JAIST Repository
https://dspace.jaist.ac.jp/

Title	計算アプローチによる肝炎の病態・治療に関する分子機構の解明
Author(s)	Ho, Tu Bao
Citation	科学研究費助成事業研究成果報告書: 1-6
Issue Date	2014-06-09
Type	Research Paper
Text version	publisher
URL	http://hdl.handle.net/10119/12169
Rights	
Description	研究種目：基盤研究(B)，研究期間：2011～2013，課題番号：23300105，研究者番号：60301199，研究分野：データマイニング，科研費の分科・細目：情報学・統計科学

Japan Advanced Institute of Science and Technology

様　式　C－１９、F－１９、Z－１９　（共通）

科学研究費助成事業　　研究成果報告書

平成２６年　６月　９日現在

機関番号：１３３０２
研究種目：基盤研究(B)
研究期間：2011～2013
課題番号：２３３００１０５
研究課題名（和文）計算アプローチによる肝炎の病態・治療に関する分子機構の解明

研究課題名（英文）Elucidation of the molecular mechanisms on the pathology and treatment of hepatitis by computational approach

研究代表者
Tu・Bao　Ho（Ho, Tu Bao）

北陸先端科学技術大学院大学・知識科学研究科・教授

研究者番号：６０３０１１９９

交付決定額（研究期間全体）：（直接経費）　15,600,000円、**（間接経費）**　4,680,000円

研究成果の概要（和文）：第1課題のHCV NS5AのIFN-RBV治療への耐性機序について、多数のNS5Aデータにおいて著効データが少数である場合に有効な準教師付アンサンブル学習手法を開発し、SVRと非SVRの2種の患者群を特徴づけるモチーフを発見した。第2課題のsiRNA抑制効果用に開発した手法では、既知の設計規則に加え2種の新siRNA設計規則を発見し、また他手法では所与のsiRNA配列についてスコアと著効情報をもつ配列と既知の規則で配列の表現力を補強し、新規のテンソル回帰手法にて予測精度を大幅に改善した。第3課題ではエピジェネティク因子と肝炎進行の相互作用に関するヒストン修飾の因果関係網等の中間結果を得た。

研究成果の概要（英文）：For the first task on HCV NS5A resistance mechanisms to interferon/ribavirin therapy, our semi-supervised ensemble method discovered the motifs that well characterizing two class of patients with SVR and non-SVR, especially in case of only a small number labelled NS5A sequences but much unlabelled sequences are available. For the second task on the knockdown efficacy of siRNA, one of our method discovered two siRNA design rules in addition to known design rules, and the other significantly improve the predictive ability for given siRNA sequences. This method employs both scored and labelled sequences as well as known design rules to enrich the poor sequence representation of siRNA by transformed matrices. The prediction is done by a novel tensor regression method on those matrices. For the third task on the interplay between epigenetic factors and hepatitis progression, we reached some intermediate results such as inferring the causual relationship network of histone modifications.

研究分野：データマイニング

科研費の分科・細目：情報学・統計科学

キーワード：医薬生物　統計解析　データマイニング

１．研究開始当初の背景

Viral hepatitis is a disease in which liver tissue is inflamed by the infection of hepatitis viruses. Hepatitis viruses persisting in the liver of most infected people can be treated with medication but the effect of the current hepatitis therapy is still quite limited. The main issue in this study is molecular mechanisms of hepatitis pathogenesis and therapies for hepatic diseases that are poorly understood and remain to be answered. Our joint project focuses on

(1) ***HCV NS5A resistance mechanisms to interferon & ribavirin therapy:*** HCV is eradicated in only nearly a half of patients treated by current standard therapy of peg-interferon combined with ribavirin (IFN/RBV). Analysis of the virus genomes and their drug resistance mechanisms to this therapy has been pursued for years. Recently, NS5A is known as the protein most reported to be implicated in the interferon resistance, and NS5A inhibitor is a hot topic in HCV research in very recent years (Gao et al., ***Nature*** 465, 2010). NS5A inhibits IFN activity via its interaction with IFN cellular antiviral pathways and the mutations in NS5A resist IFN therapy (Guillou et al.,***World J. Gastroenterology*** 13, 2007). Many questions on this topic remain unanswered such as the enigmatic role of the domains II and III of NS5A or can V3 region in NS5A be a more accurate biomarker than the ISDR region?

