

博士論文

論文題目 稀少遺伝子変異の集積が心筋梗塞発症に与える  
影響に関する研究

氏名 田島 知幸

## 要旨

心筋梗塞における遺伝の影響を明らかにするため、これまでにゲノムワイド関連解析(GWAS)で特定された19の心筋梗塞関連領域に含まれる36の遺伝子を対象とし、9,956人の患者群、8,376人の対照群に対して次世代シーケンサーを用いたPCRベースのターゲット・リシーケンスを行った。

結果、*LDLR*, *PCSK9*の2遺伝子への稀少遺伝子変異の集積が心筋梗塞と有意に関連していた。この作用はLDL-C値を介したものであった。*LDLR*への稀少変異の集積と心筋梗塞の関連に関しては2015年2月に西欧人の全エクソーム解析で既に報告されており、この結果の日本人での再現となった。*PCSK9*に関しては世界初の報告となる。

## 目次

1 序文.....	4
2 方法.....	8
2-1 研究デザイン.....	8
2-2 サンプル.....	10
2-3 候補遺伝子選択.....	11
2-4 ライブラリ作成.....	14
2-5 解析パイプラインの作成とデータ解析.....	16
2-6 アノテーション.....	17
2-7 Quality Control.....	17
2-8 レアバリエントと脂質レベル、心筋梗塞発症年齢の解析.....	19
2-9 関連解析.....	19
2-9-1 単変異ごとの関連解析.....	19
2-9-2 遺伝子ごとの解析.....	20
2-10 LDL コレステロール値、心筋梗塞発症年齢の解析.....	22
2-11 脂質異常症の有無による補正.....	22
3 結果.....	24
3-1 パイプラインのバリデーション.....	24
3-2 リシークエンス結果のサマリー.....	24
3-3 単変異ごとの関連解析の結果.....	28
3-4 遺伝子ごとの関連解析の結果.....	29
3-4-1 <i>LDLR</i> について.....	30
3-4-2 <i>PCSK9</i> について.....	31
3-4-3 <i>TCTN1</i> について.....	33
3-5 <i>LDLR</i> , <i>PCSK9</i> の低頻度変異が臨床的パラメータに及ぼす影響.....	33
3-6 <i>LDLR</i> の低頻度変異と LDL-C、心筋梗塞発症年齢との関係.....	33

3-7	<i>PCSK9</i> の低頻度変異と LDL-C、心筋梗塞発症年齢との関係 .....	36
3-8	血清 LDL-C 値と心筋梗塞リスクとの関係 .....	38
3-9	家族性高脂血症との関連 .....	40
4	考察 .....	50

## 1 序文

心筋梗塞は世界中で死因の上位を占める疾病の一つであり<sup>1</sup>、我が国においても死因の第二位を占める重大な疾患の一つである。「心筋梗塞の家族歴」が心筋梗塞発症に関して独立したリスク因子であることは古くから知られているが、その遺伝様式については明らかになっていない。疾患と遺伝の関連を考えた場合、メンデル遺伝病のような稀少疾患は非常に病原性が強くかつ浸透率の高い稀な遺伝子変異によって発症し(rare disease-rare variant hypothesis)、心筋梗塞や糖尿病のような一般的な疾患は一つ一つは病原性の低い多数の変異の影響の重ね合わせによって発症する(common disease-common variant hypothesis)、という2つのモデルが一般的である<sup>2</sup>。

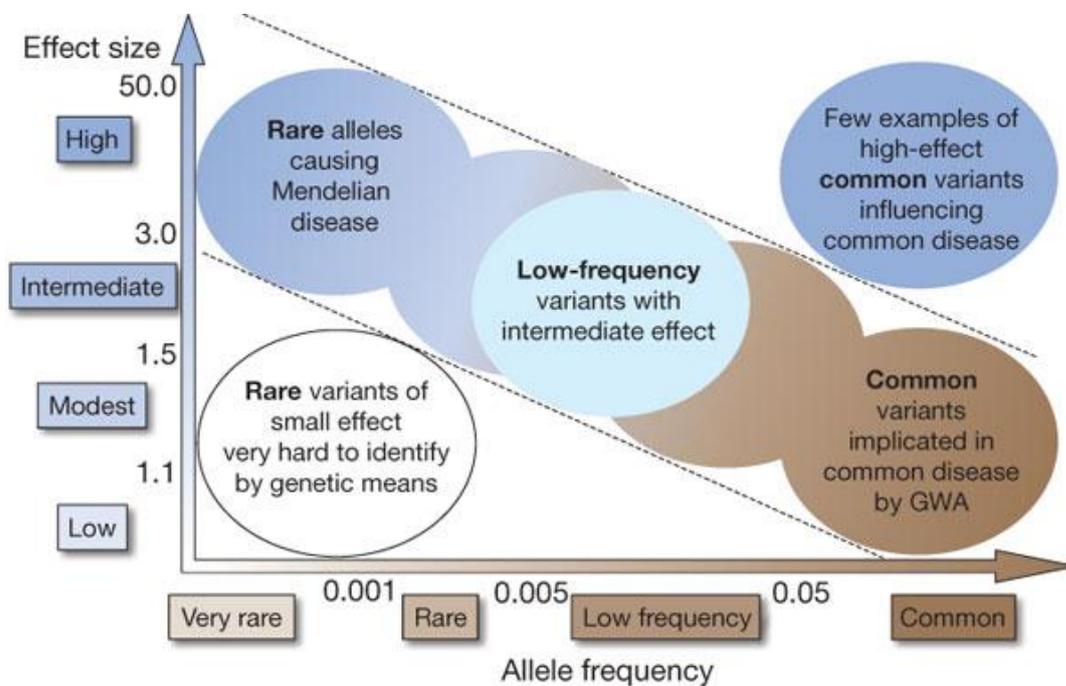


図 1 疾病と原因遺伝子変異の関係  
(参考文献 2 より引用)

現在までに数万人の患者-対照群を対象として一塩基多型 (SNP) アレイチップを用いた大規模なゲノムワイド関連解析 (GWAS) <sup>3-10</sup> が多数行われ、これにより 50 以上の心筋梗塞関連領域がヒトゲノム上に特定されたが、そのすべての影響を考慮しても、双子を対象とした研究から算出された心筋梗塞の遺伝率 <sup>11</sup> のうち 6 割弱しか説明するに至っていない。

この未知の部分(missing heritability)を説明する仮説の 1 つとして **common disease-rare variant** 仮説が提唱されている <sup>2</sup>。これは心筋梗塞のような一般的な疾患においても低頻度で比較的病原性の高い変異の集積により遺伝率の一部が説明されるという仮説である。

こうした低頻度変異と疾患との関連を調べるため、**common disease** の遺伝解析は全エクソームシーケンスやエクソームチップなどによる解析に潮流が移りつつある。その一方で **GWAS** で関連が示された領域を対象としたターゲット・リシーケンスはそのコストに比して重要な知見が得られる可能性から注目されている。

**GWAS** で有意なシグナルがある **common SNP** 周辺の遺伝子を密に調べること、領域内のどの遺伝子が病態に関して本質的な寄与をしているのかを知ることができるほか、**common disease** を対象とした **GWAS** で特定された領域の周囲から遺伝性疾患の原因となる稀少変異が特定され、病態の理解を助けるこ

とも多い<sup>12-17</sup>。

また **rare variant** の研究が病態に関する新たな知見を与え、新薬の開発に繋がる例もある。例えば家族性低コレステロール血症の家系の研究<sup>18-21</sup> から開発された抗 *PCSK9* 抗体や腎性糖尿病の家系の研究<sup>22</sup> から開発された *SGLT2* 阻害薬などがその一例である。

心筋梗塞を対象とした **rare variant** の先行研究としてはサンプル数 1000 程度の小規模のターゲット・リシーケンスが 1 報<sup>23</sup> あるが **negative study** の報告にとどまっている。2015 年 2 月に大規模な全エクソーム解析の結果の報告<sup>24</sup> があり *LDLR* と *APOA5* という脂質に関連した 2 遺伝子への **rare variant** の集積が心筋梗塞と関連することが報告されており、**positive** な報告としてはこれが唯一のものである。

次世代シーケンサーの普及によりシーケンスのコストは低減されたとはいえ、数万人規模の関連解析を行うための経済的コストはいまだに莫大であり、そう簡単に実施できるものではない。

本研究では既知の **GWAS** 関連領域に含まれる遺伝子の翻訳領域を **PCR** により増幅し、次世代シーケンサを用いてシーケンスするターゲット・リシーケンスの手法を十分なサンプルサイズで世界で初めて心筋梗塞に適用し、心筋梗塞の病態に遺伝的に大きな影響を与える変異・遺伝子を特定することでこの

common disease の遺伝様式への理解を深めること、またそれによって新規の治療ターゲットを見出すことを指向して行った。

## 2 方法

### 2-1 研究デザイン

本研究では心筋梗塞サンプルおよび対象サンプルを用いて2ステージのターゲット・リシーケンスを行った。研究デザインの概要を図2に示す。

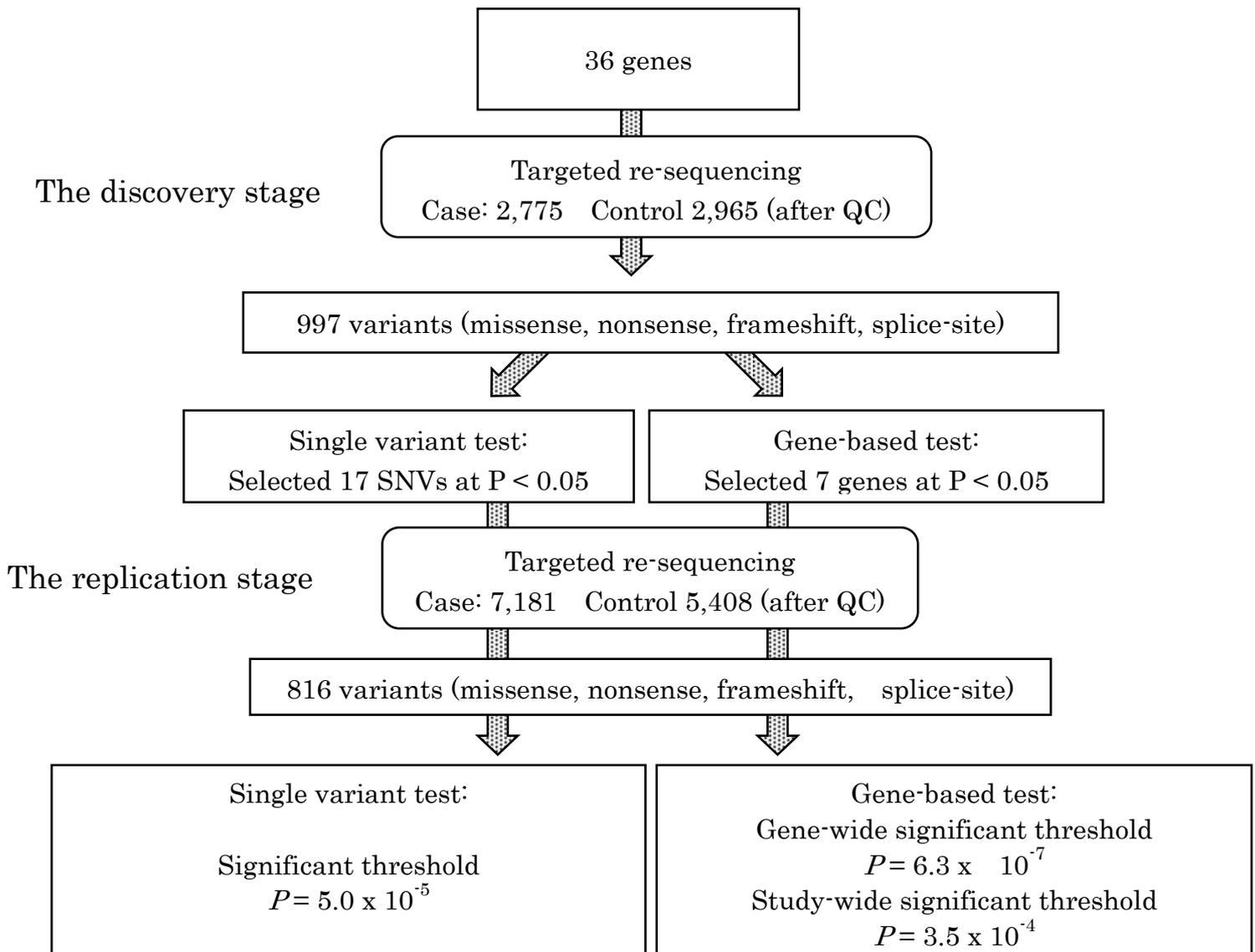


図2 研究デザイン

図のように単変異ごとの関連解析と遺伝子ごとの関連解析を独立に 2 ステージで行った。単変異ごとの関連解析では最初のステージでシーケンスされた 997 の変異で Bon Ferroni の補正をした  $P = 5.0 \times 10^{-5}$  を有意水準とし、遺伝子ごとの解析では遺伝子数を 20,000 と仮定し、変異を 4 グループに分けて解析したことにも加味して  $P = 6.3 \times 10^{-7}$  を gene-wide 有意水準に、当初ターゲットとした 36 遺伝子で Bon Ferroni の補正をした  $P = 3.5 \times 10^{-4}$  を Study-wide 有意水準と定めた。

第一ステージではターゲットとして研究開始段階で既報だった GWAS 領域 49 領域から 19 領域、36 遺伝子を PCR ターゲットとして選択し心筋梗塞患者群 2,775 人、対照群 2,965 人のサンプルに対しリシーケンスを行った。

シーケンスで特定された遺伝子変異はまず個々の変異ごとに関連解析を行うが、基本的にここで想定している変異は集団でのアレル頻度 5% 以下の低頻度変異であるため、単変異での解析で統計的に有意な  $P$  値が得られる可能性は低い。これらの低頻度変異の疾患への影響をみる方法として変異を遺伝子ごとにまとめて解析し、その遺伝子において低頻度変異が疾患群に集積しているかどうかを評価する collapsing method と呼ばれる方法が近年提唱されており、ここでは Cohort Allelic Sum Test (CAST) および Sequence Kernel Association Test (SKAT) と呼ばれる 2 つの代表的な手法で遺伝子ごとの解析も並行して行った。第一ステージが終了した段階で単変異の解析、遺伝子ごとの解析いずれかで  $P$  値が 0.05 を下回った変異・遺伝子のみを Stage2 のリシーケンスのターゲットとし、Stage1 とは独立した患者群 7,181 人、対照群 5,408 人の集団でリシー

クエンスを行った。Stage2 が終了した段階で Stage1 と Stage2 のメタ解析を行い、combined  $P$  があらかじめ設定した  $P$  値の有意水準を下回ったもののみを有意と判断した。解析の対象とした変異は missense, nonsense, frameshift, splice-site などタンパクの機能に変化をもたらす可能性のあるもののみとし、synonymous 変異などは解析から除外した上でアレル頻度 5% 以下のもののみを解析対象とした。

## 2-2 サンプル

第一ステージ、第 2 ステージともに患者群はオーダーメイド医療実現化プロジェクト(BiobankJapan)のサンプルを使用し、心筋梗塞の定義は診療記録の記載に基づき行った。対照群については BiobankJapan において以下の 5 疾患(脳動脈瘤、食道癌、子宮癌、慢性閉塞性肺疾患、緑内障)で登録されたサンプルの混合であり、狭心症、心筋梗塞および心不全の病歴を持つ者を除外した。第 1 ステージの対照群は次の 3 施設からの DNA サンプルを使用した。Pharma SNP Consortium、大阪御堂筋ロータリークラブ、東大病院検診部、前 2 施設は自己申告による持病のない健康なボランティアからの採血由来 DNA サンプルであり、東大病院検診部サンプルは人間ドック受診者のうち虚血性心疾患の既往のあるものを除外したサンプルである。表 1 にサンプルの基本的な情報を示す。

