Determination of linkage disequilibrium region suggests association of the ancient haplotype, *hX* with neural function

Makoto K. Shimada* and Tsutomu Kanasashi

Institute for Comprehensive Medical Science, Fujita Health University, Toyoake, Aichi 470-1192, Japan

**Corresponding author*: mshimada@fujita-hu.ac.jp

Abstract

Modern human populations are known to contain “ancient haplotypes” that originated from archaic humans by hybridization. Some of them had been reported before development of human genomic diversity databases, such as HapMap. Consequently, some of them have no information about linkage disequilibrium (LD) regions. Because genetic information within LD is tightly linked, to know LD region containing ancient haplotypes will be useful to estimate basic parameters of admixture events, and to infer biological functions that linked with the ancient haplotypes. One of these ancient haplotypes, haplotype X (*hX*) was found in a 10.1 kb-region located on Xp11.22, which diverged at 1.4 M years ago, with low diversity within the cluster in gene genealogy and worldwide distribution in low frequency. We determined the LD region around the ancient haplotypes using LD information obtained in the HapMap project. The LD determination presents that the LD region surrounding the *hX* is stable and contain genic region that may associate with neural and brain functions.

Introduction

Recent advancement of genome-wide genotyping projects on worldwide human populations has unveiled genetic differences among current human populations [1-3]. In the research history of human evolution, some researchers suggested haplotypes (ancient haplotypes) that have been considered to be brought into anatomically modern humans (*Homo sapiens*, AMH) before completion of these genome-wide human diversity projects [4-9]. Under this situation, genotyping of archaic human using paleontological samples revealed

introgression to AMH from archaic humans that occurred more significantly in non-Africa than in Africa [10, 11].

Because ancestral haplotypes in current human populations can be examined repeatedly in laboratories, ancestral haplotypes will be useful maker that make up disadvantage of genome-wide sequencing using paleontological samples.

Since information on some of ancient haplotypes is remain to be a piecemeal basis, information of linkage disequilibrium (LD) regions containing ancient haplotypes have not investigated. Genetic signals coded within LD region tightly linked each other and are subjected to selection pressures as a unit. To determine LD region of ancient haplotypes is an initial step to take advantage of ancient haplotypes, which will lead to estimation of basic parameters of admixture events and assumption of biological functions that linked with the ancient haplotypes.

Using HapMap information, we determined the LD region surrounding the ancient haplotype, *hX*. We present here that the LD region surrounding the *hX* shows stable LD region and it contain genic region that may associate with neural and brain functions.

Materials and methods

Ancient Haplotype Candidate Regions

The *hX* region is the ancient haplotype we have reported earlier [5]. We have found the *hX*, when we performed re-sequencing of 10.1-kbp noncoding region in the human X chromosome using the males of HGDP-CEPH Human Genome Diversity Panel (672 individuals from 52 populations) [12]. We observed that the *hX* was distributed over the world at low frequencies. Microsatellite (short tandem repeat, STR) variation within the re-sequenced region was low among copies of *hX*, even though the estimated time of ancestry of *hX* and other sequences was 1.44 M years. The region we performed re-sequencing was selected from regions showing low recombination in human chromosome X by searching recombination data at the time. Furthermore, we designed that the regions should contain one or more STR sites for gene genealogical analysis.

LD determination using HapMap data

HapMap LD Data was downloaded through the HapMap Browser (http://hapmap.org/ [13]) (Release #27 merged II+III, Feb 2009, on NCBI B36 and dbSNP b126) in JPT and CEU populations. We defined LD region when r^2 values of neighboring four combinations formed by four SNPs (i.e., neighboring two SNPs from 5' ends and two SNPs from 3' ends) that covered the widest chromosomal region overlapping the ancient haplotype candidate region. If the data from two populations (i.e., JPT and CEU) shows different region, we defined the longer one as the LD region.

Results

LD region determination

Figure 1. LD region shown by HapMap Browser. The *hX* region is represented by grey area.

Based on our criteria, HapMap data in CEU population showed distinct high-LD SNP combinations over the *hX* region (Figure 1). Table 1 shows r^2 values of neighboring four combinations formed by four SNPs those are neighboring two SNPs in both ends of the obtained LD region. The obtained LD region covers from 50,566,211 to 50,621,379 in NCBI B36, which spans 55,169 bp in length. Comparing the length of LD region with that of the *hX* region (50,583,087 to 50,593,170, 10,084 bp in length), the LD region extends to 5.47 times of the *hX* region in length.

