

博士論文の要旨

氏名 岩崎 理紗

論文題目 胃がんのリスクアレルを通して探る、日本人の遺伝的な多様性及び集団動態の解析
Study of genetic variation and demographic history of the Japanese from the viewpoint of a risk allele of gastric cancer

がんは人類にとって主要な死因の一つであり、その中でも、胃がんは人類全体で発症率が第5位、死亡率が第3位と非常に高い順位を占める。また、胃がんの発症率には人類集団の中でも偏りが見られ、特に日本をはじめとする東アジア集団での発症率が高いことはよく知られている。胃がんの中でも、特に diffuse type の胃がん (DGC) は、近年発症数が増加傾向にあることが知られている。また、日本人を対象としたゲノムワイド関連解析によって、rs2294008 (T/C) の T アレルは、DGC の発症リスクと強い関連性を持つアレル (リスクアレル) であることが報告された。このリスクアレルは 8 番染色体に位置し、様々ながんに関連する *PSCA* 遺伝子の開始コドンの一部である。このリスクアレルを含む配列がタンパク質に翻訳されると、ノンリスクアレル (C アレル) を含む配列を翻訳した場合に比べて 9 アミノ酸長くなる。この長い *PSCA* タンパク質は、細胞増殖のシグナルとして働くため、過剰な細胞増殖を招き、がんが進行すると考えられてきた。

このリスクアレルと DGC の関連性は、世界中の様々な集団で報告されているが、日本人はこのリスクアレルの頻度が、遺伝的に近縁な東アジアの集団よりも高い (日本人:0.63、中国の漢族:0.25、台湾人:0.26、韓国人:0.50) ことが知られている。rs2294008 の T アレルを持つことは、その個体の DGC の発症リスクが上がるため、生存上不利になると考えられる。そこで本研究では、集団遺伝学的な観点から、なぜ JPT だけでこのリスクアレルの頻度が高く保たれているのかについて検討した。

第1章では、DGC の発症率と rs2294008 の各アレルの生物学的な機能について、先行研究で知られていることを概観した。また、本研究で用いた正の自然選択のシグナルを検出する検定や、日本人の集団動態モデルについてまとめ、本研究の目的を示した。

第2章では、本研究で用いた各種解析手法について記載した。

第3章では、JPT でリスクアレルが何故高頻度に維持され、近縁な集団間で頻度差が保たれているのか、様々な検定を用いて解析結果を示した。まず、このリスクアレルの頻度差がどの程度なのか調べるため、ゲノムワイド SNP を比較対象として、2 集団間の頻度差を F_{ST} を指標に比較した。対象集団には、1000 人ゲノムプロジェクトの集団のうち、日本人集団 (JPT) と最も遺伝的に近い漢族 (CHB) を用いた。 F_{ST} は集団間の頻度の差を用いて、集団の分化の程度を測定し、頻度の差が大きいほど F_{ST} の値も大きくなる。この結果、rs2294008 及び連鎖関係にある近傍の SNP は、ゲノムワイドレベルで最も大きな F_{ST} 値を示し、集団の分化に貢献する領域を形成していることがわかった。

次に、非常に大きな頻度差が遺伝的に近縁な2集団間で生じた原因について調べた。近縁な集団間で大きな頻度差が維持されている原因は、中立下で強く遺伝的浮動の影響を受けた場合か、または正の自然選択によってターゲットとなるアレル頻度が上昇した場合であると考えられる。そこで、正の自然選択が頻度差に貢献している場合を次の3つの場合について検討した。(1) JPTでリスクアレルに自然選択が働いている可能性、(2) CHBでノンリスクアレルに自然選択が働いている可能性、(3) 上記のすべてである可能性についてである。その結果、いくつかの中立性検定では正の自然選択のシグナルが検出されなかったが、2D SFSを利用した解析によって、CHBでノンリスクアレルに対する正の自然選択のシグナルを検出することができた。一部の中立性検定で自然選択のシグナルが検出できなかった原因は、自然選択のターゲットサイトとなった rs2294008 の近傍に組み換えのホットスポットが存在しているため、組み換えによってターゲットサイトを含むハプロタイプの連鎖関係が崩れたためだと考えられる。さらに詳しく調べると、CHBではノンリスクアレルを持つ配列は2種類のハプロタイプ(サブハプロタイプ)で構成されており、この両方に正の自然選択が働いていることがわかった。以上より、CHBのノンリスクアレルに検出されたシグナルは、複数のハプロタイプ(C-A及びA-Gサブハプロタイプ)に対して同時に自然選択が働く、soft sweepのシグナルであることが分かった。一方で、JPTではリスクアレル単位でも、ノンリスクアレル単位でも自然選択のシグナルは検出されなかったが、CHBと同様のサブハプロタイプは集団内に存在していた。2つのサブハプロタイプに対しても2D SFSで正の自然選択のシグナルの検出を行ったところ、JPTでは正の自然選択がC-Aサブハプロタイプに対してのみ働いていた。また、JPT及びCHBで自然選択が働き始めた時期は、両集団の分岐年代よりも早いことがわかった。以上のことから、JPTではいずれかの時期に起こった選択圧の変化によって、A-Gサブハプロタイプに働いていた自然選択がリラックスまたは停止し、C-Aサブハプロタイプに対して自然選択が働き続けた(hardening)可能性が示唆された。選択がリラックス/停止したサブハプロタイプの多様性によって、残ったサブハプロタイプのselective sweepのシグナルが覆い隠され、JPTではノンリスクアレル単位の自然選択のシグナルが検出されなかったと考えられる。以上の結果は、仮説(2)を部分的に支持した。しかしながら、これらの解析結果だけでは、CHBでのノンリスクアレルの頻度の上昇を説明できても、自然選択下になくリスクアレルがJPTで何故高頻度のままであるのか説明することはできなかった。