(2) ***Selection of potent siRNAs for silencing hepatitis viruses***: RNAi is a cellular pathway wherein small RNA molecules, typically siRNAs and miRNAs, control gene expression, viewed as "one of the most exciting discoveries in biology in last couples of decades" (Fire et al., ***Nature***, 1998; Nobel prize 2006). RNAi is known as a new therapeutic strategy against hepatitis viruses, especially siRNAs target to HBV, HCV genes to inhibit their replication or host genes required for their replication. Chemically synthesized siRNAs can mimic the native siRNAs produced by RNAi but having different ability and thus selecting appropriate siRNAs before using them in experiments is crucial. It was firstly solved by using guidelines to select siRNAs from designed databases, e.g., siDirect (Naito et al., ***BMC Bioinformatics***, 2009), E-RNAi (Arziman et al., ***Nucl Acids Res.***, 2005), NEXT-RNAi (Horn, ***Genome Bio***, 2010). Recently, there are computational works such as using combined neural network and decision tree (Takasaki, ***Computers in Biology and Medicine*** 2010) or kernel regression (Qiu and Lane, ***IEEE Computational Biology and Bioinformatics***, 2009). However, they mainly focus on either knockdown efficiency or off-target effect of siRNAs. To better solve the problem, we target to computational methods for simultaneously maximum knockdown efficiency and minimum off-target effect of selected siRNAs.

(3) ***Interplay between epigenetic factors in hepatitis progression***: RNAi is a cellular pathway wherein small RNA molecules, typically siRNAs and miRNAs, control gene expression, viewed as "one of the most exciting discoveries in biology in last couples of decades" (Fire et al., ***Nature***, 1998; Nobel prize 2006). RNAi is known as a new therapeutic strategy against hepatitis viruses, especially siRNAs target to HBV, HCV genes to inhibit their replication or host genes required for their replication. Chemically synthesized siRNAs can mimic the native siRNAs produced by RNAi but having different ability and thus selecting appropriate siRNAs before using them in experiments is crucial. It was firstly solved by using guidelines to select siRNAs from designed databases, e.g., siDirect (Naito et al., ***BMC Bioinformatics***,2009), E-RNAi (Arziman et al., ***Nucl Acids Res.***, 2005), NEXT-RNAi (Horn, ***Genome Bio***, 2010). Recently, there are computational works such as using combined neural network and decision tree (Takasaki, ***Computers in Bio. and Med.***, 2010) or kernel regression (Qiu and Lane, ***IEEE Comp. Bio. Bioinf.***, 2009). However, they mainly focus on either knockdown efficiency or off-target effect of siRNAs. To better solve the problem, we target to computational methods for simultaneously maximum knockdown efficiency and minimum off-target effect of selected siRNAs.

２．研究の目的

(1) Discover the molecular mechanisms of HCV NS5A resistance to IFN/RBV therapy. We have successfully detected NS5A motifs that characterize well responders and non-respsonders in HCV sub-genotypes 1a-c and 1b. We aim at discovering the partnership of these motifs with IFN cellular antiviral pathways and mutations in NS5A to reveal its unknown IFN/RBV resistance mechanisms.

(2) Uncover the combinatorial effects of various epigenetic factors, especially DNA methylation and PTMs in hepatitis development, which can be divided into three concrete objectives: ①Identify key epigenetic factors in suppressor gene silencing, ② Characterize their combinatorial effects, and ③ Explore deeply the epigenetic regulatory network, which includes the possible correlation, dependency, or causal relationship among those factors.

