Human Immunology 86 (2025) 111272

Contents lists available at ScienceDirect

journal homepage: www.elsevier.com/locate/humimm

Research article

JoGo-LILR caller: Unveiling and navigating the complex diversity of LILRB3-LILRA6 copy number haplotype structures with whole-genome sequencing

Masao Nagasaki [a,b,*], Kouyuki Hirayasu [c,d,e,f], Seik-Soon Khor [g,h], Ryoko Otokozawa [a], Yayoi Sekiya [a], Yosuke Kawai [g], Katsushi Tokunaga [g]

[a] Division of Biomedical Information Analysis, Medical Research Center for High Depth Omics, Medical Institute of Bioregulation, Kyushu University, Fukuoka, Japan
[b] Center for Genomic Medicine, Graduate School of Medicine, Kyoto University, Kyoto, Japan
[c] Advanced Preventive Medical Sciences Research Center, Kanazawa University, Kanazawa, Japan
[d] Department of Evolutionary Immunology, Graduate School of Advanced Preventive Medical Sciences, Kanazawa University, Kanazawa, Japan
[e] Department of Immunology, Graduate School of Medical Sciences, Kanazawa University, Kanazawa, Japan
[f] Department of Immunology, School of Medical and Pharmaceutical Sciences, Kanazawa University, Kanazawa, Japan
[g] Genome Medical Science Project, National Center for Global Health and Medicine, Tokyo, Japan
[h] Singapore Centre for Environmental Life Sciences Engineering, Nanyang Technological University, 637551 Singapore, Singapore

ARTICLE INFO

Keywords:
LILRB3
LILRA6
Short read sequencing
Copy number variation
Genetic diversity

ABSTRACT

Leukocyte immunoglobulin-like receptors (LILRs), encoded on human chromosome 19q13.4, comprise a set of 11 immunoglobulin superfamily receptors known for their genetic heterogeneity. Notably, *LILRB3* and *LILRA6* within this cluster exhibit pronounced sequence homology in immunoglobulin-like domains involved in ligand binding and variable copy number (CN) states. However, understanding their precise role remains challenging. To address this difficulty, we developed an algorithm and tool named JoGo-LILR Caller, which jointly calls CNs of *LILRB3* and *LILRA6* from a population-scale whole-genome short-read sequencing dataset. This tool was applied to 2,504 international HapMap samples and yielded a global CN profile. The 100 % concordance rate corroborated this profile with the CN data obtained from 40 samples of pangenome reference assemblies provided by the Human Pangenome Reference Consortium (HPRC). The frequencies of LILRB3-LILRA6 CN haplotype structures were also estimated for five continental groups with a global CN profile. The established allele frequency profile allowed our tool to estimate LILRB3-LILRA6 CN haplotype combinations. JoGo-LILR-trio enhanced the prediction reliability for haplotype pairs within trio datasets, with trio analysis on 40 child samples demonstrating a 100 % concordance between the predicted pair of haploid CN types and the diploid reference assemblies. Its utility will extend to facilitating software advancements for imputing LILRB3-LILRA6 CN types from SNP array genotyping data, enabling subsequent association analyses that link these CN types to diverse phenotypic traits and diseases, e.g., inflammatory bowel diseases and Takayasu arteritis.

1. Introduction

The leukocyte immunoglobulin-like receptor (LILR) family, an important component encoded on the reverse strand of human chromosome 19q13.4, plays a fundamental role in modulating immune responses. This diverse family consists of inhibitory (LILRB1, LILRB2, LILRB3, LILRB4, and LILRB5) and activating (LILRA1, LILRA2, LILRA4, and LILRA5) receptors, and one secretory protein (LILRA3), intricately contributing to the dynamic equilibrium in immune regulation (Fig. 1). The LILR family is predominantly expressed in myeloid

Abbreviations: AGC, Assembled Genome Compressor; JoGo-LILR, Jonit Open Genome and Omics Platform for Leukocyte Immunoglobulin-Like Receptor; LILR, Leukocyte Immunoglobulin-Like Receptor; LILRB3, Leukocyte Immunoglobulin-Like Receptor Subfamily B Member 3; LILRA6, Leukocyte Immunoglobulin-Like Receptor Subfamily A Member 6; CN, Copy Number; CNV, Copy Number Variation; srWGS, Short-Read Whole-Genome Sequencing; HPRC, Human Pangenome Reference Consortium; 1kGP, 1000 Genomes Project; PCR, Polymerase Chain Reaction; BWA, Burrows-Wheeler Aligner; BAM, Binary Alignment Map; CRAM, Compressed Reference-Aligned Map; IGV, Integrative Genomics Viewer; HLA, Human Leukocyte Antigen; QC, Quality Control; MAPQ, Mapping Quality.

* Corresponding author at: Division of Biomedical Information Analysis, Medical Research Center for High Depth Omics, Medical Institute of Bioregulation, Kyushu University, Fukuoka, Japan.
E-mail address: nagasaki@csml.org (M. Nagasaki).

https://doi.org/10.1016/j.humimm.2025.111272
Received 23 October 2024; Revised 3 February 2025; Accepted 20 February 2025

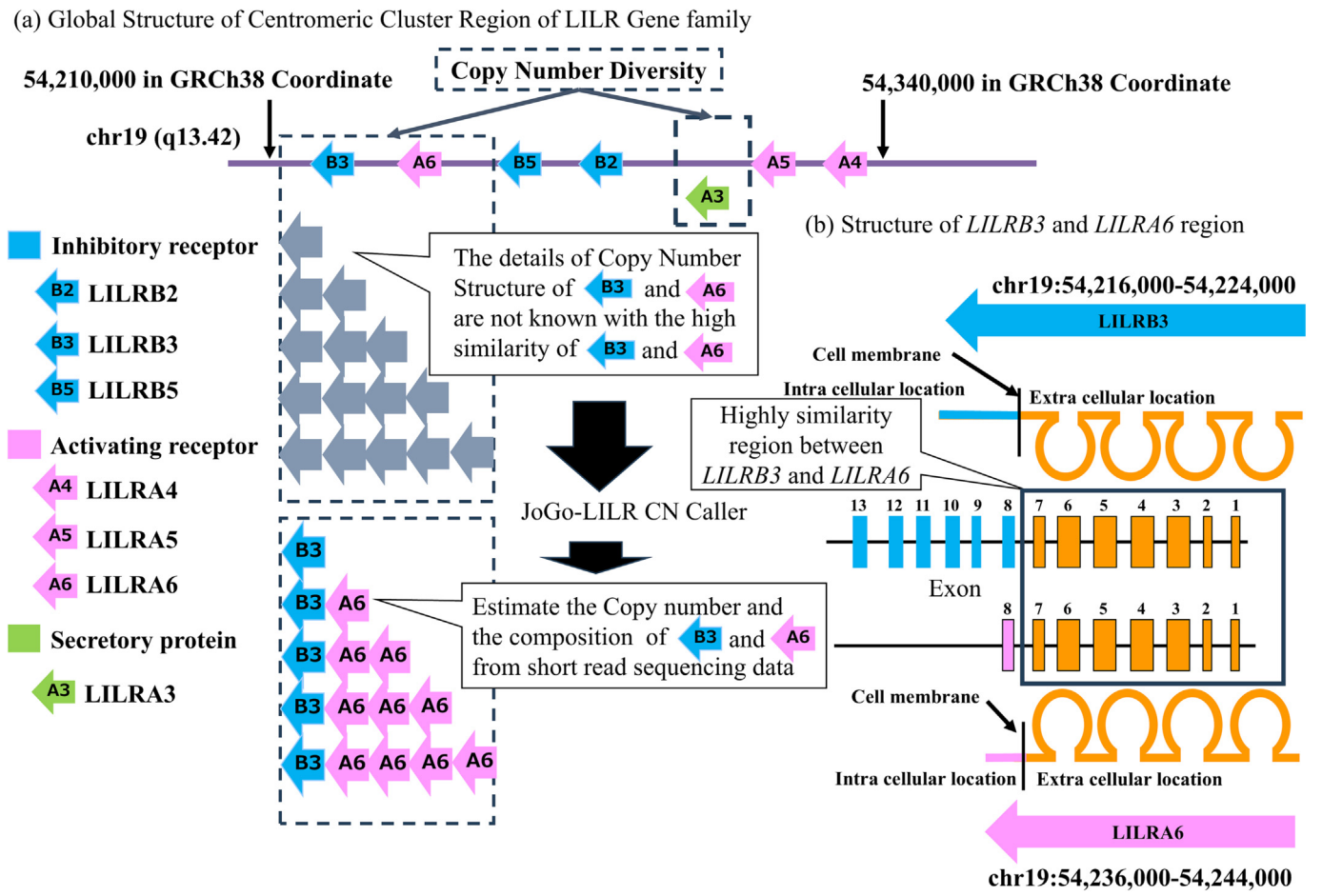

Fig. 1. Global Structure of the Centromeric Cluster Region of the LILR Gene Family and Gene Structures of LILRB3 and LILRA6. The figure provides an overview of the centromeric cluster region (GRCh38: chr19:54,210,000–54,340,000) of the LILR gene family and the *LILRB3* and *LILRA6* region with high copy number diversity. (a) The upper panel shows the genomic organization (all reverse strands) of the LILR gene family within the centromeric cluster, highlighting inhibitory receptors (LILRB2, LILRB3, LILRB5), activating receptors (LILRA4, LILRA5, LILRA6), and the secretory protein (LILRA3). Arrows represent these genes's relative positions and orientations, and regions with copy number diversity are indicated. The high similarity between *LILRB3* and *LILRA6* regions poses challenges in determining their precise total copy number (gray arrows) and the composition of *LILRB3* (blue arrows) and *LILRA6* copies (pink arrows). (b) The lower-right panel details the gene structures of *LILRB3* and *LILRA6*, including their exonic regions (GRCh38: chr19:54,216,000–54,244,000). Exons with high sequence similarity between *LILRB3* and *LILRA6* are shaded in orange, highlighting their structural resemblance. The intra- (blue and red) and extracellular locations (orange) of these gene regions are depicted.

cells, although certain members are also expressed in other cell types, including lymphoid cells (LILRB1), neurons, and hematopoietic stem cells (LILRB2) [1]. The LILR family is important for various physiological processes, including inflammation [2], immune response [3], immune tolerance [4], and cellular differentiation [5].

As depicted in Fig. 1, *LILRB3* and *LILRA6* display notable structural diversity, with highly polymorphic copy number variations compared to other LILR family members. This variability may influence immune regulation and contribute to disease susceptibility. Characterizing these structural features would enhance our understanding of immune system function.

Beyond HLA class I molecules, the LILR family interacts with a diverse array of non-HLA ligands, including apolipoprotein E4 (APOE4), angiopoietin-like proteins (ANGPTLs), galectins, cytokeratin-associated ligands, microbially cleaved immunoglobulins, and virus-derived proteins [6–12]. This extensive range of ligand recognition emphasizes the multifaceted roles of the LILR family in a variety of physiological processes.

The roles of LILR genes in health and disease are multidimensional. These genes are central to the body's response to infection, the pathology of inflammatory diseases, and cancer progression. Their interactions with primary HLA class I alleles are crucial in mediating disease associations, highlighting their importance in immune response and susceptibility to diseases, such as autoimmune and infectious diseases [1], cardiovascular diseases [13], ankylosing spondylitis [14], Takayasu arteritis [15,16], and prostate cancer [17].

Notably, variations in specific LILR genes, namely *LILRB3* and *LILRA6*, have been linked to various immune regulation and are associated with spectrum diseases, including Takayasu arteritis [15], inflammatory bowel diseases [18], ovarian cancer [19], and atopic dermatitis (AD) [20]. For instance, genome-wide copy number (CN) variation association studies estimated from the SNP array have discovered duplications in *LILRA6* to an increased risk of epithelial ovarian cancer and high-grade serous ovarian cancer [19]. In the AD study, while the copy number of *LILRA6* did not differ between unaffected and AD individuals, the lack of *LILRA6* was found to be undertransmitted within families affected by AD, suggesting a protective effect against the disease. In contrast, the presence of one copy of *LILRA6* was modestly over-transmitted, potentially due to the activating nature of this receptor triggering an immune response that promotes AD development [12]. These observations support the role of *LILRA3* and *LILRA6* CN variations in modulating immune response

and potentially influencing disease susceptibility and progression. Elucidating CN variations in these genes is useful for a better understanding the diversity of immune responses and disease susceptibility. Thus, developing of CN-calling software in these areas will accelerate these studies.

Several bioinformatics tools have been developed to determine CNs or haplotypes from short-read whole-genome sequencing data (srWGS) of other human genomic regions. Notable examples include Cyrius, which estimates the haplotypes of the *CYP2D6* and *CYP2D7* regions from srWGS data [21], HLA-VBSeq [22,23] and HLA-HD for HLA haplotyping [24], and LPA Caller, which identifies the LPA Kringle-IV-2 VNTR unit CN [25].

