

論文 / 著書情報
Article / Book Information

題目(和文)	大腸がん外科的治療における腸内環境の変化の解析
Title(English)	
著者(和文)	城間博紹
Author(English)	Hirotsugu Shiroma
出典(和文)	学位:博士(理学), 学位授与機関:東京工業大学, 報告番号:甲第12269号, 授与年月日:2022年6月30日, 学位の種別:課程博士, 審査員:山田 拓司,伊藤 武彦,北尾 彰朗,本郷 裕一,平沢 敬
Citation(English)	Degree:Doctor (Science), Conferring organization: Tokyo Institute of Technology, Report number:甲第12269号, Conferred date:2022/6/30, Degree Type:Course doctor, Examiner:,,,,
学位種別(和文)	博士論文
Type(English)	Doctoral Thesis

博士論文

令和 4 年

大腸がん外科的治療における 腸内環境の変化の解析

東京工業大学 生命理工学院
生命理工学系 生命理工学コース

城間 博紹

指導教員 山田 拓司

目次

第1章	序論	7
1.1	背景	7
1.1.1	ヒト腸内細菌のメタゲノム解析	7
1.1.2	大腸がん和腸内細菌の関係性	8
1.1.3	術後の大腸がん和腸内環境の関係性	9
1.2	目的	10
1.3	構成	10
第2章	解析に用いたデータの取得	12
2.1	序論	12
2.2	材料・手法	13
2.2.1	便の取得	13
2.2.2	内視鏡検査	14
2.2.3	大腸がんの治療	16
2.2.4	代謝産物の定量	16
2.2.5	DNAの抽出	17
2.2.6	シーケンシング	18
2.2.7	公開されているメタゲノムリード和代謝産物データの取得	18
2.2.8	被験者の疫学データの取得	19
2.3	結果・考察	19
2.3.1	内視鏡検査の結果	19
2.3.2	代謝産物・メタゲノムリードの取得	22
第3章	ヒト腸内細菌のメタゲノム解析	23
3.1	序論	23
3.2	材料・手法	24
3.2.1	高精度なリードの取得	24
3.2.2	腸内細菌の遺伝子組成和遺伝子機能組成の算出	25
3.2.3	腸内細菌の系統組成の算出	31
3.2.4	腸内細菌の生育速度の推定	32
3.2.5	<i>Bai</i> オペロンの定量	37
3.2.6	遺伝子カタログの評価	39
3.2.7	公開されているメタゲノムリードを用いたメタゲノム解析	40
3.3	結果・考察	41
3.3.1	高精度なリードの取得	41
3.3.2	再構築したゲノムのクオリティの評価	42
3.3.3	遺伝子カタログの構築とその評価	43
3.3.4	遺伝子機能組成の算出	45
3.3.5	生育速度の推定	46

3.3.6	デオキシコール酸生成菌の探索	47
第4章	外科的治療前後における腸内環境の全体像比較	50
4.1	序論	50
4.2	手法	50
4.2.1	組成データに対する多次元圧縮法の適用	50
4.2.2	組成データの変化の統計学的な比較方法	51
4.2.3	ヒトゲノムの混入量の統計学的な比較方法	52
4.3	結果・考察	53
4.3.1	大腸がん患者の外科的治療前後の組成データの変化	53
4.3.2	大腸がん患者の外科的治療前後のヒトゲノムの混入量の変化	56
第5章	外科的治療前後における腸内環境を構成する特徴量の変動比較	58
5.1	序論	58
5.2	手法	59
5.2.1	大腸がん外科的治療前後で変化する特徴量の特定	59
5.2.2	大腸がんに関連する特徴量の特定	60
5.3	結果・考察	61
5.3.1	腸内細菌の系統組成の比較	61
5.3.2	腸内細菌の生育速度の比較	65
5.3.3	腸内細菌の代謝産物の比較	67
5.3.4	腸内細菌の遺伝子の比較	71
第6章	大腸の再建法と関連する腸内環境の特徴量	73
6.1	序論	73
6.2	手法	74
6.3	結果・考察	75
6.3.1	右側大腸がんの再建法が胆汁酸代謝へ及ぼす影響	75
6.3.2	左側大腸がんの再建法が胆汁酸代謝へ及ぼす影響	81
6.3.3	大腸の再建法の違いが胆汁酸代謝へ及ぼす影響	82
第7章	外科的治療後の大腸がんのリスク推定手法の確立	84
7.1	序論	84
7.2	手法	85
7.2.1	判別器の構築	85
7.2.2	判別器への腸内環境データの適用	90
7.2.3	確率の補正	91
7.2.4	統計処理	92
7.3	結果・考察	92
7.3.1	大腸がんのリスクを推定する手法の確立	92
7.3.2	大腸がんのリスクを反映する可能性がある特徴量の探索	94
第8章	総括	97

謝辭.....	99
参考文献.....	100
付録	108

目次

図 2-1 本研究に用いたデータの取得方法の概略図	12
図 2-2 本研究と先行研究で得られた便の数	14
図 2-3 術後の患者の大腸がんのリスクの定量化	15
図 2-4 大腸がん外科的治療による大腸の再建法	16
図 3-1 メタゲノム解析の概略図	24
図 3-2 生育速度の推定手法	34
図 3-3 生育速度の推定手順の概略図	36
図 3-4 <i>Bai</i> オペロンの定量手順の概略図	38
図 3-5 高精度なリードを取得する過程におけるリード数の推移	42
図 3-6 再構築した腸内細菌のゲノムのクオリティ	43
図 3-7 遺伝子カタログの評価	45
図 3-8 デオキシコール酸生成菌の <i>bai</i> オペロンを構成する遺伝子の構造	48
図 4-1 組成データにおける大腸がん患者の外科治療前後の類似性を表した散布図	54
図 4-2 大腸がん外科的治療前後の患者の Bray-Curtis 距離を表した箱ひげ図	55
図 4-3 大腸がん外科的治療前後の患者のヒトゲノムの混入量を示した箱ひげ図	57
図 5-1 大腸がん外科的治療前後の患者の胆汁酸代謝の変化の概略図	59
図 5-2 大腸がん患者の外科的治療前後の腸内細菌の相対存在量を比較した結果	63
図 5-3 大腸がん外科的治療前後の患者の腸内細菌の相対存在量を示した箱ひげ図	64
図 5-4 大腸がん外科的治療前後の患者の腸内細菌の生育速度を示した箱ひげ図	66
図 5-5 大腸がん外科的治療前後の患者の代謝産物の量を比較した結果	69
図 5-6 大腸がん外科的治療前後の患者の代謝産物の箱ひげ図	70
図 5-7 胆汁酸代謝に関係している遺伝子とオペロンの箱ひげ図とそれらの寄与率	72
図 6-1 右側大腸がん患者と左側大腸がん患者の胆汁酸代謝の変化の概略図	74
図 6-2 右側大腸がん患者と左側大腸がん患者の血中の総コレステロールの箱ひげ図	78
図 6-3 右側大腸がん患者と左側大腸がん患者の代謝産物の箱ひげ図	79
図 6-4 右側大腸がん患者と左側大腸がん患者の腸内細菌やその遺伝子の箱ひげ図	80
図 7-1 外科的治療後の大腸がんのリスク推定手法の評価の概略図	85
図 7-2 判別器を構築するための特徴量選択の概略図	90
図 7-3 大腸がん患者の外科的治療前後の補正確率の箱ひげ図	94
図 7-4 判別器の構築に寄与した上位 10 個の特徴量を表したヒートマップ	96

表目次

表 2-1 内視鏡検査を受けた 620 人の被験者の疫学データの概要	20
表 2-2 大腸がん患者の治療前後の疫学データの概要	21
表 3-1 腸内細菌の生育速度を推定した統計値	46
表 3-2 再構築した <i>Clostridium scindens</i> と <i>Clostridium hiranonis</i> のゲノムの情報	49
表 7-1 判別器を構築に用いた特徴量の数とパラメータ	93

付録目次

付録 1 公開されているデータの Accession number	108
付録 2 公開されている <i>bai</i> 遺伝子の UniProt の Accession number	109
付録 3 大腸がん外科的治療後に減少した腸内細菌	110
付録 4 大腸がん外科的治療後に増加した腸内細菌	112
付録 5 大腸がん外科的治療後に減少した代謝産物	113
付録 6 大腸がん外科的治療後に増加した代謝産物	114

第1章 序論

1.1 背景

1.1.1 ヒト腸内細菌のメタゲノム解析

ヒトの腸管内には多くの細菌が生息しており、それらが複雑に相互作用することによって腸内細菌叢を形成している[1]。腸内細菌の多くは大腸に生息しており、酢酸や酪酸等の短鎖脂肪酸や、コール酸(Cholate)やデオキシコール酸(Deoxycholic acid; DCA)等の胆汁酸を生成することで、宿主であるヒトの免疫の制御や病原性微生物に対する防御を担っている[2]。腸内細菌やそれらが生成する代謝産物は、大腸がんや炎症性腸疾患等の様々な疾患と関与していることが報告されている[3,4]。

腸内細菌の多くは、その生態が明らかにされていないため、未だ単離・培養することが困難である[5,6]。そのため、腸内細菌の機能や特徴、さらには、腸内細菌叢を明らかにすることも困難であった。これらを明らかにするために、特定の環境由来の断片化したDNAをシーケンシングする whole-genome shotgun シーケンシングによって、単離・培養を介さずにその環境中の塩基配列(メタゲノムリード)を解読するメタゲノム解析の手法が2000年に確立した[7]。メタゲノム解析は、その環境に存在する腸内細菌の系統組成だけではなく、腸内細菌の遺伝子組成を明らかにすることが可能である。腸内細菌の系統組成の算出には、系統マーカー遺伝子の塩基配列に対してメタゲノムリードをマッピングする手法がよく用いられている[8,9]。腸内細菌の遺伝子組成の算出には、ヒト腸内細菌の遺伝子カタログである IGC(Integrated Gene Catalog)[10]に収録されている遺伝子の塩基配列に対してメタゲノムリードをマッピングする手法がよく用いられている[9]。参照配列に対してメタゲノムリードをマッピングすることで腸内細菌やその遺伝子を定量する手法は、メタゲノムサンプルに存在し、参照配列に存在しない腸内細菌やその遺伝子を検出することができない。そのため、参照配列となる腸内細菌の遺伝子やゲノムを収録してカタログ化することが重要であると考えられる。

近年、whole-genome shotgun シーケンシングによって得られたメタゲノムリードから、その環境に存在する細菌のゲノムを再構築するビニングの技術が発達した[11-13]。その技術の発達により、既に公開されているメタゲノムリードから膨大な数のヒト腸内細菌のゲノムが再構築された[14-17]。それにより、未だ単離・培養することができていないヒト腸内細菌の機能や特徴を推定することが可能になりつつある。

1.1.2 大腸がん和腸内細菌の関係性

大腸がんは、がん抑制遺伝子やがん遺伝子に変異が蓄積することにより発症・進行する疾患である[18–20]。大腸がんの発症・進行に関する遺伝子の変異のメカニズムとして、多段階発がん説が提唱されている。多段階発がん説では、がん抑制遺伝子である

APC(*Adenomatous polyposis coli*)に変異が生じることで大腸がんの前がん病変である腺腫が発生した後に、複数のがん抑制遺伝子やがん遺伝子に変異が蓄積することで大腸がんが発症すると考えられている[18,19]。先行研究において、大腸がんの多くは、多段階発がんを経て発症することが報告されている[20]。

大腸がんの発症に関与しているリスク因子を長年調査した結果、先天性の遺伝子変異ではなく環境因子が、多くの大腸がんの発症に関与していることが明らかになった[21]。その環境因子として、加齢、赤肉の摂取等の食生活の欧米化や喫煙やアルコールの摂取等の生活習慣の乱れが挙げられる[22,23]。腸内細菌はこれらの環境因子と関与していることが報告されている[24]。

近年、腸内細菌や腸内細菌が生成する代謝産物が大腸がんの発症・進行に関与しているという報告が多数発表されている。しかしながら、大腸がんの発症・進行に伴って腸内環境も変化するためその関係性は複雑である[25]。まず、大腸がん患者における腸内細菌や代謝産物の変化が大腸がんに関与する例を挙げる。先行研究において、*Fusobacterium nucleatum*の相対存在量や推定された生育速度は大腸がんの進行に従って高くなることが報告されている[4]。また、生検検体を用いた同一患者における大腸がん組織と粘膜組織に存在する代謝産物の量を比較した結果、大腸がん組織でセリン(Serine; Ser)等の一部のアミノ酸の量が多いことも報告されている[26,27]。これらの腸内細菌や代謝産物は、大腸がんの存在やそれに伴う腸内の環境(炎症や圧搾等に起因する環境)を反映しており、その環境が特定の腸内細菌や代謝産物のさらなる濃縮を引き起こしている可能性が考えられる。次に、腸内細菌が生成する代謝産物が大腸がんの発症に関連している例を挙げる。先行研究において、*Bilophila wadsworthia*や*Desulfovibrio*が生成する硫化水素は、宿主であるヒトのDNAにダメージを与えることが報告されている[28–30]。また、*Clostridium scindens*や*Clostridium hiranonis*が生成するデオキシコール酸は発がん性を有することが報告されており[31]、大腸がんの発症に関与している可能性がある[32]。そのため、腸内細菌や代謝産物と大腸がんの発症・進行との関係性を明らかにすることは、腸内細菌に由来する大腸がんのリスクを減少させるためにも重要であると考えられる。

大腸がん和腸内細菌の関連性が明らかになってきた結果、大腸がんを診断するツールとして腸内細菌を利用できる可能性も報告されている。先行研究において、5カ国で実施

されたメタゲノム解析を用いた大腸がんのコホート研究を対象としたメタアナリシスにより、コホート間で共通した大腸がんのバイオマーカーになり得る腸内細菌が同定された[8,9]。これらの腸内細菌は、便の潜血検査のように、大腸がんを検出するためのスクリーニングに用いられる可能性があるが、大腸がんの前がん病変である腺腫や早期の大腸がんである Stage 0 の大腸がん等、大腸がんのリスクが高い状態を検出することができるのかについては明らかにされていない。

大腸がんの前がん病変である腺腫や早期の大腸がんである Stage 0 の大腸がんを検出する手法は、大腸がんのリスクを長期的に減少させるための介入する機会になり得る[33]。大腸がんが進行する前に、腺腫や Stage 0 の大腸がんを検出することができた場合、侵襲的な外科的切除ではなく、内視鏡を用いた非侵襲的な方法で大腸がんを取り除くことができる[34]。そのため、これらを検出するための大腸がんのリスクを推定する手法を確立することは、患者の負担を減らすためにも重要であると考えられる。

1.1.3 術後の大腸がんと腸内環境の関係性

大腸がんの外科的治療は、大腸がん患者の生存率を高めるための根治治療の選択肢の1つである[35]。日本では、大腸がんの罹患数の増加に伴って大腸がんの外科的手術の実施数も増加傾向にある[36]。そのため、大腸がん外科的治療後の累計患者数も増加傾向にあることが予想される。

大腸がん外科的治療後の患者は、大腸がんの局所再発や、残存している大腸に原発の大腸がんとは異なる大腸がんである異時性大腸がんを発症するリスクが高いことが報告されている[37-40]。そのため、大腸がん患者の外科的治療後の大腸がんのリスクを定量的に評価することが重要であると考えられる。

大腸がんの発症・進行における腸内環境の役割が明らかになりつつあるにも関わらず、大腸がん外科的治療前後の患者を対象とした腸内環境の変化に関する先行研究は少ない[41,42]。Ohigashi et al.の研究では、81 人の大腸がん患者を対象に、外科的治療前と治療から 7 日後に便を取得し、便に含まれているいくつかの腸内細菌や代謝産物を、定量的 PCR や高速液体クロマトグラフィーを用いてそれぞれ定量したが、外科的治療前後間の間隔が短いため、腸内環境に対する外科的治療の長期的な影響は明らかにされていない[41]。Sze et al.の研究では、26 人の大腸がん患者の外科的治療前と治療からの約 1 年後の便を取得し、便に含まれている腸内細菌叢を解析したが、腸内細菌の遺伝子や代謝産物を対象にしていないため、それらに対する外科的治療の影響は明らかにされていない[42]。

1.2 目的

大腸がん外科的治療後の患者は治療により大腸がんの病変が切除されるが、大腸がんのリスクが高いことが報告されている。また、いくつかの腸内細菌やその代謝産物は大腸がんの発症への関与していることが報告されているが、腸内細菌や代謝産物に対する外科的治療の影響は明らかにされていない。

本研究は、メタゲノム解析や、代謝産物を網羅的に測定するメタボローム解析によって、外科的治療後の大腸がんに関与している腸内細菌や代謝産物を特定することを目的としている。大腸がん患者の外科的治療前と治療から約 1 年後の便を取得し、その便に含まれている腸内細菌や代謝産物を統計学的に比較することにより外科的治療が腸内環境へ及ぼす影響を明らかにすること、さらには、外科的治療後の大腸がんのリスクを推定する手法を確立することを目指した。

1.3 構成

本研究は、8 つの章で構成されている。本節では、本章を除くそれぞれの章の概要について述べる。

第 2 章 解析に用いたデータの取得

本研究は、国立がん研究センターと慶應義塾大学との共同研究により実施した。第 2 章では、共同研究先から提供して頂いた、内視鏡検査の結果や便中の代謝産物の定量データ、whole-genome shotgun シーケンシングによって得られるメタゲノムリードの取得方法について述べる。また、外科的治療から 5 年以内に、大腸がんや大腸がんの前がん病変である多発性腺腫を発症したかどうか前向きコホートを実施することで、大腸がん患者の外科的治療後の大腸がんのリスクを定量化できたため、その詳細な結果についても述べる。

第 3 章 ヒト腸内細菌のメタゲノム解析

第 3 章では、第 2 章で取得したメタゲノムリードを用いてメタゲノム解析により、腸内細菌の系統・遺伝子組成を取得する手法について述べていく。特に、遺伝子の位置関係を考慮

して、目的の遺伝子を定量することができる新たな遺伝子組成の算出手法を構築したため、その詳細な手法について述べる。

第4章 外科的治療前後における腸内環境の全体像比較

第4章では、第2章と第3章で取得した代謝産物や腸内細菌の系統・遺伝子組成データを用いて、外科的治療前後の腸内環境の全体像の変化を調査した手法や結果について述べる。

第5章 外科的治療前後における腸内環境を構成する特徴量の変動比較

第5章では、外科的治療前後で変化した腸内細菌や遺伝子・代謝産物を統計学的に比較する手法とその結果について述べる。

第6章 大腸の再建法と関連する腸内環境の特徴量

第5章にて、胆汁酸の1種であり、発がん性が報告されているデオキシコール酸やその生成に関与している遺伝子や生成菌が外科的治療後に増加したことが確認された。右側大腸がん患者の大腸の再建法は胆汁酸代謝に関連していることが報告されていたため、第6章では、大腸の再建法の違いが胆汁酸代謝へ及ぼす影響を調査し、その結果について述べる。

第7章 外科的治療後の大腸がんのリスク推定手法の確立

第2章にて、大腸がん患者の外科的治療後の大腸がんのリスクを定量化することができた。第7章では、外科的治療後の大腸がんのリスクを推定する手法を構築し、その手法で推定された大腸がんのリスクと第2章にて定量化された大腸がんのリスクを照合することで、推定手法を評価した結果について述べる。

第8章 総括

第8章では、以上の結果を統合して総括し、今後の展望を述べる。

第2章 解析に用いたデータの取得

2.1 序論

本章では、本研究の解析に用いたデータの取得方法について述べる。本研究は、国立がん研究センターと慶應義塾大学との共同研究により実施された。それぞれの研究機関において実施された項目の内訳は図 2-1 に示した。本研究では、国立がん研究センターにて、96 人の大腸がん患者の内視鏡治療または外科的治療前と治療から約 1 年後の便を取得し、外科的治療後の大腸がんに関与している腸内細菌や代謝産物を特定することを目指した(2.2.1 節、2.2.2 節、2.2.3 節)。代謝産物の組成データは、397 個の代謝産物を CE-TOF MS(Capillary Electrophoresis Time-of-Flight Mass Spectrometry)を用いて定量することで取得した(2.2.4 節)。メタゲノムサンプルのメタゲノムリードは、便から抽出された DNA を HiSeq 2500 を用いて whole-genome shotgun シーケンシングすることで取得した(2.2.5 節、2.2.6 節)。本研究では、1 サンプルあたり約 4.71×10^7 本のメタゲノムリード(中央値)を取得しており、whole-genome shotgun シーケンシングを用いたいくつかの大腸がんのコホート研究[8,9,43–46]で取得されたメタゲノムリード数と比較しても遜色がないことが確認された。

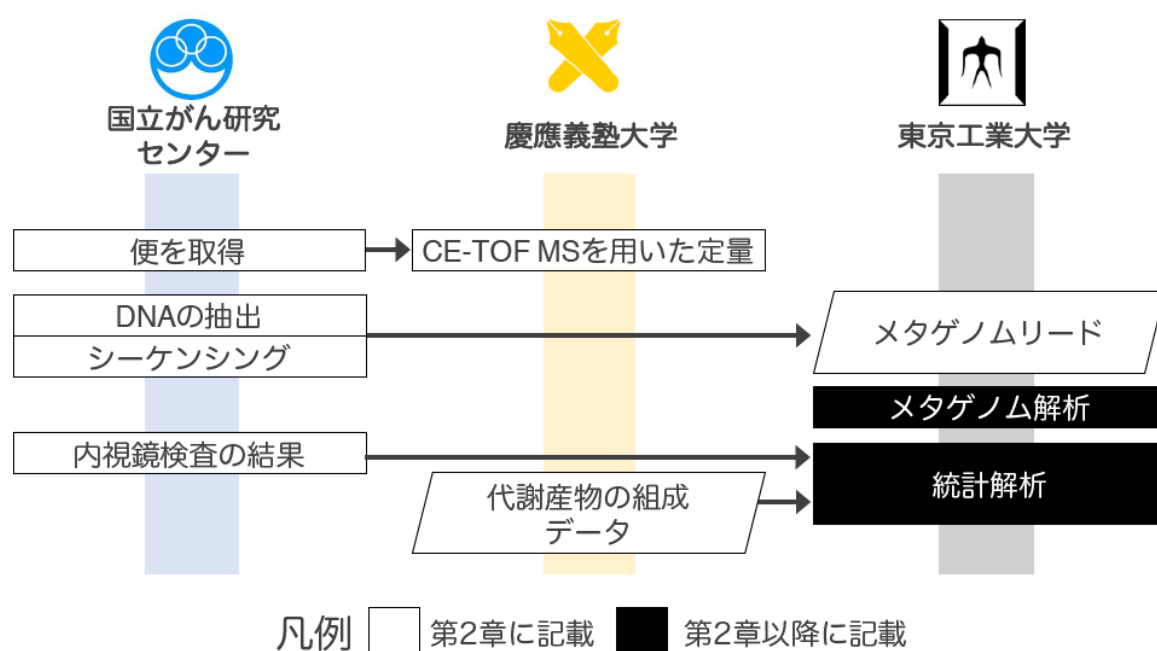


図 2-1 本研究に用いたデータの取得方法の概略図

国立がん研究センターで、内視鏡検査を受ける被験者から便を取得した。慶應義塾大学にて、便に含まれている 397 個の代謝産物を CE-TOF MS を用いて定量した。メタゲノムリードは、便から抽出された DNA を whole-genome shotgun シーケンシングすることで取得した。第 3 章で、メタゲノム解析の詳細な手法について述べる。第 4 章以降で、内視鏡検査の結果、代謝産物の組成データ、メタゲノム解析によって得られた腸内細菌叢の系統組成データや遺伝子組成データを統合した統計解析の手法や結果について述べる。

2.2 材料・手法

2.2.1 便の取得

本節の操作は共同研究先である国立がん研究センターで実施された。国立がん研究センターで大腸の内視鏡検査を受ける 620 人の被験者から 716 個の便を、先行研究[4]と本研究にて取得した。

まず、先行研究と本研究で対象とした被験者数とその被験者から得られた便の数についてそれぞれ述べる(図 2-2)。先行研究では、576 人の大腸がんの治療歴がない被験者から 604 個の便を取得した。先行研究では、576 人の被験者を内視鏡検査の結果を基に 5 つのグループ(グループの詳細は 2.2.2 節で後述する)に分類し、多段階発がんにおける腸内細菌や代謝産物の変化を明らかにした。さらに、258 人の大腸がん患者のうち 28 人の患者から外科的治療から約 1 年後に便を取得することで、小規模なデータセットではあるが、外科的治療が腸内細菌へ及ぼす影響を明らかにした。本研究では、68 人の大腸がん患者から 112 個の便を新たに取得することで、外科的治療が腸内細菌だけでなく代謝産物へ及ぼす影響をより詳細に明らかにすることを目指した。本研究では、先行研究で対象とした 258 人の大腸がん患者のうち、24 人の患者の治療から約 1 年後の便を新たに取得した。また、44 人の大腸がん患者の治療前と治療から約 1 年後の便も新たに取得した。

次に、被験者の便の取得方法について述べる。620 人の被験者は、腸管洗浄液を飲用後、最初の便を提供した。詳細は 2.2.2 節で後述するが、内視鏡検査の結果、大腸がんと診断された 302 人の患者のうち 96 人の患者も、治療から約 1 年後に腸管洗浄液を飲用後、最初の便を提供した。提供して頂いた便はすぐにドライアイスで凍結し、その後-80℃にて保管した。

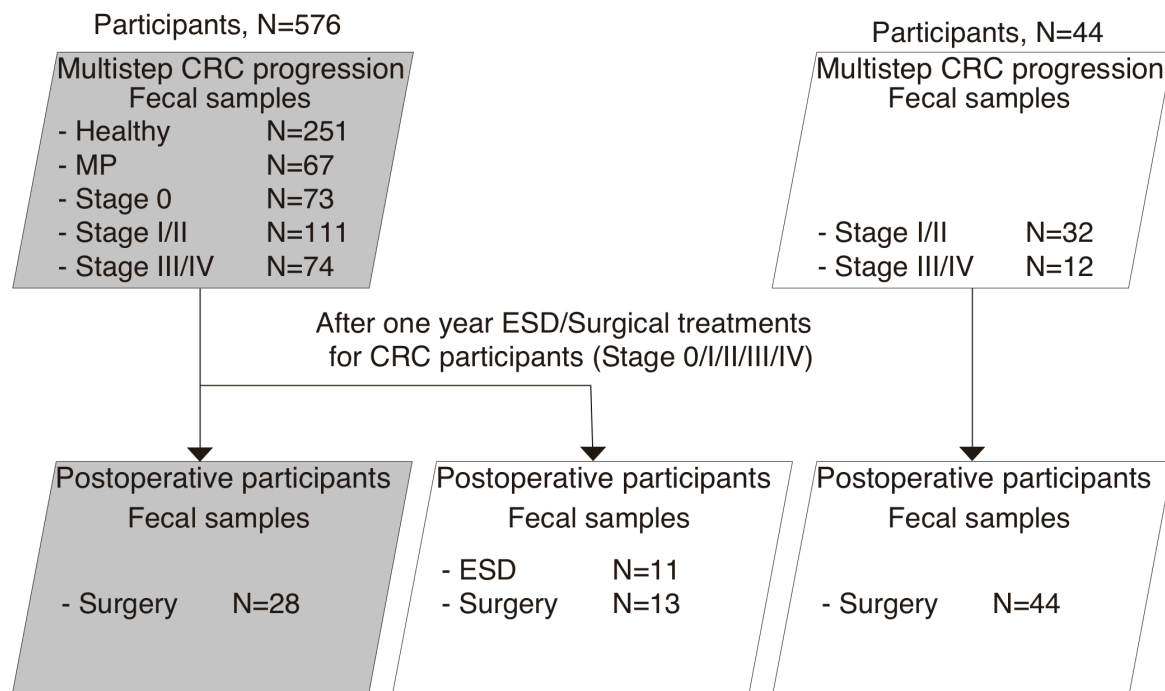


図 2-2 本研究と先行研究で得られた便の数

先行研究で得られた便は灰色で示し、本研究で新たに取得した便は白色で示した。

2.2.2 内視鏡検査

本節の操作は共同研究先である国立がん研究センターで実施された。便を提供した被験者は内視鏡検査を受けた。その結果、TNM(Tumor-Nodes-Metastasis)分類に従って、620 人の大腸がんの治療歴のない被験者を以下の 5 つのグループに分類した。腺腫が検出されなかった被験者、もしくは腺腫が検出されたが、その数が 2 個以下であった被験者は、健常者(Healthy controls; H)として定義した。腺腫が 3 個以上検出された患者は、多発性腺腫の患者(Multiple polypoid adenomas with low-grade dysplasia; MP)として定義した。大腸の粘膜内に留まっている大腸がんが検出された患者は Stage 0 の大腸がん患者(粘膜内の大腸がん患者)として定義した。粘膜より下層に大腸がんが検出されたが、大腸がんがリンパ節へ転移していない患者は、Stage I/II の大腸がん患者として定義した。粘膜より下層に大腸がんが検出され、大腸がんがリンパ節へ転移している患者は、Stage III/IV の大腸がん患者(進行性大腸がん患者)として定義した。

96 人の患者は、治療から約 1 年後に内視鏡検査を受け、その結果を基に、3 つのグループに分類された(図 2-3)。腺腫が検出されなかった患者、もしくは腺腫が検出されたが、その数が 2 個以下であった患者を、治療歴のある健常者(Postoperative Healthy

controls; PH)として定義した。治療後に、腺腫が3個以上検出された患者は、治療歴のある多発性腺腫の患者(Postoperative multiple polypoid adenomas with low-grade dysplasia; MP)として定義した。治療後に大腸の粘膜内に留まっている大腸がんが検出された患者は、治療歴のある Stage 0 の大腸がん患者(異時性大腸がん患者, Postoperative Stage 0; PS0)として定義した。

治療後の患者の大腸がんのリスクを定量化するために前向きコホートを実施した(図 2-3)。96 人の大腸がん患者のうち、78 人の患者は、治療から約 5 年間、国立がん研究センターにて定期的に内視鏡検査を受けており、治療から約 1 年後から約 5 年後の 4 年間の内視鏡検査の結果を基に、治療歴のある健常者(PH)、治療歴のある多発性腺腫の患者(PMP)に分類された。最終的に、治療から約 5 年間の内視鏡検査の結果、一度でも多発性腺腫(PMP)または大腸がん(PS0)と診断されたかどうかを基準に、患者を大腸がんのリスクが低い患者と高い患者に分類した。

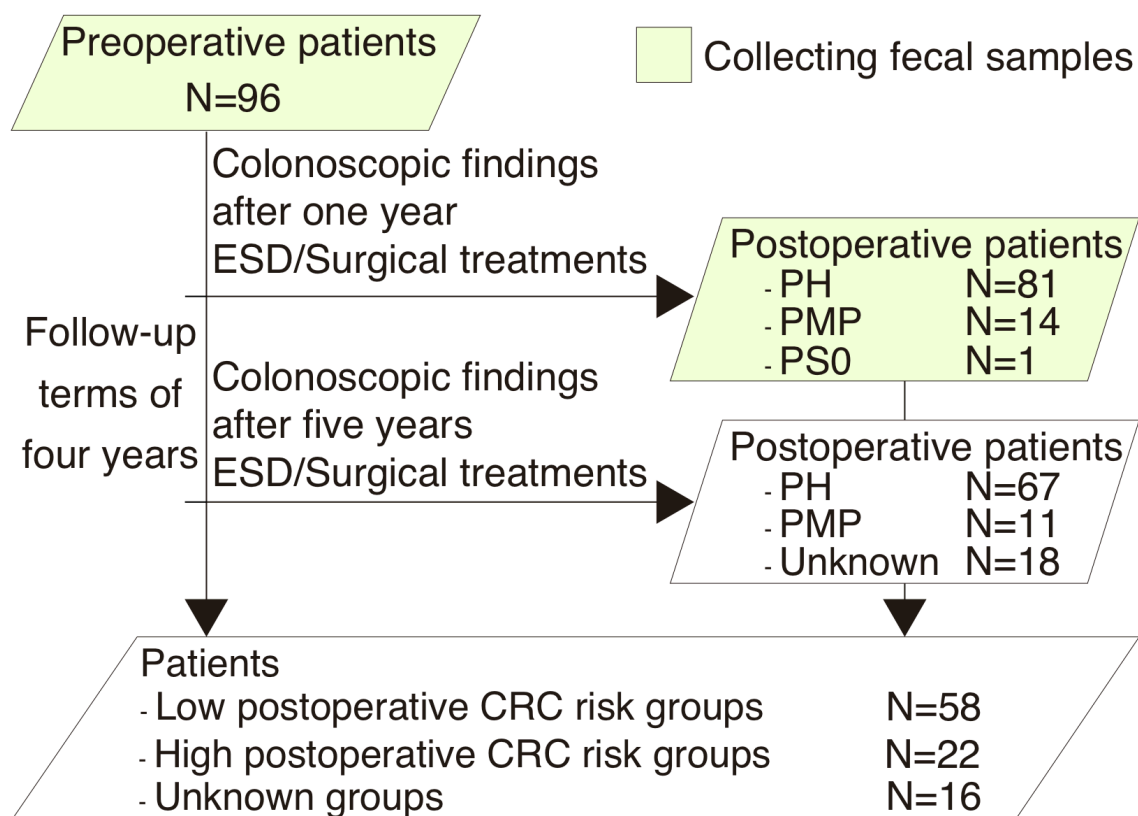


図 2-3 術後の患者の大腸がんのリスクの定量化

2.2.3 大腸がんの治療

本節の操作は共同研究先である国立がん研究センターで実施された。96 人の大腸がん患者は治療によって大腸がんの病変を切除した。そのうち、11 名の Stage 0 の大腸がん患者は内視鏡治療により大腸がんの病変を切除した。85 人の大腸がん患者は、臨床学的に内視鏡治療ができなかったため、外科的治療により大腸がんの病変を切除した。外科的治療では、手術による病原性細菌の感染を予防するために、手術前と手術後 2 日以内に抗生物質を投与した。

外科的治療による大腸の再建法は、大腸がんの病変が存在する場所によって異なった。その詳細な違いを図 2-4 に示した。大腸がんが大腸の右側に存在する場合、血流の関係上、大腸の右側をほとんど切除して大腸を再建した。大腸がんが大腸の左側(直腸・結腸)に存在する場合、大腸がんの病変のみを切除して大腸を再建した。

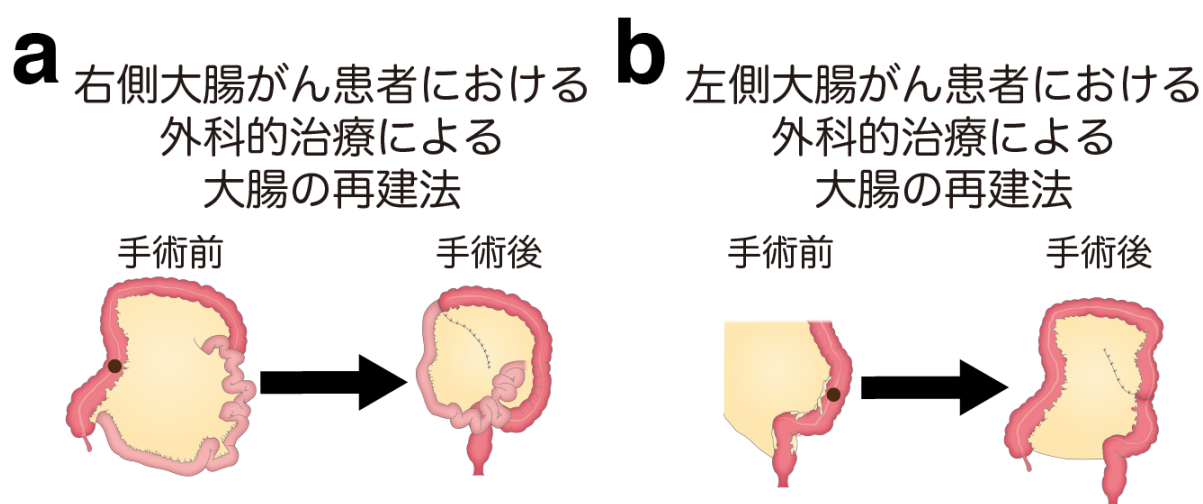


図 2-4 大腸がん外科的治療による大腸の再建法

薄いピンク色は小腸、濃いピンク色は大腸を表している。黒い丸は大腸がんの位置を示している。

2.2.4 代謝産物の定量

本節の操作は共同研究先である慶應義塾大学で実施された。CE-TOF MS を用いて便に含まれている代謝産物を定量した[47–49]。CE-TOF MS は、キャピラリー電気泳動(Capillary Electrophoresis)と質量分析計(Mass Spectrometry)を組み合わせた分析手法である。キャピラリー電気泳動は、代謝産物の電荷や分子量等の違いを利用して、便に含まれている代謝

産物を分離する役割を担っている。質量分析計は、分離したそれぞれの代謝産物を定量する役割を担っている。以下の 6 つの手順で代謝産物を定量した。

1. 内部標準物質を含む超純水に 50%のメタノール 400 μ l を加えた。内部標準物質は methionine sulfone と D-camphor-10-sulfonic acid 及び 2-morpholinoethanesulfonic acid を用いた。
2. 便を 10mg 採取し、1 に加えた後、3mm ジルコニアビーズ(BioSpec Products, Bartlesville, OK, USA)と 0.1mm ジルコニア/シリカビーズ(BioSpec Products, Bartlesville, OK, USA)を 100mg 加えてマイクロスマッシュ(TOMU, Nerima, Tokyo, Japan)を用いて 3 分間破碎した。
3. 破碎後、懸濁液を 4600xg で 15 分間遠心した。
4. 遠心後、5-kDa の限界濾過フィルター(Ultrafree MC-PLHCC 250/pk, Metabolome Analysis, Human Metabolome Technologies, Tsuruoka, Yamagata, Japan)を用いて上清のタンパク質の除去し、濾液を超純水に溶解した。
5. CE-TOF MS(Agilent Technologies, Santa Clara, CA, USA)の陽イオンと陰イオンモード用いて、4 に含まれている代謝産物を定量した。
6. CE-TOFMS によって得られた生データから、解析ソフトウェアである Master Hands を用いて、代謝産物を定量した。各代謝産物の濃度(nmole)は、それらの相対ピーク面積と標準代謝産物の濃度に基づいて計算し、便重量を用いて正規化して、サンプルの 1 グラムに含まれる代謝産物の量(nmol/g)を算出した。

