

腫瘍内不均一性を基にしたがんの予後解析

菊竹, 智恵

<https://hdl.handle.net/2324/2236052>

出版情報 : Kyushu University, 2018, 博士 (理学) , 課程博士
バージョン :
権利関係 :



平成 30 年度 博士論文

腫瘍内不均一性を基にしたがんの予後解析

九州大学大学院 システム生命科学府

菊竹 智恵

目次

1章 緒論.....	4
1.1 研究の背景および目的.....	4
2章 乳がんにおける <i>HMCN1</i> の変異アレル頻度と予後との関連.....	7
2.1 緒言.....	7
2.2 方法.....	11
2.2.1 データセット.....	11
2.2.2 変異解析.....	14
2.2.3 統計解析.....	14
2.3 結果.....	15
2.3.1 乳がんにおける高頻度の遺伝子変異.....	15
2.3.2 <i>HMCN1</i> の発現量比較.....	24
2.3.3 <i>HMCN1</i> の VAF と <i>ITH</i> との関連.....	29
2.3.4 <i>HMCN1</i> の VAF とゲノム不安定性との関連.....	31
2.3.5 <i>HMCN1</i> の VAF とドライバー遺伝子との関連.....	34
2.3.6 <i>HMCN1</i> の VAF と予後との関連.....	36
2.3.7 <i>HMCN1</i> の VAF と臨床情報との関連.....	44
2.4 考察.....	46
3章 多次元尺度を用いた患者予後に関与する腫瘍内不均一性の網羅的解析.....	49
3.1 緒言.....	49
3.2 方法.....	52
3.2.1 データセット.....	52
3.2.2 突然変異分析.....	52
3.2.3 VAF の分布の形状を定義するパラメータの計算.....	53
3.2.4 統計解析.....	53
3.3 結果.....	54
3.3.1 VAF の分布を用いたサンプルのクラスタリング.....	54
3.3.2 5つのクラスターのゲノム上の特性.....	60
3.3.3 5つのクラスターと臨床データとの関連.....	70
3.3.4 5つのクラスターと予後との関係.....	73
3.3.5 決定木による分類.....	76
3.4 考察.....	78

4章 総括および展望.....	82
参考文献.....	84

出版論文リスト

- Kikutake C, Yoshihara M, Sato T, Saito D, Suyama M. Intratumor heterogeneity of HMCN1 mutant alleles associated with poor prognosis in patients with breast cancer. *Oncotarget*. 2018; 9 (70): 33337–33347.
- Kikutake C, Yoshihara M, Sato T, Saito D, Suyama M. Pan-cancer analysis of intratumor heterogeneity associated with patient prognosis using multidimensional measures. *Oncotarget*. 2018; 9 (102): 37689–37699.

1 章 緒論

1.1 研究の背景および目的

がんは細胞の成長や増殖、細胞周期の進行の調節不全を伴う細胞集団のことである。 *TP53* や *KRAS* に代表されるドライバー遺伝子とよばれるがんの発生に直接関与する遺伝子の変異により、正常細胞ががん細胞となり、がん細胞は急速に増大してゆく。その過程でゲノム上に様々な異常が蓄積される。蓄積される変異の組み合わせは細胞によって異なるため、がん細胞は均一な集団ではなく、非常に多様な集団から構成されている[1-3]。すなわち、それぞれのがん細胞が持つ変異の組み合わせは、患者間によっても（これを腫瘍間不均一性； intertumor heterogeneity という）、ある患者の持つがん細胞間によっても（これを腫瘍内不均一性； intratumor heterogeneity という）異なっている[4]。

このような特性はがん治療において大きな障壁となりうる[5]。例えば、ある患者には非常によく効く薬剤が、別の患者では効果がないという場合がある。また、不均一な細胞集団中に薬剤耐性を持つサブクローンが存在する場合、薬剤治療後に生き残った当該サブクローンが増殖してしまうことで、薬剤耐性を持ったがんの再発につながるという事も考えられる（図 1-1）[6,7]。したがって、がんの正確な診断や適切な治療を行うために、腫瘍細胞のもつ変異の種類や変異が生じるタイミング、その結果もたらされる医学的特性を解明することが重要とされてきた。

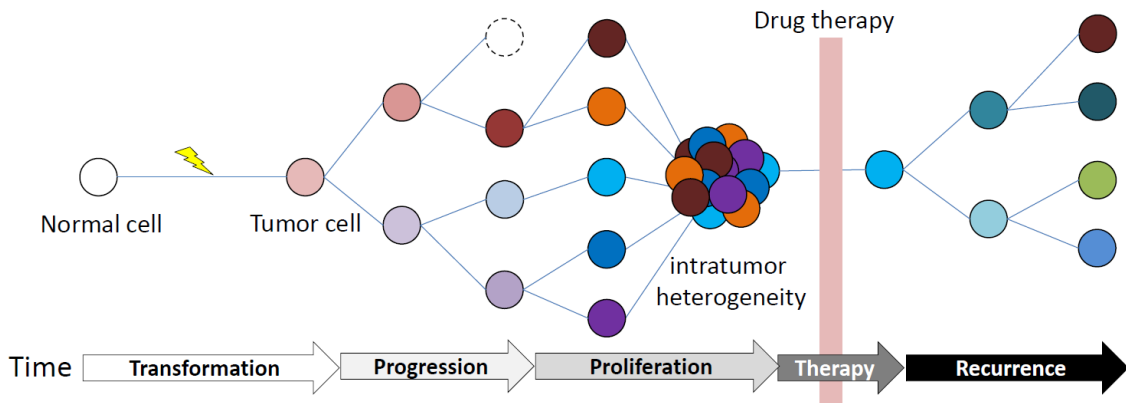


図 1-1 腫瘍内不均一性のモデル

がん細胞は様々な変異を持つ細胞の集合体と考えることができる。がん細胞中に薬剤耐性を持つ細胞集団が存在している場合、薬剤治療後に耐性を持つ細胞が増殖してしまう。

がんの研究においては、次世代シーケンサーの登場により大規模なゲノム解析が行われている。特に、The Cancer Genome Atlas (TCGA) や International Cancer Genome Consortium (ICGC) に代表されるがんゲノムシーケンシングプロジェクトにより、多くの種類のがんのシーケンスデータがデータベースに蓄積されており、誰もがこのデータを使用することができるようになっている[8, 9]。個人間の変異の違いに基づく腫瘍間不均一性を解明するため、このような大規模データを用いた解析がこれまで盛んに行われてきた。その結果、これまで知られていなかった新たなドライバー遺伝子が同定されたり、がんの新しいサブタイプ分類が提案されたりしている。近年では、腫瘍内不均一性に関しても公共のデータベースや各研究所および病院で取得されたシーケンスデータを使った研究が進められている。これらの研究の主たる目的はどのような変異ががんの進化過程のどのタイミングで出現したのかを解析することである。一方で、がんの進化過程ががん患者の予後とどのように関係しているかについては十分に解析されていない。がん患者の予後に影響を与える遺伝子変異は、診断や薬剤による治療ターゲットとして有用であることから、予後解析の重要度は非常に高い。

そこで本研究では、腫瘍内不均一性とがん患者の予後との関係についての知見を得ることを目的とした。1つ目の研究では、がんの進化過程においてどのような変異が生じると予後悪化につながるのかについて、乳がん患者のデータを用いて解析を行った[10]。2つ目の研究では、がんがどのような進化過程をたどると予後悪化につながるのかを明らかにするため、16種類のがんのデータを用いて解析を行った[11]。

2章 乳がんにおける *HMCN1* の変異アレル頻度と予後との関連

2.1 緒言

乳がんは日本人女性の中で最も罹患者数の多いがんである。2012年の時点で全世界でおよそ170万人の女性が乳がんにかかっている[12]。国立がんセンターの調査によると、日本では2012年の時点で約7.4万人の女性が乳がんにかかっており、1975年以降乳がんの患者は増加し続けていることが報告されている(図2-1)。食生活の変化や出生率の低下がこの要因の1つであると考えられている。

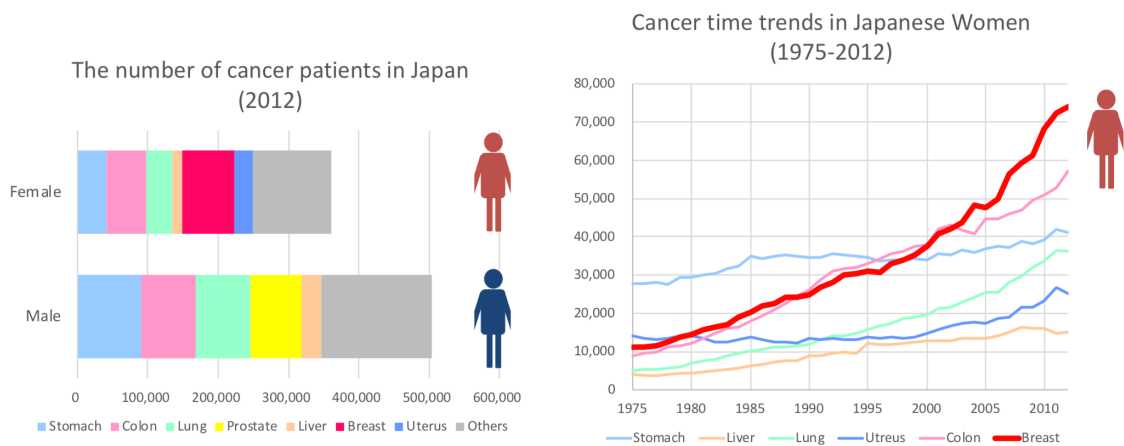


図2-1 国立がん研究センターによる日本のがん疫学調査の結果

左図は2012年における男女別の各がんの罹患者数を、右図は女性における各がんの罹患者数の推移を示す。

乳がんは医学的に不均一な性質を持つことが知られており、3種類のレセプター（エストロゲンレセプター、プロゲステロンレセプター、ヒト上皮成長因子レセプター-2）の発現状態により主として4つのサブタイプ（Luminal A, Luminal B, HER2-enriched, Triple-negative）に分けられる[13,14]（表2-1）。4つのサブタイプは免疫化学染色により判別することができ、それぞれのサブタイプに応じて効果的な治療を受けることができる。

表2-1 乳がんのサブタイプ分類と治療法

サブタイプ分類	ホルモン受容体		HER2	Ki67 値	選択される薬物療法
	エストロゲン	プロゲステロン			
Luminal A	陰性	陽性	陰性	低	ホルモン療法, 化学療法
Luminal B (HER2 陰性)	陽性 or 陰性	弱陽性 or 陰性	陰性	高	ホルモン療法, 化学療法
Luminal B (HER2 陽性)	陽性	陽性 or 陰性	陽性	低～高	ホルモン療法, 化学療法 分子標的治療
HER2-enriched	陰性	陰性	陽性	—	分子標的治療, 化学療法
Triple-negative	陰性	陰性	陰性	—	化学療法

出典：乳がん INFO ナビ

このように乳がんでは他のがんと比べて治療の先進性はあるものの、がんの再発や薬剤耐性は重要な問題として残ったままである。この要因の1つが乳がんの腫瘍内不均一性 (intratumor heterogeneity ; ITH) であり [15], がんの増殖過程により細胞内に蓄積された多様な変異によるものであると考えられている [1, 16, 17]. ITH は、乳がんをはじめ前立腺がん、腎臓がん、脳腫瘍、肝臓がん、肺がんなど様々な種類のがんで既に報告されている [18].

これまでのがん研究では、ドライバー遺伝子や薬剤ターゲットとなりうる遺伝子を含む代謝経路を同定することに重きが置かれてきた。これらの研究は主にドライバー遺伝子異常を標的とする分子標的薬や代謝経路特異的な阻害薬の開発を目的としている。乳がんでも *TP53*, *PIK3CA*, *GATA3* などに代表されるドライバー遺伝子に関する研究が盛んに進められてきたが、他の、特にパッセンジャー遺伝子と考えられてきた遺伝子の中にもその変異体のがん細胞へ影響を及ぼす可能性がある遺伝子が含まれていると考えられている [19, 20]. そこで本研究では、TCGA データを用いて乳がんの ITH を考慮した新たな予後関連遺伝子の探索を行った。この研究で用いたのは各変異における変異アレル頻度 (variant allele frequency, VAF) である。VAF は当該変異を持つがん細胞の割合を表すもので、その大小からがん細胞の進化過程において変異が生じたタイミングを推測できる (図 2-2).

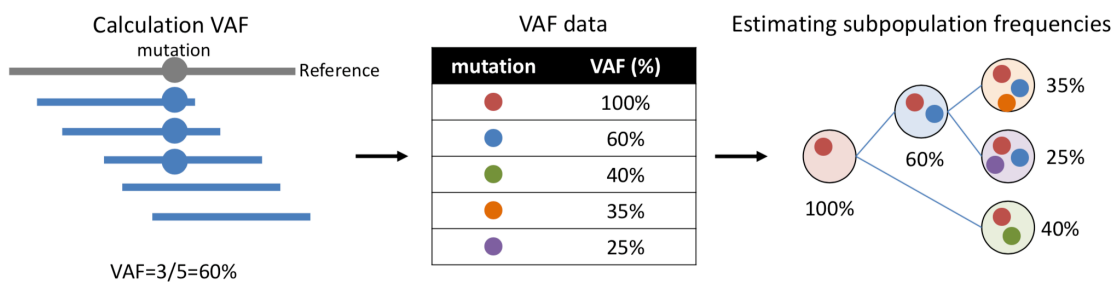


図 2-2 VAF を基にしたがんの進化過程推測のモデル図

VAF はある変異にマッピングされたリード数に対する変異を持つリード数の割合で、がん細胞中の変異の存在割合と考えることができる。VAF が大きい変異はがん細胞中で変異を持つ細胞の割合が大きいことを意味するため、がんの進化過程の比較的早い段階で生じた変異であると推測される。VAF を用いることでがんの進化過程を推測することもできる。中央に示す表から推測される進化過程の一例を右図に示す。

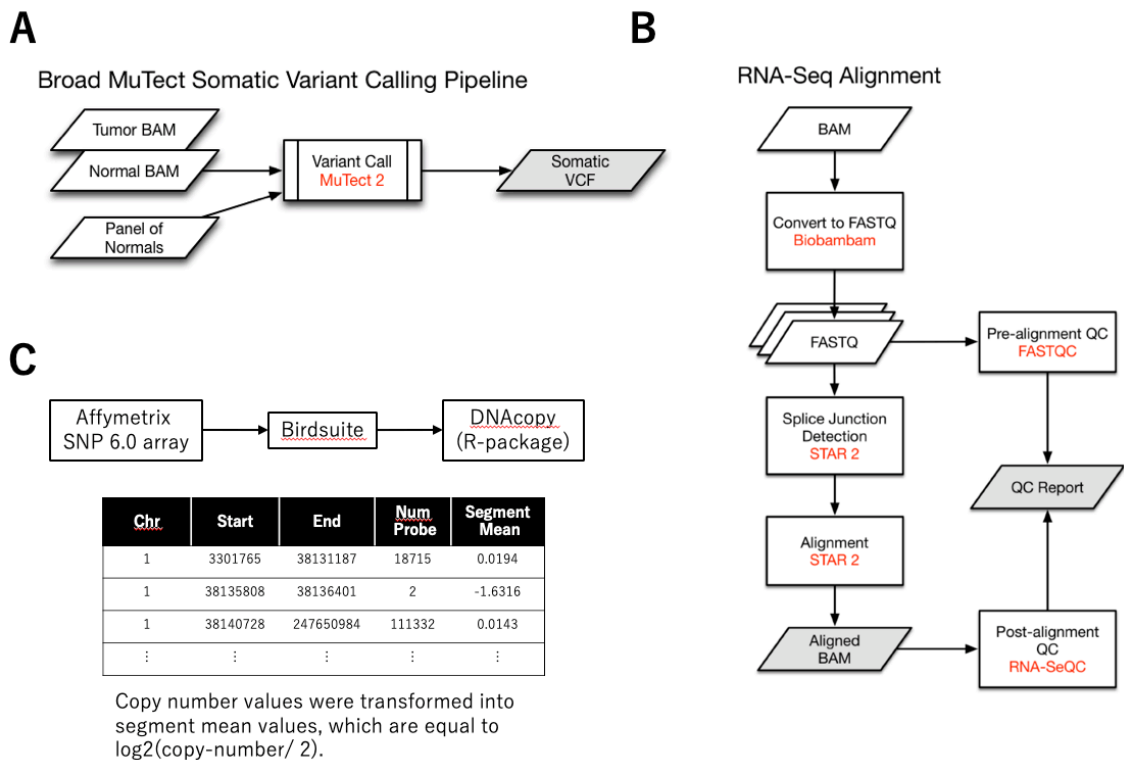
2.2 方法

2.2.1 データセット

本研究では、TCGA (<https://portal.gdc.cancer.gov/>) に蓄積されているデータを用いて解析を行った。TCGA プロジェクトは National Institutes of Health (NIH) が主導する大型がんゲノムプロジェクトであり、TCGA には、NIH 傘下のいくつかの研究所において同一のパイプラインを用いて行われた大量のシーケンスデータが蓄積されている。このうち、乳がん患者由来の 1,044 サンプルの突然変異データ (VCF ファイル) を TCGA レポジトリからダウンロードした。VCF ファイルはバリエーションコーラーの 1 つである MuTect2 によって「がん細胞」と「同一患者から得たコントロール細胞」との比較によって得られたものを用いた。これらのデータの由来となった同一の患者から得た RNA-Seq データ、コピー数多型 (CNV) データ、および臨床データもダウンロードした (表 2-2)。

表 2-2 本研究で用いたデータとフォーマット

データ	フォーマット	パイプライン	リンク
変異	VCF	図 2-3A	https://docs.gdc.cancer.gov/Data/File_Formats/VCF_Format/
RNA-Seq	BAM	図 2-3B	https://docs.gdc.cancer.gov/Data/Bioinformatics_Pipelines/Expression_mRNA_Pipeline/
CNV	TXT	図 2-3C	https://docs.gdc.cancer.gov/Data/Bioinformatics_Pipelines/CNV_Pipeline/



出典：GDC Data User's Guide

図 2-3 TCGA プロジェクトで用いられるデータ解析パイプライン

(A) MutTect2 を用いた変異コールのパイプラインを示したものである。(B) RNA-Seq データのアライメントのパイプラインを示したものである。(C) CNV データ作成のパイプライン、およびアウトプットとして得られるデータフォーマットを示したものである。

同様に、他の 15 種類のがんについても VCF ファイルおよび臨床データをダウンロードした (表 2-3)。

表 2-3 本研究で用いた乳がんを除く 15 種類のがんとそのサンプルサイズ

	がんの種類	サンプルサイズ
BLCA	bladder urothelial carcinoma (膀胱尿路上皮がん)	416
CESC	cervical squamous cell carcinoma and endocervical adenocarcinoma (子宮頸部扁平上皮がんおよび子宮頸部腺がん)	307
COAD	colon adenocarcinoma (結腸腺がん)	605
GBM	glioblastoma multiforme (多形性膠芽腫)	938
HNSC	head and neck squamous cell carcinoma (頭頸部扁平上皮がん)	512
KIRC	kidney renal clear cell carcinoma (腎明細胞がん)	697
LGG	lower grade glioma (低悪性度グリオーマ)	938
LIHC	liver hepatocellular carcinoma (肝細胞がん)	378
LUAD	lung adenocarcinoma (肺腺がん)	587
LUSC	lung squamous cell carcinoma (肺扁平上皮細胞がん)	503
OV	ovarian serous cystadenocarcinoma (卵巣漿液性腺がん)	443
PRAD	prostate adenocarcinoma (前立腺がん)	503
SKCM	skin cutaneous melanoma (皮膚黒色腫)	472
THCA	thyroid carcinoma (甲状腺がん)	504
UCEC	uterine corpus endometrial carcinoma (子宮体部類内膜がん)	604

2.2.2 変異解析

本研究で用いた変異データの平均カバレッジ深度 (average depth of coverage) は 112.1 であり, 十分な精度を確保するためカバレッジ深度が 20 以上である変異 (全体の 92.5%) を解析対象とした. この変異の中から, 乳がんのサンプルの中で 50 サンプル以上で共有されている非同義置換 (ミスセンス変異, ナンセンス変異) および挿入・欠失を持つ遺伝子を抽出した. 50 サンプルという閾値は, 一般的な乳がんのドライバー遺伝子の変異の頻度の下限値が 2~4%程度であり[19], この値を超えて統計解析のために十分なサンプルサイズを確保するために設定した. 変異ごとに, マッピングされたリードに対する変異を持つリードの割合である VAF を計算した. あるサンプルにおいて同一の遺伝子に複数の変異を有する場合には, VAF のより大きいものをその遺伝子の VAF として使用した. 一般に, 腫瘍サンプルの中にはがん細胞以外の細胞 (例えば正常細胞など) が含まれているため, がん細胞が 100%含まれているサンプルから算出された VAF となるように, 各腫瘍サンプルの純度データを用いて VAF の補正を行った. この純度は, 先行研究にてサンプルごとに免疫化学染色法により推定された値を用いた[21]. サンプル内のサブクローンの数は VCF ファイルのデータと DNA の CNV データを用いて R の SciClone パッケージ (version 1.1.0) より推定を行った[22].

2.2.3 統計解析

本研究における統計解析は R (version 3.3.1) (R Project for Statistical Computing, Vienna, Austria) および JMP Pro (version 13.0) (SAS Institute Inc., Cary, NC, USA) を用いて行った. カテゴリカルデータ (診断時年齢, 性別など) の解析には, カイ二乗検定もしくは Fisher の正確検定 (各セルの期待度数が 5 以下となるセルが 1 つ以上ある場合) を用いた. 生存時間解析および Cox 比例ハザードモデルによる解析は R の survival パッケージ (version 2.41-3) を用いた.

2 群間で発現量に差がある遺伝子の検出には Bioconductor の edgeR パッケージ (version 3.16.5) を用いた。2 群間におけるある遺伝子の発現量, MATH 値[23], tumor mutation burden (TMB), CNV abundance および PolyPhen-2 スコアの差を検定する場合には, R の exactRankTests パッケージ (version 0.8-28) を用いてウィルコクソンの順位和検定を行った。がんのステージ, 腫瘍サイズ, リンパ節への転移の有無などの順序変数に対しては, Cochran-Armitage 検定 (片側検定) を行い, VAF と直線的な関係があるかを検証した。診断時年齢の差の比較にはウェルチの t 検定を用いた。対応のある 2 群間の比較にはウィルコクソンの符号順位検定を用いた。また, 3 群以上の比較には ANOVA を用いた。連続変数同士の相関係数を算出する際には, スピアマンの順位相関係数を用いた。P 値は 0.05 以下を有意とみなした (P 値が 0.05 以下の場合には*を, 0.01 以下の場合には**を付した)。

2.3 結果

2.3.1 乳がんにおける高頻度の遺伝子変異

VAF によって予後に差がある遺伝子を探索するため, TCGA から得た乳がん 1,044 サンプルの VCF データを用いて分析を行った。1,044 サンプルのうち 50 サンプル以上で 4 種類のうちいずれかの変異 (ミスセンス変異, ナンセンス変異, フレームシフト変異 (挿入), フレームシフト変異 (欠失)) をもつ遺伝子を抽出した。その結果 17 遺伝子が抽出され (表 2-4), これらの遺伝子の変異に対して VAF を算出した。すべての VAF は, 先行研究より得た各腫瘍サンプルの純度を用いて理論上がん細胞が 100%含まれているサンプルから算出された VAF となるように補正した[21]。各遺伝子の VAF から平均値を求め, 変異を持つサンプルの数とともにプロットした (図 2-4)。また, 各遺伝子の VAF の分布をヒストグラムで表した (図 2-5)。この結果を見ると, *TP53*, *PIK3CA*, *CDH1*, *MAP3K1* といった乳が

んにおいて変異頻度が高いことが知られているドライバー遺伝子の VAF は比較的高かった。

表 2-4 TCGA の乳がんサンプルにおける高頻度な変異遺伝子の VAF

Gene	Sample count	Mean of VAF	SD of VAF
<i>PIK3CA</i>	304	0.471	0.206
<i>TP53</i>	293	0.626	0.241
<i>TTN</i>	193	0.311	0.173
<i>MUC16</i>	113	0.305	0.178
<i>CDH1</i>	104	0.468	0.221
<i>GATA3</i>	102	0.398	0.168
<i>KMT2C</i>	86	0.370	0.223
<i>MAP3K1</i>	77	0.436	0.206
<i>HMCN1</i>	64	0.251	0.166
<i>USH2A</i>	63	0.283	0.182
<i>RYR2</i>	62	0.291	0.202
<i>SYNE1</i>	56	0.272	0.171
<i>FLG</i>	53	0.264	0.201
<i>SPTA1</i>	52	0.264	0.196
<i>DMD</i>	51	0.261	0.169
<i>NEB</i>	50	0.278	0.140
<i>ZFHX4^c</i>	50	0.264	0.182

省略記号：SD, standard deviation (標準偏差)

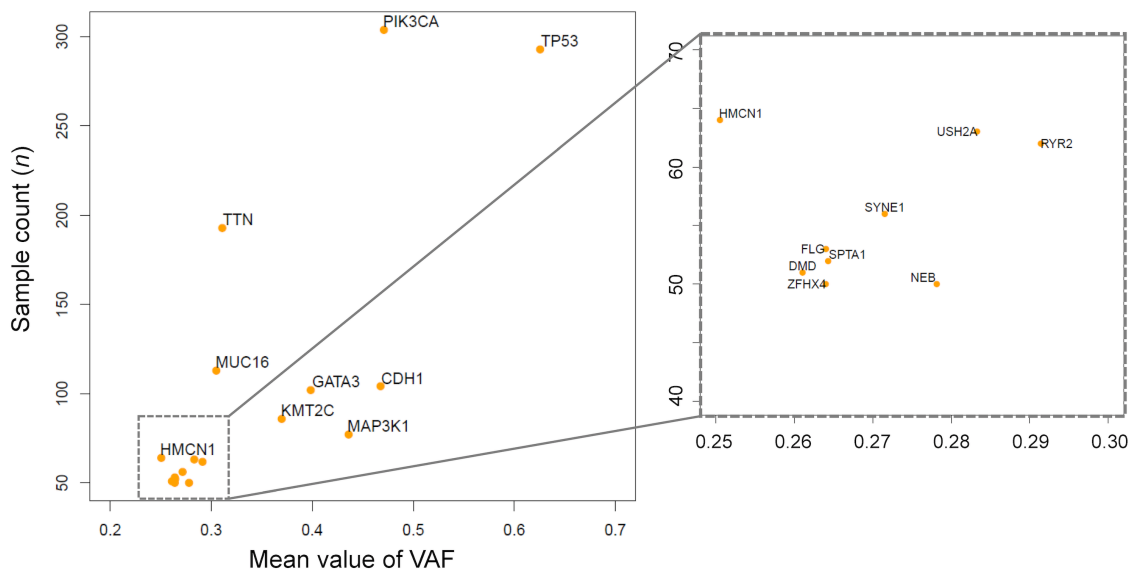


図 2-4 高頻度な遺伝子変異を持つサンプルの数とその平均 VAF

乳がんのサンプルのうち 50 サンプル以上で変異を持つ 17 遺伝子についての散布図である。x軸は変異データから算出した平均 VAF を、y軸は変異を持つサンプルの数を表している。右側の散布図は左側の破線で囲まれた領域を拡大したものである。