In the context of the challenging genomic regions of *LILRB3* and *LILRA6*, our study draws upon methodologies that have been honed in these other genomic areas. We introduce JoGo-LILR Caller (JoGo-LILR), a tool engineered to delineate the genomic structure of *LILRB3* and *LILRA6* CNs from srWGS data (Fig. 1). This tool carefully identifies and utilizes the distinct core genomic regions unique to *LILRB3* and *LILRA6*. It employs a clustering plot mechanism to enable the interpretation of LILRB3-LILRA6 CN structures through joint-calling operations on multiple samples and scaling to thousands.

We applied the JoGo-LILR method to a dataset comprising 2,504 srWGS samples from the 1000 Genomes Project (1kGP) [26]. The 1kGP, succeeding the international HapMap Project [27], aimed at providing a comprehensive catalog of human genetic variation by sequencing the genomes of 2,504 individuals from 26 geographically diverse populations across five continental groups (African, American, East Asian, European, and South Asian; Supplementary Table 1). This design ensures a balanced representation, with approximately one hundred unrelated samples from each population, to characterize variants with an allele frequency of 1 % or higher within each group. The HapMap samples were recruited at each location (most of the population cares about the grandparent's origins) and can be regarded as representative of general populations, as they were not strictly filtered based on any specific disease conditions. The 1kGP revealed that populations from the same continental group tend to share a high proportion of genetic variants due to shared ancestry and historical gene flow. Genetic differentiation is more pronounced between populations from different continents, reflecting migration history, genetic drift, and adaptation to local environments. These findings demonstrate the significant impact of geography on genetic diversity [28].

For these 2,504 samples, JoGo-LILR not only estimates the haplotype pairs of LILRB3-LILRA6 CN types by solving an optimization problem but also infers the frequencies of these CN types globally and across five continental groups. To ensure the validity of inferred haplotype pairs, the optimization process derives the frequencies of LILRB3-LILRA6 CN haploid structures by solving equations that define the relationships between diploid CN types and their constituent haploid structures. The method minimizes discrepancies between observed CN type distributions and estimated haplotype structure frequencies, ensuring biologically plausible and consistent results. Utilizing the estimated population frequency data as prior knowledge, JoGo-LILR further refines the prediction of a pair of LILRB3-LILRA6 CN haplotype structure probabilities. To evaluate its accuracy, we specifically assessed JoGo-LILR's performance against 40 pangenome reference assembly datasets from the Human Pangenome Reference Consortium (HPRC) project [29], representing a subset of the HapMap samples.

Additionally, we developed JoGo-LILR-trio, an extension of the tool designed to enhance the accuracy of LILRB3-LILRA6 CN-type estimations in trios. This version integrates Mendelian inheritance principles and parental pair of CN haplotype structure probabilities. Applied to a cohort of 602 complete trios [26], JoGo-LILR-trio predicted the most probable pair of LILRB3-LILRA6 CN haplotype structures for each child sample. These predictions were validated using 40 child samples from diploid reference assemblies with paternal or maternal origins from HPRC. JoGo-LILR is available on the Joint Open Genome and Omics Portal (https://jogo.csml.org/JoGo-LILR/).

2. Material and methods

2.1. Download of srWGS and realignment to reference assembly

The srWGS data of the original 2,504 1kGP unrelated samples, which exclude close familial relationships such as parents, children, or siblings to ensure independence in genetic analyses, along with an additional 698 related samples from a total of 3,202 samples [26] were downloaded from the International Genome Sample Resource website (https://www.internationalgenome.org/data-portal/data-collection/30x-grch38). WGS data were prepared using the TruSeq DNA PCR-Free High Throughput Library Prep Kit and sequenced on an Illumina NovaSeq 6000 system using 2 × 150 bp cycles (NovaSeq 6000 S4 Reagent Kit; NovaSeq Xp Kit; PhiX v3 Control (Illumina, San Diego, CA, USA)). For the CRAM file, the reference assembly of GRCh38DH was realigned using BWA (ver. 0.7.17-r1188) [30] with the option "mem -p t24". For the aligned BAM file, we constructed a regional BAM file between chr19:54,216,000–54,243,000 that thoroughly covered the *LILRB3* to *LILRA6* regions.

2.2. Mapping quality analysis to the aligned results

For each 2,504 1kGP unrelated sample, the mean mapQ of the sequenced reads of each chromosomal position was calculated. For the mean mapQ of 2,504 1kGP samples to a chromosomal position (in our case, a mean value from 2,504 samples), we calculated the mean mapQ to the position.

2.3. Calculation of normalized depth for LILRB3core, LILRA6core, and LILRB3 + LILRA6 regions

For the LILRB3core (chr19:54,216,000–54,219,900), LILRA6core (chr19:54,236,600–54,239,600), and LILRB3 + LILRA6 (chr19:54,216,000–54,223,100 and chr19:54,236,000–54,243,000) regions, the normalized depth was calculated using CNVNator ver. 0.4.1 through the region-specific call mode after creating the root file with a 100 bp bin, i.e., LILRB3core, LILRA6core, and LILRB3 + LILRA6 (for the autosomal region, the value would be two if the normalized depth of coverage of the region was the same as that of the reference assembly).

2.4. JoGo-LILR CN call and plot

By taking the vector (LILRB3 + LILRA6, LILRB3core/LILRA6core) from input samples, for example, 2,504 and 3,202 samples in our analysis, single linkage hierarchical clustering (using SciPy library ver.1.10.1 in Python 3.10) was applied, and the cluster group was created by the specified size (in our analysis, 20 using the fclust in the SciPy library; to avoid an infinity value, if LILRB3core/LILRA6core was higher than 4, it was treated as 4). For each cluster, the minimum distance to the theoretical anchor position of the LILRB3-LILRA6 CN type was selected as the CN type (the reference anchor position of each LILRB3-LILRA6 CN type is in Supplementary Table 2). Notably, more than one cluster may be annotated to the same LILRB3-LILRA6 CN type. After each cluster was annotated to an LILRB3-LILRA6 CN type, all samples were plotted against the LILRB3-LILRA6 CN type using a scatter plot (x-axis: LILRB3 + LILRA6, y-axis: LILRB3core/LILRA6core) constructed using the Plotly library ver. 5.18.0.

K-means clustering and DBSCAN were applied to the same input data for comparison with single linkage clustering. K-means clustering was performed using the scikit-learn library (ver.1.3.2) in Python 3.10. The number of clusters (n_clusters) was set to 20, matching the cluster size used in the single linkage. Clusters were assigned using the fit_pre-

dict method, and cluster centers were calculated. The cluster centers were then compared to the theoretical anchor positions, as in single linkage, to assign CN types. DBSCAN clustering was conducted using the scikit-learn library (ver.1.3.2) with fixed min_samples = 1 and varying eps values (e.g., 0.1, 0.5, 1.0) as the algorithm does not allow explicit specification of a fixed cluster size (e.g., n = 20). Clusters formed by DBSCAN were assigned CN types based on their center positions relative to the theoretical anchor positions.

2.5. Validation dataset of LILRA6-LILRB3 copy number structure from diploid reference assemblies

The diploid human pangenome reference dataset (v1) compressed using the assembled genome compressor (AGC) format was downloaded from the HPRC repository (https://github.com/human-pangenomics/HPP_Year1_Data_Freeze_v1.0) [29]. FASTA files were extracted from the AGC file for 40 HapMap samples shared with 3,202 HapMap samples with srWGS using APC tool (https://github.com/refresh-bio/agc). The FASTA files were aligned to GRCh38 reference assembly using Minimap2 (ver. 2.23) with '-ax asm5' option [31] and inspected using IGV (ver. 2.16.2) [32] to determine the CNs and the assembly status of the two alleles for each sample. For HG01952, HG02145, HG02818, and HG02886, which exhibited ambiguity in the CN structures, the latest HPRC assembled dataset (v1.0.1 (r2)) was additionally downloaded from the HPRC GitHub website and processed following the same procedures applied to the HPRC v1 dataset. To assess the statistical power of the algorithm, we performed a power analysis using a validation dataset of 40 samples. The CN types observed in the validation set were distributed as follows: CN2_B1A1 (0 samples), CN2_B2A0 (0 samples), CN3_B2A1 (6 samples), CN4_B2A2 (20 samples), CN5_B2A3 (9 samples), CN6_B2A4 (3 samples), CN7_B2A5 (0 samples), CN8_B2A6 (2 samples), and CN9_B2A7 (0 samples). The corresponding global population frequencies for these CN types were CN2_B1A1 (0.0012), CN2_B2A0 (0.0044), CN3_B2A1 (0.0667), CN4_B2A2 (0.5919), CN5_B2A3 (0.2672), CN6_B2A4 (0.0539), CN7_B2A5 (0.0120), CN8_B2A6 (0.0024), and CN9_B2A7 (0.0004) (these probabilities were calculated in the next section). The probability of achieving the observed concordance rate of 100 % under the global frequency distribution was calculated using the formula:

$$P(\text{All Correct}) = \prod_{i=1}^{n} f_i^{k_i},$$

where f_i is the frequency of CN type i in the global population, and k_i is the count of CN type i in the validation set.

The effect size between the observed concordance rate (100 %) and the expected concordance rate derived from the global population frequency was calculated using:

$$\text{Effect Size} = 2 \times \left(\arcsin\left(\sqrt{p_1}\right) - \arcsin\left(\sqrt{p_2}\right)\right),$$

where $p_1 = 1.0$ and $p_2 = 1.45 \times 10^{-26}$.

Using this effect size, a sample size of 40, a significance level α of 0.05, and degrees of freedom of 8 (calculated as the number of CN types minus one), the statistical power was determined.

2.6. Estimation of the frequencies of LILR CN haploid structures

To estimate the frequencies of the LILRB3-LILRA6 CN haploid structures from the nine LILRB3-LILRA6 CN types, we estimated the frequencies of seven LILRB3-LILRA6 CN haploid structures, CN0_B0-A0, CN1_B1-A0, CN2_B1-A1, CN3_B1-A2, CN4_B1-A3, CN5_B1-A4, and CN6_B1-A5 (the sum should be 1). We employed an optimization approach with sequential least quadratic programming, utilizing the optimized function in SciPy ver. 1.10.1. This method minimizes the difference between the observed result (the left term) and the estimated variables (the right term) in nine equations, aiming to accurately estimate frequencies of CN haplotype structures. The equations are as follows:

(i) CN2_B1A1 = 2 × CN0_B0-A0 × CN2_B1-A1
(ii) CN2_B2A0 = CN1_B1-A0 ×CN1_B1-A0
(iii) CN3_B2A1 = 2 × CN1_B1-A0 × CN2_B1-A1
(iv) CN4_B2A2 = 2 × CN1_B1-A0 × CN3_B1-A2 + CN2_B1-A1 × CN2_B1-A1
(v) CN5_B2A3 = 2 × CN1_B1-A0 × CN4_B1-A3 + 2 × CN2_B1-A1 × CN3_B1-A2
(vi) CN6_B2A4 = 2 × CN1_B1-A0 × CN5_B1-A4 + 2 × CN2_B1-A1 × CN4_B1-A3 + CN3_B1-A2 × CN3_B1-A2
(vii) CN7_B2A5 = 2 × CN1_B1-A0 × CN6_B1-A5 + 2 × CN2_B1-A1 × CN5_B1-A4 + 2 × CN3_B1-A2 × CN4_B1-A3
(viii) CN8_B2A6 = 2 × CN2_B1-A1 × CN6_B1-A5 + 2 × CN3_B1-A2 × CN5_B1-A4 + CN4_B1-A3 × CN4_B1-A3
(ix) CN9_B2A7 = 2 × CN3_B1-A2 × CN6_B1-A5 + 2 × CN4_B1-A3 × CN5_B1-A4

2.7. JoGo-LILR-trio: haploid CN combination and probability estimation

For trio data, the JoGo-LILR-trio algorithm determines the most probable pair of LILRB3-LILRA6 CN haploid structures for the maternal and paternal genomes, ensuring compliance with Mendelian inheritance principles. The algorithm first identifies all possible combinations of maternal and paternal haploid CNs that could result in the observed diploid CN of the child. This approach ensures that all potential combinations adhering to Mendelian inheritance rules are considered.

Next, the algorithm incorporates population-specific haplotype frequency data to calculate the likelihood of each combination. This probabilistic framework refines predictions by leveraging prior knowledge of population haplotype frequencies, ensuring consistency with the diploid CN results for both the parents and the child. Finally, the combination with the highest likelihood is reported as the most probable haploid CN structure for the trio. The algorithm also outputs the probability of the selected combination, providing a quantitative measure of the likelihood for the prediction.

This probabilistic approach allows JoGo-LILR-trio to provide robust and reliable predictions of haploid CN combinations for trio data, making it suitable for downstream analyses.

