

博士論文

肺腺癌細胞におけるマルチオミクス解析

鈴木 絢子

目次

I 章 序論.....	3
1-1. がんゲノム解析.....	3
1-2. 細胞のトランスクリプトームおよびエピゲノム解析	4
1-3. 本研究の概要	5
II 章 日本人肺腺癌 97 症例の全エキソーム解析	6
2-1. 緒言	6
2-2. 材料と方法.....	6
2-2-1. 97 症例の全エキソームシーケンシング	6
2-2-2. SNVs の探索	8
2-2-3. SNVs が濃縮しているタンパク質機能ドメインおよび領域の同定	9
2-2-4. 有意に変異しているパスウェイの同定.....	9
2-2-5. 予後関連因子候補の抽出.....	9
2-2-6. Indels の探索	10
2-2-7. 民族間の比較.....	10
2-3. 結果と考察.....	11
2-3-1. 日本人肺腺癌 97 症例の全エキソームシーケンス	11
2-3-2. 日本人肺腺癌における癌部特異的体細胞性 SNVs.....	12
2-3-3. 日本人肺腺癌における体細胞突然変異パターン	20
2-3-4. 民族間における肺腺癌ゲノム変異パターンの違い	23
2-4 小括	26
III 章 26 肺腺癌細胞株のマルチオミクス解析	27
3-1. 緒言	27
3-2. 材料と方法.....	28
3-2-1. 細胞株と培養方法.....	28

3-2-2.	全ゲノムシーケンシング	31
3-2-3.	RNA-Seq.....	31
3-2-4.	バイサルファイトシーケンシング	31
3-2-5.	ChIP-Seq.....	31
3-2-6.	ゲノム変異の探索.....	33
3-2-7.	遺伝子発現解析	34
3-2-8.	融合遺伝子の探索.....	35
3-2-9.	DNA メチル化解析.....	35
3-2-10.	クロマチンステータス解析	36
3-2-11.	マルチオミクス“hallmarks of cancer”解析	38
3-3.	結果と考察.....	40
3-3-1.	肺腺癌細胞株のゲノム変異	40
3-3-2.	トランスクリプトームの異常	51
3-3-3.	DNA メチル化異常.....	62
3-3-4.	クロマチンステータス異常	72
3-3-5.	癌細胞におけるマルチオミクス統合解析	89
3-4	小括	100
総括	102
参考文献	104
論文目録	113
謝辞	114

I 章 序論

1-1. がんゲノム解析

近年、数多くの研究グループによって、次世代シーケンサーを用いたがんゲノムの大規模な全ゲノム・エキソームシーケンシングが行われている。The Cancer Genome Atlas (TCGA) (1)は、卵巣癌(2)や肺扁平上皮癌(3)をはじめとした 20 種類を超える多様ながん種においてがんゲノムシーケンスの解析結果を報告している。また、International Cancer Genome Consortium (ICGC) (4)では、がんの原因を明らかにするために、50 タイプ・サブタイプのがんにおけるゲノム変異、トランスクリプトームおよびエピゲノム異常のカタログ化を世界的に進めている。日本のグループも ICGC に参加し、ウイルス性の肝臓癌ゲノムについて、日本人肝臓癌全ゲノム配列の解読を報告している(5-7)。

がん細胞には、single nucleotide variants (SNVs)、insertion/deletions (indels)、コピー数異常、および、転座や逆位といった染色体構造異常等のゲノム変異が見られる。多くのがん細胞で頻繁に検出されるような変異も存在する。Tumor protein p53 (TP53) 遺伝子は多くのがんで変異が検出されており、その多くは DNA 結合ドメインに集中している。一方で、がんの種類によって変異パターンが異なる遺伝子もあり、個々のがん種におけるゲノム解析は必須である。

本研究の解析対象である肺腺癌は、がんの主な原因となるいわゆるドライバー変異の同定が最も進んでいる癌種である。肺腺癌は epidermal growth factor receptor (EGFR) や Kirsten rat sarcoma viral oncogene homolog (KRAS) 遺伝子の変異が多く見られることが知られている(8)。特に、EGFR 変異陽性肺腺癌は、小分子 EGFR キナーゼ阻害剤 gefitinib および erlotinib に感受性を示し、非喫煙者や女性、アジア人に頻出することが示唆されている(9)。また、染色体構造異常による anaplastic lymphoma receptor tyrosine kinase (ALK)、ret proto-oncogene (RET) および ROS proto-oncogene 1, receptor tyrosine kinase (ROS1) の融合遺伝子が肺腺癌患者の数%において見られる。近年、ALK 融合遺伝子陽性の肺腺癌に対しては ALK 阻害剤 crizotinib が治療効果を示し、融合遺伝子も肺腺癌細胞において重要な因子の一つである証左となった。これらドライバー変異における集団中の割合は、患者の年齢や喫煙歴、民族によって異なることが報告されている(10)。また、ドライバー変異が不明である症例も依然として多く存在し、肺腺癌細胞のゲノム変異と、癌化に関わる分子メカニズムの解明が求められている。

肺腺癌における大規模なゲノム解析はいくつかの研究グループで報告されている。2008 年、Ding *et al.* (8)は 188 症例について 623 個のキナーゼを中心とした癌関連遺伝子を解析した。2012 年には、Imielinski ら(11)が 183 症例について全ゲノムまたはエ

キソームシーケンシングを行っており、肺腺癌における初の大規模かつ網羅的なゲノムシーケンス解析の報告を行った。Seo *et al.* (12)は、韓国人肺腺癌 200 症例を報告しており、アジア人肺腺癌の解析も進んでいる。2014 年には、230 症例の肺腺癌ゲノム解析が TCGA より報告され(13)、肺腺癌ゲノム研究の基盤となるゲノムデータの蓄積は急速に進められている。しかし、日本人肺腺癌を対象とした大規模ゲノム解析の報告は少なく、日本人肺腺癌におけるゲノム変異パターンの同定が求められている。

1-2. 細胞のトランスクリプトームおよびエピゲノム解析

がん細胞のゲノム変異には、U2 small nuclear RNA auxiliary factor 1 (U2AF1)、RNA binding motif protein 10 (RBM10) 等の RNA スプライシング関連遺伝子や、AT rich interactive domain 1A (SWI-like) (ARID1A)、SWI/SNF related, matrix associated, actin dependent regulator of chromatin, subfamily a, member 4 (SMARCA4) 等のクロマチンリモデリング関連遺伝子、E1A binding protein p300 (EP300) 等のヒストン修飾関連遺伝子など、遺伝子発現・転写制御に関わる遺伝子の変異が多く報告されている(11,13)。また、近年、プロモーターやエンハンサーといった遺伝子発現制御領域における変異も多数存在すると考えられるようになった(14-16)。メラノーマでは、telomerase reverse transcriptase (TERT) のプロモーターにおける変異が転写因子 E-twenty-six (ETS) の結合モチーフを生み出すことによって、TERT プロモーターの活性を上昇させている(16)。しかし、転写開始点近傍だけでも数多く見出される SNVs のうち、遺伝子発現制御を生または負に変化させるゲノム変異がどの SNVs であるか、またどの程度がん細胞のエピゲノムおよびトランスクリプトームに影響を及ぼすのか明らかでない。遺伝子発現制御に関与する SNVs の機能的同定には、がん細胞におけるエピゲノムおよびトランスクリプトームのシーケンス解析が必要不可欠である。

一方、近年の次世代シーケンサーの普及により、遺伝子発現量および転写産物の同定には RNA-Seq (17)、DNA メチル化解析にはバイサルファイトシーケンシング(18)、ヒストン修飾の状態や転写因子結合部位の同定には ChIP-Seq (19)等が広く用いられるようになってきている。

ヒトのトランスクリプトームおよびエピゲノムシーケンスデータは多種多様なプロジェクトで収集、解析されている。NIH Roadmap Epigenomics Mapping Consortium (20)では、数百を超える種類のヒトの幹細胞および正常組織のエピゲノムシーケンスデータを取得、公開している。Encyclopedia of DNA Elements (ENCODE) Consortium (21,22)でも、約百種類の主に培養細胞系におけるトランスクリプトームおよびエピゲノムデータを公開している。これらのプロジェクトでは、最終的に、取得さ

れたデータを用いた階層的な統合解析から細胞の遺伝子発現制御の網羅的な理解を目指している。がん細胞においても、このようなトランスクリプトームおよびエピゲノムデータを収集し用いることで、ゲノム解析だけでは不鮮明であったがん細胞のゲノム・エピゲノム変異と遺伝子発現制御異常の関連を明らかにすることができると考えられる。

1-3. 本研究の概要

本研究は、肺腺癌細胞におけるマルチオミクスシーケンス解析を行い、癌細胞のゲノムおよびエピゲノム変異と遺伝子発現制御異常を多階層オミクス解析の視点からとらえることを目的とした。II章では、日本人の肺腺癌臨床検体におけるゲノム変異を網羅的に同定するために行った日本人肺腺癌 97 症例の全エクソーム解析について述べる。臨床検体におけるタンパク質コード領域のゲノム変異を網羅的に解析したものである。これにより日本人肺腺癌固有のゲノム変異パターンの検出を試みた。III章では、26 種類の肺腺癌細胞株を用いて、全ゲノムデータに加えて、エピゲノムおよびトランスクリプトームデータを同時に取得、統合的解析を行った。細胞株を用いた解析は、臨床検体を用いた研究では困難である同一検体由来のマルチオミクスデータの取得が可能であった。本研究では、肺腺癌細胞において、個々の症例および細胞株に固有または共通の、ゲノム、エピゲノムおよびトランスクリプトームの関係性を明らかにすることを目指した。

II 章 日本人肺腺癌 97 症例の全エキソーム解析

2-1. 緒言

多くのがん研究において、特にタンパク質コード領域におけるアミノ酸変異を伴うゲノム変異は、重要ながん化の要因として報告されている。肺腺癌においても、EGFRやKRAS変異をはじめとした癌化に寄与するSNVsがいわゆる”ドライバー変異”として報告されており(13)、一部は分子標的薬のターゲットとして治療標的となっている(9)。しかし、一方で、これらドライバー変異が不明である症例も多数存在し、体細胞突然変異の網羅的な解析によるさらなる治療標的の探索が求められている。

アジア人肺腺癌では、EGFR 変異をドライバー変異に持つ症例が多いことが示唆されている(12)。人種によるゲノム変異パターンの違いが示唆されることから、日本人肺腺癌のゲノムを解析し、その特徴を見出すことは、日本人肺腺癌に対する治療標的の同定に向けて極めて重要である。

本章では、日本人肺腺癌 97 症例におけるタンパク質コード領域の変異に着目して、同一検体由来の癌部・非癌部ペアの全エキソーム解析を行った(23)。癌部で抽出されたゲノム変異から非癌部で抽出された **germline variants** を除くことで、癌部特異的な体細胞突然変異を抽出した。さらに同定した体細胞突然変異について、統計学的解析によって、タンパク質機能ドメイン、パスウェイ等のさまざまなレベルにおける癌関連因子候補の抽出を行い、日本人肺腺癌のゲノム変異パターンについて記述した。

2-2. 材料と方法

2-2-1. 97 症例の全エキソームシーケンシング

97 症例の日本人肺腺癌全エキソームシーケンシングデータは、国立がん研究センター東病院から提供を受けた。全ての検体は、適切なインフォームドコンセントのもと、国立がん研究センターの倫理審査委員会が承認したプロトコルに従って集められた。メタノール固定された 3 cm 以上の肺腺癌手術検体が選出され、癌部はレーザーマイクロダイゼクションで抽出された。選出された症例の生存期間や再発の有無、喫煙歴等の臨床情報を表 1 に示す。

癌部および同一検体由来の肺非癌部 DNA を用いて、全エキソームシーケンシングを行った。SureSelect ターゲットエンリッチメントシステム (アジレント) によって 1 μ g のゲノム DNA からエキソン部分の DNA を濃縮し、Genome Analyzer IIx (GAIIx ; イルミナ) によって 76 塩基長のペアエンドシーケンスタグを取得した。

表 1 日本人肺腺癌 97 症例の臨床情報

Number of patients	97
Patient age	
Average (years)	66.6
Median (years)	67
Gender	
Male	52
Female	45
Body Mass Index (BMI)	
Average	22.5
Median	22.6
Smoking history	
Smoker *	54
Non-smoker	43
Clinical information	
Lymph node metastasis	38
Vascular invasion	58
Lymphatic invasion	55
Pleural invasion	49
Recurrence	52
Carcinoembryonic antigen (CEA)	
Average	18.1
Median	5.3
Histological subtype	
Papillary	39
Bronchoalveolar carcinoma	28
Solid adenocarcinoma (with mucin)	17
Acinar	13

*過去に喫煙歴がある症例を含む。

(Table S10, Suzuki and Mimaki *et al.* 2013 *PLoS ONE* 8(9):e73484 doi:10.1371/journal.pone.0073484)

2-2-2. SNVs の探索

国立がん研究センター東病院から提供された全エクソームシーケンスデータを、Burrows-Wheeler Aligner (BWA) (24)を用いて、ヒト参照ゲノム配列 (UCSC hg19) にマップした。SAMtools (25)によって PCR duplicate の除去を行い、GATK Unified Genotyper (v1.0) (26,27)を用いて、SNVs を検出した。検出された SNVs (Confidence score が 50 以上) をタンパク質コード配列上にマップし、閾値を設定してエラーの除去を行った。タンパク質コード領域の情報は UCSC から提供されているファイルを利用した。変異を含むシーケンスタグについて、76 塩基のうち、変異以外の 75 塩基において参照ゲノム配列とミスマッチがなく、変異の base quality value (QV) が 30 以上を満たすタグ (パーフェクトタグ) が 2 以上、また、変異部分の QV が 20 以上のタグが 4 以上である変異を抽出した。また、既知のゲノム多型を取り除くために、NCBI dbSNP build 132 (OMIM-CURATED-RECORDS を除く) (28)と日本人ゲノム(29)を除去した。さらに、各症例の germline variants を除去するために、97 症例の非癌部エクソームと 73 人分の日本人エクソーム (表 2 ; JPT、1000 Genomes Project (30,31)) で検出された SNVs および 48 人分の in-house 日本人 germline データを除去した。この際に、より高感度に germline variants を取り除くために、97 症例非癌部および 73 JPT から抽出された germline SNVs は GATK で検出されたものをそのまま使用した。最終的にアミノ酸変異を伴う non-synonymous SNVs を新規体細胞突然変異候補とした。

表 2 使用した 1000 Genomes Project データ

73 JPT exome sequencing data*

NA18940, NA18942-NA18945, NA18947-NA18953, NA18959-NA18961,
NA18964, NA18966-NA18976, NA18981-NA18983, NA18985-NA18991,
NA18999-NA19000, NA19003-NA19004, NA19007,NA19012,
NA19054-NA19060, NA19062-NA19068, NA19070, NA19072,
NA19074-NA19085, NA19087-NA19088

*phase1 exome data, 20110521 ; JPT: Japanese in Tokyo, Japan ; イルミナ社のシーケンサーでシーケンスされているペアエンドのエクソームシーケンスデータを選択した。

2-2-3. SNVs が濃縮しているタンパク質機能ドメインおよび領域の同定

SNVs が有意に濃縮しているタンパク質機能ドメインを探索した。タンパク質機能ドメインは、CCDS (NCBI Consensus Coding Sequence) のアミノ酸配列に対して InterProScan (32)により検出されたものを用いた。CDS 長、タンパク質機能ドメイン長、CCDS 上の SNVs 数およびタンパク質機能ドメイン上の SNVs 数を用いて、SNVs の偏りを超幾何分布で評価した。CCDS 上の SNVs 数が 5 以上、タンパク質機能ドメイン上の SNVs 数が 3 以上の場合のみ評価の対象とし、P 値の算出には R phyper を用いた。さらに、SNVs が有意に濃縮しているタンパク質機能ドメイン上において、癌体細胞突然変異データベース Catalogue Of Somatic Mutations In Cancer (COSMIC ; v56) (33,34)で既知の SNVs ("Missense"、"Nonsense"、"Nonstop extension") が報告されているか調べた。

また特定の領域に偏って存在する SNVs を抽出した。SNVs を 20 アミノ酸ウィンドウごとに数え、CCDS 上にある全 SNVs 数に対して、20 アミノ酸以内にある SNVs 数が有意に偏っているかを超幾何分布で評価した。CCDS 上の SNVs 数が 5 以上、20 アミノ酸以内にある SNVs 数が 3 以上の場合のみ評価の対象とし、P 値の算出には R phyper を用いた。

2-2-4. 有意に変異しているパスウェイの同定

Kyoto Encyclopedia of Genes and Genomes (KEGG) (35)から提供されている 230 個のヒトのパスウェイデータを用いて、SNVs を持つ症例が有意に濃縮されているパスウェイを探索した。各パスウェイにおいて、変異遺伝子を少なくとも 1 遺伝子含む症例数とその期待値から、有意に多くの症例が変異しているパスウェイであるか、ポアソン分布を用いて評価した。症例あたりの変異遺伝子数 94.4 と総遺伝子数 17,175 から、変異率 M を算出した ($M = 94.4 / 17,175$)。下記のように、 n 個の遺伝子から構成されるパスウェイ t において、そのパスウェイに変異を持つ症例数の期待値 $\lambda_{t,n}$ を算出した。

$$\lambda_{t,n} = \{1 - (1-M)^n\} \times 97$$

k 症例がそのパスウェイに変異を持っていた場合、 k 症例以上が変異している確率 (P 値) を、ポアソン分布を用いて算出した。P 値の計算には R ppois を用いた。

2-2-5. 予後関連因子候補の抽出

各遺伝子および KEGG パスウェイに対して、97 症例を、SNVs を少なくとも 1 つ持つ症例と SNVs を持たない症例の 2 群に分け、全生存期間の比較を行った。Kaplan-Meier 法を用いて生存曲線を描き、log-rank test で 2 群の差を評価した。評価は両群が 3 症例以上になる遺伝子およびパスウェイを対象に行った。生存曲線の描画および統計値の算出は R survival ライブラリの survfit と survdiff を用いて行った。

2-2-6. Indels の探索

ヒト参照ゲノム配列にマップされた全エキソームシーケンスデータから、GATK Indel Genotyper V2 (v1.0) を用いて indels を検出した。CCDS 上にマップしたのちに、変異を含むタグが 4 以上 (Fwd および Rev タグがそれぞれ 1 以上) を満たす変異を抽出した。また、NCBI dbSNP build 131 を既知ゲノム多型として除去し、さらに、SNVs と同様に、97 症例の非癌部と 121 人分の日本人データから抽出された indels を germline variants として除去した。残った indels を新規体細胞突然変異候補とした。

2-2-7. 民族間の比較

日本人肺腺癌 97 症例から検出された体細胞性の SNVs および indels を、肺腺癌で変異が頻出しているパスウェイ(8)にマップした。さらに、Ding *et al.* (8)で報告されている 188 症例の欧米人肺腺癌のゲノム変異 (“Missense”、“Nonsense”、“Frame Shift Del”、“Frame Shift Ins”、“In Frame Del”、“In Frame Ins” ; Supplementary Table S2, Ding *et al.* 2008 *Nature*) もパスウェイ上にマップし、各遺伝子の変異頻度を民族間で比較した。さらに、Seo *et al.* (12)の 200 症例、Imielinski *et al.* (11)の 183 症例、および、TCGA (13)の 230 症例のデータとの比較も行った。

2-3. 結果と考察

2-3-1. 日本人肺腺癌 97 症例の全エキソームシーケンス

日本人肺腺癌 97 症例について、癌部全エキソームシーケンスデータを取得した。イルミナ GAIIX を用いてシーケンス解析を行い、レーン平均 32,997,883 タグのヒト参照ゲノム配列にマップ可能なシーケンスを取得した (図 1)。約 37 Mb のベイト領域に対して、平均 sequence depth は 74×、5×以上の depth を満たす領域が約 93%であった。非癌部シーケンスデータでも同様の結果が得られた。

	AD001 (cancer)	Avg. per lane	
		97 cancers	97 normal counterparts
Mapped sequences (pair)	34,059,742	32,997,883	33,316,775
Depth (avg.)	76.1	74.1	63.7
5× coverage	0.92	0.93	0.94

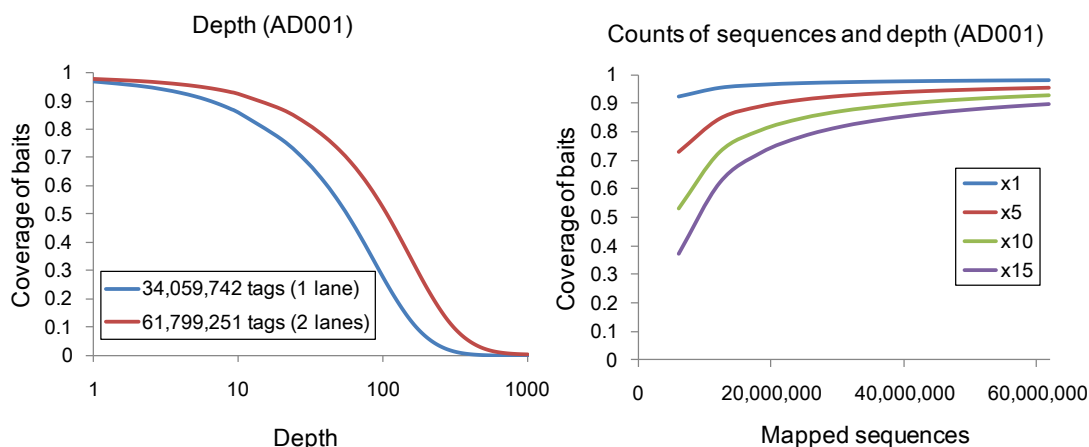


図 1 全エキソームシーケンスのシーケンスタグ数と depth および coverage

全エキソームシーケンスデータの概要を示している。上図には、症例 AD001、97 症例の癌部および非癌部の平均シーケンスタグ数、平均 depth および 5×以上の depth を満たす領域の coverage を示した。Mapping quality が 60 を満たすシーケンスタグを解析に使用した。下図では、症例 AD001 のデータを用いて、depth (左図) およびシーケンスタグ数 (右図) とベイト領域に対する coverage との関係を示した。

(Figure S1, Suzuki and Mimaki *et al.* 2013 *PLoS ONE* 8(9):e73484 doi:10.1371/journal.pone.0073484 改変)

2-3-2. 日本人肺腺癌における癌部特異的体細胞性 SNVs

得られた全エクソームシーケンススタグから、タンパク質コード領域上にある症例あたり平均 14,182 個の SNVs を検出した。そのうちアミノ酸置換を伴う non-synonymous SNVs は平均 6613 個であった。また、NCBI dbSNP や非癌部 SNVs 等の germline variants を除き、平均 102 個の体細胞性 SNVs を抽出した (図 2)。

これら SNVs の数は症例によって大きく異なっていた。特に変異数が多かった 3 症例 (平均変異数 + 2 s.d.以上) は全て男性で喫煙歴があった。肺腺癌において喫煙者の変異数は非喫煙者よりも多いという報告があり(36)、これら 3 症例の変異数は喫煙歴を反映しているのかもしれない。また、体細胞性 SNVs における塩基置換パターンを図 3 に示す。変異数が少ない症例では C>T (G>A) の割合が多く見られ、変異数の多い症例では G>T (C>A) が頻出していた。G>T 変異は喫煙との関連が報告されているため、各症例の喫煙歴と照らし合わせたところ、喫煙歴のある症例は G>T 変異の割合が多い傾向にあった。喫煙歴によって変異数および塩基置換パターンに特徴があることが示唆された。

合計 5997 遺伝子において SNVs が検出された。97 症例で最も SNVs が多く見られた遺伝子は EGFR であった (37 症例 ; 38.1%)。EGFR 変異のうち、858 番目のアミノ酸に対して lysine から arginine への置換を生じる変異 (L858R 変異) が最も頻出し、28 症例で見られた。他にも、G719A や L861Q 変異が検出された。これらの変異は EGFR チロシンキナーゼの恒常的活性化を誘起するドライバー変異であり、これらの変異を持つ肺腺癌は小分子 EGFR キナーゼ阻害剤 gefitinib および erlotinib 感受性であることが知られている(9)。また肺腺癌で頻出するといわれている KRAS の変異は、8 症例 (うち 12 番目の glycine に対する変異は 6 症例) で見られた。Phosphatidylinositol-4,5-bisphosphate 3-kinase, catalytic subunit alpha (PIK3CA) の E545K および H1047L/R 変異は 4 症例 (うち 1 症例は EGFR 変異陽性)、B-Raf proto-oncogene, serine/threonine kinase (BRAF) の G469A 変異は 1 症例で検出された。

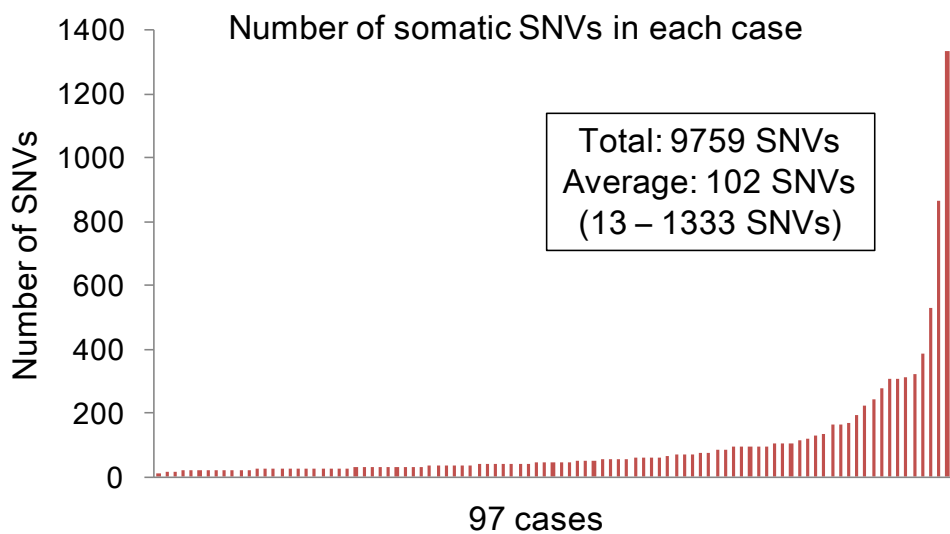


図2 各症例における体細胞性 SNVs の数

各症例における癌部特異的体細胞性 SNVs の数を示している。横軸は各症例を示しており、SNVs 数で並べられている。総計および平均値はグラフ内に記載してある。

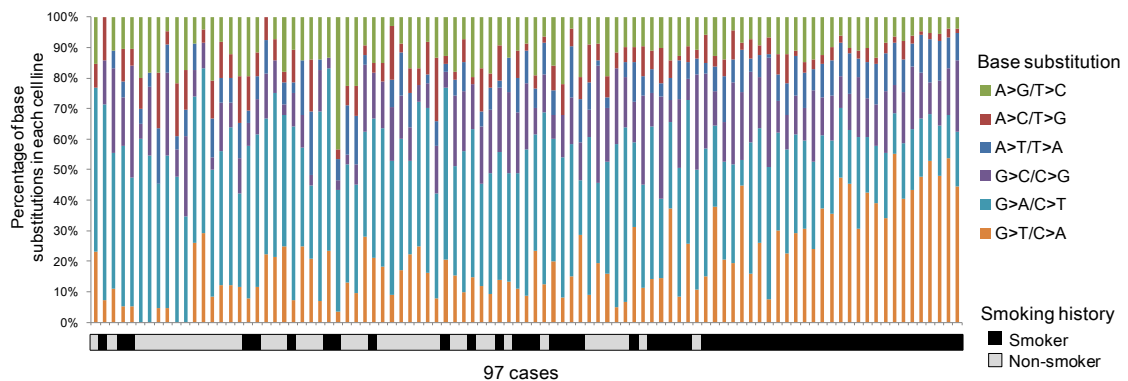


図3 体細胞性 SNVs の塩基置換パターン

各症例における体細胞性 SNVs の塩基置換パターンを示している。横軸は各症例を示しており、SNVs 数で並べられている。縦軸は各塩基置換パターンの割合を示している。各症例の喫煙歴（喫煙者：黒、非喫煙者：グレー）は横軸の下に示している。

EGFR や KRAS 等を除いて、変異遺伝子の多くは 97 症例中に低頻度でしか出現しなかった。症例間で相互重複しないゲノム変異および変異遺伝子からその変異が機能的に癌に関連する遺伝子を選択するために、SNVs が有意に濃縮されているタンパク質機能ドメインやパスウェイを統計学的に探索することで、癌関連遺伝子候補の探索を試みた。

材料と方法の章に示す統計的解析手法を用いて、SNVs が有意に濃縮されているタンパク質機能ドメインを検出したところ、11 遺伝子のタンパク質機能ドメインにおいて SNVs が統計的に有意に濃縮していた ($p < 0.02$; 表 3)。これらのタンパク質機能ドメインは、そのドメイン内に COSMIC で登録されている既知の癌関連 SNVs 候補が検出されている。また、20 アミノ酸以内に SNVs が有意に濃縮している遺伝子が 8 個見出された (表 4)。これら抽出された癌関連遺伝子の中には、EGFR、KRAS および TP53 に加えて、既に癌との関連が報告されている遺伝子が含まれていた。Tenascin N (TNN) は、97 症例中 5 個の SNVs が見出されており、うち 4 個がタンパク質機能ドメイン”IPR003961: Fibronectin, type III”上に統計的に有意に偏在していた ($p = 5.2e-5$; 図 4A)。さらに、このドメイン上には COSMIC に大腸癌の D629Y 変異が登録されており、このドメインは肺腺癌以外の癌種でも変異が検出されていた(37)。TNN は乳癌や大腸癌の腫瘍間質でその発現の異常亢進が報告されており、新規の腫瘍マーカー候補ともなっている(38-40)。同様に phosphatidylinositol-3,4,5-trisphosphate-dependent Rac exchange factor 1 (PREX1) の”IPR000219: Dbl homology (DH) domain”、および、phosphatidylinositol-3,4,5-trisphosphate-dependent Rac exchange factor 2 (PREX2) の”IPR001849: Pleckstrin homology domain”において、SNVs が統計的に有意に偏って検出された ($p = 0.00071$ および $p = 0.011$; 図 4B および 4C)。PREX1 と PREX2 は、RAC1 等の Rho ファミリーに結合し、GDP を GTP に変換することで Rho ファミリーを活性化するグアニンヌクレオチド交換因子 (RhoGEF) である(41,42)。PREX1 は乳癌における erb-b2 receptor tyrosine kinase 2 (ERBB2) パスウェイとの関連が報告されており(43)、また、PREX2 はメラノーマのゲノム解析で変異が数多く報告されている(44)。これらの遺伝子について肺腺癌との機能的相関の立証にはさらなる実験的検証が不可欠である。

個々の遺伝子についての統計解析に加え、統計的に有意に多くの症例で変異が蓄積しているパスウェイの探索を試みた。26 パスウェイが抽出された ($p < 0.05$; 表 5)。”hsa05213 Endometrial cancer” ($p = 3.1e-15$) や”hsa05223 Non-small cell lung cancer” ($p = 7.1e-12$) といった癌関連パスウェイが多く検出された。多くの EGFR 変異陽性が含まれることを反映し、EGFR を含むパスウェイが多数抽出された。一方で、EGFR、KRAS および TP53 といった既知癌関連遺伝子を含まない”hsa04977 Vitamin digestion and absorption”や”hsa02010 ABC transporters”といったパスウェイも抽出することができた。遺伝子レベルの解析では抽出できなかった癌関連因子候補を抽出することができたと考えている。

表 3 SNVs が濃縮しているタンパク質機能ドメイン

Gene	Domain	Number of SNVs		P-value*
		Domain	Gene	
EGFR	IPR001245: Serine-threonine/tyrosine-protein kinase	34	37	4.4e-21
KRAS	IPR001806: Ras GTPase	6	7	8.0e-6
TNN	IPR003961: Fibronectin, type III	4	5	5.2e-5
TP53	IPR008967: p53-like transcription factor, DNA-binding	20	23	9.5e-5
PREX1	IPR000219: Dbl homology (DH) domain	4	5	0.00071
DNAH7	IPR004273: Dynein heavy chain	5	7	0.0025
FSTL5	IPR011044: Quinoprotein amine dehydrogenase, beta chain-like	7	7	0.0043
NRXN3	IPR008985: Concanavalin A-like lectin/glucanase	5	7	0.0063
PREX2	IPR001849: Pleckstrin homology	3	7	0.011
FER1L6	IPR008973: C2 calcium/lipid-binding domain, CaLB	3	6	0.013
COL22A1	IPR008985: Concanavalin A-like lectin/glucanase	3	6	0.015

* $p < 0.02$ 、超幾何分布を用いて算出した。

(Table 1, Suzuki and Mimaki *et al.* 2013 *PLoS ONE* 8(9):e73484 doi:10.1371/journal.pone.0073484)

表 4 SNVs が濃縮している領域 (20 アミノ酸)

Gene	Window start	Window end	Number of SNVs		P-value*
			Window	Gene	
EGFR	839	858	28	37	<1.0e-11
CDC27	474	493	7	9	5.1e-11
NUP210L	740	759	4	5	4.6e-8
KRAS	1	20	6	7	4.2e-6
LRRC66	540	559	3	5	9.8e-5
PCDHB16	21	40	3	6	0.00028
TP53	218	237	6	23	0.00046
FRG1	111	130	4	7	0.00079

*超幾何分布を用いて算出した。

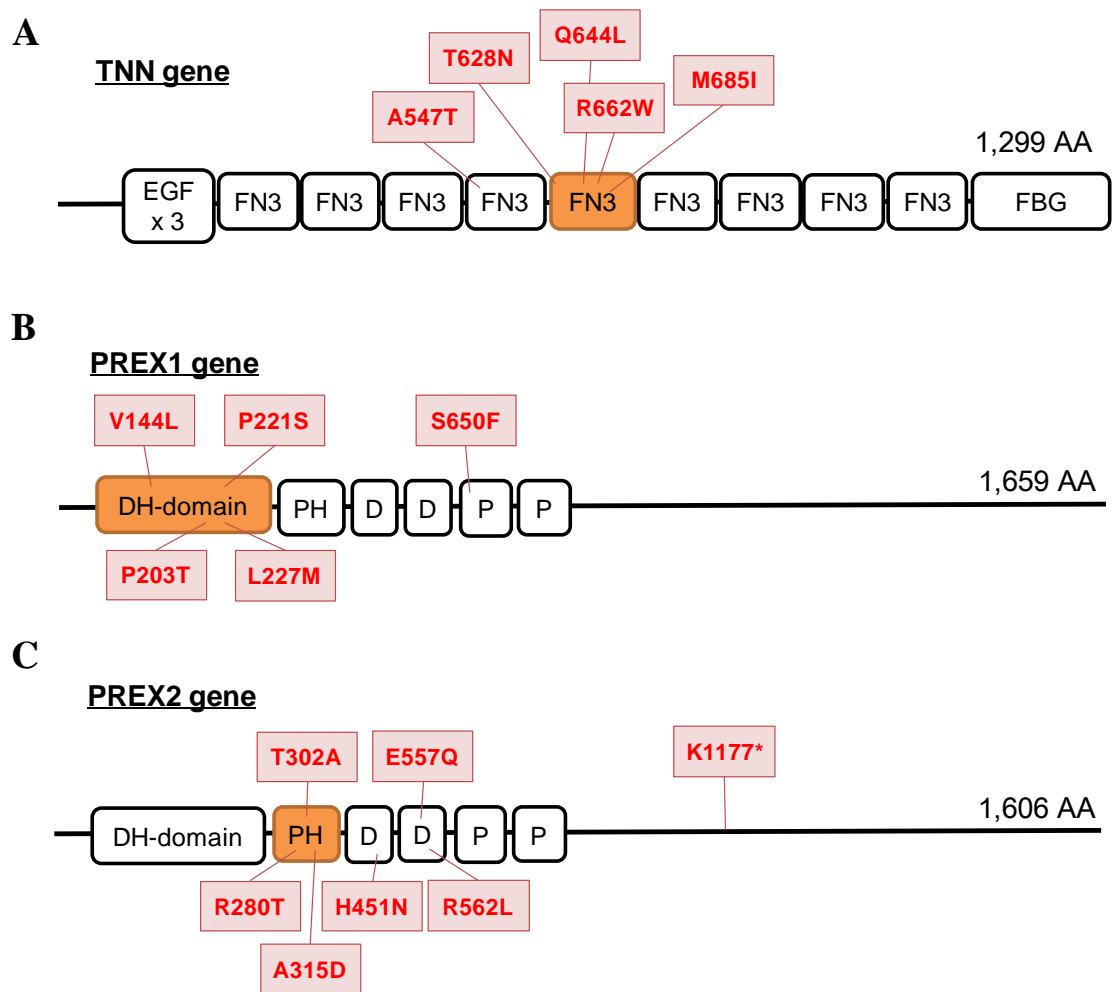


図4 タンパク質機能ドメインに SNVs が濃縮している遺伝子の例

タンパク質機能ドメインに SNVs が有意に濃縮している遺伝子 (表 3 参照) の例として、(A) TNN、(B) PREX1 および (C) PREX2 を図示した。白い箱はタンパク質機能ドメインを示し、SNVs が偏在しているドメインはオレンジで示されている。図中のタンパク質機能ドメインは以下のとおりである。EGF : EGF-like domain ; FN3 : Fibronectin, type III domain ; FBG : Fibrinogen, alpha/beta/gamma chain, C-terminal globular domain ; DH-domain : Dbl homology (DH) domain ; PH : Pleckstrin homology domain ; D : DEP domain ; P : PDZ/DHR/GLGF.

