Immunogenetics (2018) 70:381–399
https://doi.org/10.1007/s00251-017-1042-2

ORIGINAL ARTICLE

Novel full-length major histocompatibility complex class I allele discovery and haplotype definition in pig-tailed macaques

Matthew R. Semler[1] · Roger W. Wiseman[1] · Julie A. Karl[1] · Michael E. Graham[1] · Samantha M. Gieger[1] · David H. O'Connor[1,2]

Received: 8 September 2017 / Accepted: 30 October 2017 / Published online: 13 November 2017
© Springer-Verlag GmbH Germany, part of Springer Nature 2017

Abstract Pig-tailed macaques (*Macaca nemestrina*, *Mane*) are important models for human immunodeficiency virus (HIV) studies. Their infectability with minimally modified HIV makes them a uniquely valuable animal model to mimic human infection with HIV and progression to acquired immunodeficiency syndrome (AIDS). However, variation in the pig-tailed macaque major histocompatibility complex (MHC) and the impact of individual transcripts on the pathogenesis of HIV and other infectious diseases is understudied compared to that of rhesus and cynomolgus macaques. In this study, we used Pacific Biosciences single-molecule real-time circular consensus sequencing to describe full-length MHC class I (MHC-I) transcripts for 194 pig-tailed macaques from three breeding centers. We then used the full-length sequences to infer *Mane-A* and *Mane-B* haplotypes containing groups of MHC-I transcripts that co-segregate due to physical linkage. In total, we characterized full-length open reading frames (ORFs) for 313 *Mane-A*, *Mane-B*, and *Mane-I* sequences that defined 86 *Mane-A* and 106 *Mane-B* MHC-I haplotypes. Pacific Biosciences technology allows us to resolve these *Mane-A* and *Mane-B* haplotypes to the level of synonymous allelic variants. The newly defined haplotypes and transcript sequences containing full-length ORFs provide an important resource for infectious disease researchers as certain MHC haplotypes have been shown to provide exceptional control of simian immunodeficiency virus (SIV) replication and prevention of AIDS-like disease in nonhuman primates. The increased allelic resolution provided by Pacific Biosciences sequencing also benefits transplant research by allowing researchers to more specifically match haplotypes between donors and recipients to the level of nonsynonymous allelic variation, thus reducing the risk of graft-versus-host disease.

Keywords *Macaca nemestrina* · Pig-tailed macaques · MHC class I · Allele discovery · Haplotype definition · SIV/HIV

Introduction

Nonhuman primates (NHPs) are of particular importance for their use as animal models for the study of human immunodeficiency virus (HIV). NHPs infected with simian immunodeficiency virus (SIV) mount similar immune responses as humans infected with HIV (Baroncelli et al. 2008; Gardner and Luciw 2008; Joag et al. 1997). Major histocompatibility complex (MHC) class I proteins play a particularly important role in HIV/SIV immune containment by presenting peptides on the surface of virally infected cells to CD8+ killer T cells (Bontrop 2006). It is therefore important to understand the diverse MHC class I (MHC-I) transcript sequences within the population of NHPs that differ in their peptide-binding specificity. This can be difficult because many commonly used NHPs, including all macaques, have a complex MHC-I region that can contain over 20 MHC-I genes on each chromosome (Daza-Vamenta et al. 2004; Wiseman et al. 2013).

For many years, rhesus macaques (*Macaca mulatta*, *Mamu*) have been the preferred NHP model for HIV vaccine development due in part to their well-characterized MHC-I genes. However, the supply of rhesus macaques is becoming

Electronic supplementary material The online version of this article (https://doi.org/10.1007/s00251-017-1042-2) contains supplementary material, which is available to authorized users.

✉ David H. O'Connor
doconnor@primate.wisc.edu

[1] Department of Pathology and Laboratory Medicine, University of Wisconsin-Madison, 555 Science Drive, Madison, WI 53705, USA

[2] Wisconsin National Primate Research Center, University of Wisconsin-Madison, Madison, WI 53715, USA

constrained due to their popularity (Baroncelli et al. 2008). Cynomolgus macaques (*Macaca fascicularis*, *Mafa*) are also widely used for SIV research, especially in European countries. Although cynomolgus macaques are an excellent model to study viral replication, they tend to control viremia shortly after initial infection with commonly used SIV strains (many of which were initially adapted to rhesus macaques). Frequent spontaneous control makes cynomolgus macaques suboptimal for most models of HIV vaccine development (Baroncelli et al. 2008; Reimann et al. 2005). Pig-tailed macaques (*Macaca nemestrina*, *Mane*) are emerging as an important model for HIV/AIDS research and vaccine development. Pig-tailed macaques, like rhesus and cynomolgus macaques, can be infected with SIV and simian/human immunodeficiency virus (SHIV) while also progressing to simian AIDS (Hatziioannou et al. 2009, 2014; Joag et al. 1997). Moreover, pig-tailed macaques, unlike rhesus or cynomolgus macaques, can also be infected, and become symptomatic, when challenged with minimally modified strains of HIV (Del Prete et al. 2016; Frumkin et al. 1995; Hatziioannou et al. 2009, 2014). Infection with a short transcript version of HIV type 1 (HIV-1 ST) progressed to AIDS-like disease in pig-tailed macaques after depletion of CD8+ cytotoxic T cells (Hatziioannou et al. 2014). Following serial passage, HIV-1 ST acquired the ability to antagonize multiple macaque restriction factors, replicate at substantially high levels, and deplete CD4+ helper T cells in a way similar to human progression to AIDS-like diseases (Hatziioannou et al. 2014).

Additionally, infection of pig-tailed macaques with replication-competent HIV type 2 (HIV-2) led to a decline of CD4+ helper T cells, a characteristic sign of AIDS-like diseases (Baroncelli et al. 2008). This is likely due to the fact that pig-tailed macaques contain a variant of tripartite motif-containing protein 5 alpha (TRIM5α) differing in several amino acids from the versions found in other macaque species. Normal functioning TRIM5α binds to HIV capsid proteins, thus preventing uncoating and successful replication of the virus inside viable host cells (Kirmaier et al. 2010). The pig-tailed macaque TRIM5α variant binds less tightly to these proteins, leading to progression of HIV infection and release of replication-competent virus from host cells (Brennan et al. 2007; Igarashi et al. 2007; Stremlau et al. 2004). Because of this modified protein, HIV-2 infection in pig-tailed macaques has become another important model for studying human infection with HIV and progression to AIDS-like disease (Baroncelli et al. 2008; Hatziioannou et al. 2014).

Even with multiple models to mimic HIV infection, relatively little is known about pig-tailed macaque MHC-I genetics. Previous studies have identified multiple novel haplotypes and allelic variants in different cohorts (Fernandez et al. 2011; O'Leary et al. 2009; Pratt et al. 2006). However, these studies largely used short genotyping amplicons to describe haplotypes and characterize novel MHC alleles—a 195-base pair (bp) amplicon encoding the highly polymorphic region of exon two (Wiseman et al. 2009), a 367-bp amplicon encompassing exons two and three (O'Leary et al. 2009), and a 568-bp amplicon spanning from exon two into exon four (Fernandez et al. 2011). These MHC-I fragments can be useful for genotyping, but are not as informative as full-length MHC-I transcript sequences that can now be easily recovered using Pacific Biosciences (PacBio) circular consensus sequencing (CCS) technology (Karl et al. 2017; Pratt et al. 2006; Westbrook et al. 2015).

In this study, we used PacBio single-molecule real-time (SMRT) CCS technology for full-length allele discovery and haplotype definition in 194 pig-tailed macaques originating from three different institutions. We used primers to amplify the full-length ~ 1.1-kilobase (kb) sequences that encode the MHC-I proteins. Amplicons were created and amplified from complementary DNA (cDNA), thus allowing us to specifically amplify full-length, functional open reading frame (ORF) transcripts. We defined 236 novel MHC-I transcript sequences and extended 87 previously described MHC-I transcript sequences to include full-length ORF sequences. With the addition of these novel sequences, we expanded the known diversity of MHC-I in pig-tailed macaques, including allelic variants of published sequences that have previously been shown to be protective against SIV and other infectious diseases in both pig-tailed and rhesus macaques (Gooneratne et al. 2014; Loffredo et al. 2007, 2009; O'Connor et al. 2003; Pratt et al. 2006; Smith et al. 2005a, b). We also defined 192 high-resolution *Mane* MHC-I haplotypes which are useful in showing the inheritance of parental chromosomes and also reducing the risk of graft-versus-host disease (GvHD) in transplant studies using MHC-identical animals.

Materials and methods

Animal selection for full-length allele discovery

Cellular RNA was obtained from 79 pig-tailed macaques from investigators at Johns Hopkins University (JHU; Baltimore, MD). cDNA was provided from 91 pig-tailed macaques housed at the Washington National Primate Research Center (WaNPRC; Seattle, WA) and 90 pig-tailed macaques from investigators at the University of Melbourne (Melbourne, Victoria, Australia) and the Monash University Animal Research Platform (Melbourne, Victoria, Australia). Since the founding of these breeding colonies, southern (*Macaca nemestrina*) and northern (*Macaca leonina*) pig-tailed macaques have been defined as distinct species rather than subspecies of a single species on the basis of differences in morphology and mitochondrial DNA. The boundary zone of the ranges of *Macaca nemestrina* and *Macaca leonina* has been proposed to lie within southern Thailand (Malaivijitnond et al.

2012). As with recent publications characterizing the population genetics of the WaNPRC and JHU breeding colonies (Kanthaswamy et al. 2012; Zhang et al. 2017), we will refer to the individuals in this study as *Macaca nemestrina* because the vast majority of founders for these breeding colonies are thought to have originated from the geographic range of southern pig-tailed macaques. All animals were cared for according to the regulations and guidelines of the Institutional Care and Use Committee at their respective institutions.

Illumina MiSeq (San Diego, CA, USA) sequencing was performed for all three cohorts as previously described (Karl et al. 2014, 2017) in order to inform selection of animals for PacBio CCS technology (data not shown). Out of the original 260 animals, 194 were selected for full-length MHC-I sequencing. Of these 194 samples, 63 were from JHU, 43 were from the WaNPRC, and the final 88 samples were from the University of Melbourne and Monash University. Samples were selected for PacBio sequencing based upon whether they appeared to carry a novel haplotype(s) or contained a high percentage of reads not mapping to known sequences according to the Illumina MiSeq genotyping results. In order to minimize the amount of redundancy in novel allele discovery, a subset of individuals was excluded since they only carried MHC haplotypes that were shared with multiple other members of the same breeding center cohort. Animals were divided into eight separate pools depending on their institution of origin. These pools had a range of 15–64 animals per pool and each pool was run on between four and eight SMRT cells depending on the number of animals in each pool.

PCR amplification for PacBio RS II sequencing

We synthesized cDNA from RNA received from the various institutions using Superscript™ III First-Strand Synthesis System for reverse transcription polymerase chain reaction (RT-PCR) (Invitrogen, Carlsbad, CA, USA) on the samples selected for full-length sequencing. cDNAs for full-length MHC-I sequencing were amplified using a combination of two forward and three reverse primers that annealed to the 5′ and 3′ UTRs, respectively. Each amplified product also contained a unique PacBio 16-bp barcode (Menlo Park, CA, USA) that was fused to the 5′ end of the sequence-specific oligos (Supplemental Figure 1). Amplification was performed on an Applied Biosystems Veriti™ Thermal Cycler (ThermoFischer Scientific, Foster City, CA, USA) under the following conditions: initial denaturation at 98 °C for 3 min; 25 cycles of 98 °C for 5 s, 60 °C for 10 s, 72 °C for 20 s for amplification; and a final elongation of 72 °C for 5 min before being held at 4 °C until proceeding. The full-length products were confirmed on the FlashGel DNA cassette system (Lonza, Basel, Switzerland). After confirmation on the FlashGel, the full-length products were initially purified using the AMPure XP PCR purification kit (Agencourt Bioscience Corporation, Beverly, MA, USA) at a DNA to bead ratio of 1:1. Quantification was performed on the purified products using the Quant-iT dsDNA HS Assay kit and a Qubit fluorometer (Invitrogen, Carlsbad, CA, USA) using a DNA to buffer ratio of 2:198. The final purification was performed using AMPure PB beads (Agencourt Bioscience Corporation, Beverly, MA, USA) at a DNA to bead ratio of 1:1 and quantified using the Qubit fluorometer following the same protocol described earlier.

SMRTbell libraries were created using the PacBio Amplicon Template preparation protocol for CCS using P6-C4 chemistry where individual molecules are sequenced multiple times in both orientations (www.pacb.com). This protocol was described in detail previously (Karl et al. 2017; Prall et al. 2017). Briefly, the pools of PCR products were repaired and hairpin adapters were incorporated onto the ends of the products using the PacBio DNA Template Prep Kit 2.0. The products were then purified using an AMPure PB beads to DNA ratio of 0.6:1. Quality of the pool was assessed again using the Qubit dsDNA BR assay (Invitrogen, Carlsbad, CA, USA) and the Agilent 2100 bioanalyzer DNA12000 kits (Life Technologies, Madison, WI, USA) following the manufacturer's protocol. Full-length amplicons were sequenced on a PacBio RS II instrument (Menlo Park, CA, USA) as was previously described (Karl et al. 2017; Westbrook et al. 2015). Sequencing was performed by the University of Washington PacBio Sequencing Services core or at the Great Lakes Genomics Center at the University of Wisconsin—Milwaukee.

Analysis of MHC-I full-length sequences

Results of full-length sequences were analyzed as previously described (Karl et al. 2017; Prall et al. 2017). In short, we removed sequenced reads that perfectly matched to previously described full-length pig-tailed macaque alleles that were described using Sanger sequencing along with pyrosequencing. These sequences were available to us in the Immuno Polymorphism Database for the Major Histocompatibility Complex genes of Non-Human Primate (IPD-MHC NHP) (http://www.ebi.ac.uk/ipd/mhc/nhp/index.html). We excluded splice variants and chimeric sequences to leave only valid, full-length novel sequences or full-length extensions of previously described transcript sequences in the IPD. Remaining processed reads were clustered, and clusters containing three or more reads were mapped to closest related previously described sequences. Novel sequences and extensions of previously described alleles were analyzed and validated using Geneious Pro (version 9.0) (Biomatters Limited, Auckland, New Zealand) and Basic Local Alignment Search Tool (BLAST). The BLAST results between the novel and known sequences were compared and the putative novel sequences were given local names with the closest related allele

and the number of nucleotides by which they differed. A genotyping table (Supplemental Figure 2) was generated with the number of identical reads identified per candidate allele in each sample. Novel sequences confirmed by this approach as well as extensions of previously described partial sequences were submitted to IPD-MHC NHP to obtain official transcript nomenclature (Maccari et al. 2017; Robinson et al. 2013).