(3) Develop computational methods for selecting siRNA that simultaneously minimize off-target effect and maximize knockdown efficiency in silencing hepatitis viruses.

(4) Quantitatively analyze the relationships of the above three topics to better understand the hepatitis pathogenesis and therapies and prove that the multiple therapeutic strategy can increase the successful ratio of hepatitis treatment, says, more than 50%.

３．研究の方法

The three problems investigated in our project (Fig. 1) were investigated by computational approach with three main components of data, analysis methods and evaluation methods as illustrated in Figs 2-4, respectively.

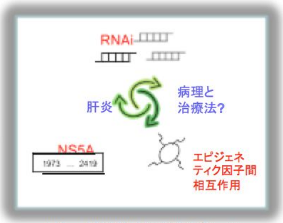
RNAi
肝炎
病理と
治療法?
NS5A
1973 ... 2419
エピジェネ
ティク因子間
相互作用

図1.　研究目的およびアプローチ

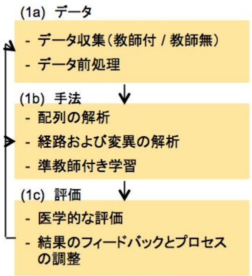
(1a) データ
- データ収集（教師付 / 教師無）
- データ前処理
(1b) 手法
- 配列の解析
- 経路および変異の解析
- 準教師付き学習
(1c) 評価
- 医学的な評価
- 結果のフィードバックとプロセスの調整

図2. NS5AとIFN/RBV療法に関して

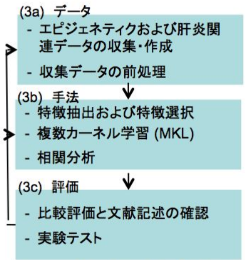
(3a) データ
- エピジェネティクおよび肝炎関連データの収集・作成
- 収集データの前処理
(3b) 手法
- 特徴抽出および特徴選択
- 複数カーネル学習 (MKL)
- 相関分析
(3c) 評価
- 比較評価と文献記述の確認
- 実験テスト

図4.　エピジェネティク因子群とHCC

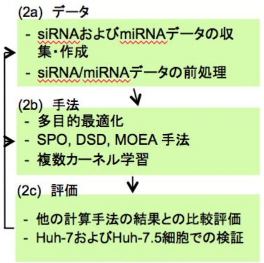
(2a) データ
- siRNAおよびmiRNAデータの収集・作成
- siRNA/miRNAデータの前処理
(2b) 手法
- 多目的最適化
- SPO, DSD, MOEA 手法
- 複数カーネル学習
(2c) 評価
- 他の計算手法の結果との比較評価
- Huh-7およびHuh-7.5細胞での検証

図3. RNAiと肝炎研究

For the first problem on NS5A, we used 77 labeled data on SVR (sustained viral response) and non-SVR NS5A sequences from Los Alamos National Database (http://hcv.lanl.gov) plus 25 labeled sequences from Chiba University Hospital (to our knowledge, only such 102 labeled sequences are available in total), and 1424 unlabeled NS5A sequences from Hepatitis Virus Database (http://s2as02.genes.nig.ac.jp) plus 168 sequences from GenBank (http://www.ncbi.nlm.nih.gov/genbank).We developed a semi-supervised ensemble method for sequence analysis that can well discover discriminative motifs (a new direction in motif learning) especially in case of only a small number labelled NS5A sequences but much unlabelled sequences are available. The evaluation is carefully done with 10-fold cross validation on two kind of found motifs. The obtained results are well characterizing two classes of patients with SVR and non-SVR.

