HapMap scanning of novel human minor histocompatibility antigens

Michi Kamei,1,2 Yasuhito Nannya,3-5 Hiroki Torikai,1 Takakazu Kawase,1,6 Kenjiro Taura,7 Yoshihiro Inamoto,8 Taro Takahashi,8 Makoto Yazaki,9 Satoko Morishima,1 Kunio Tsujimura,10 Koichi Miyamura,5,8 Tetsuya Ito,2 Hajime Togari,2 Stanley R. Riddell,11 Yoshihisa Kodera,5,8 Yasuo Morishima,5,12 Toshitada Takahashi,13 Kiyotaka Kuzushima,1 †Seishi Ogawa,4,5 and †Yoshiki Akatsuka1,5

1Division of Immunology, Aichi Cancer Center Research Institute, Nagoya, Japan; 2Department of Pediatrics and Neonatology, Nagoya City University, Graduate School of Medical Science, Nagoya, Japan; 3Department of Hematology/Oncology and *The 21st Century Center of Excellence (COE) Program, Graduate School of Medicine, University of Tokyo, Tokyo, Japan; 4Core Research for Evolutional Science and Technology, Japan Science and Technology Agency, Saitama, Japan; 5Division of Epidemiology and Prevention, Aichi Cancer Center Research Institute, Nagoya, Japan; 6Department of Information and Communication Engineering, Graduate School of Information Science, University of Tokyo, Tokyo, Japan; 7Department of Hematology, Japanese Red Cross Nagoya First Hospital, Nagoya, Japan; 8Department of Pediatrics, Higashi Municipal Hospital of Nagoya, Nagoya, Japan; 9Department of Infectious Diseases, Hamamatsu University School of Medicine, Hamamatsu, Japan; 10Program in Immunology, Clinical Research Division, Fred Hutchinson Cancer Research Center, Seattle, WA; 11Department of Hematology and Cell Therapy, Aichi Cancer Center Central Hospital, Nagoya, Japan; and 12Aichi Comprehensive Health Science Center, Aichi Health Promotion Foundation, Chita-gun, Japan

Minor histocompatibility antigens (mHags) are molecular targets of allograft immunity associated with hematopoietic stem cell transplantation (HSCT) and involved in graft-versus-host disease, but they also have beneficial antitumor activity. mHags are typically defined by host SNPs that are not shared by the donor and are immunologically recognized by cytotoxic T cells isolated from post-HSCT patients. However, the number of molecularly identified mHags is still too small to allow prospective studies of their clinical importance in transplantation medicine, mostly due to the lack of an efficient method for isolation. Here we show that when combined with conventional immunologic assays, the large data set from the International HapMap Project can be directly used for genetic mapping of novel mHags. Based on the immunologically determined mHag status in HapMap panels, a target mHag locus can be uniquely mapped through whole genome association scanning taking advantage of the unprecedented resolution and power obtained with more than 3 000 000 markers. The feasibility of our approach could be supported by extensive simulations and further confirmed by actually isolating 2 novel mHags as well as 1 previously identified example. The HapMap data set represents an invaluable resource for investigating human variation, with obvious applications in genetic mapping of clinically relevant human traits. (Blood. 2009;113:5041-5048)

Introduction

The antitumor activity of allogeneic hematopoietic stem cell transplantation (HSCT), which is a curative treatment for many patients with hematologic malignancies, is mediated in part by immune responses that are elicited as a consequence of incompatibility in genetic polymorphisms between the donor and the recipient.1,2 Analysis of patients treated for posttransplantation relapse with donor lymphocytes has shown tumor regression to be correlated with expansion of cytotoxic T lymphocytes (CTLs) specific for hematopoiesis-restricted minor histocompatibility antigens (mHags).3,4 mHags are peptides, presented by major histocompatibility complex (MHC) molecules, derived from intracellular proteins that differ between donor and recipient due mostly to the lack of an efficient method for isolation. Here we show that when combined with conventional immunologic assays, the large data set from the International HapMap Project can be directly used for genetic mapping of novel mHags. Based on the immunologically determined mHag status in HapMap panels, a target mHag locus can be uniquely mapped through whole genome association scanning taking advantage of the unprecedented resolution and power obtained with more than 3 000 000 markers. The feasibility of our approach could be supported by extensive simulations and further confirmed by actually isolating 2 novel mHags as well as 1 previously identified example. The HapMap data set represents an invaluable resource for investigating human variation, with obvious applications in genetic mapping of clinically relevant human traits. (Blood. 2009;113:5041-5048)
CTLs. To circumvent these drawbacks, we have sought to take advantage of publicly available HapMap resources. Here, we describe a powerful approach for rapidly identifying mHag loci using a large genotyping data set and LCLs from the International HapMap Project for genome-wide association analysis.