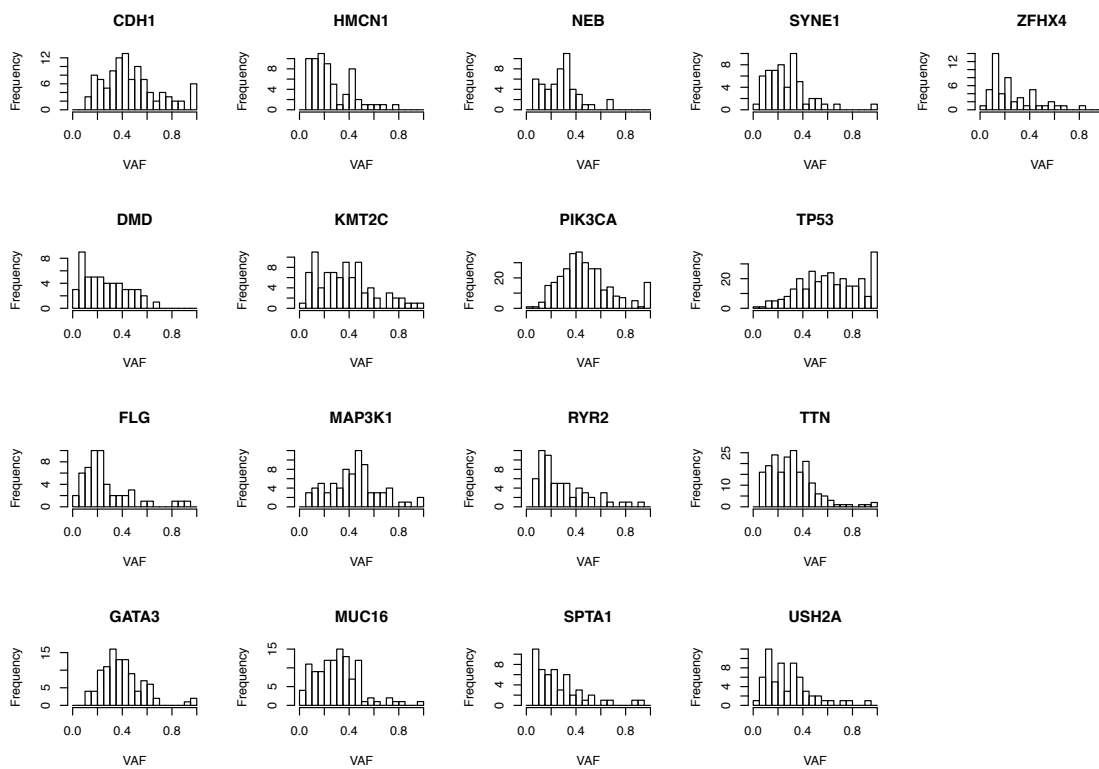


図 2-5 高頻度な遺伝子変異の VAF 分布

乳がんの 1,044 サンプルのうち 50 サンプル以上で変異を持つ 17 遺伝子における VAF の分布を表したものである。x軸は VAF を、y軸はその頻度（サンプルの数）を表している。

これら 17 種類の遺伝子の VAF が生存時間と関連があるか Cox 比例ハザードモデルを用いて検証を行った。共変量として患者の診断時年齢, がんのステージを用いた。この解析では VAF をカットオフ値 0.30 (30%) で 2 群に分けて解析を行った。これは先行研究において平均純度が 85% の腫瘍サンプルを用いた研究において, VAF が 0.25 以上の変異を clonal mutation (腫瘍細胞内に共通して現れる変異), VAF が 0.25 より小さい変異を subclonal mutation (腫瘍細胞の一部に現れる変異) と見なしていたためである [24]。すなわち, 先行研究の腫瘍サンプルの平均純度が 85% であり, 先行研究の VAF=0.25 のカットオフは, 腫瘍サンプルの純度を 100% に補正した本研究では $0.25/0.85=0.30$ に該当すると算出できることから, カットオフ値 0.30 を使用した。算出した *P* 値は Benjamini-Hochberg 法により多重比較のための補正を行った [25]。この解析の結果, *HMCN1* の VAF のみが乳がんの予後と有意に関連が見られた (FDR < 0.1) (表 2-5) ことから, 以降は *HMCN1* に着目して解析を行った。

表 2-5 TCGA の乳がんサンプルにおける高頻度な変異遺伝子の VAF を用いた生存時間解析の結果

Gene	HR ^a (95% CI)	P-value	FDR ^b
<i>PIK3CA</i>	1.780 (0.729-4.348)	0.206	0.696
<i>TP53</i>	1.276 (0.455-3.581)	0.643	0.994
<i>TTN</i>	1.850 (0.843-4.060)	0.125	0.531
<i>MUC16</i>	1.768 (0.619-5.048)	0.287	0.696
<i>CDH1</i>	1.019 (0.263-3.951)	0.979	0.999
<i>GATA3</i>	0.859 (0.245-3.010)	0.813	0.999
<i>KMT2C</i>	1.327 (0.478-3.687)	0.587	0.994
<i>MAP3K1</i>	0.114 (0.013-0.985)	0.048	0.287
<i>HMCN1</i>	11.441 (2.065-63.406)	0.005	0.090*
<i>USH2A</i>	1.185 (0.245-5.740)	0.833	0.999
<i>RYR2</i>	0.059 (0.003-1.008)	0.051	0.287
<i>SYNE1</i>	1.635 (0.110-24.181)	0.721	0.999
<i>FLG</i>	0.342 (0.040-2.923)	0.327	0.696
<i>SPTA1</i>	1.965 (0.531-7.274)	0.312	0.696
<i>DMD</i>	1.870 (0.396-8.831)	0.429	0.811
<i>NEB</i>	1.124 (0.135-9.342)	0.914	0.999
<i>ZFHX4</i> ^c	-	-	-

省略記号：95% CI, 95% confidence interval (95%信頼区間) ; FDR, false discovery rate (偽発見率)

^a HR はハザード比 (hazard ratio) を表す

^b アスタリスクは FDR < 0.1 を示す

^c *ZFHX4* VAF はサンプルサイズが小さかったため Cox 比例ハザードモデルによる解析結果が得られなかった

HMCN1 は Hemicentin-1 をコードしている遺伝子である。このタンパク質は免疫グロブリンスーパーファミリーに属する細胞外タンパク質であり、細胞の接着や遊走、増殖に関与するフィブリンファミリーの 1 つである。本研究で用いた 1,044 サンプルのうち、64 サンプルにおいて合計 78 か所の *HMCN1* の変異が見られた ($64/1,044=6.1\%$) (表 2-6 および図 2-6)。変異の内訳を見ると、6 サンプル ($6/64=9.4\%$) が 1 サンプル中に 2 か所の変異を、2 サンプル ($2/64=3.1\%$) が 3 ヶ所以上の変異を持っていた。78 か所の変異のうち 64 か所 ($64/78=82.1\%$) がミスセンス変異、8 か所 ($8/78=10.3\%$) がナンセンス変異、6 か所 ($6/78=7.7\%$) が挿入・欠失であった。さらに、これらの変異のうち 54 か所 ($54/78=69.2\%$) が免疫グロブリン様 C2 型ドメインに見られた。

表 2-6 に示した *HMCN1* の 4 種類の変異と VAF との間に関連があるかどうかを評価するため、変異の種類ごとに VAF を集計した。この結果をもとに一元配置分散分析を用いて検定を行ったが、有意な差は見られなかった ($P=0.430$) (図 2-7A)。同様に、乳がんの 4 つのサブタイプ (Luminal A, Luminal B, HER2-enriched, Basal-like (トリプルネガティブ型に含まれる)) と VAF との関連を解析したが、サブタイプによる VAF の値に有意な差は見られなかった ($P=0.379$) (図 2-7B)。この結果より、*HMCN1* の変異における VAF は乳がんのサブタイプとは独立であり、また乳がんのサブタイプの形成に影響を及ぼす可能性は低いと示唆される。

表 2-6 *HMCN1* における変異の分布

<i>HMCN1</i> Domains	Missense mutation	Nonsense mutation	Deletion	Insertion	Total
VWFA domain	2	0	0	0	2
Ig-like C2-type domains	45	5	2	2	54
TSP type-1 domains	8	1	0	0	9
Nidogen G2 beta-barrel domain	1	1	0	0	2
EGF-like domains	3	0	0	0	3
Other	5	1	1	1	8
Total	64	8	3	3	78

省略記号：EGF, epidermal growth factor（上皮細胞成長因子）；Ig, immunoglobulin（免疫グロブリン）；TSP, thrombospondins（トロンボスポンジン）；VWFA, von Willebrand factor type A（フォン・ヴィレブランド因子 A 型）。

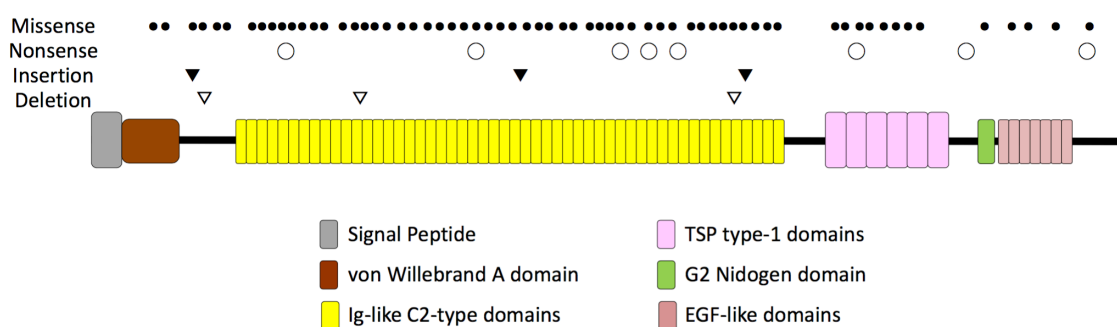


図 2-6 ヒト *HMCN1* における各ドメインの概要図

HMCN1 遺伝子概要図の上部に 78 か所の変異位置を示している。変異位置の記号は、左上に示す 4 種類の変異を表している。

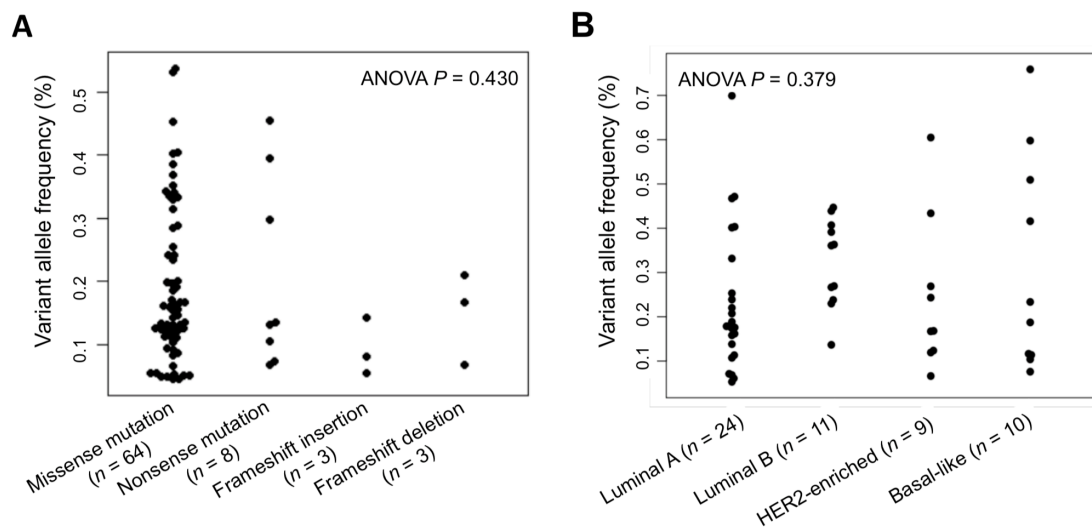


図 2-7 *HMCN1* における 4 種類の変異およびサブタイプごとの VAF

(A) 64 か所のミスセンス変異, 8 か所のナンセンス変異, 3 か所の挿入, 3 か所の欠失におけるそれぞれの VAF を示している. (B) 64 サンプルのうち, 24 サンプルの Luminal A, 11 サンプルの Luminal B, 9 サンプルの HER2-enriched, 10 サンプルの Basal-like (トリプルネガティブ) におけるそれぞれの VAF を示している. 残りの 10 サンプルについては乳がんのサブタイプが不明だったため, この図には含めていない.

2.3.2 *HMCN1* の発現量比較

VAF の大小によって転写産物の発現量に差があるかを確認するため、*HMCN1* に変異を持つサンプルを 2.3.1 に記載した VAF のカットオフ値 0.30 を用いて VAF 高値群 (VAF \geq 0.30, $n=19$) と低値群 (VAF $<$ 0.30, $n=45$) の 2 群に分け、ウィルコクソンの順位和検定により mRNA の発現量の比較を行った。その結果、2 群間で *HMCN1* の発現量に有意な差は見られなかった ($P=0.343$) (図 2-8A)。同様に、VAF のカットオフ値を 0.30 とした場合の 2 群間で *TP53* および *PIK3CA* の発現量を比較したが、両遺伝子とも有意な差は見られなかった (*TP53*; $P=0.515$, *PIK3CA*; $P=0.300$) (図 2-8B, C)。さらに、*HMCN1* の変異を持つサンプルと変異を持たないサンプル間においても mRNA の発現量の差は見られなかった ($P=0.984$) (図 2-8D)。*HMCN1*, *TP53*, *PIK3CA* の VAF とその発現量との間におけるスピアマンの順位相関係数は -0.115 ($P=0.364$), 0.016 ($P=0.787$), 0.028 ($P=0.631$) であり、いずれも有意な相関を示さなかった。したがって、*HMCN1* の発現量ではなく変異の allele dosage が乳がんの予後に影響を及ぼしている可能性がある。

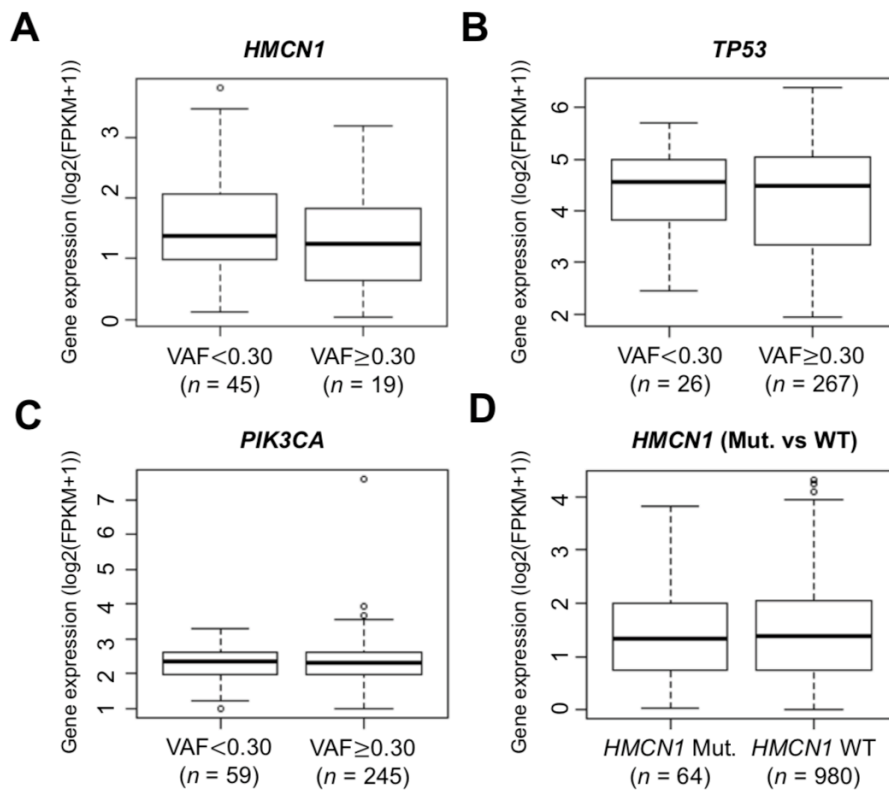


図 2-8 VAF を基にした 2 群間での発現量の比較

(A - C) *HMCN1*, *TP53*, *PIK3CA* の変異を持つサンプルにおいて、VAF のカットオフを 0.30 にして分けた 2 群間での mRNA の発現量の比較結果である。(D) *HMCN1* の変異を持つ群 (*HMCN1* Mut.) と持たない群 (WT) の 2 群間での mRNA の発現量の比較結果である。

HMCN1 の allele dosage の影響を受けている可能性のある遺伝子を同定するため、RNA-Seq データを用いて mRNA の発現量の解析を行った。*HMCN1* の VAF 高値群と低値群との間で、全遺伝子を対象に mRNA の発現量に差がある遺伝子を探索した。その結果 4 つの遺伝子で 2 群間において有意な差が見られた (図 2-9)。 *CA9* と *CASP14* は、*HMCN1* の VAF 低値群に比べて高値群において発現量が有意に高かった ($P=0.043$, $P=0.024$) (図 2-9A, B)。一方 *MTRNR2L1* と *TCN1* は、*HMCN1* の VAF 高値群に比べて低値群において発現量が有意に高かった ($P=0.024$, $P=0.043$) (図 2-9C, D)。*HMCN1* の VAF と *CA9*, *CASP14*, *MTRNR2L1*, *TCN1* の発現量との間におけるスピアマンの順位相関係数は 0.307 ($P=0.013$), 0.241 ($P=0.055$), -0.181 ($P=0.154$), -0.306 ($P=0.014$) であり、*CA9* と *TCN1* でのみ有意な相関が見られた (図 2-10)。

CA9 は炭酸脱水酵素 9 (Carbonic Anhydrase 9) をコードする遺伝子で、乳がんにおいて細胞内の低酸素状態を見るための内在性マーカーとして用いられる [26, 27]。また、*CASP14* はカスパーゼ 14 (Caspase 14) をコードしている遺伝子であり、アポトーシス関連システインペプチダーゼの一種である。近年の研究では、*CASP14* の発現量は乳がんの悪性度と関連していることが示されている [28]。*MTRNR2L1* はヒト MT-RNR2 様 1 タンパク (MT-RNR2 like 1) をコードしていることが知られているが詳細な機能は不明である。*TCN1* (Transcobalamin 1) はビタミン B12 結合タンパクファミリーの 1 つであるトランスコバラミン 1 をコードしている。

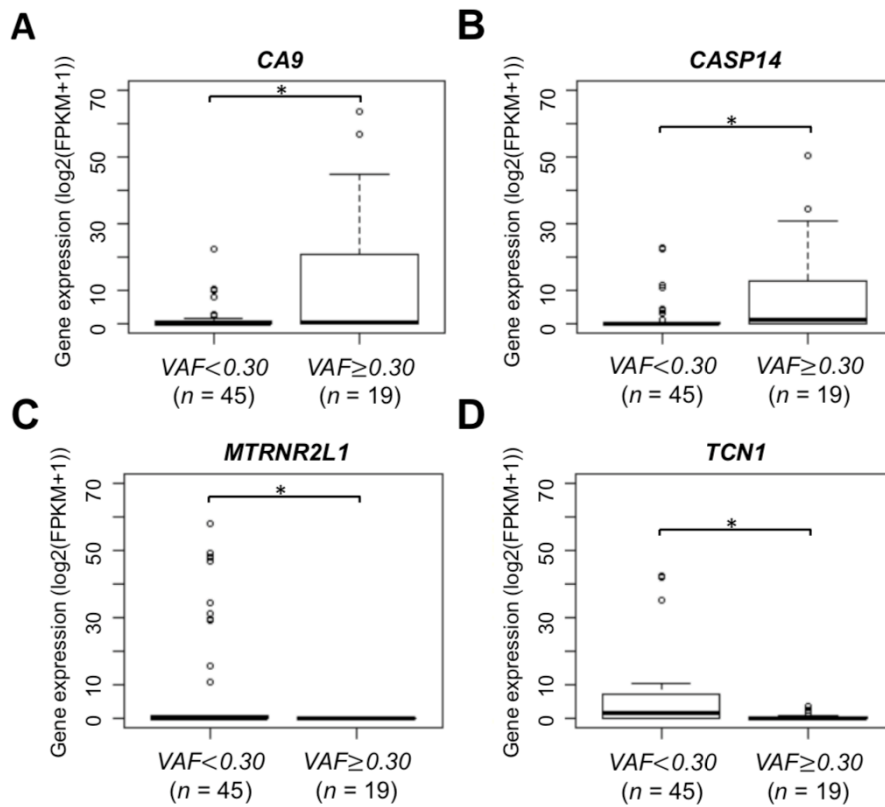


図 2-9 *HMCN1* の VAF に基づく 2 群間で mRNA の発現量に差があった 4 種類の遺伝子 *HMCN1* の変異を持つサンプルにおいて、VAF のカットオフを 0.30 にして分けた 2 群間での (A) *CA9*, (B) *CASP14*, (C) *MTRNR2L1*, (D) *TCN1* の mRNA の発現量の比較結果である。

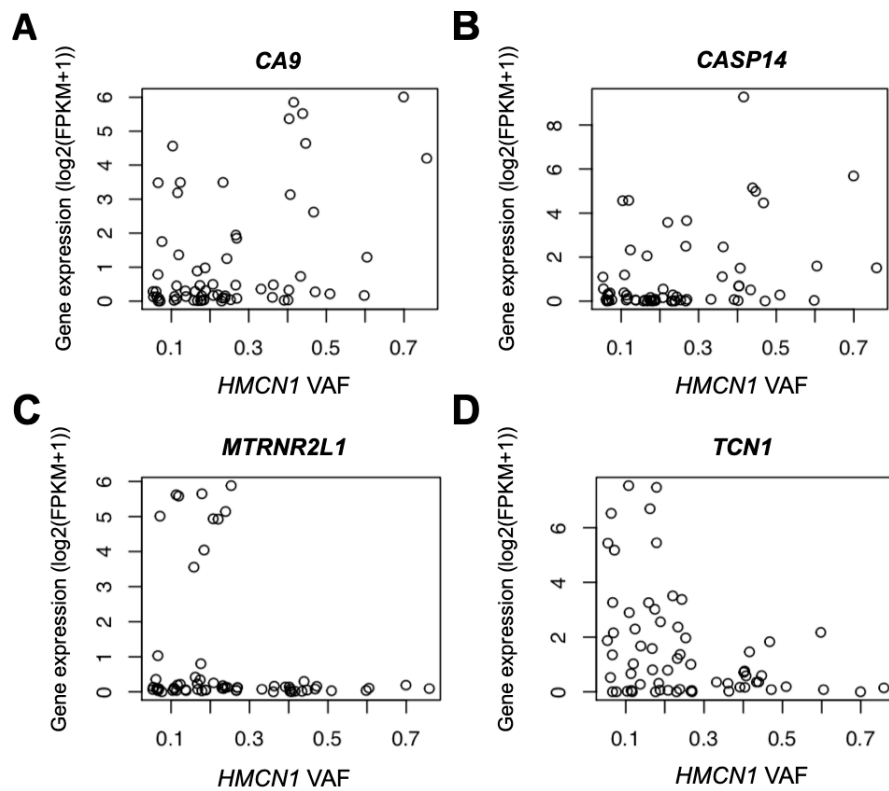


図 2-10 *HMCN1* の VAF と *CA9*, *CASP14*, *MTRNR2L1*, *TCN1* の発現量の相関
HMCN1 の変異を持つサンプルにおける VAF と (A) *CA9*, (B) *CASP14*, (C) *MTRNR2L1*,
 (D) *TCN1* の mRNA の発現量の相関を示したものである。

2.3.3 *HMCN1* の VAF と ITH との関連

HMCN1 の変異を持つサンプルにおける ITH の程度を調べるため、すでに報告されている ITH の指標であるがん細胞中のサブクローン数を用いた解析を行った。がん細胞中のサブクローン数は ITH を定量化するための値と考えられており、サブクローン数の推定を行うためのいくつかのソフトウェアが開発されている [29]。これらのアルゴリズムは変異の VAF と copy number variants (CNVs) から統計的な推測によってがん細胞内のサブクローン数を算出する。本研究ではそのうちの 1 つである SciClone というソフトウェアを使い、各サンプルの持つサブクローン数を推定した [22] (図 2-11A)。フィッシャーの正確検定によって *HMCN1* の VAF とサブクローン数の分布との関連を検証したが、有意な関係は見られなかった ($P = 0.347$)。

また、ITH の別の指標として変異アレル腫瘍不均一性指標 (mutant-allele tumor heterogeneity, MATH) が報告されている。MATH 値はあるサンプルの持つ変異から算出された VAF のばらつきを表すものであり、 x_i をあるサンプルにおける各変異の VAF とする ($i = 1, 2, \dots, n$) とき、

$$MATH = \frac{MAD}{\text{median}(x)} = \frac{\text{median}(|x_i - \text{median}(x)|)}{\text{median}(x)}$$

という式によって算出できる [23]。ただし MAD は中央絶対偏差 (median absolute deviation) を表す。がん細胞中のサブクローン数と同様に、MATH 値も ITH の程度を推測するための指標であり、先行研究では頭頸部扁平上皮がんおよび結腸がんにおいて MATH 高値群は MATH 低値群に比べて予後不良であることが示されている [23, 30]。*HMCN1* の VAF と MATH 値との間に関連があるかウィルコクソンの符号順位検定を行ったところ、VAF 高値群 ($n = 19$) では MATH の平均値は 35.014 ($SD = 11.300$)、低値群 ($n = 45$) では平均値が 33.919 ($SD = 10.557$) であり 2 群間に統計的な有意差は見られなかった ($P = 0.771$) (図 2-11B)。これらの結果は *HMCN1* の VAF の大小はがん細胞の ITH の状態とは関連がないことを示唆している。

