

## 論文の内容の要旨

論文題目 肺腺癌細胞におけるマルチオミクス解析

氏名 鈴木 絢子

### 本研究の概要

近年、国際がんゲノム解析コンソシアム ICGC や TCGA をはじめとした多くの研究グループが、がんゲノムの全ゲノム・エキソーム解析を報告している。本研究で扱う肺腺癌では、KRAS や EGFR 変異が頻出し、特に EGFR 変異陽性の患者がアジア人に多いことが示唆されている。しかし、癌化および予後因子に影響する遺伝子変異パターンは未解明な部分が多い。本研究では、日本人肺腺癌患者 97 症例の全エキソーム解析をおこない、日本人肺腺癌の数多くのゲノム変異を同定し、日本人肺腺癌特有のゲノム変異パターンを明らかにした。変異の多くは、その機能が未知であり、癌細胞の表現型に寄与する機能的な変異を抽出することは容易ではなかった。また、RNA スプライシングやクロマチンリモデリング関連遺伝子についての変異も確認され、ゲノムだけでなく、エピゲノムやトランスクリプトームをあわせて解析し、癌細胞の異常を多階層的にとらえることが重要であると考えられた。しかし、臨床検体を使用した研究では、これらマルチオミクスデータを同一検体内ですべて取得することは難しい。本研究ではさらに、腺癌細胞株をモデル系として、癌細胞のマルチオミクスシークエンスデータを取得し、ゲノム・エピゲノム変異と、遺伝子発現異常の関係性を明らかにすべく一連の解析を行った。その結果、癌関連遺伝子および細胞株固有の遺伝子発現異常およびそれらを制御するゲノム・エピゲノム変異パターンについてその一端を明らかにした。

### 材料と方法

日本人肺腺癌 97 症例の全エキソームシークエンスデータの取得と解析は、国立がん研究センター東病院との共同研究による。各症例の癌部および非癌部のシークエンスデータを用いて、体細胞突然変異の探索と解析を行った。また、26 種類の肺腺癌細胞株について、全ゲノムシークエンス、RNA-Seq、バイサルファイトシークエンス、8 種類の ChIP-Seq のデータを取得し解析を行った。鋳型の調整およびデータの一次解析は常法に従って行った。正常細胞 (SAEC) についても同様のデータを取得、解析した。

## 結果

### 日本人肺腺癌 97 症例の全エキソーム解析

日本人肺腺癌 97 症例の全エキソームシーケンスデータから、平均 112 個の体細胞突然変異を検出した。検出された体細胞突然変異のうち、症例間で同一部位および同一遺伝子上に引き起こされている変異は少なく、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った。体細胞突然変異が有意に蓄積しているタンパク質機能ドメインおよびパスウェイを抽出した。全エキソーム解析から、これら癌関連因子候補を抽出することはできたが、直接癌化に関わる因子を同定するのは依然として困難であった。

日本人肺腺癌では EGFR の変異が最も高頻出であり、56 症例で見られた。欧米人と比較したところ、EGFR 変異陽性は日本人に多く、KRAS 変異陽性は欧米人に多い傾向が見られた。また、TP53 変異の頻度に差はなかった。

### 26 肺腺癌細胞株のマルチオミクス解析

#### <全ゲノム解析>

全ゲノム解析では、細胞株あたり平均 149,209 個の体細胞突然変異候補を検出した。724 個はタンパク質コード領域（うち 536 個がアミノ酸置換を伴う）、17 個はスプライスサイト、710 個は転写開始点上流 500 塩基、1087 個は UTR において位置していた（表 1）。また、平均 143 個の遺伝子領域がコピー数増加、101 個の遺伝子領域がコピー数低下を示し、該当する領域でのゲノム増幅、欠失が示唆された。加えて合計で 552 個の染色体構造異常が検出された。