Methods

Cell lines and CTL clones

The HapMap LCL samples were purchased from the Coriell Institute (Camden, NJ). All LCLs were maintained in RPMI1640 supplemented with 10% fetal calf serum, 2 mM l-glutamine, and 1 mM sodium pyruvate. Because the recognition of a mHag requires presentation on a particular type of HLA molecule, the LCLs were stably transduced with a retroviral vector encoding the restriction HLA cDNA for a given CTL clone when necessary.

CTL lines were generated from recipient peripheral blood mononuclear cells obtained after transplantation with stimulation by those harvested before HSCT after irradiation (33 Gy), and thereafter stimulated weekly in RPMI 1640 supplemented with 10% pooled human serum and 2 mM l-glutamine. Recombinant human interleukin-2 was added on days 1 and 5 after the second and third stimulations. CTL clones were isolated by standard limiting dilution and expanded as previously described.16,17 HLA restriction was determined by conventional CRAs against a panel of LCLs sharing HLA alleles with the CTLs. All clinical samples were collected based on a protocol approved by the Institutional Review Board Committee at Aichi Cancer Center and the University of Tokyo and written informed consent was obtained in accordance with the Declaration of Helsinki.

Immunophenotyping of HapMap LCLs and high-density genome-wide scanning of mHag loci

Case (mHag+) - control (mHag-) LCL panels were generated by screening corresponding restriction HLA-transduced CHB and JPT HapMap LCL panels with each CTL clone using CRAs. Briefly, target cells were labeled with 0.1 mCi of 51Cr for 2 hours, and 10^3 target cells per well were mixed with CTL at a predetermined E/T ratio in a standard 4-hour CRA. All assays were performed at least in duplicate. The percent specific lysis was calculated by ((Experimental cpm - Spontaneous cpm) / (Maximum cpm - Spontaneous cpm)) × 100. After normalization by dividing their percent specific lysis values by that of positive control LCL (typically recipient-calculated by ((Experimental cpm - Spontaneous cpm) / (Maximum cpm - Spontaneous cpm)) - control (mHag-) assay cpm for each panel to identify those (i,j) combinations, in which f(i,j) exceeds the upper 1 percentile point of the simulated 10,000 maximum values, g(i,j)^p < 0.01.

When proxies are not complete (ie, r^2 < 1), the expected values will be decayed by the factor of r^2, and further reduced due to the probabilities of false positive (f+) and negative (f-) assays, and expressed as f(i,j) = (i + j) × r^2 through an apparent r^2 (P^2) as provided in formula 1. Under given probabilities of assay errors and maximum LD strength between markers and the mHag allele, we can expect to identify target mHag loci for those (i,j) sets that satisfy f(i,j) > g(i,j)^p < 0.01.

Empirical estimation of distributions of r^2

The maximum r^2 value (r^2max) between a given mHag allele and one or more Phase II HapMap SNPs was estimated based on the observed HapMap data set. Each Phase II HapMap SNP was assumed to represent a target mHag allele, and the (r^2max) was calculated, taking into account all the Phase II HapMap SNPs less than 500 kb apart from the target SNP.

Confirmaory genotyping

Genotyping was carried out either by TaqMan MGB technology (Applied Biosystems, Foster City, CA) with primers and probes for HA-1 mHag according to the manufacturer’s protocol using an ABI 7900HT with the aid of SDS version 2.2 software (Applied Biosystems) or by direct sequencing of amplified cDNA for the SLCA15 gene. cDNA was reverse transcribed from total RNA extracted from LCLs, and polymerase chain reaction (PCR) was conducted with cDNA with the corresponding primers. Amplified DNA samples were sequenced using BigDye Terminator version 3.1 (Applied Biosystems). The presence or absence (deletion) of the UGT2B17 gene was confirmed by genomic PCR with 2 primer sets for exons 1 and 6 as described previously14 using DNA isolated from LCLs of interest.

