

INVITED REVIEW ARTICLE

肺がんの発症に関わる遺伝子多型

白石航也¹

Genetic Polymorphisms and Susceptibility to Lung Cancer

Kouya Shiraishi¹¹Division of Genome Biology, National Cancer Center Research Institute, Japan.

ABSTRACT — Lung cancer is the leading cause of death in both men and women worldwide, and the 5-year survival rate for advanced lung cancer is as low as 20%. The development of molecular-targeted cancer therapy has significantly improved the outcomes of non-small cell lung cancer (NSCLC) patients with epidermal growth factor receptor (*EGFR*) mutation and anaplastic lymphoma kinase (*ALK*) fusion oncogenes. In addition, immune checkpoint inhibitors have also dramatically changed the treatment regimens of NSCLC. From the viewpoint of cancer prevention, the prevalence of lung cancer in men is expected to decrease in the future due to smoking cessation efforts. As the smoking prevalence in women is already lower than that in men, the incidence of lung cancer in women is not expected to decrease quite as much as that in men. However, the therapeutic effect of immune checkpoint inhibitors in smokers with NSCLC can be expected, while such an effect in never-smokers with NSCLC is not expected. Because smoking is linked to the expression of neo-antigens and increased numbers of somatic mutations, immune checkpoint inhibitors in NSCLC have suggested the smoking history to be associated with improved survival outcomes. Therefore, the importance of determining the susceptibility to lung cancer in never-smokers, especially among women, is expected to increase in the future. In this review, the author describes the risk factors of susceptibility to lung cancer that have been identified thus far and the prospects for the future of cancer prevention.

(JLCC. 2018;58:331-337)

KEY WORDS — Genetic polymorphism, Lung cancer susceptibility gene, Genetic diversity of the HLA system

Corresponding author: Kouya Shiraishi.

要旨 — 肺がんは男女ともにがんの部位別死亡数の上位を占める悪性腫瘍であり、進行肺がんの予後は20%と極めて低い。しかし *EGFR* 体細胞変異をはじめとするドライバー遺伝子変異に対する分子標的薬ならびに免疫チェックポイント阻害剤の登場により、非小細胞肺がんに対する治療法は大きな転換期を迎え、極めて予後不良だったIV期の進行肺がんの治療も現実味を帯びてきている。一方肺がんの罹患率は、禁煙の取り組みにより男性肺がんにおいて今後減少が期待できる。しかしもともと喫煙率が低い女性肺がんの発症数は、男性に比べると期待されるほど減少しないと考えられる。また喫煙者に対する免疫チェックポイント阻害剤の治療効果は十分期

待されるが、非喫煙者特に女性非喫煙者肺がんでは *EGFR* 変異といったドライバー変異を伴う場合が多い。そのため、免疫の惹起に関わる喫煙に依存するネオアンチゲンが認められないため、免疫チェックポイント阻害剤の恩恵が得られない可能性が指摘されている。したがって非喫煙者特に女性に対する肺がん発症要因の同定は、今後さらに重要性を増すと考えられる。本稿では、これまでに同定されてきた肺がんの発症リスク要因と今後の研究の展望について述べる。

索引用語 — 遺伝子多型, 肺がん感受性遺伝子, HLA 遺伝子の多様性

¹国立がん研究センター研究所ゲノム生物学研究分野。

論文責任者: 白石航也。

今までに報告されている肺がんの発症に関わる外的要因

肺がんの主要組織型である腺がんは、肺がんの危険因子である能動喫煙との関連が他の組織型（扁平上皮がんや小細胞肺がん、大細胞肺がん）に比べて比較的弱く（相対危険度は約2倍）、また約半数は非喫煙者で発症する。¹ また近年日本人の非喫煙者を対象とした受動喫煙に対するメタアナリシス研究による肺がんリスク評価が行われ、受動喫煙のある人はない人に比べて肺がんになるリスクが約1.3倍となり、国際的なメタアナリシスと同様の結果が報告されている。² また国立がん研究センターをはじめとする研究グループによる科学的根拠に基づいた『日本人のためのがん予防法』が報告されている。そこには、喫煙以外の肺がんの発症リスクとして、「肺結核が可能性あり」、「職業性アスベストがほぼ確実」とされているが、喫煙以外の主要な肺がんリスク因子は明らかではない。³ 一方で、今までの報告で米国における若年の肺がん罹患率は男性より女性で高いことが示唆されていたが、この傾向は最近の出生コホート研究でも再現されている。⁴ しかしその発症要因は、喫煙行動だけでは十分説明ができていないことから、環境要因によらない要因が若年発症に寄与していると考えられる。またアジア人では欧米人と比べて、非喫煙者、特に女性において肺がんの発症割合が高く、⁵ また腺がん組織中で認められる体細胞変異頻度の分布は、欧米人とアジア人で大きく異なる。⁶ たとえば、EGFR 体細胞変異の頻度は欧米人に比べてアジア人で高く、一方でKRAS 体細胞変異頻度はアジア人に比べて欧米人で高いことが知られている。したがって肺がんの発症要因には人種差があり、肺がん発症には、喫煙習慣や受動喫煙などによる環境要因や遺伝要因が複合的に関与していると考えられる。このような遺伝要因を同定するため、一塩基多型（single nucleotide polymorphism：SNP）をマーカーとしたゲノム網羅的関連解析（genome-wide association study：GWAS）が多数実施され、複数の肺がんリスクと関連するゲノム領域が同定されている。

