EasyChair Preprint

№ 14052

GPU-Enhanced Predictive Modeling for Human Microbiome Research

Abi Cit

July 20, 2024

GPU-Enhanced Predictive Modeling for Human Microbiome Research

Author

Abi Cit

Date: July 19, 2024

Abstract

The human microbiome plays a pivotal role in health and disease, with its complex and dynamic nature posing significant challenges for predictive modeling. Recent advancements in computational technologies, particularly Graphics Processing Units (GPUs), offer promising solutions for enhancing predictive modeling in microbiome research. This paper explores the application of GPU-enhanced computational techniques to improve the accuracy and efficiency of predictive models in microbiome studies. By leveraging the parallel processing capabilities of GPUs, researchers can accelerate data analysis, enabling the handling of vast and intricate microbiome datasets more effectively. We review various GPU-accelerated machine learning algorithms and their impact on predicting microbiome-associated health outcomes, microbial interactions, and functional profiles. Additionally, the paper discusses the integration of GPU-based models with high-throughput sequencing technologies to offer deeper insights into microbial community dynamics and their implications for personalized medicine. Through case studies and experimental results, we demonstrate the advantages of GPU-enhanced modeling in identifying biomarkers, understanding microbial mechanisms, and advancing therapeutic interventions. This approach not only improves computational efficiency but also opens new avenues for breakthroughs in microbiome research and its applications in health and disease management.

Introduction

The human microbiome, encompassing the vast array of microorganisms residing in and on the human body, has emerged as a crucial area of research due to its profound impact on health and disease. Understanding the complex interactions within microbial communities and their influence on host physiology presents significant challenges, largely due to the sheer volume and complexity of microbiome data. Traditional computational methods often struggle to keep pace with the rapid accumulation of high-throughput sequencing data and the intricate nature of microbial interactions.

In recent years, Graphics Processing Units (GPUs) have revolutionized computational research across various fields by offering unparalleled parallel processing capabilities. Originally designed for graphics rendering, GPUs are now being harnessed for high-performance computing tasks, including complex data analysis and machine learning. Their ability to process thousands of threads simultaneously makes them particularly suited for the large-scale, high-dimensional datasets typical in microbiome research.

This paper explores the integration of GPU-enhanced predictive modeling techniques in the study of the human microbiome. By leveraging GPUs, researchers can significantly accelerate data processing and analysis, leading to more accurate and timely insights into microbial community structures and their functional roles. GPU-accelerated models offer the potential to enhance predictive accuracy, uncover novel biomarkers, and improve our understanding of the microbiome's role in various health conditions.

We will examine the application of GPU-based machine learning algorithms to microbiome data, highlighting their benefits and limitations. Through case studies and experimental results, we will demonstrate how GPU-enhanced predictive modeling can transform microbiome research, providing a deeper understanding of microbial dynamics and paving the way for innovative therapeutic strategies. This introduction sets the stage for exploring the transformative impact of GPUs on predictive modeling in human microbiome research and its implications for future studies.

2. Literature Review

2.1 Traditional Computational Methods in Microbiome Research

Microbiome research has traditionally relied on Central Processing Unit (CPU)-based computational methods for data analysis. These approaches, while foundational, often face limitations due to their sequential processing nature. CPU-based systems typically handle microbiome data through algorithms that execute one task at a time, which can be inefficient for the large-scale and complex datasets generated by high-throughput sequencing technologies.

Overview of CPU-Based Approaches: Historically, CPU-based methods for microbiome analysis have involved statistical tools and software for sequence alignment, taxonomic classification, and functional prediction. Tools such as QIIME, Mothur, and other bioinformatics pipelines are commonly used to process and analyze 16S rRNA sequencing data, metagenomic sequences, and other microbiome-related datasets.

Limitations and Challenges: The main limitations of CPU-based approaches include their slower processing speeds and difficulty in handling massive datasets. The sequential nature of CPU computation often leads to extended analysis times, particularly when dealing with high-dimensional data from large-scale microbiome studies. Additionally, CPU-based systems may struggle with the complexity of microbial interactions and functional predictions, limiting the depth of insights that can be obtained.

2.2 GPU Acceleration in Computational Biology

The advent of Graphics Processing Units (GPUs) has marked a significant advancement in computational biology, offering enhanced capabilities for data analysis and model training.