(Figure 1 および Figure S9, Suzuki and Mimaki *et al.* 2013 *PLoS ONE* 8(9):e73484 doi: 10.1371/journal.pone.0073484 改変)

表 5 有意に変異しているパスウェイ

KEGG ID	Pathway definition	Number of cancers with SNVs	P-value*
hsa05213	Endometrial cancer	72	3.1e-15
hsa04320	Dorso-ventral axis formation	48	4.4e-15
hsa05219	Bladder cancer	62	4.9e-14
hsa05223	Non-small cell lung cancer	66	7.1e-12
hsa05214	Glioma	70	6.5e-11
hsa05218	Melanoma	70	1.3e-9
hsa05212	Pancreatic cancer	68	6.9e-9
hsa05215	Prostate cancer	71	4.3e-7
hsa05216	Thyroid cancer	36	1.1e-6
hsa04520	Adherens junction	59	3.7e-6
hsa05210	Colorectal cancer	53	1.8e-5
hsa04012	ErbB signaling pathway	64	2.6e-5
hsa05120	Epithelial cell signaling in Helicobacter pylori infection	53	4.8e-5
hsa04540	Gap junction	60	0.00024
hsa04912	GnRH signaling pathway	61	0.0011
hsa05217	Basal cell carcinoma	41	0.0020
hsa05222	Small cell lung cancer	52	0.0069
hsa05220	Chronic myeloid leukemia	46	0.010
hsa05160	Hepatitis C	67	0.012
hsa05014	Amyotrophic lateral sclerosis (ALS)	36	0.014
hsa04977	Vitamin digestion and absorption	20	0.015
hsa05416	Viral myocarditis	40	0.028
hsa04512	ECM-receptor interaction	47	0.034
hsa02010	ABC transporters	29	0.035
hsa04510	Focal adhesion	78	0.037
hsa05412	Arrhythmogenic right ventricular cardiomyopathy (ARVC)	40	0.039

* $p < 0.05$; Poisson 分布を用いて算出した。

(Table 2, Suzuki and Mimaki *et al.* 2013 *PLoS ONE* 8(9):e73484 doi: 10.1371/journal.pone.0073484)

ゲノム変異と予後の関係性を解析した。SNVs を持つ症例と持たない症例で有意に生存期間が異なる遺伝子およびパスウェイを探索した (表 6)。その結果、23 遺伝子および 7 パスウェイを予後関連因子候補として抽出した。ATM serine/threonine kinase (ATM) は、SNVs が検出された 4 症例で有意に全生存期間が短く、変異を持たない症例群と比して全生存期間について統計的有意差が認められた ($p=4.7e-5$; 図 5A および 5B)。ATM は、DNA 修復系や細胞周期の制御、アポトーシスに関わっている遺伝子であり、TP53 や breast cancer 1, early onset (BRCA1) といった癌抑制遺伝子に作用する (図 5C) (45,46)。ATM 変異は肺腺癌の先行研究においてもその変異が報告されており (8)、ATM が変異していることによって ATM の機能阻害または欠失が引き起こされ、細胞の癌抑制機能が低下する可能性が示唆された。

表 6 予後関連遺伝子 (A) およびパスウェイ (B)

A

Gene	Num of cases		Prognosis of SNVs(+)	P-value*
	SNVs(+)	SNVs(-)		
ATM	4	93	poor	4.7e-5
AFF4	3	94	poor	0.00015
SPTBN4	3	94	poor	0.00032
ZNF493	3	94	poor	0.00058
MTR	3	94	poor	0.0014
NF1	7	90	poor	0.0034
SLC4A3	4	93	poor	0.0035
SORCS1	4	93	poor	0.0044
HTR2C	3	94	poor	0.0059
ZNF91	3	94	poor	0.0064
DAB2	3	94	poor	0.016
HUWE1	3	94	poor	0.018
SLC34A1	4	93	poor	0.018
ANKRD50	3	94	poor	0.019
KLHL20	3	94	poor	0.020
HECW1	3	94	poor	0.025
OGDH	3	94	poor	0.026
STAB2	3	94	poor	0.030
DOCK2	7	90	poor	0.031
ANO3	4	93	poor	0.033
CDH8	4	93	poor	0.038
BNC1	3	94	poor	0.039
FER1L6	5	92	good	0.046

B

KEGG id	Pathway definition	Num of cases		Prognosis of SNVs(+)	P-value*
		SNVs(+)	SNVs(-)		
hsa04080	Neuroactive ligand-receptor interaction	57	40	poor	0.00064
hsa00670	One carbon pool by folate	6	91	poor	0.0084
hsa00260	Glycine, serine and threonine metabolism	14	83	poor	0.0085
hsa03013	RNA transport	34	63	good	0.010
hsa00380	Tryptophan metabolism	15	82	poor	0.020
hsa00020	Citrate cycle (TCA cycle)	18	79	poor	0.042
hsa04977	Vitamin digestion and absorption	20	77	good	0.046

* $p < 0.05$; log-rank test を用いて算出した。

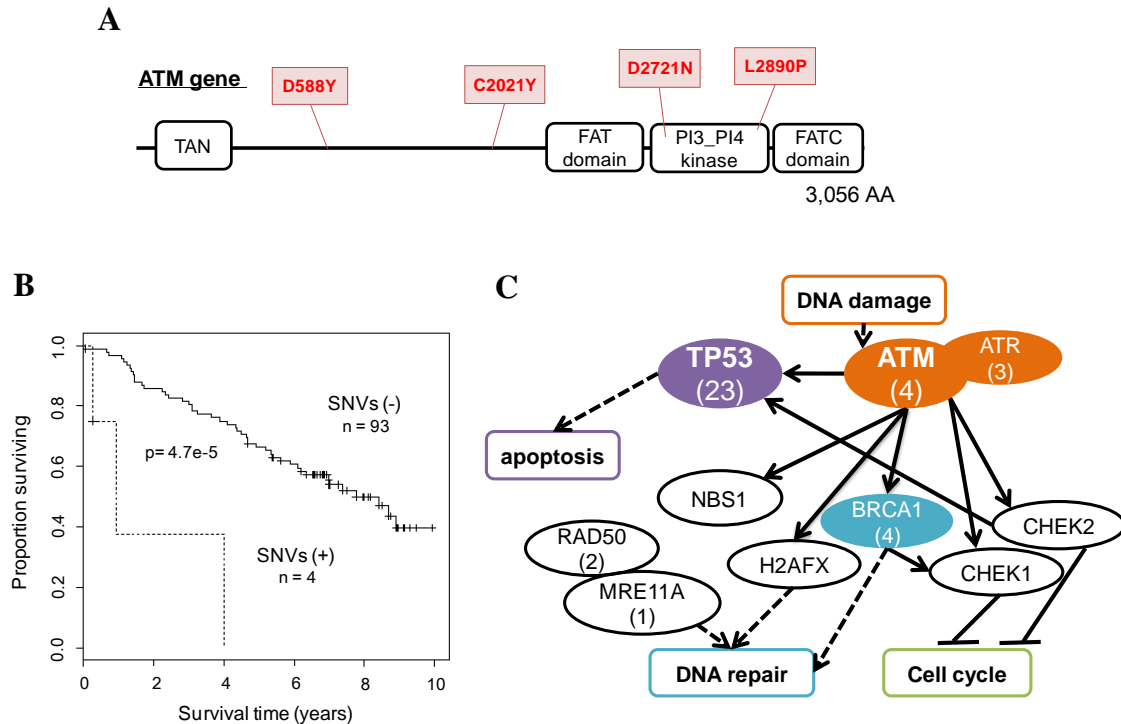


図5 予後関連遺伝子候補としての ATM 遺伝子とパスウェイ

(A) ATM で検出された体細胞性 SNVs を示している。白い箱はタンパク質機能ドメインを表している。TAN : Telomere-length maintenance and DNA damage repair ; PI3_PI4 kinase : Phosphatidylinositol 3-/4-kinase, catalytic. (B) ATM 変異陽性症例(点線)と陰性症例(実線)の全生存期間を Kaplan-Meier 法で記載した。P 値(log-rank test) は図中に示してある。(C) ATM を含む DNA 修復系およびアポトーシス、細胞周期制御に関わる遺伝子を図示した。括弧内の数値は、体細胞性 SNVs を持つ症例数である。これら SNVs はダイレクトサンガーシーケンスで体細胞突然変異であると確認されている。

(Figure 6A, 6B および Figure S11, Suzuki and Mimaki *et al.* 2013 *PLoS ONE* 8(9):e73484 doi: 10.1371/journal.pone.0073484 改変)

2-3-3. 日本人肺腺癌における体細胞突然変異パターン

SNVsに加えて、症例あたり平均 10 個（計 908 個）の体細胞性 indels を検出した。検出された indels のうち、最も高頻出だったのが EGFR における deletion であった。EGFR のエキソン 19 上に引き起こされるインフレーム deletion は、L858R 変異と同様に gefitinib および erlotinib 感受性の変異であり、ドライバー変異であることが知られている(9)。また、1 症例で EGFR のエキソン 20 における insertion (D770_N771insG) も見られた。この変異は L858R やエキソン 19 の deletion と同様の EGFR 活性を引き起こすが、小分子 EGFR キナーゼ阻害剤耐性である(9)。また、2 症例で ERBB2 のエキソン 20 における 12 塩基の insertion (GCATACGTGATG duplication) が見られた。

検出された SNVs および indels をあわせて、日本人肺腺癌における体細胞突然変異とした（図 6）。これら平均 112 個（計 10,667 個）の体細胞突然変異は、計 6353 遺伝子上に引き起こされており、EGFR や TP53 遺伝子に高頻出であった（図 7）。症例間で共通のものは少なく、個々に低頻度の多くの遺伝子上に引き起こされていた。

前述より喫煙歴と変異パターンの関連が示唆されていることから、変異遺伝子と喫煙歴との関連も調べた。EGFR 変異を持つ症例は非喫煙者に有意に多く見られた（ $p=0.00013$ ）。このことはいくつかの先行研究で報告されており、本研究でも同様の傾向が見られた。Mediator complex subunit 12 (MED12) の変異も非喫煙者に偏って引き起こされていた（ $p=0.015$ ）。MED12 変異陽性の 5 症例は全て EGFR 変異陽性であった。一方で、TP53 変異は、喫煙歴を有する症例において有意に濃縮していた（ $p=0.0036$ ）。非喫煙者および非喫煙者で有意に濃縮されている変異遺伝子は表 7 に示した。

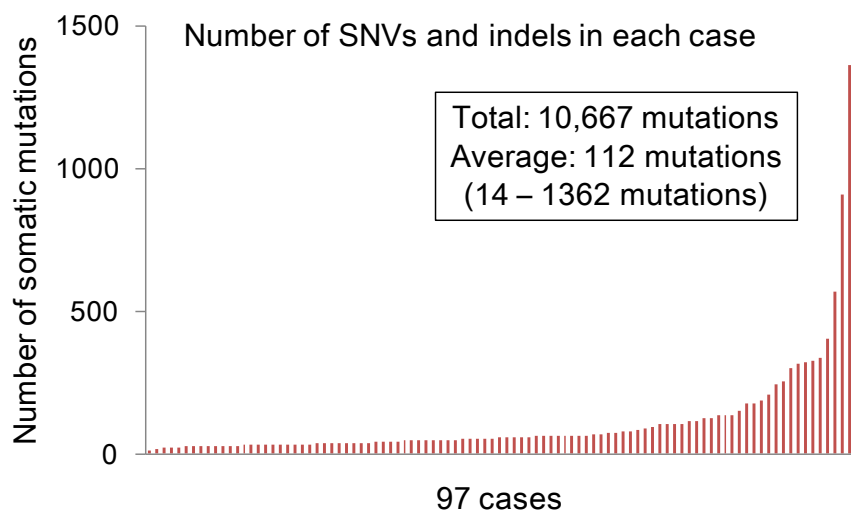


図 6 各症例における体細胞突然変異の数

各症例における癌部特異的体細胞性 SNVs と indels の合計数を示している。横軸は各症例を示しており、変異数で並べられている。総計および平均値はグラフ内に記載してある。

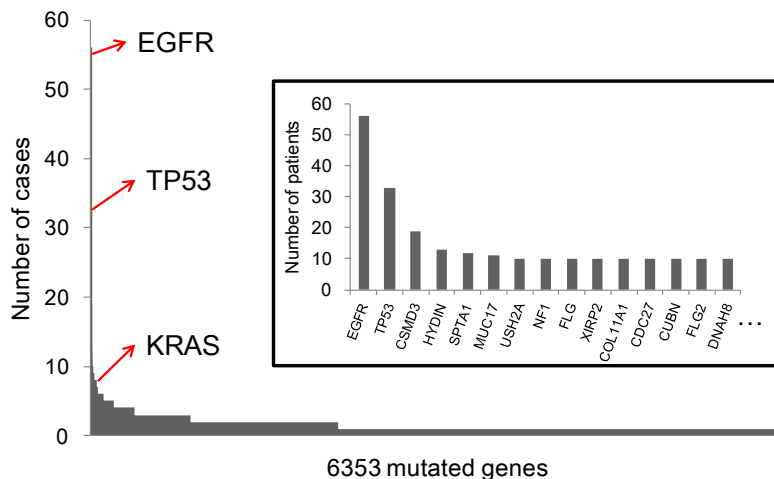


図 7 日本人肺腺癌全エクソーム解析で検出された変異遺伝子とその頻度

体細胞突然変異が検出された変異遺伝子の頻度（症例数）を示している。横軸は変異が見出された 6353 遺伝子、縦軸は各遺伝子について変異を持っていた症例数を示している。

表 7 非喫煙者 (A) および喫煙者 (B) に変異が濃縮している遺伝子

A

Gene	SNVs(+)		SNVs(-)		P-value *
	Non-smoker	Smoker	Non-smoker	Smoker	
EGFR	34	22	9	32	0.00013
MED12	5	0	38	54	0.015

B

Gene	SNVs(+)		SNVs(-)		P-value *
	Smoker	Non-smoker	Smoker	Non-smoker	
CSMD3	19	0	35	43	2.63e-6
MUC17	11	0	43	43	0.00096
COL11A1	10	0	44	43	0.0019
HYDIN	12	1	42	42	0.0034
TP53	25	8	29	35	0.0036
SPTA1	11	1	43	42	0.0063
LAMB4	8	0	46	43	0.0072
GRM8	8	0	46	43	0.0072
ANK2	8	0	46	43	0.0072
LRP2	7	0	47	43	0.014
SYNE1	7	0	47	43	0.014
EPHA3	7	0	47	43	0.014
NRXN3	7	0	47	43	0.014
KIAA1109	7	0	47	43	0.014
ANKRD30A	7	0	47	43	0.014
PRDM9	7	0	47	43	0.014

* $p < 0.015$; 超幾何分布を用いて算出した。

2-3-4. 民族間における肺腺癌ゲノム変異パターンの違い

日本人肺腺癌におけるゲノム変異の特徴を明らかにするために、他の肺腺癌ゲノムデータと比較を行った。肺腺癌で異常が頻出するパスウェイ(8)について、各遺伝子に対して日本人肺腺癌の変異と Ding *et al.* (8)に報告される 188 症例の欧米人肺腺癌の遺伝子変異頻度を比較した(図 8)。日本人肺腺癌では、EGFR 変異陽性の症例が 97 症例中 56 症例(57.7%)と特徴的であった。一方で、欧米人では KRAS を含む RAS 変異陽性の症例が多く(63 症例; 33.5%)、EGFR 変異陽性は日本人に多く RAS 変異陽性は欧米人に多い傾向が検出された。TP53 変異はどちらのデータセットでも頻出し、その頻度に差はなかった。

EGFR および KRAS の変異頻度に関して、Imielinski *et al.* (11)の 183 症例、TCGA (13)の 230 症例、および、Seo *et al.* (12)の韓国人肺腺癌 200 症例のデータとの比較も行った。日本人および韓国人肺腺癌において EGFR 変異が有意に多く見られた(図 9A)。アジア人肺腺癌に EGFR 変異が頻出していた。各症例における喫煙歴を確認したところ、日本人および韓国人のデータセットには非喫煙者が半数近く含まれており、それ以外のデータセットでは喫煙者の割合が多かった(図 9B)。図 9A で見られた差は喫煙歴によって生じている可能性が示唆されたが、喫煙者のみ、非喫煙者のみに症例を分けて変異頻度を比較したところ、アジア人では他のデータセットと比べて非喫煙者で約 2 倍、喫煙者で約 3-4 倍の EGFR 変異陽性症例が見られた(図 9C)。喫煙歴に関わらず、アジア人肺腺癌には EGFR 変異が濃縮していることが示された。肺腺癌は民族によってゲノム変異のパターンが異なり、日本人患者のデータを用いて民族ごとの解析を行うことが重要であることが示唆された。

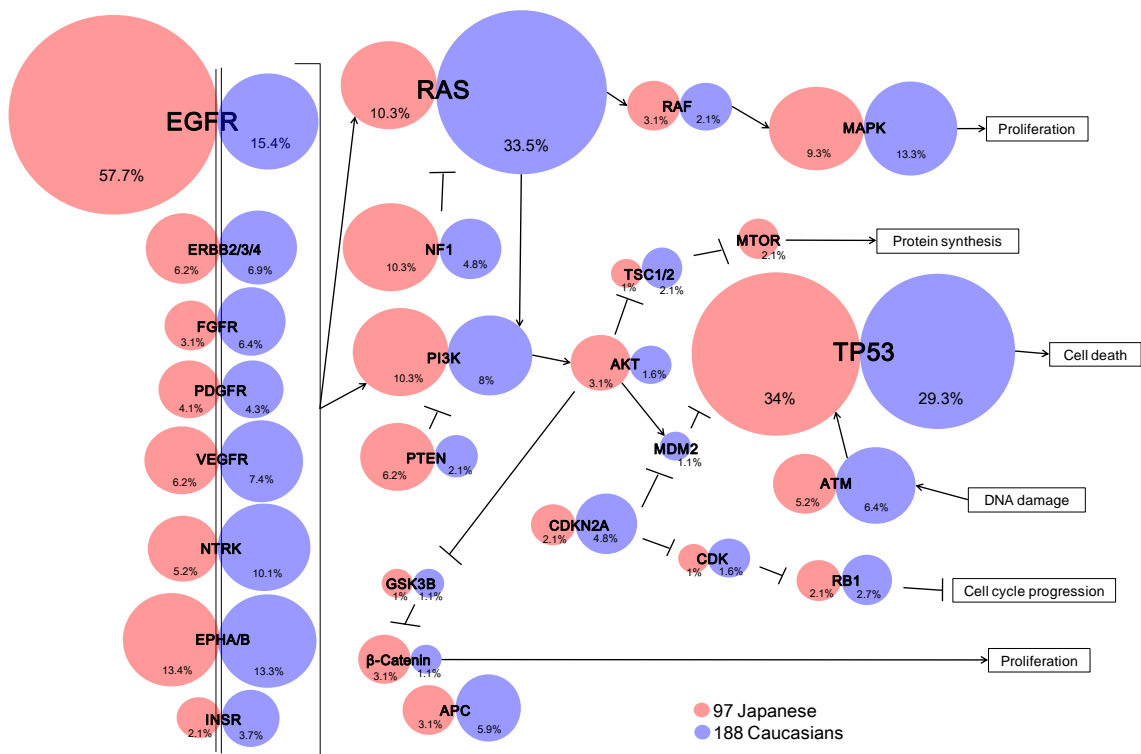


図 8 肺腺癌パスウェイにおける日本人肺腺癌と欧米人肺腺癌

日本人肺腺癌 97 症例（赤）および欧米人肺腺癌 188 症例（青）のゲノム変異を、肺腺癌で異常が頻出するパスウェイの上にマップし、その頻度を記載した。円の大きさは変異陽性症例の頻度を表しており、頻度（%）を図中に示している。パスウェイの構成遺伝子および 188 症例のゲノム変異頻度は Ding *et al.* (Figure 6 および Supplementary Table 2) (8) のデータを用いている。

(Figure 2, Suzuki and Mimaki *et al.* 2013 *PLoS ONE* 8(9):e73484 doi: 10.1371/journal.pone.0073484 参考)

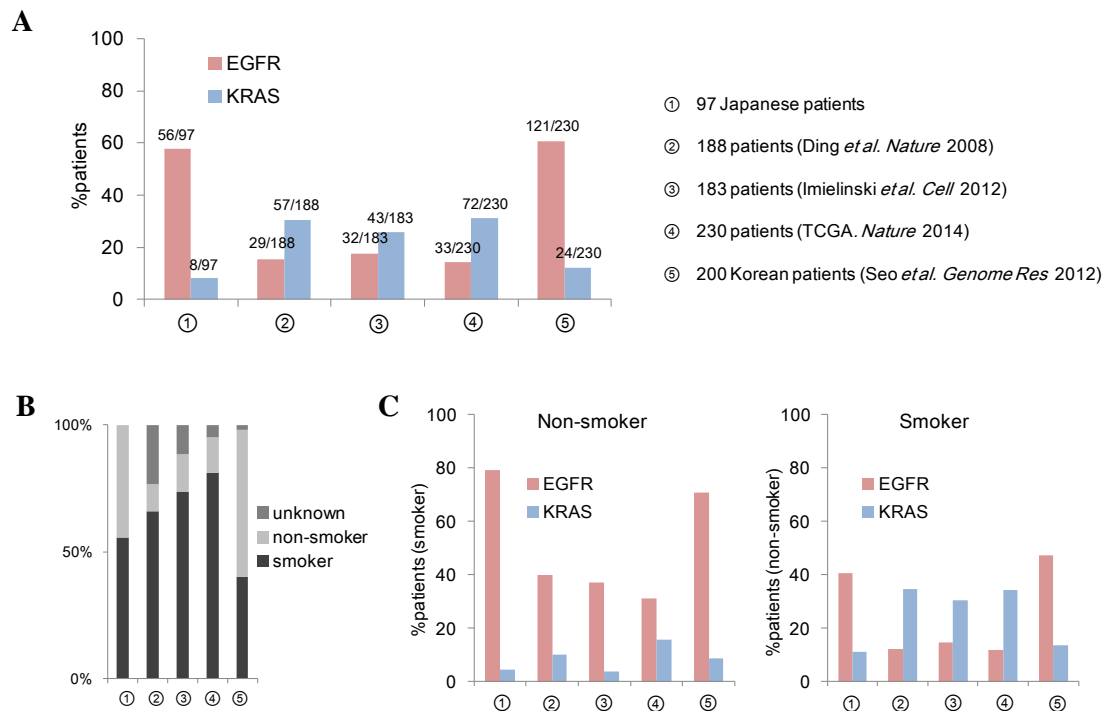


図 9 5つのデータセットにおける EGFR と KRAS 変異の頻度

(A) 日本人肺腺癌 97 症例 (①) と 4 つの肺腺癌ゲノムデータ (②-⑤) (8,11-13) における EGFR 変異と KRAS 変異の頻度を比較した。(B) 各データセットにおける喫煙者・非喫煙者の割合を図示した。喫煙歴ありの症例は全て喫煙者とした。(C) 非喫煙者 (左図) および喫煙者 (右図) における EGFR 変異と KRAS 変異の頻度を各データセットにおいて示した。喫煙歴不明の症例は解析対象から除外した。

2-4 小括

日本人肺腺癌 97 症例の全エクソームシーケンスデータから、平均 112 個、計 10,667 個の体細胞突然変異を検出した。数多くのゲノム変異から、統計学的手法を用いて、TNN や ATM といった癌関連因子候補を抽出することができた。

EGFR や TP53 に変異が頻出していた。図 8 および 9 に示したように、欧米人と日本人の肺腺癌を比較すると、日本人では EGFR の変異が多く見られ、欧米人では KRAS の変異が多く見られた。TP53 は両者で高頻度に変異が起こっており、民族間の変異頻度に違いはなかった。アジア人の肺腺癌に EGFR 変異が頻出することはすでに報告されているが(47)、本研究における日本人肺腺癌にもその傾向が見られることが示された。

癌細胞におけるゲノム変異のパターンは極めて複雑であり、どの変異が直接癌化および予後に関わっているのかを推定するのは依然として困難であった。EGFR 変異は日本人肺腺癌の特徴として顕著であるが、検出された変異の多くは EGFR 変異とは異なり数多くのさまざまな遺伝子上に引き起こされていた。日本人肺腺癌のいくつかの症例では、細胞のトランスクリプトームやエピゲノム状態に関わる遺伝子においても変異が見出された（後述）。日本人肺腺癌から見出された数多くのゲノム変異と癌細胞の関係をより明らかにするために、ゲノム変異だけでなく、トランスクリプトームおよびエピゲノムの異常といった多階層的な情報を新たに加え、癌細胞で起こっている現象を包括的に解析する必要があると考えられた。

III 章 26 肺腺癌細胞株のマルチオミクス解析

3-1. 緒言

II 章で述べた日本人肺腺癌 97 症例の全エクソーム解析(23)から、肺腺癌細胞の数多くのゲノム変異が同定された。これらゲノム変異は、多くの遺伝子に散在して観察され、EGFR や TP53、KRAS 等の一部の遺伝子を除いて、個々の遺伝子について集団中の生起頻度は低かった。これら多数のゲノム変異から、癌細胞の表現型にとって機能的な変異を、大多数を占めるとされる機能的に中立な変異から区分することは困難であった。また、検出された変異には、スプライシング関連遺伝子（例：RBM10）やヒストン修飾関連遺伝子（例：EP300）、クロマチンリモデリング関連遺伝子（例：ARID1A）の変異も含まれており、これらの症例については遺伝子発現制御が発癌および癌の生物学的形質に関与している可能性が示唆された。これら遺伝子変異について、その機能的帰結を明らかにするためには、ゲノムだけでなく、トランスクリプトームおよびエピゲノムデータを用いた多階層的解析が必要である。しかし、臨床検体を用いた研究では、解析に用いることができる試料の質・量の問題により、同一検体からこれらマルチオミクスデータを取得することは困難であった。III 章では、肺腺癌細胞株を用いたゲノム、エピゲノムおよびトランスクリプトームデータの収集と解析を行った。取得したマルチオミクスデータを用いて、肺腺癌細胞におけるゲノム・エピゲノム変異と遺伝子発現・転写制御異常の関係性を明らかにする解析モデル系の構築を行った。

3-2. 材料と方法

3-2-1. 細胞株と培養方法

26 種類のヒト肺腺癌細胞株を用いた。13 株が日本人由来、13 株が欧米人をはじめとした非アジア人由来である(表 8 および図 10)。抗生剤(Antibiotic-Antimycotic, Gibco)、非必須アミノ酸 (MEM 非必須アミノ酸溶液、Sigma)、10%ウシ胎児血清、pH 緩衝剤 (25mM HEPES-KOH pH 7.4) を添加した RPMI (RPMI 1640 培地、ニッスイ)、DMEM (ダルベッコ変法イーグル培地②、ニッスイ) または EMEM (イーグル MEM 培地①、ニッスイ) を用いて、37°C の 5% CO₂ インキュベーターで培養した。LC2/ad および A549 は DMEM、RERF-LC-MS、VMRC-LCD および ABC-1 は EMEM、その他 21 株は RPMI を用いた。PC-3、LC2/ad、H1648、H2347 は、コラーゲンディッシュ (タイプ I コラーゲンコートプレート、IWAKI) を用いて培養した。

また、正常細胞コントロールとして、正常ヒト小気道上皮細胞 (small airway epithelial cells : SAEC ; CC-2547、Takara) の培養も行った。小気道上皮細胞培地キット (無血清) (SAGM BulletKit) (CC-3118、Takara) を用いて、コラーゲンディッシュ上で培養した。

表 8 本研究で用いた 26 種類の肺腺癌細胞株

Cell line	Ethnicity	Distributor*	No.
PC-3	Japanese	JCRB	JCRB0077
PC-7	Japanese	IBL	***
PC-9	Japanese	RIKEN BRC	RCB4455
PC-14	Japanese	IBL	***
RERF-LC-Ad1	Japanese	JCRB	JCRB1020
RERF-LC-Ad2	Japanese	JCRB	JCRB1021
RERF-LC-KJ	Japanese	RIKEN BRC	RCB1313
RERF-LC-MS	Japanese	JCRB	JCRB0081
RERF-LC-OK	Japanese	JCRB	JCRB0811
VMRC-LCD	Japanese	JCRB	JCRB0814
ABC-1	Japanese	JCRB	JCRB0815
LC2/ad	Japanese	RIKEN BRC	RCB0440
II-18	Japanese	RIKEN BRC	RCB2093
A549	Caucasian†	ATCC	CCL-185
A427	Caucasian†	ATCC	HTB-53
H322	Caucasian†	ATCC	CRL-5806
H2228	Unknown†	ATCC	CRL-5935
H1299	Caucasian†	ATCC	CRL-5803
H1437	Caucasian†	ATCC	CRL-5872
H1648	Black†	ATCC	CRL-5882
H1650	Caucasian†	ATCC	CRL-5883
H1703	Caucasian†	ATCC	CRL-5889
H1819	Caucasian†	ATCC	CRL-5897
H1975	Unknown†	ATCC	CRL-5908
H2126	Caucasian†	ATCC	CCL-256
H2347	Caucasian†	ATCC	CRL-5942

*IBL: Immuno-Biological Laboratories; JCRB: Japanese Collection of Research Bioresources; RIKEN BRC: RIKEN Bio Resource Center; ATCC: American Type Culture Collection.

†参照: ATCC web ページ (<http://www.atcc.org/>).

(Supplementary Table S1, Suzuki *et al.* 2014 *Nucleic Acids Research* 42(22):13557-13572 doi: 10.1093/nar/gku885 改変)

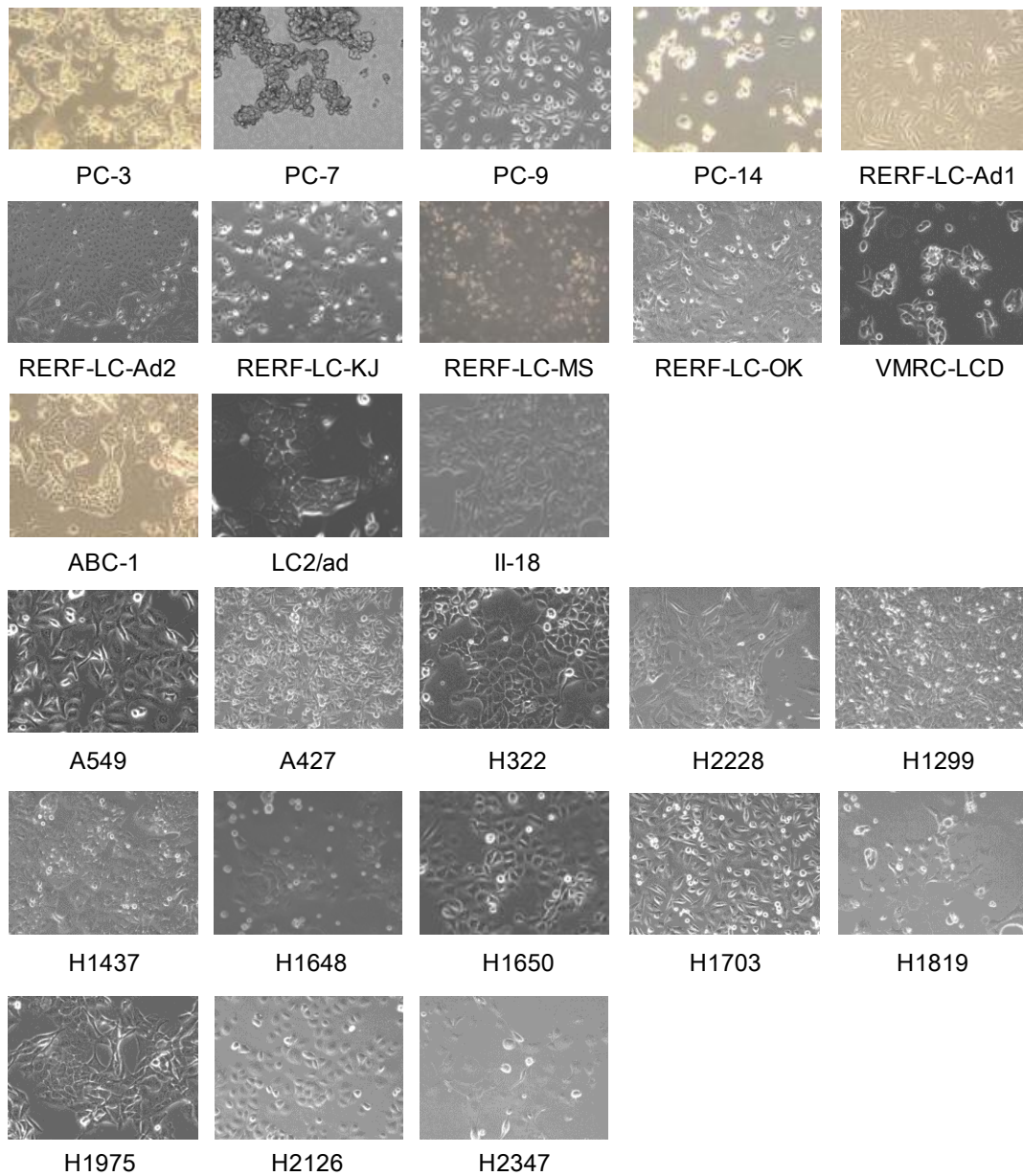


図 10 26 肺腺癌細胞株の写真

本研究で用いた肺腺癌細胞株 26 種類の写真を示している。上が日本人、下が非アジア人由来の細胞株である。PC-7 は浮遊細胞で、その他の 25 株は接着細胞である。

3-2-2. 全ゲノムシーケンシング

約 10^6 個の細胞を $1\times$ phosphate buffered saline (PBS) で洗浄し、DNeasy Kit (キアゲン) を用いて、ゲノム DNA を抽出した。ND-1000 (ナノドロップ) を用いて、抽出した DNA 量を測定した。抽出した DNA からシーケンズライブラリを作製し、イルミナ社の HiSeq2500 を用いて、各細胞株 3 - 5 レーンの 101 塩基長ペアエンド全ゲノムシーケンスタグを取得した。

3-2-3. RNA-Seq

RNeasy Maxi Kit (キアゲン) を用いて、PBS で洗浄した約 10^7 個の細胞から RNA を抽出した。BioAnalyzer (アジレント) を用いて、抽出した RNA の RNA Integrity Number (RIN) 値が 9 以上であることを確認した。イルミナ社のプロトコルに則って、ポリ A 選択をしたライブラリを作製し、HiSeq2500 を用いて、101 塩基長のペアエンド RNA-Seq タグを取得した。

3-2-4. バイサルファイトシーケンシング

SureSelect XT Methyl-Seq ターゲットエンリッチメントシステム (アジレント) を用いて、 $3\ \mu\text{g}$ のゲノム DNA からバイサルファイトシーケンシングのライブラリを作製した。コバリスを用いてゲノム DNA を断片化し、約 84 Mb のベイト領域にハイブリダイズさせて断片化 DNA を濃縮した。EZ DNA Methylation-Gold Kit (Zymo Research) を用いて、バイサルファイト変換を行った。アジレント社およびイルミナ社のプロトコルに従って作製したライブラリから、HiSeq2500 を用いて 101 塩基長のペアエンドバイサルファイトシーケンスタグを取得した。

3-2-5. ChIP-Seq

7 種類のヒストン修飾 (H3K4me1、H3K4me3、H3K9me3、H3K9/14ac、H3K27ac、H3K27me3、H3K36me3) および RNA ポリメラーゼ II (Pol II) のクロマチン免疫沈降 (Chromatin Immuno-Precipitation ; ChIP) を行った(48,49)。各 ChIP 実験に対して 10^7 - 10^8 個の細胞を用いた。1% ホルムアルデヒド溶液を培地中に添加し (PC-7 は浮遊細胞のため 0.5%)、10 分室温でクロスリンクを行った。クロスリンク反応を止めるために 125 mM グリシンを添加し、5 分室温で静置した。クロスリンク後の細胞は PBS で二度洗浄し、 -80°C フリーザーに保存した。

免疫沈降に用いた抗体は、H3K4me1 抗体 (ab8895、アブカム)、H3K4me3 抗体 (ab1012、アブカム)、H3K9me3 抗体 (ab8898、アブカム)、H3K9/14ac 抗体 (06-599、ミリポア)、H3K27ac 抗体 (ab4729、アブカム)、H3K27me3 抗体 (07-449、ミリポアおよび ab6002、アブカム)、H3K36me3 抗体 (ab9050、アブカム) および、Pol II 抗体 (ab817、アブカム) である。抗体と磁気ビーズを反応させるために、100 μl の磁

気ビーズ (Dynabeads Protein G または A、インビトロジェン) を阻害溶液 (0.5% ウシ血清アルブミンを添加した PBS 溶液) で三度洗浄し、250 μ l の阻害溶液と 10 ng の抗体 (H3K27me3 抗体は 20 ng) を加えて、4°C で 4 時間以上転倒混和した。

クロスリンクした細胞は、プロテアーゼインヒビター (ロッシュ) を加えた 3 種類の溶解バッファーを用いて溶解した。まず、5 ml の溶解バッファー1 (50 mM HEPES-KOH (pH 7.5)、140 mM 塩化ナトリウム、1 mM EDTA (pH 8.0)、0.5 mM EGTA (pH 8.0)、10% グリセロール、0.5% ノニデット P-40、0.25% Triton X-100) で細胞を溶解し、氷上で 10 分静置した。1500 rpm で 5 分遠心し、上清を捨てた。同様に 5 ml の溶解バッファー2 (200 mM 塩化ナトリウム、1 mM EDTA (pH 8.0)、0.5 mM EGTA (pH 8.0)、10 mM Tris-HCl (pH 8.0)) で細胞を溶解し、氷上で 10 分静置した。1500 rpm で 5 分遠心し、上清を捨て、1 ml の溶解バッファー3 (100 mM 塩化ナトリウム、1 mM EDTA (pH 8.0)、0.5 mM EGTA (pH 8.0)、10 mM Tris-HCl (pH 8.0)、0.1% デオキシコール酸ナトリウム、1% N-ラウロイルサルコシン) に溶かした。次に、ソニケーター (TOMY SEIKO) を用いて、氷上で 30 秒のソニケーションを 16 回繰り返した (PC-7 の場合は 10 回。インターバルは 2 分)。10% Triton X-100 を 10 μ l 加え、14,000 rpm で 10 分遠心した。上清を 1.5 ml チューブに移し、50 μ l を whole-cell extract (WCE) サンプル、残りを ChIP サンプルとして分けた。WCE サンプルは -20°C フリーザーで保存した。抗体と反応させた磁気ビーズを阻害溶液で洗浄し、ChIP サンプルに加え、4°C で一晩ローテーターにかけてクロマチン免疫沈降を行った。

クロマチン免疫沈降した ChIP サンプルを洗浄バッファー (50 mM HEPES-KOH (pH 7.5)、500 mM 塩化リチウム、1 mM EDTA (pH 8.0)、1% ノニデット P-40、0.7% デオキシコール酸ナトリウム) で 8 回、塩化ナトリウムを加えた TE バッファー (50 mM Tris-HCl (pH 8.0)、10 mM EDTA (pH 8.0)) で 1 回洗浄した。4°C、1000 rpm で 3 分遠心し、上清を除去後、200 μ l の溶出バッファー (50 mM Tris-HCl (pH 8.0)、10 mM EDTA (pH 8.0)、1% ドデシル硫酸ナトリウム) を加え、65°C のヒートブロックで 15 分インキュベート (2 分ごとにチューブをタッピング) した。室温 14,000 rpm で 1 分遠心し、上清を 1.5 ml チューブに回収した。また、保存してあった WCE サンプルに 150 μ l の溶出バッファーを加えた。脱クロスリンクのため、ChIP および WCE サンプルを 65°C のヒートブロックで 6 時間以上インキュベートした。

脱クロスリンクした ChIP および WCE サンプルに 200 μ l の TE バッファーを加えた。RNA を分解するために、8 μ l の 10 mg/ μ l の RNase A (Novagen) を加え、37°C で 2 時間静置した。次にタンパク質を分解するために、4 μ l の 20 mg/ μ l プロテアーゼ K (Takara) と 7 μ l の 300 mM 塩化カルシウムを加え、55°C で 30 分インキュベートした。フェノール・クロロホルム抽出およびエタノール沈殿を用いて DNA を回収し、35 μ l に溶解した。

BioAnalyzer を用いて、得られた ChIP および WCE サンプルの DNA 量を確認し、 $\Delta\Delta$ Ct 法を用いて ChIP サンプルにおける既知修飾領域の濃縮を確認した (表 9、プライマー)。濃縮が確認できたサンプルを用いて、ChIP-Seq ライブラリを作製し、HiSeq2500 を用いて各 ChIP および WCE サンプルのシングルエンドシーケンスタグを取得した。

表 9 $\Delta\Delta$ Ct 法による ChIP 濃縮確認のための qPCR プライマー

Target region	ChIP	Fwd primer	Rev primer
GAPDH +30, +113	H3K4me3, H3K9/14ac, Pol II	CTCTGCTCCTCCTGTTCGAC	CTAGCCTCCGGGTTTCTCT
GAPDH -359, -271	H3K9/14ac	CAATFCCCATCTCAGTCGT	GCAGCAGGACACTAGGGAGT
GAPDH down 1.5-2k	H3K36me3	CTTCGTATGACTGGGGGTGT	GGCCCAAGAGGTTGAATTTT
ACTB down 1.5-2k	H3K36me3	TTTGCACTTTCTGCATGTCC	GGGGTGTGAAGGTCTCAAA
CDKN1A +3959 +4074	H3K4me1, H3K27ac	AAAATCCAGTTGCTGCCAAG	AGGAGAGGGAAAACGGAGAG
GAPDH +2165 +2268	H3K4me1	CCCCTTCATACCCTCACGTA	GACAAGCTTCCCGTTCTCAG
ACTB +356 +460	H3K27ac	TCTTGCCAATGGGGATCG	CAGTTAGCGCCAAAGGAC
HOXD13 -925 -822	H3K27me3	CGTTCTCAGGGTCTTTCTGC	CAACGCTCTTCGCCTCTACT
HOXD13 -538 -449	H3K27me3	GGGCCTGTTTTACATCATT	TGTTTTTGCCACCTCTCCTC
MYOD1 +994, +1079	H3K27me3	CTGTCACTGGGGTAGCCTGT	CGCGCTGCAGTGTCTAATA
ZNF274 +1725 +1812	H3K9me3	GTGATGGTGCAGCTCAGAGA	CGGCCCCCTATTTCTTATTC
ACT B upstream	Negative control	CGTGTCCCCATATCAGAAC	TCAGCCTCAGTCTCCCTTGT

