

令和元年6月24日現在

機関番号：14301

研究種目：基盤研究(C) (一般)

研究期間：2016～2018

課題番号：16K08640

研究課題名(和文)日本人におけるEGFR遺伝子変異型肺がんリスクに関連する遺伝子の解明

研究課題名(英文) Isolation of genes associated with EGFR mutation-positive lung cancer risk in the Japanese

研究代表者

高橋 めい子(児玉めい子)(Takahashi, Meiko)

京都大学・医学研究科・特定講師

研究者番号：20372592

交付決定額(研究期間全体)：(直接経費) 3,700,000円

研究成果の概要(和文)：EGFR、ALK、METなど多くの肺がん感受性マーカーが明らかになってきたことにより、今後は原因遺伝子に応じた治療が開発されることが期待される。日本人における肺がんリスクに対するEGFR遺伝子変異の寄与を検討するため、申請者らは最新の技術を用いて変異の有無を高速に解析するプロトコルを確立し、肺がん患者で変異の有無と肺がんの組織型・喫煙状況・性別などの相関関係を統計学的に解析した。日本人の肺がんのおよそ3-4割は、EGFR変異が原因で起こることが知られている。感受性遺伝子を特定し発症危険群を特定することは発症予防を可能とする研究になることが期待され大きな貢献となる。

研究成果の学術的意義や社会的意義

肺がんによる死亡者数は日本人男性では胃癌を抜いて第1位となり、女性でも大腸癌・胃癌に次いで3番目を占めている。本研究では、日本人における肺がんリスクに対するEGFR遺伝子変異の寄与を検討する。EGFR変異型肺がんリスクに関連する遺伝子は、その変異を伴う肺がんの頻度が欧米人とアジア人で異なることから、人種や遺伝的背景によって異なることが予想される。この知見により、日本人における肺がんの、新たな早期診断・予防・治療を可能とし、医学の観点から大きな貢献となる。分子遺伝情報に基づいた病体解明を通して、医学の究極的目的である臨床医学への展開につなげるための情報を収集していく。

研究成果の概要(英文)：Lung cancers can be divided into two main groups: small cell (SCLC) and non-small cell lung cancer (NSCLC). NSCLC can further be classified by tissue type into adenocarcinoma, squamous cell carcinoma and large cell carcinoma, and many biomarkers such as EGFR, ALK and MET have become evident in recent years. Personalized treatment for lung cancer patients is becoming a reality and its use will only increase in the future. In order to investigate the contribution of EGFR gene mutations to lung cancer risk in the Japanese, we have established a protocol for analyzing the presence or absence of EGFR mutations at high speed and low cost using the latest MiSeq sequencing technology. The correlation of the presence of EGFR mutation with clinicopathological factors such as histology of the tumor, smoking status, and sex of lung cancer patients was analyzed statistically.

研究分野：がん遺伝学

キーワード：EGFR変異 肺がん GWAS SNP MiSeq

## 1. 研究開始当初の背景

肺がんは典型的な複合遺伝性疾患で、環境要因、遺伝的要因が複雑に関連し合って発症する。環境要因としてタバコへの曝露が最も強い因子として知られているが、タバコへの曝露者全てが肺がん罹患者ではなく、何らかの遺伝的差異が存在する。近年、遺伝子多型検討技術の大幅な進歩により、少量の DNA で遺伝子のほぼ全領域をカバーする遺伝子多型が高速で解析できるようになった。本技術を応用した肺がんに関する研究から得られた重大な知見が 2008 年前半に相次いで欧米諸国から報告がなされている。

EGFR (Epidermal Growth Factor Receptor, 上皮増殖因子レセプター)は増殖因子が結合すると、細胞の中でチロシンにリン酸が結合し、信号が次々と蛋白質へ伝えられ、細胞の増殖が引き起こされる。正常な細胞では、EGFR に増殖因子が結合した時にのみ信号が伝わるが、一部の肺がんでは EGFR 遺伝子の突然変異により、増殖因子が結合しなくてもスイッチがいつでも ON の状態になる事が明らかとなっている。EGFR 遺伝子の突然変異の原因は明らかではないが、東アジア人、非喫煙者、女性、肺がんの中でも腺癌に高頻度に見られる事が分かっている。

