Experimental and Molecular Pathology 100 (2016) 248–256

Contents lists available at ScienceDirect

Experimental and Molecular Pathology

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

Genomic landscapes of endogenous retroviruses unveil intricate genetics of conventional and genetically-engineered laboratory mouse strains

Kang-Hoon Lee [1], Debora Lim [1], Sophia Chiu, David Greenhalgh, Kiho Cho *

Department of Surgery, University of California, Davis, and Shriners Hospitals for Children Northern California, Sacramento, CA 95817, United States

ARTICLE INFO

Article history:
Received 24 September 2015
Accepted 9 January 2016
Available online 11 January 2016

Keywords:
Transposable repetitive elements (TREome)
TREome landscape
TREome gene profile
Genome complexity
Genetic surveillance

ABSTRACT

Laboratory strains of mice, both conventional and genetically engineered, have been introduced as critical components of a broad range of studies investigating normal and disease biology. Currently, the genetic identity of laboratory mice is primarily confirmed by surveying polymorphisms in selected sets of "conventional" genes and/or microsatellites in the absence of a single completely sequenced mouse genome. First, we examined variations in the genomic landscapes of transposable repetitive elements, named the TREome, in conventional and genetically engineered mouse strains using murine leukemia virus-type endogenous retroviruses (MLV-ERVs) as a probe. A survey of the genomes from 56 conventional strains revealed strain-specific TREome landscapes, and certain families (*e.g.*, C57BL) of strains were discernible with defined patterns. Interestingly, the TREome landscapes of C3H/HeJ (toll-like receptor-4 [TLR4] mutant) inbred mice were different from its control C3H/HeOuJ (TLR4 wild-type) strain. In addition, a CD14 knock-out strain had a distinct TREome landscape compared to its control/backcross C57BL/6J strain. Second, an examination of superantigen (SAg, a "TREome gene") coding sequences of mouse mammary tumor virus-type ERVs in the genomes of the 46 conventional strains revealed a high diversity, suggesting a potential role of SAgs in strain-specific immune phenotypes. The findings from this study indicate that unexplored and intricate genomic variations exist in laboratory mouse strains, both conventional and genetically engineered. The TREome-based high-resolution genetics surveillance system for laboratory mice would contribute to efficient study design with quality control and accurate data interpretation. This genetics system can be easily adapted to other species ranging from plants to humans.

1. Introduction

For more than a century, a variety of laboratory mouse strains have been identified, developed, and created all around the world. Some laboratory mouse strains ("conventional") were developed through defined breeding schemes to stabilize, isolate, and/or acquire specific traits and the others ("genetically engineered") were created by molecular and genetic manipulation of embryos, primarily to study functions of "conventional" genes (Austin et al., 2004; Rossant, 2013).

The majority of the conventional laboratory mouse strains, which are critical components of modern biomedical research efforts, were developed as inbred strains via initial cross-breeding followed by a well-defined set of contiguous brother–sister matings. In general, biomedical researchers accept the notion that homozygosity at all genomic loci in inbred strains is established by up to 40 sequential sibling matings (Green, 1981; Taft et al., 2006). However, it is not uncommon to witness visible phenotypic variations in a mouse population of a single inbred strain which are maintained at a single institution, not to mention multiple institutions (Phelan and Austad, 1994; Weichman and Chaillet, 1997, personal communication). Interestingly, a recent report that the genomic configurations of C57BL/6J inbred mice are altered temporally and spatially, resulting in numerous variant genomes in an individual mouse, contradicts the given properties of genetic homozygosity and/or uniformity within a population of an inbred strain (Lee et al., 2015). It is possible that the dynamic nature of the genomes, which has not been fully explored or understood, is responsible, at least in part, for the establishment of enough genetic heterozygosity sufficient for developing variable phenotypes in an inbred strain.