(Supplementary Table S4A, Suzuki *et al.* 2014 *Nucleic Acids Research* 42(22):13557-13572 doi: 10.1093/nar/gku885 改変)

3-2-6. ゲノム変異の探索

得られた全ゲノムシーケンスを、BWA を用いてヒト参照ゲノム配列にマップした。SAMtools を用いて PCR duplicate の除去を行い、GATK Unified Genotyper と Somatic Indel Detector (v1.6) を用いて、SNVs および indels を検出した。SNVs については、Confidence score が 50 以上、かつ、変異部分の QV が 20 以上のタグが 4 以上である変異を抽出した。indels については、変異を含むタグが 4 以上 (Fwd と Rev を必ず含む) である変異を抽出した。また、既知のゲノム多型を取り除くために、NCBI dbSNP build 137、1000 Genomes Project (> 0.1%) (30,31)、NHLBI Exome Sequencing Project (> 0.1%) (Exome Variant Server ; アクセス日時 : 2013 年 10 月 8 日 ; URL : <http://evs.gs.washington.edu/EVS/>) と in-house の日本人 germline データに登録された多型を除去した。さらに、除去された変異のうち COSMIC (v59) (33,34) に "Confirmed somatic variant" として登録されていた変異を戻し、体細胞突然変異候補とした。

また、全ゲノムシーケンスのデータからコピー数の予測を行った。Control-FREEC (50,51) を用いて、50 kb (可視化) および 1.5 kb (遺伝子領域のコピー数解析) ごとに、2 倍体に標準化したコピー数を算出した。4 以上をコピー数増加、1 以下をコピー数減少として、各細胞株における遺伝子領域のコピー数異常の数を算出した。

さらに、染色体構造異常を探索した。全ゲノムシーケンスデータの各リードをシングルエンドとしてヒト参照ゲノム配列にマップした。ペアが異なる染色体、もしくは、1 Mb 以上離れた場所にマップされ、さらに切断点候補が両方向からサポートされていた場合、その切断点候補を含むタグを参照タグとして抽出した。次に、参照タグに対して全てのシーケンスタグをマップし、参照タグにミスマッチなしにマップされたタグをサポートタグとして抽出した。参照タグが 2 以上、サポートタグが 4 以上得られた切断点を含む遺伝子ペアを染色体構造異常候補とした。

3-2-7. 遺伝子発現解析

RNA-Seq タグを ELAND (イルミナ) を用いてヒト参照ゲノム配列にマップした。遺伝子発現量として、20,598 遺伝子に対して parts per million mapped reads (PPM) および reads per kilo base million mapped reads (RPKM) (17) を算出した。また、異常な遺伝子発現を抽出するために、各細胞株において、26 細胞株平均値と比して 4 倍以上もしくは 1/16 倍以下の発現量を示す遺伝子を”differentially expressed genes”として抽出した。ハウスキーピング遺伝子、癌関連遺伝子、および、differentially expressed genes を含む 52 遺伝子の発現量について、qRT-PCR による確認を行った。用いたプライマーは Primer3Plus (52) で設計した (表 10)。また、RNA-Seq タグを Integrative Genomics Viewer (IGV) (53,54) で可視化するためのファイルは TopHat2 (55) と SAMtools を用いて作製した。

表 10 52 遺伝子と GAPDH (コントロール) の qRT-PCR プライマー

Gene	Fwd primer	Rev primer	Gene	Fwd primer	Rev primer
KRAS	CACGGTCATCCAGTGTGTGC	TTGATTTGTCAGCAGGACCA	SMAD4	GGCCTTAGTTGCCTCATCTG	AATCCAAAATGCTGCAAGG
NRAS	ATTGCCTTCAATCCCCTTCT	GGTCCATTAGCACCATGCTT	MDM2	GAGCCTCCAATGAGAGCAAC	AGGCTGCCATGTGACCTAAG
MYC	AGCGACTCTGAGGAGGAACA	CTCTGACCTTTTGCCAGGAG	WNT10A	AACCCACCACTCACTTCTG	GGCTTGTTTAGGGTGTGGA
PIK3CA	GTCAATCGGTGACTGTGTGG	GAACATGCAGTGACCTTTCA	RNF43	AGGTGAGCCCTGAGCACTAA	GTACCCACCTCAGGGACAGA
ERBB2	CTGACTTCTGTGGCATCAA	CTGGGAACCTAAGCAGGAAG	CDKN1B	CCGGCTAACTCTGAGGACAC	TTGCAGGTGCTTCTTATT
BRAF	GACGGGACTCGAGTGATGAT	CACATCACCATGCCACTTTC	GATA3	CAATGCCTGTGGGCTCTACT	TTTCTGGTCTGGATGCCCTTC
MET	CACCCACCCCTCATTCATC	GCAACAACAGCCAAACTCAA	ZFP36L1	ATGCTTTCTTCCTGGGTTT	TGATGTGCAGTGACAAGCAA
AKT1	CACACCACCTGACCAAGATG	CTGGCCGAGTAGGAGAAGT	RPL5	CAAGTGGAGGTGACTGGTGA	CCTGCATCCAAATAGCAGGT
CDKN2A	CACATTCATGTGGGCATTTT	CCCCTGAGCTTCCCTAGTTC	MYCN	TCCAGCAGATGCCACATAAG	CCACCTCTCATTCCAGGA
CDKN1A	TTAGCAGCGGAACAAGGAGT	CAACTACTCCCAGCCCATA	MYCL1	CAGGTGGTGGATGTTGTCAG	AGTGGGAGCATCCCTTTCT
STK11	CTTCGACATCGAGGATGACA	TTGTGACTGGCCTCCTCTTC	EGF	TTATGTGGAATGGGCACAGA	GAGGGCATATGAAAGCTTCG
KEAP1	AAGAAGTCCAAAGCGGGAAT	GAAGCCTGCTCTTTCCACAC	COL4A2	AGGCCAAAACCGTGCTCTTTA	GGTCTCCACTTCTGCCTGAG
NF1	CTTTTTCGACTGCGTTACA	CCGGATGGAAGGACAGTAA	ATP5F1	AGGCTCCATCCAACACATC	AAAGGTAATGGCGCTTCTGA
BRC1A	AACCATTTAAACGCCACCAA	ACTCTGTGCTTCCAGCCCTA	RPS18	GAGGATGAGGTGGAACGTGT	GGACCTGGCTGTATTTTCCA
APC	TCTTGTGTGCAACTGGGCTCTG	TTTCTGAGCATGGCAAAGTG	KIT	TGATCCCAAGTGTGAACAA	GGAGGCCCTTCTCTCAATTC
RB1	GCAGATTGTTCTCTTCCAA	TCAGACAGGCCTGGGTTATC	ELF3	AGCCCACTCTGGCTCATTTA	CCTGTGGAACCTCTTCTGTC
PTEEN	CATAACGATGGCTGTGGTTG	CCCCACTTTAGTGCACAGT	RAB25	CCCTCCTGGTGTGACCTA	TTTGTACCCACGAGCATGA
MSH6	TGAACAGAGCCTCTGGAAT	GTTCTCAACATCCCCCAAT	C1orf51	CCCCAGTCCCACCTACTACA	CTGGCAAACTGTAGCAACGA
SMARCA4	GACAGTGAAGGCGAGGAGAG	CACCTTTGACGGACCGAGATT	C17orf28	GATACAGCGGGTGTGAGGAT	AAGCTCAAGATGGTGGATGG
EP300	AATATGCTCCTGCCTTGAC	AATGTCAACCATCTGCACCA	JUND	CGCCTGGAAGAGAAAGTGAA	GTTGACCTGGCTGAGGACTT
ARID1A	GCATCCAGGACAAACATGTG	CTTTCCGGTCACTGAGGAAGC	TCF7L1	CCGCATTTAAAGGGACTCAA	TGGGATGTGAAGGGATGATG
RET	ACAGGGGATGTCAGTATCTGG	AAGCCGAAATCCGAAATCTT	LAMA5	AGCGGGGGAGTTATCACTTT	GCCCAAGTGGAAATGACATC
ALK	TGGAGTTTGTCAACAGTGGGA	TGTCTTCAGGCTGATGTTGC	REST	TGTTTGCATGGAACCTGTTGA	TCCAAAGGGAAAAATTCAGG
ROS1	AGGCTCCCAGGAATCTGAAT	CTTGCCAGAAAGGGCAGTAAG	GAPDH	CGACCACCTTTGCAAGCTCA	GGTGGTCCAGGGGCTTACT
ACTB	GATGAGATTGGCATGGCTTT	CACCTTCACCCGTTCCAGTTT	EGFR	TATGTTCCCTCCAGGTGAGC	TGTGGCCAGAGCTGTAAAGTG
IGF1R	ATCTGGCTTGTATGGTCTGG	CATGAAAACCATTTGGCTGTF	TP53	TTTGGGTCTTTGAAACCTTG	CCACAACAAAACACCAGTGC
SMARCA2	GCAGAAAATTGCCAAAGAGG	TTTTGCCTCGGACTCTGACT			

(Supplementary Table S2A, Suzuki *et al.* 2014 *Nucleic Acids Research* 42(22):13557-13572 doi: 10.1093/nar/gku885 改変)

3-2-8. 融合遺伝子の探索

TopHat-Fusion (56)を用いて、RNA-Seq データから融合遺伝子を探索した。用いたパラメータは、TopHat-Fusion のチュートリアルページ (URL : http://ccb.jhu.edu/software/tophat/fusion_tutorial.html)に記載してあるものを参考に使用した。得られた候補から、融合部位を含むタグが 10 以上、融合遺伝子の各遺伝子にそれぞれのリードがマップされたペアが 2 以上であった融合遺伝子を抽出した。また、表 11 のプライマーを用いて、RT-PCR で融合転写産物の確認を行った。

表 11 融合遺伝子の RT-PCR プライマー

Fusion transcript	Fwd primer	Rev primer
CCDC6-RET	GCAGCAAGAGAACAAGGTGC	ACCATCCTAAGTTGCTGGGC
EFHD1-UBR3	GGTCTTCAACCCCTACACGG	GCAATGCTGAGGTGGAGAGT
ERGIC2-CHRNA6	TCGCGATATTTCCGGGTACG	ACCACCATGGCCACGTATTT
ALK-PTPN3	GACAAGGATCCAGAGCCATC	AGAGTCCACGGAGTCGTCAT
EML4-ALK	CCGGCAGTCTCGATGATAGT	GCAACGTTAGGTGGGACAGT
ATAD2-ENSG00000253434	AAGATGGTGGTTCTCCGCAG	CGAGCCTACAGTGCCATAGG

(Supplementary Table S2B, Suzuki *et al.* 2014 *Nucleic Acids Research* 42(22):13557-13572 doi: 10.1093/nar/gku885 改変)

3-2-9. DNA メチル化解析

SureSelect XT Methyl-Seq のハイブリダイズキャプチャーにより、得られたバイサルファイトシークエンスタグのほとんどは参照ゲノム配列の逆鎖にマップされるため、まず、リード 1 を C から T、リード 2 を G から A に変換した。次に、G を A に変換した参照ゲノム配列を用意し、変換したシークエンスタグを BWA でマップした。得られたマッピング結果をもとに、変換前のタグをマップし、メチル化シトシン (CG、CA、CT、CC) と非メチル化シトシン (TG、TA、TT、TC) の数を数えた。CpG サイト以外の C サイトについて、CT 変換率を計算した。

5×以上の CpG サイトについて、DNA メチル化率 (各サイトの depth に対する C の割合) を算出した。全ての細胞株で 5×以上を満たす CpG サイトを用いて、階層的クラスタリングを行い、細胞株をグループ化した。また、CpG アイランド、CpG shore (CpG アイランドの外側 2 kb)、プロモーター領域 (転写開始点±1.5 kb) について、平均 DNA メチル化率を算出した。シークエンステンプレートに対して、ダイレクトサ

ンガーシーケンスや TA クローニング (pMD20-T, Takara) のサンガーシーケンスを行い、計 149 CpG サイトにおいて DNA メチル化率の確認を行った。用いた PCR プライマー配列を表 12 に示す。シーケンスプライマーは M13 (5'-GTAAAACGACGGCCAGT-3') を用いた。

また、19,323 遺伝子に対して、各遺伝子のプロモーター領域 (転写開始点±1.5 kb) の平均 DNA メチル化率を算出した。異常な DNA メチル化を抽出するために、各細胞株において、26 細胞株平均値に比して 4 倍以上もしくは 1/16 倍以下の DNA メチル化率 ("differential DNA methylation") を示す遺伝子を抽出した。

表 12 バイサルファイトシーケンステンプレートに対する PCR プライマー

Target region	Fwd primer	Rev primer
around TSS of NRAS	GAGGTAAGTTTATTTTATGT	CTCCTTCCCATTCTCCCTTC
around TSS of TTC23	GGTTTTTAAAGATTGGTTTG	CTTTATACAAATTAATAACC
around TSS of IGF1R	GGAAGTTTTGGTATTATGGA	ACTATATAAATTCTCTAATT
around TSS of ETF1	GTTTTAAGTTTAAAGTGAGT	CCTAACCAATAAATCTAAAT
around TSS of RPL21	AGGTTTTTATTTGGAAGATGG	CACAACTTTCTATATTCTC
around TSS of BRAF	ATTATAGAATAGGTTGAGGA	ATTTTTTCCCTTCTAACCAC
around TSS of CDKN2A (p16 ^{INK4A})	AGGTGGGTAGAGGGTTTGTA	ACCCCTTCTAAAAACTCCCC
around TSS of MDM2	GTATTTAAGGTAGGTTTATA	CCAACCAAACCAAAACATTC
genebody of ERAS	GGTGGTTTATATTTGTTGTA	CAAACCATCCATAAAACCTC
around TSS of MYC	AGGTTTTTAGAAGAGATAAAT	CTCTCCACTTACCCCTTTTA
around TSS of CATSPER2	GGATTGAGTGGGTAAAGAGA	TTAAACAAAATACCAAAACC
around TSS/genebody of NEDD4	GTATTTTGTAGTTAGAAGAA	ATACAATCTATCTAAAAATC
around TSS of ZNF609	GTTAATGATGAGATTTTTTA	CAAAAAAAAAACCTTAAATCT

(Supplementary Table S2C, Suzuki *et al.* 2014 *Nucleic Acids Research* 42(22):13557-13572 doi: 10.1093/nar/gku885 改変)

3-2-10. クロマチンステータス解析

8 種類の ChIP と WCE のシーケンスタグを、ELAND を用いて参照ゲノム配列にマップした。また、MACS2 (57,58) を用いて、ChIP-Seq の narrow ピークおよび broad ピークを抽出した。遺伝子とピークを対応させるために、転写開始点±1.5 kb にある H3K4me3、H3K27me3 および H3K9me3 のピークについて各遺伝子に帰属させた。また、エンハンサーマーク (H3K27ac、H3K4me1) においては、26 細胞株における narrow ピークを位置によってグループ化することで、エンハンサー領域を決定した。各遺伝子のエンハンサーは転写開始点上流 100 kb および転写領域内にあるものとし、2 個以上の遺伝子に対応するエンハンサーは近いほうの遺伝子に帰属させた。

クロマチンマーク間の相関を解析するために、20,598 遺伝子の転写開始点上流 1.5 kb

と転写領域における ChIP-Seq のシグナル強度 (ChIP PPM / WCE PPM) を算出した。各クロマチンマークペアにおけるスピアマン相関係数を計算し、クロマチンマーク間のシグナル強度の相関を解析した。

クロマチンステータスの異常を検出するために、シグナル強度が 26 細胞株いずれかで >1 PPM を示す遺伝子に対して、26 細胞株平均に比して 4 倍以上もしくは 1/16 倍以下の強度を示すクロマチンマークを”differential chromatin marks”として抽出した。H3K4me3、H3K9/14ac、H3K9me3、H3K27me3 および Pol II は転写開始点±1.5 kb、H3K36me3 は転写領域、H3K4me1 および H3K27ac は各エンハンサー領域 (転写開始点上流 100 kb 以内および転写領域内) におけるシグナル強度を用いた。得られたシグナル強度および differential chromatin marks は、いくつかの例について qPCR によって定量的・定性的に確認を行った (表 13、プライマー)。

26 細胞株の ChIP-Seq データから、ChromHMM (59)を用いて、ゲノム領域を 8 種類のクロマチンステータスに分けて記載した(60)。26 細胞株で学習したモデルを用いて、SAEC の ChIP-Seq データについてもクロマチンステータスの予測を行った。

表 13 ChIP-qPCR プライマー

Target region	ChIP	Fwd primer	Rev primer
around TSS of EGFR	H3K4me3, Pol II	CTCCTCCTCCTCTGCTCCTC	TGCCCGACGTCTAGCTC
genebody of EGFR	H3K36me3	TTTGCTGATTCAGGCTTGG	CTTGCTCCTGCCGCGTAT
around TSS of CDKN1A	H3K27me3	GGGAGGGTTCATGTCPTTGA	GAGGGAATCTGGGCTAGGTC
around TSS of CDKN1A	H3K9me3	TGCGTTCACAGGTGTTCTG	CACATCCCGACTCTCGTCAC
genebody of STK11 -1	H3K36me3	GACTGACCACGCCCTTCTTC	CCTTCCCGATGTTCTCAAAC
genebody of STK11 -2	H3K36me3	TGAGTGTGTAACGGCTTTGC	AGCTCCTGATGGCAGACAGT
around TSS of RET	H3K4me3	GCTGCTAGGCAAAGGTGAGT	ACAGAAAGGCGCTTCTGAAC
genebody of RET	H3K36me3	CAGACAAATCCCTCCCTCA	GTTCTGAGGGCCAATGTGTT
around TSS of ERBB2	H3K9/14ac	CGAAGTTGTGGACAGTCGAG	AAGCGCCAGTCTCTGGAT
around TSS of NRAS	H3K27ac	TTTCTGGCATCAGTGAAAACG	AGGCAGTGGAGCTTGAGGTA
around TSS of MYC	H3K27ac	TTCCGGTAGTGAAAACCCAG	GCATTCGACTCATCTCAGCA
MYC enhancer (upstream of TSS)	H3K27ac, H3K4me1	TTCAGGAACCCATCAAGAG	TCTGTGGAGGCTAAGGTTGG
RPL2 enhancer (upstream of TSS)	H3K4me1	AACCATATGCATGGGGAAGA	TACGTCAGCTGGCATTGTGTC
around TSS of KIT	H3K4me3	AAAGTTGCGTGTGTGTGACC	AGTAGCCGCAGAGCAGAGAA
around TSS of RAB25	H3K27ac	AGACACCAACCTGTGGAAC	ACGGAAGCTGAGAACAGGAA
around TSS of WNT10A	H3K4me3	GGGGTAGAGCCCTGAGACAC	GCCCCACTGCCTGTAGAGTA
around TSS of REST	H3K9me3	TAGACGTGAGCCCGTTACT	GGGGAGGAATTACCCTCAGA
around TSS of SMARCA2	H3K27me3	GTCTCACAAAACCCGACCT	GCGGTTTTGAGACCTCTTTG
around TSS of MYCN	H3K27ac	TCTGCTCCGAAAACAAAACC	ATTAGGGCGGGTCTCCTCTA
around TSS of LAMA5	H3K27me3	GGACTTTAATTCGCTTCC	GACTCTCTGGTTCGCAATT
around TSS of KDR	H3K4me3	CCAGCTCTACGATTCGAGT	GTGGGTAAGGAGCCCACTCT
around TSS of ZNF677	H3K4me3, Pol II	CGAGAGGAATGAGCAGGTTT	AGTCTCCGCTTTTCGGGTTT
around TSS of TTC18	H3K9/14ac	GTCATCTGCTCGGAAGAAGC	CTTGTCGCCAGACCAGACTC
around TSS of PLAU	H3K4me3	AGGACCTCTGATCAGACCA	ATTTCTGGGGAGAGTCTGG

(Supplementary Table S2D, Suzuki *et al.* 2014 *Nucleic Acids Research* 42(22):13557-13572 doi: 10.1093/nar/gku885 改変)

3-2-11. マルチオミクス“hallmarks of cancer”解析

ゲノム、エピゲノムおよびトランスクリプトームの異常と、癌細胞における表現型および生物学的意義との関連を明らかにするために、Hanahan と Weinberg が提唱した“hallmarks of cancer” (61)に基づいた解析を行った。まず、10 個の hallmark に対して、Hnisz *et al.* (62)に記載のあった Gene Ontology (GO) を用いて 2050 遺伝子を割り当てた (表 14)。さらに少なくとも 1 細胞株で >1 RPKM の発現量を示す 1840 遺伝子を選出し、各細胞株の異常 (ゲノム変異、differential DNA methylation、differential chromatin marks, differential expression を示す遺伝子の割合) を 10 個の hallmark について記載した。

また、26 細胞株について癌細胞としての特徴を見出すために、正常細胞コントロールとして SAEC の RNA-Seq および ChIP-Seq のデータを用いた。SAEC と比して 4 倍以上もしくは 1/16 倍以下の発現量や ChIP シグナル強度を示す癌細胞株が一つでもある遺伝子を、癌細胞株における異常として抽出し、10 個の hallmark について記載した。

表 14 Hallmarks of cancer に割り当てた GO term と遺伝子数

Hallmarks of cancer	GO term	遺伝子数
Inducing Angiogenesis	GO:0001525 angiogenesis	207
Enabling Replicative Immortality	GO:0032202 telomere assembly ; GO:0000723 telomere maintenance ; GO:0090398 cellular senescence ; GO:0090399 replicative senescence	79
Activating Invasion and Metastasis	GO:0045216 cell-cell junction organization ; GO:0034329 cell junction assembly ; GO:0045217 cell-cell junction maintenance ; GO:0034334 adherens junction maintenance ; GO:0016477 cell migration ; GO:0010718 positive regulation of epithelial to mesenchymal transition ; GO:0007155 cell adhesion	620
Genome Instability and Mutation	GO:0006281 DNA repair ; GO:0051383 kinetochore organization ; GO:0007062 sister chromatid cohesion ; GO:0000819 sister chromatid segregation ; GO:0051988 regulation of attachment of spindle microtubules to kinetochore ; GO:0030997 regulation of centriole-centriole cohesion ; GO:0046605 regulation of centrosome cycle ; GO:0060236 regulation of mitotic spindle organization ; GO:0090169 regulation of spindle assembly ; GO:0043146 spindle stabilization ; GO:0031577 spindle checkpoint	302
Resisting Cell Death	GO:0060548 negative regulation of cell death ; GO:0012501 programmed cell death ; GO:0010941 regulation of cell death	63
Deregulating Cellular Energetics	GO:0006091 generation of precursor metabolites and energy	57
Sustaining Proliferative Signaling	GO:0007166 cell surface receptor signaling pathway ; GO:0070848 response to growth factor	174
Tumor-Promoting Inflammation	GO:0006954 inflammatory response ; GO:0045321 leukocyte activation	302
Avoiding Immune Destruction	GO:0002507 tolerance induction ; GO:0001910 regulation of leukocyte mediated cytotoxicity ; GO:0019882 antigen processing and presentation ; GO:0002767 immune response-inhibiting cell surface receptor signaling pathway	38
Evading Growth Suppressors	GO:0007049 cell cycle ; GO:0008283 cell proliferation	518

(Supplementary Table S5A, Suzuki *et al.* 2014 *Nucleic Acids Research* 42(22):13557-13572 doi: 10.1093/nar/gku885 改変)

3-2-12. シークエンスの登録とデータベース

取得した全ゲノムシークエンス (DRA001858、DRA001859)、RNA-Seq (DRA001846、DRA002311)、バイサルファイトシークエンス (DRA001841)、および、ChIP-Seq (DRA001860、DRA002311) は、DNA Data Bank of Japan (DDBJ) から公開されている。また、解析したマルチオミクスデータは、東京大学医科学研究所ヒトゲノム解析センターのスーパーコンピュータに構築されたデータベース DBTSS (<http://dbtss.hgc.jp/>) (63)から公開されている。

3-3. 結果と考察

3-3-1. 肺腺癌細胞株のゲノム変異

26 種類の肺腺癌細胞株の全ゲノムシーケンシングより得られたシーケンスタグのうち、細胞株あたり平均 1,025,803,752 タグ（リード 1+リード 2）がヒト参照ゲノム配列にマップされた。マップされたシーケンスタグは平均 **sequence depth** が 33 × であり、全ゲノムの 91% の領域が ×5 以上の **sequence depth** でカバーされていた（図 11 および表 15）。

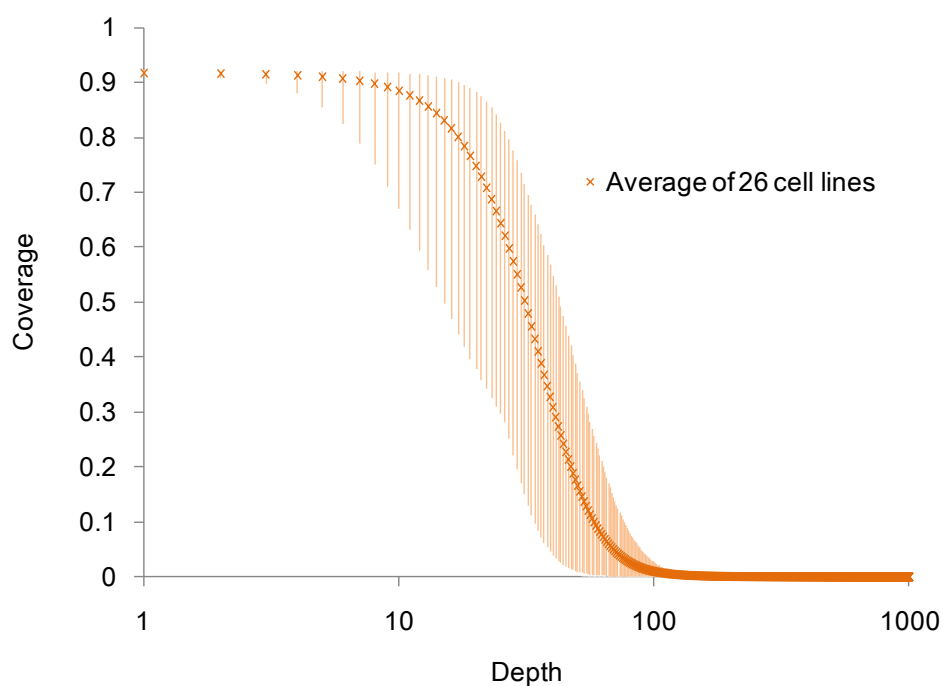


図 11 全ゲノムシーケンシングの **depth** と **coverage**

ヒト参照ゲノム配列に対する全ゲノムシーケンシングデータの平均 **depth** と **coverage** を示している。Depth および coverage は BEDTools (64) と Perl スクリプトを用いて算出した。グラフ内の × は 26 細胞株の平均値を示す。

表 15 各細胞株の全ゲノムシーケンスタグ数および depth と coverage

Cell line	Mapped sequences (read 1 + 2)	Avg. depth	Coverage (× 5)
PC-3	1,119,060,882	36.4	0.91
PC-7	1,181,752,959	38.4	0.92
PC-9	1,235,410,075	40.2	0.91
PC-14	1,377,953,696	44.8	0.91
RERF-LC-Ad1	1,189,406,566	38.7	0.91
RERF-LC-Ad2	1,204,958,194	39.2	0.92
RERF-LC-KJ	1,058,371,222	34.4	0.92
RERF-LC-MS	1,234,238,416	40.2	0.92
RERF-LC-OK	677,038,144	21.8	0.91
VMRC-LCD	1,270,060,339	41.3	0.91
ABC-1	1,131,071,585	36.8	0.91
LC2/ad	1,265,338,449	41.2	0.91
II-18	860,643,037	27.6	0.90
A549	723,563,287	22.2	0.86
A427	1,040,002,036	33.8	0.91
H322	893,332,828	28.9	0.90
H2228	830,781,519	27.0	0.91
H1299	899,909,551	29.3	0.91
H1437	711,909,693	23.1	0.91
H1648	1,065,042,096	34.6	0.92
H1650	1,031,008,238	33.5	0.91
H1703	984,465,974	31.6	0.91
H1819	1,091,791,039	35.5	0.91
H1975	1,004,161,315	32.6	0.91
H2126	640,653,382	20.8	0.91
H2347	948,973,026	30.9	0.91

*BEDTools を用いて計算した(64)。

全ゲノム解析より見出された SNVs および indels の数を表 16 に示す。細胞株あたり検出された 3,302,407 個の SNVs と 453,821 個の indels から、公共データベース等に登録されている germline variants を除去した結果、平均 125,234 個の SNVs と 23,975 個の indels が抽出された。これら体細胞突然変異候補には、公共データを用いても除去できなかった集団頻度の低い germline variants が残存していると考えられた。同一検体の B リンパ芽球由来の細胞株が入手可能であった 3 株 (H1437、H2126、H2347) を全ゲノム解析に供し、混在する germline variants の割合を算出した。これらについては、体細胞突然変異候補の約 28%が germline variants であることが確認された。

SNVs の塩基置換パターンを図 12 に示す。Germline variants の塩基置換パターンは全ての細胞株で同じ傾向が見られた。一方、体細胞突然変異候補は、C>T/G>A 変異と G>T/C>A 変異の頻度が細胞株によって異なっていた。G>T 変異と喫煙の関連性はすでに報告されており(65)、細胞株のもととなる検体の喫煙歴等がこの頻度の違いを生んでいることが示唆された。

体細胞突然変異候補のうち、平均 710 個が転写開始点上流 500 塩基以内に、1,087 個が 5' または 3' 非翻訳領域内に、724 個がタンパク質コード領域中に (うち 536 個がアミノ酸変異を伴う ; 図 13)、17 個がスプライスサイトに位置していた (表 16)。

コピー数解析の結果を表 17 に示す。平均で 143 個の遺伝子領域がコピー数 4 以上、101 個の遺伝子領域が 1 以下であり、該当する領域での遺伝子増幅、欠失が示唆された。例えば、LC2/ad では v-myc avian myelocytomatosis viral oncogene homolog (MYC) 遺伝子領域の大幅なコピー数増加が見られた。また、VMRC-LCD では v-myc avian myelocytomatosis viral oncogene neuroblastoma derived homolog (MYCN)、H1703 では EGFR、fibroblast growth factor receptor 1 (FGFR1) および platelet-derived growth factor receptor, alpha polypeptide (PDGFRA) の遺伝子領域でコピー数増加が見られた。各細胞株におけるコピー数解析結果を図 14 に示す。

これらの変異に加えて合計で 552 個の染色体構造異常が検出された (表 17)。LC2/ad では、coiled-coil domain containing 6 (CCDC6) と RET のイントロン領域における既知の染色体構造異常が検出された。肺腺癌ではこれらの染色体構造異常によって生成する融合遺伝子 (ALK、RET および ROS1 の融合遺伝子) がドライバー変異として機能することが報告されている(66-70)。肺腺癌細胞株における融合遺伝子については、3-3-2. トランスクリプトームの異常の章で後述する。

表 16 26 細胞株における SNVs と indels の数

	Total number of positions (Avg. of 26 cell lines)	
	SNVs	Indels
Total	12,732,271 (3,302,407)	1,916,622 (453,821)
Germline	10,010,429 (3,177,173)	1,597,810 (429,846)
Somatic candidates	2,721,842 (125,234)	318,812 (23,975)
Genic *	892,941 (39,695)	118,268 (8,516)
Upstream (-500 from TSS)	11,796 (551)	2,049 (159)
UTRs	24,902 (1,086)	13 (0.8)
CDS	16,354 (687)	573 (37)
Synonymous	4,505 (188)	***
Non-synonymous	11,849 (499)	***
Splice sites †	346 (14)	39 (3)
Intronic and others	839,543 (37,357)	115,594 (8,315)
Intergenic	1,828,901 (85,539)	200,544 (15,459)

*19,958 遺伝子を解析に用いた。

†イントロン領域の最初と最後の 2 塩基。

(Table 1A, Suzuki *et al.* 2014 *Nucleic Acids Research* 42(22):13557-13572 doi: 10.1093/nar/gku885)

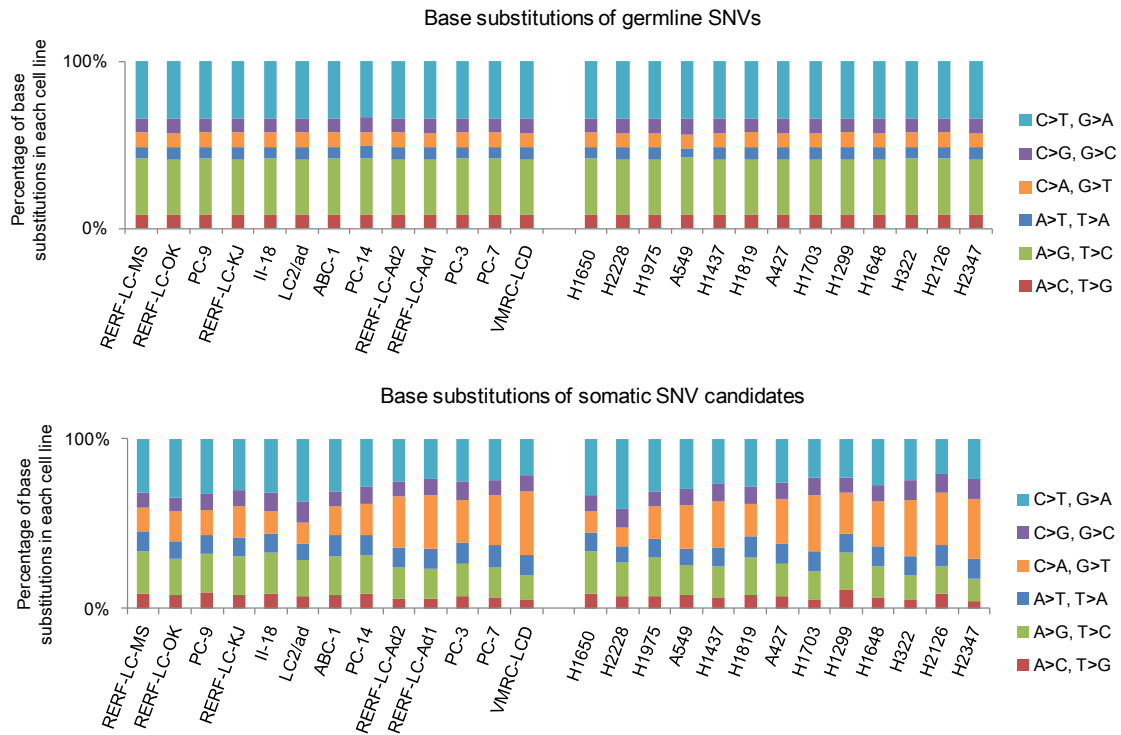


図 12 SNVs における塩基置換パターン

各細胞株の SNVs における塩基置換パターンの割合を示している。上図は germline variants、下図は体細胞突然変異候補の塩基置換パターンを示している。横軸の細胞株は図 13（後述）と同じ順番である。

(Supplementary Figure S7, Suzuki *et al.* 2014 *Nucleic Acids Research* 42(22):13557-13572 doi: 10.1093/nar/gku885)

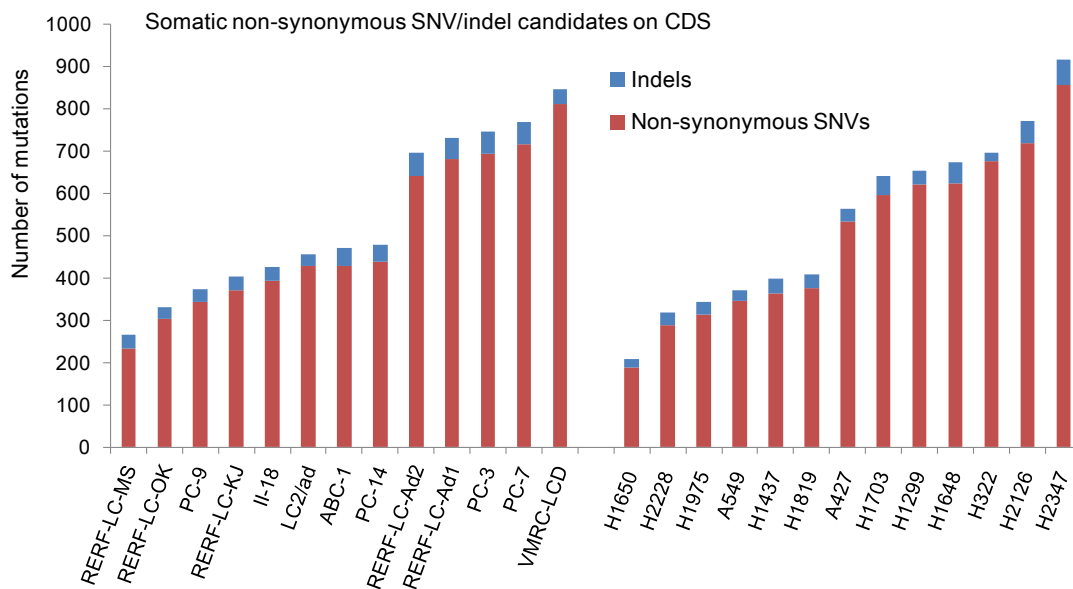


図 13 タンパク質コード領域におけるゲノム変異数

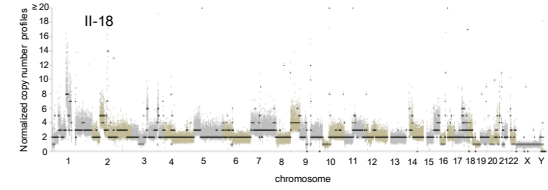
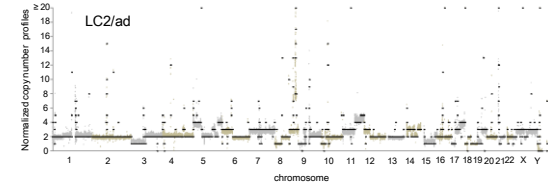
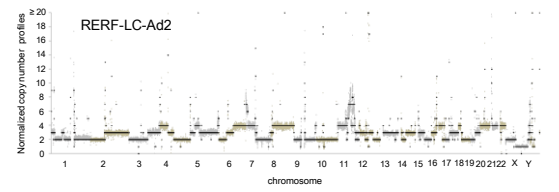
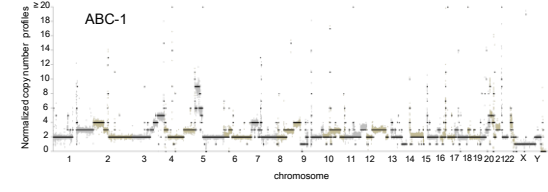
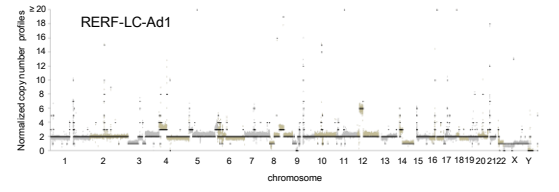
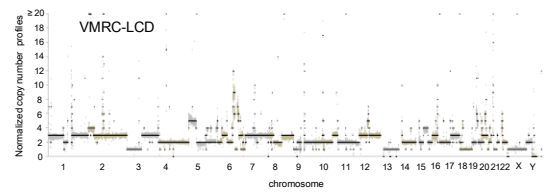
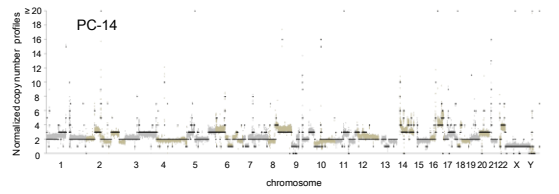
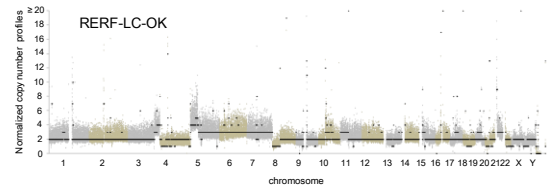
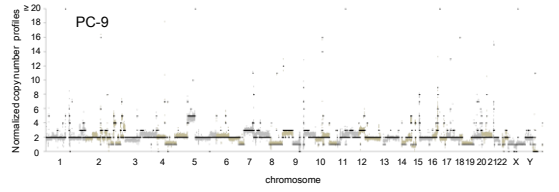
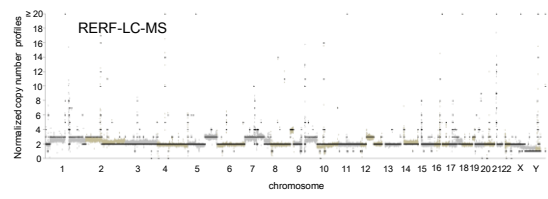
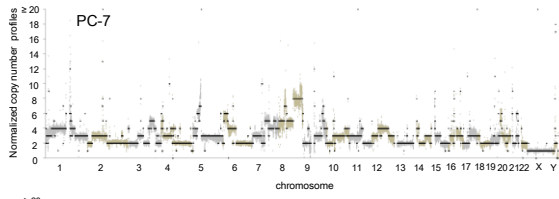
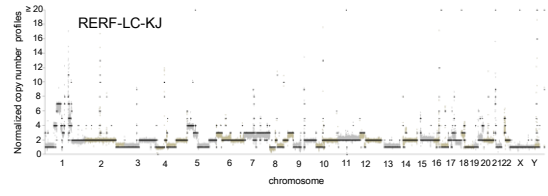
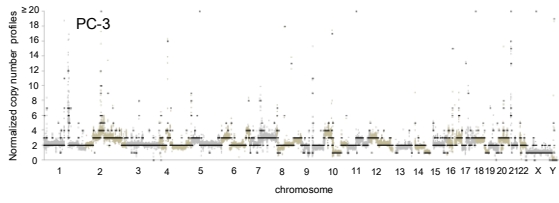
各細胞株のタンパク質コード領域における non-synonymous SNVs および indels の数を示している。横軸は細胞株を示しており、左は日本人由来、右は非アジア人由来である。横軸の細胞株の順番はそれぞれ変異数で並べられている。