遺伝要因からみる肺がん発症リスク因子

ゲノムDNAの多様性は、様々な疾患への罹患リスクマーカーとして用いられてきた。以前は、SNPチップデータより得られる5%以上のコモンなバリエーションもしくはSNPがリスクマーカーとして同定されてきたが、近年の次世代シーケンズの進歩により1%程度の頻度で認められるレアなバリエーションも同定されるようになった。また症例群と対照群の対象数、バリエーションの頻度とオッズ比を基に算出される検出力により、同定できる真

の感受性遺伝子座の閾値が設定される。そのため、多数の真の感受性遺伝子座を同定するためには莫大なサンプル数を用いた解析が必要となり、最近では数万例を対象とした解析が一般的になってきている。

欧米人のグループや我々を含めたアジア人のグループによって、数千～数万例規模の肺がん患者と非がん対照集団からなるGWASが行われ、複数のGWASによって再現性が確認されている。その代表的な肺がん感受性遺伝子座として、*CHRNA*,^{7,8} *TERT*,⁷⁻¹³ *CLPTM1L*,^{7,8} *TP63*,⁸⁻¹⁰ *BTNL2*,^{10,11} *FOXP4-AS1*,^{12,13} *ROS1*,^{8,11} *VTI1A*,^{11,12} *BPTF*,^{10,11} *CYP2A6*,^{8,14} *BRCA2*^{8,15} があり、Table 1にまとめた。

欧米人を対象としたGWASにより、ニコチン性アセチルコリン受容体サブユニットをコードする*CHRNA* 遺伝子群の多型は、組織型や喫煙の有無によらず肺がんリスクとの関連が報告されている。また*CHRNA* 遺伝子とニコチン代謝に関わる*CYP2A6* 遺伝子の多型は、喫煙習慣と関連するとともに、肺がんリスクと喫煙習慣との交互作用が報告されている。米国NCI主導で行われた、OncoArray Networkプロジェクトの一環で行われた数万人規模の肺がん患者を対象としたGWASにより、相同組換えによるDNA二本鎖切断の修復に重要な役割を果たしている*BRCA2* 遺伝子のレアバリエーションが、扁平上皮がんのリスク因子として同定された。しかしこれらのバリエーションは、アジア人では認められていない欧米人特異的なバリエーションである。また扁平上皮がんとの関連が認められた*BRCA2* 遺伝子のレアバリエーションはタンパク質短縮を生じさせる変異であり、このバリエーションは家族性を含む乳がんや卵巣がん、皮膚がん、消化管がんなど多くのがんに対して発がんリスクと関連することが報告されている。^{15,16}

テロメラーゼのサブユニットをコードする*TERT* 遺伝子の多型は、人種によらず腺がんリスクと関連する。*TERT* 遺伝子領域に近傍する*CLPTM1L* 遺伝子の多型は、欧米人の腺がんや扁平上皮がんの他に中国人の非小細胞肺がんでも発がんリスクとの関連が報告されており、今後*TERT* 遺伝子のバリエーションと独立して発がんリスクに寄与するか検討する必要がある。¹⁷ 興味深いことに、*TERT* 遺伝子の多型は、欧米人に比べてアジア人において強く腺がんリスクと関連し（Table 1）、アジア人肺腺がんの半数を占めるEGFR変異陽性腺がんの方がEGFR変異陰性腺がんと比べてより強くリスクに寄与することが報告されている。¹³ 一方で、マウス実験によるKRAS変異を導入して発生した肺がんに対して、RNAiを用いて*CLPTM1L* 遺伝子の発現量を抑制すると発がん能が低下したという報告がなされている。¹⁸ したがって、ドライバー変異型によって*TERT* や*CLPTM1L* 遺伝子の寄与

