

METHODOLOGY ARTICLE

Open Access



Complete haplotype phasing of the MHC and KIR loci with targeted HaploSeq

Siddarth Selvaraj^{1†}, Anthony D. Schmitt^{1,2†}, Jesse R. Dixon^{1,3} and Bing Ren^{1,4,5*}

Abstract

Background: The MHC and KIR loci are clinically relevant regions of the genome. Typing the sequence of these loci has a wide range of applications including organ transplantation, drug discovery, pharmacogenomics and furthering fundamental research in immune genetics. Rapid advances in biochemical and next-generation sequencing (NGS) technologies have enabled several strategies for precise genotyping and phasing of candidate HLA alleles. Nonetheless, as typing of candidate HLA alleles alone reveals limited aspects of the genetics of MHC region, it is insufficient for the comprehensive utility of the aforementioned applications. For this reason, we believe phasing the entire MHC and KIR locus onto a single locus-spanning haplotype can be a critical improvement for better understanding transplantation biology.

Results: Generating long-range (>1 Mb) phase information is traditionally very challenging. As proximity-ligation based methods of DNA sequencing preserves chromosome-span phase information, we have utilized this principle to demonstrate its utility towards generating full-length phasing of MHC and KIR loci in human samples. We accurately (~99 %) reconstruct the complete haplotypes for over 90 % of sequence variants (coding and non-coding) within these two loci that collectively span 4-megabases.

Conclusions: By haplotyping a majority of coding and non-coding alleles at the MHC and KIR loci in a single assay, this method has the potential to assist transplantation matching and facilitate investigation of the genetic basis of human immunity and disease.

Keywords: HaploSeq, MHC, HLA-Typing, KIR, Phasing

Background

The major histocompatibility complex (MHC) and the killer cell immunoglobulin-like receptor (KIR) are important regulators of human immune responses and are involved in many human diseases [1, 2]. These loci are highly polymorphic, allowing an extensive antigen-presenting repertoire that enables strong immunity against a wide range of foreign antigens, pathogens and tumor cells [1–3]. At the same time, its immunogenic heterogeneity can also create incompatibility in allotransplantation procedures, causing graft rejections and graft-versus-host disease (GVHD) [4, 5]. Furthermore, many of the hundreds of genes within these immunogenic loci are

increasingly recognized as major susceptibility genes for drug hypersensitivity reactions and appear to play a significant role in numerous diseases, including cancer [6–8]. Taken together, the clinical implications of these loci make it useful to determine the sequence type of these molecules.

Typing of human leukocyte antigen (HLA) genes, located within the MHC locus, has traditionally been achieved in low resolution using serotyping techniques [9]. With advancements in technologies including PCR and more recently, next generation DNA sequencing (NGS), molecular-based methods have now enabled more clinically significant high-resolution HLA typing [10–12]. Notably, single-molecule NGS-based DNA sequencing has been demonstrated to resolve allele ambiguity by generating haplotypes of entire genes, resulting in super high-resolution (8-digit) haplotyping of HLA genes [13, 14]. However, even precise gene-level haplotyping may not be sufficient for many

* Correspondence: biren@ucsd.edu

†Equal contributors

¹Ludwig Institute for Cancer Research, La Jolla, CA 92093, USA

⁴Department of Cellular and Molecular Medicine, and UCSD Moores Cancer Center, University of California San Diego, La Jolla, CA 92093, USA

Full list of author information is available at the end of the article

applications. For example, while gene-level haplotyping for several candidate HLA genes can reduce risk of graft failure in transplantation matching, recipients could still be susceptible to graft-versus-host disease, as the totality of transplantation associated genes have not been fully understood. In particular, reports suggest that non-HLA gene families such as inflammatory genes, immune receptors, or others across the MHC or KIR haplotype can contribute to transplantation biology [15–17]. In addition, the strong linkage disequilibrium (LD) patterns across the MHC and KIR loci can allow coordinated functional activities of alleles on the same haplotype, complicating our understanding of transplantation biology [4, 5, 9, 18, 19]. Indeed, knowledge of haplotypes across several HLA genes has been shown to generate improved transplantation outcome predictions [19, 20] and can therefore facilitate determination of novel haplotype patterns for drug discovery and genome-wide association studies [21]. In summary, it appears useful to haplotype the entirety of the MHC and KIR loci to enable better understanding of immune genetics through analyses of compound heterozygous alleles.

Several experimental protocols have been developed to construct long-range haplotypes. Specifically, methods have been developed to generate mega-base-sized haplotypes [22–25], while others can phase the entire chromosome [26–29]. However, the adaptability of these methods to generate user-defined targeted haplotypes is unclear. More recently, Targeted Locus Amplification (TLA) has been developed to accomplish targeted phasing [30], but as the haplotypes from TLA are limited to a few-hundred kilobases, they may not be amenable for phasing large mega-base scale loci such as the MHC. Here, we develop a method, referred to as targeted HaploSeq, to generate full-length complete haplotypes of MHC and KIR loci from a single assay. Specifically, targeted HaploSeq combines the previously published HaploSeq [26] method developed for genome-wide haplotype phasing, with oligo capture and sequencing. As a proof of principle, we have applied targeted HaploSeq to the MHC and KIR loci in human lymphoblastoid cells. We phased over 90 % of the alleles in MHC and KIR loci at an estimated accuracy of ~99 %. To our knowledge, targeted HaploSeq is the first method to phase the MHC and KIR loci into a single haplotype structure. These results establish the utility of targeted HaploSeq for MHC and KIR typing in biomedical research as well as clinical settings.