Results and discussion

Full-length MHC-I allele discovery

Our initial pilot study involved a subset of 16 animals from JHU in which identified *Mane* transcript sequences that perfectly mapped to known sequences were not removed in order to provide proof of concept. In total, we identified 35 previously described sequences, along with 39 extensions of previously reported partial sequences and 70 novel full-length ORF transcript sequences that differed by one or more nucleotides compared to their closest related known sequences. This study demonstrated to us that our analytical methods could successfully identify full-length ORF transcript sequences previously characterized by cDNA cloning and Sanger sequencing or by Roche/454 pyrosequencing. In our subsequent studies, all sequences that perfectly matched to previously described full-length *Mane* sequences were removed using our novel allele discovery pipeline and were not included when describing the total number of transcripts discovered. These known *Mane* transcripts were taken into account when describing new haplotypes or updating sequences associated with previously described haplotypes.

As a result of all sequencing runs, each sample across the three facilities had an average of 1225 reads identified from PacBio sequencing. An average of 15 transcripts were identified for each animal, including novel sequences without formal names at the time. Using PacBio CCS, we characterized 313 full-length novel ORF sequences and extensions of previously known *Mane* MHC-I sequences as shown in Table 1, as well as the 35 previously described alleles from our initial pilot study. These sequences were made publically available in IPD and GenBank under the accession numbers associated with each allele. Of these 313 newly characterized transcript sequences, 116 were *Mane-A* sequences, 161 were *Mane-B* sequences, and 36 were *Mane-I* sequences. This distribution matches what we have seen in previous studies of pig-tailed macaques as *Mane-B* sequences are more numerous than either *Mane-A* or *Mane-I* sequences (O'Leary et al. 2009). Of these sequences, 241 were completely novel sequences and 72 were full-length extensions of previously described alleles. We did not identify any *Mane-E* transcripts because the UTR in *Mane-E* transcripts differs significantly from the UTR sequences of *Mane-A*, *Mane-B*, and *Mane-I* regions. Thus, the primers that we use to amplify *Mane-A*, *Mane-B*, and *Mane-I* sequences fail to amplify *Mane-E* sequences efficiently.

We also discovered two novel variants of *Mane-A1*084:01* and a full-length extension of *Mane-A1*084:03* which was previously described (Fernandez et al. 2011). *Mane-A1*084*, previously named *Mane-A*10*, has been associated with delayed progression of SIV (Pratt et al. 2006; Smith et al. 2005a, b) and is therefore important to investigators studying SIV. The additional variants of *Mane-A1*084:01* may provide more insight to interactions among MHC-I and HIV proteins for specific vaccine design. A single novel variant of *Mane-B*008:01*, and one novel variant and one full-length extension of *Mane-B*017:01* were also discovered as a result of this study. In rhesus macaques, *Mamu-B*008:01* and *Mamu-B*017:01:01* transcript sequences have been associated with exceptional control of SIV replication during the chronic phase of infection (Gooneratne et al. 2014; Loffredo et al. 2007; Martins et al. 2015; O'Connor et al. 2003; Yant et al. 2006) and further studies are necessary to determine if these newly characterized allelic variants show the same protective effect as their *Mamu* counterparts.

We also compared the similarity of our newly characterized alleles to previously described alleles from rhesus and cynomolgus macaques as there is data to suggest commonly shared transcript sequences between the three species (Karl et al. 2017). We identified 37 out of 313 full-length ORF transcript sequences that were identical to previously characterized rhesus or cynomolgus nucleotide sequences, with 13 mapping to both a rhesus and cynomolgus sequence (Table 2). These shared sequences likely represent diverse MHC-I alleles that were present in a common ancestor of these three macaque species. With increased feasibility of performing high-throughput full-length ORF sequencing on large macaque cohorts, it is likely that more transcripts will appear to be shared between the same species of macaque in different geographical locations and between different macaque species.

An issue with the PacBio SMRT sequencing technology is the relatively high error rate for single-pass sequencing (http://www.pacb.com/uncategorized/a-closer-look-at-accuracy-in-pacbio/). However, this error rate is mitigated by a number of factors in the sequencing and analysis processes. First, all molecules are circularized for PacBio sequencing. For short molecules which are far below the maximum sequencing length per run, like the MHC-I amplicons studied here, this provides multiple passes over the exact same template molecule. Since the single-pass errors are essentially random, any errors observed in each pass of sequencing a template are distinct from the errors observed in any other sequencing passes. This allows for generation of a consensus sequence of a particular template

Table 1 Full-length *Mane* transcripts characterized by PacBio sequencing

Allele name	Full-length accession no.	Previous accession no.	Animals (*n*)
*Mane-A1*003:03*	LN875390	Novel	1
*Mane-A1*004:01*	LN852023	AY204725	6
*Mane-A1*004:01:02*	LN875391	Novel	2
*Mane-A1*004:02*	LN852024	Novel	4
*Mane-A1*004:03*	LT707728	Novel	1
*Mane-A1*006:03*	LN870187	HQ609862	2
*Mane-A1*006:04*	LN875392	Novel	2
*Mane-A1*006:05*	LN875393	Novel	1
*Mane-A1*009:01*	LN870188	AY204728	6
*Mane-A1*010:02*	LN875394	Novel	9
*Mane-A1*010:03*	LN875395	Novel	5
*Mane-A1*018:01*	LN852027	Novel	1
*Mane-A1*018:02*	LN870199	Novel	2
*Mane-A1*018:03*	LN899592	Novel	8
*Mane-A1*019:03*	LN852028	Novel	8
*Mane-A1*019:04*	LN870200	Novel	3
*Mane-A1*019:04:02*	LN899593	Novel	7
*Mane-A1*019:05*	LN875396	Novel	1
*Mane-A1*019:06*	LN899594	Novel	4
*Mane-A1*019:07*	LT707729	Novel	3
*Mane-A1*031:01*	LN875397	AM295836	13
*Mane-A1*032:02*	LN875398	Novel	1
*Mane-A1*038:01*	LN852029	Novel	4
*Mane-A1*038:02*	LN870201	Novel	9
*Mane-A1*040:01*	LN852030	Novel	12
*Mane-A1*047:01*	LN852031	Novel	2
*Mane-A1*053:01*	LN899595	Novel	2
*Mane-A1*054:01*	LN852025	Novel	2
*Mane-A1*054:02*	LN870198	Novel	5
*Mane-A1*054:04*	LN899596	Novel	20
*Mane-A1*060:01*	LN899597	Novel	6
*Mane-A1*061:01*	LN875400	Novel	6
*Mane-A1*061:02*	LT707730	Novel	1
*Mane-A1*063:01*	LN852033	Novel	2
*Mane-A1*063:02*	LN852034	Novel	2
*Mane-A1*063:02:02*	LN875402	Novel	1
*Mane-A1*063:03*	LN875401	Novel	5
*Mane-A1*069:01*	LN875406	Novel	1
*Mane-A1*072:03*	LN870202	Novel	3
*Mane-A1*072:04*	LN875407	Novel	2
*Mane-A1*074:01*	LN852035	Novel	2
*Mane-A1*074:02*	LN899598	Novel	19
*Mane-A1*082:01:02*	LN852038	Novel	7
*Mane-A1*082:01:03*	LN899599	Novel	32
*Mane-A1*083:01*	LN870189	AY204726	2
*Mane-A1*084:03*	LN852040	HQ609870	17
*Mane-A1*084:05*	LN852042	Novel	8
*Mane-A1*084:06*	LT707731	Novel	1
*Mane-A1*086:01N*	LN852043	Novel	1
*Mane-A1*086:02N*	LN852044	Novel	1
*Mane-A1*086:03N*	LN852045	Novel	1
*Mane-A1*087:01*	LT707713	Novel	3
*Mane-A1*090:01*	LT707732	Novel	1
*Mane-A1*093:01*	LN870203	Novel	4
*Mane-A1*095:01*	LN852026	Novel	4
*Mane-A1*114:02*	LN875409	Novel	1
*Mane-A1*114:03*	LT707733	Novel	2
*Mane-A1*115:01*	LN875410	GQ131745	1
*Mane-A1*116:01*	LN875411	GQ131748	1
*Mane-A1*130:01*	LN875412	Novel	3
*Mane-A1*131:01:01*	LN875403	Novel	1
*Mane-A1*131:01:02*	LN875404	Novel	2
*Mane-A1*131:02*	LN875405	Novel	1
*Mane-A1*132:01*	LN875408	Novel	1
*Mane-A2*01:01*	LN852046	Novel	2
*Mane-A2*05:01:01*	LN852048	EF112583	55
*Mane-A2*05:01:02*	LN870190	EF112578	12
*Mane-A2*05:02*	LN870191	EF112576	13
*Mane-A2*05:05:02*	LN870192	EF394341	4
*Mane-A2*05:08:01*	LN875414	EF112584	5
*Mane-A2*05:09*	LN899600	EF394342	16
*Mane-A2*05:19*	LN852049	Novel	22
*Mane-A2*05:19:02*	LN870206	Novel	3
*Mane-A2*05:23*	LN852050	Novel	2
*Mane-A2*05:24*	LN870204	Novel	2
*Mane-A2*05:25*	LN870205	Novel	2
*Mane-A2*05:26*	LN875413	Novel	3
*Mane-A2*05:27*	LT707714	Novel	7
*Mane-A2*05:28*	LT707734	Novel	1
*Mane-A3*13:01:02*	LN899601	GQ131746	11
*Mane-A3*13:02:01*	LN852053	EF596800	17
*Mane-A3*13:02:02*	LN852054	GQ131747	2
*Mane-A3*13:03*	LN852055	FJ178859	11
*Mane-A3*13:04*	LN852056	GQ131749	2
*Mane-A3*13:05*	LN875415	GQ131750	9
*Mane-A3*13:06*	LN870207	Novel	4
*Mane-A3*13:07*	LN899602	GQ131753	45
*Mane-A3*13:09*	LN852057	Novel	6
*Mane-A3*13:10*	LN852059	Novel	1
*Mane-A3*13:11*	LN870208	Novel	3
*Mane-A3*13:12*	LN875416	Novel	2
*Mane-A3*13:13*	LN899603	Novel	27
*Mane-A3*13:14:02*	LT707715	Novel	3
*Mane-A4*01:01:01*	LN852058	Novel	2
*Mane-A4*14:01:01*	LN875420	Novel	1

Table 1 (continued)

Allele name	Full-length accession no.	Previous accession no.	Animals (*n*)
*Mane-A4*14:01:02*	LN870209	Novel	12
*Mane-A4*14:01:03*	LN875417	Novel	13
*Mane-A4*14:01:04*	LN875418	Novel	13
*Mane-A4*14:03:02*	LN875422	Novel	2
*Mane-A4*14:05*	LN852062	HQ609878	28
*Mane-A4*14:06*	LN852060	Novel	7
*Mane-A4*14:07*	LN852063	Novel	13
*Mane-A4*14:08*	LN852064	GQ131751	12
*Mane-A4*14:09*	LN852065	Novel	2
*Mane-A4*14:10*	LN870210	Novel	2
*Mane-A4*14:11*	LN870211	Novel	6
*Mane-A4*14:12*	LN870212	Novel	2
*Mane-A4*14:13*	LN875419	Novel	3
*Mane-A4*14:14*	LN875421	Novel	1
*Mane-A4*14:15*	LN852061	Novel	4
*Mane-A4*14:16*	LN899604	Novel	4
*Mane-A6*01:01*	LN852066	AY204730	4
*Mane-A6*01:01:02*	LN899605	Novel	22
*Mane-A6*01:03*	LN852067	Novel	10
*Mane-A7*01:01*	LN852068	Novel	6
*Mane-A7*01:02*	LN870213	Novel	1
*Mane-B*004:01*	LN875423	GQ281753	9
*Mane-B*004:02*	LN875424	Novel	29
*Mane-B*007:03*	LN852069	Novel	45
*Mane-B*008:02*	LN875425	Novel	4
*Mane-B*013:01*	LN899606	JN032082	11
*Mane-B*013:02*	LN852070	Novel	31
*Mane-B*014:01N*	LN852071	Novel	3
*Mane-B*014:02:02N*	LN875427	Novel	1
*Mane-B*014:02N*	LN852072	Novel	1
*Mane-B*014:03N*	LN870214	Novel	1
*Mane-B*014:04N*	LN875426	Novel	1
*Mane-B*014:05*	LN899607	Novel	15
*Mane-B*015:01*	LN852073	AY557358	9
*Mane-B*015:04*	LN875428	Novel	1
*Mane-B*016:01*	LN852074	AY557361	4
*Mane-B*016:02*	LN852075	AY204735	9
*Mane-B*017:05*	LN875429	Novel	7
*Mane-B*019:02*	LT707716	Novel	3
*Mane-B*021:01*	LN852077	AY204736	31
*Mane-B*024:02*	LN870215	Novel	3
*Mane-B*025:01*	LN852078	Novel	12
*Mane-B*027:03:02*	LN852079	HQ609888	60
*Mane-B*027:04*	LN852080	Novel	7
*Mane-B*028:01:02*	LN875430	Novel	1
*Mane-B*028:02*	LN852082	Novel	2
*Mane-B*030:06*	LT707735	Novel	2
*Mane-B*031:01*	LN852085	Novel	8

Table 1 (continued)

