氏　　名　Lo, Chunhsuan(羅 鈞軒)

学位(専攻分野)　博士(理学)

学 位 記 番 号　総研大甲第 2421 号

学位授与の日付　2023 年 3 月 24 日

学位授与の要件　生命科学研究科　遺伝学専攻
学位規則第6条第1項該当

学位論文題目　Allelic imbalance among cancer associated genes in the pan-adenocarcinoma preferred mutations

論文審査委員　主　査　中村 保一
遺伝学専攻　教授
有田 正規
遺伝学専攻　教授
前島 一博
遺伝学専攻　教授
池尾 一穂
遺伝学専攻　准教授
中岡 博史
公益財団法人佐々木研究所　附属佐々木研究所
部長

(Form 3)

Summary of Doctoral Thesis

Name in full :

Lo, Chunhsuan (羅　鈞軒)

Title :

Allelic imbalance among cancer associated genes in the pan-adenocarcinoma preferred mutations

Cancers are controlled by multiple genes for their disease phenotype, which are supposed to be induced by the cumulative effects of both environmental carcinogens and spontaneous random genetic variations until the triggering point. Recently, a newly proposed continuum model was addressed by Knudson group to replace the classical two-hit hypothesis for the carcinogenesis triggering. Continuum model has emphasized the tumor suppression and the oncogenesis progress as dosage responsive ways, which has implied the importance of regulating mechanisms in gene expression levels. Currently, scientists are still trying to explain how heterozygous mutation carriers of cancer driver genes specifically lift up their expression amount in disease functional proteins specifically for enhancing malignancy, and my hypothesis has supposed that allelic imbalance mechanisms may have roles for it.

To investigate the biological roles of allele-specific transcription, I analyzed large-scale pan-adenocarcinoma RNA-Seq data from TCGA with paired genotype data. Heterozygous somatic mutation carrier genes from the Catalogue of Somatic Mutations in Cancer database were screened for RNA allelic imbalance (RAI) due to transcriptional efficiency. The biased aspects of allelic imbalance of those expressed mutations can further be differently grouped according to cancer driver gene and non-cancer driver gene. Among cancer driver genes, the effects of oncogenesis or the impacts of insufficiency in tumor suppression can be intensified by increasing the

expression of mutant alleles through the observed **allelic imbalance in transcriptional efficiency** (**AITE**) effect. This is in contrast to those mutated non-cancer driver genes, whose normal alleles are instead increased in the expression for buffering the mutant proteins.

Totally, my study analyzed large-scale bulk pan-adenocarcinoma samples from The Cancer Genome Atlas (TCGA) program (n=3256). The genotypes and allele-specific copy numbers inside each gene were firstly checked by SNP array and whole exome sequencing data (WESeq). Then, based on the paired RNA sequencing data (RNA-Seq), the AITE effects were screened for those heterozygous positions without existing allele-specific somatic copy number alterations (SCNA).

Among those somatic mutations, most nonsynonymous SNV mutations of well-known **cancer driver genes biased their expression toward the mutant allele** in statistically significant ways, where the average mutant allele frequency (MAF) in RNA-Seq was significantly **higher** than that in WESeq (Welch's t-test, $p = 4.86 \times 10^{-3}$ for *TP53*; $p = 3.20 \times 10^{-12}$ for *KRAS*; $p = 9.75 \times 10^{-4}$ for *CTNNB1*; $p = 3.80 \times 10^{-3}$ for *PTEN*). In contrast, the nonsynonymous SNV mutations of **passenger genes biased their expression toward the normal (reference) allele**, where the average MAF in RNA-Seq was significantly **lower** than that in WESeq (Welch' t-test, $p = 2.33 \times 10^{-20}$ for *COL3A1*; $p = 5.82 \times 10^{-23}$ for *MYH11*; $p = 2.59 \times 10^{-5}$ for *PDGFRB*; $p = 4.41 \times 10^{-3}$ for *TNC*; $p = 1.01 \times 10^{-12}$ for *PTPRC*). Especially, the bias extent of those AITE genes with increased mutant allele expression could be found in positive correlation with positive selection signals by dN/dS ratio.