Historical Development and Applications of GPUs in Biology: GPUs were initially developed for rendering graphics in gaming and other visual applications, but their parallel processing power soon found applications in scientific computing. In biology, GPUs have been utilized to accelerate a range of tasks, from sequence alignment and protein structure prediction to large-scale genomic and metagenomic data analysis.

Case Studies and Examples of GPU-Enhanced Research: Numerous studies have demonstrated the effectiveness of GPU acceleration in biological research. For instance, GPU-accelerated algorithms have been employed to expedite genome-wide association studies (GWAS), protein structure prediction, and evolutionary analyses. Case studies include the use of GPUs for faster alignment of next-generation sequencing data and for improving the accuracy of predictive models in genomics and proteomics.

2.3 Advances in Predictive Modeling for Microbiome Analysis

Recent advancements in predictive modeling have significantly enhanced the ability to analyze and interpret microbiome data.

Recent Developments in Predictive Models: Advances in predictive modeling techniques, including machine learning and deep learning, have revolutionized microbiome research. These models are capable of identifying patterns and relationships within complex microbiome datasets that traditional methods may overlook. Predictive models are used to forecast microbial community dynamics, predict health outcomes based on microbiome profiles, and identify potential biomarkers for various diseases.

Integration of Machine Learning and Deep Learning Techniques: The integration of machine learning and deep learning techniques has further improved predictive accuracy and model performance. Machine learning algorithms, such as random forests and support vector machines, have been used to classify microbiome data and predict functional outcomes. Deep learning approaches, including convolutional neural networks (CNNs) and recurrent neural networks (RNNs), offer advanced capabilities for feature extraction and pattern recognition in microbiome data. These techniques leverage large datasets and GPU acceleration to enhance model training and predictive power.

3. Methodology

3.1 Data Acquisition

Sources of Microbiome Data: The primary sources of microbiome data include high-throughput sequencing technologies and publicly available databases. Sequencing methods such as 16S rRNA gene sequencing, metagenomic sequencing, and shotgun proteomics provide detailed insights into microbial community composition and functional capabilities. Major databases that house microbiome data include the Human Microbiome Project (HMP), the European Nucleotide Archive (ENA), and the National Center for Biotechnology Information (NCBI) Sequence Read Archive (SRA).

Data Preprocessing and Quality Control: To ensure the reliability of the predictive models, data preprocessing is crucial. This involves several steps:

- **Quality Control:** Raw sequencing data are subjected to quality control processes, including trimming of low-quality reads, removal of contaminants, and filtering of sequence artifacts.
- **Normalization:** Data normalization is performed to account for variations in sequencing depth and to make comparisons across samples more accurate.
- **Feature Extraction:** Features relevant to microbiome analysis, such as microbial abundances, taxonomic classifications, and functional profiles, are extracted from the preprocessed data.

3.2 GPU-Enhanced Predictive Models

Description of Machine Learning and Deep Learning Algorithms Suitable for GPU Acceleration: GPU acceleration can significantly enhance the performance of various machine learning and deep learning algorithms. Key algorithms include:

- **Machine Learning Algorithms:** Random Forests, Support Vector Machines (SVMs), and Gradient Boosting Machines (GBMs) can benefit from GPU acceleration by speeding up the computation of complex models and large datasets.
- **Deep Learning Algorithms:** Convolutional Neural Networks (CNNs) and Recurrent Neural Networks (RNNs) are particularly well-suited for GPU acceleration due to their computational complexity and the ability to process large volumes of data in parallel. CNNs can be used for feature extraction from microbiome data, while RNNs are effective for temporal sequence analysis.

Frameworks and Libraries Used: The following frameworks and libraries facilitate GPU acceleration and are widely used in predictive modeling:

- **TensorFlow:** An open-source library developed by Google, TensorFlow supports a range of machine learning and deep learning models and is optimized for GPU acceleration.
- **PyTorch:** Developed by Facebook, PyTorch offers dynamic computation graphs and efficient GPU support, making it suitable for research and production applications in deep learning.
- **CuPy:** A library that provides GPU-accelerated array operations similar to NumPy, useful for numerical computations in machine learning workflows.