Epitope mapping

A series of deletion mutant cDNAs were designed and cloned into pcDNA3.1/V5-His TOPO plasmid (Invitrogen, Carlsbad, CA). Thereafter, 293T cells that had been transduced with restricting HLA class I cDNA for individual CTL clones were transfected with each of the deletion mutants and cocultured with the CTL clone overnight to induce interferon (IFN)-γ release, which was then evaluated by enzyme-linked immunosorbent assay (ELISA) as previously described.9

For SLC1A5, expression plasmids encoding full-length cDNA and the exon 1 of recipient and donor origin were first constructed because only the SNP in the exon 1 was found to be concordant with susceptibility to CTL-3B6. Next, amino (N)- and (carboxyl) C-terminus–truncated mini-genes encoding polypeptides around the polymorphic amino acid defined by the SNP were amplified by PCR from SLC1A5 exon 1 cDNA as template and cloned into the above plasmid. The constructs all encoded a Kozak
sequence and initiator methionine (CCACC-ATG) and for C-terminus deletions a stop codon (TAG).

For UGT2B17, a series of C-terminus deletion mutants with approximately 200 bp spacing was first constructed as above. For further mapping, N-terminus deletion mutants were added to the region that was deduced to be potentially encoding the CTL-1B2 epitope. For prediction of a CTL epitope, the HLA Peptide Binding Predictions algorithm on the Bioinformatics & Molecular Analysis Section (BIMAS) website (http://www.bimas.cit.nih.gov/molbio/hla_bind/) was used because HLA-A*0206 has a similar binding motif to that of A*0201.

Epitope reconstitution assay

The candidate mHag epitopes and allelic counterpart peptides (in case of SLC1A5) were synthesized by standard Fmoc chemistry. 

Results and discussion

Statistical approach and estimation of potential overfitting

We reasoned that the mHag locus recognized by a given CTL clone could be defined by grouping LCLs from a HapMap panel into mHag+ and mHag− subpanels according to their susceptibility to lysis by the CTL clone and then performing an association scan using the highly qualified HapMap data set containing more than 3 000 000 SNP markers. The relevant genetic trait here is expected to show near-complete penetrance, and the major concern with this approach arises from the risk of overfitting observed phenotypes to one or more incidental SNPs with this large number of HapMap SNPs under the relatively limited size of freedom due to small numbers of independent HapMap samples (90 for JPT CHB and 60 for CEU and YRI, when not including their offspring). To address this problem, we first estimated the maximum sizes of the test statistics (here, \( \chi^2 \) values) under the null hypothesis (ie, no associated SNPs within the HapMap set) by simulating 10 000 case-control HapMap panels under different experimental conditions, and compared them with the expected size of test statistic values from the marker SNPs associated with the target SNP, assuming different linkage disequilibrium (LD), or \( r^2 \) values in between. As shown in Figure 1, the possibility of overfitting became progressively reduced as the number of LCLs increased, which would allow for identification of the target locus in a broad range of \( r^2 \) values, except for those mHags having very low minor allele frequencies (MAF) below
approximately 0.05. According to our estimation using the Phase II HapMap data (see “Methods”), the majority (>90%) of common target SNPs (MAF > 0.05) could be captured by one or more HapMap SNPs with more than 0.8 of \( r^2 \) (Figure S1, available on the Blood website; see the Supplemental Materials link at the top of the online article), ensuring a high probability of detecting an association (Figure 1 left panels). The simulation of pseudo-Phase II sets generated from the ENCODE regions provided a similar estimation.13 False positive and negative immunophenotyping results could also complicate the detection, reducing the expected test statistics through the “apparent” \( r^2 \) values (\( \hat{r}^2 \)), as defined by

\[
\hat{r}^2 = r^2 \times \frac{(1 - f_P - f_N)^2}{(1 - f_P + f_N q)(1 - f_N + f_P q)}
\]

where \( f_P \), \( f_N \), and \( q \) represent false typing probabilities with positive and negative LCL panels, and the ratio of the positive to the negative LCL number, respectively. However, the high precision of cytotoxicity assays (\( f_P \sim 0.1, f_N \sim 0 \)) limits this drawback from the second term to within acceptable levels and allows for sensitive mHag locus mapping with practical sample sizes (Figure 1 middle and right panels), suggesting the robustness of our novel approach.

### Evaluation of the detection power for known mHags

Based on these considerations, we then assessed whether this approach could be used to correctly pinpoint known mHag loci (Table S1). Because the relevant mHag alleles are common SNPs and directly genotyped in the Phase II HapMap set, or if not, located within a well-defined LD block recognized in this set (Figure S2), their loci would be expected to be uniquely determined with an acceptable number of samples, as predicted from Figure 1. To test this experimentally, we first mapped the locus for HA-1\(^{14}\) mHag\(^7\) by evaluating recognition of the HLA-A*0206–transduced HapMap cell panel with HLA-A*0206–restricted CTL-4B1.20 After screening 58 well-growing LCLs from the JPT + CHB panel with CRAs using CTL-4B1 (Figure S3A; Tables S2,S3), we obtained 37 mHag\(^+\) and 21 mHag\(^-\) LCLs, which were tested for association at 3 933 720 SNP loci. The SNP (rs1801284) encoding the mHag is located within a HapMap LD block on chromosome 19q13.3, but is not directly genotyped within this data set. The genome-wide scan clearly indicated a unique association with the HA-1\(^{14}\) locus within the HMMHAI gene, showing a peak \( \hat{r}^2 \) statistic of 52.8 (not reached in 100 000 permutations) at rs10421359 (Figures 2A,3A; Tables S2,S3).
Identification of novel mHags

