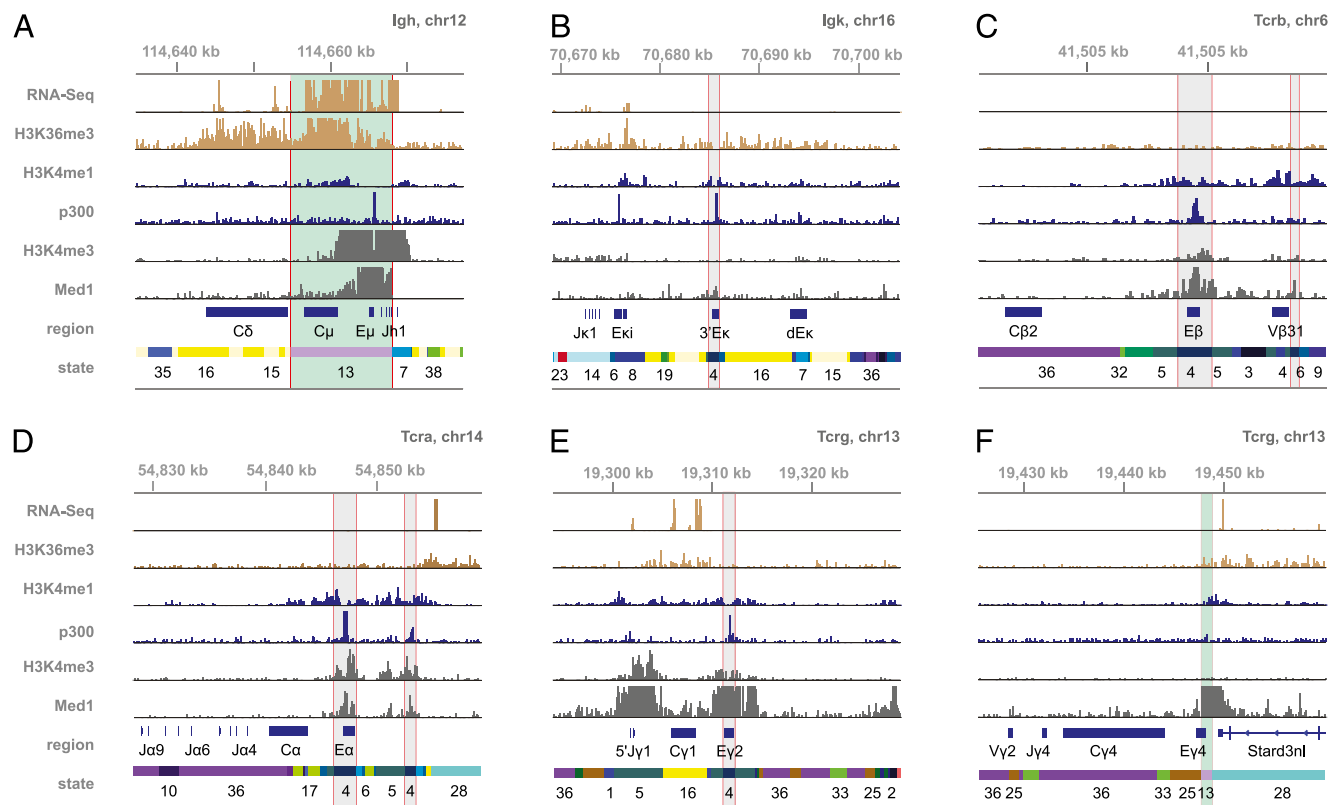
corresponds to states associated with only two or three chromatin features, which may reflect a partially active, or poised, configuration (e.g., states 1, 2, 20, and 26). Finally, 6 of the 38 chromatin states show strong enrichment for multiple histone modifications or other features of active chromatin (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 *Igh*, 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 ( $\pm 500$  bp): 1) *Igh* 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 *Igh* 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, *Igh* 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 *Igh* Vs, 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, E $\mu$ , belongs to state 13 (Fig. 3A), most likely because of its dual function as a strong promoter. A second exception, E $\gamma$ 4, 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 (E $\beta$ , E $\alpha$ , E $\gamma$ 2, 3'E $\kappa$ , E $\lambda$ 31, E $\lambda$ 24; 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, E $\mu$  (A) and E $\gamma$ 4 (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'E $\kappa$  (B), E $\beta$  (C), E $\alpha$  (D), and E $\gamma$ 2 (E). Regions identified as chromatin state 4 are highlighted in gray.



