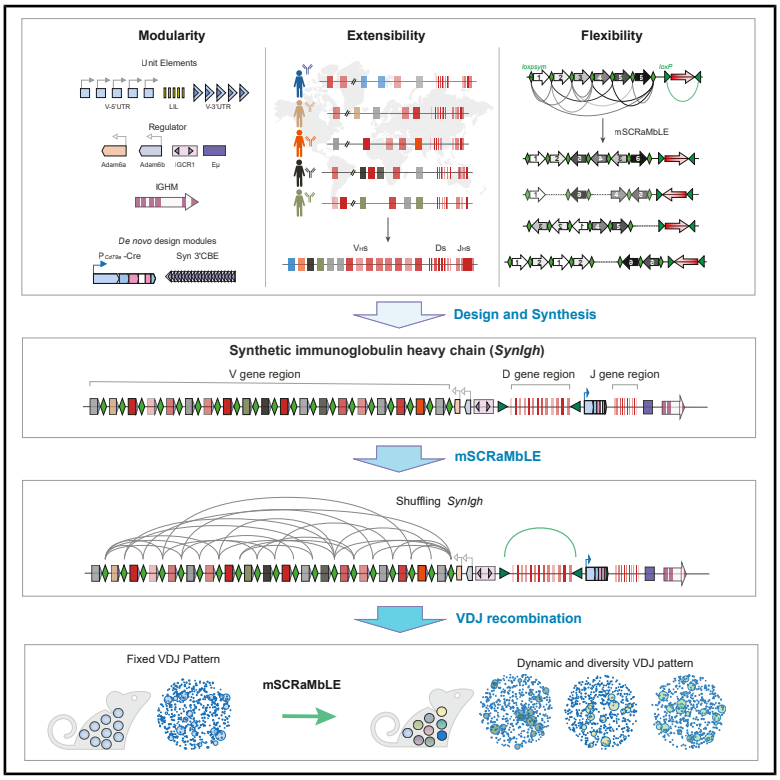
Cell Genomics

Article

Synthetic rewriting of the IGH locus

Graphical abstract



Highlights

- A human immunoglobulin heavy chain locus was designed and synthesized *de novo*
- Synthetic somatic genomic rearrangement was induced in living *Synlgh* mice
- Programmable shuffling expands the diversity of V(D)J recombination outcomes

Authors

Zong-Heng Fu, Yuan Ma, Si Cheng, ..., Zhaoqing Ba, Guang-Rong Zhao, Yi Wu

Correspondence

grzhao@tju.edu.cn (G.-R.Z.), yi.wu@tju.edu.cn (Y.W.)

In brief

Fu et al. create a *de novo*-designed synthetic human immunoglobulin heavy chain locus in mice and show that programmable locus shuffling expands V(D)J recombination diversity, establishing a generalizable framework for mammalian genome rewriting.

Fu et al., 2026, Cell Genomics *6*, 101252
June 10, 2026
https://doi.org/10.1016/j.xgen.2026.101252

CellPress

Cell Genomics

CellPress
OPEN ACCESS

Article

Synthetic rewriting of the IGH locus

Zong-Heng Fu,[1,2] Yuan Ma,[1,2] Si Cheng,[1,2] Yuan Zhong,[1,2] Nan Wang,[1,2] Zhaoqing Ba,[3,4] Guang-Rong Zhao,[1,2,*] and Yi Wu[1,2,5,*]

[1]State Key Laboratory of Synthetic Biology, Tianjin University, Tianjin 300072, China
[2]Frontiers Science Center for Synthetic Biology (Ministry of Education), School of Synthetic Biology and Biomanufacturing, Tianjin University, Tianjin 300072, China
[3]National Institute of Biological Sciences, Beijing 102206, China
[4]Tsinghua Institute of Multidisciplinary Biomedical Research, Tsinghua University, Beijing 100084, China
[5]Lead contact
*Correspondence: grzhao@tju.edu.cn (G.-R.Z.), yi.wu@tju.edu.cn (Y.W.)
https://doi.org/10.1016/j.xgen.2026.101252

SUMMARY

V(D)J recombination, which joins V, D, and J genes across immunoglobulin loci, is essential for antibody diversity. To unlock the inherent constraints of V(D)J recombination, we systematically designed, assembled, and shuffled a synthetic human immunoglobulin heavy chain locus (*SynIgh*). *SynIgh* was designed *de novo*, incorporating 60 V_H, 31 D, and 13 J_H from global populations, along with 63 synthetic rearrangement sites, endowing it with modularity, extensibility, and dynamicity. We then synthesized the repetitive *SynIgh* sequence with a stabilizer. Using this construct, we generated a *SynIgh* mouse model capable of undergoing V(D)J recombination. Moreover, we demonstrated that a dual-rearrangement strategy, combining synthetic rearrangement with natural V(D)J recombination, enables shuffling of the *SynIgh* locus in mice. This strategy expands the diversity of V(D)J recombination outcomes, thereby reconfiguring the process of antibody diversity generation. *SynIgh* establishes a generalizable framework for the design, building, testing, and learning of mammalian genome architecture and function.

INTRODUCTION

Variable (V), diversity (D), and joining (J) genes (V(D)J) recombination within the immunoglobulin (IG) locus is the fundamental process that initiates the generation of diverse antibodies, enabling the immune system to combat a wide range of pathogens.[1,2] During this process, the recombinase RAG (encoded by Recombination-activating gene 1 and 2) scans the immunoglobulin gene locus, selecting discrete V, D, and J gene segments to form V(D)J recombination products.[2–7] In recent years, substantial progress has been made in the sequencing of *Ig* loci and elucidation of the mechanisms of V(D)J recombination. Advances in sequencing technology have facilitated the resolution of immunoglobulin heavy chain locus (*IGH*) genetic variation across different populations,[8–14] including the identification of numerous V, D, and J genes and alleles, which present distinct antigen-binding epitopes and alter humoral immune responses to pathogens[15,16] and vaccine priming[17] in different populations. More importantly, *IGH* is highly enriched for large structural variants (SVs), recent haplotype-resolved sequencing has revealed that the IGH locus is driven by extensive SVs, particularly large-scale deletions spanning hundreds of kilobases that result in widespread gene copy number variations and the homozygous loss of functional genes.[14] Paired sequencing of *Ig* loci and the antibody repertoire has shown that these genetic variations influence the presence and frequency of genes in V(D)J recombination products, shaping the antibody repertoire.[14,18] To elucidate the mechanisms of V(D)J recombination, researchers have utilized mouse models with genetic modifications targeting the *Igh* locus to identify and clarify the functions of the intronic enhancer (Eμ),[19] the intergenic control region 1 (IGCR1),[6] and 3′ CTCF-binding elements (3′CBE).[20] These discoveries revealed chromatin loop extrusion during physiological V(D)J recombination at the *Igh* locus.[2–6,21] In this process, RAG binds to a J gene segment of heavy chain (J_H) recombination signal sequence (RSS) within the recombination center (RC) and then linearly scans upstream D- and V_H (V heavy gene segment)-containing chromatin via cohesin-mediated loop extrusion. Directional scanning leads RAG to selectively utilize RSS flanking D segments (D-RSSs) that are convergently oriented with RSS flanking J_H segments (J_H-RSSs) to generate DJ_H recombination products and subsequently promotes the orientational rearrangement of V_H genes during the V_H-to-DJ_H recombination step to generate V(D)J recombination products. While the gene composition and positioning of gene segments in the natural IG loci shape the patterns of V(D)J recombination, their finite number and predetermined locations present inherent constraints to the potential diversity of V(D)J recombination products.[5,14,21,22] The systematic *de novo* design and synthesis of IG loci leveraging concepts and approaches of synthetic genomics could offer a promising solution for systematic study of the *IGH* locus and enhancing V(D)J recombination diversity by altering gene composition and positioning.

Here, we have *de novo* design, assembly, and shuffling of the synthetic *IGH* locus (*SynIgh*). Compared with the natural *IGH* locus, *SynIgh* (1) features a fully modular, bottom-up hierarchical design based on studies of natural loci; (2) incorporates V, D, and J genes from diverse global human populations; and (3) includes an inducible synthetic genomic rearrangement system. Together, these characteristics endow the *SynIgh* locus with modularity, extensibility, and dynamic flexibility. To synthesize and assemble the highly repetitive sequences of *SynIgh*, we developed a "stabilizer" fusion protein that binds to frequently occurring DNA motifs within the repetitive regions in the assembly host. We then generated a *SynIgh* mouse model via pronuclear injection of repetitive DNA and tested its functions in V(D)J recombination. We further investigated the capacity of synthetic rearrangement to shuffle the *SynIgh* locus *in vivo* and assessed its impact on V(D)J recombination.

RESULTS

Design of the human *SynIgh* locus

Recent advances in identifying genetic variation within *Ig* loci and uncovering their recombination mechanisms provided the basis for the systematic *de novo* design of the *SynIgh* locus. Here, we designed the *SynIgh* with the following aims: (1) validating the current knowledge of these regulatory elements while deepening and systematizing the study of the *Igh* locus and (2) exploring the potential to overcome the limitations of gene composition and positioning in the natural locus, enhancing V(D)J diversity by broadening the gene content and altering the bias in inherent recombination patterns.

To achieve these two objectives, we propose three principles for designing *SynIgh*. (1) Modularity: the *SynIgh* locus is designed using well-defined modular elements, allowing flexibility in manipulating individual elements and making it effective to validate the existing regulatory function. (2) Extensibility: incorporation of a wider variety of V, D, and J genes from diverse human populations allows more possible combinations of V, D, and J segments, thereby expanding the set of possible V(D)J recombination products. (3) Flexibility: to decouple the inherent constraint on V(D)J recombination patterns shaped by the fixed structure of the natural locus, we introduced a dynamic genomic rearrangement system called mSCRaMbLE (mammalian synthetic chromosome rearrangement and modification by *LoxP*-mediated evolution). By integrating multiple site-specific recombination sites between each module in *SynIgh*, this system enables the generation of structural variants within the *SynIgh* locus, creating a large library of structural configurations for subsequent V(D)J recombination (Figure 1).

Under the guidance of the above design principle of modularity, *SynIgh* was designed with a hierarchical structure comprising two types of regions, namely, gene regions and regulatory regions. The gene regions include the V, D, J, and C (constant region) gene clusters of individual units. Each V gene unit, from 5′ to 3′, includes the 5′ untranslated region upstream of each V gene segment (the V-5′ UTR), signal peptide (L-I-L), coding sequence (CDS) of V gene, and the 3′ untranslated region downstream of each V gene segment (V-3′ UTR) containing the RSS (Figures S1A–S1C). Similarly, each D gene unit is composed of the D-5′ UTR, CDS of D gene, and D-3′ UTR. Each J gene unit consists of the J-5′ UTR, CDS of J gene, and J-3′ UTR (containing splicing sites). The C gene region is composed of the full-length sequences of exons and introns that serve as the constant region. The regulatory regions, which are fundamental for maintaining the proper functionality of the *IGH* locus, include the natural mouse IGCR1,[6] Eμ,[19] and Adam6a/b[23] modules, as well as the synthetic 3′CBE module,[20,21] which consists of 20 tandem CBE motifs to simplify the long natural 3′CBE sequence (Figure S1D). Overall, *SynIgh* is arranged as follows: V_H gene region, Adam6a/b, IGCR1, D gene region, J_H gene region, Eμ, C gene region, Syn 3′CBE. Following the expansion principle, the CDSs of the V, D, and J genes were selected from the intersection of the 1000 Genomes Project (1kGP)[9,13,24] and IMGT (the international ImMunoGeneTics information system) databases.[25] We selected 60 V_H, 31 D, and 13 J_H genes that are frequently presented across global populations for *SynIgh* (Table S1).

For flexibility design principle, we drew inspiration from the yeast SCRaMbLE system, which introduces numerous symmetric *loxP* (*loxPsym*) sites into the synthetic genome, permitting genomic rearrangements between *loxPsym* sites in the presence of Cre recombinase.[26–30] We developed a mSCRaMbLE system for *SynIgh*, in which each V gene unit is separated by a *loxPsym* site, with a total of 61 *loxPsym* sites across the V gene region. Thus, site-specific recombination can occur between these *loxPsym* sites, inducing inversions and deletions, shuffling the V region, and leading to the random loss of V units and changes in their linear positions. In addition, to enhance the inclusion of rare but valuable inverted D (Inv-D) gene products in V(D)J recombination,[3,31] we incorporated 2 *loxP* sites in opposite orientations flanking the D gene region. These *loxP* sites, orthogonal to *loxPsym*,[32] enable the inversion of the entire D region in the presence of Cre recombinase, allowing Inv-D gene to participate in loop extrusion-mediated recombination, further expanding the diversity of V(D)J recombination outcomes.

In summary, guided by the principles of modularity, extensibility, and flexibility design, we have *de novo* designed a synthetic human immunoglobulin heavy chain locus, *SynIgh*. Compared with the 1.2 megabase (Mb) natural human *IGH* locus (Figure S1E), the 174 kilobase (kb) *SynIgh* has simplified intergenic regions, pseudogenes, and noncoding RNAs. *SynIgh* incorporates 60 V_H, 31 D, and 13 J_H genes from diverse global populations. *SynIgh* features 61 *loxPsym* sites and 2 *loxP* sites, enabling synthetic rearrangement in the presence of Cre recombinase to generate diverse structural variations.