2.8. Validation dataset of paternal and maternal LILRA6-LILRB3 haploid CN structure from diploid reference assemblies

All 40 samples from the HPRC dataset were children from trio datasets, and the diploid reference assemblies included annotations indicating the maternal or paternal origin of each sequence. Based on these annotations, the most probable maternal and paternal haploid CN pairs estimated by JoGo-LILR-trio were compared to the corresponding pairs derived from the HPRC diploid reference assemblies.

3. Results

3.1. Global overview of the LILR centromeric cluster and LILRB3-LILRA6 structures

The centromeric cluster region of the LILR gene family, located on GRCh38: chr19:54,210,000–54,340,000, exhibits complex genomic organization, as depicted in Fig. 1. The upper panel highlights the reverse strand alignment of inhibitory receptors (LILRB2, LILRB3, LILRB5), activating receptors (LILRA4, LILRA5, LILRA6), and the secretory protein (LILRA3), along with regions of notable copy number diversity. The close structural similarity between the *LILRB3* and *LILRA6* regions introduces challenges in determining their precise

copy number and composition, as indicated by the gray, blue, and pink arrows.

The lower-right panel in Fig. 1 further details the gene structures of *LILRB3* and *LILRA6*, including their exonic regions (GRCh38: chr19: 54,216,000–54,244,000). Exons with high sequence similarity between the two genes are shaded in orange, emphasizing their structural resemblance. These features underline the significance of accurately characterizing the LILRB3-LILRA6 regions to enhance our understanding of their role in immune regulation.

3.2. Notations and abbreviations of CNs of the LILRB3 and LILRA6 regions

Building on previous discussions suggesting that *LILRB3* and *LILRA6* exhibit high copy number diversity, this study further characterizes and analyzes these regions by examining combinations of diploid copy numbers and estimating haploid CN pair structures. To achieve this, relevant terminology and notations are defined to ensure consistent representation and analysis of *LILRB3* and *LILRA6* copy number variations (Supplementary Table 2).

Fig. 2 illustrates the key concepts and terminologies used in this study. As shown in Fig. 2 (a), the haploid CN structure of *LILRB3* and *LILRA6* within a single haplotype is defined. The total CN and the individual contributions of *LILRB3* and *LILRA6* represent this structure. For example, CN3_B1-A2 represents a haploid structure with one copy of LILRB3 (denoted as B1) and two copies of LILRA6 (denoted as A2), summing to a total of three copies (CN3). This notation describes the organization of *LILRB3* and *LILRA6* in a single haplotype and serves as a basis for further analyses.

As illustrated in Fig. 2 (b), the pairing of haploid CN structures represents the diploid configuration in a sample. For instance, a sample with one haplotype of CN2_B1-A1 and another of CN3_B1-A2 is represented as [CN2_B1-A1/CN3_B1-A2]. This notation explicitly differentiates between the two haplotypes, which is important for analyses such as Mendelian inheritance modeling in trio datasets.

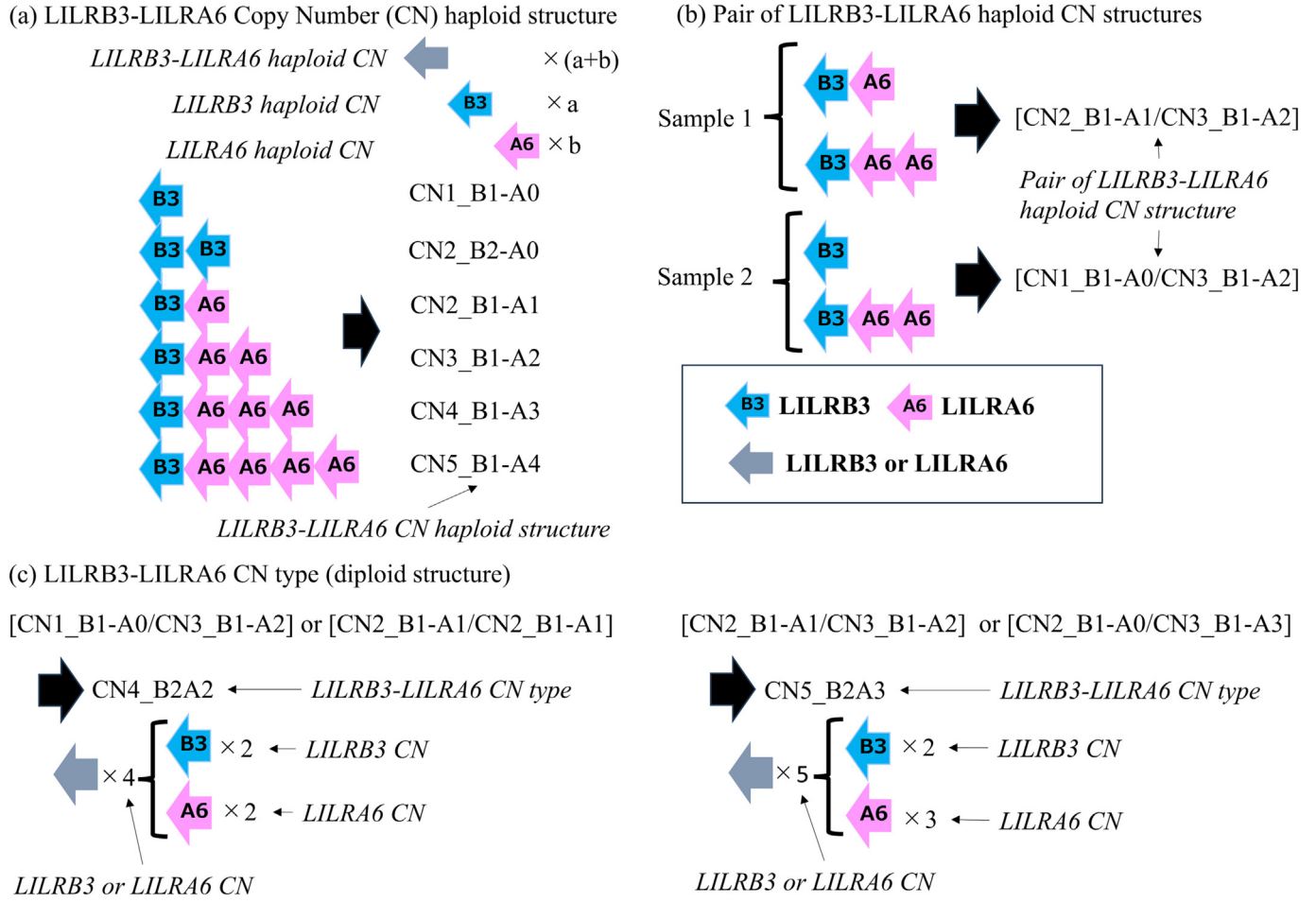

Fig. 2. Notations related to JoGo-LILR Copy Number (CN) Caller. (a) LILRB3-LILRA6 CN haploid structure: The haploid structure illustrates the copy numbers (CNs) of *LILRB3* (blue arrow) and *LILRA6* (pink arrow). The CNs of *LILRB3* and *LILRA6* in the same haplotype are concatenated with a hyphen (e.g., CN2_B1-A1 represents one copy of *LILRB3* and one copy of *LILRA6*). Additionally, the figure distinguishes between the following: *LILRB3* or *LILRA6* haplotype CN, which is the total number of copies for either *LILRB3* or *LILRA6* in a haplotype (e.g., CN2 means the total number of copies for either *LILRB3* or *LILRA6* in a haplotype is two); *LILRB3* haplotype CN, which indicates the number of *LILRB3* copies in a haplotype (e.g., one copy of *LILRB3* is indicated by B1 in CN2_B1-A1); and *LILRA6* haplotype CN, which indicates the number of *LILRA6* copies in a haplotype (e.g., one copy of *LILRA6* is indicated by A1 in CN2_B1-A1). Each row depicts an example of a haplotype structure corresponding to its CN type (e.g., CN1_B1-A0 to CN5_B1-A4). (b) Pair of LILRB3-LILRA6 CN haploid structures: Depicts pairs of haploid structures for two samples. The concatenated notations (e.g., [CN2_B1-A1/CN3_B1-A2]) represent the diploid configuration of the LILRB3-LILRA6 CN type in a given sample. (c) LILRB3-LILRA6 CN type (diploid structure): Summarizes the diploid structure by combining the total copy numbers of *LILRB3* and *LILRA6* across haplotypes. For example, CN5_B2A3 indicates a total of five copies (*LILRB3* or *LILRA6* CN), with two *LILRB3* CNs and three *LILRA6* CNs, aggregated from the haploid structures. In some cases, multiple haplotype combinations result in the same CN type. For instance, [CN1_B1-A0/CN3_B1-A2] and [CN2_B1-A1/CN2_B1-A1] both correspond to the same diploid CN type CN4_B2A2.

As depicted in Fig. 2 (c), the concept of diploid CN type is introduced, aggregating the total copy number pattern of *LILRB3* and *LILRA6* without distinguishing between individual haplotypes. For example, a total of two copies of *LILRB3* and three copies of *LILRA6* across both haplotypes is denoted as CN5_B2A3. In this representation, the total CN is five, with two copies derived from *LILRB3* and three copies from *LILRA6*. This simplified notation is particularly useful for annotating and classifying of cluster plots, allowing a structured representation of diploid CN types.

These definitions provide a practical framework for analyzing the high copy number diversity of *LILRB3* and *LILRA6*. By consistently representing haploid and diploid CN structures, this framework supports downstream analyses, including population-based haplotype frequency estimation and Mendelian inheritance modeling.

3.3. Overview of the JoGo-LILR CN Caller

The JoGo-LILR CN Caller is an algorithmic tool developed to determine LILRB3-LILRA6 CN types from whole-genome sequencing data. It follows a three-step process (Fig. 3) that integrates individual sample-level analysis (Steps 1 and 2) with population-level analysis (Step 3), ensuring both precision and consistency in CN typing.

In the first step, sequencing reads are aligned to the GRCh38 reference assembly (GRCh38DH, with decoy sequences) to generate sorted BAM or CRAM files (Step 1 in Fig. 3). This alignment standardizes the data for downstream analyses. If these files already exist, this step can be omitted for efficiency.

The second step calculates normalized read depths for predefined genomic regions using CNVNator [33] (Step 2 in Fig. 3). This step translates raw sequencing data into quantitative metrics for copy number variation analysis. These results reflect individual sample characteristics but may exhibit variability due to technical (e.g., sequencing accuracy, GC bias, mapping errors) or biological factors (e.g., actual differences in copy numbers among samples, genetic background variations).

The third step leverages data from Steps 1 and 2 to classify LILRB3-LILRA6 CN types using a population-based analysis (Step 3 in Fig. 3). Cluster plots derived from normalized read depths are used to infer CN structures. From the phased haplotype candidates, the algorithm estimates the most likely haplotype pair based on their probabilities, identifying the pair with the highest likelihood. In trio datasets (JoGo-LILR-trio), familial information refines predictions, further enhancing haplotype phasing in complex genetic datasets. Combining individual-level processing and population-level analyses improves the algorithm's robustness and ability to phase haplotypes from short-read sequencing data accurately.

3.4. Detection of core regions for LILRB3 and LILRA6

Our analysis began by aligning high-coverage PCR-free srWGS data from 2,504 unrelated samples from the 1kGP dataset [26]. These locus spanning chr19:54,216,000–54,243,000 was targeted, encompassing the complete gene bodies of *LILRB3* and *LILRA6* (Step 1 in Fig. 3).

To determine a reliable mappable region across the samples, we examined the mean mapping quality (MapQ, which reflects the confidence of read alignment to the reference genome; higher scores indicate more reliable alignment) for each chromosomal position within the extracted BAM files from the 1kGP unrelated samples. MapQ is defined as -10 log10 of the probability that the mapping position is wrong and is rounded to the nearest integer (in our analysis, 0–60, in Fig. 4). Highly similar regions, such as duplicated or paralogous sequences, typically show low mapQ scores close to 0, while uniquely mappable regions have high mapQ values, such as 60. Because the distribution of mapQ values can vary slightly among samples due to differences in sequence variation, we used the mean mapping quality to summarize these variations and assess region-specific mappability across all samples. The global distribution of the mean mapQ values, charting from the *LILRB3* to *LILRA6* regions, is illustrated in Fig. 4 (a). The analysis indicated that the upstream regions and the first halves of *LILRB3* and *LILRA6* exhibited low mapQ scores because of their sequence similarities (Blue rectangle regions in Fig. 4); notably, both genes were coded on the minus strand, with the upstream corresponding to the downstream positions on chromosome 19. More details, both genes from exon 1 to exon 7 are highly similar regions (these regions correspond to the extracellular location of these genes as depicted in Fig. 1). Conversely, the remaining downstream exons of both genes (from exon 8 to exon 13 in *LILRB3* and exon 8 in *LILRA6*) displayed significantly higher mapQ scores, i.e., high diversity between *LILRB3* and *LILRA6* (Fig. 4(a)). These regions, defined as LILRB3core and LILRA6core, enable the separation of *LILRB3* and *LILRA6* copy numbers.