Allele name	Full-length accession no.	Previous accession no.	Animals (*n*)
*Mane-B*032:01*	LN875432	Novel	1
*Mane-B*032:02*	LT707717	Novel	6
*Mane-B*035:02*	LN875433	Novel	9
*Mane-B*041:02*	LN875434	AY557364	6
*Mane-B*041:03*	LN852086	Novel	10
*Mane-B*041:04*	LN870216	Novel	3
*Mane-B*041:05*	LN875435	Novel	3
*Mane-B*041:06*	LN899608	Novel	11
*Mane-B*043:02*	LT707736	Novel	1
*Mane-B*045:01*	LN875437	AY557359	42
*Mane-B*045:02*	LN852087	GQ153500	21
*Mane-B*045:03*	LN852088	Novel	1
*Mane-B*046:01:01*	LN852089	GQ153469	6
*Mane-B*046:01:02*	LN852090	HQ609894	1
*Mane-B*046:02*	LN870217	Novel	4
*Mane-B*047:02*	LT707742	JN223301	2
*Mane-B*051:03*	LN875438	GQ153487	1
*Mane-B*051:04:05*	LN870220	Novel	2
*Mane-B*051:05*	LN870218	Novel	1
*Mane-B*051:06*	LN870219	Novel	1
*Mane-B*051:07*	LN875439	Novel	1
*Mane-B*051:08*	LT707718	Novel	10
*Mane-B*051:09*	LT707719	Novel	2
*Mane-B*054:01*	LN870193	HQ110955	1
*Mane-B*055:01*	LN875440	Novel	3
*Mane-B*056:02*	LN875441	Novel	4
*Mane-B*056:03*	LN875442	Novel	8
*Mane-B*060:02*	LN875444	HQ609902	5
*Mane-B*060:03:01*	LN870194	HQ992787	16
*Mane-B*060:03:02*	LN875445	Novel	22
*Mane-B*060:05*	LN852094	Novel	6
*Mane-B*060:06*	LT707737	Novel	3
*Mane-B*063:01Sp*	LN870221	Novel	2
*Mane-B*064:02*	LN875446	Novel	1
*Mane-B*065:01:01*	LN875447	Novel	1
*Mane-B*065:01:02*	LN875448	Novel	1
*Mane-B*068:02:01*	LN852097	AY557357	8
*Mane-B*068:02:02*	LN852098	GQ153483	6
*Mane-B*068:02:03*	LN852099	HQ609904	42
*Mane-B*068:05*	LN852100	GQ153510	33
*Mane-B*068:06*	LN852101	HQ609903	94
*Mane-B*068:07:01*	LN875450	HQ609905	28
*Mane-B*068:07:02*	LN875449	Novel	1
*Mane-B*068:08*	LN852102	Novel	19
*Mane-B*068:09*	LN852103	Novel	4
*Mane-B*068:10*	LN870222	Novel	7
*Mane-B*068:11*	LN875451	Novel	7
*Mane-B*068:12*	LN899610	Novel	3

Table 1 (continued)

Allele name	Full-length accession no.	Previous accession no.	Animals (*n*)
*Mane-B*068:13*	LT707720	Novel	2
*Mane-B*069:02*	LN852105	Novel	10
*Mane-B*069:03*	LN875452	Novel	1
*Mane-B*070:01*	LN852106	Novel	24
*Mane-B*072:02:02*	LN852107	HQ609910	12
*Mane-B*072:02:03*	LN875453	Novel	1
*Mane-B*072:02:04*	LN875456	Novel	14
*Mane-B*072:04:04*	LN875457	Novel	2
*Mane-B*072:06*	LN870223	Novel	4
*Mane-B*072:07*	LN875454	Novel	2
*Mane-B*072:08*	LN875455	Novel	1
*Mane-B*072:09*	LN899611	Novel	25
*Mane-B*078:04*	LN875458	Novel	5
*Mane-B*078:05*	LN875459	Novel	1
*Mane-B*078:06*	LT707738	Novel	1
*Mane-B*079:01:02*	LN875460	HM236285	2
*Mane-B*079:04*	LN870226	Novel	6
*Mane-B*079:07*	LN870225	Novel	3
*Mane-B*079:08*	LN870224	Novel	3
*Mane-B*079:09*	LT707721	Novel	7
*Mane-B*080:01*	LN875461	Novel	1
*Mane-B*081:01*	LN852110	Novel	1
*Mane-B*081:01:02*	LN875462	Novel	2
*Mane-B*081:02*	LN899612	Novel	6
*Mane-B*082:04*	LN852112	Novel	9
*Mane-B*082:05*	LN870227	Novel	1
*Mane-B*083:03*	LN852113	Novel	3
*Mane-B*088:02:01*	LN852115	GQ153502	30
*Mane-B*088:02:02*	LN870228	Novel	3
*Mane-B*088:03*	LN875463	GQ153488	1
*Mane-B*088:05*	LN852116	Novel	1
*Mane-B*088:06*	LN899613	Novel	9
*Mane-B*089:03*	LN852118	HQ609921	26
*Mane-B*089:05*	LN875464	JN223303	1
*Mane-B*089:06*	LN852117	Novel	2
*Mane-B*089:07*	LN870229	Novel	3
*Mane-B*098:03*	LN852120	Novel	2
*Mane-B*099:01*	LN875465	JN223300	1
*Mane-B*099:02*	LN852122	Novel	13
*Mane-B*100:01:01*	LN852093	Novel	18
*Mane-B*100:01:02*	LN875443	Novel	1
*Mane-B*101:01*	LN899614	Novel	9
*Mane-B*103:01:02*	LN870230	Novel	3
*Mane-B*103:03*	LN852123	Novel	11
*Mane-B*103:04*	LN875466	Novel	1
*Mane-B*103:05*	LT707739	Novel	1
*Mane-B*104:04*	LN852125	Novel	1
*Mane-B*105:01*	LN870231	Novel	3

Table 1 (continued)

Allele name	Full-length accession no.	Previous accession no.	Animals (*n*)
*Mane-B*105:03*	LN852126	Novel	3
*Mane-B*107:04*	LN852128	Novel	9
*Mane-B*107:05*	LN875467	Novel	5
*Mane-B*108:03*	LN875468	Novel	1
*Mane-B*111:03*	LN852129	Novel	5
*Mane-B*112:02*	LN852131	HQ609930	4
*Mane-B*112:03*	LN870232	Novel	3
*Mane-B*112:04*	LT707722	Novel	5
*Mane-B*112:05*	LT707723	Novel	3
*Mane-B*113:01*	LN852132	FJ875265	2
*Mane-B*115:03*	LN852133	JN032085	3
*Mane-B*115:05*	LN852134	Novel	5
*Mane-B*116:01:02*	LN875469	Novel	1
*Mane-B*116:02*	LN852136	Novel	5
*Mane-B*117:02*	LN852137	Novel	2
*Mane-B*118:01:02*	LN852139	Novel	50
*Mane-B*118:04N*	LN875470	Novel	2
*Mane-B*120:02:02*	LT707740	Novel	1
*Mane-B*123:01:01*	LN852143	AY557355	5
*Mane-B*125:02*	LN875436	Novel	4
*Mane-B*136:01*	LN875431	Novel	2
*Mane-B*142:01*	LT707724	HQ992785	2
*Mane-B*144:01*	LN852145	GQ153464	5
*Mane-B*145:01*	LN899609	Novel	7
*Mane-B*150:01*	LN852146	Novel	2
*Mane-B*150:02*	LN899616	Novel	20
*Mane-B*150:03*	LT707725	Novel	12
*Mane-B*162:01*	LN852147	GQ153507	34
*Mane-B*178:01:01*	LN870195	HQ992781	3
*Mane-B*178:01:02*	LN899617	Novel	15
*Mane-B*178:03*	LN852148	HQ992782	1
*Mane-B*179:01:01*	LN875472	HQ992779	2
*Mane-B*179:01:02*	LN852149	JN223304	13
*Mane-B*189:01*	LN870233	Novel	1
*Mane-B*197:01*	LN852141	Novel	2
*Mane-B*199:01*	LN875471	Novel	3
*Mane-B*199:02*	LT707726	Novel	2
*Mane-B*200:01*	LN899615	Novel	2
*Mane-I*01:01:01*	LN852150	EU429658	15
*Mane-I*01:02*	LN852151	EU429657	2
*Mane-I*01:04:02*	LN875478	Novel	4
*Mane-I*01:06:01*	LN852153	EU429653	6
*Mane-I*01:06:02*	LN852154	Novel	7
*Mane-I*01:06:03*	LN875482	Novel	2
*Mane-I*01:08:01*	LN852156	FJ178864	23
*Mane-I*01:09*	LN875481	FJ178863	2
*Mane-I*01:14*	LN852158	Novel	1
*Mane-I*01:14:02*	LN875477	Novel	1

Table 1 (continued)

Allele name	Full-length accession no.	Previous accession no.	Animals (*n*)
*Mane-I*01:14:03*	LN875487	Novel	1
*Mane-I*01:15*	LN852159	Novel	8
*Mane-I*01:16*	LN852160	Novel	5
*Mane-I*01:17*	LN852161	Novel	2
*Mane-I*01:17:02*	LN875479	Novel	4
*Mane-I*01:18*	LN852162	Novel	11
*Mane-I*01:19*	LN852163	Novel	4
*Mane-I*01:20:02*	LN870197	EU429658	6
*Mane-I*01:21*	LN852165	Novel	6
*Mane-I*01:22*	LN870196	EU429658	9
*Mane-I*01:22:02*	LN899618	Novel	12
*Mane-I*01:23*	LN870234	Novel	4
*Mane-I*01:23:02*	LT707741	Novel	1
*Mane-I*01:24*	LN870235	Novel	38
*Mane-I*01:26*	LN875473	Novel	1
*Mane-I*01:27*	LN875474	Novel	2
*Mane-I*01:28*	LN875475	Novel	1
*Mane-I*01:29*	LN875476	Novel	1
*Mane-I*01:30*	LN875480	Novel	5
*Mane-I*01:31*	LN875483	Novel	2
*Mane-I*01:32*	LN875484	Novel	5
*Mane-I*01:33*	LN875485	Novel	1
*Mane-I*01:34*	LN875486	Novel	1
*Mane-I*01:35*	LN899619	Novel	4
*Mane-I*01:36*	LT707727	Novel	4
*Mane-I*020:01*	LN852164	Novel	4

Summary of the full-length ORF sequences identified in 194 animals. This table gives the official IPD allele nomenclature, GenBank accession number, and the total number of animals in which each sequence was observed. Extensions of previously described partial *Mane* sequences also include accession numbers for the longest partial-length sequence described previously

molecule with an accuracy approaching 100%. Beyond PacBio sequencing, there is some error inevitably introduced from reverse transcription and PCR amplification of template molecules, where single nucleotide base changes and PCR chimeras may get introduced. These errors are reduced by keeping total number of PCR cycles to a minimum and using high-fidelity polymerases. Overall, any erroneous reads that may arise and pass through filtering are removed through careful curation of the sequencing results. All putative novel ORF sequences are added to the database of known sequences, and then all processed sequencing reads for each sample are mapped against that updated database to generate a genotyping table (Supplemental Figure 2). Within this table, artifact sequences are determined by looking at the full sequencing results for each sample. If two putative alleles of the same lineage group are observed within a single sample, read support for each allele is carefully examined. Sequencing artifacts will map in much smaller read numbers compared to legitimate, previously validated alleles. With this in mind, we can confidently conclude that these artifacts are not novel transcripts, but rather errors from the sequencing process.

Haplotype characterization

We used the full-length ORF sequence genotyping results to determine high-resolution haplotypes. The *Mane-A* and *Mane-B* regions were examined independently since these gene clusters are each separated by ~ 1 Mb and recombination events are relatively common on a population basis. Haplotypes were characterized by comparing samples with identical groupings of alleles, particularly animals with known relationships. Pedigree information was available for most of the animals in these cohorts. Of the 194 animals examined, 133 were known to be directly related to at least one other animal sequenced in the PacBio experiments. For those animals that lacked pedigree information, relatedness could be inferred based upon the prevalence and sharing of specific haplotypes. Haplotype designations were assigned first for combinations of alleles observed together in related animals. We then looked across the full cohort to identify all animals expressing these initially defined haplotypes. For any animals with a single haplotype defined after this process, the remaining unassigned alleles were inferred to be on the alternate parental haplotype. To define and name the haplotypes, *Mane-A* and *Mane-B* sequences were first divided roughly into major and minor transcripts—sequences that averaged greater than 4% abundance relative to the total number of sequence reads identified per sample were denoted as major transcripts, and sequences that averaged lower than 2% abundance per sample were denoted as minor transcripts. In order to be uniform across macaque species, transcripts averaging in the intermediate range between 2 and 4% abundance per sample were denoted as major or minor largely in accordance with their designations in rhesus or cynomolgus macaques (Karl et al. 2013, 2017).

Of the major, highly expressed transcripts on each haplotype, one sequence was designated as the "diagnostic" major allele, typically, the most abundant transcript on each haplotype. A few exceptions to this rule were made for sequences with putative biological significance based on studies in other macaque species, e.g., the B017 group of haplotypes. The initial four digits of each haplotype designation were derived from the diagnostic transcript sequence (B004, B008, etc.). Haplotypes with different lineage groups of major alleles traveling with the diagnostic major are denoted with suffixes of lowercase letters. Thus, the B013a haplotype consists of a *Mane-B*013* diagnostic major transcript sequence accompanied by *Mane-B*014*, *Mane-B*041*, and *Mane-B*178*

Table 2 *Mane* transcripts that are identical to previously described *Mamu* or *Mafa* class I sequences