Synthesis of highly repetitive *SynIgh* with a stabilizer

The highly modular structure of *SynIgh* involves the inclusion of numerous repetitive sequences, particularly in the V gene region. The most widely used methods for DNA assembly are Gibson[33] and yeast assembly.[34] However, the assembly and maintaining of large repetitive sequence is still a grand challenge because of its high instability.[35–37] The V gene region contains 60 modular V gene units, each featuring highly similar CDSs as well as UTRs (V-5′ UTR and V-3′ UTR). The pairwise similarity in the CDSs ranges from 51.55% to 99.67% and the average similarity for the UTR modules is 39.46% (Figure 2A). We employed an iterative assembly strategy (Figure 2B), starting with level-1

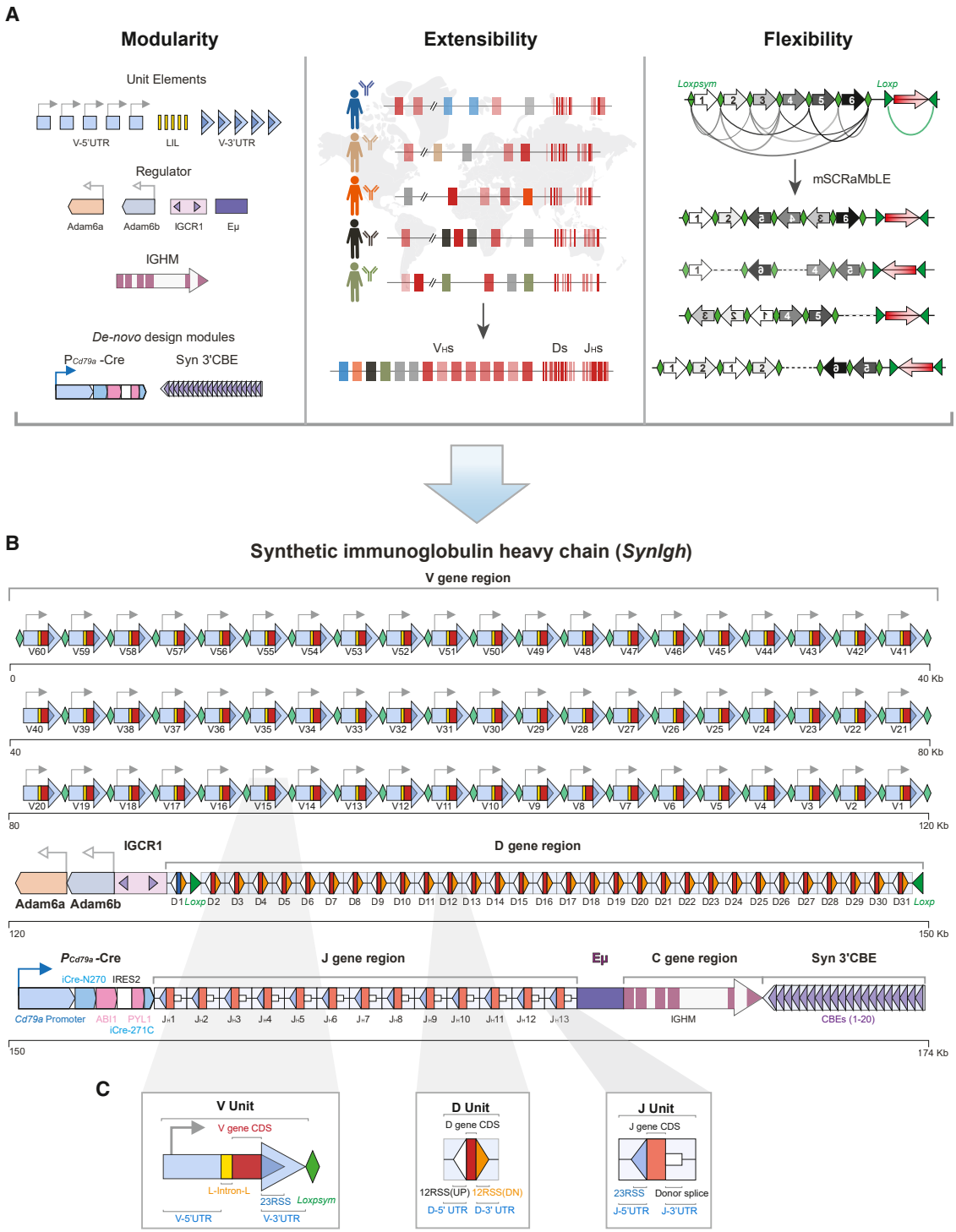


Figure 1. Design of the human *Synlgh* locus

(A) The three design principles (modularity, extensibility, and flexibility).

(B) Schematic of the synthetic immunoglobulin heavy chain locus (*Synlgh*).

(C) The design of V, D, and J units.

See also Figure S1.

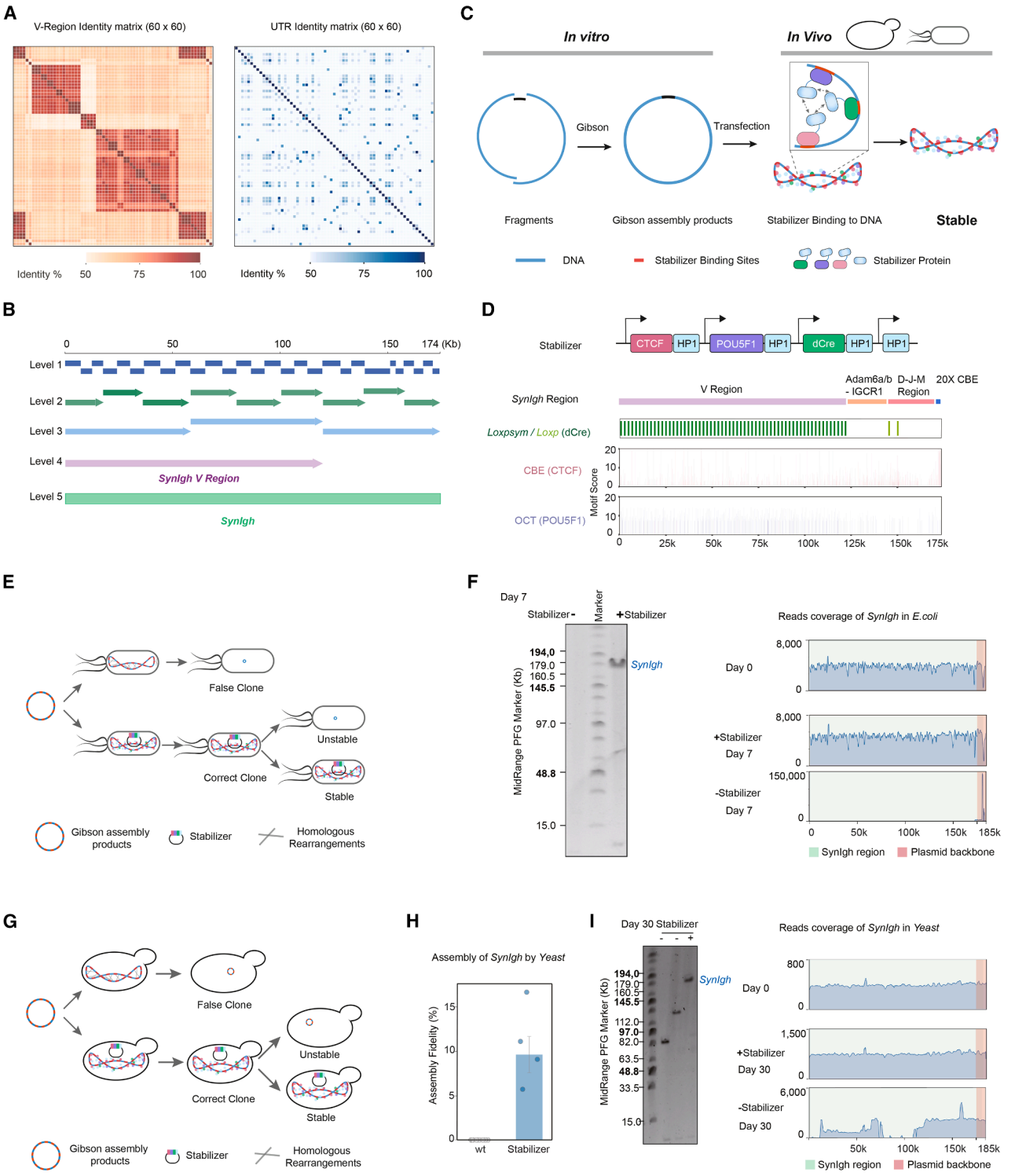


Figure 2. Synthesis of *SynIgh* with a stabilizer

(A) Homology analysis of V region in the *SynIgh*.

(B) Iterative assembly strategy for the *SynIgh* locus.

(C) Assembly process of *SynIgh* locus.

(D) Stabilizer and its binding sites in the *SynIgh* V region.

(E) Schematic illustration of the stabilizer-based assembly strategy in *E. coli*.

(legend continued on next page)

fragments of 6–8 kb, each containing 3–4 highly similar V gene CDSs. However, our attempts to use standard Gibson and yeast assembly methods expectedly failed to produce correct repetitive constructs larger than 100 kb.

In nature, the genetic stability of large-scale repetitive DNA sequences, such as centromeres, is typically supported by specialized chromatin architectures, primarily by sterically excluding bulky recombination machineries.[38–40] Inspired by these natural paradigms, we designed a synthetic "stabilizer" to maintain the integrity of assembled large repetitive *SynIgh* sequences. This stabilizer is comprised of several DNA-binding proteins, which target multiple sites within the repetitive array, and an effector heterochromatin protein 1 alpha (HP1α) capable of driving DNA compaction (Figure 2C). HP1α functions as an autonomous compaction module that utilizes its intrinsically disordered regions to drive multivalent DNA interactions.[41,42] *SynIgh* contains 61 *loxPsym* sites, 129 highly similar OCT (Octamer-binding transcription factor) motifs (motif scores >10), and 94 highly similar CBE motifs (motif scores >10). Therefore, we utilized the corresponding binding proteins dCre (dead/inactive Cre) for *loxpSym*,[43] POU5F1 (POU class 5 homeobox 1) for OCT,[44] and CTCF for CBE,[45,46] fused with the effector HP1α[47] to design a stabilizer for *SynIgh* (Figure 2D).

To test the ability of the designed stabilizer for the construction of large repetitive sequence in *E. coli*, we assembled the 174 kb *SynIgh* from three ∼60 kb fragments via Gibson assembly *in vitro* and then transformed the products into recombinant strains which harbored the stabilizer expression plasmid (Figure 2E). In earlier experiments, we failed to obtain correct clones (individual *E. coli* transformant colonies that were PCRTag positive at all expected assembly junctions) when we transformed these assembly products into *E. coli* in the absence of the stabilizer. In contrast, using the stabilizer, we screened 32 clones in a single pilot experiment and identified one clone with all correct junctions (18 PCRTags). Further analysis with pulsed-field gel electrophoresis (PFGE) confirmed that the construct size was approximately 174 kb. Next-generation sequencing (NGS) revealed uniform coverage compared with the reference sequence, with no significant deletions observed (Figure 2F). To further verify the ability of the stabilizer to maintain the long-term stability of large-scale repetitive DNA constructs, we passaged the clone containing *SynIgh* for 7 days either with or without the stabilizer expression. After 7 days, under the stabilizer selection condition, NGS revealed uniform coverage across the *SynIgh* region, similar to that of the starting clone (Figure 2F). In contrast, in the condition without the stabilizer selection, NGS sequencing revealed no reads covering the *SynIgh* region. We then sequenced the plasmid using high-depth (∼1,000× coverage) Oxford Nanopore sequencing. No evidence of large-scale rearrangements was observed relative to the reference structure, with 12 single-nucleotide variants (SNVs) and 14 indels identified (Figures S2A–S2E; Tables S2 and S3). These results show that the stabilizer-based assembly strategy can be used to assemble large repetitive sequences (>100 kb, *SynIgh*) and maintain their long-term genetic stability.

Next, to test the universality of the stabilizer, we transformed the stabilizer plasmid, which is based on the pRS42H[48] shuttle vector that replicates in both *E. coli* and *S. cerevisiae* and uses universal promoters[49] to drive stabilizer protein expression, into the yeast strain BY4742 (Figure 2G). Then, we transformed the Gibson assembly products into the stabilizer-carrying yeast strain, with the wild-type BY4742 serving as the control. In the control group, we failed to obtain any correct clones among a total of ∼1,000 clones screened. In contrast, in the stabilizer group, approximately 10% of the clones were correct (Figure 2H). Furthermore, we passaged a yeast clone containing the 174 kb *SynIgh* for 30 days. In the culture without a stabilizer, PFGE analysis revealed the *SynIgh* size reductions, and NGS revealed several deletions compared with the initial clone. In contrast, in the stabilizer group, the expected *SynIgh* size on PFGE was maintained with no significant variations, and NGS confirmed uniform coverage across the *SynIgh* region consistent with that of the starting clone (Figure 2I). In summary, the introduction of a stabilizer that binds to frequent DNA motifs within repetitive sequences enabled the successful assembly and maintaining of 174 kb *SynIgh* in *E. coli* and *S. cerevisiae*, a process that is challenging using conventional methods.

SynIgh mice undergo V(D)J recombination

To evaluate the functionality of the *SynIgh* locus, we generated a mouse model via pronuclear injection of the assembled *SynIgh* (174 kb) plasmid with *piggyBac* transposase mRNA into fertilized C57BL/6J eggs (Figure 3A). We identified *SynIgh*$^+$ F0 generation mice by PCR and nanopore sequencing. ITR (*PiggyBac* Inverted Terminal Repeats)-containing reads mapped to a single insertion locus and reads across *SynIgh* were consistent with the expected structure. (Figures 3B and S3A–S3C).