Results and discussion

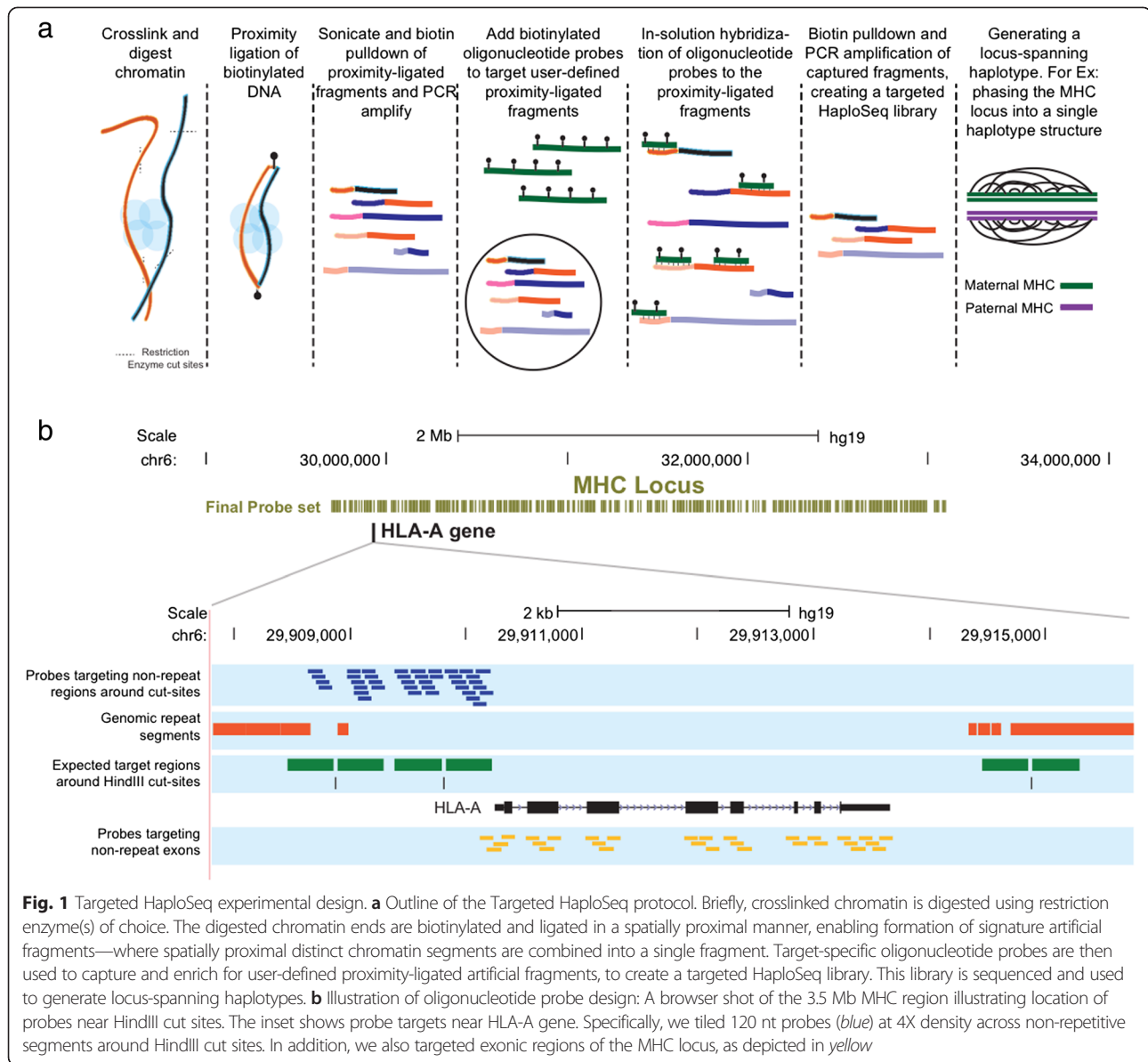
Experimental design

In the targeted HaploSeq method, a conventional Hi-C library [31] is generated using HindIII restriction digestion and amplified to obtain suitable material for oligonucleotide

probe-based enrichment of the target loci (Fig. 1a). Briefly, based on simulation results (Additional file 1: Fig. S1), we computationally generated the probe sequences, at 4X tiling density, using the SureDesign Software (Agilent Technologies) and targeted the non-repetitive +/- 400 bp regions adjacent to HindIII cut sites over the MHC and KIR loci (Fig. 1b, Additional file 2: Fig. S2a). In addition, to facilitate better phasing of genic regions, we designed probes across the exons within the MHC locus (Fig. 1a).

