INTERNATIONAL JOURNAL OF
IMMUNOGENETICS

doi: 10.1111/iji.12029

16th IHIW: Global distribution of extended HLA haplotypes

M. Askar*, J. Daghstani*, D. Thomas*, N. Leahy†, P. Dunn‡, F. Claas§, S. Doran¶, H. Saji, S. Kanangat††, M. Karoichane‡‡, A. Tambur§§, D. Monos¶¶, M. El-Khalifa***, V. Turner†††, M. Kamoun‡‡‡, M. Mustafa§§§, D. Ramon¶¶¶, M. Gandhi****, A. Vernaza††††, C. Gorodezky‡‡‡‡, D. Wagenknecht§§§§, M. Gautreaux¶¶¶¶, A. Hajeer*****, Z. Kashi†††††& M. Fernandez-Vina‡‡‡‡‡**

Summary

This report describes the project to identify the global distribution of extended HLA haplotypes, a component of 16th International HLA and Immunogenetics Workshop (IHIW), and summarizes the initial analyses of data collected. The project aims to investigate extended HLA haplotypes, compare their distribution among different populations, assess their frequency in hematopoietic stem cell unrelated donor registries and initiate an international family studies database and DNA repository to be made publicly available. HLA haplotypes compiled in immunogenetics laboratories during the evaluation of transplant candidates and related potential donors were analysed. Haplotypes were determined using the pedigree analysis tool publicly available from the National Marrow Donor Program (NMDP) website. Nineteen laboratories from 10 countries (11 laboratories from North America, five from Asia, two from Latin America and one from Australia) contributed data on a total of 1719 families comprised of 7474 individuals. We identified 10 393 HLA haplotypes, of which 1682 haplotypes included high-resolution typing at HLA-A, B, C, DRB1 and DQB1 loci. We also present haplotypes containing MICA and other HLA loci and haplotypes containing rare alleles seen in these families. The project will be extended through the 17th IHIW, and investigators interested in joining the project may communicate with the first author.

* Cleveland Clinic, Cleveland, OH, USA, † NMDP Bioinformatics, Minneapolis, MN, USA, ‡ New Zealand Blood Service, Epsom, Auckland, New Zealand, § University of Leiden, Leiden, Netherlands, ¶ Royal Perth Hospital, Perth, Australia, ** HLA Foundation Laboratory, Kyoto, Japan, †† Rush University, Chicago, IL, USA, ‡‡ Genocenter Labs, Damascus,Syria, §§ Northwestern University, Chicago, IL, USA, ¶¶ Children's Hospital of Philadelphia, Philadelphia, PA, USA, *** Hamad Medical Corporation, Doha, Qatar, ††† St. Jude Children's Research Hospital Memphis, TN, USA, ‡‡‡ University of Pennsylvania, Philadelphia, PA, USA, §§§ Sheikh Khalifa Medical City, Abu Dhabi, UAE, ¶¶¶ University of Michigan, Ann Arbor, MI, USA, **** Mayo Clinic, Rochester, MN, USA, †††† Transplant Laboratory of Panama, Panama, Republic of Panama, ‡‡‡‡ Department of Immunology and Immunogenetics, InDRE, Mexico City, Mexico, §§§§ Franciscan St. Francis Health Beech Grove, IN, USA , ¶¶¶¶ Wake Forest University, Winston-Salem, NC, USA, ***** King Abdulaziz Medical City, Riyadh, Saudi Arabia, ††††† Kashi Clinical Laboratories, Portland, OR, USA and ‡‡‡‡‡ Stanford University, Palo Alto,CA, USA

Received 22 October 2012; revised 9 November 2012; accepted 12 November 2012

Correspondence: Medhat Askar, MD, PhD, D. (ABHI), D. (ABMLI), HCLD (ABB), Allogen Laboratories, Cleveland Clinic, 9500 Euclid Avenue, C100, Cleveland, OH 44195, USA. Tel: 216-444-5918;
Fax: 216-444-8261;
E-mail: askarm@ccf.org