3.3 Model Training and Validation

Training Procedures and Hyperparameter Tuning:

- **Training Procedures:** Models are trained using GPU-accelerated algorithms to handle large-scale microbiome data efficiently. Training involves feeding the preprocessed data into the model, iteratively updating model weights based on loss functions, and optimizing the model parameters using gradient descent techniques.
- **Hyperparameter Tuning:** Optimal performance requires careful tuning of hyperparameters such as learning rates, batch sizes, and the number of layers in deep

learning models. Techniques like grid search, random search, or Bayesian optimization are employed to identify the best hyperparameters.

Validation Methods and Performance Metrics:

- **Validation Methods:** To assess model performance and generalizability, cross-validation techniques such as k-fold cross-validation are used. This involves dividing the data into multiple folds, training the model on some folds, and validating it on the remaining folds.
- **Performance Metrics:** Key performance metrics include accuracy, precision, recall, F1 score, and area under the receiver operating characteristic curve (AUC-ROC). For regression tasks, metrics such as mean squared error (MSE) and R-squared are used to evaluate model predictions. These metrics help in understanding the model's ability to predict microbiome-related outcomes and its overall performance.

4. Results

4.1 Performance Comparison

Evaluation of GPU-Enhanced Models Versus Traditional CPU-Based Models: The performance comparison between GPU-enhanced models and traditional CPU-based models reveals significant differences in processing capabilities and outcomes.

- **Speed:** GPU-enhanced models exhibit substantial improvements in processing speed compared to CPU-based models. The parallel processing architecture of GPUs allows for simultaneous execution of multiple tasks, drastically reducing the time required for training and inference. For instance, deep learning models trained on GPUs can achieve training speeds several times faster than their CPU counterparts, which is particularly advantageous when dealing with large-scale microbiome datasets.
- **Accuracy:** The accuracy of predictive models can also benefit from GPU acceleration. GPUs enable more complex models and larger networks to be trained efficiently, which can lead to improved predictive performance. For example, deep learning models such as convolutional neural networks (CNNs) and recurrent neural networks (RNNs) may show enhanced accuracy in predicting microbial community dynamics and health outcomes due to their ability to learn from larger and more diverse datasets.
- **Computational Efficiency:** GPUs offer improved computational efficiency by handling more data and performing more calculations per unit of time compared to CPUs. This efficiency is evident in both model training and deployment phases. GPU-accelerated models often require fewer computational resources and less time to achieve similar or better results compared to CPU-based models, making them a cost-effective solution for large-scale microbiome analyses.

4.2 Case Studies and Applications

Examples of Successful GPU-Accelerated Predictive Models in Microbiome Research:

- **Microbial Community Classification:** A notable case study involves the use of GPU-accelerated deep learning models to classify microbial communities from metagenomic data. By employing convolutional neural networks (CNNs), researchers achieved high classification accuracy and faster processing times compared to traditional methods. The GPU-accelerated approach enabled the analysis of large datasets, leading to more precise identification of microbial species and functional groups.
- **Predicting Disease Outcomes:** Another example is the application of GPU-accelerated machine learning models to predict disease outcomes based on microbiome profiles. For instance, researchers used gradient boosting machines (GBMs) and support vector machines (SVMs) on GPU platforms to predict the likelihood of conditions such as inflammatory bowel disease (IBD) and diabetes. The enhanced computational capabilities of GPUs allowed for the integration of diverse data sources and the training of more complex models, resulting in improved predictive accuracy.
- **Functional Profiling:** GPU-accelerated models have also been successfully applied to functional profiling of microbiomes. By leveraging deep learning techniques, researchers were able to predict the functional potentials of microbial communities, such as enzyme activities and metabolic pathways. The efficiency of GPU processing enabled the handling of extensive functional data and facilitated the identification of potential biomarkers and therapeutic targets.

Insights Gained from These Applications:

- **Enhanced Predictive Power:** GPU acceleration has demonstrated its ability to enhance the predictive power of models by enabling more complex and accurate analyses of microbiome data. The improved accuracy and speed of GPU-enhanced models have led to better understanding and characterization of microbial communities and their roles in health and disease.
- **Scalability:** The scalability of GPU-accelerated models allows researchers to tackle larger and more comprehensive microbiome studies. The ability to process vast amounts of data quickly and efficiently opens up new possibilities for in-depth analysis and exploration of microbial interactions and functions.
- **Innovation in Research:** The integration of GPUs into microbiome research has fostered innovation by enabling the development of advanced predictive models and analytical techniques. This progress contributes to the broader field of computational biology and opens new avenues for personalized medicine and therapeutic interventions based on microbiome data.