そこで、次に日本人の集団動態による可能性を検証した。検証にはアレル頻度シミュレーションを行い、集団動態のモデルには、「二重構造モデル」を利用した。このモデルは日本人の遺伝的・表現型的な多様性を、縄文系統の人類集団と、大陸由来の渡来系弥生人集団との交雑によって説明する。この結果、現生の日本人の高頻度のリスクアレルは、縄文系統の人類集団に多く由来したものである可能性があり、縄文系統の人類集団は、現生の日本人より高頻度でリスクアレルを持っていたことも示唆された。そこで、縄文系統の人類集団で現在利用可能な3ゲノム配列を利用し、JPTのリスクアレルを含んだハプロタイプと、縄文系統のハプロタイプの系統関係を調べた。この結果、縄文系統の人類集団の配列はリスクアレルを持っており、ハプロタイプはJPTに見られるハプロタイプと最も近縁であった。

以上の結果より、日本人でDGCのリスクアレルが高頻度に至ったのは、(I) A-G及び

C-A サブハプロタイプには自然選択が働いていたが、A-G サブハプロタイプに対する選択が弱まり、ノンリスクアレルの頻度上昇が抑えられ、かつ、(II) 日本人は祖先集団である縄文系統の人類集団から、高頻度のリスクアレルを含むハプロタイプを受け継いだ、複合的な要因にあると結論づけた。

更に、A-G 及び C-A サブハプロタイプに対する自然選択が世界中の他の人類集団で働いているかどうかについても調べた。この結果、C-A サブハプロタイプに対する自然選択は、韓国人集団を除くほとんどの東アジア集団で働いていた。対照的に、A-G サブハプロタイプに対する自然選択は、アフリカの集団で共通に働いていたが、ヨーロッパ集団や、非漢族の東アジア集団では働いていなかった。以上の結果は、遺伝的に近い人類集団間であっても、2つのサブハプロタイプの selection status が動的に変わることを示している。

第4章では、第3章までの結果をまとめ、以下の(1) A-G/C-A サブハプロタイプに働いている自然選択の特徴、(2) A-G/C-A サブハプロタイプを通して、世界各地の集団の selection status の多様性、(3) 東アジア集団での DGC のリスクアレルの歴史の再構築の3点について論じた。このうち、特に(2)については、rs2294008 及びこの近傍の領域では、人類の歴史上、自然選択のターゲット化とターゲットから外れることを繰り返していることがわかった。このことは、この遺伝子座に対する選択圧が、何らかの要因で大きく変動を繰り返したことを示唆している。

第5章では、本研究の意義について議論した。

最後に、第6章では、DGC のリスクアレルに関する研究の更なる発展の可能性や、本研究を通して得た知見を元にした新たな研究の可能性について論じた。

本研究は、日本人で DGC のリスクアレルがなぜ高頻度に至ったのかについて解明を試みた。本研究は、DGC を通して、生物学的にも重要な役割を持つ領域に対し、遺伝的に近い集団間でも自然選択のターゲットが変化し、あるいは遺伝的に遠い集団でも独立に同じアレルがターゲットになり得る、動的で複雑なプロセスを示せた点において、面白い知見を提示できたと考える。