To test whether the *SynIgh* locus in mice can undergo V(D)J recombination, we isolated bone marrow cells from *SynIgh*$^+$ mice and extracted DNA for VDJ-seq analysis with human J_H gene baits.[50] We detected 7,971 unique VDJ recombination products utilizing all 60 V_H genes, 31 D genes, and 13 J_H genes in the *SynIgh* locus (Figures 3C and 3D). Moreover, the V gene usage in V(D)J recombination products were highly similar among individual mice (n = 4, PCC [Pearson Correlation Coefficient] = 0.9 ± 0.04) (Figures 3E and 3F). These results indicate that the *SynIgh* locus can undergo V(D)J recombination utilizing all of the designed V_H, D, and J_H genes to produce V(D)J recombination products, and that V(D)J recombination at the *SynIgh* locus display conserved recombination pattern.

(F) PFGE size validation (left) and NGS coverage validation of the generational stability of *SynIgh* (right) in *E. coli*.
(G) Schematic illustration of the stabilizer-based assembly strategy in *S. cerevisiae*.
(H) Assembly fidelity of >100 kb constructs generated by *S. cerevisiae* (WT, n = 10 biological replicates; stabilizer, n = 4 biological replicates). Data are presented as the means ± SEM.
(I) PFGE size validation (left) and NGS validation coverage of *SynIgh* generational stability (right) in *S. cerevisiae.*
See also Figure S2.

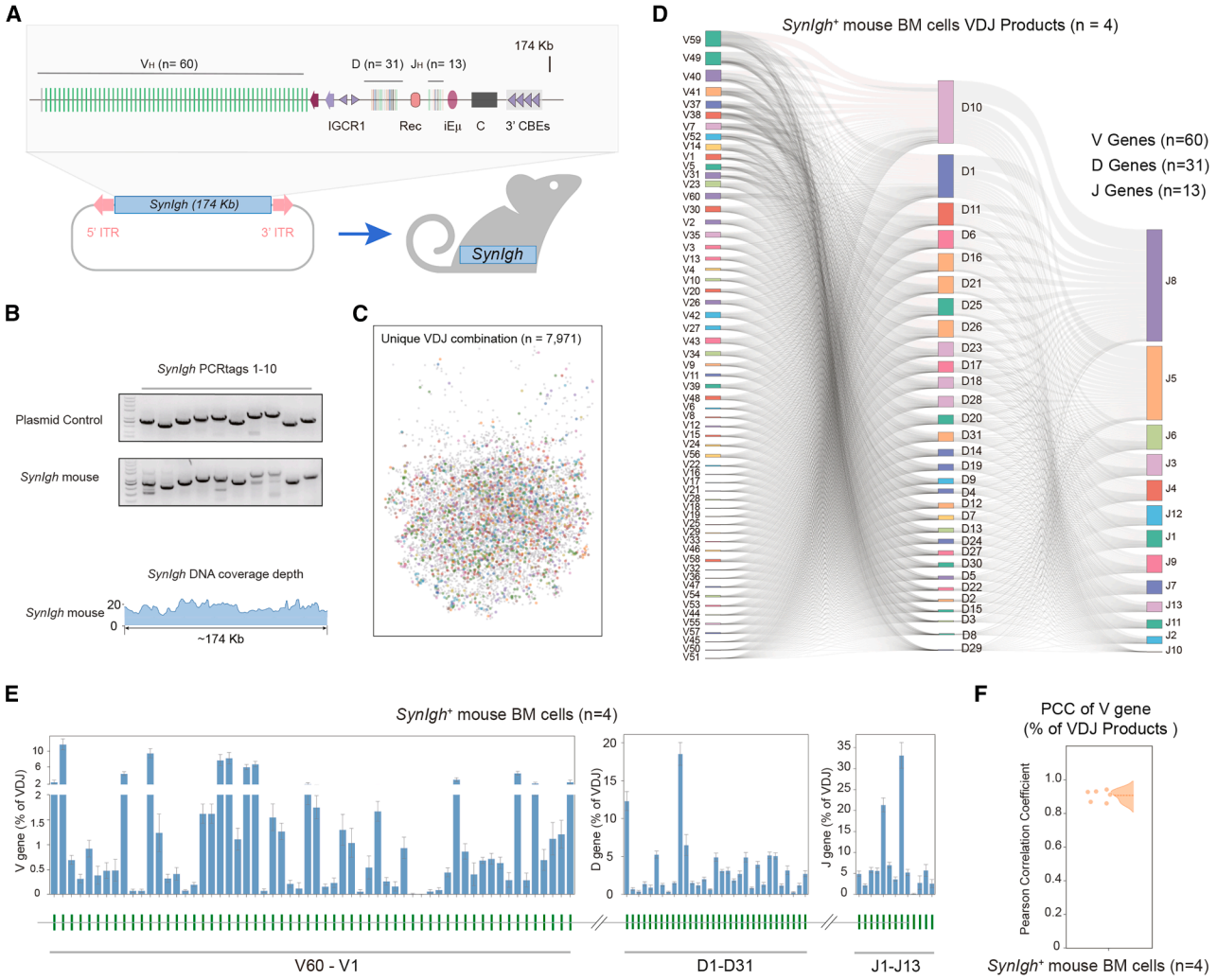


Figure 3. *SynIgh* mice undergo V(D)J recombination

(A) Schematic of the *SynIgh* mouse model.

(B) Validation of *SynIgh* integrity in the mouse genome. PCR validation of 10 junctions in *SynIgh* (top), nanopore sequencing coverage of *SynIgh* (bottom).

(C and D) VDJ-seq analysis of *SynIgh*$^{+}$ mice bone marrow cells. VDJ recombination products (C) and V, D, and J distributions of VDJ recombination products (D).

(E) VDJ-seq analysis of *SynIgh* bone marrow cells. Gene usage is presented as a percentage of all VDJ recombination products.

(F) Correlation of V gene usage in the VDJ recombination product between mouse (n = 4, PCC = 0.9 ± 0.04). The data are presented as the means ± SEM. n = 4 *SynIgh* mice.

See also Figure S3.

Synthetic rearrangement shuffling the *SynIgh* in somatic cells

The mSCRaMbLE system of *SynIgh* featuring 61 *loxPsym* sites and 2 *loxP* sites was designed to induce synthetic rearrangements within the *SynIgh* locus in the presence of Cre recombinase. To independently assess its function, *SynIgh* mice were bred with *Rag2*$^{-/-}$ mice to generate *SynIgh*$^{+}$*Rag2*$^{-/-}$ mice, in which deficiency of Rag2 prevents V(D)J recombination in pro-B cells[51] (Figure S4A). Previous work on yeast SCRaMbLE indicates that strict spatiotemporal control is essential to minimize deletion bias and preserve structural diversity.[52] To make mSCRaMbLE tightly controllable and restricted before the V-DJ stage, we placed an abscisic acid (ABA)-inducible split-Cre system (CreN-ABI and PYL-CreC)[53] under the B-cell-specific Cd79a promoter (P_{Cd79a}). The Cre cassette was inserted between the D and J segments so that productive D-J joining excises the cassette, thereby self-terminating Cre activity before V-DJ recombination. To activate mSCRaMbLE, *SynIgh*$^{+}$*Rag2*$^{-/-}$ mice were intraperitoneally injected with ABA for five consecutive days to induce split-Cre dimerization in B cells (Figures 4A and 4B).

Two weeks later, pro-B cells (B220^{+}) were sorted from the bone marrow, and genomic DNA was extracted for targeted sequencing of the *SynIgh* region. We first examined the mSCRaMbLE-mediated inversion of the *SynIgh-D* region. Using linear amplification-mediated high-throughput genome-wide

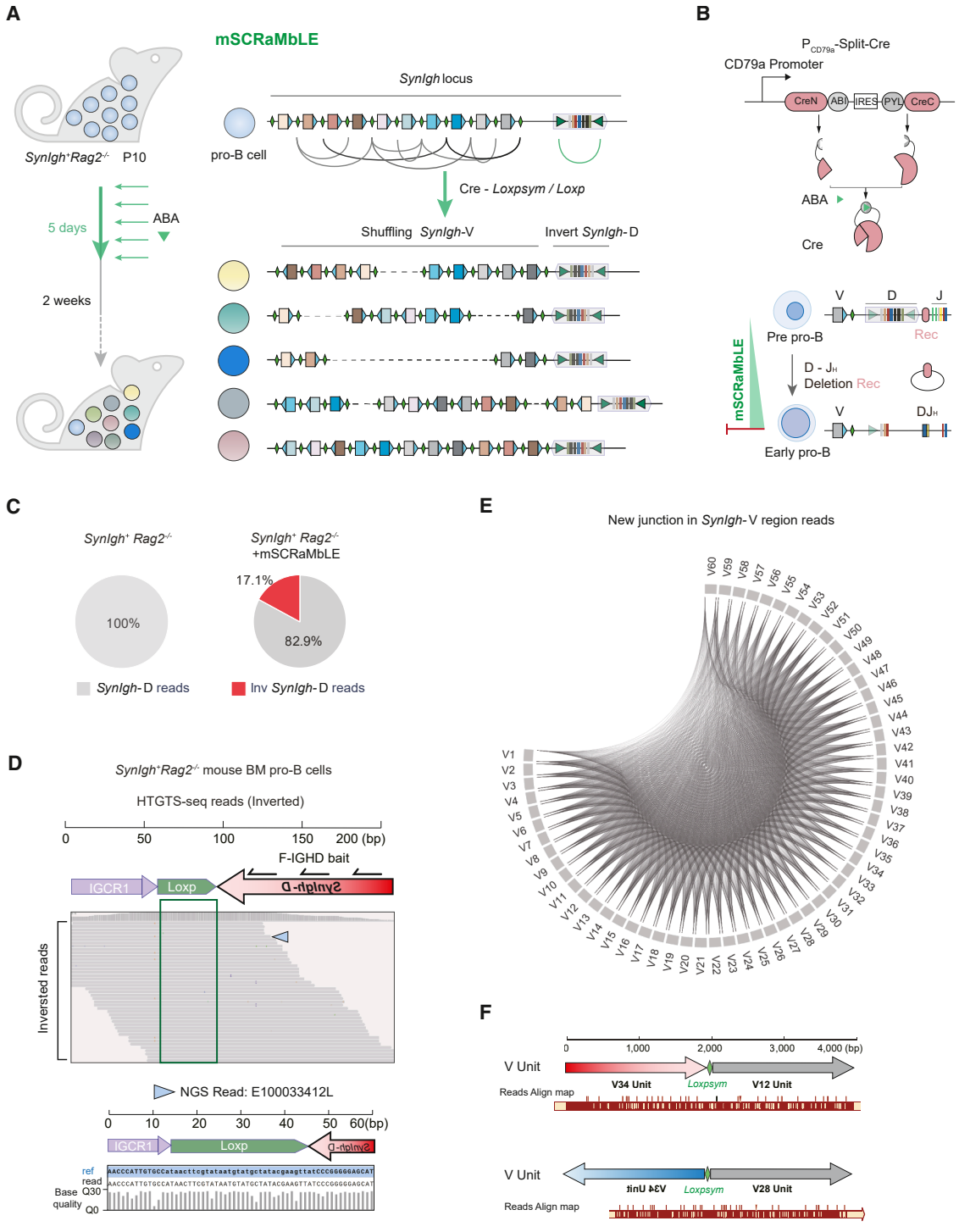


Figure 4. Synthetic rearrangement shuffling the *Synlgh* in somatic cells

(A) Schematic of the mSCRaMbLE system. Induction of synthetic rearrangement in *Synlgh*$^{+}$*Rag2*$^{-/-}$ mice via intraperitoneal injection of ABA (left). Structure of the *Synlgh* locus in bone marrow pro-B cells before and after mSCRaMbLE (right).

(B) P_{CD79a}-Cre (split) expression cassette (top). The cassette is positioned between the D and J gene clusters and is excised following D-J recombination (down).

(C) Proportion of *IGHD* reads in the bone marrow pro-B cells of *Synlgh*$^{+}$*Rag2*$^{-/-}$ mice.

(D) LAM-HTGTS-seq detection of IGCR1 region and IGHD region junctions in the bone marrow pro-B cells of *Synlgh*$^{+}$*Rag2*$^{-/-}$mice. Reads detected using F-IGHD end bait primers (black) were aligned with the IGCR1-(inv-IGHD) locus (top). Comparison of the nucleotide sequence and base quality in the region flanking the *loxP* site of E100034412L (blue triangle) with the reference sequence after inversion (bottom).

(legend continued on next page)

translocation sequencing (LAM-HTGTS) with V_H gene-specific primers serving as bait,[54] we detected new DNA junctions resulting from the inversion (Figures 4C and 4D). Since *SynIgh*$^{+}$*Rag2*$^{-/-}$ mice have only a single allele of *SynIgh*, the efficiency of *SynIgh-D* inversion in pro-B cells can be determined by calculating the ratio of the number of reads containing inverted junctions to the total number of D gene reads. In ABA-injected mice, 17.1% of the reads presented inversion of *SynIgh-D*, whereas no inversions were detected in the untreated control (Figure 4C).

Next, we assessed mSCRaMbLE-mediated shuffling of the *SynIgh*-V gene region. Due to the high similarity between the units, we adapted LAM-HTGTS for long-read sequencing to achieve sufficient coverage to identify the new junctions of V gene units. As in our previous work analyzing synthetic genome rearrangement in yeast,[29] reads containing *loxPsym* sites were extracted from the sequencing data, and then the flanking sequences of *loxPsym* were identified on the basis of alignment and the unique sequences of each V gene unit. By analysis of genomic DNA from pro-B-cell in *SynIgh*$^{+}$*Rag2*$^{-/-}$+mSCRaMbLE mice, we detected 446,758 sequencing reads representing new DNA junctions of V-unit combinations mediated by *loxPsym* sites that differ from the *SynIgh* reference sequence (Figures 4E, 4F, and S4B–S4D). These results indicated that mSCRaMbLE triggered synthetic rearrangements in the IGHD region and between V units, shuffling the *SynIgh* locus and generating a diverse library of structural variants prior to V(D)J recombination.