Introduction

HLA genes exhibit the highest degree of polymorphism in the human genome (Campbell & Trowsdale, 1993), spanning more than four megabases (Mb) within the short arm of human chromosome six (6p21.3) and contain more than 220 genes of diverse functionality (Horton *et al.*, 2004). Thousands of HLA alleles have been described (http://www.ebi.ac.uk/imgt/hla/) (Robinson *et al.*, 2011). Linkage disequilibrium (LD) between alleles at different HLA loci determines the constellations of alleles that are inherited together, otherwise known as haplotypes (Yunis *et al.*, 2003). Haplotypes, which vary in different racial and ethnic populations, identify islands of nucleotide sequence blocks between which meiotic crossing over occurs in as 'hotspots'. The number of haplotypes observed in populations is much smaller than theoretically expected, indicating that segregation of these alleles at various loci is not random (Yunis *et al.*, 2003).

In this component of the 16th International HLA and Immunogenetics Workshop (IHIW), HLA haplotypes historically compiled in 19 immunogenetics laboratories in the context of evaluating transplant candidates and their related potential donors were retrospectively analysed. These family studies were primarily for hematopoietic stem cell transplantation (HSCT) but also included some solid organ transplants.

There are three popular approaches to determine haplotypes. The first approach includes several molecular methods that allow for the construction of the haplotypes in unrelated individuals, such as allele-

specific polymerase chain reaction (AS-PCR) and somatic cell hybrids (Yan *et al.*, 2000 and Douglas *et al.*, 2001). A second less expensive and less time-consuming approach relies on statistical methods, such as the expectation–maximization (EM) algorithm, for the inference of haplotypes in unrelated individuals from large population genotypic data sets (Niu, 2004). In this project, we pursued the third method of family-based studies, which is an efficient and robust tool to establish phase and determine haplotype segregation. This method is particularly invaluable in populations with no available haplotype frequency data and in admixed populations.

The project aims to (i) investigate extended HLA haplotypes among different populations, (ii) compare the distribution of identified haplotypes among different populations, (iii) assess the frequency of identified haplotypes in major hematopoietic stem cell unrelated donor registries (preliminarily limited to the US National Marrow Donor Program (NMDP) and the Dutch population, data submitted by the Dr. Claas from the University of Leiden) and (iv) initiate an international family studies database and DNA repository to be made publicly available for further characterization of HLA haplotypes, validation of haplotype prediction algorithms and validation of HLA typing technologies.

Methods

Data collection

Data collection sheets were designed and distributed by the first author and included fields for the HLA typing at loci A, C, B, MICA, DRB1, DRB3/4/5, DQA1, DQB1, DPA1 and DPB1. The sheets also included fields for testing laboratories, family study ID, subject ID, relation, ethnicity and country of birth.

Typing method

A wide spectrum of typing methods and strategies were utilized by participating laboratories, including combinations of rSSOP, SSP, DNA sequencing and serologic confirmation of null alleles.

DNA samples

Participating laboratories were encouraged to submit DNA samples from subjects who were representative of the identified haplotypes.

Ethnicities

Ethnicities were grouped according to the data provided by the centres. Ethnicities described by centres include African American, Asian, Caucasian, Filipino, Hispanic, Indian Sub Continent, Mexican Mestizos, Middle Eastern (occurred in data from several centres, including Mexico), North African, Other (listed as 'other' by centres) and Unknown (not listed or listed as 'unknown' by centres). Individuals from Netherlands were designated as Dutch, and individuals from New Zealand were designated as of Maori or Polynesian ancestry.

Bioinformatics tools

Pedigree analysis tool

Haplotypes were identified by segregation using a pedigree analysis algorithm publicly available from The United States NMDP Bioinformatics website (http://bioinformatics.nmdp.org/HLA/Pedigree.aspx). The typings at each locus were compared among a number of related individuals to identify the HLA types that were identical by descent and part of a shared haplotype. Families analysed by this tool included (i) two parents and at least two non-HLA identical children (quartets, $n = 359$ families), (ii) at least one parent and any number of children totalling three family members (trios, $n = 635$) and (iii) other families that did not meet either criteria ($n = 725$). Because some HLA typings for individuals and families included a mix of low-, intermediate- and high-level resolution, the algorithm considered overlaps between sets of possible alleles to ascertain whether typings at a given locus were likely to be inherited by descent. The typings in common between family members make up the inherited haplotype. As this tool is designed to report results with maximum certainty, results included some partial haplotypes. Haplotypes determined by this tool were sorted by locus and by ethnicity using Microsoft Excel® spreadsheets and counted manually.