We next applied this method to mapping novel mHags recognized by CTL clone 3B6, which is HLA-B*4002–restricted; and CTL clone 1B2, which is HLA-A*0206–restricted. Both clones had been isolated from peripheral blood samples of post-HSCT different patients. In preliminary CRAs with the JPT/H11001 CHB panel, allele frequencies of target mHags for CTL-3B6 and CTL-1B2 in this panel were estimated as approximately 25% and approximately 45%, respectively (data not shown). After screening 72 JPT + CHB LCLs with CTL-3B6, 36 mHag⁺ and 14 mHag⁻ LCLs were obtained, leaving 22 LCLs undetermined based on empirically determined thresholds (> 51% for mHag⁺ LCLs and < 11% for mHag⁻ LCLs; Figure S3B, Tables S2,S4). As shown in Figure 2B, the χ² statistics calculated from the immunophenotyping data produced discrete peaks in the LCL sets. The peak in chromosome 19q13.3 for the CTL-3B6 set showed the theoretically maximum χ² value of 50 (not reached in 100 000 permutations) at rs3027952, which was mapped within a small LD block of...
approximately 182 kb, or more narrowly within its 35 kb sub-block containing a single gene, \textit{SLC1A5}, as a candidate mHag gene (Figure 3B). In fact, when expressed in 293T cells with HLA-B*4002 transgene, recipient-derived, but not donor-derived, \textit{SLC1A5} cDNA was able to stimulate IFN-\gamma secretion from CTL-3B6 (Figure 4A), indicating that \textit{SLC1A5} actually encodes the target mHag recognized by CTL-3B6. Conventional epitope mapping using a series of deletion mutants of \textit{SLC1A5} cDNA finally identified an undecameric peptide, AEATANGGLAL, as the minimal epitope (Figure 4A). The donor-type AEPTANGGLAL induced IFN-\gamma with a 2-log lower efficiency, suggesting that AEPTANGGLAL may not be transported efficiently into the ER
because endogenous expression of a minigene encoding AEPTANG-GLAL was not recognized by CTL-3B1 (Figure 4B). Unfortunately, although the peak statistic value showed the theoretically maximum value for this data set, it did not conform to the relevant SNP for this mHag (rs3027956) due to high genotyping errors of the HapMap data at this particular SNP. However, the result of our resequencing showed complete concordance with the presence of the rs3027956 SNP and recognition in the cytotoxicity assay (Table S4).

Similarly, 13 mHag- and 32 mHag+ LCLs were identified from the screening of 45 JPT LCLs from the same panel using CTL-1B2 (Figure S3C; Tables S2, S5). The $\chi^2$ statistics calculated from the immunophenotyping data produced bimodal discrete peaks with this LCL set. The target locus for the mHag recognized by CTL-1B2 was identified at a peak (max $\chi^2 = 44$, not reached in 100,000 permutations) within a 598-kb block on chromosome 4q13.1, coinciding with the locus for a previously reported mHag, UGT2B1717 (Figures 2C, 2D). In fact, our epitope mapping using UGT2B17 CDNA deletion mutants (Figure 4C), prediction of candidate epitopes by HLA-binding algorithms19 (Figure 4D) and epitope reconstitution assays (Figure 4E), successfully identified a novel nonameric peptide, CVATMIFM. Of particular note, this mHag was not defined by a SNP but by a CNV (ie, a null allele18) that is in complete LD with the SNPs showing the maximum $\chi^2$ value (Table S5). Transplanted T cells from donors lacking both UGT2B17 alleles are sensitized in recipients possessing at least 1 copy of this gene.19 Although LD between SNPs and CNVs has been reported to be less prominent,21 this is an example where a CNV trait could be captured by a SNP-based genome-wide association study.