癌遺伝子については、4 株に既知

EGFR 変異、4 株に KRAS の G12 変異、

3 株に NRAS の Q61 変異が検出され、これらは相互に排他的であった。癌抑制遺伝子については、

23 株で TP53 変異が検出された。さらに 13 株において CDKN2A 領域のゲノム欠失が検出された。

また、SMARCA4 のゲノム欠失や RBM10 のスプライスサイト変異等、クロマチンリモデリングおよびスプライシング関連遺伝子の変異も見出された。これらゲノム変異が及ぼす影響を明らかにするために、以下に同一細胞株を材料に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。

表 1. 26 細胞株における SNVs と indels の数

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

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

### <トランスクリプトーム解析>

RNA-Seq データを用いて、遺伝子発現量を算出した結果、平均 12,290 遺伝子が >1 RPKM の発現量を示した。全ゲノム解析から得られたタンパク質コード領域の変異のうち、約半数が >1 RPKM の発現量を示す遺伝子上に見られた。これらの細胞株において未同定の機能獲得変異を探索するに際し、有意な発現を示す遺伝子内での変異に着目することが可能となった。また、スプライスサイト変異が RNA スプライシングに及ぼす影響を調べた。PC-7 において NF1 については、スプライスサイト変異周辺のエクソン 19 がスキップされており、スプライスサイト変異が RNA スプライシングの異常に寄与していることが示唆された。全ゲノム解析から見出された変異についてトランスクリプトーム上での機能的帰結が明らかになった点で興味深い。当該細胞株において他の細胞株に比して遺伝子発現亢進の認められる遺伝子を細胞あたり平均 352 個、発現低下の認められる遺伝子を細胞あたり平均 1967 個見出した。EGFR は 3 株で発現量が低く、細胞株によって発現量の差が顕著であった遺伝子であるのに対し、TP53 や BRAF は、細胞株によって発現量に有意な差はなかった。これら遺伝子、細胞株に固有な発現パターンがどのように制御されているのか、エピゲノム制御因子について解析を進めた。

### <エピゲノム解析>

バイサルファイトシーケンスデータを用いて、CpG サイトの DNA メチル化率を計算した。細胞株によって DNA メチル化率が異なる遺伝子を探索した結果、平均 118 個の高 DNA メチル化遺伝子、278 個の低 DNA メチル化遺伝子が抽出された。NRAS については、25 株は低 DNA メチル化を示すのに対し、H322 は高 DNA メチル化を示した。

次に、7 種類のヒストン修飾と Pol II の ChIP-Seq を行い、細胞株によってクロマチン状態が異なる遺伝子を探索した。EGFR において、RNA の発現が見られた 23 細胞株の転写開始点付近は活性型クロマチン状態を示した。また、発現が見られない 3 株のうち、2 株において H3K4me3 のピークが検出できた。Pol II は 1 株でのみ弱いピークが見られ、H3K36me3 はこれら 3 株のいずれにおいても検出されなかった。細胞株によって転写制御の各段階の状態が異なるのではないかと示唆された。

### <統合解析>

遺伝子発現異常とゲノム・エピゲノムの変異の相関について、癌関連遺伝子を中心に統合解析を行った。癌抑制遺伝子である STK11 では、発現量が低い 3 株は転写開始点付近のゲノムが欠失していた。STK11 の発現量低下は、ゲノム変異に起因していた。CDKN2A では、13 株においてゲノム欠失、6 株において p16<sup>INK4A</sup> 転写開始点付近の高 DNA メチル化が検出された。CDKN1A では、低発現を示す細胞株においてゲノム変異は抽出されなかった。PC-14 および PC-7 では強い抑制性のヒス

トン修飾、VMRC-LCD では転写開始点付近の高 DNA メチル化が検出された。CDKN1A の遺伝子発現低下は、ゲノム変異ではなくエピゲノムの異常に起因していた。以上のことから、結果的に異常な遺伝子発現を引き起こす要因因子としてのゲノム・エピゲノム変異には、細胞株によって特徴があることが示唆された。