High-resolution haplotype frequencies

Five loci (HLA-A, C, B, DRB1 and DQB1) haplotypes that were observed in ≥ 5 independent families were examined for haplotype ranks and frequencies in the major ethnic groups in the NMDP registry (European, African American, Hispanic, and Asian/Pacific Islander, http://bioinformatics.nmdp.org/HLA/Haplotype_Frequencies/Haplotype_Frequencies.aspx). Frequencies of these haplotypes were also examined against haplotype frequencies in the Dutch population (based on maximal likelihood).

Rare alleles detector

All identified five loci (HLA-A, C, B, DRB1 and DQB1) haplotypes were examined by the web-based rare alleles detector (RAD) tool publicly available at the Allele Frequency website (http://www.allelefrequencies.net) to detect potential rare alleles (Gonzalez-Galarza *et al.*, 2011). The RAD tool defines 'very rare' alleles as never seen excluding the initial confirmation in IMGT/HLA, 'rare' alleles as seen one, two and three times, excluding initial confirmation in IMGT/HLA, and 'frequent' alleles as seen more than

Table 1. List of contributing centres and their contribution in terms of DNA contributed, families, subjects and identified HLA haplotypes

Centre	Investigator	Designation within text	DNA samples contributed	Families				Haplotypes			Individuals			Ethnicities present
				Trios	Quartets	Other	Total	High Res.	Low Res.	Total	High Res.	Low Res.	Total	
Royal Perth Hospital, Australia	Doran *et al.*	Royal Perth	0	508	52	320	880	0	6043	6043	0	3798	3798	Unknown
Cleveland Clinic Allogen Labs, OH	Askar *et al.*	Cleveland Clinic	899	12	27	140	179	490	291	781	647	253	900	African American, Asian, Caucasian, Filipino, Hispanic, Middle Eastern
HLA Foundation Laboratory, Japan	Saji & Kojima	Japan	612	34	113	8	155	429	382	811	395	211	606	Asian
Rush University, IL	Kanangat	Rush	0	1	2	74	77	92	232	324	86	148	234	African American, Asian, Caucasian, Hispanic, Unknown
Genocenter Labs, Syria	Karoichane	Syria	0	10	53	7	70	262	191	453	265	145	410	Syrian
Northwestern University, IL	Tambur	Northwestern	0	0	0	68	68	0	443	443	0	315	315	African American, Caucasian, Filipino, Hispanic, Indian Sub Continent, Unknown
Children's Hospital of Philadelphia (CHOP), PA	Monos & Lind	CHOP	0	24	24	18	66	176	86	262	184	60	244	African American, Asian, Caucasian, Hispanic, Middle Eastern, Other
Hamad Medical Corporation, Qatar	El-Khalifa & Al-Shaibi	Qatar	70	8	30	17	55	0	313	313	0	249	249	Asian, African, Middle Eastern, North African
St. Jude Children's Research Hospital, TN	Turner	St. Jude	0	14	26	0	40	60	210	270	56	110	166	Hispanic
University of Pennsylvania, PA	Kamoun	Pennsylvania	0	0	4	27	31	0	151	151	0	180	180	African American, Asian, Caucasian, Unknown
Sheikh Khalifa Medical City, UAE	Mustafa	SKMC	122	8	11	7	26	0	150	150	0	92	92	Middle Eastern
University of Michigan, MI	Ramon & Schall	Michigan	0	7	7	9	23	0	140	140	0	88	88	African American, Caucasian
Mayo Clinic, AZ	Gandhi & De Goey	Mayo	0	2	2	6	10	4	41	45	4	24	28	African American, Caucasian
Transplant Laboratory of Panama, Panama	Vernaza	Panama	31	2	4	3	9	20	23	43	15	16	31	Hispanic
Department of Immunology and Immunogenetics (InDRE), Mexico	Gorodezky *et al.*	Mexico	0	2	2	5	9	18	39	57	16	28	44	Caucasian, Mexican Mestizos, Middle Eastern
Franciscan St. Francis Health, IN	Wagenknecht	St. Francis	0	0	0	8	8	0	40	40	0	37	37	Caucasian
Wake Forest University, NC	Gautreaux & Kiger	Wake Forest	0	3	2	0	5	16	9	25	14	5	19	African American, Hispanic
King Abdulaziz Medical City, Saudi Arabia	Hajeer	KAMC	20	0	0	5	5	0	28	28	0	20	20	Middle Eastern
Kashi Clinical Labs, OR	Kashi	Kashi	0	0	0	3	3	0	14	14	0	13	13	Caucasian, Hispanic
Totals			1754	635	359	725	1719	1567	8826	10 393	1682	5792	7474	