Based on this mapQ analysis, we delineated the LILR core regions as chr19:54,216,000–54,219,900 for *LILRB3* (LILRB3core, orange rectangle in Fig. 4(b)) and chr19:54,236,600–54,239,600 for *LILRA6* (LILRA6core, green rectangle in Fig. 4(c)). Additionally, we defined the total coverage from *LILRB3* to *LILRA6* gene bodies (referred to as LILRB3 + LILRA6) as spanning chr19:54,216,000–54,223,100 and chr19:54,236,000–54,243,000. The LILRB3 + LILRA6 region was used to determine the combined copy number of *LILRB3* and *LILRA6*, as these regions exhibit high similarity and are not uniquely mappable. In high similarity regions, the alignment tool may assign reads to *LILRB3* or *LILRA6* at random when unique mapping is impossible. Still, by summing up the information from these regions, the combined copy numbers of *LILRB3* and *LILRA6* can be accurately determined. In contrast, the LILRB3core and LILRA6core regions, due to their higher mapQ values and unique mappability, were used to determine the relative proportions of *LILRB3* and *LILRA6* copy numbers individually. Normalized coverage was calculated for the specified regions using CNVNator version 0.4.1 [33] (Step 2 in Fig. 3). Normalization was essential, given the variable lengths of the core regions (3,900 bp for LILRA6core and 3,000 bp for LILRB3core), disparities in total sequenced coverage, and GC bias among srWGS samples. CNVNator considers these factors and fits into the downstream analysis.

3.5. LILR cluster plot and theoretical position of LILRB3 and LILRA6 gene copies

Using the core regions and gene body regions defined in Step 2, we considered the theoretical cluster (named LILR cluster plot) position of the normalized CN values (Step 3 in Fig. 3). The x-axis represents the total normalized CN value of the LILRB3 + LILRA6 region, reflecting the combined copy number of both genes. The y-axis indicates the ratio of the normalized CN values of the LILRB3core region to the LILRA6core region, which provides insight into the relative contributions of *LILRB3* and *LILRA6*. By integrating both the x-axis and y-axis metrics, the copy number structure of *LILRB3* and *LILRA6* for each sample can be determined. Step 3 in Fig. 3 illustrates the relationship between *LILRB3* and *LILRA6* CNs and their theoretical cluster positions in the LILR cluster plot. Crosses in the cluster plot represent individual measurements from multiple samples, which may include experimental noise (in Fig. 3, three times were measured from samples with CN3_B2A1, and other three times were measured from samples with CN4_B2A2)). Diamonds indicate the center positions of clusters formed by similar CN characteristics, and blue circles mark the theoretical anchor positions for each CN structure (full theoretical 63 anchor positions are listed in Supplementary Table 3). Clusters are annotated by comparing the center positions (diamonds) to the closest theoretical anchor positions (blue circles). For example, in Fig. 3, the red cluster center aligns with the anchor for CN3_B2A1, while the green cluster center aligns with the anchor for CN4_B2A2. These annotations indicate that the red cluster corresponds to CN3_B2A1, and the green cluster corresponds to CN4_B2A2.

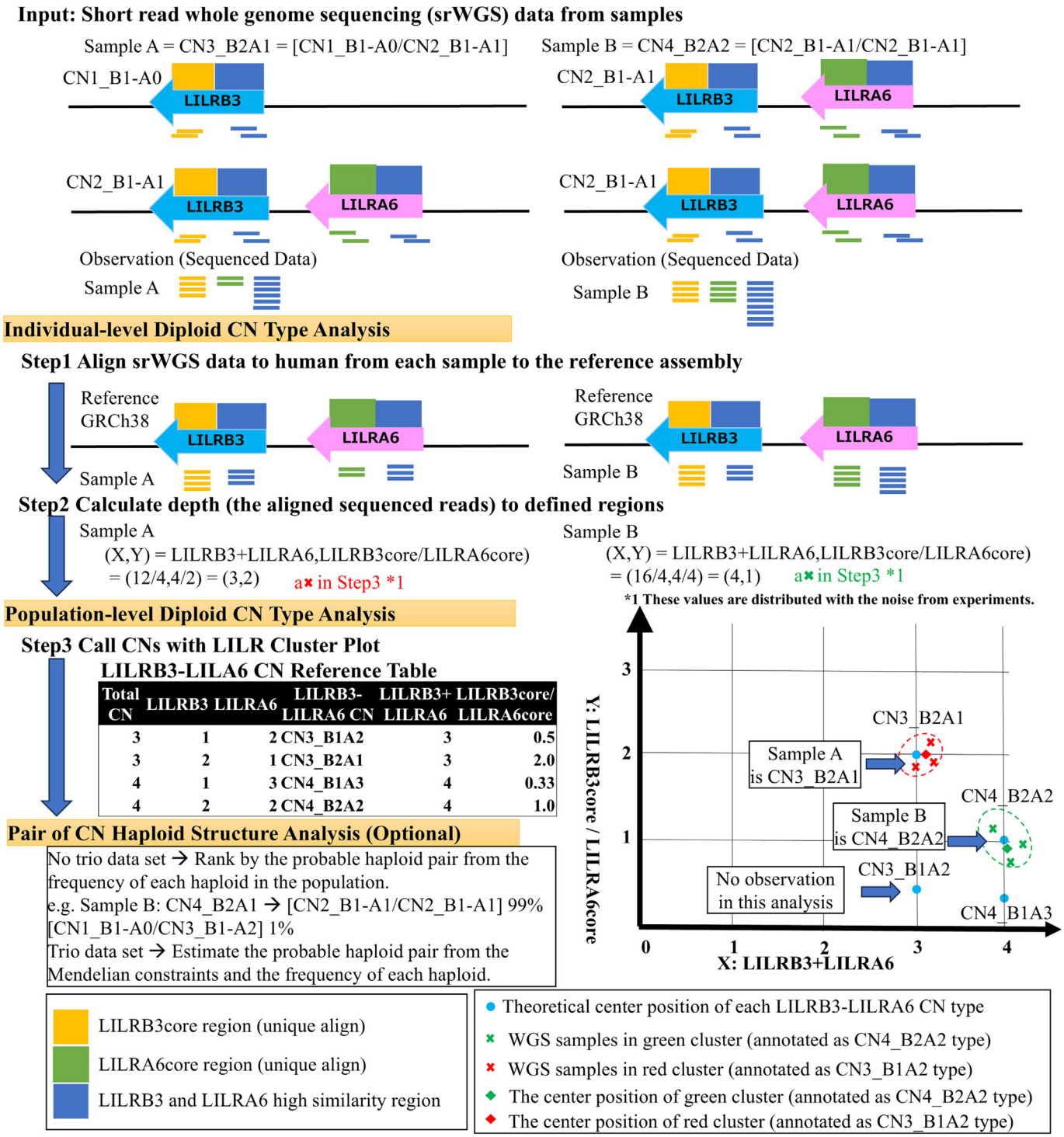

Fig. 3. Overview of JoGo-LILR CN Caller. This figure illustrates the workflow for calling LILRB3-LILRA6 copy numbers (CNs) from short-read whole-genome sequencing (srWGS) data. The input consists of three steps: alignment, normalization, and clustering. The input includes FASTQ files or aligned sequencing data for samples (e.g., Sample A and Sample B) with pair of LILRB3-LILRA6 CN haploid structures such as [CN1_B1-A0/CN2_B1-A1] and [CN2_B1-A1/CN2_B1-A1], and LILRB3-LILRA6 CN type as CN3_B2A1 and CN4_B2A2, respectively. In Step 1, sequencing reads are aligned to the human reference genome (GRCh38) to generate read-depth information for specific genomic regions. In Step 2, the depth is normalized for three distinct regions: LILRB3core (highlighted as orange), LILRA6core (green), and shared LILRB3-LILRA6 regions (blue). These normalized values are used to calculate the total copy number (LILRB3 + LILRA6) and the ratio of LILRB3core to LILRA6core. In Step 3, the normalized values are plotted (cross with red and green), with LILRB3 + LILRA6 on the x-axis and LILRB3core/LILRA6core on the y-axis to plot (right-middle plot). Clustering is applied to group samples with similar CN profiles, as shown by dotted green and red ellipses. Cluster centers (green and red diamonds) are matched to theoretical anchor positions (blue circles) from the reference table to assign CN types. For example, the green cluster corresponds to CN4_B2A2, while the red cluster corresponds to CN3_B2A1. The haplotype pair is assigned with the highest likelihood based on population frequencies. For instance, Sample B is assigned the haploid pair CN2_B1A1/CN2_B1A1 with 99 % probability. For trio data, the most probable haplotype pair is estimated based on Mendelian constraints and the frequency of each haploid.

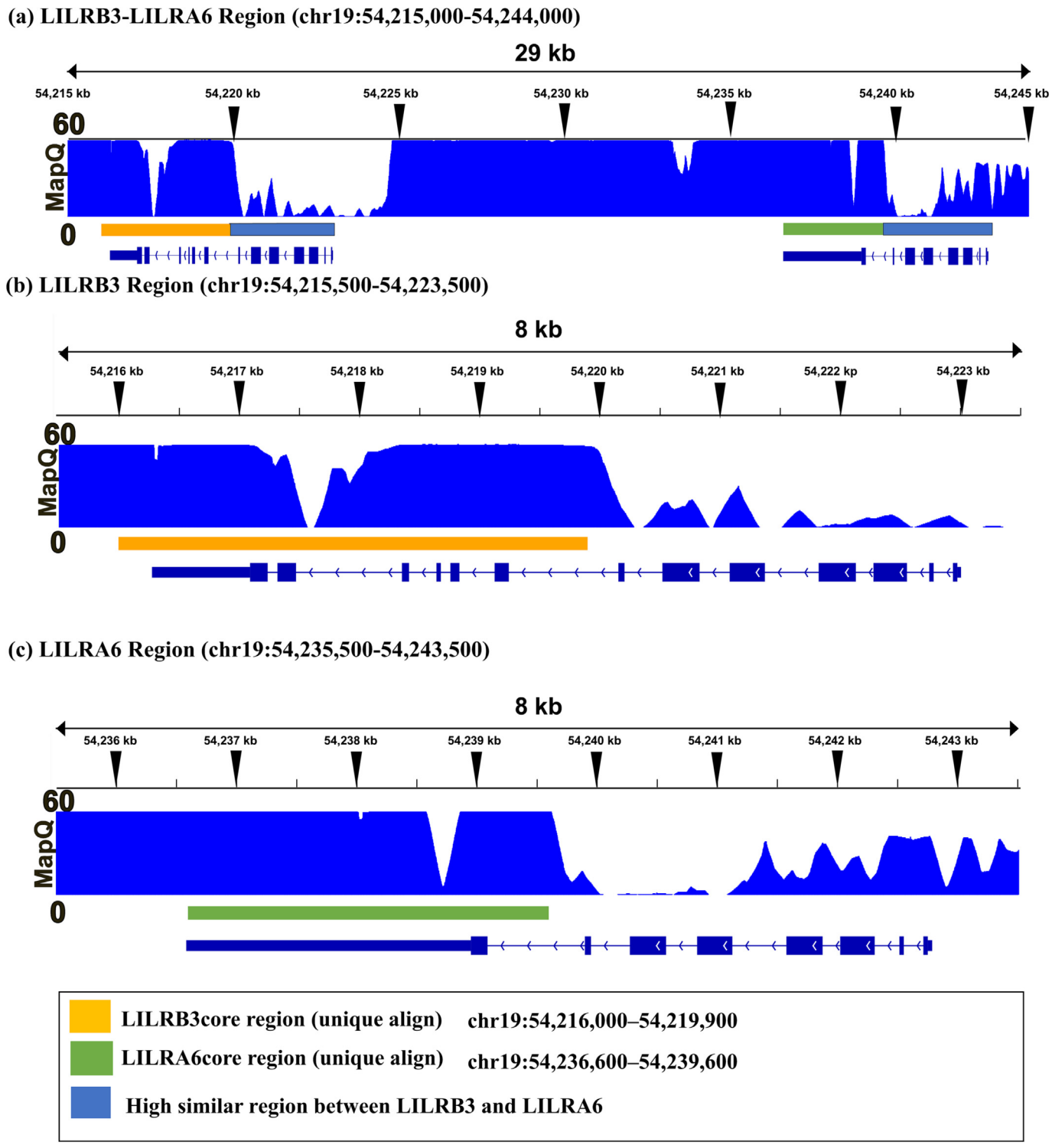

Fig. 4. The global and local mapQ distributions around LILRB3 and LILRA6. (a) The global mean mapQ distribution (from 0 (with high similarity to other regions) to 60 (unique region in GRCh38; y-axis) is shown for the 29 kb region spanning from *LILRB3* to *LILRA6* (chr19:54,215,000–54,244,000) (x-axis). Both *LILRB3* and *LILRA6* are coded on the minus strand. The regions are color-coded as green for LILRA6core, orange for LILRB3core, and blue for the high similarity regions between LILRB3 and LILRA6. The green and orange rectangles represent uniquely mappable regions for LILRA6core and LILRB3core, respectively, while the blue rectangles indicate regions with high sequence similarity between LILRB3 and LILRA6, leading to lower mapQ scores. (b) The stable mapQ region for LILRB3 (chr19:54,215,500–54,223,500) is shown in detail. This 8 kb region is identified as LILRB3core, highlighted in orange. The LILRB3core region consistently demonstrates higher mapQ scores than the flanking blue rectangle regions, which correspond to the high similarity regions shown in (a). Notably, the high similarity regions include LILRB3 exons 1–7, contributing to the lower mapQ scores in these regions. (c) The stable mapQ region for LILRA6 (chr19:54,235,500–54,243,500) is shown in detail. This 8 kb region is identified as LILRA6core, highlighted in green. The LILRA6core region also exhibits consistently higher mapQ scores than the flanking blue rectangle regions, which correspond to the high similarity regions shown in (a). Notably, the high similarity regions include LILRA6 exons 1–7, leading to lower mapQ scores in these regions.

