

論文の内容の要旨

論文題目 ゲノムワイド関連解析による虚血性心疾患の遺伝的背景に関する研究

氏名 松永 紘

<序文>

虚血性心疾患 (coronary artery disease; CAD) は、世界中で主要な死因の一つである。疫学研究により、高血圧や脂質異常症、糖尿病などのリスクファクターが報告されてきた。CAD の有病率は人種によって異なっており、心筋梗塞 (myocardial infarction; MI) での年齢調節後発症率は日本人では 10 万人あたり 27 人であるのに対し、欧米人では 42 から 142 人と報告されている。日本人の食生活の欧米化により、MI の発症率は増加傾向にはあるが、欧米人集団と比べるとまだ低値である。この違いは、環境要因だけでなく、遺伝要因による違いも関与している可能性があることを示している。

CAD など多くの遺伝子が発症に関わる多因子疾患にはゲノムワイド関連解析 (genome-wide association study; GWAS) が有効となる。現在までに CAD に対する GWAS が多く行われてきており、150 以上の疾患感受性領域が報告されている。しかし、報告されている GWAS の大部分は CARDIoGRAMplus C4D Consortium や UK Biobank など欧米からの報告であり、日本では大規模な GWAS は行われていなかった。本研究においては、約 6 万人の日本人の検体を用いて、CAD の発症に関与する遺伝子をさらに明らかにし、欧米人との人種差を検討することを目的に GWAS を行なった。

<方法>

虚血性心疾患の遺伝要因を解明するために、日本人集団における 2 種類の GWAS を行い、メタ解析を行なった。1 群では 12,494 人の MI 患者および 28,879 人の対照群を、2 群では MI 患者 2,808 人および対照群 7,261 人を用いた。さらに、日本人集団の GWAS 結果と欧米人集団の GWAS の結果を統合し、メタ解析を行なった。第 1 群および第 2 群の患者の DNA サンプルを、Illumina HumanOminExpress および HumanOminExpressExome BeadChip を用いて、一塩基多型 (single nucleotide polymorphism; SNP) のジェノタイピングを行なった。ジェノタイピングを行なった検体に対して、不適切な SNP (Call rate、マイナーアレル頻度、Hardy-Weinberg 平衡検定、クラスタープロットの結果に基づく) を除去し、厳しいクオリティーコントロール (Quality control; QC) を施行した。さらに、IBD (identity by descent) により第一および第二近親者を除外し、主成分解析により東アジア人集団から外れている検体を除去した。QC 後に 1000 人ゲノムプロジェクト (1000 Genomes Project) の東アジア人集団の遺伝子型データを参照にインピュテーション (Imputation) を行なった。Imputation の精度が低い SNP ($R^2 < 0.7$) を除外し、最終的に第 1 群で 5,968,989 個の SNP、第 2 群で 5,936,278 個の SNP を対象に解析を行なった。各 SNP にお

いて、ロジスティック回帰分析 (additive model) を行った。GWAS における有意水準である、 $P < 5 \times 10^{-8}$ を満たした SNP を有意な SNP と判断した。また、メタ解析では fixed effect model や random effect model よりも検出力が高いと示されている MR-MEGA という方法を使用した。

GWAS で検出されたシグナルに関連した遺伝子を明らかにするために、遺伝子発現量やメチル化に影響を与える領域を複数のデータベースで検索した。さらに、われわれは GWAS で得られたシグナルと臨床検査値との関連を調べるために、日本人集団での 60 個の検査値の GWAS の結果の探索も行った。

同定されたシグナルに関連したパスウェイを発見するために、Gene Set Enrichment Analysis (GSEA) という手法を用いて、CAD 発症に関連したパスウェイを探索した。また、CAD の発症に関与が大きい組織や細胞を検討した。欧米人のデータを加えたメタ解析の結果より $P < 1.0 \times 10^{-5}$ を満たす独立した SNP を取り出し、これらの SNP の近傍にある遺伝子を検索した。37,427 人のマイクロアレイの結果より、これらの遺伝子が多く発現する組織や細胞を検索した。最後に、日本人の GWAS と欧米人の GWAS の結果より、遺伝要因における人種間の違いを検討した。

<結果>

はじめに日本人集団で 2 つのコホートを用いて GWAS を行なった。クオリティーコントロールを行ったにもかかわらず、集団の構造化を評価する genomic inflation factor (λ_{GC}) の上昇が見られたため、補正を行なった。そして、2 種類の GWAS を固定効果モデル (fixed effect model) を用いて、5,932,413 個の SNP でメタ解析を施行した。メタ解析では GWAS での有意水準である $P < 5.0 \times 10^{-8}$ を満たした領域が 18 個検出された。検出された領域のうち、16 個は既報の領域であった。一方、11q22 と 12q24 は新規の疾患感受性領域の候補であった。しかしながら、12q24 領域のリード SNP である rs11066542 は 1MB 以内に既報の領域はなかったが、2MB 以内に既報の *BRAP-ALDH2* 領域が存在した。この領域の影響の可能性を考慮し、条件的分析 (conditional analysis) を行なったところ、シグナルは消失した。したがって、日本人集団でのメタ解析では rs1848599 をリード SNP とする 11q22 領域のみが新規領域として検出された。