Table 1. Summary of Previously Reported Cases of GWAS with Lung Cancer Risk

Chromosome position	Gene symbol	Gene name	Strongest SNP-risk allele	Location	Frequency of risk allele*		Odds ratio		Association between histological type and lung cancer risk
					EAS	EUR	Asian	Caucasian	
3q28	<i>TP63</i>	Tumor protein p63	rs10937405-C	Intron	0.689	0.572	1.23-1.31	1.12	ADC, EGFR-mutated ADC
5p15.33	<i>TERT</i>	Telomerase reverse transcriptase	rs2736100-G	Intron	0.415	0.499	1.27-1.46	1.12-1.25	ADC
5p15.33	<i>CLPTMIL</i>	CLPTM1-like	rs31489-C	Intron	0.810	0.591	-	1.12-1.20	ADC, SQC
6p21.32	<i>BTNL2</i>	Butyrophilin-like 2	rs3817963-G	Intron	0.231	0.276	1.18-1.29	-	ADC
	<i>HLA-DPB1</i>	Major histocompatibility complex, class II, DP beta 1	rs2179920-A	Inter-genic	0.141	0.233	1.36	-	EGFR-mutated ADC
6p21.1	<i>FOXP4-AS1</i>	FOXP4 antisense RNA 1	rs2495239-A	Intron	0.429	0.086	1.18-1.19	-	ADC, EGFR-mutated ADC
6q22.1	<i>ROS1</i>	ROS proto-oncogene 1, receptor tyrosine kinase	rs9387478-C	Inter-genic	0.502	0.495	1.18	1.09	ADC
10q25.2	<i>VTI1A</i>	Vesicle transport through interaction with t-SNAREs 1A	rs7086803-A	Intron	0.294	0.028	1.28	-	ADC
13q13.1	<i>BRCA2</i>	BRCA2, DNA repair-associated	rs11571833-T	Stop gained	0.000	0.011	-	1.60-2.47	SQC, SCC
15q25.1	<i>CHRNA3</i>	Cholinergic receptor nicotinic alpha 3	rs1051730-T	Synonymous	0.027	0.369	-	1.31-1.35	ADC, SQC, SCC
17q24.2	<i>BPTF</i>	Bromodomain PHD finger transcription factor	rs7216064-A	Intron	0.692	0.214	1.16-1.20	-	ADC
19q13.2	<i>CYP2A6</i>	Cytochrome P450 family 2 subfamily A member 6	rs56113850-C	Intron	0.340	0.592	-	1.12-1.16	ADC, SQC

Asians and Caucasians have different genetic susceptibilities to lung cancer. Variants in *CHRNA3*, *CYP2A6* and *BRCA2* are strongly associated with lung cancer risk in Caucasians. Variants in *TERT*, *TP63*, *ROS1* and HLA region are associated with lung cancer risk both in Asians and Caucasians. On the other hand, variants in *BTNL2*, *FOXP4-AS1*, *VTI1A* and *BPTF* are strongly associated with lung cancer risk in Asians.

*EAS, East Asia in 1000 genomes super population; EUR, Europe in 1000 genomes super population.

ADC, adenocarcinoma; SQC, squamous cell carcinoma; SCC, small cell carcinoma.

が異なる可能性も考えられる。

当初アジア人集団を対象としたGWASでのみ、p53がん抑制遺伝子のファミリーに属する*TP63*とインスリン受容体ファミリーの受容体チロシンキナーゼである*ROS1*遺伝子が肺がん感受性遺伝子座として同定されていたが、後に欧米人を対象としたGWASにより*TP63*や*ROS1*も肺がん感受性遺伝子として同定された。しかし、肺がんリスクへの寄与度は欧米人と比べてアジア人で高い傾向を示しており、人種によって感受性遺伝子の寄与度が異なると考えられる。またアジア人を対象に同定された感受性遺伝子座としては、免疫応答に関わる*BTNL2*と*FOXP4-AS1*、ヌクレオソームリモデリング構成因子である*BPTF*、小胞輸送において膜融合に関わる*VTI1A*が報告されている。

我々は*EGFR*変異を伴う腺がんを対象とした全ゲノム関連解析を実施し、HLAクラスII領域を含む6つの感受性遺伝子座を同定している (Figure 1)。さらに本稿

