

論文の内容の要旨

論文題目 Analysis of the correlation between cancer genomics and histology

(腫瘍における遺伝子変異と病理組織像との相関に関する研究)

氏名 河辺 昭宏

背景と目的

病理医が顕微鏡を用いて病理組織を観察することは病理診断の基礎であり、長年にわたる知見の積み重ねは学問的進歩と臨床への応用に大きな役割を果たしてきた。しかし、病理医の眼に依存する性質上、病理診断における客観性と再現性の担保は必然的な課題であり続けている。

近年の情報技術の進歩により人間の眼に頼らない機械学習による画像認識アルゴリズムの開発が進んでいる。この技術と Whole slide image(WSI)の技術でデジタル化された病理画像を組み合わせることで、病理診断における客観性と再現性の課題を克服できる可能性が指摘され、このため機械学習を病理診断の補助に活用する研究が盛んに行われている。

また、病理組織所見から癌の遺伝子変異を推定する研究については、上記のような画像認識アルゴリズムが開発される以前より行われているが、この分野においても機械学習の応用が可能であると考えられる。

本研究では病理医が病理診断の際に行うような病理画像の判別を機械学習でも実行可能であるという仮説を立て、機械学習の一手法である畳み込みニューラルネットワークを用いて病理組織画像の解析を行い、既知の遺伝子変異と形態学的特徴との関係と比較することで上記仮説の立証を試み、また現状の技術的限界についても考察を行った。

方法

1. 病理画像における特徴ベクトルの抽出と病理組織学的類似性の対応関係

オンラインで公開された The Cancer Genome Atlas (TCGA)の 32 がん種 9,662 枚の診断用スライド上の腫瘍領域のうち、損傷、不鮮明、染色不良でない領域を、スライドごとに3か所ずつ選択し、それぞれの選択領域から 128*128 μ m、512*512 ピクセルのパッチを10枚ずつ位置と向きをランダムに独立に選択・獲得し、計 271,000 枚のパッチを入手した。一般画像のみの学習を行った畳み込みニューラルネットワーク (Convolutional Neural Network; CNN) を用いてこれらの病理画像パッチから 1024 次元の特徴ベクトルを抽出した。非線形次元削減アルゴリズム (Uniform Manifold Approximation and Projection; UMAP) を用いて病理画像パッチの特徴ベクトル間の距離を保ったまま二次元情報に圧縮し、症例ごとの病理画像の特徴の関係性を二次元図上に配置し、病理学的な組織の類似性を再現した配置を認めるかどうか検証した。

2. 病理画像と相関を有する遺伝子変異の抽出及び組織学的所見の抽出

1.と同じ畳み込みニューラルネットワークを用いてがん種ごとにロジスティック回帰モデルによる学習を行い、各症例の病理画像ごとに遺伝子変異を有する確率を判定し、変異の確率上位の組織画像と下位の組織画像を比較することで、変異遺伝子に対応する組織学的特徴の抽出を試みた。遺伝子検索の対象としては OncoKB データベースに登録されているドライバー遺伝子を選定し、うち Actionable genes として 44 遺伝子、All alterations として 322 遺伝子を選定した。その後、がん種の組織画像と変異遺伝子の関係性について、①既存の組織学的分類と明確に相関するもの、②既存の組織学的分類との相関を認めないが、ROC 曲線の Area Under the Curve (AUC)が高く、かつ形態学的特徴が明瞭なものの、③組織学的所見は明瞭でないが、何らかの形態学的特徴が想定されるもの、④組織学的所見が明瞭でなく、組織の形態学的特徴と無関係な情報（色調など）に影響されたものの、の4つのカテゴリーに分類した。

結果

1. 病理画像における特徴ベクトルの抽出と病理組織学的類似性の対応関係

二次元平面状に病理画像を配置した結果、腺癌の症例が集まる領域、扁平上皮癌の症例が集まる領域、肉腫が集まる領域と、病理学的分類と同様に類似した組織像のパッチが近接する領域が多く認められた。一方で、異型上皮が腺腔を形成して浸潤する像と腺腔を形成せず小胞巣状に浸潤する像が混在する領域も存在し、病理学的所見の違いと機械学習による特徴抽出が完全には合致しない例と考えられた。

また、UMAP 上で各がん種が分布する領域を図示した結果、同一種のがん種で症例間の組織像の差が小さいがん種が比較的狭い領域に集簇するのに対し、症例間の組織像の差が大きいがん種や同一がん種でも複数の組織型をとるがん種は広い領域に散在する傾向を認めた。

2. 病理画像と相関を有する遺伝子変異の抽出及び組織学的所見の抽出

① *FGFR3* 遺伝子変異陽性と判定された尿路上皮癌の症例では非浸潤性尿路上皮癌に特徴的な像を呈する。一方、陰性と判定された症例は浸潤性尿路上皮癌の形態をとる。尿路上皮癌の *KRAS* 遺伝子変異症例画像も類似した傾向を認める。*ETV1* 遺伝子変異陽性と判定された尿路上皮癌の症例では腫瘍細胞に高悪性度の所見を認め、また一部症例では浸潤の像を認める。陰性判定症例では明らかな浸潤の所見は認めず、細胞異型も陽性判定症例より軽度である。