表 1 使用したサンプルの基本的情報

Stage	Phenotype	N	Mean Age (SD)	Mean onset-age of MI (SD)	Source	Mean T-Chol (mg/dL) (SD)	Mean LDL-C (mg/dL) (SD)	Mean HDL-C (mg/dL) (SD)	Mean TG (mg/dL) (SD)
	Cases	2,775	62.7 (10.9)	57.6 (10.8)	BBJ	190.6 (35.9)	111.8 (31.8)	47.5 (12.6)	156.2 (110.5)
Discovery					PSC (N = 963)	NA	NA	NA	NA
	Controls	2,965	56.4 (12.1)	-	MRC (N = 1,060)	NA	NA	NA	NA
					Univ. of Tokyo (N = 942)	208.8 (33.0)	125.7 (46.6)	65.6 (17.4)	113.8 (70.4)
Replication	Cases	7,181	68.0 (9.82)	62.2 (10.4)	BBJ	189.9 (36.5)	111.6 (31.6)	48.9 (14.6)	147.5 (93.7)
	Controls	5,408	62.8 (9.81)	-	BBJ	199.5 (40.6)	119.2 (34.7)	57.5 (17.0)	132.0 (94.8)

PSC、および MRC 由来サンプルについては健常者ボランティアから提供された DNA を使用しており血液検査データは存在しない。

BBJ: Biobank Japan、PSC: ファーマ SNP コンソーシアム、MRJ: 大阪御堂筋ロータリークラブ

### 2-3 候補遺伝子選択

本研究は開始段階で心筋梗塞の新規治療ターゲットの発見を目指したため、これに沿ってターゲット遺伝子の絞り込みを行った。基本的にはこれまでに

われた GWAS の結果、世界的にコンセンサスが得られていると考えられた領域をベースとし、ここからリシーケンスの対象遺伝子を選択した。これまで多くの GWAS が虚血性心疾患や心筋梗塞を対象に実施されているが、研究開始段階で最も大規模なものが CARDIOGRAMplusC4D コンソーシアムによる 2013 年の GWAS メタ解析<sup>8</sup>であった。患者群約 6 万人、対照群約 12 万人の解析で 47 の領域が虚血性心疾患と関連するとして挙げられている。また同年に報告された中国人(漢民族)における GWAS<sup>7</sup>で前出のコンソーシアムの解析ではあがっていないアジア人固有の領域が 2 つ挙げられている。

これらの計 49 のゲノム上の領域には 66 遺伝子が含まれる。GWAS の結果はゲノム上の特定の領域が疾患と関連することを示唆するが、連鎖不平衡から決定されるその領域は複数の遺伝子を含み、いずれの遺伝子が心筋梗塞に影響を与えるのか判断が難しいことも多い。本研究においては上記の論文のなかで名前が挙げられている遺伝子以外にも GWAS の top signal と連鎖不平衡にある遺伝子を広く対象として選定した (9 遺伝子の追加)。またヒト血管内皮細胞を用いた eQTL 解析の報告<sup>25</sup>があり、それにより心筋梗塞との関連が指摘されている遺伝子で上記の論文中では記載のない遺伝子についても対象に追加した (8 遺伝子の追加)。最終的に 83 の遺伝子を対象とし、以下の 2 基準でスコアリングを行った。

(1) コードされたタンパクに対する測定系やそれを対象とした治療薬がすでに存在するもの

(2) ノックアウトマウスモデルが心筋梗塞を想起させるもの

これらの 2 基準のうち 1 つ以上に該当するもののみを選定し、次の 36 遺伝子が含まれる 19 のゲノム上の領域を本研究のターゲットとして設定した。(表 2)

表 2 ターゲット遺伝子選定基準

Chr	GWAS SNP	Reported genes	Other genes in LD	eQTL genes	Criteria1	Criteria2
1	rs17114036	<i>PPAP2B</i>		<i>PPAP2B</i>		1
1	rs4845625	<i>IL6R</i>			1	
1	rs11206510	<i>PCSK9</i>			1	
2	rs1561198	<i>VAMP5, VAMP8, GGCX</i>	<i>MAT2A</i>			1
2	rs6725887	<i>WDR12</i>	<i>ICA1L, NBEAL1, CYP20A1, CARF</i>		1	
2	rs2123536a	<i>TTC32, WDR35</i>			1	
4	rs7692387	<i>GUCY1A3</i>	<i>GUCY1B3</i>			1
4	rs1878406	<i>EDNRA</i>				1
6	rs4252120	<i>PLG</i>			1	1
6	rs12190287	<i>TCF21</i>				1
7	rs2023938	<i>HDAC9</i>			1	1
9	rs1333049	<i>CDKN2BAS1</i>	<i>CDKN2A, CDKN2B</i>			1
10	rs501120	<i>CXCL12</i>			1	
11	rs974819	<i>PDGFD</i>				1
12	rs3184504	<i>SH2B3</i>	<i>BRAP</i>	<i>MAPKAPK5, TCTN1</i>		1

13	rs9319428	<i>FLT1</i>		1
15	rs17514846	<i>FURIN, FES</i>		1
17	rs2281727	<i>SMG6</i>	<i>SRR</i>	1
19	rs1122608	<i>LDLR</i>	<i>ICAM1</i>	1

reportedGene は GWAS 論文中で関連遺伝子として挙げられている遺伝子を記載した。Other genes in LD は GWAS での top SNP と  $r^2 < 0.8$  以内の連鎖不平衡にある SNP を記載。eQTL Gene は参考文献 24 で関連が示唆されている遺伝子を記載した。Criteria1 はコードされたタンパクが測定可能であるか、それを対象とした治療薬が存在するもの。Criteria2 はノックアウトマウスが心筋梗塞を想起させるものであり、該当する遺伝子が存在する場合 1 と記載した。LD: linkage equilibrium, eQTL: expression quantitative trait loci

## 2-4 ライブラリ作成

今回用いた手法は PCR ベースのターゲット・リシーケンスである。上記のとおり選定したゲノム上のターゲット遺伝子の CDS 領域を PCR で増幅し、次世代シーケンサー (HiSeq2500, Illumina San Diego, CA, USA) を用いてシーケンスを行った。上記の 36 遺伝子には 534 の CDS 領域が含まれる。CDS 領域の決定は CCDS データベース (Release 15, <https://www.ncbi.nlm.nih.gov/CCDS/>)<sup>26</sup> を参照して行ったほか、一部 CCDS データベースに未収録であった *CDKN2B-AS1* に関しては NCBI データベース (RefSeq, <http://www.ncbi.nlm.nih.gov/refseq/>) を参照して決定した。CDS 領域が 180 塩基対 (bp) を超えるものに関してはシーケンサーの試薬の関係上、180 bps 以下となるように配列を分割した。こうして 690 のゲノム配列を PCR 増幅のターゲットとして決定し、それぞれに対して膨大な数のプライマーの組

わせが考えられるため、Primer3(<http://bioinfo.ut.ee/primer3/>)をベースとした自製ソフトウェアにより自動的にプライマー作成を行った。プライマー配列の末尾には PCR 反応後に個体識別配列をつなげるため、設計段階で共通アダプター配列を付与しておいた。

上述のように作成した領域は一部重複する部分があり、すべてのターゲット領域を一度に増幅しようとするるとプライマーダイマーが増幅され、関心領域が効率的に増幅されないため、対象領域が重複せず、互い違いになるようにプライマーを 2 群にわけ、2 回の PCR 反応ですべての領域を増幅できるようなデザインで実験を行った。これでも 400 弱の PCR ターゲットを同時になるべくプロダクト量のばらつきが少なくなるように増幅する必要がある。これを実現するために様々な実験条件の調整を行ったが最終的に投入プライマー濃度を極端に低くし、1 プライマーペアあたり 40 nM の投入に留めることで HiSeq でシーケンス可能な量のプロダクトを許容範囲内のばらつきで得ることができた。

実際の実験は以下のように行った。384 well 乾燥ゲノムプレートを使用し、1 回目の PCR 反応(【95°C 30 秒】 – 【60°C 90 秒】 – 【72°C 60 秒】 x30 サイクル)によって関心領域の増幅を行い、2 回目の PCR 反応によって 1 回目の PCR でアンプリコンの両端に付与されたアダプター配列に 8 bp からなるサンプル識別用バーコード配列を付与した。こうして 1 個体ごとに識別可能になった PCR

産物を1本のチューブにまとめ、磁性ビーズ(AmpureXP, Beckman Coulter, Brea, CA, USA)による精製、Bioanalyzer system(Agilent, Santa Clara, CA, USA)および qPCR assay(KAPA lifesystems, Wilmington, MA, USA)によるライブラリ定量を行った後に HiSeq2500 を使用してシーケンスを行った。

なお Stage2 においては Stage1 で予め設定した P 値を下回った遺伝子のみをターゲットとするプライマーのみを用いて PCR 反応を行った。

## 2-5 解析パイプラインの作成とデータ解析

HiSeq から出力されたシーケンスデータは、まず bcl2fastq ソフトウェア (Illumina)を用いて fastq 形式に変換した。この段階の fastq ファイルはプライマーの合成オリゴ配列を含むため cutadapt ソフトウェア<sup>27</sup>を用いてあらかじめ登録しておいたプライマー配列を read から除去した。Reference genome には hg19 を使用し、BWA ソフトウェア<sup>28</sup>(Version 0.7.5)を用いてアラインメントを行った。SAMTools ソフトウェア(Version 0.1.19, <http://samtools.sourceforge.net/>)を用いてバイナリ化およびインデックス作成を行った後、以下のように Genome Analysis ToolKit<sup>29</sup>(GATK version 3.2.2)を使用してデータ解析を行った。

まず GATK IndelRealigner を使用して既知の InDel 周囲のリードの再アラインメントを行った後に GATK を使用して変異のコールを行った。GATK は whole

genome や whole exome といった実験のデータ解析をすることを目的に実装されており、今回のような amplicon シークエンスのデータをそのまま解析すると特定のバリエントをうまくコールできないことが予備実験からわかっていたため、その問題を解決するため以下のようなパイプラインを作成して用いた。

まず、GATK UnifiedGenotyper と GATK HaplotypeCaller の 2 種の変異コールプログラムを独立に使用してそれぞれ VCF (Variant Call Format) ファイルを生成し、両方のファイルの和集合となるようにファイルの統合を行う。この統合されたファイルをソースとして使用し、GATK UnifiedGenotyper を GENOTYPE\_GIVEN\_ALLELES モードで再度走らせて genotyping を行う。こうすることでコールの精度を上げることができた。

## 2-6 アノテーション

得られたバリエントは SnpEff ソフトウェア<sup>30</sup>を用いてアノテーションを付与した。転写産物が複数あり、アノテーションも複数存在するものに関してはもともと重大な影響を与えると思われるアノテーションを採用した

## 2-7 Quality Control

偽陽性のバリエントを除外するために、バリエントレベルでの QualityControl

(QC)を行った。バリエントが Depth (DP) 20 以下でコールされている場合、また Hardy-Weinberg 平衡から大きく逸脱している場合 ( $HWE\_P < 1 \times 10^{-5}$ )、バリエントの missing rate が 5%を上回る場合、genotype を no\_call とした。さらに、特定のバリエントについてすべてのサンプルの Minor Allele をもつ read の割合をヒストグラムにした。このヒストグラムは自然なバリエントであればそのアレル頻度に応じて 0, 0.5, 1.0 の 3つの位置にピークをもつ 3峰性のヒストグラムになることが期待される。一方でバリエントがコールのエラーや PCR のエラーによる偽陽性であった場合、ヒストグラムは 3峰性にならず、連続的な分布を呈することが予想される。これを用いてすべてのバリエントについてヒストグラムを目視でチェックし、偽陽性と思われるバリエントを除外した。

また保存状態が悪い DNA サンプルなどによる偽陽性を除外するためサンプルレベルでの QC も併せて行った。まず特定のサンプルについて得られた readのうち hg19 に正しくマップされた read の割合を計算し、これが 0.6 に満たないものを除外した。また増幅対象としたすべての塩基対に対して 20 以上の DP が得られた割合が 95%以下のものも除外した。

今回使用した PCR ベースの解析パイプラインは新規に確立したものであるため、その精度を検証する必要があると考えた。このため本実験サンプルと並行して

国際 Hapmap プロジェクトで使用された日本人-中国人サンプル(JCH) 76 サンプルを Stage1 で使用したのと同じプライマーセットを用いて増幅し、得られた解析結果のデータを 1000 genomes project Phase3 (以下 1000g)の成果物として公開されているシーケンスデータとの比較を行った。それに加え、今回使用したサンプルの一部は以前に行われた GWAS でも使用されており、Beads-chip で genotype されたデータが存在していたため、そのデータを今回の解析との genotype 一致率を確認するために使用した。

## 2-8 レアバリエントと脂質レベル、心筋梗塞発症年齢の解析

*LDLR* および *PCSK9* のレアバリエントが臨床パラメータに与える影響を検討するため、心筋梗塞群 9,951 人、対照群 5,407 人の LDL コレステロール値および、心筋梗塞群 9,285 人の発症年齢の臨床データを参照した。

## 2-9 関連解析

### 2-9-1 単変異ごとの関連解析

変異ごとの関連解析は、ケースかコントロールか、その変異を持つか持たないかで 2x2 分割表を作り、フィッシャーの正確性検定を用いて分布の偏りを評価

した。Stage1 の結果、P 値が 0.05 を下回った変異のみを Stage2 にすすめ、2 つのステージの解析が終了したところで 2 つの結果を

Cochran-Mantel-Haenzsel 法によって統合した  $P$  値を算出した。これがあらかじめ定めた有意水準を下回るか否かで有意性を判断した。統計には plink ソフトウェア<sup>31</sup>を使用した。有意水準は Stage1 の QC 終了時点でコールされたすべての変異数 (997) で 0.05 を除した値である  $P = 5.0 \times 10^{-5}$  を用いた。

### 2-9-2 遺伝子ごとの解析

レアバリアントの解析ではバリアントの頻度が非常に小さいため、単変異ごとの解析では統計的な有意性を示すために膨大なサンプルサイズが必要とされる。

一方、ケース群で一定の遺伝子に稀少な変異の集積が見られれば、その遺伝子の機能が疾患に対して何らかの影響をもっているはずと考えることができる。

これを示すために collapsing method と呼ばれる、稀少変異を属する遺伝子ごとにまとめてその影響を解析する手法が検討されている。10 以上の様々に異なる

方法が提唱されているが、いずれも決定的なものはない。本研究では Cohort

Allelic Sum Test (CAST) と Sequence Kernel Association Test (SKAT) とよ

ばれる 2 つの手法を採用した。CAST においてはオッズ比および 95%信頼区間

を R (<https://www.r-project.org/>)を用いて算出した。レアバリアントをまとめる

手法に関しては先行研究<sup>24, 32</sup>に参考にして、以下の4つのグループに分ける手法を採用した。すなわち(1) All non-synonymous 群:すべての non-synonymous 変異を含む (2) damaging (broad)群:すなわち、すべての disruptive な変異および以下の5つのタンパク機能予測プログラム(PolyPhen-2 HumDiv, PolyPhen2-HumVar, SIFT, MutationTester, LRT score)のいずれか1つに damaging な変異であると判定された missense 変異 (3) damaging (strict)群:すべての disruptive な変異および上記の5つのプログラムのすべてで damaging な変異であると予測された missense 変異 (4) disruptive 群: disruptive な変異のみの4グループである。

しかし、こういったタンパク機能予測プログラムは基本的にタンパクの機能が損なわれるかどうかを判定するものであり、loss-of-function 変異を予想することは可能でも gain-of-function 変異を特定することは難しいことが知られている<sup>33</sup>。このため上記のグループ(1)に対しては基本的に loss-of-function と gain-of-function が混在しているものと考えて両方向の変異の影響を勘案することのできる SKAT を適応し、あとの3つのグループに関しては基本的に機能欠失変異が絞り込まれているものと考えて1方向の影響のみを想定している CAST を適応した。2ステージのメタアナリシスを行う段階においては SKAT に関しては R の MetaSKAT パッケージを使用し、CAST に関しては R の