Data from New Zealand Blood Services (designated as New Zealand in text) and the University of Leiden (designated as Netherlands in text) are not included. Dr. Dunn of New Zealand Blood services contributed 120 DNA samples. For individuals and haplotypes, 'high res.' refers to a haplotype that was high resolution on A, B, C, DRB1 and DQB1, and 'low res.' refers to a haplotype that contained at least one loci with incomplete or low-resolution data on the same five loci.

Table 2. Identified 5-Loci high-resolution HLA haplotypes and their distribution in major US populations and in the Dutch population

A	B	C	DRB1	DQB1	Number of Families	Centres	Ethnicities	European Frequency	European Rank	African American Frequency	African American Rank	Asian/ Pacific Islander Frequency	Asian/ Pacific Islander Rank	Hispanic Frequency	Hispanic Rank	Dutch Rank	Dutch Frequency
01:01	08:01	07:01	03:01	02:01	31	CHOP, Cleveland Clinic, Rush, St. Jude	Caucasian, Hispanic	0.07408	1	0.01391	2	0.00089	313	0.01784	1	1	0.1052
24:02	52:01	12:02	15:02	06:01	29	Japan	Asian	0.00000	NA	0.00000	NA	0.00977	7	0.00000	NA		
03:01	07:02	07:02	15:01	06:02	23	CHOP, Cleveland Clinic, Mayo, Rush, Syria	African American, Caucasian, Syrian	0.03547	2	0.00915	3	0.00000	NA	0.01278	3	2	0.0533
33:03	44:03	14:03	13:02	06:04	19	Japan	Asian	0.00000	NA	0.00000	NA	0.00667	12	0.00000	NA		
24:02	07:02	07:02	01:01	05:01	14	Japan	Asian	0.00133	102	0.00011	1330	0.00530	15	0.00000	NA	269	0.0007
02:07	46:01	01:02	08:03	06:01	11	Japan	Asian	0.00000	NA	0.00000	NA	0.00677	10	0.00000	NA		
02:01	07:02	07:02	15:01	06:02	10	CHOP, Cleveland Clinic, Rush, St. Jude	Caucasian, Hispanic	0.02341	4	0.00305	22	0.00048	368	0.00705	8	3	0.0379
24:02	54:01	01:02	04:05	04:01	8	Japan	Asian	0.00000	NA	0.00000	NA	0.00363	26	0.00000	NA		
02:01	08:01	07:01	03:01	02:01	7	CHOP, Cleveland Clinic, Rush	Caucasian	0.00978	10	0.00167	86	0.00000	NA	0.00283	30	18	0.0063
24:02	59:01	01:02	04:05	04:01	7	Japan	Asian	0.00000	NA	0.00000	NA	0.00444	17	0.00000	NA		
01:01	57:01	06:02	07:01	03:03	6	CHOP, Cleveland Clinic, Rush	Caucasian	0.01273	6	0.00174	70	0.00755	9	0.00327	27	8	0.0121
02:01	15:01	03:04	04:01	03:02	6	CHOP, Cleveland Clinic, Rush	African American, Caucasian	0.01216	8	0.00087	205	0.00000	NA	0.00187	58	5	0.0255
02:01	44:02	05:01	04:01	03:01	6	Cleveland Clinic	Caucasian	0.02436	3	0.00430	11	0.00000	NA	0.00266	38	9	0.0112
03:01	35:01	04:01	01:01	05:01	6	CHOP, Cleveland Clinic	Caucasian	0.01259	7	0.00174	75	0.00178	69	0.00322	28	4	0.0289
11:01	15:01	04:01	04:06	03:02	6	Japan	Asian	0.00000	NA	0.00000	NA	0.00348	28	0.00000	NA		
02:01	18:01	05:01	03:01	02:01	5	CHOP, Cleveland Clinic, Mexico, St. Jude	Caucasian, Hispanic, Mexican Mestizos	0.00227	51	0.00044	718	0.00000	NA	0.00187	60	82	0.0019
02:01	40:01	03:04	13:02	06:04	5	Cleveland Clinic, Rush	Caucasian	0.01003	9	0.00218	53	0.00000	NA	0.00047	1002	7	0.0171
24:02	07:02	07:02	15:01	06:02	5	CHOP, Cleveland Clinic, Japan	Asian, Caucasian	0.00795	13	0.00131	131	0.00178	65	0.00106	157	10	0.0109