The recent generation of the HapMap has had a profound impact on human genetics.12,13 In the field of medical genetics, the HapMap is a central resource for the development of theories and methods that have made well-powered, genome-wide association studies of common human diseases a reality.22-28 The HapMap samples provide not only an invaluable reference for genetic variations within human populations, but highly qualified genotypes that enable gene-wide scanning. Here, we have demonstrated how effectively HapMap resources can be used for genetic mapping of clinically relevant human traits. No imputations and tagging strategies are required25,28 and the potential loss of statistical power due to very limited sample sizes is circumvented by accurate immunologic detection of the traits.

Using publicly available HapMap resources, high-throughput identification of mHag genes is possible without highly specialized equipment or expensive microarrays. Except for clinically irrelevant mHags with very low allele frequencies (eg, MAF < 5%), the target of a given CTL can be sensitively mapped within a mean LD block size, typically containing just a few candidate genes. The methodology described here will facilitate construction of a large panel of human mHags including those presented by MHC class II molecules, and promote our understanding of human allo-immunity and development of targeted allo-immune therapies for hematologic malignancies.1,2 The HapMap scan approach may be useful for exploring other genetic traits or molecular targets (eg, differential responses to some stress or drugs), if they can be discriminated accurately through appropriate biologic assays. In this context, the recent report that we may reprogram the fate of terminally differentiated human cells29 is encouraging, indicating possible exploration of genotypes that are relevant to cell types other than immortalized B cells.

Acknowledgments

We thank Drs P. Martin and W. Ho for critically reading the manuscript; and Ms Keiko Nishida, Dr Ayako Demachi-Okamura, Dr Yukiko Watanabe, Ms Hiromi Tamaki, and the staff members of the transplant centers for their generous cooperation and technical expertise.

This study was supported in part by a grant for Scientific Research on Priority Areas (B01; no. 17016089) from the Ministry of Education, Culture, Science, Sports, and Technology, Japan; grants for Research on the Human Genome, Tissue Engineering Food Biotechnology and the Second and Third Team Comprehensive 10-year Strategy for Cancer Control (no. 26) from the Ministry of Health, Labor, and Welfare, Japan; and a grant-in-aid from Core Research for Evolutional Science and Technology (CREST) of Japan.

Authorship

Contributions: M.K. performed most of the immunologic experiments and analyzed the data; Y.N. performed the majority of genetic analyses and analyzed the data; H.T., T.K., M.Y., S.M. and K.Tsujimura performed research; K.Taura contributed to the computational simulation; Y.I., Taro T., K.M., Y.K. and Y.M. collected clinical data and specimens; T.I., H.T., S.R.R., Toshitada T. and K.K. contributed to data analysis and interpretation, and writing of the article; and Y.A. and S.O. supervised the entire project, designed and coordinated most of the experiments in this study, and contributed to manuscript preparation.

Conflict-of-interest disclosure: The authors declare no competing financial interests.

Correspondence: Seishi Ogawa, MD, PhD, Department of Hematology and Oncology, Department of Regeneration Medicine for Hematopoiesis, The 21st Century COE Program, Graduate School of Medicine, University of Tokyo, 7-3-1, Hongo, Bunkyo-ku, Tokyo 113-8655, Japan; e-mail: sogawa-tky@umin.ac.jp; or Yoshihiko Akatsuka, MD, PhD, Division of Immunology, Aichi Cancer Center Research Institute, 1-1 Kanokoden, Chikusa-ku, Nagoya 464-8681, Japan; e-mail: yakatsuk@aichi-cc.jp.

References


HapMap scanning of novel human minor histocompatibility antigens

Michi Kamei, Yasuhiro Nannya, Hiroki Torikai, Takakazu Kawase, Kenjiro Taura, Yoshihiro Inamoto, Taro Takahashi, Makoto Yazaki, Satoko Morishima, Kunio Tsujimura, Koichi Miyamura, Tetsuya Ito, Hajime Togari, Stanley R. Riddell, Yoshihisa Kodera, Yasuo Morishima, Toshitada Takahashi, Kiyotaka Kuzushima, Seishi Ogawa and Yoshiki Akatsuka

Updated information and services can be found at:
http://www.bloodjournal.org/content/113/21/5041.full.html

Articles on similar topics can be found in the following Blood collections
  Plenary Papers (499 articles)
  Transplantation (2233 articles)

Information about reproducing this article in parts or in its entirety may be found online at:
http://www.bloodjournal.org/site/misc/rights.xhtml#repub_requests

Information about ordering reprints may be found online at:
http://www.bloodjournal.org/site/misc/rights.xhtml#reprints

Information about subscriptions and ASH membership may be found online at:
http://www.bloodjournal.org/site/subscriptions/index.xhtml