これまでに申請者らは、肺がんの発症リスクに関連する遺伝的因子を同定するため、日本人集団の肺がん患者と対照群検体の全ゲノム関連解析 (GWAS) を実施している。この GWAS 結果を用いて、日本人における EGFR 遺伝子変異の有無を考慮した関連解析を実施する。

## 2. 研究の目的

本研究は日本人における EGFR 遺伝子変異型肺がんリスクに関連する遺伝子の解明を目的とし、以下のとおり 2 つのテーマから構成される。

- (1) まず申請者らは、EGFR 遺伝子変異を検出する実験系をデザインする。EGFR 遺伝子変異は細胞内のチロシンキナーゼドメインに集中しているが、特に頻度が高いものはエクソン 19 のコドン 746-750 を中心とする部位の欠失変異と、エクソン 21 のコドン 858 においてロイシンからアルギニンに変化する点突然変異 L858R である。しかしながら、まれな点突然変異等がエクソン 18 と 20 に少数認められるため、当初予定していたエクソン 19 と 21 に加えて 18 と 20 もシーケンスの対象に加えることにし、ローコストで実験可能なプロトコールを作成することにした。このプロトコールの確立により多検体から変異を起こしているものの抽出がこれまでと比べ飛躍的に効率的に行えるようになる。
- (2) 申請者らは既に日本人集団の肺がん患者と、対照群検体の全ゲノム関連解析 (GWAS) を実施している。この結果を用いて、日本人における EGFR 遺伝子変異の有無を考慮した関連解析、さらに肺がんの組織型・喫煙状況・性別など臨床病理学的因子との相関関係を統計学的に解析する。本課題が成功し EGFR 遺伝子変異型肺がんリスク感受性遺伝子の全貌を明らかにすることで、人種間や病態・環境因子・生活習慣における相対的貢献度を解明することが可能になる。

### 3. 研究の方法

本課題では最新の技術を用いてEGFRの変異の有無を高速に解析するプロトコルを確立し、肺がんの組織型と変異の有無や、喫煙状況、性別など臨床病理学的因子との相関関係を統計学的に解析する。これまでに報告されたEGFR 遺伝子変異の頻度は、東洋人(32%)・非東洋人(7%)、男性(10%)・女性(38%)、非喫煙者(47%)・喫煙者(7%)、腺癌(30%)・非腺癌(2%)など臨床背景に強く関連していることが分かっている。本解析に使用する検体の臨床情報は全て揃っているため、性別・喫煙情報等でのsubtype解析も実施する。

- (1) 96 穴プレートを使用したプロトコルをデザインし、肺がん組織由来 DNA を用いてチロシンキナーゼドメインに集中している EGFR 遺伝子変異同定のため、エクソン 18~21 の変異の有無を MiSeq を用いてシーケンスする。
- (2) 変異の有無と GWAS 結果をもとに subtype 解析を実施し、日本人における EGFR 遺伝子変異・欠失誘発の機構解明に寄与する。

### 4. 研究成果

#### (1) 新しいプロトコルの確立

まず、一度に 96 サンプル分の EGFR エクソン 18 から 21 までをローコストで増幅できるプロトコルを作った。シーケンサーにかけるまでの主な実験の流れとして、エクソン 18 の 189 塩基、エクソン 19 の欠失部分を含む 265 塩基、エクソン 20 の 235 塩基とエクソン 21 の突然変異を含む 210 塩基の領域を PCR で増幅させた。異種性癌サンプルに含まれる体細胞変異は、ばらつきが多く出現頻度も低いため、従来のサンガー法では検出できない可能性がある。このような変異をより確実に検出するためには、3,000X 程度のより深いカバレッジの読み取りが必要になってくる。申請者らが所有する Illumina 社 MiSeq は、高速・高精度のベンチトップ型のシーケンサーである。99%の野生型に混在した 1%の変異であっても検出可能になることが期待できる。使用する Illumina 社のキットをプロトコル通りに実験すると莫大な費用を要することになるため、まずは 1 キットあたり数百サンプルの実験ができるような条件検討を実施した。サンプル調製では、まずターゲット領域の両側にシーケンスを行うためのアダプターとサンプルを識別するためのバーコード配列を結合させた。分子バーコードは、各読み取り領域の一部を形成する固有の配列の集合体であり、特定のサンプルのすべての読み取りデータを同定することができる。上記領域を増幅するプライマーの末尾にアダプターが結合可能な共通配列をオーバーハングで持たせ、それぞれを Tailed PCR することでオーバーハング領域を付加した。PCR はマルチプレックス PCR 法を用いて 1 反応液内で各エクソンを同時に増幅させることに成功した。また、反応液を標準プロトコルの約 1/3 の系に縮小したことによって試薬コストの削減を行うことができた。384 サンプル分のバーコード配列は illumina 社より購入し、アダプターとバーコード配列の付加は illumina 社のキットを使用し PCR によって付加した。残りの 384 サンプル分は申請者らが開発したバーコードを付加することに成功した。バーコード付加に関しても反