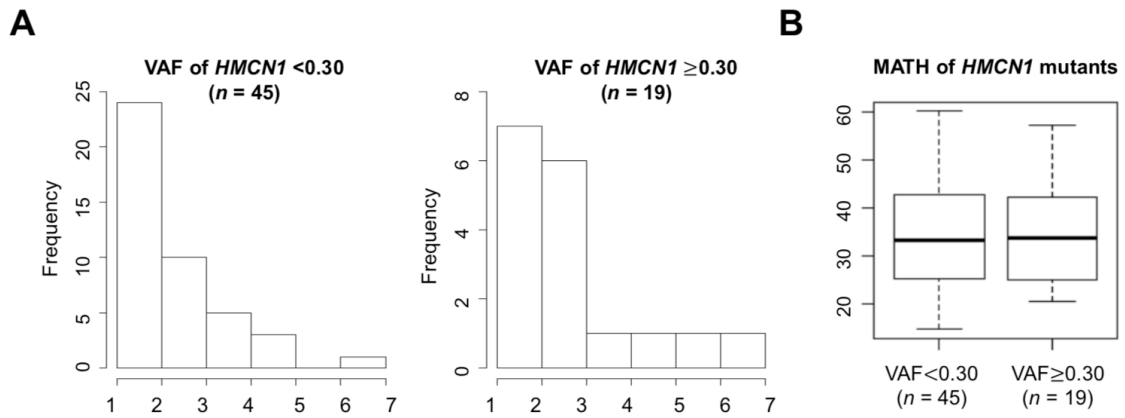


図 2-11 *HMCN1* の VAF に対するサブクローン数と MATH 値

(A) *HMCN1* の VAF 低値群と高値群におけるサブクローンの数を表したものである。x軸は SciClone によって推定されたがん細胞中のサブクローンの数を表している。(B) *HMCN1* の VAF 低値群と高値群における MATH 値の分布を表したものである。y軸は MATH 値を 100 倍した値として表している。

2.3.4 *HMCN1* の VAF とゲノム不安定性との関連

ゲノム不安定性もまた ITH に影響をもたらす要因の 1 つである。ゲノム不安定性には突然変異と CNV が含まれる。最近の研究により、これら双方が共に多い場合に高い ITH を示すことが分かっている[31]。そこで、2.3.3 の結果を検証するため、*HMCN1* 高値群と低値群とで突然変異の数の比較を行った。この比較に用いたのは TMB で、あるサンプルの持つ非同義置換の総数である。統計解析の結果 $P = 0.253$ となり、*TP53* ($P = 0.510$) および *PIK3CA* ($P = 0.469$) と同様に統計的な有意差が検出されなかった (図 2-12)。*HMCN1*, *TP53*, *PIK3CA* の VAF と TMB との間におけるスピアマンの順位相関係数は -0.073 ($P = 0.569$), 0.304 ($P < 0.001$), -0.108 ($P = 0.002$) であり、*TP53*, *PIK3CA* の VAF においてのみ有意な相関を示した。

次に、*HMCN1* 高値群と低値群とでコピー数変異量 (CNV abundance) の比較を行った。CNV との関連解析では、CNV abundance という全ゲノム領域に対するコピー数異常のある領域の割合を用いた。*HMCN1* の変異を持つサンプルにおいて CNV abundance を算出し統計解析を行った結果、VAF 高値群と低値群の 2 群間に有意な差が見られた ($P = 0.004$) (図 2-13A)。同様の比較を *TP53*, *PIK3CA* に対しても行ったところ *HMCN1* と同様に有意な結果が得られた (*TP53*; $P < 0.001$, *PIK3CA*; $P < 0.001$) (図 2-13B, C)。また、*HMCN1*, *TP53*, *PIK3CA* の VAF と CNV abundance との間におけるスピアマンの順位相関係数は 0.018 ($P = 0.752$), 0.539 ($P < 0.001$), 0.328 ($P < 0.001$) であり、TMB と同様に *TP53*, *PIK3CA* の VAF においてのみ有意な相関を示した。*HMCN1* の VAF が高いサンプルでは染色体不安定性が有意に高かったが、変異の数には差が見られなかったことから、先行研究で示された「突然変異頻度と CNV abundance が共に高い」という高 ITH の条件を満たすとは言えない。したがって、本結果は *HMCN1* の allele dosage は ITH の状態とは関連がないという結果を裏付けるものであると考えられる。

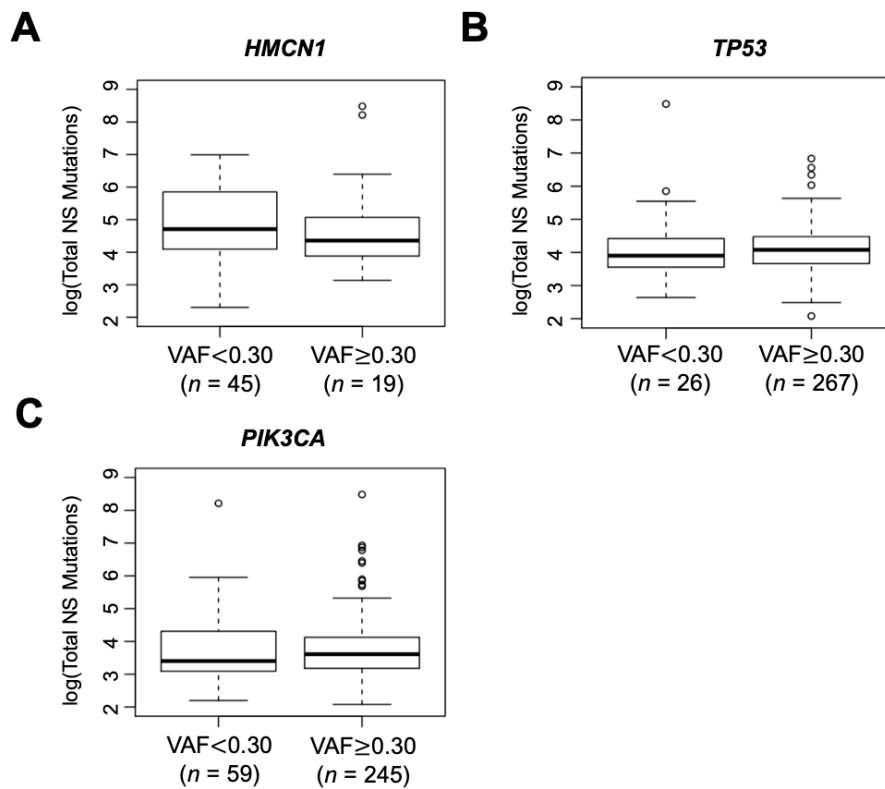


図 2-12 *HMCN1* の VAF と TMB との関連

(A - C) *HMCN1*, *TP53*, *PIK3CA* の変異を持つサンプルにおいて、VAF のカットオフを 0.30 にして分けた 2 群間での非同義置換数の比較結果である。

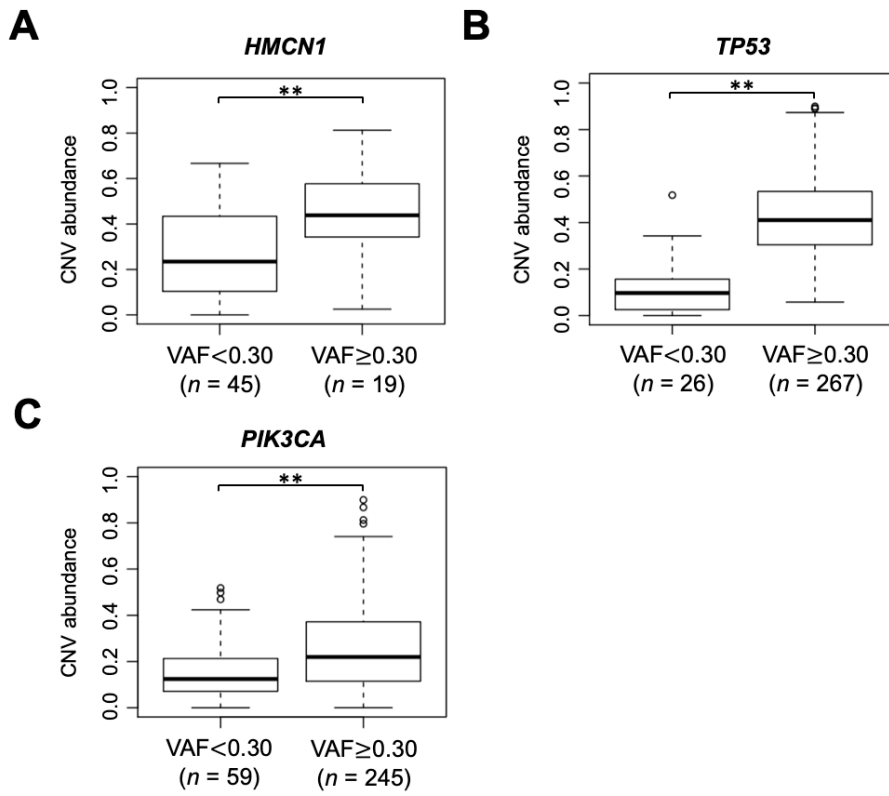


図 2-13 *HMCN1* の VAF と染色体不安定性との関連

(A - C) *HMCN1*, *TP53*, *PIK3CA* の変異を持つサンプルにおいて、VAF のカットオフを 0.30 にして分けた 2 群間での CNV abundance の比較結果である。アスタリスクは統計的有意性を示す。

2.3.5 *HMCN1* の VAF とドライバー遺伝子との関連

乳がんにおいて *TP53* や *PIK3CA* は最も一般的なドライバー遺伝子として知られている [19]。 *HMCN1* の変異を持つサンプルにおいて、 *HMCN1* の VAF とこれらのドライバー遺伝子の変異における VAF との間に関係があるかを検証するために、まず *HMCN1* 変異を持つサンプルのうち上記 2 遺伝子の変異の有無を調べた。 その結果、 64 サンプルのうち、 22 サンプル (34%) と 23 サンプル (36%) でそれぞれ *TP53* と *PIK3CA* の変異を持っており、うち 5 サンプルは両方の遺伝子の変異を持っていた。 これらの頻度は TCGA の乳がんデータにおける *TP53*, *PIK3CA* の変異頻度 (37%, 36%) と同程度であった [19]。

次に、 *TP53*, *PIK3CA* の変異を持つサンプルについて VAF の値を調べた。 *TP53* の変異を持つ 22 サンプルでは、 *TP53* の VAF の平均値は 0.697 (SD = 0.249)、 *HMCN1* の VAF の平均値は 0.288 (SD = 0.201) であり、同一サンプルにおける VAF の差の平均値は 0.409 (SD=0.288) だった。 一方、 *PIK3CA* の変異を持つ 23 サンプルでは、 *PIK3CA* の VAF の平均値は 0.442 (SD = 0.269)、 *HMCN1* の VAF の平均値は 0.230 (SD = 0.148) であり、同一サンプルにおける VAF の差の平均値は 0.212 (SD=0.273) だった。 ウィルコクソンの符号順位検定の結果、 2 種類のドライバー遺伝子の VAF の値は *HMCN1* の VAF よりも有意に高かったことから (*TP53*; $P < 0.01$, *PIK3CA*; $P < 0.01$) (図 2-14), *HMCN1* の変異ががんの発生時ではなく、 *TP53* や *PIK3CA* をはじめとするドライバー遺伝子の変異によるがん発生後、進化過程の途中で出現した可能性を示唆している。

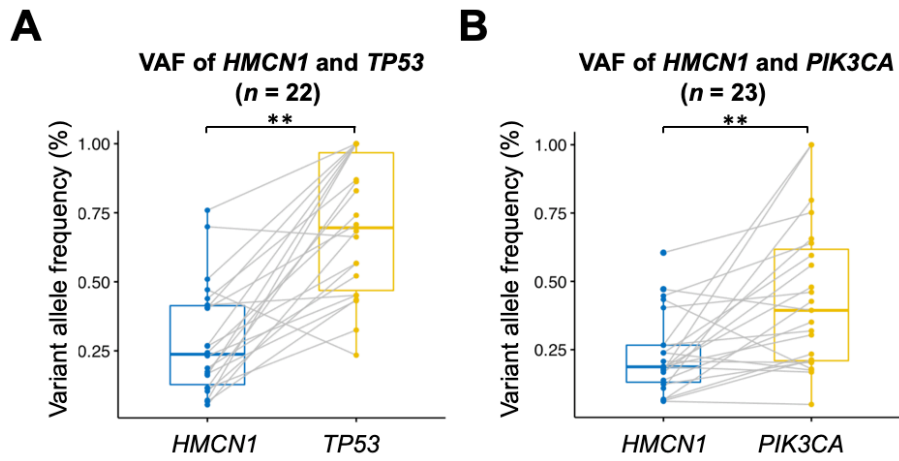


図 2-14 *HMCN1n* の VAF と *TP53* および *PIK3CA* の比較

(A) *HMCN1* と *TP53* の VAF ($n = 22$) と (B) *HMCN1* と *PIK3CA* ($n = 23$) の VAF との比較結果である。アスタリスクは統計的有意性を示す。

2.3.6 HMCN1 の VAF と予後との関連

HMCN1 の VAF と予後との関連を評価するため、HMCN1 変異を持つ 64 サンプルをカットオフ値 0.30 を用いて 2 群に分け、生存時間分析を行った。さらにこれらの 2 群と HMCN1 変異を持たないサンプル群 (WT) との比較も行った。Log-rank 検定の結果、HMCN1 の VAF 高値群は低値群および WT に比べて有意に予後不良であった (対 WT; $P = 0.022$, 対 VAF < 0.30; $P = 0.015$) (図 2-15)。

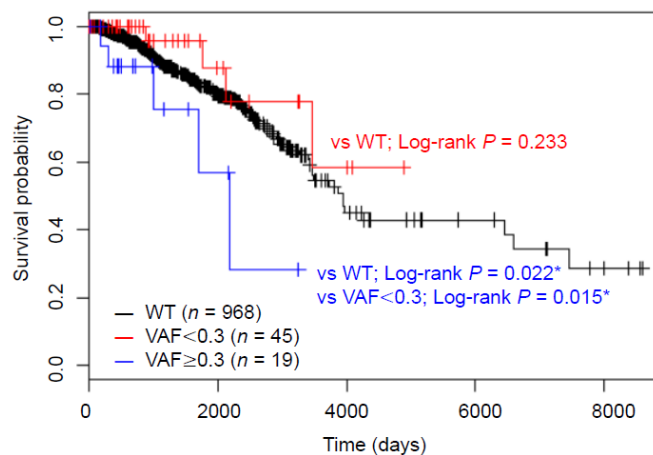


図 2-15 HMCN1 の VAF に基づく Kaplan-Meier 曲線

HMCN1 の変異を持つサンプルを、VAF のカットオフ 0.30 にして 2 つのグループに分けた (< 0.30; 赤色, $n = 45$ および ≥ 0.30 ; 青色, $n = 19$)。また、比較として HMCN1 変異を持たない WT (黒色, $n = 968$) も用いた。Log-rank 検定により、2 つの生存曲線間の差の統計的有意性を評価した結果を示している (VAF ≥ 0.30 対 VAF < 0.30, VAF ≤ 0.30 対 WT および VAF ≥ 0.30 対 WT)。アスタリスクは統計的有意性を示す。

次に共変量（リンパ節転移の有無，ステージ，腫瘍の大きさ，診断時年齢）（表 2-7）の影響を考慮した解析を行うため，Cox 比例ハザードモデルを用いた解析を行った．この解析に先立って比例ハザード性の検定を行ったところ，すべての変数において棄却されなかった（カイ二乗検定; $P \geq 0.05$ ）ため，比例ハザード性を満たしていると判断した．これらの変数を共変量とし，Cox 比例ハザード回帰分析を行った結果，VAF 高値群で死亡リスクが有意に高かった（表 2-8）．

表 2-7 本研究で用いた乳がん 1,044 サンプルの臨床情報

Variables	Overall	WT	MT	<i>P</i> -value ^a
	<i>n</i> = 1,044	<i>n</i> = 980	<i>n</i> = 64	
	No. (%)	No. (%)	No. (%)	
Lymph node status				0.175
Negative	485 (46.5)	450 (45.9)	35 (54.7)	
Positive	540 (51.7)	513 (52.3)	27 (42.2)	
Unknown	19 (1.8)	17 (1.7)	2 (3.1)	
Tumor stage				0.080
1	172 (16.5)	163 (16.6)	9 (14.1)	
2	582 (55.7)	536 (54.7)	46 (71.9)	
3	239 (22.9)	231 (23.6)	8 (12.5)	
4	20 (1.9)	19 (1.9)	1 (1.6)	
Unknown	31 (3.0)	31 (3.2)	0	
Tumor size (cm)				0.028 *
< 2	267 (25.6)	255 (26.0)	12 (18.8)	
2-5	603 (57.8)	556 (56.7)	47 (73.4)	
≥ 5	171 (16.4)	166 (16.9)	5 (7.8)	
Unknown	3 (0.3)	3 (0.3)	0	
Molecular subtype				0.055
Luminal A	479 (45.9)	455 (46.4)	24 (37.5)	
Luminal B	192 (18.4)	181 (18.5)	11 (17.2)	
HER2-enriched	78 (7.5)	69 (7.0)	9 (14.1)	
Basal-like	165 (15.8)	155 (15.8)	10 (15.6)	
Normal	33 (3.2)	28 (2.9)	5 (7.8)	
Unknown	97 (9.3)	92 (9.4)	5 (7.8)	
Age (year)				0.450
Median (range)	59 (27-90)	59 (27-90)	59 (34-90)	
< 50	276 (26.4)	261 (26.6)	15 (23.4)	
≥ 50	743 (71.2)	649 (66.2)	49 (76.6)	
Unknown	25 (2.4)	25 (2.6)	0 (0)	

省略記号：MT, *HMCN1* mutant (*HMCN1* の変異を持つサンプル) ; WT, wild-type (*HMCN1* の変異を持たないサンプル)

^a アスタリスクは統計的有意性を示す。

表 2-8 64 人の乳がん患者の臨床情報を用いた Cox 比例ハザード回帰分析の結果

Variables	HR (95% CI)	P-value ^a
Lymph node status		
Positive vs Negative	97.931 (2.709-3539.805)	0.012 *
Tumor grade		
3-4 vs 1-2	9.468 (0.042-2147.487)	0.417
Tumor size, cm		
2-5 vs ≤ 2	1.281 (0.159-10.302)	0.816
> 5 vs ≤ 2	29.032 (0.284-2966.692)	0.154
Age		
≥ 50 vs < 50	0.114 (0.011-1.169)	0.068
VAF		
≥ 0.30 vs < 0.30	17.950 (1.216-264.976)	0.036 *

省略記号：HR, hazard ratio (ハザード比)；95% CI, 95% confidence interval (95%信頼区間)

^a アスタリスクは統計的有意性を示す。

HMCN1 の VAF の予後への有意な関連が 2 つの群における突然変異のタンパク質の立体構造への影響の偏りに起因するかどうかを検証するために、PolyPhen-2 スコアを用いて *HMCN1* の点突然変異のタンパク質構造および機能への影響を評価した [32]. PolyPhen-2 スコアは 0~1 の範囲をとり、「probably damaging (> 0.85)」、「possibly damaging (0.85-0.15)」、または「benign (< 0.15)」のいずれかの予測をもたらす。*HMCN1* の VAF と PolyPhen-2 スコアとの間のスピアマンの順位相関係数は -0.211 であり、有意な相関を示さなかった ($P=0.138$). また、*HMCN1* の VAF のカットオフを 0.30 にして分けた 2 群間においても PolyPhen-2 スコアの有意差は見られなかった ($P=0.052$). さらに予後と PolyPhen-2 スコアとの関係を分析するために、*HMCN1* 変異を持つ 64 サンプルを 0.85 の PolyPhen-2 スコアをカットオフ値とし、2 つのグループに分類した (PolyPhen-2 スコア高値; $n=27$ および低値群; $n=24$). この解析の結果、*HMCN1* 変異の Polyphen-2 スコアは乳がんの予後と有意な関連が見られなかった (PolyPhen-2 スコア < 0.85 対 WT; $P=0.801$ および PolyPhen-2 ≥ 0.85 対 WT; $P=0.671$) (図 2-16). これらの結果は、*HMCN1* の VAF が独立した予後因子である可能性を示唆している.

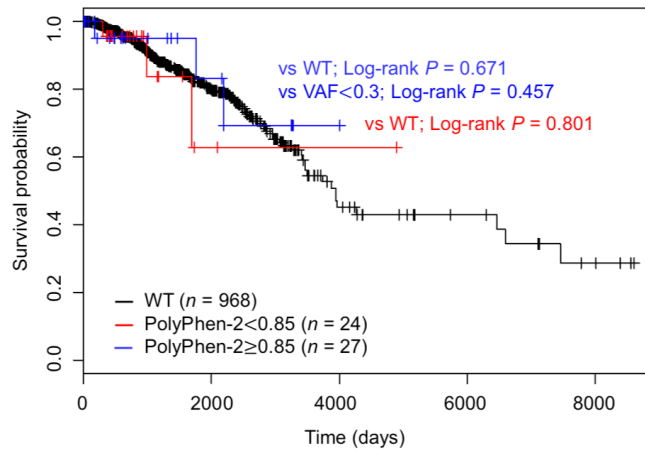


図 2-16 *HMCN1* における変異の PolyPhen-2 スコアに基づく Kaplan-Meier 曲線

HMCN1 の変異を持つサンプルを、PolyPhen-2 スコアのカットオフを 0.85 にして 2 つのグループに分けた (< 0.85, 赤色, $n = 24$ および ≥ 0.85 , 青色, $n = 27$)。また、比較として *HMCN1* 変異を持たない WT (黒色, $n = 968$) も用いた。Log-rank 検定を用いて、2 つの生存曲線間の差の統計的有意性を評価した。

独立した別データによって、乳がんサンプルにおける *HMCN1* の VAF の予後への影響を確認するため、TCGA の他の 15 種類のがん由来のデータセットを用いて *HMCN1* の VAF と予後との関係性を評価した。15 種類のがんの中で、子宮頸部扁平上皮がんおよび子宮頸部腺がん (CESC) サンプルのみが、VAF 低値群 ($n = 15$, $VAF < 0.30$) に比べて高値群 ($n = 7$, $VAF \geq 0.30$) において予後不良であった (Log-rank 検定: $P = 0.048$) (図 2-17)。さらになんかのステージ、診断時年齢を共変量とし、Cox 比例ハザード回帰分析を行ったが、子宮頸部扁平上皮がんおよび子宮頸部腺がんの *HMCN1* の VAF の予後との関連は見いだせなかった (HR = 5.436, 95% CI: 0.543–54.432, $P = 0.150$)。子宮がんは乳がんと同様にエストロゲンホルモンが関与するがんである。したがって、これらの結果は乳がんの *HMCN1* の VAF が予後因子であることの妥当性を一部証明するものであると考えられる。

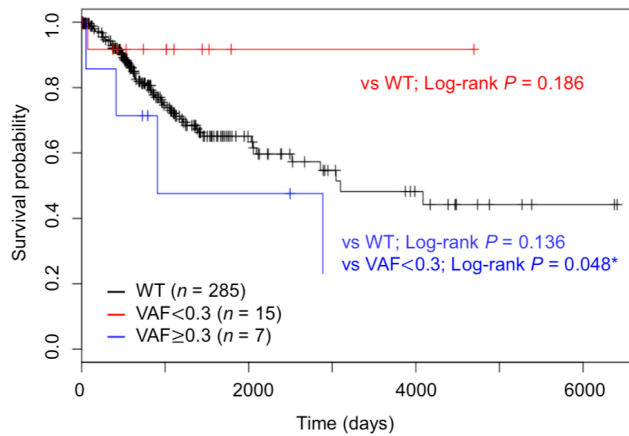


図 2-17 CESC における HMCN1 の VAF に基づく Kaplan-Meier 曲線

HMCN1 の変異を持つサンプルを、VAF のカットオフ 0.30 にして 2 つのグループに分けた (<0.30; 赤色, $n = 15$ および ≥ 0.30 ; 青色, $n = 7$)。また、比較として *HMCN1* 変異を持たない WT (黒色, $n = 285$) も用いた。Log-rank 検定を用いて、2 つの生存曲線間の差の統計的有意性を評価した結果を示している (VAF ≥ 0.30 対 VAF < 0.30, VAF ≤ 0.30 対 WT および VAF ≥ 0.30 対 WT)。アスタリスクは統計的有意性を示す。

2.3.7 *HMCN1* の VAF と臨床情報との関連

HMCN1 の変異を持つ 64 サンプルにおける個々の臨床的特徴と *HMCN1* の VAF との関連を評価するために、リンパ節転移の有無、がんのステージ、腫瘍サイズおよび診断時年齢のデータを用いた。64 サンプルは、リンパ節の転移状態によって 3 つの群 (N0, N1, N2-N3) に、ステージによって 3 つの群 (ステージ 1, 2, 3-4) に分けた。Cochran-Armitage 傾向検定を用いた解析の結果、リンパ節の転移状態 ($P = 0.029$) と *HMCN1* の VAF との間で有意な関連を見出した (図 2-18A)。一方、ステージ ($P = 0.151$) および腫瘍サイズ ($P = 0.283$) は、*HMCN1* の VAF と有意な関連が見られなかった (図 2-18B, C)。VAF 高値群と VAF 低値群の患者の平均診断時年齢は、それぞれ 58.05 歳 ($SD = 17.95$) および 61.41 歳 ($SD = 11.85$) であり、t 検定の結果、2 群間の平均年齢に有意差は見られなかった ($P = 0.461$) (図 2-18D)。 *HMCN1* の VAF と患者の診断時年齢との間におけるスピアマンの順位相関係数は -0.056 ($P = 0.659$) であり、有意な相関を示さなかった。これらの結果は、*HMCN1* の変異が乳がんの転移に影響を与えており、それによって予後の悪化がもたらされる可能性を示唆している。