Table 1. HapMap LD data representing LD region of *hX*

pos_SNP1	pos_SNP2	pop	rsID_SNP1	rsID_SNP2	*D'*	r^2	*LOD*
50,566,211	50,619,102	CEU	rs5915314	rs5915333	1	0.851	17.61
50,566,211	50,621,379	CEU	rs5915314	rs2382650	1	0.851	17.61
50,571,456	50,619,102	CEU	rs6614576	rs5915333	1	0.898	18.56
50,571,456	50,621,379	CEU	rs6614576	rs2382650	1	0.898	18.56

Figure 2. Genomic annotation of the LD region containing *hX* shown by UCSC genome browser. The *hX* region is represented by black bar with red gaps at the top of the view.

Observation of *hX* LD region

Figure 2 shows annotations on the obtained LD region in the UCSC genome browser [14]. According to the annotation shown in the UCSC genome browser, the LD region contains the first exon and intron of SHROOM4 gene, and its promoter and/or enhancer, those are not included in the 10.1-kb *hX* region. SHROOM4 is thought to regulate cytoskeletal architecture by modulating the spatial distribution of myosin II [15]. SHROOM4 gene is expressed various types of cells. It is noteworthy that SHROOM4 gene functioned in neurulation, which is based on the fact that SHROOM4 gene is related to the Stocco dos Santos X-linked mental retardation syndrome. Furthermore, the annotation of OMIM [16] data indicates that the LD region is associated with mental disorders suggested by at least 17 studies.

Discussion

We present that LD region containing the *hX* also contain functional regions that related to

brain and/or central nerve system (CNS), though *hX* region itself is noncoding region. Because strong LD region inherits as a block, LD region act as a genomic regional unit in evolution. The variation on the LD region containing ancient haplotype may represent ancient variation. Using knowledge of relationship between sequence variation and function, the LD region may be used to predict functional difference between archaic human and majority of AMH. Moreover, because evolution of brain/CNS is significant to human evolution, the maintenance mechanism of ancestral haplotype in introgression process may be a clue to solve brain/CNS evolution in human lineage. The some other ancient haplotypes suggesting introgression from archaic human to AMH have association with brain/CNS phenotypes, such as the H2 haplotype on the MAPT gene locus [8, 9] and the D haplotype of the microcephalin (MCPH1)[7]. These examples may encourage developing a hypothesis that introgression from archaic human have a certain role in human evolution. The LD region obtained in the present study will be expected to help researchers in further studies.

Acknowledgements

We are grateful to Michiyo Murase for her help in the analysis. This work is based on the study M.K.S performed with Jody Hey and his colleagues, and financially supported by Grant-in-Aid for Scientific Research on Innovative Areas (22101001).

Reference

1. The 1000 Genomes Project Consortium: **A map of human genome variation from population-scale sequencing.** *Nature* 2010, **467**:1061-1073.
2. The International HapMap 3 Consortium: **Integrating common and rare genetic variation in diverse human populations.** *Nature* 2010, **467**:52-58.
3. Levy S, Sutton G, Ng PC, Feuk L, Halpern AL, Walenz BP, Axelrod N, Huang J, Kirkness EF, Denisov G, Lin Y, MacDonald JR, Pang AWC, Shago M, Stockwell TB, Tsiamouri A, Bafna V, Bansal V, Kravitz SA, Busam DA, Beeson KY, McIntosh TC, Remington KA, Abril JF, Gill J, Borman J, Rogers Y-H, Frazier ME, Scherer SW, Strausberg RL, Venter JC: **The Diploid Genome Sequence of an Individual Human.** *PLoS Biol* 2007, **5**:e254.
4. Yotova V, Lefebvre J-F, Moreau C, Gbeha E, Hovhannesyan K, Bourgeois S, Bédarida S, Azevedo L, Amorim A, Sarkisian T, Avogbe P, Chabi N, Dicko MH, Santa Amouzou ESK, Sanni A, Roberts-Thomson J, Boettcher B, Scott RJ, Labuda D: **An X-linked haplotype of Neandertal origin is present among all non-**

African populations. *Mol Biol Evol* 2011, **28**:1957-1962.