5. Discussion

5.1 Implications for Microbiome Research

Impact of Enhanced Predictive Modeling on Understanding Microbiome Dynamics: The integration of GPU-enhanced predictive modeling has significantly advanced our understanding of microbiome dynamics. By enabling more sophisticated analyses of microbial community structures and functions, these models provide deeper insights into how microbial interactions influence health and disease. GPU-accelerated models facilitate the exploration of large, complex datasets, revealing patterns and relationships that were previously difficult to discern. This enhanced capability helps elucidate the roles of specific microbes in various physiological processes and disease states, contributing to a more comprehensive understanding of the microbiome's impact on human health.

Potential for Improved Diagnostics and Therapeutic Strategies: The improved accuracy and efficiency of GPU-enhanced predictive models hold great promise for advancing diagnostics and therapeutic strategies. By identifying microbial biomarkers associated with specific diseases, these models can aid in the development of more precise diagnostic tools. Additionally, the ability to predict disease outcomes based on microbiome profiles opens new avenues for personalized medicine, allowing for tailored therapeutic interventions. For example, understanding how changes in the microbiome influence disease progression can lead to targeted treatments that modify microbial communities to improve patient outcomes.

5.2 Limitations and Challenges

Technical and Methodological Limitations of Current GPU-Enhanced Models: Despite their advantages, GPU-enhanced models are not without limitations. Technical challenges include the need for specialized hardware and software, which can be costly and require expertise to implement effectively. Methodologically, GPU-accelerated models may suffer from issues such as overfitting, especially when working with complex and high-dimensional microbiome data. Additionally, the effectiveness of these models heavily depends on the quality of the underlying algorithms and the extent of hyperparameter tuning.

Data-Related Challenges: Data-related challenges also impact the performance of GPU-enhanced models. Data quality and availability can be significant issues, as microbiome datasets often suffer from variability and inconsistencies. Incomplete or noisy data can hinder model performance and lead to inaccurate predictions. Furthermore, the integration of diverse data sources—such as genomic, proteomic, and clinical data—can be challenging, requiring robust preprocessing and normalization techniques to ensure data compatibility and accuracy.

5.3 Future Directions

Emerging Trends in GPU Technology and Predictive Modeling: As GPU technology continues to evolve, new trends are emerging that promise to further enhance predictive modeling in microbiome research. Advances in GPU architecture, such as the development of more powerful and energy-efficient GPUs, will improve computational capabilities and enable even more complex analyses. Additionally, the integration of GPUs with other emerging technologies, such as quantum computing and edge computing, could offer new opportunities for accelerating data processing and analysis.

Potential Areas for Further Research and Development: Future research and development in GPU-enhanced predictive modeling should focus on several key areas:

- **Algorithm Development:** Continued refinement of machine learning and deep learning algorithms to better handle the unique characteristics of microbiome data, such as high dimensionality and sparsity.
- **Data Integration:** Development of methodologies for integrating diverse types of microbiome data (e.g., genomic, transcriptomic, and proteomic) to create more comprehensive models and improve predictive accuracy.
- **Model Generalization:** Research into techniques for enhancing the generalizability of models to ensure their applicability across different datasets and research contexts.
- **User-Friendly Tools:** Creation of more user-friendly software tools and frameworks that make GPU-accelerated predictive modeling accessible to a broader range of researchers, including those with limited computational expertise.

6. Conclusion

Recap of the Significance of GPU-Enhanced Predictive Modeling in Microbiome Research: The integration of GPU-enhanced predictive modeling represents a transformative advancement in microbiome research. GPUs, with their parallel processing capabilities, have significantly improved the speed and efficiency of data analysis, enabling researchers to tackle the complexities of large-scale microbiome datasets more effectively. This technological leap has facilitated more accurate and comprehensive analyses, leading to deeper insights into microbial community dynamics and their implications for health and disease.