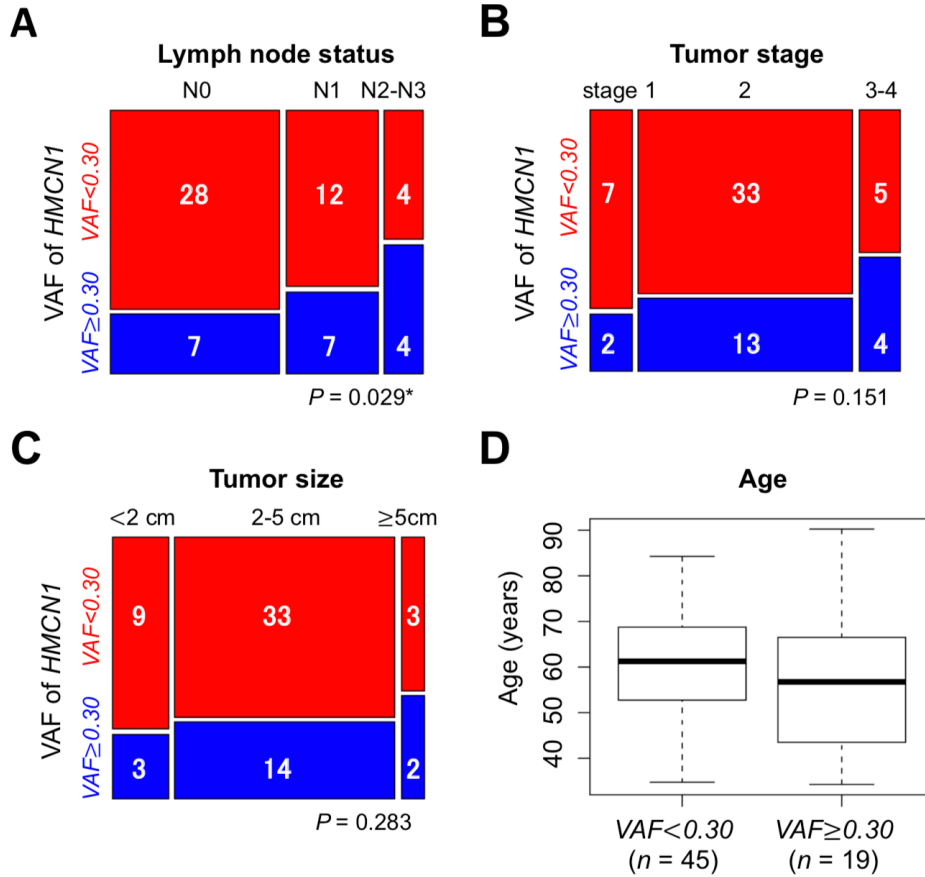


図 2-18 臨床的特徴と *HMCN1* の VAF との関連

(A) リンパ節の転移状態、(B) ステージ、(C) 腫瘍サイズ、および (D) 患者の診断時年齢と *HMCN1* の VAF との関連を示した。*HMCN1* の変異を持つサンプルにおいて、VAF のカットオフを 0.30 とし 2 つのグループに分けた (< 0.30, 赤色, $n = 45$ および ≥ 0.30 , 青色, $n = 19$)。モザイクプロット内の青色と赤色の四角は、それぞれ VAF 高値群と低値群におけるサンプルの数を示す。TCGA データセットでは、2 例のリンパ節の転移状態は不明であった。アスタリスクは統計的有意性を示す。

2.4 考察

乳がんは、その生物学的特徴によりいくつかのサブタイプに分類されるが、さらに同一腫瘍内において非常に高い ITH を示すことも知られている。本研究では、ITH を考慮した乳がんの新しい予後関連因子を同定することを目的に、TCGA に登録されているサンプルの持つ変異情報から得た VAF を用いて解析を行った。VAF に基づく生存時間解析により、*HMCN1* の VAF が乳がんの予後と関連している可能性があることを示した。さらに、*CA9* および *CASP14* の発現量の上昇に基づく腫瘍内の悪性度の亢進や[26-28]、リンパ節への転移との関連についても確認されたことから、*HMCN1* の変異はがん細胞の浸潤および転移と関連している可能性も示唆された。これらの結果を独立した別データ（乳がん患者の DRIVE データセット、CIMBA データセット、FoundationOne データセットのデータ）を用いて検証しようと試みたが、VAF もしくは生存時間情報を得ることができなかったため解析ができなかった。しかし、TCGA の子宮頸部扁平上皮がんおよび子宮頸部腺がんのデータを用いて解析を行ったところ、*HMCN1* の VAF に基づく 2 つのグループ間で生存時間に有意差が見られた。子宮頸部扁平上皮がんおよび子宮頸部腺がんは乳がんと同様に女性特有のがんであり、体内のエストロゲンホルモンの量と関連していることが知られている。したがってこの結果は、乳がんに対する *HMCN1* の予後への影響を裏付ける可能性があるが、より多くのコホートを蓄積することによって結果の妥当性をさらに評価することが必要である。

HMCN1 遺伝子は、免疫グロブリンスーパーファミリーに属する細胞外タンパク質である Hemicentin-1 をコードし、フォン・ヴィレブランド因子および免疫グロブリン様 C2 型ドメインを含むいくつかの異なるドメインからなる[33]。ヒトにおける *HMCN1* の詳細な機能は依然として不明であるが、*HMCN1* の変異は加齢性黄斑変性に関連することが示されている[34]。最近の研究によると、*HMCN1* は胆嚢がん転移の抑制因子として作用し[35]、頭頸部扁平上皮がんのサンプルにおいて共通して変異を持つことが分かっている[36]。

本研究の結果を踏まえ、乳がんの浸潤および転移に対して HMCN1 の突然変異体が及ぼす影響について次のような考察を行った：HMCN1 (別名 FBLN6 (フィブリン 6)) は細胞外マトリックス (ECM) タンパク質の 1 つであり [37, 38], 細胞接着機能を有するエストロゲン受容体結合部位を含むことが知られている [39, 40]. フィブリンは、基底膜および細胞間相互作用の形成に関与し, ECM 構造の組織化および安定化を導くことが示されている [41]. 以前の研究では、エピジェネティック修飾によってサイレンシングされたフィブリン 5 が肺がんにおける浸潤および転移を促進することが報告されている [42]. したがって、遺伝子の変異に起因する HMCN1 の不安定化の結果、十分な細胞接着が阻害され、がんの浸潤や転移が促進されると考えられる。

がんの転移は乳がん患者の主要な死因であるが、転移に関与する因子やメカニズムの解明は十分に進んでいない。これは、転移がんサンプルの入手が困難であることに加え、がんはその増殖に伴い数多くの変異を蓄積した複雑な集団になっており、その中から転移を引き起こす原因遺伝子を同定するのが困難なためである [43]. 転移腫瘍内で形成される ITH の構成は原発腫瘍内の ITH の構成と大きく異なっていることが知られているが、この違いが治療効果に影響を与える可能性がある [44]. したがって、転移メカニズムの解明はがんの効果的な治療のために強く望まれている。本研究で用いたアプローチは、がんの転移や再発に関わるがん関連遺伝子を探索するために有用であると考えられる。

この研究では、非同義置換および短い挿入・欠失に焦点を当てた。しかしながら、ITH にはシス調節領域やスプライシング領域のような非コード領域における突然変異や染色体の CNV も関与していることから [45-48], これらの変異を含めた網羅的な解析を行うことでより多くのがん関連遺伝子を同定することができるだろう。さらに、エピジェネティック修飾の変化もまたがんの進行を促進する可能性がある [49]. 正常細胞から得た全ゲノムのエピジェネティックデータセットが急速に増加してきており [50], いずれがん細胞のエピゲノムデータ解析によって、エピジェネティック因子の ITH への影響を評価することがで

きるようになると考えられる。

本研究は TCGA に登録されたゲノムおよび発現データの比較分析によって、*HMCN1* を乳がんの転移に関連する候補因子として同定した最初の研究である。現在用いられている分子タイプに基づく乳がんの 4 分類に加えて、がんの進化過程を反映する VAF を用いることで乳がんサンプルの新たな遺伝的プロファイル情報を得ることができる。本研究のアプローチにより、新しい診断マーカーもしくは治療標的となる候補遺伝子を同定することができ、これにより Precision Medicine が促進されると期待される。

3 章 多次元尺度を用いた患者予後に関与する腫瘍内不均一性の網羅的解析

3.1 緒言

がんは様々な変異の蓄積により細胞増殖や細胞周期の調節に異常をきたした細胞群のことである。近年登場した次世代シーケンシング技術により、がん細胞における変異の蓄積過程を推測することが可能になった。例えば、1つの腫瘍の複数の領域から得たサンプルを用いた「多領域シーケンシング」により、ある変異が腫瘍の様々な場所で共通して保持されている変異（クローナル突然変異）か、一部の場所で生じた変異（サブクローナル突然変異）かを推測することができる。前者はがんの進化初期に、後者はがんの進化後期に生じた可能性が高いと考えられる。これらの情報をもとにすると、クローナルおよびサブクローナル変異を系統樹の幹および枝として描く「がんの進化系統樹」を作成することができる[4, 24, 51]。また、がん細胞における変異蓄積過程を推測する別の方法として、変異アレル頻度 (variant allele frequency, VAF) を使用する方法がある。VAFにより、がん細胞中である変異を含む細胞集団の割合を推定することができる[52, 53]。すなわち、VAFが高いほど腫瘍内の変異の頻度が高いことを表すため、がんの進化過程において初期に出現した変異であると推測できる。これらの研究により、がんの進化ががんの種類間においても、さらには同じがん種の個体間においても非常に多様であることが明らかになった。

これまでの研究や臨床的知見から、がんの進化が進むほど腫瘍内不均一性 (ITH) が進むため、患者の予後が悪くなると考えられてきた。そのため、ITHを予後予測因子として用いるためのいくつかの指標が開発されてきた。例えば、腫瘍内のサブクローナル集団の数を推定し、がん細胞のITHの程度を定量化するアルゴリズムが開発され[22, 54]、予後との関係が解析された。その結果、ITHが高いサンプルほど（細胞内のサブクローン数が多い

ほど) 予後が悪いという線形の関係ではなく, ITH の程度と予後との間には非線形の関係があることが示された[55]. すなわち, 中程度の ITH であるサンプルにおいて最も予後が悪いという結果であった. このほか, VAF のばらつきを表すための変異アレル腫瘍不均一性指標 (mutant-allele tumor heterogeneity, MATH) (2.3.3 参照) [23]や, 体細胞突然変異パターンの不確実性を表すためのエントロピーに基づく突然変異対立遺伝子率 (entropy-based mutation allele fraction, EMAF) [56]と予後との関係についても報告された.

MATH は下式から算出される.

$$MATH = \frac{MAD}{\text{median}(x)} = \frac{\text{median}(|x_i - \text{median}(x)|)}{\text{median}(x)}$$

ここで, x_i をあるサンプルにおける各変異の VAF とする ($i = 1, 2, \dots, n$). ただし MAD は中央絶対偏差 (median absolute deviation) を表す.

また, EMAF は下式から算出される.

$$EMAF = - \sum_{i=1}^n \Delta f(x_i) \log f(x_i) = - \int f(x) \log f(x) dx$$

ここで, $f(x)$ を x の確率密度関数とする.

MATH が高い場合には, 頭頸部扁平上皮がんにおいて予後不良と有意に関連しており, EMAF が高い場合には, 非小細胞肺癌において予後不良と有意に関連していることが報告されている.

このように, がんの進化によってもたらされる ITH と患者の予後との関連について様々な研究結果が蓄積されてきているにもかかわらず, 十分な知見が得られているとは言い難い. なぜなら, これらの ITH の指標は「ITH が高いほど予後が悪い」という関係をすべてのがん種で示すことができる頑健なものではないためである. これは, がんの種類によってその進化の特徴が異なっていることによるものであり, これまで用いられてきた1次元の指標では予後との関係を十分に表すことができていないのではないかと考えられる.

そこで本研究では, ITH を多次元的に捉えて患者の予後との関係を解析するため

に、各サンプルの持つ変異由来の VAF の分布を用いることとした。VAF の分布の形状を表すパラメータとして、各サンプルの持つ突然変異の数、VAF の分布のピークの位置、VAF のばらつき (MATH) という 3 つのパラメータを採用した。TCGA に登録されている 16 種類のがん由来の 6,064 サンプルのデータを用いて本解析を行った (図 3-1)。

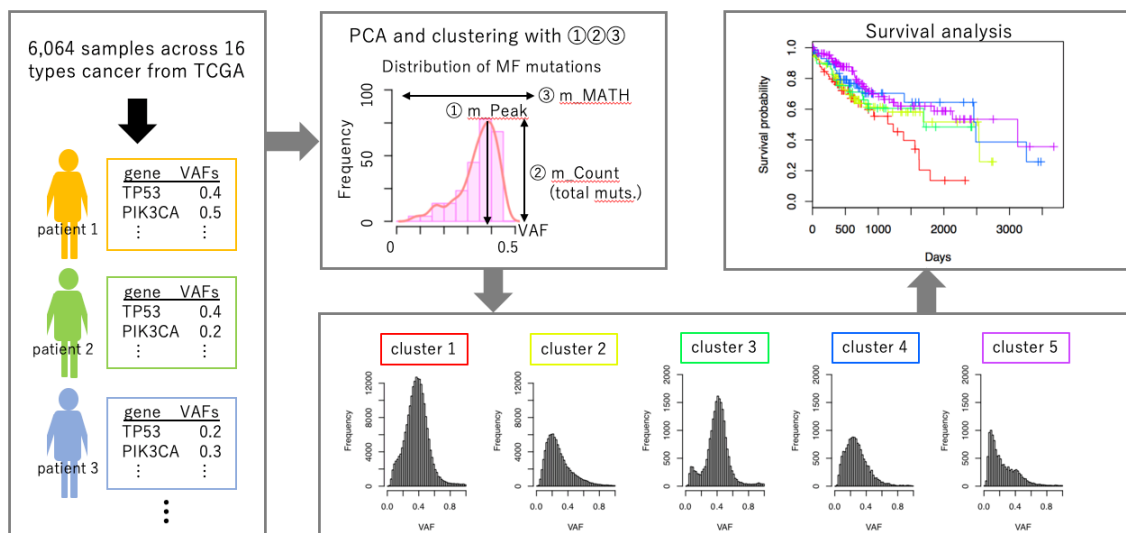


図 3-1 本研究の要約図

TCGA から得たサンプル中の各突然変異について VAF を計算した。各サンプルの VAF の分布の形状から算出された 3 つのパラメータ (①②③) を使用して、サンプルを 5 つのクラスターに分類した。各クラスターに属するサンプル間の予後と比較するために生存時間分析を行った。

3.2 方法

3.2.1 データセット

体細胞変異の VCF データは TCGA (<https://portal.gdc.cancer.gov/>) からダウンロードした。VCF ファイルはバリエントコーラーの 1 つである MuTect2 によって「がん細胞」と「同一患者から得たコントロール細胞」との比較によって得られたものを用いた。本研究で用いたがんは以下の 16 種類である：膀胱尿路上皮がん (BLCA)、浸潤性乳がん (BRCA)、子宮頸部扁平上皮がんおよび子宮頸部腺がん (CESC)、結腸腺がん (COAD)、多形性膠芽腫 (GBM)、頭頸部扁平上皮がん (HNSC)、腎明細胞がん (KIRC)、低悪性度グリオーマ (LGG)、肝細胞がん (LIHC)、肺腺がん (LUAD)、肺扁平上皮細胞がん (LUSC)、卵巣漿液性腺がん (OV)、前立腺がん (PRAD)、皮膚黒色腫 (SKCM)、甲状腺がん (THCA)、および子宮体部類内膜がん (UCEC)。また、体細胞変異データと同じサンプル由来の CNV データと臨床データもダウンロードした[8]。

3.2.2 突然変異分析

本研究で用いた変異データの平均カバレッジ深度 (average depth of coverage) は 99.4 であり、十分な精度を確保するためカバレッジ深度が 20 以上 (全体の 93.3%) である点突然変異 (SNV) を研究に用いた。さらに、PolyPhen-2 によって「probably damaging (> 0.85)」または「possibly damaging (0.15–0.85)」と分類された変異を「機能的 (more functional, MF) 変異」と見なし、以降の研究に用いた[32, 57]。MF 変異は、タンパク質の構造もしくは機能に影響を与えると考えられることから、がんの発生や進行、転移などに関与する可能性がある。これらの変異情報に対して CNV データとの統合を行い、CNV が含まれる変異は除外した。CNV の領域にある SNV の場合、CNV と SNV のうちどちらが先に生じたのかを予測するのは難しく、そのため SNV の VAF の値が正確に推定できない可能性があるためである。変異の抽出基準として、統合した CNV のセグメント平均値 (segment mean)

が-0.2 と 0.2 との間（コピー数が $2 \times 2^{-0.2} = 1.74$ から $2 \times 2^{0.2} = 2.30$ に該当する）であり、CNV コールの偽陽性率を低減させるため、ターゲットキャプチャーに用いたプローブを 10 個以上含む領域にあるという条件を設定した[58]. 以上の条件により抽出された変異を 2 つ以上持つサンプルのみを本研究の解析に用いた. 各変異について VAF を算出し、VAF を ESTIMATE R パッケージによってサンプルごとに推定された腫瘍の純度で調整した[21].

3.2.3 VAF の分布の形状を定義するパラメータの計算

サンプルごとに、VAF の分布の形状を定義する 3 つのパラメータを算出した. パラメータは次の 3 つである：MF 変異の VAF の分布の確率密度関数の最大値に対応する VAF (m_Peak), \log_2 [変異の総数] (m_Count) および MF 変異の MATH スコア (m_MATH) [23]. m_Peak は VAF の分布のピーク位置を表し、 m_Count は分布の大きさを表し、 m_MATH は VAF のばらつきを表す.

3.2.4 統計解析

統計解析は、R (version 3.3.1) (R Project for Statistical Computing, Vienna, Austria), および JMP Pro (version 13.0) (SAS Institute Inc., Cary, NC, USA) を用いて行った. サンプルをクラスタリングするために、R の cluster パッケージ (version 2.0.6) を使用して 2 乗ユークリッド距離から k -medoids クラスタリングを行った.

カイ二乗検定またはフィッシャーの正確検定（分割表において 1 つ以上のセルの期待度数が 5 以下であった場合）を用いて、カテゴリカルデータを比較した. 連続データの場合、3 群以上の比較は ANOVA を用いて行い、Tukey 法を用いて多重比較を行った.

生存時間分析においては、R の survival パッケージ (バージョン 2.41-3) を用いて Cox 比例ハザード回帰分析を行い、ハザード比 (HR) および 95%信頼区間 (95%CI) を算出した. 同パッケージを用いて比例ハザード性の仮定の評価も行った.

16 種類のがんの間でゲノムデータや臨床データの比較を行う際には, Benjamini-Hochberg 法により P 値の補正を行った[25]. 本研究では P 値が 0.05 以下の場合を統計的に有意であると見なした (P 値が 0.05 以下の場合には*を, 0.01 以下の場合には**を, 0.001 以下の場合には***を付した).

3.3 結果

3.3.1 VAF の分布を用いたサンプルのクラスタリング

TCGA レポジトリより 16 種類のがん由来の 6,064 サンプルの突然変異データ (VCF ファイル) を取得した (表 3-1).

表 3-1 本研究で用いた 16 種類のがんとそのサンプルサイズ

	がんの種類	サンプルサイズ
BLCA	bladder urothelial carcinoma (膀胱尿路上皮がん)	400
BRCA	breast invasive carcinoma (乳腺腺がん)	935
CESC	cervical squamous cell carcinoma and endocervical adenocarcinoma (子宮頸部扁平上皮がんおよび子宮頸部腺がん)	274
COAD	colon adenocarcinoma (結腸腺がん)	265
GBM	glioblastoma multiforme (多形性膠芽腫)	137
HNSC	head and neck squamous cell carcinoma (頭頸部扁平上皮がん)	478
KIRC	kidney renal clear cell carcinoma (腎明細胞がん)	304
LGG	lower grade glioma (低悪性度グリオーマ)	474
LIHC	liver hepatocellular carcinoma (肝細胞がん)	349
LUAD	lung adenocarcinoma (肺腺がん)	471
LUSC	lung squamous cell carcinoma (肺扁平上皮細胞がん)	459
OV	ovarian serous cystadenocarcinoma (卵巣漿液性腺がん)	155
PRAD	prostate adenocarcinoma (前立腺がん)	459
SKCM	skin cutaneous melanoma (皮膚黒色腫)	445
THCA	thyroid carcinoma (甲状腺がん)	295
UCEC	uterine corpus endometrial carcinoma (子宮体部類内膜がん)	164

本研究では、PolyPhen-2 によって「probably damaging」または「possibly damaging」のいずれかに分類された突然変異を MF 変異と定義し、以降の解析に用いた。16 種類のがんにおける MF 変異の合計は 469,553 であった（補足表 3-1）。各サンプルの MF 変異の VAF の分布の形状を定義するために、「分布のピークの位置」、「分布の大きさ」、「VAF のばらつき」という 3 つのパラメータを用いた。「分布のピーク位置」は「m_Peak」と表し、各サンプルの持つ MF 変異の VAF の分布から確率密度関数を算出し、密度関数が最大値を取るときの VAF と定義した。「分布の大きさ」は「m_Count」と表し、各サンプルの持つ変異の総数を log2 変換したものと定義した。「VAF のばらつき」は「m_MATH」と表し、各サンプルの持つ MF 変異の VAF を用いて算出した MATH と定義した。まずはじめに、3 つのパラメータの相関関係を調べるために、すべての可能な組み合わせにおける相関係数を計算した。m_Peak と m_MATH、m_Peak と m_Count、m_Count と m_MATH の間に観測された相関係数はそれぞれ -0.44、0.03、0.00 であった。この結果より、これら 3 変数の間には強い相関がないことが確認された。

次に、16 種類のがんの個々のサンプルの VAF の分布パターンの分類を行った。6,064 サンプルから得た VAF の分布の形状を定義する 3 つのパラメータを用いて主成分分析（principal component analysis, PCA）を行った（表 3-2）。第 1 主成分と第 2 主成分の寄与率は 48.2% および 33.3% であり、2 つの主成分によって累積寄与率が 80% 以上となることから、第 1 主成分及び第 2 主成分（PC1、PC2）を以後の分析に用いた（図 3-2A）。これら 2 つの主成分を用いて、6,064 のサンプルを *k*-medoids アルゴリズムによって 5 つのクラスターに分割した（図 3-2B）。すなわち、本研究では *k*=5 を採用した。各クラスターに属するすべてのサンプルの持つ変異を統合した VAF のヒストグラムを図 3-3A に示した。さらに、16 種類のがんについて 5 つのクラスターの割合（図 3-3B）、および 5 つのクラスターにおける 3 つのパラメータの中央値を計算した（表 3-3）。

表 3-2 各クラスターのパラメータの主成分負荷量

	PC1	PC2	PC3
m_Peak	0.707	-0.007	0.707
m_Count	0.055	0.997	-0.045
m_MATH	-0.705	0.071	0.706

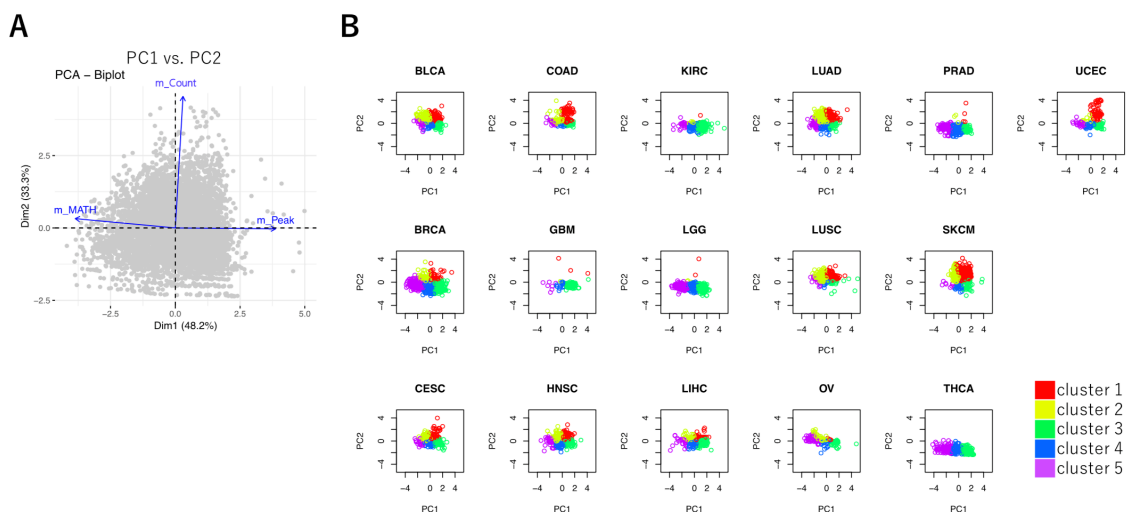


図 3-2 16 種類のがんのサンプル由来の 3 つのパラメータを用いた PCA の結果

(A) 16 種類のがんのサンプルを使用した PCA の結果を表す。PCA によって得られた PC1 (x 軸) と PC2 (y 軸) の値をプロットした。各軸ラベルのパーセンテージは、寄与率を示している。(B) 16 種類のがんにおいて、5 つのクラスターのいずれかに分けられたサンプルを表す。x 軸は PC1 を、y 軸は PC2 を示している。散布図の右側に 5 つのクラスターのカラーコードを示している。

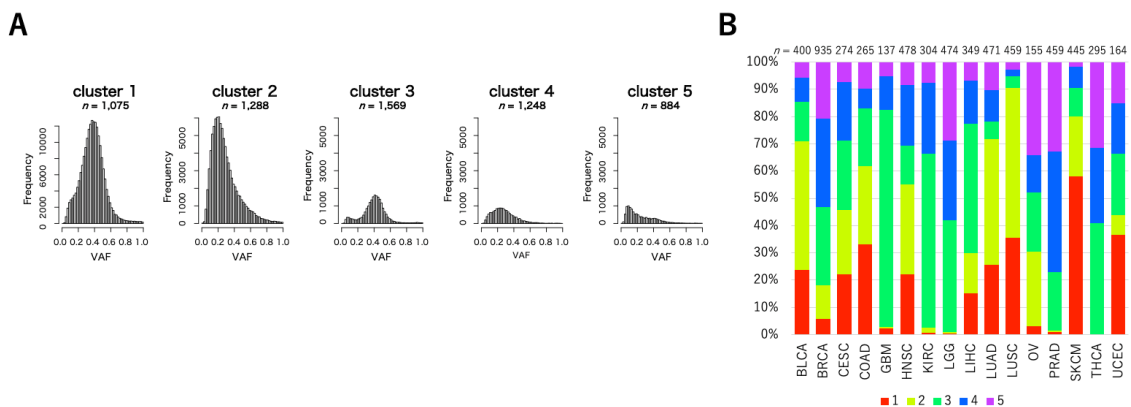


図 3-3 各クラスターの VAF の分布と 16 種類のがんにおける各クラスターの割合

(A) 5つのクラスターに属するサンプルの持つ変異の VAF のヒストグラムを示す。横軸は VAF を、縦軸は変異の数を示す。(B) 各がん種の 5つのクラスターの割合 (サンプル頻度) を示す。グラフ上部の数は、本研究で分析を行ったサンプルの数を表す。棒グラフの下部に 5つのクラスターのカラーコードを示している。

表 3-3 各クラスターのパラメータの中央値

パラメータ	クラスター				
	1	2	3	4	5
m_Peak の中央値	0.376	0.229	0.424	0.271	0.140
m_Count の中央値	6.794	6.119	4.000	3.807	3.807
m_MATH の中央値	0.186	0.320	0.133	0.251	0.473