Mane allele	IPD accession no.	Identical *Mamu* and/or *Mafa* sequence(s)
*Mane-A1*004:01:01*	IPD0006235	*Mamu-A1*004:02:01* (IPD0001811—1098 bp)
*Mane-A1*004:01:02*	IPD0007156	*Mamu-A1*004:02:02* (IPD0001693—1098 bp)
*Mane-A1*032:02*	IPD0007162	*Mafa-A1*032:04* (IPD0002384—1003 bp)
*Mane-A2*05:19:01*	IPD0006268	*Mamu-A2*05:04:03* (IPD0001911—1098 bp), *Mafa-A2*05:06:02* (IPD0001218—1098 bp)
*Mane-A2*05:25*	IPD0007191	*Mamu-A2*05:10* (IPD0001871—1098 bp), *Mafa-A2*05:57:01* (IPD0006411—1098 bp)
*Mane-A3*13:04*	IPD0004048	*Mafa-A3*13:17:01* (IPD0006415—822 bp)
*Mane-A4*01:01:01*	IPD0004054	*Mamu-A4*01:02:01* (IPD0001930—862 bp), *Mafa-A4*01:09* (IPD0003394—1074 bp)
*Mane-A4*14:01:03*	IPD0007183	*Mafa-A4*14:01* (IPD0001798—1098 bp), *Mamu-A4*14:04:02* (IPD0006920—692 bp)
*Mane-A4*14:06*	IPD0007050	*Mamu-A4*14:04:02* (IPD0006920—692 bp)
*Mane-A4*14:15*	IPD0007051	*Mamu-A4*14:04:02* (IPD0006920—692 bp)
*Mane-A4*14:16*	IPD0007298	*Mamu-A4*14:04:02* (IPD0006920—692 bp)
*Mane-B*007:03*	IPD0007057	*Mamu-B*007:04:01* (IPD0002243—1089 bp), *Mafa-B*007:08* (IPD0006970—1089 bp)
*Mane-B*013:02*	IPD0007058	*Mamu-B*013:01* (IPD0002244—1089 bp), *Mafa-B*013:18* (IPD0007441—1089 bp)
*Mane-B*014:02:02N*	IPD0007196	*Mamu-B*014:01* (IPD0002245—1089 bp), *Mafa-B*014:01* (IPD0002745—1089 bp)
*Mane-B*027:03:02*	IPD0004174	*Mafa-B*044:06* (IPD0004326—1087 bp), *Mamu-B*027:02* (IPD0002261—1054 bp)
*Mane-B*031:01*	IPD0006352	*Mafa-B*031:03* (IPD0007452—1095 bp)
*Mane-B*045:03*	IPD0007065	*Mamu-B*045:08:01* (IPD0008784—1089 bp)
*Mane-B*046:01:01*	IPD0004143	*Mafa-B*046:01:01* (IPD0002699—1080 bp)
*Mane-B*051:07*	IPD0007212	*Mafa-B*051:04:01* (IPD0003827—425 bp)
*Mane-B*055:01*	IPD0007214	*Mafa-B*055:01* (IPD0005568—659 bp)
*Mane-B*060:02*	IPD0004151	*Mafa-B*060:01* (IPD0001682—1080 bp)
*Mane-B*064:02*	IPD0007221	*Mamu-B*064:01* (IPD0002308—1077 bp), *Mafa-B*064:05* (IPD0007462—1077 bp)
*Mane-B*065:01:01*	IPD0007223	*Mafa-B*065:05* (IPD0007463—1089 bp)
*Mane-B*068:02:03*	IPD0004155	*Mamu-B*068:02:01* (IPD0006301—1089 bp), *Mafa-B*068:06:01* (IPD0006510—1089 bp)
*Mane-B*068:07:01*	IPD0006370	*Mafa-B*068:03* (IPD0002713—1089 bp)
*Mane-B*082:04*	IPD0007072	*Mafa-B*082:01:01* (IPD0003662—1107 bp)
*Mane-B*089:02*	IPD0002914	*Mamu-B*089:01:02* (IPD0004042—822 bp)
*Mane-B*089:03*	IPD0004167	*Mafa-B*089:01:02* (IPD0002039—1116 bp), *Mamu-B*089:01:02* (IPD0004042—822 bp)
*Mane-B*105:01*	IPD0002765	*Mafa-B*105:01* (IPD0004004—1089 bp), *Mamu-B*105:01* (IPD0003536—649 bp)
*Mane-B*105:03*	IPD0007088	*Mamu-B*105:01* (IPD0003536—649 bp)
*Mane-B*136:01*	IPD0007202	*Mafa-B*136:02* (IPD0001651—1089 bp)
*Mane-B*144:01*	IPD0004181	*Mafa-B*144:02* (IPD0002698—1089 bp)
*Mane-I*01:01:01*	IPD0000633	*Mafa-I*01:11:01* (IPD0006272—1089 bp)
*Mane-I*01:20:01*	IPD0007109	*Mafa-I*01:01:01* (IPD0000353—1089 bp)
*Mane-I*01:22:01*	IPD0007176	*Mamu-I*01:06:01* (IPD0001264—1089 bp)
*Mane-I*01:27*	IPD0007246	*Mamu-I*01:29* (IPD0007009—1089 bp), *Mafa-I*01:09* (IPD0002494—1089 bp)
*Mane-I*01:36*	IPD0009089	*Mamu-I*01:30* (IPD0007011—1089 bp)

Summary of 37 *Mane* full-length ORF sequences that are identical to previously described Mamu and/or Mafa nucleotide sequences

transcripts while the B013b haplotype consists of a *Mane-B*013* diagnostic major transcript plus a *Mane-B*007* transcript. Finally, for the high-resolution haplotypes defined here, any sub-haplotypes differing from each other by any allelic variants across the haplotype are assigned a different Roman numeral suffix. For instance, the B013a.i sub-haplotype consists of the major alleles *Mane-B*013:01*, *Mane-B*014:01N*, *Mane-B*041:05*, and *Mane-B*178:01:02* while the B013a.ii sub-haplotype is defined by the major alleles *Mane-B*013:01*, *Mane-B*014:05*, *Mane-B*041:05*, and *Mane-B*178:01:02*. Therefore, this pair of sub-haplotypes differs only in the specific variants of *Mane-B*014* allele that were observed. Most of these major haplotype variants also differed in the minor MHC-I transcripts that were detected to be co-inherited, but that level of distinction was not considered when subdividing haplotypes with letter and/or Roman numeral suffixes. Figure 1a, b illustrates how haplotypes are assigned within related groups of animals.

In total, we characterized 192 *Mane* class I haplotypes; 86 of these were *Mane-A* haplotypes and 106 of these were *Mane-B* haplotypes. The definitions of the *Mane-A* and *Mane-B* haplotypes are shown in Figs. 2 and 3, respectively. The primary PacBio genotyping results for all 194 individuals (Supplemental Figure 2) that were used to define these haplotypes are also available on our public web portal (https://dholk.primate.wisc.edu/project/dho/public/begin.view?) along with lists of major and minor transcripts that are associated with these *Mane-A* and *Mane-B* haplotypes. This distribution is consistent with previous macaque data sets (Karl et al. 2013, 2017) where MHC haplotypes of the B genes are more numerous; presumably, this is due to the additional rounds of complex duplication events that macaque B regions have experienced. Of the *Mane-A* haplotypes, A052a.i was the most common, being observed in 13.4% of the 388 total chromosomes. Likewise, B118d was the most common *Mane-B* haplotype; it was present in 7.5% of the total chromosomes examined. Although phased combinations of *Mane-A* and *Mane-B* haplotypes may be defined when examining animals from a particular breeding facility as illustrated in Fig. 1a, b, this is generally not possible when genotyping experimental cohorts that lack directly related individuals. Recombination between the *Mane-A* and *Mane-B* chromosomal regions has scrambled these haplotype combinations of the population level despite the fact that recombination events between these gene clusters appear to occur at a rate below 1% per meiosis. As an example, the A052a.i haplotype that is the most common within the Australian breeding colony segregates with at least four different *Mane-B* haplotypes including B101, B004a, B024a.ii, and B013a.i in the 79 individuals evaluated in this study (Supplemental Figure 2). The *Mane-A* and *Mane-B* haplotypes reported here can be inferred based on shared allele profiles even in the absence of closely related individuals or detailed pedigree information.

Haplotype diversity among breeding centers

The primary goal of these studies was to maximize the discovery of novel *Mane-A* and *Mane-B* sequence variants and to characterize additional high-resolution *Mane-A* and *Mane-B* haplotypes. With this goal in mind, cDNA from animals at the JHU and WaNPRC breeding centers was prescreened by Illumina MiSeq analyses in order to enrich

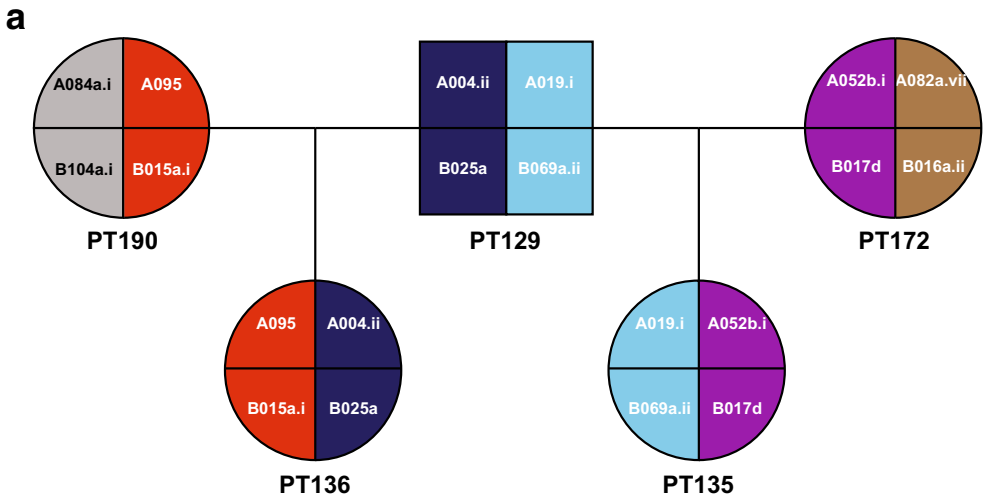

Fig. 1 **a** Sample pedigree information visually showing the segregation of MHC-I haplotypes from sires (PT129) and dams (PT172 and PT190) to their respective progeny (PT135 and PT136). **b** A subsequent representation of the pedigree information showing the passage of the common transcripts seen in each haplotype to the various offspring. The alleles are also represented as percent of the total sequences observed in each animal providing a way to distinguish between major and minor transcript sequences in the haplotype and are color coordinated as to whether they originated from the sire or dam

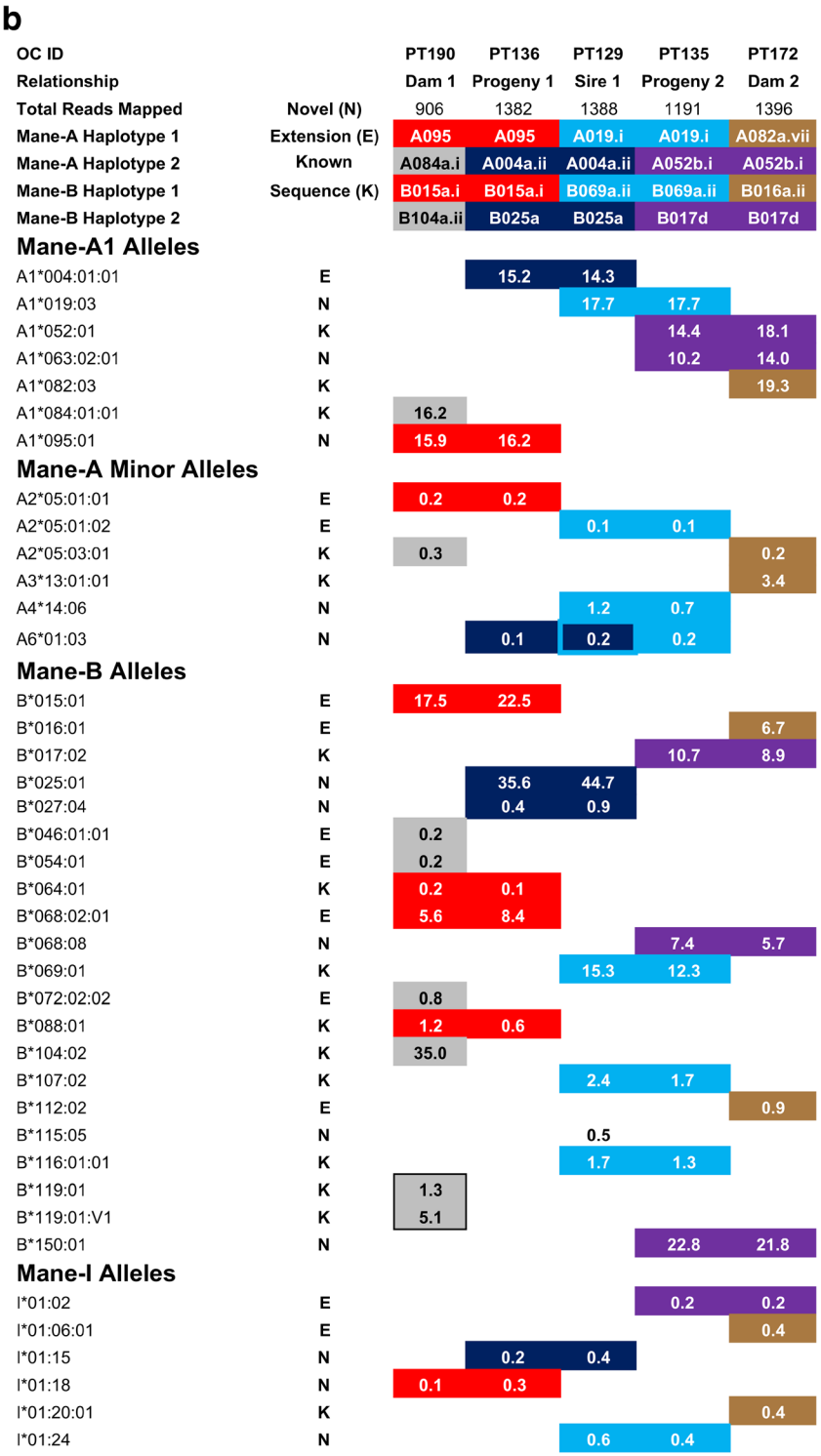

Fig. 1 continued.

these PacBio cohorts with individuals expected to carry novel MHC-I alleles and haplotypes. A subset of MHC-identical animals from common sires and dams in these breeding groups was excluded in order to limit redundancy in the PacBio analyses. In contrast, all cDNA samples from 2 consecutive years of health screening for the pig-tailed macaques in the University of Melbourne and Monash University breeding program were analyzed sequentially without restricting the number of offspring from common sires and dams that were evaluated. Compared to the other