three times, excluding the initial confirmation in IMGT/HLA. The high-resolution HLA typings of unrelated individuals of Maori and Polynesian ancestry (contributed by Dr. Dunn, New Zealand Blood Services) were also examined by RAD.

Results and discussion

Participating centres

Nineteen laboratories from 10 countries (11 laboratories from North America, five from Asia, two from Latin America and one from Australia) contributed data on a total of 1719 families comprised of 7474 individuals. Table 1 summarizes the data of participating centres, families, individuals, identified haplotypes and abbreviations of populations used throughout this text. It is of note that all families contributed from Japan were designated as Asian, from St. Jude were designated as Hispanic, from SKMC and KAMC were designated as Middle Eastern and 42 families from Qatar were designated as Middle Eastern. In addition, Netherlands contributed 295 haplotypes identified by maximum likelihood in the Dutch population that were used for comparison with pedigree identified haplotypes. New Zealand contributed high-resolution typings of 157 unrelated individuals of Maori and Polynesian ancestry that were used for comparison to rare alleles observed in the analysed families. DNA samples were contributed for further characterization of extended HLA haplotypes by the centres listed in Table 1.

Identified five loci haplotypes

The number of haplotypes identified using the pedigree analysis tool is 10 393 HLA haplotypes (Table 1). Of those haplotypes, 1567 haplotypes included high-resolution typing at HLA-A, B, C, DRB1 and DQB1 loci. Table 2 presents the group of these high-resolution five loci haplotypes that appeared in ≥ 5 independent families ($n = 18$). This table also presents the ranks of these haplotypes in the four major US populations (Source: NMDP website, accessed on Oct 20, 2012) and in the Dutch population. It is worth mentioning that the frequently observed haplotypes are skewed towards Asian populations because centres from this continent, particularly Japan, contributed about 25% of the high-resolution data submitted to the project. Figure 1 illustrates the distribution of ethnicity among high-resolution haplotypes.

Other loci containing haplotypes

We also identified haplotypes containing MICA ($n = 625$), DRB3 ($n = 290$), DRB4 ($n = 192$), DRB5 ($n = 122$), DQA1 ($n = 118$), DPA1 ($n = 17$) and DPB1 ($n = 771$) in the analysed families. Table 3 presents the most frequently observed haplotypes containing loci MICA, DRB3, DRB4, DRB5, DQA1 and DPB1. All haplotypes with complete A, B, C, DRB1 and DQB1 typings that also contained DPA1 were observed only once (data not shown).

Partial haplotype associations (blocks)

Most of the MICA data came from the Japanese group, and the most commonly observed B-MICA association were B*52:01-MICA*009 (65 Japanese families), B*51:01-MICA*009 (49 Japanese and one Hispanic), B*35:01-MICA*002 (49 Japanese), B*15:01-MICA*010 (48 Japanese), B*40:02-MICA*027 (44 Japanese), B*44:03-MICA*004 (41 Japanese and one Hispanic) and B*40:06-MICA*027