さらに体系的に、各細胞株のゲノム・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との関連性を明らかにするために、Hanahan と Weinberg が提唱した”Hallmarks of cancer”をもとに、癌細胞の多階層的な特徴を記述した (図 1)。”Genome Instability and Mutation”や”Enabling Replicative Immortality”に関連する遺伝子の発現量は、正常細胞に比べて癌細胞株で

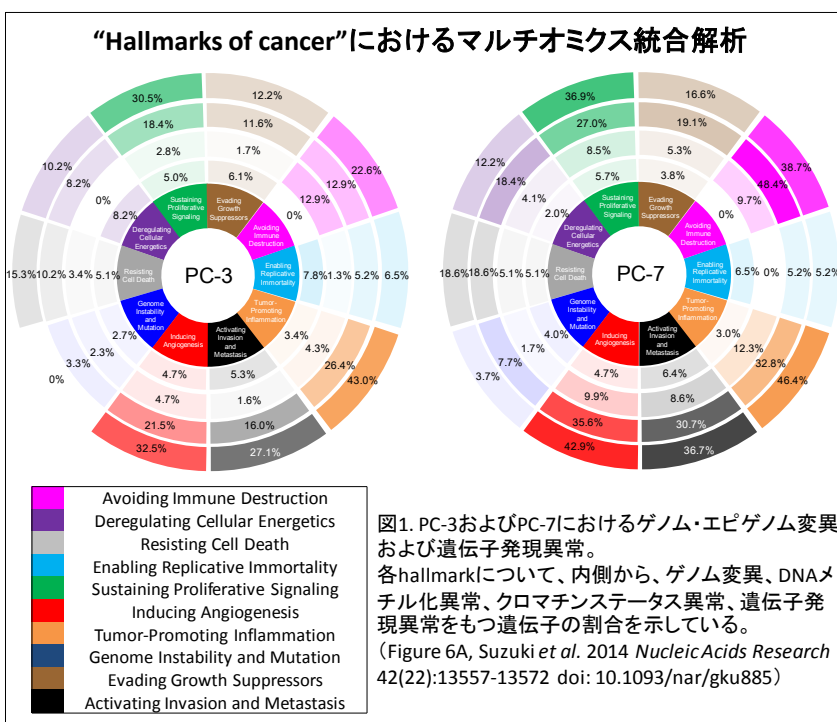


図1. PC-3およびPC-7におけるゲノム・エピゲノム変異および遺伝子発現異常。各hallmarkについて、内側から、ゲノム変異、DNAメチル化異常、クロマチンステータス異常、遺伝子発現異常をもつ遺伝子の割合を示している。(Figure 6A, Suzuki et al. 2014 *Nucleic Acids Research* 42(22):13557-13572 doi: 10.1093/nar/gku885)

高いが、癌細胞株間の差は少なかった。また、”Avoiding Immune Destruction”のエピゲノムステータスにおいて、PC-7ではクロマチンステータス異常が多いのに対し、PC-3ではDNAメチル化異常も多く、エピゲノム異常の状態は細胞株によって異なることが示唆された。各 hallmark について、細胞株によって異なる異常を網羅的に記述することは、多様な癌細胞の生物学的特性あるいは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。

## 結論

本研究では、日本人肺腺癌 97 症例の全エクソーム解析を行い、肺腺癌細胞のゲノム変異を数多く同定し、日本人肺腺癌特有のゲノム変異パターンを明らかにした。また、26 種類の肺腺癌細胞株におけるマルチオミクスシーケンスデータを用いて、臨床検体では困難である、癌細胞を多階層的な視点から解析する解析モデルを試行した。癌細胞の遺伝子発現・転写制御異常と、それらを引き起こすゲノム・エピゲノム変異をより詳細に明らかにするために、マルチオミクスデータをさまざまな角度から解析する有効性が示されたと考えている。