Haplotype	Variant	Major 1	Major 2	Minor 1	Minor 2	Minor 3
A003	i	*A1*003:01*	-	*A2*05:18*	*A3*13:10*	-
A003	ii	*A1*003:03*	-	*A2*05:18*	*A3*13:09*	-
A004a	i	*A1*004:01:01*	-	*A4*14:01:04*	*A6*01:03*	-
A004a	ii	*A1*004:01:02*	-	*A6*01:03*	-	-
A004a	iii	*A1*004:02*	-	*A2*05:01:01*	*A6*01:03*	-
A004b	-	*A1*004:03*	*A1*090:01*	*A4*14:01:02*	-	-
A006	i	*A1*006:02*	-	*A2*05:01:01*	-	-
A006	ii	*A1*006:03*	-	*A2*05:24*	*A4*14:10*	-
A006	iii	*A1*006:04*	-	*A4*14:01:02*	-	-
A006	iv	*A1*006:05*	-	-	-	-
A009	i	*A1*009:01*	-	*A2*05:02*	*A4*14:01:02*	-
A009	ii	*A1*009:01*	-	*A2*05:02*	*A4*14:01:04*	-
A010	i	*A1*010:02*	-	*A2*05:19:01*	*A3*13:05*	-
A010	ii	*A1*010:03*	-	-	-	-
A016	-	*A1*016:nov:01*	-	*A3*13:nov:02*	-	-
A018	-	*A1*018:03*	-	*A2*05:01:01*	-	-
A019	i	*A1*019:03*	-	*A2*05:01:02*	*A4*14:06*	-
A019	ii	*A1*019:04:01*	-	*A4*14:12*	-	-
A019	iii	*A1*019:04:02*	-	*A1*086:02N*	*A3*13:01:02*	-
A019	iv	*A1*019:05*	-	*A4*14:01:03*	-	-
A019	v	*A1*019:06*	-	*A2*05:01:01*	*A4*14:16*	-
A019	vi	*A1*019:07*	-	*A2*05:01:01*	*A3*13:07*	-
A031	-	*A1*031:01*	-	*A4*14:01:04*	-	-
A038	i	*A1*038:01*	-	*A2*05:01:01*	*A4*14:15*	-
A038	ii	*A1*038:02*	-	*A2*05:01:01*	*A3*13:07*	*A4*14:11*
A047	-	*A1*047:01*	*A2*01:01*	*A2*05:25*	*A4*01:01:01*	-
A052a	i	*A1*052:01*	-	*A2*05:01:01*	*A3*13:13*	*A4*14:05*
A052a	ii	*A1*052:01*	-	*A2*05:05:02*	*A4*14:05*	-
A052a	iii	*A1*052:01*	-	*A2*05:19:01*	*A4*14:07*	-
A052a	iv	*A1*052:01*	-	*A4*14:01:03*	-	-
A052b	i	*A1*052:01*	*A1*063:02:01*	*A6*01:03*	-	-
A052b	ii	*A1*052:01*	*A1*063:02:02*	*A6*01:03*	-	-
A053	-	*A1*053:01*	-	*A2*05:26*	*A4*14:01:03*	-
A054a	i	*A1*054:01*	-	*A2*05:01:01*	*A4*14:09*	-
A054a	ii	*A1*054:02*	-	*A2*05:01:02*	*A6*01:02*	-
A054b	-	*A1*054:04*	*A1*074:02*	*A2*05:09*	*A6*01:01:02*	-
A060	-	*A1*060:01*	-	*A4*14:01:03*	-	-
A061	i	*A1*061:01*	-	*A3*13:09*	-	-
A061	ii	*A1*061:01*	-	*A3*13:12*	-	-
A061	iii	*A1*061:02*	-	*A4*14:01:02*	-	-
A063	i	*A1*063:01*	-	*A2*05:08:01*	*A3*13:04*	-
A063	ii	*A1*063:03*	-	*A1*086:03N*	*A2*05:01:01*	*A3*13:02:02*
A063	iii	*A1*063:03*	-	*A3*13:07*	-	-
A063	iv	*A1*063:03*	-	*A3*13:12*	-	-
A072	i	*A1*072:03*	-	*A2*05:19:02*	*A3*13:11*	-
A072	ii	*A1*072:04*	-	*A4*14:03:02*	-	-
A074	-	*A1*074:01*	-	*A2*05:23*	-	-
A082a	i	*A1*082:01:01*	-	*A3*13:01:01*	-	-
A082a	ii	*A1*082:01:02*	-	*A2*05:01:01*	-	-
A082a	iii	*A1*082:01:03*	-	*A2*05:03:01*	*A3*13:01:01*	-
A082a	iv	*A1*082:01:03*	-	*A2*05:27*	*A3*13:07*	-
A082a	v	*A1*082:02*	-	*A2*05:19:01*	*A3*13:01:01*	*A4*14:15*
A082a	vi	*A1*082:03*	-	*A2*05:01:02*	*A3*13:01:01*	-
A082a	vii	*A1*082:03*	-	*A2*05:03:01*	*A3*13:01:01*	-
A082b	-	*A1*082:01:02*	*A1*031:01*	*A2*05:03:01*	*A4*14:01:04*	-
A082c	-	*A1*082:03*	*A1*069:01*	*A2*05:03:01*	*A3*13:01:01*	*A4*14:13*
A083	-	*A1*083:01*	-	-	-	-
A084a	i	*A1*084:01:01*	-	*A2*05:03:01*	*A4*14:08*	-
A084a	ii	*A1*084:01:02*	-	*A2*05:01:02*	*A4*14:08*	-
A084a	iii	*A1*084:01:02*	-	*A2*05:02*	*A4*14:08*	-
A084a	iv	*A1*084:01:02*	-	*A2*05:03:01*	*A4*14:08*	-
A084a	v	*A1*084:01:02*	-	*A2*05:08:01*	*A4*14:08*	-
A084a	vi	*A1*084:03*	-	*A2*05:01:01*	*A3*13:02:01*	*A4*14:01:01*
A084a	vii	*A1*084:04*	-	*A2*05:01:02*	*A3*13:03*	-
A084a	viii	*A1*084:05*	-	*A2*05:01:02*	*A3*13:03*	-
A084a	ix	*A1*084:06*	-	*A2*05:28*	*A4*14:08*	-
A084b	-	*A1*084:01:02*	*A1*003:01*	*A2*05:18*	-	-
A084c	-	*A1*084:01:01*	*A7*01:02*	*A2*05:03:01*	*A3*13:06*	-
A085	-	*A1*085:nov:01*	-	*A2*05:08:01*	*A4*14:11*	-
A087	-	*A1*087:01*	-	*A3*13:14:02*		-
A093	-	*A1*093:01*	-	*A2*05:05:02*	-	-
A095	-	*A1*095:01*	-	*A2*05:01:01*	-	-
A114	i	*A1*114:01*	-	*A2*05:02*	*A4*14:01:02*	-
A114	ii	*A1*114:02*	-	*A2*05:02*	-	-
A114	iii	*A1*114:03*	-	*A3*13:01:01*	-	-
A115	-	*A1*115:01*	-	-	*A4*14:14*	-
A116	-	*A1*116:01*	-	*A2*05:20*	-	-
A130	-	*A1*130:01*	-	*A2*05:26*	-	-
A131a	i	*A1*131:01:01*	-	*A2*05:01:01*	*A4*14:01:04*	-
A131a	ii	*A1*131:01:02*	-	*A2*05:01:01*	*A4*14:01:03*	-
A131b	-	*A1*131:02*	*A1*032:02*	-	*A4*14:01:03*	-
A132	-	*A1*132:01*	-	-	-	-
A701a	i	*A7*01:01*	-	-	-	-
A701a	ii	*A7*01:02*	-	*A2*05:01:01*	*A3*13:06*	-
A701b	-	*A7*01:02*	*A1*018:02*	*A2*05:01:01*	*A3*13:06*	-
A701c	-	*A7*01:01*	*A1*040:01*	*A2*05:01:01*	*A4*14:08*	-

◀ **Fig. 2** A description of the sequences that define each *Mane-A* haplotype. The haplotypes are named for the first major transcript described and contain other major and minor alleles that are passed onto offspring together. Major transcripts were defined as having over 4% of the total reads associated with each animal and minor alleles had less than 2% of the total reads. Sequences with an abundance between 2 and 4% were defined as major or minor in accordance with published alleles from rhesus or cynomolgus macaques (Karl et al. 2013, 2017)

breeding centers, this methodological difference resulted in apparently lower MHC haplotype diversity for animals in the Australian cohort where haplotypes A052a.i and A082a.iv together made up almost half of the *Mane-A* haplotype distribution (Fig. 4a). The distribution of *Mane-B* haplotypes was slightly more diverse, but four haplotypes, B118d, B004a, B101, and B013b.iv, accounted for just over 50% of the total haplotypes observed in this population (Fig. 4b).

The majority of *Mane-A* and *Mane-B* haplotypes defined in this study were specific to a single institution. However, multiple haplotypes were present across more than one cohort, as 16% of the *Mane-A* haplotypes and 14% of the *Mane-B* haplotypes were shared between at least two breeding centers. Most of the shared *Mane-A* and *Mane-B* haplotypes were observed between JHU and WaNPRC which was expected because a subset of the founding pig-tailed macaques for the JHU breeding program was obtained from WaNPRC. In contrast, only six *Mane-A* and *Mane-B* haplotypes observed in individuals from the Australian cohort were also shared with animals evaluated from JHU and WaNPRC. Additional studies will be required in order to determine whether these differences reflect distinct geographic origins for the pig-tailed macaques at each of these breeding centers.

Advantages of using full-length cDNA sequencing over short genotyping amplicons

In understudied populations of macaques and other NHPs, PacBio full-length sequencing provides multiple advantages over use of short amplicons that have been used routinely for MHC genotyping analyses over the past several years (Karl et al. 2013, 2017). For one, pig-tailed macaques and other understudied NHP populations have been historically used less often in infectious disease research compared to rhesus macaques, so the MHC region of these species is generally less well characterized. PacBio full-length sequencing provides an opportunity to not only discover novel allelic variants, but also extend partial sequences to include complete ORFs. It also provides us with improved resolution between related sequences, allowing for alleles to be defined to the level of synonymous variation in the genetic sequence. By defining these transcript sequences with higher specificity, it becomes possible to study whether functional differences arise due to subtle variations in the genetic sequence. Another advantage of defining full-length ORF sequences is in regard to restriction of infectious diseases. Previous studies have identified multiple MHC-I variants that bind to specific epitopes of SIV (Gooneratne et al. 2014). In this study, we identified multiple closely related allelic variants of these MHC-I gene products that have been documented to restrict SIV epitopes. Because of the subtle, but possibly important differences in these newly discovered allelic variants, they may present distinct target epitopes from SIV/HIV and other infectious agents. Further research is needed to characterize immune responses associated with these allelic variants, but their discovery may be an important piece of information for the use of an effective long-term immune response against HIV or novel vaccine development.

Matching MHC-identical animals to prevent GvHD

An additional advantage of using full-length sequencing over shorter amplicons involves transplant research. Full-length allele discovery defines transcript sequences to the level of nonsynonymous and synonymous polymorphisms in all coding regions of the gene. This specificity is of great use in transplant research where having MHC-identical donors and recipients is necessary to reduce the chances of GvHD, in which the host cells recognize the newly transplanted cells as a foreign antigen and develop an immune response (Anasetti et al. 1990; Ayala García et al. 2012). It is also advantageous to match MHC or human leukocyte antigen (HLA) donors and recipients in terms of stem cell transplants. A recent study described the use of induced pluripotent stems cells in the brain to reduce the effects of Parkinson's disease (PD) by regenerating and increasing the survival of dopamine-releasing neurons. The MHC-identical NHPs used in this study showed a reduced immune response from microglia and lymphocytes to the graft (Morizane et al. 2017). By continuing to describe MHC-I transcripts and haplotypes to specific levels of allelic variation, animals can be matched to minor allelic variation and GvHD-like diseases can be reduced in studies. It also provides a means for future studies to be undertaken in the hopes that these resources and therapies can provide a means for treating difficult neurological disorders.

However, with this specific level of definition, a paradoxical situation may arise in which as we identify more allelic variants, it may become difficult to find donors and recipients that have genetically matching MHC. As a counter argument, there may be a difference between being genetically identical and being functionally identical in terms of the host response in GvHD. There is a possibility that, although there are amino acid differences between two variants of the same MHC-I sequence, there may not be significant differences in their functionality and phenotypic appearance to host cells. The discovery of novel transcripts and further haplotype definition