(Figure 1A, Suzuki *et al.* 2014 *Nucleic Acids Research* 42(22):13557-13572 doi: 10.1093/nar/gku885 改変)

表 17 コピー数異常および染色体構造異常の数

Cell line	Copy number aberration				Number of rearrangements (gene region)
	≥ 4		≤ 1		
	Total	Gene region	Total	Gene region	
PC-3	316	133	148	95	4
PC-7	366	202	133	79	11
PC-9	265	107	187	127	1
PC-14	326	124	198	120	7
RERF-LC-Ad1	332	102	175	104	1
RERF-LC-Ad2	326	170	81	52	13
RERF-LC-KJ	294	126	198	129	2
RERF-LC-MS	343	102	140	76	3
RERF-LC-OK	345	253	117	82	91
VMRC-LCD	336	158	188	133	5
LC2/ad	310	142	146	94	6
ABC-1	274	92	174	100	4
II-18	364	284	114	81	130
A549	168	79	140	90	74
A427	311	119	210	133	9
H322	344	163	159	103	9
H2228	325	86	243	153	11
H1299	303	109	204	118	6
H1437	346	191	149	105	8
H1648	290	126	139	86	4
H1650	315	113	237	132	3
H1703	274	220	52	33	128
H1819	435	199	217	114	9
H1975	348	124	139	74	5
H2126	325	132	199	115	4
H2347	312	85	201	117	4

(Supplementary Table 7 および 8, Suzuki *et al.* 2014 *Nucleic Acids Research* 42(22):13557-13572 doi: 10.1093/nar/gku885 改変)



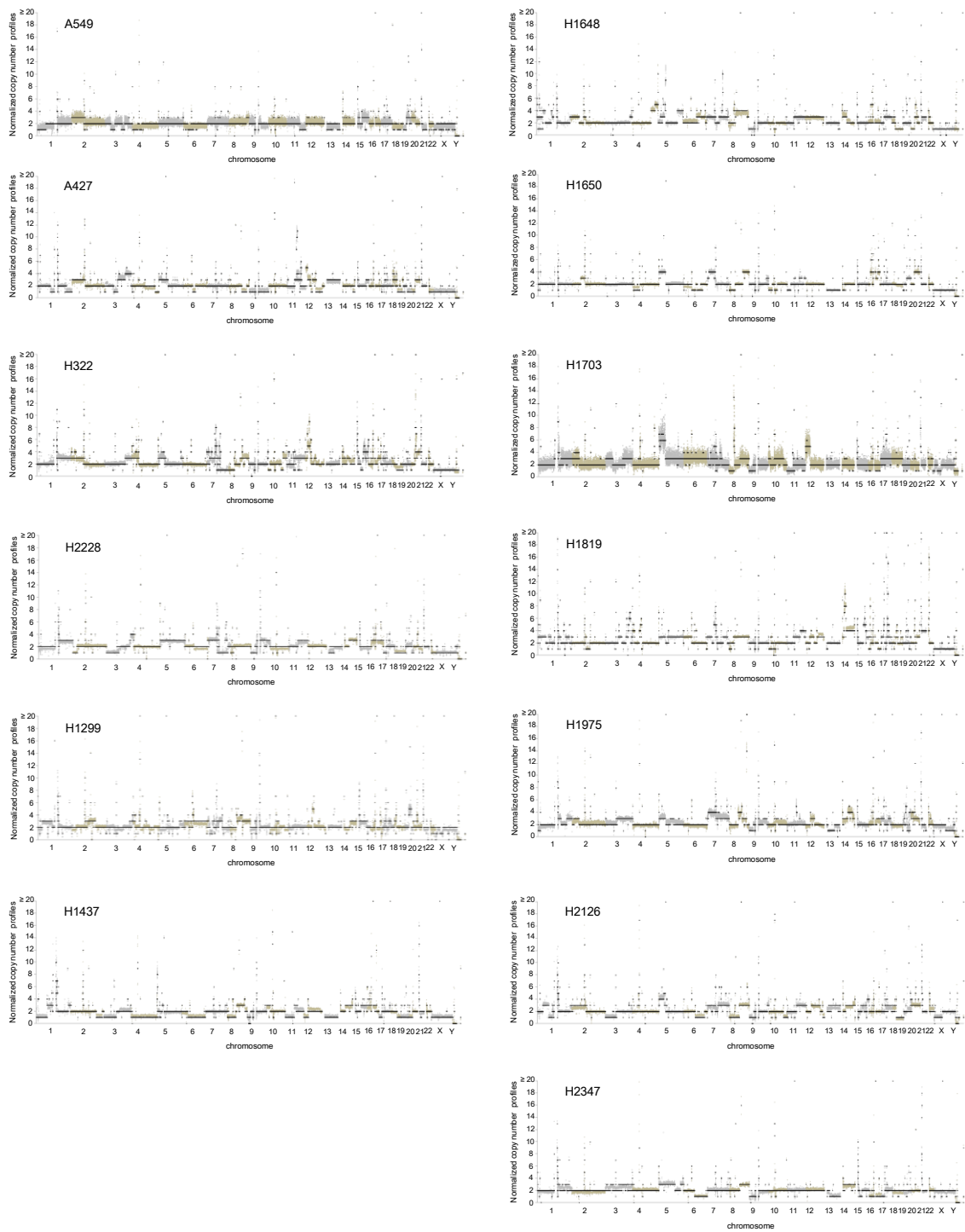


図 14 各細胞株におけるコピー数の状態

ゲノム領域に対して、Control-FREEC (50,51)を用いて 50 kb ごとに算出した標準化コピー数を各細胞株で図示した。コピー数は 2 に標準化されている。

(Figure 1B, Suzuki *et al.* 2014 *Nucleic Acids Research* 42(22):13557-13572 doi: 10.1093/nar/gku885 を含む)

検出したゲノム変異に対して、特に既知の癌関連遺伝子について解析を行った。文献に基づき、9個の癌遺伝子、11個の癌抑制遺伝子、3個のヒストン修飾・クロマチンリモデリング関連遺伝子、3個の融合遺伝子を含む計26癌関連遺伝子を選出した(8,11,71)。各細胞株におけるタンパク質コード領域およびスプライスサイトの変異やコピー数異常を図15に示す。ドライバー変異に関しては、EGFRについて、既知の gefitinib 感受性変異 (L858R および E746_A750del 変異) が4株 (PC-9、II-18、H1650、H1975) において検出された(9)。そのうち、H1975 はさらに gefitinib 耐性獲得変異である T790M 変異を有していた。KRAS について G12 変異が4株 (RERF-LC-Ad1、RERF-LC-Ad2、A549、A427)、neuroblastoma RAS viral (v-ras) oncogene homolog (NRAS) について Q61 変異が3株 (PC-14、H1299、H2347) において検出された(72)。また、MYC について3株 (PC-7、LC2/ad、H1975)、ERBB2 について1株 (H1819)、MET proto-oncogene, receptor tyrosine kinase (MET) について1株 (H1648) において、顕著なゲノム増幅が見られた。癌抑制遺伝子変異に関しては、TP53 について14株でミスセンス変異、5株でナンセンスおよびフレームシフト変異、3株 (H1650、H1703、H1819) でスプライスサイト変異、1株 (H1299) でゲノム領域の欠損が見られた(73)。Cyclin-dependent kinase inhibitor 2A (CDKN2A) については、13株でゲノム欠失が見られた(74-76)。癌抑制遺伝子においては、スプライスサイト変異も多く認められた。これらは、遺伝子機能解析にしばし利用されるこれら代表的な細胞株において、それぞれの有する癌関連遺伝子の変異パターンを網羅的に記載したものとして意義深いと考えている。

ヒストン修飾・クロマチンリモデリング関連遺伝子にもゲノム変異が多数検出された(77,78)。クロマチンリモデリング関連遺伝子である SMARCA4 について、スプライスサイト変異を含む SNVs および indels が5株、1 kb 以上のゲノム欠失が5株で検出された。これらゲノム変異が、癌細胞におけるエピゲノム、トランスクリプトームステータスに及ぼす影響を明らかにするために、以下に同一細胞株を用いたトランスクリプトームおよびエピゲノム解析を行った。

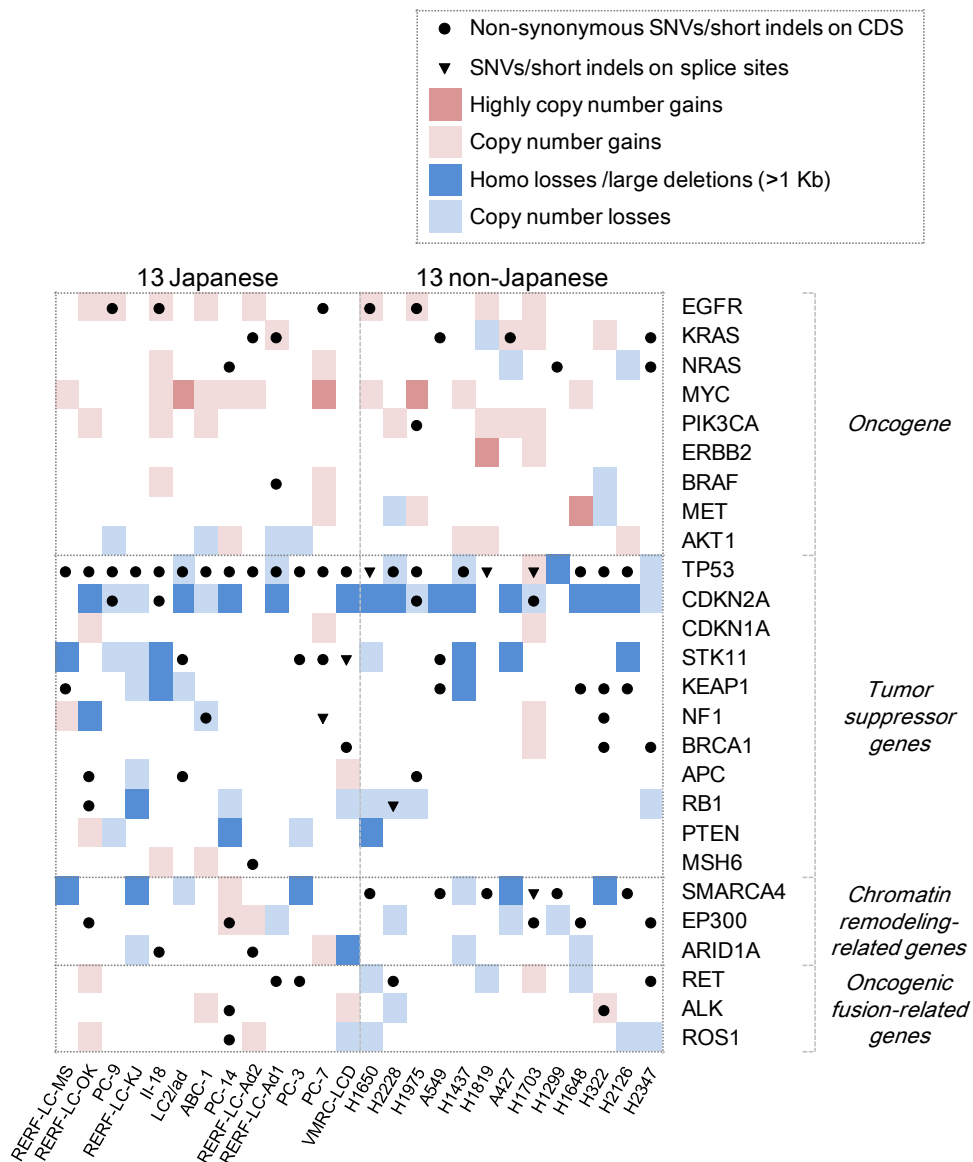


図 15 26 癌関連遺伝子のゲノム変異

26 細胞株において、26 種類の癌関連遺伝子のゲノム変異（タンパク質コード領域の non-synonymous 変異、スプライスサイト変異およびコピー数異常）について図示している。コピー数増加は標準化コピー数 4 以上、コピー数減少は 1 以下を示しており、コピー数 8 以上を特に高いコピー数増加、コピー数 0 をゲノム欠失として濃い色で示している。TP53 における 2 個の変異は、dbSNP に登録されていたため除去されていたが、体細胞突然変異として戻した。

(Figure 1D, Suzuki *et al.* 2014 *Nucleic Acids Research* 42(22):13557-13572 doi: 10.1093/nar/gku885)

3-3-2. トランスクリプトームの異常

RNA-Seq を行い、26 細胞株について細胞株あたり平均 135,469,256 タグ (67,734,628 ペア) を得た。参照ゲノム配列にマップされた平均 48,663,351 タグ (リード 1 のみ) を遺伝子発現解析に用いた結果、平均 12,290 遺伝子について >1 RPKM の発現量を示していた (表 18)。52 遺伝子について qRT-PCR を行い、検出された遺伝子発現量の妥当性を検証した。合計 1352 (52 遺伝子×26 細胞株) アッセイについて、それぞれ 3 回実験を行い、qRT-PCR による遺伝子発現量に関するデータを得た。これらを用いて RPKM 値と qRT-PCR との相関を求めたところ、ピアソンの相関係数 $r=0.89$ の相関が確認された (図 16)。

得られた RNA-Seq データを、全ゲノム解析で得られたゲノム変異と比較検討した。ゲノム変異が検出された遺伝子について、その発現量を検証したところ、タンパク質コード領域において SNVs および indels を有する遺伝子の約半数が、有意な遺伝子発現 (>1 RPKM) を示した (図 17)。機能的なドライバー変異の検出を試みるには、これら発現遺伝子上の変異に検証を絞ることが可能であると考えられた。一方で、発現が確認されない遺伝子上の変異については、非発現遺伝子上に生起する機能的に中立なパッシブ変異か、発現低下を惹起する変異であることが考えられるが、これらについても他の細胞株での発現パターンを参照することで区分することが可能であると考えている。

全ゲノム解析で検出されたゲノム変異が、転写産物にどのような影響を及ぼすか解析を行った。染色体構造異常による融合遺伝子の探索を行い、RNA-Seq データから 135 個の融合遺伝子候補を得た。LC2/ad における CCDC6-RET (79,80) および H2228 における ALK-PTPN3 (81) といった既知の融合遺伝子も含まれていた。しかし、H2228 における癌関連融合遺伝子 EML4-ALK (81) は、本論文で用いた計算機的手法によって抽出できなかった。これについては発現量が低いために RNA-Seq による検出ができなかったと考えている。7 個の融合遺伝子については、RT-PCR のよる転写産物の確認を行った (図 18)。これらについて転写産物が確認できたが、今後さらなる機能解析が必要であると考えられた。

スプライスサイト変異について解析を行った。癌抑制遺伝子を中心に、スプライスサイト変異が RNA スプライシングパターンに影響を及ぼしていることが、RNA-Seq 解析結果として直接検出される例が数多く認められた。(図 19)。PC-7 では、neurofibromin 1 (NF1) のイントロン 19 にスプライスサイト変異 (GT>TT) が抽出され、RNA-Seq 解析結果から、実際にエクソン 19 がスキップされていた (図 19A)。同様に TP53 にスプライスサイト変異を持つ 3 細胞株 (H1650、H1703 および H1819) では、TP53 におけるスプライシング異常が確認された (図 19D、19E および 19F)。また、H2347 における RBM10 変異(11)や VMRC-LCD における UPF1 regulator of

nonsense transcripts homolog (UPF1) 変異(82)など、RNAのスプライシングや分解に関わる遺伝子自体にもスプライスサイト変異が見られ、これらの転写産物はスプライシング異常を起こしていた。これらの細胞株において、遺伝子全体としてどのようなスプライシングの異常を生じているのか解析することは興味深い。全ゲノムシーケンスと RNA-Seq データを併用することで、ゲノム変異による異常な転写産物を抽出することが可能であることを実証することができたと考えている。

表 18 RNA-Seq のタグ数と発現量ごとの遺伝子数

Cell line	Used sequences (Read1)	Num of genes	
		> 1 RPKM	> 5 RPKM
PC-3	49,914,547	12,205	9,240
PC-7	50,925,975	12,129	9,009
PC-9	34,167,521	12,817	9,532
PC-14	53,977,381	12,169	9,037
RERF-LC-Ad1	56,406,046	12,298	9,206
RERF-LC-Ad2	45,580,359	12,392	8,804
RERF-LC-KJ	60,803,665	12,054	8,938
RERF-LC-MS	52,715,099	13,045	9,090
RERF-LC-OK	33,086,988	12,309	8,954
VMRC-LCD	45,944,953	12,502	8,711
ABC-1	37,993,504	11,715	8,384
LC2/ad	43,665,988	12,366	9,206
II-18	63,869,445	11,955	9,038
A549	20,440,396	12,155	8,998
A427	41,895,881	11,866	9,011
H322	54,487,583	12,457	9,351
H2228	56,465,940	12,409	9,106
H1299	51,120,991	11,735	8,958
H1437	49,890,034	12,275	8,921
H1648	38,908,100	12,604	9,317
H1650	26,635,691	12,716	9,595
H1703	87,705,180	11,736	8,695
H1819	75,262,673	12,494	9,185
H1975	36,195,247	12,715	9,634
H2126	46,862,796	12,143	9,016
H2347	50,325,156	12,278	9,030
SAEC	180,054,144	12,126	8,809

(Supplementary Table S10, Suzuki *et al.* 2014 *Nucleic Acids Research* 42(22):13557-13572 doi: 10.1093/nar/gku885)

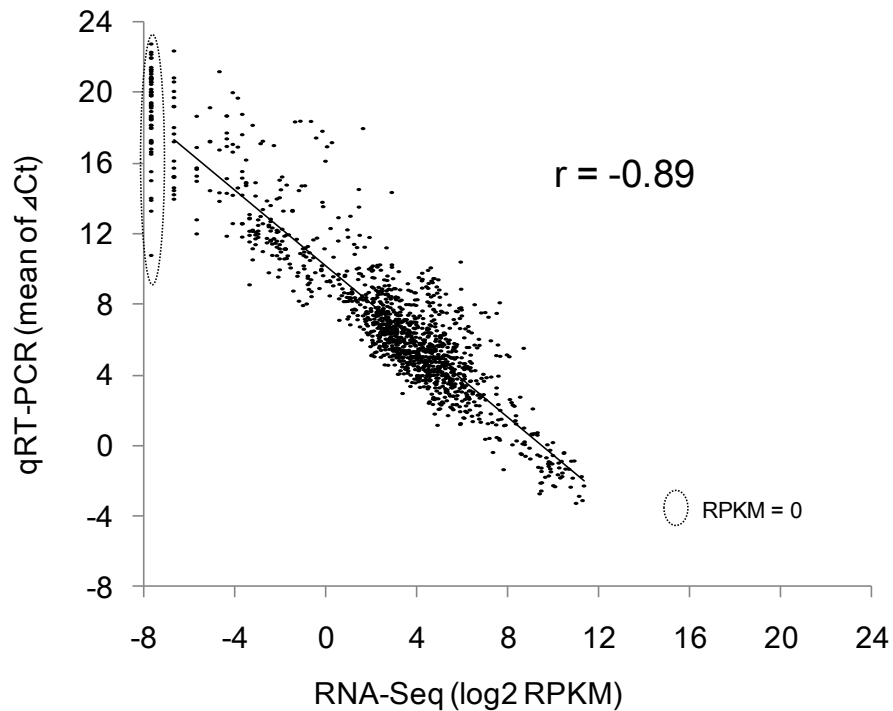


図 16 52 遺伝子における RNA-Seq と qRT-PCR 結果の比較

26 細胞株における 52 遺伝子の発現量 (1352 プロット) を示した。52 遺伝子は、癌関連遺伝子、differentially expressed genes および Housekeeping 遺伝子から選出した。横軸は RNA-Seq データから算出した RPKM、縦軸は qRT-PCR の ΔCt 値 ($Ct - Ct_{GAPDH}$, $n=3$) を示している。Ct 値が "Undetermined" の場合は 40 とし、RPKM が 0 であった 54 プロットについては相関係数の計算に使用していない。

(Supplementary Figure S2, Suzuki *et al.* 2014 *Nucleic Acids Research* 42(22):13557-13572 doi: 10.1093/nar/gku885)

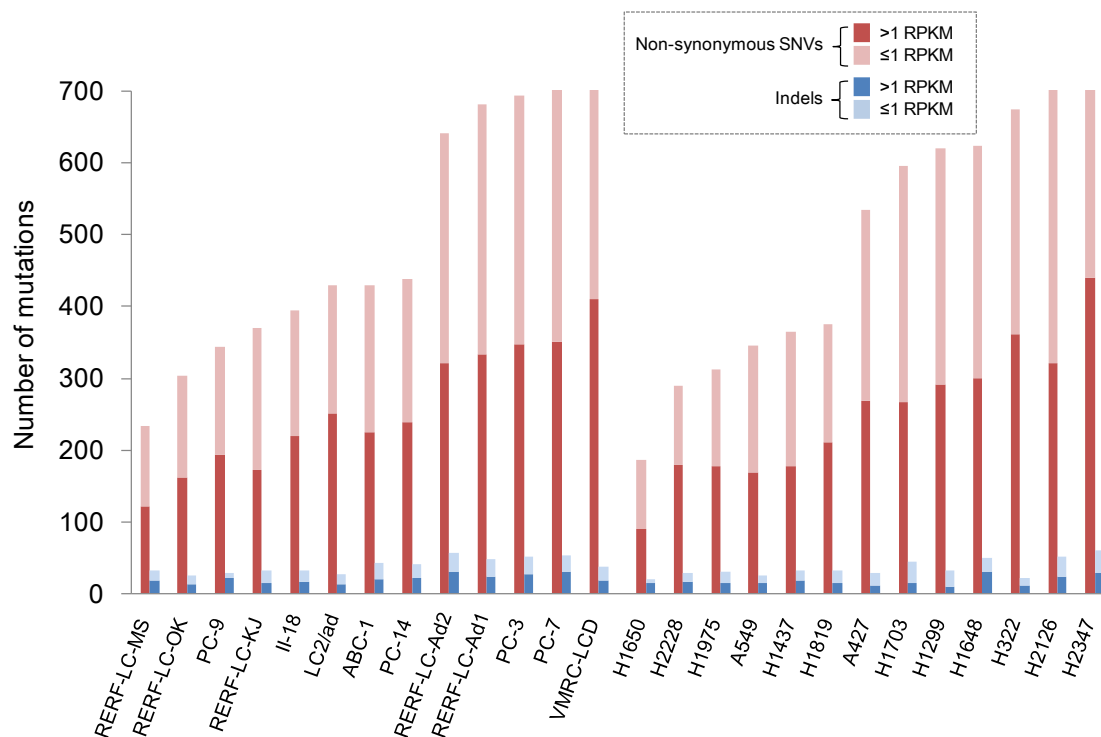


図 17 遺伝子発現量と SNVs および indels の数

各細胞株で検出されたタンパク質コード領域の non-synonymous SNVs および indels について、mRNA が発現している遺伝子 (>1 RPKM) の変異と、発現していない遺伝子 (≤1 RPKM) の変異に分けて、それぞれの変異数を図示した。どの細胞株も、約半数の変異が発現している遺伝子上に引き起こされていた。横軸の細胞株の順番は図 13 と同じである。

(Figure 2A, Suzuki *et al.* 2014 *Nucleic Acids Research* 42(22):13557-13572 doi: 10.1093/nar/gku885)

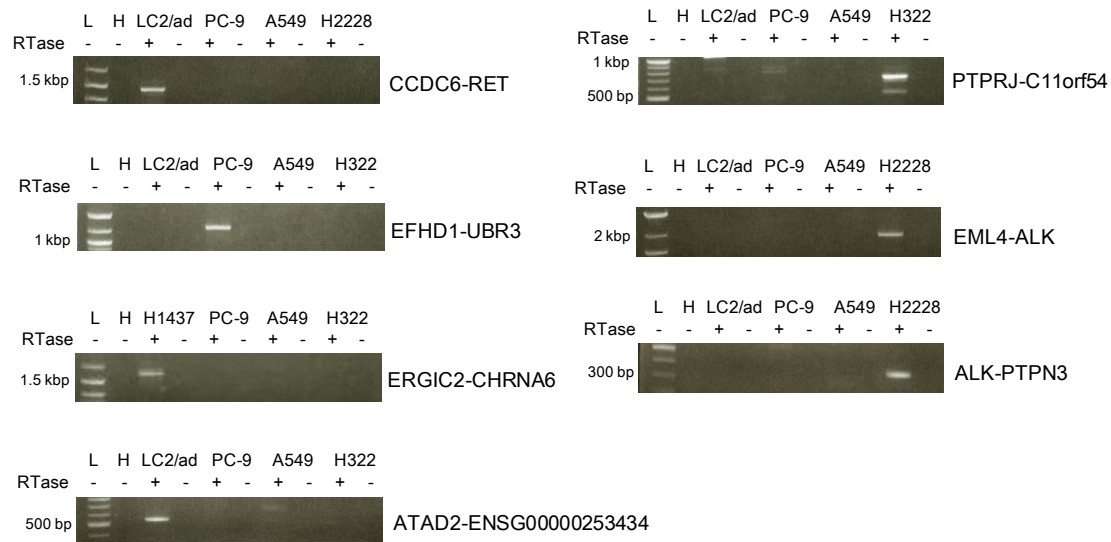


図 18 融合遺伝子の RT-PCR

7 個の融合遺伝子候補 (LC2/ad の CCDC6-RET および ATAD2-ENSG00000253434、PC-9 の EFHD1-UBR3、H1437 の ERGIC2-CHRNA6、H322 の PTPRJ-C11orf54、H2228 の EML4-ALK および ALK-PTPN3) について、RT-PCR で転写産物の確認を行った。ENSG00000253434 は非コード RNA である。EML4-ALK は計算機的に検出されなかったが、RT-PCR で転写産物が確認された。L はラダー、H は H₂O (ブランクレーン) を示す。ラダーは 2-Log DNA Ladder (N3200S、BioLabs) を使用した。RT-PCR は 40 サイクルで行った。

(Figure 2C および Supplementary Figure S13, Suzuki *et al.* 2014 *Nucleic Acids Research* 42(22):13557-13572 doi: 10.1093/nar/gku885 改変)

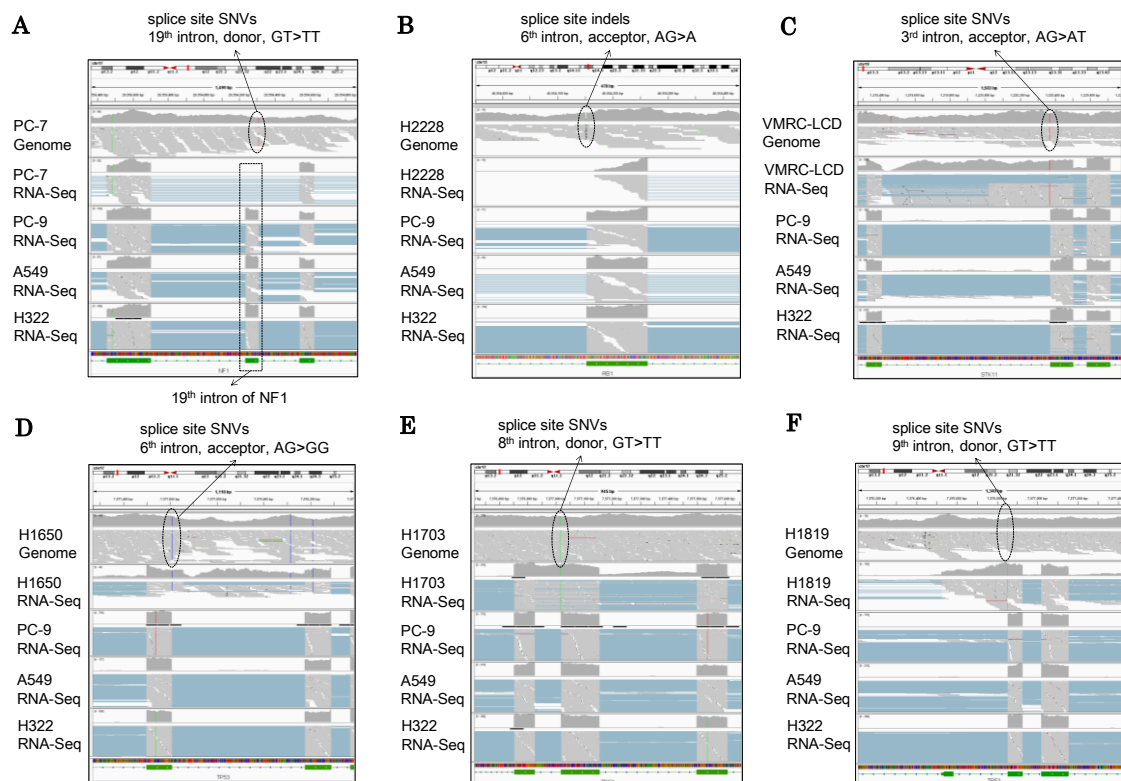


図 19 スプライスサイト変異とスプライシング異常

癌抑制遺伝子で見られたスプライスサイト変異と RNA スプライシング異常の例を示す。

(A) PC-7 の NF1 変異、(B) H2228 の retinoblastoma 1 (RB1) 変異、(C) VMRC-LCD の serine/threonine kinase 11 (STK11) 変異、(D - F) H1650、H1703 および H1819 の TP53 変異について示しており、スプライスサイト変異近傍の全ゲノムおよび RNA-Seq データを、IGV を用いて可視化している。また、PC-9、A549 および H322 の RNA-Seq データを正常なスプライシングのコントロールとして載せている。

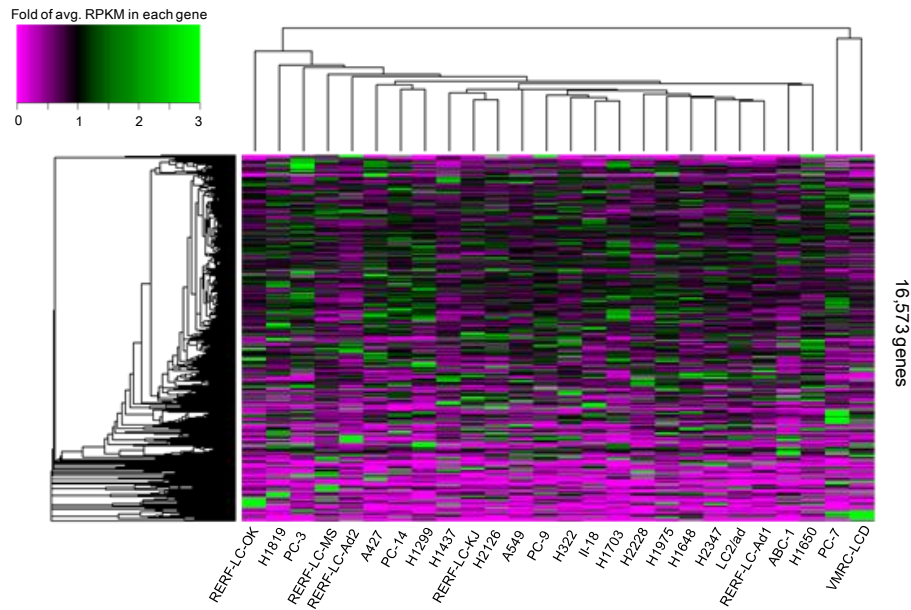
(Figure 2B および Supplementary Figure S12, Suzuki *et al.* 2014 *Nucleic Acids Research* 42(22):13557-13572 doi: 10.1093/nar/gku885 改変)

26 細胞株における遺伝子発現パターンを網羅的に解析するために、階層的クラスタリングを行った (図 20A)。PC-7 と VMRC-LCD が外側にクラスタリングされた。PC-7 は唯一の浮遊細胞である。また、この 2 株はともに日本人由来の細胞株であり、EGFR および KRAS 変異陰性である。これらの細胞株については、特にゲノム変異が多く蓄積している。これら他の細胞株と異なったゲノム変異上の特徴が遺伝子発現パターンに反映されているとすれば興味深い。さらに正常細胞 (SAEC) を加えて同様に階層的クラスタリングを行った (図 20B)。SAEC は PC-7 や VMRC-LCD および RERF-LC-OK の 3 株とその他の 23 細胞株の間にクラスタリングされた。SAEC の外側にクラスタリングされた 3 株は、他の肺腺癌とは異なる特殊な発現パターンを持つことが示唆された。

個々の発現制御異常解析に供するために、各細胞株について differentially expressed genes を定義した。これらは、当該細胞株において、26 細胞株平均発現量に比して 4 倍以上あるいは 1/16 倍以下の発現量を示すものである。細胞株当たり平均 352 個の高発現遺伝子および 1967 個の低発現遺伝子が抽出された (図 21)。高発現遺伝子、低発現遺伝子の総数は細胞株によって大きく異なっていた。

癌関連遺伝子に着目して細胞株によって発現量が異なる遺伝子を精査した (図 22)。EGFR について、3 株 (PC-3、PC-7、VMRC-LCD) において有意な発現量低下が見られた。一方で、BRAF や TP53 等は細胞株間で発現量に有意な差がなかった。これら遺伝子発現異常のバリエーションがどのようなエピゲノム異常と関連しているのかを明らかにするために、同一細胞株によるエピゲノム解析を行った (後述)。

A



B

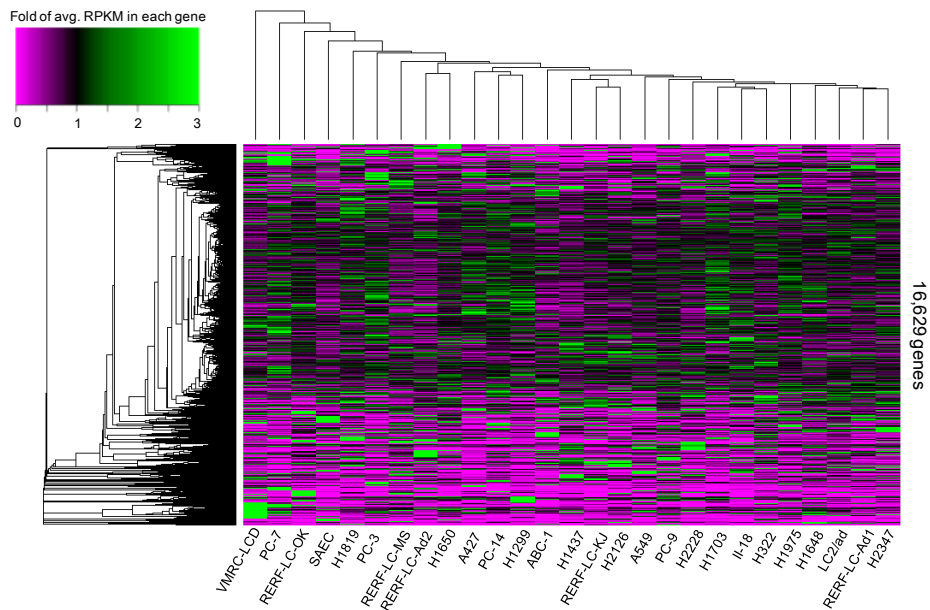


図 20 遺伝子発現パターン

(A) 26 細胞株のいずれかで >1 RPKM を示す 16,573 遺伝子について、各細胞株の RPKM に対して 26 細胞株平均発現量の fold 値を算出し、階層的クラスタリングを行った。(B) 正常細胞 (SAEC) を加え、16,629 遺伝子に対して同様の解析を行った。

(Supplementary Figure S14, Suzuki *et al.* 2014 *Nucleic Acids Research* 42(22):13557-13572 doi: 10.1093/nar/gku885 改変)

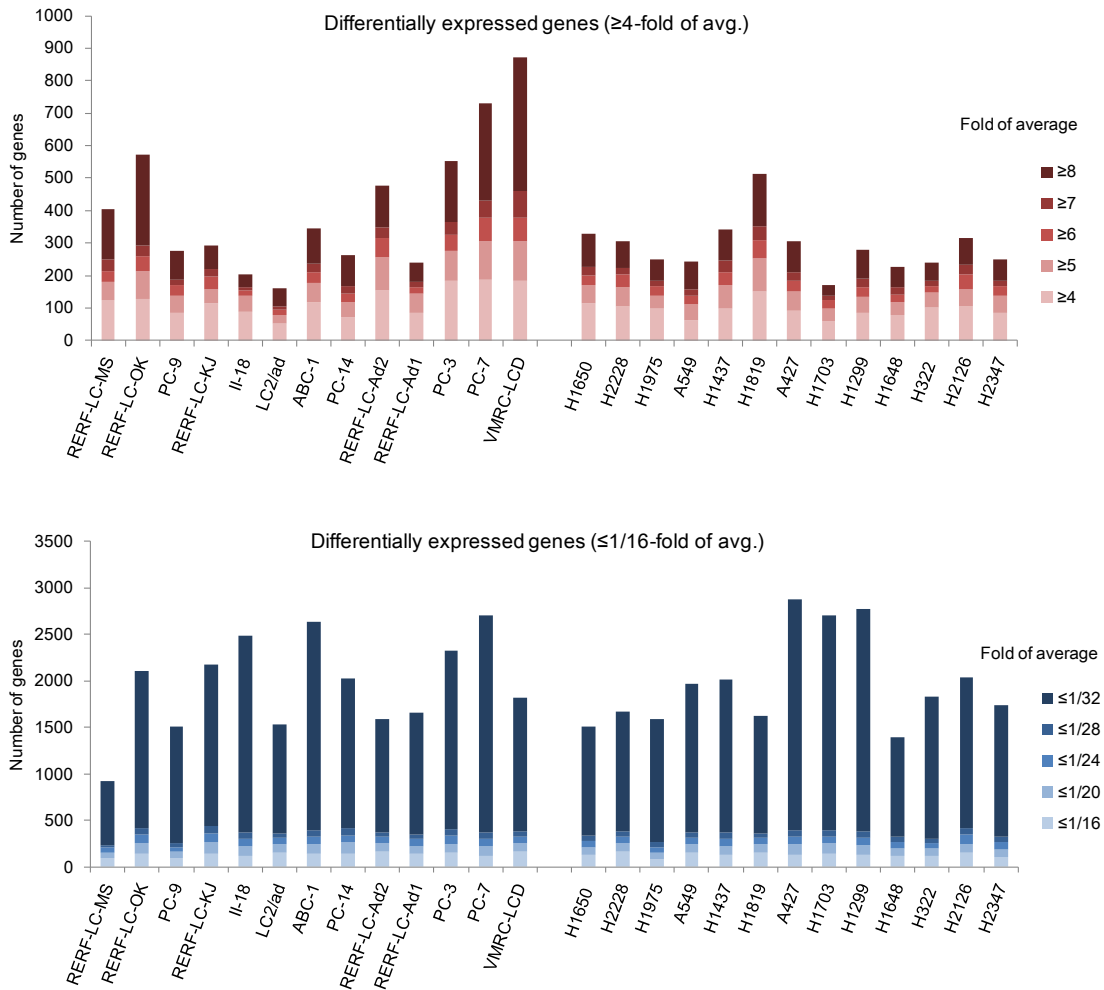


図 21 26 細胞株の differentially expressed gene の数

高発現遺伝子（上図）は平均発現量の 4 倍以上、低発現遺伝子（下図）は 1/16 倍以下の発現量を示す遺伝子を指し、各細胞株の高発現および低発現遺伝子の数を図示している。平均発現量の fold ごとに色分けしており、横軸の細胞株の順番は図 13 と同じである。