By leveraging this clustering approach, Step 3 mitigates the impact of experimental noise and stabilizes CN structure determination through joint-calling. This statistical property ensures that even if individual samples exhibit variability, the center positions of clusters remain stable as sample size increases, providing robust annotations of CN structures.

3.6. LILR cluster plot to estimate CNs by the 1kGP samples

To elucidate the patterns of the international srWGS dataset, an LILR cluster plot was created using the 1kGP unrelated samples (Fig. 5(a), Supplementary Table 4). The analysis revealed nine patterns corresponding to LILRB3 + LILRA6 and LILRA6core/LILRB3core combinations. These patterns were anchored at specific options: (infinity (capped at 4 in our analysis), 2), (1.0, 2), (2.0, 3), (1.0, 4), (0.67, 5), (0.5, 6), (0.4, 7), (0.33, 8), and (0.286, 9), each representing unique LILRB-LILRA6 CN types: CN2_B2A0, CN2_B1A1, CN3_B2A1, CN4_B2A2, CN5_B2A3, CN6_B2A4, CN7_B2A5, CN8_B2A6, and CN9_B2A7, respectively. We further clarify the assignment from a cluster to a specific LILRB-LILRA6 CN type (Supplementary Fig. 1).

To achieve this classification, single linkage hierarchical clustering was employed in JoGo-LILR CN Caller. This method was selected based on its suitability for handling non-spherical cluster shapes, as observed in the LILRB3-LILRA6 CN dataset. Additionally, by setting the number of clusters higher than the theoretical CN patterns, the influence of outliers was mitigated by assigning them to separate clusters (e.g., Cluster ID 20 in Supplementary Fig. 1). This approach minimized the impact of outliers on the primary clusters, ensuring higher purity and improving the accuracy of CN type assignments.

Notably, the y-axis metric (LILRB3core/LILRA6core ratio) was critical for differentiating earlier clusters, such as CN2_B1A1 and CN2_B2A0, by capturing variations in the proportional relationship between *LILRB3* and *LILRA6*. In contrast, the x-axis metric (LILRB3 + LILRA6) played a primary role in distinguishing clusters for CN4_B2A2 and subsequent CN patterns, where the total copy number becomes increasingly distinct across samples. While the x-axis was dominant in later clusters, the y-axis still contributed additional resolution by reflecting subtle proportional differences between *LILRB3* and *LILRA6*, ensuring robust and accurate cluster assignments.

For comparison, we evaluated alternative clustering methods, including K-means and DBSCAN (Supplementary Figs. 2 and 3). In the K-means result, two key issues were observed. First, the CN9_B2A7 cluster, which was distinctly identified by the single linkage, was merged into the CN8_B2A6 cluster (Supplementary Fig. 2). Second, K-means divided groups with the same CN into multiple unintended clusters when the cluster size was set to the same value as the single linkage. This behavior was particularly evident for CN4_B2A2, CN5_B2A3, and CN6_B2A4 (Supplementary Fig. 4). While it is common for a single CN type to be divided into multiple subclusters (as observed in K-means, DBSCAN, and single linkage), K-means uniquely tended to assign samples more evenly across subclusters. This feature often led to subclusters being biased in their (x, y) positions relative to the reference anchor, with some subclusters appearing to the left of the anchor and others to the right. As a result, K-means was less effective in utilizing the stability provided by averaging the positions of all samples within a subcluster when compared to the reference anchor. DBSCAN, a density-based clustering method, was also tested. In DBSCAN, the eps parameter defines the radius within which neighboring points are considered part of the same cluster, and a smaller eps value generally leads to a greater number of clusters. In this study, we fixed the minimum number of samples required to form a cluster at 1 to ensure small clusters could still be detected. While DBSCAN produced results comparable to single linkage when eps was set to a small value (e.g., eps = 0.1), increasing eps (e.g., to eps = 0.5 or

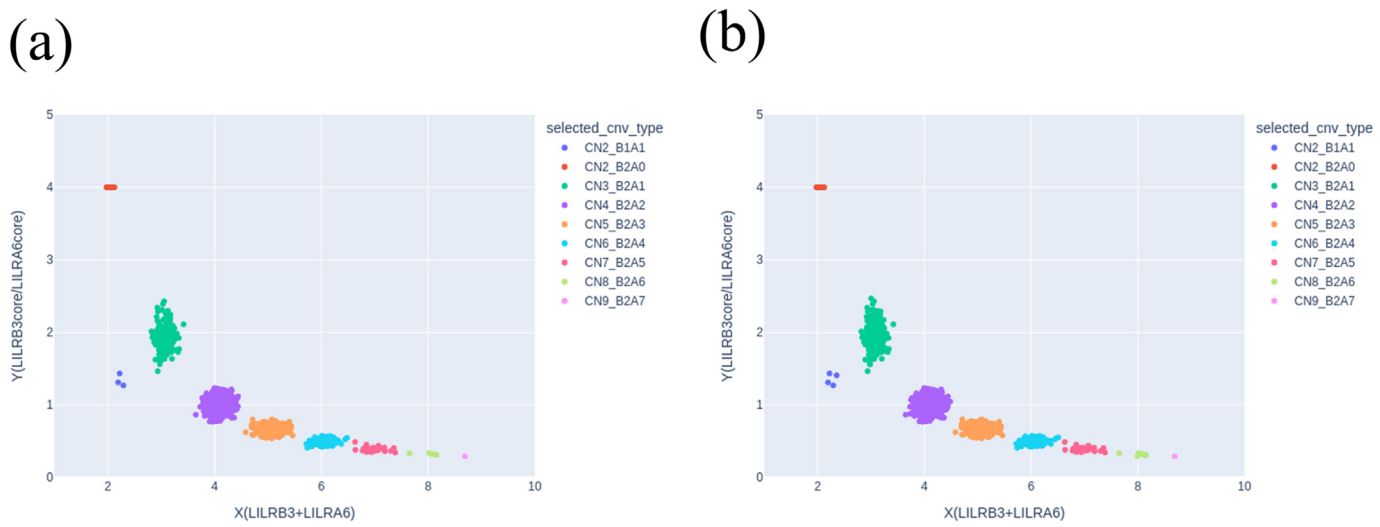

Fig. 5. The LILRB3-LILRA6 CN cluster plots. (a) The LILR cluster plot shows the calculated position (LILRB3 + LILRA6, LILRB3core/LILRA6core) of 2,504 samples. The x-axis represents the total normalized CN value of the LILRB3 + LILRA6 region, reflecting the combined copy number of both genes. The y-axis indicates the ratio of the normalized CN values of the LILRB3core region to the LILRA6core region, providing insight into the relative contributions of *LILRB3* and *LILRA6*. By integrating both the x-axis and y-axis metrics, the copy number structure of *LILRB3* and *LILRA6* for each sample can be determined. Each point in the plot represents the calculated results for a single sample based on the procedures described in Step 1 and Step 2 of Fig. 3. Samples with the same CN structure form clusters but do not land on identical points due to sequencing biases and sequence-based haplotype differences, which introduce slight deviations in normalized CN values. These deviations explain the observed spread within clusters. Notably, the y-axis metric effectively separates earlier clusters (e.g., CN2_B1A1 vs. CN2_B2A0) by capturing variations in the proportional relationship between *LILRB3* and *LILRA6*. In contrast, the x-axis metric plays a primary role in distinguishing clusters in later patterns (e.g., CN4_B2A2 onwards) by reflecting the total copy number differences. Although the y-axis metric has a smaller influence in later clusters, it still provides additional resolution by capturing subtle proportional differences between *LILRB3* and *LILRA6*. The clusters corresponding to each CN type are further detailed in Supplementary Fig. 1, illustrating their composition and characteristics. (b) The LILR cluster plot of 3,202 samples includes the 2,504 samples in (a) and their 698 related individuals. This plot expands the dataset to include additional family members, which enables further exploration of inheritance patterns and familial CN relationships. The clusters corresponding to each CN type are further detailed in Supplementary Fig. 7.

eps = 1.0) led to the merging of multiple CN clusters into a single group (Supplementary Fig. 3 and Supplementary Fig. 5). This result demonstrates the sensitivity of DBSCAN to eps values, which can significantly affect the clustering outcome. Considering the need for parameter simplicity and stability, single linkage was considered a more suitable approach for this study.

Fig. 6 (a) shows the frequency of the identified CN types for *LILRB3* and *LILRA6* in five continental groups: East Asian (EAS), South Asian (SAS), European (EUR), American (AMR), and African (AFR) (Supplementary Table 5 (a) and (b)). Our comprehensive analysis indicated that CN4 was the most prevalent CN globally, accounting for 59.2 % of cases, followed by CN5 at 26.7 %. Notably, the EAS population exhibited a higher frequency of CN4 (82.5 %) than the other groups. In contrast, the AFR population demonstrated a greater propensity for higher CN states, with instances of more than five copies constituting 13.5 % of the population, compared to a global average of 6.9 %.

3.7. Performance assessment of JoGo-LILR CN calls using high-quality diploid human pangenome reference

To evaluate the performance of the JoGo-LILR CN calls in Step 3, we utilized 40 HapMap samples with high-quality diploid human pangenome reference from HPRC [29]. The human pangenome reference used for our new validation was constructed using high-fidelity long-read sequencing (PacBio HiFi), ultra-long-read sequencing (Oxford Nanopore Technologies, ONT), Bionano Optical Maps, and Hi-C sequencing. These technologies collectively provide comprehensive genomic coverage and high resolution for structural variants, including CNVs. These assemblies, with a QV of 52 (indicating less than one error per 200,000 bases), are widely recognized as a gold standard for genome analysis [29].

The pairs of LILRB3-LILRA6 CN haplotype structures for 40 samples are detailed in Table 1 (The integrative Genome Viewer (IGV) [32] snapshots are in Supplementary Fig. 6). For these 40 samples, we compared the LILRB3-LILRA6 reference structure from HPRC and the LILRB3-LILRA6 CN types deduced from the LILR cluster plot in Step 3. This analysis yielded a perfect match for all 40 samples, confirming the accuracy of estimating LILRB3-LILRA6 CN types from the 1kGP srWGS data, as documented in Table 1.

The observed concordance rate between short-read and pangenome reference data was 100 %. Using the observed concordance rate and the expected concordance rate derived from the global population frequency distribution, the probability of achieving the observed concordance rate under the global distribution was calculated as 1.54×10^{-26}, the effect size was calculated as 3.142 (Cohen's h). This large effect size demonstrates a significant deviation from the expected concordance rate, confirming the high accuracy of JoGo-LILR Caller. Based on this effect size and the sample of 40 specimens, the statistical power was 1.000.

3.8. LILRB3-LILRA6 CN haplotype structures in diverse populations

Using the LILR cluster plot, the total number of possible patterns of LILRB3-LILRA6 CNs per sample was nine (Supplementary Table 5 (a) and (b)). This observation was further corroborated through a comparative analysis with the lrWGS dataset, which validated the accuracy of the LILRB3-LILRA6 CN determinations. By employing the frequencies of the LILRB3-LILRA6 CN types, where the sum of the two alleles was considered, we deduced the frequencies of the LILRB3-LILRA6 CN haplotype structures. This estimate was achieved by solving optimal equations for the frequencies of LILRB3-LILRA6 CNs globally or in five continental groups: EAS, SAS, EUR, AMR, and AFR (Fig. 6(b); Supplementary Table 5(c)). According to the global profile, the CN2_B1-A1 variant emerged as the most prevalent, accounting for 76.0 % of the observed variants, followed by CN2_B1-A2 at 4.4 %. In the EAS population, CN1_B1-A0 was more frequent (4.0 %) than in the other populations, except for AFR (12.2 %). Interestingly, our analysis uniquely identified the CN0_B0-A0 type within the EAS cohort at a frequency of 0.19 % (Supplementary Table 5(c)). The CN0_B0-A0 detected by our JoGo-LILR tool was subsequently experimentally validated, and its functional implications were also delineated [34].