Haplotype	Variant	Major 1	Major 2	Major 3	Major 4	Major 5	Minor 1	Minor 2	Minor 3	Minor 4	Minor 5	Minor 6	Minor 7	Minor 8
B004a	-	B*004:02	B*030:01	B*043:01	B*045:01	-	B*051:08	B*072:03	B*088:01	I*01:18	-	-	-	-
B004b	-	B*004:02	B*065:01:02	B*080:01	-	-	B*051:02:01	I*01:34	I*01:24	-	-	-	-	-
B008	-	B*008:02	B*125:02	-	-	-	B*072:02:02	B*119:01	I*01:24	-	-	-	-	-
B013a	i	B*013:01	B*014:01N	B*041:06	B*178:01:02	-	B*116:01:01	I*01:24	-	-	-	-	-	-
B013a	ii	B*013:01	B*014:05	B*041:06	B*178:01:02	-	B*032:02	B*088:02:01	B*103:01:01	-	-	-	-	-
B013b	i	B*013:02	B*007:03	-	-	-	B*051:01	B*060:02	B*070:01	B*072:03	-	-	-	-
B013b	ii	B*013:02	B*007:03	-	-	-	B*051:01	B*060:05	B*070:01	B*072:03	B*082:04	-	-	-
B013b	iii	B*013:02	B*007:03	-	-	-	B*060:05	B*070:01	B*072:04:02	B*082:04	I*01:16	-	-	-
B013b	iv	B*013:02	B*007:03	-	-	-	B*051:01	B*060:03:02	B*070:01	B*072:09	B*079:09	B*082:02	I*01:36	-
B015a	i	B*015:01	B*068:02:01	-	-	-	B*064:01	B*082:02	B*088:01	I*01:19	I*01:18	-	-	-
B015a	ii	B*015:01	B*068:02:02	-	-	-	B*064:02	I*01:19	I*01:18	-	-	-	-	-
B015a	iii	B*015:01	B*068:07:02	-	-	-	B*072:02:03	I*01:28	-	-	-	-	-	-
B015a	iv	B*015:04	B*068:02:03	-	-	-	-	-	-	-	-	-	-	-
B015b	i	B*015:02	B*030:02	B*068:07:01	-	-	B*035:02	B*051:09	B*072:04:02	B*088:02:01	B*112:04	I*01:01:01	-	-
B015b	ii	B*015:02	B*030:02	B*068:07:01	-	-	B*082:04	I*01:26	-	-	-	-	-	-
B015b	iii	B*015:02	B*030:06	B*068:07:01	-	-	B*035:02	B*088:02:01	-	-	-	-	-	-
B015c	-	B*015:02	B*039:02	B*108:02	-	-	B*035:02	B*063:01Sp	B*072:04:02	-	-	-	-	-
B016a	i	B*016:01	-	-	-	-	B*089:03	B*112:02	B*189:01	I*01:20:01	I*01:06:01	-	-	-
B016a	ii	B*016:01	-	-	-	-	B*089:05	B*112:02	I*01:20:01	I*01:06:01	-	-	-	-
B016b	-	B*016:02	B*041:01	-	-	-	B*082:02	B*088:01	B*089:02	B*162:01	I*01:21	-	-	-
B016c	-	B*016:02	B*100:01:01	-	-	-	B*054:02	B*088:01	B*089:03	B*162:01	I*01:16	I*01:21	-	-
B017a	-	B*017:01	B*024:01	B*068:10	-	-	B*089:02	I*01:06:03	-	-	-	-	-	-
B017b	-	B*017:01	B*099:01	B*100:01:02	-	-	I*01:22:01	-	-	-	-	-	-	-
B017c	-	B*017:01	B*107:05	B*120:02:01	-	-	B*060:03:01	B*072:02:04	B*078:04	I*01:06:02	I*01:17:02	-	-	-
B017d	-	B*017:02	B*068:08	B*150:01	-	-	I*01:02	-	-	-	-	-	-	-
B017e	-	B*017:05	B*031:01	B*041:02	B*056:03	-	B*046:02	B*082:01	B*089:03	B*162:01	I*01:24	-	-	-
B017f	-	B*017:05	B*041:02	B*056:03	-	-	B*082:01	I*01:24	-	-	-	-	-	-
B024a	i	B*024:01	B*068:10	-	-	-	B*060:03:01	B*089:02	B*171:01	-	-	-	-	-
B024a	ii	B*024:01	B*068:12	-	-	-	B*088:01	B*162:01	-	-	-	-	-	-
B024b	-	B*024:02	B*068:10	B*178:01:01	-	-	B*089:07	B*116:02	-	-	-	-	-	-
B025a	-	B*025:01	-	-	-	-	B*115:05	I*01:15	-	-	-	-	-	-
B025b	-	B*025:01	B*027:04	-	-	-	B*051:06	B*072:06	B*115:05	I*01:15	-	-	-	-
B028	i	B*028:01:01	B*021:01	B*061:02	B*068:05	B*124:01	B*045:02	B*060:05	B*088:01	B*098:01	B*162:01	I*01:08:01	-	-
B028	ii	B*028:01:02	B*021:01	B*061:02	B*068:05	B*124:01	B*045:02	B*060:02	B*098:01	I*01:08:01	-	-	-	-
B028	iii	B*028:02	B*021:01	B*061:02	B*068:05	B*124:01	B*045:02	B*088:01	I*01:08:01	-	-	-	-	-
B031	-	B*031:01	-	-	-	-	B*035:02	B*116:01:01	I*01:24	-	-	-	-	-
B039	-	B*039:02	B*108:03	-	-	-	B*051:03	B*082:01	I*01:14:02	-	-	-	-	-
B043	i	B*043:01	B*030:01	B*045:01	-	-	B*060:06	B*162:01	I*01:07	-	-	-	-	-
B043	ii	B*043:01	B*030:03	B*045:01	-	-	I*01:04:01	-	-	-	-	-	-	-
B043	iii	B*043:02	B*030:01	B*045:01	-	-	B*072:02:04	B*078:06	-	-	-	-	-	-
B047a	i	B*047:01:01	B*068:02:02	-	-	-	B*079:01:02	B*112:01	I*01:01:01	I*01:06:01	-	-	-	-
B047a	ii	B*047:01:01	B*068:02:02	-	-	-	B*112:01	I*01:06:02	I*01:07	-	-	-	-	-
B047a	iii	B*047:01:01	B*068:02:03	-	-	-	B*112:05	B*162:01	-	-	-	-	-	-
B047a	iv	B*047:02	B*068:02:03	-	-	-	B*112:03	I*01:01:01	-	-	-	-	-	-
B047b	-	B*047:01:01	-	-	-	-	B*098:01	B*112:03	B*162:01	I*01:01:01	-	-	-	-
B052a	-	B*052:01	B*058:01	-	-	-	B*055:01	B*060:03:01	B*115:01	I*01:25	-	-	-	-
B052b	-	B*052:01	B*031:01	B*058:01	-	-	B*055:01	B*115:01	I*01:25	-	-	-	-	-
B056a	-	B*056:01	B*011:01	-	-	-	B*072:02:02	-	-	-	-	-	-	-
B056b	i	B*056:02	B*041:05	-	-	-	I*01:23:01	I*01:24	-	-	-	-	-	-
B056b	ii	B*056:02	B*041:05	-	-	-	I*01:23:02	I*01:24	-	-	-	-	-	-
B065	-	B*065:01:01	B*030:03	B*122:01	-	-	B*051:02:01	B*088:01	B*162:01	I*01:30	-	-	-	-
B068a	-	B*068:02:03	-	-	-	-	B*072:04:02	B*089:03	I*01:11	-	-	-	-	-
B068b	-	B*068:05	B*021:01	B*124:01	-	-	I*01:29	-	-	-	-	-	-	-
B068c	-	B*068:13	B*004:01	-	-	-	B*060:02	B*072:02:04	I*01:22:02	-	-	-	-	-
B069a	i	B*069:01	B*107:02	-	-	-	B*072:04:02	B*079:04	B*082:04	B*115:03	I*01:11	I*01:16	-	-
B069a	ii	B*069:01	B*107:02	-	-	-	B*072:03	B*079:04	B*089:03	B*116:01:01	I*01:24	-	-	-
B069b	i	B*069:02	B*081:01:01	B*179:01:01	-	-	B*032:01	I*01:31	-	-	-	-	-	-
B069b	ii	B*069:02	B*081:01:01	B*179:01:02	-	-	B*079:07	B*082:03	B*189:01	I*01:22:01	-	-	-	-
B069b	iii	B*069:02	B*081:02	B*179:01:02	-	-	B*060:03:01	B*082:02	-	-	-	-	-	-
B069b	iv	B*069:03	B*081:01:02	B*179:01:01	-	-	B*032:01	B*060:03:01	B*082:02	I*01:31	-	-	-	-
B091	-	B*091:01	B*068:02:03	-	-	-	B*113:01	I*01:04:01	I*01:22:01	-	-	-	-	-
B099	-	B*099:02	B*041:03	-	-	-	B*079:08	B*088:02:01	B*103:03	I*01:04:01	-	-	-	-
B101	-	B*101:01	B*068:02:03	-	-	-	B*051:02:01	B*072:04:02	B*079:07	B*089:03	I*01:22:02	-	-	-
B104a	i	B*104:02	-	-	-	-	B*035:02	B*046:01:01	B*072:01	B*079:01:05	B*144:01	I*01:22:01	-	-
B104a	ii	B*104:02		-	-	-	B*035:02	B*046:01:01	B*051:02:01	B*072:02:02	B*119:01	B*144:01	I*01:01:01	I*01:07
B104a	iii	B*104:02	-	-	-	-	B*046:01:01	B*078:02	B*116:02	B*144:01	I*01:24	-	-	-
B104a	iv	B*104:03	-	-	-	-	B*046:02	B*051:07	B*072:04:02	B*082:02	B*189:01	I*01:14:03	I*01:06:01	-
B104b	-	B*104:04	B*031:01	-	-	-	B*046:01:02	B*088:05	B*144:01	I*01:14:01	-	-	-	-
B105a	-	B*105:01	B*041:04	-	-	-	B*063:01Sp	B*088:02:02	B*103:01:02	B*162:01	I*01:20:02	I*01:23:01	-	-
B105b	-	B*105:03	B*107:04	B*197:01	-	-	B*089:06	B*098:03	I*01:17:01	-	-	-	-	-
B111a	i	B*111:01	B*027:03:02	-	-	-	I*01:32	-	-	-	-	-	-	-
B111a	ii	B*111:03	B*027:03:02	-	-	-	B*072:04:02	I*01:06:02	-	-	-	-	-	-
B111b	-	B*111:01	B*117:01	-	-	-	B*051:01	B*072:04:02	I*01:22:01	-	-	-	-	-
B111c	-	B*111:03	B*068:02:01	-	-	-	B*074:04:04	I*01:32	-	-	-	-	-	-
B118a	i	B*118:01:01	B*027:03:02	B*030:03	B*122:01	-	B*079:07	B*088:02:01	B*103:04	I*01:04:01	-	-	-	-
B118a	ii	B*118:01:01	B*027:03:02	B*030:03	B*122:01	-	B*072:08	B*082:03	I*01:30	-	-	-	-	-
B118a	iii	B*118:01:02	B*027:03:02	B*030:03	B*122:01	-	B*079:07	B*082:03	B*088:02:01	B*103:01:01	I*01:04:02	-	-	-
B118a	iv	B*118:01:02	B*027:03:02	B*030:03	B*122:01	-	B*088:02:01	B*103:01:01	I*01:35	-	-	-	-	-
B118a	v	B*118:02	B*027:03:02	B*030:03	B*122:01	-	B*088:01	B*098:01	B*162:01	I*01:04:01	-	-	-	-
B118b	i	B*118:01:02	B*027:03:02	B*030:01	B*043:01	B*122:01	B*045:02	B*060:03:01	B*072:01	B*079:07	B*082:03	B*088:02:01	B*103:01:01	I*01:07
B118b	ii	B*118:01:02	B*027:03:02	B*030:01	B*043:01	[B*122:01]	B*060:02	B*072:01	I*01:07	-	-	-	-	-
B118c	-	B*118:01:02	B*027:03:02	B*030:03	B*122:01	B*178:03	B*045:02	B*162:01	I*01:07	-	-	-	-	-
B118d	-	B*118:01:02	B*027:03:02	B*122:01	-	-	B*060:03:02	B*116:01:01	I*01:24	-	-	-	-	-
B118e	-	B*118:02	B*027:03:02	B*030:03	B*068:02:03	B*122:01	B*088:01	B*112:05	I*01:04:01	-	-	-	-	-
B118f	-	B*118:03	B*004:01	B*122:01	-	-	I*01:30	-	-	-	-	-	-	-
B118g	i	B*118:03	B*030:03	B*122:01	-	-	B*072:09	I*01:18	-	-	-	-	-	-
B118g	ii	B*118:04N	B*030:03	B*122:01	-	-	B*060:03:02	B*088:02:01	B*103:01:01	B*162:01	I*01:04:01	I*01:21	-	-
B118h	-	B*118:03	-	-	-	-	B*045:02	B*082:02	B*112:01	I*01:22:01	I*01:30	-	-	-
B120a	i	B*120:01	B*107:01	-	-	-	B*078:01	B*082:02	B*103:01:01	I*01:04:01	-	-	-	-
B120a	ii	B*120:01	B*107:01	-	-	-	B*082:02	B*116:01:01	I*01:24	-	-	-	-	-
B120a	iii	B*120:02:01	B*107:05	-	-	-	B*051:07	B*060:03:01	B*072:02:04	B*078:04	B*082:03	B*098:01	I*01:17:02	-
B120b	-	B*120:01	B*031:01	B*107:01	-	-	B*078:05	B*088:02:02	I*01:33	-	-	-	-	-
B120c	-	B*120:02:02	B*014:02:01N	B*107:05	-	-	B*060:03:01	B*072:02:04	B*078:04	B*082:03	I*01:17:02	-	-	-
B121	-	B*121:01	B*007:01	-	-	-	B*064:01	B*089:03	B*103:05	I*01:20:02	-	-	-	-
B122	-	B*122:01	B*027:03:02	-	-	-	B*072:02:04	I*01:30	-	-	-	-	-	-
B123a	i	B*123:01:01	B*030:04	B*068:09	B*083:02	-	I*01:01:03	-	-	-	-	-	-	-
B123a	ii	B*123:01:01	B*030:04	B*068:09	B*083:03	-	B*051:02:01	B*072:02:02	B*119:01	I*01:24	-	-	-	-
B123b	-	B*123:01:02	B*027:03:01	-	-	-	B*051:07	B*088:02:01	B*098:01	B*103:01:01	I*01:04:02	-	-	-
B136	-	B*136:01	B*004:01	-	-	-	B*072:02:02	I*01:27	I*01:09	-	-	-	-	-
B145	-	B*145:01	B*068:11	-	-	-	B*060:03:01	B*089:03	B*116:01:01	I*01:24	-	-	-	-
B150a	-	B*150:02	B*068:08	-	-	-	B*072:04:02	B*088:06	B*089:02	B*162:01	-	-	-	-
B150b	-	B*150:03	B*004:01	B*068:08	-	-	B*072:02:04	B*089:03	B*142:01	I*01:09	-	-	-	-
B199	i	B*199:01	B*107:04	-	-	-	B*051:02:01	B*072:07	B*089:03	-	-	-	-	-
B199	ii	B*199:02	B*107:04	-	-	-	B*060:03:01	B*072:02:02	-	-	-	-	-	-
B200a	-	B*200:01	B*019:02	B*030:03	-	-	B*088:01	I*01:22:02	-	-	-	-	-	-
B200b	-	B*200:01	B*030:03	-	-	-	B*088:02:01	B*103:01:01	I*01:35	-	-	-	-	-

◀ **Fig. 3** A description of the sequences that define each *Mane-B* haplotype. The haplotypes are named for the first major transcript described and contain other major and minor alleles that are passed onto offspring together. Major transcripts were defined as having over 4% of the total reads associated with each animal and minor alleles had less than 2% of the total reads. Sequences with an abundance between 2 and 4% were defined as major or minor in accordance with published alleles from rhesus or cynomolgus macaques (Karl et al. 2013, 2017)

will enhance the specificity to which the MHC-I proteins are defined but further research is necessary to define whether this level of definition gives rise to phenotypic differences noticeable to host immune cells.

In contrast, GvHD may be a beneficial component to potential HIV cure strategies. Timothy Ray Brown, better known as the "Berlin Patient," was "cured" of HIV when he received an allogeneic hematopoietic stem cell transplant from a homozygous CCR5Δ32 donor after intensive chemotherapy. Almost 10 years after his transplant, he maintains undetectable levels of HIV DNA and RNA without combination antiretroviral therapy (cART) (Yukl et al. 2013). It has been postulated that one of the mechanisms involved in treating, and eventually controlling, his infection with HIV was essentially a graft-versus-host effect that targets latently infected cells, thus depleting the HIV reservoir (Mavigner et al. 2014; Zou et al. 2013). Pig-tailed macaques have recently been used to model this case by infection with SHIV. After treatment with cART, a subset of macaques was given an autologous hematopoietic stem cell transplant before cART removal. This transplant, while leading to similar peak viral loads as the control group, also led to higher set points after cART removal. These findings suggest that the transplant disrupted the adaptive immune response after rebound and further highlights the role of the HIV-resistant cells received by the "Berlin Patient" (Reeves et al. 2017). In other transplant research, pig-tailed macaques have been utilized in studies showing organ regeneration after embryonic stem cell transplants (Chong et al. 2014). Due to their infectability with HIV and progression to AIDS (Baroncelli et al. 2008; Hatziioannou et al. 2014), pig-tailed macaques provide an important model to further test this HIV cure theory as well as model other important regenerative procedures.

Pig-tailed macaques and their use in infectious disease research

Pig-tailed macaques are important models for infectious disease research, and especially for research involving HIV, SIV, and AIDS-like diseases. Pig-tailed macaques express a variant of the TRIM5α protein that allows them to be infected with minimally modified forms of HIV-2 and multiple forms of SIV (Brennan et al. 2007; Hatziioannou et al. 2009, 2014; Igarashi et al. 2007; Kirmaier et al. 2010). Because of this, pig-tailed macaques provide a more accurate representation of the course of HIV infection in humans and progression to AIDS-like diseases (Baroncelli et al. 2008; Hatziioannou et al. 2014).