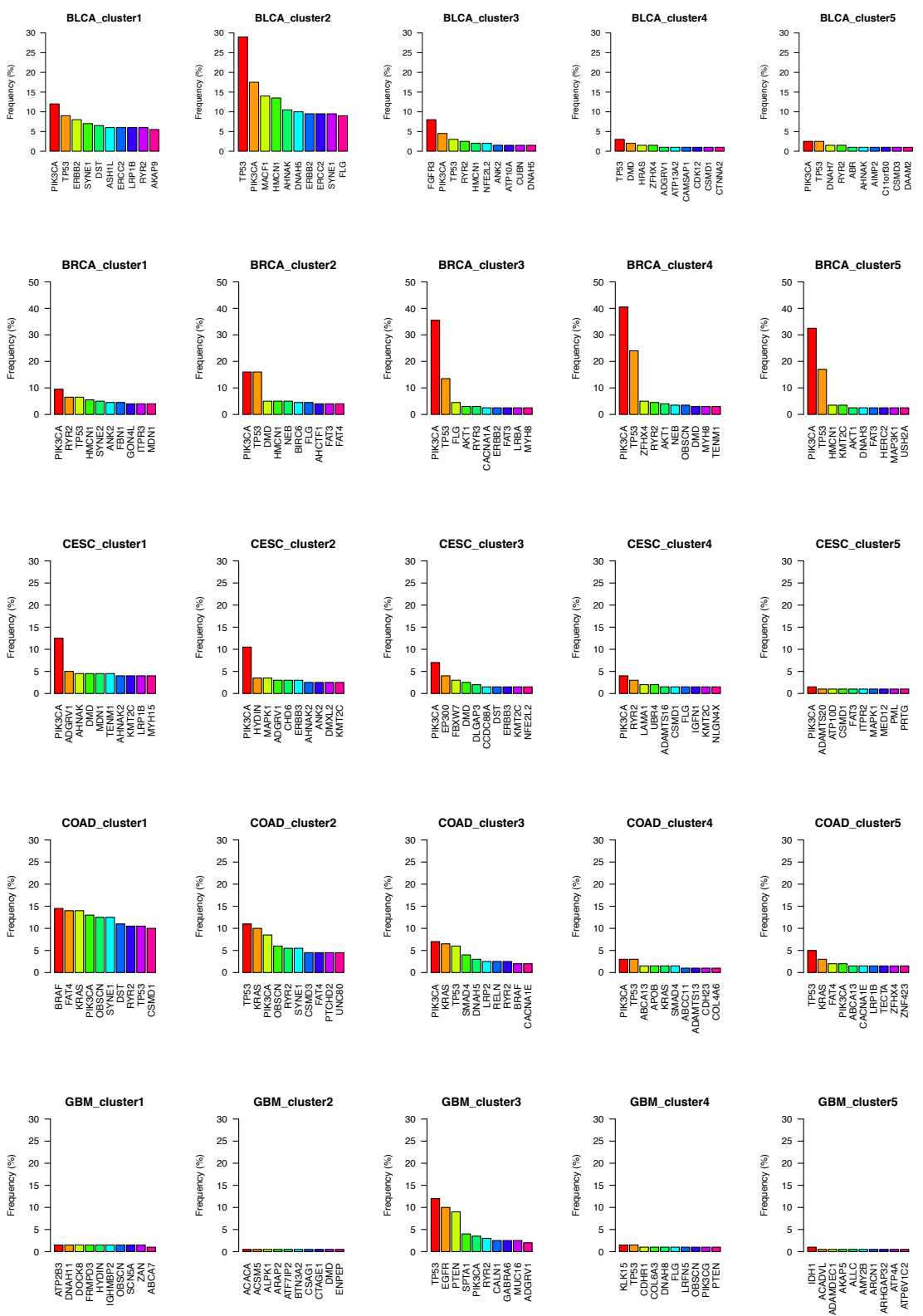
クラスタリングの結果、クラスター1および2のサンプルは、他の3つのクラスターよりも多くの変異を有していた。VAFの分布を見ると、クラスター1のサンプルは、低いVAFに比べて高いVAFの変異をより多く持っていた。一方、クラスター2のサンプルは、高いVAFに比べて低いVAFの変異がより多く生じていた。これらの結果から、クラスター1ではクローナルな変異が、クラスター2ではサブクローナルな変異がより多く蓄積していることが予想された[59, 60]。図3-3Bに示すように、クラスター1の頻度はSKCM（メラノーマ）およびLUSC（肺扁平上皮がん）において比較的高かった。この結果は、メラノーマおよび肺がんを含むいくつかの種類のがんにおいて、発がん前に突然変異が蓄積されているというMcGranahanらの研究結果をサポートするものであった[24]。クラスター2の頻度はBLCA（膀胱がん）、LUAD（肺腺がん）、およびLUSCで高かった。これらのがんでは、多くのサブクローナル変異が発生することが先行研究にて既に確認されている[24]。

クラスター3, 4, 5のサンプルは、クラスター1, 2に比べて変異の数が少なかった。これらのクラスターの頻度は、GBM（膠芽腫）、KIRC（腎臓がん）、LGG（グリオーマ）、PRAD（前立腺がん）、THCA（甲状腺がん）において比較的高かった。これまでの研究において、腎臓、脳、前立腺および甲状腺の腫瘍では突然変異の数が他のがんと比べて比較的少ないということが知られており[61]、本研究の結果はこの知見をサポートするものであった。次に、これら3つのクラスター間の違いに着目した。クラスター3のサンプルはm_Peakが高く、m_MATHが低かったのに対し、クラスター5のサンプルはm_Peakが低く、m_MATHが高かった（表3-3）。この結果は、クラスター3のサンプルでは主たる変異ががん発生の初期段階で起こり、がんの進行に従って大量の突然変異が積極的に蓄積されずに、初期に生じた突然変異が維持される細胞集団であると解釈することができる。クラスター4は、クラスター2と同様に中程度のVAFのピークを持つサンプル群である。このピークは、強い正の選択下で増大した特定の変異を持つサブクローンの存在に起因する可能性がある[62, 63]。すなわち、がんの進化中～後期に出現した、がんの生存にとって好ましい変異を

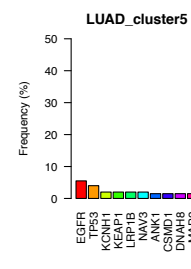
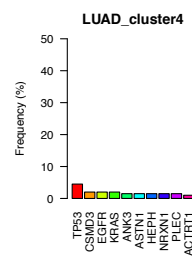
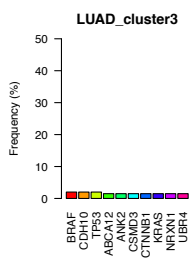
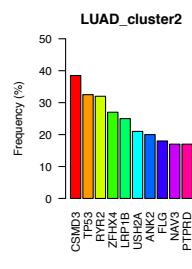
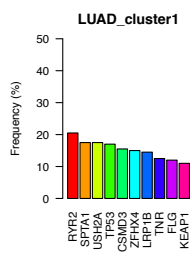
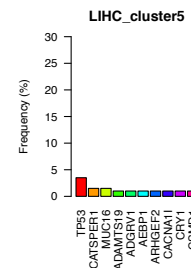
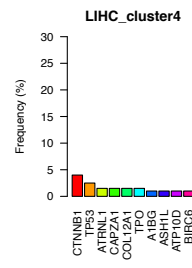
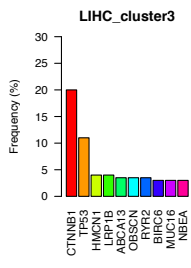
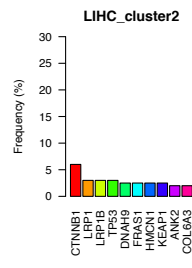
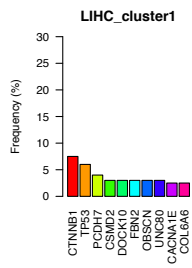
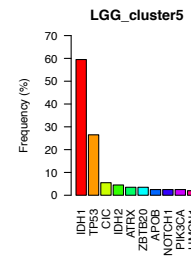
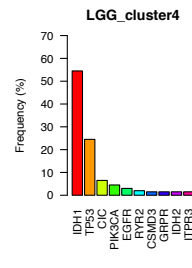
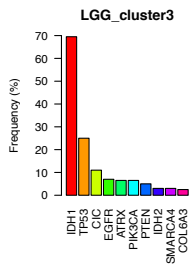
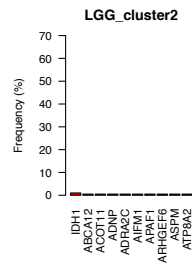
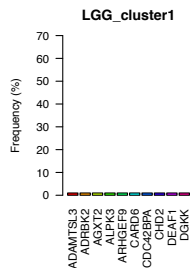
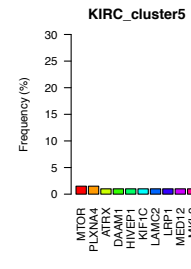
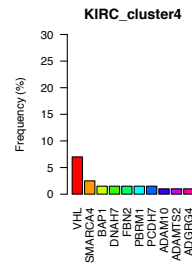
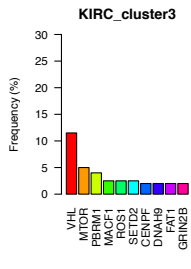
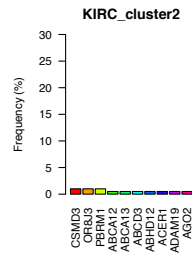
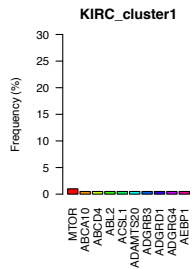
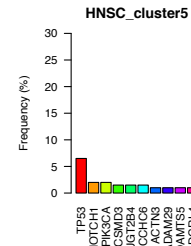
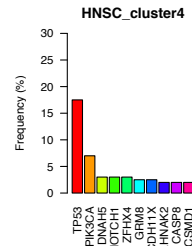
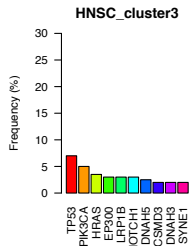
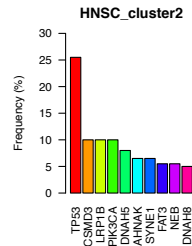
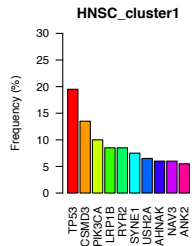
持つ細胞が増殖した結果であると考えられる。クラスター5 のサンプルは非常に低い VAF のピークを持つサンプル群であり、これらはがんの中立進化過程において起こりうる変異を多く有していると考えられる[64]。PRAD においてはクラスター4 および 5 の頻度が特に高かった。前立腺がんは進化の過程で多くのサブクローンを持つ枝分かれ型のがんであることが知られており、本研究の結果はこの知見と一致するものであった[65]。

3.3.2 5つのクラスターのゲノム上の特性

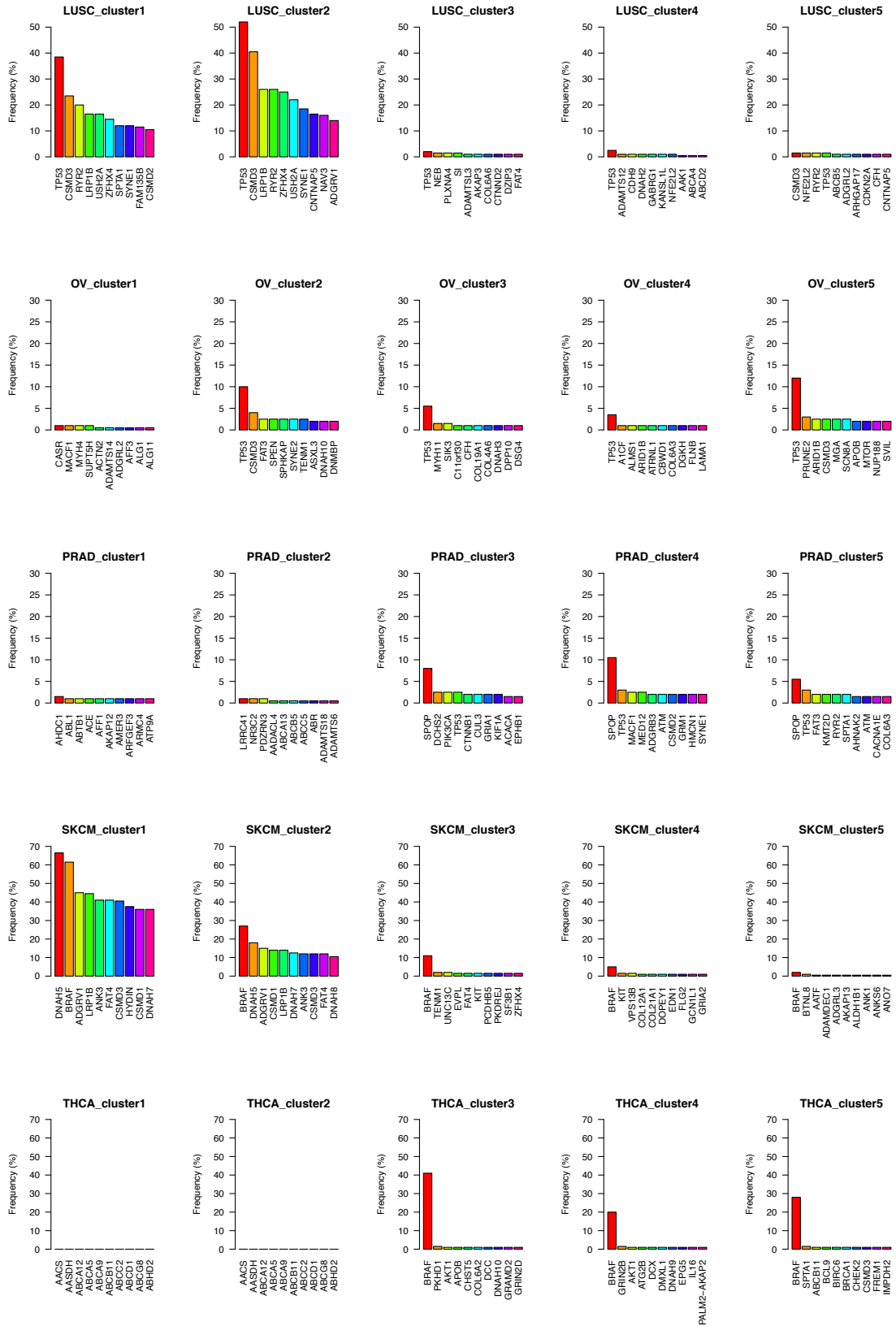
5つのクラスターのゲノム上の特性を評価するために、16種類のがんそれぞれに対して各クラスター内で発生頻度の高い上位10遺伝子を集計した(図3-4)。



(ㄅㄅ)



(つづく)



(つづく)

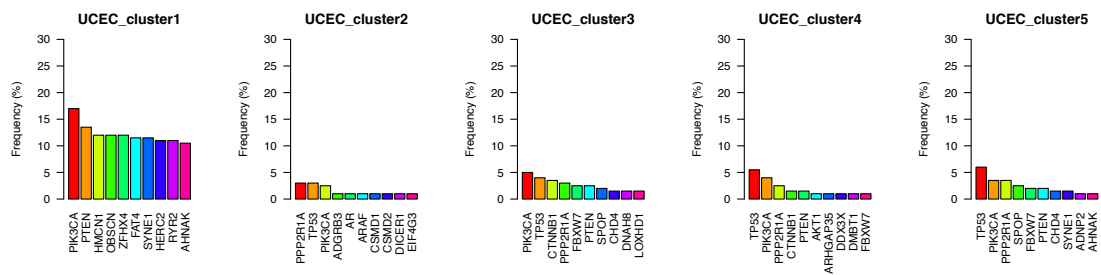


図 3-4 16 種類のがんの各クラスターにおける変異発生頻度上位 10 遺伝子

16 種類のがんの各クラスターにおいて MF 変異の発生頻度が高い上位 10 個の遺伝子を示す。縦軸は変異の発生頻度（変異を持つサンプルの割合）を表している。

BRCA（乳がん）ではクラスター3, 4, 5において *PIK3CA* の変異発生頻度が比較的高かった（5つのクラスターにおける発生頻度はそれぞれ 9.5%, 16.0%, 35.5%, 40.5%, 32.5%）。これらのクラスターはクラスター1, 2に比べて変異の数が少ないクラスターである。BRCAのクラスター3, 4, 5において *PIK3CA* の VAF の分布を見ると、VAF の平均がそれぞれ 0.469, 0.432, 0.390 であり VAF の分布のピーク位置から想定されるよりも高かった（図 3-5）。これらの結果より、乳がんでは多くの場合がんの発生初期にドライバー遺伝子である *PIK3CA* に変異が起こった後、積極的な変異の増加が起こらず、*PIK3CA* の変異を持つ細胞が細胞集団内で比較的安定に維持される可能性がある。同様に、LIHC（肝臓がん）ではクラスター3においてドライバー遺伝子である *CTNNB1* の変異発生頻度が高かった（5つのクラスターにおける発生頻度はそれぞれ 7.5%, 6.0%, 20.0%, 4.0%, < 1.0%）。LIHC のクラスター3において *CTNNB1* の VAF の分布を見ると VAF の平均が 0.446 であることから、*CTNNB1* の突然変異は多くの場合がんの進化過程初期に発生し、がん細胞集団中の大部分を占拠していると考えられる（図 3-5）。SKCM（メラノーマ）やUCEC（子宮がん）では主としてクラスター1において発生頻度の高い変異が検出されたが、同様に変異の数が多いクラスター2ではこれらの変異の発生頻度はクラスター1ほど高くなかった。したがって、SKCM やUCEC ではがん発生初期に生じる突然変異に共通の特徴がある可能性を示唆している。

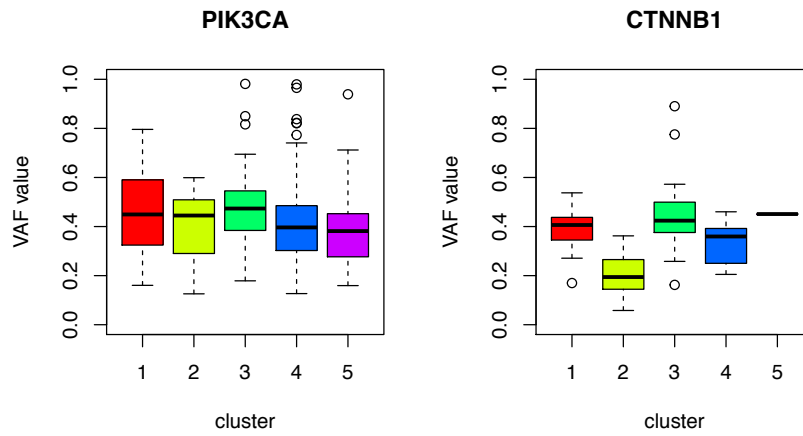


図 3-5 *PIK3CA* および *CTNNB1* の VAF の分布

BRCA における *PIK3CA* のクラスターごとの VAF と LIHC における *CTNNB1* のクラスターごとの VAF を示す。

次に ITH の程度とゲノム不安定性との関連について評価を行った。以前の研究では ITH が高いほど（サブクローン数が多いほど）ゲノム不安定性が高いという結果が得られている [55]。そこで先行研究に従いコピー数変異量 (CNV abundance) を算出し、ゲノム不安定性の指標とした (2.3.4 参照)。この解析では、クラスター2 のサンプルは MATH 値が比較的高く、VAF の低い変異を多く持つサンプル群であることから、5 つのクラスターの中で最も ITH の高いサンプルであると仮定した。すなわち、クラスター2 において CNV abundance は最も高い値になることが予想されたが、実際には BLCA (膀胱がん)、BRCA および LGG (グリオーマ) の 3 種類のがんでのみクラスター2 の CNV abundance が他のクラスターよりも有意に高かった (図 3-6A, 補足図 3-1)。これらの結果は、ITH の定義ががんによって異なること、ITH を分析する際には CNV の不均一性や CNV が SNV に及ぼす影響についても考慮すべきであることを示唆している。

最後に各クラスターの変異スペクトラムの分布の集計を行った (図 3-6B, 補足図 3-2)。変異スペクトラムとは塩基の変異パターンを 6 種類 (T>G, T>C, T>A, C>T, C>G, C>A) に分類したものであり、加齢や紫外線、タバコなどの変異をもたらす原因によってそれぞれ異なる変異パターンが生じる。変異パターンをもとに変異 signature と呼ばれる約 30 種類のパターンを用いて各サンプルに生じた変異を分類する方法も使われる [66]。本研究では 6 種類の変異パターンの頻度をクラスター間で比較した。メラノーマでは紫外線が主な原因である C>T 変異の割合が高いが、幹型突然変異 (クローナル変異) に比べて枝型突然変異 (サブクローナル変異) において C>T のトランジション頻度が低下し、T>G トランスバージョン頻度が増加することが知られている [67, 68]。本研究では SKCM のクラスター2, 3, 4, 5 ではクラスター1 と比較して C>T 変異の頻度が有意に低下していた (cluster 2; FDR-adjusted $P < 0.001$, cluster 3; FDR-adjusted $P < 0.001$, cluster 4; FDR-adjusted $P < 0.001$, and cluster 5; FDR-adjusted $P < 0.001$) ことから、クラスター1 より少ない変異を持つクラスターでは枝型変異、すなわちがんの進化過程の初期よりも後期に生じるサブク

ーナルな変異を多く含んでいると示唆される[2, 69]. 一方, 肺がんでは喫煙の影響によると考えられる C>A 変異が多いが, LUAD (肺腺がん) や LUSC (肺扁平上皮がん) では幹型変異に比べて枝型変異において C>A トランスバージョンが低下することが知られている[70]. 本研究では LUAD のクラスター2, 3, 4, 5 および LUSC のクラスター3, 4 においてクラスター1 と比較して C>A の頻度の低下が見られた (LUAD : cluster 2; FDR-adjusted $P < 0.001$, cluster 3; FDR-adjusted $P < 0.001$, cluster 4; FDR-adjusted $P < 0.001$, and cluster 5; FDR-adjusted $P < 0.001$) (LUSC : cluster 3; FDR-adjusted $P < 0.001$ and cluster 4; FDR-adjusted $P < 0.001$). これらの結果より, 上記のクラスターに属するサンプルではがんの進化過程のより後期に生じた変異を多く含んでいると考えられる. LUAD における変異スペクトラムの結果は喫煙歴の結果からも明らかであった (図 3-6C) [70, 71].

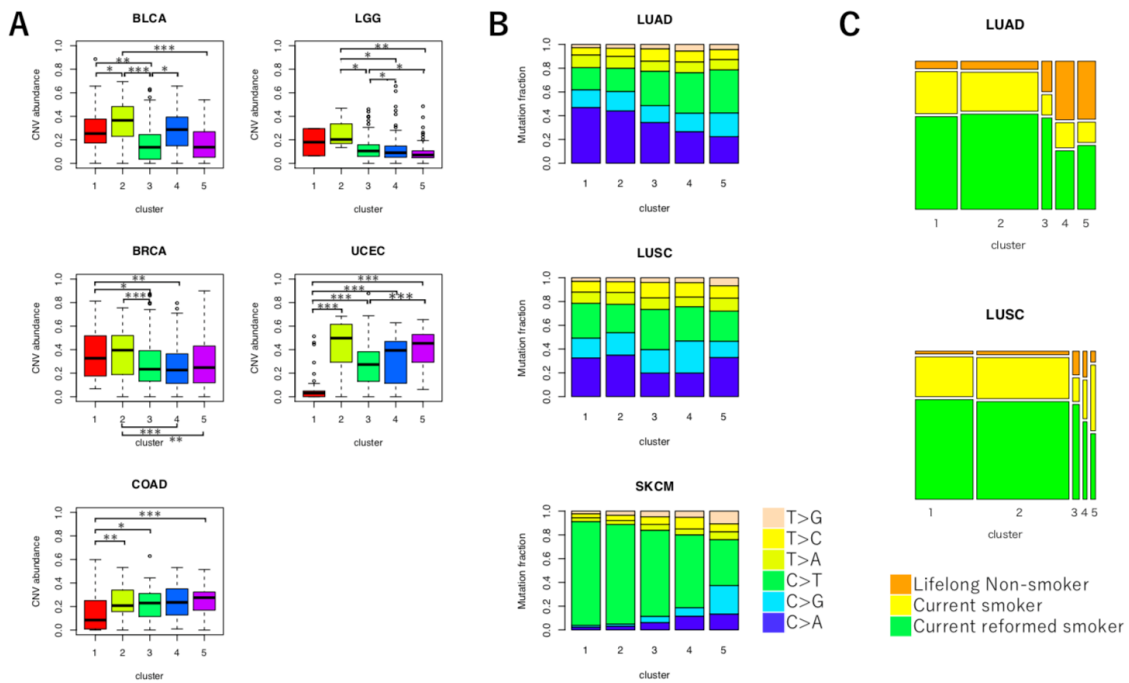


図 3-6 VAF の分布の 5 つのクラスター間の遺伝的特徴の比較

(A) 各クラスターにおける CNV abundance の分布を示す. ANOVA により平均年齢の比較を行い, 続いて Tukey 法によって多重比較を行った. ここに示したのはクラスター間で顕著な差が見られた 5 種類のがんのプロットである. P 値に応じて*を付した: * $P < 0.05$, ** $P < 0.01$, *** $P < 0.001$. (B) 3 種類のがんにおける 6 種類の変異スペクトラムの割合を示す. (C) LUAD および LUSC の患者の喫煙歴の割合を示す.