Metafor パッケージを使用して Cochran-Mantel-Haenszel 法により計算を行った。遺伝子レベルの解析における有意水準の設定はまず遺伝子数を 20,000 とし、4 つの独立したグループ分けに対してテストを行っていることを考慮し、 $0.05/20000/4 = 6.3 \times 10^{-7}$  を gene-wide 有意水準に設定した。また当初 36 遺伝子をターゲットとしたため  $0.05/36/4 = 3.5 \times 10^{-4}$  を study-wide 有意水準とした。

### 2-10 LDL コレステロール値、心筋梗塞発症年齢の解析

また *LDLR* および *PCSK9* は両者とも脂質を介して虚血性心疾患に関連することがよく知られている遺伝子であるため得られたレアバリエントが LDL コレステロール値および心筋梗塞発症年齢に与える影響を検討するため、*LDLR* に関しては上記の 4 グループについてそれぞれ LDL コレステロール値、心筋梗塞発症年齢を検討した。また *PCSK9* については関連解析の結果に基づいて、E32K と R93C の 2 つの影響の大きい変異の有無、および disruptive 変異の有無についてそれらの臨床パラメータを検討した。非キャリアと disruptive 変異のキャリアの比較には *t* 検定を用いた。また脂質異常症の治療薬の有無によってサブ解析を行った。

### 2-11 脂質異常症の有無による補正

さらに *LDLR* と *PCSK9* が心筋梗塞に与える影響が脂質を介したものであることを確認するため、ロジスティック回帰分析を行った。この解析においては Stage によらずすべてのサンプルを同時に解析する手法をとり、高脂血症の有無を共変量に加えたモデルにおいて P 値の変化を観察した。ここでの高脂血症は (1) LDL コレステロール値 > 140mg/dL 、 (2) 脂質異常症治療薬の内服 のいずれかを満たすもの、とした。

### 3 結果

#### 3-1 パイプラインのバリデーション

1000g のデータでは対象の関心領域に 185 のバリエーションがコールされており、一方、私の解析パイプラインでは 188 がコールされた。内訳をみると 184 はコールが一致しており、1 は 1000g のみでコール、4 は当パイプラインのみでコールされていた。すなわち、1000g のデータを基準とみなした場合、当パイプラインのバリエーション検出に関する感度は 99.5%(184/185)、陽性的中率は 97.9%(184/188)と計算される。また Beadschip との genotype 一致率は 99.94%(440,765/41,051)であり、以上をもって当パイプラインは充分高い精度を達成していると判断した。

#### 3-2 リシーケンス結果のサマリー

Stage 1 は 36 の対象遺伝子に対して心筋梗塞 2,811 人、対照群 2,974 人について実施した。シーケンスの結果、ターゲットとした塩基の 98.9%の塩基が DP 20 以上でカバーされた。平均の DP は 590 であった。

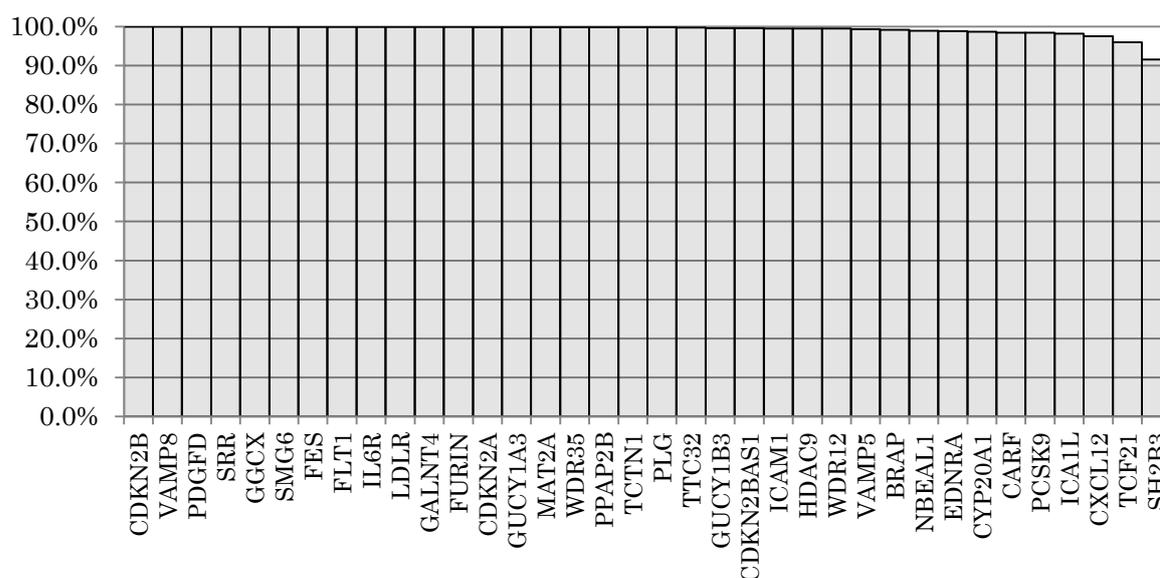


図 3 Stage1 の遺伝子のカバー率

Stage1 のシーケンス結果。横軸に対象遺伝子、縦軸にその遺伝子内のターゲット塩基対のうち何%が全てのサンプルで Depth 20 以上で解析可能であったかを示した。

QC 後に残ったサンプルは心筋梗塞群 2,775 人、対照群 2,965 であって 997 の変異が検出された。これらの変異に対して単変異ごとの関連解析、および遺伝子ごとの関連解析を行った。この Stage 1 においては P 値の閾値を 0.05 に設定し、これを下回った 12 遺伝子上の 17 SNV を Stage 2 へと進めた (表 3)。遺伝子ごとの解析においては上述のような 4 つの異なるグルーピングを用いたが、いずれか 1 つのグルーピングにおいて P 値が 0.05 を下回った遺伝子を Stage 2 へと進めた。ただしこの場合 Stage2 での評価は Stage1 で 0.05 を下回ったグルーピングのみについてだけ行った。この閾値により 20 の単変異、7 の遺伝子が選択された (表 4)。結果を以下に示す。

表 3 単変異関連解析 Stage1 の結果

Chr	Position	rsID (dbSNP138)	A1 A2	MAF in cases	MAF in controls	OR	<i>P</i>	Gene	Type	AA change
1	55505604	-	A G	0.016	0.0079	2	$8.6 \times 10^{-5}$	<i>PCSK9</i>	missense	E32K
1	55509585	rs151193009	T C	0.0047	0.011	0.44	$2.7 \times 10^{-4}$	<i>PCSK9</i>	missense	R93C
1	154426970	rs2228145	C A	0.37	0.4	0.9	$7.5 \times 10^{-3}$	<i>IL6R</i>	missense	D358A
2	203684462	-	A C	0.0016	0.00034	4.8	0.033	<i>ICAIL</i>	missense	D174Y
2	203948147	-	G A	0.0022	0.00034	6.4	$5.9 \times 10^{-3}$	<i>NBEAL1</i>	missense	N297S
2	203996741	-	C A	0.0049	0.0083	0.59	0.028	<i>NBEAL1</i>	missense	K1175Q
2	204003431	rs141142182	C T	0.012	0.016	0.723	0.048	<i>NBEAL1</i>	missense	I1574T
2	204032045	-	A G	0.0013	0	NA	$6.2 \times 10^{-3}$	<i>NBEAL1</i>	missense	V195I
2	204131371	-	C G	0	0.0013	0	$8.0 \times 10^{-3}$	<i>CYP20A1</i>	missense	Q189H
4	156715112	-	T A	0.0013	0	NA	$6.2 \times 10^{-3}$	<i>GUCY1B3</i>	missense	E200D
6	161159625	rs121918027	A G	0.022	0.015	1.5	$3.5 \times 10^{-3}$	<i>PLG</i>	missense	A620T
9	21971043	-	T G	0.00018	0.0013	0.13	0.04	<i>CDKN2A</i>	missense	D105E
11	103814218	-	T C	0.0012	0.00017	7.5	0.034	<i>PDGFD</i>	missense	R245Q
12	111082836	rs118096349	T G	0.0072	0.011	0.64	0.031	<i>TCTN1</i>	missense	G466C
17	1989027	rs187319098	G C	0.002	0.0046	0.43	0.022	<i>SMG6</i>	missense	E268Q
17	2186947	rs372239404	G A	0.0016	0.00017	9.6	$9.5 \times 10^{-3}$	<i>SMG6</i>	missense	M807T
19	11226885	-	G C	0.0016	0.00017	9.6	$9.5 \times 10^{-3}$	<i>LDLR</i>	missense	L568V

Stage1 の結果、Fisher の正確性検定による *P* 値が 0.05 を下回ったもののみを表示した。Chr: chromosome, OR: odds ratio, MAF: minor allele frequency, AA: amino acid, NA: not available

表 4 遺伝子ごと関連解析 Stage1 の結果

Genes	Non-Synonyous (SKAT)	Damaging(broad) (CAST)	Damaging(strict) (CAST)	Disruptive (CAST)
<i>PCSK9</i>	$3.7 \times 10^{-5}$	$1.6 \times 10^{-3}$	0.67	0.69
<i>GUCY1B3</i>	$7.3 \times 10^{-3}$	0.047	0.033	0.055
<i>PLG</i>	0.017	$4.3 \times 10^{-3}$	1	1
<i>ICAIL</i>	0.039	0.13	0.018	NA
<i>NBEAL1</i>	0.043	0.41	1	0.84
<i>TCTN1</i>	0.044	0.040	1	NA
<i>LDLR</i>	0.049	0.15	$2.6 \times 10^{-5}$	$5.0 \times 10^{-3}$

Stage1 の結果、SKAT または CAST による  $P$  値が 0.05 を下回ったもののみを表示した。いずれかのグループで 1 回でも 0.05 を下回っている遺伝子をすべて表示している。

これらの変異、遺伝子をターゲットとし、心筋梗塞群 7,316 人、対照群 5,828 人で Stage2 のリシーケンスを行った。ターゲットとした塩基の 98.2%が DP 20 以上でカバーされ、平均 DP は 980 であった。QC 後に残ったサンプルは心筋梗塞群 7,181 人、対照群 5,408 人であった。816 の変異が特定された。

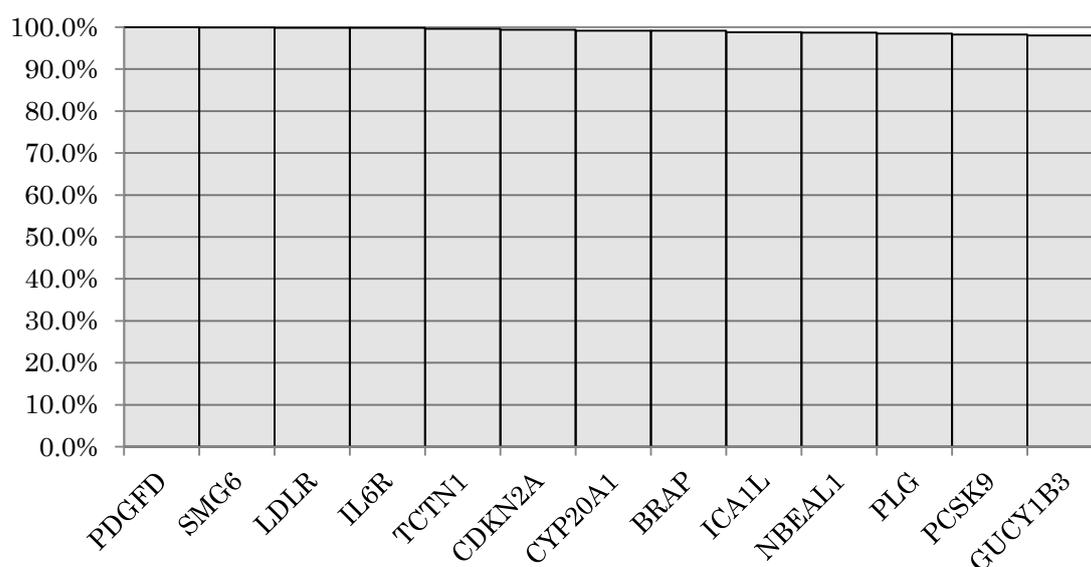


図 4 Stage2 の遺伝子のカバー率

Stage2 のシーケンス結果。横軸に対象遺伝子、縦軸にその遺伝子内のターゲット塩基対のうち何%が全てのサンプルで Depth 20 以上で解析可能であったかを示した。単変異の解析のみで  $P < 0.05$  を下回った場合も遺伝子全てをシーケンスの対象としている。ただし Stage1 で  $P < 0.05$  を満たしていない変異については Stage2 での結果は検討対象から外した。

次に Stage1 と Stage2 の結果について上に述べた手法でメタアナリシスを行った。

### 3-3 単変異ごとの関連解析の結果

単変異ごとの解析について予め設定した有意水準を下回る変異を 2 つ確認した。

表 5 単変異ごとの関連解析結果

Chr	Position	rsID (dbSNP138)	Alt Ref	Gene	Type / grouping	AA change	Stage	MAF in cases	MAF in controls	OR	(95%CI)	P
							Discovery	0.016	0.0079	2.0	(1.4-1.9)	8.6 x 10 <sup>-5</sup>
1	55505604	-	A G	<i>PCSK9</i>	Missense Non-synonymous	E32K	Replication	0.015	0.0095	1.5	(1.2-1.9)	3.5 x 10 <sup>-4</sup>
							Combined	0.015	0.0090	1.7	(1.4-2.0)	3.5 x 10 <sup>-7</sup>
							Discovery	0.0047	0.011	0.44	(0.27-0.70)	2.7 x 10 <sup>-4</sup>
1	55509585	rs151193009	T C	<i>PCSK9</i>	Missense damaging(broad)	R93C	Replication	0.0075	0.011	0.71	(0.54-0.93)	9.8 x 10 <sup>-3</sup>
							Combined	0.0067	0.011	0.63	(0.50-0.78)	3.4 x 10 <sup>-5</sup>

AA: amino acid, MAF: minor allele frequency, OR odds ratio, CI: confidence interval

いずれもアレル頻度 1%以下の低頻度変異で *PCSK9* の変異が 2 つであった。なお、この 2 変異は OR>1 のリスクとして働く変異 (E32K) と OR<1 の抵抗性変異 (R93C) であった。

### 3-4 遺伝子ごとの関連解析の結果

遺伝子ごとの関連解析の結果は *LDLR*, *PCSK9* の 2 遺伝子が gene-wide 有意水準を下回った。また *TCTN1* が 1 つのグルーピングで study-wide 有意水準を下回った。

表 6 遺伝子ごとの関連解析結果

Gene	Grouping (test method)	Stage	Freq in cases (count / %)	Freq in controls (count / %)	OR	(95%CI)	<i>P</i>
<i>LDLR</i>	Damaging(strict) (CAST)	Discovery	23 (1.0%)	5 (0.17%)	6.0	(2.3-20)	$2.6 \times 10^{-5}$
		Replication	61 (0.85%)	12 (0.22%)	3.9	(2.1-7.9)	$2.1 \times 10^{-6}$
		Combined	-	-	4.4	(2.6-7.4)	$7.2 \times 10^{-10}$
<i>PCSK9</i>	All non-synonymous (SKAT)	Discovery	-	-	-	-	$3.7 \times 10^{-5}$
		Replication	-	-	-	-	$7.5 \times 10^{-4}$
		Combined	-	-	-	-	$2.3 \times 10^{-7}$
<i>TCTN1</i>	Damaging(broad) (CAST)	Discovery	67 (2.4%)	99 (3.3%)	0.72	(0.51-0.99)	0.040
		Replication	182 (2.5%)	187 (3.5%)	0.73	(0.59-0.90)	0.0028
		Combined	-	-	0.72	(0.61-0.86)	$2.2 \times 10^{-4}$