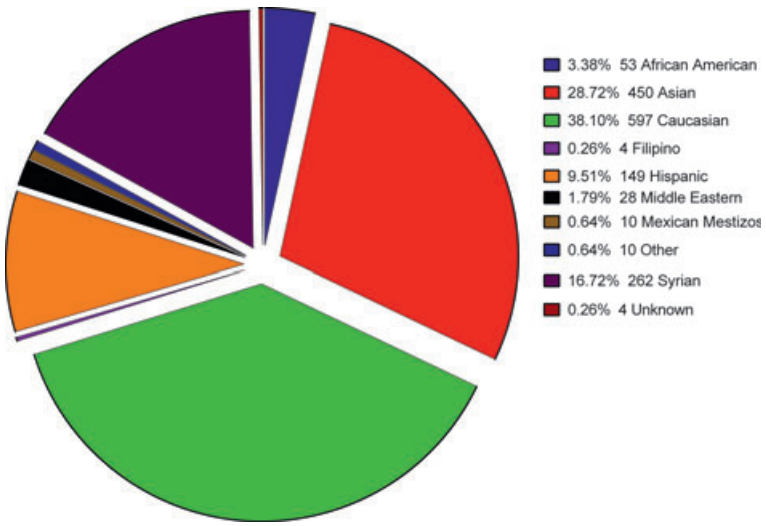

Figure 1. Distribution of identified high-resolution HLA haplotypes by ethnicity ($n = 1567$). Percentage of high-resolution haplotypes for each ethnicity shown.

Table 3. High-resolution haplotypes containing MICA, HLA-DRB3, DRB4, DRB5, DQA1 and DPB1 alleles

Extra locus	A	C	B	MICA	DRB1	DRB3	DRB4	DRB5	DQA1	DQB1	DPA1	DPB1	Subjects	Ethnicities	Testing laboratory
MICA	24:02	12:02	52:01	**009**	15:01	–	–	–	–	06:01	–	–	43	Japanese	Japan
	33:03	14:03	44:03	**004**	13:02	–	–	–	–	06:04	–	–	25	Japanese	Japan
	24:02	07:02	07:02	**008**	01:01	–	–	–	–	05:01	–	–	18	Japanese	Japan
DRB3	01:01	07:01	08:01	007	03:01	**01:01**	–	–	–	02:01	–	–	35	African American, Caucasian	Cleveland Clinic, Mayo, Rush
	24:02	04:01	35:02	–	11:04	**02:02**	–	–	–	03:01	–	–	6	Caucasian, Hispanic	Cleveland Clinic, Rush
	02:01	07:01	08:01	–	03:01	**01:01**	–	–	–	02:01	–	–	5	Caucasian	Cleveland Clinic, Rush
DRB4	02:01	03:04	15:01	–	04:01	–	**01:03**	–	–	03:02	–	–	4	Caucasian	Cleveland Clinic
	29:01	16:01	44:03	–	07:01	–	**01:01**	–	–	02:02	–	–	4	Caucasian, Hispanic	Cleveland Clinic, Rush
	23:01	04:01	44:03	–	07:01	–	**01:01**	–	–	02:01	–	–	4	Caucasian	Cleveland Clinic, Rush
DRB5	03:01	07:02	07:02	–	15:01	–	–	**01:01**	–	06:02	–	–	14	Caucasian	Cleveland Clinic, Rush
	02:01	07:02	07:02	–	15:01	–	–	**01:01**	–	06:02	–	–	9	Caucasian	Cleveland Clinic, Rush
	02:01	05:01	44:02	–	15:01	–	–	**01:01**	–	06:02	–	–	4	Caucasian	Cleveland Clinic, Rush
DQA1	01:01	07:01	08:01	–	03:01	–	–	–	**05:01**	02:01	–	–	14	Caucasian, Hispanic	Cleveland Clinic, Mayo, Rush
	02:01	04:01	35:01	–	01:01	–	–	–	**01:01**	05:01	–	–	3	African American, Caucasian	Cleveland Clinic, Rush
	03:01	04:01	35:01	–	01:01	–	–	–	**01:01**	05:01	–	–	3	Caucasian	Cleveland Clinic, Rush
	24:02	04:01	35:01	–	01:01	–	–	–	**01:01**	05:01	–	–	3	Caucasian	Mayo, Rush
	02:01	07:02	07:02	–	15:01	–	–	–	**01:02**	06:02	–	–	3	Caucasian	Cleveland Clinic, Rush
	03:01	07:02	07:02	–	15:01	–	–	–	**01:02**	06:02	–	–	3	Caucasian	Rush
	23:01	04:01	44:03	–	07:01	–	–	–	**02:01**	02:02	–	–	3	Caucasian, Hispanic	Rush, Wake Forest
	29:01	16:01	44:03	–	07:01	–	–	–	**02:01**	02:02	–	–	3	Caucasian, Hispanic	Rush
DPB1	24:02	12:02	52:01	009	15:02	–	–	–	–	06:01	–	**09:01**	33	Japanese	Japan
	33:03	14:03	44:03	004	13:02	–	–	–	–	06:04	–	**04:01**	17	Japanese	Japan
	24:02	07:02	07:02	008	01:01	–	–	–	–	05:01	–	**04:02**	15	Japanese	Japan