3.3.3 5つのクラスターと臨床データとの関連

VAF の分布と臨床データとの関連を見るため、がんの診断時年齢、診断時のがんのステージ、性別、および乳がんのサブタイプのデータを解析した (補足表 3-2)。クラスター間で診断時年齢に有意な差が見られたのは BLCA (膀胱がん), HNSC (頭頸部がん), LIHC (肝臓がん), LUAD (肺腺がん) の 4 種類のがんのみであった (BLCA: FDR-adjusted $P = 0.041$, HNSC: FDR-adjusted $P = 0.041$, LIHC: FDR-adjusted $P = 0.041$, and LUAD: FDR-adjusted $P = 0.041$) (図 3-7A)。このうち LUAD を除く 3 種類のがんではクラスター 1 もしくは 2 のいずれかのサンプルにおいてがんの診断時年齢が有意に高かった。年齢とともに活性酸素による DNA の損傷や細胞分裂時の複製エラーの機会が増加するため、がんの診断時年齢が上がると細胞内に蓄積されている変異の数も増えることが知られており、上記の結果はこの知見と一致するものであった [66, 72]。

診断時のがんのステージとクラスターの分布との関連を解析するため、がんのステージを I と II (低ステージ群)、および III と IV (高ステージ群) の 2 群に分けて 5 つのクラスターの頻度に差があるかどうかをカイ二乗検定により検証した。その結果 SKCM (メラノーマ) でのみクラスターの分布に有意差が確認された (FDR-adjusted $P = 0.009$) (図 3-7B)。SKCM の低ステージ群ではクラスター 1 のサンプルの頻度が有意に高く、一方、高ステージ群ではクラスター 2 と 4 のサンプルの頻度が有意に高かった (cluster 1: FDR-adjusted $P = 0.001$, cluster 2: FDR-adjusted $P = 0.024$, and cluster 4: FDR-adjusted $P = 0.042$)。この結果は、メラノーマの悪性度はがんの進化過程後期におけるサブクロンの増大によって高まる可能性を示唆している。

上記の結果とは反対に、いずれのがんにおいても性別間での 5 つのクラスターの分布に差は見られなかった。さらに、BRCA (乳がん) のサブタイプとの関連についても調べた。乳がんには主たるサブタイプが 4 種類あり (Basal-like, HER2-enriched, Luminal A, Luminal B)、うち Basal-like はトリプルネガティブ乳がんに含まれるサブタイプである [19]。

各クラスターの頻度を比較すると、Basal-like タイプの乳がんはクラスター2 で最も頻度が高かった（5 つのクラスターにおける Basal-like タイプの乳がんの頻度はそれぞれ 23.1%, 34.8%, 11.7%, 15.9%, 16.7%）（図 3-7C）。さらに HER2-enriched タイプの乳がんはクラスター1 で最も頻度が高かった（5 つのクラスターにおける HER2-enriched タイプの乳がんの頻度はそれぞれ 26.9%, 14.7%, 3.9%, 9.8%, 6.1%）。Luminal A タイプの乳がんの割合はクラスター1, 2 と比べてクラスター3, 4, 5 において比較的が高かった。これらの結果は、乳がんの進化過程の違いはサブタイプに依存する可能性があるが[73], サブタイプによってすべての事象を説明できるわけではないことを示唆している。すなわち、同一のサブタイプであっても進化過程をもとにすることで新しい乳がんの分類が可能になるのではないかと考えられる。

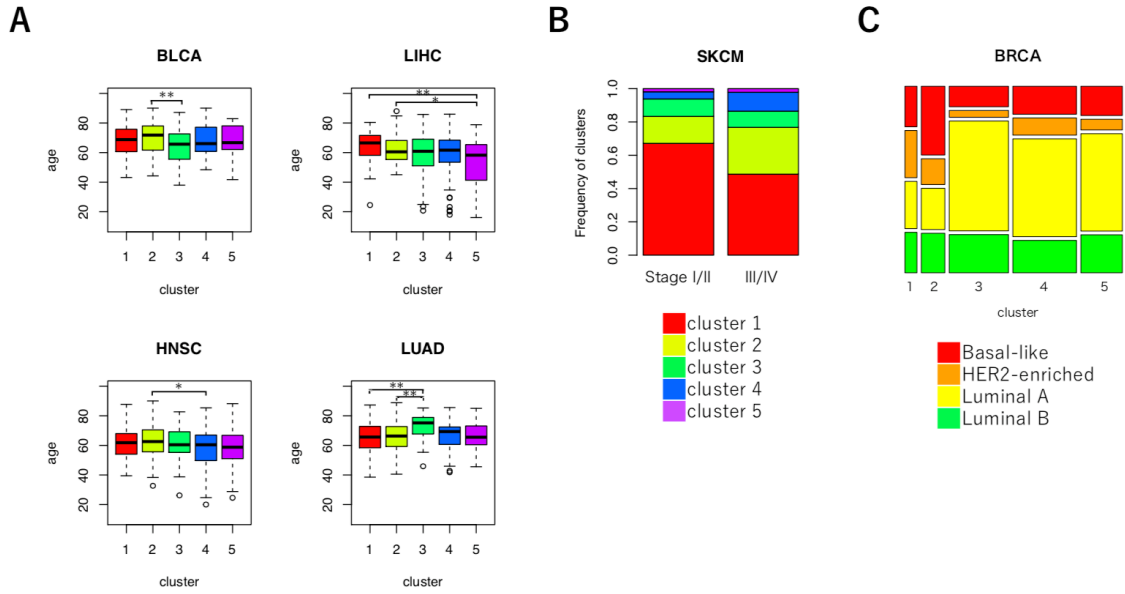


図 3-7 VAF の分布の 5 つのクラスター間の臨床的特徴の比較

(A) 各クラスターの診断時年齢の分布を示した。ANOVA により平均診断時年齢の比較を行い、続いて Tukey 法によって多重比較を行った。ここに示したのは ANOVA により統計的に有意な結果となった 4 種類のがんである。P 値に応じて*を付した： $P < 0.05$ ， $** P < 0.01$ ， $*** P < 0.001$ 。(B) SKCM のサンプルを低ステージ群（ステージ I / II）または高ステージ群（ステージ III / IV）に分けた場合の 5 つのクラスターの割合を示す。(C) BRCA における各クラスターの分子サブタイプの割合を示す。

3.3.4 5つのクラスターと予後との関係

VAF の分布と予後との関係を評価するために、Cox 比例ハザードモデルを用いた単変量解析を行った。この解析ではクラスター2のサンプルをリファレンスとして用いたが、これは3.3.2で述べたようにクラスター2が最も高度なITH状態であると予想されたためである。これまでの研究においてITHが高いほど予後が悪いことが報告されていることから、クラスター2のサンプルが最も予後が悪いと予想した。Cox解析の結果、有意な結果が得られたのはLUSC（肺扁平上皮がん）とSKCM（メラノーマ）の2種類のがんのみであった（補足表3-3）。これら2種類のがんは、仮説とは反対にクラスター2のサンプルが他のクラスターのサンプルと比較して有意に予後が良かった（LUSC：クラスター4 vs 2, HR = 4.75, $P < 0.001$ ）（SKCM：クラスター1 vs 2, HR = 1.56, $P = 0.021$, クラスター3 vs 2, HR = 3.91, $P < 0.001$, クラスター4 vs 2, HR = 2.71, $P = 0.001$ ）（補足表3-3）。この結果は、がんの種類によってITHの程度の定義が異なる可能性があること、あるいはITHが予後へもたらす影響が異なる可能性があることを示唆している。

VAF の分布と予後との間の関係を詳細に調べるために、共変量を考慮した解析も行った。この分析の目的は、診断時年齢、性別、およびがんのステージの3つの共変量を用いた多変量Cox比例ハザード回帰分析によって、予後不良をもたらすがん細胞集団の状態について評価することである。この分析では、がんごとに特定のクラスターのサンプルをリファレンスとして使用した。すなわち、HRが全て1以上（ $HR > 1$ ）となるクラスターをリファレンスクラスターとして選択した（補足表3-4）。例えばBLCA（膀胱がん）ではクラスター1をリファレンスとして使用した。解析の結果、16種類のがんのうち7種類のがん（BLCA, LGG（グリオーマ）, LIHC（肝臓がん）, LUAD（肺腺がん）, LUSC, SKCM, UCEC（子宮がん））において少なくとも1つのクラスターが予後と関連していた（図3-8）。

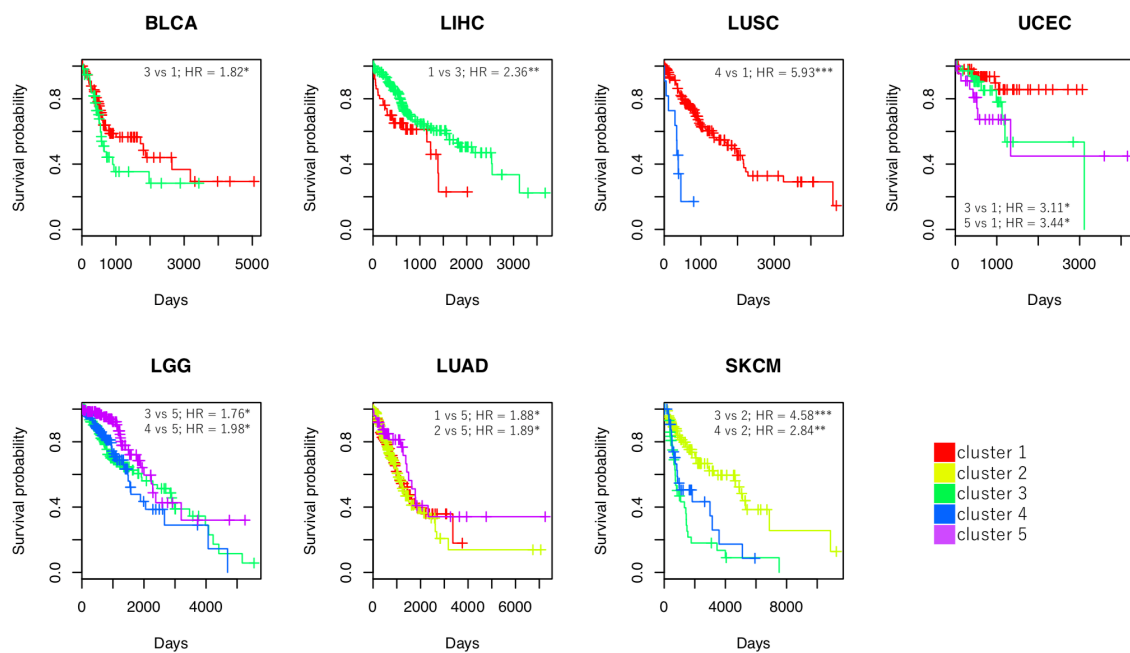


図 3-8 生存曲線および Cox 比例ハザード回帰分析から算出された HR

Cox 比例ハザード回帰分析により有意な結果を得た 7 種類のがんのクラスターとその生存曲線を示す。ここに示した生存曲線はリファレンスとしたクラスターと、リファレンスとしたクラスターに対して有意となった 1 つ以上のクラスターの生存時間をプロットしたものである。本分析は、各がん種において 10 以上のサンプルサイズを有するクラスターのみを用いて行った。横軸は時間 (日) を、縦軸は生存確率を示す。P 値に応じて*を付した：* $P < 0.05$, ** $P < 0.01$, *** $P < 0.001$ 。

LIHC におけるクラスター1のサンプルは、クラスター3のサンプルと比較して予後が悪かった ($P=0.002$)。また、LUADのクラスター1および2のサンプルは、クラスター5のサンプルと比較して予後が悪かった ($P=0.034, P=0.024$)。これは、LIHC や LUAD においてはより多くの突然変異の蓄積が予後の悪化をもたらす可能性を示唆している。逆に、BLCAのクラスター3のサンプルは、クラスター1のサンプルと比較して予後が悪かった ($P=0.020$)。また、UCECのクラスター3および5のサンプルは、クラスター1のサンプルと比較して予後が悪かった ($P=0.049, P=0.045$)。これらの結果は、多数の突然変異の蓄積が、予後の悪化と必ずしも関連しないことを示している。BLCA および UCEC においては、がん発生の初期に起こる少数の突然変異を持つ細胞ががん細胞を占拠することによって予後が悪化する可能性がある。すなわち、がん発生時に生じた少数の突然変異を持つ細胞が、がんの進化過程において残存し、進化後期にサブクローナル変異の著しい増加が積極的に起こらない場合に予後が悪くなると考えられる。BLCA と UCEC の結果は、先行研究で示されたようにゲノムの不安定性がコスト (がん細胞のゲノム不安定性を維持すること) と利益 (生存競争に対して有利であること) のトレードオフであることを支持するものである [55]。

LGG のクラスター3 および 4 のサンプルでは、クラスター5 のサンプルと比較してより予後が悪かった ($P=0.023, P=0.010$)。したがって、より高い m_Peak およびより低い m_MATH を有する LGG のサンプルは予後不良であると考えられる。LGG のドライバー遺伝子の1つである *IDH1* の変異の頻度は、クラスター3, 4, 5 いずれのサンプルの中においても高かった(クラスター3, 4, 5 における *IDH1* の発生頻度はそれぞれ 69.5%, 54.5%, 59.5%) ことから、がんの早期から中期において突然変異の数を増加させるような *IDH1* の変異以外の因子が患者の予後に影響を与えている可能性がある。しかしながら、本研究ではクラスター3 と 4 のサンプルで特異的に変異が生じた遺伝子を同定することはできなかった。

LUSC および SKCM の VAF の分布が予後に及ぼす効果は、上記の 5 つのがんよりも複雑であった。LUSC のクラスター 4 のサンプルは、クラスター 1 のサンプルと比較して、予後が悪かった ($P < 0.001$)。SKCM のクラスター 3 および 4 のサンプルは、クラスター 2 のサンプルよりも予後が悪かった ($P < 0.001$, $P = 0.001$)。これらのがんでは、がんの悪性度を増大させるサブクローナルな細胞集団の増加によって予後が悪化すると推測された。

3.3.5 決定木による分類

本研究で用いた 3 つのパラメータを用いることで、ITH に基づいた予後予測を行うことができる。新規サンプルを、本研究で用いた 5 つのクラスターのいずれかに分類するための簡単な分割基準を生成するために、決定木分析 (classification and regression tree, CART) を行った。まず、複雑な決定木を構築し、複雑度パラメータ (complex parameter, cp) = 0.1 を用いて分岐を枝刈りした (図 3-9) [74]。さらに、サンプルを 5 つのクラスターのいずれか 1 つに分類するアルゴリズムの分類精度を検証するために、10-fold 交差検定を行った。交差検証による最大精度は 80.7% であり、平均精度 ($\pm SD$) は 76.8% ($\pm 2.0\%$) であった。この結果は、3 つのパラメータから計算された決定木モデルが、任意のサンプルを 5 つのクラスターのうちいずれか 1 つのクラスターに分類するためにある程度まで使用できることを示している。

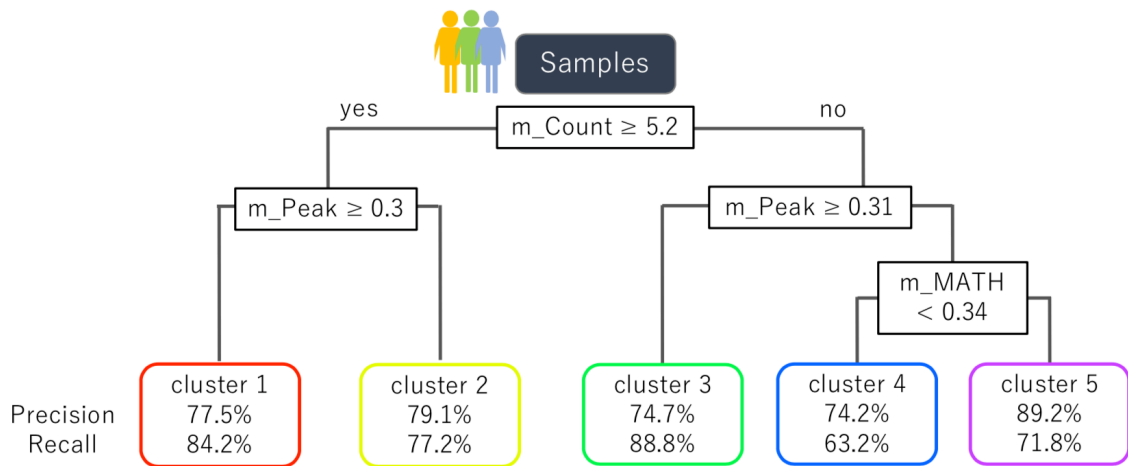


図 3-9 サンプルを VAF の分布に基づく 5 つのクラスターに分類するための決定木

MF 変異の VAF の分布に由来する 3 つのパラメータを用いて構築した決定木を示す。決定木中の数式は、サンプルの分割基準を表す。ツリーの下部にあるノードは、5 つのクラスターとそれに対応する分類の精度（適合率と再現率）を示す。

3.4 考察

がんは一般的に、ITHが進むほど予後が悪くなると考えられている[1, 6, 75, 76]. これまでの研究では、ITHを定量化するための1次元の指標を用いてITHと予後との関係を分析してきた[23, 56, 77]. これらの研究の中でAndorらは、腫瘍細胞のサブクローン数と予後との間に非線形の関係があることを示した[55]. そのため、ITHと予後との関係をより詳細に分析するためには一次元の指標ではなく、多次元的な観点から評価する必要があると考えた. 本研究では多次元的にITHを評価するために、各サンプルの持つ変異のVAFの分布の形状を用いて解析を行った. VAFの分布の形状を定義する3種類のパラメータを使用し、これらのパラメータに基づいてサンプルを5つのクラスターに分類した. クラスター数に関しては3から6まで検討を行ったが、クラスター数を5つにしたときに最も予後に差が出るがんの種類が多かったことから、本研究ではクラスター数=5を用いた. 各クラスターに属するサンプルには特徴的な遺伝的性質および臨床的性質があり、各がん種における進化経路を推測するためにこれらの特徴を使用した. さらに、各がん種について生存時間分析を行い、7種類のがんで予後と有意に関連しているクラスターがあることを見出した. この結果より、VAFの分布を用いることで、がん細胞にとって好ましい進化の状態を予測することが可能であると考えられた. また、予後と関連するVAFの分布はがんの種類によって大きく異なることを見出した. すなわち、がん細胞はその進行に伴い複数の突然変異を蓄積すると考えられているが、より多くの突然変異が必ずしもより悪い予後に関連するとは限らないことが分かった. 本研究の結果に基づき、以下3種類のがん（メラノーマ、肺腺がん、肺扁平上皮がん）の進化についての考察を行う.

SKCM（メラノーマ）については、VAFの分布に基づく分析から顕著な結果を得た. これまでの研究から、メラノーマは非常に悪性度の高いがんであり、主たる発がん原因であるUVの影響により発がん時に数多くの突然変異を有していることが示されている[61, 78-80]. メラノーマはがんの中でも進行が早く、体内の様々な組織に転移しやすい傾向があ

り、がん細胞内の変異組成の変化による薬剤耐性をもたらす[55, 81]. 本研究の結果は、多くのメラノーマサンプルが発がんに先立って蓄積された多数の突然変異を有することを示した. 興味深いことに、変異の数が少ないサンプルは、多数の突然変異を有するサンプルよりも予後が悪かった. 変異スペクトラムの結果を考慮すると、突然変異の少ないサンプルが持つ変異の多くは、がん発生の初期段階ではなく後期に起こると推測された. また、クラスター2, 4 のサンプルの割合は、低ステージのがんよりも高ステージのがんで有意に高かった. これらの結果から、次のようなメラノーマの遺伝的進化の仮説を提唱した: メラノーマは、ドライバー遺伝子である *BRAF* を含む多数の遺伝子変異によって生じ (クラスター1 および 2) [80], 進化過程の選択圧下で特定の突然変異を有する細胞が選択される. その中で、より少数の変異を有する悪性度の高い細胞が細胞集団を占有する結果、予後不良となる (クラスター3 および 4). この仮説を元にするると、変異の少ないサンプルではドライバー遺伝子の変異頻度が非常に低かったことから、ドライバー遺伝子以外のゲノム異常がメラノーマの進化に関与している可能性が考えられる.

LUAD (肺腺がん) と LUSC (肺扁平上皮がん) は肺がんの中で非小細胞肺がん (non-small cell lung cancer, NSCLC) に分類されるがんであり、これらのがんはメラノーマと同様に多数のクローナル変異を有していることが知られている [24]. 近年では、NSCLC 治療における免疫チェックポイント阻害剤に対する応答を評価するためのバイオマーカーとして、TMB が使用されている [82]. 以前の研究では、NSCLC における高 TMB が予後不良と関連していることが示されており [83], 肺腺がんではより多くの突然変異を有するサンプル (クラスター1 および 2) が予後不良であるという本研究の結果と一致していた. したがって、肺腺がんでは変異が生じるタイミングではなく変異の数が予後に大きな影響を及ぼす可能性がある. この原因の一部は喫煙習慣であり、喫煙により細胞内に徐々に変異が蓄積し、がん細胞の悪性度が上昇していくと考えられる. 肺腺がんとは逆に肺扁平上皮がんでは、選択圧下でより悪性度の高いがん細胞が細胞集団中で増大した場合 (クラスター4),

多数のクローナル変異（クラスター1）を有するサンプルよりも予後が悪くなると推測された。

本研究で用いた 3 つのパラメータをそれぞれ個別に使用した場合においても、がん患者の予後をある程度予測することは可能である。しかし、単一のパラメータではがん細胞の進化過程を同時に推測することは難しい。例えば、変異の数と予後との間に関連が検出された場合、その変異の出現タイミングを推測することは困難であると考えられる。したがって、予後とがんの進化過程との関連を明らかにするためには、がん細胞の多次元的な分析を行う必要がある。本研究の手法を用いることでこの関係をより容易に推測することができることから、この手法を応用することでがんの予後予測や治療ターゲットの探索に適用することができると考えられる。

この研究では、各サンプルの SNV を使用して 16 種類のがんを分析したが、16 種類すべてのがんにおける VAF の分布と予後との関係を見出すことはできなかった。この理由として、PRAD（前立腺がん）や THCA（甲状腺がん）は非常に予後が良いがんであるため、クラスター間で予後に差が見られなかったと考えられる。また、BRCA（乳がん）には生物学的特徴の異なる複数のサブタイプが含まれており、サブタイプによって予後に差があることが知られているため[13, 14]、各サンプルに付随するサブタイプ情報を考慮した解析が必要であったと考えられる。加えて、本研究で用いた変異は SNV のみであるが、この情報だけでは予後との関係を解析するには不十分であった可能性がある。がんの進化過程と予後への影響をより詳細に理解するためには、短い挿入・欠失や CNV などのゲノムに生じうる網羅的な変異を用いて解析する必要がある。最近の研究によると、CNV はがん遺伝子の変異アレルに対して多くの場合正の選択圧がかかっていることが示されており[84]、CNV ががんの進化に大きな影響を及ぼしていることが示唆されている。また、変異スペクトラムよりも詳細な変異パターンである変異 signature や個々の遺伝子の発現パターンを考慮することで、VAF の分布の形状に対する生物学的な意味付けができると考えられる。VAF

の分布の意味付けに加え、特定の形状の VAF の分布を特徴づける遺伝的性質（例えば突然変異遺伝子、転写調節領域における突然変異の蓄積、エピジェネティック修飾の変化等）を同定することができれば、その特徴をより正確な予後予測因子または診断マーカーとして使用することができるだろう。

4 章 総括および展望

本研究の最終的な目標は、「がんの ITH や進化過程に基づく最適な治療法の確立」である。すなわち、がん細胞の増殖に伴う変異の蓄積過程において、生物学的特徴を大きく変化させる可能性のある遺伝子やその出現タイミングを同定することが重要であると考えられる。2 章では乳がんのサンプルを用いて ITH を考慮した予後解析を行ったが、がんの種類によって予後に影響を及ぼす進化過程は異なるという知見に基づくと、他のがんについても個別の解析を行うことでがん種特異的な因子の同定が可能になるだろう。このような解析により、治療ターゲットとなる新たな遺伝子が見つかるだけでなく、進化過程を元にした新しいがんの分類ができるのではないかと考える。例えば、乳がんのサブタイプの 1 つであるトリプルネガティブ型の乳がんは、他のサブタイプの乳がんと異なりホルモン療法や抗 HER2 療法の効果が低く、治療が難しいとされる。最近の研究ではトリプルネガティブ型乳がんの中には様々なタイプの乳がんが含まれており、薬剤感受性が異なることが示唆されている [85, 86]。そのため、がんの進化過程を元にするすることで、効果的かつ副作用の少ない治療を行うためのトリプルネガティブ型乳がんの新しい分類が可能になるかもしれない。

また 3 章では、がんの進化過程を考察するにあたって VAF の分布と予後との関係を解析した。このような知見に基づいてがん患者にとって最適な治療法を探索する場合には、変異遺伝子の情報に加えて各患者がどのような治療法を受け、どのような薬を使用したかについても考慮する必要がある。先行研究では、低悪性度グリオーマ患者にテモダール（一般名：テモゾロミド）を投与した後にがん細胞内でサブクローナル変異が著しく増加したという結果が示されている [24]。一方で、変異の数が増えることで体内の免疫システムが過剰に反応し、免疫チェックポイント阻害薬の効果が上昇するという知見もある [87]。したがって、遺伝子変異ががん細胞に及ぼす効果だけでなく、治療法に対するがん細胞や体内の反応についても知見を得る必要があるだろう。

本研究の妥当性を評価するためには独立した別データを用いた確認を行う必要が

ある。一般的な ITH の解析では腫瘍の多領域サンプリングデータを用いたり、シングルセル解析由来のデータを用いたりすることで、ある変異ががん細胞においてがん発生初期に生じた変異なのか後期に局所的に生じた変異なのかを判断しがんの進化過程を推測する。このようなデータは TCGA のようなバルクデータと異なりがんの進化過程に関して明らかな知見を得ることができるが、十分なサンプルサイズを得にくいという欠点がある。したがって、多領域サンプリングデータやシングルセル解析データによって推測されたがんの進化過程を用いて、TCGA や ICGC 等に蓄積されている大量のバルクデータによる解析での妥当性を示す必要があると考える。

参考文献

1. Nowell PC. The clonal evolution of tumor cell populations. *Science*. 1976; 194: 23–28.
2. Gerlinger M, Rowan AJ, Horswell S, Larkin J, Endesfelder D, Gronroos E, Martinez P, Matthews N, Stewart A, Tarpey P, Varela I, Phillimore B, Begum S, et al. Intratumor heterogeneity and branched evolution revealed by multiregion sequencing. *N Engl J Med*. 2012; 366: 883–892.
3. Gupta RG, Somer RA. Intratumor Heterogeneity: Novel Approaches for Resolving Genomic Architecture and Clonal Evolution. *Mol Cancer Res*. 2017; 15: 1127–1137.
4. Swanton C. Intratumor heterogeneity: Evolution through space and time. *Cancer Res*. 2012; 72: 4875–4882.
5. McGranahan N, Swanton C. Biological and therapeutic impact of intratumor heterogeneity in cancer evolution. *Cancer Cell*. 2015; 27: 15–26.
6. Dexter DL, Kowalski HM, Blazar BA, Fligel Z, Vogel R, Heppner GH. Heterogeneity of tumor cells from a single mouse mammary tumor. *Cancer Res*. 1978; 38: 3174–3181.
7. Nik-Zainal S, Van Loo P, Wedge DC, Alexandrov LB, Greenman CD, Lau KW, Raine K, Jones D, Marshall J, Ramakrishna M, Shlien A, Cooke SL, Hinton J, et al. The life history of 21 breast cancers. *Cell*. 2012; 149: 994–1007.
8. Weinstein JN, Collisson E a, Mills GB, Shaw KRM, Ozenberger BA, Ellrott K, Shmulevich I, Sander C, Stuart JM. The Cancer Genome Atlas Pan-Cancer analysis project. *Nat Genet*. 2013; 45: 1113–1120.
9. Hudson TJ, Anderson W, Aretz A, Barker AD. International network of cancer genome projects. *Nature*. 2010; 464: 993–998.
10. Kikutake C, Yoshihara M, Sato T, Saito D, Suyama M. Intratumor heterogeneity of HMCN1 mutant alleles associated with poor prognosis in patients with breast cancer. *Oncotarget*. 2018; 9: 33337–33347.
11. Kikutake C, Yoshihara M, Sato T, Saito D, Suyama M. Pan-cancer analysis of intratumor heterogeneity associated with patient prognosis using multidimensional measures. *Oncotarget*. (in press)
12. Torre LA, Bray F, Siegel RL, Ferlay J, Lortet-tieulent J, Jemal A. Global Cancer Statistics, 2012. *CA a cancer J Clin*. 2015; 65: 87–108.
13. Perou CM, Sørlie T, Eisen MB, van de Rijn M, Jeffrey SS, Rees CA, Pollack JR, Ross DT, Johnsen H, Akslen LA, Fluge O, Pergamenschikov A, Williams C, et al. Molecular portraits of human breast tumours. *Nature*. 2000; 406: 747–752.