For the second problem on siRNAs, we all two kinds of available scored and labeled siRNA sequences obtained by different laboratories. In particular, we employ the following datasets: ① The Huesken dataset of 2431 siRNA sequences targeting 34 human and rodent mRNAs, commonly divided into the training set HU train of 2182 siRNAs and the testing set HU test of 249 siRNAs (Huesken et al., Nature 2005); ② Three independent datasets for evaluation, including the Reynolds set of 240 siRNAs (Reynolds et al., 2004), the Vicker dataset of 76 siRNA sequences targeting two genes (Vicker et al., 2003), and the Harborth dataset of 44 siRNA sequences targeting one gene (Harborth et al., 2003). Our developed method to predict the knowckdown efficacy of a given siRNA based on the following key ideas to overcome the limitations of previous mothods: ① Exploiting both available scored and labeled siRNA datasets as well as the discovered siRNA design rules to enrich siRNA sequences by converting them into matrices with more integrated information, and ② Learning a tensor regression on those transformed matrices and predict the knockdown efficacy of new siRNA with this tensor regression.

For the third problem on epigenetic factors in hepatitis progression, we have yet reached the final target of understanding the impact of epigenetic factors in hepatitis disease. There are two main reasons of that limitation. One is a good dataset on such relationship has yet available during the research period, and the other is finding such relationship much be based on several well known other processes such as functional linkage between nucleosome dynamics and protein binding profiles or transcriptional relationships. Instead, we focused on investigating the reconstruction of those relation networks from data. In fact, we particularly use different available next-generation sequencing data of the research community for the purpose, such as ChIP-Chip data to learn transcriptional relationships. Different statistical learning methods have been employed with adaptation such as multiple kernel learning, especially methods of probability graphical models to learn structures and parameters of the reconstructed networks.

４．研究成果

For the first task on HCV NS5A resistance mechanisms to interferon/ribavirin therapy, our semi-supervised ensemble method discovered the motifs that well characterizing two class of patients with SVR and non-SVR, especially in case of only a small number labelled NS5A sequences but much unlabelled sequences are available. For the second task on the knockdown efficacy of siRNA, one of our methods discovered two siRNA design rules in addition to known design rules, and the other significantly improved the predictive ability for given siRNA sequences. The latter employs both scored and labelled sequences as well as known design rules to enrich the poor sequence representation of siRNA by transformed matrices, on which a novel tensor regression method makes the prediction task. For the third task on the interplay between epigenetic factors and hepatitis progression, we reached some intermediate results such as inferring the causual relationship network of histone modifications.

５．主な発表論文等

（研究代表者、研究分担者及び連携研究者には下線）

〔雑誌論文〕（計 12 件）

(1) Nakamoto, S., Kanda, T., Wu, S., Shirawasa, H., Yokosuka, O., Hepatitis C virus NS5A inhibitors and drug resistance mutations, World Journal of Gastroenterol, 20(11): 2902-12, 2014. 査読有

(2) Than, K., Ho, T.B., Modeling the diversity and log-normal of data, ***Intelligent Data Analysis***, Volume 18(6), 2014 (in press). 査読有

(3) Le, N., Ho, T.B., Kanda, T., Kawasaki, S., Takabayashi, K., Wu, S., Yokosaka, O., A Semi-Supervised Learning Method for Discriminative Motif Finding and Its Application, ***Journal of Universal Computer Science***, Vol. 19. No. 4, 563-580, 2013. 査読有

(4) Kanda T, Yokosuka O, Omata M., Antiviral therapy for " difficult -to-treat " hepatitis C virus-infected patients, ***Chin Med J (Engl),*** 126(23): 4568-74, 2013. 査読有

(5) Kanda T, Yokosuka O, Omata M., Treatment of hepatitis C virus infection in the future, Clinical Translation Medicine, 11, 2(1):9. doi: 10.1186/2001 -1326-29, 2013. 査読有

(6) Kandat T., Nakamoto S., Arai, M., Miyamura, T., Wu, S., Fujiwara, K., Yokosuka O., Natural interferon-beta plus ribavirin therapy led to sustained virological response after seven unsuccessful courses of anti-viral treatment in a chronic hepatitis C patient, ***Clinical Journal Gastroentorology***, 6(2), DOI:10.1007/ s12328- 013-0366-1, 2013. 査読有