でも紹介する HLA imputation 法を用いることで、HLA-DPB1タンパク質の57番目のアミノ酸の置換を起こす多型が、*EGFR*変異陽性の腺がん発症リスクに関わることを見出している。今後は、HLAアレルやアミノ酸の違いにより免疫活性化能の違いがあるかを検討する必要がある。今回報告した肺がん感受性遺伝子の機能的意義は未だ不明な部分が多く、今後培養実験やがんゲノム解析を通して、さらなる機能的な役割を明らかにする必要がある。

また今までに報告したバリエーションは、個々のオッズ比が1.1~1.5と低いものの、効果サイズの推定値で重み付けしたリスクアレル数の和を算出するポリジェニック・スコアを算出することで、遺伝的な発症リスク予測モデルを構築することができると考えられる。これらの解析手法を用いることで、高リスク群を捕捉するための手段の1つになると考えられる。

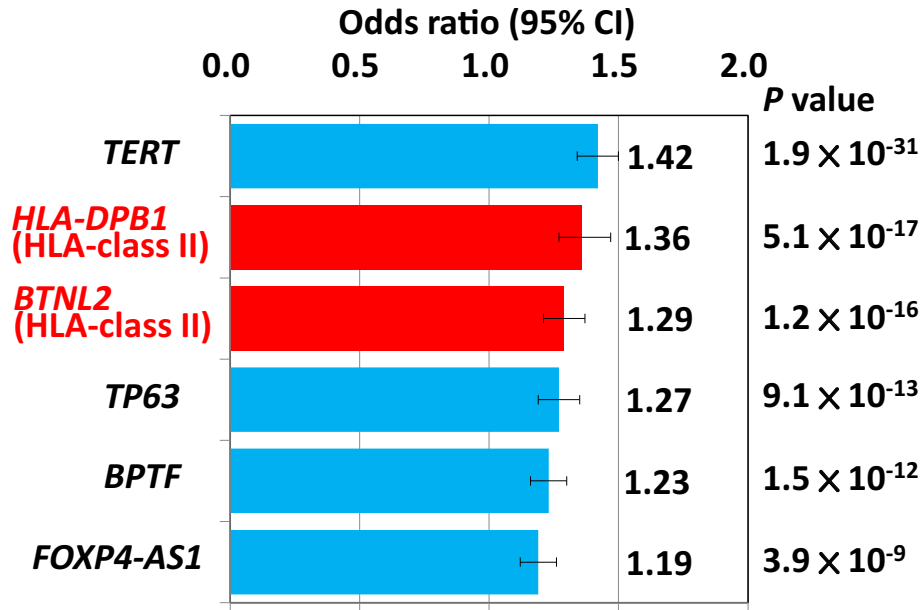


Figure 1. Association of six SNPs with the risk of lung adenocarcinoma with an *EGFR* mutation. Four loci—5p15.33 (*TERT*: rs2736100), and 6p21.3 (*BTNL2*: rs3817963), 3q28 (*TP63*: rs7636839), and 17q24.2 (*BPTF*: rs7216064)—previously shown to be strongly associated with the overall lung adenocarcinoma risk in East Asians were re-discovered as loci more strongly associated with the risk of lung adenocarcinomas with *EGFR* mutations than with those without such mutations. In addition, the loci HLA-class II at 6p21.32 (*BTNL2*: rs2179920) and 6p21.1 (*FOXP4-AS1*: rs2495239) were newly identified as being strongly associated with the overall lung adenocarcinoma risk in East Asians.

HLA (human leukocyte antigen) 遺伝子領域の多様性と肺がんリスクとの関連

先に述べた欧米人を対象とした GWAS でもう 1 つ着目すべき点として、扁平上皮がんにおいて HLA 領域に多数の感受性遺伝子が同定されたことである (Table 2)。しかし HLA 領域は非常に強い連鎖不平衡を形成することから、Table 2 に示されているバリエーションが全て独立したリスク因子となり得るのかについては、さらなる検証が必要である。HLA は主に HLA クラス I (*HLA-A*, *HLA-B*, *HLA-C* など) と HLA クラス II (*HLA-DR*, *HLA-DQ*, *HLA-DP* など) に分けられる。MHC (major histocompatibility complex) クラス I は、たとえば腫瘍細胞などから産生される外来性抗原から細胞障害性 T 細胞への抗原提示に用いられ、MHC クラス II は抗原提示細胞からヘルパー T 細胞への抗原提示に用いられる。近年のがんゲノム解析により、成人 T 細胞白血病・リンパ腫¹⁹ や非小細胞肺癌²⁰ において、HLA 遺伝子の欠失変異や欠損などが他の遺伝子よりも多く蓄積されており、がん免疫からの回避に重要な役割を果たしていると考えられている。したがって肺がんの発症要因にも、HLA の多様性が大きな影響を与えている可能性が考えられる。