Ethnicity in this table listed as 'Japanese' refers to haplotypes found only in individuals from Japan; this ethnicity is referred to as Asian otherwise. Alleles of the locus of interest are in bold.

Table 4. High-resolution haplotypes containing alleles designated as rare or very rare by the rare allele detection tool

	Ethnicity	A	B	C	DRB1	DQB1
Rare	Caucasian	**02:43N**	15:01	01:02	15:01	06:02
Rare	Caucasian	**02:43N**	08:01	07:01	03:01	02:01
Rare	Caucasian	**02:43N**	40:02	02:02	03:01	02:01
Rare	Caucasian	**02:43N**	07:02	07:02	11:01	03:01
Rare	Syrian	**02:65**	18:01	07:01	11:04	03:01
Rare	Caucasian	**23:07N**	44:03	04:09N	07:01	02:01
Rare	Syrian	**23:15**	41:01	08:01	13:01	03:01
Rare	Syrian	**26:19**	52:01	17:01	15:02	05:01
Rare	Syrian	**26:19**	52:01	17:01	15:02	06:01
Rare	Syrian	24:02	**35:87**	16:01	11:01	03:02
Rare	Syrian	24:02	**35:87**	16:01	11:04	03:01
Rare	Syrian	02:01	**40:88**	07:01	10:01	05:02
Rare	Syrian	02:01	**40:88**	07:01	10:01	05:01
Rare	Syrian	11:01	52:01	**12:08**	09:01	03:02
Rare	Syrian	32:01	35:01	04:01	**11:60**	03:01
Rare	Syrian	03:01	41:01	12:04	**13:35**	05:01
Rare	Caucasian	68:01	18:01	07:01	04:07	**03:13**
Very rare	Syrian	**02:65**	18:01	07:01	11:04	03:01
Very rare	Mexican Mestizos	**02:366**	39:06	07:02	14:06	03:01
Very rare	Syrian	**23:15**	41:01	08:01	13:01	03:01
Very rare	Hispanic	24:02	**35:67**	01:02	04:01	03:02
Very rare	Syrian	24:02	**35:87**	16:01	11:04	03:01
Very rare	Syrian	02:05	**40:88**	06:01	07:02	02:01
Very rare	Syrian	32:01	35:01	04:01	**11:60**	03:01
Very rare	Syrian	03:01	41:01	12:04	**13:35**	05:01

Ethnicity in this table listed as 'Syrian' refers to haplotypes found only in individuals from Syria; this ethnicity is referred to as Middle Eastern otherwise. Rare alleles are in bold. Each haplotype was only seen in one family.

(32 Japanese). Interestingly, the Class II haplotype DRB1*04:07, DRB4*01:01, DQA1*03:01, DQB1*03:02, DPA1*01:03 and DPB1*04:02 was observed in two families tested in different parts of the world (Mexico and Wake Forest). Both of the two families were identified as Hispanic. The same DRB1-DRB4-DQB1-DPB1 haplotype was also seen in four additional Hispanic families tested in the same two laboratories with missing DRB4 and/or DQA1 information.

Very rare and rare alleles observed in the project families

Table 4 shows the alleles designated as 'rare' and 'very rare'. Each allele is shown within the haplotype it was found within. Each haplotype was only seen in one family. Examination of the New Zealand data designated DPA1*01:05 in one Polynesian individual with admixed history as a 'rare' allele. That allele was not seen in any analysed other family. Table 4 presents the haplotypes containing these alleles designate by RAD as 'rare' and 'very rare' alleles.

Results from this project are limited by the skewed representation of populations from the data that were submitted by the participating laboratories. Additionally, recording and calculation of ethnic identification in less homogenous populations may also be a potential source of inaccuracies.