### 3-4-1 *LDLR*について

*LDLR*については設定した4つのグルーピング(1) All non-synonymous に振り分けられる変異が138、(2) Damaging (broad) に振り分けられる変異が108、(3) Damaging (stcic)に振り分けられる変異が52、(4) disruptive に振り分けられる変異が14得られた(重複あり)。(表9参照) 関連解析の結果としては Damaging (strict)と disruptive 群において、それぞれ gene-wide 有意な関連がみられた(それぞれ  $OR = 4.4$ ,  $P = 7.2 \times 10^{-10}$  と  $OR = 15$ ,  $P = 5.8 \times 10^{-7}$ )、全

でのグループの結果を表 7 に示す。

表 7 *LDLR*, *PCSK9*の結果

gene	grouping/test method	OR	95%CI	P
<i>LDLR</i>	All non-synonymous/SKAT	-	-	$7.7 \times 10^{-4}$
	Damaging (broad)/CAST	1.2	1.0-1.3	0.032
	Damaging (strict)/CAST	4.4	2.6-7.4	$7.2 \times 10^{-10}$
	Disruptive/CAST	15	3.7-64	$5.8 \times 10^{-7}$
<i>PCSK9</i>	All non-synonymous/SKAT	-	-	$2.3 \times 10^{-7}$
	Damaging (broad)/CAST	0.75	0.66-0.86	$2.1 \times 10^{-5}$
	Damaging (strict)/CAST	0.84	0.48-1.5	0.54
	Disruptive/CAST	0.59	0.22-1.6	0.29
	Gain-of-function/CAST	1.3	1.1-1.5	$1.0 \times 10^{-4}$
	Loss-of-function/CAST	0.66	0.53-0.81	$1.1 \times 10^{-4}$

これを見るとグルーピングの絞りこみ基準が厳しくなり、**damaging** な変異が絞りこまれていると思われるグループでより心筋梗塞に対する OR が高くなる傾向が確認できる。*LDLR* の領域に関しては現在まで日本人および中国人での common SNP を対象とした GWAS では関連が示されておらず、この関連はレアバリアントの集積による効果であるといえることができる。

### 3-4-2 *PCSK9*について

*PCSK9* の心筋梗塞との関連は *LDLR* より複雑である。*PCSK9* では(1) All non-synonymous に振り分けられる変異が 94、(2) Damaging (broad) に振り分けられる変異が 64、(3) Damaging (strict)に振り分けられる変異が 18、(4)

disruptive に振り分けられる変異が 6 得られた。リスク変異と抵抗性変異双方の効果を加味できる SKAT で低い  $P$  値が得られている一方、他の 3 つのグループにおいては Damaging (broad) で study-wide な関連が見られているのみで他の 2 群に関しては有意な関連が見られていない。これは *PCSK9* の変異はリスク変異と抵抗性変異が入り混じっていることを示唆する。単変異の解析でみられた E32K と R93C の関係もこれを支持するし、過去の脂質異常症家系の解析でもこの遺伝子の gain-of-function 変異は LDL-C を上げる方向に働き、loss-of-function 変異が LDL-C を下げる方向に働いていることが広く知られている。このためここで行ったタンパク機能予測プログラムを用いた 3 つのグループは *PCSK9* の変異を分類する際には適切でない可能性がある。

Gain-of-function 変異を適切に予測する手法は知られていないため、ここでは *PCSK9* のバリエーションを家族性高脂血症のデータベースに登録されている変異についてそれが gain-of-function 変異としての登録であるか loss-of-function 変異としての登録であるかによってバリエーションを 2 つに分け、それぞれのグループと心筋梗塞との関連を見た。結果はそれぞれ  $OR = 1.3$ ,  $P = 1.0 \times 10^{-4}$  と  $OR = 0.66$ ,  $P = 1.1 \times 10^{-4}$  といずれも study-wide な関連がみられ、OR も予想された通りの結果が示された。これらの関連のうち、単変異の解析で有意な関連を示した E32K, R93C からの影響を評価する目的でこの 2 つの変異を除外して再度

関連解析を行ったところ、いずれも有意性は失われた。これらの結果からは *PCSK9* と心筋梗塞の関連においては互いに効果の方向が異なる 2 つの低頻度変異 E32K, R93C の影響が大きいということが言える。

### 3-4-3 *TCTN1* について

さらに我々は *TCTN1* 遺伝子と心筋梗塞の study-wide な関連を見出した。

*TCTN1* においては Damaging (broad) 群において  $OR = 0.72$ ,  $P = 2.2 \times 10^{-4}$  の関連が示された。一方、その他の群においては得られた変異数が極端に少なく有意性は示されなかった。

### 3-5 *LDLR*, *PCSK9* の低頻度変異が臨床的パラメータに及ぼす影響

*LDLR*, *PCSK9* はともに血清 LDL コレステロールに影響を与え、心筋梗塞のリスクに関連することが知られている。そこで、今研究で特定された稀少変異が血清 LDL 値、及び心筋梗塞群の中での心筋梗塞発症年齢にどのような影響を与えているかを調べた。

### 3-6 *LDLR* の低頻度変異と LDL-C、心筋梗塞発症年齢との関係

遺伝子ごとの解析で使用した 4 群について、それぞれの群に分類される変異の

保持者について、それぞれの血液検査データ、診療録情報から LDL-C 値、心筋梗塞発症年齢を平均値で比較した。(図 5 および図 6)

これらを参照すると分かるように、よりタンパク機能に対して有害と思われる変異が多く含まれる群になればなるほど、LDL-C 値は高い傾向があり、また心 non-carrier と比較して disruptive variant の carrier の血清 LDL コレステロー

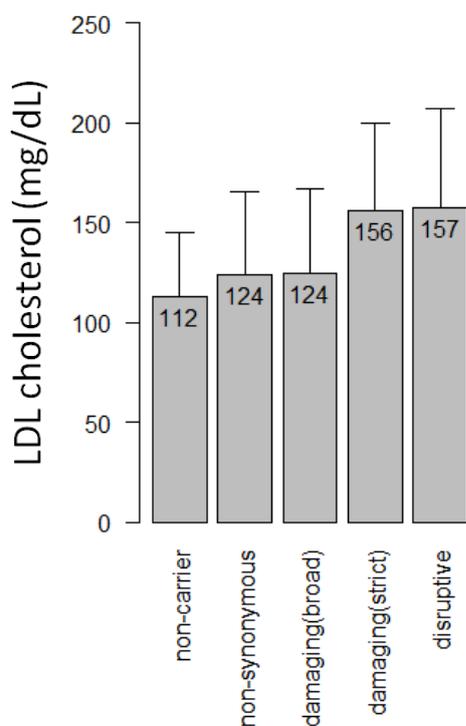


図 5 LDLR の変異と LDL-C の関係

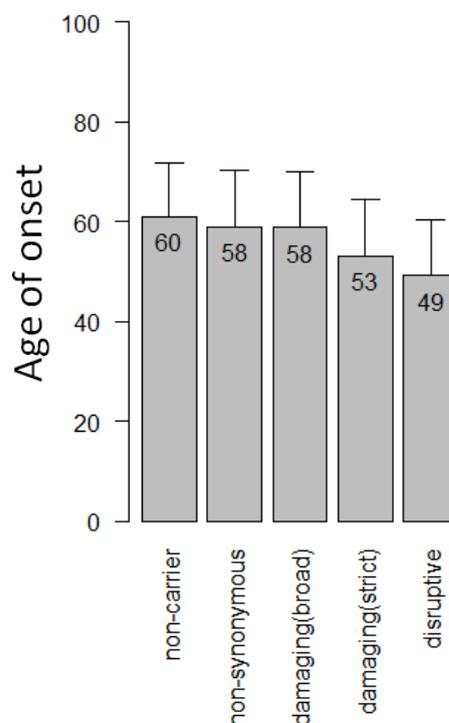


図 6 LDLR の変異と発症年齢の関係

ル値は 112.7mg/dL に対して 157.2mg/dL と有意に高く ( $P= 3.4 \times 10^{-13}$ )、発症年齢は 61.0 歳に対して 49.2 歳と有意に若かった ( $P= 1.1 \times 10^{-10}$ )。

Damaging (strict) 群および disruptive 群に分類される変異の carrier と non-carrier の心筋梗塞発症年齢を比較したヒストグラムを示す。(図 7、8)

Damaging(strict)群および Disruptive 群の発症年齢が平均より若年側に偏位している傾向が見られる。Disruptive 群では一部 2 峰性の分布にもみえるがこれはこれらの変異のキャリア数自体が非常に少ない(n=36)ためであると考えている。

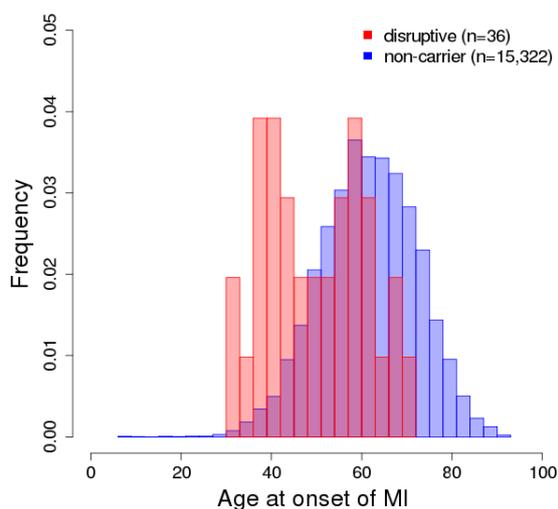


図 7 Disruptive 群の発症年齢

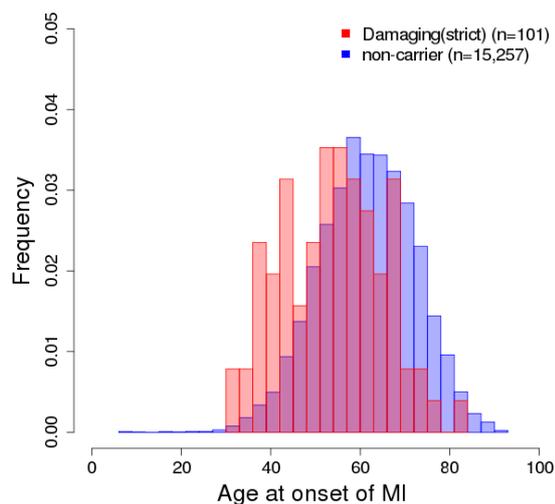


図 8 Damaging (strict)群の発症年齢

血清 LDL コレステロール値がスタチンを含む脂質異常症に対する治療によって変化することは明らかであるので脂質異常症治療薬の内服有無によって群をわけたサブ解析も合わせて行ったが、こちらも同様の傾向が観察された。(図 9、10)

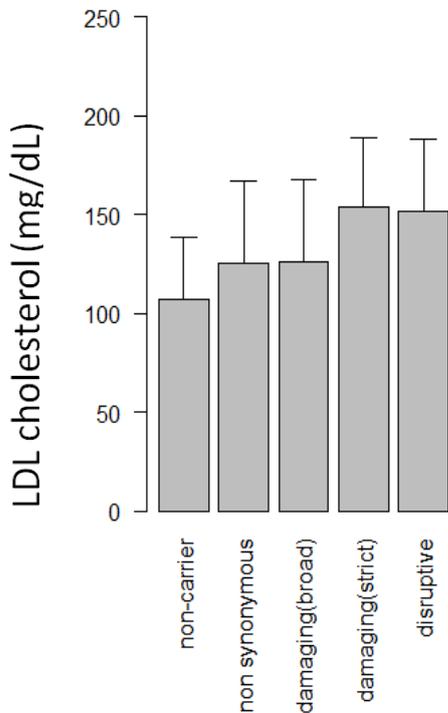


図 9 治療群の LDL-C

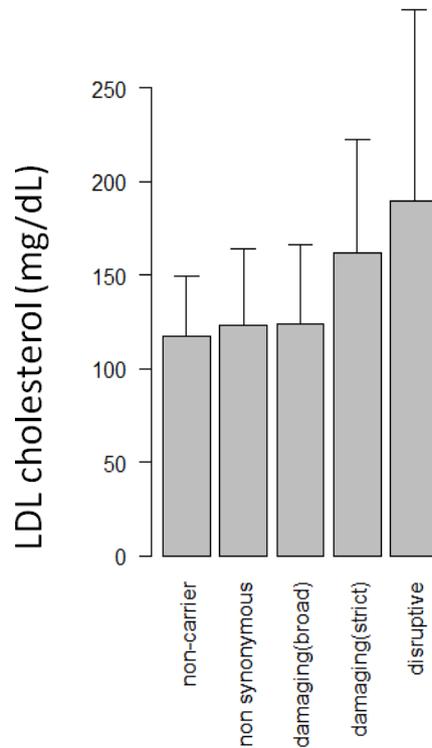


図 10 非治療群の LDL-C

### 3-7 PCSK9の低頻度変異と LDL-C、心筋梗塞発症年齢との関係

*PCSK9*についても *LDLR*と同様にそれぞれの変異のキャリアにおいて LDL-C 値、心筋梗塞発症年齢の 2つのパラメータを評価した。ただし、*PCSK9*においては risk と protective 両方向のバリエントが混在しており、機械的に分類することは困難であること、遺伝子ごとの解析でみたように E32K、R93C の 2つのバリエントの影響が *PCSK9* と心筋梗塞の関連において大きな比率を占めていることを踏まえ、E32K のキャリア、R93C のキャリア、disruptive variant の

キャリアの 3 群において比較を行った。結果は図 11,12 に示されるように E32K キャリアにおいて LDL-C 値が高く、発症年齢は若い傾向が見られ、R93C キャリア、disruptive キャリアにおいては LDL-C 値は低く、発症年齢は遅い傾向がみられた。

Non-carrier と disruptive carrier を比較すると LDL-C 値は 112.5 mg/dl に対して 73.8 mg/dl ( $P=7.4 \times 10^{-3}$ ) と低く、発症年齢は 61.0 歳に対して 73.8 歳 ( $P=3.3 \times 10^{-3}$ ) と有意に遅かった。

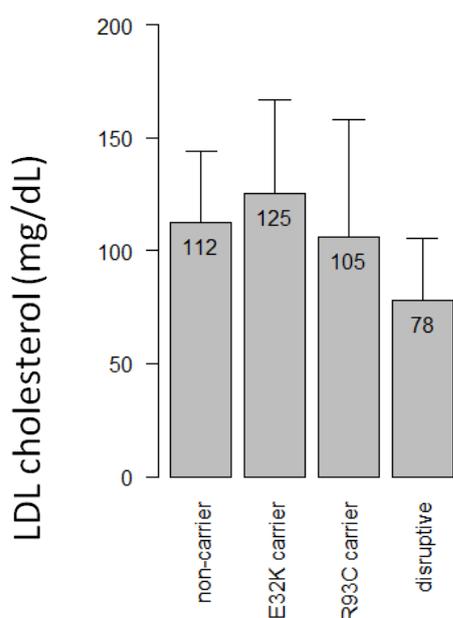


図 11 PCSK9 の変異と LDL-C の関係

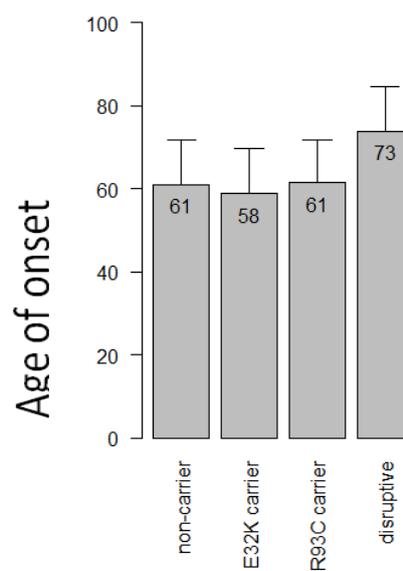


図 12 PCSK9 の変異と発症年齢の関係

脂質異常症治療薬の影響を除外するため、ここでも治療群、非治療群に分けて同様の解析を行ったが同じ傾向が観察された。(図 13、14)