Synthetic rearrangement transforms the V(D)J recombination pattern

We then tested the influence of mSCRaMbLE-induced shuffling of the *SynIgh* on V(D)J recombination patterns. After administering ABA to *SynIgh*$^{+}$ mice for five consecutive days to induce mSCRaMbLE, we isolated bone marrow cells from these mice (*SynIgh*$^{+}$+mSCRaMbLE) and extracted their DNA for VDJ-seq analysis (Figures 5A and 5B). We detected 9,353 unique VDJ recombination products utilizing all 60 V_H genes, 31 D genes, and 13 J_H genes from the *SynIgh* locus. To assess the impact of the inversion of the *SynIgh*-D region, we analyzed D gene utilization within VDJ recombination products. We found that, in addition to the expected forward-oriented D genes, all inverted D genes (Inv-D) were also present in the products, contributing to 41% (3,826 out of 9,353) of the recombination products (Figure 5C). Next, to evaluate the effect of the shuffling in the V region, we analyzed the V gene utilization in VDJ recombination products across different *SynIgh*$^{+}$+mSCRaMbLE mice. Earlier, we observed that the V gene utilization of *SynIgh*$^{+}$ mice was highly similar across individuals (PCC = 0.9 ± 0.04). Therefore, using this as a control, we evaluated V gene utilization in *SynIgh*$^{+}$+mSCRaMbLE mice. We observed that in VDJ recombination products some V genes that were highly utilized in the *SynIgh*$^{+}$ controls showed decreased usage, whereas less utilized V genes exhibited increased usage, in the *SynIgh*$^{+}$+mSCRaMbLE mice. Specifically, genes such as V12 and V52, which were utilized at only 0.5% in *SynIgh*$^{+}$, presented more than a 10-fold increase in utilization in *SynIgh*$^{+}$+mSCRaMbLE (mouse #1, #2), whereas frequently used genes such as V7 and V40 presented more than a 10-fold decrease. Furthermore, we performed a correlation analysis of V gene utilization rates across different individuals. Among the *SynIgh*$^{+}$+mSCRaMbLE and *SynIgh*$^{+}$ mice, the V gene utilization rates were weakly correlated (PCC = 0.24 ± 0.2, n = 7). Among *SynIgh*$^{+}$+mSCRaMbLE mice, gene utilization correlation values ranged widely, from 0.09 to 0.74 (n = 7) (Figure 5D). Moreover, to visualize the V(D)J recombination patterns, we performed PCA (principal component analysis) on the types and frequencies of the VDJ recombination products. The 4 *SynIgh*$^{+}$ mice clustered closely in PCA space, indicating their similar VDJ product compositions. In contrast, the VDJ recombination products of the 7 *SynIgh*$^{+}$+mSCRaMbLE mice were widely dispersed in PCA space, and their positions varied from those of the *SynIgh*$^{+}$ cluster and among themselves (Figure 5E). These results indicate that V gene usage in VDJ recombination products from *SynIgh*$^{+}$+mSCRaMbLE mice differs not only from that in *SynIgh*$^{+}$ mice but also among the individual mice. Overall, synthetic rearrangement extensively shuffled the *SynIgh* locus, leading to the transformation of the V(D)J recombination pattern and the enrichment of V(D)J recombination diversity.

DISCUSSION

While somatic DNA recombination is generally constrained to preserve genomic integrity, V(D)J recombination stands as an evolutionarily conserved exception in vertebrate lymphocytes. Here, by introducing the mSCRaMbLE system into the *SynIgh* locus, we developed a synthetic and programmable version of somatic DNA recombination. Compared to recent efforts that randomized the human genome via engineered DNA recombination in cultured cells,[55,56] our study establishes the instance of synthetic somatic DNA recombination in living somatic cells. Admittedly, *loxPsym*-mediated recombination inevitably favors deletions due to continuous rearrangement, potentially constraining the structural diversity compared to theoretical schematics; however, future integration of eukaryotic multiplexed site-specific inversion systems, such as RCI(Recombinase Rci)51-5/multi-sfxal01,[57,58] will further mitigate this bias by favoring invertible rearrangements. Coupling synthetic mSCRaMbLE with natural V(D)J recombination releases compositional and positional constraints within the *IGH* locus, allowing inversion of the D region and rearrangement of V genes prior to recombination. This facilitates the inclusion of rare inverted D genes and repositions V genes, effectively transforming the one-to-many V(D)J recombination process into a many-to-many framework that enhances recombination diversity. In addition to enhancing V(D)J

(E) Circos plots showing new junctions between 60 V_H units in the *SynIgh* locus V region in the bone marrow pro-B cells of *SynIgh*$^{+}$*Rag2*$^{-/-}$ mice. V units (gray) are arranged clockwise, matching the orientation of the *SynIgh* locus. The connections between units represent new junctions detected by sequencing (gray line). (F) ONT reads aligned with the V34-V12 locus (top) and (inverted-V34)-V28 (bottom), with nucleotide comparison and base quality at the flanking *loxPsym* sites. Purified pro-B cells from 4 mice were pooled together for mSCRaMbLE-seq experiment (C–E). See also Figure S4.

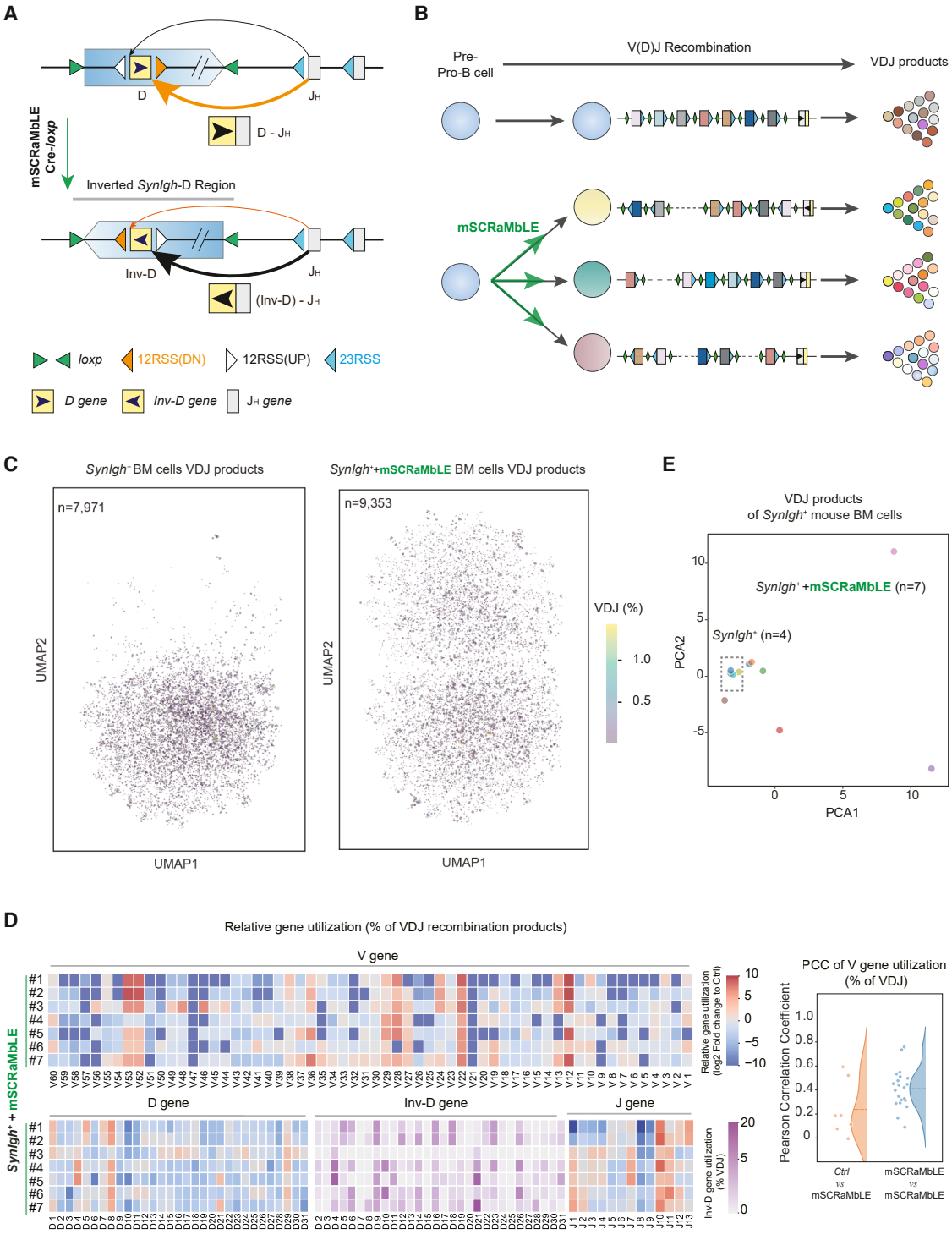


Figure 5. Synthetic rearrangement transforms the V(D)J recombination pattern

(A) Schematic illustration of the DJ recombination product generated by recombination between D-RSS-DN and J_H-RSS in *Synlgh*. Synthetic rearrangement induces *Cre-loxP*-mediated inversion in IGHD, followed by recombination between D-RSS-UP and J_H-RSS to generate (Inv-D)J recombination products.

(B) Schematic illustration of synthetic rearrangement to shuffle the *Synlgh* structure prior to V(D)J recombination, affecting the subsequent V(D)J recombination pattern.

(C) VDJ-seq analysis of the VDJ recombination products in bone marrow cells from *Synlgh*$^+$ (left) and *Synlgh*$^+$+mSCRaMbLE (right) mice.

(legend continued on next page)

recombination diversity, mSCRaMbLE in *SynIgh* provides a systematic approach to dissect how genetic variation shapes this process. Specifically, compared to genomic diversification driven by natural evolution, this synthetic approach provides a more efficient route for the generation of diverse genomic architectures.[14] Unlike conventional methods that rely on multiple genetically modified models, a single induction event generates a vast library of structural variants. Coupled with tools such as *in vitro* differentiation and VDJ-seq,[4,5] this system facilitates functional dissection of modular elements and higher-order regulatory interactions within the *IGH* locus, offering new avenues for understanding the regulatory architecture of the *IGH* locus and how genetic variants drive immune diversity.

Mammalian synthetic genomics works, such as synthetic Hox,[59] Sox2,[60] α-globin,[61] and ACE2[62] loci, have demonstrated their advanced capability to elucidate the principles of regulatory mechanisms and to construct disease models. However, while systematic design and rewriting of genetic loci to expand natural functions has been achieved in microorganisms,[27,63,64] such efforts have not yet been realized in mammals. Our synthetic reconfiguration and shuffling leverage *de novo* modular design and multiplex genomic locus manipulation, establishing a systematic genome rewriting paradigm characterized by modularity, extensibility, and flexibility. In summary, the design, synthesis, and shuffling of *SynIgh* enhance V(D)J recombination, reconfiguring the process of antibody diversity generation and offering a scalable framework for the rewriting of larger and more intricate mammalian loci.

Limitations of the study

SynIgh was designed as a simplified synthetic version of the human IGH locus. Although this design supports V(D)J recombination in mice, additional intergenic sequences, regulatory elements, and structural variation not included in *SynIgh* may influence recombination efficiency and subsequent B cell development. In addition, the mSCRaMbLE-seq analysis was performed on pooled pro-B cells using targeted long-read sequencing. This approach captures rearrangement patterns in the population, but it does not directly resolve shuffled *SynIgh* structures in individual cells or fully link a given structure to its downstream V(D)J recombination product. Finally, the findings of this study are based mainly on sequence and recombination outputs, and the downstream biological consequences of *SynIgh* rewriting, such as antibody repertoire diversity and B cell functional maturation, were not examined here.

RESOURCE AVAILABILITY

Lead contact

Further information and requests for resources and reagents should be directed to and will be fulfilled by the lead contact, Yi Wu (yi.wu@tju.edu.cn).

Materials availability

All reagents generated in this study are available from the lead contact with a completed Materials Transfer Agreement.

Data and code availability

- The raw sequencing data reported in this paper have been deposited in the Genome Sequence Archive of the National Genomics Data Center, China National Center for Bioinformation/Beijing Institute of Genomics, Chinese Academy of Sciences (GSA: CRA019373, CRA019453, CRA063111) and are publicly accessible at https://ngdc.cncb.ac.cn/gsa.
- The custom analysis code generated for this study has been deposited in Zenodo and is publicly available at https://doi.org/10.5281/zenodo.18576074.
- Any additional information required to reanalyze the data reported in this paper is available from the lead contact upon request.

ACKNOWLEDGMENTS

We thank the lab members for stimulating discussions. This work was supported by the National Key R&D Program of China, Synthetic Biology Research (2019YFA0903800), the National Natural Science Foundation of China (32551084, 32471483), Scientific Research Innovation Capability Support Project for Young Faculty, the National Natural Science Foundation of China, Young Scientists Fund (C Class) (32500783), and the National Natural Science Foundation of Tianjin (24JCQNJC01250). Z.-H.F. is supported by the China Postdoctoral Science Foundation (2024M762365).

AUTHOR CONTRIBUTIONS

Z.-H.F., Y.W., and G.-R.Z. conceptualized the study. Z.-H.F., Y.W., and G.-R.Z. designed the experiments. Z.-H.F. designed the locus. Z.B. aided with the locus design and experimental design. Z.-H.F. and Y.M. assembled the constructs. Z.-H.F. established and validated the mouse model. Z.-H.F. performed WGS, VDJ-seq, and other sequencing assays, as well as data analysis. S.C., Y.Z., and N.W. assisted with the mouse experiments. Z.-H.F. wrote the manuscript, with Y.W. and G.-R.Z. reviewing and editing the manuscript.