This project aimed to generate pedigree-determined extended HLA haplotypes in subjects of several different ethnicities and to compare the global distribution of these haplotypes. These data are invaluable to facilitate rapid identification of matched donors for HSCT by inferring unknown phase and allele assignment (Gourraud *et al.*, 2005). HLA haplotype analysis has also been used for disease associations and used for a better understanding of many other processes, such as anthropology genetics, due to its relation with immune response and its high degree of polymorphism (Crawford & Nickerson, 2005).

Future directions

This report summarizes the initial analyses of data collected during the first phase of the project. The project will be extended through the 17th IHIW. During the next phases of the project, completion of HLA typing at all loci of haplotypes of interest will be conducted and a select group of haplotypes will be tested by next-generation sequencing (NGS) for characterization of the whole MHC region. Concurrently, the project will continue to accept family data and DNA samples: investigators interested in joining the project should communicate directly with the first author.

Acknowledgements

The authors acknowledge the support of Martin Maiers and Abeer Madbouly from the NMDP Bioinformatics group (USA) in utilizing the NMDP bioinformatics tools. We also thank Aiwen Zhang, Patrick Reville, John McMichael and Stanley Mistak from the Cleveland Clinic (USA) for their support with data organization. We also acknowledge the following investigators for participating in the project: Campbell Witt and Frank Christiansen from Royal Perth Hospital (Australia); Edinur Atan and Geoff Chambers from Victoria University (New Zealand); Hiroto Kojima from the HLA Foundation Laboratory (Japan); Yasuo Morishima from Aichi Cancer Center Research Institute (Japan); Curt Lind from Children's Hospital of Philadelphia (USA); Nisreen Al-Shaibi from Hamad Medical Corporation (Qatar); Cynthia Schall from the University of Michigan (USA); Steve De Goey from Mayo Clinic (USA); Andrea Munguía, Hilario Flores and Carmen Alaez from InDRE (Mexico); Sylvia Piggot and Ina Kurbegovic from Rush University (USA); David Kiger from Wake Forest University (USA);and Marion Alvares and Anitha Santhosh from Sheikh Khalifa Medical City (UAE).

References

Campbell, R.D. & Trowsdale, J. (1993) Map of the human MHC. *Immunology Today*, **14**, 349.

Crawford, D.C. & Nickerson, D.A. (2005) Definition and clinical importance of haplotypes. *Annual Review of Medicine*, **56**, 303.

Douglas, J.A., Boehnke, M., Gillanders, E., Trent, J.M. & Gruber, S.B. (2001) Experimentally-derived haplotypes substantially increase the efficiency of linkage disequilibrium studies. *Nature Genetics*, **28**, 361.

Gonzalez-Galarza, F.F., Christmas, S., Middleton, D. & Jones, A. R. (2011) Allele frequency net: a database and online repository for immune gene frequencies in worldwide populations. *Nucleic Acids Research*, **39**, D913.

Gourraud, P.A., Lamiraux, P., El-Kadhi, N., Raffoux, C. & Cambon-Thomsen, A. (2005) Inferred HLA haplotype information for donors from hematopoietic stem cells donor registries. *Human Immunology*, **66**, 563.

Horton, R., Wilming, L., Rand, V., Lovering, R.C., Bruford, E. A., Khodiyar, V.K. *et al.* (2004) Gene map of the extended human MHC. *Nature Reviews Genetics*, **5**, 889.

Niu, T. (2004) Algorithms for inferring haplotypes. *Genetic Epidemiology*, **27**, 334.

Robinson, J., Mistry, K., McWilliam, H., Lopez, R., Parham, P. & Marsh, S.G. (2011) The IMGT/HLA database. *Nucleic Acids Research*, **39**, D1171.

Yan, H., Papadopoulos, N., Marra, G., Perrera, C., Jiricny, J., Boland, C.R. *et al.* (2000) Conversion of diploidy to haploidy. *Nature*, **403**, 723.

Yunis, E.J., Larsen, C.E., Fernandez-Vina, M., Awdeh, Z.L., Romero, T., Hansen, J.A. & Alper, C.A. (2003) Inheritable variable sizes of DNA stretches in the human MHC: conserved extended haplotypes and their fragments or blocks. *Tissue Antigens*, **62**, 1.