The genetically engineered mouse strains are mainly categorized into two types: 1) transgenic strain in which an unknown number of copies of an exogenous gene is introduced during embryonic development and 2) knock-out strain in which a specific gene locus, occasionally including unintended targets, is partially manipulated using a variety of protocols to disturb its presumed coding potentials (Li et al., 2013; Thomas and Capecchi, 1987). Similar to the conventional laboratory mouse strains, incoherent, often visible, phenotypes are manifested within a population of a transgenic or knock-out strain (Doetschman, 2009; Montagutelli, 2000; Sigmund, 2000). This phenotypic inconsistency among the population of

* Corresponding author at: Department of Surgery, University of California, Davis, and Shriners Hospitals for Children Northern California, 2425 Stockton Blvd., Sacramento, CA 95817, United States.
E-mail address: kcho@ucdavis.edu (K. Cho).
[1] K-HL and DL contributed equally to this study.

http://dx.doi.org/10.1016/j.yexmp.2016.01.005
0014-4800/

a genetically engineered strain might be linked to dissimilarity outside of the target locus or loci which are not thoroughly interrogated during genotyping processes. In fact, not all genetically engineered strains are subjected to a defined series of brother–sister matings to establish a recombinant inbred strain (Eisener-Dorman et al., 2009). It is recommended that if the backcross strain is an inbred, the specific inbred strain or its congenic strain serves as a control for the genetically engineered strain (Lambert, 2007). As such, it would be logical to question the genetic uniformity of the areas outside the target locus or loci between a genetically engineered strain and its inbred backcross-control strain.

The genetic integrity of conventional inbred as well as genetically engineered mouse strains is primarily monitored by examining: 1) pathophysiologic characteristics (*e.g.*, coat color, tumor, susceptibility, lifespan) and 2) molecular profiles of the genome focusing on specific marker sets (Davisson, 1990; Fox et al., 1997; Silver, 1995). In particular, the molecular monitoring protocols typically survey for single nucleotide polymorphisms (SNPs) of conventional genes, microsatellite variations, and engineered mutant gene loci (Petkov et al., 2004; http://jaxmice.jax.org/geneticheath/monitoring). In consideration of the absence of a single complete mouse genome sequence in any genome database and the estimated size of the mouse genome of over 2.5 Gb, the information gathered from a survey of the small sets of gene SNPs and microsatellites may be insufficient for confirming the genetic integrity and uniformity of any mouse strain, either conventional or genetically engineered.

Exons of conventional genes consist of a fraction (~1.2%) of the mouse and human genomes, and the vast majority of the residual genome is occupied by a plethora of repetitive elements (REs) which include transposable repetitive elements (TREs), named the TREome (Mouse Genome Sequencing Consortium, 2002; Lander et al., 2001). TREs have the potential to dynamically shape the genomic landscape of an individual through “copy and paste” functions. Accumulation of acquired TRE activity in an individual's genomes due to both acute and chronic stress signals from the environment (*e.g.*, onset of puberty, injury, infection) could conceivably be one genetic mechanism shaping the uncharacterized variations in a population of mice (Bohne et al., 2008; Cho et al., 2008b; Lee et al., 2012). As such, polymorphisms of conventional genes and a minute and biased set of microsatellites may play somewhat limited roles in defining the variable phenotypes of an inbred strain. Uncharacterized and/or unexplored genetic elements (*e.g.*, TREs) and their variations, which are inherently and adaptively embedded in the individual mouse's genomes, are likely to contribute to the processes determining specific and dynamic phenotypes.

Our recent finding of highly polymorphic TRE profiles in the mouse population is inconsistent with the current assumption of high genome identity with regard to their conventional gene sequences (Frazer et al., 2007; Kao et al., 2012; Kirby et al., 2010; Lee et al., 2011; Mekada et al., 2009). In addition, recent findings of age- and organ-specific variations in genomic configurations due to TRE activity in C57BL/6J inbred mice contradicts the perceived genetic staticity and uniformity of a given inbred mouse population (Lee et al., 2012). The genetic integrity, stability, and/or uniformity of laboratory mouse strains are typically confirmed by a set of polymorphism data (*e.g.*, morphophysiologic phenotypes, conventional genes, microsatellites). From a logical and efficient research perspective, the current systems for monitoring mouse genetics need to be re-evaluated to account for the uncharacterized properties and effects of the inherent diversity and acquired activity of TREs and other explored elements, on a dynamic genome platform.