(7) Kandat T., Kato K., Tsubota, A., Takada, N., Nishino, T., Mikami, S., Miyamura, T., Maruoka D., We, S., Nakamoto, S., Arai, M., Fujiwara, K., Imazeki, F., Yokosuka O., Platelet count and sustained virological response in hepatitis C treatment,

World Journal of Hepatology, 5(4):182-188, 2013. 査読有

(8) Le, N.T, Ho, T.B., Ho, B.H., Computational reconstruction of transcriptional relationships from ChIP-Chip data, ***IEEE-ACM Trans. On Computational Biology and Bioinformatics***, Vol. 10. No. 2, 300-307, 2012. 査読有

(9) Ho, B.H., Le, N.T., Ho, T.B., Quantitatively assessing the effect of regulatory factors on nucleosome dynamics, ***Journal of Ambient Intelligence and Humanized Computing***, Springer, Vol. 3, Issue 4, 265-280, 2012. 査読有

(10) Takabayashi K., The cutting-edge of medicine. Application of medical informatics in internal medicine, ***Nihon Naika Gakkai Zasshi.***,10;101 (11):3239-46, 2012. 査読有

(11) Nguyen, T.P., Ho, T.B., Detecting Disease Genes Based on Semi-Supervised Learning and Protein-Protein Interaction Networks, ***Artificial Intelligence in Medicine***, Vol. 54, 63-71, 2011. 査読有

(12) Tran, D.H., Ho, T.B., Pham, T.H., Satou, K., microRNA expression profiles for classification and analysis of tumor samples, ***IEICE Trans. Information Systems***, E94.D (3), 416-422, 2011. 査読有

〔学会発表〕（計 14 件）

(1) Le, N.T, Ho, T.B., Ho, B.H., Tran, D.H. A nucleosomal approach to inferring causal rela- tionships of histone modifications, ***Asia-Pacific Bioinformatics Conference APBC 2014***, 17-19 January 2014, Shanghai, China.

(2) Bui N.T., Ho. T.B., A novel framework to improve siRNA efficacy prediction, ***Pacific-Asia Conference on Knowledge Discovery and Data Mining PAKDD'04***, 13-16 May 2014, Tainan, Taiwan.

(3) Nguyen, D.K., Than, K., Ho, T.B., Simplicial Nonnegative Matrix Factorization, ***IEEE International Conference on Research, Innovation and Vision for Future***, RIVF 2013, November 10-13, 2013, Hanoi, Vietnam.

(4) Nguyen, Q.D., Pham, T.H., Ho, T.B., Nguyen, V.H., Tran, D.H., Reconstruction of Triple-wise Relationships in Biological Networks from Profiling Data, ***The 9th International Conference on Computing and Information Technology*** (IC2IT2013), Bangkok, 9-10 May 2013, Springer AISC 209, 205-215, Thailand.

(5) Bui, N.T., Ho, T.B., Kawasaki, S., An Effective Method for Generating siRNA Design Rules, ***The 5th Asian Conference On Intelligent Information and Database Systems, ACIID 2013***, Kuala Lumpur, 18-20 March 2013, LNAI 7803, 196-205, Malaysia.

(6) Ho, T.B., Takabayashi, T., Kanda, T., Kawasaki, S., Le, T.N., Bui, N.T., Than, Q.K., From Clinical to Genomics Data in Hepatitis Study, ***The First Asian Conference on Information Systems***, SiemRiep 6-8 December 2012, Cambodia.

(7) Bui, N.T., Ho, T.B., Kawasaki, S., A Sequential Apriori Algorithm for Discriminative Design Rules of Effectiv siRNA Sequences, ***13th International Symposium on Knowledge and Systems Science,*** Kanazawa 19-20 November 2012.

(8) Ho, B.H., Ho, T.B., Analysis of gene expression behaviour by nucleosome and protein binding profiles, ***The 10th Asia Pacific International Conference on Bioinformatics and Biomedical Technology, ICBBT 2012***, 26-28 February 2012, Singapore.