Overall, the tumors showed the significant higher AITE occurring rates in their genomes when being compared to the paired-solid-normal tissues (Welch's t-test, $p < 2.2 \times 10^{-16}$, n=152). By further comparing their whole genome sequencing data (WGSeq), more *de novo* transposable elements (TE) insertion sites were also detected inside tumors than that inside normal tissues with significance (Welch's t-test, p=0.0456,

n=27). Within 1 Mbp around of AITE genes, tumor-specific *de novo* TE insertions were discovered to be connected to AITE occurrence.

Finally, the single-cell RNA-Seq analysis has further revealed the intra-tumor heterogeneity in cancer cell transcriptomes. By using liver cancer tissues as a model, my data suggest that the abnormal parenchymal cells and the abnormal hepatic stellate cells inside liver tumors seem to be able to simultaneously express the cancer driver genes and non-cancer driver genes, whose bias sides support the current conclusions in my allelic imbalance studies. In the near future, with the improved single molecular sequencing technology, the single-cells sequencing data in pan-adenocarcinoma are expected to be deposited on the online data bases. And how the allele-specific *de novo* TE insertions can affect the non-stochastic allele-specific expression can be the next research targets for those heterozygous mutated cancer driver genes to enhance the carcinogenesis phenotypes.

Form8 : Separate Sheet
（様式８・別紙１）

Results of the doctoral thesis defense

博士論文審査結果

Name in Full
氏 名　　Lo, Chunhsuan（羅 鈞軒）

Title
論文題目　　Allelic imbalance among cancer associated genes in the pan-adenocarcinoma preferred mutations

申請者は The Cancer Genome Atlas (TCGA) 由来のデータベースから得た RNA-Seq データと遺伝子型データを解析し、汎腺がん pan-adenocarcinoma におけるがんドライバー遺伝子の突然変異対立遺伝子に対する対立遺伝子の不均衡を解析した。汎腺がんのデータとして TCGA から 3,256 サンプルを取得し、SNP アレイと全エキソームシーケンスデータ（WESeq）により、各遺伝子内の遺伝子型とアレル特異的コピー数を算出し、次に RNA-Seq データから申請者が「転写効率における対立遺伝子不均衡（Allelic Imbalance in Transcriptional Efficiency; AITE)」と呼ぶ効果をアレル特異的体細胞コピー数変化のないヘテロ接合部位に於いて抽出した。これらの体細胞変異のうち、よく知られたがんドライバー遺伝子の非同義一塩基変異サイトの大部分において、正常アレルに比べて変異アレルの発現量が亢進しており、RNA-Seq の平均変異アレル頻度（Mutant Allele Frequency; MAF）は WESeq の MAF より有意に高かった。一方、パッセンジャー遺伝子の非同義一塩基変異サイトにおいては、正常アレルに発現が偏り、RNA-Seq の平均 MAF は WESeq の MAF より有意に低かった。全体として、腫瘍組織は正常組織と比較してゲノム中の AITE 発現率が有意に高いことが示された。さらに、全ゲノム塩基配列データを比較すると、腫瘍内には正常組織内よりも多くの *de novo* の transposable elements (TE) 挿入が検出された。AITE 遺伝子の周囲 1 Mbp 以内では、腫瘍特異的な *de novo* TE 挿入が AITE の発生と関連していることが示唆された。最後に申請者は、シングルセル RNA-Seq 解析によりがん細胞のトランスクリプトームにおける腫瘍内部での不均一性を示し、肝がん組織をモデルとした場合に肝がん内部の異常な実質細胞や肝星細胞はがんドライバー遺伝子と非がんドライバー遺伝子を同時に発現するが、その偏りが申請者の結論を支持すると考察した。

申請者はがんドライバー遺伝子のヘテロ接合型変異キャリアが特異的に疾患機能性タンパク質の発現量を上昇させ悪性度を高めるメカニズムについて、対立遺伝子不均衡メカニズムがその役割を担っている可能性を想定して研究を計画し、公開データを用いて細胞のがん化とその進行要因を理解するための大規模な情報処理と統計解析を行い、上記の結論を導いた。以上の理由により審査委員会は、本論文が学位の授与に値すると判断した。