The mouse *Ig* locus has two highly conserved enhancers, termed E $\lambda$ 13 and E $\lambda$ 24, located distal to each of the V $\lambda$ J $\lambda$  cassettes (24). In addition, we identified two state 4 regions lying even more distal to the V $\lambda$ J $\lambda$  cassettes (Fig. 4B;  $\lambda$ RE1 and  $\lambda$ RE3), which are highly conserved (1.054 and 0.706, respectively). These regions may represent shadow enhancers, which are suspected to serve as booster or redundancy elements for the regulation of many genes (25). Indeed, in conjunction with the V $\lambda$ 1 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,  $\lambda$ RE1 also boosts the function of its nearby enhancer, E $\lambda$ 31, in an additive manner. As a control, the  $\lambda$ RE2 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*

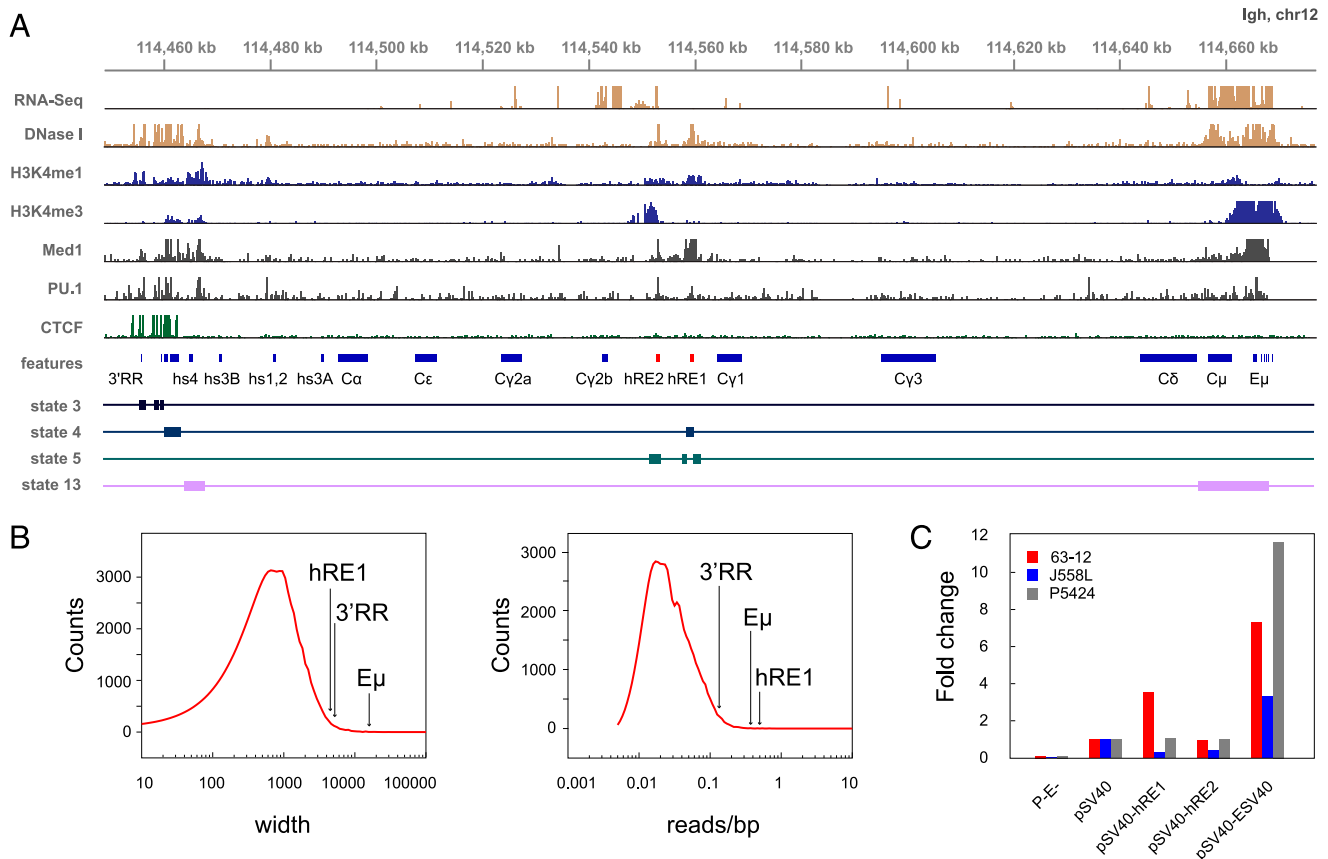
Transcription, V(D)J recombination, and *Igh* class switching are controlled by a set of enhancers and promoters, most of which have presumably been uncovered (26). Our chromatin analysis assigned the two classical *Igh* enhancers as state 13 ( $E\mu$  and  $hs4$ ) (Fig. 5A). Several independent enhancers 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 ( $E\mu$ ) and control of class-switch recombination ( $E\mu$  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  $C\gamma 1$  and  $C\gamma 2b$ , 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,  $E\mu$ , 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).