In JoGo-LILR, the haplotype option allows for the probability of a pair of LILRB3-LILRA6 CN haplotype structures.

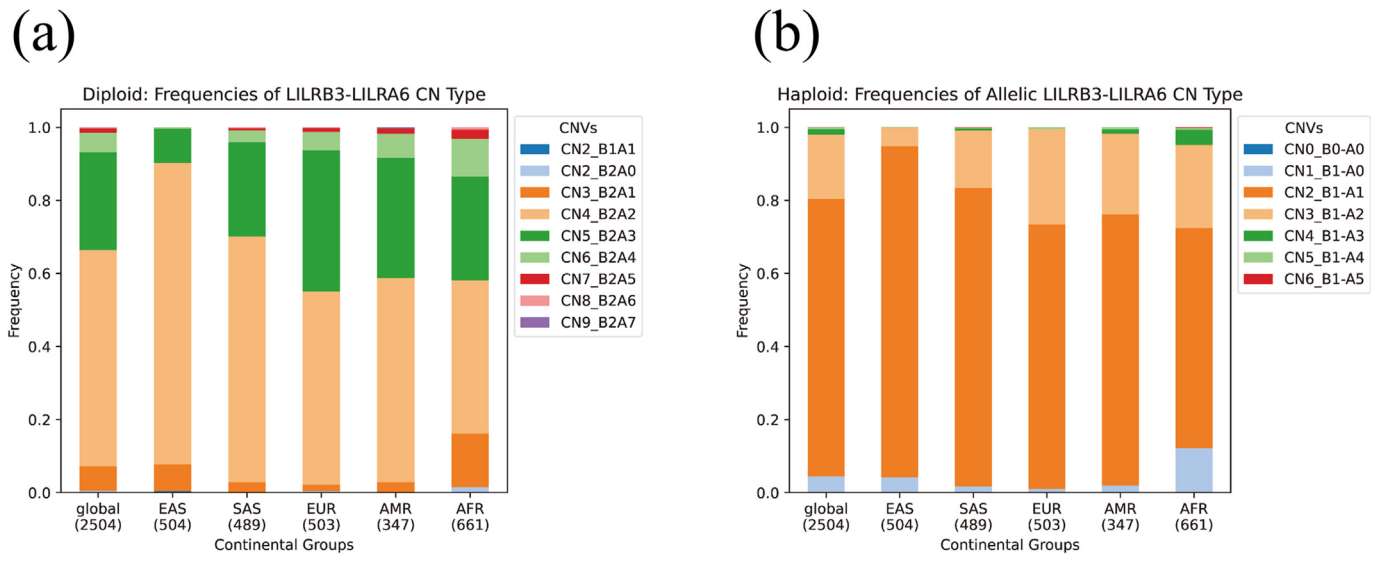

Fig. 6. Distribution of LILRB3-LILRA6 CNs in Global Populations. (a) The called frequencies of the LILRB3-LILRA6 CN types across the global population (2,504 individuals) and five continental groups: East Asia (EAS; 504 individuals), South Asia (SAS; 489 individuals), Europe (EUR; 503 individuals), America (AMR; 347 individuals), and Africa (AFR; 661 individuals). The x-axis represents the continental groups, and the y-axis shows the normalized frequencies of each CN type (e.g., CN2_B1A1, CN3_B1A2).(b) The estimated frequencies of the LILRB3-LILRA6 CN haploid structures across the global population and five continental groups. Each bar represents the haploid structure frequencies for the corresponding group.

Table 1
Comparison of LILRB3-LILRA6 copy number (CN) types determined using the JoGo-LILR Caller based on short-read whole-genome sequencing (srWGS) and the diploid reference assembly from HPRC for the shared same samples. This table summarizes results from 40 samples, including the population and continental group classification, the total copy number of LILRB3 and LILRA6 genes (X), the ratio of core regions (Y), and concordance status between the srWGS and pangenome reference assembly results. The CN haplotype structures for both alleles from diploid reference.

Sample ID	Population	Continental Group	X (LILRB3 + LILRA6 Copy Number)	Y (LILRB3core/ LILRA6core Ratio)	By JoGo-LILR Caller CN type	Concordance (JoGo-LILR vs. Diploid Reference Assembly)	HPRC validation version	LILRB3-LILRA6 CN type by HPRC	By Diploid Reference Assembly in HPRC	
									1st LILRB3-LILRA6 CN haplotype structure	2nd LILRB3-LILRA6 CN haplotype structure
HG00438	CHS	EAS	3.0346	2.0472	CN3_B2A1	YES	v1	CN3_B2A1	CN1_B1-A0	CN2_B1-A1
HG00621	CHS	EAS	3.2692	1.7841	CN3_B2A1	YES	v1	CN3_B2A1	CN1_B1-A0	CN2_B1-A1
HG00673	CHS	EAS	4.1803	0.8772	CN4_B2A2	YES	v1	CN4_B2A2	CN2_B1-A1	CN2_B1-A1
HG00733	PUR	AMR	4.1995	0.9578	CN4_B2A2	YES	v1	CN4_B2A2	CN2_B1-A1	CN2_B1-A1
HG00735	PUR	AMR	4.2288	1.0067	CN4_B2A2	YES	v1	CN4_B2A2	CN2_B1-A1	CN2_B1-A1
HG00741	PUR	AMR	5.0314	0.6312	CN5_B2A3	YES	v1	CN5_B2A3	CN2_B1-A1	CN3_B1-A2
HG01071	PUR	AMR	5.8592	0.4906	CN6_B2A4	YES	v1	CN6_B2A4	CN3_B1-A2	CN3_B1-A2
HG01106	PUR	AMR	4.2343	1.0796	CN4_B2A2	YES	v1	CN4_B2A2	CN2_B1-A1	CN2_B1-A1
HG01109	PUR	AMR	3.9571	0.9288	CN4_B2A2	YES	v1	CN4_B2A2	CN2_B1-A1	CN2_B1-A1
HG01175	PUR	AMR	4.0591	0.9904	CN4_B2A2	YES	v1	CN4_B2A2	CN2_B1-A1	CN2_B1-A1
HG01243	PUR	AMR	2.923	2.0696	CN3_B2A1	YES	v1	CN3_B2A1	CN2_B1-A1	CN1_B1-A0
HG01258	CLM	AMR	8.0174	0.3264	CN8_B2A6	YES	v1	CN8_B2A6	CN5_B1-A4	CN3_B1-A2
HG01358	CLM	AMR	4.186	0.8926	CN4_B2A2	YES	v1	CN4_B2A2	CN2_B1-A1	CN2_B1-A1
HG01361	CLM	AMR	3.9737	1.0852	CN4_B2A2	YES	v1	CN4_B2A2	CN2_B1-A1	CN2_B1-A1
HG01891	ACB	AFR	6.1392	0.4765	CN6_B2A4	YES	v1	CN6_B2A4	CN3_B1-A2	CN3_B1-A2
HG01928	PEL	AMR	4.1236	0.8732	CN4_B2A2	YES	v1	CN4_B2A2	CN2_B1-A1	CN2_B1-A1
HG01952	PEL	AMR	4.8906	0.7058	CN5_B2A3	YES	v1.0.1(r2)	CN5_B2A3	CN3_B1-A2	CN2_B1-A1
HG01978	PEL	AMR	5.2689	0.6476	CN5_B2A3	YES	v1	CN5_B2A3	CN3_B1-A2	CN2_B1-A1
HG02055	ACB	AFR	4.9656	0.6475	CN5_B2A3	YES	v1	CN5_B2A3	CN2_B1-A1	CN3_B1-A2
HG02080	KHV	EAS	4.1604	0.9357	CN4_B2A2	YES	v1	CN4_B2A2	CN2_B1-A1	CN2_B1-A1
HG02145	ACB	AFR	5.0924	0.7207	CN5_B2A3	YES	v1.0.1(r2)	CN5_B2A3	CN1_B1-A0	CN4_B1-A3
HG02148	PEL	AMR	2.9282	1.8576	CN3_B2A1	YES	v1	CN3_B2A1	CN1_B1-A0	CN2_B1-A1
HG02257	ACB	AFR	3.1123	1.7829	CN3_B2A1	YES	v1	CN3_B2A1	CN1_B1-A0	CN2_B1-A1
HG02572	GWD	AFR	4.1399	0.9647	CN4_B2A2	YES	v1	CN4_B2A2	CN2_B1-A1	CN2_B1-A1
HG02622	GWD	AFR	4.9871	0.5998	CN5_B2A3	YES	v1	CN5_B2A3	CN2_B1-A1	CN3_B1-A2
HG02630	GWD	AFR	4.1385	0.8934	CN4_B2A2	YES	v1	CN4_B2A2	CN1_B1-A0	CN3_B1-A2
HG02717	GWD	AFR	2.9996	2.2262	CN3_B2A1	YES	v1	CN3_B2A1	CN2_B1-A1	CN1_B1-A0
HG02723	GWD	AFR	3.737	0.8766	CN4_B2A2	YES	v1	CN4_B2A2	CN2_B1-A1	CN2_B1-A1
HG02818	GWD	AFR	6.1893	0.5448	CN6_B2A4	YES	v1.0.1(r2)	CN6_B2A4	CN1_B1-A0	CN5_B1-A4
HG02886	GWD	AFR	5.0884	0.6826	CN5_B2A3	YES	v1.0.1(r2)	CN5_B2A3	CN2_B1-A1	CN3_B1-A2
HG03098	MSL	AFR	4.1686	1.0414	CN4_B2A2	YES	v1	CN4_B2A2	CN2_B1-A1	CN2_B1-A1
HG03453	MSL	AFR	4.1121	0.9352	CN4_B2A2	YES	v1	CN4_B2A2	CN2_B1-A1	CN2_B1-A1
HG03486	MSL	AFR	8.1359	0.3022	CN8_B2A6	YES	v1	CN8_B2A6	CN4_B1-A3	CN4_B1-A3
HG03492	PJL	SAS	4.8823	0.6745	CN5_B2A3	YES	v1	CN5_B2A3	CN2_B1-A1	CN3_B1-A2
HG03516	ESN	AFR	4.0381	0.9679	CN4_B2A2	YES	v1	CN4_B2A2	CN2_B1-A1	CN2_B1-A1
HG03540	GWD	AFR	5.0642	0.6576	CN5_B2A3	YES	v1	CN5_B2A3	CN2_B1-A1	CN3_B1-A2
HG03579	MSL	AFR	4.1252	0.9578	CN4_B2A2	YES	v1	CN4_B2A2	CN2_B1-A1	CN2_B1-A1
NA18906	YRI	AFR	4.1369	0.9625	CN4_B2A2	YES	v1	CN4_B2A2	CN2_B1-A1	CN2_B1-A1
NA19240	YRI	AFR	3.9916	0.9999	CN4_B2A2	YES	v1	CN4_B2A2	CN2_B1-A1	CN2_B1-A1
NA20129	ASW	AFR	4.3382	1.0711	CN4_B2A2	YES	v1	CN4_B2A2	CN2_B1-A1	CN2_B1-A1

3.9. JoGo-LILR-trio and the performance evaluation using pedigree srWGS dataset

In assessing LILRB3-LILRA6 CN types for parental genotypes, the JoGo-LILR-trio algorithm probabilistically determines the most likely pairs of LILRB3-LILRA6 haploid CN structures for parents and their child. The algorithm first identifies all possible haploid CN pairings for the parents that comply with Mendelian inheritance patterns observed in the offspring. It then integrates population-specific haplotype frequency data to prioritize the most probable combination as the primary candidate for the child's pair of haploid CN structures.

When multiple parental CN-type pairings are feasible, JoGo-LILR-trio evaluates the likelihood of each pairing and outputs the combination with the highest probability along with its confidence score. This approach ensures robust predictions of haploid CN structures for the trio, aligning with both Mendelian principles and population frequency data.

To evaluate the performance of the called results of LILRB3-LILRA6 CN types and the estimated CN haplotype structures for the children, we used 602 completed trio data by integrating the 698 HapMap pedigree datasets [26] with the former 2,504 non-pedigree datasets. JoGo-LILR identified LILRB3-LILRA6 CN types through joint-calling on 3,202 samples (the LILR cluster plot is shown in Fig. 6(b); the LILRB3-LILRA6 CN types are summarized in Supplementary Table 6).

The JoGo-LILR-trio further incorporates trio structures and evaluates the probability of pairs of LILRB3-LILRA6 haplotype structures. For example, the LILRB3-LILRA6 CN type CN4_B2A2 of EUR can correspond to two possible pairs of LILRB3-LILRA6 CN haplotype structures: (CN2_B1-A1, CN2_B1-A1) or (CN1_B1-A0, CN1_B1-A2), with probabilities of 0.995 and 0.005, respectively. The JoGo-LILR-trio tool was applied to the LILRB3-LILRA6 results generated by JoGo-LILR (Supplementary Table 7).