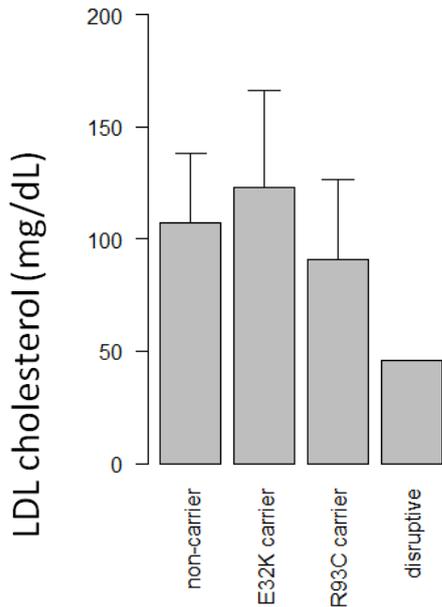


図 13 治療群の LDL-C

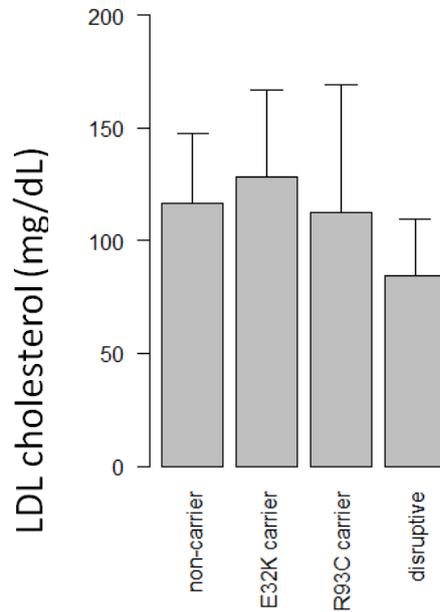


図 14 非治療群の LDL-C

### 3-8 血清 LDL-C 値と心筋梗塞リスクとの関係

ここまで *LDLR* および *PCSK9* の低頻度遺伝子変異が心筋梗塞発症に影響を与えること、それらの変異と LDL-C 値が密接に結びついていることを見てきたが、これらの遺伝子変異が心筋梗塞に与える影響は全て LDL を介したものであるのか、それとも未知の機序により心筋梗塞に直接の影響を与えているのであろうか。これを検討するためにロジスティック回帰解析を行い、脂質異常症の存在を共変数としたときに遺伝子ごとの解析でみられた心筋梗塞との相関の P 値が

どのように変化するのかを見た。

表 8 脂質異常症の有無による調整

Gene	Gene grouping / variant	<i>P</i>	adjusted <i>P</i>
<i>LDLR</i>	Damaging (strict)	$8.4 \times 10^{-7}$	0.018
	Disruptive	$9.5 \times 10^{-6}$	0.028
<i>PCSK9</i>	E32K	$4.7 \times 10^{-5}$	0.93
<i>TCTN1</i>	Damaging (broad)	$6.5 \times 10^{-4}$	$1.3 \times 10^{-3}$

*P*値は血液データが存在するすべてのサンプルについての logistic regression により得られたもの。adjusted *P*は高脂血症の有無を covariate とした場合の *P* 値を示した。著しく相関が弱まることが確認できる。

表 8 において、遺伝子ごとの解析で gene-wide 有意な関連が見られた 2 つの遺伝子の 3 つのグルーピングを取り上げ、単変量解析による *P* 値が脂質異常症の有無を共変数としたモデルでロジスティック回帰分析を行った結果、*P* 値がどう変化したかを右に示している。補正前は非常に強い関連が見られていた 3 つのグループの関連のシグナルは脂質異常症の有無による調整により、いずれも非常に減弱していることが見て取れる。一方で脂質異常との関連が知られていない *TCTN1* の関連シグナルはこの補正によってほとんど影響を受けていない。

この結果から、今回見られた *LDLR*, *PCSK9* と心筋梗塞との関連は LDL コレステロール値を介した影響により多くの部分が説明されるものと考えられた。

### 3-9 家族性高脂血症との関連

*LDLR*, *PCSK9* は共に家族性高脂血症の原因遺伝子として広く知られている。

これまで多くの家族性高脂血症の家系のシーケンスが行われ、多数の変異が原因変異としてデータベースに登録されている<sup>34, 35</sup>。今回の研究で特定した

*LDLR* および *PCSK9* の変異をこのデータベースと照合すると、*LDLR* につい

て 42% (59/138)、*PCSK9* については 24% (23/94) が既にデータベースに収

載されていた。今回、新規に特定された変異の効果について調べるため、デー

タベースに収載された変異を除外した解析を行った。アレル頻度が高いものほ

ど既にデータベースに載っている可能性が高いため、これらを除外すると有効

なサンプルの数が著しく減少し、遺伝子ごとの解析で gene-wide 有意な関連を

示した 3 グループの関連シグナルはいずれも減弱が見られた。(*LDLR*

disruptive  $P=0.13$ , *LDLR* damaging (strict)  $P=0.027$ , *PCSK9* all

non-synonymous  $P=0.58$ )。しかし、*LDLR* の新規の変異について LDL-C 値、

心筋梗塞発症年齢を調べると、上で見てきたのと変わらない傾向が観察された。

(図 15、16)

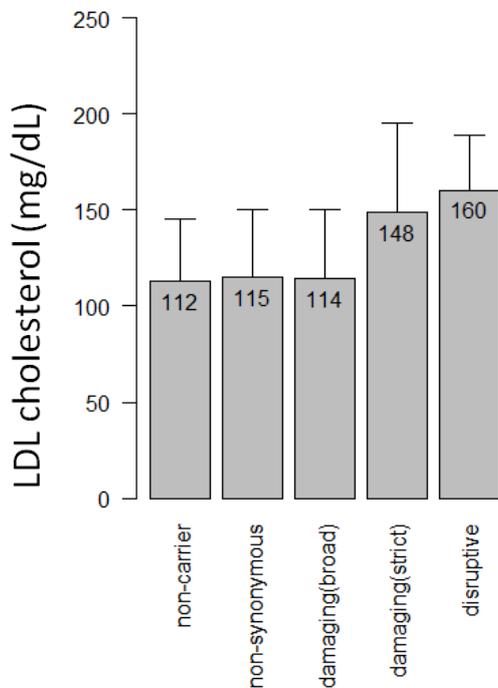


図 15 新規変異と LDL-C の関係

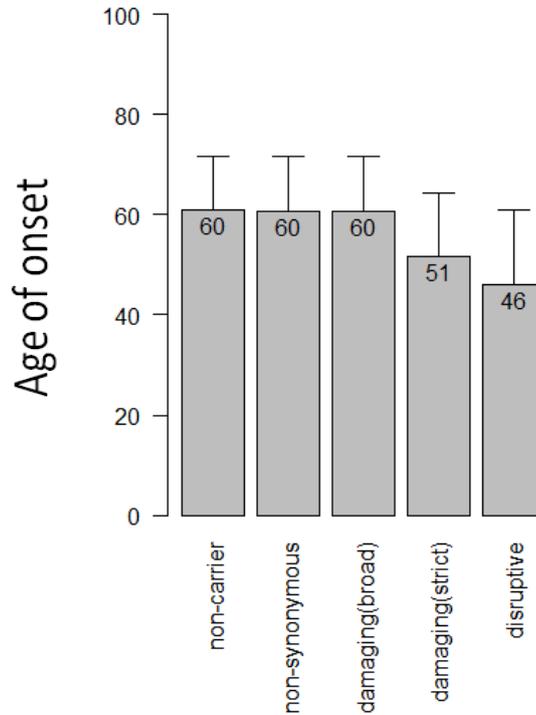


図 16 新規変異と発症年齢の関係

これは今回新規に特定された変異も既に家族性高脂血症の原因として特定されている変異と同様の特性を持っていることを示唆する結果と言える。一方で *PCSK9* の変異については多くの部分が **E32K** と **R93C** の 2 つの影響によって説明されるが、この 2 つの変異はいずれも家族性高脂血症の原因としてデータベースに収載されており、*LDLR* で示したような解析を行うことは困難であった。今回のリシーケンスで特定され解析対象とした *LDLR* 及び *PCSK9* の全変異のリストを以下に示す。

表 9 LDLR の変異リスト

Chr	BP	R A		Type	Ca ses	Cont rols	Group	AA change	LDLR FH database ID	FH allele name
		e l	f t					or splice-site change		
19	11200232	C	T	missense	0	1	Damaging(broad)	P3L		
19	11200247	T	G	missense	1	0	Damaging(broad)	L8W		
19	11200270	C	T	missense	1	1	Non synonymous	L16F		
19	11200280	C	T	missense	0	1	Non synonymous	A19V		
19	11200282	G	A	missense	0	1	Non synonymous	G20R	LDLR_00868	
19	11210898	G	C	splice-sit e	1	0	Disruptive	-1G>C	LDLR_00277	
19	11210902	G	A	missense	1	0	Non synonymous	G24D		
19	11210910	T	C	missense	1	0	Damaging(strict)	C27R		
19	11210923	A	T	missense	1	0	Damaging(broad)	E31V		
19	11210928	C	T	nonsense	1	0	Disruptive	Q33*	LDLR_00005	FH Turkey/Milan-4
19	11210952	T	A	missense	4	3	Non synonymous	S41T		
19	11210970	G	A	missense	3	1	Damaging(strict)	D47N	LDLR_00383	FH Hyogo
19	11210982	G	A	missense	0	2	Damaging(strict)	E51K		
19	11213346	T	C	missense	3	2	Damaging(broad)	V66A		
19	11213360	G	A	missense	5	1	Non synonymous	G71R		
19	11213391	G	A	missense	1	0	Damaging(broad)	R81H		
19	11213402	C	T	nonsense	1	0	Disruptive	Q85*	LDLR_00184	
19	11213429	G	A	missense	0	1	Damaging(strict)	D94N		
19	11213432	T	G	missense	1	0	Damaging(strict)	C95G	LDLR_00454	
19	11213434	C	A	nonsense	1	0	Disruptive	C95*	LDLR_00306	
19	11213441	G	A	missense	1	1	Damaging(broad)	G98S	LDLR_01032	
19	11213462	C	T	missense	2	0	Non synonymous	P105S	LDLR_00442	
19	11215926	G	A	missense	61	62	Damaging(broad)	R74H	LDLR_00366	
19	11215943	T	G	missense	2	0	Damaging(strict)	C80G	LDLR_00935	
19	11215952	C	T	missense	0	2	Damaging(broad)	R83W	LDLR_00692	
19	11215964	T	T	indel G frameshif t	1	0	Disruptive	C128fs		
19	11215970	T	A	missense	4	0	Non synonymous	S89T	LDLR_00936	

19	11215976	C	T	missense	1	0	Non synonymous	R91W		
19	11215977	G	A	missense	7	2	Non synonymous	R91Q		
19	11215989	A	A	indel	1	0	Disruptive	D136fs		
		C		frameshift						
19	11216000	G	A	missense	2	1	Damaging(strict)	E99K	LDLR_00413	FH Sassari-2
19	11216007	C	T	missense	0	1	Damaging(broad)	S142F		
19	11216009	T	G	missense	1	0	Damaging(strict)	C143G		
19	11216033	G	A	missense	4	2	Non synonymous	A151T	LDLR_01038	
19	11216060	T	C	missense	1	0	Damaging(strict)	C160R	LDLR_00721	
19	11216090	G	A	missense	1	0	Damaging(strict)	D170N		
19	11216099	T	C	missense	1	0	Damaging(strict)	C173R	LDLR_00023	FH Greece-1, FH Canada
19	11216147	G	A	missense	1	0	Non synonymous	V189M	LDLR_00938	
19	11216243	G	A	missense	1	0	Damaging(strict)	D221N	LDLR_00490	
19	11216255	A	G	missense	2	1	Damaging(broad)	K225E		
19	11216258	T	C	missense	1	0	Damaging(strict)	S226P	LDLR_00035	FH Miami-1
19	11216263	C	A	missense	1	0	Damaging(strict)	D227E		
19	11216264	G	C	missense	0	1	Damaging(strict)	E228Q	LDLR_00038	FH Tulsa-2, FH Iraq
19	11217303	C	T	missense	0	1	Damaging(broad)	R253W	LDLR_00375	
19	11217315	C	T	missense	2	1	Damaging(broad)	R257W	LDLR_00712	
19	11217352	G	A	missense	1	0	Damaging(broad)	G269D	LDLR_00504	FH Rome-3
19	11218074	T	C	missense	0	1	Non synonymous	L275P		
19	11218076	T	C	missense	0	1	Damaging(strict)	C276R		
19	11218079	G	A	missense	0	1	Damaging(broad)	E277K	LDLR_00195	FH Walloon, Genoa-3
19	11218104	A	G	missense	3	1	Damaging(broad)	H285R		
19	11218154	T	G	missense	1	0	Damaging(strict)	C302G		
19	11218157	C	T	missense	1	1	Damaging(broad)	R303W	LDLR_00512	
19	11218158	G	A	missense	1	0	Non synonymous	R303Q		
19	11221334	A	G	missense	1	0	Damaging(broad)	N316S		
19	11221357	G	A	missense	1	3	Damaging(broad)	G324S	LDLR_00640	
19	11221364	C	A	missense	1	0	Damaging(broad)	S326Y		
19	11221397	A	G	missense	5	4	Damaging(broad)	E337G		
19	11221399	T	A	missense	8	1	Damaging(strict)	C338S	LDLR_00280	FH Wakayama
19	11222190	A	A	indel	1	0	Disruptive	D354fs		
		T		frameshift						
				t						

19	11222192	A	T	missense	2	2	Damaging(strict)	I187F		
19	11222213	G	A	missense	1	0	Damaging(broad)	D362N		
19	11222253	A	T	missense	1	0	Damaging(broad)	Y375F		
19	11222265	G	A	missense	1	0	Damaging(strict)	C379Y	LDLR_00364	
19	11222295	C	T	missense	0	1	Damaging(strict)	T389M	LDLR_00884	
19	11222305	C	A	nonsense	1	0	Disruptive	C392*	LDLR_00157	
19	11223954	G	C	missense	1	0	Damaging(broad)	G228A		
19	11223962	G	A	missense	2	0	Damaging(broad)	A399T	LDLR_00534	FH Nuoro
19	11223969	T	G	missense	1	0	Damaging(strict)	L401R		
19	11223974	T	C	missense	1	0	Damaging(strict)	F403L	LDLR_00536	FH Wakayama
19	11223977	A	G	missense	4	1	Damaging(broad)	T404A		
19	11223983	C	T	missense	1	0	Damaging(strict)	R406W	LDLR_00451	FH Rumania
19	11224013	C	T	missense	1	0	Damaging(strict)	R416W	LDLR_00216	
19	11224014	G	A	missense	5	0	Damaging(broad)	R416Q	LDLR_00379	
19	11224016	A	T	missense	2	1	Damaging(broad)	S417C		
19	11224017	G	A	missense	1	0	Damaging(broad)	S417N		
19	11224019	G	A	missense	2	3	Damaging(broad)	E418K	LDLR_00942	
19	11224025	A	G	missense	11	6	Damaging(broad)	T420A		
19	11224030	C	G	missense	0	1	Damaging(broad)	S421R		
19	11224095	G	C	missense	1	0	Damaging(strict)	W443S	LDLR_00779	
19	11224233	G	A	missense	1	1	Non synonymous	G461S		
19	11224270	T	A	missense	2	0	Damaging(broad)	I473N	LDLR_00295	
19	11224276	C	T	missense	1	0	Damaging(strict)	A475V		
19	11224320	T	T	indel	1	0	Disruptive	W490fs		
		G		frameshift						
				t						
19	11224326	G	A	missense	1	0	Damaging(strict)	D492N	LDLR_00074	FH Cincinnati-2
19	11224336	T	C	missense	0	1	Non synonymous	L495P		
19	11224354	C	T	missense	4	3	Damaging(broad)	A501V	LDLR_00899	
19	11224368	G	A	missense	4	0	Non synonymous	V506M		
19	11224398	G	A	missense	9	3	Damaging(broad)	G516S		
19	11224434	C	G	missense	0	4	Non synonymous	H528D		
19	11226775	T	C	missense	1	0	Damaging(strict)	M531T		
19	11226798	C	T	missense	0	1	Damaging(broad)	P539S		
19	11226801	G	A	missense	1	0	Damaging(strict)	A540T	LDLR_00243	
19	11226829	G	A	missense	0	1	Damaging(broad)	G549D	LDLR_00232	
19	11226844	C	T	missense	1	0	Non synonymous	S554L		