HLA 遺伝子型と免疫状態の推定

HLA 遺伝子の多型は、その高度な多様性のため人種間で大きく異なるため、HLA アリルの同定は非常に煩雑であった。しかし近年のゲノム解析の進歩により、がん免疫機構を解明するための 1 つの手法として、HLA アリルの推定やネオアンチゲン解析などが行われている。そのゲノム解析手法について簡単に概説する。

SNP チップデータをベースにした HLA 遺伝子型推定アルゴリズム (HLA imputation) には、主に SNP2HLA²¹ と HIBAG²² が知られており、SNP チップデータがあれば、HLA アリルを推定することができる。両ソフトウェアの HLA アリルの推定精度はともに 90% 以上と高く、また日本人特異的なアリルやハプロタイプを含む日本人参照データが既に構築されている。また次世代シーケンスを用いた HLA 遺伝子型推定アルゴリズムは既に多数報告されており、主なアルゴリズムとして HLAmirer,²³ OptiType²⁴ などが知られている。報告されているアルゴリズムの精度評価により、HLA クラス I に限ると OptiType が、最も精度よく HLA 遺伝子型を推定できるとされている。

腫瘍特異的に発現した体細胞変異を伴うペプチド (ネ

Table 2. Genetic Variants of the HLA Region Identified in GWAS in European and American Populations

Chromosome position	Gene symbol	Gene name	Strongest SNP-risk allele	Location	Frequency of risk allele*		Odds ratio	Association between histological type and lung cancer risk
					EAS	EUR		
6p22.2	<i>BTN3A2</i>	Butyrophilin subfamily 3 member A2	rs34107459-C	Regulatory region	0.000	0.067	1.26	SQC
6p22.1	<i>PRSS16</i>	Serine protease 16	rs35768595-T	Intergenic	0.000	0.069	1.24	SQC
6p22.1	<i>HLA-A</i>	Major histocompatibility complex, class I, A	rs147097402-A	Intron	ND	ND	1.25	SQC
6p21.33	<i>HLA-C</i>	Major histocompatibility complex, class I, C	rs115390513-C	Intergenic	0.000	0.087	1.26	SQC
6p22.1	<i>HLA-E</i>	Major histocompatibility complex, class I, E	rs116310021-G	Intergenic	0.000	0.080	1.27	SQC
6p22.1	<i>HLA-G</i>	Major histocompatibility complex, class I, G	rs116506680-T	Intergenic	0.340	0.439	1.10	ADC
6p21.32	<i>HLA-DQA1</i>	Major histocompatibility complex, class II, DQ alpha 1	rs74942078-A	Intron	0.052	0.106	1.27	SQC
6p21.32	<i>HLA-DQB2</i>	Major histocompatibility complex, class II, DQ beta 2	rs148861668-T	Intergenic	0.000	0.084	1.22	SQC

McKay et al reported that variants in HLA region were associated with risk to lung squamous cell carcinoma in Caucasians. On the other hand, Shiraishi et al reported that variants in HLA region, including *BTNL2* and *HLA-DPB1*, were associated with lung adenocarcinoma risk in Asians.

*EAS, East Asia in 1000 genomes super population; EUR, Europe in 1000 genomes super population.