DECLARATION OF INTERESTS

The authors declare no competing interests.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- EXPERIMENTAL MODEL AND STUDY PARTICIPANT DETAILS
  - Mice
  - Yeast strains
  - Bacterial strains
- METHOD DETAILS
  - Design and synthesis of *SynIgh*
  - Pulsed-field gel electrophoresis
  - DNA-seq
  - Generation of the *SynIgh* mouse model
  - Induction of mSCRaMbLE
  - Bone marrow cell purification
  - VDJ-seq

(D) VDJ-seq analysis of *SynIgh*$^+$+mSCRaMbLE bone marrow cells. Gene usage is presented as a percentage of all VDJ recombination products (left). Correlation of V gene usage in the VDJ recombination products between *SynIgh*$^+$ and *SynIgh*$^+$+mSCRaMbLE (*SynIgh*$^+$+mSCRaMbLE, $n = 7$ vs. *SynIgh*: PCC = 0.24 ± 0.2; *SynIgh*$^+$+mSCRaMbLE vs. *SynIgh*$^+$+mSCRaMbLE: $n = 7$, PCC = 0.09 to 0.74) (right).
(E) PCA of the VDJ recombination products from the bone marrow of *SynIgh*$^+$ ($n = 4$) and *SynIgh*$^+$+mSCRaMbLE ($n = 7$) mice, comparing individual VDJ product composition by type and frequency.

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at https://doi.org/10.1016/j.xgen.2026.101252.

Received: November 19, 2025
Revised: February 28, 2026
Accepted: April 23, 2026
Published: May 18, 2026

REFERENCES

1. Tonegawa, S. (1983). Somatic generation of antibody diversity. Nature *302*, 575–581. https://doi.org/10.1038/302575a0.
2. Zhang, Y., Zhang, X., Dai, H.-Q., Hu, H., and Alt, F.W. (2022). The role of chromatin loop extrusion in antibody diversification. Nat. Rev. Immunol. *22*, 550–566. https://doi.org/10.1038/s41577-022-00679-3.
3. Zhang, Y., Zhang, X., Ba, Z., Liang, Z., Dring, E.W., Hu, H., Lou, J., Kyritsis, N., Zurita, J., Shamim, M.S., et al. (2019). The fundamental role of chromatin loop extrusion in physiological V(D)J recombination. Nature *573*, 600–604. https://doi.org/10.1038/s41586-019-1547-y.
4. Ba, Z., Lou, J., Ye, A.Y., Dai, H.-Q., Dring, E.W., Lin, S.G., Jain, S., Kyritsis, N., Kieffer-Kwon, K.-R., Casellas, R., and Alt, F.W. (2020). CTCF orchestrates long-range cohesin-driven V(D)J recombinational scanning. Nature *586*, 305–310. https://doi.org/10.1038/s41586-020-2578-0.
5. Dai, H.-Q., Hu, H., Lou, J., Ye, A.Y., Ba, Z., Zhang, X., Zhang, Y., Zhao, L., Yoon, H.S., Chapdelaine-Williams, A.M., et al. (2021). Loop extrusion mediates physiological Igh locus contraction for RAG scanning. Nature *590*, 338–343. https://doi.org/10.1038/s41586-020-03121-7.
6. Jain, S., Ba, Z., Zhang, Y., Dai, H.-Q., and Alt, F.W. (2018). CTCF-Binding Elements Mediate Accessibility of RAG Substrates During Chromatin Scanning. Cell *174*, 102–116.e14. https://doi.org/10.1016/j.cell.2018.04.035.
7. Schatz, D.G., Zhang, Y., Xiao, J., Zha, S., Zhang, Y., and Alt, F.W. (2024). Chapter 2 - The Mechanism, Regulation and Evolution of V(D)J Recombination. In Molecular Biology of B Cells, Third Edition, T. Honjo, M. Reth, A. Radbruch, F. Alt, and A. Martin, eds. (Academic Press), pp. 13–57. https://doi.org/10.1016/B978-0-323-95895-0.00004-0.
8. Lees, W.D., Christley, S., Peres, A., Kos, J.T., Corrie, B., Ralph, D., Breden, F., Cowell, L.G., Yaari, G., Corcoran, M., et al. (2023). AIRR community curation and standardised representation for immunoglobulin and T cell receptor germline sets. ImmunoInformatics *10*, 100025. https://doi.org/10.1016/j.immuno.2023.100025.
9. Auton, A., Abecasis, G.R., Altshuler, D.M., Durbin, R.M., Abecasis, G.R., Bentley, D.R., Chakravarti, A., Clark, A.G., Donnelly, P., Eichler, E.E., et al. (2015). A global reference for human genetic variation. Nature *526*, 68–74. https://doi.org/10.1038/nature15393.
10. Engelbrecht, E., Rodriguez, O.L., Lees, W., Vanwinkle, Z., Shields, K., Schultze, S., Gibson, W.S., Smith, D.R., Jana, U., Saha, S., et al. (2025). Germline polymorphisms in the immunoglobulin kappa and lambda loci underpinning antibody light chain repertoire variability. Nat. Commun. *16*, 11707. https://doi.org/10.1038/s41467-025-66759-9.
11. Jana, U., Rodriguez, O.L., Lees, W., Engelbrecht, E., Vanwinkle, Z., Peres, A., Gibson, W.S., Shields, K., Schultze, S., Dorgham, A., et al. (2026). The human IG heavy chain constant gene locus is enriched for large structural variants and coding polymorphisms that vary among human populations. Cell Genom. *6*, 101058. https://doi.org/10.1016/j.xgen.2025.101058.
12. Corcoran, M., Narang, S., Kaduk, M., Chernyshev, M., Färnert, A., Sundling, C., and Hedestam, G.B.K. (2026). Ultra-high-throughput IGH genotyping of 25 global populations reveals population-biased allelic diversity and homozygous V and D gene deletions. Immunity *59*, 1107–1122.e5. https://doi.org/10.1016/j.immuni.2026.01.026.
13. Khatri, I., Berkowska, M.A., van den Akker, E.B., Teodosio, C., Reinders, M.J.T., and van Dongen, J.J.M. (2021). Population matched (pm) germline allelic variants of immunoglobulin (IG) loci: Relevance in infectious diseases and vaccination studies in human populations. Genes Immun. *22*, 172–186. https://doi.org/10.1038/s41435-021-00143-7.
14. Rodriguez, O.L., Safonova, Y., Silver, C.A., Shields, K., Gibson, W.S., Kos, J.T., Tieri, D., Ke, H., Jackson, K.J.L., Boyd, S.D., et al. (2023). Genetic variation in the immunoglobulin heavy chain locus shapes the human antibody repertoire. Nat. Commun. *14*, 4419. https://doi.org/10.1038/s41467-023-40070-x.
15. Scheepers, C., Shrestha, R.K., Lambson, B.E., Jackson, K.J.L., Wright, I.A., Naicker, D., Goosen, M., Berrie, L., Ismail, A., Garrett, N., et al. (2015). Ability to develop broadly neutralizing HIV-1 antibodies is not restricted by the germline immunoglobulin gene repertoire. J. Immunol. *194*, 4371–4378. https://doi.org/10.4049/jimmunol.1500118.
16. Pushparaj, P., Nicoletto, A., Sheward, D.J., Das, H., Castro Dopico, X., Perez Vidakovics, L., Hanke, L., Chernyshev, M., Narang, S., Kim, S., et al. (2023). Immunoglobulin germline gene polymorphisms influence the function of SARS-CoV-2 neutralizing antibodies. Immunity *56*, 193–206.e7. https://doi.org/10.1016/j.immuni.2022.12.005.
17. deCamp, A.C., Corcoran, M.M., Fulp, W.J., Willis, J.R., Cottrell, C.A., Bader, D.L.V., Kalyuzhniy, O., Leggat, D.J., Cohen, K.W., Hyrien, O., et al. (2024). Human immunoglobulin gene allelic variation impacts germline-targeting vaccine priming. npj Vaccines *9*, 58. https://doi.org/10.1038/s41541-024-00811-5.
18. Zaslavsky, M.E., Craig, E., Michuda, J.K., Sehgal, N., Ram-Mohan, N., Lee, J.-Y., Nguyen, K.D., Hoh, R.A., Pham, T.D., Röltgen, K., et al. (2025). Disease diagnostics using machine learning of B cell and T cell receptor sequences. Science *387*, eadp2407. https://doi.org/10.1126/science.adp2407.
19. Perlot, T., Alt, F.W., Bassing, C.H., Suh, H., and Pinaud, E. (2005). Elucidation of IgH intronic enhancer functions via germ-line deletion. Proc. Natl. Acad. Sci. USA *102*, 14362–14367. https://doi.org/10.1073/pnas.0507090102.
20. Liang, Z., Zhao, L., Ye, A.Y., Lin, S.G., Zhang, Y., Guo, C., Dai, H.-Q., Ba, Z., and Alt, F.W. (2023). Contribution of the IGCR1 regulatory element and the 3′Igh CTCF-binding elements to regulation of Igh V(D)J recombination. Proc. Natl. Acad. Sci. USA *120*, e2306564120. https://doi.org/10.1073/pnas.2306564120.
21. Hill, L., Ebert, A., Jaritz, M., Wutz, G., Nagasaka, K., Tagoh, H., Kostanova-Poliakova, D., Schindler, K., Sun, Q., Bönelt, P., et al. (2020). Wapl repression by Pax5 promotes V gene recombination by Igh loop extrusion. Nature *584*, 142–147. https://doi.org/10.1038/s41586-020-2454-y.
22. Lin, S.G., Ba, Z., Du, Z., Zhang, Y., Hu, J., and Alt, F.W. (2016). Highly sensitive and unbiased approach for elucidating antibody repertoires. Proc. Natl. Acad. Sci. USA *113*, 7846–7851. https://doi.org/10.1073/pnas.1608649113.
23. Macdonald, L.E., Meagher, K.A., Franklin, M.C., Levenkova, N., Hansen, J., Badithe, A.T., Zhong, M., Krueger, P., Rafique, A., Tu, N., et al. (2020). Kappa-on-Heavy (KoH) bodies are a distinct class of fully-human antibody-like therapeutic agents with antigen-binding properties. Proc. Natl. Acad. Sci. USA *117*, 292–299. https://doi.org/10.1073/pnas.1901734117.
24. Byrska-Bishop, M., Evani, U.S., Zhao, X., Basile, A.O., Abel, H.J., Regier, A.A., Corvelo, A., Clarke, W.E., Musunuri, R., Nagulapalli, K., et al. (2022). High-coverage whole-genome sequencing of the expanded 1000 Genomes Project cohort including 602 trios. Cell *185*, 3426–3440.e19. https://doi.org/10.1016/j.cell.2022.08.004.
25. Manso, T., Folch, G., Giudicelli, V., Jabado-Michaloud, J., Kushwaha, A., Nguefack Ngoune, V., Georga, M., Papadaki, A., Debbagh, C., Pégorier, P., et al. (2022). IMGT® databases, related tools and web resources

through three main axes of research and development. Nucleic Acids Res. *50*, D1262–D1272. https://doi.org/10.1093/nar/gkab1136.

26. Dymond, J.S., Richardson, S.M., Coombes, C.E., Babatz, T., Muller, H., Annaluru, N., Blake, W.J., Schwerzmann, J.W., Dai, J., Lindstrom, D.L., et al. (2011). Synthetic chromosome arms function in yeast and generate phenotypic diversity by design. Nature *477*, 471–476. https://doi.org/10.1038/nature10403.

27. Richardson, S.M., Mitchell, L.A., Stracquadanio, G., Yang, K., Dymond, J.S., DiCarlo, J.E., Lee, D., Huang, C.L.V., Chandrasegaran, S., Cai, Y., et al. (2017). Design of a synthetic yeast genome. Science *355*, 1040–1044. https://doi.org/10.1126/science.aaf4557.

28. Wu, Y., Li, B.-Z., Zhao, M., Mitchell, L.A., Xie, Z.-X., Lin, Q.-H., Wang, X., Xiao, W.-H., Wang, Y., Zhou, X., et al. (2017). Bug mapping and fitness testing of chemically synthesized chromosome X. Science *355*, eaaf4706. https://doi.org/10.1126/science.aaf4706.

29. Zhou, S., Wu, Y., Zhao, Y., Zhang, Z., Jiang, L., Liu, L., Zhang, Y., Tang, J., and Yuan, Y.-J. (2023). Dynamics of synthetic yeast chromosome evolution shaped by hierarchical chromatin organization. Natl. Sci. Rev. *10*, nwad073. https://doi.org/10.1093/nsr/nwad073.

30. Zhou, S., Wu, Y., Xie, Z.-X., Jia, B., and Yuan, Y.-J. (2021). Directed genome evolution driven by structural rearrangement techniques. Chem. Soc. Rev. *50*, 12788–12807. https://doi.org/10.1039/D1CS00722J.

31. Prabakaran, P., Gupta, A., Rao, S.P., Rajpal, D., Wendt, M., Qiu, Y., and Chowdhury, P.S. (2025). Unveiling inverted D genes and D-D fusions in human antibody repertoires unlocks novel antibody diversity. Commun. Biol. *8*, 133. https://doi.org/10.1038/s42003-024-07441-6.

32. Luo, Z., Wang, L., Wang, Y., Zhang, W., Guo, Y., Shen, Y., Jiang, L., Wu, Q., Zhang, C., Cai, Y., and Dai, J. (2018). Identifying and characterizing SCRaMbLEd synthetic yeast using ReSCuES. Nat. Commun. *9*, 1930. https://doi.org/10.1038/s41467-017-00806-y.