14. Sorlie T, Perou CM, Tibshirani R, Aas T, Geisler S, Johnsen H, Hastie T, Eisen MB, van de Rijn M, Jeffrey SS, Thorsen T, Quist H, Matese JC, et al. Gene expression patterns of breast carcinomas distinguish tumor subclasses with clinical implications. *Proc Natl Acad Sci U S A*. 2001; 98: 10869–10874.
15. Michor F, Polyak K. The origins and implications of intratumor heterogeneity. *Cancer Prevention Research*. 2010. 3: 1361–1364.
16. Vogelstein B, Fearon ER, Hamilton SR, Kern SE, Preisinger AC, Leppert M, Nakamura Y, White R, Smits AMM, Bos JL. Genetic Alterations during Colorectal-Tumor Development. *N Engl J Med*. 1988; 319: 525–532.
17. Burrell RA, McGranahan N, Bartek J, Swanton C. The causes and consequences of genetic heterogeneity in cancer evolution. *Nature*. 2013; 501: 338–345.
18. Suzuki Y, Ng SB, Chua C, Leow WQ, Chng J, Liu SY, Ramnarayanan K, Gan A, Ho DL, Ten R, Su Y, Lezhava A, Lai JH, et al. Multiregion ultra-deep sequencing reveals early intermixing and variable levels of intratumoral heterogeneity in colorectal cancer. *Mol Oncol*. 2017; 11: 124–139.
19. Network TCGA, Cancer T, Atlas G, Muzny DM, Bainbridge MN, Chang K, Dinh HH, Drummond JA., Fowler G, Kovar CL, Lewis LR, Morgan MB, Newsham IF, et al. Comprehensive molecular portraits of human breast tumours. *Nature*. 2012; 487: 61–70.
20. McFarland CD, Korolev KS, Kryukov G V, Sunyaev SR, Mirny LA. Impact of deleterious passenger mutations on cancer progression. *Proc Natl Acad Sci U S A*. 2013; 110: 2910–2915.
21. Aran D, Sirota M, Butte AJ. Systematic pan-cancer analysis of tumour purity. *Nat Commun*. 2015; 6: 8971.
22. Miller CA, White BS, Dees ND, Griffith M, Welch JS, Griffith OL, Vij R, Tomasson MH, Graubert TA, Walter MJ, Ellis MJ, Schierding W, DiPersio JF, et al. SciClone: Inferring Clonal Architecture and Tracking the Spatial and Temporal Patterns of Tumor Evolution. *PLoS Comput Biol*. 2014; 10: e1003665.
23. Mroz EA, Rocco JW. MATH, a novel measure of intratumor genetic heterogeneity, is high in poor-outcome classes of head and neck squamous cell carcinoma. *Oral Oncol*. 2013; 49: 211–215.
24. McGranahan N, Swanton C. Clonal Heterogeneity and Tumor Evolution: Past, Present, and the Future. *Cell*. 2017. p. 613–628.
25. Benjamini Y, Hochberg Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J R Stat Soc Ser B*. 1995; 57: 289–300.
26. Guimbellot JS, Erickson SW, Mehta T, Wen H, Page GP, Sorscher EJ, Hong JS.

- Correlation of microRNA levels during hypoxia with predicted target mRNAs through genome-wide microarray analysis. *BMC Med Genomics*. 2009; 2: 15.
27. Wykoff CC, Beasley NJP, Watson PH, Turner KJ, Pastorek J, Sibtain A, Wilson GD, Turley H, Talks KL, Maxwell PH, Pugh CW, Ratcliffe PJ, Harris AL. Hypoxia-inducible expression of tumor-associated carbonic anhydrases. *Cancer Res*. 2000; 60: 7075–7083.
 28. Handa T, Katayama A, Yokobori T, Yamane A, Horiguchi J, Kawabata-Iwakawa R, Rokudai S, Bao P, Gombodorj N, Altan B, Kaira K, Asao T, Kuwano H, et al. Caspase14 expression is associated with triple negative phenotypes and cancer stem cell marker expression in breast cancer patients. *J Surg Oncol*. 2017; 116: 706–715.
 29. Jiang Y, Qiu Y, Minn AJ, Zhang NR. Assessing intratumor heterogeneity and tracking longitudinal and spatial clonal evolutionary history by next-generation sequencing. *Proc Natl Acad Sci U S A*. 2016; 113: E5528–5537.
 30. Rajput A, Bocklage T, Greenbaum A, Lee JH, Ness SA. Mutant-Allele Tumor Heterogeneity Scores Correlate With Risk of Metastases in Colon Cancer. *Clin Colorectal Cancer*. 2017; 16: e165–170.
 31. Raynaud F, Mina M, Tavernari D, Ciriello G. Pan-cancer inference of intra-tumor heterogeneity reveals associations with different forms of genomic instability. *PLoS Genet*. 2018; 14: e1007669.
 32. Adzhubei IA, Schmidt S, Peshkin L, Ramensky VE, Gerasimova A, Bork P, Kondrashov AS, Sunyaev SR. A method and server for predicting damaging missense mutations. *Nat Methods*. 2010; 7: 248–249.
 33. Vogel BE, Muriel JM, Dong C, Xu X. Hemicentins: what have we learned from worms? *Cell Res*. 2006; 16: 872–878.
 34. Schultz DW, Klein ML, Humpert AJ, Luzier CW, Persun V, Schain M, Mahan A, Runckel C, Cassera M, Vittal V, Doyle TM, Martin TM, Weleber RG, et al. Analysis of the ARMD1 locus: Evidence that a mutation in HEMICENTIN-1 is associated with age-related macular degeneration in a large family. *Hum Mol Genet*. 2003; 12: 3315–3323.
 35. Li M, Liu F, Zhang Y, Wu X, Wu W, Wang XA, Zhao S, Liu S, Liang H, Zhang F, Ma Q, Xiang S, Li H, et al. Whole-genome sequencing reveals the mutational landscape of metastatic small-cell gallbladder neuroendocrine carcinoma (GB-SCNEC). *Cancer Lett*. 2017; 391: 20–27.
 36. Ledgerwood LG, Kumar D, Eterovic AK, Wick J, Chen K, Zhao H, Tazi L, Manna P, Kerley S, Joshi R, Wang L, Chiose SI, Garnett JD, et al. The degree of intratumor mutational heterogeneity varies by primary tumor sub-site. *Oncotarget*. 2016; 7:

- 27185–27198.
37. Sisto M, D'Amore M, Lofrumento DD, Scagliusi P, D'Amore S, Mitolo V, Lisi S. Fibulin-6 expression and anoikis in human salivary gland epithelial cells: Implications in Sjogren's syndrome. *Int Immunol.* 2009; 21: 303–311.
 38. Chowdhury A, Herzog C, Hasselbach L, Khouzani HL, Zhang J, Hammerschmidt M, Rudat C, Kispert A, Gaestel M, Menon MB, Tudorache I, Hilfiker-Kleiner D, Mühlfeld C, et al. Expression of fibulin-6 in failing hearts and its role for cardiac fibroblast migration. *Cardiovasc Res.* 2014; 103: 509–520.
 39. Thippeswamy H, Paul P, Purushottam M, Philip M, Jain S, Chandra PS. Estrogen pathway related genes and their association with risk of postpartum psychosis: A case control study. *Asian J Psychiatr.* 2017; 26: 82–85.
 40. Vogel BE, Hedgecock EM, Albertson DG, Thomson JN, Carpten JD, Makalowska I, Robbins CM, Scott N, Sood R, Connors TD, Bonner TI, Smith JR, Faruque MU, et al. Hemicentin, a conserved extracellular member of the immunoglobulin superfamily, organizes epithelial and other cell attachments into oriented line-shaped junctions. *Development.* 2001; 128: 883–894.
 41. Timpl R, Sasaki T, Kostka G, Chu ML. Fibulins: A versatile family of extracellular matrix proteins. *Nat Rev Mol Cell Biol.* 2003. 479–489.
 42. Yue W, Sun Q, Landreneau R, Wu C, Siegfried JM, Yu J, Zhang L. Fibulin-5 suppresses lung cancer invasion by inhibiting matrix metalloproteinase-7 expression. *Cancer Res.* 2009; 69: 6339–6346.
 43. Valastyan S, Weinberg RA. Tumor metastasis: Molecular insights and evolving paradigms. *Cell.* 2011. p. 275–292.
 44. Yates LR, Knappskog S, Wedge D, Farmery JHR, Gonzalez S, Martincorena I, Alexandrov LB, Van Loo P, Haugland HK, Lilleng PK, Gundem G, Gerstung M, Pappaemmanuil E, et al. Genomic Evolution of Breast Cancer Metastasis and Relapse. *Cancer Cell.* 2017; 32: 169-184.e7.
 45. Zack TI, Schumacher SE, Carter SL, Cherniack AD, Saksena G, Tabak B, Lawrence MS, Zhang C-Z, Wala J, Mermel CH, Sougnez C, Gabriel SB, Hernandez B, et al. Pan-cancer patterns of somatic copy number alteration. *Nat Genet.* 2013; 45: 1134–1140.
 46. Shlien A, Malkin D. Copy number variations and cancer. *Genome Med.* 2009; 1: 62.
 47. Khurana E, Fu Y, Chakravarty D, Demichelis F, Rubin MA, Gerstein M. Role of non-coding sequence variants in cancer. *Nat Rev Genet.* 2016; 17: 93–108.
 48. Scacheri CA, Scacheri PC. Mutations in the non-coding genome. *Curr Opin Pediatr.* 2016; 27: 659–664.

49. Jones PA, Baylin SB. The fundamental role of epigenetic events in cancer. *Nat Rev Genet.* 2002; 3: 415–428.
50. Feingold E, Good P, Guyer M, Kamholz S, Liefer L, Wetterstrand K, Collins F, Gingeras T, Kampa D, Sekinger E, Cheng J, Hirsch H, Ghosh S, et al. The ENCODE (ENCyclopedia Of DNA Elements) Project. *Science.* 2004; 306: 636–640.
51. Kuipers J, Jahn K, Beerenwinkel N. Advances in understanding tumour evolution through single-cell sequencing. *Biochim Biophys Acta.* 2017; 1867: 127–138.
52. Hajirasouliha I, Mahmoody A, Raphael BJ. A combinatorial approach for analyzing intra-tumor heterogeneity from high-throughput sequencing data. *Bioinformatics.* 2014; 30: i78-86.
53. Williams MJ, Werner B, Barnes CP, Graham TA, Sottoriva A. Identification of neutral tumor evolution across cancer types. *Nat Genet.* 2016; 48: 238–244.
54. Andor N, Harness JV., Müller S, Mewes HW, Petritsch C. Expands: Expanding ploidy and allele frequency on nested subpopulations. *Bioinformatics.* 2014; 30: 50–60.
55. Andor N, Graham TA, Jansen M, Xia LC, Aktipis CA, Petritsch C, et al. Pan-cancer analysis of the extent and consequences of intratumor heterogeneity. *Nat Med.* 2015; 22: 105–113.
56. Shen S, Wei Y, Zhang R, Du M, Duan W, Yang S, Zhao Y, Christiani DC, Chen F. Mutant-allele fraction heterogeneity is associated with non-small cell lung cancer patient survival. *Oncol Lett. Spandidos Publications;* 2018; 15: 795–802.
57. Ostrow SL, Barshir R, DeGregori J, Yeger-Lotem E, Hershberg R. Cancer Evolution Is Associated with Pervasive Positive Selection on Globally Expressed Genes. *PLoS Genet.* 2014; 10: e1004239.
58. Laddha S V., Ganesan S, Chan CS, White E. Mutational Landscape of the Essential Autophagy Gene BECN1 in Human Cancers. *Mol Cancer Res.* 2014; 12: 485–490.
59. Greaves M, Maley CC. Clonal evolution in cancer. *Nature.* 2012; 481: 306–313.
60. Davis A, Gao R, Navin N. Tumor evolution: Linear, branching, neutral or punctuated? *Biochim Biophys Acta.* 2017; 1867: 151–161.
61. Lawrence MS, Stojanov P, Polak P, Kryukov G V, Cibulskis K, Sivachenko A, Carter SL, Stewart C, Mermel CH, Roberts SA, Kiezun A, Hammerman PS, McKenna A, et al. Mutational heterogeneity in cancer and the search for new cancer-associated genes. *Nature.* 2013; 499: 214–218.
62. Sun R, Hu Z, Sottoriva A, Graham TA, Harpak A, Ma Z, Fischer JM, Shibata D, Curtis C. Between-region genetic divergence reflects the mode and tempo of tumor evolution. *Nat Genet.* 2017; 49: 1015–1024.

63. Williams MJ, Werner B, Heide T, Curtis C, Barnes CP, Sottoriva A, Graham TA. Quantification of subclonal selection in cancer from bulk sequencing data. *Nat Genet.* 2018; 50: 895–903.
64. Park Y, Lim S, Nam J-W, Kim S. Measuring intratumor heterogeneity by network entropy using RNA-seq data. *Sci Rep.* 2016; 6: 37767.
65. Gundem G, Van Loo P, Kremeyer B, Alexandrov LB, Tubio JMC, Papaemmanuil E, Brewer DS, Kallio HML, Högnäs G, Annala M, Kivinummi K, Goody V, Latimer C, et al. The evolutionary history of lethal metastatic prostate cancer. *Nature.* 2015; 520: 353–357.
66. Alexandrov LB, Jones PH, Wedge DC, Sale JE, Campbell PJ, Nik-Zainal S, Stratton MR. Clock-like mutational processes in human somatic cells. *Nat Genet.* 2015; 47: 1402–1407.
67. Drobetsky EA, Turcotte J, Châteauneuf A. A role for ultraviolet A in solar mutagenesis. *Proc Natl Acad Sci U S A.* 1995; 92: 2350–2354.
68. Harbst K, Lauss M, Cirenajwis H, Isaksson K, Rosengren F, Törngren T, Kvist A, Johansson MC, Vallon-Christersson J, Baldetorp B, Borg Å, Olsson H, Ingvar C, et al. Multiregion whole-exome sequencing uncovers the genetic evolution and mutational heterogeneity of early-stage metastatic melanoma. *Cancer Res.* 2016; 16: 4765–4774.
69. Yap TA, Gerlinger M, Futreal PA, Pusztai L, Swanton C. Intratumor heterogeneity: Seeing the wood for the trees. *Sci Transl Med.* 2012; 4: 127ps10.
70. de Bruin EC, McGranahan N, Mitter R, Salm M, Wedge DC, Yates L, Jamal-Hanjani M, Shafi S, Murugaesu N, Rowan AJ, Grönroos E, Muhammad MA, Horswell S, et al. Spatial and temporal diversity in genomic instability processes defines lung cancer evolution. *Science.* 2014; 346: 251–256.
71. Govindan R, Ding L, Griffith M, Subramanian J, Dees ND, Kanchi KL, Maher CA, Fulton R, Fulton L, Wallis J, Chen K, Walker J, McDonald S, et al. Genomic landscape of non-small cell lung cancer in smokers and never-smokers. *Cell.* 2012; 150: 1121–1134.
72. Vijg J. Somatic mutations, genome mosaicism, cancer and aging. *Curr. Opin. Genet. Dev.* 2014; 26: 141–149.
73. Wang Y, Waters J, Leung ML, Unruh A, Roh W, Shi X, Chen K, Scheet P, Vattathil S, Liang H, Multani A, Zhang H, Zhao R, et al. Clonal evolution in breast cancer revealed by single nucleus genome sequencing. *Nature.* 2014; 512: 155–160.
74. Feldesman MR. Classification trees as an alternative to linear discriminant analysis. *Am J Phys Anthropol.* 2002; 119: 257–275.

75. Fidler IJ, Kripke ML. Metastasis results from preexisting variant cells within a malignant tumor. *Science*. 1977; 197: 893–895.
76. Heppner GH. Tumor heterogeneity. *Cancer Res*. 1984; 44: 2259–2265.
77. Morris LGT, Riaz N, Desrichard A, Şenbabaoğlu Y, Hakimi AA, Makarov V, Reis-Filho JS, Chan TA. Pan-cancer analysis of intratumor heterogeneity as a prognostic determinant of survival. *Oncotarget*. 2016;7:10051–10063.
78. Tsao H, Chin L, Garraway LA, Fisher DE. Melanoma: From mutations to medicine. *Genes Dev*. 2012; 26: 1131–1155.
79. Akbani R, Akdemir KC, Aksoy BA, Albert M, Ally A, Amin SB, Arachchi H, Arora A, Auman JT, Ayala B, Baboud J, Balasundaram M, Balu S, et al. Genomic Classification of Cutaneous Melanoma. *Cell*. 2015; 161: 1681–1696.
80. Hodis E, Watson IR, Kryukov GV, Arold ST, Imielinski M, Theurillat JP, Nickerson E, Auclair D, Li L, Place C, Dicara D, Ramos AH, Lawrence MS, et al. A landscape of driver mutations in melanoma. *Cell*. 2012; 150: 251–263.
81. Grzywa TM, Paskal W, Włodarski PK. Intratumor and Intertumor Heterogeneity in Melanoma. *Transl. Oncol*. 2017; 10: 956–975.
82. Spigel DR, Schrock AB, Fabrizio D, Frampton GM, Sun J, He J, Gowen K, Johnson ML, Bauer TM, Kalemkerian GP, Raez LE, Ou S-HI, Ross JS, et al. Total mutation burden (TMB) in lung cancer (LC) and relationship with response to PD-1/PD-L1 targeted therapies. *J Clin Oncol*. 2016; 34: 9017–9017.
83. Owada-Ozaki Y, Muto S, Takagi H, Inoue T, Watanabe Y, Fukuhara M, Yamaura T, Okabe N, Matsumura Y, Hasegawa T, Ohsugi J, Hoshino M, Shio Y, et al. Prognostic Impact of Tumor Mutation Burden in Patients With Completely Resected Non–Small Cell Lung Cancer: Brief Report. *J Thorac Oncol*. 2018; 13: 1217–1221.
84. Bielski CM, Donoghue MTA, Gadiya M, Hanrahan AJ, Won HH, Chang MT, Jonsson P, Penson AV, Gorelick A, Harris C, Schram AM, Syed A, Zehir A, et al. Widespread Selection for Oncogenic Mutant Allele Imbalance in Cancer. *Cancer cell*. 2018; 34: 852–862.e4.
85. Perou CM. Molecular Stratification of Triple-Negative Breast Cancers. *The Oncologist*. 2011; 16: 61–70.
86. Bareche Y, Venet D, Ignatiadis M, Aftimos P, Piccart M, Rothe F, Sotiriou C. Unravelling triple-negative breast cancer molecular heterogeneity using an integrative multiomic analysis. *Ann Oncol*. 2018; 29: 895–902.
87. Goodman AM, Kato S, Bazhenova L, Patel SP, Frampton GM, Miller V, Stephens PJ, Daniels GA, Kurzrock R. Tumor Mutational Burden as an Independent

Predictor of Response to Immunotherapy in Diverse Cancers. *Mol Cancer Ther.* 2017; 16: 2598–2608.

謝辞

本研究を遂行する上で、ご指導、ご鞭撻、ご援助をいただいた方々に感謝の意を表します。

本論文の作成にあたり、終始適切な助言を賜り、また丁寧に指導して下さった須山幹太先生に心より感謝いたします。

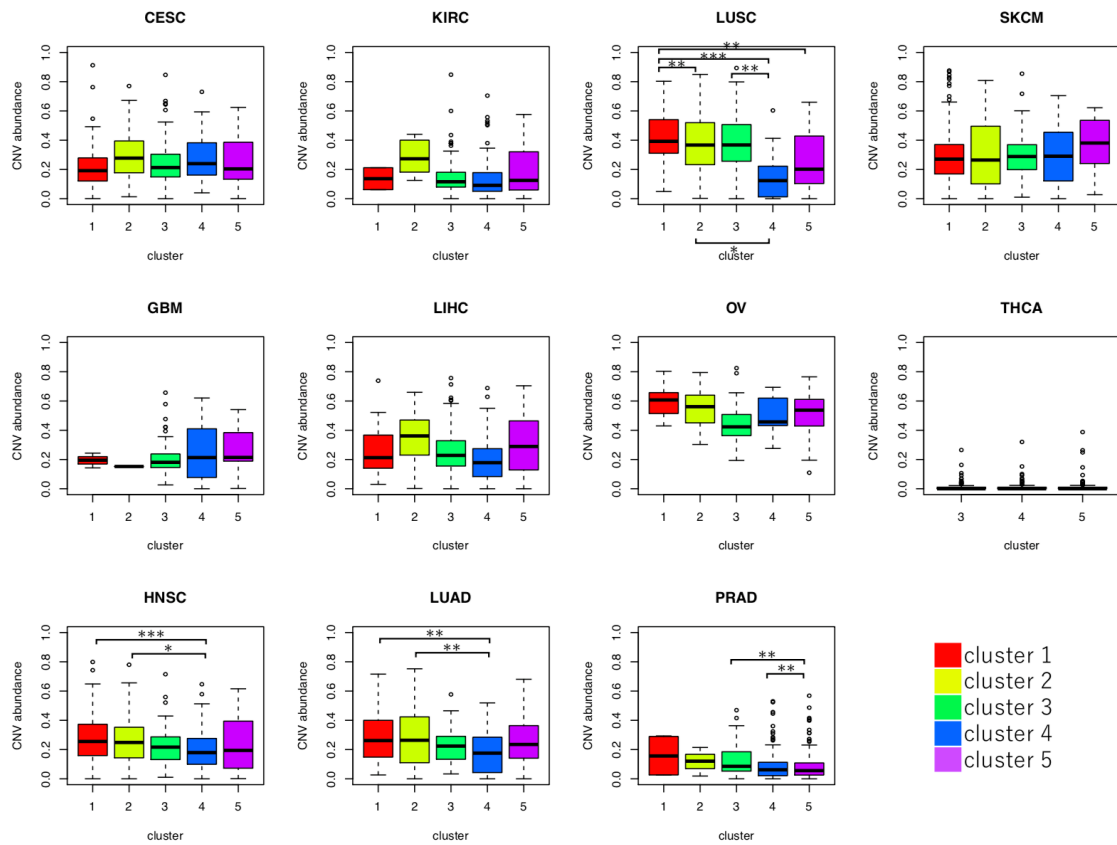
データ分析方法、研究の取り組み方など多方面にわたって指導いただいた佐藤哲也先生に感謝いたします。

研究の進め方や解析方法など適切なアドバイスを頂いた齋藤大助先生に感謝申し上げます。

吉原美奈子先生にはがんについて臨床的な面からの詳細なご助言、考察をいただきました。心より感謝いたします。

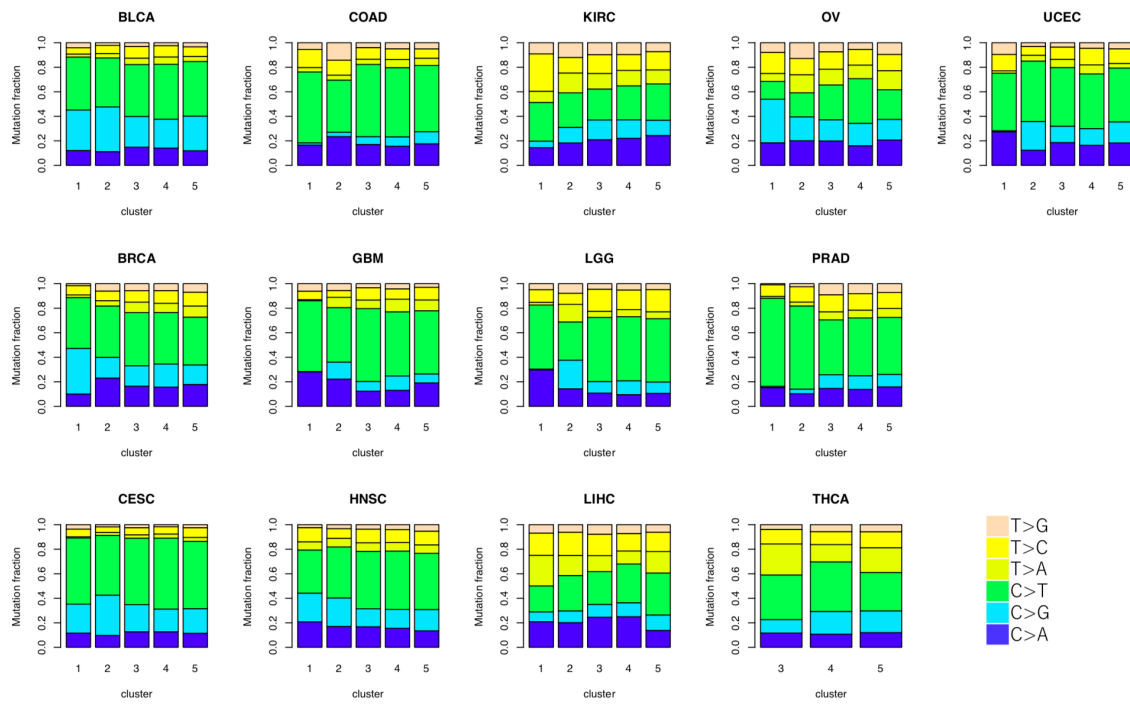
研究室のメンバーには常に刺激的な議論やアドバイスを頂き、精神的にも支えられました。ありがとうございます。

ここに記しきれない多くの方々のご支援によって本研究が成立していることを銘記し、深く感謝いたします。



補足図 3-1 VAF 分布の 5 つのクラスター間の CNV abundance の比較

各クラスターにおける CNV abundance を示す。ANOVA により CNV abundance の比較を行い、続いて Tukey 法によって多重比較を行った。 P 値に応じて*を付した：* $P < 0.05$, ** $P < 0.01$, *** $P < 0.001$ 。



