

Project Code: RMU5080066

Project Title: A Search for Tumor Markers for Cholangiocarcinoma by Serum Protein Profiling

Investigator: Dr. Rutaiwan Tohtong, Department of Biochemistry, Faculty of Science, Mahidol University

Email address: [rutaiwan.toh@mahidol.ac.th](mailto:rutaiwan.toh@mahidol.ac.th)

Project Period: 3 years (2007-2010)

.....

### **Abstract**

Cholangiocarcinoma (CCA) is a highly lethal disease. The high mortality rate of this disease is largely due to lack of early detection, hence, in most cases, the disease has already advanced to late stages with local or distant metastasis at the time of detection. If the disease can be detected, hence effectively controlled, at early stages, the survival rate of CCA patients is likely to be improved significantly. Currently available tumor markers such as CA19-9 and CEA are rather inefficient and of limited usage due to low sensitivity and specificity. As such, a tumor marker with high sensitivity and specificity is urgently needed. Here, we attempt to identify a serum tumor marker for CCA that can effectively distinguish CCA from CBD (common bile duct disease) samples.

The serum samples from 16 CCA patients and 14 controls (CBD, common bile duct disease) were separated by SDS-PAGE followed by LC-MS/MS and subjected to statistical analysis to identify peptides whose abundance are significantly elevated or suppressed in CCA samples compared to controls. Eight peptides were identified with AUC (Area Under the ROC curve, ROC = Receiver Operating Characteristic) above 0.80, suggesting that these peptides should serve a more efficient marker for CCA compared to CA19-9 (AUC is 0.70). From the 8 potential markers, IL-28 was chosen for validation using an independent set of subjects comprising 32 CCA sera and 32 CBD sera and ELISA technique. In contrast to the data obtained from the training set, IL-28 exhibited low sensitivity, specificity and AUC score when tested in the validation set. Our data showed that IL-28 could not serve as a tumor marker for CCA, and it performs worse than the currently available marker such as CA19-9. We explained that failure of IL-28a to reproduce in the validation set may be, at least in part, due to 'overfitting', a phenomenon which arises when a large amount of 'predictors' are used to discriminate among a small outcome events. Due to the limitation in the technology and lack of well-developed rules of evidence for data analysis, we suggest that it is premature to use proteomics for 'discovery-based' research as it could generate false-positive data which cannot be reproduced.

รหัสโครงการ: **RMU5080066**

ชื่อ โครงการ: การหา Tumor Markers ของมะเร็งท่อน้ำดีด้วยการทำ Serum Protein Profiling

ชื่อหลักวิจัย: นางสาวฤทัยวรรณ โต๊ะทอง ภาควิชาชีวเคมี คณะวิทยาศาสตร์ มหาวิทยาลัยมหิดล

Email address: [rutaiwan.toh@mahidol.ac.th](mailto:rutaiwan.toh@mahidol.ac.th)

ระยะเวลาโครงการ : 3 ปี (2550-2553)

---

#### บทคัดย่อ

มะเร็งท่อน้ำดี (CCA) เป็นโรคที่มีอัตราการเสียชีวิตของผู้ป่วยสูงมาก ทั้งนี้เนื่องจากขาดวิธีการตรวจหาโรคในระยะเริ่มต้น ดังนั้นผู้ป่วยส่วนใหญ่ตรวจพบโรคเมื่อโรคได้ลุกลามหรือแพร่กระจายไปมากแล้ว หากสามารถตรวจพบโรคได้ตั้งแต่ระยะแรก ผู้ป่วยอาจมีโอกาสรอดชีวิตเพิ่มมากขึ้น Tumor marker ที่มีอยู่ในปัจจุบัน เช่น CA19-9 และ CEA มีประสิทธิภาพต่ำ ดังนั้น จึงมีความจำเป็นที่จะต้องค้นหา Tumor marker ที่มีประสิทธิภาพสูง ทั้งทางด้านความไว (sensitivity) และ ความจำเพาะ (specificity) โครงการวิจัยนี้มุ่งที่จะหา Tumor marker จากซีรัมของผู้ป่วย CCA ที่สามารถจำแนก CCA ออกจาก โรคท่อน้ำดีอุดตัน (CBD, common bile duct disease) ได้อย่างมีประสิทธิภาพ

ตัวอย่างซีรัมจากตัวอย่างที่เป็น training set ซึ่งประกอบด้วยผู้ป่วย CCA จำนวน 16 รายและ ผู้ป่วย CBD จำนวน 14 รายถูกนำมาแยกโดยเทคนิค SDS-PAGE และ LC-MS/MS จากนั้นวิเคราะห์ด้วยวิธีการทางสถิติเพื่อค้นหาเปปไทด์ที่มีระดับแตกต่างกันอย่างมีนัยสำคัญในระหว่างสองกลุ่มตัวอย่างนี้ เราพบว่าเปปไทด์ 8 ชนิดมีค่า AUC (Area Under the ROC curve, ROC = Receiver Operating Characteristic) สูงกว่า 0.80 ซึ่งบ่งชี้ว่าเปปไทด์เหล่านี้มีความสามารถในการจำแนก CCA ได้สูงกว่าความสามารถของ CA19-9 จากจำนวน 8 เปปไทด์ IL-28a ได้ถูกเลือกมาเพื่อทดสอบ (validate) ในอีกกลุ่มตัวอย่างหนึ่ง (validation set) ซึ่งประกอบด้วยซีรัมจากผู้ป่วย CCA จำนวน 32 ราย และ ผู้ป่วย CBD จำนวน 32 ราย แต่เมื่อทดสอบด้วย validation set พบว่าค่า AUC, sensitivity และ specificity ของ IL-28a ต่ำมาก และ ต่ำกว่าค่าของ CA19-9 ซึ่งเป็น tumor marker ที่ใช้อยู่ในปัจจุบัน ดังนั้น IL-28a จึงไม่สามารถใช้เป็น tumor marker ได้ สาเหตุที่เราไม่สามารถ validate IL-28a ได้นั้น อาจเป็นเพราะเกิดปรากฏการณ์ 'overfitting' ซึ่งเกิดขึ้นเมื่อมีตัวแปร (predictors) เป็นจำนวนมากที่ใช้เพื่อจะจำแนกกลุ่มตัวอย่างจำนวนน้อย (กรณีนี้สองกลุ่ม คือ กลุ่ม CCA และ กลุ่ม CBD) เทคนิค proteomics ยังมีข้อจำกัดอีกมากทั้งทางด้านเทคโนโลยีและการวิเคราะห์ข้อมูล ทำให้ไม่เหมาะที่จะใช้ในการวิจัยแบบ 'discovery-based' เนื่องจากมีความเป็นไปได้สูงที่จะให้ผลที่เป็น false positive และไม่สามารถ reproduce ผลการทดลองได้