19	11226856	A	C	missense	0	1	Damaging(broad)	E558A		
19	11226883	C	A	missense	1	0	Damaging(broad)	T567N		
19	11226885	C	G	missense	33	5	Damaging(broad)	L568V	LDLR_00558	
19	11227540	C	T	missense	3	1	Damaging(broad)	L571F		
19	11227547	G	A	missense	0	1	Non synonymous	G573D		
19	11227550	G	A	missense	0	1	Damaging(strict)	R574H		
19	11227594	G	A	missense	1	1	Damaging(broad)	D589N	LDLR_01066	
19	11227601	A	G	missense	3	1	Damaging(broad)	N591S		
19	11227612	C	T	missense	5	2	Damaging(strict)	R595W	LDLR_01207	
19	11227613	G	A	missense	1	0	Damaging(strict)	R595Q	LDLR_00563	
19	11227649	A	A	indel	1	0	Disruptive	H607fs		
		C		frameshift						
				t						
19	11227651	C	T	missense	3	0	Damaging(strict)	P608S	LDLR_00662	
19	11227663	G	T	missense	6	2	Damaging(broad)	A612S		
19	11227666	G	A	missense	1	0	Damaging(broad)	V613I	LDLR_01070	
19	11227676	T	C	splice-site	14	1	Disruptive	+2T>C	LDLR_00087	FH Niigata
				e						
19	11230790	T	A	missense	1	0	Damaging(broad)	I623N		
19	11230798	G	A	missense	0	2	Damaging(broad)	E626K		
19	11230802	C	T	missense	1	0	Damaging(broad)	A627V	LDLR_00658	
19	11230820	G	A	missense	1	0	Damaging(broad)	R633H	LDLR_01073	
19	11230880	T	A	missense	1	0	Damaging(broad)	V653D		
19	11230909	G	A	missense	1	0	Damaging(broad)	E536K		
19	11231056	G	C	missense	1	0	Damaging(strict)	W666C		
19	11231084	G	A	missense	3	0	Damaging(strict)	G676S	LDLR_00946	
19	11231112	C	T	missense	2	1	Damaging(strict)	P685L	LDLR_00094	FH Gujerat, FH Frosinone1, FH Kanazawa-2
19	11231130	C	T	missense	1	0	Damaging(strict)	S691L		
19	11231146	C	A	nonsense	1	0	Disruptive	C696*	LDLR_00906	
19	11233888	G	A	missense	2	0	Non synonymous	V727I		
19	11233907	C	T	missense	0	1	Damaging(broad)	S733F		
19	11233915	G	A	missense	3	0	Non synonymous	V736I		
19	11233931	C	G	missense	0	1	Non synonymous	T741R		
19	11233948	C	T	missense	0	1	Non synonymous	P747S		
19	11233966	C	T	missense	1	0	Non synonymous	P753S		

19	11238695	G	A	missense	1	1	Non synonymous	V775I		
19	11238705	G	T	missense	1	0	Non synonymous	R778I		
19	11238713	G	G	indel	1	2	Non synonymous	EK781E		
		A								
		G								
		A								
19	11238731	G	A	missense	0	1	Non synonymous	V787M		
19	11238761	G	A	missense	1	0	Damaging(broad)	V619M	LDLR_00259	
19	11240215	G	A	indel	0	1	Disruptive	V806fs		
		T	G	frameshift						
		C	A	t						
		A								
		G								
19	11240230	A	T	nonsense	11	0	Disruptive	K811*	LDLR_00281	FH Tokyo
19	11240240	G	A	missense	1	0	Damaging(broad)	R814Q	LDLR_00376	FH Xhosa
19	11240278	G	A	missense	2	0	Damaging(broad)	V827I	LDLR_00101	FH New York-5
19	11241972	C	G	missense	9	13	Damaging(broad)	L855V		
19	11241988	C	T	missense	136	152	Damaging(broad)	A860V		

表 10 PCSK9 の変異リスト

Chr	Position	rsID(dbSNP138)	Ref	Alt	Type	cases	controls	Group	AA change	PCSK9 FH database ID	Predicted Effect <sup>a</sup>
1	55505520	rs186669805	G	A	missense	245	204	non synonymous	V4I	PCSK9_00006	GoF
1	55505532		C	G	missense	0	1	non synonymous	R8G		
1	55505552	rs113330492	A	ACTG	insertion	2310	2038	non synonymous	L15LL		
1	55505581		G	A	missense	2	1	non synonymous	G24D		
1	55505596		G	A	missense	2	0	non synonymous	R29H		
1	55505601		C	A	missense	1	0	non synonymous	Q31K		
1	55505604		G	A	missense	292	147	non synonymous	E32K	PCSK9_00008	GoF
1	55505625		G	A	missense	1	0	damaging (broad)	E39K		
1	55505668	rs11583680	C	T	missense	2336	2056	non synonymous	A53V	PCSK9_00011	
1	55505671		A	C	missense	3	6	non synonymous	E54A	PCSK9_00012	GoF
1	55505684		C	A	missense	1	0	non synonymous	H58Q		
1	55505692		CA	C	indel frameshift	1	0	disruptive	T61fs		

1	55505712		G	A	missense	9	4	damaging (broad)	A68T	PCSK9_00014	
1	55509520		C	T	missense	0	2	damaging (broad)	P71L		
1	55509543		G	A	missense	1	2	damaging (broad)	V79M		
1	55509561		G	A	missense	4	1	damaging (broad)	E85K		
1	55509577		A	G	missense	1	0	damaging (broad)	Q90R		
1	55509582		G	C	missense	1	0	damaging (broad)	E92Q		
1	55509585	rs151193009	C	T	missense	134	177	damaging (broad)	R93C	PCSK9_00019	LoF
1	55509594	rs185392267	C	T	missense	1	1	damaging (broad)	R96C		
1	55509598	rs376385276	G	A	missense	0	1	damaging (broad)	R97H		
1	55509606		G	C	missense	0	1	damaging (broad)	A100P		
1	55509618	rs369067856	C	T	missense	1	0	damaging (broad)	R104C	PCSK9_00021	GoF
1	55509630		C	T	missense	1	0	damaging (broad)	L108F		
1	55509648		G	A	missense	0	1	damaging (broad)	V114I		
1	55509661		T	C	missense	1	1	damaging (broad)	L118P		
1	55509704		G	C	missense	2	0	damaging (broad)	E132D		
1	55512254		G	A	missense	0	1	damaging (broad)	S153N		
1	55512299		C	T	missense	5	9	non synonymous	A168V		
1	55517990		G	T	missense	0	0	damaging (strict)	S188I		
1	55518006		C	A	missense	1	0	damaging (strict)	H193Q		
1	55518016		G	A	missense	0	1	damaging (broad)	E197K		
1	55518029		T	G	missense	0	1	damaging (broad)	M201R		
1	55518070		C	T	missense	1	1	damaging (strict)	R215C		
1	55518071		G	A	missense	2	0	damaging (broad)	R215H	PCSK9_00036	GoF
1	55518082		C	G	missense	3	5	damaging (broad)	Q219E	PCSK9_00039	LoF
1	55518329		A	G	missense	0	1	damaging (broad)	K222E		
1	55518362	rs150169598	G	T	missense	1	0	damaging (broad)	V233L		
1	55518374	rs148195424	C	T	missense	0	3	damaging (broad)	R237W	PCSK9_00047	
1	55518375		G	T	missense	0	1	damaging (broad)	R237L		
1	55518381		C	A	missense	0	1	damaging (strict)	A239D	PCSK9_00048	LoF
1	55518383		G	A	missense	0	1	damaging (strict)	G240S		
1	55518386		G	C	missense	0	1	damaging (strict)	V241L		
1	55518407		C	T	missense	0	1	damaging (broad)	R248C		
1	55518417	rs376945520	G	A	missense	2	0	damaging (broad)	R251H		
1	55518452	rs200146448	G	A	missense	130	120	damaging (broad)	G263S	PCSK9_00051	
1	55518456	rs201789841	C	T	missense	124	100	non synonymous	T264I		
1	55521683		A	G	missense	0	1	non synonymous	K273E		

1	55521716		G	A	missense	1	0	damaging (broad)	V284M		
1	55521783		G	T	missense	0	1	non synonymous	R306M		
1	55521794		G	A	missense	1	1	damaging (broad)	V310M		
1	55523034		G	A	missense	0	1	damaging (strict)	D343N		
1	55523076	rs148562777	C	T	missense	0	1	damaging (broad)	R357C		
1	55523119		C	T	missense	0	0	damaging (broad)	A371V		
1	55523141		C	G	missense	0	1	damaging (strict)	C378W		
1	55523188		G	A	splice-site	1	2	disruptive			
1	55523798		A	G	missense	111	129	damaging (broad)	I424V	PCSK9_00064	
1	55523808		C	A	missense	1	0	damaging (broad)	A427D		
1	55523812		G	A	nonsense	4	5	disruptive	W428*	PCSK9_00066	LoF
1	55523829		G	A	missense	1	0	non synonymous	R434Q		
1	55523873		C	T	missense	0	1	non synonymous	H449Y		
1	55524237	rs562556	A	G	missense	625	511	non synonymous	V474I	PCSK9_00074	
1	55524244	rs376388695	G	A	missense	1	0	non synonymous	R476H		
1	55524246		T	A	missense	1	0	damaging (strict)	C477S		
1	55524300		C	T	missense	0	2	damaging (broad)	R495W		
1	55524303	rs374603772	C	T	missense	1	2	damaging (broad)	R496W	PCSK9_00076	GoF
1	55524304	rs139669564	G	A	missense	1	0	non synonymous	R496Q	PCSK9_00077	
1	55524312	rs201395805	C	T	missense	2	6	damaging (broad)	R499C		
1	55525165	rs374455190	G	T	missense	1	1	damaging (broad)	G504W		
1	55525183		C	T	missense	1	0	damaging (broad)	R510W		
1	55525195		G	A	missense	6	3	damaging (broad)	A514T	PCSK9_00078	GoF
1	55525219		G	A	missense	4	2	damaging (strict)	A522T	PCSK9_00080	GoF
1	55525241		T	G	missense	1	0	damaging (broad)	L529R		
1	55525261		G	A	missense	0	1	non synonymous	V536I		
1	55525295		G	T	missense	0	1	non synonymous	G547V		
1	55527059		C	T	missense	1	1	damaging (broad)	H565Y		
1	55527062		T	A	missense	2	0	non synonymous	W566R		
1	55527093	rs72646525	C	T	missense	4	4	non synonymous	P576L		
1	55527110	rs373323910	C	T	nonsense	1	0	disruptive	R582*		
1	55527131	rs372586224	G	A	missense	5	5	non synonymous	V589M		
1	55527141		G	C	missense	2	1	non synonymous	R592T		
1	55527158	rs367606156	G	A	missense	3	4	damaging (broad)	A598T		
1	55527174		C	G	missense	1	1	damaging (broad)	A603G		
1	55527190		CAA	C	indel	0	1	disruptive	K609fs		

				frameshift							
1	55527204	A	C	missense	0	1	non synonymous	H613P			
1	55529048	G	A	missense	1	1	damaging (broad)	V624M	PCSK9_00089	GoF	
1	55529064	G	A	missense	10	7	damaging (strict)	G629D			
1	55529108	rs143291739	G	A	missense	65	66	non synonymous	V644I	PCSK9_00091	
1	55529132	rs201280059	A	G	missense	2	1	damaging (broad)	N652D		
1	55529153		C	G	missense	10	8	non synonymous	R659G		
1	55529182		C	A	missense	15	8	non synonymous	S668R	PCSK9_00093	LoF
1	55529187	rs505151	A	G	missense	932	720	non synonymous	G670E	PCSK9_00094	
1	55529217		G	A	missense	0	1	damaging (broad)	R680Q		
1	55529243		C	T	nonsense	0	2	disruptive	Q689*		

---

## 4 考察

本研究では心筋梗塞を対象とし大規模サンプルを使用したターゲットリシークエンスを行った。ターゲット遺伝子は過去の GWAS に基づいて決定した。結果として *LDLR* と *PCSK9* の稀少遺伝子変異の集積が心筋梗塞と関連していることを示した。これらの関連は血清 LDL コレステロール値の変化によってもたらされていることが示唆された。

*LDLR* の稀少遺伝子変異の集積が心筋梗塞のリスクとなることは西欧人を用いた whole-exome study により 2015 年 2 月に CARDIOGRAMplusC4D コンソーシアムによって報告<sup>24</sup>されたため、本研究の結果は同様の関連を日本人において独立に示したものとなる。本研究においても先行研究と同様のグルーピングを用いて結果を示したが、同じグルーピングの OR は非常によく一致しており、*LDLR* の心筋梗塞に与える影響はヨーロッパ集団とアジア集団の間で共通していることが示唆される。一方で、個々の変異を見ていくと、その内訳は大きく異なっている。先行論文において公開されている変異のリストと本研究で特定した変異のリスト（表 9）を比較した場合、本研究では *LDLR* の変異

（synonymous 変異は除く）を 138 特定したが、コンソーシアムは 156 を特定している。そのうち 28 変異（20%）が共通であるが残りは異なっている。*LDLR* の心筋梗塞に与える影響はヨーロッパ集団、アジア集団の間で一定であるが、

個々の rare variant は集団によって大きく異なる、という事実は心筋梗塞という common disease の遺伝様式に関する示唆を与える。

*PCSK9* と心筋梗塞との関連も今回示された。*PCSK9* については高脂血症家系の解析<sup>18</sup> から gain-of-function 変異のキャリアが高い血清 LDL コレステロール値を持ち、心筋梗塞のハイリスクであることが知られている。一方で loss-of-function 変異のキャリアは低い血清 LDL コレステロール値を持ち、心筋梗塞含む虚血性心疾患のリスクが低い<sup>19-21</sup> ことも知られている。本研究においても E32K と R93C という 2 つの低頻度変異がそれぞれ risk と protective の 2 方向で日本人における心筋梗塞のリスクに影響を与えていることが示唆された。

Disruptive 変異のキャリアの血清コレステロール値が非常に低く、心筋梗塞群における発症年齢も高齢であったことは *PCSK9* における稀少変異の集積が心筋梗塞に対して protective な効果を持っていることを強く示唆する。ただ、*PCSK9* の disruptive 変異は *LDLR* におけるそれと比較してさらに頻度が低く、今回の研究で統計学的有意差を持ってこれを示すことはできなかった。これを示すためにはさらに大規模な meta-analysis などが待たれる。

*PCSK9* の disruptive 変異のキャリアの血清コレステロール値、および心筋梗塞発症年齢からは *PCSK9* の機能低下が生涯に渡るコレステロール低下作用、ひいては虚血性心疾患からの抵抗的な効果を持つことを示唆している。これは既に

家系の解析では示されている事実ではあるが、本研究において一般集団における心筋梗塞においてこれが示されたことは、よりこの事実を補強するものと考ええる。現在、新規の脂質異常症治療薬として米国で治験中の抗 *PCSK9* 抗体は血清 LDL コレステロール値低下効果が既に示され、現在虚血性心疾患に対する低減効果が評価されている段階<sup>36</sup>であるが、本研究の結果から判断すると虚血性心疾患に対する効果は **gain-of-function** 変異の有無によって大きく異なる可能性があり、特に今後 **E32K** キャリアの頻度が比較的高い日本で使用される段階に至った場合には特別な注意が必要と考えられる。

