in silico analysis of gene expression dynamics during human iPS cell generation

氏　　名	久野　朗広
学位の種類	博士（　人間生物学　）
学位記番号	博甲第　　８７７０　　号
学位授与年月	平成　３０年　４月　３０日
学位授与の要件	学位規則　第４条第１項該当（昭和２８年４月１日文部省令第９号）
審査組織	グローバル教育院
学位論文題目	*in silico* analysis of gene expression dynamics during human iPS cell generation （ヒトiPS細胞形成過程における遺伝子発現変動の生物情報学解析）

	（職名）	（学位）	（氏名）
主　　査	筑波大学教授	医学博士	柳沢　正史
副　　査	筑波大学教授（ｸﾞﾛｰﾊﾞﾙ教育院）	医学博士	佐藤　孝明
副　　査	筑波大学教授	医学博士	高橋　智
副　　査	筑波大学准教授	Ph.D.	Kiong Ho

論文の要旨

Purpose

The generation of human induced pluripotent stem cells (hiPSCs) from somatic cells by induction of OCT3/4, SOX2, KLF4, and MYC (OSKM) has revolutionized not only stem cell biology but also clinical medicine. hiPSCs have provided the new strategy to investigate how cell fate is controlled, to clarify mechanisms of disease from patients-derived iPSCs, and to apply regenerative medicine, to uncover ways to apply stem cell treatment for several diseases. Because various types of somatic cells from patients are available for iPSCs source, there would be a shared reprogramming route in various cell types. Recent study indicated that reprogramming process from fibroblasts to iPSCs could be categorized 3 phases: initiation, maturation, and stabilization. Initiation is characterized by mesenchymal-to-epithelial transition (MET). Maturation is marked by the beginning of the expression of pluripotency genes including Nanog, Sall4, Esrrb, which occurs at about day 8 of reprogramming in mouse embryonic fibroblasts (MEFs). Stabilization is the phase that pluripotency genes are stably expressed without exogenous OSKM gene expression, which starts at around day 21 in MEFs. Previous study suggested maturation step is a main roadblock for reprogramming from human dermal fibroblasts into hiPSCs. However, the obstruction of reprogramming process of other cell types is largely unknown although various types of cells are available as hiPSCs source. For the reason, the applicant investigated whether the common reprogramming routes exist across various human cell types and to address whether the maturation process is a major blockage for reprogramming regardless of the somatic cell types.

Methods

For the analysis to identify genes with dynamic expression during reprogramming in various human cell types, the applicant used microarray data of five human somatic cell types: HDF (Human Dermal Fibroblasts), ASC (Adipose-derived Stem Cells), HA (Human Astrocyte), NHBE (Normal Human Bronchial Epithelial cells) and PrEC (Prostate Epithelial Cells), as well as two stem cells: hiPSC, and hESC. To find a common route of reprogramming,

dynamically expressed genes across different cell types during reprogramming process were identified. The applicant proceeded analysis of covariance (ANOVA) and constructed a regression model to discover genes with significant differential expression during reprogramming in five cell types (P-value < 0.01, FDR < 0.05, $R^2 > 0.6$). Consequently, the filtration yielded 3615 genes. Next, the applicant applied principal component analysis (PCA) and hierarchical clustering analysis (HCA) using these 3615 genes in order to analyze the similarity of transcriptome in each sample. In addition, 3615 genes were categorized into 5 clusters by HCA and the biological roles of each cluster were analyzed using an online tool (http://metascape.org). To identify the critical transcription factors (TFs) during reprogramming, the applicant applied CoRegNet, which provides an algorithm based on association analysis to measure correlations of TFs and non-TFs gene expression.

Results

The applicant identified 3615 genes which have dynamic expression changes during reprogramming process. The applicant evaluated the data for overall similarity between samples using PCA and HCA. The results indicate there are 3 distinct transcriptomic phases after induction of OSKM reprogramming factors: the early phase from day 0 to 3, the mid phase from day 7 to 15, and the late phase from day 20 to later. The greatest phase-to-phase distances exist between mid to late phase.

To study molecular events during reprogramming, the applicant categorized 3615 genes into 5 separate groups according to their gene expression patterns during reprogramming. The functional annotation of the gene lists in each group shows common reprogramming events in 5 cell types. The first group containing 816 genes had a higher expression in the early phase and remained repressed during reprogramming process. The 816 genes were mainly annotated as the extracellular matrix organization, which could directly influence cell proliferation and differentiation. The second group including 536 genes had high expression during the early and the mid-phase but decreased their expression in the late phase. These 536 genes were annotated as immune response-related genes, which can be caused by the effect of retroviral OSKM induction. The gene expressions in the third cluster were transiently up regulated only in the mid-phase, which was enriched by hemidesmosome and epidermal development-related genes. The genes in the fourth cluster had a sharp up regulation in the late phase of reprogramming, which was annotated as trans-synaptic signaling related genes. The genes in the fifth cluster showed gradually increased in expression as the reprogramming progressed. They were highly annotated as cell cycle related genes. Briefly, the functional annotation of gene clusters classified by gene expression patterns showed the mesenchymal-epithelial transition from day 0 to 3, transient up regulation of epidermis related genes from day 7 to 15, and up regulation of pluripotent genes from day 20, partially similar to the reprogramming process of mouse embryonic fibroblasts.

Because TFs could regulate cell fate by controlling their target gene expression, the applicant focused on identifying candidate TFs which play critical role in each reprogramming phase. For this purpose, the applicant scored influences of TFs and reconstructed TF network by CoRegNet. TF influences clearly exhibited two distinct clusters dependent on the days of reprogramming. The pluripotency-related TFs such as NANOG, SALL4, OCT4 and SOX2 had the influence in the late phase. On the other hand, tissue morphogenesis associated TFs, such as EHF, MEF2C, and FOXE1, had the influences in the early phase. Next, the applicant visualized the co-regulatory network of the TFs for each time point of the reprogramming process. The time-course TF network illustrated that the positive influence TFs from day 0 and 15 had a sparse network compared to the negative influence TFs. In contrast, positive influence TFs from day 20 network was denser than negative influence TFs. This would reflect the heterogeneous cell status in different phases. Furthermore, no co-regulatory networks were observed between positive influence TFs and negative influence TFs in all phases, especially between the mid and late phase. These results suggested that the transition of TF influence occurs between the mid and late phase.

審査の要旨

【批評 Review】

The dissertation suggests that human cellular reprogramming process could be traced into 3 different phases upon OSKM induction across different cell types: the early phase from day 0 to 3, the mid phase from day 7 to 15, and the late phase from day 20 to later. As the late phase exhibited the greatest dissimilarity based on transcriptome and transcription factor activity analysis and the pluripotency genes such as NANOG and SALL4 start to express in mid phase in the microarray data, the mid phase in this study could be correspond to the maturation. This study suggests that maturation step could be a major roadblock during human cellular reprogramming, regardless of cell types. Furthermore, expression of 71 TFs were dynamically altered during reprogramming, and thus, these TFs may play important roles in regulating the reprogramming process. This dissertation provides important information about the molecular mechanism of human cellular reprogramming process.

【最終試験の結果 Result】

The final examination committee conducted a meeting as a final examination on 26 January, 2018. The applicant provided an overview of dissertation, addressed questions and comments raised during Q&A session. All of the committee members reached a final decision that the applicant has passed the final examination.

【結論 Conclusion】

The final examination committee approved that the applicant is qualified to be awarded Doctor of Philosophy in Human Biology.