5. Shimada MK, Panchapakesan K, Tishkoff SA, Nato AQ, Jr., Hey J: **Divergent haplotypes and human history as revealed in a worldwide survey of X-linked DNA sequence variation.** *Mol Biol Evol* 2007, **24**:687-698.
6. Cox MP, Mendez FL, Karafet TM, Pilkington MM, Kingan SB, Destro-Bisol G, Strassmann BI, Hammer MF: **Testing for Archaic Hominin Admixture on the X Chromosome: Model Likelihoods for the Modern Human RRM2P4 Region From Summaries of Genealogical Topology Under the Structured Coalescent.** *Genetics* 2008, **178**:427-437.
7. Evans PD, Mekel-Bobrov N, Vallender EJ, Hudson RR, Lahn BT: **Evidence that the adaptive allele of the brain size gene microcephalin introgressed into Homo sapiens from an archaic Homo lineage.** *Proc Natl Acad Sci USA* 2006, **103**:18178-18183.
8. Stefansson H, Helgason A, Thorleifsson G, Steinthorsdottir V, Masson G, Barnard J, Baker A, Jonasdottir A, Ingason A, Gudnadottir VG, Desnica N, Hicks A, Gylfason A, Gudbjartsson DF, Jonsdottir GM, Sainz J, Agnarsson K, Birgisdottir B, Ghosh S, Olafsdottir A, Cazier JB, Kristjansson K, Frigge ML, Thorgeirsson TE, Gulcher JR, Kong A, Stefansson K: **A common inversion under selection in Europeans.** *Nat Genet* 2005, **37**:129-137.
9. Hardy J, Pittman A, Myers A, Gwinn-Hardy K, Fung HC, de Silva R, Hutton M, Duckworth J: **Evidence suggesting that Homo neanderthalensis contributed the H2 MAPT haplotype to Homo sapiens.** *Biochem Soc Trans* 2005, **33**:582-585.
10. Green RE, Krause J, Briggs AW, Maricic T, Stenzel U, Kircher M, Patterson N, Li H, Zhai W, Fritz MH-Y, Hansen NF, Durand EY, Malaspinas A-S, Jensen JD, Marques-Bonet T, Alkan C, Prufer K, Meyer M, Burbano HA, Good JM, Schultz R, Aximu-Petri A, Butthof A, Hober B, Hoffner B, Siegemund M, Weihmann A, Nusbaum C, Lander ES, Russ C, Novod N, Affourtit J, Egholm M, Verna C, Rudan P, Brajkovic D, Kucan Z, Gusic I, Doronichev VB, Golovanova LV, Lalueza-Fox C, de la Rasilla M, Fortea J, Rosas A, Schmitz RW, Johnson PLF, Eichler EE, Falush D, Birney E, Mullikin JC, Slatkin M, Nielsen R, Kelso J,

Lachmann M, Reich D, Paabo S: **A Draft Sequence of the Neandertal Genome.** *Science* 2010, **328**:710-722.

11. Reich D, Green RE, Kircher M, Krause J, Patterson N, Durand EY, Viola B, Briggs AW, Stenzel U, Johnson PLF, Maricic T, Good JM, Marques-Bonet T, Alkan C, Fu Q, Mallick S, Li H, Meyer M, Eichler EE, Stoneking M, Richards M, Talamo S, Shunkov MV, Derevianko AP, Hublin J-J, Kelso J, Slatkin M, Paabo S: **Genetic history of an archaic hominin group from Denisova Cave in Siberia.** *Nature* 2011, **468**:1053-1060.
12. Cann HM, de Toma C, Cazes L, Legrand MF, Morel V, Piouffre L, Bodmer J, Bodmer WF, Bonne-Tamir B, Cambon-Thomsen A, Chen Z, Chu J, Carcassi C, Contu L, Du R, Excoffier L, Ferrara GB, Friedlaender JS, Groot H, Gurwitz D, Jenkins T, Herrera RJ, Huang X, Kidd J, Kidd KK, Langaney A, Lin AA, Mehdi SQ, Parham P, Piazza A, Pistillo MP, Qian Y, Shu Q, Xu J, Zhu S, Weber JL, Greely HT, Feldman MW, Thomas G, Dausset J, Cavalli-Sforza LL: **A human genome diversity cell line panel.** *Science* 2002, **296**:261-262.
13. Thorisson GA, Smith AV, Krishnan L, Stein LD: **The International HapMap Project Web site.** *Genome Res* 2005, **15**:1592-1593.
14. Fujita PA, Rhead B, Zweig AS, Hinrichs AS, Karolchik D, Cline MS, Goldman M, Barber GP, Clawson H, Coelho A, Diekhans M, Dreszer TR, Giardine BM, Harte RA, Hillman-Jackson J, Hsu F, Kirkup V, Kuhn RM, Learned K, Li CH, Meyer LR, Pohl A, Raney BJ, Rosenbloom KR, Smith KE, Haussler D, Kent WJ: **The UCSC Genome Browser database: update 2011.** *Nucleic Acids Res* 2010, **39**:D876.
15. Yoder M, Hildebrand JD: **Shroom4 (Kiaa1202) is an actin-associated protein implicated in cytoskeletal organization.** *Cell Motil Cytoskeleton* 2007, **64**:49-63.
16. Hamosh A, Scott AF, Amberger JS, Bocchini CA, McKusick VA: **Online Mendelian Inheritance in Man (OMIM), a knowledgebase of human genes and genetic disorders.** *Nucleic Acids Res* 2005, **33**:D514-517.