2.2.5 DNA の抽出

本節の操作は外部機関で実施された。凍結した便を融解後、GNOME® DNA Isolation Kit(MP Biomedicals, Santa Ana, California, USA)を用いて DNA を抽出した。DNA の吸光度(230nm, 260nm, 280nm)は、4200 TapeStation(Agilent technology, Santa Clara, CA)を用いて測定した。DNA の精度および濃度は、測定した吸光度の比 (260nm/280nm, 260nm/230nm)をそれぞれ計算することで算出した。極端に DNA の精度や濃度が低いサンプルは含まれていなかったため、全てのサンプルを-80℃にて保管した。

2.2.6 シーケンシング

本節の操作は外部機関で実施された。凍結した DNA は融解後、Nextra XT DNA Library Preparation Index Kit(Illumina)および、Nextra XT DNA Library Preparation kit(Illumina)を用いてタグメンテーションされた。4200 TapeStation(Aglient technology, Santa Clara, CA)を用いてタグメンテーション後の DNA のピークサイズを評価した。タグメンテーション後の DNA を、シーケンサー(HiSeq 2500, Illumina)を用いて 150 bp のペアエンドシーケンシングすることで、メタゲノムリードを取得した。

2.2.7 公開されているメタゲノムリードと代謝産物データの取得

本研究では健常者の時系列データを取得していないため、2つのコホート研究由来の公開されている健常者の時系列データをダウンロードした[50,51](付録 1)。この公開データは、外科的治療前後間での腸内細菌やその遺伝子、代謝産物の組成の変化を定量化するためのコントロールに用いた。詳細な手法・結果については第 4 章で述べる。

メタゲノムリードは、Voigt et al.のコホートのデータを ENA(European Nucleotide Archive)からダウンロードした(PRJEB8347) [50]。Voigt et al.の DNA 抽出方法(G'NOMEs kit, MP Biomedicals, Illkirch, France)及びシーケンシングプラットフォーム(HiSeq 2000/2500, Illumina)は本研究のそれらとほぼ同じであるため、本研究のデータと比較する際にこれらの手法によるバイアスが小さい可能性がある。Voigt et al.のコホートのデータに含まれている、抗生物質を使用した 1 人の参加者(alien)に由来するメタゲノムリードと、本研究のメタゲノム解析の手法で解析することができなかった 1 つのメタゲノムサンプル(bugkiller-11-7-0)を、本研究の解析対象から取り除いた。Voigt et al.のコホートの 6 人由来の 29 サンプルのメタゲノムリードを本節以降の解析に用いた。

代謝産物の組成データは、Nagata et al.のコホートのデータを用いた[51]。Nagata et al.の代謝産物の定量方法も、本研究と同じ手法(2.2.4 節を参照)であるため、本研究のデータと比較する際にこれらの手法によるバイアスが小さい可能性がある。Nagata et al.の代謝産物の組成データは、その定量を担当した共同研究先から提供して頂いた。Nagata et al.の 8 人由来の 16 サンプルの代謝産物の組成データを本節以降の解析に用いた。

2.2.8 被験者の疫学データの取得

本節の操作は共同研究先である国立がん研究センターで実施された。内視鏡検査を受けた 620 人の被験者の年齢、性別、BMI(Body Mass Index, Kg/m²)等の疫学データを取得した。大腸がんの治療前と治療後に便を提供した 96 人の患者の治療後のこれらのデータは取得することができなかった。第 6 章にて詳細に後述するが、外科的治療後に便中の総胆汁酸(定量できた 4 つの胆汁酸の量の総和)が増加したメカニズムを明らかにするために、大腸がんの治療前と治療後に便を提供した 96 人の治療前後の血中の総コレステロールを定量した。

2.3 結果・考察

2.3.1 内視鏡検査の結果

先行研究[4]と本研究では、国立がん研究センターで内視鏡検査を受ける 620 人の被験者から、腸管洗浄液飲用後の最初の便を 716 個取得した。620 人の被験者は、内視鏡検査の結果、251 人の健常者、67 人の多発性腺腫の患者、73 人の Stage 0 の大腸がん患者、143 人の Stage I/II の大腸がん患者、86 人の Stage III/IV の大腸がん患者に分類した(表 2-1)。先行研究で対象とした患者やその患者の便も含めて、本研究では、内視鏡検査の結果大腸がんと診断された 302 人の患者のうち、96 人の大腸がん患者から治療前と治療から約 1 年後に便を取得した。大腸がんの大きさや進行度によって治療方法が異なった(表 2-2)。Stage 0 の大腸がん患者 11 人は、大腸がんが小さかったため内視鏡治療により大腸がんの病変を切除した。Stage 0 の大腸がん患者 2 人は、臨床学的に内視鏡治療が難しかったため、外科的治療により大腸がんの病変を切除した。57 人の Stage I/II の大腸がん患者と 26 人の Stage III/IV の大腸がん患者も臨床学的に内視鏡治療が難しかったため、外科的治療によって大腸がんの病変を切除した。

85 人の大腸がん外科的治療後の患者は、治療から約 1 年後の内視鏡検査の結果を基に、73 人の治療歴のある健常者(PH)、11 人の治療歴のある多発性腺腫の患者(PMP)、1 人の治療歴のある Stage 0 の大腸がん患者(異時性大腸がん患者, PS0)に分類された(表 2-2)。また、85 人の大腸がん外科的治療後の患者のうち、76 人の患者は国立がん研究センターにて定期的に内視鏡検査を受けており、治療から 1 年から約 5 年までの 4 年間の内視鏡検査の結果に基づいて、65 人の治療歴のある健常者(PH)、11 人の治療歴のある多発性腺腫の患者(PMP)にも分類された。最終的に、治療から約 5 年間の内視鏡検査の結果、一度でも多発

性腺腫または大腸がんと診断されたかどうかを基準に 76 人の患者を、57 人の大腸がんのリスクが低い患者と 19 人の大腸がんのリスクが高い患者に分類した。治療から約 5 年間で大腸がんを局所再発した患者はいなかった。

外科的治療では、大腸がんが存在する場所によって切除する大腸の領域が異なった。20 人の右側大腸がん患者は大腸の右側を切除した(表 2-3)。65 人の左側大腸がん患者(直腸・結腸がん)は、大腸がんが存在する病変周辺のみを切除した(表 2-3)。先行研究において、便の潜血検査等のスクリーニング等で発見された大腸がん患者の 27%が右側大腸がん患者である[52]ことから、本研究の大腸がん患者を選抜する過程においてバイアスがないことが示唆された。

先行研究において、腸管洗浄液を飲用せずに正常な排便によって取得した便に含まれる腸内細菌叢と、腸管洗浄液飲用後の最初の排便によって取得した便に含まれる腸内細菌叢を比較した結果、これらに大きな差が確認されなかったことが報告されている[53]。また、腸管洗浄液を使用せずに正常な排便によって取得した便に含まれる代謝産物の量と、腸管洗浄液を飲用した被験者の最初の排便によって取得した便に含まれる代謝産物の量を比較した結果、バリン、ロイシン、イソロイシン等のアミノ酸を含む 32 個の代謝産物の量が異なることが報告されているが、詳細は後述するが本研究で着目する全ての胆汁酸、アミノ酸の 1 種であるセリンやアミノ酸関連代謝産物(Gly-Leu, Urocanate, N,N-Dimethylglycine)については大きく異なることも報告されている[51]。そのため、腸管洗浄液を飲用後の最初の便に含まれている腸内細菌や代謝産物は、正常な排便によって得られる便と同様に、その被験者の腸内環境を反映していると考えられる。

表 2-1 内視鏡検査を受けた 620 人の被験者の疫学データの概要

特徴量	Healthy controls	MP	Stage 0	Stage I/II	Stage III/IV
メタゲノムサンプル (N)	251	67	73	143	86
メタボロームサンプル (N)	245	61	68	140	85
年齢 (中央値±標準偏差)	63±12.6	64±9.12	65±7.54	65±9.67	60.5±11.0
性別 (N, 女:男)	115:136	19:48	28:45	50:93	36:50
BMI (kg/m ² , 中央値±標準偏差)	22.4±3.04	22.1±4.70	22.5±3.57	22.7±3.17	23.1±3.43

表 2-2 大腸がん患者の治療前後の疫学データの概要

特徴量	内視鏡治療	外科的治療
メタゲノムサンプル (N)	11	85
メタボロームサンプル (N)	9	83
治療前後の期間 (日, 中央値±標準偏差)	356±69.4	374±189.2
治療前の内視鏡検査の結果 (N, 0:I:II:III:IV)	11:0:0:0:0	2:20:37:24:2
治療後の大腸がんのリスクが低い患者数 (N)	1	57
治療後の大腸がんのリスクが高い患者数 (N)	3	19
再建法 (N, 左側:右側)	-	65:20
治療から約1年後の内視鏡検査の結果 (N, PH:PMP:PS0)	8:3:0	73:11:1
治療前の血中の総コレステロール (mg/dL, 中央値±標準偏差)	210±33.3	208±40.5
治療から約1年後の血中の総コレステロール (mg/dL, 中央値±標準偏差)	195±12.7	212±38.1
治療前の年齢 (中央値±標準偏差)	69±8.24	62±9.63
性別 (N, 女:男)	5:6	39:46
BMI (kg/m ² , 中央値±標準偏差)	22.5±3.15	22.5±3.37

2.3.2 代謝産物・メタゲノムリードの取得

先行研究[4]と本研究では、620 人の被験者から 716 個の便を取得した。716 個の便から抽出された DNA を whole-genome shotgun シーケンシングすることで、1 サンプルあたり、 4.71×10^7 本(中央値)のメタゲノムリードを取得した。604 個の便由来のメタゲノムリードは先行研究で既に取得しており、112 個の便由来のメタゲノムリードを本研究で新たに取得した。本研究で取得したメタゲノムリード数は、いくつかの大腸がんのコホート研究[8,9,43-46]で取得されたメタゲノムリード数(中央値, 約 4×10^7 から約 7×10^7)と比較しても遜色がないことが確認された。

先行研究[4]と本研究で得られた 716 個の便のうち、694 個の便に含まれている 397 個の代謝産物を CE-TOF MS を用いて定量したデータを取得した。317 個の便に含まれている代謝産物は既に先行研究で測定しており、377 個の便に含まれている代謝産物を本研究で新たに測定した。22 個の便はその量が少なかったため、それらの便に含まれている代謝産物を定量することができなかった。

第3章 ヒト腸内細菌のメタゲノム解析

3.1 序論

本章では、シーケンサーから得られたメタゲノムリードを用いたメタゲノム解析により、716 サンプルの腸内細菌の系統や遺伝子・機能組成データを算出した手法について述べる。また、メタゲノムリードを用いて特定の腸内細菌の生育速度を推定したため、その手法についても述べていく(3.2.4 節)。さらに、メタゲノムリードを用いてデオキシコール酸を生成する *bai* オペロンを定量したため、その手法についても述べる(3.2.5 節)。メタゲノム解析は、主に 3 つの工程で構成されている(図 3-1)。まず、シーケンサーから得られたメタゲノムリードに含まれるクオリティの低いリードや宿主であるヒト由来のリード等を取り除くことで高精度なリードを取得した(3.2.1 節)。腸内細菌の系統組成は、**mOTU2 profiler** を用いて算出した(3.2.3 節)。腸内細菌の遺伝子組成と遺伝子機能組成は、メタゲノムリードから腸内細菌のゲノムを再構築後、そのゲノムに由来する遺伝子カタログを利用して遺伝子を定量する手法を構築し、それを用いて算出した(3.2.2 節)。本研究で構築した遺伝子カタログに収録されている遺伝子の塩基配列に対する 1 サンプルあたりのメタゲノムリードのマッピング率は 73.0%(中央値)であることが確認された(3.2.6 節)。そのため、本研究で構築した手法によってメタゲノムリードに含まれている腸内細菌の遺伝子を十分に検出することができた可能性が示唆された。

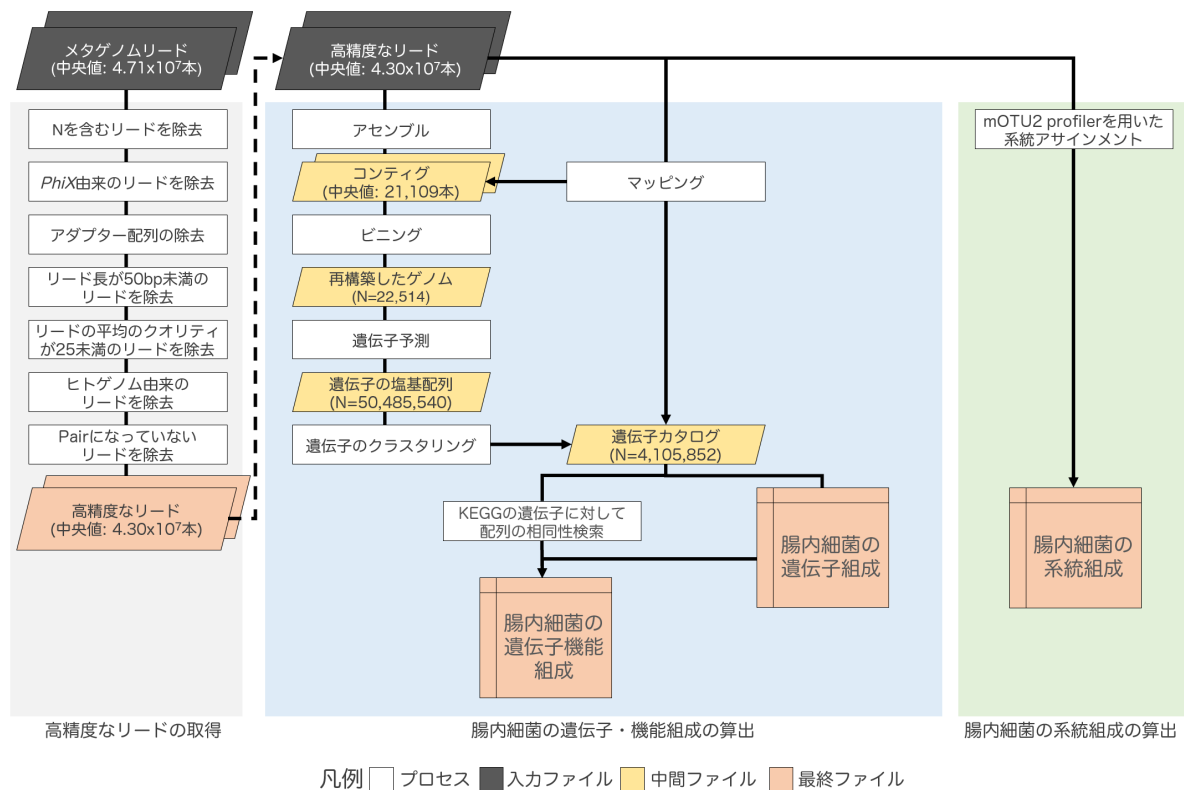


図 3-1 メタゲノム解析の概略図

Whole-genome shotgun シーケンシングによって得られたメタゲノムリードを用いてメタゲノム解析を実施した結果、腸内細菌の系統組成や遺伝子、遺伝子機能組成を取得した。

3.2 材料・手法

3.2.1 高精度なリードの取得

以下の操作はメタゲノムサンプル毎に独立して実施された。2.2.6 節にて取得したシーケンサーから得られたメタゲノムリードを用いて、以下の 7 つの処理で高精度なリードを取得した(図 3-1)。

1. シーケンサーから得られたリードから、N を 1 塩基でも含むリードを除去した。
2. マッピングツールである Bowtie2 (version 2.2.9)[54]を用いて *PhiX* の塩基配列に対してリードをマッピングし、マッピングされた *PhiX* 由来のリードを除去した。*PhiX* の塩基配列は、2017 年に NCBI のデータレポジトリからダウンロードした (GI: 9626372)。パラメータは `--fast-local` を用いた。

3. それぞれのリードに付加されているアダプター配列を cutadapt (version 1.9.1)[55]を用いて除去した。また、リードの 3'末端側はクオリティが低くなる傾向があるため、3'末端側から連続してクオリティ値が 17 以下の塩基配列を除去した。Forward 側からシーケンシングされたリードに対してのパラメータは `-a CTGTCTCTTATACACATCTCCGAGCCCACGAGAC -O 33 -q 17` を用いた。Reverse 側からシーケンシングされたリードに対してのパラメータは `-a CTGTCTCTTATACACATCTGACGCTGCCGACGA -O 32 -q 17` を用いた。
4. 50 bp 未満のリードを除去した。
5. リードのクオリティの平均値が 25 未満のリードを除去した。
6. Bowtie2 を用いてリードをヒトゲノムの塩基配列に対してマッピングし、マッピングされたヒト由来のリードを除去した。ヒトゲノムの塩基配列は、2017 年に NCBI のデータレポジトリからダウンロードした (GI: 568336000 から 568336023 までの 24 個)。パラメータは `--fast-local` を用いた。
7. ペアエンドシーケンシングしたリードのうち、片方でのみ存在するリードを除去した。

それぞれの処理後のリード数は、seqkit(version v0.5.5)[56]の stats モードを用いて算出した。また、それぞれの処理前後のリード数をシーケンサーから得られたリード数で割ることで、それぞれの処理で取り除かれたリード数の割合を算出した。

3.2.2 腸内細菌の遺伝子組成と遺伝子機能組成の算出

本節では、ヒト腸内細菌の研究における遺伝子組成と遺伝子機能組成の算出手法の概要と、本研究におけるその手法について説明する(図 3-1)。

ヒト腸内細菌の研究における遺伝子・機能組成の算出手法は主に 2 つに分類される。参照配列になり得る遺伝子の塩基配列がない場合、アセンブルによって得られたコンティグに対してリードをマッピングし、コンティグに存在する遺伝子を定量する手法がよく用いられる[4]。この手法は、コンティグが参照配列の代替になっているため、コンティグの生成が重要である。コンティグを生成することで、参照配列に存在しないメタゲノムサンプル由来の遺伝子を定量できる可能性がある。しかしながら、相対存在量が少ない腸内細菌由来のコンティグを取得することが難しいため、それらの遺伝子を定量することができないことが問題となっている。

参照配列となる遺伝子の塩基配列がある場合、その参照配列に対してリードをマッピングして遺伝子を定量する手法がよく用いられる[9]。2016年に、ヒト腸内細菌の遺伝子カタログである IGC(Integrated Gene Catalog)が構築され[10]、参照配列として最もよく利用されている。この遺伝子カタログは、1267人の欧米や中国人のメタゲノムリードに由来する約1000万個の遺伝子の塩基配列やアミノ酸配列が収録されているが、収録されている遺伝子の由来となった腸内細菌や遺伝子の位置関係を明らかにすることができない。

上記の問題を解決するために、本研究では、メタゲノムリードから、腸内細菌のゲノムを再構築し、そのゲノムに存在する遺伝子の塩基配列を参照配列として利用することで遺伝子を定量する新たな手法を構築した(図 3-1)。腸内細菌の相対存在量は個人差が大きい。そのため、相対存在量が全体的に少ない腸内細菌でも特定の個人では相対存在量が多いことがある。そのため、メタゲノムサンプル毎にアセンブル・ビニングすることで再構築したゲノムの遺伝子を全てのサンプルで共有後、それを参照配列とすることで、全体的に相対存在量が少ない腸内細菌の遺伝子も定量することが可能である。また、再構築したゲノムの系統を推定することで、それぞれのゲノムに存在する遺伝子と腸内細菌の系統を対応させ、遺伝子の由来となる腸内細菌を調べることも可能である。さらに、参照配列となった遺伝子はアセンブルによって得られているため、遺伝子の位置関係を考慮して、正確に目的の遺伝子を定量することも可能である。本研究で新たに構築した遺伝子を定量するための手法は、以下の4つの工程で構成されている。

1. ゲノムの再構築
2. 再構築したゲノムを基盤とした遺伝子カタログの構築
3. 遺伝子組成の算出
4. 遺伝子機能組成の算出

1. ゲノムの再構築

以下の操作はメタゲノムサンプル毎に独立して実施された。3.2.1 節で取得した 610 個のメタゲノムサンプル由来の高精度なリードを、metaSPAdes(version 3.12.0)[57]を用いてアセンブルした。パラメータは`--meta -t 4`を用いた。計算機資源の問題上、metaSPAdes を用いたアセンブルは、900G のメモリを確保し、4 スレッドで 24 時間実施された。

metaSPAdes を用いてアセンブルすることができなかった 106 個のメタゲノムサンプル由来の高精度なリードを MEGAHIT(version 1.1.3)[58]を用いてアセンブルした。パラメータは`--min-contig-len 1500`を用いた。先行研究において、MEGAHIT と比較して metaSPAdes の方がヒト腸内細菌由来のメタゲノムリードを用いたアセンブルの性能が良いことが報告されているため[14]、metaSPAdes を用いて多くのメタゲノムサンプルをアセンブルした。アセンブルによって得られたコンティグのうち、1500bp 以下のコンティグを seqkit の seq モードを用いて取り除いた。パラメータは`-m 1500`を用いた。

アセンブルによって得られたコンティグと高精度なリードを用いてビニングし、ゲノムを再構築した。ビニングするためのツールはいくつか存在するが、多くのツールでは、コンティグの k -mer の出現頻度や塩基の出現頻度であるテトラヌクレオチド頻度、アセンブルに用いたメタゲノムリードをコンティグにマッピングすることによって得られる、各コンティグのカバレッジの情報を統合し、コンティグを細菌種毎に分類することでゲノムを再構築する。ビニングによって再構築されたゲノムは、そのビニングするツール毎に、その結果が多少異なることが予想される。本研究では、3 つのビニングツールを用いてゲノムを再構築した。その後、それぞれのツールから再構築されたゲノムの中から最もクオリティの良いゲノムを選抜するゲノムの精製も実施した。以下の操作も、メタゲノムサンプル毎に独立して実施された。まず、metaWRAP(version 1.2.1)[59]の binning モードを利用して、ビニングツールである MaxBin2(version 2.2.6)[12]、MetaBAT2(version 2.2.6)[13]、CONCOCT(version 1.0.0)[11]を用いてビニングした。パラメータは、`-l 1500, --metabat2, --maxbin2, --concoct`を用いた。次に、metaWRAP の bin_refinement モードを利用して、MaxBin2、MetaBAT2、CONCOCT によって再構築したゲノムを精製した。パラメータは、`-c 50 -x 10 --quick`を用いた。ゲノムを精製するために、まず、bin_refiner[60]を用い

て、2つの異なるビニングツールから構築されたゲノムから新たなゲノムを構築した。その後、3つのビニングツールを用いて構築されたゲノムと、`bin_refiner`を用いて構築された4つのゲノムから最もクオリティの良いゲノムを選抜した。ゲノムのクオリティは、`CheckM(version 1.0.12)`[61]を、`lineage_wf`モードでパラメータは`--quick`を用いて実施することによって取得されるゲノムの完成率(`completeness`)とゲノムの汚染率(`contamination`)を用いた以下の式によって定義された。

$$\text{Score} = \text{completeness} - 5 \times \text{contamination}$$

`CheckM`は、与えられたゲノムの系統を推定し、与えられたゲノムに含まれている系統特異的なマーカー遺伝子の有無を調べることで、ゲノムの完成率と汚染率を推定するツールである。`CheckM`に用いられている系統特異的なマーカー遺伝子は、ゲノム上でのその遺伝子の位置も保存されているため、与えられたゲノムにおけるその遺伝子の保有率を調べることでゲノムの完成率を推定することができる。また、系統特異的なマーカー遺伝子は、それぞれの系統で1コピーしか存在しないため、与えられたゲノム上でその遺伝子のコピー数を調べることでゲノムの汚染率を推定することができる。

ゲノムの完成率が50%未満またはゲノムの汚染率が10%以上であるゲノムはそのクオリティが低いと判断され、解析の対象から除外した。再構築したゲノムは、`MIMAG(Minimum Information about a Metagenome-Assembled Genome)`[62]に基づいて2つのグループに分類された。ゲノムの完成率が90%以上かつゲノムの汚染率が5%未満であるゲノムはクオリティが高いゲノム(HQ; High-quality)、ゲノムの完成率が90%未満またはゲノムの汚染率が5%以上であるゲノムはクオリティの低いゲノム(LQ; Low-quality)に分類された。

2. 再構築したゲノムを基盤とした遺伝子カタログの構築

1で再構築したゲノム毎に、`Prokka(version 1.14.0)`[63]を用いて、ゲノムから遺伝子の塩基配列を予測した。パラメータは変更せずデフォルトを用いた。再構築したゲノムから予測された全ての遺伝子の塩基配列を`CD-HIT EST(version 4.7)`[64]を用いて、`identity=95%`、`coverage=90%`の閾値でクラスタリングした。パラメータは、先行研究においてヒト腸内細菌の遺伝子カタログである`IGC[10]`を構築したパラメータと同じパラメータである`-c 0.95 -G 0 -aS 0.9 -g 1 -r 1 -d -M 0`を用いた。クラスタリングによって得られた代表配列を参照配列とした。以上の操作で、遺伝子カタログを構築することができた。

参照配列になった遺伝子の由来となる腸内細菌を特定するためには、全ての再構築したゲノムの系統を推定する必要がある。しかしながら、それには膨大な計算時間が要求され

るため、計算量を減らすために、dRep(version 2.2.3)[65]を用いてゲノムの類似性に基づいたクラスタリングを行い、それぞれのクラスターで代表となったゲノムの系統を推定した。

dRep は、2 種類のクラスタリングで構成されている。まず、与えられたゲノムセットにおける全てのゲノム間の類似性を Mash(version 2.2)[66]を用いて算出後、90%の類似性でクラスタリングした。Mash は、MinHash 法を用いて 2 つのゲノムを構成する k -mer の有無を比較することで 2 つのゲノムの類似性を算出するツールである。Mash は、計算速度は速いが、その正確性は、ゲノムをアラインメント等して比較していないため後述する ANI(Average Nucleotide Identity)と比較して低いことが知られている。そのため、Mash を用いたクラスタリングによって得られたクラスターを構成する全てのゲノム間の類似性を MUMmer(version 3.1)[67]を用いて算出し、95%の類似性で再度クラスタリングした。MUMmer は、2 つのゲノムをアラインメントすることで、2 つのゲノム間の ANI を算出するツールである。MUMmer を用いたクラスタリングの結果、それぞれのクラスターで最もクオリティの良いゲノムを選抜し、代表ゲノムとした。ゲノムのクオリティの算出方法は以下の式により算出した。

$$\text{Score} = \text{completeness} - 5 \times \text{contamination} + 0.5 \times \log(\text{N50})$$

それぞれのゲノムの completeness と contamination は、前述した CheckM を用いて算出した。

計算機資源の問題上、再構築した 22,547 個の腸内細菌のゲノムセットを、5,000 個のゲノムセット毎にランダムに分割し、dRep を用いてクラスタリングした。パラメータは、--S_ani 0.95 を用いた。ランダムに分割したゲノムセットをクラスタリングして得られた全てのゲノムを、dRep を用いて再度クラスタリングし、代表ゲノムを取得した。

代表ゲノムの系統は、GTDB-Tk(version 1.3.0)[68]の classify_wf モードを用いて推定した。パラメータは特に変更せずデフォルトを用いて GTDB-Tk を実行し、DB は version 95 を用いた。GTDB-Tk は、多くの細菌が保有し、それぞれの細菌で 1 コピーしか存在しない遺伝子の配列を与えられたゲノムから予測し、それらを用いて、参照となるゲノムの系統樹における系統を推定するツールである。系統推定した結果、与えられたゲノムと、最も近い系統である参照となるゲノムの類似性を、MinHash 法を基盤とした FastANI(version 1.31)[69]を用いて算出し、その類似性によって、その細菌の種や属等の系統のレベルを推定することが可能である。例えば、与えられたゲノムと参照となったゲノムの類似性が 95%以上であった場合、与えられたゲノムは参照となった細菌と同じ種である。GTDB-Tk によって推定された系統は、GTDB-Tk の github (<https://github.com/Ecogenomics/GTDBTk/>)に収録されている

gtdb_to_ncbi_majority_vote.py を用いて、NCBI Taxonomy に変換した。以上の操作で、ゲノムに存在する遺伝子とそのゲノムの系統を対応させることができた。

3. 遺伝子組成の算出

以下の操作はメタゲノムサンプル毎に独立して実施された。2 で構築した遺伝子カタログに収録されている遺伝子の塩基配列に対して、3.2.1 節で取得した高精度なリードを BWA(version 0.7.17-r1188)[70]を用いてペアエンドマッピングした。BWA のパラメータはデフォルトで、mem アルゴリズムを使用した。マッピングした結果、参照配列となった遺伝子の塩基配列とメタゲノムリードの一致率が 95%未満であった結果を取り除いた後に、参照配列となった遺伝子毎に、マッピングされたメタゲノムリード数を算出した。参照配列となった遺伝子にマッピングされたメタゲノムリード数を参照配列となった遺伝子長で割ることで遺伝子の量を算出した。遺伝子の相対存在量は、算出した遺伝子の量の総和で、それぞれの遺伝子の量を割ることで算出した。

4. 遺伝子機能組成の算出

2 で構築した遺伝子カタログに収録されている遺伝子の塩基配列をアミノ酸配列に変換した。そのアミノ酸配列を、DIAMOND(version 0.9.14.115)[71]を用いて、遺伝子機能データベースである KEGG(Kyoto Encyclopedia of Genes and Genomes; 2017 年に取得)[72]に収録されている原核生物の遺伝子のアミノ酸配列に対して、配列の相同性検索を実施した。DIAMOND のパラメータは、blastp --sensitive を用いた。配列の相同性検索の結果、KEGG に収録されている遺伝子に対して identity < 40%または bit Score < 70 または Coverage < 80 未満の検索結果を取り除き、遺伝子カタログに収録されている遺伝子と KEGG に収録されている遺伝子を対応させた。

3 で算出した遺伝子の相対存在量とその遺伝子を KEGG に収録されている遺伝子に対応させて、KEGG に収録されている遺伝子毎に遺伝子の相対存在量を足し合わせることで、KEGG に収録されている遺伝子組成を算出した。遺伝子カタログに収録されている遺伝子が複数の KEGG に収録されている遺伝子に対応した場合、その遺伝子の相対存在量をその対応した数で割ることで補正した。

KEGG に収録されている遺伝子と KEGG ORTHOLOGY(KO)を対応させることで、KO レベルの遺伝子機能組成を算出した。KO とは、KEGG に収録されている遺伝子の機能に基づいて定義されたオーソログである。遺伝子機能組成を構成する特定の KO の相

対存在量を調べることで、特定の反応を担う遺伝子の相対存在量を明らかにすることができる。

3.2.3 腸内細菌の系統組成の算出

Whole-genome shotgun シーケンシングによって得られたメタゲノムリードを用いて、その環境における細菌の系統組成を算出するツールがこれまでにいくつか開発されてきた。それらのツールの性能を評価した結果、メタゲノムリードを *k-mer* に分解し、参照配列となるゲノムに対してアラインメントする **Kraken 2**[73]が最も性能が良いことが報告されている[74]。しかしながら、前述した、メタゲノム解析を用いた大腸がんのコホート研究の1つ[9]は、**mOTU2 profiler**(version 2.0.0)[75]を用いて腸内細菌の系統組成を算出したため、本研究でも、**Kraken 2**ではなく **mOTU2 profiler**(version 2.0.0)を用いて、属及び種レベルの腸内細菌の系統組成データを取得した(図 3-1)。本研究で使用したツールを先行研究で使用したツールに統一することで、先行研究で明らかになった結果と本研究で明らかになった結果を比較する際に発生するツール間の齟齬を解消することができる。

mOTU2 profiler は、多くの細菌が保有し、それぞれの細菌で1コピーしか存在しない10個の系統マーカー遺伝子の塩基配列に対してメタゲノムリードをマッピングすることで、その環境における細菌の系統組成を算出するツールである。系統マーカー遺伝子は、25,000以上の細菌のゲノムと3,100以上のメタゲノムサンプルのコンティグから生成されているため、メタゲノムサンプルに含まれていてゲノムが明らかにされていない細菌を検出することが可能である。例えば、*unknown Dialister*や*unknown Peptostreptococcaceae*は、ゲノムが明らかにされていない腸内細菌であり、多くの大腸がんのコホートにおいて、健常者と比較して大腸がん患者でその相対存在量が多いことが報告されている[9]。

以下の操作はメタゲノムサンプル毎に独立して実施された。**mOTU2 profiler** の **map_tax** モードのデフォルトパラメータを用いて、10個の系統マーカー遺伝子の塩基配列に対して3.2.1節で取得した高精度なリードをマッピング後、**mOTU2 profiler** の **calc_mgc** モードのデフォルトパラメータを用いて、それぞれの系統マーカー遺伝子に対してマッピングされたリード数を算出し、**calc_motu** モードの **k** のパラメータを **mOTU** または **genus** にそれぞれ設定して、種レベル、属レベルの腸内細菌の系統組成を算出した。

3.2.4 腸内細菌の生育速度の推定

第5章にて、詳細な手法や結果について後述するが、大腸がん外科的治療前後で変化する腸内細菌を特定するために、大腸がん患者の外科的治療前後の腸内細菌の相対存在量を統計学的手法を用いて比較した。その結果、患者の外科的治療後に24種の腸内細菌の相対存在量が有意に増加したことが確認された($P < 0.005$)。本節では、生育速度を推定する意義とその手法の原理、外科的治療後に相対存在量が増加した腸内細菌の生育速度の推定手法、腸内細菌の推定された生育速度とその相対存在量の関係性を調査した手法について説明する。

1. 生育速度を推定する意義とその手法の原理

相対存在量の比較解析において、ある細菌の相対存在量の変化が他の細菌の相対存在量の変化によって引き起こされる可能性があり、それによって、ある細菌の菌数が変化していないにも関わらず、相対存在量としては変化しているように観察されてしまう偽陽性が生じる可能性がある[76]。そのため、相対存在量を統計学的に比較することで特定した細菌の菌数を定量的PCRすることで定量することで、その細菌の菌数も異なることを確認することが重要である[4,8,9]。しかしながら、1回の定量的PCRで1つのサンプルの1つの細菌の菌数しか定量できず、1回の定量的PCRに必要な金銭的なコストが高いため、多くの細菌の菌数を定量することができないことが問題となっている。

これらの問題を解決する1つ方法として、細菌のゲノムと、その細菌のDNAをwhole-genome shotgunシーケンシングすることで得られるリードを用いて、その細菌のゲノム上のリードの分布を明らかにすることで細菌の生育速度を推定する手法が2015年に確立された[77]。この手法では、細菌のゲノムが環状であり、複製開始点である*ori*から複製の終点である*ter*まで両方向からゲノムが複製されるという仮定に基づいて生育速度を推定する。一般的に、細胞のゲノムが複製された後に、細胞が分裂するため、ゲノムの複製率はその細胞の生育速度を反映していると考えられている。生育速度が高い細菌のDNAをwhole-genome shotgunシーケンシングした場合、その細菌のゲノムの複製の速度に応じて既にゲノムが複製されている領域がまだ複製されていない領域よりも多くシーケンシングされるため、ゲノム上のリードの分布に勾配が生じる(図3-2a)。一方で、生育速度が低い細菌のDNAをwhole-genome shotgunシーケンシングした場合、複製されているゲノムの領域が少ないため、ゲノム全体の領域がほぼ均等にシーケンシングされる(図3-2a)。細菌の生育速度は、ゲノム上の領域毎にリードがどれくらいマッピングされたのかを表すカバレッジを算出後、線形回帰を基盤とした手法により複製の終点周辺のカバレッジに対する複製開始点周辺のカバレッジの比率(PTR; peak-to-trough ratio)を算出することで推定される。単一の

細菌の培養系における細菌の増殖の推移は、その培養系における推定された生育速度の推移と対応していることも報告されている[77]。この手法は、単一の細菌の培養系だけではなく、多くの細菌が混在する環境でも適用することができるが、細菌の環状のゲノムとその細菌由来の多くのリードが必要である。これらの問題を解決するために、いくつかの生育速度を推定するツールが開発されてきた[78–80]が、本研究では、環状のゲノムではないドラフトゲノムを用いて低カバレッジでも生育速度を推定できるツールである GRiD(version 1.2)[81]を用いて細菌の生育速度を推定した。

GRiD は、特定の細菌のドラフトゲノムに対してメタゲノムリードをマッピングし、ドラフトゲノムを構成するそれぞれのコンティグ毎にカバレッジを算出後、そのカバレッジの情報に基づいてコンティグを並び替えた後に、平滑化曲線を基盤とした手法により、複製の終点周辺のカバレッジに対する複製開始点周辺のカバレッジの比率を算出することで生育速度を推定するツールである(図 3-2b)。コンティグの並び替えは、カバレッジが最大となるコンティグが最初と最後に位置し、カバレッジが最小となるコンティグが中心に位置するように対称に並び替えられるため、ゲノム上のリードの分布の勾配を調べることができる。