2. Materials and methods

2.1. Animal experiments

The following mouse strains were purchased from the Jackson Laboratory: female C57BL/6J, C3H/HeJ, C3H/HeOuJ, and CD14 knock-out (B6.129S4-*Cd14*$^{tm/frm}$/J). All animals were provided with water and food *ad libitum* at a University of California Davis facility and some of the mice were aged for a period of time. The experimental protocol was approved by the Animal Use and Care Administrative Advisory Committee of the University of California Davis. Animals were sacrificed to collect tissues followed by snap-freezing in liquid nitrogen.

2.2. Genomic DNA of various laboratory mouse strains

Snap-frozen tissue samples were subjected to genomic DNA isolation using a DNeasy Tissue kit (Qiagen, Valencia, CA) and the DNA samples were normalized to 20 ng/μl. In addition, genomic DNA from 63 laboratory mouse strains, which include nine 129 substrains, were purchased from the Jackson Laboratory (Bar Harbor, ME). In addition, genomic DNA from a C57BL/6J × 129S1/SvImJF2/J (B6129SF2/J) mouse, a F2 hybrid from F1 × F1 whose parents were C57BL/6J (female) and 129S1/SvImJ (male), was obtained from the Jackson Laboratory. According to the information from the Jackson Laboratory's website, the genomic DNA was isolated from either the brain or spleen of respective mouse strains (http://jaxmice.jax.org/jaxnotes/archive/499b.html). Gender identity of each DNA sample was confirmed by amplifying a region specific for mouse chromosome Y by PCR using a pair of primers (Supplementary Table 1) followed by agarose gel electrophoresis.

2.3. Polymorphism analysis of genomic TREome (murine leukemia virus-type endogenous retrovirus [MLV-ERV]) long terminal repeats (LTRs)

The polymorphic regions of the MLV-ERV LTRs were identified from the genomic DNAs of 12 laboratory mouse strains (Jackson Laboratory) by PCR using a set of primer pairs (Supplementary Table 1) which were designed from a well-conserved region. Following ligation into a TA vector (Promega, Madison, WI), 24 colonies were picked from the MLV-ERV amplicons of each strain and plasmid DNAs were prepared using a QIAprep Spin Miniprep kit (Qiagen) before sequencing (Molecular Cloning Laboratories, South San Francisco, CA). A set of unique MLV-ERV LTR sequences was compiled for each mouse strain by multiple alignment analysis using the Vector NTI program (Invitrogen, Carlsbad, CA). Within a set of unique MLV-ERV LTR sequences for each mouse strain, the occurrence frequency of 256 four-nucleotide “word” (a nucleotide sequence of specific length) combinations at all four possible reading frames were counted using a program developed in our laboratory (unpublished). Within each strain, the occurrence frequency data for the individual words were normalized and converted into probability distribution function (PDF) values. For each word, the average and standard deviation of the PDF values from all 12 strains were calculated using Excel (Microsoft, Redmond, WA). Based on an assumption that the higher the standard deviation in a word, the more variation in the word, the extent of variations in each four-nucleotide word within the 12 strain-population was visualized with a schedule of gray shades (white-lowest variation; black-highest variation) on a 16 × 16 (=256) matrix. To examine/simulate diversity in conventional gene sequences in comparison to the MLV-ERV LTR sequences, the single nucleotide polymorphism (SNP) data for the GAPDH gene (~4.7 kb) among 19 laboratory mouse strains (A/J, C57BL/6J, 129X1/SvJ, AKR/J, BALB/cByJ, C3H/HeJ, CAST/EiJ, DBA/2J, FVB/NJ, MOLF/EiJ, NOD/ShiLtJ, SM/J, BTBR $T^{+}Itpr3^{tf}$/J, KK/HlJ, LG/J, NZW/LacJ, PWD/PhJ, WSB/EiJ, and 129S1/SvImJ; http://www.informatics.jax.org/javawi2/servlet/WIFetch?page=snpQF) was subjected to the same PDF analysis as above.

2.4. TREome landscaping of mouse genomes using MLV-ERV sequences as a probe

Genomic DNA (20 ng) was cut with Nco-I (New England Biolab, Ipswich, MA) at 37 °C for 4 h followed by self-ligation of the cut fragments using T4 ligase (Promega) overnight at 4 °C. The TREome landscape data was collected by I-PCR amplification of the junctions spanning putative MLV-ERV integration loci using 2 μl of the ligation products, Taq