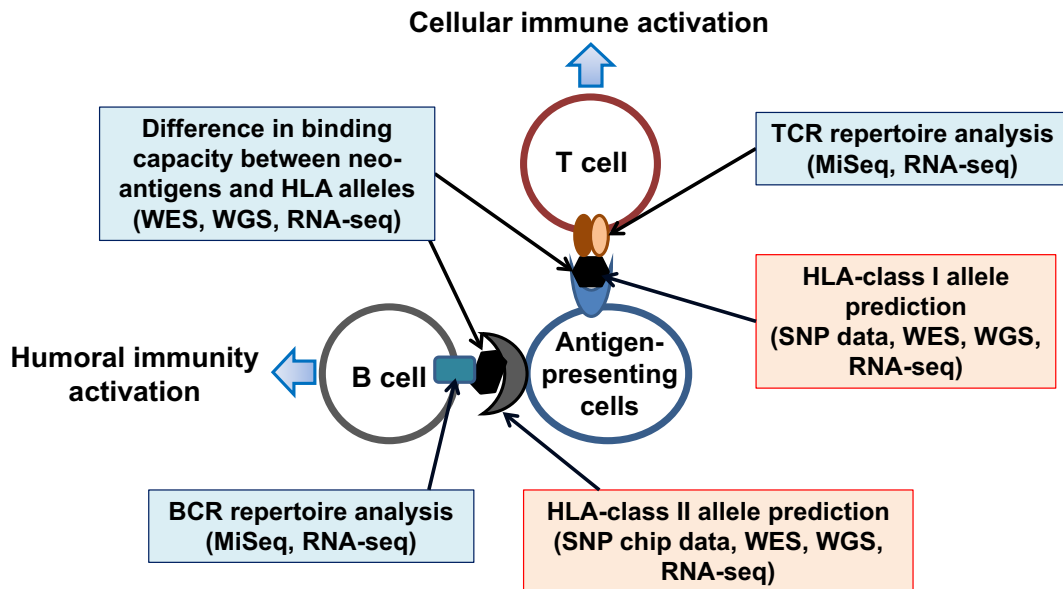


Figure 2. Overview of the neo-antigen landscape and identification strategies using next-generation sequencing. To predict each individual's HLA alleles, HLA imputation is performed using SNP chip data, whole-exome sequencing (WES), whole-genome sequencing (WGS) or RNA sequencing (RNA-seq). T cell receptor (TCR) and B cell receptor (BCR) repertoire analyses are useful for evaluating the diversity of a patient's immune system. To predict the binding capacity between neo-antigens and HLA alleles, some analyses are performed using WES, WGS and RNA-seq.

オアンチゲン)がMHCに提示されると、腫瘍特異的な免疫を誘導される。よってネオアンチゲンの多様性を評価することは、個々人の免疫状態を明らかにするために必

要な解析である。ネオアンチゲンを同定するために、MHCクラスIと変異ペプチドとの結合能を推定するアルゴリズムが多数報告されており (Figure 2)、主に

NetMHCpan,²⁵ NetCTLpan²⁶などが知られている。具体的には、全エクソンシーケンスなどで認められたアミノ酸変化を伴うバリエーション情報を基にエピトープを推定し、さらに immune epitope database (IEDB) によるニューラルネットワークなどのデータ解析技術を組み合わせることで、MHC クラス I とネオアンチゲンとの結合親和性を推定する。一方 Neopepsee²⁷ は、腫瘍と非がん組織検体由来 DNA を用いた全エクソンシーケンスデータなどを基に変異ペプチドを推定するとともに、RNA-seq により実際にがん組織中で発現しているネオアンチゲンに絞って解析する方法である。また MHC クラス II へのペプチドの結合能の推定には、NetMHCII が最も精度が高いと言われている。しかし、MHC クラス II に結合する変異ペプチドの長さは、MHC クラス I に比べて厳格に規定されていないためアルゴリズムの推定が難しく、MHC クラス I と II ともに培養実験などを用いた生物学的な検証が必須である。

免疫システムはネオアンチゲンの多様性に対応するため、 $\alpha\beta$ T 細胞受容体 (T cell receptor : TCR)・B 細胞受容体 (B cell receptor : BCR) 遺伝子や免疫グロブリン遺伝子を再構成することで、抗原多様性に対処している。MHC が提示した抗原と結合する相補性決定領域 (complementarity determining region : CDR) では無作為に塩基の欠失や挿入が発生することで、抗原認識に対して多様性を持たせている。TCR/BCR 遺伝子配列を網羅的に調べることで、疾患に関連する T 細胞や B 細胞の同定につながることも、TCR を通してどのような抗原を認識し、腫瘍細胞に対して免疫を誘導しているか免疫系の多様性を評価するためにも有用である。TCR レパトア解析は、CDR をロングリードでシーケンスする必要がある。それに対応可能な次世代シーケンサー (Illumina 製 MiSeq など) を用いて行われる。一方で、RNA シーケンスといった次世代シーケンスなどを用いた推定アルゴリズムが開発されており、MiXCR,²⁸ MIGEC²⁹ やその後の解析ツールとしての VDJtools³⁰ が知られている。しかしまだこれらのアルゴリズムに対して、異なるサンプルセットを用いた精度評価がなされていないため、実際に免疫応答が惹起されるかについてはさらなる生物学的な検証実験が必要である。