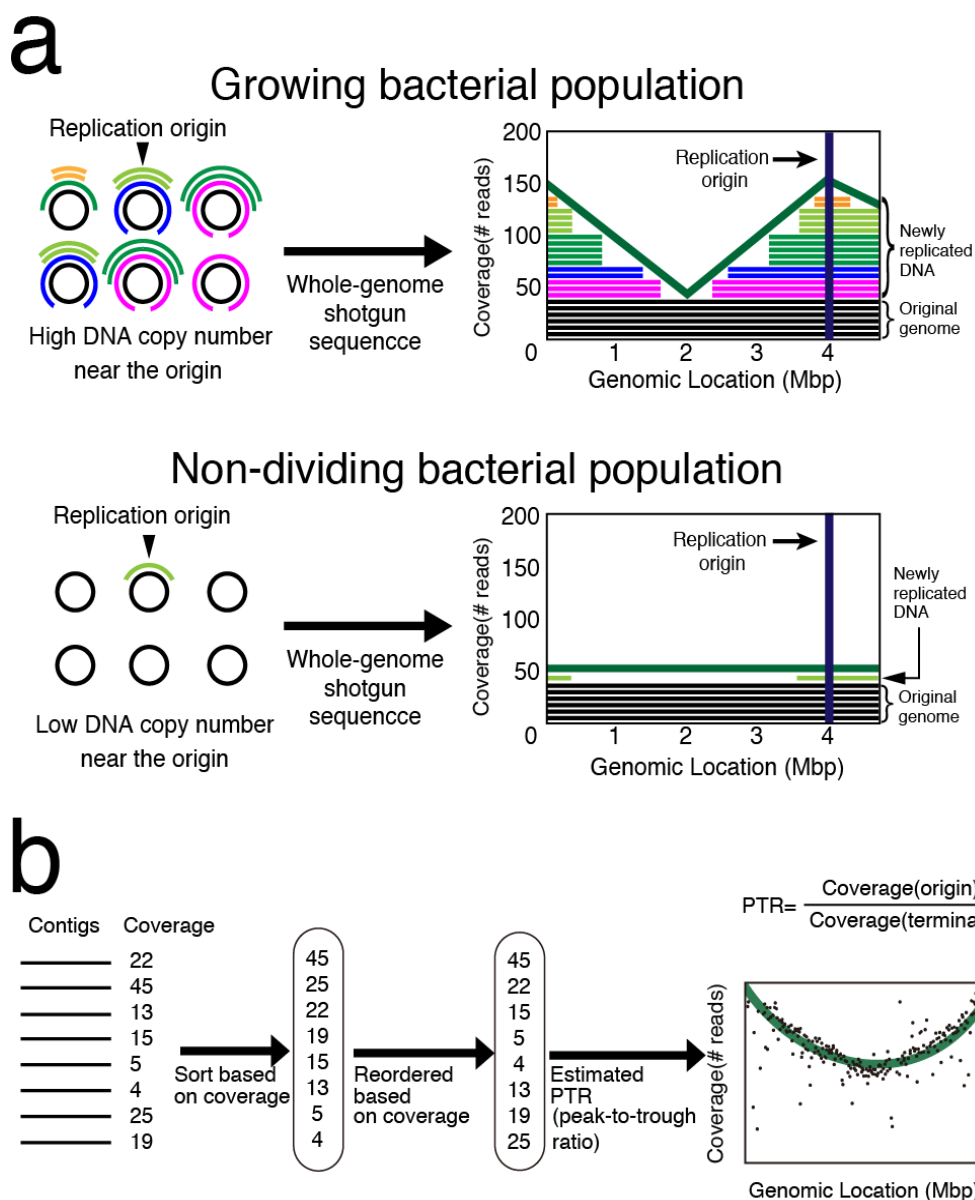


図 3-2 生育速度の推定手法

a は生育速度が速い細菌と遅い細菌の DNA を whole-genome shotgun シーケンシングすることによって得られるリードをその細菌のゲノムにマッピングし、ゲノム上のリードの分布を示した図である。b は GRiD を用いた生育速度の推定手順について示した図である。

2. 外科的治療後に相対存在量が増加した腸内細菌の生育速度の推定手法

以下の操作はそれぞれの腸内細菌に対してメタゲノムサンプル毎に独立して実施された。

GRiD の single モードのデフォルトパラメータを用いて、外科的治療後に相対存在量が増加した腸内細菌の生育速度を推定した(図 3-3)。GRiD を用いた生育速度の推定には、相対存在量が増加した 24 種の腸内細菌のゲノムが必要である。相対存在量が増加した腸内細菌のゲノムは、再構築した腸内細菌の種名と mOTU2 profiler における 24 種の腸内細菌の種

名を NCBI Taxonomy を用いて照合することで取得した。24 種の腸内細菌のうち、7 種の腸内細菌については、ゲノムや mOTU2 profiler における NCBI Taxonomy が明らかにされていない等の理由で再構築したゲノムと照合することができなかった。

生育速度の推定には、その腸内細菌由来の多くのメタゲノムリードが必要である。メタゲノムサンプル毎に腸内細菌のゲノムに対して 3.2.1 節で取得した高精度なリードをマッピングした結果、一部のメタゲノムサンプルでは、ゲノムのカバレッジが 0.2 未満であったため、その腸内細菌の生育速度を推定することができなかった。生育速度を推定することができなかったメタゲノムサンプルは、その腸内細菌の統計解析から取り除いた。

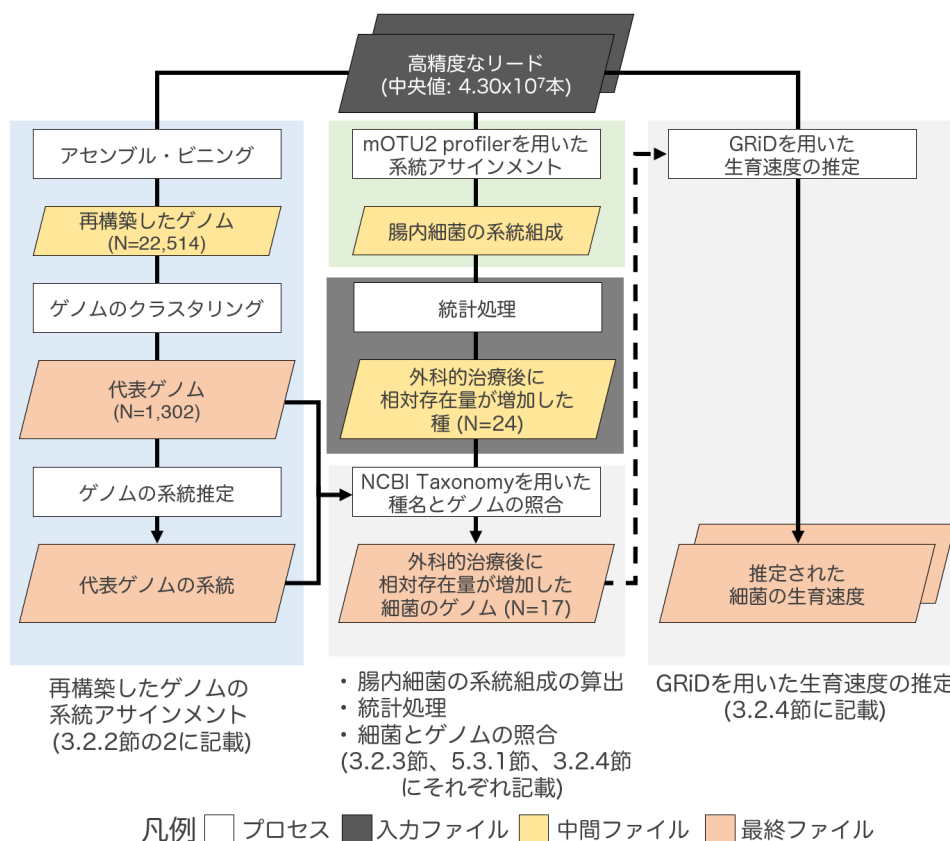


図 3-3 生育速度の推定手順の概略図

背景の色は実施された章や節毎に異なることを示している。水色は、3.2.2 節、緑色は 3.2.3 節、濃い灰色は 5.3.1 節、薄い灰色は 3.2.4 節にて実施されたことを示している。

3. 腸内細菌の生育速度と相対存在量の関係性

腸内細菌の相対存在量と推定された生育速度の関係性を明らかにするために、大腸がん外科的治療前の患者と大腸がん外科的治療後の患者における、腸内細菌の相対存在量と腸内細菌の推定された生育速度の spearman の相関係数を、R の cor を用いてそれぞれ算出した。Method のパラメータは spearman を用いた。また、これらが統計学的に相関関係にあるのかを調べるために R の cor.test を用いて無相関検定し P 値をそれぞれ取得した。Method のパラメータは spearman を用いた。それぞれの腸内細菌において、生育速度を推定することができなかった患者を取り除いて相関係数や P 値を算出した。

3.2.5 *Bai* オペロンの定量

第6章にて詳細に後述するが、大腸がん患者の外科的治療前後の便に含まれている代謝産物の量を比較した結果、コール酸と発がん性が報告されているデオキシコール酸の量が外科的治療後に増加したことが確認された。コール酸からデオキシコール酸への変換は、コール酸のステロイド核の C-7 位に付加されている α ヒドロキシ基を外す反応である。この反応は複雑であるため、特定の1つの遺伝子ではなく、隣接した8つの *bai* 遺伝子(*baiB*, *baiCD*, *baiE*, *baiA*, *baiF*, *baiG*, *baiH*, *baiI*)から構成される *bai* オペロンによってその反応が制御されている。2019年に、*bai* オペロンを構成する8つの *bai* 遺伝子のうち、6つの *bai* 遺伝子(*baiB*, *baiCD*, *baiE*, *baiF*, *baiA*, *baiH*)がコール酸からデオキシコール酸に変換する反応に必須であることが報告されている[82]。

Bai オペロンを構成する8つの *bai* 遺伝子の配列は、NCBI等の公共データベースで公開されているが、その数が非常に少ないため、メタゲノムリードからこれらの遺伝子を定量するのは困難であった。2019年に、隠れマルコフモデル(Hidden Markov Model; HMM)や配列の相同性検索を基盤とした手法によって、ヒト腸内細菌の遺伝子カタログである IGC[10]に収録されている遺伝子配列とその遺伝子組成から、これらの遺伝子やオペロンを特定し、それらを定量する手法が確立された[9]。この手法では、遺伝子の相対存在量の共起関係を基にオペロンを定義し、それを定量しているが、遺伝子の位置関係を明らかにしていないため、デオキシコール酸の生成に必須である全ての遺伝子が含まれていないオペロンを定量している可能性がある。

本研究では、この手法を改良し、以下の3つの工程で、隣接する6つの *bai* 遺伝子(*baiB*, *baiCD*, *baiE*, *baiF*, *baiA*, *baiH*)を保有する腸内細菌のゲノムを再構築し、その遺伝子と遺伝子カタログに収録されている遺伝子に対応させて *bai* オペロンを定量した(図3-4)。

1. 隠れマルコフモデルの構築
2. 隠れマルコフモデルを用いたデオキシコール酸生成菌の探索
3. 遺伝子カタログに収録されている *bai* オペロンの定量

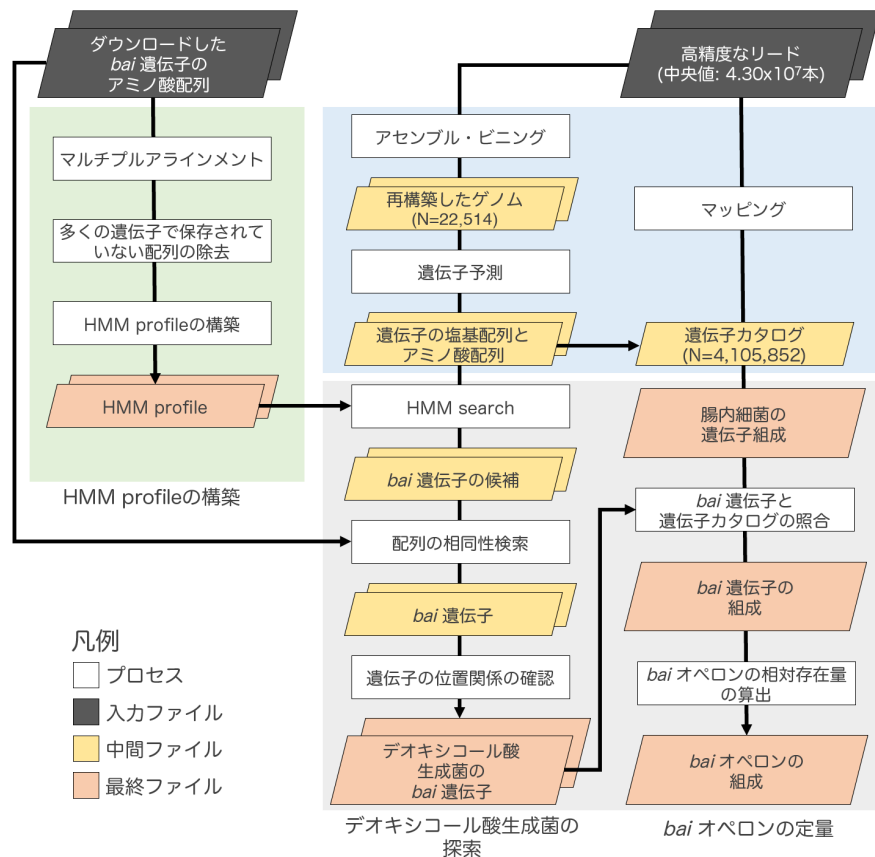


図 3-4 *Bai* オペロンの定量手順の概略図

背景の色は実施された章や節毎に異なることを示している。水色は、3.2.2 節、緑色は 3.2.5 節の 1、薄い灰色は 3.2.5 節の 2 と 3 にて実施されたことを示している。

1. 隠れマルコフモデルの構築

以下の操作は *bai* オペロンを構成する 8 つの遺伝子毎に独立して実施された。まず、*bai* オペロンを構成する遺伝子のアミノ酸配列を UniProt から 2020 年にダウンロードした(付録 2)。ダウンロードした遺伝子から断片化されている遺伝子を取り除いた後、mafft(version 7.427)[83]を用いてマルチプルアライメントした。パラメータは--autoを用いた。マルチプルアライメントした遺伝子のアミノ酸配列を trimAI(version1.2rev59)[84]を用いて、多くの遺伝子で保存されているアミノ酸配列以外のアミノ酸配列を取り除いた。パラメータは-gt 0.9 -cons 60を用いた。HMMER software(version 3.2)[85]の hmmbuild モードを用いて、hmm profile を構築した。

2. デオキシコール酸の生成菌の探索

以下の操作は 3.2.2 節の 1 で再構築した 22,547 個の腸内細菌のゲノム毎に独立して実施された。ゲノムに含まれている全ての遺伝子のアミノ酸配列に対して、1 で構築した 8 つの

hmm profile を HMMER software の hmm scan モードを用いてそれぞれ適用することで、*bai* オペロンを構成する *bai* 遺伝子の候補を探索した。パラメータは $E\ 1 \times 10^{-10}$ を用いた。次に、その候補となった遺伝子を 1 でダウンロードした遺伝子に対して BLAST+(version 2.2.30)[86] の BLASTP モードを用いて配列の相同性検索を実施した。パラメータはデフォルトパラメータを用いた。配列相同性検索の結果、1 でダウンロードした遺伝子に対して identity < 70% 未満の検索結果を取り除き、*bai* オペロンを構成する *bai* 遺伝子を特定した。最後に、これらの *bai* 遺伝子の位置関係を調べることで、ゲノム上で 6 つの *bai* 遺伝子 (*baiB*, *baiCD*, *baiE*, *baiF*, *baiA*, *baiH*) が隣接しているかどうかを調査した。これら 6 つの *bai* 遺伝子を保有する腸内細菌を、デオキシコール酸生成菌として定義した。

3. 遺伝子カタログに収録されている *bai* オペロンの定量

デオキシコール酸生成菌の *bai* オペロンを構成する 8 つの *bai* 遺伝子 (*baiB*, *baiCD*, *baiE*, *baiA*, *baiF*, *baiG*, *baiH*, *baiI*) と、3.2.2 節の 2 で構築した遺伝子カタログに収録されている遺伝子をそれぞれ対応させた。遺伝子カタログに収録されている遺伝子の相対存在量を 3.2.2 節の 3 で算出した遺伝子組成から抽出後、それぞれの生成菌の 8 つの *bai* 遺伝子の相対存在量を遺伝子毎に足し合わせることで、8 つの *bai* 遺伝子の組成を算出した。この *bai* 遺伝子の組成に含まれている 8 つの遺伝子の相対存在量の総和をメタゲノムサンプル毎に算出することで、*bai* オペロンの組成を取得した。

3.2.6 遺伝子カタログの評価

本研究で構築した遺伝子カタログと既存のヒト腸内細菌の遺伝子カタログである IGC[10] に収録されている遺伝子の重複率と、それらに対するメタゲノムリードのマッピング率を評価した。

遺伝子の重複率の評価

まず、既存のヒト腸内細菌の遺伝子カタログである IGC に収録されている遺伝子の塩基配列を以下の url

(ftp://climb.genomics.cn/pub/10.5524/100001_101000/100064/1.GeneCatalogs/IGC.fa.gz)

からダウンロードした。IGC に収録されている遺伝子に対する、本研究によって得られた遺伝子カタログに収録されている遺伝子の割合を明らかにするために、IGC に収録されている遺伝子の塩基配列と、3.2.2 節の 2 で構築した遺伝子の塩基配列を統合後、CD-HIT

EST を用いて identity=95%、coverage=90%の閾値でクラスタリングした。クラスタリングのパラメータは、-c 0.95 -G 0 -aS 0.9 -g 1 -r 1 -d -M 0 を用いた。遺伝子の重複率の可視化は、R の VennDiagram(version 1.6.20)パッケージの venn.diagram を用いた。

遺伝子カタログに対するメタゲノムリードのマッピング率の評価

本節の操作はメタゲノムサンプル毎に実施された。まず、3.2.2 節の 2 で構築した遺伝子カタログに収録されている遺伝子の塩基配列に対して、3.2.1 節で取得した高精度なリードを、BWA を用いてペアエンドマッピングした。次に、IGC に収録されている遺伝子の塩基配列に対して、3.2.1 節で取得した高精度なリードを、BWA を用いてペアエンドマッピングした。最後に、IGC と本研究で構築した遺伝子カタログを統合した遺伝子カタログに収録されている遺伝子の塩基配列に対して、3.2.1 節で取得した高精度なリードを、BWA を用いてペアエンドマッピングした。BWA のパラメータはデフォルトで、mem アルゴリズムを使用した。マッピングした結果、参照配列となった遺伝子の塩基配列とメタゲノムリードの一致率が 95%未満であった結果を取り除き、それぞれの遺伝子カタログに収録されている遺伝子の塩基配列にマッピングされたリード数を算出した。それぞれの遺伝子カタログに収録されている遺伝子の塩基配列にマッピングされたリード数を高精度なリード数で割ることで、それぞれの遺伝子カタログに対するメタゲノムリードのマッピング率を算出した。

3.2.7 公開されているメタゲノムリードを用いたメタゲノム解析

2.2.7 節で取得した Voigt et al.由来の 29 サンプルのメタゲノムリードは、3.2.1 節で前述した高精度なリードの取得と同じ手法で高精度なリードを取得した。高精度なリードを用いたアセンブル・ビニング等を実施せずに、3.2.2 節の 3 と 4 で前述した遺伝子組成及び遺伝子機能組成の算出と同じ手法で、高精度なリードを 3.2.2 節の 2 で構築した遺伝子カタログに収録されている遺伝子の塩基配列に対してマッピングすることで、遺伝子組成と遺伝子機能組成を算出した。

3.3 結果・考察

3.3.1 高精度なリードの取得

シーケンサーから、1 サンプルあたり 4.71×10^7 本(中央値)のメタゲノムリードを取得した。それに含まれているクオリティの低いリードや、宿主であるヒト由来のリード等を取り除くことで、1 サンプルあたり 4.30×10^7 本(中央値)の高精度なリードを取得した。シーケンサーから得られたメタゲノムリードの 6.55%(中央値)が、高精度なリードを取得する過程で取り除かれたことが確認された(図 3-5a)。また、高精度なリードを取得する過程において、ヒト由来のリードを除去する工程で最も多くのメタゲノムリードが除去されたことも確認された(図 3-5b)。さらに、ヒト由来のリードの割合は、特定のメタゲノムサンプルでのみ多いことも確認された(図 3-5b)。これらの結果は、特定のメタゲノムサンプルにおいて便に多くのヒト由来の DNA が混入したことを反映している可能性が考えられる。先行研究において、炎症性腸疾患の患者の便にはヒトの DNA が多く混入しており、炎症等による出血がヒトゲノムの供給源である可能性が報告されている[87]。また、便の潜血検査は大腸がんを発見するためのスクリーニングに最も用いられていることも報告されている[33]。以上のことから、ヒトゲノムの混入量は大腸がんに関係し、その供給源は大腸がん起因する出血である可能性が示唆された。

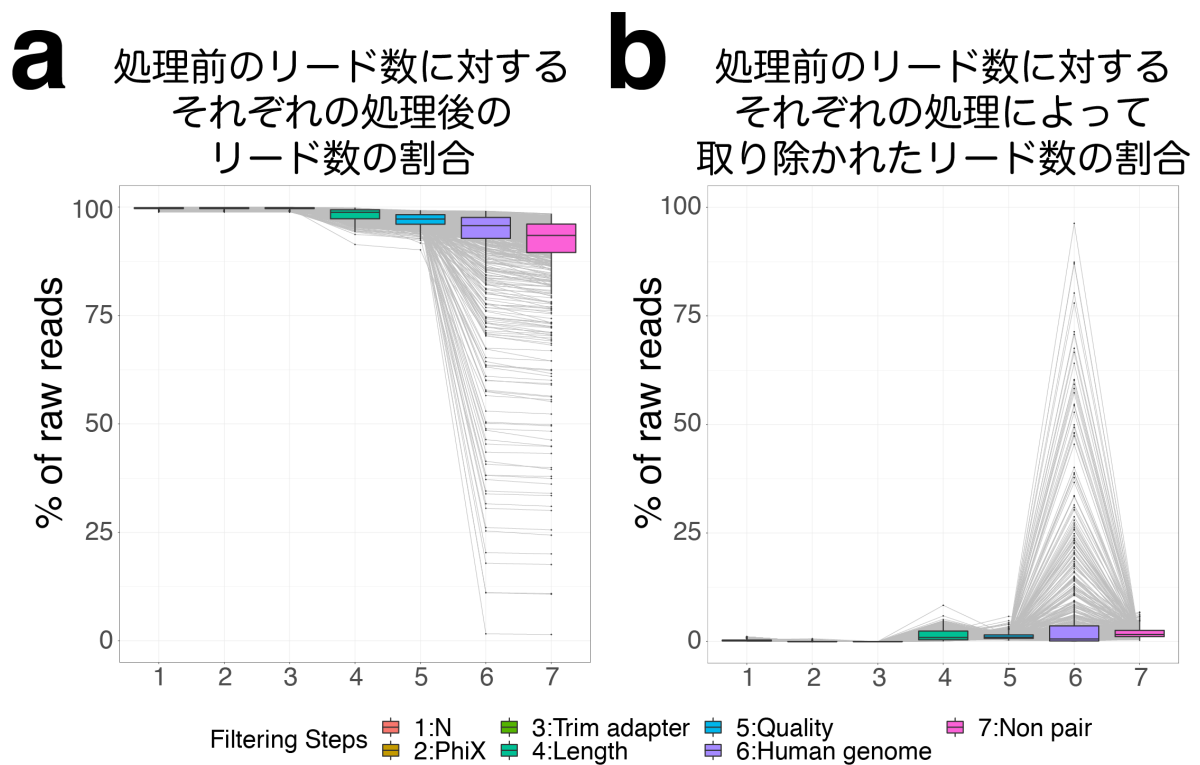


図 3-5 高精度なリードを取得する過程におけるリード数の推移

横軸の番号は、高精度なリードを取得する過程で、それぞれの処理を実施した順番に対応している(凡例参照)。灰色の線は、同一のメタゲノムサンプルであることを示している。

3.3.2 再構築したゲノムのクオリティの評価

メタゲノムサンプル毎にそのメタゲノムサンプルの高精度なリードをアセンブルし、コンティグを取得した。これらのコンティグのうち、1500bp以下のコンティグを取り除いた結果、1サンプルあたり、21,109本(中央値)のコンティグを取得した。1サンプルあたりのコンティグのN50は、13150.5(中央値)であった。

高精度なリードとそれから得られたコンティグを用いて、メタゲノムサンプル毎に、ゲノムを再構築した。それらのゲノムのクオリティを評価した結果、1サンプルあたり、14個のクオリティが低いゲノム(中央値)と、17個のクオリティが高いゲノム(中央値)を再構築することができた。各メタゲノムサンプルから再構築したゲノムを統合した結果、716個のメタゲノムサンプルから10,148個のクオリティが低いゲノムと、12,399個のクオリティが高いゲノムを再構築することができた(図 3-6b)。ヒト腸内細菌は、同じ種でも株毎に保有している遺伝子が異なることがあるため、メタゲノムリードから再構築した腸内細菌

のゲノムもカタログ化することが提案されている[14–17]。そのため、本研究で再構築したゲノムも、ヒト腸内細菌のゲノムのリソースとして利用できる可能性が示唆された。

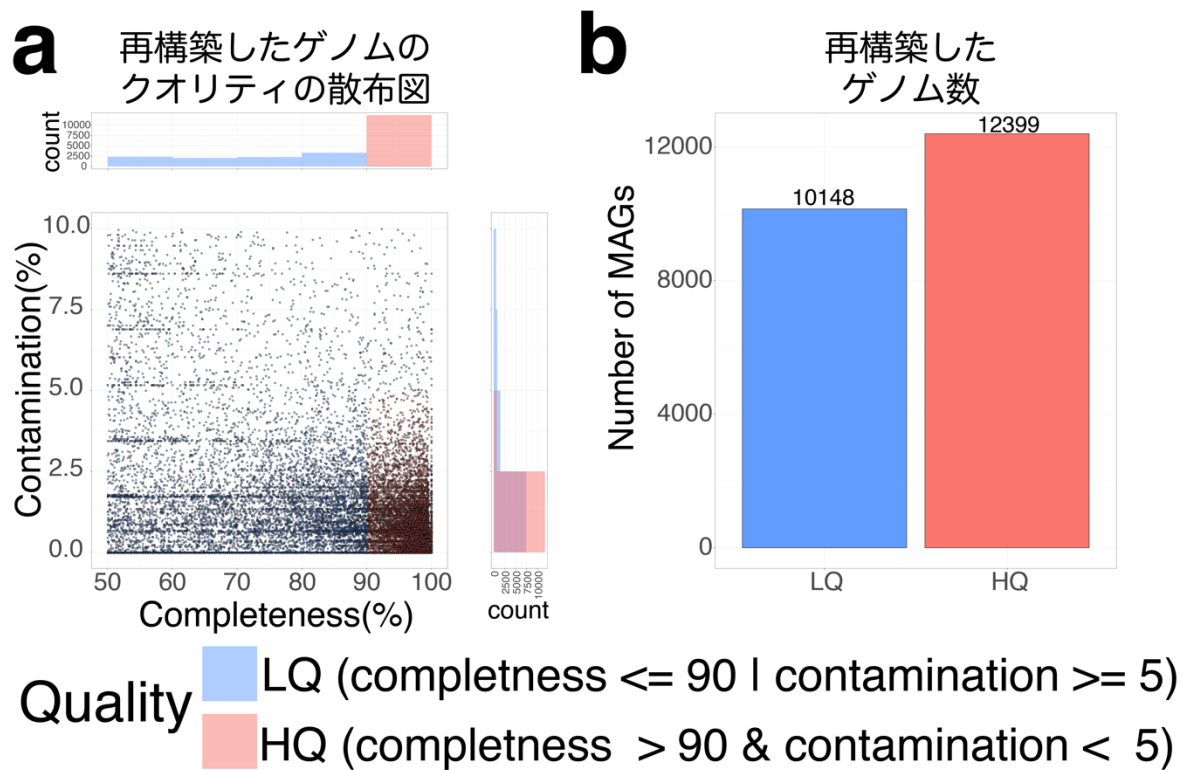


図 3-6 再構築した腸内細菌のゲノムのクオリティ

a は再構築した 22,547 個の腸内細菌のゲノムの完成率とゲノムの汚染率を示した散布図である。b の棒グラフはそれぞれのグループに属するゲノム数を示している。色はゲノムのクオリティによって分類したグループを表している(凡例参照)。

3.3.3 遺伝子カタログの構築とその評価

再構築した 22,547 個のゲノムから 50,485,540 個の遺伝子が予測された。これらの遺伝子を、その塩基配列の類似性に基づいてクラスタリングした結果、4,105,853 個の遺伝子が収録されている遺伝子カタログを構築することができた。

まず、本研究で構築した遺伝子カタログと既存のヒトの腸内細菌の遺伝子カタログである IGC に収録されている遺伝子の重複率を調査した。その結果、本研究で構築した遺伝子カタログは IGC の 29.4%しか網羅できていないことが確認された(図 3-7a)。一方で、本研究で構築した遺伝子カタログの 32.1%の遺伝子は、IGC に収録されている遺伝子と重複しないことも確認された(図 3-7a)。

次に、遺伝子カタログに収録されている遺伝子の塩基配列に対するメタゲノムリードのマッピング率を調査した。本研究で構築した遺伝子カタログに対する1サンプルあたりのメタゲノムリードのマッピング率は73.0%(中央値)であることが確認された(図 3-7b)。IGC や、IGC と本研究で構築した遺伝子カタログを統合した遺伝子カタログに対する1サンプルあたりのメタゲノムリードのマッピング率は、それぞれ74.9%(中央値)、76.9%(中央値)であることが確認された(図 3-7b)。本研究で構築した遺伝子カタログの遺伝子は、IGC に収録されている遺伝子の29.4%しか網羅できていないにも関わらず、本研究で構築された遺伝子カタログと IGC に対する1サンプルあたりのメタゲノムリードのマッピング率の差分はわずか1.9%であった。以上の結果から、本研究で構築した遺伝子カタログを用いた遺伝子組成の算出手法によって、メタゲノムリードに含まれている腸内細菌の遺伝子を十分に検出することができた可能性が示唆された。

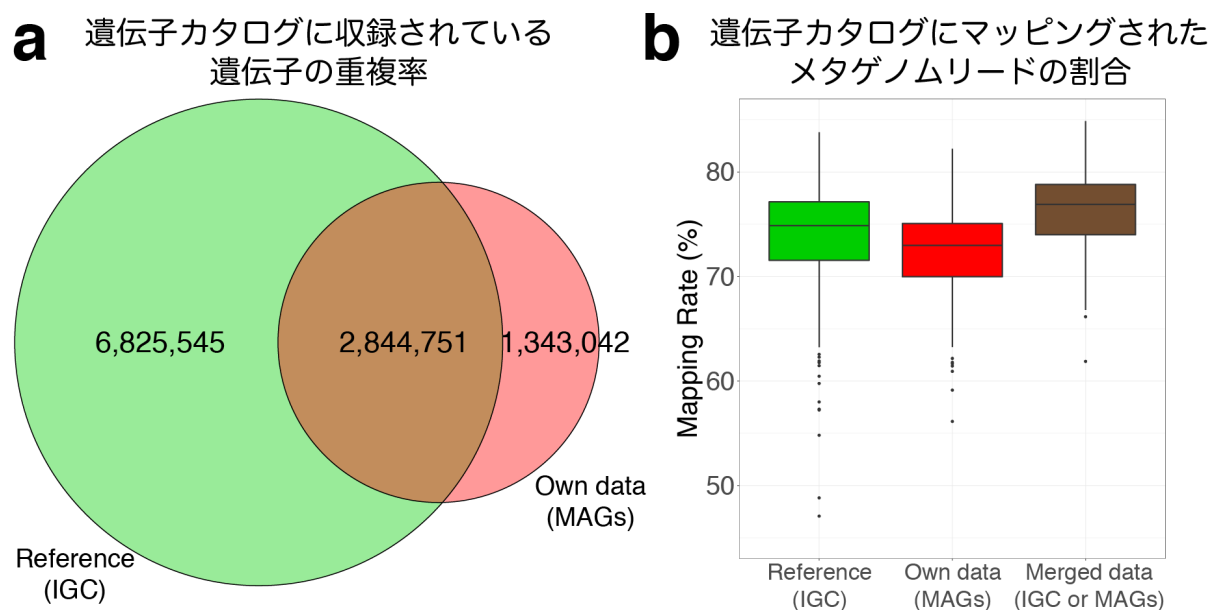


図 3-7 遺伝子カタログの評価

a は既存の遺伝子カタログである IGC に収録されている遺伝子と本研究で取得した遺伝子カタログに収録されている遺伝子の重複した数を表したベン図である。b は、それぞれの遺伝子カタログに対するメタゲノムリードのマッピング率を示した箱ひげ図である。緑色が IGC、赤色が本研究で取得した遺伝子カタログ、茶色がそれらを統合した遺伝子カタログに対するマッピング率を示している。

3.3.4 遺伝子機能組成の算出

遺伝子カタログに収録されている約 4.11×10^6 (4,105,853) 個の遺伝子の塩基配列をアミノ酸配列に変換後、遺伝子機能データベースである KEGG に収録されている原核生物の遺伝子のアミノ酸配列に対して配列の相同性検索を実施した。その結果、約 2.72×10^6 (2,718,565) 個の遺伝子が、KEGG に収録されている遺伝子と対応したことが確認された。また、約 1.46×10^6 (1,460,503) 個の遺伝子が、KO と対応したことも確認された。遺伝子カタログに収録されている遺伝子の塩基配列に対してメタゲノムリードをマッピングすることで算出した遺伝子組成と、その遺伝子カタログの遺伝子と対応する KO を対応させることで、6081 個の KO を特徴量とした遺伝子機能組成を算出することができた。

3.3.5 生育速度の推定

外科的治療後に相対存在量が増加した 17 種の腸内細菌の生育速度を推定した。生育速度の推定には、その腸内細菌由来の多くのリードが必要であるため、一部のメタゲノムサンプルにおいて、特定の腸内細菌の生育速度を推定することができなかった。それぞれの腸内細菌において生育速度を推定することができたサンプル数を表 3-1 に示した。

大腸がん外科的治療前の患者と大腸がん外科的治療後の患者において、相対存在量とその生育速度に正の相関関係がある腸内細菌を探索した。その結果、それぞれの患者で、4 種の腸内細菌(*Parabacteroides merdae*, *Bifidobacterium longum*, *C. scindens*, *Ruminococcus gnavus*)の相対存在量とその生育速度に正の相関関係があることが確認された(相関係数 > 0.4)。そのため、これらの 4 種については、大腸がん外科的治療前の患者と比較して、大腸がん外科的治療後の患者で生育速度が高い可能性が示唆された。

表 3-1 腸内細菌の生育速度を推定した統計値

Species	Reference genome	N	Coverage (median)	GRiD (Pre, median)	GRiD (Post, median)	Cor (Pre)	P value (Pre)	Cor (Post)	P value (Post)
<i>Ruminococcus gnavus</i>	H5NKWBCX2_PG2320_551A1521_H1_L002.bin.10	711	5.65	1.14	1.26	0.797	7.60×10^{-20}	0.605	1.13×10^{-9}
<i>Clostridium scindens</i>	mix23_mix24-N711-S502_AAGAGGCA-CTCTCTAT.bin.30	586	0.204	1.17	1.45	0.690	1.04×10^{-10}	0.594	1.71×10^{-6}
<i>Bifidobacterium longum</i>	HFFY5BCX3_PG3817_269A1221_H1_L001.bin.2	669	3.30	1.23	1.24	0.634	2.66×10^{-10}	0.572	1.32×10^{-8}
<i>Parabacteroides merdae</i>	HC2YKBCX2_PG2516_11A1004_H1_L001.bin.10	709	9.73	1.13	1.17	0.414	9.23×10^{-5}	0.582	5.14×10^{-9}
<i>Clostridiales bacterium</i> VE202-14	10307-1-2_S9.bin.21	650	0.37	1.49	1.49	0.412	2.18×10^{-4}	0.0909	0.429
<i>Dialister invisus</i>	mix3_N703-N505_AGGCAGAA-GTAAGGAG.bin.2	343	0.26	1.00	1.21	0.331	0.0347	0.507	1.09×10^{-4}
<i>Clostridium clostridioforme</i>	HGKYNADXX_PG0662_263A0617_H1_L002.bin.6.fa	711	2.53	3.50	3.78	0.294	6.29×10^{-3}	-0.163	0.139
<i>Bifidobacterium dentium</i>	HGKY7ADXX_PG0662_178A0104_H1_L002.bin.11	662	1.78	1.09	1.10	0.278	0.0130	0.254	0.0207
<i>Bifidobacterium breve</i>	HFG22BCX3_PG3817_261A1106_H1_L001.bin.3	661	1.63	1.00	1.00	0.212	0.0621	0.401	1.72×10^{-4}
<i>Oscillibacter</i> sp. KLE 1728	H5MWWBCX2_PG2320_544A0321_H1_L002.bin.1	710	3.66	1.21	1.18	0.0612	0.578	0.259	0.0181
<i>Anaerostipes caccae</i>	BHCNJJADXX.N704-S503_TCCTGAGC-TATCCTCT.bin.15	437	0.400	1.57	1.64	0.0527	0.740	-0.0959	0.470
<i>Blautia hansenii</i>	10532-1-2_S9.bin.24.fa	708	1.81	1.00	1.00	0.0401	0.716	0.0145	0.898
<i>Eggerthella lenta</i>	HFFY5BCX3_PG3817_221A1005_H1_L001.bin.19	698	0.450	1.79	1.74	0.0210	0.849	-0.273	0.0125
<i>Clostridium innocuum</i>	HF7LNBCX3_PG3817_041A0318_H1_L002.bin.20	567	0.311	4.21	3.46	-0.125	0.305	-0.586	1.76×10^{-8}
<i>Bacteroides xylanisolvens</i>	mix1_N702-N505_CGTACTAG-GTAAGGAG.bin.20	714	15.7	1.74	1.81	-0.415	7.87×10^{-5}	-0.224	0.0396
<i>Bilophila wadsworthia</i>	HFFY5BCX3_PG3817_269A1221_H1_L001.bin.3	582	2.40	2.09	1.90	-0.748	1.48×10^{-13}	-0.659	4.00×10^{-10}
<i>Flavonifractor plautii</i>	HFFY5BCX3_PG3817_269A1221_H1_L001.bin.8	712	4.47	1.85	1.59	-0.769	8.99×10^{-18}	-0.737	1.33×10^{-15}