33. Gibson, D.G., Young, L., Chuang, R.-Y., Venter, J.C., Hutchison, C.A., and Smith, H.O. (2009). Enzymatic assembly of DNA molecules up to several hundred kilobases. Nat. Methods *6*, 343–345. https://doi.org/10.1038/nmeth.1318.

34. Gietz, R.D., and Schiestl, R.H. (2007). High-efficiency yeast transformation using the LiAc/SS carrier DNA/PEG method. Nat. Protoc. *2*, 31–34. https://doi.org/10.1038/nprot.2007.13.

35. Ostrov, N., Beal, J., Ellis, T., Gordon, D.B., Karas, B.J., Lee, H.H., Lenaghan, S.C., Schloss, J.A., Stracquadanio, G., Trefzer, A., et al. (2019). Technological challenges and milestones for writing genomes. Science *366*, 310–312. https://doi.org/10.1126/science.aay0339.

36. Hossain, A., Lopez, E., Halper, S.M., Cetnar, D.P., Reis, A.C., Strickland, D., Klavins, E., and Salis, H.M. (2020). Automated design of thousands of nonrepetitive parts for engineering stable genetic systems. Nat. Biotechnol. *38*, 1466–1475. https://doi.org/10.1038/s41587-020-0584-2.

37. He, B., Ma, Y., Tian, F., Zhao, G.-R., Wu, Y., and Yuan, Y.-J. (2023). YLC-assembly: large DNA assembly via yeast life cycle. Nucleic Acids Res. *51*, 8283–8292. https://doi.org/10.1093/nar/gkad599.

38. Pal, S., Postnikoff, S.D., Chavez, M., and Tyler, J.K. (2018). Impaired cohesion and homologous recombination during replicative aging in budding yeast. Sci. Adv. *4*, eaaq0236. https://doi.org/10.1126/sciadv.aaq0236.

39. Haws, S.A., Simandi, Z., Barnett, R.J., and Phillips-Cremins, J.E. (2022). 3D genome, on repeat: Higher-order folding principles of the heterochromatinized repetitive genome. Cell *185*, 2690–2707. https://doi.org/10.1016/j.cell.2022.06.052.

40. Mitrentsi, I., Lou, J., Kerjouan, A., Verigos, J., Reina-San-Martin, B., Hinde, E., and Soutoglou, E. (2022). Heterochromatic repeat clustering imposes a physical barrier on homologous recombination to prevent chromosomal translocations. Mol. Cell *82*, 2132–2147.e6. https://doi.org/10.1016/j.molcel.2022.03.033.

41. Larson, A.G., Elnatan, D., Keenen, M.M., Trnka, M.J., Johnston, J.B., Burlingame, A.L., Agard, D.A., Redding, S., and Narlikar, G.J. (2017). Liquid droplet formation by HP1α suggests a role for phase separation in heterochromatin. Nature *547*, 236–240. https://doi.org/10.1038/nature22822.

42. Keenen, M.M., Brown, D., Brennan, L.D., Renger, R., Khoo, H., Carlson, C.R., Huang, B., Grill, S.W., Narlikar, G.J., and Redding, S. (2021). HP1 proteins compact DNA into mechanically and positionally stable phase separated domains. eLife *10*, e64563. https://doi.org/10.7554/eLife.64563.

43. Lu, X., Shaw, W.M., Sutradhar, A., Stracquadanio, G., and Ellis, T. (2024). Synthetic genome modules designed for programmable silencing of functions and chromosomes. Preprint at bioRxiv. https://doi.org/10.1101/2024.03.22.586311.

44. Shrinivas, K., Sabari, B.R., Coffey, E.L., Klein, I.A., Boija, A., Zamudio, A.V., Schuijers, J., Hannett, N.M., Sharp, P.A., Young, R.A., and Chakraborty, A.K. (2019). Enhancer Features that Drive Formation of Transcriptional Condensates. Mol. Cell *75*, 549–561.e7. https://doi.org/10.1016/j.molcel.2019.07.009.

45. Ong, C.-T., and Corces, V.G. (2014). CTCF: an architectural protein bridging genome topology and function. Nat. Rev. Genet. *15*, 234–246. https://doi.org/10.1038/nrg3663.

46. Wei, C., Jia, L., Huang, X., Tan, J., Wang, M., Niu, J., Hou, Y., Sun, J., Zeng, P., Wang, J., et al. (2022). CTCF organizes inter-A compartment interactions through RYBP-dependent phase separation. Cell Res. *32*, 744–760. https://doi.org/10.1038/s41422-022-00676-0.

47. Gao, Y., Han, M., Shang, S., Wang, H., and Qi, L.S. (2021). Interrogation of the dynamic properties of higher-order heterochromatin using CRISPR-dCas9. Mol. Cell *81*, 4287–4299.e5. https://doi.org/10.1016/j.molcel.2021.07.034.

48. Taxis, C., and Knop, M. (2006). System of Centromeric, Episomal, and Integrative Vectors Based on Drug Resistance Markers for SacCharomyces Cerevisiae. Biotechniques *40*, 73–78. https://doi.org/10.2144/000112040.

49. Patel, J.R., Oh, J., Wang, S., Crawford, J.M., and Isaacs, F.J. (2022). Cross-kingdom expression of synthetic genetic elements promotes discovery of metabolites in the human microbiome. Cell *185*, 1487–1505.e14. https://doi.org/10.1016/j.cell.2022.03.008.

50. Chovanec, P., Bolland, D.J., Matheson, L.S., Wood, A.L., Krueger, F., Andrews, S., and Corcoran, A.E. (2018). Unbiased quantification of immunoglobulin diversity at the DNA level with VDJ-seq. Nat. Protoc. *13*, 1232–1252. https://doi.org/10.1038/nprot.2018.021.

51. Alt, F.W., Zhang, Y., Meng, F.-L., Guo, C., and Schwer, B. (2013). Mechanisms of Programmed DNA Lesions and Genomic Instability in the Immune System. Cell *152*, 417–429. https://doi.org/10.1016/j.cell.2013.01.007.

52. Jia, B., Wu, Y., Li, B.-Z., Mitchell, L.A., Liu, H., Pan, S., Wang, J., Zhang, H.-R., Jia, N., Li, B., et al. (2018). Precise control of SCRaMbLE in synthetic haploid and diploid yeast. Nat. Commun. *9*, 1933. https://doi.org/10.1038/s41467-018-03084-4.

53. Weinberg, B.H., Cho, J.H., Agarwal, Y., Pham, N.T.H., Caraballo, L.D., Walkosz, M., Ortega, C., Trexler, M., Tague, N., Law, B., et al. (2019). High-performance chemical- and light-inducible recombinases in mammalian cells and mice. Nat. Commun. *10*, 4845. https://doi.org/10.1038/s41467-019-12800-7.

54. Hu, J., Meyers, R.M., Dong, J., Panchakshari, R.A., Alt, F.W., and Frock, R.L. (2016). Detecting DNA double-stranded breaks in mammalian genomes by linear amplification–mediated high-throughput genome-wide translocation sequencing. Nat. Protoc. *11*, 853–871. https://doi.org/10.1038/nprot.2016.043.

55. Koeppel, J., Ferreira, R., Vanderstichele, T., Riedmayr, L.M., Peets, E.M., Girling, G., Weller, J., Murat, P., Liberante, F.G., Ellis, T., et al. (2025). Randomizing the human genome by engineering recombination between repeat elements. Science *387*, eado3979. https://doi.org/10.1126/science.ado3979.

56. Pinglay, S., Lalanne, J.-B., Daza, R.M., Kottapalli, S., Quaisar, F., Koeppel, J., Garge, R.K., Li, X., Lee, D.S., and Shendure, J. (2025). Multiplex

generation and single-cell analysis of structural variants in mammalian genomes. Science *387*, eado5978. https://doi.org/10.1126/science.ado5978.

57. Li, J., Gong, S., Ma, Y., Han, P., Wang, N., Fu, Z., Zhang, X., Huang, X., Yang, T., Tong, H., et al. (2025). Creation of a eukaryotic multiplexed site-specific inversion system and its application for metabolic engineering. Nat. Commun. *16*, 1918. https://doi.org/10.1038/s41467-025-57206-w.
58. Han, P., Ma, Y., Fu, Z., Guo, Z., Xie, J., Wu, Y., and Yuan, Y.J. (2021). A DNA Inversion System in Eukaryotes Established via Laboratory Evolution. ACS Synth. Biol. *10*, 2222–2230. https://doi.org/10.1021/acssynbio.1c00132.
59. Pinglay, S., Bulajić, M., Rahe, D.P., Huang, E., Brosh, R., Mamrak, N.E., King, B.R., German, S., Cadley, J.A., Rieber, L., et al. (2022). Synthetic regulatory reconstitution reveals principles of mammalian Hox cluster regulation. Science *377*, eabk2820. https://doi.org/10.1126/science.abk2820.
60. Brosh, R., Coelho, C., Ribeiro-dos-Santos, A.M., Ellis, G., Hogan, M.S., Ashe, H.J., Somogyi, N., Ordoñez, R., Luther, R.D., Huang, E., et al. (2023). Synthetic regulatory genomics uncovers enhancer context dependence at the Sox2 locus. Mol. Cell *83*, 1140–1152.e7. https://doi.org/10.1016/j.molcel.2023.02.027.
61. Blayney, J.W., Francis, H., Rampasekova, A., Camellato, B., Mitchell, L., Stolper, R., Cornell, L., Babbs, C., Boeke, J.D., Higgs, D.R., and Kassouf, M. (2023). Super-enhancers include classical enhancers and facilitators to fully activate gene expression. Cell *186*, 5826–5839.e18. https://doi.org/10.1016/j.cell.2023.11.030.
62. Zhang, W., Golynker, I., Brosh, R., Fajardo, A., Zhu, Y., Wudzinska, A.M., Ordoñez, R., Ribeiro-dos-Santos, A.M., Carrau, L., Damani-Yokota, P., et al. (2023). Mouse genome rewriting and tailoring of three important disease loci. Nature *623*, 423–431. https://doi.org/10.1038/s41586-023-06675-4.
63. Schindler, D., Walker, R.S.K., Jiang, S., Brooks, A.N., Wang, Y., Müller, C.A., Cockram, C., Luo, Y., García, A., Schraivogel, D., et al. (2023). Design, construction, and functional characterization of a tRNA neochromosome in yeast. Cell *186*, 5237–5253.e22. https://doi.org/10.1016/j.cell.2023.10.015.
64. Ma, Y., Su, S., Fu, Z., Zhou, C., Qiao, B., Wu, Y., and Yuan, Y.-J. (2024). Convenient synthesis and delivery of a megabase-scale designer accessory chromosome empower biosynthetic capacity. Cell Res. *34*, 309–322. https://doi.org/10.1038/s41422-024-00934-3.
65. Bolotin, D.A., Poslavsky, S., Mitrophanov, I., Shugay, M., Mamedov, I.Z., Putintseva, E.V., and Chudakov, D.M. (2015). MiXCR: software for comprehensive adaptive immunity profiling. Nat. Methods *12*, 380–381. https://doi.org/10.1038/nmeth.3364.
66. Chen, S., Zhou, Y., Chen, Y., and Gu, J. (2018). fastp: an ultra-fast all-in-one FASTQ preprocessor. Bioinformatics *34*, i884–i890. https://doi.org/10.1093/bioinformatics/bty560.
67. Li, H. (2018). Minimap2: pairwise alignment for nucleotide sequences. Bioinformatics *34*, 3094–3100. https://doi.org/10.1093/bioinformatics/bty191.
68. Danecek, P., Bonfield, J.K., Liddle, J., Marshall, J., Ohan, V., Pollard, M.O., Whitwham, A., Keane, T., McCarthy, S.A., Davies, R.M., and Li, H. (2021). Twelve years of SAMtools and BCFtools. GigaScience *10*, giab008. https://doi.org/10.1093/gigascience/giab008.
69. McInnes, L., Healy, J., Saul, N., and Großberger, L. (2018). UMAP: Uniform Manifold Approximation and Projection. J. Open Source Softw. *3*, 861. https://doi.org/10.21105/joss.00861.
70. Harris, C.R., Millman, K.J., van der Walt, S.J., Gommers, R., Virtanen, P., Cournapeau, D., Wieser, E., Taylor, J., Berg, S., Smith, N.J., et al. (2020). Array programming with NumPy. Nature *585*, 357–362. https://doi.org/10.1038/s41586-020-2649-2.
71. Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., et al. (2011). Scikit-learn: Machine Learning in Python. J. Mach. Learn. Res. *12*, 2825–2830.
72. Virtanen, P., Gommers, R., Oliphant, T.E., Haberland, M., Reddy, T., Cournapeau, D., Burovski, E., Peterson, P., Weckesser, W., Bright, J., et al. (2020). SciPy 1.0: fundamental algorithms for scientific computing in Python. Nat. Methods *17*, 261–272. https://doi.org/10.1038/s41592-019-0686-2.
73. Waskom, M. (2021). seaborn: statistical data visualization. J. Open Source Softw. *6*, 3021. https://doi.org/10.21105/joss.03021.
74. Hunter, J.D. (2007). Matplotlib: A 2D Graphics Environment. Comput. Sci. Eng. *9*, 90–95. https://doi.org/10.1109/MCSE.2007.55.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Antibodies		
Anti-B220–APC (clone RA3-6B2)	eBioscience	Cat# 17-0452-83; AB_467254
Bacterial and virus strains		
E. coli: *Stbl4*	Invitrogen	Cat# 11635018
E. coli: *SynIgh* assembly strain	This paper	N/A
Chemicals, peptides, and recombinant proteins		
Exonuclease V	New England Biolabs (NEB)	Cat# M0345L
Yeast tRNA	Sigma-Aldrich	Cat# R8508
NotI-HF	New England Biolabs (NEB)	Cat# R0189L
Abscisic acid (ABA)	MedChemExpress (MCE)	Cat# HY-100560
Propanediol	Sigma-Aldrich	Cat# 82280
PEG-400	Sigma-Aldrich	Cat# 91893
Tween 80	Sigma-Aldrich	Cat# P8074
PBase mRNA	Genscript	Cat# RP-A00064–0.2
MidRange PFG Marker	New England Biolabs (NEB)	Cat# N0342S
2 kb Plus II DNA Ladder	Transgen-biotech	Cat# BM121-01
Megabase Agarose gel	Bio-Rad	Cat# 1613108
Sodium acetate	Invitrogen	Cat# AM9740
Ethidium bromide	Invitrogen	Cat# 15585011
Chloramphenicol	Sigma-Aldrich	Cat# R4408
Kanamycin	Sigma-Aldrich	Cat# K1876
Carbenicillin	Sigma-Aldrich	Cat# C3416
Hygromycin B	Gibco	Cat# 10687010
Critical commercial assays		
Red Blood Cell Lysing Buffer	Invitrogen	Cat# 00-4333-57
2×Rapid Taq Master Mix	Vazyme	Cat# P222
KOD One™ PCR Master Mix	toyobo	Cat# KMM-101
Large-Construct Kit	QIAGEN	Cat# 12462
NEBNext Ultra II FS DNA Library Prep Kit	New England Biolabs (NEB)	Cat# E7805
AMPure XP beads	Beckman Coulter	Cat# A63881
MagAttract HMW DNA Kit	QIAGEN	Cat# 67563
Ligation Sequencing Kit V14	Oxford Nanopore Technologies	Cat# SQK-LSK114
MinION/GridION Flow Cell (DNA)	Oxford Nanopore Technologies	Cat# FLO-MIN114
PromethION Flow Cells (DNA)	Oxford Nanopore Technologies	Cat# FLO-PRO114M
DNeasy Blood & Tissue Kit	QIAGEN	Cat# 69504
Dynabeads MyOne Streptavidin T1 beads	Thermo Fisher Scientific	Cat# 65601
Deposited data		
Raw FASTQs, *SynIgh*_NGS_*E.coli*_Clone1_Day0	This paper	GSA: SAMC4170447
Raw FASTQs, *SynIgh*_NGS_*E.coli*_Clone1_Day7_stabilizer+	This paper	GSA: SAMC4170446
Raw FASTQs, *SynIgh*_NGS_*E.coli*_Clone1_Day7_stabilizer-	This paper	GSA: SAMC4170445
Raw FASTQs, *SynIgh*_WGS_*S.C*_Clone1_Day30_stabilizer-	This paper	GSA: SAMC4170444
Raw FASTQs, *SynIgh*_WGS_*S.C*_Clone1_Day30_stabilizer+	This paper	GSA: SAMC4170443
Raw FASTQs, *SynIgh*_plasmid_ONT	This paper	GSA: SAMC6857405
Raw FASTQs, SynIgh_WGS_S.C_Clone1_Day0	This paper	GSA: SAMC4170442
Raw FASTQs, *SynIgh*_mus_ONT_WGS	This paper	GSA: SAMC4170788