(Figure 2D, Suzuki *et al.* 2014 *Nucleic Acids Research* 42(22):13557-13572 doi:10.1093/nar/gku885)

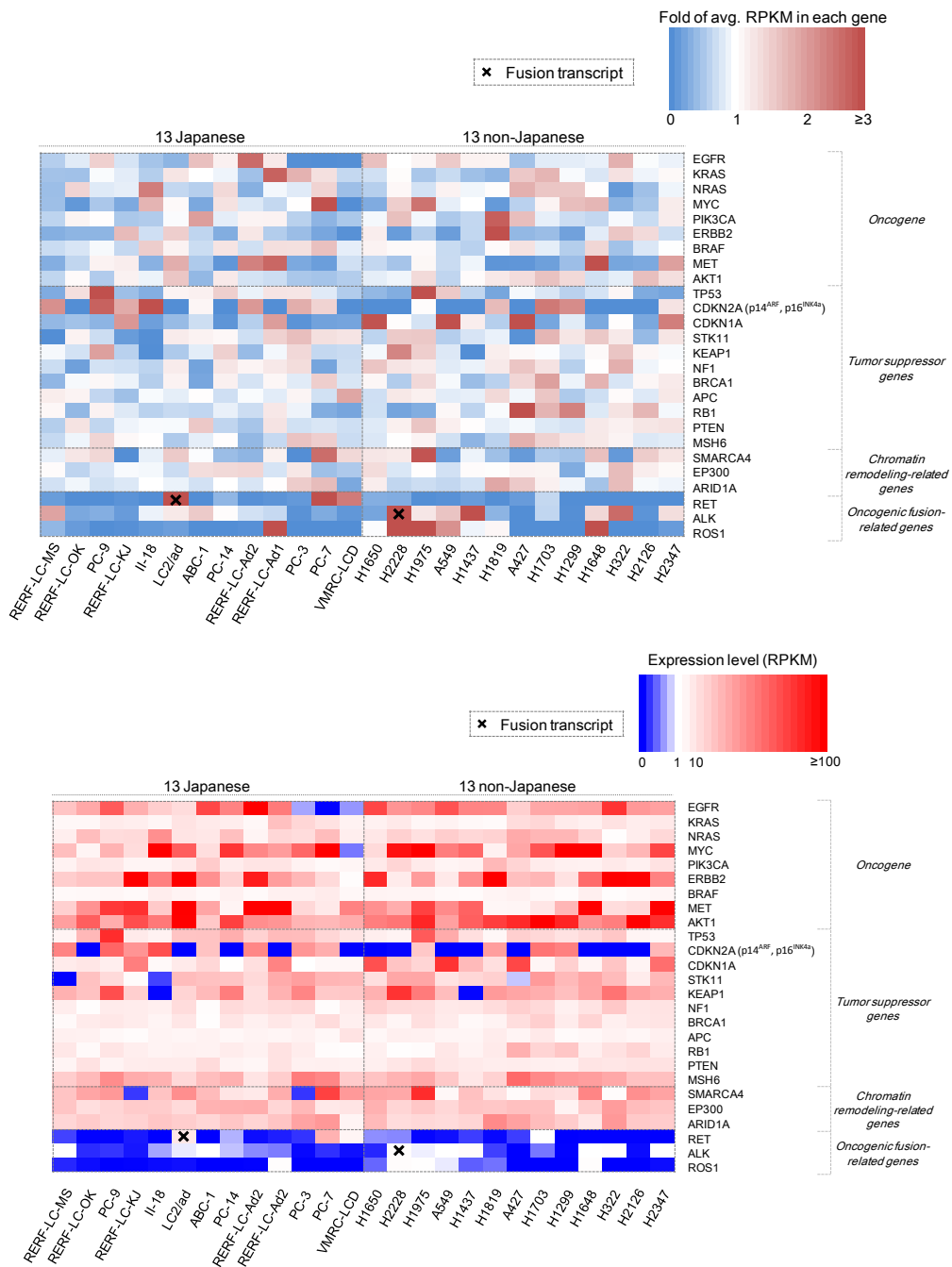


図 22 26 癌関連遺伝子の遺伝子発現量

26 癌関連遺伝子について、上図は 26 細胞株の平均発現量の fold 値、下図は発現量 (RPKM) のヒートマップを示している。融合遺伝子を×印で示す。

(Figure 2E, Suzuki *et al.* 2014 *Nucleic Acids Research* 42(22):13557-13572 doi: 10.1093/nar/gku885 改変)

3-3-3. DNA メチル化異常

癌細胞におけるシトシン塩基に対する DNA メチル化状態を解析するために、バイサルファイトシーケンシング(18)を行った。解析にはアジレント社が提供する SureSelect XT Methyl-Seq を用いた。得られたシーケンスタグのうち、平均 113,946,186 タグが参照ゲノム配列にマップされた(表 19)。SureSelect XT Methyl-Seq において設定された 84 Mb のベイト領域に対して、各細胞株の平均 sequence depth は 110×であり、10×sequence depth での coverage は 91%であった(表 19 および図 23)。バイサルファイト反応による C>T 変換率は、全ての細胞株で 99.2%以上であった。ダイレクトサンガーシーケンスおよびTA クローニングのサンガーシーケンスによって、DNA メチル化率の確認を行った(図 24)。83% (124 / 149 箇所の CpG サイト) について、DNA メチル化のパターンが相互に一致していた。

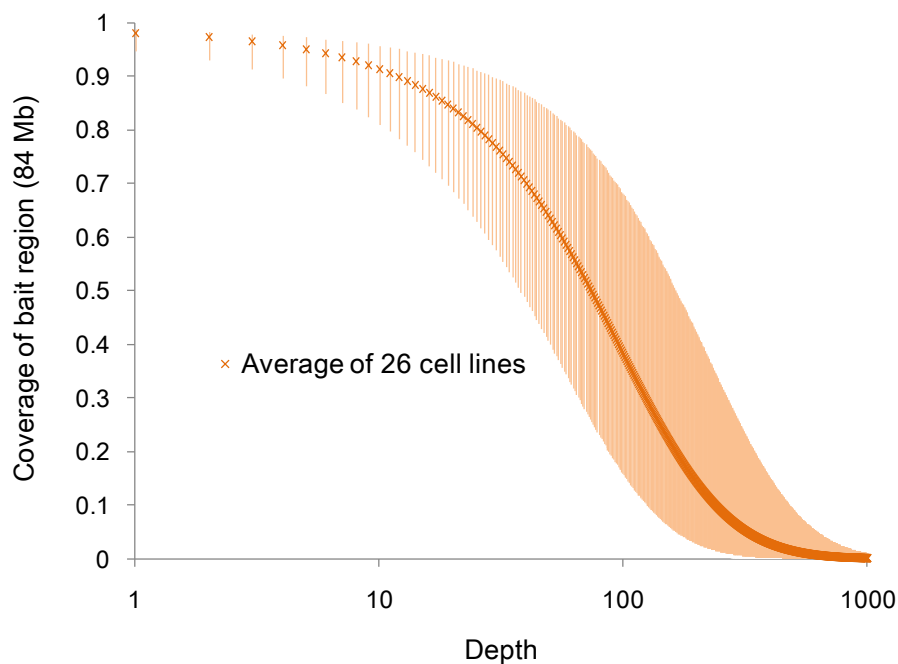


図 23 バイサルファイトシーケンスの depth と coverage

約 84 Mb のベイト領域に対するバイサルファイトシーケンスタグの平均 depth と coverage を示している。Depth および coverage は BEDTools (64) と Perl スクリプトを用いて算出した。グラフ内の×は 26 細胞株の平均値を示す。

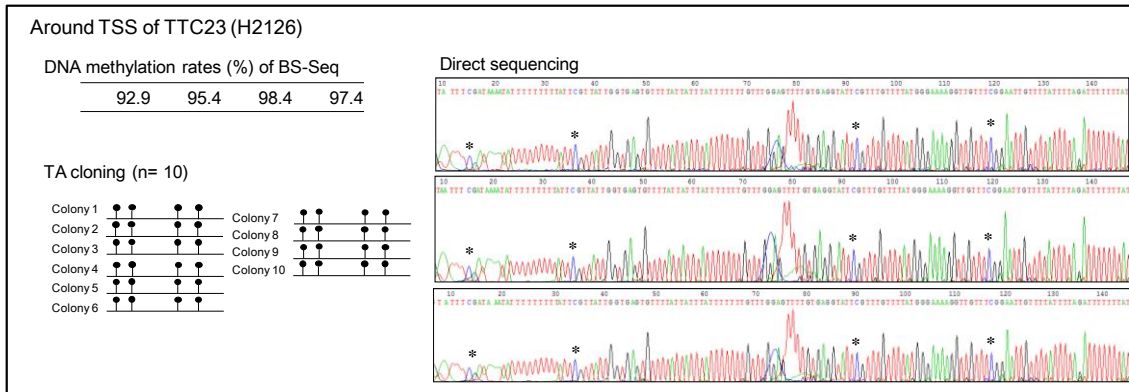
表 19 バイサルファイトシーケンスの統計

Cell line	Mapped sequences	Avg. of depths	Conversion rate (x5)*	CpG sites (> x5)
PC-3	157,902,653	161.4	0.994	3,673,159
PC-7	109,919,011	110.9	0.994	3,418,929
PC-9	87,012,056	89.6	0.994	3,231,320
PC-14	204,216,479	210.3	0.994	4,064,068
RERF-LC-Ad1	87,043,746	89.1	0.992	3,264,395
RERF-LC-Ad2	78,300,691	83.0	0.994	3,448,211
RERF-LC-KJ	72,844,738	74.9	0.993	3,068,971
RERF-LC-MS	102,938,936	109.0	0.994	3,598,662
RERF-LC-OK	161,552,507	165.0	0.993	3,758,532
VMRC-LCD	84,681,570	89.5	0.992	3,136,774
LC2/ad	112,097,386	116.0	0.988	3,548,548
ABC-1	93,158,547	93.1	0.993	3,493,903
II-18	99,682,438	165.0	0.993	3,327,001
A549	87,966,180	91.0	0.991	3,324,364
A427	53,499,542	54.3	0.992	2,614,641
H322	153,896,186	165.8	0.989	4,161,775
H2228	122,705,759	81.6	0.993	4,815,543
H1299	118,923,875	82.2	0.994	4,533,930
H1437	98,311,209	63.1	0.993	4,382,225
H1648	102,033,841	104.4	0.989	3,357,747
H1650	105,694,196	109.4	0.994	3,460,378
H1703	127,897,486	81.6	0.994	5,513,896
H1819	220,008,485	223.4	0.986	4,085,231
H1975	79,688,628	81.7	0.993	3,274,116
H2126	124,651,437	80.2	0.993	4,991,289
H2347	115,973,241	76.1	0.993	4,661,415

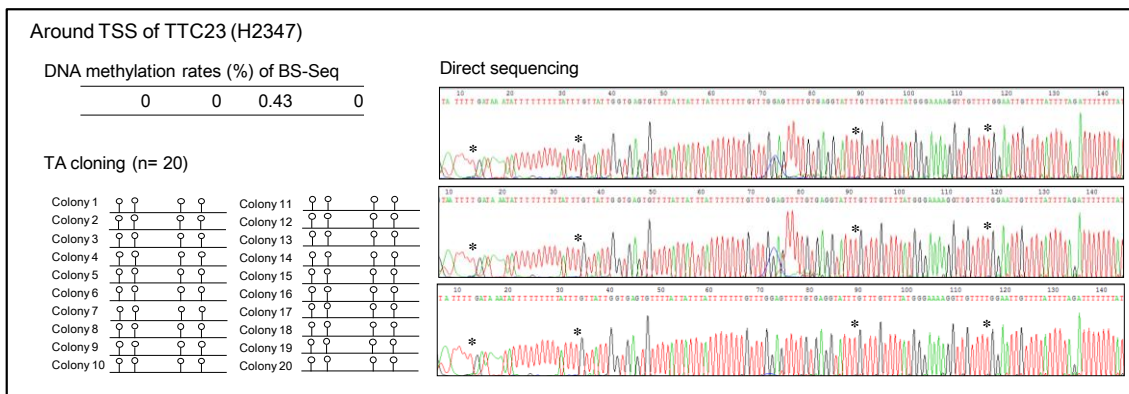
*C>T 変換率: $(TA + TT + TC) / (CA + CT + CC + TA + TT + TC)$

(Figure 2E, Suzuki *et al.* 2014 *Nucleic Acids Research* 42(22):13557-13572 doi: 10.1093/nar/gku885)

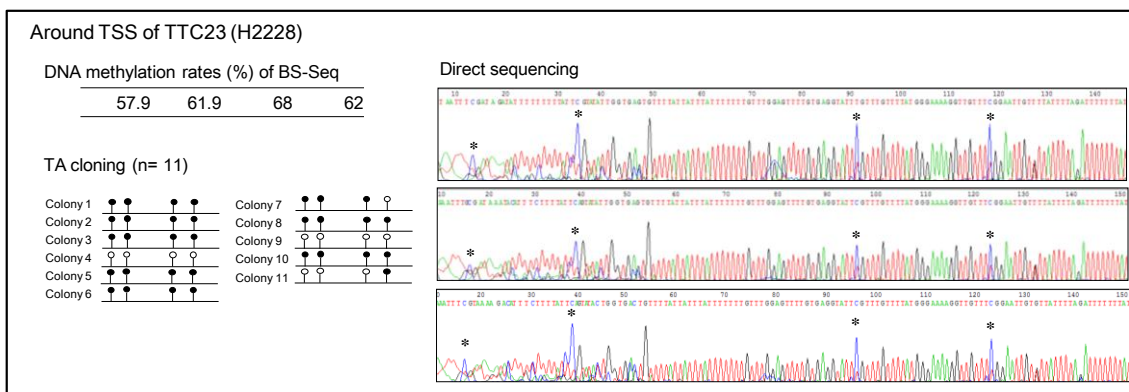
A



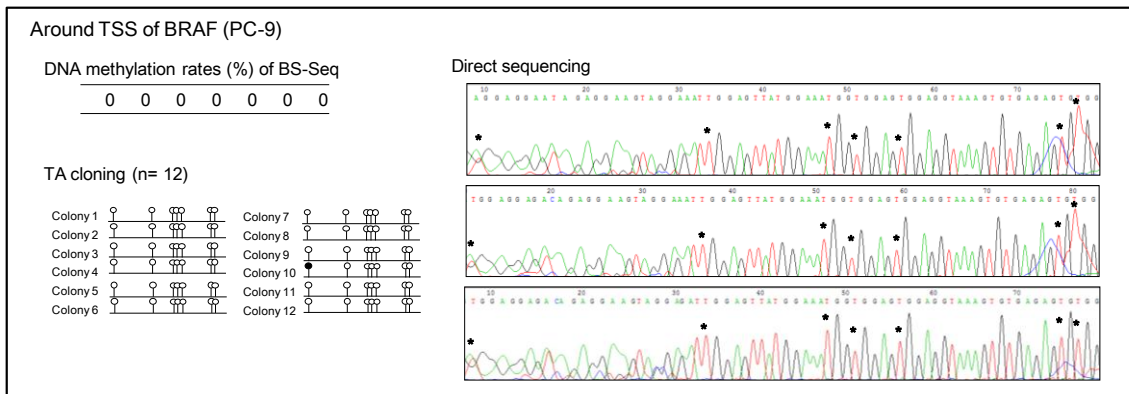
B



C



D



E

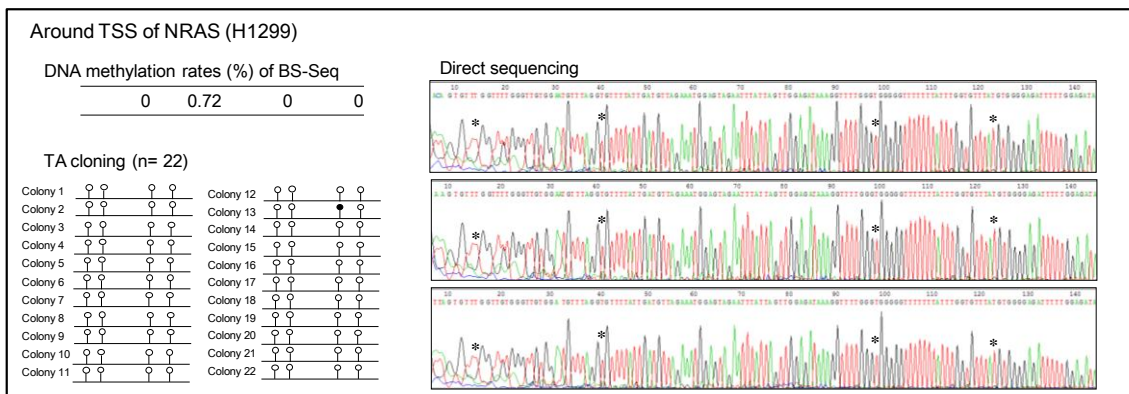


図 24 サンガーシーケンスによるパイサルファイトシーケンスの確認

パイサルファイトシーケンシングのテンプレートを用いて、ダイレクトサンガーシーケンスおよび TA クローニングを行った。(A-C) H2126、H2347、H2228 における tetratricopeptide repeat domain 23 (TTC23) 転写開始点付近 (4 CpG サイト)、(D) PC-9 の BRAF 転写開始点付近 (7 CpG サイト)、(E) H1299 の NRAS 転写開始点付近 (4 CpG サイト) の結果を例として示している。

(Supplementary Figure S3, Suzuki *et al.* 2014 *Nucleic Acids Research* 42(22):13557-13572 doi: 10.1093/nar/gku885 改変)

各細胞株において、×5以上の sequence depth を満たす CpG サイトは細胞株あたり平均 3,777,270 サイトであった。そのうち、1,273,909 サイトが CpG アイランド上に見られ、その他の CpG サイトに比べて DNA メチル化率が低かった（上図、図 25）。

各領域の平均 DNA メチル化率を図 25 の下図に示す。CpG アイランドは低 DNA メチル化を示すのに対し、CpG shore は細胞株によって DNA メチル化率が大きく異なっていた(83)。同様に、CpG アイランドを含むプロモーター領域（転写開始点±1.5 kb）についてもその他のプロモーターより DNA メチル化率が低い傾向にあった。また階層的クラスタリングの結果、9 株の高 DNA メチル化群と、17 株の低 DNA メチル化群に 26 細胞株が分割された（図 26）。EGFR 変異陽性株および KRAS G12 変異陽性株は全て低 DNA メチル化群に位置していた。

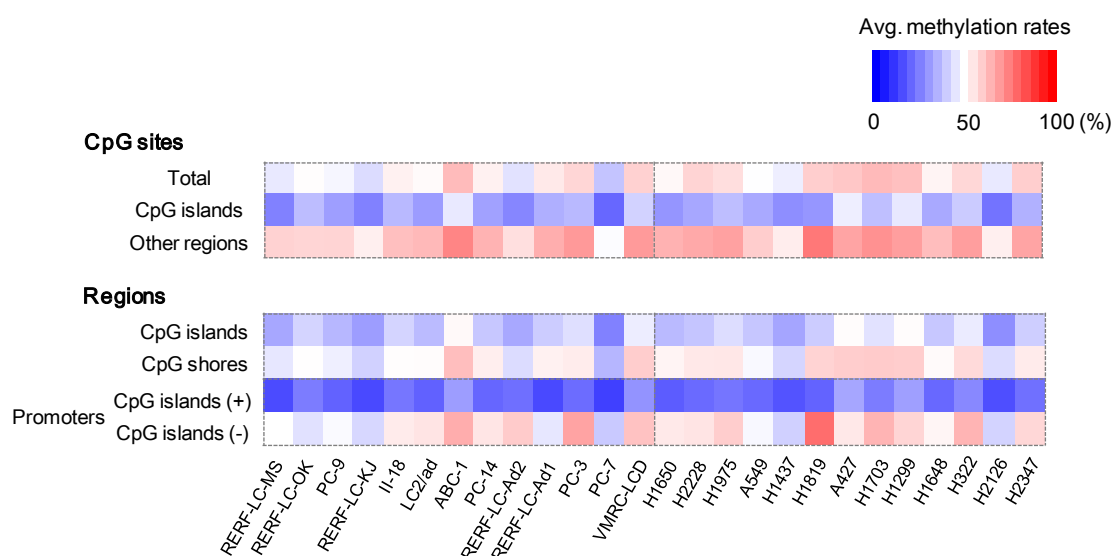


図 25 各細胞株における CpG サイトの平均 DNA メチル化率

各細胞株における DNA メチル化状態をヒートマップとして図示している。上図は、各 CpG サイトの平均 DNA メチル化率を示しており、下図は、各領域の平均 DNA メチル化率を示している。CpG shore は CpG アイランドの外側 2 kb の領域を指す。プロモーター領域は転写開始点±1.5 kb の領域とし、1 塩基でも CpG アイランドを含むプロモーターを CpG アイランド (+)、全く含まないプロモーターを CpG アイランド (-) とした。

(Figure 3A, Suzuki *et al.* 2014 *Nucleic Acids Research* 42(22):13557-13572 doi:10.1093/nar/gku885)

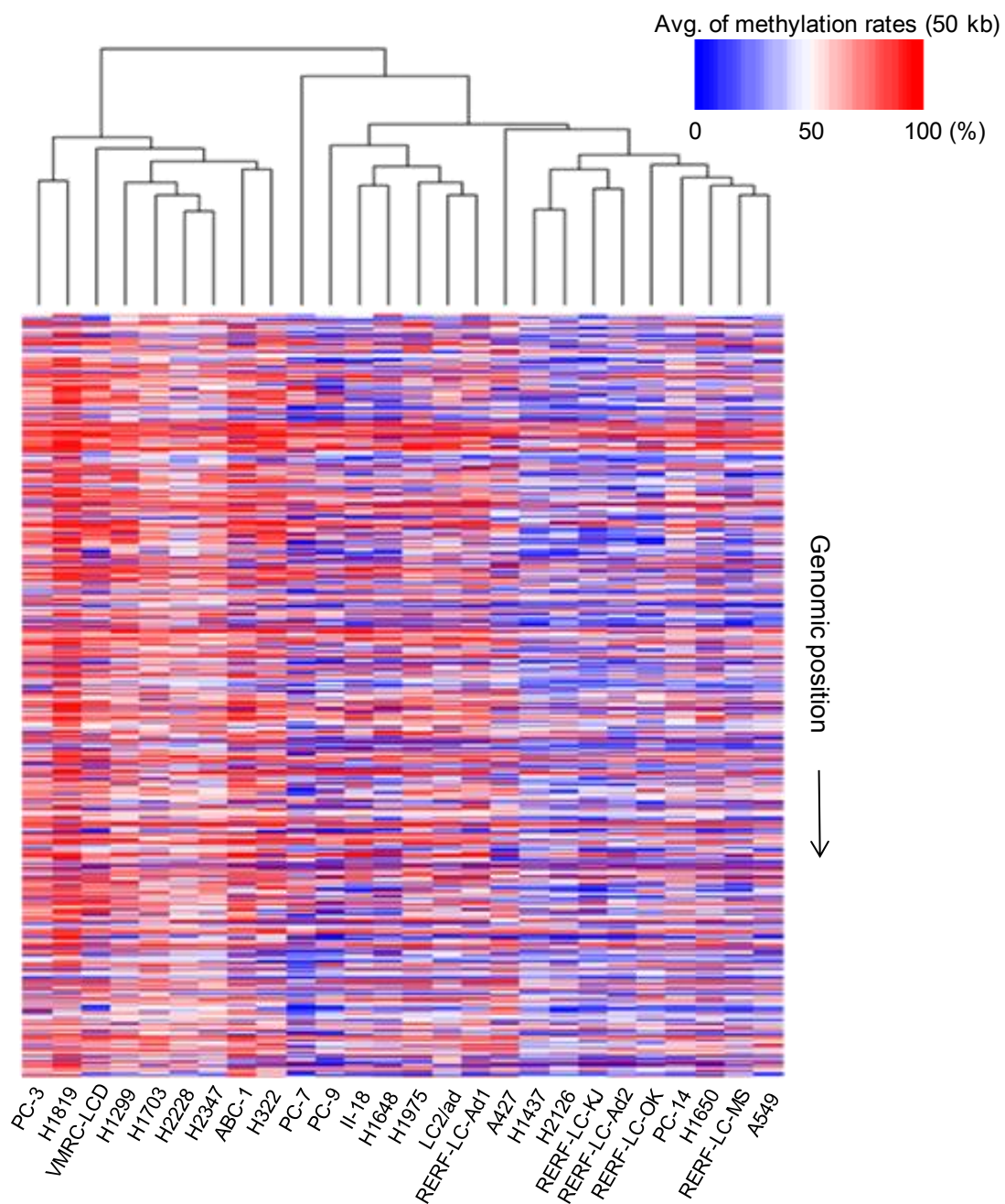


図 26 26細胞株における DNA メチル化パターン

50 kb ごとの平均 DNA メチル化率を用いた階層的クラスタリングの結果を図示している。平均 DNA メチル化率の算出には、26 細胞株全てで DNA メチル化率が算出された CpG サイトのみを用いている。

(Supplementary Figure S16, Suzuki *et al.* 2014 *Nucleic Acids Research* 42(22):13557-13572 doi: 10.1093/nar/gku885)

各細胞株における DNA メチル化異常(84-87)の指標として、プロモーター領域（転写開始点±1.5 kb)の differential DNA methylation を定義した。当該細胞株において、26 細胞株平均に比して 4 倍以上あるいは 1/16 倍以下に DNA メチル化率が変動している遺伝子群である。細胞株あたり平均 118 個の高 DNA メチル化遺伝子と 278 個の低 DNA メチル化遺伝子を抽出した（図 27）。PC-7 においては、特に低 DNA メチル化遺伝子が濃縮していた。また、differential DNA methylation を示すプロモーター上に、平均 61 個のゲノム変異が位置していた。これらのゲノム変異の中には、変異が生起した結果、DNA メチル化に影響を及ぼしたものがあると考えている。

Differential DNA methylation の例として、insulin-like growth factor 1 receptor (IGF1R) と NRAS における例を図 28 に示す。IGF1R のプロモーター領域では、3 株が高 DNA メチル化、5 株が低 DNA メチル化を示した。また、NRAS のプロモーター領域では、H322 が高 DNA メチル化を示した。ほとんどの場合、高 DNA メチル化を示した細胞株では、他の細胞株に比して遺伝子発現の低下が認められた。RNA-Seq 解析により見出された遺伝子発現低下遺伝子について、DNA メチル化に起因するものを抽出することが可能であると考えている。

前述した 26 種類の癌関連遺伝子について、転写開始点付近の DNA メチル化状態を解析した（図 29）。癌遺伝子および癌抑制遺伝子の多くは低 DNA メチル化傾向を示しており、RNA-Seq 解析からもこれらについては有意な遺伝子発現が観測されている。H322 における NRAS（前述）、PC-7 および VMRC-LCD における ERBB2、6 株における CDKN2A について、高い DNA メチル化が観察された。融合遺伝子関連の RET、ALK および ROS1 では細胞株によって DNA メチル化率に違いが見られた。これらは多くの細胞株において遺伝子発現は極めて低い。一方で、転写開始点付近で低 DNA メチル化を示す細胞株も存在することから、DNA メチル化以外にもその転写抑制機構が存在することが示唆される。

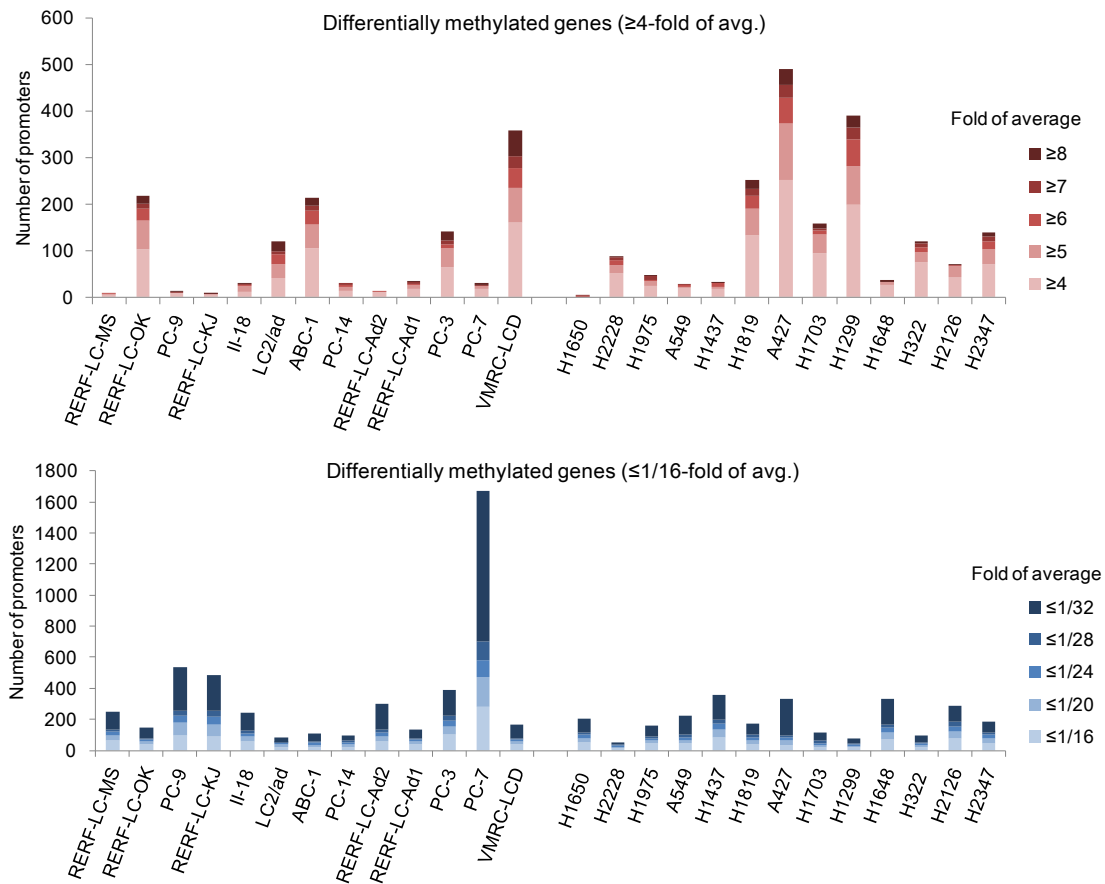


図 27 26 細胞株の differentially DNA methylated genes の数

各細胞株において、平均 DNA メチル化率の 4 倍以上を示す高 DNA メチル化遺伝子(上図)、1/16 倍以下を示す低 DNA メチル化遺伝子 (下図) の数を表示しており、fold ごとに色分けしている。横軸の細胞株の順番は図 13 と同じである。

(Figure 3B, Suzuki *et al.* 2014 *Nucleic Acids Research* 42(22):13557-13572 doi: 10.1093/nar/gku885)

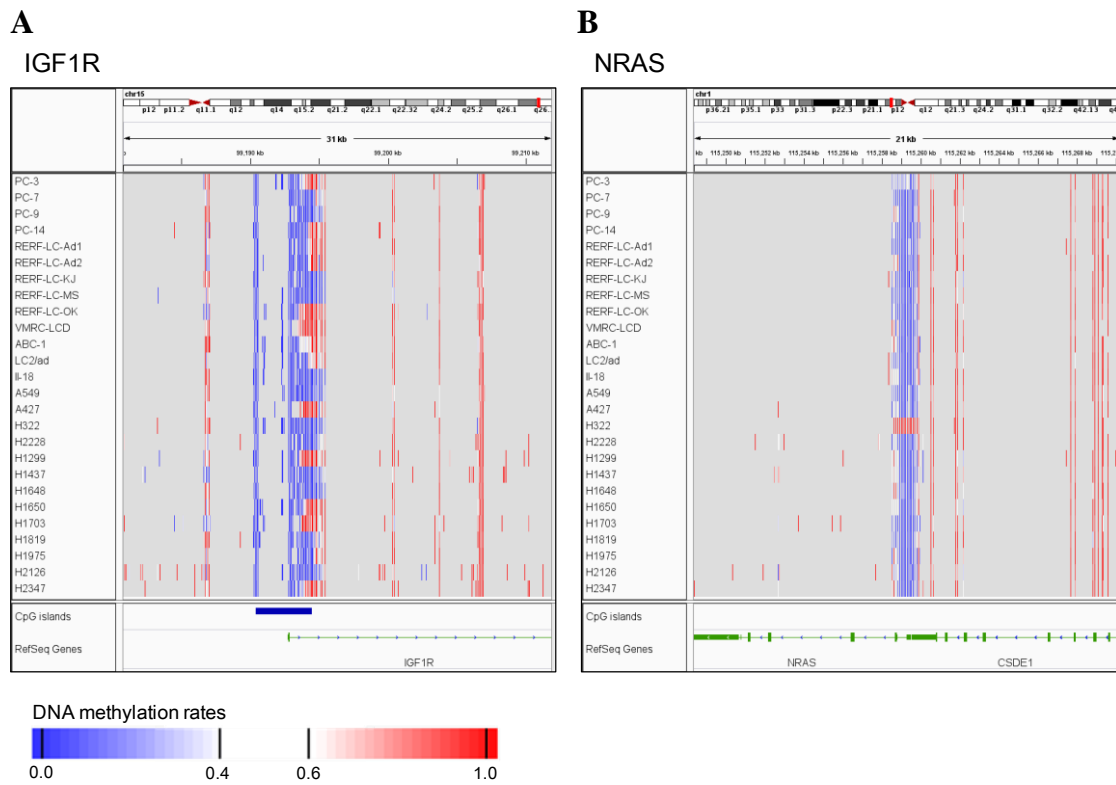


図 28 IGF1R および NRAS 転写開始点付近の DNA メチル化状態

IGF1R (A) および NRAS (B) の転写開始点付近の DNA メチル化パターンを IGV で可視化している。DNA メチル化率が算出された CpG サイトについて、色で DNA メチル化率を示している。

(Supplementary Figure S17, Suzuki *et al.* 2014 *Nucleic Acids Research* 42(22):13557-13572 doi: 10.1093/nar/gku885)

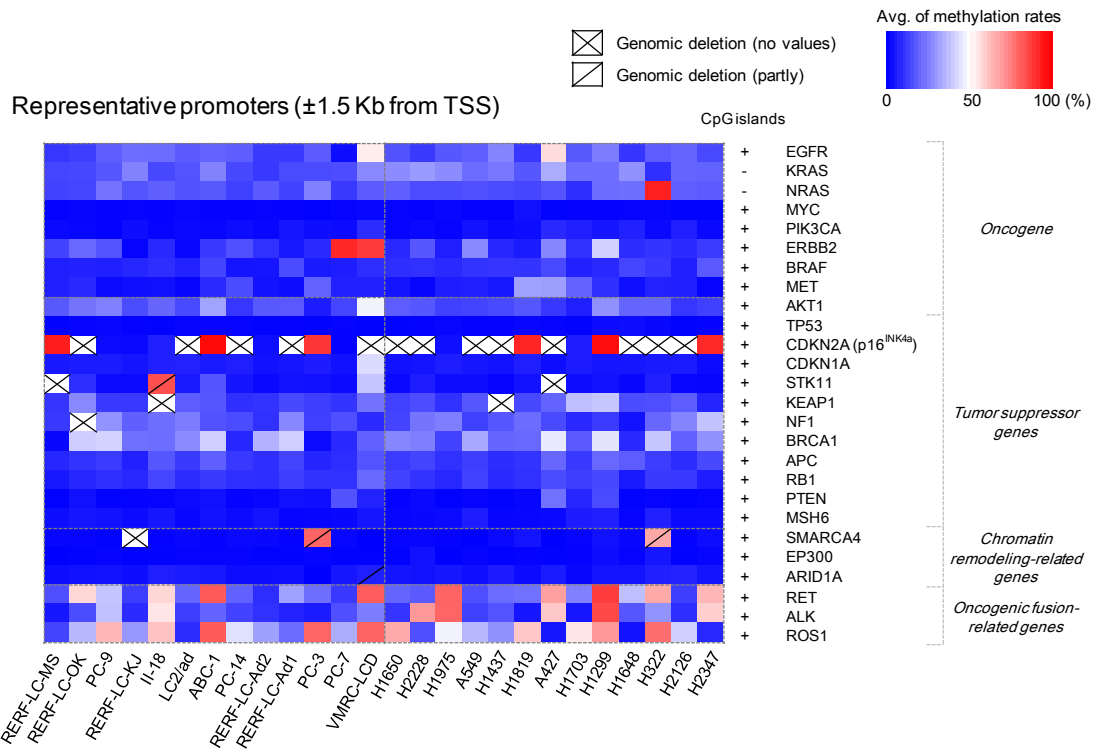


図 29 26 癌関連遺伝子の転写開始点付近における DNA メチル化状態

26 個の癌関連遺伝子について、各細胞株における転写開始点±1.5 kb の平均 DNA メチル化率を図示した。細胞株は、図 13 と同じ順番に並べられている。転写開始点±1.5 kb の領域が完全に欠損しているもしくは部分的に欠損している場合は、それぞれゲノム欠失、部分的なゲノム欠失としてヒートマップ内に記載した。各遺伝子の CpG アイランドの有無は、ヒートマップの横に記述した。

(Figure 3C, Suzuki *et al.* 2014 *Nucleic Acids Research* 42(22):13557-13572 doi:10.1093/nar/gku885)

3-3-4. クロマチンステータス異常

7 種類の代表的なヒストン修飾 (H3K4me1、H3K4me3、H3K9/14ac、H3K9me3、H3K27ac、H3K27me3、H3K36me3) (78,88)および Pol II の ChIP-Seq 解析(19)を行った。得られたシーケンスタグのうち、平均 23,276,610 タグが参照ゲノム配列にマップされた (表 20)。得られた ChIP-Seq データの精度を確認するために、反復実験および公共データ (A549 H3K4me3 ; ENCODE プロジェクト(21,22)) との比較を行った (図 30 および図 31)。図 30 に示す通り、H1975 における H3K4me3 および LC2/ad における Pol II の ChIP-Seq について 2 回実験を行い、シグナル強度を比較した。ピアソン相関係数 r は両者とも 0.9 以上であり、再現性が示されたといえる。A549 における H3K4me3 の ChIP-Seq データについて、本研究で得たデータと ENCODE プロジェクトのデータ (wgEncodeEH001905 および wgEncodeEH001904) との比較を図 31 に示す。87%以上のピークにおいてオーバーラップがあり、シグナル強度においてピアソンの相関係数 $r > 0.9$ の相関が得られた。

MACS2 を用いて ChIP-Seq データのピークを検出した結果を表 20 に示す。各遺伝子に対してピークを帰属させたところ、細胞株あたり平均 12,239 遺伝子 (59%) について H3K4me3 のピーク、2835 遺伝子 (14%) について H3K27me3 または H3K9me3 のピークが検出された。前述の RNA-Seq の結果から、各細胞株において平均 12,290 遺伝子について有意な遺伝子発現 (> 1 RPKM) が検出されている。H3K4me3 のピークが見られた遺伝子数と発現遺伝子数はほぼ同数であった。