Next, by performing capture-sequencing [32, 33], we generated targeted HaploSeq data in GM12878 lymphoblastoid cells at 2× whole-genome sequencing depth with 30–50 fold target enrichment across the MHC and KIR loci (Fig. 2a, Additional file 2: Fig. S2b). More than 90 % of probes had at least 5-fold sequence coverage compared to data from virtual probes with an average of ~100 fold enrichment. This highlights the sensitivity of the probes from our targeted HaploSeq protocol. Next, to validate the quality of our targeted HaploSeq data, we compared it to a previously published HaploSeq dataset [26] generated from the same cell line. As HaploSeq utilizes chromatin interaction patterns to reconstruct haplotypes, we compared these between the two datasets and observed a high concordance ($r^2 = 0.8$, Fig. 2b, Additional file 3: Fig. S3a, b). By using haplotype inference from the parent-child trio whole-genome sequencing (WGS) data [34], we examined the fraction of chromatin interactions between the homologous chromosomes (h-trans interactions), whose rarity is critical for accurate *de novo* haplotyping. Similar to HaploSeq, targeted HaploSeq data rarely exhibit h-trans interactions (Additional file 4: Fig. S4a).

Of note, the MHC locus appears to have a higher h-trans ratio in both HaploSeq and targeted HaploSeq datasets, but several lines of evidence suggest that these might be systematic errors from sequencing and analysis protocols. First, reads supporting h-trans interactions are primarily observed in complex regions with high variant density (Additional file 4: Fig. S4b). Second, >85 % of h-trans interactions from targeted HaploSeq dataset originate from the same end of a given paired-end fragment. Lastly, about 95 % of these same-end h-trans interactions are also observed in long-fragment reads (LFR) in previously published Molecule datasets [25] from the same individual, indicating that a significant fraction of these h-trans interactions could have arisen from incorrect local haplotype inferences from the parent-child trio WGS data (Fig. 2c, d, Additional file 5). Taken together, our targeted HaploSeq data is of high quality and therefore enables accurate analyses of haplotype structures across the MHC and KIR loci.



High-resolution and accurate phasing of MHC and KIR loci

By utilizing heterozygous genotype identifications (SNVs) from the trio-based WGS data [34], we used the HaploSeq and LCP protocols to perform *de novo* haplotyping. We generated a single haplotype structure over the MHC locus resolving over 90 % of ~9,400 heterozygous alleles and we used the trio-based haplotype structure to estimate the accuracy of our approach to be ~97.7 % (Additional file 6: Fig. S5). However, as the parent-child trio data could have accumulated incorrect phasing at regions with high variant density, we repeated the *de novo* haplotyping protocol after ignoring variants that we found to be h-trans in both our and LFR datasets. Consequently, our phasing accuracy improved to 98.94 % (Additional file 6: Fig. S5). Despite reducing the phasing error by over

50 %, from 2.3 to 1.06 %, we still observe a majority of phasing errors occurring in the high variant density regions (Fig. 2e). This suggests that the accuracy can potentially be further improved by using long-read or single molecule technologies that may be more suitable for mapping such complex regions. Of note, unlike switch errors—the standard method to calculate phasing error rates where an incorrect haplotype block is penalized only once, we estimate error by testing each variant independently and therefore our error rate represents worst-case scenario. To this end, as the density of variants affects the resolution of HaploSeq-based haplotyping, we observed a relatively lower resolution phasing for the KIR locus (Additional file 1: Fig. S1b). Regardless, we obtained accurate phasing of 348 out of 353 variants resolved at the KIR