今後の展望

近年のゲノム解析により、多数の肺がん感受性遺伝子座が同定できるようになった。しかし当初期待されていたような高精度の肺がん高危険度群の捕捉には未だ至っていない。その原因としては、同定されたりスクアリアルがオッズ比 1.1~1.5 程度と低いことと、喫煙をはじめとする環境要因が複合的に関与しているためと考えられ

る。今後はより疫学的な側面を含めた、たとえば喫煙依存的な発がんリスク要因の同定、非喫煙者における受動喫煙による発がんリスクに寄与する遺伝的要因の同定が重要になると考えられる。またがんゲノム解析により、HLA の失活が発がん重要な役割を果たしているという報告がなされており、今後 HLA に着目した発がん機構を解明する必要がある。したがって、胚細胞系列変異の他に体細胞変異や生活習慣情報を組み合わせて解析することが重要であり、既に取得されている大規模ゲノム情報の統合解析が、プレジジョン・メディシンを推進する上で有益なツールとなり得る。

本論文内容に関連する著者の利益相反：なし

REFERENCES

1. Sobue T, Suzuki T, Fujimoto I, Matsuda M, Doi O, Mori T, et al. Case-control study for lung cancer and cigarette smoking in Osaka, Japan: comparison with the results from Western Europe. *Jpn J Cancer Res.* 1994;85:464-473.
2. Hori M, Tanaka H, Wakai K, Sasazuki S, Katanoda K. Secondhand smoke exposure and risk of lung cancer in Japan: a systematic review and meta-analysis of epidemiologic studies. *Jpn J Clin Oncol.* 2016;46:942-951.
3. Sasazuki S, Inoue M, Shimazu T, Wakai K, Naito M, Nagata C, et al. Evidence-based cancer prevention recommendations for Japanese. *Jpn J Clin Oncol.* 2018;48:576-586.
4. Jemal A, Miller KD, Ma J, Siegel RL, Fedewa SA, Islami F, et al. Higher Lung Cancer Incidence in Young Women Than Young Men in the United States. *N Engl J Med.* 2018;378:1999-2009.
5. Thun MJ, Hannan LM, Adams-Campbell LL, Boffetta P, Buring JE, Feskanich D, et al. Lung cancer occurrence in never-smokers: an analysis of 13 cohorts and 22 cancer registry studies. *PLoS Med.* 2008;5:e185.
6. Kohno T, Nakaoku T, Tsuta K, Tsuchihara K, Matsumoto S, Yoh K, et al. Beyond ALK-RET, ROS1 and other oncogene fusions in lung cancer. *Transl Lung Cancer Res.* 2015;4:156-164.
7. Landi MT, Chatterjee N, Yu K, Goldin LR, Goldstein AM, Rotunno M, et al. A Genome-wide Association Study of Lung Cancer Identifies a Region of Chromosome 5p15 Associated with Risk for Adenocarcinoma. *Am J Hum Genet.* 2011;88:861.
8. McKay JD, Hung RJ, Han Y, Zong X, Carreras-Torres R, Christiani DC, et al. Large-scale association analysis identifies new lung cancer susceptibility loci and heterogeneity in genetic susceptibility across histological subtypes. *Nat Genet.* 2017;49:1126-1132.
9. Miki D, Kubo M, Takahashi A, Yoon KA, Kim J, Lee GK, et al. Variation in TP63 is associated with lung adenocarcinoma susceptibility in Japanese and Korean populations. *Nat Genet.* 2010;42:893-896.
10. Shiraishi K, Kunitoh H, Daigo Y, Takahashi A, Goto K, Sakamoto H, et al. A genome-wide association study