Certain MHC haplotypes provide protection against SIV and may be able to slow progression of SIV into AIDS-like diseases in macaques. Notably, in pig-tailed macaques, *Mane-A1*084:01*, previously named *Mane-A*10*, restricts the Gag KP9 and several additional epitopes of SIV, thus slowing SIV viral escape in T cell lymphocytes (Gooneratne et al. 2014). SIV viral loads in the plasma are also shown to be significantly reduced in macaques expressing *Mane-A1*084* compared to those macaques without it (Smith et al. 2005a, b). In the current study, we identified two additional variants of *Mane-A1*084* lineage (*Mane-A1*084:05* and *Mane-A1*084:06*) that differed from *Mane-A1*084:01* by one and two nonsynonymous substitutions, respectively. The amino acid substitution in *Mane-A1*084:05* that changes lysine to glutamic acid was outside the α1 and α2 domains. One of the amino acid substitutions in *Mane-A1*084:06* occurred in an F pocket key residue in the α1 domain, changing from glutamic acid to glycine, while the other substitution, changing from serine to asparagine, was in the α1 domain, but not in either a B or F pocket residue. A previous study was published that described how different MHC-I molecules with amino acid substitutions in F and B pocket residues bound the same peptides as *Mamu-B*008* and may provide the same level of protection (Loffredo et al. 2009). Because of these results, the possibility exists that our novel transcripts may provide the same level of protection, but further studies are required in order to determine whether this is true in vivo. If our novel sequences do provide the same level of protection, there is incentive for researchers to use animals containing these alleles for further studying HIV progression and escape in pig-tailed macaques.

Other species of macaques have documented alleles that are shown to restrict progression of SIV to AIDS-like disease. Notably in rhesus macaques, both *Mamu-B*008:01* and *Mamu-B*017:01:01* have been shown to control SIV replication and progression to disease; *Mamu-B*008:01* restricts the Vif RL8, Vif RL9, and Nef RL10 epitopes of SIV and *Mamu-B*017:01:01* restricts the Nef IW9 epitope (Loffredo et al. 2007; Martins et al. 2015; O'Connor et al. 2003; Yant et al. 2006). In our studies, we identified one novel variant of both *Mane-B*017:01* (*Mane-B*017:05*) and *Mane-B*008:01* (*Mane-B*008:02*). *Mane-B*008:02* was seen in four animals from the WaNPRC at relatively high numbers of the total reads from each of the four animals. This observation is consistent with high levels of transcription for the specific sequence, as has been observed for *Mamu-B*008:01* in rhesus macaques. *Mane-B*017:05* was identified in six animals coming from both the JHU and the WaNPRC colonies, again in relatively high read numbers. Both *Mane-B*008:02* and *Mane-B*017:05* differed from the previously characterized

a

Haplotype	Variant	Diagnostic Allele	JHU	WaNPRC	Melbourne
A003	i	*A1*003:01*	1		
A003	ii	*A1*003:03*		1	
A004a	i	*A1*004:01:01*	3	3	
A004a	ii	*A1*004:01:02*	2		
A004a	iii	*A1*004:02*	2		
A004b	-	*A1*004:03*	1		
A006	i	*A1*006:02*		4	
A006	ii	*A1*006:03*	2		
A006	iii	*A1*006:04*		2	
A006	iv	*A1*006:05*		1	
A009	i	*A1*009:01*	1	4	
A009	ii	*A1*009:01*		1	
A010	i	*A1*010:02*	4	5	
A010	ii	*A1*010:03*		5	
A016	-	*A1*016:nov:01*		2	
A018	-	*A1*018:03*			7
A019	i	*A1*019:03*	6	2	
A019	ii	*A1*019:04:01*	3		
A019	iii	*A1*019:04:02*			7
A019	iv	*A1*019:05*		1	
A019	v	*A1*019:06*			4
A019	vi	*A1*019:07*	3		
A031	-	*A1*031:01*	3	5	
A038	i	*A1*038:01*	1	1	
A038	ii	*A1*038:02*	1	3	5
A047	-	*A1*047:01*	2		
A052a	i	*A1*052:01*			52
A052a	ii	*A1*052:01*	4		
A052a	iii	*A1*052:01*	11	2	
A052a	iv	*A1*052:01*			2
A052b	i	*A1*052:01*	2		
A052b	ii	*A1*052:01*		1	
A053	-	*A1*053:01*			2
A054a	i	*A1*054:01*	2		
A054a	ii	*A1*054:02*	5		
A054b	-	*A1*054:04*			20
A060	-	*A1*060:01*			6
A061	i	*A1*061:01*	2	3	
A061	ii	*A1*061:01*		1	
A061	iii	*A1*061:02*	1		
A063	i	*A1*063:01*	2		
A063	ii	*A1*063:03*	2		
A063	iii	*A1*063:03*			1
A063	iv	*A1*063:03*		1	

Haplotype	Variant	Diagnostic Allele	JHU	WaNPRC	Melbourne
A072	i	*A1*072:03*	3		
A072	ii	*A1*072:04*		2	
A074	-	*A1*074:01*	2		
A082a	i	*A1*082:01:01*		1	
A082a	ii	*A1*082:01:02*	2		
A082a	iii	*A1*082:01:03*	2		
A082a	iv	*A1*082:01:03*			32
A082a	v	*A1*082:02*	2	1	
A082a	vi	*A1*082:03*		1	
A082a	vii	*A1*082:03*	2		
A082b	-	*A1*082:01:02*		5	
A082c	-	*A1*082:03*		1	
A083	-	*A1*083:01*	2		
A084a	i	*A1*084:01:01*	1	5	
A084a	ii	*A1*084:01:02*	2		
A084a	iii	*A1*084:01:02*			11
A084a	iv	*A1*084:01:02*	1	1	
A084a	v	*A1*084:01:02*			2
A084a	vi	*A1*084:03*	14	4	
A084a	vii	*A1*084:04*	3		
A084a	viii	*A1*084:05*	8		
A084a	ix	*A1*084:06*	1		
A084b	-	*A1*084:01:02*			4
A084c	-	*A1*084:01:01*	1		
A085	-	*A1*085:nov:01*		2	
A087	-	*A1*087:01*			3
A093	-	*A1*093:01*	3	1	
A095	-	*A1*095:01*	4		
A114	i	*A1*114:01*	1		4
A114	ii	*A1*114:02*		1	
A114	iii	*A1*114:03*	2		
A115	-	*A1*115:01*		1	
A116	-	*A1*116:01*		1	
A130	-	*A1*130:01*		3	
A131a	i	*A1*131:01:01*		1	
A131a	ii	*A1*131:01:02*		2	
A131b	-	*A1*131:02*		1	
A132	-	*A1*132:01*		1	
A701a	i	*A7*01:01*		1	
A701a	ii	*A7*01:02*		1	
A701b	-	*A7*01:02*	3		
A701c	-	*A7*01:01*	1	1	14

Fig. 4 **a** Summary of the distribution of *Mane-A* haplotypes across the three cohorts in the study. Each haplotype and variant is represented by the diagnostic transcript for which it is named with the number of chromosomes each haplotype was observed and indicated. **b** Summary of the distribution of *Mane-B* haplotypes across the three cohorts in the study. Each haplotype and variant is represented by the diagnostic transcript for which it is named with the number of chromosomes each haplotype was observed and indicated

Mamu versions by two nonsynonymous amino acid substitutions. The two amino acid substitutions in *Mane-B*008:02* occurred in the α1 domain at B pocket key residues, altering an arginine to asparagine and a histidine to asparagine. One amino acid substitution in *Mane-B*017:05* occurred at an F pocket residue in the α2 domain, changing an isoleucine to leucine, while the other occurred outside the α1 and α2 domains, changing from valine to isoleucine. As described earlier, these novel variants may show the same protective effect as is seen in their *Mamu* counterparts (Gooneratne et al. 2014; Loffredo et al. 2007; Martins et al. 2015; O'Connor et al. 2003; Yant et al. 2006). Subsequent studies are required to show a correlation between these newly discovered alleles and control of SIV replication as has been seen in rhesus macaques.

This newly characterized PacBio MHC data is an important resource for colony managers and researchers to use when designing further studies. For example, a recent study described the use of pig-tailed macaques to study neurocognitive disorders associated with SIV infection (Beck et al. 2017).

b

Haplotype	Variant	Diagnostic Allele	JHU	WaNPRC	Melbourne
B004a	-	*B*004:02*			24
B004b	-	*B*004:02*		1	
B008	-	*B*008:02*		2	
B013a	i	*B*013:01*			2
B013a	ii	*B*013:01*			7
B013b	i	*B*013:02*			6
B013b	ii	*B*013:02*	1	3	
B013b	iii	*B*013:02*	1		
B013b	iv	*B*013:02*			19
B015a	i	*B*015:01*	3	2	
B015a	ii	*B*015:01*	1		
B015a	iii	*B*015:01*		1	
B015a	iv	*B*015:04*		1	
B015b	i	*B*015:02*			15
B015b	ii	*B*015:02*		1	
B015b	iii	*B*015:02*	3		
B015c	-	*B*015:02*		5	
B016a	i	*B*016:01*	1	1	
B016a	ii	*B*016:01*	1	1	
B016b	-	*B*016:02*	3	1	
B016c	-	*B*016:02*	2		
B017a	-	*B*017:01*		2	
B017b	-	*B*017:01*		1	
B017c	-	*B*017:01*		2	
B017d	-	*B*017:02*	2		
B017e	-	*B*017:05*	4		
B017f	-	*B*017:05*		3	
B024a	i	*B*024:01*	2		
B024a	ii	*B*024:01*			2
B024b	-	*B*024:02*	1	2	
B025a	-	*B*025:01*	2		
B025b	-	*B*025:01*	7		
B028	i	*B*028:01:01*	20	7	
B028	ii	*B*028:01:02*		1	
B028	iii	*B*028:02*	2		
B031	-	*B*031:01*		1	
B039	-	*B*039:02*		1	
B043	i	*B*043:01*	4	1	
B043	ii	*B*043:01*		1	
B043	iii	*B*043:02*	1		
B047a	i	*B*047:01:01*		2	
B047a	ii	*B*047:01:01*	2		
B047a	iii	*B*047:01:01*		1	2
B047a	iv	*B*047:02*	2		
B047b	-	*B*047:01:01*	2		
B052a	-	*B*052:01*		3	
B052b	-	*B*052:01*	1		
B056a	-	*B*056:01*			2
B056b	i	*B*056:02*		2	
B056b	ii	*B*056:02*	1		
B065	-	*B*065:01:01*		1	
B068a	-	*B*068:02:03*			3
B068b	-	*B*068:05*		1	
B068c	-	*B*068:13*			2
B069a	i	*B*069:01*	2	1	9
B069a	ii	*B*069:01*	6		
B069b	i	*B*069:02*		1	
B069b	ii	*B*069:02*	2	2	
B069b	iii	*B*069:02*			4
B069b	iv	*B*069:03*		1	
B091	-	*B*091:01*	2		
B099	-	*B*099:02*	10		
B101	-	*B*101:01*			22
B104a	i	*B*104:02*	1		
B104a	ii	*B*104:02*	1	2	
B104a	iii	*B*104:02*	2		
B104a	iv	*B*104:03*	3	1	
B104b	-	*B*104:04*	1		
B105a	-	*B*105:01*	3		
B105b	-	*B*105:03*	2		
B111a	i	*B*111:01*		3	
B111a	ii	*B*111:03*	2		
B111b	-	*B*111:01*	1		
B111c	-	*B*111:03*	3		
B118a	i	*B*118:01:01*		1	
B118a	ii	*B*118:01:01*		1	
B118a	iii	*B*118:01:02*		2	
B118a	iv	*B*118:01:02*			4
B118a	v	*B*118:02*	6		
B118b	i	*B*118:01:02*		2	
B118b	ii	*B*118:01:02*		1	
B118c	-	*B*118:01:02*	1		
B118d	-	*B*118:01:02*			29
B118e	-	*B*118:02*	1		
B118f	-	*B*118:03*		1	
B118g	i	*B*118:03*	1	1	
B118g	ii	*B*118:04N*		1	
B118h	-	*B*118:03*		2	
B120a	i	*B*120:01*	1		
B120a	ii	*B*120:01*			1
B120a	iii	*B*120:02:01*		2	
B120b	-	*B*120:01*		1	
B120c	-	*B*120:02:02*	1		
B121	-	*B*121:01*	1		
B122	-	*B*122:01*			5
B123a	i	*B*123:01:01*	1		
B123a	ii	*B*123:01:01*	3		
B123b	-	*B*123:01:02*		2	
B136	-	*B*136:01*		2	
B145	-	*B*145:01*		2	4
B150a	-	*B*150:02*			8
B150b	-	*B*150:03*			3
B199	i	*B*199:01*		3	
B199	ii	*B*199:02*			1
B200a	-	*B*200:01*			1
B200b	-	*B*200:01*			1

Fig. 4 continued.

These researchers described how pig-tailed macaques that express *Mane-A1*084* were much less likely to develop encephalitis than macaques without *Mane-A1*084* after SIV challenge in this model of HIV neuropathogenesis. MHC genotyping allowed these investigators to refine their model and reduce the number of animals required per study group by excluding *Mane-A1*084*-positive animals when encephalitis was the desired study endpoint. MHC genotyping is also

important for colony managers when determining which animals should be bred in order to diversify the MHC haplotypes available in their colony. Alternatively, other researchers may require animals with MHC-identical haplotypes for transplant studies that require alternative breeding strategies.

The increased allelic resolution and discovery provided by PacBio CCS adds to our knowledge of pig-tailed macaques while also representing an important step for advancing their use in biomedical research. As an important model for studying HIV infection and progression to AIDS-like disease, the use of full-length sequencing technology and the discovery of novel allelic variants that may be protective against infection promote the use of the pig-tailed macaque model for infectious diseases.

Acknowledgements The authors gratefully acknowledge the support of Suzanne Queen at Johns Hopkins University, Rita Cervera Juanes and Betsy Ferguson at the Oregon Health & Science University, and Thakshila Amarasena and Stephen Kent at the University of Melbourne for their assistance in the collection and shipping of pig-tailed macaque RNA and cDNA. We also would like to thank Nel Otting, Natasja de Groot, and Ronald Bontrop at the Biomedical Primate Research Centre for their assistance in submitting and naming of novel sequences in the Immuno Polymorphism Database for the Major Histocompatibility Complex genes of Non-Human Primate.

Funding sources This research was supported by contracts HHSN272201600007C and HHSN272201100013C from the National Institute of Allergy and Infectious Diseases and the National Institutes of Health, and was conducted at a facility constructed with support from the Research Facilities Improvement Program (RR15459-01, RR20141-01).