*PCSK9* は先述のコンソーシアムの **whole-exome study** では挙がっていない遺伝子である。ほぼ同等の規模である本研究でこれを指摘することができた理由は以下の2つであると考ええる。まず第一に今回観察された関連に大きな影響を与えている **E32K** 変異が東アジア人に固有であり、西欧人では存在していないこと（この変異は単独で **OR 1.7** であり頻度も心筋梗塞群で **1.5%**、対照群でも **0.9%** と比較的高い）、第二に *PCSK9* は **loss-of-function** 変異の他に **gain-of-function** 変異が比較的多く見られることが知られているが、両方向の効果は互いに打ち消し合ってしまうため今回用いた **SKAT** のような手法を用いない限り見落としてしまう可能性がある。

*TCTN1* は本研究において **gene-wide** 有意な関連は観察されなかったものの

suggestive な関連が観察された遺伝子である。Tctn1 は絨毛関連疾患の原因遺伝子として知られているが虚血性心疾患の関連ではこれまでほとんど知見がない遺伝子である。TCTN1 は染色体の 12q24 領域に存在するが、GWAS ではこの領域のシグナルは SH2B3 によるものとして説明されることが多い<sup>3-6, 8</sup>、一方日本を含むアジアではこの領域のシグナルは BRAP によって説明されると考えられている<sup>37</sup>。一方で Erbilgin らはヒト・マウスの血管内皮細胞を用いた eQTL 解析で、この領域の proxy SNP は MAPKAPK5 および TCTN1 の転写レベルと関連していると報告<sup>25</sup>している。本研究では SH2B3, BRAP, MAPKAPK5, TCTN1 全てをターゲットとしてリシーケンスを行ったが TCTN1 のみで suggestive な関連が見られている。この結果から 12q24 領域においては TCTN1 についても注目する必要があると考える。

今回有意な関連が見られた 2 つの遺伝子、LDLR と PCSK9 はいずれも家族性脂質異常症の家系の解析から原因遺伝子として広く知られている遺伝子であった。実際 LDLR については 42% (59/138)、PCSK9 については 24% (23/94) の変異がシーケンスバリエントのデータベースに既に登録されている変異であった。本研究では家族性高脂血症の患者を特定して除外するようなデザインは不可能であったことから、そもそものサンプルに家族性高脂血症の患者が含まれていたに過ぎないという解釈も可能ではある。しかし、これらの変異を持

つものを家族性高脂血症と定義することになると、その割合は対照群の 0.44% に及ぶ。またデータベースに登録されていない *LDLR*, *PCSK9* の変異

(synonymous 変異は含まない) を持つものも含めるとその割合は 1.3% に及ぶこととなる。いずれの数値も日本人における家族性脂質異常症の罹患率の推測値である 0.2% を大きく上回っている。これには以下の 2 通りの解釈が考えられる、1 つは common disease である心筋梗塞の遺伝性の一部は家族性の脂質異常症の原因としても知られる変異が孤発的に生じることで説明されるという考え方。もう 1 つはこれらの変異のキャリアは正に家族性脂質異常症のキャリアであって、家族性脂質異常症が一般に考えられているより遥かに多数存在しており、そうした患者が心筋梗塞のハイリスクである、という考え方。いずれかが正しいのか、本研究の結果のみで結論することはできない。しかし、いずれにしてもこれらの効果は血清 LDL コレステロール値を介したものであるため、そのリスクは内服治療などによる治療介入で相殺することができるものである。本研究はこうしたリスク患者に対する積極的な治療介入の有効性を遺伝の観点からも強調したものといえる。

最後に本研究の限界について述べる、本研究はターゲットシーケンスであり、すべての GWAS 関連遺伝子を対象としていない。このため心筋梗塞の missing heritability について本論文で何らかの結論を出すことはできない。また直近の

大規模 GWAS メタ解析の報告<sup>9</sup>によると心筋梗塞の heritability は rare variant によってというよりは一つ一つは影響の大きくない非常に多数の variant によって説明されるという仮説のほうが支持されるようになってきている。ただし、本研究で示したように心筋梗塞においてもレアバリエントによって説明されるものが一部はあると考えられる。

また、本研究の **Replication Stage** で使用したコントロールサンプルは 5 疾患のケースミックスであり健常者コントロールではない。仮にこれらの 5 疾患と関連するレアバリエントがこれらのサンプルに集積していた場合、結果に影響を与えてしまう可能性が考えられる。この影響を検討するために **Discovery Stage** と **Replication Stage** で使用したコントロールサンプルの間で、本実験で行ったのと同様の遺伝子ごとの関連解析(SKATのみ)を行ったがいずれの遺伝子も有意な相関は示さなかった。ただし、**Replication Stage** に進んだ 11 遺伝子でしか評価ができていないこと、本実験と比較してサンプルサイズがほぼ半分であり検出力が低いという限界は残っている。

まとめると、本研究では日本人心筋梗塞患者の大規模サンプルを用いてターゲット・リシークエンスを行った。結果として *LDLR*, *PCSK9* の 2 遺伝子の稀少変異の集積が心筋梗塞のリスクと関連していることを示した。また *LDLR* と *PCSK9* の心筋梗塞への影響の違いを明らかにすることを通じて、複雑で理解の

進んでいなかった common disease の遺伝様式についての新しい知見をもたらしたと言える。

1. Ruff CT, Braunwald E. The evolving epidemiology of acute coronary syndromes. *Nat Rev Cardiol.* 2011;8:140-147
2. Manolio TA, Collins FS, Cox NJ, Goldstein DB, Hindorff LA, Hunter DJ, McCarthy MI, Ramos EM, Cardon LR, Chakravarti A, Cho JH, Guttmacher AE, Kong A, Kruglyak L, Mardis E, Rotimi CN, Slatkin M, Valle D, Whittemore AS, Boehnke M, Clark AG, Eichler EE, Gibson G, Haines JL, Mackay TFC, McCarroll SA, Visscher PM. Finding the missing heritability of complex diseases. *Nature.* 2009;461:747-753
3. McPherson R, Pertsemlidis A, Kavaslar N, Stewart A, Roberts R, Cox DR, Hinds DA, Pennacchio LA, Tybjaerg-Hansen A, Folsom AR, Boerwinkle E, Hobbs HH, Cohen JC. A common allele on chromosome 9 associated with coronary heart disease. *Science.* 2007;316:1488-1491
4. Samani NJ, Erdmann J, Hall AS, Hengstenberg C, Mangino M, Mayer B, Dixon RJ, Meitinger T, Braund P, Wichmann HE, Barrett JH, Konig IR, Stevens SE, Szymczak S, Tregouet DA, Iles MM, Pahlke F, Pollard H, Lieb W, Cambien F, Fischer M, Ouwehand W, Blankenberg S, Balmforth AJ, Baessler A, Ball SG, Strom TM, Braenne I, Gieger C, Deloukas P, Tobin MD, Ziegler A, Thompson JR, Schunkert H. Genomewide association analysis of coronary artery disease. *N Engl J Med.* 2007;357:443-453
5. Kathiresan S, Voight BF, Purcell S, Musunuru K, Ardissino D, Mannucci PM, Anand S, Engert JC, Samani NJ, Schunkert H, Erdmann J, Reilly MP, Rader DJ, Morgan T, Spertus JA, Stoll M, Girelli D, McKeown PP, Patterson CC, Siscovick DS, O'Donnell CJ, Elosua R, Peltonen L, Salomaa V, Schwartz SM, Melander O, Altshuler D, Merlini PA, Berzuini C, Bernardinelli L, Peyvandi F, Tubaro M, Celli P, Ferrario M, Fetiveau R, Marziliano N, Casari G, Galli M, Ribichini F, Rossi M, Bernardi F, Zonzin P, Piazza A, Yee J, Friedlander Y, Marrugat J, Lucas G, Subirana I, Sala J, Ramos R, Meigs JB, Williams G, Nathan DM, MacRae CA, Havulinna AS, Berglund G, Hirschhorn JN, Asselta R, Duga S, Spreatico M, Daly MJ, Nemes J, Korn JM, McCarroll SA, Surti A, Guiducci C, Gianniny L, Mirel D, Parkin M, Burt N, Gabriel SB, Thompson JR, Braund PS, Wright BJ, Balmforth AJ, Ball SG, Hall A, Linsel-Nitschke P, Lieb W, Ziegler A, Konig I, Hengstenberg C, Fischer M, Stark K, Grosshennig A, Preuss M, Wichmann HE, Schreiber S, Ouwehand W, Deloukas P, Scholz M, Cambien F, Li M, Chen Z, Wilensky R, Matthai W, Qasim A, Hakonarson HH, Devaney J, Burnett MS, Pichard AD, Kent KM, Satler L, Lindsay JM, Waksman R, Knouff CW, Waterworth DM, Walker MC, Mooser V, Epstein SE, Scheffold T, Berger K, Huge A, Martinelli N, Olivieri O, Corrocher R, McKeown P, Erdmann E, Konig IR, Holm H, Thorleifsson G, Thorsteinsdottir U, Stefansson K, Do R, Xie C, Siscovick D. Genome-wide association of early-onset myocardial

- infarction with single nucleotide polymorphisms and copy number variants. *Nat Genet.* 2009;41:334-341
6. Schunkert H, König IR, Kathiresan S, Reilly MP, Assimes TL, Holm H, Preuss M, Stewart AF, Barbalic M, Gieger C, Absher D, Aherrahrou Z, Allayee H, Altshuler D, Anand SS, Andersen K, Anderson JL, Ardissino D, Ball SG, Balmforth AJ, Barnes TA, Becker DM, Becker LC, Berger K, Bis JC, Boekholdt SM, Boerwinkle E, Braund PS, Brown MJ, Burnett MS, Buysschaert I, Carlquist JF, Chen L, Cichon S, Codd V, Davies RW, Dedoussis G, Dehghan A, Demissie S, Devaney JM, Diemert P, Do R, Doering A, Eifert S, Mokhtari NE, Ellis SG, Elosua R, Engert JC, Epstein SE, de Faire U, Fischer M, Folsom AR, Freyer J, Gigante B, Girelli D, Gretarsdottir S, Gudnason V, Gulcher JR, Halperin E, Hammond N, Hazen SL, Hofman A, Horne BD, Illig T, Iribarren C, Jones GT, Jukema JW, Kaiser MA, Kaplan LM, Kastelein JJ, Khaw KT, Knowles JW, Kolovou G, Kong A, Laaksonen R, Lambrechts D, Leander K, Lettre G, Li M, Lieb W, Loley C, Lotery AJ, Mannucci PM, Maouche S, Martinelli N, McKeown PP, Meisinger C, Meitinger T, Melander O, Merlini PA, Mooser V, Morgan T, Muhleisen TW, Muhlestein JB, Munzel T, Musunuru K, Nahrstaedt J, Nelson CP, Nothen MM, Olivieri O, Patel RS, Patterson CC, Peters A, Peyvandi F, Qu L, Quyyumi AA, Rader DJ, Rallidis LS, Rice C, Rosendaal FR, Rubin D, Salomaa V, Sampietro ML, Sandhu MS, Schadt E, Schafer A, Schillert A, Schreiber S, Schrezenmeier J, Schwartz SM, Siscovick DS, Sivananthan M, Sivapalaratnam S, Smith A, Smith TB, Snoop JD, Soranzo N, Spertus JA, Stark K, Stirrups K, Stoll M, Tang WH, Tennstedt S, Thorgeirsson G, Thorleifsson G, Tomaszewski M, Uitterlinden AG, van Rij AM, Voight BF, Wareham NJ, Wells GA, Wichmann HE, Wild PS, Willenborg C, Wittteman JC, Wright BJ, Ye S, Zeller T, Ziegler A, Cambien F, Goodall AH, Cupples LA, Quertermous T, Marz W, Hengstenberg C, Blankenberg S, Ouwehand WH, Hall AS, Deloukas P, Thompson JR, Stefansson K, Roberts R, Thorsteinsdottir U, O'Donnell CJ, McPherson R, Erdmann J, Samani NJ. Large-scale association analysis identifies 13 new susceptibility loci for coronary artery disease. *Nat Genet.* 2011;43:333-338
7. Lu X, Wang L, Chen S, He L, Yang X, Shi Y, Cheng J, Zhang L, Gu CC, Huang J, Wu T, Ma Y, Li J, Cao J, Chen J, Ge D, Fan Z, Li Y, Zhao L, Li H, Zhou X, Chen L, Liu D, Duan X, Hao Y, Lu F, Liu Z, Yao C, Shen C, Pu X, Yu L, Fang X, Xu L, Mu J, Wu X, Zheng R, Wu N, Zhao Q, Liu X, Wang M, Yu D, Hu D, Ji X, Guo D, Sun D, Wang Q, Yang Y, Liu F, Mao Q, Liang X, Ji J, Chen P, Mo X, Li D, Chai G, Tang Y, Li X, Du Z, Dou C, Yang Z, Meng Q, Wang D, Wang R, Yang J, Schunkert H, Samani NJ, Kathiresan S, Reilly MP, Erdmann J, Peng X, Chen R, Qiang B, Gu D. Genome-wide association study in Han Chinese identifies four new susceptibility loci for coronary

- artery disease. *Nat Genet.* 2012;44:890-894
8. the CARDIoGRAMplusC4D Consortium, Deloukas P, Kanoni S, Willenborg C, Farrall M, Assimes TL, Thompson JR, Ingelsson E, Saleheen D, Erdmann J, Goldstein BA, Stirrups K, Konig IR, Cazier JB, Johansson A, Hall AS, Lee JY, Willer CJ, Chambers JC, Esko T, Folkersen L, Goel A, Grundberg E, Havulinna AS, Ho WK, Hopewell JC, Eriksson N, Kleber ME, Kristiansson K, Lundmark P, Lyytikäinen LP, Rafelt S, Shungin D, Strawbridge RJ, Thorleifsson G, Tikkanen E, Van Zuydam N, Voight BF, Waite LL, Zhang W, Ziegler A, Absher D, Altshuler D, Balmforth AJ, Barroso I, Braund PS, Burgdorf C, Claudi-Boehm S, Cox D, Dimitriou M, Do R, Doney AS, El Mokhtari N, Eriksson P, Fischer K, Fontanillas P, Franco-Cereceda A, Gigante B, Groop L, Gustafsson S, Hager J, Hallmans G, Han BG, Hunt SE, Kang HM, Illig T, Kessler T, Knowles JW, Kolovou G, Kuusisto J, Langenberg C, Langford C, Leander K, Lokki ML, Lundmark A, McCarthy MI, Meisinger C, Melander O, Mihailov E, Maouche S, Morris AD, Muller-Nurasyid M, Nikus K, Peden JF, Rayner NW, Rasheed A, Rosinger S, Rubin D, Rumpf MP, Schafer A, Sivananthan M, Song C, Stewart AF, Tan ST, Thorgeirsson G, van der Schoot CE, Wagner PJ, Wells GA, Wild PS, Yang TP, Amouyel P, Arveiler D, Basart H, Boehnke M, Boerwinkle E, Brambilla P, Cambien F, Cupples AL, de Faire U, Dehghan A, Diemert P, Epstein SE, Evans A, Ferrario MM, Ferrieres J, Gauguier D, Go AS, Goodall AH, Gudnason V, Hazen SL, Holm H, Iribarren C, Jang Y, Kahonen M, Kee F, Kim HS, Klopp N, Koenig W, Kratzer W, Kuulasmaa K, Laakso M, Laaksonen R, Lind L, Ouwehand WH, Parish S, Park JE, Pedersen NL, Peters A, Quertermous T, Rader DJ, Salomaa V, Schadt E, Shah SH, Sinisalo J, Stark K, Stefansson K, Tregouet DA, Virtamo J, Wallentin L, Wareham N, Zimmermann ME, Nieminen MS, Hengstenberg C, Sandhu MS, Pastinen T, Syvanen AC, Hovingh GK, Dedoussis G, Franks PW, Lehtimäki T, Metspalu A, Zalloua PA, Siegbahn A, Schreiber S, Ripatti S, Blankenberg SS, Perola M, Clarke R, Boehm BO, O'Donnell C, Reilly MP, Marz W, Collins R, Kathiresan S, Hamsten A, Kooner JS, Thorsteinsdottir U, Danesh J, Palmer CN, Roberts R, Watkins H, Schunkert H, Samani NJ. Large-scale association analysis identifies new risk loci for coronary artery disease. *Nat Genet.* 2013;45:25-33
  9. Majid Nikpay, Anuj Goel, Hong-Hee Won, Leanne M Hall, Christina Willenborg, Stavroula Kanoni, Danish Saleheen, Theodosios Kyriakou, Christopher P Nelson, Jemma C Hopewell, Thomas R Webb, Lingyao Zeng, Abbas Dehghan, Maris Alver, Sebastian M Armasu, Kirsi Auro, Andrew Bjornes, Daniel I Chasman, Shufeng Chen, Ian Ford, Nora Franceschini, Christian Gieger, Christopher Grace, Stefan Gustafsson, Jie Huang, Shih-Jen Hwang, Yun Kyoung Kim, Marcus E Kleber, King Wai Lau, Xiangfeng Lu, Yingchang Lu, Leo-Pekka Lyytikäinen, Evelin Mihailov,