CAD に関連した領域をさらに同定することを目的に、欧米人集団も加えたメタ解析を行なった。日本人の 2 種類の GWAS と CARDIoGRAMplusC4D と UK Biobank をメタ解析した欧米人集団の結果を統合し、5,167,567 個の SNP を用いた。この解析では、76 個の領域を検出し、そのうち 3 個が新規の領域であった。11q22 領域は欧米人集団とのメタ解析でも検出され、それぞれのメタ解析のリード SNP が 1MB 以内であった。そのため、conditional analysis を行なったところ、この 2 つの領域は同一領域と考えられた。最終的に 3 領域 (1q21、10q26、11q22) が新規領域と考えられた。

rs6587520 は 1q21 領域のリード SNP であり、CTSS のイントロンに存在する。CTSS はリ

ソソームに局在しているシステインプロテアーゼであるカテプシン S をコードしている。rs6587520 は左室と大動脈において CTSS の遺伝子発現を、白血球において CTSK の遺伝子発現を調節している領域だとわかった。また、rs6587520 は複数の免疫細胞で CTSS の発現上昇と関連していた。

rs2257129 は 10q26 領域のリード SNP であり、*WDR11* と *FGFR2* の遺伝子間領域に存在する。今回の領域はエンハンサー領域を示唆するモノメチル化 (H3K4me1) のピークとオーバーラップしていたが、*WDR11* や *FGFR2* などの遺伝子発現量を調節する領域であることは確認できなかった。しかし、臨床検査値との関係を探索したところ、CAD 発症のリスクファクターである HDL コレステロールの減少と中性脂肪の上昇との関連がみられた。

rs10488763 は 11q22 領域のリード SNP であり、*RDX* と *FDX1* の遺伝子間領域に存在している。rs10488763 にヒストン修飾は確認できなかったが、rs10488763 と完全な連鎖関係にある rs1443120 は脂肪細胞や血球細胞でプロモーター領域に存在していた。また、rs10488763 は白血球と単球で *RDX* の遺伝子発現を、白血球で *RDX* のメチル化に影響を与えている領域であるとわかった。

パスウェイ解析では、25 個の有意なパスウェイを同定した。脂質関連がもっとも頻度が高く、輸送系やシグナリング経路が続いた。この結果は CAD の進展に脂質代謝以外にも様々な生物学的プロセスが関わっていることを示唆した。CAD 発症に関与が大きい組織や細胞を探索したところ、もっとも関連が大きかったのは動脈であった ($P = 3.05 \times 10^{-6}$)。また、代謝や免疫、血圧の調整に重要な役割を果たす副腎も有意であった ($P = 1.31 \times 10^{-5}$)。さらに、脂質代謝に関わる脂肪細胞とも関連が見られた ($P = 1.78 \times 10^{-5}$)。

日本人集団と欧米人集団の MI への遺伝的感受性の相違を検討した。欧米人も加えたメタ解析で検出された 76 個の疾患感受性領域に関して、リード SNP の MI 発症に対する効果を日本人集団と欧米人集団でオッズ比を用いて比較した。大部分の遺伝子座でオッズ比は類似しており、2 人種間でオッズ比は中等度の相関を認めた (スピアマンの相関係数 = 0.56、 $P = 1.21 \times 10^{-7}$)。しかし、*ABO* は欧米人集団で MI 発症に関わる影響が大きく、7 個の領域 (*APOB*、*FNI*、*ATF6B*、*HDAC9*、*UBE3B*、*RPH3A*、*ADAMTS7*) では日本人集団で影響が大きいという結果が得られた。

最後に CAD 発症に関わる組織や細胞で人種差があるか検討した。リスクアレルの頻度により、日本人を含む東アジア人集団で頻度が多いグループと欧米人集団で頻度が多いグループに分割した。東アジア人集団では副腎や副腎皮質が優位となり、欧米人集団では動脈や脂肪組織が優位となった。リスクアレルの頻度の違いは人種間の有病率の違いに寄与する可能性がある。そのため、これらの結果は東アジア人集団と欧米人集団での相違を示唆している可能性が考えられた。

<考察>

われわれは現在までに報告されている中で日本人集団で最大規模の CAD に対する GWAS

を行った。さらに、日本人の GWAS に CARDIoGRAMplusC4D と UK Biobank をメタ解析した欧米人集団の結果を統合することにより、人種横断的な大規模なメタ解析も行なった。最終的に CAD に対する疾患感受性領域を 76 個発見し、そのうち 3 個が新規であった。検出されたリード SNP を用いて、日本人集団と欧米人集団を比較し、遺伝要因による人種差を明らかにした。

新規の疾患感受性領域の 3 個のうち 2 個は動脈硬化の進展に重要な役割を果たす免疫系の関与が示唆された。ゆえに、われわれが発見した新規の疾患感受性領域は CAD 発症に免疫と動脈硬化が関与していることが強調された。また、パスウェイ解析や CAD 発症に強く関わる組織・細胞を検索したところ、脂質だけでなく、動脈や副腎などの関連も示唆された。これらの結果より、CAD の進展には様々なメカニズムが関与すると考えられた。

また、日本人集団と欧米人集団で各領域の MI 発症への効果を調べたところ、概ね効果は同等であったが、一部の領域で人種差を認めた。さらに CAD 発症に関わる組織や細胞を比較したところ、日本人を含む東アジア集団では副腎、欧米人集団では動脈や脂肪組織の関与が大きいことが示された。疫学研究からも 2 人種間で脂肪代謝の違いが示唆されており、遺伝要因での関与が考えられた。

この研究は日本人集団で十分な検出力を持った CAD の GWAS であり、日本人集団での多遺伝子リスクスコアを作成するなどオーダーメイド医療に役立つと考えられる。さらに、欧米人集団も含めたメタ解析では日本人集団と欧米人集団での遺伝要因の共通点と相違点を示した。人種間で遺伝的構造の違いがあるため、CAD 発症の遺伝要因を明らかにするには欧米人集団以外での解析がさらに必要となってくるだろう。