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  $C\gamma 2b$  germline promoter, which is active in pro-B cells based on its enrichment for H3K4me3 and the presence of sterile  $I\gamma 2b$  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.,  $E\beta$  and  $E\lambda 24$ , 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



**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.

for *Igh* V region promoters (state 5). This region corresponds to the germline Cg2b promoter (located near the hRE1 enhancer), which is clearly active in pro-B cells judging from RNA-seq and ChIP-seq data (H3K4me3). Indeed, recent studies have shown that pro-B cells can execute class-switch recombination (CSR) to Cg2b (28). We suspect that enhancer region hRE1 plays a role in stabilizing the active conformation of *Igh* required for CSR (29), V(D)J recombination (26), or both in precursor B cells. The primary enhancer region directing CSR in activated, mature B cells is thought to be the 3'RR. However, deletion of 3'RR abrogates recombination to all *Igh* isotypes except Cg1, the C region lying most proximal to the hRE1 enhancer (30). As such, our chromatin state analysis of pro-B cells provides at least four new enhancer 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 E $\mu$  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

- Osipovich, O., and E. M. Oltz. 2010. Regulation of antigen receptor gene assembly by genetic-epigenetic crosstalk. *Semin. Immunol.* 22: 313–322.
- Degner-Leisso, S. C., and A. J. Feeney. 2010. Epigenetic and 3-dimensional regulation of V(D)J rearrangement of immunoglobulin genes. *Semin. Immunol.* 22: 346–352.
- Whyte, W. A., D. A. Orlando, D. Hnisz, B. J. Abraham, C. Y. Lin, M. H. Kagey, P. B. Rahl, T. I. Lee, and R. A. Young. 2013. Master transcription factors and mediator establish super-enhancers at key cell identity genes. *Cell* 153: 307–319.
- Lovén, J., H. A. Hoke, C. Y. Lin, A. Lau, D. A. Orlando, C. R. Vakoc, J. E. Bradner, T. I. Lee, and R. A. Young. 2013. Selective inhibition of tumor oncogenes by disruption of super-enhancers. *Cell* 153: 320–334.
- Langmead, B., C. Trapnell, M. Pop, and S. L. Salzberg. 2009. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol.* 10: R25.
- Trapnell, C., L. Pachter, and S. L. Salzberg. 2009. TopHat: discovering splice junctions with RNA-Seq. *Bioinformatics* 25: 1105–1111.
- Zang, C., D. E. Schones, C. Zeng, K. Cui, K. Zhao, and W. Peng. 2009. A clustering approach for identification of enriched domains from histone modification ChIP-Seq data. *Bioinformatics* 25: 1952–1958.
- Trapnell, C., A. Roberts, L. Goff, G. Pertea, D. Kim, D. R. Kelley, H. Pimentel, S. L. Salzberg, J. L. Rinn, and L. Pachter. 2012. Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks. *Nat. Protoc.* 7: 562–578.
- Song, J. S., W. E. Johnson, X. Zhu, X. Zhang, W. Li, A. K. Manrai, J. S. Liu, R. Chen, and X. S. Liu. 2007. Model-based analysis of two-color arrays (MA2C). *Genome Biol.* 8: R178.
- Quinlan, A. R., and I. M. Hall. 2010. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* 26: 841–842.
- Ernst, J., and M. Kellis. 2012. ChromHMM: automating chromatin-state discovery and characterization. *Nat. Methods* 9: 215–216.
- Gopalakrishnan, S., K. Majumder, A. Predeus, Y. Huang, O. I. Koues, J. Verma-Gaur, S. Loguerio, A. I. Su, A. J. Feeney, M. N. Artyomov, and E. M. Oltz. 2013. Unifying model for molecular determinants of the preselection V $\beta$  repertoire. *Proc. Natl. Acad. Sci. USA* 110: E3206–E3215.
- Barski, A., S. Cuddapah, K. Cui, T. Y. Roh, D. E. Schones, Z. Wang, G. Wei, I. Chepelev, and K. Zhao. 2007. High-resolution profiling of histone methylations in the human genome. *Cell* 129: 823–837.
- Abeel, T., Y. Van de Peer, and Y. Saey. 2009. Toward a gold standard for promoter prediction evaluation. *Bioinformatics* 25: i313–i320.
- Yip, K. Y., C. Cheng, N. Bhardwaj, J. B. Brown, J. Leng, A. Kundaje, J. Rozowsky, E. Birney, P. Bickel, M. Snyder, and M. Gerstein. 2012. Classification of human genomic regions based on experimentally determined binding sites of more than 100 transcription-related factors. *Genome Biol.* 13: R48.