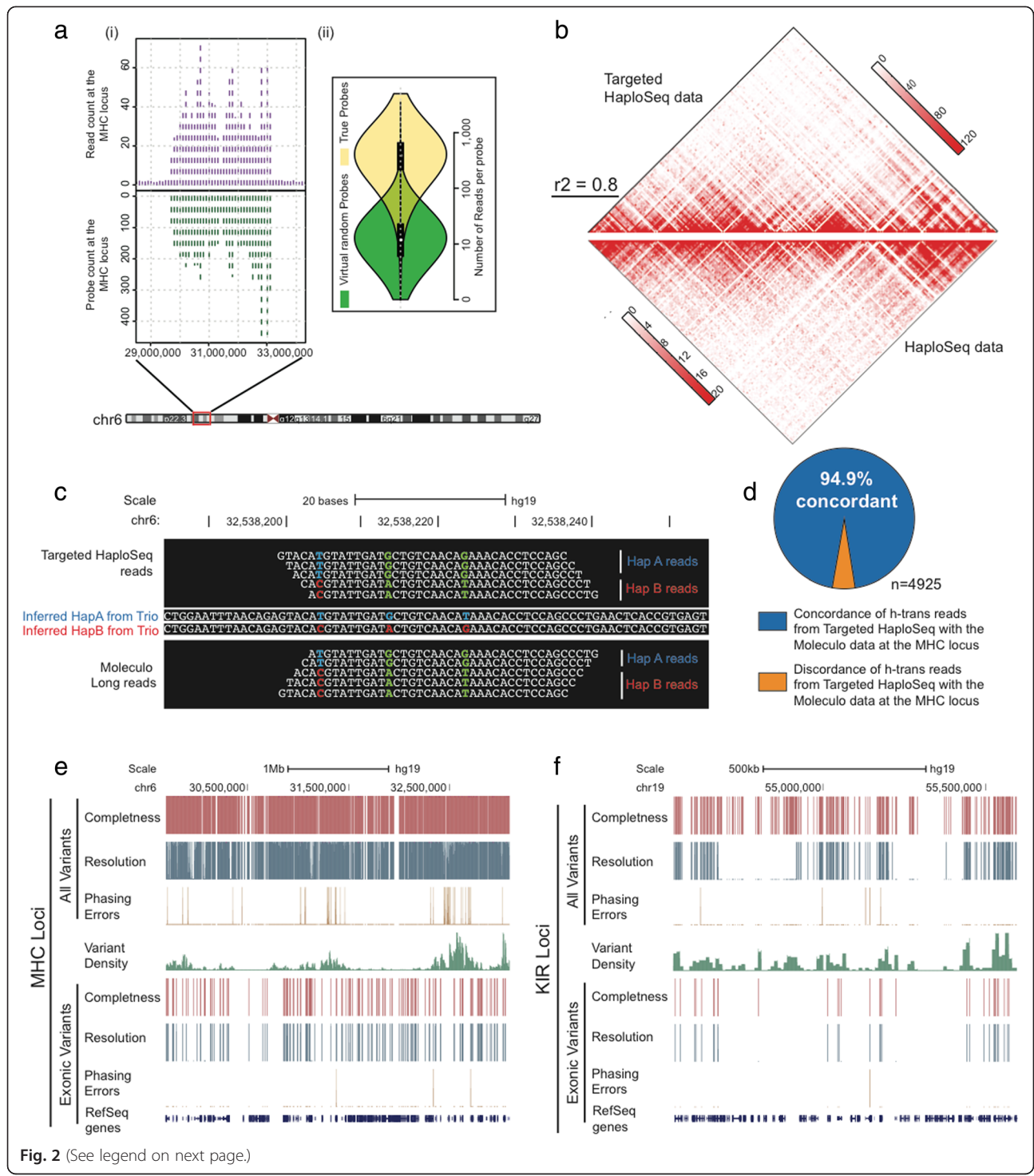


Fig. 2 (See legend on next page.)

(See figure on previous page.)

Fig. 2 High-resolution and accurate phasing of MHC and KIR loci. **a** (i) Top chart demonstrates enrichment of targeted HaploSeq reads at the 100 kb binned MHC locus and the bottom plot shows number of probes in 100 kb bins used across the MHC locus. Visually, we can observe a high correlation between these plots, demonstrating the expected relationship between density of probes and the sequencing depth of targeted HaploSeq reads. (ii) To illustrate the sensitivity of probes, we virtually created random probes flanking HindIII cut sites and compared the enrichment in targeted HaploSeq data from these regions to the data from regions containing true probes. We observe ~100 fold more reads from true regions (on target, *yellow*) than the random regions (off target, *green*) and this fold-enrichment suggests high-sensitivity of our probes. **b** High correlation of targeted HaploSeq and the previously published HaploSeq datasets from GM12878 cells at the MHC locus ($r^2 = 0.8$). **c** An example of haplotype inconsistency in the parent-child trio WGS data. Specifically, HapA (TGT-blue) and HapB (CAG-red) represent two haplotypes inferred from the trio dataset. Single-end reads from targeted HaploSeq (*top*) and Moleculo long-fragment reads (*bottom*) support a case of an inter-haplotype adjacent SNP-pair (*green*) and therefore raises an inconsistency with the parent-child trio haplotype inference. **d** Overall, ~95 % of the targeted HaploSeq reads representing homologous-trans (h-trans) interacting SNVs are concordant with the Moleculo LFR data. **e** High-resolution phasing capabilities of targeted HaploSeq method at the MHC locus. Completeness represents the collection of all heterozygous SNVs (*red*) within the MHC locus. Resolution represents the set of phased or resolved heterozygous SNVs in a single haplotype structure. While we observe ~1 % error, these errors are highly concentrated in the high variant density regions. The bottom section represents phasing of only exonic variants. **f** Similar figure as e) for the KIR locus

loci (Fig. 2f). Together, we resolved ~90 % of alleles among the MHC and KIR loci at ~99 % accuracy (Additional file 4: Fig. S4), demonstrating that our approach can generate complete, high-resolution and accurate haplotypes.