- Alanna C Morrison, Natalia Pervjakova, Liming Qu, Lynda M Rose, Elias Salfati, Richa Saxena, Markus Scholz, Albert V Smith, Emmi Tikkanen, Andre Uitterlinden, Xueli Yang, Weihua Zhang, Wei Zhao, Mariza de Andrade, Paul S de Vries, Natalie R van Zuydam, Sonia S Anand, Lars Bertram, Frank Beutner, George Dedoussis, Philippe Frossard, Dominique Gauguier, Alison H Goodall, Omri Gottesman, Marc Haber, Bok-Ghee Han, Jianfeng Huang, Shapour Jalilzadeh, Thorsten Kessler, Inke R König, Lars Lannfelt, Wolfgang Lieb, Lars Lind, Cecilia M Lindgren, Marja-Liisa Lokki, Patrik K Magnusson, Nadeem H Mallick, Narinder Mehra, Thomas Meitinger, Fazal-ur-Rehman Memon, Andrew P Morris, Markku S Nieminen, Nancy L Pedersen, Annette Peters, Loukianos S Rallidis, Asif Rasheed, Maria Samuel, Svati H Shah, Juha Sinisalo, Kathleen E Stirrups, Stella Trompet, Laiyuan Wang, Khan S Zaman, Diego Ardissino, Eric Boerwinkle, Ingrid B Borecki, Erwin P Bottinger, Julie E Buring, John C Chambers, Rory Collins, L Adrienne Cupples, John Danesh, Ilja Demuth, Roberto Elosua, Stephen E Epstein, Tõnu Esko, Mary F Feitosa, Oscar H Franco, Maria Grazia Franzosi, Christopher B Granger, Dongfeng Gu, Vilmundur Gudnason, Alistair S Hall, Anders Hamsten, Tamara B Harris, Stanley L Hazen, Christian Hengstenberg, Albert Hofman, Erik Ingelsson, Carlos Iribarren, J Wouter Jukema, Pekka J Karhunen, Bong-Jo Kim, Jaspal S Kooner, Iftikhar J Kullo, Terho Lehtimäki, Ruth J F Loos, Olle Melander, Andres Metspalu, Winfried März, Colin N Palmer, Markus Perola, Thomas Quertermous, Daniel J Rader, Paul M Ridker, Samuli Ripatti, Robert Roberts, Veikko Salomaa, Dharambir K Sanghera, Stephen M Schwartz, Udo Seedorf, Alexandre F Stewart, David J Stott, Joachim Thiery, Pierre A Zalloua, Christopher J O'Donnell, Muredach P Reilly, Themistocles L Assimes, John R Thompson, Jeanette Erdmann, Robert Clarke, Hugh Watkins, Sekar Kathiresan, Ruth McPherson, Panos Deloukas, Heribert Schunkert, Nilesh J Samani & Martin Farrall for the CARDIoGRAMplusC4D Consortium. A comprehensive 1000 genomes-based genome-wide association meta-analysis of coronary artery disease. *Nat Genet.* 2015;47:1121-1130
10. Ozaki K, Ohnishi Y, Iida A, Sekine A, Yamada R, Tsunoda T, Sato H, Hori M, Nakamura Y, Tanaka T. Functional snps in the lymphotoxin-alpha gene that are associated with susceptibility to myocardial infarction. *Nat Genet.* 2002;32:650-654
  11. Mayer B, Erdmann J, Schunkert H. Genetics and heritability of coronary artery disease and myocardial infarction. *Clin Res Cardiol.* 2007;96:1-7
  12. Bonnefond A, Clement N, Fawcett K, Yengo L, Vaillant E, Guillaume J-L, Dechaume A, Payne F, Roussel R, Czernichow S, Hercberg S, Hadjadj S, Balkau B, Marre M, Lantieri O, Langenberg C, Bouatia-Naji N, Charpentier G, Vaxillaire M, Rocheleau G, Wareham NJ, Sladek R, McCarthy MI, Dina C, Barroso I, Jockers R, Froguel P.

- Rare *MTNR1B* variants impairing melatonin receptor 1b function contribute to type 2 diabetes. *Nat Genet.* 2012;44:297-301
13. Sanna S, Li B, Mulas A, Sidore C, Kang HM, Jackson AU, Piras MG, Usala G, Maninchedda G, Sassu A, Serra F, Palmas MA, Wood WH, Njølstad I, Laakso M, Hveem K, Tuomilehto J, Lakka TA, Rauramaa R, Boehnke M, Cucca F, Uda M, Schlessinger D, Nagaraja R, Abecasis GR. Fine mapping of five loci associated with low-density lipoprotein cholesterol detects variants that double the explained heritability. *PLoS genetics.* 2011;7:e1002198
  14. Johansen CT, Wang J, Lanktree MB, Cao H, McIntyre AD, Ban MR, Martins RA, Kennedy BA, Hassell RG, Visser ME, Schwartz SM, Voight BF, Elosua R, Salomaa V, O'Donnell CJ, Dallinga-Thie GM, Anand SS, Yusuf S, Huff MW, Kathiresan S, Hegele RA. Excess of rare variants in genes identified by genome-wide association study of hypertriglyceridemia. *Nature genetics.* 2010;42:684-687
  15. Seddon JM, Yu Y, Miller EC, Reynolds R, Tan PL, Gowrisankar S, Goldstein JI, Triebwasser M, Anderson HE, Zerbib J, Kavanagh D, Souied E, Katsanis N, Daly MJ, Atkinson JP, Raychaudhuri S. Rare variants in *CFI*, *C3* and *C9* are associated with high risk of advanced age-related macular degeneration. *Nat Genet.* 2013;45:1366-1370
  16. Zhan X, Larson DE, Wang C, Koboldt DC, Sergeev YV, Fulton RS, Fulton LL, Fronick CC, Branham KE, Bragg-Gresham J, Jun G, Hu Y, Kang HM, Liu D, Othman M, Brooks M, Ratnapriya R, Boleda A, Grassmann F, von Strachwitz C, Olson LM, Buitendijk GH, Hofman A, van Duijn CM, Cipriani V, Moore AT, Shahid H, Jiang Y, Conley YP, Morgan DJ, Kim IK, Johnson MP, Cantsilieris S, Richardson AJ, Guymer RH, Luo H, Ouyang H, Licht C, Pluthero FG, Zhang MM, Zhang K, Baird PN, Blangero J, Klein ML, Farrer LA, DeAngelis MM, Weeks DE, Gorin MB, Yates JR, Klaver CC, Pericak-Vance MA, Haines JL, Weber BH, Wilson RK, Heckenlively JR, Chew EY, Stambolian D, Mardis ER, Swaroop A, Abecasis GR. Identification of a rare coding variant in complement 3 associated with age-related macular degeneration. *Nat Genet.* 2013;45:1375-1379
  17. Rivas MA, Beaudoin M, Gardet A, Stevens C, Sharma Y, Zhang CK, Boucher G, Ripke S, Ellinghaus D, Burt N, Fennell T, Kirby A, Latiano A, Goyette P, Green T, Halfvarson J, Haritunians T, Korn JM, Kuruvilla F, Lagace C, Neale B, Lo KS, Schumm P, Torkvist L, Dubinsky MC, Brant SR, Silverberg MS, Duerr RH, Altshuler D, Gabriel S, Lettre G, Franke A, D'Amato M, McGovern DPB, Cho JH, Rioux JD, Xavier RJ, Daly MJ. Deep resequencing of GWAS loci identifies independent rare variants associated with inflammatory bowel disease. *Nat Genet.* 2011;43:1066-1073

18. Abifadel M, Varret M, Rabes JP, Allard D, Ouguerram K, Devillers M, Cruaud C, Benjannet S, Wickham L, Erlich D, Derre A, Villegier L, Farnier M, Beucler I, Bruckert E, Chambaz J, Chanu B, Lecerf JM, Luc G, Moulin P, Weissenbach J, Prat A, Krempf M, Junien C, Seidah NG, Boileau C. Mutations in *PCSK9* cause autosomal dominant hypercholesterolemia. *Nat Genet.* 2003;34:154-156
19. Benn M, Nordestgaard BG, Grande P, Schnohr P, Tybjaerg-Hansen A. *PCSK9* R46L, low-density lipoprotein cholesterol levels, and risk of ischemic heart disease: 3 independent studies and meta-analyses. *J Am Coll Cardiol.* 2010;55:2833-2842
20. Cohen J, Pertsemlidis A, Kotowski IK, Graham R, Garcia CK, Hobbs HH. Low ldl cholesterol in individuals of African descent resulting from frequent nonsense mutations in *PCSK9*. *Nat Genet.* 2005;37:161-165
21. Cohen JC, Boerwinkle E, Mosley TH, Jr., Hobbs HH. Sequence variations in *PCSK9*, low LDL, and protection against coronary heart disease. *N Engl J Med.* 2006;354:1264-1272
22. Santer R, Calado J. Familial renal glucosuria and *SGLT2*: From a mendelian trait to a therapeutic target. *Clin J Am Soc Nephrol.* 2010;5:133-141
23. Beaudoin M, Lo KS, N'Diaye A, Rivas MA, Dube MP, Laplante N, Phillips MS, Rioux JD, Tardif JC, Lettre G. Pooled DNA resequencing of 68 myocardial infarction candidate genes in French Canadians. *Circ Cardiovasc Genet.* 2012;5:547-554
24. Do R, Stitzel NO, Won HH, Jorgensen AB, Duga S, Angelica Merlini P, Kiezun A, Farrall M, Goel A, Zuk O, Guella I, Asselta R, Lange LA, Peloso GM, Auer PL, Girelli D, Martinelli N, Farlow DN, DePristo MA, Roberts R, Stewart AF, Saleheen D, Danesh J, Epstein SE, Sivapalaratnam S, Hovingh GK, Kastelein JJ, Samani NJ, Schunkert H, Erdmann J, Shah SH, Kraus WE, Davies R, Nikpay M, Johansen CT, Wang J, Hegele RA, Hechter E, Marz W, Kleber ME, Huang J, Johnson AD, Li M, Burke GL, Gross M, Liu Y, Assimes TL, Heiss G, Lange EM, Folsom AR, Taylor HA, Olivieri O, Hamsten A, Clarke R, Reilly DF, Yin W, Rivas MA, Donnelly P, Rossouw JE, Psaty BM, Herrington DM, Wilson JG, Rich SS, Bamshad MJ, Tracy RP, Cupples LA, Rader DJ, Reilly MP, Spertus JA, Cresci S, Hartiala J, Tang WH, Hazen SL, Allayee H, Reiner AP, Carlson CS, Kooperberg C, Jackson RD, Boerwinkle E, Lander ES, Schwartz SM, Siscovick DS, McPherson R, Tybjaerg-Hansen A, Abecasis GR, Watkins H, Nickerson DA, Ardissino D, Sunyaev SR, O'Donnell CJ, Altshuler D, Gabriel S, Kathiresan S. Exome sequencing identifies rare *LDLR* and *APOA5* alleles conferring risk for myocardial infarction. *Nature.* 2015;518:102-106
25. Erbilgin A, Civelek M, Romanoski CE, Pan C, Hagopian R, Berliner JA, Lusis AJ. Identification of cad candidate genes in GWAS loci and their expression in vascular

- cells. *J Lipid Res.* 2013;54:1894-1905
26. Pruitt KD, Harrow J, Harte RA, Wallin C, Diekhans M, Maglott DR, Searle S, Farrell CM, Loveland JE, Ruef BJ, Hart E, Suner MM, Landrum MJ, Aken B, Ayling S, Baertsch R, Fernandez-Banet J, Cherry JL, Curwen V, Dicuccio M, Kellis M, Lee J, Lin MF, Schuster M, Shkeda A, Amid C, Brown G, Dukhanina O, Frankish A, Hart J, Madaid BL, Mudge J, Murphy MR, Murphy T, Rajan J, Rajput B, Riddick LD, Snow C, Steward C, Webb D, Weber JA, Wilming L, Wu W, Birney E, Haussler D, Hubbard T, Ostell J, Durbin R, Lipman D. The consensus coding sequence (CCDS) project: Identifying a common protein-coding gene set for the human and mouse genomes. *Genome Res.* 2009;19:1316-1323
  27. Marcel M. Cutadapt removes adapter sequences from high-throughput sequencing reads. 2011;17:10-12
  28. Li H, Durbin R. Fast and accurate short read alignment with burrows-wheeler transform. *Bioinformatics.* 2009;25:1754-1760
  29. McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernytsky A, Garimella K, Altshuler D, Gabriel S, Daly M, DePristo MA. The genome analysis toolkit: A mapreduce framework for analyzing next-generation dna sequencing data. *Genome Res.* 2010;20:1297-1303
  30. Cingolani P, Platts A, Wang le L, Coon M, Nguyen T, Wang L, Land SJ, Lu X, Ruden DM. A program for annotating and predicting the effects of single nucleotide polymorphisms, Snpeff: Snps in the genome of drosophila melanogaster strain w1118; iso-2; iso-3. *Fly (Austin).* 2012;6:80-92
  31. Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MA, Bender D, Maller J, Sklar P, de Bakker PI, Daly MJ, Sham PC. Plink: A tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet.* 2007;81:559-575
  32. Purcell SM, Moran JL, Fromer M, Ruderfer D, Solovieff N, Roussos P, O'Dushlaine C, Chambert K, Bergen SE, Kahler A, Duncan L, Stahl E, Genovese G, Fernandez E, Collins MO, Komiyama NH, Choudhary JS, Magnusson PK, Banks E, Shakir K, Garimella K, Fennell T, DePristo M, Grant SG, Haggarty SJ, Gabriel S, Scolnick EM, Lander ES, Hultman CM, Sullivan PF, McCarroll SA, Sklar P. A polygenic burden of rare disruptive mutations in schizophrenia. *Nature.* 2014;506:185-190
  33. Flanagan SE, Patch AM, Ellard S. Using SIFT and Polyphen to predict loss-of-function and gain-of-function mutations. *Genet Test Mol Biomarkers.* 2010;14:533-537
  34. Leigh SE, Foster AH, Whittall RA, Hubbart CS, Humphries SE. Update and analysis of the University college London low density lipoprotein receptor familial

- hypercholesterolemia database. *Ann Hum Genet.* 2008;72:485-498
35. Leigh SE, Leren TP, Humphries SE. Commentary *PCSK9* variants: A new database. *Atherosclerosis.* Ireland; 2009:32-33.
  36. Wu NQ, Li S, Li JJ. Update of clinical trials of anti-*PCSK9* antibodies. *Cardiovasc Drugs Ther.* 2015;29:159-169
  37. Ozaki K, Sato H, Inoue K, Tsunoda T, Sakata Y, Mizuno H, Lin TH, Miyamoto Y, Aoki A, Onouchi Y, Sheu SH, Ikegawa S, Odashiro K, Nobuyoshi M, Juo SH, Hori M, Nakamura Y, Tanaka T. Snps in *BRAP* associated with risk of myocardial infarction in asian populations. *Nat Genet.* 2009;41:329-333