応液を 1/4 に縮小してコスト削減を行った。1 回のシーケンス反応で 768 サンプル（768 種類のバーコード）同時にシーケンス可能となった。

これらの 768 サンプルをプールする際、サンプル間での量を均一化するためビーズを使用し、DNA 量の normalization を行った。リアルタイム PCR で定量後、ペアエンドシーケンスを実施した。次に、得られたデータに品質管理を施した。GATK の Haplotype Caller と Unified Genotyper で参照配列と比較し mutation call を実行した。サンプルの平均リード数は 3,200 ほど、そのうち平均 3,000 ほどが正しいゲノム上の位置にマップされた。

768 検体のうち、EGFR の 18~21 エクソンに変異が見つかったサンプルは 214 検体、そのうち 25 検体には 2 つ以上のエクソンで変異が同定された。24 件体でエクソン 18 に、95 件体でエクソン 19 に、32 件体でエクソン 20 に、88 件体でエクソン 21 に変異が見つかった。今回未知の変異は同定できなかった。GWAS を実施している肺がん検体のうち、551 件体が EGFR 変異あり、1008 検体が変異なし、ということが判明した。GWAS 実施済みだが EGFR 変異は不明なサンプル 856 件体については血液由来の DNA であり、同一患者のがん部から抽出した DNA が手に入らなかった。

## (2) Subtype 解析

過去に申請者らは、肺がんの発症リスクに関連する遺伝的因子を同定するため、イルミナ社の Human610k アレイを用いて肺がん患者検体 2,415 例、対照群検体 4,794 例の全ゲノム関連解析 (GWAS) を実施した。結果の品質管理は、検体については、タイピング成功率が 90%以上、近縁度が高いペアの片方を除外、また SNP についてはタイピング成功率が 95%以上のもののみを解析に使用した。患者と対照群の間で、アリル頻度、ジェノタイプ分布を統計解析した。その結果、5p15 にゲノムワイドで有意となる閾値 ( $5.0 \times 10^{-7}$ ) を下回る領域の他、13q21、1p34 等の複数の候補領域が同定されている。