- Zhang, S., Q. Li, J. Liu, and X. J. Zhou. 2011. A novel computational framework for simultaneous integration of multiple types of genomic data to identify microRNA-gene regulatory modules. *Bioinformatics* 27: i401–i409.
- Matthews, A. G., A. J. Kuo, S. Ramon-Maiques, S. Han, K. S. Champagne, D. Ivanov, M. Gallardo, D. Carney, P. Cheung, D. N. Ciccone, et al. 2007. RAG2 PHD finger couples histone H3 lysine 4 trimethylation with V(D)J recombination. *Nature* 450: 1106–1110.
- Hoffman, M. M., O. J. Buske, J. Wang, Z. Weng, J. A. Bilmes, and W. S. Noble. 2012. Unsupervised pattern discovery in human chromatin structure through genomic segmentation. *Nat. Methods* 9: 473–476.
- Ernst, J., and M. Kellis. 2010. Discovery and characterization of chromatin states for systematic annotation of the human genome. *Nat. Biotechnol.* 28: 817–825.
- Ebert, A., S. McManus, H. Tagoh, J. Medvedovic, G. Salvaggio, M. Novatchkova, I. Tamir, A. Sommer, M. Jaritz, and M. Busslinger. 2011. The distal V(H) gene cluster of the *Igh* locus contains distinct regulatory elements with Pax5 transcription factor-dependent activity in pro-B cells. *Immunity* 34: 175–187.
- Verma-Gaur, J., A. Torkamani, L. Schaffer, S. R. Head, N. J. Schork, and A. J. Feeney. 2012. Noncoding transcription within the *Igh* distal V(H) region at PAIR elements affects the 3D structure of the *Igh* locus in pro-B cells. *Proc. Natl. Acad. Sci. USA* 109: 17004–17009.
- Choi, N. M., S. Loguerio, J. Verma-Gaur, S. C. Degner, A. Torkamani, A. I. Su, E. M. Oltz, M. Artyomov, and A. J. Feeney. 2013. Deep sequencing of the murine IgH repertoire reveals complex regulation of nonrandom v gene rearrangement frequencies. *J. Immunol.* 191: 2393–2402.
- Mercer, E. M., Y. C. Lin, C. Benner, S. Hjunhwan, J. Dutkowski, M. Flores, M. Sigvardsson, T. Ideker, C. K. Glass, and C. Murte. 2011. Multilineage priming of enhancer repertoires precedes commitment to the B and myeloid cell lineages in hematopoietic progenitors. *Immunity* 35: 413–425.
- Hagman, J., C. M. Rudin, D. Haasch, D. Chaplin, and U. Storb. 1990. A novel enhancer in the immunoglobulin lambda locus is duplicated and functionally independent of NF kappa B. *Genes Dev.* 4: 978–992.
- Hobert, O. 2010. Gene regulation: enhancers stepping out of the shadow. *Curr. Biol.* 20: R697–R699.
- Medvedovic, J., A. Ebert, H. Tagoh, I. M. Tamir, T. A. Schwickert, M. Novatchkova, Q. Sun, P. J. Huis In 't Veld, C. Guo, H. S. Yoon, et al. 2013. Flexible long-range loops in the VH gene region of the *Igh* locus facilitate the generation of a diverse antibody repertoire. *Immunity* 39: 229–244.
- Perlot, T., and F. W. Alt. 2008. Cis-regulatory elements and epigenetic changes control genomic rearrangements of the IgH locus. *Adv. Immunol.* 99: 1–32.
- Kumar, S., R. Wuerffel, I. Achour, B. Lajoie, R. Sen, J. Dekker, A. J. Feeney, and A. L. Kenter. 2013. Flexible ordering of antibody class switch and V(D)J joining during B-cell ontogeny. *Genes Dev.* 27: 2439–2444.
- Han, J. H., S. Akira, K. Calame, B. Beutler, E. Selsing, and T. Imanishi-Kari. 2007. Class switch recombination and somatic hypermutation in early mouse B cells are mediated by B cell and Toll-like receptors. *Immunity* 27: 64–75.
- Vincent-Fabert, C., R. Fiancette, E. Pinaud, V. Truffinet, N. Cogné, M. Cogné, and Y. Denizot. 2010. Genomic deletion of the whole IgH 3' regulatory region (hs3a, hs1.2, hs3b, and hs4) dramatically affects class switch recombination and Ig secretion to all isotypes. *Blood* 116: 1895–1898.
- Shiina, T., H. Inoko, and J. K. Kulski. 2004. An update of the HLA genomic region, locus information and disease associations: 2004. *Tissue Antigens* 64: 631–649.