(*Continued on next page*)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Raw FASTQs, *SynIgh* _BMcell_VDJseq_4	This paper	GSA: SAMC4170792
Raw FASTQs, *SynIgh* _BMcell_VDJseq_3	This paper	GSA: SAMC4170791
Raw FASTQs, *SynIgh* _BMcell_VDJseq_2	This paper	GSA: SAMC4170790
Raw FASTQs, *SynIgh* _BMcell_VDJseq_1	This paper	GSA: SAMC4170789
Raw FASTQs, *SynIgh*$^{+}$*Rag2*$^{-/-}$_pro-B_D	This paper	GSA: SAMC4170799
Raw FASTQs, *SynIgh*$^{+}$*Rag2*$^{-/-}$_pro-B _D_scramble	This paper	GSA: SAMC4170798
Raw FASTQs, *SynIgh*$^{+}$*Rag2*$^{-/-}$_pro-B _V_scramble	This paper	GSA: SAMC4170797
Raw FASTQs, *SynIgh*_BM _Scramble_VDJseq_7	This paper	GSA: SAMC4170806
Raw FASTQs, *SynIgh*_BM _Scramble_VDJseq_6	This paper	GSA: SAMC4170805
Raw FASTQs, *SynIgh*_BM _Scramble_VDJseq_5	This paper	GSA: SAMC4170804
Raw FASTQs, *SynIgh*_BM _Scramble_VDJseq_4	This paper	GSA: SAMC4170803
Raw FASTQs, SynIgh_BM _Scramble_VDJseq_3	This paper	GSA: SAMC4170802
Raw FASTQs, SynIgh_BM _Scramble_VDJseq_2	This paper	GSA: SAMC4170801
Raw FASTQs, SynIgh_BM _Scramble_VDJseq_1	This paper	GSA: SAMC4170800
Experimental models: Organisms/strains		
Mouse: *Rag2*$^{-/-}$	Jiangsu Cyagen Biosciences	Cat# C001324
Mouse: *SynIgh*$^{+}$	This paper	N/A
Mouse: *SynIgh*$^{+}$ *Rag2*$^{-/-}$	This paper	N/A
Yeast: *S. cerevisiae* strain BY4742	ATCC	Cat#4040004
Yeast: *SynIgh* assembly strain	This paper	N/A
E. coli: *Stbl4*	Invitrogen	Cat# 11635018
E. coli: *SynIgh* assembly strain	This paper	N/A
Oligonucleotides		
Biotin-labelled JH-specific primers	This paper	Table S4
Nested JH-specific primers	This paper	Table S4
mSCRaMbLE-seq primers	This paper	Table S5
Recombinant DNA		
SynIgh	This paper	Table S2
Software and algorithms		
Python v3.12.5	Python Software Foundation	https://www.python.org/RRID: SCR_008394
R v4.4.1	R Core Team	https://www.r-project.org/RRID: SCR_001905
MiXCR v4.7.0	Bolotin et al.[65]	https://mixcr.com/RRID: SCR_004797
fastp v0.23. 4	Chen et al.[66]	https://github.com/OpenGene/fastp
minimap2 v2.28	Li[67]	https://github.com/lh3/minimap2 RRID: SCR_018550
SAMtools v1.20	Danecek et al.[68]	https://www.htslib.org/RRID: SCR_002105
UMAP v0.5.6	McInnes et al.[69]	https://github.com/lmcinnes/umap RRID: SCR_018217
NumPy v2.0.0	Harris et al.[70]	https://numpy.org/ RRID: SCR_008633
pandas v2.2.2	pandas development team	https://pandas.pydata.org/ RRID: SCR_018214
scikit-learn v1.5.0	Pedregosa et al.[71]	https://scikit-learn.org/ RRID: SCR_002577
SciPy v1.14.0	Virtanen et al.[72]	https://scipy.org/ RRID: SCR_008058
seaborn v0.13.2	Waskom[73]	https://seaborn.pydata.org/ RRID: SCR_018132
matplotlib v3.9.1	Hunter[74]	https://matplotlib.org/ RRID: SCR_008624

(*Continued on next page*)

Continued		
REAGENT or RESOURCE	SOURCE	IDENTIFIER
plotly v5.23.0	Plotly development team	https://plotly.com/python/ RRID: SCR_013991
Analysis code generated for this manuscript	This paper	Zenodo: https://doi.org/10.5281/zenodo.18576074
Other		
MicroPulser Electroporator	Bio-Rad	Cat# 1652100
CHEF-DR II system	Bio-Rad	Cat# 1703670
NovaSeq 6000 System	Illumina	RRID: SCR_016387
GridION System	Oxford Nanopore Technologies	RRID: SCR_017987
PromethION 2 Solo	Oxford Nanopore Technologies	N/A
FACSAria III	BD Biosciences	RRID: SCR_005996
Qubit	Thermo Fisher Scientific	RRID: SCR_020553
Agilent 2100 Bioanalyzer	Agilent Technologies	RRID: SCR_018043

EXPERIMENTAL MODEL AND STUDY PARTICIPANT DETAILS

Mice

All mice were maintained on a C57BL/6 background under specific pathogen-free (SPF) conditions in the animal facility of Tianjin University. *Rag2*$^{-/-}$ (*Rag2*-KO, Cat. No. C001324) mice were purchased from Jiangsu Cyagen Biosciences. *SynIgh*$^{+}$ and *SynIgh*$^{+}$*Rag2*$^{-/-}$ mice were generated as described below. Mice were housed under a 12 h light/dark cycle at 22 ± 2°C and 50% ± 5% relative humidity and were provided standard chow and water *ad libitum*. Equal numbers of male and female mice were used where applicable. Mice used in this study were experimentally naive before the procedures described here and were not subjected to previous procedures. All animal experiments were approved by the Institutional Animal Care Committee of Tianjin University (TJUE-2024-390).

Yeast strains

Saccharomyces cerevisiae strain *BY4742* was used for SynIgh assembly experiments. *SynIgh* assembly strains were generated by transformation of the stabilizer plasmid and/or *SynIgh* assembly products as described in method details. Yeast was cultured under standard growth conditions at 30°C.

Bacterial strains

Escherichia coli Stbl4 was used for cloning and maintenance of *SynIgh* constructs. *SynIgh* assembly strains were generated by transformation of assembly products into *Stbl4* cells and were cultured under standard conditions at 30°C with the appropriate antibiotics as described below.

METHOD DETAILS

Design and synthesis of *SynIgh*

The germline V, D, and J gene sequences used for SynIgh construction were selected based on population-matched allelic variants derived from the 1000 Genomes Project (May 2013 release).[13] Gene nomenclature and sequence mapping were performed using IMGT/GENE-DB Version 3.1.34. We confirmed that the sequences and nomenclature for the selected 60 V, 31 D, and 13 J genes remain consistent across these versions. The explicit correspondence between *SynIgh* gene names, IMGT identifiers, and their nucleotide sequences is provided in Table S1. The *SynIgh* sequence is provided in Table S2.

The DNA fragments used for assembly were synthesized by Synbio Technologies (Tianjin). The Gibson method was used for DNA assembly. The assembled product was treated with Exonuclease V (NEB, M0345L) according to the manufacturer's instructions to remove unassembled linearized fragments. The product was then mixed with 1 μL of 1 μg/μL yeast tRNA (Sigma–Aldrich, R8508). Subsequently, ethanol precipitation was performed by adding 2–2.5 volumes of 100% ethanol and a 1/10 volume of 3 M sodium acetate (pH 4.8). After incubation at −20°C for 15 min, the sample was centrifuged at 14,000 × g for 10 min at 4°C. The resulting pellet was washed once with 100 μL of 70% ethanol, air-dried, and resuspended in 5 μL of ultrapure water. The assembly products were then introduced into *Stbl4 E. coli* by electroporation via a MicroPulser Electroporator (Bio-Rad) according to the manufacturer's instructions. Plasmids were purified from bacterial cultures grown in Luria broth (LB) supplemented with the appropriate antibiotics (12.5 μg/mL chloramphenicol, 25 μg/mL kanamycin, 50 μg/mL carbenicillin, or 500 μg/mL hygromycin B).

All yeast experiments were conducted with the *BY4742* strain under standard media and growth conditions, whereas all *E. coli* experiments utilized the *Stbl4* strain under standard media and growth conditions. Both yeast and bacteria were cultured at 30°C. Yeast transformations were performed using the lithium acetate method. For yeast colony PCR, a single *Saccharomyces cerevisiae* colony was gently picked with a toothpick and transferred into 30 μL of 20 mM NaOH. The mixture was gently agitated to fully resuspend the colonies. This suspension was then lysed by boiling in a PCR machine at 99°C for 5 min, followed by cooling to 4°C for 1 min, repeated three times. The resulting lysate was used directly as the template for PCR. PCR was performed using Vazyme 2×Rapid Taq Master Mix (Vazyme, P222) according to the manufacturer's instructions. The PCR products were separated on a 1% agarose gel containing ethidium bromide, with a 2 kb Plus II DNA Ladder (Transgen-biotech, BM121-01) used as a molecular weight standard.

Pulsed-field gel electrophoresis

SynIgh assembly constructs were verified by digesting approximately 250–500 ng of plasmid DNA purified by alkaline lysis from small-scale (5–10 mL) saturated bacterial cultures with NotI-HF (New England Biolabs, R0189L). The digestion reactions were conducted at 37°C for 1–2 h. The entire reaction mixture was then separated using the CHEF-DR system (Bio-Rad, 1703670) on a 1% Megabase Agarose gel (Bio-Rad, 1613108) in 0.5X TBE buffer. Electrophoresis was performed at 6 V/cm with ramped pulse times ranging from 1 to 25 s for 24 h at 14°C. The MidRange PFG Marker (NEB, N0342S) was used as a molecular weight standard. After separation, the gels were stained in a 0.5 μg/mL ethidium bromide solution for 20–30 min, followed by destaining in water for 20–30 min before imaging.

DNA-seq

For validation of the *SynIgh* assembly, constructs were extracted from yeast and *E. coli* using the QIAGEN Large-Construct Kit (QIAGEN, 12462) according to the manufacturer's instructions. Approximately 100 ng of DNA was used for library construction with the NEBNext Ultra II FS DNA Library Prep Kit (E7805). DNA purification was performed using AMPure XP beads (Beckman Coulter, A63881) on a magnetic stand. Paired-end 150-bp sequencing was conducted on a NovaSeq 6000 (Illumina) sequencing platform.