また、エンハンサーマークである H3K4me1 および H3K27ac について、各細胞株で抽出されたピークから、総計 847,766 個の H3K4me1 および 426,224 個の H3K27ac エンハンサー領域が抽出された。そのうち、683,606 個の H3K4me1 エンハンサーと 337,545 個の H3K27ac エンハンサーは、遺伝子領域およびその近傍 (上流 100 kb または転写領域) に存在した。

H3K27me3 と H3K9me3 はともに転写抑制性のヒストン修飾であり、H3K4me3 の鋭いピークに比べてブロードなパターンを示した。解析に用いた MACS2 は narrow ピークと broad ピークを抽出することができ、H3K4me3 は narrow ピーク、H3K27me3 および H3K9me3 は broad ピークを使用して解析を行った。しかし、目視確認を行った結果、抑制性のヒストン修飾についてシグナルをピークとして検出することは困難であると考えられたために、ピークとしてシーケンスタグ濃縮領域を検出するのではなく、該当領域にマップされるシーケンスタグ数として検出したシグナル強度を直接解析に用いることとした。

表 20 ChIP-Seq の平均タグ数およびピーク数

ChIP	Average of mapped sequences	Average number of peaks (MACS2)	
		Narrow peaks	Broad peaks
H3K4me3	26,140,455	21,209	16,208
H3K9/14ac	19,596,187	34,374	23,753
Pol II	26,056,772	15,715	13,997
H3K36me3	24,264,604	107,708	47,710
H3K4me1	25,900,257	108,882	75,854
H3K27ac	25,690,276	61,061	38,297
H3K27me3	21,584,812	53,587	42,163
H3K9me3	21,155,573	39,559	51,760
WCE	19,100,553	***	***

(Supplementary Table S12A, Suzuki *et al.* 2014 *Nucleic Acids Research* 42(22):13557-13572 doi: 10.1093/nar/gku885)

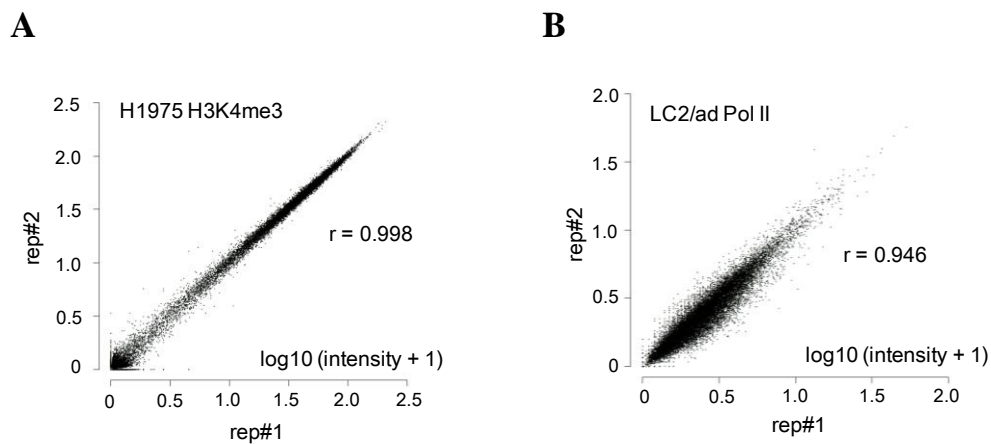
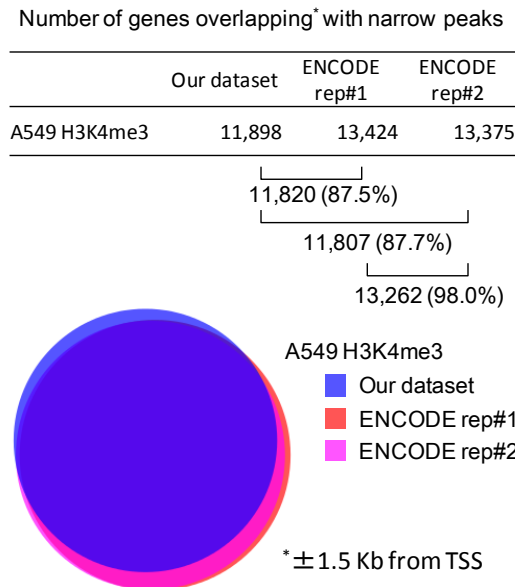
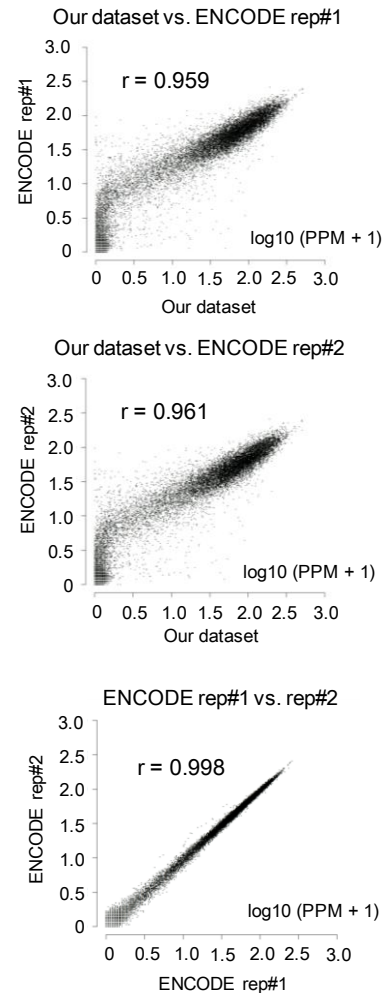


図 30 ChIP-Seq の再現性の確認

H1975 の H3K4me3 および LC2/ad の Pol II について、2 回 ChIP 実験を行い、再現性を確認した。20,598 遺伝子について、転写開始点±1.5 kb の ChIP シグナル強度 (intensity ; ChIP PPM / WCE PPM) を算出し、1 回目と 2 回目の実験のシグナル強度を比較した。両者ともピアソン相関係数 r は 0.9 以上であり、再現性が示された。

(Supplementary Figure S5, Suzuki *et al.* 2014 *Nucleic Acids Research* 42(22):13557-13572 doi: 10.1093/nar/gku885)

A**B****図 31 ENCODE との比較**

A549 の H3K4me3 について、ENCODE プロジェクトのデータ (wgEncodeEH001905 および wgEncodeEH001904) と比較した。(A) 転写開始点付近の narrow ピークを比較した。抽出されたピークの大部分 (87%) がオーバーラップしていた。(B) 20,598 遺伝子の転写開始点±1.5 kb の ChIP シグナル強度 (PPM) を比較した。ピアソン相関係数 r が 0.9 以上であり、公共データとの強い正の相関が示された。

(Supplementary Figure S6, Suzuki *et al.* 2014 *Nucleic Acids Research* 42(22):13557-13572 doi: 10.1093/nar/gku885)

各 ChIP-Seq データにおけるシグナル強度の相関を図 32 に示す。H3K9/14ac と H3K27ac の間に最も強い正の相関が認められた ($r= 0.878$)。H3K4me3 と H3K9/14ac、H3K4me3 と H3K27ac の間にも相関係数 0.8 を超える高い正の相関がみられた ($r= 0.869$ および $r= 0.844$)。これらは全て転写活性性のヒストン修飾であるため、そのシグナル強度に高い相関が認められると考えられた。同様に転写抑制性のヒストン修飾である H3K27me3 と H3K9me3 は正の相関を示し ($r= 0.647$)、一方で、H3K4me3 と H3K27me3、および、H3K4me3 と H3K9me3 は負の相関を示した ($r= -0.524$ および $r= -0.405$)。H3K36me3 と Pol II は、H3K9me3 とは特に相関が検出されなかった ($r= -0.126$ および $r= -0.150$)。全体としてこれらの活性性、抑制性の因子間で一定の相関が認められたものの、明確な相関が認められない遺伝子も一定数存在することは興味深い。これらについては、異なる因子ごとに独立した制御が行われているのかもしれない。

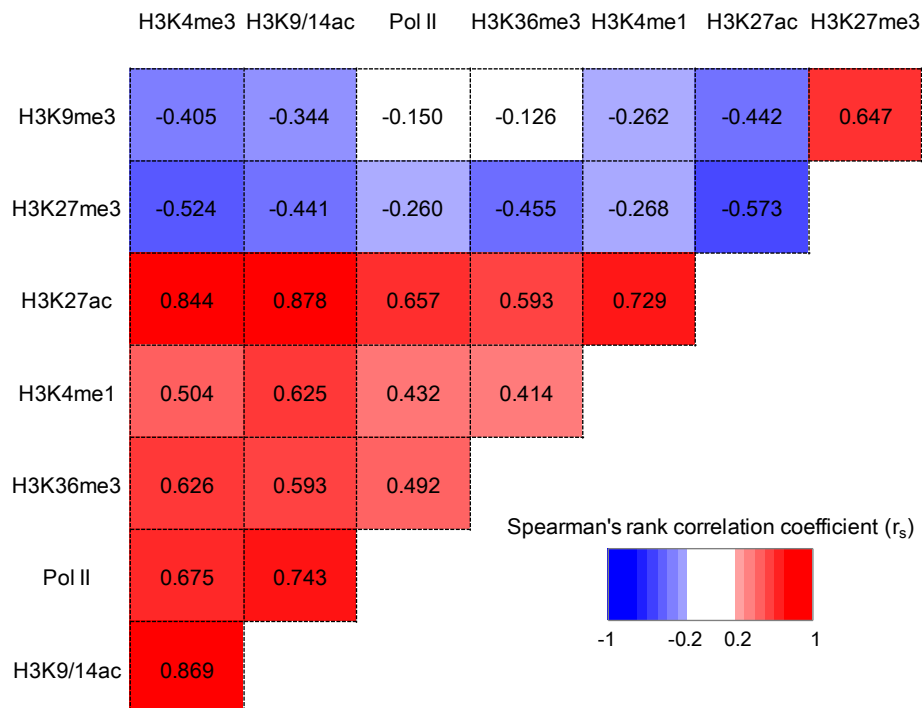


図 32 各 ChIP シグナル強度の相関

ChIP-Seq データを用いて 20,598 遺伝子のシグナル強度 (ChIP PPM / WCE PPM) を算出し、各 ChIP シグナルのスピアマン相関係数を細胞株ごとに算出した。図中の数値は 26 細胞株の平均値を示す。赤が正の相関、青が負の相関を示し、色が濃くなるほど相関が強いことを示す。

(Figure 4A, Suzuki *et al.* 2014 *Nucleic Acids Research* 42(22):13557-13572 doi:10.1093/nar/gku885)

ChIP-Seq データから、癌細胞におけるクロマチンステータスの異常(78,88)を明らかにするために、細胞株によってシグナル強度が異なる **differential chromatin marks** を抽出した (図 33)。これらは当該細胞株でのシグナル強度が 26 細胞株平均に比して 4 倍以上あるいは 1/16 倍以下に変動しているものである。これらのうち、いくつかの例については qPCR での検証実験を行った (図 34)。ChIP-Seq データから得られたシグナル強度と qPCR を比較したところ、相関係数 $r = 0.63$ の相関が得られた (図 34A)。また図 34E に示す通り、シグナル強度が 4 倍以上異なる 42 ペアについて、qPCR で定性性を確認した。85.7% (36/42 ペア) において、ChIP-Seq のシグナル強度と qPCR が定性的に一致した。

転写抑制性のヒストン修飾である H3K27me3 および H3K9me3 については、PC-7 や II-18 においてそのシグナルが高い遺伝子が多く検出された。一方で H1299 においては H3K9me3 について、H1819 においては H3K27me3 について強いシグナルが検出された。細胞株によって転写抑制に関わるエピゲノムに特徴があることが示唆された。DNA メチル化も含めた例を後述する。

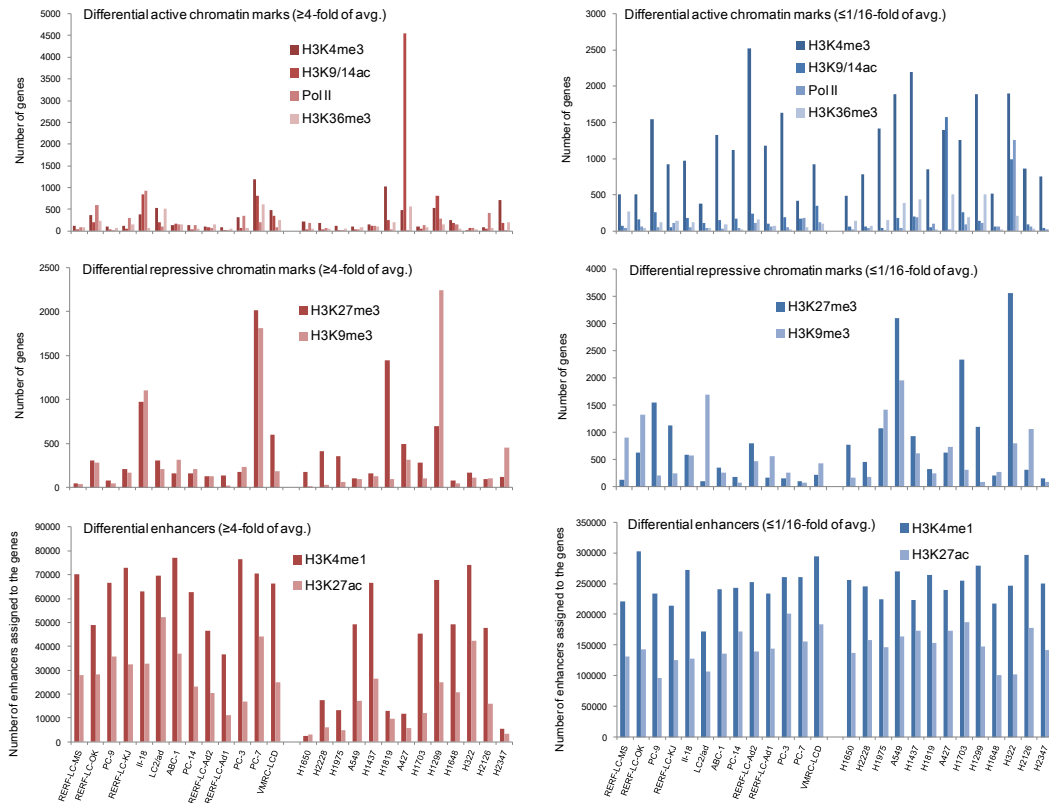
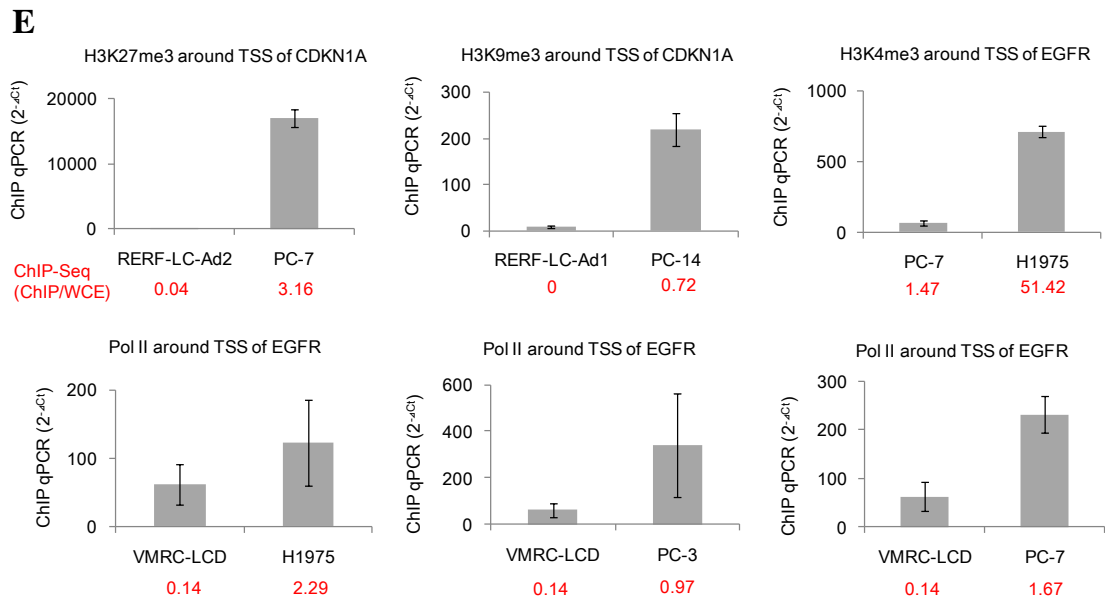
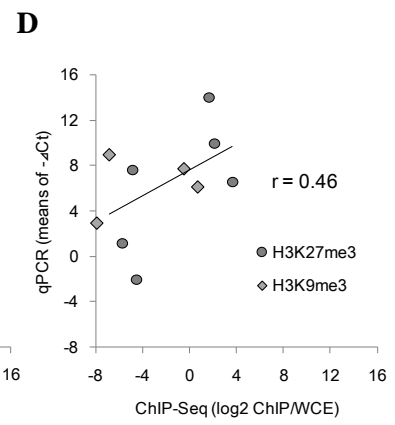
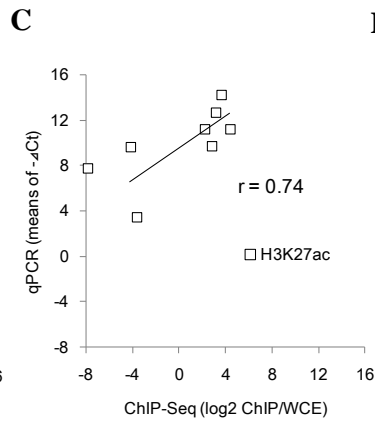
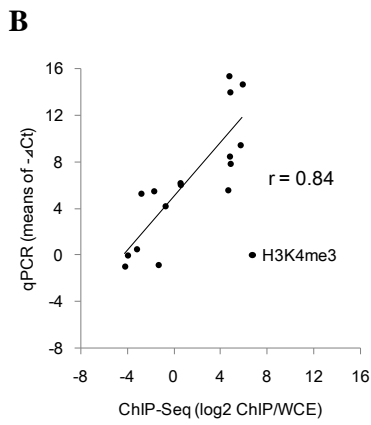
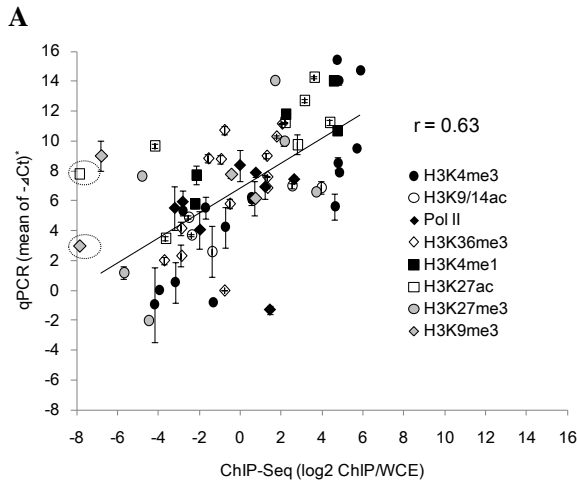
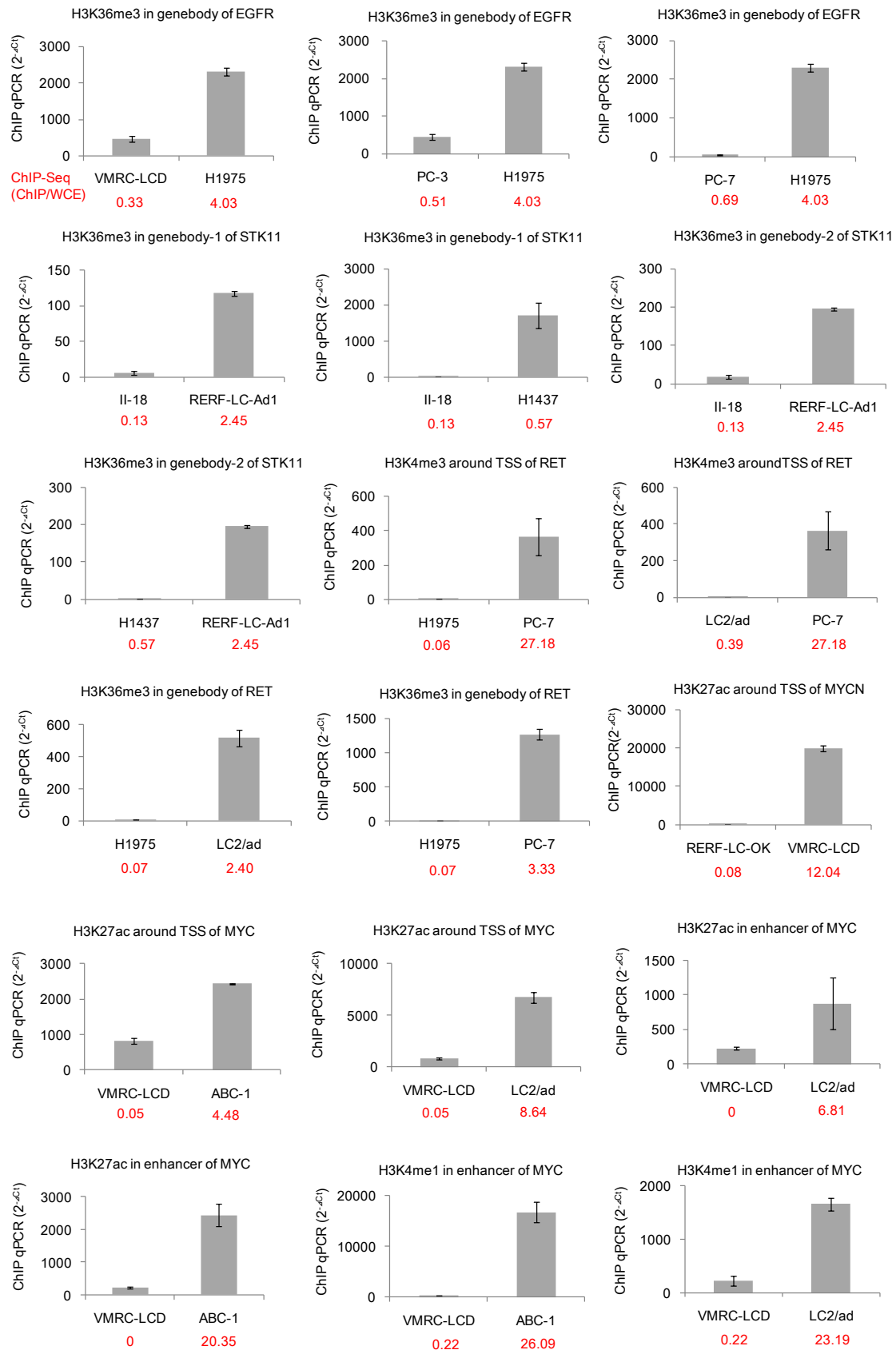


図 33 各細胞株における differential chromatin marks の数

各細胞株における differential chromatin marks の数をグラフに示している。上段が転写活性性 (H3K4me3、H3K9/14ac、Pol II および H3K36me3)、中段が転写抑制性 (H3K27me3 および H3K9me3)、下段がエンハンサー (H3K4me1 および H3K27ac) の differential chromatin marks の数を示している。横軸は各細胞株を示しており、順番は図 13 と同様である。

(Figure 4B, Suzuki *et al.* 2014 *Nucleic Acids Research* 42(22):13557-13572 doi: 10.1093/nar/gku885)





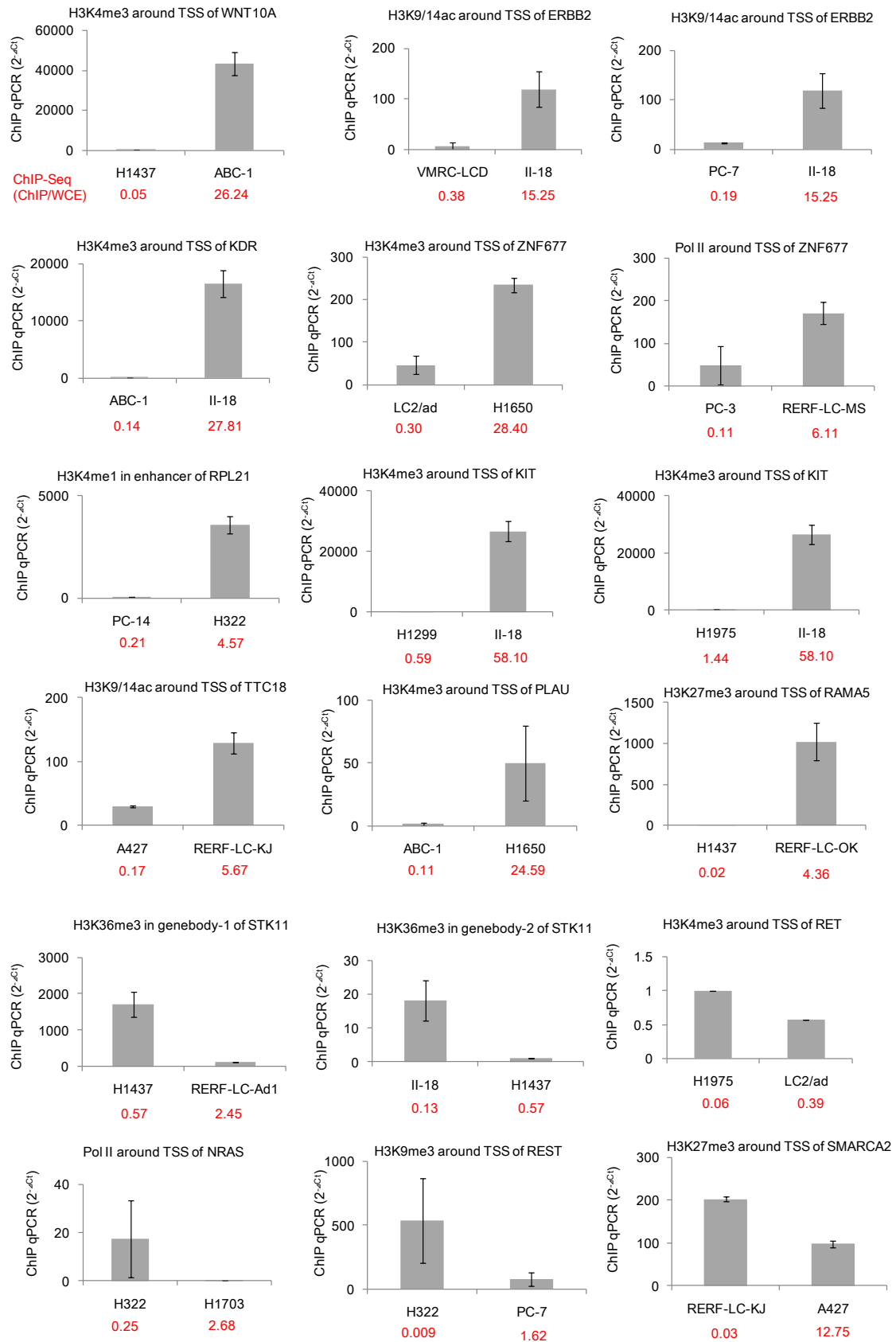


図 34 ChIP-qPCR によるシグナル強度の確認

ChIP-Seq データから算出されたシグナル強度について、ChIP-qPCR で確認を行った。

(A) ChIP-Seq のシグナル強度と qPCR の Ct 値を比較した (n=3)。 ΔCt は $Ct_{\text{ChIP}} - Ct_{\text{WCE}}$ で算出されている。Differential chromatin marks を含む 65 点の結果から、正の相関 ($r=0.63$) が見出された。特に H3K4me3 (B) および H3K27ac (C) では強い正の相関が見られ、転写活性化ヒストン修飾の ChIP-Seq から算出されたシグナル強度は定量的であると示された。一方で抑制性のヒストン修飾 (D) では弱い正の相関が見られた。抑制性のヒストン修飾はブロードなパターンを示すので、ピーク型のヒストン修飾に比べてシグナル強度を算出するのが困難であることが示唆された。(E) ChIP-Seq のシグナル強度が 4 倍以上異なる 42 ペアについて、qPCR で定性性を確認した。グラフは qPCR の結果を示しており、グラフの下に ChIP-Seq のシグナル強度が赤字で記載してある。85.7% (36 ペア) において、ChIP-Seq のシグナル強度と qPCR の結果が定性的に一致した。

(Supplementary Figure S4, Suzuki *et al.* 2014 *Nucleic Acids Research* 42(22):13557-13572 doi: 10.1093/nar/gku885)

前述で検出されたゲノム変異と、得られたクロマチンステータスを関連付けるために、プロモーターやエンハンサーといった制御領域に存在するゲノム変異 (SNVs および indels) を抽出した (表 21)。プロモーター領域 (転写開始点±1.5 kb) に細胞株あたり平均 2683 個のゲノム変異が存在し、このうち、当該細胞において H3K4me3 について高あるいは低シグナルパターンが見出されたものが 397 個存在した。同様に、エンハンサー領域では、H3K4me1 エンハンサーに平均 8848 個、H3K27ac エンハンサーに 6654 個のゲノム変異が見られ、そのうち、当該細胞において H3K4me1 について高あるいは低シグナルパターンが見出されたものが 3687 個、H3K27ac について高あるいは低シグナルパターンが見出されたものが 3112 個検出された。これら differential chromatin marks と共起するゲノム変異が細胞株平均で 6257 個存在し、癌細胞におけるクロマチンステータスの異常の生起に関わっている可能性が示唆された(15,89)。

表 21 制御領域におけるゲノム変異と differential chromatin marks

	Average number of positions in the 26 cell lines	
	SNVs	Short indels
Regulatory regions*	11,413	2,160
Promoter†	2,187	497
Promoter with differential H3K4me3	181	36
Enhancer (H3K4me1)‡	7,543	1,305
Enhancer with differential H3K4me1	3,163	524
Enhancer (H3K27ac)‡	5,549	1,006
Enhancer with differential H3K27ac	2,647	465

*プロモーター (転写開始点±1.5 kb) およびエンハンサー領域

†20,598 遺伝子

‡683,606 H3K4me1 および 337,545 H3K27ac のエンハンサー

(Table 1B, Suzuki *et al.* 2014 *Nucleic Acids Research* 42(22):13557-13572 doi:10.1093/nar/gku885)

俯瞰的にゲノム領域でのクロマチン状態を観測するために、8種類のChIP-Seqデータに対して、ChromHMM (59)を用いてクロマチンステータスを記載した(60)。クロマチンステータスとChIP-Seqとの関係を図35に示す。検出された8種類のクロマチンステータスをそれぞれ”active promoter”、”weak/poised promoter”、”strong enhancer”、”weak enhancer”、”transcriptional elongation”、”inactive region”、”inactive region/heterochromatin”、”low/no signal”と定義した。図36にEGFRの例を示す。23株において、転写開始点付近に”active promoter”ステータスが検出された。これら23株においては、RNA-Seq解析からEGFRの遺伝子発現が確認されており、転写の活性化がクロマチンステータスと一致していた。残りの3株(PC-3、PC-7およびVMRC-LCD)については、RNA-Seq解析からはEGFRの発現は検出されなかったが(3-3-2参照)、ChromHMMの結果から、これらの細胞株間で相互に異なる転写開始点付近のクロマチンステータスが検出された。PC-3においては、クロマチンステータスが”weak/poised promoter”を示すのに対し、VMRC-LCDとPC-7は”inactive region”のステータスを示した。さらに個々の転写活性性のクロマチンマークに着目したところ、PC-3とVMRC-LCDではH3K4me3のシグナルが認められたが、PC-7では検出されなかった(図37)。また、Pol IIについては、PC-3においてのみ弱いシグナルが検出された。転写伸長マークであるH3K36me3は3株のいずれにおいても検出されなかった。RNA-Seqにおいて遺伝子発現低下が検出された場合でも、細胞株によってその分子機構が異なる例が見出されたと考えている。

EGFRを含む26癌関連遺伝子のクロマチンステータスについて図38に示す。ChIP-Seqのデータをゲノムデータ、遺伝子発現データおよびDNAメチル化データに加えて解析に用いることで、遺伝子発現異常を引き起こす制御因子をより詳細かつ段階的に明らかにすることが可能になったと考えている。

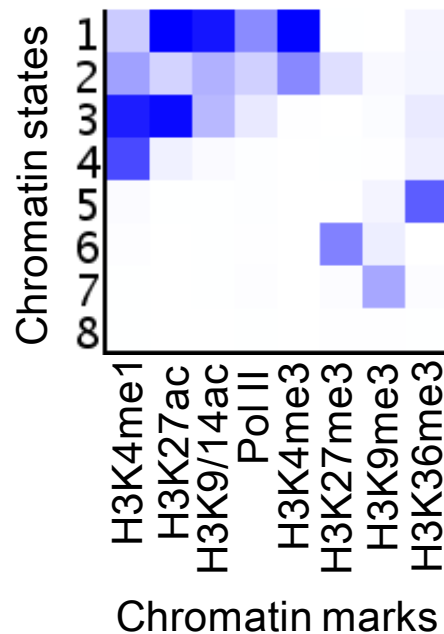


図 35 ChromHMM による 8 種類のクロマチンステータス

ChromHMM の出力ファイルの一つで、ChromHMM によるクロマチンステータスと ChIP-Seq シグナルとの関係性を示している。ChIP-Seq シグナルの寄与度から、H3K4me3、H3K9/14ac および H3K27ac が強いステータス 1 を“active promoter”、H3K4me3 が少し弱いステータス 2 を“weak/poised promoter”とした。また、H3K27ac および H3K4me1 が強いステータス 3 を“strong enhancer”、H3K4me1 のみが強いステータス 4 を“weak enhancer”、H3K36me3 が強いステータス 5 を“transcriptional elongation”とした。転写抑制性の H3K27me3 が強いステータス 6 は“inactive region”、H3K9me3 が特徴的であるステータス 7 は“inactive region/heterochromatin”とした。ステータス 8 は ChIP-Seq のシグナルが検出されていないので“low/no signal”とした。

(Supplementary Figure S20, Suzuki *et al.* 2014 *Nucleic Acids Research* 42(22):13557-13572 doi: 10.1093/nar/gku885 一部)

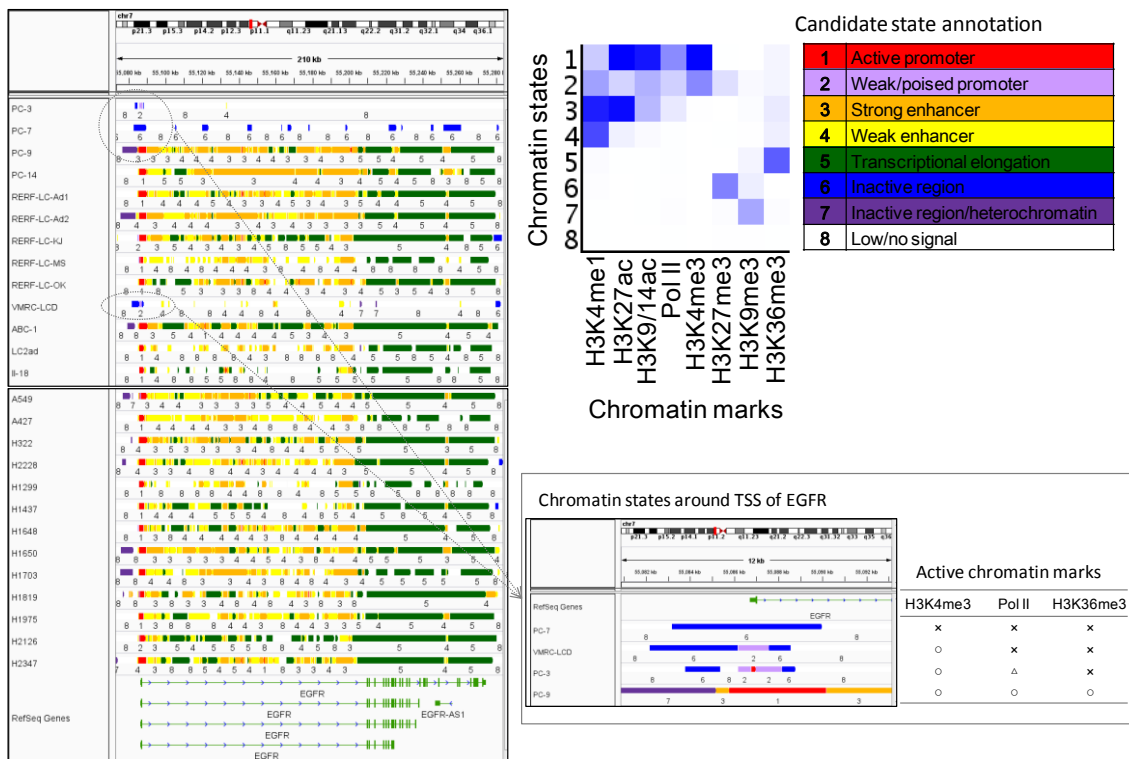


図 36 ChromHMM による EGFR 遺伝子領域のクロマチンステータス

26 細胞株における EGFR のクロマチンステータスを IGV で表示している。ChromHMM の 8 種類のクロマチンステータスで EGFR の領域を色分けしており、各クロマチンステータスの色は右上に示すとおりである。PC-9 のように EGFR の転写が活性化している細胞株では転写開始点付近が”active promoter”として赤く塗られている。PC-7、VMRC-LCD および PC-3 では転写が活性化しておらず、転写開始点付近のステータスが”weak/poised promoter”および”inactive region”となっている。

(Supplementary Figure S20, Suzuki *et al.* 2014 *Nucleic Acids Research* 42(22):13557-13572 doi: 10.1093/nar/gku885)