Reference genome: 生育速度の推定に用いたゲノム; Pre: 大腸がん外科的治療前の患者; Post: 大腸がん外科的治療後の患者

3.3.6 デオキシコール酸生成菌の探索

まず、22,547 個の再構築したゲノムから、デオキシコール酸の生成に必須である 6 つの *bai* 遺伝子(*baiB*, *baiCD*, *baiE*, *baiF*, *baiA*, *baiH*)を保有するゲノムを探索した。その結果、6 つの *bai* 遺伝子を保有するゲノムを 11 個特定した(表 3-2)。次に、この 11 個のゲノムの 6 つの *bai* 遺伝子の位置関係を調査した。位置関係を調査した結果、7 個のゲノムにおいて、6 つの *bai* 遺伝子が同じコンティグに存在していることが確認された。この 7 個のゲノムは、*baiI* 遺伝子以外の 7 個の *bai* 遺伝子を保有することも確認された(図 3-8)。

次に、この 7 個のゲノムの系統を推定した。その結果、5 個のゲノムは *C. scindens*、2 個のゲノムは *C. hiranonis* にアサインメントされた。この 2 種はデオキシコール酸を生成することが報告されているため[88]、本研究で探索したデオキシコール酸生成菌は新種ではないことが確認された。

最後に、再構築した全ての *C. scindens* と *C. hiranonis* のゲノムに、6 つの *bai* 遺伝子が隣接して存在するのかどうかを調査した(表 3-2)。*C. hiranonis* については、再構築した全てのゲノムで、6 つの *bai* 遺伝子が隣接して存在したことが確認された。一方で、*C. scindens* については、再構築した 13 個のゲノムのうち、4 つのゲノムには 6 つの *bai* 遺伝子が存在しないことが確認された。また、その 4 つのゲノムの 1 つ(HGKY7ADXX_PG0662_183A0202_H1_L002.bin.24)は、そのクオリティが高いことも確認された(ゲノムの完成率=95.02%, ゲノムの汚染率=0.765%)。この結果は、アセンブルでできなかったため *bai* 遺伝子が存在しなかった可能性が低いことを示唆している。以上のことから、*C. scindens* のデオキシコール酸の生成能は株毎に異なる可能性が示唆された。*C. scindens* の相対存在量が増加したことは、必ずしも *bai* オペロンの相対存在量が増加したことを保証できないため、遺伝子を定量することが重要であると考えられる。

外科的治療後にデオキシコール酸が増加したことに伴い、*C. scindens* や *C. hiranonis* が外科的治療後に *bai* オペロンを獲得した可能性も考えられた。そのため、同一患者由来の再構築した *C. scindens* と *C. hiranonis* のゲノムの *bai* オペロンの有無を調査することを試みたが、同一患者で *C. scindens* と *C. hiranonis* のゲノムを再構築することができなかった。今後、Oxford nanopore を用いた long-reads を用いたメタゲノム解析によって、*C. scindens* と *C. hiranonis* のゲノムを再構築し、同一の患者で、*C. scindens* や *C. hiranonis* が外科的治療後に *bai* オペロンを獲得したかどうか検証することで、胆汁酸と *bai* オペロンの獲得の関係性をより詳細に明らかにすることが期待される。

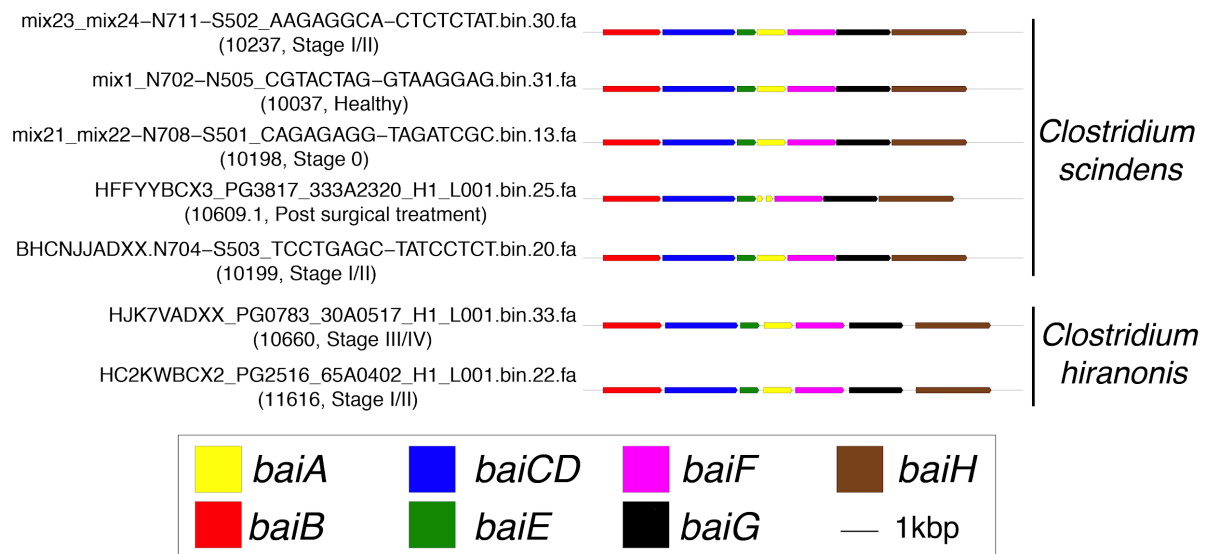


図 3-8 デオキシコール酸生成菌の *bai* オペロンを構成する遺伝子の構造
 それぞれの色は遺伝子の種類を表している。括弧の中の数字はメタゲノムサンプルとそのメ
 タゲノムサンプルのカテゴリーを表している。

表 3-2 再構築した *Clostridium scindens* と *Clostridium hiranonis* のゲノムの情報

MAGs	Species	Sample	Category	Bai gene	Arrangement	Completeness (%)	Contamination (%)
BHCNJADXX.N704-S503_TCCTGAGC-TATCCTCT.bin.20	<i>Clostridium scindens</i>	10199	Stage I/II, Pre-surgical treatment	O	O	90.11	0.643
HFFYYBCX3_PG3817_333A2320_H1_L001.bin.25	<i>Clostridium scindens</i>	10609.1	Post-surgical treatment	O	O	94.06	0
mix1_N702-N505_CGTACTAG-GTAAGGAG.bin.31	<i>Clostridium scindens</i>	10037	Healthy control	O	O	96.08	0.779
mix21_mix22-N708-S501_CAGAGAGG-TAGATCGC.bin.13	<i>Clostridium scindens</i>	10198	Stage 0	O	O	94.94	0.974
mix23_mix24-N711-S502_AAGAGGCA-CTCTCTAT.bin.30	<i>Clostridium scindens</i>	10237	Stage I/II	O	O	98.73	0.584
HC2KWBCX2_PG2516_65A0402_H1_L001.bin.22	<i>Clostridium hiranonis</i>	11616	Stage I/II, Pre-surgical treatment	O	O	88.69	0.582
HJK7VADXX_PG0783_30A0517_H1_L001.bin.33	<i>Clostridium hiranonis</i>	10660	Stage III/IV	O	O	98.6	0.699
H5MWWBCX2_PG2320_317A0116_H1_L002.bin.2	<i>Clostridium scindens</i>	11521	Healthy control	X	X	59.37	1.559
H5W27BCX2_PG2320_257A1908_H1_L002.bin.11	<i>Clostridium scindens</i>	11103	Stage I/II	X	X	68.04	0.487
HC2T7BCX2_PG2516_39A0704_H1_L001.bin.22	<i>Clostridium scindens</i>	11279	Stage I/II, Pre-surgical treatment	O	X	77.21	0
10307-1-2_S9.bin.38	<i>Clostridium scindens</i>	10307.1	Post-surgical treatment	O	X	70.25	0.584
10545-1-2_S10.bin.1	<i>Clostridium scindens</i>	10545.1	Post-ESD treatment	O	X	76.8	1.228
HF7LCBCX3_PG3817_106A0502_H1_L002.bin.30	<i>Clostridium scindens</i>	11002.1	Post-surgical treatment	X	X	83.74	0.584
HFG2FBCX3_PG3817_273A0218_H1_L001.bin.18	<i>Clostridium scindens</i>	11367.1	Post-surgical treatment	O	X	70.37	1.286
HGKY7ADXX_PG0662_183A0202_H1_L002.bin.24	<i>Clostridium scindens</i>	10545	Pre-ESD treatment	X	X	95.02	0.765

MAGs: 再構築したゲノム; Bai gene: デオキシコール酸の生成に必須な6つの*bai*遺伝子を保有している場合○; Arrangement: 遺伝子が隣接していた場合○

第4章 外科的治療前後における腸内環境の全体像比較

4.1 序論

本章では、85 人の大腸がん患者の外科的治療前後における腸内環境の全体像の比較手法やその結果について述べる。まず、大腸がん患者の腸内細菌の系統組成や、それらの遺伝子機能組成、代謝産物の組成が外科的治療前後で異なるのかどうかを明らかにするために、Bray-Curtis 距離を基盤としたサンプル間の類似性をそれぞれの組成毎に算出し、個人間の対応のない統計学的な手法と個人間の対応のある統計学的な手法を用いて、外科的治療前の患者と外科的治療後の患者のサンプル間の類似性をそれぞれ比較した。最後に、大腸がんに関連している可能性があるヒトゲノムの混入量(シーケンサーから得られたメタゲノムリードに含まれているヒト由来のリードの割合)が、外科的治療前後でどのように変化したのかを調査した。統計解析の結果、外科的治療後の患者の腸内環境は、外科的治療前の患者の腸内環境と比較して有意に異なることが確認された($P < 0.005$)。また、ヒトゲノムの混入量は大腸がんの進行に従ってその量が多くなり、外科的治療後にその量が有意に減少したことも確認された($P < 0.005$)。以上の結果から、大腸がんの外科的治療は、大腸がんを切除するだけでなく腸内環境を大きく変化させる可能性が示唆された。

4.2 手法

4.2.1 組成データに対する多次元圧縮法の適用

本節では、多次元圧縮法に用いたデータの取得方法とそれを用いた多次元圧縮法について説明する。

まず、多次元圧縮法に用いたデータの取得方法について説明する。各被験者の疫学データ、属レベルの腸内細菌叢の系統組成データ、腸内細菌の遺伝子機能組成データ、代謝産物の組成データを、2.2.8 節、3.2.3 節、3.2.2 節、2.2.4 節にてそれぞれ取得した。代謝産物の組成データについては、被験者毎に代謝産物の定量値の総和を算出後、その総和でそれぞれの代謝産物の定量値を割ることで、それぞれの代謝産物の定量値を相対値に補正した。

次に、多次元圧縮法について説明する。多次元圧縮法は、組成データ等に由来するサンプル間の類似性等を表した距離行列を2次元または3次元の空間に圧縮する手法である。この手法により、それぞれの次元におけるサンプル間の類似性を反映した各サンプルの座標を取得することができる。本研究では、それぞれの組成データにおける、大腸がん患者の外科的治療前後の類似性を可視化するために、それぞれの組成データに由来するサンプル間の距離行列に対して多次元圧縮法を適用し、2次元上にサンプル間の類似性を可視化した。Rのveganパッケージ(version 2.5-3)のmetaMDSを用いて、それぞれの組成におけるサンプル間の距離行列の算出し、それを多次元圧縮することで各被験者の2次元空間上の座標を取得した。パラメータはk=2、trymax=20、distance='bray'を用いた。2つのサンプル間の距離は、Bray-Curtis 距離によって算出した。

4.2.2 組成データの変化の統計学的な比較方法

まず、組成データの統計学的な比較に用いたデータの取得方法について説明する。各被験者の疫学データ、属レベルの腸内細菌叢の系統組成データ、腸内細菌の遺伝子機能組成データ、代謝産物の組成データを、2.2.8 節、3.2.3 節、3.2.2 節、2.2.4 節にてそれぞれ取得した。本研究では、健常者の時系列データを取得していないため、公開されているコホートの健常者の時系列データを統計学的な比較に用いた。Voigt et al.の健常者の腸内細菌叢の系統組成データと遺伝子機能組成データ、Nagata et al.の健常者の代謝産物の組成データを3.2.7 節、3.2.7 節、2.2.7 節にてそれぞれ取得した。代謝産物の組成データについては、被験者毎に代謝産物の定量値の総和を算出後、その総和でそれぞれの代謝産物の定量値を割ることで、それぞれの代謝産物の定量値を相対値に補正した。個人間の対応のない手法と個人間の対応のある手法を用いて、組成データの変化を統計学的に評価した。

個人間の対応のない統計学的な比較手法

大腸がん患者の外科的治療前後のそれぞれの組成データを、個人間の対応のない PERMANOVA(Permutational multivariate analysis of variance)検定を用いて比較した。PERMANOVA 検定は、Rのveganパッケージのadonisを用いて実施し、*P*値を取得した。パラメータは、method='bray'、permutation=9999を用いた。PERMANOVA 検定では、患者のラベル(外科的治療前、外科的治療後)をランダムに並び替えたときのラベル間の分散が、並び替えないときの分散と比較してどれくらい異なるのかを評価している。この検定によって得られた *P* 値の有意水準は0.005とした。

個人間の対応のある統計学的な比較手法

まず、個人間の対応のある統計学的な比較手法の評価指標である Bray-Curtis 距離の算出方法について説明する。大腸がん患者の外科的治療前後のそれぞれの組成を用いて、外科的治療前後間の Bray-Curtis 距離を患者毎に算出した。また、大腸がん患者の内視鏡治療前後のそれぞれの組成を用いて、内視鏡治療前後間の Bray-Curtis 距離も患者毎に算出した。さらに、公開されているコホートである Voigt et al.の健常者の腸内細菌叢の系統組成データと遺伝子機能組成データに対して、被験者の異なる時点で取得されたサンプル間の Bray-Curtis 距離も被験者毎に算出した。Voigt et al.のコホートでは、各被験者の異なる時点で取得されたデータが3つ以上存在したため、各被験者の異なる時点で取得されたサンプルの組み合わせが複数存在した。そのため、その全ての組み合わせの Bray-Curtis 距離を算出した。公開されているコホートである Nagata et al.の健常者の代謝産物の組成に対しても、被験者の異なる時点で取得されたサンプル間の Bray-Curtis 距離を被験者毎に算出した。これらの手法によって算出した Bray-Curtis 距離を以下の統計学的な比較に用いた。

外科的治療前後の患者毎に算出された Bray-Curtis 距離と内視鏡治療前後の患者毎に算出された Bray-Curtis 距離と、公開されている健常者の時系列データから算出された Bray-Curtis 距離をウィルコクソンの順位和検定を用いてそれぞれ比較した。ウィルコクソンの順位和検定は片側検定で実施し、R の coin パッケージの `wilcox_test` を用いて P 値を取得した。alternative のパラメータは `less` または `greater` を用いた。この検定によって得られた P 値の有意水準は 0.005 とした。

4.2.3 ヒトゲノムの混入量の統計学的な比較方法

本節では、ヒトゲノムの混入量の取得方法と、それらの比較に用いた統計学的な手法について説明する。

まず、ヒトゲノムの混入量の取得方法について説明する。3.2.1 節にて、シーケンサーから得られたメタゲノムリードから高精度なリード取得する過程で、各被験者のヒトゲノムの混入量(シーケンサーから得られたメタゲノムリード対するヒト由来のリードの割合)を算出した。各被験者のヒトゲノムの混入量と、それに含まれている各被験者のラベル(健常者、多発性腺腫、Stage 0 の大腸がん、Stage I/II の大腸がん、Stage III/IV の大腸がん、大腸がん外科的治療前の患者、大腸がん外科的治療後の患者)を対応させて、このデータを以下の統計学的な比較に用いた。

健常者と多段階発がんのそれぞれの状態の患者(多発性腺腫の患者、Stage 0 の大腸がんの患者、Stage I/II の大腸がんの患者、Stage III/IV の大腸がんの患者)のヒトゲノムの混入量を、ウィルコクソンの順位和検定を用いてそれぞれ比較した。ウィルコクソンの順位和検定は片側検定で実施し、R の coin パッケージの wilcox_test を用いて P 値を取得した。alternative のパラメータは less または greater を用いた。次に、大腸がん患者の外科的治療前後のヒトゲノムの混入量を個人間の対応のある検定であるウィルコクソンの符号順位検定を用いて比較した。ウィルコクソンの符号順位検定は片側検定で実施し、R の coin パッケージの wilcoxsign_test を用いて P 値を取得した。distribution のパラメータは、exact、alternative のパラメータは、less または greater を用いた。これらの検定によって得られた P 値の有意水準は 0.005 とした。

4.3 結果・考察

4.3.1 大腸がん患者の外科的治療前後の組成データの変化

大腸がん患者の腸内細菌の系統組成、それらの遺伝子機能組成、代謝産物の組成が外科的治療前後で異なるのかどうかを明らかにするために、個人間の対応のない統計学的手法と個人間の対応のある統計学的手法によって、患者の外科的治療前後の組成をそれぞれ比較した。

まず、個人間の対応のない統計学的手法によって患者の外科的治療前後のそれぞれの組成を比較した結果に着目した。外科的治療前の患者の腸内細菌の系統組成、遺伝子機能組成、代謝産物の組成を、外科的治療後の患者のそれらと比較した結果、全ての組成において、外科的治療前の患者と外科的治療後の患者は異なることが確認された($P=1.0\times 10^{-4}$, $P=1.0\times 10^{-4}$, $P=2.0\times 10^{-4}$, 図 4-1a, 図 4-1b, 図 4-1c)。次に、腸内細菌の系統組成、遺伝子機能組成、代謝産物の組成におけるサンプル間の類似性を多次元圧縮して可視化した結果に着目した。全ての組成データにおいて、外科的治療前の患者の座標の分布と外科的治療後の患者の座標の分布は異なることが確認された(図 4-1a, 図 4-1b, 図 4-1c)。一方で、外科的治療後の一部の患者は、外科的治療前の患者と同じ位置に可視化されることも確認されたため、個人間の対応のある統計学的手法を用いて評価することが重要であると考えられる。

最後に、個人間の対応のある統計学的手法によって大腸がん患者の外科的治療前後の腸内細菌の系統組成、遺伝子機能組成、代謝産物の組成を比較した結果に着目した。腸内細菌の系統組成、それらの遺伝子機能組成、代謝産物の組成における大腸がん患者の外科

的治療前後の Bray-Curtis 距離は、公開されている健常者の Bray-Curtis 距離と比較して有意に高いことが確認された($P=1.42 \times 10^{-15}$, $P=1.25 \times 10^{-8}$, $P=1.48 \times 10^{-5}$, 図 4-2a, 図 4-2b, 図 4-3c)。また、腸内細菌の系統組成における、大腸がん患者の外科的治療前後の Bray-Curtis 距離は、大腸がん患者の内視鏡治療前後の Bray-Curtis 距離よりも高いことが確認された($P=0.0263$, 図 4-2a)。以上の結果から、大腸がんの外科的治療は、大腸がんを切除するだけでなく腸内環境を大きく変化させる可能性が示唆された。

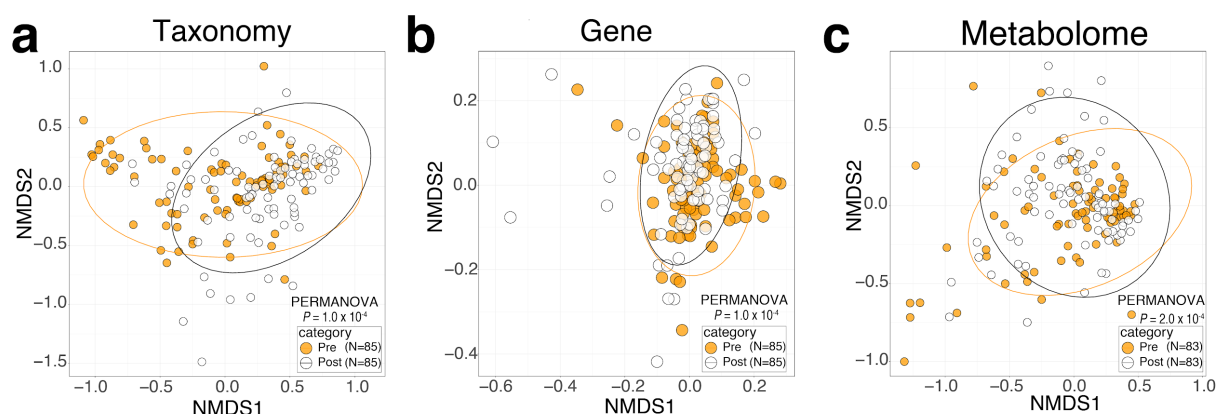


図 4-1 組成データにおける大腸がん患者の外科治療前後の類似性を表した散布図
それぞれの丸はそれぞれの組成におけるサンプル間の類似性を反映している(a: 腸内細菌の系統組成データ; b: 腸内細菌の遺伝子機能組成データ; c: 代謝産物の組成データ)。楕円の線は外科的治療前の患者と外科的治療後の患者の座標の分布を表している(オレンジ色: 外科的治療前の患者; 白色: 外科的治療後の患者)。外科的治療前の患者と外科的治療後の患者のそれぞれの組成を PERMANOVA 検定によって比較して P 値を取得した。

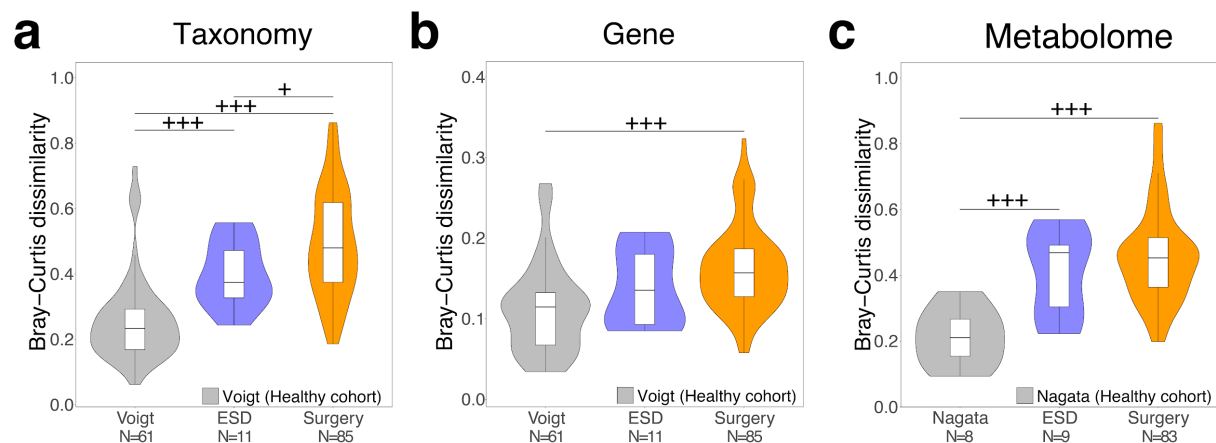


図 4-2 大腸がん外科的治療前後の患者の Bray-Curtis 距離を表した箱ひげ図

それぞれのバイオリン図と箱ひげ図は、それぞれの被験者毎に算出された Bray-Curtis 距離を示している(a: 腸内細菌の系統組成データ; b: 腸内細菌の遺伝子機能組成データ; c: 代謝産物の組成データ)。色は被験者のグループを反映している(灰色: 公開されているコホートの健常者; 青色: 内視鏡治療前後の患者; オレンジ色: 外科的治療前後の患者)。それぞれの被験者のグループのサンプル数(N)を箱ひげ図の下に示している。それぞれのバイオリン図と箱ひげ図の外れ値を省略して可視化している。統計学的に Bray-Curtis 距離が高かったことが確認された場合、以下のように表記している($P < 0.005$: +++; $P < 0.01$: ++; $P < 0.05$: +)。

4.3.2 大腸がん患者の外科的治療前後のヒトゲノムの混入量の変化

大腸がんに関係している可能性があるヒトゲノムの混入量が、大腸がん患者の外科的治療前後でどのように変化するかを調査した。まず、健常者のヒトゲノムの混入量と多段階発がんのいずれかの状態の患者(多発性腺腫の患者、Stage 0 の大腸がん患者、Stage I/II の大腸がん患者、Stage III/IV の大腸がん患者)のヒトゲノムの混入量をそれぞれ比較した結果に着目した。ヒトゲノムの混入量は、健常者と比較して、Stage 0 の大腸がん患者、Stage I/II の大腸がん患者、Stage III/IV の大腸がん患者で有意に高いことが確認された($P=1.75 \times 10^{-3}$, $P=1.63 \times 10^{-5}$, $P=2.90 \times 10^{-7}$, 図 4-3 の左側の箱ひげ図)。また、ヒトゲノムの混入量は、大腸がんの進行に従ってその量が多くなる傾向にあることも確認された。これらの結果から、ヒトゲノムの混入量は大腸がんの進行に従って変化する腸内環境を反映している可能性が示唆された。次に、大腸がん患者の外科的治療前後のヒトゲノムの混入量を比較した結果に着目した。ヒトゲノムの混入量は、外科的治療後に有意に減少したことが確認された($P=2.47 \times 10^{-4}$, 図 4-3 の右側の箱ひげ図)。この結果は、大腸がんの外科的治療により、大腸がんの進行に起因する腸内環境が改善されたことを反映している可能性があると考えられる。

ヒトゲノムの混入量 (上限値 : 20)

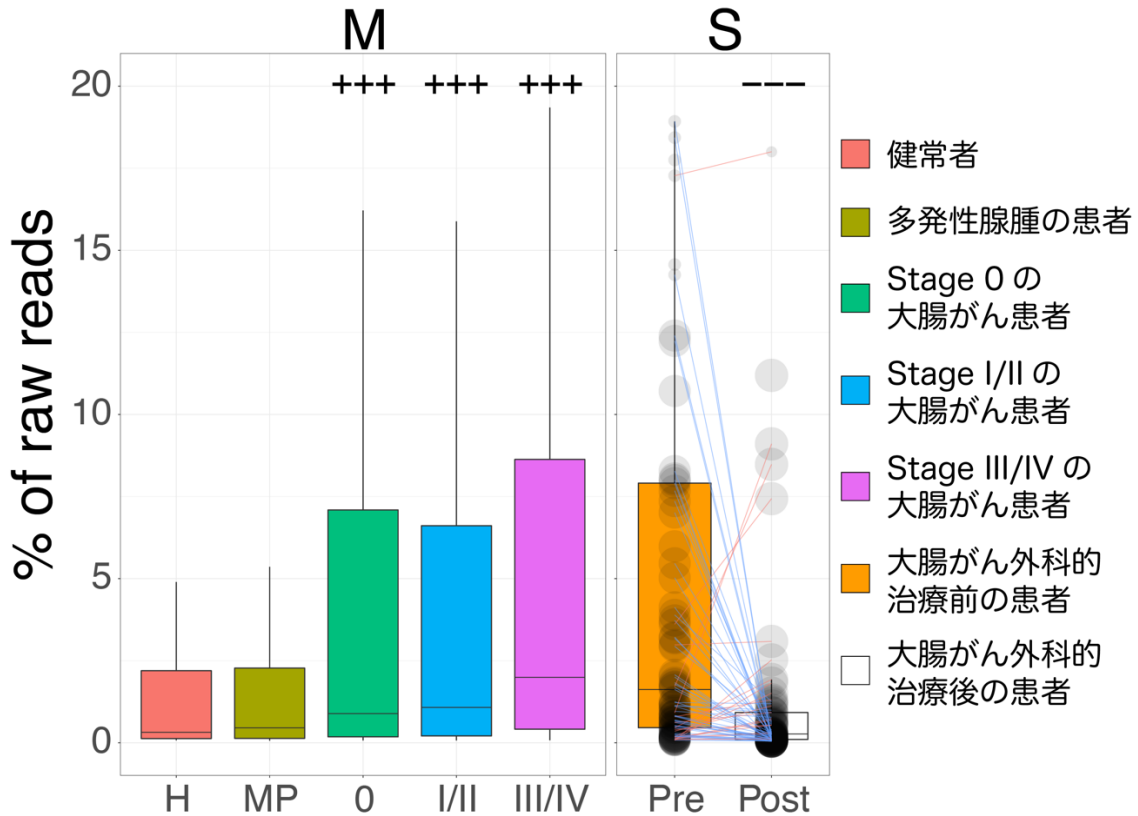


図 4-3 大腸がん外科的治療前後の患者のヒトゲノムの混入量を示した箱ひげ図

左側の箱ひげ図は、健常者と多段階発がんのいずれかの状態の患者のヒトゲノムの混入量を示している。右側の箱ひげ図は、大腸がん患者の外科的治療前後のヒトゲノムの混入量を示している。左側の箱ひげ図は外れ値を省略している。右側の箱ひげ図の丸の大きさは、外科的治療前後のそれぞれの患者におけるそれらの分布を反映している。右側の箱ひげ図の線の色は、同一患者の外科的治療前後の変化を表しており、外科的治療後に増加した場合は赤色、外科的治療後に減少した場合は青色、外科的治療前後で増加も減少も確認されなかった場合は緑色で可視化している。左側の箱ひげ図と右側の箱ひげ図は、上限値以下のサンプルのみを可視化している。統計学的に増加またはその量が高かったことの場合、以下のように表記している ($P < 0.005$: +++; $P < 0.01$: ++; $P < 0.05$: +)。統計学的に減少またはその量が少なかったことが確認された場合、以下のように表記している ($P < 0.005$: ---; $P < 0.01$: --; $P < 0.05$: -)。

第5章 外科的治療前後における腸内環境を構成する特徴量の変動比較

5.1 序論

本章では、大腸がん患者の外科的治療前後で変化する腸内細菌やそれらの遺伝子、代謝産物を特定する手法やその結果について述べる。まず、大腸がん患者の外科的治療前後の腸内細菌やその遺伝子の相対存在量と代謝産物の量を、個人間の対応のあるウィルコクソンの符号順位検定を用いて統計学的に比較した。次に、健常者のそれらと、多発性腺腫の患者、Stage 0 の大腸がん患者、Stage I/II の大腸がん患者、Stage III/IV の大腸がん患者のそれらをウィルコクソンの順位和検定を用いてそれぞれ比較し、大腸がん外科的治療前後で変化したそれらが大腸がんに関連しているのかを調査した。これらの統計解析の結果、Stage III/IV の患者と健常者の分類に寄与しているバイオマーカーである *F. nucleatum* や *Parvimonas micra* 等の腸内細菌の相対存在量や、アミノ酸の 1 種であるセリン等の代謝産物の量は外科的治療後に有意に減少したことが確認された($P < 0.005$)。一方で、コール酸や発がん性が報告されているデオキシコール酸等の胆汁酸の量や、コール酸からデオキシコール酸に変換する遺伝子(*bai* オペロン)と、その変換を担う腸内細菌(*C. scindens*)の相対存在量は、外科的治療後に有意に増加したことが確認された(図 5-1, $P < 0.005$)。また、*C. scindens* の推定された生育速度も外科的治療後の患者で高いことも確認された($P < 0.05$)。以上の結果から、大腸がんの進行に起因する腸内環境は外科的治療により改善されたが、腸内細菌に由来する大腸がんのリスクが外科的治療後も高い可能性が示唆された。そのため、外科的治療後の患者の腸内環境を対象とした介入の必要性が示唆された。

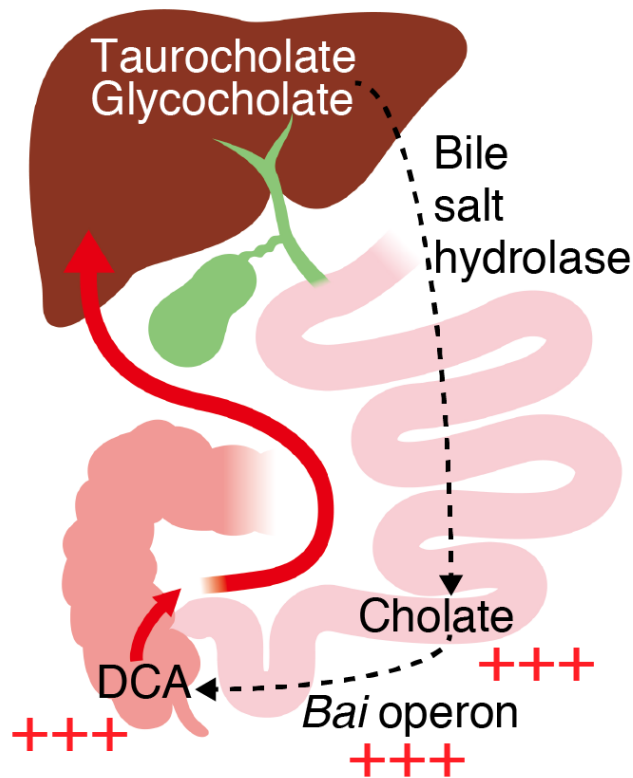


図 5-1 大腸がん外科的治療前後の患者の胆汁酸代謝の変化の概略図

図中の代謝産物はその生成された位置に配置されている。コール酸は肝臓で生成されるが、すぐにグリシンもしくはタウリンと結合し、グリココール酸(Glycocholate)、タウロコール酸(Taurocholate)にそれぞれ変換されるため、小腸に配置されている。黒色の矢印は、腸内細菌による変換を表している。赤い矢印は胆汁酸の再吸収の流れを表している。統計学的に有意に増加した場合($P < 0.005$)、+++と示した。

5.2 手法

5.2.1 大腸がん外科的治療前後で変化する特徴量の特定

本節では、統計学的な比較に用いたデータの取得方法と、その比較に用いた手法について説明する。

まず、統計解析するためのデータの取得方法について説明する。大腸がん患者の外科的治療前後の疫学データ、種レベルの腸内細菌叢の系統組成データ、腸内細菌の推定された生育速度のデータ、腸内細菌の遺伝子機能組成データ、*bai* オペロンの組成データ、代謝産物の組成データを、2.2.8 節、3.2.3 節、3.2.4 節、3.2.2 節、3.2.5 節、2.2.4 節にてそれぞれ

れ取得した。それぞれのデータとそのデータに含まれている各患者のラベル(外科的治療前、外科的治療後)を対応させて、これらのデータを以下の統計学的な比較に用いた。

大腸がん外科的治療前後で変化する腸内細菌やその遺伝子、代謝産物を特定するために、個人間の対応のある検定であるウィルコクソンの符号順位検定を用いて、大腸がん患者の外科的治療前後のそれらの相対存在量や量を比較した。また、大腸がん患者の外科的治療前後の便中の総胆汁酸の量をウィルコクソンの符号順位検定を用いて比較した。ウィルコクソンの符号順位検定は片側検定で実施し、Rのcoinパッケージのwilcoxsign_testを用いてP値を取得した。distributionのパラメータは、exact、alternativeのパラメータは、lessまたはgreaterを用いた。生育速度の推定には、その腸内細菌由来のメタゲノムリードが必要であるため、特定のサンプルで相対存在量が少ない腸内細菌の生育速度を推定することができなかった。そのため、推定された生育速度は、個人間の対応のないウィルコクソンの順位和検定を用いて外科的治療前の患者と外科的治療後の患者を比較した。ウィルコクソンの順位和検定は片側検定で実施し、Rのcoinパッケージのwilcox_testを用いてP値を取得した。alternativeのパラメータはlessまたはgreaterを用いた。これらの検定によって得られたP値の有意水準は0.005とした。

それぞれの腸内細菌や代謝産物が外科的治療前後でどの程度変化したのかを可視化するためにfold changeを算出した。Fold changeは患者毎に、外科的治療後の腸内細菌の相対存在量や代謝産物の量を、外科的治療前のそれらの相対存在量や量で割ることで算出した。患者の外科的治療前または外科的治療後のそれらの相対存在量や量が0であった場合は、fold changeを算出することができなかった。そのため、その患者のfold changeは0としてfold changeの平均値を算出した。

5.2.2 大腸がんと関係する特徴量の特定

本節では、統計学的な比較に用いたデータの取得方法と、その比較に用いた手法について説明する。

まず、統計解析するためのデータの取得方法について説明する。健常者と多段階がんのいずれかの状態の患者(多発性腺腫の患者、Stage 0の大腸がん患者、Stage I/IIの大腸がん患者、Stage III/IVの大腸がん患者)の疫学データ、種レベルの腸内細菌叢の系統組成データ、腸内細菌の推定された生育速度のデータ、腸内細菌の遺伝子機能組成データ、baiオペロンの組成データ、代謝産物の組成データを、2.2.8節、3.2.3節、3.2.4節、3.2.2

節、3.2.5 節、2.2.4 節にてそれぞれ取得した。それぞれのデータとそのデータに含まれている各被験者のラベルを対応させて、以下の統計学的な比較に用いた。

大腸がんと関係する腸内細菌やその遺伝子、代謝産物を特定するために、ウィルコクソンの順位和検定を用いて、健常者のそれらと多段階発がんのいずれかの状態の患者のそれらをそれぞれ比較した。ウィルコクソンの順位検和検は片側検定で実施し、R の coin パッケージの `wilcox_test` を用いて P 値を取得した。alternative のパラメータは less または greater を用いた。これらの検定によって得られた P 値の有意水準は 0.005 とした。

5.3 結果・考察

5.3.1 腸内細菌の系統組成の比較

大腸がん患者の外科的治療前後の腸内細菌の相対存在量を、個人間の対応のあるウィルコクソンの符号順位検定を用いて比較した。その結果、外科的治療後に有意に減少した 90 種の腸内細菌と、外科的治療後に有意に増加した 24 種の腸内細菌を特定した($P < 0.005$, 図 5-2)。