Raw reads were processed using fastp (v0.23.4) for adapter trimming and quality filtering. Specifically, adapter trimming was enabled for paired-end reads, and reads were filtered using a Q30 criterion (Phred quality ≥30 for qualified bases; reads failing the Q30 filter after trimming were removed). Post-QC reads were aligned to the expected SynIgh reference sequence using minimap2 (v2.28) in short-read mode (-ax sr). Alignments were coordinate-sorted and indexed, and per-base coverage profiles were computed using SAMtools (v1.20) (e.g., samtools sort/index followed by samtools depth; summary statistics such as mean/median depth were calculated from the depth output). These short-read alignments and coverage metrics were used to confirm sequence concordance and uniformity across the construct.

For nanopore long-read sequencing, assembly, and variant characterization. Construct DNA was extracted as above and 1 μg input DNA was used for Nanopore library preparation. High-molecular-weight DNA was mechanically fragmented by needle shearing prior to library construction with the Ligation Sequencing Kit V14 (Oxford Nanopore Technologies, SQK-LSK114). Sequencing was performed on a MinION device using an R10.4.1 flow cell, achieving approximately ∼1000× depth of coverage across the construct. Basecalling was performed using Dorado (v1.3.0) with the dna_r10.4.1_e8.2_400bps_sup@v5.2.0 model. Reads were filtered to retain Qscore ≥10 for downstream analyses. Long reads were assembled *de novo* using Flye (v1.3.1), and internal polishing was performed using three iterative polishing rounds (–iterations 3) to refine consensus contigs. Assembled contigs were aligned to the expected reference using minimap2 in assembly-to-reference mode (-cx asm5 –cs) to obtain base-level difference strings. Sequence differences (substitutions/insertions/deletions and larger discrepancies) were parsed from the alignment output using custom scripts (see data and code availability) to generate the final set of construct variants and to verify overall sequence concordance (Table S3).

To identify piggyBac insertion junctions, Nanopore reads were scanned for internal terminal repeat (ITR) motifs using seqkit grep with strand-aware matching and mismatch tolerance (seqkit grep -j 32 -s -i -m 2 -f), using ITR-specific sequences: ITR5: ccctagaaagatagtctgcgtaaaattgacgcatg; ITR3-1: catgcgtcaattttacgcatgattatctttaacgtacgtcacaatatgattatctttctaggg; ITR3-2: cctcgatatacagaccgataaaacac. Reads containing ITR sequence were extracted, and the internal ITR portion was trimmed to retain flanking genomic/construct sequence adjacent to the ITR boundary. The resulting flanks were mapped by BLAST against the appropriate reference sequence to determine the insertion locus and orientation, and integration sites were summarized across all supporting reads.

Generation of the *SynIgh* mouse model

SynIgh (174 kb) DNA was co-injected with *PBase* mRNA into C57/B6 zygotes. Founders were identified via PCR, and those verified to have all 30 correct PCR junctions were selected for WGS. For mouse whole-genome sequencing, high-molecular-weight DNA was extracted from mouse tail samples using the MagAttract HMW DNA Kit (QIAGEN, 67563). Libraries were prepared using the Ligation Sequencing Kit V14 (Nanopore, SQK-LSK114). Sequencing was performed on a PromethION system (Nanopore) using a PromethION Flow Cell R10.4.1, achieving ∼30x coverage per mouse. Mice with confirmed sequencing results were then mated

with wild-type (WT) mice to establish lines bearing single-copy transgenes. To construct *SynIgh*$^{+}$*Rag2*$^{-/-}$ mice, *SynIgh*$^{+}$ mice were bred with *Rag2*$^{-/-}$ mice to produce *SynIgh*$^{+}$*Rag2*$^{+/-}$ offspring, which were then bred with *Rag2*$^{-/-}$ mice to produce homozygotes (*SynIgh*$^{+}$*Rag2*$^{-/-}$).

Induction of mSCRaMbLE

Abscisic acid (ABA, MCE, HY-100560) was dissolved at a concentration of 40 mg/mL in a carrier solution containing 16.7% propanediol (Sigma–Aldrich, 82280), 22.5% PEG-400 (PEG-400, Sigma–Aldrich, 91893), and 1.25% Tween 80 (Sigma–Aldrich, P8074), which was then diluted in physiological saline. The solution was sonicated in a water bath to ensure complete dissolution. The drug was administered via intraperitoneal injection at a dosage of 200 mg/kg ABA. For mice aged 10–14 days (P10–14), 100 μL of the ABA solution was injected intraperitoneally once daily for five consecutive days. After the injections, the mice were maintained for an additional two weeks, after which their bone marrow were harvested for further analysis.

Bone marrow cell purification

For bone marrow VDJ-seq, single-cell suspensions were derived from the bone marrow of 4–6-week-old *SynIgh*$^{+}$ mice and incubated in Red Blood Cell Lysing Buffer (Invitrogen, 00-4333-57) to deplete erythrocytes. For the detection of synthetic rearrangement in bone marrow pro-B cells, B220^{+} bone marrow pro-B cells were purified via FACS sorting (FACSAria III, BD Biosciences) from 4–6-week-old *SynIgh*$^{+}$*Rag2*$^{-/-}$ mice using anti-B220–APC (1:1,000 dilution; eBioscience, 17-0452-83).

VDJ-seq

VDJ-seq analysis of the *SynIgh* locus was conducted as previously described[50] with the following modifications. (1) genomic DNA was extracted from approximately 10 ^6 bone marrow cells using the Qiagen DNeasy Blood & Tissue Kit (QIAGEN, Cat. No. 69504), incorporating an RNase A treatment step to ensure purity. DNA quality was verified ($A_{260}/A_{280} \approx 1.8$) prior to library preparation. (2) Custom adapters (Table S4) were ligated to the fragmented DNA to enable absolute molecular counting and PCR error correction. Instead of physical shearing, DNA was enzymatically fragmented, end-repaired, and adapter-ligated in a single workflow using the NEBNext Ultra II FS DNA Library Prep Kit (NEB, E7805). Post-ligation purification was performed using 1X AMPure XP beads (Beckman Coulter, A63881) to efficiently remove unligated adapters and small fragments. (3) Single-cycle primer extension using J-specific biotinylated primers. Adapter-ligated DNA was split into multiple 50-μL reactions ($\leq$1 μg DNA per reaction). Primer extension was performed using a pooled set of outer JH reverse biotinylated primers and an exonuclease-deficient DNA polymerase. Extension was carried out with a single extension cycle using the following thermal program: 95°C for 4 min; 59°C for 5 min, 72°C for 15 min. Reactions were pooled and purified using a PCR purification column kit, and DNA was eluted in EB buffer. (4) Streptavidin capture and on-bead PCR amplification. Biotinylated primer-extension products were captured using Dynabeads MyOne Streptavidin T1 beads (Thermo Fisher Scientific, 65601). Beads were washed three times in 1× B&W buffer, resuspended in 2× B&W buffer, and incubated with the primer-extended DNA overnight at room temperature with rotation. Beads were then magnetically separated, and the bound products were washed twice with 1× B&W buffer and once with EB buffer before resuspension in EB buffer. PCR was performed directly on the beads using nested JH-specific primers and an adapter-targeting primer; Illumina flow-cell binding sequences and multiplex indices were incorporated using Illumina indexing primers in the final PCR step (15 cycles). After amplification, libraries were purified and size-selected using AMPure XP beads to remove residual primer-dimers and to retain fragments >400 bp (bead-based size selection). (5) Libraries were quantified using a Qubit fluorometer, and fragment size distribution was assessed on an Agilent 2100 Bioanalyzer. Libraries showed a predominant size range of 200–1,000 bp with no prominent primer-dimer peak. Paired-end 250-bp sequencing was performed on an Illumina NovaSeq 6000 platform.

Raw sequencing data were demultiplexed, adapter-trimmed, and quality-filtered (>Q30). Alignment and clone assembly were performed using MiXCR (v4.7.0) against a custom reference library containing all *SynIgh* V, D, and J genes. The analysis utilized the generic-amplicon preset with the following parameters: –library custom.json –species humouse –dna –floating-left-alignment-boundary –floating-right-alignment-boundary J. Clonotypes were assembled based on the region spanning CDR1 to CDR3 using the parameter -OassemblingFeatures = ''[CDR1+FR2+CDR2+FR3+CDR3]'', with a minimum sequence length cutoff of 200 bp (-OminimalClonalSequenceLength = 200). Final clonotype data, including gene hits (vHit, dHit, jHit), family usage, and productivity status (isProductive, isOOF, hasStops), were exported for downstream analysis. The utilization of each gene along with the dimensionality reduction of VDJ products was analyzed using a custom pipeline (see data and code availability statement).

mSCRaMbLE-seq

Purified pro-B cells from 4 *SynIgh*$^{+}$*Rag2*$^{-/-}$ mice were pooled together for mSCRaMbLE-seq experiment. To achieve higher coverage and unbiased identification of mSCRaMbLE rearrangement types, we modified the LAM-HTGTS-seq[54] method and adapted it for Nanopore long-read sequencing. The steps were as follows: (1) Genomic DNA was extracted from pooled pro-B cells using the MagAttract HMW DNA Kit (QIAGEN, Cat. No. 67563) following the manufacturer's high-molecular-weight protocol. (2) Linear amplification. Junction-containing DNA sequences were enriched via a single cycle of Linear Amplification-Mediated PCR (LAM-PCR). A 5′-biotinylated primer targeting the *SynIgh* V-region and carrying Adapter 1 at the 5′ end was utilized (Table S5). The reaction was performed using KOD One PCR Master Mix (Toyobo, Cat. No. KMM-101) with the following thermal profile: denaturation at 98°C for 10 min, followed by a long annealing and extension step at 68°C for 10 min. Post-reaction, the products were purified using 0.5X

AMPure XP beads (Beckman Coulter) to remove excess primers and small fragments. The purified DNA was then heat-denatured at 95°C for 5 min and immediately snap-cooled on ice to generate single-stranded DNA (ssDNA) for capture. (3) Streptavidin bead capture. Biotinylated LAM-PCR products were captured using Dynabeads MyOne Streptavidin C1 (Life Technologies) according to the manufacturer's recommendations. DNA–bead complexes were washed extensively with 1× Binding & Washing (B&W) buffer (1 M NaCl, 5 mM Tris-HCl, 0.5 mM EDTA) to remove unbound DNA and residual reagents. (4) On-Bead Ligation: While bound to beads, Adapter 2 was ligated to the captured DNA using T4 DNA Ligase under standard ligation conditions. Beads were washed following ligation to remove free adapters and enzymes. (5) The adapter-ligated templates were amplified using primers targeting Adapter 1 and Adapter 2 for 20 PCR cycles. Subsequently, the amplicons were size-selected to enrich for fragments larger than 2 kb.

High–molecular weight genomic DNA was prepared using the Ligation Sequencing Kit V14 (Oxford Nanopore Technologies; SQK-LSK114) according to the manufacturer's instructions. Sequencing was performed on a PromethION 2 Solo instrument using PromethION R10.4.1 flow cells. Basecalled FASTQ reads with a mean quality score >10 were retained for downstream analyses.

A custom computational pipeline was used to reconstruct and classify mSCRaMbLE-seq data (see data and code availability). The workflow consisted of four key steps: (1) Anchor scanning and subsequence extraction. Reads were scanned for a anchor motif (*loxPsym*, ATAACTTCGTATAATGTACATTATACGAAGTTAT) using sliding-window matching allowing up to six mismatches. Inter-anchor subsequences were extracted (anchors excluded) with base qualities retained, and labeled by parent read ID and start/end coordinates. (2) Mapping to a block reference. SynIgh V units were split into individual blocks to build block60.fasta. Subsequences were mapped to block60.fasta using minimap2 (-ax map-ont), and SAM files were generated for downstream parsing. (3) Trajectory reconstruction and filtering. The primary alignment per subsequence was retained. Subsequences were aggregated to parent reads, ordered by subsequence start coordinate, and concatenated into signed block trajectories (±block index). Reads with unmapped segments were removed, and only trajectories spanning ≥2 V units per read were kept. (4) Junction classification. Signed segments were parsed and analyzed as adjacent pairs (window size 2). Pairs were classified as *unchanged* (same sign, adjacent indices), *Del* (same sign, non-adjacent indices), or *Inv* (opposite signs, different indices).

QUANTIFICATION AND STATISTICAL ANALYSIS

Unless otherwise indicated, data are presented as mean ± s.e.m. The exact n, the definition of n, and the statistical details for each experiment are provided in the figure legends and below. Biological replicates refer to independent mice or independently analyzed colonies, as specified. No statistical methods were used to predetermine sample size. Data collection and analysis were not performed blind to the conditions of the experiments.

For yeast assembly fidelity assays (Figure 2H), data are represented as means ± s.e.m. (n = 10 biological replicates for wild-type; n = 4 biological replicates for stabilizer constructs). Figure 3: VDJ-seq analysis was performed on bone marrow cells from n = 4 independent *SynIgh* mice. Figure 4: For mSCRaMbLE-seq analysis, purified pro-B cells from n = 4 *SynIgh*$^{+}$*Rag2*$^{-/-}$ mice were pooled to generate a representative library. Figure 5: VDJ repertoire analysis compared *SynIgh*$^{+}$ mice (n = 4) and *SynIgh*$^{+}$+mSCRaMbLE mice (n = 7).

The Pearson correlation coefficient (PCC) was calculated using the scipy.stats.pearsonr[71] function to quantify the similarity of V gene utilization between individual mice or groups (Figures 3F and 5D). Principal Component Analysis (PCA) was performed using the sklearn.decomposition.PCA[72] module to visualize the divergence of VDJ recombination product repertoires based on gene usage type and frequency (Figure 5E). UMAP analysis was performed using the umap-learn package[69] (v0.5.6) for dimensionality reduction of VDJ products.