氏 名	深 川 竜 郎

学位（専攻分野）　　　博士（理学）

学 位 記 番 号　　　総研大甲第１５９号

学 位 授 与 の 日 付　　　平成 7 年 9 月２８日

学 位 授 与 の 要 件　　　生命科学研究科　遺伝学専攻

　　　　　　　　　　　学位規則第４条第１項該当

学 位 論 文 題 目　　　Studies on a boundary of long-range G+C% mosaic

　　　　　　　　　　　domains in the human genome: characterization of

　　　　　　　　　　　pseudoautosomal boundary-like sequence(PABL)

　　　　　　　　　　　found near the boundary

論 文 審 査 委 員　　　主 査　教 授　　五 條 堀　　　孝

　　　　　　　　　　　　　　　教 授　　廣 瀬　　進

　　　　　　　　　　　　　　　教 授　　小 川 智 子

　　　　　　　　　　　　　　　助教授　　城 石 俊 彦

　　　　　　　　　　　　　　　教 授　　岡 田 典 弘（東京工業大学）

論文内容の要旨

Studies on a boundary of long-range G+C % mosaic domains in the human genome: characterization of pseudoautosomal boundary-like sequence (PABL) found near the boundary

The human genome, like those of warm-blooded vertebrates in general, is composed of long-range G+C% (GC%) mosaic structures related to chromosome bands. Several groups showed that the Giemsa-dark G bands are composed mainly of AT-rich sequences, and T bands (a subgroup of Giemsa-pale R bands) mainly of GC-rich sequences: ordinary R bands are heterogeneous and appear to be intermediate. Gene density, CpG island density, codon usage, chromosome condensation, DNA replication timing, repeat sequence density, and other chromosome behaviors such as recombination and mutation rate are related to chromosome bands and to long-range GC% mosaic domains. Gene-dense T bands with loose chromatin structures replicate early in S phase and are rich in *Alu* repeats, while G bands with condensed chromatin structures replicate late and are rich in LINE-1 repeats. Because chromosome bands are structures observed with microscopes, precise location of their boundaries may seem meaningless. However, considering various genome behaviors connected with chromosome bands, it appears possible to precisely locate their boundaries by putting informative landmarks on genome DNA. Boundaries may be structurally assigned as clear GC% transition points, and signals for punctuating and/or differentiating respective functions (e.g., a switching signal from early to late DNA replication) may be found in the boundaries.

The human major histocompatibility complex (MHC) has classes I (about 2 Mb), III (1 Mb), and II (1 Mb) from telomere to centromere. Ikemura and his colleagues had found the human MHC to be a typical example of long-range GC% mosaic structures by analyzing MHC sequences compiled by GenBank database and predicted a possible boundary of the Mb-level domains within an under-characterized 450 kb containing the junction of MHC classes II and III. To clone the mosaic boundary, bidirectional chromosome walking from the class III CYP21 to the class II and from the class II HLA-DRA to the class III was conducted in this study, and contiguous clones covering the 450-kb region bridging classes II and III were obtained. To analyze base-compositional distribution of the walked area, insert DNAs of the clones were purified and digested by nuclease P1, and GC% was measured by a HPLC method. About 150 kb from the HLA-DRA was fairly homogeneously AT-rich (mostly less than 40% GC) showing extension of AT-rich sequences from the class II side. Then a sharp transition to about 50% GC occurred and this GC-rich level continued to the class III CYP21. To analyze the structures near and at the GC% transition, the cosmid and λ clones harboring the transitional region were sequenced. The following three types of characteristic structures were found; *Alu* repeats densely clustered in a 20-kb region, five LINE-1 repeats also

clustered in a 30-kb region, and a sequence highly homologous with the pseudoautosomal boundary sequence of the short arms of the human sex chromosomes (PABX1 and PABY1); PABX1 and PABY1 are the interface sequences between sex-specific and pseudoautosomal regions. The author designated the sequence highly homologous with PABXY1 "PABL". There exists a possibility that the organization, a dense *Alu* cluster - a dense LINE-1 cluster - a PABL, is one characteristic of certain types of long-range GC% mosaic boundaries and of band boundaries. The author focused on characterization of the newly found sequence, PABL.

Human sex chromosomes are divided into two functionally distinct regions, sex-specific sequences and pseudoautosomal regions (PARs); within each male meiosis, X and Y chromosomes exchange DNA sequences by homologous recombination in PARs and thus PAR sequences are practically identical between X and Y chromosomes because of this obligatory recombination. The interface between PAR1 (about 2.6 Mb) and the sex-specific region is the pseudoautosomal boundary (PAB1) and therefore PAB1 is the proximal (centromeric) limit to X-Y homologous recombination in PAR1: PAB1 is very unique in the human genome as a strict physical site at which unusually high frequency recombination in the 2.6 Mb of PAR1 (known to be 20-fold greater than the genome average) terminates abruptly. Ellis and his colleagues reported sequences around the interface, i.e., PABX1 and PABY1 sequences. Interestingly, the sequence found near the boundary of the long-range GC% mosaic domains in the MHC is highly homologous (about 80% nucleotide identity) with the PABXY1 sequences which constitute the functional interface in the sex chromosomes. Using the sequence in the MHC as a probe, multiple copies of pseudoautosomal boundary-like sequences (PABLs) were detected through Southern blot hybridization against genomic DNAs and cosmid cloning. The author defined a ca. 650-nt consensus sequence of the PABL core by determining and comparing eleven independent PABL sequences.