補足図3-2 各クラスターの変異スペクトラムの比較

13種類のがんにおける6種類の突然変異スペクトラムの割合を示す。

補足表1 各がんの臨床データとVAF分布のパラメータ

	BLCA	BRCA	CESC	COAD	GBM	HN5C	KIRC	LGG	LHIC	LUAD	LUSC	OV	PRAD	SKCM	THCA	UCEC	
sample size (total n=6,178)	400	935	274	265	137	478	304	474	349	471	459	155	459	445	295	164	
gender																	
male (n (%))	105 (26)	924 (99)	274 (100)	122 (46)	48 (35)	127 (27)	109 (36)	213 (45)	112 (32)	250 (53)	116 (25)	155 (100)	0 (0)	167 (38)	213 (72)	164 (100)	
female (n (%))	295 (74)	11 (1)	0 (0)	143 (54)	89 (65)	351 (73)	195 (64)	261 (55)	237 (68)	221 (47)	343 (75)	0 (0)	459 (100)	278 (62)	82 (28)	0 (0)	
age (mean ± sd ^a)	68.62 ± 10.53	59.06 ± 13.05	48.66 ± 13.76	65.9 ± 13.17	61.26 ± 13.08	61.59 ± 11.84	60.62 ± 12.3	44.43 ± 13.07	59.71 ± 13.32	65.94 ± 10.03	67.83 ± 8.67	58.85 ± 11.23	61.67 ± 6.76	58.75 ± 15.71	50.8 ± 15.89	66.29 ± 11.37	
tumor stage																	
stage I	1	154	0	39	0	24	167	0	164	251	222	0	0	75	142	0	
stage II	128	541	0	102	0	66	30	0	79	110	151	0	0	135	33	0	
stage III	137	202	0	78	0	75	65	0	81	79	76	0	0	165	76	0	
stage IV	132	18	0	36	0	247	40	0	5	25	6	0	0	20	44	0	
unknown	2	20	274	10	137	66	2	474	20	6	4	155	459	50	0	164	
MF mutations																	
total number	36,906	30,380	20,951	39,774	7,867	26,676	6,018	9,927	13,939	53,256	49,181	7,331	7,839	98,939	1,685	58,884	
median number (median (range))	66 (5-1,397)	17 (2-2,213)	34 (5-4,278)	46 (15-3,594)	18 (3-4,730)	38.5 (3-1,056)	18 (3-183)	11 (2-4,301)	32 (3-542)	76 (3-769)	85 (3-1,009)	34 (2-328)	10 (2-2,432)	118 (2-5,292)	5 (2-31)	32 (3-5,086)	
median of VAF peaks (median (range))	0.28 (0.07-0.63)	0.28 (0.08-1.05)	0.33 (0.11-0.69)	0.33 (0.04-0.53)	0.45 (0.09-0.91)	0.26 (0.07-1)	0.38 (0.08-1.28)	0.34 (0.05-0.63)	0.37 (0.07-0.67)	0.26 (0.08-0.87)	0.27 (0.09-1.18)	0.22 (0.06-0.95)	0.21 (0.05-0.56)	0.36 (0.13-0.98)	0.34 (0.04-0.62)	0.37 (0.06-0.89)	
median of MATH scores (median (range))	0.28 (0.04-0.58)	0.26 (0-0.83)	0.23 (0.01-0.6)	0.22 (0.07-0.6)	0.13 (0.02-0.64)	0.23 (0.06-0.73)	0.17 (0.02-0.72)	0.26 (0.01-0.76)	0.18 (0.03-0.78)	0.29 (0-0.66)	0.28 (0.01-0.66)	0.37 (0.01-0.74)	0.25 (0.01-0.72)	0.19 (0.01-0.65)	0.24 (0-0.87)	0.22 (0-0.67)	

a. sd means standard deviation

補足表2 各がんの5クラスターの臨床的特徴と遺伝的特徴

Cancer type	Cluster					P-value ^a
	1	2	3	4	5	
BLCA (n = 400)	95	189	58	35	23	
gender						
male (%)	74 (25)	143 (48)	44 (15)	16 (5)	18 (6)	
female (%)	21 (20)	46 (44)	14 (13)	19 (18)	5 (5)	0.078
age (mean ± sd)	68.56 ± 10.16	70.12 ± 10.38	64.45 ± 10.95	67.9 ± 9.9	68.24 ± 11.03	0.041**
tumor stage						
stage I/II	32	54	26	13	4	
stage III/IV	63	134	31	22	19	0.194
unknown	0	1	1	0	0	
CNV abundance (mean ± sd)	0.281 ± 0.165	0.346 ± 0.176	0.174 ± 0.167	0.286 ± 0.188	0.179 ± 0.162	0.001**
BRCA (n = 935)	54	115	268	304	194	
gender						
male (%)	0 (0)	2 (18)	4 (36)	3 (27)	2 (18)	
female (%)	54 (6)	113 (12)	264 (29)	301 (33)	192 (21)	-
age (mean ± sd)	62.06 ± 13.18	60.24 ± 12.67	60.39 ± 13.11	57.88 ± 12.78	57.48 ± 13.29	0.061
tumor stage						
stage I/II	42	84	188	233	148	
stage III/IV	11	26	71	69	43	0.682
unknown	1	5	9	2	3	
CNV abundance (mean ± sd)	0.357 ± 0.205	0.361 ± 0.207	0.271 ± 0.173	0.252 ± 0.176	0.285 ± 0.191	<0.001***
CESC (n = 274)	61	64	70	59	20	
gender						
male (%)	0 (-)	0 (-)	0 (-)	0 (-)	0 (-)	
female (%)	61 (22)	64 (23)	70 (26)	59 (22)	20 (7)	-
age (mean ± sd)	48.58 ± 13.25	53.1 ± 13.75	46.54 ± 13.89	47.23 ± 13.03	46.26 ± 14.89	0.084
tumor stage						
stage I/II	0	0	0	0	0	
stage III/IV	0	0	0	0	0	-
unknown	61	64	70	59	20	
CNV abundance (mean ± sd)	0.217 ± 0.160	0.296 ± 0.164	0.256 ± 0.167	0.288 ± 0.159	0.260 ± 0.186	0.190
COAD (c = 265)	88	76	56	19	26	
gender						
male (%)	47 (33)	39 (27)	34 (24)	11 (8)	12 (8)	
female (%)	41 (34)	37 (30)	22 (18)	8 (7)	14 (11)	0.740
age (mean ± sd)	67.03 ± 13.19	66.57 ± 13.18	66.86 ± 12.47	61.82 ± 14.11	61.11 ± 13.28	0.196
tumor stage						
stage I/II	57	35	33	8	8	
stage III/IV	28	37	23	11	15	0.100
unknown	3	4	0	0	3	
CNV abundance (mean ± sd)	0.144 ± 0.159	0.238 ± 0.136	0.223 ± 0.123	0.238 ± 0.154	0.271 ± 0.128	<0.001***
GBM (n = 137)	3	1	109	17	7	
gender						
male (%)	1 (1)	0 (0)	75 (84)	7 (8)	6 (7)	
female (%)	2 (4)	1 (2)	34 (71)	10 (21)	1 (2)	0.181
age (mean ± sd)	40.37 ± 14.72	69.65 ± NA	61.9 ± 12.05	59.77 ± 16.75	62.7 ± 13.97	0.110
tumor stage						
stage I/II	0	0	0	0	0	
stage III/IV	0	0	0	0	0	-
unknown	3	1	109	17	7	
CNV abundance (mean ± sd)	0.194 ± 0.050	0.152	0.202 ± 0.098	0.238 ± 0.182	0.272 ± 0.178	0.122
HNSC (n = 478)	106	158	68	106	40	
gender						

	male (%)	77 (22)	111 (32)	53 (15)	77 (22)	33 (9)	
	female (%)	29 (23)	47 (37)	15 (12)	29 (23)	7 (6)	0.648
	age (mean ± sd)	62.59 ± 10.64	63.69 ± 11.24	61.04 ± 11.14	59.14 ± 12.71	58.12 ± 14.28	0.041*
	tumor stage						
	stage I/II	19	25	11	23	12	
	stage III/IV	68	115	45	70	24	0.404
	unknown	19	18	12	13	4	
	CNV abundance (mean ± sd)	0.281 ± 0.166	0.254 ± 0.155	0.229 ± 0.122	0.195 ± 0.141	0.241 ± 0.187	0.001**
KIRC (n = 304)		2	6	194	79	23	
	gender						
	male (%)	0 (0)	4 (2)	129 (66)	48 (25)	14 (7)	
	female (%)	2 (2)	2 (2)	65 (60)	31 (28)	9 (8)	621.000
	age (mean ± sd)	65.87 ± 17.14	73.86 ± 10.07	60.52 ± 12.12	59.61 ± 12.39	61.01 ± 12.6	0.117
	tumor stage						
	stage I/II	1	3	129	52	12	
	stage III/IV	1	3	63	27	11	0.559
	unknown	0	0	2	0	0	
	CNV abundance (mean ± sd)	0.137 ± 0.106	0.282 ± 0.123	0.140 ± 0.100	0.145 ± 0.147	0.190 ± 0.176	0.513
LGG (n = 474)		2	3	194	139	136	
	gender						
	male (%)	1 (0)	1 (0)	101 (39)	81 (31)	77 (30)	
	female (%)	1 (0)	2 (1)	93 (44)	58 (27)	59 (28)	0.740
	age (mean ± sd)	41.55 ± 9.03	63.15 ± 8.82	45.61 ± 13.04	44.28 ± 13.71	42.52 ± 12.14	0.075
	tumor stage						
	stage I/II	0	0	0	0	0	
	stage III/IV	0	0	0	0	0	-
	unknown	2	3	194	139	136	
	CNV abundance (mean ± sd)	0.180 ± 0.164	0.269 ± 0.177	0.118 ± 0.080	0.120 ± 0.105	0.087 ± 0.070	0.001**
LIHC (n = 349)		53	51	166	55	24	
	gender						
	male (%)	44 (19)	31 (13)	109 (46)	37 (16)	16 (7)	
	female (%)	9 (8)	20 (18)	57 (51)	18 (16)	8 (7)	0.250
	age (mean ± sd)	64.06 ± 10.44	62.14 ± 10.68	59.23 ± 12.97	57.69 ± 15.33	52.91 ± 17.66	0.041*
	tumor stage						
	stage I/II	37	37	114	34	21	
	stage III/IV	15	11	37	20	3	0.316
	unknown	1	3	15	1	0	
	CNV abundance (mean ± sd)	0.249 ± 0.151	0.349 ± 0.155	0.251 ± 0.136	0.196 ± 0.164	0.301 ± 0.199	0.152
LUAD (n = 471)		121	217	30	54	49	
	gender						
	male (%)	69 (31)	95 (43)	14 (6)	26 (12)	17 (8)	
	female (%)	52 (21)	122 (49)	16 (6)	28 (11)	32 (13)	0.241
	age (mean ± sd)	64.98 ± 10.00	65.35 ± 10.09	72.25 ± 9.33	66.57 ± 9.69	66.39 ± 9.54	0.041*
	tumor stage						
	stage I/II	90	169	23	46	33	
	stage III/IV	27	47	6	8	16	0.404
	unknown	4	1	1	0	0	
	CNV abundance (mean ± sd)	0.287 ± 0.162	0.277 ± 0.187	0.221 ± 0.121	0.182 ± 0.149	0.257 ± 0.170	0.006**
LUSC (n = 459)		163	253	19	11	13	
	gender						
	male (%)	114 (33)	199 (58)	15 (4)	8 (2)	7 (2)	
	female (%)	49 (42)	54 (47)	4 (3)	3 (3)	6 (5)	0.250
	age (mean ± sd)	67.13 ± 8.2	68.46 ± 8.66	64.55 ± 12.38	72.76 ± 4.25	64.76 ± 8.94	0.075
	tumor stage						
	stage I/II	123	209	18	11	12	
	stage III/IV	38	42	1	0	1	0.194

	unknown	2	2	0	0	0	
	CNV abundance (mean ± sd)	0.416 ± 0.154	0.360 ± 0.193	0.384 ± 0.244	0.164 ± 0.195	0.251 ± 0.207	<0.001***
OV (n = 155)		5	42	34	21	53	
	gender						
	male (%)	0 (-)	0 (-)	0 (-)	0 (-)	0 (-)	
	female (%)	5 (3)	42 (27)	34 (22)	21 (14)	53 (34)	-
	age (mean ± sd)	61.72 ± 16.28	57.89 ± 11.92	54.82 ± 8.57	62.3 ± 12.65	60.56 ± 10.65	0.117
	tumor stage						
	stage I/II	0	0	0	0	0	
	stage III/IV	0	0	0	0	0	-
	unknown	5	42	34	21	53	
	CNV abundance (mean ± sd)	0.602 ± 0.142	0.556 ± 0.121	0.445 ± 0.147	0.501 ± 0.122	0.500 ± 0.138	0.123
PRAD (n = 459)		4	3	98	204	150	
	gender						
	male (%)	4 (1)	3 (1)	98 (21)	204 (44)	150 (33)	
	female (%)	0 (-)	0 (-)	0 (-)	0 (-)	0 (-)	-
	age (mean ± sd)	64.88 ± 9.74	67.24 ± 8.55	62.04 ± 6.75	61.35 ± 6.97	61.64 ± 6.38	0.455
	tumor stage						
	stage I/II	0	0	0	0	0	
	stage III/IV	0	0	0	0	0	-
	unknown	4	3	98	204	150	
	CNV abundance (mean ± sd)	0.158 ± 0.152	0.118 ± 0.098	0.128 ± 0.102	0.082 ± 0.091	0.088 ± 0.099	0.004**
SKCM (n = 445)		258	98	47	34	8	
	gender						
	male (%)	165 (59)	61 (22)	27 (10)	18 (6)	7 (3)	
	female (%)	93 (56)	37 (22)	20 (12)	16 (10)	1 (1)	0.621
	age (mean ± sd)	58.79 ± 16.59	58.12 ± 15.31	56.21 ± 14.69	63.85 ± 11.49	58.62 ± 9.03	0.309
	tumor stage						
	stage I/II	141	34	22	9	4	
	stage III/IV	90	52	18	21	4	0.007**
	unknown	27	12	7	4	0	
	CNV abundance (mean ± sd)	0.298 ± 0.174	0.305 ± 0.224	0.311 ± 0.169	0.301 ± 0.207	0.371 ± 0.198	0.490
THCA (n = 295)		0	0	121	81	93	
	gender						
	male (%)	0 (0)	0 (0)	34 (41)	19 (23)	29 (35)	
	female (%)	0 (0)	0 (0)	87 (41)	62 (29)	64 (30)	0.648
	age (mean ± sd)	-	-	47.29 ± 14.84	52.67 ± 15.57	52.06 ± 17.02	0.075
	tumor stage						
	stage I/II	0	0	65	48	62	
	stage III/IV	0	0	56	33	31	0.316
	unknown	0	0	0	0	0	
	CNV abundance (mean ± sd)	-	-	0.018 ± 0.057	0.012 ± 0.033	0.015 ± 0.043	0.363
UCEC (n = 164)		60	12	37	30	25	
	gender						
	male (%)	0 (-)	0 (-)	0 (-)	0 (-)	0 (-)	
	female (%)	60 (37)	12 (7)	37 (23)	30 (18)	25 (15)	-
	age (mean ± sd)	64.25 ± 12.7	70.73 ± 12.31	69.91 ± 9.73	64.2 ± 9.26	66.15 ± 11.06	0.112
	tumor stage						
	stage I/II	0	0	0	0	0	
	stage III/IV	0	0	0	0	0	-
	unknown	60	12	37	30	25	
	CNV abundance (mean ± sd)	0.058 ± 0.109	0.437 ± 0.232	0.285 ± 0.211	0.307 ± 0.211	0.426 ± 0.158	<0.001***

a. FDR adjusted P values; * $P < 0.05$, ** $P < 0.01$, and *** $P < 0.001$

補足図3 単変量Cox回帰分析の結果

Cancer	Variables	HR (95% CI)	<i>P</i> -value ^{a, b}
BLCA	cluster 1 (vs cluster 2)	0.866 (0.589-1.273)	0.464
	cluster 3 (vs cluster 2)	1.288 (0.832-1.994)	0.256
	cluster 4 (vs cluster 2)	1.274 (0.764-2.124)	0.353
	cluster 5 (vs cluster 2)	1.510 (0.840-2.715)	0.168
BRCA	cluster 1 (vs cluster 2)	0.869 (0.343-2.203)	0.767
	cluster 3 (vs cluster 2)	0.759 (0.425-1.355)	0.351
	cluster 4 (vs cluster 2)	0.964 (0.559-1.664)	0.896
	cluster 5 (vs cluster 2)	0.914 (0.505-1.656)	0.768
CESC	cluster 1 (vs cluster 2)	1.256 (0.603-2.614)	0.543
	cluster 3 (vs cluster 2)	1.707 (0.811-3.593)	0.159
	cluster 4 (vs cluster 2)	1.272 (0.588-2.751)	0.541
	cluster 5 (vs cluster 2)	1.485 (0.529-4.172)	0.453
COAD	cluster 1 (vs cluster 2)	1.054 (0.557-1.993)	0.871
	cluster 3 (vs cluster 2)	0.678 (0.318-1.448)	0.316
	cluster 4 (vs cluster 2)	1.459 (0.570-3.738)	0.431
	cluster 5 (vs cluster 2)	1.555 (0.664-3.644)	0.310
HNSC	cluster 1 (vs cluster 2)	0.943 (0.658-1.351)	0.749
	cluster 3 (vs cluster 2)	0.810 (0.522-1.256)	0.346
	cluster 4 (vs cluster 2)	0.764 (0.521-1.122)	0.170
	cluster 5 (vs cluster 2)	0.688 (0.382-1.239)	0.213
LIHC	cluster 1 (vs cluster 2)	1.706 (0.899-3.238)	0.102
	cluster 3 (vs cluster 2)	0.907 (0.526-1.564)	0.726
	cluster 4 (vs cluster 2)	1.043 (0.526-2.066)	0.905
	cluster 5 (vs cluster 2)	0.789 (0.326-1.911)	0.600
LUAD	cluster 1 (vs cluster 2)	0.929 (0.646-1.335)	0.691
	cluster 3 (vs cluster 2)	0.737 (0.382-1.425)	0.365
	cluster 4 (vs cluster 2)	0.951 (0.576-1.569)	0.844
	cluster 5 (vs cluster 2)	0.663 (0.391-1.125)	0.128
LUSC	cluster 1 (vs cluster 2)	0.889 (0.654-1.207)	0.449
	cluster 3 (vs cluster 2)	0.991 (0.461-2.129)	0.982
	cluster 4 (vs cluster 2)	4.752 (2.280-9.904)	<0.001***
	cluster 5 (vs cluster 2)	1.428 (0.663-3.075)	0.363
OV	cluster 3 (vs cluster 2)	1.338 (0.761-2.354)	0.312
	cluster 4 (vs cluster 2)	1.616 (0.798-3.274)	0.183
	cluster 5 (vs cluster 2)	1.052 (0.596-1.858)	0.861
SKCM	cluster 1 (vs cluster 2)	1.557 (1.068-2.269)	0.021*
	cluster 3 (vs cluster 2)	3.908 (2.384-6.408)	<0.001***
	cluster 4 (vs cluster 2)	2.709 (1.543-4.755)	0.001**
UCEC	cluster 1 (vs cluster 2)	0.604 (0.115-3.176)	0.552
	cluster 3 (vs cluster 2)	1.883 (0.393-9.026)	0.429
	cluster 4 (vs cluster 2)	1.812 (0.376-8.720)	0.459
	cluster 5 (vs cluster 2)	2.254 (0.467-10.882)	0.312

a. * $P < 0.05$, ** $P < 0.01$, and *** $P < 0.001$

b. Where variables did not satisfy the proportional-hazards assumption for the Cox model ($P < 0.05$), the P -values are colored red.

補足図4 多変量Cox回帰分析の結果

Cancer	Variables	HR (95% CI)	<i>P</i> -value ^{a, b}
BLCA	cluster 2 (vs cluster 1)	1.121 (0.761-1.651)	0.564
	cluster 3 (vs cluster 1)	1.827 (1.099-3.035)	0.020*
	cluster 4 (vs cluster 1)	1.592 (0.896-2.826)	0.113
	cluster 5 (vs cluster 1)	1.398 (0.740-2.643)	0.302
	age	1.033 (1.017-1.050)	<0.001***
	stage (I, II vs III, IV)	2.080 (1.431-3.024)	<0.001***
	gender (male vs female)	0.919 (0.658-1.284)	0.622
BRCA	cluster 1 (vs cluster 3)	1.872 (0.764-4.590)	0.170
	cluster 2 (vs cluster 3)	1.720 (0.929-3.185)	0.084
	cluster 4 (vs cluster 3)	1.583 (0.973-2.577)	0.065
	cluster 5 (vs cluster 3)	1.689 (0.978-2.919)	0.060
	age	1.035 (1.021-1.050)	<0.001***
	stage (I, II vs III, IV)	2.758 (1.910-3.984)	<0.001***
CESC	cluster 1 (vs cluster 2)	1.293 (0.620-2.695)	0.494
	cluster 3 (vs cluster 2)	1.878 (0.887-3.977)	0.099
	cluster 4 (vs cluster 2)	1.369 (0.629-2.981)	0.428
	cluster 5 (vs cluster 2)	1.664 (0.586-4.723)	0.339
	age	1.015 (0.997-1.035)	0.110
COAD	cluster 1 (vs cluster 3)	1.566 (0.766-3.203)	0.219
	cluster 2 (vs cluster 3)	1.221 (0.554-2.691)	0.621
	cluster 4 (vs cluster 3)	2.016 (0.73-5.57)	0.176
	cluster 5 (vs cluster 3)	2.048 (0.776-5.409)	0.148
	age	1.027 (1.004-1.05)	0.019*
	stage (I, II vs III, IV)	2.866 (1.637-5.017)	<0.001***
	gender (male vs female)	1.3 (0.756-2.237)	0.343
GBM	cluster 4 (vs cluster 3)	1.626 (0.89-2.969)	0.114
	age	1.027 (1.009-1.045)	0.004**
	gender (male vs female)	0.862 (0.568-1.309)	0.486
HNSC	cluster 1 (vs cluster 4)	1.058 (0.658-1.702)	0.815
	cluster 2 (vs cluster 4)	1.273 (0.834-1.943)	0.263
	cluster 3 (vs cluster 4)	1.159 (0.679-1.980)	0.588
	cluster 5 (vs cluster 4)	1.181 (0.618-2.257)	0.615
	age	1.016 (1.001-1.031)	0.039*
	stage (I, II vs III, IV)	1.899 (1.254-2.875)	0.002**
	gender (male vs female)	0.750 (0.537-1.045)	0.089
KIRC	cluster 3 (vs cluster 5)	1.469 (0.578-3.731)	0.419
	cluster 4 (vs cluster 5)	1.829 (0.686-4.875)	0.228
	age	1.057 (1.033-1.081)	<0.001***
	stage (I, II vs III, IV)	4.047 (2.454-6.676)	<0.001***
	gender (male vs female)	1.136 (0.696-1.856)	0.610
LGG	cluster 3 (vs cluster 5)	1.762 (1.081-2.870)	0.023*
	cluster 4 (vs cluster 5)	1.977 (1.176-3.326)	0.010*
	age	1.056 (1.040-1.072)	<0.001***
	gender (male vs female)	1.108 (0.769-1.598)	0.582

LIHC	cluster 1 (vs cluster 3)	2.359 (1.378-4.04)	0.002**
	cluster 2 (vs cluster 3)	1.262 (0.706-2.255)	0.433
	cluster 4 (vs cluster 3)	1.297 (0.721-2.335)	0.386
	cluster 5 (vs cluster 3)	1.178 (0.527-2.63)	0.690
	age	1.008 (0.993-1.024)	0.315
	stage (I, II vs III, IV)	2.563 (1.727-3.804)	<0.001***
	gender (male vs female)	0.743 (0.493-1.12)	0.156
LUAD	cluster 1 (vs cluster 5)	1.875 (1.048-3.354)	0.034*
	cluster 2 (vs cluster 5)	1.889 (1.089-3.276)	0.024*
	cluster 3 (vs cluster 5)	1.395 (0.622-3.128)	0.419
	cluster 4 (vs cluster 5)	1.994 (0.988-4.025)	0.054
	age	1.014 (0.999-1.03)	0.076
	stage (I, II vs III, IV)	2.681 (1.936-3.711)	<0.001***
	gender (male vs female)	0.92 (0.673-1.258)	0.601
LUSC	cluster 2 (vs cluster 1)	1.203 (0.882-1.641)	0.243
	cluster 3 (vs cluster 1)	1.524 (0.691-3.363)	0.297
	cluster 4 (vs cluster 1)	5.929 (2.765-12.71)	<0.001***
	cluster 5 (vs cluster 1)	1.799 (0.819-3.948)	0.143
	age	1.017 (1-1.035)	0.054
	stage (I, II vs III, IV)	1.779 (1.265-2.502)	0.001**
	gender (male vs female)	1.195 (0.857-1.667)	0.295
OV	cluster 2 (vs cluster 5)	1.057 (0.597-1.873)	0.849
	cluster 3 (vs cluster 5)	1.238 (0.715-2.142)	0.446
	cluster 4 (vs cluster 5)	1.389 (0.684-2.821)	0.364
	age	1.034 (1.013-1.055)	0.001**
SKCM	cluster 1 (vs cluster 2)	1.484 (0.984-2.239)	0.060
	cluster 3 (vs cluster 2)	4.58 (2.684-7.814)	<0.001***
	cluster 4 (vs cluster 2)	2.837 (1.518-5.302)	0.001**
	age	1.022 (1.012-1.033)	<0.001***
	stage (I, II vs III, IV)	1.619 (1.183-2.217)	0.003**
	gender (male vs female)	1.106 (0.803-1.525)	0.537
THCA	cluster 4 (vs cluster 1)	3.42 (0.807-14.498)	0.095
	cluster 5 (vs cluster 1)	0.378 (0.038-3.752)	0.406
	age	1.111 (1.039-1.188)	0.002**
	stage (I, II vs III, IV)	2.536 (0.515-12.489)	0.252
	gender (male vs female)	0.739 (0.141-3.863)	0.720
UCEC	cluster 2 (vs cluster 1)	1.47 (0.271-7.972)	0.655
	cluster 3 (vs cluster 1)	3.112 (1.005-9.641)	0.049*
	cluster 4 (vs cluster 1)	2.887 (0.942-8.848)	0.064
	cluster 5 (vs cluster 1)	3.441 (1.031-11.489)	0.045*
	age	1.008 (0.973-1.045)	0.652

a. * $P < 0.05$, ** $P < 0.01$, and *** $P < 0.001$

b. Where variables did not satisfy the proportional-hazards assumption for the Cox model ($P < 0.05$), the P -values are colored red.