As current HLA typing protocols primarily type candidate genes across the MHC loci, we analyzed our method's phasing capabilities across heterozygous genes from MHC and KIR loci. In total, we resolve ~92 % of heterozygous variants, representing over 92 % of heterozygous genes, at an accuracy of 99.34 % (Fig. 2e, f, Additional file 7: Fig. S6). In this regard, we generate highly accurate phasing for several "classical" genes used in conventional HLA typing protocols. For example, in the case of genes such as HLA-B, HLA-C, HLA-DRB1, HLA-DQA1, HLA-DQB1, HLA-DPA1 and HLA-DPB1, we resolve phasing of >99.5 % of the heterozygous variants at 100 % accuracy. Similarly at the KIR loci, we accurately predict all but one exonic variant (Additional file 7: Fig. S6). To our knowledge, our method is the first to demonstrate high-resolution and accurate haplotyping across the entire MHC and KIR loci, phasing not only the highly diverse major and minor alleles, but also other important immunological genes and variants at non-genic regions across the locus together in a single haplotype structure.

Conclusions

Here, we describe the targeted HaploSeq method to generate large mega-base scale haplotypes in human cells. Using this technology, we reconstruct complete phase information of MHC and KIR loci. In principle, targeted HaploSeq is blind to genotyping and can be used to identify genetic variants *de novo* within the targeted loci. For example at the MHC locus, our method identified ~27 % of variants at an accuracy of 99.76 and 89.21 % for heterozygous and homozygous genotypes, respectively. This performance can be further improved with the use of multiple 4-base or 6-base cutters during Hi-C

library preparation [35], instead of a single 6-base recognizing restriction enzyme as demonstrated in this manuscript. Alternatively, computational strategies such as population-based imputation can be also be used to generate comprehensive genotyping [36].

High-resolution genotyping and phasing of immunogenic loci such as MHC and KIR has several applications. First, it has the potential to greatly improve the practice of HLA typing/matching for clinical transplantation procedures [13, 15, 20, 37], as this method provides access to alleles that are otherwise un-typed using current methods. In addition, with population-scale MHC and KIR haplotyping, our method can help to elucidate a refined set of minimal alleles that confer the highest risk for GVHD, thereby informing follow-up cost-effective selective typing of these most informative alleles. Second, as our method phases coding and non-coding cis-regulatory sequences together, one can study patterns of compound heterozygosity and linkage of human immune variation [7, 16, 17]. Finally, several studies have uncovered numerous disease-associated HLA and KIR alleles and by understanding long-range haplotypes, we can now start to unravel mechanistic underpinnings of human immune disorders [21, 38, 39].

Recently, proximity-ligation methods such as Hi-C have been demonstrated to be useful in assembling genomes *de novo* [40, 41]. As targeted HaploSeq obtains high-quality chromatin interaction datasets, similar to Hi-C [31], this methodology can potentially be used to generate diploid assembly of complex regions, such as the MHC or T-cell receptor beta (Tcrb) locus [42], of human and other large genomes. Similarly, Hi-C has also recently been used in metagenomics studies to deconvolute the species present in complex microbiome mixtures [43, 44]. With the advent of targeted HaploSeq, it is now possible to capture distinct loci that are informative and discriminative enough to delineate species mixtures based on the captured proximity-ligation fragments.

Taken together, we present targeted HaploSeq and demonstrate its application for targeted phasing of HLA and KIR loci in the human genome. We believe that this method will lead to new avenues in biomedical research and in personalized clinical genomics.

Data access

All sequencing data have been submitted to the Gene Expression Omnibus (GEO) database and will be publicly available upon publication. Data has been made available under the accession number GSE65726.

Ethics

Not applicable, non-human subjects.

Additional files

Additional file 1: Figure S1. Targeting regions around HindIII cut sites allows complete and high-resolution haplotyping of MHC and KIR loci. a) (i) and (ii) depict completeness and resolution at MHC locus, respectively. We simulated reads across \pm 400 bp from HindIII cut sites in the MHC region to study our ability to obtain complete and high-resolution haplotypes. As the MHC region has a high-density of het. variants (a het. variance every \sim 300 bases), 2X sequencing coverage is enough to generate complete haplotypes, regardless of read length. On the same lines, we obtain high-resolution seed haplotypes at low sequencing coverage. b) (i) and (ii) depict completeness and resolution at KIR locus respectively. On the contrary, as the KIR locus has a lower density of variants, high sequencing coverage is required to obtain complete haplotypes. In particular, 40 bp reads are not enough to obtain complete phasing even at 50X coverage and therefore is omitted in the resolution plot. Similarly, even at high sequencing coverage, resolution is very limited regardless of read length. (TIFF 8219 kb)

Additional file 2: Figure S2. Targeted enrichment at the KIR genomic locus. a) Genome browser shot of the \sim 1 Mb KIR region. The inset shows targets near KIR3DL2 gene, depicting target regions (green) around HindIII cut sites and repeat segments (red). We tiled 120-bp probes (blue) at 4X density across these non-repeat target regions. b) (i) Top Plot demonstrates enrichment of GM12878 Targeted-HaploSeq reads at the 100 kb binned KIR locus while the bottom plot shows number of probes used across the KIR locus. Together, these plots show a high correlation among probes and read enrichment. (ii) Plot demonstrating sensitivity of capture probes—the true probes capture reads \sim 100 fold than random probes created virtually near HindIII cut sites (TIFF 8219 kb)