まず、外科的治療後にその相対存在量が有意に減少した 90 種の腸内細菌に着目した(付録 3)。ヒトゲノムの混入量は、大腸がんの進行に従ってその量が多くなり、外科的治療前と比較して外科的治療後にその量が有意に減少したことが確認されたため、大腸がんの進行に関与している腸内細菌の相対存在量も外科的治療前と比較して外科的治療後に減少する可能性があるという仮説を立てた。その仮説を検証するために、相対存在量が減少した腸内細菌と大腸がんの関連性を調査した。腸内細菌と大腸がんの関連性を明らかにするために、健常者の腸内細菌の相対存在量と多段階発がんのいずれかの状態の患者(多発性腺腫の患者、Stage 0 の大腸がん患者、Stage I/II の大腸がん患者、Stage III/IV の大腸がん患者)の腸内細菌の相対存在量をそれぞれ比較した。その結果、90 種の腸内細菌が健常者と比較してその量が有意に多いことが確認された。多段階発がんのいずれかの状態の患者で健常者と比較してその相対存在量が多く、外科的治療前の患者と比較して外科的治療後の患者でその相対存在量が減少した腸内細菌は 14 種であった。この 14 種の腸内細菌のうち、7 種の腸内細菌(*Parvimonas micra* [ref_mOTU_v2_1145], *Gemella morbillorum* [ref_mOTU_v2_4513], *Fusobacterium nucleatum* subsp. *animalis* [ref_mOTU_v2_0776], *Peptostreptococcus stomatis* [ref_mOTU_v2_4614], unknown *Dialister* [meta_mOTU_v2_5867], unknown *Peptostreptococcaceae* [meta_mOTU_v2_5742], *Solobacterium moorei*

[ref_mOTU_v2_0531])は、いくつかのコホート間で共通した大腸がんのバイオマーカーとして報告されている[9]。この7種の腸内細菌は、健常者と比較して Stage I/II と Stage III/IV の大腸がん患者でその相対存在量が有意に多く、外科的治療前と比較して外科的治療後にその相対存在量が有意に減少したことが確認された($P < 0.005$, 図 5-3a-g)。先行研究において、4種の腸内細菌(*P. micra*, *G. morbillorum*, *F. nucleatum*, *P. stomatis*)は、Stage III/IV の大腸がん患者と健常者の分類に寄与しているバイオマーカーとして報告されている[4]。また、*F. nucleatum* は、大腸がん組織に多く存在し[89]、大腸がんの細胞の増殖を促進すること[90]が報告されている。以上のことから、これらの腸内細菌の相対存在量が外科的治療後に減少したことも大腸がんの外科的治療により大腸がんの進行に起因する腸内環境が改善されたことを反映しているという仮説を支持していると考えられる。

次に、外科的治療後にその相対存在量が有意に増加した24種の腸内細菌に着目した(付録4)。この24種の腸内細菌と腸疾患の関係性について調査した結果、*R. gnavus*、*B. wadsworthia*、*C. scindens*が大腸がんの発症と関係している可能性がある代謝産物の生成に関与していることが確認された(図 5-3h の右側の箱ひげ図, 図 5-3i の右側の箱ひげ図, 図 5-3j の右側の箱ひげ図)。先行研究において、*C. scindens* は、コール酸を発がん性が報告されているデオキシコール酸に変換する機能を担うことが報告されている[88]。*B. wadsworthia* は発がん性が報告されている硫化水素の生成菌であり、胆汁酸の1種であるタウロコール酸から生成されるタウリンを栄養源として生育することが報告されている[91]。*R. gnavus* は、デオキシコール酸を3 β -ヒドロキシデオキシコール酸に変換する機能を担うことも報告されている[92]。以上のことから、これらの腸内細菌の変化は外科的治療によって胆汁酸代謝が変化したことを反映している可能性が示唆された。

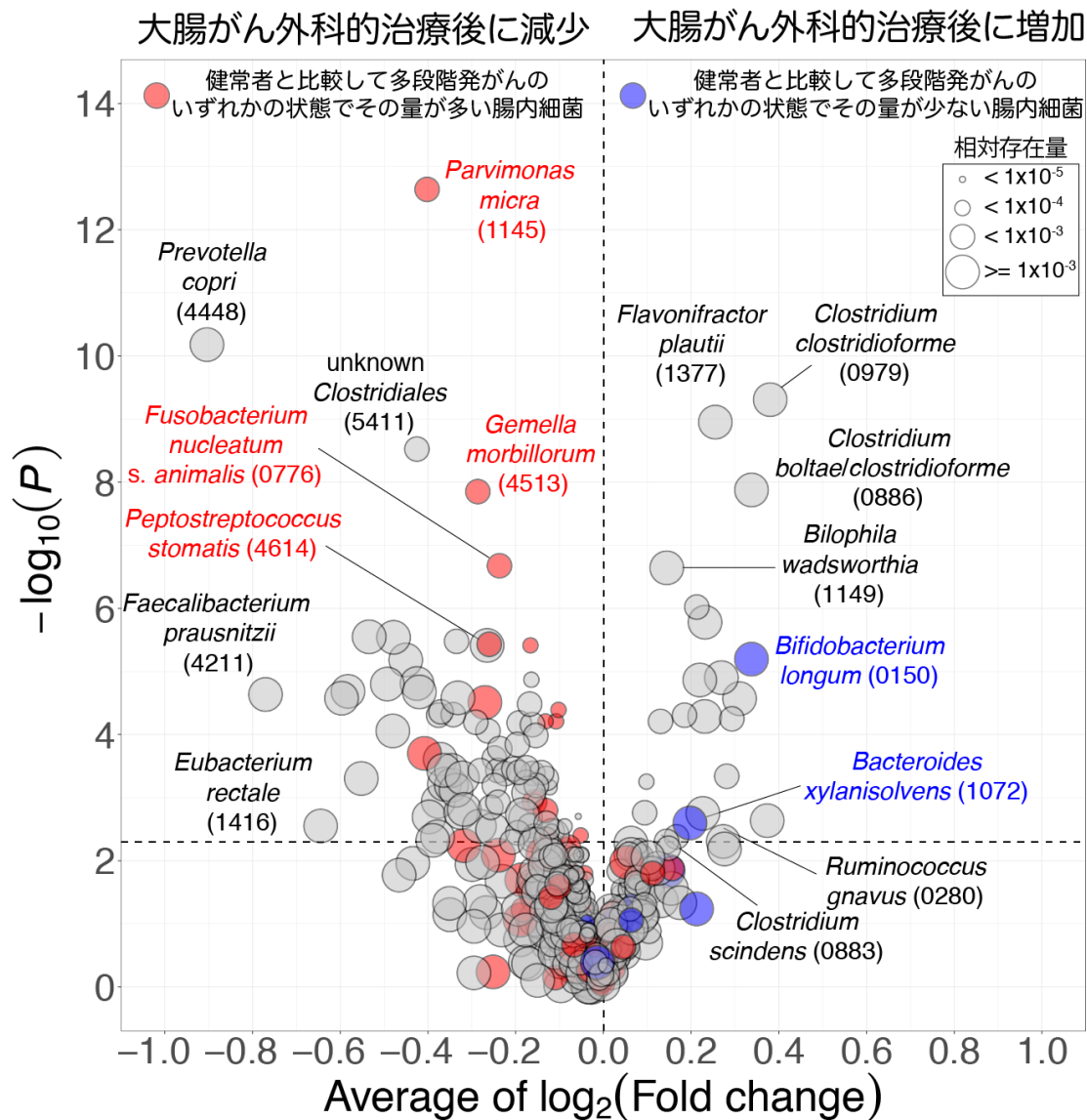


図 5-2 大腸がん患者の外科的治療前後の腸内細菌の相対存在量を比較した結果
縦軸は、片側のウィルコクソンの符号順位検定を用いて取得した P 値を $-\log_{10}$ で変換した値を示している。横軸は、85 人の大腸がん患者毎に算出した外科的治療前後の fold change の平均値を示している。横の点線は、 P 値の統計学的な有意水準である 0.005 を $-\log_{10}$ で変換した値を示している。それぞれの丸の大きさは大腸がん患者の外科的治療前後の腸内細菌の相対存在量の平均値を反映している(凡例参照)。赤色の丸は、健常者と比較して多段階発がんのいずれかの状態の患者でその相対存在量が有意に多いことを示している($P < 0.005$)。青色の丸は、健常者と比較して多段階発がんのいずれかの状態の患者でその相対存在量が有意に少ないことを示している($P < 0.005$)。それぞれの腸内細菌の括弧の中の番号は mOTU2 profiler の OTU の番号を示している。

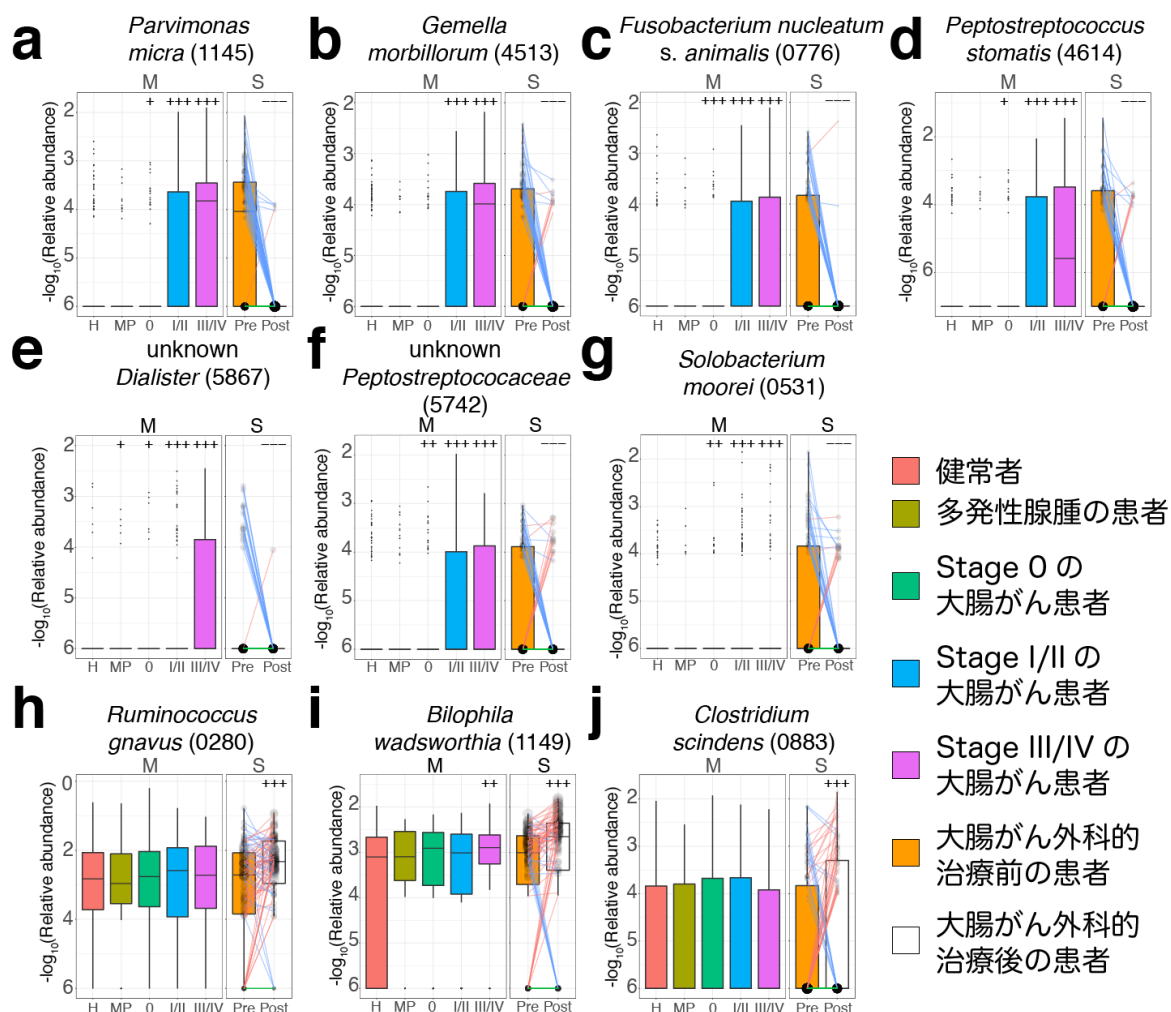


図 5-3 大腸がん外科的治療前後の患者の腸内細菌の相対存在量を示した箱ひげ図

左側の箱ひげ図は、健康者と多段階発がんのいずれかの状態の患者の腸内細菌の相対存在量を示している。右側の箱ひげ図は、大腸がん患者の外科的治療前後の腸内細菌の相対存在量を示している。それぞれの箱ひげ図の縦軸は、 $-\log_{10}$ で変換した腸内細菌の相対存在量を示している。腸内細菌の相対存在量が 0 であった場合は、その相対存在量を 1×10^{-6} に置換後、 $-\log_{10}$ で変換して可視化した。右側の箱ひげ図の線の色は同一患者での外科的治療前後の変化を表しており、外科的治療後に増加した場合は赤色、外科的治療後に減少した場合は青色、外科的治療前後で増加も減少も確認されなかった場合は緑色で可視化している。右側の箱ひげ図の丸の大きさは、外科的治療前後のそれぞれの患者におけるそれらの分布を反映している。統計学的に増加またはその量が多かったことが確認された場合、以下のように表記している ($P < 0.005$: +++; $P < 0.01$: ++; $P < 0.05$: +)。統計学的に減少またはその量が少なかったことが確認された場合、以下のように表記している ($P < 0.005$: ---; $P < 0.01$: --; $P < 0.05$: -)。それぞれの腸内細菌の括弧の中の番号は mOTU2 profiler の OTU の番号を示している。

5.3.2 腸内細菌の生育速度の比較

大腸がん外科的治療後に相対存在量が増加した 17 種の腸内細菌の生育速度を推定した。外科的治療前の患者と外科的治療後の患者の 17 種の腸内細菌の推定された生育速度をウィルコクソンの順位和検定を用いて比較した。その結果、外科的治療前の患者と比較して外科的治療後の患者で推定された生育速度が高い 4 つの腸内細菌を特定した($P < 0.05$, 図 5-4 の箱ひげ図)。この 4 つの腸内細菌のうち、大腸がん外科的治療前後の患者において、*C. scindens* と *R. gnavus* の推定された生育速度とそれらの相対存在量は正の相関関係にあることが確認された(詳細な結果は 3.3.5 節を参照)。これらの結果から、大腸がん外科的治療前の患者と比較して大腸がん外科的治療後の患者で、*C. scindens* と *R. gnavus* の細菌数も増加した可能性が示唆された。細菌の生育速度はその細菌の活性を反映している可能性がある[78]が、一概にそうとは限らないため、その細菌の遺伝子の発現やその細菌が生成する代謝産物も定量することが重要であると考えられる。

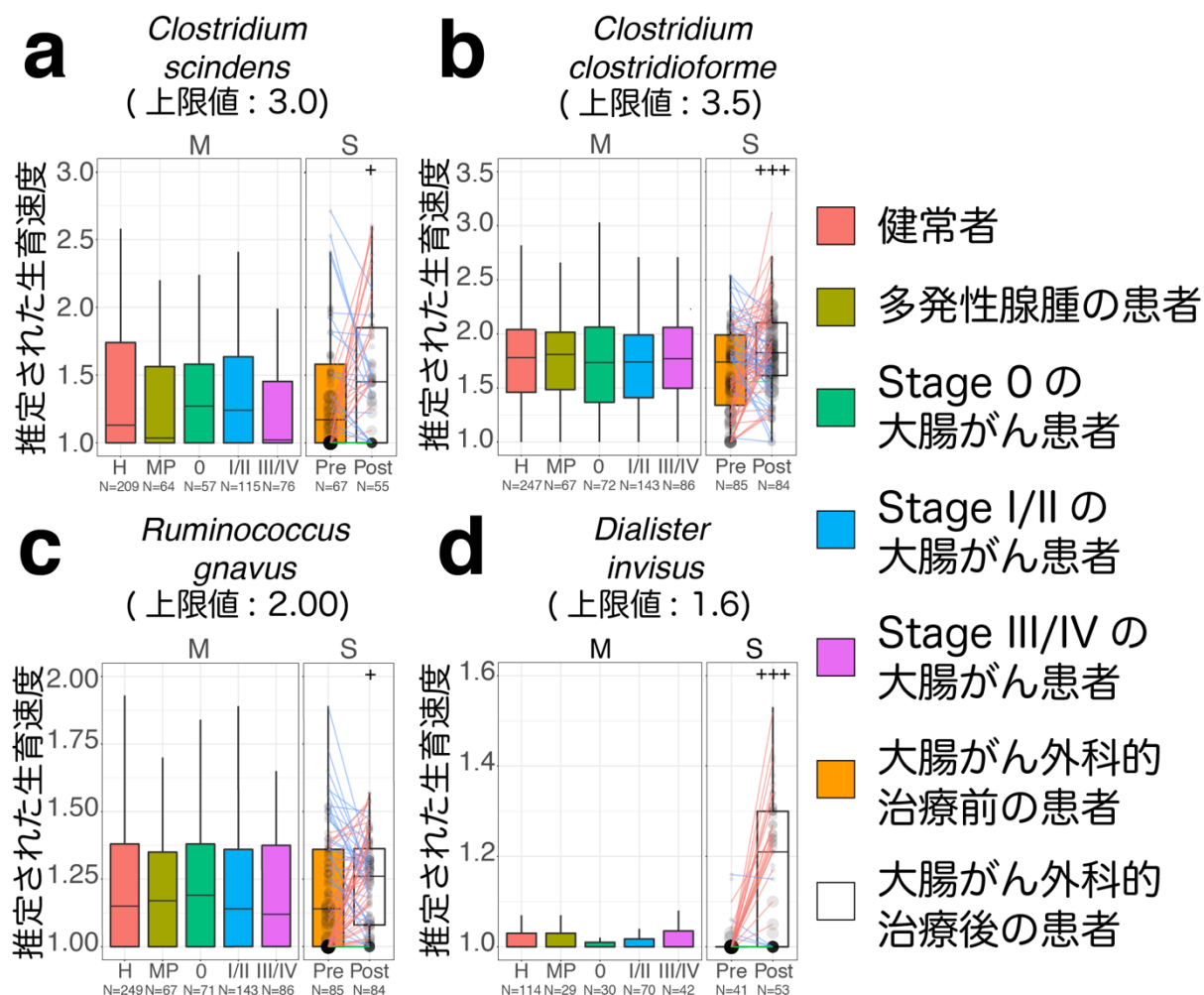


図 5-4 大腸がん外科的治療前後の患者の腸内細菌の生育速度を示した箱ひげ図

左側の箱ひげ図は、健常者と多段階発がんのいずれかの状態の患者の腸内細菌の推定された生育速度を示している。右側の箱ひげ図は、大腸がん患者の外科的治療前後の腸内細菌の推定された生育速度を示している。左側の箱ひげ図は外れ値を省略している。右側の箱ひげ図の丸の大きさは、外科的治療前後のそれぞれの患者におけるそれらの分布を反映している。右側の箱ひげ図の線の色は、同一患者での外科的治療前後の変化を表しており、外科的治療後に増加した場合は赤色、外科的治療後に減少した場合は青色、外科的治療前後で増加も減少も確認されなかった場合は緑色で可視化している。左側の箱ひげ図と右側の箱ひげ図は、上限値以下のサンプルのみを可視化している。統計学的に推定された生育速度が高いことが確認された場合、以下のように表記している($P < 0.005$: +++; $P < 0.01$: ++; $P < 0.05$: +)。生育速度を推定することができたサンプル数(N)を箱ひげ図の下に示している。

5.3.3 腸内細菌の代謝産物の比較

大腸がん患者の外科的治療前後の代謝産物の量を、個人間の対応のあるウィルコクソンの符号順位検定を用いて比較した。その結果、外科的治療後に有意に減少した 43 個の代謝産物と外科的治療後に有意に増加した 17 個の代謝産物を特定した($P < 0.005$, 図 5-5)。

まず、外科的治療後にその量が有意に減少した 43 個の代謝産物に着目した(付録 5)。アミノ酸の 1 種であるセリンやアミノ酸関連代謝産物である Gly-Leu や Urocanate、N,N-Dimethylglycine は、健常者と比較して Stage I/II と Stage III/IV の大腸がん患者でその量が有意に多く、外科的治療前と比較して外科的治療後でその量が有意に減少したことが確認された($P < 0.005$, 図 5-6a, 図 5-6b, 図 5-6c, 図 5-6d)。これらの代謝産物は、Stage III/IV の大腸がん患者と健常者の分類に寄与しているバイオマーカーとして報告されている[4]。また、セリンは大腸がんの細胞増殖を促すピルビン酸を生成するための供給源であり[93]、粘膜組織と比較して大腸がん組織でその量が多いことも報告されている[26,27]。以上のことから、これらの代謝産物が外科的治療後に減少したことも大腸がんの外科的治療により大腸がんの進行に起因する腸内環境が改善されたことを反映しているという仮説を支持していると考えられる。

次に、外科的治療後にその量が有意に増加した 17 個の代謝産物に着目した(付録 6)。コール酸やデオキシコール酸等の胆汁酸の量は、外科的治療前と比較して外科的治療後に有意に増加したことが確認された($P=4.71 \times 10^{-7}$, $P=2.00 \times 10^{-5}$, 図 5-6e の箱ひげ図, 図 5-6f の箱ひげ図)。先行研究において、デオキシコール酸は発がん性を有することが報告されており[31]、大腸がんの発症に関与している可能性があることが報告されている[32]。原発の大腸がんとは異なる大腸がんである異時性大腸がんの発症メカニズムは、多段階発がんにおける大腸がんの発症メカニズムと異なる可能性があるが、外科的治療後の患者においてデオキシコール酸の量が多いことは、外科的治療後の患者は大腸がんのリスクが高いという臨床学的な知見[37,38]と整合性があることが確認された。腸内細菌がデオキシコール酸を生成するため、大腸がん外科的治療後の患者は腸内細菌に由来する大腸がんの発症リスクが高い可能性が示唆された。そのため、腸内環境を対象とした介入の必要性が示唆された。

最後に、大腸がん患者の外科的治療前後のタウロコール酸、グリココール酸、総胆汁酸の量と血中の総コレステロール値をそれぞれ比較した結果に着目した。大腸がん患者の外科的治療前後の血中の総コレステロール値を比較した結果、それらに統計学的な有意な差は確認されなかった($P=0.128$)。先行研究において、コレステロールは主に肝臓で生成され、タウロコール酸やグリココール酸等の胆汁酸の生成に用いられることが報告されている[94]。以上のことから、大腸がん外科的治療後の患者の肝臓における、タウロコール酸やグ

リココール酸等の胆汁酸の生成量は外科的治療前と比較して変化していないことが示唆された。一方で、総胆汁酸の量は、外科的治療前と比較して外科的治療後に有意に増加したことが確認された($P=1.86 \times 10^{-7}$, 図 5-6g の箱ひげ図)。先行研究において、肝臓で生成されたタウロコール酸やグリココール酸等の胆汁酸は十二指腸に輸送された後、これらの胆汁酸の 95%は回腸末端で再吸収され、再吸収されなかった残りの 5%の胆汁酸は大腸に流入後、便として排出されることが報告されている[95]。以上のことから、大腸がん外科的治療後の患者の肝臓における胆汁酸の生成量は変化していない可能性が高いが、便から排出される総胆汁酸の量が増加したため、これらの患者において胆汁酸の再吸収障害が生じた可能性が示唆された。先行研究において、右側大腸がん外科的治療後の患者は外科的治療による大腸の再建により胆汁酸の再吸収障害が生じる可能性があることが報告されている[96]ため、大腸がん患者を大腸の再建法毎に層別化し、胆汁酸が変化したメカニズムを解明することが重要であると考えられる。大腸がん患者の外科的治療前後のタウロコール酸とグリココール酸の量を比較した結果、それらに統計学的な有意な差は確認されなかった($P=0.284$, $P=0.101$, 図 5-5h の箱ひげ図, 図 5-6i の箱ひげ図)。コール酸やデオキシコール酸等の腸内細菌の代謝に関係する胆汁酸の量は外科的治療前と比較して、外科的治療後に増加したが、ヒトの肝臓で生成されるタウロコール酸やグリココール酸等の胆汁酸の量は大腸がん患者の外科的治療前後で変化していなかったため、これらの胆汁酸代謝を担う腸内細菌の遺伝子の量を詳細に調査する必要があると考えられる。

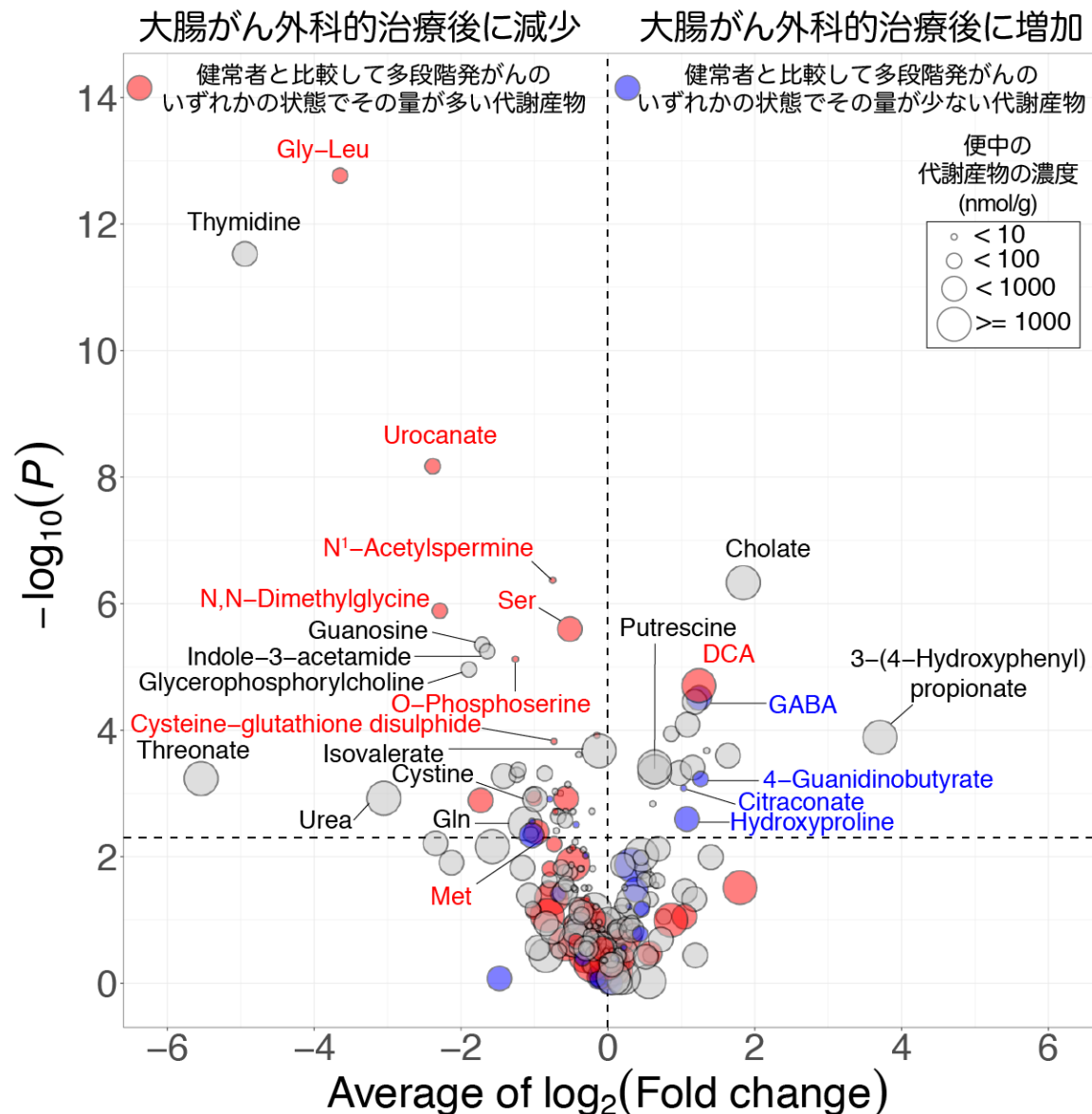


図 5-5 大腸がん外科的治療前後の患者の代謝産物の量を比較した結果

縦軸は、片側のウィルコクソンの符号順位検定によって取得した P 値を $-\log_{10}$ で変換した値を示している。横軸は、83 人の大腸がん患者毎に算出した外科的治療前後の fold change の平均値を示している。横の点線は、 P 値の統計学的な有意水準である 0.005 を $-\log_{10}$ で変換した値を示している。それぞれの丸の大きさは大腸がん外科的治療前後の患者の代謝産物の量の平均値を反映している(凡例参照)。赤色の丸は、健常者と比較して多段階発がんのいずれかの状態の患者でその量が有意に多いことを示している。青色の丸は、健常者と比較して多段階発がんのいずれかの状態の患者でその量が有意に少ないことを示している。

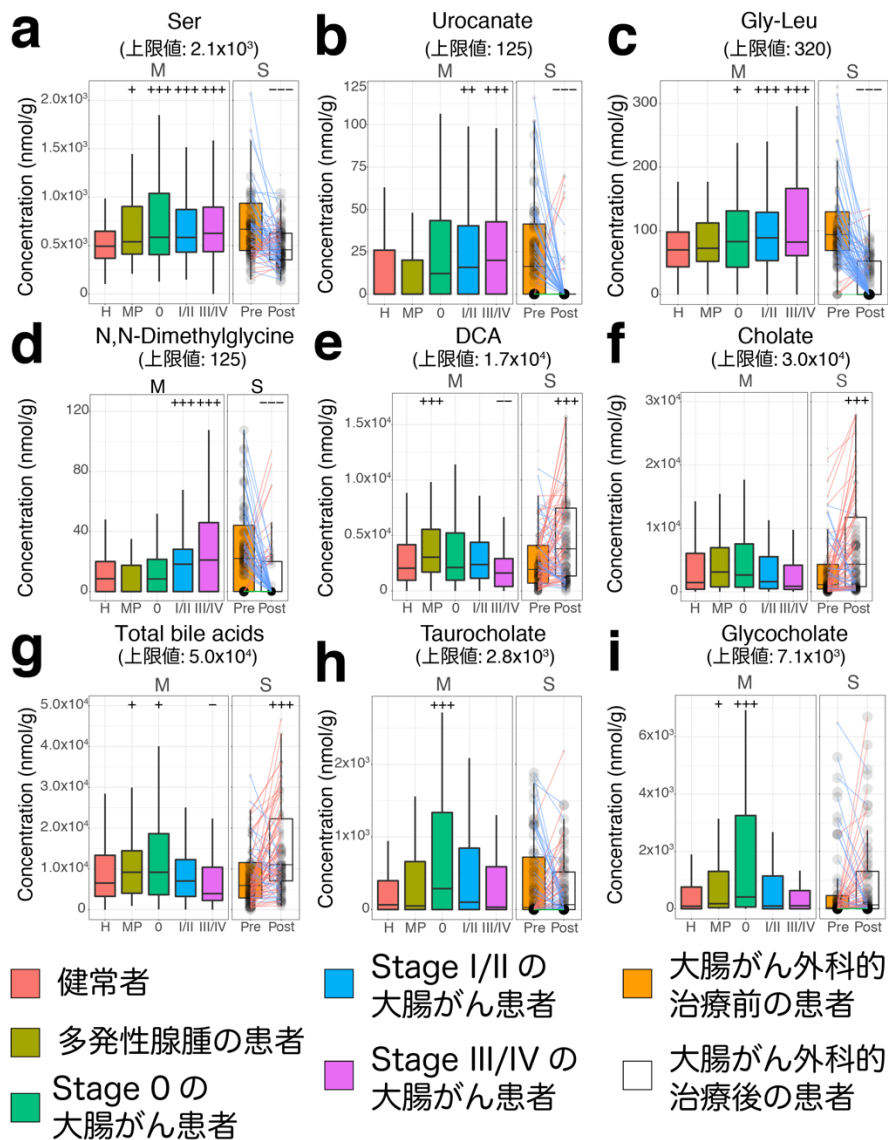


図 5-6 大腸がん外科的治療前後の患者の代謝産物の箱ひげ図

左側の箱ひげ図は、健常者と多段階発がんのいずれかの状態の患者の代謝産物の量を示している。右側の箱ひげ図は、大腸がん患者の外科的治療前後の代謝産物の量を示している。右側の箱ひげ図の丸の大きさは、外科的治療前後のそれぞれの患者におけるそれらの分布を反映している。左側の箱ひげ図は外れ値を省略している。右側の箱ひげ図の線の色は、同一患者の外科的治療前後の変化を表しており、外科的治療後に増加した場合は赤色、外科的治療後に減少した場合は青色、外科的治療前後で増加も減少も確認されなかった場合は緑色で可視化している。左側の箱ひげ図と右側の箱ひげ図は、上限値以下のサンプルのみを可視化している。統計学的に増加またはその量が多かったことが確認された場合、以下のように表記している($P < 0.005$: +++; $P < 0.01$: ++; $P < 0.05$: +)。統計学的に減少またはその量が少なかったことが確認された場合、以下のように表記している($P < 0.005$: ---; $P < 0.01$: --; $P < 0.05$: -)。

5.3.4 腸内細菌の遺伝子の比較

大腸がん患者の外科的治療前後の代謝産物の量を比較した結果、デオキシコール酸やコール酸等の胆汁酸の量が外科的治療後に増加したため、それらの生成に関与する遺伝子の相対存在量を調査した。

まず、大腸がん患者の外科的治療前後のタウロコール酸またはグリココール酸をコール酸に変換する遺伝子である *bile salt hydrolase*(K01442)の相対存在量を比較した結果に着目した。タウロコール酸とグリココール酸の量は外科的治療前後で変化していないが、コール酸の量は外科的治療後に増加したため、*bile salt hydrolase* の相対存在量も増加していることが予想された。しかしながら、患者の外科的治療前後の *bile salt hydrolase* の相対存在量は、統計学的に変化していないことが確認された($P=0.288$, 図 5-7a の箱ひげ図)。遺伝子の相対存在量が代謝産物の変化を反映していない理由として、この遺伝子を保有しているが、反応には寄与していない腸内細菌が存在する可能性が考えられる。そのため、この遺伝子の発現量を定量することで、この反応を担う腸内細菌を特定する研究が重要である。

次に、大腸がん患者の外科的治療前後のコール酸からデオキシコール酸に変換する *bai* オペロンの相対存在量を比較した結果に着目した。コール酸とデオキシコール酸の量は外科的治療後に増加したため、*bai* オペロンの相対存在量も増加していることが予想された。*bai* オペロンの相対存在量は、外科的治療後に有意に増加したことが確認された($P=4.88 \times 10^{-3}$, 図 5-7b の箱ひげ図)。このオペロンを保有する *C. scindens* の相対存在量も外科的治療後に増加したことも確認されたため、このオペロンの担い手を調査することで、腸内細菌とその機能を担う遺伝子の変化に整合性があるかどうかを確認することが重要であると考えられる。

最後に、これらの遺伝子とオペロンの担い手を調査した。遺伝子の担い手とその割合は、その遺伝子を担う腸内細菌の相対存在量を基に算出した(遺伝子とその遺伝子の系統の対応は 3.2.2 節を参照)。その結果、*bile salt hydrolase* は多くの腸内細菌が保有しており、その中でも特に *Bacteroides* 属が最もその機能を担っていることが確認された(図 5-7c)。一方で、*bai* オペロンは *C. scindens* が主にその機能を担っていることが確認された(図 5-7d)。この結果から、*C. scindens* と *bai* オペロンの相対存在量の変化に整合性があることが確認された。

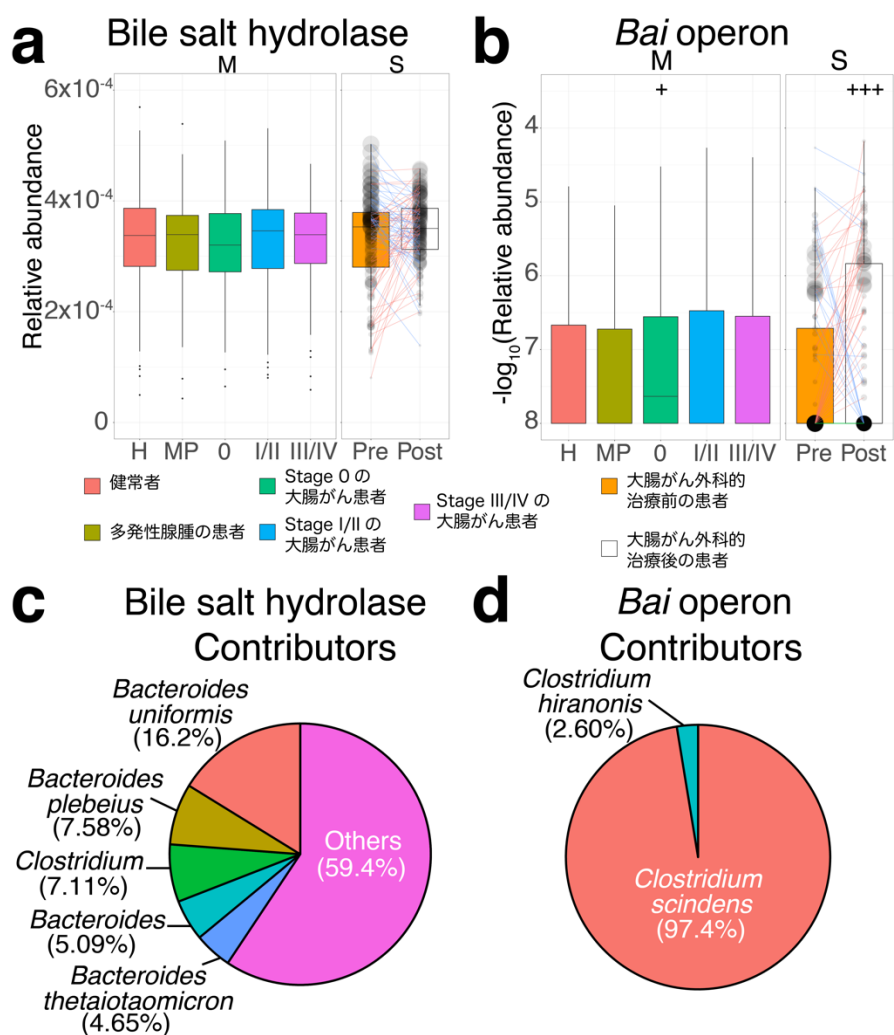


図 5-7 胆汁酸代謝に関係している遺伝子とオペロンの箱ひげ図とそれらの寄与率

胆汁酸代謝に関連している bile salt hydrolase(a)と *bai* オペロン(b)の相対存在量を示している。左側の箱ひげ図は、健常者と多段階発がんのいずれかの状態の患者のそれらの相対存在量を示している。右側の箱ひげ図は、大腸がん患者の外科的治療前後のそれらの相対存在量を示している。b の縦軸は、 $-\log_{10}$ で変換した *bai* オペロンの相対存在量を示している。*Bai* オペロンの相対存在量が 0 であった場合は、その相対存在量を 1×10^{-8} に置換後、 $-\log_{10}$ で変換した。右側の箱ひげ図の線の色は、同一患者での外科的治療前後の変化を表しており、外科的治療後に増加した場合は赤色、外科的治療後に減少した場合は青色、外科的治療前後で増加も減少も確認されなかった場合は緑色で可視化している。右側の箱ひげ図の丸の大きさは、外科的治療前後のそれぞれの患者におけるそれらの分布を反映している。統計学的に増加またはその量が多かったことが確認された場合、以下のように表記している($P < 0.005$: +++; $P < 0.01$: ++; $P < 0.05$: +)。円グラフ(c, d)は、遺伝子やオペロンの担い手となる腸内細菌の割合を示している。円グラフの面積はその割合を反映しており、括弧の中の数字はその割合を示している。