Although GenBank genomic sequences showing significant homology with the PABLs were confined to PABXY1, several human ESTs (expressed sequence tags) showed evident homology with separate portions of PABLs, indicating some, if not all, PABLs are transcribable. To clarify characteristics of the predicted PABL transcripts, six human cDNA libraries of different tissues and cells were screened using the PABL segment as a probe. Positive clones were obtained from all six libraries. Sequence analysis of the six cDNA clones showed there exists again a 650-nt core sequence which corresponds to that defined by the genomic PABLs. No ORFs with significant sizes could be found for the obtained cDNAs, not only for the PABL core sequences but also their flanks. When the cDNA sequences were searched with the BLASTX program against the protein sequence database, no significant homology with known proteins was detected. GRAIL, a computer program trained to identify protein-coding ORFs in human DNA, also could not detect reliable protein-coding capacity. These may suggest that their functional form, if present, is RNA molecules. To estimate intact sizes of PABL transcripts, northern blot

analysis of human total or polyA$^+$ RNA fraction was conducted using the PABL probe. Broad bands, estimated to be 5-10 kb in length, were detected.

For study of evolutionary processes involved in forming the present PABXY1 and PABLs, their phylogenetic relationships were examined. In order to estimate evolutionary rates, the reported PABXY1 sequences of great apes and Old World monkeys were included; divergence between great apes and Old World monkeys is postulated here to be 25 million years ago. Phylogenetic trees were constructed using the neighbor-joining method. Using the evolutionary distance between human and Old World monkey PABX1, as well as that for PABY1, divergence time of PABLs including PABXY1 was estimated to be 60-120 million years: this is consistent with the result obtained through Southern hybridization that PABLs are present in the bovine genome. The evolutionary rates of individual PABLs and PABXY1 were then calculated using the divergence time. The rates of some PABLs were far less than 1×10^{-9} substitutions per site per year, indicating evolutionary and functional constrains were executed on PABL sequences. Taking phylogenetic relationship between PABLs and PABXY1 into consideration, evolutionary process in the formation of the present pseudoautosomal boundary PAB1 is proposed by postulating an illegitimate recombination between two PABLs.

ヒトゲノムに存在する巨大Ｇ+Ｃ％区分構造の境界に関する研究：境界に見い出した偽常染色体部位境界様配列(PABL)の解析

　　ヒトを始めとする温血脊椎動物のゲノムでは、Ｇ＋Ｃが豊富なＤＮＡ領域とそうでない領域から構成されるモザイク構造になっていることが知られている。この構造は、巨大Ｇ＋Ｃ％区分構造ともいわれ、染色体のバンド構造とも関係すると考えられている。しかし、これらのモザイク構造が進化的にどのように形成されたものであるのか、またそれらが染色体構造における機能、ＤＮＡ複製や遺伝子発現などの関連をはじめとして、どのような生物学的機能をもっているのかも全くわかっていない。

　　申請者は、ヒトゲノムに存在するこのような巨大Ｇ＋Ｃ％区分構造の生物学的意義を解明する試みの一つとして、Ｇ＋Ｃが豊富な領域とそうでない領域の境界に注目し、（１）これらの境界領域では、Ｇ＋Ｃ％が穏やかに変化しているのかあるいは急激に変化しているのか、（２）これらの境界領域を特定する何らかの遺伝的シグナルが存在するかどうか、（３）巨大Ｇ＋Ｃ％区分構造の進化的形成過程における境界領域の役割は何か、を明らかにするという研究目的を具体的に設定した。この研究目的を具体的に達成するため、申請者は、巨大Ｇ＋Ｃ％区分構造が存在することが知られているヒトの主要組織適合性複合体（ＭＨＣ）領域のクラスⅡとクラスⅢの境界領域を実験的に解析した。特に、クラスⅢのＣＹＰ２１遺伝子からクラスⅡの方へとクラスⅡのＨＬＡ－ＤＲＡ遺伝子からクラスⅢの方へと、両方の染色体ウォーキングを実行し、クラスⅡとⅢを繋ぐ４５０ｋｂのＤＮＡの領域をカバーする隣接クローンの作成に成功した。そして、Ｐ１ヌクレアーゼとＨＰＬＣ法による解析によって、それらの境界領域のＧ＋Ｃ％の直接的測定を実行した。その結果、以下のような興味深い知見を得た。（１）境界領域ではＧ＋Ｃ％が急激に変化することを明らかにし、（２）Ｇ＋Ｃ％の変移点には大規模な*Ａｌｕ*とＬＩＮＥ－１クラスターならびに偽常染色体部位境界配列（ｐｓｅｕｄｏａｕｔｏｓｏｍａｌ　ｂｏｕｎｄａｒｙ：　ＰＡＢ）と相同性の高い配列が存在することを発見した。（３）ＰＡＢは、性染色体の性特異的部位と偽常染色体部位（ｐｓｅｕｄｏａｕｔｏｓｏｍａｌ　ｒｅｇｉｏｎ：　ＰＡＲ）の境界として定義され、ＰＡＲが頻繁に相同組換えを起こすことが知られていることから、その区分構造の進化的形成に関する新しいモデルを提唱し意義ある議論を行なった。

　　以上のように、本論文は、ヒトゲノムにおける巨大Ｇ＋Ｃ％区分構造を理解する上で、新しくかつ重要な知見を明らかにしたもので、学位論文として価値があると判断した。特に、本論文は、研究業績が特に優れていると認定した。なお、これらの成果の一部は、Ｇｅｎｏｍｉｃｓに発表されている。