Additional file 3: Figure S3. Targeted HaploSeq data has large pool of long insert fragments. a) Insert-size distribution of targeted HaploSeq (green) and b) HaploSeq (purple) in GM12878 LCLs. Both these datasets have similar amount of long-insert fragments which is critical for long range haplotyping. (TIFF 8219 kb)

Additional file 4: Figure S4. Homologous chromosomal interactions are rare and most of them are enriched in high variant density regions of the MHC loci. Using haplotypes identified from the parent-trio whole genome sequencing data, we define homologous trans (h-trans) interactions in the Targeted HaploSeq (green) and HaploSeq—from our previous publication (purple). a) h-trans interactions are rare $<$ 1 % in whole genome (i), about 5–6 % in the MHC locus (ii) and $<$ 0.5 % in KIR locus (iii). While h-trans interactions are $<$ 1 % whole-genome, we see them in significantly higher fractions at the MHC locus (\sim 5 %). Interestingly, majority of these are found at regions with very high variant density (b), suggesting that the haplotype predictions from parent-trio data at these regions could be error-prone, which in-turn results in higher h-trans in HaploSeq datasets. (TIFF 8219 kb)

Additional file 5: Online Methods. (DOCX 149 kb)

Additional file 6: Figure S6. Targeted HaploSeq generates a single (complete) haplotype structure across MHC/KIR locus. The performance

metric of the Targeted HaploSeq protocol, measured by completeness (span of the haplotype bloc), resolution (fraction of het. alleles resolved), and accuracy. While each of these metrics were defined after performing read-based as well as population based haplotyping, seed resolution is estimated only based on read-based haplotyping. The overall resolution is defined as the weighted average among all alleles across the MHC and KIR loci together. We observe over 50 % decrease in error rate from 2.3 to 1.06 % after correcting for potential incorrect local haplotypes from parent-trio data. (TIFF 8219 kb)

Additional file 7: Figure S7. Targeted HaploSeq generates high quality phasing of heterozygous genes. Over 92 % of exonic het. variants are phased at an accuracy of 99 %. (TIFF 8219 kb)

Competing interests

S.S., A.D.S, J.R.D., and B.R. are named inventors on a patent application on the technology described in this manuscript. S.S., J.R.D. and B.R. are co-founders of Arima Genomics, Inc.

Authors' contributions

B.R., S.S. and A.D.S conceived the strategy. A.D.S performed the experiments and optimized the targeted aspects of HaploSeq. J.R.D assisted in the experiments. S.S. conducted the analysis. S.S. prepared the manuscript with assistance from A.D.S and B.R. All authors read and approved the final manuscript.

Authors' information

Not applicable.

Availability of data and materials

Not applicable.

Acknowledgements

We thank members of the Ren laboratory for helpful suggestions throughout the course of this work.

Funding

Research is supported by funds from NIH (R01ES024984), LICR and UCSD provided to B. R. A.D.S is supported in part by the UCSD Genetics Training Grant (T32 GM008666).

Author details

¹Ludwig Institute for Cancer Research, La Jolla, CA 92093, USA. ²Biomedical Sciences Graduate Program, University of California San Diego, La Jolla, CA 92093, USA. ³Medical Scientist Training Program, University of California San Diego, La Jolla, CA 92093, USA. ⁴Department of Cellular and Molecular Medicine, and UCSD Moores Cancer Center, University of California San Diego, La Jolla, CA 92093, USA. ⁵Institute of Genomic Medicine, University of California San Diego, La Jolla, CA 92093, USA.

Received: 16 February 2015 Accepted: 16 September 2015

Published online: 05 November 2015

References

- Jin P, Wang E. Polymorphism in clinical immunology - From HLA typing to immunogenetic profiling. *J Transl Med.* 2003;1:8. doi:10.1186/1479-5876-1-8.
- Middleton D, Gonzelez F. The extensive polymorphism of KIR genes. *Immunology.* 2010;129:8–19. doi:10.1111/j.1365-2567.2009.03208.x.
- Horton R, Gibson R, Coggill P, Miretti M, Allcock RJ, Almeida J, et al. Variation analysis and gene annotation of eight MHC haplotypes: the MHC Haplotype Project. *Immunogenetics.* 2008;60:1–18. doi:10.1007/s00251-007-0262-2.
- Petersdorf EW. The major histocompatibility complex: a model for understanding graft-versus-host disease. *Blood.* 2013;122:1863–72. doi:10.1182/blood-2013-05-355982.
- Proll J, Danzer M, Stabentheiner S, Niklas N, Hackl C, Hofer K, et al. Sequence capture and next generation resequencing of the MHC region highlights potential transplantation determinants in HLA identical haematopoietic stem cell transplantation. *DNA Res.* 2011;18:201–10. doi:10.1093/dnares/dsr008.