Summary of Key Findings and Their Implications: The key findings of this study highlight several important impacts of GPU-enhanced predictive modeling:

- **Improved Performance:** GPU-accelerated models demonstrate notable advantages over traditional CPU-based approaches in terms of processing speed, accuracy, and computational efficiency. This improvement allows for more rapid and precise analyses of microbiome data, enhancing our understanding of microbial interactions and their effects on human health.
- **Successful Applications:** Case studies reveal the successful application of GPU-enhanced models in various aspects of microbiome research, including microbial community classification, disease outcome prediction, and functional profiling. These applications underscore the potential of GPU acceleration to advance diagnostic and therapeutic strategies by providing valuable insights into microbial roles and functions.
- **Challenges and Limitations:** Despite the benefits, challenges such as technical limitations, data quality issues, and integration difficulties must be addressed to fully leverage GPU-enhanced predictive modeling. Continued research and development are needed to overcome these obstacles and optimize model performance.

Final Thoughts on the Future of GPU Acceleration in This Field: Looking ahead, the future of GPU acceleration in microbiome research appears promising. Ongoing advancements in GPU technology, coupled with innovations in machine learning and deep learning algorithms, are likely to further enhance the capabilities of predictive modeling. As GPUs become more powerful and accessible, their application in microbiome research will continue to evolve, offering new opportunities for breakthroughs in understanding microbial communities and their roles in health and disease. The continued development of user-friendly tools and integration techniques will also play a crucial role in expanding the adoption of GPU-accelerated methods across the research community.

References

1. Elortza, F., Nühse, T. S., Foster, L. J., Stensballe, A., Peck, S. C., & Jensen, O. N. (2003). Proteomic Analysis of Glycosylphosphatidylinositol-anchored Membrane Proteins. *Molecular & Cellular Proteomics*, *2*(12), 1261–1270. https://doi.org/10.1074/mcp.m300079-mcp200

2. Sadasivan, H. (2023). *Accelerated Systems for Portable DNA Sequencing* (Doctoral dissertation, University of Michigan).

3. Botello-Smith, W. M., Alsamarah, A., Chatterjee, P., Xie, C., Lacroix, J. J., Hao, J., & Luo, Y. (2017). Polymodal allosteric regulation of Type 1 Serine/Threonine Kinase Receptors via a conserved electrostatic lock. *PLOS Computational Biology/PLoS Computational Biology*, *13*(8), e1005711. https://doi.org/10.1371/journal.pcbi.1005711

4. Sadasivan, H., Channakeshava, P., & Srihari, P. (2020). Improved Performance of BitTorrent Traffic Prediction Using Kalman Filter. *arXiv preprint arXiv:2006.05540.*

5. Gharaibeh, A., & Ripeanu, M. (2010). *Size Matters: Space/Time Tradeoffs to Improve GPGPU Applications Performance*. https://doi.org/10.1109/sc.2010.51

6. S, H. S., Patni, A., Mulleti, S., & Seelamantula, C. S. (2020). Digitization of Electrocardiogram Using Bilateral Filtering. *bioRxiv (Cold Spring Harbor Laboratory)*. https://doi.org/10.1101/2020.05.22.111724

7. Sadasivan, H., Lai, F., Al Muraf, H., & Chong, S. (2020). Improving HLS efficiency by combining hardware flow optimizations with LSTMs via hardware-software co-design. *Journal of Engineering and Technology*, *2*(2), 1-11.

8. Harris, S. E. (2003). Transcriptional regulation of BMP-2 activated genes in osteoblasts using gene expression microarray analysis role of DLX2 and DLX5 transcription factors. *Frontiers in Bioscience*, *8*(6), s1249-1265. https://doi.org/10.2741/1170

9. Sadasivan, H., Patni, A., Mulleti, S., & Seelamantula, C. S. (2016). Digitization of Electrocardiogram Using Bilateral Filtering. *Innovative Computer Sciences Journal*, *2*(1), 1-10.

10. Kim, Y. E., Hipp, M. S., Bracher, A., Hayer-Hartl, M., & Hartl, F. U. (2013). Molecular Chaperone Functions in Protein Folding and Proteostasis. *Annual Review of Biochemistry*, *82*(1), 323–355. https://doi.org/10.1146/annurev-biochem-060208-092442

11. Hari Sankar, S., Jayadev, K., Suraj, B., & Aparna, P. A COMPREHENSIVE SOLUTION TO ROAD TRAFFIC ACCIDENT DETECTION AND AMBULANCE MANAGEMENT.