Compliance with ethical standards All animals were cared for according to the regulations and guidelines of the Institutional Care and Use Committee at their respective institutions.

Conflict of interest The authors declare they have no conflict of interest.

References

Anasetti C, Beatty PG, Storb R, Martin PJ, Mori M, Sanders JE, Thomas ED, Hansen JA (1990) Effect of HLA incompatibility on graft-versus-host disease, relapse, and survival after marrow transplantation for patients with leukemia or lymphoma. Hum Immunol 29(2): 79–91

Ayala García MA, González Yebra B, López Flores AL, Guaní Guerra E (2012) The major histocompatibility complex in transplantation. J Transp Secur 2012:842141

Baroncelli S, Negri DR, Michelini Z, Cara A (2008) Macaca mulatta, fascicularis and nemestrina in AIDS vaccine development. Expert Rev Vaccines 7(9):1419–1434

Beck SE, Queen SE, Metcalf Pate KA, Mangus LM, Abreu CM, Gama L, Witwer KW, Adams RJ, Zink MC, Clements JE, Mankowski JL (2017) An SIV/macaque model targeted to study HIV-associated neurocognitive disorders. J Neuro-Oncol. https://doi.org/10.1007/s13365-017-0582-4

Bontrop RE (2006) Comparative genetics of MHC polymorphisms in different primate species: duplications and deletions. Hum Immunol 67(6):388–397

Brennan G, Kozyrev Y, Kodama T, Hu SL (2007) Novel TRIM5 isoforms expressed by Macaca Nemestrina. J Virol 81(22):12210–12217

Chong JJ, Yang X, Don CW, Minami E, Liu YW, Weyers JJ, Mahoney WM, Van Biber B, Cook SM, Palpant NJ, Gantz JA, Fugate JA, Muskheli V, Gough GM, Vogel KW, Astley CA, Hotchkiss CE, Baldessari A, Pabon L, Reinecke H, Gill EA, Nelson V, Kiem HP, Laflamme MA, Murry CE (2014) Human embryonic-stem-cell-derived cardiomyocytes regenerate non-human primate hearts. Nature 510(7504):273–277

Daza-Vamenta R, Glusman G, Rowen L, Guthrie B, Geraghty DE (2004) Genetic divergence of the rhesus macaque major histocompatibility complex. Genome Res 14(8):1501–1515

Del Prete GQ, Lifson JD, Keele BF (2016) Nonhuman primate models for the evaluation of HIV-1 preventive vaccine strategies: model parameter considerations and consequences. Curr Opin HIV AIDS 11(6):546–554

Fernandez CS, Reece JC, Saepuloh U, De Rose R, Ishkandriati D, O'Connor DH, Wiseman RW, Kent SJ (2011) Screening and confirmatory testing of MHC class I alleles in pig-tailed macaques. Immunogenetics 63(8):511–521

Frumkin LR, Patterson BK, Leverenz JB, Agy MB, Wolinsky SM, Morton WR, Corey L (1995) Infection of Macaca nemestrina brain with human immunodeficiency virus type 1. J Gen Virol 76(Pt 10): 2467–2476

Gardner MB, Luciw PA (2008) Macaque models of human infectious disease. ILAR J 49(2):220–255

Gooneratne SL, Alinejad-Rokny H, Ebrahimi D, Bohn PS, Wiseman RW, O'Connor DH, Davenport MP, Kent SJ (2014) Linking pig-tailed macaque major histocompatibility complex class I haplotypes and cytotoxic T lymphocyte escape mutations in simian immunodeficiency virus infection. J Virol 88(24):14310–14325

Hatziioannou T, Ambrose Z, Chung NP, Piatak M, Yuan F, Trubey CM, Coalter V, Kiser R, Schneider D, Smedley J, Pung R, Gathuka M, Estes JD, Veazey RS, KewalRamani VN, Lifson JD, Bieniasz PD (2009) A macaque model of HIV-1 infection. Proc Natl Acad Sci U S A 106(11):4425–4429

Hatziioannou T, Del Prete GQ, Keele BF, Estes JD, McNatt MW, Bitzegeio J, Raymond A, Rodriguez A, Schmidt F, Mac Trubey C, Smedley J, Piatak M, KewalRamani VN, Lifson JD, Bieniasz PD (2014) HIV-1-induced AIDS in monkeys. Science 344(6190): 1401–1405

Igarashi T, Iyengar R, Byrum RA, Buckler-White A, Dewar RL, Buckler CE, Lane HC, Kamada K, Adachi A, Martin MA (2007) Human immunodeficiency virus type 1 derivative with 7% simian immunodeficiency virus genetic content is able to establish infections in pig-tailed macaques. J Virol 81(20):11549–11552

Joag SV, Adany I, Li Z, Foresman L, Pinson DM, Wang C, Stephens EB, Raghavan R, Narayan O (1997) Animal model of mucosally transmitted human immunodeficiency virus type 1 disease: intravaginal and oral deposition of simian/human immunodeficiency virus in macaques results in systemic infection, elimination of CD4+ T cells, and AIDS. J Virol 71(5):4016–4023

Kanthaswamy S, Ng J, Penedo MC, Ward T, Smith DG, Ha JC (2012) Population genetics of the Washington National Primate Research Center's (WaNPRC) captive pigtailed macaque (Macaca nemestrina) population. Am J Primatol 74(11):1017–1027

Karl JA, Bohn PS, Wiseman RW, Nimityongskul FA, Lank SM, Starrett GJ, O'Connor DH (2013) Major histocompatibility complex class I haplotype diversity in Chinese rhesus macaques. G3 (Bethesda) 3(7):1195–1201

Karl JA, Heimbruch KE, Vriezen CE, Mironczuk CJ, Dudley DM, Wiseman RW, O'Connor DH (2014) Survey of major histocompatibility

complex class II diversity in pig-tailed macaques. Immunogenetics 66(11):613–623

Karl JA, Graham ME, Wiseman RW, Heimbruch KE, Gieger SM, Doxiadis GG, Bontrop RE, O'Connor DH (2017) Major histocompatibility complex haplotyping and long-amplicon allele discovery in cynomolgus macaques from Chinese breeding facilities. Immunogenetics 69(4):211–229

Kirmaier A, Wu F, Newman RM, Hall LR, Morgan JS, O'Connor S, Marx PA, Meythaler M, Goldstein S, Buckler-White A, Kaur A, Hirsch VM, Johnson WE (2010) TRIM5 suppresses cross-species transmission of a primate immunodeficiency virus and selects for emergence of resistant variants in the new species. PLoS Biol 8(8): e1000462

Loffredo JT, Maxwell J, Qi Y, Glidden CE, Borchardt GJ, Soma T, Bean AT, Beal DR, Wilson NA, Rehrauer WM, Lifson JD, Carrington M, Watkins DI (2007) Mamu-B*08-positive macaques control simian immunodeficiency virus replication. J Virol 81(16):8827–8832

Loffredo JT, Sidney J, Bean AT, Beal DR, Bardet W, Wahl A, Hawkins OE, Piaskowski S, Wilson NA, Hildebrand WH, Watkins DI, Sette A (2009) Two MHC class I molecules associated with elite control of immunodeficiency virus replication, Mamu-B*08 and HLA-B*2705, bind peptides with sequence similarity. J Immunol 182(12):7763–7775

Maccari G, Robinson J, Ballingall K, Guethlein LA, Grimholt U, Kaufman J, Ho CS, de Groot NG, Flicek P, Bontrop RE, Hammond JA, Marsh SG (2017) IPD-MHC 2.0: an improved inter-species database for the study of the major histocompatibility complex. Nucleic Acids Res 45(D1):D860–D864

Malaivijitnond S, Arsaithamkul V, Tanaka H, Pomchote P, Jaroenporn S, Suryobroto B, Hamada Y (2012) Boundary zone between northern and southern pig-tailed macaques and their morphological differences. Primates 53(4):377–389

Martins MA, Tully DC, Cruz MA, Power KA, Veloso de Santana MG, Bean DJ, Ogilvie CB, Gadgil R, Lima NS, Magnani DM, Ejima K, Allison DB, Piatak M, Altman JD, Parks CL, Rakasz EG, Capuano S, Galler R, Bonaldo MC, Lifson JD, Allen TM, Watkins DI (2015) Vaccine-induced simian immunodeficiency virus-specific CD8+ T-cell responses focused on a single Nef epitope select for escape variants shortly after infection. J Virol 89(21):10802–10820

Mavigner M, Watkins B, Lawson B, Lee ST, Chahroudi A, Kean L, Silvestri G (2014) Persistence of virus reservoirs in ART-treated SHIV-infected rhesus macaques after autologous hematopoietic stem cell transplant. PLoS Pathog 10(9):e1004406

Morizane A, Kikuchi T, Hayashi T, Mizuma H, Takara S, Doi H, Mawatari A, Glasser MF, Shiina T, Ishigaki H, Itoh Y, Okita K, Yamasaki E, Doi D, Onoe H, Ogasawara K, Yamanaka S, Takahashi J (2017) MHC matching improves engraftment of iPSC-derived neurons in non-human primates. Nat Commun 8(1): 385

O'Connor DH, Mothe BR, Weinfurter JT, Fuenger S, Rehrauer WM, Jing P, Rudersdorf RR, Liebl ME, Krebs K, Vasquez J, Dodds E, Loffredo J, Martin S, McDermott AB, Allen TM, Wang C, Doxiadis GG, Montefiori DC, Hughes A, Burton DR, Allison DB, Wolinsky SM, Bontrop R, Picker LJ, Watkins DI (2003) Major histocompatibility complex class I alleles associated with slow simian immunodeficiency virus disease progression bind epitopes recognized by dominant acute-phase cytotoxic-T-lymphocyte responses. J Virol 77(16):9029–9040

O'Leary CE, Wiseman RW, Karl JA, Bimber BN, Lank SM, Tuscher JJ, O'Connor DH (2009) Identification of novel MHC class I sequences in pig-tailed macaques by amplicon pyrosequencing and full-length cDNA cloning and sequencing. Immunogenetics 61(10):689–701

Prall TM, Graham ME, Karl JA, Wiseman RW, Ericsen AJ, Raveendran M, Alan Harris R, Muzny DM, Gibbs RA, Rogers J, O'Connor DH (2017) Improved full-length killer cell immunoglobulin-like receptor transcript discovery in Mauritian cynomolgus macaques. Immunogenetics 69(5):325–339

Pratt BF, O'Connor DH, Lafont BA, Mankowski JL, Fernandez CS, Triastuti R, Brooks AG, Kent SJ, Smith MZ (2006) MHC class I allele frequencies in pigtail macaques of diverse origin. Immunogenetics 58(12):995–1001

Reeves DB, Peterson CW, Kiem HP, Schiffer JT (2017) Autologous stem cell transplantation disrupts adaptive immune responses during rebound simian/human immunodeficiency virus viremia. J Virol 91(13):e00095–e00017

Reimann KA, Parker RA, Seaman MS, Beaudry K, Beddall M, Peterson L, Williams KC, Veazey RS, Montefiori DC, Mascola JR, Nabel GJ, Letvin NL (2005) Pathogenicity of simian-human immunodeficiency virus SHIV-89.6P and SIVmac is attenuated in cynomolgus macaques and associated with early T-lymphocyte responses. J Virol 79(14):8878–8885

Robinson J, Halliwell JA, McWilliam H, Lopez R, Parham P, Marsh SG (2013) The IMGT/HLA database. Nucleic Acids Res 41(Database issue):D1222–D1227

Smith MZ, Dale CJ, De Rose R, Stratov I, Fernandez CS, Brooks AG, Weinfurter J, Krebs K, Riek C, Watkins DI, O'connor DH, Kent SJ (2005a) Analysis of pigtail macaque major histocompatibility complex class I molecules presenting immunodominant simian immunodeficiency virus epitopes. J Virol 79(2):684–695

Smith MZ, Fernandez CS, Chung A, Dale CJ, De Rose R, Lin J, Brooks AG, Krebs KC, Watkins DI, O'Connor DH, Davenport MP, Kent SJ (2005b) The pigtail macaque MHC class I allele Mane-A*10 presents an immunodominant SIV Gag epitope: identification, tetramer development and implications of immune escape and reversion. J Med Primatol 34(5–6):282–293

Stremlau M, Owens CM, Perron MJ, Kiessling M, Autissier P, Sodroski J (2004) The cytoplasmic body component TRIM5alpha restricts HIV-1 infection in Old World monkeys. Nature 427(6977):848–853

Westbrook CJ, Karl JA, Wiseman RW, Mate S, Koroleva G, Garcia K, Sanchez-Lockhart M, O'Connor DH, Palacios G (2015) No assembly required: full-length MHC class I allele discovery by PacBio circular consensus sequencing. Hum Immunol 76(12):891–896

Wiseman RW, Karl JA, Bimber BN, O'Leary CE, Lank SM, Tuscher JJ, Detmer AM, Bouffard P, Levenkova N, Turcotte CL, Szekeres E, Wright C, Harkins T, O'Connor DH (2009) Major histocompatibility complex genotyping with massively parallel pyrosequencing. Nat Med 15(11):1322–1326

Wiseman RW, Karl JA, Bohn PS, Nimityongskul FA, Starrett GJ, O'Connor DH (2013) Haplessly hoping: macaque major histocompatibility complex made easy. ILAR J 54(2):196–210

Yant LJ, Friedrich TC, Johnson RC, May GE, Maness NJ, Enz AM, Lifson JD, O'Connor DH, Carrington M, Watkins DI (2006) The high-frequency major histocompatibility complex class I allele Mamu-B*17 is associated with control of simian immunodeficiency virus SIVmac239 replication. J Virol 80(10):5074–5077

Yukl SA, Boritz E, Busch M, Bentsen C, Chun TW, Douek D, Eisele E, Haase A, Ho YC, Hütter G, Justement JS, Keating S, Lee TH, Li P, Murray D, Palmer S, Pilcher C, Pillai S, Price RW, Rothenberger M, Schacker T, Siliciano J, Siliciano R, Sinclair E, Strain M, Wong J, Richman D, Deeks SG (2013) Challenges in detecting HIV persistence during potentially curative interventions: a study of the Berlin patient. PLoS Pathog 9(5):e1003347

Zhang X, Kanthaswamy S, Trask JS, Ng J, Oldt RF, Mankowski JL, Adams RJ, Smith DG (2017) Genetic characterization of a captive colony of pigtailed macaques (Macaca nemestrina). J Am Assoc Lab Anim Sci 56(4):390–395

Zou S, Glynn S, Kuritzkes D, Shah M, Cook N, Berliner N (2013) Hematopoietic cell transplantation and HIV cure: where we are and what next. Blood 122(18):3111–3115