GWAS 結果を用いて、8 つの subtype 解析を実施した。

Study (1) 肺腺がん・喫煙歴あり vs 肺腺がん・禁煙

Study (2) 肺腺がん vs コントロール群

Study (3) 肺腺がん・EGFR 変異陽性 vs コントロール群

Study (4) 肺腺がん・EGFR 変異陰性 vs コントロール群

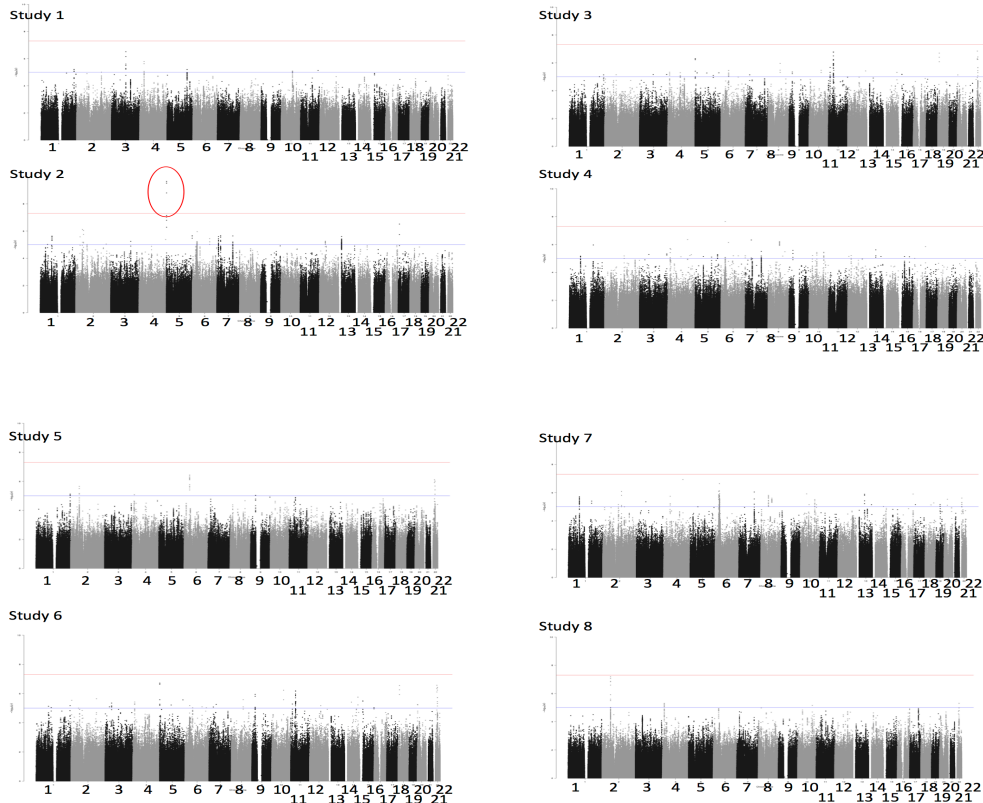
Study (5) 肺腺がん・EGFR 変異陽性 vs 肺腺がん・EGFR 変異陰性：同じ腺癌という表現型の中、どの時点で分岐して EGFR 変異の無い癌と変異のある癌に分かれるのかの解明につながる。

Study (6) EGFR 変異陽性 vs コントロール群：EGFR 変異有りの肺がんリスクと関連する SNP を同定することにより、EGFR 遺伝子変異・欠失誘発の機構解明に寄与する。

Study (7) EGFR 変異陰性 vs コントロール群

Study (8) EGFR 変異陽性 vs EGFR 変異陰性

ゲノムワイド関連解析で得られた各 SNP の P 値 ( $-\log_{10}$ ) を縦軸に、染色体上の位置を横軸にとった以下のような図をマンハッタンプロットと呼ぶ。赤色の線は GWAS での有意レベルにあることを示している。



関連のある SNP が存在する場合、近傍の連鎖不平衡の関係にある SNP も関連を示すため、プロット上、突出したピークが形成される。Study2 にのみ GWAS 有意を示す SNP があるが、これはすでに肺腺がんとの関連が報告されている TERT 遺伝子領域である ( $p=2.20 \times 10^{-10}$ )。

それぞれの Study でゲノムワイドには達していないが患者とコントロール群の間で有意傾向を示す SNP が多数あった。これらは、疾患感受性遺伝子の候補となるものであるが、独立した集団の検体を用いた再現性検証が必須である。そのため今後は検体数の増加につとめる。新たに収集された検体を用いて候補多型のタイピングを行い、結果の比較を通して、真の感受性遺伝子の同定を目指す。また、GWAS 済みの検体でまだ EGFR のシークエンスができていない検体も残っている。我々が確立したプロトコールでエクソン 18~21 のシークエンスを実施する予定だ。

本研究開始時は日本人における EGFR 遺伝子変異を考慮した関連解析はまだ報告されていなかったのだが、平成 28 年 8 月に国立研究開発法人国立がん研究センター、愛知県がんセンターなどからなる共同研究グループにより、EGFR 遺伝子変異陽性肺腺がんについて罹りやすさを決める遺伝子領域が報告された。肺腺がん患者 3 千人と、がんにかかっていない 1 万 5 千人について GWAS を実施した結果、6 つの遺伝子領域の個人差が EGFR 変異陽性の肺腺がんへの罹りやすさを決めていることが明らかになった。その中には、免疫反応の個人差の原因となる HLA クラス II 遺伝子領域が含まれていた。まずはこの 6 つの遺伝子が今回どのような結果になっているのかを調べてみた。患者群は女性・非喫煙者・EGFR 遺伝子変異陽性肺腺がんとし、コントロール群は女性・非喫煙者とした。

Table 1. Association of GWAS identified SNPs and lung adenocarcinoma risk among never-smoking Asian women (Seow et al., 2017)