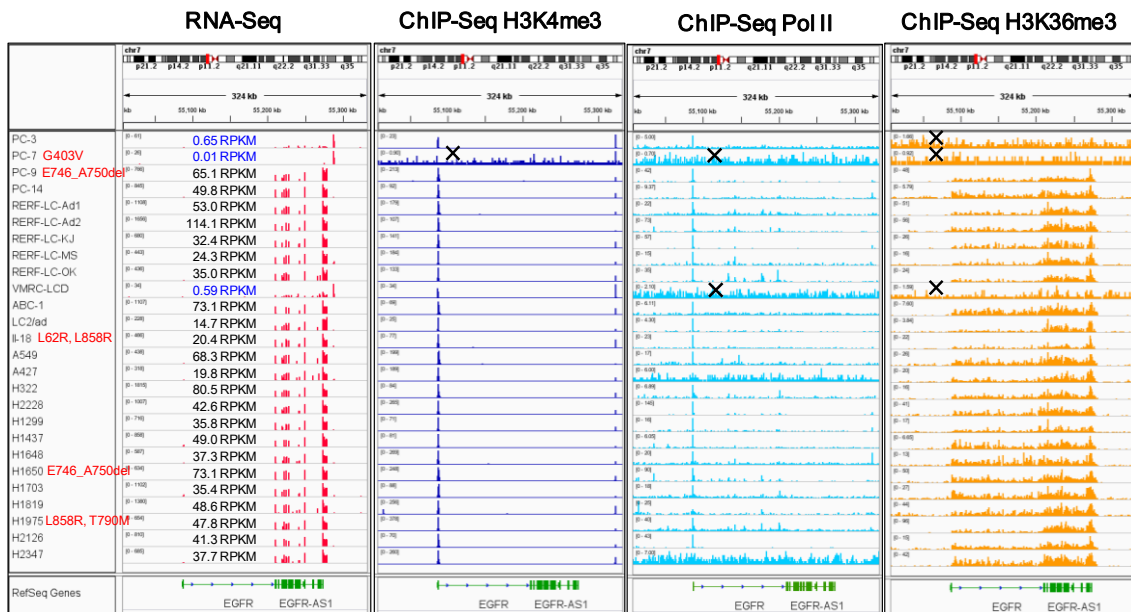
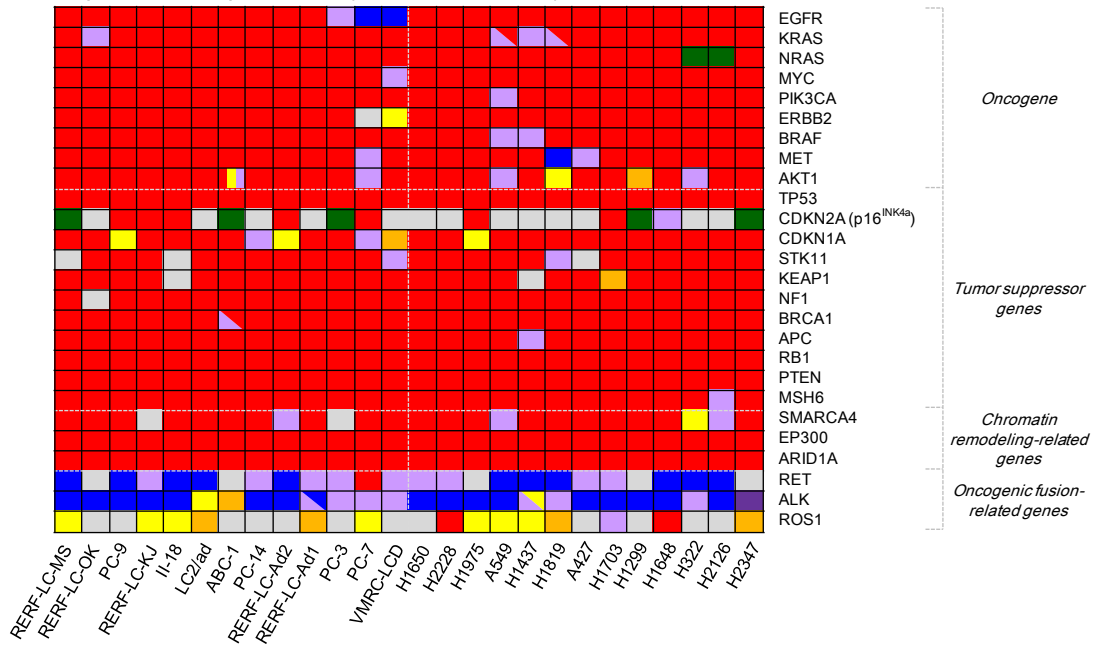


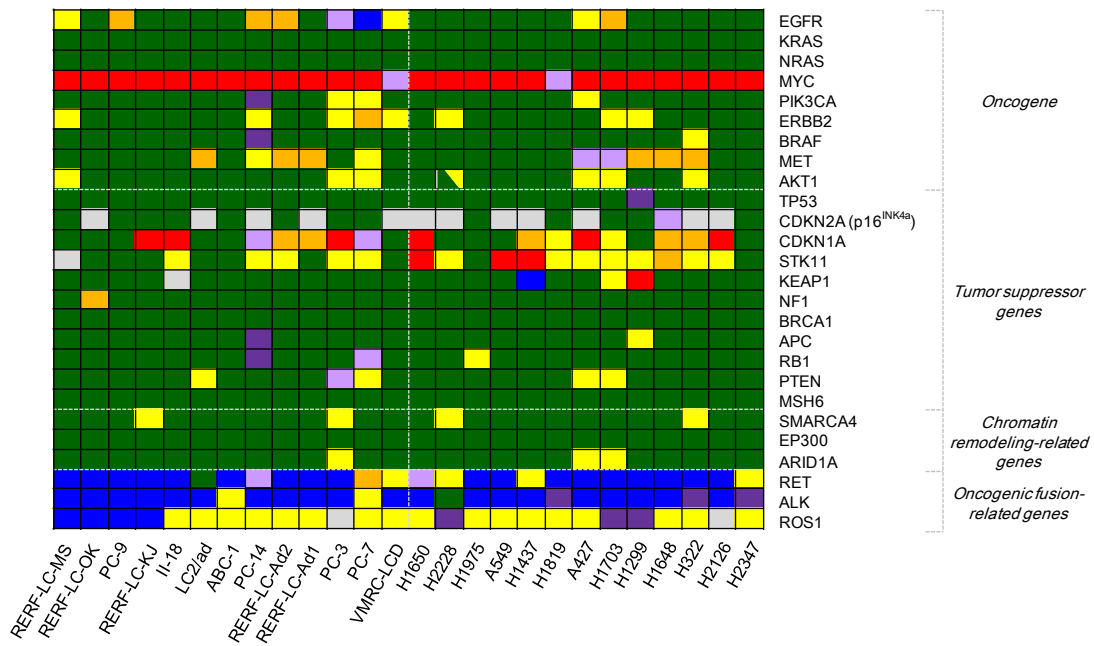
図 37 EGFR における遺伝子発現および転写活性性のクロマチンマークのパターン
EGFR の領域について、各細胞株における RNA-Seq および活性性クロマチンマークの
ChIP-Seq シグナルを IGV で可視化した。RNA-Seq の図中に RPKM を記載してある
(発現量が低い 3 細胞株については青字で示している)。赤字は EGFR のゲノム変異を
示している。図中の×は有意なシグナルが検出されていないことを示している。
(Supplementary Figure S21, Suzuki *et al.* 2014 *Nucleic Acids Research*
42(22):13557-13572 doi: 10.1093/nar/gku885)

1	Active promoter
2	Weak/poised promoter
3	Strong enhancer
4	Weak enhancer
5	Transcriptional elongation
6	Inactive region
7	Inactive region/heterochromatin
None	No significant signals

Representative promoters (± 1.5 Kb from TSS)



Gene bodies



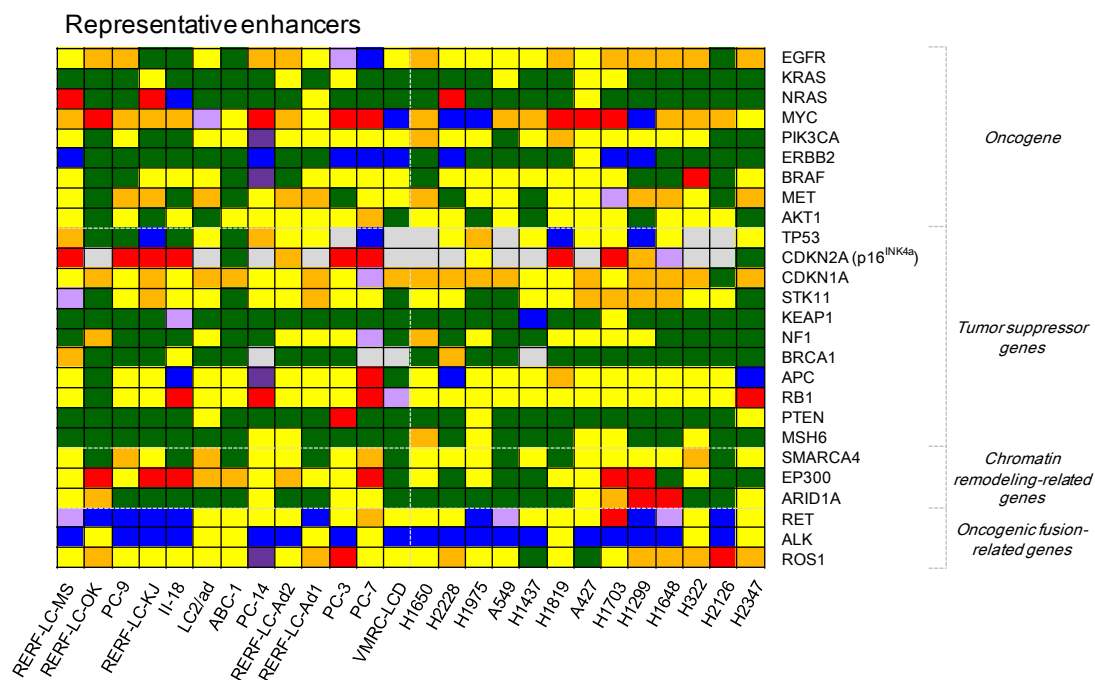


図 38 26 癌関連遺伝子におけるクロマチンステータス

26 個の癌関連遺伝子について、各細胞株におけるクロマチンステータスを図示している。細胞株は、図 13 と同じ順番に並べられている。各領域（プロモーター、転写領域、および、エンハンサー）において、ステータス 1 から 7 のうち最頻出のステータスを図に示している。有意なステータスがない領域に関しては灰色で示している。

(Figure 4C, Suzuki *et al.* 2014 *Nucleic Acids Research* 42(22):13557-13572 doi: 10.1093/nar/gku885)

3-3-5. 癌細胞におけるマルチオミクス統合解析

26 肺腺癌細胞株における全ゲノムシーケンス、RNA-Seq、バイサルファイトシーケンス、ChIP-Seq 解析の結果、ゲノム、エピゲノムおよびトランスクリプトームにおける多様な異常が検出された。遺伝子発現・転写制御異常を誘起するゲノム・エピゲノム変異を同定するために、RNA-Seq 解析から遺伝子発現異常が検出された遺伝子について、多階層的な視点からその要因を説明すべく、該当領域でのゲノム変異、エピゲノム異常の検出を試みた。

発現低下が認められた癌関連遺伝子に着目して解析を行った。癌抑制遺伝子 STK11 (90)、CDKN2A p16^{INK4A} (91)、および、cyclin-dependent kinase inhibitor 1A (p21, Cip1) (CDKN1A) (92)の例を図 39-41 に示す。STK11 について、遺伝子発現の喪失が認められた 3 株 (RERF-LC-MS、II-18、A427) において、大規模なゲノム欠失が見られた (図 39)。これらのゲノム欠失においては共通して転写開始点が欠損していた。これら 3 株では、近傍の Pol II および H3K36me3 のシグナルも抽出されなかった。ゲノム欠失変異が転写開始反応を阻害し、遺伝子発現喪失に繋がっていると考えられた。また、H1437 と H2126 では転写開始点付近のゲノムは正常であったが、転写領域にゲノム欠失が見られた。一方、RERF-LC-KJ では、明確なゲノム欠失は見られないにも関わらず、発現低下 (平均発現量の 1/4 倍以下) が検出された。ChIP-Seq データを精査したところ、転写開始および Pol II のリクルートは行われるが、H3K36me3 のシグナルが転写領域の途中で失われていることが示された。RERF-LC-KJ では STK11 の転写領域の一部が重複を引き起こしている可能性が示唆され、このゲノム変異が原因で転写伸長が阻害されていると考えている。これらいずれの例においても、STK11 における異常な遺伝子発現低下は、ゲノム自体の変異に起因していた。

CDKN2A p16^{INK4A} については、13 株に大規模なゲノム欠失が見られた (図 40)。さらに 2 株においてナンセンスおよびフレームシフト変異が見られ、半数以上が発現低下に関わるゲノム変異を引き起こしていることが分かった。別の 6 株では、p16^{INK4A} 転写開始点付近における高 DNA メチル化が見られ、p16^{INK4A} の発現低下を引き起こしていた。CDKN2A p16^{INK4A} は、ゲノム変異に加えて高 DNA メチル化によっても、異常な発現低下を引き起こされている例が存在することが示された。

CDKN1A では、STK11 および CDKN2A p16^{INK4A} とのいずれの場合とも異なり、低発現株 (平均発現量の 1/16 倍以下: PC-3; 1/4 倍以下: II-18、PC-7、PC-14、VMRC-LCD、H1703、H1648 および H2126) において、ゲノム変異は検出されなかった。エピゲノム異常を確認したところ、図 41 に例示するように、PC-14 や PC-7 では転写抑制性のヒストン修飾である H3K27me3 および H3K9me3 が転写開始点付近で高いシグナル強度を示した。また VMRC-LCD においては、転写開始点付近に高 DNA メチル化が見られた。CDKN1A の発現低下は、ゲノム変異ではなくエピゲノム異常によってのみ生起していた。

これらの例が示すように、ゲノム変異単独、ゲノム変異および DNA メチル化異常の組み合わせ、あるいはゲノム変異を伴わないヒストン修飾異常および DNA メチル化異常の組み合わせ等、遺伝子発現低下を引き起こす多様なゲノム・エピゲノム変異が見出された。興味深いことに、それらは遺伝子によって特徴的であると考えられた。

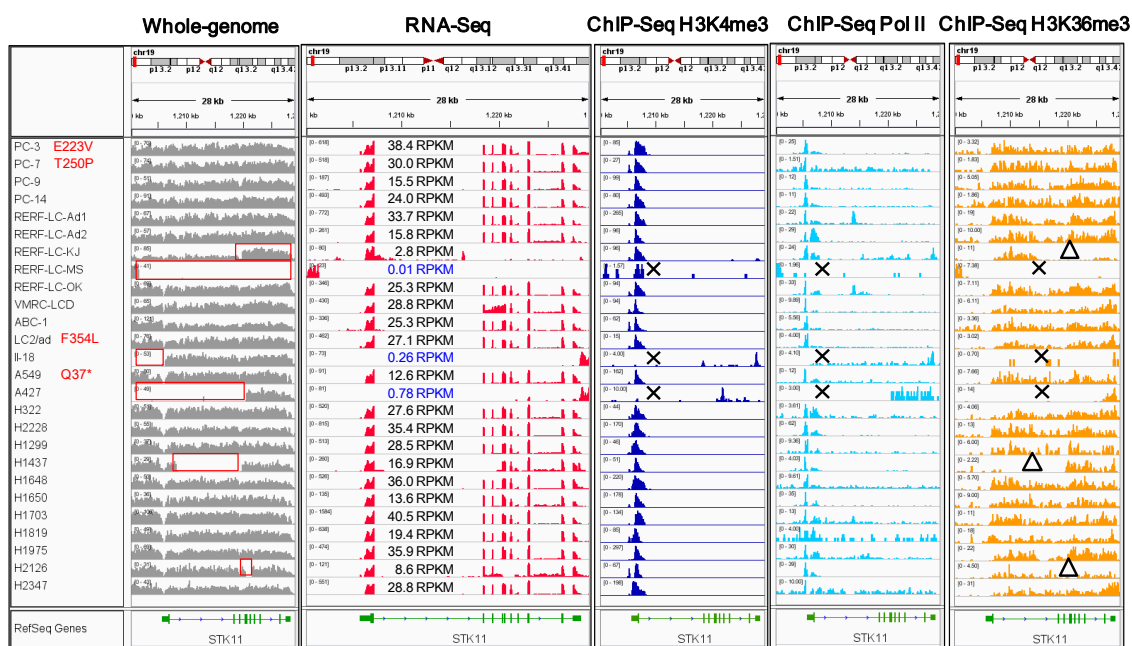


図 39 STK11 におけるゲノム変異と、エピゲノムおよびトランスクリプトームステータス

STK11 の領域について、各細胞株における全ゲノムシーケンス、RNA-Seq および活性性クロマチンマークの ChIP-Seq のシグナルを IGV で可視化した。全ゲノムシーケンスの図中に赤線で囲まれた領域でゲノム欠失および重複が起こっている。RNA-Seq の図中に RPKM を記載してある（特に発現量が低い 3 細胞株については青字で示している）。赤字は STK11 の non-synonymous SNVs を示している。図中の×は有意なシグナルが検出されていないことを示している。△はゲノム異常によって H3K36me3 のシグナルが欠失している部分を示す。

(Supplementary Figure S22, Suzuki *et al.* 2014 *Nucleic Acids Research* 42(22):13557-13572 doi: 10.1093/nar/gku885)

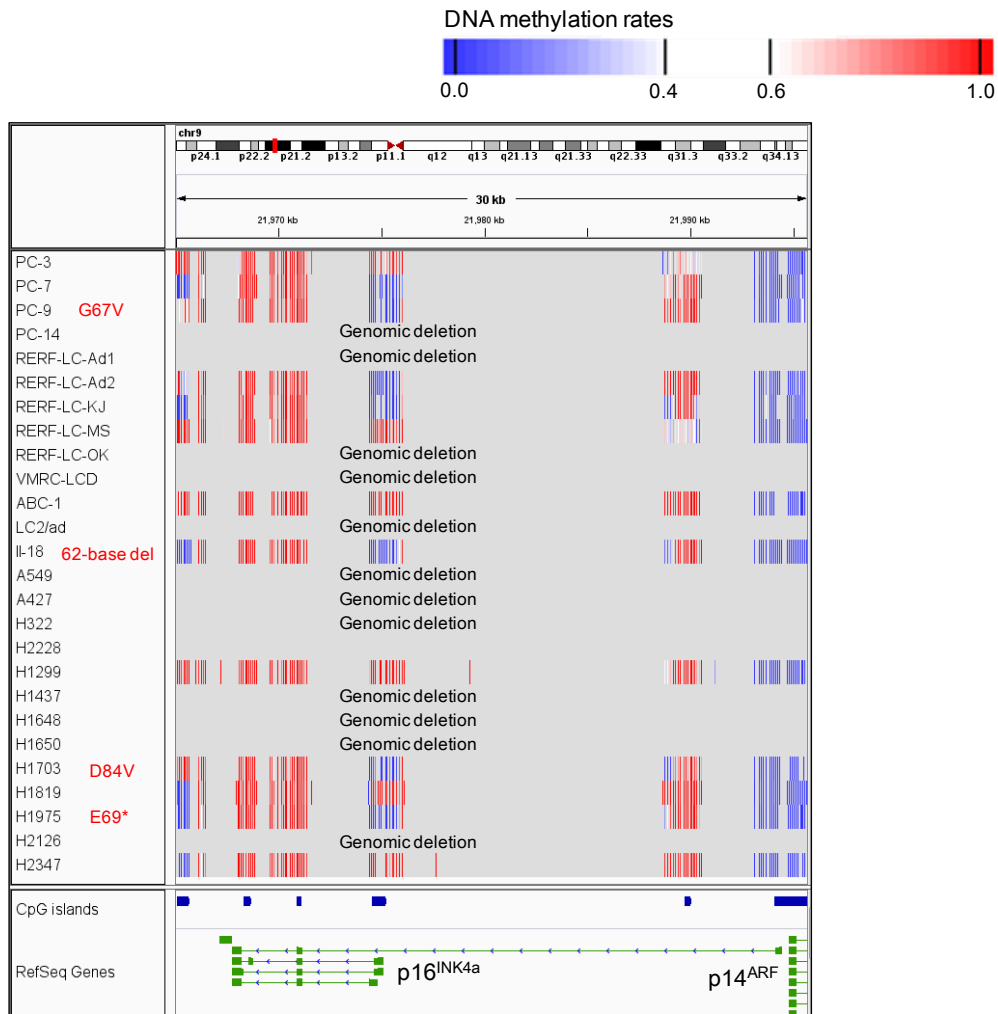


図 40 CDKN2A p16^{INK4}におけるゲノム変異および DNA メチル化異常
 CDKN2A の領域について、各細胞株における DNA メチル化状態を IGV で可視化した。
 13 株ではゲノム欠失が起こっており、DNA メチル化率が算出されていない。赤字は
 CDKN2A p16^{INK4A} の non-synonymous SNVs および indels を示している。
 (Figure 3D, Suzuki *et al.* 2014 *Nucleic Acids Research* 42(22):13557-13572 doi:
 10.1093/nar/gku885)

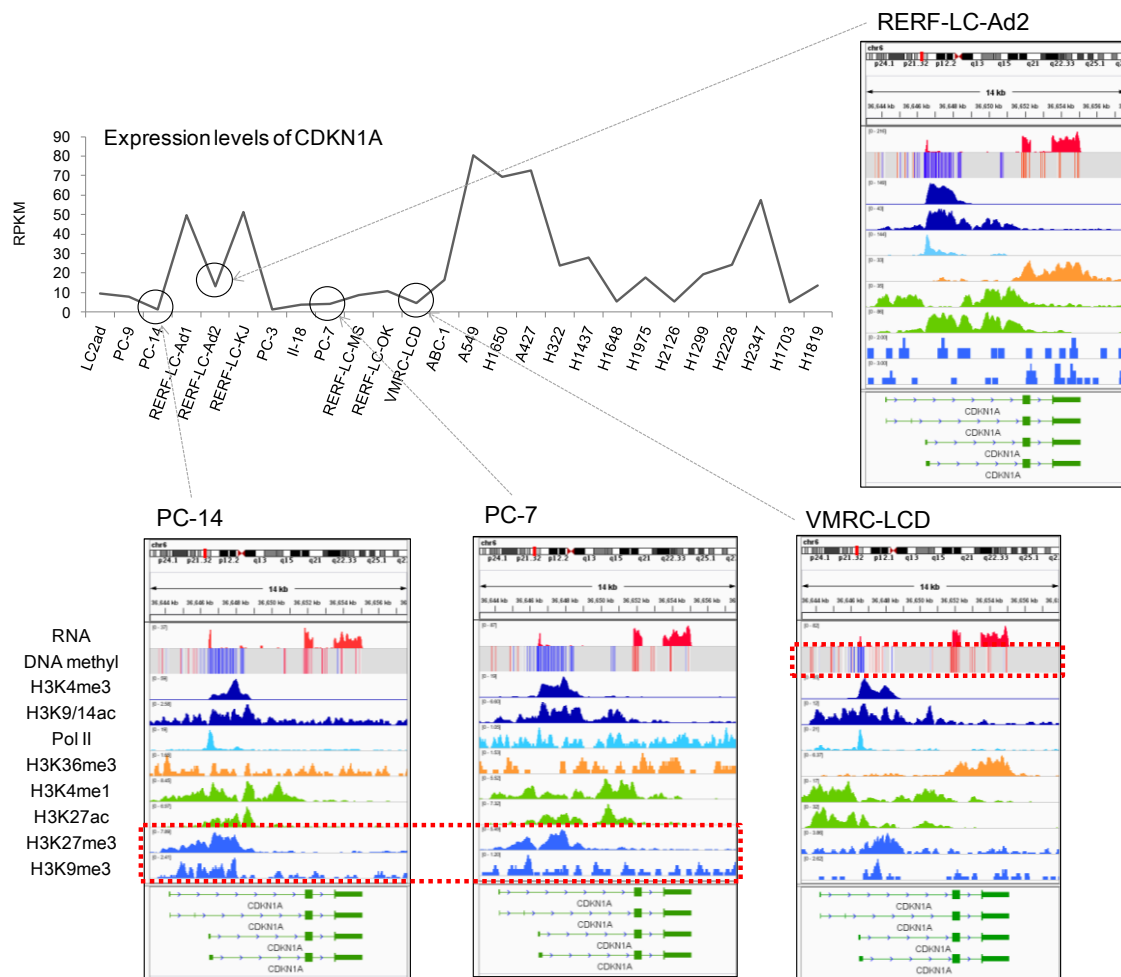


図 41 CDKN1A におけるエピゲノム異常

グラフは各細胞株における CDKN1A の発現量を示している。日本人由来の細胞株のうち、発現量が低い PC-14、PC-7 および VMRC-LCD と、発現量が中程度の RERF-LC-Ad2 について、トランスクリプトームとエピゲノムステータスを IGV で示した。各細胞株においてエピゲノム異常が起こっている部分を赤い点線で示した。

(Figure 5, Suzuki *et al.* 2014 *Nucleic Acids Research* 42(22):13557-13572 doi:10.1093/nar/gku885)

細胞株ごとに特徴的なエピゲノム異常のパターンを俯瞰するために、それぞれの細胞株において顕著なエピゲノム異常、特に転写抑制因子について解析した(93)。各細胞株において、転写開始点付近の DNA メチル化、H3K27me3 および H3K9me3 のシグナルが高い遺伝子に着目し、その数を細胞株ごとに算出した (図 42)。その結果、PC-7 および II-18 においては H3K27me3 と H3K9me3 といった抑制性のヒストン修飾について強いシグナルが検出された。一方で H1299 においては、H3K9me3 と DNA メチル化が顕著であることが明らかとなった。また、VMRC-LCD、A427 および H1819 においては H3K27me3 と DNA メチル化が特徴的であった。転写抑制に関わるエピゲノム異常は細胞株によって大きく異なり、それぞれに特徴的であることが明らかになった。

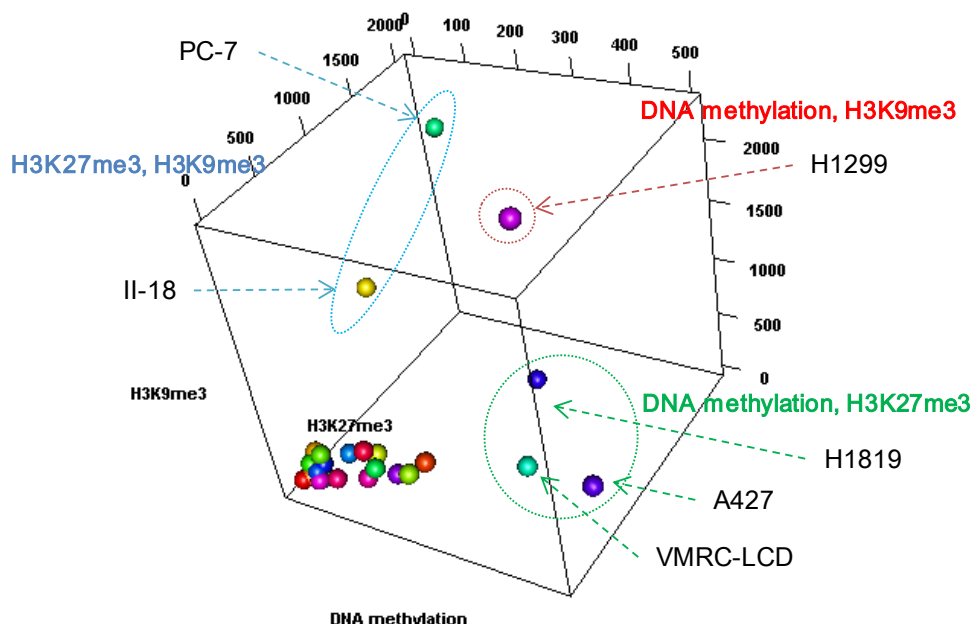


図 42 各細胞株における転写抑制性的なエピゲノム異常の特徴

各細胞株において、DNA メチル化、H3K27me3 および H3K9me3 のシグナルが平均値よりも 4 倍以上高い遺伝子の数を算出し、図示した。各プロットはそれぞれの細胞株を示している。

(Supplementary Figure S19, Suzuki *et al.* 2014 *Nucleic Acids Research* 42(22):13557-13572 doi: 10.1093/nar/gku885)

細胞株におけるゲノム、エピゲノムおよびトランスクリプトームの異常の生物学的意義あるいは細胞表現型との相関を明らかにするために、Hanahan と Weinberg が提唱した”hallmarks of cancer” (61)をもとに、各 hallmark について、マルチオミクス階層上での異常を投影した。先行研究に基づいて(62)、10 個の hallmark に合計 2050 遺伝子を割り当て、それぞれに生起するゲノム、エピゲノムおよびトランスクリプトームの異常（ゲノム変異、differential DNA methylation、differential chromatin marks、differentially expressed genes）を抽出、hallmark ごとに異常を持つ遺伝子の割合を検討して、そのパターンを hallmark ごとにあるいは細胞株間で比較解析した（図 43 および 44）。”Enabling Replicative Immortality”と”Genome Instability and Mutation”については、他の hallmark に比べてトランスクリプトーム階層において異常を示す遺伝子の割合が少なかった。また、”Avoiding Immune Destruction”ではゲノム階層において変異の割合が少なく、遺伝子発現異常の割合が高いという異なる傾向が見られた。

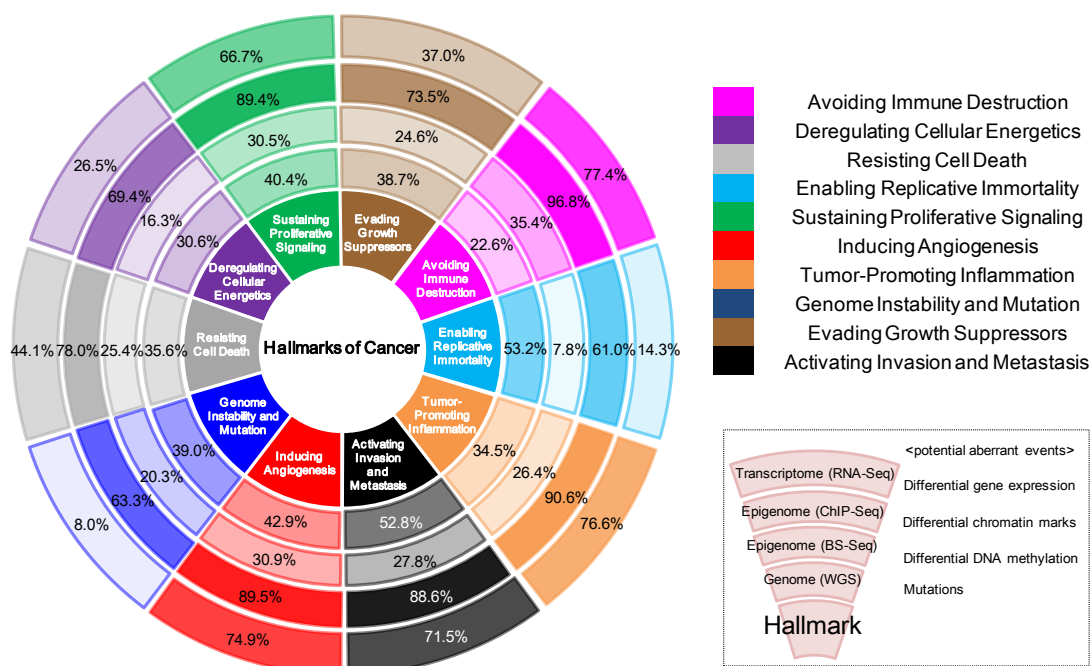
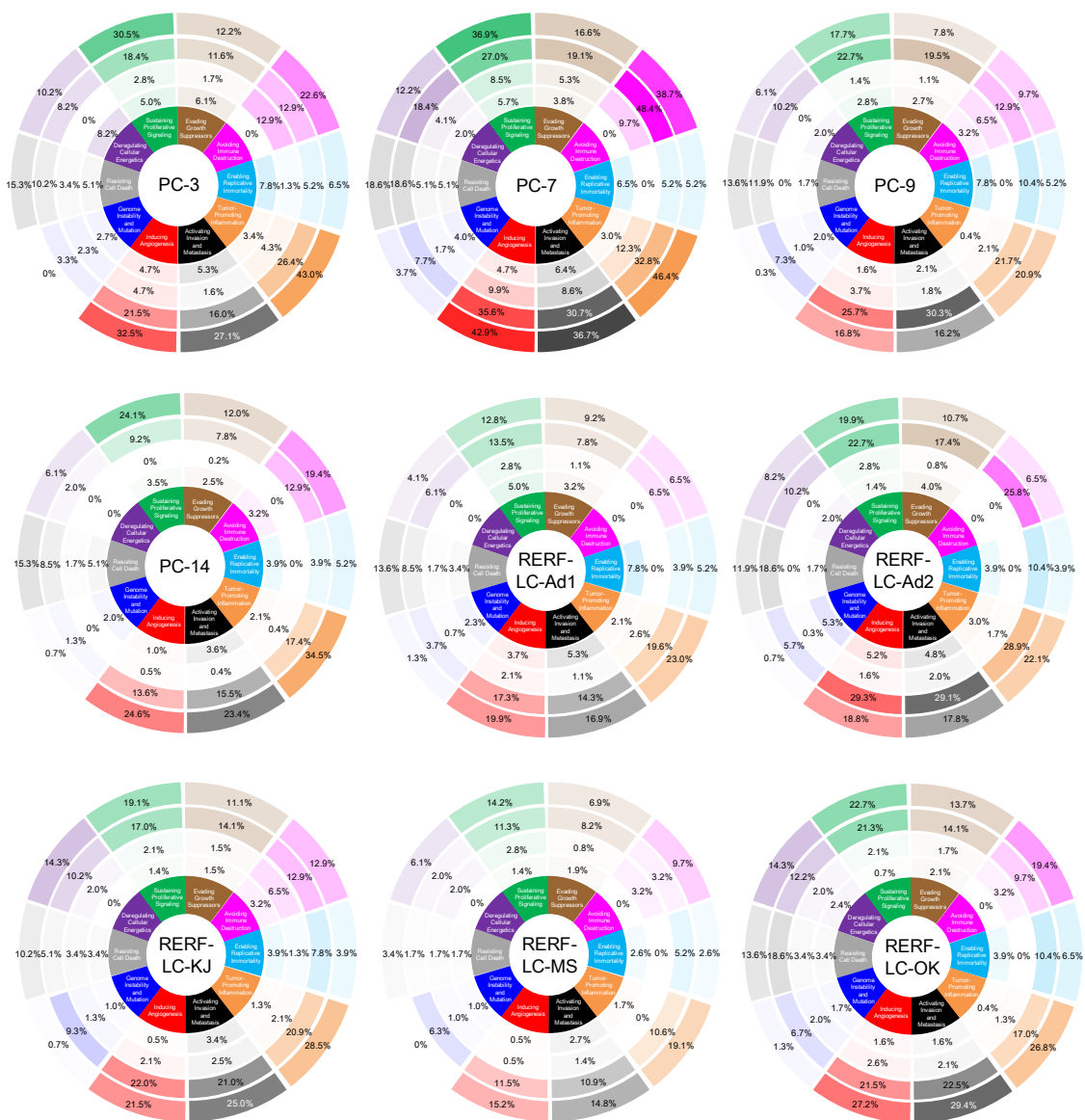


図 43 26 肺腺癌細胞株におけるマルチオミクス”hallmarks of cancer”
 10 個の”hallmarks of cancer”について、26 細胞株におけるゲノム変異、differential DNA methylation、differential chromatin marks および differentially expressed genes の割合を記載した。
 (Supplementary Figure S23A, Suzuki *et al.* 2014 *Nucleic Acids Research* 42(22):13557-13572 doi: 10.1093/nar/gku885)

図 44 に示すように、VMRC-LCD において”Deregulating Cellular Energetics”については、遺伝子発現異常を示す遺伝子が存在していたが、他の細胞株と異なり DNA メチル化およびクロマチンの階層での異常は検出されなかった。このことから、この hallmark は VMRC-LCD においては他の細胞株とは異なった遺伝子発現制御異常が生起している可能性が示唆された。PC-3 と PC-7 において、”Avoiding Immune Destruction”のゲノム・エピゲノム異常の割合に着目した。この hallmark において、PC-3 および PC-7 はともにゲノム変異は検出されなかった。エピゲノム異常においては、PC-7 ではクロマチンステータスの異常の割合が多いのに対して、PC-3 ではクロマチンステータスと DNA メチル化の異常は同程度であった。Hallmark ごと、細胞株ごとに固有のこれら特徴量に着目することにより、それぞれの由来する症例の癌化のメカニズムの解明に対して、またその治療戦略の策定に対して有用な知見が得られるのではないかと考えている。



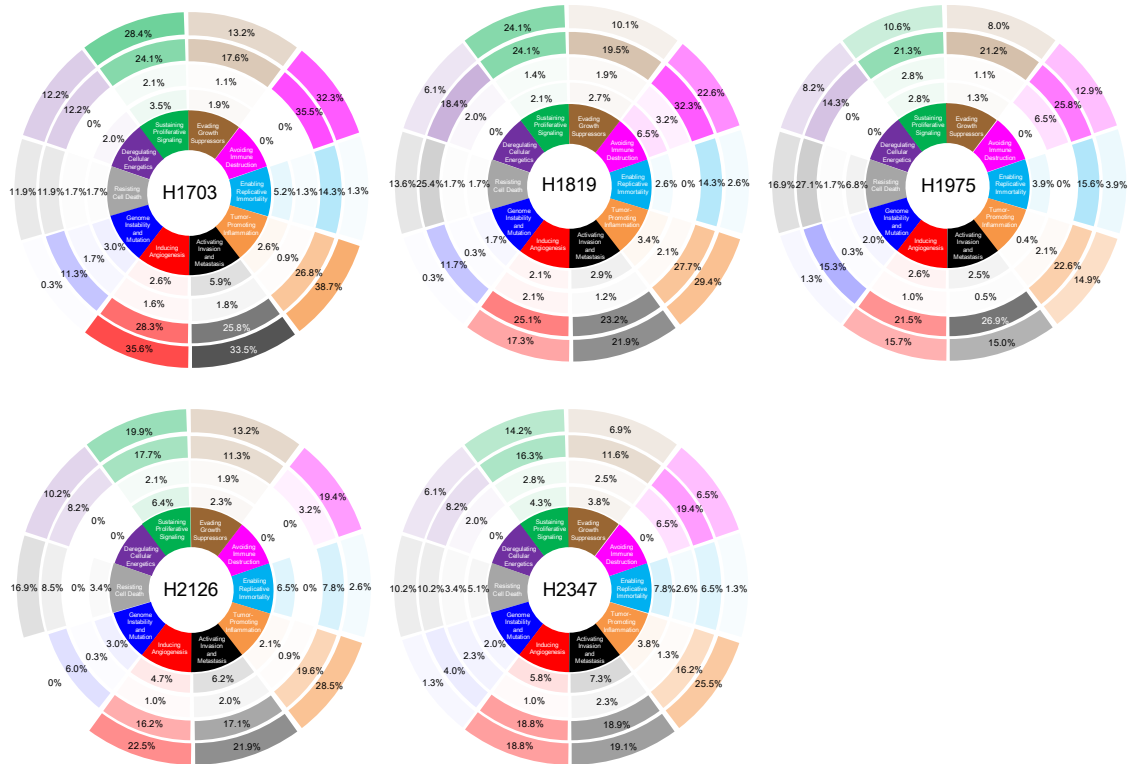


図 44 各細胞株におけるマルチオミクス”hallmarks of cancer”

各細胞株について、図 43 と同様に、ゲノム変異、differential DNA methylation、differential chromatin marks および differentially expressed genes の割合を記載した。円の中心に細胞株名が記載してある。

(Supplementary Figure S23B, Suzuki *et al.* 2014 *Nucleic Acids Research* 42(22):13557-13572 doi: 10.1093/nar/gku885)

これまでの解析結果は、癌細胞株の特徴量を相互に比較したものであったが、さらに、肺腺癌細胞株の癌細胞としての特徴を見出すために、トランスクリプトーム (RNA-Seq) およびエピゲノム (ChIP-Seq) について、正常ヒト小気道上皮細胞 (SAEC ; small airway epithelial cells) から同様のデータを取得し、26 癌細胞株のデータと比較した。SAEC の RNA-Seq と ChIP-Seq データに関する統計について表 22 に示す。SAEC に比して癌細胞株において高いまたは低い ChIP-Seq シグナルおよび発現量を示した遺伝子を抽出し、クロマチンステータス異常および遺伝子発現異常として、10 個の hallmark 上に投影した (図 45A)。その結果、癌細胞株間での比較において相互に発現変動を示す遺伝子の濃縮が認められなかった”Enabling Replicative Immortality” と”Genome Instability and Mutation”について、SAEC に比して癌細胞株で遺伝子発現変動を示す遺伝子の濃縮が認められた。これら 2 つの hallmark には、癌細胞株間で比較すると発現量の差は少ないが、正常細胞に比べて癌細胞株で発現亢進が認められる遺伝子が濃縮していた (図 45B)。SAEC のデータをさらに加えることで、個々の癌細胞株で相互に異なる特徴に加え、正常細胞と比較して癌細胞の特徴をとらえることも可能となった。