Remarkably, all 602 trios adhered to Mendelian inheritance rules. Among these, 549 child samples (91.2 %) were accurately estimated to have the most probable pair of LILRB3-LILRA6 haploid CN structure with probabilities exceeding 99 %, and 583 child samples (96.8 %) with probabilities exceeding 90 % (Supplementary Table 8). These probabilities reflect the likelihood of the most probable haplotype pair

being the correct one among all possible pairings, rather than predictive accuracy. For instance, sample HG01258 was determined to have the most probable CN haplotype pair CN3_B1-A2 and CN5_B1-A4 with a probability of 99.99 %. In comparison, sample HG02055 was identified with the most probable CN haplotype pair CN2_B1-A1 and CN3_B1-A2, with a probability of 96.4 %.

The average likelihood of the most probable haplotype pair across all 602 samples was 0.9901, with a median likelihood of 1.0000. Taken together, these results strongly support the validity and reliability of identifying the top candidate haplotype pair with high confidence (Supplementary Table 8).

To assess the accuracy of our method for estimating the most probable pair of LILRB3-LILRA6 CN haplotype structures in a child, we analyzed 40 child samples sourced from the HapMap trio dataset, validated against diploid reference assemblies provided by the HPRC [29]. Table 2 summarizes the comparison between the JoGo-LILR-trio estimated haplotype pairs and the HPRC-derived true haplotype pairs, as well as the concordance result (Supplementary Fig. 7; Supplementary Table 8). The analysis demonstrated 100 % concordance between JoGo-LILR-trio predictions and HPRC-derived diploid haplotype pairs for all 40 child samples, affirming the tool's ability to identify true CN haplotype pairs accurately.

Due to the limited number of diploid reference assemblies, we could not fully evaluate the estimated performance. However, these results provide robust evidence of the reliability and accuracy of JoGo-LILR-trio in estimating LILRB3-LILRA6 CN haplotype structures when applied to trio datasets.

3.10. Software

The software inputs the fastq or aligned files, BAM or CRAM format, and calls the LILRB3-LILRA6 CN types and LILRB3-LILRA6 CN haplotype structures. The prior option allows for joint CN calling with 1kGP samples. If the population type of each sample is specified, the population-specific CN frequency is used a priori (EAS, SAS, EUR, AMR, and AFR; the default is global). The JoGo-LILR-trio takes the JoGo-LILR and Plink ped format [35] (family structure format) and outputs the most probable pair of LILRB3-LILRA6 CN haplotype struc-

Table 2

Comparison of LILRB3-LILRA6 haploid CN structures inferred by JoGo-LILR-trio and those determined from diploid reference assemblies in HPRC. This table summarizes the comparison of LILRB3-LILRA6 haploid CN structures inferred for 40 trio families using the JoGo-LILR-trio Caller based on srWGS and those assigned as paternal or maternal in the HPRC diploid reference assemblies. The most probable pair of haploid CN structures for each child was inferred by JoGo-LILR-trio and compared to the haploid CN structures annotated in the HPRC reference. The table includes the population and continental group classification, sample ID, inferred haploid CN structures for both parents, and the probability of the estimated pair. Concordance between the JoGo-LILR-trio inference and the HPRC diploid reference data is also indicated.

Sample ID	Population	Continental Group	Trio	Concordance (JoGo-LILR vs. Diploid Reference Assembly)	By JoGo-LILR-trio		By Diploid Reference Assembly in HPRC
					Pair of LILRB3-LILRA6 haploid CN structure inherit from (mother/father)	Probability	Pair of LILRB3-LILRA6 haploid CN structure inherit from (mother/father)
HG00438	CHS	EAS	Child	YES	CN1_B1-A0/CN2_B1-A1	99.8 %	CN1_B1-A0/CN2_B1-A1
HG00621	CHS	EAS	Child	YES	CN1_B1-A0/CN2_B1-A1	99.8 %	CN1_B1-A0/CN2_B1-A1
HG00673	CHS	EAS	Child	YES	CN2_B1-A1/CN2_B1-A1	100.0 %	CN2_B1-A1/CN2_B1-A1
HG00733	PUR	AMR	Child	YES	CN2_B1-A1/CN2_B1-A1	100.0 %	CN2_B1-A1/CN2_B1-A1
HG00735	PUR	AMR	Child	YES	CN2_B1-A1/CN2_B1-A1	100.0 %	CN2_B1-A1/CN2_B1-A1
HG00741	PUR	AMR	Child	YES	CN2_B1-A1/CN3_B1-A2	100.0 %	CN2_B1-A1/CN3_B1-A2
HG01071	PUR	AMR	Child	YES	CN3_B1-A2/CN3_B1-A2	96.9 %	CN3_B1-A2/CN3_B1-A2
HG01106	PUR	AMR	Child	YES	CN2_B1-A1/CN2_B1-A1	100.0 %	CN2_B1-A1/CN2_B1-A1
HG01109	PUR	AMR	Child	YES	CN2_B1-A1/CN2_B1-A1	100.0 %	CN2_B1-A1/CN2_B1-A1
HG01175	PUR	AMR	Child	YES	CN2_B1-A1/CN2_B1-A1	100.0 %	CN2_B1-A1/CN2_B1-A1
HG01243	PUR	AMR	Child	YES	CN2_B1-A1/CN1_B1-A0	99.9 %	CN2_B1-A1/CN1_B1-A0
HG01258	CLM	AMR	Child	YES	CN5_B1-A4/CN3_B1-A2	100.0 %	CN5_B1-A4/CN3_B1-A2
HG01358	CLM	AMR	Child	YES	CN2_B1-A1/CN2_B1-A1	100.0 %	CN2_B1-A1/CN2_B1-A1
HG01361	CLM	AMR	Child	YES	CN2_B1-A1/CN2_B1-A1	100.0 %	CN2_B1-A1/CN2_B1-A1
HG01891	ACB	AFR	Child	YES	CN3_B1-A2/CN3_B1-A2	96.7 %	CN3_B1-A2/CN3_B1-A2
HG01928	PEL	AMR	Child	YES	CN2_B1-A1/CN2_B1-A1	100.0 %	CN2_B1-A1/CN2_B1-A1
HG01952	PEL	AMR	Child	YES	CN3_B1-A2/CN2_B1-A1	99.2 %	CN3_B1-A2/CN2_B1-A1
HG01978	PEL	AMR	Child	YES	CN3_B1-A2/CN2_B1-A1	99.9 %	CN3_B1-A2/CN2_B1-A1
HG02055	ACB	AFR	Child	YES	CN2_B1-A1/CN3_B1-A2	96.4 %	CN2_B1-A1/CN3_B1-A2
HG02080	KHV	EAS	Child	YES	CN2_B1-A1/CN2_B1-A1	100.0 %	CN2_B1-A1/CN2_B1-A1
HG02145	ACB	AFR	Child	YES	CN1_B1-A0/CN4_B1-A3	100.0 %	CN1_B1-A0/CN4_B1-A3
HG02148	PEL	AMR	Child	YES	CN1_B1-A0/CN2_B1-A1	99.2 %	CN1_B1-A0/CN2_B1-A1
HG02257	ACB	AFR	Child	YES	CN1_B1-A0/CN2_B1-A1	96.5 %	CN1_B1-A0/CN2_B1-A1
HG02572	GWD	AFR	Child	YES	CN2_B1-A1/CN2_B1-A1	99.4 %	CN2_B1-A1/CN2_B1-A1
HG02622	GWD	AFR	Child	YES	CN2_B1-A1/CN3_B1-A2	92.9 %	CN2_B1-A1/CN3_B1-A2
HG02630	GWD	AFR	Child	YES	CN1_B1-A0/CN3_B1-A2	92.9 %	CN1_B1-A0/CN3_B1-A2
HG02717	GWD	AFR	Child	YES	CN2_B1-A1/CN1_B1-A0	100.0 %	CN2_B1-A1/CN1_B1-A0
HG02723	GWD	AFR	Child	YES	CN2_B1-A1/CN2_B1-A1	98.5 %	CN2_B1-A1/CN2_B1-A1
HG02818	GWD	AFR	Child	YES	CN1_B1-A0/CN5_B1-A4	98.9 %	CN1_B1-A0/CN5_B1-A4
HG02886	GWD	AFR	Child	YES	CN2_B1-A1/CN3_B1-A2	92.9 %	CN2_B1-A1/CN3_B1-A2
HG03098	MSL	AFR	Child	YES	CN2_B1-A1/CN2_B1-A1	100.0 %	CN2_B1-A1/CN2_B1-A1
HG03453	MSL	AFR	Child	YES	CN2_B1-A1/CN2_B1-A1	100.0 %	CN2_B1-A1/CN2_B1-A1
HG03486	MSL	AFR	Child	YES	CN4_B1-A3/CN4_B1-A3	91.3 %	CN4_B1-A3/CN4_B1-A3
HG03492	PJL	SAS	Child	YES	CN2_B1-A1/CN3_B1-A2	99.9 %	CN2_B1-A1/CN3_B1-A2
HG03516	ESN	AFR	Child	YES	CN2_B1-A1/CN2_B1-A1	99.4 %	CN2_B1-A1/CN2_B1-A1
HG03540	GWD	AFR	Child	YES	CN2_B1-A1/CN3_B1-A2	92.9 %	CN2_B1-A1/CN3_B1-A2
HG03579	MSL	AFR	Child	YES	CN2_B1-A1/CN2_B1-A1	99.4 %	CN2_B1-A1/CN2_B1-A1
NA18906	YRI	AFR	Child	YES	CN2_B1-A1/CN2_B1-A1	99.4 %	CN2_B1-A1/CN2_B1-A1
NA19240	YRI	AFR	Child	YES	CN2_B1-A1/CN2_B1-A1	92.9 %	CN2_B1-A1/CN2_B1-A1
NA20129	ASW	AFR	Child	YES	CN2_B1-A1/CN2_B1-A1	99.4 %	CN2_B1-A1/CN2_B1-A1

tures for parents and the child. The JoGo-LILR is available on the Joint Open Genome and Omics Portal (https://jogo.csml.org/JoGo-LILR/).

4. Discussion

JoGo-LILR can estimate the LILRB3-LILRA6 CN type and identify probable pairs of LILRB3-LILRA6 CN haplotype structures from the srWGS dataset from individual cases to large-scale cohort samples. The calling results of the general and diseased populations can be used for downstream association studies of phenotypes and diseases, such as immune diseases. Furthermore, the JoGo-LILR Caller has the potential to be applied to populations with genetic abnormalities, enabling targeted studies to explore the relationship between LILRB3-LILRA6 CN types and specific genetic conditions or disease susceptibility. The estimated LILRB3-LILRA6 CN types perfectly matched the 40 samples from diploid reference assemblies in HPRC. JoGo-LILR-trio successfully estimated the haploid CN pairs using Mendelian principles and CN haplotype frequencies, with the results perfectly aligned with diploid reference assemblies with the source of paternal and maternal.

Although the copy number structure information provided by the JoGo-LILR Caller can already facilitate association studies, integrating base-resolution level haplotype structure information alongside CN data will enable more detailed analyses in the future. The estimated pairs of LILRB3-LILRA6 CN haplotype structures allow researchers to develop a method to impute the LILRB3-LILRA6 CN haplotype structure not only from srWGS but also from genotyping SNP array technology.

Application of JoGo-LILR to worldwide populations clearly indicates that LILRB3-LILRA6 CN types exhibit less variability in East Asians, consistent with the previous finding of a diminished repertoire of activating receptors within the leukocyte receptor complex in these populations [36]. The underlying mechanism for this observation remains unclear, while ligands for activating *LILRA6* and inhibitory *LILRB3*, such as apolipoprotein (APOE) 4, angiopoietin-like proteins (ANGPTL), galectins, and cytokeratin-associated ligands [6–9]), provide a possible explanation. Variations in the ratio of *LILRA6* and *LILRB3* for specific ligands represent an adaptive evolution, allowing distinct populations to fine-tune immune responses to particular pathogens in their local environments. While this hypothesis highlights the potential role of environmental pressures, the bottleneck effect observed in East Asians cannot be excluded as a contributing factor.

We have previously shown that the 1kGP genotyping call set contains a significant number of genotyping errors that can be attributed to several factors including short-read sequencing, mapping and alignment issues inherent srWGS technology [36]. Addressing these challenges at the basepair resolution level remains a critical next step. Recent advancements in telomere-to-telomere (T2T) sequencing have demonstrated the ability to accurately determine basepair level phased haplotypes. Our study confirmed part of the LILRB3-LILRA6 CN and CN haplotype structures using diploid reference assemblies in HPRC. By increasingly cataloging these regions, JoGo-LILR can infer probable pairs of LILRB3-LILRA6 CN types and haplotypes even from srWGS data. For example, our tools, HLA-VBSeq v1 [22] and v2 [23] can estimate the most probable pair of haplotypes of HLA class I and II regions from a cataloged haplotype panel using variational Bayesian statistics, which could be adopted for similar purposes in LILRB3-LILRA6 analysis.