- identifies two new susceptibility loci for lung adenocarcinoma in the Japanese population. *Nat Genet.* 2012;44:900-903.
11. Lan Q, Hsiung CA, Matsuo K, Hong YC, Seow A, Wang Z, et al. Genome-wide association analysis identifies new lung cancer susceptibility loci in never-smoking women in Asia. *Nat Genet.* 2012;44:1330-1335.
 12. Wang Z, Seow WJ, Shiraishi K, Hsiung CA, Matsuo K, Liu J, et al. Meta-analysis of genome-wide association studies identifies multiple lung cancer susceptibility loci in never-smoking Asian women. *Hum Mol Genet.* 2016;25:620-629.
 13. Shiraishi K, Okada Y, Takahashi A, Kamatani Y, Momozawa Y, Ashikawa K, et al. Association of variations in HLA class II and other loci with susceptibility to EGFR-mutated lung adenocarcinoma. *Nat Commun.* 2016;7:12451.
 14. Thorgeirsson TE, Gudbjartsson DF, Surakka I, Vink JM, Amin N, Geller F, et al. Sequence variants at CHRN3-CHRNA6 and CYP2A6 affect smoking behavior. *Nat Genet.* 2010;42:448-453.
 15. Wang Y, McKay JD, Rafnar T, Wang Z, Timofeeva MN, Broderick P, et al. Rare variants of large effect in BRCA2 and CHEK2 affect risk of lung cancer. *Nat Genet.* 2014;46:736-741.
 16. Meeks HD, Song H, Michailidou K, Bolla MK, Dennis J, Wang Q, et al. BRCA2 Polymorphic Stop Codon K3326X and the Risk of Breast, Prostate, and Ovarian Cancers. *J Natl Cancer Inst.* 2015;108. pii: djv315.
 17. Hu Z, Wu C, Shi Y, Guo H, Zhao X, Yin Z, et al. A genome-wide association study identifies two new lung cancer susceptibility loci at 13q12.12 and 22q12.2 in Han Chinese. *Nat Genet.* 2011;43:792-796.
 18. James MA, Vikis HG, Tate E, Rymaszewski AL, You M. CRR9/CLPTMIL regulates cell survival signaling and is required for Ras transformation and lung tumorigenesis. *Cancer Res.* 2014;74:1116-1127.
 19. Kataoka K, Nagata Y, Kitanaka A, Shiraishi Y, Shimamura T, Yasunaga J, et al. Integrated molecular analysis of adult T cell leukemia/lymphoma. *Nat Genet.* 2015;47:1304-1315.
 20. McGranahan N, Rosenthal R, Hiley CT, Rowan AJ, Watkins TBK, Wilson GA, et al. Allele-Specific HLA Loss and Immune Escape in Lung Cancer Evolution. *Cell.* 2017;171:1259-1271.e11.
 21. Okada Y, Momozawa Y, Ashikawa K, Kanai M, Matsuda K, Kamatani Y, et al. Construction of a population-specific HLA imputation reference panel and its application to Graves' disease risk in Japanese. *Nat Genet.* 2015;47:798-802.
 22. Khor SS, Yang W, Kawashima M, Kamitsuji S, Zheng X, Nishida N, et al. High-accuracy imputation for HLA class I and II genes based on high-resolution SNP data of population-specific references. *Pharmacogenomics J.* 2015;15:530-537.
 23. Warren RL, Choe G, Freeman DJ, Castellarin M, Munro S, Moore R, et al. Derivation of HLA types from shotgun sequence datasets. *Genome Med.* 2012;4:95.
 24. Szolek A, Schubert B, Mohr C, Sturm M, Feldhahn M, Kohlbacher O. OptiType: precision HLA typing from next-generation sequencing data. *Bioinformatics.* 2014;30:3310-3316.
 25. Hoof I, Peters B, Sidney J, Pedersen LE, Sette A, Lund O, et al. NetMHCpan, a method for MHC class I binding prediction beyond humans. *Immunogenetics.* 2009;61:1-13.
 26. Stranzl T, Larsen MV, Lundegaard C, Nielsen M. NetCTLpan: pan-specific MHC class I pathway epitope predictions. *Immunogenetics.* 2010;62:357-368.
 27. Kim S, Kim HS, Kim E, Lee MG, Shin EC, Paik S, et al. Neopepsee: accurate genome-level prediction of neoantigens by harnessing sequence and amino acid immunogenicity information. *Ann Oncol.* 2018;29:1030-1036.
 28. Bolotin DA, Poslavsky S, Mitrophanov I, Shugay M, Mamedov IZ, Putintseva EV, et al. MiXCR: software for comprehensive adaptive immunity profiling. *Nat Methods.* 2015;12:380-381.
 29. Melnik S, Caudron-Herger M, Brant L, Carr IM, Rippe K, Cook PR, et al. Isolation of the protein and RNA content of active sites of transcription from mammalian cells. *Nat Protoc.* 2016;11:553-565.
 30. Shugay M, Bagaev DV, Turchaninova MA, Bolotin DA, Britanova OV, Putintseva EV, et al. VDJtools: Unifying Post-analysis of T Cell Receptor Repertoires. *PLoS Comput Biol.* 2015;11:e1004503.