第6章 大腸の再建法と関連する腸内環境の特徴 微量

6.1 序論

第5章で、コール酸や発がん性が報告されているデオキシコール酸等の胆汁酸の量が外科的治療後に増加したことが確認された。先行研究において、右側大腸がん外科的治療後の患者は外科的治療による大腸の再建により、右側大腸だけでなく、胆汁酸の再吸収を担う回腸の一部も取り除かれるため、その影響で胆汁酸の再吸収阻害が生じる可能性が報告されている[96]。そのため、本章では、右側大腸がん患者と左側大腸がん患者、それぞれの外科的治療前後で変化する胆汁酸代謝に関与する腸内細菌やそれらの遺伝子、代謝産物を特定する手法やその結果について述べる。右側大腸がん患者における外科的治療前後のそれらを比較した結果、本研究において定量することができた全ての胆汁酸(タウロコール酸、グリココール酸、コール酸、デオキシコール酸)の量が外科的治療後に増加したことが確認された($P < 0.05$, 図 6-1a)。一方で、左側大腸がん患者における外科的治療前後のそれらを比較した結果、コール酸やデオキシコール酸の量、コール酸からデオキシコール酸に変換する遺伝子(*bai* オペロン)やそれを担う腸内細菌(*C. scindens*)の相対存在量が外科的治療後に有意に増加したことが確認された($P < 0.005$, 図 6-1b)。また、左側大腸がん患者、右側大腸がん患者、それぞれで、血中の総コレステロール値が外科的治療後に増加していないにも関わらず、コレステロールから生成される総胆汁酸の量が外科的治療後に有意に増加したことも確認された($P < 0.005$)。以上の結果から、右側大腸がん外科的治療後の患者では、胆汁酸の再吸収を担う回腸末端の一部を切除したことにより多くの胆汁酸が回腸末端で再吸収されずに大腸内に流入したため、定量できた全ての胆汁酸の量が増加したことが示唆された。一方で、左側大腸がん外科的治療後の患者において便中の総胆汁酸やコール酸の量が増加したメカニズムを明らかにすることができなかったが、コール酸の量が多くなったため *C. scindens* の活性が高くなった結果、デオキシコール酸の量が増加した可能性が示唆された。そのため、大腸内でコール酸を減少させることは大腸内のデオキシコール酸を減少させるための介入のターゲットの1つになり得ることが示唆された。

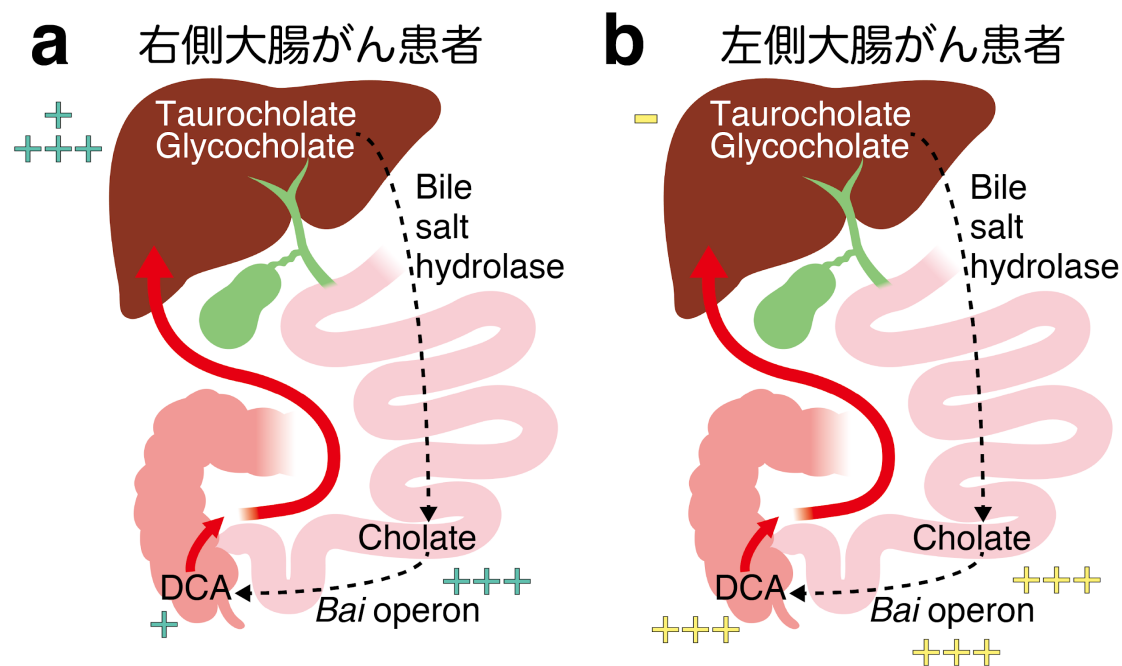


図 6-1 右側大腸がん患者と左側大腸がん患者の胆汁酸代謝の変化の概略図

図中の代謝産物はその生成された位置に配置されている。コール酸は肝臓で生成されるが、すぐにグリシンもしくはタウリンと結合し、グリココール酸、タウロコール酸にそれぞれ変換されるため、小腸に配置されている。黒色の矢印は、腸内細菌による変換を表している。赤い矢印は胆汁酸の再吸収の流れを表している。統計学的に増加したことが確認された場合、以下のように表記している($P < 0.005$: +++; $P < 0.01$: ++; $P < 0.05$: +)。統計学的に減少したことが確認された場合、以下のように表記している($P < 0.005$: ---; $P < 0.01$: --; $P < 0.05$: -)。

6.2 手法

本節では、統計学的な比較に用いたデータの取得方法とその比較に用いた手法について説明する。

まず、統計学的な比較に用いたデータの取得方法について説明する。大腸がん患者の外科的治療前後の疫学データ、種レベルの腸内細菌叢の系統組成データ、腸内細菌の推定された生育速度のデータ、腸内細菌の遺伝子機能組成データ、*bai* オペロンの組成データ、代謝産物の組成データを、2.2.8 節、3.2.3 節、3.2.4 節、3.2.2 節、3.2.5 節、2.2.4 節にてそれぞれ取得した。これらのデータから、血中の総コレステロール値と胆汁酸代謝に関係している 4 つの代謝産物(タウロコール酸、グリココール酸、コール酸、デオキシコール酸)の

量、それら 4 つの胆汁酸の総和である総胆汁酸の量、コール酸の量に対するデオキシコール酸の量の比、bile salt hydrolase や *bai* オペロンの相対存在量、腸内細菌(*C. scindens*)の相対存在量とその推定された生育速度を統計学的な比較に用いた。それぞれのデータとそのデータに含まれている各患者のラベル(外科的治療前、外科的治療後、右側大腸がん、左側大腸がん)を対応させることで、それぞれのデータを層別化した。これらのデータを以下の統計学的な比較に用いた。

次に、統計学的な比較に用いた手法について説明する。右側大腸がん患者、左側大腸がん患者、それぞれの外科的治療前後で変化する胆汁酸に関連する特徴量を特定するために、個人間の対応のある検定であるウィルコクソンの符号順位検定を用いてそれらの量を比較した。コール酸の量に対するデオキシコール酸の量の比は、患者毎にデオキシコール酸の量をコール酸の量で割ることで算出したが、それぞれの患者のコール酸の量またはデオキシコール酸の量が 0 であった場合はその比を算出することができなかった。そのため、これらの患者を取り除いて、ウィルコクソンの符号順位検定を用いて外科的治療前後のその比を比較した。ウィルコクソンの符号順位検定は片側検定で実施し、R の `coin` パッケージの `wilcoxsign_test` を用いて *P* 値を取得した。`distribution` のパラメータは、`exact`、`alternative` のパラメータは、`less` または `greater` を用いた。生育速度の推定にはその腸内細菌由来のメタゲノムリードが必要であるため、特定のサンプルで相対存在量が少ない腸内細菌の生育速度を推定することができなかった。そのため、推定された生育速度は、個人間の対応のないウィルコクソンの順位和検定を用いてそれぞれの外科的治療前の患者と外科的治療後の患者を比較した。右側大腸がん外科的治療後の患者と左側大腸がん外科的治療後の患者でその量が異なる胆汁酸に関連する特徴量を特定するために、ウィルコクソンの順位和検定を用いて比較した。ウィルコクソンの順位和検定は片側検定で実施し、R の `coin` パッケージの `wilcox_test` を用いて *P* 値を取得した。`alternative` のパラメータは `less` または `greater` を用いた。これらの検定によって得られた *P* 値の有意水準は 0.005 とした。

6.3 結果・考察

6.3.1 右側大腸がんの再建法が胆汁酸代謝へ及ぼす影響

右側大腸がん患者の外科的治療前後の血中コレステロール値と、胆汁酸代謝に関連する代謝産物の量、腸内細菌の相対存在量やその推定された生育速度を統計学的に比較した。まず、右側大腸がん患者の外科的治療前後の血中の総コレステロールの値を統計学的に比較した結

果に着目した。右側大腸がん患者の外科的治療前後の血中の総コレステロールの値を比較した結果、それらに統計学的な有意差は確認されなかった($P=0.327$, 図 6-2)。先行研究において、コレステロールは主に肝臓で生成され、タウロコール酸やグリココール酸等の胆汁酸の生成に用いられることが報告されている[94]。以上のことから、右側大腸がん患者の肝臓における、タウロコール酸やグリココール酸等の胆汁酸の生成量は外科的治療前後で変化していないことが示唆された。

次に、右側大腸がん患者の外科的治療前後の代謝産物の量を統計学的に比較した結果に着目した。右側大腸がん患者のタウロコール酸とグリココール酸の量は、外科的治療後に増加したことが確認された($P=0.0327$, $P=1.54 \times 10^{-3}$, 図 6-3a, 図 6-3b)。また、右側大腸がん患者のタウロコール酸またはグリココール酸から変換されるコール酸の量と、コール酸から変換されるデオキシコール酸の量も外科的治療後に増加したことが確認された($P=1.68 \times 10^{-4}$, $P=0.0301$, 図 6-3c, 図 6-3d)。さらに、右側大腸がん患者の総胆汁酸の量も外科的治療後に有意に増加したことが確認された($P=3.62 \times 10^{-5}$, 図 6-3e)。先行研究において、肝臓で生成されたタウロコール酸やグリココール酸等の胆汁酸は十二指腸に輸送された後、これらの胆汁酸の 95%は回腸末端で再吸収され、再吸収されなかった残りの 5%の胆汁酸は大腸に流入後、便として排出されることが報告されている[95]。また、右側大腸がん外科的治療後の患者は外科的治療による大腸の再建により、右側大腸だけでなく、胆汁酸の再吸収を担う回腸の一部も取り除かれるため、その影響で胆汁酸の再吸収阻害が生じる可能性があることも報告されている[96]。以上のことから、本研究において、右側大腸がん外科的治療後の患者でこれらの胆汁酸の量が増加したことは、これらの患者で胆汁酸の再吸収阻害が生じる可能性があるという先行研究の結果と整合性があることが確認された。先行研究において、大腸内で胆汁酸の量が多いことは、大腸の蠕動運動や水分の分泌を促進し下痢を誘導する可能性が報告されている[97]。また、水分の吸収の主な部位である右側の腸を切除することは、便の緩みと関連している可能性が報告されている[98]。これらの先行研究の結果は、右側大腸がん外科的治療後の患者の下痢の有病率が高いことと関連している可能性が示唆されている[99]。そのため、本研究における右側大腸がん外科的治療後の患者の下痢の有病率を調べることでこの仮説を検証する研究が必要であると考えられる。

次に、右側大腸がん患者の外科的治療前後の胆汁酸代謝を担う腸内細菌や遺伝子の相対存在量、*C. scindens* の推定された生育速度を統計学的に比較した結果に着目した。右側大腸がん患者の外科的治療前後のタウロコール酸またはグリココール酸をコール酸に変換する遺伝子である bile salt hydrolase の相対存在量を比較した結果、それらに統計学的な有意な差は確認されなかった($P=0.478$, 図 6-4a)。右側大腸がん患者の外科的治療前後のコー

ル酸からデオキシコール酸に変換する遺伝子である *bai* オペロンの相対存在量を比較した結果、それらに統計学的な有意差は確認されなかったが、外科的治療後に減少する傾向にあったことが確認された($P=0.301$, 図 6-4b)。右側大腸がん患者の外科的治療前後の *bai* オペロンの担い手である *C. scindens* の相対存在量を比較した結果、それらに統計学的な有意差は確認されなかったが、外科的治療後に減少する傾向にあったことも確認された($P=0.445$, 図 6-4c)。一方で、右側大腸がん患者の外科的治療前後の *C. scindens* の推定された生育速度を比較した結果、それらに統計学的な有意差は確認されなかったが、外科的治療後にそれが高いことが確認された($P=0.237$, 図 6-4d)。先行研究において、デオキシコール酸は大腸の右側でコール酸から変換されることが報告されており[100]、デオキシコール酸生成菌である *C. scindens* も大腸の右側に生息していると仮定すると、大腸の右側を切除することにより *C. scindens* の生息に適した環境が変化したため、右側大腸がん患者の *C. scindens* の相対存在量は外科的治療後に減少する傾向にあったことが示唆された。また、右側大腸がん外科的治療後の患者の *C. scindens* の相対存在量は減少したにも関わらず、*C. scindens* の推定された生育速度が高い傾向にあったことについては、6.3.2 節で後述するが、コール酸の量が増加したため、*C. scindens* の活性が高くなったという仮説と矛盾していないと考えられる。

最後に、右側大腸がん患者の外科的治療前後のコール酸の量に対するデオキシコール酸の量の比を統計学的に比較した結果に着目した。右側大腸がん患者の外科的治療前後のコール酸の量に対するデオキシコール酸の量の比を比較した結果、統計学的な有意差は確認されなかったが、外科的治療後にその比が低い傾向にあることが確認された($P=0.106$, 図 6-3f)。この結果から、右側大腸がん外科的治療後の患者の大腸において、コール酸の量は多いがコール酸からデオキシコール酸への変換率が低い傾向にある可能性が示唆された。また、右側大腸がん外科的治療後の患者において、コール酸の量に対するデオキシコール酸の量の比が低い傾向にあることは、コール酸からデオキシコール酸に変換する *bai* オペロンとその担い手である *C. scindens* の相対存在量が減少した傾向にあったことが原因である可能性が示唆された。

血中の 総コレステロール値

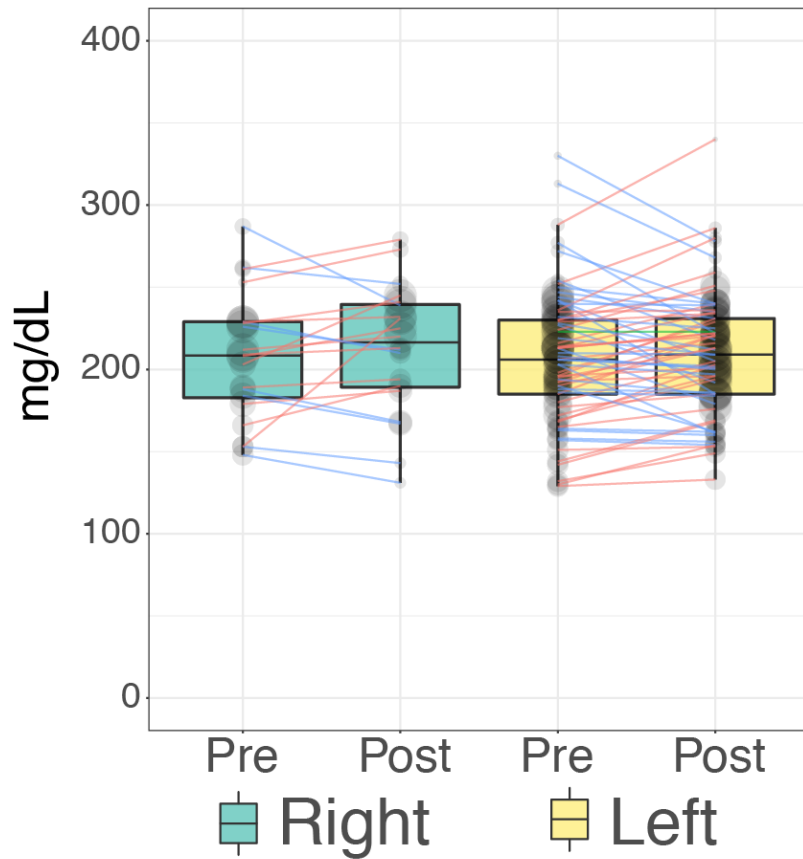


図 6-2 右側大腸がん患者と左側大腸がん患者の血中の総コレステロールの箱ひげ図
丸の大きさは、外科的治療前後のそれぞれの患者における血中総コレステロール値の分布を反映している。箱ひげ図の色は、右側大腸がん患者(緑色)、左側大腸がん患者(黄色)であることを表している。線の色は同一患者での外科的治療前後の変化を表しており、外科的治療後に増加した場合は赤色、外科的治療後に減少した場合は青色、外科的治療前後で増加も減少も確認されなかった場合は緑色で可視化している。

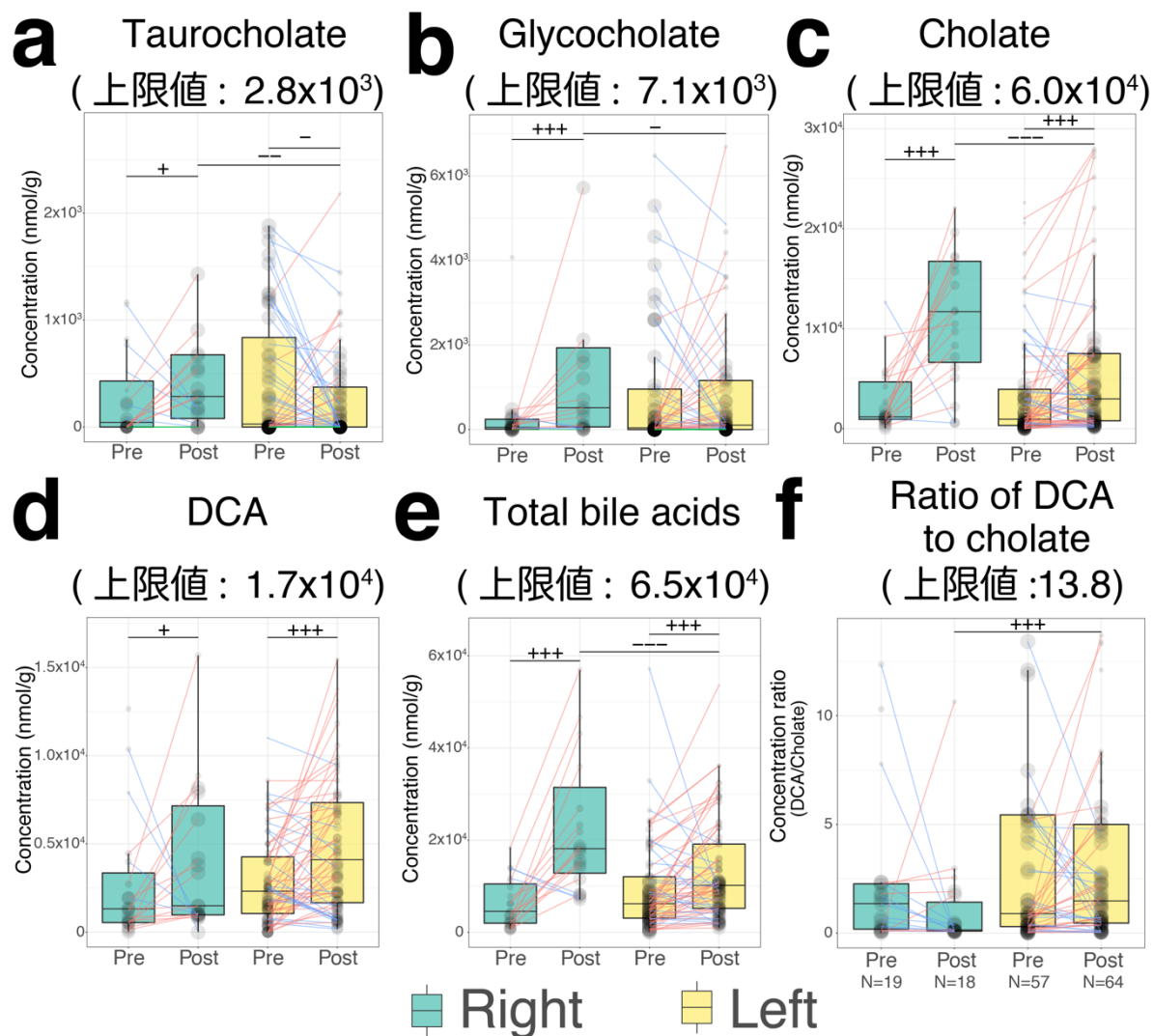


図 6-3 右側大腸がん患者と左側大腸がん患者の代謝産物の箱ひげ図

箱ひげ図は上限値以下のサンプルのみを可視化している。丸の大きさは、外科的治療前後のそれぞれの患者におけるそれらの分布を反映している。箱ひげ図の色は、右側大腸がん患者(緑色)、左側大腸がん患者(黄色)であることを表している。線の色は同一患者での外科的治療前後の変化を表しており、外科的治療後に増加した場合は赤色、外科的治療後に減少した場合は青色、外科的治療前後で増加も減少も確認されなかった場合は緑色で可視化している。統計学的に増加またはその量が多かったことが確認された場合、以下のように表記している($P < 0.005$: +++; $P < 0.01$: ++; $P < 0.05$: +)。統計学的に減少またはその量が少なかったことが確認された場合、以下のように表記している($P < 0.005$: ---; $P < 0.01$: --; $P < 0.05$: -)。コール酸の量に対するデオキシコール酸の量の比(f)を算出することができたサンプル数(N)を箱ひげ図の下に示している。

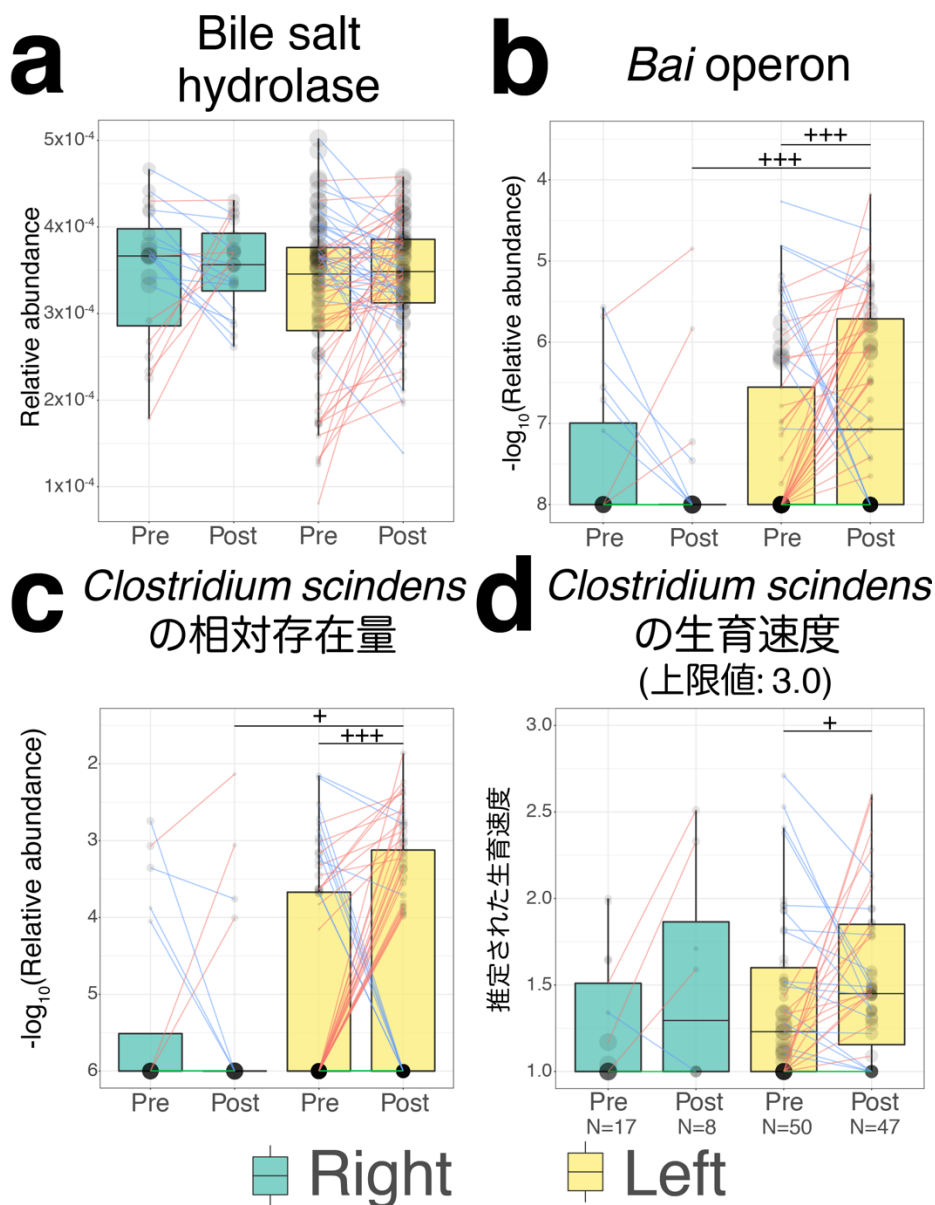


図 6-4 右側大腸がん患者と左側大腸がん患者の腸内細菌やその遺伝子の箱ひげ図

d は上限値以下のサンプルのみを可視化している。丸の大きさは、外科的治療前後のそれぞれの患者におけるそれらの分布を反映している。箱ひげ図の色は、右側大腸がん患者(緑色)、左側大腸がん患者(黄色)であることを表している。線の色は同一患者での外科的治療前後の変化を表しており、外科的治療後に増加した場合は赤色、外科的治療後に減少した場合は青色、外科的治療前後で増加も減少も確認されなかった場合は緑色で可視化している。統計学的に増加またはその量が多かったことが確認された場合、以下のように表記している($P < 0.005$: +++; $P < 0.01$: ++; $P < 0.05$: +)。統計学的に減少またはその量が少なかったことが確認された場合、以下のように表記している($P < 0.005$: ---; $P < 0.01$: --; $P < 0.05$: -)。生育速度を推定することができたサンプル数(N)を箱ひげ図の下に示している(d)。

6.3.2 左側大腸がんの再建法が胆汁酸代謝へ及ぼす影響

左側大腸がん患者の外科的治療前後の血中コレステロール値と胆汁酸代謝に関連する代謝産物の量、腸内細菌の相対存在量やその推定された生育速度を統計学的に比較した。まず、左側大腸がん患者の外科的治療前後の総胆汁酸の量と血中の総コレステロールの値を統計学的に比較した結果に着目した。左側大腸がん患者の総胆汁酸の量は、外科的治療後に有意に増加したことが確認された($P=1.30 \times 10^{-4}$, 図 6-3e)。一方で、左側大腸がん患者の外科的治療前後の血中の総コレステロールの値を比較した結果、それらに統計学的な有意な差は確認されなかった($P=0.203$, 図 6-2)。これらのことから、左側大腸がん外科的治療後の患者の肝臓における胆汁酸の生成量は増加していない可能性が高いにも関わらず、便中の総胆汁酸の量は増加したことが確認された。左側大腸がん外科的治療後の患者における総胆汁酸の量が増加したメカニズムを明らかにすることはできなかったが、手術によって神経が損傷した結果、胆汁酸の再吸収能が低下したため、便中の総胆汁酸の量が増加した可能性が示唆された。

次に、左側大腸がん患者の外科的治療前後のタウロコール酸、グリココール酸、コール酸の量と、タウロコール酸またはグリココール酸をコール酸に変換する *bile salt hydrolase* の相対存在量を統計学的に比較した結果に着目した。左側大腸がん患者の外科的治療前後のグリココール酸の量を比較した結果、それらに統計学的な有意な差は確認されなかった($P=0.431$, 図 6-3b)。左側大腸がん患者のタウロコール酸の量は外科的治療後に減少し、コール酸の量は外科的治療後に有意に増加したことが確認された($P=0.0366$, $P=5.31 \times 10^{-4}$, 図 6-3a, 図 6-3c)。これらの結果から、左側大腸がん外科的治療後の患者において、タウロコール酸からコール酸への変換率が高い可能性が示唆された。しかしながら、左側大腸がん患者の外科的治療前後のタウロコール酸またはグリココール酸をコール酸に変換する遺伝子である *bile salt hydrolase* の相対存在量を比較した結果、それらに統計学的な有意な差は確認されなかった($P=0.0897$, 図 6-4a)。この遺伝子は、多くの腸内細菌が保有しているため、この遺伝子の相対存在量はその変換率を反映していなかった可能性がある。そのため、この遺伝子の発現を定量することで、この反応を担う腸内細菌を特定する研究が必要であると考えられる。

最後に、左側大腸がん患者の外科的治療前後のコール酸の量に対するデオキシコール酸の量の比や、コール酸からデオキシコール酸に変換する *bai* オペロンやその担い手である *C. scindens* の相対存在量と *C. scindens* の推定された生育速度を統計学的に比較した結果に着目した。左側大腸がん患者の外科的治療前後のコール酸の量に対するデオキシコール酸の量の比は、統計学的な有意差は確認されなかったが、外科的治療後に増加する傾向にあ

ることが確認された($P=0.332$, 図 6-3f)。この結果から、左側大腸がん外科的治療後の患者の大腸において、コール酸からデオキシコール酸への変換率が高い傾向にある可能性が示唆された。左側大腸がん患者のコール酸からデオキシコール酸に変換する *bai* オペロンや、その担い手である *C. scindens* の相対存在量は外科的治療後に有意に増加したことが確認された($P=1.91 \times 10^{-3}$, $P=2.74 \times 10^{-3}$, 図 6-4b, 図 6-4c)。また、左側大腸がん外科的治療後の患者の *C. scindens* の推定された生育速度も、左側大腸がん外科的治療前の患者のそれと比較して高いことが確認された($P=0.0360$, 図 6-4d)。これらの結果から、左側大腸がん外科的治療後の患者では、コール酸の量が多くなったため *C. scindens* の活性が高くなった結果、デオキシコール酸の量が増加した可能性が示唆された。右側大腸がん外科的治療後の患者で、コール酸の量が増加したことに伴って、*C. scindens* の推定された生育速度が高くなったことも、この仮説を支持していると考えられる。以上の結果から、大腸内でコール酸を減少させることは、大腸内のデオキシコール酸を減少させるための介入のターゲットの 1 つになり得ることが示唆された。

6.3.3 大腸の再建法の違いが胆汁酸代謝へ及ぼす影響

右側大腸がん外科的治療後の患者と左側大腸がん外科的治療後の患者の胆汁酸代謝に関連する代謝産物の量、腸内細菌の相対存在量やその推定された生育速度を統計学的に比較した。

まず、右側大腸がん外科的治療後の患者と左側大腸がん外科的治療後の患者の総胆汁酸の量とタウロコール酸、グリココール酸、コール酸の量を統計学的に比較した結果に着目した。総胆汁酸、タウロコール酸、グリココール酸、コール酸の量は、左側大腸がん外科的治療後の患者と比較して右側大腸がん外科的治療後の患者でそれらの量が多いことが確認された($P=1.84 \times 10^{-3}$, $P=9.58 \times 10^{-3}$, $P=0.0192$, $P=1.08 \times 10^{-3}$, 図 6-3e, 図 6-3a, 図 6-3b, 図 6-3c)。これら結果から、左側大腸がん外科的治療後の患者と比較して右側大腸がん外科的治療後の患者の方が胆汁酸の再吸収能が低い可能性が示唆された。

次に、右側大腸がん外科的治療後の患者と左側大腸がん外科的治療後の患者のデオキシコール酸の量と、コール酸の量に対するデオキシコール酸の量の比、コール酸からデオキシコール酸に変換する *bai* オペロンやその担い手である *C. scindens* の相対存在量と *C. scindens* の推定された生育速度を統計学的に比較した結果に着目した。デオキシコール酸の量を比較した結果、それらに統計学的な有意差は確認されなかったが、右側大腸がん外科的治療後の患者と比較して左側大腸がん外科的治療後の患者でその量が多い傾向にあることが確認された($P=0.157$, 図 6-3d)。また、コール酸の量に対するデオキシコール酸の量の比

は、右側大腸がん外科的治療後の患者と比較して左側大腸がん外科的治療後の患者でその比が有意に高いことが確認された($P=1.86 \times 10^{-3}$, 図 6-3f)。これらの結果から、右側大腸がん外科的治療後の患者と比較して左側大腸がん外科的治療後の患者の方がデオキシコール酸の生成に適した腸内環境である可能性が示唆された。コール酸からデオキシコール酸に変換する *bai* オペロンや、その担い手である *C. scindens* の相対存在量も、右側大腸がん外科的治療後の患者と比較して左側大腸がん外科的治療後の患者でその相対存在量が多いことが確認された($P=1.99 \times 10^{-3}$, $P=0.0106$, 図 6-4b, 図 6-4c)。また、*C. scindens* の推定された生育速度は、統計学的な有意差は確認されなかったが、右側大腸がん外科的治療後の患者と比較して左側大腸がん外科的治療後の患者でそれが高い傾向にあったことが確認された($P=0.341$, 図 6-4d)。これらの結果も、右側大腸がん外科的治療後の患者と比較して、左側大腸がん外科的治療後の患者の方がデオキシコール酸の生成に適した腸内環境であるという仮説を支持していると考えられる。

第7章 外科的治療後の大腸がんのリスク推定 手法の確立

7.1 序論

大腸がん患者は外科的治療後も大腸がんのリスクが高いことが報告されているため、そのリスクを評価することが重要であると考えられる。本研究では、治療から約5年間で大腸がんの前がん病変である多発性腺腫や大腸がんを発症したかどうかを基に、大腸がん患者の外科的治療後の大腸がんのリスクを定量化することができた(第2章参照)。先行研究において、腸内環境の組成データ(腸内細菌の系統組成やそれらの遺伝子機能組成、代謝産物の組成)を用いた教師あり学習によって構築された判別器は、大腸がんを発見するためのスクリーニングに用いることができる可能性が示唆されている[8,9,43,44,46]。本章では、大腸がんのリスクが低い患者と高い患者の外科的治療前後の腸内環境の組成データをこの手法に適用することで、判別器を用いて外科的治療後の大腸がんのリスクを推定できるかどうかを検討した手法や、その結果について述べる(図7-1)。

まず、外科的治療後の大腸がんのリスクを推定する手法の構築方法について述べる。先行研究において、大腸がんの発症・進行に応じて腸内細菌の相対存在量や代謝産物の量が変わることが報告されている[4]ため、健常者以外の患者を多段階発がんのそれぞれの状態(多発性腺腫、Stage 0 の大腸がん、Stage I/II の大腸がん、Stage III/IV の大腸がん)に層別化し、それぞれの状態の患者と健常者を分類する判別器をそれらの被験者の腸内環境の組成データから構築した。次に、それぞれの判別器に大腸がん患者の外科的治療前後の腸内環境の組成データを適用し、判別器から出力される確率を取得した。最後に、判別器から得られた確率を、大腸がんのリスクが低い患者と高い患者で比較し、大腸がんのリスクが高い患者でその確率が高いことを確認することで、その判別器が外科的治療後の大腸がんのリスクを推定できるかどうかを評価した。その結果、健常者と多発性腺腫の患者または、健常者と Stage 0 の大腸がん患者を分類する判別器に大腸がんのリスクが低い患者と高い患者の外科的治療前の腸内環境の組成データを適用することで得られた確率は、大腸がんのリスクが低い患者と比較して大腸がんのリスクが高い患者で高いことが確認された($P < 0.05$)。

以上のことから、健常者と多発性腺腫の患者または、健常者と Stage 0 の大腸がん患者を分類する判別器を用いることで、外科的治療後の腸内環境に由来する大腸がんのリスクを評価できる可能性が示唆された。