The JoGo-LILR caller was developed based on the srWGS data from the PCR-Free protocol. Theoretically, the PCR-Free protocol is less GC-biased than the PCR protocol. This is because the PCR process can introduce bias by preferentially amplifying certain regions of the genome based on their GC content, which can lead to underrepresentation or overrepresentation of certain regions, particularly those with high GC content. On the other hand, the PCR-Free approach avoids amplification altogether and directly sequences the DNA, thereby reducing such biases and providing more uniform coverage across the genome. For this reason, we recommend using the JoGo-LILR caller with PCR-Free protocol datasets. If the user has already obtained the srWGS data using the PCR protocol, we recommend using the cnv2ratio metric. This metric represents the estimated proportion of genes carrying two copies on autosomal chromosomes. A cnv2ratio of 0.85, for instance, means 15 % of genes are detected with zero, one, or more than two copies. Typically, the cnv2ratio exceeds 0.9 with PCR-Free data. For PCR-amplified srWGS data, it is preferable to use JoGo-LILR with samples having a cnv2ratio above 0.85.

Author contributions

Masao Nagasaki and Katsushi Tokunaga designed the study. Masao Nagasaki mainly analyzed the data and wrote the manuscript. Kouyuki Hirayasu was responsible for summarizing and sharing the scientific background knowledge of complex LILR structures. Kouyuki Hirayasu, Seik-Soon Khor, Yosuke Kawai, and Katsushi Tokunaga intensively discussed the LILRB3-LILRA6 CN results. Ryoko Otokozawa and Yayoi Sekiya assisted with the data analyses. All authors read and approved the final manuscript.

CRediT authorship contribution statement

Masao Nagasaki: Writing – review & editing, Writing – original draft, Software, Resources, Methodology, Investigation, Funding acquisition, Formal analysis, Conceptualization. **Kouyuki Hirayasu:** Writing – review & editing, Investigation, Funding acquisition, Conceptualization. **Seik-Soon Khor:** Writing – review & editing, Investigation, Conceptualization. **Ryoko Otokozawa:** Software. **Yayoi Sekiya:** Review & Software. **Yosuke Kawai:** Writing – review & editing, Supervision, Methodology, Investigation, Funding acquisition. **Katsushi Tokunaga:** Writing – review & editing, Supervision, Project administration, Investigation, Funding acquisition, Conceptualization.

Funding

Masao Nagasaki received grants from the Japan Agency for Medical Research and Development (AMED) (Grant Numbers JP20ek0109492, JP21wm0425009, JP20ek0109485, JP21ek0109548, JP23ek0109675, JP23ek0109672, JP22tm0424222, JP24gm2010001) and JST NBDC Grant Number JPMJND2302, and JSPS KAKENHI Grant Number JP21H02681. This work was partially supported by the “Joint Usage/Research Center for Interdisciplinary Large-scale Information Infrastructures” and “High Performance Computing Infrastructure” in Japan (Project ID: jh200047-NWH, jh210018-NWH, jh220014, jh230016, and jh240015). Kouyuki Hirayasu received grants from AMED (Grant Number JP23wm0225036), JSPS KAKENHI grant number JP23H02714 and 23H04774, and the Naito Foundation.

Data availability

JoGo-LILR Caller v1 is available at https://jogo.csml.org/JoGo-LILR/.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

The infrastructure of Omics Science Center Secure Information Analysis System, Medical Institute of Bioregulation at Kyushu University provides the (part of) computational resource (https://sis.bioreg.kyushu-u.ac.jp/). This work was supported in part by the MEXT Cooperative Research Project Program, Medical Research Center Initiative for High Depth Omics, and CURE:JPMXP1323015486 for MIB, Kyushu University. We sincerely appreciate the reviewers for their insightful comments and constructive suggestions, which have significantly contributed to enhancing the quality of this manuscript. We also extend our gratitude to the editor for their thoughtful handling of the review process and support throughout the revision of this manuscript.

Appendix A. Supplementary material

Supplementary data to this article can be found online at https://doi.org/10.1016/j.humimm.2025.111272.

References

[1] K. Hirayasu, H. Arase, Functional and genetic diversity of leukocyte immunoglobulin-like receptor and implication for disease associations, J. Hum. Genet. 60 (2015) 703.
[2] H.R. Katz, Inhibition of inflammatory responses by leukocyte Ig-like receptors, Adv. Immunol. 91 (2006) 251.
[3] F. Abdallah, S. Coindre, M. Gardet, F. Meurisse, A. Naji, N. Suganuma, et al, Leukocyte immunoglobulin-like receptors in regulating the immune response in infectious diseases: a window of opportunity to pathogen persistence and a sound target in therapeutics, Front. Immunol. 12 (2021) 717998.
[4] H. Cheng, F. Mohammed, G. Nam, Y. Chen, J. Qi, L.I. Garner, et al, Crystal structure of leukocyte Ig-like receptor LILRB4 (ILT3/LIR-5/CD85k): a myeloid inhibitory receptor involved in immune tolerance, J. Biol. Chem. 286 (2011) 18013.
[5] N.T. Young, E.C. Waller, R. Patel, A. Roghanian, J.M. Austyn, J. Trowsdale, The inhibitory receptor LILRB1 modulates the differentiation and regulatory potential of human dendritic cells, Blood 111 (2008) 3090.
[6] J. Zhou, Y. Wang, G. Huang, M. Yang, Y. Zhu, C. Jin, et al, LilrB3 is a putative cell surface receptor of APOE4, Cell Res. 33 (2023) 116.
[7] J. Zheng, M. Umikawa, C. Cui, J. Li, X. Chen, C. Zhang, et al, Inhibitory receptors bind ANGPTLs and support blood stem cells and leukaemia development, Nature 485 (2012) 656.
[8] D.C. Jones, C.R. Hewitt, M.R. Lopez-Alvarez, M. Jahnke, A.I. Russell, V. Radjabova, et al, Allele-specific recognition by LILRB3 and LILRA6 of a cytokeratin 8-associated ligand on necrotic glandular epithelial cells, Oncotarget 7 (2016) 15618.
[9] R. Huang, X. Liu, J. Kim, H. Deng, M. Deng, X. Gui, et al, LILRB3 supports immunosuppressive activity of myeloid cells and tumor development, Cancer Immunol. Res. 12 (2024) 350.
[10] K. Hirayasu, F. Saito, T. Suenaga, K. Shida, N. Arase, K. Oikawa, et al, Microbially cleaved immunoglobulins are sensed by the innate immune receptor LILRA2, Nat. Microbiol. 1 (2016) 16054.
[11] D. Cosman, N. Fanger, L. Borges, M. Kubin, W. Chin, L. Peterson, et al, A novel immunoglobulin superfamily receptor for cellular and viral MHC class I molecules, Immunity 7 (1997) 273.
[12] K.R. Chan, E.Z. Ong, H.C. Tan, S.L. Zhang, Q. Zhang, K.F. Tang, et al, Leukocyte immunoglobulin-like receptor B1 is critical for antibody-dependent dengue, PNAS 111 (2014) 2722.
[13] T.A. Nakada, W. Takahashi, E. Nakada, T. Shimada, J.A. Russell, K.R. Walley, Genetic polymorphisms in sepsis and cardiovascular disease: do similar risk genes suggest similar drug targets?, Chest 155 (2019) 1260
[14] H. Wang, Y. Wang, Y. Tang, H. Ye, X. Zhang, G. Zhou, et al, Frequencies of the LILRA3 6.7-kb deletion are highly differentiated among Han Chinese subpopulations and involved in ankylosing spondylitis predisposition, Front. Genet. 10 (2019) 869.
[15] P.A. Renauer, G. Saruhan-Direskeneli, P. Coit, A. Adler, K. Aksu, G. Keser, et al, Identification of susceptibility Loci in IL6, RPS9/LILRB3, and an intergenic locus on chromosome 21q22 in Takayasu arteritis in a genome-wide association study, Arthritis Rheumatol. 67 (2015) 1361.
[16] C. Terao, H. Yoshifuji, T. Matsumura, T.K. Naruse, T. Ishii, Y. Nakaoka, et al, Genetic determinants and an epistasis of LILRA3 and HLA-B*52 in Takayasu arteritis, PNAS 115 (2018) 13045.
[17] J. Xu, Z. Mo, D. Ye, M. Wang, F. Liu, G. Jin, et al, Genome-wide association study in Chinese men identifies two new prostate cancer risk loci at 9q31.2 and 19q13.4, Nat. Genet. 44 (2012) 1231.
[18] Z. Liu, R. Liu, H. Gao, S. Jung, X. Gao, R. Sun, et al, Genetic architecture of the inflammatory bowel diseases across East Asian and European ancestries, Nat. Genet. 55 (2023) 796.
[19] B.M. Reid, J.B. Permuth, Y.A. Chen, B.L. Fridley, E.S. Iversen, Z. Chen, et al, Genome-wide analysis of common copy number variation and epithelial ovarian cancer risk, Cancer Epidemiol. Biomark. Prev. 28 (2019) 1117.
[20] M.R. Lopez-Alvarez, W. Jiang, D.C. Jones, J. Jayaraman, C. Johnson, W.O. Cookson, et al, LILRA6 copy number variation correlates with susceptibility to atopic dermatitis, Immunogenetics 68 (2016) 743.
[21] X. Chen, F. Shen, N. Gonzaludo, A. Malhotra, C. Rogert, R.J. Taft, et al, Cyrius: accurate CYP2D6 genotyping using whole-genome sequencing data, Pharmacogenomics J. 21 (2021) 251.
[22] N. Nariai, K. Kojima, S. Saito, T. Mimori, Y. Sato, Y. Kawai, et al, HLA-VBSeq: accurate HLA typing at full resolution from whole-genome sequencing data, BMC Genomics 16 (Suppl. 2) (2015) S7.
[23] Y.Y. Wang, T. Mimori, S.S. Khor, O. Gervais, Y. Kawai, Y. Hitomi, et al, HLA-VBSeq v2: improved HLA calling accuracy with full-length Japanese class-I panel, Hum. Genome Var. 6 (2019) 29.
[24] S. Kawaguchi, K. Higasa, M. Shimizu, R. Yamada, F. Matsuda, HLA-HD: An accurate HLA typing algorithm for next-generation sequencing data, Hum. Mutat. 38 (2017) 788.
[25] S. Behera, J.R. Belyeu, X. Chen, L.F. Paulin, N.Q.H. Nguyen, E. Newman, et al. Identification of allele-specific KIV-2 repeats and impact on Lp(a) measurements for cardiovascular disease risk, bioRxiv 2023:2023.04.24.538128.
[26] M. Byrska-Bishop, U.S. Evani, X. Zhao, A.O. Basile, H.J. Abel, A.A. Regier, et al, High-coverage whole-genome sequencing of the expanded 1000 Genomes Project cohort including 602 trios, Cell 185 (2022) 3426.
[27] International HapMap C, D.M. Altshuler, R.A. Gibbs, L. Peltonen, D.M. Altshuler, R.A. Gibbs, et al, Integrating common and rare genetic variation in diverse human populations, Nature 467 (2010) 52.
[28] Genomes Project C, A. Auton, L.D. Brooks, R.M. Durbin, E.P. Garrison, H.M. Kang, et al, A global reference for human genetic variation, Nature 526 (2015) 68.
[29] W.W. Liao, M. Asri, J. Ebler, D. Doerr, M. Haukness, G. Hickey, et al, A draft human pangenome reference, Nature 617 (2023) 312.
[30] H. Li, R. Durbin, Fast and accurate short read alignment with Burrows-Wheeler transform, Bioinformatics 25 (2009) 1754.
[31] H. Li, Minimap2: pairwise alignment for nucleotide sequences, Bioinformatics 34 (2018) 3094.
[32] J.T. Robinson, H. Thorvaldsdottir, W. Winckler, M. Guttman, E.S. Lander, G. Getz, et al, Integrative genomics viewer, Nat. Biotechnol. 29 (2011) 24.
[33] A. Abyzov, A.E. Urban, M. Snyder, M. Gerstein, CNVnator: an approach to discover, genotype, and characterize typical and atypical CNVs from family and population genome sequencing, Genome Res. 21 (2011) 974.
[34] K. Hirayasu, S.-S. Khor, Y. Kawai, M. Shimada, Y. Omae, G. Hasegawa, et al, Identification of the hybrid gene LILRB5-3 by long-read sequencing and implication of its novel signalling function, Front. Immunol. 15 (2024).
[35] S. Purcell, B. Neale, K. Todd-Brown, L. Thomas, M.A. Ferreira, D. Bender, et al, PLINK: a tool set for whole-genome association and population-based linkage analyses, Am. J. Hum. Genet. 81 (2007) 559.
[36] K. Hirayasu, J. Sun, G. Hasegawa, Y. Hashikawa, K. Hosomichi, A. Tajima, et al, Characterization of LILRB3 and LILRA6 allelic variants in the Japanese population, J. Hum. Genet. 66 (2021) 739.