また、*PTEN* 遺伝子変異陽性と判定された子宮体癌の症例では子宮体癌のサブカテゴリーである類内膜腺癌を示す組織像を認める一方、陰性判定症例についてはいずれも子宮体癌のサブカテゴリーである漿液性腺癌を示す組織像を示す。

CDH1 遺伝子変異陽性と判定された乳癌症例では、核腫大の目立たない小型の腫瘍細胞が増殖し、腫瘍細胞は列を成すように索状に浸潤し、乳癌のサブカテゴリーである浸潤性小葉癌として矛盾しない組織像を示す。一方、陰性判定症例では同様の組織像を認め

ない。

以上の報告は過去の文献報告と一致しており、機械学習が既存の組織学的分類に相当する特徴を認識可能であることを示している。

- ② *KDM6A* 遺伝子変異陽性と判定された胃癌症例では好塩基性の細胞体と類円形の核を有する腫瘍細胞が網状に増殖し、その間隙に炎症細胞が浸潤する像を認め、*Epstein-Barr virus* 関連胃癌で典型的な像に類似する。*PPM1D* 遺伝子変異陽性と判定された尿路上皮癌症例では核異型の強い腫瘍細胞が比較的小型の胞巣を形成して増殖する一方、変異陰性判定症例では、陽性判定症例を上回る核異型を呈する症例は見出されない。

PIK3R1 遺伝子変異陽性と判定された尿路上皮癌の症例では、個々の腫瘍細胞は変異陰性症例と比較して核腫大が目立つ。*KDM5A* 遺伝子変異陽性と判定された乳癌の症例の大部分では、異型度の高い腫瘍細胞が線維性間質をほとんど伴わずに密に増殖するが、陰性判定症例では、間質組織が腫瘍組織間に比較的多く介在する。*TERT* 遺伝子変異陽性と判定された乳癌症例では泡沫状の類円形腫大核を有する腫瘍細胞が様々な太さの線維性間質を伴って浸潤性に増殖する症例と、核周囲が抜けて見える扁平上皮への分化を示した腫瘍細胞が混在する一方、陰性判定症例では、線維性間質は概ね均一の太さを呈し腫瘍細胞の核は小さく細胞質も乏しい。*NRAS* 遺伝子変異陽性と判定された大腸癌の症例では好酸性の細胞質を有する腫瘍細胞が明瞭な腺管を形成して増殖しており、粘液産生が目立たない。

以上は遺伝子変異と関連した組織所見としては既報のない新規の知見である。

- ③ 子宮頸癌の *NF1* 遺伝子変異については、陽性判定症例と陰性判定症例の双方に扁平上皮癌と腺癌が含まれる。卵巣癌の *MCL1* 遺伝子変異についても、陽性判定症例と陰性判定症例の双方に小型の腺腔を少数形成する症例、管腔を形成せずシート状に増殖する症例、スリット状の間隙を形成して増殖する症例、比較的大型の腺管を形成する症例が混在し、両者で組織構築の明瞭な差異を見出すことができない。

これらの遺伝子変異症例は病理医が認識できず機械学習のみが認識できる特徴を含むものと考えられる。

- ④ 乳癌の *EGFR* 遺伝子変異に関する特徴抽出を示す。全体として陽性判定症例ではスライド全体の色調が濃く、陰性判定症例ではスライド全体の色調が薄いという差があり、機械学習が染色条件の差というアーティファクトを誤抽出した可能性を反映したものと考えた。

考察

機械学習によって病理画像から特徴ベクトルを抽出し、二次元平面状に投影した結果、原発巣が離れていても病理学的に類似した腫瘍が機械学習でもおおむね近い座標に集められる様子を認めたことから、広範な症例において機械学習によって既存の組織学的分類に準じた特徴抽出が行われているといえる。一部で異なる特徴を有する症例が隣接・混在す

る状況を認めるが、本研究では特徴抽出までの段階ではニューラルネットワークは病理画像による学習を行っておらず、より病理画像を学習したアルゴリズムを用いることで改善できる可能性は存在する。

病理組織の特徴抽出に TCGA のデータを用いた研究は過去の文献にも存在するが、本研究では原発巣を限定せず多種多様ながん種において同一のアルゴリズムに基づくニューラルネットワークを用いて遺伝子変異に相関する病理画像の特徴抽出とそれに基づく解析を行った。このような研究は私の知る限り初のものであり、結果として病理医と同様の組織分類が可能であることを示した。

本研究では実際には遺伝子変異を有さない症例でも機械学習によって変異の可能性が高いと判定されるケースが散見され、診断の不確実性は無視できないが、多くの症例の病理画像を用いて学習を行い、また腫瘍領域の選択の方式を工夫することでより成績が向上する可能性はあり、臨床応用に向けさらなる検討を要する。