表 22 SAEC の RNA-Seq および ChIP-Seq

A. RNA-Seq

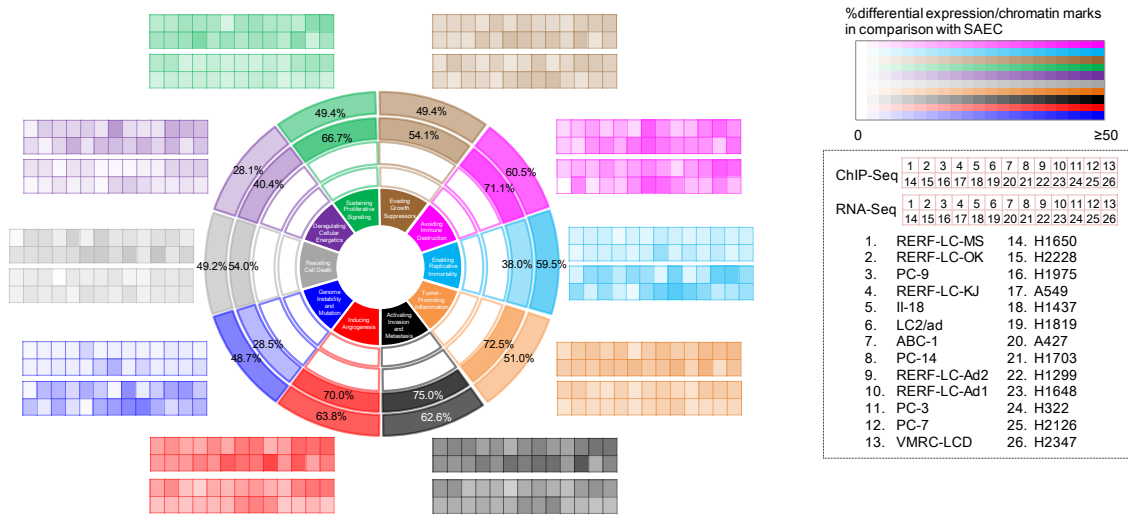
Used sequences (Read1)	Num of genes	
	> 1 RPKM	> 5 RPKM
180,054,144	12,126	8,809

B. ChIP-Seq

ChIP	Mapped sequences	Number of peaks (MACS2)	
		Narrow peaks	Broad peaks
H3K4me3	43,579,277	15,626	14,093
H3K9/14ac	21,603,337	50,674	45,159
Pol II	21,986,637	16,703	15,234
H3K36me3	56,493,935	321,485	107,145
H3K4me1	52,851,492	226,330	154,297
H3K27ac	45,848,952	170,013	88,659
H3K27me3	29,626,299	88,943	83,095
H3K9me3	40,496,823	316,142	148,544
WCE	45,429,763	***	***

(Supplementary Table S10 および S12B, Suzuki *et al.* 2014 *Nucleic Acids Research* 42(22):13557-13572 doi: 10.1093/nar/gku885 改変)

A



B

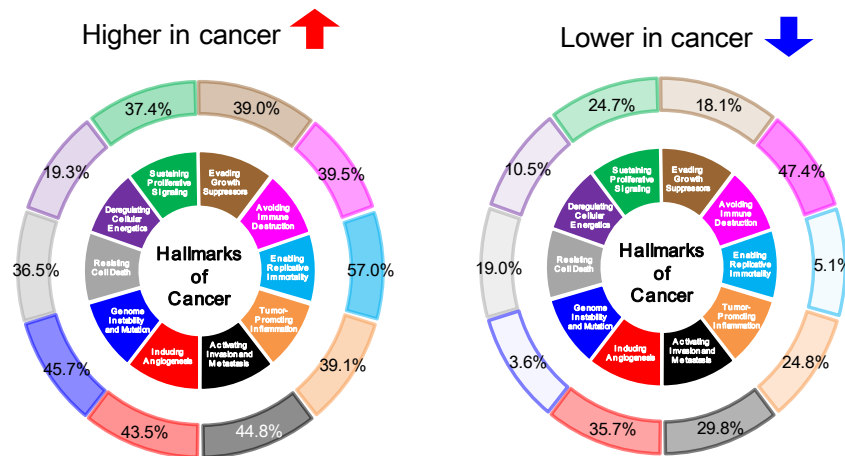


図 45 SAEC との比較

(A) 10 個の”hallmarks of cancer”について、SAEC と比較した際の 26 細胞株におけるクロマチンステータス異常および遺伝子発現異常の割合を記載した。周囲のヒートマップは各細胞株と SAEC を比較した際の異常の割合を示している（上：クロマチンステータス異常；下：遺伝子発現異常）。(B) SAEC と比べて 4 倍以上および 1/16 倍以下の発現量を示す遺伝子を、癌における高発現遺伝子および低発現遺伝子とし、各 hallmark についてその割合を記載した。

(Figure 6C および Supplementary Figure S24, Suzuki *et al.* 2014 *Nucleic Acids Research* 42(22):13557-13572 doi: 10.1093/nar/gku885)

3-4 小括

本研究では、26 種類の肺腺癌細胞株について、ゲノム、エピゲノムおよびトランスクリプトームシーケンスデータを取得し、癌細胞におけるトランスクリプトームの異常がどのようなゲノム・エピゲノム変異によるものなのか明らかにするために多階層的解析を試みた。肺腺癌という同一の癌種由来の細胞株 26 種類について、全ゲノムシーケンス、RNA-Seq、バイサルファイトシーケンスおよび 8 種類の ChIP-Seq データが、全て材料を同一とする細胞株から取得されているという点において、独自性の高いデータを産出することができたと考えている。

取得したオミクスデータを用いてマルチオミクス解析を行った結果、癌細胞において遺伝子発現異常を引き起こすゲノム・エピゲノム変異は、遺伝子によって特徴があることが明らかとなった。STK11 についてはゲノム変異、CDKN2A p16^{INK4A} についてはゲノム変異および DNA メチル化異常、CDKN1A については DNA メチル化異常およびヒストン修飾異常が遺伝子発現の異常低下を生起していた。他の遺伝子についても、遺伝子発現異常の原因となりうるゲノム・エピゲノム変異の特徴を明らかにするべく、詳細なパターン分類を行う必要がある。

特にエピゲノム異常については、細胞株によってその要因因子が大きく異なることも示された。例えば PC-7 では全体として H3K27me3 のシグナルが高く、PC-7 の多くの遺伝子に対してその発現低下に寄与していると考えられた。PC-7 ではヒストン修飾関連遺伝子 enhancer of zeste 2 polycomb repressive complex 2 subunit (EZH2) のゲノム変異が検出されており、この変異が H3K27me3 の亢進に関与しているとするれば興味深い。同様に多くの遺伝子について、細胞株ごとのエピゲノム、トランスクリプトームの異常との相関解析を行うことで、その成因となるゲノム変異を抽出する試みも可能かもしれない。各オミクス階層での異常を他のオミクス階層に成因を求めることができるという意味でも、本研究により同一材料を用いて作出されたデータは意義深い。

実際、本研究では、Hanahan と Weinberg が提唱した“hallmarks of cancer”を用いて、各細胞株のマルチオミクスデータと癌細胞における生物学的特徴を関連付けた。Hallmark を用いた解析については、従来、ゲノムもしくはトランスクリプトームのみといった単層での解析が主であったが、本研究では各 hallmark を多階層的な視点でとらえることができた。

本研究で用いた細胞株は、一般に広く遺伝子機能解析あるいは抗がん剤スクリーニング等に用いられている細胞株である。しかし、多くの場合、その細胞特性およびオミクス異常についての情報なしにそれらの実験が計画されるために、その有効性、普遍性が問題になる局面も多い。本研究で収集された情報は、さまざまな細胞株を用いた実験系、特に合理的な薬剤スクリーニングに向けて、重要な知見を与えると考えている。本研究

で取得および解析したマルチオミクスデータは、データベース DBTSS (<http://dbtss.hgc.jp/>) (63)より公開している。

総括

本研究では、日本人肺腺癌 97 症例の全エキソーム解析を行い、日本人固有のゲノム変異パターンを記載した。また、26 種類の肺腺癌細胞株を用いたマルチオミクスシーケンス解析によって、ゲノム解析だけでは不鮮明であった癌細胞のゲノム・エピゲノム変異と遺伝子発現制御異常の関係性について多くの知見を得た。

ゲノム、エピゲノムおよびトランスクリプトーム解析といった多階層的解析を行う上で、培養細胞系をモデル解析系として用いるのには多くの利点があると考えられる。得られた知見をいかにして臨床検体癌細胞での遺伝子発現・転写制御異常の解明へと応用するかは大きな課題である。

II 章で解析した日本人肺腺癌 97 症例と III 章で用いた 26 細胞株のゲノム変異を比較した結果、共通部位に引き起こされているゲノム変異は EGFR や KRAS、TP53 変異を除いてごく少数であった。しかし一方で、細胞株でも検出された EP300 や ARID1A といったエピゲノムおよびトランスクリプトームのパターンに影響を及ぼしうる遺伝子の変異は、臨床検体でも見出されている (図 46)。

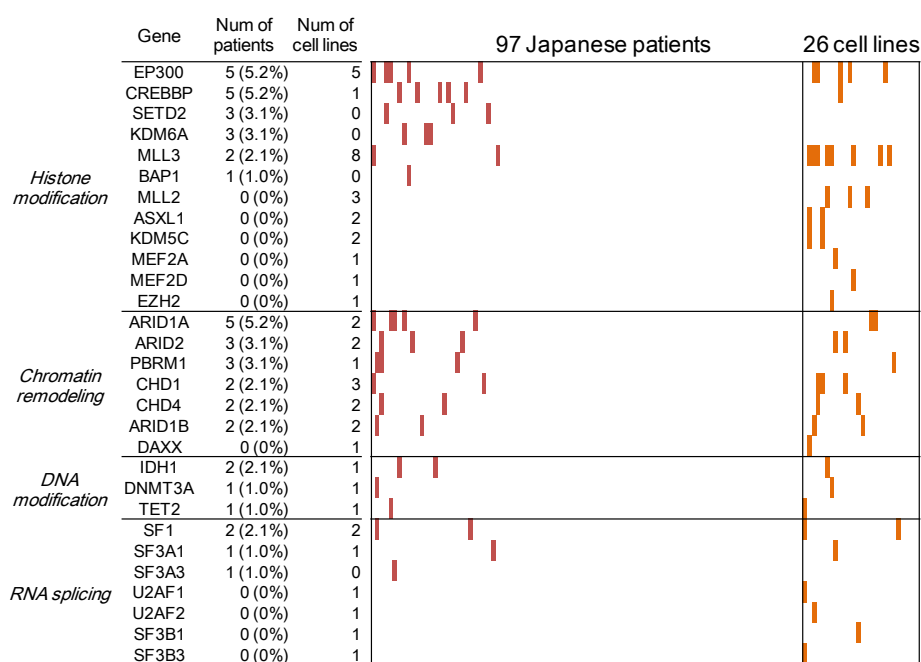


図 46 エピゲノム制御およびスプライシング関連遺伝子における変異

Watson *et al.* (94)の Figure 1 に記載のあった 42 個のエピゲノム制御およびスプライシング関連遺伝子のうち、日本人肺腺癌 97 症例または 26 肺腺癌細胞株で non-synonymous 変異が見られた 29 個の遺伝子について、変異の頻度を数値で示している。また、変異が見られた症例および細胞株を色で示している。

これらの変異を有する症例におけるエピゲノムおよびトランスクリプトーム状態を細胞株のデータから予測することができれば、臨床応用上有用な知見が得られる可能性がある。例えば ARID1A に変異を持つ 5 症例は、同様に ARID1A 変異陽性の RERF-LC-Ad2 や II-18（もしくは ARID1A の一部が欠失している VMRC-LCD）といった細胞株と似たクロマチン状態を示すかもしれない。また、LC2/ad で検出された U2AF1 変異(95)は、Imielinski *et al.* (11)や TCGA (13)の肺腺癌臨床検体で同一部位の変異が報告されており、LC2/ad のトランスクリプトームパターン（特にスプライシング情報）を U2AF1 変異陽性症例に写像することが可能ではないかと考えている。代表的ないくつかの遺伝子について検証実験を行うことが必要である。

細胞株から得られたマルチオミクスデータの臨床検体への写像を目指した情報学的手法の開発は意義深い課題である。培養細胞より得られた知見が、臨床検体を直接用いた解析では依然として困難な多階層的解析の端緒となることを期待する。

参考文献

1. The Cancer Genome Atlas Research Network, Weinstein, J.N., Collisson, E.A., Mills, G.B., Shaw, K.R., Ozenberger, B.A., Ellrott, K., Shmulevich, I., Sander, C. and Stuart, J.M. (2013) The Cancer Genome Atlas Pan-Cancer analysis project. *Nat Genet*, **45**, 1113-1120.
2. The Cancer Genome Atlas Research Network. (2011) Integrated genomic analyses of ovarian carcinoma. *Nature*, **474**, 609-615.
3. The Cancer Genome Atlas Research Network. (2012) Comprehensive genomic characterization of squamous cell lung cancers. *Nature*, **489**, 519-525.
4. International Cancer Genome Consortium. (2010) International network of cancer genome projects. *Nature*, **464**, 993-998.
5. Totoki, Y., Tatsuno, K., Yamamoto, S., Arai, Y., Hosoda, F., Ishikawa, S., Tsutsumi, S., Sonoda, K., Totsuka, H., Shirakihara, T. *et al.* (2011) High-resolution characterization of a hepatocellular carcinoma genome. *Nat Genet*, **43**, 464-469.
6. Totoki, Y., Tatsuno, K., Covington, K.R., Ueda, H., Creighton, C.J., Kato, M., Tsuji, S., Donehower, L.A., Slagle, B.L., Nakamura, H. *et al.* (2014) Trans-ancestry mutational landscape of hepatocellular carcinoma genomes. *Nat Genet*, **46**, 1267-1273.
7. Fujimoto, A., Totoki, Y., Abe, T., Boroevich, K.A., Hosoda, F., Nguyen, H.H., Aoki, M., Hosono, N., Kubo, M., Miya, F. *et al.* (2012) Whole-genome sequencing of liver cancers identifies etiological influences on mutation patterns and recurrent mutations in chromatin regulators. *Nat Genet*, **44**, 760-764.
8. Ding, L., Getz, G., Wheeler, D.A., Mardis, E.R., McLellan, M.D., Cibulskis, K., Sougnez, C., Greulich, H., Muzny, D.M., Morgan, M.B. *et al.* (2008) Somatic mutations affect key pathways in lung adenocarcinoma. *Nature*, **455**, 1069-1075.
9. Sharma, S.V., Bell, D.W., Settleman, J. and Haber, D.A. (2007) Epidermal growth factor receptor mutations in lung cancer. *Nat Rev Cancer*, **7**, 169-181.
10. El-Telbany, A. and Ma, P.C. (2012) Cancer genes in lung cancer: racial disparities: are there any? *Genes Cancer*, **3**, 467-480.
11. Imielinski, M., Berger, A.H., Hammerman, P.S., Hernandez, B., Pugh, T.J.,

- Hodis, E., Cho, J., Suh, J., Capelletti, M., Sivachenko, A. *et al.* (2012) Mapping the hallmarks of lung adenocarcinoma with massively parallel sequencing. *Cell*, **150**, 1107-1120.
12. Seo, J.S., Ju, Y.S., Lee, W.C., Shin, J.Y., Lee, J.K., Bleazard, T., Lee, J., Jung, Y.J., Kim, J.O., Yu, S.B. *et al.* (2012) The transcriptional landscape and mutational profile of lung adenocarcinoma. *Genome Res*, **22**, 2109-2119.
 13. The Cancer Genome Atlas Research Network. (2014) Comprehensive molecular profiling of lung adenocarcinoma. *Nature*, **511**, 543-550.
 14. Muratani, M., Deng, N., Ooi, W.F., Lin, S.J., Xing, M., Xu, C., Qamra, A., Tay, S.T., Malik, S., Wu, J. *et al.* (2014) Nanoscale chromatin profiling of gastric adenocarcinoma reveals cancer-associated cryptic promoters and somatically acquired regulatory elements. *Nat Commun*, **5**, 4361.
 15. Weinhold, N., Jacobsen, A., Schultz, N., Sander, C. and Lee, W. (2014) Genome-wide analysis of noncoding regulatory mutations in cancer. *Nat Genet*, **46**, 1160-1165.
 16. Huang, F.W., Hodis, E., Xu, M.J., Kryukov, G.V., Chin, L. and Garraway, L.A. (2013) Highly recurrent TERT promoter mutations in human melanoma. *Science*, **339**, 957-959.
 17. Mortazavi, A., Williams, B.A., McCue, K., Schaeffer, L. and Wold, B. (2008) Mapping and quantifying mammalian transcriptomes by RNA-Seq. *Nat Methods*, **5**, 621-628.
 18. Lister, R., Pelizzola, M., Dowen, R.H., Hawkins, R.D., Hon, G., Tonti-Filippini, J., Nery, J.R., Lee, L., Ye, Z., Ngo, Q.M. *et al.* (2009) Human DNA methylomes at base resolution show widespread epigenomic differences. *Nature*, **462**, 315-322.
 19. Furey, T.S. (2012) ChIP-seq and beyond: new and improved methodologies to detect and characterize protein-DNA interactions. *Nat Rev Genet*, **13**, 840-852.
 20. Bernstein, B.E., Stamatoyannopoulos, J.A., Costello, J.F., Ren, B., Milosavljevic, A., Meissner, A., Kellis, M., Marra, M.A., Beaudet, A.L., Ecker, J.R. *et al.* (2010) The NIH Roadmap Epigenomics Mapping Consortium. *Nat Biotechnol*, **28**, 1045-1048.
 21. Bernstein, B.E., Birney, E., Dunham, I., Green, E.D., Gunter, C., Snyder, M. and Consortium, E.P. (2012) An integrated encyclopedia of DNA elements in the human genome. *Nature*, **489**, 57-74.
 22. ENCODE Project Consortium. (2004) The ENCODE (ENCyclopedia Of DNA

- Elements) Project. *Science*, **306**, 636-640.
23. Suzuki, A., Mimaki, S., Yamane, Y., Kawase, A., Matsushima, K., Suzuki, M., Goto, K., Sugano, S., Esumi, H., Suzuki, Y. *et al.* (2013) Identification and characterization of cancer mutations in Japanese lung adenocarcinoma without sequencing of normal tissue counterparts. *PLoS One*, **8**, e73484.
 24. Li, H. and Durbin, R. (2009) Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics*, **25**, 1754-1760.
 25. Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G. and Durbin, R. (2009) The Sequence Alignment/Map format and SAMtools. *Bioinformatics*, **25**, 2078-2079.
 26. McKenna, A., Hanna, M., Banks, E., Sivachenko, A., Cibulskis, K., Kernytisky, A., Garimella, K., Altshuler, D., Gabriel, S., Daly, M. *et al.* (2010) The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res*, **20**, 1297-1303.
 27. DePristo, M.A., Banks, E., Poplin, R., Garimella, K.V., Maguire, J.R., Hartl, C., Philippakis, A.A., del Angel, G., Rivas, M.A., Hanna, M. *et al.* (2011) A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nat Genet*, **43**, 491-498.
 28. Sherry, S.T., Ward, M.H., Kholodov, M., Baker, J., Phan, L., Smigielski, E.M. and Sirotkin, K. (2001) dbSNP: the NCBI database of genetic variation. *Nucleic Acids Res*, **29**, 308-311.
 29. Fujimoto, A., Nakagawa, H., Hosono, N., Nakano, K., Abe, T., Boroevich, K.A., Nagasaki, M., Yamaguchi, R., Shibuya, T., Kubo, M. *et al.* (2010) Whole-genome sequencing and comprehensive variant analysis of a Japanese individual using massively parallel sequencing. *Nat Genet*, **42**, 931-936.
 30. The 1000 Genomes Project Consortium. (2010) A map of human genome variation from population-scale sequencing. *Nature*, **467**, 1061-1073.
 31. The 1000 Genomes Project Consortium. (2012) An integrated map of genetic variation from 1,092 human genomes. *Nature*, **491**, 56-65.
 32. Quevillon, E., Silventoinen, V., Pillai, S., Harte, N., Mulder, N., Apweiler, R. and Lopez, R. (2005) InterProScan: protein domains identifier. *Nucleic Acids Res*, **33**, W116-120.
 33. Forbes, S.A., Bhamra, G., Bamford, S., Dawson, E., Kok, C., Clements, J., Menzies, A., Teague, J.W., Futreal, P.A. and Stratton, M.R. (2008) The Catalogue of Somatic Mutations in Cancer (COSMIC). *Curr Protoc Hum*

Genet, **Chapter 10**, Unit 10 11.

34. Forbes, S.A., Bindal, N., Bamford, S., Cole, C., Kok, C.Y., Beare, D., Jia, M., Shepherd, R., Leung, K., Menzies, A. *et al.* (2011) COSMIC: mining complete cancer genomes in the Catalogue of Somatic Mutations in Cancer. *Nucleic Acids Res*, **39**, D945-950.
35. Kanehisa, M., Goto, S., Furumichi, M., Tanabe, M. and Hirakawa, M. (2010) KEGG for representation and analysis of molecular networks involving diseases and drugs. *Nucleic Acids Res*, **38**, D355-360.
36. Govindan, R., Ding, L., Griffith, M., Subramanian, J., Dees, N.D., Kanchi, K.L., Maher, C.A., Fulton, R., Fulton, L., Wallis, J. *et al.* (2012) Genomic landscape of non-small cell lung cancer in smokers and never-smokers. *Cell*, **150**, 1121-1134.
37. Wood, L.D., Parsons, D.W., Jones, S., Lin, J., Sjöblom, T., Leary, R.J., Shen, D., Boca, S.M., Barber, T., Ptak, J. *et al.* (2007) The genomic landscapes of human breast and colorectal cancers. *Science*, **318**, 1108-1113.
38. Degen, M., Brellier, F., Kain, R., Ruiz, C., Terracciano, L., Orend, G. and Chiquet-Ehrismann, R. (2007) Tenascin-W is a novel marker for activated tumor stroma in low-grade human breast cancer and influences cell behavior. *Cancer Res*, **67**, 9169-9179.
39. Degen, M., Brellier, F., Schenk, S., Driscoll, R., Zaman, K., Stupp, R., Tornillo, L., Terracciano, L., Chiquet-Ehrismann, R., Rüegg, C. *et al.* (2008) Tenascin-W, a new marker of cancer stroma, is elevated in sera of colon and breast cancer patients. *Int J Cancer*, **122**, 2454-2461.
40. Brellier, F., Martina, E., Degen, M., Heuzé-Vourc'h, N., Petit, A., Kryza, T., Courty, Y., Terracciano, L., Ruiz, C. and Chiquet-Ehrismann, R. (2012) Tenascin-W is a better cancer biomarker than tenascin-C for most human solid tumors. *BMC Clin Pathol*, **12**, 14.
41. Welch, H.C., Coadwell, W.J., Ellson, C.D., Ferguson, G.J., Andrews, S.R., Erdjument-Bromage, H., Tempst, P., Hawkins, P.T. and Stephens, L.R. (2002) P-Rex1, a PtdIns(3,4,5)P₃- and Gbetagamma-regulated guanine-nucleotide exchange factor for Rac. *Cell*, **108**, 809-821.
42. Donald, S., Hill, K., Lecureuil, C., Barnouin, R., Krugmann, S., John Coadwell, W., Andrews, S.R., Walker, S.A., Hawkins, P.T., Stephens, L.R. *et al.* (2004) P-Rex2, a new guanine-nucleotide exchange factor for Rac. *FEBS Lett*, **572**, 172-176.
43. Sosa, M.S., Lopez-Haber, C., Yang, C., Wang, H., Lemmon, M.A., Busillo,

- J.M., Luo, J., Benovic, J.L., Klein-Szanto, A., Yagi, H. *et al.* (2010) Identification of the Rac-GEF P-Rex1 as an essential mediator of ErbB signaling in breast cancer. *Mol Cell*, **40**, 877-892.
44. Berger, M.F., Hodis, E., Heffernan, T.P., Deribe, Y.L., Lawrence, M.S., Protopopov, A., Ivanova, E., Watson, I.R., Nickerson, E., Ghosh, P. *et al.* (2012) Melanoma genome sequencing reveals frequent PREX2 mutations. *Nature*, **485**, 502-506.
 45. Bartkova, J., Horejsi, Z., Koed, K., Kramer, A., Tort, F., Zieger, K., Guldborg, P., Sehested, M., Nesland, J.M., Lukas, C. *et al.* (2005) DNA damage response as a candidate anti-cancer barrier in early human tumorigenesis. *Nature*, **434**, 864-870.
 46. Dasika, G.K., Lin, S.C., Zhao, S., Sung, P., Tomkinson, A. and Lee, E.Y. (1999) DNA damage-induced cell cycle checkpoints and DNA strand break repair in development and tumorigenesis. *Oncogene*, **18**, 7883-7899.
 47. Li, C., Fang, R., Sun, Y., Han, X., Li, F., Gao, B., Iafrate, A.J., Liu, X.Y., Pao, W., Chen, H. *et al.* (2011) Spectrum of oncogenic driver mutations in lung adenocarcinomas from East Asian never smokers. *PLoS One*, **6**, e28204.
 48. Tanimoto, K., Tsuchihara, K., Kanai, A., Arauchi, T., Esumi, H., Suzuki, Y. and Sugano, S. (2010) Genome-wide identification and annotation of HIF-1 α binding sites in two cell lines using massively parallel sequencing. *Hugo J*, **4**, 35-48.
 49. Kanai, A., Suzuki, K., Tanimoto, K., Mizushima-Sugano, J., Suzuki, Y. and Sugano, S. (2011) Characterization of STAT6 target genes in human B cells and lung epithelial cells. *DNA Res*, **18**, 379-392.
 50. Boeva, V., Zinovyev, A., Bleakley, K., Vert, J.P., Janoueix-Lerosey, I., Delattre, O. and Barillot, E. (2011) Control-free calling of copy number alterations in deep-sequencing data using GC-content normalization. *Bioinformatics*, **27**, 268-269.
 51. Boeva, V., Popova, T., Bleakley, K., Chiche, P., Cappo, J., Schleiermacher, G., Janoueix-Lerosey, I., Delattre, O. and Barillot, E. (2012) Control-FREEC: a tool for assessing copy number and allelic content using next-generation sequencing data. *Bioinformatics*, **28**, 423-425.
 52. Untergasser, A., Nijveen, H., Rao, X., Bisseling, T., Geurts, R. and Leunissen, J.A. (2007) Primer3Plus, an enhanced web interface to Primer3. *Nucleic Acids Res*, **35**, W71-74.
 53. Robinson, J.T., Thorvaldsdottir, H., Winckler, W., Guttman, M., Lander, E.S.,

- Getz, G. and Mesirov, J.P. (2011) Integrative genomics viewer. *Nat Biotechnol*, **29**, 24-26.
54. Thorvaldsdóttir, H., Robinson, J.T. and Mesirov, J.P. (2013) Integrative Genomics Viewer (IGV): high-performance genomics data visualization and exploration. *Brief Bioinform*, **14**, 178-192.
55. Kim, D., Pertea, G., Trapnell, C., Pimentel, H., Kelley, R. and Salzberg, S.L. (2013) TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. *Genome Biol*, **14**, R36.
56. Kim, D. and Salzberg, S.L. (2011) TopHat-Fusion: an algorithm for discovery of novel fusion transcripts. *Genome Biol*, **12**, R72.
57. Zhang, Y., Liu, T., Meyer, C.A., Eeckhoute, J., Johnson, D.S., Bernstein, B.E., Nusbaum, C., Myers, R.M., Brown, M., Li, W. *et al.* (2008) Model-based analysis of ChIP-Seq (MACS). *Genome Biol*, **9**, R137.
58. Feng, J., Liu, T., Qin, B., Zhang, Y. and Liu, X.S. (2012) Identifying ChIP-seq enrichment using MACS. *Nat Protoc*, **7**, 1728-1740.
59. Ernst, J. and Kellis, M. (2012) ChromHMM: automating chromatin-state discovery and characterization. *Nat Methods*, **9**, 215-216.
60. Ernst, J., Kheradpour, P., Mikkelsen, T.S., Shores, N., Ward, L.D., Epstein, C.B., Zhang, X., Wang, L., Issner, R., Coyne, M. *et al.* (2011) Mapping and analysis of chromatin state dynamics in nine human cell types. *Nature*, **473**, 43-49.
61. Hanahan, D. and Weinberg, R.A. (2011) Hallmarks of cancer: the next generation. *Cell*, **144**, 646-674.
62. Hnisz, D., Abraham, B.J., Lee, T.I., Lau, A., Saint-André, V., Sigova, A.A., Hoke, H.A. and Young, R.A. (2013) Super-enhancers in the control of cell identity and disease. *Cell*, **155**, 934-947.
63. Suzuki, A., Wakaguri, H., Yamashita, R., Kawano, S., Tsuchihara, K., Sugano, S., Suzuki, Y. and Nakai, K. (2014) DBTSS as an integrative platform for transcriptome, epigenome and genome sequence variation data. *Nucleic Acids Res.*
64. Quinlan, A.R. and Hall, I.M. (2010) BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics*, **26**, 841-842.
65. Alexandrov, L.B., Nik-Zainal, S., Wedge, D.C., Aparicio, S.A., Behjati, S., Biankin, A.V., Bignell, G.R., Bolli, N., Borg, A., Børresen-Dale, A.L. *et al.* (2013) Signatures of mutational processes in human cancer. *Nature*, **500**, 415-421.

66. Soda, M., Choi, Y.L., Enomoto, M., Takada, S., Yamashita, Y., Ishikawa, S., Fujiwara, S., Watanabe, H., Kurashina, K., Hatanaka, H. *et al.* (2007) Identification of the transforming EML4-ALK fusion gene in non-small-cell lung cancer. *Nature*, **448**, 561-566.
67. Ju, Y.S., Lee, W.C., Shin, J.Y., Lee, S., Bleazard, T., Won, J.K., Kim, Y.T., Kim, J.I., Kang, J.H. and Seo, J.S. (2012) A transforming KIF5B and RET gene fusion in lung adenocarcinoma revealed from whole-genome and transcriptome sequencing. *Genome Res*, **22**, 436-445.
68. Kohno, T., Ichikawa, H., Totoki, Y., Yasuda, K., Hiramoto, M., Nammo, T., Sakamoto, H., Tsuta, K., Furuta, K., Shimada, Y. *et al.* (2012) KIF5B-RET fusions in lung adenocarcinoma. *Nat Med*, **18**, 375-377.
69. Lipson, D., Capelletti, M., Yelensky, R., Otto, G., Parker, A., Jarosz, M., Curran, J.A., Balasubramanian, S., Bloom, T., Brennan, K.W. *et al.* (2012) Identification of new ALK and RET gene fusions from colorectal and lung cancer biopsies. *Nat Med*, **18**, 382-384.
70. Takeuchi, K., Soda, M., Togashi, Y., Suzuki, R., Sakata, S., Hatano, S., Asaka, R., Hamanaka, W., Ninomiya, H., Uehara, H. *et al.* (2012) RET, ROS1 and ALK fusions in lung cancer. *Nat Med*, **18**, 378-381.
71. Blanco, R., Iwakawa, R., Tang, M., Kohno, T., Angulo, B., Pio, R., Montuenga, L.M., Minna, J.D., Yokota, J. and Sanchez-Cespedes, M. (2009) A gene-alteration profile of human lung cancer cell lines. *Hum Mutat*, **30**, 1199-1206.
72. Bos, J.L. (1989) ras oncogenes in human cancer: a review. *Cancer Res*, **49**, 4682-4689.
73. Iwakawa, R., Kohno, T., Enari, M., Kiyono, T. and Yokota, J. (2010) Prevalence of human papillomavirus 16/18/33 infection and p53 mutation in lung adenocarcinoma. *Cancer Sci*, **101**, 1891-1896.
74. Sato, M., Takahashi, K., Nagayama, K., Arai, Y., Ito, N., Okada, M., Minna, J.D., Yokota, J. and Kohno, T. (2005) Identification of chromosome arm 9p as the most frequent target of homozygous deletions in lung cancer. *Genes Chromosomes Cancer*, **44**, 405-414.
75. Nagayama, K., Kohno, T., Sato, M., Arai, Y., Minna, J.D. and Yokota, J. (2007) Homozygous deletion scanning of the lung cancer genome at a 100-kb resolution. *Genes Chromosomes Cancer*, **46**, 1000-1010.
76. Kohno, T., Otsuka, A., Girard, L., Sato, M., Iwakawa, R., Ogiwara, H., Sanchez-Cespedes, M., Minna, J.D. and Yokota, J. (2010) A catalog of genes

- homozygously deleted in human lung cancer and the candidacy of PTPRD as a tumor suppressor gene. *Genes Chromosomes Cancer*, **49**, 342-352.
77. Elsässer, S.J., Allis, C.D. and Lewis, P.W. (2011) Cancer. New epigenetic drivers of cancers. *Science*, **331**, 1145-1146.
 78. Füllgrabe, J., Kavanagh, E. and Joseph, B. (2011) Histone onco-modifications. *Oncogene*, **30**, 3391-3403.
 79. Matsubara, D., Kanai, Y., Ishikawa, S., Ohara, S., Yoshimoto, T., Sakatani, T., Oguni, S., Tamura, T., Kataoka, H., Endo, S. *et al.* (2012) Identification of CCDC6-RET fusion in the human lung adenocarcinoma cell line, LC-2/ad. *J Thorac Oncol*, **7**, 1872-1876.
 80. Suzuki, M., Makinoshima, H., Matsumoto, S., Suzuki, A., Mimaki, S., Matsushima, K., Yoh, K., Goto, K., Suzuki, Y., Ishii, G. *et al.* (2013) Identification of a lung adenocarcinoma cell line with CCDC6-RET fusion gene and the effect of RET inhibitors in vitro and in vivo. *Cancer Sci*, **104**, 896-903.
 81. Jung, Y., Kim, P., Keum, J., Kim, S.N., Choi, Y.S., Do, I.G., Lee, J., Choi, S.J., Kim, S., Lee, J.E. *et al.* (2012) Discovery of ALK-PTPN3 gene fusion from human non-small cell lung carcinoma cell line using next generation RNA sequencing. *Genes Chromosomes Cancer*, **51**, 590-597.
 82. Liu, C., Karam, R., Zhou, Y., Su, F., Ji, Y., Li, G., Xu, G., Lu, L., Wang, C., Song, M. *et al.* (2014) The UPF1 RNA surveillance gene is commonly mutated in pancreatic adenosquamous carcinoma. *Nat Med*, **20**, 596-598.
 83. Irizarry, R.A., Ladd-Acosta, C., Wen, B., Wu, Z., Montano, C., Onyango, P., Cui, H., Gabo, K., Rongione, M., Webster, M. *et al.* (2009) The human colon cancer methylome shows similar hypo- and hypermethylation at conserved tissue-specific CpG island shores. *Nat Genet*, **41**, 178-186.
 84. Ehrlich, M. (2002) DNA methylation in cancer: too much, but also too little. *Oncogene*, **21**, 5400-5413.
 85. Belinsky, S.A. (2004) Gene-promoter hypermethylation as a biomarker in lung cancer. *Nat Rev Cancer*, **4**, 707-717.
 86. Jones, P.A. and Baylin, S.B. (2002) The fundamental role of epigenetic events in cancer. *Nat Rev Genet*, **3**, 415-428.
 87. Herman, J.G. and Baylin, S.B. (2003) Gene silencing in cancer in association with promoter hypermethylation. *N Engl J Med*, **349**, 2042-2054.
 88. Chi, P., Allis, C.D. and Wang, G.G. (2010) Covalent histone modifications--miswritten, misinterpreted and mis-erased in human cancers.

- Nat Rev Cancer*, **10**, 457-469.
89. Schuster-Böckler, B. and Lehner, B. (2012) Chromatin organization is a major influence on regional mutation rates in human cancer cells. *Nature*, **488**, 504-507.
 90. Sanchez-Cespedes, M., Parrella, P., Esteller, M., Nomoto, S., Trink, B., Engles, J.M., Westra, W.H., Herman, J.G. and Sidransky, D. (2002) Inactivation of LKB1/STK11 is a common event in adenocarcinomas of the lung. *Cancer Res*, **62**, 3659-3662.
 91. Andujar, P., Wang, J., Descatha, A., Galateau-Sallé, F., Abd-Alsamad, I., Billon-Galland, M.A., Blons, H., Clin, B., Danel, C., Housset, B. *et al.* (2010) p16INK4A inactivation mechanisms in non-small-cell lung cancer patients occupationally exposed to asbestos. *Lung Cancer*, **67**, 23-30.
 92. Abbas, T. and Dutta, A. (2009) p21 in cancer: intricate networks and multiple activities. *Nat Rev Cancer*, **9**, 400-414.
 93. Baylin, S.B. and Jones, P.A. (2011) A decade of exploring the cancer epigenome - biological and translational implications. *Nat Rev Cancer*, **11**, 726-734.
 94. Watson, I.R., Takahashi, K., Futreal, P.A. and Chin, L. (2013) Emerging patterns of somatic mutations in cancer. *Nat Rev Genet*, **14**, 703-718.
 95. Yoshida, K., Sanada, M., Shiraishi, Y., Nowak, D., Nagata, Y., Yamamoto, R., Sato, Y., Sato-Otsubo, A., Kon, A., Nagasaki, M. *et al.* (2011) Frequent pathway mutations of splicing machinery in myelodysplasia. *Nature*, **478**, 64-69.

論文目録

1. Ayako Suzuki, Sachiyo Mimaki, Yuki Yamane, Akikazu Kawase, Koutatsu Matsushima, Makito Suzuki, Koichi Goto, Sumio Sugano, Hiroyasu Esumi, Yutaka Suzuki and Katsuya Tsuchihara. Identification and characterization of cancer mutations in Japanese lung adenocarcinoma without sequencing of normal tissue counterparts. 2013 *PLoS One* 8(9): e73484. doi: 10.1371/journal.pone.0073484.
2. Ayako Suzuki, Hideki Makinoshima, Hiroyuki Wakaguri, Hiroyasu Esumi, Sumio Sugano, Takashi Kohno, Katsuya Tsuchihara and Yutaka Suzuki. Aberrant transcriptional regulations in cancers: genome, transcriptome and epigenome analysis of lung adenocarcinoma cell lines. 2014 *Nucleic Acids Research* 42(22): 13557-13572. doi: 10.1093/nar/gku885.
3. Ayako Suzuki, Hiroyuki Wakaguri, Riu Yamashita, Shin Kawano, Katsuya Tsuchihara, Sumio Sugano, Yutaka Suzuki and Kenta Nakai. DBTSS as an integrative platform for transcriptome, epigenome and genome sequence variation data. 2015 *Nucleic Acids Research* 43(Database issue): D87-D91. doi: 10.1093/nar/gku1080.

謝辞

日頃の研究および本論文の作成において指導して下さった東京大学新領域創成科学研究科情報生命科学専攻 鈴木穰教授に深く感謝いたします。また、指導教員としてご指導くださった同研究科メディカルゲノム専攻 菅野純夫教授に深く感謝いたします。

本論文審査の副査をお引き受けくださった東京大学医科学研究所 中井謙太教授、古川洋一教授、柴田龍弘教授、松田浩一准教授には貴重なご助言を多数いただきました。深く感謝いたします。

日本人肺腺癌 97 症例の全エクソームシーケンスのデータは、国立がん研究センターの土原一哉博士、江角浩安博士、後藤功一博士の研究チームより提供していただきました。同研究チームの先生方には定期ミーティングにて貴重なご助言を多数いただきました。

肺腺癌細胞株マルチオミクス解析は、国立がん研究センターの土原一哉博士、河野隆志博士、牧野嶋秀樹博士との共同研究として行われました。解析を進める際には何度もディスカッションに応じていただき、また、数多くのアドバイスをいただきました。

鈴木研究室 若栗浩幸博士、東京大学医科学研究所 中井謙太教授には、解析した肺腺癌細胞株のマルチオミクスデータをデータベース DBTSS にて公開していただきました。本当にありがとうございました。

細胞株から得られたシーケンスデータの解析は、東京大学医科学研究所ヒトゲノム解析センタースーパーコンピュータを用いて行いました。本研究は、独立行政法人日本学術振興会 (JSPS) 科研費 特別研究員奨励費 24・5406 の助成を受けて行われました。

今村聖実様、昆布恵美様、阿部佳澄様、登坂真紀子様、石川由似様、Jakub Makalowski 様には、シーケンシングおよび数多くの実験を行っていただきました。また、堀内映実様、富所布紗乃様、菊池泰司様、張捷様、井口富久美様、多田智史様には、シーケンスデータの解析を手助けしていただきました。ご助力くださった鈴木研究室および菅野研究室の皆様には深く感謝いたします。本当にありがとうございました。

最後に、いつも応援してくれた両親、妹、弟に心より感謝します。

鈴木 絢子