6. Chung WH, Hung SJ, Chen YT. Human leukocyte antigens and drug hypersensitivity. *Curr Opin Allergy Clin Immunol*. 2007;7:317–23. doi:10.1097/ACI.0b013e3282370c5f.
7. Rizzo R, Bortolotti D, Baricordi OR, Fainardi E. New insights into HLA-G and inflammatory diseases. *Inflamm Allergy Drug Targets*. 2012;11:448–63.
8. Zeestraten EC, Reimers MS, Saadatmand S, Dekker JW, Liefers GJ, van den Elsen PJ, et al. Combined analysis of HLA class I, HLA-E and HLA-G predicts prognosis in colon cancer patients. *Br J Cancer*. 2014;110:459–68. doi:10.1038/bjc.2013.696.
9. Mahdi BM. A glow of HLA typing in organ transplantation. *Clin Transl Med*. 2013;2:6. doi:10.1186/2001-1326-2-6.
10. Chang CJ, Chen PL, Yang WS, Chao KM. A fault-tolerant method for HLA typing with PacBio data. *BMC bioinformatics*. 2014;15:296. doi:10.1186/1471-2105-15-296.
11. Boegel S, Lower M, Schafer M, Bukur T, de Graaf J, Boisguerin V, et al. HLA typing from RNA-Seq sequence reads. *Genome medicine*. 2012;4:102. doi:10.1186/gm403.
12. Bai Y, Ni M, Cooper B, Wei Y, Fury W. Inference of high resolution HLA types using genome-wide RNA or DNA sequencing reads. *BMC Genomics*. 2014;15:325. doi:10.1186/1471-2164-15-325.
13. Hosomichi K, Jinam TA, Mitsunaga S, Nakaoka H, Inoue I. Phase-defined complete sequencing of the HLA genes by next-generation sequencing. *BMC Genomics*. 2013;14:355. doi:10.1186/1471-2164-14-355.
14. Shiina T, Suzuki S, Ozaki Y, Taira H, Kikkawa E, Shigenari A, et al. Super high resolution for single molecule-sequence-based typing of classical HLA loci at the 8-digit level using next generation sequencers. *Tissue Antigens*. 2012;80:305–16. doi:10.1111/j.1399-0039.2012.01941.x.
15. Furst D, Muller C, Vucinic V, Bunjes D, Herr W, Gramatzki M, et al. High-resolution HLA matching in hematopoietic stem cell transplantation: a retrospective collaborative analysis. *Blood*. 2013;122:3220–9. doi:10.1182/blood-2013-02-482547.
16. Mullighan C, Heatley S, Doherty K, Szabo F, Grigg A, Hughes T, et al. Non-HLA immunogenetic polymorphisms and the risk of complications after allogeneic hemopoietic stem-cell transplantation. *Transplantation*. 2004;77:587–96.
17. Guo Z, Hood L, Malkki M, Petersdorf EW. Long-range multilocus haplotype phasing of the MHC. *Proc Natl Acad Sci U S A*. 2006;103:6964–9. doi:10.1073/pnas.0602286103.
18. Traherne JA. Human MHC architecture and evolution: implications for disease association studies. *Int J Immunogenet*. 2008;35:179–92. doi:10.1111/j.1744-313X.2008.00765.x.
19. Petersdorf EW, Malkki M, Horowitz MM, Spellman SR, Haagenson MD, Wang T. Mapping MHC haplotype effects in unrelated donor hematopoietic cell transplantation. *Blood*. 2013;121:1896–905. doi:10.1182/blood-2012-11-465161.
20. Petersdorf EW, Malkki M, Gooley TA, Martin PJ, Guo Z. MHC haplotype matching for unrelated hematopoietic cell transplantation. *PLoS Med*. 2007;4, e8. doi:10.1371/journal.pmed.0040008.
21. Larsen CE, Alford DR, Trautwein MR, Jalloh YK, Tarnacki JL, Kunnenkeri SK, et al. Dominant sequences of human major histocompatibility complex conserved extended haplotypes from HLA-DQA2 to DAXX. *PLoS Genet*. 2014;10, e1004637. doi:10.1371/journal.pgen.1004637.
22. Peters BA, Kermani BG, Sparks AB, Alferov O, Hong P, Alexeev A, et al. Accurate whole-genome sequencing and haplotyping from 10 to 20 human cells. *Nature*. 2012;487:190–5. doi:10.1038/nature11236.
23. Kaper F, Swamy S, Klotzle B, Munchel S, Cottrell J, Bibikova M, et al. Whole-genome haplotyping by dilution, amplification, and sequencing. *Proc Natl Acad Sci*. 2013;110:5552–7.
24. Kitzman JO, Mackenzie AP, Adey A, Hiatt JB, Patwardhan RP, Sudmant PH, et al. Haplotype-resolved genome sequencing of a Gujarati Indian individual. *Nat Biotechnol*. 2011;29:59–63. doi:10.1038/nbt.1740.
25. Kuleshov V, Xie D, Chen R, Pushkarev D, Ma Z, Blauwkamp T, et al. Whole-genome haplotyping using long reads and statistical methods. *Nat Biotechnol*. 2014;32:261–6. doi:10.1038/nbt.2833.