(9) Ho, B.H., Le, N.T., Ho, T.B., Analysis of functional linkages between nucleosome dynamics and trans-regulatory factors, ***4th International Conference on Bioinformatics and Biomedical Technology 2012, BICoB -2012***, 12-14 March 2012, Las Vegas, USA.

(10) Le, N.T., Ho, T.B., Ho, B.H., Computational reconstruction of transcriptional relationships from ChIP-Chip data, ***The 10th Asia Pacific Bioinformatics Conference***, 17-19 January 2012, Melbourne, Australia.

(11) Pham, T.H., Ho, T.B., Nguyen, Q.D., Nguyen, V.H., Tran, D.H., Multivariate Mutual Information Measures for Discovering Biological Networks, ***IEEE RIVF International Conference on Information and Communication Technologies***, Ho Chi Minh City, February 27-March 2, 2012, Vietnam.

(12) Le, N., Ho, T.B., A Semi-Supervised Method for Discriminative Motif Finding and Its Application to Hepatitis C Virus Study, ***4th Asian Conference on Intelligent Information and Database Systems ACIIDS 2012***, 19-21 March 2012, Kaohsiung, Taiwan.

(13) Ho, T.B., Kawasaki, S., Le, N.T., Kanda, T., Le, T.N., Takabayashi, K., Yokosuka, O., Finding HCV NS5A Discriminative Motifs for Assessment of IFN/Ribavarin Therapy Effect, ***Workshop Data Mining in Genomics and Proteomics***, International Conference ECML/PKDD, September 5-9, 2011, Athens, Greece.

(14) Le, N.T., Ho, T.B., Reconstruction of histone modification network from next-generation sequencing data, ***IEEE International Conference on Bioinformatics and Bioengineering BIBE 2011,*** October 24-26 2011, Taichung, Taiwan.

〔図書〕（計２件）

(1) Tseng, V.S., Ho, T.B., Zhou, Z.-H., Chen, A.L.P., Kao, H.-Y., Lecture Notes in Artificial Intelligence LNAI 8443, Springer. ***Advances in Knowledge Discovery and Data Mining*** (Eds.), Proceedings of the 18th Pacific-Asia Conference on Knowledge Discovery and Data Mining, 2014, 2500 pages.

(2) Nguyen, T.P., Ho, T.B., Springer-Verlag, Mining multiple biological data for reconstructing signal transduction networks (book chapter), ***Data Mining: Foundations and Intelligent Paradigms***, D.E. Holmes, L.C. Jain (Eds.), 163-185, 2011.

〔産業財産権〕
○出願状況：（計０件）
名称：
発明者：
権利者：
種類：
番号：
出願年月日：
国内外の別：

○取得状況（計０件）

名称：
発明者：
権利者：
種類：
番号：
取得年月日：
国内外の別：

〔その他〕
ホームページ等

６．研究組織

(1)研究代表者
ﾎｰﾂｰﾊﾞｵ（HO TU BAO）
北陸先端科学技術大学院大学・知識研究科・教授
研究者番号：６０３０１１９９

(2)研究分担者
高林　克日己（TAKABAYASHI KATSUHIKO）
千葉大学・医学部付属病院・教授
研究者番号：　９０１８８０７９

横須賀　収（YOKOSUKA OSAMU）
千葉大学・医学部・教授
研究者番号：　９０１８２６９１

神田　達郎（KANDA TATSUO）
千葉大学・医学部・特任講師
研究者番号：　２０３４５００２

DAM HIEU CHI（DAM HIEU CHI）
北陸先端科学技術大学院大学・知識研究科・准教授
研究者番号：７０３９７２３０

河崎さおり（KAWASAKI SAORI）
北陸先端科学技術大学院大学・先端領域社会人教育院・特任准教授
研究者番号：４０３７７４３７

(3)連携研究者
なし