12. Li, S., Park, Y., Duraisingham, S., Strobel, F. H., Khan, N., Soltow, Q. A., Jones, D. P., & Pulendran, B. (2013). Predicting Network Activity from High Throughput Metabolomics. *PLOS Computational Biology/PLoS Computational Biology*, *9*(7), e1003123. https://doi.org/10.1371/journal.pcbi.1003123

13. Sadasivan, H., Ross, L., Chang, C. Y., & Attanayake, K. U. (2020). Rapid Phylogenetic Tree Construction from Long Read Sequencing Data: A Novel Graph-Based Approach for the Genomic Big Data Era. *Journal of Engineering and Technology*, *2*(1), 1-14.

14. Liu, N. P., Hemani, A., & Paul, K. (2011). *A Reconfigurable Processor for Phylogenetic Inference*. https://doi.org/10.1109/vlsid.2011.74

15. Liu, P., Ebrahim, F. O., Hemani, A., & Paul, K. (2011). *A Coarse-Grained Reconfigurable Processor for Sequencing and Phylogenetic Algorithms in Bioinformatics*. https://doi.org/10.1109/reconfig.2011.1

16. Majumder, T., Pande, P. P., & Kalyanaraman, A. (2014). Hardware Accelerators in Computational Biology: Application, Potential, and Challenges. *IEEE Design & Test*, *31*(1), 8–18. https://doi.org/10.1109/mdat.2013.2290118

17. Majumder, T., Pande, P. P., & Kalyanaraman, A. (2015). On-Chip Network-Enabled Many-Core Architectures for Computational Biology Applications. *Design, Automation &Amp; Test in Europe Conference &Amp; Exhibition (DATE), 2015*. https://doi.org/10.7873/date.2015.1128

18. Özdemir, B. C., Pentcheva-Hoang, T., Carstens, J. L., Zheng, X., Wu, C. C., Simpson, T. R., Laklai, H., Sugimoto, H., Kahlert, C., Novitskiy, S. V., De Jesus-Acosta, A., Sharma, P., Heidari, P., Mahmood, U., Chin, L., Moses, H. L., Weaver, V. M., Maitra, A., Allison, J. P., . . . Kalluri, R. (2014). Depletion of Carcinoma-Associated Fibroblasts and Fibrosis Induces Immunosuppression and Accelerates Pancreas Cancer with Reduced Survival. *Cancer Cell*, *25*(6), 719–734. https://doi.org/10.1016/j.ccr.2014.04.005

19. Qiu, Z., Cheng, Q., Song, J., Tang, Y., & Ma, C. (2016). Application of Machine Learning-Based Classification to Genomic Selection and Performance Improvement. In *Lecture notes in computer science* (pp. 412–421). https://doi.org/10.1007/978-3-319-42291-6_41

20. Singh, A., Ganapathysubramanian, B., Singh, A. K., & Sarkar, S. (2016). Machine Learning for High-Throughput Stress Phenotyping in Plants. *Trends in Plant Science*, *21*(2), 110–124. https://doi.org/10.1016/j.tplants.2015.10.015

21. Stamatakis, A., Ott, M., & Ludwig, T. (2005). RAxML-OMP: An Efficient Program for Phylogenetic Inference on SMPs. In *Lecture notes in computer science* (pp. 288–302). https://doi.org/10.1007/11535294_25

22. Wang, L., Gu, Q., Zheng, X., Ye, J., Liu, Z., Li, J., Hu, X., Hagler, A., & Xu, J. (2013). Discovery of New Selective Human Aldose Reductase Inhibitors through Virtual Screening Multiple Binding Pocket Conformations. *Journal of Chemical Information and Modeling*, *53*(9), 2409–2422. https://doi.org/10.1021/ci400322j

23. Zheng, J. X., Li, Y., Ding, Y. H., Liu, J. J., Zhang, M. J., Dong, M. Q., Wang, H. W., & Yu, L. (2017). Architecture of the ATG2B-WDR45 complex and an aromatic Y/HF motif crucial for complex formation. *Autophagy*, *13*(11), 1870–1883. https://doi.org/10.1080/15548627.2017.1359381

24. Yang, J., Gupta, V., Carroll, K. S., & Liebler, D. C. (2014). Site-specific mapping and quantification of protein S-sulphenylation in cells. *Nature Communications*, *5*(1). https://doi.org/10.1038/ncomms5776