26. Selvaraj S, Dixon J R, Bansal V, Ren B. Whole-genome haplotype reconstruction using proximity-ligation and shotgun sequencing. *Nat Biotechnol*. 2013;31:1111–8. doi:10.1038/nbt.2728.
27. Kirkness EF, Grindberg RV, Yee-Greenbaum J, Marshall CR, Scherer SW, Lasken RS, et al. Sequencing of isolated sperm cells for direct haplotyping of a human genome. *Genome Res*. 2013;23:826–32. doi:10.1101/gr.144600.112.
28. Fan HC, Wang J, Potanina A, Quake SR. Whole-genome molecular haplotyping of single cells. *Nat Biotechnol*. 2011;29:51–7. doi:10.1038/nbt.1739.
29. Yang H, Chen X, Wong H. Completely phased genome sequencing through chromosome sorting. *Proc Natl Acad Sci*. 2012;109:3190–3190. doi:10.1073/pnas.1200309109.
30. de Vree PJ, de Wit E, Yilmaz M, van de Heijning M, Klous P, Versteeg MJ, et al. Targeted sequencing by proximity ligation for comprehensive variant detection and local haplotyping. *Nat Biotechnol*. 2014;32:1019–25. doi:10.1038/nbt.2959.
31. Lieberman-Aiden E, van Berkum NL, Williams L, Imakaev M, Ragoczy T, Telling A, et al. Comprehensive mapping of long-range interactions reveals folding principles of the human genome. *Science*. 2009;326:289–93. doi:10.1126/science.1181369.
32. Gnirke A, Melnikov A, Maguire J, Rogov P, LeProust EM, Brockman W, et al. Solution hybrid selection with ultra-long oligonucleotides for massively parallel targeted sequencing. *Nat Biotechnol*. 2009;27:182–9. doi:10.1038/nbt.1523.
33. Ng SB, Turner EH, Robertson PD, Flygare SD, Bigham AW, Lee C, et al. Targeted capture and massively parallel sequencing of 12 human exomes. *Nature*. 2009;461:272–6. doi:10.1038/nature08250.
34. Genomes Project C, Abecasis GR, Altshuler D, Auton A, Brooks LD, Durbin RM, et al. A map of human genome variation from population-scale sequencing. *Nature*. 2010;467:1061–73. doi:10.1038/nature09534.
35. Rao SS, Huntley MH, Durand NC, Stamenova EK, Bochkov ID, Robinson JT, et al. A 3D Map of the Human Genome at Kilobase Resolution Reveals Principles of Chromatin Looping. *Cell*. 2014;159:1665–80. doi:10.1016/j.cell.2014.11.021.
36. Browning BL, Browning SR. Improving the Accuracy and Efficiency of Identity-by-Descent Detection in Population Data. *Genetics*. 2013;194:459–71. doi:10.1534/genetics.113.150029.
37. Lee SJ, Klein J, Haagenson M, Baxter-Lowe LA, Confer DL, Eapen M, et al. High-resolution donor-recipient HLA matching contributes to the success of unrelated donor marrow transplantation. *Blood*. 2007;110:4576–83. doi:10.1182/blood-2007-06-097386.
38. Traherne JA, Horton R, Roberts AN, Miretti MM, Hurler ME, Stewart CA, et al. Genetic analysis of completely sequenced disease-associated MHC haplotypes identifies shuffling of segments in recent human history. *PLoS Genet*. 2006;2, e9. doi:10.1371/journal.pgen.0020009.
39. Romero V, Larsen CE, Duke-Cohan JS, Fox EA, Romero T, Clavijo OP, et al. Genetic fixity in the human major histocompatibility complex and block size diversity in the class I region including HLA-E. *BMC Genet*. 2007;8:14. doi:10.1186/1471-2156-8-14.
40. Burton JN, Adey A, Patwardhan RP, Qiu R, Kitzman JO, Shendure J. Chromosome-scale scaffolding of de novo genome assemblies based on chromatin interactions. *Nat Biotechnol*. 2013;31:1119–25. doi:10.1038/nbt.2727.
41. Kaplan N, Dekker J. High-throughput genome scaffolding from in vivo DNA interaction frequency. *Nat Biotechnol*. 2013;31:1143–7. doi:10.1038/nbt.2768.
42. Spicuglia S, Pekowska A, Zacarias-Cabeza J, Ferrier P. Epigenetic control of Tcrb gene rearrangement. *Semin Immunol*. 2010;22:330–6. doi:10.1016/j.smim.2010.07.002.
43. Beitel CW, Froenicke L, Lang JM, Korf IF, Micheltore RW, Eisen JA, et al. Strain- and plasmid-level deconvolution of a synthetic metagenome by sequencing proximity ligation products. *PeerJ*. 2014;2, e415. doi:10.7717/peerj.415.
44. Burton JN, Liachko I, Dunham MJ, Shendure J. Species-level deconvolution of metagenome assemblies with Hi-C-based contact probability maps. *G3*. 2014;4:1339–46. doi:10.1534/g3.114.011825.