SNP	Nearest gene(s)	Chr	Major/minor allele	Human Molecular Genetics 2017					Our Results				
				MAF	No.	No.	OR	P value	MAF	No.	No.	OR	P value
rs4488809	<i>TP63</i>	3q28	T/C	0.47/0.53	7,448	7,007	0.80	$4.30 \times 10^{-17}$	0.46/0.51	516	578	0.83	0.0366
rs2736100	<i>TERT</i>	5p15.33	A/C	0.49/0.46	7,505	7,070	1.43	$6.12 \times 10^{-43}$	0.46/0.38	516	578	1.35	$5.11 \times 10^{-04}$
rs3817963	<i>BTNL2</i>	6p21.3	T/C	0.33/0.29	7,255	6,745	1.16	$1.63 \times 10^{-7}$	0.39/0.33	516	578	1.28	$6.35 \times 10^{-03}$
rs2179920	<i>HLA-DPBI</i>	6p21.32	C/T	0.14/0.13	7,457	7,020	1.17	$1.69 \times 10^{-5}$	0.20/0.17	516	578	1.18	0.148
rs2395185	HLA class II	6p21.32	G/T	0.41/0.38	7,757	9,637	1.16	$2.04 \times 10^{-9}$	0.42/0.38	516	578	1.19	0.0452
rs9387478	<i>ROSI/DCBLD1</i>	6q22.2	C/A	0.47/0.50	8,022	9,970	0.86	$5.25 \times 10^{-11}$	0.45/0.49	516	578	0.86	0.0672
rs7086803	<i>VTG1A</i>	10q25.2	G/A	0.30/0.26	7,964	9,914	1.25	$9.22 \times 10^{-17}$	0.27/0.21	516	578	1.37	$2.28 \times 10^{-03}$
rs11610143		12q13.13	C/G	0.30/0.33	10,267	10,634	0.85	$3.55 \times 10^{-13}$	0.32/0.35	516	578	0.90	0.252
rs7216064	<i>BPTF</i>	17q24.3	A/G	0.40/0.42	7,720	8,630	0.86	$6.19 \times 10^{-9}$	0.25/0.31	516	578	0.74	$1.48 \times 10^{-03}$

我々の検体数が Seow et al の論文に比べると圧倒的に少ないので p 値に差があるものの、同じ傾向は見られた。残念ながらゲノムワイドに有意な遺伝子座位はなかった。

本研究の特色として一度に数百検体の EGFR の変異の有無を高速に解析するプロトコルを確立したことが挙げられる。EGFR 遺伝子変異の原因は明らかではないが、日本人の肺がんのおよそ 30~40%は、この EGFR 遺伝子変異が原因で起こることが知られている。日本でその感受性遺伝子を特定し発症危険群を特定することは発症予防を目的とした治療を可能とする研究になることが期待され、医学の観点からも大きな貢献となるであろうと考えている。EGFR のシーケンスを進めるとともに、今後は候補 SNP の再現性を確認するために必要な検体を最大限に増やす。

GWAS を用いた subtype 解析の結果、候補として挙がってきた SNP のほとんどは遺伝子外やイントロンに存在していたため、疾患感受性への直接的な関与については不明確であった。これらのゲノム領域におけるイルミナチップでは解析できない稀少 SNP、及び遺伝子に直接的な影響を与える SNP や変異を迅速に同定するため、次世代遺伝子解析装置を用いたエクソーム解析を行うことを検討中である。また全エクソン塩基配列の決定を介して遺伝子に直接的な影響を与える SNP や変異に関しても評価し、発症リスク群の特定を可能とすることを目的とする。

EGFR 遺伝子変異型肺がんリスク感受性遺伝子の全貌を明らかにすることで、人種間や病態・環境因子・生活習慣における相対的貢献度を解明することが可能になる。また、新たな分子診断や予後の予測、新薬の開発等につながれば社会的に大きな貢献も期待できる。遺伝的多型性とがんリスクの検討は、発がん機構の解明に寄与すると共に、ハイリスク者を特定しがん予防を実践する上にも非常に有用である。

## 5. 主な発表論文等

〔雑誌論文〕 (計 0 件) 〔学会発表〕 (計 0 件) 〔図書〕 (計 0 件)  
〔産業財産権〕 出願状況 (計 0 件) 取得状況 (計 0 件) 〔その他〕 該当なし

## 6. 研究組織

(1) 研究分担者：該当なし (2) 研究協力者：該当なし

※科研費による研究は、研究者の自覚と責任において実施するものです。そのため、研究の実施や研究成果の公表等については、国の要請等に基づくものではなく、その研究成果に関する見解や責任は、研究者個人に帰属されます。