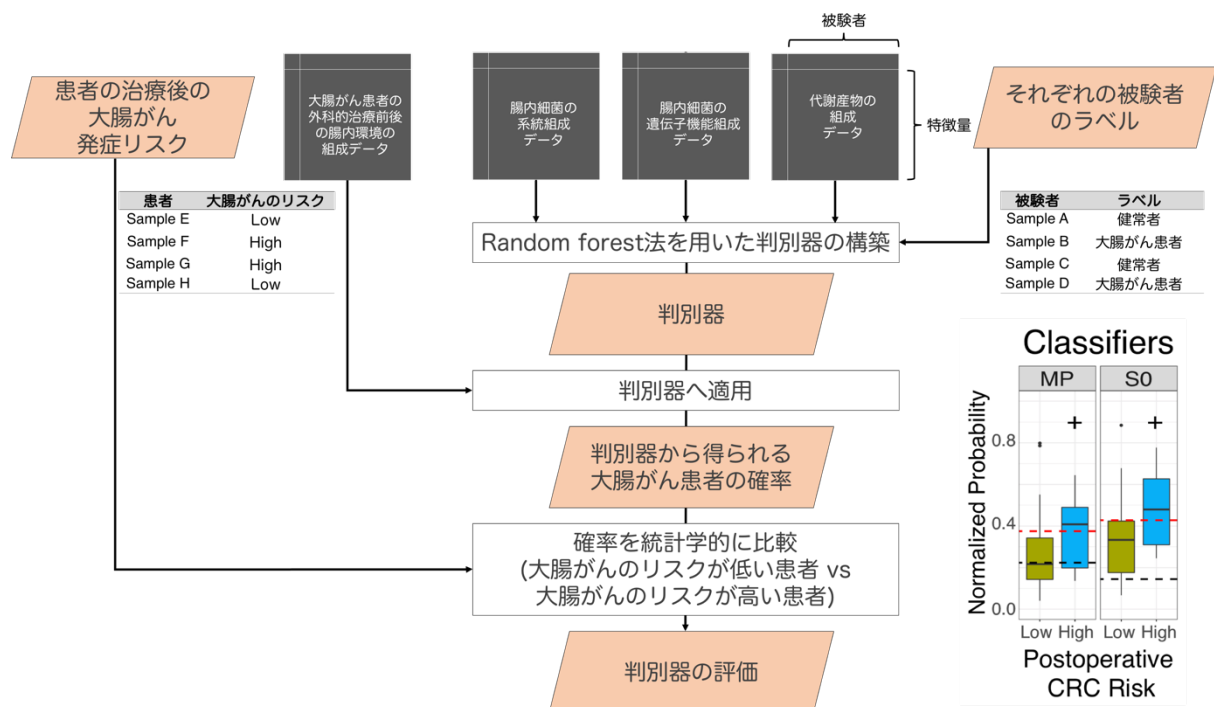


図 7-1 外科的治療後の大腸がんのリスク推定手法の評価の概略図

腸内細菌の系統組成データ、腸内細菌の遺伝子機能組成データ、代謝産物の組成データとそれらのデータに対応する各被験者のラベルを用いて、Random forest 法を基盤とした判別器を構築した。その判別器へ大腸がん患者の外科的治療前後の腸内環境の組成データを適用することで、確率を取得した。大腸がんのリスクが低い患者と比較して大腸がんのリスクが高い患者でその確率が高いかどうかを調査することで、大腸がんのリスクを推定するうえでの判別器の性能を評価した。

7.2 手法

7.2.1 判別器の構築

本節では、判別器の構築に用いた腸内環境の組成データの取得方法、判別器の構築に用いたアルゴリズム、判別器のパラメータ選択のための精度の評価方法、判別器の構築方法について説明する。

判別器の構築に用いる腸内環境の組成データの取得方法

先行研究において、判別器の構築に用いた被験者のメタゲノムサンプルは既に公開されているが[4]、一部の被験者については便に含まれている代謝産物は測定されていなかった(2.3.2

節参照)。本研究では、その被験者の便に含まれている代謝産物を新たに測定したため、その被験者の代謝産物の組成データを取得することができた。判別器を構築するうえで、対象となる被験者数が多い方が判別器の構築に用いていない患者の外科的治療後の大腸がんのリスクを推定する精度が高いことが予想されたため、判別器を再度構築する必要があった。

腸内環境の組成データは、腸内細菌の系統組成データ、それらの遺伝子機能組成データ、代謝産物の組成データから構築されており、3.2.3 節、3.2.2 節、2.2.4 節にてそれぞれの組成を取得した。腸内環境の組成データとそのデータに含まれている各被験者のラベル(健常者、多発性腺腫の患者、Stage 0 の大腸がん患者、Stage I/II の大腸がん患者、Stage III/IV の患者)を対応させた。判別器の構築には、腸内環境の組成データと、そのデータに対応する各被験者のラベルが必要である。構築した判別器は、腸内環境の組成データを入力として受け取ることで、それに対応する被験者の各ラベルを予測することが可能である。判別器によって出力される各ラベルに属する確率が一定の閾値を超えた場合、その被験者は特定のラベルに属すると予測される。

本研究では、245 人の健常者、61 人の多発性腺腫の患者、57 人の Stage 0 の大腸がん患者、85 人の Stage I/II の大腸がん患者、59 人の Stage III/IV の大腸がん患者と、そのデータに対応する各被験者のラベルを用いて、健常者とそれぞれの患者を分類する 4 つの判別器を構築した。代謝産物を測定できなかった患者の腸内環境の組成データは、判別器の構築に使用しなかった。また、大腸がん外科的治療前後の患者の腸内環境の組成データを用いて判別器を評価するため、83 人の外科的治療前後の患者の腸内環境の組成データも判別器の構築に使用しなかった。

判別器の構築に用いたアルゴリズム

線形回帰を基盤とした Lasso(least absolute shrinkage and selection operator)回帰や、非線形回帰である決定木を基盤とした Random forest 法等のアルゴリズムは判別器の構築によく用いられている[8,9]。Lasso 回帰は、特徴量選択の過程で共起関係にある特徴量を取り除かれてしまうため、判別器の構築に寄与した特徴量の解釈が難しいことがある[4]。そのため、本研究では、Python(version 3.5.3)の機械学習のライブラリーである Scikit-learn(version 0.19.1)の RandomForestClassifier の fit を用いて、Random forest 法を基盤とした判別器を構築した。max_features のパラメータは sqrt、class_weight のパラメータは balanced、random_state のパラメータは 6 を用いた。後述するが、決定木の数である n_estimators と、決定木の深さである max_depth は、最も精度が高くなるパラメータを用いた。

Random forest 法は、複数回ランダムに分割されたデータから構築された判別器である弱学習器を組み合わせたアルゴリズムである[101]。ランダムに分割されたデータ毎に構築された弱学習器は、その構築に用いたデータがそれぞれ異なるため、その多様性が高くなる。そのため、これらの弱学習を組み合わせることで予測能が高い判別器を構築することが可能である。判別器の構築に寄与した特徴量の寄与率は、判別器の `feature_importances_` を用いて算出した。Random forest 法では、ランダムに分割されたデータ毎に特徴量を 1 つ取り除いた判別器を構築後、その判別器の構築に用いていないデータをその判別器へそれぞれ適用することで算出した予測精度が、特徴量を取り除いていない場合の予測精度と比較してどれくらい低下したのかを表した `mean decrease gini` を指標に判別器の構築に寄与した特徴量の寄与率を算出する。そのため、特徴量間の共起関係に依存せずに、それぞれの特徴量の寄与率を算出することができる。寄与率が高い特徴量は、大腸がんを診断するためのバイオマーカーとして利用できる可能性がある[8,9]。

判別器のパラメータ選択のための精度の評価方法

適切な判別器のパラメータや特徴量を選択せずに判別器を構築した場合、判別器の構築に用いた腸内環境の組成データに対して過剰に適合し、判別器の構築に用いていない腸内環境の組成データに対する予測能が著しく低下する過学習が生じる可能性がある。そのため、判別器に構築に用いる特徴量やパラメータを変更して多くの判別器を構築し、その中から最も予測精度が良い判別器を選択した。

判別器の予測精度は 10 分割交差検証することで評価した。10 分割交差検証は、腸内環境の組成データを 10 個に分割後、判別器の構築に用いる 9 個のトレーニングデータと判別器の予測精度の検証に用いる 1 個のテストデータの組み合わせを 10 個作成することで、予測精度を評価する手法である。

本研究では、Python の機械学習のライブラリーである Scikit-learn の `StratifiedKFold` を用いて 10 分割交差検証した。`StratifiedKFold` のパラメータは、`n_splits=10`、`shuffle=True` を用いた。乱数のパラメータである `random_state` を 0 から 9 まで変更して、10 分割交差検証を 10 回繰り返すことで、100 個のトレーニングデータとテストデータの組み合わせを取得した。それぞれの組み合わせにおいて、トレーニングデータを用いて判別器を構築後、判別器の `predict_proba` を用いてその判別器へテストデータを適用することで、テストデータである腸内環境の組成に対応する各被験者の各ラベルに属する確率を取得した。判別器から出力されたテストデータに含まれている各被験者のそれぞれの各ラベルに属する確率と、その被験者のラベルを照合することで予測精度を算出した。各被

験者の各ラベルに属する確率を各被験者のラベルと照合するためには、その確率の閾値を定義する必要があるが、判別器の精度や判別器を構築するときに用いたラベルの偏り等によって、判別器毎に適切な閾値が異なる。そのため、AUC(Area Under The Curve)を指標に予測精度を評価した。AUC は、Scikit-learn の `roc_auc_score` を用いてテストデータに含まれている各被験者の健常者以外のラベルに属する確率とその被験者のラベルから算出した。それぞれのトレーニングデータとテストデータの組み合わせにおいて算出される 100 個の AUC の平均値をその判別器の予測精度とした。

判別器の構築方法

腸内環境の組成データは、特徴量が非常に多いため、以下の 10 個の手順で特徴量を選択し、判別器を構築した(図 7-2)。この手順を、健常者と多発性腺腫の患者、健常者と Stage 0 の大腸がん患者、健常者と Stage I/II の大腸がん患者、健常者と Stage III/IV の大腸がん患者の腸内環境の組成データに対してそれぞれ適用し、最終的に 4 つの判別器を構築した。

1. 腸内環境の組成データを構成するそれぞれの組成データ(腸内細菌の系統組成データ、遺伝子機能組成データ、代謝産物の組成データ)に対して 10 分割交差検証を 10 回繰り返し、100 個のトレーニングデータとテストデータの組み合わせを作成した。
2. それぞれのトレーニングデータを用いて、決定木の数である `n_estimators` と決定木の深さである `max_depth` を変更して多くの判別器を構築後、その判別器へそれぞれのテストデータを適用し精度の評価指標である AUC を取得した。この工程はそれぞれの組成データ毎に実施した。
3. それぞれのパラメータ毎に 100 個の AUC の平均値を算出し、最も AUC が高くなるパラメータを決定した。この工程もそれぞれの組成データ毎に実施した。
4. 最も精度が高くなった判別器の構築に寄与した特徴量の寄与率を算出した。この工程もそれぞれの組成データ毎に実施した。
5. それぞれのトレーニングデータを用いて、決定木の数である `n_estimators` と決定木の深さである `max_depth` と、それぞれの組成データから構築された判別器の寄与率が高い特徴量の数を変更して多くの判別器を構築後、その判別器へそれぞれのテストデータを適用し AUC を取得した。
6. それぞれのパラメータ毎に 100 個の AUC の平均値を算出し、最もそれが高くなるパラメータを決定した。

7. 最も精度が高くなる判別器の構築に寄与した特徴量の寄与率を算出した。
8. 100 個のトレーニングデータから構築された判別器の構築に寄与した特徴量毎にその寄与率の中央値を算出し、その寄与率が 0 であった特徴量を取り除いた。
9. 8 で得た特徴量に対応する腸内環境の組成データから最終的な判別器を構築した。
10. 最終的な判別器の構築に寄与した特徴量の寄与率を算出した。

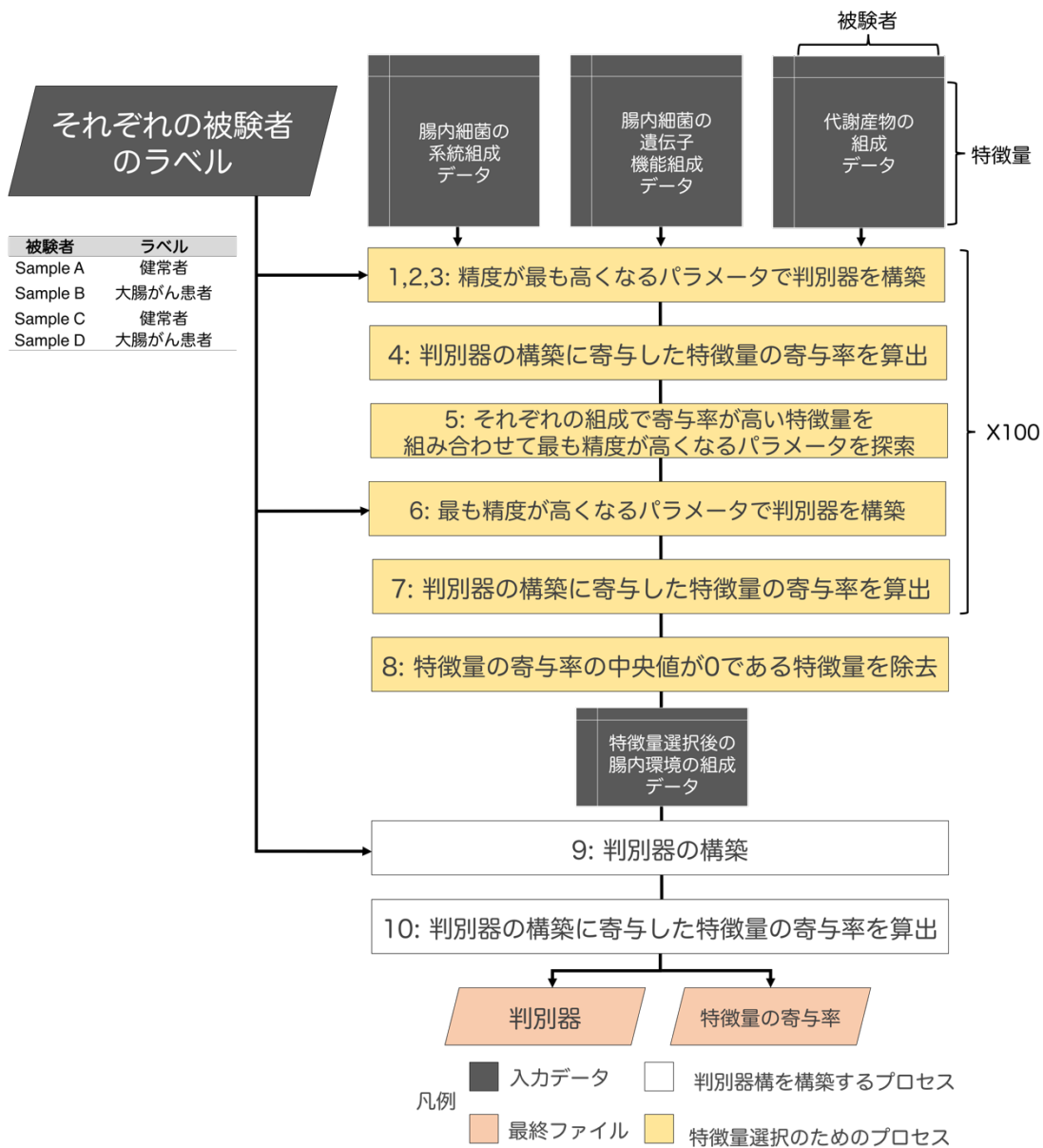


図 7-2 判別器を構築するための特徴量選択の概略図

腸内細菌の系統組成データ、腸内細菌の遺伝子機能組成データ、代謝産物の組成データに対して 10 分割交差検証を 10 回繰り返して 100 個のトレーニングデータとテストデータの組み合わせを作成した。その 100 個のトレーニングデータを用いて特徴量や判別器の精度が最も高くなるパラメータを探索した。特徴量選択後の組成データとそれぞれの被験者のラベルを用いて、そのパラメータで判別器を構築した。

7.2.2 判別器への腸内環境データの適用

構築した 4 つの判別器へ本研究で取得した各被験者の腸内環境データをそれぞれ適用することで、各被験者の各ラベルに属する確率を取得することが可能である。それぞれの判別器

の構築に用いていない多段階発がんのいずれかの状態の患者(多発性腺腫の患者、Stage 0 の大腸がん患者、Stage I/II の大腸がん患者、Stage III/IV の大腸がん患者)や大腸がん患者の外科的治療前後の腸内環境の組成データを、構築した 4 つの判別器へそれぞれ適用することでそれらの各ラベルに属する確率を算出した。判別器の構築に用いた各被験者の腸内環境の組成のデータもその判別器へ適用することが可能であるが、それらの被験者の腸内環境の組成データは既に判別器の構築に用いているため、この手法で算出される各ラベルに属する確率は予測能を正確に反映しないことが予想される。そのため、1 人の被験者の腸内環境の組成データを取り除いて判別器を構築し、その判別器へ 1 人の被験者の腸内環境の組成データを適用し、その被験者の各ラベルに属する確率を取得する工程を判別器の構築に用いた全ての被験者に対して実施する Leave-one-out 法により、それぞれの判別器の構築に用いた各被験者の各ラベルに属する確率を算出した。Leave-one-out 法により構築した複数の判別器は、7.2.1 節で構築した判別器と比較して、1 人分の腸内環境の組成データが判別器の構築に使用されていないが、それらの大部分は共通している。そのため、Leave-one-out 法により構築された判別器から出力される各ラベルに属する確率も予測能を反映していると考えられる。

7.2.3 確率の補正

それぞれの判別器から出力される各ラベルに属する確率は、判別器の精度や判別器を構築するときに用いたラベルの偏り等によって分布が異なる可能性がある。そのため、それぞれの判別器から得られた確率を、以下の式で補正した。

$$\text{Normalized probability} = \frac{\text{probability} \cdot \min(\text{probability})}{\max(\text{probability}) \cdot \min(\text{probability})}$$

Normalized probability は各被験者の補正された健常者以外のラベルに属する確率である。probability は判別器から得られた各被験者の健常者以外のラベルに属する確率である。min(probability)は判別器から得られた全ての被験者の健常者以外のラベルに属する確率の最小値である。max(probability)は判別器から得られた全ての被験者の健常者以外のラベルに属する確率の最大値である。

7.2.4 統計処理

腸内環境の組成データをそれぞれの判別器と補正式へ適用することで、その腸内環境の組成データに対応する各被験者の補正確率を取得した。各被験者の補正確率と各被験者のラベルを対応させて、以下の統計学的な比較に用いた。

大腸がんのリスクが低い患者と比較して大腸がんのリスクが高い患者でその補正確率が高いかどうかを調べるために、ウィルコクソンの順位和検定を用いて、それらの補正確率を比較した。ウィルコクソンの順位和検定は片側検定で実施し、R の `coin` パッケージの `wilcox_test` を用いて P 値を取得した。`alternative` のパラメータは `less` または `greater` を用いた。これらの検定によって得られた P 値の有意水準は 0.005 とした。

7.3 結果・考察

7.3.1 大腸がんのリスクを推定する手法の確立

外科的治療後の大腸がんのリスクを推定する手法を確立するために、判別器を構築し、その判別器に大腸がんのリスクが低い患者と高い患者の腸内環境の組成データを適用することで得られた確率を取得し、大腸がんのリスクが低い患者の確率と大腸がんのリスクが高い患者の確率を比較した。

まず、判別器を構築した結果について述べる。健常者、多発性腺腫の患者、Stage 0 の大腸がん患者、Stage I/II の大腸がん患者、Stage III/IV の大腸がん患者の腸内環境の組成データとそのデータに対応する被験者のラベルを用いて、健常者とそれぞれの患者を分類する 4 つの判別器を構築した。10 分割交差検証を 10 回繰り返すことで得られた判別器の性能は表 7-1 に示した。構築した 4 つの判別器の中で、健常者と Stage III/IV の大腸がん患者を分類する判別器の精度が最も高いことが確認された ($AUC=0.849$)。先行研究においても、対象となった被験者数が本研究とは異なるが、健常者と Stage III/IV の大腸がん患者を分類する判別器の精度が高いことが報告されているため[4]、判別器を構築する手法の再現性を確認することができた。

次に、56 人の大腸がんのリスクが低い患者と 18 人の大腸がんのリスクが高い患者の外科的治療前後の腸内環境の組成データを判別器に適用することで、確率を取得し、大腸がんのリスクが低い患者の確率と大腸がんのリスクが高い患者の確率を比較した結果について述べる。健常者と多発性腺腫の患者、健常者と Stage 0 の大腸がん患者を分類する判別器

に、大腸がんのリスクが低い患者と大腸がんのリスクが高い患者の外科的治療前の腸内環境の組成データを適用し得られた確率は、大腸がんのリスクが低い患者と比較して大腸がんのリスクが高い患者で高いことが確認された(図 7-3a, $P=0.0240$, $P=0.0129$)。このことから、これらの判別器に大腸がん患者の外科的治療前の腸内環境の組成データを適用することで、外科的治療後の大腸がんのリスクを推定できる可能性が示唆された。一方で、健常者と多発性腺腫の患者、健常者と Stage 0 の大腸がん患者を分類する判別器に、大腸がんのリスクが低い患者と大腸がんのリスクが高い患者の外科的治療後の腸内環境の組成データを適用し得られた確率を、大腸がんのリスクが低い患者と大腸がんのリスクが高い患者で比較した結果、それらに有意な差は確認されなかった(図 7-3b, $P>0.05$)。大腸がん患者の外科的治療後の腸内環境の組成データを適用した結果外科的治療後の大腸がんのリスクを推定できなかった原因については明らかにすることはできなかった。

表 7-1 判別器を構築に用いた特徴量の数とパラメータ

Classifier	Species (N)	KO (N)	Metabolome (N)	n_estimator	max_depth	Average of AUC
Healthy vs MP	5	0	15	2500	20	0.640
Healthy vs Stage 0	20	5	25	500	5	0.686
Healthy vs Stage I/II	15	0	25	500	15	0.708
Healthy vs Stage III/IV	5	0	25	2500	20	0.848

Nは特徴量の数を示している

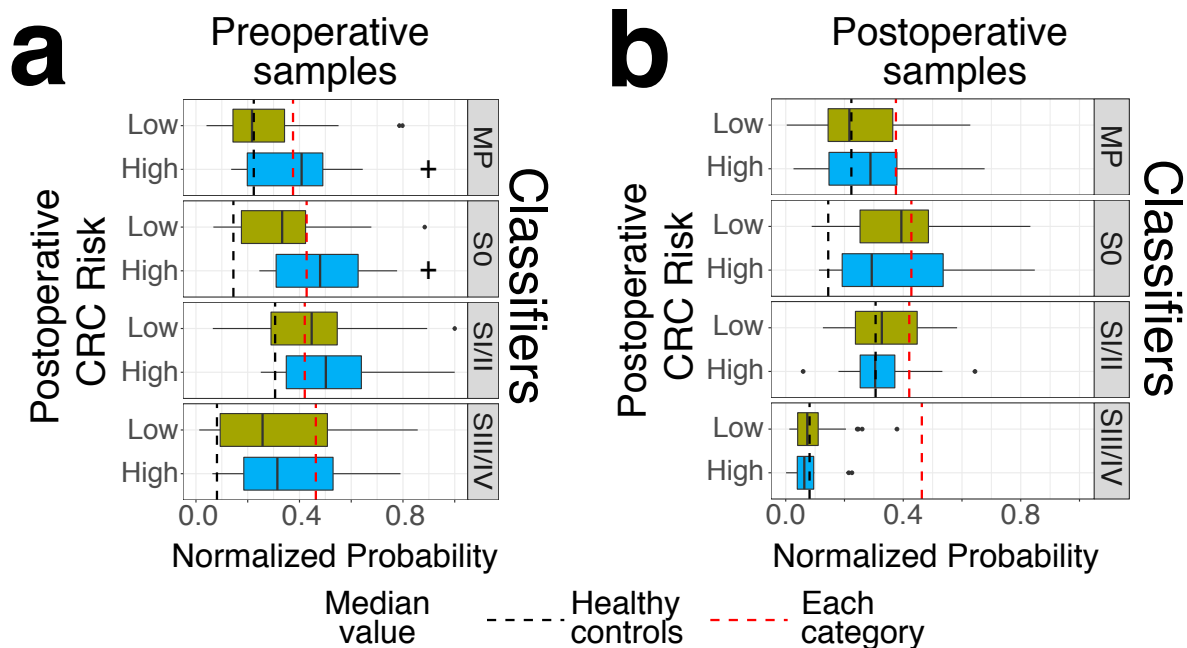


図 7-3 大腸がん患者の外科的治療前後の補正確率の箱ひげ図

多段階発がんのいずれかの状態の患者(MP: 多発性腺腫の患者; S0: Stage 0 の大腸がん患者; SI/II: Stage I/II の大腸がん患者; SIII/IV: Stage III/IV の大腸がん患者)と健常者を分類する 4 つの判別器に、大腸がんのリスクが低い患者(Low; 黄色)と高い患者(High; 青色)の外科治療前(a)と外科的治療後(b)の腸内環境の組成データを適用し、得られた補正確率を箱ひげ図で示した。横線はそれぞれの判別器の構築に用いた健常者(黒色)とそれ以外の患者(赤色)の補正確率を表している。統計学的に補正確率が高いことが確認された場合、以下のように表記している($P < 0.05$; +)。

7.3.2 大腸がんのリスクを反映する可能性のある特徴量の探索

健常者と多発性腺腫の患者、健常者と Stage 0 の大腸がん患者を分類する判別器に大腸がん患者の外科的治療前の腸内環境の組成データを適用することで、外科的治療後の大腸がんのリスクを推定できる可能性が示唆された。本説では、それぞれの判別器の構築に寄与した特徴量を調査することで、大腸がんのリスクを反映する可能性のある特徴量を探索した。

まず、健常者と多発性腺腫の患者を分類する判別器の構築に寄与した特徴量を調査した。その結果、アラニン(Ala)やスレオニン(Thr)等のアミノ酸や、胆汁酸の 1 種であるデオキシコール酸(DCA)、胆汁酸関連の代謝産物であるタウリン(Taurine)の寄与率が高いことが確認された(図 7-4)。これらの代謝産物は大腸がんのリスクを反映する可能性があると考えられる。寄与率が高い特徴量として、発がん性を有することが報告されているデオキシ

コール酸[31]が含まれていることも、この仮説を支持していると考えられる。デオキシコール酸だけではなく、外科的治療前後でその量が統計学的に有意に減少しなかったアミノ酸も大腸がんのリスクを反映する可能性があるが、大腸がんの発症や進行との関連性が報告されていないため、その詳細なメカニズムを追求することはできなかった。

次に、健常者と Stage 0 の大腸がん患者を分類する判別器の構築に寄与した特徴量を調査した。その結果、グルタミン酸(Glu)やチロシン(Tyr)等のアミノ酸や、ペプチド分解酵素である gamma-glutamyl transpeptidase(*ggt*, K00681)の寄与率が高いことが確認された(図 7-4)。これらの代謝産物も大腸がんのリスクを反映している可能性があると考えられる。しかしながら、これらの代謝産物も大腸がんの発症や進行との関連性が報告されていないため、その関連性を追求することができなかった。今後、これらの代謝産物が大腸がんの細胞にどのような影響を及ぼすのか実験することで、これらの代謝産物と大腸がんの関係性を明らかにする必要があると考えられる。

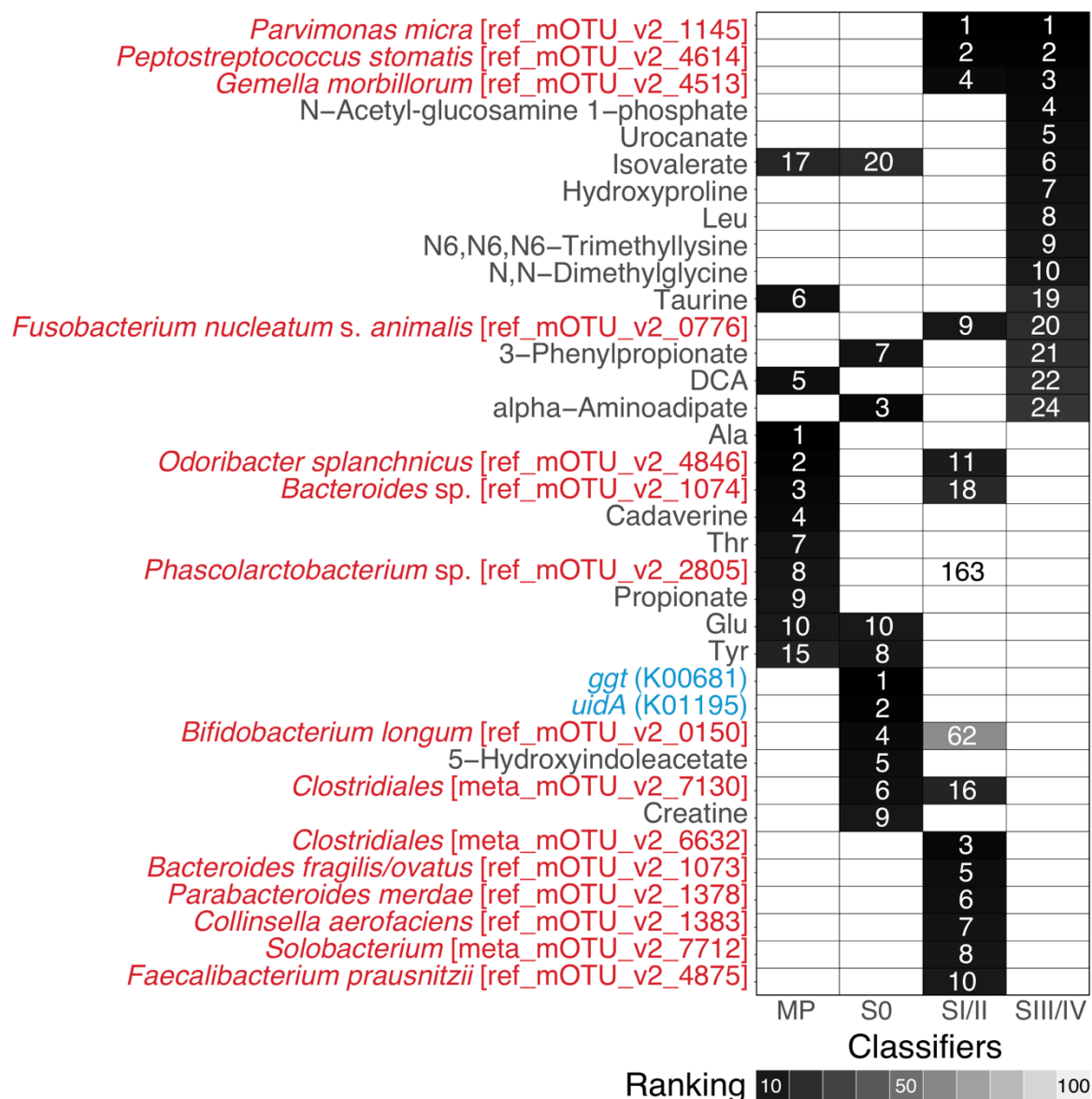


図 7-4 判別器の構築に寄与した上位 10 個の特徴量を表したヒートマップ

横軸は、多段階発がんのいずれかの状態の患者(MP: 多発性腺腫の患者; S0: Stage 0 の大腸がん患者; SI/II: Stage I/II の大腸がん患者; SIII/IV: Stage III/IV の大腸がん患者)と健常者を分類する 4 つの判別器の種類、縦軸は特徴量を示している。数字はそれぞれの判別器におけるそれぞれの特徴量をその寄与率を高い順に並び替えたときの順位を示しており、色はその順位を反映している(凡例参照)。縦軸の特徴量の色は、特徴量の種類を表しており、赤色は腸内細菌、青色は遺伝子(KO)、黒色は代謝産物を示している。

第8章 総括

本研究では、外科的治療後の大腸がんに関与している可能性がある腸内細菌や代謝産物を特定することを目的として、85 人の大腸がん患者の外科的治療前と治療から約 1 年後の便を取得し、その便に含まれている腸内細菌や代謝産物を統計学的に比較することにより、外科的治療が腸内環境へ及ぼす影響を明らかにすること、さらには、外科的治療後の患者の大腸がんのリスクを推定する手法を確立することを目指した。

大腸がん外科的治療前後の患者の腸内細菌叢や代謝産物の組成は大きく変化したことが確認された。Stage III/IV の大腸がん患者と健常者の分類に寄与しているバイオマーカーである *F. nucleatum* や *P. micra* 等の腸内細菌の相対存在量や、アミノ酸の 1 種であるセリン等の代謝産物の量は外科的治療後に減少したことが確認された。一方で、コール酸や発がん性が報告されているデオキシコール酸等の胆汁酸の量と、コール酸からデオキシコール酸に変換する遺伝子である *bai* オペロンとその変換の担い手である *C. scindens* の相対存在量は外科的治療後に増加したことが確認された。また、*C. scindens* の推定された生育速度も外科的治療後の患者で高いことが確認された。以上の結果から、大腸がんの進行に起因する腸内環境は外科的治療により改善されたが、腸内細菌に由来する大腸がんのリスクが高い可能性が示唆された。

外科的治療前後で胆汁酸代謝が変化したメカニズムを解明するために、20 人の右側大腸がん患者と 65 人の左側大腸がん患者、それぞれで外科的治療前後に変化した胆汁酸代謝に関連した腸内細菌や代謝産物を特定した。右側大腸がん外科的治療後の患者は、胆汁酸の再吸収を担う回腸末端を一部切除したことにより、多くの胆汁酸が回腸末端で再吸収されずに大腸内に流入したため、定量できた全ての胆汁酸の量が外科的治療後に増加したことが示唆された。一方で、左側大腸がん患者は、外科的治療によって大腸がんの病変のみを切除したため、腸内環境は変化しないことが予想されたが、胆汁酸代謝に係る腸内細菌や代謝産物が大きく変化したことが確認された。左側大腸がん外科的治療後の患者で便中の総胆汁酸やコール酸の量が増加したメカニズムを明らかにすることができなかったが、コール酸の量が多くなったため *C. scindens* の活性が高くなった結果、デオキシコール酸の量が増加した可能性が示唆された。そのため、大腸内でコール酸を減少させることは大腸内のデオキシコール酸を減少させるための介入のターゲットの 1 つになり得ることが示唆された。腸内細菌の代謝によりタウロコール酸またはグリココール酸はコール酸に変換されるが、その変換を担う遺伝子の相対存在量は変化していなかったため、この遺伝子の発現量を定量することで、この反応を担う腸内細菌を特定する研究が重要であると考えられる。

外科的治療後の大腸がんのリスクを推定する手法を確立するために、腸内環境の組成(腸内細菌の系統組成や遺伝子機能組成、代謝産物の組成)データを用いて大腸がんのリスクを推定する手法を開発した。まず、外科的治療から約 5 年間で大腸がんの前がん病変である腺腫や大腸がんを発症したかどうかを基に、大腸がん患者を大腸がんのリスクが低い患者と高い患者に分類した。次に、健常者と多発性腺腫の患者、健常者と Stage 0 の大腸がん患者の腸内環境の組成データからそれらを分類する判別器を構築し、大腸がんのリスクが低い患者と高い患者の外科的治療前の腸内環境の組成データを適用した。最後に、判別器から得られた確率を大腸がんのリスクが低い患者と高い患者で比較し、大腸がんのリスクが低い患者と比較して高い患者でその確率が高いことが確認された。このことから、これらの判別器に大腸がん患者の外科的治療前の腸内環境の組成データを適用することで、外科的治療後の大腸がんのリスクを推定できる可能性が示唆された。

最後に、この研究の今後の展望について述べる。まず、本研究によって得られた結果を検証するためのコホート研究が必要である。患者を選抜する過程でのバイアスや、代謝産物の定量手法、統計学的な解析手法等の違いにより、本研究で得られた結果が他の機関で実施された研究では確認されない可能性がある。そのため、本研究とは独立した機関で、大腸がん外科的治療前後の患者を対象とした新たなコホート研究を実施する必要があると考えられる。次に、本研究で構築した大腸がんのリスクを推定する手法を検証するための大規模な前向きコホート研究が必要であると考えられる。本研究では、外科的治療から約 5 年間に大腸がんの前がん病変である腺腫や大腸がんを発症したかどうかを基に、外科的治療後の大腸がんのリスクを定量化できたが、対象とした患者数が少ないため、対象とする患者数を増やして再解析した場合に同じ結果が得られない可能性がある。そのため、対象とする患者数を増やし、引き続き前向きコホート研究を実施することが重要であると考えられる。

謝辞

本研究を遂行するにあたり、お世話になった方々にこの場を借りて感謝申し上げます。

指導教官として幅広いご指導・ご鞭撻をして頂きました東京工業大学の山田 拓司准教授に心より感謝致します。山田准教授には、**CapsLock** キーを **Tab** に置換する等の研究を遂行するうえでの効率化の技術や、英文の論文の執筆方法、データの可視化、プレゼンテーションの資料の作成等多くのことをご教授して頂きました。また、ヒトの臨床学的なデータを対象とした共同研究の機会も提供して頂きました。心より感謝申し上げます。

本研究の共同研究先の方々にも、研究に関する多くのことをご指導して頂きました。メタゲノムデータの取得に関しては、被験者の便の取得をして頂きました谷内田 真一教授(当時)、柴 知史博士をはじめとする国立がん研究センターの方々に感謝致します。便中の代謝産物の組成データの取得に関しては、代謝産物の定量をして頂いた慶應義塾大学の福田 真嗣特任教授に感謝致します。谷内田教授には、研究を遂行するために多くの議論をして頂きました。柴博士と福田教授にはコロナ禍にも関わらず、被験者の疫学データの調査や代謝産物の定量をして頂きました。心より感謝致します。

研究室に所属しているの方々にも、研究に関する多くのことをご指導頂きました。水谷 紗弥佳博士には、統計処理や機械学習、メタゲノムデータと代謝産物を統合するマルチオミクス解析の基礎についてご指導頂き、感謝の念に堪えません。英文の論文執筆に関してご指導して頂いた Pande Putu Erawijantari 博士(当時)に感謝致します。研究結果に関する詳細な議論をして頂いた中川 善一氏に感謝申し上げます。

研究室のスタッフには、研究を遂行するための環境をサポートして頂きました。研究室の計算機資源を管理して頂いた土倉 大尚氏に感謝致します。また、本研究を学会で発表するにあたり、宿泊先の手続きをして頂きました秘書の渡辺 夕紀子氏に感謝致します。

最後に、学生生活を献身的に支えて頂いた両親と婚約者である山里 美希氏に心より感謝致します。

参考文献

- 1 Fan Y, Pedersen O. Gut microbiota in human metabolic health and disease. *Nat Rev Microbiol* 2020;**19**:55–71.
- 2 Clemente JC, Ursell LK, Parfrey LW, *et al.* The impact of the gut microbiota on human health: An integrative view. *Cell* 2012;**148**:1258–1270.
- 3 Duvallet C, Gibbons SM, Gurry T, *et al.* Meta-analysis of gut microbiome studies identifies disease-specific and shared responses. *Nat Commun* 2017;**8**:1784.
- 4 Yachida S, Mizutani S, Shiroma H, *et al.* Metagenomic and metabolomic analyses reveal distinct stage-specific phenotypes of the gut microbiota in colorectal cancer. *Nat Med* 2019;**25**:968–976.
- 5 Walker AW, Duncan SH, Louis P, *et al.* Phylogeny, culturing, and metagenomics of the human gut microbiota. *Trends Microbiol.* 2014;**22**:267–274.
- 6 Stewart EJ. Growing unculturable bacteria. *J. Bacteriol.* 2012;**194**:4151–4160.
- 7 Béjà O, Suzuki MT, Koonin E V., *et al.* Construction and analysis of bacterial artificial chromosome libraries from a marine microbial assemblage. *Environ Microbiol* 2000;**2**:516–529.
- 8 Thomas AM, Manghi P, Asnicar F, *et al.* Metagenomic analysis of colorectal cancer datasets identifies cross-cohort microbial diagnostic signatures and a link with choline degradation. *Nat Med* 2019;**25**:667–678.
- 9 Wirbel J, Pyl PT, Kartal E, *et al.* Meta-analysis of fecal metagenomes reveals global microbial signatures that are specific for colorectal cancer. *Nat Med* 2019;**25**:679–689.
- 10 Li J, Jia H, Cai X, *et al.* An integrated catalog of reference genes in the human gut microbiome. *Nat Biotechnol* 2014;**32**:834–841.
- 11 Alneberg J, Bjarnason BS, De Bruijn I, *et al.* Binning metagenomic contigs by coverage and composition. *Nat Methods* 2014;**11**:1144–1146.
- 12 Wu YW, Simmons BA, Singer SW. MaxBin 2.0: An automated binning algorithm to recover genomes from multiple metagenomic datasets. *Bioinformatics* 2016;**32**:605–607.
- 13 Kang D, Li F, Kirton ES, *et al.* MetaBAT 2 : an adaptive binning algorithm for robust and efficient genome reconstruction from metagenome assemblies. *PeerJ* 2019;**26**:e7359.

- 14 Pasolli E, Asnicar F, Manara S, *et al.* Extensive Unexplored Human Microbiome Diversity Revealed by Over 150,000 Genomes from Metagenomes Spanning Age, Geography, and Lifestyle. *Cell* 2019;**176**:649-662.e20.
- 15 Forster SC, Kumar N, Anonye BO, *et al.* A human gut bacterial genome and culture collection for improved metagenomic analyses. *Nat Biotechnol* 2019;**37**:186–192.
- 16 Almeida A, Mitchell AL, Boland M, *et al.* A new genomic blueprint of the human gut microbiota. *Nature* 2019;**568**:499–504.
- 17 Nayfach S, Shi ZJ, Seshadri R, *et al.* New insights from uncultivated genomes of the global human gut microbiome. *Nature* 2019;**568**:505–510.
- 18 Fearon ER, Vogelstein B. A genetic model for colorectal tumorigenesis. *Cell* 1990;**61**:759–767.
- 19 B V, ER F, SR H, *et al.* Genetic alterations during colorectal-tumor development. *N Engl J Med* 1988;**319**:525–532.
- 20 Yamagishi H, Kuroda H, Imai Y, *et al.* Molecular pathogenesis of sporadic colorectal cancers. *Chin. J. Cancer.* 2016;**35**:4.
- 21 Lichtenstein P, Holm N V., Verkasalo PK, *et al.* Environmental and Heritable Factors in the Causation of Cancer — Analyses of Cohorts of Twins from Sweden, Denmark, and Finland. *N Engl J Med* 2000;**343**:78–85.
- 22 Johnson CM, Wei C, Ensor JE, *et al.* Meta-Analyses of colorectal cancer risk factors. *Cancer Causes Control* 2013;**24**:1207–1222.
- 23 Huxley RR, Ansary-Moghaddam A, Clifton P, *et al.* The impact of dietary and lifestyle risk factors on risk of colorectal cancer: A quantitative overview of the epidemiological evidence. *Int J Cancer* 2009;**125**:171–180.
- 24 Song M, Chan AT. Environmental Factors, Gut Microbiota, and Colorectal Cancer Prevention. *Clin. Gastroenterol. Hepatol.* 2019;**17**:275–289.
- 25 Tjalsma H, Boleij A, Marchesi JR, *et al.* A bacterial driver–passenger model for colorectal cancer: beyond the usual suspects. *Nat Rev Microbiol* 2012;**10**:575-582.
- 26 Hirayama A, Kami K, Sugimoto M, *et al.* Quantitative Metabolome Profiling of Colon and Stomach Cancer Microenvironment by Capillary Electrophoresis Time-of-Flight Mass Spectrometry. *Cancer Res* 2009;**69**:4918–4926.
- 27 Mal M, Koh PK, Cheah PY, *et al.* Metabotyping of human colorectal cancer using two-dimensional gas chromatography mass spectrometry. *Anal Bioanal Chem* 2012;**403**:483–493.

- 28 Attene-Ramos MS, Wagner ED, Gaskins HR, *et al.* Hydrogen sulfide induces direct radical-associated DNA damage. *Mol Cancer Res* 2007;**5**:455–459.
- 29 Attene-Ramos MS, Nava GM, Muellner MG, *et al.* DNA damage and toxicogenomic analyses of hydrogen sulfide in human intestinal epithelial FHs 74 int cells. *Environ Mol Mutagen* 2010;**51**:304–314.
- 30 Yazici C, Wolf PG, Kim H, *et al.* Race-dependent association of sulfidogenic bacteria with colorectal cancer. *Gut* 2017;**66**:1983–1994.
- 31 Yoshimoto S, Loo TM, Atarashi K, *et al.* Obesity-induced gut microbial metabolite promotes liver cancer through senescence secretome. *Nature* 2013;**499**:97–101.
- 32 Ajouz H, Mukherji D, Shamseddine A. Secondary bile acids: An underrecognized cause of colon cancer. *World J. Surg. Oncol.* 2014;**12**:164.
- 33 Levin B, Lieberman DA, McFarland B, *et al.* Screening and Surveillance for the Early Detection of Colorectal Cancer and Adenomatous Polyps, 2008: A Joint Guideline from the American Cancer Society, the US Multi-Society Task Force on Colorectal Cancer, and the American College of Radiology. *CA Cancer J Clin* 2008;**58**:130–160.
- 34 Tanaka S, Kashida H, Saito Y, *et al.* Japan Gastroenterological Endoscopy Society guidelines for colorectal endoscopic submucosal dissection/endoscopic mucosal resection. *Dig Endosc* 2020;**32**:219–239.
- 35 Al Bandar MH, Kim NK. Current status and future perspectives on treatment of liver metastasis in colorectal cancer (Review). *Oncol. Rep.* 2017;**37**:2553–2564.
- 36 Kakeji Y, Takahashi A, Udagawa H, *et al.* Surgical outcomes in gastroenterological surgery in Japan: Report of National Clinical database 2011–2016. *Ann Gastroenterol Surg* 2018;**2**:37–54.
- 37 Mulder SA, Kranse R, Damhuis RA, *et al.* The Incidence and Risk Factors of Metachronous Colorectal Cancer. *Dis Colon Rectum* 2012;**55**:522–531.
- 38 Yabuuchi Y, Imai K, Hotta K, *et al.* Higher incidence of metachronous advanced neoplasia in patients with synchronous advanced neoplasia and left-sided colorectal resection for colorectal cancer. *Gastrointest Endosc* 2018;**88**:348–359.e1.
- 39 Bouvier AM, Latournerie M, Jooste V, *et al.* The lifelong risk of metachronous colorectal cancer justifies long-term colonoscopic follow-up. *Eur J Cancer* 2008;**44**:522–527.

- 40 Yun JA, Huh JW, Kim HC, *et al.* Local recurrence after curative resection for rectal carcinoma: The role of surgical resection. *Medicine (Baltimore)* 2016;**95**:e3942.
- 41 Ohigashi S, Sudo K, Kobayashi D, *et al.* Significant Changes in the Intestinal Environment After Surgery in Patients with Colorectal Cancer. *J Gastrointest Surg* 2013;**17**:1657–1664.
- 42 Sze MA, Baxter NT, Ruffin MT, *et al.* Normalization of the microbiota in patients after treatment for colonic lesions. *Microbiome* 2017;**5**:150.
- 43 Zeller G, Tap J, Voigt AY, *et al.* Potential of fecal microbiota for early-stage detection of colorectal cancer. *Mol Syst Biol* 2014;**10**:766.
- 44 Feng Q, Liang S, Jia H, *et al.* Gut microbiome development along the colorectal adenoma–carcinoma sequence. *Nat Commun* 2015;**6**:6528.
- 45 Vogtmann E, Hua X, Zeller G, *et al.* Colorectal cancer and the Human gut microbiome: Reproducibility with Whole-Genome Shotgun Sequencing. *PLoS One* 2016;**11**:e0155362.
- 46 Yu J, Feng Q, Wong SH, *et al.* Metagenomic analysis of faecal microbiome as a tool towards targeted non-invasive biomarkers for colorectal cancer. *Gut* 2017;**66**:70–78.
- 47 Soga T, Baran R, Suematsu M, *et al.* Differential metabolomics reveals ophthalmic acid as an oxidative stress biomarker indicating hepatic glutathione consumption. *J Biol Chem* 2006;**281**:16768–16776.
- 48 Soga T, Igarashi K, Ito C, *et al.* Metabolomic profiling of anionic metabolites by capillary electrophoresis mass spectrometry. *Anal Chem* 2009;**81**:6165–6174.
- 49 Ishii C, Nakanishi Y, Murakami S, *et al.* A metabologenomic approach reveals changes in the intestinal environment of mice fed on american diet. *Int J Mol Sci* 2018;**19**:4079.
- 50 Voigt AY, Costea PI, Kultima JR, *et al.* Temporal and technical variability of human gut metagenomes. *Genome Biol* 2015;**16**:73.
- 51 Nagata N, Tohya M, Fukuda S, *et al.* Effects of bowel preparation on the human gut microbiome and metabolome. *Sci Rep* 2019;**9**:4042.
- 52 Loupakakis F, Hurwitz HI, Saltz L, *et al.* Impact of primary tumour location on efficacy of bevacizumab plus chemotherapy in metastatic colorectal cancer. *Br J Cancer* 2018;**119**:1451–1455.

- 53 Nishimoto Y, Mizutani S, Nakajima T, *et al.* High stability of faecal microbiome composition in guanidine thiocyanate solution at room temperature and robustness during colonoscopy. *Gut*. 2016;**65**:1574–1575.
- 54 Langmead B, Salzberg SL. Fast gapped-read alignment with Bowtie 2. *Nat Methods* 2012;**9**:357–360.
- 55 Martin M. Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet.journal* 2011;**17**:10–12.
- 56 Shen W, Le S, Li Y, *et al.* SeqKit : A Cross-Platform and Ultrafast Toolkit for FASTA / Q File Manipulation. *PLoS One* 2016;**11**:1–10.
- 57 Nurk S, Meleshko D, Korobeynikov A, *et al.* MetaSPAdes: A new versatile metagenomic assembler. *Genome Res* 2017;**27**:824–834.
- 58 Li D, Liu CM, Luo R, *et al.* MEGAHIT: An ultra-fast single-node solution for large and complex metagenomics assembly via succinct de Bruijn graph. *Bioinformatics* 2015;**31**:1674–1676.
- 59 Uritskiy G V, Diruggiero J, Taylor J. MetaWRAP — a flexible pipeline for genome-resolved metagenomic data analysis. *Microbiome* 2018;**6**:1–13.
- 60 Song WZ, Thomas T. Binning-refiner: Improving genome bins through the combination of different binning programs. *Bioinformatics* 2017;**33**:1873–1875
- 61 Parks DH, Imelfort M, Skennerton CT, *et al.* CheckM: Assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes. *Genome Res* 2015;**25**:1043–1055.
- 62 Bowers RM, Kyrpides NC, Stepanauskas R, *et al.* Minimum information about a single amplified genome (MISAG) and a metagenome-assembled genome (MIMAG) of bacteria and archaea. *Nat. Biotechnol.* 2017;**35**:725–731.
- 63 Seemann T. Prokka: rapid prokaryotic genome annotation. *Bioinformatics* 2014;**30**:2068–2069.
- 64 Fu L, Niu B, Zhu Z, *et al.* CD-HIT: Accelerated for clustering the next-generation sequencing data. *Bioinformatics* 2012;**28**:3150–3152.
- 65 Olm MR, Brown CT, Brooks B, *et al.* DRep: A tool for fast and accurate genomic comparisons that enables improved genome recovery from metagenomes through de-replication. *ISME J* 2017;**11**:2864–2868.
- 66 Ondov BD, Treangen TJ, Melsted P, *et al.* Mash: Fast genome and metagenome distance estimation using MinHash. *Genome Biol* 2016;**17**:132.

- 67 Kurtz S, Phillippy A, Delcher AL, *et al.* Versatile and open software for comparing large genomes. *Genome Biol* 2004;**5**:R12.
- 68 Chaumeil P-A, Mussig AJ, Hugenholtz P, *et al.* GTDB-Tk: a toolkit to classify genomes with the Genome Taxonomy Database. *Bioinformatics* 2019;**36**:1925–1927.
- 69 Jain C, Rodriguez-R LM, Phillippy AM, *et al.* High throughput ANI analysis of 90K prokaryotic genomes reveals clear species boundaries. *Nat Commun* 2018;**9**:1–8.
- 70 Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* 2009;**25**:1754–1760.
- 71 Buchfink B, Xie C, Huson DH. Fast and sensitive protein alignment using DIAMOND. *Nat. Methods*. 2014;**12**:59–60.
- 72 Kanehisa M, Goto S. KEGG: Kyoto Encyclopedia of Genes and Genomes. *Nucleic Acids Res*. 2000;**28**:27–30.
- 73 Wood DE, Lu J, Langmead B. Improved metagenomic analysis with Kraken 2. *Genome Biol* 2019;**20**:257.
- 74 Ye SH, Siddle KJ, Park DJ, *et al.* Benchmarking Metagenomics Tools for Taxonomic Classification. *Cell*. 2019;**178**:779–794.
- 75 Milanese A, Mende DR, Paoli L, *et al.* Microbial abundance, activity and population genomic profiling with mOTUs2. *Nat Commun* 2019;**10**:1–11.
- 76 Knight R, Vrbanac A, Taylor BC, *et al.* Best practices for analysing microbiomes. *Nat. Rev. Microbiol*. 2018;**16**:410–422.
- 77 Korem T, Zeevi D, Suez J, *et al.* Growth dynamics of gut microbiota in health and disease inferred from single metagenomic samples. *Science* 2015;**349**:1101–1106.
- 78 Brown CT, Olm MR, Thomas BC, *et al.* Measurement of bacterial replication rates in microbial communities. *Nat Biotechnol* 2016;**34**:1256–1263.
- 79 Gao Y, Li H. Quantifying and comparing bacterial growth dynamics in multiple metagenomic samples. *Nat Methods* 2018;**15**:1041–1044.
- 80 Suzuki S, Yamada T. Probabilistic model based on circular statistics for quantifying coverage depth dynamics originating from DNA replication. *PeerJ* 2020;**8**:e8722.
- 81 Emiola A, Oh J. High throughput in situ metagenomic measurement of bacterial replication at ultra-low sequencing coverage. *Nat Commun* 2018;**9**:1–8.

- 82 Funabashi M, Grove TL, Wang M, *et al.* A metabolic pathway for bile acid dehydroxylation by the gut microbiome. *Nature* 2020;**582**:566–570.
- 83 Katoh K, Standley DM. MAFFT multiple sequence alignment software version 7: Improvements in performance and usability. *Mol Biol Evol* 2013;**30**:772–780.
- 84 Capella-Gutiérrez S, Silla-Martínez JM, Gabaldón T. trimAl: A tool for automated alignment trimming in large-scale phylogenetic analyses. *Bioinformatics* 2009;**25**:1972–1973.
- 85 Eddy SR. Accelerated profile HMM searches. *PLoS Comput Biol* 2011;**7**:e1002195.
- 86 Camacho C, Coulouris G, Avagyan V, *et al.* BLAST+: Architecture and applications. *BMC Bioinformatics* 2009;**10**:421.
- 87 Lewis JD, Chen EZ, Robert N, *et al.* Inflammation, Antibiotics, and Diet as Environmental Stressors of the Gut Microbiome in Pediatric Crohn's Disease. *Cell Host Microbe* 2015;**18**:489–500.
- 88 Ridlon JM, Alves JM, Hylemon PB, *et al.* Cirrhosis, bile acids and gut microbiota: Unraveling a complex relationship. *Gut Microbes* 2013;**4**:382.
- 89 Li YY, Ge QX, Cao J, *et al.* Association of *Fusobacterium nucleatum* infection with colorectal cancer in Chinese patients. *World J Gastroenterol* 2016;**22**:3227–3233.
- 90 Kostic AD, Gevers D, Pedamallu CS, *et al.* Genomic analysis identifies association of *Fusobacterium* with colorectal carcinoma. *Genome Res* 2012;**22**:292–298.
- 91 Devkota S, Wang Y, Musch MW, *et al.* Dietary-fat-induced taurocholic acid promotes pathobiont expansion and colitis in *Il10^{-/-}* mice. *Nature*. 2012;**487**:104–108.
- 92 Devlin AS, Fischbach MA. A biosynthetic pathway for a prominent class of microbiota-derived bile acids. *Nat Chem Biol* 2015;**11**:685–690.
- 93 Ohshima K, Nojima S, Tahara S, *et al.* Serine racemase enhances growth of colorectal cancer by producing pyruvate from serine. *Nat Metab* 2020;**2**:81–96.
- 94 Kenny DJ, Plichta DR, Shungin D, *et al.* Cholesterol Metabolism by Uncultured Human Gut Bacteria Influences Host Cholesterol Level. *Cell Host Microbe* 2020;**28**:245–257.
- 95 Ticho AL, Malhotra P, Dudeja PK, *et al.* Intestinal absorption of bile acids in health and disease. *Compr Physiol* 2020;**10**:21–56.
- 96 Hope C, Reilly J, Lund J, *et al.* Systematic review: the effect of right hemicolectomy for cancer on postoperative bowel function. *Support. Care Cancer*. 2020;**28**:4549.

- 97 Mekjian HS, Phillips SF, Hofmann AF. Colonic secretion of water and electrolytes induced by bile acids: perfusion studies in man. *J Clin Invest* 1971;**50**:1569–1577.
- 98 Fornaro R, Belcastro E, Lo Presti G, *et al.* The ileocecal valve as a prognostic factor in extensive resection of the small intestine. *Chir Ital* 1991;**43**:49–54.
- 99 Bertelsen CA, Larsen HM, Neuenschwander AU, *et al.* Long-term functional outcome after right-sided complete mesocolic excision compared with conventional colon cancer surgery: A population-based questionnaire study. *Dis Colon Rectum* 2018;**61**:1063–1072.
- 100 Thomas LA, Veysey MJ, Bathgate T, *et al.* Mechanism for the transit-induced increase in colonic deoxycholic acid formation in cholesterol cholelithiasis. *Gastroenterology* 2000;**119**:806–815.
- 101 Breiman L. RANDOM FORESTS. *Mach Learn* 2001;**45**:5–32.

付録

付録 1 公開されているデータの Accession number

Metagenome samples	Participant IDs	Cohorts	Accession number
bugkiller-11-0-0	bugkiller	Voigt et al.	ERS652544
bugkiller-11-2-0	bugkiller	Voigt et al.	ERS652545
bugkiller-11-60-0	bugkiller	Voigt et al.	ERS652549
bugkiller-11-392-1	bugkiller	Voigt et al.	ERS652550
bugkiller-11-600-0	bugkiller	Voigt et al.	ERS652556
daisy-11-0-0	daisy	Voigt et al.	ERS652589
daisy-11-2-0	daisy	Voigt et al.	ERS652590
daisy-11-7-0	daisy	Voigt et al.	ERS652591
halbarad-11-0-0	halbarad	Voigt et al.	ERS652579
halbarad-11-2-0	halbarad	Voigt et al.	ERS652580
peacemaker-11-0-0	peacemaker	Voigt et al.	ERS652558
peacemaker-11-2-0	peacemaker	Voigt et al.	ERS652559
peacemaker-11-7-0	peacemaker	Voigt et al.	ERS652560
peacemaker-11-60-0	peacemaker	Voigt et al.	ERS652563
peacemaker-11-392-1	peacemaker	Voigt et al.	ERS652564
peacemaker-11-600-0	peacemaker	Voigt et al.	ERS652570
scavenger-11-0-0	scavenger	Voigt et al.	ERS652572
scavenger-11-2-0	scavenger	Voigt et al.	ERS652573
scavenger-11-7-0	scavenger	Voigt et al.	ERS652574
scavenger-11-60-0	scavenger	Voigt et al.	ERS652577
tigress-11-0-0	tigress	Voigt et al.	ERS652585
tigress-11-2-0	tigress	Voigt et al.	ERS652586
tigress-11-7-0	tigress	Voigt et al.	ERS652587

付録 2 公開されている *bai* 遺伝子の UniProt の Accession number

Gene	Accession number
<i>baiA</i>	Q9RB46_9FIRM
<i>baiA</i>	B4YST2_9FIRM
<i>baiA</i>	BAIA2_CLOSV
<i>baiB</i>	Q9RB49_9FIRM
<i>baiB</i>	B4YST9_9FIRM
<i>baiB</i>	BAIB_CLOSV
<i>baiCD</i>	Q9RB48_9FIRM
<i>baiCD</i>	B4YSU0_9FIRM
<i>baiCD</i>	BAICD_CLOSV
<i>baiE</i>	Q9RB47_9FIRM
<i>baiE</i>	B4YSU1_9FIRM
<i>baiE</i>	BAIE_CLOSV
<i>baiE</i>	A0A0M1UWR2_PAESO
<i>baiF</i>	Q9RB45_9FIRM
<i>baiF</i>	B4YSU2_9FIRM
<i>baiF</i>	BAIF_CLOSV
<i>baiG</i>	Q9RB44_9FIRM
<i>baiG</i>	B4YSU3_9FIRM
<i>baiG</i>	BAIG_CLOSV
<i>baiH</i>	A5A8R6_9FIRM
<i>baiH</i>	B4YSU4_9FIRM
<i>baiH</i>	BAIH_CLOSV
<i>baiI</i>	B4YSU5_9FIRM
<i>baiI</i>	BAII_CLOSV

付録 3 大腸がん外科的治療後に減少した腸内細菌

Species	P value
Parvimonas micra [ref_mOTU_v2_1145]	2.27x10 ⁻¹³
Prevotella copri [ref_mOTU_v2_4448]	6.55x10 ⁻¹¹
unknown Clostridiales [meta_mOTU_v2_5411]	2.96x10 ⁻⁹
Gemella morbillorum [ref_mOTU_v2_4513]	1.39x10 ⁻⁸
Fusobacterium nucleatum s. animalis [ref_mOTU_v2_0776]	2.09x10 ⁻⁷
unknown Faecalibacterium [meta_mOTU_v2_6631]	2.79x10 ⁻⁶
unknown Clostridiales [meta_mOTU_v2_5669]	2.84x10 ⁻⁶
Coprococcus catus [ref_mOTU_v2_4874]	3.29x10 ⁻⁶
Peptostreptococcus stomatis [ref_mOTU_v2_4614]	3.66x10 ⁻⁶
Bacteroides coprocola [ref_mOTU_v2_4312]	3.81x10 ⁻⁶
unknown Dialister [meta_mOTU_v2_5867]	3.81x10 ⁻⁶
Dorea longicatena [ref_mOTU_v2_2893]	6.48x10 ⁻⁶
unknown Clostridiales [meta_mOTU_v2_7061]	1.34x10 ⁻⁵
bacterium LF-3 [ref_mOTU_v2_3608]	1.53x10 ⁻⁵
unknown Clostridiales [meta_mOTU_v2_5712]	1.56x10 ⁻⁵
unknown Clostridiales [meta_mOTU_v2_6832]	2.04x10 ⁻⁵
unknown Eubacterium [meta_mOTU_v2_6657]	2.08x10 ⁻⁵
Faecalibacterium prausnitzii [ref_mOTU_v2_4211]	2.29x10 ⁻⁵
unknown Clostridiales [meta_mOTU_v2_7093]	2.58x10 ⁻⁵
unknown Ruminococcaceae [meta_mOTU_v2_6905]	2.62x10 ⁻⁵
Holdemanella biformis [ref_mOTU_v2_4395]	3.05x10 ⁻⁵
unknown Clostridiales [meta_mOTU_v2_6049]	3.24x10 ⁻⁵
Fusobacterium periodonticum [ref_mOTU_v2_1402]	4.01x10 ⁻⁵
unknown Clostridiales [meta_mOTU_v2_7031]	4.31x10 ⁻⁵
unknown Clostridiales [meta_mOTU_v2_7014]	4.77x10 ⁻⁵
unknown Clostridiales [meta_mOTU_v2_6632]	4.87x10 ⁻⁵
Fusobacterium nucleatum s. polymorphum [ref_mOTU_v2_0131]	6.10x10 ⁻⁵
Fusobacterium nucleatum s. vincentii [ref_mOTU_v2_0754]	6.10x10 ⁻⁵
unknown Clostridiales [meta_mOTU_v2_6787]	6.30x10 ⁻⁵
Clostridium spiroforme [ref_mOTU_v2_4235]	6.45x10 ⁻⁵
unknown Clostridiales [meta_mOTU_v2_5806]	6.87x10 ⁻⁵
Oscillibacter sp. 57_20 [meta_mOTU_v2_5351]	8.63x10 ⁻⁵
Oscillibacter sp. ER4 [ref_mOTU_v2_3624]	8.64x10 ⁻⁵
uncultured Clostridium sp. [meta_mOTU_v2_6381]	1.02x10 ⁻⁴
unknown Clostridiales [meta_mOTU_v2_6916]	1.41x10 ⁻⁴
unknown Clostridiales [meta_mOTU_v2_5603]	1.64x10 ⁻⁴
Ruminococcus torques [ref_mOTU_v2_4718]	1.96x10 ⁻⁴
Eubacterium ramulus [ref_mOTU_v2_2795]	2.33x10 ⁻⁴
Megamonas funiformis/rupeellensis [ref_mOTU_v2_0502]	2.42x10 ⁻⁴
unknown Clostridiales [meta_mOTU_v2_7527]	3.32x10 ⁻⁴
unknown Clostridiales [meta_mOTU_v2_6808]	3.46x10 ⁻⁴
unknown Tyzzerella [meta_mOTU_v2_5947]	3.64x10 ⁻⁴
Eubacterium hallii [ref_mOTU_v2_4207]	3.65x10 ⁻⁴
Blautia obeum [ref_mOTU_v2_4719]	3.66x10 ⁻⁴
unknown Ruminococcaceae [meta_mOTU_v2_6478]	3.67x10 ⁻⁴
unknown Peptostreptococcaceae [meta_mOTU_v2_7331]	3.91x10 ⁻⁴
unknown Clostridiales [meta_mOTU_v2_6819]	3.96x10 ⁻⁴
unknown Clostridiales [meta_mOTU_v2_5417]	4.07x10 ⁻⁴
unknown Clostridiales [meta_mOTU_v2_5826]	4.11x10 ⁻⁴
unknown Clostridiales [meta_mOTU_v2_6088]	4.78x10 ⁻⁴

unknown Clostridiales [meta_mOTU_v2_6044]	4.88x10 ⁻⁴
unknown Clostridium [meta_mOTU_v2_6792]	4.89x10 ⁻⁴
unknown Clostridiales [meta_mOTU_v2_6926]	5.63x10 ⁻⁴
Blautia obeum [ref_mOTU_v2_4202]	6.02x10 ⁻⁴
Dorea longicatena [ref_mOTU_v2_4203]	6.26x10 ⁻⁴
unknown Clostridiales [meta_mOTU_v2_7157]	6.32x10 ⁻⁴
unknown Clostridiales [meta_mOTU_v2_6575]	6.33x10 ⁻⁴
uncultured Clostridium sp. [meta_mOTU_v2_5907]	6.75x10 ⁻⁴
unknown Clostridiales [meta_mOTU_v2_5805]	7.54x10 ⁻⁴
unknown Clostridiales [meta_mOTU_v2_7158]	8.31x10 ⁻⁴
Dorea formicigenerans [ref_mOTU_v2_0973]	1.13x10 ⁻³
Gemella sanguinis [ref_mOTU_v2_1151]	1.14x10 ⁻³
Solobacterium moorei [ref_mOTU_v2_0531]	1.16x10 ⁻³
unknown Clostridiales [meta_mOTU_v2_7531]	1.17x10 ⁻³
unknown Clostridiales [meta_mOTU_v2_7180]	1.40x10 ⁻³
Oscillibacter sp. 57_20 [meta_mOTU_v2_6676]	1.51x10 ⁻³
unknown Clostridiales [meta_mOTU_v2_6807]	1.57x10 ⁻³
Roseburia hominis [ref_mOTU_v2_4572]	1.70x10 ⁻³
Holdemanella biformis [meta_mOTU_v2_7589]	1.83x10 ⁻³
unknown Clostridiales [meta_mOTU_v2_7630]	1.95x10 ⁻³
Faecalibacterium prausnitzii [ref_mOTU_v2_1379]	1.99x10 ⁻³
Prevotella stercorea [ref_mOTU_v2_1551]	2.09x10 ⁻³
unknown Clostridiales [meta_mOTU_v2_7782]	2.51x10 ⁻³
Clostridium sp. L2-50 [ref_mOTU_v2_4212]	2.75x10 ⁻³
Eubacterium rectale [ref_mOTU_v2_1416]	2.77x10 ⁻³
unknown Clostridiales [meta_mOTU_v2_5396]	2.79x10 ⁻³
unknown Clostridiales [meta_mOTU_v2_6000]	2.81x10 ⁻³
unknown Prevotella [meta_mOTU_v2_5555]	3.17x10 ⁻³
unknown Bacteroidales [meta_mOTU_v2_5375]	3.36x10 ⁻³
unknown Clostridiales [meta_mOTU_v2_6525]	3.40x10 ⁻³
unknown Clostridiales [meta_mOTU_v2_6699]	3.55x10 ⁻³
unknown Peptostreptococcaceae [meta_mOTU_v2_5742]	3.90x10 ⁻³
Lachnoanaerobaculum sp. [ref_mOTU_v2_1537]	3.91x10 ⁻³
unknown Prevotella [meta_mOTU_v2_5502]	3.91x10 ⁻³
unknown Bacteroidales [meta_mOTU_v2_5329]	4.03x10 ⁻³
Coproccoccus comes [ref_mOTU_v2_4313]	4.10x10 ⁻³
unknown Clostridiales [meta_mOTU_v2_6028]	4.49x10 ⁻³
unknown Ruminococcaceae [meta_mOTU_v2_5330]	4.56x10 ⁻³
unknown Ruminococcaceae [meta_mOTU_v2_6557]	4.68x10 ⁻³
unknown Clostridiales [meta_mOTU_v2_6852]	4.94x10 ⁻³

付録 4 大腸がん外科的治療後に増加した腸内細菌

Species	P value
<i>Clostridium clostridioforme</i> [ref_mOTU_v2_0979]	4.88x10 ⁻¹⁰
<i>Flavonifractor plautii</i> [ref_mOTU_v2_1377]	1.11x10 ⁻⁹
<i>Clostridium boltae/clostridioforme</i> [ref_mOTU_v2_0886]	1.31x10 ⁻⁸
<i>Bilophila wadsworthia</i> [ref_mOTU_v2_1149]	2.25x10 ⁻⁷
unknown Clostridiales [meta_mOTU_v2_6672]	9.41x10 ⁻⁷
<i>Parabacteroides merdae</i> [ref_mOTU_v2_1378]	1.64x10 ⁻⁶
<i>Bifidobacterium longum</i> [ref_mOTU_v2_0150]	6.29x10 ⁻⁶
<i>Eggerthella lenta</i> [ref_mOTU_v2_0642]	1.24x10 ⁻⁵
<i>Bacteroides fragilis/ovatus</i> [ref_mOTU_v2_1073]	1.35x10 ⁻⁵
Enterobacteriaceae sp. [ref_mOTU_v2_0036]	2.66x10 ⁻⁵
<i>Bifidobacterium dentium</i> [ref_mOTU_v2_0631]	4.87x10 ⁻⁵
<i>Bifidobacterium breve</i> [ref_mOTU_v2_0157]	5.09x10 ⁻⁵
Clostridiales bacterium VE202-14 [ref_mOTU_v2_2689]	5.56x10 ⁻⁵
<i>Lactobacillus casei/paracasei</i> [ref_mOTU_v2_0226]	6.10x10 ⁻⁵
<i>Anaerostipes caccae</i> [ref_mOTU_v2_1381]	4.51x10 ⁻⁴
<i>Pyramidobacter pisolens</i> [ref_mOTU_v2_4064]	5.49x10 ⁻⁴
<i>Blautia hansenii</i> [ref_mOTU_v2_1428]	1.71x10 ⁻³
<i>Coprococcus</i> sp. [ref_mOTU_v2_0303]	1.75x10 ⁻³
<i>Dialister invisus</i> [ref_mOTU_v2_4598]	2.23x10 ⁻³
<i>Bacteroides xylanisolvens</i> [ref_mOTU_v2_1072]	2.48x10 ⁻³
<i>Clostridium innocuum</i> [ref_mOTU_v2_0643]	4.12x10 ⁻³
<i>Oscillibacter</i> sp. KLE 1728 [ref_mOTU_v2_0858]	4.47x10 ⁻³
<i>Clostridium scindens</i> [ref_mOTU_v2_0883]	4.85x10 ⁻³
<i>Ruminococcus gnavus</i> [ref_mOTU_v2_0280]	4.87x10 ⁻³

付録 5 大腸がん外科的治療後に減少した代謝産物

Metabolite	P value
C02155_Gly-Leu	1.75x10 ⁻¹³
C00214_Thymidine	3.03x10 ⁻¹²
C00785_Urocanate	6.83x10 ⁻⁹
C02567_N1-Acetylspermine	4.32x10 ⁻⁷
C01026_N,N-Dimethylglycine	1.31x10 ⁻⁶
C00065_Ser	2.55x10 ⁻⁶
C00387_Guanosine	4.52x10 ⁻⁶
C02693_Indole-3-acetamide	5.72x10 ⁻⁶
C01005_O-Phosphoserine	7.63x10 ⁻⁶
C00670_Glycerophosphorylcholine	1.11x10 ⁻⁵
C07151_Metformin	1.22x10 ⁻⁴
Cysteine-glutathione disulphide	1.52x10 ⁻⁴
Isovalerate	2.13x10 ⁻⁴
C00371_trans-Zeatin	2.44x10 ⁻⁴
C00385_Xanthine	4.27x10 ⁻⁴
C02704_Methyl sulfate	4.88x10 ⁻⁴
C00197_3PG	5.11x10 ⁻⁴
C00140_N-Acetylglucosamine	5.36x10 ⁻⁴
C01620_Threonate	5.86x10 ⁻⁴
C01081_Thiamine monophosphate	7.32x10 ⁻⁴
C02494_1-Methyladenosine	9.37x10 ⁻⁴
C00314_Pyridoxine	9.94x10 ⁻⁴
C01104_Trimethylamine N-oxide	1.06x10 ⁻³
C00491_Cystine	1.20x10 ⁻³
C00086_Urea	1.20x10 ⁻³
C00327_Citrulline	1.21x10 ⁻³
C00630_Isobutyryl CoA	1.23x10 ⁻³
C10172_Proline betaine	1.24x10 ⁻³
C05629_3-Phenylpropionate	1.30x10 ⁻³
C00860_Histidinol	1.94x10 ⁻³
C01762_Xanthosine	1.95x10 ⁻³
C00074_PEP	1.95x10 ⁻³
C00519_Hypotaurine	1.95x10 ⁻³
C06369_2-Deoxyglucose 6-phosphate	1.95x10 ⁻³
C00956_alpha-Aminoadipate	2.32x10 ⁻³
C02918_1-Methylnicotinamide	2.67x10 ⁻³
C01046_N-Methylglutamate	2.80x10 ⁻³
C00064_Gln	3.03x10 ⁻³
C04259_N-alpha,N-alpha-Dimethylhistidine	3.12x10 ⁻³
C01959-Taurocyamine	3.90x10 ⁻³
C00073_Met	4.06x10 ⁻³
C00378_Thiamine	4.45x10 ⁻³
C02656_Pimelate	4.63x10 ⁻³

付録 6 大腸がん外科的治療後に増加した代謝産物

Metabolite	P value
C00695_Cholate	4.71x10 ⁻⁷
C04483_DCA	2.00x10 ⁻⁵
C00334_GABA	3.13x10 ⁻⁵
C02714_N-Acetylputrescine	3.60x10 ⁻⁵
C05771_Isopropanolamine	8.23x10 ⁻⁵
N1,N8-Diacetylspermidine	1.15x10 ⁻⁴
C01744_3-(4-Hydroxyphenyl)propionate	1.31x10 ⁻⁴
C06337_Terephthalate	2.11x10 ⁻⁴
C05135_4-(beta-Acetylaminoethyl)imidazole	2.54x10 ⁻⁴
C01181_gamma-Butyrobetaine	3.83x10 ⁻⁴
C00483_Tyramine	3.93x10 ⁻⁴
C00134_Putrescine	4.58x10 ⁻⁴
C00099_beta-Ala	4.80x10 ⁻⁴
C01035_4-Guanidinobutyrate	5.96x10 ⁻⁴
C02226_Citraconate	8.31x10 ⁻⁴
C02627_2-Deoxystreptamine	1.47x10 ⁻³
C01015_Hydroxyproline	2.56x10 ⁻³